APPLICATION

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

`layeranalyzer`: Inferring correlative and causal connections from time series data in R

Trond Reitan[1,2] 🆔 | Lee Hsiang Liow[1,2] 🆔

[1]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

[2]Natural History Museum, University of Oslo, Oslo, Norway

Correspondence
Trond Reitan
Email: trond.reitan@ibv.uio.no

Abstract

1. Distinguishing correlative and causal connections among time series is an important challenge in evolutionary biology, ecology, macroevolution and palaeobiology.

2. Here, we present `layeranalyzer`, an R package that uses linear stochastic differential equations as a tool for parametrically describing evolutionary and ecological processes and for modelling temporal correlation and Granger causality between two or more time series.

3. We describe the basic functions in `layeranalyzer` and briefly discuss modelling strategies by demonstrating our tool with three disparate case studies. First, we model a single time series of phenotypic evolution in a bird species; second, we extract cyclical connections in the well-known hare-lynx dataset; third, we infer the correlative and causal connections among the genus origination and extinction rates of brachiopods and bivalves.

4. We summarize the advantages and limitations of using linear stochastic differential equations and `layeranalyzer` for studying correlative and causal connections.

KEYWORDS
causal connections, correlative connections, hidden layers, layeranalyzer, processes, stochastic differential equations, time series

1 | INTRODUCTION

A common goal in ecological and evolutionary studies is to understand what drivers and factors might be involved in population, species, community and palaeontological changes (e.g. Brook, Sodhi, & Bradshaw, 2008; Charmantier et al., 2008; Ezard, Aze, Pearson, & Purvis, 2011; Turchin, 1999), and their mechanisms of effecting such changes. In some cases, it is possible to devise controlled experiments to dissect the contributions of various variables to ecological or evolutionary changes. However, where data stem from natural populations and communities, inferences must often be made from time series data that are patchy, uncertain and mere proxies for underlying processes (Houle, Pelabon, Wagner, & Hansen 2011) that we are attempting to measure.

Several tools are available for inferring connections, if any, between two or more time series, including transfer entropy (Schreiber, 2000) and cross convergent mapping (Sugihara et al., 2012), both of which are non-parametric (see review in Hannisdal & Liow, 2018). Here, we present software in the form of an R package, `layeranalyzer`, for a third approach, namely linear stochastic differential equations (SDEs, see Allen, 2003; Øksendal, 2003). Linear SDEs in the form of Ornstein–Uhlenbeck (Lande, 1976) and Wiener processes (Raup, 1977) have long been used for modelling phenotypic evolution and in phylogenetic comparative analyses (Hansen, 1997; Hansen, Pienaar, & Orzack, 2008). However, more general linear SDEs have also

been applied to questions of causality in biology relatively recently (Hannisdal, Haaga, Reitan, Diego, & Liow, 2017; Liow, Reitan, & Harnik, 2015; Reitan & Liow, 2017; Reitan, Schweder, & Henderiks, 2012).

Linear SDEs offer process models of how a given set of time series were generated, such that correlations and Granger causality (Granger, 1969) can be distinguished, especially for continuous time processes (Schweder, 1970). Granger causality is based on information flow. Roughly speaking, process B (Granger) causes process A if the predictive probability function for future states of process A is different when conditioning on all past states of all processes, and on conditioning on all past states of all processes, except process B (e.g. Eichler, 2013 and Supporting Information for details on causal connections).

Some advantages of linear SDEs as a time series tool are as follows. (a) The stochasticity of the processes can be separated from measurement noise, as the processes have temporal correlations while the latter does not. (b) Measurements do not need to be taken at equidistant time steps, unlike in ARIMA-based methods (Box, Jenkins, & Reinsel, 1994, see Chapter 2) and analysed time series can in fact be 'gappy'. This is especially useful for ecological and palaeontological data which are rarely equally spaced in time. (c) Process connections can be explored where measurements are not made at the same time points, as long as there is overlap in the time series. (d) Hidden processes ('layers') affecting the measured time series can be explicitly modelled. (e) Causal feedback loops can be modelled. (f) Measurement uncertainty can be incorporated in analyses. (g) Non-stationarity such as linear trends and random walks are permitted such that detrending data is not necessary. (h) Linear SDEs are analytically tractable (Reitan et al., 2012).

Here, we describe a new R package `layeranalyzer` that parametrically characterizes time series and models temporal correlation and Granger causality using linear SDEs. We provide three different examples to illustrate the types of ecological and evolutionary questions `layeranalyzer` can be used to answer.

