A comparison of VMS and AIS data: the effect of data coverage and vessel position recording frequency on estimates of fishing footprints

Shepperson, Jennifer; Hintzen, Niels T.; Szostek, Claire; Bell, Ewen; Murray, Lee; Kaiser, Michel

ICES Journal of Marine Science

Published: 01/05/2018

Peer reviewed version

A comparison of VMS and AIS data; the effect of data coverage and vessel position recording frequency on estimates of fishing footprints

Jennifer L. Shepperson[1,3], Niels T. Hintzen[2], Claire L. Szostek[1], Ewen Bell[3], Lee G. Murray[1], Michel J. Kaiser[1]

[1] School of Ocean Sciences, Askew Street, Menai Bridge, Anglesey, LL59 5AB

[2] Wageningen Marine Research, PO Box 68, 1970AB IJMUIDEN

[3] Centre for Environment, Fisheries, and Aquaculture Science, Pakefield Road, Lowestoft, Suffolk, NR33 0HT

Corresponding Author: Jennifer L. Shepperson, j.shepperson@bangor.ac.uk, Tel: +44 1248 388501

Abstract

Understanding the distribution of fishing activity is fundamental to quantifying its impact on the seabed. Vessel monitoring system (VMS) data provides a means to understand the footprint (extent and intensity) of fishing activity. Automatic Identification System (AIS) data could offer a higher resolution alternative to VMS data, but differences in coverage and interpretation need to be better understood.

VMS and AIS data for individual vessels in the English Channel scallop fishery were compared. There were substantial gaps in the AIS data coverage; AIS data only captured 26% of the time spent fishing compared to VMS data. The amount of missing data varied substantially between vessels (45-99% of each individuals' AIS data were missing). If data were reduced to include only directly temporally matching data (i.e. only for segments of time in which there were VMS and AIS records for a vessel), the time and extent of fishing was very similar between the two types of data, although a slightly greater amount of fishing effort (+2.6% fishing hours) was still determined by VMS data compared to AIS data. Using a cubic Hermite spline interpolation of VMS data provided the greatest similarity in the extent of fishing predicted by directly matching VMS and AIS data, but the scale at which the data were analysed (i.e. size of the grid cells) had the greatest influence on the estimates of fishing extent and intensity compared to the data type (AIS, VMS) and data treatment (point, straight line interpolation, or cubic Hermite spline interpolation).

The present gaps in coverage of AIS may make it inappropriate for absolute estimates of fishing extent and intensity. VMS already provides a means of collecting more complete fishing position data, shielded from public view. Hence, there is a clear incentive to increase the VMS poll frequency to calculate more accurate fishing footprints, which would ultimately benefit both fishers and scientists.

Key Words: vessel monitoring system; automatic identification system; fisheries; footprint; extent; scallop dredging

1. Introduction

1.1. We need to understand fishing footprints to understand fishing impacts

Physical disturbance by mobile bottom contacting fishing gears is the largest cause of human disturbance to continental shelves in all areas of the world (Foden et al., 2011). In order to understand the extent and consequences of these disturbances it is necessary to have an accurate understanding of the distribution in space and time of that disturbance. For these reasons, the use of vessel tracking data to analyse patterns of fishing effort and the impact of fishing pressure on marine environments is a key area of fisheries science (Mccauley *et al.*, 2016; Russo *et al.*, 2016; Hintzen *et al.*, 2012; Lee *et al.*, 2010).

1.2. VMS data is increasingly used to analyse fishing activity, but has limitations

Satellite based Vessel Monitoring Systems (VMS) were introduced as an enforcement tool, but the resulting data are increasingly important for scientific research and management (Murray *et al.*, 2011, 2013; Lambert *et al.*, 2012). Despite the importance of these data, the temporal resolution of VMS is relatively low in Europe usually with a 1 or 2 hourly poll rate. This poll rate is designed as a compromise between adequate resolution and costs to fishers. Interpolation of VMS data is typically used to fill in the gaps between successive VMS points to produce a continuous track. VMS data can be joined to grids to analyse the extent of, and patterns in fishing intensity, either as raw point data, or as interpolated tracks. However, the methodology used to analyse VMS data can influence the estimates of fishing intensity (Piet and Hintzen, 2012), and the relationship between fishing intensity and epifaunal biomass (Lambert *et al.*, 2012). In particular, the grid cell size used for analysis influences the intensity estimates (Hinz *et al.*, 2013; Lambert *et al.*, 2012; Piet and Quirijns, 2009; Dinmore *et al.*, 2003).

1.3. AIS data has a higher temporal resolution, and could be used to investigate fishing activity.

There has been a recent increase in interest in the potential for using publicly available Automatic Identification System (AIS) vessel tracking data to investigate fishing activity (Mccauley *et al.*, 2016; Russo *et al.*, 2016; Natale *et al.*, 2015). AIS data is openly available to the public, at high resolution, whilst VMS data is subject to strict confidentiality regulations, which can mean only highly aggregated data is available to scientists outside regulatory bodies (Hinz et al., 2013). High resolution analyses of VMS readily demonstrate the problems and limitations of using aggregated data (Hinz et al., 2013). In order to obtain raw VMS data, non-governmental scientists need to approach each vessel individually, which is impractical to get coverage of the whole UK fleets, and makes extension to the European wide fleet unfeasible. While AIS data may provide a useful alternative source of information with which to understand patterns in fishing activity, the present study reveals the many problems with interpreting AIS data due to gaps in the data.

VMS is mandatory on fishing vessels >12m in length in the European Union for enforcement purposes (EC, 2009), whilst AIS is required on vessels >15m for safety purposes. Nevertheless, aspects of the AIS technology, and legislation, mean fishing activity may not be completely recorded by AIS (McCauley et al., 2016). Thus, whilst access to VMS data is subject to confidentiality issues that degrades its utility for research purposes (Hinz *et al.*, 2013), AIS data has different disadvantages, as it can lack consistent coverage. AIS signals are recorded in a different way to VMS data, in that they are broadcast omni-directionally and can be picked up by receivers on land, or by other vessels, as the system was designed to reduce collisions and offer safety mainly when near other traffic or near ports. If a vessel is out of reach of a land based station, the signal must be transmitted from vessel to vessel until it reaches a land station. In areas with relatively low vessel densities, this could cause gaps in coverage. Signals can also be 'lost' in areas of very high density traffic. In addition, skippers can turn down the power on the AIS, which reduces the range of the signal, further increasing the likelihood of gaps in coverage. McCauley et al., (2016) argue that having an AIS system on board a vessel, but failing to use it properly, should no longer be viewed as legal compliance. It is also possible for skippers to falsify AIS data, and provide incorrect vessel IDs (McCauley et al., 2016), with the vessel identity of AIS signals not subjected to the same validation process by inspection agencies as VMS data. Despite the positional accuracy of AIS data being comparable to VMS data, there can be considerable variation in spatial coverage between different fleets of vessels (Russo *et al.*, 2016).

1.4. High resolution AIS data might improve footprint estimates of scallop dredging

Besides the lower overall fleet coverage of AIS data (i.e. number of vessels with AIS), it would be useful to understand more about the differences in inferred fishing activity between VMS and AIS data, where the coverage is concurrent. There can be considerable gaps in AIS data coverage in space and time within fleets (Russo *et al.*, 2016; Natale *et al.*, 2015), but there has been little comparison of the recorded activity by the two data sources specifically on trips where both VMS and AIS were actively transmitting data, or assessment of individual variation in coverage. In situations where AIS is the primary data source available to scientists, e.g. in areas where no, or only highly aggregated VMS data is available, it is important to know how the conclusions drawn from AIS would correlate with those drawn from VMS data. In addition, an analysis of concurrent VMS and AIS data would enable us to understand better the complexity in patterns of fishing that may be missed by VMS data due to the issue of temporal position frequency. The structure of the VMS and AIS data itself is essentially the same, a file with coordinates, speed, heading, and vessel ID, which means the same data processing techniques can simply be applied. Nevertheless, despite using the same processing techniques, there could be differences in the resulting conclusions due to differences in the way the data were generated. Primarily, AIS data is available at a much higher poll frequency than VMS data. Finer scale patterns in activity may therefore be seen with AIS data, for example, using a longer 2 hourly poll frequency in VMS could miss shorter steaming sections between tows, giving the impression of long continuous fishing activity, potentially overestimating fishing activity. Alternatively, due to the difficulty in accurately interpolating the tracks between 2 hourly position records, the VMS could also lead to an under-estimate of the extent and intensity of fishing.

