Article Type: Concepts & synthesis

Running title: Intrinsic predictability & forecasting

Title: The intrinsic predictability of ecological time series and its potential to guide forecasting

Frank Pennekamp[1,§,*], Alison C. Iles[2,5,6 §], Joshua Garland[3], Georgina Brennan[4], Ulrich Brose[5,6], Ursula Gaedke[7], Ute Jacob[8], Pavel Kratina[9], Blake Matthews[10], Stephan Munch[11, 12], Mark Novak[2], Gian Marco Palamara[1,13], Björn C. Rall[5,6], Benjamin Rosenbaum[5,6], Andrea Tabi[1], Colette Ward[1], Richard Williams[14], Hao Ye[15], Owen L. Petchey[1]

[§] F. Pennekamp and A. Iles contributed equally to this work

* Corresponding author: frank.pennekamp@ieu.uzh.ch

[1]University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland,

[2]Oregon State University, 3029 Cordley Hall, Corvallis, OR 97331,

[3]Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501

[4]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor LL57 2UW, United Kingdom

[5]EcoNetLab - Theory in Biodiversity Science, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103, Leipzig, Germany

[6]Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger-Str. 159, 07743, Jena, Germany

[7]Institute for Biology, University of Potsdam
[8]University of Hamburg
[9]Queen Mary University of London
[10]Eawag, Department of Aquatic Ecology, Center for Ecology, Evolution and Biogeochemistry, Seestrasse 79, 6047, Kastanienbaum, Switzerland
[11]Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 110 Shaffer Rd., Santa Cruz, CA 95060, United States
[12]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, United States
[13]Eawag, Department Systems Analysis, Integrated Assessment and Modelling, Überlandstrasse 133, 8600, Dübendorf, Switzerland
[14]Slice Technologies, San Mateo
[15]University of Florida, Wildlife Ecology and Conservation, 110 Newins-Ziegler Hall, PO Box 110430, Gainesville, FL 32611-0430

Abstract

Successfully predicting the future states of systems that are complex, stochastic and potentially chaotic is a major challenge. Model forecasting error (FE) is the usual measure of success; however model predictions provide no insights into the potential for improvement. In short, the *realized* predictability of a specific model is uninformative about whether the system is inherently predictable or whether the chosen model is a poor match for the system and our observations thereof. Ideally, model proficiency would be judged with respect to the systems' *intrinsic* predictability – the highest achievable predictability given the degree to which system dynamics are the result of deterministic v. stochastic processes. Intrinsic predictability may be quantified with permutation entropy (PE), a model-free, information-theoretic measure of the complexity of a time series. By means of simulations we show that a correlation exists between estimated PE and FE and show how stochasticity, process error, and chaotic dynamics affect the relationship. This relationship is verified for a dataset of 461 empirical ecological time series. We show how deviations from the expected PE-FE relationship are related to covariates of data quality and the nonlinearity of ecological dynamics. These results demonstrate a theoretically-grounded basis for a model-free evaluation of a system's intrinsic predictability. Identifying the gap between the intrinsic and realized predictability of time series will enable researchers to understand whether forecasting proficiency is limited by the quality and quantity of their data or the ability of the chosen forecasting model to explain the data. Intrinsic predictability also provides a model-free baseline of forecasting proficiency against which modeling efforts can be evaluated.

Key words: time series analysis, Empirical Dynamic Modelling, permutation entropy, information theory, population dynamics, forecasting

Introduction

Understanding and predicting the dynamics of complex systems are central goals for many scientific disciplines (Weigend and Gershenfeld 1993, Hofman et al. 2017). Ecology is no exception as environmental changes across the globe have led to repeated calls to make the field a more predictive science (Clark et al. 2001, Petchey et al. 2015, Dietze 2017, Dietze et al. 2018). One particular focus is *anticipatory predictions*, forecasting probable future states in order to actively inform and guide decisions and policy (Mouquet et al. 2015, Maris et al. 2018). Robust anticipatory predictions require a quantitative framework to assess ecological forecasting and diagnose when and why ecological forecasts succeed or fail.

Forecast performance is measured by **realized predictability** (see glossary), often quantified as the correlation coefficient between observations and predictions, or its complement, **forecasting error (FE)** measures, such as root mean squared error (RMSE). Hence, realized predictability is in part determined by the model used as for any given system, different models will give different levels of realized predictability. Furthermore, it can be unclear, from realized predictability alone, whether the system is stochastic or the model is a poor choice.

By contrast, the **intrinsic predictability** of a system is an absolute measure that represents the highest achievable predictability (Lorenz 1995, Beckage et al. 2011). The intrinsic predictability of a system can be approximated with model-free measure of time series complexity, such as Lyapunov exponents or **permutation entropy** (Boffetta et al. 2002, Bandt and Pompe 2002, Garland et al. 2014). In principle, intrinsic predictability has the potential to indicate whether the model, data, or system are limiting realized predictability. Thus, if we know the intrinsic predictability of a system and its realized predictability under specific models, the difference between the two is indicative of how much predictability can be improved (Beckage et al. 2011).

Here we formalize a conceptual framework connecting intrinsic predictability and realized predictability. Our framework enables comparative investigations into the intrinsic predictability across systems and provides guidance on where and why forecasting is likely to succeed or fail. We use simulations of the logistic map to demonstrate the behaviour of PE in response to time series complexity and the effects of both process and measurement noise. We confirm a general relationship between PE and FE, using a large dataset of empirical time series and demonstrate how the quality, length, and **nonlinearity** in particular of these time series influences the gap between intrinsic and realized predictability and the consequences for forecasting.

Conceptual framework

The foundation for linking intrinsic and realized predictability lies in information theory and builds on research demonstrating a relationship between PE and FE for complex computer systems (Garland et al. 2014). Information theory was originally developed by Claude Shannon as a mathematical description of communication (Shannon 1948) but has since been applied across many disciplines. In ecology, several information-theoretic methods have proved useful, including the Shannon biodiversity metric in which the probability of **symbol** occurrences (see Box 1) is replaced by the probability of species occurrences (Jost 2006, Sherwin et al. 2017), and the Akaike Information Criterion (Akaike 1974) which is widely used for comparing the performance of alternative models (Burnham and Anderson 2002). Given its importance to our framework, we first provide an introduction to information theory with special attention to applications for ecological time series. Since our goal is to inform where, when and why forecasting succeeds or fails; we then i) describe how **information** may be partitioned into **new and redundant information** based on permutation entropy, ii) demonstrate how redundant information is exploited by different forecasting models, and iii) examine the relationship between permutation entropy and realized predictability and how it can inform forecasting.

An information-theoretic perspective

A first step towards predicting the future of any system is understanding *if* the observations of that system contain **information** about the future, i.e. does the system have a memory. The total information in each observation can be thought of as a combination of information that came from past states (i.e., **redundant information**) and information that is only available in the present state (i.e., **new information**).

When there is a substantial amount of information transmitted from the past to the present (figure 1Aiii), the system is said to be highly redundant. In other words, future states depend greatly on the present and past states. In these cases, very little new information is generated during each subsequent observation of the system and the resulting time series is, in theory, highly predictable (has high intrinsic predictability).

Conversely, in systems dominated by stochasticity, the system state at each time point is mostly independent of past states (figure 1Ai). Thus, all of the information will be "new" information, and there will be little to no redundancy with which to train a forecasting model. In this case, regardless of model choice, the system will not be predictable (has low intrinsic predictability).

Imperfect observations introduce uncertainty or bias into time series, and thereby affect the redundant information that is available or perceived. Observation errors in particular will reduce the redundant information available to forecasting models, thus lowering the realized predictability. We refer to this reduction as **lost information**, which is not an innate property of the system but is the result of the practical limitations of making measurements and any information-damaging processing of the data (figure 1B, Box 2). As such, lost information can be mitigated and is an important leverage point for ecologists to improve their forecasts. For example, replicate measurements or other forms of data integration that increase estimation accuracy and reduce bias will reduce information loss and can improve forecasts.

Permutation entropy

Permutation entropy (PE) is a measure of time series complexity that approximates the rate at which new information is being generated along a time series (Box 1). PE approximates and is inversely related to intrinsic predictability by quantifying how quickly the system generates new information. Time series with low permutation entropy have high redundancy and are expected to have high intrinsic predictability (Garland et al. 2014).

PE uses a symbolic analysis that translates a time series into a frequency distribution of **words** (see glossary for definition). The frequency distribution of words is then used to assess the predictability of the time series. For example, a time series in which a single word (i.e. a specific pattern) dominates the dynamics has high redundancy and thus future states are well predicted by past states. In contrast, a random time series, in which no single pattern dominates, would produce a nearly uniform frequency distribution of words, with future states occurring independently from past states. Hence, by quantifying the frequency distribution of words, PE approximates how much information is transmitted from the past to the present, corresponding to the intrinsic predictability of a time series.

When observations are imperfect PE measures the joint influence of new information (from either internal or external processes) and lost information (due to the observation process as well as data processing). We refer to the redundant information that is not lost and remains available as **active information**, which is the information that can be exploited by forecasting models.

Forecasting and redundant information

Realized predictability is highest when the chosen forecasting model exploits all the active information contained in a time series. For illustration, we forecast the oscillating abundance of a laboratory ciliate population (Veilleux 1976) with three different approaches (figure 1C): i) the mean of the time series (a model which uses relatively little of the active information), ii) a linear autoregressive integrated moving-average model (ARIMA) that uses the local-order structure of the time series in addition to the mean (a model which uses an intermediate amount of the active information), and iii) empirical dynamic modelling (EDM) that can incorporate **nonlinearities**, when present, in addition to the mean and local-order structure (a model which can feasibly use more active information). The time series was split into training data and test data. Forecasting models were fit to the training data and used to make forward predictions among the test data. The forecast performance of the models (i.e. the realized predictability) varied with the amount of information they used, which depended on structural differences among the models that exploit the active information coming from the past. EDM and ARIMA had similar performance suggesting that the time series entailed little nonlinearities for the EDM to exploit.

