Deliverable 4.3 (D4.3)

Quantification on sources of uncertainty in previous analyses

M54

Project acronym:	EU BON
Project name:	EU BON: Building the European Biodiversity Observation Network
Call:	ENV.2012.6.2-2
Grant agreement:	308454
Project duration:	01/12/2012 – 31/05/2017 (54 months)
Co-ordinator:	MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany
Delivery date from Annex I:	M54 (May 2017)
Actual delivery date:	M54 (May 2017)
Lead beneficiary:	UFZ
Authors:	Andreas Menzel, Aliénor Jeliazkov, Dirk Schmeller, Gudrun Carl, Klaus Henle, Jean-Baptiste Mihoub, Guy Pe'er, Ingolf Kühn (UFZ, Helmholtz Centre for Environmental Research, Germany) Duccio Rocchini, Carol X. Garzon-Lopez, Matteo Marcantonio (FEM, Fundazione Edmund Mach, Italy) Yoni Gavish, Tim Benton, Charles Marsh, William Kunin (UnivLeeds) Mathias Kuemmerlen, Stefan Stoll, Peter Haase (SGN, Senckenberg Gesellschaft für Naturforschung, Germany) Patricia Mergen (MRAC, Royal Museum for Central Africa, Belgium) Kathleen Kesner-Reyes, Rainer Froese, Cristina Garilao (FIN, FishBase Information and Research Group, Philippines) Christos Arvanitidis, Nicolas Bailly (HCMR, Hellenic Center of Mediterranean Research, Greece) Quentin Groom (BGM, Botanic Garden Meise, Belgium) Corinne Martin, Lauren Weatherdon (UNEP-WCMC, UNEP World Conservation Monitoring Centre, UK) Nicolas Titeux, Lluís Brotons, (EBCC-CTFC, European Bird Census Council - Centre Tecnològic Forestal de Catalunya, Spain) Neil Brummitt (NHM, The Natural History Museum, UK) Yu-Pin Lin, (NTU, National Taiwan University, Taiwan) Anke Hoffmann, Johannes Penner, Kathrin Vohland, Florian Wetzel (MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany) + V. Amici, G. Bacaro, L. Bastin, A. Chiarucci, G. Coro, D. Doktor, G. M. Foody, H. C. Hauffe, K. S. He, K. Kaschner, J. O'Connell, C. Ricotta, A. Rizzoli, R. Rosá, O. Schweiger (multiple institutions, see content of the Deliverable)

This project is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union **Dissemination Level**		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308454.

Executive Summary

Introduction

The recent rapid, ubiquitous and global environmental changes require close exchange between knowledge holders, decisionmakers and policymakers to inform and support key decisions on the management of biodiversity and natural resources. Science, its approaches, results and recommendations are frequently associated with uncertainty, while stakeholders and practitioners often require clear and certain information; a situation that limits the communication between scientists and the aforementioned groups and therefore also restricts efforts regarding conservation and management of biodiversity.

As part of WP4 '**Link environment to biodiversity: analyses of patterns, processes and trends**', task 4.5 aimed at identifying and summarising existing sources of uncertainties alongside the biodiversity modelling process and finally at quantifying those uncertainties in terms of analyses and criteria of decision-making. The latter turned out to be a challenging task. The different existing approaches and frameworks in biodiversity modelling as well as the involvement of different scientific communities itself are too heterogeneous to gain a general directive to quantify uncertainty at this point. Therefore, the focus of task 4.5 was re-oriented; the partners worked on reviewing these heterogeneous sources of uncertainty and on assessing how these are considered and addressed in current research on biodiversity. The following three focal points were set: (1) **The development of a conceptual framework integrating the existing sources of uncertainty** that are linked to the modelling process to set a baseline for prioritisation and potential future quantification of uncertainty, which is based on the current state of recognition and incorporation of these sources. This also includes the identification of gaps in current data and methodologies leading to future improvements. (2) **The development of coherent and straightforward tools and (statistical) methods to explicitly account for uncertainty in biodiversity models** and to start closing the identified gaps. This task was approached in close collaboration with WP3 'Improving tools and methods for data analysis and interface' to utilize overlaps and synergies in both topics and involved partners. (3) As a perspective, we further provide some reflection on the **main difficulties identified in the communication of uncertainties** surrounding scientific results **towards stakeholders and decision-makers** of different levels. As this aim is a main objective of WP6 'Stakeholder engagement and science-policy dialogue', we here focus on the communication, and especially the visualisation of uncertainties directly stemming from biodiversity modelling rather than from interactions within realms at the interface of science and policy.

Progress towards objectives

A large body of existing literature deals with issues of uncertainties alongside the biodiversity modelling process, but there are only few attempts to integrate all sources of uncertainty in one conceptual framework. Additionally, the issue of quantifying those uncertainties is rarely addressed overall. This is insufficient while facing a constant increase in velocity of decision-making, especially concerning those decisions that target future environmental changes and societal developments.

As a first point, we will briefly place the process of biodiversity modelling within the socio-political and socio-economic context, in which the need of clarification as well the formulation of research questions arise. Following, we will summarise the **four essential sources of uncertainty** in (biodiversity) modelling, i.e. 1) **data** – both environmental and biological, (2) **calibration** – i.e. characteristics of the modelling process itself, (3) **validation** – i.e. the process of testing the accuracy of the assigned models and (4) **projections**, and jointly integrate them into the conceptual framework.

As a next step, we will discuss the issue of **propagation of uncertainty** with increasing complexity alongside the modelling process and the current possibilities to actually **quantify** those uncertainties. As a last point, we will discuss **communication strategies (including visualisation of model uncertainty)** to integrate the matter of uncertainties to the interface of scientist and decision-maker involvement as well as the overall process of informing conservation and management of biodiversity.

Achievements and current status

The deliverable follows the structure of the developed conceptual framework and aligns along the abovementioned sources of uncertainty. Thereby, the partners reviewed the aspects and implications of all these sources and developed tools to address these questions as illustrated by **eight case studies** supporting the reduction and/or incorporation of uncertainty in biodiversity and species distribution modelling.

1) Concerning the availability and bias in biodiversity data as well as environmental predictors, there was a strong increase in publications during the last years in both, making data sets available and revealing gaps in data coverage. Nevertheless, it has to be noted that a sheer larger amount of available data does not readily translate to a greater knowledge. A stratified and non-random sampling that is congruent among countries and continents is necessary to systematically tackle questions in biodiversity research that then can inform management decisions and conservation action.

> Our ability to define relevant temporal baselines for biodiversity is still limited and this creates uncertainty especially due to the lack of knowledge about biodiversity states prior to the rise of harmful anthropogenic activities. **Case study I** summarises an assessment of such temporal baselines based on European monitoring schemes.
>
> **Case study II** gives an example on how to assess biodiversity metrics directly via satellite remote sensing to overcome limitations by deriving them from field observations. **This case study comes with a ready-to-use open-source software implementation.**
>
> **Case study III** presents a statistical method for a scale-specific regression to assess the importance of several environmental variables on ecosystem processes at different spatial scales.

2) Species distribution and other biodiversity models have experienced a documented strong rise and advancement, which is also true for accounting for uncertainty within the models. Nevertheless, there is still room for improvement, especially if more biodiversity data will be available in the future that will support the parameterisation of even more complex models.

3) Using an appropriate accuracy measure is essential for assessing the prediction accuracy of biodiversity models. Still, accuracy measures undergo much less development and discussion than the previous sources of uncertainty.

> **Case study IV** introduces spatially corrected versions of current and commonly used accuracy measures that will contribute to evaluate prediction errors in presence/absence models, especially in case of medium or high degree of similarity of adjacent data, i.e. aggregated (clumped) or continuous species distributions. **This case study comes with a ready-to-use open-source software implementation.**

4) Reporting uncertainty in projections provides confidence in model results that supports decision-making in conservation-related recommendations and policies. Uncertainties, especially in the

future distributions of species render any decisions about where or how to implement conservation actions difficult and may increase expenses whenever large areas must be managed.

> **Case study V** takes an overall assessment of uncertainties surrounding different taxonomic freshwater groups, by applying different scenarios of climate and land-use change to data from the Rhine-Main-Observatory (EU BON) test site.
>
> **Case study VI** presents an overall assessment of the uncertainty of predictions using the AquaMaps framework.

Whenever one takes the steps from the iterative procedure of data (collection), to modelling, measuring its accuracy and projection, uncertainty surrounding each of the corresponding techniques and approaches propagates through the whole process. This is generally true for an increase in complexity of modelling frameworks, as each parameter that has to be parameterised adds its own uncertainty to the model outcome. So far, complex models are often also surrounded by a high model uncertainty, as each inherent parameterisation is surrounded by uncertainty on its own.

> **Case study VII** presents a Bayesian modelling framework that explicitly accounts for bias due to different sampling effort and demonstrates (1) how to incorporate this information on uncertainty directly into a modelling framework and (2) how to propagate this uncertainty throughout the model. **This case study comes with a ready-to-use open-source software implementation.**

Quantification of uncertainties alongside the modelling process is an elaborative, but yet inevitable task, if the ultimate goal is to inform the scientific community, stakeholders and the general public. At the moment, quantification often means assessing the importance of each of the beforementioned sources of uncertainty, by having several ‘treatments’ for each group.

> **Case study VIII** is an example at the forefront of quantification of uncertainty in habitat/land-cover classification models or species distribution models in a general way for future assessments.

Science-policy audiences are highly diverse and often receive information that differs in both quantity and quality compared to what science typically provides. Strategies of communication, i.e. which information are inevitable to the audience and how can they be best communicated, change with the addressed audience.

Future developments

Reporting uncertainties alongside each step of biodiversity modelling is essential and should be the golden standard within and outside the scientific community. An uncertainty assessment should be one of the preliminary steps in any related decision-making process, such as actions based on biodiversity modelling results or the delineation of a biodiversity conservation area. Therefore, conservation planners should identify uncertainties in the planning process and, when necessary, evaluate the sensitivity of conservation planning outcomes to the different sources of uncertainty. Additionally, identified uncertainties may require further targeted monitoring to incorporate them into the management development process. This deliverable aims at creating awareness for discussion and integration of model uncertainty among all parties involved in informing and developing key decisions in biodiversity conservation and management.

Among biodiversity modelling procedures, species distribution modelss are standard and essential tools for understanding factors that affect species geographical ranges and for predicting their response to current and future global changes and have already substantially improved during the

last decades. Although species distribution models are strongly established in informing management and conservation decisions, raising the awareness regarding sources of uncertainty and the development of new methods to directly incorporate uncertainty at different levels of the modelling process will finally improve the communication of uncertainties surrounding each scientific result. This will then support a higher level of information concerning these decisions and lead to a higher acceptance among stakeholders and decision-makers, as well as lastly within the general public.

After summarising the current state and limitations of this field, we will be able to further improve tools and methods to account for uncertainty in future biodiversity modelling. In this respect, several activities are at different stages of advancement, from already initiated to well-advanced.

Table of content

1. General Introduction 8
2. Sources of uncertainty 14
2.1 Data 14
2.1.1 Biological response 15
Case study I: *Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures* 17
Case Study II: *Measuring Rao's Q diversity index from remote sensing: An open source solution* 22
2.1.2 Environmental information and predictors 25
Case study III: *Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis* 27
2.2 Model calibration: model types and procedures 29
2.3 Model validation / Measures of accuracy 32
Case study IV: *Spind: a package for computing spatially corrected accuracy measures* 33
2.4. Projections 36
Case Study V: *Uncertainty in distribution predictions of freshwater biodiversity under land-use and climate change scenarios in the RMO test site* 37
Case study VI: *Quantifying uncertainty in AquaMaps* 39
3. Propagation of uncertainty and model complexity 41
Case study VII: *Anticipating species distributions: Handling sampling effort bias under a Bayesian framework* 42
4. Prioritisation of sources of uncertainty 48
Case study VIII: *Assessing the transferability of habitat/land-cover classification models in space and time* 49
5. Outlook: Communication and visualisation of uncertainty 53
6. References 55
7. Annexes 66
7.1. Setting temporal baselines for biodiversity could be an impediment for capturing the full impact of anthropogenic pressures 66
7.2. Measuring Rao’s Q diversity index from remote sensing: An open source solution 82
7.3. Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis 94
7.4. Spind: a package for computing spatially corrected accuracy measures 110
7.5. Uncertainties in distribution predictions of freshwater biodiversity under land-use and climate change scenarios in the RMO test site 125
7.6. Anticipating species distributions: Handling sampling effort bias under a Bayesian framework 138

1. General Introduction

Biodiversity is a multidimensional cluster concept and uncertainties are inherent to many of the dimensions of the research surrounding it. The recent rapid and ubiquitous environmental changes require close exchange between knowledge holders, decisionmakers and policymakers to inform and support key decisions on the management of biodiversity and natural resources. This is of particular importance when considering that ongoing and future changes in the Anthropocene (Steffen et al. 2007, Ellis & Ramankutty 2008, Sarrazin & Lecomte 2016) will possibly result in novel conditions and complex dynamics, so that recommendations and predictions will remain embedded within an unprecedented level of uncertainty in a future with no current analogues. Specifications are needed as to what, precisely, is uncertain, what is the reason for uncertainty and whether this uncertainty matters (Funtowicz & Ravetz 1990, Haila & Henle 2014); specifications and in-depth analyses that recent assessments (e.g. MA - Millenium Ecosystem Assessment 2005, Rio+20 – Cardinale et al. 2012, Ehrlich et al. 2012) in biodiversity science were lacking.

Already in 1921, Knight links measurable (thus quantifiable) uncertainties directly to the notion of risk and the probability for a phenomenon to occur, going as far as almost synonymizing the notions measurable uncertainties and risk. So called unmeasurable uncertainties are doubted to be actually ucertainties at all and associated later with the notion of the "unknown". Ritholtz (2012) warns against the unconsidered synonymous use of the terms risk and uncertainty both in science and in the media, because it can be misleading for decision-makers. Following the German psychologist Gerd Gigerenzer (Director of the Center for Adaptive Behavior and Cognition (ABC) at the Max Planck Institute for Human Development) both terms have to be distinguished (Ramnath 2017):

> *risk*: You are dealing with risk when you know all the alternatives, outcomes and their probabilities.
>
> *uncertainty*: You are dealing with uncertainty when you don't know all the alternatives, outcomes or their probabilities.

Uncertainty has always surrounded and will always surround human decision-making processes and has been one of the main drivers of scientific development (Funtowicz & Ravetz 1990, Pe'er et al. 2014). Three major categories of uncertainties can be classified (Walker et al. 2003, Howell et al. 2013): inexactness (technical uncertainty), unreliability (methodological uncertainty) and "border with ignorance" (epistemic uncertainty). The latter is seen as most problematic, especially in predictive biodiversity modelling, as it lacks data to make it quantifiable. Despite the efforts done and the number of parameters measured, there will always be a part of unexplained uncertainty in each model and result; citing Donald Rumsfeld's famous quote (12th of February 2012):

> *'There are known knowns; there are things we know we know. We also know there are known unknowns, that is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.'*

The latter, inherent and unquantifiable uncertainty has a proven impact on the outcomes of models (Regan et al. 2002, Walker et al. 2003) but should not prevent good science to happen (Pe'er et al. 2014). Both scientists and decision makers have to accept that uncertainty is systematically underestimated and that obtained results only represent a subset of total uncertainty that may be quantifiable. Reducing or avoiding uncertainty has been traditionally targeted in science and decision making issues, but uncertainty will never be completely reducible. Therefore, we stress

that one should rather embrace it than trying to avoid it (Haila & Henle 2014). As critical aspects of interpreting and transferring scientific outcomes rely on uncertainty, a careful communication of the uncertainties involved with any analysis or projection into the future as well as the usage of exact, agreed-upon definitions of concepts and terms is needed (linguistic uncertainty; Regan et al. 2002, Kujala et al. 2013). To reduce such linguistic uncertainty within the framework of the present deliverable, we provide a glossary of important terms that are used throughout the document (**see glossary box at the end of this introduction**).

For these reasons, WP4 designed a separate task on uncertainties in biodiversity distribution and trend assessments. Here, we aim at classifying the different sources of uncertainty when assessing current and future trends in distribution and abundance of species. We focus on quantifiable sources of uncertainty but will also reflect on unquantifiable ones that are relevant for management or decision-making. We will consider the propagation of uncertainty in hierarchical models or series of combined models and sketch the potential to quantify uncertainty. We will use results from our own research within WP4 as case studies and complement them with existing literature to provide expert judgement on the (relative) importance of each source of uncertainty as guidance for handling them in research, management, and decision-making. Lastly, we provide some reflections on how to communicate uncertainty to peers and to stakeholders, including verbal, numerical and graphical means.

At this point it has to be explicitly embraced that the process of biodiversity modelling is fundamentally coupled with and embedded in its socio-economic context and furthermore an integral part of the iterative cycle of science (**Figure 1.1**; Schmolke et al. 2010). Biodiversity models hold the strong potential of supporting recommendations and limitations regarding the entities they model, which potentially lead on guiding further scientific and societal actions such as experimental, monitoring and/or conservation efforts. Those efforts then lead to a process of collective learning and the direction of further research, which again lead to the ermergence of new problems and research questions, which are addressed by biodiversity modelling. Addison et al. (2013) summarise practical solutions to make those models visible and valuable for decision-makers. In doing so, they identify common objectives to the use of models in decision-making. More importantly, they call for (1) modellers to involve decision-makers and stakeholders throughtout the modelling process and (2) decision-makers to involve modellers early on in the process of problem formulation to jointly promote mutual understanding of the underlying perspectives and concepts. Note that these challenges are in line with the aim of improving science-policy interface, which is also actively pursued by WP8 “Dissemination and outreach of EU BON activities”.

Task 4.5 is predominantly focused on the development of tools and methods to assess and incorporate uncertainty into frameworks of biodiversity modelling. Therefore, the socio-economic realm of collective decision-making shall not be the major topic of this Deliverable. Nevertheless, it is important to acknowledge that the process of biodiversity modelling is never detached from other scientific components, policy- and decision-making or the general public. Concluding, this is also a call for biodiversity researchers at any level to engage in theory and practice of decision sciences (Polasky et al. 2011, Beale & Lennon 2012).

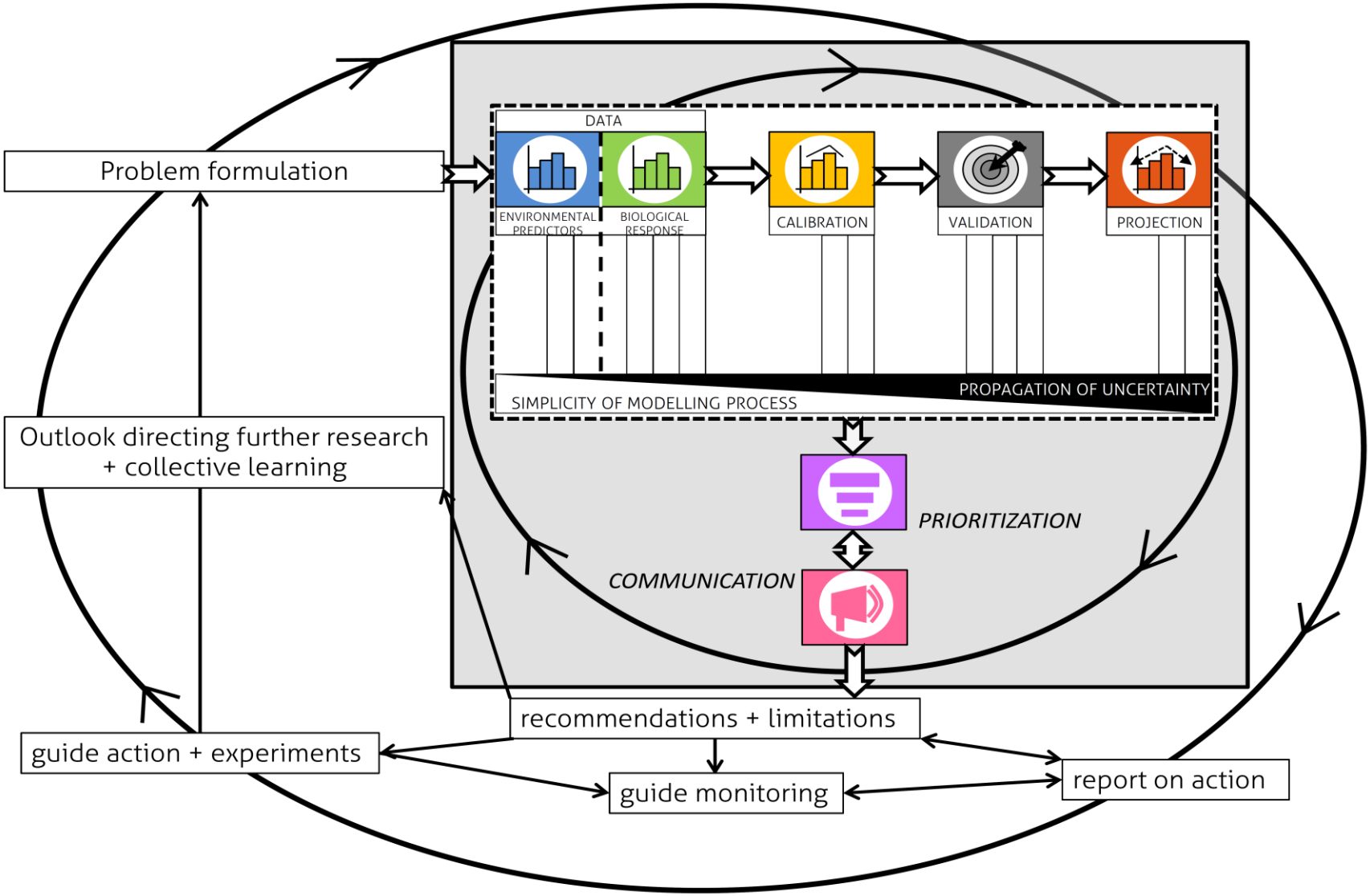


Figure 1.1: *Conceptual framework illustrating the biodiversity modelling process within the iterative cycle of general decision-making processes in the socio-political/economic realm.*

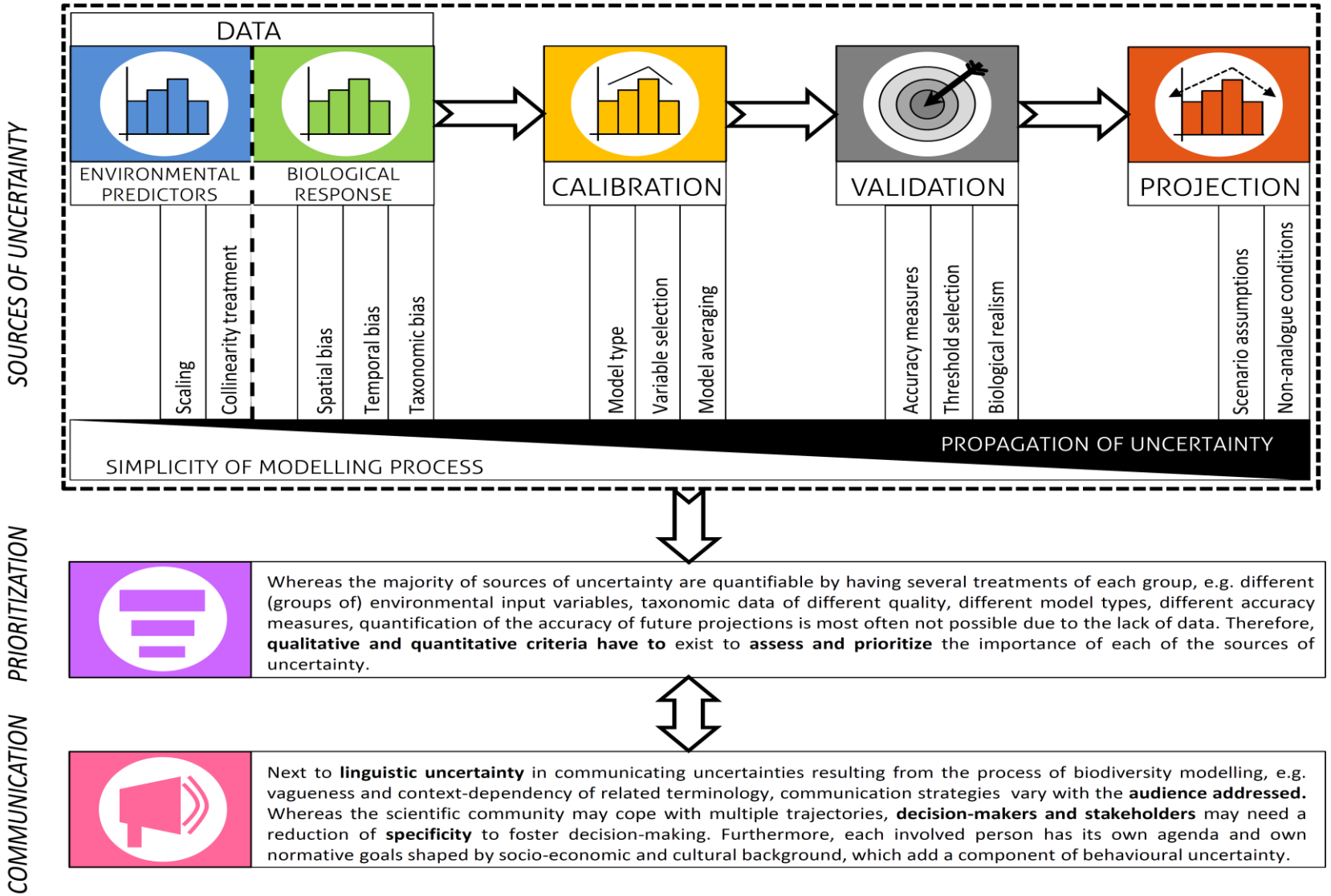


Figure 1.2: *Conceptual framework integrating the sources of uncertainty surrounding the biodiversity modelling process as well as the process of prioritisation and communication of uncertainties originating from such analyses.*

GLOSSARY BOX

Uncertainty: Uncertainty is the absence or inadequacy of knowledge regarding the description of a current state or future outcome and the absence of knowledge regarding probabilities of one or more possible future outcomes. In this context, it refers specifically to the uncertainty of *input* and *output* state and process variables of models.

Uncertainty can be subdivided into *technical uncertainty* (inexactness), *methodological uncertainty* (unreliability), and *epistemic uncertainty* ("border with ignorance"). Whereas the first two can be represented as *quantifiable* errors, the latter often lacks data and is characterised by *qualitative* statements at the most.

Funtowicz S., Ravetz J. (1990). Uncertainty and quality in science for policy. Kluwer, Dordrecht.

Regan H. M., Colyvan M., Burgman M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications 12: 618-628.

Walker W. E. et al. (2003). Defining uncertainty – a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4: 5-17.

Refsgaard J. C., can der Sluijs J. P., Hojberg A. L., Vanrolleghem P. A. (2007). Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling & Software 22: 1543-1566.

Kujala H., Burgman M. A., Moilanen A. (2013). Treatment of uncertainty in conservation under climate change. Conservation Letters 6: 73-85.

Haila Y., Henle, K. (2014). Uncertainty in biodiversity science, policy and management: a conceptual overview. Nature Conservation 8: 27-43.

Risk: Risk results from the absence or inadequacy of knowledge regarding the description of a future outcome. It can be formalized as the probability of a possible future outcome times its impact (severeness).

Prediction: A prediction is a statement about the future that is purely based on a set of quantities derived from a statistical model and its inherent assumptions, without using further assumptions, e.g. concerning future environmental and socio-economic developments. Predictions are based on data and evidence and are therefore less broad than projections or scenarios, often address a smaller scale or shorter time period, but involve a lesser degree of uncertainty at the same time. Frequently, though, the term "prediction" is used when "projection" is meant.

Dormann C. F., et al. (2008a). Prediction uncertainty of environmental change effects on temperate European biodiversity. Ecology Letters 11: 235-244.

Buisson L., Thuiller W., Casajus N., Lek S., Grenouillet G. (2010). Uncertainty in ensemble forecasting of species distribution. Global Change Biology 16: 1145-1157.

Evans M. R. et al. (2013a). Predictive Systems Ecology. Proceedings of the Royal Society B.

Guisan A. et al. (2013). Predicting species distributions for conservation decisions. Ecology Letters 16: 1424-1435

Projection: A projection is a statement about the potential future of a quantity or set of quantities, often derived from a statistical model. In contrast to predictions, projections involve assumptions beyond modelling, e.g. concerning future environmental and socio-economic developments, that may or may not be realised. Projections may be the raw material for scenarios.

IPCC glossary: https://www.ipcc.ch/pdf/glossary/tar-ipcc-terms-en.pdf

Keyfitz N. (1972) On Future Population. Journal of the American Statistical Association 67: 347-363

Singer A. et al. (2016). Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? Ecological Modelling 326: 63-74.

Scenario: A scenario is a synopsis of a plausible sequence of possible future actions, events or developments. It is a qualitative, often simplified set of assumptions about key driving forces and their relationships to support thinking about the future. A set of scenarios depicts different interpretations of the current situation (baseline) and illustrates as well as compares alternative pathways of uncertain environmental and socio-economic developments and their diverging consequences related to biodiversity.

Zurek M. B., Henrichs T. (2007). Linking scenarios across geographical scales in international environmental assessments. Technological Forecasting and Social Change 74: 1282-1295.

Spangenberg J. H. et al. (2012). Scenarios for investigating risk to biodiversity. Global Ecology and Biogeography 21: 5-18.

Validation: Validation is a process of assessing whether a value of a data item derived from a statistical technique is accurately predicted, i.e. it comes from a given set of defined and acceptable values. It is a test of correctness, completeness, intentional implementation and structural plausibility.

Rykiel, E. J. (1996). Testing ecological models: the meaning of validation. Ecological Modelling 90: 229-244.

2. Sources of uncertainty

Spatial, temporal and taxonomic information related to species occurrences is at the core of modelling biodiversity data and is at present predominantly conducted via species distribution models (SDMs; also called niche or habitat suitability models) that assess the relationship between species ranges and potential changes of these ranges as a response to different factors (Elith et al. 2006, Dormann et al. 2008a). Therefore, SDMs act as the major source informing decisions on management and conservation of biodiversity. SDMs are typically based on several assumptions that simplify the biological reality and/or help to satisfy statistical presumptions. Those assumptions occur at any step of the modelling process, starting with raw data and ending with potential projections to distant, non-analogous spaces or futures and therefore, each of these steps contributes to the emergence of uncertainties while modelling biodiversity (Dormann et al. 2008b). Thereby, uncertainty propagates throughout the modelling process (Guillera-Arroita et al. 2015) and substantially influences the decision-making process that may be derived from it.

Awareness of different sources of uncertainty in the described modelling process is essential for an unbiased and strong transfer of scientific results to inform decision-makers and stakeholders. In the following, we will characterise all potential sources of uncertainty (**Figure 1.2**): 1) *data* – both environmental and biological, (2) *calibration* – i.e. characteristics of the modelling process itself, (3) *validation* – i.e. the process of testing the accuracy of the assigned models and (4) *projections*, and integrate them in a joint conceptual framework. We summarise the current state of awareness for each of these sources' uncertainties within the scientific community and the current developments of methods to directly account for the diverse set of uncertainties while modelling biodiversity.

2.1 Data

The societal and scientific interest in understanding patterns and processes of biodiversity loss and changes in species ranges at larger scales (from regional to country to European and up to global scale) fundamentally increased during the past decades, as more and more processes such as climate and land-use change, habitat fragmentation, urbanisation and the introduction of alien species were found to (inter)act at these scales (Sala et al. 2000, Ellis & Ramankutty 2008, IPCC 2013). At the same time, new technologies and global initiatives have produced and consolidated a huge amount of data and the analyses of these processes were highly promoted by the availability of large European and global data sets (species distribution data/biodiversity inventories, e.g., GBIF - Edwards 2000, Map of Life – Jetz et al. 2012, Atlas of European Breeding Birds - Hagemeijer and Blair, 1997, Atlas Florae Europaeae - Lahti and Lampinen, 1999; trait data, e.g., TRY - Kattge et al., 2011; climate data, e.g., CliMond - Kriticos et al., 2012). Currently, we are living in an age, in which the term 'big data' is established in both scientific and non-scientific realms and scientists most certainly face the biggest information availability of all times (Kelling et al. 2009, Hampton et al. 2013). This was made possible due to a strong progress in digitalisation of biodiversity data and the development of data standards (Guralnick et al. 2007) as well as an increased effort in utilising citizen science initiatives (Bell et al. 2008, Dickinson et al. 2012).

Despite these advances, global biodiversity data are generally incomplete and still biased in many ways (for comprehensive reviews see Hortal et al. 2015, Meyer 2016 and Meyer et al. 2016a) and the scientific community faces the urgent question on how to distribute resources to balance those biases, to decide where to invest in monitoring infrastructure and which biodiversity units to monitor. There is still insufficient consistency among national or regional monitoring and sharing of biodiversity data (Pereira et al. 2013). Those gaps have to be filled

systematically and biases have to be explicitly addressed to ensure indisputable management strategies and policies in biodiversity conservation. Nevertheless, it is important to note that although it is a crucial intention to increase data availability for the purpose of analysing biodiversity responses to global change, the mere increase of data availability is no guarantee for a likewise increase in knowledge, as available data, no matter how detailed, will always have an inevitable component of uncertainty (no matter of which of the below-mentioned sources). Therefore, next to taking the effort of a systematic observation and collection of biodiversity entities that needs to encompass local, national, regional and global scales (Pereira et al. 2013), there is also a strong need to explicitly account for data biases and limitations in existing and future modelling frameworks (Meyer 2016; Meyer et al. 2016a, 2016b).

This section will state the potential biases that have to be taken into account, whenever data are used for biodiversity modelling. We distinguish between (1) biological variables, which are the response of biodiversity models and usually a species or taxa (but also traits or community compositions) of certain interest and (2) environmental variables, which are used as predictors in biodiversity models – as both parts of available data face different difficulties in processing. Overall, data is the source of uncertainty, for which the largest body of literature investigating the effect of its surrounding uncertainty has been assessed. We will summarise important components of this source of uncertainty in the following, especially as relatively new approaches such as the comprehensive use of remote sensing and the conceptualisation of Essential Biodiversity Variables (EBVs) are recent advances in this realm.

2.1.1 Biological response

A large body of literature assessing the uncertainties surrounding biological response data exists and comprehensive reviews and assessments have been published recently (Hortal et al. 2015; Meyer 2016, Meyer et al. 2015, 2016a). Those authors thoroughly discuss the several shortfalls in the current state of biodiversity knowledge and make this source of uncertainty in biological modelling one of the best documented. There are various reasons for the limitation of data accessibility, which often are uncorrelated and dominant in different regions in Europe and of the world (Meyer et al. 2015). New data can and should be generated by establishing straightforward and unifying biodiversity monitoring schemes to close critical data gaps as well as by the integration of already existing, still non-digitalised data in museums and other collections, by making the best use of data derived from ecological field studies and by utilising satellite remote sensing techniques (Proença et al. 2016).

Most niche-based models are based on species occurrence data, by either using presence/absence or presence-only data. Raw occurrence data arising from recorded presences of species are usually gathered in monitoring schemes, museums, herbaria, field surveys or volunteer observation networks (Graham et al. 2004; Dickinson et al. 2010, 2012). While these data are increasingly used to generate species distribution maps, they come along with various kinds of uncertainties (Rocchini et al. 2011). Uncertainty may arise from differences in sampling protocols and design, the lack of systematic or exhaustive surveys and failure to report absences (Yoccoz et al. 2001). Guillera-Arroita et al. (2015) summarise three probabilities that influence the chance of recording a species at a particular site: (1) the probability that a species occupies the site, (2) the probability that the site is sampled and (3) the probability that the species is detected given that it is present. If the detection of species is imperfect, which is predominantly the case even for immobile species (Garrard et al. 2008, Chen et al. 2013), SDMs are confounded and rather estimate the likelihood of a species being observed rather than being present. Species' detection probabilities may vary across habitats

and create false absences, if absences are reported at all (Tyre et al. 2003, MacKenzie et al. 2006, Chen et al. 2013), and/or false positives due to misidentification (Miller et al. 2011, Chambert et al. 2015). In addition, specific occurrences or observations are more certain than others (e.g. recording a bird's nest > a bird's song > a bird's sight).

The inherent complexity and dynamics of species distributions are reflected in the quality of the available data (Jiguet et al. 2005; Dormann et al. 2008b; Elith & Graham 2009; Rocchini et al. 2011); considering species abundance data that are essential for understanding population dynamics, they are generally rather scarce. These shortfalls in data availability finally affect estmates of minimum viable populations (Reed et al. 2003), once again highlighting the crucial importance of structures promoting long-term monitoring of population trends (Hortal et al. 2015).

Attempts exist to combine the efforts underlying the existing data sets into an informative distribution map (Jetz et al. 2012), but available maps are still predominantly biased presence-only maps and the harmonisation of and collection of unbiased data remains an important challenge for the future.

Spatial uncertainty

Despite those efforts, biodiversity data are still rarely collected in a stratified and regular manner, but predominantly arise from random sampling that is opportunistic and spatially biased (Albert et al. 2010, Guillera-Arroita et al. 2015). If the spatial varation in sampling effort corresponds to the spatial varation in environmental data (e.g. due to differences in detectability per habitat), this might generally lead to misleading estimates of species' actual and potential ranges (Loiselle et al. 2008, Beale & Lennon 2012).

Different localities obtain different data qualities, which results in different, spatially structured (and therefore biased) quality of biological response data (Boakes et al. 2010, Rocchini et al. 2011, Yang et al. 2013, Hortal et al. 2015, Meyer et al. 2015). This spatial bias or *spatial uncertainty* may moreover arise from different reasons: some areas are better mapped than others and are spatially biased in this regard (Manceur & Kühn 2014, Kuemmerlen et al. 2016), which increases uncertainty and decreases predictive power of biodiversity models. This might be due to differences in sampling effort and mapping schemes in these localities (Albert et al. 2010), which might in turn be a result of differences in the overall socio-economic status (Amano & Sutherland 2013; Meyer 2016; Meyer et al. 2016a), differences in scientific infrastructures (e.g. the density of institutes or the proximity to the next research centre; Moerman & Estabrook 2006) and/or the simple accessibility of certain areas (e.g. via roads; Barve et al. 2011). This leads to the fact that countries with weaker positions regarding these measures are often those with highest biodiversity (Amano & Sutherland 2013).

The great majority of occurrence data, although increasingly available freely and in a digital way, furthermore carry uncertainty in their geographical locations (Naimi et al. 2011). Location uncertainty arises from various factors, including inaccuracy in the measurement of location, failure to specify the geographic datum, errors in geo-referencing, and operator errors (Graham et al. 2004; Naimi et al. 2011). Recent studies have addressed the impact of positional errors in species occurrences on the accuracy of SDMs (Naimi et al. 2011). Graham et al. (2008) explored whether positional error in species occurrence data influenced SDM performance, and they compared various models in this respect. They concluded that common modelling techniques (see **2. 2. Model calibration: model types and procedures**) are particularly robust to a moderate level of positional error. Those techniques would allow

useful predictions of species distributions even when occurrence data are not free of sampling errors (Graham et al. 2008; Naimi et al. 2011).

The spatial context of species presences (and absences) in SDMs additionally has to be considered from a statistical point of view to avoid misinterpretations, as typically spatially organised biodiversity data are often spatially autocorrelated, i.e. adjacent locations or data points share more similar values than distant ones (Dormann et al. 2007, Kühn & Dormann 2012). If this pattern propagates to the residuals of a model, one of the key assumptions of standard statistical analyses, i.e. that residuals are independent from each other identically distributed, is violated. This violation in turn might lead to biases in model parameter estimation. If the presence of spatial autocorrelation is detected in model residuals (e.g. using *Moran's I,* or *Geary's C)*, it is therefore strongly recommended to use methods that account for this phenomenon. Otherwise there is a high chance of misinterpreting observed and detected patterns (Kühn 2007).

Temporal uncertainty

Under the current rate of global change, successful biodiversity conservation and management increasingly depends on a meaningful comparison among recent and historic conditions (Barnosky et al. 2017). In contrast to spatial biases, where baselines to inform conservation strategies are identified more commonly, such baselines are predominantly missing for temporal uncertainty, i.e. it is uncertain which changes and progresses about which time frames have to be evaluated. Lacking such temporal baselines is a major source of uncertainty while implementing effective biodiversity conservation strategies (Mihoub et al. 2017, see **Case Study I**). Moreover, predictions of biodiversity trends will substantially benefit from such baselines; one of the major calls of Deliverable 4.2 'Report on projections of range and biodiversity changes using improved European data sources'. Within the scope of this deliverable, UFZ followed up on this topic and assessed the beginning and temporal consistency of European monitoring schemes over time.

Case study I: *Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures*

Partners involved: UFZ (Jean-Baptiste Mihoub, Klaus Henle, Dirk Schmeller), EBCC-CTFC (Nicolas Titeux, Lluis Brotons), NHM (Neil A. Brummitt).

- For more details, see: Annex 7.1
- For the original paper, see: J.-B. Mihoub, K. Henle, N. Titeux, L. Brotons, N. A. Brummitt and D.S. Schmeller. Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. 2017. Scientific Reports 7: 41591.

a) Context

Our ability to define relevant temporal baselines for biodiversity is still limited. Nonetheless, temporal baselines are needed for biodiversity, in order for the change in biodiversity to be measured over time, the targets for biodiversity conservation to be defined and conservation to be implemented and evaluated. In this respect, the lack of knowledge about biodiversity

states prior to the rise of harmful anthropogenic activities is a critical limitation for our understanding of the full impact of such pressures as well as of past and therefore current changes.

b) Concept / Objective

Most structured biodiversity monitoring schemes have been initiated within the last few decades, whereas most of the anthropogenic pressures that are currently impacting biodiversity have been operating over centuries or even millennia. Although the limitations of biodiversity information available from monitoring schemes are widely recognized, a comprehensive and quantitative evaluation of the potential of monitoring schemes to identify temporal baselines capturing the impacts of major anthropogenic pressures on biodiversity is still lacking. We conducted a quantitative evaluation of the temporal baselines that could be identified using comprehensive information on biodiversity monitoring schemes sourced from several meta-databases.

c) Data and methods

We focus on Europe as one of the regions of the world with the oldest and most intensive biodiversity monitoring efforts. We report the start of European biodiversity monitoring schemes to examine the possibilities offered by available data for documenting past states of biodiversity with respect to different (i) taxonomic groups, (ii) EBV classes and (iii) types of data collected. Then, we compare the onset of biodiversity monitoring schemes with historical time-series or reconstructions of the main anthropogenic pressures that are currently acting on biodiversity at global or regional scales.

For each taxonomic group studied, type of data collected and EBV class targeted, we calculated descriptive metrics of the temporal baseline that could be drawn for biodiversity based on the starting year of the biodiversity monitoring schemes in Europe. We then compared the start of biodiversity monitoring schemes with global or regional long-term time-series reflecting the major anthropogenic pressures that are known to impact biodiversity the most.

We first identified the value of the pressure p_i corresponding to the starting year of each scheme i by projecting the intersect between the starting year of the scheme i and the regression trend of the pressure on the pressure axis. We then determined the level of pressure reached at that time, expressed as the percentage of the pressure range already reached when the schemes started, as follows:

$$\% \text{ pressure range reached} = \text{medP} - \text{minP}/\text{rangP}$$

where the medP is the median of all p_i, minP is the minimum value of the pressure over time and rangP is the known range of that pressure, which was calculated as the difference between the maximum and minimum values of the pressure along the time-series.

d) Main results

Most of the major anthropogenic pressures that are known to impact biodiversity began hundreds of years earlier than the start of biodiversity monitoring schemes (**Figure 2.1**). More importantly, anthropogenic pressures started to escalate exponentially from the beginning or the middle of the 20th century, while the vast majority of biodiversity monitoring schemes

started only after these pressures had already reached more than half of their present-day order of magnitude or had already peaked and decreased.

Setting temporal baselines from biodiversity monitoring data would therefore underestimate the full range of impacts of major anthropogenic pressures. In addition, biases among taxa and organization levels provide a truncated picture of biodiversity over time. In terms of median starting dates, birds and fishes are the focus of the oldest schemes, whereas schemes focusing on amphibians, molluscs, plants and reptiles are more recent (approximately a decade later).

Comparisons of starting years among EBV classes and types of data collected were only possible for a reduced set of monitoring schemes. The overall picture of the start of monitoring schemes dating back to the mid 1990's is consistent with the findings resulting from all databases previously found for the taxonomic groups.

e) Conclusion & Implications for considering uncertainties in biodiversity modelling

We demonstrate that most of the data currently available from European biodiversity monitoring schemes have been collected from the 1950's onwards, i.e. long after modern anthropogenic pressures might have started to impact species populations and communities. These limitations need to be explicitly acknowledged when designing management strategies and policies as they seriously constrain our ability to identify relevant conservation targets aimed at restoring or reversing biodiversity losses.

We argue that information derived solely from current biodiversity monitoring schemes is not well suited to setting relevant temporal baselines. To face this important challenge, we encourage both scientists and policy-makers to adopt a more conservative attitude toward temporal baselines for biodiversity by explicitly recognizing the uncertainties associated with current limitations. This implies acknowledging limits to our ability to document past biodiversity states from monitoring schemes, and that the changes measured from these schemes may seriously underestimate the full impact that major anthropogenic pressures have had on biodiversity. A consistent integration of fragmentary information across disciplines are critical if we are to set temporal baselines for biodiversity that reflect past states of biodiversity before the rise of major anthropogenic pressures.

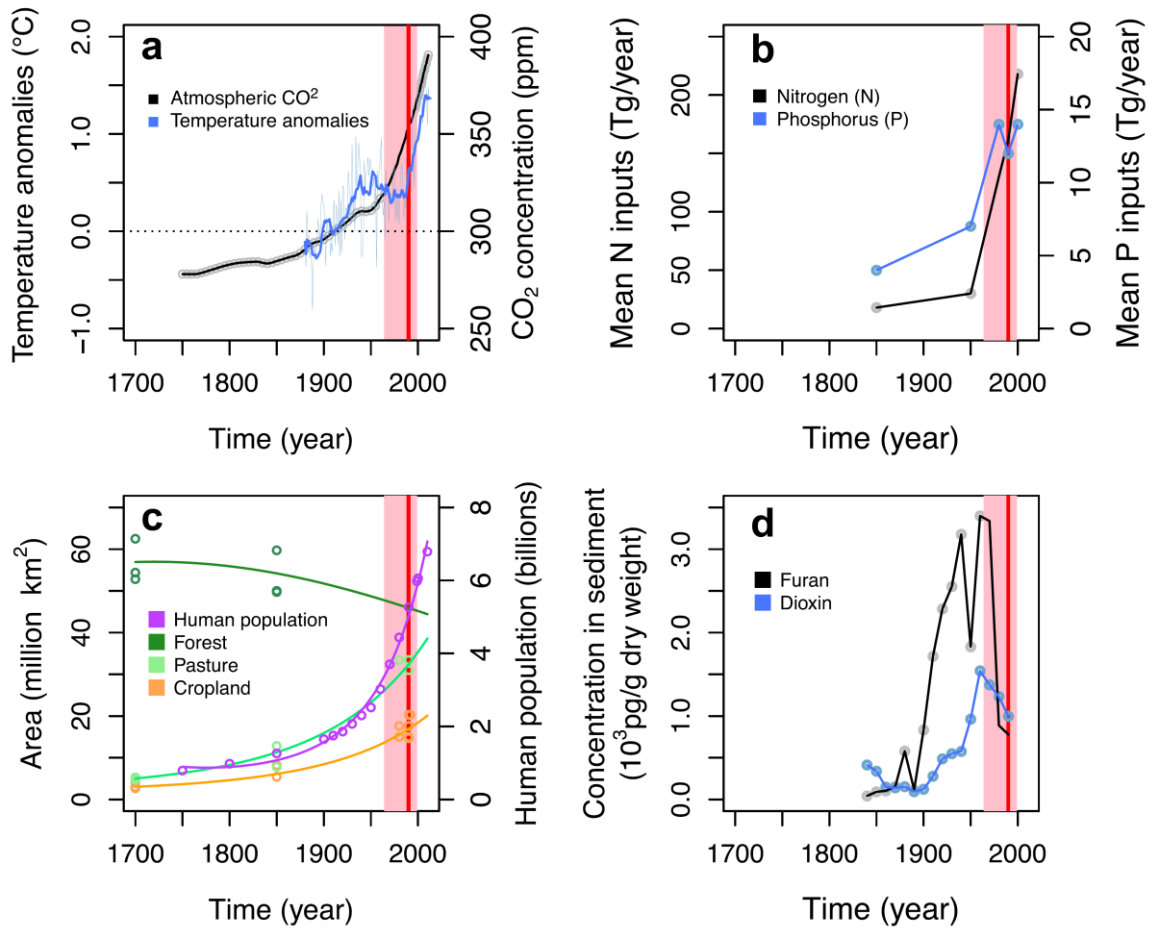


Figure 2.1: *Temporal mismatch between biodiversity monitoring schemes in Europe and major global or regional anthropogenic pressures known to impact biodiversity. The onset of biodiversity monitoring is represented using the median value (vertical red line) and the first and third quartiles (light red area) of the starting years of biodiversity monitoring schemes. Major pressures include (a) climate: global temperature anomalies and European atmospheric concentrations of carbon dioxide, (b) global anthropogenic nitrogen and phosphorus, (c) global human population sizes and global land use changes and (d) pollutant emissions in the United Kingdom (UK).*

Taxonomic uncertainty

Different taxonomic groups respond differently to environmental drivers (*inter-taxon differences*; Mora et al. 2008, Meyer et al. 2015, Meyer 2016, Kuemmerlen et al. 2016 → **case study VI**). Therefore, it is of crucial importance that we drive efforts to harmonise data quality between different taxonomic groups. Typically, plants, butterflies, mammals and birds are the most prominently investigated groups, as they historically hold the highest information density at least in Europe, due to long-lasting scientific and public interest. Nevertheless, assessing the status of 'less popular' taxonomic groups will lead to a full understanding of biodiversity and ecosystem change over space and time. Until then, a lower availability of information will continuily result in a smaller feed of information into the modelling process and finally in a shortfall regarding the prediction of future changes in biodiversity entities. This is of special importance, when you consider that groups of species do not act independently of each other, but are tied in numerous interactions, which will be altered when changes in biodiversity occur. Striving for a more complete data coverage across taxonomic groups will enhance our understanding in this regard.

As a second point, also subspecies and populations can respond differently to environmental drivers (*intra-taxon differences*). Therefore, uncertainties and errors arising from inconsistent usage of taxonomic classification and/or falsely conducted identifications (Jansen & Dengler 2010, Ahrends et al. 2011) will lead to biased and potentially misleading implications that inform management and conservation decisions.

Insights from Remote Sensing in mitigating uncertainty

Satellite remote sensing can deliver long-term data sets with an extremely high sampling frequency as well as an extensive geographic coverage (Pettorelli 2014, Skidmore et al. 2015, Proença et al. 2016). Whereas remote sensing has already been established as a powerful tool to gain global information on environmental predictors, it becomes increasingly common and approved in the direction of biological response variables (Schimel et al. 2013, Skidmore et al. 2015, Pettorelli et al. 2016, Rocchini et al. 2017) as a time and cost effective way for providing explicit maps of species distributions. Therefore, remote sensing acts as a key tool for deriving spatially explicit ancillary variables, such as climate-related drivers or biomass estimates (Feilhauer et al. 2011). Meaningful information derived from remote sensing might also be based on the classification of remotely sensed satellite or airborne multi- or hyper-spectral images to create habitat maps, which are related to species distributions.

Mapping and modelling the complexity of ecosystems and their changes over time is a key issue in spatial ecology and biogeography. Evidence exists that abrupt classification of vegetation types, especially at the species level, can present misleading or even erroneous results (Schmidtlein & Sassin 2004). Usually, vegetation assemblages show changes along environmental gradients (e.g., moisture or soil type) and therefore transitions are normally not abrupt. Alternative approaches like ordination methods aim to extract major floristic gradients describing the variation of the assemblages as metric variables, thus still retaining the continuous character of the data (Schmidtlein & Sassin 2004). Gradients can be related to any sort of remote sensing data using regression approaches, such as generalized linear models or partial least square regression (Feilhauer et al. 2011).

Regardless of the method being used, but also true for (satellite) remote sensing techniques, the assumptions for carrying out classification are associated with one major drawback: classes are mutually exclusive and have discrete boundaries separating each other. Hence, processing and classifying images can result in a substantial loss of information, due to the degradation of continuous quantitative information into discrete classes (Palmer et al. 2002; Foody & Cutler 2003). Classification can implicitly degrade information and increase uncertainty in the data and related outcomes. The uncertainty related to the classification process often remains hidden in the output maps, thus it cannot be readily accounted for. In other words, the error produced during the classification process is not accounted for in the output. From that, two sources of uncertainty can be defined in the classification of remote sensing data; (i) vagueness, namely the lack of sharpness of relevant distinctions, and (ii) ambiguity, arising from conflicting distinctions (discordance; Klir & Wierman 1999, Rocchini & Ricotta 2007).

It has to be noted that remote sensing information on the distribution of species and other entities of biodiversity modelling provides not only valuable information on their current association with different environments, but this information may serve as a baseline for conditions that will exceed the assessed steady-state conditions (Schimel et al. 2013). As the rate of environmental change is high and will lead to new combinations of environmental conditions it is of upmost importance to gain information on such baselines as soon as possible (Schimel et al. 2013). Therefore, a current additional challenge is to link remote

sensing data to local observations, to finally upscale those measures by generating models on ecosystem processes (GEO BON 2015, Proença et al. 2016) and potentially Essential Biodiversity Variables (Skidmore et al. 2015, Pettorelli et al. 2016).

The following case study by FEM gives an example on how biodiversity metrics can be assessed directly via satellite remote sensing to overcome limitations by directly deriving them from field observation (**Case Study II**). In Detail, FEM introduces an R package to apply Rao's Q to remotely sensed data to furthermore overcome limitations of commonly used diversity metrics.

Case Study II: *Measuring Rao's Q diversity index from remote sensing: An open source solution*

Partners involved: FEM (Duccio Rocchini)
+External Partners:

Technische Universität Berlin, Department of Landscape Architecture and Environmental Planning, Geoinformation in Environmental Planning Lab, Berlin, Germany (Matteo Marcantonio)
Department of Environmental Biology, University of Rome "La Sapienza", 00185 Rome, Italy (Carlo Ricotta)

- For more details, see: Annex 7.2
- For the original paper, see: Rocchini D., Marcantonio M., and C. Ricotta. 2017. Measuring Rao's Q diversity index from remote sensing: An open source solution. Ecological Indicators **72**: 234-238.

(a) Context

Estimating biodiversity from field data presents a number of drawbacks mainly related to time and costs, together with intrinsic difficulties to build standardised procedures for reproducible data gathering (Palmer et al. 2002).

It has been demonstrated that the measure being used can lead to very different (and sometimes misleading) results. As an example, one of the mostly used diversity measures of the landscape based on spectral remotely sensed data, i.e. the Shannon's entropy (Shannon 1948), has a number of implicit drawbacks like: i) the difficulty to discriminate between differences in richness or relative abundance (Nagendra 2002) or ii) the impossibility to consider spectral values as numbers instead of classes (Rocchini & Neteler 2012a). Concerning the second point, Shannon's entropy accounts for richness and relative abundance of spectral values but it does not explicitly consider the numerical magnitude (values) of pixels.

