Understanding global climate change scenarios through bioclimate stratification
Soteriades, Andreas; Trabucco, A; Murray-Rust, D; Metzger, MJ

Environmental Research Letters

DOI:
10.1088/1748-9326/aa7689

Published: 02/07/2017

Publisher's PDF, also known as Version of record

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Hawliau Cyffredinol / 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.
Understanding global climate change scenarios through bioclimate stratification

To cite this article: A D Soteriades et al 2017 Environ. Res. Lett. 12 084002

View the article online for updates and enhancements.
Understanding global climate change scenarios through bioclimate stratification

A D Soteriades¹, D Murray-Rust², A Trabucco³ and M J Metzger¹,4

¹ School of GeoSciences, The University of Edinburgh, Drummond Street, Edinburgh, EH8 9XP, United Kingdom
² Design Informatics, The University of Edinburgh, Evolution House, Edinburgh, EH1 2LE, United Kingdom
³ Impacts on Agriculture, Forests and Ecosystem Services Division, Euro-Mediterranean Centre on Climate Change, via de Nicola 25, I-07100 Sassari, Italy
⁴ Author to whom any correspondence should be addressed.
E-mail: marc.metzger@ed.ac.uk

Keywords: climate change, RCP, stratification, climate analogue, climate service, climate change scenario

Supplementary material for this article is available online

Abstract

Despite progress in impact modelling, communicating and understanding the implications of climatic change projections is challenging due to inherent complexity and a cascade of uncertainty. In this letter, we present an alternative representation of global climate change projections based on shifts in 125 multivariate strata characterized by relatively homogeneous climate. These strata form climate analogues that help in the interpretation of climate change impacts. A Random Forests classifier was calculated and applied to 63 Coupled Model Intercomparison Project Phase 5 climate scenarios at 5 arcmin resolution. Results demonstrate how shifting bioclimate strata can summarize future environmental changes and form a middle ground, conveniently integrating current knowledge of climate change impact with the interpretation advantages of categorical data but with a level of detail that resembles a continuous surface at global and regional scales. Both the agreement in major change and differences between climate change projections are visually combined, facilitating the interpretation of complex uncertainty. By making the data and the classifier available we provide a climate service that helps facilitate communication and provide new insight into the consequences of climate change.

1. Introduction

The last decades have revealed overwhelming evidence that the world’s climate is changing (Fischer and Knutti 2015, Foster and Rahmstorf 2011, IPCC 2013). Since the 1970s an increasing effort has been placed in modelling the future climate to understand the magnitude of potential change (Coumou and Robinson 2013, IPCC 2013, Manabe and Wetherald 1975, McNeall et al 2016, Yip et al 2011). General Circulation Models (GCMs) have been run for a multitude of scenarios following alternative emissions trajectories, providing a range of plausible climate change with mean global surface temperature likely to exceed 1.5 °C- or even 3 °C in some scenarios- by the end of the century (Collins et al 2013, IPCC 2013). However, it remains difficult to understand and communicate a straightforward implication of these projected changes in generic climate variables (e.g. global mean surface temperature or annual precipitation; Leary et al 2009, Saxon et al 2005, Veloz et al 2012), especially given the great spatial and temporal variation in the projections (Flato et al 2013).

Climate impacts modelling is providing an improved understanding of potential consequences of climate change at global and continental scales (e.g. Schröter et al 2005, Sitch et al 2003, Thuiller et al 2005, Zia et al 2016). For example, dynamic global vegetation models (e.g. Sitch et al 2003) help understand consequences for primary production (Kucharik et al 2000) and global carbon balances (Piao et al 2009), while species distribution models identify vulnerable species (Thuiller et al 2011) and support nature conservation (Franklin 2010). Furthermore,
integrated modelling approaches combine socio-economic and climate scenarios to assess global environmental change impacts for a wide range of domains (e.g. IMAGE Team 2001, Schröter et al 2005, Wei et al 2009). Although these approaches provide insight in specific potential impacts, these more advanced models are also paired with an increased complexity that limit their applications to a limited number of climate projections and yet present a cascade of model uncertainties that complicate understanding of their results (Jones 2000, Visser et al 2000).