Understanding this error becomes particularly important when attempting to understand the environmental footprint of different fishing activities. In this paper, the focus is on scallop fishing, which is considered one of the least environmentally compatible forms of mobile bottom contacting fishing gears (Kaiser et al. 2006). European Union Directives such as the Marine Strategy Framework Directive (MFSD) and Good Ecological Status (GES) use the fishing footprint (spatial distribution of fishing activity) as an indicator of ecosystem health. Understanding the distribution of fishing activity is fundamental to understanding and quantifying the impact that fishing has on the seabed (Kaiser et al., 2016). VMS data provides a means to understand the spatial footprint (extent and intensity) of scallop dredging and other fisheries and this has become a research field of its own (Eigaard et al., 2017; Hinz et al., 2013;Lambert et al., 2012;). However, there is a conflict between the requirement for high temporal and spatial resolution data needed for scientific research, and the publicly available lower temporal and spatial resolution data to uphold confidentiality of commercially sensitive data (Hinz et al., 2012, Lambert et al., 2012). In many trawl fisheries the gear is towed for in excess of four hours and often in a single direction with few deviations. In contrast, scallop dredgers can make short tows

(~20 minutes), make tight turns and often tow parallel to their previous tracks, which can make the prediction of trajectories using interpolation methods difficult when the resolution of data is low (Lambert et al., 2012). With scallop dredging activity, higher resolution AIS data may therefore be better able to capture (1) the true footprint of fishing by better capturing the sharp turns made by vessels; and (2) the true fishing effort level, by better capturing the time spent in each activity state (i.e. fishing cf. steaming). This could provide insights into the appropriate treatment of lower resolution VMS data.

Nevertheless, this benefit of higher poll frequency in AIS data may be counteracted by gaps in coverage. Before AIS data can be used as a data source for management, these differences in coverage and interpretation must be understood and addressed. This paper seeks to address this gap in understanding, by comparing the fishing activity of vessels in the English Channel Scallop fishery, on days for which it was possible to obtain both VMS and AIS position records.

1.5. Aims

The main aims of this paper were to: (1) determine the relative coverage of AIS data in relation to VMS data at both the fleet and an individual vessel level; and (2) for matching data (from the same vessels in the same time period), determine whether the fishing extent and intensity predicted by three common methods of VMS data analysis (point density, straight line interpolation (Stelzenmuller et al. 2008), and cubic Hermite spline interpolation (Hintzen et al. 2010)) showed a comparable accuracy to the higher poll frequency AIS data. Conclusions were drawn about the accuracy of the two data sources (AIS vs VMS) for estimating fishing extent and intensity, considering the relative coverage of each data type, differences in temporal poll frequency, and the level of spatial aggregation used during data analysis.

2. Methods

2.1. Data Coverage

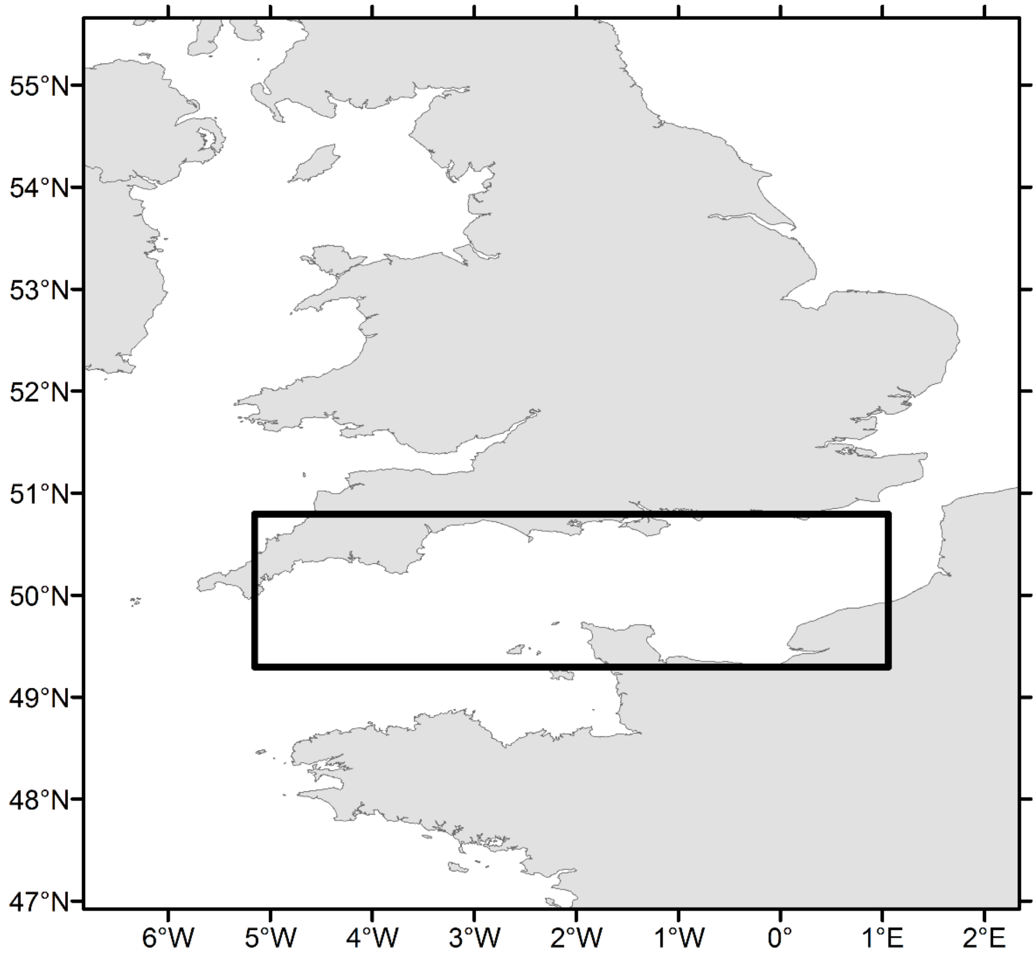


Figure 1. The spatial window in which the VMS and AIS records were recorded. The actual positions of data points are concealed for confidentiality.

VMS and AIS data were obtained for vessels in the English Channel scallop fishery, in the calendar year 2012, in the spatial window shown in Figure 1. Eight scallop dredgers (all >15 m L.O.A.) in the

English Channel gave permission for their raw VMS data to be used in this analysis. The VMS data included vessel identification data, position, time, speed and heading. AIS position, time, speed, heading, and vessel identification data for the same eight vessels over the same time period were obtained from the company AstraPaging Ltd (http://www.astrapaging.com/), a private AIS data provider. The Maritime Mobile Service Identity (MMSI) field, a nine digit number uniquely identifying a ship radio station installed on each vessel, was used to request the AIS data for the eight vessels. VMS data was provided at a poll frequency of approximately 2 hours, and the AIS data was provided at a poll frequency of approximately 5 minutes.

