

## RESEARCH ARTICLE

# Diagnosing the Dynamics of Observed and Simulated Ecosystem Gross Primary Productivity with Time Causal Information Theory Quantifiers

Sebastian Sippel<sup>1\*</sup>, Holger Lange<sup>2,3</sup>, Miguel D. Mahecha<sup>1,4,5</sup>, Michael Hauhs<sup>6</sup>, Paul Bodesheim<sup>1</sup>, Thomas Kaminski<sup>7</sup>, Fabian Gans<sup>1</sup>, Osvaldo A. Rosso<sup>3,8,9</sup>

**1** Max Planck Institute for Biogeochemistry, Jena, Germany, **2** Norwegian Institute of Bioeconomy Research, Ås, Norway, **3** Instituto de Física, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil, **4** German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany, **5** Michael Stifel Center Jena for Data-Driven and Simulation Science, Jena, Germany, **6** University of Bayreuth, Bayreuth, Germany, **7** The Inversion Lab, Hamburg, Germany, **8** Instituto Tecnológico de Buenos Aires (ITBA) and CONICET, Ciudad Autónoma de Buenos Aires, Argentina, **9** Complex Systems Group, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Las Condes, Santiago, Chile

\* [ssippel@bgc-jena.mpg.de](mailto:ssippel@bgc-jena.mpg.de)



## OPEN ACCESS

**Citation:** Sippel S, Lange H, Mahecha MD, Hauhs M, Bodesheim P, Kaminski T, et al. (2016) Diagnosing the Dynamics of Observed and Simulated Ecosystem Gross Primary Productivity with Time Causal Information Theory Quantifiers. PLoS ONE 11(10): e0164960. doi:10.1371/journal.pone.0164960

**Editor:** Ben Bond-Lamberty, Pacific Northwest National Laboratory, UNITED STATES

**Received:** April 28, 2016

**Accepted:** October 4, 2016

**Published:** October 20, 2016

**Copyright:** © 2016 Sippel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** We have gathered the data from various data providers specifically to conduct the complexity analysis as presented in this paper. Therefore, the original data are indeed third party data, hence it is not our property and we cannot make it public. However, we give very specific information below, and all data is available from the data providers as indicated below: 1. The JRC-TIP FAPAR dataset is made available by the European Commission and the Inversion Lab, as indicated in the manuscript. The JRC-TIP FAPAR

## Abstract

Data analysis and model-data comparisons in the environmental sciences require diagnostic measures that quantify time series dynamics and structure, and are robust to noise in observational data. This paper investigates the temporal dynamics of environmental time series using measures quantifying their information content and complexity. The measures are used to classify natural processes on one hand, and to compare models with observations on the other. The present analysis focuses on the global carbon cycle as an area of research in which model-data integration and comparisons are key to improving our understanding of natural phenomena. We investigate the dynamics of observed and simulated time series of Gross Primary Productivity (GPP), a key variable in terrestrial ecosystems that quantifies ecosystem carbon uptake. However, the dynamics, patterns and magnitudes of GPP time series, both observed and simulated, vary substantially on different temporal and spatial scales. We demonstrate here that information content and complexity, or Information Theory Quantifiers (ITQ) for short, serve as robust and efficient data-analytical and model benchmarking tools for evaluating the temporal structure and dynamical properties of simulated or observed time series at various spatial scales. At continental scale, we compare GPP time series simulated with two models and an observations-based product. This analysis reveals qualitative differences between model evaluation based on ITQ compared to traditional model performance metrics, indicating that good model performance in terms of absolute or relative error does not imply that the dynamics of the observations is captured well. Furthermore, we show, using an ensemble of site-scale measurements obtained from the FLUXNET archive in the Mediterranean, that model-data or model-model mismatches as indicated by ITQ can be attributed to and interpreted as differences in the temporal structure of the respective ecological time series. At global scale, our

dataset was introduced in the following publication: Pinty B., Clerici M., Andreadakis I., Kaminski T., Taberner M., Verstraete M. M., Gobron N., Plummer S. and Widlowski J.-L. (2011) Exploiting the MODIS Albedos with the Two-stream Inversion Package (JRC-TIP) Part II: Fractions of transmitted and absorbed fluxes in the Vegetation and Soil layers, *Journal of Geophysical Research-Atmospheres*, 116, D09106, doi:[10.1029/2010JD015373](https://doi.org/10.1029/2010JD015373). 2. For the site-based time series of GPP dataset, we use the well-known La-Thuile Synthesis Dataset (as described in the manuscript), available from <http://fluxnet.fluxdata.org/data/la-thuile-dataset/>. It's available based on a fair-use policy, hence everyone can request, use and download the dataset from the given URL. 3. The specific URL to access the GPP model runs data is <https://www.bgc-jena.mpg.de/geodb/projects/Data.php>. Everyone can download the data after a data usage agreement has been made with the owners of the data. 4. The specific URL to access CMIP5 data is <https://pcmdi.llnl.gov/search/cmip5/>, and the related publication is: Taylor, K.E., Stouffer, R.J. and Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), p.485. DOI: [http://dx.doi.org/10.1175/BAMS-D-11-00094.1](https://dx.doi.org/10.1175/BAMS-D-11-00094.1).

**Funding:** This work was supported by German National Academic Foundation, <http://www.studienstiftung.de> (no grant number), SS (Studienstiftung des Deutschen Volkes); Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, <http://www.conicet.gov.ar> (no grant number), OAR; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), <http://cnpq.br>, Grant number 310003/2016-4; HL; European Space Agency, STSE project CAB-LAB ("Coupled Atmosphere Biosphere virtual LABoratory"), <http://earthsystemdatacube.net>, MDM FG PB SS; European Commission, BACI project ("Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index"), <http://baci-h2020.eu/index.php/> (Grant number 640176), MDM SS.

**Competing Interests:** The authors have declared that no competing interests exist.

understanding of C fluxes relies on the use of consistently applied land models. Here, we use ITQ to evaluate *model structure*: The measures are largely insensitive to climatic scenarios, land use and atmospheric gas concentrations used to drive them, but clearly separate the structure of 13 different land models taken from the CMIP5 archive and an observations-based product. In conclusion, diagnostic measures of this kind provide data-analytical tools that distinguish different types of natural processes based solely on their dynamics, and are thus highly suitable for environmental science applications such as model *structural* diagnostics.

## Introduction

Understanding the feedback between terrestrial ecosystems and changing environmental conditions is a key prerequisite to model the impact of global change. Of particular relevance are changes in the global carbon (C from now on) cycle, i.e. in land-atmosphere fluxes and thus biosphere carbon stocks (e.g. [1]). Today, empirical and process-based models of varying structural and numerical complexity are used to assess and predict past, present and future ecosystem-atmosphere carbon exchange (e.g. [2]).

The state of current process-based understanding is in part implemented in terrestrial biosphere models. The process representations in these models differ widely, which is in part due to insufficient mechanistic understanding of the underlying processes or uncertain parameter estimates. This lack of consensus leads to major uncertainties in the predictions [3]. The starting point to reduce uncertainties in modelled fluxes or physiological state variables is to quantify their mismatch with observations [4].

Model *benchmarking* seeks to quantify this mismatch and to rank the models accordingly [5–9]. Obviously, model rankings depend on the metric(s) considered; it is unlikely that one single model will rank best with respect to multiple metrics that may emphasize different aspects.

Activities such as the International Land Model Benchmarking Project (ILAMB, <http://ilamb.org>) have formalized this idea by providing a range of benchmarks. In order to standardize the benchmarking approaches further, several platforms for automatic benchmarking have been developed [2, 7]. Most of the metrics used for benchmarking so far are straightforward, for instance focusing on long-term mean values per variable and pixel (or region), or the amplitude of some specific seasonal cycle (e.g. of some land-atmosphere flux). However, metrics of this kind can only provide a limited insight into the dynamics of the models under scrutiny. Consequently, latest benchmarking efforts intend to consider relevant observed and simulated patterns, such as process efficiencies or turnover-rates [10], i.e. these approaches attempt to derive a pattern-oriented strategy to model evaluation and model-data integration [11–13]. Hence, various initiatives to land surface model benchmarking are still working towards developing a widely acceptable set of benchmarks [8] that target various relevant aspects of model behaviour.

One difficulty in model benchmarking is that reference data are potentially affected by observational noise or biases. Ideally, models should exhibit similar dynamics as the observations on a range of time scales [14]. Disagreements of models and observations indicate either inaccurate parameter estimates, structural deficits, or other inadequacies of the models; or they originate in low quality of the reference observations, with e.g. noise contamination at high frequencies and sensor ageing processes at long time scales. Hence, in any model-data comparison

such as benchmarking, a distinction between signal and noise is necessary. In the interpretation of data analysis, signal extraction from the noise background is key; process-oriented models should be able to reproduce the signal but not the noise part of observations.

Here, we contribute to the discussion of model evaluation and benchmarking by investigating the potential of *Information Theory Quantifiers* (ITQ) as an additional set of benchmarking tools, with a special focus on the dynamics and structure of model simulated vs. observed time series. For example, questions such as '*How much information do the observations reveal about the dynamics of the underlying system or processes? Do observed time series resemble stochastic or deterministic processes? Are models reproducing the observed process classes?*' arise naturally from an Information Theory perspective and could potentially be tackled using ITQ. These measures draw a distinction between deterministic and stochastic processes, are complementary to the present set of tools and might thus provide a balanced investigation of model evaluation, e.g. *sensu* [6].

Among the set of variables established to describe C dynamics, the gross CO<sub>2</sub> uptake of terrestrial vegetation from the atmosphere ('Gross Primary Productivity', or GPP) is of particular importance for terrestrial ecosystems. GPP can be derived from in-situ measurements of net ecosystem exchange fluxes of CO<sub>2</sub>. GPP is predominantly controlled by radiation, but likewise sensitive to temperature, CO<sub>2</sub> concentrations, water availability and the phenological status of the vegetation, amongst other factors [15]. GPP is routinely obtained as the difference of Net Ecosystem Productivity (NEP) measurements and estimates for ecosystem respiration ( $R_{eco}$ ) at individual sites as part of a global monitoring network [16]. GPP is expected to show seasonal variation; abrupt changes or systematic trends in GPP can be induced by land use change, extreme events, and changing climate, among other factors. Numerous attempts to quantify GPP at larger spatial scales exist, but many of these rely on dynamical global vegetation and biogeochemical models. Based on assumed biological and physical processes, the latter either try to reconstruct historical GPP time series, based on observed precipitation, radiation, temperature and other drivers, or project the future dynamics of GPP using climate scenarios (Representative Concentration Pathways, RCPs [17]). Generally, GPP simulations respond globally to transient climate change, but locally and regionally also to abrupt land use changes, or extreme hydrometeorological anomalies [18, 19].

This study aims to introduce and interpret ITQ in the context of environmental time series analysis with particular emphasis on model evaluation and benchmarking of GPP dynamics at site, continental and global scale. The paper is organized as follows: the Section Quantifiers from Information Theory" provides an introduction to the field of ordinal pattern statistics and the associated Information Theory Quantifiers. "Data Sets Investigated" introduces the data: observations and model simulations of GPP at the site, continental and global scale; and a remotely sensed widely used proxy for vegetation activity ('Fraction of Absorbed Photosynthetically Active Radiation' (FAPAR)). The subsection "An Information Theory perspective on spatio-temporal environmental datasets" provides an intuitive introduction to ITQ in the context of environmental science using i) a FAPAR dataset at continental-scale and high spatial resolution; and ii) in-situ GPP measurements from flux tower sites to illustrate the effects of temporal resolution from the FLUXNET database, cf. <http://fluxnet.fluxdata.org> [16]. Then, we proceed by comparing the complexity and information content of GPP time series simulated by two process-based ecosystem models and an observations-based dataset at continental scale in a spatially explicit manner; these results are augmented with site-scale flux tower measurements. In "Global analysis of land surface models", ITQ calculated from global-scale model simulations of GPP obtained from a climate model intercomparison project (CMIP5) comprising both historical reconstructions and climate scenario runs (1861-2090; 'scenarios') are evaluated to diagnose model structure. Finally, we conclude on the suitability of Information

Theory Quantifiers for environmental science applications, such as diagnosing model structure or model benchmarking.

## Quantifiers from Information Theory

Given a time series or other observational data, a natural question arises: how much information are these data revealing about the dynamics of the underlying system or processes? The information content of data sets is typically evaluated via characterizing a value distribution or a probability density function (PDF)  $P$  describing the apportionment of some measurable or observable quantity [20]. These quantifiers represent metrics on the space of PDFs for data sets, allowing to compare different sets and to classify them according to the properties of underlying processes—broadly, stochastic vs. deterministic. Here, we refer to stochastic processes like correlated noise on one hand, and to deterministic chaotic maps which do not contain any noise on the other. Time series generated by these two categories of processes are difficult to discern by conventional analysis, but are clearly separated by ITQ [21].

In our case, we are interested in the temporal dynamics of GPP, and our data are *time series*  $x(t)$ . Thus, we are mostly interested in metrics which take the temporal order of observations explicitly into account; i.e. the approach is fundamentally a *causal* rather than a *statistical* one. In a purely statistical approach, correlations between successive values from the time series are ignored or simply destroyed via construction of the PDF; while a causal approach focuses on the PDFs of data sequences.

The quantifiers selected are based on ordinal pattern statistics (cf. e.g. [22]). For an application of alternative quantifiers based on Symbolic Dynamics to environmental data, we refer to [23]. The quantifiers used here belong to either of two broad categories: those which quantify the *information content* of data versus those related to their *complexity* on one hand; and metrics related to *global* properties of the appropriate PDFs versus ones which take *local* properties into account. Note that we are referring to the space of probability density functions here, not physical space. For the sake of clarity and simplicity, we introduce Information Theory quantifiers that are defined on discrete PDFs in this section, since we are only dealing with discrete data (time series). However, all the quantifiers also have definitions for the continuous case [24, 25].

## The Shannon entropy as a measure for information content

The permutation Shannon entropy is a measure for the information content of the time series [24] and quantifies uncertainty, disorder, state-space volume, and lack of information [26]. Let  $P = \{p_i; i = 1, \dots, N\}$  with  $\sum_{i=1}^N p_i = 1$ , be a probability distribution, with  $N$  possible states of the system under study. The Shannon information measure (entropy) reads

$$S[P] = - \sum_{i=1}^N p_i \ln [p_i] . \quad (1)$$

The Shannon entropy is minimal when all but one of the  $p_i$ 's vanish, and maximal when all  $p_i$ 's are equal, i.e. for the uniform distribution  $P_e = \{p_i = \frac{1}{N}, \forall i = 1, \dots, N\}$ . In this case,  $S[P_e] = S_{\max} = \ln N$ . However, these two situations are extreme cases, unlikely to occur in any natural phenomenon considered here. In the following, we focus on the ‘normalized’ Shannon entropy,  $0 \leq \mathcal{H} \leq 1$ , given as

$$\mathcal{H}[P] = S[P]/S_{\max} . \quad (2)$$

## Statistical complexity measures

Contrary to information content, there is no universally accepted definition of complexity. Here, we focus on describing the *complexity of time series* and do not refer to the complexity of the underlying *systems*. In fact, “simple” models might generate complex data, while “complicated” systems might produce output data of low complexity [27].

An intuitive notion of a quantitative complexity attributes low values both to perfectly ordered data (i.e. with vanishing Shannon entropy) as well as to uncorrelated random data (with maximal Shannon entropy), as both cases can be well described in a compact manner. For example, the statistical complexity of a simple oscillation or trend (ordered), but also of uncorrelated white noise (disordered) would be classified as low. Hence, linear transient trends and measurement noise of some geophysical variable would exhibit small complexity values, but the two processes would differ considerably in their Shannon entropy (Fig 1). Between the two cases of minimal and maximal entropy, data are more difficult to characterize and hence the complexity should be higher. We seek some functional  $\mathcal{C}[P]$  quantifying structures present in the data which deviate from these two cases. These structures relate to organization, correlational structure, memory, regularity, symmetry, patterns, and other properties [28].

One suitable way to guarantee the desired properties for a complexity measure is to build the product of a measure of information and a measure of disequilibrium, i.e. some kind of distance from the uniform ('equilibrium') distribution of the accessible states of a system. In this respect, in [29] an effective *Statistical Complexity Measure* (SCM)  $\mathcal{C}$  was introduced, that is able to detect and discern basic dynamical properties of datasets.