In this letter, we present an alternative representation of global climate change projections for 2050 based on shifts in 125 multivariate strata with a relatively homogeneous climate. The approach has similarities to earlier global work on biomes (Leemans et al 1996, Tchebakova et al 1993), and regional studies in US (Saxon et al 2005, Veloz et al 2012) and Europe (Metzger et al 2008), but is based on the recent Global Environmental Stratification (GEnS; Metzger et al 2013b). The approach provides a convenient statistical summary of a climate change scenario by quantifying shifts in strata identifying climatic analogues (see Hallegatte et al 2007, Veloz et al 2012) that help illustrate and understand the magnitude of the projected change (e.g. desert climates from Africa moving into Spain). Such an approach can support a variety of impacts studies including crop modelling (e.g. Ewert et al 2005, Hermans et al 2010), biodiversity modelling (e.g. Verboom et al 2007), and land use change modelling (Zomer et al 2014a, 2014b). All data and algorithms are available through the University of Edinburgh data repository DataShare (Metzger et al 2017).

2. Data sources

2.1. The global environmental stratification

The construction of the Global Environmental Stratification, described in detail by Metzger et al (2013b) is based on statistical clustering to ensure that subjective choices are explicit, their implications are understood and the strata can be seen in the global context (Metzger et al 2005). Statistical screening of 42 gridded bioclimatic variables derived from the WorldClim database (Hijmans et al 2005) for baseline (1960–2000) produced a subset of four variables: the annual sum of daily temperatures above 0°C; an annual aridity index (AI) (Trabucco et al 2008); temperature seasonality; and potential evapotranspiration seasonality. These were compacted in uncorrelated dimensions using principal components analysis. Statistical clustering was subsequently used to classify the principal bioclimate gradients in temperature, aridity and seasonality into relatively similar biophysical environments. To provide structure and support a consistent nomenclature these strata were aggregated into global environmental zones based on the attribute distances between strata. The GEnS delineates 125 strata for current conditions (1961–2000), which have been aggregated into 18 global environmental zones and have a 30 arcsec resolution (approximately 1 x 1 km at the equator). The zones are displayed on the global map in figure 1 (a). The zones and their containing strata are listed in table 1 with bioclimate summary statistics, which is used as a reference in the sections to follow. Added value of the GEnS compared to existing global classifications include the rigorous statistical methods used to delineate the strata, the high spatial resolution which allows for the identification of regional gradients, and the increased number of strata (e.g. in major mountain systems where altitudinal gradients are subdivided). Comparison with existing global, continental, and national stratifications confirms that it successfully partitions important environmental gradients (Metzger et al 2013b, van Wart et al 2013).

