Ecology, 96(5), 2015, pp. 1174–1181

Reports

Spatial convergent cross mapping to detect causal relationships from short time series

Adam Thomas Clark,[1,5] Hao Ye,[2] Forest Isbell,[1,3] Ethan R. Deyle,[2] Jane Cowles,[1] G. David Tilman,[1,4] and George Sugihara[2]

[1]*University of Minnesota, Department of Ecology, Evolution, and Behavior, 1987 Upper Buford Circle, Saint Paul, Minnesota 55108 USA*
[2]*Scripps Institution of Oceanography, University of California–San Diego, 9500 Gilman Drive, La Jolla, California 92093 USA*
[3]*Department of Plant Biology, 2502 Miller Plant Sciences, University of Georgia, Athens, Georgia 30602 USA*
[4]*Bren School of Environmental Science and Management, University of California, 2400 Bren Hall, Santa Barbara, California 93106 USA*

Abstract. Recent developments in complex systems analysis have led to new techniques for detecting causal relationships using relatively short time series, on the order of 30 sequential observations. Although many ecological observation series are even shorter, perhaps fewer than ten sequential observations, these shorter time series are often highly replicated in space (i.e., plot replication). Here, we combine the existing techniques of convergent cross mapping (CCM) and dewdrop regression to build a novel test of causal relations that leverages spatial replication, which we call multispatial CCM. Using examples from simulated and real-world ecological data, we test the ability of multispatial CCM to detect causal relationships between processes. We find that multispatial CCM successfully detects causal relationships with as few as five sequential observations, even in the presence of process noise and observation error. Our results suggest that this technique may constitute a useful test for causality in systems where experiments are difficult to perform and long time series are not available. This new technique is available in the multispatialCCM package for the R programming language.

Key words: causality; convergent cross mapping; dewdrop regression; multispatialCCM; spatial replication; time series.

Introduction

Detecting causal relationships in complex systems is one of the fundamental and most challenging goals of science. Convergent cross mapping (CCM) has recently been introduced as a practical numerical approach for identifying causal relationships in weakly coupled nonlinear systems (Sugihara et al. 2012). This approach is potentially of great utility in ecology, where many systems appear to be weakly coupled and complex and can therefore be difficult to analyze using traditional techniques (e.g., repeated-measures or mixed-effects ANOVA, general algebraic modeling systems, neural networks, or autoregressive models). Typically, CCM can be applied to time series of roughly 30 or more sequential observations (Sugihara et al. 2012). However, ecological time series are often even shorter. In this report, we present a method that expands the application of CCM to very short time series that are spatially replicated (e.g., data from multiple plots).

CCM is based on an algorithm that compares the ability of lagged components of one process to estimate the dynamics of another. In ecology, these processes might represent time series observations of environmental data, such as temperature, or of species data, such as population abundance. There are three basic ways in which two processes can be causally linked: neither influences the others' temporal dynamics, and the variables are therefore causally unrelated; a forcing process influences the temporal dynamics of a response process, but the response process has no effect on the forcing process in return, meaning that there is unidirectional causality; or there is bidirectional causality where each variable influences the others' dynamics. In practice, the ability of any method to accurately distinguish between these cases depends not only on the strength of underlying relationships, but also on the amount of data available, the presence of process noise and observation error, and the sensitivity and assumptions of the method. We explore cases for which the underlying strength and direction of causal interactions, process noise, and observation error are known, and use these to assess the detection limits of CCM when applied to short, replicated observational records.

As described by Sugihara et al. (2012), the dynamics of a unidirectional forcing process can be more accurately estimated using information from a response process than is true for the reverse case. This counter-

Manuscript received 4 August 2014; revised 17 December 2014; accepted 22 December 2014. Corresponding Editor: B. D. Inouye.

[5] E-mail: adam.tclark@gmail.com

intuitive phenomenon occurs because the response process necessarily contains information about the forcing process whereas the reverse may not be true. For example, consider the dynamics of a species responding to an exogenous environmental variable, such as length of day. In a weakly coupled system, many processes influence species abundance, and consequently, day length alone will not be a good predictor of the organism's endogenous abundance dynamics. However, changes in day length (the forcing process) necessarily propagate to and influence the dynamics of the species (the response process), and day length therefore can be predicted from species abundance. Thus, observed states of a forcing process should be significantly better described by observed states of a response process than would be expected by random chance, and the accuracy of that description should improve with increasing time series length, as this increases the amount of information available with which to estimate system dynamics (Fig. 1a). This joint criterion therefore provides a metric to test for significant causal relationships among processes. Algorithmic and visual descriptions of this approach are available in the supplement of Sugihara et al. (2012).

Although CCM may perform relatively well on short time series of 30+ observations, longer time series are generally preferable, particularly when causal relations are weak or process noise and observation error are large (Sugihara et al. 2012). However, ecological time series are often much shorter, and experiments and observations generally focus on spatial rather than temporal replication. Thus, many potential applications of CCM may not be feasible using the currently published framework. Nevertheless, spatially replicated data do contain temporal information, and it seems plausible that sufficient replication might compensate for brevity in time series. We present a method (multispatial CCM) to recover this information for application in CCM through a bootstrapping technique that has previously been applied to single variable simplex projection (dewdrop regression; Hsieh et al. 2008). Dewdrop regression combines information from many short time series of systems that share similar dynamical forms to predict their future dynamics. We adapt this method for CCM in order to leverage the spatial breadth in ecological data sets. We also present the first comprehensive test of detection power and error rates for CCM across a broad range of forcing strengths, observation error, and process noise magnitudes.

Methods

Multispatial CCM algorithm

To test the predictive ability of variables, we use simplex projection, an application of Takens' theorem (Sugihara 1990). Simplex projection predicts the dynamics of a process that is part of a larger system by using information from multiple lagged observations of that single process. Consider as an example a predator–prey system, where both predator and prey dynamics influence one another. Because interacting processes contain information about one another, accurate predictions of future predator population dynamics can be made based on knowledge of previous predator dynamics, even if no information on prey populations is available (Schaffer 1985, Sugihara and May 1990, Sugihara 1994). Three consecutive years of predator population dynamics might be compared to three-year periods from historical data. To predict next year's population size, simplex projection identifies a subset of observed three-year dynamics that are most similar to the current three-year trend, and averages predictions across these observations weighted by their similarity. Recent work has shown these techniques to be applicable under a wide range of circumstances (Sugihara and May 1990, Deyle and Sugihara 2011) and perform well both on simulated and observed data (Deyle et al. 2013, Perretti et al. 2013).

Classical simplex projection uses a single long time series, which means that most observations have many corresponding historical observations, and many lagged time steps that can be used to make predictions and estimates. In the multispatial approach that we propose, we instead leverage information from many short time series, by drawing samples from the pool of all spatially replicated observations. We then look across these samples to find observations with similar historical dynamics and use a weighted average of those observations to estimate expected dynamics for subsequent time steps. Based on these estimates, we can characterize how well the dynamics of a putative forcing process are described by the dynamics of a suspected response process and thus assess their causal relatedness.

To test for significant causal relationships between two candidate processes, the multispatial CCM algorithm proceeds in five steps, which are described briefly below. An annotated implementation of these steps for the R programming environment (R Core Development Team 2014) is available in the multispatialCCM package available through the Comprehensive R Archive Network (CRAN) (*available online*).[6] Worked examples of applying these functions to real and simulated data are available in the Appendix.