Based on the seminal notion advanced by Lopez-Ruiz *et al.* [30], this statistical complexity measure [29, 31] is defined through the functional product form

$$\mathcal{C}[P] = \mathcal{Q}_j[P, P_e] \cdot \mathcal{H}[P] \quad (3)$$

of the normalized Shannon entropy  $\mathcal{H}$ , see Eq (2), and the disequilibrium  $\mathcal{Q}_j[P, P_e]$  defined in terms of the Jensen-Shannon divergence  $\mathcal{J}[P, P_e]$ . The latter is given by

$$\mathcal{Q}_j[P, P_e] = Q_0 \mathcal{J}[P, P_e] = Q_0 \{S[(P + P_e)/2] - S[P]/2 - S[P_e]/2\}, \quad (4)$$

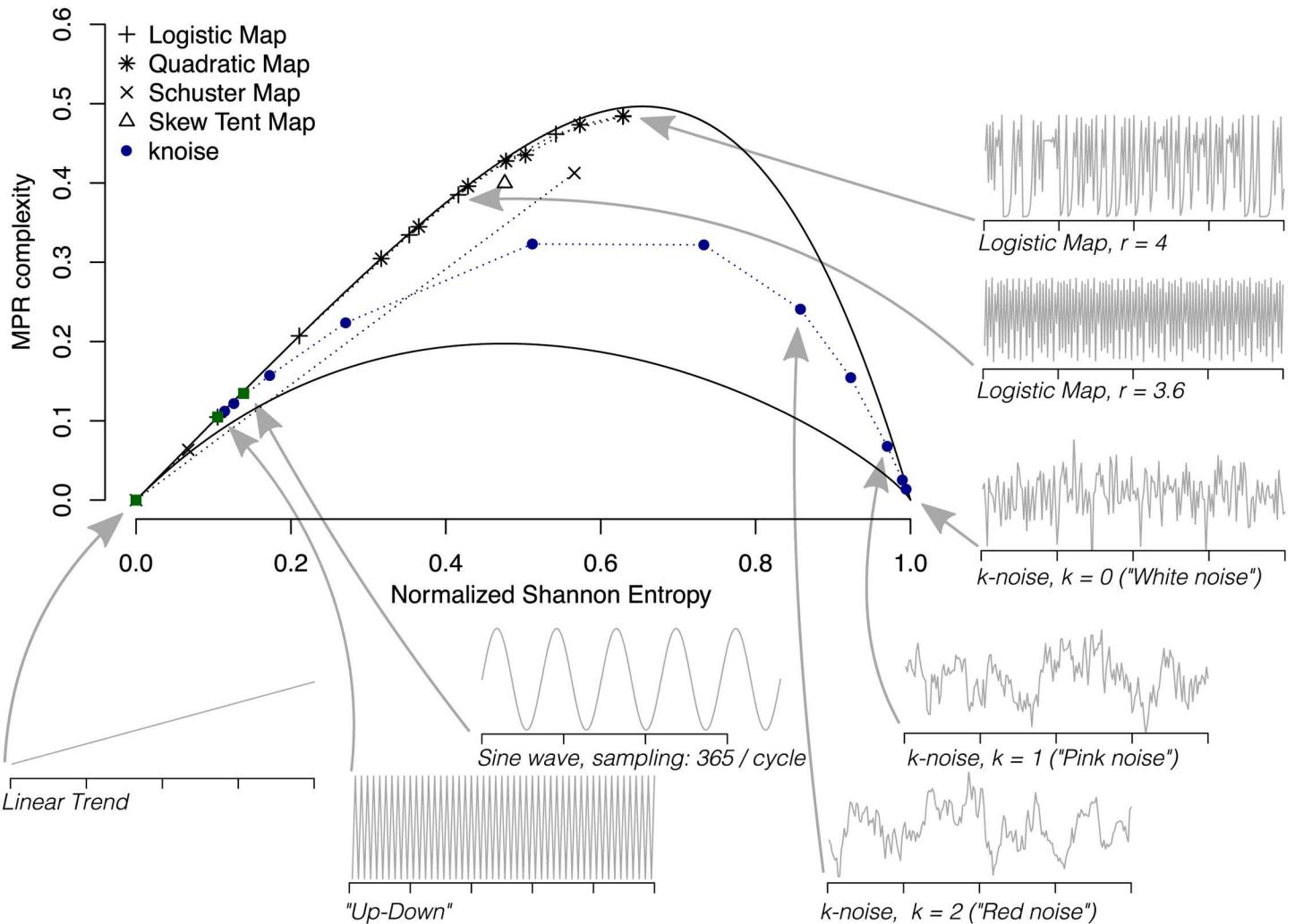
where  $Q_0$  is a normalization constant chosen such that  $0 \leq \mathcal{Q}_j \leq 1$ :

$$Q_0 = -2 \left\{ \frac{N+1}{N} \ln(N+1) - \ln(2N) + \ln N \right\}^{-1}. \quad (5)$$

The inverse of this value is obtained for the non-normalized Jensen-Shannon divergence when one of the components of  $P$ , say  $p_m$ , is equal to one and the remaining  $p_j$ 's are zero.

The Jensen-Shannon divergence quantifies the difference between probability distributions and is especially useful to compare the symbolic composition of different sequences [32]. Note that the above introduced Information Theory quantifier depends on two different probability distributions: one associated with the system under analysis,  $P$ , and the other the uniform distribution  $P_e$ . For the latter, other reference distributions can be chosen to test whether an observed distribution is close to a target distribution. It has been shown that there are *limit curves* for complexity: for a given value of  $\mathcal{H}$  and any data set, the possible  $\mathcal{C}$  values vary between a minimum  $\mathcal{C}_{min}(\mathcal{H})$  and a maximum  $\mathcal{C}_{max}(\mathcal{H})$ , restricting the possible values of the complexity measure [33].

An alternative measure which is local in distribution space is the Fisher Information Measure (FIM). For calculating FIM, we use probability amplitudes as a starting point [25] and



**Fig 1. Illustration of causal Information Theory quantifiers as data-analytical tools.** The causality entropy–complexity plane exhibits upper and lower limit curves, and the distances to these can be used to classify relevant processes. For all calculations,  $D = 6$ ,  $\tau = 1$ , and  $n = 10^4$  were chosen.

doi:10.1371/journal.pone.0164960.g001

obtain a discrete normalized FIM,  $0 \leq \mathcal{F} \leq 1$ , convenient for our purposes [34]:

$$\mathcal{F}[P] = F_0 \sum_{i=1}^{N-1} [\sqrt{p_{i+1}} - \sqrt{p_i}]^2 . \quad (6)$$

Here the normalization constant  $F_0$  reads

$$F_0 = \begin{cases} 1, & \text{if } p_{i^*} = 1 \text{ for } i^* = 1 \text{ or } i^* = N \text{ and } p_i = 0, \forall i \neq i^*, \\ 1/2, & \text{otherwise.} \end{cases} \quad (7)$$

The general behavior of this discrete version of the FIM (Eq (6)) is opposite to that of the Shannon entropy, except for periodic motions. The local sensitivity of FIM for discrete PDFs is reflected in the fact that it depends on the specific ‘ $i$ -ordering’ of the discrete values  $p_i$ .

[34, 35]. The summands in Eq (6) can be regarded as a ‘distance’ between two contiguous probabilities. Thus, a different ordering of the patterns would lead to a different FIM-value, demonstrating its local nature. In the present work, we follow the so-called Keller sorting scheme [36] for the generation of the Bandt and Pompe PDF discussed in the next section.

### The Bandt and Pompe approach to generate a causal PDF

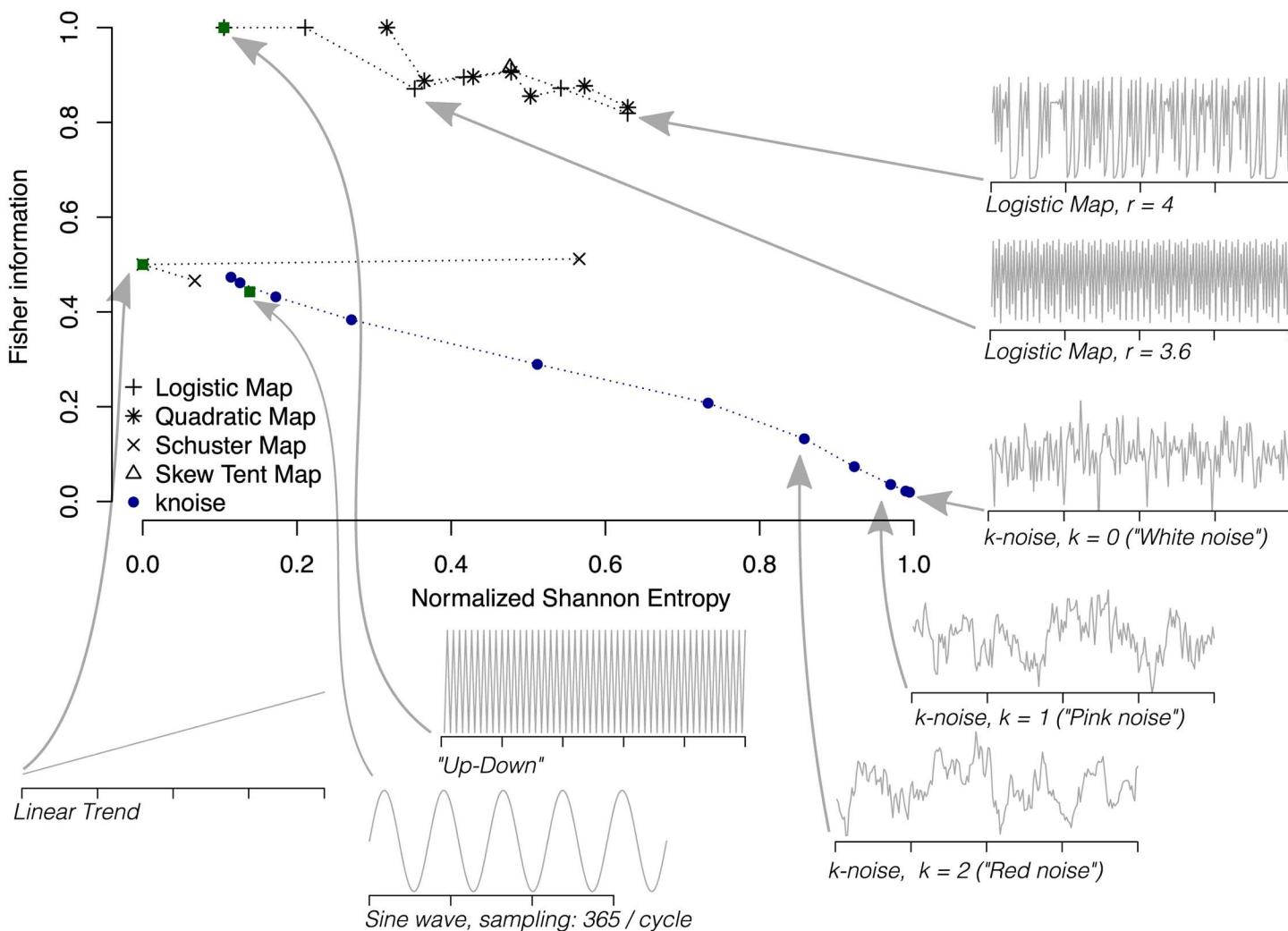
The quantifiers from Information Theory rely on a probability distribution associated to the time series. The determination of the most adequate PDF is a fundamental problem because the PDF  $P$  and its sample space  $\Omega$  are inextricably linked. The usual histogram technique is inadequate since the data are treated purely stochastic and the temporal information is completely lost. Bandt and Pompe (BP) [22] introduced a simple and robust symbolic methodology that takes into account time ordering of the time series by comparing neighboring values in a time series. The symbolic data are created by first ranking the values of the series within windows of a fixed length, and then reordering these embedded data in ascending order. This is tantamount to a phase space reconstruction with embedding dimension (pattern length)  $D$  and time lag  $\tau$  (see Text 1 in [S1 File](#) for a more detailed description). In this way, it is possible to quantify the diversity of the ordering symbols (patterns) derived from a scalar time series. Note that the appropriate symbol sequence arises naturally from the time series, and no model-based assumptions are needed. As such, it allows to uncover important details concerning the ordinal structure of the time series [21] and can also yield information about temporal correlation [37].

This type of analysis of a time series entails losing details of the original series’ amplitude information. However, the symbolic representation of time series by recourse to a comparison of consecutive ( $\tau = 1$ ) or nonconsecutive ( $\tau > 1$ ) values allows for an accurate empirical reconstruction of the underlying phase-space, even in the presence of weak (observational and dynamic) noise [22]. Furthermore, the ordinal patterns associated with the PDF are invariant with respect to nonlinear monotonous transformations; nonlinear drifts or scaling artificially introduced by a measurement device will not modify the estimation of quantifiers, a nice property if one deals with experimental data (see, e.g., [38]). The only condition for the applicability of the BP method is a very weak stationarity assumption: for  $k \leq D$ , the probability for  $x_t < x_{t+k}$  should not depend on  $t$ . For a review of the BP methodology and its applications to physics, biomedical and econophysics signals, see [39].

Regarding the selection of the parameters, Bandt and Pompe suggested working with  $4 \leq D \leq 6$  for typical time series lengths, and specifically considered a time lag  $\tau = 1$  in their cornerstone paper [22]. For the artificially generated time series shown below (Figs 1 and 2), we chose  $D = 6$  and follow the Lehmer-permutation scheme [35] to calculate the Fisher Information. For the measured and modelled time series analyzed here, the embedding dimension is chosen as  $D = 4$  throughout due to time series length requirements ([S1 File](#)), in particular at coarse temporal resolution, and to achieve comparability across the different analyses. In all cases, the delay parameter has been set to  $\tau = 1$ .

### Incorporating amplitude information: Weighted ordinal pattern distribution

Recently, the permutation entropy was extended to incorporate also amplitude information [40]. Hence, a potential disadvantage of ordinal pattern statistics, namely the loss of amplitude information, can be addressed by introducing weights in order to obtain a ‘weighted permutation entropy (WPE)’. In the context of environmental time series, which typically exhibit a pronounced seasonal cycle and thus seasonally varying signal to noise ratios, this idea might be



**Fig 2. Illustration of the causality Shannon-Fisher plane plane ( $\mathcal{H} \times \mathcal{F}$ ) as data-analytical tool.**  $\mathcal{H} \times \mathcal{F}$  is used to classify relevant processes, such as constant/periodic signals, white and coloured noise (noise with a power spectrum proportional to  $1/f^k$ ), and deterministic signals (e.g. chaotic maps), similar to Fig 1. For all calculations,  $D = 6$ ,  $\tau = 1$ , and  $n = 10^4$  were chosen.

doi:10.1371/journal.pone.0164960.g002

particularly useful to address (noisy) low-variance patterns (e.g. during dormancy periods in winter).

Weighting the probabilities of individual patterns according to their variance alleviates potential issues regarding to ‘high noise, low signal’ patterns, because low-variance patterns that are strongly affected by noise are down-weighted in the resulting ‘weighted ordinal pattern distributions’. For example, [40] show that a weighted entropy measure is sensitive to sudden changes in the variance of the time series. Here, we extend the idea of WPE following [40] to derive a weighted permutation entropy ( $\mathcal{H}_w$ ), weighted statistical complexity ( $\mathcal{C}_w$ ), and weighted Fisher Information ( $\mathcal{F}_w$ ).

Non-normalized weights are computed for each temporal window for the time series  $X$ , such that

$$w_j = \frac{1}{D} \sum_{k=1}^D (x_{j+k-1} - \bar{X}_j^D)^2. \quad (8)$$

Here, the embedding dimension is denoted by  $D$  and  $\bar{X}_j^D$  denotes the arithmetic mean of the time series in the current window with index  $j$ . Thus, the weight of each window of length  $D$  is given by its variance in the equation above. The weights  $w_j$  are then used to modify the relative frequencies of each ordinal pattern  $\pi_j(X_j^D)$ , with  $N - D + 1$  windows, given by

$X_j^D = (x_j, x_{j+1}, \dots, x_{j+D-1})$ , for a time series of length  $N$  [40]:

$$p_w(\pi_i) = \frac{\sum_{j=1}^{N-D+1} \delta_{\pi_i, \pi_j} w_j}{\sum_{k=1}^{N-D+1} w_k} \quad (9)$$

The denominator of this equation provides the normalization, and the Kronecker delta  $\delta_{\pi_i, \pi_j}$  in the numerator serves to indicate which pattern occurs in each window  $j$ :

$$\delta_{\pi_i, \pi_j} = \begin{cases} 1, & \text{if } \pi_i = \pi_j, \\ 0, & \text{if } \pi_i \neq \pi_j. \end{cases}$$

After having calculated the appropriate  $\pi_i$  for each  $i = 1, \dots, D!$ , the same Eqs (2), (3) and (6) are applied to obtain the weighted versions  $\mathcal{H}_w$ ,  $\mathcal{C}_w$  and  $\mathcal{F}_w$  of the ITQ.

The weighting of the ITQ can be considered as noise filter, provided that noise is characterized by relatively low variance, and thus enhances the signal contained in the time series. In any case, rare patterns are suppressed in favour of more frequent ones.

Obviously, the window-based variance is but one out of many weighting recipes, others are easily conceivable. There is also a connection to the celebrated Rényi entropy ([41])

$$H_q(X) = \frac{1}{1-q} \log_2 \left( \sum_{i=1}^{D!} p_i^q \right)$$

where the Rényi exponent  $q$  suppresses low-frequency patterns for  $q > 1$ ; for  $q = 1$ , the Shannon entropy is obtained. The resulting ‘Rényi ordinal pattern distribution’ could be considered as a special case of the weighted pattern distribution, characterized by a single exponent; this has not been investigated so far, however.

Summarizing, ITQ computed based on an appropriately weighted form of the ordinal pattern distribution are suitable to analyse data sets with considerable amplitude information (e.g. seasonal variation) from an information-theoretic viewpoint. Since this is an issue for the time series investigated here, we will mostly perform our ITQ analysis using the weighted versions described here.

## Causal information planes

A particularly useful visualization of the quantifiers from Information Theory is their juxtaposition in two-dimensional graphs (‘causal information planes’), e.g. a) The *causality entropy-complexity plane*,  $\mathcal{H} \times \mathcal{C}$ , or its variance-weighted variant  $\mathcal{H}_w \times \mathcal{C}_w$ , is based only on global characteristics of the associated time series PDF (both quantities are defined in terms of Shannon entropies); while b) the *causality Shannon-Fisher plane*,  $\mathcal{H} \times \mathcal{F}$ , or its variance-weighted variant ( $\mathcal{H}_w \times \mathcal{F}_w$ ), is based on global and local characteristics of the PDF. In the case of  $\mathcal{H} \times$

$\mathcal{C}$  (or its weighted version) the value range is  $[0, 1] \times [\mathcal{C}_{min}(\mathcal{H}), \mathcal{C}_{max}(\mathcal{H})]$ , while in the causality plane  $\mathcal{H} \times \mathcal{F}$  (or with weights) the range is presumably  $[0, 1] \times [0, 1]$ ; no limit curves have been shown to exist so far.

These diagnostic planes are particularly efficient to distinguish between the deterministic chaotic and stochastic nature of a time series [21, 34] since the permutation quantifiers have distinct behaviours for different types of processes, see Figs 1 and 2, respectively.

Chaotic maps have intermediate entropy  $\mathcal{H}$  and Fisher  $\mathcal{F}$  values, while their complexity  $\mathcal{C}$  reaches larger values, very close to the upper complexity limit [21, 34]. For regular processes, entropy and complexity have small values, close to zero, while the Fisher information is close to one. Uncorrelated stochastic processes have  $\mathcal{H}$  near one and  $\mathcal{C}, \mathcal{F}$  near zero, respectively. It has also been found that correlated stochastic noise processes with a power spectrum proportional to  $1/f^k$ , where  $1 \leq k \leq 3$ , are characterized by intermediate permutation entropy and intermediate statistical complexity values [21], as well as intermediate to low Fisher information [34]. In both causal information planes ( $\mathcal{H} \times \mathcal{C}$ , see Fig 1 and  $\mathcal{H} \times \mathcal{F}$ , see Fig 2), stochastic data are clearly localized at different planar positions than deterministic chaotic ones. These two causal information planes have been profitably used to visualize and characterize different dynamical regimes when the system parameters vary [34, 35, 42–51]; to study temporal dynamic evolution [52–54]; to identify periodicities in natural time series [55]; to identify deterministic dynamics contaminated with noise [56, 57]; to estimate intrinsic time scales and delayed systems [58–60]; for the characterization of pseudo-random number generators [61, 62]; to quantify the complexity of two-dimensional patterns [63]; and for ecological [45], biomedical and econophysics applications (see [39] and references therein).