(b) Concept / Objective

On the contrary, Rao's Q index (Rao 1982) does take into account *i* and *j* value by considering their pairwaise distance d_{ij}:

$$Q = \sum \sum d_{ij} \times p_i \times p_j$$

The aim of this part of the Deliverable is to solve the aforementioned issue, by the application of Rao's Q to remotely sensed data in an open source environment (e.g. Rocchini & Neteler 2012b), providing a straightforward R function to calculate it in 2D systems. As far as we know, this is the first attempt to measure Rao's Q in a 2D space applied to remotely sensed data.

(c) Data and methods

The function *spectralrao()* to derive Rao's Q, written in the R statistical language (R Core Team 2016), is stored in the GitHub repository https://github.com/mattmar/spectralrao. The function accepts matrix, RasterLayer or SpatialGridDataFrame object as input (or a list of them).

(d) Main results

For the synthetic set of data of **Figure 2.2**, H is of low applicability, due to the high heterogeneity in the input data. On the contrary, Rao's Q meaningfully highlighted the areas with higher heterogeneity since it allows considering distances together with relative abundance of values, the intersection between the simulated submatrices.

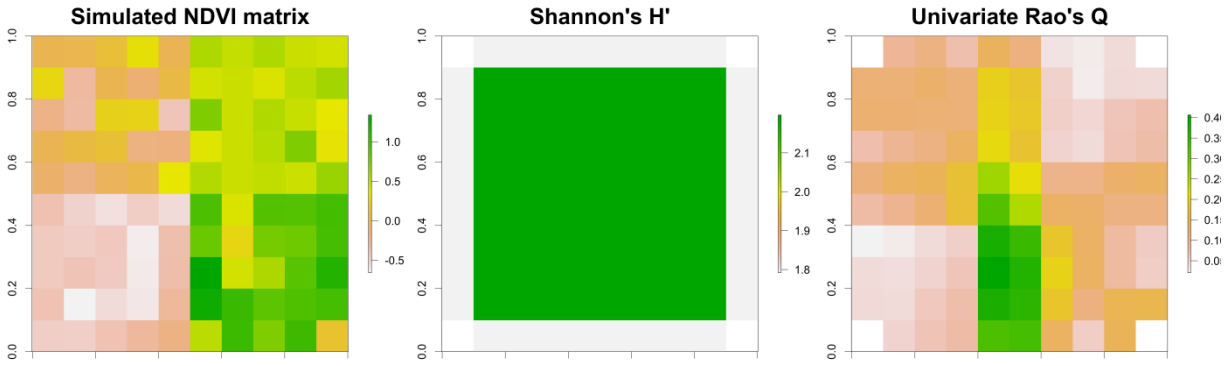


Figure 2.2: *An example of the calculation of Rao and Shannon indices on a hypothetical NDVI image. In this case, Shannon index tends to overestimate diversity since it considers the differences in the abundance of classes, while Rao Q seems to be more reliable taking into account their distance.*

In **Figure 2.3**, it is apparent that H tends to saturate in case of high diversity since in the local 9x9 pixels window of analysis all the pixel values, even though similar among them, are still different. As a consequence, since H does not take into account their distances but only their relative abundances; its value will always approximate saturation. On the contrary, Rao's Q overcomes this limitation by the pairwaise distance term.

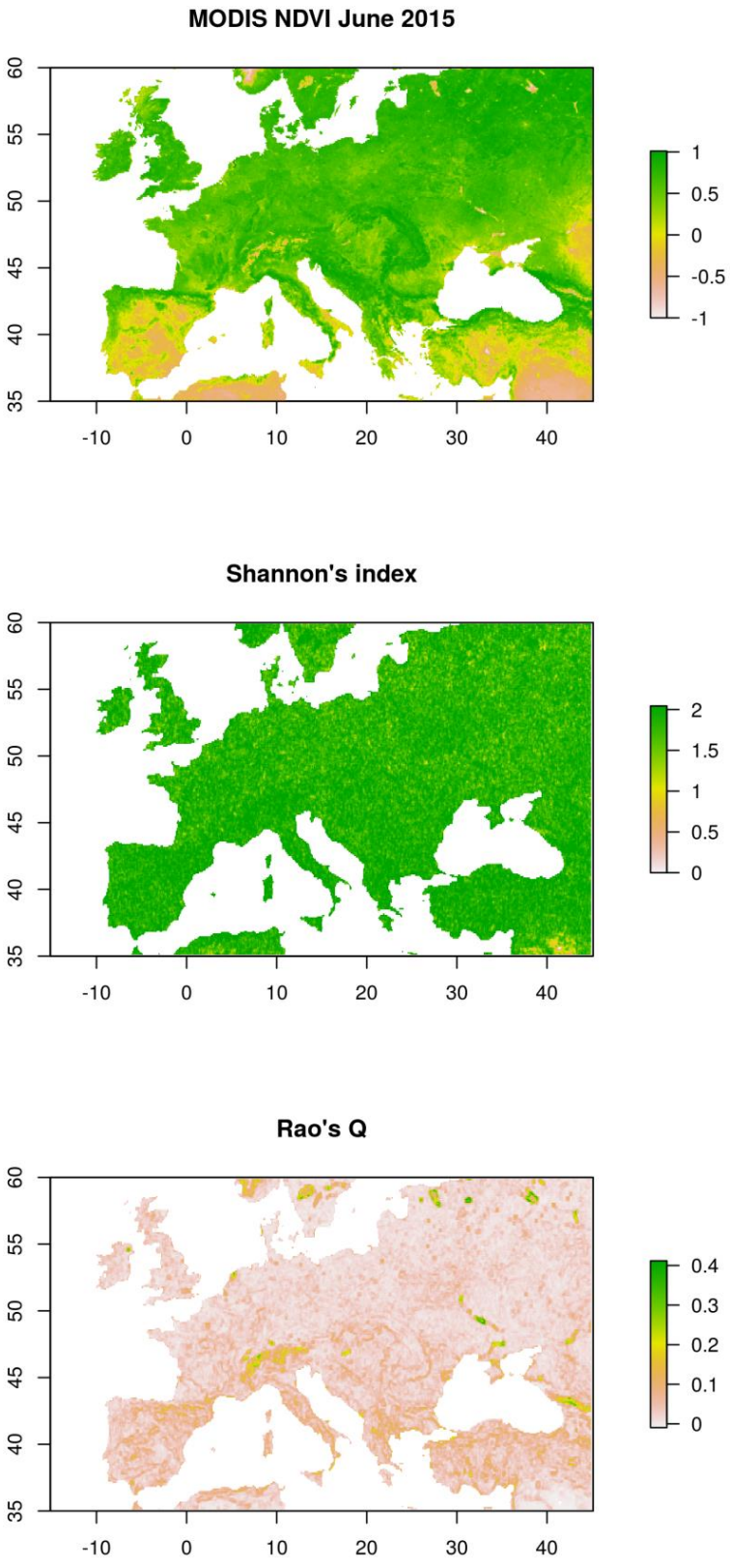


Figure 2.3: *In this MODIS Normalised Difference Vegetation Index (NDVI) image at 0.1 degrees resolution of June 2015, the Shannon and the Rao indices are calculated. While Shannon tends to saturate towards higher values, Rao's Q is not affected by small differences between pairs of pixel values.*

(e) Conclusion & implications for considering uncertainties in biodiversity modelling

Dealing with digital images, an advantage of Rao's Q over more conventional diversity measures, is that, while the calculation of H relies solely on the relative proportion of the digital numbers (DN), Q takes also into account their pairwise differences. Moreover, while H is usually calculated on one single band at a time, Rao's Q can accommodate multivariate differences between DNs. Indeed, it can be calculated on multiple bands, hence representing the DNs dispersion in a multivariate space.

Rao's Q has been estensively used in functional diversity application (Botta-Dukat 2005; Ricotta & Moretti 2011). Functional ecologists make use of a wide set of functional traits (plants functional characteristics) to assess the diversity of natural systems. Rao's Q has been shown to be a valide candidate to summarize them in a single diversity value (Botta-Dukat 2005). However, as previously stated, this is the first application of the Rao's Q in a 2D space with remotely sensed data.

Due to its flexibility, Rao's Q based on the aforementioned multivariate distances may be helpful to optimize the relationship between biodiversity values recorded from remote sensors and species inventories recorded from field observations.

Essential Biodiversity Variables

The concept of Essential Biodiversity Variables (EBVs) was developed by the GEO BON framework in order to measure changes in biodiversity, to fill existing gaps in available biodiversity information and to form consensus and basis for future monitoring programs of the world (Pereira et al. 2013, Brummitt et al. 2016). The concept is derived from Essential Climate Variables and aims at identifying a minimum set of biodiversity variables to inform scientists, managers and the general public in global biodiversity change (Proença 2016). Thereby, an EBV is defined as '*a measurement required for study, reporting, and management of biodiversity change'* (Pereira et al. 2013) and is sofar mainly related to patterns but not processes. Currently, the propsed EBVs can be clustered into six classes: (1) genetic composition, (2) species populations, (3) species traits, (4) community composition, (5) ecosystem structure and (6) ecosystem function.

To this point, EBVs are not yet a modelled entity; as it is still debated which EBVs will be the most representative and available primary data are still being assessed for this purpose (Proença et al. 2016). For this reason, EBVs are not a particular focus of this deliverable. Nevertheless, following the present and perspective importance in stimulating and unifying global monitoring efforts, the EBV concept is a crucial part of future issues regarding biological response data in biodiversity modelling and has to be taken into account in discussions of future recommendations and actions in conservation and management of biodiversity.

2.1.2 Environmental information and predictors

Contrasting biological response data, environmental predictors do not suffer as strongly from any kind of detection bias, as they are in general comprehensively measurable entities. Nevertheless, uncertainties surrounding environmental predictors still occur due to measurement errors (related to the variety of used instruments and measurement techniques/schemes) and due to the basic selection of the dimension of raw environmental information we choose to measure at all (Wu and Li 2006). Furthermore, it has to be remembered that, especially at larger scales, environmental predictors are often themselves

predictions from models based on data interpolation, which are in turn based on fewer sample locations or satellite-derived data (Beale & Lennon 2012).

Moreover, downscaling of coarse resolution predictors plays an important role (Wilby et al. 1998, Tabor & Williams 2010, McInerny & Purves 2011) and collinearity of predictor variables has been identified as a crucial point in variable selection (Dormann et al. 2013); with the selection of methods to account for collinearity influencing the model's certainty itself (Dormann et al. 2013).

Collinearity treatment

Treating collinearity of environmental predictors can also be seen as a part of the modelling calibration process. Nevertheless, we argue that it already contributes to the step of predictor selection prior to the actual model calibration and therefore belongs to the section of handling the available data.

Including highly collinear environmental variables induces the risk of reducing the accuracy of parameter estimation and superimposing the detection of important environmental predictors that significantly shape the modelled distribution range. Therefore, not using a treatment that accounts for collinearity is a form of misspecification of a (statistical) biodiversity model (Beale & Lennon 2012). Braunisch et al. (2013) and Dormann et al. (2013) compile the most comprehensive reviews regarding the effect of the method being used to account for collinearity in SDMs.

Dormann et al. (2013) review the existing methods to account for collinearity and distinguish between collinearity treatments that are either (1) based on clustering algorithms (**P**rincipal **C**omponent **A**nalysis, **P**rincipal **C**oordinate **A**nalysis, **C**orrespondence **A**nalysis, **N**on-**M**etric **M**ultidimensional **S**caling, **V**ariance **I**nflation **F**actor Analysis), (2) not based on clustering algorithms (absolute value of correlation coefficients, sequential regression) or (3) modelling techniques that directly incorporate collinear predictors (latent variable approaches – **P**rincipal **C**omponent **R**egression, **P**artial **L**east **S**quares, Dimension Reduction). They run a simulation study using several degrees of collinearity and conclude with a list of most promising approaches for each of the abovementioned groups.

Braunisch et al. (2013) go one step further and explore the influence of the selection of collinear climate variables onto the projections of future ranges of bird species in Europe in 2050. They found that the selection of climatic variables were an important source of uncertainty for future range predictions that are difficult to control by using contemporary information and conclude that substantial differences based on predictor selection have to be stated, especially concerning possibly derived conservation actions.

Uncertainties associated with spatial resolution

Spatial resolution is critical to any study of species distributions (Lauzeral et al. 2011; Jimenez-Alfaro et al. 2012) and uncertainty arises from the fact that environmental predictors, which are used to describe species' niches in SDMs, often are recorded at different spatial resolutions (cell or grain sizes; Lauzeral et al. 2011). Furthermore, it has to be acknowledged that different environmental predictors differ in their importance on species diversity (e.g. Jeliazkov et al., 2014, 2016) and distributions and ecosystem functions with scale (Carl et al. 2016, see **Case Study III**). Thus, spatial uncertainty related to issues of scale should be taken into account in the interpretation of species distribution maps, and to prioritise areas in which further monitoring must be performed to obtain reliable knowledge of the distribution of a particular species, supporting conservation decisions and allocation of efforts (Rocchini et al. 2011).

Issues of scale in predictions of species distribution have clear effects on the implementation of management activities in biodiversity conservation. Coarse grain sizes are associated with high uncertainty in the spatial distribution of species and habitats (Hermoso & Kennard 2012). For example, a species could be expected to be present throughout a large area when coarse grains are used to represent its distribution, even though the species occupies a small portion of that area only (Hermoso & Kennard 2012). Such issues related to spatial resolution in species distribution modelling can in turn strongly influence conservation management and policies. For instance, coarse resolution can bring significant degree of uncertainty in species category listing for the IUCN Red List (Rocchini et al. 2011) through directly influencing measurement of Extent of Occurrence (EOO) or Area of Occupancy (AOO) (Jiménez-Alfaro et al. 2012) used for IUCN criteria A, B and D.

Overall, the selection of scale for data collection (and inference) is crucial in statistical modelling. At the same time, different predictors will be relevant at different scales and these scales may in turn differ from previously specified collection units. So far, a study that assesses such a scale-dependency as well as the statistical tool to do so is missing. The following case study will close this gap by presenting such a method utilising a two-dimensional wavelet analysis (**Case Study III**).

Case study III: *Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis*

Partners involved: UFZ (Gudrun Carl and Ingolf Kühn + Daniel Doktor, Oliver Schweiger)

- For more details, see: Annex 7.3
- For the original paper, see: Carl G., Doktor D., Schweiger O. and I. Kühn. 2016. Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis. Journal of Biogeography **43**: 2502-2512.

(a) Context

Assessing the relationship between a spatial process and environmental variables as a function of spatial scale is a challenging problem. Data collection for biogeographic and environmental data is frequently carried out with reference to a gridded map of a specific resolution. A statistical model based on these data will provide statistical inferences at this specific spatial scale.

(b) Concept / Objective

Because different (e.g., biological) processes act at different scales, multiple relationships are scale-specific as well. Hence, the selection of scale for data collection and inference is crucial in statistical modelling. In general, however, different scales will be relevant in such multiple relationships and some of them will be different from the pre-specified collection unit. As a consequence, conclusions based on regressions of these data, i.e. its parameter estimates, hypotheses tests and P-values, may be misleading and can result in incorrect inferences. At least, this is the case if we ignore that these conclusions are restricted to a particular scale and disregard the complexity and multi-scaled structure of the problem. Therefore, there is a need for a valid and reliable tool to examine and evaluate scale dependencies.

(c) Data and methods

Analysing scale dependency, one has to be very clear about the four different components of scale (Scheiner et al. 2000): (i) sample unit, (ii) grain, (iii) focus and (iv) extent. Sample unit refers to the spatial dimension of the collection unit (e.g. sampling plot). Grain is the smallest unit to which all sample units are standardized for a specific analysis (i.e. finest resolution). The units of grain can then be aggregated to coarser units of analysis, that is, focus (i.e. coarser resolution). Extent in this context is the complete geographic area sampled.

We present a method for applying two-dimensional wavelet analysis to a generalized linear model. The main advantage of scale-specific wavelet regression is that it differs from previous methods, which simply upscale data by averaging of aggregated cells and thus regress variables of enlarged grain size. Instead, wavelet analysis is able to extract scale-specific variations of both dependent and independent variables. Therefore, a wavelet regression can measure how a change in environmental variables at a given resolution (i.e. focus) influences change in the response variable at the same resolution (Ye et al. 2015).
Our method has the advantage that all calculations were done in a single framework. Firstly, the wavelet approach is carried out by means of multiresolution analysis, which is able to decompose gridded data (maps or images) into components at different resolutions. This data decomposition is embedded into the framework of a multiple regression analysis (Carl and Kühn 2008, Keitt and Urban 2005). This way we are able to develop a regression technique that allows for scale-specific regressions. Note that this kind of regression aims at scale dependent investigations. In detail, this means that a scale-specific regression accounts for fluctuations or spatial variations at a specific spatial resolution. This wavelet multiresolution regression (WMRR) also allows response vectors of binary or Poisson distribution. Therefore, our WMRR approach is a method for applying two-dimensional wavelet analysis to generalized linear models. Secondly, applying all regressions in a multimodel inference approach circumvents a common problem: Using separate regressions for each scale will result in multiple testing. Due to decreasing sample size, hypothesis tests have declining power. Therefore, results cannot be compared by means of hypothesis testing. The multimodel inference approach does not suffer from this problem. It calculates variable importance by using an information-theoretic approach based on Akaike weights (Burnham and Anderson 2002). Since results obtained from finer to coarser scaled data can then be compared, one is able to examine the effect of scale dependencies and to evaluate the relative importance of several environmental variables across different spatial scales. Therefore, we provided an answer to the key question whether similar mechanisms act at different spatial scales.

To illustrate our new up-scaling method, it is necessary to use data at medium to large extent and fine sample unit because sample unit acts as a preset for the grain (i.e. finest resolution) in the analysis. Scale dependency is then studied by leaving extent and grain constant and aggregating grains to coarser resolutions (i.e. foci). Hence, to discuss scale dependency, i.e. alterations in the relative importance of different environmental factors caused by increasingly coarser resolutions (foci), we need data collected over a regular grid consisting of sufficient grid cells. Therefore, in our case study, we examine data sampled on a map of 1024×1024 grid cells at $0.01° \times 0.01°$ resolution (grain), i.e. approximately 1×1 km², in Central Europe. We focus on the relationship of vegetation greening to climate, topography and land use. Remote-sensing vegetation indices based on satellite observations indicating the vegetation activity (Yang et al. 2012) were used to estimate the vegetation period per year (White et al. 2003). We use vegetation period as a response variable that is regressed on climate, topographic and land use data.

(d) Main results

To the best of our knowledge, this is the first study analysing the scale dependency of an ecosystem process, clearly distinguishing between the different components of scale, namely extent, grain and focus, having an extremely large sample size (n = 1048576), and covering a large range of different resolutions (c. 1 km² to c. 1000000 km²).

Our results indicated that the relative variable importance detectable by scale-specific regressions is strongly scale-dependent. Moreover, for two different approaches, (i) leaving grain and extent constant and changing focus and (ii) leaving extent constant and changing grain, we were able to demonstrate how 2-D scale dependencies can be systematically analysed. It was shown at which "scale" the turning point is where drivers change in importance.

(e) Conclusion & implications for considering uncertainties in biodiversity modelling

We believe that our newly proposed method is particularly suitable for studying scale dependencies of various spatial processes on environmental drivers having gridded data with sufficiently large sample size.

2.2 Model calibration: model types and procedures

SDMs are used to estimate species' spatial distribution from available data sets, including data on observed species occurrences and environmental information. SDMs have long been established as essential tools in biodiversity conservation (Guisan & Zimmermann 2000, Guisan & Thuiller 2005, Elith et al. 2006, Elith & Leathwick 2009) and their estimates of species distributions have been widely used to assess the effects of climate change and to develop adaptive strategies for managing biodiversity and conserving habitats. Accounting for uncertainty within these model frameworks is critical for an effective planning of such actions (Wenger et al. 2013).

The fast development of SDMs, that are based on a variety of statistical and also machine-learning techniques and are supported by the fact that computational efficiency is no longer a major problem, has formalised the process of applying them to questions of current states and changes of species ranges (Beale & Lennon 2012). Nevertheless, it has to be constantly assessed whether the objectives of these applications do match the purpose they are meant to inform (Guillera-Arroita et al. 2015). Selected models should not only match objectives of a study, but also the available expertise, knowledge and the type and amount of data available. Additionally, model calibration includes several inevitable sources of uncertainty in model-based statistical analyses of species ranges. Among these are (1) variable selection, i.e. uncertainty resulting from different variable selection approaches and (2) model types, i.e. uncertainty resulting from differences in selected algorithms.

Variable selection

Different approaches of variable selection and/or model simplification (i.e. identifying non-significant predictors) are surrounded by different levels of uncertainty and can be conducted manually or by automated stepwise procedures. Murtaugh (2009) compiled a comprehensive assessment of the performance of several variable selection methods.

A *forward selection* of predictors is generally not recommended, as the selection of which predictor comes in first is often arbitrary and therefire based on a too high uncertainty. Contrastingly, a *backward selection* starts with a full model including all selected predictors and reduces the model based on the chosen criteria. At this point, it has to be highlighted that variable selection/model simplification by error probabilities or by *information criteria (e.g. Akaike Information Criterion - AIC or Bayesian Information Criterion - BIC)* are two different conceptual (or even philosophical) approaches and should not be combined. The former are more suitable for hypothesis testing and the latter are more suitable for descriptive models; depending on the particular modelling approach, one must make the choice of one or the other – a guideline that is still not followed by the entirety of biodiversity modellers.

Within the selection criteria, a selection process following BIC was shown to favour small models in comparison to AIC. A corrected AIC (AICc) can be used to account for a small ratio between sample size and number of predictors (rule of thumb: sample size/number of predictors < 40).

Model types

A diverse set of modelling algorithms exists and is currently used in species distribution modelling. As it is the case for literature regarding data in biodiversity modelling, several publications deal with the issue of assessing the different performances of those algorithms (e.g. Dormann et al. 2008b, Watlling et al. 2015) under different backgrounds of research questions and possible conservation targets. Indeed, different modelling algorithms inherit different level of uncertainties and especially forecasts of future distributions vary with selected model type (Peterson et al. 2007, Gritti et al. 2013). Therefore, we will characterise basic groups of SDM approaches in the following. It has to be noted that the comparison of these model types can lead to insights into processes and mechanisms, as they differ in their underlying assumptions (Franklin 2013).

Correlative or niche models

A vast majority of SDMs rely on niche-based models that are based on correlative methods, i.e. linking current species occurrences (presence only or presence-absence data) with environmental predictors of recent and potential future ranges. Correlative SDMs assess the realised niche of the target species. They are most commonly using a mapped landscape for which the predictor values are known and are able to produce a similar map of species probability of presence (and potentially absence). Correlative SDMs are relatively easy to implement, as they are only making basic statistical assumptions, use most common type of information on species occurrences and belonging statistical methods include general statistical approaches. Dormann et al. (2008b) classified these techniques into three groups: i) 'traditional' methods, such as Generalized Linear Models (GLM), Generalized Additive Models (GAM), Discriminant Analysis, Classification And Regression Trees (CART), ii) 'machine-learning' methods, such as Artificial Neural Networks (ANN), Boosted Regression Trees (BRT), Random Forest, Multiple Adaptive Regression Splines, Support Vector Machines (SVM), among others, and iii) 'presence-only' techniques, such as Genetic Algorithms and Rule-set Prediction, Maximum Entropy, and Environmental Niche Factor Analysis (ENFA) (Dormann et al. 2008b).

Nevertheless, classic correlative SDMs do not incorporate representations of biology/physiology of the target species, assume stationarity (i.e. model parameters are constant in space and time) and are therefore prone to overfitting and misleading predictions while extrapolating to different locations or times (Dormann et al. 2012, Heikkinen et al.

2012). It is therefore important to note that predictions based on purely correlative SDMs are usually limited in their transferability to other environments. Within frameworks of correlative SDMs, their transferability would certainly benefit from accounting for the abundance of species instead of their mere presence/absence as well as from high-resolution data on environmental predictors, to capture fine-scale processes affecting species distributions (McInerny & Purves, 2011). A typical validation technique of correlative SDMs is cross-validation with an independent data set. Nevertheless, independent data sets, which are then not used to train the model, are still often rare.

Process-based or mechanistic models

Contrasting correlative models, process based SDMs specify responses of selected (physiological) traits or processes to environmental predictors and estimate proxies of occurrences that are related to fitness parameters (Gritti et al. 2013); assessing the fundamental niche of the target species. Moreover, such SDMs aim at incorporating biotic interactions and feedbacks into the modelling frameworks (Boulangeat et al. 2012, Dormann et al. 2012, Singer et al. 2016), i.e.intra- and interspecific interactions, dispersal limitations/abilities and evolutionary and life-history components. The development of these process-based models in distribution modelling is a promising way to cope with exacerbated structural uncertainty related to future changes. Following that thought, process-based models allow the unification of the strategic vs. tactical modelling frameworks and provide both higher understanding and predictive capability (Evans et al. 2012, 2013a, 2013b). Additionally, they are said to be more robust against extrapolation to non-analogue conditions, i.e. novel climates or distant spaces (Dormann et al. 2012, Gritti et al. 2013), as they incorporate a higher degree of biological or structural realism. At the same time, process-based models require a higher number of paramaters and therefore different data sets and data types than correlative models, as well as more accurate knowledge of important underlying processes driving the dynamics of the system under investigation. Therefore, process-based models are to a higher degree subjected to uncertainty propagation arising from parameter estimation error or from the conceptual formalisation of underlying processes into a modelling framework, which might decrease their practicality (Beale & Lennon 2012). Singer et al. (2016) argue that because of the limitations in knowledge that leads to strong restrictions in parameterisation of process-based SDMs, as well as the sheer complexity of the included biotic relations, they are currently still limited in delivering reliable projections; a trade-off between structural realism and projection reliability. Yet, the authors additionally suggest a protocol to fill knowledge gaps to increase reliability of SDM projections by combining empirical and modelling effort (Singer et al. 2016).

Hybrid models

Hybrid models are the result of a combination of correlative and process-based models, by associating correlative approaches to describe the realised niche of a target species and by inferring from population dynamics or physiological processes (Peterson et al. 2011, De Cáceres & Brotons 2012, Gritti et al. 2013).

Model averaging / Ensemble modelling

While modelling the distribution of species, taxa or other biodiversity entities there most often exist more than one well-founded model as basis for statistical inferences, which introduces a different kind of model uncertainty. The technique of model averaging was developed to improve the predictive ability of statistical models by combining predictions from a set of

models and therefore incorporating the uncertainty surrounding the final model selection. Since the influential publication by Burnham & Anderson (2002) on the concept of multi-model inference using AIC, it has experienced a strong increase regarding its application in biodiversity modelling (Banner & Higgs 2017). Whenever there is uncertainty about which model is the best, it is recommended that interpretations are based on a set of models (Burnham & Anderson 2002). Nevertheless, the interpretation of a set of several best models is often difficult, especially if the sets of predictors, they are each based on, differs substantially. In such cases it is typically suggested to find a consensus model that persists of parameters that are averaged across the different models (see also Araujo & New 2007).

While the concept of incorporating model selection uncertainty via model averaging is a promising approach, more and more authors raise concerns regarding a statistically sound way of using it (Cade 2015, ver Hoef & Boveng 2015, Banner & Higgs 2017). Cade (2015) identifies three flawed practices associated with model averaging and focuses on multicollinearity among explanatory predictors as point of major concern (see also Dormann et al. 2013). The author strongly argues for discontinuing simple averaging of regression coefficients stemming from different models (as suggested by Burnham & Anderson 2002) if ignoring multicollinearity among predictors, as it may result in misleading inferences. Furthermore, Cade (2015) suggests using partial standard deviations to account for the changings scales of regression coefficients among models. Banner & Higgs (2017) pick up this suggestion and developed model averaged posterior plots (MAP) that provide visual summaries of all components going into averaging of partial regression coefficients across different models. Those authors emphasise that the added model complexity due to model averaging has to be considered thoughtfully and that it has to be reviewed whether (1) inferences from model averaging change compared to conditioning on one reasonable model and (2) model averaging is needed when a well-defined research question can be soundly addressed directly by a single model. Overall, model averaging by multi-model inference, especially relying on sum-of weights has been shown to misleading assumptions about this statistical approach (Galipaud et al. 2014).

Accounting for gaps in modelled data sets

As already mention in the previos section on data in biodiversity modelling (→ **2.1. Data**), gaps are still a typical characteristic of data sets obtaining biodiversity data. One possibility to close these gaps is the allocation of resources to collect the missing data by expanding the effort on collecting raw data at sampling sites. Still, as data sets will never be complete, (1) data should be collected in a way, so that any detection bias can be recognized (Lahoz-Monfort et al. 2014) and modelled and (2) advanced statistical methods should be developed, to incorporate uncertainty stemming from gaps in the available data sets. In an excellent example, Manceur and Kühn (2014) use Bayesian Image Restoration in combination with expert knowledge to include non-detections of focal species to reduce uncertainty in SDMs (see also Bierman et al. 2010).

2.3 Model validation / Measures of accuracy

The level of uncertainty of presence/absence predictions originating from biodiversity models is influenced by the chosen threshold to transform probabilities of occurrence (habitat suitability) into binomial predictions (Rocchini et al. 2011). High threshold values infer that most locations included in the model indeed harbour the species, whereas lower thresholds will include increasingly more locations where the species may be absent in reality. The

uncertainty associated with the classification of continuous species occurrence probabilities should be considered and discussed in conjunction with SDM predictions (Peterson et al. 2007; Rocchini et al. 2011) to improve the reliability (lower global uncertainty) of the mapped locations of species occurrences. Based on this accompanied uncertainty, some authors argue against the conversion of maps of occurrence probability to presence-absence maps, which is predominantly done, as it discards information on uncertainty and gives a false sense of the confidence in the shown predictions (Wenger et al. 2013, Calabrese et al. 2014, Guillera-Arroita et al. 2015). Guillera-Arroita et al. (2015) point out that it is rarely assessed whether this binary output is indeed required for the particular research question and application or whether this loss of information (on uncertainty) could also be detrimental for the purpose of the analysis.

The common approach is to use performance or accuracy indicators (Allouche et al. 2006). The choice of model performance or accuracy indicators is an important source of uncertainty, as they frequently serve as main argument to select between different model runs or set-ups. The choice of indicators varies greatly between analyses. For SDMs, a considerable number of publications have dealt with the topic. Kappa has been identified as unsuitable for such models (McPherson et al. 2004) and AUC (area under the ROC curve) and ROC (receiver operating characteristic curve) have attracted criticism more recently (Lobo et al. 2008, Beale & Lennon 2012). So far, TSS (True Skill Statistics) has not been observed to be misleading. Those accuracy measures are useful functions for the assessment of prediction errors in presence/absence models. However, in a spatial context, the traditional non-spatial measures are not appropriate and can be misleading in distribution modelling. This is due to the fact that, in these non-spatial measures, a false prediction at a single location is simply assigned as being wrong, regardless of its distance to a correct prediction. In other words, the classical abovementioned measures suffer from the problem, that accuracy is not a function of spatial distribution. Instead, they weight all errors equally, the falsely predicted positive errors as well as the falsely predicted negative errors.

As part of the activities in WP4, UFZ developed spatial accuracy measures, which are (1) sensitive to the spatial distribution of the predictions and (2) comparative to the classical ones. Using these spatially corrected accuracy measures is proposed to account for spatial arrangements of predictors to benefit the assessment of prediction accuracy of biodiversity models (see **Case Study IV**).

Case study IV: *Spind: a package for computing spatially corrected accuracy measures*

Partners involved: UFZ (Gudrun Carl and Ingolf Kühn)

- For more details, see: Annex 7.4
- For the original paper, see: Carl G., and I. Kühn. 2016. Spind: a package for computing spatially corrected accuracy measures. Ecography **40**: 675-682.

(a) Context

Using an appropriate accuracy measure is essential for assessing prediction accuracy in species distribution modelling. Therefore, model evaluation as an analytical uncertainty is a challenging problem. Accuracy measures such as Cohen's kappa coefficient (or Kappa for

short) are coefficients useful to assess prediction errors in presence/absence models (such as species distribution models). In a spatial context, however, the traditional non-spatial measures are not appropriate and can thus be misleading in species distribution modelling (Fielding 2002). The reason is that a false prediction has simply the quality of being false regardless of its distance to an appropriate actual value and thus true prediction. One can argue, though, that a false prediction of presence in close proximity to a true (observed) presence is better than a false presence far away from an observed presence (Fielding & Bell 1997, Fielding 2002).

This is particularly the case when sampling at nearby locations leads to sample values that are not statistically independent from each other. If so, then it is to be expected that predictions have the same nature. This phenomenon of statistical dependence caused by spatial dependence should be considered as relevant. This applies particularly to sampling on raster maps, where original data maps are sectioned into grids (Hagen-Zanker 2009). Due to a relatively arbitrary specification of cell size and grid orientation, discretization will generally cause a loss of information. Occurrences at grid cell boundaries, for instance, must be allocated to a specific grid cell (ignoring proximity to the neighbour cell) (Shekhar et al. 2002).

(b) Concept / Objective

We present 'spind', a new software package (based on the R software program), which introduces several spatial accuracy measures that are sensitive to the spatial arrangement of predictions (Carl & Kühn 2016). As alternative measures for the evaluation of grid-based models, they take into account that a false prediction may not be completely wrong if it is in a certain spatial proximity to the correct result. The degree of dependency can be measured and analysed by correlograms, i.e. computations of spatial autocorrelation of both predicted and actual values.

(c) Data and methods

We were not interested in developing totally new spatial measures. Instead, the aim of our study was to generalize classical measures. To enable efficient comparisons, we modified and improved well-known measures (i.e., Kappa, as well as sensitivity, specificity, true skill statistic and other ones) to spatially corrected versions. These classical measures are based on a calculation and evaluation of a confusion matrix. This is a 2x2 contingency table that cross-classifies observed occurrences (i.e. actual presence / actual absence) and predicted ones according to two classes (i.e. predicted presence / predicted absence). The threshold dividing into classes of predicted presences and absences has frequently the value *threshold* = 0.5, but any other threshold value within the interval from 0 to 1 could be chosen, e.g. based on prevalence or maximizing traditional accuracy measures such as Kappa or true skill statistic. When setting the threshold to 0.5, then the probability of presences is the same as the probability of absences.

We implement proximity as the same amount of spatial autocorrelation in both actual and predicted values. For spatial data, the amount of spatial autocorrelation can be calculated by means of the Moran's I (e.g., Lichstein et al. 2002). This formula measures the strength of two-dimensional autocorrelation based on the assumption that it is isotropic (i.e., independent of direction). Autocorrelation is computed as a function of "lag distance", therefore, one has to introduce lag distance intervals for the spatial structure under consideration. For a square grid, the first distance class can be defined to comprise lags between 0 and 1 and thus be assigned to nearest neighbours, i.e. to the (generally) four adjacent grid cells located at distance unit 1 (in relation to coordinates of cell centres) in the cardinal directions.

Autocorrelation at lag distance 1 is generally higher than that at greater distances because close observations are more likely to be similar to one another than those far away from each other. Therefore, the autocorrelation value ac(1) is most important. It is noteworthy that the spatial autocorrelation ac(1) of predicted values (i.e., predictions before dividing into groups by a threshold) is generally higher than that of actual values. The reason is that predictions are continuous values varying between the extremes 0 and 1, whereas actual values simply consist of 0's and 1's. This autocorrelation deficit of actuals can be considered as a measure to what extent actual values can be adjusted to reflect a spatial context. Therefore, we generate "adjusted actuals" having the same amount of autocorrelation as predictions. These adjusted actual values are softened compared to the original ones and, accordingly, appear widened in spatial mapping. Therefore, a prediction at a single location can be registered to be in the proximity (i.e. widened area) of an actual value. It is to remark, that, computationally, it is difficult to increase the autocorrelation of actuals in one step to a certain level. Here, we use a step-by-step procedure incorporating autocorrelation until it is balanced with the autocorrelation of predictions.

For evaluation, one has to summarise the results for predicted and adjusted actual values in a generalized confusion matrix. In order to ensure that the additional information captured in adjusted actual values is not completely lost again, it is necessary to make the contingency table "finer". If we cross-classify the distributions of the variables in a 4x4 contingency table then we are able to distinguish different kinds of misclassification. Therefore, the predicted values have to be classified into 4 classes separated at the following 3 levels: (1) upper split: *us = (1+threshold)/2*, (2) threshold: *th = threshold*, *and* (3) lower split: *ls = threshold*/2. Since the total of elements remains constant, a comparison to the results of a 2x2 contingency table is possible. The appropriate weighting pattern for the 4x4 contingency table can be described as follows: three of its cells (i.e. n_{13}, n_{14}, n_{24}) contain false positive errors and another three cells (i.e. n_{31}, n_{41}, n_{42}) contain false negative errors. Note that n_{23} and n_{32} would be classified as false in the classical approach but as true here due to the close match.

Having specified the values of this refined cross-classification as well as refined weighting pattern, we can calculate measures such as weighted Kappa, sensitivity, and specificity for evaluation of prediction accuracy. Moreover, by computing *sensitivity* and *specificity* as functions of *threshold*, other measures such as receiver operating characteristic (ROC), the area under the ROC curve (AUC), and maximum true skill statistic (TSS) can be calculated as usual.

(d) Main results

In summary, our new method for evaluation of prediction accuracy consists of the following steps: (1) incorporate additional autocorrelation into binary observation data until spatial autocorrelation in predictions and actuals is balanced, (2) cross-classify predictions and adjusted actuals in a 4x4 contingency table, (3) use a refined weighting pattern for errors, and (4) calculate weighted Kappa, sensitivity, specificity and subsequently ROC, AUC, TSS to get spatially corrected indices.

We provide all tools for calculating spatially corrected indices in our newly created package 'spind'. It is open-source software (published under the GPL public license, ver. 2), and is available as both a package spind_1.0.zip (windows version) and a source package spind.1.0-1.tar.gz. Both R packages, together with documentation, are available on GitHub (< https://github.com/carl55/spind >).

To illustrate the impact of our spatial method we present an example of simulated data as well as an example of presence/absence data of the plant species *Dianthus carthusianorum* across

Germany (Carl & Kühn 2016). Our analysis includes a statistic for the comparison of spatial and classical (non-spatial) indices. We find that our spatial indices tend to result in higher values than classical ones. These differences are statistically significant at medium and high autocorrelation levels.

(e) Conclusion & implications for considering uncertainties in biodiversity modelling

We conclude that these spatial accuracy measures may contribute to evaluate prediction errors in presence/absence models, especially in case of medium or high degree of similarity of adjacent data, i.e. aggregated (clumped) or continuous species distributions.

2.4. Projections

Reporting uncertainty in projections provides confidence in model results that supports decision-making in conservation-related recommendations and policies (Moilanen 2006). Uncertainties in the distributions of species render any decisions about where or how to implement conservation actions difficult, and may increase expenses whenever large areas must be managed (Hermoso & Kennard 2012).

Uncertainty in model projections can result from several of the so far mentioned sources (see Araujo & New 2007, Dormann et al. 2008b). Not all of them are necessarily unwanted and can help evaluating the projections if they are properly documented and communicated. For instance, if you consider differences in differently outlined scenarios of a distant future, such uncertainty is welcome, part of human society and inherent to management and conservation action (Real et al. 2010). Nevertheless, such scenarios are based on different assumptions and although all the scenarios are equally likely in theory, this does not hold true in reality. Additionally, model projections of input data of the same scenarios differ among models, e.g. different Global Circulation Models (GCM) (Goberville et al. 2015). Moreover, already within the Deliverable D4.2 ‘Report on projections of range and biodiversity changes using improved European data sources’ the EU BON partners stated that scenarios and therefore projections are suffering from substantial limitations concering the accounting of land-use information, the inclusion of ecological traits and the scaling of developed scenario frameworks – aspects that contribute to overall and conceptual uncertainty in projections of biodiversity models.

The following two case studies assess uncertainty in projections originating from models calibrated with data from (1) the Rhine-Main observatory (EU BON) test site – comparing different scenarios and taxonomic groups (**Case Study V**) and (2) the AquaMaps framework (**Case Study VI**).

Case Study V: *Uncertainty in distribution predictions of freshwater biodiversity under land-use and climate change scenarios in the RMO test site*

Partners involved: SGN (Matthias Kuemmerlen, Stefan Stoll, Peter Haase)

- For more details, see: Annex 7.5

1. Context

Freshwater biodiversity is particularly sensitive to changes in the environment. Their use as indicators of water quality can also be extended to additional, broader anthropogenic impairments such as those expected to be brought about by global environmental change. Previous studies have documented strong relationships between the occurrence of freshwater biota and environmental factors such as climate, hydrology and land use. These links have been exploited to set up species distribution models (SDMs) in the Rhine-Main Observatory (RMO), an EU BON test site. These models have been adapted for the particular conditions of freshwater ecosystems: they are limited to individual catchments; make predictions on the stream network and incorporate a wide range of environmental predictors that are highly relevant to these ecosystems (see more details in Deliverable 3.2). In a further step, these SDMs have been used to project the potential distribution of freshwater biota under future scenario conditions: one land-use and two climate scenarios (IPCC RCP 4.5 and 8.5), as well as the combination of each one of the climate forecasts with the land-use scenario. Three very different taxonomic groups were modelled: benthic macroinvertebrates, freshwater fish and aquatic macrophytes. Such a modelling framework is subject to uncertainty stemming from many potential sources. We assessed some of the sources of uncertainty for SDM projections under future scenarios.

2. Concept / Objective

We projected both land use and climate change scenarios to assess the uncertainty stemming from each. We also projected the combination of these scenarios to observe if there were any interactions. Finally, we analysed the different taxonomic groups independently to determine possible differences in uncertainty.

3. Data and methods

A SDM was set up for the catchment of the Kinzig River (RMO) at a spatial resolution of 25 m. Input data comprised occurrence data for the species modelled and environmental predictors of the following categories: topography, geology, climate, hydrology and land use. Two of the most recent IPCC climate scenarios were implemented, namely the RCP 4.5 and 8.5 scenarios, with data stemming from the EURO-CORDEX modelling initiative. The climate scenarios were also used to modify the hydrological predictor used in the model: river discharge. A land use scenario was also implemented with data stemming from a national model for Germany. Predictions were made for the year 2030. Coefficients of variation between projections were interpreted as a measure of uncertainty.

4. Main results

Scenario projection always implied a significant increase in uncertainty, compared to the equivalent present projection. However, no relationship between uncertainty and taxonomical group was found.

Uncertainty was more related to individual scenarios, than to either climate or land use scenarios, or the combination of both (**Figure 2.4**). This is probably due to the magnitude of the predicted change, specific to each scenario. We found evidence that choice of individual predictors can also carry significant consequences for model projections in terms of uncertainty.

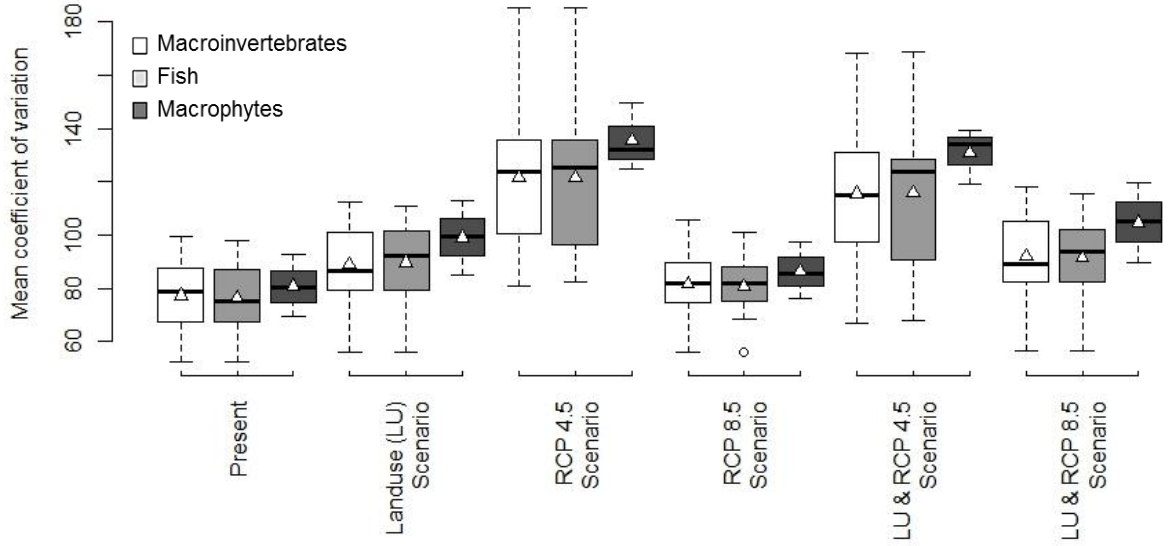


Figure 2.4: *Coefficients of variance for macroinvertebrates, fish and macrophytes for the present and four scenario projections*

e) Conclusion & Implications for considering uncertainties in biodiversity modelling

Uncertainty in future scenarios could be related to the magnitude of the forecasted changes in the environment. For example, the RCP 4.5 scenario introduced very high variability in the scenario values for the predictor temperature of the driest quarter, potentially influencing the higher uncertainty in this scenario projection. Thus, choice of predictors is also an important source of uncertainty. Complex predictors such as the so-called bioclim variables are of particular attention.

Case study VI: *Quantifying uncertainty in AquaMaps*

Partners involved: FIN (Rainer Froese, Cristina Garilao, Kathleen Kesner-Reyes)

+External Partners:

Department of Biometry and Environmental System Analysis, Albert-Ludwigs-University of Freiburg, Tennenbacher Strasse 4, 79106 Freiburg, Germany (Kristin Kaschner)
Instituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche (CNR), via Moruzzi 1, 56124 Pisa, Italy (Gianpaolo Coro)

- Annex unavailable because the manuscript is currently in preparation

a) Context

AquaMaps is a species distribution model, adapted from Kaschner et al. (2006), that generates model-based, large scale predictions of the natural occurrences of aquatic species. Like other ecological niche models, it uses estimates of environmental tolerances of a species with respect to environmental factors that can be used as predictors of species presence. In AquaMaps, these estimates of species presence (also called the environmental envelopes) are derived from point data and from independent knowledge about a species.

One of AquaMaps' main strengths is its ability to generate thousands of reliable species maps using minimum amount of information. However, other than expert reviews done on individual species maps, a close scrutiny into the overall uncertainty of AquaMaps map predictions and the sources of this uncertainty had not been investigated so far within EU BON.

b) Concept / Objective

To remedy this, in AquaMaps a species' known distribution (FAO area or bounding box limits) is used to verify the validity of occurrence records harvested from huge repositories like the Global Biodiversity Information Facility (GBIF; gbif.org). "Good point data" are then selected and assigned to a corresponding half-degree cell in a global grid of 259000 cells. These "good cells" are used to extract the range and frequency of environmental parameters within the species' native range and generate the species' environmental envelopes with respect to temperature, salinity, primary production, sea ice concentration and distance to land (Kesner-Reyes et al. 2016).

Confidence in AquaMaps species distribution therefore relies on (1) the goodness of the environmental envelopes which is a function of how many good cells contributed to its estimation, (2) the proportion of cells in the native range that are supported by observations and (3) expert reviews.

c) Data and Methods

Uncertainty for every mapped species in the AquaMaps for EU BON interface (i.e. all marine fishes and marine mammals predicted to occur in Europe) was obtained by computing for confidence for each species map. Confidence is based on the ratio of "good cells" to the total number of native cells with a probability of >0.5 (where 0.5 would cover a species' core habitat). As the probability of having observations for all cells decreases with the number of

cells, the equation for confidence is log(# good cells) / log(# cells > 0.5). Maps that are expert-reviewed are assigned a confidence value of 0.9. Uncertainty is then 1-confidence.

d) Main results

Figure 2.5 shows overall AquaMaps uncertainty for each half-degree cell in Europe, based on a scale from 0-1, representing the median uncertainty for all species predicted to occur in a given cell. The map nicely shows high-uncertainty areas within Europe where species maps are generally less "good" (uncertainty is higher compared to other areas) because of fewer "good cells" available to generate species environmental envelopes.

The map also shows known patterns of point data collection. The North Sea, Baltic and Black Seas are relatively shallow areas that have been heavily fished for centuries and are regularly visited by research vessels that report their catch to GBIF. The overall low uncertainty in these areas is the result of species having many "good cells" being available for species envelope generation. In contrast, offshore areas are relatively deep and populated by deep-sea and oceanic wide-ranging species that are caught or sampled mostly on the shelves. The higher uncertainty in these areas reflect the fewer number of "good cells" available for species envelope generation.

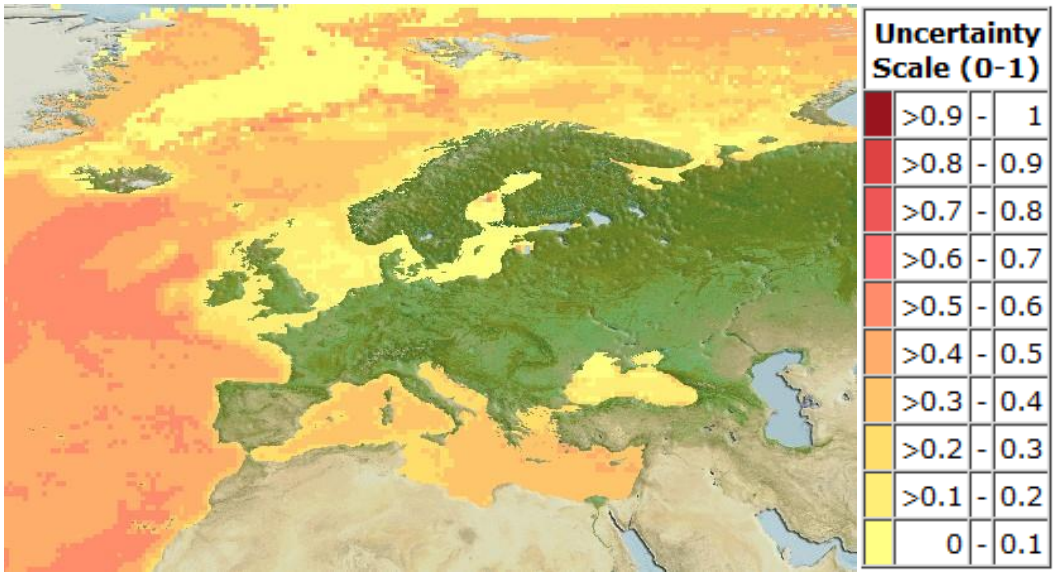


Figure 2.5 *Map showing overall AquaMaps uncertainty for each half-degree cell in Europe. Uncertainty ranges from 0-1. Yellow cells are areas where uncertainty is low (i.e., predicted species occurrence is generally good) due to the fact that many "good cells" were available to generate species environmental envelopes. These are typical of areas like the North Sea, Baltic and Black Seas that are shallow and have been sampled for many years. Orange and red cells are areas where uncertainty is higher (i.e., predicted species occurrence may not be as good) because of fewer "good cells" available to generate species environmental envelopes. These are typical of relatively deeper areas like the Mediterranean Sea and offshore areas.*

e) Conclusion & Implications for considering uncertainties in biodiversity modelling

The results of the effort to find a way to estimate uncertainty in AquaMaps is useful for the many users of AquaMaps. For the first time, each computer-generated map will be assigned a confidence/uncertainty value to qualify the goodness of the environmental envelopes generated and ultimately the goodness of the species predictions.

This exercise will also be expanded to produce a global map of uncertainty, including all taxonomic group to show general areas where more data or research is needed. Likewise, uncertainty maps can also be generated for certain species groups to answer similar questions.

3. Propagation of uncertainty and model complexity

Each of the sources of uncertainty contributes to the overall uncertainty of the entire modelling process. Therefore, whenever one takes the iterative procedure from data (collection), to modelling, measuring its accuracy and projection, uncertainty surrounding each of the corresponding techniques and approaches propagates through the whole process. This is generally true for an increase in complexity of modelling frameworks, as each parameter that has to be parameterised adds its own uncertainty to the model outcome.

Coming back to the Introductory chapter (→ **General Introduction**), Gerd Gigerenzer (Director of the Center for Adaptive Behavior and Cognition (ABC) at the Max Planck Institute for Human Development) introduces the concept of "ecological rationality", i.e. knowing which (simple) heuristics will work in which environments (Ramnath 2017). The basis for this thought is that if dealing with risk, complex mathematical models will work for optimisation. But if dealing with uncertainty, they potentially will not work as well, as there environment is dynamic. Under such conditions, more complex models might be outperformed by more simple ones, as the former accumulate a higher degree of uncertainty. Nevertheless, Gigerenzer does not argue for neglecting complex models, but warns against the automatic use of complex models for description of complex relationships, as complex models are built upon a larger number of estimations and therefore introduce additional uncertainty (Ramnath 2017). The concept of 'ecological rationality' therefore suggests having complex models when you have good knowledge about your system, few alternative states or scenarios and a high amount of data. If dealing with plenty of alternative scenarios, a small amount of data and high uncertainty regarding your system in general, more simple models or even rule of thumbs should be preferred, as they are robust and provide you with an adaptive toolbox.

Concerning approaches in biodiversity modelling, we can distinguish between two modes of accounting for uncertainty propagation: (1) in a series of different non-hierarchical models and (2) in (Bayesian) hierarchical models.

(1) It is possible to couple (already) described SDMs with e.g. climate projections (i.e. using scenario data from Global Circulation Models and Regional Circulation Models) in traditional frequentist approaches, but they can only assess 'variances' of different metrics. As described earlier, SDMs have associated uncertainties related to data, calibration and validation processes and climate projections that are model outputs themselves suffer from the same uncertainties. It is a so far unresolved issue of how to propagate uncertainty from one model (or module or sequence) to another one, for instance propagating climate projection uncertainties to ecological projections originating from SDMs.

(2) A hierarchical model usually consists of several modules (sub-models) that integrate different processes (e.g. a SDM module, a population dynamics module, a dispersal module, a projection module). In opposition to the more traditional frequentist methods, hierarchical Bayesian frameworks can be used to propagate uncertainties, to incorporate knowledge about uncertainties directly into the model assessment and help defining on how to combine the different sources of uncertainty in a final output.

The following case study contributed by FEM touches more issues discussed within this Deliverable than error propagation (**Case Study VII**). But by using a hierarchical Bayesian framework, FEM accounts for sampling effort (see **2.1.1 Biological response**) within the modelling framework and finally both visualises it (see **5. Outlook: Communication and visualisation of uncertainty**) via cartograms and directly incorporates existing knowledge about this source of uncertainty. Overall, this case study gives a good example of integration of uncertainty into the biodiversity modelling process.

Case study VII: *Anticipating species distributions: Handling sampling effort bias under a Bayesian framework*

Partners involved: FEM (Duccio Rocchini, Carol X. Garzon-Lopez, Matteo Marcantonio, Heidi C. Hauffe, Carolo Ricotta, Annapaola Rizzoli; Roberto Rosà), NHM (Neil A. Brummitt)

External partners:

Department of Life Sciences, University of Siena, Via P.A. Mattioli 4, Siena 53100, Italy (Valerio Amici)
Department of Life Sciences, University of Trieste, Via L. Giorgieri 10, Trieste 34127, Italy (Giovanni Bacaro)
School of Computer Science, Aston University, UK (Lucy Bastin)
BIGEA, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum, University of Bologna, Via Irnerio 42, Bologna 40126, Italy (Alessandro Chiarucci)
University of Nottinhgam, University Park, Nottingham NG7 2RD, UK (Giles M. Foody)
Department of Biological Sciences, Murray State University, Murray, KY 42071, USA (Kate S. He)

- For more details, see: Annex 7.7
- For the original paper, see: Rocchini et al. 2017. Anticipating species distributions: Handling sampling effort bias under a Bayesian framework, *Science of the Total Environment* 72: 234-238

a) Context

FEM and NHM dealt with the uncertainty related to sampling effort when anticipating species distributions, considering two main issues, such as: sampling effort uncertainty mapping and sampling effort uncertainty incorporation into the modelling procedure. Anticipation is an important topic in ecological fields such as community ecology, species distribution

modeling, landscape ecology, and biological invasion science. Anticipatory methods are also crucial for developing effective management practices to deal with invasive species.

