

Bangor University

DOCTOR OF PHILOSOPHY

Explorative coastal oceanographic visual analytics : oceans of data

George, Richard

Award date:
2013

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 26. Apr. 2024

Explorative Coastal Oceanographic Visual Analytics: Oceans of Data

Author: Richard Llwyd Simon Frost George

School of Computer Science, Bangor University
August, 2013

Supervisor: Jonathan C. Roberts

Acknowledgements

I would like to thank the many people who have supported me throughout the past four years. At times it has been challenging, and their help has been greatly appreciated in enabling to reach this point.

Firstly, my supervisor, Dr Jonathan Roberts has been an invaluable source of advice, guidance and encouragement throughout my studies, particularly when I began to doubt I would ever finish, and has challenged me to extend my knowledge and understanding into areas of visualization I had never anticipated pursuing. Thanks to him, I now understand more about statistical analysis than I ever thought would be possible (and I am sure my 'A' level Maths teacher would appreciate that).

I would also like to thank Professor Alan G Davies and Dr Peter Robins of the Centre for Applied Marine Studies at Bangor University, for providing the datasets for my project; for being willing guinea pigs for my theories of how to improve on the analytical capabilities of their models and for providing me with a challenging 'real life' problem: I hope my proposed solution will reduce some of their calculating drudgery.

Finally, I would not have arrived at the completion of this research without family and friends - thanks James, Lukas, Matt, Steve and Tom for putting up with me being grumpy :). My father, a few years ago, acquired for me extremely old oceanographic text book, and as I flicked over the pages, my eye was caught by what seemed to be some very early examples of visual analytics. So, in conjunction with my supervisor, I have him to thank for some of the inspiration for this project. My partner, Sophie, has been extremely supportive, despite working to complete her own studies, and I thank her for not abandoning me to a life chained to a computer. I would like to say a massive thank you to my mother, for proof reading my dyslexic sprawl. She has always been a shining star in my life and I have her to thank for always encouraging me.

Thank you all

Richard

Statement of originality & release

Statement of Originality

The work presented in this dissertation is entirely from the studies of the individual student, except where otherwise stated. Where derivations are presented and the origin of the work is either wholly or in part from other sources, then full reference is given to the original author. This work has not been presented previously for any degree, nor is it at present under consideration by any other degree awarding body.

Student:

.....

Statement of Availability

I hereby acknowledge the availability of any part of this dissertation for viewing, photocopying or incorporation into future studies, providing that full reference is given to the origins of any information contained herein.

Student:

.....

Company Representative (if applicable)

.....

Abstract

The widely acknowledged challenge to data analysis and understanding, resulting from the exponential increase in volumes of data generated by increasingly complex modelling and sampling systems, is a problem experienced by many researchers, including ocean scientists. The thesis explores a visualization and visual analytics solution for predictive studies of coastal shelf and estuarine modelled, hydrodynamics undertaken to understand sea level rise, as a contribution to wider climate change studies, and to underpin coastal zone planning, flood prevention and extreme event management.

But these studies are complex and require numerous simulations of estuarine hydrodynamics, generating extremely large datasets of multi-field data. This type of data is acknowledged as difficult to visualize and analyse, as its numerous attributes present significant computational challenges, and ideally require a wide range of approaches to provide the necessary insight. These challenges are not easily overcome with the current visualization and analysis methodologies employed by coastal shelf hydrodynamic researchers, who use several software systems to generate graphs, each taking considerable time to operate, thus it is difficult to explore different scenarios and explore the data interactively and visually.

The thesis, therefore, develops novel visualization and visual analytics techniques to help researchers overcome the limitations of existing methods (for example in understanding key tidal components); analyse data in a timely manner and explore different scenarios. There were a number of challenges to this: the size of the data, resulting in lengthy computing time, also many data values becoming plotted on one pixel (overplotting).

The thesis presents: (1) a new visualization framework (VINCA) using caching and hierarchical aggregation techniques to make the data more interactive, plus explorative, coordinated multiple views, to enable the scientists to explore the data. (2) A novel estuarine transect profiler and flux tool, which provides instantaneous flux calculations across an estuary. Measures of flux are of great significance in oceanographic studies, yet are notoriously difficult and time consuming to calculate with the commonly used tools. This derived data is added back into the database for further investigation and analysis. (3) New views, including a novel, dynamic, spatially aggregated Parallel Coordinate Plots (Sa-PCP), are developed to provide different perspectives of the spatial, time dependent data, also methodologies for develop-

ing high-quality (journal ready) output from the visualization tool. Finally, (4) the dissertation explored the use of hierarchical data-structures and caching techniques to enable fast analysis on a desktop computer and to overcome the overplotting challenge for this data.

Contents

1	Introduction	1
1.1	The Grand Scientific Challenges	5
1.2	Role of Sci Vis	7
1.2.1	Visual Analytics	8
1.3	The Oceanographic Domain	9
1.3.1	Overview	9
1.3.2	Coastal shelf studies and modelling	10
1.3.3	Visualization and visual analytics challenges	11
1.4	Collaboration with OS	12
1.5	Hypothesis	12
1.6	Contributions of the thesis	14
1.6.1	Publications	17
1.7	Structure	18
2	The Data	21
2.1	The Coastal Domain	21
2.2	The Estuarine Case Studies	24
2.3	Data Sampling or Simulation	27
2.4	Numerical Models and Current Visualization Approaches	29
2.5	TELEMAC-2D	31
2.6	Graphical interfaces for TELEMAC	33
2.7	The case study – Vis challenges	34
2.7.1	An important hydrodynamic component - Tidal Flux	37
2.8	Summary	38
3	OS Visualization	39
3.1	Review of vis tools	40
3.2	OS Vis Systems	43
3.3	Ocean Visualization	44

3.4	Coastal shelf visualization	45
3.5	Visualization for EDSS	47
3.6	Ocean Observatories	50
3.7	Workflow, including data management, system design and architecture	51
3.7.1	Recent developments	52
3.7.2	Workflow systems	53
3.7.3	Workflow systems in ocean sciences	54
3.8	Summary	55
4	Related Work	57
4.1	In Collaboration OS with CS...	58
4.1.1	Understanding the Coastal Shelf Researcher	60
4.2	Oceanographic Visualization	64
4.2.1	Scope of the Oceanographic Related Work	65
4.2.2	Multi-field oceanographic data & the challenges they present to visualization	66
4.2.3	The value of visualization and visual analytics for ocean sci- ence data	71
4.3	Large data management	77
4.3.1	Overview of data management methodologies	77
4.3.2	Defining the problem and identifying solutions	80
4.3.3	Spatial Indexing	82
4.4	Interaction & Interfaces	90
4.4.1	Query Driven Visualization	91
4.4.2	Interaction	93
4.4.3	Current approaches to interfaces and interaction in ocean sciences	94
4.5	Coordinated Multiple Views	96
4.5.1	The value of Coordinated Multiple Views (CMV) for multi- field data	99
4.5.2	Views	99
4.5.3	Use of Coordinated Multiple Views in ocean sciences and the wider geosciences	102
4.6	Visual Analytic Tools	106
4.6.1	Parallel Coordinate Plot	107
4.6.2	Other visual analysis techniques	108

4.6.3	Bespoke problem solving tools	110
4.6.4	Resolving spatial clutter for PCP and similar clustering techniques	113
4.6.5	Axis/dimension reordering	119
4.6.6	Data analysis tools in the geosciences	120
4.7	Towards requirements of an investigative coastal data visual analytics system	122
4.7.1	Context	122
4.7.2	Requirements	124
4.8	Summary	125
5	VINCA	129
5.1	Establishing system requirements	130
5.1.1	Identified Challenges	132
5.1.2	Summary of requirements	136
5.2	Design and Initial Implementation	136
5.2.1	Selection of programming tool	137
5.2.2	Early concept and design work	138
5.3	Data Import	139
5.3.1	Stage 1 – Data-Import	140
5.3.2	Stage 2 – Database loading	142
5.4	Spatial indexing	143
5.4.1	Selection of spatial indexing	144
5.4.2	Development and testing of PR quadtree and k-d tree approaches	146
5.4.3	Uses of PR quadtree algorithm	149
5.5	Interaction and Interfaces	149
5.5.1	GUI Elements	150
5.5.2	Data Selection within the views	151
5.6	Coordinated Multiple Views – CMV	152
5.6.1	Developing the windows for CMV	153
5.6.2	Development of CMV management tool: Hierarchical CMV Manager.	154
5.7	Views	156
5.7.1	(Object Abstraction) Abstract View Class	158
5.7.2	Control Views	159

5.7.3	Context Views	161
5.7.4	Overlays	163
5.7.5	Attribute plots and graphs	165
5.7.6	PCP Query View	168
5.7.7	Transect/Flux View	170
5.8	Improving System Performance	170
5.8.1	Caching	170
5.8.2	QuickDraw	171
5.8.3	Export of publishable quality outputs.	172
5.9	Summary	172
6	PCP Query	174
6.1	PCPs & hydrodynamic research	175
6.2	PCP Architecture	176
6.2.1	Processes common to the PCP variants:	177
6.3	Raw Data PCP	180
6.3.1	Visualizing the Raw Data PCP	181
6.3.2	Query Results	182
6.4	Frequency PCP	182
6.4.1	Frequency PCP	183
6.4.2	Development of a variant frequency PCP	183
6.4.3	Spatially Indexed Binning	184
6.4.4	Evaluation of the variant frequency PCP	184
6.5	Sa-PCP (Spatially aggregated PCP)	185
6.5.1	Utilising Spatial Indexing to pre-process the data	186
6.5.2	Construction of the Sa-PCP	187
6.5.3	Visualizing the Sa-PCP from the range data	188
6.5.4	Display Ranges	188
6.5.5	Display using Envelopes	189
6.5.6	Transfer Functions	190
6.5.7	Associated Heat Map	190
6.6	Evaluation of the techniques	190
6.7	Summary	191
7	Deriving New Data	195
7.1	Transects	196

7.2	Flux	196
7.2.1	Definition	198
7.2.2	How it is currently calculated	199
7.2.3	Flux in other work	200
7.3	Developing the tool	200
7.3.1	Transect Profiler	201
7.3.2	Flux Calculator	203
7.4	Flux along an estuary	205
7.5	Transect and Flux Vis	209
7.5.1	Scaling Challenge	211
7.5.2	Positioning Challenge	211
7.6	Evaluation of Techniques	212
7.7	Summary	213
8	Conclusions	215
8.1	User Survey	215
8.1.1	User Survey Results	216
8.1.2	Discussion of the Results	218
8.2	Review of the Hypotheses	219
8.3	Summary of achievements	220
8.3.1	Achievement 1: Development and implementation of a coastal oceanographic Visual Analytics framework	221
8.3.2	Achievement 2: production of derived data supporting more effective data analysis	224
8.3.3	Achievement 3: Implementation of a visual analytics sys- tem based on multiple views	225
8.3.4	Achievement 4: Data Structures	226
8.4	Future Work	228
8.4.1	Future Work – Short Term	229
8.4.2	Future Work – Medium Term	230
8.4.3	Future Work – Long Term	231
8.5	Conclusions	233
	Bibliography	234
	Appendix	261

List of Figures

1.1	Direct reproduction from Maury [196] “ Plate VI. illustrates the position of the channel of the Gulf Stream for summer and winter. The diagram A shows a thermometrical profile presented by cross-sections of the Gulf Stream, according to observations made by the hydrographical parties of the United States Coast Survey.”	2
1.2	Direct reproduction from Maury [196] “ Plate XII. exhibiting a vertical section of the Atlantic, and showing the contrasts of its bottom with the sea-level in a line from Mexico across Yucatan, Cuba, San Domingo, and the Cape de Verds, to the Coast of Africa.”	3
1.3	Direct reproduction from Maury [196] “ Plate V. is a section taken from one of the manuscript charts at the Observatory. It illustrates the method adopted there for co-ordinating for the Pilot Charts the winds as reported in the abstract logs. For this purpose the ocean is divided into convenient sections, usually five degrees of latitude by five degrees of longitude. These parallelograms are then subdivided into a system of engraved squares, the months of the year being the ordinates, and the point of the compass being the abscissae. As the wind is reported by a vessel that passes through any part of the parallelogram, so it is assumed to have been at that time all over the parallelogram. From such investigation as this the Pilot Charts are constructed.”	4
2.1	Oceanographic domain showing the area of interest – coastal shelf and estuaries.	22
2.2	Relationship between tidal flow and sediment transport. Direct copy from Davies and Thorne [67]	23

2.3	Figure showing the location of Dyfi Estuary and Burry Inlet. These are the test sites for datasets that are used in this dissertation. The data is simulated using TELEMAC-2D, with real-world data (sampled from these two locations) to become the boundary conditions for the simulations	25
2.4	The Dyfi estuary. The diagram shows the bathymetry from a model simulation by Telemac and visualized with VINCA. Several features are annotated on the picture, including the embankment, salt marsh, channel and railway embankment.	26
2.5	The Figure depicts the bathymetry of the Burry Inlet, with the bathymetry shown in colour. Various features can be seen, from the Railway line and Port in the North, to the Tidal flats and Salt Marshes in the South position.	27
2.6	TELEMAC-2D grid for the Dyfi Estuary, showing difference in resolution for area of greater interest. The offshore resolution is 500m, gradually reducing to 50m in the estuary and 25m in the river channels. Direct copy from Robins [243].	32
2.7	Screenshots from Blue Kenue TM that depict several visualizations that are typical from the use of this software.	34
2.8	Table of variables loaded into VINCA	35
2.9	Burry Estuary visualization (generated by VINCA) showing scale and complexity of simulated data of the hydrodynamic flow. Eight time-steps are shown. The timesteps are four hours apart, cover the M2-S2 tidal cycle and the colours represent the magnitude of the hydrodynamic flow.	36
3.1	copied from Stein et al. [277]: multivariate data visualization of salinity, isosurface of dissolved oxygen and flow vectors.	46
3.2	Copied from Zhang et al. [326], multiple level of detail visualization of tsunami simulation, placed into global context.	49
4.1	Snapshot of Macquarie Ridge region, undersea canyon between Tasmania and Victoria Australia, produced for display not for data analysis. The VRML model and visualization allowed scientists to display the connectivity and continuation of various features to support jurisdiction negotiations. Reproduced from Watford et al. [312] . . .	59

4.2	Reproduced from He and Hamblin [121]: multislice plot of modelled currents (vectors) and temperatures (colour contours) at 08:54h, 24 July 1996 with illumination	62
4.3	Copied from Trouw et al. [291] : Vertical suspended sediment concentration. field during wave cycle (asymmetric wave)	62
4.4	Data representation (using model post-processor and analysis software) reproduced from Paradis et al. [218], of a TELEMAC-2D storm surge model for the Gironde Estuary.	63
4.5	Representation of comparisons of glacial inundation simulations, showing the difficulty of undertaking a detailed comparison of the outcomes, and understanding the variances, without advanced analytics. Reproduced from Alho and Aaltonen [4]	64
4.6	copied from [13] showing the difference between linear and cyclical spiral representation of cyclical data, to illustrate a weekly pattern. .	74
4.7	Howe's solution to the challenges presented by the unstructured grids of finite element models. The gridfield algebra is implemented as modules, grouped by kind and visualized using VTK. Copied from [136].	84
4.8	Representation of a quadtree, copied from Popenet et al. [228], showing, on the left hand side, a geometrical description, and on the right hand side a logical description, with numbers indicating the level of the corresponding cell.	87
4.9	Direct copy from Samet [248] showing a PR quadtree and its records.	88
4.10	k-d tree partitioning, direct copy from Sinha et al [269].	89
4.11	Reproduced from Blower et al. [40]: GUI for selecting model output from a consortium member, annotated by selection region.	97
4.12	copied from Doleisch et al. [76]: Coordinated Multiple View of diesel exhaust emissions	98
4.13	Reproduced from Blass et al. [38]: CMV main user interface of application to study meteorological data. The top part of the screen contains spatial viewing components (two slice viewers and a 3D isosurface view). The parallel coordinate plot is positioned in the middle, and the control interface is positioned at the bottom.	100
4.14	Reproduced from Ware et al. [311]. CMV with focus+context: 3D visualization showing the path of data sampling to acquire high resolution data in relation to bathymetry	104

4.15	Reproduced from Ho et al. [135]: linked coordinate multiple view interface for 4D oceanographic data visual analytics, comprising 3D context view for providing an overview, together with 2D image focus and attribute views. The isosurface embedded in the 3D view represents water temperature changes in relation to depth.	105
4.16	Reproduced from Stein et al. [277]: visualization of transport flux data in the Chesapeake Bay.	112
4.17	Reproduced from Howe et al. [136]: salt flux visualization in the Columbia Estuary	113
4.18	Reproduced from [274]: Dynamic axis scaling: (a) before, (b) after. Moving the mouse wheel in the focus area of the axis moves the values of the upper and lower limits closer together, facilitating zooming into the central axis.	116
4.19	Copied from [230]. This shows PCP with S-shaped axis derived from polar graph, to encode wind direction.	120
4.20	copied from Kehrner et al. [161]: a) scatterplot showing the percentage of total outliers (x axis), and measure to determine how outliers are distributed (y axis)	121
5.1	Prototype 1(P1)	131
5.2	Prototype 2 (P2)	131
5.3	Dyfi estuary data, showing data association problem	132
5.4	shows the model mesh for the Dyfi Estuary and river channel – ranging from resolution of 500m offshore , reducing to 50m in the estuary and 25m in the river channel, reproduced from Robins [243] . . .	133
5.5	prototype 3D VTK visualization of Dyfi Estuary TELEMAC-2D data with water level at full opacity, showing problem of coincident topology	133
5.6	Reproduced from Robins [243]. Static plots of velocity vectors generated using Blue Kenue.	135
5.7	comparative screen shots of ebb and peak flow velocity, generated using Blue Kenue, reproduced from Robins [243]	135
5.8	Reproduced from Silva and Freire’s [265]; schematic of visual analytics and visualization pipeline, illustrates the basis upon which the pipeline for this system has been developed.	138

5.9	The schematic demonstrates the processes involved in the Data Im- porter Pipeline	141
5.10	Data Loading Pipeline	143
5.11	TELEMAC-2D finite element mesh for the Burry Inlet, reproduced from Robins [242] The resolution in the river channels (the dark- est areas) is 25m, increasing to 50m in the outer estuary and 100m offshore.	144
5.12	PR Quadtree applied to Dyfi Estuary TELEMAC-2D dataset	147
5.13	Build times in milli secs for PR Quadtree and k-d trees with 26693 and 51183 points.. . . .	148
5.14	Number of leaf nodes for PR quadtree and k-d trees with 26693 and 51183 points.	148
5.15	Distribution of leaf node aspect ratios for k-d trees with 26693 points and 51182 points: the dark band, which denotes aspect ratios pro- viding regular shaped nodes, encompasses fewer nodes in the larger dataset.	148
5.16	PR Quadtree (right) and k-d tree (left) for 26693 points.	149
5.17	Reproduced from Pike et al. [226], schematic illustrating the user goals and tasks facilitated by interactive visual analytics.	150
5.18	Simple List visualization of points contained within leaf node of spatial data structure.	151
5.19	zoomed in view of PR Quadtree cell, showing properties of the cell.	153
5.20	Model-View-Controller type window manager used for VINCA. . .	154
5.21	Example of simple list management tool.	155
5.22	Outline of the Hierarchical CMV Manger.	156
5.23	VINCA	157
5.24	Table of views within VINCA.	157
5.25	Schematic of a generic View	158
5.26	Colour Mapper in Simple Mode	160
5.27	Colour Mapper in CM mode.	160
5.28	Main view showing the bathymetry of the Burry data set	162
5.29	Dual view to allow side by side comparisons, the partition on the left depicts the bathymetry, the partition on the right depicts the speed of flow.	162
5.30	Example of the spatially indexed heat map overlay used within the system.	163

5.31	Output from the dynamic vector filter using the quadtree. View a shows no filtering, then progressing through b and c to heavy filtering at view d.	164
5.32	Point Graph View, with the 1D partition on the left, the temporal vector partition in the middle and the vector rose plot on the right. .	166
5.33	Output from VINCA showing the tidal harmonics with normalised scaling. The blue line is the water elevation and the red is the speed of flow.	166
5.34	demonstrates the selection method used in the polar partitions	167
6.1	Flow chart diagram of PCP architecture	176
6.2	a point on the mesh, related to its PR quadtree polyline	180
6.3	Picture illustrating the RAW PCP still blown out even though each polyline has only a transparency value of 1 (out of 255).	181
6.4	Because of the multi-resolution grid, it is unsuitable to utilize a fixed bin to create a frequency-bin of the data. Although the visual depiction of the parallel coordinate plot looks appropriate (as depicted left); when the data is plotted to an associated heat-map the sampling problems are clearly shown (right).	183
6.5	Spatially indexing binning	184
6.6	Spatial Bin to Regular bin	184
6.7	Visualization of the frequency PCP variant using the quadtree to provide binning for the spatial dimension (the two leftmost in the figure)	185
6.8	outlines the flow and generation of the Sa-PCP	186
6.9	This figure represents a very high level of aggregation.	187
6.10	This figure represents a very low level of aggregation.	187
6.11	This shows the envelopes using the full range of data contained at each node in the PR quadtree at an aggregation level X using a depth based sort.	189
6.12	This shows the envelopes using 1SD either side of mean at each node in the PR quadtree at an aggregation level X using a depth based sort	189
6.13	Visual depiction of the transfer functions used.	190
6.14	This shows a visual depiction of the tubes transfer function without depth sorting	191

6.15	This shows a visual depiction of the tubes transfer function with non pairwise depth sorting.	191
6.16	The figures show different levels of aggregation of the tree. In the left hand column is the PCP while the right column shows the plot window. Both are linked together. The user can interactively change the level of aggregation to explore. While it may not be useful to display the tree at such low levels, the figures demonstrate how the user can choose the level of aggregation. These views depict a single time step, however the user can step or animate through time to see how the data changes over time.	194
7.1	The schematic represents a typical transect across an estuary (left) with points that are regularly sampled to interpolate the data across this transect (right)	197
7.2	Multiple fluxes from various locations in the Dyfi estuary. Reproduced from Robins [243]	198
7.3	The Transect is divided into n cross sectional trapezoids	203
7.4	Illustrates the conversion from point based velocities to cell based perpendicular velocity.	204
7.5	Illustrates the calculation of the perpendicular (normal) component of the segment's velocity.	204
7.6	the skeleton as defined by the user.	207
7.7	initial rectangular boundaries placed around each of the piece-wise segments of the skeleton.	207
7.8	the joining of the two boundary ends to form continuous region. . .	208
7.9	T showing the trapezoid boundary.	208
7.10	boundary after user manipulation	208
7.11	the interpolated points that will form the starting and ending points for the interpolated transects.	209
7.12	the interpolated transects the algorithm will generate.	209
7.13	Diagram (i) Shows the Burry estuary with the region transect/flux before calculation. Diagram (ii) demonstrates the result from the algorithm, with all the transect positions calculated. It shows the regular distribution of the generated transects.	209
7.14	shows the publishable output from VINCA of a simple transect profile and associated flux curve.	210

7.15 shows the region based transect / flux view in VINCA.	210
--	-----

CHAPTER 1

Introduction

Preamble This thesis could be said to originate, in part, from a gift from my father: a few years ago, well aware of my interest in anything to do with physical oceanography and also my firmly held views on the value of more visualization and fewer words, my father gave me a ‘vintage’ oceanography book, he discovered in a dusty second hand book shop. Imagine my surprise when I opened the book and one of the first plates to leap out of it was an 1860 hand drawn representation of the Gulf Stream using what can only be described as an extremely early version of a Line Integral Convolution – LIC (Figure 1.1), followed by a representation of a transect of the North Atlantic Ocean (Figure 1.2), and a frequency representation (tally chart) of the winds used for creating Pilot Charts (Figure 1.3). The book, “The Physical Geography of the Sea and its Meteorology”, by Captain M F Maury [196], which is over 150 years old, seemed almost contemporary in some of the visual analysis techniques it was using to convey information and understanding of complex concepts for large volumes of sampled data. I reflected that if oceanographers were applying such techniques in the middle of the 19th century to gain insight into their data, using painstakingly sampled data, pen and ink, we should be capable of producing a visual analytics system for coastal shelf and estuarine research, with the wealth of knowledge of visualization at our disposal. In fact, Collins [64], in ‘Data Visualization has it all been seen before?’ concluded that the major data visualization techniques had all been developed before the advent of computers, albeit for real (sampled) data. Whilst accepting that such a tool today requires represen-

tation of vastly more data, what was being achieved then was comparative. Thus, it would be following in a long visual tradition to explore and try to identify advanced visual means to enable coastal and estuarine scientists to analyse their data more efficiently and effectively. Out of that, this thesis was born.

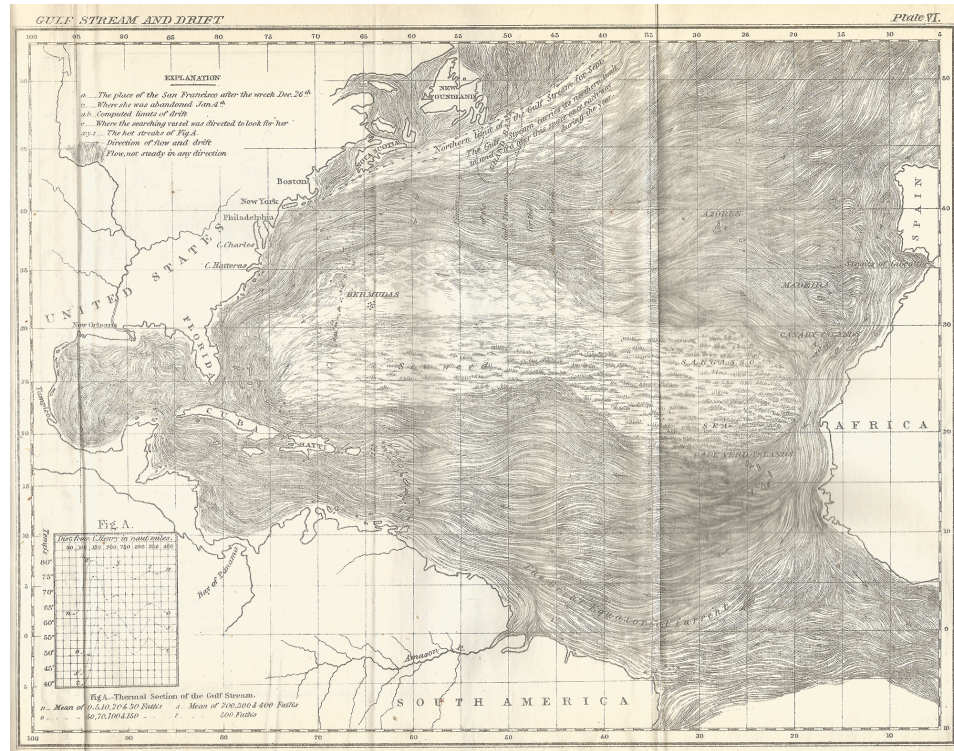


Figure 1.1: Direct reproduction from Maury [196] “**Plate VI.** illustrates the position of the channel of the Gulf Stream for summer and winter. The diagram A shows a thermometrical profile presented by cross-sections of the Gulf Stream, according to observations made by the hydrographical parties of the United States Coast Survey.”

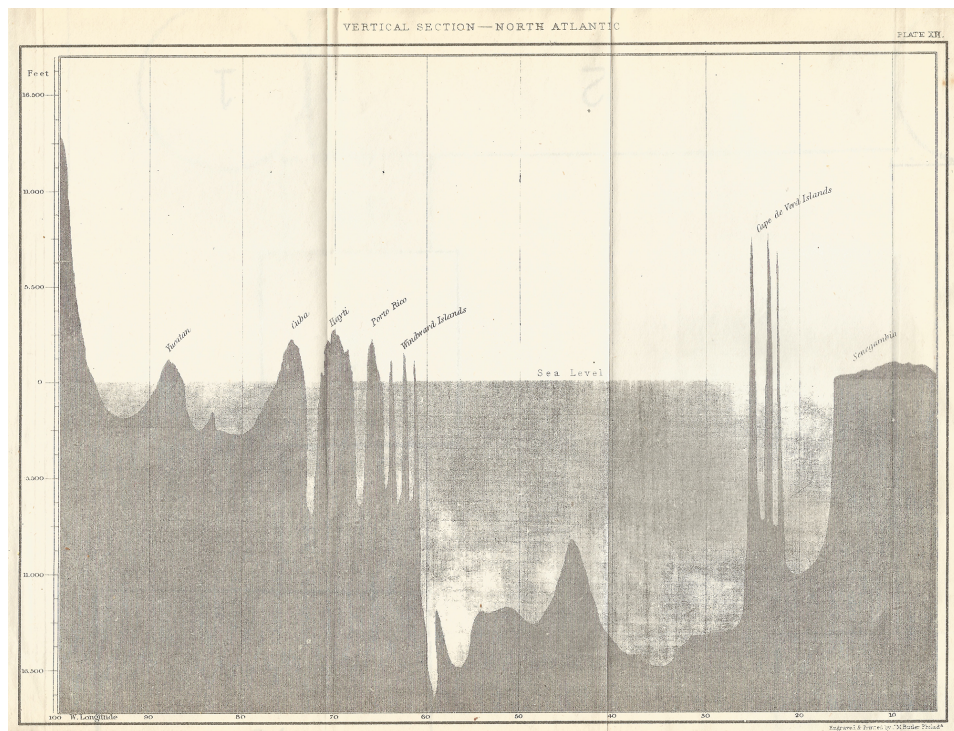


Figure 1.2: Direct reproduction from Maury [196] “**Plate XII.** exhibiting a vertical section of the Atlantic, and showing the contrasts of its bottom with the sea-level in a line from Mexico across Yucatan, Cuba, San Domingo, and the Cape de Verdes, to the Coast of Africa.”

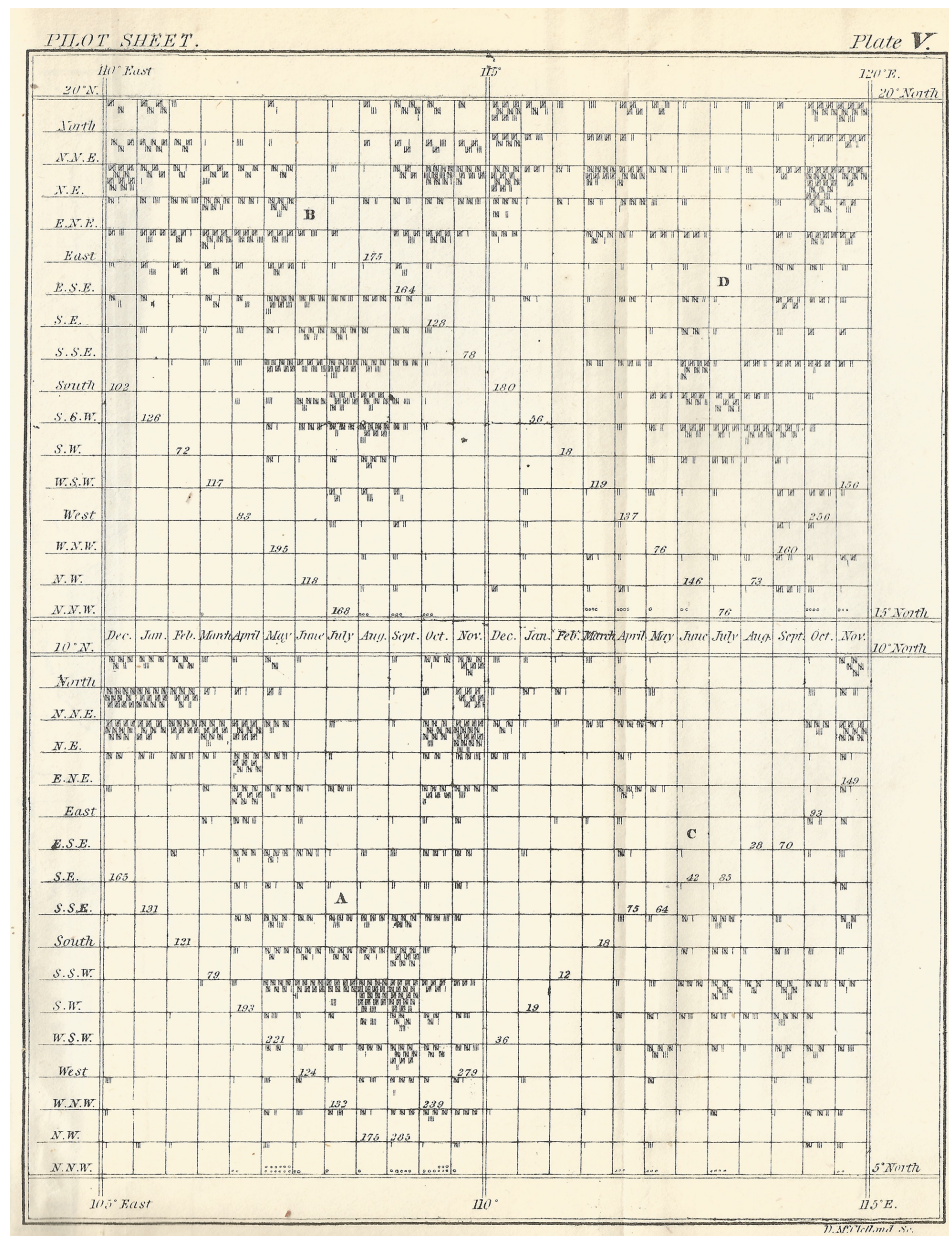


Figure 1.3: Direct reproduction from Maury [196] “**Plate V.** is a section taken from one of the manuscript charts at the Observatory. It illustrates the method adopted there for co-ordinating for the Pilot Charts the winds as reported in the abstract logs. For this purpose the ocean is divided into convenient sections, usually five degrees of latitude by five degrees of longitude. These parallelograms are then subdivided into a system of engraved squares, the months of the year being the ordinates, and the point of the compass being the abscissae. As the wind is reported by a vessel that passes through any part of the parallelogram, so it is assumed to have been at that time all over the parallelogram. From such investigation as this the Pilot Charts are constructed.”

1.1 A Wider Context: The Grand Scientific Challenges

In a transformational period for scientific researchers, computer simulation and analysis is replacing traditional ways of exploring scientific phenomena. The evolution of scientific studies over many millennia has relied on dependable data to enable theories to be developed and tested, to extend knowledge and to support predictions. This thirst for data has driven continuous improvements in the methods of taking and recording data – from the humble wax or stone tablets used by ancient scribes, to the classical scientists notebooks, and now to the modern digital age of the computer. Each advancement enhances the level of detail of experimentation and study, but in turn generates ever increasing quantities of data.

The scale of the analytical problem is discussed in *The Fourth Paradigm* [114, 130]: gigantic projects such as CERN’s Large Hadron Collider ¹, or the Pan-STARRS camera array of telescopes with four cameras each generating 1.4 billion pixels ² are each capable of creating Petabytes of data per day [27]. Similar challenges face university researchers, whose day-to-day research consistently generates Gigabyte datasets. Discussing geoscience modelling, Erlebacher et al. [87] described the exponential increase in data volumes as ‘data flooding’, leading to ‘information overload’ [169] Thus, whilst data intensive projects bring huge opportunities, they also bring challenges to existing methods of study, as conventional analytical methods have limited effectiveness [26, 286]. The struggle to handle the amount of information available for analysis threatens the clarity and insight which improved data production should support. Paradoxically, more data leads to less understanding, with data analysis techniques lagging behind the ability to generate data, necessitating new interpretative tools supporting knowledge discovery [85, 130].

Earth Sciences (including Ocean Sciences) exemplify this. Enhanced computing power has led to improved methods of gathering both observed and modelled data (in mathematical models the resolution of calculation lattices has increased, more variables are incorporated and wider time periods are investigated). Advanced simulation tools are capable of generating Terabytes of multi-field data, with complex inter-relationships, and along with understanding one simulation, frequently, there is a need to explore and compare simulations and possibilities. This presents challenges to the scientist to organise the data in a meaningful way; undertake query

¹<http://public.web.cern.ch/public/en/lhc/lhc-en.html>

²<http://pan-starrs.ifa.hawaii.edu/>

and analysis supported by visual tools; document the results, and share and present the results to others. Nonetheless, despite the analytical challenges, models afford benefits, for example, the need to run physical experiments may be obviated.

However, current tools increasingly lack the capability both to analyse the data and visualize it appropriately. In ocean sciences, users employ tools such as MATLAB and Excel to visualize their data, and whilst the results are accurate, the tools do not work well on large datasets, they are not interactive, can often be time-consuming to set up, and rely on specialist knowledge to operate them. Therefore, they do not permit rapid exploration of different scenarios.

Gray [109] reflects on this in his talk on eScience to the NRC-CSTB in Mountain View in 2007:

“Some research communities use MATLAB, for example, but the funding agencies in the U.S. and elsewhere would do a lot more to foster the building of tools to make scientists more productive. When you go and look at what scientists are doing, day in day out, in terms of data analysis, it is truly dreadful. Where essentially the only tools I have at my disposal are MATLAB and Excel!”

To advance from this, scientists will increasingly rely upon new techniques to gain understanding [114,286], and a goal for computer scientists is to provide more interactive and analytical tools to aid domain scientists. Science will be increasingly done in the database, finding relationships among existing data rather than focusing on data collection [282]. This will require the data to be reorganised to clarify and illuminate – to “hide complexity” [40] to reveal spatial and temporal relationships and enable researchers to interrogate data in a timely manner. Thus databases will become an intrinsic element of scientific study – a case argued by Emmott and Risson [85], in “Towards 2020 Science”: integration of computer science concepts, methodology and data analysis into scientific theory, experimentation and analysis will create a golden triangle of data exploration, where novel conceptual and technological tools from computer science will support enhanced scientific problem solving.

In “The Fourth Paradigm” [130], the contributing authors discuss how evolving data exploration practices are leading to a new scientific research paradigm. Gray’s case [109] is that unifying theory, experimentation and simulation into a data intensive pipeline represents a new “Fourth Paradigm” of data exploration (the previous three paradigms being empirical, theoretical and computational). Here, computa-

tional science and visualization may support exploration of bigger datasets; correlation of data to other datasets and collaboration between researchers to improve insight and make new discoveries that might not be possible with traditional ways of working.

1.2 The Role of Scientific Visualization

Scientific research relies on display methodologies both to illustrate and explain insights and conclusions, and to support analysis, understanding and comparison of datasets. For the past 25 years or more, this included computational scientific visualization systems, which proved valuable in integrating disparate datasets; facilitated comparison with other sources of data [321]; allowed researchers to view datasets in their entirety [99, 122]; detected features, and revealed relationships. Hence, scientific visualization played an important role in data exploration, with the discipline rapidly reaching a state of some maturity [189, 282].

However, the 'unprecedented and overwhelming' [152] requirement to support increasingly complex modelling and simulations generating ever larger datasets has challenged models' analytical tools to provide the required insight. Thus, Thomas and Cook, in "Illuminating the Path" [286] conclude that fundamental limits are being approached, requiring a reduction of the scale of the data to a level which can be studied and understood by researchers.

Concurrently, there has been a period of introspection and review of the future of scientific visualization – Johnson's visualization research challenges and Lorensen's 'On the death of visualization' among others [152, 189, 299]. Whilst the important role of scientific visualization in providing insight into data is recognised, there is a sense that future advances may be incremental, and may not involve the step changes of earlier years. Although Johnson [152] and Ertl [88] consider the development of a theory of visualization an important goal, many researchers conclude that visualization is not necessarily an end in itself, but an important element of its applications. Scientific visualization has to adapt to new models and simulations, so a more useful approach may be to focus on user needs in this data intensive era [299]. Many visualization techniques are now highly advanced, but quite often well-established techniques are not usable with complex applications, such as the extremely large, domain specific, multi-field datasets generated by, for example, ocean science research.

However, any advanced visualization system for complex applications must pro-

vide added value, for example by incorporating tools permitting accurate calculations and analysis not possible with more conventional means. They must also be capable of evaluation through user testing [88, 189, 299], with usability and evaluation regarded as a major challenge for the future development of visualization [180]. There is, therefore, a strong case for a more collaborative, user-focused approach between scientific researchers (the users) and the computer scientists developing visualization and analytic systems for advanced data exploration, supporting the predictive research necessary for resolving some of the major challenges facing scientists today. Such studies might also, through the challenging visualization questions they pose, lead to further insight into core visualization research and enable new application areas to be developed.

1.2.1 Visual Analytics

The above debate is particularly relevant for study domains incorporating extremely large, multi-field data, where analytically important detail may be buried within the mass of data [286]. Analysing this data is challenging but valuable, as it supports detailed exploration and reduces uncertainty, as large data may provide more complete information, thus there is a strong case for systems to exploit and understand data hidden within datasets [165]. However Earth and Ocean science research is such a domain, where the size and complexity of the data necessitates:

- functions to filter and reduce the amount of data.
- software and display scalability.
- multi-resolution representation of the data.

These needs may be met by visual analytics, defined by Thomas and Cook [286] as:

“the science of analytical reasoning facilitated by interactive visual analytics ... enabling information to be synthesised and insight to be derived.”

Hansen et al. [114] considers visual analytics to be an integral part of science discovery, enabling:

“..the detection and validation of expected results while also enabling unexpected discoveries in science. It allows for the validation of new

theoretical models, provides comparison between models and datasets, enables quantitative and qualitative querying, improves interpretation of data, and facilitates decision making. Scientists can use visual data analysis systems to explore what if scenarios, define hypotheses, and examine data using multiple perspectives and assumptions”.

To achieve those outcomes, visual analytics system developers will need to work with users to:

- identify their needs and problems in studying specific aspects of their data
- suggest new ways of investigating the domain data
- develop appropriate, interactive novel tools to facilitate analysis
- synthesise tools with visualization techniques to transform the data
- develop representations of the data at appropriate levels of complexity for differing audiences.

In meeting these requirements, an interactive, investigative environment is created, where visual analytics may enable information overload to be transformed into opportunity [165].

1.3 The oceanographic domain & its data challenges

1.3.1 Overview

Through its dynamic nature and the size and complexity of its data, Oceanography suggests itself as a useful domain for studying aspects of data exploration in a data intensive world. Its multi-scale geological, biological, chemical and physical processes require data analysis over extremely lengthy periods, produce gigabytes of data for each time step and involve simultaneous measurement of multi-disciplinary, dynamic processes over areas ranging from vast ocean expanses to small estuarine and river flows. Methods of collecting data include physical and satellite observation systems, but also numerical modelling. Comprehensive observation systems require huge resources, but because of the vast domain, they cannot be all encompassing, so modelling has become important in simulating real physical processes [238].

Furthermore, Ocean Science presents its own grand challenges: Brandt [44] links the goals of long term ocean research with methods to ameliorate human

predicaments, and with applied research to manage ocean resources, thereby identifying a number of significant priorities. These include, among others, understanding, modelling and predicting climate; understanding and predicting changes to ocean ecosystems; understanding and predicting natural catastrophes and understanding the long-term effects of events such as storm surges, hurricanes and flooding.

1.3.2 Coastal shelf studies and modelling

Climate change is a major scientific concern of our era, and studying the oceans' physical mechanisms is a key to understanding the changes and planning for the future. Studies of the coastal shelf are important because of the impact of climate change on the coastal margin. Fifty percent of the world's population lives within 60kms of the shoreline [67], and the Intergovernmental Panel on Climate Change, in its 2007 report, projected a global rise in sea levels of $18 - 59\text{cms}$ between 1990 and 2090³. In the area of study, Wales – DEFRA (the Department for Environment, Food and Rural Affairs) [91] predicted sea level rise exceeding 10mm pa., in the second half of the twenty first century. The socio-economic impact of sea level rise on coastal zones worldwide is well documented, and estuaries and associated coastal shelves present a combination of attributes requiring careful management, including protected eco-systems, centres of population, and industries essential to the economic well-being of the area, such as leisure, tourism, fishing etc.

Coastal shelf studies often use simulation and predictive modelling, as historic rates of change to sea level may not be reliable guides to predicting changes in the future [48, 206], although agreement between sea level rise models has been poor, and there are significant regional variations because of non-uniform future ocean warming. Nonetheless, models are crucial to understanding how sea level changes affect the near shore region, to support flood prevention and extreme event management; dredging; port operations, and major engineering projects, such as offshore wind farms.

Thus, taking into account the demands of studying large data, its complexity in this domain and the need to resolve a significant real world problem, modelled coastal and estuarine studies presents an appropriately challenging area of study on which to test the hypotheses of this thesis. The many hydrodynamic variables include fresh river water interaction with saline water to produce strong baroclinic

³www.ipcc.ch

currents; shallow water depth generating non-linear tidal movements; the impact on water circulation of a typically complex topography, plus human intervention (e.g., coastal defences) [239].

Furthermore, coastal shelf models are often more complex, unpredictable and detailed than deep ocean models [181, 321], producing ever larger datasets, with many interconnecting factors. The models tend to be high resolution and cover only part of the coast although recently, one of the ocean science collaborators on this project has developed modelled simulations extending well out into the Irish Sea, which will generate even larger, and more complex datasets ⁴. Thus, within this context, researchers exploring, comparing and contrasting different simulations and parameterisations of the data must integrate visualized data with analytical methods and store these intermediate results and investigations in a database. Such data intensive discovery processes add to the visualization and analytical challenges by exerting even greater computational and visual computing pressures.

1.3.3 Visualization and visual analytics challenges

Whilst visualization techniques are commonplace in ocean science studies, research demonstrates that current processes are ad hoc and piecemeal. Szalay and Gray [282] acknowledges that few visualization systems are able to visualize data across multiple scales and numerous data sets, similarly Keim et al. [165] considers this a challenging research area in relation to the scale and uncertainty of the data: data are often incomplete, occur at different scales and are interpolated or based on differing assumptions. Furthermore, many visualizations are static plots of transects or sample points from the entire dataset, making it difficult to obtain an overview of the dataset or to explore and contrast different datasets (methods for visualizing the data often do not seem to have a high priority with model developers [299]). This constrains comparative investigations, and there is a lack of visual analytics tools within the coastal shelf and estuarine domain.

New techniques and tools (of which visual data analysis is one example), are, therefore, needed to bring clarity and perception into this data intensive world. Advanced visual analytics techniques have been recognised as valuable for studies of large volumes of multi-field data [167]. Hence, by collaborating with ocean scientists developing estuarine flooding scenarios, which study the diverse relationships between flow hydrodynamics, sediment transport, bottom morphology and coastal

⁴private conversation with Dr P Robins, Summer 2011

erosion, the goal is to create and test a data exploration system. This will incorporate intuitive, domain-based, visual analytic interfaces, which deal with the data generated at an appropriate speed, and which support complex querying. The hypothesis is that these novel visual analytics will support a greater level of knowledge discovery and inference, to enable the researcher to converge on a particular outcome and seek a specific answer, as well as permitting analysis to support comprehensive investigation of plausible alternatives.

1.4 Collaboration with Ocean Scientists

The thesis investigates an oceanographic problem from the perspective of a computer scientist, bridging the divide between these two domains by utilising novel visual analytic techniques for a real world problem. The work has been undertaken in collaboration with the Centre for Applied Marine Sciences (CAMS), Bangor University, using case studies based on coastal and estuarine regions in South and West Wales, which are increasingly prone to flooding. CAMS researchers, in conjunction with the Countryside Council for Wales, have studied a number of estuaries to model and analyse the potential for flooding, consequent upon generally accepted views of climate change and sea level rise ⁵ [78, 231] .

Professor Alan G Davies and Dr Peter Robins of CAMS kindly provided datasets from their flooding scenarios for the system development. Discussions with these end-users enriched the motivations, helped identify the challenges in studying the data, and supported the development of an appropriate visualization and visual analytics system. They have given a particular focus to the identification and development of tools, provided a case-study for implementing and testing the tool, and given direct input over the evaluation of the created technologies, algorithms and tools.

1.5 Hypothesis

Initially, **limitations** in current display and analytical practices were identified which drive the motivation for this thesis:

1. Coastal shelf researchers utilise several different visualization tools e.g. MATLAB, Blue Kenue, Excel to display information, which can take several min-

⁵www.ipcc.ch

utes to generate a single plot, and several hours to explore and visualize different scenarios and data parameterisations.

2. Scientists wish to perform complex analysis on their data, which currently requires a sequential process involving several tools. For example, they may use MATLAB, R and Excel sequentially to analyse the data, each time storing intermediate results in files. This, too, is time consuming, and the use of several tools is neither convenient nor effective for large datasets, resulting in difficulty in displaying and exploring time, and specifically in creating animations of the visual depictions. Bespoke animations can be made for specific scenarios, but again they take time and effort to create.
3. The limitations of the analytical and visualization process described in 2, hinders the creation of high quality output renderings in a vector format to enable the visualization to be viewed at any resolution.
4. As a consequence of increasing amounts of data, scientists are losing the ability to effectively visualize and analyse the underlying data in a timely and effective manner.

Building on these limitations, the hypothesis has **four propositions**:

1. If it is possible to develop a Visual Analytic tool for estuarine and coastal shelf analysis that integrates several visualization forms, then researchers will be able to perform concurrent, complex analysis and visualization of their data. A conglomerated and multiple-view system would support more rapid discovery and enable complex operations and calculations to be performed on the data, which were previously not possible.
2. If the processed or derived data is made available to the Visual Analytics tool then scientists will be able to be more efficient and less ad hoc in their analysis.
3. If the visual analytic tool is able to generate several different views of the data, then one of the renderings might be an output rendering in a vector format to support viewing and study of the visualization at any resolution.
4. There are many structures and methods in computer science that enable rapid calculations. If these can be integrated within a Visual Analytics tool then

the scientists may be able to analyse their large datasets on a fast desktop computer.

Linked to the four propositions of the hypothesis are the **four aims** of the thesis, which are: to develop a visual analytics tool for Coastal and Estuarine studies; to allow derived (new) data to be made available to the domain scientist for analysis; to generate different views of the data including an output rendering in a vector format, and finally, to use Computer Science methodologies to enable rapid and effective visualization and analysis of the data.

1.6 Contributions of the thesis

The application of novel visual analytics, including computational tools, to large, complex oceanographic datasets permits interactive interrogation of the data and allows more expressive data-exploration tasks to be performed. A collaborative approach with coastal shelf scientists to identify their study needs allows the development of systems and tools, which will add to understanding, and support the predictive capabilities of numerical model simulations.

In defence of this statement, the thesis makes the following contributions:

1. **Researches how visualization and visual analytics might aid understanding of coastal shelf and estuarine flow, and reviews current practices in relation to coastal shelf and estuarine visualization and visual analytics.**

Hydrodynamic flow (tides and currents) is a key to understanding the physical processes affecting the coastal shelf, including estuaries. However, it is extremely complex, as it is influenced by many factors, both hydrodynamic and non-hydrodynamic, and it in turn, influences the evolution of estuaries, the surrounding coastline and their associated eco-systems. But coastal shelf researchers, themselves, acknowledge the limitations of current methods of analysing this domain, which requires studies of these inter-relating factors on many scales.

2. **Collaborates with ocean science researchers to understand their needs.**

Many ocean scientists identify the importance of their research in relation to the implications of climate change, in particular sea level rise. At the same time, they recognise the challenges of interrogating, analysing and understanding their extremely large datasets and the limitations of existing methodologies for predicting future trends and the projected impact of extreme events. Yet it is

argued that coastal shelf and estuarine studies are even more complex than deep ocean studies, with very many more closely interrelated factors at play.

Estuarine studies assume great importance in relation to sea level rise, as the coastal fringes of estuaries often contain large populations. For this reason, understanding the physical processes within the estuary and the increased potential for flooding are necessary for planning and management of the area.

A review identified a number of visualization and visual analysis challenges in supporting the exploration and understanding of these extremely large, complex, multi-field data sets. Whilst the problems relating to some aspects of the visualization pipeline have been quite widely addressed, there were few studies identified for the application of visual analytics techniques relating to coastal shelf studies, and even fewer applications for estuarine research.

The need for researchers to be able to drill down into these large and complex data, together with the limited availability of visual analytics systems for modelled estuarine hydrodynamics, led to its selection as an appropriate area of study.

3. **Proposes a number of new approaches to enable the development of a novel, interactive, data querying and visual analytics system for coastal shelf and estuarine studies.**

The system permits coastal shelf and estuarine researchers to be able to explore, query, interactively investigate and quantitatively compare different runs and parameterisations of simulations of estuarine hydrodynamic flow; enables detailed study of key elements of the data, not easily accomplished by traditional methods, and supports the predictive requirement for their research.

The approach demonstrates how interactive visual analysis aids detailed exploration of flooding scenarios, to help identify areas which will be prone to flooding under certain circumstances, and thus supports the development of coastal flood management processes.

4. **Proposes a linked Coordinated Multiple View Interface (CMV), to permit complex querying of ocean science data.**

The system examines and compares differing scenarios of multi-field, structured and unstructured data types, to establish the impact of differing variables. Thus, the interface provides a multiple linked view interface (Hierarchical-Model-View), based on the Model-View-Controller pattern; several plot windows; temporal

graphs and parallel coordinate plots (PCPs) to investigate the highly complex data. There are links between views and animation and users are able to easily add new views and select new values for display in another graph.

5. Develops a new rendering technique for Parallel Coordinate Plots (PCPs).

PCP has been shown to be useful for displaying special and large multi-time point data, but there are difficulties due to the size of the data, causing clutter which affects performance and inhibits interactive spatial exploration.

A solution is proposed, that extends the framework by Fua et al. [93]. This algorithm performs a hierarchical parallel coordinate that bins the PCP on the spatial nature of the data, using a data space methodology to aggregate the information in the geo-spatial domain based on the level of detail in the data, rather than the frequency of the data in the PCP. The proposed work creates a binned PCP by developing several associated (linked) data structures (using a quadtree to spatially index the data). Whilst other researchers have explicitly followed Fua et al's. framework, few other systems have used spatial structures to relate the spatial nature of data flow to exploration of the information. The work presented in this thesis uses the PR-quadtrees, linked with a geospatial plot of the data to enable both interaction and exploration of the simulations.

6. Proposes a new tool to enable researchers to undertake the calculation of tidal flux.

The tool demonstrates the ability of visual analytics to provide enhanced analysis and computational facilities, by applying a user defined, path-based algorithm to calculate derived data. A significant challenge for ocean science researchers is investigating simulated data and applying analytic techniques, such as calculating maximum tidal flow, or the limit of potential flooding events, by utilising different metrics and equations to perform specific quantitative analysis.

Flux calculation was selected to prove the effectiveness of the tool. It is considered by ocean science researchers to be one of the most important hydrodynamic calculations, and is essential to support prediction of future trends, but it is a complex calculation not easily accomplished through existing post-processing systems, and requires numerous transects across the estuary.

The proposed tool permits all locations to be calculated and analysed in one step. It is a coordinated, multiple view tool that provides several plot windows and parallel coordinate plots (PCPs) for rapid exploration, (using a frequency

binning approach to deal with the problem of overplotting, based on Novotny and Hauser's [211] algorithm) and temporal graphs. It loads the simulation data; derives diagnostic data calculated from a metric or equation, which is stored in a database, and visualizes the derived data. This enables the user to investigate whole series of transects (where previously they would have selected just a few), and then to focus on supplementary analysis of the derived data.

1.6.1 Publications

There are four publications that have been made from this dissertation: In 2009, a conference paper was published that described comparison between different software programming tools. Part of the background and motivation published in this paper is recorded in this dissertation.

- **R.L.S.F George, J.C. Roberts, "Coastal Shelf Visualization using VTK and OpenDX of Hydro-Informatic Numerical Models", Proc. of Theory and Practice of Computer Graphics 2009, Cardiff, pp223-232, ISBN 978-3-905673-71-5.**

Second, the Processing language has been used as the graphics library to generate the ocean science visualizations. In fact, the developed software discussed in this dissertation developed incrementally. Initially this was named iCove, and later was re-named VINCA. The initial design and development of iCove and how Processing was used to implement some of the parts was published in the TPCG 2010 conference proceedings. This material is included in the design Chapter.

- **R.L.S.F.George, J.C.Roberts and A.G.Davies. "Using Processing to develop iCove: a tool for interactive coastal oceanographic visualization", EG UK Theory and Practice of Computer Graphics (2010), J. Collo-mosse, I. Grimstead (Editors), Eurographics Association, University of Sheffield, September 2010, pp.99-105.**

Third, a poster was presented on VINCA at VisWeek2012 in the LDAV workshop. This is also published as a two-page paper and recorded in the IEEE Digital Library.

- **R. George, P. Robins, A. G. Davies, J.C. Roberts, "Interactive Visual Analytics of Coastal Oceanographic Simulation data", IEEE VisWeek 2012, Seattle, Washington, October 2012. IEEE Computer Press.**

Fourth paper was presented at the EnvirVis Workshop at Euroviz 2013, Leipzig, Germany, June 2013, which discusses the development of a novel tool, as part of the VINCA visual analytics system, for measuring estuary transects and calculating tidal flux. This material is included in Chapter 7, which discusses the VINCA Flux tools. An extended version of the paper has been invited for publication in a topical issue of the ISI journal Environmental Earth Sciences.

- **R. George, P. Robins, A. G. Davies, J.C. Roberts, “Visual Analytics of the Hydrodynamic Flux for Coastal Flooding Prediction and Management”, Workshop on Visualisation in Environmental Sciences (EnvirVis), Euro-Vis 2013 , pp. 1-5. O. Kolditz, K. Rink and G. Scheuermann (Editors)**

1.7 Structure

The thesis is structured as follows:

Chapter 2. The Data & Casestudy locations: The data used for the case studies are described and discussed, and placed in the context of similar research worldwide. The chapter explains the role of modelled and sampled data in the research, and provides an overview of the TELEMAC-2D hydrodynamic model which generated the simulated data used in this thesis. The limitations of the existing display and analysis systems for the model are reviewed. The chapter also considers the nature of the data and the challenges they present to effective visualization and analysis, which informs the development of ideas and concepts for the proposed visualization system. Visual analytics methodologies, tools and techniques are introduced, reflecting on the important aspects of a visual analytics system for coastal shelf and estuarine research and producing a summary of requirements for such a system.

Chapter 3. The Visualization Context: review of practices in ocean science visualization and vis
the current state of evolution of visualization systems is reviewed under a number of headings to support the identification of approaches in developing an appropriate visualization and visual analytics system. This takes into account the needs of coastal shelf researchers, the attributes and challenges presented by the data, and the strengths and limitations of visualization software systems. The influence of climate change studies, including the evolution of large scale observatories and Environmental Decision Support Systems, on

the current state of the art in coastal oceanographic visualization is discussed. The review also considers architecture and data management systems in relation to ocean science visualization and visual analytics.

Chapter 4. Related Work: This research based chapter underpins the thesis and provides the principles upon which the visualization and visual analytics system are developed. It studies how ocean scientists conduct their research and the analytic tools they use; reviews oceanographic visualization practices; identifies challenges and techniques relating to the specific attributes of the data; considers the issues facing researchers with regard to large data management and scalability, and identifies and discusses appropriate visual analytics tools for dealing with the data challenges, whilst meeting the ocean science research goals

Chapter 5. System Overview: The scope of this chapter covers both initial design and prototyping, together with building the system. Initial design thoughts are tested through prototyping, challenges identified and solutions proposed, leading to a summary of system requirements and the identification of an appropriate programming tool. One of the significant contributions of this thesis is discussed: the development of a PR-Quadtree spatial index for scaling the data, which is utilised in different forms within the visual analytics system. The remainder of the chapter is devoted to the process of building the visualization system – the GUI, and the visual analytics system, including the linked Coordinated Multiple Views, the views, plots and graphs. Issues of interaction and interfaces are also discussed.

Chapter 6. Parallel Coordinate Plots – Drilling into the Detail: one of the principle components of the visual analytics system is discussed in detail. Three variants of a PCP are described and tested to assess their ability to explore and interactively visualize the datasets; to provide derived data for further iterative analysis and to identify trends and correlations. Particular focus is placed on the development and implementation of methodologies to deal with the acknowledged issue of overplotting in relation to PCPs and large datasets.

Chapter 7. Deriving New Data – Completing the Visual Analytics Loop: this describes the design and development of a flux calculation tool as a case study to demonstrate the benefits of a collaborative approach with ocean scientists. Their needs were identified, their research methods studied and a custom tool

developed which would significantly improve the speed and effectiveness of their ability to study and calculate a key measure for their predictive capabilities (tidal flux). The aim of this tool in demonstrating the iterative visual analytics paradigm is also discussed, in relation to its ability to provide derived data.

Chapter 8. Discussion & Conclusions: this reflects on all stages of the projects.

The chapter reviews the work in relation to the hypotheses; considers users' evaluation of the system; discusses the main contributions and makes proposals for the development of a best working practice for the analysis of coastal shelf and estuarine modelled data, including the need for close collaboration between domain researchers and computer scientists. The chapter concludes with suggestions for taking the work forwards.

CHAPTER 2

The Data & Case-study locations

The datasets for this dissertation are created from a two year project undertaken by Bangor University's Centre for Applied Marine Sciences (CAMS), in support of coastal zone management, including flood prevention, in two estuarine environments in Mid Wales and South West Wales [242,243]. The data comprises a number of predictive flooding scenarios derived from the TELEMAC-2D numerical model of hydrodynamic flow.

Ocean science researchers wish to investigate and analyse this simulated data with the aim, for example, of calculating maximum tidal flow or the limits of potential flooding events, by applying different metrics and equations over the data. Thus, not only is there a need for visualization of the simulations, but also a requirement to apply algorithms and equations over selected areas of the data to support quantitative analysis. This chapter describes the data, particularly focusing on the data creation and numerical models (sections 2.3 and 2.4), the modelling software (TELEMAC) (sections 2.5 and 2.6) and details and locations where the case-studies take place (section 2.2).

2.1 The Coastal and Estuarine Domain: the context

Ocean science encompasses both the deep ocean and the shallower continental shelf and is a complex, vast, and multi-disciplinary area of study incorporating biolog-

ical, chemical, geological and physical sciences. All these studies rely on an understanding of physical principles, such as hydrodynamics, with the dynamics of coastal shelf waters being more complex than those of the deep ocean.

The area of interest for this work, the coastal zone, encompasses that part of the ocean which has coastal waters less than 200 metres deep, estuaries, and rivers up to the tidal limit – effectively that part of the ocean where physical, biological and geochemical processes are directly affected by the land adjacent to it. Here, there is a dynamic interaction between the tides, currents and waves of the coastal shelf seas and the land, which results in constant change within coastal zones. Figure 2.1 indicates the area of interest for this research project: the coastal shelf and estuaries.

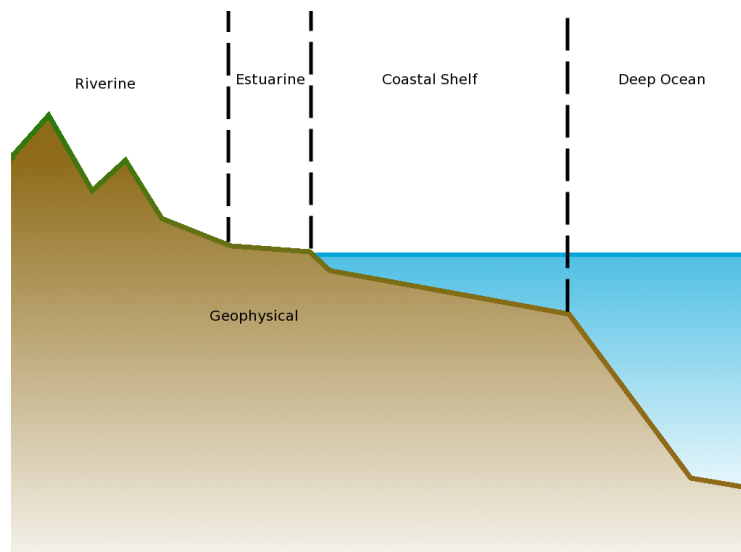


Figure 2.1: Oceanographic domain showing the area of interest – coastal shelf and estuaries.

Estuaries begin at the point where a river becomes tidal (experiences the rise and fall of the tide each day). They are also subject to the other influences common to the coastal shelf. These include bottom friction, where the seabed interacts with the flow of water, through friction. Bottom friction is not a factor in deep ocean studies.

Cameron and Pritchard [52] describe an estuary as a semi-enclosed coastal body of water having a free connection with the open sea. Its water salinity differs from that of the coastal shelf due to interaction (mixing) with freshwater from rivers feeding into the estuary, and the influence of temperature, which may cause evaporation in the shallower estuarine waters. This thermohaline forcing (influence of temperature and salinity) causes stratification of layers of water with different salinity, a

major factor in driving the current (creating baroclinic currents) within the estuary. Thus, in hydrodynamic terms, estuaries are regions with strong baroclinic currents; where shallow water depths generate non-linear tidal perturbations, and where complex topography and man made structures control circulation [239].

Estuaries are frequently surrounded by significant populations and important eco-systems, so understanding their hydrodynamics is critical to successful coastal zone management, particularly in areas prone to flooding. Fundamental to this are the relationships and interplay between current, wave flow and sediment transport. Figure 2.2 shows the interactions between tidal flow/current, sediments and seabed.

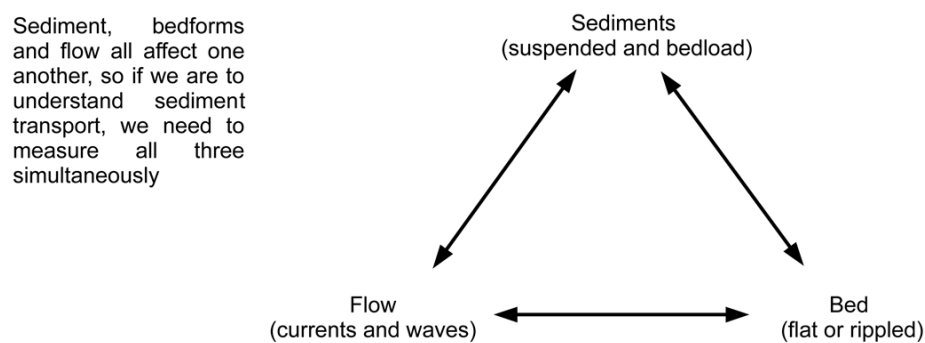


Figure 2.2: Relationship between tidal flow and sediment transport. Direct copy from Davies and Thorne [67]

This interplay means that estuaries are believed to change their characteristics over lengthy periods of time, when they revert from erosion to deposition, through the interaction of tides and sediment transport, and Pethick [225] classified estuaries as:

Type I rapid infill of deep, wide estuaries

Type II inter-tidal flats, and reduced sediment supply

His belief was that estuaries oscillate between these two types, and there is a dynamic equilibrium, where erosion of inter-tidal flats causes a **Type II** estuary to revert to a **Type I**. So, hydrodynamic factors such as tidal asymmetry (causing flood or ebb dominance) together with wave/tide interaction; the spring/neap cycle; frictional effects, and slack water asymmetry all play a role in sediment transport, redistribution and loss. When sea level rise is factored in, this hydrodynamically driven evolution of the estuary may make it more susceptible to extreme events such as storm surges and may lead to flooding [49, 243].

Bearing in mind the socio-economic imperatives, the characteristics common to all estuaries outlined above, and the potential consequences of sea level rise, it is unsurprising that similar work (albeit on differing scales) is being undertaken in estuarine regions worldwide, to deal with real world storm surge and flooding problems. Examples include the Columbia River Coastal Observatory [136, 148]; Chesapeake Bay [277]; river basin flooding [41]; storm disaster assessment [328] and North Carolinas Coastal Hazards Centre dealing with potential flooding and storm surges in the Mississippi Delta [118, 233].

2.2 The Estuarine Case Studies

The two estuarine case studies – the Dyfi Estuary in Cardigan Bay, Mid Wales; and the Burry Inlet (Loughar Estuary) in South West Wales (Figure 2.3) – are of a type which encompass river channels, salt marshes, creeks and sandbanks and demonstrate the estuarine characteristics outlined above. They represent a good fit with the estuarine work referred to above, and although on a smaller scale, are studying the same problems. Both estuaries urgently require management to reduce the flood risk, as these relatively shallow estuaries have been subject to previous heavy management, thus their natural tidal prisms have been reduced by earlier coastal realignment. As a result, coastal defences are increasingly over-topped, because of sea level rise, increased storm surges and high river flow. Furthermore, flood waters drain slowly, which has an impact both on their ecosystems and local infrastructure. Both estuaries have a main ebb dominant channel, flanked by extensive flood dominant tidal flats and salt marshes. The mouths of the estuaries are constricted by the presence of a spit, such that the strongest tidal flow is experienced through the mouth in both cases [241].

Whilst each estuary and its coastal zone presents different management challenges, the research has focused on:

- The impact of coastal defence management strategies (eg realignment, reinforcing embankments).
- Predicting the impact of extreme scenarios on velocity and sediment transport.
- Predicting future salt marsh migration and coastal squeeze.
- Researching sediment transport patterns in estuaries.

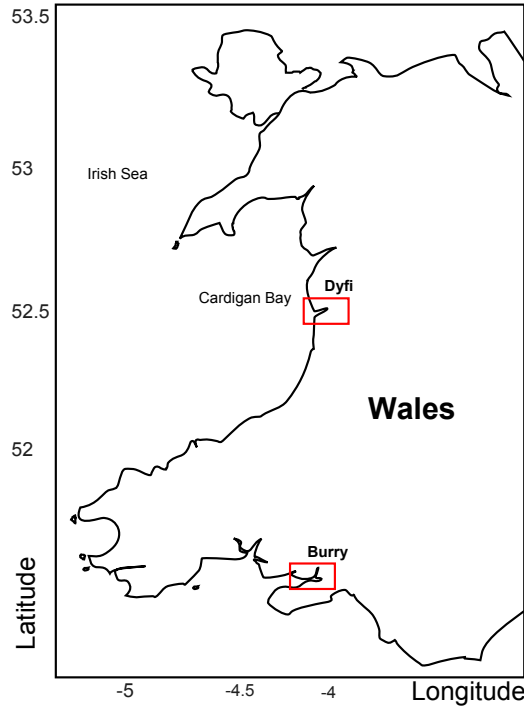


Figure 2.3: Figure showing the location of Dyfi Estuary and Burry Inlet. These are the test sites for datasets that are used in this dissertation. The data is simulated using TELEMAC-2D, with real-world data (sampled from these two locations) to become the boundary conditions for the simulations

Robins and Davies [241], and Robins et al. [239] summarise the main features of each estuary:

The **Dyfi** is a typical macro-tidal sandy estuary, (Figure 2.4 bathymetry of the Dyfi Estuary) with extensive tidal flats and salt-marshes exposed at mid tide, contrasting with the deeper channels, which reach approximately seven metres at a high spring tide, whereas the tidal flats reach a maximum of two metres. It is some eight kms long, and two kms wide, with the width constrained to one km by the Ynyslas Spit at its mouth. This spit is composed of sand, shingle and more stable ships' ballast, dumped there. The waters of the estuary cover an area of approx 17.3km^2 . The two principle rivers which flow into the estuary – Afon Dyfi (mean flow rate $25\text{m}^3/\text{s}$) and Afon Leri (mean flow rate $5\text{m}^3/\text{s}$) – have been incorporated into the simulations and model domain includes all land below 10 metres.

The size of the estuary was reduced considerably by the 19th century construction of the Borth-Dyfi Junction railway embankment, which currently acts as a sea defence for the southerly, low lying areas. Both the railway embankment, and other embankments on the river flood plains (protecting agricultural land), severely re-

strict the natural tidal prism. Furthermore, the embankments contribute to flooding at the town of Machynlleth, when high river flow coincides with high spring tides. Simulations are needed to support decisions on whether to reinforce the embankments to reduce flood risk, or realign them, which would threaten protected marsh ecosystems.

A particularly significant feature is the Borth Bog (Cors Forchno), a National Nature Reserve, SSSI and Special Area of Conservation, and one of the finest examples of a raised peat bog in the UK. However, it is at risk of flooding during extreme tidal and rainfall events, and its existence might be threatened by sea level rise, especially if flooding events become more frequent [239].

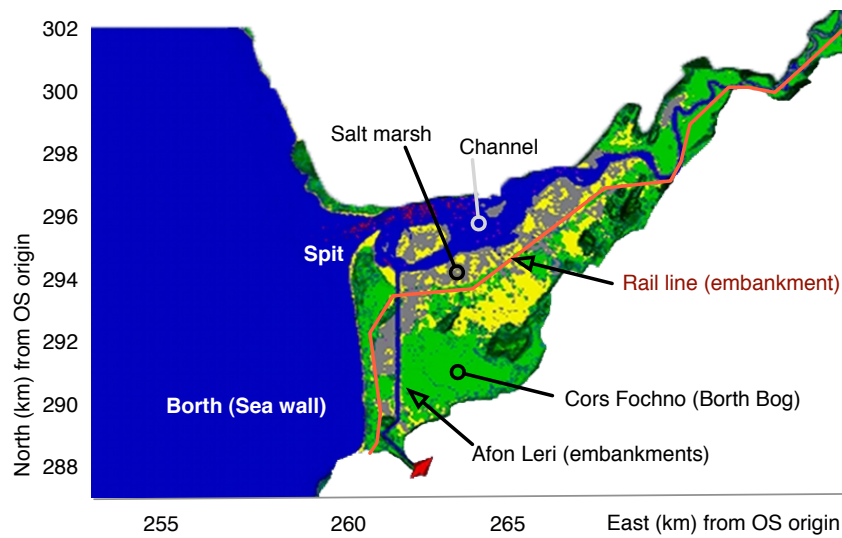


Figure 2.4: The Dyfi estuary. The diagram shows the bathymetry from a model simulation by Telemac and visualized with VINCA. Several features are annotated on the picture, including the embankment, salt marsh, channel and railway embankment.