## Quantification of distance between models and observations

The overall objective of this paper is to provide a methodology for analyzing time series and comparing models with data based on Information Theory Quantifiers. However, the quantification of visually observed differences in causal information planes is not completely straightforward. Since the information planes are not Euclidean spaces, the Euclidean distance between pairs of points is not suitable. We need a distance metric that takes the nonlinear structure of the manifold into account. As we are working in the space of ordinal pattern distributions, a distance measure between PDF's is appropriate. We quantify the discrepancy between observations and the model outputs by calculating the Jensen-Shannon divergence [64] between them:

$$\mathcal{J}[P_{obs}, P_{mod}] = S[(P_{obs} + P_{mod})/2] - S[P_{obs}]/2 - S[P_{mod}]/2, \quad (10)$$

where  $P_{obs}$  and  $P_{mod}$  are the ordinal pattern distributions of the observations (or ‘observation-based products’, see next section) and the model outputs, respectively.

We note that a model evaluation based on Information Theory Quantifiers directly on the model residuals ( $X_{t,Mod} - X_{t,Obs}$ ) is not meaningful in itself: the ordinal pattern distribution of the model residuals for a poor, but noisy model with high variance would exhibit low complexity and high entropy values (‘white noise’)—despite an inadequate model structure in this simple example. That model residuals exhibit white noise behaviour is not an indicator for a good model performance per se.

## Data analysis and software

The open source R-package “statcomp” (Statistical Complexity Analysis) has been written to facilitate an easy access to ITQ’s and is available on CRAN (<http://CRAN.R-project.org/package=statcomp>) and R-Forge (<http://r-forge.r-project.org/projects/statcomp/>). A short

installation guide, links to a detailed manual and a code tutorial to reproduce Figs 1 and 2 are given in [S1 Code](#).

## Data Sets Investigated

### A remotely sensed vegetation activity proxy: FAPAR

The ‘Fraction of Absorbed Photosynthetically Active Radiation’ (FAPAR) is a variable that describes the ratio of absorbed to total incoming solar radiation in the photosynthetically active wavelength range at the land surface [65]. As the solar radiation is the driver for photosynthetic activity, FAPAR is used to diagnose vegetation productivity (e.g. [66]) and a so-called essential climate variable for global monitoring of the land surface and the terrestrial biosphere [67]. Here we use the JRC-TIP FAPAR data set [65] to provide an intuitive introduction and interpretation of continental-scale gradients and structure as obtained from analyses using ITQ (cf. above). This FAPAR product is derived (together with a set of further land surface variables such as the effective Leaf Area Index and the albedo of the soil background) by the Joint Research Centre Two-Stream Inversion Package (JRC-TIP), which is based on a one-dimensional (Two-Stream) representation of the canopy-soil system. The products were retrieved in an inversion procedure that combines the information in the MODIS broadband albedo product and prior information on the model’s state variables. For the purpose of our analysis it is worth noting that the prior information is constant in space and time with the exception of snow events, i.e. the spatio-temporal structure in the FAPAR data set is solely imposed by observations from space. The use of the two-stream model ensures physical consistency of all derived variables, as long as the products are used in the native resolution of the albedo input product [68], which in our case is 1km. In the temporal domain, as for the output of the terrestrial model (see below) we use monthly averages.

### GPP time series at site-scale: Flux tower measurements

Fluxes across the atmosphere-biosphere boundary (directly above the canopy) are measured routinely in a global network of flux tower sites (FLUXNET) using the Eddy-Covariance (EC) method [16]. Net ecosystem fluxes of carbon were partitioned into GPP and ecosystem respiration by using nighttime data that consist only of respiratory fluxes [69].

In this study, the dynamics of GPP time series from an ensemble of European FLUXNET sites with each more than five years of continuous measurements are investigated at monthly resolution. In addition, three sites are selected to illustrate the effect of temporal resolution (i.e. aggregation from half-hourly to monthly resolution) on complexity and entropy of the respective time series. These three sites represent different major European climate regions, i.e. Mediterranean evergreen broadleaf (Puechabon, France), temperate and boreal evergreen needleleaf (Tharandt, Germany and Hyttiälä, Finland, respectively) forest sites. A table containing detailed information about all investigated sites is available (Table 1 in [S1 File](#)).

### Continental-scale estimates of GPP: comparison of model runs and an observations-based product

**Model Tree Ensembles (MTE).** An empirical upscaling of GPP fluxes from the site scale to global scales was conducted by [70]. These authors used Fluxnet site measurements with local meteorological observations and remotely sensed vegetation indices to train an ensemble of model trees. In a subsequent step, the model trees were used to predict spatially explicit, fully data-driven GPP fluxes (at 0.5° spatial and monthly temporal resolution, 1982–2011) using global, gridded meteorological data and remote sensing observations. These interpolated

and upscaled GPP fluxes comprise the ‘observations’, as opposed to the model runs described in the following paragraph which are ‘simulations’. Two variants of the MTE dataset are used in this study, which differ in the method applied for partitioning the tower-based net flux measurements (Net Primary Productivity, or NPP) into GPP and ecosystem respiration,  $R_{eco}$ . These include the widely used extrapolation of night-time respiration into daytime [69] (‘MTE-MR’), and a separation method that uses a light response curve [71] (‘MTE-GL’).

**LPJmL.** The Lund-Potsdam-Jena managed Land dynamic global vegetation model (LPJmL) simulates dynamic vegetation development and structure of 10 natural plant functional types, two of which are herbaceous and eight are woody [72]. The human land use scheme consists of 13 crop functional types, including both grazing lands and agricultural crops [73]. Photosynthetic carbon assimilation in LPJmL follows the process-oriented coupled photosynthesis and water balance scheme of the BIOME3 model [74]. Photosynthesis is simulated at the canopy scale depending on seasonally varying nitrogen content and carboxylation capacity, which are functions of absorbed photosynthetically active radiation, temperature, atmospheric CO<sub>2</sub>, day length, and canopy conductance (*ibid.*).

**JSBACH.** The Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH) is a modular land surface scheme based on the ‘Biosphere Energy-Transfer and Hydrology Model’ (BETHY, [75]). It comprises 13 natural plant functional types that are distinguished by plant (eco-) physiological properties, and the relevant spatial characteristics are prescribed as maps [76]. The model operates internally with 30 minutes temporal resolution. Model grid cells are covered by at most four different plant functional types [77]. Additionally, five vegetation phenotypes are specified, namely managed (non-forest) lands, grassland, raingreen forest or shrubland, evergreen and summergreen. Photosynthesis is simulated using distinct physiologically based submodules for C3 and C4 plants [78], and includes an explicit representation of the interdependence between the carbon assimilation rate and stomatal conductance [76]. Both variables are a function of temperature, soil moisture, water vapour, CO<sub>2</sub> concentration and the absorption of visible solar radiation, the latter of which is resolved in three canopy layers.

**Modelling Protocol.** Both vegetation model simulations (LPJmL and JSBACH) were conducted at 0.25° spatial resolution and at the daily time scale for Europe [79]. Subsequently, the output was linearly aggregated to the monthly time scale and 0.5° spatial resolution for comparability with the MTE dataset. It is important to note that aggregation (taking mean values) and decimation (thinning to the desired resolution) of the time series are operations with very different consequences for the complexity analysis: whereas aggregation increases correlation, decreases Shannon entropy and in general induces a shift in the entropy-complexity plane to the left and upwards, decimation diminishes correlation, increases Shannon entropy and leads to a shift in the opposite direction.

## Global-scale simulated GPP dynamics: CMIP5

In order to evaluate the influence of different model structures vs. different climatic scenarios and trends on GPP dynamics, the behaviour of global GPP dynamics as simulated in the Fifth Climate Model Intercomparison Project (CMIP5) multimodel ensemble [3] is analyzed. The two representative concentration pathways (RCPs) 4.5 and 8.5 [17] are used in 13 different models and model variants (i.e. ensemble members) in monthly resolution from 1860–2099 (see Table 2 in [S1 File](#) for a detailed overview).

ITQ are calculated for each land grid cell in each of 59 model simulations (comprising combinations of model variants, different emission scenarios and ensemble members) and for each of the twenty-one 30-year periods within 1860–2099, yielding in total 1239 simulated 30-year

periods. In addition, the two MTE datasets for the period 1982–2011 are included in the comparison. Subsequently, for each 30-year period, all grid cells are visualized in the causal information planes (see [S6](#) and [S7](#) Figs for an example), where each model-scenario combination generates a point cloud in the causal information plane.

To compare these point clouds in a rigorous manner, i.e. to take the different local point densities into account, the planes are rasterized using a regular grid of 25x25 pixels. Subsequently, the number of land grid cells that fall into each of the respective (up to 625) ‘causal information classes’ are counted and then normalized (to account for different spatial resolution), yielding one count ‘vector’ for each model-scenario combination. Note that due to the existence of the limit curves in  $\mathcal{H} \times \mathcal{C}$ , some of the pixels are empty by necessity. All pixels with no points across all 1239 vectors are omitted, which yields an effective reduction in the vector’s dimensionality. Then, a principal component analysis (PCA) is conducted on these vectors that is used to illustrate in a simple manner the separation of CMIP5 models and scenarios along the first and second principal component. For  $\mathcal{H} \times \mathcal{F}$ , we proceed in a similar manner.

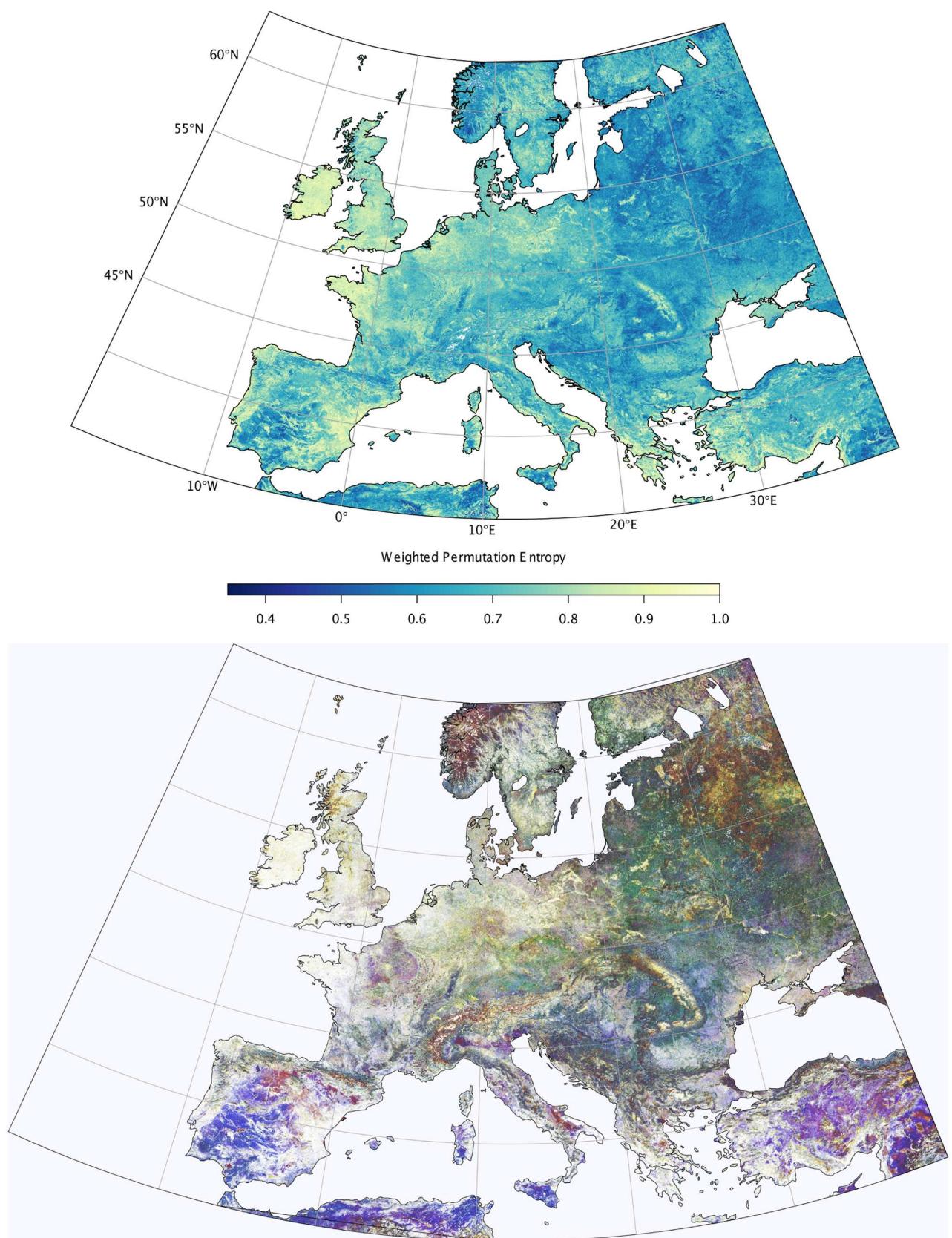
## Results and Discussion

In this section, we first illustrate and interpret continental-scale gradients in ITQ as obtained from a high-resolution vegetation productivity proxy (FAPAR) and proceed by investigating the effect of temporal resolution on ITQ at site scale. Subsequently, we compare two models and an observations-based product, including an interpretation that might offer hints for model improvements. Finally, a global-scale application of complexity measures based on CMIP5 model simulations is presented.

### An Information Theory perspective on spatio-temporal environmental datasets

**Continental-scale gradients in Information Theory Quantifiers.** ITQ from time series of vegetation activity proxies such as FAPAR show a large variety of spatial patterns from local to continental scales due to differences in vegetation type, land use, climate, and other factors. For example, the Weighted Permutation Entropy ( $\mathcal{H}_w$ ) for monthly FAPAR time series over Europe shows large spatial gradients that are mainly related to the regularity of the seasonal cycle in vegetation activity ([Fig 3](#), upper panel). Western parts of the continent and many coastal regions show rather high entropy values, e.g. over the British Isles, North-Western France, and the Netherlands, indicating a relatively stochastic behaviour of the respective time series. On the other hand, (North-)Eastern parts of the continent, but also seasonally dry regions like the Iberian peninsula, Southern Italy and North Africa exhibit low entropy values and thus monthly time series appear to be predictable. However, in [Fig 3](#) it becomes already obvious that a quantification of entropy alone is not able to distinguish between climatologically and ecologically well-separated regions such as Southern Spain and Eastern Europe (both displaying low entropy).

Therefore, additional ITQ are needed and useful to distinguish further features in the time series, e.g. related to the separation of stochastic vs. deterministic processes [21]. Therefore, we exemplify the potential of combining several ITQ in the context of environmental time series in an RGB color image ([Fig 3](#), bottom panel), where three ITQ (Weighted permutation entropy, statistical complexity, and Fisher Information) are encoded in the red, green, and blue channel, respectively ([Fig 3](#), bottom panel). This picture ([Fig 3](#), bottom panel, see [S3 Fig](#) for a high-resolution version, maps of individual ITQ are available in [S1](#) and [S2](#) Figs) illustrates that ITQ indeed capture a large amount of structure in environmental time series: Continental-scale gradients are clearly visible, but now structurally distinct regions (e.g. Southern Spain vs.

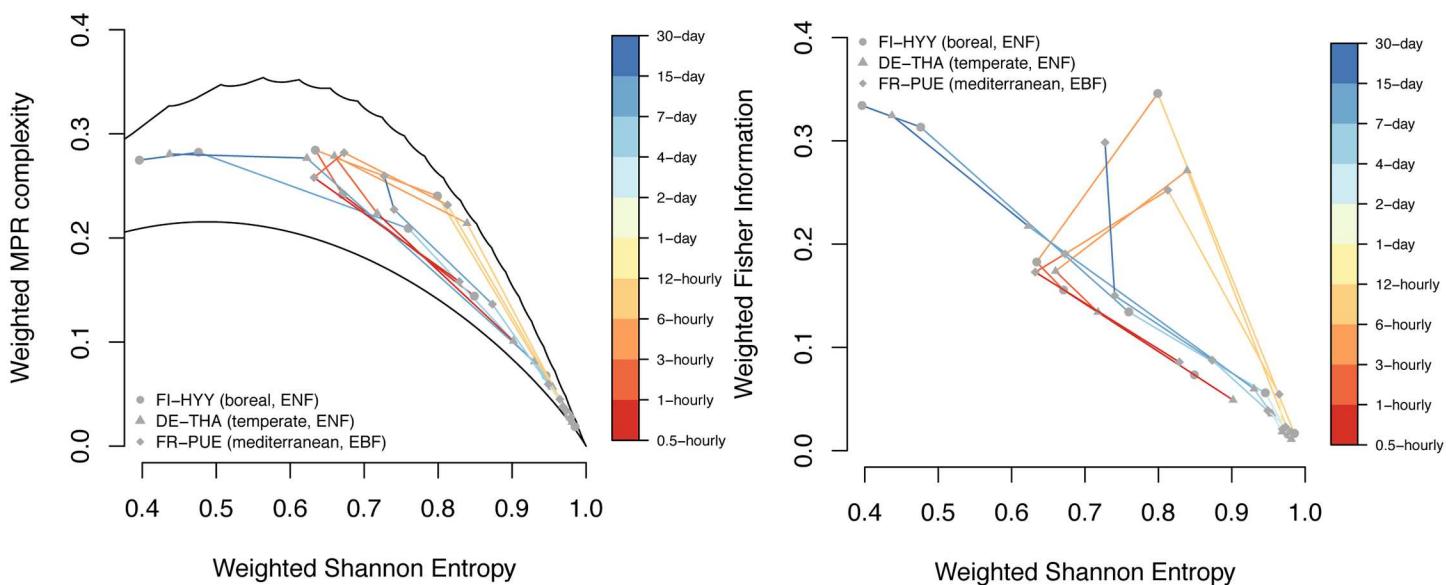


**Fig 3. ITQ derived from monthly time series from the JRC-TIP FAPAR dataset.** (top) Weighted Permutation Entropy, (bottom) RGB plot of weighted permutation entropy ( $\mathcal{H}_w$ , red channel), weighted complexity ( $\mathcal{C}_w$ , green channel), and weighted Fisher Information ( $\mathcal{F}_w$ , blue channel). An intense color indicates a high value of the respective ITQ.

doi:10.1371/journal.pone.0164960.g003

Eastern Europe, see above) become well separated. Moreover, regional structures related to ecosystem type, vegetation seasonality, and topography (e.g. mountain ranges) appear. It should be noted that the patterns and gradients obtained result from dynamical behaviour, i.e. quantify properties of vegetation in time, whereas the relation to other aspects like water availability, soil type, climate zone, or topography is non trivial and they do not enter the calculation at all. This is a demonstration of the strength of the link between dynamical properties and environmental conditions for plant communities (the basis of plant biogeography). These results thus highlight the potential of ITQ as indicators of ecological structure, and for model-data comparison exercises (see next subsection).