Determine the best embedding dimension for the analysis.—In simplex projection, the accuracy of predictions varies as a function of embedding dimension, E, which describes the size of the time windows (i.e., number of time steps) that are used for prediction ($E = 3$ in the predator–prey example in *Methods: Multispatial CCM algorithm*). One can estimate an optimal embedding dimension by using simplex projection to test the ability of a process to predict its own dynamics through leave-one-out cross-validation (i.e., removing one observation from the time series and using the rest of the time series to predict its dynamics).

[6] http://CRAN.R-project.org

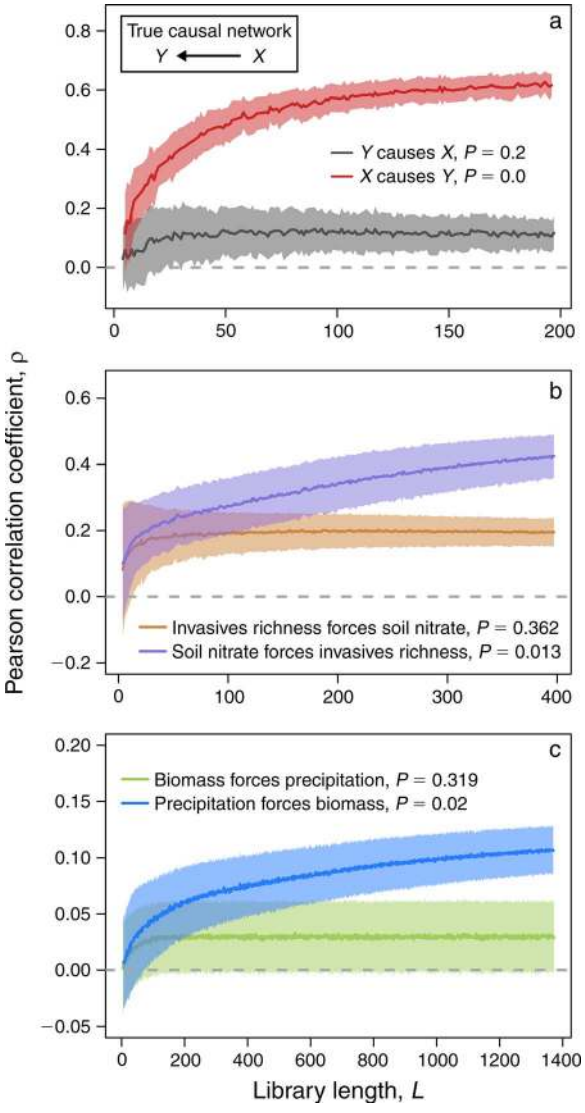


FIG. 1. Example applications of multispatial CCM. Lines and shaded regions show mean ± SD from bootstrapped iterations. Causal forcing is indicated when the Pearson correlation coefficient ρ is significantly greater than zero for large library length L (number of historical observations, including observation time and number of spatial replicates included in the composite time series) and that ρ increases significantly with increasing L. (a) A simulated system of interacting processes, where X causes Y, but Y does not cause X, as correctly indicated by the test (causal network shown in inset), is based on 20 observations in 10 plots, as in the multispatialCCM R package example files (Supplement). (b) A real-world example (Appendix: Fig. A14b) comparing soil nitrate and invading plant species richness in a plant biodiversity experiment at Cedar Creek, indicating that nitrate dynamics force invader species richness (purple line and shading), but the reverse is not true (brown line and shading). (c) A second example (Appendix: Fig. A11a) comparing annual summer precipitation (June–August) and *Elymus* (*Agropyron*) *repens* aboveground biomass dynamics, indicating that biomass is forced by precipitation (blue line and shading), but the reverse is not true (green line and shading). See Appendix for details.

In multispatial simplex projection, we sample across many short time series from spatial replicates. Because of this, it is important that time lags should not cross gaps between plots in the composite time series. Even though multispatial CCM assumes that data from multiple plots come from the same dynamical system, observations from the end of the time series of one plot will have no bearing on the beginning of the time series of another plot. As a result, lagged dimensions are only considered when they come from the same plot (i.e., do not cross any gaps), which therefore limits the potential embedding dimension E for an analysis. For example, if only five observations are recorded in each plot, then $E \leq 4$, as the analysis requires four observations to describe historical dynamics followed by one observation against which to test estimated values.

Test for nonlinearity and stochastic noise.—Because CCM applies in coupled nonlinear systems, it is important that the system under consideration is not purely random, in which case other tests for causality (e.g., Granger's test; see Detto et al. 2012) rather than CCM might instead be considered (Sugihara et al. 2012). Furthermore, it is important to test that dynamics are not so dissimilar among plots, or so strongly influenced by stochastic noise, that causal links cannot be recovered. To accomplish this, the same simplex projection algorithm described in *Methods: Multispatial CCM algorithm: Determine the best embedding dimension for the analysis* can be applied. Using the E identified with techniques described above, predictions should be made based on historical dynamics for observations that are increasingly far into the future. If the system is nonlinear, then predictive power should significantly decrease with increasing prediction time (Sugihara 1994). Predictive power for short time intervals should also be better than expected by random chance. Note that this is a necessary but insufficient criterion for determining nonlinearity. For a completely random process, predictive ability should be equally low regardless of prediction interval. However, an autocorrelated system dominated by stochastic noise could show the same decreasing pattern that we use to detect nonlinearity. In these cases, CCM should (correctly) return that there is no causal forcing in either direction, as increasing information about the system should not increase predictive ability.

Calculate two processes' abilities to describe each others' dynamics using CCM.—In addition to varying with E, predictive skill also varies as a function of library length, L, which describes the number of historical observations that are used to generate predictions. In multispatial CCM, L increases both with increased observation time in each spatial replicate and with increased number of spatial replicates included in the composite time series. As L increases, it is more likely that previous trends will have been observed that are similar to current trends, thus improving the predictive ability of simplex projection. Because processes that are

PLATE 1. A small bunch of *Andropogon gerardii* in an old field at Cedar Creek, Minnesota, USA. *A. gerardii* reduces local soil nitrate concentrations to levels significantly below those required by many other species, leading to reduced recruitment around the individual. Similar effects occur in plots with low soil nitrate concentrations in the Big Biodiversity experiment, which reduces invasion success by non-planted species. Photo credit: A. T. Clark.

responding to the effects of a forcing process should be good indicators of the dynamics of the forcing process (just as species abundance should be a good indicator of day length in the example in *Methods: multispatial CCM algorithm*), the strength of that description should consequently increase as L increases, approaching some finite limit for very large L (Sugihara et al. 2012).

Use bootstrapping to leverage spatial information.—To prevent the order in which spatial replicates are sampled from influencing the outcome of the test, we use a simple, nonparametric bootstrapping routine. Given n spatial replicates, we draw n samples (with replacement) from among all spatial replicates and repeat tests for the best E, the relationship between predictive power and interval, and for CCM for this resampled assemblage. We then repeat this procedure for many iterations in order to average across many potential combinations and orderings of spatial data. This also provides estimates for uncertainty around the predictive power estimated in each test. In the annotated examples we present, we use 100 iterations per test and for each iteration calculate the Pearson correlation coefficient ρ, comparing predicted estimates from the CCM algorithm to observed values at each library length. For analysis of real data, the number of iterations required will almost always exceed 100 (e.g., we use 1000 replicates for our real-world examples). We discuss how to determine whether sufficient iterations have been conducted in the Appendix.