AIS coverage can vary between fleets (Natale et al., 2015), and has also been shown to capture a smaller amount of fishing activity than VMS data (Russo et al., 2016). Therefore, following the initial assessment of data coverage, data were excluded from the analysis if they could not be matched to corresponding VMS or AIS data from that same vessel over the same time period. Initially, for each date, a vessel's VMS data were removed from the analysis if that same vessel had not also recorded AIS data on that date, and vice versa; therefore the term 'comparable date' is used to signify a date on which a particular vessel had recorded both VMS and AIS, which generated 'comparable data'. Nevertheless, even if there were some VMS and AIS for a vessel on a particular day/trip, either dataset may not be complete within the trip. Thus a further category of matching data was identified, by extracting trips where the ratio of the duration of VMS:AIS points was between 0.8 and 1.2, i.e. there was less than 20% mismatch in the duration of VMS compared to AIS, so substantial sections of either data were not missing within a trip. There were therefore 2 categories of data: comparable data, which refers to trips for which there is some VMS and AIS for that vessel, but within trip completeness has not been quantified; and matching data, which refers to trips for which the ratio of VMS:AIS is between 0.8 – 1.2, meaning that only a subsection of a trip may be included, which has more complete data from both sources. Only comparable or matching data were used in the comparisons of fishing activity, extent, intensity, and track interpolation.

2.2. Data Processing

Both VMS and AIS datasets were subjected to the same data cleaning and processing strategy, using the *VMStools* packages in R (Hintzen *et al.*, 2012). Duplicate VMS records and records close to (within 1km of) port were removed, along with erroneous position records allocated to land (Lee *et al.*, 2010). Following examination of the frequency distribution of the recorded speeds, position records between 1 – 5knots were classed as fishing activity (Figure 2); these fishing points were either analysed as raw point data, or interpolated to reconstruct higher resolution estimates of the fishing tracks (Section 2.3). The level of data retention of VMS and AIS data was recorded at each stage of data cleaning and processing, to identify differences and similarities in the data, and identify any substantial loss of data.

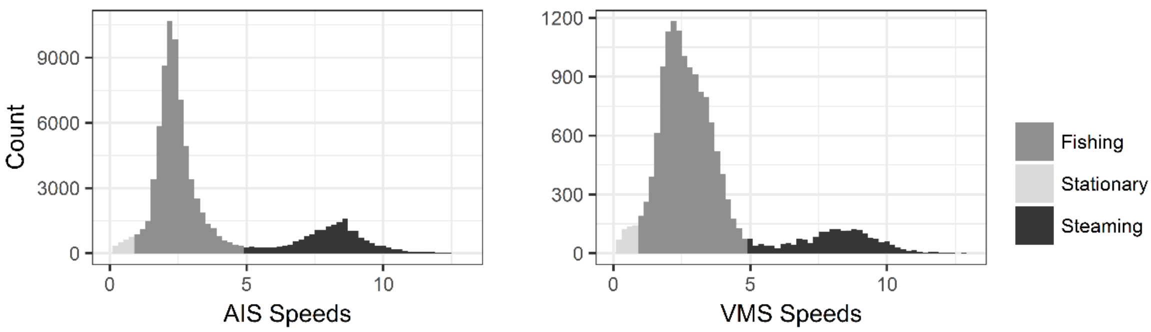


Figure 2. Speed frequency distributions of vessel data.

2.3. Interpolation of tracks

VMS and AIS data can be analysed in the raw point data format, or vessel tracks can be reconstructed using a straight line (SL) interpolation, or a cubic Hermite spline (cHs) interpolation

(Hintzen et al., 2012). Succinctly, the cHs method uses information on vessel position, heading and speed at times t and t+1 to define a trajectory. The combination of speed and heading are represented by vectors, and vector length is multiplied by a parameter *fm* that influences the curvature of the interpolations (Lambert et al., 2012). The VMStools package in R provides a function for *fm* parameter optimisation; the high resolution AIS data was used to determine the optimal *fm* parameter for cHs interpolation of the VMS data. CHs interpolation of the AIS data was not possible, as there was no higher resolution data for the optimisation process, nevertheless the AIS points were recorded at a high 5 minute poll frequency, so a SL interpolation would give a sufficient level of spatial detail in the tracks. The SL interpolation of the AIS data can be assumed as the most robust estimate of the path taken by the vessels due to its high poll frequency.

2.4. Data Analysis

The number and proportion of points classed as fishing activity were compared between data types and vessels, to identify differences between the data types, and whether these differences varied between individual vessels. To investigate the footprint of fishing activity, points that were classed as fishing activity were joined to spatial grids of 1km, 3km, 5km, and 10km in cell size. These grids were used to calculate the fishing extent and intensity. The interpolated tracks were turned into a series of points approximately every 30 seconds along the track, to analyse in the same way as the raw point data.

The extent of fishing was calculated by counting grid cells which had at least 1 fishing point in them using each data type and interpolation method. The intensity of fishing was compared by calculating the area swept in each grid cell, by summing the area swept per point in the cell, using each data type and interpolation method. Area swept per point was calculated as:

$$\text{Area swept (km}^2\text{) = speed (km/h) * time fishing (h) * total dredge width (km)}$$

where the total dredge width is assumed to be 0.018km for all vessels, corresponding to the width of 24 individual dredges each measuring 0.75m across. In this analysis a fixed dredge width was used as the actual dredge width was unknown. Total dredge width has been shown to increase with vessel length (Eigaard et al., 2016), but as the estimates of fishing effort in this analysis are relative and not absolute (as only a subsection of vessels were included), a fixed dredge width was considered appropriate.

In addition, the swept area ratio (SAR) was calculated for each grid cell (Gerritsen *et al.,* 2013). The SAR indicates what proportion of the cell has been dredged at least one time, calculated as:

$$\text{Swept area ratio = Area swept (km}^2\text{) / area of cell (km}^2\text{)}$$

In this study, one year of AIS and VMS data were used, therefore these SAR values relate to a one year time period. A SAR of 1 therefore indicates that on average each part of a grid cell has been dredged one time over the year, a SAR of 2 indicates the whole cell has been swept twice, a SAR of 0.5 indicates that on average half of the cell has been dredged one time. Each VMS interpolation method (point data, SL interpolation, and cHs interpolation) was compared to the SL AIS interpolations (assumed as the truest fishing tracks) and AIS point data.

2.5. Data Confidentiality

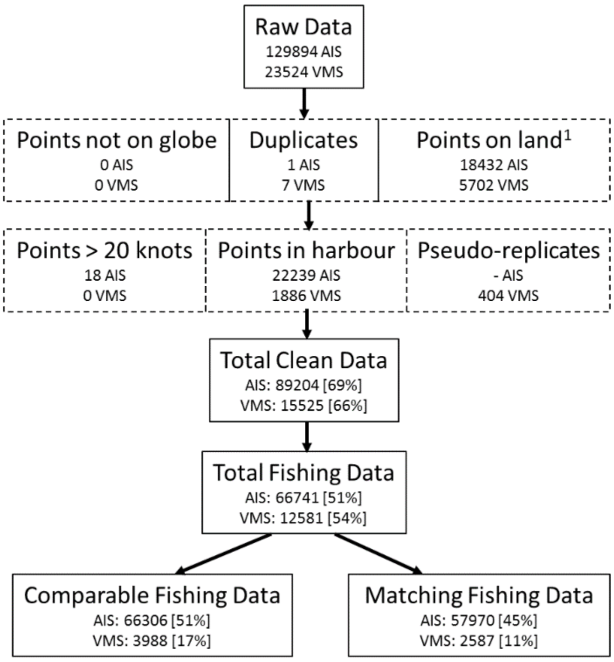
Vessels that contributed to this study are anonymous throughout the analysis, as VMS and AIS data are commercially sensitive, and therefore confidentiality is an important issue. VMS data were provided by fishermen under the condition that the location of fishing activity would not be displayed, therefore the spatial reference has been removed from any maps. The same level of confidentiality has been afforded to the AIS data.