Sea level rise predictions around the **Burry Inlet** (figure 2.5 bathymetry of Burry Inlet) are similar to those of the Dyfi Estuary [243], but the Burry faces different coastal management issues. It too has protected ecosystems, including vast tidal flats, sand banks and extensive salt marshes on the south coast (SAC, SPA, SSSI, National Nature Reserve and others), which are at risk from increased tidal inundation, together with sediment erosion.

Larger than the Dyfi, the Burry Inlet is 16kms long with a spring tidal prism of $1 \times 10^{12} m^3$, but, unlike the Dyfi, it has a large, urbanised population on its north coast, together with rail networks and a sewage works at considerable risk of flooding. Afon Loughor (and its tributaries), and Afon Llan (including Afon

Lliw) are its primary river systems, producing a mean freshwater discharge into the estuary of $1.1 \times 10^4 m^3/s$. The freshwater inputs, however, are insignificant when compared with the tidal prism. The tidal spit of the Burry is also threatened and might be breached, as a result of wave-induced erosion on its seaward side. This would expose the highly protected salt marsh to the open sea [241,243].

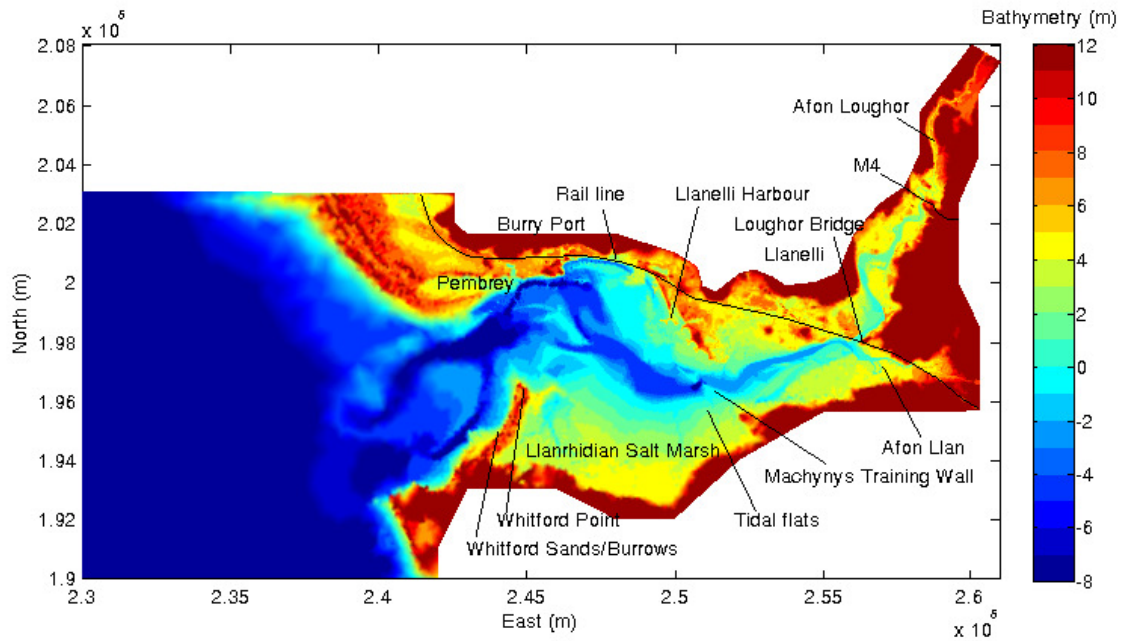


Figure 2.5: The Figure depicts the bathymetry of the Burry Inlet, with the bathymetry shown in colour. Various features can be seen, from the Railway line and Port in the North, to the Tidal flats and Salt Marshes in the South position.

These profiles illustrate that the risks of flooding to populations and ecosystems in both areas are significant, and understanding the complex hydrodynamic factors at play is a key to flood prevention and management of the coastal zone.

2.3 Data Sampling or Simulation

Ocean science research data are derived from two main sources: sampling and modelling. Both have seen dramatic advances over the past fifty years or so, through computing and technological developments, leading to enhanced scientific understanding of the physical processes affecting oceans, seas and continental shelves. Increasingly sophisticated methods of capturing (sampling) data include remote sensing from satellites, weather radar and aircraft, which provide more accurate

measurements of water and wave levels, surface winds, and wave amplitude over the surface of the water than earlier data gathering techniques [140].

Concurrently, enhanced computing capacity and speed have enabled theoretical modelling to provide better measurements; give increased accuracy (demonstrated when tested against historic sampled data) and support more detailed analysis, interpretation and understanding of the data, thereby handling increasingly complex areas of study [140, 141]. This, in part, is due to more accurate representation of land, river and coastal forms.

Coastal shelf models may be used for predictive purposes including wave height and storm surge; sediment transport; changes to sea-bed morphology, and flood prediction, thus supporting coastal zone management, including flood prevention/alleviation measures, and predicting the impact of man made structures, such as harbours and marinas, on hydrodynamics and sediment morphology.

Robins [238] explains that physical processes are simulated by calculating rates of change in time and space of the multiple variables, enabling hindcasts (useful for validation with sampled data) and forecasts to be made. Tidal propagation in an estuary may be simulated by calculating changes in free surface elevation and velocities over a period of time, by creating a computational grid or mesh of the area under investigation, in which the bathymetry and the numerous variables of interest for each grid cell are represented by a single node value. The evolution of the variables may be tracked in space and time by connecting adjacent nodes and performing mathematical calculations at each node.

A modelling solution was preferred for the CAMS project because of the capabilities already described, also there are so many variables in estuarine studies which require a large number of highly detailed simulations to provide accurate and high quality quantitative results. Many of the phenomena simulated cannot be predicted by non-modelling means: for example, complex combinations and balances of spring tides, high storm surge and increased freshwater run-off in conjunction with sediment erosion and other local factors. These are important as they may cause significant local flooding, and their impact on sediment erosion may have further hydrodynamic consequences [239]. Already mentioned - another factor supporting the use of models is that, within this complex domain increasingly influenced by sea level rise, historic samples do not necessarily provide a sound basis for future predictions of change, so forecasting the evolution of shape and form of coastlines is still not well developed and presents challenges [48, 67]. Nonetheless, sampled data were used to validate the accuracy of the model by comparing its

outcomes with historic, observed hydrodynamic flow patterns and sediment transport.

2.4 Numerical Models and Current Visualization Approaches

Oceanographic models are complex, with much regional variation, and research indicates there are a wealth of them. An appreciation of the sheer number, diversity and complexity of coastal shelf models and modelling techniques may be gained from Jones' 2002 study [155] of operational modelling in Europe. Nearly 200 models were reported by respondents, with at least eight methods of describing and dividing the domain, and it was acknowledged there may be many more unreported models in use. There are many one-off models, together with models for particular aspects of coastal shelf studies, for example, TELEMAC¹, a widely used model simulating coastal, estuarine and river hydrodynamics, which provides the datasets for this work.

The development of a high performance visualization system has to take into account the variables relating to the model. These include the equations used; approximations; methods of describing and dividing the domain of study; types of vertical coordinates; dimensionality; time discretisation; advection; boundaries; included processes and so on. Features (or objects), for example sandbanks, tidal channels, scour pits, may be affected by a number of factors, and may change position with the passage of time, thus adding to the complexity and challenges of analysing and understanding the data. Further, datasets are also subject to abnormal distribution, because of the nature of the sampling.

However, there are common factors amongst modelling systems: they tend to produce extremely large multi-field datasets, and many models use unstructured, adaptive grids, with objects represented by unconnected three-dimensional coordinates. Another common factor is the widely recognised difficulty in dealing with uncertainty, and thus validating the simulations and their associated graphical representations. All data is uncertain to an extent [7] – and ocean science data more so, because of its inherently variable nature [181]. To this can be added the uncertainties introduced during the modelling and visualization stages. Brown et al. [48] considers the lack of processes to communicate uncertainty in studies of coastal ero-

¹www.opentelemac.org

sion, also the impact on uncertainty of changes to spatial resolution are a concern for a number of researchers [25, 92, 116]. Furthermore, displaying uncertainty in multi-field data, (such as that in this study), presents challenges, because of the difficulties of defining, representing and controlling the uncertainty [217, 257]. Such uncertainties may have an impact on the predictive capability of a model: an important factor if it is to be relied upon, for example, by coastal zone managers. Thus, a role for an advanced visualization and visual analytics techniques system may be to provide an accurate representation of such uncertainty.

However, for this work, the main area of interest is post processing. How are the data displayed and explored? Do ocean science researchers still rely on the models post-processor and visualization system for displaying datasets? Have high performance visualization systems been developed either generically for models, or for the particular oceanographic area of study? What type of visualization system is used?

Despite the need for advanced visualization and visual analytics systems to support models, identified in Chapter 1, researchers suggest that current practice amongst ocean scientists is to rely on static visualization systems, rather than novel, sophisticated data exploration methods [13, 130], although Song et al. [271] recognises that there is some acceptance that current tools are lacking, albeit in the related area of atmospheric studies. Methods for visualizing the data include the simple visualization tools provided by the model developers; other customised post processing tools, or MATLAB or Excel, or a combination of these. Other, more sophisticated tools have been used (such as VTK, Vis5D+ and Voyager) [13, 102]. Collaboration with ocean science researchers on this work suggests that decisions on where to locate the visualizations spatially is based on their own assumptions and experience, rather than on exploring all possible positions.

However, non-interactive exploration may not be a question of the functionality of the tools available, but more the scale of the data. At the current state of evolution of visualization and analytics systems for coastal shelf models, familiarity is probably the key factor by which ocean science researchers select a visualization system, despite the availability of custom built tools (e.g. Cotter and Gorman [65] describe a tool for handling problems with unstructured meshes which may resolve some of the difficulties they experience with their models).

Thus, the visualization systems routinely used are limited: the data is saved from individual runs of the simulation, processed and then visualized at widely varying levels of sophistication. These systems tend to use two dimensional snapshots or

contour plots; it is difficult to visualize multiple variables on the same plot; animation is limited and the user has to switch from snapshot to snapshot to examine the data. Viewing static images simultaneously provides little scope for comparison of such dynamic data. There are, therefore, limited opportunities for visualizing multiple variables in the same space, and for integrating larger scale hydrodynamic effects with the small-scale hydrodynamics of detailed areas of study [277].

These factors suggest there is a need for novel methods of exploring and analysing coastal shelf data, using interactive visual analytics techniques to provide holistic views of the data; to query parts of them; to facilitate quantitative measurements of specific positions and make calculations; to study different parameterisations, and to look at the data from different viewpoints.

2.5 TELEMAC-2D

The case study data are derived from TELEMAC-2D, a numerical hydro-informatic system, with an integrated tool used to model free surface flows, which is based on the shallow water equations. It calculates the temporal and spatial variations in the hydrodynamics of the area of study, through a finite element grid representing the bathymetry [238]. TELEMAC (now open source software, TELEMAC-MASCARET²) was developed by the Laboratoire National d’Hydraulique, to provide numerical solutions to complex, multi-disciplinary hydrology problems, and has widely accepted provenance, founded on its sound theoretical base. Described as “a whole processing chain for the calculation of water, solute and sediment motions in the fluvial, estuarine, lacustrine and groundwater domains” [127], it has been used for numerous studies worldwide, such that it is a common and documented standard in its field [128].

TELEMAC’s large library of numerical algorithms (shared by all modules to ensure consistency) includes: 2D hydrodynamic flow; 3D hydrodynamic flow; wave agitation in harbours; wave propagation in the coastal zone; 2D water quality; 2D sediment transport; 2D suspended sediment transport; graphical post-processing; grid generation; grid interface; 2D sections through the results of a 3D simulation and 2D flows². Consecutive linking of modules such as tidal hydrodynamics and sediment transport establishes relationships between the two, which are fed into successive modules, thus supporting research.

²www.opentelemac.org

TELEMAC-2D solves the St Venant equations of momentum and continuity (derived from the Navier-Stokes equations), by initially solving advection terms, followed by propagation, diffusion and source terms [25, 129]. It uses finite element, high-capacity algorithms, discretising space through its Matisse module as an unstructured grid of triangular elements, (with variables calculated at the corner nodes of the triangles comprising the mesh), and also defining boundary conditions. The system treats the computational domain as a continuous field, rather than a series of cross sections, which, together with the finite element methodology, enables complex topography to be represented with a minimum number of elements, whilst supporting accuracy in solving the St Venant equations and reducing computational time.

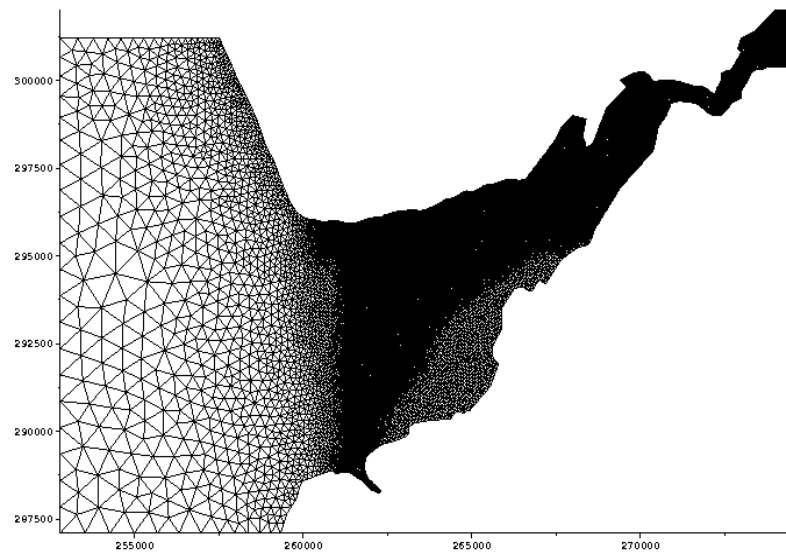


Figure 2.6: TELEMAC-2D grid for the Dyfi Estuary, showing difference in resolution for area of greater interest. The offshore resolution is 500m, gradually reducing to 50m in the estuary and 25m in the river channels. Direct copy from Robins [243].

Robins and Davies [241] considers TELEMAC-2D to be an “ideal modelling framework for the estuarine environment”, because its finite element grid permits graded mesh resolution, see Figure 2.6.

Similarly, Jones [155] believes it to be particularly useful for areas of complicated geometry, as it enables areas of interest to be refined. Areas of considerable complexity, requiring high levels of bathymetric accuracy, such as coastlines, river channels and features like embankments and sea defences, can be the subject of detailed resolution, whereas areas of less interest, eg offshore grids may be increased, to optimise computing power [241]. TELEMAC-2D’s other value to this research

is that it is able to simulate wetting and drying of specific areas (cells), important when modelling tidal flats in the Dyfi Estuary and Burry Inlet.

TELEMATAC uses an object oriented programming paradigm where modules can be modified to meet the needs of specific simulations, making it extremely flexible [127]. It is able to specify boundary conditions; introduce new functions and link with other modelling systems. However, like all numerical modelling systems, many researchers believe there are uncertainty challenges with regard to validation of the models [47, 92], but these are not the focus of this thesis.

2.6 Graphical interfaces and post processing for TELEMATAC

TELEMATAC incorporates a number of graphical interfaces and post processors for visualizing the data, but in common with other models, these tools are limited. Principal tools include Rubens (TELEMATAC's post processing software), Blue Kenue, produced by the National Research Council of Canada ³ and Fudaa-Prepro ⁴. However, they do not provide the required levels of interactivity, and do not compare well against the higher level of analysis possible with scientific visualization tools. For example, Rubens utilizes 2D snapshots or contour plots; it is difficult to visualize multiple variables on the same plot and little animation is provided. The user has to switch from snapshot to snapshot to examine the data, so viewing static images of highly dynamic data limits the scope for comparison.

CETMEF's (part of the French Ministry of Ecology and Sustainability) Fudaa-Prepro GUI for TELEMATAC is little reported upon: there are few research papers documenting its use and capabilities in supporting detailed analysis of simulation. Observation of the system suggests that although it provides a Java front end for TELEMATAC, it supports only limited interactivity and visual analytics.

The CAMS researchers selected Blue Kenue, see Figure 2.7, as their preferred visualization tool for its ability to generate triangular meshes using constrained Delaunay triangulation (essential because of undersampling at the mouth of the Dyfi Estuary) [103, 238]. The system inputs lines plus other regular and triangular grids; hard points and break lines can be specified in the mesh generator, which are preserved when the nodes are created, and node density is specified from a user provided density map (which can be rectangular grids, triangular meshes or polygonal data).

³www.nrc-cnrc.gc.ca/eng/ibp/chc/software/kenue/blue-kenue.html

⁴prepro.fudaa.fr/

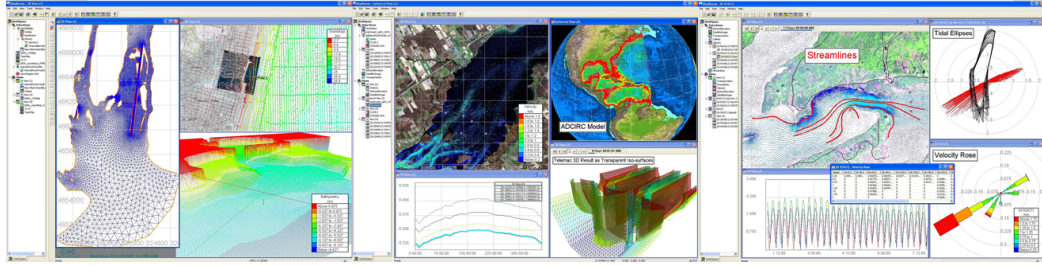


Figure 2.7: Screenshots from Blue KenueTM that depict several visualizations that are typical from the use of this software.

Whilst Blue Kenue supports data preparation and visualization, and has proved to be extremely useful (in research in Bangor’s School of Ocean Sciences, in particular), neither it, nor Rubens, nor Fudaa Prepro are able to provide the detailed level of analysis required. For example, they cannot undertake the instantaneous computation of water and sediment flux between defined points, which are key components of tidal prism, an essential measure in understanding flood risk. In practice, to compute the flux data, a custom MATLAB script had to be developed which, itself, proved time consuming to operate and complex to control.

Despite its widespread use, there appears to be few visualization tools for TELEMAC, incorporating advanced visual analytics tools. Therefore, this thesis proposes to study the effectiveness of quantitative analysis techniques using VINCA (Visual Environment for Coastal Environments), a tool developed for visual analytics environments, and which is described in more detail in the following chapters. The goal of this research is to provide a tool where all the locations might be calculated and analysed in one step, and includes a flux calculator, which has been developed to enable scientists to enhance the speed and efficiency of specific calculations and analysis.

2.7 The case study data and the visualization and analytic challenges

Modelled earth science simulations are often complex and lengthy [87], and estuarine studies are no exception, tending to produce very large datasets with many variables. Ledoux and Gold [181] and Robins and Davies [241] believe the challenges of modelling and visualizing riverine and estuarine models to be greater than most other areas of ocean studies for this reason. Furthermore, conclusions drawn

from micro-scale analysis might not apply on a larger scale [47].

Two datasets are used: the Dyfi Estuary and the Burry Inlet, (Figure 2.8 provides an overview of the data and the number of variables) which are simulated using the M2-S2 tidal components. They are of interest because of their potential for flooding, they are dynamic environments, and they demonstrate complex inter-relationships between geological features, weather patterns, tidal flow, sediment transport (and the nature of the sediment), vegetation patterns and the influence of human intervention (e.g. flood defences and other man-made structures) [241]. These inter-relationships present significant challenges to understanding processes and predicting change, particularly with the addition of a further variable - sea level rise. Some changes occur swiftly, whereas others evolve slowly over lengthy periods, and man made structures may also affect the complex balance [287].

Time Independent		Time Dependant					
X	Y	Bottom	Water Depth	Free Surface	Velocity U	Velocity V	Speed

Figure 2.8: Table of variables loaded into VINCA

The simulations are created on unstructured grids, using depth averaged values, storing higher resolution points along and in the estuary. The grids include man-made features such as coastal breakwaters; railway embankments and sea walls; major river channels, tidal flats and low lying land susceptible to flooding, and extend offshore to simulate tidal propagation [239, 241]. The need to calculate over the entire flood plain for this particular project requires a wider high resolution area than normal for such modelling exercises.

Robins et al. [239] describes the simulation data: present day mean and extreme tidal/fluvial scenarios have been simulated, exploring potential flooding events over a 100 year period from the present day; modelling extreme tidal and fluvial events (worst case scenarios) over that period, and focusing on features such as tidal channels, sandbanks, scour pit etc., to identify how they change and what impact this has on predicted flooding. The simulations incorporate potential flood management options, such as new sea defences and embankment configurations and simulate the consequences of these on tidal prism, tidal velocities and further potential for

flooding.

For the boundary conditions, at open sea boundaries predicted tidal harmonics of elevations and velocities are used to force the model; at the river boundaries, river run-off represents the freshwater input, and frictional resistance to the flow is used on the sea bed. This enables barotropic and baroclinic velocity simulations to be made. The model also incorporates atmospheric conditions that model heat exchange, wind stress and precipitation.

The standard prognostic variables are those calculated by the simulation, including temperature, water depth and salinity, but the diagnostic variables are derived from the data, which include depth-averaged velocities and the tidal flux.

A typical run of the tidal elements of a simulation for this project comprises over 50,000 points for each time step, and usually 100 time steps per run. This is the data which will be imported into a database within VINCA for visualization and analysis. Thus, 10^6 data points require to be plotted if all of these time steps are to be displayed at the same time.

But when these points are plotted onto a typical screen resolution of 1200×800 , each of the data points in the estuary becomes overplotted. At this resolution, and for this dataset each pixel represents about 200 points per pixel. The detail of these plots is demonstrated in Figure 2.9, which shows a set of high resolution images generated by VINCA of the Burry Inlet. Each of the eight time-steps is four hours apart and the figure covers the entire M2-S2 tidal cycle. The colours in the figure represent the magnitude of the hydrodynamic flow.

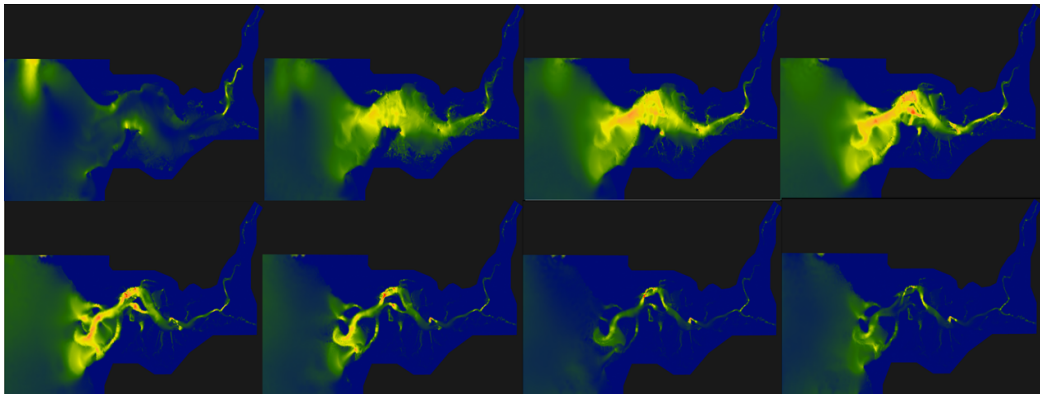


Figure 2.9: Burry Estuary visualization (generated by VINCA) showing scale and complexity of simulated data of the hydrodynamic flow. Eight time-steps are shown. The timesteps are four hours apart, cover the M2-S2 tidal cycle and the colours represent the magnitude of the hydrodynamic flow.

Although the simulation extends the higher resolution area further than normal,

there are still areas of under-sampling at the mouth. However, use of a constrained Delaunay triangulation provides the desired topological equivalence.

A further problem encountered with the data during initial prototyping was co-incident topology – multiple surfaces at the same point in 3D space – which probably hindered early attempts to develop analytic tools, because it proved challenging to accurately select, or ‘pick’ individual points upon which to perform analytic task [103]. Viewing geometry is an important element in any visualization and visual analytics tool, so VINCA had to be able to support accurate point selection.

2.7.1 An important hydrodynamic component - Tidal Flux

One of the most important hydrodynamic aspects of study within the modelled data is that of velocity change: Robins et al. [239] regards understanding the changes to water flow (velocity), caused by the differing variables within each tidal/fluvial scenario, to be a key to establishing the conditions which will result in flooding in different areas around the estuaries. To enable the impact of differing hypothetical flood management strategies (e.g. sea defences or embankment configurations) on tidal prism to be established, tidal velocities needs to be investigated.

Thus, tidal flux (in the case of an estuary the volume of water per second generated by tidal forces flowing across a given cross section of the estuary) is a fundamental element of the work of the CAMS researchers. There have been references throughout this chapter to tidal prism (which measures the amount of water flowing in and out of the estuary at any given point). Coastal ocean scientists regard tidal prism as a single measurable parameter providing insight into the hydrodynamics and morphodynamics of an estuary [239]. However, tidal prism cannot be calculated without first calculating the flux across the mouth of an estuary at high and low tides, throughout a period.

Models, including TELEMAC-2D provide prognostic variables (height, depth, speed, vector flow), but they do not calculate these other required quantities. Thus, these extra calculations are undertaken as separate post-processing tasks – and the limitations of the routinely used post-processing systems for TELEMAC-2D have been described earlier in this chapter. Bearing in mind the importance of tidal prism to their simulations and predictions, researchers have undertaken complex and lengthy, but by no means efficient, processes to calculate tidal flux. Thus, the development of an efficient and accurate flux calculator for estuarine studies was identified as a priority for this thesis.

2.8 Summary

The chapter presents the two principle datasets, which are predictive flooding scenarios of the Burry and Dyfi Estuaries in Wales, simulated using the TELEMAC-2D model. The attributes of the model are presented, together with the reasoning for using simulated data. The datasets are relevant and important because they provide a complex scenario with specific, unique challenges studied by ocean science researchers in support of coastal zone management and flood prevention. The research also uses sampled data to underpin and verify the simulations.

The datasets represent regions that flood: by analysing the hydrodynamic simulations, researchers are able to study sediment transport over time, leading to an understanding of the changes which lead to flooding. Similar work to resolve real life problems in flood prevention and coastal zone management is being undertaken by ocean science researchers in estuarine domains worldwide. Holistically, therefore, these databases are representative geographic locations which may influence research into climate change, and associated rises in sea level.

Finally, the datasets are large and complex: ocean scientists produce several runs of their simulations in TELEMAC-2D, using different parameterisations over many years. The scale of this represents a significant challenge for ocean scientists, and thus for this thesis, where new methods for analysing, manipulating and visualizing the data may be tested.

CHAPTER 3

The Visualization Context: review of practices in ocean science visualization and visual analysis

This chapter provides a context for the development of an appropriate visualization and visual analytics system for modelled, estuarine hydrodynamic data, by considering the current state of evolution of ocean science visualization. It seeks to establish what progress there is towards the use of exploratory visualization tools, integrating data-intensive visual analytics methods and discusses examples of visualization and analytics systems developed for the ocean sciences, including both the coastal shelf and estuarine domain and the deep ocean. Although the review refers to the development tools and systems encountered in ocean science visualization, these are not the focus of the research or this section, but are discussed as a means of identifying approaches which have been successfully applied to the oceanographic domain.

The review will demonstrate that developers are starting to create interactive visualization systems and techniques to address the data intensive needs of coastal and estuarine ocean scientists, for example the ocean observatory workflow models which are discussed in Section 3.6. However, a growing trend identified in this review is the focus on visualization and visual analytics for modelled simulations of ocean, climate and meteorological studies. This has possibly occurred because of the need to understand extreme events such as storm surges, hurricanes and flooding (probably the consequence of climate change) to support disaster management plan-

ning, within environmental decision support systems (EDSS) and also to develop preventative measures. This, in turn, has led to a greater focus on the development of bespoke or turnkey visualization systems, through working in collaboration with domain researchers to identify their study requirements, and taking into account the recognised challenges of visualizing modelled multi-field data, particularly if these are unstructured and multi-resolution, which generic visualization tools struggle to handle.

Nonetheless, both this review and Chapter 4 will show that such systems are mainly confined to large scale projects, and coastal shelf and estuarine hydrodynamic visualization remains a scantily researched area, with many domain researchers continuing to rely on traditional post-processing methodologies.

The section incorporates:

1. Brief review of ocean science development environments;
2. Bespoke (turnkey) systems for ocean sciences;
3. Deep ocean visualization;
4. Coastal shelf and estuarine visualization;
5. Hydrodynamic predictive studies, e.g. visualization systems for preventative management and disaster planning (Environmental Decision Support Systems, EDSS);
6. Ocean Observatories;
7. Workflow Systems.

3.1 Brief review of visualization environments associated with ocean sciences

Whilst by no means a comprehensive review, this section provides a sense of the numerous systems and tools which have been identified in relation to oceanographic visualization and visual analysis. Most of the visualization systems mentioned incorporate hydrodynamic elements to differing degrees, whether or not the hydrodynamics is the focus of the research.

Andrews [9] review of benthic mapping visualization identifies GIS and visualization systems including ESRI ¹, MapInfo ², Erdas ³, Envi ⁴, Surfer ⁵, Fledermaus [197], GMT ⁶, and ArcExplorer ⁷ as useful for 2D benthic data. However, concerns were also expressed that these systems, at that time, were not able to support 3D visualization, which enables scientists in differing disciplines to better understand the spatial relationships, and within a visual analytics system, this is useful for providing “context”.

Borodin et al. [42] reviews visualization tools in relation to the volume, flow and terrain visualization and presentational needs of the multi-disciplinary MERCW Project, which studied the ecological risks posed by sea-dumped chemical weapons. This incorporated hydrodynamic elements similar to those relevant to this project, so this review provided a useful starting point. ESRI ArcGIS ESRI ArcGIS [183], Fledermaus [197], GeuZui3D [20,311], GeoNav3D [19], GeoVR [139], GeoVRML [198], X-VISION [159], and Vis5D [199] were identified as useful tools for oceanographic research, and Virtual Geographic Information System [186] was singled out as capable both of 3D visualization and also the capacity to handle large datasets, through level of detail algorithms.

Research into oceanographic physical processes, including hydrodynamics, are intrinsic to understanding weather patterns and climate, thus visualization systems used in climate and meteorological research are mentioned here, albeit accepting that in weather and climate research, the emphasis is on 3D flow and volume visualization, so some systems may not be appropriate for 2D modelled hydrodynamic data. However, there is a great wealth of visualization systems in use in climate and weather studies, including many specific, domain based applications. Middleton et al. [201] in a review of visualization in weather and climate research identify Vis5D ⁸, VisAD ⁹, IDV ¹⁰, AVS ¹¹ [297], Data Explorer (OpenDX) ¹², Iris Ex-

¹<http://www.esri.com>

²<http://www.pbinsight.com/welcome/mapinfo/>

³<http://geospatial.intergraph.com/products/ERDASIMAGINE/ERDASIMAGINE/Details.aspx>

⁴<http://www.exelisvis.com/envi/>

⁵<http://www.goldensoftware.com/products/surfer/surfer.shtml>

⁶<http://www.gmt.soest.hawaii.edu>

⁷<http://www.esri.com/software/arcgis/explorer/arcexplorer>

⁸<http://vis5d.sourceforge.net/doc/>

⁹<http://www.ssec.wisc.edu/~billh/visad.html>

¹⁰www.unidata.ucar.edu/software/idv/

¹¹<http://www.avs.org/>

¹²<http://www.opendx.org/>

plorer ¹³, Wavefront Visualizer ¹⁴, and VTK ¹⁵ as useful tools. However, many climate researchers prefer fourth generation data languages such as IDL ¹⁶, Ferret ¹⁷, GrADS ¹⁸, NCL ¹⁹ and CDAT ²⁰. These have all been developed for use in this domain of study, even though their visualization capabilities are widely acknowledged to be more limited. To the Middleton et al., Nocke et al. [208] adds MAGICs ²¹; McIDAS ²²; Metview ²³; Ocean Data View ²⁴, and GMT ²⁵, commenting that Ocean Data View, Ferret and GMT are also useful for oceanographic research.

However, Nocke et al. [208] qualifies the usefulness of many of the application systems mentioned above: they are quite old and possess limitations which hamper insight into hidden structures in multivariate modelled data, as they focus on spatial visualization of individual models. Thus, notwithstanding their difficulty of use, Nocke et al. [208] concluded that the functionality of current commercial and open source software such as AVS, OpenDX, VTK, InfoVis-Toolkit and Spotfire, was more useful for current visualization requirements.

Conversely, despite the availability of a wealth of visualization development tools, Clyne et al. [62] argues that most generic systems target broad needs, and lack tools and techniques (including numerical techniques for quantitative analysis), which will enable them to tackle domain-specific problems (e.g., mapping transformations in the geosciences). At the same time, these systems offer features and capabilities of little interest to various groups, but which probably contribute to the overall complexity of the system.

Nonetheless, some of the more advanced visualization and analysis systems do use generic visualization toolkits (for example VTK in [136]), often as part of a complex workflow package integrating multi-disciplinary, multi-source data acquisition, pre-processing, visualization and analytics of the type described by Howe et al. [136] for an ocean observatory.

The review will suggest that the nature of oceanographic data in general, and

¹³http://www.nag.com/Welcome_iec.asp

¹⁴http://www.en.wikipedia.org/wiki/The_Advanced_Visualizer

¹⁵<http://www.vtk.org/>

¹⁶<http://www.exelisvis.com/idl/>

¹⁷<http://www.ferret.wrc.noaa.gov/>

¹⁸<http://www.iges.org/grads/>

¹⁹<http://www.ncl.ucar.edu/overview.shtml>

²⁰<http://www2-pcmdi.llnl.gov/cdat>

²¹<http://www.ecmwf.int/publications/manuals/magics/>

²²<http://www.unidata.ucar.edu/software/mcidas/current/mclearn/mcidas.html>

²³<http://www.ecmwf.int/products/data/software/metview.html#description>

²⁴<http://odv.awi.de/en/home/>

²⁵<http://gmt.soest.hawaii.edu/>

modelled estuarine hydrodynamic data in particular, will necessitate a combination of a number of systems and tools, and it is possible that a single, off the shelf visualization system may not be able to meet all the research requirements, no matter how extensible or flexible. Thus, in view of the goals and challenges presented by this thesis, a turnkey approach is of particular interest for this work.

3.2 Bespoke (turnkey) systems for ocean sciences

In the context of this thesis, turnkey systems are developed specifically to support an area of research or study, to meet researchers' needs. Many of the examples identified – and indeed the more recent ones – also fit into the categories of predictive/preventative management; workflow management or ocean observatories and are discussed in those sections.

One of the best known early turnkey visualization and analysis system for marine studies is GeoZui3D [311], which supports interpretation of multi-disciplinary marine data, including hydrodynamics. Its unique, centred zooming user interface, linked views and analysis tools are familiar methodologies in many visual analytics systems today. However, unsurprisingly, one of its limitations is speed, such that it is bettered by many newer systems running on PCs [42], capable of tracing hundreds of thousands of particles in real time, a feat beyond even recent versions of GeoZui3D. Furthermore, even though it has evolved, it does not possess the tools for rich, interactive data exploration [111].

An early visualization system flow visualization system for studies of global scale, ocean and acoustic propagation (modelled and sampled) data was AGP [149]. The developers found it particularly useful for hydrodynamic data, but, of course, this is an old system now and was developed for deep sea studies, so is unlikely to have the functionality required for estuarine studies.

Another early system, Wang et al. [306] developed a Windows based visualization and analysis system (PC-Vis5D) for coastal, estuarine and atmospheric modelling. The rapid growth in the size of datasets has proved problematic for researchers using PCs, and the system aimed to provide enhanced interrogation, including stereoscopic viewing, and analysis tools, whilst improving performance. The system was successfully tested on the tidal circulation pattern of the Pearl River, where it provided new insights into hydrodynamic flow in the estuary.

Liang and Molkenthin [185] describe a distributed visualization system for hydrodynamic simulations of a large river estuary, where the requirement is for a col-

laborative approach in a distributed environment. The visualization system was built in Iris Performer, which in conjunction with OpenGL and Cave Library also enabled exploration of the data through ImmersaDesk.

More recently, Ho and Jern [135] presented a state of the art visual analytics system. This customised Volume Data Explorer, used VTK, with the GAV toolkit to produce an integrated Infovis and Scivis system for studying large scale climate models, in particular the deep ocean circulation. Whilst the focus was on volume visualization and analytics, the developers concluded that the use of 2D views, graphs and plots facilitated exploration and discovery.

VAPOR ²⁶ is of interest as it provides analytical and visualization capabilities for extremely large, modelled datasets in earth and space sciences, using desktop computing capabilities, through a wavelet-based progressive access data model. Its GUI supports a wide range of visual analytics tools and methodologies, although by its own admission, it does not have the quantitative analysis capabilities of systems such as MATLAB. Most of the research output thus far is focused on solar research, astrophysical simulation and hurricane analysis, so it is not possible to comment on how far it would be a useful development system for estuarine hydrodynamics. However, its major constraint is that it is limited to structured grids, and does not support unstructured grids, so is unlikely to be of use for this work. Nevertheless it illustrates the growing trend towards the creation of analytical and visualization systems directed towards resolving the challenges inherent in modelled data.

BlueKenue ²⁷ is a good example of a turnkey visualization tool. It was developed specifically for hydraulic modellers, and associated primarily with TELEMAC numerical models. However, there are few published examples of its use [178,239], and little discussion of its strengths, weaknesses and capabilities as a visualization tool able to support interactive visualization and visual analytics. It is mentioned in this review as it is used to both model and visualize data, and, importantly, is used by the ocean scientists collaborating on this project. It is discussed more fully in later chapters.

3.3 Ocean Visualization

An early example of an interactive, integrated system for handling, analysing and visualizing multi-source, multi-modal oceanographic data, is Knudsen [173] which

²⁶<http://www.vapor.ucar.edu>

²⁷www.nrc-cnrc.gc.ca/eng/ibp/chc/software/kenue/blue-kenue.html

integrates GIS functionality with image processing and data management systems to provide full spatio-temporal functionality. The MERCW project [42, 105], also uses GIS integrated with advanced flow visualization techniques to study hydrodynamic flow in relation to chemical pollution in the Baltic. Similarly, Kemp and Meaden [170] use GIS as the basis of a fisheries management, analysis and visualization system because of the need to import data from a number of sources (including cartographic information), and the requirement for a flexible system meeting the needs of a disparate range of users.

As part of the Poseidon multi-disciplinary ocean observation programme, Patrikalakis et al. [221] used OpenDX to provide greater visualization functionality, than that provided by MATLAB and similar systems, including interactive, graphically complex, distributed workflows, incorporating feature extraction and estimations of uncertainty. Of interest in OpenDX is its ability to support unstructured, scattered data [132], but, unfortunately, this is counterbalanced by its limitations with regard to large data management [103, 222], and thus interactivity – something it shares with many other generic visualization environments [305].

VTK, a comparatively widely used visualization system in geosciences and ocean science, was used, in conjunction with Maya Vi, by Cotter and Gorman [65], for a visualization system on which to prove an algorithm for calculating density fluxes in modelled unstructured oceanographic data, finding it provided many of the high-level primitives needed to perform 3D visualization and evaluate diagnostic quantities.

3.4 Coastal shelf and estuarine visualization and visual analytics

An early system developed specifically for understanding complex, modelled hydrodynamic features in Lake Eyrie, (which would be similar to those in an estuarine environment) was developed using apE [194]. Despite its age, it offered functionality which many present day ocean scientists do not routinely have access to, including interactive computational steering and feature extraction. Similarly, other early visualization systems presented techniques and tools to compensate for the limitations of existing analytical methods, including visual analysis techniques.

Galloway et al. [99] used a visualization system developed in Data Explorer to study tropical ecosystems off the coast of SE Asia and Australia, regarding it as

a significant improvement on static plots. Kitsiou et al. [172] describes a highly portable data assimilation, visualization and analysis tool built in IDL, for simultaneous acquisition and study of sampled data derived from multiple sensors in the Gulf of Lions to understand marine pollution. The interactive system permits analysis of spatial and temporal variability of scalar and vector data, data filtering, spectral analysis, multiple views, stick plots and statistical analysis. However, He and Hamblin [121] found visualization of finite element modelled hydrodynamic data challenging, using Iris Explorer, because of its inability to visualize irregular data.

The ability to handle modelled unstructured, multi-resolution data is a problem for a number of visualization systems. VTK is an exception [65]: it is able to deal with AMR data. As a result, it has been used in a number of coastal shelf and estuarine visualization systems, including an interactive system for visualizing simulated, multi-field datasets studying estuarine water quality, (See Figure 3.1).

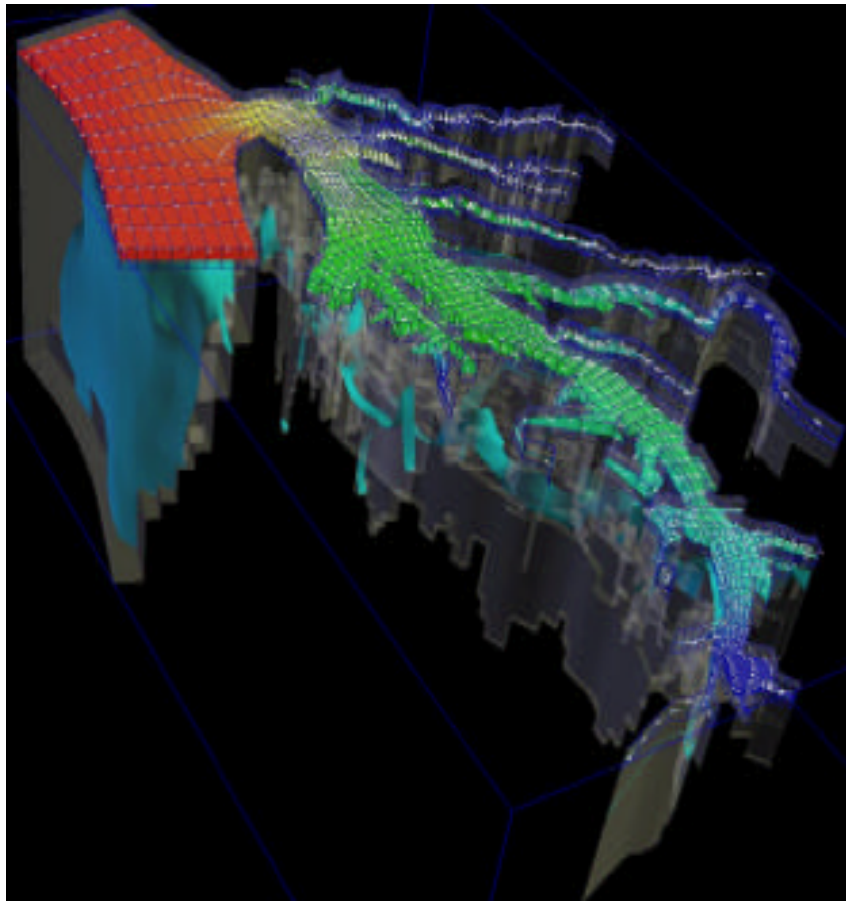


Figure 3.1: copied from Stein et al. [277]: multivariate data visualization of salinity, isosurface of dissolved oxygen and flow vectors.

However, Jiminez et al. [148] experienced difficulties with VTK in developing 3D and 4D visualization and analysis tools for a complex environmental and observation system, producing modelled and sampled/sensed data and working in close collaboration with users. One of the main goals was to achieve interactivity: a challenge considering the scale and multi-disciplinary nature of the data. However, despite its claimed functionality for handling large, 3D, time varying unstructured data, the developers had to augment VTK with custom code, including complex processes for data import; volume rendering and visualizing the velocity field. Furthermore, the speed/quality dichotomy proved an issue: VTK could not deliver real time frame rates. Similarly, whilst Keen et al. [159], found VTK to be helpful in a suite of visualization and analytic tools for an immersive visualization system studying sediment transport, the requirement for interactive frame rates necessitated the use of IRIS Performer to optimise the application.

Whilst most estuarine hydrodynamic studies continue to be conducted using traditional numerical models such as TELEMAC-2D, in conjunction with MATLAB style tools and limited visualization systems, GIS is increasingly used in this domain. Andrews [9] reports that GIS is the main tool for analysing and visualizing data relating to benthic habitats, which include hydrodynamic datasets amongst others. He concludes that whilst the lack of a 3D perspective may prove limiting to establishing relationships, this is not an issue in shallow estuarine environments.

3.5 Coastal shelf and estuarine visualization for Environmental Decision Support Systems, (EDSS)

This is an area of particular interest in view of the nature of the data for this thesis. The coastal zone is often mentioned in conjunction with management, and extreme event management (e.g. floods and storm surges) is the focus of much predictive coastal shelf and estuarine modelled hydrodynamic research. Here, the ability of GIS to combine data of various types has resulted in its use in many Environmental Decision Support Systems, in conjunction with visual analytics tools [200]: Zheng et al.'s [328] storm surge decision support system for the Yhangtze Estuary; Bonazountas et al.'s [41] integrated modelled and sensed visual analytics system for river basin disaster planning and management involving 14 countries; Lee and Choi's [182] GIS-based VR system for hydro-environmental modelling and visualization for communicating environmental impact issues to the public and stake-

holders, and Allen et al.'s. [6] SLOSH storm surge system, which is widely used in coastal locations worldwide.

However, whilst its use for predictive studies seems to be well established, limitations have been identified in such systems, as GIS struggles to handle unstructured data. Wright and Halpin [322] also discusses its limitations in representing dynamic data with constant changes in location and attributes, where it is constrained by its convention of combining data of various types by assigning coordinates and displaying the layers together. This suggests it is of limited applicability for research involving studies of tides, currents and sediment transport. Similarly Allen et al. [6], expresses concerns in relation to storm surge studies, citing inherent 3D space-time dynamics limitations, together with computational problems, and generalisation and conflation errors when integrating coarse grain model output with fine scale GIS data. A limited ability to visualize temporal data and large datasets is also raised by Bonazountas et al. [41]. Thus, for the purposes of this thesis, the disadvantages of a GIS approach would seem to outweigh the advantages.

Other turnkey visualization and visual analytics approaches for this domain include Pais et al. [214] which focuses on dyke flood control and analytic tools to optimise sensor control room management, and North Carolina's Coastal Hazards Centre is developing novel visual analytics tools for managing large scale natural disasters, including storm surges, flooding, severe winds and rainfall [233]. Likewise, the FloodViz [80] visualization and analytics tool, has been developed to aid assessment and interpretation of modelled river flooding scenarios, including storm surge and flood wave propagation, but again with the emphasis on event management.

Harrison et al. [118] describes a highly flexible, interactive visualization system for analysing LIDAR data of coastal regions to support both exploratory and predictive analysis, based on chains of multiple linked scatter plots, managed by an interactive tree view. However, these studies focus on understanding the evolution of disasters in real time, rather than prediction and the development of preventative management strategies, which may require a different analytical approach.

A combined deep ocean and coastal shelf application includes Zhang et al.'s. [326] modelled tsunami wave propagation forecasts, for emergency management: a challenging area of research involving studies of very large scale data relating both to the formation of waves and their impingement on the coast, encompassing a number of different types of models. As with most areas of ocean science, the system developers regard visualization for this domain as not very advanced. Amira's

interactive 3D volume visualization system was used in conjunction with Google Earth²⁸ to provide geographical information and to enable the visualization results to be displayed on a virtual globe (see Figure 3.2).

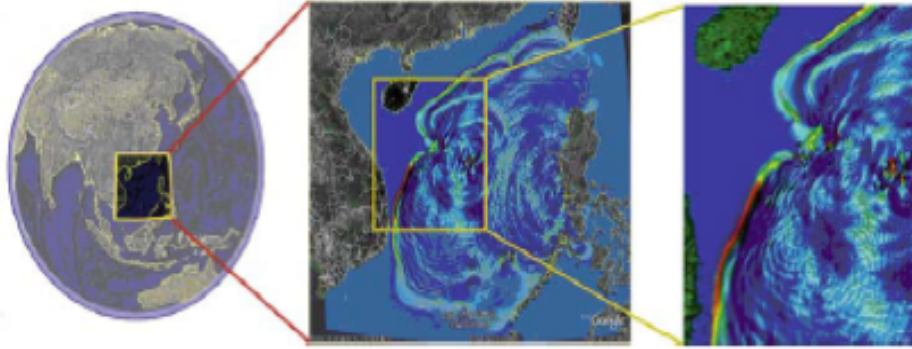


Figure 3.2: Copied from Zhang et al. [326], multiple level of detail visualization of tsunami simulation, placed into global context.

In support of climate change studies, Theron's [284] Parallel Coordinate Plot (PCP) based visual analytics tool provides insight into paleoceanographic data, by reconstructing environmental features over thousands of years and Turdukulov and Blok [295] applies visual analytics to tracking iceberg movement and breakup. For coastal margin research, Howe et al. [136] describe a domain specialised, platform integrating workflow systems, 3D visualization and a remote query engine, which studies the impact of climate change in an estuarine environment.

However, despite an exhaustive search of the literature, it was not possible to identify a significant body of visual analytics systems supporting predictive estuarine hydrodynamic modelling in relation to flood prevention and coastal zone management, although as noted above, visual analytics systems have been developed as integral elements of post flood disaster management systems.

Furthermore, many of the examples cited for EDSS relate to the much studied areas of 3D data, visualized through 3D volume visualization and flow visualization. There were comparatively few examples of 2D modelled data studied through interactive visual analysis (yet the researchers preferred to use 2D models, when undertaking the studies which provided the data for this thesis).

In general, in all the domains discussed here, (not just EDSS) where 2D data was modelled, for example with TELEMAC-2D, they were generally studied through 3D visualization [185, 281, 283, 321], or by using the visualization capability of

²⁸<http://earth.google.com>

the simulation system and the limited visualization capabilities of analysis systems such as MATLAB (examples include Carrivick [53] and Alho and Aaltonen [4] for glacial flooding; Lam et al. [178] and Paradis et al. [218] for estuarine operational storm surge forecasting).

3.6 Multi-disciplinary ocean observatories and holistic regional studies

Throughout this thesis, the state of the art in ocean science visualization will be seen to be closely associated with extremely large scale projects such as ocean observatories and holistic regional studies. These Environment Observation and Forecasting Systems (EOFS) are a new class of large scale distributed system designed to monitor, model, visualize, analyse and forecast the physical processes within a wide area [276], for example an estuary and its associated river system, or a particular coastal area. Examples include: the Center for Coastal Margin Observation and Prediction (CMOP formerly CORIE), whose studies are centred on the extremely large Columbia River Estuary [23, 136, 148]; ECOOP the European Coastal Oceanography Project [100] – which are discussed in more detail below. These projects adopt an holistic approach to multi-disciplinary, multi-source data acquisition, pre-processing, analysis and visualization, encompassed within workflow systems, of which the visualization element is just one element.

Such large scale studies involving collaborative research across disparate fields have been made possible by advances in computing and analytic techniques and widespread connectivity, together with the emergence of tools such as Google Earth and Google Maps, which have transformed the way geospatial data is presented on the internet [100].

CMOPS, studying the massive Columbia River estuary is one of the longest established ocean observatories, and out of this has evolved systems, which represent the current state of the art in estuarine hydrodynamic visualization. Dataflow pipelines are widely used in these systems [302], developed with visualization systems such as AVS, SCIRun, and VTK based systems such as Paraview, VisIT, VisTrails and DeVide. A more recent estuarine application is Howe et al's. [136] workflow system and remote query engine for visualizing fisheries data integrated with observed and modelled oceanographic data to create a visual analysis tool, enabling researchers to explore relationships between ocean conditions and fish distribution

and to gain an understanding of a complex salt wedge.

The Collaborative Ocean Visualization Environment (COVE) [111] supports a regional ocean observatory, encompassing multi-disciplinary deep ocean and coastal shelf studies of sensed and modelled data. The approach here is different to that of CMOPS, in that the goal was not to provide an extremely complex visualization system, but one which domain scientists and engineers building a sensing array might find intuitive and usable. Existing geo-browsers and web browsers, also off the shelf visualization systems, did not provide the required levels of functionality, so a turnkey system was developed using C++ on top of an OpenGL graphics interface, with a cross platform interface built using the Fast Light Tool Kit (FLTK) and incorporating shared storage and access systems for collaborative working. A quadtree is used to deal with overplotted data.

ECOOP (the European Coastal Sea Operational Observing and forecasting system Project) [100], provides distributed data, modelling, analytics and visualization services through a web portal. This visualizes and compares physical (and biological) modelled and sampled/sensed marine data from 23 different modelled data feeds and one observational source (merging data from many sources), using GODIVA2 (a web-based system for visualizing and exploring 4D terabyte scale scientific data) [39], at interactive speeds, thus supporting visual analytics. The system enables comparison of numerous datasets and identification of features of interest, through its interoperability with other data visualization systems. Together with ECOOP, GODIVA2 is also used by the MERSEA (now MyOcean) system.

Other similar visualization/workflow systems for ocean observatories include NANOOS [5], and LOOKING [55].

Clearly, with these extremely large scale projects, the demands on system architecture and on the visualization systems themselves are of a significantly greater magnitude than the current research for this thesis, nonetheless, they may provide direction to the development of an appropriate visualization and visual analytics system for this research, and for its future development.

3.7 Workflow, including data management, system design and architecture

While the focus of this research is on the visual analytics, this section sets it in the context of the wider visualization pipeline, considering how to achieve a balance be-

tween a number of often dichotomous factors in providing the estuarine researcher with fast and usable visualization and visual analytics systems for extremely large, modelled, unstructured, AMR multi-field datasets.

Factors include speed of operation; rendering quality (whether journal quality is required); maintenance of fast, stable rendering speeds for interactivity, and the need to provide a range of views, plots and graphs, which will support data analysis and knowledge discovery (which in themselves are computationally intensive). This complex balancing act has to be achieved in the context of the datasets which present many challenges to effective visualization. Visual analytics solutions will be discussed in Chapter 4, but other solutions lie in system design and architecture, which for the sake of completeness are briefly discussed here.

The preoccupation with these issues is not new: as the research examples show, the almost unending exponential increase in the size and complexity of datasets continues to leave both hardware and systems, including visualization systems, hard pressed to keep up with them. Even before the evolution of visual analytics techniques, researchers were seeking means of improving speed and performance, and achieving stable interactivity. Ocean science examples include Coelho et al. [63], Chao et al. [54] and Park et al. [219], all adopting a parallel processing approach. Distributed/grid computing solutions include Patrikalakis et al.'s [221] dynamic workflow coupling data sensing; distributed computing; information retrieval and management and visualization interfaces.

3.7.1 Recent developments

This review focuses on developments post 2005, the generally acknowledged date when visual analytics solutions evolved. As a principle informing the development of numerous visualization systems, Haber and McNabb's [113] visualization pipeline paradigm, of filter/map/render has stood the test of time, and is still valid today, in achieving concurrency with large scale data [203]. However, there is a growing recognition that its structure does not support the extreme concurrency required for exascale computing, requiring the data to be broken into billions of partitions [36]. Thus, visualization system developers for domains such as ocean sciences, which generate massive datasets, are increasingly concerned with the development of innovative new hardware and software architectures to manage the increasing data intensity, particularly as much of the work in these domains involve simulations for predictive purposes, which require interactive rendering speeds.

For extremely large datasets some researchers are pursuing methods which enhance the speed and quality of rendering to achieve greater perception, through the use of commodity graphics cards and algorithms to improve classification of data, lighting and compositing [191]. Others are continuing to focus on parallel visualization: Ma et al. [191] and also Moreland et al. [203] DAX system, which proposes an alternative to the traditional visualization pipeline, presenting algorithms with ‘pervasive parallelism’ for extreme scale visualization. However, these solutions do not acknowledge the fact that much research takes place on more limited desktop systems. iRun [301] focuses on the challenges presented to visualizing large volumes of unstructured data of the type associated with this project, for which an algorithmic solution in conjunction with distributed hardware accelerated, parallel volume rendering is proposed which enables a standard desktop PC to pre-process around 36 million tetrahedra in about an hour, and achieve interactive rendering.

A further approach is Bernholdt et al.’s. [32] component architecture, which proposes a ‘plug and play’ environment, whereby the developer selects useful components required to solve the visualization and analytics problem. Childs et al. [58] has proposed a modular approach, where components for a visual analysis system may be operated in any of four architectures. Similarly, Koop et al. [175] has developed a database of visualization pipelines to enable developers to select the optimum system, using whole or part pipelines.

3.7.2 Workflow systems

However, many researchers are adopting a more holistic approach: Grochow et al. [110] concludes there should be greater flexibility in how and where visualization pipelines are executed and they should incorporate many of the data abstraction, querying and manipulation roles (i.e. visual analytics) traditionally associated with data management. The case is thus made for a close integration of visualization system, analytics tools and data retrieval. In this respect, workflow and similar optimisation systems are increasingly used in visualization and visual analytics systems for extremely large scale projects, which also incorporate the further development of parallel processing.

Workflow systems encompass the whole pipeline, from modelling, through data input and pre-processing to visual analytics, visualization and display, and also consider hardware and architecture. Typically, they adopt a client-server approach (powerful servers provide parallel processing, but cannot enable interactive visu-

alization, which is provided by a desktop server), with grid, cloud or other systems providing resources. At the same time, they optimise performance at all stages of the process.

Examples of workflow systems include Childs et al. [58] VisIt, which the developers believe is able to handle datasets up to billions of elements and which adaptively optimises performance at every operation; Brodlie et al's. [45] which proposes a more focused adaptive infrastructure, optimising usage of the underlying structure to achieve interactive rendering speeds; Benzaken et al. [30] Ediflow, and numerous others, which have not been tested in the oceanographic domain.

Grochow et al. [110] argues the advantages of workflow systems are that they enable reasoning about computational tasks to be undertaken visually, rather than as scripts; they aim to produce reproducible research, which is essential for testing hypotheses (easier to exactly replicate work) and allow tasks to be performed on a variety of platforms.

3.7.3 Workflow systems in ocean sciences

Interestingly, the domain of ocean science visualization, which has tended to lag behind in the use of cutting edge, advanced visualization and analytics systems is at the forefront of the use of workflow systems. Over the past decade, with the inception of the huge, multi-disciplinary ocean observatories, operating in estuarine and coastal shelf waters, (discussed in Section 3.6) there has been a requirement for sophisticated systems to set up the infrastructure and then manage the extremely large, multi-disciplinary, multi-source, distributed, collaborative projects which incorporate both sensed and modelled data. But how far the techniques used in these observatories trickle out to the wider coastal research community remains to be seen, and certainly, the research in this chapter and in Chapter 4 suggests that advanced visualization and visual analytics systems are still pretty rare, particularly in smaller scale coastal shelf and estuarine studies.

The extremely large Columbia Estuary ocean observatory [136] exemplifies the current state of the art for large scale estuarine analysis and visualization, although it has to be recognised that not all researchers will have access to resources such as these. A provenance aware workflow system, 3D visualization and analysis system and remote query engine are integrated with both modelling and sensed data input. Similarly, Grochow et al. [111] COVE provides a flexible workflow system, which assigns computing resources dependent on the scale and complexity of the

research project: thus data analysis pipelines span a variety of platforms and may be dynamically adapted to differing needs. The system was developed in collaboration with ocean science researchers to provide a visual environment to meet all their analytical and presentational needs.

Already mentioned, other ocean observatory workflow systems include; the web-based GODIVA2 [39], in conjunction with ECOOP [100], the European coastal oceanography model and visualization system, and LOOKING [55], which describes a design for ocean observatory cyber-infrastructure encompassing sensor networks; models; power/compute/storage/ visualization grids and retrospective informatics and analytics services which couple real-time interactive sensing networks with predictive modelling.

3.8 Summary

This chapter provides the oceanographic visualization context underpinning the related work of Chapter 4, which is concerned with the development of an appropriate visual analytics system meeting the needs of coastal shelf and estuarine researchers, and the attributes of the data. This chapter focuses on the current state of evolution of visualization for the ocean sciences (both deep ocean and coastal shelf and estuarine studies), with a view to identifying avenues of approach for the related work, and also discusses issues of system architecture, including workflow systems.

Following a brief overview of the range of visualization development environments available, it groups and discusses examples of oceanographic visualization under a number of headings, identifying their strengths and weaknesses and their possible relevance to this thesis. The headings are bespoke, turnkey systems; deep ocean visualization; coastal shelf and estuarine visualization; hydrodynamic predictive studies (Environmental Decision Support Systems, EDSS); Ocean Observatories; Workflow Systems.

The conclusion is that while the development of extremely large ocean observatory systems have certainly provided a clear motivation to researchers to advance visualization and visual analytics techniques for coastal shelf studies, notwithstanding this, coastal shelf hydrodynamic visualization in its own right still remains relatively immature. In addition, despite the widespread development of Environmental Decision Support Systems, these focus on event management, but there is also a need to support predictive studies to inform the development of strategies for flood prevention and associated coastal zone management.

The review identifies the challenges that visualization systems face in handling the type of data associated with modelled, multi-field hydrodynamic data, where many types of system (for example GIS based visualization tools) were unable to handle the multi-resolution, unstructured nature of the data. It concludes that no single system can be seen to convey an outright benefit: all have advantages and disadvantages in dealing with such data, and discusses the use of turnkey systems, encompassing a range of tools and systems, including generic and bespoke tools, to resolve these challenges.

Finally, whilst acknowledging the focus of the thesis is visual analytics, the review briefly discusses issues of system architecture and the evolving use of workflow management systems to optimise performance and support big data visualization and analysis.

CHAPTER 4

Related Work

The thesis focuses on the little studied area of interactive visualization and visual analytics for modelled coastal shelf and estuarine hydrodynamics [15, 121]. It adopts a user-focused, collaborative approach by working with ocean scientists to identify analytical capabilities which would support enhanced knowledge discovery, from which new visualization methods and a visualization infrastructure might be developed. This chapter identifies and discusses the related work underpinning the thesis, and includes:

1. **Collaboration with ocean scientists,**
2. **Oceanographic visualization and Visual Analytics**
3. **Large data management & Scalability**
4. **Interaction and Interfaces**
5. **Coordinated Multiple Views.**
6. **Multi-field visualization and analytics techniques.**
7. **Requirements** of an investigative visual analytics system for coastal shelf and estuarine studies.

4.1 In Collaboration with Ocean Scientists and Computer Scientists

In Chapters 1 and 2, the importance of coastal and estuarine flood simulations in relation to sea level rise, in areas of ecological importance or large populations, was identified. At the same time, it was noted that the visualization techniques used for displaying and studying the simulation outcomes are limited [120, 121] and some analysis packages perform poorly, particularly for very large data volumes [62]. Indeed, Hibbard et al's [132] ten year old conclusion, that the functionality of visualization for earth system sciences is limited, still applies today in the coastal shelf domain, where the ability to drill into the data to identify the trends, relationships and correlations essential for predictive studies is still not widely available.

Chapters 1 and 2 also discussed the general need within sciences, including geosciences, for more advanced visualization and analytic techniques, consequent upon the exponential increase in the size and complexity of the databases requiring study. This encompasses new methodologies to relate data to other datasets; to fuse datasets; to investigate different models; to show and analyse areas of interest in different ways, and to perform complex calculations and do this in both time and space. In fact, this closely correlates to the 'shopping list' for single system attributes identified by Hibbard et al. [132]: static 1D, 2D, 3D; state of the art visualization; high levels of interactivity through an advanced interface; data import for all major file formats; fast, high quality processing of extremely large datasets; analytics functionality; publishable quality output; platform independence and freely available.

Ocean science visualization is a large discipline, and although there are common factors, the challenges facing coastal shelf and estuarine visualization are very different from deep ocean studies in aspects such as scale, number of fields, complexity of the domain, among others. A review revealed a bewildering array of tools, systems and analytic techniques used to process, visualize, interact with and analyse the data. However, as the focus of the thesis is on the development of useful and usable visual analytics systems to support big data analysis, and not the visualization programming environments themselves, this related work does not incorporate an extensive analysis of visualization systems and tools and their use in the oceanographic and coastal shelf domain. Nonetheless, as Chapter 3 demonstrates, such a review has informed the thesis, as the development of appropriate and usable visualization and analytics tools for the work has required not only a

considerable understanding of the nature of the data, user requirements and visual analytics techniques, but also an appreciation of the strengths and limitations of the available software and tools to establish a suitable development environment.

Visualization has long been used in ocean science but, until comparatively recently, primarily for display. Animation displays the temporal aspects of the data [74], and hedgehog and vector plots, contour plots, and statistical visualization techniques such as line graphs and histograms [72, 149, 319] are also used. However, these techniques are difficult to apply to support detailed analysis of increasingly large data generated for predictive simulation. Figure 4.1 illustrates this: Watford et al. [312] combined fly-through techniques using VRML models which show general trends, and enable non-technical audiences to understand the major features of an undersea canyon, but they do not support data analysis.

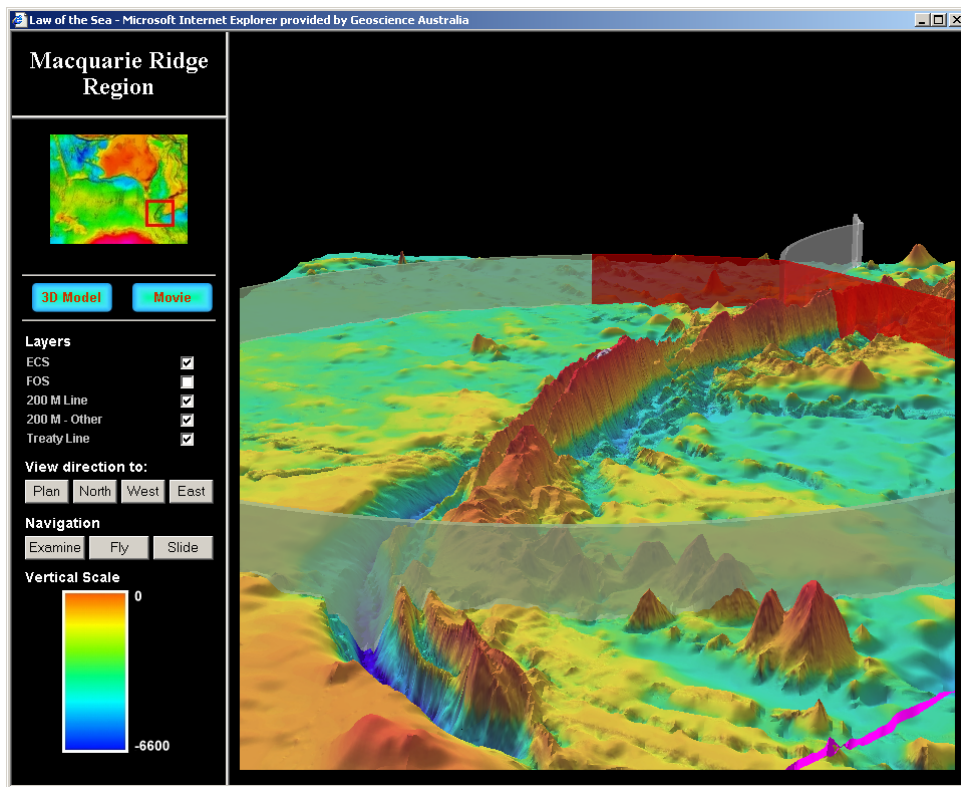


Figure 4.1: Snapshot of Macquarie Ridge region, undersea canyon between Tasmania and Victoria Australia, produced for display not for data analysis. The VRML model and visualization allowed scientists to display the connectivity and continuation of various features to support jurisdiction negotiations. Reproduced from Watford et al. [312]

Furthermore, there are severe computational and visual-computing challenges

to data-intensive discovery using modelled data. The models are complex and take time to compute, whilst unstructured grids and multi-resolution data add to the widely recognised challenges of visualizing and studying multi-field output. Researchers need to compare and contrast different investigations and parameterisations of the data to develop predictions, but many visual depictions are static plots: for example, one or two transects or specific sample points are created from the data and displayed in a graphic using MATLAB (for instance) [208]. This technique does not easily support exploration of different views of the same data, provide an overview of the whole dataset, or allow several datasets to be explored, contrasted, and relationships established.

Ma et al [192] reflects on the need to harness the knowledge of domain researchers (in this case ocean scientists), in developing visualization systems, to ensure the system is able to identify those interesting subsets of information essential to knowledge discovery. In the case of this thesis, discussions with coastal shelf researchers identified a number of issues relating to visualization and analysis of modelled estuarine hydrodynamic data: these include the visualization pipeline and architecture-related challenges centred around data processing, speed and rendering; the need to bring clarity to the data, without losing or missing important features; the usefulness of an overview of the data able to establish trends (currently they have no means of viewing the data holistically), also validation and uncertainty.

Primarily, the preliminary work identified a significant need by ocean scientists for analytics and for tools to undertake complex, detailed calculations, usually only made with some difficulty using the typical range of post-processing and visualization systems.

4.1.1 Understanding the Coastal Shelf Researcher

For a proposed visualization and visual analytics system for estuarine hydrodynamic studies to be relevant, usable and useful, it may be helpful to understand why many ocean scientists in general, and coastal shelf researchers in particular, have tended to rely on static plots, rather than taking advantage of advanced visualization systems. Although largely true until recently, the statement should be slightly qualified as a number of extremely large scale, multi-disciplinary ocean observatories now undertake large data oceanographic research, including the estuarine environment [136]. Advanced visualization and visual analytics tools are intrinsic to the workflow systems of these observatories.

Notwithstanding this, the majority of coastal shelf and estuarine researchers still rely on traditional methodologies. To place this into context: in 2000, the journal *Estuarine and Coastal Shelf Science* produced a special edition devoted to Visualization in Marine Science. Seventeen papers were presented. Focusing on the nine papers which studied hydrodynamic flow (modelled and sampled), five simply relied on the models visualization tool for visualizing and displaying the output [95, 260, 264, 281, 324]. Of the others, Wolanski and Spagnol [320] described a visualization system based on Data Explorer, but used for display only, not for analysis, concluding that visualization illustrated the outcomes more effectively and clearly, particularly to non-ocean scientists. Signell [264] came to similar conclusions.

The remaining three papers recognised the value of visualization as an analytic tool in providing solutions to challenges presented by the data. Acknowledging the limitations of the models visualization capability, and because of the difficulties of handling 3D irregularly gridded output computationally, He and Hamblin [121] used Iris Explorer to evaluate the performance of a 3D fluid dynamics model used for water quality management. The visualization system validated the model against sampled data and identified and studied areas of interest, using a multiple slicing tool and illumination (see figure 4.2) although it was recognised that comparison of up to four slices led to information overload.

Similarly, Boyer et al.'s [43] water quality monitoring system used visualization and analysis to establish relationships. A box-and-whisker plot (used as a graphical/statistical tool) supports comparison and correlation of temporal variables, whilst 2D contour maps show sources and mixing of fresh water and nutrient loading, demonstrate the impact of different processes on water quality, and illuminate causal relationships between spatial variables. The authors concluded the system provided a unique understanding of processes in the local ecosystem, particularly circulation, and also supported the questioning of relationships between variables to generate new hypotheses.

Trouw et al. [291] used visualization to examine sediment entrainment, vortex formation and advection. A 2.5D visualization of sediment re-suspension demonstrated simultaneous temporal and spatial variation in suspended sediment concentration and wave induced flow velocities, thus providing insight into the mechanisms. Animations were able to show why the direction of net sediment transport due to asymmetric waves is opposite to the direction of the highest velocities (Figure 4.3). However, some ocean scientists do use simple statistical graphs and some

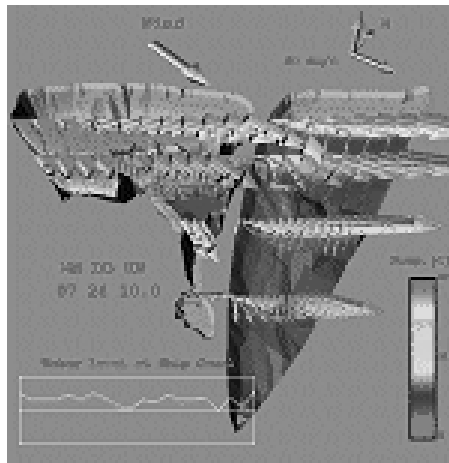


Figure 4.2: Reproduced from He and Hamblin [121]: multislice plot of modelled currents (vectors) and temperatures (colour contours) at 08:54h, 24 July 1996 with illumination

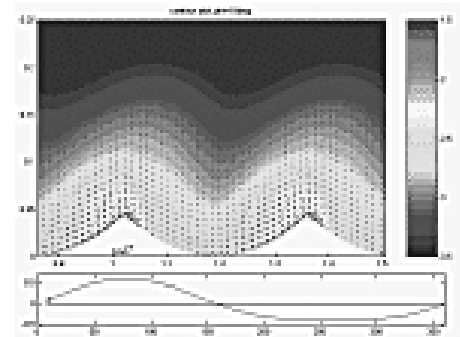


Figure 4.3: Copied from Trouw et al. [291] : Vertical suspended sediment concentration field during wave cycle (asymmetric wave)

animation [4, 53, 72, 207, 218] for example: Figure 4.4, shows model post processor and analysis software used for a TELEMAC-2D storm surge model. Analysis tools such as MATLAB or Excel are also used.