Test whether these predictions indicate a significant causal relationship.—Finally, it is necessary to determine whether the calculated ρ is significantly greater than zero and whether it increases significantly with L. Testing for this increase can be complicated by two factors. First, because the relationship between ρ and L is often not well described by simple parametric curves, statistical tests, such as regression based on particular functions, may not be a good indicator of significance in all cases. Second, a very rapid rise of ρ with L may be an indication of synchrony (see Sugihara et al. 2012 and citations therein). Synchrony arises when forcing of one process by another is so strong that the response process dynamics become subordinate to the forcing process. Consequently, both processes become good predictors of one another's dynamics, even if forcing is solely unidirectional, which can confound CCM analysis and should not necessarily be taken as an indication of a causal link.

In the analyses presented here, we interpret CCM results for which ρ is greater at the longest L available than at the shortest L available and where ρ at the longest L is greater than zero as indicating causal forcing. Note that the longest L is determined by data availability, whereas the shortest is determined by E. To test for statistical significance of this signal, we use the nonparametric bootstrapping conducted in *Methods: Multispatial CCM algorithm: Use bootstrapping to leverage spatial information* to determine whether the ρ vs. L relationship passes both of these criteria for at least 95% of bootstrapped iterations.

Simulating system dynamics

We applied the multispatial CCM method to a simulated dynamical system as considered in Sugihara et al. (2012). The model describes competition between two species, where X and Y signify the abundance of the two species, t and $t+1$ are times, r_X and r_Y are species' intrinsic growth rates, α describes the effect of species X on the dynamics of species Y, and β describes the effects of species Y on species X:

$$X(t+1) = X(t)\Big(r_X - r_X X(t) - \beta Y(t)\Big)$$

$$Y(t+1) = Y(t)\Big(r_Y - r_Y Y(t) - \alpha X(t)\Big).$$

We used the parameters $\beta = 0$ and $\alpha \geq 0$ for all simulations presented here. This corresponds to a system where process Y is influenced by process X, but X is not influenced by Y, except for the case where $\alpha = 0$, in which case neither process influences the other.

To introduce process noise into our simulation, we drew r_X from a random normal distribution centered on 3.8 with SD σ_P, and r_Y from a random normal distribution centered on 3.5 also with SD σ_P for each simulation of each plot. Consequently, species dynamics varied among plots based on σ_P, but did not vary through time within a single plot. To introduce observation error into the system that was proportional to species abundance, we multiplied each observed value of X and Y by a random, log-normally distributed variable with mean 1 (0 on the log scale) and SD σ_O after simulating the entire time series. Consequently, the observation error did not change the actual trajectory of the processes, but did alter the values that were used for subsequent analyses.

To test the ability of multispatial CCM to detect different causal signals across a range of forcing strengths and magnitudes of process noise and observation error, we simulated the system for five different values of α spaced evenly between 0 and 2.5, for five values of σ_P spaced evenly between 0 and 0.2, and for five values of σ_O spaced evenly between 0 and 0.2. In these simulations, the average values for $X(t)$ and $Y(t)$ are both around 0.7. Thus, across the range of α values that we tested (0.000, 0.625, 1.250, 1.875, and 2.500), roughly 0%, 7.1%, 14.1%, 21.2%, and 28.2%, respectively, of the dynamics of $Y(t)$ were determined by the dynamics of $X(t)$. Similarly, the range of σ_P and σ_O values that we tested (0.00, 0.05, 0.10, 0.15, and 0.20) corresponded to average differences of roughly 0%, 1.6%, 3.1%, 4.6%, and 6.0%, respectively, in individual species' growth rates among plots (for σ_P), and 0.0%, 5.9%, 11.4%, 17.0%, and 23.4%, respectively, between observed and actual values of $X(t)$ and $Y(t)$ for σ_O.

We compared these parameters in a fully factorial design ($5 \times 5 \times 5 = 125$ combinations), simulated each parameter set 1200 times, and bootstrapped each simulated library 100 times. This resulted in 15 million individual multispatial CCM tests for causality in each direction. In each simulation, we chose the first value for X and Y in each plot from a random, uniform distribution between 0 and 1. We considered three scenarios, representing different kinds of plot data. In scenario 1, we simulated 20 observations from each of 100 separate systems, representing 20 sequential observations in 100 plots. Second, we simulated 10 observations from each of 100 plots, and third, 5 observations in each of 100 plots to assess the effect of shrinking time series length on multispatial CCM performance. These time periods roughly correspond to the number of annual observations that might be expected from an established long-term ecological research site, from a single long-term research project, and from a PhD dissertation, respectively. In the Appendix, we consider a further set of cases in the same model with bidirectional causation (i.e., where $\beta > 0$).

Real-world examples

We also apply CCM to two examples of real-world data. First, we consider a short time series (4–7 sequential samples in each of 72 spatial replicates) describing the dynamics of invading plant species and soil nitrate in the Big Biodiversity experiment at Cedar Creek, Minnesota, USA. Experimental results suggest that soil nitrate dynamics force invasion dynamics, while the reverse should not be true (Fargione and Tilman 2005; see Plate 1)). Second, we consider a longer time series (62 replicates of about 20 observations each) recording annual dynamics for summer precipitation and aboveground biomass for the cool-season, drought-intolerant grass *Elymus* (*Agropyron*) *repens*. In this system, *E. repens* dynamics should be forced by precipitation, but the reverse should not be true. Eight further examples and detailed methods, diagnostic tests, and descriptions of the data are presented in Appendix A.

Results

Our analyses of simulated data using multispatial CCM correctly recovered causal forcing of process Y by process X under most cases considered for all three plot lengths that we tested, particularly when observation error and process noise were small (Fig. 2a). It also