3. Results

3.1. How do the basic VMS and AIS datasets compare?

The eight studied vessels recorded 129894 AIS points, and 23524 VMS points during the calendar year 2012. After cleaning, there were 89204 AIS and 15525 VMS points remaining (Figure 3) representing a 69% and 66% retention respectively. 51% of the raw uncleaned AIS data and 54% of the raw uncleaned VMS data were retained as representing fishing activity, comprising 66741 and 12581 points respectively. This represented 75% of the clean AIS data and 81% of the clean VMS data.

For the comparable trip data (i.e. records which had corresponding AIS or VMS for that vessel on that day) there were 66306 AIS points and 3988 VMS points from seven vessels (thus one vessel was excluded from further analysis). This retained 51% of the raw AIS data and 17% of the raw VMS data, or 99% of the cleaned AIS data and 32% of the cleaned VMS data that represented fishing points. For the matching data (i.e. only trips with a ratio of VMS:AIS within the threshold 0.8 – 1.2, to reduce missing data within trips), this was reduced to 57970 AIS points and 2587 VMS points (45% and 11% of the raw data, 65% and 17% of the clean data, respectively). A substantial amount of AIS data were therefore missing, i.e. there were a lot of days for which there were VMS data but no corresponding AIS data, but there were comparable VMS records for almost all AIS records. When the data were reduced further to only trips with a high VMS:AIS ratio, 13% of the comparable AIS data and 35% of the comparable VMS data were removed. This indicates that whilst there were more missing AIS data within trips, there were also missing VMS data within trips. The average time interval between all VMS points was 130 minutes, and between AIS was 13 minutes, but when only fishing points were used, the average time interval between VMS points was 114 minutes, and between AIS points was 5 minutes.



[1] Points classed as being on land are often just in harbour, but the GIS map used to define the land/harbour area may not be high enough resolution to accurately distinguish the harbour boundaries.

Figure 3. Preparation and cleaning of AIS and VMS data. Dashed boxes indicate data that is removed, solid boxes indicate retained data.

3.2. How does the assignment of fishing activity compare between VMS and AIS data?

The raw VMS data indicated 29701 hours of fishing, but the raw AIS data only estimated 7647 hours of fishing, which constitutes a substantial gap in the coverage of AIS data. Despite using data from vessels which have both VMS and AIS on-board, the AIS data only captured 26% of the time spent fishing compared to VMS data.

The proportion of time each vessel spent fishing, steaming, and effectively stationary was then compared between the comparable AIS and VMS data for each vessel on each day. Overall, using comparable trip data, the AIS data indicated a total of 5661 hours fishing by all vessels across the study period, and the VMS data a total of 7751 hours, suggesting 2090 extra fishing hours with the VMS than the AIS data. The AIS data indicated a total area swept (calculated as area swept per fishing point: area swept = speed * dredge width * time) of 469km^2, and the VMS data indicated a total area swept of 651km^2.

For the matching data, the AIS data indicated a total of 4924 hours fishing across the study period, the VMS a total of 5053 hours, suggesting 129 extra hours fishing by the VMS data. The AIS data indicated an area swept of 405km^2, and the VMS an area swept of 406km^2. If using all available data, the overall extent of fishing is under-estimated by 74% by AIS data compared to VMS data, using comparable data (i.e. trips for which there is some VMS and some AIS data) the overall extent of fishing is under-estimated by 39% when using AIS data, but if using only data that directly matches in time, the extent of area affected by fishing was very similar.

3.3. How does the data coverage and activity assignment differ between individuals?

A substantial amount of AIS data were missing, however the amount of missing data differed between individual vessels (Table 1). In all cases, more AIS data were missing than VMS data. Thus when retaining only comparable trip data, for some vessels the removal of VMS data was large, e.g. a reduction from 2176 points to 151 points (93% loss); the smallest loss was 34%. In contrast, the greatest loss of AIS data due to having no corresponding VMS data was 2%, and the smallest loss was nil. However, when using only trips with matching data (a high ratio of VMS:AIS), considerably more data were removed; 45-99% of VMS data, and 7-51% of AIS data. This suggests that overall there were substantially more AIS data missing, but there were some trips with considerable amounts of VMS data missing as well.

The comparable trip VMS data gave a higher estimate of time spent fishing and area swept than the AIS data for all individuals (Table 1). The magnitude of this difference varied between 18-92% between individual vessels. When using the more closely matched data, the time spent fishing for each individual varied from +2% to -26%, and the area swept from +3% to -17%, but the values of the absolute difference were relatively low.

Table 1. Number of VMS and AIS points per vessel when using all available data, only comparable data, and only matching data. The time spent fishing and area swept columns indicate the total for each vessel using comparable and matching VMS and AIS data. The % dif column indicates how much smaller the AIS value was than the VMS value.

Comparable Data

ID	Number of fishing points (all available)		Number of Comparable Fishing points		VMS lost	AIS lost	Time fishing (hours)		% dif	Area swept (km^2)		% dif
	VMS	AIS	VMS	AIS			VMS	AIS		VMS	AIS	
1	1539	7249	616	7227	60%	0.3%	1171	646	-45%	115	64	-44%
2	1963	942	251	942	87%	0.0%	514	91	-82%	57	10	-82%
3	2176	1038	151	1036	93%	0.2%	267	96	-64%	25	9	-64%
4	1581	18086	1023	18075	35%	0.1%	1957	1521	-22%	149	117	-21%
5	1798	12291	701	12241	61%	0.4%	1328	1043	-21%	100	82	-18%
6	946	13369	620	13322	34%	0.4%	1263	1120	-11%	105	94	-10%
7	1921	13766	626	13463	67%	2.0%	1251	1145	-8%	98	93	-5%

Matching Data

ID	Number of fishing points (all available)		Number of Matching Fishing points		VMS lost	AIS lost	Time fishing (hours)		% dif	Area swept (km^2)		% dif
	VMS	AIS	VMS	AIS			VMS	AIS		VMS	AIS	
1	1539	7249	259	5448	83%	25%	497	477	-4%	49	47	-4%
2	1963	942	29	460	99%	51%	58	43	-26%	6	5	-17%
3	2176	1038	40	753	98%	27%	77	69	-10%	7	7	
4	1581	18086	735	16202	54%	10%	1410	1362	-3%	107	105	-2%
5	1798	12291	463	10209	74%	17%	889	869	-2%	67	69	+3%
6	946	13369	521	12464	45%	7%	1042	1048	+1%	87	88	+1%
7	1921	13766	540	12434	72%	10%	1079	1055	+2%	84	85	+1%

The correlation between the number of VMS and number of AIS points per day varied between vessels, and depended on the data treatment (Figure 4), but there was a considerable amount of missing fishing data for all vessels. Using comparable trip data only, the correlation between the duration of VMS and AIS fishing records per vessel per day improved significantly. Nevertheless, the duration of AIS data is slightly lower than the VMS data for all vessels. For some vessels there is considerably less AIS data than expected on comparable days, suggesting some gaps in coverage within a trip. The dashed line indicates the 1:1 ratio between VMS and AIS data; matching data were identified as trips that had a ratio of 0.8 – 1.2, therefore matching data showed a strong correlation between the duration of VMS and AIS points by definition (Figure 5c).