The relationship between realized and intrinsic predictability

With a perfect forecasting model, realized predictability - measured by forecasting error (FE) - and intrinsic predictability - measured by permutation entropy (PE) - will be positively related. More specifically, the relationship will pass through the origin and monotonically increase up to the maximum limit of PE = 1 (figure 1D, the boundary between the white and grey regions; Garland et al. 2014). In the top right of this figure are systems with high PE and therefore low redundancy and high forecasting error. In the bottom-left of the figure are systems with low PE and therefore high redundancy and low forecast error. The boundary is the limit for a perfect model that maximizes the use of active information.

Lost information complicates the interpretation of the PE - FE relationship by obscuring the system's actual intrinsic predictability. We illustrate this case in figure 1D using two hypothetical systems: one with high intrinsic predictability and a large amount of lost information, and one with lower intrinsic predictability but relatively little lost information. Despite the differences in the redundancy of two systems, the PE of their time series can be very similar (even identical) because PE does not differentiate between new and lost information.

For this example, both systems in figure 1D start with high FE relative to their PE. Selecting more appropriate forecasting models causes a reduction in FE but no change in PE. Reducing lost information (e.g. by increasing the frequency of measurements) decreases both PE and FE. The system with a high redundancy and a low **Shannon entropy rate** has a greater overall potential for improving forecasting skill through the recovery of lost information. In contrast, the system with low redundancy has limited scope to further improve forecasting skill; forecasting is less limited by lost information, but rather by its lower redundancy. As such, the lowest possible forecast error will be substantially higher in the second system than in the first system because the intrinsic predictability of the second is inherently lower and cannot be changed.

Materials & Methods

Forecasting with EDM

Empirical dynamic modelling is a set of nonlinear forecasting techniques brought to the attention of ecologists through the work of Sugihara (1994). The method is based on the idea that a system's attractor generating the dynamics of a time series can be reconstructed via delay coordinate embedding (Takens 1981, Sauer et al. 1991), which can then be used to forecast system dynamics (Lorenz 1969, Farmer and Sidorowich 1987, Sauer et al. 1991, Casdagli and Eubank 1992, Smith 1992, Weigend and Gershenfeld 1993, Garland and Bradley 2011, 2015, Garland et al. 2014). These methods are rooted in a deterministic dynamic system's paradigm and hence require at least some determinism in the temporal course of the system and hence are unsuitable for purely stochastic systems. However, they have proven to reliably forecast ecological systems even in the presence of process and measurement noise typical for ecological systems (Sugihara & May 1990, Ye et al. 2015) and are constantly improved to deal with issues such as observation error and nonstationarity of ecological systems (Munch et al. 2017). The variant of these methods we use in this manuscript is based on the simplex projection and S-map method (Sugihara 1994) through the rEDM package (https://ha0ye.github.io/rEDM/).

The EDM approach first identifies the optimal embedding dimension E of the training data by fitting a model using simplex projection (Sugihara 1994). The embedding dimension E determines the number of temporal lags used for the delay coordinate embedding. We tested values for E between 1 and 10 and selected the value of E with the highest forecast skill using leave-one-out cross validation (Sugihara 1994). We then fitted the tuning parameter $\boldsymbol{\theta}$ on the training data using the S-map model. $\boldsymbol{\theta}$ describes the nonlinearity of the system and was varied in 19 steps (0, 0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 0.5, 0.75, 1.0, 1.5, 2, 3, 4, 6, 8, and 10) to find the lowest error using leave-one-out cross validation on the training data.

In contrast to other forecasting methods such as ARIMA, the EDM approach searches across multiple time series models by finding the optimal in-sample combination of embedding dimension and tuning parameter using cross-validation. Due to this model selection step, EDM tests a suite of forecasting models equal to the number of combinations of $\boldsymbol{\theta}$ and E. When $\boldsymbol{\theta}$ is 0, the EDM model corresponds to an autoregressive model of the order of the embedding dimension (i.e. an AR3 model if $E = 3$). Values of $\boldsymbol{\theta}$ greater than 0 can account for increasing degrees of state-dependence.

Assessment of forecast error

We quantified forecasting error with the root mean squared error (Hyndman and Koehler 2006),

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{k}(c_i - p_i)^2}{k}},$$

where k is the number of observed c_i values (i.e. abundances) and p_i are their corresponding predicted values. To compare forecast errors across time series that vary widely in units and variability, we normalized their RMSE by the range of observed values using

$$\text{nRMSE} = \frac{RMSE}{max(c_i) - min(c_i)}.$$

Smaller nRMSE corresponds to smaller forecasting error.

Calculation of permutation entropy

We calculated the weighted permutation entropy (WPE) of time series using the methods outlined in Box 1.

Logistic Map Time Series

To demonstrate how both intrinsic and realized predictability change along a continuum from simple to complex and chaotic time series, we applied permutation entropy to time series from a well known population dynamic model, the Logistic Map:

$$x_{t+1} = r\, x_t\, (1 - x_t).$$

This model maps the current year's population size to next year's population size with simple density-dependence between non-overlapping generations. Although simple, this first-order, nonlinear function produces a wide range of dynamical behavior, from stable and oscillatory equilibria to chaotic dynamics (May and others 1976). We include this range of behavior by simulating the logistic map for 500 incremental growth rates between $r = 3.4$ and $r = 3.9$. We simulated each growth rate for 10,000 time steps keeping the last 3000 times steps for analysis. Weighted permutation entropy of time series was calculated for permutation order, *m,* from 3 to 5 and for time delay, τ, from 1 to 4. For simplicity, we will refer to weighted permutation entropy only in the results section and use the generic term permutation entropy everywhere else. Forecasting error for each time series was calculated using the normalized root mean squared error of an EDM forecast of the last 200 time steps.

Because ecological systems are influenced by both deterministic and stochastic drivers and the logistic map is purely deterministic, we sought to evaluate how stochasticity (noise) affects weighted permutation entropy and forecast error. To do so, we independently added both observational noise and process noise to the simulated population sizes by drawing random values from Gaussian distributions with standard deviations of either 0, 0.0001, 0.001, or 0.01 (Bandt and Pompe 2002). We also investigated the effect of non-Gaussian noise distributions on WPE and FE, although in this case we applied it to the Ricker model which does not have an upper bound of 1 like the logistic map (see appendix S1 for details). If the new population size was not between 0 and 1, a new value was drawn. Observational noise was added to the population size time series after the simulation, whereas process noise was added to population size at each time step during the simulation.

Empirical Time Series Data

For empirical evidence of a relationship between permutation entropy and forecasting error, we examined a large variety of ecological time series that differ widely in complexity and data quality. We further investigated whether deviations from the expected general relationship can be explained by time series covariates such as measurement error (proxied here by whether the data originated from field versus lab studies), the nonlinearity of the time series (as quantified by the theta parameter of an EDM), or time series length. This allowed us to identify possible predictors of time series complexity and the potential with which the time series of a system can be moved along the permutation-forecasting error (PE-FE) relationship to maximize realized predictability.

Time series databases and processing

We compiled laboratory time series from the literature and field time series from the publicly available Global Population Dynamic Database (GPDD) for our analysis. The GPDD is the largest compilation of univariate time series available, spanning a wide range of geographic locations, biotopes and taxa (NERC Centre for Population Biology, 1999, Inchausti & Halley 2001). The GDPP database was accessed via the rGDPP package in R (https://github.com/ropensci/rgpdd). We added laboratory time series from studies by Becks et al. (2005), Fussmann et al. (2000), and the datasets used in a meta-analysis by Hiltunen et al. (2014). Time series with less than 30 observations, gaps greater than 1 time step and more than 15% of values being equal (and hence having the same rank in the ordinal analysis, i.e. ties) were excluded, resulting in a total of 461 time series. Each time series was divided into training (initial ⅔ of the time series) and test data (the last ⅓ of the time series), with the EDM model performing best on the training set being used to estimate forecast error in the test set. We calculated the weighted permutation entropy (WPE) of each empirical time series using a permutation order, m, of 3 and a time delay, $\boldsymbol{\tau}$, of 1. Results were robust to the choice of $m \in [2, 5]$ and $\boldsymbol{\tau} \in [1, 4]$. The three different ways to deal with ties (i.e. “random”, “first”, “average”) did not qualitatively affect the results, with results being robust to variation in time series minimum length and tie percentage

Statistical analysis

All analyses were performed in the statistical computing environment R (R Development Core Team 2016). We used the lme4 package to fit mixed models to investigate the relationship between forecast error and permutation entropy (Bates et al. 2015), with forecasting error being the dependent variable. We included the data source (i.e. publication) as a random grouping variable to account for possible non-independence across time series from the same study. The independent variables were permutation entropy, the data type, the number of observations (N), the proportion of zeros in the time series (zero_prop), the proportion of ties in the time series (ties_prop), and, from the EDM analysis, the nonlinearity ($\boldsymbol{\theta}$) and the embedding dimension (E) of the time series. The data type, i.e. whether time series were measured in the lab or in the field, was included with our hypothesis being that lab measurements have lower observation error. Zero and tie proportions were included as they pose problems to the estimation of PE, as do short time series (see Box 2). Three of our predictor variables, namely PE, $\boldsymbol{\theta}$ and E are potentially measured with error violating an assumption of linear models (Quinn and Keough 2002). However, alternative approaches such as reduced major axis regression are only advised if the relationship between response and predictors is symmetric (Smith 2009). We therefore did not adjust for error, but note that the strength of the relationship of our predictors may be potentially underestimated due to measurement error in the predictors (Quinn and Keough 2002). Model diagnostics showed normally and homogeneously distributed residuals. Code to reproduce the analysis can be found on Github: XXX.

Results

Logistic Map Time Series

The expected relationship between weighted permutation entropy and forecasting error occurred in the simulations of the logistic map. Both WPE and FE generally increase as the growth rate, r, increases and the dynamics of the logistic map enter the realm of deterministic chaos (figure 4D). Correspondingly, both WPE and FE decline when chaotic dynamics converge to limit cycles (figure 4, gold example with $r \approx 3.84$).