**The influence of seasonal and diurnal oscillations on statistical complexity.** Temporal resolution strongly affects the position of time series with oscillatory components in causal information planes. This property is illustrated for GPP time series from the boreal, temperate and Mediterranean climate regions (Fig 4). At the highest temporal resolution (half-hourly), GPP time series have a high  $\mathcal{H}_w$  and correspondingly small  $\mathcal{C}_w$  value, but this is distinctively different from the purely random case. At first, aggregating increases the complexity, until the strong daily cycle of the series is ‘detected’ at a resolution of 6 hours (due to  $D = 4$ , when one window covers just one day). The simple oscillatory behaviour reduces the complexity. The random extreme ( $\mathcal{H}_w$  close to one,  $\mathcal{C}_w$  close to zero) in the entropy-complexity plane is almost reached at daily resolution. Further aggregation increases  $\mathcal{C}_w$  and reduces  $\mathcal{H}_w$ ; at monthly resolution, we find an  $\mathcal{H}_w$  value around 0.65 only. We expect the classification in  $\mathcal{H}_w \times \mathcal{C}_w$  to be



**Fig 4. The effect of seasonal and diurnal oscillations on ITQ.** GPP dynamics at three flux tower measurement sites from half-hourly to monthly resolution as quantified by a)  $\mathcal{H}_w \times \mathcal{C}_w$ , and b)  $\mathcal{H}_w \times \mathcal{F}_w$ .

doi:10.1371/journal.pone.0164960.g004

most sensitive in this region in the center of the plane, since the difference between upper and lower limit curves is largest here. Thus, the monthly scale seems to be especially suitable for the subsequent regional assessment. It is not possible to further aggregate the series due to the length requirements—but at 3 months resolution, the annual cycle would be found, and another ‘loop’ in the diagram would be expected.

Similar to  $\mathcal{H}_w \times \mathcal{C}_w$ , the position of a time series in the Shannon-Fisher plane ( $\mathcal{H}_w \times \mathcal{F}_w$ ) strongly depends on the temporal resolution (Fig 4). At the highest temporal resolution, the observed, EC-based GPP time series has a high  $\mathcal{H}_w \simeq 0.95$  and a very small Fisher information. Aggregation decreases entropy and increases  $\mathcal{F}_w$  (Fig 4, upper part), until again the daily cycle of the series comes into reach at a resolution of 6 hours. Further aggregation increases entropy and decreases Fisher information, towards the “random corner” ( $\mathcal{H}_w$  close to one,  $\mathcal{F}_w$  close to zero) which is almost reached at daily and two-daily resolution. Aggregating further leads to a more and more steep increase in  $\mathcal{F}_w$ , where the slope of this part of the curve resembles that of the  $k$ -noise shown in Fig 2.

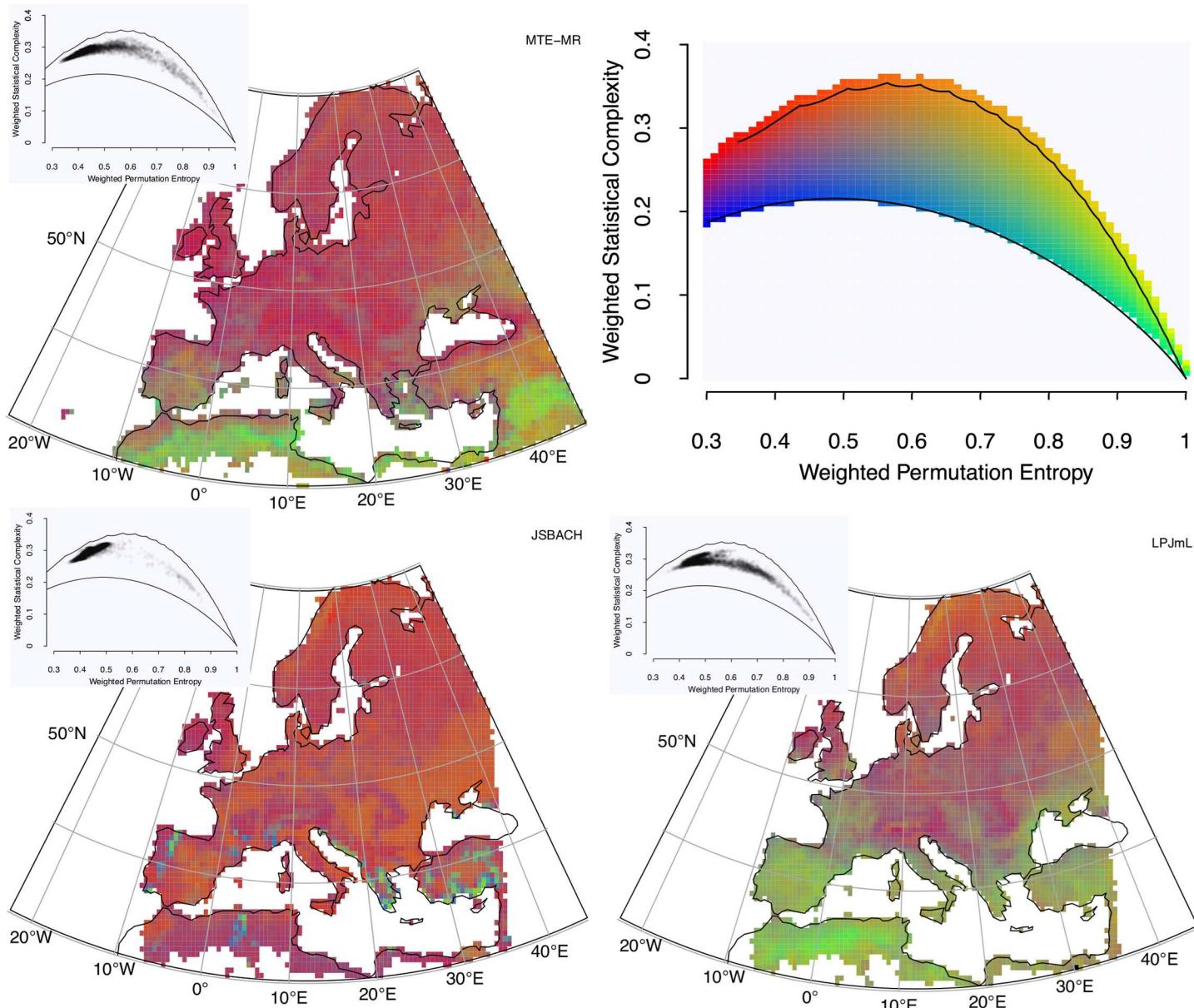
## Model-data comparison using ITQ

ITQ calculated from long-term GPP time series (1982–2011) at monthly resolution across Europe from two models and an observations-based product are shown in Fig 5 ( $\mathcal{H}_w \times \mathcal{C}_w$  and Fig 6  $\mathcal{H}_w \times \mathcal{F}_w$ ), using a 2-dimensional colour scheme for visualization in a single map [80]. The observations-based product (MTE reconstructions, upper panels) indicate large, comparatively homogeneous regions across major biogeographical gradients, where the entropy of GPP time series is intermediate and their complexity medium to high. In the Mediterranean and in North Africa, the entropy is higher and the complexity and Fisher Information lower.

In the context of this comparison, MTE results are considered as “surrogate observations” or benchmarking datasets to be reproduced by the land surface models—a typical scenario in carbon cycle research [81, 82]. Accepting this assumption for the sake of comparison, models are considered “realistic” if their output is close to the MTE results (Figs 5 and 6). The first observation is that JSBACH produces a spatially very homogeneous ITQ field, whereas LPJmL exhibits a pronounced North-South gradient. A striking difference between the models is that JSBACH is more or less reproducing the ITQ from MTE, whereas LPJmL generates output which appears with higher information content than MTE. However, in seasonally dry regions such as southern parts of the Iberian Peninsula and North Africa, MTE and LPJmL seem to point in a similar direction towards higher entropy time series, whereas JSBACH does not show this feature. Thus, in these regions LPJmL performs better. These patterns likewise become obvious when looking at the density of points in the planes (insets), where JSBACH has much lower density in the high-entropy and low-complexity/low Fisher Information region than the observations; for LPJmL, the reverse holds true.

An interesting question in the context of model-data comparison using ITQ relates to whether the derived mismatch patterns (Figs 5 and 6) would be reproduced by traditional model performance metrics (e.g. the root mean squared error, RMSE).

In Fig 7, colour-coded Jensen-Shannon divergences for the ordinal pattern distributions between the observation dataset MTE-MR and either of the two models are visualized and can be seen as an ITQ-based model performance metric. These maps demonstrate that JSBACH model output is rather close to the MTE-reconstructions in large parts of Central and Eastern Europe, but shows some deviations in Scandinavia, North-West Russia, and Northern Africa. In contrast, LPJmL is in better agreement with the MTE in Scandinavia, but LPJmL deviates from MTE in the Mediterranean and in several temperate European regions. Overall, LPJmL is the model with higher entropy values (at the monthly time scale) having a bias towards the

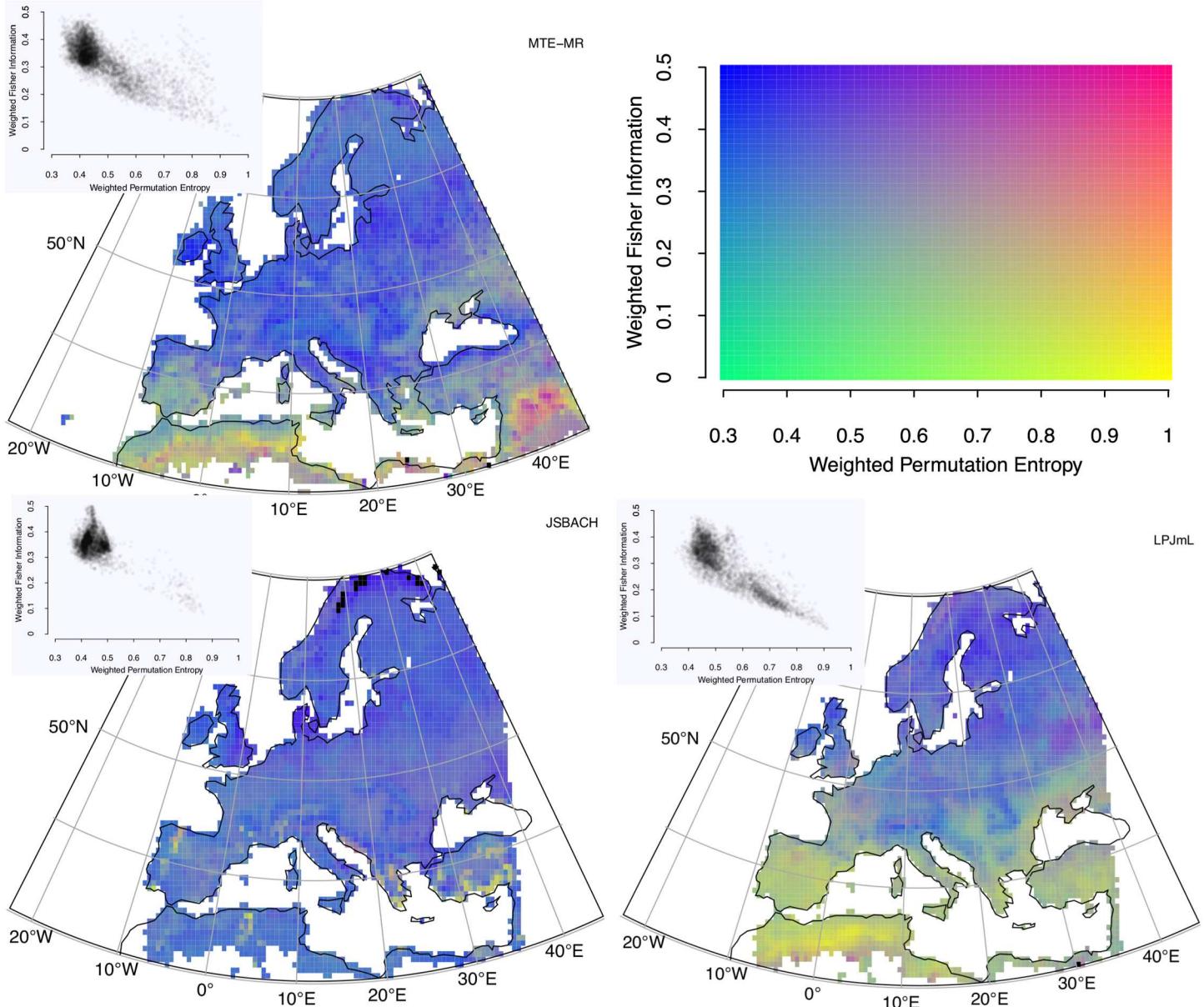


**Fig 5. Model-Data comparison of simulated and observations-based GPP over the European continent.** Colour-coded  $\mathcal{H}_w \times \mathcal{C}_w$  (top right), as obtained from (top left) MTE-MR, (bottom left) JSBACH, and (bottom right) LPJmL. The causal information planes illustrating the point densities are given in the insets.

doi:10.1371/journal.pone.0164960.g005

random case (at least relative to MTE). JSBACH largely reproduces the behaviour of the MTE dataset, although it overemphasizes deterministic components in the Mediterranean and in the northern boreal.

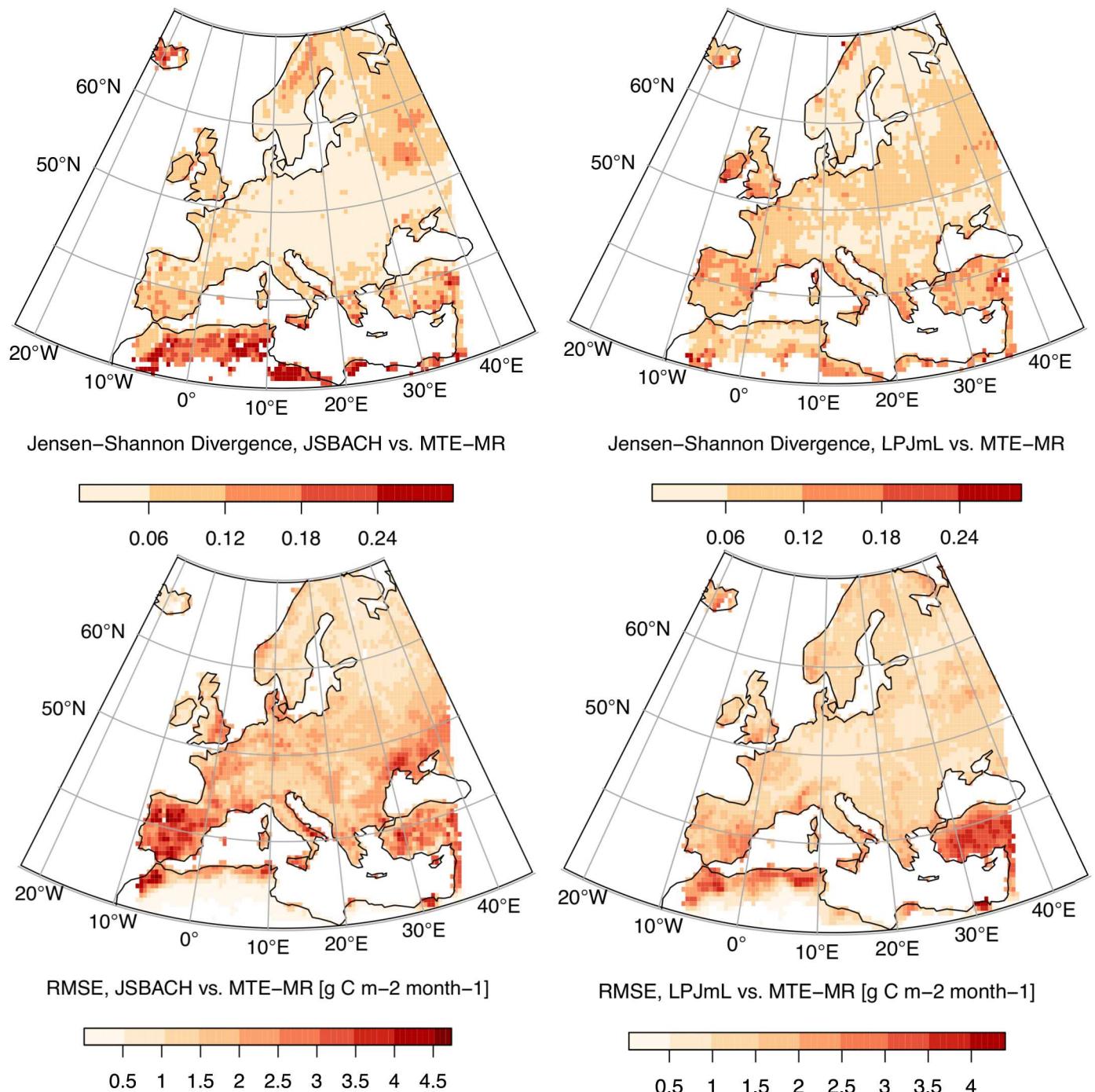
While this analysis is useful to investigate deviations in the dynamics, as quantified by the ordinal pattern distributions, we emphasize that it can complement, but not replace traditional model evaluation metrics. This is because the ITQ analysis performed here cannot account for biases in the mean, variance or higher statistical moments of model output. A direct



**Fig 6. Model-Data comparison of simulated and observations-based GPP over the European continent.** Colour-coded  $\mathcal{H}_w \times \mathcal{F}_w$  (bottom right), as obtained from (top left) MTE-MR, (bottom left) JSBACH, and (bottom right) LPJmL. The causal information planes illustrating the point densities are given in the insets.