Although not estuarine studies, Alho and Aaltonen [4] visualised simulations of extreme glacial outburst flow illustrate the difficulties in comparing areas of interest through single static images alone, without the benefit of interactive, visual analytics functionality, see Figure 4.5.

“Off the shelf” visualization systems have been used in coastal shelf studies [13, 102, 103, 121, 277] and others, but they often encounter difficulties in coping with domain specific demands. This may act as a deterrent to researchers, particularly if customised tools are required, which may make the systems complex to use. Nocke et al. [208] and Clyne and Rast [62] point out the lack of numerical capabilities for quantitative analysis within such systems, and whilst these are available in packages such as MATLAB, IDLTH, or MathematicaTM, there are few closely integrated advanced 3D visualization and quantitative numerical tools. Furthermore, the need for advanced visualization techniques is not fulfilled by the analysis packages, which tend to have limited 3D visualization algorithms – indeed Chave et al [55] regards typical visualization and analytical tools as being relatively primitive, a comment which still applies today, with some notable exceptions. Additionally, ocean sci-

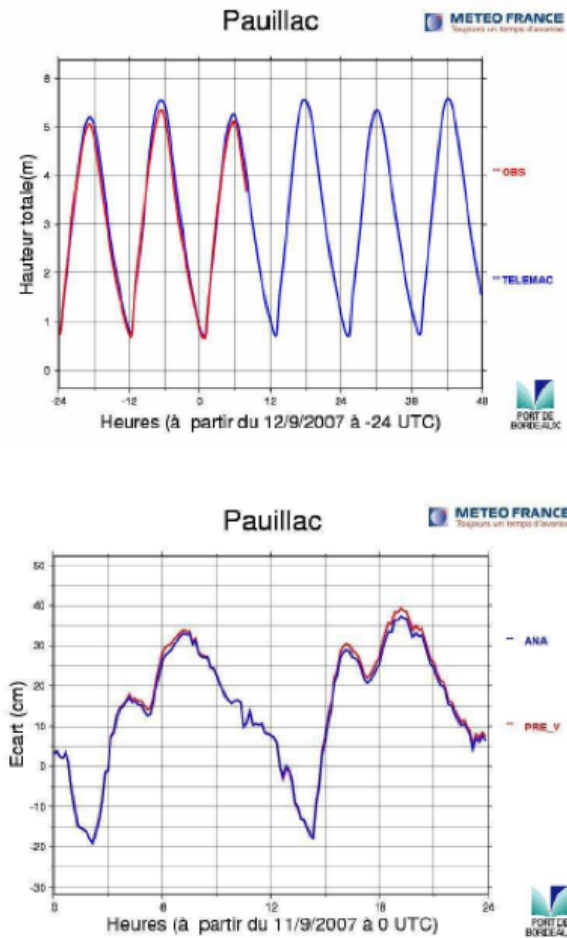


Figure 4.4: Data representation (using model post-processor and analysis software) reproduced from Paradis et al. [218], of a TELEMAT-2D storm surge model for the Gironde Estuary.

entists may also believe that the nature of their data, with their many inter-relating variables, requires high levels of domain specific knowledge on the part of the visualization developer.

However, in a domain which has traditionally relied on the visualization capability of their models (albeit of limited effectiveness), it is possible that coastal shelf researchers may simply not be aware of the advanced alternatives, particularly if, historically, collaboration between visualization researchers and domain scientists has been limited. Such an argument is considered by Johnson [152] who also wonders whether the increasing scale and complexity of the data presents such challenges that physical scientists either may not wish to, or may not be able to tackle them with their current tools. Thus the research is bounded by the limitations of

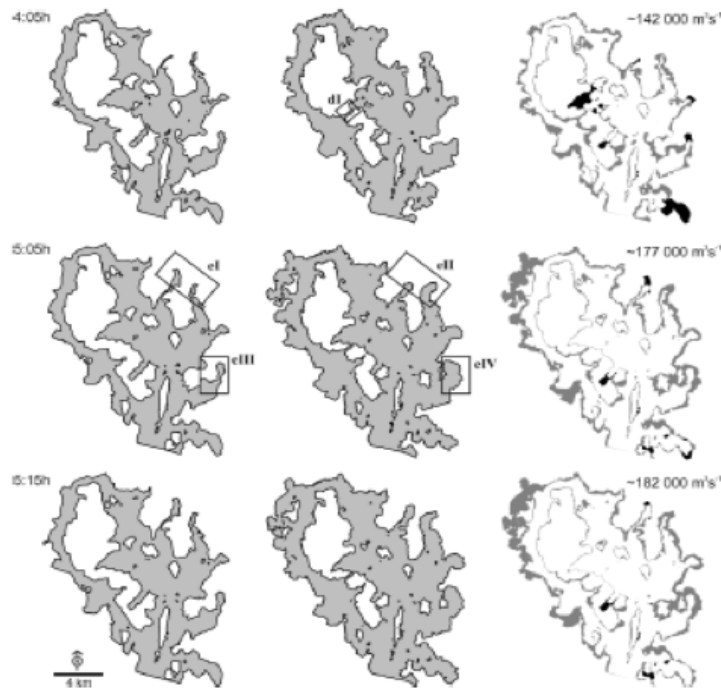


Figure 4.5: Representation of comparisons of glacial inundation simulations, showing the difficulty of undertaking a detailed comparison of the outcomes, and understanding the variances, without advanced analytics. Reproduced from Alho and Aaltonen [4] .

their current display and analytic tools.

Nonetheless, there are exceptions: Cotter and Gorman [65] make the case for a collaborative approach between ocean scientists applying unstructured grid models, and visualization researchers, to resolve the challenges to visualization and analysis presented by data from this type of model. Although there is a rich source of methods and tools for analysing model results, these struggle to handle the output from unstructured grids (such as TELEMAC-2D). Thus Ho and Jern [135], having developed a visual analytics tool for exploring volumetric data, regard the next stage as further collaboration with users to develop specific task appropriate tools – an approach this research is seeking to adopt.

4.2 Oceanographic Visualization

Earlier sections have discussed a paradox: that despite the increasing requirement to undertake big data analysis, and the availability of advanced visualization and

analytics tools, many coastal shelf researchers, (with a few exceptions), still tend to rely on traditional, but limited methodologies. Chapter 3 provided an overview of the level of evolution of visualization in ocean sciences, in general and coastal shelf and estuarine studies in particular, and based on that, this section identifies the challenges for visualization presented by multi-field data; considers the value of visualization and visual analytics for oceanographic studies, including the coastal shelf, and identifies potential avenues of approach for exploratory visualization tools integrating data-intensive visual analytics methods, including systems, tools and techniques for the data and domain.

4.2.1 Scope of the Oceanographic Related Work

There are a number of domains of study relevant to this thesis, not simply ocean science visualization. These include those where the data possess similar attributes to oceanographic hydrodynamic data, or where the hydrodynamics have a similar desirable outcome. Particularly relevant areas include geosciences such as meteorology, climate studies, geology, astronomy, and other applied physical sciences. For example:

- In meteorological and climate studies, users need to understand long time series of sensor and modelled data) [38, 144, 160, 230, 234, 271, 275, 294];
- In astronomy and cosmology researchers need to be able to separate relevant data from noise to gain an understanding of the Universe, or verify their models [61, 62, 117, 138];
- In geology: Jones et al. [156] and Hanyk et al. [115] describes interactive, immersive visualization and analytics systems for studying multi-scale data, using commercially available software, whereas Billen et al. [37] utilises Visualizer software designed specifically for an immersive VR environment, and for supporting analytics, such as feature identification and quantitative analysis within that environment.

Lipsa et al's. [187] review of Visualization in the Physical Sciences considers a range of physical science domains and the reader is directed to this work, to discover the breadth of visualization techniques within related domains, many of which are concerned with fluid flow analysis, in common with oceanographic visualization. As fluid flow visualization is not the main focus of this thesis, an in-depth review of

its techniques is outside the scope of this work. However, some examples of fluid flow research are mentioned, where they may have an application for ocean hydrodynamic visualization. For example, of relevance to coastal shelf visualization, and to the case study data for this work, is the visualization of micro-hydrodynamics, where in recent years, relationships between small scale topography and larger scale physics have been questioned [68].

For a comprehensive review of the state of the art in flow visualization and current research issues, the reader is referred to Peng and Laramée [224], which discusses research issues for flow visualization techniques as does [223,224] and Ware et al. [311]. Of particular relevance Peng et al. [223,224] considers visualization of flow on surfaces in 2.5D is still a challenge, (as is interactive, 3D spatio-temporal visualization and data exploration [135]).

Equally, volume visualization algorithms have been used in ocean science to display vortices and eddies, and other glyph-based visualization techniques have been used to represent the twists and turns of the water column. However, this thesis is not concerned with processes within the water column, thus, whilst interesting, volume visualization systems and techniques, are not studied in detail. Useful visualization techniques also include line charts, plots, the use of colours, all to aid quantitative analysis of the simulation data.

The focus of the thesis is on velocities, and the generation of aggregated and averaged values by time, which are then displayed as graphs or maps, and animated to demonstrate change. The model data is 2D, and depth-averaged velocities enable 2.5D representation and calculation, avoiding the distortion which may result from 3D representation of 2D data.

4.2.2 Multi-field oceanographic data & the challenges they present to visualization

This thesis relates to multi-field data: Fuchs and Hauser [94], in their comprehensive review of the visualization of multi-variate scientific data, discuss data types, describing multi-field as data where multiple attributes are grouped into multiple physical fields, and each component of the field requires the other components to be interpreted. Thus, development of visualization and visual analytics tools for these data types with their complex inter-relationships, require an understanding of the inherent attributes of the datasets generated by the simulations (which in the case of this work are multi-dimensional, multi-variate, multi-resolution, spatio-temporal).

This process is described by Kirby et al [171] as breaking down the data into components; exploring the relationships between them and visually expressing both the components and the relationships. Brooks and McDonnell [47] expand on this: by seeking spatio-temporal relationships and patterns, visualization provides more intuitive analysis in a data-rich environment, where the concern is with dynamic processes and the interaction and linkages of many multi-scaled entities.

Spatio-Temporal data

Spatio-temporal data are ubiquitous in the real world no more so than in the geosciences. Their attributes, the benefits they convey, the challenges they present to data analysis, and appropriate visualization and visual analytics methods have been widely researched. Unsurprisingly, there is an abundance of literature.

Aigner et al. [2] and Andrienko et al. [13] discuss the attributes of time at length. Time is considered to be an “outstanding dimension”, pointing to Schneiderman’s [259] identification of temporal data as one of the seven basic data types, yet time is still regarded as a common quantitative primitive, rather than a special dimension in many visualization systems. Time is categorised as having two primitives (discrete time points and time intervals, which use different sets of temporal relations) and also as being linear, cyclical (e.g. the tidal cycle) or branching – the last mentioned introduces additional variables to a model, and is the attribute which facilitates alternative predictive simulations.

Other important characteristics are whether there are single or multiple variables for each temporal primitive, and whether there is a spatial dimension. All these factors must be considered when developing the visualization and analytics system, particularly in relation to its validity, as techniques used to visualize or study one time characteristic may not work for another.

The data’s spatial dimension is governed by the fundamental concept of spatial dependence: everything is related to everything else, with near things more related than distant things. Consequently, there is either positive or negative correlation between data points at proximal locations, which equally applies to temporal data points, an important factor when applying statistical analysis to spatio-temporal data, as standard statistical techniques may not apply. Thus, spatial and temporal components of data are inextricably linked [13, 157], but the number of variables means that a common framework for visualizing time and space is unlikely to meet the full needs of researchers, in view of the differing ways the attributes of the

data need to be represented, or areas of interest identified and studied. The combined attributes of spatio-temporal data present the greater challenges for visualization [163] and one of the goals of visual analytics is to provide insight by creating systems which encompass the heterogeneous needs of the data.

Nonetheless, although challenging to visualize, both Aigner et al. [2], and Andrienko et al. [13] consider the similarities between the two data types present analytic opportunities to understanding relationships. Temporal components permit predictive analysis: thus ‘what if’ scenarios may be studied to develop and investigate hypotheses supporting prediction and management in applications such as natural disaster management, flood prevention etc., where it is acknowledged that visualization enables analysis of both spatial and temporal data components [3, 11]. Examples include flood and wild fire disaster planning and management system [41]; dike monitoring systems for The Netherlands [214]; hurricane prediction and management [157]; river flooding management [80]; urban disaster management [118], and interactive visual analysis for paleoceanography [284], which reconstructs and analyses climate and ocean conditions over many thousands of years to identify climate change trends.

A further opportunity is that their geo-location naturally permits interpolation, enabling gaps in the data or incomplete data to be rectified – a frequent occurrence with modelled hydrodynamic data [157, 168, 283]. Additionally, multivariate data may be integrated by using references to space/time and spatio-temporal inferences made [13].

However, the significant challenges should not be ignored. Geoscience spatiotemporal datasets are very large, both data items and observed attributes, yet the limitations of screens displaying the data mean it is impossible to present more than 10^7 records on a screen [162] necessitating careful selection of data scaling methods, which do not lose essential characteristics, or relationships.

Variates of spatio-temporal data are highly interconnected, with vast ranges of variability in the hierarchy of scales: in oceanographic data, spatial variability ranges from microscopic to thousands of miles [13, 168], similarly, temporal variability ranges from fractions of a second to many thousands of years [20, 28]. The spatial hierarchy is particularly relevant to estuarine studies, where a smaller scale area, such as a river, may be nested within a larger one, or small (and dynamic) features such as tidal channels, sand-flats and scour pits may also appear within the larger scale estuary.

This requires tools capable of accurate analysis at multiple scales, and is an ac-

knowledge challenge within coastal shelf and estuarine visualization and analysis, where the reliability of scaled up or scaled down analytics have been questioned [47] – one size does not necessarily fit all. Patterns or relationships detected at one scale or over a period of time, may not be identified at different ones, or may even be totally opposite to the original relationship. Furthermore, as Andrienko et al. [13] reports, this is further hindered by a lack of any systematic method to detect the scale (both in space and time) at which a feature undergoes change.

In other geosciences, the multi-scale complexity of the spatio-temporal range (between two to five orders of magnitude) challenges the accuracy of cosmological simulations [138], requiring both quantitative and qualitative methods of study, through visual analytics. Similarly, understanding hurricane structure presents significant challenges because of the difficulty in analysing both spatial relationships between hurricane features in a single time step and also temporal relationships [157].

A related problem is the need to integrate data measured at different scales, from different sources (sampled or modelled) – for example time intervals. Here Andrienko et al. [13] argues for linked data initiatives and open data APIs to resolve this problem, but believes that current solutions are not adequate. The Kemp and Meaden [170] multi-purpose visualization tool for fisheries’ management tackles this problem: it integrates data from multiple sources, to which different spatial representations are applied for different user groups. Additionally, in recent years, the large ocean observatories referred to in Chapter 3, which integrate multi-disciplinary, field observation with modelled cross-scale data, have implemented linked data initiatives to study specific problems. For example, Howe et al. [136] describes a visualization and analytics system, which integrated fisheries data with oceanographic observations and model results to understand how physical variables, probably linked to climate change, might have an impact on fish populations. Baptista et al. [24] and Grochow et al [110, 111] discuss the visualization and data analysis environment in greater detail.

Earlier, dependency was described as an opportunity but it is also a challenge. Dependencies between data are weakened by the heterogeneous nature of the space. However, estuaries have numerous complex, dynamic, evolving boundaries (and thus dependencies): land/sea; man-made structure/sea; seabed/water column; tidal channels / sandflats; fresh/salt water; between waters of different temperature and salinity and numerous others. Unsurprisingly, Baptista [23] describes estuaries as ‘poster cases’ for the need for new paradigms, because of their strong spatio-temporal variability at multiple scales. Whilst some of the boundaries may be mod-

elled into hydrodynamic simulations [238] in such a domain, further visual analytics techniques may be needed to support boundary identification and study.

Multi-Variate Data

Geoscience studies generally require detailed understanding of the relationships and interplay of the data variates, in conjunction with their spatio-temporal dimensions, but as discussed in Chapter 2, in relation to the case study data, estuarine studies are even more complex than other areas of ocean science, with numerous inter-connecting variates. Earlier, Howe et al's. [136] study of changes to estuarine fish populations was discussed in the context of integration of disparate datasets. However, this study is also a good example of estuarine multi-field data, with variables including temperature, depth, salinity and hydrodynamic flow, over time where with the addition of various attributes, scalar data rapidly becoming high dimensional, necessitating filtering to reduce the number of variables to enable features of interest to be identified.

In the light of such complexity, a number of key questions need to be resolved in developing visualization systems. Fuchs and Hauser [94] identify these as:

1. How to capture all the relevant information;
2. How to combine quantities of information belonging to the same locations in space;
3. How to identify quantities that may be derived from the data to support understanding;
4. How to perform feature identification (both automated and interactive);
5. The need to identify, balance and combine visualization techniques to meet the needs of each data type.

These issues are, of course, common to spatio-temporal data, where several attributes are contained in the same point in space, and the scale of these time-dependent datasets means that visualization approaches alone will not achieve the rapid interaction required for exploratory analysis [20]. However, the representation techniques used for analysing spatio-temporal data are very different from the attribute graphs which apply to multi-variate data – hence the requirement for a visualization and analytics approach which integrates different methods of exploring the data.

4.2.3 The value of visualization and visual analytics for ocean science data

The section considers the role and value of visualization and visual analytics to the type of data associated with ocean and coastal shelf research and provides pointers for approaches in developing an appropriate visual analytics system for the case study data.

Visualization for Ocean Science Data

Brooks and McDonnell [47] believe visualization provides more intuitive analysis in data-rich environments such as earth sciences, where the concern is with dynamic processes and the interaction and relationships between many multi-scaled entities. Visualization seeks spatial and temporal relationships from the data and reveals the patterns within, supporting new insight and inference. Similarly, in a study of vortex detection in both coastal shelf and deep ocean hydrodynamic flows, Sadarjoen et al. [247] expresses the importance of visualization for understanding the underlying physics, and for simulating and modifying designs. Andrews [9] argues that visualization also facilitates multidisciplinary studies through pattern detection: data analysis techniques between biological, geological and hydrodynamic data may be very different, but visualization's more intuitive medium enables scientists in the different disciplines to recognise and understand spatial relationships. However, this has to be set in the context of the large data challenges discussed in earlier chapters.

Taking meteorological studies as a case study: both modelling and remote sensing generates continuous streams of complex spatio-temporal data, which existing atmospheric tools struggle to handle. Thus, Song et al's. [271] useful overview of meteorological visualisation, cites the value of systems such as Visad [131, 134] or VIS-5D [133, 289, 290] in resolving the limitations of atmospheric tools. However, the size of the datasets is such that only small sub-sets may be visualized, which affects analytical capability, and relationships within multiple fields are undiscovered or unrecognised.

Furthermore, the systems may not permit interactive and flexible exploration of the data. Even though volume rendering is routinely used in visualizing atmospheric data, systems have not been able to keep pace with the increasing volume of remote sensed data, and the greater complexity and sophistication of simulation models (3D cloud simulations are common) [157, 271, 294]. Coastal shelf and es-

tuarine hydrodynamics research has not been at the forefront in the application of visualization systems, and these comments apply equally to it, by the very nature of the data being studied.

Nonetheless, advanced, traditional techniques continue to be applied within geosciences and Janicke et al. [144], in relation to climate variability studies, suggests that most systems still rely on conventional visualization techniques. For example, ocean science and meteorological systems focus on immersive interfaces to aid understanding and analysis including Keen et al. [159] – studies of complex spatial and temporal variability); Lu et al. [190] – hurricane evolution; Wilde et al. [317] – CUMULVS – Collaborative, User Migration, User Library for Visualization and Steering.

However, even these more advanced visualization techniques are limited. The GODIVA system [39, 40] (Grid-based Diagnostics, Interactive Visualization and Data Analysis) for ocean circulation and climate modelling, proposes an architecture/hardware based solution for interactive desktop visual analysis and Web services through traditional visualization tools. Whilst conferring a number of benefits, such as speedy initial processing and some data manipulation /calculation, the system does not provide the degree of interactivity, feature identification, and detailed interrogation and analysis possible with a full scale visual analytics system. Similarly, Billen et al's. [37] Visualizer software supports VR, including immersion, through CAVE and GeoWall, as well as desktop use, to interact with very large datasets that cannot easily be viewed or manipulated by other means. Although in the immersive environment the system supported feature identification, and improved the speed and effectiveness of knowledge discovery, there were interactivity problems in applying a system developed for a 3D VR environment on a 2D desktop, which made the approach less effective for desktop analysis.

Turdukukov et al's. [294], convective cloud studies, also highlights the big data problem for the interactive animation of large series of spatiotemporal data, traditionally regarded as the only way to explore these data. Here, information overload means that features may be overlooked.

Thus although visualization confers benefits to ocean science data analysis, there are still unresolved challenges in managing the data to enable understanding and knowledge discovery. One of these challenges, visualization of the spatiotemporal attributes of the data, is discussed in more detail below:

Visualization tools, techniques and systems for spatio-temporal, coastal shelf data

The material in the preceding sections suggests that creating a usable and useful visualization system which accounts for all the attributes and complexities of the data is a demanding process, with visualization of the temporal component providing the greatest concern [13]. In understanding the evolution of their data over time, researchers compare data located at different points of the time axis to detect trends and patterns [2, 20, 56]. To support this process, many of the widely used visualization systems for scientific visualization, generic visualization frameworks, and analytic techniques (derived mainly from information visualization) have been successfully applied to temporal data. For simple data analysis, these have been seen to outperform specialized systems or techniques, because of their ease of use. But often visualization tools and methodologies for representing time are applicable only to specific categories of temporal and other data. A small sample of examples include SimViz [76], for simulated data; TimeWheel's axis based system visualizes multi-variate data focusing on temporal dependencies [288]; ThemeRiver [119] supports static visualization of document collections but may be suitable for other quantitative data with the introduction of interactivity, however it is not suitable for branching time.

Generic approaches treat time as a quantitative dimension (mapped to a quantitative domain). Examples include XmdvTool [308], Visage [174] based on Inselberg's parallel coordinate plot [142, 143]. However, in treating time as a variable, they do not support the direct visual connection between the variables and the time axis required for interactive visual analytics, where time has a particular meaning [2]. Thus, although generic visualization tools and methods may offer greater capability than the post-processing modules of many numerical models, they may fail because they do not pay sufficient regard to the nature of the data and its requirements. For example, Grochow et al. [111] reports on the limited ability of geo-referencing browsers to provide data manipulation or analytical capabilities, which also may not be able to meet the challenges of such large and complex datasets [56, 187].

In the absence of a single visualization framework to deal with all data types and attributes, and the limitations of generic frameworks, there is an argument that spatio-temporal data require a special approach [13], based on understanding the characteristics of the time axis, what is to be analysed and how it is to be repre-

sented. Aigner et al. [2] describes this as the time, data, representation mantra, where any visualization interface studying temporal data needs to include specific facilities for temporal queries [16]. With generic systems they are mostly built around linear time, sometimes cyclical. Figure 4.6 shows the importance of selecting an analysis method which reflects the attributes of the temporal data. The cyclical time data is better represented in the spiral view rather than the linear plot [13].

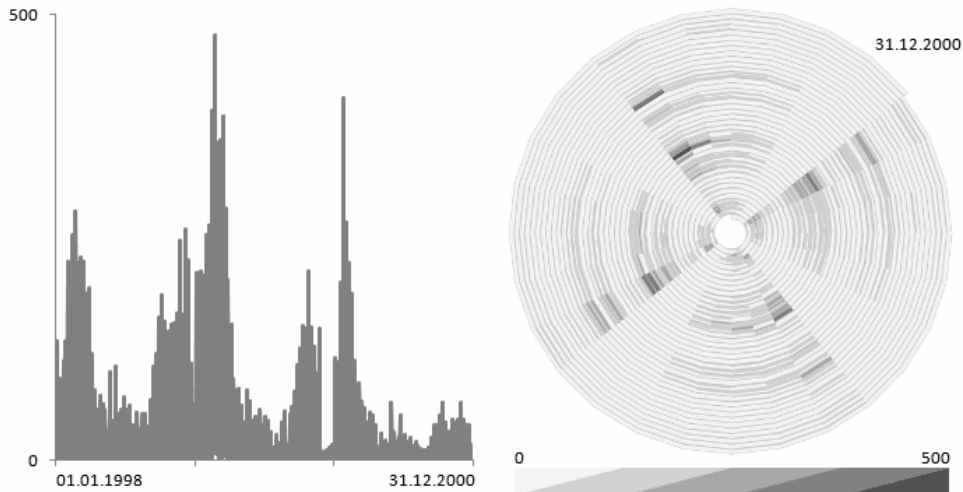


Figure 4.6: copied from [13] showing the difference between linear and cyclical spiral representation of cyclical data, to illustrate a weekly pattern.

However, branching time is also of interest for this work, in enabling predictive analysis, but techniques for representing both this and multi-perspective time are limited in their ability to represent multiple time-dependent attributes or combine with spatial visual representations, thus necessitating more research into techniques to visualize and analyse branching data [2].

While there has been much discussion about temporal data, the spatial component is also important: the scale of spatial analysis is measured in the size of units of measurement, but as already discussed earlier, patterns or relationships noted at one scale may not apply in another [13]. Thus scale variance is an important property of the data, and effective visualization depends on the scale of the analysis matching the scale of the phenomenon being studied, whilst at the same time reflecting the user's research needs.

These factors, therefore, strengthen the case for an application-based, task specific, data-focused approach to the development of a suitable system, which strongly

reflects the research needs of the user – an approach which may well become a paradigm for handling spatio-temporal, multi variate /dimensional data. This brings an increasing focus on synthesising scientific visualization and information visualization techniques, applied interactively through visual analytics.

Visual Analytics considerations

Visual analytics is an application-orientated discipline [165], responsive to users' needs to understand very large datasets, (categorised by Fua et al. [93] as those comprising approximately $10^6 - 10^9$ data elements). The technique synthesises interactive visual surfaces with analytical reasoning, using scientific visualization and information visualization techniques. As discussed in earlier sections, the extremely large, multi-field datasets typical of coastal shelf and other geosciences require a visual analytics' system with a combination of methods to enable patterns to be discovered [14], which are also capable of handling the spatio-temporal nature of the data.

Visual analytics depends on dynamic data manipulation through a visual user interface, using techniques such as interactive feature identification through brushing; linking; on-demand data derivation and focus+context or other viewing methodology. Roberts [236] provides an overview of these methodologies. These techniques support task-related, comprehensible coordinated windows for interaction, data exploration and discovery of clusters, gaps, outliers, trends and relationships [135, 146, 166, 258].

Keim et al. [167] considers the value of visual analytics for large volumes of unstructured, spatio-temporal data (typically associated with models of hydrodynamic flow), as being to bring clarity to the data. Whilst visualization of these high resolution datasets is frequently more effective than algorithmic analysis in providing insight [46], the acknowledged challenges of size and complexity, already discussed, necessitate methodologies to 'to hide complexity' [40] and to facilitate analysis and insight. The question to address, therefore, is how visualization and visual analytics may meet the identified needs of coastal oceanographers studying modelled estuarine hydrodynamic data by overcoming the challenges identified as inherent in the nature of the data.

The aftermath and security concerns of the 9/11 tragedy, drove much of the agenda (and money) for Visual Analytics in the early days. However, the evolution of geoscience visual analytics was rather more a product of Hurricane Katrina in

2005 [285].

Geovisual analytics tools support interactive spatial data exploration and analysis of very large datasets from multiple perspectives, and permit the identification of relationships across all the scales of the data [147]. But such systems and tools themselves make intensive processing demands [62, 135, 160], so the need to optimise performance cannot be ignored, if the system is to achieve interactive speeds - an essential component of visual analytics. Thus, any consideration of tools and techniques has to be balanced with architecture and hardware requirements and optimised against the needs of the analyst if the optimum balance between understanding and knowledge, and usability is to be achieved. These are not the principle focus of this thesis, which is concerned with proving the value of visual analytics.

Keim et al. [163] expresses the view that visual analytics has not produced a single best practice: rather there is a repertoire of approaches, depending on the needs of the domain and the research – thus, the scope of this research has been extended to the wider ocean sciences and geosciences, and includes techniques from flow and volume visualization. However, a degree of caution is required in selecting techniques from other domains: the geosciences and particularly, ocean sciences, experience widely differing scales, types, dimensionality, variates and sources of data. For example, in deep ocean visual analytics, one aim is to find techniques for studying relatively sparse information within a large data space. Thus, cropping may be used to visually emphasise a more interesting region [135], but this may not be appropriate in the much more densely sampled and complex estuarine environment, where cropping may cause correlations or relationships to be missed. Ho and Jern [135], when discussing the application of 2D visual interaction techniques to 3D volumetric data, also draw attention to the importance of identifying what is perceptually possible when interacting with the data. This may equally apply when considering whether the interactivity developed for 3D visual analytics systems will apply to 2D data and 2.5D visual analytics systems. Whilst not a matter of assuming that techniques will apply across differing dimensions, it is likely the principles of a system will be portable, but the appropriateness of specific data analysis techniques will need to be assessed individually, in relation to the data.

Thus, the value of studying a breadth of research, in the absence of depth within the specific area of interest, lies in suggesting promising avenues of approach which may merit further investigation, rather than obtaining definitive solutions for modelled estuarine, hydrodynamic data. The review, therefore, considers a number of visual analytics methods found to be useful for filtering, presenting and exploring

oceanographic and geoscience data, and which may prove of value for this thesis. It does not attempt a full survey of the development and current state of the art of either visualization or visual analytics, but it provides sufficient level of detail to place the current work in context, and the reader is referred to Keim et al's. [166], survey of the state of the art in visual analytics, also see Andrienko and Andrienko [14] for a more detailed appraisal.

4.3 Large data management

It is widely acknowledged that ocean science simulations typically generate extremely large datasets, including the marine and coastal environment, with their continuing increase in size and complexity presenting new challenges for domain researchers and visualization system developers. Yet an almost universally expressed requirement by researchers in all areas of geosciences (for example [118, 160, 206, 271, 284, 294, 295] is for methodologies to aid understanding and analysis of the complex relationships within their domains of study, particularly with the need to determine the implications of climate change and predict its effects.

However, the review of the data completed in Section 4.2 identified a challenging visualization and analysis requirement for ocean sciences, including modelled coastal shelf and estuarine data, arising from the size of the datasets; the numerous, heterogenous attributes of the multi-field data (including their spatio-temporal nature), and their many inter-relationships. But whilst it was identified there was a role for visual analytics in aiding exploration and knowledge discovery, many of the techniques by which this is achieved, themselves make significant computational and processing demands. For instance, Andrienko and Andrienko [12] highlights the significant computational power required for Coordinated Multiple Views (CMV) for very large datasets, similarly Clyne and Rast [62] point to the burden placed on achieving interactivity by both CMV and Parallel Coordinate Plot (PCP). Thus, an integral part of the visualization and analytics system are methods to resolve these problems.

4.3.1 Overview of data management methodologies

There are numerous potential solutions, either through manipulation of the data or through computer science techniques. Whilst most visual analysis is concerned with showing the data [293], data structures may be used to achieve fast filtering (through

both data abstraction and aggregation) to avoid issues of spatial clutter; enable the interactive visualization speeds necessary for visual analytics and allow the user to identify subsets of interesting data points. These types of solutions, which include hierarchical data structures, are the principal subject of this and the following two sections.

However to place this into context, performance may also be improved by creating fast architectures, using methods such as parallel processing algorithms; GPU programming techniques to make the rendering perform faster, or grid and cloud computing. Another method is to pre-process the data through caching techniques, and taking this idea further, the datasets can be aggregated together and categorised into parts or features. Then, performance is enhanced by merely displaying the aggregation, rather than the detail. Whilst recognising the optimum system would incorporate all or a number of the types of solutions identified above, this research focuses on data structures, particularly as an important facet of the work is to identify means of resolving over-plotting. Thus, the related work highlights scaling methodologies.

Achieving the required interactivity in the visualization of large data sets is acknowledged as a significant problem [301], particularly as some visualization solutions may not support the necessary level of understanding of the data [46]. Keim et al. [166], and Bertini et al. [34] regard scalability as a key challenge to resolve.

Data aggregation and abstraction permit filtering and manipulation of large and complex datasets, enabling the researcher to interact with the data to identify subsets of data (features of interest) for further analysis to establish correlations and relationships, and propose hypotheses. Through these processes, data may be derived for further iterative study. Data scaling is particularly relevant to visual analytics tools for spatio-temporal datasets of the type associated with this research, where correlation analysis is computationally expensive, as the datasets contain many locations and many time points [327]. Not only does filtering allow faster rendering speeds supporting interactivity, it also allows the use of tools and techniques to reduce data overload, suppress irrelevant detail and spatial clutter – which hinders insight, and aids in conveying the important detail [286].

Many data abstraction approaches have been described for multi-variate, spatio-temporal data, for example grouping techniques such as binning and clustering, which partition data into subsets displaying similarities; statistical means (calculating data characteristics); the hierarchical data structure process of spatial indexing, which moves data from the geographic location to fill the graphic space more effi-

ciently [13]; principle component analysis [154] – dimension reduction, or reducing the number of variables in multi-variate datasets to reveal trends, or wavelet analysis, which is a commonly used technique for data abstraction and data analysis, particularly in meteorology and climate studies.

Data scaling is required at different stages of the visualization and analytics pipeline. Numerous systems have been proposed for identifying features of interest through visual queries, including database exploration tools, such as VQE [73], Visage [245], FastBit [323], which rapidly identifies and visualizes regions of interest, through compression and accelerated algorithms. DEX [279] extends FastBit into the visualization pipeline by focusing visualization algorithms onto potentially, topologically complex regions of interest, defined by the query. Its application to astronomy suggests it may be suitable for modelled ocean science data, however, its development for regular gridded data, may not be useful for unstructured, adaptive mesh refinement (AMR) data.

A number of clutter reduction techniques lend themselves to combination with the attribute plots, graphs and analytical techniques used in visual analytics [38, 125]. These include brushing; focus + context; setting the parameters for data queries as intrinsic elements of the analytic tool (e.g. sliders); binning etc. Stolte et al. [280] reports on XmdvTool [308], Spotfire ¹ and Xgobi [50], which provide attribute tools such as scatterplots and parallel coordinates, augmented with interactive techniques such as brushing and zooming, but of these, only XmdvTool supports exploration of hierarchically structured data, such as that associated with adaptive mesh refinement models. However, in practice, with the extremely large data sets associated with multi-variate, spatiotemporal data, several scaling processes are often applied. Blass et al. [38], for example, extends the use of parallel coordinates to large multi-timepoint datasets, with a system which uses data quantization and compression as preprocessing, then fast brushing techniques in coordinated multiple views.

Nonetheless, despite the challenges, within many areas of the geosciences, researchers have adopted interactive approaches in conjunction with filtering and analysis. For example, within geology, field researchers are developing direct querying systems, which could be applied to coastal shelf and estuarine models. The direct-manipulation interface developed by Billen et al. [37] allows users to grab and rotate an object in the VR environment with the mouse pointer, to investigate iso-surfaces passing through any data point. Similarly, Jones et al. [156] argues for an

¹<http://www.spotfire.tibco.com>

immersive VR approach to analysis of lengthy, multi-scale geological time series, using commercially available software. In the oceanographic domain, to understand the complex spatial and temporal variability in the marine bottom boundary layer, and to study its dynamics, Keen et al. [159] also adopted an immersive approach, coupled with an integrated hydrodynamic modelling environment, encompassing 3D visualization, animation and analysis.

Increasingly, there are approaches which deal with the scaling and data management problem holistically [117, 166]. These include the growing number of workflow systems presented for geosciences, which incorporate visual analytics techniques, and novel architecture solutions, such as parallel programming or MapReduce techniques [71, 137]. Within ocean sciences, as mentioned earlier in the review, the development of large-scale coastal and ocean simulations undertaken by ocean observatories has led to an increasing focus on workflow management, which incorporate extremely large scale query-driven visual analytics. The querying solutions proposed by Howe et al. [136, 137], for the massively parallel data analytics and visualization system for the Columbia River Estuary adopt a database-centric approach, arguing that it is preferable to express more computation in the data management layer, and the system incorporates both sub-setting and database-style algebraic manipulation. The Collaborative Ocean Visualization Environment [111], provides workflow, data and asset management, geo-browsers and visualization systems supporting intensive studies of regions of the deep ocean. Whilst not yet tested on geoscientific data, Benzaken et al.'s. [30] EdiFlow integrates a workflow management system into a visualization system, database management, and visual analytics.

4.3.2 Defining the problem and identifying solutions

The goal of visual analytics systems is to provide correct, clear and informative output and interrogation of the underlying data, built on interactive rate visualization. But this is hindered by two difficulties: a visualization problem of over-plotting, and a selection issue. The screen space of display systems is regular and has a finite number of square pixels to handle an ever increasing volume of data, composed of large-scale, unstructured adaptive grids (irregular data space). Thus overplotting is an issue [270], particularly as visual analytics tends to use techniques derived from information visualization like PCP, scatterplots, glyphs etc. The challenge is to visualize this data at a satisfactory speed, to resolve the data association problem

without losing the required level of detail and enable areas of interest to be identified for quantitative analysis.

This results in the requirement for scaling, either by focusing on subsets of variables or by use of descriptive modelling techniques [193] for which Bertini et al. [34] has identified a number of quality metrics involved in the scaling and visual exploration process. These include clustering, correlation, outliers, complex patterns, image quality (e.g., spatial clutter) and feature preservation. Typically, two strategies have been adopted by researchers to reduce the number of points and provide a rendering which is representative of the data image (or visual) space and data space. The first, visual space algorithms such as Johansson and Cooper's [151] screen space method, focus on de-cluttering and provide a visual quality metric, bypassing data analysis, while frequency binning methods convey the density of the underlying data [101].

The second applies hierarchical data structures, (and these have been used in the geosciences) to support interactive visualization. Quality metrics calculated in data space detect features within the data, without using information from the view which displays the results. Thus, the data space method aggregates the information in the geospatial domain in relation to the level of detail in the data, by creating associated, linked, data-structures. These, when based on the principle of regular deconstruction through spatial indexing, enable computing resources to be focused on areas of interest, and facilitate easy data querying [252].

However, Bertini et al. [34] suggests that both data space and screen space strategies present different scaling problems: with visual space methods, as the data increase in size, visualization may be degraded and may actually hinder rather than support the identification of patterns. On the other hand, data space methods, which are probably more effective in pattern detection, may present computation problems because of the data size. Nonetheless, the two methods are not mutually exclusive and a number of applications combine both approaches [33,222], but decisions on a suitable approach will also depend on the attributes of the data, including with modelled data, the type of numerical model and the challenges it presents to effective visualization.

In the case of modelled data, whilst interactive rendering for structured grids has been well studied, this is not the case for unstructured grids [301], as their complexity challenges effective rendering [121]. This problem is further compounded by the use of adaptive mesh refinement in many modelled scenarios, which generates multi-resolution data. However, as the following section demonstrates, spatial

indexing through hierarchical data structures has been demonstrated to be useful in resolving these problems, and is thus an avenue worth exploring.

4.3.3 Spatial Indexing

The purpose of spatial indexing is to support spatial selection – to retrieve spatial objects from large datasets in response to a query value [112]. The hierarchical data structures of the type reviewed in this chapter sort the data according to the space they occupy, organising that space and the objects in it, so only parts of the space and a subset of the objects are required to answer a query. This is achieved by decomposing the space occupied by the data into buckets where each bucket has an associated bucket region, containing all the objects stored in buckets. Applying the principle of recursive decomposition, hierarchical data structures are regarded as useful because they are compact, they facilitate search, and are capable of saving space and time, depending on the nature of the data [252].

Spatial indexing and the geosciences

The use of spatial indexing to associate data between views and also to reduce the data is explored, with a view to establishing its relevance for the case study data. The value of the approach for data querying through association with place is widely recognised [251, 255] and a number of spatial indexing solutions proposed for geoscience data, including some ocean science applications, are discussed below, although there is only a little evidence of their use for coastal shelf and estuarine visualization.

Quadtrees and their variants (described in detail in Section 4.4.3c), which are most frequently used for partitioning two dimensional space, are proven methods for the multi-field datasets associated with the geosciences, because of their explicitly spatial nature [205, 248]; their ease of access and speed of manipulation [69, 90, 296]; their flexibility in dealing with non-uniform data and their usefulness in applications requiring searches [252]. Using geo-spatial data, Keim and Hermann [164], compared a hierarchical, data partitioning algorithm based on a quadtree (Gridfit) with a nearest neighbour algorithm, and a curve-based algorithm, concluding that Gridfit was more efficient in resolving overplotting problems, and was also mathematically and visually more effective. Conversely, Stockinger et al. [278] expressed concerns that quadtrees separate cells which are neighbours, preferring a data extraction approach based on Bitmap indices, in conjunction with

3D visualization, but this technique is based on an assumption of uniform grids, which is not always the case with modelled oceanographic data.

The PR (point region) quadtree [251], is regarded as the most relevant for modelling fields and is particularly useful for the multi-resolution grids associated with numerical modelling, offering an adaptive solution, as more squares are present to capture greater detail when there are an increased number of points. Examples of quadtree use in the geosciences include the SAND spatial database system [89,253]; applications for mobile, handheld GIS technology, where limitations such as small screen display, bandwidth, colour resolution and application capabilities need to be surmounted to achieve faster response times [255]; interactive visualization systems for large digital terrain models [215, 216, 292]; increasing the accuracy of terrain modelling triangulation [254]; representation of vector data in a GIS [205].

A small number of oceanographic applications have used similar hierarchical structures: as already mentioned, these include Keim and Hermann's [164] hierarchical data partitioning algorithm based on a quadtree. A quadtree has also been used to optimise the speed and quality of rendering for hydrographic data visualization [18] and the C-squares spatial indexing system was developed for oceanographic databases based on the quadtree [232]. The hypothesis was that the algorithm would be particularly useful for irregular gridded marine data, although limitations for computing areas were identified, as were issues of ambiguity at boundaries – which might present a problem for its application to estuarine data, where there are numerous boundaries. Johnson-Robertson et al's. [153] application of a quadtree supports visualization and analysis of vast quantities of data generated from underwater robotic surveys and the GeoZui3D [311] oceanographic visualization system for interpreting multiple sources of 3D data incorporates a quadtree to support statistical binning techniques and enable real-time display. Vo et al. [301] applies an octree (a 3D variant of a quadtree) to a hardware-accelerated volume rendering system for massive estuarine data sets to support interactive visualization and visual analytics.

K-d trees, a variant of binary search trees, are particularly useful for multi-dimensional searches [112, 220, 255], including feature extraction, so the technique may also translate to the coastal oceanographic domain. Lamphere et al. [179] applied a k-d tree to scale simulated 3D volume data, improve the interactivity of very large scale visualizations performed on massively parallel supercomputers and to resolve speed/quality/level of detail issues. Similarly, Hsu et al. [138] achieves significant improvements in performance using a balanced k-d tree for feature extrac-

tion in massive cosmological simulations. Correlation queries in extremely large, long term, modelled climate data sets are facilitated by Zhang et al's. [327] spatial cone tree (which utilizes a k-d tree type clustering scheme) to reduce computational cost. Duvenage [79] proposes an implicit min/max k-d tree algorithm to evaluate line of sight queries, focusing on boundary information, on large scale terrain simulations to achieve significant improvements in performance.

Numerical Models and Spatial Indexing

It has already been mentioned that unstructured grids present challenges for interactive speed visualization, yet much ocean science modelling is undertaken on unstructured grids, with Bernardon et al. [31] reporting that little work has been proposed for scaling such data. The focus of research has been on volume rendering time-varying data on regular grids, and although many solutions are hardware related, spatial data structures, such as octrees have been used to optimize rendering, for example Ellsworth et al's. [84] Time-Space Partitioning tree for time varying data, and similarly Vo et al's. [301] system for volume rendering of tens of millions of tetrahedra at interactive rates on a PC mentioned above.

The workflow system of Howe et al. [136], for the Columbia River Estuary proposes an algorithm for unstructured grids, incorporating gridfield algebra into the workflow, before visualization using VTK(see figure 4.7).

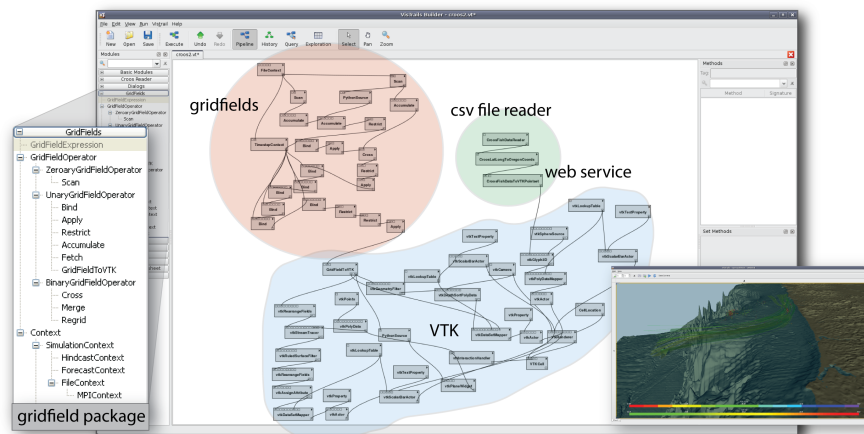


Figure 4.7: Howe's solution to the challenges presented by the unstructured grids of finite element models. The gridfield algebra is implemented as modules, grouped by kind and visualized using VTK. Copied from [136].

Modelled data is often multi-resolution, generated from Adaptive Mesh Refinement (AMR) grids. Instead of applying uniform, high density grids across the whole data area, which require greater computing capacity, the meshes are comparatively coarse in areas of less interest, and are progressively refined both spatially and temporally in areas of greater interest. These hierarchical data types present a significant challenge to visualization. They refine the domain space of a simulation into a hierarchy of nested, sequentially refined grids, both spatially and temporally, such that each timestep consists of multiple levels of grid cell refinement. However, modelling the data to a uniform grid is sometimes not helpful as this may introduce grid artefacts not present in the original data, and may increase storage requirements [234]. Further, the disparity in refinement between timesteps prevents the evaluation of multi-temporal queries and also creates rendering problems. Yet, as Weber et al. [313] reports, despite the growing popularity of AMR simulations, little research into their effective visualization has been undertaken, and the standard way of describing subdivision has become k-d tree algorithms.

Park et al. [220] proposes a solution incorporating multi-resolution splatting techniques using k-d tree and octree data structures, together with a GUI, interactive determination of transfer functions and viewing /rendering parameters, which significantly improved rendering speeds of cosmological data. Similarly Kreylos et al. [176] make AMR data homogeneous by changing them to blocks of constant resolution using a k-d tree, whereas Kahler et al. [158] applies a k-d tree on the GPU to order data and accelerate rendering, also using the algorithm on a volume visualization system for massive cosmological data. However, most k-d tree examples identified related to cosmological data, and none related to oceanographic AMR modelling.

Non spatial indexing solutions for visualization of AMR data have also been proposed. The approach by Gosinck et al. [107] enabling query driven visualization of hurricane data, proposes a composite template of the timesteps. Each grid is synchronised to the composite, before applying a GPU based, Query Driven Visualization approach for accelerating querying to better than interactive levels. The VisIt visualization and analysis tool [58] marks coarse cells, which are refined at a lower level as ‘ghosts’, enabling AMR data to be rendered for subsequent analysis by ignoring all the ghost cells. But many non-spatial indexing solutions (such as Weber et al. [313] and [314]) do not take into account the query driven nature of visual analytics. However, the flow visualization approach by Peng et al. [223], on studies of the flow past a marine turbine, which incorporates visual analytics tech-

niques, is to use the CFD mesh manager to compute mesh adjacency information as a pre-processing step.

Quadtrees & k-d trees

This section is informed by Samet [248–252];

The review focuses on two spatial indexing variants: quadtrees and k-d trees because of the identified association with geoscientific data, and evidence they have been successfully applied to Adaptive Mesh Refinement (AMR) data. Both are binary space partitioning trees, and their regular decomposition approach has been proved effective for arbitrarily distributed data, as it is more flexible than many other indexing methods (the depth of the tree, and therefore the density of the information can be adjusted from one area to another in relation to the variability of the data being represented) [106]. K-d trees were also seen to be useful for volume visualization of unstructured, multi-resolution grids.

Quadtree The term quadtree refers to a class of hierarchical data structures used in spatial indexing, differentiated by the type of data they represent; the principle governing the decomposition process and the resolution (variable or not). They may be used for points, rectangles, regions, curves, surfaces and volumes, and the decomposition may be regular (equal parts on each level) or governed by the input. Resolution (the number of times the decomposition process is applied) may be pre-determined or adapt to the properties of the input data.

Quadtrees are used for 2D data, where each internal node can have up to four children, each child representing one of the four quadrants of the parents square, and obtained by decomposing the coordinate space. The root of the tree reflects the complete picture, and the leaf nodes the tessellated picture segments, which, if recombined into groups indicated by the internal nodes, would reproduce the original image. Cells may be added or removed by moving up and down the hierarchy. Figure 4.8 shows the quadtree data structure [228].

A hierarchical description of picture patterns, elements and their relationships is retained explicitly in the data [195]. At any level, the cells are equal in area and shape, and adjacency relationships between cells are correctly encoded [106] – of value for data derived from finite element models, such as the case study data for this thesis.

The quadtree is used in the form of a square, as it is a planar decomposition

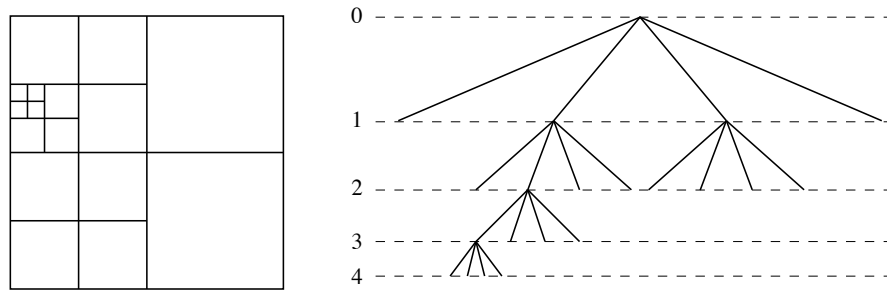


Figure 4.8: Representation of a quadtree, copied from Popenet et al. [228], showing, on the left hand side, a geometrical description, and on the right hand side a logical description, with numbers indicating the level of the corresponding cell.

which produces partitions in an infinitely repetitive pattern, thus it can be used for images of any size and it is high resolution, by virtue of being decomposable into increasingly finer patterns.

Quadtree type methods easily support location based queries and searches by descending the tree until the object is found, and also nearest neighbour, if this is required. It distinguishes object from background and is thus useful for focusing on interesting sub-sets of data [195]. Quadtree type methods have limitations: in particular, placing objects in relation to the decomposition lines of the space where they are embedded, (which affects storage costs and the amount of decomposition taking place), nevertheless Samet [252] concludes that this can be overcome by bucketing which decomposes a block only if it contains more than n objects.

The most widely studied variant of the quadtree, the region quadtree, is based on regular decomposition of the space into four equal size quadrants, and is characterised as a variable resolution data structure.

PR-Quadtree The PR (point region) variant (proposed by Samet [250] of the region quadtree is of acknowledged value for modelling fields, and is recognised as useful for the multi-resolution data associated with adaptive mesh refinement modelling, as more squares are present to capture greater detail, when there are more points. The PR quadtree adapts the region quadtree to point data, by associating data points (which do not need to be discrete) with quadrants. It is organised in the same way as the region quadtree and decomposes the total space into quadrants and sub-quadrants, until the number of points in each quadrant is within a set limit (the node capacity). In this case the decomposition continues until only one point and its coordinates exist in the quadrant (hence ‘point region’ quadtree) [269]. The

shape of the resulting PR quadtree is independent of the order that data points are inserted. (See Figure 4.9 illustrating a PR quadtree and the records it represents).

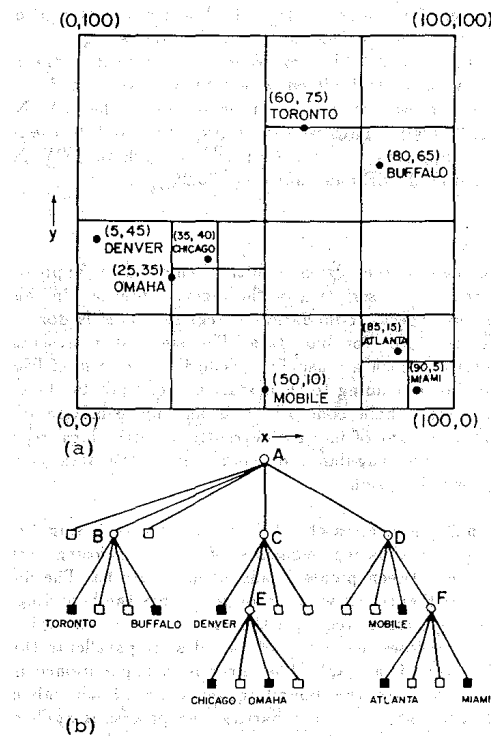


Figure 4.9: Direct copy from Samet [248] showing a PR quadtree and its records.

K-d Tree K-d trees, described by Bentley [29], as a k-dimensional binary tree, is a data structure which decomposes multidimensional spaces and is useful both for rendering and efficient storage of AMR data hierarchically [220]. The root cell represents the entire simulation volume, and individual nodes in the tree represent a data point and direction of search. Each node contains a key, drawn from one of the k dimensions, with a discriminator which governs the branching decision. Leaves contain the points to be stored, with the recursive decomposition ending when each cell contains just a single point [212, 269]. (Figure 4.10, illustrates k-d tree partitioning). The manner in which it divides the data space into smaller subspaces, results in less data having to be stored, than the method used by the quadtree (potentially placing less demand on memory).

Quadtree / k-d tree summary The respective attributes of the PR quadtree and the k-d tree can thus be summarized as follows:

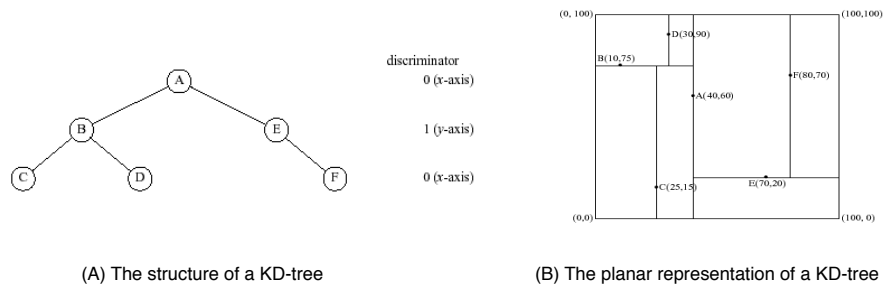


Figure 4.10: k-d tree partitioning, direct copy from Sinha et al [269].

- The PR Quadtree partitions 2 dimensional space by recursively dividing it into four quadrants or regions, whilst the k-d tree is a k-dimensional data structure which organises space in a k-dimensional space.
- Each node of a PR quadtree represents a particular region in 2D coordinate space, but each node in a k-d tree is a k-dimensional point.
- Non-leaf nodes do not store data in the PR quadtree, and internal nodes have just four children (of which some may be empty) representing a different, congruent quadrant of the region represented by their parent node. In the k-d tree, non-leaf nodes are k-dimensional points, and every non-leaf node generates a splitting hyper plane that divides the space into two subspaces.
- Only leaf nodes contain data points in the PR Quadtree, but in the k-d tree every node is a k-dimensional point.
- The structure of the PR quadtree is determined by the data values it contains and is not governed by the order of their insertion, but the shape of the k-d tree depends on the order in which the points are inserted (which may present difficulties in the development of tools permitting users to select level of detail by growing/pruning the tree).
- In the PR quadtree dense data will require more partitions and thus may not be balanced, whereas in the k-d tree, the worst case is that the number of the levels in the tree becomes equal to the number of points to be inserted [269].

Thus both variants will present advantages and disadvantages for scaling the data in this work, although they are similar in their ability to handle AMR data and their demonstrable value for geoscience data, including the ocean sciences. Whilst the benefits of a spatial indexing approach were apparent, no clear preferred

option emerged from the review as to the method of spatial indexing, suggesting a requirement for prototyping.

4.4 Interaction & Interfaces

As the earlier survey of multi-field data demonstrated, it is apparent that no one approach lends itself to understanding all the characteristics of the data and the relationships between them, thus requiring a number of views and analysis tools to locate and highlight salient features or provide statistical and quantitative information through a visual analytics system. But this has considerable implications for the requirements of the visualization system: as the earlier sections have identified, visual analytics relies on an interface which facilitates user interactivity with the data, and also supports a rich diversity of views, plots and graphs for studying the different attributes of large, multi-field datasets. Methods of achieving interactive visualization speeds were considered in the previous section, and this section therefore:

- discusses interaction and interfaces in relation to visual analytics and to coastal shelf and estuarine studies;
- considers their application to coastal shelf and estuarine data (and the wider oceanographic and geoscience context);
- identifies an approach which supports the needs of a visual analytics system for coastal shelf and estuarine studies.

To enhance researchers' capacity for understanding big data, the role of Visual Analytics is to improve the human ability to capture associations and complex relationships within the data, in support of decision-centred visualization, but this process depends on an enabling mechanism – interaction between the user and the data, via an interface which presents the data in a number of ways to enable analysis and stimulate insight.

Coastal ocean scientists examine their data through comparisons of representations of the data, although comparisons of different runs of a simulation, or different simulations are still largely undertaken with image level comparison, usually static plots, as previous sections have demonstrated. There is therefore, scope for visual analytics systems with enhanced interactivity and querying systems to be usefully applied [208].

The concept of the ‘mental model’, first proposed by Craik [66] is relevant here: this proposes that a mental model of an interactive visualization may be stimulated in the ‘working memory’ of the user to facilitate reasoning [188]. Visual analytics, by providing complementary views, plots and graphs of the data is intended to stimulate this mental model. Increasingly, coastal shelf research is undertaken in support of environmental decision making, for example in relation to flood prevention strategies as part of coastal zone management. However, establishing relationships and key interactions within geoscience data is problematical using mathematical extraction, hence, visual analytics facilitates comparison between different modelled ‘what if’ scenarios to capture associations and complex relationships [166, 262].

Furthermore, the traditional reliance by coastal ocean scientists on the comparison of static plots is unlikely to meet the complex analytical requirements of the researchers, as their perceptual judgements may be affected by change blindness [268, 307]. This proposes that the ability to detect changes is hindered by small interruptions between images, whilst shifting focus from one image to another. A number of geoscience researchers have recognised this and applied coordinated multiple plots to enhance perception.

In addition to the potential for change blindness, single complex views can be cognitively overwhelming to the viewer, so Simons and Rensick [268] consider multiple views aid memory by reducing the amount of data needed to be considered at any one time. For example, Aoyama et al. [17] presents a novel, hierarchical system, with views and attributes represented on different levels, rather than side by side, to minimise the potential for change blindness.

However, visual analytics, using coordinated multiple view aiding qualitative visual examination may support the development of a conceptual framework for developing quantitative analysis [8, 37, 262].

4.4.1 Query Driven Visualization

Coordinated Multiple Views require a query driven system (Query Driven Visualization, QDV) to identify the areas or features of interest, which is founded on assessing data quality, and Andrienko et al. [16], consider QDV to be a cornerstone of visual analytics. However, not all data is of equal quality or of equal interest or importance [180], thus QDV, first described by Stockinger et al. [278], is based on the premise that the derived subsets of scientifically interesting data often provide insight, and support the generation of hypotheses.

Features or ‘areas of interest’ will vary, dependent on the domain of study: for example, estuarine hydrodynamic researchers are interested in the formation of channels, spits and scour pits; meteorologists and climate scientists in the structure of convective clouds, and deep ocean scientists in the formation of eddies, each with its own very specific definition. However, a definition which might be applied across all domains is required. Silver [266] describes a feature as a coherent structure (an ‘effect’) persisting for a period of time, whereas Bethel et al. [35] develops and quantifies the qualitative term of scientifically interesting as a combination of data range values (Boolean range predicates), which QDV enables to be defined, extracted and analysed, using scientific and information visualization techniques [278].

Query driven visualization also supports undirected exploration. This process is illustrated by Kehrer et al. [160] through an analytics system identifying regions in the atmosphere which may act as indicators for climate change. Although user focused, to the extent that it deals with research into real world challenges, it is not directed towards resolving a specific problem. Rather more exploratory in nature, it seeks out relationships in data which may suggest promising hypotheses, to be tested and confirmed through visual analytics. Traditionally, such hypotheses have been generated through intuition or trial and error, but Kehrer et al. sees an opportunity for visualization to support what has been a cumbersome process, through the iterative capacity of visual analytics and its ability to set parameters and boundaries for analysis.

Within ocean sciences, as mentioned earlier, the development of large ocean observatories has led to an increasing focus on workflow management, which incorporates extremely large scale query-driven visual analytics. The querying solutions proposed by Howe et al. [136, 137], for the massively parallel data analytics and visualization system for the Columbia River Estuary adopt a database-centric approach, arguing that it is preferable to express more computation in the data management layer, and the system incorporates both sub-setting and database-style algebraic manipulation.

However, what is apparent is that there is no single ‘correct’ approach: the multi-field nature, complexity and scale of oceanographic and geo-scientific data require a flexible system, offering a number of approaches to support efficient storage and access for spatial objects and their time varying characteristics [296]. Here, parameters may be readily changed, data easily visualized and then queried in a number of different ways, using methods such as coordinated multiple views, to make com-

parisons, and reveal new perspectives and relationships. This requires high levels of interaction.

4.4.2 Interaction

Fuchs and Hauser [94] describe interaction as probably the most important tool for understanding complex data, because it is able to support direct querying and QDV. The entire iterative process of drilling into the data and consequent knowledge discovery depends on the ability to interact with representations of the data to re-parameterise, navigate, manipulate, search and compare [3].

Appropriate perception-driven interactive systems rely on an understanding of theories of cognition as applied to visualization [165] and relating these to the needs of the user and the tools and techniques available in the visual analytics environment. For example, decisions such as whether to support data selection through direct querying techniques (for example brushing) as opposed to indirect techniques, such as sliders [322], are intrinsic to developing the visualization and analytics system. Brushing might be the preferred option by researchers, enabling them to undertake the more complex knowledge discovery tasks associated with visual analytics, requiring high levels of interactivity in identifying relationships in the data, whereas sliders might be used to select parameters prior to complex comparison [184].

Thus, the usability of the system depends, in large part, on the effectiveness of the user interface, initially in providing researchers with suitable views and plots to facilitate data examination, and secondly by presenting it in such a way as to stimulate the reasoning process outlined earlier. But the development of user interfaces supporting techniques such as query driven visualization, has only taken centre stage in recent years, particularly with the evolution of visual analytics, and, within the geosciences, the need to develop ways of studying ever increasing large volumes of spatio-temporal, multi-variate data [191].

Pike et al. [226] concludes that interaction and inquiry are linked and presents a number of goals for the evolution of interactive systems and interfaces for visual analytics. These include embodied interaction (creating core analytic capabilities); capturing user's intentions (what is the user's goal); knowledge-based interfaces (developing the human knowledge/visual analytics synergy); multi-analyst collaboration; applying principles of design and perception to interface design and interactivity and evaluating interaction (how beneficial is it). The goals also include interoperability – utilising new tools. Steed et al. [275] exemplifies this, where a

number of enhancements to PCPs are incorporated into an advanced PCP system for studying hurricanes and tropical cyclones – a system which stimulated ideas for the PCP proposed in this research. In totality, Pike’s goals are wide ranging and beyond the scope of this project other than in a limited way, but they may inform future development of the system. Whilst an interesting avenue of research, the focus of this thesis does not permit a detailed study of interaction, instead, the reader is referred to the work by Liu and Stasko [188], and Pike et al. [226] for an understanding of cognition, interaction and human interfaces relating to visualization, visual analysis and the exploration of large datasets.

4.4.3 Current approaches to interfaces and interaction in ocean sciences

Earlier sections of this chapter have painted a picture of limited use of advanced interaction, display and analytic techniques, particularly in Ocean Sciences, where cross-reference and comparison of data proves time consuming and cumbersome, using the limited visualization capability of the simulation model’s software, and often relying on non-specific systems such as MATLAB.

Many early coastal shelf and oceanographic visualizations relied mainly on indirect interaction through menus and buttons, and limited direct interaction, through a mouse, for example Jiminez et al. [148] and Stein et al. [277]. However, because of the delay whilst displays are updated, indirect techniques may preclude operation in real time, which is increasingly required in ocean science modelling, particularly for those systems supporting environmental incident and disaster management. Nonetheless, there are a number of exceptions. For example, GeoZui3D [20, 311] describe a highly interactive system for 3D visualization of multi-disciplinary, multi-field ocean data. The unique zooming interface was designed with the user in mind and placed selected points at the centre of the workspace. It also offered linked Coordinated Multiple Views and focus+context.

Visual analytics aside, much of the focus of attention relating to interaction in the geosciences in general, and ocean sciences in particular, has been on the use of VR techniques including immersive and stereoscopic viewing systems. These harness human perception to aid understanding of the data, by enabling the user to interface directly with the increasingly large datasets, and to gain context from the surrounding environment [37]. Depth information may be ambiguous in 3D visualization [306], thus an immersive environment may aid understanding of spatial and

dynamic relationships.

Examples include the use of a CAVE VR system [159] to explore the dynamics of sediment mixing within the marine bottom layer, believing this will provide a better understanding of the spatial variability. Borodin et al. [42] describes stereo rendering in conjunction with a PowerWall or CAVE to gain an additional depth cue, in studying the impact of dumped chemical weapons on marine ecosystems in the Baltic. In a state of the art review of immersive displays for visualizing and exploring coastal ecosystems, Hermann and Moore [126] conclude that immersive techniques provide valuable depth cues lacking in 2D plots, when visualizing velocity with vectors. In another example, researchers used head mounted displays to study modelled tidal current data, where it was concluded that the more spatially complex the model, the greater the benefit of stereoscopic presentation [213].

These examples illustrate the value of immersive and stereoscopic systems in the study of 3D hydrodynamic data, especially to provide additional insights into the data through manipulation and viewing, that are generally believed not to be possible through visualization alone. However, whether this insight is greater than that of a visual analytics system is not known. Of course, there is also the matter of access to immersive facilities such as a CAVE or PowerWall (particularly as estuarine researchers may require access to computing and display facilities in the field). Nonetheless, were a 3D version of the visual analytics system for this thesis to be developed, it may be a useful further step to assess whether VR techniques would complement visual analytics sufficiently to justify their incorporation into the system, particularly in regard to the difficulties associated with the scale of the datasets and the challenges they present to display technologies. Of course, there would necessarily be a computing cost.

The large ocean observatories mentioned in earlier sections have had an impact on the evolution of interaction techniques, interfaces and display methodologies for ocean science data, through the development of visualization and analysis systems for multi-source, multi-disciplinary, distributed, collaborative projects. For example, Shen et al. [263] demonstrate the feasibility of integrating a complex ocean modelling system with interactive 3D visualization, through a user-friendly GUI, which enables researchers to make real time observations; control wind speed/direction and rate of flow, and use check boxes and sliders to select variables for visualization. The Collaborative Ocean Visualization Environment (COVE) [110], provides a number of highly useful features, including the facility to layer several types of data. But, whilst it supports interactive visualization

of both modelled and sensed data from many sources, it does not yet provide the detailed visual analytics techniques which would be required for this thesis. However, as discussed earlier, the workflow management systems proposed by Howe et al. [136, 137], for the Columbia River Estuary system do support highly complex, large scale, interactive visualization and visual analytics, through a system integrating 3D visualization and analytics and remote querying for both modelled and sensed data.

GODAE (the Global Ocean Data Assimilation Experiment) [40], supported by MERSEA, provides interfaces for users to access and visualize the data at various levels: fast visualization to decide whether to acquire the data, then rapid interactive visualization to view daily updates (but limited ability to manipulate or study the data). Finally, there is an interface permitting interactive, web-based visualization; panning and zooming; adjustments to colour palettes and scale; changing map projections and creation of animations. The interface supports display of maps, scientific graphics, tables and data subsets, also inter-comparison through its slide sorter (for comparing fields within a single model). Datasets from different institutions may be compared by re-gridding and differencing. Figure 4.11, shows the GODAE/MERSEA GUI for selecting model output.

Finally, VAPOR (a visualization and analytics platform for ocean, atmospheric and astro scientists) [61], supports desktop visualization and analysis of terra-scale datasets, including real time animation. Interactivity is improved through data caching and a GUI, which facilitates simultaneous views of multi-field variables. However, there are currently no examples of the application of VAPOR to ocean science data. But as the Related Work has also demonstrated, the level of sophistication of the large ocean observatories does not apply throughout ocean science and coastal shelf visualization.

Both in this section and earlier ones, the technique of Coordinated Multiple Views has been mentioned. The next section discusses its attributes and its role as an interactive interface in providing coastal shelf researchers with the analytical tools they require to aid understanding of their data.

4.5 Coordinated Multiple Views

Coordinated multiple views (CMV), in conjunction with linking and brushing are well established means in the geosciences of enabling detailed, interactive analysis of data, through linked, simultaneous, side-by-side displays of different data vari-

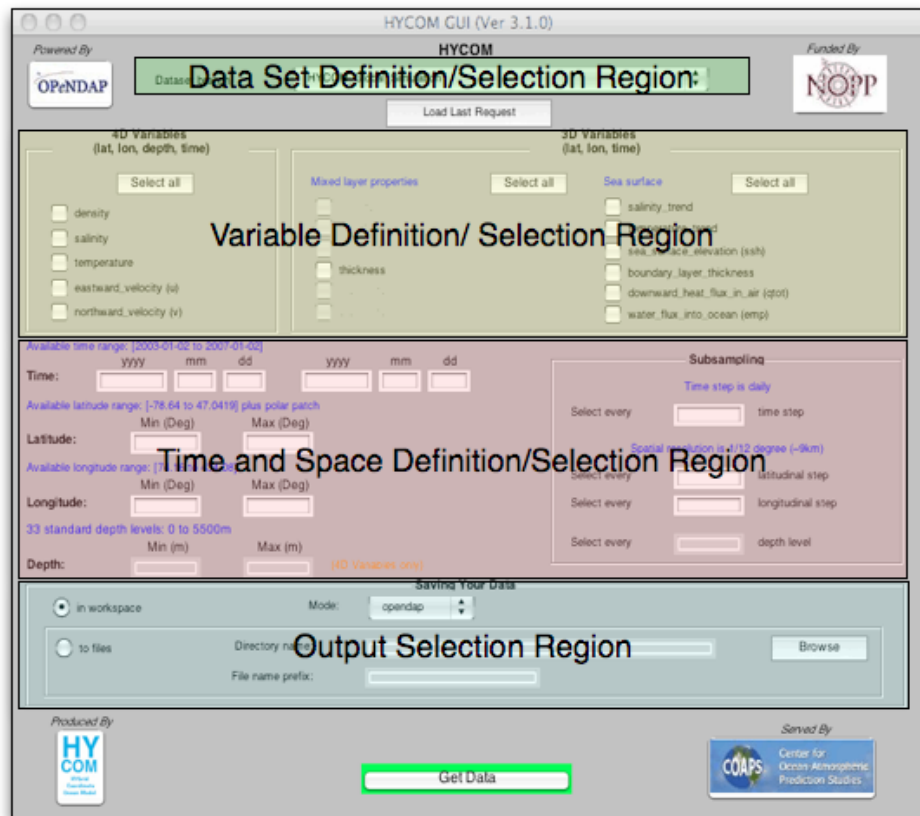


Figure 4.11: Reproduced from Blower et al. [40]: GUI for selecting model output from a consortium member, annotated by selection region.

ates [160,166,236,258]. The benefits of the system are recognised as providing improved user performance: by explicitly linking and highlighting like features from one view to another, the ability to establish relationships between data is supported, and a further benefit is unification of the desktop [209].