Invasive species can modify the structure and functioning of ecosystems, altering biotic interactions and homogenizing previously diverse plant and animal communities over large spatial scales, ultimately resulting in a loss of genetic, species and ecosystem diversity. The annual economic impact of invasive species has been estimated at over 100 billion dollars just within the USA, an order of magnitude higher than those caused by all natural disasters put together. Given the massive negative economic and ecological effects of invasive species, a robust method for predicting species' distributions is crucial for an early assessment of species invasions and effective application of appropriate management actions malanson.

Investigating how biodiversity is distributed spatially and temporally across the globe has long been a central theme in ecology and the methods developed to answer this question have become key tools for biodiversity monitoring. For example, species distribution models (SDMs) have been used to map the current distribution of a single species, model the potential distribution of native and invasive species.

In combination with remote sensing products and current global data sets, SDMs have become the method of choice for monitoring biodiversity at multiple spatial and temporal scales. However, the strength of this combination depends on the careful selection and application of integrative modeling approaches, in combination with a thorough assessment of uncertainty in both data inputs and modeling methods.

b) Concept and objective

The aim of this part of the Deliverable was to propose coherent and straightforward methods to explicitly account for uncertainty when mapping species distributions in the light of anticipating the spread of invasive species. In particular we will cover: i) explicitly mapping uncertainty in sampling bias, ii) mitigating uncertainty in data through prior beliefs and Bayesian inference and iii) reporting uncertainty in species distribution maps through Markov Chain Monte Carlo methods. The findings of this manuscript should be of particular interest to landscape managers and planners attempting to predict the spread of species and deal with errors in species distribution maps in a straightforward manner.

One of the main problems with field data on species distributions is related to "sampling effort bias', namely the bias inherent in some areas being under-sampled with respect to others. Quantifying and mapping the uncertainty derived from variation in the number of observations due to sampling effort can be achieved using cartograms, in which the shape of spatial objects (e.g. polygons, cells, etc.) is directly related to a determined property, in our case to uncertainty.

As an example, we show a cartogram of the distribution of *Abies alba* overlapping a grid to the set of records obtained from the Global Biodiversity Information Facility (GBIF, http://www.gbif.org, **Figure 3.1**)

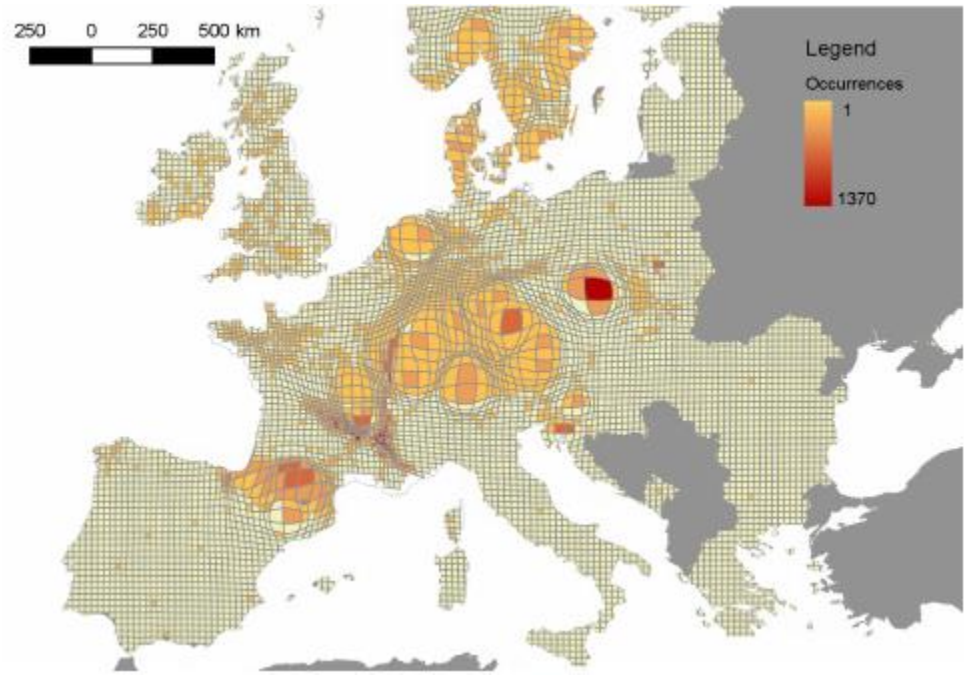


Figure 3.1: *Cartogram representing the sampling effort bias (cell distortion) of the GBIF dataset related to Abies alba. This species is not native in Northern Europe, although it is widely cultivated as a timber tree, as thus present in the GBIF dataset.*

c) Data and methods

We built a multi-level model to take into account the different resolution of the predictor variables (**Figure 3.2**) and the differential sampling effort of *Abies alba* occurrences in each NUTS3 polygon. The sampling effort was used to re-scale the precision of the likelihood at pixel level, multiplying the scaled sampling effort by the standard deviation of the Gaussian likelihood. As a result, the likelihood estimate of pixels in regions with a higher number of samples was expected to be more precise. The theoretical model (**Figure 3.2**) was coded in JAGS language and run in JAGS 4.2.0 through R (R Core Team, 2016) using the R2jags and CODA packages.

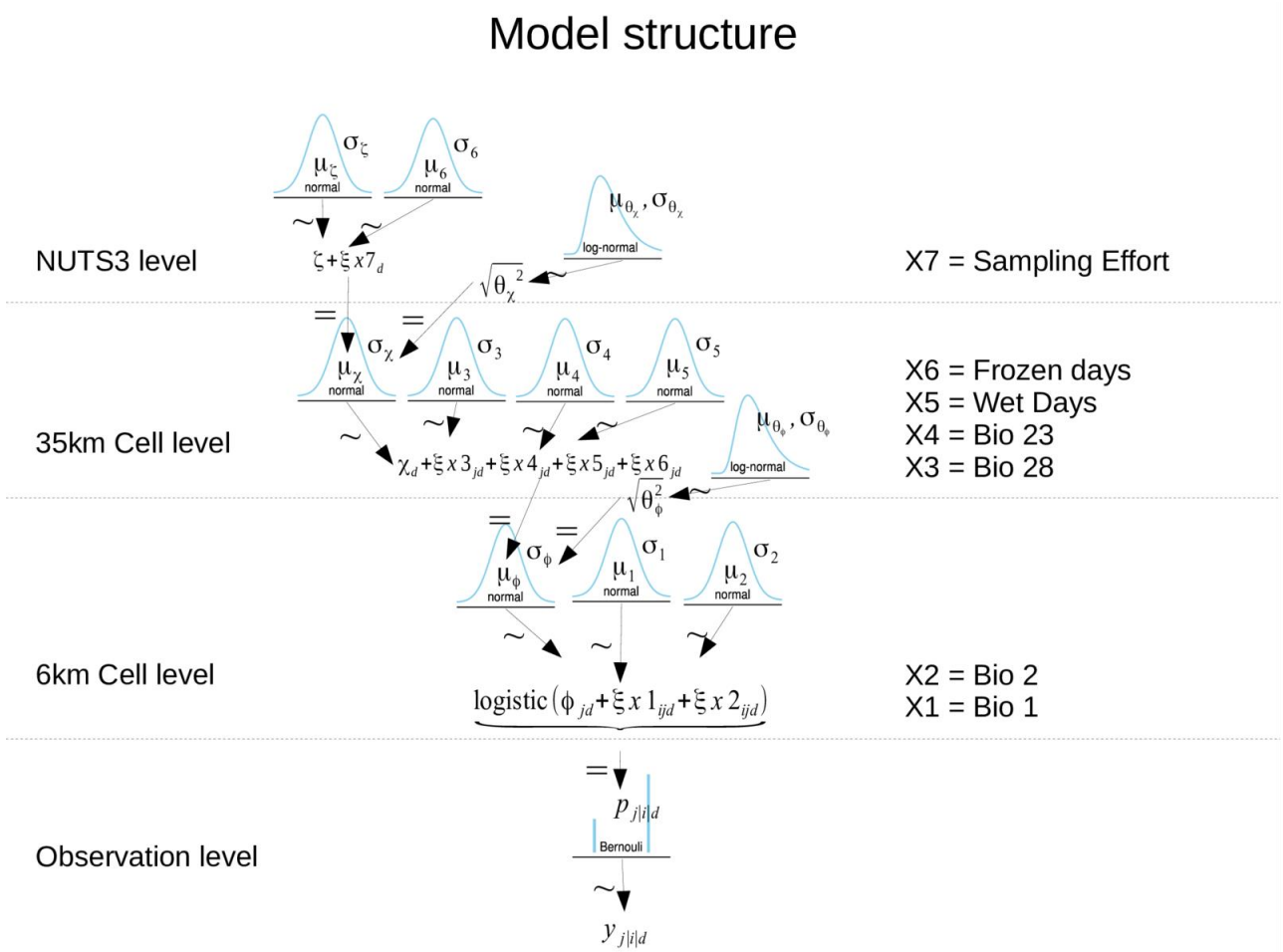


Figure 3.2 *Multi-level model represented through a pictogram. To select the predictor variables, we performed a literature review on the ecology of the species by finally selecting radiation seasonality (Bio23), the annual mean moisture index (Bio28), the number of wet days during summer and the frost days during winter and early spring, the annual mean temperature (Bio1), the mean diurnal temperature range (Bio2). Sampling effort was calculated as the diversity of dates of survey recorded in the GBIF dataset per each NUTS3 country. Symbols used in this figure: μ, σ = mean and standard deviation of prior and hyperprior distributions; ζ, χ, φ = intercepts for NUTS3, 35km, 6km level of the model; subscript d,j,i,o = index for NUTS3, 35km, 6km and observation level; weightijd = scaled weights for sampling effort; logistic(ψ) = logistic transformation of the model output (link function); pi|j|d = probability of occurrence; yo|i|j|d = presence or absence.*

d) Main results

The Posterior Probability Distributions (PPDs) of model parameters for the three models (with different priors on sampling effort) are reported in **Figure 3.3**. All the models agreed on the direction and effect size of the predictors (**Figure 3.3**). Credible effects (no intersection with 0 in **Figure 3.3**) were attained for those variables directly related to temperature. In particular, annual mean temperature (Bio1 and Bio12) and radiation seasonality (Bio23) showed negative effects while mean diurnal temperature range (Bio2) showed positive effects (**Figures 3.3 and 3.4**). The negative credible effect of Bio12 implies that the relationship between the probability of presence (suitability) of *Abies alba* and annual mean temperature has a "bell shape", by rising slowly to the left of the annual mean temperature average (7.8° C) and decreasing rapidly when on its right (**Figure 3.4**). On the contrary, the distribution of wet days, annual mean moisture index (Bio28) and frost days included 0, showing a non-credible effect on the presence of *Abies alba*.

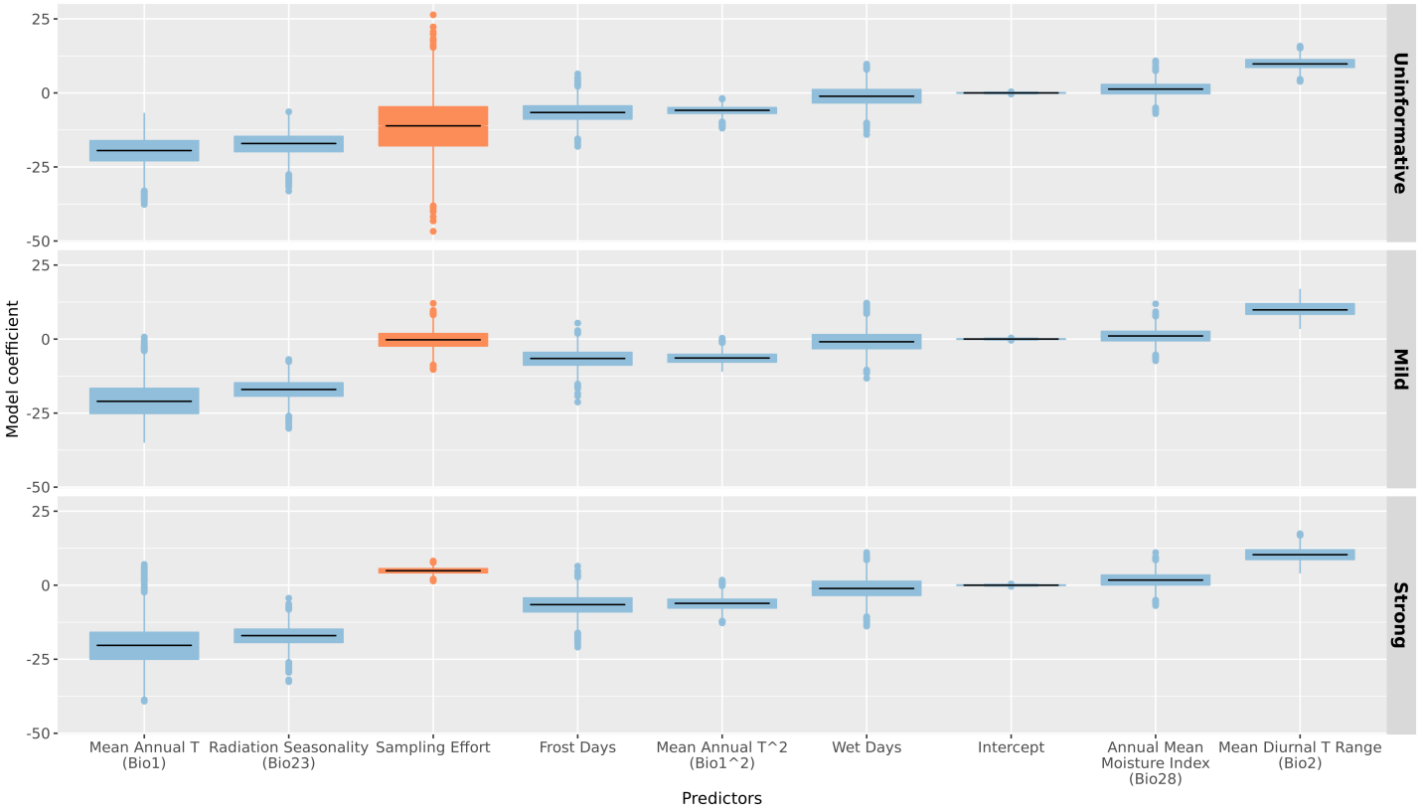


Figure 3.3: *Boxplots of the β coefficient PPDs for the three models. Each box represents the 1st and 3rd quartiles of a coefficient distribution, the black horizontal line the distribution median, the whiskers the limits of the 1.5*interquartile range, while the filled circles represent the outlying points. If whiskers did not overlap 0 we inferred as "credible effect". We showed in red the boxplots reportingthe distribution of the β coefficient of the sampling effort.*

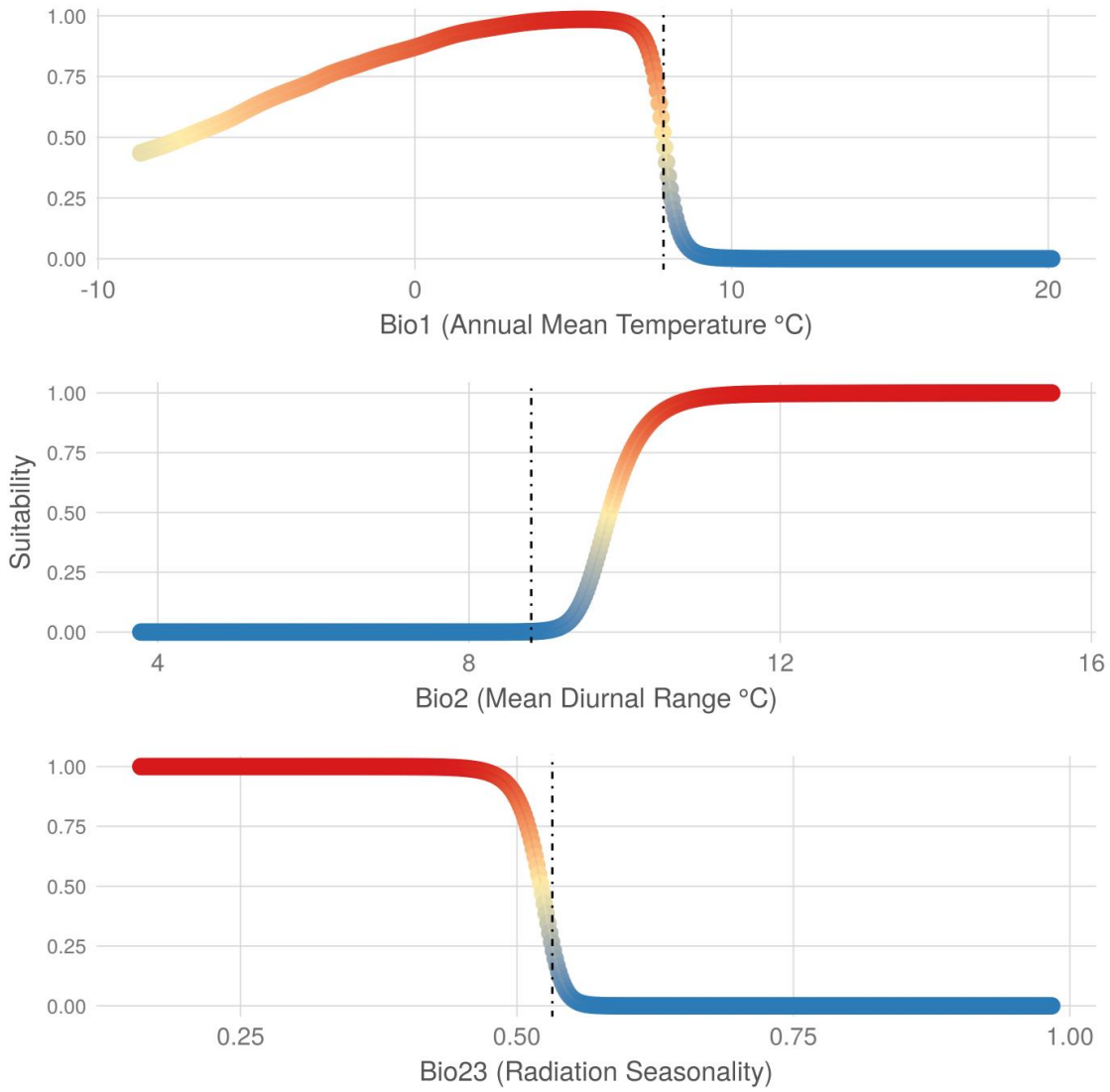


Figure 3.4: *In this figure the average probability of presence (suitability) of Abies alba is plotted against the three variables with the highest average coefficient effect size in the model (top: range of annual mean temperature Bio1, middle: mean diurnal range Bio2, bottom: Radiation Seasonality or Bio23). The relationship between the probability of presence (suitability) of Abies alba and annual mean temperature has a "bell shape", rising slowly moving from the left of the study area average (7.8 °C), peaking just before the average and decreasing rapidly when on its right. The shape of the relationship between the probability of presence and the mean diurnal temperature range is inverted. A low diurnal temperature range is associated with a low suitability while a wide temperature variability is associated with high suitability. The highest suitability is reported for Bio2 values higher than 11 °C. The Radiation Seasonality (the standard deviation of the weekly solar radiation estimates expressed as a percentage of the mean of those estimates) shows a negative pattern with respect to suitability. Areas with a very high average difference in solar radiation during the year (i.e. Northern Europe) are reported as weakly suitable for Abies alba. All the curves were obtained varying the value and the model coefficient of Bio1, Bio2 and Bio23 while keeping the values of the other predictors at their average.*

e) Conclusion & Implications for considering uncertainties in biodiversity modelling

In summary, the model with the strong prior showed an improved precision of sampling effort, basically maintaining that of the others (**Figure 3.3**). Based on this and since the DIC did not show differences for the strong prior model with respect to the uninformative prior-model (with a DDIC ≤ 4), we further focused on the model with a strong prior to build the output distribution map. The resulting potential niche distribution of *Abies alba* is thus shown in **Figure 3.5**.

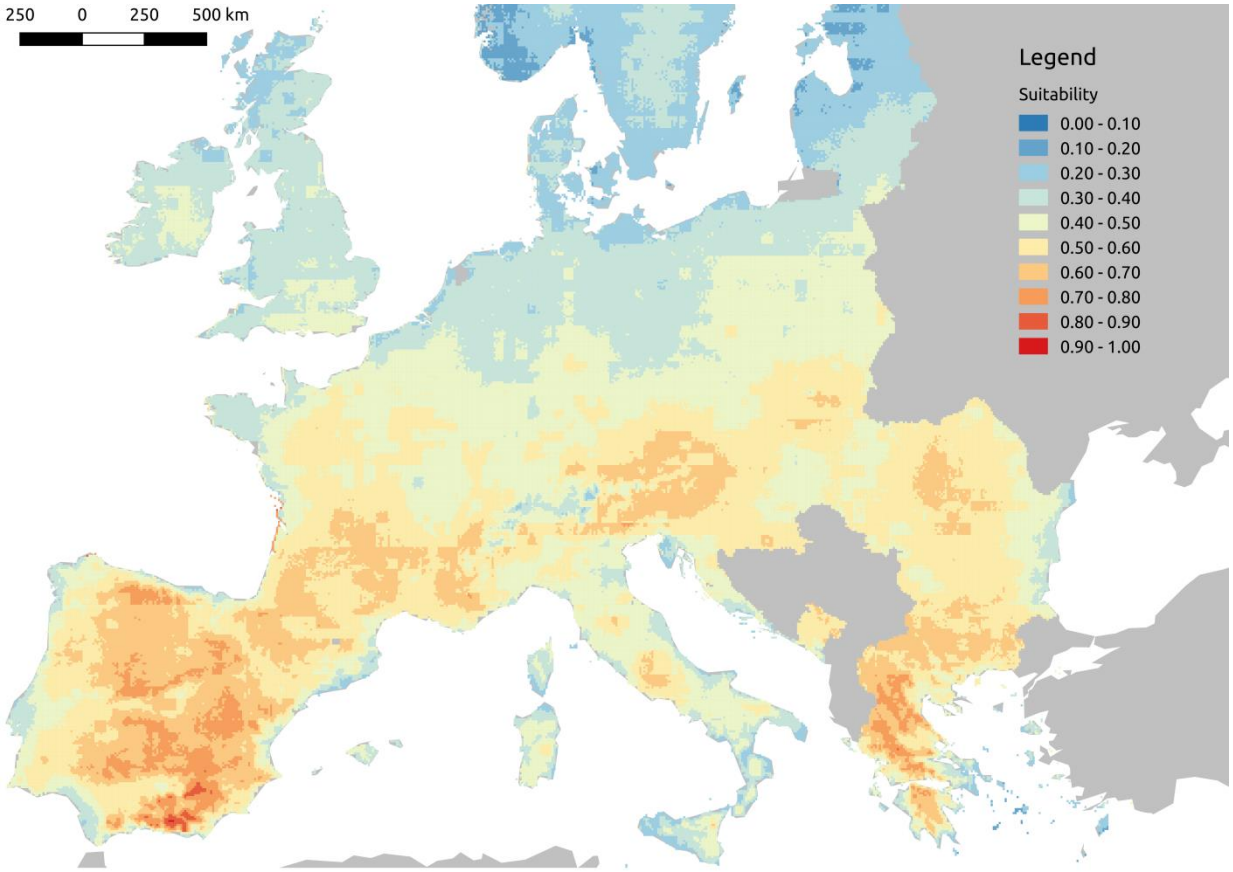


Figure 3.5: *Abies alba suitability distribution as derived from the multi-level model with strong prior on sampling effort. The pixel value is the average of the PPDs for that pixel.*

4. Prioritisation of sources of uncertainty

Embracing and communicating uncertainties alongside the biodiversity/species distribution modelling process is not possible without prioritisation of the different sources of uncertainty. In turn, this needs understanding of the quantification of uncertainty in such models. Buisson et al. (2010) demonstrated that the uncertainty stemming from data and selected modelling frameworks is relatively small compared to the uncertainty stemming from scenarios used to project these models to the future. This aligns with the amount of literature dealing with each of the sources of uncertainties. Whereas uncertainty surrounding data (biological response and environmental predictors), model calibration and model validation is in general widely covered and understood, quantification and capture of model uncertainty stemming from projections and scenario assumptions as well as limitations are underrepresented.

Quantification

Quantification of uncertainties alongside the modelling process is an elaborative, but yet inevitable task, if the ultimate goal is to inform the scientific community, stakeholders and the general public. At the moment, quantification often means assessing the importance of each of the sources of uncertainty, by having several "treatments" for each group as shown above (e.g. Uusitalo et al., 2007, Dormann et al., 2008; Goberville et al., 2015), i.e. different environmental input as well as response variables (e.g. based on different scales or sampling schemes), different model types and variable selection approaches, different scenarios, and different accuracy measures. This includes approaches of sensitivity analyses, scenario ranking and partitioning of uncertainty. Nevertheless, quantification of the accuracy of future projections is in the majority of cases not possible due to the lack of data. Therefore, qualitative as well as quantitative criteria have to exist next to each other to assess and prioritise the importance of each of the sources of uncertainty.

The following case study contributed by UnivLeeds is an example at the forefront of how to quantify uncertainty in habitat/land-cover classification models or SDMs in a general way in future assessments (**Case Study VIII**).

Case study VIII: *Assessing the transferability of habitat/land-cover classification models in space and time*

Partners involved: UnivLeeds (Yoni Gavish, Tim Benton)

+ External Partners:

School of Biosystems and Food Engineering, University College Dublin, Dublin, D04 N2E5, Ireland (Jerome O'Connell)

- Annex unavailable because the manuscript is currently under review.

a) Context

In many cases, data collected at a certain place/time is used to train habitat/land-cover (H/LC) classification models or species distribution models (SDMs) that predict at another place/time. Such extrapolation usually carries a cost of reduced model performance. Quantifying this decrease in model performance with distance may allow predicting the uncertainty associated with transferring models in space and time.

b) Concept / Objective

Providing reliable habitat or land-cover maps at fine thematic and spatial resolutions over wide extents is crucial for various conservation related issues. With the increasing availability of open-access satellite images (Kerr & Ostrovsky 2003; Xie et al. 2008) and the development of advanced machine-learning algorithms that produce reliable rule-sets (Bradter et al. 2011; Rodriguez-Galiano et al. 2012), the limiting factor in producing H/LC maps is the ground-truth data required to train the models. Similarly, in SDMs the availability of occurrence records is the main factor limiting the production of high quality species distribution maps. Therefore, to optimize the usage of the rarely available ground-truth data, it is often required to use a model trained in a local site to predict in other sites. However, various studies of H/LC classification (Pax-Lenney et al. 2001; Olthof et al. 2005; Knorn et al. 2009;

Rokitnicki-Wojcik et al. 2011; Brenning 2012) and SDMs (Randin et al. 2006; Tuanmu et al. 2011; Heikkinen et al. 2012; Wenger & Olden 2012; Wang & Jackson 2014; Duque-Lazo et al. 2016; Huang & Frimpong 2016; Huang et al. 2016) have shown that extrapolation in space and time carries a cost in model performance.

Interestingly, in the SDM literature model transferability was explored mainly as a function of model complexity and not as a quantitative decay curve against the distance between the source site (where the model is trained) and target site (where the model is predicting). In the H/LC classification literature, model transferability was explored in relation to distance, however, no one have tried using advanced statistical methods to explore the shape of the decay curve. In addition, in neither of the two fields, none have tried to use the performance decay curves to predict the performance a model trained in a known source sites will have in a novel target sites or vice versa (novel source site and known target site). The ability to predict the decay in performance links directly to model uncertainty, as it may allow associating uncertainty values (i.e. performance indices such as kappa or true-skills-statistics) to models projected in time or space. As far as we know, we present here the first attempt to do so.

c) Data and Methods

A manuscript summarising this analysis is currently under review, such that this deliverable will only contain a brief summary of the methods and only a subsection of the results. In short, we have used ground truth data for ten sites in the UK to explore the transferability of H/LC classification models with spatial, temporal, spectral and environmental distance. The classification scheme included 9 classes, representing crop and non-cropped land-covers in agricultural landscapes. After a segmentation procedure, we have created a total of 66 spectral variables and 31 environmental variables for each object in each of the ten sites. These variables were used to train a local classification model in each site, using the randomForest algorithm (Breiman 2001). Then, the local model trained in each site was used to predict the H/LC in all other sites. We assessed model performance using the kappa statistic. At the end of this stage we had the observed kappa for 100 pairs of sites, including 90 pairs in which the source and target sites differed and 10 additional pairs where they were identical.

Next, we used various techniques to estimate 1 spatial, 2 temporal, 27 spectral and 38 environmental distance measures between each pair of sites. This distance measures where used as explanatory variables in generalized dissimilarity models (GDM) with the observed transferred kappa as the dependent variable. GDMs are statistical models developed specifically to predict the decay of community similarity with various distance measures, using I-splines (Ferrier et al. 2007). We have fitted a separate GDM model for each site, by excluding it from the list of sites pairs. We then plugged the distances between the left-aside site and each of the other sites into the GDM output, and gained predicted kappa values. These kappa values allowed exploring two complimentary questions: 1. Can we predict the performance of a novel source site in a known target site? 2. Can we predict the performance of a known source site in a novel target site? We also explored the relative importance and the shape of the performance decay curves with spatial, temporal, spectral and environmental distances.

d) Main results

We have found performance to decay with various spatial, temporal, spectral and environmental distances between source and target sites (**Figure 4.1**). We also found that the expected performance of novel sites in target sites (and vice versa) can be predicted with the GDMs (**Figure 4.2**)

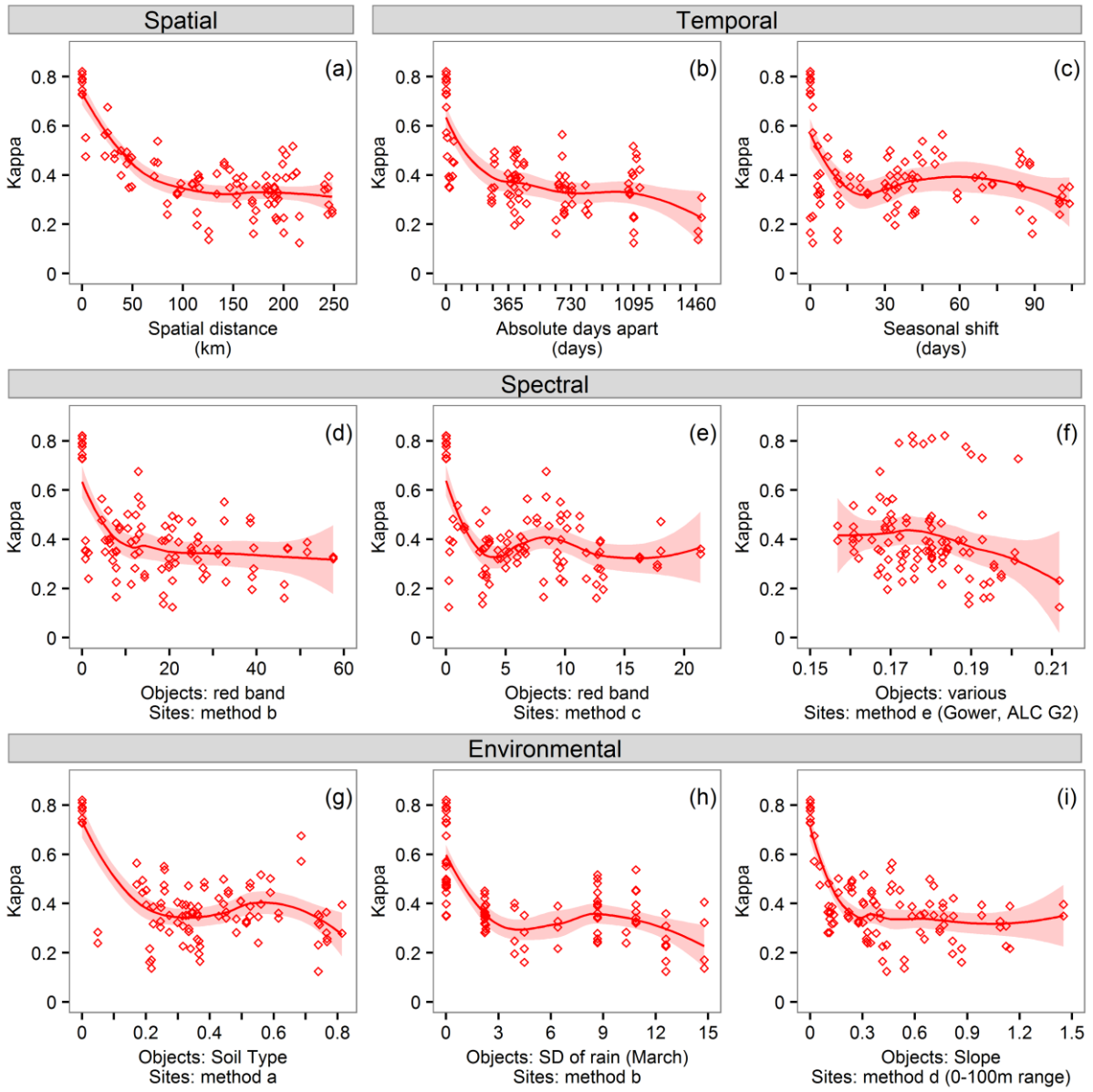


Figure 4.1: *Examples of the change of observed kappa with (a) spatial, (b, c) temporal, (d, e, f), spectral and (g, h, i) environmental distances between sites. The solid line is a loess curve (±confidence intervals). For the spectral and environmental distance, the Objects line in the axis title describes the object level variable on which the distance measure is based. Methods a-e in the 'Sites' line of the axis title refers to various ways by which the object level variables where aggregated to create distance measures at the sites scale. ALC G2, ALC G4 are the agricultural land classification grades 2 and 4, respectively.*

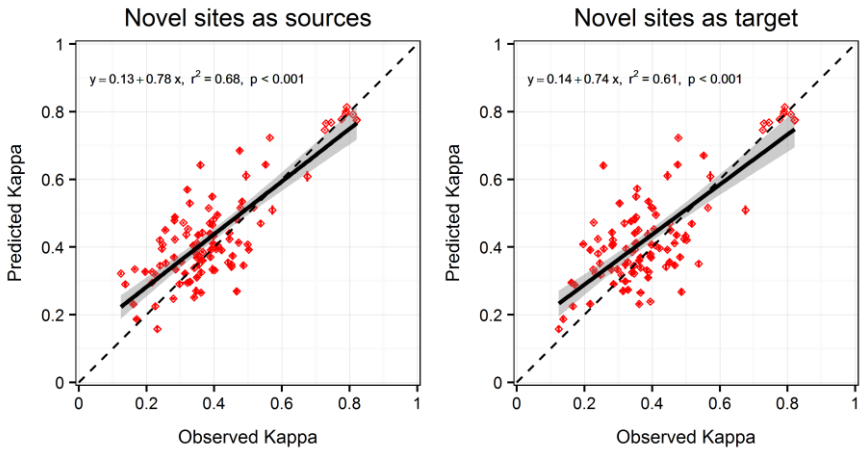


Figure 4.2: *The predicted vs. observed kappa when novel sites are source and when novel sites are targets. Results of linear regression model are given within the plot (solid line ± confidence intervals). Dashed line is the line of unity.*

e) Conclusion & Implications for considering uncertainties in biodiversity modelling

One of the main outcomes of this study is that the model transferability to or from a novel site is predictable (**Figure 4.2**). These are encouraging results that should be taken with a grain of salt, since we also found considerable decrease in performance with distance. For example, we found kappa to decrease from 0.8 to around 0.4 once the spatial distance is larger than around 100 km (**Figure 4.1**). None-the-less, we can see several potential usages for the performance decay curves in conservation and management of habitats and species:

1. *Mapping uncertainty:* we can predict the level of accuracy that each known site will have in any novel site. Therefore, we can map classification uncertainty and identify areas for which we currently cannot provide a reliable classification model. Temporaly, we can predict the performance of a model projected into the future, providing uncertainty estimates for various future scenarios.
2. *Optimize site selection:* similarly, we can predict the performance any novel site will have in any other site and we can use this information to optimize site selection for ground truthing. That is, we can select sites that are expected to provide models that can be transferred with the required accuracy to other sites that currently we cannot predict in.
3. *Averaging of local models:* the predicted performance of known sites in a novel site may be used as weights when averaging the prediction of the different local models, thereby giving higher weights to sites that are likely to provide more accurate classification model in the novel site.
4. *Comparing classification algorithms:* much on the work in the SDM literature have focused on the trade-off between model generality and model transferability, with a notion that the more complex algorithms provide higher local performance but lower transferability than the simpler algorithms. Our work puts this trade-off on a continuum, in which a more complex algorithms start at higher performance but decays first, while a simpler algorithms start with a lower performance but will decay slower. If the two lines intersect, we can identify range of distances in which one algorithm outperforms the other. Alternatively, we can use the predicted performance from the decay curves to weight the algorithms when averaging their outputs.
5. *Variable selection:* Most emphasis in selecting variables is traditionally put on identifying those that are important for increasing local performance. However, as far as we know, no

one explored the effect of including or excluding certain variables on model transferability. Comparing the decay curves of two sets of models, one that include a certain variable and one that do not, may allow identifying variables that may have little effect on model performance, but carry considerable influence on model transferability. Alternatively, we have found here that some variables are not necessarily important in the classification, but different between sites in their values are important for predicting transferability.

5. Outlook: Communication and visualisation of uncertainty

Science-policy audiences are highly diverse and often receive information that differs in both quantity and quality compared to what science typically provides (McInerny 2013). Strategies of communication, i.e. which information are inevitable to the audience and how can they be best communicated, change with the addressed audience. Therefore, the communication of uncertainties alongside the modelling process to different stakeholders and decision-makers is of special importance (Maxim & van der Sluijs 2011, Polasky et al. 2011, Knights 2014). Lastly, the general public has to be kept in mind. Many conservation actions nowadays rely on SDMs or other biodiversity models and if scientists are perceived as overstating problems or downplaying uncertainties, the public is likely to lose confidence in the scientific community (Beale & Lennon 2012).

Overall, identifying and communicating uncertainty is crucial as soon as knowledge has to be transferred, translated and shared (Pe'er et al. 2014). The identification of gaps, imprecision, inaccuracy or any kind of uncertainty is the starting point for improving knowledge accuracy and methodological efficiency. Besides, acknowlegding uncertainty may significantly improve the quality of information delivered to stakeholders (Jenouvrier 2013; Augusiak et al. 2014), and can help assessing different management options. Alternatively, neglecting communicating substantial aspects of uncertainty may lead to the oversimplification of scientific outcomes and weaken their interpretations, ultimately biasing management decisions (Polasky et al. 2011). Promising frameworks (e.g. process-based models, (Evans et al. 2013a); and approaches, e.g. risk management analysis, worst-case analysis, decision making theory, scenario or resilience thinking; (Polasky et al. 2011)) might be particularly relevant to deal with a broad range of uncertainty sources and to aid policymakers and practitioners for defining strategies of relevance to society (Polasky et al. 2011). Better identifying and communicating about uncertainty sources could allow faster and more adapted reactions to changing conditions. For instance, incorporating extreme scenarios in routine forecasting exercises or adopting a resilience thinking approach can help to quickly adopt new reaction paradigm if the old one becomes untenable (Carpenter & Brock 2006; Polasky et al. 2011). Consistently reporting uncertainty beyond the scientific field, e.g. in producing explicit uncertainty assessment of projections as transparent communication tools to stakeholders, would both manifest scientists' credibility and keep stakeholders aware of the evolving nature of knowledge to reinforcing the adaptive potential along the chain of decision making.

Visualisation of uncertainty

The power of visualisation as a direct way to transport scientifically derived information to a diverse audience has been largely underexplored in biodiversity research in general (Spiegelhalter et al. 2011, McInerny 2013), but should be "among the main priorities for developing modern science and science policy" (McInerny 2013). Visualisation of information is crucial to permeate scientific work and results as well as to inform science-

policy interfaces. As it is true for communication of uncertainties in general, different audiences, spanning from the general public to experts, may also require different visualising features (McInerny 2013).

Concerning biodiversity inventories and the depiction of species distribution ranges, maps are probably the most intuitive and widely-used tool of visualisation (Wenger et al. 2013), as scientists, stakeholders and decision-makers, as well as the general public are used to exposure to geographical/topographical maps. Nevertheless, the usage of maps to illustrate model-derived projections or scenarios comprises the danger of imparting false accuracy, as geographical/topographical maps are generally seen as fixed and certain (Wenger et al. 2013). Therefore, maps for this kind of purpose have to be easy-to read and dense in information at the same time, as otherwise there is potential for misinterpretation and/or miscommunication, which finally might lead to biases in the perception (Grilli et al. 2013, Wenger et al. 2013). Visualising complex data sets in a 2D-manner, as is common for most maps, is difficult and depicting measures of uncertainty is no exception from this. One way to circumvent this problem would be the juxtaposition of several maps next to each other (e.g. one of the raw data, one with projections, one with uncertainty surrounding this projection), but this may lead to the separation of the parts and again lead to miscommunication. This is related to the actual model-derived values in the 2D-space of a map, often depicted in the underlying point- or grid cell system. Contrastingly, those point or grid records could be altered in another way, to include another layer of information to the visualised data. Cartograms are one possibility, in which the size of specific data points is distorted, e.g. according to the underlying sampling effort of each raw data location (Rocchini et al. 2017, → **Case study VII**).

Another argument for the common use of fixed maps is their relatively small production effort. Another possibility to illustrate biodiversity assessments and to include information on the different sources of uncertainty, which also accounts for different background knowledge, political orientation and cultural background of users, are interactive maps or other interactive visualisation forms (McInerny et al. 2013). This is another field of scientific visualisation and communication that is largely underexplored in biodiversity research, although it would make a big impact on the communication of scientific results and especially the surrounding uncertainties. To overcome this lack of interest and/or awareness, it will need institutional support of structures to incorporate communication science into biodiversity research.

6. References

Addison, P. F. E., Rumpff, L., Bau, S. S., Carey, J. M., Chee, Y. E., Jarrad, F. C., et al. (2013). Pratical solutions for making models indispensable in conservation decision-making. Diversity and Distributions, 19, 490-502.

Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C., et al. (2011). Conservation and the botanist effect. Biological Conservation, 144, 131-140.

Albert, C. H., Yoccoz, N. G., Edwards, T. C., Graham, C. H., Zimmermann, N. E. & Thuiller, Wi. (2010). Sampling in ecology and evolution - bridging the gap between theory and practice. Ecography, 33, 1028-1037.

Allouche, O., Tsoar, A., Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43, 1223–1232.

Amano, T. & Sutherland, W. J. (2013). Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security. Proceedings of the Royal Society B-biological Sciences, 280, 20122649.

Araujo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends In Ecology & Evolution, 22, 42-47.

Augusiak, J., van den Brink, P. J. & Grimm, V. (2014). Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. Ecological Modelling, 280, 117-128.

Banner, K. M., & Higgs, M. D. (2017). Considerations for assessing model averaging of regression coefficients. Ecological Applications, 27, 78-93.

Barnosky, A. D., Hadly, E. A., Gonzalez, P., Head, J., Polly, P. D., Lawing, A. M., et al. (2017). Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. Science, 355.

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., et al. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling, 222, 1810-1819.

Beale, C. M., & Lennon, J. J. (2012). Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367, 247-258.

Bean, W. T., Stafford, R., & Brashares, J. S. (2012). The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography, 35, 250-258.

Bell, S., Marzano, M., Cent, J., Kobierska, H., Podjed, D., Vandzinskaite, D., et al. (2008). What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. Biodiversity and Conservation, 17, 3443-3454.

Bierman, S. M., Butler, A., Marion, G., & Kühn, I. (2010). Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. Ecography, 33, 451-460.

Boakes, E. H., McGowan, P. J., Fuller, R. A., Ding, C. Q., Clark, N. E., O'Connor, K., et al. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. Plos Biology, 8, e1000385.

Botta-Dukat, Z. (2005). Rao's quadratic entropy as a measure of functional diversity based on multiple traits. Journal of Vegetation Science, 16, 533-540.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecology Letters, 15, 584-593.

Bradter, U., Thom, T. J., Altringham, J. D., Kunin, W. E., & Benton, T. G. (2011). Prediction of National Vegetation Classification communities in the British uplands using

environmental data at multiple spatial scales, aerial images and the classifier random forest. Journal of Applied Ecology, 48, 1057-1065.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013). Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. Ecography, 36(9), 971-983.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package 'sperrorest'. IEEE International Symposium on Geoscience and Remote Sensing IGARSS.

Brummitt, N., Regan, E. C., Weatherdon, L. V., Martin, C. S., Geijzendorffer, I. R., Rocchini, D., et al. (2016). Taking stock of nature: Essential biodiversity variables explained . Biological Conservation, https://doi.org/10.1016/j.biocon.2016.09.006

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. Global Change Biology, 16, 1145-1157.

Burnham, K. P. (2002). Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Cade, B. S. (2015). Model averaging and muddled multimodel inferences. Ecology, 96, 2370-2382.

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. Global Ecology and Biogeography, 23, 99-112.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). Biodiversity loss and its impact on humanity. Nature, 486, 59-67.

Carl, G., & Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. Ecological Modelling, 207, 159-170.

Carl, G., & Kühn, I. (2010). A Wavelet-Based Extension of Generalized Linear Models to Remove the Effect of Spatial Autocorrelation. Geographical Analysis, 42, 323-337.

Carl, G., Doktor D., Schweiger O, & Kühn, I. (2016). Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis. Journal of Biogeography, 43: 2502-2512.

Carl, G., & Kühn, I. (2008). Analyzing spatial ecological data using linear regression and wavelet analysis. Stochastic Environmental Research and Risk Assessment, 22, 315-324.

Carl, G., & Kühn, I. (2017). Spind: a package for computing spatially corrected accuracy measures. Ecography, 40, 675-682.

Carpenter, S. R., & Brock, W. A. (2006). Rising variance: a leading indicator of ecological transition. Ecology Letters, 9, 311-318.

Chambert, T., Miller, D. A., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. Ecology, 96, 332-339.

Chen, G., Kery, M., Plattner, M., Ma, K., & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. Journal of Ecology, 101, 183-191.

De Cáceres, M., & Brotons, L. (2012). Calibration of hybrid species distribution models: the value of general-purpose vs. targeted monitoring data. Diversity and Distributions, 18, 977-989.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment,10, 291-297.

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. In D. Futuyma, H. Shafer, & D. Simberloff

(Eds.), Annual Review of Ecology, Evolution, and Ssystematics, Vol. 41, pp. 149-172).

Dormann, C. F. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36, 27-46.

Dormann, C. F. (2012). Correlation and process in species distribution models: bridging a dichotomy. Journal of Biogeography, 39, 2119-2131.

Dormann, C. F., Schweiger, O., Arens, P., Augenstein, I., Aviron, S., Bailey, D., et al. (2008a). Prediction uncertainty of environmental change effects on temperate European biodiversity. Ecology Letters, 11, 235-244.

Dormann, C. F., Purschke, O., Márquez, J. R., Lautenbach, S., & Schröder, B. (2008b). Components of uncertainty in species distribution analysis: a case study of the great grey shrike. Ecology, 89, 3371-3386.

Dormann, C. F., McPherson, J. M., Araujo, M., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30, 609-628.

Duque-Lazo, J., van Gils, H., Groen, T. A., & Navarro-Cerrillo, R. M. (2016). Transferability of species distribution models: The case of Phytophthora cinnamomi in Southwest Spain and Southwest Australia. Ecological Modelling, 320, 62-70.

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. Science, 289, 2312-2314.

Ehrlich, P. R., Kareiva, P. M., & Daily, G. C. (2012). Securing natural capital and expanding equity to rescale civilization. Nature, 486, 68-73.

Elith, J., Graham C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29, 129-151.

Elith, J., & Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography, 32, 66-77.

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology Evolution and Systematics, 40, 677-697.

Ellis, E. C. & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. Frontiers in Ecology and the Environment, 6, 439--447.

Evans, M. R. (2012). Modelling ecological systems in a changing world. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367, 181-190.

Evans, M. R., Bithell, M., Cornell, S. J., Dall, S. R., Diaz, S., Emmott, S., et al. (2013a). Predictive systems ecology. Proceedings of the Royal Society B-biological Sciences, 280.

Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., et al. (2013b). Do simple models lead to generality in ecology? Trends In Ecology & Evolution, 28, 578-583.

Feeley, K. J., & Silman, M. R. (2011). Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions, 17, 1132-1140.

Feilhauer, H., Thonfeld, F., Faude, U., He, K.S., Rocchini, D. & Schmidtlein, S., (2013). Assessing floristic composition with multispectral sensors – a comparison based on monotemporal and multiseasonal field spectra. International Journal of Applied Earth Observation and Geoinformation, 21, 218–229.

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions, 13, 252-264.

Fielding, A. H. (2002). What are the appropriate characteristics of an accuracy measure? In J. M. Scott (Hrsg.), Predicting species occurrences (S. 271-280). Island Press.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24, 38-49.

Foody, G. M., & Cutler, M. E. (2003). Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. Journal of Biogeography, 30, 1053-1066.

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. Diversity and Distributions, 19, 1217-1223.

Funtowicz, S. O., & Ravetz, J. R. (1990). Uncertainty and quality in science for policy. Kluwer Academic Publishers Dordrecht, the Netherlands; Norwell, MA, U.S.A.

Galipaud, M., Gillingham, M. A., David, M., & Dechaume-Moncharmont, F.-X. (2014). Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. Methods in Ecology and Evolution, 5, 983-991.

Garrard, G. E., Bekessy, S. A., McCarthy, M. A., & Wintle, B. A. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. Austral Ecology, 33, 986-998.

GEO BON. (2015). Global Biodiversity Change Indicators. Global Biodiversity Change Indicators. Version 1.2.

Goberville, E., Beaugrand, G., Hautekèete, N.-C., Piquot, Y., & Luczak, C. (2015). Uncertainties in the projection of species distributions related to general circulation models. Ecology and Evolution, 5, 1100-1116.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T., Loiselle, B. A., et al. (2008). The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology, 45, 239-247.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution , 19, 497-503.

Gritti, E. S., Duputie, A., Massol, F., & Chuine, I. (2013). Estimating consensus and associated uncertainty between inherently different species distribution models. Methods in Ecology and Evolution, 4, 442-452.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., et al. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography, 24, 276-292.

Guisan, A. & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. { Ecological Modelling, 135, 147-186.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8, 993-1009.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., et al. (2013). Predicting species distributions for conservation decisions. Ecology Letters, 16, 1424-1435.

Guralnick, R. P., Hill, A. W., & Lane, M. (2007). Towards a collaborative, global infrastructure for biodiversity assessment. Ecology Letters, 10, 663-672.

Hagemeijer, E. & Blair, M. J. (1997). The EBCC Atlas of European Breeding Birds: their distribution and abundance. T. & A.D. Poyser, London.

Hagen-Zanker, A. (2009). An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. International Journal of Geographical Information Science, 23, 61-73.

Haila, Y., & Henle, K. (2014). Uncertainty in biodiversity science, policy and management: a conceptual overview. Nature Conservation, 8, 27-43.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., et al. (2013). Big data and the future of ecology. Frontiers in Ecology and the Environment, 11, 156-162.

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., et al. (2015). Will remote sensing shape the next generation of species distribution models? Remote Sensing in Ecology and Conservation, 1, 4-18.

Heikkinen, R. K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography, 35, 276-288.

Hermoso, V., & Kennard, M. J. (2012). Uncertainty in coarse conservation assessments hinders the efficient achievement of conservation goals. Biological Conservation, 147, 52-59.

Hoffmann, A., Penner, J., Vohland, K., Cramer, W., Doubleday, R., Henle, K., et al. (2014). Improved access to integrated biodiversity data for science, practice, and policy - the European Biodiversity Observation Network (EU BON). Nature Conservation, 6, 49-65.

Hortal, J., de Bello, F., Diniz-Filho, J. A., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46, 523-549.

Howell, D., Filin, A. A., Bogstad, B., & Stiansen, J. E. (2013). Unquantifiable uncertainty in projecting stock response to climate change: Example from North East Arctic cod. Marine Biology Research, 9, 920-931.

Huang, J., & Frimpong, E. A. (2016). Limited transferability of stream-fish distribution models among river catchments: reasons and implications. Freshwater Biology, 61, 729-744.

Huang, J., Frimpong, E. A., & Orth, D. J. (2016). Temporal transferability of stream fish distribution models: can uncalibrated SDMs predict distribution shifts over time? Diversity and Distributions, 22, 651-662.

IPCC (2013). IPCC, 2013: Summary for Policymakers. (Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A. Xia, Y., Bex, V. & Midgley, P. M., eds.) Cambridge University Press, Cambridge, United Kingdom and New York, USA.

Jansen, F., & Dengler, J. (2010). Plant names in vegetation databases - a neglected source of bias. Journal of Vegetation science, 21, 1179-1186.

Jeliazkov A., Chiron F., Garnier J., Besnard A., Silvestre M., & Jiguet F. (2014) Level-dependence of the relationships between amphibian biodiversity and environment in pond systems within an intensive agricultural landscape. Hydrobiologia, 723, 7–23.

Jeliazkov A., Mimet A., Chargé R., Jiguet F., Devictor V., & Chiron F. (2016) Impacts of agricultural intensification on bird communities: New insights from a multi-level and multi-facet approach of biodiversity. Agriculture, Ecosystems & Environment, 216, 9–22.

Jenouvrier, S. (2013). Impacts of climate change on avian populations. Global Change Biology, 19, 2036-2057.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. Trends n Ecology \& Evolution, 27, 151-159.

Jiguet, F., Julliard, R., Couvet, D., & Petiau, A. (2005). Modeling Spatial Trends in Estimated Species Richness using Breeding Bird Survey Data: A Valuable Tool in Biodiversity Assessment. Biodiversity & Conservation, 14(, 3305-3324.

Jimenez-Alfaro, B., Draper, D., & Nogues-Bravo, D. (2012). Modeling the potential area of occupancy at fine resolution may reduce uncertainty in species range estimates. Biological Conservation, 147, 190-196.

Jimenez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica, 31, 361-369.

Kaschner, K., Watson, R., Trites, A. W., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Marine Ecology Progress Series, 316, 285-310.

Kattge, J., Diáz, S., Lavorel, S., Prentice, C., Leadley, P., Bönisch, G., et al. (2011). TRY - a global database of plant traits. Global Change Biology, 17, 2905--2935.

Keitt, T. H., & Urban, D. L. (2005). Scale-specific inference using wavelets. Ecology, 86, 2497-2504.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., et al. (2009). Data-intensive Science: A New Paradigm for Biodiversity Studies. Bioscience, 59, 613-620.

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. Trends in Ecology & Evolution, 18, 299-305.

Kesner-Reyes, K. K., & Froese, R. (2016). AquaMaps: algorithm and data sources for aquatic organisms. AquaMaps: algorithm and data sources for aquatic organisms.

Keyfitz, N. (1972). On future population. Journal of the American Statistical Association, 67, 347-363.

Klir, G. J., & Wierman, M. J. (1999). Uncertainty-Based Information: Elements of Generalized Information Theory. (Kacprzyk, J., ed.) Physica-Verlag.

Knight, F. H. (1921). Risk, Uncertainty and Profit. Boston, MA: Houghton Mifflin Co.

Knights, A., Culhane, F., Hussain, S., Papadopoulou, K., Piet, G., Raakær, J., et al. (2014). A step-wise process of decision-making under uncertainty when implementing environmental policy . Environmental Science & Policy, 39, 56-64.

Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. Remote Sensing of Environment, 113, 957-964.

Kriticos, D. J. (2012). CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. Methods in Ecology and Evolution, 3, 53-64.

Kuemmerlen, M., Stoll, S., Sundermann, A., & Haase, P. (2016). Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. Ecological Indicators , 65, 122-132.

Kühn, I. & Dormann, C. F. (2012). Less than eight (and a half) misconceptions of spatial analysis. Journal of Biogeography, 39, 995-998.

Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distributions 13, 66-69.

Kujala, H., Burgman, M. A., & Moilanen, A. (2013). Treatment of uncertainty in conservation under climate change. Conservation Letters, 6, 73-85.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. Global Ecology and Biogeography, 23, 504-515.

Lahti, T. & Lampinen, R. (1999). From dot maps to bitmaps: Atlas Florae Europaeae goes digital. Acta Botanica Fennica, 162, 5-9.

Lauzeral, C., Leprieur, F., Beauchard, O., Duron, Q., Oberdorff, T., & Brosse, S. (2011). Identifying climatic niche shifts using coarse-grained occurrence data: a test with non-native freshwater fish. Global Ecology and Biogeography, 20, 407-414.

Lawson, C. R., Hodgson, J. A., Wilson, R. J., & Richards, S. A. (2014). Prevalence, thresholds and the performance of presence-absence models. Methods in Ecology and Evolution, 5, 54-64.

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs, 72, 445-463.

Lobo, J. M., Jimenez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145-151.

Loiselle, B. A., Jorgensen, P. M., Consiglio, T., Jimenez, I., Blake, J. G., Lohmann, L. G., et al. (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography, 35, 105-116.

MacKenzie, D. I. (2006). Modeling the probability of resource use: The effect of, and dealing with, detecting a species imperfectly. Journal of Wildlife Management, 70, 367-374.

Manceur, A. M., & Kühn, I. (2014). Inferring model-based probability of occurrence from preferentially sampled data with uncertain absences using expert knowledge. Methods in Ecology and Evolution, 5, 739-750.

Maxim, L., & van der Sluijs, J. P. (2011). Quality in environmental science for policy: Assessing uncertainty as a component of policy analysis. Environmental Science & Policy, 14, 482-492.

McInerny, G. J. (2013). Embedding visual communication into scientific practice. Trends in Ecology & Evolution, 28, 13-14.

McInerny, G. J. & Purves, D. W. (2011). Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. Methods in Cell Biology, 2, 248-257.

McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Journal of Applied Ecology, 41(5), 811-823.

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature Communications, 6, 8221.

Meyer, C. (2016). Limitations in global information on species occurrences. Frontiers of Biogeography, 8, e28195

Meyer, C., Weigelt, P., & Kreft, H. (2016a). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 19, 992-1006.

Meyer, C., Jetz, W., Guralnick, R. P., Fritz, S. A., & Kreft, H. (2016b). Range geometry and socio-economics dominate species-level biases in occurrence information. Global Ecology and Biogeography, 25, 1181-1193.

Mihoub, J.-B., Henle, K., Titeux, N., Brotons, L.s., Brummitt, N. A., & Schmeller, D. S. (2017). Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. Scientific Reports, 7, 41591.

Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology, 92, 1422-1428.

Moerman, D. E., & Estabrook, G. F. (2006). The botanist effect: counties with maximal species richness tend to be home to universities and botanists. Journal of Biogeography, 33, 1969-1974.

Moilanen, A., Runge, M. C., Elith, J., Tyre, A., Carmel, Y., Fegraus, E., et al. (2006). Planning for robust reserve networks using uncertainty analysis. Ecological Modelling, 199, 115-124.

Mora, C., Tittensor, D. P., & Myers, R. A. (2008). The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proceedings of the Royal Society B-Biological Sciences, 275, 149-155.

Murtaugh, P. A. (2009), Performance of several variable-selection methods applied to real ecological data. Ecology Letters, 12, 1061–1068.

Nagendra, H. (2002). Opposite trends in response for the Shannon and Simpson indices of landscape diversity. Applied Geography, 22

Naimi, B., Skidmore, A. K., Groen, T. A., & Hamm, N. A. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. Journal of Biogeography, 38, 1497-1509.

Olthof, I., Butson, C., & Fraser, R. (2005). Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. Remote Sensing of Environment, 95, 290-302.

Palmer, M. W., Earls, P. G., Hoagland, B. W., White, P. S., & Wohlgemuth, T. (2002). Quantitative tools for perfecting species lists. Environmetrics, 13, 121-137.

Pax-Lenney, M., Woodcock, C. E., Macomber, S. A., Gopal, S., & Song, C. (2001). Forest mapping with a generalized classifier and Landsat TM data. Remote Sensing of Environment, 77, 241-250.

Pe'er, G., Mihoub, J.-B., Dislich, C., & Matsinos, Y. G. (2014). Towards a different attitude to uncertainty. Nature Conservation 8, 95-114.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H., Scholes, R. J., et al. (2013). Essential Biodiversity Variables. Science, 339, 277-278.

Peterson, T. A., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography, 30, 550-560.

Peterson, T. A., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., et al. (2011). Ecological Niches and Geographic Distributions. Princeton University Press.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. Journal of Applied Ecology, 51, 839-848.

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T. P., Fernandez, M., et al. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. Remote Sensing in Ecology and Conservation, 2, 122-131.

Polasky, S., Carpenter, S. R., Folke, C., & Keeler, B. (2011). Decision-making under great uncertainty: environmental management in an era of global change. Trends in Ecology & Evolution, 26, 398-404.

Proença, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., et al. (2016). Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. Biological Conservation, https://doi.org/10.1016/j.biocon.2016.07.014

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ramnath, N. S. (2017) Gigerenzer's simple rules. https://www.foundingfuel.com/article/gigerenzers-simple-rules/, accessed at 28rd of March, 2017

Randin, C. F., Dirnbock, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? Journal of Biogeography, 33, 1689-1703.

Rao, C. R. (1982). Diversity and Dissimilarity Coefficients - A Unified Approach. Theoretical Population Biology, 21, 24-43.

Real, R., Marquez, A. L., Olivero, J., & Estrada, A. (2010). Species distribution models in climate change scenarios are still not useful for informing policy planning: an uncertainty assessment using fuzzy logic. Ecography, 33, 304-314.

Reed, D. H., O'Grady, J. J., Brook, B. W., Ballou, J. D., & Frankham, R. (2003). Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. Biological Conservation, 113, 23-34.

Refsgaard, J. C., van der Sluijs, J. P., Hojberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process - A framework and guidance. Environmental Modelling & Software, 22, 1543-1556.

Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications, 12, 618-628.

Ricotta, C., & Moretti, M. (2011). CWM and Rao's quadratic diversity: a unified framework for functional ecology. Oecologia, 167, 181-188.

Ritholtz, B. (2012). Defining Risk Versus Uncertainty. Defining Risk Versus Uncertainty. http://ritholtz.com/2012/12/defining-risk-versus-uncertainty/

Rocchini, D., & Neteler, M. (2012a). Spectral rank-abundance for measuring landscape diversity. International Journal of Remote Sensing, 33, 4458-4470.

Rocchini, D., & Neteler, M. (2012b). Let the four freedoms paradigm apply to ecology. Trends in Ecology & Evolution, 27, 310-311.

Rocchini, D., & Ricotta, C. (2007). Are landscapes as crisp as we may think? Ecological Modelling , 204, 535-539.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jimenez-Valverde, A., Ricotta, C., et al. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Progress in Physical Geography, 35, 211-226.

Rocchini, D., Marcantonio, M., & Ricotta, C. (2017). Measuring Rao's Q diversity index from remote sensing: An open source solution. Ecological Indicators, 72, 234-238.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93-104.

Rokitnicki-Wojcik, D., Wei, A. H., & Chow-Fraser, P. (2011). Transferability of object-based rule sets for mapping coastal high marsh habitat among different regions in Georgian Bay, Canada. Wetlands Ecology and Management, 19, 223-236.

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. Ecological Modelling, 90, 229-244.

Sala, O. E.-S. (2000). Biodiversity - Global biodiversity scenarios for the year 2100. Science, 287, 1770-1774.

Sarrazin, F., & Lecomte, J. (2016). Evolution in the Anthropocene. Science, 351, 922-923.

Scheiner, S. M., Cox, S. B., Willig, M., Mittelbach, G. G., Osenberg, C., & Kaspari, M. (2000). Species richness, species-area curves and Simpson's paradox. Evolutionary Ecology Research, 2, 791-802.

Schimel, D. S., Asner, G. P., & Moorcroft, P. (2013). Observing changing ecological diversity in the Anthropocene. Frontiers in Ecology and the Environment, 11, 129-137.

Schmidtlein, S., & Sassin, J. (2004). Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. Remote Sensing of Environment, 92, 126-138.

Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology & Evolution, 25, 479-486.

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27, 623-656.

Shekhar, S., Schrater, P. R., Vatsavai, R. R., Wu, W. L., & Chawla, S. (2002). Spatial contextual classification and prediction models for mining geospatial data. IEEE Transactions in Multimedia, 4.

Singer, A., Johst, K., Banitz, T., Fowler, M. S., Groeneveld, J., Gutierrez, A. G., et al. (2016). Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? Ecological Modelling, 326, 63-74.

Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., et al. (2015). Environmental science: agree onbiodiversity metrics to track from space: comment. Nature, 523, 403–405.

Spangenberg, J. H., Bondeau, A., Carter, T. R., Fronzek, S., Jaeger, J., Jylha, K., et al. (2012). Scenarios for investigating risks to biodiversity. Global Ecology and Biogeography, 21, 5-18.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing Uncertainty About the Future. Science, 333, 1393-1400.

Steffen, W., Crutzen, P. J., & McNeill, J. R. (2007). The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature? Ambio, 36, 614-621.

Tabor, K., & Williams, J. W. (2010). Globally downscaled climate projections for assessing the conservation impacts of climate change. Ecological Applications, 20, 554-565.

Tuanmu, M. N., Vina, A., Roloff, G. J., Liu, W., Ouyang, Z. Y., Zhang, H. M., et al. (2011). Temporal transferability of wildlife habitat models: implications for habitat monitoring. Journal of Biogeography, 38, 1510-1523.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. Ecological Applications, 13, 1790-1801.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling, 203, 312-318.

Ver Hoef, J. M., & Boveng, P. L. (2015). Iterating on a single model is a viable alternative to multimodel inference. The Journal of Wildlife Management, 79, 719-729.

Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., et al. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment, 4, 5-17.

Wang, L. F., & Jackson, D. A. (2014). Shaping up model transferability and generality of species distribution modeling for predicting invasions: implications from a study on Bythotrephes longimanus. Biological Invasions, 16, 2079-2103.

Watling, J. I., Brandt, L. A., Bucklin, D. N., Fujisaki, I., Mazzotti, F. J., Romanach, S. S., et al. (2015). Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. Ecological Modelling, 309, 48-59.

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3, 260-267.

Wenger, S. J., Som, N. A., Dauwalter, D. C., Isaak, D. J., Neville, H. M., Luce, C. H., et al. (2013). Probabilistic accounting of uncertainty in forecasts of species distributions under climate change. Global Change Biology, 19, 3343-3354.

White, M. A., Brunsell, N., & Schwartz, M. D. (2003). Vegetation Phenology in Global Change Studies. (Schwartz, M. D., ed), Phenology: An Integrative Environmental Science, pp. 453-466. Dordrecht: Springer Netherlands.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. Proceedings of the National Academy of Sciences of the United States of America, 106, 19729-19736.

Wilby, R. L., Wigley, T. M., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., et al. (1998). Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research, 34, 2995-3008.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., et al. (2010). An ecologist's guide to the animal model. Journal of Animal Ecology, 79, 13-26.

Wu J. & Li H. (2006) Concepts of scale and scaling. Scaling and uncertainty analysis in ecology (Wu, J., Jones, K. B., Li, H. & Loucks, O. L., eds.), pp. 3–15. Springer Science & Business Media, Netherlands.

Xie, Y. C., Sha, Z. Y., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology, 1, 9-23.

Yang, W., Ma, K., & Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species richness-environment models. Journal of Biogeography, 40, 1415-1426.

Yang, Y., Xu, J. H., Hong, Y. L., & Lv, G. H. (2012). The dynamic of vegetation coverage and its response to climate factors in Inner Mongolia, China. Stochastic Environmental Research and Risk Assessment, 26, 357-373.

Ye, X. P., Wang, T. H., Skidmore, A. K., Fortin, D., Bastille-Rousseau, G., & Parrott, L. (2015). A wavelet-based approach to evaluate the roles of structural and functional landscape heterogeneity in animal space use at multiple scales. Ecography, 38, 740-750.

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. Trends in Ecology & Evolution, 16, 446-453.

Zurek, M. B., & Henrichs, T. (2007). Linking scenarios across geographical scales in international environmental assessments. Technological Forecasting and Social Change , 74, 1282-1295.

7. Annexes

7.1. Setting temporal baselines for biodiversity could be an impediment for capturing the full impact of anthropogenic pressures

Partners involved: UFZ, EBCC-CTFC, NHM

Project leaders: Jean-Baptiste Mihoub, Dirk Schmeller (UFZ)

Manuscript title

Setting temporal baselines for biodiversity could be an impediment for capturing the full impact of anthropogenic pressures [published in *Scientific Reports*]

One sentence summary

We report on the temporal baselines that could be drawn from biodiversity monitoring schemes in Europe and compare those with the rise of important anthropogenic pressures.

Authors

J.-B. Mihoub[1,2], K. Henle[1], N. Titeux[3,4], L. Brotons[3,4,5], N. A. Brummitt[6] and D.S. Schmeller[1,7,8]

Affiliations

[1] *UFZ - Helmholtz Centre for Environmental Research, Department of Conservation Biology, Permoserstr. 15, 04318 Leipzig, Germany.*

[2] *Sorbonne Universités, UPMC Université Paris 06, Muséum National d'Histoire Naturelle, CNRS, CESCO, UMR 7204, 61 rue Buffon, 75005, Paris, France.*

[3] *European Bird Census Council (EBCC) and Forest Sciences Centre of Catalonia (CEMFOR-CTFC), InForest Joint Research Unit (CSIC-CTFC-CREAF), Ctra. Sant Llorenç de Morunys km 2, 25280 Solsona, Spain.*

[4] *Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), 08193 Cerdanyola del Vallés, Spain.*

[5] *Consejo Superior de Investigaciones Científicas (CSIC), 08193 Cerdanyola el Vallés, Spain.*
[6] *Department of Life Sciences, The Natural History Museum, Cromwell Road, South Kensington, London SW75BD, UK.*

[7] *CNRS, EcoLab, 31062 Toulouse, France.*

[8] *Université de Toulouse, UPS, INPT, EcoLab (Laboratoire Ecologie Fonctionnelle et Environnement), 118 route de Narbonne, 31062 Toulouse, France.*

*Correspondence to: jean-baptiste.mihoub@mnhn.fr

Abstract

Temporal baselines are needed for biodiversity, in order for the change in biodiversity to be measured over time, the targets for biodiversity conservation to be defined and conservation progress to be evaluated. Limited biodiversity information is widely recognized as a major barrier for identifying temporal baselines, although a comprehensive quantitative assessment of this is lacking. Here, we report on the temporal baselines that could be drawn from biodiversity monitoring schemes in Europe and compare those with the rise of important anthropogenic pressures. Most biodiversity monitoring schemes were initiated late in the 20th century, well after anthropogenic pressures had already reached half of their current magnitude. Setting temporal baselines from biodiversity monitoring data would therefore underestimate the full range of impacts of major anthropogenic pressures. In addition, biases among taxa and organization levels provide a truncated picture of biodiversity over time. These limitations need to be explicitly acknowledged when designing management strategies and policies as they seriously constrain our ability to identify relevant conservation targets aimed at restoring or reversing biodiversity losses. We discuss the need for additional research efforts beyond standard biodiversity monitoring to reconstruct the impacts of major anthropogenic pressures and to identify meaningful temporal baselines for biodiversity.

Introduction

A comprehensive understanding of biodiversity responses to anthropogenic pressures is necessary if human development is to remain within planetary boundaries[1], and for assessing its impact on biological evolution in the Anthropocene[2]. Temporal baselines are essential for reliably measuring changes in biodiversity over time[3], for instance by mitigating the consequences of the shifting reference syndrome[4–6]. Further, temporal baselines also frame conservation objectives by identifying the biodiversity reference states aimed for guiding the feasibility of and efforts required to reach those objectives[7], and by defining the time-period within which progress and change are to be evaluated[8].

In this respect, the lack of knowledge about biodiversity states prior to the rise of harmful anthropogenic activities is a critical limitation for understanding the full impact of such pressures and, therefore, for implementing appropriate conservation goals and strategies. Failing to set relevant temporal baselines for biodiversity represents a major risk for implementing effective biodiversity conservation. It may decrease our understanding of past and therefore current changes, misinform conservation objectives and restrict our ability to assess progress. Nonetheless, there are several obstacles that limit our ability to define relevant temporal baselines for biodiversity.

Monitoring schemes provide an important source of information on biodiversity change, guiding further research, conservation assessment and planning[9]. Monitoring schemes are typically used to document changes in biodiversity over time, making the implicit assumption that the state of biodiversity when the scheme started is an appropriate temporal baseline against which to measure that change. However, most structured biodiversity monitoring schemes have been initiated within the last few decades, whereas most of the anthropogenic pressures that are currently impacting biodiversity have been operating over centuries or even millennia[10–12]. Current drivers of biodiversity decline, such as habitat loss and fragmentation, exploitation, pollution, climate change or species introductions result from processes initiated long ago by accelerating agricultural, technological and industrial developments, driven by an increasing human population and its societal needs[13–16]. This mismatch between the restricted

temporal coverage of biodiversity monitoring and the long history of anthropogenic pressures inevitably limits any assessment of the full impacts of such pressures on biodiversity[12,17].

Furthermore, the biodiversity data from these schemes remain scattered, suffer from geographic and taxonomic bias and from strong methodological heterogeneity across space and time[18–20]. These issues make such data difficult to access, to assemble and to analyze over large spatial and temporal scales[9,21,22]. Although significant efforts are underway to mobilize and standardize biodiversity data globally[23], progress towards the fully operational integration of information across scales is still insufficient to provide unbiased knowledge of the status and trends of biodiversity[24]. The recently proposed Essential Biodiversity Variables (EBVs), encompassing six EBV classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem function, and Ecosystem structure), provide a framework for comprehensively representing the different components of biodiversity in order to measure change over time[24,25], to identify the most important gaps in data coverage and to improve monitoring practices across time and space[3,26].

Although the limitations of biodiversity information available from monitoring schemes are widely recognized, a comprehensive and quantitative evaluation of the potential of monitoring schemes to identify temporal baselines capturing the impacts of major anthropogenic pressures on biodiversity is still lacking. Yet such an assessment is urgently required as it would help provide stakeholders with precise information on the knowledge gaps in currently available biodiversity data. Here, we conduct such a quantitative evaluation of the temporal baselines that could be identified using comprehensive information on biodiversity monitoring schemes sourced from several meta-databases. We focus on Europe as one of the regions of the world with the oldest and most intensive biodiversity monitoring efforts. We report the start of European biodiversity monitoring schemes to examine the possibilities offered by available data for documenting past states of biodiversity with respect to different (i) taxonomic groups, (ii) EBV classes and (iii) types of data collected. Then, we compare the onset of biodiversity monitoring schemes with historical time-series or reconstructions of the main anthropogenic pressures that are currently acting on biodiversity at global or regional scales. We show that the past biodiversity states that may be estimated from available biodiversity monitoring data are unlikely to reflect the full impact of anthropogenic pressures on biodiversity. We highlight the implications for setting appropriate temporal baselines and the consequences for biodiversity conservation management practices and policies, and we provide recommendations on possible ways to move forward with this.

Methods

Biodiversity monitoring databases

The databases considered in this study were selected according to the following criteria: they provide meta-data on biodiversity monitoring schemes, they are representative of monitoring practices in Europe and they contain relevant information across taxa. We considered primarily the most comprehensive meta-database describing standard information on biodiversity monitoring practices in Europe (hereafter DaEuMon). DaEuMon is based on questionnaires and was compiled under the FP6-project EuMon[9,46]. We considered here all schemes focusing on species monitoring that were reported in DaEuMon up to 2009 (N = 452). Since DaEuMon may only report a fraction of biodiversity monitoring schemes in Europe[9,46], we considered other independent sources of data documenting biodiversity monitoring schemes in order to provide the most representative overview of existing biodiversity information in Europe. We selected two additional databases with high quality

control, consistent standards and compatible meta-data structure with regard to sourced references and taxonomic, temporal and spatial coverage: The Participatory Monitoring Networks in Europe database (PMN[47]) and the Global Population Dynamics Database (GPDD) Version 2.048. Like DaEuMon, the PMN database has been compiled within the FP6-project EuMon. The PMN database gathered information related to biodiversity monitoring schemes in Europe (N = 326) based on a different questionnaire structure from DaEuMon, with a very marginal overlap of schemes between the two databases. The GPDD database is one of the largest, freely available databases on species population dynamics worldwide, from which we considered only schemes conducted in Europe (see Supplementary Methods; N = 177). We combined the different biodiversity monitoring schemes from the three meta-databases whenever data interoperability allowed (see below for details).

General approach and assumptions

We considered the starting year of each biodiversity monitoring scheme as a surrogate of the oldest state of biodiversity that can be estimated from that scheme. We broke these metrics down with respect to the (i) taxonomic group studied (ii) type of data collected (e.g. species occurrence record or count) and (iii) EBV class targeted by the schemes (for a comprehensive description of the EBV considered within each of the EBV classes see ref. 49).

Including the PMN and GPDD databases helped to improve the comprehensiveness of biodiversity monitoring when compared to the use of DaEuMon only. Combining the different databases helped counterbalance potential biases in each individual database in terms of temporal, geographical and taxonomic coverage (see Supplementary Methods and Supplementary Fig. S1 for the taxonomic coverage). Nevertheless, integrating complementary information was only possible for the comparison between taxonomic groups due to limitations in data interoperability between the three databases. As the three databases partially differed in terms of taxonomic resolution – for example, plants were mostly mentioned as "Plants" within PMN, and as "Orchids", "Mosses, liverworts & ferns" and "Other plants" in DaEuMon – we aggregated schemes to the lowest common taxonomic level of the three databases for each taxonomic group. In contrast to taxonomy, there was no information available about the EBV class targeted and the type of data collected in PMN. In addition, GPDD almost exclusively contains biodiversity monitoring schemes that have collected count data and that have targeted the EBV class 'Species Populations' (specifically through the EBV 'Population abundance'). The comparison between data types and EBV classes was thus not possible from the PMN database, and integrating information from PDD would have strongly skewed the analysis toward one type of data and one EBV class. Consequently, the comparison between the types of data collected and the EBV classes targeted by biodiversity monitoring schemes in Europe was only carried out using DaEuMon. PMN and GPDD push back the starting years of biodiversity monitoring schemes compared to the use of DaEuMon only, but the latter provides the most representative and comprehensive overview of biodiversity monitoring practices in Europe.

Nevertheless, it is possible that the number of monitoring schemes collecting data on phenology and focusing on the EBV classes 'Species Traits' and 'Genetic Composition' are under-represented in DaEuMon. However, most trait or DNA databases do not contain structured monitoring data that allow documenting changes over time and are restricted to specific taxonomic groups (e.g. Polytraits for marine polychaetes[50] or YouTHERIA[51] for mammals). While trait-based monitoring databases documenting changes over time do exist, they remain scattered, difficult to access and to our knowledge are not currently compiled in

any meta-database, so that such trait-based monitoring databases could not be considered in this study.

Analysis

For each taxonomic group studied, type of data collected and EBV class targeted, we calculated descriptive metrics of the temporal baseline that could be drawn for biodiversity based on the starting year of the biodiversity monitoring schemes in Europe (median, mean, minimum or maximum). We then compared the start of biodiversity monitoring schemes with global or regional long-term time-series reflecting the major anthropogenic pressures that are known to impact biodiversity the most[1]: global human population size[52], European temperature anomalies[53], global land use changes[54,55], global anthropogenic nitrogen and phosphorus[56], atmospheric concentration of carbon dioxide[57] and contaminant emissions in the United Kingdom (furan and dioxin58, considered as representing emissions in other European countries). In order to provide a quantitative assessment of the mismatch between the start of biodiversity monitoring schemes and the onset of anthropogenic pressures but without making any assumption about the causal relationship between the pressure and its impact on biodiversity, we here report the level that each pressure had already reached when biodiversity monitoring schemes were initiated. We first identified the value of the pressure p_i corresponding to the starting year of each scheme i by projecting the intersect between the starting year of the scheme i and the regression trend of the pressure on the pressure axis (see Supplementary Figure S3). We then determined the level of pressure reached at that time, expressed as the percentage of the pressure range already reached when the schemes started, as follows:

$$\% \text{ pressure range reached} = \text{medP} - \text{minP}/\text{rangP}$$

where the medP is the median of all pi, minP is the minimum value of the pressure over time and rangP is the known range of that pressure, which was calculated as the difference between the maximum and minimum values of the pressure along the time-series.

We used a non-parametric Kruskal-Wallis test to assess differences between the different categories considered in the biodiversity monitoring schemes (i.e. taxonomic groups studied, types of data collected and EBV classes targeted). For categories in which significant heterogeneity was found using the Kruskal-Wallis test, we performed a post-hoc analysis using the Conover-Iman multiple pair-wise comparisons test[59]. Adjustments of multiple pair-wise comparisons were made using the Benjamini-Hochberg procedure controlling for false discovery rate, which are more reliable than classical Bonferoni procedures[60]. All statistical analysis were performed using the R software[61] (including the package conover.test for post-hoc analysis). Importantly, a single biodiversity monitoring scheme may have included several taxonomic groups, collected different types of data or targeted several EBV classes, and information might have been provided for some components of the questionnaires but not for others within a single monitoring scheme. Therefore, the number of monitoring schemes considered may differ between the different topical comparisons as well as the total number of schemes contained in the three databases.

Results

Biodiversity monitoring and the history of major anthropogenic pressures

Most of the major anthropogenic pressures that are known to impact biodiversity began hundreds of years earlier than the start of biodiversity monitoring schemes (Fig. 1). In Europe, most of these schemes started in the late 20th century (Fig. 1 and Table 1). Only a small proportion of these schemes were initiated before the middle of the 20th century (c.a. 12.5% before 1950, N = 210) and c.a. 50.6% (N = 857) started 1990 or later. More importantly, anthropogenic pressures started to escalate exponentially from the beginning or the middle of the 20th century, while the vast majority of biodiversity monitoring schemes started only after these pressures had already reached more than half of their present-day order of magnitude or had already peaked and decreased (Fig. 1 and Supplementary Table S1). As a consequence, a large part of the anthropogenic pressures on biodiversity have operated long before any data on the past states of biodiversity was recorded by monitoring schemes in Europe.

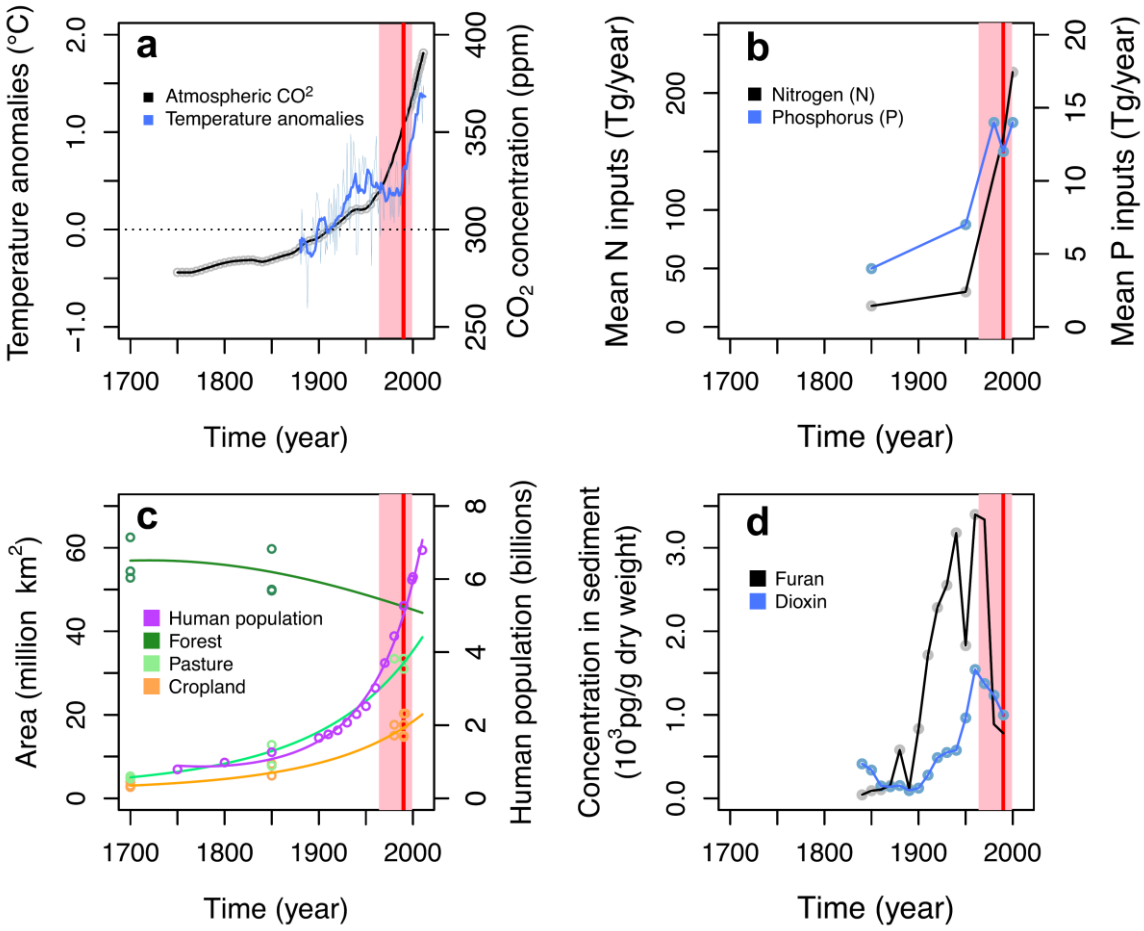


Figure 1 *Temporal mismatch between biodiversity monitoring schemes in Europe and major global or regional anthropogenic pressures known to impact biodiversity. The onset of biodiversity monitoring is represented using the median value (vertical red line) and the first and third quartiles (light red area) of the starting years of biodiversity monitoring schemes (see Table 1). Major pressures include* ***(a)*** *climate: global temperature anomalies and European atmospheric concentrations of carbon dioxide,* ***(b)*** *global anthropogenic nitrogen and phosphorus,* ***(c)*** *global human population sizes and global land use changes and* ***(d)*** *pollutant emissions in the United Kingdom (UK) (sourced from*[52–58]*).*

Table 1 *Temporal baselines of biodiversity monitoring schemes in Europe: the summary statistics of the starting years for the schemes are described for each of (a) the taxonomic groups studied, (b) the EBV classes targeted and (c) the type of data collected.*

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N
Taxonomic group							
Amphibians	1817	1974	1994	1983	2000	2008	155
Birds	1634	1962	1986	1974	1998	2007	458
Fish	1634	1962	1979	1971	1994	2008	154
Insect	1804	1966	1989	1977	2000	2009	265
Mammals	1538	1964	1990	1974	2000	2008	339
Mollusc	1907	1971	2001	1981	2002	2008	18
Plants	1634	1974	1993	1975	2000	2008	245
Reptiles	1817	1982	1993	1982	2000	2008	51
Overall	**1538**	**1964**	**1990**	**1976**	**1999**	**2009**	**1685**
EBV class							
Genetic Composition	1980	1982	1985	1990	1994	2004	3
Community Composition	1901	1991	2000	1995	2005	2009	164
Species Populations	1800	1990	1999	1994	2004	2009	433
Species Traits	1950	1956	1979	1980	2007	2008	5
Overall	**1800**	**1990**	**1999**	**1995**	**2004**	**2009**	**605**
Type of data							
Capture Mark Recapture	1901	1988	1992	1989	2000	2005	43
Count	1933	1990	1999	1995	2005	2008	300
Occurrence	1800	1993	2000	1994	2003	2009	72
Phenology	1960	1990	1999	1993	2000	2005	21
Population Structure	1952	1994	2000	1994	2004	2008	16
Overall	**1800**	**1990**	**1999**	**1994**	**2004**	**2009**	**452**

Taxonomic groups

Biodiversity monitoring schemes in Europe focus on amphibians, birds, fishes, insects, mammals, molluscs, plants and reptiles (Table 1 and Supplementary Fig. S1). We found strong heterogeneity among taxonomic groups in the start of biodiversity monitoring schemes (Chi2 7 = 33.314, N = 1635, $p < 0.001$, Figs 2a and 3), with an exponential overall increase in the number of schemes starting from the 1950's (Fig. 4). In terms of median starting dates, birds and fishes are the focus of the oldest schemes, whereas schemes focusing on amphibians, molluscs, plants and reptiles are more recent (approx. a decade later; Table 1 and Supplementary Table S2). Birds and mammals have been the most common focus of the schemes (27%, N = 458 and 20%, N = 339, respectively). Other taxonomic groups such as amphibians, fish, plants and insects were less studied but reptiles and molluscs were the least monitored groups (3%, N = 51 and 1%, N = 18 respectively; Table 1, Figs 3 and 4 and

Supplementary Fig. S1). A very few monitoring schemes were implemented before or near the onset of major anthropogenic pressures, e.g. mammals in 1538, and birds and plants in 1634 (Table 1, Figs 2a and 4) but these mostly entailed non-systematic monitoring approaches or covered relatively small spatial extents.

EBV classes and type of data collected

Comparisons of starting years among EBV classes and types of data collected were only possible for a reduced set of monitoring schemes (see Methods). Although using this restricted set meant ignoring some of the oldest schemes, the overall picture of the start of monitoring schemes dating back to the mid 1990's is consistent with the findings resulting from all databases previously found for the taxonomic groups (Table 1 and Fig. 2).

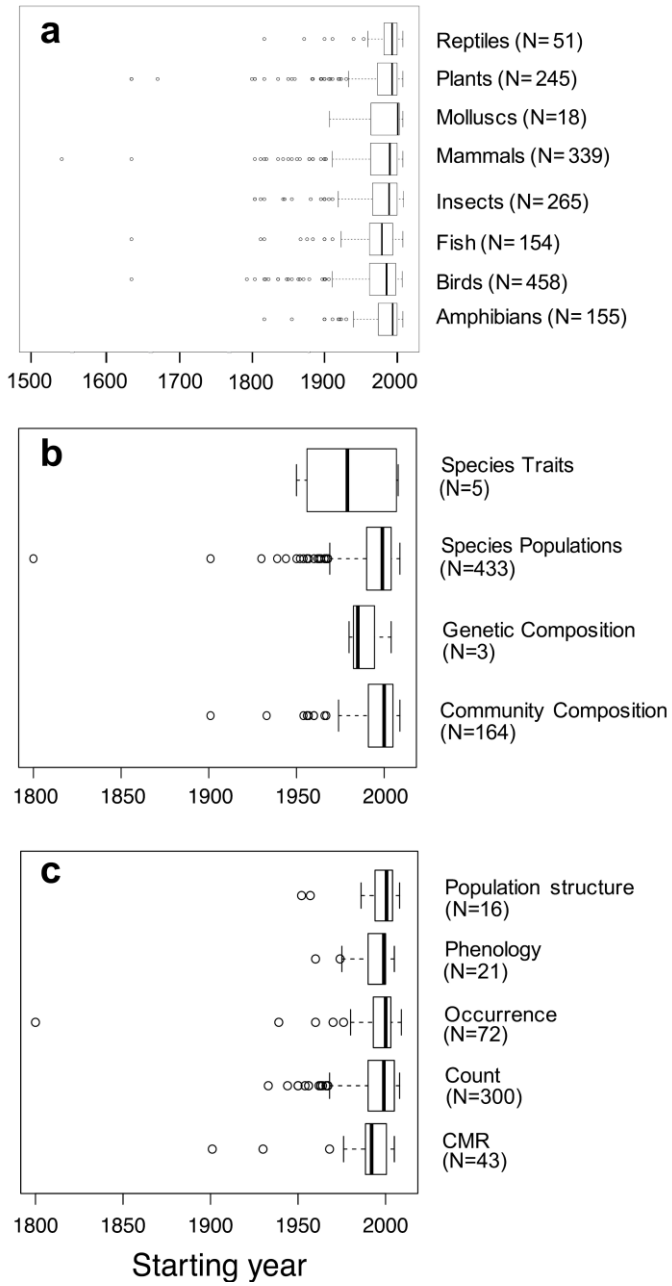


Figure 2 *Univariate boxplots based on the starting year of biodiversity monitoring schemes in Europe for each of (a) the taxonomic groups studied (from entire database), (b) the EBV classes targeted and (c) the type of data collected (from reduced dataset using DaEuMon only; see Methods).*

The monitoring schemes have targeted 4 out of the 6 EBV classes from the EBV framework[24]: Genetic Composition, Species Populations, Species Traits and Community Composition. The types of data collected in the monitoring schemes include abundance of individuals (count), records of species' presence/absence (occurrence), capture-mark-recapture data (CMR), phenological events (phenology) and measures of the population structure (population structure).

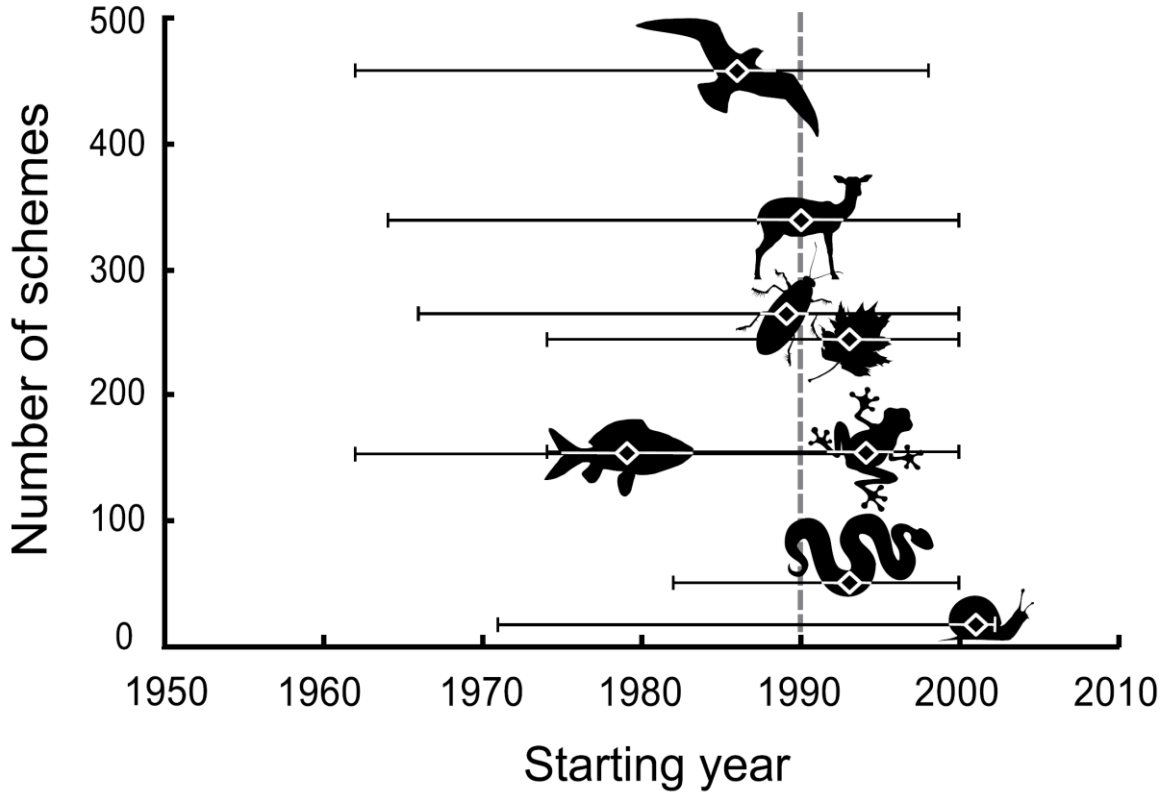


Figure 3 *Taxonomic heterogeneity in the start of biodiversity monitoring schemes in Europe (median starting dates ± first and third quartiles) with respect to the number of schemes. The eight taxonomic groups (amphibians, birds, fishes, insects, mammals, molluscs, plants and reptiles) are represented with schematic icons. Dashed line indicates the overall median starting date across all taxonomic groups.*

Starting years of biodiversity monitoring schemes differed among the types of data collected ($Chi^2_4 = 10.422$, $p = 0.034$, $N = 452$; Fig. 2c). Even though the oldest schemes collected CMR data (Table 1, Fig. 2c and Supplementary Table S2), the focus shifted towards the collection of count data from the 1950's onwards (Fig. 5).

Overall, the majority of the information available from biodiversity monitoring schemes are count data (66.4%, N = 300) and, to a lesser extent, occurrence data (15.9%, N = 72; Table 1; see also Supplementary Fig. S2). In comparison, data on phenology and population structure are collected in only 4.6% (N = 21) and 3.5% (N = 16) of the schemes, respectively (Table 1, Fig. 5 and Supplementary Fig. S2).

We did not find any significant difference in the starting years of the monitoring schemes among the EBV classes targeted (Chi^2_3 = 2.271, p = 0.518, N = 605; Fig. 2b). However, biodiversity monitoring schemes have focussed disproportionately on only two EBV classes: Species Populations (71.6%, N = 433) and Community Composition (27.1%, N = 164; Table 1). In contrast, the EBV classes Species Traits and Genetic Composition have been the focus of only a very small number of schemes (respectively 0.8%, N = 5 and 0.5%, N = 3; Fig. 6).

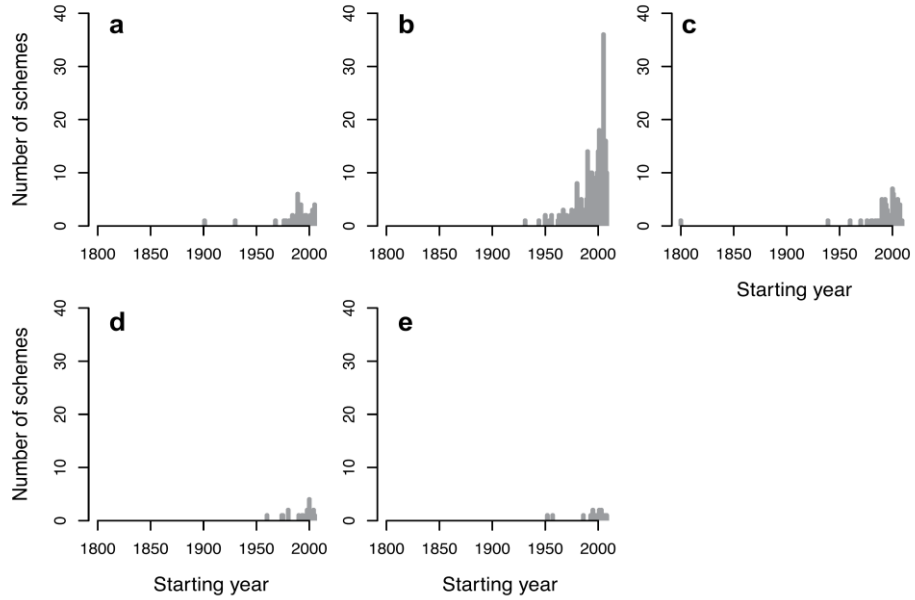


Figure 4 *Number of monitoring schemes initiated over time according to their starting year for each taxonomic group studied: (**a**) amphibians (N = 155), (**b**) birds (N = 458), (**c**) fishes (N = 154), (**d**) insects (N = 265), (**e**) mammals (N = 339), (**f**) molluscs (N = 18), (**g**) plants (N = 245) and (**h**) reptiles (N = 51).*

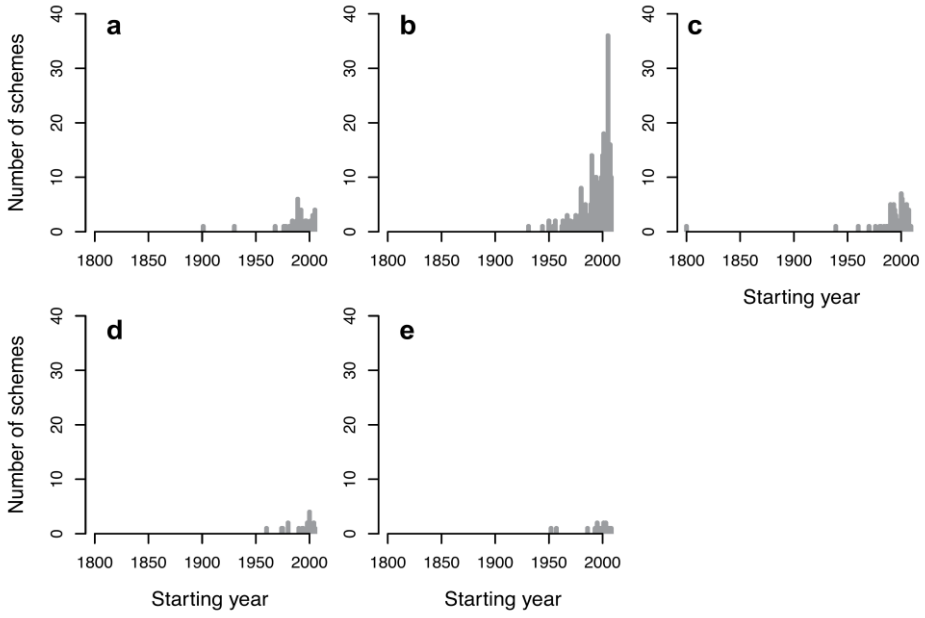


Figure 5 *Number of monitoring schemes initiated over time according to their starting year for each type of data collected: (**a**) CMR (N = 43), (**b**) Count (N = 300), (**c**) Occurrence (N = 72), (**d**) Phenology (N = 21) and (**e**) Population structure (N = 16).*

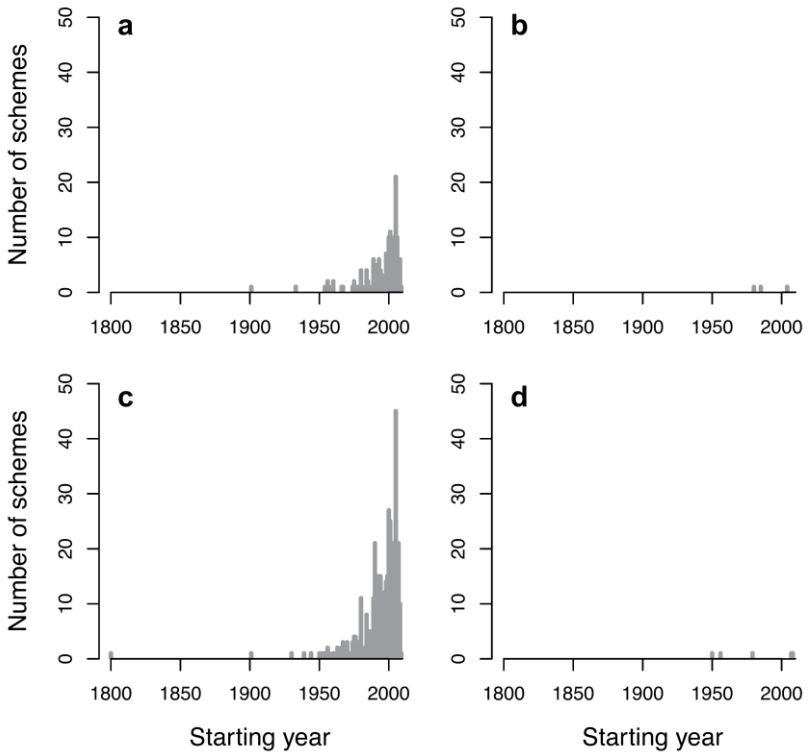


Figure 6 *Number of monitoring schemes initiated over time according to their starting year for each EBV class targeted: (**a**) Community composition (N = 164), (**b**) Genetic composition (N = 3), (**c**) Species population (N = 433) and (**d**) Species traits (N = 5).*

Discussion

We provide here a first quantitative evaluation of the limitations of setting temporal baselines to fully assess the impact of major anthropogenic pressures on biodiversity. Our analysis shows that structured biodiversity monitoring data in Europe do not date back far enough in time to document and assess the full impact of anthropogenic pressures on biodiversity, even for popular taxonomic groups such as birds and mammals. Major anthropogenic pressures have continuously accelerated and escalated since the Quaternary period[13,15], most remarkably during the Industrial Revolution in the middle of the 19th century and from the "Great acceleration" in the 1950's[16,27]. Species extinction rates reported during the last decades are considered to be comparable to those of an extinction crisis[28]. Nevertheless, extinction rates in vertebrates had exceeded the background rates as early as the 18th and 19th centuries, and even before this for some mammal and bird groups[29]. We demonstrate that most of the data currently available from European biodiversity monitoring schemes have been collected from the 1950's onwards, i.e. long after modern anthropogenic pressures might have started to impact species populations and communities[29–31]. The sharp increase in the number of monitoring schemes from the 1990's likely reflecting a response to the reporting commitments outlined in the European Nature Directives32,33 or similar obligations from international conventions, such as the Convention on Biological Diversity or the Convention for Migratory Species34. Our findings are line with previous studies showing that structured biodiversity monitoring schemes have been recently implemented[11,12,35] and that accurate biodiversity data for major realms is not available before the 1960's (marine[12,14,20,36], terrestrial or freshwater[9,10,37]). Despite biodiversity monitoring schemes contributing to an increased understanding of recent anthropogenic impacts, the changing states of biodiversity

since the rise of these pressures are mostly unknown and might be seriously underestimated[28,38].

Beyond the time-series limitations of biodiversity monitoring, our analysis further illustrates a range of different sources of heterogeneity that can further diminish the relevance of available biodiversity data. We implicitly assume in this study that the starting year of monitoring schemes can be considered as a surrogate of the past states of biodiversity to document changes over time. This statement supposes a temporal continuity in monitoring, implying that any scheme ever started is still running today and that there is no temporal gap in the time-series. In practice, however, available biodiversity datasets are, at best, fragmented[37] and most schemes are conducted on a relatively short-term basis10,39 (mean duration of schemes in this study = 15.42 ± 16.34 years, N = 452). Similarly, most biodiversity monitoring schemes are conducted at small geographical scales9,10 so that opportunities to assess past states of biodiversity at global, regional or even national scales remain limited. In addition to limited temporal coverage, inconsistencies in the temporal and spatial continuity of biodiversity monitoring schemes may therefore impose critical constraints for the assessment of biodiversity change over time.

Our analysis also highlights different sources of heterogeneity among biodiversity monitoring schemes, such as the biased representation of some taxonomic groups, the collection of only a few types of data and the relative neglect of several EBV classes. Therefore, in addition to being limited in time, the available data only reflect a fraction of the biodiversity. Existing biases in taxonomic coverage are known limitations that prevent the assessment of the changing state of the whole of biodiversity[19,20,40], but the biases within the types of data collected or biological organisation levels that are the focus of monitoring schemes are much less frequently reported. Even if the emphasis on count and occurrence data does not systematically translate into a bias among EBV classes, the data collected in biodiversity monitoring schemes disproportionately document only two EBV classes (‘Species Populations’ and ‘Community Composition’), and overlook other EBV classes, such as ‘Species Traits’ and ‘Genetic Composition’.

Altogether, irregular temporal coverage and biases in taxonomic groups, types of data collected and EBV classes targeted offer a very truncated picture of biodiversity. Limited temporal coverage only allows a limited subset of the changing state of biodiversity needed to represent the full impact of anthropogenic pressures to be documented[41]. Besides, the majority of available biodiversity information remains inconsistent and incomplete for accurate and consistent estimates of past[12,17] and changing states of biodiversity across taxa or biological organisation levels. This may promote asymmetries in biodiversity assessments and conservation objectives. For instance, if a temporal baseline was to be drawn from available data, the baseline for birds, mammals and fish would have to be set further in the past compared to reptiles, amphibians or molluscs. Consequently, previous global biodiversity assessments have been forced to use various temporal baselines41. In addition, the lack of consistent information about past biodiversity states is likely to maintain vagueness and promote the shifting baseline reference syndrome[4–6] by creating uncertainty about past states of biodiversity[14,42]. Altogether, the temporal limitations and bias in biodiversity monitoring data represent a risk to misinform on the actual states and trends of biodiversity in response to anthropogenic pressures and to misguide the definition of sustainable conservation objectives.

We argue that information derived solely from current biodiversity monitoring schemes is not well suited to setting relevant temporal baselines. To face this important challenge, we encourage both scientists and policy-makers to adopt a more conservative attitude toward

temporal baselines for biodiversity by explicitly recognizing the uncertainties associated with current limitations. This implies acknowledging limits to our ability to document past biodiversity states from monitoring schemes, and that the changes measured from these schemes may seriously underestimate the full impact that major anthropogenic pressures have had on biodiversity. In addition, cross-disciplinary research areas such as bio-archaeology and paleo-ecology offer promising approaches to reconstructing past states and histories of biodiversity using alternative sources of information[17,43,44]. More reliable indicators of biodiversity change could be provided by integrating historical or archeological data with recent biodiversity monitoring data. Additional mobilization and digitization of biodiversity data45 is needed to ensure consistent available data over large spatial extents, but strengthening research efforts to improve the linkage between monitoring, archeological and historical information[17,43,44] is also an important way forward to extend the temporal coverage of available information. These developments and a consistent integration of fragmentary information across disciplines are critical if we are to set temporal baselines for biodiversity that reflect past states of biodiversity before the rise of major anthropogenic pressures.

Acknowledgements

This work was initiated under the framework of the FP7 project EU BON (contract 308454), which provided financial support to J.B.M., D.S.S., N.T., K.H. and N.B. Additional support was provided to L.B. and N.T. through the FORESTCAST project (CGL2014-59742, Ministry of Economy and Competitiveness of Spain).

References

1. Steffen, W. et al. Planetary boundaries: Guiding human development on a changing planet. Science 347, doi: 10.1126/science.1259855 (2015).
2. Sarrazin, F. & Lecomte, J. Evolution in the Anthropocene. Science 351, 922–923 (2016).
3. Schmeller, D. S. et al. Towards a global terrestrial species monitoring program. J. Nat. Cons. 25, 51–57 (2015).
4. Price, A. R. G. et al. Shifting environmental baselines in the Red Sea. Mar. Pollut. Bull. 78, 96–101 (2014).
5. Kai, Z. et al. Shifting Baselines on a Tropical Forest Frontier: Extirpations Drive Declines in Local Ecological Knowledge. Plos One 9, 8 (2014).
6. Pauly, D. Anecdotes and the shifting baseline syndrome of fisheries. Trends Ecol. Evol. 10, 430 (1995).
7. Bull, J. W. et al. Importance of Baseline Specification in Evaluating Conservation Interventions and Achieving No Net Loss of Biodiversity. Cons. Biol. 28, 799–809 (2014).
8. Balmford, A. & Bond, W. Trends in the state of nature and their implications for human well-being. Ecol. Lett. 8, 1218–1234 (2005).
9. Schmeller, D. et al. Bird-monitoring in Europe – a first overview of practices, motivations and aims. Nat. Cons. 2, 41–57 (2012).
10. Marsh, D. M. & Trenham, P. C. Current Trends in Plant and Animal Population Monitoring. Cons. Biol. 22, 647–655 (2008).
11. Lepetz, V., Massot, M., Schmeller, D. S. & Clobert, J. Biodiversity monitoring: some proposals to adequately study species' responses to climate change. Biodiv. Cons. 18, 3185–3203 (2009).
12. Lotze, H. K. & Worm, B. Historical baselines for large marine animals. Trends Ecol. Evol. 24, 254–262 (2009).