The effect of stochastic noise on the WPE-FE relationship depended on the type of noise considered. While process noise strongly affects both WPE and FE (figure 5A) observational noise affects forecasting error more strongly than WPE (figure 5B). Indeed, the relationship between WPE and FE is largely obscured at high process noise but remains positive at high observational noise (figure 5A, B, top panels), particularly when dynamics are chaotic. When the dynamics are chaotic, the weighting in WPE is very effective at reducing the influence of observational noise on estimates of permutation entropy. However, when the dynamics exhibit stable limit cycles, WPE is sensitive to noise and this depends strongly on the chosen time delay, $\boldsymbol{\tau}$, and word length, m. This effect is a statistical artefact caused by tied ranks in the words that are then influenced by noise. For instance, applying $\boldsymbol{\tau}=2$ for a 2-point limit cycle with a small amount of noise produces a WPE close to one, appearing as white noise as all permutations occur with equal probability. Limit cycles are best analyzed with $\boldsymbol{\tau}=1$ to capture the oscillations, although with $m=3$ small amounts of noise still result in two permutations occurring with equal frequency (1-3-2 or 2-3-1) and so WPE is elevated with respect to the no-noise case despite the high redundancy of the limit cycles (figure 5B, dark blue and gold points; see appendix S2: Fig. S1 for details). The effect of stochasticity on the WPE-FE relationship is generally robust to the chosen model and noise distribution (see appendix S1: Fig. S1, S2 for the analysis of the Ricker model with multiplicative lognormal noise).

Empirical Time Series Results

The 461 empirical time series vary in length (median = 50, min = 30, max = 197) and, as measured by WPE, complexity (median = 0.84, min = 0.076, max = 1). Forecasting error (nRMSE) ranges from 0.0000093 to 1.37, with a median of 0.19. Our analysis shows the expected positive relationship between permutation entropy and forecast error, with more complex time series (high WPE) yielding higher forecasting error (Table 1, center panel of figure 6). No difference in mean forecast error nor a difference in slope is detected between time series originating from lab or field studies (Table 1). Exploring the effects of time series covariates indicates that longer time series had lower FE, whereas time series with larger dimensionality (E) and greater nonlinearity (θ) as measured by EDM show higher FE (Table 1). These covariates increase the amount of variation in FE explained across time series to 35% (CI: 29 - 42%). An analysis of the partial R^2 of all fixed effects in the model revealed that PE individually explained the largest amount of variation among predictors (21%, CI: 15 - 27%), followed by time series length (18%, CI: 12 - 24%), time series nonlinearity θ (6%, CI: 2 - 10%) and the chosen embedding dimension E (4%, 1 - 9%). Zero and tie proportions, as well as whether time series were from the field or the lab (type) explained less than 1% of the observed variation.

The PE v. FE relationship allows us to identify time series which were predicted better, equal to or worse than expected regarding their complexity (figure 6 a-f). Time series ‘b’ and ‘c’ fall along the expected relationship and hence are well predicted despite large differences in complexity. Time series ‘a’ shows a clear trend which is well predicted. In contrast, time series ‘d’-’f’ have higher than expected forecast error. Time series ‘d’ shows higher than expected error due to a strong outlier in the predicted values early in the test dataset. Time series ‘e’ is consistently poorly predicted, potentially due to wrong model choice or due to the short time series length. Time series ‘f’ is complex (high PE) with predictions missing the ongoing downward trend in the test data.

Discussion

The urgent need for ecologists to provide operational forecasts to managers and decision makers requires that we understand when and why forecasts succeed or fail (Clark et al. 2001, Petchey et al. 2015, Dietze 2017). We propose that the measurement of the intrinsic predictability of an ecological system can help reveal the origin of predictive uncertainty and indicate whether and how it can be reduced.

Our results show that realized and intrinsic predictability positively covary. The simulation study revealed that the relationship can be obscured by stochastic process noise, while measurement noise led to more scatter but preserved the positive slope (figure 5). Although process noise often dominates over measurement noise in ecological time series (Ahrestani et al. 2013), the positive relationship between intrinsic and realized predictability we revealed across a wide range of empirical time series supports the applicability of our framework. In our analysis, permutation entropy explained the largest amount of variation (21%) in forecast error, followed by time series length, dimensionality and nonlinearity, jointly accounting for 35% of the variation. Time series that fell onto the expected relationship (figure 6b,c) were well predicted given their complexity, whereas clear outliers (e.g. figure 6e) would not require the use of PE to be identified as such. The relationship however allowed us to identify potential problems with forecasts of time series that have reasonable forecasts error, but which may be affected by overfitting (figure 6a), outliers (figure 6d) or regime shifts (figure 6e) that may have gone unnoticed when looking at FE alone, particularly if applying automated or semi-automated forecasting methods across hundreds or thousands of time series (White et al. 2018).

The value of intrinsic predictability to guide forecasting

A major advantage of permutation entropy is the independence from any assumed underlying model of the system, which makes this "model-free" method highly complementary to existing model-based approaches. For instance, Dietze (2017) recently proposed a model-based framework that partitions the contribution of various factors to predictive uncertainty, including the influence of initial conditions, internal dynamics, external forcing, parameter uncertainty and process error at different scales. If, for example, the dominant factor affecting near-term forecasts is deemed to be internal dynamics, then insight into intrinsic predictability would demonstrate how stable those internal dynamics are. Similarly, if a lot of variation remains unexplained by the model (i.e. the process error not explained by the known internal dynamics, initial conditions, external drivers, and estimated parameters), then "model-free" methods can provide insight into whether that variation is largely stochastic or contains unexploited structure that could be captured with further research into the driving deterministic processes. Finally, permutation entropy could be applied to the predicted dynamics of models to ascertain whether they accurately reflect properties of the observed dynamics, such as their complexity, similar to comparing the nonlinearity of a time series with the dynamics of the best model using the EDM framework (Storch et al. 2017). Thus, intrinsic predictability provides diagnostic insights into predictive uncertainty and guidance for improving predictions.

Comparative assessments of intrinsic predictability

The model-free nature of permutation entropy is advantageous in cross-system and cross-scale comparative studies of predictability. Whereas comparing all available forecasting methods on a given time series and predicting with the best-performing method would give us the best realized predictability (e.g., Ward et al. 2014), we would miss out on the comparative insight gained from aligning very different time series along the complexity gradient quantified by permutation entropy. Such a comparison could afford insight into whether intrinsic predictability differs across levels of ecological organization, taxonomic groups, habitats, geographic regions or anthropogenic impacts (Petchey et al. 2015). Determining the most appropriate covariates of monitored species (e.g. body size, life history traits, and trophic position) that minimize lost information would also inform monitoring methods. Furthermore, monitoring how realized and intrinsic predictability converge over time provides a means to judge improvements in predictive proficiency (Petchey et al. 2015, Houlahan 2016, Dietze 2017). To do so, we need to apply available forecasting models to the same time series and measure their forecast error in combination with their intrinsic predictability. The monitoring of predictive proficiency has greatly advanced weather forecasting as a predictive science (Bauer et al. 2015). The analysis of univariate time series presented here only begins to put the intrinsic predictability of different systems into perspective. A primary goal is hence to expand the availability of long-term, highly resolved time series to determine potential covariates and improve our general understanding of ecological predictability (Ward et al. 2014, Petchey et al. 2015).

Reliable assessment of intrinsic predictability

Permutation entropy requires time series data of suitable length and sampling frequency to infer the correct permutation order and time delay (Riedl et al. 2013). Given the complexity of many ecological time series, the method rarely works with less than 30 data points (see recommendations in box 1). We acknowledge these as fairly stringent requirements for ecological time series. Time series measured at the appropriate time scales over long periods of time are rare, despite the knowledge that they are among the most effective approaches at resolving long-standing questions regarding environmental drivers (Lindenmayer et al. 2012, Hughes et al. 2017, Giron-Nava et al. 2017). This problem is beginning to be resolved with automated measurements of system states, such as chlorophyll-a concentrations in aquatic systems (Blauw et al. 2018, Thomas et al. 2018), assessment of community dynamics in microbiology (Trosvik et al. 2008, Faust et al. 2015, Martin-Platero et al. 2018), and phenological (Pau Stephanie et al. 2011) and flux measurements (Dietze 2017). Such high-frequency, long-term data are likely to provide a more accurate picture of the range of possible system states, even when systems are non-ergodic and change through time (e.g. figure 6f). In fact, given the ease with which it is computed, PE can be assessed with a moving window across time or space to determine if a system is stationary or changing. As such, PE may be used as an early warning signal for system tipping points and critical transitions (Scheffer et al. 2009, Dakos and Soler-Toscano 2017) or to evaluate the effect of a management intervention on the system state.

Currently, there is no generally accepted approach to calculate uncertainty in PE values and compare whether two PE values are statistically different. Approaches such as comparing empirical estimates of PE to white-noise time series or parametric bootstrapping have been suggested (Little and Kane 2016, Traversaro and O. Redelico 2018), however, these approaches are not free from challenges and may provide an overconfident picture of uncertainty. One suggestion is for the practitioner to rely on persistence over parameter space. That is, slightly modify the parameters of the calculation (change m and $\boldsymbol{\tau}$) and see if the results change. If the results do not change, this should suggest a higher degree of reliability. Nevertheless, this limitation does not diminish the usefulness of PE for regression-based applications such as those presented and we are confident that increased usage of PE will result in methodological advances such as uncertainty estimation.

Although the full potential of permutation entropy to guide forecasting is not yet realized, many other fields are starting to take advantage of its diagnostic potential. In paleoclimate science, permutation entropy has proven useful for detecting hidden data problems caused by outdated laboratory equipment (Garland et al. 2016, 2018), and in the environmental sciences it has provided insight into model-data deviations of gross primary productivity to further understand the global carbon cycle (Sippel et al. 2016). In epidemiology a recent study on the information-theoretic limits to forecasting of infectious diseases concluded that for most diseases the forecast horizon is often well beyond the time scale of outbreaks, implying prediction is likely to succeed (Scarpino and Petri 2017). Our result showing that permutation entropy covaries with forecast error highlights the potential of using permutation entropy to better understand time series predictability in ecology and other disciplines.