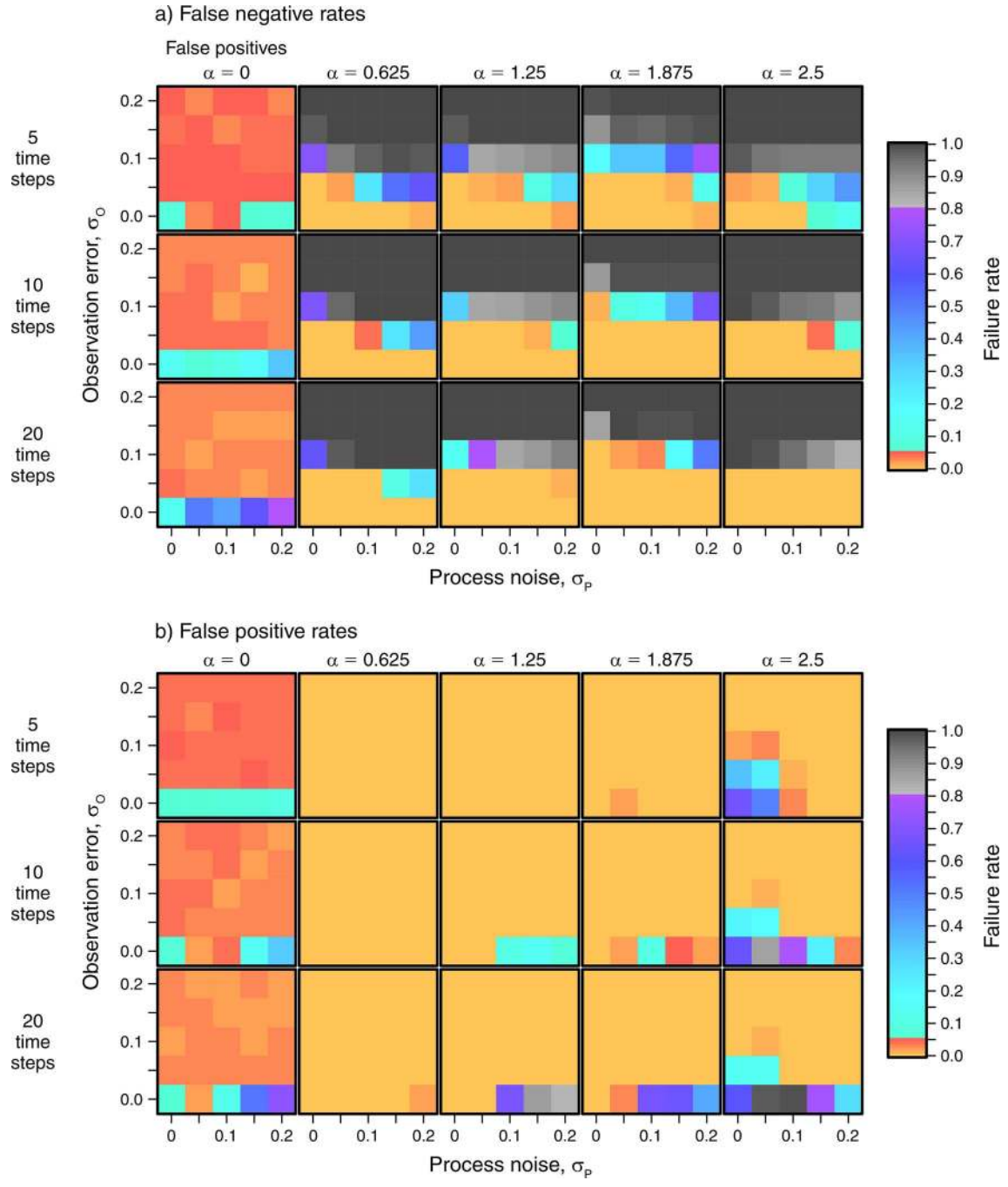


FIG. 2. Failure rates of multispatial CCM. Panel (a) shows the failure to detect an existing causal effect (i.e., false positives when $\alpha = 0$, false negatives otherwise), while panel (b) shows the false detection of a nonexistent effect (i.e., false positives). Colors indicate failure rate, the proportion of simulations for which the algorithm gave the incorrect result, with orange indicating 0–5% failures, blue indicating 5–80%, and gray/black indicating 80–100%. The horizontal axis in all panels shows increasing rates of process noise (σ_P) at five evenly spaced levels between 0 and 0.2. The vertical axis in all panels shows increasing rates of observation error (σ_O) at five evenly spaced levels between 0 and 0.2. From left to right, panels vary in forcing strength (α) of Y by X at five evenly spaced levels between 0 and 2.5. From top to bottom, panels vary in number of time steps included in the simulation (100 plots with 5, 10, and 20 observation time steps each).

performed well for both real-world examples, correctly detecting the presumed direction of causal association for both systems (Fig. 1b, c).

When no observation error or process noise were present, the test correctly indicated causal relations in almost every instance. In most cases where σ_O (strength of observation error) was less than or equal to 0.05, failures to detect causal signals were below 5% (and were generally around 0%), though for very high observation error ($\sigma_O = 0.2$), causal signals were almost never detected. Detection strength also declined in most cases with process noise, though the effect was not as large as for observation error. In general, detection rates remained higher even with process noise when time series were long (maximum library length, 20) and for moderately strong forcing ($\alpha = 1.875$). Performance for bidirectional simulations was similar (Appendix: Fig. A16).

The multispatial algorithm correctly indicated no forcing of process X by process Y (i.e., in the wrong direction) in most cases, as well (Fig. 2b). For weak forcing ($\alpha = 0.625$), error rates were around 0% for all levels of observation error and process noise. For systems with stronger forcing, false detection rates rose when observation error was not present, particularly when process noise was high, reaching 60–100% for the case with the strongest forcing ($\alpha = 2.5$). When no forcing was present ($\alpha = 0$), the test correctly indicated no causal forcing in most cases. In all cases with any observation error ($\sigma_O > 0$), failure rates were below 5%. Without observation error or process noise, failure rates were somewhat higher, but remained below 20%. The test performed worst when there was no observation error, but process noise was strong and showed failure rates as high as 80% in both directions, particularly for long libraries.

Discussion

Multispatial CCM appears to perform well for moderately short and very short time series (20, 10, and 5 observations per plot) when there is no observation error or process noise and when forcing is weak. Without error or noise, false detection rates (type I error) were less than 5% for all forcing strengths considered, except for the strongest forcing strength tested. Failures to detect causal links (type II error) were also around 0% when process noise and observation error were absent. These results indicate that multispatial CCM performs best under the same sort of circumstances that classical CCM does: low observation error and process noise relative to forcing strength (Sugihara et al. 2012). Encouragingly, errors only increased moderately as observational periods decreased, suggesting that short time series can be analyzed without a large decrease in test performance. Note that shorter observational periods in our simulations also lead to shorter L (i.e., total number of observations), as we retained the same number of spatial replicates among all tests. Some of the decrease in algorithm performance could therefore be alleviated by increasing spatial replication to compensate for decreasing observational period.

The test is conservative when observation error and process noise are present, and thus, failures to detect causal links in systems where large observation error is suspected should not necessarily be taken as evidence for a lack of causal relationship. When observation error and process noise are present, there appears to be a trade-off between type I and type II errors. Increased observation error decreases correct detections of causality, but also decreases spurious detections. Process noise slightly decreased the rate of detection of causal signals, but only when forcing was strong and had mixed effects on false detection rates. Our results suggest a steep decline in algorithm performance around $\sigma_O = 0.1$, which represents roughly an 11.4% difference between true and observed values in our system.

The test is anticonservative when there is strong forcing, particularly in the absence of error and noise, and thus, detection of apparently bidirectional causality in systems where strong forcing is suspected might instead simply be unidirectional causality. These detection failures are demonstrative of a failure of our convergence detection method, not necessarily of the CCM algorithm itself and are likely a result of synchrony. As processes become tightly coupled, each tends to become a good estimators of the others' dynamics even though causation is unidirectional (Sugihara et al. 2012). Signals that are confounded by synchrony are characterized by a sharp rise in ρ over the first few time steps, followed by a long flat plateau (see Fig. 3e in Sugihara et al. 2012), which is therefore a useful diagnostic that can warn of confounding. When synchrony is the product of an exogenous variable (e.g., joint response to temperature), it can often be minimized by taking first differences of the cross-correlated variables, where the preceding observation is subtracted from each observation prior to analysis (Granger and Newbold 1974). Where it cannot be eliminated (i.e., very strong forcing), synchrony may preclude data from CCM analysis.

The performance of multispatial CCM, based on a simple simulation of competing species, does indicate that algorithm performance is sensitive to properties of the dynamical systems under investigation. We chose this functional form because of its computational simplicity and because it makes the direction and strength of interactions, process noise, and observation error easy to interpret and quantify. However, a variety of other dynamical forms were considered by Sugihara et al. (2012) with very little variation in algorithm performance for classical CCM. Encouragingly, the algorithm performs well for a number of real-world cases, suggesting that at least some real systems meet the assumptions required for CCM analysis. In particular, synchrony is rarer than in the simple simulated system

that we consider. Nevertheless, where the assumptions of multispatial CCM are not met or where manifolds are sparsely sampled (e.g., a system with an exogenous time trend or nonstationary processes), the method is unlikely to work. In the Appendix, we include five examples that our diagnostic tests show do not meet all the required assumptions and show how these problems lead to method failures (in addition to five examples where the method performs well).