2 | OVERVIEW OF `LAYERANALYZER`

`layeranalyzer` characterizes measurements of continuous time processes in the form of linear SDEs (for which each realized state is normally distributed, conditioned on the previous state), where measurement noise is also assumed to be normal. The analytically calculated expected values and covariance matrices (see Supplementary Information in Reitan et al., 2012) are used in combination with the Kalman filter, a recursive algorithm that performs process state inference (Särkkä, Vehtrari, & Lampinen, 2004) and calculates likelihood values simultaneously. Our Supporting Information contains a short verbal description of this algorithm. The Kalman smoother is used for process state inference (see Särkkä et al., 2004 and Supporting Information for details), giving inference for the process state at each given point in time, conditioned on all parameter values and all measurements. While both maximum likelihood (ML) estimation and Bayesian inference are possible in `layeranalyzer`, we focus on

the latter using Markov chain Monte Carlo (MCMC) sampling. This is because Bayesian inference was more reliable than ML for varied scenarios we have explored and we found that ML estimation in our examples (and likely in general for `layeranalyzer`) requires good starting points provided by the MCMC samples, regardless.

`layeranalyzer` also allows for detecting whether an observed process (such as changing phenotypes, demographic dynamics or diversification rates) is affected correlatively or causally (Granger, 1969) by other measured and unmeasured processes. An unmeasured process with a causal connection to another process (itself measured or unmeasured) is termed a hidden layer, hence the name of the package. The measured process is designated the top layer (i.e. layer 1), while the unmeasured, hidden process is the 'lower' layer (i.e. layer 2) in a 2-layered system. If another process (i.e. layer 3, the lowest layer in this case) drives layer 2, the system would be 3-layered, and so on. Unmeasured processes leave signatures on the autocorrelation structure of observed processes, allowing us to characterize such unmeasured time series. In addition to giving greater flexibility in describing the correlation structure of a time series, including layers in one's models can be motivated by the empirical system being studied. For instance, our earliest application aimed at distinguishing between the measured phenotype of fossil coccoliths (layer 1) and their optimum phenotype (layer 2, see Reitan et al., 2012). See the Supporting Information for more on layered systems, and for figures of process realizations.

`layeranalyzer` has four general output types:

1. Model selection. This is relevant when the structure of the process (e.g. is it stationary, with or without hidden layers) is of interest, and/or if the plausibility of connections between two or more processes is examined.

2. Parameter inference. Given the structure and/or connections of one or more processes, this output gives estimates and 95% credibility intervals for the expected values, stochastic contribution strength, the characteristic time of the process(es) and possible causal or correlative connection terms. The characteristic time describes the auto-correlation within each process. Note that characteristic time and half-life, an alternative parametrization thereof, are related to lag time for causal stochastic processes (see Supporting Information).

3. Inference of the process states. This gives estimates and uncertainties for a reconstruction of how the processes developed over time. For each time at which the original measurements of the observed time series were made, and for unmeasured, but regularly spaced time points of interest, an estimate and 95% credibility interval can be inferred for the state of each process (measured or hidden) conditioned on the measurements. This is done by applying Kalman smoothing (which infers processes conditioned on both the measurements and parameter values) while varying the parameters over the MCMC samples.

4. Process realizations. These are simulations of the processes (measured or hidden) given the measurements. Process realizations can be of interest when studying the effect of uncertainties in the process in subsequent analyses.

TABLE 1 Functions and objects in `layeranalyzer`. See Supporting Information for code and details

	Description
Objects	
`layer.data.series`	Contains single time series and associated information
`layer.prior`	Specifies prior knowledge concerning the processes in question
`layer.series.structure`	Specifies process structure (number of layers, lowest layer = OU, linear trend OU or Wiener process, deterministic layers etc.) and `layer.data.series` and `layer.prior`
Functions	
`layer.analyzer`	Used for single model analysis. It takes as input one or more `layer.series.structure` objects, connection specifications and run options and returns parameter estimates and model selection criteria
`summary.layered`	Outputs summary of a single `layer.analyzer` object
`compare.layered`	Outputs model comparison of multiple `layer.analyzer` objects
`traverse.standalone.layered`	Performs multiple calls to `layer.analyzer` for a `layer.data.series` object in order to explore the process structure of a single time series
`traverse.connections.layered`	Performs multiple calls to `layer.analyzer` for 2 or more `layer.series.structure` objects to explore connections between process structures belonging to multiple time series. Causal connections, causal feedback loops and correlative connections can be tested
`read.layer.data.series`	Reads and converts a file containing a table with time points, values and possibly other valid time series information (sites, standard deviation, number of samples) directly into a `layer.data.series` object