Acknowledgements

This paper originates from the “sPRED - Synthesizing Predictability Research of Ecological Dynamics” working group, supported by the Synthesis Centre of the German Centre for Integrative Biodiversity Research (DFG-FZT-118). FP, AT and OP benefitted from funding by the Swiss National Science Foundation (grant 31003A_159498 to OP). AI was supported by the Alexander von Humboldt Foundation. JG was supported by an Omidyar Fellowship from the Santa Fe Institute. HY is supported by the Gordon and Betty Moore Foundation’s Data-Driven Discovery Initiative through grant GBMF4563 to Ethan P. White. BR, BCR and UB acknowledge support by the German Research Foundation (FZT 118). We thank Gregor Fussmann and Lutz Becks for generously sharing time series data from microcosm experiments.

References

Ahrestani, F. S., M. Hebblewhite, and E. Post. 2013. The importance of observation versus process error in analyses of global ungulate populations. Scientific Reports 3:3125.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.

Amigó, J. M., M. B. Kennel, and L. Kocarev. 2005. The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. Physica D: Nonlinear Phenomena 210:77–95.

Bandt, C. 2005. Ordinal time series analysis. Ecological Modelling 182:229–238.

Bandt, C., and B. Pompe. 2002. Permutation Entropy: A Natural Complexity Measure for Time Series. Physical Review Letters 88:174102.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:1–48.

Bauer, P., A. Thorpe, and G. Brunet. 2015. The quiet revolution of numerical weather prediction. Nature 525:47–55.

Beckage, B., L. J. Gross, and S. Kauffman. 2011. The limits to prediction in ecological systems. Ecosphere 2:art125.

Becks, L., F. M. Hilker, H. Malchow, K. Jürgens, and H. Arndt. 2005. Experimental demonstration of chaos in a microbial food web. Nature 435:1226–1229.

Blauw, A. N., E. Benincà, R. W. P. M. Laane, N. Greenwood, and J. Huisman. 2018. Predictability and environmental drivers of chlorophyll fluctuations vary across different time scales and regions of the North Sea. Progress in Oceanography 161:1–18.

Boffetta, G., M. Cencini, M. Falcioni, and A. Vulpiani. 2002. Predictability: a way to characterize complexity. Physics Reports 356:367–474.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer, New York.

Casdagli, M., and S. Eubank, editors. 1992. Nonlinear Modeling and Forecasting. 1 edition. CRC Press, Redwood City, Calif.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. P. Jr, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological Forecasts: An Emerging Imperative. Science 293:657–660.

Dakos, V., and F. Soler-Toscano. 2017. Measuring complexity to infer changes in the dynamics of ecological systems under stress. Ecological Complexity 32:144–155.

Dietze, M. C. 2017. Prediction in ecology: a first-principles framework. Ecological Applications 27:2048–2060.

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers, and E. P. White. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences:201710231.

Fadlallah, B., B. Chen, A. Keil, and J. Príncipe. 2013. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. Physical Review E 87:022911.

Farmer, J. D., and J. J. Sidorowich. 1987. Predicting chaotic time series. Physical Review Letters 59:845–848.

Faust, K., L. Lahti, D. Gonze, W. M. de Vos, and J. Raes. 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology 25:56–66.

Fussmann, G. F., S. P. Ellner, K. W. Shertzer, and N. G. Hairston. 2000. Crossing the hopf bifurcation in a live predator-prey system. Science (New York, N.Y.) 290:1358–1360.

Garland, J., and E. Bradley. 2011. Predicting Computer Performance Dynamics. Pages 173–184 *in* J. Gama, E. Bradley, and J. Hollmén, editors. Advances in Intelligent Data Analysis X. Springer Berlin Heidelberg.

Garland, J., and E. Bradley. 2015. Prediction in projection. Chaos: An Interdisciplinary Journal of Nonlinear Science 25:123108.

Garland, J., R. James, and E. Bradley. 2014. Model-free quantification of time-series predictability. Physical Review E 90:052910.

Garland, J., T. R. Jones, E. Bradley, R. G. James, and J. W. C. White. 2016. A First Step Toward Quantifying the Climate's Information Production over the Last 68,000 Years. Pages 343–355 Advances in Intelligent Data Analysis XV. Springer, Cham.

Garland, J., T. R. Jones, E. Bradley, M. Neuder, and J. W. C. White. 2018. Climate entropy production recorded in a deep Antarctic ice core. arXiv:1806.10936 [physics].

Giron-Nava, A., C. James, A. Johnson, D. Dannecker, B. Kolody, A. Lee, M. Nagarkar, G. Pao, H. Ye, D. Johns, and G. Sugihara. 2017. Quantitative argument for long-term ecological monitoring. Marine Ecology Progress Series 572:269–274.

Hiltunen, T., N. G. Hairston, G. Hooker, L. E. Jones, and S. P. Ellner. 2014. A newly discovered role of evolution in previously published consumer–resource dynamics. Ecology letters 17:915–923.

Hofman, J. M., A. Sharma, and D. J. Watts. 2017. Prediction and explanation in social systems. Science 355:486–488.

Houlahan, J. E. 2016. The priority of prediction in ecological understanding. Oikos.

Hughes, B. B., R. Beas-Luna, A. K. Barner, K. Brewitt, D. R. Brumbaugh, E. B. Cerny-Chipman, S. L. Close, K. E. Coblentz, D. Nesnera, K. L, S. T. Drobnitch, J. D. Figurski, B. Focht, M. Friedman, J. Freiwald, K. K. Heady, W. N. Heady, A. Hettinger, A. Johnson, K. A. Karr, B. Mahoney, M. M. Moritsch, A.-M. K. Osterback, J. Reimer, J. Robinson, T. Rohrer, J. M. Rose, M. Sabal, L. M. Segui, C. Shen, J. Sullivan, R. Zuercher, P. T. Raimondi, B. A. Menge, K. Grorud-Colvert, M. Novak, and M. H.

Carr. 2017. Long-Term Studies Contribute Disproportionately to Ecology and Policy. BioScience 67:271–281.

Hyndman, R. J., and A. B. Koehler. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22:679–688.

Jost, L. 2006. Entropy and diversity. Oikos 113:363–375.

Lindenmayer, D. B., G. E. Likens, A. Andersen, D. Bowman, C. M. Bull, E. Burns, C. R. Dickman, A. A. Hoffmann, D. A. Keith, M. J. Liddell, A. J. Lowe, D. J. Metcalfe, S. R. Phinn, J. Russell-Smith, N. Thurgate, and G. M. Wardle. 2012. Value of long-term ecological studies. Austral Ecology 37:745–757.

Little, D. J., and D. M. Kane. 2016. Permutation entropy of finite-length white-noise time series. Physical Review E 94:022118.

Lorenz, E. N. 1969. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. Journal of the Atmospheric Sciences 26:636–646.

Lorenz, E. N. 1995. Predictability: a problem partly solved.

Maris, V., P. Huneman, A. Coreau, S. Kéfi, R. Pradel, and V. Devictor. 2018. Prediction in ecology: promises, obstacles and clarifications. Oikos 127:171–183.

Martin-Platero, A. M., B. Cleary, K. Kauffman, S. P. Preheim, D. J. McGillicuddy, E. J. Alm, and M. F. Polz. 2018. High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nature Communications 9:266.

May, R. M., and others. 1976. Simple mathematical models with very complicated dynamics. Nature 261:459–467.

Mouquet, N., Y. Lagadeuc, V. Devictor, L. Doyen, A. Duputié, D. Eveillard, D. Faure, E. Garnier, O. Gimenez, P. Huneman, F. Jabot, P. Jarne, D. Joly, R. Julliard, S. Kéfi, G. J. Kergoat, S. Lavorel, L. Le Gall, L. Meslin, S. Morand, X. Morin, H. Morlon, G. Pinay, R. Pradel, F. M. Schurr, W. Thuiller, and M. Loreau. 2015. Predictive ecology in a changing world. Journal of Applied Ecology 52:1293–1310.

Munch, S. B., V. Poynor, and J. L. Arriaza. 2017. Circumventing structural uncertainty: A Bayesian perspective on nonlinear forecasting for ecology. Ecological Complexity 32:134–143.

Pau Stephanie, Wolkovich Elizabeth M., Cook Benjamin I., Davies T. Jonathan, Kraft Nathan J. B., Bolmgren Kjell, Betancourt Julio L., and Cleland Elsa E. 2011. Predicting phenology by integrating ecology, evolution and climate science. Global Change Biology 17:3633–3643.

Petchey, O. L., M. Pontarp, T. M. Massie, S. Kéfi, A. Ozgul, M. Weilenmann, G. M. Palamara, F. Altermatt, B. Matthews, J. M. Levine, D. Z. Childs, B. J. McGill, M. E. Schaepman, B. Schmid, P. Spaak, A. P. Beckerman, F. Pennekamp, and I. S. Pearse. 2015. The ecological forecast horizon, and examples of its uses and determinants. Ecology Letters 18:597–611.

Quinn, G. G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Riedl, M., A. Müller, and N. Wessel. 2013. Practical considerations of permutation entropy. The European Physical Journal Special Topics 222:249–262.

Sauer, T., J. A. Yorke, and M. Casdagli. 1991. Embedology. Journal of Statistical Physics 65:579–616.

Scarpino, S. V., and G. Petri. 2017. On the predictability of infectious disease outbreaks. arXiv:1703.07317 [physics, q-bio].

Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara. 2009. Early-warning signals for critical transitions. Nature 461:53–59.

Shannon, C. E. 1948. A mathematical theory of communication. The Bell System Technical Journal 27:379–423.

Sherwin, W. B., A. Chao, L. Jost, and P. E. Smouse. 2017. Information Theory Broadens the Spectrum of Molecular Ecology and Evolution. Trends in Ecology & Evolution 0.