13. Waters, C. N. et al. The Anthropocene is functionally and stratigraphically distinct from the Holocene. Science 351, doi: 10.1126/science.aad2622 (2016).
14. Pinnegar, J. K. & Engelhard, G. H. The 'shifting baseline' phenomenon: a global perspective. Rev. Fish Biol. Fish. 18, 1–16 (2008).
15. Lorenzen, E. D. et al. Species-specific responses of Late Quaternary megafauna to climate and humans. Nature 479, 359–364 (2011).
16. Ostberg, S., Schaphoff, S., Lucht, W. & Gerten, D. Three centuries of dual pressure from land use and climate change on the biosphere. Envir. Res. Lett. 10, 044011 (2015).
17. Turvey, S. T., Crees, J. J. & Di Fonzo, M. M. I. Historical data as a baseline for conservation: reconstructing long-term faunal extinction dynamics in Late Imperial–modern China. Proc. R. Soc. B: Biol. Sci. 282, doi: 10.1098/rspb.2015.1299 (2015).
18. Pereira, H. M. et al. Global biodiversity monitoring. Front. Ecol. Env. 8, 459–460 (2010).
19. Collen, B., Ram, M., Zamin, T. & McRae, L. The tropical biodiversity data gap: addressing disparity in global monitoring. Trop. Cons. Sci. 1, 75–88 (2008).
20. Costello, M. J. et al. A census of marine biodiversity knowledge, resources, and future challenges. PLoS ONE 5, e12110 (2010).
21. Henry, P.-Y. et al. Integrating ongoing biodiversity monitoring: potential benefits and methods. Biodiv. Cons. 17, 3357–3382 (2008).
22. Hoffmann, A. et al. The need for an integrated biodiversity policy support process – Building the European contribution to a global Biodiversity Observation Network (EU BON). Nat. Cons. 6, 49–65 (2014).
23. Guralnick, R. P., Hill, A. W. & Lane, M. Towards a collaborative, global infrastructure for biodiversity assessment. Ecol. Lett. 10, 663–672 (2007).
24. Pereira, H. M. et al. Essential Biodiversity Variables. Science 339, 277–278 (2013).
25. Brummitt, N. et al. Taking Stock of Nature: Essential Biodiversity Variables Explained. Biol. Cons. doi: 10.1016/j.biocon.2016.09.006 (2016).
26. Geijzendorffer, I. R. et al. Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. J. Appl. Ecol. 53, 1351–1350 (2015).
27. Steffen, W. et al. The trajectory of the Anthropocene: The Great Acceleration. The Anthropocene Rev. 2, 81–98 (2015).
28. De Vos, J. M. et al. Estimating the Normal Background Rate of Species Extinction. Cons. Biol. 29, 452–462 (2014).
29. Ceballos, G. et al. Accelerated modern human–induced species losses: Entering the sixth mass extinction Science Adv. 1, e1400253 (2015).
30. Steadman, D. W. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. Science 267, 1123–1131 (1995).
31. Barnosky, A. D. et al. Assessing the Causes of Late Pleistocene Extinctions on the Continents. Science 306, 70–75 (2004).
32. European Commision (EEC). Directive of the European Parliament and of the Council on the Conservation of wild birds (79/409/EEC, amended 2009: 2009/147/EC). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri= CELEX:32009L0147, (Date of access: 21/04/2016) (2009).
33. European Commision (EEC). Directive of the European Parliament and of the Council on the Conservation of natural habitats and of wild fauna and flora (92/43/EEC consolidated 2007). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri= CELEX:32009L0128,
(Date of access: 21/04/2016) (2007).
34. UN (United Nation). Convention on Wetlands of International Importance especially as Waterfowl Habitat. http://portal.unesco.org/en/ev.php-URL_ID= 15398&URL_DO= DO_TOPIC&URL_SECTION= 201.html, (Date of access: 21/04/2016) (1971).

35. Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. Citizen science as an ecological research tool: Challenges and benefits. Annual Rev.Ecol. Evol. Syst. 41, 149–172 (2010).
36. Mora, C., Tittensor, D. P. & Myers, R. A. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proc. R. Soc. B: Biol. Sci. 275, 149–155 (2008).
37. Bonebrake, T. C., Christensen, J., Boggs, C. L. & Ehrlich, P. R. Population decline assessment, historical baselines, and conservation. Cons. Lett. 3, 371–378 (2010).
38. Pimm, S. L. et al. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344, doi: 10.1126/science.1246752 (2014).
39. Houlahan, J. E. et al. Quantitative evidence for global amphibian population declines. Nature 404, 752–755 (2000).
40. May, R. M. Ecological science and tomorrow's world. Phil. Trans. R. Soc. B: Biol. Sci. 365, 41–47 (2010).
41. Butchart, S. H. M. et al. Global Biodiversity: Indicators of Recent Declines. Science 328, 1164–1168 (2010).
42. Regan, H. M., Colyvan, M. & Burgman, M. A. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol. Appl. 12, 618–628 (2002).
43. Rick, T. C. & Lockwood, R. Integrating Paleobiology, Archeology, and History to Inform Biological Conservation. Cons. Biol. 27, 45–54 (2013).
44. Clavero, M. & Revilla, E. Biodiversity data: Mine centuries-old citizen science. Nature 510, 35–35 (2014).
45. Edwards, J. L. Research and societal benefits of the Global Biodiversity Information Facility. Bioscience 54, 485–486 (2004).
46. Schmeller, D. S. et al. Advantages of Volunteer-Based Biodiversity Monitoring in Europe. Cons. Biol. 23, 307–316 (2009).
47. Bell, S. et al. What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. Biodiv. Cons. 17, 3443–3454 (2008).
48. NERC (Centre for Population Biology, Imperial College). The Global Population Dynamics Database Version 2. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html, (Date of access: 21/04/2016) (2010).
49. GEOBON, Essential Biodiversity Variables. http://geobon.org/essential-biodiversity-variables/ebv-classes-2/, (Date of access: 21/04/2016) (2016).
50. Faulwetter, S. et al. Polytraits: A database on biological traits of marine polychaetes. Biod. Data Journal 2, e1024 (2014).
51. YouTHERIA, Web portal containing data on the life-history, ecology and geography of all mammals. http://www.utheria.co.uk/home.aspx, (Date of access: 13/06/2015) (2010).
52. UN (United Nations). The World at Six Billion. http://www.un.org/esa/population/publications/sixbillion/sixbillion.htm, (1999) (Date of access: 19/03/2016).
53. European Environment Agency (EEA). European average air temperature anomalies (1850 to 2012) in °C over land areas only, for annual (upper), winter (middle) and summer (lower) periods. http://www.eea.europa.eu/data-and-maps/figures/annual-winterdecember-
january-february-and-summer-june-july-august-mean-temperature-deviations-in-europe-1860-2007-oc-the-linesrefer-to-10-year-moving-average-european-land-5, (2013) (Date of access: 19/03/2016).
54. Klein Goldewijk, K. & Ramankutty, N. Land cover change over the last three centuries due to human activities: The availability of new global data sets. GeoJournal 61, 335–344 (2004).

55. Klein Goldewijk, K. & Ramankutty, N. Land Use Changes During the Past 300 Years. In: Encyclopedia of Life Support Systems (EOLSS), Natural Resources Policy and Management - Land Use, Land Cover and Soil Sciences (eds Verheye, W.) (UNESCO, 2004).
56. Peñuelas, J. et al. Human-induced nitrogen–phosphorus imbalances alter natural and managed ecosystems across the globe. Nat. Comm. 4, doi: 10.1038/ncomms3934 (2013).
57. European Environment Agency (EEA). Atmospheric concentration of Carbon Dioxide, Methane and Nitrous Oxide. http://www.eea.europa.eu/data-and-maps/figures/atmospheric-concentration-of-co2-ppm-1, (2014) (Date of access: 19/03/2016).
58. Green, N. J. L., Jones, J. L. & Jones, K. C. PCDD/F Deposition Time Trend to Esthwaite Water, UK, and Its Relevance to Sources. Envir. Sci. Tech. 35, 2882–2888 (2001).
59. Conover, W. J. Practical Nonparametric Statistics. 3rd edition (Wiley, 1999).
60. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B (Methodological) 57, 289–300 (1995).
61. R Core Team R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (2013).

Author Contributions

J.B.M. and D.S.S. wrote the manuscript and performed the analysis. K.H., N.T., L.B. and N.B. provided decisive feedback and suggestions to improve the manuscript. All authors contributed to scientific discussions and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep
Competing financial interests: The authors declare no competing financial interests.

7.2. Measuring Rao's Q diversity index from remote sensing: An open source solution

Partners involved: FEM

Project leaders: Duccio Rocchini (FEM)

Manuscript title

Measuring Rao's Q diversity index from remote sensing: An open source solution [publishe in *Ecological Indicators*]

One sentence summary

We propose the use of Rao's Q applied to remotely sensed data, providing a straightforward R-package function to calculate it in 2D systems.

Authors

D. Rocchini[1], M. Marcantonio[1,2] and C. Ricotta[3]

Affiliations

[1] *Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, Via E. Mach 1, 38010 S. Michele all'Adige (TN), Italy*

[2] *Technische Universität Berlin, Department of Landscape Architecture and Environmental Planning, Geoinformation in Environmental Planning Lab, Berlin, Germany*

[3] *Department of Environmental Biology, University of Rome "La Sapienza", 00185 Rome, Italy*

*Correspondence to: duccio.rocchini@fmach.it

Abstract

Measuring biodiversity is a key issue in ecology to guarantee effective indicators of ecosystem health at different spatial and time scales. However, estimating biodiversity from field observations might present difficulties related to costs and time needed. Moreover, a continuous data update for biodiversity moni-toring purposes might be prohibitive. From this point of view, remote sensing represents a powerful tool since it allows to cover wide areas in a relatively low amount of time. One of the most common indicators of biodiversity is Shannon's entropy H_, which is strictly related to environmental heterogeneity, and thus to species diversity. However, Shannon's entropy might show drawbacks once applied to remote sensing data, since it considers relative abundances but it does not explicitly account for distances among pixels' numerical values. In this paper we propose the use of Rao's Q applied to remotely sensed data, providing a straightforward R-package function to calculate it in 2D systems. We will introduce the theoretical rationale behind Rao's index and then provide applied examples based on the proposed R function.

Introduction

Measuring biodiversity as an indicator of ecosystem healthhas been recognized by major initiatives worldwide, includ-ing the Group on Earth Observation (GEO BON, http://www.earthobservations.org/geobon.shtml) initiative, the InternationalGeosphere Biosphere Programme (IGBP, http://www.igbp.net/),the World Climate Research Programme (WCRP, http://wcrp-climate.org/), the Committee on Earth Observation Systems (CEOS)Biodiversity task (http://ceos.org/), among others.

However, estimating biodiversity from field data presents anumber of drawbacks mainly related to time and costs, togetherwith intrinsic difficulties to build standardized procedures forreproducible data gathering (Palmer et al., 2002).

For this purpose, using maps in a GIS environment orheterogeneity-related maps derived from remotely sensed imagery(e.g. Carranza et al., 2007) might help in finding hotspots of diversity over space and track their variation over time (Boyd and Foody,2011), from local (Feilhauer et al., 2013) to global (Rocchini et al., 2010) spatial scales. This is true under the light of the Spectral Vari-ation Hypothesis (Palmer et al., 2002) which states that the higherthe environmental heterogeneity, the higher will be the speciesdiversity of a certain area. The rationale under the Spectral Varia-tion Hypothesis is that a higher spatial variability (measured byspectral diversity from remotely sensed images) is related to ahigher amount of ecological niches for species living therein. Hence, measuring the heterogeneity of a landscape is critical since it isdirectly related to its diversity (Gillespie et al., 2008; Skidmore et al., 2015). Moreover, landscape diversity is related to the diversity atother ecosystem levels such as species diversity.

It has been demonstrated that the measure being used can leadto very different (and sometimes misleading) results. As an example, one of the mostly used diversity measures of the landscapebased on spectral remotely sensed data, i.e. the Shannon's entropy (Shannon, 1948), has a number of implicit drawbacks like: (i) thedifficulty to discriminate between differences in richness or relativeabundance (Nagendra, 2002) or (ii) the impossibility to considerspectral values as numbers instead of classes (Rocchini and Neteler, 2012b). Concerning the second point, Shannon's entropy accountsfor richness and relative abundance of spectral values but it doesnot explicitly consider the numerical magnitude (values) of pixels.

The aim of this paper is to solve the aforementioned issue, bythe application of Rao's Q to remotely sensed data, providing a straightforward R function to calculate it in 2D systems. We willfirst introduce the theoretical rationale behind Rao's index and thenprovide applied examples based on the proposed R function. As faras we know, this is the first attempt to measure Rao's Q in a 2D space applied to remotely sensed data.

Theory under the use of the Rao's Q index

Methods for measuring landscape diversity have mostly reliedon the classification of remotely sensed image. However, imageclassification has several drawbacks which should be seriouslytaken into account, e.g.: (i) the accuracy assessment should be per-formed in a robust manner, thus requiring time and costs overallwhen field assessment is involved (Foody, 2002), (ii) it is dif-ficult to build practically sound accuracy assessment protocols (Foody, 2008), (iii) the classification should be performed only byrobust algorithms avoiding as much as possible manual digitization(Burnett and Blaschke, 2003), (iv) several issues have to be bypassedwhen choosing pure training samples in order to avoid mixingeffects (Small, 2004).

Besides these technical shortcomings, classi-fication is a subjective task in its very nature and it inevitably leadsto the degradation of continuous information (Palmer et al., 2002).

Rocchini et al. (2010) summarize several approaches to measureecosystem diversity from remotely sensed images, mainly basedon the continuous variability of pixel values (e.g. original digitalnumbers of a satellite image). Such approaches show their fullpower when relying on Free and Open Source algorithms. OpenSource algorithms allow indeed robustness and reproducibilitythanks to the public availability of the used code (Rocchini andNeteler, 2012a).

Among the most spread diversity indexes used in ecology thereis the Shannon entropy index (H', Shannon, 1948). This index canbe easily applied to remote sensed data. Given a certain numberof reflectance values in a remotely sensed image, also referred toas digital numbers, H' can be calculated as $H' = -\sum p_i \times \log(p_i)$. Inthis particular application H' takes into account the relative pro-portion p of each reflectance value i. Generalizing, H' considers theequitability of the system. Furthermore, when transformed to the Pielou evenness index J (Pielou, 1969), calculated as J = H'/Hmax, itshows the maximum possible diversity within the same numberof reflectance values. Quoting Ricotta and Avena (2003), who provided an elegant mathematical dissertation about the Pielou index applied to both species and landscape classes:

> "The normalization of H' with respect to maximum entropy ($J = H'/H_{max}$) is termed 'evenness' because it measures deviationfrom an even distribution of individuals amongst the N species".

Translating the sentence from species to spectral diversity, it turnsout to be:

> "The normalization of H' with respect to maximum entropy($J = H' /H_{max}$) is termed 'evenness' because it measures deviationfrom an even distribution of individual pixels amongst the N reflectance values."

However, Shannon and Pielou indices only rely on the relativeabundance of reflectance, not considering the numerical value of reflectance per se. Facing the problem from a mathematical pointof view, let M be an image of 3 × 3 pixels (indicated by c to avoidconfusion with p, used in this manuscript to indicate the proportionof area of each category):

$$M = \begin{pmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{pmatrix} \quad (1)$$

Let i and j be two different pixel values, e.g. two Digital Numbers (DNs) of a 8 bit image with i /= j, as:

$$M = \begin{pmatrix} i & i & j \\ i & j & j \\ j & j & j \end{pmatrix} \quad (2)$$

In this case H =−∑p × lnp =− (3/9 × ln(3/9) + 6/9 × ln(6/9)) = 0.637. Shannon entropy does not take into account the value of i andj but just the proportion of i and j values. Therefore, it does notdiscriminate among different contexts such as, (a) i = 1 and j = 200 or (b) i = 201 and j = 200.

On the contrary, Rao's Q index does take into account i and j value by considering their pairwaise distance d_{ij}:

$$Q = \sum \sum d_{ij} \times p_i \times p_j \quad (3)$$

As an example, in case of (a) Q = 88.444, while in case of (b) Q = 0.444. As a consequence, deriving Rao's Q involves calculating a distance matrix M_d for all the pixel values:

$$M_d = \begin{pmatrix} d_{1,1} & d_{1,2} & d_{1,3} & \cdots & d_{1,n} \\ d_{2,1} & d_{2,2} & d_{2,3} & \cdots & d_{2,n} \\ d_{3,1} & d_{3,2} & d_{3,3} & \cdots & d_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & d_{n,3} & \cdots & d_{n,n} \end{pmatrix} \quad (4)$$

or more simply $\overbrace{123N}^{\text{}}$, when N pixels are considered (see also Rocchini (2007) on distance matrices in a spectral space). Thus,Rao's Q is related to the sum of all the pixel values pairwise dis-tances, each of which is multiplied by the relative abundance ofeach pair of pixels in the analysed image d × ($1/N^2$). In other words, Rao's Q is the expected difference in reflectance values betweentwo pixels drawn randomly with replacement from the consideredevaluated pixels set. The distance matrix can be built in severaldimensions (layers), thus allowing to consider more than one bandat a time. As a consequence Rao's Q can be calculated in a multidi-mensional (multi-layers) system.

In remote sensing applications the derivation of syntheticindexes of any sort (i.e., diversity) is often performed consideringsmall chunks of the whole image per time, commonly defined as'windows' or 'moving windows'. From now on, we will use thisterminology to indicate the local space of analysis.

Coding Rao's Q in R

The function spectralrao() to derive Rao's Q, written in the Rstatistical language (R Core Team, 2016), is reported in Appendix 1and stored in the GitHub repository https://github.com/mattmar/spectralrao. The function accepts matrix, RasterLayer or Spatial-GridDataFrame object as input (or a list of them). It can be ran withtwo different settings, using (i) a single matrix (mode="classic") or (ii) more matrices (mode="multidimension") as input. Distancecan be calculated relying on Euclidean, Manhattan and Canberradistances by the distance_m parameter. Appendix 2 provides acomplete description of such distances, with their advantages anddisadvantages, together with proper reference to previous ecolog-ical papers using them. Further, a user-defined distance matrix canbe also provided through the function argument distance_m.

In this manner, it is possible to obtain H' as output, together with Rao's Q, setting the option shannon=TRUE. On the other hand, if mode="multidimensional", a list of matrices must be providedas input. The overall distance matrix is thus calculated in a multi- or hyper-dimensional system by using the previously stated measuresthrough the function argument distance m. Each distance is thenmultiplied by the inverse of the squared number of pixels in theconsidered moving window (as in Eq. (3)), and the Rao's Q is finallyderived by applying the summation (see Eq. (3)).
In the following section, we describe a step-by-step workflow to derive Rao's Q for both modes, using simulated or spectral-likereal matrices as input data.

Firstly, the R function can be loaded by relying on the source file available in Appendix 1, as:

```
source("yourpath/spectralrao.r")
```

Synthetic data, i.e. two matrices r1 and r2, can be generated by:

```
###Clustered simulated spectral matrices r1
xy1 <- matrix(rnorm(25, 0, .25), ncol=5,nrow=5)
xy2 <- matrix(rnorm(25, -.5, .1), ncol=5,nrow=5)
xy3 <- matrix(rnorm(25, .5, .1), ncol=5,nrow=5)
xy4 <- matrix(rnorm(25, .8, .25), ncol=5,nrow=5)
r1 <- cbind(rbind(xy1, xy2),rbind(xy3,xy4))
###Clustered simulated spectral matrices r2
xy11 <- matrix(rnorm(20, 4.5, .01), ncol=5,nrow=4)
xy21 <- matrix(rnorm(30, 5.5, .1), ncol=5,nrow=6)
xy31 <- matrix(rnorm(40, 6.5, .1), ncol=5,nrow=8)
xy41 <- matrix(rnorm(10, 7.0, .05), ncol=5,nrow=2)
r2 <- t(cbind(rbind(xy11, xy21),rbind(xy31,xy41)))
```

Using (i) r1 as input data, (ii) the Euclidean distance as the metric to calculate the distance matrix and (iii) an operational moving window of 3 × 3 pixels, Rao's Q can be derived as:

```
raomatrix <- spectralrao(r1, mode="classic",
distance_m="euclidean", window=3, shannon=T)
```

For this particular set of data (r1), H' is of low applicability (Fig. 1), due to the high heterogeneity in the input data. On the contrary, Rao's Q meaningfully highlighted the areas with higher heterogeneity, the intersection between the simulated submatri-ces.

In a slightly different way, using both r1 and r2 as input data, we can derive the multidimensional form of Rao's Q:

```
raomatrix <- spectralrao(list(r1,r2),
mode="multidimension", distance_m="euclidean", window=3,
shannon=F)
```

When visualizing this example in Fig. 2, it is evident how Rao's Q, interpreted in its multidimensional meaning, sintetizes the con-trasting signal from two different datasets in a single index.

Additional arguments implemented in the R code (Appendix 1) are related to the tolerance of retaining NAs at the border of theimages by using the argument na.tolerance and the rescaling and centering of input data, relying on the (true or false) argument rescale based on commonly used mean and standard deviation subtraction techniques (see Appendix 1).

The function accepts also R "spatial objects" as input data. Toillustrate this application, in the following chunk of code a June2015 MODIS Normalized Difference Vegetation Index (NDVI) imageat 0.1 degrees resolution was downloaded from the Nasa Earth Observation dataset for Europe and Rao's Q was calculated.

Downloading the raster file using we get:

```
cd ~
wget -O example_modis_ndvi_2015.tiff
http://neo.sci.gsfc.nasa.gov/servlet/RenderData?
si=1690249&cs=rgb&format=TIFF&width=3600&height=1800
```

Setting the no data pixels as NAs and calculating Rao's Q:

```
ndvi2015 <- raster("~/example_modis_ndvi_2015.tiff")
ndvi2015[ndvi2015==255]<-NA
raomatrix <- spectralrao(ndvi2015, distance_m="euclidean",
window=9, shannon=T)
```

In Fig. 3, it is apparent that H' tends to saturate in case of high diversity since in the local 9 × 9 pixels window of analysis all the pixel values, even though similar among them, are still different. As a consequence, since H_ does not take into account their distance but only their relative abundance, its value will always approximate saturation. On the contrary, Rao's Q overcomes this limitation by the pairwaise distance term.

The output of spectralrao() R function is a list of objects. The output list has dimension 1 if shannon=FALSE or if mode="multidimensional", or dimension 2 if shannon=TRUE and mode="classic". If RasterLayer or SpatialGridDataFrame R objects are provided as input, the function output will be a list of Raster-Layer object(s).

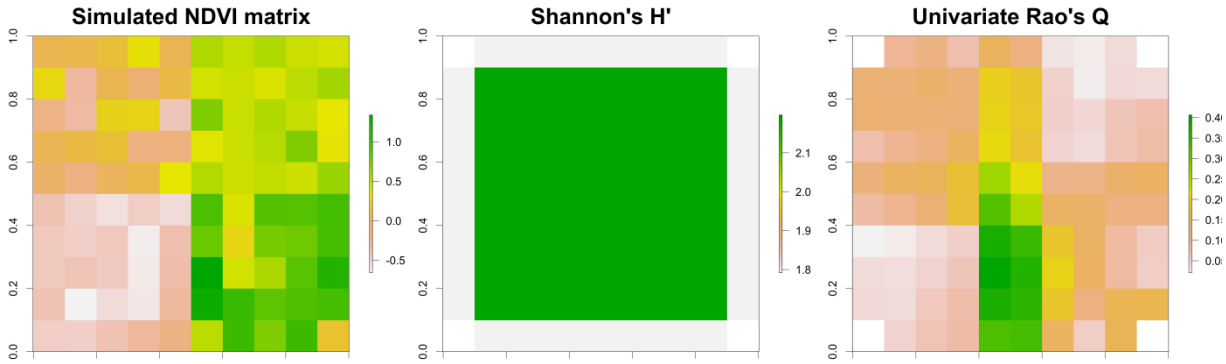


Figure 1 *An example of the calculation of Rao and Shannon indices on a hypothetical NDVI image. In this case, Shannon index tends to overestimate diversity since it considers the differences in the abundance of classes, while Rao Q seems to be more reliable taking into account their distance.*

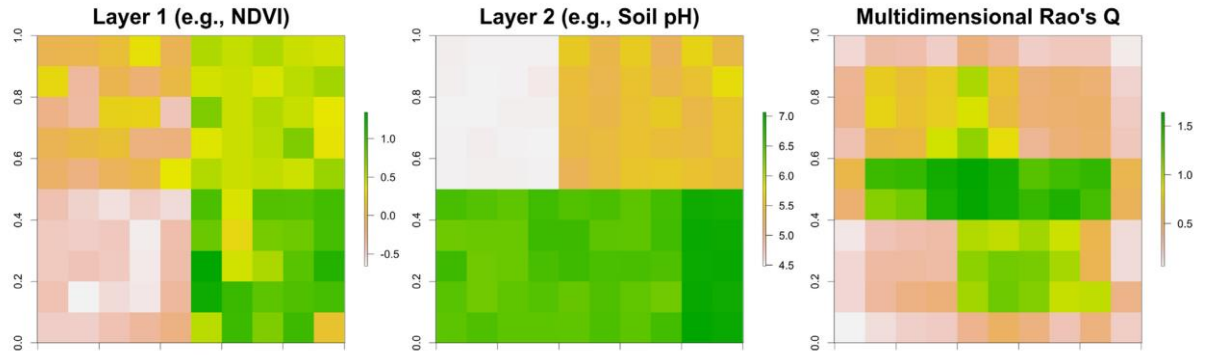


Figure 2 *Calculation of Rao's Q in a multidimensional set. Refer to the main text for additional information.*

Discussion

In this paper, we demonstrated the potential advantages ofapplying Rao's Q to calculate diversity in digital imagery, thus avoiding the non-dimensionality of other more common indices like the Shannon's index.

Dealing with digital images, an advantage of Rao's Q over moreconventional diversity measures is that, while the calculation of H' relies solely on the relative proportion of the digital numbers (DN), Q takes also into account their pairwise differences. Moreover, while H' is usually calculated on one single band at a time, Rao's Qcan accommodate multivariate differences between DNs. Indeed, it can be calculated on multiple bands, hence representing the DNsdispersion in a multivariate space.

Rao's Q has been extensively used in functional diversity appli-cation (Botta-Dukát, 2005; Ricotta and Moretti, 2011; Marcantonio et al., 2014). Functional ecologists make use of a wide set offunctional traits (plants functional characteristics) to assess thediversity of natural systems. Rao's Q has been shown to be avalid candidate to summarize them in a single diversity value (Botta-Dukát, 2005). However, as previously stated, this is the firstapplication of the Rao's Q in a 2D space with remotely sensed data.

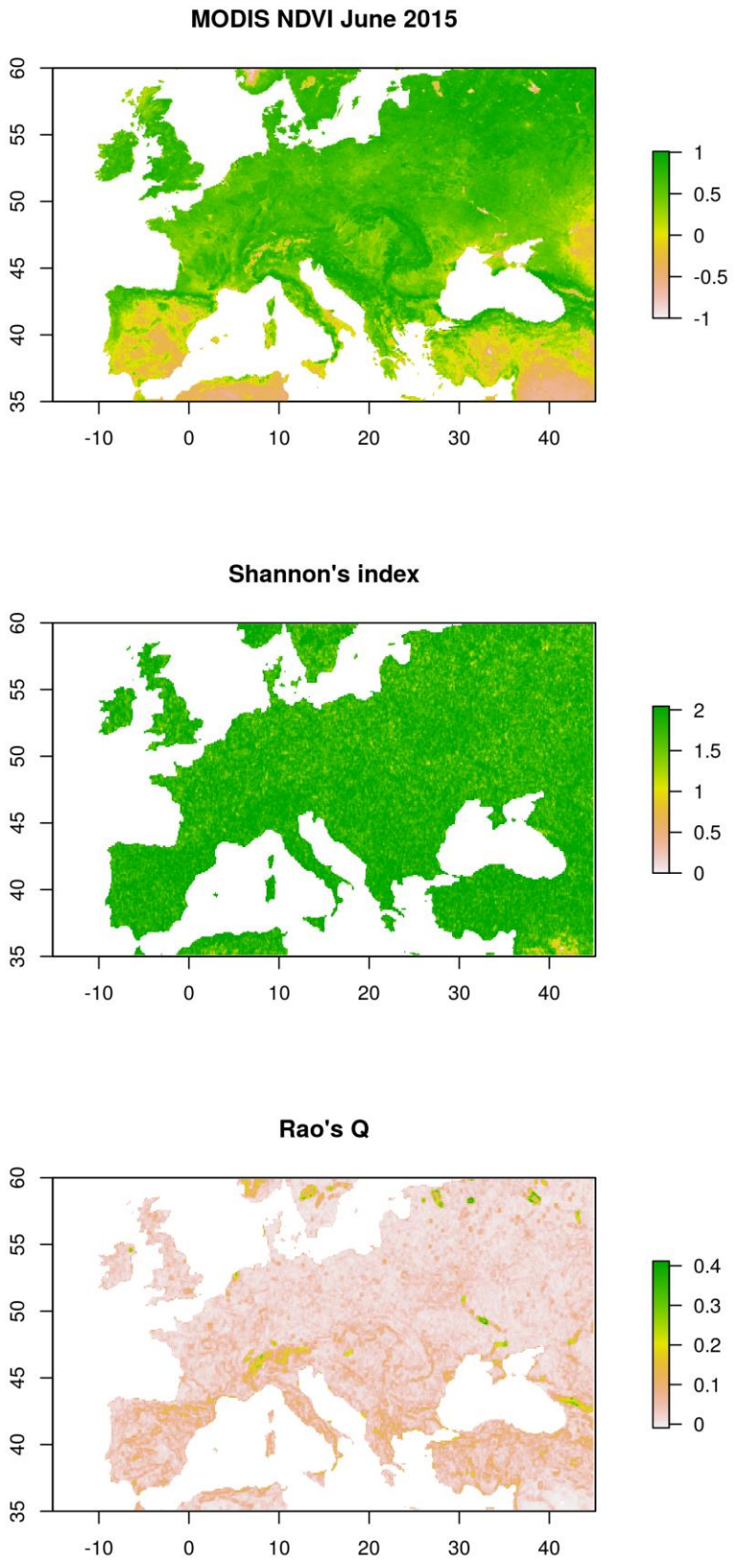


Figure 3 *In this MODIS Normalized Difference Vegetation Index (NDVI) image at 0.1 degrees resolution of June 2015, the Shannon and the Rao indices are calculated. While Shannon tends to saturate towards higher values, Rao's Q is not affected by small differences between pairs of pixel values.*

Note here that if, for a single band, Rao's Q is calculated usinghalf the squared Euclidean distance $d_{ij}= \frac{1}{2} (i - j)^2$ the resulting index $\frac{1}{2} \sum\sum p_i \times p_j \times (i - j)^2$ reduces to the well-known formulaof variance (expressed as the average squared difference amongDN values; see Pavoine (2012)), which is routinely used in remotesensing for summarizing the spatial complexity of digital images (Rocchini et al., 2010). Accordingly, Rao' Q can be interpreted as a multivariate generalization of the variance of a quantitative vari-able such as the DNs of a spectral band, thus bridging the gapbetween remotely sensed measures of diversity and spatial complexity.

Note also that in principle, apart from the Euclidean distance, Rao' Q can be calculated with a plethora of different multivariatemeasures of dissimilarity (see Podani, 2000 for review) that may beselected according to the specific users' needs. In this framework, due to the additive property of Q for which if $d_{ij}= d_{ij(1)} + d_{ij(2)}$ then$Q = Q_{(1)}+ Q_{(2)}$, it is possible to calculate a number of multivariateindices of Rao' Q based on different combinations of single-bandindices. For example, given two bands U and V, with the proposed R code we can calculate the Rao's Q for both bands separately. Inthis one-dimensional case, the distance used is simply the abso-lute difference between the DNs of each band. The resulting indexvalues $Q_{(U)}$ and $Q_{(V)}$ can be then additively recombined into one single index as $Q = Q_{(U)} + Q_{(V)}$. This is tantamount saying that, first,the univariate distances associated to the single bands are recom-bined into the multivariate Manhattan or city-block distance (see Podani, 2000) such that $d_{ij}= d_{ij(U)} + d_{ij(V)}$, and next the Rao index Q iscalculated directly from the multivariate Manhattan distances d_{ij}.

If the Manhattan distance is divided by the number of variables such that $d_{ij} = \frac{1}{2}d_{ij(U)} + \frac{1}{2}d_{ij(V)}$ and hence $Q = \frac{1}{2}Q_{(U)} + \frac{1}{2}Q_{(V)}$, the so-called mean character difference or Czekanowski dissimilar-ity is obtained. Finally, if the simple average is substituted by theweighted average $d_{ij} = w_{(U)}d_{ij(U)} + w_{(V)}d_{ij(V)}$ (i.e. $Q = w_{(U)}Q_{(U)} +w_{(V)}Q_{(V)}$ with $0 \leq w \leq 1$ and $\sum w = 1$) we obtain a highly flexible generalization of the Czekanowski index (Pavoine et al., 2009) in which the weights w of single bands can be determined accordingto the reflectance properties of single bands or to the specific user'srequirements. For example, the weights can be set proportional tothe range of DN values in each band.

Due to its flexibility, Rao's Q based on the aforementioned mul-tivariate distances may be helpful to optimize the relationshipbetween biodiversity values recorded from remote sensors andspecies inventories recorded from field observations.

Acknowledgements

We are grateful to the handling Editor Prof. Giovanni Zurlini andtwo anonymous referees for comments on a previous version of thismanuscript. DR was partially funded by: (i) the EU BON (Building theEuropean Biodiversity Observation Network) project, funded bythe European Union under the 7th Framework programme, Contract No. 308454, (ii) by the ERA-Net BiodivERsA, with thenational funders ANR, BelSPO and DFG, part of the 2012–2013 BiodivERsA call for research proposals, (iii) the EU-LIFE projectLIFE14ENV/IT/000514 FutureForCoppices.The PhD Scholarship of MM is supported by FIRS > T (FEM International Research School Trentino).

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2016.07.039.

References

Botta-Dukát, Z., 2005. Rao's quadratic entropy as a measure of functional diversitybased on multiple traits. J. Veg. Sci. 16, 533–540, http://dx.doi.org/10.1111/j.1654-1103.2005.tb02393.x.

Boyd, D.S., Foody, G.M., 2011. An overview of recent remote sensing and GIS basedresearch in ecological informatics. Ecol. Inf. 6, 25–36.

Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationshipmodelling methodology for landscape analysis. Ecol. Modell. 168, 233–249.

Carranza, M.L., Acosta, A., Ricotta, C., 2007. Analyzing landscape diversity in time:the use of Rènyi's generalized entropy function. Ecol. Indic. 7, 505–510.

Feilhauer, H., Thonfeld, F., Faude, U., He, K.S., Rocchini, D., Schmidtlein, S., 2013.Assessing floristic composition with multispectral sensors – a comparisonbased on monotemporal and multiseasonal field spectra. Int. J. Appl. Earth Obser. Geoinf. 21, 218–229.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. RemoteSens. Environ. 80, 185–201.

Foody, G.M., 2008. Harshness in image classification accuracy assessment. Int. J.Remote Sens. 29, 3137–3158.

Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuringand modelling biodiversity from space. Progress Phys. Geogr. 32, 203–221.

Marcantonio, M., Rocchini, D., Ottaviani, G., 2014. Impact of alien species on dunesystems: a multifaceted approach. Biodiv. Conserv., 1–24, http://dx.doi.org/10.1007/s10531-014-0742-2.

Nagendra, H., 2002. Opposite trends in response for the Shannon and Simpsonindices of landscape diversity. Appl. Geogr. 22, 175–186.

Palmer, M.W., Earls, P., Hoagland, B.W., White, P.S., Wohlgemuth, T., 2002.Quantitative tools for perfecting species lists. Environmetrics 13, 121–137.

Pavoine, S., 2012. Clarifying and developing analyses of biodiversity: towards ageneralisation of current approaches. Methods Ecol. Evol. 3, 509–518.

Pavoine, S., Vallet, J., Dufour, A.B., Gachet, S., Daniel, H., 2009. On the challenge oftreating various types of variables: application for improving themeasurement of functional diversity. Oikos 118, 391–402.

Pielou, E.C., 1969. An Introduction to Mathematical Ecology. John Wiley, New York.Podani, J., 2000. Introduction to the Exploration of Multivariate Biological Data.Backhuys, Leiden, Freely available at: http://ramet.elte.hu/podani/books.html.

R Core Team, 2016. R: A Language and Environment for Statistical Computing.Foundation for Statistical Computing, Vienna, Austria.

Ricotta, C., Avena, G., 2003. On the relationship between Pielou's evenness andlandscape dominance within the context of Hill's diversity profiles. Ecol. Indic, 361–365.

Ricotta, C., Moretti, M., 2011. CWM and Rao's quadratic diversity: a unifiedframework for functional ecology. Oecologia 167, 181–188.

Rocchini, D., 2007. Distance decay in spectral space in analysing ecosystem _-diversity. Int. J. Remote Sens. 28, 2635–2644.Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark,S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C.,Southworth, J., Neteler, M., 2010. Remotely sensed

spectral heterogeneity as aproxy of species diversity: recent advances and open challenges. Ecol. Inform.5, 318–329.

Rocchini, D., Neteler, M., 2012a. Let the four freedoms paradigm apply to ecology.Trends Ecol. Evol. 27, 310–311.

Rocchini, D., Neteler, M., 2012b. Spectral rank-abundance for measuring landscapediversity. Int. J. Remote Sens. 33, 4458–4470.Shannon, C., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27,379–423.

Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R.,Müncher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., Schaepman, M.E.,Turner, W., Wang, T.J., Wegmann, M., 2015. Environmental science: agree onbiodiversity metrics to track from space: comment. Nature 523, 403–405.

Small, C., 2004. The Landsat ETM+ spectral mixing space. Remote Sens. Environ. 93, 1–17.

7.3. Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis

Partner involved: UFZ

Project leader: Gudrun Carl and Ingolf Kühn (UFZ)

Manuscript title

Assessing relative variable importance across different spatial scales: a two-dimensional wavelet analysis [published in Journal of Biogeography]

One sentence summary

We present a method for applying two-dimensional wavelet analysis to a generalized linear model and the obtained results demonstrate that it is particularly suitable for studying scale dependencies of various spatial processes on environmental drivers keeping grain and extent constant and changing resolution.

Authors

Gudrun Carl[1]*, Daniel Doktor[2], Oliver Schweiger[1], Ingolf Kühn[1,3,4]

Affiliations

[1] *Helmholtz Centre for Environmental Research - UFZ, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany*

[2] *Helmholtz Centre for Environmental Research - UFZ, Department of Computational Landscape Ecology, Permoserstr. 15, 04318 Leipzig, Germany*

[3] *Martin Luther University Halle-Wittenberg, Geobotany & Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany*

[4] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*

*Correspondence to: gudrun.carl@ufz.de

Abstract

Aim Assessing the relationship between a spatial process and environmental variables as a function of spatial scale is a challenging problem. Therefore, there is a need for a valid and reliable tool to examine and evaluate scale dependencies in biogeography, macroecology and other earth sciences.

Location Central Europe (latitude 43.99°- 54.22° N, longitude 4.79°- 15.02° E).

Methods We present a method for applying two-dimensional wavelet analysis to a generalized linear model. This scale-specific regression is combined with a multimodel inference approach evaluating the relative importance of several environmental variables across different spatial scales. We apply this method to data of climate, topographic and land cover variables to explain variation in annual greening of vegetation (i.e. phenology) in Central Europe.

Results Land use is more important to explain the variation in greening than climate at smaller resolution while climate is more important at larger resolution with a shift at approx. 1000 km^2.

Main conclusions To the best of our knowledge, this is the first study analysing the scale dependency of an ecosystem process, clearly distinguishing between the different components of scale, namely grain, focus and extent. The obtained results demonstrate that our newly proposed method is particularly suitable for studying scale dependencies of various spatial processes on environmental drivers keeping grain and extent constant and changing focus (i.e. resolution).

Keywords Discrete wavelet transform, generalized linear model, multimodel inference, remote sensing signal, spatial scales, vegetation period

Introduction

The importance of spatial patterns and spatial scales has often been cited as a key issue in biogeography, macroecology, and beyond that, all earth sciences (Levin, 1992; Dale, 1999; Wu & Hobbs, 2002; Fortin & Dale, 2005; Schröder & Seppelt, 2006). Data collection for biogeographic and environmental data is frequently carried out with reference to a gridded map of a specific resolution. A statistical model based on these data will provide statistical inferences at this specific spatial scale. Because different (e.g., biological) processes act at different scales, multiple relationships are scale-specific as well (Pearson & Dawson, 2003; Pearson *et al.*, 2004; Guisan & Thuiller, 2005; Keil *et al.*, 2012). Hence, the selection of scale for data collection and inference is crucial in statistical modelling. In general, however, different scales will be relevant in such multiple relationships and some of them will be different from the pre-specified collection unit (i.e. focus and extent of analysis *sensu* Scheiner *et al.* (2000) ; see below). As a consequence, conclusions based on regressions of these data, i.e. its parameter estimates, hypotheses tests and *P*-values, may be misleading and can result in incorrect inferences. At least, this is the case if we ignore that these conclusions are restricted to a particular scale and disregard the complexity and multi-scaled structure of the problem. Therefore, there is a need for a valid and reliable tool to examine and evaluate scale dependencies (Wu & Hobbs, 2002; Borcard *et al.*, 2004). The principal coordinates of neighbour matrices (PCNM) analysis (e.g., Borcard *et al.*, 2004) as well as its generalization, the Moran's eigenvector maps (MEM) approach (Dray *et al.*, 2006), provide a spectral

decomposition of the spatial relationships. Both methods have in common that the eigenvectors used for spatial filtering purposes are sinusoidal waves of increasing frequency in case of regular sampling. Hence, the methods are basically statistical versions of Fourier analysis. Accordingly, *n-1* eigenvectors are generally needed in order to decompose a centred series of *n* observations and to completely capture its variability. In the general case of a large number *n*, however, the question arises as to how to identify the main scales of spatial variation and to select appropriate subsets of eigenvectors (Jombart *et al.*, 2009).

To overcome these problems, recent publications recommend the use of wavelet transforms as a tool for scale-specific regression analysis (Dong *et al.*, 2008; Lookingbill *et al.*, 2011), which is expected to be useful to identify scale-specific relationships between predictor and response variables, and thus to provide deeper insights in multiple scale dependencies. Wavelet analysis is an extension and refinement of Fourier analysis (Percival & Walden, 2000; Stark, 2005). Like Fourier analysis, it can be used to detect scale-specific (or frequency-specific) features of a signal. However, different from Fourier analysis, it is able to make any necessary local adjustments, providing different coefficients for different positions (or times). In fact, a Fourier analysis provides frequency or scale components whose amplitudes are the same at all positions (or times), whereas a wavelet analysis is able to provide additional information about which component is present at which spatial (or time) interval. This is because Fourier analyses are based on sinusoidal waves, whereas wavelet analyses use so-called wavelets, i.e. small waves visualizable as localized oscillations (Daubechies, 1992; Torrence & Compo, 1998; Cazelles *et al.*, 2008). Due to their much better local adaptation, such a wavelet analysis requires only a few frequency components to completely capture the variability of a signal.

Different approaches have been proposed for applying wavelet transforms to multiple linear regressions. On the one hand, Keitt & Urban (2005) developed a scale-dependent regression and found evidence for scale-specific relationships and inferences regarding predictor variables and the response variable. However, their approach is limited in its application to one-dimensional data analyses and response vectors of Gaussian distribution. On the other hand, wavelets have been used to remove spatial autocorrelation in multiple regressions affected by correlated errors (Carl & Kühn, 2008, 2010). Our method (Carl & Kühn, 2010) allows regular two-dimensional (2-D) sampling grids as well as different distributions (e.g., binomial or Poisson). In both cases, it has proved fruitful to insert wavelet transforms into the regression analysis of spatial data. Most recently, Ma and Zhang (2015) as well as Ye *et al.* (2015) followed the idea and performed a regression analysis using 2-D wavelet transforms to describe scale-specific patterns. Their results have demonstrated that such regressions are appropriate tools for exploring spatial variations at multiple spatial scales. However, the approaches described by Ma and Zhang (2015) and Ye *et al.* (2015) are only applicable to Gaussian response models and therefore exclude, e.g., logistic regressions. Moreover, the fact alone that different slopes at different scales can be discovered by means of wavelets is not sufficient. Ma and Zhang (2015) ranked explanatory variables at a given spatial scale in terms of the magnitude of the standardized coefficients. Ranking without any rank order weights is, however, a rather poor method. Instead, the regression analysis should be followed by any assessment, i.e. the calculation of an appropriate index for the strength of evidence. Also, Ma and Zhang (2015) scaled species richness. Since species richness does not scale additively and cannot be averaged, wavelets are inappropriate to scale such data. Therefore, the major objective of this study is to combine the advantages of the previous methods and to develop a two-dimensional wavelet regression applicable to various distributions. Moreover, our wavelet multiresolution regression will lead to scale-dependent inferences by means of rank order weights.

Analysing scale dependency, one has to be very clear about the four different components of scale (Scheiner *et al.*, 2000): (i) sample unit, (ii) grain, (iii) focus and (iv) extent. Sample unit refers to the spatial dimension of the collection unit (e.g. sampling plot). Grain is the smallest unit to which all sample units are standardized for a specific analysis (i.e. finest resolution). The units of grain can then be aggregated to coarser units of analysis, that is, focus (i.e. coarser resolution). Extent in this context is the complete geographic area sampled. The main advantage of scale-specific wavelet regression is that it differs from previous methods, which simply upscale data by averaging of aggregated cells and thus regress variables of enlarged grain size. Instead, wavelet analysis is able to extract scale-specific variations of both dependent and independent variables. Therefore, a wavelet regression can measure how a change in environmental variables at a given resolution (i.e. focus) influences change in the response variable at the same resolution (Ye *et al.*, 2015). To illustrate our new up-scaling method, it is necessary to use data at medium to large extent and fine sample unit because sample unit acts as a preset for the grain (i.e. finest resolution) in the analysis. Scale dependency is then studied by leaving extent and grain constant and aggregating $2^j \times 2^j$ (with *j* being a level of analysis) grains to coarser resolutions (i.e. foci). Hence, to discuss scale dependency, i.e. alterations in the relative importance of different environmental factors caused by increasingly coarser resolutions (foci), we need data collected over a regular grid consisting of sufficient grid cells. Therefore, in our case study, we examine data sampled on a map of 1024×1024 grid cells at $0.01° \times 0.01°$ resolution (grain), i.e. approximately 1×1 km^2, in Central Europe. We focus on the relationship of vegetation greening to climate, topography and land use. Remote-sensing vegetation indices based on satellite observations indicating the vegetation activity (Yang *et al.*, 2012) were used to estimate the vegetation period per year (White *et al.*, 2003). We use vegetation period as a response variable that is regressed on climate, topographic and land use data. To the best of our knowledge, this is the first study at intermediate extent and fine grain (and hence large sample size), which covers a very large range of different foci, clearly differentiating among the different components of scale. In principle, it can be assumed that: (i) The impact of climate, topographic as well as land use variables on vegetation period is scale-dependent. (ii) Land use variables are more relevant in relation to vegetation period than climate and topographic ones for models with fine resolution, and vice versa for those with coarse resolution (cf. Pearson *et al.*, 2004). Hence, our goal is to demonstrate how variations or fluctuations at multiple spatial scales can be systematically analysed, and to draw specific conclusions regarding these assumptions.

Methods

Wavelets

The crucial idea behind wavelet analysis can be formulated as follows: wavelets are small waves, i.e. localized oscillating functions (Daubechies, 1992; Ma & Zhang, 2015). In a one-dimensional spatial context, one can imagine that such a brief oscillation is locally aligned with a segment of the given transect, thus enabling a comparison between template (wavelet) and original (transect). If there is high similarity, then the absolute value of the corresponding wavelet coefficient is high. If there is low similarity, it is low (Dale & Mah, 1998; Csillag & Kabos, 2002; Ye *et al.*, 2015). By translation, i.e. shifting the wavelet along the transect line, one is able to stepwise evaluate the whole transect. Moreover, this transect can be scanned several times by gradually stretched or compressed wavelets, and thus varying in width and oscillating behaviour, which corresponds to scale or resolution. Based on a set of wavelets derived from a prototype (i.e. mother wavelet) and generated by scaling and translation of this original, it is possible to capture the complete information of any transect. Such wavelets of

different dilations and locations and their associated scaling functions constitute the so-called wavelet family (Daubechies, 1992). Each wavelet acts as both a window and a filter. One can show that the information of any discrete function *f* is codable by a wavelet transform, i.e. it can be captured by coefficients belonging to a certain wavelet family. If the used wavelet family is a family of orthogonal wavelets, then there exists a minimal set of wavelets, enabling a complete information transfer (Mallat, 1989; Percival & Walden, 2000). The number and kind of coefficients in discrete wavelet transforms (DWT) depends on the number and kind of wavelets used in the analysis, and thus not only on the number of observations, but also on a pre-specified number of resolution levels (Bruce & Gao, 1996). There are two kinds of coefficients: detail and smooth ones, reflecting different oscillating behaviour of mother wavelets and scaling functions and representing the highly-varying (detailed) and slowly-varying (smooth) parts of function *f*, respectively (Bruce & Gao, 1996; Ma & Zhang, 2015). Subsequently, it is possible to reconstruct the original function *f* by applying the back transform, i.e. the inverse wavelet transform. Moreover, by means of wavelet transform and back transform, one is able to decompose a function into orthogonal components at different scales. These components can be visualized as parts of the function at different resolutions. Therefore, this method is called multiresolution analysis (MRA) (Mallat, 1989; Dong *et al.*, 2008). The MRA algorithm always provides detail components at levels gradually incremented up to a preset limit (D_1, D_2, ..., D_J) and one smooth component at the upper level (S_J).

For illustration, we present the results of a wavelet MRA decomposition stopped at level 3 (Fig. 1b) in comparison to a PCNM analysis limited to four components (Fig. 1a). Both analyses are performed on the same signal vector *f*, which is a time series (or spatial transect). The example illustrates that PCNM and wavelet decompositions differ in their ability to detect local variations. Only in case of wavelet analysis, the components reveal that signal variability increases with time (or spatial variable). As a consequence, this method yields a perfect reconstruction of the signal from just four components. In general, wavelet analysis is locally more accurate compared to Fourier analysis, which requires many more components.

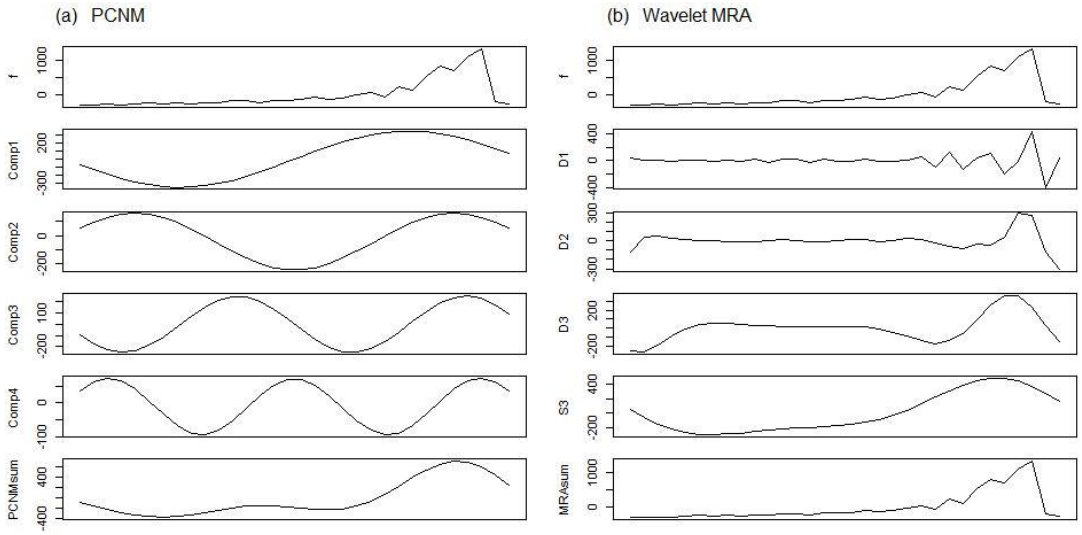


Figure 1: *Comparison of signal decomposition and reconstruction obtained for two different methods and constructed for a time series (or spatial transect) containing 32 observations. (a) PCNM is used to decompose the signal f (top panel) into its first four components (mid panels) and to reconstruct the signal by the sum of these four components (lower panel). (b) Wavelet MRA is used to decompose the same signal f (top panel) into four components (mid panels) and to reconstruct the signal by the sum of these four components (lower panel).*

Two-dimensional wavelet analysis

The use of wavelets in the fields of geophysics, biology, ecology and agriculture is rapidly developing (Kumar & Foufoula-Georgiou, 1997; Torrence & Compo, 1998; Dong *et al.*, 2008). However, most of the concepts for wavelet analysis apply to either time signals (Cazelles *et al.*, 2008) or one-dimensional spatial data, which are much like time series (Dale & Mah, 1998). The need to explore and assess images and landscapes requires a more comprehensive, two-dimensional wavelet analysis (Csillag & Kabos, 2002). The 2-D DWT enables us to transform a data matrix into a matrix of wavelet coefficients. Therefore, 2-D wavelet analysis allows us to analyse data such as discrete images or geographical patterns of ecological or environmental variables (Csillag & Kabos, 2002). Note that the increased dimensionality results in newly formed wavelets. Four shapes are formed out of the two (detailed and smooth) ones, which are used in case of one dimension (i.e., mother wavelet and scaling function). These new four types of 2-D wavelets are three mother wavelets (applied in different directions: vertically, horizontally, and diagonally) and one scaling function. The scaling procedure is dyadic, i.e. it is a stepwise enlargement of wavelets by scale factor 2^j in both dimensions: $2^j \times 2^j,\ j = 1,2, \ldots$ If a data matrix has size $2^K \times 2^K$, then level j can range from 1 to K without any edge effects. Therefore, the 2-D MRA decomposition of a matrix F is

$$F = \sum_{j=1}^{J} D_j^v + \sum_{j=1}^{J} D_j^h + \sum_{j=1}^{J} D_j^d + S_J \qquad (1)$$

This means that it is an additive decomposition into $3J + 1$ components, where the matrices D_j represent the detail (namely high-frequency) parts and the matrix S_J represents the smooth (namely low-frequency) part of matrix F (cf. also Fig.1b). The matrices D_j^v, D_j^h, D_j^d, and S_J are linear combinations of corresponding 2-D wavelets. The limit parameter J is used to constrain the number of multiresolution components. Note that the smooth matrix S_J exclusively emerges at the coarsest resolution, whereas the matrices D_j accumulate over all scales up to resolution level J. This means that, as a scale-by-scale decomposition, the (particularly resulting) smooth matrices can be decomposed again and again. This finally resulting matrix S_J reflects the averaged or smoothed part of matrix F at maximum resolution, whereas the detail matrices D_j^v, D_j^h, D_j^d arising at every resolution level represent its multiple spatial variations or fluctuations. The three directions of 2-D wavelets are indexed by v, h, d for vertical, horizontal, and diagonal, respectively. If index m corresponds to these different spatial directions, equation (1) can be written as follows

$$F = \sum_{m=1}^{3} \sum_{j=1}^{J} D_j^m + S_J \qquad (2)$$

For our purpose, the application of a modified version of the DWT is more appropriate. It is called the maximal overlap DWT (MODWT) (Percival & Walden, 2000). The MODWT is a redundant non-orthogonal transform but has the advantage that it provides as many wavelet coefficients per scale and wavelet-type as elements of F are given. Therefore, the wavelet coefficients give information about which frequencies are dominant at which positions in matrix F. The described properties (Eqs. 1 and 2) hold for DWT as well as for MODWT.