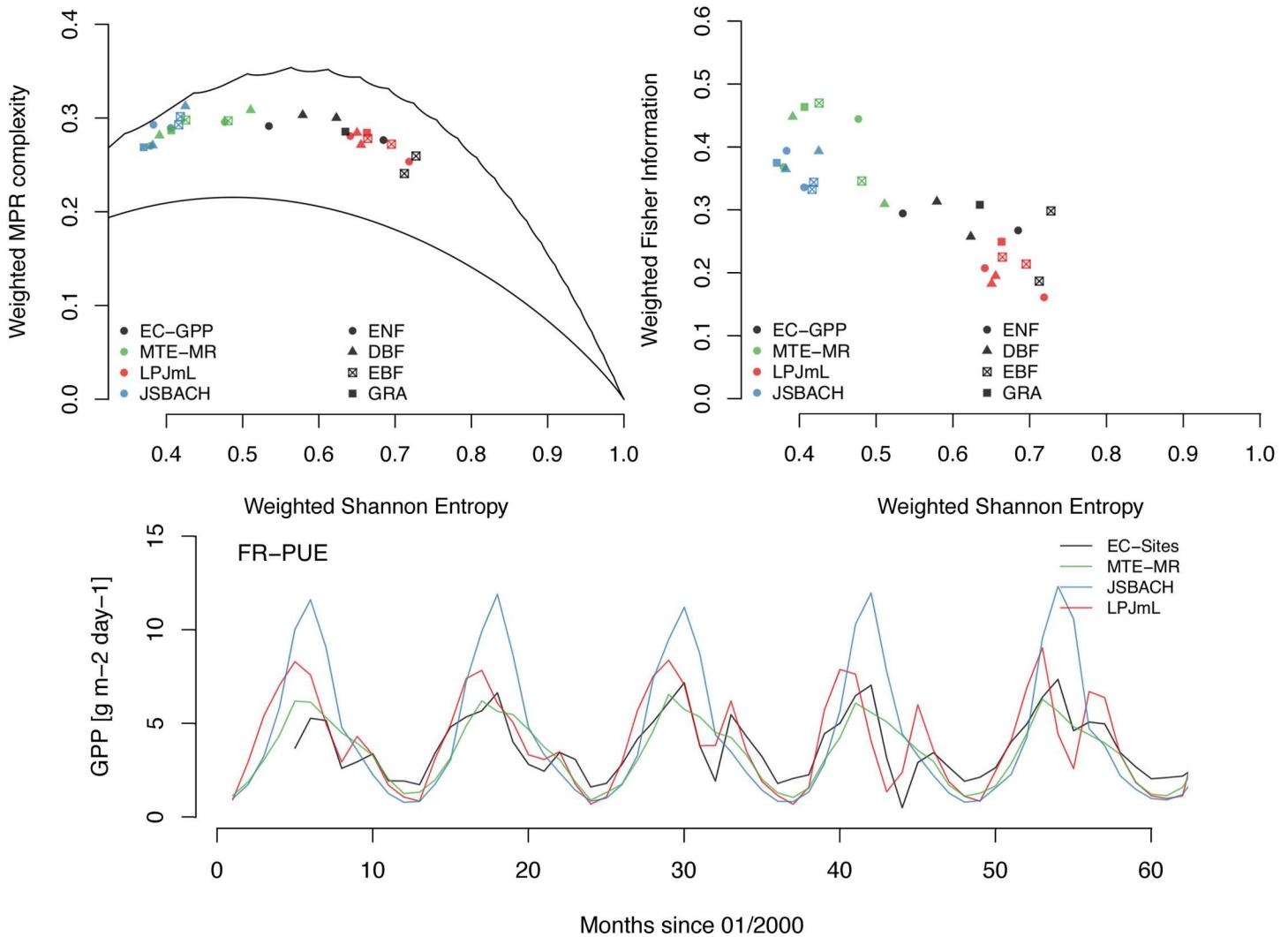
doi:10.1371/journal.pone.0164960.g006

comparison between a conventional model evaluation metric (root mean squared error, RMSE) and the complexity-based metrics (i.e. Jensen-Shannon divergence) reveals that the latter indeed complements traditional model evaluation tools (Fig 7): For the majority of simulated grid cells, the two metrics behave opposite to each other, i.e. the model with a lower RMSE performs worse on the dynamics (i.e. higher Jensen-Shannon divergence), and vice versa (S4 Fig). Yet, the two metrics are not strongly anticorrelated ( $R = -0.296$ ) and thus not redundant to each other.



**Fig 7. Model-Data comparison of simulated and observed GPP over the European continent.** Jensen-Shannon Divergence between (A) JSBACH and MTE-MR, and (B) LPJmL and MTE-MR. RMSE between (C) JSBACH and MTE-MR, and (D) LPJmL and MTE-MR.

doi:10.1371/journal.pone.0164960.g007



**Fig 8. ITQ-based model-data comparison at site scale for Mediterranean flux tower sites.** (A)  $\mathcal{H}_w \times \mathcal{C}_w$ , and (B)  $\mathcal{H}_w \times \mathcal{F}_w$ . (C) Time series of GPP at Puechabon, France.

doi:10.1371/journal.pone.0164960.g008

Given the observed qualitative discrepancies between the “surrogate observations” and both models in their representation of GPP dynamics using ITQ, a crucial question arises: *Can these findings be explained or interpreted in an ecologically meaningful way?*

To this end, we focus on the Mediterranean region because here the qualitative differences are most pronounced (Figs 5–7). Fig 8 shows an ensemble of Mediterranean flux tower observations (all sites with more than 5 years of continuous data, see Table 1 in S1 File), in addition to the respective pixels in MTE and the two models. The ensemble of sites in  $\mathcal{H}_w \times \mathcal{C}_w$  and  $\mathcal{H}_w \times \mathcal{F}_w$  (Fig 8) largely confirms the observed (spatial) pattern: In the Mediterranean, JSBACH simulations are very homogeneous, close to MTE, and these two point clouds cluster towards the more deterministic parts of the causal information planes. On the other hand, LPJmL tends to produce time series with higher information content, leading to a poor ITQ-based performance relative to MTE (see above). Surprisingly though, site-based GPP measurements tend to fall closer to LPJmL than to the other two datasets in the planes. This indicates

that the observations-based MTE product (and JSBACH) do not necessarily match the site-level GPP measurements from an ITQ perspective. However, a ‘process-oriented’ interpretation of these ITQ-based patterns and discrepancies is feasible, considering, for example, the monthly GPP time series of an evergreen broadleaf forest site (Puechabon, France, Fig 8, bottom panel): Here, MTE and JSBACH indicate a very simple seasonal oscillation with only one (summer) peak per season. In contrast, flux tower measurements and LPJmL exhibit a two-peaked seasonal structure, with an early summer peak, subsequent GPP reduction due to water limitation, followed by a (smaller) autumn peak once water limitation ceases. Hence, in the model-data comparison presented above, ITQ readily diagnose these two contrasting dynamics, which can thus be used as a starting point to improve or optimize models and observations-based datasets.

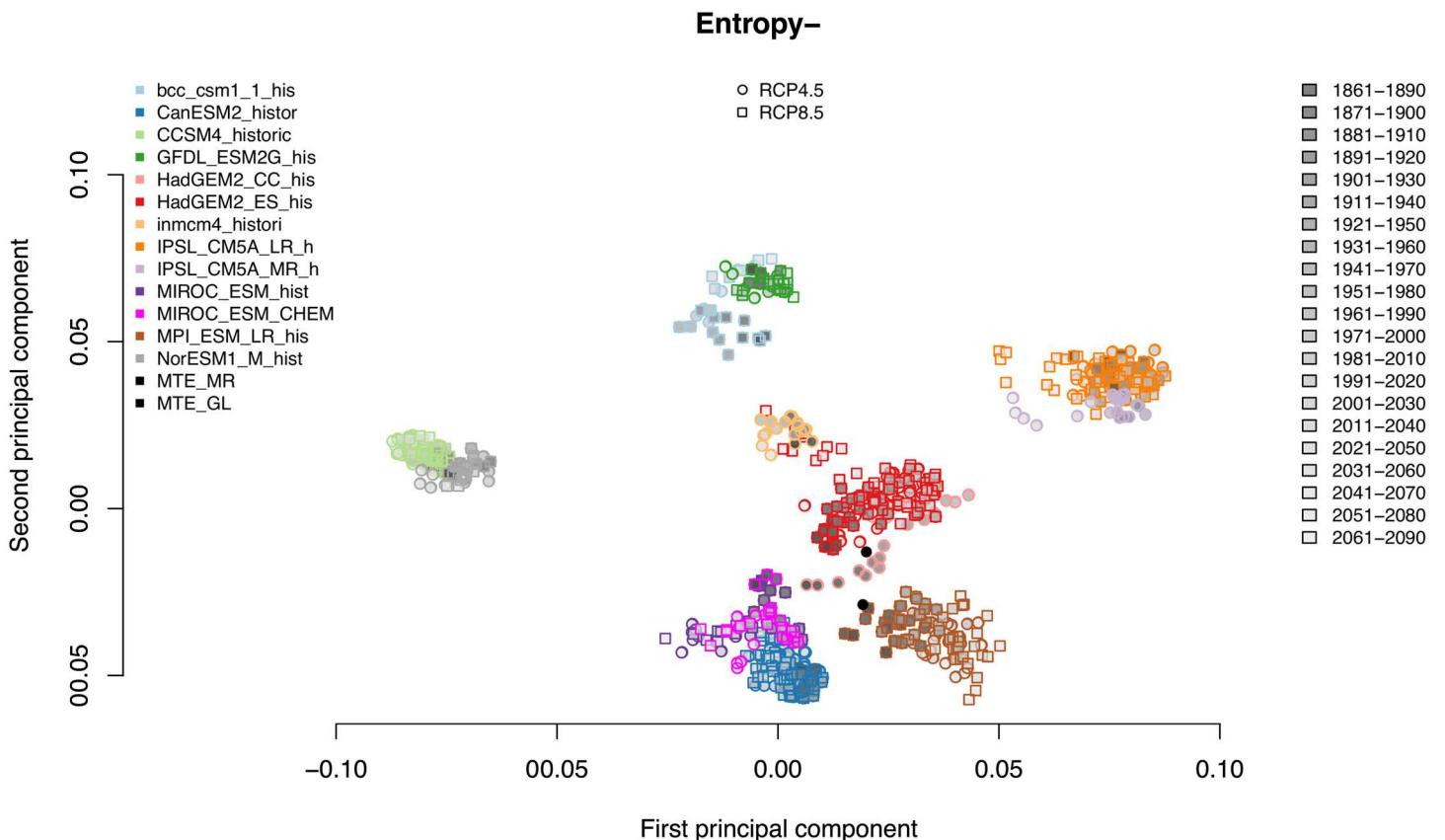
In summary, these results show that model evaluation and improvement based solely on absolute or relative error metrics could be misleading if the dynamics are misrepresented in the model; in contrast a joint consideration of a variety of benchmarking metrics might provide useful hints to scrutinizing various aspects of model behaviour.

## Global analysis of land surface models

The CMIP5 runs allow a detailed evaluation of the relative importance of the model choice and the scenario used (cf. Table 2 in S1 File). Global maps of the Weighted Permutation Entropy (S6 Fig) and the causal information planes for one of the 30-years periods and all 59 model runs are shown in S7 and S8 Figs. These plots show a large diversity of patterns and point densities in  $\mathcal{H}_w \times \mathcal{C}_w$  across the runs. Some of the models extend over the whole entropy range, others are more constrained to higher entropy values. Although looking quite similar at first glance for most of the 59 model simulations, it is obvious that the local density of points is rather different between them. None of the simulations is very close to the upper limit curve, i.e. do not achieve the highest complexity possible, which would be the realm of deterministic dynamical systems. This might be explained by the fact that these models are not autonomous dynamical systems, but driven by stochastic drivers such as precipitation. In  $\mathcal{H}_w \times \mathcal{F}_w$ , many simulations exhibit a “fork” structure, or a separation of the values into (at least) two different “curves” with two different slopes (S8 Fig). The width of the point clouds around these “curves” is, however, rather different between the simulations. Global maps of the Weighted Permutation Entropy confirm that different models indeed show considerable differences in the simulated dynamics of GPP across various regions of the globe (S6 Fig). Using the rastering presented in the Methods section, we performed a PCA of the point clouds. The result for the first two dimensions is given by Fig 9. This figure shows that the entropy and Fisher information measures separate models according to model *structural* properties, and not scenarios, which is the core result of this section. Note that in all cases where clusters contain points of different colours, the colour-encoded models are actually variants of each other, for example, differing in spatial resolution or with or without a module for atmospheric chemistry. The classification is not sensitive to the two scenarios, which were identical for all models. The linear or bow-shaped structures of the points for some of the models indicate a systematic development during the simulation period.

## Conclusions

The main intention of the current paper was to test the usefulness of Information Theory Quantifiers as tools for data analysis and model benchmarking in the environmental sciences, and thus to reach a wider dissemination of these tools. Time series of simulated and observed GPP as a key ecosystem variable were investigated to show that the dynamics varies as a



**Fig 9. Differentiating model structure with statistical complexity measures.** The different models are coded by different colours, filling indicates the analyzed time period for all grid cells of the globe.

doi:10.1371/journal.pone.0164960.g009

function of the underlying generating model or process, spatio-temporal scale and climate and ecosystem type, reflecting that this variable has an intricate temporal behaviour and is relevant for applications on local to global scales. Hence, statistical complexity measures can be used as model and data analytical tools that diagnose the short-term dynamics of the underlying time series. This is complementary to most conventional model diagnostics quantifying global statistical properties of the data, such as bias, correlation, variance, or higher statistical moments, which are in part insensitive to the temporal order (non-causal).

Three central results were presented in the present paper: First, ITQ derived from time series of remotely sensed vegetation activity proxies (FAPAR in our case) show both climate- and ecosystem-related continental-scale gradients and regional structures, and thus might serve as useful, efficient and robust data-analytical tools. Second, a model-data comparison based on ITQ reveals qualitatively different results compared to traditional model evaluation metrics; e.g., as demonstrated for European GPP simulations, a model with low absolute or relative error does not necessarily reproduce the dynamics of the observations well, and vice versa. A striking result is that the GPP data set that is constructed by the MTE approach has different, much more deterministic ITQ than the EC observations that enter the MTE procedure. Discrepancies between models and/or data provide useful hints to model structural deficiencies (as shown for Mediterranean FLUXNET sites). Third, the complexity of model simulated GPP is highly sensitive to model structure (i.e. differing between models), but largely insensitive to external climate forcing,

atmospheric greenhouse gas concentrations or land use changes, as demonstrated for global- and biome-scale GPP simulations from the CMIP5 archive. These results highlight the benefits of model benchmarking and evaluation against a variety of model evaluation metrics, including model structural diagnostics such as the ITQ presented here.

ITQ allow a ranking of models according to their ability to reproduce the observed dynamical behaviour within the time span given by the embedding dimension and the delay parameter. The crucial limiting factor for this time span is the length of the time series. In the future, the complexity indicators could serve as objective functions to improve model performance, e.g. in an iterative or machine-learning setting to find optimal parameter sets or model structures (Ilie et al., submitted to *Geoscientific Model Development*) or in a data assimilation system (see e.g. [83]). In general terms, the methods employed here analyze and compare data sets, not models. They are close to non-parametric—the choice of the embedding dimension is dictated by the length of the time series—and do not make any specific assumptions about properties of the data, statistical or otherwise. Our results indicate that for land surface models, it is likely not sufficient to change the parameters of the model to reproduce the observed behaviour; rather, the model structure has to be revised, since the same model structure produces similar patterns independent of e.g. their initialization and the details of the parametrization.

Finally, a broad range of future applications of Information Theory Quantifiers in environmental science is conceivable: These could consist of efficiently diagnosing satellite data streams of very high spatio-temporal resolution in an increasingly data-rich era, as well as to use statistical complexity of environmental variables for classification purposes. Therefore, we propose that these tools could be taken up more widely by the community for model evaluation and benchmarking activities.

## Supporting Information

**S1 Fig. Weighted statistical complexity for the JRC-TIP FPAR dataset.**  
(PDF)

**S2 Fig. Weighted Fisher Information for the JRC-TIP FPAR dataset.**  
(PDF)

**S3 Fig. High-resolution version of Fig 3 (bottom panel) in the main text.**  
(PDF)

**S4 Fig. ITQ-based and RMSE-based model comparison of JSBACH and LPJmL vs. MTE.**  
(left) Jensen-Shannon Divergence, and (right) RMSE.  
(PDF)

**S5 Fig. ITQ's for temperate and boreal FLUXNET sites.** (top left) Boreal sites,  $\mathcal{H}_w \times \mathcal{C}_w$ , and (top right) boreal sites  $\mathcal{H}_w \times \mathcal{F}_w$ . (bottom left) Temperate sites,  $\mathcal{H}_w \times \mathcal{C}_w$ , and (bottom right) temperate sites,  $\mathcal{H}_w \times \mathcal{F}_w$ .  
(PDF)

**S6 Fig. Global maps of the Weighted Permutation Entropy of GPP fluxes across 59 CMIP5 models and simulations for all land grid cells.** Plots are illustrative for the 30-yr period 1981–2010.  
(PDF)

**S7 Fig.  $\mathcal{H}_w \times \mathcal{C}_w$  of GPP fluxes across 59 CMIP5 models and simulations for all land grid cells.** Plots are illustrative for the 30-yr period 1981–2010.  
(JPG)

**S8 Fig.**  $\mathcal{H}_w \times \mathcal{F}_w$  of GPP fluxes across 59 CMIP5 models and simulations for all land grid cells. Plots are illustrative for the 30-yr period 1981–2010.

(JPG)

**S1 File.** Supporting Information file that contains additional Text and Tables.

(PDF)

**S1 Code.** Supporting Code Tutorial that introduces the R-package on Statistical Complexity (“statcomp”) and allows to reproduce Figs 1 and 2.