2.2. WorldClim Climate change scenarios

To ensure consistency with the GEnS climate predictors, monthly climate averages from the WorldClim future climate scenarios (Hijmans et al 2005, Ramirez and Jarvis 2010) were used for the analysis to reconstruct the GEnS bioclimate variables representative for for 2050 (i.e. average over 2041–2060). These datasets were developed by downscaling a suite of 19 GCMs projections, provided by Phase 5 of the Coupled Model Intercomparison Project (CIMP5; Mehl and Bony 2011), and four Representative Concentration Pathways (RCPs; Vuuren et al 2011) emissions scenarios to 30 arcsec resolution, as described by Ramirez and Jarvis (2010). The four main RCPs were taken into consideration (i.e. RCP 2.6, RCP 4.5, RCP 6 and RCP 8.5), representing increased greenhouse gas concentrations, and thus radiative forcing. The resulting ensemble of GCMs/ RCPs projections accounted for 63 elements, which represent convergence and agreement (i.e. uncertainty) among different future climate projections. Such downscaling was carried out using the Delta method, a statistical downscaling technique, which is based on the interpolation of future simulated GCM monthly climate anomalies with high resolution ‘actual’ monthly climate surface distributions from WorldClim (Ramirez and Jarvis 2010). It is assumed that the change in climate is relatively stable over space within a GCM pixel, while changes in future climate conditions (averages and seasonality) are driven at GCM pixel level by the GCMs projections. Statistical downscaling techniques have emerged to satisfy the need for interpolating regional-scale atmospheric predictor variables (e.g. area averages of precipitation and/or temperature; Wilby et al 1998) to resolutions.
high enough to represent trends needed for bioclimate studies, at least for lower order statistical moments, i.e. monthly averages (Sunyer et al 2012, Sunyer et al 2010). Furthermore, the simplicity of the assumptions behind the Delta method favour fast computing efficiency, allowing for downscaling of multi-GCMs ensemble at very high resolution to fully represent climate projections modelling uncertainty.
2.3. Data preparation
To facilitate computation, all datasets were re-sampled to 5 arcmin resolution. Although the experiment could have been carried out at 30 arcsec resolution, the lower computational cost made it possible to use multiple GCMs, allowing the maximum representation of uncertainties in climate change assessments (Hein et al 2009, Viner 2002). The resampling was carried out in ArcGIS version 9.3 (ESRI 2011), using the mean values for the bioclimate variables and BlockMajority for the nominal GEnS data. The data were then combined into a single data file with 2221696 rows, one for each 5 arcmin land surface grid cell, with the following attributes: the GEnS stratum, latitude, longitude, the four climate variables for the baseline, and for each of the 63 future scenarios (i.e. 19 GCMs and 4 RCPs, although not all 4 RCPs were available for each GCM).

3. Mapping global climate change through bioclimatic classification

3.1. Classification
To map the future location of the GEnS strata a classification problem needs to be solved by finding a model, or classifier, that can adequately describe the GEnS strata based on the climate variables in baseline (i.e. the training dataset), to subsequently model their future distribution using the climate scenarios (as described above). The WEKA Data Mining Software in Java (Hall et al 2009, Witten et al 2011) was used to explore alternative classification algorithms (see appendix). Random Forests (Breiman 2001) showed superior performance in classifying the GEnS and was thus selected. The method uses a combination of multiple classification trees (i.e. rules for predicting a class given the values of its predictor variables) to provide more accurate classifications (Breiman 2001, Cutler et al 2007). Random Forests has a high classification accuracy (Cutler et al 2007) and has been used in other climate change-related modelling exercises (e.g. Cutler et al 2007, Prasad et al 2006, Rehfeldt et al 2012). Stratified cross-validation (Zhang et al 1999) was used to train the baseline dataset with Random Forests, a resampling technique using multiple random training and test subsamples. Exploratory comparisons between random subsets of the whole training dataset were used to determine the optimal settings for the Random Forests parameters, as described in detail in the appendix, resulting in a classifier that correctly classified 83.61% of the strata.

3.2. How do the global bioclimatic zones shift?
The maps resulting from the classification show notable spatial shifts in the environmental zones between baseline and future (2050) conditions, as illustrated in figures 1(a) and (b). General observations are the significant spatial expansion of the Extremely Hot and Xeric zones (cf table 1) and the colonisation of cold mountaintops by warmer zones. These spatial shifts are examined below in greater detail.