Wavelet multiresolution regression

Based on the 2-D MRA decomposition (Eq. 2), i.e. the decomposition of matrices into scale-specific subcomponents, we are able to develop a regression technique, which allows scale-specific regressions. Different from other methods such as PCNM or MEM, our approach is applied to the response variable as well as all explanatory variables in a multiple regression.

This is possible since the components of all these variables occur in a spatial context as they were sampled in a plane. Thus, we first must convert these vectors into matrices reflecting the spatial layout, i.e. where all the components were sampled. Then the 2-D MRA decomposition (2) can be applied to each matrix built in this way. Subsequently, all scales which are not to be analysed in the model have to be eliminated. Therefore, for instance, transform P_{D_j}

$$F \to P_{D_j} F = \sum_{m=1}^{3} D_j^m \qquad (3)$$

provides a tool keeping only detail matrices of level *j*.

Finally, we revert to vectors (i.e. convert each matrix of specific scale into a vector) that allow us to continue as usual in a linear regression where *y* is a vector of responses and *X* is a matrix of predictors. According to the above-mentioned interpretation of detail matrices, this means that this regression, keeping only detail matrices of level *j*, accounts for fluctuations or spatial variations at a specific spatial resolution (i.e. focus).

Note that this wavelet analysis is applicable not only to normal linear models, but also to regressions in which the response variable has a non-normal distribution. In that case, the canonical link function is $\eta_i = g(\mu_i) = x_i' \cdot \beta, \quad i = 1,2,\cdots,n$, with the expected value of the response being $\mu_i = E(y_i)$, n is sample size, and β is a vector of unknown parameters. The matrix

$$W = diag\left\{ v_{ii}^{-1} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right\} \qquad (4)$$

denotes a diagonal weight matrix, where the variance of the response is $A = diag\{var(y_i)\} = diag\{v_{ii}\}$.

In case of generalized linear models, the process of decomposition and scale selection needs to be restarted within each step of the iterative procedure (Carl & Kühn, 2010). Therefore, the generalized iterative solution for parameter β is

$$b^{(m)} = \left(\left(P_{D_j} W^{\frac{1}{2}} X \right)' P_{D_j} W^{\frac{1}{2}} X \right)^{-1} \left(P_{D_j} W^{\frac{1}{2}} X \right)' P_{D_j} W^{\frac{1}{2}} z\,, \qquad (5)$$

where

$$P_{D_j} W^{\frac{1}{2}} z = P_{D_j} W^{\frac{1}{2}} X \; b^{(m-1)} + P_{D_j} A^{-\frac{1}{2}} (y - \mu)\,. \qquad (6)$$

Quality Criteria

On the one hand, it has been suggested that wavelet covariance could provide a useful measure for scale-dependent interactions between, e.g., an explanatory variable and the response variable in a regression (Kumar & Foufoula-Georgiou, 1997; White *et al.*, 2003). The scale-specific contributions to sample variance or covariance can be efficiently estimated by MODWT wavelet coefficients (Lark & Webster, 2004). The 2-D wavelet variance can be defined as an extension to the 1-D approach as

$$var(j) = \frac{1}{n} \sum_{m=1}^{3} \sum_{i=1}^{n} (d_{im}^F(j))^2 \qquad (7)$$

where $d_{im}^{F}(j)$ are the detail wavelet coefficients of matrix F at level j and n is the sample size. Likewise, the 2-D wavelet covariance, which is a scale-dependent component of covariance for two matrices F and G, is given by

$$cov(j) = \frac{1}{n}\sum_{m=1}^{3}\sum_{i=1}^{n}|d_{im}^{F}(j) \cdot d_{im}^{G}(j)| \quad (8)$$

Here the matrices F and G have to be produced from vectors f and g as described above. The vectors f and g have to be standardized, to ensure that the wavelet transforms are comparable to each other.

On the other hand, regression methods may give us deeper insight into the variety of factors and its relations to a response variable. Generally, regression methods do not only aim for estimation of slope coefficients, but also for hypothesis testing and *P*-values in order to decide whether the predictors are significant or which subset of predictors is relevant. However, problems can arise if we want to compare *P*-values of different models, in particular, models of different sample size. This is due to the fact that the power of the test depends on sample size (McDonald, 2009). When decreasing sample size, the power of the test declines. Since we perform a multi-scale analysis eliminating step by step scale-specific subcomponents and thus rendering certain resolutions ineffective, as a matter of fact the sample size does change. To avoid comparisons of significance tests across scales and instead to provide a consistently good quality criterion, we use the approach of model selection based on multimodel inference (MMI) developed by Burnham & Anderson (2002). Our scale-dependent regression analysis outlined above (Eqs. 5 and 6) allows us to separately apply the method at each scale. This is required because the set of candidate models must always be related to the same dataset. Therefore, at each scale, MMI can make statistical inference via the full set of candidate models and model ranking by means of Akaike's information criterion (AIC). In order to estimate the relative importance of a variable, so-called Akaike weights, i.e. normalized likelihoods of AIC differences, are introduced as model weights. The sum of Akaike weights over the subset of models that include a certain variable can then be considered as a measure for the importance of this variable. Only these values of relative variable importance, i.e. relative instead of absolute values, are the measures that are eventually used for evaluations and comparisons across scales. In our application, it is necessary not only to estimate regression coefficients, but also to calculate an effective sample size for log-likelihoods and thus Akaike weights. That is because most of the information in the 1024 × 1024 grid is redundant due to the positive spatial dependence in the data. This means that individual observations include information already present in nearby observations, so that neighbouring grid cells are highly correlated and the effect or value of sample size is less than the number of observations (Dale & Fortin, 2009). The estimated degree of spatial autocorrelation can therefore be used to adjust the sample size, i.e. to determine how much smaller this effective sample size is (Dale & Fortin, 2005). As shown by Dale and Fortin (2009), the effective sample size n_{eff} for correcting for autocorrelation can approximately be calculated by the following formula:

$$n_{eff} = \frac{n^2}{n+2\sum_{k=1}^{n-1}(n-k)r(k)} \quad (9)$$

where $r(k)$ is the autocorrelation at lag k calculated on $n-k$ pairs of observations. We use Moran's *I* values of the residuals of the full model as an two-dimensional, radially symmetric approximation for the function $r(k)$. For the redundant MODWT based on sample size n across all levels, n_{eff} is a good approximation for all levels.

All statistical analyses were performed using R x64 3.0.1 (R Development Core Team, 2013). The R code is given in Appendix S2. The tools for calculating wavelet transforms are available in package *waveslim* (Whitcher, 2005). Except where explicitly noted otherwise, the results given in the following sections were calculated using d4 wavelets. We tested other types of wavelets as well, e.g., Haar and d16 wavelets. These analyses yielded almost the same results (not presented here for brevity) as those for d4 wavelets. Differences are hardly detectable, except for the highest levels where the loss of information generally causes expanded uncertainties.

Data

We extracted satellite data from the MEDOKADS NOAA AVHRR data archive provided by the Meteorological Institute of the Freie Universität Berlin. Signals from satellite observations are available as so-called normalized difference vegetation index (NDVI) values. NDVI based on the daily reflectance in the red (Red) and near infrared (NIR) AVHRR bands

$$NDVI = (NIR - Red)/(NIR + Red) \quad (10)$$

detects the part of photosynthetically relevant radiation absorbed by plants. Thus NDVI is accepted as a good indicator of the vitality and photosynthetic activity of the vegetation, i.e., NDVI indirectly indicates seasonal changes in leaf and shoot growth and in the greenness of the vegetation. Therefore, it can be used to estimate the length of the vegetation period (White *et al.*, 2003). We computed estimates of vegetation period (see Appendix S1 in Supporting Information), averaged the values over the years 1989-2007, and provided a map representative of Central Europe (Fig. 2). This map has a resolution of $0.01° \times 0.01°$ (approximately 1×1 km^2) and consists of $2^{10} \times 2^{10} = 1024 \times 1024$ grid cells, allowing a dyadic up-scaling from level 1 to level 10 (The original matrix can be indexed by level 0). The selected area ranges across 10.24 degrees of both latitude and longitude. The Alps as well as coastal regions are included. The altitude ranges from sea level to 4300 m. Thus the extent of environmental variation in the region is remarkable. Vegetation period is generally highest for the land-cover type grassland especially in southern Germany followed by forests. Agricultural areas generally display a shorter vegetation period particularly in areas with extensive irrigation or even controlled flooding such as rice fields in northern Italy. Here, the satellite cannot receive vegetation signals due to surfaces covered extensively by water until shortly before harvest. Elevated areas exhibit shorter vegetation periods, most notably for the Alps. Lakes, glaciers and bare land obviously do not display any vegetation period.

Moreover, we extracted climate variables from the WorldClim database Version 1.4 (Hijmans *et al.*, 2005), elevation data from the WorldClim data base, and land cover data from Corine Land Cover 2006 vector data Version 17 (EEA, 2013) and assigned them to the vegetation grid (see Appendix S1 in Supporting Information). This enables us to use vegetation period as the response or outcome variable for further analyses. It is regressed on climate, topographic and land use data, i.e. annual mean temperature (Bio1), annual precipitation (Bio12), altitude, and the land cover categories artificial areas, agricultural areas, forests and grass/scrublands.

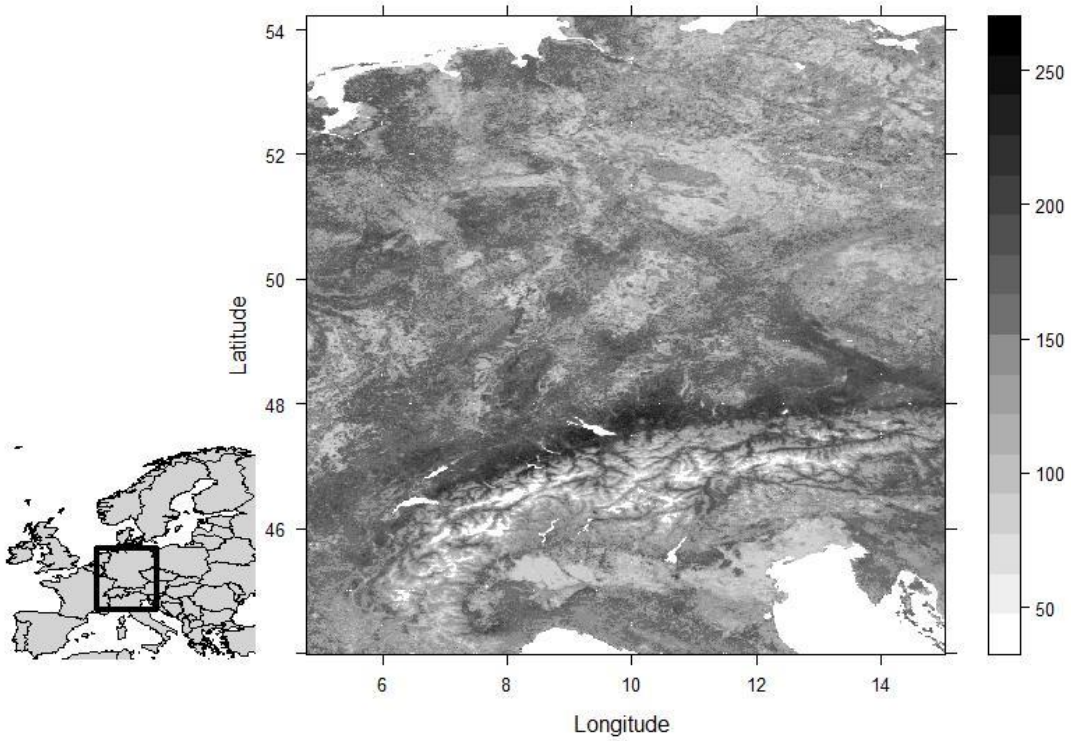


Figure 2: *Map of vegetation periods in Central Europe (latitude 43.99°- 54.22° N, longitude 4.79°- 15.02° E). The data are estimates based on remote-sensing vegetation indices sampled on a* 1024×1024 *grid at a resolution of approximately* 1×1 km^2. *The vegetation period given in days is an average over the years 1989-2007 (greyscale to the right of the map). A map of Europe displaying the selected map section for Central Europe is shown in the inset.*

Results

The extent of variation as a function of increasing spatial resolutions is represented as wavelet variance (Eq. 7) (Fig. 3a). This procedure of up-scaling related to the resolution level can be imagined as a two-dimensionally gradual (i.e. dyadic) enlargement of sizes of grid cells. Roughly speaking, all variances of land use variables show decreasing values with increasing resolution level, i.e. increasing cell size, while variances of other variables show opposite trends. This becomes particularly evident for the levels from 1 (approx. 4 km^2) to 5 (approx. $2^5 \times 2^5$ $km^2 \approx 1000$ km^2 resolution), whereas higher levels yield other results. Note that because of the loss of information with increasing level, the levels higher than 7 provide less reliable results than others.

Wavelet covariances (Eq. 8) evaluating the relationship between explanatory variables and response variables are informative as well (Fig. 3b). One can say at least that both Figs. 3a and 3b reveal that the contributions of climatic, topographic as well as land use variables vary considerably across resolutions. Moreover, the difference of land use variables compared to other ones is detectable in both cases. The relevance of land use variables compared to others switches approximately at level 5 or 6 (cell length approx. 2^6= 64 km, cell size approx. 4000 km^2).

A more accurate analysis of the true amount of relative variable importance can be based on 2-D MRA decompositions, scale-specific regressions, and Akaike weights as described above.

This relative variable importance shows that it indeed provides more detailed curves and thus a deeper insight into what variances and covariances roughly reflect (Fig. 4a). All land use variables appear clearly dominant at the levels 3 (cell size approx. 64 km^2) to 5 (cell size approx. 1000 km^2). The switch of importance between land use variables and other ones occurs between level 5 and 6.

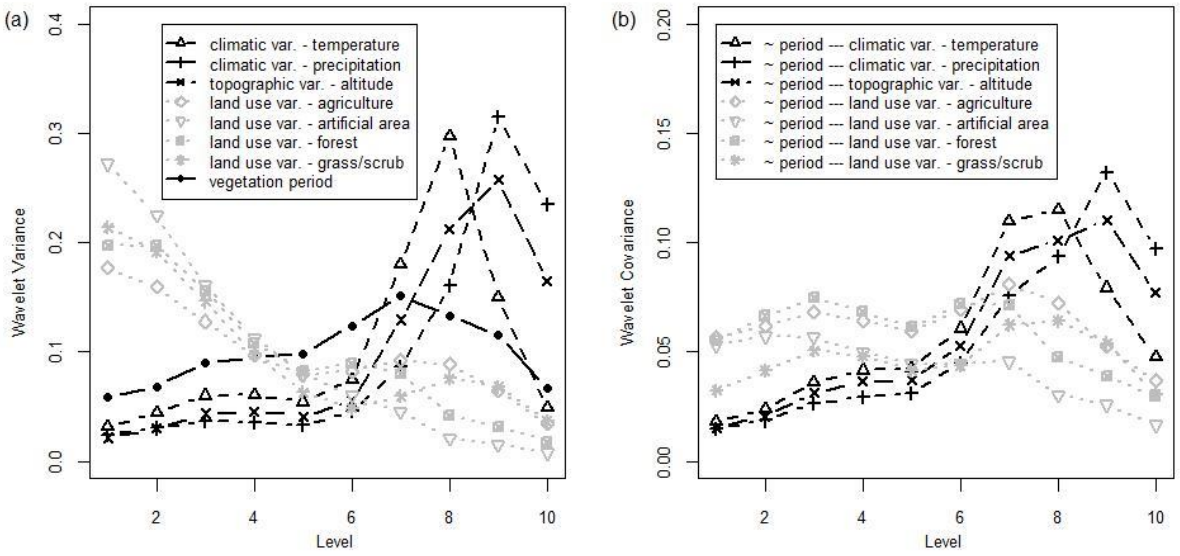


Figure 3: *Wavelet variance (a) and covariance (b) for different variables. Levels range from 1 (cell size approx. 4 km^2) to 10 (coarsest resolution, cell size approx. 1000000 km^2). (Relationships with) Land use variables are indicated by grey lines and symbols, all others are indicated by black ones.*

Discussion

The relative variable importance visualized in Fig. 4a shows clear and strong dependency on resolution. Thus our first assumption, stated at the end of the Introduction, is supported: the impact of climate, topographic as well as land use variables on vegetation period is scale-dependent. Our second assumption was: land use variables are more relevant in relation to vegetation period than climate and topographic ones for models with fine resolution, and vice versa for those with coarse resolution. Although Fig. 3b supports this assumption and Fig. 4a also detects this switch of variable importance between level 5 and 6, this statement cannot be accepted without qualification. Figure 4a provides more details. It shows that, at resolution levels 1 and 2, temperature as climatic variable is more important than all the others, especially than land use variables. One reason for this could be that land scape structure and grid cell structure for remote sensing images are different, i.e. patches of certain land use are not spatially congruent with AVHRR pixels. This leads us to conclude that this difference between landscape and raster format becomes increasingly relevant at lower levels. Note that substituting sample size n with effective sample size n_{eff} in log-likelihoods has no impact on the ranking order in MMI, only the relative distance of weights is reduced. A change in sample size cannot inflate the importance of a single variable, enabling it to dominate and earn the top-ranking position. Therefore, the importance of temperature at fine scales (i.e., levels 1 and 2) is probably an artefact of misassigned land use/land cover.

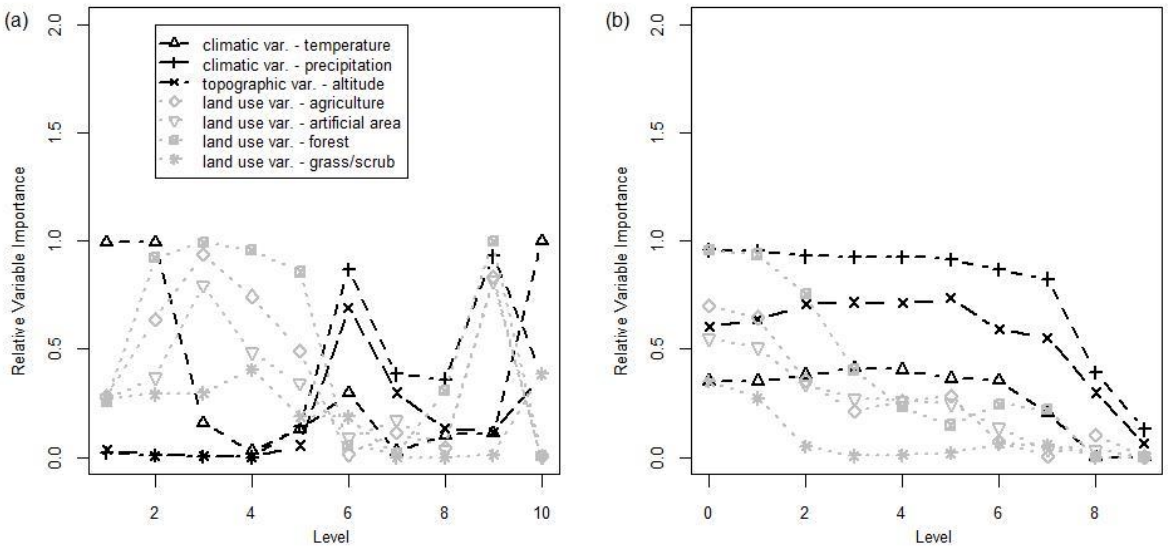


Figure 4: *Relative variable importance (measured as Akaike weights) as a function of scale. Land use variables are indicated by grey lines and symbols, all the others are indicated by black ones. (a) The analysis is based on scale-specific regressions capturing only detail components as scale-specific subcomponents. Levels range from 1 (cell size approx. 4 km^2) to 10 (coarsest resolution, cell size approx. 1000000 km^2). (b) The analysis is based on scale-specific regressions capturing only smooth components as scale-specific subcomponents. Levels range from 0 (grain size approx. 1 km^2, finest resolution, no wavelet transform) to 9 (grain size approx. 250000 km^2).*

On the whole, our examination of resolution-specific variability has shown that there is a marked change in the importance of drivers of the ecosystem process of vegetation greening at a cell length slightly above 32 km. Considering previous knowledge (e.g., Willis & Whittaker, 2002; Pearson & Dawson, 2003) this does not seem unexpected. Similarly to our result, Luoto *et al.* (2007) found that species distribution models of birds in Finland improved notably by including land use data in addition to climate data at resolutions of 10 km and 20 km while at resolutions of 40 km and 80 km climate was sufficient. In this context, it is important to note that our analyses are examinations of the scale-specific variability. Such analyses capture and evaluate resolution-specific variations of variables and their relations, i.e. relations of local fluctuations at a given scale. This is because the previously used wavelet analysis captures only the detail (i.e. high-frequency) components.

However, it seems that many previous studies did not clearly differentiate between resolution-specific (i.e. focus-specific) variation analysis and upscaling by averaging of aggregated cells. As explained above, one has to distinguish between four different components of scale (Scheiner *et al.*, 2000): sample unit, grain, focus and extent. Whereas Willis & Whittaker (2002), inter alia, simply discussed the importance, very generally, of "scale", Pearson & Dawson (2003) were more concrete and translating them into different extents. Still, it is clear that the importance of specific ecosystem processes does not only depend on extent but also on grain and focus. Analysing a process with a grain of 1 km^2 at an extent of 100 km^2 will quite likely yield a different result than having the same grain and continental extent. Hence, we hypothesize that not only varying focus (i.e. resolution level) and keeping grain and extent constant (as we did) will have an effect but also varying grain and keeping extent constant or keeping grain constant and varying extent will have an effect on the hierarchy of drivers of

ecosystem processes (see also Rahbek, 2004). It hence is clear that studies using the same datasets can come to different conclusions (Thuiller *et al.*, 2004; Pompe *et al.*, 2008). Previous studies using cross-scale analyses frequently employ several generalized additive or linear models with intermediate (Luoto *et al.*, 2007) or large resolutions (Rahbek & Graves, 2001) but are often unclear on how the scaling precisely was done. We can assume that many authors simply averaged or lumped data from finer to coarser units.

It is worth noting that data averaging of aggregated cells changes grain size, but not focus. It can be visualized as a smoothing. Based on this interpretation, one might ask the question, what if matrix S_J is always included in scale-specific regressions. The matrix S_J is the smooth or low-frequency part the MRA decomposition of matrix F (see Eq. 2). The interpretation that can be given in this case is as follows: As a consequence of the relation

$$\sum_{m=1}^{3} D_J^m + S_J = S_{J-1}\,, \qquad (11)$$

only smooth components would be under consideration. A scaling procedure based on smooth components can be imagined as a smoothing over gradually enlarged two-dimensional grid cells. In particular, smooth MRA components of Haar wavelets can be seen as a series of averaging operations. The results for scale-specific regressions for such smooth components are given in Fig. 4b. (Note that the level shift is a consequence of the index shift in Eq. 11.) These results reflect what we have to expect when the analysis is carried out for data sampled on increasingly larger grid cells, i.e., it is an analysis quantifying the effect of increasing grain. It can clearly be seen that if the map is split up into increasingly larger grains, the relative importance of land-use variables decreases, whereas precipitation as the variable with the smoothest curve is dominant across all levels and also the relative importance of temperature and altitude remains stable across all levels except for the highest, most uncertain ones.

Conclusions

To the best of our knowledge, this is the first study analysing the scale dependency of an ecosystem process, clearly distinguishing between the different components of scale, namely extent, grain and focus, having an extremely large sample size (n = 1048576), and covering a large range of different resolutions (c. 1 km^2 to c. 1000000 km^2). In summary, our method has the advantage that all calculations were done in a single framework. Firstly, the wavelet approach is carried out by means of multiresolution analysis, which is able to decompose gridded data (maps or images) into components at different resolutions. This data decomposition is embedded into the framework of a multiple regression analysis (Keitt & Urban, 2005; Carl & Kühn, 2008). This wavelet multiresolution regression (WMRR) also allows response vectors of binary or Poisson distribution. Therefore, our WMRR approach is a method for applying two-dimensional wavelet analysis to generalized linear models. Secondly, applying all regressions in a multimodel inference approach circumvents a common problem: Using separate regressions for each scale will result in multiple testing. Due to decreasing sample size, hypothesis tests have declining power. Therefore, results cannot be compared by means of hypothesis testing. The multimodel inference approach does not suffer from this problem. It calculates variable importance by using an information-theoretic approach based on Akaike weights (Burnham & Anderson, 2002). Since results obtained from finer to coarser scaled data can then be compared, one is able to examine the effect of scale dependencies and to evaluate the relative importance of several environmental variables across different spatial scales. Therefore, we provided an answer to the key question whether similar mechanisms act at different spatial scales. We applied our method to data on climate

variables and land cover data to explain variation in vegetation greening as an example of an ecosystem process. Our results indicated that the relative variable importance detectable by scale-specific regressions is strongly scale-dependent. Moreover, for two different approaches, (i) leaving grain and extent constant and changing focus and (ii) leaving extent constant and changing grain, we were able to demonstrate how 2-D scale dependencies can be systematically analysed. It was shown at which "scale" the turning point is where drivers change in importance.

Finally, we believe that our newly proposed method is particularly suitable for studying scale dependencies of various spatial processes on environmental drivers having gridded data with sufficiently large sample size. Roughly speaking, this means that not less than 5 levels should be analysed to check whether scale-dependent changes occur in variable importance. Therefore, a quadratic matrix of at least $2^5 \times 2^5 = 1024$ elements is needed for each of the predictor and response variables.

Acknowledgements

We acknowledge support by the EU infrastructure project "Distributed Infrastructure for EXPErimentation in Ecosystem Research – EXPEER" (project reference 262060) and the Integrated Projects SCALES (grant 226 852) funded under the European Commission Framework Programme (FP) 7. We are grateful to Richard Pearson, Jonathan Belmaker and two referees for constructive critiques of the manuscript.

References

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. Ecology, 85, 1826-1832.

Bruce, A. & Gao, H.Y. (1996) Applied wavelet analysis with S-PLUS. Springer, New York.

Burnham, K.P. & Anderson, D.R. (2002) Model selection and multimodel inference. Springer, New York.

Carl, G. & Kühn, I. (2008) Analyzing spatial ecological data using linear regression and wavelet analysis. Stochastic Environmental Research and Risk Assessment, 22, 315-324.

Carl, G. & Kühn, I. (2010) A wavelet-based extension of generalized linear models to remove the effect of spatial autocorrelation. Geographical Analysis, 42, 323-337.

Cazelles, B., Chavez, M., Berteaux, D., Menard, F., Vik, J.O., Jenouvrier, S. & Stenseth, N.C. (2008) Wavelet analysis of ecological time series. Oecologia, 156, 287-304.

Csillag, F. & Kabos, S. (2002) Wavelets, boundaries, and the spatial analysis of landscape pattern. Ecoscience, 9, 177-190.

Dale, M.R.T. (1999) Spatial pattern analysis in plant ecology. Cambridge Univ. Press, Cambridge.

Dale, M.R.T. & Mah, M. (1998) The use of wavelets for spatial pattern analysis in ecology. Journal of Vegetation Science, 9, 805-814.

Dale, M.R.T. & Fortin, M.J. (2005) Spatial Analysis: A Guide For Ecologists. Cambridge University Press, Cambridge.

Dale, M.R.T. & Fortin, M.J. (2009) Spatial autocorrelation and statistical tests: some solutions. Journal of Agricultural, Biological, and Environmental Statistics, 14, 188-206.

Daubechies, I. (1992) Ten lectures on wavelets. CSBM-NSF series application mathematics, vol. 61. SIAM Publication, Philadelphia.

Dong, X., Nyren, P., Patton, B., Nyren, A., Richardson, J. & Maresca, T. (2008) Wavelets for agriculture and biology: A Tutorial with applications and outlook. BioScience, 58, 445-453.

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196, 483-493.

EEA (2013) Corine Land Cover 2006 seamless vector data. Available at: http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-3#tab-gis-data.

Fortin, M.J. & Dale, M. (2005) Spatial Analysis. A guide for ecologists. Cambridge Univ. Press, Cambridge.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8, 993-1009.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Jombart, T., Dray, S. & Dufour, A.B. (2009) Finding essential scales of spatial variation in ecological data: a multivariate approach. Ecography, 32, 161-168.

Keil, P., Schweiger, O., Kühn, I., Kunin, W.E., Kuussaari, M., Settele, J., Henle, K., Brotons, L., Pe'er, G., Lengyel, S., Moustakas, A., Steinicke, H. & Storch, D. (2012) Patterns of beta diversity in Europe: the role of climate, land cover and distance across scales. Journal of Biogeography, 39, 1473-1486.

Keitt, T.H. & Urban, D.L. (2005) Scale-specific inference using wavelets. Ecology 86, 2497–2504.

Kumar, P. & Foufoula-Georgiou, E. (1997) Wavelet analysis for geophysical applications. Reviews of Geophysics, 35, 385-412.

Lark, R.M. & Webster, R. (2004) Analysing soil variation in two dimensions with the discrete wavelet transform. European Journal of Soil Science, 55, 777-797.

Levin, S.A. (1992) The problem of pattern and scale in ecology. Ecology, 73, 1943-1067.

Lookingbill, T.R., Rocca, M.E. & Urban, D.L. (2011) Focused Assessment of Scale-Dependent Vegetation Pattern. Predictive Species and Habitat Modeling in Landscape Ecology (ed. by C.A. Drew, Y.F. Wiersma and F. Huettmann), pp. 111-138. Springer, New York.

Luoto, M., Virkkala, R. & Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. Global Ecology and Biogeography 16, 34–42.

Ma, Z. & Zhang, L. (2015) Modeling Bird Species Richness at Multiple Spatial Scales Using Two-Dimensional Wavelet Analysis. For. Sci. , 61, 1-16.

Mallat, S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674-693.

McDonald, J.H. (2009) Handbook of Biological Statistics. Sparky House Publishing, Baltimore, Maryland.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography, 12, 361-371.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography, 27, 285-298.

Percival, D.B. & Walden, T. (2000) Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge.

Pompe, S., Hanspach, J., Badeck, F., Klotz, S., Thuiller, W. & Kühn, I. (2008) Climate and land use change impacts on plant distributions in Germany. Biology Letters 4, 564-567.

R Development Core Team (2013) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, {ISBN} 3-900051-07-0, Available at: http://www.R-project.org.

Rahbek, C. (2004) The role of spatial scale and the perception of large-scale species-richness patterns. Ecology Letters, 8, 224-239.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. Proc Natl Acad Sci U S A 98, 4534-4539.

Scheiner, S.M., Cox, S.B., Willig, M., Mittelbach, G.G., Osenberg, C. & Kaspari, M. (2000) Species richness, species-area curves and Simpson's paradox Evolutionary Ecology Research 2, 791-802.

Schröder, B. & Seppelt, R. (2006) Analysis of pattern-process-interactions based on landscape models – overview, general concepts, methodological issues. Ecological Modelling 199, 505-516.

Stark, H.G. (2005) Wavelets and signal processing. An application-based introduction. Springer, Berlin.

Thuiller, W., Araújo, M.B. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? . Journal of Biogeography 31, 353-361.

Torrence, C. & Compo, G.P. (1998) A practical guide to wavelet analysis. Bulletin of the American Meteorological Society, 79, 61-78.

Whitcher, B. (2005) Waveslim: basic wavelet routines for one-, two- and three-dimensional signal processing. R package version 1.5.

White, M.A., Brunsell, N. & Schwartz, M.D. (2003) Vegetation phenology in global change studies. Phenology: an integrative environmental science (ed. by M.D. Schwartz), pp. 453-466. Kluwer Academic Publishers.

Willis, K.J. & Whittaker, R.J. (2002) Ecology. Species diversity--scale matters. Science 295, 1245-1248.

Wu, J. & Hobbs, R. (2002) Key issues and research priorities in landscape ecology: An idiosyncratic synthesis. Landscape Ecology 17, 355–365.

Yang, Y., Xu, J., Hong, Y. & Lv, G. (2012) The dynamic of vegetation coverage and its response to climate factors in Inner Mongolia, China. Stochastic Environmental Research and Risk Assessment, 26, 357–373.

Ye, X., Wang, T., Skidmore, A.K., Fortin, D., Bastille-Rousseau, G. & Parrott, L. (2015) A wavelet-based approach to evaluate the roles of structural and functional landscape heterogeneity in animal space use at multiple scales. Ecography, 38, 740–750.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article

Appendix S1 Additional information about datasets.

Appendix S2 R code for calculating scale-specific regressions.

7.4. Spind: a package for computing spatially corrected accuracy measures

Partners involved: UFZ

Project leaders: Gudrun Carl (UFZ)

Manuscript title

Spind: a package for computing spatially corrected accuracy measures [published in *Ecography*]

One sentence summary

We present 'spind', a new software package (based on the R software program) that provides spatial performance measures for grid-based models.

Authors

G. Carl[1] and I. Kühn[1,2,3]

Affiliations

[1] *Helmholtz Centre for Environmental Research - UFZ, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany*

[2] *Martin Luther University Halle-Wittenberg, Geobotany & Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany*

[3] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*

*Correspondence to: gudrun.carl@ufz.de

Abstract

Using an appropriate accuracy measure is essential for assessing prediction accuracy in species distribution modelling. Therefore, model evaluation as an analytical uncertainty is a challenging problem. Although a variety of accuracy measures for the assessment of prediction errors in presence/absence models is available, there is a lack of spatial accuracy measures, i.e. measures that are sensitive to the spatial arrangement of the predictions. We present 'spind', a new software package (based on the R software program) that provides spatial performance measures for grid-based models. These accuracy measures are generalized, spatially corrected versions of the classical ones, thus enabling comparisons between them. Our method for evaluation consists of the following steps: (1) incorporate additional autocorrelation until spatial autocorrelation in predictions and actuals is balanced, (2) cross-classify predictions and adjusted actuals in a 4x4 contingency table, (3) use a refined weighting pattern for errors, and (4) calculate weighted Kappa, sensitivity, specificity and subsequently ROC, AUC, TSS to get spatially corrected indices. To illustrate the impact of our spatial method we present an example of simulated data as well as an example of presence/absence data of the plant species *Dianthus carthusianorum* across Germany. Our

analysis includes a statistic for the comparison of spatial and classical (non-spatial) indices. We find that our spatial indices tend to result in higher values than classical ones. These differences are statistically significant at medium and high autocorrelation levels. We conclude that these spatial accuracy measures may contribute to evaluate prediction errors in presence/absence models, especially in case of medium or high degree of similarity of adjacent data, i.e. aggregated (clumped) or continuous species distributions.

Introduction

Accuracy measures such as Cohen's kappa coefficient (or Kappa for short) are coefficients useful to assess prediction errors in presence/absence models (such as species distribution models). In a spatial context, however, the traditional non-spatial measures are not appropriate and can thus be misleading in species distribution modelling (Fielding 2002). The reason is that a false prediction has simply the quality of being false regardless of its distance to an appropriate actual value and thus true prediction. One can argue, though, that a false prediction of presence in close proximity to a true (observed) presence is better than a false presence far away from an observed presence (Fielding & Bell 1997; Fielding 2002).

This is particularly the case when sampling at nearby locations leads to sample values that are not statistically independent from each other. If so, then it is to be expected that predictions have the same nature. This phenomenon of statistical dependence caused by spatial dependence should be considered as relevant. This applies particularly to sampling on raster maps, where original data maps are sectioned into grids (Hagen-Zanker 2009). Due to a relatively arbitrary specification of cell size and grid orientation, discretization will generally cause a loss of information. Occurrences at grid cell boundaries, for instance, must be allocated to a specific grid cell (ignoring proximity to the neighbour cell) (Shekhar *et al.* 2002). This is one reason for analysing spatial neighbourhoods and incorporating spatial dependence into accuracy assessment.

There are also ecological reasons for integrating spatial context. Given a species range, searching for new off-range occurrences, one would look probably more frequently and expect more likely to find a new occurrence close to its range margin than further away from its range margin. One of the reasons is that many expanding species have more frequently range advances close to its range margin than those rare long-distance dispersals resulting in occurrences far away (Nathan & Muller-Landau 2000; Nathan *et al.* 2002). Another is dispersal limitation, resulting in new occurrences close to known occurrences even under suitable environmental conditions further away (Svenning, Normand & Skov 2006). Hence models exist to account for sampling bias, giving location further away from currently know occurrences a lower likelihood of being occupied (Bierman *et al.* 2010; Manceur & Kühn 2014).

Classical accuracy measures do not take into consideration the spatial context of any mispredictions. They neglect the degree of similarity of adjacent data. In reality, however, maps of both *actual* and *predicted values* have some degree of spatial autocorrelation (Hagen-Zanker 2009). In the presence of spatial autocorrelation of *model residuals*, the use of methods accounting for this is recommended (e.g., Carl & Kühn 2007; Dormann *et al.* 2007; Carl & Kühn 2010). These approaches account for problems in parameter estimation and realized degrees of freedom resulting in non-autocorrelated residuals. Hence they make sure that no fundamental assumptions of hypotheses testing and statistical approaches are violated. They do not yield, though, uncorrelated *predictions*. Hence the results of such models, when using traditional, non-spatial measure of accuracy, can potentially also suffer from the

problems outlined above. Therefore, the use of spatial metrics of accuracy is even necessary when using methods to account for autocorrelation in model calibration.

For illustration purposes, the maps in Fig. 1 show details of the results of two grid-based models. Although the prediction in Fig. 1b is located in closer proximity to actuals than the prediction in Fig 1a, classical performance measures assign both predictions to the class of false positive errors. In other words, classical measures suffer from the problem that accuracy is not a function of spatial arrangement. Instead, all falsely predicted positive errors rank equally as well as all falsely predicted negative errors, independent of the distance to actual (observed) values.

Here, we present and describe the new software package ‚spind', which introduces several spatial accuracy measures that are (a) sensitive to the spatial arrangement of predictions and (b) comparable to classical measures (Carl and Kühn 2016).

(a) As alternative measures for the evaluation of grid-based models, they will take into account that a false prediction may not be completely wrong if it is in a certain spatial proximity to the correct result. The degree of dependency can be measured and analysed by correlograms, i.e. computations of spatial autocorrelation of both predicted and actual values. Moreover, a new classification and weighting scheme for predictions is needed.

(b) We are not interested in developing totally new spatial measures. Such spatial measures already exist, as for instance, the Average Distance to Nearest Prediction (ADNP) and the Spatial Accuracy Measure (SAM) (Shekhar *et al.* 2002). They have the disadvantage that their results cannot be compared to those of non-spatial measures. Instead, the aim of our study is to generalize classical measures. To enable efficient comparisons, we modify and improve well-known measures (i.e., Kappa, as well as sensitivity, specificity, true skill statistic and other ones) to spatially corrected versions.

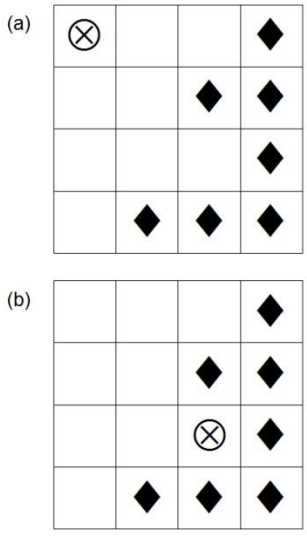


Figure 1: *Example showing a prediction of presence as a result of two different models in relation to the same actual values, i.e. observed presences. Cells with/without diamond ♦ indicate presence/absence of actual values and cells with/without circlemultiply ⊗ refer to presence/absence of predicted values. (a) Locations in the first model, (b) Locations in the second model. In spatial sense, the prediction in (b) might be more accurate than the prediction in (a).*

Methods

Spatially corrected method

The performance of a presence/absence model is often summarized in a confusion matrix (Tab. 1). This is a 2x2 contingency table that cross-classifies observed occurrences (i.e. actual presence / actual absence) and predicted ones according to two classes (i.e. predicted presence / predicted absence). Several classical measures are based on a calculation and evaluation of this confusion matrix. The threshold dividing into classes of predicted presences and absences has frequently the value *threshold* = 0.5, but any other threshold value within the interval from 0 to 1 could be chosen, e.g. based on prevalence or maximizing traditional accuracy measures such as Kappa or true skill statistic. When setting the threshold to 0.5, then the probability of presences is the same as the probability of absences.

Table 1: *Confusion matrix as a 2x2 contingency table. Threshold is the threshold used to transform predicted probability of occurrence of species distribution models into 1's and 0's, for instance, for presence/absence maps.*

	Actual (Presence) *1*	Actual (Absence) *0*	Total
Predicted (Presence) *1 – threshold*	True positive n_{11}	False positive n_{12}	$n_{1.}$
Predicted (Absence) *threshold – 0*	False negative n_{21}	True negative n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

Fielding & Bell (1997) used two simple approaches of weighting in a spatial framework. These are methods that weight false positive errors n_{12} by a function of their distance/ proximity to actual positive locations and thus provide adjusted false positive errors. In this way, the roughly weighted proximity relationships reflect autocorrelation for locations in a two-dimensional gridded dataset. As a result the ratio of adjusted errors to actual errors is recommended for assessment. The magnitude of weights (and thus the strength of autocorrelation), however, was chosen relatively arbitrarily. To circumvent this problem, one can propose new map similarity measures without any weights. One of such measures is the Average Distance to Nearest Prediction (ADNP) (Shekhar *et al.* 2002). This value (i.e. arithmetic mean of distances), however, is not related to a confusion matrix and its corresponding evaluation measures. Conversely, one can try to incorporate a spatial weights matrix reflecting the real proximity relationships into the confusion matrix. Shekhar *et al.* (2002) developed the Spatial Accuracy Measure (SAM) based on such a generalized confusion matrix. Because a direct combination of different distance measures within one confusion matrix is problematic, the spatial weights matrix is incorporated into all elements of the confusion matrix. But as a consequence of this, all totals change in comparison to the classical confusion matrix and thus renormalization limiting their comparability to the

classical confusion matrix is necessary. Hagen-Zanker (2009) introduced an improved Kappa statistic with particular focus on neighbour cells. This extension of the weighted Kappa takes the effect of spatial autocorrelation into consideration, however, without directly quantifying spatial autocorrelation. Instead, the approach tries to estimate its effects by counting adjacent neighbour cells and distinguishing between different degrees of belonging.

To overcome all these problems, we (i) implement proximity as the same amount of spatial autocorrelation in both actual and predicted values and (ii) summarize the results in a weighted 4x4 contingency table.

(i) For spatial data, the amount of spatial autocorrelation can be calculated by means of the Moran's I (e.g., Lichstein *et al.* 2002). This formula measures the strength of two-dimensional autocorrelation based on the assumption that it is isotropic (i.e., independent of direction). Autocorrelation is computed as a function of "lag distance", therefore, one has to introduce lag distance intervals for the spatial structure under consideration. For a square grid underlying all maps used here, the first distance class can be defined to comprise lags between 0 and 1 and thus be assigned to nearest neighbours, i.e. to the (generally) four adjacent grid cells located at distance unit 1 (in relation to coordinates of cell centres) in the cardinal directions. Autocorrelation at lag distance 1 is generally higher than that at greater distances because close observations are more likely to be similar to one another than those far away from each other. Therefore, the autocorrelation value ac(1) is most important. It is noteworthy that the spatial autocorrelation ac(1) of predicted values (i.e., predictions before dividing into groups by a threshold) is generally higher than that of actual values. The reason is that predictions are continuous values varying between the extremes 0 and 1, whereas actual values simply consist of 0's and 1's. This autocorrelation deficit of actuals can be considered as a measure to what extent actual values can be adjusted to reflect a spatial context. Therefore, we generate "adjusted actuals" having the same amount of autocorrelation as predictions. These adjusted actual values are softened compared to the original ones and, accordingly, appear widened in spatial mapping. Therefore, a prediction at a single location can be registered to be in the proximity (i.e. widened area) of an actual value. It is to remark, that, computationally, it is difficult to increase the autocorrelation of actuals in one step to a certain level. Here, we use a step-by-step procedure incorporating autocorrelation until it is balanced with the autocorrelation of predictions.

(ii) For evaluation, one has to summarize the results for predicted and adjusted actual values in a generalized confusion matrix (Tab. 2). In order to ensure that the additional information captured in adjusted actual values is not completely lost again, it is necessary to make the contingency table "finer". If we cross-classify the distributions of the variables in a 4x4 contingency table then we are able to distinguish different kinds of misclassification. Therefore, the predicted values have to be classified into 4 classes separated at the following 3 levels: (1) upper split: *us = (1+threshold)/2*, (2) threshold: *th = threshold, and* (3) lower split: *ls = threshold/*2. Since the total of elements remains constant, a comparison to the results of a 2x2 contingency table is possible. Three cells n_{ij} in the upper right corner (for: $j - i \geq 2$: n_{13}, n_{14}, n_{24}, displayed in dark-grey) contain false positive errors, whereas three cells in the bottom left corner (for: $i - j \geq 2$: n_{31}, n_{41}, n_{42}, displayed in dark-grey) contain false negative errors. This refined weighting pattern can simply be written in matrix notation, i.e. by means of the weighting matrix W

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Having specified the values of this refined cross-classification, we can calculate measures such as weighted Kappa, sensitivity, and specificity for evaluation of prediction accuracy. The weighted Kappa κ is defined as

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where $p_o = \sum_i \sum_j w_{ij}\, p_{ij}$ and $p_e = \sum_i \sum_j w_{ij}\, p_{i.} p_{.j}$ with $p_{ij} = n_{ij}/n$ (Fleiss & Cohen 1973; Fleiss 1981; Sachs & Hedderich 2006). Accordingly, the formulas for the weighted *sensitivity* and weighted *specificity* can be given by

$$sensitivity = (\sum_i \sum_k w_{ik}\, n_{ik}) / \sum_i \sum_k n_{ik} \quad \text{for } k = 1{,}2$$

and

$$specificity = (\sum_i \sum_l w_{il}\, n_{il}) / \sum_i \sum_l n_{il} \quad \text{for } l = 3{,}4\,.$$

By computing *sensitivity* and *specificity* as functions of *threshold*, other measures such as receiver operating characteristic (ROC), the area under the ROC curve (AUC), and maximum true skill statistic (TSS) can be calculated as usual (Hanley & McNeil 1982; Franklin 2009).

Table 2: *Generalized confusion matrix as a 4x4 contingency table. As in Tab. 1, dark grey cells are considered as false while light grey ones as true. Please note that n_{32} and n_{23} would be classified as false in the classical approach but as true here due to the close match. us: upper split, ls: lower split, th: threshold used.*

	Adjusted Actual *1 – 0.75*	Adjusted Actual *0.75 – 0.5*	Adjusted Actual *0.5 – 0.25*	Adjusted Actual *0.25 – 0*	Total
Predicted *1 – us*	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1.}$
Predicted *us – th*	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2.}$
Predicted *th – ls*	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3.}$
Predicted *ls – 0*	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	n

In summary, our new method for evaluation of prediction accuracy consists of the following steps: (1) incorporate additional autocorrelation into binary observation data until spatial autocorrelation in predictions and actuals is balanced, (2) cross-classify predictions and adjusted actuals in a 4x4 contingency table, (3) use a refined weighting pattern for errors, and (4) calculate weighted Kappa, sensitivity, specificity and subsequently ROC, AUC, TSS to get spatially corrected indices.

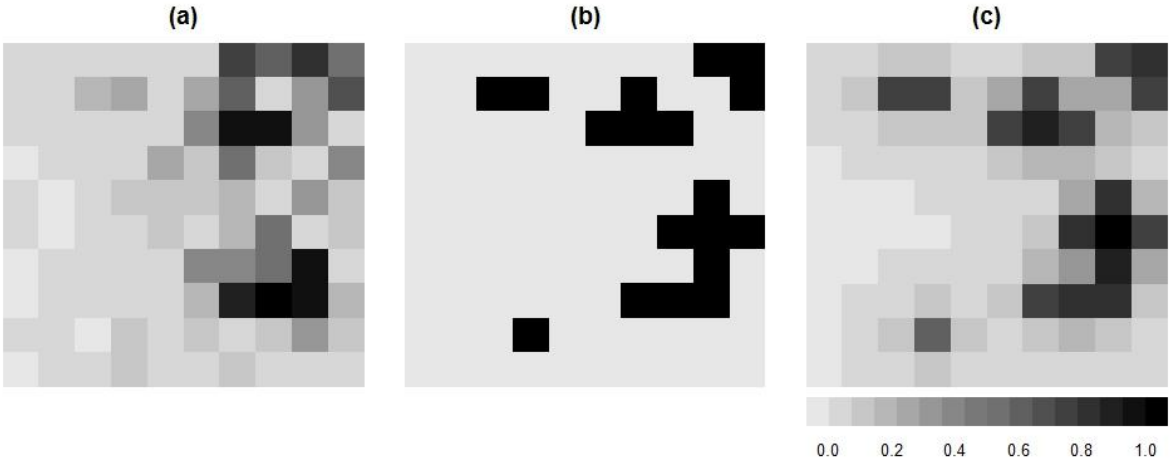


Figure 2: *Example map of simulated data. (a) Predicted values, (b) Actual values, and (c) Adjusted actual values within their spatial context of a 10x10 grid.*

Package overview

All statistical analyses were performed using the R x64 software version 3.1.2 (R Core Team 2014). We provide all tools for calculating spatially corrected indices in our newly created package 'spind' (Carl and Kühn 2016). It is open-source software (published under the GPL public license, ver. 2), and is available as both a package spind_1.0.zip (windows version) and a source package spind.1.0-1.tar.gz. Both R packages, together with documentation, are available on GitHub (< https://github.com/carl55/spind >).

The R package depends on the package 'lattice', which produces Trellis graphics for **R,** as well as 'splancs' with function areapl, which calculates an area of a polygon (Rowlingson & Diggle 1993 ; Bivand & Gebhardt 2000). Spind contains four functions. Function th.dep calculates threshold-dependent metrics (kappa and confusion matrix), i.e. it depends on a cutoff value used for splitting predictions, whereas function th.indep calculates threshold-independent metrics (ROC, AUC, and (max)TSS). Both functions are based on a 2D analysis taking the grid structure of datasets into account (for a regular gridded dataset, grid cells are assumed to be square). Therefore, another two functions are used internally. Function adjusted.actuals provides adjusted actual values reflecting spatial autocorrelation balanced to predictions. Function acfft calculates spatial autocorrelation. Moreover, an example data set (Fig.2) is given to demonstrate how one can use the functions.

Illustration and validation

Application to simulated data

Just in order to visualize the effect of step (1) in our analysis, we firstly present an example of simulated data based on a small grid. The model predictions (Fig. 2a) as well as the actual values (Fig. 2b) are displayed within their spatial context, i.e. the 10x10 grid. When we calculate spatial autocorrelation of predicted and actual values and increase the autocorrelation in actuals (Fig. 3), we produce adjusted actuals (Fig 2c). Fig. 2c shows that grid cells in the immediate proximity of the original agglomeration presented in Fig. 2b have now increased values, whereas a few actual presences are slightly reduced. In our example (Fig. 2b, in the bottom right hand corner), a hook-shaped group of adjoined actuals is to be found just as displayed in Fig. 1b. The prediction for the cell surrounded by this hook has the value 0.52. If we use, for instance, a threshold of 0.5, such a value is classified as false positive error in classical theory. For spatially corrected measures, however, we compute an adjusted actual value of 0.35 at this position. In the 4x4 contingency table (Tab. 2), therefore, this prediction is assigned to element n_{23} and thus is no longer considered an error.

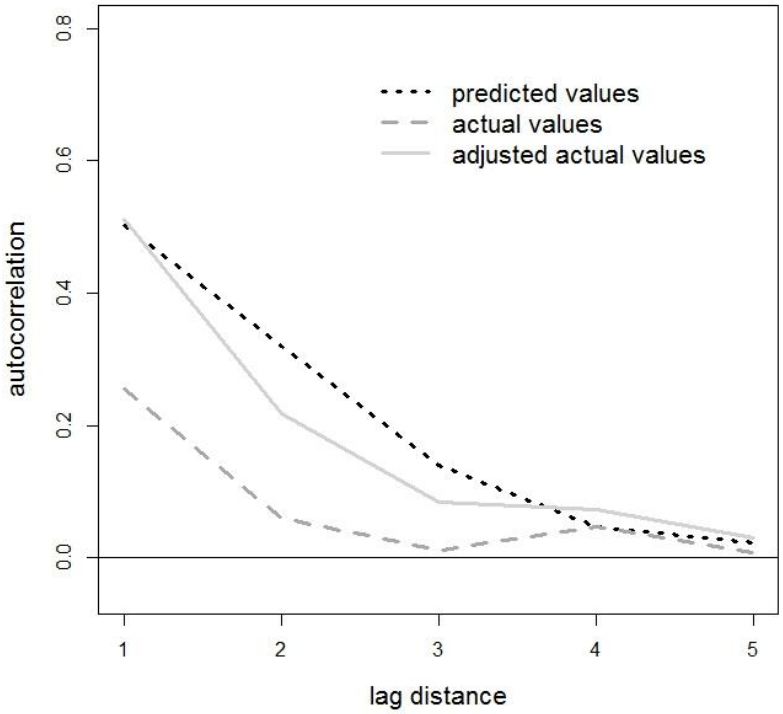


Figure 3: *Example correlograms of simulated data. Spatial autocorrelation of predicted, actual, and adjusted actual values.*

Application to real macroecological data

Secondly, we compute these spatially corrected indices for a real macroecological dataset. Therefore, we selected data for presence/ absence of the plant species *Dianthus carthusianorum* across Germany. This is an example already used in a previous paper (Carl & Kühn 2008). The distribution of actual values of *D. carthusianorum* is given in Fig. 4b. To

produce predicted values (Fig. 4a), we related environmental variables (which need not be the most appropriate ones) to these actual values and performed a logistic regression. Information on species distribution is available from FLORKART (see http://www.floraweb.de) which contains species location in a grid of 2,995 grid cells. The cells of this lattice are 10‘ longitude x 6‘ latitude, i.e. about 11 x 11 km^2, and therefore almost square cells. Moreover, we extracted climate variables (temperature, precipitation) provided by the ‘‘Deutscher Wetterdienst, Department Klima und Umwelt’’, elevation data from the ARCDeutschland500 dataset provided by ESRI, land use data from Corine Land Cover (1990) raster data, and geology digitized from data provided by the “Bundesanstalt für Geowissenschaften und Rohstoffe“ (1993). As explained above, the spatial method modifies actuals until autocorrelation of actuals and predictions is balanced. In our example, the value of spatial autocorrelation of the actuals is 0.63 at lag distance 1, whereas this value for predictions is 0.87. Due to this difference, the method has to produce adjusted actual values of nearly the same magnitude of autocorrelation, i.e. $ac(1) \approx 0.87$. These adjusted actuals softened compared to the original ones are given in Fig. 4c.