Sippel, S., H. Lange, M. D. Mahecha, M. Hauhs, P. Bodesheim, T. Kaminski, F. Gans, and O. A. Rosso. 2016. Diagnosing the Dynamics of Observed and Simulated Ecosystem Gross Primary Productivity with Time Causal Information Theory Quantifiers. PLOS ONE 11:e0164960.

Smith, L. A. 1992. Identification and prediction of low dimensional dynamics. Physica D: Nonlinear Phenomena 58:50–76.

Smith, R. J. 2009. Use and misuse of the reduced major axis for line-fitting. American Journal of Physical Anthropology 140:476–486.

Storch, L. S., S. M. Glaser, H. Ye, and A. A. Rosenberg. 2017. Stock assessment and end-to-end ecosystem models alter dynamics of fisheries data. PLOS ONE 12:e0171644.

Sugihara, G. 1994. Nonlinear Forecasting for the Classification of Natural Time Series. Philosophical Transactions: Physical Sciences and Engineering 348:477–495.

Takens, F. 1981. Detecting strange attractors in turbulence. Pages 366–381 Dynamical Systems and Turbulence, Warwick 1980. Springer, Berlin, Heidelberg.

Thomas, M. K., S. Fontana, M. Reyes, M. Kehoe, F. Pomati, and T. Coulson. 2018. The predictability of a lake phytoplankton community, over time-scales of hours to years. Ecology Letters 21:619–628.

Traversaro, F., and F. O. Redelico. 2018. Confidence intervals and hypothesis testing for the Permutation Entropy with an application to epilepsy. Communications in Nonlinear Science and Numerical Simulation 57:388–401.

Trosvik, P., K. Rudi, T. Næs, A. Kohler, K.-S. Chan, K. S. Jakobsen, and N. C. Stenseth. 2008. Characterizing mixed microbial population dynamics using time-series analysis. The ISME Journal 2:707–715.

Veilleux, B. 1976. The analysis of a predatory interaction between Didinium and Paramecium. Master’s Thesis, University of Alberta, Canada.

Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos 123:652–661.

Weigend, A. S., and N. A. Gershenfeld. 1993. Time Series Prediction: Forecasting The Future And Understanding The Past. 1 edition. Routledge, Reading, MA.

White, E. P., G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, and S. M. Ernest. 2018. Developing an automated iterative near-term forecasting system for an ecological study. bioRxiv:268623.

Glossary

Active information: The amount of information that is available to forecasting models (redundant information minus lost information; figure 1).

Forecasting error (FE): A measure of the discrepancy between a model's forecasts and the observed dynamics of a system. Common measures of forecast error are root mean squared error and mean absolute error.

Entropy: Measures the average amount of information in the outcome of a stochastic process.

Information: Any entity that provides answers and resolves uncertainty about a process. When information is calculated using logarithms to the base two (i.e. information in bits), it is the minimum number of yes/no questions required, on average, to determine the identity of the symbol (Jost 2006). The information in an observation consists of information inherited from the past (redundant information), and of new information.

Intrinsic predictability: the maximum achievable predictability of a system (Beckage et al. 2011).

Lost information: The part of the redundant information lost due to measurement or sampling error, or transformations of the data (figure 1).

New information, Shannon entropy rate: The Shannon entropy rate quantifies the average amount of information per observation in a time series that is unrelated to the past, i.e., the new information (figure 1).

Nonlinearity: When the deterministic processes governing system dynamics depend on the state of the system.

Permutation entropy (PE): permutation entropy is a measure of the complexity of a time series (Bandt and Pompe 2002) that is negatively correlated with a system's predictability

(Garland et al. 2014). Permutation entropy quantifies the combined new and lost information. PE is scaled to range between a minimum of 0 and a maximum of 1.

Realized predictability: the achieved predictability of a system from a given forecasting model.

Redundant information: The information inherited from the past, and thus the maximum amount of information available for use in forecasting (figure 1).

Symbols, words, permutations: symbols are simply the smallest unit in a formal language such as the letters in the English alphabet i.e., {“A”, “B”,..., “Z”}. In information theory the alphabet is more abstract, such as elements in the set {“up”, “down”} or {“1”, “2”, “3”}. Words, of length *m* refer to concatenations of the symbols (e.g., up-down-down) in a set. Permutations are the possible orderings of symbols in a set. In this manuscript, the words are the permutations that arise from the numerical ordering of *m* data points in a time series.

Weighted permutation entropy (WPE): a modification of permutation entropy (Fadlallah et al. 2013) that distinguishes between small-scale, noise-driven variation and large-scale, system-driven variation by considering the magnitudes of changes in addition to the rank-order patterns of PE.

Boxes

Box 1 Theory and estimation of PE and WPE

Information theory provides several measures for approximating how much new information is expected per observation of a system (e.g. the Shannon-entropy rate and the Kolmogorov-Sinai entropy). However, these measures are only well defined for infinite sequences of discrete random variables and can be quite challenging to approximate for continuous random variables, especially if one only has a finite set of imperfect observations. Permutation entropy is a measure aimed at robustly approximating the Shannon-entropy rate of a times series (or the Kolmogorov-Sinai entropy if the time series is stationary).

Rather than estimating probability mass functions from **symbol** frequencies or frequencies of sequences of symbols, as is done with traditional estimates of the Shannon-entropy rate, permutation entropy uses the frequencies of orderings of sequences of values; it is an ordinal analysis (see figure 2 for a visual explanation). The ordinal analysis of a time series maps the successive time-ordered elements of a time series to their value-ordered permutation of the same size. As an example, if $[x_1, x_2, x_3] = [11,6,8]$then its ordinal pattern, or word, $\phi([x_1, x_2, x_3])$, is 2-3-1 since $x_2 \leq x_3 \leq x_1$(see red time series fragment in figure 2A). PE is calculated by counting the frequencies of these words (or **permutations**) that arise after the time series undergoes this ordinal analysis. That is, given a time series (figure 2A), let $S_m$be defined as the set of all permutations (possible words) π of order (word length) m and time delay τ, describing the delay between successive points in the time series (figure 2B for $m = 3$ and $\tau = 1$). For each permutation $\pi \in S_m$we estimate its relative frequency of occurrence for the observed time series after performing ordinal analysis on each delay vector, $p(\pi) = \{i \vee i \leq N - m, \phi$, where $|\cdot|$ denotes set cardinality (figure 2C). Then permutation entropy of order $m \geq 2$ is calculated as $h(m) = -\sum_{\pi \in S_m} p(\pi) log_2\big(p(\pi)\big)$.

Since, $0 \leq h(m) \leq log_2(m!)$, it is common to normalize permutation entropy by dividing by $log_2(m!)$. With this convention, maximal $h(m) = 1$ and minimal $h(m)$ is equal to 0. Since in the infinite word length limit, permutation entropy is equivalent to the Kolmogorov-Sinai entropy as long as the observational uncertainty is sufficiently small (Amigó et al. 2005), we can approximate the intrinsic predictability of an ecological time series by computing $1 - h(m)$.

For the ordinal analysis of a time series, ranks are only well defined if all values are different. If some values are equal (so called ‘ties’), the ordinal analysis is not possible. Several approaches are available to break the ties: the "first" method results in a permutation with increasing values at each index set of ties, and analogously "last" with decreasing values. The "random" method puts these in random order whereas the "average" method replaces them by their mean, and "max" and "min" replaces them by their maximum and minimum respectively, the latter being the typical sports ranking.

In contrast, an ordinal analyses is also affected by small scale fluctuations due to measurement noise which can obscure the influence of large scale system dynamics. **Weighted permutation entropy (WPE)** reduces the influence of small-scale fluctuations by taking into account the relative magnitudes of the time series values within each word (Fadlallah et al. 2013). That is, each word’s ($X_t = \left[x\middle| \quad \middle| t, x_{t-\tau}, \dots, x_{t-\tau(m-1)}\right]$ contribution to the probability mass function is weighted by its variance, *viz.*, $w(X_t) \equiv var(X_t)$. Using this weighting function, the weighted probability of each permutation is estimated by:$p_w(\pi) = \sum_{t \leq N-m} w(X_t) \cdot \delta$ where $\delta(x, y) = 1$if and only if $x = y$ and $\delta(x, y) = 0$otherwise. The weighted permutation entropy of order $m \geq 2$ is then defined as $h_w(m) = -\sum_{\pi \in S_m} p_w(\pi) log_2\left(p_w(\pi)\right)$. Similar to PE, the weighted permutation entropy is normalized by $log_2(m!)$. We use weighted permutation entropy for all analyses presented in this manuscript.

The estimation of PE to time series requires specifying a order m and time delay τ. The shorter the chosen word length, the fewer possible words there are and the better we can estimate permutation frequencies. However, the ability to distinguish patterns is limited by the possible number of unique permutations. Hence, when word lengths are too short or too long, the frequency distribution is more uniform. In practice the total length of the time series limits the choice of possible word lengths and hence the number of unique words that can be resolved (Riedl et al. 2013). Regarding the time delay τ, most applications to study the complexity of a time series use a $\tau = 1$ (Riedl et al. 2013). If $\tau > 1$, Bandt (2005) notes the interesting property of the permutation entropy to be small, if the series has main period p for $\tau = p / 2$ and $3\,p / 2$, and to be large for $\tau = p$ and $\tau = 2\,p$. We refer to Riedl et al. (2013) who provide practical considerations regarding setting permutation order m and time delay τ.