The scope of this section does not permit a detailed appraisal of CMV development and techniques, instead the reader is referred to Roberts [236,237], which reviews CMV methodologies and the state of the art. However, the premise of CMV is that users will obtain a better understanding of their data by interacting with them and viewing different representations to gain insight, through the greater flexibility the technique and its tools provides. CMV enables dependencies for example spatial or temporal dependencies, or dependencies resultant on other attributes of the data - to be identified and studied. The nature of spatial data has meant that traditionally they have been visualized through scientific visualization techniques (flow or volume visualization or with GIS), but the advent of visual analytics, and the resultant requirement for query driven visualization, has seen information visu-

alization methodologies and tools adopted in conjunction with scientific visualization, to facilitate exploration of the data. The process permits users to formulate a problem and concurrently solve it [272], by studying the linked representations of the data. Techniques include statistical analysis; identification of areas of interest and attribute plots and graphs (for example function graphs and hierarchical clustering techniques such as scattergraphs, histograms and parallel coordinate plots), together with temporal views, which allow users to investigate the relations between data variates [2,107,258]. These different data representations are discussed in more detail in Section 4.6.

Areas of interest may be selected by sliders, menus and buttons, but the interactive, direct manipulation technique of brushing is most commonly used [237] for linking. Its value for visual analytics is that it provides instantaneous querying, and also supports the derivation of sub-sets of data. Directly applied to the visual display, logical combinations of brushes enable specification of complex features [3]; identified areas of interest to be simultaneously represented, and relationships explored in the linked views. For example, Figure 4.12, Doleisch et al. [76] , shows multiple views of a visual analytics system for a diesel exhaust system where temperature values in relation to time are colour mapped to features of interest and velocities compared.

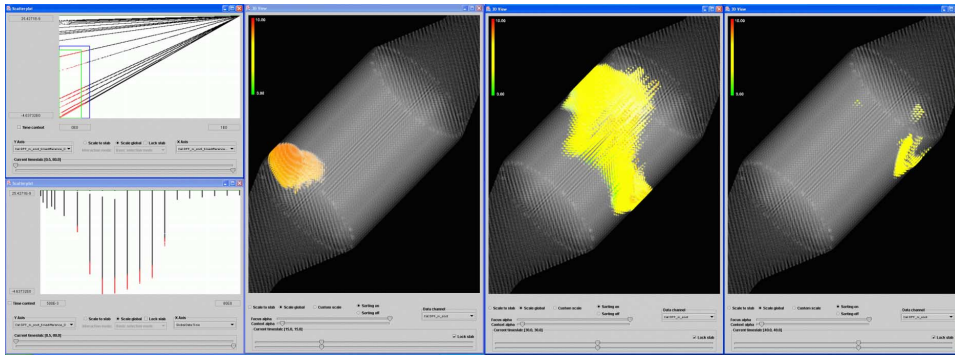


Figure 4.12: copied from Doleisch et al. [76]: Coordinated Multiple View of diesel exhaust emissions

An interface based on Coordinated Multiple Views was identified as a useful avenue of study for this work because of the attributes outlined above which support a rich functionality of interaction, enabling analysis and comparison of multi-field data using views, plots and graphs closely related to the data attributes. The coordination provides additional value by explicitly linking (and highlighting) like

features in one view to another.

4.5.1 The value of Coordinated Multiple Views (CMV) for multi-field data

As Visual Analytics approaches have evolved, a number of methodologies have demonstrated themselves to be useful for multi-field data, including a Coordinated Multiple Views interface. Ho and Jern [135] considers Coordinated Multiple Views to be helpful for understanding spatio-temporal relationships. Such data are difficult to analyse through a single visualization form, particularly if this is volume representation, because parts of the volume can be occluded from the user. However, CMV possesses the flexibility to overcome the limitations of a particular view, and typically, views for studying spatio temporal data will consist of a number of computationally different depictions [16] enabling different aspects of the data to be identified through the linked views [2, 236]. Attribute plots and graphs, linked to the views through CMV, will permit investigation of variates of the data in the spatio-temporal context. Figure 4.13 reproduced from Blass et al. [38], exemplifies the approach in a meteorological system, which combines both 2D and 3D views, together with attribute views.

The interactive nature of the CMV interface allows users to navigate the time axis, provides techniques to convey different levels of abstraction and enables selection and comparison of different views and features of the data values [3, 237]. However, Aigner et al [2] suggest the need for a degree of caution: whilst acknowledging the ability of CMV to represent different parts of the time axis, the coordination of visual and analytical methods that may not share parameters (for example derived principle components and predicted future trends) is considered to be challenging.

4.5.2 Views

Whilst the next section considers in detail some of the attribute tools and techniques used in visual analytics (which vary dependent on the attributes of the data), this section focuses on the methods for visualizing and viewing the data.

Animation and static plot views

Spatio-temporal data may be visualized by animation by displaying the data at individual time steps, a process which is ubiquitous in geoscience visualization [16]

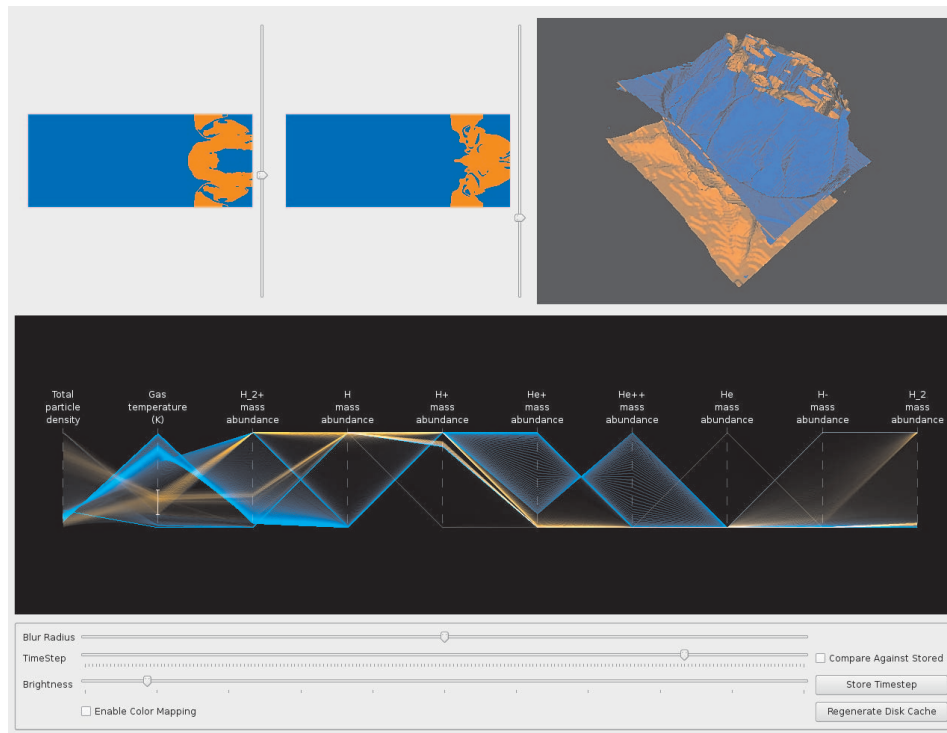


Figure 4.13: Reproduced from Blass et al. [38]: CMV main user interface of application to study meteorological data. The top part of the screen contains spatial viewing components (two slice viewers and a 3D isosurface view). The parallel coordinate plot is positioned in the middle, and the control interface is positioned at the bottom.

– Vis5D [133] for earth and space data, being a representative case. Oceanographic researchers often perform animations of their data – which are inherently temporal and consequently naturally afford animation to be used. Data may also be visualized by drawing function graphs to display them at individual time points, and the line-graph that shows time along the x-axis demonstrates the complete dataset.

There are advantages and disadvantages to both animation and static plots, but a goal of this research is to identify representations which will improve data comparison, considering the limitations of the existing representation methods used in conjunction with hydrodynamic modelling and other oceanographic studies.

Whilst a static snapshot of an individual time step provides only limited understanding of trends and transformations, including changes in the structure and position of features in the data set, over time [157], it is more effective than animation in providing quantitative analysis, showing different views of attributes of the data and supporting comparisons, say of time spans, or identifying spatial and

temporal patterns [13] – particularly when incorporated into the CMV framework. Animation is also not as effective in supporting studies of longer multi-variate time series [268, 295], where researchers may not identify all the changes, and struggle to retain the details. The strength of animation, however, lies in qualitative analysis, for showing the dataset in its entirety [321], for conveying its time dependency and evolution and for supporting insight into events and temporal processes [13].

Thus, each type of view supports different aspects of spatio-temporal inference and the use of a CMV interface enables systems to be developed which incorporate both. Examples include [38, 135, 161].

Dimensionality of representation

The dimensionality of representation – 2D or 3D – is widely discussed within the visualization community, and there are strongly made arguments in both camps, although it is not proposed to review this debate extensively. Many researchers regard 2D as sufficient for effective analysis, particularly in view of problems, such as occlusion, which arise when incorporating the third dimension [2]. Yet whilst other researchers see 3D representation as an opportunity to encode further information, it remains a significant challenge in certain domains, for example high dimensional, 2.5D surface flow visualization, volumetric flow (3D) and flow with several attributes [223].

Nonetheless, for many years, 2D graphical representations of complex data have provided insight into quantitative data [37]. However, Joshi et al. [157] argues the case for 3D representation for hurricane visualization, as it supports greater insight into the spatial structure of the hurricane, where understanding spatio-temporal relationships between hurricane features is a major challenge. GeoZuid3D [20, 311], (mentioned earlier in this section), a three dimensional system, interprets multiple sources of 3D data, to establish inter-relationships between 3D features in heterogeneous, spatio-temporal oceanographic data for better understanding of biological and physical processes. Similarly Howe et al. [136] found that 3D viewing proved more effective than 2D methods in identifying the source of upstream salt flux.

Yet although many 2D representations are not able to correlate features with surrounding spatial features or adjacent time steps, Baptista [23] pragmatically prefers 2D representation, because the processing demands are not as high as with 3D – regarding functionality as being of greater importance. An earlier view, Van Wijk and van Liere's [298] considers that while 3D animated, volume rendered visualiza-

tion might be an obvious choice (because of its ability to show as many dimensions as possible simultaneously), it is only really appropriate for 3D data sets, because of the challenges of the approach. Speed is an issue, as is interactivity, but importantly, volume rendered views are more difficult to interpret than simpler representation methods, and will probably require fine tuning. This is also an issue for non computer-science users, where ease of use is a factor. However, the work on this dissertation has focused on the use of two dimensional slices, which offer fast rendering, permit easy visual interpretation, and direct interaction. The counter-argument is presented by Dos Santos and Brodlie [77], who propose solutions to a number of the problems with 3D visualization identified by van Wijk and van Liere's [298]. Nonetheless, Peng et al. [223] consider that many of the issues still present challenges today, for example visualization of flow on surfaces in 2.5D and 3D flow visualization.

However, both 2D and 3D confer benefits: 3D views may be used to gain context and to provide an overall understanding of the data space, and 2D views will provide the precise data access, thus supporting both qualitative and quantitative analysis. A CMV interface is able to support both representations.

4.5.3 Use of Coordinated Multiple Views in ocean sciences and the wider geosciences

The value of linked coordinated multiple views in supporting complex data analysis is widely acknowledged within the geosciences, albeit there is a degree of caution among some researchers. Andrews [9] expresses a concern that developers of CMV-based visual analysis systems must not lose sight of the user: all systems should enable the user to add to understanding of the data, and serve to focus attention on the relationships within them, rather than adding to clutter and causing distraction because of the complexity of the views.

Ho and Jern [135] discusses the many benefits of coordinated multiple linked views to volume visualization, so it is unsurprising that the majority of the examples identified are 3D applications for 3D+ data [38,75,135,160] for example. A number of these are discussed in greater detail below:

Kehrer et al. [160], describes a system for studying atmospheric climate change identifiers, with linked CMV showing 3D views of volumetric, temporal data, together with 2D scatterplots, histograms and a function graph view, aggregated using frequency binmaps to retain responsiveness. Features are specified in the attribute

views, and brushed sub-sets reintegrated into the data as a synthetic data attribute used in the 3D view, with a focus+context approach to discriminate specified features from the remaining data. Explicitly expressed brush attributes may be interactively adjusted and saved for application to other datasets, supporting rapid comparison. The system was considered particularly useful for establishing correlations and giving improved understanding of the impact of different parameters analysis.

A CMV system for meteorological data describes a two-way linkage between the top and bottom of the screen (the top part contains spatial views and the bottom half a PCP), with point selection by probe on the spatial views overlaying the selected voxel on the PCP [38]. In other work, Turdukolov et al. [294] links selections from the PCP to slice viewers. The system also includes an object brushing tool for use with the spatio-temporal data, to select and view objects in a number of ways (for example their object history). The brushing operates across all plots – abstract graphics, images and a 3D event graph.

In cosmology, researchers have used CMV of 3D visualizations and 1 or 2D attribute plots to compare cosmological simulations to verify time evolution [138]. In such a massive field of study, the ability to gather quantitative evidence in a structured manner enabled greater productivity.

Whilst not relating to 2D hydrodynamic data, Harrison et al. [118] demonstrates a visual analytics system for exploratory and predictive analysis of LIDAR data, combining 2D and 3D scatter plots into linked CMVs. One way coordination links a 3D view of the data with the scatter plots to reduce computational resources and aid clarity during initial exploration. Users may create multiple scatter plots of varying axes, which are linked through logical operators and selected brushing may be viewed both in 3D and in the scatter plots.

An ocean science example is Theron [284], which describes a CMV and brushing approach for analysing paleoceanographic data to understand geolocation. Probing is used to gain details about an object and understand relationships between the views; selecting is used to mark objects of short term interest and painting for objects of long term interest. GeoZui3D [311] once again provides an early, but rare example of CMV in support of coastal ocean science data analysis. A novel, zooming centre of workspace interaction is used to link multiple views (Figure 4.14), in a focus+context approach to a complex visualization system for multidisciplinary, multi-field oceanographic data (modelled and sampled). By tethering the overview to the magnified view, the relationship between the different views is established – regarded as challenging with 3D CMV by the authors.

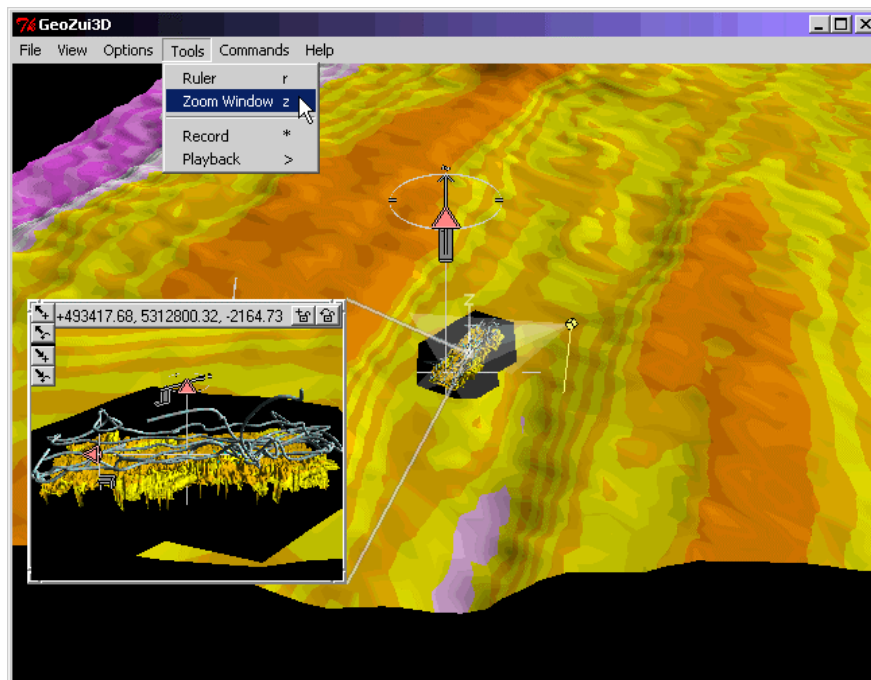


Figure 4.14: Reproduced from Ware et al. [311]. CMV with focus+context: 3D visualization showing the path of data sampling to acquire high resolution data in relation to bathymetry .

Ho and Jern [135] describes an interactive visualization and visual analytics system for studying 4D deep ocean volume data, which illustrates the increasing maturity of visual analytics in some domains of oceanographic studies (Figure 4.15). The system incorporates 3D visualization and 2D image focus and attribute views to support the analysis of spatio-temporal volume data, (regarded as challenging because of rendering time) with a single type of visual representation. Seven, time-linked 2D and 3D coordinated multiple views incorporate navigation including zoom, rotate, pan, brush and highlight. Direct data manipulation includes dynamically moveable orthogonal planes in 3D space, and changing isovalues and opacity. Colours and scales in all views are changed by moving the splitters in the colour legend. Changes take place simultaneously in all views. Attribute views include depth profile, a time graph (controlled by a slider or mouse).

The complementary views highlight different patterns and features of the data. 2D views (focus) support identification of detailed features, and precise navigation and measurement of distance and location, whereas the 3D (context) view provides

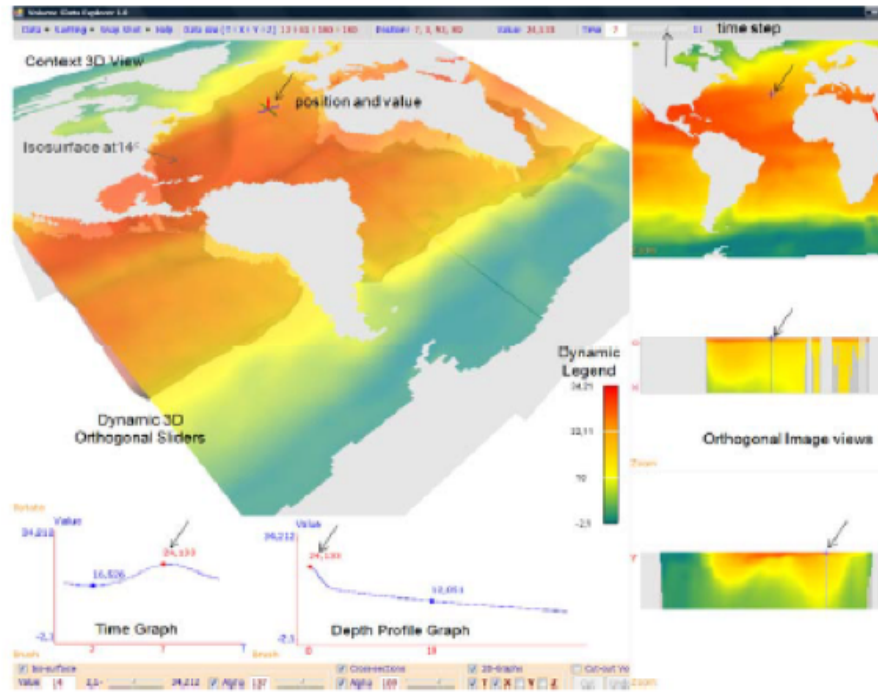


Figure 4.15: Reproduced from Ho et al. [135]: linked coordinate multiple view interface for 4D oceanographic data visual analytics, comprising 3D context view for providing an overview, together with 2D image focus and attribute views. The isosurface embedded in the 3D view represents water temperature changes in relation to depth.

an overall understanding of the data space. In such a complex system, performance might be an issue, and a number of techniques were introduced to optimize this, and maintain immediate interactive response speeds (defined as less than 200 milliseconds). These included data pre-calculation and reuse of data items in attribute views; a lazy calculation method executes a process only after all inputs are changed to avoid unnecessary repetitions (a problem also identified by Peng et al. [223]), plus optimizing rendering performance through the GPU.

However, whilst numerous methods for analysing the data may be incorporated into a CMV system, there are two qualifying factors: as recognised by Ho and Jern [135], the need to maintain interactive visualization speeds (and many of the analytical techniques are computationally demanding), and also an awareness of over-complicating, and possibly contributing to the information overload that visual analytics is intended to resolve. Thus, there is a need for a balanced approach.

4.6 Visual Analytic tools: Parallel Coordinate Plots and other techniques

Earlier sections have discussed the need for complementary techniques to analyse multi-field data. The previous section considered a methodology for visual presentation of the views – Coordinated Multiple Views (CMV) – which is able to facilitate comparison of different aspects of the data. This section considers data analysis tools which might be incorporated into a CMV interface for the type of data associated with this project, focusing primarily on information visualization techniques which have been adopted by visual analytics.

Exploration of large multi-field datasets requires facilities for querying, either by interrogating the whole data as it is visualized, or by placing constraints, such that only specific subsets of the data satisfy the query to enable identification of features of interest. Techniques and tools for spatio-temporal data usually include facilities within the user interface for temporal queries, which filter the data and select time moments or intervals to be represented on the screen [16]: for example TEMPEST’s ‘time wheel’ [82], the range slider of Time Graph [135]. Techniques which support query driven visualization, and which are useful for multi-field data include brushing [237]; probing [284] (enabling users to specify a location or data range); zooming [311] and focus+context [211].

Visual analytics graphs and plots have typically included statistical analysis and attribute views such as Parallel Coordinate Plots (PCP), also other hierarchical clustering techniques, such as scatter-plots and histograms, which allow users to investigate relationships between the data variates and support identification and understanding of attributes of the data [2, 107, 258].

Many approaches for describing correlations have been described, and the review article by Gleicher et al. [104] provides a good overview of comparison techniques for visualization, these include scalar field gradients [81, 256]. Gosink et al. [107] describes a data querying system which represents the cumulative distribution function of all variables graphically, and incorporates a correlation field to reveal trends between any three variables, but this has not been tested on geoscience data.

Similar approaches combining statistical analysis and visual analytics have been applied successfully to studies of tropical cyclone analysis [275]. The statistical analysis highlights data associations, whilst parallel coordinate plots provided a deeper understanding of the relationships. Steed et al. [275] were motivated by the

limited capability of widely used methodologies, making it challenging to detect features of interest. These methodologies also do not lend themselves to standard model comparison and detailed analysis.

The remainder of this section considers some of these tools and techniques in greater detail; identifies the challenges they may present, suggests solutions, and provides a short review of their use in the geosciences.

4.6.1 Parallel Coordinate Plot

Since its introduction by Inselberg [142, 143], and extension by Wegman [315] to multi-variate data, the hierarchical clustering technique of parallel coordinate plots [PCP] has become a widely, and successfully used technique for creating compact 2D visual representations of large, high-dimensional datasets [211, 275, 304]. Each attribute of the data is represented by an individual axis, with axes aligned in parallel and interconnected by polylines representing data events, which cross each axis at a position corresponding to its value for that dimension.

Typically used to display non-spatial, multi-variate data, nonetheless, PCPs have been demonstrated to be successful in demonstrating trends in spatial data, also large, multi-time-point data – characteristics potentially of value in studying the data for this work, where there is a need to retain the spatial context. Coastal shelf hydrodynamic modellers, in simulating flooding scenarios, need to understand the interactions between numerous variables. But identifying and quantifying associations between interrelated variables is widely recognised as a significant challenge, particularly in large data domains such as ocean sciences and climate studies [275].

Researchers believe that PCP confers several advantages over many other techniques. For example, while statistical analysis highlights data associations, PCP are able to provide a deeper understanding of the relationships [208]. The PCP's ability to identify correlations and clusters in multifield data is routinely identified as a major strength, and the only limitation to the number of dimensions visualised is the horizontal resolution of the screen. However, it is recognised that the user may find it difficult to perceive structures or data relations, as the axes become closer [93, 223]. dos Santos and Brodlie [77] adds that all dimensions are treated equally, and also points to a reciprocal duality between Cartesian and parallel coordinates, where points are mapped to lines, rotations become translations and inflection points become cusps. Blass et al. [38] consider the advantages of PCPs to include no loss of data due to projections (unlike scatter plots); the ability to visual-

ize the geometry of high dimensional objects, not simply the data, and, by explicitly linking all points over all dimensions, linking is independent of selection, and valid for all points. Furthermore, of value to the datasets for this work, PCPs have been successfully used for unstructured datasets up to a few million points.

Thus, theoretically, PCPs should be useful to demonstrate trends in spatial and large, multi-time point data; to provide visual clues supporting the discovery of trends; to enable easier identification of areas of interest; to select data for further investigation and to help develop hypotheses for more detailed study of the selected data. The PCP should support scientists in focusing on specific outcomes and queries and ensuring that plausible alternatives have been investigated comprehensively.

However, there are also issues to resolve if the technique is to aid understanding of the data: traditional PCPs suffer from spatial clutter (overplotting): the associated impact on performance inhibits interactive spatial exploration and the clutter hinders perception – for example a mass of overlapping lines precludes identification of trends, correlations or anomalies [93, 211, 304]. Furthermore, PCP requires interaction to be fully effective (i.e. to re-order, repeat, scale, flip and distort the axes to support detailed exploration of the data), limiting the use of PCPs to a few thousands of data items per view, unless techniques to scale the data are applied to large datasets. Fua et al. [93] regards these issues as serious handicaps to query driven visualization. Thus, in developing PCP systems, there is considerable focus on data scaling methodologies, which are discussed later in this section.

4.6.2 Other visual analysis techniques

Scatter and Line Plots

Scatter plots are widely used to visualize both 2D and 3D statistical aggregations of the data [161], and 2D scatter plots have been proved a versatile technique. They lend themselves to direct interaction – it is easier for users to interact in 2D [118] and are thus able to support both overview and detailed understanding of the data, essential for visual analytics. However, Nocke et al. [208] notes the limited capabilities of methodologies relying on techniques such as clustering, as they do not lend themselves to standard model comparison and detailed analysis, and also lead to the loss of overview. Piringer et al. [227] also reports a number of problems associated with their use. Many of these are the consequence of overplotting, where too many points plotted on the same pixel make it difficult to accurately judge distribution.

Furthermore, techniques such as scatter plots, histograms and line plots will require separate or layered plots to analyse multiple variables, as they do not show correlation in the spatial coordinates of the dataset [107], and it is also felt they may not present the optimal approach for perceptual reasons [274]. Equally scatter plot matrices [60], an extension of scatter plots, which present multiple adjacent scatter plots in a single display have also been proposed, although these are extremely resource demanding, and still do not resolve the perceptual issues [275]. Layered plots may be used to condense scatter plot information into a single display, but these too are limited, as a result of layer occlusion and interference [124]. Equally, 3D scatter-plots do not provide an easy solution: whilst providing an additional dimension for separating structures, they do not enable easy interactivity. Furthermore, point display is problematical and lacks sufficient depth cues, making it difficult to judge the three dimensional structure of the points, and binning is challenging [227].

In common with PCPs, these visual forms do not scale well with large datasets, similarly resulting in much overplotting, requiring scaling solutions, which are also discussed below.

Glyphs

Glyphs are described by Fuchs and Hauser [94] as powerful communications items, enabling a large number of data dimensions to be incorporated into geometric or colour attributes of a single shape or symbol, and may be customised to a domain of study to facilitate interpretation. Flow vector glyphs are often used in ocean sciences, as glyph based visualization is recognised as an effective tool for depicting multi-field data [59]. In the case of this work, the use of glyphs is explored as a means of compensating for the weakness of PCPs in encoding direction [230, 304]. One of the requirements of a visual analytics system for hydrodynamic simulations is to support basic calculation of tidal ellipses (to show how tidal currents vary, over time). This requires flow vector glyphs, as the tidal ellipse calculation is made by plotting tidal current vectors with their tails at a single point, as a function of time. Over the tidal cycle, the vectors typically trace a tidal ellipse.

However, there are many different styles of glyph, and, although a mature technique, much research is still taking place in this area [98, 202]: indeed the development of an optimised, advanced vector mapping system for analysing modelled hydrodynamic data for predictive uses would represent a major project in its own

right, and there is no single right answer. This is evidenced by Mitchell et al's. [202] experiment in trying to produce an optimal display of an ocean flow field, where 22 users created 22 different styles, with particular differences in the use of colour. However, a comprehensive review of glyph based visualization is beyond the scope of this work and readers are referred to Ward [309, 310].

Polar Graphs

Polar graphs are widely used in the geosciences, particularly weather studies, as a representation for vectors, where Qu et al. [230] incorporates the polar graph into a PCP. This technique might translate to studies of hydrodynamic vectors. The traditional compass rose is also a widely understood means of conveying direction, which users are able to interpret intuitively. Again, this plot might be used to compensate for the limitations of PCP in encoding direction.

4.6.3 Bespoke problem solving tools

Coordinated Multiple Views provides the flexibility for custom developed tools to be developed for resolving particular analytic problems experienced by researchers, thereby producing derived data for further detailed analysis. A representative example of this in the case of coastal shelf visual analytics is a flux calculation tool. As earlier chapters discussed, early collaboration with coastal shelf scientists identified the need for an efficient and speedy methodology for calculating tidal flux - a calculation not easily undertaken through the simulation model, but one which is central to understanding the potential for flooding.

Coastal oceanographic models have many qualities, in particular their acknowledged ability to support prediction, as many of the phenomena cannot be predicted by non-modelled means: for example the balance of the variables of spring tides, high storm surge, increased freshwater run off, in conjunction with sediment erosion and other local factors, which would result in significant local flooding [239] where historic samples cannot provide reliable predictions of future change because of the additional variable of sea level rise. But whilst modelling is the best tool coastal ocean scientists have, the models have limitations: they cannot currently provide a complete understanding of all the physical processes required to produce the simulations, nor can they totally emulate the real world.

Simulation tools such as TELEMAC-2D provide prognostic variables such as height of bed, water depth, speed and vector flow. But other quantities are usually

required for a detailed analysis of the domain, and these extra calculations and analysis often take place as separate post processing tasks, rather than being included within the visualization system. This derived (diagnostic) data can be a time costly and difficult process, especially as the size and complexity of datasets has increased, placing further strain on the post-processing tools and systems used by ocean scientists. Yet, a common theme throughout this thesis has been the limitations of the visualization and analytics tools routinely used by coastal shelf and estuarine researchers. So many domain scientists have often needed to gain an understanding of programming to be able to coax limited tools and systems to provide them with the derived data they need to complete their analysis.

What is becoming the accepted standard suite of tools in visual analytics, of the range and variety already discussed is likely to be helpful to ocean scientists in understanding their data: based solely on a knowledge of the data types, dimensions and variates, it would be possible for a system developer to produce a visual analytics system to meet many of the research requirements of coastal ocean scientists. But this would not necessarily meet the entire needs of a predictive project such as this, where specific calculations and analysis techniques are essential for the required level of understanding. This is where collaboration between computer and ocean scientists is essential for the development of custom built visual analytics components, targeted at resolving particular analytic problems – which a computer scientist might not even begin to identify simply by studying the data types. An example of this is the calculation of an important hydrodynamic measure, tidal flux, whose attributes are discussed in Chapter 2.

Flux in other work

In researching this thesis, numerous visualization and analysis systems (albeit very few visual analytics systems) for deep ocean and coastal shelf/estuarine hydrodynamics research were identified and studied. Interestingly, only four dealt with this challenging and time consuming calculation, despite its acknowledged importance by ocean scientists. Each example illustrates the importance of adopting a collaborative working methods between ocean scientists and computer scientists.

Stein et al.'s [277] visualization system for investigating modelled water quality data (including pollutants and contaminants) in Chesapeake Bay, the USA's largest estuary, was developed to support scientists investigating the impact of pollution on an extremely valuable fishing industry. Commercial visualization tools did not meet

the needs of the researchers, thus the system developers worked closely with ocean scientists to identify their needs.

The eutrophication model studies the addition of both artificial and natural substances to an aquatic system, and hydrodynamic data was necessary to understand the transport of pollutants throughout the estuary, and its impact on the fishery habitats. The model supplied vector constituent fields, with their values reflecting fluxes entering and exiting the cells (the paper does not discuss the methodology for calculating flux). The role of the visualization system here was to visualize the flux. Using OpenGL, a tool was developed to display the transport flux data, which had not previously been available in earlier visualization systems. See Figure 4.16 reproduced from Stein et al. [277]. Researchers found it useful for clarifying concerns about boundary conditions and their formulation, and how these impacted upon the model.

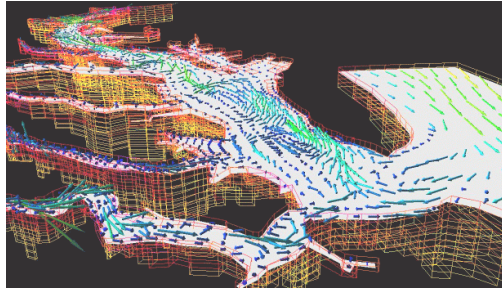


Figure 4.16: Reproduced from Stein et al. [277]: visualization of transport flux data in the Chesapeake Bay.

Similarly, Kitsiou et al. [172] were interested in the visualization of water quality data, and matter transport, using sampled data from the Gulf of Lyons. Whilst not calculating transport flux, the system incorporated a wide range of analysis tools. More recently, as part of the large scale Columbia River Ocean Observatory, Howe et al. [136] describe a visualization system for fishery data, where researchers are trying to understand the factors affecting fish stocks and distribution. One major area of study was a salt intrusion into the estuary, where researchers traditionally visualized 2D slices of the salinity flux, selected by trial and error, and studied them as static plots. Howe et al. [136] developed an interactive 3D tool, which calculated salt flux, allowed users to sweep a plane through the field arbitrarily and visualize the salt flux vectors as barbs, see Figure 4.17. The 3D visualization provided greater insight into the data, but no further analytical functionality was provided, and the

method of calculating flux was not described.

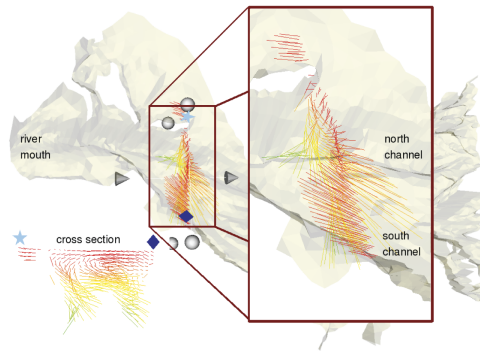


Figure 4.17: Reproduced from Howe et al. [136]: salt flux visualization in the Columbia Estuary

Finally, Cotter and Gorman [65] expresses the view that flux measures are among the most complicated diagnostic quantity to calculate and visualize, and describes diagnostic tools for unstructured oceanographic data (in this case deep ocean convection), including a tool for calculating flux in the water column. They explore a number of the difficulties inherent in the data, particularly the complexity of developing systems for unstructured grids, and the algorithm stimulated much of the flux tool development for this thesis.

4.6.4 Resolving spatial clutter for PCP and similar clustering techniques

The short survey of data analysis tools above indicates that many of these tools are subject to overplotting, requiring data aggregation and abstraction methods to maintain interactivity, reduce spatial clutter and suppress irrelevant details, whilst not losing important information. Creating derived data sets reduces the quantity of data requiring rendering, thereby supporting interactive visualization speeds. This might be achieved by setting the parameters for the data queries as an intrinsic element of the analytic tool (for example using sliders on a PCP to identify subsets of data [273,275]). Alternatively, aggregation and binning techniques [211] are frequently used to reduce the clutter, with the scalability and reusability of binning conferring the advantage not having to re-bin the original data. However, simplification algorithms such as binning may lose the spatial context of the data. Alternatively, data reduction and abstraction may be used to create a compressed but

representative dataset, prior to visualization and analysis [166].

Thus, much of the attention of the system developer will focus on methodologies for optimising the analytical capability of the tool to ensure it is both usable and useful. However, as the examples later in this section demonstrate, the analytic tools often require a suite of methods at various stages of the pipeline to maintain interactive rendering speeds and to provide the necessary degree of clarity within the views, whilst also retaining the characteristics of the data. Hence, many of the extensions and enhancements proposed to PCPs, for example, have focused on methodologies to achieve this, such as Blass et al. [38], which extends the use of parallel coordinates to large multi-timepoint datasets, with a system which uses data quantization and compression as preprocessing, then fast brushing techniques in coordinated multiple views.

Many data abstraction approaches have been described for multi-field data and a principle techniques is brushing, which has already been discussed in the previous section. However, brushing apart, Walker et al. [304] identifies four main categories: Alpha Blending; Clustering; Focus+Context and Frequency and Density plots. To these Fua et al. [93], and dos Santos and Brodlie [77] add dimension reduction techniques. These categories are not clear cut, as many systems incorporate a range of tools and techniques from across the categories to achieve an optimum solution for the needs of a specific application.

These categories are discussed below, together with other techniques, although a comprehensive review of the wealth and combinations of tools available is not possible within the scope of this thesis.

Alpha blending - Transparency:

Wegman's [315] and Wegman and Luos [316] transparency based alpha blending represents the density of the plots with transparency. High levels of transparency are used for extremely dense data ranging to opaque for less dense data. Alternatively, brightness/luminance or colour saturation, may be used in a similar way to transparency, in conjunction with additive blending. However, Walker et al. [304] concludes this is not useful for large datasets, where it is difficult to gain an understanding of the patterns and clusters, thus outliers may be lost, which may be of greater significant to the ocean science researchers, than the general trend. Conversely Nocke et al. [208] found a transparency approach useful for comparison of dense simulation results.

Clustering:

Grouping techniques such as binning and clustering partition data into subsets displaying similarities. Fua et al's. [93] hierarchical clustering algorithm, an extension of the XmdvTool (Ward (1994), [334], uses a data space methodology based on level of detail, rather than data frequency. It structures and presents the data at different levels of extraction, enabling the user to vary the level of detail to achieve multi-resolution visualization, supporting identification of trends and hidden patterns. It uses nested clusters of lines based on proximity, with colour demonstrating clusters, and transparency showing the mean and extent of the cluster. The hierarchical system, based on a tree structure, is categorized as implicit where similar objects are grouped on a specific metric, for example Euclidian distance. Thus, an approach based on spatial indexing [249], which moves data from the geographic location to fill the graphic space more efficiently [13] may well be appropriate for this type of algorithm.

Likewise Gabriel et al. [97] demonstrate a similar approach, using oceanographic data as exemplars. Johannsons [150] approach is to transform each cluster into three textures (animation, outliers and structure), which are then combined into a polygon.

However, there are a number of concerns about clustering techniques: they may cause information about the original data to be lost, thus arguing the case for views displaying the original data on demand, so they may be compared with the derived data [208].

Focus + context

Focus + context is an important technique for PCP development: first described by Furnas [96], it is widely used to overcome display limitations. Within the same image, context is provided by a general overview of the data, whilst focus is derived by visually highlighting a sub-set of the data. The tool integrates a visually accentuated representation (through greater detail and opacity) of selected data items in focus with a visually de-emphasised representation of the rest of the data (the context) [211].

There are numerous methods for providing focus + context. For example, a Sampling Lens, together with an automatic sampling algorithm might be used in conjunction with parallel coordinate plots and scatter-plots to study patterns and trends within dense areas of the data whilst retaining context [83]. Alternatively,

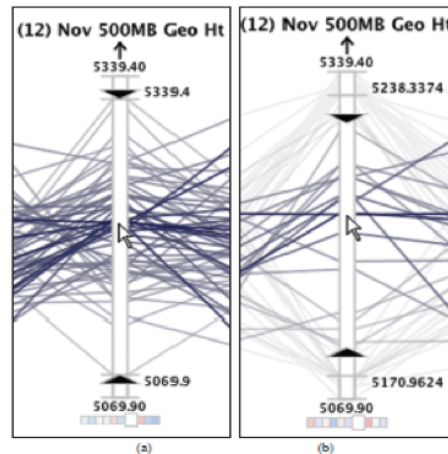


Figure 4.18: Reproduced from [274]: Dynamic axis scaling: (a) before, (b) after. Moving the mouse wheel in the focus area of the axis moves the values of the upper and lower limits closer together, facilitating zooming into the central axis.

dynamic axis scaling has been used in PCP to identify data sub-sets [275], accompanied by aerial shading, and also double ended sliders on each end of the axis to highlight lines for selection and more detailed study. See Figure 4.18.

Similarly, Jern et al. [147] uses dynamic axis scaling within multiple linked views for large geospatial datasets. Kehrer et al. [160] describes a four level focus+context visualization, with three levels of focus in every attribute view, where the viewer specifies logical combinations of features. A flow visualization approach is described by Doleisch et al. [76] – SimViz 3D’s focus+context visualization of flow features modulates transfer functions for colours and opacities, plus sizes of singular visualization objects per cell.

Zooming

Zooming is also recognised as a useful tool for providing focus + context, and is able to rapidly change the viewpoint of the data to support better understanding of interrelationships [110, 311]. Zooming in provides detail, whereas zooming out gives context. Fua et al. [93] uses dimension zooming to support distortion techniques for exploring dense information, through the selective enlargement of data subsets. At the same time, context is maintained with the surrounding data, through a mini map showing the position of the zoomed information within the entire data space. By animating the zooming process, differences in scaling across the dimensions are shown, also the effect on data points neighbouring the area of interest.

Thus, the user is able to study localized trends, whilst retaining context.

Novotny and Hauser’s [211] algorithm

The algorithm is potentially useful for coastal shelf hydrodynamic data, where outliers are features of interest (albeit, it is more limited in demonstrating relationships between the axes) [304]. It uses a binning process on the PCP to permit visualization at several levels of abstraction, enabling the representation of clusters, outliers and trends. This is particularly important when using focus + context tools, as outliers may have a detrimental effect on the accuracy of the context element. Binning converts the data to a frequency based representation, by dividing the data space into a set of multidimensional intervals (bins). Each bin is assigned an occupancy value which determines the number of data records that belong to the bin [267]. Novotny and Hauser’s output orientated approach, bins not to the original n dimensional data but to the 2D visualization space. The algorithm places the two dimensional subspace (a pair of adjacent axes representing a pair of dimensions in the data) into bins, which form a bin map, represented as a 2D histogram of the distribution of all line segments between the two axes. This achieves a scalable visualization of the large data, without enormous memory demands. Other advantages include:

- The size and number of the bins determine the precision of the aggregation.
- The bins are reusable without the need to re-bin the original data (e.g. to generate a coarser representation), making hierarchical binning possible
- The algorithm lends itself to performing binning on the GPU.

An extension of Novotny and Hauser’s [211] algorithm, by Rubel et al. [246] incorporates adaptive histogram bins providing a higher resolution in highly dense areas of data.

Frequency and density plots:

A number of clutter reduction algorithms have been proposed based on data frequency, involving aggregating and filtering the data through binning. Rodrigues et al. [244] uses frequency plots, computing frequency in 1D space defined by each attribute. Artero et al. [21] adopts this type of approach, subsequently used by Novotny and Hauser [211], similarly Blass et al. [38]. Following the binning process, the data frequency is able to be represented by a histogram, which is useful for

revealing clusters and outliers. However, whilst Walker et al. [304], regard 1D point based histograms as useful for data overview, another solution is required to show the relationships between the data axes, for example by extending the histogram with a vector based approach [101].

Dimensional reduction techniques:

These reduce the amount of data presented, either by studying only sub-sets of the data, or by mathematical means to derive new data whilst retaining relationships - the premise being that studying filtered sub-sets may reveal insight into patterns and trends. Examples include coordinated multiple views with brushing or other feature identification device (which have already been discussed); the 2D HyperSlice by van Wijk and van Liere [298], and an extension of this, the 3D HyperCell by dos Santos and Brodlie [77]. Both are useful for visualizing multi-variate scalar data, and the direct relation between screen space and data space supports interactive feature identification. dos Santos and Brodlie [77] considers HyperCell improves the ability to identify features, and also supports analytic tools such as brushing, rotation and cell splitting.

Heat maps

Heat maps and associated choropleth maps have been used in conjunction with scatter plots and PCP. Harrison et al. [118] reports on the use of heat maps showing brighter colours for over-written points, to resolve the problems of overplotting commonly associated with scatter-plots, in a tree based, hierarchical tool for studying LIDAR data, This enables features of interest to be more easily identified, for further, iterative study. Equally, Andrienko and Andrienko [10] uses choropleth maps and spatial references with a PCP. The PCP is dynamically linked to the choropleth map, which identifies attributes by colour, with the strength of the attributes dominance shown by the degree of darkness of the colour.

Wavelet analysis

Wavelet analysis is a commonly used technique for data abstraction and data analysis, particularly for analysing variability in meteorology and climate studies. Janicke et al. [144] extends the approach from a limited number of time series to large, multi-variate, 2D climate datasets, by coupling with clustering to identify contiguous spatial sub-sets. The data is visualized as a colour map and presented in multiple

views. But within ocean sciences, wavelet analysis is mostly confined to deep ocean and climate study models, and its use has been extremely limited in estuarine and riverine domains [177].

4.6.5 Axis/dimension reordering

Nocke et al. [208] believes there is scope for enhancing the interaction capabilities in Visual Analytics systems with tools such as PCP, scatter-plots, glyphs using interactive filtering, where more sophisticated interaction techniques, such as reordering and highlighting would improve navigation in high dimensional space. Similarly, Peng et al. [222] believes that varying the dimension order can enhance the visualization's expressiveness, without reducing information content or disturbing the data.

From the papers reviewed in this survey, it appears that reordering techniques are becoming more commonplace, for example a number of PCP applications using reordering have been identified [38, 230, 256, 275, 284], although not all of these apply to modelled data.

However, dos Santos and Brodlie [77] expresses concern that when there is no pre-defined order for arranging the dimensions, relationships may be obscured by an inappropriate arrangement of the dimensions. This view is shared by Peng et al. [222], believing that completely different conclusions might be drawn, dependent on the order of axes, and therefore proposes a solution based on a clutter reduction algorithm.

Equally, there is a need for flexibility within the axes to meet the analytic needs of researchers, and Blass et al. [38] regards axis ordering as a vital part of the PCP, with Qu et al. [230] concluding that axes presenting a potential correlation should be placed close together. To achieve this, a weighted complete graph (a variation of Sauber's [256] multifield graph) shows the relationship between the dimensions, to generate an optimised axis order.

Other solutions tend to be varied, ranging from Blass et al. [38] and Theron's [284] simple interactive axis reordering through drag-and-drop, to quantifiable statistical approaches. Steed et al's. [275] approach is based on the statistical computations which are an intrinsic part of the climate model. Selection from correlation coefficient, interquartile or standard deviation range, MLR coefficient, or SLR r^2 values - with positive correlations placed to the right of target axis in descending order - enables the user to focus quickly and easily on the strongest correlations.

4.6.6 Data analysis tools in the geosciences

Many algorithms involving a wide array of solutions have been proposed, including geoscience applications, although there are scant few examples of the use of PCPs solely for oceanographic research, particularly coastal shelf and estuarine data. However, there are numerous examples for meteorological and climate research, which present similar analytical challenges. Qu et al. [230] describes the data challenges of weather data visualization, which may be resolved by the use of PCPs: the simulations are spatio-temporal, typically multivariate vector data, often consisting of more than 10 dimensions. The wind speed and direction vector data differentiates them from most scalar multivariate data, so conventional visualization techniques, such as scatter-plots and glyphs, do not provide satisfactory solutions.

Nocke et al. [208] applies parallel coordinates, transparencies and graphical tables for mapped variables to compare simulated climate data. However, the effectiveness of PCPs is regarded by some researchers as more limited for vector data [304], thus Qu et al. [230] proposes a combined approach incorporating a PCP, a polar system (widely used for representing vectors) and a weighted graph for ordering the axes of the PCP, Figure 4.19.

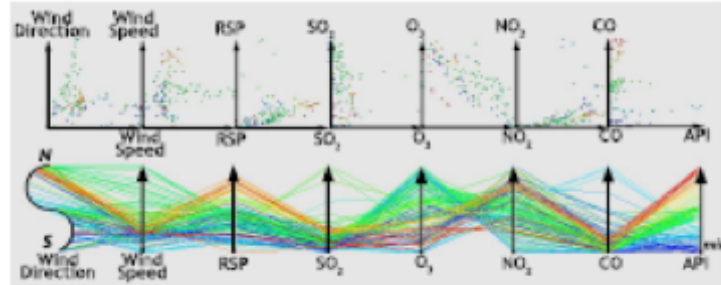


Figure 4.19: Copied from [230]. This shows PCP with S-shaped axis derived from polar graph, to encode wind direction.

Blass et al. [38] extends the use of PCPs to visualize and explore multi-time point, volumetric, meteorological data with tens of millions of data points, enabling relationships to be easily identified, through the explicit linking of the PCP.

PCPs have been used in climate modelling with an oceanographic component, which presents similar challenges to modelled estuarine flood simulations in relation to their complexity and data characteristics. The work of Steed et al. [273–275], who have presented parallel coordinates systems with a rich array of tools for highly

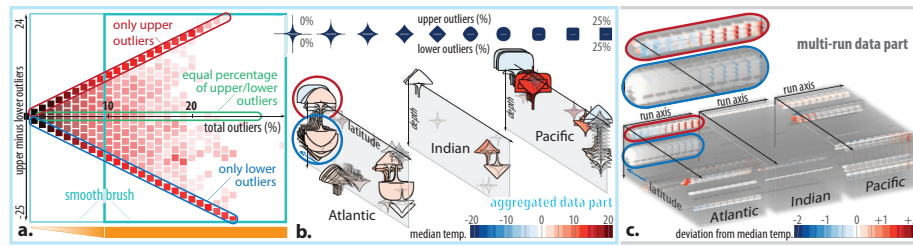


Figure 4.20: copied from Kehrer et al. [161]: a) scatterplot showing the percentage of total outliers (x axis), and measure to determine how outliers are distributed (y axis)

complex, statistically based, predictive studies of tropical cyclone and hurricane systems, incorporates many of the more recent extensions and innovations to PCPs. Like many estuarine researchers, climate scientists have traditionally relied on simple plots and graphs and side by side comparison of static plots for analysis of their data. These have been found to be particularly detrimental for seeking out combinations of conditions leading to events. However, in evaluating Steed et al.'s. [275] PCP system, climate researchers concluded that a single frame of it enabled observations which would have only been possible by examining hundreds of plots using traditional methods.

The array of techniques includes dynamic axis scaling to modify the focus area of a selected axis, with each axis divided into three partitions to provide a central focus and outer context areas; frequency information display; histograms; dynamic axis reordering; axis inversion; details on demand, and aerial perspective shading - an approach similar to that of Jern et al. [147] for large geospatial data sets.

The PCP is also used to analyse inconsistencies between cosmological particle simulations, where the tool extends the limitations of a spatial view in uncertain data that is too complex for the spatial view alone [117]. In paleoceanography, researchers have developed an interactive focus+context PCP, together with brushing, to allow researchers to visually reconstruct paleoenvironmental features, in support of climate change research [284].

Scatter-plots are frequently incorporated into visual analytics systems for the geosciences. Kehrer et al. [161], in a visual analytics system for multi-run climate data, uses scatter-plots (Figure 4.20) to show aggregated properties for all grid cells and timesteps, in the statistical analysis of outliers, as does Wilkinson et al.'s. [318] weather data system.

Harrison et al. [118] uses scatter-plots in a linked feature space approach sup-

porting studies of non-uniform LIDAR data for terrain exploration and urban disaster management. The iterative system creates multiple scatter-plots, establishes chains of plots based on selection domains then links plots using logical operators, to identify features of interest and subject these to detailed analysis.

These examples cited are representative of the analytical tools and techniques, also the geoscientific domains where they may be found, and reinforce the conclusion that coastal shelf and estuarine studies might prove useful areas of study for extending their use.

4.7 Towards requirements of an investigative coastal data visual analytics system

This short overview of the methodologies, tools and techniques typically associated with visual analytics systems for the geosciences, suggests that many of them are common to most applications and domains. What also emerged is that the main criteria in developing an application based system should be identifying and balancing the challenges presented by the data and the research priorities of the user. Thus, the main focus of this section is to draw together all elements of the related work to establish an approach for developing an appropriate visual analytics system based on these criteria.

The previous sections have identified material used in shaping decisions and guiding development of the tools built in this dissertation. A full survey of the state of the art of either visualization or visual analytics was not attempted, but the aim was to provide sufficient level of detail to place the current work into context, and to seek to establish the requirements of a visualization and visual analytics system for coastal shelf simulations.

4.7.1 Context

Given the variety, range and extent of visualization and analysis systems applied to the geosciences and its numerous domains, identification of appropriate requirements may present a daunting task. Unsurprisingly, in an area of study with such complex and challenging data, and in the absence of an established and recognised overarching advanced visualization framework, methodologies have been proposed for analysing the individual attributes of the data type to establish the most appro-

priate solution [2, 16, 57, 94]. The starting point for these is the resolution of a problem relating to an attribute or inter-relating attributes of the data its type (spatial, temporal, dimensionality, variate); its derivation (sensed, sampled or modelled) or the physics of the data (e.g. fluid flow). Proposed solutions or techniques are then tested using geoscience domains as case studies. Examples include [108, 306] – spatio-temporal data, [123] – multi-dimensional data; [303] – type of numerical model, [283] – interpolation of sparsely sampled data, [197] – complex, multi-component spatial datasets; [118] – Visualization of LIDAR data, [262] – scaling extreme scale datasets.

Whilst such tightly focused approaches have the value of developing and testing hypotheses or methodologies, which may then be applied to a number of domains, they may be too narrowly focused to be of value outside the area of study for which they were developed. Equally, they may not consider visualization pipeline issues which may arise as a consequence of the approach, and which may affect the usability of the whole system.

Thus, other approaches focus on visualization pipeline methodologies for managing the inherently challenging nature of the data (its volume, complexity and attributes), either to support analysis of the data, and add to its understanding, or else tools and techniques to optimise performance. Examples include the Global Ocean Assimilation Experiment (GODAE), with its suite of tools for discovering, sharing, analysing and optimising oceanographic visualization; Blower et al. [39, 40] – who presents systems for optimum management of data transport and processing; Stockinger et al. [279] – which describes a data scaling tool; Coelho et al. [63] – parallel volume rendering to achieve interactive visualization speeds for volume visualization; Liang and Molkenthin [185] – web based collaborative visualization environment; Wilde et al. [317] – immersive and 3D viewers for 3D data; Patrikalakis et al. [221] – coupling biological and physical oceanography with acoustics through distributed/Grid computing in support ocean modelling and forecasting.

However, another approach is user-focused, and one which supports the development of bespoke application based visualization systems, to enable scientists to solve problems [65, 121, 229, 277], understand key relationships and correlations [1], and where needed, develop predictive tools, which will provide the scientific underpinning of, for example, environmental disaster planning. Examples include Harrison et al. [118] storm surges and flooding, Dyer [80] flooding and DeAmicis et al. [70] pollution incidents.

This approach requires data and visualization pipeline needs to be considered in

the context of the ocean scientists' research goals, rather than in isolation of them. It turns on its head the process of developing a new visualization tool or algorithm, which is verified by a domain related case study, to one of placing the needs of the user at the centre. The following requirements have been developed for this approach.

A number of comprehensive methodologies for evaluating and categorising the nature and attributes of multi-field data in the context of visualization and analytics systems have been proposed [2, 13, 16, 94]. These have informed the identification of these requirements and also the development of VINCA.

4.7.2 Requirements

Requirement 1: Consider the right user/system methodology This encapsulates the user focused, collaborative approach of this work discussed above, balanced with a detailed understanding of the attributes of the data and their visualization and analytic challenges, underpinning an optimized application based system, as exemplified by Cotter and Gorman's [65] diagnostic tool for 3D unstructured oceanographic data.

Requirement 2: Create a system that is able to deal with temporal data; operate at different scales and deal with all the complexities of multi-field data. The related work has demonstrated that one of the major limiting factors in model simulations is analyzing and understanding the complexities of multi-field data, where the temporal component often presents the greatest concern, and also where the tools and techniques required for spatio-temporal data, are not helpful for understanding other attributes of the data. Thus any system supporting prediction and planning (such as flood management), must incorporate a range of tools for locating salient features and understanding all components of the data. These will typically include animation and static plots; 2D and 3D visualization and attribute graphs and plots such as PCP, mostly derived from information visualization, but also including bespoke tools, developed to meet unique aspects of researchers' needs.

Requirement 3: Enable multiple, linked views of the data. Once it is accepted that different types of tools (both views and attribute plots) are required to understand all fields of the data, then an appropriate interface is required which will permit interaction, cross-reference and comparison to enable correlations

to be made and relationships established. A technique such as Coordinated Multiple Views enables users to view their data through different projections. The coordination provides additional value by explicitly linking (and highlighting) like features in one view to the other.

Requirement 4: Ensure the system and tools are underpinned by an understanding of theories of cognition and the recognition of interaction as a query tool. Query driven visualization is a central principle of visual analytics, enabling the identification of features of interest for further detailed study and analysis through an iterative process. This requires an understanding of cognition and human interfaces as applied to visualization to develop appropriate perception driven, interactive system.

Requirement 5: Consider performance, ensure the system includes techniques and systems to maintain interactivity, and ensure its usability and usefulness. Whilst the focus of the system is its visual analytics, there is a need to balance speed and performance with optimum analytical capability. Not only must the system be capable of consistently achieving interactive processing speeds to be able to support query driven visualization of the extremely large datasets, but it must also be able to provide methods of dealing with the frequently computationally heavy demands of the visual analytics tools themselves. Thus, together with solutions based on architecture or hardware or workflow management, techniques such as hierarchical data structures (spatial indexing) and the whole array of aggregation and abstraction solution should be considered – but once again, within the context of the particular challenges presented by the hydrodynamic data (AMR, multi-field, modelled).

4.8 Summary

The need for advanced visualization and, more recently, visual analytics techniques for detailed investigation and understanding of large datasets of multi-field data is discussed in relation to ocean science research in particular estuarine hydrodynamic studies, and the wider geosciences context. Whilst visualization is well established to explore, analyse and communicate information about ocean science data, the increasing size, complexity and level of detail of both sampled and modelled datasets has necessitated visual analytics techniques to aid knowledge discovery, as this ‘big data’ has, paradoxically, resulted in a diminished ability to gain insight. A further

strong impetus for the evolution of visual analytics has been the requirement to understand and predict the likely consequences of climate change and associated sea level rise.

The challenges presented by a particular ocean science research problem, and how to resolve them forms the basis of the related work: the need for estuarine researchers to compare and contrast investigations and parameterizations of their data, and thus understand their flooding simulations, in support of coastal zone and flood prevention management. These challenges include a research environment where traditionally, researchers have not taken full advantage of advanced visualization and visual analytics tools, preferring to rely on model post-processors and analysis software with limited visualization capabilities so there is a dichotomy between their techniques and the available tools.

In a recent review of the visual analytics of movement, Andrienko and Andrienko [14] do not include ocean science hydrodynamics, arguing they have scarcely been addressed in visual analytics. This supports the case that there is a real need to develop visualization and visual analytics techniques for estuarine, modelled hydrodynamics.

There is also a requirement to make this process collaborative: however, a collaborative application-based approach with domain researchers is recognised by Johnson [152], van Wijk [299] and Keim et al. [165] as one of visualization and visual analytics major challenges. This data-centric, user focused approach is reinforced by a study of the attributes and challenges of visualizing and analyzing heterogeneous data of this type, for which no overarching framework, systems or common methodologies exist [2].

Reviewing the attributes of the data in the context of the ocean science research objectives suggested the need for an interactive visualization and visual analytics approach to support advanced analysis and prediction based on:

- an understanding of human cognition and theories of human/computer interaction
- an interface with a range of tools facilitating interaction with the data and permitting comparison and study of linked, complementary views and graphs of areas of interest within the data
- methodologies for abstracting and aggregating the data without losing its essential characteristics and relationships, to hide complexity, support analysis and maintain interactive visualization speeds

- query driven visualization enabling the identification of features of interest and the derivation of sub-plots of data for further, iterative analysis
- application of advanced analysis techniques, reflecting the attributes of the data, derived from data mining and information visualization
- custom development of tools to improve researchers' abilities to undertake difficult calculations
- an visualization system architecture which optimises both analysis capabilities and performance (speed and quality of rendering)

The review of visual analytics for the domain and the wider geosciences revealed a commonality of approach which focuses on finding location relation patterns and relationships [147], together with a wide variety of individual tools and techniques, dependent on the specific challenges of the data and the research requirements – suggesting there is no one right answer [163,300].

A number of techniques and tools were identified which present promising avenues of approach in meeting the requirements for a visual analytics approach identified above. These include:

1. Data abstraction and aggregation techniques, including visual space algorithms, also data space methodologies, e.g. spatial indexing, being one, wavelet analysis.
2. Linked, coordinated multiple views views (e.g. 2D and 3D, static plots and animation) for providing different representations of the data and enabling dependencies to be studied
3. Selection of areas of interest, using techniques such as brushing, zooming, focus+context, transparency, binning to further refine the data
4. Data analysis techniques and attribute plots, scatterplots, histograms, parallel coordinates, polar graphs, heat maps

The state of the art review demonstrated a rich variety, across all the geosciences, of visual analytics tools meeting complex research needs in meteorology and climate studies, geology, cosmology/astronomy and some aspects of deep ocean studies (many associated with meteorology). However, one of the main challenges identified remained overplotting, and many of the analytical methods used themselves contribute to overplotting.

Thus, the challenge is to develop an interactive visualization and visual analytics system which will provide all the analytic tools researchers require to meet their predictive needs, but at acceptable levels of speed (to retain the essential interactivity) and quality of rendering.

CHAPTER 5

The VINCA visualization system

This chapter presents the development of VINCA (VIsualization eNvironment for Coastal Analysis). It establishes system requirements and explains how the system was built and operated. The system has been developed in close collaboration with ocean scientists at the Centre for Applied Marine Sciences, Bangor University, who are using the TELEMAC-2D numeric model to research estuarine hydrodynamic flow in the context of flood prevention. Cotter and Gorman [65] recognises the importance of collaborative working between ocean scientists using numerical models and visualization researchers, because of the complexity of modelled data, and the challenges they present.

The collaboration process for this project not only included discussing scientists' needs, but close observation of how they analyse their datasets; identifying the analytical problems they encounter, and producing two, 3D visualization prototypes using generic MVEs (OpenDX and VTK). These provided familiarity with the challenges of visualizing TELEMAC-2D hydrodynamic data, and enabled further feedback from users to be obtained. The process informed the establishment of specific requirements for what would become a turnkey system; decision making about the nature and type of visualization and analysis system to be developed and the choice of development tools and systems.

Therefore the chapter contains nine sections, focusing on the development of the tools.

- Establishing system requirements
- Design and initial implementation
- Data Import
- Spatial Indexing
- Interfaces and interaction
- Coordinated and Multiple Views
- Views, attribute graphs and plots
- System performance
- Quality results

5.1 Establishing system requirements

Related to other projects, earlier prototypes had been developed to visualize coastal shelf tidal flow data (using depth averaged data), generated by the TELEMAC-2D modelling system, with both sets of data presenting similar challenges. The development of these prototypes helped influence the design choices when developing the Visual Analytics tools for the main part of this thesis.

Prototype One (P1) Figure 5.1 was a visualization of the immensely complex hydrodynamics of the Menai Strait, using OpenDX, and was an application tool developed specifically for that oceanographic domain. Prototype Two (P2) Figure 5.2 aimed to develop an extensible system for use with any TELEMAC-2D dataset, which would also permit bespoke tools to be added to meet the specific analytic requirements of the research. This prototype was built with VTK, through Java Bindings, using Dyfi Estuary data, which, in part is used for the current work. A detailed review of the strengths and weaknesses of both systems and the MVEs used to develop them is found in George and Roberts [103].

A number of limitations were observed which informed the identification of system requirements for this work. Although only two generic development systems were tested through the prototypes, they supported the view that there is not much to choose between generic visualization tools when considering the balance of their strengths and weaknesses. Each is capable of resolving some, but not all the problems encountered with the data. Often, the tools require complex solutions, and

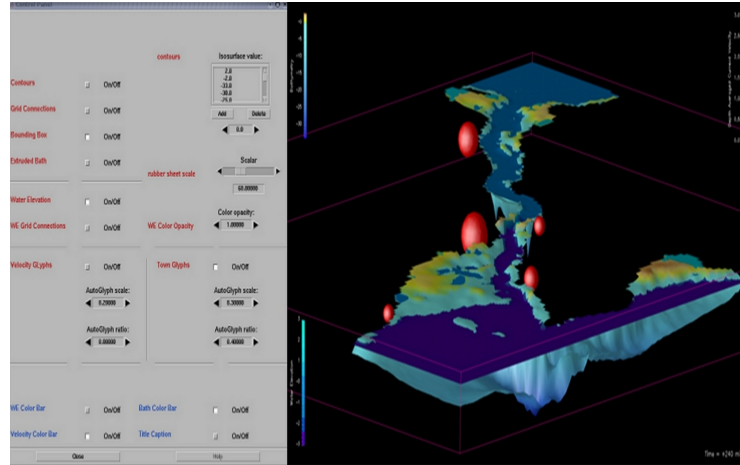


Figure 5.1: Prototype 1(P1)

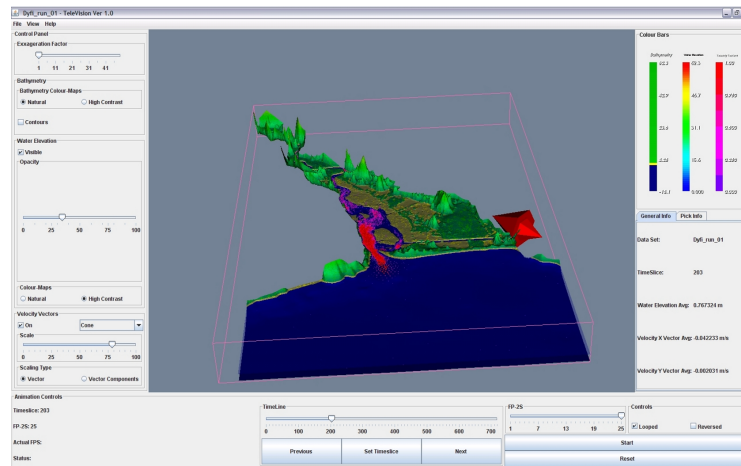


Figure 5.2: Prototype 2 (P2)

modifications to the visualization system to resolve problems to achieve full functionality, particularly when dealing with extremely large datasets; the complexities of modelled AMR data, and when attempting to develop bespoke tools to deal with specific research requirements.

Whilst both prototypes were regarded by ocean science researchers as useful in providing insight into their data, the speed/rendering quality dichotomy was an issue, particularly in view of the size of the datasets, and it was recognised that both architecture/hardware and visual analysis solutions might be necessary to achieve significant improvements.

5.1.1 Identified Challenges

Earlier work developing prototypes P1 and P2 from other projects, and the data sources described in Chapters 2 and 4 identified several challenges, which were borne out by the prototype implementations:

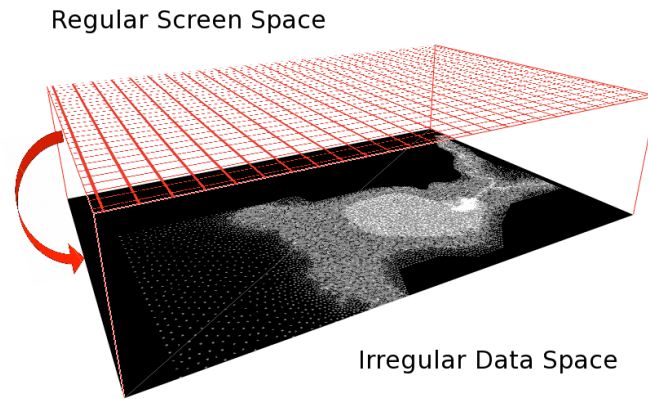


Figure 5.3: Dyfi estuary data, showing data association problem

1. it is difficult to manage extremely large and complex data-sets of multi-variate, spatio-temporal data;
2. unstructured scattered data proved challenging in the current prototypes;
3. overplotting was a major issue: the dataspace for this project is irregular composed of large scale, unstructured adaptive grids, which create a data association challenge requiring an efficient algorithm to associate regular screen space with adaptive, irregular data space, Figure 5.3;
4. the generation of multi-resolution grids caused problems with the rendering. As a result of the use of Adaptive Mesh Refinement (AMR) techniques, resulting in undersampling offshore and massive overplotting in areas of significant interests. Figure 5.4, reproduced from Robins [243], shows the mesh configuration for the present day Dyfi Estuary and river channel, used for simulations of velocity. The mesh resolution varies from 500m offshore, to 50m in the estuary and 25m in the river channels).
5. There was much coincident topology. This was a particular problem with the 3D visualization prototype developed using VTK, where multiple surfaces at exactly the same point in 3D space, affects the ability of the renderer to

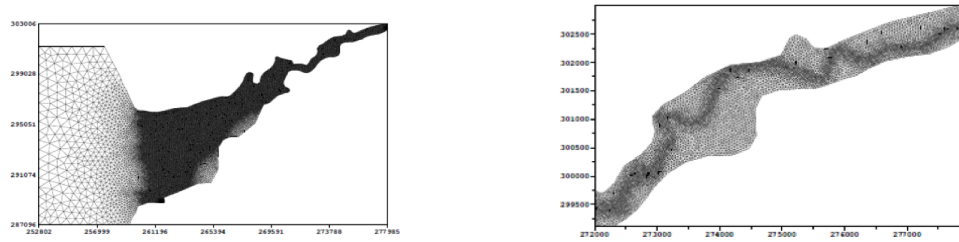


Figure 5.4: shows the model mesh for the Dyfi Estuary and river channel – ranging from resolution of 500m offshore, reducing to 50m in the estuary and 25m in the river channel, reproduced from Robins [243]

make the correct selection (Figure 5.5, illustrates the outcome – a flickering, speckled effect). Whilst a solution for the VTK prototype proved challenging, a similar problem with the earlier OpenDX prototype (a 3D visualization of TELEMAC-2D hydrodynamic data of the Menai Straits) did not arise.

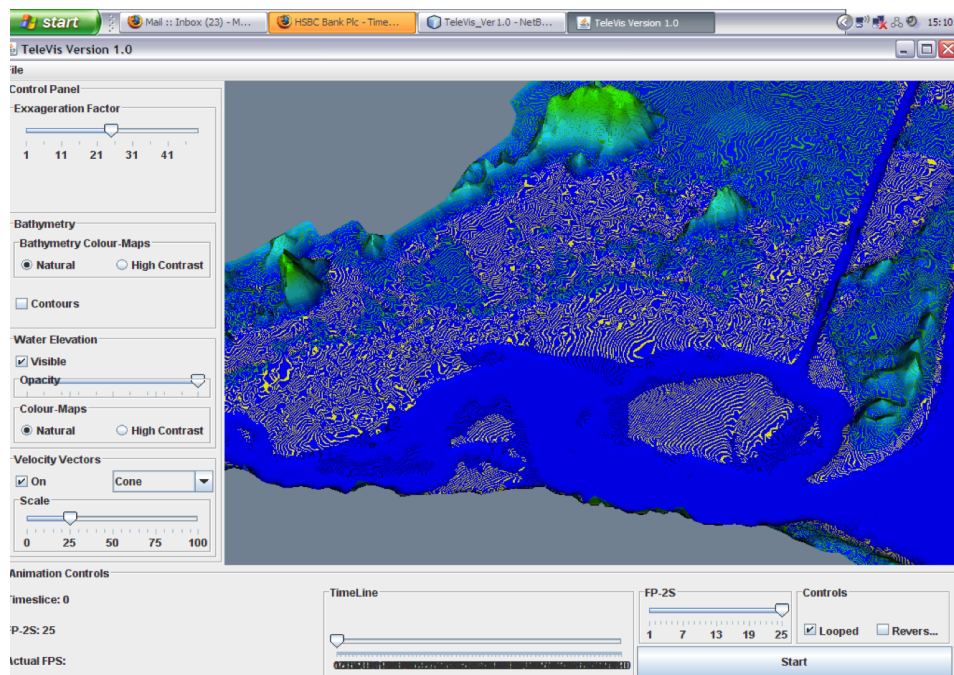


Figure 5.5: prototype 3D VTK visualization of Dyfi Estuary TELEMAC-2D data with water level at full opacity, showing problem of coincident topology

In addition, TELEMAC-2D, although a widely used model, has limited post-visualization capabilities, as discussed in earlier chapters. At the outset of this PhD, the visualization capability of TELEMAC-2D was achieved through the Rubens post-processor. This was used (as part of a suite of tools) by the School of Ocean Sciences, Bangor, especially in conjunction with the tools and systems available

in MATLAB. The limitations of MATLAB as a visualization tool are discussed in Chapter 2, where Rubens shortcomings are succinctly summarised by Hayir et al. [120] as not user friendly; unstable and slow. Therefore, users typically used Blue Kenue ¹, that is able import TELEMAC-2D data sets.

Very few papers have been identified which discuss the use of Blue Kenue as a visualization tool, or provide a considered review of its visualization and analytics capabilities, except for a technical report relating to the use of TELEMAC-2D and Blue Kenue to study lake flow data. This concluded that its animated visualization of the data was regarded by users as overly complex for general interpretation and communication [145]. In the absence of objective user appraisal, as part of this research, a sample visualization of the data was produced using Blue Kenue. It was found to be difficult to learn and not intuitive, thus it was difficult to see how users would be able to take full advantage of its features. This seems to be borne out by the lack of published research evaluating its capabilities. This conclusion is also shared by the collaborating oceanographers.

An important requirement for the visual analytics system has been identified as the ability to calculate tidal flux. It has been highlighted that flux calculation is difficult to accomplish using tools such as MATLAB, yet, as has been described in Chapter 2, tidal flux is essential for the calculation of the tidal prism to establish the conditions under which areas around an estuary flood [239]. Discussions with the researchers identified that the complex and time consuming nature of the calculations (which can take up to two hours per calculation) meant they were limiting the number of flux calculations made along an estuary, and thus possibly missing important comparisons and relationships.