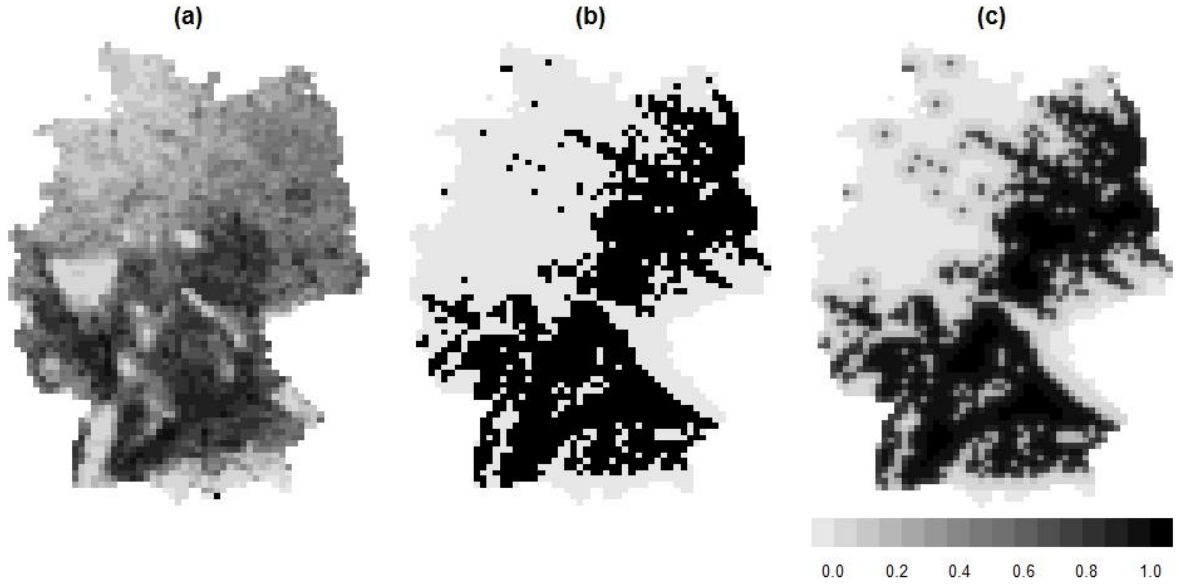


Figure 4: *Example of real macroecological data, i.e. the distribution of Dianthus carthusianorum in Germany. (a) Predicted values, (b) Actual values, and (c) Adjusted actual values*

Statistics

Lastly, several steps are undertaken to verify our spatial indices, in detail. At first we generate data of the kind given in Fig. 2a. For this purpose, values for two predictors and errors are randomly generated and provided with a certain degree of spatial autocorrelation. They are linearly combined using specified parameters (intercept and two slopes). This linear combination is scaled and transformed into outcomes ranging from 0 to 1. For normally distributed variables, the mean value of outcomes is 0.5 on average. Hence the prevalence of the simulated species is set to 0.5. If we subsequently split the values in 0’s and 1’s using a threshold value 0.5, then, of course, these ex-post created “actuals” match perfectly with the “predictions” generated prior to this. The values for classical Kappa, AUC, and TSS are 1 in

this case of perfect match. This should also be valid for spatial indices. Additionally, in order to check the effect of a certain mismatch, we modify the map of actuals by shifting all columns to the left adjacent position (except for the leftmost column, which is shifted to the rightmost position). One can expect that such a displacement or pattern of "shifted match" will result in lower fitting accuracy and thus lower values for Kappa, AUC, and TSS. The values depend on the index used for evaluation and, in addition, the degree of spatial autocorrelation. This is because the degree of adjacent similarity is relevant. If values in neighbourhoods are similar, then shifting may be less problematic than if values are randomly distributed and independent. To compare classical measures with our spatially corrected ones (i.e. spatial indices), we generate 30x30 maps as described above for both perfect and shifted match at 10 different levels of autocorrelation. Using 100 randomly generated datasets in each case, we run all settings 100 times to produce 100 solutions for the indices in each case.

Results

Application to real macroecological data

To demonstrate the impact of our method in a real-world example, the results for classical and spatially corrected measures for presence/absence data of the plant species *Dianthus carthusianorum* are given in Tab. 3. One can clearly see that the numbers of both false positive errors and false negative errors are less for spatial indices compared to those of classical ones. As a consequence, the values for Kappa (threshold = 0.5), AUC, and TSS increase when incorporating spatial corrections.

Table 3: *Predictions for plant species Dianthus carthusianorum across Germany. Results for classical measures and spatially corrected measures (i.e. spatial index).*

	classical index	spatial index
false positive errors	397	331
false negative errors	462	406
Kappa	0.42	0.46
AUC	0.80	0.85
TSS	0.48	0.57

Statistics

Using 100 randomly generated datasets to compare classical and spatial indices, we find that the values for both classical and spatial indices reach their maximum of approximately 1 if we use data of perfect match. In case of shifted match, all indices are functions of autocorrelation. Starting at autocorrelation level 0, all indices increase as a function of autocorrelation. As expected and methodologically intended, spatial indices are, on average, equal or higher than classical ones. Higher values occur at medium and high autocorrelation levels due to the increasing degree of adjacent similarity being taken into consideration by spatial indices. We can see that especially when having strong autocorrelation, spatial indices tend to result in higher values and classical indices would indicate a poorer fit. Mean values

and error bars for Kappa, AUC, and TSS are given in Fig. 5a. For testing the null hypothesis that the value for classical Kappa is equal to the mean value of spatial Kappa values, we use a 95% confidence interval. It is obtained by $(\hat{\kappa} - 1.96 \cdot \sigma(\hat{\kappa}), \hat{\kappa} + 1.96 \cdot \sigma(\hat{\kappa}))$, where $\hat{\kappa}$ is the mean value of classical Kappa and $\sigma(\hat{\kappa})$ is its standard deviation (Kanga *et al.* 2013). We found that for an autocorrelation value of 0.7, the null hypothesis is rejected and the difference for Kappa values is thus statistically significant. Accordingly, hypothesis tests can be used to evaluate differences between classical and spatial AUC values and between classical and spatial TSS values. In both cases, we found statistically significant differences at autocorrelation values of 0.6, 0.7, and 0.8.

One might still ask whether spatial autocorrelation of predicted values is appropriate to estimate the autocorrelation deficit of actuals and thus to define their neighbourhoods. More specifically, the question arises whether the predictions are appropriate as a basis for adjusting the observations. We can respond with a counter question: How can one get better estimates for actual values than model predictions? Note that we do not use the predictors themselves such as environmental variables. Instead, outcomes predicted by statistical models such as species distribution models are used here. As a consequence, predictors that are not significant will usually have no impact, or only a minor impact, on predictions. To gain deeper insight into the adjustment of actuals and to discuss the risks of our method, we present a further example of simulated data. For this purpose, values for two non-autocorrelated predictors and for a non-autocorrelated error are randomly generated. They are linearly combined using specified parameters (intercept and two slopes). This linear combination is scaled and transformed into outcomes ranging from 0 to 1. Subsequently, we split the values in 0's and 1's as above. To produce (non-autocorrelated) predicted values, we relate the two predictors to these actual values and perform a logistic regression. Having non-autocorrelated data, we find the same values for both classical and spatial indices. If we instead regress these actuals on predictors affected by a certain degree of autocorrelation, then the fitting accuracy decreases. Note that, in this case, autocorrelation acts as a disturbing factor. The classical indices are the correct ones, and the spatial indices, which falsely impose autocorrelation, result in higher values. Therefore, this example investigates to what extent our method adjusts the observations wrongly towards an incorrect pattern. Mean values and error bars for Kappa, AUC, and TSS are given for 100 randomly generated datasets (Fig. 5b). As can be seen from Figure 5b, the differences in fitting accuracy are less than the standard deviations of classical indices and thus not significant.

Discussion

Our results show that especially under medium to high levels of spatial autocorrelation of predicted data spatial measures of accuracy yielded different results compared to classical measures. We therefore advocate the use of the proposed metrics.

There were several assumptions we made: (i) We corrected the actuals by adding autocorrelation to the same degree as that of the predicted values in order to change binary data to continuous data and to be able to define a neighbourhood. For technical reasons, it is impossible to do it the other way round. Still, this frequently results in lower values than 1 (being absolutely present). (ii) While we used quartiles to classify adjusted actuals, we used varying thresholds and, dependent on them, upper and lower splits to classify predicted values. The flexibility for predictions is needed, because it is not always useful to use a threshold of 0.5 (Liu *et al.* 2005; Hanberry & He 2013). Using the same flexibility for actuals, though, turned out not to be useful when developing the method. (iii) It is in the logic of the spatial index method to regard cells with close by values (n_{23} and n_{32} in the generalized confusion matrix) as true, rather than false. But this decision is arbitrary. Not doing so would

result in the classical measures of accuracy. (iv) The generalized confusion matrix could in principle also have more elements than 4x4 cells. It turned out, however, that this gets computationally difficult, especially with varying thresholds. Further, defining "true" and "false" would get very arbitrary. Still, due to defining the spatial neighbourhood, the 4x4 cell confusion table might result in a higher susceptibility to very high or very low prevalences. This means that at small prevalences the number "present" adjusted actuals (n_{32}) might increase and at very high prevalences the number "absent" adjusted actuals (n_{23}) might increase at disproportionate rate compared to prevalence (as an effect of an unfavourable edge/area ratio). In such cases, though, with just very few observed presences or observed absences, robust models are inherently difficult to fit. Hence it is warned against the parameterisation of data deficient models, anyhow (Coudun & Gégout 2006; Franklin 2009, p. 63).

As briefly outlined in the methods section, there are measures available that consider spatial proximity. Fielding and Bell (1997) weight the false positives errors by a distance function to actual positive locations. The advantage of our approach is that the degree of spatial weighting is estimated as autocorrelation deficit rather than set arbitrarily. Further, in our approach the marginal sums of the confusion matrix remain the same compared to non-spatial metrics and we also consider the distance of false negatives to actual negatives. The methods of Shekhar *et al.* (2002) introduced completely different metrics which cannot be compared to the classical metrics, by design. It is our utmost aim to retain comparability between spatial and non-spatial metric but minimize arbitrary decisions.

The results of our simulations (Fig. 5) suggest that the proposed spatial measures of model accuracy only increase accuracy and do not decrease accuracy. This, however, is not inevitable. Because known presences get down-weighted and known absences get up-weighted by adjusting the actuals, in principle fit could (slightly) decrease, but will probably very rarely happen. So on average model accuracy increases when using spatial metrics. One could then argue that this is not helpful and does not warrant using the new approach. Using our metrics will help to formalize spatial uncertainty and may even account (partially) for unobserved, though present, actuals, i.e. occurrences that indeed are there but were not yet observed. To some degree observer bias (Manceur & Kühn 2014) can thus be minimized when assessing accuracy. Further, our approach increases the comparability of results between autocorrelated and non-autocorrelated data. And lastly we argue that the use of spatial measures of accuracy is better, since we think it is the correct measure, compared to the use of non-spatial measures, in case of autocorrelated data.

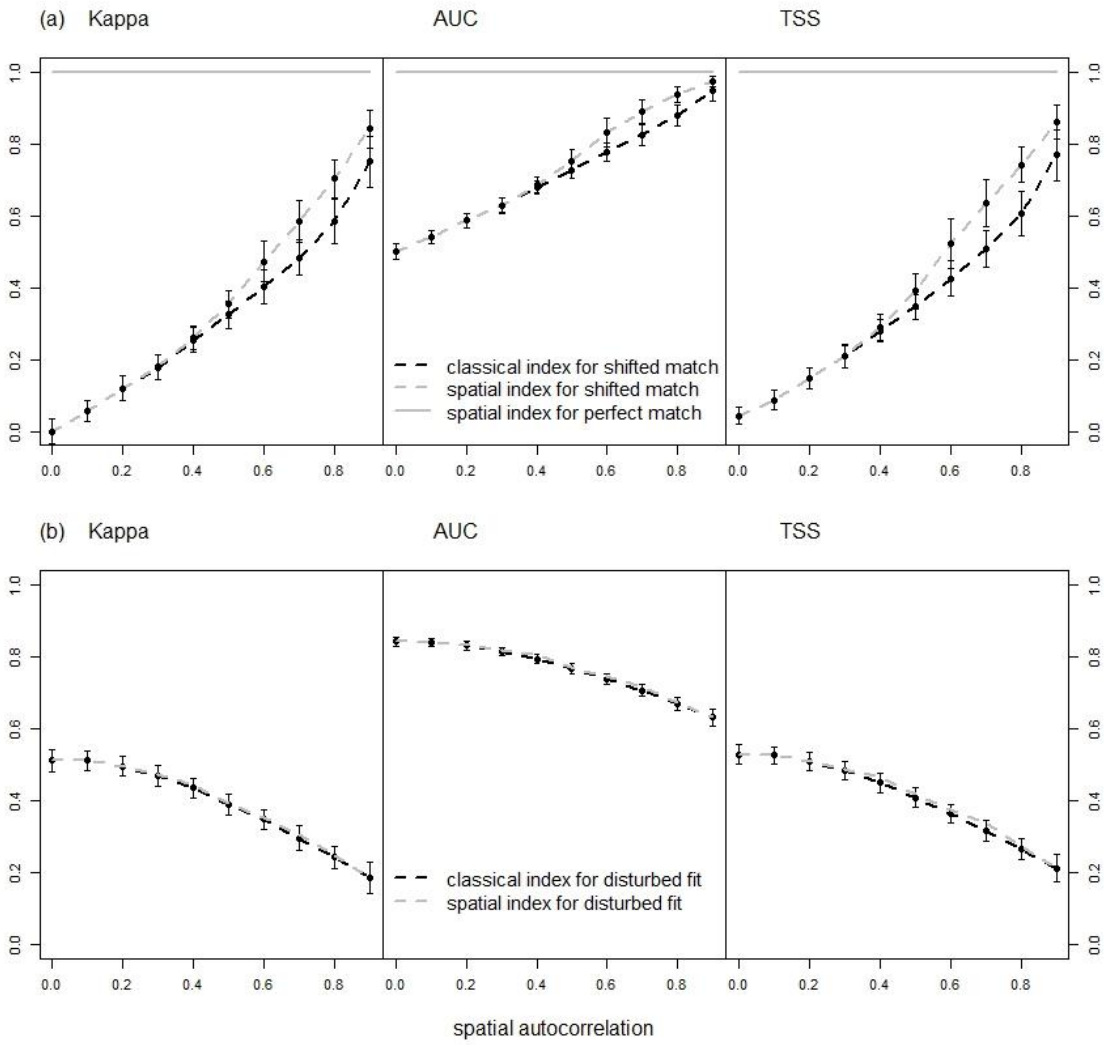


Figure 5: *Statistic for comparing classical measures to spatially corrected ones (i.e. spatial indices). The indices Kappa, AUC, and TSS are given as a function of autocorrelation (a) for both perfect and shifted match and (b) for disturbed fit. Spatial autocorrelation is measured as Moran's I. The strength of autocorrelation is indicated by ac(1), i.e. the value of autocorrelation related to nearest neighbours. The error bars indicate the interval delimited by mean value ± standard deviation.*

Having non-autocorrelated data, our simulations suggest that there is no difference between the spatial and the classical measures of accuracy. So one could use them but it is not necessary. In non-autocorrelated situations, therefore, spatial arrangements of predictions and actuals become irrelevant. In the presence of autocorrelated data, however, one is advised to already use spatial metrics of accuracy.

One issue that still remains unsolved is to properly measure accuracy when having presence-only data. Since both, classical and spatial metrics, need presence as well as true absence data, they are inappropriate when using presence-only data. The results heavily depend on the

choice of the algorithm used to select pseudo-absences (e.g., Barbet-Massin *et al.* 2012). This is a fruitful and rewarding topic for future research.

We conclude that these spatial accuracy measures are useful, especially in case of medium or high degree of similarity of adjacent data. They are primarily intended as goodness-of-fit measures for the evaluation of species distribution models based on high resolution maps.

Acknowledgements

We acknowledge support by the EU Collaborative project "EU BON: Building the European Biodiversity Observation Network" (grant 308454) funded under the European Commission Framework Programme (FP) 7. We are grateful to Thiago Rangel and two referees for constructive critiques of the manuscript.

References

Barbet-Massin, M., et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? . - Methods in Ecology and Evolution 3: 327-338.

Bierman, S. M., et al. 2010. Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. - Ecography 33: 451-460.

Bivand, R. and Gebhardt, A. 2000 Implementing functions for spatial statistical analysis using the R language. - Journal of Geographical Systems 2: 307-317.

Bundesanstalt für Geowissenschaften und Rohstoffe 1993. Geologische Karte der Bundesrepublik Deutschland 1:1 000 000. Hannover: Bundesanstalt für Geowissenschaften und Rohstoffe.

Carl, G. and Kühn, I. 2007. Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. - Ecological Modelling 207: 159-170.

Carl, G. and Kühn, I. 2008. Analyzing spatial ecological data using linear regression and wavelet analysis. - Stoch Environ Res Risk Assess 22: 315–324.

Carl, G. and Kühn, I. 2010. A wavelet-based extension of generalized linear models to remove the effect of spatial autocorrelation. - Geographical Analysis 42: 323-337.

Carl, G. and Kühn, I. 2016. Spind: a package for computing spatially corrected accuracy measures. Ecography. DOI: 10.1111/ecog.02593

Corine Land Cover 1990. raster data. Available at: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-1990-raster.

Coudun, C. and Gégout, J.-C. 2006. The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. - Ecological Modelling 199: 164-175.

Dormann, C. F., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. - Ecography 30: 609-628.

Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. - Environmental Conservation 24: 38-49.

Fielding, A. H. 2002. What are the appropriate characteristics of an accuracy measure? - In: J. M. Scott, et al. (eds), Predicting species occurrences. Issues of accuracy and scale. Island Press, pp. 271-280.

Fleiss, J. L. and Cohen, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. - Educational and Psychological Measurement 33: 613-619.

Fleiss, J. L. 1981. Statistical methods for rates and proportions. - Wiley.

Franklin, J. 2009. Mapping species distributions. Spatial inference and prediction. - Cambridge University Press.

Hagen-Zanker, A. 2009. An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. - International Journal of Geographical Information Science 23: 61-73.

Hanberry, B. B. and He, H. S. 2013. Prevalence, statistical thresholds, and accuracy assessment for species distribution models. - Web Ecology 13: 13-19.

Hanley, J. A. and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristics curve. - Radiology 143: 29-36.

Kanga, C., et al. 2013. Kappa statistic for the clustered dichotomous responses from physicians and patients. - Stat Med. 32: 3700-19.

Lichstein, J. W., et al. 2002. Spatial autocorrelation and autoregressive models in ecology. - Ecol. Monogr. 72: 445-463.

Liu, C., et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. - Ecography 28: 385-393.

Manceur, A. M. and Kühn, I. 2014. Inferring model-based probability of occurrence from preferentially sampled data with uncertain absences using expert knowledge. - Methods in Ecology and Evolution 5: 739-750.

Nathan, R. and Muller-Landau, H. C. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. - Trends in Ecology & Evolution 15: 278-285.

Nathan, R., et al. 2002. Mechanisms of long-distance dispersal of seeds by wind. - Nature 418: 409-413.

R Core Team 2014. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available at: http://www.R-project.org.

Rowlingson, B. and Diggle, P. 1993 Splancs: spatial point pattern analysis code in S-Plus. - Computers and Geosciences 19: 627-655.

Sachs, L. and Hedderich, J. 2006. Angewandte Statistik, Methodensammlung mit R. - Springer.

Shekhar, S., et al. 2002. Spatial contextual classification and prediction models for mining geospatial data. - IEEE Transactions on Multimedia 4: 174-188.

Svenning, J.-C., et al. 2006. Range filling in European trees. - Journal of Biogeography 33: 2018-2021.

7.5. Uncertainties in distribution predictions of freshwater biodiversity under land-use and climate change scenarios in the RMO test site

Partners involved: SGN

Project leader: Mathias Kuemmerlen (SGN)

One sentence summary

We found evidence that choice of individual predictors can carry significant consequences for model projections in terms of uncertainty.

Authors

Matthias Kuemmerlen[1,2,3], Stefan Stoll[2,3], Peter Haase[2,3]

Affiliations

[1]*Eawag, Überlandstraße 133, 8600 Dübendorf, Switzerland*

[2] *Senckenberg Research Institute and Natural History Museum Frankfurt, Department of River Ecology and Conservation, Clamecystr. 12, D-63571Gelnhausen, Germany*

[3] *Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt am Main, Germanya*

*Correspondence to: mathias.kuemmerlen@eawag.ch

Abstract

Global environmental changes are predicted to modify significantly the global distribution patterns of biodiversity. Freshwater ecosystems are particularly sensitive to anthropogenic impairments and represent an optimal system to investigate the possible effects of future environmental conditions. Species distribution models are frequently used to predict potential changes in the distribution of biodiversity under climate change conditions. Uncertainty is always associated with such models, particularly because models inherit it from different sources: environmental predictor variables, biotic data, modelling methods, among others. We set up for benthic macroinvertebrates, freshwater fish and aquatic macrophytes in a catchment in central Germany. The model was projected using five scenarios for the year 2030: one on land-use change, two on climate change scenarios (RCP 4.5 and 8.5) and two further ones depicting the combined effect of each climate scenario and the land-use scenario. Uncertainties were investigated according to taxonomic group, type of future scenario and spatial structuring along the stream network. Uncertainty stemming from individual projection was evaluated by estimating the coefficient of variation across repetitions. Uncertainty in this study was more strongly related to individual scenarios, rather than to climate or land use scenarios. This is probably due to the magnitude of the predicted change, specific to each scenario. We found evidence that choice of individual predictors can also carry significant consequences for model projections in terms of uncertainty.

Introduction

Anthropogenic environmental change consists of several ecosystem impairments of which land-use and climate change are the two best known ones (Vitousek 1992). The effects of these impairments on ecosystems and their impacts on biodiversity have been investigated for over a decade (Sala et al. 2000). However, the vast majority of research has focused on the consequences of climate change, ignoring most of the remaining components of global environmental change (Titeux et al. 2016). In addition, most studies have focused on large areas such as continents, leading to results that cannot be implemented in terms of local conservation initiatives and management decisions.

In freshwater ecosystems, the most relevant spatial unit for management and conservation purposes is easy to define: the catchment (Saunders et al. 2002; Palmer et al. 2009). The topography in the landscape dictates how water is divided into different valleys and these are then connected through a stream network. This hierarchical ecologically relevant dependency, as well as the relatively uniform environmental properties and biotic communities make catchments the optimal spatial scales to investigate the effects of environmental change. Nevertheless, only few studies have addressed the effects of more environmental factors beyond climate change (CC): urbanization (Nelson et al. 2009) and land-use change (Kuemmerlen et al. 2015).

Freshwater habitats are regarded as the most threatened ecosystems globally considering the relatively small surface they occupy and the very high number of species that inhabit them (Malmqvist and Rundle 2002; Dudgeon et al. 2006; Strayer and Dudgeon 2010). Future environmental conditions have been predicted to modify the availability of water as a resource (Kundzewicz et al. 2008), to affect freshwater fisheries (Ficke et al. 2007) and to reduce the distribution ranges of benthic macroinvertebrates (Domisch et al. 2013), just to name a few. However, no study has compared the responses of several taxonomic groups under the same future environmental conditions.

In the present study, for the first time, we compare the effects of future environmental conditions on distribution predictions of benthic macroinvertebrates, freshwater fish and freshwater macrophytes simultaneously. For this purpose we set up a species distribution model (SDM) for the catchment of the Kinzig River in Central Germany, using topographical, geological, climatic, hydrologic and land-use predictors. Calibrated models were then projected for one land use change (LUC) scenario and two CC scenarios, as well as and the combination of them. Results allow us to assess which models performed the best and which predictors are most important in explaining the distributions for each one of the taxonomic groups. Further, we analyse how predicted distributions respond to LUC and CC scenarios, as well as the combinations of both and compare them between the three taxonomic groups.

Predicting the distribution of species using SDMs has become a widespread method, implemented in all biomes (Elith and Leathwick 2009). The exponential increase in its usage has also sparked criticisms, aimed mainly at simplistic approaches (Araújo and Peterson 2012; Merow et al. 2014). While SDM predictions can only approximate the fundamental niche of species (Peterson and Soberón 2012), it remains a powerful tool to understand current distribution patterns and to shed some light on the potential changes that can be expected for the future. The strength of SDMs relies mostly on the amount of biological information required to set up a model: it is much more modest when compared to alternative, mechanistic models (Singer et al. 2016). Further, SDM have been successfully implemented to assess the effects of future predicted LUC and CC, both individually and combined (Kuemmerlen et al. 2015).

Uncertainties are not frequently assessed in SDMs, although there have been studies assessing their importance. Uncertainty can be an important factor to take into account when evaluating distribution predictions, particularly when considering their spatial distribution. This is why Rocchini et al. (2011) make a call to address uncertainty and explicitly plot its distribution in so called *maps of ignorance*.

Materials and Methods

A SDM was set up for the Kinzig catchment with biological and environmental data stored at repositories of the Rhine-Main-Observatory (RMO), a long-term ecological research site operating in the Kinzig catchment since 2007. The catchment has an area of approx. 1 060 km^2, with elevations ranging from 98 to 731 m.a.s.l. Additional information on the RMO LTER can be found in Tonkin *et al* (2016) and Kuemmerlen *et al* (2016).

The delineated catchment and the stream network were obtained from a sink-corrected digital elevation model (DEM; ©GeoBasis-DE; BKG, 2011) at a spatial resolution of 25 m. In order to obtain the stream network, an existent stream vector (©GeoBasis-DE; BKG, 2011) was burnt into the DEM, followed by the computation of flow direction and flow accumulation. Then, cells representing subcatchments of size 1 km^2 or larger, were defined as streams. Based on the obtained stream network (28,205 grid cells), the catchment for the Kinzig River, from its confluence into the Main River, was delineated (1,740,742 grid cells). All geo-processing procedures were carried out using the software ArcGIS (ESRI, Redlands CA, USA).

Biological records of macrophytes, fish and macroinvertebrates were compiled from material collected in the regular monitoring of the RMO by the staff of the Senckenberg Research Station Gelnhausen and from data provided by the Hessian Authority for Nature Conservation, Environment and Geology (Hessisches Landesamt für Naturschutz, Umwelt und Geologie: HLNUG), as well as Hessian Forestry Authority (HessenForst). Only species with sufficient observations were included in the model: 20 or more for macroinvertebrates and fish; 10 or more for macrophytes. The latter were subjected to a lower threshold because of the general scarceness of records and the low number occurrences available for all species.

Preselection of environmental variables was based on (i) correlation analysis to rule out linear dependence between variables ($r < |0.7|$; Dormann et al. 2013); (ii) previous model runs identifying relevant predictors; and (iii) expert knowledge. This resulted in 16 predictors that were applied to all taxa groups.

Aspect (orientation) and slope (inclination) were included as topographic predictors in the model. These were calculated from the DEM at a spatial scale of 25m. Aspect serves as a proxy for sunlight exposure, which influences water temperature and primary production, while slope is a surrogate for flow velocity and oxygen saturation of streams (Kuemmerlen et al. 2014).

The geological predictors basalt, sandstone, limestone, as well as fine and coarse sediments, were extracted from the GÜK300 database for the German state of Hesse (©HLUG, 2007). These particular predictors resulted from reclassifying existing categories. Further, for every predictor, the proportion in the upper subcatchment was calculated for every grid cell in the stream network. This approach depicts the relative importance of the geology and has been applied successfully before in freshwater SDMs, particularly for fish (Kuemmerlen et al. 2016; Kuemmerlen and Haase, in prep).

Climatic predictors were calculated as *bioclim* variables (Nix 1986) and included mean annual temperature, mean diurnal temperature range and temperature of the warmest quarter. For this purpose, maximum and minimum monthly temperature data was obtained from the Land Surface Temperature (LST) dataset for Europe with a 250 m spatial resolution (Metz et al. 2014) and mean monthly precipitation from a dataset of the German Meteorological Institute (Deutscher Wetter Dienst [DWD], unpublished data) with a 1 km spatial resolution and for the time period between 2003 and 2012. Monthly averages were calculated for both datasets and downscaled to 25 m through a geographically weighted regression using the elevation from the DEM, implemented with the package "raster" for R (R Development Core Team 2014; Hijmans 2016). For the CC scenario, predicted air temperature (2 m above ground) and precipitation data were obtained from the EURO-CORDEX initiative (Jacob et al. 2013) at the 0.11 degree spatial resolution. Both temperature and precipitation were averaged across three different models: the MPI-ESM comprehensive earth-system model (Giorgetta et al. 2013), the Regional Atmospheric Climate Model (RACMO; van Meijgaard et al. 2008) and the regional climate model COSMO-CLM (Keuler et al. 2016). The average was implemented using the delta change method: differences between average values of a baseline period (2003-2012) and a future period (2020-2040) were calculated and added to current climatic conditions. In this way, only the anomalies of predicted climate conditions are applied to present climatic data. This is not trivial, as data of current conditions has a fine spatial resolution, while future scenarios have much coarser spatial resolutions. Data for two different climatic scenarios of the Intergovernmental Panel on Climate Change was prepared: the Representative Concentration Pathways (RCPs) 4.5 and 8.5 (IPCC 2013). They represent a realistic (RCP 4.5) and a severe (RCP 8.5) scenario and are also the only ones provided by EURO-CORDEX for both temperature and precipitation

In the present model, hydrology is represented by one surrogate of discharge: accumulated precipitation only. Previous freshwater SDMs have used data from hydrological models for present and future predictions with high temporal resolution, which allows calculating additional hydrological indicators (e.g. Kuemmerlen et al. 2015). However, for each scenario investigated, a new hydrological model is required. In the current study we assessed five different future scenarios, in addition to the present projection, which hampered the usage of such modelled hydrological data. Instead, discharge is approximated by adding the annual precipitation in the upper subcatchment, for each grid cell in the stream network. This method has the advantage that it requires much lower effort as it can be derived from precipitation data, but it also has the disadvantage that it only takes into account possible changes to discharge caused by CC.

Five different land uses were included in the model: agriculture, industrial, open areas with natural vegetation (e.g. windthrow), pasture and urban. Data for both present and future land use was sourced from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR; Hoymann and Goetzke 2014). Future land use was modelled for the year 2030 focusing on the development of settlement and transport area, but also considering socio-economic development and spatial planning regulations. For both present and future land uses, the relative proportion in the upper subcatchment was calculated for each single grid cell.

To compare the performance indicators and present predicted ranges between taxonomic groups, Wilcoxon two-sample tests were performed. To compare present and future predicted ranges within taxonomic groups, paired Wilcoxon two-sample tests were used.

Results

All individual models showed high performance with average values of 0.79 ± 0.08, 0.94 ± 0.03, 96.5 ± 3.5 and 82.0 ±6.5, for True Skill Statistic (TSS; Allouche et al. 2006), area under curve (AUC), Sensitivity and Specificity, respectively.

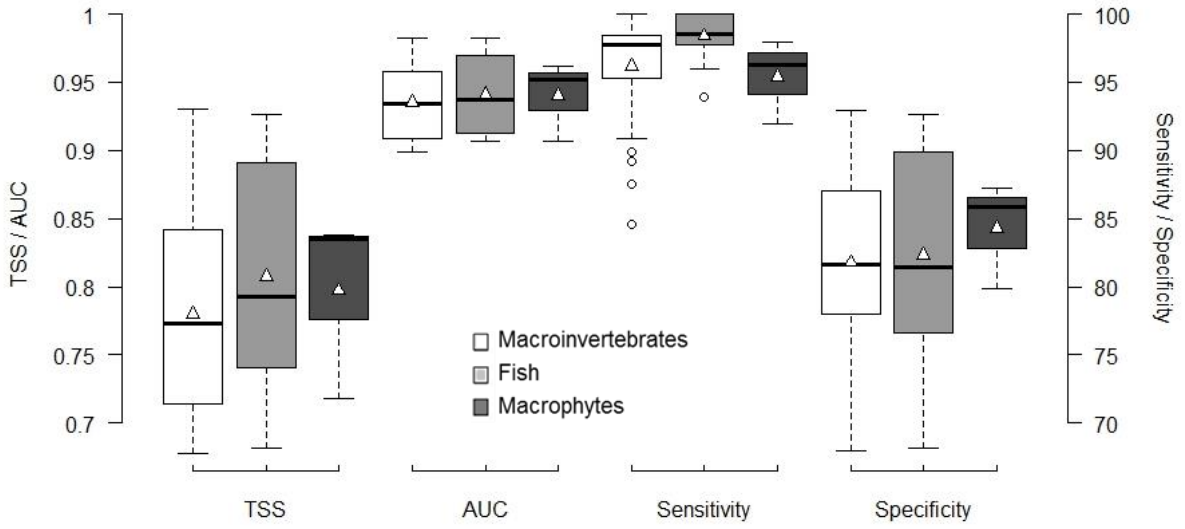


Figure 1: *Model performance indicators for macroinvertebrates, fish and macrophytes. TSS = true skill statistic; AUC = area under curve.*

The predictor category explaining most of the variance in the model for all taxa groups was discharge. For macroinvertebrates and fish, the second most important predictor group was land use, while for macrophytes it was geology.

Present range projections did not diverge in size between macroinvertebrates, fish and macrophytes. Future projections for macroinvertebrates showed significant increments in predicted range size under the LUC scenario ($p < 0.001$), as well as significant decreases in the CC RCP 8.5 scenario ($p < 0.001$). However, both positive and negative effects seem to balance each other out in the combined scenario (LUC + CC RCP 4.5), as no difference were observed. For fish, no significant effect of any scenario on predicted range size was detected.

Concerning uncertainty, no differences were detected between taxonomic groups in the present or future projections (Fig. 2, Table 1). For macroinvertebrates, as well as fish, every scenario projection had a significantly higher coefficient of variation that the present projection (all $p < 0.01$). Uncertainty stemming from the LU scenario was significantly lower than that of the CC RCP 4.5 scenario, but significantly higher than that of the CC RCP 8.5 scenario. The pattern was identical when macroinvertebrates or fish only were considered. When CC and LU were merged into a combined scenario, uncertainty from the resulting scenarios was also balanced out: the LUC + CC RCP 4.5 scenario resulted in significantly lower, and the LUC + CC RCP 8.5 in significantly higher uncertainty than their CC-only counterparts.

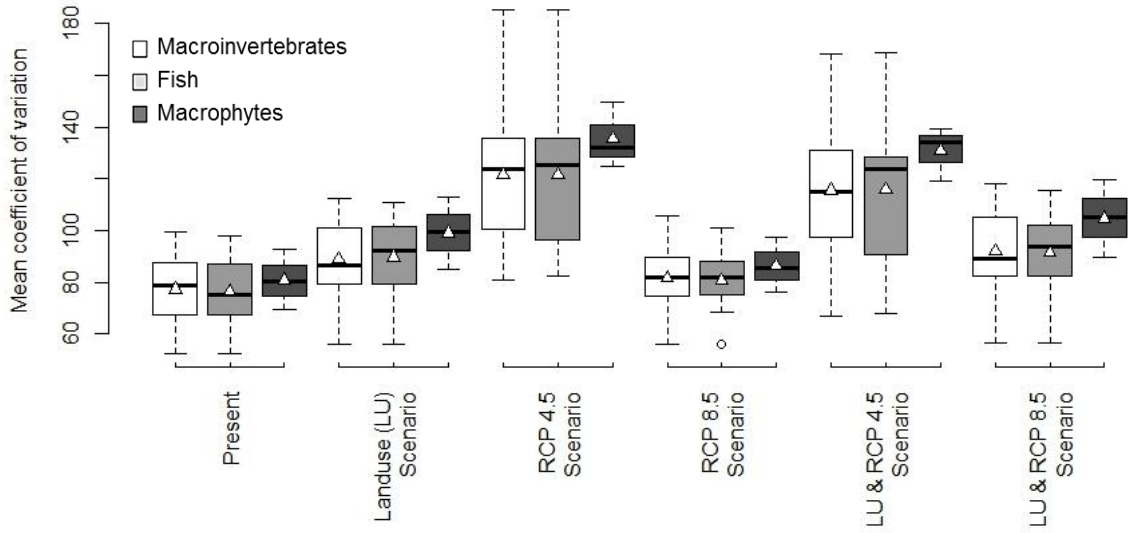


Figure 2: *Coefficients of variance for macroinvertebrates, fish and macrophytes for the present and four scenario projections*

The spatial structuring of uncertainty in the predicted stream network is arranged as a gradient from large to small rivers (Fig. 3, Table 1). In the mainstem (stream of order 4), average uncertainty is smallest, while it is largest in headstreams (streams of order 1). Uncertainty is significantly higher for all scenarios when compared to the present prediction, for the entire community, as well as for macroinvertebrates and fish (all $p < 0.01$). However, the pattern observed and described previously for the single scenarios, persists across space in streams of all sizes (orders 1 through 4): the effect of the CC RCP 8.5 scenario always induces the lowest uncertainty, followed by the LU and the CC RCP 4.5 scenarios.

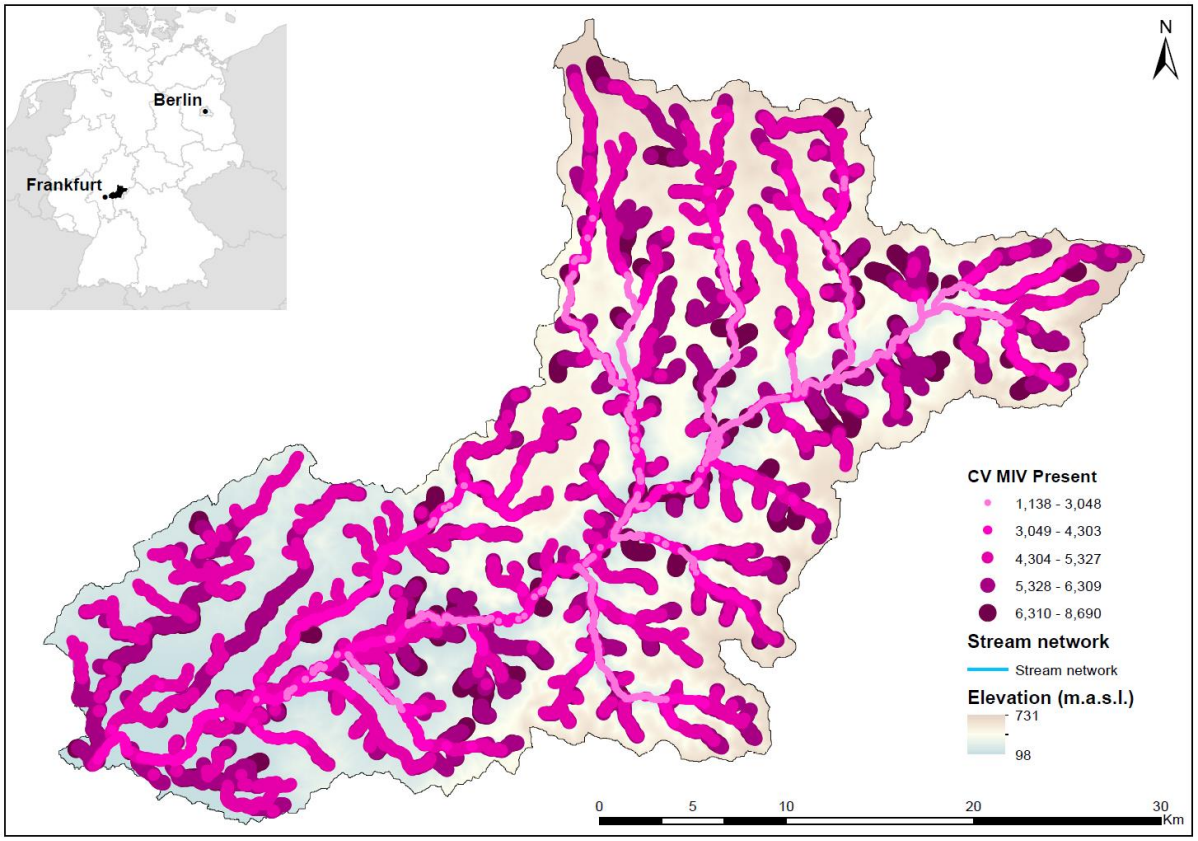


Figure 3: *Spatial distribution of coefficients of variance in the stream network for macroinvertebrates*

Discussion

Models set up in this study performed similarly to previous SDMs for the RMO (Kuemmerlen et al. 2016). The observed high importance of hydrological factors is well established for freshwater ecosystems (Poff et al. 2010) and supports discharge as the most relevant predictor in the models. The importance of hydrological factors for the distribution of freshwater biota has been observed in previous catchment SDMs in the RMO and beyond (Kuemmerlen et al. 2012; Kuemmerlen et al. 2014; Kuemmerlen et al. 2016). The difference in predicted range size for macroinvertebrates between the current and the future scenarios, suggests a higher habitat specialization. More specific habitat suitability requirements could induce smaller predicted ranges and therefore a higher sensitivity to environmental change.

The uncertainty stemming from the individual scenarios was dominant across all taxonomic groups and spatial contexts. Thus, uncertainty seems to be most strongly linked to the environmental predictors used to model and project the distribution predictions. However, it is independent of the type of scenario, as climate alone induced both the highest (RCP 4.5) and the lowest (RCP 8.5) levels of uncertainty in the projections. The environmental data for the 2030 scenario for both of these scenarios (see Table A1 in the Annex) reveal that for RCP 4.5 there was a high variability in the temperature of the driest quarter (Bioclim 09). This is induced by changes in precipitation patterns, leading to strong modifications in the value ranges of this predictor. Thus, choice of predictors can also be an important source of uncertainty in scenario projections.

We found no relationship between taxonomy and uncertainty. This means that uncertainty is probably not dependant taxonomical group and the associated uncertainty stemming from each taxon-specific sampling method. Interestingly, the taxonomic group with the smallest number of taxa, aquatic macrophytes, had the lowest variability in uncertainty across all projections. This suggests, that the uncertainty of a modelled community may be inversely related to the number of species considered.

Table 1: *Mean coefficients of variance for all the stream network and for four different stream orders, for macroinvertebrates, fish and macrophytes*

	Projection	Macroinvertebrates	Fish	Macrophytes
All stream network	Present (pred. area ± SD)	77.3 (12.3)	76.4 (12.3)	80.8 (9.6)
	Landuse (rel. change)	*15.2 ****	*17.3 ****	*22.6*
	RCP 4.5 (rel. change)	*57.1 ****	*59.1 ****	*67.7*
	RCP 8.5 (rel. change)	*5.8 ****	*5.7 ***	*6.8*
	Landuse + RCP 4.5 (rel. change)	*49.5 ****	*51.4 ****	*61.7*
	Landuse + RCP 8.5 (rel. change)	*19 ****	*19.8 ****	*29.6*
Stream order 1	Present (pred. area ± SD)	77.7 (12.4)	76.7 (12.4)	81.1 (9.1)
	Landuse (rel. change)	*15.1 ****	*17.1 ****	*22.9*
	RCP 4.5 (rel. change)	*56.4 ****	*58.2 ****	*67.3*
	RCP 8.5 (rel. change)	*5.8 ****	*6*	*6.6*
	Landuse + RCP 4.5 (rel. change)	*48.8 ****	*50.6 ****	*61.5*
	Landuse + RCP 8.5 (rel. change)	*18.8 ****	*19.8 ****	*29.7*
Stream order 2	Present (pred. area ± SD)	77.2 (12.2)	76.1 (11.9)	80.9 (9.7)
	Landuse (rel. change)	*16.2 ****	*18.7 ****	*23.4*
	RCP 4.5 (rel. change)	*58.2 ****	*60.7 ****	*68.3*
	RCP 8.5 (rel. change)	*6 ****	*5.9 ****	*6.9*
	Landuse + RCP 4.5 (rel. change)	*50.2 ****	*52.7 ****	*61.6*
	Landuse + RCP 8.5 (rel. change)	*19.9 ****	*21.1 ****	*30.2*
Stream order 3	Present (pred. area ± SD)	76.7 (12.4)	75.7 (12)	80 (10.3)
	Landuse (rel. change)	*15.8 ****	*18.2 ****	*23.4*
	RCP 4.5 (rel. change)	*58.1 ****	*60.4 ****	*69.2*
	RCP 8.5 (rel. change)	*5.7 ****	*5.3*	*7.3*
	Landuse + RCP 4.5 (rel. change)	*50.6 ****	*52.8 ****	*63*
	Landuse + RCP 8.5 (rel. change)	*19.8 ****	*20.5 ****	*31.1*
Stream order 4	Present (pred. area ± SD)	76.6 (12.3)	77.2 (13.1)	80.5 (11.3)
	Landuse (rel. change)	*11.9 ****	*12.6 ***	*16.9*
	RCP 4.5 (rel. change)	*56.2 ****	*56.4 ****	*65.8*
	RCP 8.5 (rel. change)	*5.1 ****	*4.1*	*7.2*
	Landuse + RCP 4.5 (rel. change)	*49.7 ****	*49 ****	*61.1*
	Landuse + RCP 8.5 (rel. change)	*15.6 ****	*14.6 ***	*24.3*

There was a constant relationship of stream size and uncertainty across taxa groups and future environmental scenarios. The observed gradient in spatial distribution of the uncertainty may well be a product of the distribution of sampling sites in the stream network: most sites are found in larger streams (order four) and are also most frequently samples (Kuemmerlen et al. 2016). Thus, streams with the thin occurrence information get high levels of uncertainty in predictions.

Uncertainty was balanced out in combined scenarios, indicating that uncertainty can be controlled for, if the uncertainty of individual scenarios is known. This is in line with previous findings, where the effects of LU and CC scenarios projections counter-acted, neutralizing their effects, instead of adding on to each other (Kuemmerlen et al. 2014).

Conclusions

Uncertainty in future scenarios could be related to the magnitude of the forecasted changes in the environment. The proportion of LUC in the RMO is modest, having only a moderate effect on some species, while others remain mostly unaffected. In general, predicted LUC in Germany is very low for the year 2030 (Hoymann and Goetzke 2014), which is very likely to be a realistic estimate. Moreover, the scarcity of land-use models in general and in particular for other, later time periods, is the most important limiting factor in the applicability of SDMs to estimate potential changes and shifts in species distributions.

Uncertainty stemming from climate was the highest for the RCP 4.5 scenario, most likely because of the very high variability in the scenario values for the predictor temperature of the driest quarter. While the predictor itself was implemented correctly, it may have strongly influenced the outcome of the projection in this scenario. Therefore we can conclude that the choice of predictors is also an important factor in scenario projections. Complex predictors such as the so-called bioclim variables are of particular attention.

Uncertainty due to modelling technique was not assessed in this exercise, as it represented an additional level of complexity, for which time was not sufficient. Also, uncertainty due to individual climate models was not assessed. Assessing the effect of individual GCMs would be interesting, in addition to the effect of averaging across all of them

Acknowledgements

The authors are grateful to the following institutions for the data provided: Hessisches Landesamtfür Umwelt und Geologie (HLUG), Deutscher Wetterdienst (DWD), HessenForst. This study was supported by the EU BON project (Building the European Biodiversity Observation Network), a European Union 7th Framework Programme (grant No 308454).

References

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93:1527–1539. doi: 10.1890/11-1930.1

BKG (2011) Bundesamt für Kartographie und Geodäsie, Frankfurt am Main. http://www.geodatenzentrum.de

Domisch S, Araújo MB, Bonada N, et al (2013) Modelling distribution in European stream macroinvertebrates under future climates. Glob Change Biol 19:752–762. doi: 10.1111/gcb.12107

Dormann CF, Elith J, Bacher S, et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Dudgeon D, Arthington AH, Gessner MO, et al (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev 81:163–182. doi: 10.1017/S1464793105006950

Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annu Rev Ecol Evol Syst 40:677–697. doi: 10.1146/annurev.ecolsys.110308.120159

Ficke AD, Myrick CA, Hansen LJ (2007) Potential impacts of global climate change on freshwater fisheries. Rev Fish Biol Fish 17:581–613. doi: 10.1007/s11160-007-9059-5

Giorgetta MA, Jungclaus J, Reick CH, et al (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. J Adv Model Earth Syst 5:572–597. doi: 10.1002/jame.20038

Hijmans RJ (2016) raster: Geographic Data Analysis and Modeling.

HLNUG (2007) Geologische Übersichtskarte von Hessen (GÜK300), 5. überarbeitete, digitale Ausgabe. http://www.hlnug.de

Hoymann DJ, Goetzke DR (2014) Die Zukunft der Landnutzung in Deutschland – Darstellung eines methodischen Frameworks. Raumforsch Raumordn 72:211–225. doi: 10.1007/s13147-014-0290-y

IPCC (2013) Summary for Policymakers. In: Stocker TF, Qin D, Plattner G-K, et al. (eds) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,

Jacob D, Petersen J, Eggert B, et al (2013) EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg Environ Change 14:563–578. doi: 10.1007/s10113-013-0499-2

Keuler, Klaus; Radtke, Kai; Kotlarski, Sven; Lüthi, Daniel (2016) Regional climate change over Europe in COSMO-CLM: Influence of emission scenario and driving global model, Meteorol. Z., 25: 121–136

Kuemmerlen M, Domisch S, Schmalz B, et al (2012) Integrierte Modellierung von aquatischen Ökosystemen in China: Arealbestimmung von Makrozoobenthos auf Einzugsgebietsebene. Hydrol Wasserbewirtsch 56:185–192. doi: 10.5675/HyWa_2012,4_3

Kuemmerlen M, Schmalz B, Cai Q, et al (2015) An attack on two fronts: predicting how changes in land use and climate affect the distribution of stream macroinvertebrates. Freshw Biol 60:1443–1458. doi: 10.1111/fwb.12580

Kuemmerlen M, Schmalz B, Guse B, et al (2014) Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. Ecol Model 277:77–86. doi: 10.1016/j.ecolmodel.2014.01.020

Kuemmerlen M, Stoll S, Sundermann A, Haase P (2016) Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. Ecol Indic 65:122–132. doi: http://dx.doi.org/10.1016/j.ecolind.2015.08.008

Kundzewicz Z, Mata L, Arnell NW, et al (2008) The implications of projected climate change for freshwater resources and their management.

Malmqvist B, Rundle S (2002) Threats to the running water ecosystems of the world. Environ Conserv 29:134–153. doi: 10.1017/S0376892902000097

van Meijgaard, E., L. van Ulft, W. van de Berg, F. Bosveld, B. van den Hurk, G. Lenderink, and A. Siebesma (2008) The KNMI regional atmospheric climate model RACMO version 2.1, Tech. Rep. TR-302.

Merow C, Smith MJ, Edwards TC, et al (2014) What do we gain from simplicity versus complexity in species distribution models? Ecography 37:1267–1281. doi: 10.1111/ecog.00845

Metz M, Rocchini D, Neteler M (2014) Surface Temperatures at the Continental Scale: Tracking Changes with Remote Sensing at Unprecedented Detail. Remote Sens 6:3822–3840. doi: 10.3390/rs6053822

Nelson KC, Palmer MA, Pizzuto JE, et al (2009) Forecasting the combined effects of urbanization and climate change on stream ecosystems: from impacts to management options. J Appl Ecol 46:154–163. doi: 10.1111/j.1365-2664.2008.01599.x

Nix HA (1986) A biogeographic analysis of Australian elapid snakes. In: Longmore R (ed) Atlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra, pp 4–15

Palmer MA, Lettenmaier DP, Poff NL, et al (2009) Climate Change and River Ecosystems: Protection and Adaptation Options. Environ Manage 44:1053–1068. doi: 10.1007/s00267-009-9329-1

Peterson AT, Soberón J (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. Nat Conserv 10:102–107.

Poff NL, Richter BD, Arthington AH, et al (2010) The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshw Biol 55:147–170. doi: 10.1111/j.1365-2427.2009.02204.x

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rocchini D, Hortal J, Lengyel S, et al (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Prog Phys Geogr 35:211–226. doi: 10.1177/0309133311399491

Sala OE, Chapin FS, Iii, et al (2000) Global Biodiversity Scenarios for the Year 2100. Science 287:1770–1774. doi: 10.1126/science.287.5459.1770

Saunders DL, Meeuwig JJ, Vincent ACJ (2002) Áreas Protegidas de Agua Dulce: Estrategias para la Conservación. Conserv Biol 16:30–41. doi: 10.1046/j.1523-1739.2002.99562.x

Singer A, Johst K, Banitz T, et al (2016) Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? Ecol Model 326:63–74. doi: 10.1016/j.ecolmodel.2015.11.007

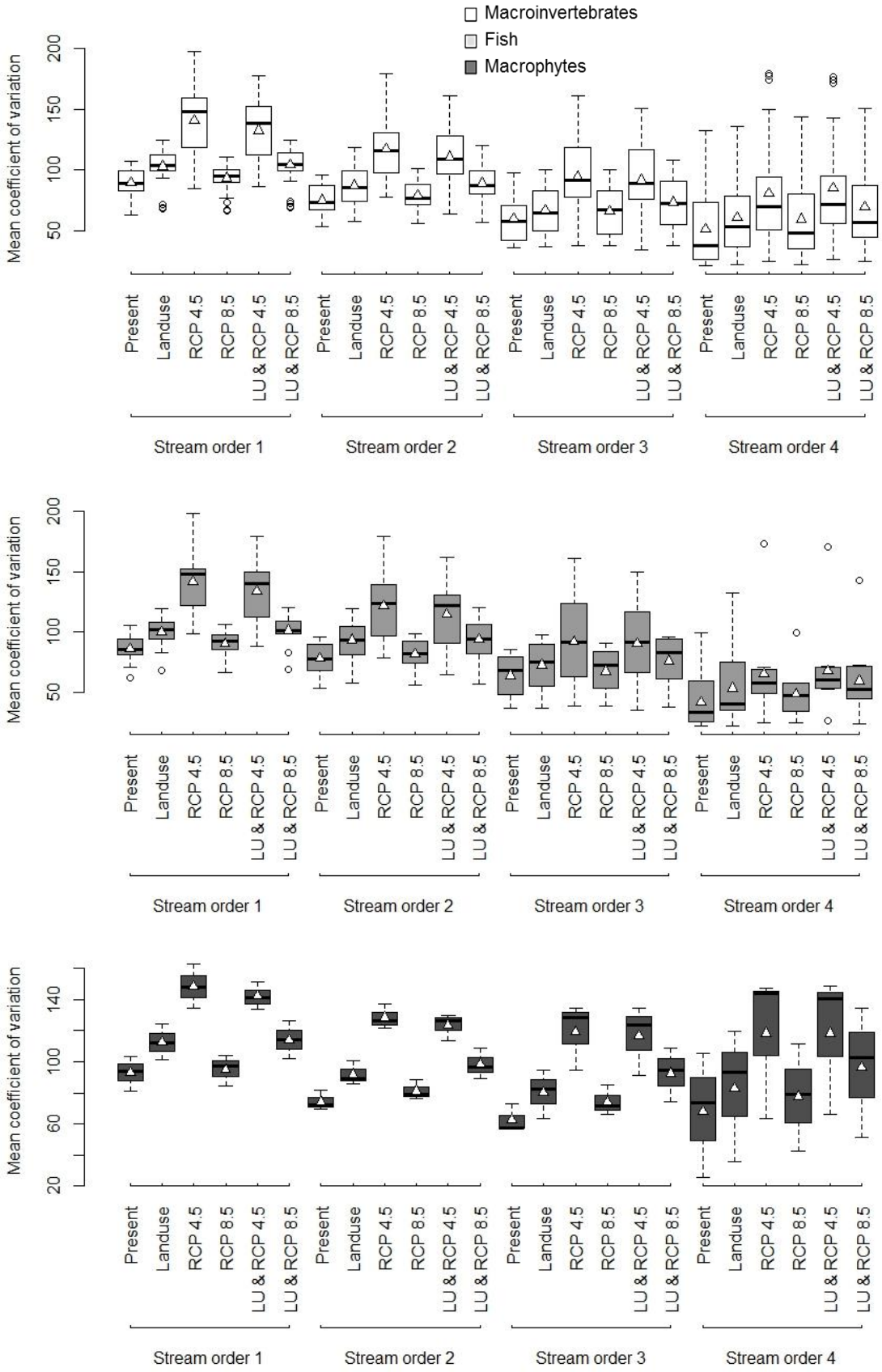
Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future challenges. J North Am Benthol Soc 29:344–358.