Box 2 with information on the limitations of PE / WPE

When analyzing time series, ecologists typically employ a number of data pre-processing methods. These methods are used to reduce low-frequency trends or periodic signals (detrending), reduce high-frequency variation (smoothing), standardize across the time series or reduce the influence of extreme values (transformation), deal with uncertain or missing data points (gap or sequence removal, and interpolation), to examine specific time step sizes (downsampling), or to combine different time series (aggregation). Table 2 summarizes the anticipated effects on permutation entropy of a suite of commonly used pre-processing methods. In many cases, whether a method increases or decreases permutation entropy will depend on the specific attributes of the time series (e.g., its embedding dimension, autocorrelation, covariance structure, etc.) and the permutation order (m) at which its permutation entropy is approximated. This is illustrated by specific examples in figure 3 which contrasts the permutation entropies (using $m = 3$) of three hypothetical time series before (top row) and after (bottom row) the application of (a-b) linear detrending, (c-d) log-transformation, (e-f) interpolation of a missing or removed data point with a cubic smoothing spline. As these examples illustrate, with the exception of affine transformations, every pre-processing method discussed has the ability to alter our estimation of how much predictive information is contained in a time series. As such, performing pre-processing of a time series before permutation entropy is determined is not recommended. If the question to be addressed depends on such pre-processing, then care must be taken to understand how preprocessing is affecting the information estimate.

Tables

Table 1: Model table presenting fixed effects of the mixed model analysis relating forecasting error to permutation entropy (PE), and additional time series covariates. Parameter estimates (*B*), 95% confidence intervals (CI) and p-values are provided. Forecasting error increases with weighted permutation entropy across 461 ecological time series.

	Forecasting error (nRMSE)		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	0.0893	0.0106 – 0.1681	.027
PE	0.4796	0.3944 – 0.5648	<.001
Type (lab)	-0.0751	-0.2988 – 0.1486	.511
Sample size (N)	-0.0017	-0.0021 – -0.0013	<.001
Zero prop.	0.4062	0.0719 – 0.7405	.018
Ties prop.	-0.3344	-0.7698 – 0.1009	.133
Embedding dimension (E)	0.0088	0.0051 – 0.0124	<.001
Nonlinearity (theta)	0.0113	0.0072 – 0.0154	<.001
PE:type (lab)	0.1006	-0.1714 – 0.3726	.469

Table 2. Summary of the anticipated effects on permutation entropy of a suite of commonly used pre-processing methods.

Data processing method	**Examples**	**Effect on**		**Remark**
		Permutation entropy (PE)	**Weighted permutation entropy (WPE)**	
Detrending	Linear, nonlinear (e.g., GAM), differencing	Increase or decrease	Increase or decrease	Effect will depend on attributes of the time series for any chosen permutation order > 2.
Transformation	$(x - \overline{x})/\sigma$, log(x), $\sqrt[4]{(x)}$, Fisher, etc.	None	Increase, decrease, or none	Normalization or rescaling will have no effect as long as the transformation is linear. Nonlinear transformations that compress large values (e.g., log(*x*)) will increase WPE. Nonlinear transformations that amplify large values (e.g., Fisher) will decrease WPE.
Gap or	Missing data	Increase or	Increase or	Zeros should be retained if they represent true species absences

sequence removal	(NAs), below detection level (zeros), species absences (zeros), constant values (poor precision)	decrease	decrease	(decreasing PE and WPE). Otherwise zeros and constant values can be removed (increasing or decreasing PE and WPE, ***see main text***) or replaced by uncorrelated noise (increasing PE and WPE). The effect of concatenation will depend on attributes of the time series and gap size. Better to not count words that bridge gaps.
Interpolation	To infer gaps or to make time series equidistant	Increase or decrease	Increase or decrease	More likely to decrease than increase. Increases may occur for some nonlinear methods depending on attributes of the time series and the chosen permutation length. Better to ignore time-step uncertainty, assume equidistance, and not count words that bridge gaps.
Smoothing	Time-averaging, time-summation	Decrease	Decrease	Like linear interpolation decreases PE and WPE by increasing the count of only-ascending or only-descending permutations.

Downsampling	Regular subsetting to increase time-step size	Increase or decrease	Increase or decrease	Effect will depend on attributes of the time series (particularly its intrinsic embedding dimension) and the chosen permutation length.
Time series aggregation	Combining species to functional group	Increase or decrease	Increase or decrease	Effect will depend on attributes of the time series being aggregated (e.g., their relative magnitudes, covariance, etc.).

Figure legends

Figure 1.

A) The total information content of an observation of a system at a given state in time, S_t, is depicted by filled circles with past states (S_{t-1} and S_{t-2}) represented by shades of grey. i) lack of overlap between past and present states illustrating a case where no information is transmitted from past states (i.e. a purely stochastic system), with low redundancy and high Shannon entropy rate, ii) intermediate overlap indicating a case when some information is transferred from past to present (i.e. a deterministic system strongly driven by stochastic forcing), with intermediate redundancy and Shannon entropy rate, iii) large overlap indicating a case when the current state is mostly determined by the previous state (i.e. a highly deterministic system), with high redundancy and low Shannon entropy rate. Note that both the redundancy and Shannon entropy rate of a system are intrinsic properties of the system and will only change if the system itself changes.

B) The total information of an observation (black circle) is composed of new information and redundant information; redundant information is composed of active and lost information. A system's redundancy determines its intrinsic predictability. Information may be lost due to observation error and data processing (*lost information*). This reduces the redundant information that can be used for forecasting (*active information*). Lost information is not an intrinsic property of the system but rather represents practical limitations on our ability to make accurate measurements. The rate at which new information is being generated (Shannon entropy rate) may be approximated with permutation entropy. Because permutation entropy quantifies the joint contribution of the Shannon entropy rate and the lost information, efforts that minimize the amount of lost information not only maximize the redundant information that can actively be used for forecasting but also improve the estimation of the intrinsic Shannon entropy rate.

C) The realized predictability is the degree to which forecast models can exploit the active information of a time series. Consider, for example, a time-series on the abundance of a species (black line) of which the first 21 days are used to train (parameterize) three forecasting models: a forecast that uses the simple mean of the training data set (red), an Autoregressive integrated moving average (ARIMA) model (green), and an Empirical Dynamical Model (EDM, blue). The forecasting performance of these models is assessed using the remaining time series (after day 22). The inset shows the normalized root mean squared error (nRMSE) as a measure of deviation between predicted and observed values (i.e. forecast error) for each of the three forecasting models. ARIMA and EDM exploit the available structure in the data better than the mean forecast, as illustrated by the coloured wedges filling different amounts of the area of active information.

D) In the relationship between PE and FE a system can be moved toward the ideal grey boundary with forecast models that make better use of active information or by reducing information loss, not necessarily in that order. The two panels depict how to reach the greatest achievable forecasting skill in two different systems that have the same initial permutation entropy but differ in their relative amounts of new and redundant information (i.e. they differ in their intrinsic predictability). As these intrinsic properties of the system cannot be changed, improvements to forecast skill rely on fully exploiting the active information available (e.g., improved forecasting model) and minimizing information loss (e.g., improved sampling) to better approximate the true Shannon entropy rate, which establishes the lower boundary (grey area).

Figure 2. We illustrate how to estimate permutation entropy from an empirical time series (A) assuming m = 3 and $\tau = 1$. A permutation order m = 3 allows for a set of 6 (i.e. 3!) permutations, shown in panel B. The occurrence of each permutation π is then counted and

divided by the total number of permutations as an estimate of their proportional frequency (panel C). For example, permutations 2-3-1 (shown in red) and 3-2-1 are each only found once in the time series, whereas 1-2-3 and 3-1-2 are found twice, leading to frequencies of 0.17, 0.17, 0.33 and 0.33, respectively. The permutation entropy is then calculated as the Shannon entropy of proportional frequencies. For the given time series this is 1.92, which is normalized by $\log_2(3!)$ yielding a permutation entropy of 0.74.

Figure 3. Anticipated effects of a suite of commonly used pre-processing methods on (non-weighted) permutation entropy (PE) using three hypothetical time series before (top row) and after (bottom row) the application of (a-b) linear detrending, (c-d) log-transformation, and (e-f) interpolation of a missing or removed data point with a cubic smoothing spline.

Figure 4. Simulations of the deterministic logistic map with no added process or observation noise. A) The last 30 time steps of three times series are plotted to demonstrate different behaviors, including 2-point limit cycles ($r \approx 3.41$; dark blue), chaotic behavior ($r \approx 3.73$; green), and 3-point limit cycles within the chaotic realm ($r \approx 3.84$; gold). B) A bifurcation diagram of the logistic map attractor for growth rates between $r = 3.4$ and 3.9. C) Weighted permutation entropy (WPE) of the logistic map time series as the growth rate, r, changes for permutation order, m, of 3 (light grey), 4 (dark grey) and 5 (black), and time delay, $\boldsymbol{\tau}$ of 1. D) forecast error quantified by the normalized root mean squared error (nRMSE) of an EDM forecast ($E = 2$, $\boldsymbol{\tau} = 1$) of the last 200 time steps of each simulation plotted against WPE (m=5, $\boldsymbol{\tau}$=1). The color coding corresponds to the growth rates in ‘B’.

Figure 5. The relationship between weighted permutation error (WPE; m=5, τ=1) and forecasting error (measured as nRMSE) at three levels of A) process noise and B) observational noise. As the y-axis range and scale changes between subplots, the 'no noise' case is plotted in grey as a visual reference. The color coding corresponds to the growth rates in figure 4B. Systems with higher process noise exhibit both higher WPE and higher forecasting error. WPE is robust to observational noise when dynamics are chaotic, however limit cycles cause elevated estimates of WPE dependent on the choice of m and τ.

Figure 6. Relationship between weighted permutation entropy and forecast error (nRMSE, note square root scale of y axis) across 461 time series (middle panel) and specific exemplary time series (observations in black, forecasts in red, a-f). Forecast error increases with complexity of the time series as indicated by the higher permutation entropy value. The slope of the relationship was the same for time series from field and laboratory systems. The upper panels (a-c) show time series with forecast error lower than (a) or as expected (b-c) given their level of complexity, whereas the lower panels (d-f) illustrate time series which have higher than expected forecast error.