The key function in the `layeranalyzer` package (Table 1) is `layer.analyzer`, which analyses a single model (which may include multiple time series). Options for the `layer.analyzer` function include:

a. Process structure: The default model fitted for a single time series is a one-layered Ornstein–Uhlenbeck process (Table 1). However, non-stationary Wiener processes, linear trends or periodicity are alternative options available. Hidden layers can also be imposed.

b. Causal connections: One process $x_1(t)$ (measured or hidden) can affect another, $x_2(t)$, where $\beta_{1,2}$ denotes the strength of the causal connection. Specifically, if $x_1(t)$ changes by one unit, $x_2(t)$ will gradually change by $\beta_{1,2}$ units. This can be modelled with the option 'causal'. Causal connections can be uni-or bidirectional.

c. Correlative connections: A correlative connection is inferred if $x_1(t)$ and $x_2(t)$ are affected by the same stochastic perturbatio and this is modelled using the option 'corr'.

d. Group options. Comparable measurements from grouped locations can be analysed as a clustered set of processes with the same structure (see Reitan et al., 2012).

3 | CASE STUDY 1: ONE TIME SERIES OF PHENOTYPIC CHANGE

The data analysed in this example are the body size measurements of individuals of *Acrocephalus scirpaceus* over a 19-year period (Sætre et al., 2017). The authors suggested that these reed warblers evolved rapidly to a smaller optimal body size as described by an Ornstein–Uhlenbeck process.

We performed an automatic model search for process structure using the function `traverse.standalone.layered` after conforming the size measurements and their uncertainties into a `layer.data.series` object. We summarized the ML based (option `do.maximum.likelihood=True` in `traverse.standalone.layered`) model search output using the function `compare.layered`, so we could compare our model selection results with those of Sætre et al. (2017). `traverse.standalone.layered` has an option for the maximum number of layers explored. In this example, we set this to two (i.e. maximum one hidden layer), with the lowest layer being either an OU process, an OU process with linear time trend or a Wiener process.

According to the `layeranalyzer` AICc model choice, there were no hidden layers in the body size data. However, an OU process with a linear time trend was slightly better than a pure OU process (see Supporting Information Table S1). This suggests that the size of *A. scirpaceus* started above the optimal body size, but the optimal body size was increasing (Figure 1), a more nuanced view of what Sætre et al. (2017) concluded. The characteristic time, which summarizes the autocorrelation (which drops to 1/e for a time interval equal to the characteristic time) of the process, was estimated to 2.5 years. In other words, the process lags 2.5 years behind the slow-moving trend for increasing body size. (See Supporting Information for other parameter estimates.)

4 | CASE STUDY 2: TWO TIME SERIES OF POPULATION GROWTH

The Canadian hare–lynx system, first described by MacLulich (1937) and later Elton and Nicholson (1942), is a classic predator–prey cycle with feedback loops that has been repeatedly reanalysed (e.g. Trostel, Sinclair, Witers, & Krebs, 1987; Vik, Brinch,
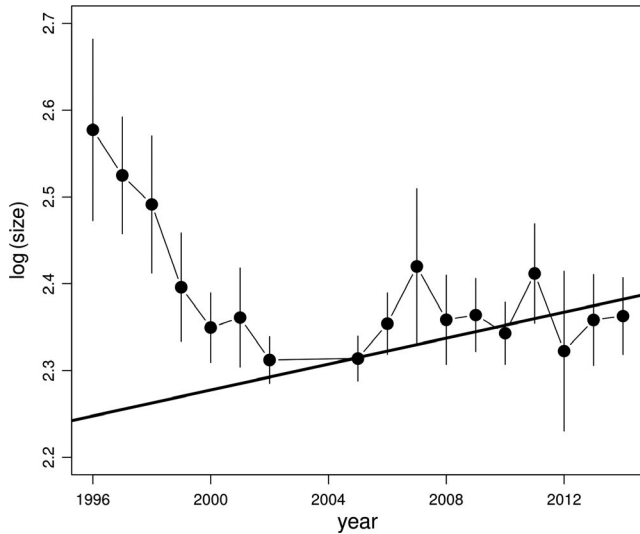
FIGURE 1 Changes in *Acrocephalus scirpaceus* body size. Yearly mean natural log body mass (g) of individuals of *A. scirpaceus*, with their 95% confidence intervals. The solid line shows the inferred, linearly changing optimal body size from the best model

Boutin, & Stenseth, 2008). We use this system to illustrate how `layeranalyzer` can be used to infer feedback loops. Note that we here use count data from traps as a proxy for yearly abundance (as done by previous authors), which is in turn used as a proxy for the continuous expected abundance process in our models (see May, 1973).