The zonal flow diagram in figure 2 shows that warmer environments generally replace cooler ones. Hot zones move towards becoming Extremely Hot & Xeric, consequently leading to a near doubling of the latter (+96% on average over the 63 scenarios). Likewise, a great decrease in the tundra-type Extremely Cold & Mesic zone (–37%) is observed, largely shifting to warmer, drier, and more temperate zones (Cool Temperate & Dry; Cool Temperate & Moist). Areas with Warm Temperate & Mesic, and with Cool Temperate & Xeric, zones are expected to increase, replacing colder zones. In figure 3 we observe that zones generally shift poleward. The Arctic is invaded by wetter zones ([Extremely] Cold & Mesic; [Extremely] Cold & Wet). Areas between the Tropic of Cancer and the Tropic of Capricorn are becoming warmer. Warm Temperate environments are becoming less mesic and more xeric. Notable areas that do not experience shift are the Extremely Hot & Xeric zone, parts of the tropic Extremely Hot & Moist zone, the coldest areas of the polar regions, and some wet oceanic regions (see supplementary data stacks.iop.org/ERL/12/084002/mmmedia).

3.3. What do shifts look like at a regional scale?
It is not possible to provide an extensive range of regional examples in this letter, but we provide several noteworthy examples that illustrate how the shifts of individual environmental zones and GEnS strata provide a convenient and informative summary of global and regional environmental change. The examples were selected to demonstrate climate change impacts on, among others, major agricultural regions, deserts, mountain ranges, permafrost environments and tundra biomes (figure 4).

At a regional scale, the shifts in strata illustrate that major agricultural regions becoming hotter and drier. For example, the Cerrado in Brazil shifts from Extremely Hot & Moist to Extremely Hot & Xeric (figure 4(a)), with strata 122 and 123 expanding greatly in the region (supplementary data); and in California, USA, shifts from Warm Temperate & Mesic to Warm Temperate & Xeric (figure 4(b)) and Hot & Dry (especially strata 78, 82, 84, 86, 89, 91, 92 and 96; supplementary data).

The expansion of desert-like climates is clearly visible in the maps. For example, the Sahara shows a northward expansion of the Extremely Hot & Xeric zone in (strata 119, 121, 124 and 125; supplementary data), replacing zone Extremely Hot & Arid, which also shifts northward (figures 4(c) and (d)). Meanwhile, the Mediterranean countries Spain, Portugal, Italy, Greece
and Cyprus are occupied by zone Hot & Dry, currently prevalent in the north of Africa (figure 4(d)).

Conversely, the maps illustrate the extent to which colder climes contract. For example, in the Himalayas zone Extremely Cold & Mesic is largely replaced by the warmer zone Cold & Mesic (figure 4(e); strata 37 and 42; supplementary data). Similarly, permafrost environments and tundra biomes show a great decline. Zone Extremely Cold & Mesic in the Northern Hemisphere is replaced by the warmer zone Cold & Mesic (figure 4(f); strata 30, 31, 34, 35, 36, 40, 41, 44 and 48; supplementary data).

4. Discussion

The examples above demonstrate how shifting bioclimate strata summarize future environmental shifts at global and regional scales. Both the agreement in major change and differences between climate change projections are visually combined (figures 1–4), facilitating the interpretation of complex uncertainty. The data reported here present a major advance over earlier attempts at mapping bioclimate shifts (e.g. Alessandri et al 2014, Beck et al 2005, Belda et al 2016, Jylhä et al 2010, Leemans et al 1996, Mahlstein et al 2013, Metzger et al 2008, Rohli et al 2015, Saxon et al 2005, Wang and Overland 2004) due to its greater thematic resolution (125 strata) and spatial resolution (5 arcmin). Its global coverage also provides consistency for comparative analyses, and provides the potential for the dataset to be used as a consistent tool for global assessments. Zomer et al have already demonstrated how the GEnS can be used to understand potential vegetation change in the Himalayas (2014b), and climate change impacts on biodiversity and rubber production in Yunnan, China (2014a).