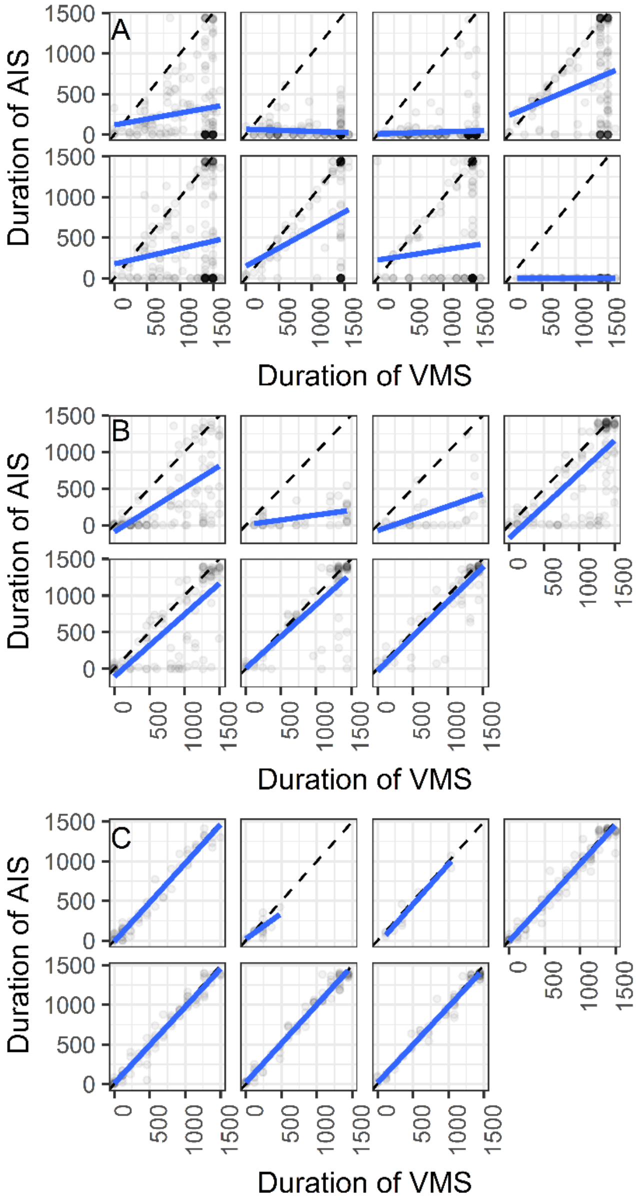


Figure 4. The correlation between the duration of VMS and AIS fishing points per vessel per day (A), the correlation between the duration of VMS and AIS fishing points using only data on comparable days (B), and the correlation between the duration of VMS and AIS fishing points using only data that were classed as matching (i.e. high ratio AIS:VMS) (C). Blue solid line indicates the correlation between the numbers of points per day, black dashed line indicates the 1:1 correlation. Points are translucent such that darker areas indicate a concentration of points.

3.4. How does the spatial footprint of fishing compare between VMS and AIS data?

3.4.1. Interpolation of VMS and AIS fishing tracks

A straight line interpolation of the matching AIS data was used to create the best estimate of the vessels' tracks. The matching VMS data was interpolated using both the straight line (SL) and cubic Hermite spline (cHs) approach. Parameter optimisation in *VMStools* gave an *fm* parameter of 0.19, which suggested that a non-linear interpolation gave a more appropriate interpolation of the VMS tracks than a SL interpolation, based on the distance between the interpolated VMS positions and the higher frequency AIS positions.

Three days of data were selected at random to display individual interpolated tracks (Figure 5). From visual observation of the three different types of track interpolations (Figure 5), when using a cHs interpolation, the AIS fishing tracks display shorter sections of fishing activity than with the VMS fishing tracks. The low temporal resolution of the VMS data (2 hours) may have forced the interpolations to be continuous such that they potentially missed sections of time in which fishing did not occur. In contrast, as the AIS data has a higher temporal resolution (5 minutes) it can account for shorter periods of fishing and steaming within this 2 hour window.

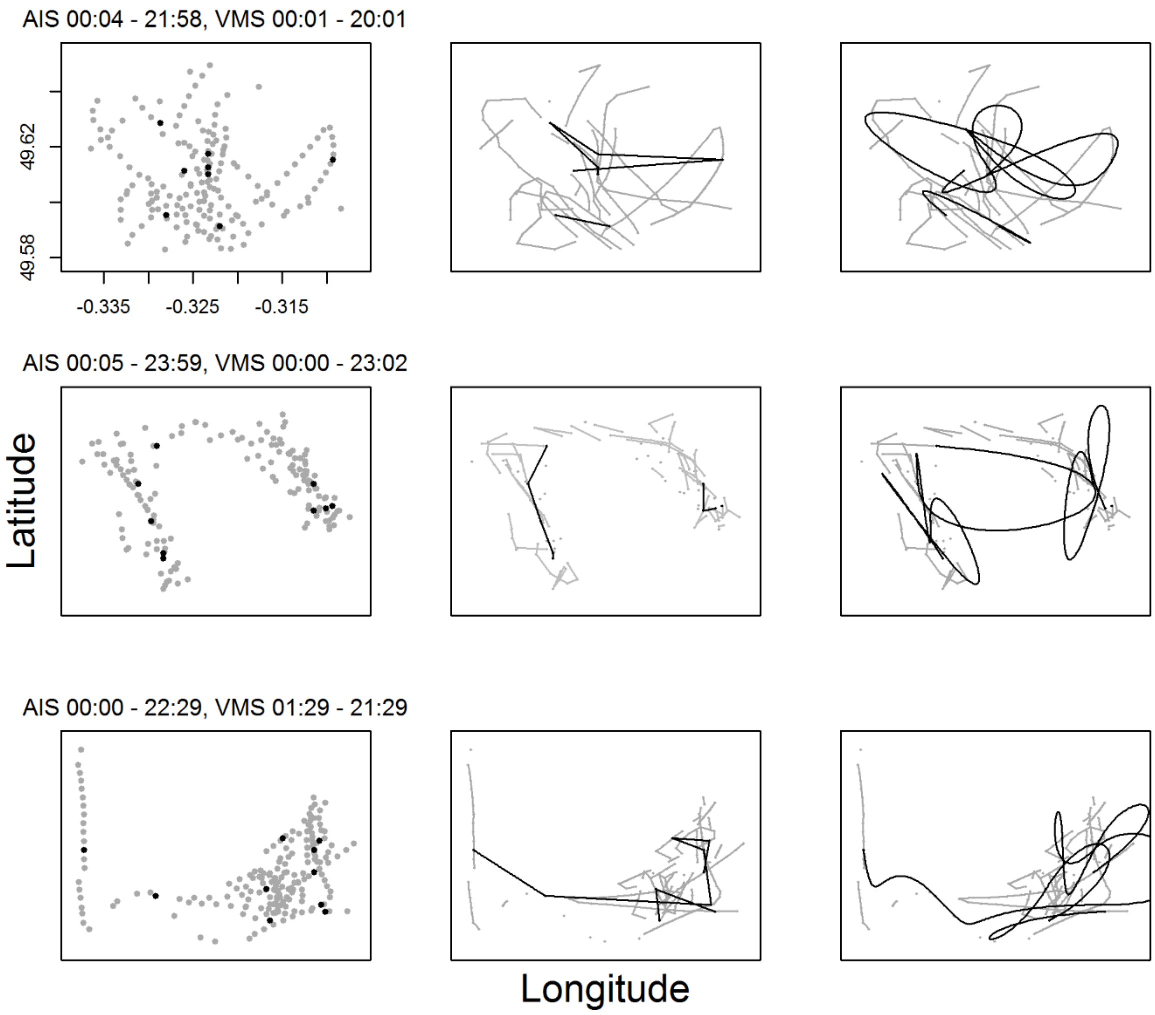


Figure 5. Each row = one trip by one vessel. The time span of each data set is displayed for each row. Black = VMS, grey = AIS. The first column displays the point data, the second column displays straight line (SL) interpolated data, and the third column displays the cubic Hermite spline (cHs) interpolated data for the VMS data and straight line interpolated for the AIS data. The extent of fishing activity appears to be underestimated using a straight line interpolation, but they show a similar extent with a cHs interpolation, albeit with a lower resolution in the VMS tracks.