For a wide range of forcing strengths, and for moderate levels of observation error and process noise, multispatial CCM performs well in detecting causality, even for time series as short as five observations per plot. Our findings suggest that multispatial CCM is a useful analytical tool for empirical studies and can help inform questions about causal links in systems where experimental studies are not possible or available or where the results from experiments are difficult to interpret. Based on our experience with the technique and the results that we present here, we suggest that multispatial CCM can be applied as a useful analytical method, but caution that results based on the convergence detection algorithm that we present here are often very conservative. Further study with other dynamical systems will be important for determining the general applicability of multispatial CCM and for identifying convergence detection algorithms that are less conservative for a wide variety of applications. We look forward to future examples of multispatial CCM applications and to corresponding with parties that are interested in applying it to novel systems.

Acknowledgments

We thank R. Barnes, S. Binder, M. Burgess, M. Kosmala, C. Lehman, C. Perretti, K. Thompson, and P. Wragg for insightful discussions about this project. We also thank B. Inouye for helping us restructure our original manuscript, and T.J. Hefley and an anonymous reviewer for helpful comments, particularly their suggestions to add real-world analysis examples. Computing time for A. T. Clark was supported through NSF/XSEDE and the University of Minnesota Supercomputing Institute. Data collection was supported by the NSF LTER program, including DEB-0620652 and DEB-1234162, and by the Cedar Creek Ecosystem Science Reserve and the University of Minnesota. A. T. Clark, E. Deyle, and H. Ye were supported by the NSF GRFP. E. Deyle was supported by the EPA STAR Fellowship. G. Sugihara and H. Ye were supported by NSF DEB-1020372. G. Sugihara was supported by NSF-NOAA CAMEO program grant NA08OAR4320894. Additional support to G. Sugihara was provided by the Sugihara Family Trust, the Deutsche Bank-Jameson Complexity Studies Fund, and the McQuown Chair in Natural Sciences.

Literature Cited

Detto, M., A. Molini, G. Katul, P. Stoy, S. Palmroth, and D. Baldocchi. 2012. Causality and persistence in ecological systems: a nonparametric spectral Granger causality approach. American Naturalist 179:524–535.

Deyle, E. R., M. Fogarty, C. Hsieh, L. Kaufman, A. D. MacCall, S. B. Munch, C. T. Perretti, H. Ye, and G. Sugihara. 2013. Predicting climate effects on Pacific sardine. Proceedings of the National Academy of Sciences USA 110: 6430–35.

Deyle, E. R., and G. Sugihara. 2011. Generalized theorems for nonlinear state space reconstruction. PLoS ONE 6(3):e18295.

Fargione, E. J., and D. Tilman. 2005. Diversity decreases invasion via both sampling and complementarity effects. Ecology Letters 8:604–11.

Granger, C. W. J., and P. Newbold. 1974. Spurious regressions in econometrics. Journal of Econometrics 2:111–120.

Hsieh, C., C. Anderson, and G. Sugihara. 2008. Extending nonlinear analysis to short ecological time series. American Naturalist 171:71–80.

Perretti, C. T., S. B. Munch, and G. Sugihara. 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proceedings of the National Academy of Sciences USA 110:5253–5257.

R Core Development Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org

Schaffer, W. M. 1985. Order and chaos in ecological systems. Ecology 66:93–106.

Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time-series. Philosophical Transactions of the Royal Society 348:477–495.

Sugihara, G., and R. M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344:734–741.

Sugihara, G., R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, and S. Munch. 2012. Detecting causality in complex ecosystems. Science 338:496–500.

Supplemental Material

Ecological Archives

The Appendix and Supplement are available online: http://dx.doi.org/10.1890/14-1479.1.sm

Appendix A: Supplemental examples of multispatial CCM analyses for: Spatial "convergent cross mapping" to detect causal relationships from short time-series

Adam Clark *et al.*
Department of Ecology, Evolution, and Behavior
University of Minnesota, Twin Cities
St. Paul, MN 55108, USA
adam.tclark@gmail.com

December 23, 2014

Abstract

This Appendix contains detailed information for the real world examples provided in the text, along with some additional examples. The first section describes how to interpret results and diagnostic statistics (**Diagnostics and interpretation**). The second section includes information about the data and methods for each real world example (**Data and methods**). The third section presents the real world examples and interprets relevant results and diagnostic statistics (**Real world examples**). The fourth and final section shows simulation methods and results for a system with bi-directional causality (**Bi-directional causality**).

1 Diagnostics and interpretation

Here, we show how to interpret the results from the same simulated system described in the text in Figure 1, and in the example files for the R package multispatialCCM. In this system, the dynamics of process X influence the

dynamics of process Y, but the reverse is not true. Before we can commence with running the actual CCM test, we first need to find the best embedding dimension for the system (as described in step (1) in the methods section in the main manuscript), and determine whether the system has sufficient information for analysis (step (2) in the methods section) (Figure A1a and A1b respectively). Once the diagnostic requirements have been met, we can move on to implementing the bootstrapped CCM algorithm (steps 3-4 in the methods section of the main manuscript). In order to interpret the results, we also need to ensure that we have completed sufficient bootstrapped iterations to generate stable estimates of the relationship between ρ and library length L (Figures A1c-d).

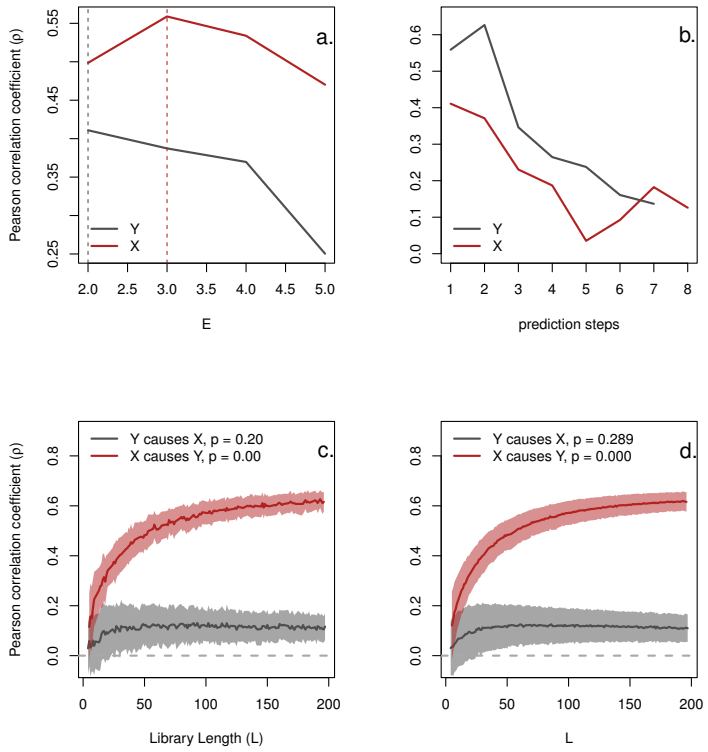


Figure A1: Results from the same simulated example from Figure 1a in the main text. Pearson Correlation coefficient ρ describes the ability of the algorithm to predict system dynamics. Shows test for best embedding dimension E (a), fit vs. prediction time step (b), and CCM results based on 100 bootstrapped iterations (c) and 1000 bootstrapped iterations (d). Solid lines show mean, and shaded region shows +/- 1 standard deviation. See main text for details.