Titeux N, Henle K, Mihoub J-B, et al (2016) Biodiversity scenarios neglect future land-use changes. Glob Change Biol 22:2505–2515. doi: 10.1111/gcb.13272

Tonkin JD, Stoll S, Jähnig SC, Haase P (2016) Anthropogenic land-use stress alters community concordance at the river-riparian interface. Ecol Indic 65:133–141. doi: 10.1016/j.ecolind.2015.08.037

Vitousek PM (1992) Global Environmental Change: An Introduction. Annu Rev Ecol Syst 23:1–14.

Supplementary Information



Figures A1a-c: *Mean coefficient of variation by taxon group and stream order*

Table A1: *Mean and standard deviation of environmental predictors for the present and scenario projections. Mean differences between scenario and present values are shown in relative terms.*

	Present		RCP 4.5 Scenario 2030			RCP 8.5 Scenario 2030			LU Scenario 2030		
	Mean	Std. Dev.	Mean	Std. Dev.	Mean diff. (%)	Mean	Std. Dev.	Mean diff. (%)	Mean	Std. Dev.	Mean diff. (%)
Mean annual temperature (°C)	9.59	0.66	10.04	0.66	4.7%	10.11	0.65	5.4%	-	-	-
Mean diurnal temperature range (°C)	9.06	0.19	9.06	0.19	0.0%	9.06	0.19	0.0%	-	-	-
Mean temperature warmest quarter (°C)	1.10	1.33	15.21	0.69	12173.9%	1.92	1.32	693.9%	-	-	-
Estimated discharge (m^3/s)	1.48	4.18	1.58	4.46	6.9%	1.56	4.41	6.0%	-	-	-
Urban LU in subcatchment (%)	0.05	0.06	-	-	-	-	-	-	0.05	0.06	6.2%
Industrial LU in subcatchment (%)	0.01	0.04	-	-	-	-	-	-	0.01	0.04	3.3%
Agricultural LU in subcatchment (%)	0.21	0.22	-	-	-	-	-	-	0.21	0.22	-3.4%
Pasture LU in subcatchment (%)	0.23	0.18	-	-	-	-	-	-	0.23	0.18	-1.2%
Natural LU in subcatchment (%)	0.03	0.03	-	-	-	-	-	-	0.02	0.03	-11.8%

7.6. Anticipating species distributions: Handling sampling effort bias under a Bayesian framework

Partner involved: FEM

Project leader: Duccio Rocchini, Carol X. Garzon-Lopez, Matteo Marcantonio

Manuscript title

Anticipating species distributions: Handling sampling effort bias under a Bayesian framework [published in *Science of the Total Environment*]

One sentence summary

We propose an innovative method to i) map sampling effort bias using cartogram models and ii) explicitly consider such uncertainty in the modeling procedure under a Bayesian framework, which allows the integration of multilevel input data with prior information to improve the anticipation species distributions.

Authors

D. Rocchini[1], C.X. Garzon-Lopez[2], M. Marcantonio[1,3], V. Amici[4], G. Bacaro[5], L. Bastin[6,7], N. Brummit[8], A. Chiarucci[9], G.M. Foody[10], H.C. Hauffe[1], K.S. He[11], C. Ricotta[1], A. Rizzoli[1], and R. Rosà[1]

Affiliations

[1] *Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, Via E. Mach 1, S. Michele all'Adige 38010, TN, Italy*
[2] *UR "Ecologie et Dynamique des Systemes Anthropises" (EDYSAN, FRE 3498 CNRS), 9 Universite de Picardie Jules Verne, 1 rue des Louvels, Amiens Cedex 1 FR-80037, France*
[3] *Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, USA*
[4] *Department of Life Sciences, University of Siena, Via P.A. Mattioli 4, Siena 53100, Italy*
[5] *Department of Life Sciences, University of Trieste, Via L. Giorgieri 10, Trieste 34127, Italy*
[6] *School of Computer Science, Aston University, UK*
[7] *Joint Research Centre, European Commission, Italy*
[8] *Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[9] *BIGEA, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum, University of Bologna, Via Irnerio 42, Bologna 40126, Italy*
[10] *University of Nottinhgam, University Park, Nottingham NG7 2RD, UK*
[11] *Department of Biological Sciences, Murray State University, Murray, KY 42071, USA*
[12] *Department of Environmental Biology, University of Rome "La Sapienza", Rome 00185, Italy*

*Correspondence to: duccio.rocchini@fmach.it

Abstract

Anticipating species distributionsin space and time is necessary for effective biodiversity conservation and for prioritizing management interventions. This is especially true when considering invasive species. In such a case, anticipating their spread is important to effectively plan management actions. However, considering uncertainty in the output of species distribution models is critical for correctly interpreting results and avoiding inappropriate decision-making. In particular, when dealing with species inventories, the bias resulting from sampling effort may lead to an over- or under-estimation of the local density of occurrences of a species. In this paper we propose an innovative method to i) map sampling effort bias using cartogram models and ii) explicitly consider such uncertainty in the modeling procedure under a Bayesian framework, which allows the integration of multilevel input data with prior information to improve the anticipation species distributions.

Introduction

Anticipation is an important topic in ecological fields such as food science (Lobell et al., 2012), community ecology (Keddy, 1992), species distribution modeling (Willis et al., 2009), landscape ecology (Tattoni et al., in press), and biological invasion science (Rocchini et al., 2015). Anticipatory methods are also crucial for developing effective management practices to deal with invasive species (Rocchini et al., 2015).

Invasive species can modify the structure and functioning of ecosystems, altering biotic interactions and homogenizing previously diverse plant and animal communities over large spatial scales, ultimately resulting in a loss of genetic, species and ecosystem diversity (Winter et al., 2009). The annual economic impact of invasive species has been estimated at over 100 billion dollars just within the USA (NRC, 2002), an order of magnitude higher than those caused by all natural disasters put together (Ricciardi et al., 2011); some authors go as far as to claim that the economic impact of invasive species is incalculable (Mack et al., 2000).

Given the massive negative economic and ecological effects of invasive species, a robust method for predicting species' distributions is crucial for an early assessment of species invasions and effective application of appropriate management actions (Malanson and Walsh, 2013).

Investigating how biodiversity is distributed spatially and temporally across the globe has long been a central theme in ecology (Gaston, 2000) and the methods developed to answer this question have become key tools for biodiversity monitoring (Ferretti and Chiarucci, 2003). For example, species distribution models (SDMs) have been used to map the current distribution of a single species (Rocchini et al., 2011), model the potential distribution of native and invasive species (Rocchini et al., 2015), investigate the statistical performance of different models to infer the distribution of species under various ecological conditions (Elith and Graham, 2009; Guisan and Zimmermann, 2000), test the transferability in space of modeled distribution patterns (Heikkinen et al., 2012; Randin et al., 2006), predict long term changes to species distributions (Pearman et al., 2008) and make inferences on future biodiversity scenarios (Engler et al., 2009; Pompe et al., 2008), evaluate the potential of satellite imagery bands as predictors of biodiversity patterns (Mathys et al., 2009), analyse spatial autocorrelation in species distributions (Carl and Kühn, 2007; Dormann, 2007), and understand biogeographical patterns (Sax, 2001).

In combination with remote sensing products (e.g. Feilhauer et al. 2013, Rocchini, 2007) and current global data sets on in situ species observations, SDMs have become the method of choice for monitoring biodiversity at multiple spatial and temporal scales. However, the strength of this combination depends on the careful selection and application of integrative modeling approaches, in combination with a thorough assessment of uncertainty in both data inputs and modeling methods.

Reliable anticipation of species invasions depends on the quality of input data on one hand and robustness of the predictive SDM on the other. As an example, Rocchini et al. (2011) demonstrated theoretically that input data arising from biased species distribution maps could potentially lead to unsuitable management strategies. In addition, Elith and Leathwick (2009) demonstrated that, given the same input data set, different SDMs might lead to dissimilar results (see also Bierman et al., 2010, Manceur and Kuhn, 2014).

The aim of this manuscript is to propose coherent and straightforward methods to explicitly account for uncertainty when mapping species distributions in the light of anticipating the spread of invasive species. In particular we will cover i) explicitly mapping uncertainty in sampling bias, ii) mitigating uncertainty in data through prior beliefs and Bayesian inference and iii) reporting uncertainty in species distribution maps through Markov Chain Monte Carlo methods. The findings of this manuscript should be of particular interest to landscape managers and planners attempting to predict the spread of species and deal with errors in species distribution maps in a straightforward manner.

2. Mapping input uncertainty related to sampling effort bias

In anticipating species distributions a first step is to ensure that the information indicating where species are present is bias-free or, at least, that the uncertainty of input data is explicitly taken into account in further modeling steps.

One of the main problems with field data on species distributions is related to "sampling effort bias" (Rocchini et al., 2011), namely the bias inherent in some areas being under-sampled with respect to others. Quantifying and mapping the uncertainty derived from variation in the number of observations due to sampling effort can be achieved using cartograms (Gastner and Newman, 2004), in which the shape of spatial objects (e.g. polygons and cells) is directly related to a determined property, in our case to uncertainty.
Cartograms build on the standard treatment of diffusion theory by Gastner and Newman (2004), in which the current spatial density of a population is given by the population under study, respectively, at position r and time t.

$$J = v(r,t)p(r,t) \quad (1)$$

Cartograms facilitate the visualization of spatial uncertainty in the data by varying the size of each polygon according to the density of information contained (e.g. number of observations and variation). As an example, we show a cartogram of the distribution of *Abies alba* Miller overlapping a grid to the set of records obtained from the Global Biodiversity Information Facility (GBIF, http://www.gbif.org, Fig. 1). GBIF offers free and open access to hundreds of millions of records from over 30,000 species datasets which are collated from around the world and stored with a common Darwin Core data standard. The cartogram was developed using the free and open source software ScapeToad (http://scapetoad.choros.ch/). Since cells with a higher number species occurrences might be biased by the effort spent visiting them, in

Fig. 1, the shape of each cell is determined by the number of times it was visited (i.e. number of different dates recorded in GBIF for the species in that cell). From now on, we will refer to this as sampling effort. The colour represents the spatial distribution (density of occurrences, sensu) of the species in each cell. Therefore, cartograms allow uncertainty to be shown explicitly in a straightforward manner. Furthermore, sampling effort might be considered as a variable in the SDM procedure, as described in the next section.

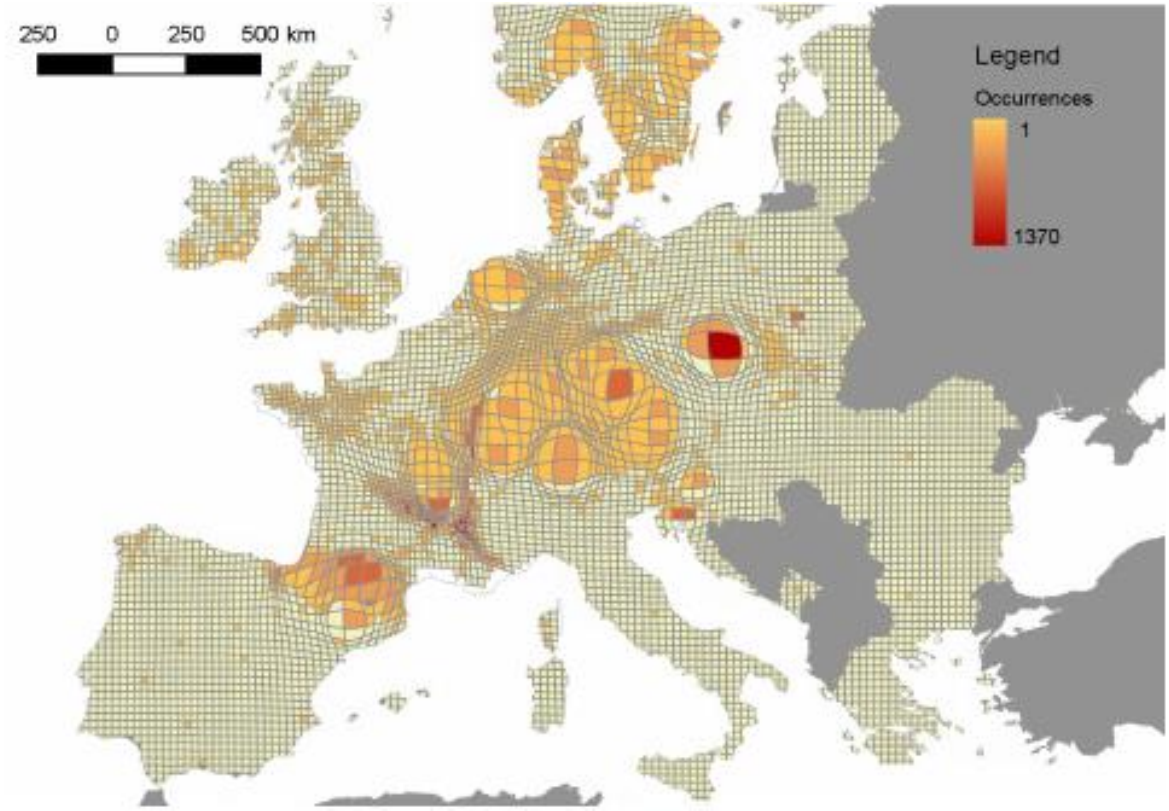


Figure 1 *Cartogram representing the sampling effort bias (cell distortion) of the GBIF dataset related to Abies alba. This species is not native in Northern Europe, although it is widely cultivated as a timber tree, as thus present in the GBIF dataset.*

3. Accounting for input uncertainty in the modeling procedure: multi-level models, prior beliefs and probability distribution surfaces

Species observation records are often heterogeneous and incomplete because, for example, they are unevenly distributed by year or area, or were collected by different field operators. In addition, there is wide variation in recording behaviours.

GBIF is a classic example of such heterogeneity: GBIF data is opportunistically gathered from a mixture of systematic surveys and volunteer projects, and the intensity of publishing effort is strongly influenced by the membership of the organisation. In terms of geographic coverage, GBIF contains plentiful data from Northern Europe and America, parts of Latin and Central America, South Africa, Australiaand Oceania – but by contrast, there are significant gaps in other regions, and there is a large variation in sampling effort even between neighbouring European countries (see Appendix 1, Fig. S1). This heterogeneity makes it difficult to estimate the underlying variable (actual species presence and density of

occurrences) and potentially has an enormous impact on the information content of any one species observation or set of observations (Isaac and Pocock, 2015). This paper proposes methods by which ancillary knowledge about a species and its environment might be exploited in a Bayesian framework to increase that information content.

Multi-level models can be essential for detecting (spatially) clustered data by considering the variation between groups (clusters). This approach is more efficient and powerful than standard linear modelling techniques as it provides a coherent and flexible method for modelling the effects of sampling variation and allows uncertainty to be elegantly accounted for at all levels of data structure (Gelman and Hill, 2006).

Furthermore, environmental variables with different spatial or temporal resolution (i.e., country, regional or pixel level) are often used as predictors in SDMs. Multi-levels models can simultaneously and coherently incorporate multi-level predictors allowing effects to be modelled at the appropriate scale (Gelman and Hill, 2006). Hierarchical models are naturally handled using Bayesian methods, which provide intuitive and direct estimates of uncertainty around parameter estimates (Link and Sauer, 2002).

Despite tremendous effort by ecologists, collecting unbiased and reliable data on the presence of species in a determined area/time to assess their potential distribution through SDMs is sometimes not feasible since systematic field work is inherently expensive, time-consuming, and often involves logistical hurdles, if the species under study is, for example, rare, elusive, inhabits remote areas, or is in transitional equilibrium with its ecological niche (as is the case with invasive species). Even for less problematic species, presence/absence data may also be distorted by several potential flaws, such as sampling errors and subjectivity. As a result, SDM outputs may show high uncertainty and be difficult to interpret, jeopardizing their utility in conservation applications. However, besides the availability of observation data directly exploitable for modeling purposes, there is a wider set of ecological data that can be used in SDMs, the so called “prior knowledge”. This data is very often neglected and comprises information represented in different formats; for example, previously conducted experiments, scientific literature on the studied species or similar species, or even as “prior beliefs” (basic ecological principles). Bayesian inference allows basic ecological principles and prior data to be incorporated in a straightforward manner with potential cost-effective consequences in increasing confidence of SDMs (Bierman et al., 2010; Manceur and Kuhn, 2014; McCarthy and Masters, 2005). The prior information needs to be translated into a probability distribution, which is then combined under Bayes' rule with the likelihood information contained in the original data to estimate a “posterior belief” or posterior probability distribution (PPD). The contribution of the prior and the data to the posterior distribution depends on their relative precision, with the more precise of the two having the greatest effect. A prior distribution can be non-informative (flat prior), mildly informative (vague prior) or infor mative (strong prior). In any case, the prior must be clearly described and justified according to the context under investigation (Kruschke, 2015).

The result of the interaction between the likelihood of the data and the prior distribution is itself a probability distribution (posterior probability distribution or PPD). In an SDM, the advantage of having model parameter estimations expressed as probability distributions, and not as point estimation of the mean, is that the predicted suitability of the species in each prediction unit (pixel) is itself a probability distribution. The suitability of the PPD in each spatial unit represents the uncertainty of the prediction in that unit. This uncertainty is stored in the Markov Chain Monte Carlo (MCMC) model and can be re-used in future modeling exercises that, for example, use a different set of data.

As an example, we applied a multi-level logistic regression with Bayesian inference to model the distribution of *Abies alba* in Europe. We chose this species due to its well known autoecology and actual distribution in Europe (Farjon, 1998; Gazol et al., 2015; Tinner et al., 2013). We derived 44375 *Abies alba* presence records from the GBIF database, as points in vector format (see Appendix 1, Figs. S3 and S4). We generated an equal number of pseudoabsences using the following strategy: we selected random points a) within areas where conifers have been sampled (conifer occurrences in the GBIF dataset) to pick the same areas that have been surveyed using the sampling protocol used to record *Abies alba* presences, b) outside dry climatic zones (e.g. Mediterranean climate) derived from the Koppen-Geiger climatic zones map (Koppen and Geiger, 1930) where this species is not found and c) outside a radius of 100 m around the presence points to avoid overlap with presence points.

We generated an equal number of absence locations at areas within which conifers have been sampled (conifer occurrences in the GBIF dataset) and outside a 100 m radius from the presence points and the from the Koppen-Geiger climatic zones map.

To select the predictor variables, we performed a literature review on the ecology of the species (Aussenac, 2002; Gazol et al., 2015; Rolland et al., 2009; Tinner et al., 2013; Wolf, 2003). Hence, we relied on three different datasets by selecting i) the annual mean temperature (Bio1), and mean diurnal temperature range (Bio2) obtained from the WorldClim dataset (Hijmans et al., 2005), ii) radiation seasonality (Bio23) and the annual mean moisture index (Bio28), obtained from the CliMond dataset (Kriticos et al., 2012), and iii) the number of wet days during summer and frost days during winter (and early spring) derived from the wet-days and ground-frost data in the climate research unit dataset (Mitchell et al., 2004) (see Fig. 2). Considering sampling effort as a predictor, the sampling of the GBIF dataset is clearly opportunistic. As a result, the unevenness of sampling effort is particularly evident, with the Northern European region being more sampled than other European regions (see Appendix 1, Fig. S1). This bias in GBIF data could generate unreliable predictions.

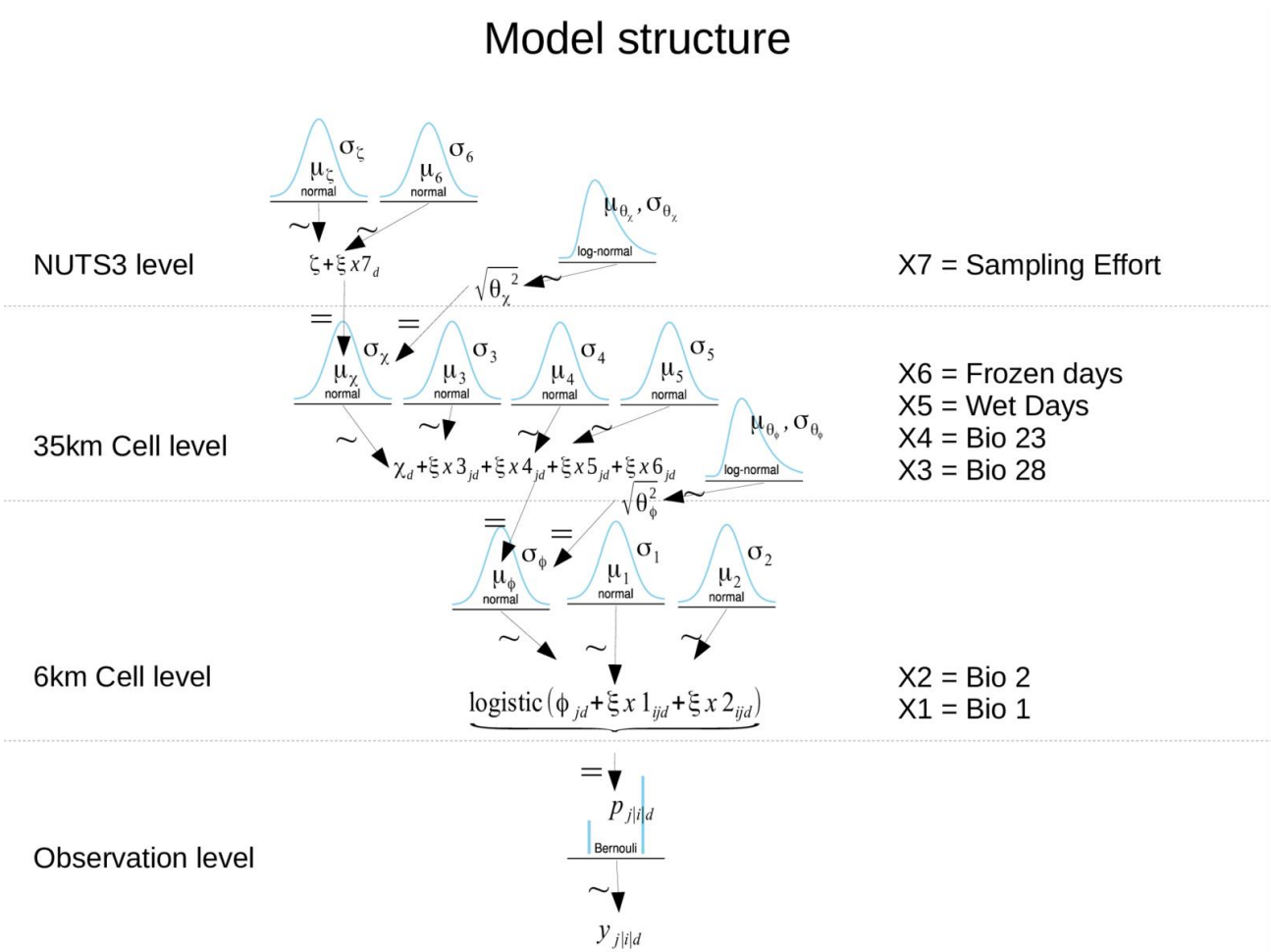


Figure 2 *Multi-level model represented through a pictogram. To select the predictor variables, we performed a literature review on the ecology of the species by finally selecting radiation seasonality (Bio23), the annual mean moisture index (Bio28), the number of wet days during summer and the frost days during winter and early spring, the annual mean temperature (Bio1), the mean diurnal temperature range (Bio2). Sampling effort was calculated as the diversity of dates of survey recorded in the GBIF dataset per each NUTS3 country. Refer to the main text for additional information on the source of each set.*

The clustering of GBIF data mainly derives from differences in surveys at national and subnational level (Appendix 1, Fig. S1). Thus, the sampling effort was derived as the number (richness) of dates of survey recorded in the GBIF dataset per polygon of the official administrative division of European countries using the Nomenclature of Territorial Units for Statistics level 3 (NUTS 3).

We built a multi-level model to take into account the different resolution of the predictor variables (Fig. 2) and the differential sampling effort was used to re-scale the precision of the likelihood at pixel level, multiplying the scaled sampling effort by the standard deviation of the Gaussian likelihood. As a result, the likelihood estimate of pixels in regions with a higher number of samples was expected to be more precise. The theoretical model (Fig. 2) was coded in JAGS language and run in JAGS 4.2.0 through R (Team, 2016) using the R2jags (Su and Yajima, 2016) and CODA (Plummer et al., 2006) packages. In order to allow reproducibility (Rocchini and Neteler, 2012) of our approach we have included the complete R code in Appendix 2.

As previously stated, in heterogeneous datasets like the GBIF set, the sampling effort in a certain region may be correlated with the presence of the species under study. Therefore, a more highly sampled region should have also a higher probability of hosting the species. However, our data showed a weak sampling effort signal, with a high number of very low-sampled regions showing presence of *Abies alba*. This may result from errors, or low numbers of records not being representative of the distribution of the species under study. Therefore, we applied uninformative priors (μ = 0,SD = 1/10 −2) for all the predictors but not for sampling effort, whose prior distribution $p(\theta)$ was given three different sets of parameters:

$$p(\theta) = \begin{cases} dnorm(0, 1/10^{-2}), & \text{uninformative prior.} \\ dnorm(1, 10), & \text{mild positive prior.} \\ dnorm(5, 5), & \text{strong positive prior.} \end{cases} \quad (2)$$

Such distributions were chosen as examples under the hypothesis that *i)* data alone were enough to account for heterogeneity in sampling effort; *ii)* a mildly informative (vague) prior knowledge about the positive correlation of sampling effort was useful for improving the model; *iii)* imposing strong prior knowledge on the positive influence of the prior would improve the model output. These three hypotheses were translated in three models that shared the same structure (Fig. 2) except for the prior distribution imposed on sampling effort. All the predictors were scaled and centered in order to improve the efficiency of the MCMC process. PPDs for all parameters were sampled from each of two chains with 10,000 MCMC iterations using 1000 burn-in and 1000 adaptation iterations, with a thinning set of 20. Convergence was assessed by the statistic of Gelman and Rubin (1992). Each model was then used to estimate the suitability PPDs in each pixel of the study area. The parameter estimates for the three models will show if different prior belief on the role of sampling effort changed the model parameter estimates. Furthermore, the Deviance Information Criterion (DIC, see Spiegelhalter et al., 2014) was used to assess the model with the best predictive power.

The posterior probability distributions (PPDs) of model parameters for the three models (with different priors on sampling effort, see Eq. 2) are reported in Fig. 3. All the models agreed on the direction and effect size of the predictors (Fig. 3). Credible effects (no intersection

with 0 in Fig. 3) were attained for those variables directly related to temperature. In particular, annual mean temperature (Bio1and Bio12) and radiation seasonality (Bio23) showed negative effects while mean diurnal temperature range (Bio2) showed positive effects (Figs. 3 and 4). The negative credible effect of Bio12 implies that the relationship between the probability of presence (suitability) of *Abies alba* and annual mean temperature has a "bell shape", by rising slowly to the left of the annual mean temperature average (7.8 °C) and decreasing rapidly when on its right (Fig. 4). On the contrary, the distribution of wet days, annual mean moisture index (Bio28) and frost days included 0, showing a non-credible effect on the presence of *Abies alba*.

The sampling effort coefficient changed heavily between models. In the first model with an uninformative prior, the coefficient average was slightly negative but with its high density interval comprising 0 (Fig. 3). Therefore we concluded that according to the data the sampling effort had a non-credible effect. In the second model (Fig. 3) a mildly informative positive prior affected the estimate of the parameters, but yet was not enough to derive a credible effect of the prior estimate. In the last model, the strong informative prior pulled the estimation of sampling effort coefficient towards positive values. This showed that, according to the data and to the "prior knowledge", the sampling effort was positively affecting the probability of presence of *Abies alba*.

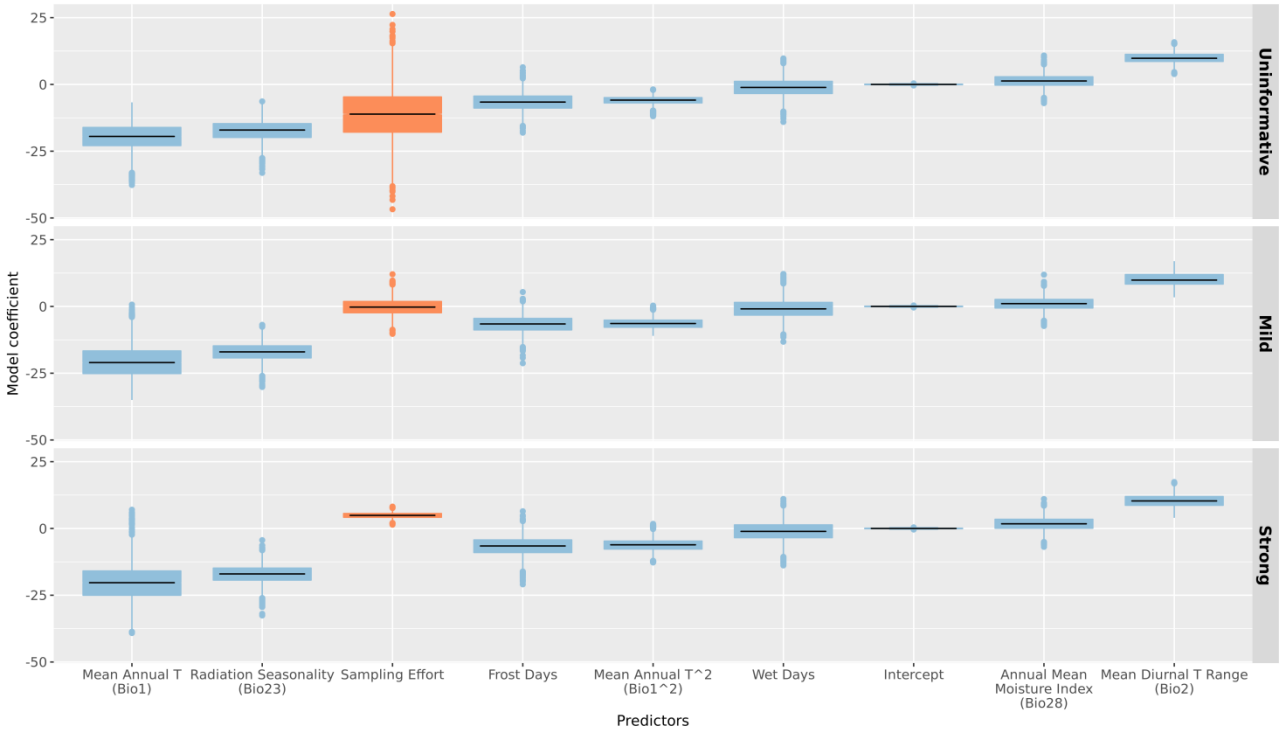


Figure 3: *Boxplots of the β coefficient PPDs for the three models (in the three figure facets). Each box represents the 1st and 3rd quartiles of a coefficient distribution, the black horizontal line the distribution median, the whiskers the limits of the 1.5*interquartile range, while the filled circles represent the outlying points. If whiskers did not overlap 0 we inferred as "credible effect". We showed in red the boxplots reporting the distribution of the β coefficient of the sampling effort. It is clear that the major difference among models was related to the precision of sampling effort, which increased passing from the model with an uninformative prior on sampling effort, through that with a mild prior, reaching its highest value in the model with a strong prior.*

In summary, the model with the strong prior showed an improved precision of sampling effort, basically maintaining that of the others (Fig. 3). Based on this and since the DIC did not show differences for the strong prior-model with respect to the uninformative prior-model (Table 1, $\delta DIC \leq 4$, see Burnham and Anderson, 2002), we further focused on the model with a strong prior to build the output distribution map. The resulting potential niche distribution of *Abies alba* is thus shown in Fig. 5.

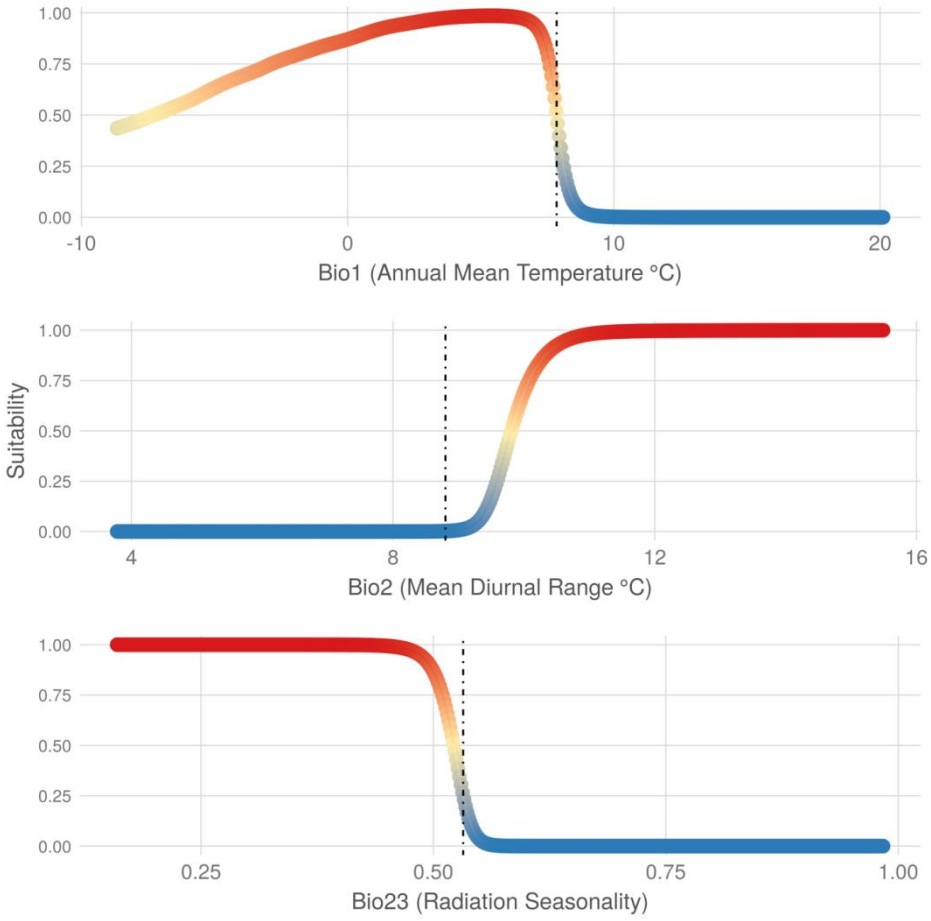


Figure 4: *In this figure the average probability of presence (suitability) of Abies alba is plotted against the three variables with the highest average coefficient effect size in the model (top: range of annual mean temperature Bio1, middle: mean diurnal range Bio2, bottom: Radiation Seasonality or Bio23). The relationship between the probability of presence (suitability) of Abies alba and annual mean temperature has a "bell shape", rising slowly moving from the left of the study area average (7.8 °C), peaking just before the average and decreasing rapidly when on its right. The shape of the relationship between the probability of presence and the mean diurnal temperature range is inverted. A low diurnal temperature range is associated with a low suitability while a wide temperature variability is associated with high suitability. The highest suitability is reported for Bio2 values higher than 11 °C. The Radiation Seasonality (the standard deviation of the weekly solar radiation estimates expressed as a percentage of the mean of those estimates) shows a negative pattern with respect to suitability. Areas with a very high average difference in solar radiation during the year (i.e. Northern Europe) are reported as weakly suitable for Abies alba. All the curves were obtained varying the value and the model coefficient of Bio1, Bio2 and Bio23 while keeping the values of the other predictors at their average. As reported in the main text, this results as well as that in Figure 5 is derived from the model with a strong prior on sampling effort.*

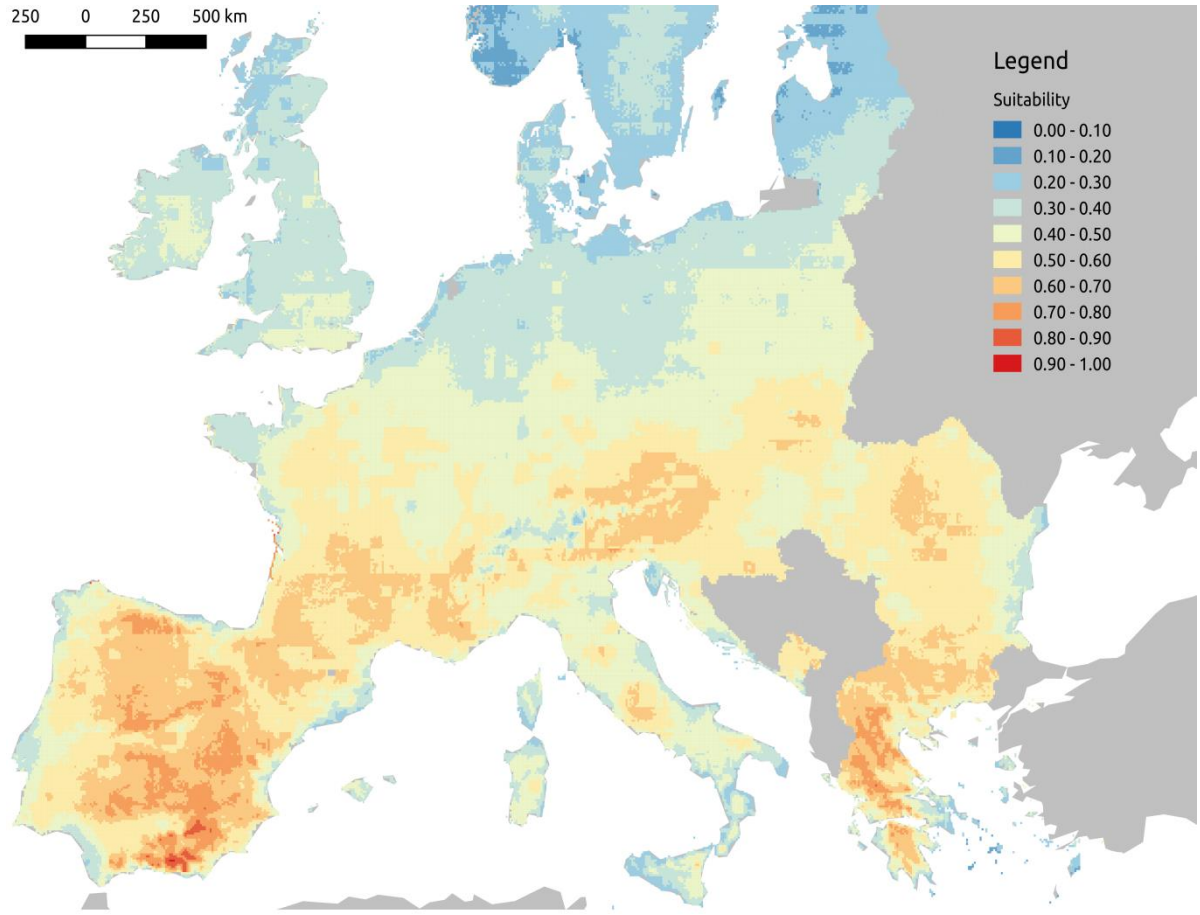


Figure 5: *Abies alba suitability distribution as derived from the multi-level model with strong prior on sampling effort. The pixel value is the average of the PPDs for that pixel.*

Table 1 *Deviance Information Criterion (DIC) used to assess the prior with the best predictive power. Notice that dDIC <= 4 using an uninformative prior and a strong prior on sampling effort. Therefore, a strong prior allowed us to decrease uncertainty and maintain high model quality. Refer to the main text for additional information.*

Model	DIC	Gelman diagnostic	Burn In	Iterations	Chains
Uninformative prior	1938	1.13	2000	10000	2
Mild prior	2133	1.15	2000	10000	2
Strong prior	1940	1.22	2000	10000	2

Discussion

In this paper, we have demonstrated the importance of i) mapping uncertainty derived from varying sampling effort and ii) considering it in an explicit manner in order to anticipate species' potential distributions. We have provided a case study with a plant species widespread throughout Europe (*Abies alba*) where the observed data (Fig. 1) and the modeled potential niche (Fig. 5) differed mainly because of tree plantations recorded in the GBIF dataset. For example, Northern Europe was shown to be unsuitable for the natural spread of

the species in our Bayesian model (Fig. 5), as well as in previous studies on the distribution of the species (e.g. the European Forest Genetic Resources Programme, http://www.euforgen.org/, see Appendix 1, Fig. S2), corroborating our results. However, it appeared to be present in the GBIF field-based dataset (Fig. 1, see also Appendix 1, Fig. S3), mainly because of human-related conifer plantations.

Notably, when we associated a stronger prior to sampling effort, model coefficient estimates had lower uncertainty, and in addition, the model DIC did not differ from the model with the uninformative prior. Therefore, a strong prior allowed us to decrease uncertainty and maintain high model quality ($\delta DIC \leq 4$, see Burnham and Anderson, 2002).

We have shown that multilevel models coupled with Bayesian inference can be used to account for variability in sampling effort, integrating external data on prior knowledge with species observations, to model species distribution more accurately and with higher certainty than previous methods. The priors considered in the reported case study were only examples generated here to illustrate how the precision of parameter estimates can potentially be increased using prior knowledge about the system under study. However, in order to have scientifically sound results, the priors considered should obviously be fully justified and rooted in ecological theory.

Anticipating species potential distributions based on prior information (Bayesian modeling) can help to predict the potential future spread of a species in space (and time) in a robust manner (Bierman et al., 2010; Manceur and Kuhn, 2014). Using sampling effort bias among priors was important in our case since it allowed such uncertainty to be considered explicitly in the model. This can help to accommodate the error rate directly into the modeling procedure.

Hence, calibrating models conditioned on previous knowledge and/or observations might be feasible when relying on a Bayesian framework in which

$$P(Y \mid H) \tag{3}$$

where P = the probability of occurrence of patterns Y given a hypothesis H is substituted by

$$P(H \mid Y) \tag{4}$$

i.e. the probability P that a hypothesis H is true in light of the available data.

Bayesian statistics have long been used in independent scientific disciplines and topics such as trait loci mapping (Ball, 2001), environmental science (Clark, 2005), machine learning approaches in computer science (Dietterich, 2000), classification of remotely-sensed images (Goncalves et al., 2009), conservation genetics (Bertorelle et al., 2004), statistical algorithm evelopment (Hoeting et al., 1999) and sampling strategies (Mara et al., 2016).

In the framework of ecological patterns and processes, Ellison (2004) makes an explicit quest for using known information to build a model, relying on prior rather than posterior probabilities. This reinforces the view of Ginzburg et al. (2007) that biology should constrain mathematical constructions. Quoting the authors, “While mathematics provides an incredibly

vast set of possible equations, logic dictates that only a small subset of these equations can represent a given ecological phenomenon. A large number of constructions, while mathematically sound, should be excluded based on their inconsistency with biology."

This is especially true when the results of model construction impact decision-making, which could be more focused and effective if uncertainty was explicitly taken into account based on previous literature regarding the main drivers that shape the distribution of species (Ellison, 1996). Our approach reduces the danger of relying on misleading predictions of alien species invasions with high model errors, which are hidden or unrecognizable using previous approaches (Rocchini et al., 2015).

In the framework of species distribution modeling it has been demonstrated that prior probabilities in the observation of a certain species might improve model performance. This is true at various hierarchical levels, from species to entire communities. Thus, applying Bayes' theorem to predict values at a certain site might thus allow known environmental properties to be accounted for. If Bayesian models do not outperform other modeling techniques, they at least better reflect the theory under the realized niche of a certain species. A number of examples are provided in Guisan and Zimmermann (2000), modelling different plant species in different habitat types.

Conclusions

In the light of the importance of anticipating species future distributions, especially for economically important invasive species, it is crucial to detect those areas into which such a species might be expected to disperse. Anticipating their spread based on the suitability of environmental conditions can lead to more effective management strategies, allowing timely actions to be initiated and preventing further spread (Rocchini et al., 2015).

This can be summarized by the following equation:

$$Decision = \begin{pmatrix} <E_m|>I & <E_m|<I \\ >E_m|>I & >E_m|<I \end{pmatrix} \quad (5)$$

In this case, a high (or low) invasion rate I might be related to high or low error Em in the output model being observed by decision makers. The most dangerous situation is when a low predicted invasion rate is related to a high error in the modeling procedure. In this case decision makers might underestimate the effort against the likelihood of invasion that, from the species distribution map, is suspected to be low.

In this paper we have demonstrated the power of incorporating sampling bias into the model being used by relying on prior probabilities of distribution of a plant species widely spread in Europe. We believe this is a good example to further encourage species distribution modelers and environmental planners and conservationists to account for uncertainty and bias in the sampling effort in anticipating the spatial spread of species, instead of relying on distribution maps with potentially hidden uncertainty.

Acknowledgments

We are particularly grateful to the Handling Editor Rocco Scolozzi and to two anonymous reviewers who provided useful insights which improved a first draft of this manuscript. We thank Ingolf Kühn for precious suggestions. DR was partially supported by the EU BON (Building the European Biodiversity Observation Network) project, funded by the European Union under the Seventh Framework Programme (Contract No. 308454), by the ERANET BioDiversa FP7 project DIARS, funded by the European Union and by the LIFE project Future For CoppiceS.

References

Alba-Sanchez, F., Lopez-Saez, J.A., Pando, B.B., Linares, J.C., Nieto-Lugilde, D., Lopez-Merino, L., 2010. Past and present potential distribution of the Iberian Abies species: a phytogeographic approach using fossil pollen data and species distribution models. Divers. Distrib. 16, 214–228. http://dx.doi.org/10.1111/j.1472-4642.2010.00636.x.

Aussenac, G., 2002. Ecology and ecophysiology of circum-Mediterranean firs in the context of climate change. Ann. For. Sci. 59, 823–832.

Ball, R.D., 2001. Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. Genetics 159, 1351–1364.

Barbosa, A., 2015. Fuzzysim: applying fuzzy logic to binary similarity indices in ecology.
Methods Ecol. Evol. in press.

Beck, J., Ballesteros-Mejia, L., Buchmann, C.M., Dengler, J., Fritz, S.A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.K., Winter, M., Dormann, C.F., 2012. What's on the horizon for macroecology?. Ecography 35, 673–683.

Beck, J., Boller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Eco. Inform. 19, 10–15.

Bertorelle, G., Bruford, M., Chemini, C., Vernesi, C., Hauffe, H.C., 2004. New, flexible Bayesian approaches to revolutionize conservation genetics. Conserv. Biol. 18, 1–2.

Bierman, S.M., Butler, A., Marion, G., Kuhn, I., 2010. Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. Ecography 33, 451–460.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Springer.

Carl, G., Kuhn, I., 2007. Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. Ecol. Model. 207, 159–170.

Chen, G., Kery, M., Plattner, M., Ma, K., Gardner, B., 2012. Imperfect detection is the rule rather than the exception in plant distribution studies. J. Ecol. 101, 183–191.

Clark, J., 2005. Why environmental scientists are becoming Bayesians. Ecol. Lett. 8, 2–14.

Dietterich, T.G., 2000. Ensemble methods in machine learning. Lect. Notes Comput. Sci 1857, 1–15.

Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob. Ecol. Biogeogr. 16, 129–138.

Elith, J., Graham, C.H., 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Ellison, A., 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol. Appl. 6, 1036–1046.

Ellison, A., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520.
Engler, R., Randin, C.F., Vittoz, P., Czaka, T., Beniston, M., Zimmermann, N.E., Guisan, A., 2009. Predicting future distributions of mountain plants under climate change: does dispersal capacity matter?. Ecography 32, 34–45.
Farjon, A., 1998. World Bibliography and Checklist of Conifers, RBG Kew. Feilhauer, H., Thonfeld, F., Faude, U., He, K.S., Rocchini, D., Schmidtlein, S., 2013.
Assessing floristic composition with multispectral sensors - a comparison based on monotemporal and multiseasonal field spectra. Int. J. Appl. Earth Obs.Geoinf. 21, 218–229.
Ferretti, M., Chiarucci, A., 2003. Design concepts adopted in long-term forest monitoring
programs in Europe - problems for the future?. Sci. Total Environ. 310, 171–178.
Gastner, M.T., Newman, M.E.J., 2004. Diffusion-based method for producing density-equalizing maps. Proc. Natl. Acad. Sci. U. S. A. 101, 7499–7504.
Gaston, K.J., 2000. Global patterns in biodiversity. Nature 405, 220–227.
Gazol, A., Camarero, J.J., Gutierrez, E., Popa, I., Andreu-Hayles, L., Motta, R., Nola, P., Ribas, M., Sanguesa-Barreda, G., Urbinati, C., Carrer, M., 2015. Distinct effects of climate warming on populations of silver fir (*Abies alba*) across Europe. J. Biogeogr. 42, 1150–1162.
Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences.
Stat. Sci. 7, 457–511.
Ginzburg, L.R., Jensen, C.X.J., Yule, J.V., 2007. Aiming the "unreasonable effectiveness of mathematics" at ecological theory. Ecol. Model. 207, 356–362.
Goncalves, L., Fonte, C., Julio, E., Caetano, M., 2009. A method to incorporate uncertainty in the classification of remote sensing images. Int. J. Remote Sens. 30 (20), 5489–5503.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.
Heikkinen, R.K., Marmion, M., Luoto, M., 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability?. Ecography 35, 276–288.
Hijmans, R., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution
interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.
Hoeting, J.A., Volinsky, C.T., Madigan, D., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–417
Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records. Biol. J. Linn. Soc. 115, 522–531.
Keddy, P.A., 1992. Assembly and response rules: two goals for predictive community ecology. J. Veg. Sci. 3, 157–164.
Koppen, W., Geiger, R., 1930. Handbuch Der Klimatologie. Gebruder Borntraeger, Berlin, Germany.
Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Bathols, J., Macadam, I., Scott, J.K., 2012. Climond: global high resolution historical and future scenario climate surfaces for bioclimatic modelling. Methods Ecol. Evol. 3, 53–64.
Kruschke, J.K., 2015. Doing Bayesian Data Analysis, Second Elsevier, Amsterdam.
Link, W.A., Sauer, J.R., 2002. A hierarchical analysis of population change with application to Cerulean Warblers. Ecology 83, 9.
Lobell, D.B., Sibley, A., Ortiz-Monasterio, J.I., 2012. Extreme heat effects on wheat senescence in India. Nat. Clim. Chang. 2, 186–189.
Mack, R.N., Simberloff, D., Lonsdale, W.M., et al., 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecol. Appl. 10, 689–710.

Manceur, A.M., Kuhn, I., 2014. Inferring model-based probability of occurrence from preferentially sampled data with uncertain absences using expert knowledge. Methods Ecol. Evol. 5, 739–750.

Mathys, L., Guisan, A., Kellenberger, T.W., Zimmermann, N.E., 2009. Evaluating effects of spectral training data distribution on continuous field mapping performance. ISPRS J. Photogramm. Remote Sens. 64, 665–673.

Malanson, G.P., Walsh, S.J., 2013. A geographical approach to optimization of response to invasive species. In: Walsh, S.J., Mena, C. (Eds.), Science and Conservation in the Galapagos Islands: Frameworks and Perspectives. Springer, New York, USA, pp. 199–215.

Mara, T.A., Delay, F., Lehmann, F., Younes, A., 2016. A comparison of two Bayesian approaches for uncertainty quantification. Environ. Model Softw. 82, 21–30.

McCarthy, M.A., Masters, P., 2005. Profiting from prior information in Bayesian analyses of ecological data. J. Appl. Ecol. 42, 1012–1019. http://dx.doi.org/10.1111/j.1365-2664.2005.01101.x.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., New, M., 2004. A Comprehensive Set of Climate Scenarios for Europe and the Globe: The Observed Record (1900–2000) and 16 Scenarios (2000–2100). University of East Anglia, Norwich, UK, 30.

NRC (Committee on the Scientific Basis for Predicting the Invasive Potential of Non-indigenous Plants and Plant Pests in the United States), 2002. In: Predicting Invasions of Non-indigenous Plants and Plant Pests, National Academy Press, Washington, DC.

Pearman, P.B., Randin, C.F., Broennimann, O., Vittoz, P., van der Knaap, W.O., Engler, R., Le Lay, G., Zimmermann, N.E., Guisan, A., 2008. Prediction of plant species distributions across six millennia. Ecol. Lett. 11, 357–369.

Pompe, S., Hanspach, J., Badeck, F., Klotz, S., Thuiller, W., Kuhn, I., 2008. Climate and land use change impacts on plant distributions in Germany. Biol. Lett. 4, 564–567.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R News. 6, 7–11.

Core Team, R., 2016. R: a language and environment for statistical computing. R Foundation for Statistical computing, Vienna, Austria.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space?. J. Biogeogr. 33, 1689–1703.

Ricciardi, A., Palmer, M.E., Yan, N.D., 2011. Should biological invasions be managed as natural disasters?. Bioscience 61, 312–317.

Rocchini, D., 2007. Distance decay in spectral space in analysing ecosystem b-diversity. Int. J. Remote Sens. 28, 2635–2644.

Rocchini, D., Andreo, V., Forster, M., Garzon-Lopez, C.X., Gutierrez, A.P., Gillespie, T.W., Hauffe, H.C., He, K.S., Kleinschmit, B., Mairota, P., Marcantonio, M., Metz, M., Nagendra, H., Pareeth, S., Ponti, L., Ricotta, C., Rizzoli, A., Schaab, G., Zebisch, M., Zorer, R., Neteler, M., 2015. Potential of remote sensing to predict species invasions - a modeling perspective. Prog. Phys. Geogr. 39, 283–309.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog. Phys. Geogr. 35, 211–226.

Rocchini, D., Neteler, M., 2012. Let the four freedoms paradigm apply to ecology. Trends Ecol. Evol. 27, 310–311.

Rolland, C., Michalet, R., Desplanque, C., Petetin, A., Aime, S., 2009. Ecological requirements of *Abies alba* in the French Alps derived from dendro-ecological analysis. J. Veg. Sci. 10, 297–306.

Sax, D.F., 2001. Latitudinal gradients and geographic ranges of exotic species: implications for biogeography. J. Biogeogr. 28, 139–150.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2014. The Deviance Information Criterion: 12 years on (with discussion). J. R. Stat. Soc. Ser. B 76, 485–493.

Su, Y.S., Yajima, M., 2016. R2jags: a package for running jags from R. http://CRAN.R-project.org/package=R2jags.

Tattoni, C., Ianni, E., Geneletti, D., Zatelli, P., Ciolli, M., 2016. Landscape changes, traditional ecological knowledge and future scenarios in the alps: a holistic ecological approach. Sci. Total Environ. (in press).

Tinner, W., Colombaroli, D., Heiri, O., Henne, P.D., Steinacher, M., Untenecker, J., Vescovi, E., Allen, J.R.M., Carraro, G., Conedera, M., Joos, F., Lotter, A.F., Luterbacher, J., Samartin, S., Valsecchi, V., 2013. The past ecology of *Abies alba* provides new perspectives on future responses of silver fir forests to global warming. Ecol. Monogr. 83, 419–439.

Willis, S.G., Thomas, C., Hill, J.K., Collingham, Y., Telfer, M.G., Fox, R., Huntley, B., 2009. Dynamic distribution modelling: predicting the present from the past. Ecography 32, 5–12.

Winter, M., Schweiger, O., Klotz, S., Nentwig, W., Andriopoulos, P., Arianoutsou, M., Basnou, C., Delipetrou, P., Didziulis, V., Hejda, M., Hulme, P.E., Lambdon, J.P., Pysek, P., Royl, D.B., Kuhn, I., 2009. Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. Proc. Natl. Acad. Sci. U. S. A. 106, 21721–21725.

Wolf, H., 2003. EUFORGEN technical guidelines for genetic conservation and use for silver fir (*Abies alba*). In: International Plant Genetic Resources Institute, Rome, Italy p. 6.