The most notable methodological limitation is the fact that the strata are fixed on the current climate. It does not allow for the emergence of hotter and drier strata in the future, i.e. the hottest parts of the Sahara will always remain in the Extremely Hot & Xeric stratum 125. Furthermore, novel climates with annual profiles that do not currently exist will not be identified as such, but assigned to the most similar current climate. These issues are inherent in

Figure 2. Zonal flow diagram, showing both the areal transitions between zones and the overall and relative changes in area averaged over all 63 GCMs for the 2050 time-slice. Transition amount is calculated by comparing the starting and ending class of each grid cell and by aggregating. It is represented by the intensity of arrows between each zone (transitions with less than 50000 km² are removed). Areal change is calculated by comparing the total area of each class to the baseline value. Each class is represented by a circle which is coloured according to whether area increases (green) or decreases (red). Spatial structure is provided by aridity and temperature, with each class centred on its baseline average. This figure was produced in R (R Core Team 2015).
any approach using climate analogues (e.g. Halle-gatte et al 2007, Veloz et al 2012) to understand change.

The efficiency of the Random Forests classifier (appendix) was important to be able to run analyses on desktop computers. Although 16% of grid cells were misclassified due to heterogeneity within the strata, these cells would have been assigned to strata that are climatically similar (Metzger et al 2008), and consequently any interpretation errors will be small. The classifier is freely available (Metzger et al 2017), and can be used to map the GEnS strata for future WorldClim climate change projections, as well as for the 30 arcsec (approximately 1 km²) data. The latter proved too detailed for global calculations, but provide valuable detail in regional applications (cf Zomer et al 2014a, 2014b).

5. Conclusion

The bioclimate shifts presented here form a middle ground, conveniently summarizing current knowledge of climate change with the interpretation advantages of categorical data but a level of detail that resembles a continuous surface. By making the shifting GEnS datasets and the classifier available (Metzger et al 2017) we provide a climate service that helps facilitate communication and provide new insight into the consequences of climate change.

Appendix. Methods for calculating strata shifts

1. Training the classifiers

1.1. Classifier selection

Alternative models, or classifiers, for mapping the location of the GEnS strata were explored in the WEKA Data Mining Software in Java (Hall et al 2009, Witten et al 2011), version 3-6-4. Two algorithms were deemed especially promising: a decision tree (Random Forests; Breiman 2001) and a Bayesian classifier (Naïve Bayes; John and Langley 1995). Random Forests comprises a combination of classification trees to provide more accurate classifications (Breiman 2001, Cutler et al 2007) and is considered an effective tool in prediction (Breiman 2001) due to its high classification accuracy (Cutler et al 2007). Naïve Bayes is based on Bayes’ rule of conditional probability and the algorithm operates under the simplistic assumptions that attributes in a given dataset are independent, and follow the normal (Gaussian) probability distribution (Witten et al 2011). Despite these simplistic assumptions, it is generally agreed that Naïve Bayes is highly effective when tested on actual datasets (e.g. Friedman et al 1997, John and Langley 1995, Turhan and Bener 2009, Vaidya 2008).

Stratified multiple-fold cross-validation was used to train the baseline dataset with the two algorithms,
by resampling multiple random training and test subsamples (Zhang et al. 1999). Specifically, it creates a sub-set of the dataset (training set) for each fold that is then trained on the remaining data (test set) to test the algorithm’s ability to predict the actual dataset. The method has the advantage over other training methods (e.g. data splitting into training and test sets) that all observations of the datasets can be used for training and testing the model, thus reducing any uneven representation in training and test sets (Witten et al. 2011, Zhang et al. 1999). Ten folds were used as this is the recommended number of folds to get the best error
estimate from the cross-validation technique (e.g. Borra and Di Ciaccio 2010, Kim 2009, Li et al 2010, Witten et al 2011).