Embedding dimension When choosing the best embedding dimension, we are looking for the dimension corresponding to the highest predictive power for one time step into the future. Ideally, this will either be at the lowest dimension tested (as for process X here, with $E = 3$, Figure A1a), or at an intermediate dimension representing a "hump" in predictive abilities between higher and lower dimensionalities (as for process Y here, with $E = 2$, Figure A1a). However, it is also important not to over-fit the model, and in

some cases it may be prudent to choose a smaller embedding dimension that has moderately lower predictive power than a higher dimensional model. We will discuss this further in the "Real world examples" section.

Prediction steps We conduct this test for two reasons. First, we wish to ensure that we have reasonable predictive ability for short time steps, or else there is likely not enough information (or too much stochastic noise) present in the data for CCM to work. Second, the predictive power should drop as the length of the prediction time step increases. Both of these conditions are met for the simulated data (Figure A1b). If this is not the case, then the system is probably purely linear, and CCM should not be used (see manuscript for other suggestions). Particularly bad is a system where predictive ability decreases, and then increases with prediction time step (a "U-shaped" ρ vs. time plot), as this suggests either some form of periodicity in system dynamics, or more likely, that data are not uniformly sampled, such that only a few observations are being predicted for longer prediction intervals. We will discuss this further in the "Real world examples" section. It should be noted that this is a "necessary but insufficient" test for non-linearity, since a linear system dominated by stochastic noise could still show the same response. In cases dominated by stochastic noise, CCM should (correctly) return "no causal link in either direction", since increasing information about the system will not increase predictive power.

Bootstrap iterations We increase the number of bootstrapped iterations for two reasons. First, it increases the precision of p-value estimates, as the lower detection limit is determined by 1/(number of iterations). This is not so important for the tests that we present here, as we only need to show $p<0.05$ for significance tests. Second, and more importantly, we need to increase iteration number in order to reduce Monte Carlo stochasticity (i.e. significantly different results from the same simulation parameters). The "best practice" for testing this is to increase the number of iterations until the mean and standard deviation estimates for ρ stabilize. Here, we show results for 100 and 1000 bootstraps (Figures A1c and A1d respectively). Though 1000 replicates provides a smoother estimate, there is no significant change in the distribution of rho estimates between the two iterations, so we chose to use only 100 iterations for our simulated performance tests in the manuscript. For real data, more iterations are generally required.

2 Data and methods

All data and detailed methods for the real world examples that we present in the main text and below are available on the Cedar Creek LTER data webpage at
`www.cbs.umn.edu/explore/field-stations/cedarcreek/research/data`.
Here, we give brief descriptions of each dataset that we use, and discuss existing findings and knowledge about the direction of causality for those systems. The experiment numbers that we reference can be used to look up each data set in the Cedar Creek webpage. For convenience, we also include ".csv" files that pre-collate relevant data tables and insert gaps in the time series between plot observations. These files can be used without further modification to run the analyses that we present here.

E001: Nitrogen addition experiment This experiment was established in 1982 to test how long-term nitrogen addition would alter plant species composition in an old field community. For full methods, see [8]. We analyze data from three fields (A, B, and C) and four nitrogen treatment levels, which we have grouped into "low" (0 $g\mathrm{NH_4NO_3}yr^{-1}$, with and without micronutrients), and "high" (25 and 40 $g\mathrm{NH_4NO_3}yr^{-1}$) treatments. Each field contained 12 plots in each treatment category. Above-ground species-level biomass in one field (C) was sampled annually from 1982-2011, while the others were sampled annually from 1982-2004. Because of some missing data, we retained 11 "plots" with 30 sequential measurements each, and 24 "plots" with 23 measurements each for the analyses presented here.

We focus on the dynamics of two of the most common species in the experiment, the cool-season, early-successional grass *Elymus (Agropyron) repens*, and the warm-season, late-successional grass *Schizachyrium scoparium*. In the absence of nitrogen fertilization, *A. repens* typically arrives early in succession, but is replaced by *S. scoparium* in the first 30 years of succession (see Figure 1 in [9]), because *S. scoparium* is a superior competitor for nitrogen [16]. However, in the presence of high levels of nitrogen addition, *A. repens* tends to re-colonize plots and out-compete *S. scoparium*, either for light or through indirect effects of leaf litter accumulation [3]. Fields A, B, and C were abandoned around 1968, 1957, and 1934 respectively. Thus, in "low" addition plots, we would expect to find that *A. repens* should be

disappearing or absent in all fields, and that its dynamics should be forced by those of *S. scoparium*. Conversely, in "high" addition plots, we expect that *S. scoparium* should be declining through time, and that its dynamics should be forced by *A. repens*.

Considerations for E001 analysis

- System is non-stationary (Figures A4, A6)
- Accurate prediction requires high embedding dimensions (Figures A3, A5)
- "U-shaped" prediction strength vs. interval relationship (Figures A3, A5)
- Conclusion: CCM should not be applied to these data, results are suspect

E026: Competition plots on a soil gradient This experiment was established in 1986 to test resource reduction and competitive outcome for monoculture and two-species prairie plant mixtures grown across a soil nitrogen fertility gradient. Full methods are available in [16]. Here, we focus on two soil fertility levels ("high" and "high + NH_4NO_3 fertilizer") in monoculture plots of the cool-season, early-successional grass *Agrostis scabra*. Aboveground biomass data for both species and leaf litter was collected annually from subsets of plots between 1986 and 1993. For *A. scabra* grown in "high" treatments, this included 12 plots with 8 sequential samples, 6 with 7, 3 with 4, and 9 with 3. For "high + NH_4NO_3 fertilizer" treatments, this included 4 plots with 8 sequential observations, 2 with 7, 1 with 4, and 3 with 3.

When grown on rich soils, *A. scabra* produces copious amounts of leaf litter. Because plots in this experiment were never burned, populations grown on very fertile soils (or soils with added fertilizer) show strong oscillations, possibly following chaotic dynamics, and ultimately crash, likely because they smother themselves in their own leaf litter [14]. Thus, we expect to find that *A. scabra* forces litter dynamics in less rich soils, but that it should also be forced by litter dynamics for very fertile soils.

Considerations for E026 analysis

- Different soil treatments follow different dynamics (Figure A7)
- Suspect ρ vs. prediction steps relationship for leaf litter (Figure A9a)
- Insufficient data for the case with added fertilizer (Figure A8b)
- Conclusion: Analysis in Figure A8a is likely legitimate, analysis in Figure A8b requires more data

E054: Plant biomass in old fields E054 is a subset of the long-term observational E014 study of old field succession at Cedar Creek. Full methods for E014 can be found in [2]. E054 includes annual species-level above-ground biomass samples taken from 1988-2011, taken at four subplots in each of 15 fields. Fields were abandoned between 1927 and 1998. Mean successional dynamics follow those described in [9]. Because of incomplete sampling and staggered abandonment of fields across years, we include data for 45 plots with 24 sequential samples, 2 with 23, 3 with 22, 4 with 20, 1 with 15, 4 with 11, 1 with 8, and 2 with 3.

Though there are many potential ecological questions to test in this system, we use these data to test a rather simple hypothesis about precipitation. Using total summer annual precipitation data (June - August), we test the extent to which *A. repens* and *S. scoparium* are forced by (or force) precipitation patterns at Cedar Creek. Because *A. repens* is a cool-season drought-intolerant species, its dynamics should strongly depend on water availability, and it should therefore be forced by precipitation. *S. scoparium* is a warm-season drought-tolerant species, and should therefore be more resistant to water stress, and its dynamics should not be as strongly forced by precipitation [11]. In both cases, precipitation should be an exogenous process, and its dynamics should not be forced by those of either species.