(R)

## Acknowledgments

We thank Kirsten Thoncke and Susanne Rolinski for providing LPJmL model output, Christian Beer for providing JSBACH results, data management and the modelling protocol. Furthermore, we thank Anja Rammig and Markus Reichstein for comments and discussions. S.S. is grateful to the ‘German National Academic Foundation’ (Studienstiftung des Deutschen Volkes) for support and the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC) for training. H.L. is grateful to the hospitality at the Instituto de Física, Universidade Federal de Alagoas in Maceió, Brazil, and also to the guest researcher funding provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, fellowship no. 310003/2016-4) during a sabbatical stay. O.A.R. acknowledges support from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. S.S., F.G., and M.D.M. thank the European Space agency for funding the STSE project CAB-LAB; S. S., F.G., and M.D.M. are also grateful to the European Commission for funding the BACI project; grant agreement No 640176. The JRC-TIP FAPAR product based on MODIS broadband albedo observations was provided by The Inversion Lab and the Joint Research Centre of the European Commission, and we thank Michael Voßbeck, Bernhard Pinty, and Nadine Gobron for making the dataset available. This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: CarboEuropeIP, CarboItaly, CarboMont, GreenGrass. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval, Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California—Berkeley and the University of Virginia. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table 2 in [S1 File](#)) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

## Author Contributions

**Conceptualization:** SS HL MDM MH OAR.

**Formal analysis:** SS HL.

**Funding acquisition:** MDM.

**Investigation:** MDM FG TK.

**Methodology:** SS HL.

**Resources:** MDM TK.

**Software:** SS TK HL OAR.

**Supervision:** MDM HL.

**Validation:** SS MDM PB FG.

**Visualization:** TK SS.

**Writing – original draft:** SS HL MH OAR.

**Writing – review & editing:** SS HL.

## References

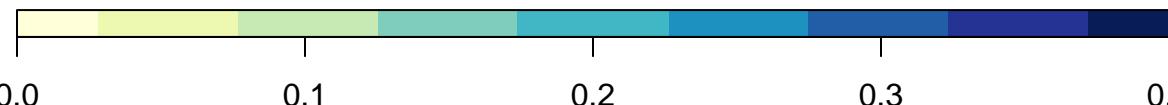
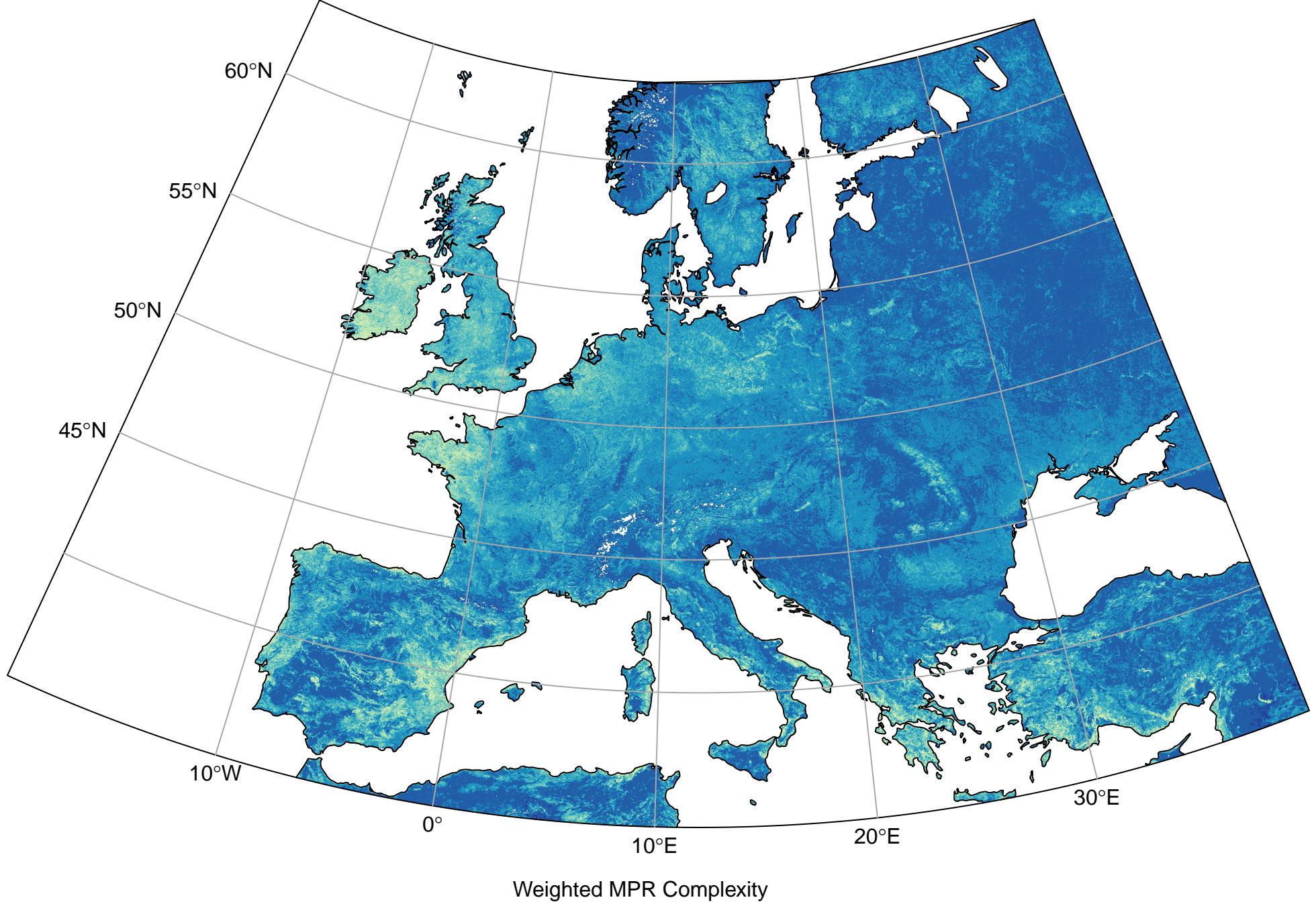
1. Ciais P, Sabine C, Bala G, Bopp L, Brovkin V, Canadell J, et al. Carbon and other biogeochemical cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press; 2014. p. 465–570.
2. Eyring V, Righi M, Evaldsson M, Lauer A, Wenzel S, Jones C, et al. ESMValTool (v1.0): a community diagnostic and performance metrics tool for routine evaluation of Earth System Models in CMIP. Geoscientific Model Development. 2016; 9:1747–1802. doi: [10.5194/gmd-9-1747-2016](https://doi.org/10.5194/gmd-9-1747-2016)
3. Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society. 2012; 93(4):485–498. doi: [10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
4. Kelley D, Prentice IC, Harrison S, Wang H, Simard M, Fisher J, et al. A comprehensive benchmarking system for evaluating global vegetation models. Biogeosciences. 2013; 10:3313–3340. doi: [10.5194/bg-10-3313-2013](https://doi.org/10.5194/bg-10-3313-2013)
5. Abramowitz G. Towards a benchmark for land surface models. Geophysical Research Letters. 2005; 32(22). doi: [10.1029/2005GL024419](https://doi.org/10.1029/2005GL024419)
6. Gupta HV, Clark MP, Vrugt JA, Abramowitz G, Ye M. Towards a comprehensive assessment of model structural adequacy. Water Resources Research. 2012; 48(8). doi: [10.1029/2011WR011044](https://doi.org/10.1029/2011WR011044)
7. Abramowitz G. Towards a public, standardized, diagnostic benchmarking system for land surface models. Geoscientific Model Development. 2012; 5(3):819–827. doi: [10.5194/gmd-5-819-2012](https://doi.org/10.5194/gmd-5-819-2012)
8. Luo YQ, Randerson JT, Abramowitz G, Bacour C, Blyth E, Carvalhais N, et al. A framework for benchmarking land models. Biogeosciences. 2012; 9(10):3857–3874. doi: [10.5194/bg-9-3857-2012](https://doi.org/10.5194/bg-9-3857-2012)
9. Best M, Abramowitz G, Johnson H, Pitman A, Balsamo G, Boone A, et al. The plumbing of land surface models: benchmarking model performance. Journal of Hydrometeorology. 2015; (2015: ). doi: [10.1175/JHM-D-14-0158.1](https://doi.org/10.1175/JHM-D-14-0158.1)
10. Carvalhais N, Forkel M, Khomik M, Bellarby J, Jung M, Migliavacca M, et al. Global covariation of carbon turnover times with climate in terrestrial ecosystems. Nature. 2014; 514(7521):213–217. doi: [10.1038/nature13731](https://doi.org/10.1038/nature13731) PMID: [25252980](#)
11. Grimm V, Railsback SF. Pattern-oriented modelling: a ‘multi-scope’ for predictive systems ecology. Phil Trans R Soc B. 2012; 367(1586):298–310. doi: [10.1098/rstb.2011.0180](https://doi.org/10.1098/rstb.2011.0180) PMID: [22144392](#)
12. Reichstein M, Beer C. Soil respiration across scales: the importance of a model–data integration framework for data interpretation. Journal of Plant Nutrition and Soil Science. 2008; 171(3):344–354. doi: [10.1002/jpln.200700075](https://doi.org/10.1002/jpln.200700075)
13. Reichstein M, Mahecha MD, Ciais P, Seneviratne SI, Blyth EM, Carvalhais N, et al. Elk–testing climate–carbon cycle models: a case for pattern–oriented system analysis. iLEAPS Newsletter. 2011; 11:14–21.
14. Mahecha MD, Reichstein M, Jung M, Seneviratne SI, Zaehle S, Beer C, et al. Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales. Journal of Geophysical Research: Biogeosciences. 2010; 115(G2). doi: [10.1029/2009JG001016](https://doi.org/10.1029/2009JG001016)
15. Chapin SF III, Matson P P A and Vitousek. Principles of Terrestrial Ecosystem Ecology. 2nd ed. Springer; 2011.
16. Baldocchi D, Falge E, Gu L, Olson R, Hollinger D, Running S, et al. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux

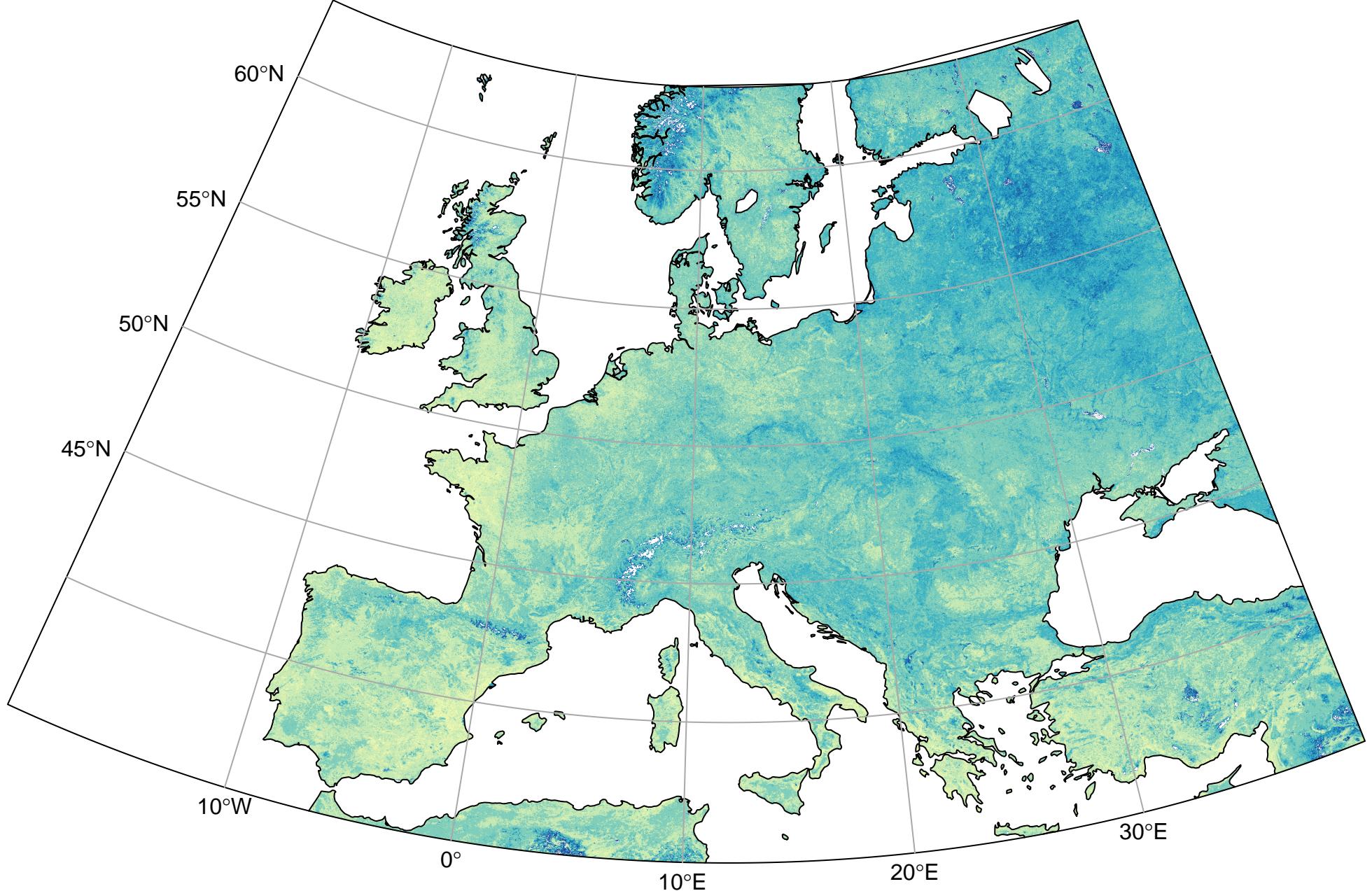






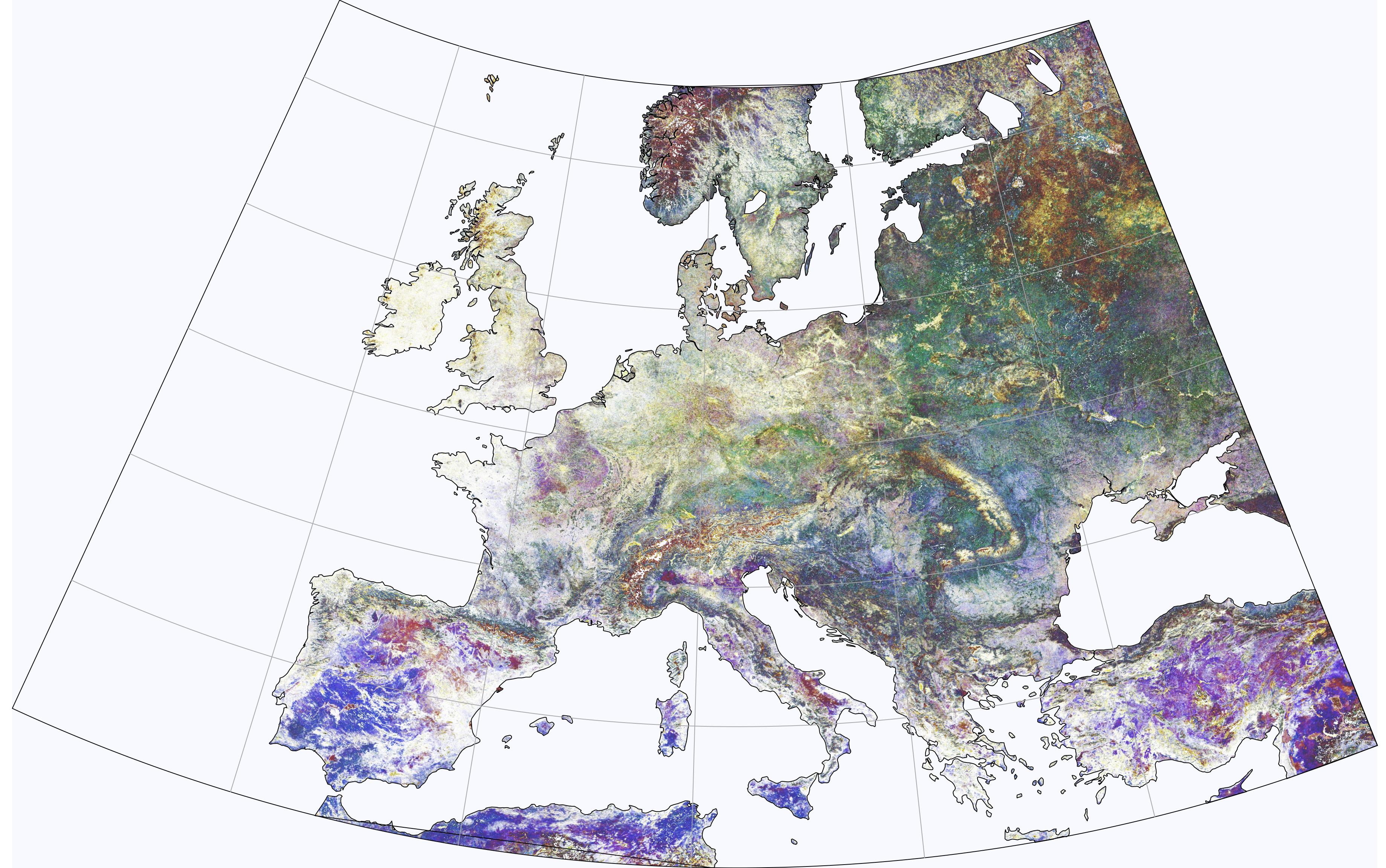
81. Piao S, Sitch S, Ciais P, Friedlingstein P, Peylin P, Wang X, et al. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO<sub>2</sub> trends. *Global Change Biology*. 2013; 19(7):2117–2132. doi: [10.1111/gcb.12187](https://doi.org/10.1111/gcb.12187) PMID: 23504870
82. Anav A, Friedlingstein P, Kidston M, Bopp L, Ciais P, Cox P, et al. Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth System Models. *Journal of Climate*. 2013; 26(18):6801–6843. doi: [10.1175/JCLI-D-12-00417.1](https://doi.org/10.1175/JCLI-D-12-00417.1)
83. Schürmann GJ, Kaminski T, Köstler C, Carvalhais N, Voßbeck M, Kattge J, et al. Constraining a land surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System. *Geoscientific Model Development Discussions*. 2016; 2016:1–24.