Concurrently, with the development work undertaken for this thesis, the ocean scientists have been developing analytic tools using Blue Kenue, and over the three to four year period, they have produced an alternative flux calculator. However, the tool requires manual calculations, is complex and time consuming and does not have the same level of functionality as the tool presented in this thesis ². Thus, a further requirement of the system was identified as the need to produce a more intuitive and user friendly visualization system, which would enable improved insight into the data. This would also include the flux calculator, which, based on visual analytics tools, would be able to calculate and analyse all locations in one step, and would, therefore, present a significant improvement upon either existing method.

¹http://www.nrc-cnrc.gc.ca/eng/solutions/advisory/blue_kenue_index.html

²e-mail exchange 19.11.2012 with Dr P Robins

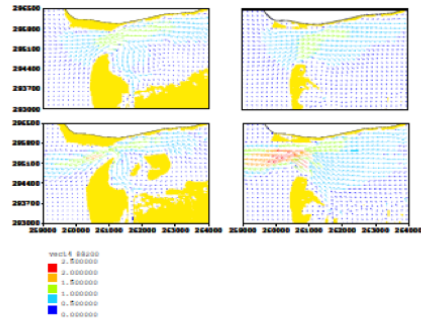


Figure 4.15: Velocity vectors (magnitude and direction) during peak flood (top) and ebb (bottom) flow for Scenario 1 (left) and Scenario 4 (right). The vectors have been interpolated from their nodal positions onto an orthogonal (regular) grid for clearer visualisation. The yellow areas indicate dry land.

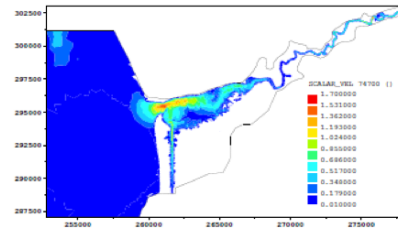


Figure 4.10: Contour map showing scalar velocity (m s⁻¹) during peak flood flow from Scenario 1 (mean Spring tide, mean annual flowrates, no storm surge).

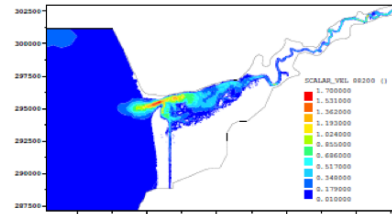


Figure 4.11: Contour map showing scalar velocity (m s⁻¹) during peak ebb flow from Scenario 1 (mean Spring tide, mean annual flowrates, no storm surge).

Figure 5.6: Reproduced from Robins [243]. Static plots of velocity vectors generated using Blue Kenue.

Figure 5.7: comparative screen shots of ebb and peak flow velocity, generated using Blue Kenue, reproduced from Robins [243]

Discussions with ocean scientists and observation of their analysis techniques revealed that much of their time is spent comparing and analysing simulations of different flooding scenarios, which often require side by side comparison of static screenshots and attribute graphs. However, as the Related Work indicated, such a technique is widely regarded as problematic: with the rapidly increasing size and complexity of the data, it is unlikely to provide the required levels of insight [268, 307]

Figures 5.6 and 5.7 illustrate typical screen shots generated using TELEMAC-2D and Blue Kenue, and demonstrates the difficulties in observing the data to gain the detailed understanding of underlying trends and patterns required for predictive studies, where there is no facility to zoom into the data and identify areas of interest for further detailed analysis. Thus, enhanced, interactive analytics systems and tools would be useful, for example coordinated multiple views, to support a range of views and attributes plots to provide quantitative information and to facilitate comparison and aid understanding of different attributes of the data; to support data query and facilitate study of specific areas of interest and to reduce the amount of data needed to be considered at any one time.

Finally, the two prototypes also indicated the contribution that more enhanced visualization might play in demonstrating and explaining highly complex hydro-

dynamic processes (this was recognised by the ocean scientists in particular relating to the OpenDX visualization system for the highly complex flow of the Menai Strait [103]. Thus, another requirement for the system would be to support the ability to export publishable quality outputs.

5.1.2 Summary of requirements

By considering the factors outlined above, and the hypothesis in Chapter 1 (Introduction), the following set of requirements were identified:

1. Development of a useful and user friendly interactive visualization and visual analytics system, which would provide higher levels of functionality for estuarine researchers than their current model analysis and presentational systems.
2. The ability to deal with the challenges presented by modelled data.
3. The ability to support knowledge discovery, through query driven visualization, enabling areas of interest to be identified, selected and subjected to detailed study, including comparison and identification of underlying trends.
4. The ability to obtain quantitative results from these queries.
5. Development of a tool to extract user defined transects and then obtain the hydrodynamic flux moving across them.
6. A system capable of balancing speed/rendering/analytics to optimise performance levels, including the achievement of interactive visualization speeds.
7. Ability to export publishable quality outputs.

5.2 Design and Initial Implementation

The process of design and implementation for the system was an evolving one, particularly as collaboration continued with ocean scientists to retain the user focus. As a result, initial ideas were superseded by later considerations and new tools were included. Thus, an agile software development approach was adopted: new tools and ideas were discussed with ocean scientists which were then included into the system. This section discusses the overall design process and early developmental phases. It includes:

- Selection of an appropriate programming tool for development of the system
- Early concept and design work

5.2.1 Selection of programming tool

VINCA is written in Java, using Processing as the renderer. This section discusses the factors which influenced that decision. The cautionary message of Clyne and Rast [62] has been at the forefront of the decision making process, together with the mantra that the system needs to be both usable and useful. Clyne and Rast argues that despite many of the claims made in favour of visualization as an indispensable part of knowledge discovery, the reality is somewhat different, because of the limited capabilities of the visualization software available, a view echoed by Bernholdt et al. [32].

Whilst there were attractions to using something like VTK or OpenDX – including a strong familiarity with their use for visualizing modelled estuarine data, the problems encountered with the earlier prototypes, including the speed/quality dichotomy proved a deterrent. Taking this into consideration, together with the complex and highly challenging nature of the data and the specific problems presented by modelled AMR data (which generic tools struggle to handle), and an understanding of the needs of the researchers, it was concluded a turnkey approach, based on an API was preferred.

The focus of the research is to prove the value of a visual analytics approach to enhance the analytical and knowledge discovery capabilities of, what, in visualization terms, is a little researched domain of study. The system therefore required a close coupling between the data, the visuals and the analysis, which, it was concluded, current visualization tools do not easily provide. Open source software was preferred to maximize availability for users within the ocean science domain

Thus, Processing was selected as the development tool for this thesis. Processing originated from the MIT Media Lab³. It offers a comprehensive graphics library which may be used for visualizing both 2D and 3D data. Processing builds on the graphical capabilities of JOGL (Java OpenGL Bindings⁴), and is capable of running on very large screens. Use of the OpenGL renderer makes it quick to render. Additionally, Processing has off-screen render capabilities, making it easy to create a pdf version of the output (necessary for publishable quality output). Finally, it

³www.processing.org

⁴<http://jogamp.org/jogl/www/>

has been used for coastal shelf visualization, for example the University of Washington's Applied Physics Lab system to visualize the dynamics of coastal marine ecosystems as a part of the NSF RISE project [22]. More recently, there is evidence of Processing's successful use for iterative visual analytics, albeit not in relation to geoscience data, and for significantly smaller datasets [204].

5.2.2 Early concept and design work

The overarching design was to follow Keim et al's. [166] knowledge discovery process of Analyse First; Show the Important; Zoom, Filter and Analyse Further; Details on Demand, which is effectively represented by Silva and Freire's [265] schematic as shown in Figure 5.8), itself an enhancement of Van Wijk's model [299].

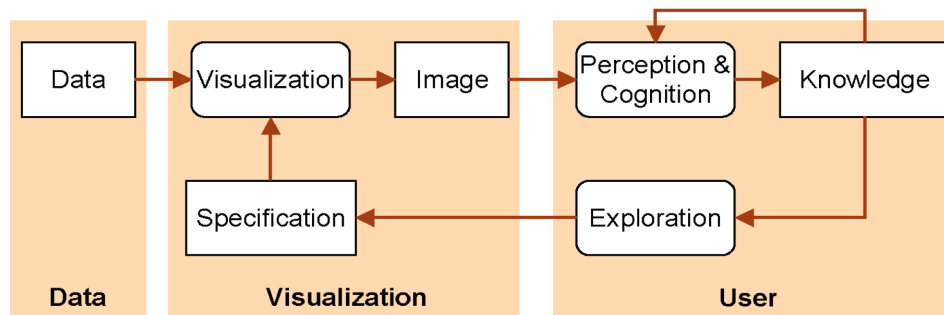


Figure 5.8: Reproduced from Silva and Freire's [265]; schematic of visual analytics and visualization pipeline, illustrates the basis upon which the pipeline for this system has been developed.

Doleisch et al's. [75] reflection that visualization of large, simulated, high dimensional datasets is one of the most challenging fields in scientific visualization, requiring the resolution of a number of complex problems, is certainly true of hydrodynamic, estuarine modelled data. Throughout the development of VINCA, the software engineering challenge has been to develop an interactive visualization and visual analytics system, which achieves a balance between functionality, speed, performance and quality balance in relation to very large, complex datasets.

Therefore the design of the system, also had to consider:

1. Manipulation and filtering of the data by a variety of means, whilst maintaining its characteristics to derive datasets which may then be examined and analysed, through interactive visual analytics techniques;
2. Methods of implementing query based visualization, achieved through the identification of features/areas of interest;

3. Linked coordinated multiple views facilitating data analysis through a suite of complementary views, plots and graphs focusing on attributes of the data, and designed to provide different perspectives, and reveal patterns and relationships.

Finally, a Model View Controller approach was taken in the development of the system, such as to separate the model (the data) and the analysis of the data, from the visual front-end. This enables multiple views of the data to be easily created, see Figure 5.20.

5.3 Data Import

The ASCII data consists of a file for each of the variables i.e. X position, Y position, Bottom, Water Elevation, Free Surface, Velocity U, Velocity V, Scalar Velocity and a couple of other associated meta files that explain the mesh structure and the boundary. The size of the data files is highly dependent on the number of points in the mesh and the number of timesteps. The ASCII case study data used ranged from 300MB in the case of a really low resolution Burry Mesh with 76 timesteps; to 600MB for a higher resolution Burry Mesh with 76 timesteps; to finally 2GB for a high resolution Dyfi Mesh with 120 timesteps.

The Binary Data used to test the data importer came to 1.1GB for a high resolution mesh of the Mawddach Estuary – these data-sets were just used for testing the importer and never actually visualized

Initially, the data import routine from the VTK prototype, written in Java, was re-used. This translated the TELEMAC data into an ASCII format for import into VINCA. However, this was not intended as a permanent solution, because it is inefficient. At a later stage, a TELEMAC specific data importer was written for the data, to make pre-processing calculations, thus speeding up the visualization and analysis phases. The importer also created a MySQL database containing the data for future investigation, which would, additionally, be used for storing session data. The session data includes the derived data, and also the current view hierarchy, so that after a break, the user is able to recommence investigation again at the same point, rather than having to undertake the discovery trail from the beginning every time the data-set is loaded.

The Data import was split into two distinct operations:

1. Data Import: reading, pre-processing and importing raw data files into a MySQL

database, and generating the XML loading file.

2. Database Loading: loading the database into working memory and pre-visualization processes.

Tasks were split to enable the data to be easily re-loaded from the database without having to return to the raw files and repeat the reading/pre-processing and importing stages. It also allowed for a central MySQL repository containing all of the visualized databases to be used. This central repository might be stored on a dedicated Database server, and by means of a high speed network connection, a thin client/visualization workstation could then be used for investigating the data. This would allow for multiple researchers to investigate the same data easily, and, furthermore, would mean that the data would only need to be imported into the data-base once.

5.3.1 Stage 1 – Data-Import

This stage read the raw data files, pre-processed the data, then subsequently populated the database and generated the XML loading file. TELEMAT-2D data comes in two main formats: Binary and ASCII and it was an aim to allow for both types to be imported natively. At the onset of the project, the ASCII dataset was the only one available for use, so initially just the ASCII importer was built. However, as the project progressed, a greater number of Binary datasets became available, and a Binary importer was developed to accommodate these datasets.

The ASCII data came in the form of multiple files each containing the data for a single variable, together with the mesh data file; a file providing the connections between the mesh file, and another file containing the boundary nodes. (The raw ASCII files are represented in the figure 5.9 as a filestack and the Binary as a single file)

After researching third party libraries capable of importing this type of data, it was discovered that TELEMAT and its sister products (TOMOWAC and SYSPHE etc) use an uncommon format, unique to them, but there were no libraries for importing the data type. A MATLAB script, and some Fortran routines were identified, but none of these were ideal as each went through time-consuming translations. For example, one Fortran code used a legacy VTK format, while another went through a custom ASCII format, so neither would have an advantage over the current prototype setup, and would, in fact, be similar to the original prototype. This led to a

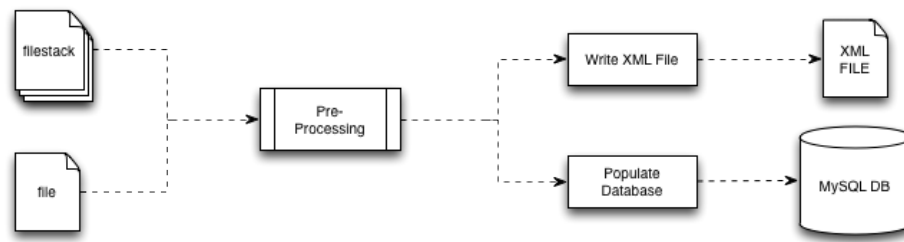


Figure 5.9: The schematic demonstrates the processes involved in the Data Importer Pipeline

decision to build the binary importer from scratch. Another consideration was to use a Java Native Interface to call the functions of one of the Fortran routines, but this created issues in its own right: it would be necessary to compile the Fortran routine for various different operating systems, which, in turn, would require different distributions of the Visual Analytics system.

So, the approach was to reverse-engineer the binary importer in native Java from the Fortran routines and MATLAB scripts, using the file structure from the TELEMAC users manual as a guide. This removed the need for other software and extra processing steps, thus improving the speed of the import process, which met the desired outcome to be achieved from preferring binary data over ACSCII, and would contribute to optimising the speed and performance of the system.

There are three stages to pre-processing the data: (1) reconstructing the mesh; (2) calculating other information relating to each cell in the mesh (such as its centroid (and all interpolated variables over all timesteps, and its area; (3) creating an edge table to hold information just about the edges of the mesh. Figure 5.9 indicates the positioning of pre-processing in the Data Importer Pipeline, but to avoid complicating the schematic, the three pre-processing stages are here described in detail:

Stage 1: Reconstruction of the mesh. The results file from TELEMAC-2D includes a file describing the connectivity of the mesh. Therefore, through reading this file along with the point-data, the mesh can be re-created in the system.

Stage 2: Calculating other information. To enable different views of the same data and allow various types of exploration, other information has to be calculated during pre-processing. This additional information supports cell based study of the data, achieved by using each of the individual cells in the

model as the data source, as opposed to point based study. Calculation of the supplementary information requires determining the centre point (centroid), and interpolating its associated variables over all timesteps.

Stage 3: Other metrics. Other metrics such as the area of each cell might also be calculated. The edge table was generated but never actually used. It comprises a list of the edges in the mesh, with no duplicates (e.g. the cells contain three edges).

A multi-table custom schema was developed for MySQL, which was populated with the pre-processed data. The schema consisted of:

Points Table: Holds the x and y position of the individual points, and a unique ID for each of the points.

Point Data Table: Contains the scalar and vector data for all the points held within the Points Table, for all timesteps.

Cells Table: Holds three PointID references which define the cell, a CellID, and the x and y position for each cell's centroid.

Cell Data Table: Contains the scalar and vector data for all the centroids held in the Cell Table for all timesteps.

Edge Table: Holds two PointID references which define the edge. Even though the system currently does not use the edge data, it was left within the data-base for the development of future tools as needed.

An XML file was also generated to hold the meta information for loading the MySQL database [database name, username etc] into the visualization environment. This XML file allows the user to load the database into the system without having to re-import the data, once the data has been imported into the MySQL database.

5.3.2 Stage 2 – Database loading

Following data import into the MySQL database, the data are loaded into the visualization environment immediately, read from the database and loaded into working memory, where a second preprocessing stage takes place, which includes the generation of the interaction quad-tree. However, the user may opt to reload the dataset

into the visualization environment at a later stage, for further investigation by using the XML file. This means the user does not then have to repeat the process of importing the data into the MySQL database. See Figure 5.10.

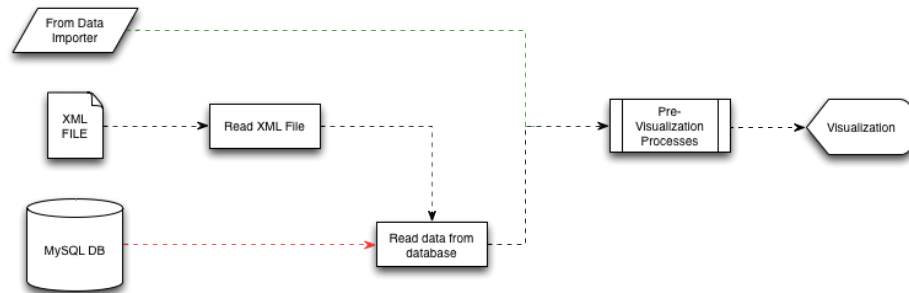


Figure 5.10: Data Loading Pipeline

5.4 Spatial indexing

At this point in describing the pipeline, it is appropriate to discuss the use of spatial indexing: it is applied a number of times throughout the system and is central to the development of a usable system meeting the researchers needs.

In particular, as identified in the Related Work (Chapter 4), to meet the requirements of the system, data structures need to be used that can create visualizations at interactive speeds, operate on large datasets, enable fast filtering (both data abstraction and data aggregation) and enable the user to identify subsets of interesting data points (or features within the data). In addition, the use of Coordinated Multiple Views and parallel coordinates were demonstrated to be particularly challenging for interactive visualization and put an extra burden on the chosen data structures.

A number of options for improving speed and performance were discussed in Chapter 4 including parallel programming, hardware, grid and cloud computing, caching techniques and pre-calculation or prediction algorithms to speed-up the calculations. Indeed, the optimum system would ideally incorporate solutions from all of these. However, while aspects such as GPU programming would be useful, this work has deliberately steered away from these solutions. It is unclear as to what the computing architectures of the end users would be, and, furthermore, one of the long term aims was to enable web-based operation. Therefore this thesis utilises a range of techniques. In addition, the techniques developed also aim to overcome other challenges such as overplotting.

5.4.1 Selection of spatial indexing

Overplotting is a severe problem on large datasets. The estuarine data analysis requires the simulation of extreme tidal/fluviat events and estuarine sediment transport up to 100 years from the present and the hydrodynamic models studying the tidal flow also enable features such as sandbanks, scour pits and tidal channels throughout the flood plain to be visualized (and how they change over time). Adaptive (non uniform) meshes, widely used in modelling the estuary to optimise computing efficiency, are used in the datasets for this dissertation. A typical data run is up to 50,000 points for each time-step and up to 100 time-steps per run. If all the time-steps were to be displayed at the same time, this would require 10^6 data points to be plotted, see Figure 5.11 a TELEMAC-2D finite element mesh for the Burry Inlet [242]. The resolution in the river channels is 25m, increasing to 50m in the outer estuary and 100m offshore.

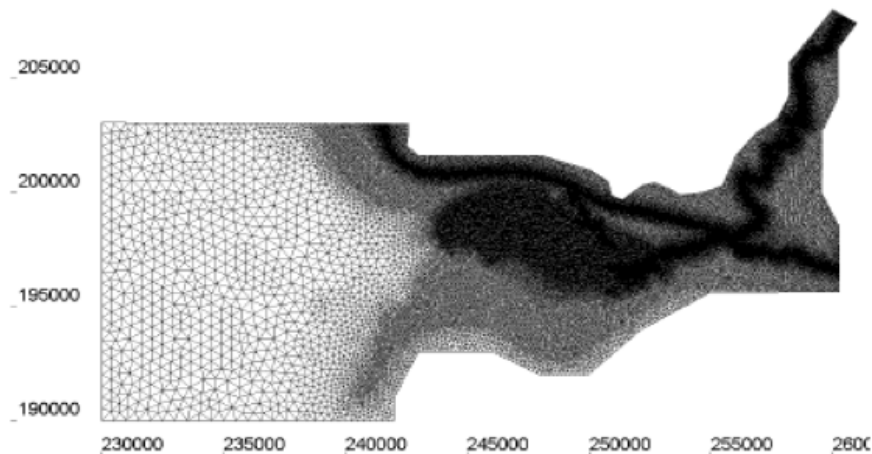


Figure 5.11: TELEMAC-2D finite element mesh for the Burry Inlet, reproduced from Robins [242] The resolution in the river channels (the darkest areas) is 25m, increasing to 50m in the outer estuary and 100m offshore.

If these data points are plotted on a typical screen resolution of 1280×800 , each of the points inside the estuary becomes overplotted: the amount of screen real estate available to each view in a Coordinated Multiple View is actually much less and, as such, the points inside the estuary becomes overplotted – in fact, a typical estuary has about 50 points per pixel. The overplotted data hinders the achievement of interactive visualization speeds and also selection of underlying data points, as there are too many data points on the same pixel, making selection challenging.

The system requires data to screen space conversion metrics to be calculated at several points in the visualization pipeline:

- at the pre-visualization stage to associate the regular screen space to the adaptive, irregular data space and provide access to all of the data;
- to support data selection and speed up queries, thus enabling meaningful, interactive visualization;
- to resolve overplotting of the dynamic vector flow data;
- in conjunction with the parallel coordinates to aggregate the data to reduce spatial clutter and support detailed examination of the data;

Proposed solutions to reduce the number of points tend to fall into two categories: visual space and data space, and the decision to adopt spatial indexing, a hierarchical data space solution, was based on a number of factors. The Related Work (Chapter 4) indicated their attributes would meet the needs of this research, including the specific problems presented by the data. Furthermore, during prototyping, experimentation with spatial indexing algorithms demonstrated them to be efficient, effective in facilitating search and selection, easy to use, fast, visually pleasing, and they easily resolved the overplotting problem.

Non spatial indexing solutions for dealing with the specific challenges of visualizing modelled, unstructured, multi-resolution data were also discounted for other reasons. The desire was to develop a simple, elegant system, which would not confuse the users. Indeed Samet [252] argues that the advantage of hierarchical data structures lies in their conceptual clarity and ease of manipulation. Furthermore, many of the non spatial indexing solutions do not take into account the query driven nature of visual analytics.

However, the primary reason for the selection of spatial indexing was its documented applicability to large volumes of unstructured, multi-resolution modelled data produced by Adaptive Mesh Resolution techniques – the type of data associated with this research. These hierarchical data types present a significant challenge to visualization, documented in Chapters 2 and 4. They refine the domain space of a simulation into a hierarchy of nested, sequentially refined grids, both spatially and temporally, such that each timestep consists of multiple levels of grid cell refinement. Additionally, modelling the data to a uniform grid is sometimes not a helpful solution as this method may introduce grid artefacts not present in the original data, and, furthermore, may increase storage requirements [234].

There is a vast array of spatial indexing algorithms, but it was decided to focus on two of these, both binary space partitioning trees: the PR quadtree and related k-d tree. Both of these have been proved to be effective for arbitrarily distributed data [252] because of their regular decomposition approach, which is regarded as more flexible than many other indexing methods. The related work (Chapter 4.3) illustrated the value of spatial indexing (including both quadtrees and k-d trees) for the data types and evidence of the use of spatial indexing algorithms in geosciences, also for deep ocean, coastal shelf and estuarine applications, for which its recursive nature is particularly appropriate.

The decision to use the **PR (point region)** variant of the quadtree was made because of its acknowledged value for modelling fields, and its usefulness for the multi-resolution data associated with adaptive mesh refinement modelled data. [251]. Similarly, k-d trees, were also seen to be useful for unstructured, multi-resolution grids.

After considering the attributes of spatial indexing algorithms, as part of the early development work for this system, it was decided to develop both PR quadtree and k-d tree algorithms, to identify strengths and weaknesses in their use for this data, and for testing by ocean scientists.

5.4.2 Development and testing of PR quadtree and k-d tree approaches

Prototypes were developed using both approaches to compare their effectiveness in resolving the overplotting problem and facilitating data point selection, analysis and further visualization. By associating the spatial data structure to screen space, the spatial indexing algorithm allows the user to interrogate the data structure visually and thus have quick access to all the underlying data, (see Figure 5.12, which demonstrates a representation of the PR quadtree).

Key criteria were established for reviewing the suitability of both spatial indexing algorithms for a visual analytics system. These included the standard metrics such as build time and tree size, but also metrics for screen space size distribution and aspect ratio distribution.

The algorithms were tested using datasets of the Burry Estuary, which contained 26693 points and 51186 points respectively, and were provided by the Centre for Applied Marine Science, Bangor University. A stochastically generated dataset was used to provide a uniform/control dataset. All of the testing was performed on a

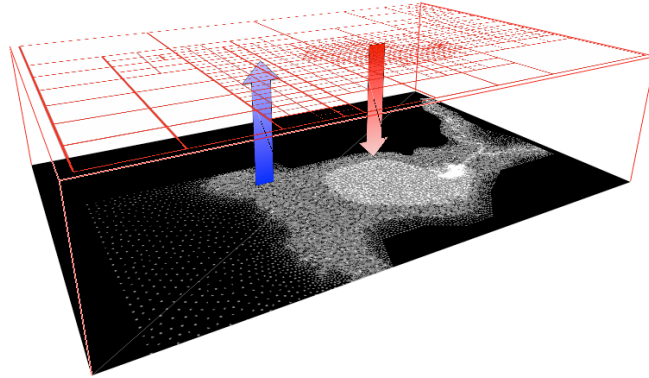


Figure 5.12: PR Quadtree applied to Dyfi Estuary TELEMAC-2D dataset

2.53GHz Dual Core Mac Mini with 4GB RAM.

Using the standard metrics as a basis, Figure 5.13 demonstrates that the PR quadtree was quicker to build in both the real and uniform data cases, and proved to scale better in the real data case. The k-d tree is slower as it has to sort the data on each recursive step. Figure 5.13 also demonstrates that both the PR quadtree and k-d tree (in the smaller test case) proved quick enough to be used at interactive frame rates (10fps+), but the PR quadtree's superior scaling would provide a better solution for generating spatial indexing on the fly for larger data-sets.

As the amount of available RAM increases, memory requirements become less of a concern but assessing the number of leaf nodes gives a rough estimate of the speed of interrogation of the data structure, without performing detailed access speed analysis. It can be noted from Figure 5.14 that the PR quadtree increases significantly in size between the two test cases, whereas the k-d tree remains the same size. It was also seen from the uniform datasets, that the k-d tree requires significantly fewer leaf nodes than a PR quadtree. This implies that the k-d tree will be quicker to interrogate than the PR quadtree. The k-d tree also proves to be more space efficient in that all of the leaf nodes contain some data points, whereas with the PR quadtree, some of the leaf nodes are empty, see Figure 5.16.

However, the aspect ratio of the k-d tree varied greatly, and increasing the number of nodes causes fewer of the nodes to fall within the regular shape required for an intuitive, user friendly layout. Figure 5.15, illustrates this. The dark band denotes the aspect ratio which provides regular shaped nodes.

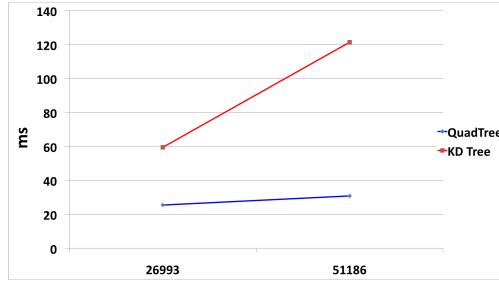


Figure 5.13: Build times in milli secs for PR Quadtree and k-d trees with 26693 and 51183 points..

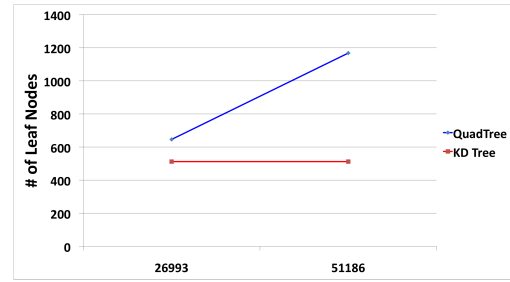


Figure 5.14: Number of leaf nodes for PR quadtree and k-d trees with 26693 and 51183 points.

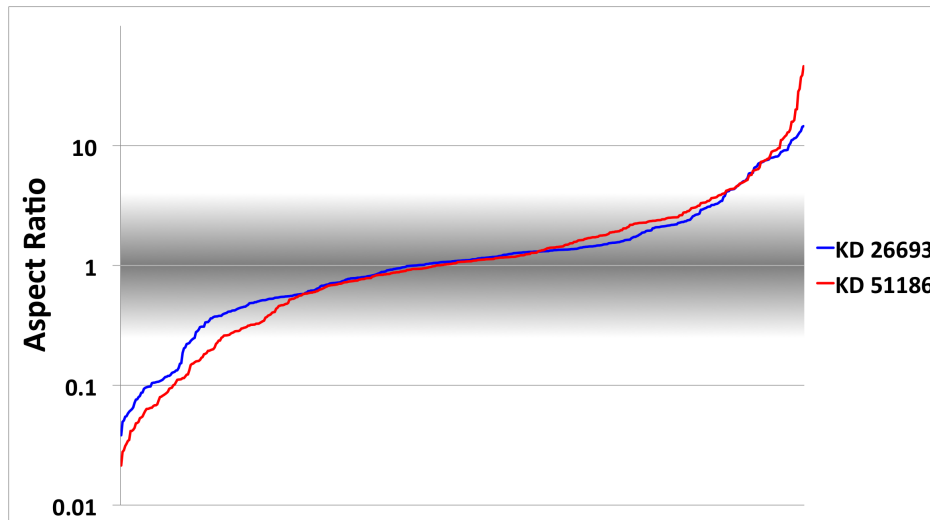


Figure 5.15: Distribution of leaf node aspect ratios for k-d trees with 26693 points and 51182 points: the dark band, which denotes aspect ratios providing regular shaped nodes, encompasses fewer nodes in the larger dataset.

Even though there were aspects of performance that were better with the k-d tree, balancing this against the fact that the aspect ratio for the PR quadtree stayed constant, also considering the visual layout of both spatial indexing methods, (Figure 5.16), it is apparent that the PR quadtree gives a neater, more visually appealing result, which lends itself to better interaction and a more understandable visualization, if the grid is visualized.

Although not as space efficient as the k-d tree and having slower access times, the PR Quadtree still achieved fast build times, and produced intuitive, neat layouts which lent themselves to a scalable, interactive visual analytics framework. Finally, given the fact that both algorithms have benefits and disadvantages, the balance further tipped in favour of the PR quadtree, as it was the approach preferred by the users, who found the PR quadtree clearer and easier to understand, when demon-

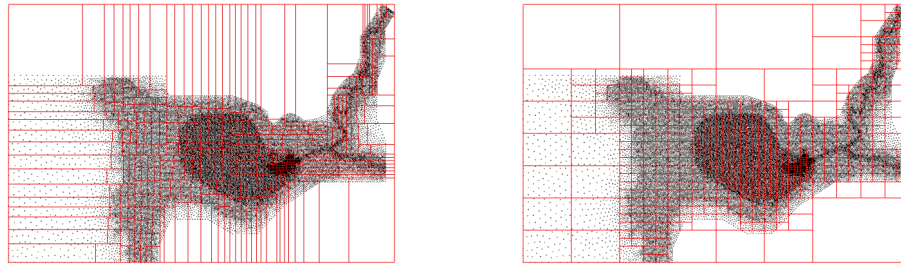


Figure 5.16: PR Quadtree (right) and k-d tree (left) for 26693 points.

strated at an early stage of development (a conclusion borne out by the analysis of leaf node aspect ratio distribution).

5.4.3 Uses of PR quadtree algorithm

Having established a preference for a PR quadtree algorithm, it was decided to use it in a number of ways throughout the system, and for this two variants were developed: the first, a static PR quadtree, with a bucket capacity of 100 points for filtering raw data to resolve the data to screen space association problem and for data selection. The second version, a dynamic PR quadtree, is used to resolve the problem of massively overplotted vector flow data (by supporting multiple level of detail) and also to create a spatially aggregated PCP. The various uses of the PR Quadtree for data selection, and in its dynamic variant, are discussed later in this chapter, when the system's views and analytic tools are considered in the context of interactivity and interfaces.

5.5 Interaction and Interfaces

Interaction and interfaces are discussed throughout the thesis, in relation to specific aspects of VINCA. This section provides a context to this and also focuses on some of the more important elements. Pike et al. [226] makes a distinction between low level user/interface interaction (where the aim is to manipulate data representations to identify correlations, patterns etc.) and high level user/information space interaction (to generate understanding). This is presented schematically in Figure 5.17. Thus, whilst the challenges inherent within the data have governed many elements of developing the system and tools, the focus in regard to interfaces and means of interacting with them has been upon the user. Intuitive usability was a key consider-

ation, especially as users had little previous experience of working with high performance visualization and visual analytics systems. This required a system where it proved easy to interact with representations of the data, thus inviting further investigation, suggesting inter-relationships and hypotheses and stimulating reasoning.

There are several important aspects of interaction that are included in VINCA. Brushing and highlighting are two features that are used predominantly in a multiple-view system. Selection is also important, which enables the user to extract different elements from the dataset and see how they correspond in other views. The system allows different forms of selection and filtering. For example, the user can determine the ranges of the selection points along an axis of a Parallel Coordinate Plot, and can also select ranges over the line graphs. Specific flux points can be placed on the screen etc. These are discussed further as they are used within the system.

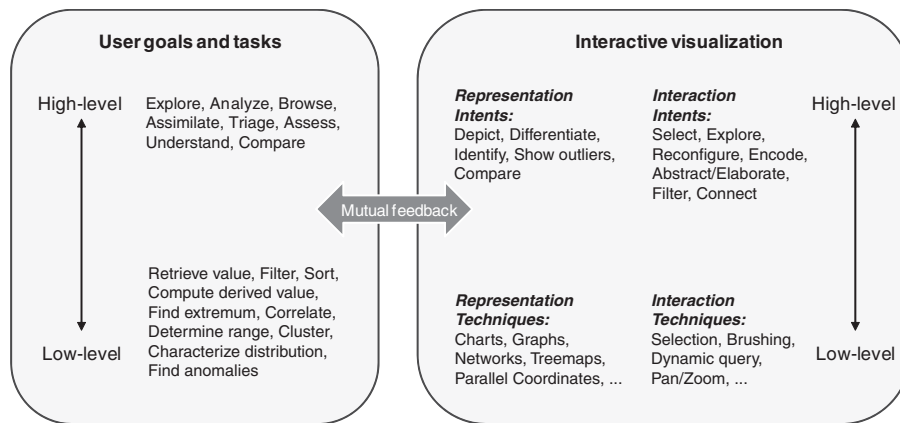


Figure 5.17: Reproduced from Pike et al. [226], schematic illustrating the user goals and tasks facilitated by interactive visual analytics.

5.5.1 GUI Elements

Before considering data selection methods and the various views, the construction of the GUI elements should be mentioned. In general, direct querying methods through the use of the mouse were preferred as these tend to be more intuitive and provide greater interactivity with the data. Basic GUI elements (sliders and buttons) and their interaction were incorporated into VINCA using Andreas Schlegel's ControlP5⁵ GUI and controller library for Processing. This library provides controllers such as sliders, buttons, toggles, knobs, textfields, radio buttons, checkboxes etc. to

⁵<http://www.sojamo.de/libraries/controlP5/>

be added in separate control windows for organisation in tabs or groups. All basic user interface elements can be easily plugged into specific panels and can be drawn directly to the main window without the need for a separate control panel.

5.5.2 Data Selection within the views

The challenges of the data cannot be ignored here: query driven visualization demands interactive visualization speeds, which are inhibited by overplotting. Furthermore, the size and complexity of the datasets require techniques such as rapid zooming, to enable users to quickly identify and select areas of interest. Thus, spatial indexing, using the PR quadtree, is an essential element supporting data selection for study and analysis in the linked views and graphs presented in the CMV interface. The PR quadtree provides a clean method of spatially indexing the massively overplotted data, especially visually: all the nodes have the same aspect ratio as the overall bounding box of the data set. This provides control of the number of data points required to be searched when users select data. Testing demonstrated that a maximum of 100 points per PR quadtree node proved to most effective, achieving the optimum speed versus number of searches to find a node within the dataset.

A methodology for displaying the points in the PR quadtree cell to users, when they interact with the system, was required. Initial thoughts were based on using lists. These included a simple list (see Figure 5.18); a list based on a spiral using the distance from the mouse click, also a radial-based display, again using the distance as a metric, but incorporating the direction metric from the mouse click.

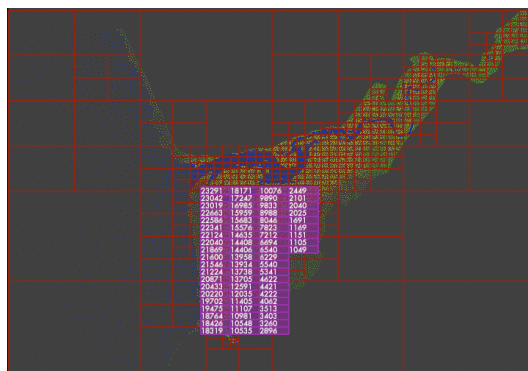


Figure 5.18: Simple List visualization of points contained within leaf node of spatial data structure.

The simple list, Figure 5.18, presents a drop-down menu incorporating all of

the data points contained within the node selected by the user. However, it does not give any spatial indication of where the data points are in the node. A second style of list based on a spiral was developed. The idea was to provide a visual clue of the distance of the data point from where the user clicked, but it did not resolve overplotting, or enable the location of the data to be assessed correctly. This required alternative solutions, which included the Data Zoom Function.

Data Zoom Function

The related work identified zooming as an extremely useful tool for enabling users to rapidly change the viewpoint of the data to support better understanding of interrelationships. Zooming in provides detail, whereas zooming out gives context and the zooming tool enables the user to quickly zoom in and out of areas of interest. Based on the PR quadtree selection, and treating each of the leaf nodes of the PR quadtree as a window to zoom into, this function overcomes the massive overplotting, which would prevent the zooming tool from providing any meaningful examination of the data, and gives an easy method of selecting underlying data. The quadtree came into its own, as it has the same aspect ratio as the bounding box of the data; it is visually appealing and fits neatly.

There are two modes of operation. Initially, as the user scrolls over the main view of the data with the mouse, it partly zooms into the current PR Quadtree node, where the mouse is located. This not only provides a less overplotted view of the data, but also gives a clear visual reference to the users search location within the data. The user selects data by clicking on the PR Quadtree cell to reveal a view of that node alone. This allows for a clear view with no overplotting, permitting the user to select the underlying data easily. To provide the user with a spatial reference of this node's location in the context of the whole dataset, within the zoomed-in view, a small floating thumbnail graphic is placed over the view, see Figure 5.19.

5.6 Coordinated Multiple Views – CMV

With such complex data, single views are not able to provide insights into all its attributes of the data: for example, the related work discussed the difficulty of analyzing spatio temporal data using a single visualization technique, thus analysis requires multiple complementary techniques and views and attribute plots and graphs. With the use of CMV for this project, it is hoped to provide a significantly richer

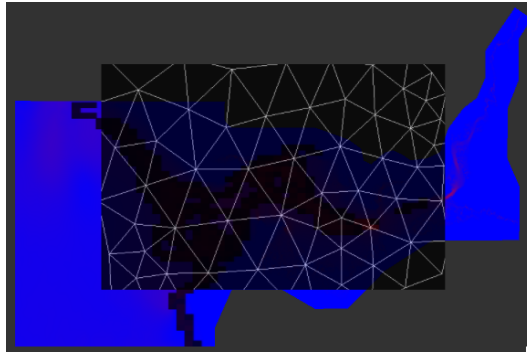


Figure 5.19: zoomed in view of PR Quadtree cell, showing properties of the cell.

analysis environment, and to encourage a shift away from the study of juxtaposed static plots, which researchers in a number of areas of the geosciences increasingly regard as problematical, ineffective and prone to change blindness [126, 138, 208]. The user interface will include a linked CMV environment, where complex queries may be undertaken, through a diverse range of views, plots and graphs.

Users are able to easily add a new view, and their actions are linked between views. The estuary may be viewed holistically on a map (with colour representing the scalar, speed) and then specific transects along the estuary may be selected and displayed in a linked view. Animation is linked between views, so that as the animation plays, the other views reflect the dynamic changes to the data. Values from one view or graph may be selected from one window and shown in another graph. Finally, iterative, oceanographic analytic tools are provided including the transect profiler and flux calculator.

5.6.1 Developing the windows for CMV

As the Processing library does not natively support and manage multiple windows, a bespoke window management system was built.

The manager controls the position and scale of the windows; how they open and close and how the data is passed between the windows. This bespoke system provides several advantages: it enables the use of large display screens, and it is possible to annotate lines and other information on the top of any view to show the relationships between data within different windows. A Model View Controller pattern, (Figure 5.20) was first developed to enable the multiple views to be readily created and linked together, and to manage the windows.

On this was built a hierarchical management structure, which is discussed in

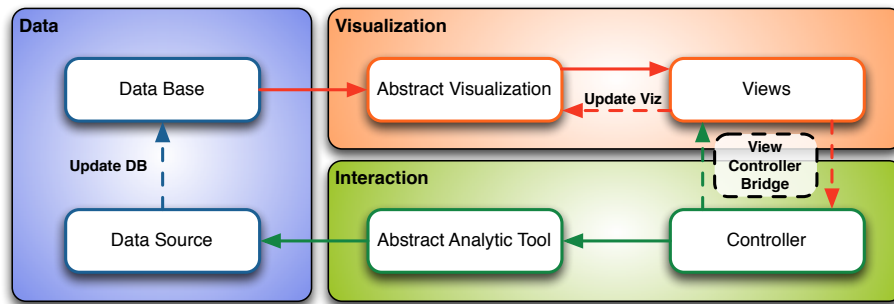


Figure 5.20: Model-View-Controller type window manager used for VINCA.

detail in the next section. The root node of the graph is the first window, which, in the case of this project, is usually a colour plot of the estuary. This model allows sub-views to be linked to other views, such that the sub-views depend on other views. For instance, users may require a graph of a transect: they select a position on one window and the profile of that transect appears in a new window. The user may then delete the new window, or select information from it to display in another window. The system gives users the option to depict these dependencies by lines. However, it is recognised that these might quickly become cluttered, so lines are only added when they are required. The Model-View-Controller window manager is thus transformed into a Hierarchical Model View pattern [51], with information being passed from controller to controller to communicate and update the linked information.

5.6.2 Development of CMV management tool: Hierarchical CMV Manager.

Having decided on a CMV approach, unless a coherent management structure for linking the views is developed, a configuration of views and sub-views will evolve, which will not necessarily represent the most efficient or effective manner of analysing data. Thus, there is a need for a management structure. There are numerous methods for linking and managing configuration of the views, details of which may be found in Roberts [235, 236] and Erbacher and Frincke [86]. The most straightforward method is to use a simple list, and this was investigated to give an understanding of the potential problems of developing a linking and management tool for the CMV.

Simple List Window-Manager

A view is selected by the user and placed at the head of a list, which contains either a list of views or a list of references to the views, see Figure 5.21.

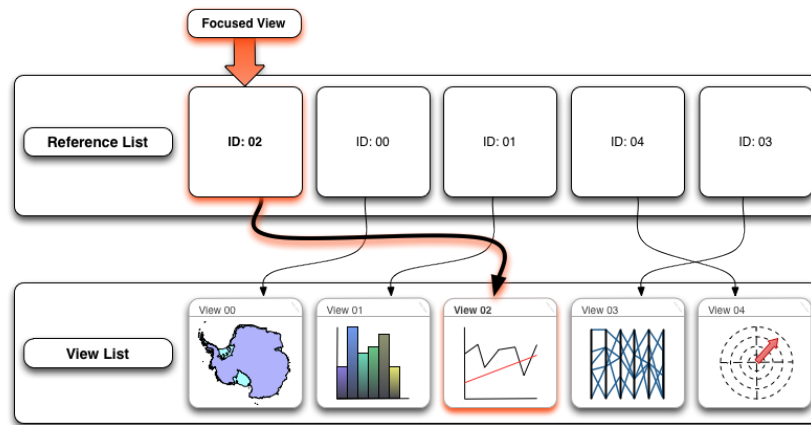


Figure 5.21: Example of simple list management tool.

Whilst it worked well for a simple visual analytics system, where there are a limited number of views and when the coordination between the views is relatively basic, it is easy to see that a proliferation of views, and a large, highly complex dataset will soon result in a confusing system, which becomes time consuming to navigate and explore, and fails to provide the desired insight, as the links are not meaningful. The problem is further exacerbated when views are not coordinated to all other views, but to only specific views.

Research led to consideration of concepts such as North and Schneidermann's [210] Snap-Together Visualization (which provides a more finely controlled version of sibling window linkages) and Erbacher and Frincke's [86] hierarchical model. This builds on North and Schneiderman and provides for any number of nodes at each of five levels, progressing from overview, to visual facilitation, probing/zooming and then to a perceptual layer. As this model seemed to fit well with Pike et al. [226] approach to user goals and tasks for visual analytics (a progression from low level to high level), illustrated in Figure 5.17, it was decided to adopt a similar hierarchical approach.

Hierarchical CMV Window-Manager

Whatever the selected coordination and linking solution, a CMV manager was required. At the commencement of the project, Processing did not provide an off

the shelf solution, thus one had to be developed. The approach allows for views, together with their branch of the hierarchy, to be focused easily. The value of the approach is that it enables the user, when focussing on a particular view, to place it in context. Thus, if a high level view is focused, then all of its detailed views are also focused, and this subsequently works in reverse. If a detailed view is selected then the context, (the parent of this view) is also focused. Figure 5.22 illustrates the approach.

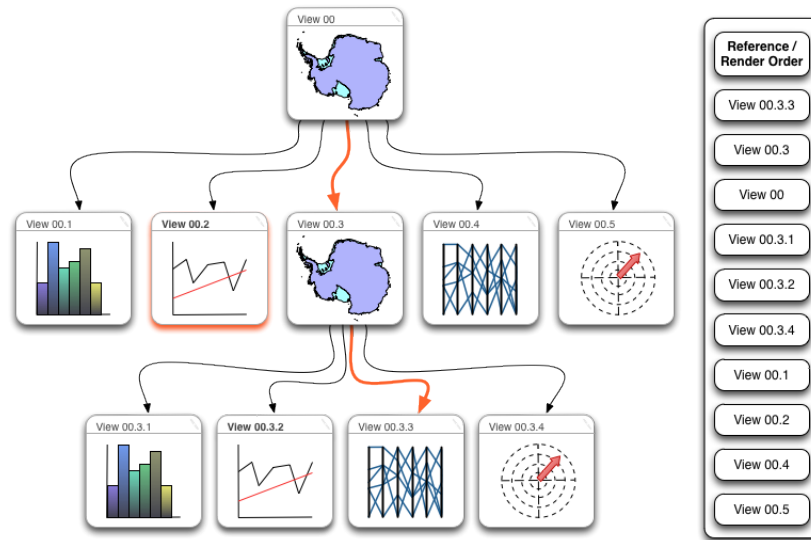


Figure 5.22: Outline of the Hierarchical CMV Manger.

To achieve this, a tree data structure is used to store the references to the individual views. This required the development of a generic view, which all other specific views might be built from, plus a tree data structure to hold and manage the interaction for the views. The hierarchical structure of the CMV manager brings greater cohesiveness and order to the CMV, such that meaningful linkages enable more intuitive and rapid searching. Additionally, views are easier to generate and remove, again providing for more efficient use.

5.7 Views

A range of complementary, but different, views and plots are required to enable the visualization system to meet the research requirements relating to the estuarine hydrodynamics i.e. understanding and predicting flooding. The ultimate goal for a visual analytics system such as this, would be to support inter comparison and

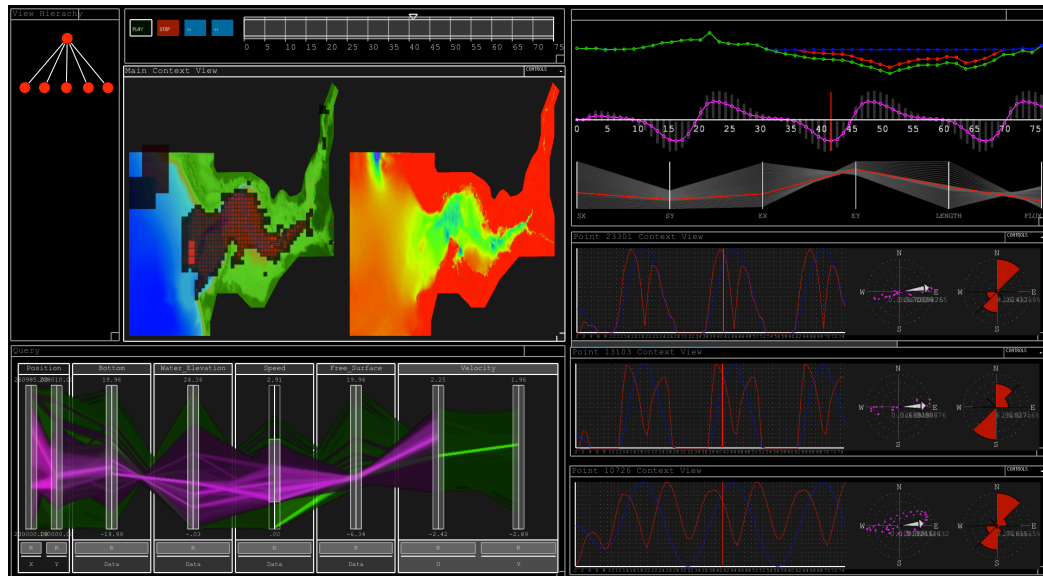


Figure 5.23: VINCA

analysis of a number of runs of the simulation with differing variables, or else to compare the output of different models. The plan is to add more views. However, the aim of this research has been proof of concept, thus the views for the system are focused on a requirement to compare and analyse simulated data from a single run to establish its usability and usefulness before considering its extension.

View Type	Views
Control View	Hierarchy Management, Timeline, Colour-Mapper
Context View	Main Context, 3D Context (Future)
Hybrid Context / Attribute View	Sub Context, Region Transect/Flux
Attribute View	Point Graph, Single Transect/Flux

Figure 5.24: Table of views within VINCA.

Figure 5.24 shows the current views available within the system. At its current state of evolution, the visual analytics system for this project includes:

- the control views, which are used to manage the system and include the CMV hierarchy management view; the timeline and the colour mapper.
- 2D/2.5D views including the main visualization view; data overview; dual

view.

- 1D and 2D attribute plots including PCP (with a number of sub-views); point graph view; 1D partition; temporal polar vector partition; polar vector rose partition; transect/flux view.
- Overlays including heat map overlay and vector overlay.

5.7.1 (Object Abstraction) Abstract View Class

The generic view (see Figure 5.25) is the foundation for all the other views, which are built from it, and contains the parameters and methods required by all the views. Figure 5.25 illustrates how these parameters are related to the visualization environment. This permits a developer to create different types of views, and easily extend the Coordinated Multiple View management system. It also allows for partitioning the Plot Space to create multiple partitions (if required). This enables ease of access and interaction from a single view, with different visualizations of the same data, which are held together. It was decided to group visualizations of the same data together, rather than separating them out into their own views, as a method of reducing visual clutter, because too many floating views can add to confusion. It also enabled the size of the tree map to be decreased, such that quick interaction is retained, as deep levels of investigation might cause an explosion in number of views. The views developed to date are outlined below, together with their features.



Figure 5.25: Schematic of a generic View

5.7.2 Control Views

The control views give the user access to the system controls, they consist of: (1) Hierarchy Management View (HMF), (2) Timeline View, and (3) the Colour Mapper View.

The HMF and the Timeline View are visible constantly and the Colour Mapper View pops up on request by the user. See figure 5.23.

The HMF has two main functions:

1. It provides a graphical representation of the CMV hierarchy, which allows users to see the discovery trail they following.
2. It acts as the central control panel for the system by allowing the users to:
 1. switch views on and off,
 2. delete views,
 3. close the current data-set and return to the startup menu
 4. exit the system.

The Timeline View provides the playback functions, which include, Start/Pause, Stop, Step Forward and Step Backward. A graphical timeline is also drawn, which doubles as a range slider allowing the user to easily select a specific time period to investigate. See figure 5.23.

The Colour Mapper View provides an interface for the user to generate and edit the colour-maps used within the program. It has three modes of operation:

1. Colour-Picker Model
2. Colour-Mapper Model
3. Internal Colour-Picker Mode

Depending on the context at which the colour-mapper is launched it will choose the appropriate mode and display the desired GUI. The Colour Picker Mode launches the Colour Mapper View into a basic colour picker, where the user is presented with a Hue / Saturation colour wheel and two sliders to set the Brightness and the Alpha component.

The ColourMapper Mode provides the user with a set of tools for generating and editing variable or component specific colour-maps (e.g. heat-maps). The tool

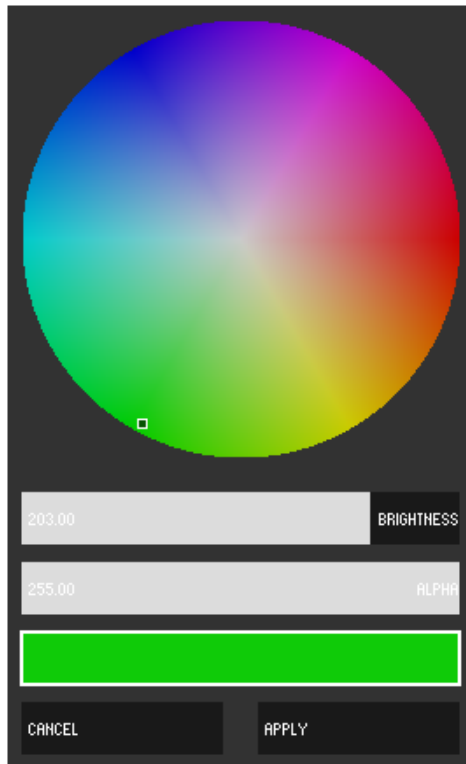


Figure 5.26: Colour Mapper in Simple Mode

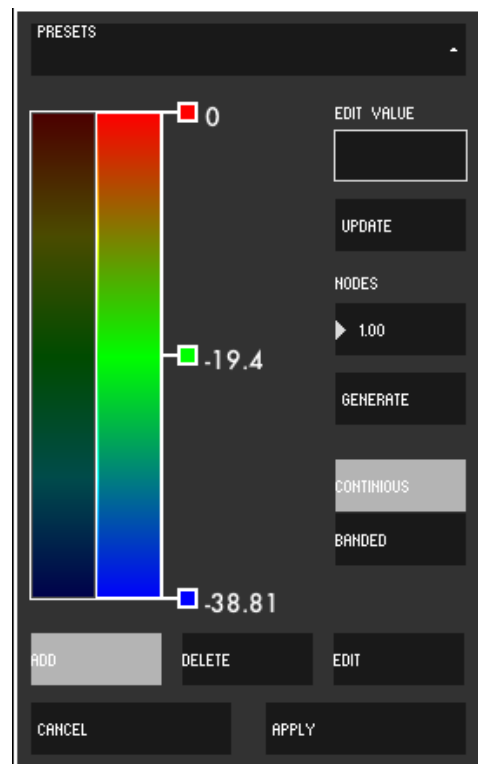


Figure 5.27: Colour Mapper in CM mode.

is launched in this mode when the user clicks on a variables colour-map button or a heat-map button in the Main Context View.

Figure 5.27 shows a double colour bar on the left hand side, the left part represents an HSVA colour map and the one on the right depicts a RGBA colour map, to provide users with a choice of colouring the data. The user can select the style of map by right-clicking on the colour bars, which will grey out the unselected option. The user can add colour handles to the map, and change the colour by right-clicking them. This launches the internal colour picker (which is the same as the simple colour picker, but when a colour is selected, it returns to the colour map editor rather than shutting the window).

The buttons on the right of the control panel in Figure 5.27 give the user: (i) more precise control of the values of the colour handles; (ii) a method of generating a standard Rainbow colour-map with a set number of divisions; (iii) a method of switching between a continuous colour map and a banded or discrete colour-map.

Once a colour-map is selected and applied, it is then added to a list of preset colour-maps, which may be reused with other variables, by scaling into their data

range.

5.7.3 Context Views

The context views provide the overview of the data whilst allowing users to interact and create attribute plots.

Main View

This is the main overview for the system (see figure 5.28), and all other detailed views are derived from it. This allows users to view and interact with the data holistically, and provides the overarching context for feature selection. The view may be studied at various levels of detail and provides the mechanisms allowing users to study the underlying data contained within the database, permitting switching between different visualizations of the underlying data and selection of techniques which will best support investigation.

It permits:

- switching between each of the variables held within the data base, for a list of the variables see section 5.3.
- editing colour maps of the data with the colour mapper window see section 5.7.4.
- different types of rendering – points, lines etc.
- various types of interaction with the underlying data including methods which deal with overplotting, such that all data can be accessed even on a small screen.
- development of a fast draw library, originally developed for the data view but was used elsewhere.

Dual View

The dual view (Figure 5.29) is a second view placed next to the main view to allow for side by side comparisons, and to enable the user to see the context view in multiple manners. A side by side view, that can easily be switched on and off, reduces the total number of views required to be studied by the user, and thus minimises

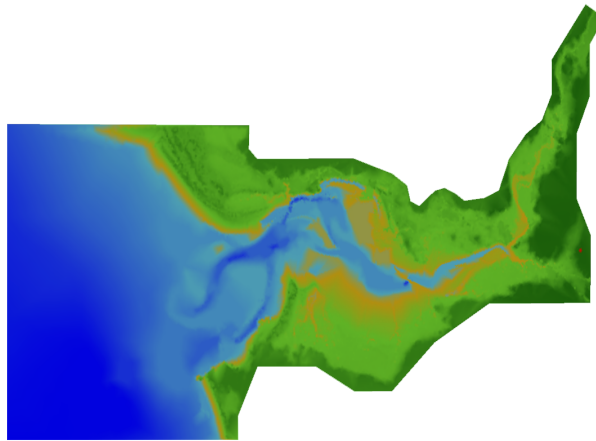


Figure 5.28: Main view showing the bathymetry of the Burry data set

spatial clutter. At the same time it enables multiple versions of a view to be easily visualized in different ways, thus providing different insight into the data. The dual view supports the same type of interaction as the main view, allowing the user to seamlessly interact with either partition.

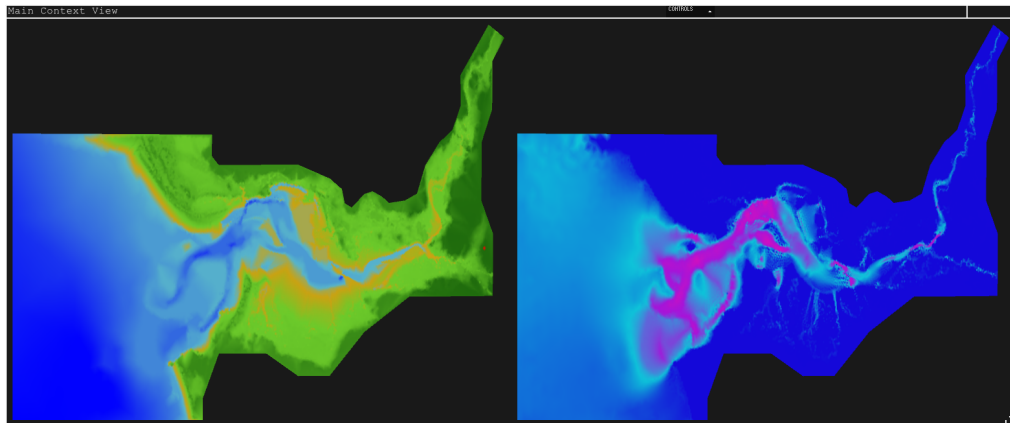


Figure 5.29: Dual view to allow side by side comparisons, the partition on the left depicts the bathymetry, the partition on the right depicts the speed of flow.

Sub View

The sub view is derived from the main view, and, by zooming, the user is able to obtain a subset of the main overview, from which a region of interest can be selected, which then provides the context for the user to undertake detailed study.

When the sub view is taken, a new spatial index is built for it providing access to finer and finer levels of detail within the analysis.

5.7.4 Overlays

Two different overlays were developed to support specific aspects of analysis. These included a heat map overlay for use with the parallel coordinates plots and the vector overlay. Their development is summarised here, but how they are used is covered in the relevant view discussion.

Heat Map Overlay

PCPs often suffer from occlusion (the result of cluttering) so a heat map overlay was used to demonstrate the results from the PCP query view. However, a degree of caution is necessary in that learning to read and understand heat maps may present a challenge, and care is needed in selecting hue and intensity, as colour is not perceived quantitatively. Thus, techniques such as gradients of colour intensity or hues, or diverging scales are utilised.

The type of heat map selected was dependent on the type of query being undertaken, and variants used include raw data; a static bin-sized spatially indexed heat map (see figure 5.30), and a variable bin-size spatially aggregated heat map. Each of these is detailed in greater detail in chapter 6. The colourmap of the heat map can be edited through the Colourmap editor.

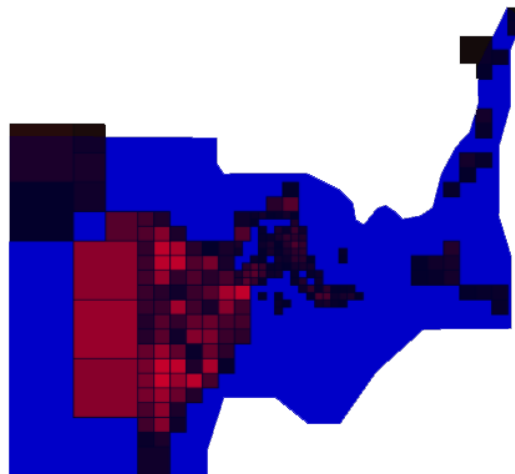


Figure 5.30: Example of the spatially indexed heat map overlay used within the system.

Vector visualization overlays

Flow vector glyphs are often used in ocean sciences. Therefore to provide context, and to support identification of features of interest, vector fields have been included. However, there are many different styles, and this is a much researched area, which is beyond the scope of this research. Time would not have permitted the development of an advanced mapping system for these studies, a major project in its own right. Thus, it was decided for the current version of this system, to adopt a simple arrow glyph approach to display the vector information, whilst recognising it was far from an ideal solution.

Dynamic Vector Filtering

The size of the modelled datasets, together with their spatio-temporal, multi-variate nature makes producing meaningful vector visualization problematical, and the density of the meshes in the estuary and tidal channels in this project resulted in massively overplotted arrows when visualized. Figure 5.31, view (a), shows a visualization of unfiltered vector arrows, which are so dense as to be virtually meaningless.

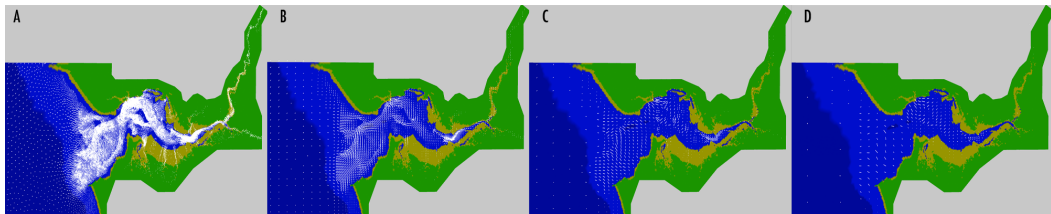


Figure 5.31: Output from the dynamic vector filter using the quadtree. View a shows no filtering, then progressing through b and c to heavy filtering at view d.

Rather than undertaking a detailed study of optimum methods of visualizing the vectors, as spatial indexing was already a key element of the system, it was decided to extend the use of the PR-quadtree to filter the vector arrows, because of its quick build nature, its ability to retain the spatial context of the data, and its ability to operate dynamically at interactive frame rates. The approach was similar to that adopted by Yu et al. [325] who used a basic quadtree to resolve the problem of overplotted data near the ground surface, when rendering the vector fields of earthquake simulations. The surface vector field was to be visualized using LIC, and there was a need to simplify the LIC calculations, and the quadtree was used to derive a 2D regular grid from the unstructured 3D data.

Thus, in this case, a separate dynamic quadtree was developed to aggregate

the data, whereby the user could determine the required level of detail for study through specifying the bucket size of the quadtree by using a slider. This process is illustrated in figure 5.31 which from left to right shows increasingly heavy filtering. The bucket size determined the maximum number of data points the quadtree leaf node might contain. Then, the vector was filtered by averaging all the data contained within the bucket.

5.7.5 Attribute plots and graphs

The adoption of information visualization tools and techniques, including statistical techniques and attribute views such as function graphs, scatterplots, histograms, polar graphs add significantly to the functionality of the visual analytics system, in enabling relationships between variates of the data to be studied. This visual analytics system includes a number of these, the most significant being the Parallel Coordinate Plots. However scatter and polar graphs are also used in a number of different forms.

A weakness of PCPs is their ability to encode direction [230,304], but one of the requirements of the visual analytics system is to support basic calculation of tidal ellipses (a way of illustrating how tidal currents vary, over time). This is undertaken by plotting tidal current vectors with their tails at a single point, as a function of time. Over the tidal cycle, the vectors typically trace a tidal ellipse. Polar graphs are widely used in weather studies as a representation for vectors, but they are also relevant for studies of hydrodynamic vectors. The traditional compass rose is a widely understood means of conveying direction, which users are able to interpret intuitively. In the case of this research, two uses for polar graphs in relation to vectors data were identified (the frequency of vector data, and tidal ellipse). But rather than adopting a similar approach to Qu et al. [230], where a single polar graph was integrated into the PCP, two separate polar graphs were developed to provide understanding of different attributes of the vector data.

Point Graph View

The point graph view (see Figure 5.32) allows the user to see the data from an individual point within the data set, and three different views are provided:

- A 1D graph partition represents the data temporally

- A temporal polar vector partition provides a vector trail, which leads to an understanding of the tidal ellipse
- A polar vector rose partition gives a frequency based view of the vector data held at this point.

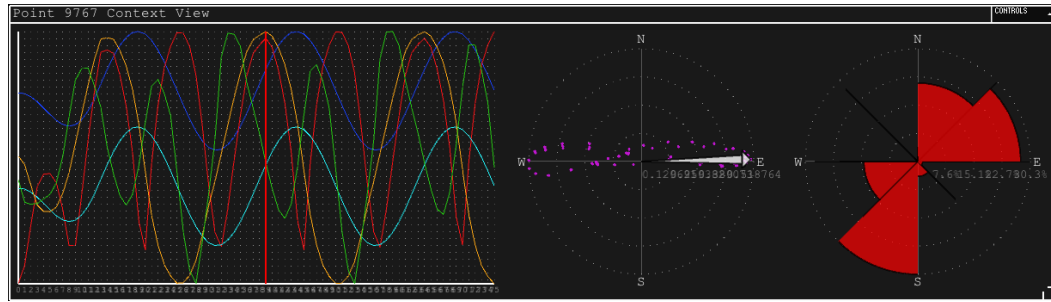


Figure 5.32: Point Graph View, with the 1D partition on the left, the temporal vector partition in the middle and the vector rose plot on the right.

The **1D partition** is based on a scatter plot (a useful plot for visualizing unstructured data), where the x dimension (axis) of the plot is always the temporal dimension (i.e. time) and the y dimension is the variable being investigated. This enables the user to see the tidal harmonics (variation of the tidal elevation and free surface), and how the other variables change temporally. The system allows users to switch each of the variables on and off and change the colour of them, thus providing a means of flexibly interpreting and understanding the impact of the variables.

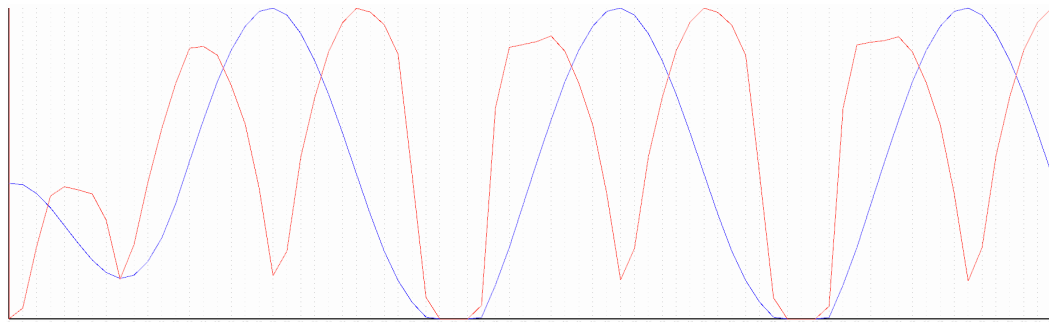


Figure 5.33: Output from VINCA showing the tidal harmonics with normalised scaling. The blue line is the water elevation and the red is the speed of flow.

The visualization of the 1D view allows for two methods of scaling for the vertical spatial dimensions (bottom, water depth and free surface) which can be selected by the user. The first method is based on the max / min of the individual dimensions (normalised scaling), see Figure 5.33. The other is based on the holistic

max/min for the spatial dimensions i.e. the maximum and minimum levels the water level or bathymetry reach over all time steps. This allows the user to see the spatial dimensions (bathymetry, water elevation and free surface) relative to one another, thus providing a 1D cross section through time.

Figure 5.32–left shows the 1D partition with holistic scaling: the x axis represents the bathymetric bottom, the dark blue line depicts the water surface, the light blue line depicts the free surface, the red line depicts the speed, the orange and green line depicts the u and v components of the velocity.

The **Temporal Polar Vector Plot/Partition** (see Figure 5.32–middle) provides the user with both an instantaneous and holistic view of the vector data at an individual point. Using a polar coordinate system the plot traces out the tip of the vector over time (providing the instantaneous view), and by connecting the points traced out by the tip of the vector, a basic representation of the tidal ellipse for this point is achieved, which provides the holistic view. An improvement to the system would be to calculate the tidal ellipse parameters from the vector data (as a pre-processing stage), and then superimpose the calculated tidal ellipse including all parameters over the polar graph. This would also link in with the vector overlays, where an option might be to visualize all of the tidal ellipses in the main view as well.

However, one problem encountered with both the temporal vector plot and the vector rose plot was how to select data from these plots, as normal drag and select only works for Cartesian screen space. A custom selection method was developed, which allows users to select data through polar space by allowing them to drag around the plot and select the points (see Figure 5.34).

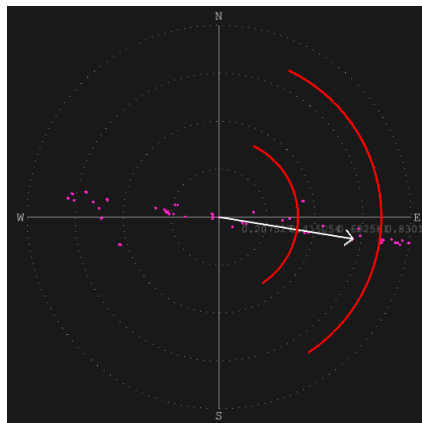


Figure 5.34: demonstrates the selection method used in the polar partitions

The **Polar Vector Rose Partition** provides the user with an holistic view of the distribution of the speed and direction of the vector data at this specific point, see Figure 5.32–Right . Each of the spokes on the plot represents a bin and the length of each spoke indicates the frequency of that bin. The bin with the spoke of longest length shows which direction had the greatest frequency. Other statistical information can also be encoded into the spoke such as: the maximum, average or median speed, through colour mapping.

The VRP within the system was visualized as arc sections, where the width (angle) of the arc represented the width of the bin and the length of the section represented the frequency. The user can also change the colour mapping of the spoke / arcs to encode either the maximum, average or median speed of the bins into the visualization.

5.7.6 PCP Query View

The PCP query view is one of the visual analytics systems main plots, and is one of the important contributions of this work. It provides the method of querying the data holistically (per time step) and temporally (animation). It provides a number of different methods of representing the PCP including:

- a raw PCP
- a frequency based PCP
- a spatially aggregated PCP.

The techniques used to develop this view and the underlying analytics are the subject of Chapter 6, Parallel Coordinate Plot Query, where they are described in greater detail. In this chapter the focus is on the use of the PR quadtree within the PCP.

Use of PR quadtree in PCP

A methodology was required to resolve the widely acknowledged problem with regard to parallel coordinates of clutter, which affects performance and inhibits interactive spatial exploration. A number of binning or aggregation simplification algorithms were considered to reduce the clutter, but discounted, as the spatial context of the data a highly important factor in oceanographic investigation - is lost. Thus,

a methodology for aggregating the data based on its spatial nature was required to retain the spatial context.