Figures

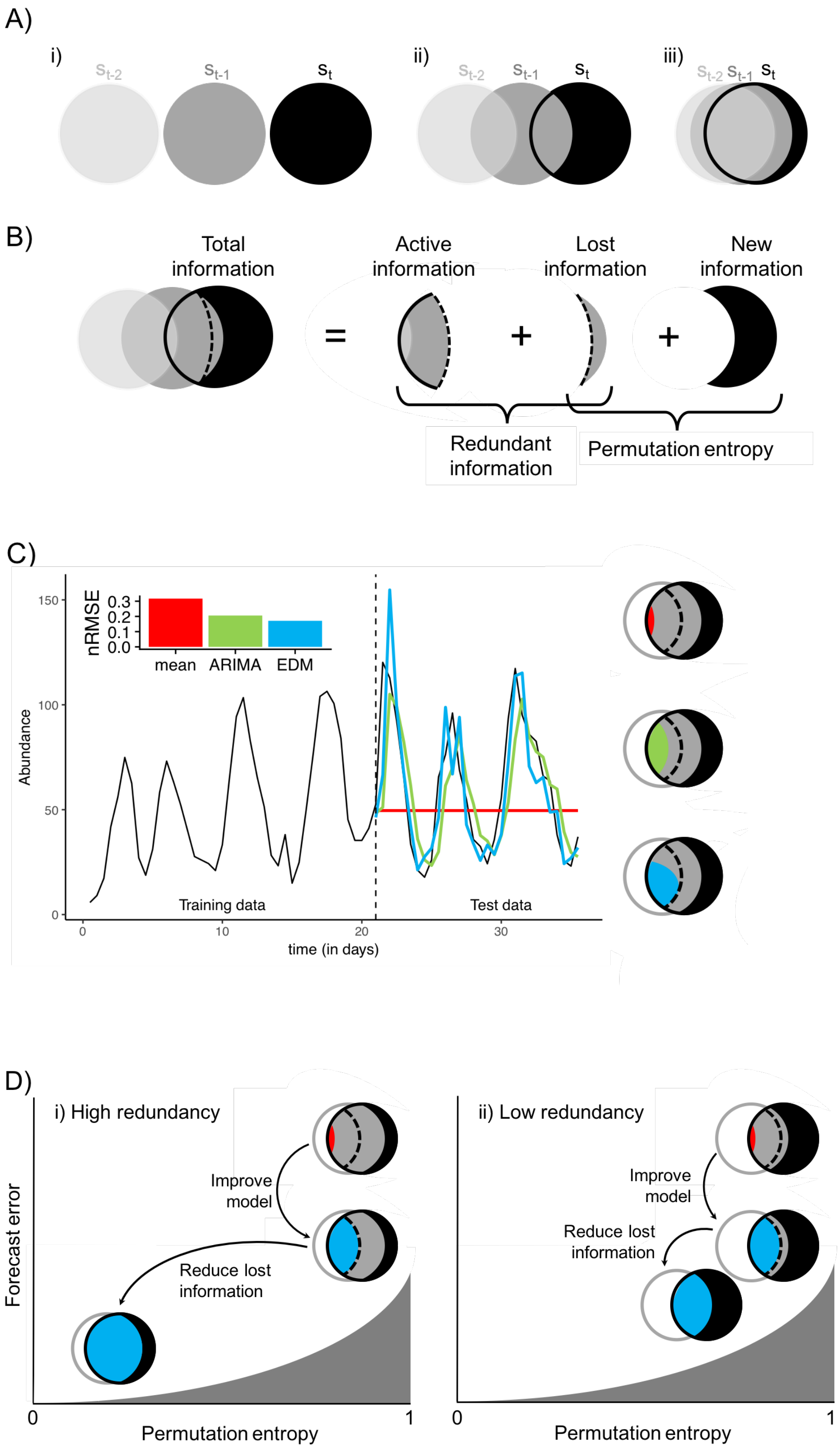


Figure 1

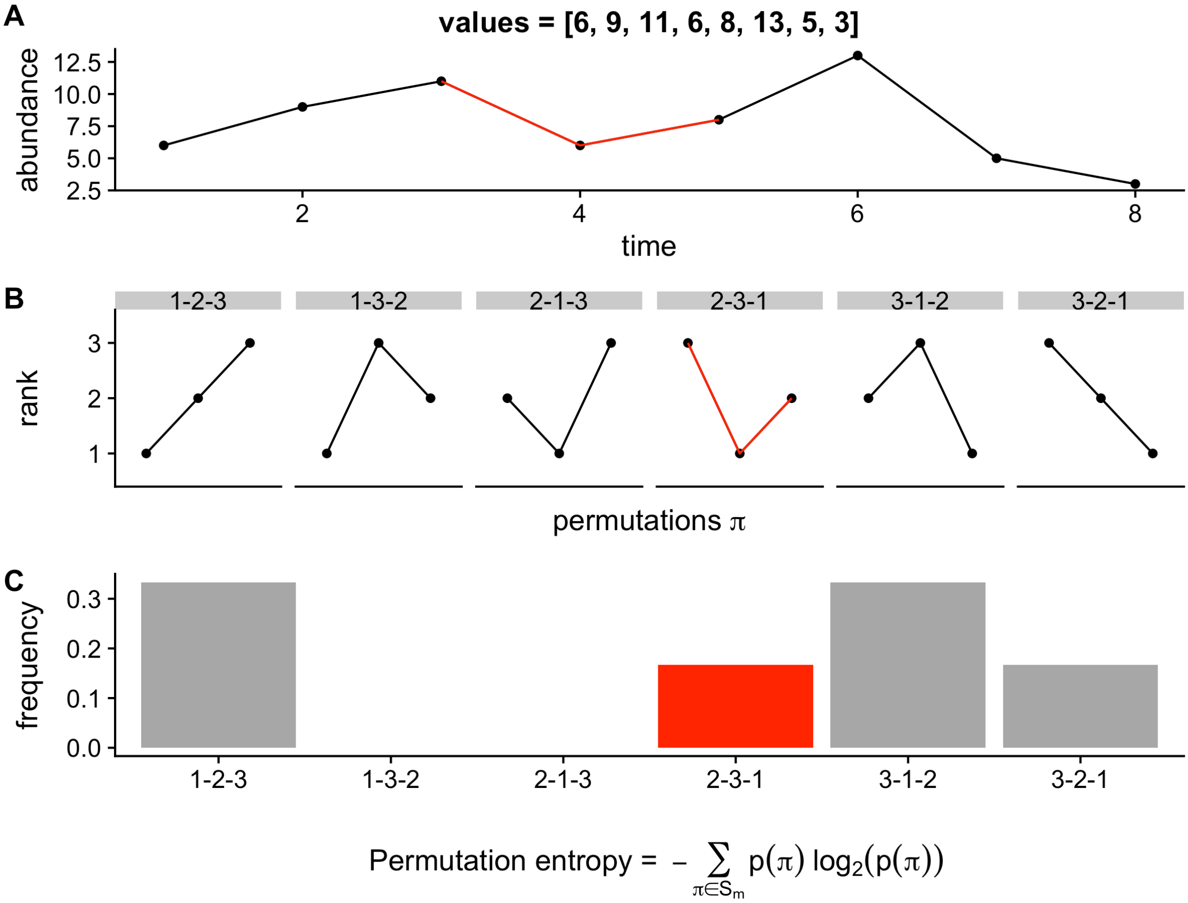

A
values = [6, 9, 11, 6, 8, 13, 5, 3]
abundance
12.5
10.0
7.5
5.0
2.5
2
4
6
8
time
B
1-2-3
1-3-2
2-1-3
2-3-1
3-1-2
3-2-1
rank
3
2
1
permutations π
C
frequency
0.3
0.2
0.1
0.0
1-2-3
1-3-2
2-1-3
2-3-1
3-1-2
3-2-1
Permutation entropy = $-\sum_{\pi \in S_m} p(\pi)\log_2(p(\pi))$