Before analyses, we normalized the count data (see Supporting Information). We then applied the function `traverse.stand-alone.layered` on hares and lynx separately. These independent searches revealed that the normalized hare and lynx counts each conform to a two-layered model with negative feedback from first to second layer, suggesting a negative-positive feedback loop with an unmeasured process. In other words, a positive-negative feedback loop leading to cyclical behaviour can be gleaned from just one time series with `layeranalyzer`. We know from previous work that lynx abundance might be driving hare abundance. To explicitly explore this, we imposed an OU-like process as the structure for each of these two measured time series. Then we searched for connections between the hare and lynx data using the function `traverse.`

`connections.layered` and compared the resulting models using `compare.layered`.

By rerunning the `layer.analyzer` function on the model deemed best using the `smoothing.specs` option, we provide estimates for the processes (Figure 2). We specified in `smoothing.specs`, that in addition to performing inference for the measured points, we wanted inference for regularly spaced time points (ten per year), in order to get an almost continuous process state inference over time. We do not have estimates for the standard deviation of each measurement error for these datasets. For such cases, `layer.analyzer` introduces a parameter for each time series that represents a common standard deviation for all the measurement errors.

5 | CASE STUDY 3: THREE TIME SERIES OF DIVERSIFICATION RATES

In this example, we reproduce the analyses in Reitan and Liow (2017), with simplifications, to illustrate how three or more time series can be investigated simultaneously using `layeranalyzer`. Reitan and Liow (2017) investigated how the diversification dynamics of whole clades might influence those of different clades with similar niches using genus origination and extinction rates of bivalves and brachiopods estimated in Liow et al. (2015). The best model describing these processes is one where high bivalve extinction rates drove brachiopod origination rates through more than 450 million years of their evolutionary history (Reitan & Liow, 2017). There was also a causal connection from bivalve to brachiopod extinction rates.

We constrain the enormous number of joint process models examined in Reitan and Liow (2017) to 25 by (a) reducing the original six observed time series (Reitan & Liow, 2017) to only three (b) only analysing models where one connection has been added, changed or removed relative to the model considered best by Reitan and Liow (2017) and (c) assuming the same prior knowledge of the system for the purpose of demonstration. Specifically, we assume that the structure of these three time series, namely extinction rates for brachiopods and bivalves (one hidden layer each) and the origination rates for brachiopods (only one measured layer) are known, to
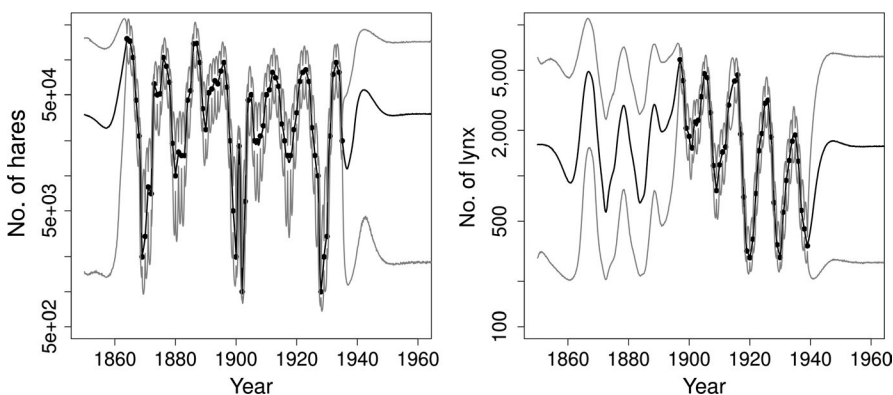

FIGURE 2 Lynx and hare process state inference. The plotted points are empirical data while black lines are estimated means and grey lines are 95% credibility bands. Lynx data start in 1897 and end in 1939 (with 1914 missing), while hare data start in 1864 and end in 1935. Note that the uncertainty bands make small 'bubbles' between each years as the models are in continuous time and the uncertainty increases between measurements

reduce analysis time. Note that in addition to affecting their corresponding observed processes, hidden layers also function as common causes in a connection analysis. Since we are examining a fixed set of models rather than traversing all possible models, we use `layer.analyzer` to examine each of the 25 models and then use `compare.layered` to compare and evaluate them afterwards (see Supporting Information for details).