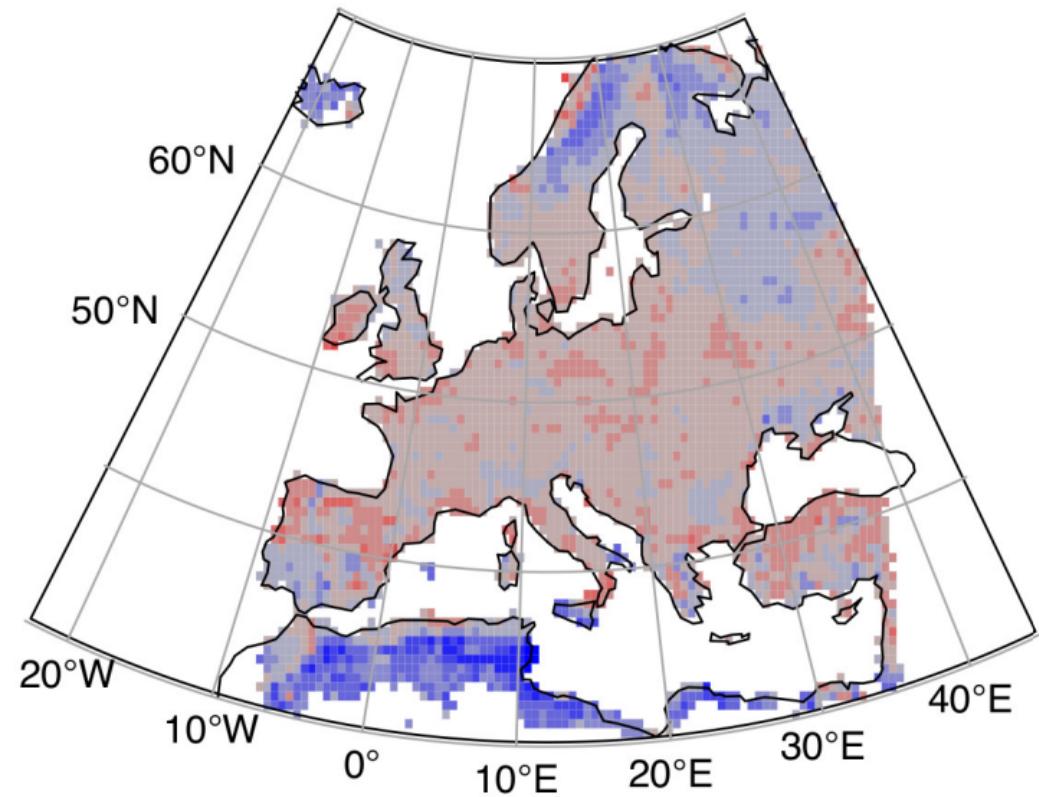




Weighted Fisher Information



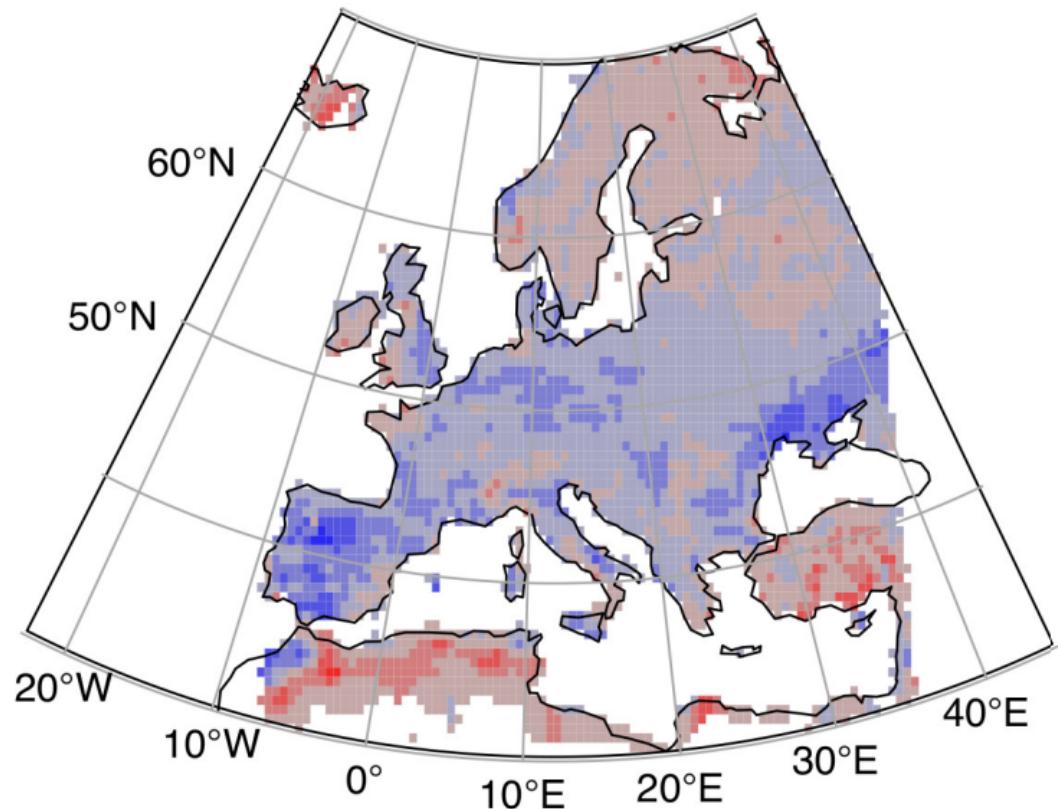




JS (JSBACH vs. MTE) – JS (LPJmL vs. MTE)



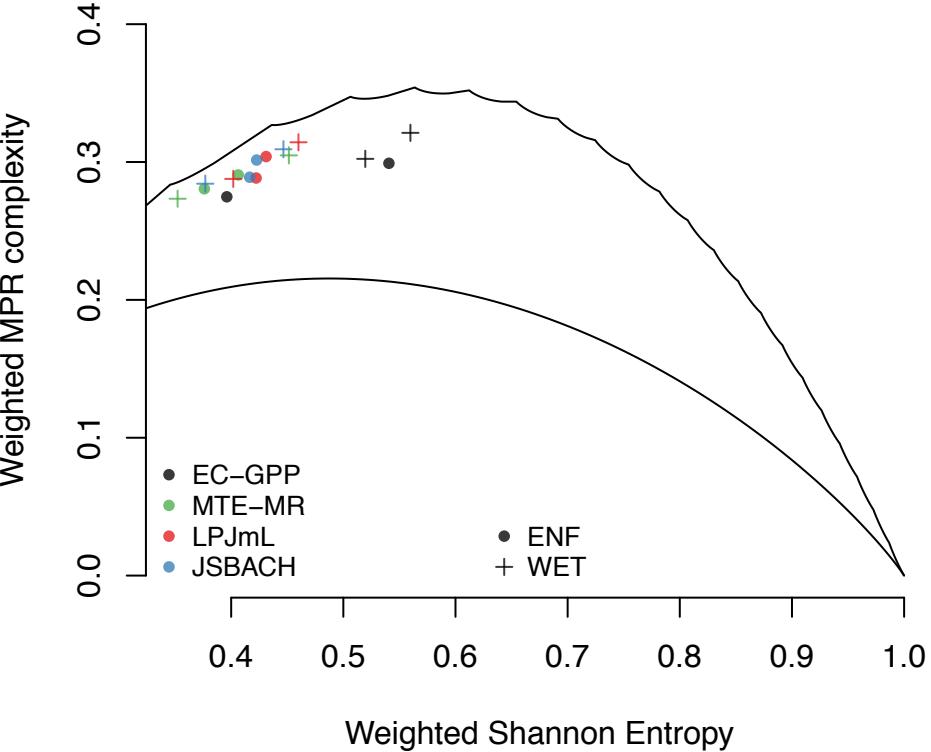
-0.15 -0.05 0.05 0.15 0.25



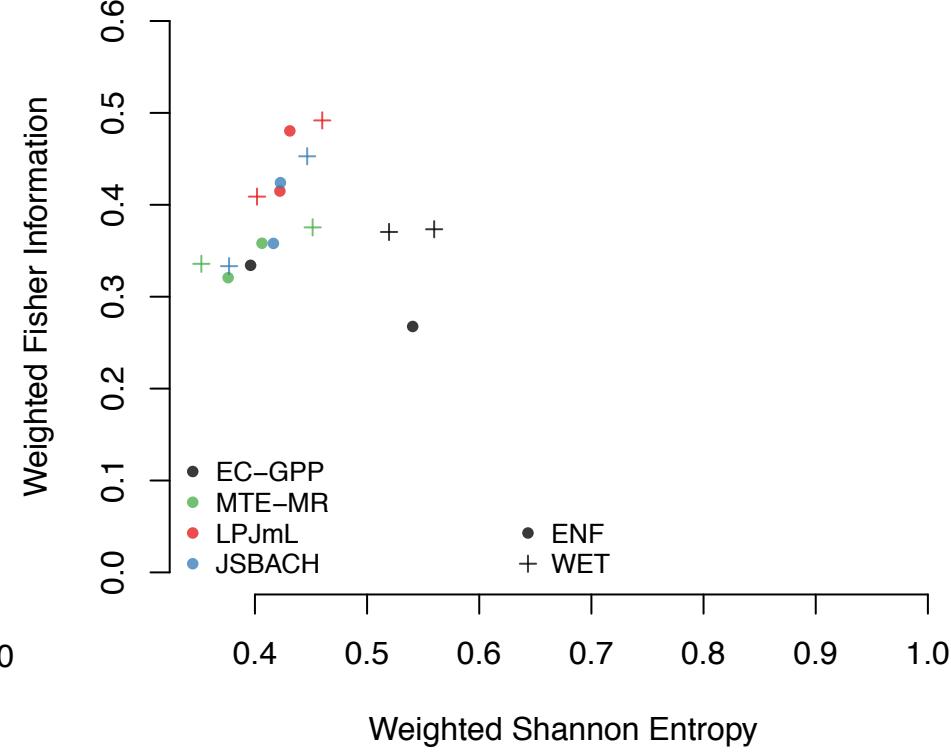
RMSE (JSBACH vs. MTE) – RMSE (LPJmL vs. MTE)