The solution was to use hierarchical spatial indexing. The related work highlighted the recognised value of spatial indexing for data querying. Thus, it was decided to develop a binning solution based on the spatial nature of the data, by first building a PR quadtree, which would be used for spatially indexing the data, and would then support the creation of several associated (linked) data structures. The approach adopted is similar to that of Fua et al's. [93] framework for hierarchical parallel coordinates, which bins the PCP on the spatial nature of the data (using a data space methodology based on level of detail, rather than data frequency) to aggregate the data in the geo-spatial domain. However, it differs from Fua et al. [93] in its approach to range selection, to ensure a range more representative of the data (explained in Chapter 6.5.4). Furthermore, Fua et al's. [93] system extends the generic XmdvTool [308], whereas that of this thesis is built on a turnkey system which takes into account the challenges presented by the data and the model – for example the use of adaptive mesh refinement.

The PR quadtree spatially indexes the unstructured mesh see Figure 5.12. The root of the tree represents the entire geospatial region: if a quadrant contains a datapoint, then the quadrant is split further. This occurs recursively until all the points are stored. On the leaf nodes, the coordinates; multivariate datapoints; a link to the unstructured grid of the data, and a reference to the original data are stored. This represents a full PR quadtree.

By associating the quadtree with the original data, the need to re-mesh is reduced and users are also able to utilise the fastest method when searching for points: when the user selects a point on the screen, this is mapped to the data through the PR quadtree. Additionally, a k-d tree has been implemented, using the median to split the points. Other trees, such as R and R^* trees have been considered, but because the quadtree grid retains the same aspect ratio throughout, it was preferred because when demonstrated, it was seen to be easier to understand by the ocean scientists.

When displaying the main window, the screen resolution is calculated, and the node capacity, c , is set to be dependent on the pixel size. Given the size of the data, each pixel may represent approximately 200 data points. When the bucket capacity, $c = 1$, the data in the quadtree matches directly to the unstructured mesh. However, this is unwieldy and reduces the frame rate of the data from 30Hz to less than 2Hz.

One problem with quadtrees is that a large fan-out may occur, when the datapoints are close together. But by increasing the bucket capacity, this subdivision can

be reduced. To select a specific point from the dataset, the user may either select the data by using a lens view, which provides a magnifying glass visualization onto the unstructured mesh, or by displaying a list of the actual values (of a specified parameter).

When the bucket capacity is not 1 then each of the nodes in the tree represent a range of values. Because multivariate data is being stored, there are different ranges for each parameter. Subsequently, summary information of the succeeding nodes is also stored. Computations are undertaken of the quantity of data points; the median; mean; standard deviation and also the quartile ranges, This node data creates the spatial bucketing used to create the aggregated PCP.

5.7.7 Transect/Flux View

This tool represents a significant contribution of this work. Developed specifically for studying estuarine flux, it provides the user with an iterative view of the data, which the underlying data set does not contain or cannot provide. This is achieved by deriving new data from the original data set, and is an important crucial step in achieving the knowledge discovery goals set by the thesis. The whole process of simulating the flux is discussed in greater detail in Chapter 7.

5.8 Improving System Performance

Despite the numerous methods described above for scaling and aggregating the data at various stages of the visual analytics process, there were still the expected performance issues, associated with large data sets of this type. The need for a smooth flow of data, without the distraction of lags in speed, bottlenecks, etc., at speeds promoting effective interaction is essential to the knowledge discovery process. Accepting that major architecture and hardware solutions were outside the scope of this stage of the project, a number of methods were developed to manage the problem.

5.8.1 Caching

Caching the results proved necessary to deal with the bottlenecks caused by the sheer quantity of data to draw. This required a speed/memory trade off. Each of the computationally intensive views are cached, thus permitting the fast playback required for initially overviewing the data.

The caching is undertaken at various stages:

- VertexList models (see Quickdraw) for the Main View are cached for each time-step ;
- heat maps are cached for each time-step;
- VertexList models for the raw PCP are cached;
- Many of the queries are cached: for example if a user queries the PCP view, then the results of the query are cached in both the PCP (brushed view) and the heat maps (main view).

5.8.2 QuickDraw

During the development of the system one major bottleneck emerged at the rendering stage, especially rendering the large numbers of triangles required to draw the underlying mesh. A couple of possible solutions emerged for this problem, either to render the data to an offscreen buffer and then display the resulting image, cache each time step and thus have quick playback. However, this still left undesirable effects. For example, if the view was scaled, the offscreen buffers would have to be refreshed and this took too much time, and actually caching the image also took too long. So another solution was developed which would prove to have further benefits at a later stage.

The solution was to use GLVertexLists, instead of the native Processing primitives, or a call to GLVertex from inside Processing. At this stage of working on the software development, Processing was not as fully evolved, and didn't have as many contributed libraries as it now has, when there was no specific library to meet the system requirements. This meant developing a custom library to build GLVertexList models, update and display them. Processing itself does not support VertexLists, but through its nature, (built on top of JOGL), it allows developers to hook into the lower level OpenGL framework and use native OpenGL methods and commands such as calls to the VertexLists.

By developing the Quickdraw library, the triangles are quickly rendered, with the ability to render up to approximately four million triangles at interactive frame rates (12*fps*). This also provided additional benefits: the VertexList models could also be used to resolve other problems. Another proposed view was the 3D context

view, (not yet developed due to time constraints, although it is recognised as necessary), which might use the same VertexList model as that developed for the 2D main context view, saving on memory and rendering time.

A decision was taken to cache each time step of the data as a new VertexList model, rather than having to update a single model at each time step, which would save the update time – one of the major time consuming factors in the rendering. A pragmatic decision was taken to adopt this approach as it was concluded that speed was more important than memory imprint. One of the key facets of a visual analytics system is the need for rapid interaction (important in convincing ocean scientists that such a system offered advantages their traditional analysis techniques could not confer), especially as desktop systems are gradually pushing 16Gb of RAM.

5.8.3 Export of publishable quality outputs.

A Visual Analytics system needs a method of being able to output its finding, which may take place through many differing forms, especially with modern techniques of information dissemination. A system is now not just limited to providing static images or even animations, but this data can now be uploaded directly to webpages or blogs. However, as this PhD is concerned with the analysis and generation of new data rather than methods of delivery, the simplest method of saving the output is provided through the use of a image exporter for each view, where the view can be saved as either a raster based image (jpeg, png, tiff) or as a vector based PDF.

A method of saving animations is also provided again, either by saving multiple single images which can be combined using any third party video editing software, or by creating a video file. A further feature which has been discussed but not yet implemented for time reasons, was a report builder, which would allow the user to generate a publishable report with different views arranged to the user's choice.

5.9 Summary

The chapter presents VINCA, a turnkey visualization and visual analytics system developed to meet the analytical needs of coastal shelf researchers in studying modelled hydrodynamic flooding simulations. It incorporates discussions of initial design and prototyping; identification of system requirements; selection of an appropriate programming tool and building the system. The requirements were seen to

be a fast, interactive visualization system and a visual analytics interface based on linked coordinated multiple views (CMV), which would provide researchers with complementary analytic tools for understanding the complex relationships and correlations of their multi-field data. These include visualizations, views, plots, and graphs, made available through a linked hierarchical CMV framework and interface, which facilitates interaction with and exploration of the data, including derived data.

The challenges in actually achieving a system which provided a stable, reliable platform; fast rendering; the required interactive and analytic capabilities and also publication quality images are also discussed. The enabling methodology for achieving this is considered in detail: a number of methods for scaling, aggregating, abstracting and otherwise filtering the data at almost every stage of the pipeline, which would retain the characteristics of the data, including its spatial nature and enable rapid visualization on a standard, fast desktop computer. Thus, the chapter devotes considerable attention to the selection and implementation of an appropriate algorithm, settling on a hierarchical data structure, spatial indexing, in view of its ability to handle unstructured AMR data, amongst other reasons.

CHAPTER 6

Parallel Coordinate Plot (PCP) Query

The techniques developed for this research focus on maintaining the spatial context of the data, thus the tools and techniques for aggregating one of the PCP variants were based on a hierarchical data structure, using a PR quadtree for spatial aggregation (an implicit hierarchical clustering strategy, which fits into Fua et al's. [93] theoretical framework of hierarchical PCPs). Three different types of PCP are described, to enable data exploration and analysis in a variety of ways. These are:

- A raw data PCP;
- Frequency based PCP (based on the work of Novotny and Hauser [211]);
- A novel spatially aggregated PCP;

Each variant of the PCP is plugged into a custom developed PCP architecture (framework), which allows the PCP to be manipulated and organised by the user, and includes an associative layout which groups dimensions with similar properties together.

The remainder of this chapter is structured as follows:

- Background to the value of PCPs to estuarine hydrodynamic research;
- PCP architecture;
- Raw Data PCP;

- Frequency PCP;
- Sa-PCP;
- Analysis;
- Chapter Summary;

6.1 Selection of PCPs & their value to estuarine hydrodynamic research

In simulating estuarine hydrodynamics, the ocean science researchers collaborating in this project wish to understand the interactions between sediments, bottom morphology and flow hydrodynamics, particularly the submarine processes of erosion and deposition. They wish to investigate how sediment is transported in an estuary, how tides, waves and sediment suspension affect erosion, and how these patterns change over a 100 year period. Thus, in relation to the hydrodynamics of the estuaries, they have produced a number of realistic mean and extreme tidal/fluvial scenarios to examine the specific model variables affecting circulation. These variables are the magnitudes of tidal amplitude, storm surge, river flow rates and sea level rise [240,243].

However, as the Related Work in Chapter 4 has indicated, identifying and quantifying associations between interrelated variables is a significant challenge for ocean scientists, because of the complexity and size of the datasets. Having described an interactive visualization and query driven visual analytics system, capable of supporting detailed analysis, the attention turned to tools for studying the attributes of the data.

Undoubtedly, parallel coordinate plots (PCP) are one of the main tools used widely in visual analytics, including the geosciences, for querying and exploring large, multi-dimensional datasets, through their ability to create compact 2D visual representations. Their relevance to this type of project is considered in detail in Chapter 4.

Currently, the estuarine ocean scientists have no means of viewing their data holistically so the ability of PCPs to provide visual clues should enable them to discover trends, identify areas of interest more easily, select data for further investigation and to develop hypotheses for more detailed study of the selected data.

Nonetheless, despite the anticipated benefits of the tool, Chapter 4 also highlighted the problem that PCPs share with many of the techniques used in visual analysis that of spatial clutter. The raw data PCP developed as part of this project illustrates the challenge, see Figure 6.3 which demonstrates intensive visual clutter. Thus much of the work in this section describes methods of resolving this problem to maintain interactivity; to provide clarity within the views and to retain the characteristics of the data.

As the Related Work illustrated, there are very few examples of the use of PCPs for coastal shelf and estuarine data. However, PCPs have been used in climate modelling, which presents similar challenges to modelled estuarine flood simulations. Thus, the development of the PCP tool for this thesis has been influenced by Steed, Swan, Jankun Kelly and Fitzpatrick [273–275], whose PCP system for studies of tropical cyclone and hurricane systems, incorporates a rich array of tools, in dealing with similar analytical problems to those encountered with the data for this work.

6.2 PCP Architecture

As with the majority of the system the architecture of the PCP is split into a number of separate components (see Figure 6.1):

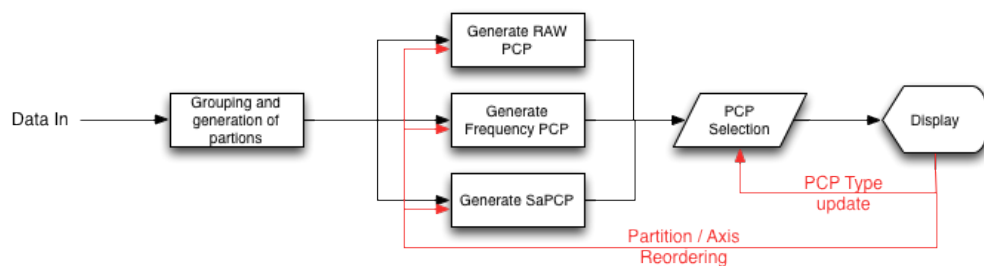


Figure 6.1: Flow chart diagram of PCP architecture

- Initial import of the data into the PCP framework;
- Grouping the data into defined partitions;
- Division into three distinct processes for the PCP variates;
 1. Generation of Raw Data PCP.
 2. Generation of the Frequency PCP.

3. Generation of the Sa-PCP.

- Individual visualization of each variant of the PCP in the Query view within the CMV environment;
- User selection of a version of the PCP for exploring the data.

6.2.1 Processes common to the PCP variants:

There are a number of processes which are common to all the PCP variants:

Data grouping

Each axis represents an individual dimension (variate) of the data, and variates displayed on the PCP include position (x,y components of the mesh), height of sea bed, water elevation above sea bed, speed, velocity and free surface height.

After building the PCP, the data is placed into separate groups of axes, dependent on its nature: the x and y components of the mesh become a spatial group; each of the scalar variables becomes a scalar group, and the two or more components comprising the vector data become a vector group. The goal of this is to provide the basis for a logical ordering for the axes, by relating them to the nature of the data, such that each group represents different semantic information.

Grouping axes together allows users to move several axes together, as one unit. Other operations can be performed by group, such as removing a group from the display, or viewing the 1st and 2nd derivatives of a variable. This supports speedier, more intuitive interaction, rather than a slow, painstaking process where the user has move axes individually to obtain a logical ordering. Each of the groups was assigned a separate partition in the PCP, which would dynamically generate the correct number of axes for the group, dependent on the number of data variables identified for each group. The partition also holds all of the working data and derivatives associated with this partition/group.

Axis Ordering

The purpose of grouping the axes in this way, and presenting a default setting is more than simply providing a quick and easy method for users to manipulate the axes. Effective ordering of the axes is regarded as an essential component of any

good PCP, with Qu et al. [230] concluding that axes presenting a potential correlation should be placed close together. This system starts from the premise that grouping axes by data type will provide an initial, logical basis for ordering the axes. It is designed to deal with the concerns of dos Santos and Brodlie [77] that a lack of a pre-defined order, and a resultant inappropriate arrangement of axes (data variates) may obscure rather than clarify relationships.

The system is flexible and users may change the order of axes as their inquiry and data exploration process dictates, by a simple interactive drag and drop facility (of proven effectiveness in a number of systems [38, 284]). Furthermore, grouping the data makes it easier for the user to reorganise axes, as it reduces the number of axis movements necessary to retain a logical ordering of the data. Additionally, each of the partitions is able to be dragged horizontally such that multiple arrangements might be made and the axes can be reordered within each group, thus allowing for many possible reorderings.

Other methodologies for reordering axes were considered: Chapter 4 discusses some of the other approaches adopted by PCP developers, as well as drag and drop, for example the quantifiable statistical approach used by Steed et al. [274]. However, the complexity of the PCP, with its three very different variants, presented challenges to alternatives. The temporal nature of the data in conjunction with the multiple PCP variants, would lead to a different optimum axis order for each time step and each variant, which might make for a very confusing discovery trail for the user (and a user unfamiliar with advanced analytic techniques).

Setting the axis scale

The scale on each axis of the PCP has a default range set to the holistic minimum and maximum of its dimension, which represents the minimum and maximum of a particular variate throughout the complete time series, and is calculated in the pre-processing step of data import.

The second scaling option is to normalise the axes for each time-step by using the maximum and minimum values for each variate, from the current time step being played. The user is able to select which option they prefer for each individual axis, enabling them to normalise the data to a particular time step or else study how the time step relates to the whole time range in general.

Other facilities

Another option provided is inversion of the axes, so that trends may be made more obvious. Steed et al. [274], for example, in their cyclone and hurricane PCP system, use inversion.

Querying the PCP

Querying the PCP is undertaken using a range slider over each axis. When the slider is moved, it performs a MySQL query, using the different ranges specified for each dimension/axis as the limits for the query. The result set returned by the query is then highlighted and the associated heat map in the main view is updated to display the data returned by the query.

The second mode of query is driven from the main view, when a user selects a particular PR quadtree cell. This is passed downstream to the PCP view, automatically setting the spatial ranges which fit the PR quadtree cell, then queries the data, thus providing a spatially indexed query. Once the user de-selects the PR quadtree cell, the PCP reverses to the full spatial range.

Further ideas for data querying and manipulation

Another option investigated but not implemented due to time-constraints is dynamic axis scaling, similar to Steed et al. [274]. Use of a second set of range sliders, to dynamically increase or decrease the axis scaling, enables the user to progressively zoom into the data to focus on a small subset of the data, and is based on the widely used focus+context enhancement to PCPs [93, 211]. This, in conjunction with an approach similar to Steed et al. which draws the focus lines to the viewer's attention through the use of a darker hue for them (based on human perception of aerial perspective) would reduce line clutter and make it easier to analyse lines of interest. This could be considered as a future enhancement to the system.

Focus+Context

Focus + context is an important concept with regard to PCPs, in its ability to overcome display limitations, and it is successfully used in the PCP variants of this project. The tool integrates a visually accentuated representation (through greater detail and opacity) of selected data items in focus with a visually de-emphasised

representation of the rest of the data (the context). The aim is to improve user orientation, provide an overview of the data, and support the identification of features of interest. Novotny and Hauser's [211] algorithm re-uses the binning previously used for outlier separation and trend clustering. Focus+context is used in several of the attribute plots, including the vector visualization overlays, and also the raw data PCP brushed view, through the use of transparency.

6.3 Raw Data PCP

A raw data traditional PCP was built, as a control, to provide the basis for testing methods of presenting the data to manage spatial clutter and obtain the required results to support interactivity and provide a PCP which would enable exploration and analysis. The datasets for this project contain around 10^6 data points, which is the level at which Fua et al. [93] concludes that axis configuration techniques such as parallel coordinates become unsuitable, without tools to further scale and aggregate the data.

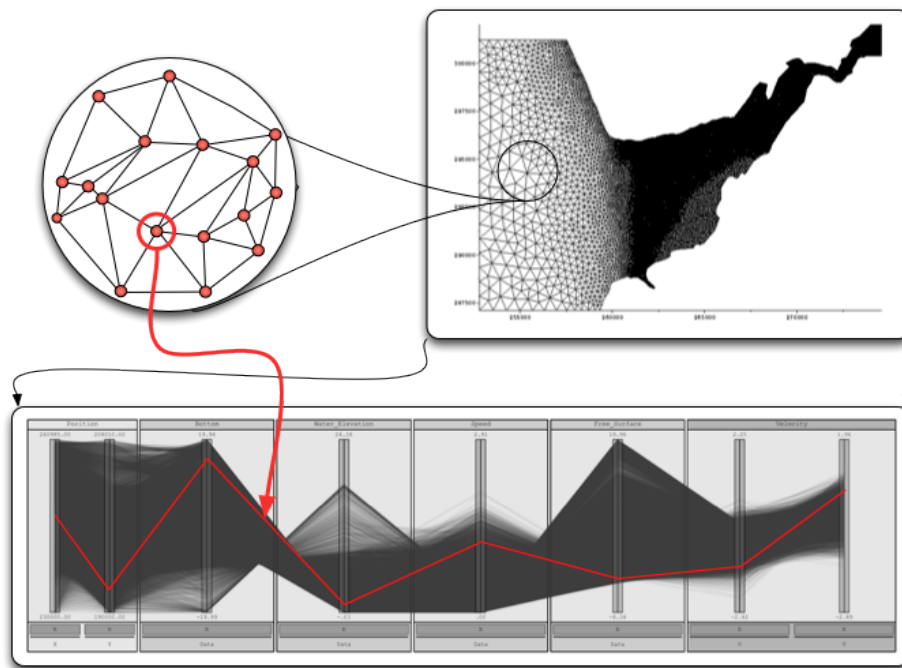


Figure 6.2: a point on the mesh, related to its PR quadtree polyline

However, the size of the dataset not only hindered rendering, because of the

number of lines and thus calls to GLVertex, but also produced a PCP which was virtually meaningless and impossible to interpret. But whilst the clutter, which hindered exploration of the data, could not be resolved, it was possible to achieve interactive frame rates when the PCP was rendered, through a tool developed to solve another problem in the visual analytics system. A novel manner of storing and displaying the polylines making up the PCP was implemented through the custom QuickDraw library, initially developed for the mesh data in the main view. This library was adapted to allow for polylines. This success permitted further development work to focus on reducing clutter to improve the data analysis capability of the system.

6.3.1 Visualizing the Raw Data PCP

The main issue in visualizing the Raw PCP is that the size of the data sets makes it very difficult to spot trends or outliers, because of the cluttered graph. Approaches to solve this included the use of transparency: Wegman’s [315] alpha blending represents the density of the plots with transparency, which Nocke et al. [208] believes to be very useful for comparing dense simulation results. High levels of transparency were used for extremely dense data, ranging to opaque for less dense data. Alternatively, brightness/ luminance or colour saturation, may be used in a similar way to transparency, in conjunction with additive blending. However, these techniques do not solve the problem with extremely large data bases, as either the alpha channel (in the case of the transparency), or the colour channels are very quickly blown out (i.e. the alpha or colour values maxes out at 255), and the clutter is only slightly reduced. Figure 6.3 shows the use of transparency on the raw data PCP, where even the lowest transparency level is not capable of reducing the clutter.

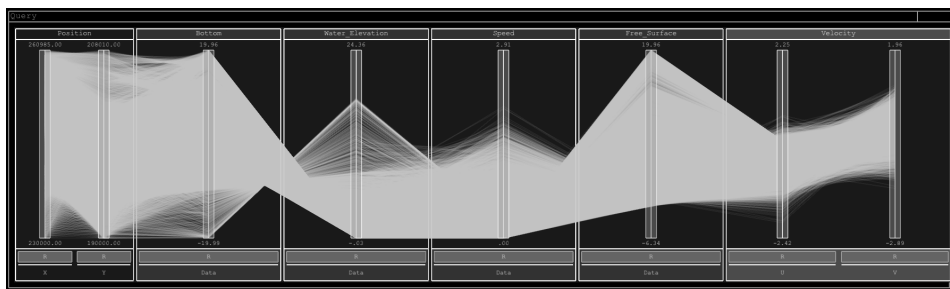


Figure 6.3: Picture illustrating the RAW PCP still blown out even though each polyline has only a transparency value of 1 (out of 255).

Furthermore, and importantly, by reducing the visibility of individual results, it

becomes difficult to assess whether there are outliers within the data, and these may be of considerable significant to the ocean science researchers, rather more than the general trend.

6.3.2 Query Results

Querying the PCP is undertaken indirectly by adjusting the range sliders on the axes, and the query results for the Raw PCP are represented in two ways: firstly, as a brushed view within the PCP, and secondly a heat-map within the main context view and all subsequent down-stream sub-views. Brushing within the PCP shows the user which points have been selected in the query, and the selected points are presented in a different colour to the remaining points in the PCP. Generation of subsets of data through direct brushing (mouse selection) has not yet been implemented, but would certainly be a useful addition to enhance the user's ability to drill into the data.

The heat map acts as the link between the PCP and the main context view, and all subsequent downstream views, as PCP queries update the heat map in the main context view. It is a copy of the mesh, where the points/cells returned in the query are emphasised by colour, and the remaining cells become transparent, thus placing the focus on the data of interest.

6.4 Frequency PCP

Having produced the raw data PCP, it was readily apparent that further methods of data manipulation would be required to enable users to see the distribution of the data and identify trends more effectively. A frequency-based, binning approach, based on a variant of the frequency and outlier preservation algorithm of Novotny and Hauser [211] was implemented, with the aim of overcoming the overplotting issues, thus enabling the user to gain a clear overview of data distribution and identify trends. Novotny and Hauser's algorithm was attractive in that it enabled the identification of both trends and outliers, which are often of considerable significance in hydrodynamic studies and may require detailed examination. From the system development perspective, identification of outliers, especially when using focus+context, is important as they may have a detrimental impact on the accuracy of the context.

6.4.1 Using the Novotny and Hauser algorithm to build a Frequency PCP

The algorithm relies on binning the data with a regular bin map, which is then used to build the frequency based PCP. A PCP was built using this method, including Novotny and Hauser's [211] approach to detecting outliers, but after testing the algorithm, it was discovered that even though it dealt neatly with an over-plotted PCP in a standard case of non-spatial data, there were drawbacks and issues when used with the complex modelled hydrodynamic datasets, and it struggled with unstructured, multi-resolution spatial data.

In the case of this data, the algorithm's binning strategy (the fixed bin map), which is employed on a per axis / dimension basis is not suitable for the spatial nature of the data, and sampling problems occur. This is not evident in the visualization of the PCP, but when the binned data is plotted as an associated heat map overlay in the main view, the sampling discrepancies becomes extremely visible in the form of a chequering pattern in the low resolution part of the mesh, see Figure 6.4.

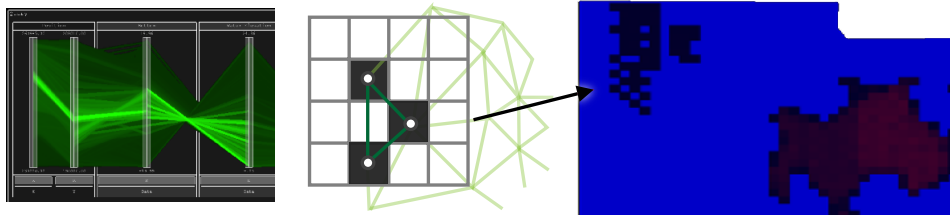


Figure 6.4: Because of the multi-resolution grid, it is unsuitable to utilize a fixed bin to create a frequency-bin of the data. Although the visual depiction of the parallel coordinate plot looks appropriate (as depicted left); when the data is plotted to an associated heat-map the sampling problems are clearly shown (right).

6.4.2 Development of a variant frequency PCP

Moving on from the Novotny and Hauser's [211] algorithm, to remove the sampling discrepancies, a solution was proposed which used the PR quadtree generated for the interaction within the main view, to bin the spatial dimensions only. This would allow for an adaptive bin map to be generated (with varying bin sizes), having finer resolutions in areas of higher mesh density, which then utilised regular binning for all other dimensions within the PCP.

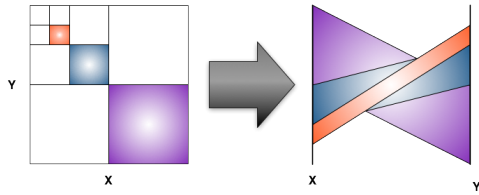


Figure 6.5: Spatially indexing binning

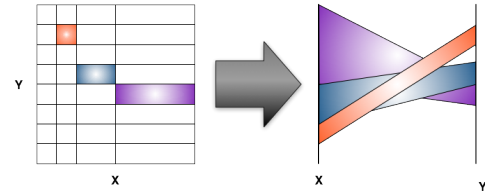


Figure 6.6: Spatial Bin to Regular bin

This enables the data to be binned at a LOD in the spatial dimension that is suitable and fitting for the space. However, this algorithm did not recognise the range of the data contained within the PR quadtree cells. Thus, as each of the non-spatial dimensions were binned in screen/PCP space, which is a regular bin, they were losing detail regarding individual records (individual spatial bins), whilst gaining information about how busy specific regions of the PCP were. As a result, the information about the individual points contained in the spatial bin was split up into the regular frequency space. So, resolving the sampling discrepancy resulted in a different issue.

6.4.3 Spatially Indexed Binning

As the variant utilized the same PR quadtree to perform the binning as that used for the interaction in the main view, so the associated heat map fitted directly in with the PR quadtree's cell structure used in the main view, and thus provided a neat and more closely linked visual depiction to interaction methodology, see Figure 6.5.

For displaying the joining section between the two types of binned axes (spatially indexed axes and regular axes), an approach of creating a bin map that represented the co-joining of the two different data structures was required. An irregular bin map was developed for this purpose, see Figure 6.6. The X axis represents the spatial indexed axis, with a non-fixed bin dimension, and the Y represents the axis with a fixed regular bin dimension. However whilst the bin map successfully removed the sampling discrepancies, in doing so, it lost the information about the range of values contained within each spatial bin, see Figure 6.7.

6.4.4 Evaluation of the variant frequency PCP

The variant dealt with the re-sampling issue, which made the Novotny and Hauser [211] algorithm unsuitable, but in so doing, it failed to recognise the range of the values contained within each cell of the PR quadtree. As such, this represented more of

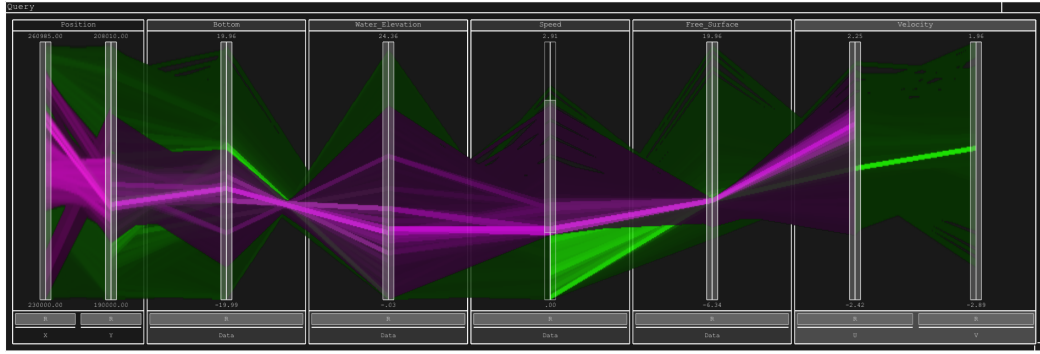


Figure 6.7: Visualization of the frequency PCP variant using the quadtree to provide binning for the spatial dimension (the two leftmost in the figure)

a hybrid visual/data space solution rather than a pure data space solution. Consequently, the development of the Frequency PCP was not completed to include the clustering technique. However, having pursued a spatial data approach in developing a PR quadtree for the pre-processing of the data and for other tools, there was an interest in examining whether the PR quadtree algorithm might be extended to aggregating the data for the PCP.

As Novotny and Hauser [211] suggests that using “hierarchical or adaptive binning is possible” which might “improve the situation in a data-dependent way”, it was decided to explore this by developing an algorithm which would enable binning in a data-dependant manner, thus allowing the use of a hierarchical structure to aggregate the data for the PCP.

6.5 Sa-PCP (Spatially aggregated PCP)

The Sa-PCP was designed around an interactive architecture which, in conjunction with use of the PR quadtree to dynamically aggregate the data, would allow the user to select various levels of detail quickly and easily. It was developed to prove the validity of a spatial frequency approach to dealing with these large ocean science datasets. This would bin using just the spatial dimension, and then use this to draw the PCP. Thus, the range of the data would be retained, whilst reducing the complexity and clutter of the traditional PCP. Building the Sa-PCP was split into two sections:

1. Pruning the PR quadtree to the desired level and calculation of required statistics and metrics.

2. Building the Sa-PCP using the selected range (either the full range or 1SD either side of the mean).

Splitting building into the two separate stages allowed interaction and updating to be downstream of the filtering (pruning) stage, thus this operation would not have to be performed repeatedly each time the Sa-PCP was built after a rendering update.

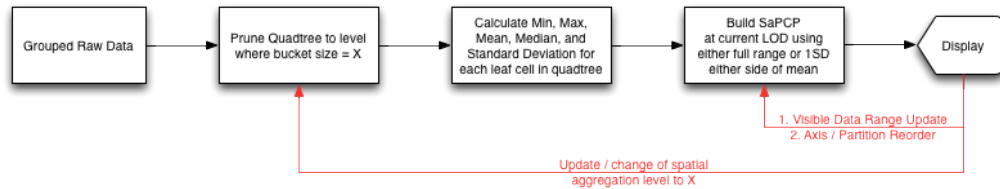


Figure 6.8: outlines the flow and generation of the Sa-PCP

For the rest of this section, two distinct processes are described, starting with pre-processing the data required to build the Sa-PCP using dynamic filtering/pruning by a PR quadtree. Secondly, construction of the Sa-PCP is considered.

6.5.1 Utilising Spatial Indexing to pre-process the data

A PR quadtree, built to the level where the bucket size of the PR quadtree $c = 1$ and associated to the raw data was utilised. This gave a spatially indexed reference to the whole dataset. When the bucket capacity $c = 1$, the data in the PR quadtree match directly to the unstructured mesh. However, this does not aggregate the data and thus does not solve the problem of overplotting, clutter or speeding up the rendering. But, by increasing the size of the bucket capacity to > 1 the subdivision of the PR quadtree is reduced, and a spatially aggregated dataset is obtained. When the bucket capacity is not 1, then each of the nodes in the tree represents a range of values. This is because multi-variate data is being stored, and, there are different ranges for each variable contained within the dataset. Subsequently, other statistical information relating to the nodes is computed, to better understand the distribution of the data, and stored. This includes:

1. the quantity of data points;
2. the median for each variable;
3. the mean for each variable;

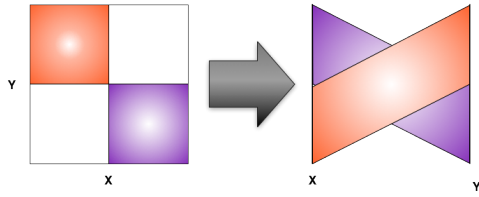


Figure 6.9: This figure represents a very high level of aggregation.

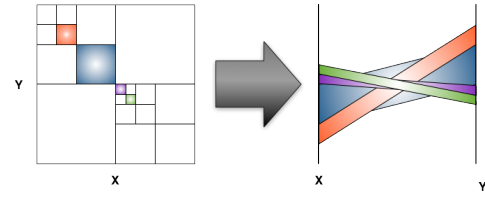


Figure 6.10: This figure represents a very low level of aggregation.

4. the standard deviation for each variable;
5. the interquartile range for each variable.

Each of these, apart from the quantity of nodes, is computed on a per time-step basis so the data may be displayed temporally. The holistic range (minimum and maximum of all cells) for each variable are also computed on a per time-step basis and over the complete time range. This will allow for the two methods of scaling the PCP's axes, either normalised to the current time-step or to the complete time-range.

6.5.2 Construction of the Sa-PCP

As already discussed, the data are arranged into groups of axes, with each of the axes/groups representing each of the variates contained within the dataset. During the data-loading phase, one of the pre-visualization processes is construction of the PR quadtree used in the system. Here, the PR quadtree used for managing the Sa-PCP is built with a bucket capacity of 1, which will give full access to all the underlying points contained within the data, and allow for many levels of aggregation. Once built, all of the statistical metrics are calculated for every single node in the PR quadtree. This data structure is then passed on to build the Sa-PCP at a level of detail specified by the user, dependent on the research processes being undertaken — to identify features of interest, drill down into the data, or study particular data points, for example outliers.

As shown earlier, when the bucket capacity $c = 1$, an analogue of the full traditional PCP is built. But drawing/building with $c = 1$ is not efficient, because the estuary data are huge, and the problems associated with traditional PCPs, will not have been resolved, see Figure 6.3.

However, when $c \neq 1$, there are several choices of how to display the data in the PCP. The average of all the values within each cell (mean or median), of the spatial

aggregations might be displayed with a single line. But this has limited use, and does not present the full range/spread of data values, as it is an over simplification. To overcome this, the data can be presented as a range of values, either the full range, the inter-quartile range or as one standard deviation either side of the mean.

There is, however, an optimum balance between aggregating the values on the PR quadtree and the spatial resolution of the visual depictions, and the user can select the level of detail – the aggregation of the PCP, based on a slider. When the slider value v is updated by the user, this causes the PR quadtree to be pruned to a level of detail where bucket capacity $c = v$, and rebuilding the Sa-PCP at this new level of detail is initiated. As the slider value changes, this causes the visual depiction of the Sa-PCP to change and also forces an update of the associated heat map in the Main View to the same level of detail as in the Sa-PCP. It may be possible to aggregate the data using a different spatial scheme, not simply changing the bucket capacity. While the PR quadtree (and k-d tree) do provide a spatial aggregation mechanism, the aggregated cells are square (or rectangular, in the kd-tree case) and at higher aggregations this would not be representative of the data. Therefore, it may be possible to design a tree-like structure that aligns better with the data. For instance, a local split method may be designed which splits along the length of the estuary, but that is left to future work.

6.5.3 Visualizing the Sa-PCP from the range data

Use of the spatial aggregation PR quadtree algorithm prunes the PCP such that the level of detail fits the space. This then enables the three different ranges (full range, interquartile range, and 1 standard deviation either side of the mean) to be displayed either as:

- envelopes.
- transfer functions.

6.5.4 Display Ranges

Whilst the PR quadtree fits into Fua et al's. [93] theoretical framework of hierarchical parallel coordinates, the ranges selected for this project differ. To provide a range which is more representative of the data, it was decided to display the mean value (as the centre) plus or minus one standard deviation. Fua et al. displays the range and uses a ratio of $n1 : v1$ to change the opacity of the variable width bands

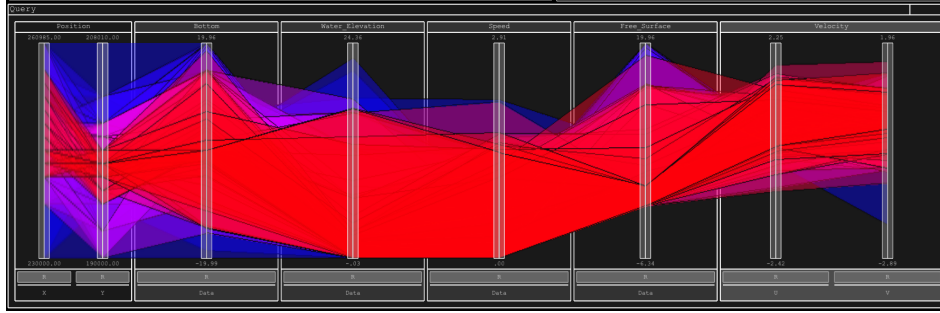


Figure 6.11: This shows the envelopes using the full range of data contained at each node in the PR quadtree at an aggregation level X using a depth based sort.

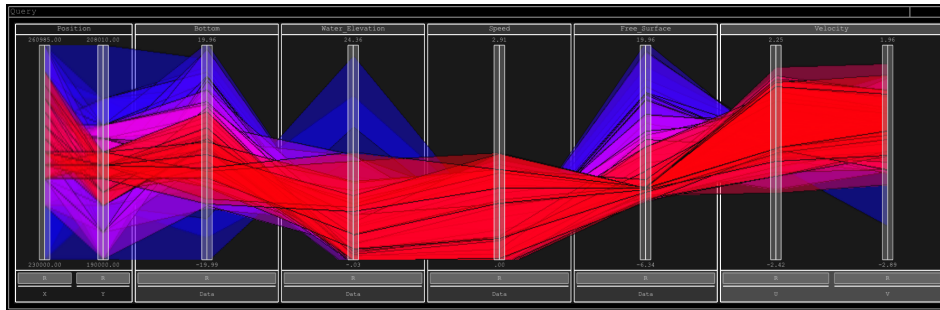


Figure 6.12: This shows the envelopes using 1SD either side of mean at each node in the PR quadtree at an aggregation level X using a depth based sort

to represent the range better (where $n1$, is the number of points, and $v1$ is the size of the cluster of the node of the tree).

6.5.5 Display using Envelopes

Display of the range data using envelopes is shown in Figure 6.11. Because the bands overlap, both transparency and a simple sorting method are used that display the larger bands at the back, followed by the smaller bands. This presents a depth sort (or topological sorting) challenge, and it is accepted the adopted heuristic may not provide the optimal solution. However, the other alternative method is to calculate the ordering in a pairwise manner, but this then introduces discontinuities between the axes.

Nonetheless, the creation of envelopes is visually appealing, and provides a means whereby the user is able to understand the range of the data contained within each bin. However, overplotting is still a problem with this display technique.

6.5.6 Transfer Functions

A number of transfer functions, including line-intensity and tube-like (Figure 6.13) were implemented. They may be applied to any of the variants of the PCPs and by changing them, different facets of the data is able to be highlighted. Figure 6.16 shows four different aggregation values on the same time-step and a full rendering of the tube-like transfer function is shown in Figure 6.14.

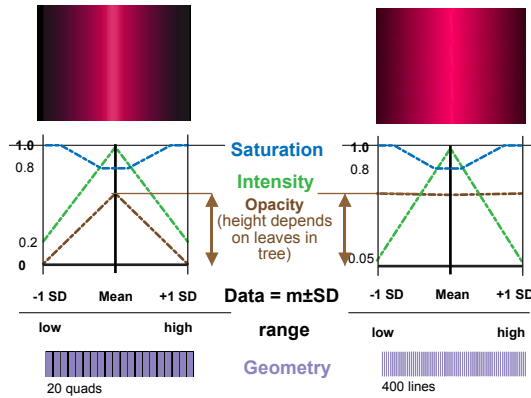


Figure 6.13: Visual depiction of the transfer functions used.

6.5.7 Associated Heat Map

The associated heat maps see figure 6.16 (right hand column) in the main view give the user a visual depiction of the currently level of aggregation. Figure 6.16 shows that the heat maps and thus the Sa-PCP retains the spatial context of the data. As with the other heat maps, when the Sa-PCP is queried only the cells that are returned in the results are displayed.

6.6 Evaluation of the techniques

Whilst it has been demonstrated that it is possible to use a spatial PR quadtree to generate a hierarchical bin map for use with a PCP, which provides a useful tool for exploratory visualization of large, complex, multi-resolution and multi-variate oceanographic and other spatial data, there are some limitations. For instance, the aggregations appear the most elegant when the aggregation level is high, but this represents a large area, and there are still some, although less, issues with overplotting at lower aggregation levels. A solution might be to use an R^* tree with a topological split, however, the R tree variants may not be suitable for the types of

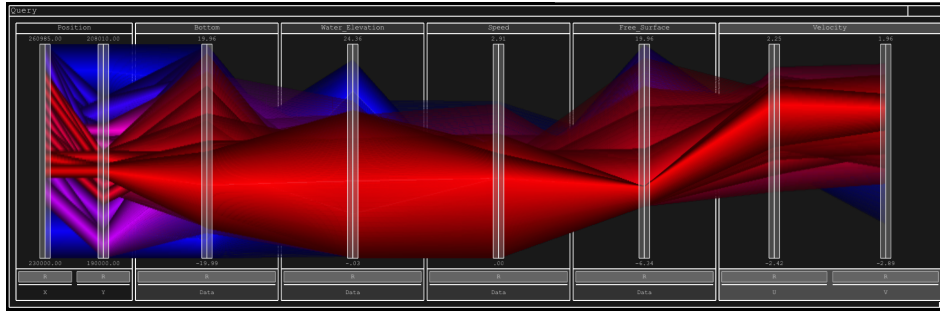


Figure 6.14: This shows a visual depiction of the tubes transfer function without depth sorting

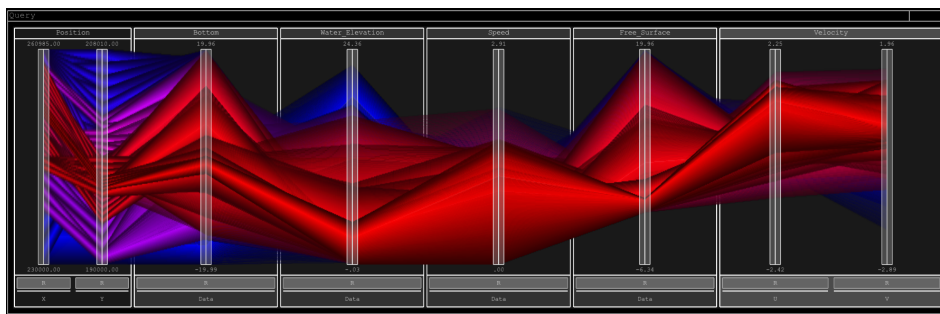


Figure 6.15: This shows a visual depiction of the tubes transfer function with non pairwise depth sorting.

queries and the structure of the modelled hydrodynamic data. Another extension may be to investigate temporal tree variants as this would enable the time nature of the data to be included.

The challenge is further complicated by the requirement to investigate several data runs. This adds another dimension, but to provide full support to these predictive studies, it would be useful to incorporate multiple data runs into the PCP. Some researchers include these as categories within the PCP, but it would certainly increase the requirement for a wide variety of methods for filtering and aggregating the data, and at this point, it may be appropriate to consider other means of demonstrating the characteristics of the data revealed by the PCP.

6.7 Summary

The focus of the chapter is PCP, parallel coordinate plots, a widely used visual analytics tool, and one with a record of successful application in the related area of climate and meteorological studies. It forms one of the principal components of the

visual analytics system, and is thus, a key element of proving the hypothesis that it is possible to build a visual analytics system for large, complex, modelled, coastal shelf and estuarine datasets and that this might support knowledge discovery. The selection of PCP as an exemplar of the value of visual analytics is discussed and placed into the context of the research needs for studying simulations, where the requirement is to establish trends and correlations.

Three variants of a PCP were developed and tested to assess their ability to explore and interactively visualize the datasets, and to prove the capacity to provide derived data for further analysis. The three PCP variants described are:

- Raw data PCP;
- Frequency PCP;
- Sa-PCP;

The chapter discusses their architecture; the query, exploration and analytical processes available to users; the challenges presented in developing this component of the VA system and proposed solutions for resolving these. The attributes of the PCP are also described and include a logical grouping for the axes based on the nature of the data; axis re-ordering capability; options for setting the axis scale, focus+context; axis inversion; some derived data through statistical measures.

In selecting PCP as a case study, it was acknowledged that overplotting was an issue with regard to large data, so much of the focus of the chapter was on developing a PCP which would support insight through the use of data abstraction and aggregation techniques. The raw data PCP was developed to show the extent of the challenges to rendering and interpretation through overplotting, and whilst interactive rendering was achieved, problems remained with clutter, and consequent outlier detection.

Various methods for dealing with this, including transparency, brightness/luminescence and colour saturation were discussed but did not easily support the identification of outliers. A frequency PCP, based on the frequency and outlier preservation algorithm of Novotny and Hauser [211], is also described and also discarded: two variants were developed, which both proved unsatisfactory, one of the problems being the unstructured, multi-resolution nature of the data.

A novel spatially aggregated PCP, and associated heat map is proposed, based on the dynamic variant of the PR quadtree described in Chapter 4, which allows the user to select level of detail.

Finally the Sa-PCP is evaluated as providing the capacity to investigate the data temporally and holistically for trend analysis and proposals for developing and extending its capabilities are discussed.

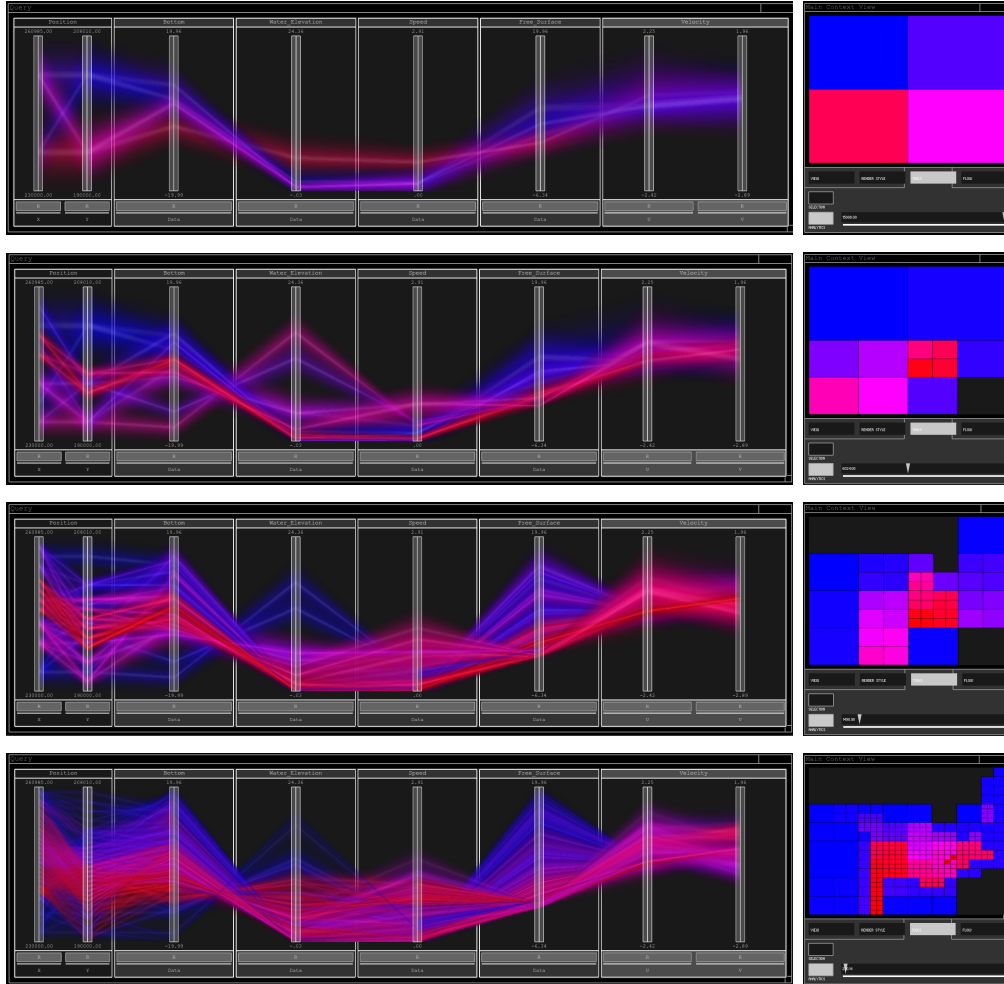


Figure 6.16: The figures show different levels of aggregation of the tree. In the left hand column is the PCP while the right column shows the plot window. Both are linked together. The user can interactively change the level of aggregation to explore. While it may not be useful to display the tree at such low levels, the figures demonstrate how the user can choose the level of aggregation. These views depict a single time step, however the user can step or animate through time to see how the data changes over time.

CHAPTER 7

Deriving New Data, Completing the VA Loop

As a case study of creating derived data for visual analytics, this chapter of the thesis focuses on the development of a flux calculator for estuarine hydrodynamic studies. The Data and Related Work chapters have already discussed how flux calculation presents significant challenges to the researchers, yet it is a measure which is extremely important to understanding their simulations and producing flood predictions to aid coastal zone management.

Tidal flux is an exemplar of the limited ability of models to provide all the quantities required by researchers to undertake detailed analysis of the estuarine domain and fully understand the simulations. Earlier chapters have also shown how time costly and difficult a process the production of these derived diagnostic data can be. However, this type of calculation is not something which the typical suite of visual analytics tools is likely to support – hence the need for a bespoke approach.

This chapter therefore considers:

- Transects;
- Flux: what is it, why is it important and how is it currently calculated;
- Related work;
- Architecture, tool components and developing the tool;
- Transect profiler;

- Flux calculator;
- Calculating flux along an estuary;
- Visualizing the transect and the flux;
- Evaluation of tool and techniques;
- Summary;

7.1 Transects

A transect is simply a cross sectional profile of a coastal zone/estuary (see Figure 7.1). Oceanographers will survey multiple transects in a research area using both classical survey equipment (theodolite and staff) and modern differential GPS techniques. Classical survey equipment is still preferred in some situations when surveying transects, and in many cases a volunteer stands in the middle of a channel to find the deepest point.

Multiple transects give a visual depiction of the shape of the bed through the research area, and these subsequently allow for quantitative metrics, such as cross sectional area of the water column, to be calculated when used in conjunction with other experimental / observational tools such as tide gauges and current profilers. This allows oceanographers to understand the dynamics throughout the whole research area.

One of the aims of the transect profiler in this work is to enable oceanographers to simulate field practice, particularly as within the world of computational oceanographic simulation, the method of producing a transect is mathematically more complex than standing in the middle of a channel, yet requires much less physical effort. A cross section is cut/resampled from a mesh depending on the data type (regular, unstructured etc., which will govern the type of interpolation technique). A starting point and ending point is provided and then an algorithm interpolates a set of regularly spaced points across the transect containing the position and underlying data variable values see Figure 7.1.

7.2 Flux

A measure of hydrodynamic flux is highly desirable to oceanographers, as it enables a detailed picture of the volume of water moving through an area to be developed.

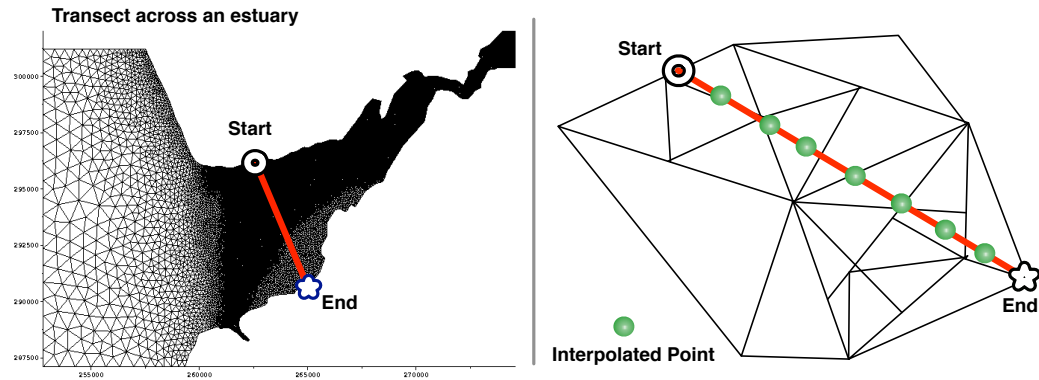


Figure 7.1: The schematic represents a typical transect across an estuary (left) with points that are regularly sampled to interpolate the data across this transect (right)

The flux of water moving across a transect (see Chapters 2 and 4) of an area provides a clear indication of the net transport of water, and with it, the net transport of sediment (suspended in the water) within a coastal or estuarine domain. It will help validate the coastal model as these transects can be compared to observational data, where researchers use either mechanical, electromagnetic or acoustic current profilers to measure the flow along the transect in the real world.

However, whilst it is a valuable measure in its own right, in providing a greater understanding of hydrodynamic and sediment movements, as the earlier chapters have discussed, it is also a component of the tidal prism (a measure of the difference in the volume of water between high and low tide i.e. the amount of water flowing in and out of the estuary at any given point).

Figure 7.2 is an example of flux graphs of one of the test data sets for this work, produced with the traditional post processing methods used by oceanographers [238]. The graphs show test flux profiles for comparing two scenarios: (a) – which represents a hypothetical model domain and (b) the present day estuary shape. Cross-estuary profiles were calculated every 2 kms east from the mouth of the Dyfi (just five calculations per scenario) for a hypothetical spring/neap tidal cycle (15 days). Apparently identical, they are not, but it is difficult to identify the variations, without detailed and painstaking comparison of the graphs.

Flux and tidal prism may appear somewhat obscure calculations, but understanding them has many applications, especially in coastal zone and estuaries. In the case of this research it is being used to assess the confluence of many factors which might lead to flooding. Other applications include: port management (for dredging and maintaining shipping channels); fisheries (for understanding and con-

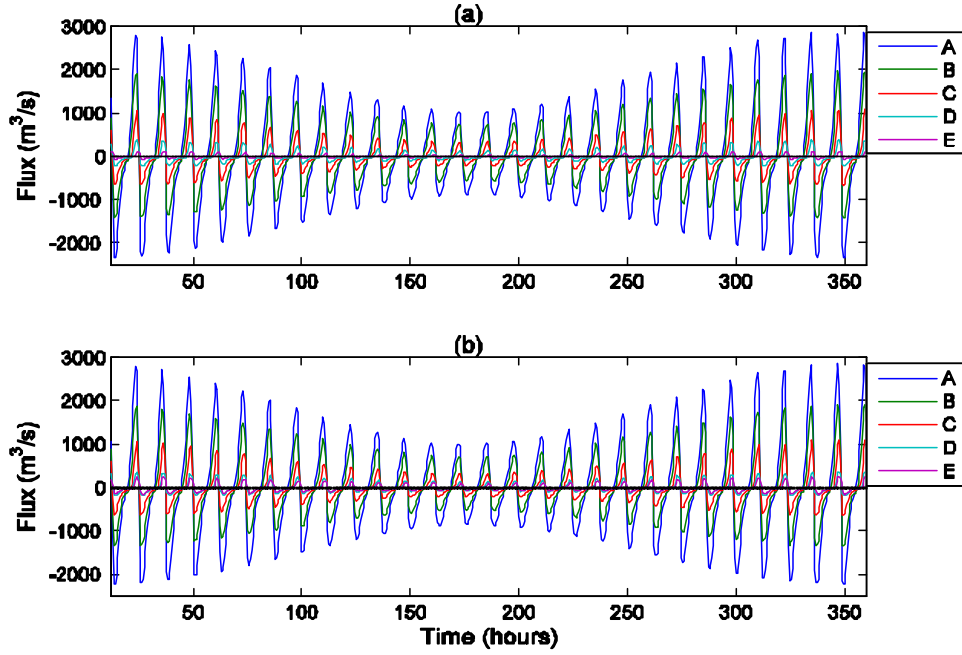


Figure 7.2: Multiple fluxes from various locations in the Dyfi estuary. Reproduced from Robins [243]

serving habitats to preserve fish stocks); studies of water quality and water pollution, including the management of polluting events, to name but a few examples.

7.2.1 Definition

Specifically, hydrodynamic flux is the flow of water through a given surface per unit of time. It is measured in cubic meters per second. Flux is calculated across a transect.

If $U(x, y, t)$ is the depth-averaged velocity component in the x direction, at time t and at point (x, y) in the domain, and a transect is taken across the channel in the y -direction at location $x = x_0$, then the instantaneous water flux, normal to this transect is given by:

$$Q(t) = \int_{y1}^{y2} U(x_0, y, t) [z_b(x_0, y) + \eta(x_0, y, t)] dy$$

where $y2$ and $y1$ are the points of intersection of the water surface with dry land at time t , z_b is the water depth below the Mean Water Level (MWL), and η is the

tidally varying elevation around the MWL.

This is a simplification of the equation, and deals only with the x component of the velocity, i.e. the transects would all have to be vertical. The new tool would allow for fluxes along transects with arbitrary direction to be taken and calculated. However to create a model for the flux where the direction of the transect is arbitrary, the y component of the velocity has to be included into the equation, as well as the y component of the cross sectional area. In practice, the above integral is implemented in a discrete manner based upon the numerically simulated data from TELEMAC-2D.

7.2.2 How it is currently calculated

Currently, ocean science researchers use the TELEMAC-2D output data and either:

- run this through a custom MATLAB script that calculates the flux between two pre-defined points within the mesh: this process of retrieving the flux can take a couple of hours depending of the size of the data. (Until very recently, this was the only means of calculating the flux).
- Use a new multi-step manual process in Blue Kenue, the post processing and visualization system for TELEMAC-2D, which is reviewed in section 5.1 of this thesis. Dr Peter Robins, who developed this procedure, describes it:

“Using Blue Kenue, I add a transect, create a time series along the line of velocity and the concentration I am interested in (e.g. water depths), then multiply the two time series (velocity*water depth). Finally, I have to integrate the new time series along the transect to calculate the flux. This line then needs to be summed at a time period to calculate net flux values. The transect has to start/end within the model domain (i.e. not on land) and the intervals along the transect have to be manually assigned.” Dr P.Robins.

Neither process is ideal: the time consuming nature of the MATLAB process has meant that researchers have tended to minimise, rather than optimise the number of calculations along the estuary, as demonstrated in Figure 7.2, where only 5 transects per scenario are used, and thus may possibly miss important comparisons. The new Blue Kenue system is also acknowledged to have limitations. Again, referring to

the conversation with Dr Robins, he says: “However, it’s not straight forward and a certain amount of processing needs to be done.”

This processing involves a number of manual computations, and the whole procedure is certainly not instantaneous, whilst not providing the same level of functionality as the tool developed as part of this thesis.

The flux calculator proposed as an important element of this thesis derives the flux across a transect from interactive user input, using a custom algorithm. Not only does this provide researchers with a swift measure of flux, but the derived flux data is passed back into the database for storage, thus completing the iterative cycle of visual analytics outlined in Chapter 4. This new data can then be visualized and analysed alongside the raw data. The derived data visualization is presented in the linked Coordinated Multiple Views, where the researchers are able to study the derived data further through the different views and plots, including a PCP, to observe how the patterns are located spatially and non spatially. Such a close integration of the data with the visualization and analysis, enables the user to quickly drill down and investigate different areas of interest within the scenarios.

7.2.3 Flux in other work

The related work provides scant evidence of visualization and visual analytics solutions for the calculation of flux (whether it be tidal, sediment or salinity flux). These examples are discussed in detail in Chapter 4. However, Cotter and Gorman [65] should be mentioned here as their algorithm for calculating flux directly from unstructured data (although for deep ocean convection) informed the development of the flux tool in this work.

7.3 Architecture: tool components and developing the tool

The overall aim is to provide oceanographers with a tool to simulate their field practice by providing transects and calculating fluxes at multiple positions along an estuary or in a bay. This may be achieved in two distinct manners: firstly, the user might select each of the transects individually for investigation, but as with real surveying, this could become a time consuming process. The second method would be for the user to define a range/boundary for investigating the flux, then the

algorithm would generate multiple transects and fluxes through this region, which provides an automatic method to calculate several fluxes within a frustum.

With the overall aim in mind, the flux is calculated at every position along the estuary, which is a three stage process:

1. Calculation of the flux over a specific transect of an estuary: this process itself occurs in two parts: first the extraction of a transect (transect profiler), then the calculation of the flux across this transect.
2. Calculation of n fluxes along the estuary. A user interface is provided that enables users to define the centre of the estuary (or the path the fluxes will calculate across). The transects and fluxes are then calculated at regular distances along the path.
3. The data is stored to the database, for generation of the Transect/Flux View. This view enables the user to filter and explore the flux information, for subsequent analysis within VINCA.

7.3.1 Transect Profiler

The user defines the start and end point for the transect profiler, which then cuts a straight line cross section out of the mesh. Figure 7.3 depicts a typical transect cross-section. This transect line is then divided into m sample points, where m is determined by the length of the transect and a global resolution parameter (Gr). Therefore:

$$m = LengthOfTransect / Gr.$$

Gr is a relaxed parameter and is usually set at between $1m$ and $10m$, depending on the resolution of the underlying data (this is also user-assignable), such that a whole number of sample points between the starting and ending points of the transect is achieved. Barycentric interpolation is used to interpolate all the data values at each one of the sample points along the transect line.

Barycentric Coordinates are defined ¹ as: The mesh comprises a triangle (A,B,C), which defines a plane. If one of the points, A, of the triangle is taken to be the origin on the plane, all other positions are now relative to this origin. So coordinate values

¹<http://www.blackpawn.com/texts/pointinpoly/default.html>

can be given for any position on plane. Psuedo Code for Barycentric interpolation function is given in Algorithm 7.1

Algorithm 7.1 Psuedo Code for Barycentric Interpolation:

```

function BARYCENTRICINTERPOLATION( $\vec{A}, \vec{B}, \vec{C}, \vec{P}$ )
     $\vec{v0} \leftarrow \vec{C} - \vec{A}$ 
     $\vec{v1} \leftarrow \vec{B} - \vec{A}$ 
     $\vec{v2} \leftarrow \vec{P} - \vec{A}$ 
     $\text{dot00} \leftarrow \vec{v0} \cdot \vec{v0}$ 
     $\text{dot01} \leftarrow \vec{v0} \cdot \vec{v1}$ 
     $\text{dot02} \leftarrow \vec{v0} \cdot \vec{v2}$ 
     $\text{dot11} \leftarrow \vec{v1} \cdot \vec{v1}$ 
     $\text{dot12} \leftarrow \vec{v1} \cdot \vec{v2}$ 
     $\text{invDenom} \leftarrow 1/(\text{dot00} \times \text{dot11} - \text{dot01} \times \text{dot01})$ 
     $u \leftarrow (\text{dot11} \times \text{dot02} - \text{dot01} \times \text{dot12}) \times \text{invDenom}$ 
     $v \leftarrow (\text{dot00} \times \text{dot12} - \text{dot01} \times \text{dot02}) \times \text{invDenom}$ 
    if  $u \geq 0 \wedge v \geq 0 \wedge (u + v) < 1$  then
        return  $u, v$ 
    else
        return null
    end if
end function

```

▷ Compute Vectors

▷ Compute dot products

▷ Compute barycentric coordinates

▷ Check if point is in triangle

▷ Return the barycentric co-ordinates if inside

▷ Return nothing if outside

Once the all the sample points have been interpolated, this provides a transect composed of n segments where $n = m - 1$, see figure 7.3. From this piecewise description of the transect, other important quantitative metrics may be calculated, such as the length over ground of the transect, and the cross-sectional area of the water column (wetted area between the sea bed and water elevation).

The length over ground of the transect was calculated using Pythagoras Theorem iteratively through all $n - 1$ segments. The straight line distance d between each sample point and the interpolated height of the sea bed for each sample point h_i was known, thus the difference in height between each sample point could be calculated, and from this the length l_i of each segment between each point could be derived as:

$$l_i = \sqrt{d^2 + (h_i - h_{i+1})^2}$$

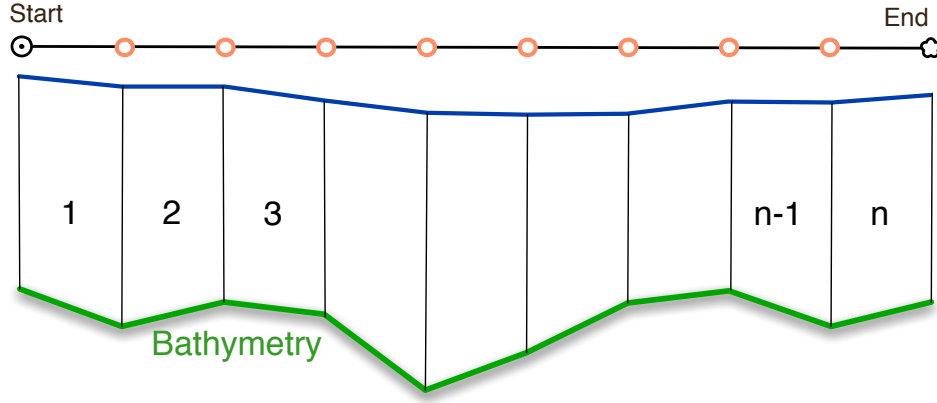


Figure 7.3: The Transect is divided into n cross sectional trapezoids

Therefore, the total length L over ground of the transect is given by a summation of the lengths from all n segments in the transect.

$$L = \sum_{i=1}^n l_i$$

The cross-sectional area is also calculated in a piecewise manner, by calculating the area for each of the n trapezoidal segments that are formed between the sea bed and water elevation (water surface). These are then summed together to derive the complete cross-sectional area of water between the sea bed and water surface (the water column).

The area of a single trapezoid is given as:

$$Area_i = ((h_i + h_{i+1})/2) \times d$$

$$\therefore TotalArea = \sum_{i=1}^n Area_i$$

7.3.2 Flux Calculator

The flux calculation is made along a user selected transect (cross section) of the domain, Figure 7.1. The transect profiler generates a transect, with interpolated

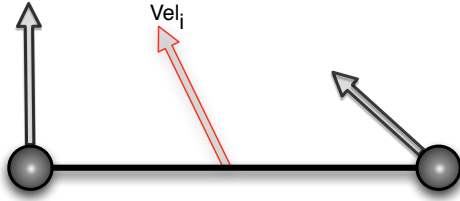


Figure 7.4: Illustrates the conversion from point based velocities to cell based perpendicular velocity.

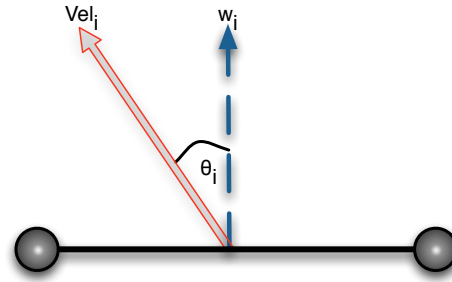


Figure 7.5: Illustrates the calculation of the perpendicular (normal) component of the segment's velocity.

points and other data such as cross-sectional area, which the flux calculator takes and uses to produce an instantaneous calculation of hydrodynamic flux for all time-steps to provide a temporal depiction of the flux across this transect.

To calculate the flux, the velocity, which is moving perpendicular (normal) to the transect, is multiplied by the cross-sectional area. In a piecewise method, this involves multiplying the cross-sectional areas of each of the n segments by their corresponding perpendicular (normal velocity), and then summing the individual fluxes to obtain the total flux across the complete estuary.

The velocity data is depth averaged, i.e. an average value for the velocity of the whole water column at that point, so it could be assumed that the velocity was the same at the water surface as it was at the sea bed (although this is not the case in real life). However, this assumption is routinely made in coastal shelf and estuarine modelling.

Calculating the depth averaged velocity vel_i for each of the segments is an average of the two velocities from the interpolated sample points that define the segment (see Figure 7.4). To calculate the flux, the perpendicular (Normal) component w_i of vel_i to the transect cross section is required, see Figure 7.5.

$$w_i = vel_i \times \cos\theta_i$$

where θ_i is the angle between vel_i and the normal to the transect segment.

The flux for each segment q_i can now be calculated as:

$$q_i = w_i \times Area_i$$

The total flux Q can therefore be calculated as the summation of the n segments' fluxes q_i .

$$Q = \sum_{i=1}^n q_i$$

This calculation can then be performed for all the time-steps to produce a time varying flux profile for the transect. See figure 7.14.

7.4 Flux along an estuary

Comprehensive analysis of all the fluxes in the coastal domain requires calculation of the flux at any transect through a continuous region running through the domain. An evenly spaced set of transects is required, following the shape of the estuary/domain, which are not too asymmetric. All of the derived fluxes from these transects will then be added to the database and subsequently visualized using the Parallel Coordinate Plots and other views for further investigations. This enables the user to analyse the data in a way that was previously not possible, and represents a significant evolution from the few available examples of flux visualization and analytics that have been described in the Related Work.

However, in establishing a requirement for a continuous region, there is a need to define this such that it fits the shape and nature of the domain. One method could be to take a transect across the mouth of the estuary and propagate this back to the start/riverine end of the estuary, to generate a region bounding the skeleton of the domain. However this is problematic for several reasons:

- the estuary / domain may have several tributaries (affluents) - so which is the start?
- it may be braided into very small channels, meander or break into islands (bars) and join back into several channels.
- Curvature of the channels usually makes one side longer than the other.

In addition, the sea bed shape changes with time: sediment, such as sand, gravel and clays are deposited and washed away creating a highly morphodynamic environment. At its current state of development, and because the need to prove the

concept took priority, the model does not include sea bed evolution. However, it is intended to include a sediment transport model into the system in the future, and to move away from the fixed bathymetry model. For this study a fixed bottom topography was utilised for every timestep.

The process of generating a region bounding the skeleton of the domain is difficult to achieve automatically. Other possible solutions could involve either hill-climbing or gradient descent algorithms on the bathymetry data to generate a skeleton for the domain, then querying the data to find points that reside at the tidal maximum (i.e. highest tide contour), for generating the boundary of the region. However, for the aforementioned reasons it was decided not to use these algorithms, and instead to have a user-defined skeleton and boundary.

Consequently, the user is given control over the area they wish to study, with the transects automatically calculated along a user defined skeleton path. This seems to be the most appropriate option, as the oceanographers will know where the real data transects were taken and, furthermore, they may wish to validate the visual tool/model with the real world observed data. Therefore, a semi-automatic approach is taken, which allows the user to define key locations which are used to calculate the fluxes. This process is divided into five steps:

1. The user defines the skeleton, using a point and click strategy.
2. Generate rectangular boundaries around each of the segments making up the skeleton.
3. Merge and then Edit boundary joining points to create a piece-wise trapezoid region. It is also at this stage the user has the option to move the boundary points in a constrained manner, if the boundary was not correctly fitted.
4. Calculate intermediate starting and ending points along the boundary to give a set of sample lines.
5. Generate Transects and Fluxes for each of the sample lines in the region.

Steps 1-3 happen iteratively as the user defines the skeleton, and then stage 4 and 5 occurs when the user decides that the skeleton and boundary fit and wishes to generate the result set.

Stage 1: The user generates a skeleton by clicking in key locations in the main view that will define the shape of the domain. This creates a desired piecewise-linear skeleton which the transects will occur along. Figure 7.6.

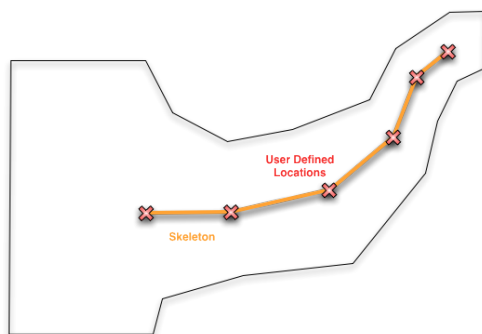


Figure 7.6: the skeleton as defined by the user.

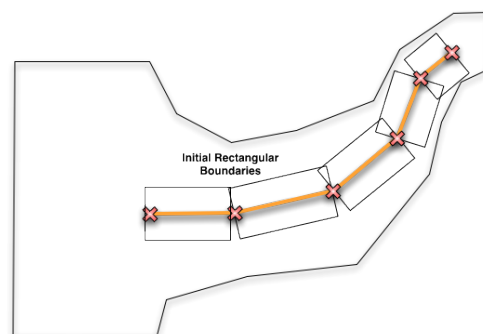


Figure 7.7: initial rectangular boundaries placed around each of the piece-wise segments of the skeleton.