Using `layeranalyzer`, we find that the process model presented in Reitan and Liow (2017), was indeed the best one among the 25 models examined, according to the Bayesian model likelihood and resulting model probability (see Supporting Information for parameter estimates and comparisons with estimates from Reitan and Liow (2017)).

6 | CAUTIONARY NOTES AND FUTURE DEVELOPMENT FOR LINEAR SDES AND LAYERANALYZER

Linear SDEs perform well in recovering correlative and causal processes under the set of scenarios we have explored (see Supporting Information in Liow et al., 2015). While a comprehensive exploration of the performance of this tool is out of the scope of the current work, we have provided simulation tools for users to investigate the behaviour of `layeranalyzer` (see the project web page https://github.com/trondreitan/layeranalyzer).

Since linear SDEs are Gaussian processes, the input data are assumed to be normally distributed, and two of our examples required transformation to meet this assumption. Additionally, in order to calculate an analytical likelihood, measurement noise should be also normally distributed.

Nonlinear connections between processes, such as in the Lotka–Volterra model for predator–prey relationships (Lotka, 1910) breaks the underlying assumption of linear SDEs. That said, such a linear model can still capture the salient features of nonlinear connection models, as demonstrated in the hare-lynx example. However, highly nonlinear connections cannot be correctly modelled using this tool, even with data-transformation. By using linear SDEs, there is also an implicit assumption that each process depends on its previous state (as well as the state of other connected processes) in a linear fashion. It is worth exploring how each residual might depend on the previous residual where this is a concern, by using, for instance, nonlinear regression tools. Where nonlinearity is a serious concern, nonparametric tools such as cross convergent mapping (Sugihara et al., 2012; Ye, Deyle, Gilarranz, & Sugihara, 2015) could be explored (see Hannisdal & Liow, 2018 for a review).

Even when the processes are normal and depend linearly on the previous state values, the measurements might be highly non-normal, such as in the case of counts (especially low counts) and presence/absence data. In our second example, we have followed previous literature in simply using (large) counts as proxies for continuous processes. However, for future applications to such data, we plan on developing a GLM version of the linear SDE toolbox, where data can be linked to the underlying linear process(es) using a set of standard distribution families via a link function.

In summary our tool `layeranalyzer` provides a means to characterize and quantify ecological and evolutionary processes from time series data.

AUTHORS' CONTRIBUTIONS

T.R. led the coding and method development and L.H.L. helped to test the code and wrote the first draft of the manuscript; T.R. and L.H.L. developed the examples. All authors contributed to revisions of the manuscript and gave final approval for submission.

DATA AVAILABILITY STATEMENT

The project page is https://doi.org/10.5281/zenodo.3406161 (Reitan & Liow, 2017) and package can be installed from Github using install_github(repo="trondreitan/layeranalyzer"). The data and example code are supplied in the package. The package will be maintained on github and on the first author's website.

ORCID

Trond Reitan (iD) https://orcid.org/0000-0003-0793-8731
Lee Hsiang Liow (iD) https://orcid.org/0000-0002-3732-6069

REFERENCES

Allen, L. J. S. (2003). *An introduction to stochastic processes with applications to biology.* Upper Saddle River, NJ: Pearson Education Inc.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis, Forecasting and control* (3rd ed.). Englewood Cliff, NJ: Prentice-Hall.

Brook, B. W., Sodhi, N. S., & Bradshaw, C. J. A. (2008). Synergies among extinction drivers under global change. *Trends in Ecology & Evolution*, *23*, 453–460. https://doi.org/10.1016/j.tree.2008.03.011

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & Sheldon, B. C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, *320*, 800–803. https://doi.org/10.1126/science.1157174

Eichler, M. (2013). Causal inference with multiple time series: Principles and problems. *Philosophical Transactions of the Royal Society A:*

Mathematical, Physical and Engineering Sciences, 371, 20110613. https://doi.org/10.1098/rsta.2011.0613

Elton, C., & Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in Canada. *The Journal of Animal Ecology*, 11, 215–244. https://doi.org/10.2307/1358