Considerations for E054 analysis

- Embedding dimension for both *A. repens* and *S. scoparium* should be reduced below the "best-fitting" E to preserve sample size and avoid over-fitting (Figure A10a-b)
- Conclusion: Both analyses in Figure A11 are likely legitimate

E120: "Big Biodiversity" experiment This experiment is the longest-running randomized test for the effects of plant diversity on ecosystem functions. Plots were established in 1994 and planted with 1, 2, 4, 8, or 16 species, and have since then been sampled annually for above-ground plant biomass. Full methods are described in [12]. The most well-known result from the experiment is that planted species number strongly, positively influences above-ground biomass production. However, because the diversity treatments are fixed, rather than dynamical variables, they do not lend themselves to CCM analysis. Instead, we focus on three other published results from the experiment: soil nitrate effects on invasion by non-planted species, biomass effects on soil nitrate, and biomass effects on insect abundance and diversity.

A number of studies in E120 have found significant increases in insect diversity as a function of increased planted species richness [4, 5]. A posited cause of this is that increased plant diversity increases above-ground biomass, which in turn increases the foraging space and habitat structure available to insect species. Interestingly, results do not agree on the effects of above-ground biomass on insect abundance. In one case [5], diversity was found to have no significant influence on insect abundance, whereas another study [4] found significant effects of both planted diversity and above-ground biomass on insect abundance. Here, we test for the causal relationships between above-ground plant biomass and both insect species richness and abundance. Across all diversity levels, we included 162 plots with 5 observations each, 136 with 4, and 24 with 3.

Though soil nitrate was not sampled as frequently as above-ground biomass, most plots were measured for 7 sequential years between 1996 and 2002. Empirical and theoretical results show that soil nitrate levels should be reduced in high diversity mixtures compared to low diversity mixtures because of more complete utilization of niche space [12, 13]. Additionally, increased biomass is associated with decreases in soil nitrate levels, even in monocultures [15]. Consequently, we expect that above-ground biomass dynamics should influence soil nitrate dynamics. However, because species are hypothesized to maintain soil nitrate at a relatively constant level regardless of their own biomass, classical models of resource competition [6, 7] suggest that above-ground biomass should not be influenced by soil nitrate dynamics. Across all

diversity levels, we included 132 plots with 7 sequential observations, 4 with 5, and 41 with 4.

High diversity has long been associated with decreases in invasion success. Though there is much debate about this relationship in natural systems, decreased invasion by non-planted species as a function of increased planted species richness has been described by a number of studies in E120 [1, 4]. A posited mechanism for this is soil nitrate: increased diversity leads to decreased soil nitrate, which in turn reduces invader success [10]. Based on diagnostic plots of system dynamics (details in the "Real world examples" section), we combined planted diversity treatments into "low diversity" (1-2 species), "intermediate diversity" (4-8 species), and "high diversity" (16 species) for this analysis. The "low diversity" treatments had 7 sequential samples in 57 plots, 5 in 4, and 4 in 11. The "intermediate diversity" treatments had 7 in 45, and 4 in 13. The "high diversity" treatments had 7 in 27, and 4 in 7.

Considerations for E120 analysis

- Time series for soil nitrate and insects are too short to test prediction power vs. interval
- Only some combinations of diversity treatments combine to form tractable manifolds for multispatial CCM analysis
- Analyses in Figure A12b and A14a show decrease in predictive ability with library length, suggesting that plots combined in the analysis may be too dissimilar
- Conclusion: Analyses in Figures A12a and A14b are likely legitimate. Analyses in Figure A12b and Figure A14a are suspect.

3 Walk-through of real world examples

For the following examples, we used 1,000 bootstrapped iterations of the multispatial CCM algorithm, after testing for both 100 and 1,000 iterations and finding no significant difference in the mean or distribution of ρ. In each case, we walk through the diagnostic statistics that we used to validate the

CCM analysis, and discuss whether or not the results of the analyses are to be believed. We hope that this section demonstrates that naively applying the multispatial CCM algorithm without checking for continuity of manifolds across plots, proper embedding dimensions, and predictive abilities will rarely yield meaningful results. Just as with any other analysis technique, making sure to meet the assumptions of the method is half the battle.

E001: Nitrogen addition experiment Based solely on the relationship between L and ρ, results for this CCM test suggest that *S. scoparium* dynamics force *A. repens* dynamics in non-fertilized plots (Figure A2a), while neither process forces the other in fertilized plots (Figure A2b). However, both of these analyses have a number of problems that we can pick up with a few diagnostic tests, and both of these results should probably not be believed.

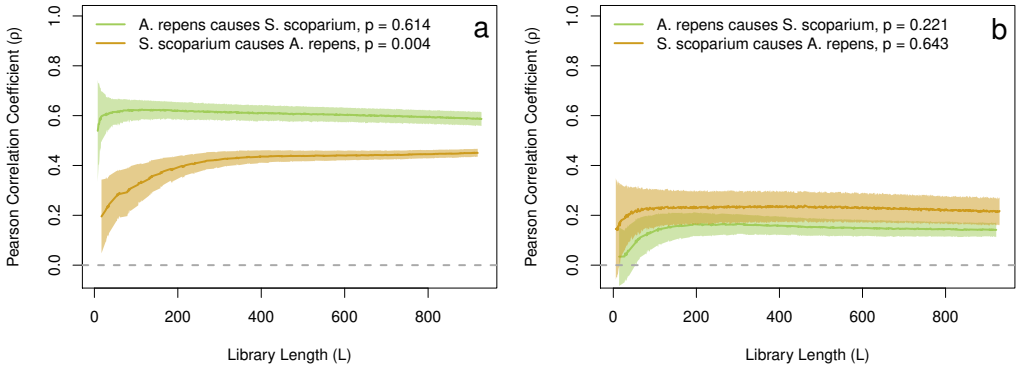


Figure A2: Test of causal forcing between *Agropyron repens* and *Schizachyrium scoparium* above-ground biomass dynamics in (a) unfertilized control plots, and (b) fertilized plots receiving 25 or 40 $g\mathrm{NH_4NO_3}yr^{-1}$.

First, let us consider the diagnostic plots for the non-fertilized case (Figure A3). For *A. repens*, the embedding dimension plot looks alright, with a relatively high predictive power achieved with 6 embedding dimensions. *S. scoparium* is somewhat more problematic because such a high embedding dimension is required to achieve moderate predictive power ($E_{max} = 14$). Even more problematic is the increase in predictive power with prediction interval. Both plots suggest that we can better predict plot dynamics 15 years into the future than we can 10 years into the future, which suggests either cyclical

patterns in the data, or a temporal trend in the data. Separating data by field (not shown) does not alleviate this problem.