Stage 2: This process involves generating rectangular boundaries around each one of the segments that makes up the skeleton. This is an initial step to align the boundary with the direction of the transect. See Figure 7.7. The rectangular boundaries are the length of the skeleton segment, and are initially setup with a constant width that is determined by the overall scale of the spatial dimension of the dataset.

Stage 3: The edges of the discontinuous rectangular boundaries need to be adjusted such that a joining section can be merged to form a continuous bounded region. This is done by averaging the orthogonal components to obtain the average angle at the joining ends of adjacent rectangular boundaries. See Figure 7.8.

The user can continue adding to the skeleton until the shape of the domain has been fitted (as shown in Figure 7.9). Once the user is happy with the skeleton, the envelope (bounded region) can be edited by moving the boundary control points (these are depicted as red circles in Figure 7.10).

Stage 4: Now that a bounded region has been defined, the intermediate sample points that make up the sample lines need to be interpolated, see Figure 7.11. The intermediate points given are automatically calculated, by a user-defined constant Cd , but this has two contradictory constraints. Whilst wishing to allow the user to define Cd , and so to have the transects evenly spaced, there is a need to constrain some of the transect locations to the key user-specified positions. Consequently, the constraint of the distance factor Cd , is loosened and the maximum quantity of transects is evenly spaced out between the control points.

Stage 5: This is the process of generating the transect and flux for each of the intermediate sample lines. It follows the method devised for a single transect,

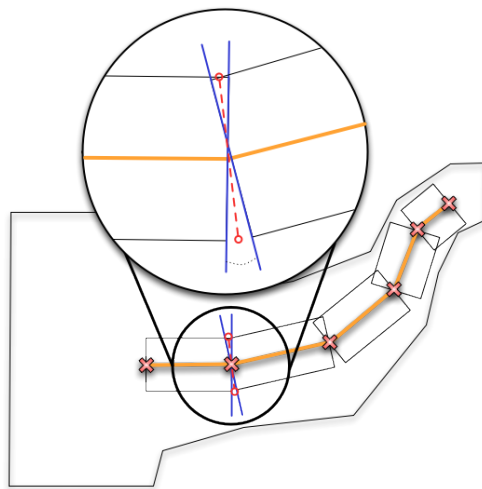


Figure 7.8: the joining of the two boundary ends to form continuous region.

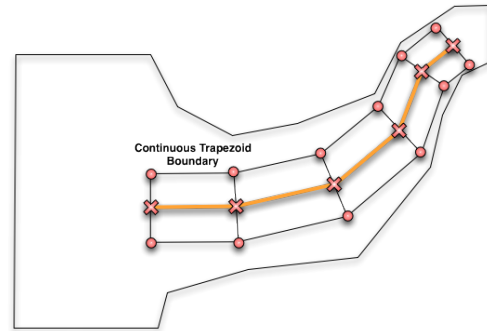


Figure 7.9: T showing the trapezoid boundary.

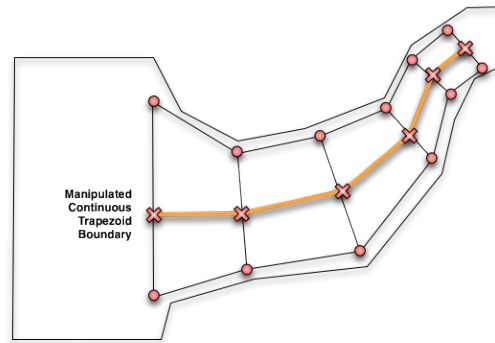


Figure 7.10: boundary after user manipulation

but instead of having user defined points, it uses the intermediate sample points as the basis for the starting and ending points of the transect. The transect therefore becomes each intermediate sample line. Once the transect has been generated, the flux data for it can then be calculated for all time-steps. Whilst iterating along the path (through all the sample lines), the transect and flux data is stored to the database for visualization and analysis.

Taking the user-defined skeleton as depicted in Figure 7.13(i) the results of the transect generation are shown in Figure 7.13(ii). This creates a suite of profiles and flux that are depicted in Figure 7.15.

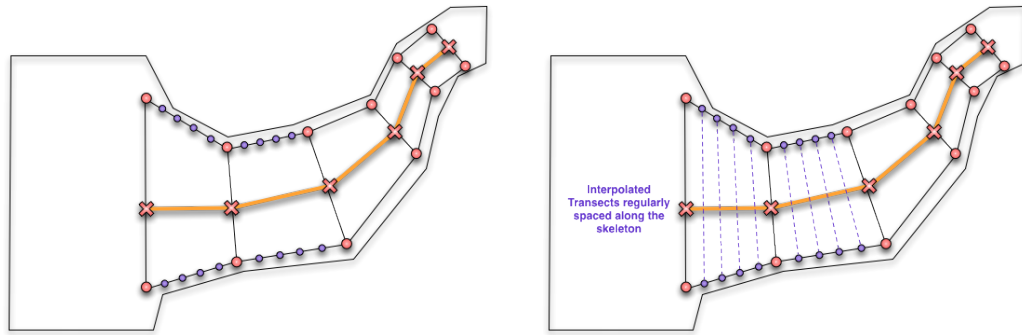


Figure 7.11: the interpolated points that will form the starting and ending points for the interpolated transects. Figure 7.12: the interpolated transects the algorithm will generate.

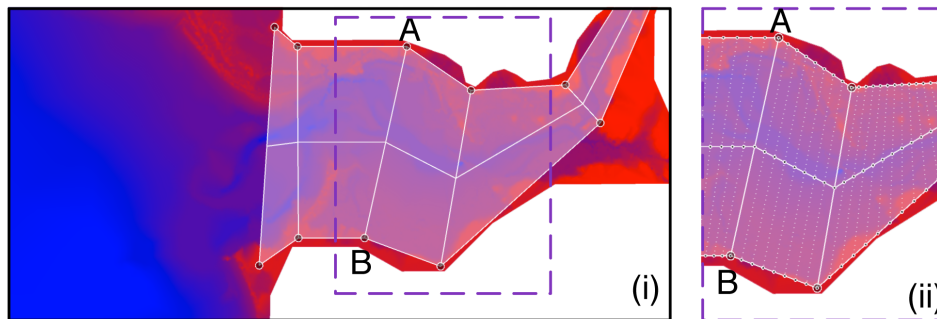


Figure 7.13: Diagram (i) Shows the Burry estuary with the region transect/flux before calculation. Diagram (ii) demonstrates the result from the algorithm, with all the transect positions calculated. It shows the regular distribution of the generated transects.

7.5 Visualizing the Transect and Flux

In the case of a single transect and the region based transect, both the transect and flux are represented with a line graph. With the transect, this is a 2D cross sectional view, and with the flux graph the data is displayed as a time varying flux curve, see Figure 7.14.

With the case of the transect profile it is also possible to encode other information into the visualization, such as the tidal range, the free surface range, and possibly in the future, the bed evolution range. The other information is encoded by allowing the user to display an envelope for both the water elevation and free surface. The actual value at the current time-step is still also displayed.

The region based view has all of the possible transects and fluxes composited together to show the range of values through the whole region, it is also possible to

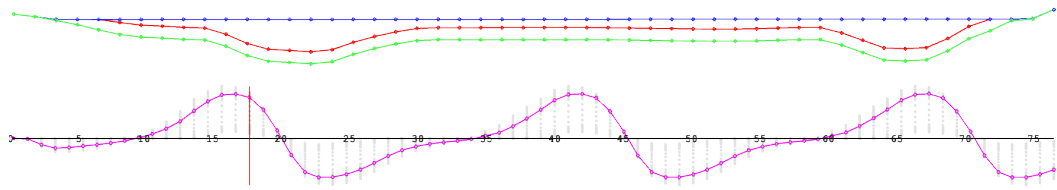


Figure 7.14: shows the publishable output from VINCA of a simple transect profile and associated flux curve.

focus in on one transect / flux, either through selecting it in the main view, or by using a slider attached to the view. This also updates the main view by highlighting the current transect within the range, see Figure 7.15.

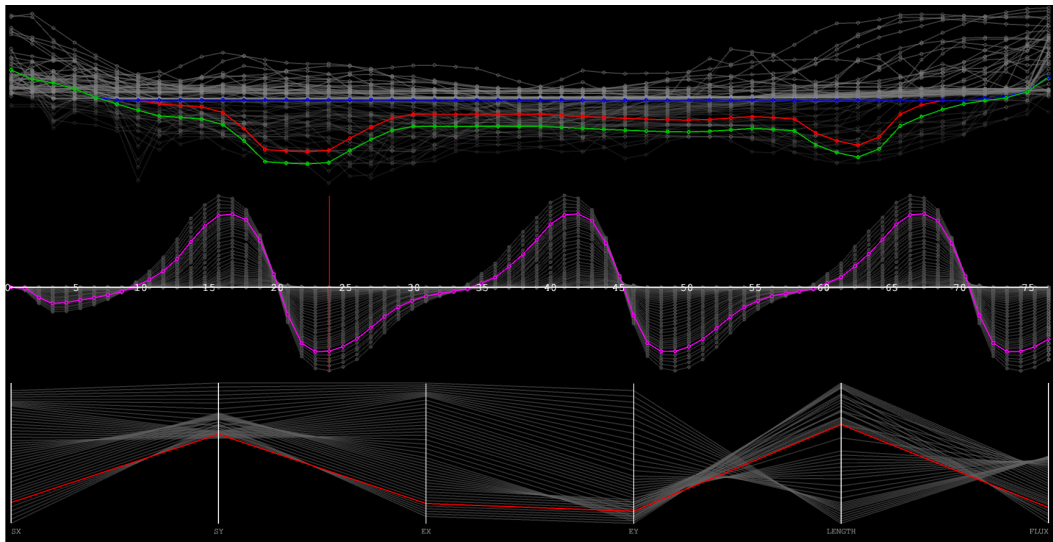


Figure 7.15: shows the region based transect / flux view in VINCA.

The transect partition has all of the same functionality as the single version: the ability to show the tidal ranges etc, but in this case it is limited to the focused transect, as otherwise it would have caused too much visual clutter.

As the region based view had to deal with multiple transects of varying length and in different spatial locations, this provided a number of interesting display challenges: 1. How to scale / map each transect into the display panel / output canvas. 2. Positioning – where to locate the scaled transect within the plot (i.e. along the horizontal or vertical axis).

The region based view also has an additional tool, which the single view does not have, which is an additional traditional PCP. This gives users an holistic view of the transects and fluxes, allowing them to compare different transects more easily.

See figure 7.15. To give the user a spatial reference within the PCP, an extra axis is used to define the position on the skeleton on which the transect resides, starting at one end and ending with the other. The starting point (x, y) and ending point (x, y) were originally used as the spatial reference, but found that this caused too much visual clutter / over plotting and just confused the matter.

Other axes within this PCP include: Length of Transect, Length over Ground, current flux (time varying), running net flux (time varying), net flux, tidal range and free surface range. As with the main PCP view the transect / flux PCP (TF PCP) reorders the axes by drag and drop, but in this case they are not grouped, as with the main PCP view. Range sliders are also provided so that the user is able to query the transect / flux PCP by brushing. The results of these queries can be displayed by highlighting, using coloured lines in the transect / flux plots and TF PCP. Another display method is to filter the data, by only drawing the data that is returned from the query, this has the effect of reducing the number of lines.

7.5.1 Scaling Challenge

As the length of each of the transects was highly variable, there were multiple ways the transects could be scaled / mapped into the display panel, each giving the user a different method of viewing the data. The methods of scaling currently implemented include: 1. Individual Scaling - all transects are scaled / mapped individually into the display panel. 2. Normalized scaling - all transects are scaled / mapped into the display panel by normalising their length against the longest transect in the region.

7.5.2 Positioning Challenge

The positioning challenge arises out of the scaling challenge, and the positioning of the plot is inherently linked with its scaling. In the first case of Individual scaling, the position challenge is extremely simple: the plot is simply positioned at the top left hand corner of the display window. With the second case, that of normalised scaling, this is where the transects are positioned around their centre lines, such that the longest transect covers the complete width of the display panel, whereas all smaller transects are located within this range.

7.6 Evaluation of Techniques

The technique outlined in this section provides a rapid, but rigorous method of extracting and analysing information that has been derived / calculated from the underlying data, and allows for users to drill down into new information that is inaccessible or slow to retrieve using traditional methodologies. For example, the human-in-the-loop specification of the flux envelope and the automatic creation of multiple fluxes has led to faster analysis. This has significantly speeded up an important knowledge discovery process, whilst also allowing ocean science researchers to perform ocean science research instead of having to solve software engineering challenges.

The results generated by the tool in both the single and region based methods provide an output that is of publishable quality suitable for screen and print display. Furthermore, it is believed that this method within the VINCA system is likely to provide the quickest and neatest solution to extracting multiple transects and fluxes from TELEMAC-2D data. The method has been tested by the ocean scientists and has been proved to provide quicker results than their current methodologies.

Developing this tool has proved the iterative visual analytics paradigm at the centre of this thesis, based on derived data, is achievable and works. Users have studied and selected an area of interest that they would like to investigate further, and have been provided with a set of tools that will enable them to gain insights that would be difficult to achieve with the raw data alone. Equally, the importance of working collaboratively with domain researchers to identify what they require to be able to function more efficiently as researchers has also been demonstrated. What has been developed is a tool which is both useful and usable, and will significantly improve their capabilities in relation to a key aspect of their research.

With this in mind, it follows that other tools fitting the same paradigm are necessary to enable even more detailed analysis of these complex and large data sets, as they may enable the rate at which knowledge may be processed and obtained to be enhanced. The current flexible VINCA framework allows for new tools to be bolted on and used within the visual analytics process and a number of ideas have already been explored with the ocean scientists. One example is a Tidal Ellipse Parameter calculator and analyzer, which, in the same spirit as the flux calculator, would add functionality not easily achieved with their current tools. It would add to understanding of tidal aspects of the data, and furthermore, underpin a fluid flow analysis tool to improve the fluid flow visualization capabilities of VINCA.

These reflections are supported by the Standard Usability Survey (SUS) study conducted with ocean science and computer science researchers. Full details of the study and the outcomes are presented in Chapter 8, Discussion and Conclusions.

7.7 Summary

The chapter describes and discusses one of the major contributions of this thesis: the development, implementation and use of a novel tool comprising a transect profiler and flux calculator. This serves as a case study for several important aspects of the thesis: it is an exemplar of the benefits to be derived from a collaborative approach with ocean scientists encompassing analysis of user needs, enabling a suite of customised visual analytics tools to be developed which meet their specific research requirements. These include the production of derived data which supports an iterative approach to analysis.

The importance of calculations of flux are discussed and a definition of tidal flux is presented: it is regarded by ocean scientists in general as a very important measure, and for estuarine scientists, tidal flux is a key measure in understanding their predictive, modelled flooding scenarios. Whilst numeric models provide many useful variables supporting prediction, there are other calculations and analysis which require to be derived as part of the analytic process. Flux is one of these. Not only is it important in its own right, but its calculation enables a further derived measure to be calculated, tidal prism. Existing methods of calculating flux are discussed and found to be slow, inefficient and complex. However, research demonstrates that it is a problem which has been little addressed through a visualization approach.

The standard range of techniques, plots, graphs and views which have come to be associated with visual analytics tools would not be able to provide flux calculations. Thus, custom built visual analytics components/views were required. Calculation of flux requires cross-sectional profiles (transects) of the estuary and the chapter describes algorithms for interpolating points across the estuary, selecting an interpolation technique dependent on the data type, and for calculating the flux.

The tool permits the extraction of n transects across the estuary, and thus n flux calculations and their analysis through a three stage process:

1. Calculation of the flux across a specific transect.
2. Calculation of n fluxes, through a user interface which enables the user to define the path by which the fluxes will be calculated.

3. Storage of the calculations to a database for exploration and analysis using the VINCA visual analytics tool.

The tool confers a number of benefits: it is able to provide a rapid, and rigorous method of undertaking a traditionally lengthy and complex calculation, providing virtually instantaneous results; it supports iterative analysis of the derived data permitting researchers to drill into their data and its ease of use easily facilitates calculations, whereas the computational complexity of the traditional methods had been an inhibiting factor, even though the dynamic estuarine environment required significant numbers of calculations. The flux calculator thus proves the visual analytics paradigm at the heart of this thesis.

CHAPTER 8

Discussion & Conclusions

This chapter evaluates the research in great detail, examining the outcome of each of the four aims identified in conjunction with the hypotheses. It is divided into five parts: Review of User Survey; Review of Hypothesis, Summary of Achievements, Future Work, and Final Conclusions.

8.1 User Survey

The discussion, evaluation and conclusions which follow is informed by a SUS study, which provides feedback from ocean scientists and computer scientists as to the level of success of VINCA's visual analytics system in solving some of the data analysis challenges confronted by coastal shelf and estuarine researchers.

The user survey can be found in appendix 1.

The evaluation was conducted in the laboratory, and on a provided laptop. This meant that users did not need to setup the software, nor import the data. Although neither of these tasks are complex, they are time consuming and it was decided that they were not needed in the evaluation phase.

There were three parts to the evaluation process: (i) introduction and background, (ii) tutorial and tasks, and (iii) post evaluation questionnaire; detailed as follows:

(i) Users were welcomed to the study and the evaluation process was described

to them. Each of the users was given an evaluation pack, this can be found in Appendix 1. Users could take away the pack and they detached the consent form for our records. A brief introduction to the visualization tool VINCA was provided as well as a short description of oceanography and the movement of water on an estuary (for the benefit of computer science reviewers). The users then completed the consent form, detached the form from the pack, and moved to the tutorial part.

Users were led through three sets of tasks. Each of these tasks required the user to operate increasingly more complex features of VINCA. The first set of tasks enabled the users to open a window in VINCA, move the window, make it larger/smaller and close it, load specific data, display the data in a map plot and manipulate the colour map of the displayed data.

The second set of tasks led the user to operate the exploration and query functionality of VINCA. They were given a task to analyse the deep-water sections of the dataset, and see how they compared to shallow water. Finally, the users were asked to perform a flux calculation that specifically included the data from the area of interest which the user had focused on in the second task.

(iii) Finally users were asked to fill in a demographic questionnaire, a set of questions on usability and write some open-ended comments.

8.1.1 User Survey Results

This evaluation represents an initial pilot study, and currently additional users are being sought to perform a much larger study. Four participants completed the pilot evaluation: two were ocean science (OS) researchers, who are simulating and investigating flooding events in an estuary. Two participants were visual analytics / visualization (CS) experts, with expertise in coding visualization tools. The following details the questions and answers to the questions:

Describe some positive aspects of VINCA:

CS-01 Nice functionality for manipulating specific values and seeing the result, intuitive way to filter the map and look into the waters behaviour.

CS-02 The system is extremely well implemented with clear visual links and an easy to use workflow embedded within. The application is well thought out with the original selection of brushed data integrated into each feature perfectly.

OS-01 Transect Visualizations. Flux Visualization. Independent controls for each figure. Easy to use colour scaling (mapper).

OS-02 Speed of post-processing. Computation of flow-visualization. New post processing features such as classification of statical type analysis of data.

Describe some negative aspects of VINCA:

CS-01 I generally prefer lighter colour schemes, Title font could be clearer. Radio Buttons need to be smaller.

CS-02 No negative aspects

OS-01 Cannot produce animations, Black background is a bit depressing, Cannot delete figures – n.b. this is a bug that needs fixing.

OS-02 Some modellers like to have complete control on data post processing via programming like MATLAB. I didn't feel that I have full control of the data processing.

Briefly describe any type of enhancement that you would like to see in VINCA:

CS-01 3D model of the sea bed and water volume

CS-02 Undo operations for the FLUX, bigger screen for visual analysis (perhaps a power-wall)

OS-01 Grid generation

OS-02 Publishing tool (print and plots into a report). It is a useful idea to apply it to NetCDF format which is used by many models like SWAN, DOM, ROMS ...

Explain whether you think VINCA will be a useful tool for aiding coastal oceanographic researchers knowledge discovery?

CS-01 Indeed it will be helpful – Viz tools are much faster in allowing analysts to see effects / behaviours it is a matter of speed, compared to alternatives.

CS-02 A variety of features allow the scientist to dig deep into the data which will surely help with knowledge discovery. From the start specific data can be brushed into ranges the scientist is interested in, whereby multiple visual analysis options are provided from this brushing. These multiple visual depictions of the data will surely aid knowledge discovery of scientists as it provides many forms of visual analysis.

OS-01 Definitely. The flux / transect options are very useful for visualizing 3D data in time i.e. 4D.

OS-02 I think it is a good idea but it needs a team to make a professional tool. It is good at this stage and has potential.

What is your opinion of using an interactive visual oceanographic analytics environment, such as VINCA:

CS-01 It felt that it can be a powerful tool, helped me understand a few things about tides.

CS-02 It's great fun to use and rewarding to answer questions that you have about the data.

OS-01 Well I already use one regularly – Blue Kenue and also MATLAB. BK has many advantages, but also some significant shortcomings that VINCA can address. If polished up, VINCA could complement BK and MATLAB, maybe one day supersede it.

OS-02 It depends, many like it. I personally prefer a tool which has programming features like Matlab to control data. But it is good for some purposes

What is your opinion on the future of interactive visual oceanographic analytic environments?

CS-01 Interactive vis tools will be essential in the future for analysis. The familiarity of humans in reading maps, coupled with the power of data analysis and interaction make such tools intuitive and soon essential.

CS-02 If more ocean scientists learn computing and build and use tools such as these, then I'm sure the domain will benefit greatly.

OS-01 There are many improvements that can be made such as performance, data handling capacity, data manipulation, exporting methods.

OS-02 NO USEFUL ANSWERS HERE

8.1.2 Discussion of the Results

There are certainly limitations with this pilot study, particularly because of the small number of participants, and a much larger study is needed to evaluate the techniques and benefits of the research ideas in VINCA. It would be ideal to incorporate VINCA into the data-to-day workflow of the oceanographers, and while this is planned for future work, it was impossible to perform such an evaluation in the duration of this thesis.

However, the results from this initial evaluation are positive. Generally users were extremely excited over the tool and, as can be seen from their comments above,

they were keen to know more about VINCA and see how it can be used in their day-to-day workflow.

It is clear, just from the comments, that VINCA does enable certain visualization commands to be performed faster and more efficiently, that users can locate specific points of interest and explore them further, and gain quantitative results on their data from the visualization.

Some of the comments are fairly minor, and matters of personal preference, and many of the suggestions for improving the system have already been identified and referred to at various points in this thesis. Time, however, precluded the implementation of some of this additional functionality, for example 3D, building a report containing a number of views, export facilities.

It is also clear that other tools (such as Blue Kenue) have additional functionality. This is not surprising, as much effort, money and time has been given to Blue Kenue as a commercial tool.

What was interesting was the comments from ocean science users, which bear out many of the conclusions about the study of the way ocean science researchers work, conducted at the beginning of this thesis. There is a dichotomy: whilst accepting the value of a system like VINCA, there is a mistrust of a tool which requires them to lose some control of their data – yet systems like MATLAB, which provide the control, are not easily able to provide the same level of bespoke functionality as demonstrated by VINCA.

It is a matter of familiarity and education, but in fostering collaborative working between the two domains, through the continuing development of useful tools, then there is no question why systems such as VINCA cannot become a tool of choice in the coastal shelf and estuarine domain. However, this pilot study does demonstrate the potential of VINCA, and its novelty in visual techniques and exploration methodologies.

8.2 Review of the Hypotheses

This thesis developed out of a real sense of opportunity: that ocean scientists were not making the best use of the rich variety of visualization and visual analytics tools available to support advanced data exploration and enable knowledge discovery. Thus, they were handicapped by a limited view of the role of visualization simply as the means of displaying their data. Throughout this research there has been considerable emphasis on the aspiration to achieve a significant enhancement of the

visualization and analytic tools used in studying oceanographic data, through an effective collaboration with ocean scientists, derived from a personal understanding of their research priorities and methodologies. Achieving this, it was hoped would enable a framework to be established, which might form a paradigm for the development of visualization and visual analytics systems to inform other computer and ocean scientists undertaking similar collaborative projects.

It was clear from the outset that there were limitations in current techniques in coastal shelf and estuarine simulation studies. Therefore the overarching goal of the thesis was to develop novel analytics and data exploration techniques and tools in the extremely complex domain of modelled estuarine hydrodynamics. However, it was equally clear that this is a challenging area, which suggests why (in common with coastal shelf studies in general) this seems to be a sparsely studied area, with less published research than would be anticipated in a domain of such environmental importance.

This thesis demonstrates that visual analytics tools can be built for the field of estuarine hydrodynamics and that considerable opportunity for further research still arises. Indeed a comprehensive visual analytics tool was developed (VINCA). The tool provides much functionality to demonstrate the hypothesis: it includes the generation of multiple views and is capable of generating the visual results in publication quality output, it offers effective data management processes, is fast (relatively), and provides new interactive data exploration and analysis capacity. However, it is accepted that areas of work remain to be done to transform VINCA into a fully formed system, which would be the tool of choice for ocean scientists seeking to analyse complex modelled hydrodynamic data. This chapter therefore includes a road map for its further development and enhancement, as well as wider considerations as to how the paradigm might be extended to related domains.

Personal understanding of the physical oceanography concepts and theoretical underpinning has supported day to day verification and evaluation of the system's implementation and refinement, through a knowledge of what the expected outcomes should be.

8.3 Summary of achievements

In Chapter 1, the four propositions of the hypothesis were set out, together with the associated aims. These are discussed further within this section. For ease of reference these are re-stated below:

- to develop a visual analytics tool (framework) for Coastal and Estuarine studies;
- to allow derived (new) data to be made available to the domain scientist for analysis;
- to generate different views of the data including an output rendering in a vector format to support viewing and study of the visualization at any resolution;
- to use Computer Science methodologies to enable rapid and effective visualization and analysis of the data.

8.3.1 Achievement 1: Development and implementation of a coastal oceanographic Visual Analytics framework

A framework for visualizing and analysing certain aspects of the hydrodynamics of coastal and estuarine regions has been developed. It allows for domain specific visual analytics tools to be plugged into a coordinated multiple view environment and for detailed, iterative, study and analysis to be undertaken. Chapter 5 provides full details of the development of the framework whereas chapters 6 and 7 contribute evidence of the analytic functionality of the system.

The development of the framework was central to achievement of all the aims of the project: aims two, three and four depended on a well researched, designed and appropriate interactive visualization system, and visual analytics platform. Then, the necessary tools to provide advanced analytics and derived data for further calculation might be introduced and tested. Although very large, the size of the datasets did not preclude the use of a pipeline, and having decided on this structure, it was essential first to develop a fully functioning, interactive visualization system and appropriate interface which would effectively support the required analytic tools.

However, the components of the system could not be considered in isolation of each other: the research had demonstrated that the proposed techniques and tools for the visual analytics system in themselves would make computationally intensive demands on the system, for example QDV (query driven visualization), the linked coordinated multiple views and the PCP. Thus, it was impossible to consider the visualization system and visual analytics interface in isolation of considerations of what the analytics tools might be.

This made the research phase of the project particularly important: not simply the research into the attributes and challenges presented by the data and the numerical model, the desired insights required of the simulations and the various options for the visualization systems and visual analytics tools, but also the collaboration with the ocean scientists to establish what they required to improve their analytical capabilities. It was an object lesson in the validity of the paradigm of the close coupled system, essential from the development stage onwards.

The research and review phase of the project were instrumental in the achievement of the aims and proving the hypothesis, such that the research and collaborative approach with coastal shelf and estuarine researchers are regarded as among the major contributions of this thesis, as set out in Chapter 1 (Contributions 1 and 2). As far as can be discovered, no similar, equally comprehensive review of oceanographic visualization and visual analytics has been undertaken, particularly one which focuses on coastal and estuarine modelled data; which seeks to understand the challenges and difficulties the domain presents, and which also aims to set the state of maturity of the discipline into context with other, closely related geosciences. Chapters 3 and 4 detail these studies.

Furthermore, if a framework was to be produced to inform computer and ocean scientists in developing their own systems and tools, there was a need for a knowledge based paradigm, built on a good appreciation of current practices and an understanding of the challenges presented by the data. Thus, in Chapter 4, requirements for an interactive visualization system are identified and attributes of a visual analytics approach are proposed, together with suggestions for appropriate methodologies, tools and techniques.

It is clear from this research that while other researchers have developed visualization tools, and visual analytics systems are starting to be used in oceanographic visualization, there remains much work to be done, and many opportunities for research to be undertaken in this area.

Through Chapters 3 and 4, together with the initial prototyping in Chapter 5, it was established that the main requirements were for a fast, interactive visualization system and a visual analytics interface based on linked coordinated multiple views, because the needs of the data required different approaches to enable researchers to fully understand the complex relationships and correlations of their multi-field nature. The main challenges lay in actually achieving a system which provided a stable, reliable platform; fast rendering and views, plus plots and graphs which the researchers, unused to such methods of interacting with their data, would find

intuitive, and would be encouraged to explore their data in different ways.

The prototyping, described in Chapter 5, effectively identified the main issues, which would have to be dealt with in successfully producing a framework. These centred around the size and complexity of the data which resulted in extremely slow rendering; its unstructured multi-resolution nature; massive overplotting and coincident topology. Thus, the system had to be capable of handling these problems, otherwise it simply would not be usable or useful. These factors influenced every aspect of the decision making and subsequent development processes. For example, the choice of Processing, a low level graphics tool, was made after prototyping demonstrated the limitations of toolkits or modular visual programming tools, as it enables different renderers to be used.

Much time was devoted to developing appropriate methods of scaling, aggregating, abstracting and otherwise filtering the data at almost every stage of the pipeline. This required a methodology which would be able to handle the essential characteristics of the data, particularly the spatial nature, whilst retaining its integrity; support interactive visualization and enable the identification of features of interest. Furthermore, any method had to be able to handle unstructured AMR data which are recognised as presenting particular challenges.

The development of a PR quadtree spatial indexing algorithm proved highly successful and its flexibility enabled it to be adapted for use in a number of variants throughout the pipeline, including its novel use in conjunction with the PCP (its specific use with regard to the views and the data management are considered in Sections 7.2.3 and 7.2.4 of the chapter). However, without an effective means of scaling the data at the pre-processing stage it would not have been possible to have created an interactive speed visualization system, thus no visual analytics, as the attempts to run the data without pre-processing through the PR quadtree demonstrated.

Successfully scaling the data enabled the use of linked CMV, which were regarded as essential to the visual analytics system because of the requirement to support multiple complementary views (Chapter 5). Reflecting the hierarchical structure of the spatial indexing, the CMV were linked through a hierarchical manager, as a simple list was found to be confusing in what might evolve to be quite a large and complex system, as further views and plots may well be added to meet the evolving requirements of researchers.

At this stage of the project, the range of views is limited to sufficient to prove that the system is able to operate at interactive speeds and to demonstrate that it is

capable of detailed, iterative study and supports feature identification, trend analysis, and enables relationships to be established. The work in Chapter 5.7, and testing the system as it was developed, demonstrated it was capable of supporting the required level of visualization and analysis. As such its contributions include new approaches for studying coastal and estuarine modelled data, through the linked CMV interface.

8.3.2 Achievement 2: production of derived data supporting more effective data analysis

A tool for deriving data was developed, enabling ocean science researchers to be more efficient and less ad hoc in their analyses.

The derivation of data is regarded as one of the main novelties of the system, which is achieved through the Flux calculator (Chapter 7) particularly. This permits new data to be derived, and subsequently visualized and analysed. It gives added value to the system and enables a richer knowledge discovery. Without the tools described in Chapter 8, ocean scientists would not be able to make their calculations other than by extremely complex methods.

Chapters 6 and 7 both demonstrate the calculation and use of derived data. Within Chapter 6 (PCP), a small amount of derived data is provided to the user - the 1st and 2nd time derivatives of the scalar variables can be displayed, when grouping the axes. This has not yet been developed to its full potential, simply sufficient to prove the potential for the use of the PCP with this data, but in consultation with ocean scientists, it would certainly be possible to include further statistical measures, dependent on their requirements. But in considering flooding scenarios, and the prediction of extreme events, further statistical derivations are likely to be useful.

Chapter 7 provides two of the most significant contributions of this thesis. Not only does it demonstrate a novel analytical tool, but it also demonstrates the feasibility of generating derived data specifically for the coastal zone domain, and the development of a custom built tool to aid estuarine researchers in complex, but highly important calculations. The successful implementation of the transect profiler and flux calculator (including the region based method, providing a near continuous method of getting the flux at any point in the domain), which also allows for clear and accessible results for the Ocean Science researchers, provides a strong case for further tools to be developed in this manner.

Scientists are now able to understand tidal flux at any number of transects across the estuary, something previously not possible. Current methods are ad hoc, where scientists calculate a few fluxes at specific points along the estuary and they also need to estimate where the best location of these transects are. Now they can calculate all the fluxes and plot graphs to thus demonstrate which is the best position. This adds significant benefit and makes for much time saving for the scientist.

This also demonstrates the value of collaborative working between computer and ocean scientists, where the combined knowledge and skills are able to develop a tool which is fit for purpose and provides results which significantly outstrip previous methods of making the calculation, in speed and quality of output.

8.3.3 Achievement 3: Implementation of a visual analytics system based on multiple views

A visual analytics system was developed based on CMV (coordinated, multiple views) which presented complementary views, plots and graphs of the data, with an output enabling publishable quality rendering.

Chapters 5, 6 and 7 demonstrate that a multiple view visual analytics system was built with a range of views which represented the attributes of the data, although it was accepted that there were many more novel ways of representing the data than are currently presented. The goal at this stage of the project was to demonstrate that the system was usable and did provide useful insight, whilst also showing the potential for the development of a full, rich set of views, focused to the specific needs of the research. For example, in studying how the researchers analysed their data, it was apparent that they need to compare how changing a variable might reflect in a greater or lesser flooding risk. Thus, it is recognised that for predictive studies the ability to closely compare different simulations, different runs of the same simulation, or simulations produced by different models would be helpful and a further enhancement of the system will include facilities for side-by-side comparison. Several view types were included.

Chapters 5, 6 and 7 show that spatial, temporal and holistic views were developed, both within the system and as high quality publishable output. Chapter 5 details the majority of views and their various capabilities: not only are there a variety of views but overlays which provide a different visualization of the same data, for example heat-maps and vector overlays. Other views could be imagined, such as a 3D visualization of the data. These are left to future work.

Chapter 6 demonstrates the ability to investigate the data, both temporally and holistically for trend analysis, by studying the data by means of three different PCPs. This is a completely novel tool for the ocean science researchers collaborators, but as a proven technique for spatio temporal data, particularly in meteorological studies, one of the aims of this research was to demonstrate its usefulness in this domain of study. Currently, the researchers have no means of viewing their data holistically and the PCP views enable them to identify areas of interest more easily for further detailed examination. Much thought and discussion with the ocean science researchers went into the facility to order the axes, to provide a logical ordering, which would support clarification of relationships rather than obscuring them, whilst also providing the researchers with the ability to manipulate the axes themselves, hence it was decided to group the axes by data type. However, possibly further development of the work might include different means of grouping and ordering the axes, for example using statistically derived quantities, and a comparison of the effectiveness of each in revealing trends, areas of interest etc.

Finally, the linked CMV interface with its hierarchical structure (Chapter 5) did provide a clear, user friendly, logical framework for enabling overview and detailed analysis of the data without being confusing.

The views in Chapters 5, 6 and 7 demonstrate that publishable quality views were produced. Although publishable quality views are desirable, and identified as important by the researchers, in view of the need to present their conclusions to disparate audiences, it was not regarded as the most significant element of the project, which was to prove the methodology for studying and analysing the data. However, various means of saving and exporting the output are included (Chapter 5) and a plan for creation of a report builder to customise a suite of views for individual reports has been developed.

Thus achievement of this aim supported a number of the contributions of this thesis by new approaches to the analysis of estuarine hydrodynamic data through a linked CMV structure.

8.3.4 Achievement 4: Data Structures

Computer Science methodologies were used to enable rapid and effective visualization and analysis of the data on a fast desktop computer. A major challenge of the project was the need to devote a significant period of time on resolving the many issues resultant from the nature of the data and also the computationally intensive

nature of the analytics tools. Throughout, there had to be a balance between implementing analytical tools which would provide a significant improvement over existing analysis methods, whilst achieving the required speed and quality metrics. As the focus of this research was on the visual analytics, means of optimising the system through overall architecture, hardware and interface solutions were not a priority.

Thus multiple, novel data-structures were used throughout the system, for overall data scaling and for enhancing the performance of the tools and techniques. These supported the achievement of rapid calculations. Production of a visualization using unprocessed data demonstrated the scale of the problem, and as has been mentioned earlier, spatial indexing through a PR quadtree was identified as an appropriate means of scaling the data at the pre-processing stage, and the reasoning for this is fully discussed in Chapter 5. Whilst other methods such as screen space algorithms were considered, the nature and attributes of the modelled data required a solution with proven effectiveness in dealing with unstructured, multi-resolution spatio-temporal data. The widely recognised challenge of accomplishing this suggested the need for a solution of already proven effectiveness for the data types. Hence, the novel aspects of this work lay rather more in developing an holistic, workable system, of demonstrable value in providing insight, rather than in investigating and proving a new means of scaling the data.

The PR quadtree proved effective in enabling fast, interactive visualization and, furthermore, supported the presentation of high quality views in its ability to deal with overplotted data. However, the addition of the visual analytics tools necessitated further data abstraction and aggregation techniques, as the tools themselves were subject to problems of overplotting, and also made computationally significant demands on the system, threatening its ability to maintain interactive rendering.

Again solutions were sought mainly through further variants of the static PR quadtree and bucketing mechanism used to filter the raw data, as it had proven its effectiveness for this task. Chapter 5 discusses the development of a novel, dynamic PR quadtree, which supports the zooming function for providing focus and context, enabling features of interest to be identified, studied and set into broader context. It also aggregates the data, by specifying level of detail to resolve the problem of the massively overplotted data in the vector overlay, as discussed in Chapter 5.

However, the main use of the dynamic PR quadtree is in enabling the PCP view to provide usable insight, rather than meaningless, massively overplotted axes, as described in Chapter 6, which had the potential to reduce the ability to undertake

rapid feature identification, or indeed achieve interactivity. This would nullify the value of the system. Original attempts to apply a frequency based approach to aggregating the data failed because of the use of a regular bin map which would not reliably support detection of outliers. Thus a novel solution through the dynamic PR quadtree was felt would provide a more reliable solution, which, on testing proved to be the case.

Together with the use of spatial data structures to resolve the challenge of selection of over-plotted data, discussed in Chapter 5, other hierarchical algorithmic solutions included the use of a hierarchical data structure to hold the axes for the PCP, which proved to be beneficial in manipulation and ordering of the PCP. A further use of data structures which proved effective was the hierarchical view manager. Storing the views in such a manner allowed for easier and better coordination between views, permitting local and global coordination.

Without the various data structure algorithms developed and applied, the raw data or non-aggregated versions of the system and tools demonstrate that it would not have been possible to have produced a system which would operate effectively on the standard, fast desktop computer typically used by the estuarine researchers. In using data space algorithms, the variants of the PR quadtree played a part in many of the contributions of the thesis and allowed the hypotheses to be proved. It was thus the major enabling process of the project.

8.4 Future Work

There is much work still remaining. This section is divided into three parts, in relation to the immediacy of the period. First considerations are made over what work could be done in the short period (i.e., the next year or so), then in the medium term and finally the long term goals.

Throughout the process of developing this research, the focus had to remain on developing a proof of concept, rather than a commercial system, and, thus, many ideas which merited further investigation and inclusion into the system were identified throughout the course of development and implementation (Chapters 5, 6, 7), and some are discussed in the sections above which evaluate the achievement of the aims.

8.4.1 Future Work – Short Term

There are several improvements that would be useful in the current implementations that would add much value to the current tools. These include the following:

Develop the 3D capability of VINCA. One of the main priorities is to develop the 3D capability of the developed tools. This would not be too difficult to achieve, and would be easily implemented as another view.

Improve the Flux calculator. The current Flux calculator enables the user to automatically generate multiple fluxes across the domain (at user defined points). However it would be useful to select how these fluxes are spread along the estuary in greater detail e.g. by using a curved (spline based) representation of the skeleton rather than a linear one. Therefore, a more functionally rich interface would be beneficial to the user. Again, this implementation would not be too difficult to achieve, and could be easily included in the current implementation.

Include a new analytical tool. A further tool identified by ocean science users is a *Tidal Ellipse Parameter calculator* and analyser. This would enable oceanographers to understand the tidal aspects of their data in more detail, and would form the basis for a more detailed fluid flow analysis tool to improve the fluid flow visualization and analytic capabilities of Vinca. These tools might be implemented immediately to increase functionality, as the sole reason for not doing so as part of this project was they are time consuming, although not necessarily challenging to develop.

Include further flow-visualization methods. The Related Work section steered the emphasis away from flow-visualization techniques, however a number have subsequently been researched and discussed with ocean scientist users. It would be useful to include these as either new views or overlays onto existing views (whatever is most suitable for the technique.) For instance, LIC techniques would be useful, along with ribbons and other abstraction techniques. In particular it would be helpful to include more analytics capability, so a user could select particular flows. For instance, it would then be possible to show all the regions where the flows are forming eddies, converging or diverging.

Investigate the further development of comparison. Another concept that would be useful to include in the short-term is to investigate how comparison can be

incorporated and used. As discussed earlier in this chapter, ocean scientists wish to compare between different runs or scenarios. Currently VINCA provides side-by-side comparison through the use of the coordinated and multiple views. However a direct comparative view would be beneficial.

Extend the User Base. Finally, in starting to think about the longer term Future Work aspirations, it would be useful to extend the user base. The plan is to encourage other domain scientists to use the tool, to generate a development schedule and identify additional views and tools to enhance the system further.

8.4.2 Future Work – Medium Term

The first consideration of Future Work is to transfer the ideas to other oceanographic fields, or other domains.

This research has pushed the boundaries of visual analytics for estuarine (and coastal) visual analysis, and has developed a VA tool for a specific domain of ocean science. However, VINCA has many features that could be applied to other oceanographic domains e.g. there are other oceanographic areas that hold similar data, such as fluvial and deep ocean. It would be easily feasible to adapt the current tool to visualize these different datasets. Fundamentally, this would mean that a more generic data reader would be required, and also that the main map view would need to be adapted to operate on different geographic locations.

However, in reality, little work would need to be applied to move to domains with much the same attributes, for example:

- Other physical oceanographic domains e.g. in support of offshore wind farm, fluvial, inland waterways and lakes, deep ocean, tsunami studies (combined geological, deep ocean and coastal shelf).
- Other types of oceanography e.g. Paleo, chemical, biological, geological, and marine engineering.

Furthermore, the system might be developed to support TELEMAT users in other areas of research, for example with HECToR (High End Computing Terascale Resource), the UKs national supercomputing service, which has been working to improve TELEMAT-2D performance, consequent upon the limitations imposed by the unstructured nature of the data [261].

From that point, many questions remain: what other domains would benefit? How could these techniques be transferred into other domains? What would need to change in the current implementation? Nonetheless, other earth and natural science disciplines suggest themselves, as their data are similar, for example:

- Geophysics – Tectonic Activity, paleo, seismic studies;
- Social Geography and Demographics;
- Atmospheric and climate studies including meteorology;

In addition, the techniques could be adapted to other non-similar (non-spatial or non-geographical) domains by using a hierarchical data structure linked with a spatial data structure could be adopted. For instance, it may be possible to use it to join Functional MRI scans (i.e., spatial data) with EEG type (temporal) data. Perhaps other types of chemical, biological or geological, and marine engineering would benefit from the ideas.

Specifically, the concepts included in the PCP could be adapted to other domains. There is much power in coordinating two different views together such that the user has a better understanding of the data, and also can operate the visualization more effectively. Consequently, it may be possible to adapt or generalise the PCP/PR-Tree arrangement to other domains.

The second Future Work proposal (in the medium phase) is to develop the concepts into a client-server system.

The modular structure of the system would enable the research to be easily set into a client/server arrangement, allowing VINCA to be operated over the Internet. Such remote operation may also permit users to work collaboratively. Obviously the exact method of synchronisation would need to be investigated; but the MVC pattern used in the CMV development affords extension to a client-server model. Similarly, the use of Processing would help to move the system to an App based approach, which may be useful in some instances (for example researchers working in the field), although probably a simpler cut/down version of the tool would need to be implemented for an App version, for instance.

8.4.3 Future Work – Long Term

Looking at the longer term is always difficult. However there are a few grand challenges that are facing the oceanographic community: including the manipulation of Big Datasets, Collaboration and Expressibility.

Big data: the investigation of Big Datasets is not only a challenge directed to oceanographic research. Indeed, space sciences, security, medical, banking and finance, commerce, are all areas where big data research is occurring. However, if the ocean scientists wish to answer some of the questions such as predicting weather or climate change, flooding events or tsunamis then they need to work with big datasets. However, many of these in the world are sparse datasets. For instance, recommender systems for purchasing habits are sparse, whilst oceanographic datasets are dense. This then adds additional challenges to how the information is aggregated and abstracted. Thus many challenges still remain: including manipulating several datasets that exist over several runs, and also incorporate real data along with derived and simulated data. Cloud based visual analytics tools may help here, to enable these big datasets to be shared easily and investigated anywhere in the world with very little computing power locally.

Collaboration: this is another issue. Most modern research is collaborative, and likewise so is oceanographic research. In particular, there are a number of big oceanographic projects that are working with researchers (and data simulations) from around the world. These are funded by several governments and are trying to address big questions. For instance, the Global Oceanographic Data Archaeology and Rescue (GODAR) and the World Ocean Database Project aims to gather historical and current oceanographic data in one place. However, it remains a challenge to integrate all the different real (sampled) data, along with the simulated data, in a way that several scientists can collaborate and explore the data simultaneously and manipulate it together. Some projects to exist to start to address these challenges, but overall, this remains another big challenge.

Expressibility: this is a big challenge for oceanographers. These scientists wish to ask questions of their data, that are tied-up with specific domain knowledge. In this research, general tools such as VTK or Iris Explorer can help to visualize the data, but it is difficult to generate an expressive set of analysis tasks, and to perform complex queries that are specific to the domain. Therefore, it may be possible to create a Domain Specific Language or library for Oceanography and subsequently Earth Sciences visual analytics, which will allow researchers of the future to perform highly complex analysis very rapidly on extremely large domain specific datasets.

8.5 Conclusions

The initial chapters to this thesis, established that it fits into the wider context of the grand scientific and scientific visualization challenges which have been identified and much discussed over the past decade. Johnson [152] provided 15 challenges; and the thesis focuses on a number of these.

‘Think about the science’: The science has been very much the focus of this thesis, particularly in view of the combined knowledge and experience of both ocean and computer sciences, and the aim of developing a system with the end users in mind, by collaborating with them from the outset. Within the system itself, the exemplar is the transect calculator and flux tool which meets the real needs of the users, but which would not have emerged as a requirement if the system had been developed by a computer scientist working in isolation of users.

Quantify effectiveness: Throughout the work there has been collaboration with oceanographers. An incremental approach of developing tools ready for oceanographers to use, has enabled better tools to be created. The next stage is clearly a long-term study where the oceanographers start to use the tools in their every-day practices. Effort is currently being placed on this next stage.

Specific Visualization challenges: Global/local visualization (details within context) – the research has developed an problem solving environment that integrates scientific visualization methodologies with information visualization techniques, it incorporates temporal visualization methods and allows visual abstraction and selections to be performed. Therefore, the research demonstrates that it is possible to develop and implement a visual analytics system for the complex area of estuarine hydrodynamics, which provides insight into the data, using a multiplicity of views, graphs and plots to give different perspectives of the spatial, time dependent data, and conveying the detail of the data. The integrated problem solving environment achieved through linked CMV, uses both scientific and information visualization tools and techniques, permits feature detection, supports focus+context, allows iterative study of the data to develop and investigate new hypotheses and enables knowledge discovery. All of this is achieved through the use of general models of computing, in this case a data space method for scaling and aggregating the data.

Bibliography

- [1] P Adams and D Monnermuth. Visualization of steep breaking waves and thin spray sheets around a ship. In *Proceedings, 14th IEEE Visualization Conference*, pages 555–558, 2003.
- [2] W Aigner, S Miksch, W Muller, H Schumann, and C Tominski. Visualizing time oriented data: a systematic view. *Computers and Graphics*, 31:401–409, 2007.
- [3] W Aigner, S Miksch, W Muller, H Schumann, and C Tominski. Visual methods for analysing time oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14 (1):47–60, 2008.
- [4] P Alho and J Aaltonen. Comparing a 1D hydraulic model with a 2D hydraulic model for the simulation of extreme glacial outburst floods. *Hydrological Processes*, 22:1537–1574, 2008.
- [5] A Allan, R Blair, and A Jaramillo. The NANOOS Visualization System: Aggregating, displaying and serving data. In *Oceans 2009: Marine Technology for our future, Global and Local Challenges*, pages 1–9, 2009.
- [6] A Allen, S Sanchagrin, and N Lee. Geovisualization strategies to analyze and communicate inundation risk from storm surges. In *Shifting Shorelines Conference, TCS 22, The Coastal Societys 22nd International Conference*, 2010.
- [7] R Allendes Osorio and K Brodlie. Uncertain flow visualization using LIC. In *Proceedings. 7th Theory and Practice of Computer Graphics*, 2009.
- [8] J Anderson, L Gosink, M Cuchaineau, and K Joy. Feature identification and extraction in function fields. In *Eurographics/IEEE VGTC Symposium on Visualization*, 2007.

- [9] B Andrews. Techniques for spatial analysis and visualization of benthic mapping data. Technical report, NOAA Coastal Services Centre, Charleston, South Carolina, USA (SAIC Report No 623), 2003.
- [10] G Andrienko and N Andrienko. Exploring spatial data with dominant attribute map and parallel coordinates. *Computers, Envi*, 25:5–15, 2001.
- [11] G Andrienko and N Andrienko. Visual exploration of the spatial distribution of temporal behaviours. In *Proceedings of the International Conference on Information Visualization*, pages 799–806, 2005.
- [12] G Andrienko and N Andrienko. Coordinated Multiple Views: a critical view. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007) 2007 Proceedings of the International Conference on Information Visualization*, pages 799–806, 2007.
- [13] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S.I. Fabrikant, M. Jern, M.J. Kraak, H. Schumann, and C. Tominski. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.
- [14] N Andrienko and G Andrienko. A visual analytics framework for spatio-temporal analysis and modelling. *Data Mining and Knowledge Discovery*, 27 (1):pp 55–83, 2013. Data Mining and Knowledge Discovery, accepted 2012 published version: <http://dx.doi.org/10.1007/s10618-012-0285-7>.
- [15] N Andrienko and G Andrienko. Visualization of movement: an overview of methods, tools and procedures. *Information Visualization*, 12 (1):3–24, 2013. Accepted for publication in Information Visualization, 2012 volume unknown.
- [16] N Andrienko, G Andrienko, and P Gatalsky. Tools for visual comparison of spatial development scenarios. In *Seventh International Conference on Information Visualization*, pages 237–244, 2003.
- [17] D. A Aoyama, J. T. T Hsiao, A. F Cardenas, and R. K Pon. TimeLine and visualization of multiple-data sets and the visualization querying challenge. *Journal of Visual Languages and Computing*, 18:1–21, 2007.
- [18] R Arsenault. Real time 3D display of hydrographic data for QA and QC. In *US. Hydrographic. Conference*, 2001.
- [19] R Arsenault, M Plumlee, S Smith, C Ware, R Brennan, and L Mayer. Fusing information in a 3D chart of the future display. In *Fusing information in a 3D chart of the future display*, 2003.
- [20] R Arsenault, C Ware, M Plumtree, S Martin, L Whitcomb, D Wiley, T Gross, and A Bilgili. A system for visualizing time varying oceanographic 3D data.

- In *Proceedings of the Oceans 2004. MTS/IEEE Conference*, pages 743–747, 2004.
- [21] A Artero, M de Oliveira, and F Levkowitz. Uncovering clusters in crowded parallel coordinates visualization. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 81–88, 2004.
- [22] N Banas. NPZVisualizer. Web, 2008.
- [23] A Baptista. Modern paradigms for modeling and visualization of environmental systems. *Encyclopedia of Physical Science and Technology*, Vol 5:565–581, 2002.
- [24] A Baptista, B Howe, J Freire, D Maier, and C Silva. Scientific exploration in the era of ocean observatories. *Computing in Science and Engineering*, 10 (3):53–58, 2008.
- [25] P Bates, M Anderson, J M Hervouet, and J Hawkes. Investigating the behaviour of two-dimensional finite element models of compound channel flow. *Earth Surface Processes and Landforms*, 22:3–17, 1997.
- [26] B Belaton and K Brodlie. Model Centred Approach to Scientific Visualization. *Journal of WSCG*, 10(1):63–70, 2002.
- [27] G Bell, J Gray, and A Szalay. Petascale Computational Systems. *IEEE Computer*, 39:110–112, 2006.
- [28] J Bellingham and M Godin. Exploring Ocean Data. *SIGMOD Record*, 37 (2):78–82, 2008.
- [29] J.L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [30] V Benzaken, J D Fekete, P L Hmery, W Khemiri, and I Manolescu. Edi-Flow: data intensive interactive workflows for visual analytics. In *IEEE 27th International Conference on Data Engineering*, pages 780–791, 2011.
- [31] F Bernardon, S Callahan, J Comba, and C Silva. An adaptive framework for visualizing unstructured grids with time varying scalar fields. *Parallel Computing*, Vol 33 (6):391–405, 2007.
- [32] D E Bernholdt, W R Elwasif, J S Kohl, and T G W Epperly. A component architecture for high performance computing. In *Proceedings of the Workshop on Performance Optimization via High-Level Languages and Libraries*, 2002.
- [33] E Bertini and G Santucci. Quality metrics for 2D scatterplot graphics: Automatically reducing spatial clutter. In *Proceedings, Smart Graphics*, 2004.

- [34] E Bertini, A Tatu, and D Keim. Quality metrics in high dimensional data visualization: an overview and systematization. *IEEE Transactions on Visualization and Computer Graphics*, 17 (12):2203–2212, 2011.
- [35] E Bethel, S Campbell, E Dart, J Shalf, K Stockinger, and K Wu. High performance visualization using Query Driven Visualization and Analytics. Technical report, Lawrence Berkeley National Laboratory, 2006.
- [36] E. W Bethel, C Johnson, K Joy, S Ahern, V Pascucci, H Childs, J Cohen, M Duchaineau, B Hamann, C Hanson, D Laney, P Lindstrom, J Meredith, G Ostrouchov, S Parker, C Silva, A Sanderson, and X Tricoche. SciDAC Visualization and Analytics Center for Enabling Technology. In *J. Phys*, 2007.
- [37] M. I Billen, O Kreylos, B Hamann, A Jadamec, M, L. H Kellog, O Staadt, and D. Y Summer. A geoscience perspective on immersive 3D gridded data visualization. *Computers and Graphics*, 34:1056–1072, 2008.
- [38] J Blass, C Botha, and F Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14 (6):1436–1451, 2008.
- [39] J Blower, K Haines, A Santokhee, and C Liu. GODIVA2: Interactive visualization of environmental data on the Web. *Phil Trans R Soc A*, 367:1035–1039, 2009.
- [40] J. D Blower, F Blanc, P Cornillon, S. C Hankin, and T Loubrieu. Underpinning technologies for oceanographic data sharing, visualization and analysis: Review and future outlook. In *GODAE Final Symposium*, 2008.
- [41] M Bonazountas, S Kanellopoulos, J Schaller, D Kallidromitou, and G Martirano. GIS and Visual Analytics on grid technologies. In *Proceedings of the NATO Advanced Research Workshop on Geographical Information Processing and Visual Analytics for Environmental Security*, pages 265–385, 2008.
- [42] P Borodin, A Greß, and K Reinhard. Visualization Aspects in the MERCW Project. In *Proceedings of The 2nd US/EU-Baltic International Symposium*, 2006.
- [43] J. N Boyer, P Sterling, and R. D Jones. Maximising information from a water quality monitoring network through visualization techniques. *Estuarine, Coastal and Shelf Science*, 50:39–48, 2000.
- [44] A Brandt. Grand challenges for Ocean Science Research. *Oceanography*, 19:14–17, 2006.
- [45] K Brodlie, J Brooke, M Chen, D Chisnall, C Hughes, N John, M Jones, M Riding, N Roard, M Turner, and J Wood. Adaptive Infrastructure for

- Visual Computing. In *Proc. of Theory and Practice of Computer Graphics*, pages 147–156. Eurographics, June 2007.
- [46] J. M Brooke, J Marsh, S Pettifer, and L. S Sastry. The importance of locality in the visualization of large datasets. *Concurrency and computation practice and experience*, 19:195–205, 2007.
- [47] S Brooks and R McDonnell. Research advances in geocomputation for hydrological and geomorphological modeling towards the twenty-first century. *Hydrological processes*, 14(11-12):1899–2108, 2000.
- [48] I Brown, S Jude, S Koukoulas, R Nicholls, M Dickson, and M Walkden. Dynamic simulation and visualization of coastal erosion. *Computers, Environment and Urban Systems*, 30(6):840–860, 2006.
- [49] J Brown and A G Davies. Flood/ebb tidal dominance in an estuary: sediment transport and morphology. In *Proc of 5th IAHR International Symposium on river, coastal and estuarine morphodynamics*, volume 2, pages 779–787, 2007.
- [50] A Buja, D Cook, and D F Swayne. Interactive High-Dimensional Data Visualization. *Journal of Computational and Graphical Statistics*, 5 (1):78–99, 1996.
- [51] J Cai, R Kapila, and G Pal. HMVC: The layered pattern for developing strong client tiers. In *Java World*, 2000.
- [52] W Cameron and D Pritchard. *The Sea: Estuaries*. Wiley, New York, 1963. Hill, M N (Editor).
- [53] J Carrivick. Application of 2D hydrodynamic modelling to high magnitude outburst floods: An example from Kverkfjoll, Iceland. *Journal of Hydrology*, 321:187–199, 2006.
- [54] Y Chao, P Li, P Wang, D Katz, and Cheng B. Ocean modelling and visualization on massively parallel computers. Technical report, DSpace at Jet Propulsion Laboratory, 2004.
- [55] A Chave, M Arrott, L Smarr, J Orcutt, E Lazowska, J Delaney, and Abbot M. LOOKING: Cyberinfrastructure for Ocean Observatories. In *SSC06*, 2006.
- [56] C Chen. An information-theoretic view of visual analytics. *IEEE Computer Graphics and Applications*, 28 (1):18–23, 2008.
- [57] J Chen, A MacEachren, and D Guo. Visual Inquiry Toolkit: an integrated approach for exploring and interpreting space-time multivariate patterns. In *Proceedings AutoCarta 2006*, 2006.

- [58] H Childs, E Brugger, B Whitlock, J Meredith, S Ahern, K Bonnell, M Miller, G H Weber, C Harrison, D Pugmire, T Fogal, C Garth, A Sanderson, E W Bethel, M Durant, D Camp, F M Favre, O Reubel, P Navratil, M Wheeler, P Selby, and F Vivodtzev. Visit: An End User Tool for Visualising and Analyzing Very Large Data. In *Proceedings of SciDAC*, 2011.
- [59] D Chung, P Legg, M Parry, R Bown, I Griffiths, R Laramée, and M Chen. Glyph Sorting: Interactive Visualization for Multi-dimensional Data. ArXiv e-prints, April 2013. accepted for publication in Information Visualization.
- [60] W Cleveland. *The Elements of Data Graphing*. Wadsworth, 1985.
- [61] J Clyne, P Mininni, A. Norton, and A M Rast. Interactive desktop analysis of high resolution simulations: Application to turbulent plume dynamics and current sheet formation. *New Journal of Physics*, 9:301, 2007.
- [62] J Clyne and M Rast. A prototype discovery environment for analyzing and visualizing terascale turbulent fluid flow simulations. In *Proceedings of Visualization and Data Analysis 2005*, pp. 284-294, (2005)., pages 284–294, 2005.
- [63] A Coelho, M Nascimento, C Bentes, M Stelling de Castro, and R Farias. Parallel volume rendering for ocean visualization in a cluster of PCs. In *Brazilian Symposium*, 2004.
- [64] B Collins. *Animation and Scientific Visualization Tools and Applications*, chapter Data visualization - has it all been seen before?, pages pp 3–28. D Watson, Academic Press, London, 1993.
- [65] C. J Cotter and G. J Gorman. Diagnostic tools for 3D unstructured oceanographic data. *Ocean Modelling*, 20:170–182, 2008.
- [66] K Craik. *The Nature of Explanation*. Cambridge University Press, Cambridge, 1943.
- [67] A. G Davies and P. D Thorne. Advances in the Study of Moving Sediments and Evolving Seabeds. *Surveys in Geophysics*, 29:1–36, 2008.
- [68] A. M Davies and L. Y Oey. Recent developments in physical oceanographic modelling: Part III. *Continental Shelf Research*, 26:1359, 2006.
- [69] M De Berg, M Van Krefeld, M Overmars, and O Schwarzkopf. *Computational Geometry (2nd Revised Ed)*., chapter Quadrees, pages 291–306. Springer-Verlag, 2000.
- [70] R DeAmicis, G Conti, S Pifferrn, and B Simoes. Geospatial visual analytics. In *Proceedings of the NATO Advanced Research Workshop on Geographical Information Processing and Visual Analytics for Environmental Security*, pages 265–385, 2008.

- [71] J Dean and S Ghemawat. MapReduce: simplified data processing on large clusters. In *Proceedings of the 6th Conference of Symposium in Operating Systems Design and Implementation*, 2004.
- [72] Z Defne, K Haas, and M Fritz. Numerical modeling of tidal currents and the effects of power extraction on estuarine hydrodynamics along the Georgia Coast, USA. *Renewable Energy*, 36:3461–3471, 2011.
- [73] M Derthick, J Kolojejchick, and S F Roth. An Interactive Visualization Environment for Data Exploration. In *Proceedings of Knowledge Discovery in Databases*, pages 2–9, 1997.
- [74] D Dibase, A MacEachren, J Krygier, and C Reeves. Animation and the role of map design in scientific visualization. *Cartog. and Geographic Info Science*, 19 (4):201–214, 1992.
- [75] H Doleisch, M Gasser, and H Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *Proceedings of the 5th Joint IEEE VGTC-EUROGRAPHICS Symposium on Visualization (Vis-Sym 03)*, pages 239–248, 2003.
- [76] H Doleisch, M Mayer, M Gasser, R Wanker, and H Hauser. Visual Analysis of complex time dependent simulation results of a diesel exhaust system. In *Data Visualization: Proceedings of the 6th Joint IEEE TCVG Eurographics Symposium on Visualization*, pages 91–96, 2004.
- [77] S dos Santos and K Brodlie. Visualizing and investigating multidimensional functions. In *Joint Eurographics-IEEE TCVG Symposium on Visualization*, pages 1–10, 2002.
- [78] B Douglas. Global sea level rise. *J Geophysical Research*, 96 (C4):6981–6992, 1991.
- [79] B Duvenage. Using an implicit min/max kd-tree for doing efficient terrain line of sight calculations. In *Afrigraph09. Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa*, pages 81–90, 2009.
- [80] J Dyer. Development and application of an operational software tool for assessment of flood inundation. In *Northern Gulf Institutes Annual Conference*, 2012.
- [81] H Edelsbrunner, J Harer, V Natarajan, and V Pascucci. Local and global comparison of continuous functions. In *IEEE Visualization*, pages 275–288. IEEE Computer Society, 2004.
- [82] R Edsall and D Peuquet. A graphical user interface for the integration of time into GIS. In *Proceedings of the 1997 American Congress of Surveying and Mapping Annual Convention and Exhibition*, pages 182–189, 1997.

- [83] G Ellis, E Bertini, and A Dix. The sampling lens: Making sense of saturated visualizations. In *Proceedings CHI05, Extended abstracts on human factors in computing systems*, pages 1351–1354, 2005.
- [84] D Ellsworth, L J Chang, and H W Shen. Accelerating time varying hardware volume rendering using tsp trees and colour based error metrics. In *VVS '00 Proceedings of the 2000 IEEE symposium on Volume Visualization*, pages 119–128, 2000.
- [85] S Emmott and S Rison. Towards 2020 science. Technical report, The 2020 Science Group, Microsoft Research, 2005.
- [86] R Erbacher and D Frincke. Hierarchical linked views. In *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pages 35–46, 2007.
- [87] Gordon Erlebacher, David A Yuen, and Fabien Dubuffet. Current trends and demands in visualization in the geosciences. *Visual Geosciences*, 6(3):1–59, 2001.
- [88] T Ertl. Interactive visualisation, beyond standard techniques for irrelevant data sets. In *Distinguished Lecture, Visual Computing Graduate School*, 2010.
- [89] C Esperanca, G Hjaltason, H Samet, F Brabec, and E Tanin. An overview of the sand spatial database system. In *Communications of the ACM*, 2001.
- [90] R Finkel and J. L Bentley. Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974.
- [91] Department for Environment; Food and Rural Affairs. Flood and coastal defence appraisal guidance. fcdpag3 economic appraisal. supplementary note on operating authorities, climate change impacts, 2006.
- [92] J French and N Clifford. Hydrodynamic modelling as a basis for explaining estuarine environmental dynamics: some computational and methodological issues. *Hydrological Processes*, 13:2089–2108, 2000.
- [93] Y H Fua, M Ward, and E Rudensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the Conference on Visualization, Viz 99*, pages 43–50, 1999.
- [94] R Fuchs and H Hauser. Visualization of multi-variate scientific data. *Computer Graphics Forum*, 28:1670–1690, 2009.
- [95] K Furakawa. The role of man made headland in generating patches in coastal waters of ise bay, japan. *Estuarine, Coastal and Shelf Science*, 50 (1):33–37, 2000.

- [96] G W Furnas. Generalized fisheye views. In *CHI '86 Conference on Human Factors in Computing Systems*, 1986.
- [97] T Gabriel, A Pintilie, and M Berthold. Exploring hierarchical rule systems in parallel coordinates. In *IDA 2005. Proceedings of the 6th International Conference on Advances in Intelligent Data Analysis*, pages 97–108, 2005.
- [98] K Gaither and R Moorhead. Visualizing vector information in ocean environments. In *OCEANS '95. MTS/IEEE. Challenges of Our Changing Global Environment*, 1995.
- [99] D Galloway, E Wolanski, and B King. Coastal oceanography data visualization using data explorer. In *Proceedings of the 1996 IBM Visualization Data Explorer Symposium*, 1996.
- [100] A Gemmell, R Barciela, J Blower, K Haines, Q Harpham, K Millard, M Price, and A. Saulter. An ecoop web portal for visualizing and comparing distributed coastal oceanography model and in situ. *Ocean Science*, 7:445–454, 2011.
- [101] Z Geng, Z Peng, R. S Laramée, R Walker, and J. C Roberts. Angular histograms: Frequency-based visualizations for large, high dimensional data. visualization and computer graphics. *IEEE Transactions*, 17 (12):2572–2580, 2011.
- [102] R George, J Roberts, and A G Davies. Using processing to develop icove: a tool for interactive coastal oceanographic visualization. In J P Collomosse and I J Grimstead, editors, *TPCG*, pages 99–105, 2010.
- [103] R. L George and J. C Roberts. Coastal shelf visualization: A comparative study of the use of opendx and vtk in visualizing hydro-informatic numerical models. In *Eurographics UK - Theory and Practise of Computer Graphics*, 2009.
- [104] M Gleicher, D Albers, R Walker, I Jusufi, C Hansen, and J Roberts. Visual comparison for information visualization. *Info Vis*, 10 (4):289–309, 2011.
- [105] N Goncharova, P Borodin, and A Gress. Gis for planning, navigation, acquisition and visualization of results for the study of chemical munitions dumpsites in the baltic sea. In *Proceedings of 2nd International Conference of GIS/RS in Hydrology, Water Resources and Environment (ICGRHWE '07)*, 2007.
- [106] M Goodchild and Y Shiren. A hierarchical spatial data structure for global geographic information systems. *Graphical Models and Image Processing*, 54:31–44, 1992.

- [107] L Gosink, J Anderson, E Bethel, and K. Joy. Query driven visualization of time varying adaptive mesh refinement data. *Transactions on Visualization and Computer Graphics/IEEE Visualization*, 14 (6):1715–1722, 2007.
- [108] J Grant, G Erlebacher, and J O’Brien. Visualizing ocean flow vertical motions using lagrangian-eulerian time surfaces. In *IEEE Visualization 2002*, 2002.
- [109] J Gray. *The Four Paradigm: Data Intensive Scientific Discover*, chapter Jim Gray on eScience: A Transformed Scientific Method, pages xvii–xxxi. Microsoft Research, Redmond WA, 2009.
- [110] K Grochow, B Howe, M Stoermer, R Bargao, and E Lazowska. Client + cloud: Seamless architectures for visual data analytics in the ocean sciences. In *SSDBM’10 Proceedings of the 22nd international conference on Scientific and statistical database management*, pages 114–131, 2010.
- [111] K Grochow, M Stoermer, D Kelley, J Delaney, and E Lazowska. Cove: a visual environment for ocean observatory design. *Journal of Physics, Conference Series*, 125:1–8, 2008.
- [112] R Guting. An introduction to spatial database systems. *Invited contribution to a special issue on spatial database systems of the VLDB Journal*, 3 (4):357–399, 1994.
- [113] R Haber and D McNabb. *Visualization Idioms: A conceptual model for scientific visualization systems*, *Visualization in Scientific Computing*. IEEE CS Press, 1990.
- [114] C Hansen, C Johnson, V Pascucci, and C Silva. *The Fourth Paradigm: Data Intensive Scientific Disco*, chapter Visualization for Data-Intensive Science, pages 153–164. Microsoft Research, 2009.
- [115] L Hanyk, D. A Yuen, C Matyska, and J Velonsky. Visualization of time-dependent dynamics of postglacial rebound. *Vis Geoscience*, 13:37–47, 2008.
- [116] R. J Hardy and P. D Bates. The importance of spatial resolution in hydraulic models for floodplain environments. *Journal of Hydrology*, 216:124–136, 1999.
- [117] S Haroz, K Ma, and K Heitmann. Multiple uncertainties in time-variant cosmological particle data. In *IEEE Pacific Visualization Symposium*, 2008.
- [118] L Harrison, T Burkiewicz, Z Wang, W Ribarsky, and R Chang. A linked feature space approach to exploring lidar data. In *Visual Analytics for Homeland Defense and Security II, Proceedings of the SPIE 2010*, volume 7709, 2010.

- [119] S Havre, E Hetzler, P Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8 (1):9–20, 2002.
- [120] Q Hayir, B Senongeullari, I Klinc, A Erturk, k Cigizoglu, A, M. S Kabdasli, O Yagci, and K Day. Scenarios of tsunami amplitudes in the north eastern coast of sea of marmara, generated by submarine mass failure. *Coastal Engineering*, 55:333–356, 2008.
- [121] C He and P. F Hamblin. Visualization of model results and field observations in irregular coastal models. *Estuarine, Coastal and Shelf Science*, 50:73–80, 2000.
- [122] M Head, L Phu, O Costolo, K Countryman, and C Szczechowski. Applications of 3d visualizations of oceanographic databases. *Oceans, MTS/IEEE Conference Proceedings*, 2:1210–1215, 1997.
- [123] C Healey. Formalising artistic techniques and scientific visualization for painted renditions of complex information spaces. In *Proceedings International Joint Conference on Artificial Intelligence*, volume 50, pages 73–80, 2001.
- [124] C Healey, J Enns, L Tateosian, and M. Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics*, 23 (1):64–69, 2004.
- [125] J Heinrich and D Weiskopf. Continuous parallel coordinates. *IEE Tran*, 15, issue 6:1531–1538, November 2009.
- [126] A Hermann and C Moore. *Computers in Fisheries Research*, chapter Visualization in fisheries oceanography: New approaches for the rapid exploration of coastal ecosystems, pages 317–336. Springer Science and Business Media B.V, 2009. Megrey, B A and Mokness, E (Editors).
- [127] J Hervouet. Telemac modelling system: an overview. *Hydrological Processes*, 14:2209–10, 2000.
- [128] J Hervouet and P Bates. The telemac modelling system, introduction. *Hydrological Processes.*, 14:2207–8, 2000.
- [129] J M Hervouet. *Hydrodynamics of Free Surface Flows*. John Wiley and sons, 2007.
- [130] T Hey, S Tansley, and S Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington, 2009.
- [131] B Hibbard. Visad: Connecting people to computations and people to people. *ACM SIGGRAPH Computer Graphics*, 32 (3):10–12, 1998.