3.4.2. How do the estimates of fishing extent and intensity differ between data types?

To investigate how the spatial scale of analysis (i.e. the grid cell size used to aggregate data) influenced the estimates of fishing extent, points that were classed as fishing (cf. steaming or stationary) were joined to spatial grids of varying cell size. Only matching data were used for this part of the analysis. The estimates of fishing extent presented here are not absolute efforts of scallop dredging activity in the English Channel, as data was only available from 8 vessels, and only subsections of data which are temporally matched between AIS and VMS data were used. The analysis is therefore relative, providing a comparison between the data types, and not absolute estimates of fishing extent and intensity. Increasing the grid cell size for analysis increased the estimated extent of fishing activity (Figure 6, Table 2). The total extent estimates were most similar when using the cubic Hermite spline interpolation method for the VMS data.

In this analysis an assumption was made that the straight line interpolation of the AIS data using a grid with 1 x 1km grid cells provided the most accurate estimate of the extent of fishing, as it was the highest resolution data treatment. In this case, using a grid with 10 x 10km grid cells substantially overestimated the extent of fishing. The cHs interpolation of the VMS data at 1km cell resolution gave a very similar value for the extent of fishing as the SL AIS at 1km cell resolution, but the point VMS data at 1km cell resolution greatly underestimated the extent of fishing activity. This suggests that the poll frequency of the VMS data is too low to give an accurate estimate of fishing extent unless either points are either aggregated to a low resolution grid, or if using a high resolution (e.g. 1km resolution) grid the points should be interpolated using a cHs interpolation. The method of data treatment had a substantial impact on the estimated extent of fishing when using a high resolution grid (1km resolution) (Figure 6). At a coarse resolution (10km resolution) the method of VMS data treatment had less impact on the estimate of extent, but the overall extent estimate was significantly higher than when using a 1km grid resolution. The data treatment had little effect on the estimate of extent determined from the AIS data (Figure 6).

The amount of the study area that was perceived as un-trawled (i.e. swept area ratio (SAR) ≈ 0) decreased as the grid size increased (Figure 7). Generally, there was very little difference between the different data types (VMS, AIS) or data treatments (point data, SL interpolation, cHs interpolation). There were slight differences in the areas trawled very lightly (SAR < 0.25) when using a high resolution analysis grid (1km cells). Only directly matching data from 7 vessels over 12 months was used, therefore although the differences are small, if scaled up to a fleet and using more complete data, these differences would be multiplied. A 3km grid showed little variation in the pattern of fishing intensity between each data treatment.

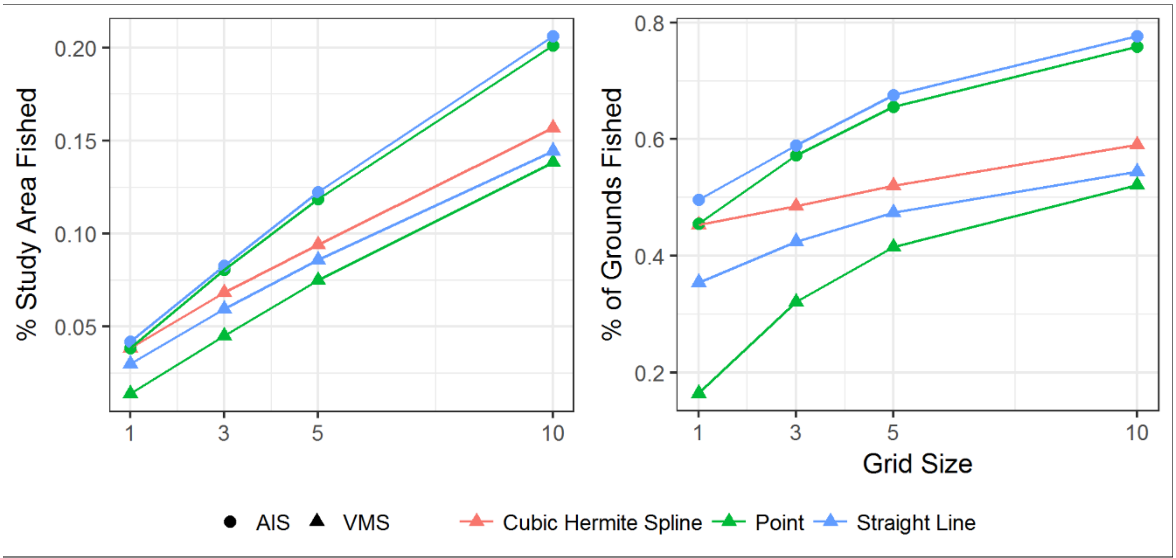


Figure 6. Comparison of the extent of fishing activity across the whole study area using each interpolation method, and the extent of fishing activity across fishing grounds (i.e. only areas which had recorded fishing activity by any of the data).

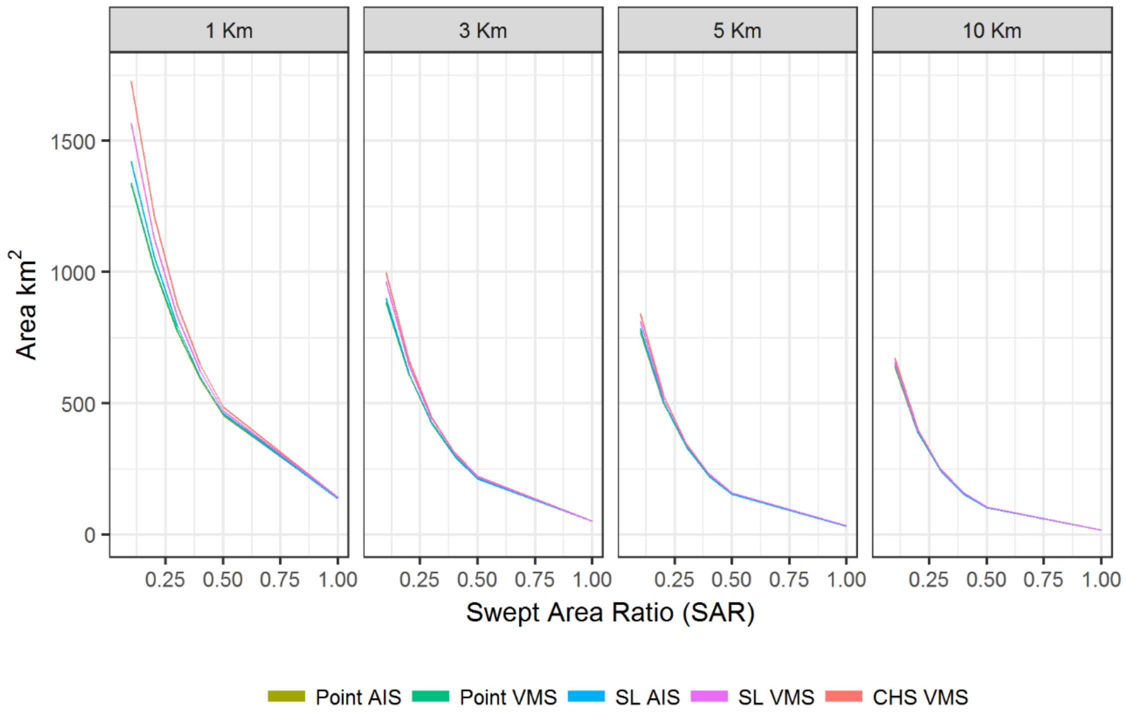


Figure 7. The area of seabed swept 0 – 1 times during the study period (swept area ratio, SAR), by grid size (1, 3, 5, and 10km), using each method of data interpolation (point, SL and cHs), with only matching VMS and AIS data. Zero SAR data is not included.

Table 2. Comparison of VMS and AIS data using different interpolation methods at different spatial scales. The Difference (km²) column indicates how much larger or smaller the extent was using AIS data, in km^2, and the percentage difference indicates this difference as a percentage of the study area or the fishing grounds.