Ezard, T. H. G., Aze, T., Pearson, P. N., & Purvis, A. (2011). Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science*, 332, 349–351. https://doi.org/10.1126/science.1203060

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424. https://doi.org/10.2307/1912791

Hannisdal, B., Haaga, K. A., Reitan, T., Diego, D., & Liow, L. H. (2017). Common species link global ecosystems to climate change: Dynamical evidence in the planktonic fossil record. *Proceedings of the Royal Society B-Biological Sciences*, 284, 20170722. https://doi.org/10.1098/rspb.2017.0722

Hannisdal, B., & Liow, L. H. (2018). Causality from palaeontological time series. *Palaeontology*, 61, 495–509. https://doi.org/10.1111/pala.12370

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51, 1341–1351. https://doi.org/10.1111/j.1558-5646.1997.tb01457.x

Hansen, T. F., Pienaar, J., & Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62, 1965–1977. https://doi.org/10.1111/j.1558-5646.2008.00412.x

Houle, D., Pelabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and meaning in biology. *Quarterly Review of Biology*, 86, 3–34. https://doi.org/10.1086/658408

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30, 314–334. https://doi.org/10.1111/j.1558-5646.1976.tb00911.x

Liow, L. H., Reitan, T., & Harnik, P. G. (2015). Ecological interactions on macroevolutionary time scales: Clams and brachiopods are more than ships that pass in the night. *Ecology Letters*, 18, 1030–1039. https://doi.org/10.1111/ele.12485

Lotka, A. J. (1910). Contribution to the theory of periodic reaction. *Journal of Physical Chemistry*, 14(3), 271–274. https://doi.org/10.1021/j150111a004

Maclulich, D. A. (1937). Fluctuations in the numbers of the varying hare (*Lepus americanus*). University of Toronto Studies, Biological series 43.

May, R. M. (1973). *Stability and complexity in model ecosystems*. Princeton, NJ: Princeton University Press.

Øksendal, B. (2003). *Stochastic differential equations. An introduction with applications* (6th ed.). Berlin: Springer-Verlag.

Raup, D. M. (1977). Stochastic models in evolutionary paleontology. In A. Hallam (Ed.), *Patterns of evolution, as illustrated by the fossil record* (pp. 123–158). New York: Elsevier.

Reitan, T., & Liow, L. H. (2017). An unknown Phanerozoic driver of brachiopod extinction rates unveiled by multivariate linear stochastic differential equations. *Paleobiology*, 43(4), 537–549. https://doi.org/10.1017/pab.2017.11

Reitan, T., Schweder, T., & Henderiks, J. (2012). Phenotypic evolution studied by layered stochastic differential equations. *Annals of Applied Statistics*, 6, 1531–1551. https://doi.org/10.1214/12-AOAS559

Sætre, C. L. C., Coleiro, C., Austad, M., Gauci, M., Sætre, G., Voje, K. L., & Eroukhmanoff, F. (2017). Rapid adaptive phenotypic change following colonization of a newly restored habitat. *Nature Communication*, 8, 14159. https://doi.org/10.1038/ncomms14159

Särkkä, S., Vehtrari, A., & Lampinen, J. (2004). Time series prediction by Kalman Smoother with cross-validated noise density. *IEEE International Joint Conference on Neural Networks*, 1–4, 1615–1619.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85, 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Schweder, T. (1970). Composable Markov processes. *Journal of Applied Probability*, 7(2), 400–410. https://doi.org/10.2307/3211973

Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338, 496–500. https://doi.org/10.1126/science.1227079

Trond, R., & Liow, L. H. (2019). layeranalyzer v0.1.0 (Version 0.1.0). *Zenodo*. http://doi.org/10.5281/zenodo.3406161

Trostel, K., Sinclair, A. R. E., Witers, C. J., & Krebs, C. J. (1987). Can predators cause the 10-year hare cycle? *Oecologia (Berlin)*, 74, 185–192.

Turchin, P. (1999). Population regulation: A synthetic view. *Oikos*, 84, 153–159. https://doi.org/10.2307/3546876

Vik, J. O., Brinch, C. N., Boutin, S., & Stenseth, N. C. (2008). Interlinking hare and lynx dynamics using a century's worth of annual data. *Population Ecology*, 50, 267–274.

Ye, H., Deyle, E. R., Gilarranz, L. J., & Sugihara, G. (2015). Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, 5, 14750. https://doi.org/10.1038/srep14750

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.