- [132] B Hibbard, M Bottinger, M Schultz, and J Biercamp. Visualization in earth system science. *ACM SIGGRAPH Computer Graphics*, 36 (4):5–9, 2002.
- [133] B Hibbard and D Santek. The vis-5d system for easy interactive visualization. In *Proc IEEE Visualization*, pages 28–35, 2000.
- [134] W Hibbard, C Dyer, and B Paul. Display of scientific data structures for algorithm visualization. In *Proceedings of IEEE Visualization Viz 92*, 1992.
- [135] Q Ho and M Jern. Interacting with 4d oceanographic volume data using geoanalytics tools. In *GI Science 2008, Workshop on Geospatial Visual Analytics*, 2008.
- [136] B Howe, P Lawson, R Bellinger, E Anderson, E Santos, J Freire, C Scheidegger, A Baptista, and C Silva. End-to-end, escience: Integrating workflow, query, visualization and provenance at an ocean observatory. In *eScience '08. IEEE Fourth International Conference*, pages 127–134, 2008.
- [137] B Howe, H Vo, and C Silva. Query driven visualization in the cloud, with mapreduce. In *Fourth Annual Workshop on Ultrascale Visualization, 2009, co-located with Supercomputing*, 2009.
- [138] C Hsu, J Ahrens, and K Heitmann. Verification of the time evolution of cosmological simulations of hypothesis-driven comparative and quantitative visualization. In *Pacific Viz 2010*, pages 81–88, 2010.
- [139] B Huang and H Lin. Geovr: a web-based tool for virtual reality presentation from 2d gis data. *Computers and Geosciences*, 25 (10):1167–1175, 1999.
- [140] J Hunt. Floods in a changing climate: a review. *Phil. Trans. R Soc. Lond. A*, 360:1531–1543, 2002.
- [141] J Hunt. Inland and coastal flooding: developments in prediction and prevention. *Phil. Trans. R. Soc. Lond A*, 363:1475–1491, 2005.
- [142] A Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1 (2):69–92, 1985.
- [143] A Inselberg and B Dinsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Vis 90. Proceedings of the 1st conference on Visualization*, pages 361–378, 1990.
- [144] H Janicke, M Bottinger, U Milolajawewicz, and G Scheurmann. Visual exploration of climate variability changes using wavelet analysis. *IEEE Transactions on Visualization and Computer Graphics (2009)*, 15 (6):1375–1382, 2009.

- [145] R Jenkinson. Rainy river 2d hydrodynamic modelling study phase ii. Technical report, National Research Council, Canada, Canadian Hydraulics Centre, Ottawa, 2011.
- [146] M Jern. *Visualization user interface a third generation GUI paradigm. Visualization and Modelling*. Academic Press, 1997. Earnshaw, R and Vince, J and Jones, H (Editors).
- [147] M Jern, T Astrom, and S Johansson. Analytic tools applied to large geospatial datasets. In *12th International Conference Information Visualization, IEEE 2008*, pages 362–372, 2008.
- [148] W. H Jiminez, W. T Correa, C. T Silva, and A Baptista. Visualizing spatial and temporal variability in coastal observatories. In *VIS '03: Proceedings of the conference on Visualization*, 2003.
- [149] A Johannsen and R. J Moorhead. Agp: Ocean model flow visualization. *IEEE Computer Graphics*, 15:28–33, 1995.
- [150] J Johansson, P Liung, M Jern, and M Cooper. Revealing structure within clustered parallel coordinates displays. In *Symposion of Information Visualization, InfoViz05*, pages 125–132, 2005.
- [151] J Johansson and M Cooper. A screen space quality method for data abstraction. In A Vilanova, A Telea, G Scheuermann, and T Moller, editors, *Eurographics/IEEE VGTC Symposium on Visualization 2008*, volume 27 (3), pages 1039–1046, 2008.
- [152] C Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24 (4):13–17, 2004.
- [153] M JohnsonRobertson, O Pizarro, S. B Williams, and I Mahon. Generation and visualization of largescale three dimensional reconstructions for underwater robotic surveys. *Journal of Field Robotics*, 27(1):21–51, 2010.
- [154] I Joliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
- [155] J. E Jones. Coastal and shelf-sea modelling in the european context. *Oceanography and Marine Biology: an Annual Review*, 40:37–41, 2002.
- [156] R. R Jones, K. J. W McCaffrey, P Clegg, R. W Wilson, R. E Holdsworth, J Imber, and S Waggott. Integration of regional to outcrop digital data: 3d visualization of multi-scale geological models. *Computers and Geosciences*, 35:4–18, 2009.
- [157] A Joshi, J Caban, and P Rheingans. Case study on visualizing hurricanes using illustration inspired techniques. In *IEEE Transactions on Visualization and Computer Graphics*, volume 15 (5), pages 709–718, 2009.

- [158] R Kahler, J Wise, T Abel, and H. C Hege. Gpu assisted raycasting for cosmological adaptive mesh refinement simulation. In *Eurographics Volume Graphics*, 2006.
- [159] T. R Keen, R Vickery, and P Flynn. Scientific visualisation of sediment dynamics in the bottom boundary layer. In *Estuarine and Coastal Modeling Proceedings of the Seventh International Conference*, 2001.
- [160] J Kehrer, F Ladstadter, P Muigg, H Doleisch, A Steiner, and H Hauser. Hypothesis generation in climate research with interactive visual data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 14 (6):1579–1586, 2008.
- [161] K Kehrer, P Muigg, H Doleisch, and H Hauser. Interactive visual analysis of heterogeneous scientific data across an interface. In *IEEE Transactions on Visualization and Computer Graphics, IEEE VisWeek 2011*, volume 17 (7), page 934946, 2011.
- [162] D Keim. Designing pixel-oriented visualization techniques: theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, 6 (1):59–78, 2000.
- [163] D Keim, G Adrienko, J-D Fekete, C Gorg, J Kohlhammer, and Melancon G. *Visual Analytics: Definition, processes and challenges*. In *Information Visualization, Chapter 7, of LNCS State-of-the-Art Survey*, volume 4950. Springer, 2008. Kerren, A. and Stasko, J T and Fekete, J D and, North C (Editors).
- [164] D Keim and A Hermann. The gridfit algorithm: an efficient and effective approach to visualizing large amounts of spatial data. In *Proceedings of the IEEE, Visualization 98.*, 1998.
- [165] D Keim, J Kohlhammer, G Santucci, F Mannsmann, F Wanner, and M Schaefer. Visual analytics challenges. In *Proceedings, Echallenges 2009, Istanbul*, 2009.
- [166] D Keim, F Mansmann, D Oelke, and H Ziegler. Visual analytics: Combining automated discovery with interactive visualizations. In *Discovery Science*, pages 2–14, 2008.
- [167] D Keim, F Mansmann, A Stoffel, and H Zeigler. *Visual analytics*. Springer, 2009.
- [168] D Keim, W Muller, and H Schumann. Visual data mining. In *Eurographics 2002, State of the Art Report*, pages 49–68, 2002.
- [169] Daniel A. Keim, Florian Mansmann, Andreas Stoffel, and Hartmut Ziegler. Visual analytics. In *Encyclopedia of Database Systems*, December 2009. Springer, December, 2009.

- [170] Z Kemp and G Meaden. Visualization for fisheries management from a spatiotemporal perspective. *ICES Journal of Marine Science*, 50:190–202, 2002.
- [171] R Kirby, H Marmansis, and H Laidlaw. Visualizing multi-valued data from 2d incompressible flows using concepts from painting. In *Proceedings IEEE Visualization 99*, pages 333–340, 1999.
- [172] D Kitsiou, X de Madron, and P Arnau. Development of a data visualization and analysis tool to study the particle dynamics in the coastal zone. *Marine Pollution Bulletin*, 43 (7 to 12):262–269, 2001.
- [173] T Knudsen. Busstop – an integrated system for handling, analysis and visualization of ocean data. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geology*, 24:411–414, 1999.
- [174] J Kolojejchick, S Roth, and P Lucas. Information appliances and tools in visage. *IEEE Computer Graphics and Applications*, 17 (4):32–41, 1997.
- [175] D Koop, C Scheidegger, S Callaghan, J Freire, and C Silva. Viscomplete: Automating suggestions for visualization pipelines. *IEEE Transactions on Visualization and Computer Graphics*, 14 (6):1691–1698, 2008.
- [176] O Kreylos, G Weber, E Bethel, J Shalf, B Hamann, and K Joy. Remote interactive direct volume rendering of amr data. Technical report, Lawrence Berkeley National Laboratory, USA, 2002.
- [177] D Labat. Recent advances in wavelet analysis: part one, a review of concepts. *Journal of Hydrology*, 314:275–288, 2005.
- [178] D Lam, L Leon, S Hamilton, N Crookshank, D Bonin, and D Swayne. Multi-modal integration in a decision support system: a technical user interface approach for watershed and lake management scenarios. *Environmental Modelling and Software*, 19:317–324, 2004.
- [179] P. B Lamphere, J. M Linebarger, and A. R Breckenridge. Dynamic isosurface extraction and level-of-detail in voxel space. Technical report, National Labs, Albuquerque, NM (US) and Sandia National Labs, Livermore, CA (US), 1999.
- [180] R S Laramée and R Kosara. Future challenges and unsolved problems in human centred visualization. In A. Kerren, A Ebert, and J Meyer, editors, (*GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5-8, 2006, Revised Papers*) *Chapter 5 of Human-Centered Visualization Environments - Lecture Notes in Computer Science (LNCS)*, pages 231–254. Springer Verlag, 2006.
- [181] H Ledoux and C Gold. Modelling oceanographic data with the three-dimensional voronoi diagram. *Review Internationale de Geomatique*, 16:51–70, 2006.

- [182] J Lee and K Choi. Real-time hydroenvironmental modelling and visualization system for public engagement. *Environmental Fluid Mechanics*, 8:411–421, 2008.
- [183] J Li, N Jing, and M Sun. Spatial database techniques oriented to visualization in 3d gis. In *Proceedings of the 2nd International Symposium on Digital Earth*, 2001.
- [184] Q Li, X Bao, C Song, J Zhang, and J North. Dynamic query sliders vs brushing histograms. In *CHI '03 Extended abstracts on human factors in computing systems*, 2004.
- [185] S. J Liang and F Molkenhuth. A virtual gis-based hydrodynamic model system for tamshui river. *Journal of Hydroinformatics*, 3:195–202, 2001.
- [186] P Lindstrom, D Koller, W Ribarsky, L Hodges, A den Bosch, and N Faust. An integrated global gis and visual simulation system. Technical report, Georgia Institute of Technology, 1997.
- [187] S Lipsa, R Laramée, S Cox, J Roberts, and R Walker. Visualization for the physical sciences. In N John and B Wyvill, editors, *Eurographics State of the Art reports (Llandudno, UK, 2011)*, pages 49–73, 2011.
- [188] Z Liu and J Stasko. Mental models, visual reasoning and interaction in information visualization: a top down perspective. *IEEE Transactions on Visualization and Computer Graphics*, 16 (6):999–1008, 2010.
- [189] W Lorenson. On the death of visualization. In *Position Papers NIH/NSF Proc. Fall Workshop Visualization Research Challenges*, 2004.
- [190] D Lu, P Croft, R Reddy, S Zeigler, and R Moorhead. Using immersive virtual reality to study hurricane cases. In *Preprint for AMS Annual Meeting, Orlando, Florida, 2002*, 2002.
- [191] K L Ma, E Lum, H Yu, H Akiba, M Y Huang, and Y Wang. Scientific discovery through advanced visualization. In *J. Phys.: Conf. Ser.*, volume 16, pages 491–500, 2005.
- [192] K. L Ma, C Wang, H Yu, K Moreland, J Huang, and R Ross. Next generation visualization technologies: enabling discoveries at extreme scale. Technical report, SCIDAC Institute for Ultra Scale Visualization, 2009.
- [193] B Mann, R Williams, M Atkinson, K Brodlie, A Storkey, and C Williams. Scientific data mining, integration and visualization. Technical report, e-Science Institute, Edinburgh, 2002.
- [194] R Marshall, J Kempf, Dyer S, and Y Chieh-Cheng. Visualisation methods and simulation steering for a 3d turbulence model of lake erie. In *Proceedings of the 1990 Symposium, Interactive 3D Graphics*, pages 89–98, 1990.

- [195] P Masumder. Planar decomposition for quadtree data structure. *Computer Vision, Graphics and Image Processing.*, 38:258–274, 1987.
- [196] M. F Maury. *The Physical Geography of the Sea and its Meteorology*. Sampson Low, Son & Marston, 1869.
- [197] L Mayer, M Paton, L Gee, J Gardner, and C Ware. Interactive 3d visualization: A tool for seafloor navigation, exploration and engineering. In *OCEANS 2000 MTS/IEEE Conference and Exhibition*, volume 2, pages 913–919, 2000.
- [198] M McCann. Creating 3d oceanographic data visualizations for the web. In *Proceedings of the 7th International Conference on Web3D Technology*, pages 179–184, 2002.
- [199] D McClurg, C Moore, and N Soreide. Real time 3d el nino, la nina visualizations and animations from the tao buoy network in the tropical pacific. In *roceedings of the Ocean 99 MTS/IEEE Conference*, volume 2, pages 951–954, 1999.
- [200] N Merati, C Chamberlin, C Moor, C Titov, and T Vance. *Integration of Tsunami analysis tools into a GIS workspace research modelling and hazard mitigation efforts within NOAA's Center for Tsunami Research*, in *Geospatial Techniques in Urban Hazard and Disaster Analysis Geotechnologies and the Environment, Chapter 14*. Springer Science+Business Media BV, 2010. Showalter, P and Lu, Y (Editors).
- [201] D Middleton, T Scheitlin, and B Wilhelmson. *Visualization in weather and Climate Research*. In *The Visualization Handbook*. Academic Press, 2005. Hansen, C and Johnson, C (Editors).
- [202] P Mitchell, C Ware, and J Kelley. Designing flow visualizations for oceanography and meteorology using interactive design space hill climbing. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 355–361, 2009.
- [203] K Moreland, U Ayachit, B Geveci, and K L Ma. Dax: Data analysis at extreme. In *SciDAC 2011*, 2011.
- [204] C Muller and L Birn. Combining iterative analytical reasoning and software development using the visualization language processing. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 251–252, 2009.
- [205] R Nelson and H Samet. A consistent hierarchical representation for vector data. In *Proceedings, SIGGRAPH 86. Proceedings of the 13th annual conference on computer graphics and interactive techniques*, pages 197–206, 1986.

- [206] R J Nicholls and A Cazenave. Sea-level rise and its impact on coastal zones. *Science*, 328 (5985):1517–1520, 2010.
- [207] O Nielsen, D Gray, A McPherson, and A Hitchman. Hydrodynamic modelling of coastal inundation. In *MODSIM 2005 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia & New Zealand*, pages 518–523, 2005.
- [208] T Nocke, M Flechsig, and U Bohm. Visual exploration and evaluation of climate related simulation data. In *Proceedings of the 2007 Winter Simulation Conference*, 2007.
- [209] C North and B Schneiderman. A taxonomy of multiple window coordinations. Technical report, University of Maryland Computer Science Department, Technical Report, No CS-TR-3854, 1997.
- [210] C North and B Schneiderman. Snap-together visualization: Coordinating multiple views to explore information. Technical report, University of Maryland Computer Science Department Technical Report CT-TR-4020, 1999.
- [211] M Novotny and H Hauser. Outlier preserving focus+content visualization in parallel coordinates. *IEE Transactions on Visualization and Computer Graphics*, 12 (5):893–900, 2006.
- [212] B Ooi, K McDonell, and R Sacks-Davis. Spatial k-d tree: An indexing mechanism for spatial databases. In *Proceedings, IEEE COMPSAC Conference. Tokyo*, pages 433–438, 1987.
- [213] R Ostnes, V Abbott, and S Lavender. Visualization techniques: an overview part 2. *The Hydrographic Journal*, 114:3–8, 2004.
- [214] N Pais, A De Vries, A De Jong, and A Boertjes. Remote and in situ sensing for dyke monitoring: The ijkdiik experience. In *Proceedings of the NATO Advanced Research Workshop on Geographical Information Processing and Visual Analytics for Environmental Security*, pages 465–476, 2008.
- [215] R Pajarola. Large scale terrain visualization using the restricted quadtree triangulation. In *IEEE Visualization*, pages 19–26, 1998.
- [216] R Pajarola, M Antonijuan, and Lario R. Quadtin: Quadtree based triangulated irregular networks. In *IEEE Visualization*, pages 395–402, 2002.
- [217] A. T Pang, C. M Wittenbrink, and S. K Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13:370–390, 1997.
- [218] D Paridis, P Ohl, and P Daniel. Operational storm surges forecasting in an estuary. *Marine Geodesy*, 32 (2):166–177, 2009.

- [219] S Park, C Bajaj, and I Ihm. Visualization of very large oceanography time-varying volume datasets. *Lecture Notes in Computer Science*, 3037:419–426, 2004.
- [220] S Park, C. L Bajaj, and V Siddavanahali. Case study: interactive rendering of adaptive mesh refinement data. In *IEEE Visualization*, pages 521–524, 2002.
- [221] N Patrikalakis, J McCarthy, A Robinson, H Schmidt, C Evangelinos, P Haley, S Lalis, P Lermusiaux, R Tian, W Leslie, and W Cho. *Towards a dynamic data driven system for rapid adaptive interdisciplinary ocean forecasting in Dynamic Data Driven Application Systems*. Kluwer Academic Publishers, Netherlands, 2004.
- [222] W Peng, M Ward, and E Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proceedings, IEEE Symposium Information Visualization (InfoViz) 2004*, 2004.
- [223] Z Peng, X Geng, and R Laramée. Design and implementation of a system for interactive higher dimension vector field visualization. In *Eurographics 2011, poster*, 2011.
- [224] Z Peng and R Laramée. Higher dimensional vector field visualization: A survey. *Theory and Practice of Computer Graphics*, 2:149–163, 2009.
- [225] J S Pethick. *Estuaries and wetlands, function and form. Wetland Management*. Thomas Telford, 1994.
- [226] W Pike, J Stasko, R Chang, and T OConnell. The science of interaction. *Information Visualization*, 8 (4):263–274, 2009.
- [227] H Piringer, R Kosara, and H Hauser. Interactive focus+context visualization with linked 2d/3d scatter plots. In *Proc of the International Conference on Coordinated and Multiple Views in Exploratory Visualization. (CMV2004)*, pages 49–60, 2004.
- [228] S Popenet, R Gorman, G Rickard, and H Tolman. A quadtree-adaptive spectral wave model. *Ocean Modelling*, 34:36–39, 2010.
- [229] K Potter, A Wilson, P Bremer, D Williams, C Doutriaux, V Pascucci, and C Johnson. Visualization of uncertainty and ensemble data: exploration of climate modelling and weather forecast data with integrated visus-cdat systems. In *Journal of Physics: Conference Series*, volume 180 (1), pages 1–5, 2009.
- [230] H Qu, W Y Chan, A Xu, K L Chun, K H Lau, and P Guo. Visual analysis of the air pollution problem in hong kong. *IEEE Transactions on Visualization and Computer Graphics*, 13 (6):1408–1415, 2007.

- [231] S Rahmstorf. A new view on sea level rise. *Nature Reports Climate Change*, 4:44–45, 2010.
- [232] T Rees. C-squares, a new spatial indexing system and its applicability to the description of oceanographic datasets. *Oceanography*, 16(1):11–19, 2003.
- [233] W Ribarsky. Advanced visual analytics. In *CHC 4th Annual Meeting*. Charlotte Visualization Center, Coastal Hazards Center of Excellence, 2012.
- [234] K Riley, Y Song, M Kraus, D Ebert, and J Levit. Visualization of structured, non uniform grids. *IEEE Computer Graphics and Application*, 26 (1):46–55, 2006.
- [235] J Roberts. *Exploring Geovisualization*, chapter Exploratory visualization with multiple linked views, pages 159–180. Elsevier Ltd, 2004.
- [236] J Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *CMV 07*,, pages 61–71, 2007.
- [237] J Roberts and A Wright. Towards ubiquitous brushing for information visualization. In *Tenth International Conference on Information Visualization*, pages 151–156, 2006.
- [238] P Robins. Present and future flooding scenarios in the dyfi estuary, wales, uk. Report for the countryside council for wales (number 3), Centre for Applied Marine Sciences, Bangor University, 2008.
- [239] P Robins, A Davies, and R Jones. Application of a coastal model to simulate present and future inundation and aid coastal management. *Journal of Coastal Conservation*, 15:1–14, 2011.
- [240] P Robins and A. G Davies. Morphodynamic processes in shallow estuaries: influence of tidal flats and channels on sand transport. In *Proceedings of the 6th IAHR Symposium on River, Coastal and Estuarine Morphodynamics (RCEM 2009)*, volume 1, pages 161–167, 2009.
- [241] P Robins and A G Davies. Application of telemac-2d and sisyph to complex estuarine regions to inform future management decisions. In *XVIIIth Telemac & Mascaret User Club Chatou, France*, pages 19–21, 2011.
- [242] P E Robins. Development of a morphodynamic model of the burry inlet to inform future management decisions. Technical Report 1805, Bangor University, CAMS Report: CAMS 2009-5, 2009.
- [243] P E Robins. Development of a morphodynamic model of the dyfi estuary to inform future management decisions. Technical report, Bangor University, CAMS Report: CAMS 2009 Countryside Council for Wales, 2009.

- [244] J Rodrigues, A Traina, and C Traina. Frequency plot and relevance plot to enhance visual data exploration. In *SIBGRAPI Brazilian Conference on Graphics, Patterns and Images*, volume 117, 2003.
- [245] S F Roth, P Lucas, J A Senn, C C Gomberg, P J Burks, M B Stroffolino, J Kolojejchick, and C Dunmire. Visage: A user interface environment for exploring information. In *Proceedings of Information Visualization*, pages 3–12, 1996.
- [246] O Rubel, W Prabhat, H Childs, J Meredith, C Geddes, E CormierMichel, S Ahern, G Weber, P Messmer, H Hagen, B Hamann, and E Bethel. High performance multivariate visual data exploration for extremely large data. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, volume 12, pages 51–55, 2008.
- [247] I Sadarjoen, F Post, B Ma, D Banks, and H G Pagendarm. Selective visualization of vortices in hydrodynamic flows. In *Visualization 98 Proc Vis*, pages 419–422, 1998.
- [248] Samet. *Theoretical Foundations of Computer Graphics and CAD.*, volume 40 of *NATO ASI Series*, chapter An overview of quadtrees, octrees and related hierarchical data structures, pages 51–68. Springer Berlin, 1988.
- [249] H Samet. Hierarchical spatial data structures. In A Buchmann, O Gnther, T R Smith, and Y F Wang, editors, *Design and Implementation of Large Spatial Databases 1st Symposium, SSD'89*, volume 409. Springer-Verlag Lecture Notes in Computer Science, 1989.
- [250] H Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing and GIS*. Addison-Wesley, Reading MA, 1990.
- [251] H Samet. *The design and analysis of spatial data structures*. Addison-Wesley series in computer science, 1990.
- [252] H Samet. *Modern Database Systems: The Object Model, Interoperability and Beyond*, chapter Spatial Data Structures, pages 361–385. Addison Wesley/ACM Press, Reading, MA, 1995.
- [253] H Samet and R Webber. Extending sand spatial database system for the visualization of three dimensional scientific data. *Geographical Analysis*, 38:87–101, 2006.
- [254] R Samet and E Ozsavas. Optimisation of quadtree triangulation for terrain models. In J Blanc-Talon, editor, *Proceedings ACIVS 2007, Lecture Notes in Computer Science*, pages 48–59, 2007.
- [255] M. M Sardadi, M. S Rahim, Jupir Z, and D Daman. Quadtree spatial indexing use to make faster showing geographical map in mobile gis technology, using

- an oracle 10g application server and mapviewer applications. *International Journal of Computer Science and Network Security*, 8(10):253–262, 2008.
- [256] N Sauber, H Thiesel, and H P Seidel. Multi-field graphs: an approach to visualizing correlations in multi-field scalar data. *IEEE Transactions on Visualization and Computer Graphics*, 12 (5):917–924, 2006.
- [257] G Schmidt, S-L Chen, A Bryden, M Livingston, and B Osborn. Multi-dimensional visual representations for underwater environmental uncertainty. *EEE Computer Graphics and Applications*, 24 (5):56–65, 2004.
- [258] B Schneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11 (6):70–77, 1994.
- [259] B Schneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings IEEE Symp. Visual Languages*, pages 336–343, 1996.
- [260] D Schwab, Beletsk D, and J Lou. The 1998 coastal turbidity plume in lake michigan. *Estuarine, Coastal and Shelf Science*, 50 (1):49–58, 2000.
- [261] Z Shang. Developing the TELEMAC system for HECToR (phase 2b & beyond). Technical report, Distributed CSE, 2011.
- [262] H Shen, T Lee, A Chaudhuri, and B Nouanesengse. Visual analytics for enabling extreme scale scientific discovery. In *Proceedings of SciDac Conference 2011, Denver, CO*, 2011.
- [263] Y Shen, J. R Crouch, J. A Austin, and M. S Dinniman. An educational interactive numerical model of the chesapeake bay. *Computers & Geosciences*, 34:247–258, 2008.
- [264] R Signell, H Jenter, and A Blumberg. Predicting the physical effects of relocating bostons sewage outfall. *Estuarine, Coastal and Shelf Science*, 50 (1):59–72, 2000.
- [265] C Silva and J Freire. Software infrastructure for exploratory visualization and data analysis: past present and future. In *SCIDAC 2008 Journal of Physics Conference Series*, volume 125, 2008.
- [266] D Silver. *Feature Visualization. Scientific Visualization, Overviews, Methodologies and Techniques*. IEEE Press, Los Alamitos, USA, 1997. Nielsen, G and Hagen, H and Muller, H (Editors).
- [267] B Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.
- [268] D Simons and R Rensick. Change blindness: past, present and future. *Trends in Cognitive Science*, 9 (1):16–20, 2005.

- [269] R Sinha, S Samaddar, D Bhattacharuyya, and T H Kim. A tutorial on spatial data handling. *International Journal of Database Theory and Application*, 3 (1):1–12, 2010.
- [270] M Sips, J Schneidewind, D. A. Keim, and H. Schumann. Scalable pixel-based visual interfaces: Challenges and solutions. In *Tenth International Conference on Information Visualization (IV06)*, 2006.
- [271] Y Song, J Ye, N Svakhine, S LasherTrapp, M Baldwin, and D Ebert. An atmospheric visual analysis and exploration system. *IEEE Transactions on Visualization and Computer Graphics*, 12 (5):1157–1164, 2006.
- [272] R Spence. *Information Visualization*. Harlow: Addison Wesley/ACM Press, 2001.
- [273] C Steed, P Fitzpatrick, T JankunKelly, and J Swann. North atlantic hurricane trend analysis using parallel coordinates and statistical techniques. Technical report, North Atlantic hurricane trend analysis using parallel coordinates and statistical technique, 2008.
- [274] C Steed, P Fitzpatrick, J Swann, and T JankunKelly. Tropical cyclone trend analysis using enhanced parallel coordinates and statistical analysis. *Cartography and Geographic Information Science*, 36 (3):252–265, 2009.
- [275] C. A Steed, P. J Fitzpatrick, T. J Jankun-Kelly, and J. E Swann. An interactive parallel coordinates technique applied to a tropical cyclone climate analysis. *Computers & Geosciences*, 35:1529–1539, 2009.
- [276] D Steere, A Baptista, D McNamee, C Pu, and J Walpole. Research challenges in environmental observation and forecasting systems. In *Proceedings MobiCom 00 Proceedings of the 6th annual international conference on mobile computing and networking.*, pages 292–299, 2000.
- [277] R Stein, A Shih, P Baker, C Cerco, and M Noel. Scientific visualization of water quality in the chesapeake bay. In *Vis '00: Proceedings of the Conference on Visualization*, pages 509–512. IEEE Computer Society, 2000.
- [278] K Stockinger, J Shalf, E Bethel, and K Wu. Query driven visualization of large data sets. In *IEEE Visualization*, pages 115–121, 2003.
- [279] K Stockinger, J Shalf, W Bethel, and K. Wu. DEX, increasing the capability of scientific data analysis pipelines by using efficient bitmap indices to accelerate scientific visualizations. In *International Conference on Scientific and Statistical Database Management*, pages 35–44. IEEE Computer Society, 2005.
- [280] C Stolte, D Tang, and P Hanrahan. Query, analysis and visualization of hierarchically structured data using polaris. Technical report, Stanford University, Computer Science Department, 2000.

- [281] T Suzuki and M Matsuyama. Numerical experiments on stratified wind-induced circulation in tokyo bay. *Estuarine, Coastal and Shelf Science*, 50 (1):17–25, 2000.
- [282] A Szalay and J Gray. Science in an Exponential World. *Nature*, 440 (7083):413–414, 2006.
- [283] H Takahashi, Y Hosokawa, K Furukawa, and H Yoshimura. Visualization of an oxygen-deficient bottom water circulation in osaka bay, japan. *Estuarine, Coastal and Shelf Science*, 50:81–84, 2000.
- [284] R Theron. Visual analytics of paleoceanographic conditions. In *IEEE Symposium On Visual Analytics Science And Technology*, pages 19–26, 2006.
- [285] J J Thomas and K Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26 (1):10–13, 2006.
- [286] J J Thomas and K A Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE CS Press, 2005.
- [287] P. D Thorne and A. G Davies. Sounding out sediment transport: Observations and modelling of ripple generated vortices. *Ocean Challenge*, 14(3):15–21, 2006.
- [288] C Tominski, J Abello, and H Schumann. Axes-based visualizations with radial layouts. In *Proceedings 2004 Symposium on Applied Computing*, pages 1242–1247, 2004.
- [289] L Treinish. Task specific visualization design: a case study in operational weather forecasting. In *Proceedings IEEE Visualization*, pages 405–409, 1998.
- [290] L Treinish. Multi-resolution visualization techniques for nested weather models. In *Proceedings IEEE Visualization*, pages 513–516, 2000.
- [291] K Trouw, J Williams, and C Rose. Modelling sand resuspension by waves over a rippled bed. *Estuarine, Coastal and Shelf Science*, 50 (1):137–142, 2000.
- [292] T Tsai and H C Chui. Adaptive level of detail for large terrain visualization. In J Chen, J Jiang, and S Nayak, editors, *XXIst ISPRS Congress Technical Commission IV*, volume XXXVII part 4, pages 579–584, 2008.
- [293] E Tufte. *The visual display of quantitative information, 2nd Ed.* Graphics Press, 2001.
- [294] U Turdukolov, M J Kraak, and C. Blok. Designing a visual environment for exploration of time series of remote sensing data: in search for convective clouds. *Computers and Graphics*, 31:371–379, 2007.

- [295] Ulanbek Turdukulov and Connie Blok. Visual analytics to explore iceberg movement. Technical report, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 2009.
- [296] T Tzouramanis, M Vassilakopoulos, and Y Manolopoulos. Multiversion linear quadtree for spatio temporal data. In *DBIS-DASFAA '00 Proceedings of the East-European Conference on Advances in Databases and Information Systems Held Jointly with International Conference on Database Systems for Advanced Applications: Current Issues in Databases and Information Systems*, pages 279–292, 2000.
- [297] C Upson, T Faulhaber, D Kamins, D Laidlaw, D Schlegel, J Vroom, R Gurwitz, and A van Dam. The application visualization system: A computational environment for scientific visualization. *IEEE Computer Graphics and Applications*, 9 (4):30–42, 1989.
- [298] J van Wijk and R van Liere. Hyperslice visualization of scalar function of many variables. In *Proceedings of the IEEE Conference on Visualization (Visualization 93)*, 1993.
- [299] J J van Wijk. The value of visualization. In *IEEE Visualization*, pages 79–86, 2005.
- [300] J. J van Wijk, J. S Hin, A, W. C de Leeuw, and F. H Post. Three ways to show 3d fluid flow. *IEEE CG&A*, 14(5):33–39, September 1994.
- [301] H Vo, S Callahan, N Smith, C Silva, D Owen, and D Weinstein. irun: Interactive rendering of large unstructured grids. In *Eurographics Symposium on Parallel Graphics and Visualization*, 2007.
- [302] H Vo, D Osmari, B Summa, J Comba, V Pascucci, and C Silva. Stream enabled parallel dataflow architecture for multicore systems. *Computer Graphics Forum*, 10 (3):1073–1082, 2010.
- [303] R Wain and M Ashworth. A java gui and distributed corba client-server interface for a coastal ocean model. Technical report, Council for the Central Laboratory of the Research Councils, 2005.
- [304] J Walker, Z Gheng, M Jones, and R Laramée. Visualization of large, time-dependent, abstract data with integrated spherical and parallel coordinates. In *Eurographics Conference on Visualization (EuroVis)*, pages pp 43–47, 2012.
- [305] M Walkley, J Wood, and K Brodlie. A distributed co-operative problem solving environment. In *Proceedings Computational Science ICCS 2002*, pages 853–861, 2000.
- [306] H Wang, K. H Lau, W. M Chan, and L. A Wong. A pc-based visualization system for coastal ocean and atmospheric modeling. In *Conference Proceeding Paper, Estuarine and Coastal Modelling 1999*, pages 1233–1243, 1999.

- [307] M WangBaldonado, A Woodruff, and A Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI01)*, pages 110–119, 2001.
- [308] M Ward. Xmdvtool: integrating multiple methods for visualizing multivariate data. In *Proceedings of IEEE Visualization*, 1994.
- [309] M Ward. A taxonomy of glyph placement strategies for multi-dimensional information visualization. *Information Visuali*, 1:3–4, 2002.
- [310] M Ward. *Handbook of Data Visualization*, chapter Multivariate data glyphs: Principles and practice, pages 179–198. Springer Handbooks, 2008.
- [311] C Ware, M Pluntree, R Arsenault, L Mayer, and S Smith. Geozui3d: Data fusion for interpreting oceanographic data. In *MTS IEEE Conference and Exhibition*, volume 3, pages 1960–1964, 2001.
- [312] F Watford, D Beard, R Hay, and K Ambrose. Innovative visualization for australias marine zone. In *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The national Biennial Conference of the Spatial Sciences Institute*, 2005.
- [313] G Weber, V Beckner, H Childs, T Licocki, M Miller, B van Straalen, and E Bethel. Visualization tools for adaptive mesh refinement data. In *Proceedings of the 4th High End Visualization Workshop*, 2007.
- [314] G Weber, V Beckner, H Childs, T Ligocki, M Miller, B van Straalen, and E Bethel. Visualization of scalar adaptive mesh refinement data. *Numerical Modeling of Space Plasma Flows: Astronum-2007 (Astronomical Society of the Pacific Conference Series)*, 385:309–320, 2008.
- [315] E Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85 (411):664–675, 1990.
- [316] E Wegman and Q Luo. High dimensional clustering using parallel coordinates and the grand tour. *Computing Science and Statistics*, 2:361–368, 1996.
- [317] T Wilde, J. A Kohl, and R. E Flanery. Immersive and 3d viewers for cumulvs: Vtk/cave and avs/express. *Future Generation Computer Systems*, 19(5):701–719, 2003.
- [318] L Wilkinson, A Anand, and R Grossman. High dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *IEEE Transactions on Visualization and Computer Graphics*, 12 (6):1362–1372, 2006.
- [319] S Williams, M Hecht, M Petersen, R Strelitz, M Maltrud, J Ahrens, M Hkawitshka, and B Hamann. Visualization and analysis of eddies in a global ocean simulation. *Computer Graphics Forum*, 30 (3):991–1000, 2011.

- [320] E Wolanski and S Spagnol. Sticky waters in the great barrier reef, visualization in marine science. *Estuarine, Coastal and Shelf Science*, 50 (1):27–32, 2000.
- [321] E Wolanski, S Spagnol, P Gentian, M Spaulding, and D Prandle. Visualization in marine science. *Estuarine, Coastal and Shelf Science*, 50(1):7–9, 2000.
- [322] D. J Wright and P. N Halpin. *Place Matters: Geospatial Tools for Marine Science, Conservation, and Management in the Pacific Northwest*, chapter Spatial reasoning for "terra incognita": Progress and grand challenges in marine GIS. Oregon State University Press, 2005.
- [323] K Wu, S Koegler, J Chen, and A Shoshani. Using bitmap index for interactive exploration of large datasets. In *Proceedings Scientific and Statistical Database Management, IEEE Computer Society*, pages 65–74, 2003.
- [324] E Yassuda, S Davies, D Mendelsohn, T Isaji, and S Peene. Development of a waste load allocation model for the charleston harbour estuary, phase ii estuarine. *Coastal and Shelf Science*, 50 (1):99–107, 2000.
- [325] H Yu, K L Ma, and J Welling. A parallel visualization pipeline for terascale earthquake simulations sc04. In *Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, 2004.
- [326] H Zhang, Y Shi, and D Yuen. Modeling and visualization of tsunamis. *Pure and Applied Geophysics*, 165:475–496, 2008.
- [327] P Zhang, S Shekhar, Huang. Y, and V Kumar. Spatial cone tree: An indexing structure for correlation queries in spatial time series data. In *Proceedings of International Workshop on Next Generation Geospatial Information 2003*, 2003.
- [328] Z Zheng, D Huang, J Zhang, and S He. A soa-based decision support geographic information system for storm disaster assessment. In *18th International Conference on Geoinformatics*, pages 1–6, 2010.

Appendix

VINCA
Usability Evaluation



PRIFYSGOL
BANGOR
UNIVERSITY

August 2013

SCHOOL OF COMPUTER SCIENCE
BANGOR UNIVERSITY



Consent Form

VINCA Usability Evaluation

This is to certify that I hereby agree to participate as a volunteer in an experiment as an authorised part of the research undertakings within the School of Computer Science at Bangor University, under the supervision of Jonathan C. Roberts.

The experiment, and my role in the investigation have been fully explained to me by Richard L. S. F. George, and I fully understand his explanation. The procedures and risks have been explained to me fully and I have been able to have all questions answered to my satisfaction.

I understand that all data that is a result of my participation will remain strictly confidential. I understand that I may request a summary of the results of this study by contacting one of the researchers (Richard L. S. F. George (r.l.george@bangor.ac.uk), Jonathan C. Roberts (j.c.roberts@bangor.ac.uk)).

I understand that I am free to withdraw my consent and terminate my participation at anytime without prejudice.

My responsibility as a participant is to take part in the experiment actively, and willingly and if I choose to do so I will exercise my right to withdraw.

Date

Participant's Signature

I certify that I have fully explained the investigation to the above individual.

Date

Experimenter's Signature

This form will be produced in duplicate. One copy should be retained by the participant and the other by the researcher.



Consent Form

VINCA Usability Evaluation

This is to certify that I hereby agree to participate as a volunteer in an experiment as an authorised part of the research undertakings within the School of Computer Science at Bangor University, under the supervision of Jonathan C. Roberts.

The experiment, and my role in the investigation have been fully explained to me by Richard L. S. F. George, and I fully understand his explanation. The procedures and risks have been explained to me fully and I have been able to have all questions answered to my satisfaction.

I understand that all data that is a result of my participation will remain strictly confidential. I understand that I may request a summary of the results of this study by contacting one of the researchers (Richard L. S. F. George (r.l.george@bangor.ac.uk), Jonathan C. Roberts (j.c.roberts@bangor.ac.uk)).

I understand that I am free to withdraw my consent and terminate my participation at anytime without prejudice.

My responsibility as a participant is to take part in the experiment actively, and willingly and if I choose to do so I will exercise my right to withdraw.

Date

Participant's Signature

I certify that I have fully explained the investigation to the above individual.

Date

Experimenter's Signature

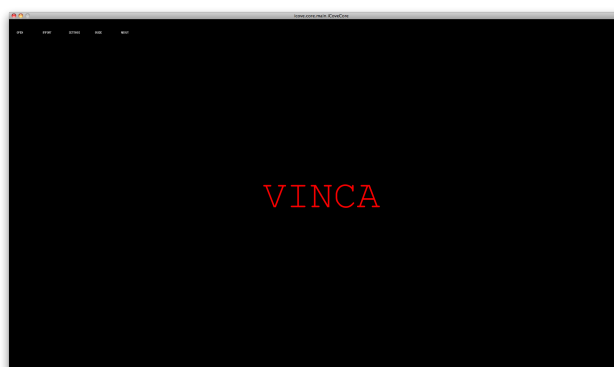
This form will be produced in duplicate. One copy should be retained by the participant and the other by the researcher.

Evaluation Tasks

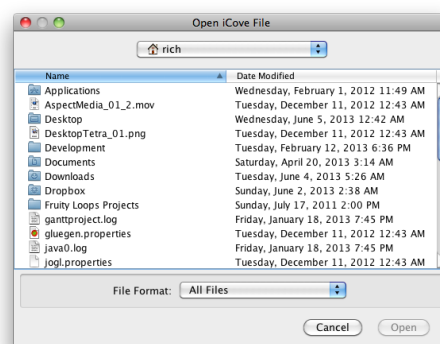
Please read each task description before you start the task. Feel free to refer back to the task description at any point.

Task 1. Loading the data into VINCA and preparing for exploration.

The first task requires you to load the test data into VINCA. To do this, you need to click on the the 'Load Data' button in the top left of the startup screen shown below (a). An open file dialog box will launch(b), use this to navigate to the Users/rich/Telemac Data/burry_05.xml file and then press 'Open'.

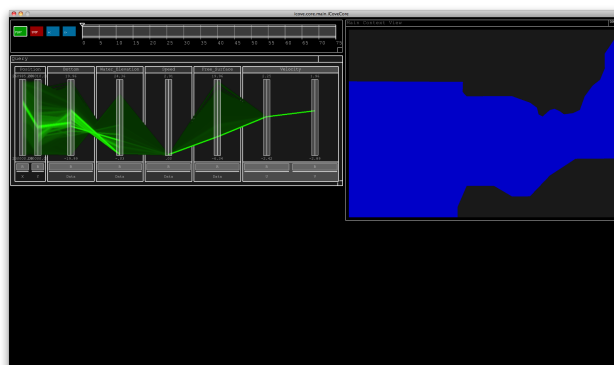


(a)

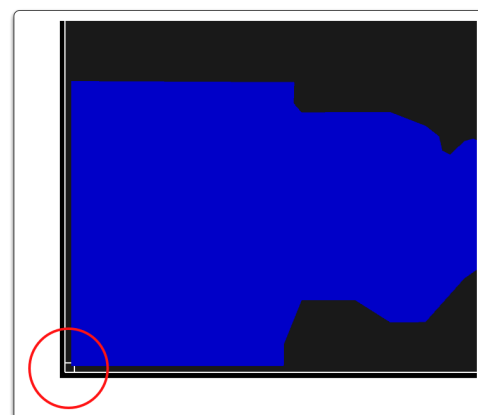


(b)

The data will take approximately 3 minutes to load, and once it has loaded you will be presented with three views (Timeline, Spatial Context, and Query) ready for exploration(a). You will notice in the spatial context view that the data is shaded with a single solid colour, this give a depiction of the shape of the domain but doesn't provide any information about the data contained within the mesh. So as an initial starting point for the exploration trail, you will change the colour-mapping of the spatial context view. To do this, you need to click the bottom left hand corner of the panel containing the spatial context view(b), this will open the views's control panel.



(a)



(b)

The control panel will open with the colour-mapping options as default, to turn on a colour-mapping for a specific variable you will need to click the button underneath it's data label(a). So change the colour-mapping to 'Bottom' and you will see an instant change in the spatial context view, this is now depicting the bathymetry of the domain with a basic RGB colour-map. It would be more useful to add some extra detail to the colour-map, so to edit the colour-map click the colour button associated with the 'bottom' variable at the base of the control panel(b).

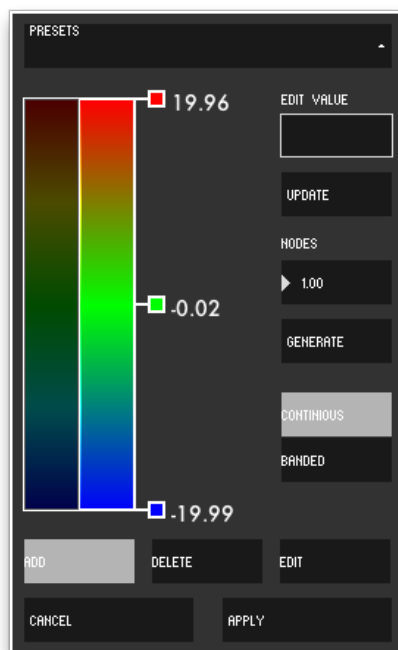


(a)

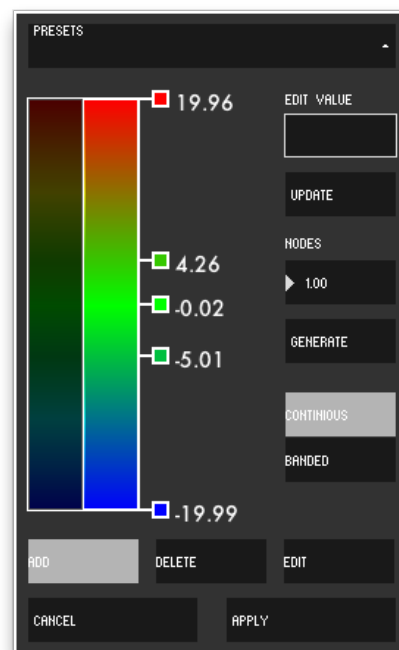


(b)

This will open the colour mapping tool(a). To edit the colour map: firstly click on the 'add' button this will allow you to place a couple more markers on the colour bar by clicking at the approximate height desired on the colour bar. For this task click roughly once underneath the middle bar and once above the middle bar(b).

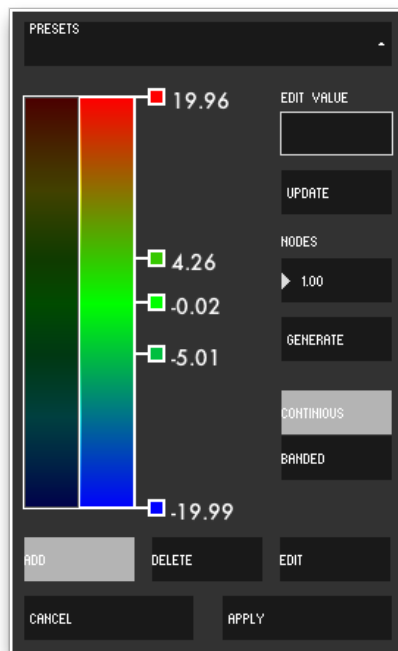


(a)

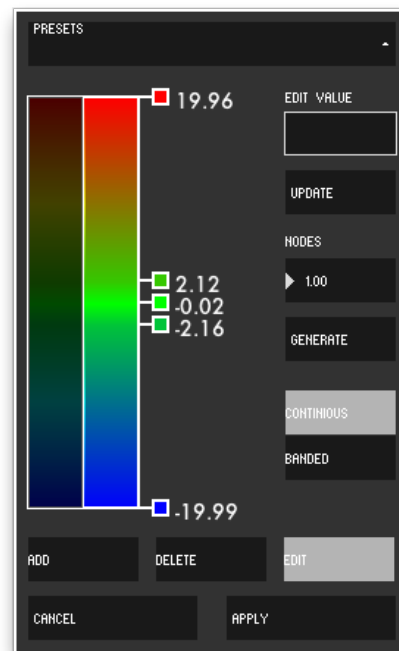


(b)

Then switch the tool to edit mode by pressing the 'edit' button(a). This will allow you you both change the colour of the markers and thus the colour map, and finely position the markers. Firstly position the bottom of the the three middle markers at approximately -2.00 and the upper of the three middle markers at 2.00, to position the markers press and drag under place in the desired location(b). Now change the colour of the markers by right clicking on the marker



(a)



(b)

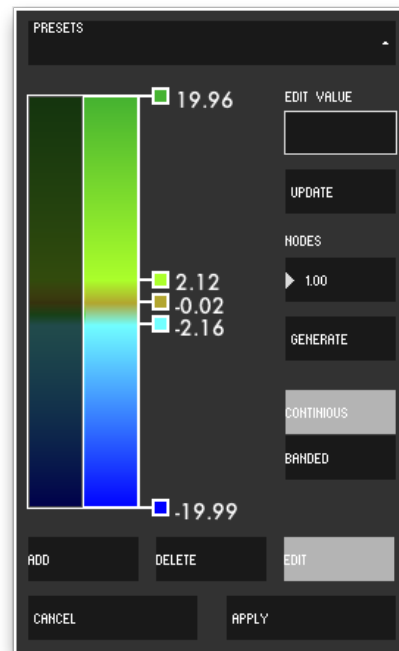
Now change the colour of the markers by right clicking on the marker, this will launch the colour chooser. Using the colour chooser wheel and brightness slider Select a colour and press 'apply'. This will update the colour bar.

Repeat this process for the rest of the markers. In this task use a dark blue for the bottom marker, a light blue for the marker at -2.0, an yellowy orange marker for the marker at 0.0, a light yellowy green for the marker at 2.0 and a dark green for the top marker.

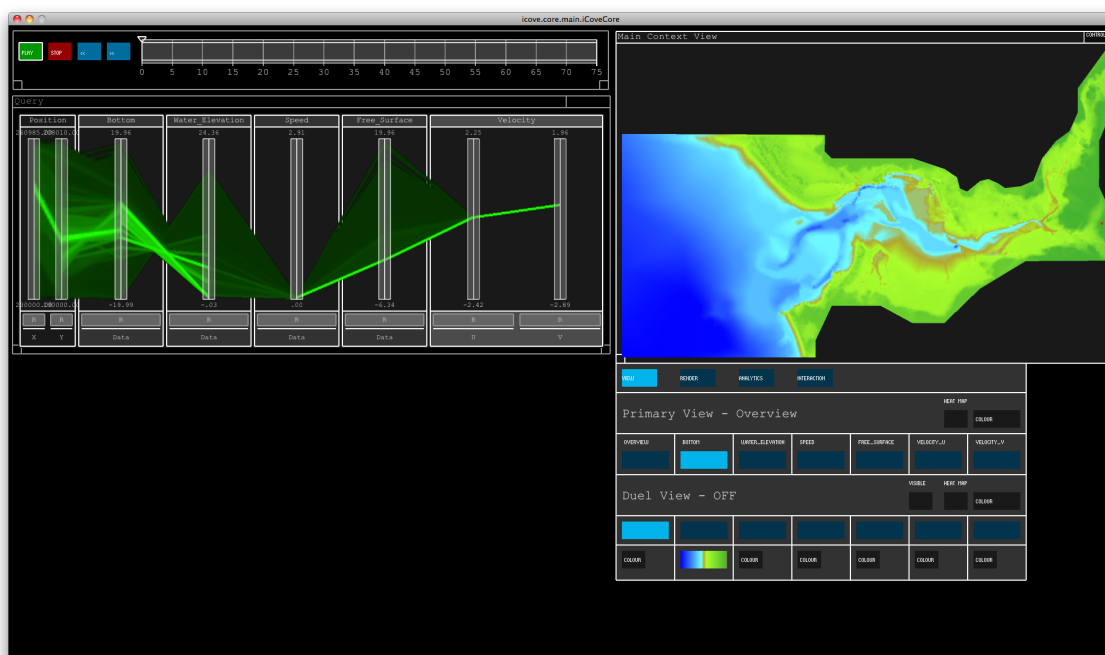
Once all of the colour markers have been edited, press the 'apply' button in the colour mapper tool and this will update the colour map in the spatial context.



(a)



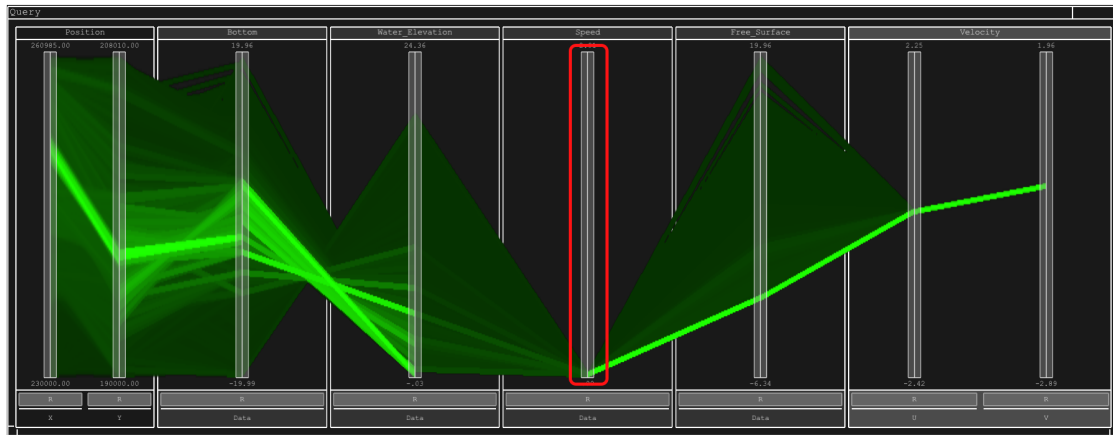
(b)



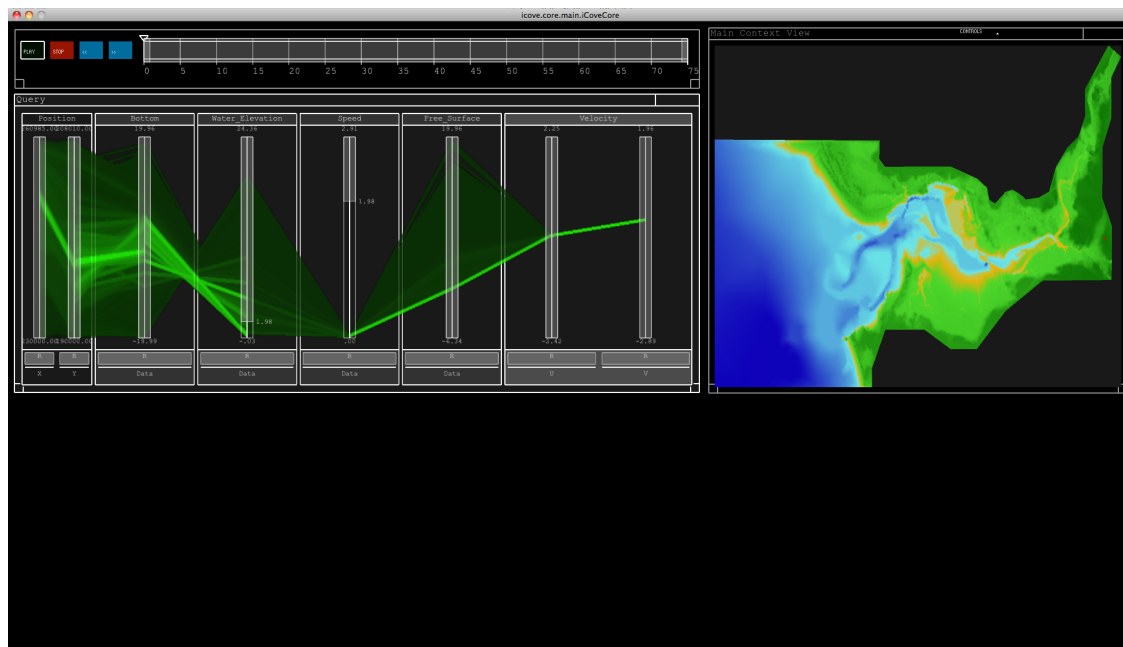
Task 2. Basic Exploration and querying

In the second task you will start to explore the data and learn how to get answers from the system, so along with the task a few questions will be asked and it required that you note your answers down in the places provided.

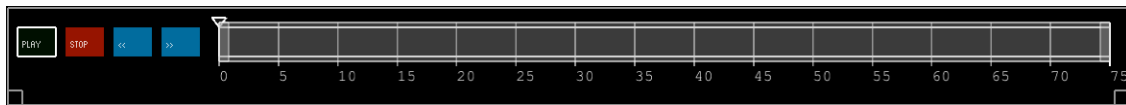
Exploration of the data is centred around the Query View, this view allows you to view the data holistically and filter the data quickly. To filter the data, each of the dimensions in the Query View has a range slider attached to the axis.



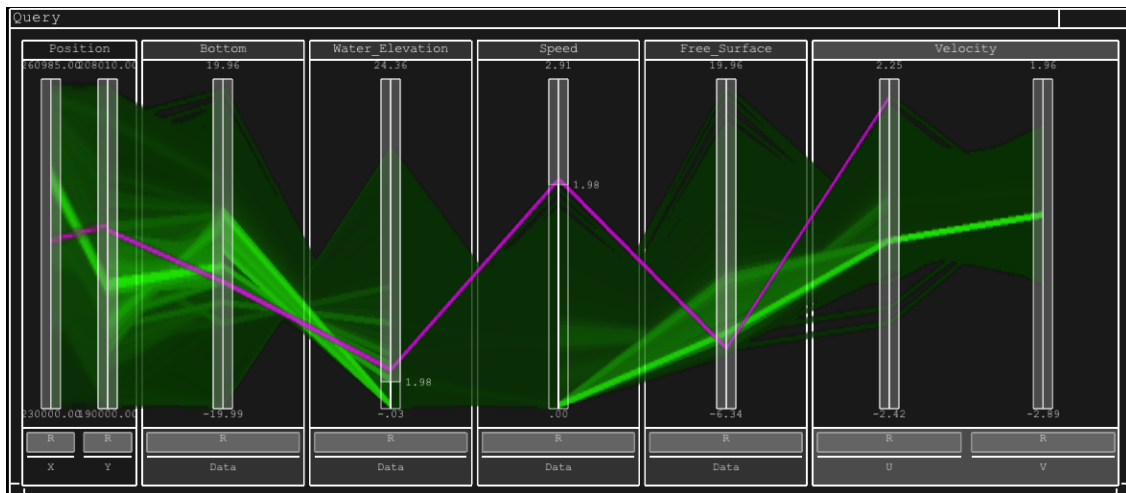
An example query an oceanographer may have is: which deep water channels in the estuary have the maximum flow speed through them? To answer this question using the Query View we need to first adjust the range slider of the speed dimension so that it filters only the fast moving areas of flow, so drag the lower marker of the range slider with the mouse to approximately 2.0 m/s. To help improve our search adjust the lower water elevation range to 2.0m, this will allow for us to look at areas that have at least 2.0m of water which in general a deep water channel should have.



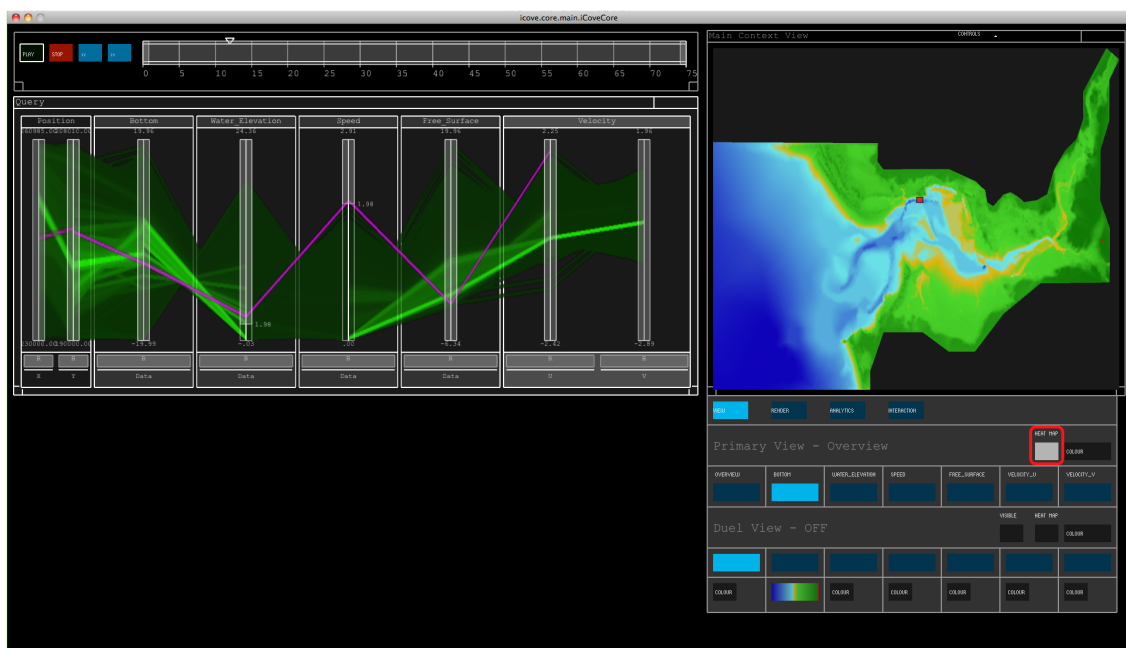
Now its time to find at which time-steps the maximum flow speeds occur at. To do this, press the play button – this will play through each time-step of the data automatically filtering it and caching the results. (This will take approximately 1 – 2mins on the first run) Once it has played back through the data once, press the stop button to reset the timeline back to the start.



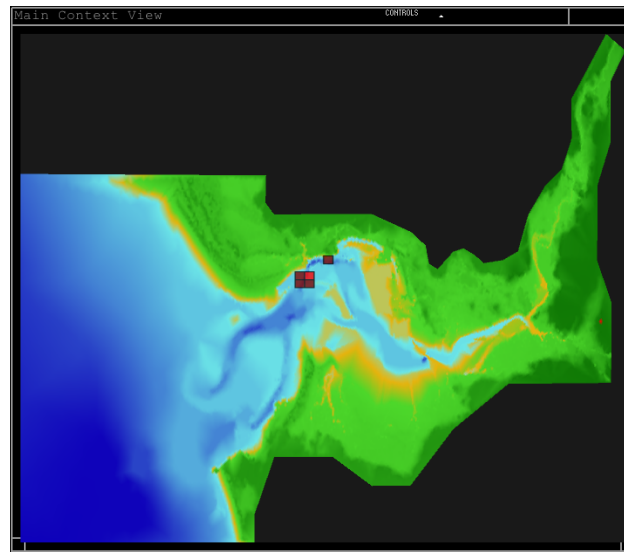
Using the stepper buttons in the timeline view, locate one of the time-steps that meets the search criteria. This is indicated in the Query View with the purple brushed lines.



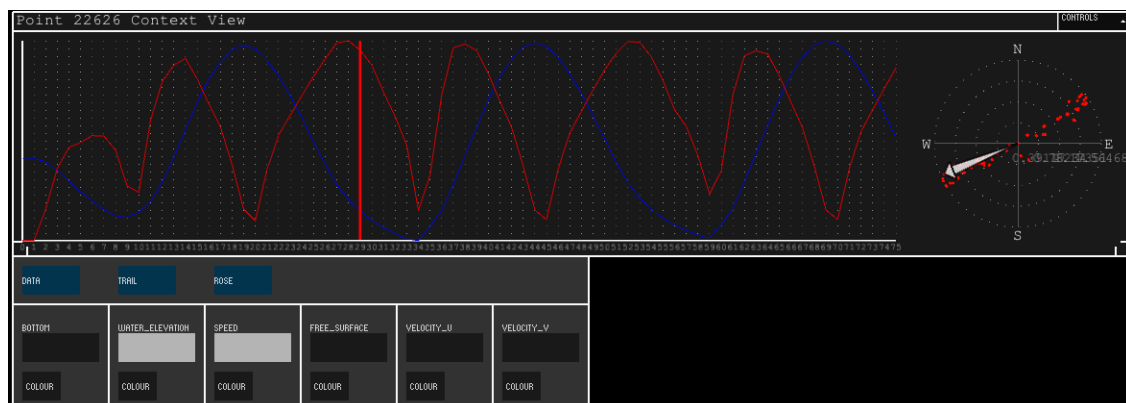
To get a spatial representation of the results for where the greatest flow speed occurs within the data set we need to switch on the heat map, which is located in the view tab of the Main View control panel see image below. This will switch on a heat-map in the Main View that is dynamically filtered with the results from the Query View. Continue to step through the data using the steppers and it will be relatively easy to find the channels that have the fastest flows.



Now that you've found the locations of the channels with the fastest flows (for an example see image below), let's sample the raw data in these regions to have a closer look at the underlying data. Click the mouse in one of the filtered rectangles of the heat map as indicated in image below – if yours does not look exactly the same do not worry just choose one of the filtered rectangles displayed in your display.



This will generate a point attribute graph. The point attribute graph shows the data at a particular point within the data-set. To get a more understandable representation of the vector data contained at this point, the point attribute graph also has a polar representation, this can be switched on using the trails button in the trails tab. This will give an approximate representation of a tidal ellipse at this point.



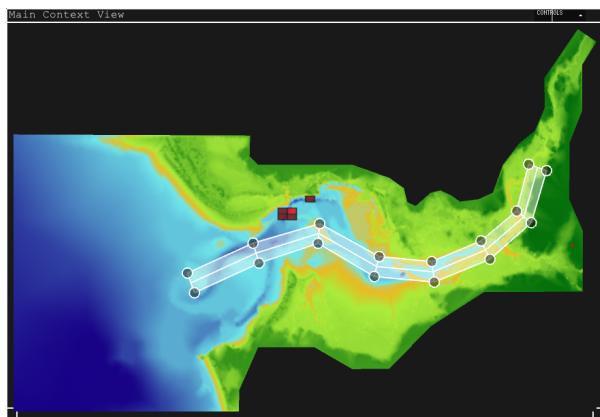
Repeat the process of sampling the data and generating a Point attribute graph for another area highlighted in the search.

Task 3. Flux Calculation

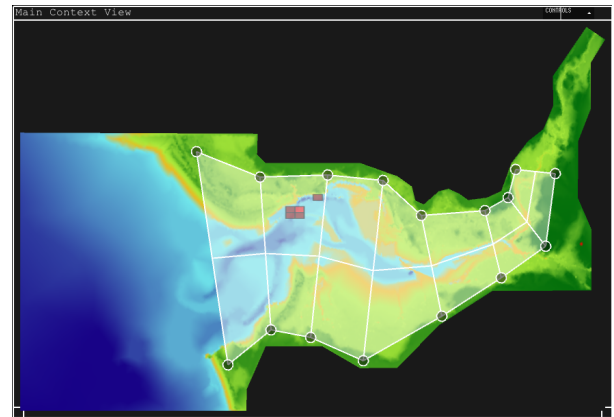
This task will demonstrate one of the more complex analytical features of the system and show how extra value can be added to simulation data through visuals analysis. So for this task we will be generating multiple transects (cross sections) and associated fluxes within the estuary, for quick analysis of the change or variation in flux throughout the estuary. The first step is to switch the analytics to Region Flux mode - this is done by selecting the “Analytics” tab in the Main View control panel, then selecting “Transects”, and finally “Region”. See image below



Creating the boundary uses a point and click strategy, so move the mouse to a location at the mouth of the estuary and click, then move a short distance up the estuary (to the right) and click again. A box will appear, this is the initial boundary definition, repeat the process of moving up the estuary and clicking to extend the initial boundary definition, do this another 5 or 6 times so that the initial boundary definition looks something like image a.



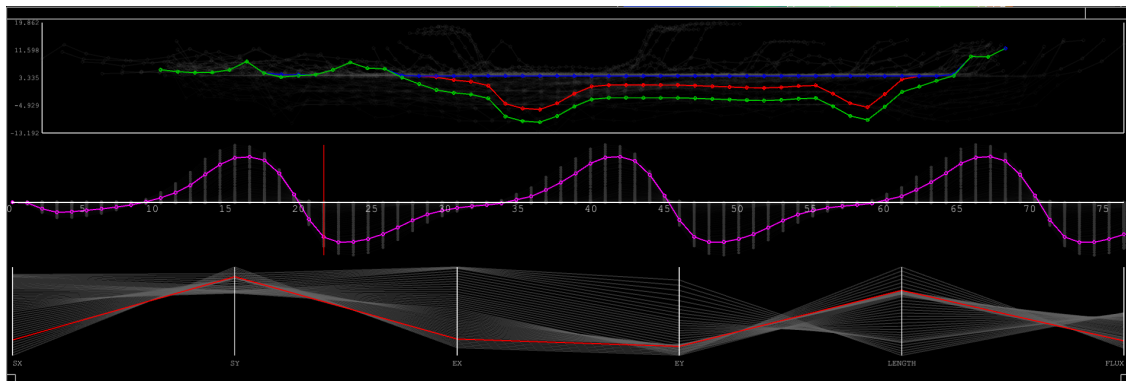
(a)



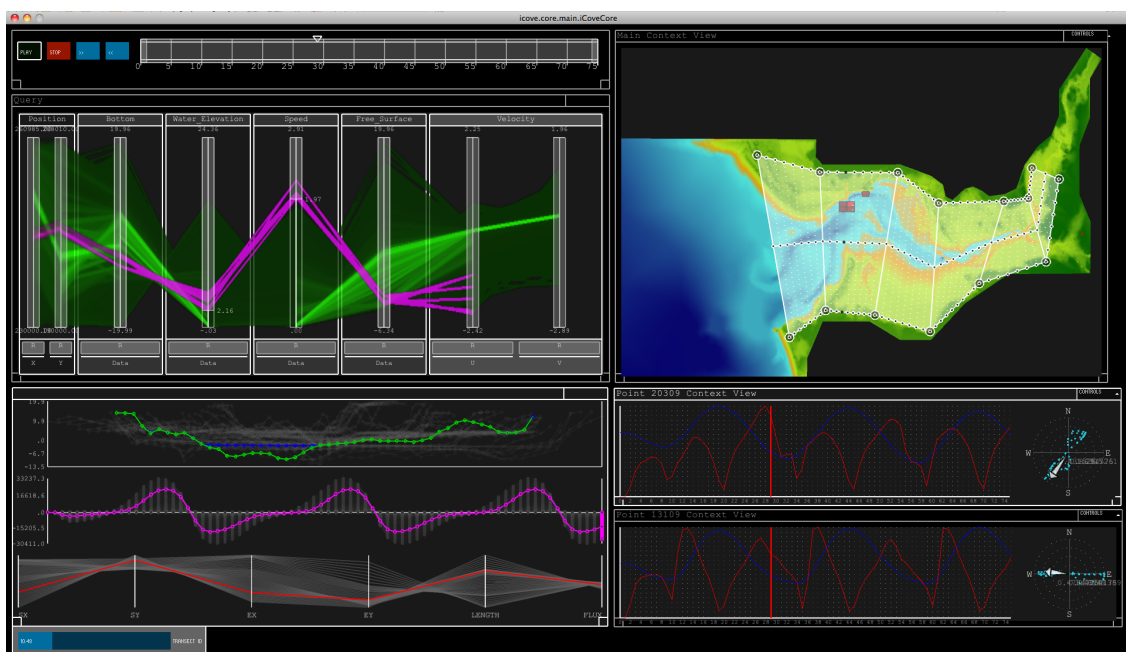
(b)

The next step in creating the boundary involves manipulating it to fit the shape of the estuary. Press the “edit” button in the flux menu of the Main View control panel, this will allow you to manipulate the initial boundary using a click and drag strategy, by picking up the corner handles and moving them to the appropriate location. See image b. Make sure that the boundary stays within the confines of the data i.e. The white boundary line does not cross into the black areas.

Now that the boundary has been defined the last step is to generate the transects and associated fluxes. To do this press the “generate” button located in the flux menu of the Main View control panel. This will take a few seconds to generate the multi transect and flux view.



To view each individual transects and associated flux use the slider in the Transect View control panel. The transect that is currently being displayed is indicated in the Main View by the red line.



Go to next page

Demographic Questionnaire

Please mark your responses with an 'X' in the appropriate box.

1. Gender

Female ☐Male ☐

2. Age Group

- 29 ☐30 ⇒ 39 ☐40 ⇒ 49 ☐50 ⇒ 59 ☐59 + ☐

3. How would you rate your expertise with computers?

No experience ☐Novice ☐Intermediate ☐Advanced ☐Expert ☐

4. How would you rate your expertise with oceanographic analysis?

No experience ☐Novice ☐Intermediate ☐Advanced ☐Expert ☐

5. How would you rate your expertise with visual analytics environments?

No experience ☐Novice ☐Intermediate ☐Advanced ☐Expert ☐

6. How would you rate your expertise with oceanographic analysis environments?

No experience ☐Novice ☐Intermediate ☐Advanced ☐Expert ☐

7. How familiar are you with the use of visual analysis:

Not at all ☐Slightly ☐Moderately ☐Very ☐Extremely ☐

8. Explain your experience with visual analysis:

.....

.....

.....

.....

Usability Assessment – System Usability Scale (SUS)

Please mark your responses with an 'X' in the appropriate box.

9. I think that I would like to use this system frequently

Strongly disagree					Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
1	2	3	4	5	

10. I found the system unnecessarily complex

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

11. I thought the system was easy to use

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

12. I think that I would need the support of a technical person to be able to use this system

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

13. I found the various functions in this system were well integrated

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

14. I thought there was too much inconsistency in this system

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

15. I would imagine that most people would learn to use this system very quickly

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

16. I found the system very cumbersome to use

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

17. I felt very confident using the system

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

18. I needed to learn a lot of things before I could get going with this system

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5

Post-Experiment Questionnaire

19. Describe some positive aspects of VINCA:

.....

.....

.....

.....

20. Describe some negative aspects of VINCA:

.....

.....

.....

.....

21. Briefly describe any type of enhancement that you would like to see in VINCA:

.....

.....

.....

.....

22. Explain whether you think VINCA will be a useful tool for aiding coastal oceanographic researchers knowledge discovery?

.....

.....

.....

.....

23. What is your opinion of using an interactive visual oceanographic analytics environment, such as VINCA:

.....

.....

.....

.....

24. What is your opinion on the future of interactive visual oceanographic analytic environments?:

.....

.....

.....

.....

25. Any other points:

.....

.....

.....

.....