		Direct match VMS			Direct Match AIS					
	Grid Size	Extent (km²)	Extent as % of study area	Extent as % of grounds	Extent (km²)	Extent as % of study area	Extent as % of grounds	Differemce (km²)	% difference (study area)	% difference (grounds)
P	1km	1127	1	16	3123	4	46	1996	2	29
	3km	3654	5	32	6516	8	57	2862	4	25
	5km	6075	8	41	9600	12	66	3525	4	24
	10km	11200	14	52	16300	20	76	5100	6	24
SL	1km	2432	3	35	3407	4	50	975	1	14
	3km	4833	6	42	6714	8	59	1881	2	17
	5km	6950	9	47	9900	12	68	2950	4	20
	10km	11700	14	54	16700	21	78	5000	6	23
cHs	1km	3111	4	45				296	0.4	4
	3km	5526	7	49				1188	1	10
	5km	7625	9	52				2275	3	16
	10km	12700	16	59				4000	5	19

4. Discussion

Neither VMS nor AIS were designed as tools to aid our understanding of fisheries science, but whilst both data sources offer valuable data for understanding fishing activity, their use must be based on an informed understanding of the most appropriate way to process and interpret the data. Of greatest importance, AIS data can have a substantially lower coverage than VMS data and thereby provides a potential underestimate of overall activity (Russo et al., 2016, Natale et al., 2015). This study has contributed to this understanding by demonstrating how even when only looking at VMS and AIS from the same individual vessels (i.e. accounting for the lower fleet coverage of AIS), and even reducing the data to only include fishing trips with concurrent AIS and VMS data logging, there were still substantial gaps in AIS data coverage.

4.1. Can AIS data be an appropriate fisheries monitoring tool?

Whilst AIS data is attracting attention as a promising tool for analysing fishing activity, because it provides publicly available high resolution vessel tracking data, the gaps in its coverage present a substantial hurdle. In this study considerable gaps in the coverage of AIS data compared to the VMS data were found, which concurs with similar studies (Russo et al., 2016, Natale et al., 2015). A similar proportion of VMS and AIS data were retained following data cleaning (i.e. removing incorrect coordinates, points on land, points in harbour etc.), but after only matching data were retained, only 11% of VMS fishing points had corresponding AIS data. This coverage also varied substantially between individual vessels, with 1 – 55% retention of VMS data when using only matching data; for one vessel, there were no directly matching AIS data for 99% of the VMS. For the whole of the studied fleet, this translated as AIS data capturing only 26% of the duration of fishing activity captured by VMS data in 2012. Clearly, this is a substantial gap in the AIS data coverage, and would be a cause for concern if using AIS data to analyse fishing activity without VMS data. It is likely inappropriate to use AIS data for absolute estimates of fishing extent or intensity, because the gaps in coverage are too substantial. It may be possible to use AIS for comparative studies, making an assumption that the gaps in coverage are relatively consistent over short time frames and on average between fleets. Nevertheless, more comparative studies of AIS and VMS may be needed, to better understand the differences and gaps in coverage. Caution would be required when using historical AIS data compared to more recent or future AIS data, where compliance, increased uptake, or technological advances may potentially lead to higher coverage in recent years.

Aside from addressing the more technical limitations to spatial coverage (Russo et al., 2016; Natale et al., 2015), these results support the suggestions from McCauley et al., (2016) that to gain the full benefits of AIS data for fisheries science, policy interventions would also be required, for example, to reduce the gaps in AIS coverage from fishers turning down the AIS transmitter. There are, however, two principle reasons why fishers may wish to conceal their activity from an AIS system; 1) detection avoidance whilst undertaking illegal fishing activity, and 2) preventing other fishers from using AIS data to identify prime fishing grounds. Real time AIS data is openly available to view, including to other fishermen so it is understandable that fishermen may be reluctant for such high resolution tracking data to be openly and instantaneously available, due to the commercial sensitivity of such data. It is difficult to envisage how this issue would be overcome. If it became a legal requirement that the AIS unit was functioning at full strength and an openly available high resolution high coverage dataset of fishing activity was achieved, it could lead to conflict or negative economic consequences for fishers.

4.2. The Impact of grid cell size and interpolation method

This analysis was designed to highlight any systematic under- or over-estimation of fishing effort due to the low temporal frequency, through comparing it with the higher temporal frequency AIS data.

When VMS and AIS data were directly matching in temporal coverage, the scale of analysis (i.e. size of the grid cells) had the greatest influence on the estimates of fishing extent and intensity compared to the data-type (AIS, VMS) and data treatment (point, SL, cHs). The SAR estimates were very similar, which is perhaps not surprising for temporally matching data over the same time period, unless there were inaccuracies in the GPS positions themselves. Therefore, if the data were truly concurrent, AIS could provide a viable alternative to VMS, and offer the benefit of higher temporal resolution. Nevertheless, the most considerable differences in the data come from the substantial and significant gaps in coverage of AIS data.

When using directly matching data, the VMS data still slightly overestimated the time spent fishing compared to the AIS data, by 129 hours. These differences in recorded fishing duration could be related to the bias from a 2 hourly ping rate of VMS data. Nonetheless, even with directly matching data this constituted 10 extra days of fishing with VMS data (assuming 12 hours continuous fishing per day), or 3% of the total fishing hours recorded. At a small scale of 7 vessels over 1 year of fishing, this is a small value, but scaled up to a whole fleet across multiple years, this could represent substantial fishing effort. The temporal resolution of the VMS data may have missed the shorter hauling/moving sections in between scallop dredge tows, which would be less than two hours in duration, and could therefore overestimate the fishing effort. Identifying fishing activity is much more sensitive under VMS than AIS data, as a 2 hour time frame could be incorrectly classified as fishing or non-fishing, whilst with AIS only seconds or minutes would be misclassified. Technology that would provide information on when the gear is in the water would improve estimates of scallop fishing activity and would address on of the issues of poll frequency for VMS.

4.3. Predicting unpredictable scallop dredging activity

The issues associated with interpolating low poll frequency VMS data are particularly relevant to scallop dredging, due to the idiosyncratic movements of the vessels, such as sharp turns and re-towing over the same areas, which may be missed by the lower resolution VMS data (Lambert et al., 2012). The optimal *fm* parameter (a parameter that determines how much the tracks should curve in a cHs interpolation) was markedly different in this study compared to that undertaken by Lambert et al., (2012) with scallop dredgers in the Isle of Man. They concluded that the optimal *fm* parameter was close to zero, i.e. a straight line. Here an optimal fm parameter of 0.19 was reported, which is considerably different from a straight line. For scallop dredgers, and in other fisheries with shorter haul durations, the availability of a higher resolution dataset is perhaps even more important (c.f. trawlers) due to the unpredictable movement patterns associated with these fisheries.

The results presented here therefore relate to scallop dredge fisheries, which represent an atypical part of the European (and global) fishing fleet in terms of towing behaviour characteristics. It may not be appropriate to extrapolate the comparison of matching AIS and VMS data to other mobile bottom contacting fishing gears, such as otter trawlers and beam trawlers. As other trawl vessels perform longer tows, and are characterised by fewer idiosyncratic movements and sharp turns, the discrepancies in estimates from VMS and AIS may be smaller; similar work to compare AIS and VMS data for other types of vessels would enable a better understanding of the value of AIS data for fisheries research.

When providing VMS data in an aggregated format, it is likely not appropriate to provide it at a grid resolution less than 3km by 3km, unless the data has been interpolated, because the extent could be underestimated. It should, however, be noted that the underestimate of fishing extent with point data may be exacerbated by the small sample sizes; where a larger amount of data is available a better estimate of the footprint may be produced without interpolation. Providing data aggregated at a low resolution can overestimate the extent of fishing activity, and interpolation of 2 hourly VMS pings may be unable to resolve the more complex fishing tracks that some scallop vessels follow. When considering the influence of these factors on fishery management options, there is the need for a balance between data cost (i.e. resolution) and accuracy of the results. Over-estimation of the impacted area may result in more draconian action than is necessary whereas underestimation of the impacted area, whilst of short term benefit to the industry may have longer term repercussions for sustainability.