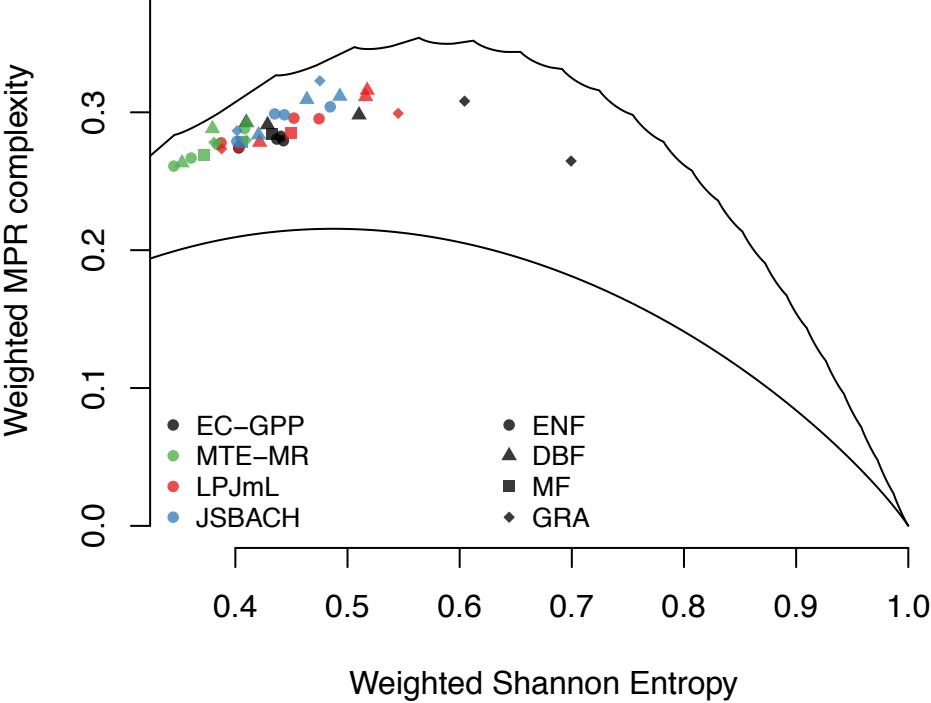
-3 -2 -1 0 1 2 3



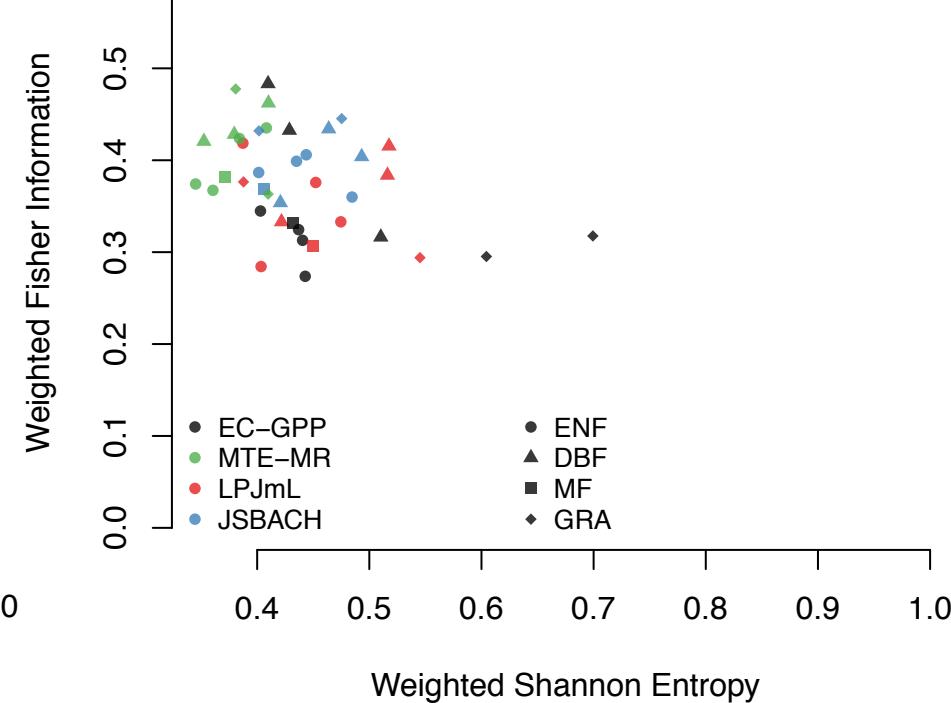
Weighted Shannon Entropy



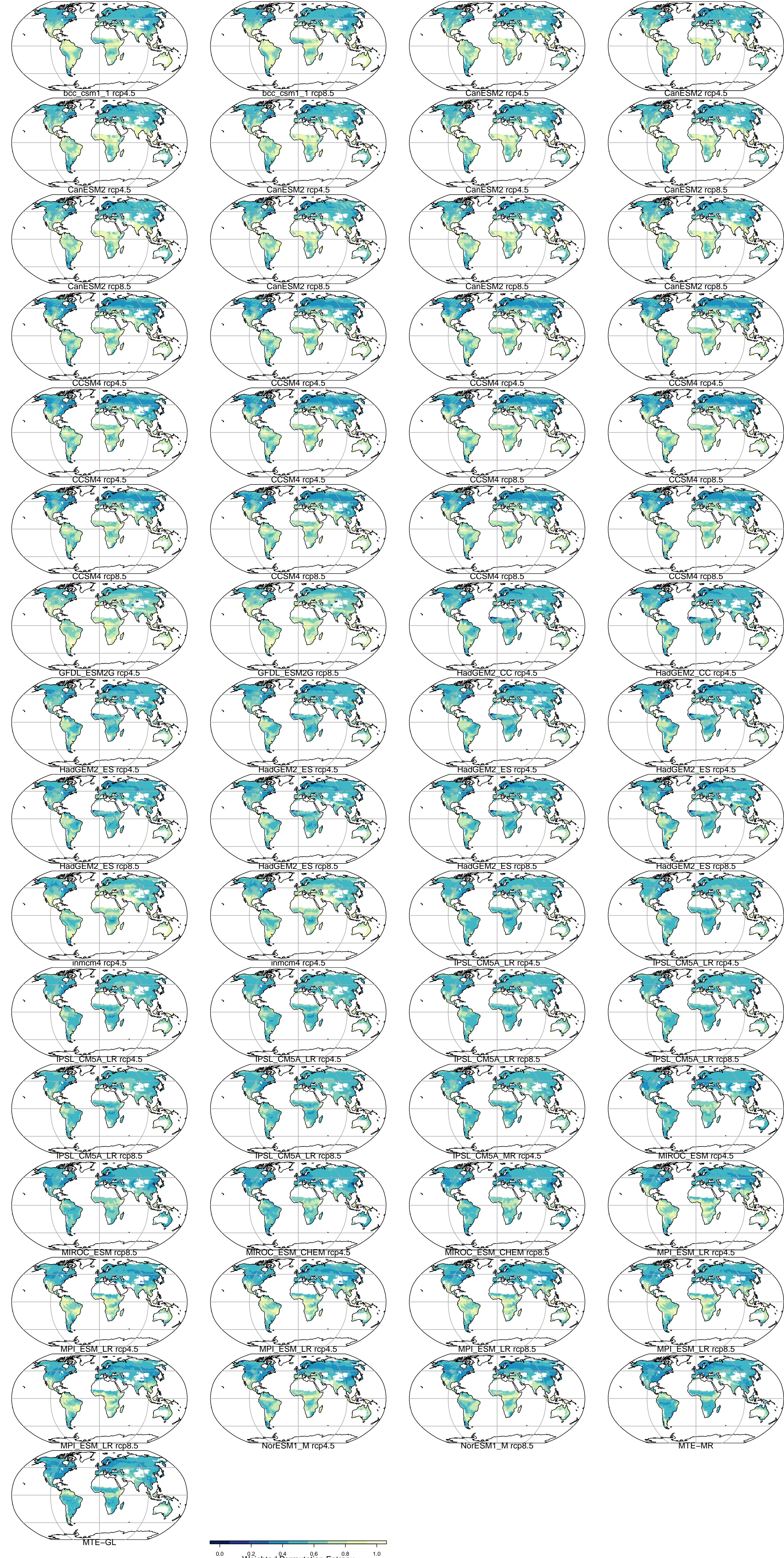
Weighted Shannon Entropy

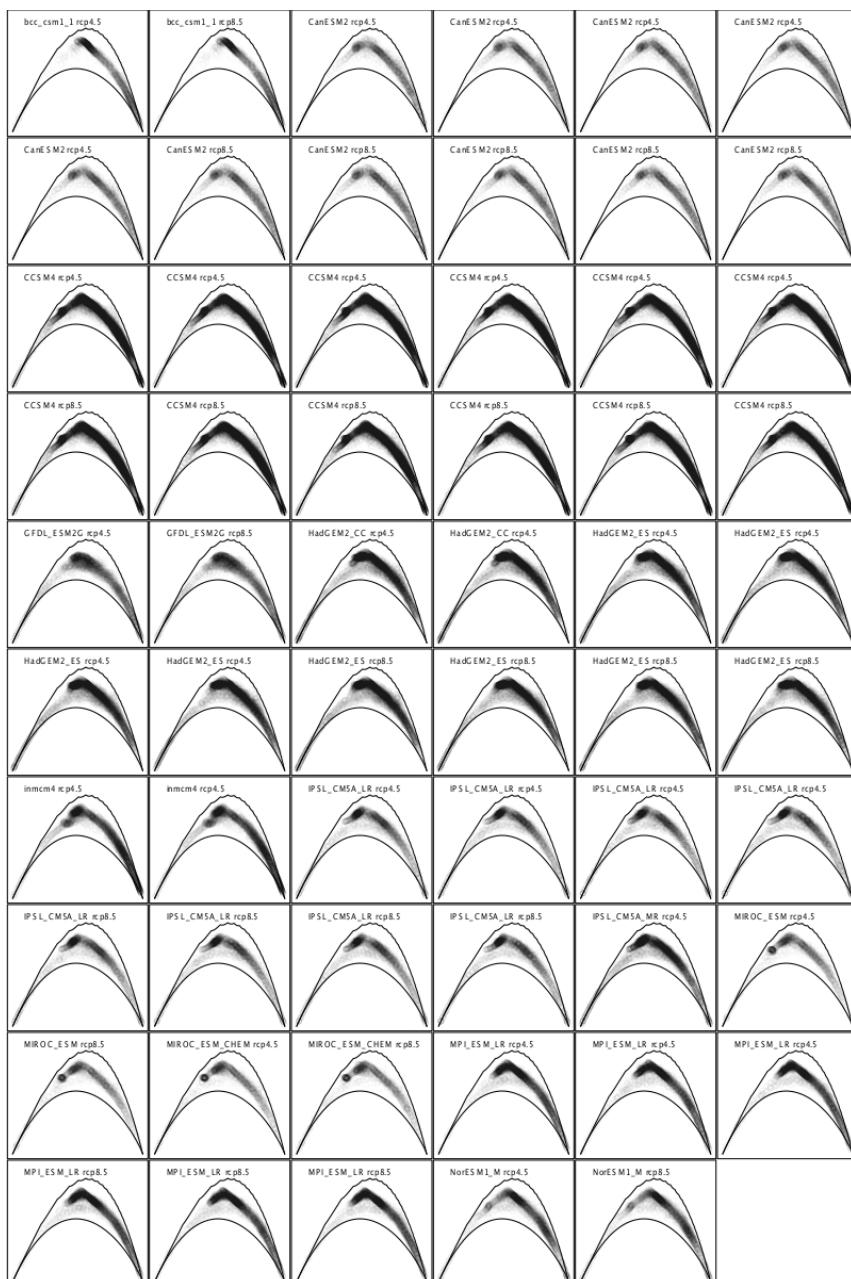


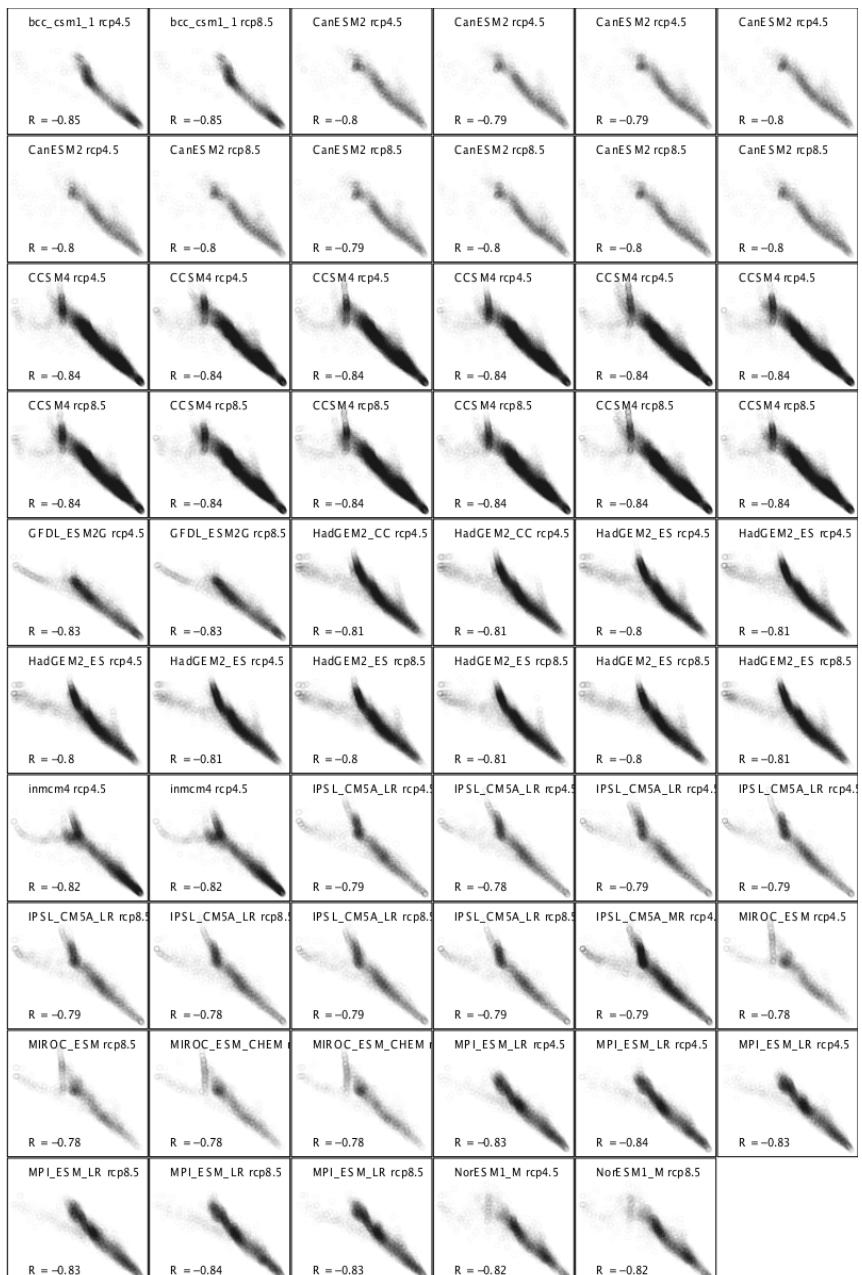
Weighted Shannon Entropy



Weighted Shannon Entropy







---

# S1 Supporting File to ‘Diagnosing the Dynamics of Observed and Simulated Ecosystem Gross Primary Productivity with Time Causal Information Theory Quantifiers’

Sebastian Sippel<sup>1\*</sup>, Holger Lange<sup>2,3</sup>, Miguel D. Mahecha<sup>1,4,5</sup>, Michael Hauhs<sup>6</sup>, Paul Bodesheim<sup>1</sup>, Thomas Kaminski<sup>7</sup>, Fabian Gans<sup>1</sup>, Osvaldo A. Rosso<sup>3,8,9</sup>

**1** Max Planck Institute for Biogeochemistry, Jena, Germany.

**2** Norwegian Institute of Bioeconomy Research, Ås, Norway.

**3** Instituto de Física, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil.

**4** German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany.

**5** Michael Stifel Center Jena for Data-Driven and Simulation Science, Jena, Germany.

**6** University of Bayreuth, Bayreuth, Germany.

**7** The Inversion Lab, Hamburg

**8** Instituto Tecnológico de Buenos Aires (ITBA) and CONICET, Ciudad Autónoma de Buenos Aires, Argentina.

**9** Complex Systems Group, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Las Condes, Santiago, Chile.

\* ssippel@bgc-jena.mpg.de

---

## Contents

1

<b>Tables</b>	<b>3</b>	2
<b>Text</b>	<b>5</b>	3
Ordinal pattern statistics . . . . .	5	4

## Tables

5

**Table 1.** All FLUXNET sites used in this study.

Name	Years	Long.	Lat.	PFT	Climate Region	Ref.
BE-Vie	1996-2006	6.00	50.31	MF	Temperate	[2]
CH-Oe1	2000-2010	7.73	47.29	GRA	Temperate	[3]
CZ-Bk1	2000-2006	18.54	49.50	ENF	Temperate	[4]
DE-Hai	2000-2006	10.45	51.08	DBF	Temperate	[5]
DE-Tha	1996-2006	13.57	50.96	ENF	Temperate	[6]
DE-Wet	2002-2006	11.46	50.45	ENF	Temperate	[5]
DK-Sor	1996-2006	11.64	55.49	DBF	Temperate	[6]
ES-ES1	1999-2006	-0.32	39.35	ENF	Mediterranean	[4]
FI-Hyy	1996-2006	24.29	61.85	ENF	Boreal	[7]
FI-Kaa	2000-2006	27.30	69.14	WET	Boreal	[8]
FI-Sod	2000-2006	26.64	67.36	ENF	Boreal	[9]
FR-Hes	1997-2006	7.06	48.67	DBF	Temperate	[10]
FR-Pue	2000-2006	3.60	43.74	EBF	Mediterranean	[4]
HU-Bug	2002-2006	19.60	46.69	GRA	Temperate	[11]
IT-Amp	2002-2006	13.61	41.90	GRA	Mediterranean	[11]
IT-Col	1996-2006	13.59	41.85	DBF	Mediterranean	[4]
IT-Cpz	1997-2006	12.50	41.71	EBF	Mediterranean	[4]
IT-Ren	1999-2006	11.43	46.59	ENF	Temperate	[12]
IT-Ro1	2000-2006	11.93	42.41	DBF	Mediterranean	[13]
IT-Sro	1999-2006	10.28	43.73	ENF	Mediterranean	[14]
NL-Loo	1996-2006	5.74	52.17	ENF	Temperate	[15]
RU-Fyo	1998-2006	32.92	56.46	ENF	Temperate	a.
SE-Deg	2001-2005	19.55	64.18	WET	Boreal	b.

a. <http://www.fluxdata.org:8080/sitepages/siteInfo.aspx?RU-Fyo>

b. <http://www.fluxdata.org:8080/sitepages/siteInfo.aspx?SE-Deg>

**Table 2.** All CMIP5 models used in this study. Number of ensemble members includes both representative concentration pathways (RCP4.5 and RCP 8.5).

Modeling Center (or Group)	Inst. ID	Model Name	# Ens.
Beijing Climate Center, China Meteorological Administration	BCC	bcc-csm1-1	2
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2	10
National Center for Atmospheric Research	NCAR	CCSM4	12
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-ESM2G	2
Met Office	MOHC	HadGEM2-CC	2
Hadley Centre	MOHC	HadGEM2-ES	8
Institute for Numerical Mathematics	INM	inmcm4	2
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-LR	8
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM	2
Max Planck Institute for Meteorology	MPI-M	MIROC-ESM-CHEM	2
Norwegian Climate Centre	NCC	NorESM1-M	2

---

## Text

6

### Ordinal pattern statistics

7

Given a one-dimensional time series  $\mathcal{X}(t) = \{x_t; t = 1, \dots, M\}$  and a chosen window length  $D$ ,  
“ordinal patterns” of order  $D$  are generated by first embedding the time series:  
8  
9

$$(s) \mapsto (x_{s-(D-1)\tau}, x_{s-(D-2)\tau}, \dots, x_{s-\tau}, x_s), \quad (1)$$

which assigns to each time  $s$  the  $D$ -dimensional vector of values at times  $s - (D - 1)\tau, \dots, s - \tau, s$ .  
Clearly, the greater  $D$ , the more information on the past is incorporated into the embedding vectors.  
By “ordinal pattern” related to the time  $(s)$ , we mean the permutation  $\pi = (r_0, r_1, \dots, r_{D-1})$  of  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
[0, 1, …,  $D - 1$ ] defined by

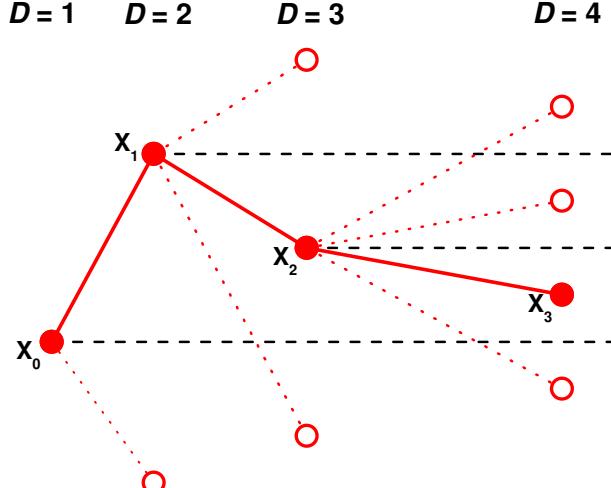
$$x_{s-r_{D-1}\tau} \leq x_{s-r_{D-2}\tau} \leq \dots \leq x_{s-r_1\tau} \leq x_{s-r_0\tau}. \quad (2)$$

In this way the vector defined by Eq. (1) is converted into a unique symbol  $\pi$ . We set  $r_i < r_{i-1}$  if  
 $x_{s-r_i} = x_{s-r_{i-1}}$  for uniqueness.  
14  
15

In order to illustrate the Bandt-Pompe (BP) method, we will consider a simple example: a time  
series with seven ( $M = 7$ ) values  $\mathcal{X} = \{4, 7, 9, 10, 6, 11, 3\}$  only, and we evaluate the BP-PDF for  
 $D = 3$  and  $\tau = 1$ . In this case the state space is divided into  $3!$  partitions and 6 mutually exclusive  
permutation symbols are considered. The triplet  $(4, 7, 9)$  and  $(7, 9, 10)$  represent the permutation  
pattern  $[012]$  since they are in increasing order. On the other hand,  $(9, 10, 6)$  and  $(6, 11, 3)$   
correspond to the permutation pattern  $[201]$  since  $x_{s+2} < x_s < x_{s+1}$ , while  $(10, 6, 11)$  has the  
permutation pattern  $\{102\}$  with  $x_{s+1} < x_s < x_{s+2}$ . Then, the associated probabilities to the 6  
patterns are:  $p([012]) = p([201]) = 2/5$ ;  $p([102]) = 1/5$ ;  $p([021]) = p([120]) = p([210]) = 0$ .  
For all the  $D!$  possible orderings (permutations)  $\pi_i$ , their relative frequencies can be naturally  
computed according to the number of times this particular order sequence is found in the time series,  
divided by the total number of sequences,  
27  
28

$$p(\pi_i) = \frac{\#\{s | s \leq N - (D - 1)\tau; (s) \text{ is of type } \pi_i\}}{N - (D - 1)\tau}, \quad (3)$$

where  $\#$  denotes cardinality. Thus, an ordinal pattern probability distribution  
 $P = \{p(\pi_i), i = 1, \dots, D!\}$  is obtained from the time series.  
27  
28



**Fig 1.** Illustration of the construction principle for ordinal patterns of length  $D$  [1]. If  $D = 4$  and  $\tau = 1$ , full circles and continuous lines represent the sequence of values  $x_0 < x_1 > x_2 > x_3$  which lead to the pattern  $\pi = [0321]$ .

Figure 1 illustrates the construction principle of the ordinal patterns of length  $D = 2, 3$  and  $4$  with  $\tau = 1$  [1]. Consider the sequence of observations  $\{x_0, x_1, x_2, x_3\}$ . For  $D = 2$ , there are only two possible directions from  $x_0$  to  $x_1$ : up and down. For  $D = 3$ , starting from  $x_1$  (up) the third part of the pattern can be above  $x_1$ , below  $x_0$ , or between  $x_0$  and  $x_1$ . A similar situation can be found starting from  $x_1$  (down). For  $D = 4$ , for each one of the six possible positions for  $x_2$ , there are four possible localizations for  $x_3$ , yielding  $D! = 4! = 24$  different possible ordinal patterns. In Fig. 1, full circles and continuous lines represent the sequence values  $x_0 < x_1 > x_2 > x_3$ , which leads to the pattern  $\pi = [0321]$ . A graphical representation of all possible patterns corresponding to  $D = 3, 4$  and 5 can be found in Fig. 2 of Parlitz *et al.* [1].

The embedding dimension  $D$  plays an important role in the evaluation of the appropriate probability distribution, because  $D$  determines the number of accessible states  $D!$  and also conditions the minimum acceptable length  $M \gg D!$  of the time series that one needs in order to work with reliable statistics [16].

---

## References

1. Parlitz U, Berg S, Luther S, Schirdewan A, Kurths J, Wessel N. Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Computers in biology and medicine*. 2012;42(3):319–327.
2. Aubinet M, Chermanne B, Vandenhaute M, Longdoz B, Yernaux M, Laitat E. Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes. *Agricultural and Forest Meteorology*. 2001;108(4):293–315.
3. Ammann C, Flechard C, Leifeld J, Neftel A, Fuhrer J. The carbon budget of newly established temperate grassland depends on management intensity. *Agriculture, Ecosystems & Environment*. 2007;121(1):5–20.
4. Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*. 2005;11(9):1424–1439.
5. Anthoni P, Knohl A, Rebmann C, Freibauer A, Mund M, Ziegler W, et al. Forest and agricultural land-use-dependent CO<sub>2</sub> exchange in Thuringia, Germany. *Global Change Biology*. 2004;10(12):2005–2019.
6. Grünwald T, Bernhofer C. A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt. *Tellus B*. 2007;59(3):387–396.
7. Suni T, Rinne J, Reissell A, Altimir N, Keronen P, Rannik U, et al. Long-term measurements of surface fluxes above a Scots pine forest in Hyytiala, southern Finland, 1996–2001. *Boreal Environment Research*. 2003;8(4):287–302.
8. Aurela M, Laurila T, Tuovinen JP. Seasonal CO<sub>2</sub> balances of a subarctic mire. *Journal of Geophysical Research: Atmospheres*. 2001;106(D2):1623–1637.
9. Suni T, Berninger F, Vesala T, Markkanen T, Hari P, Mäkelä A, et al. Air temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. *Global Change Biology*. 2003;9(10):1410–1426.
10. Granier A, Ceschia E, Damesin C, Dufrêne E, Epron D, Gross P, et al. The carbon balance of a young beech forest. *Functional ecology*. 2000;14(3):312–325.

- 
11. Gilmanov T, Soussana J, Aires L, Allard V, Ammann C, Balzarolo M, et al. Partitioning European grassland net ecosystem CO<sub>2</sub> exchange into gross primary productivity and ecosystem respiration using light response function analysis. *Agriculture, ecosystems & environment*. 2007;121(1):93–120.
  12. Cescatti A, Marcolla B. Drag coefficient and turbulence intensity in conifer canopies. *Agricultural and forest meteorology*. 2004;121(3):197–206.
  13. Arain MA, Restrepo-Coupe N. Net ecosystem production in a temperate pine plantation in southeastern Canada. *Agricultural and Forest Meteorology*. 2005;128(3):223–241.
  14. Chiesi M, Maselli F, Bindi M, Fibbi L, Cherubini P, Arlotta E, et al. Modelling carbon budget of Mediterranean forests using ground and remote sensing measurements. *Agricultural and Forest Meteorology*. 2005;135(1):22–34.
  15. Dolman A, Moors E, Elbers J. The carbon uptake of a mid latitude pine forest growing on sandy soil. *Agricultural and Forest Meteorology*. 2002;111(3):157–170.
  16. Kowalski A, Martín M, Plastino A, Rosso O. Bandt–Pompe approach to the classical-quantum transition. *Physica D: Nonlinear Phenomena*. 2007;233(1):21–31.