Figure 2

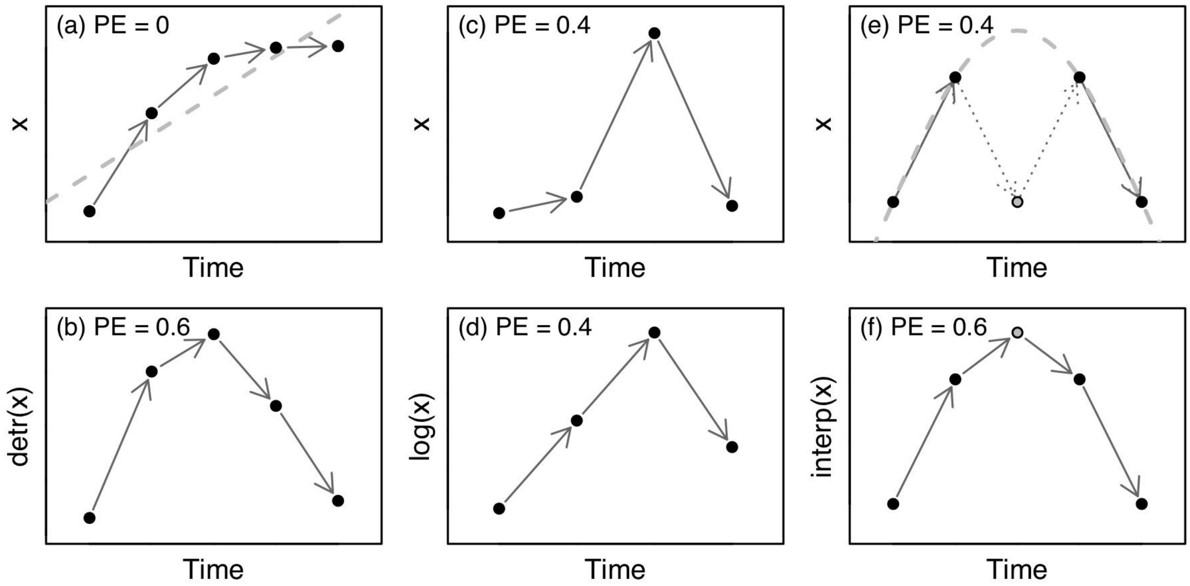
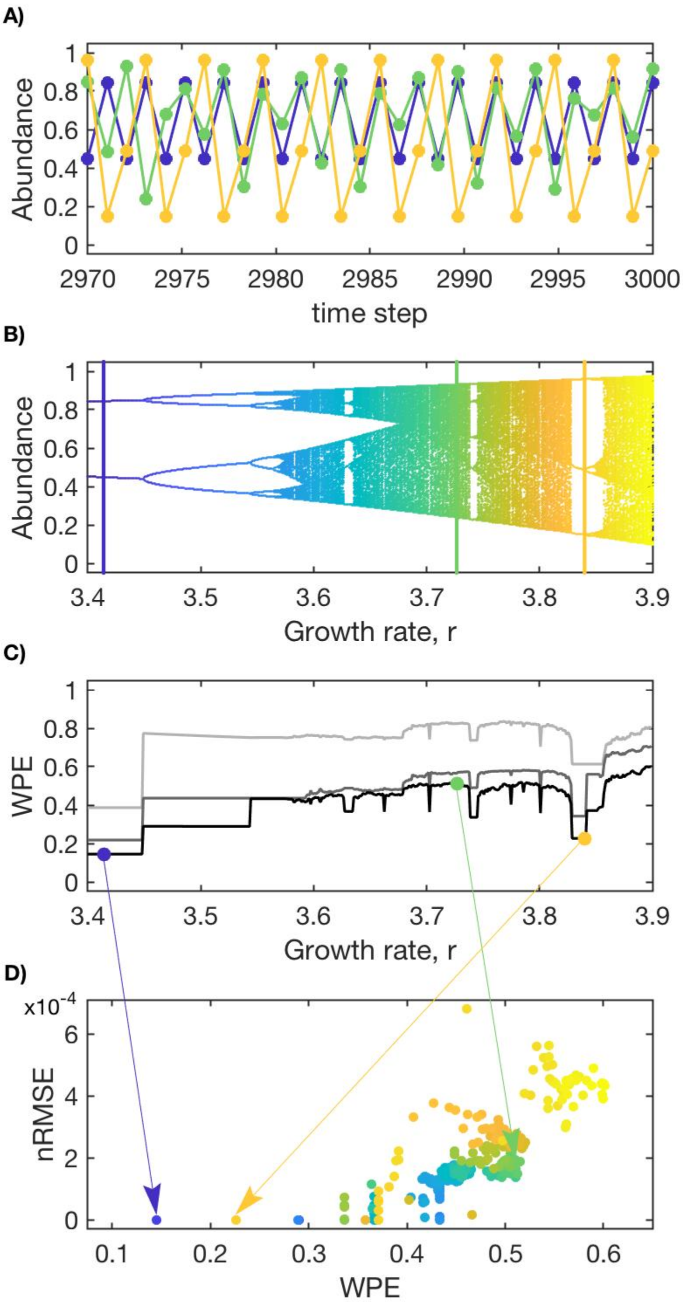

(a) PE = 0
x
Time
(c) PE = 0.4
x
Time
(e) PE = 0.4
x
Time
(b) PE = 0.6
detr(x)
Time
(d) PE = 0.4
log(x)
Time
(f) PE = 0.6
interp(x)
Time

Figure 3



902

903 **Figure 4**

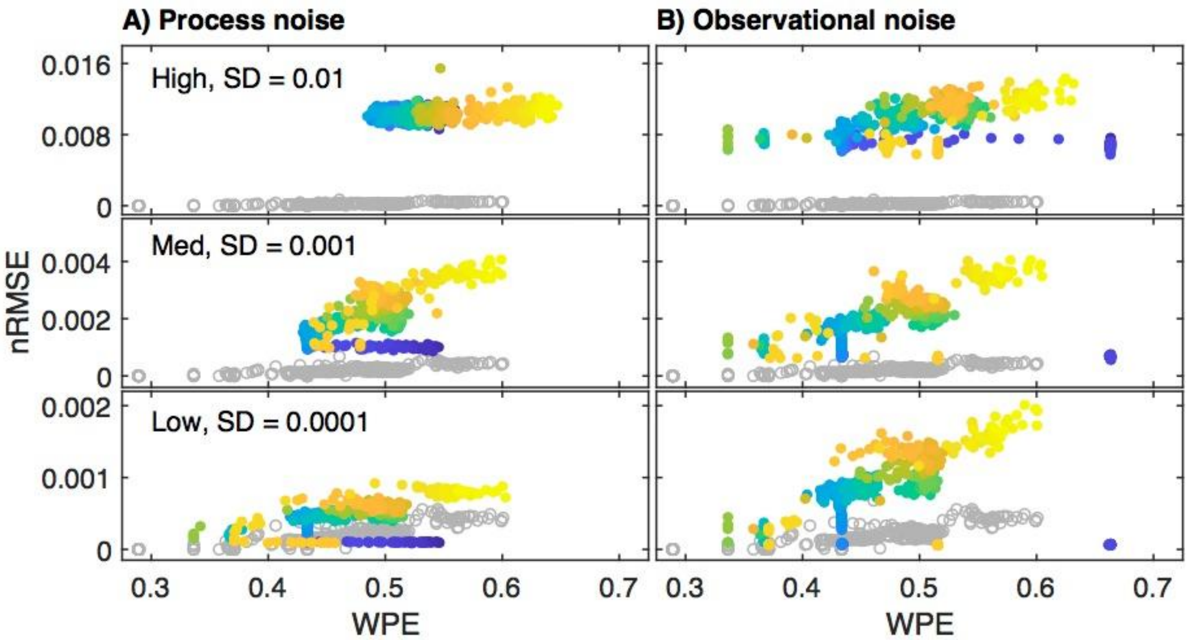

A) Process noise
B) Observational noise
High, SD = 0.01
Med, SD = 0.001
Low, SD = 0.0001
nRMSE
0.016
0.008
0
0.004
0.002
0
0.002
0.001
0
0.3
0.4
0.5
0.6
0.7
WPE
WPE

Figure 5

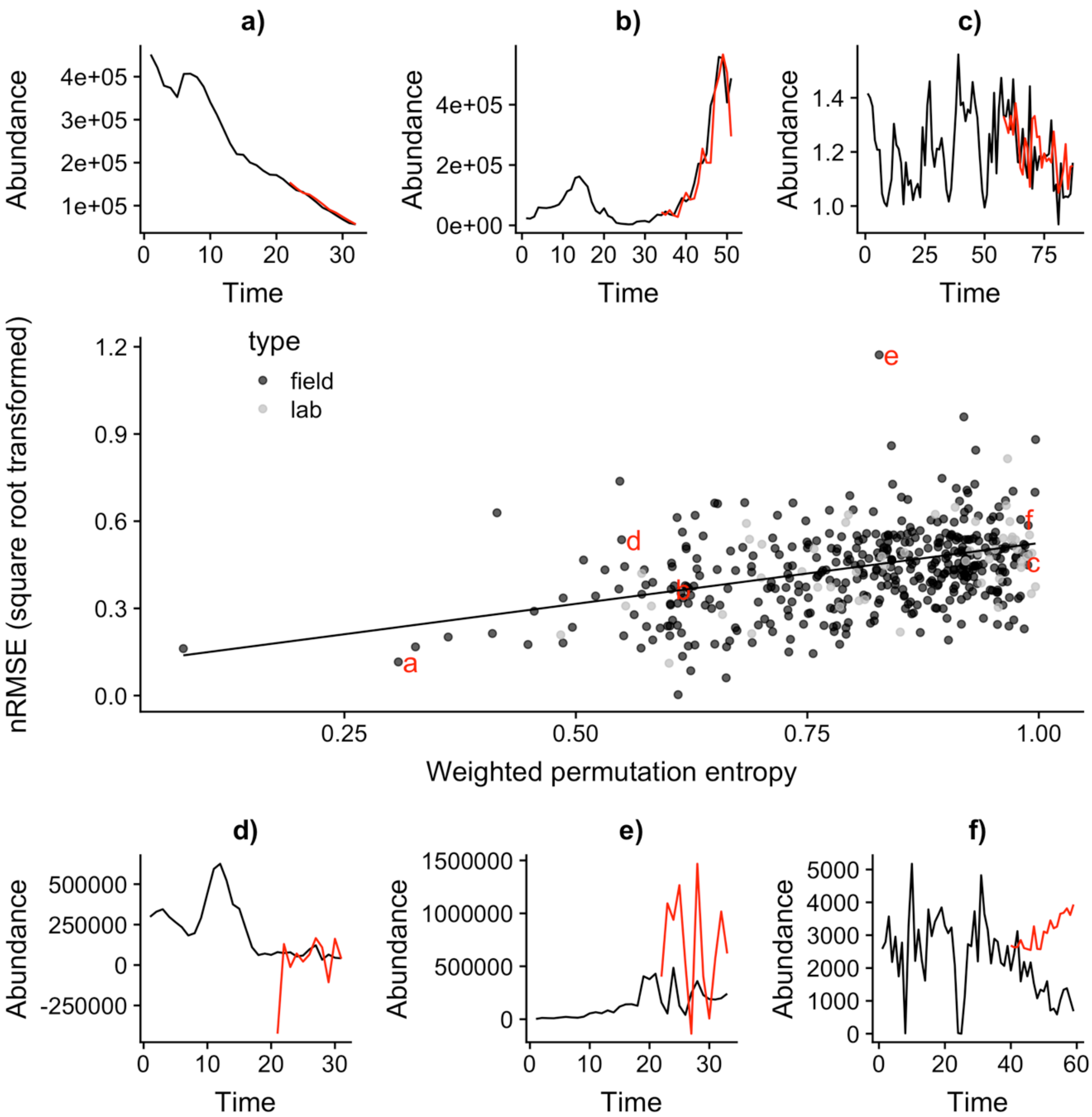

a)
b)
c)
Abundance
Time
4e+05
3e+05
2e+05
1e+05
0e+00
1.4
1.2
1.0
type
field
lab
nRMSE (square root transformed)
Weighted permutation entropy
0.25
0.50
0.75
1.00
d)
e)
f)
500000
250000
0
-250000
1500000
1000000
5000
4000
3000
2000
1000

Figure 6