VMS is often only available in an aggregated format due to confidentiality issues, which can overestimate the extent (Hinz et al., 2013). Highly aggregated VMS data at coarse resolutions is highly limiting in the ability of science to draw reasonable conclusions about fishing footprints and impacts (Hinz *et al.*, 2013). This study has provided a strong argument for the creation of comprehensive positional information at higher temporal resolution than is currently available in order to make robust estimates of fishery activities in space and time. The solution which offers the optimal increase in data accuracy, and therefore accuracy of footprint estimation would be to increase the rate at which VMS data are collected (i.e. higher polling rates). Nevertheless higher quality positional data could also be achieved through more rigorous implementation of AIS units, or through partnership agreements between the scallop industry and scientists. The simultaneous use of AIS and VMS data could also increase the quality of outputs in situations where both data sets are available for analysis (Russo et al., 2016).

5. Conclusions

The present study has highlighted the issues surrounding substantial gaps in coverage by AIS data. McCauley et al., (2016) described AIS as currently a 'service that best observes vessels that don't mind being seen'. This likely arises from a lack of desire to be seen by competing vessels and in some cases may be linked to legal infringements. Nonetheless, VMS already provides a means of collecting such data in a manner shielded from public view and hence represents an appropriate pathway for the more accurate calculation of fishing footprints through increased polling rates. At present, given the current low frequency of VMS polling there remains the potential for over-reporting of fishing effort, which could lead to a worse assessment of the state of the marine environment in relation to this metric. As AIS was developed for the purpose of safety and collision avoidance, unless additional legislation is effected to regulate the use of AIS by fishing vessels, designed to specifically increase the coverage of fishing activity, it seems unlikely that AIS data could be considered as an equal alternative to VMS data. If the gaps in coverage were addressed, the increased poll frequency of AIS data would allow more accurate analysis of fishing activity, but increasing the poll frequency of VMS data may be a more viable option. The use of reliable high resolution AIS or VMS data would ultimately benefit fishers and scientists, through generating more accurate fishing footprints and a better understanding of the ecosystem impacts of fishing, and thus more sustainable management.

6. Acknowledgements

We would like to thank the fishermen from the English Channel Scallop fishery who provided their raw VMS data for this analysis. The data used in this study are used under strict conditions of confidentiality and hence explains the lack of reference points used in the images shown to illustrate the results. All relevant fishers were sent a copy of this manuscript prior to submission to the journal. This work was conducted as part of a PhD studentship funded jointly by Bangor University and the Centre for Environment, Fisheries and Aquaculture Science.

7. References

Campbell, M. S., Stehfest, K. M., Votier, S. C., and Hall-Spencer, J. M. 2014. Mapping fisheries for marine spatial planning: Gear-specific vessel monitoring system (VMS), marine conservation and offshore renewable energy. Marine Policy, 45: 293–300.

Dinmore, T. A., Duplisea, D. E., Rackham, B. D., Maxwell, D. L., and Jennings, S. 2003. Impact of a large-scale area closure on patterns of fishing disturbance and the consequences for benthic communities. ICES Journal of Marine Science, 60: 371–380.

EC. 2009. Council Regulation (EC) No. 1224/ 2009 of 20 November 2009 Establishing a Community control system for ensuring compliance with the rules of the Common Fisheries Policy, amending Regulations (EC) No. 847/96, (EC) No 2371/2002, (EC) No. 811/2004, (EC) No.

Eigaard, O. R., Bastardie, F., Breen, M., Dinesen, G. E., Hintzen, N. T., Lafargue, P., ... Rijnsdorp, A. D. (2016). Estimating seabed pressure from demersal trawls, seines, and dredges based on gear design and dimensions. ICES Journal of Marine Science, 73(Suppl. 1), 27-43

Eigaard, O. R., Bastardie, F., Hintzen, N.T., Buhl-Mortensen, L., Buhl-Mortensen, P., Catarino, R., Dinesen, G. R. 2017. The footprint of bottom trawling in European waters: distribution, intensity and seabed integrity. ICES Journal of Marine Science, 74(3): 847 - 865

Foden, J., Rogers, S.I., Jones. A.P. 2011. Human pressures on UK seabed habitats: a cumulative impact assessment. Mar Ecol Prog Ser 428: 33-47.

Gerritsen, H. D., Minto, C., and Lordan, C. 2013. How much of the seabed is impacted by mobile fishing gear? Absolute estimates from Vessel Monitoring System (VMS) point data. – ICES Journal of Marine Science, 70: 523–531.

Hintzen, N. T., Piet, G. J., and Brunel, T. P. A. 2010. Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data. Fisheries Research, 101: 108 -

115.

Hintzen, N. T., Bastardie, F., Beare, D., Piet, G. J., Ulrich, C., Deporte, N., Egekvist, J., *et al.* 2012. VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. Fisheries Research, 115–116: 31–43.

Hinz, H., Murray, L. G., Lambert, G. I., Hiddink, J. G., and Kaiser, M. J. 2013. Confidentiality over fishing effort data threatens science and management progress. Fish and Fisheries, 14: 110–117.

Joo, R., Salcedo, O., Gutierrez, M., Fablet, R., and Bertrand, S. 2015. Defining fishing spatial strategies from VMS data: Insights from the world's largest monospecific fishery. Fisheries Research, 164: 223–230.

Kaiser, M., Clarke, K., Hinz, H., Austen, M., Somerfield, P., and Karakassis, I. 2006. Global analysis of response and recovery of benthic biota to fishing. Marine Ecology Progress Series, 311: 1–14.

Kaiser, M. J., Hilborn, R., Jennings, S., Amaroso, R., Andersen, M., Balliet, K., Barratt, E., *et al.* 2016. Prioritization of knowledge-needs to achieve best practices for bottom trawling in relation to seabed habitats. Fish and Fisheries, 17: 637–663.

Lambert, G. I., Jennings, S., Hiddink, J. G., Hintzen, N. T., Hinz, H., Kaiser, M. J., and Murray, L. G. 2012. Implications of using alternative methods of vessel monitoring system (VMS) data analysis to describe fishing activities and impacts. ICES Journal of Marine Science, 69: 682–693.

Lee, J., South, a. B., and Jennings, S. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. ICES Journal of Marine Science, 67: 1260–1271.

Mccauley, B. D. J., Woods, P., Sullivan, B., Bergman, B., Jablonicky, C., Roan, A., Hirshfi, M., *et al.* 2016. Ending hide and seek at sea. Science, 351: 1148–1150.

Murray, L. G., Hinz, H., Hold, N., and Kaiser, M. J. 2013. The effectiveness of using CPUE data derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop biomass. ICES Journal of Marine Science, 70: 1330–1340.

Murray, L., Hinz, H., and Kaiser, M. 2011. Functional response of fishers in the Isle of Man scallop fishery. Marine Ecology Progress Series, 430: 157–169.

Natale, F., Gibin, M., Alessandrini, A., Vespe, M., and Paulrud, A. 2015. Mapping fishing effort through AIS data. PLoS ONE, 10: 1–16.

Piet, G. J., and Quirijns, F. J. 2009. The importance of scale for fishing impact estimations. Canadian Journal of Fisheries and Aquatic Sciences, 66: 829–835.

Piet, G. J., and Hintzen, N. T. 2012. Indicators of fishing pressure and seafloor integrity. ICES Journal of Marine Science, 69 (10): 1850 - 1858.

Russo, T., D'Andrea, L., Parisi, A., Martinelli, M., Belardinelli, A., Boccoli, F., Cignini, I., *et al.* 2016. Assessing the fishing footpring using data integrated from different tracking devices: Issues and opportunities. Ecological Indicators, 69: 818–827.

Stelzenmuller, V., Rogers, S. I., and Mills, C. M. 2008. Spatio-temporal patterns of fishing pressure on UK marine landscapes, and their implications for spatial planning and management. ICES Journal of Marine Science, 65 (6): 1081: 1091.