Given the substantial size of the baseline dataset (2221696 rows by 4 bioclimatic variables + 125 GEnS strata), the optimal parameter settings for the Random Forests algorithm (i.e. the number of trees selected, their maximum depth and the number of attributes used in the random selection) were determined using a subset of the data. A stratified 3% random sample without replacement was used as the training set to obtain an estimate of the classifier’s performance for different combinations of parameter values, a common method in data mining to save computing time (Borra and Di Ciaccio 2010). This exercise revealed that: (1) infinite maximum tree depth significantly increased calculation time and model size without necessarily improving the classifier’s accuracy; (2) selecting more than 16 trees gave negligible changes in the classifier’s accuracy despite increasing computing time; (3) classification accuracy stabilized around 82% when maximum depth ranged from 10–15 and number of trees ranged from 10–25. The final parameter settings were chosen following further training experiments with stratified 5%, 7%, 14% and 25% random samples without replacement of the baseline dataset (i.e. selecting three random features and ten trees with a maximum depth of 11 for each tree), which were then used to run the entire baseline dataset resulting in a 29 Mb model that could correctly predict 83.61% of the grid cells.

Conversely, Naïve Bayes is a very simple classifier with fixed parameters, making it possible to use the whole baseline dataset for training without initial testing. Naïve Bayes’ simplicity allowed for a fast training time (approximately 40 min) on the whole baseline dataset, correctly classifying 78.31% of the strata using a 44 Kb model. Although the classification accuracy was only slightly lower than for Random Forests, there were significant misclassifications of certain strata. Further inspection revealed strata with relatively large variation to points in the extremes of the distribution. Others have previously noted that Naïve Bayes is best suited for smaller datasets, an opinion supported by various studies (e.g. Catal and Diri 2009, Chandra and Gupta 2011).

Based on this analysis the Random Forests classifier was chosen for further analysis of the shifting strata.

Acknowledgments

The authors would like to thank the two anonymous reviewers for their constructive feedback and suggestions. ADS would like to thank Dr Eibe Frank from The University of Waikato for his invaluable support with the WEKA software. ADS gratefully acknowledges support from the A.G. Leventis Foundation (personal grant) and The University of Edinburgh MSc Integrated Resource Management (publication bursary). MJM would like to acknowledge support from the European Commission Seventh Framework Programme under Grant Agreement No. FP7-ENV-2012-308393-2 (OPERAs).

ORCID

A Trabucco https://orcid.org/0000-0002-0743-3680
M J Metzger https://orcid.org/0000-0002-5119-5894

References

Breiman L 2001 Random Forests Mach. Learn. 45 5–32
Catal C and Diri B 2009 Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem Inform. Sci. 179 1040–58
Coumou D and Robinson A 2013 Historic and future increase in the global land area affected by monthly heat extremes Environ. Res. Lett. 8 034018
ESRI 2011 ArcGIS 9.3 (California: Environmental Research Systems Incorporated)
Franklin I 2010 Moving beyond static species distribution models in support of conservation biogeography Divers. Distrib. 16 321–30
Friedman N, Geiger D and Goldszmidt M 1997 Bayesian network classifiers Mach. Learn. 29 131–63


Hijmans R J, Cameron S E, Parra J L, Jones P G and Jarvis A 2005 Very high resolution interpolated climate surfaces for global land areas Int. J. Climatol. 25 1963–78

IMAGE Team 2001 The IMAGE 2.2 Implementation of the SRES Scenarios: A Comprehensive Analysis of Emissions, Climate Change and Impacts in the 21st Century (Bilthoven: National Institute of Public Health and the Environment [RIVM])


Schröter D et al 2005 Ecosystem service supply and vulnerability to global change in Europe Ecol. Model. 181 133–7


Sunyer M A, Madsen H and Ang P H 2012 A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change Atmos. Res. 103 119–28

Sunyer M A, Madsen H and Yamagata K 2010 On the use of statistical downscaling for assessing climate change impacts on hydrology Int. Workshop Advances in Statistical Hydrology (Taormina, Italy, 23–25 May 2010)


Turban B and Bené A 2009 Analysis of naive bays’ assumptions on software fault data: an empirical study Data Knowl. Eng. 68 278–90
Viner D 2002 A qualitative assessment of the sources of uncertainty in climate change impacts assessment studies: a short discussion paper Advances in Glob. Change Res. 10 139–51