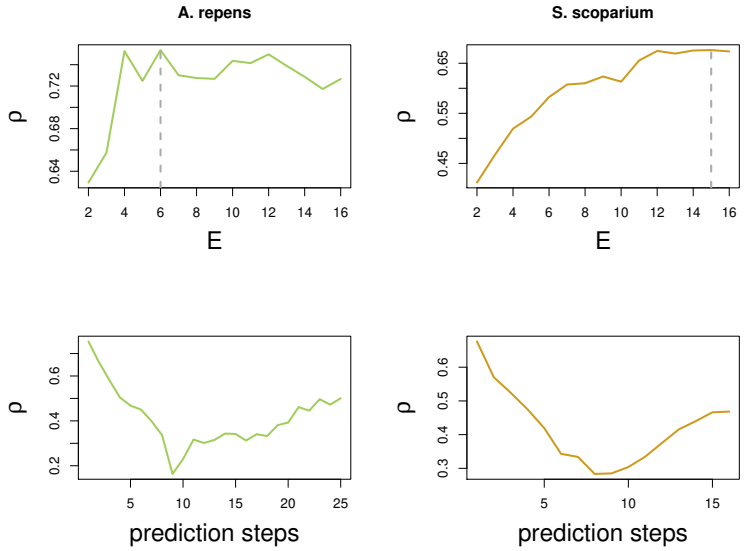
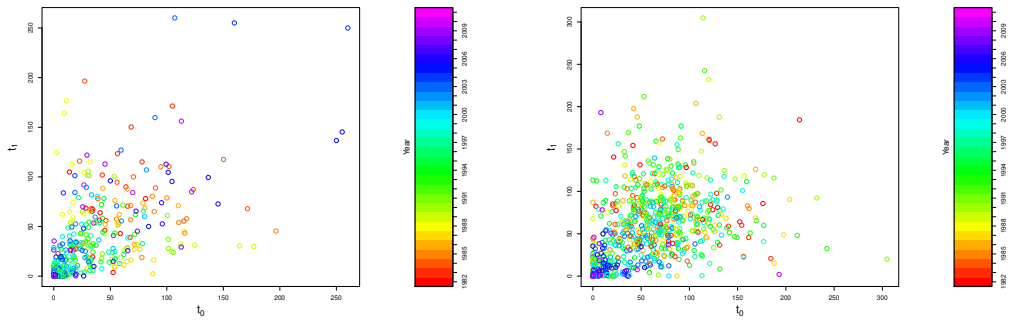


Figure A3: Diagnostic plots for test in Figure A2a (unfertilized control plots), showing relationship between predictive power ρ and embedding dimension E, or length of prediction interval.

To investigate, we can plot the dynamics of each process in two lagged dimensions (i.e. *A. repens* population this year, vs. next year, Figure A4).



(a) *A. repens* (b) *S. scoparium*

Figure A4: Lagged population dynamics for unfertilized plots in two dimensions.

This reveals a potential cause of the problem – both dynamics have a slight temporal trend. *A. repens* is abundant only early and late in the time series (in the early 1980's, and again in the 2000's). Similarly, *S. scoparium* is relatively rare late in the time series (after about 2000). Consequently, dynamics are much easier to predict across long intervals, as long intervals always predict population sizes late in the time series. The problem is even more apparent in the fertilized plots. Here, we find similar patterns in the diagnostic plots (Figure A5), and an even more rapid decline in *S. scoparium* (Figure A6).

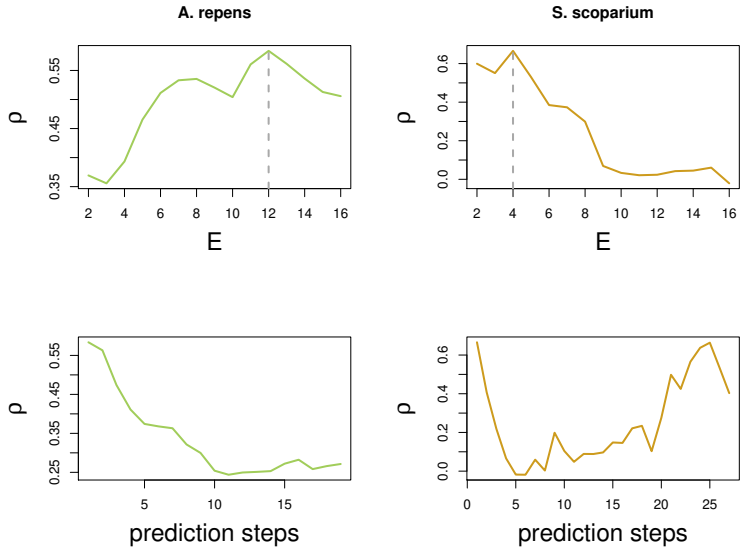
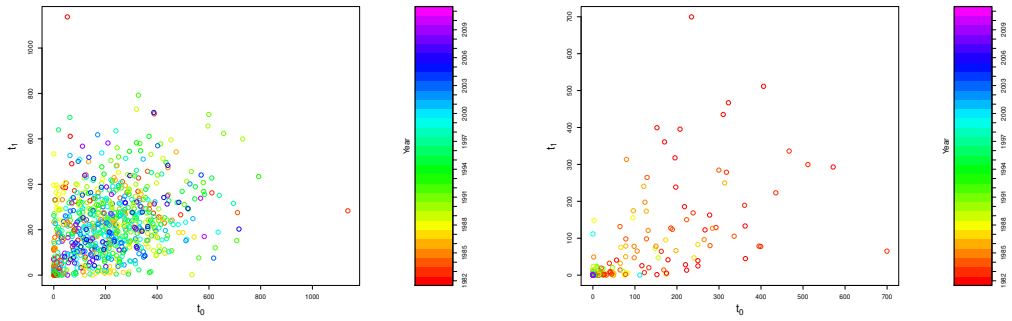


Figure A5: Diagnostic plots for test in Figure A2b (fertilized plots), showing relationship between predictive power ρ and embedding dimension E, or length of prediction interval.



(a) *A. repens* (b) *S. scoparium*

Figure A6: Lagged population dynamics for fertilized plots in two dimensions.

In cases where the time trend is caused by an exogenous variable (e.g. climate change), this problem might be alleviated by detrending the data.

However, in this case the dynamics are likely driven by simple interactions between the variables that we are testing. For *A. repens*, the temporal trend means that we have very low sampling density across the manifold (i.e. we have many samples along the time trend, but we do not have a lot of samples for any particular location along the time trend). For *S. scoparium*, the systems both collapse to a population size near zero, leaving minimal meaningful dynamics for the algorithm to test. Ultimately, our diagnostics suggest that these data should not be analyzed using CCM.

E026: Competition plots on a soil gradient In order to properly analyze this system, we first need to determine which plots can be included together as part of a single manifold. To do this, we again plot the dynamics of *A. scabra* in two lagged dimensions (Figure A7).

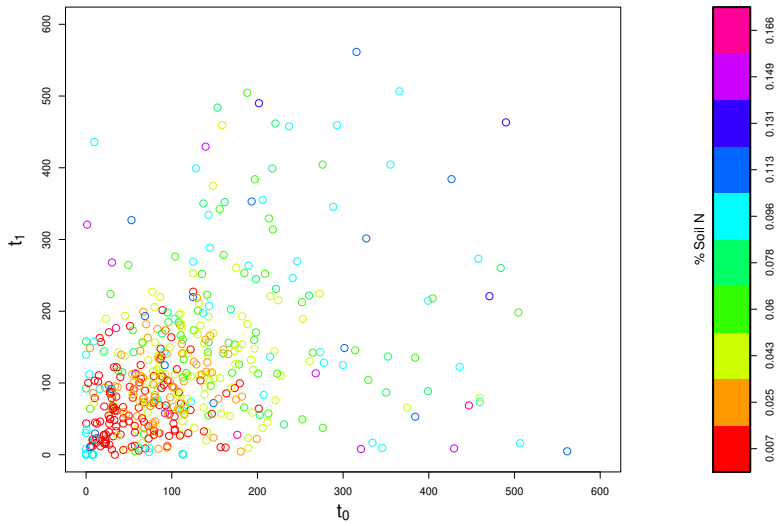


Figure A7: Lagged population dynamics *Agrostis scabra* in two dimensions.

This reveals substantially different dynamics among plots depending on total soil nitrogen, suggesting that we should analyze each soil mixture separately. Furthermore, because plots differ greatly in their above-ground biomass, we do not standardize the time series, as this could distort the manifold.

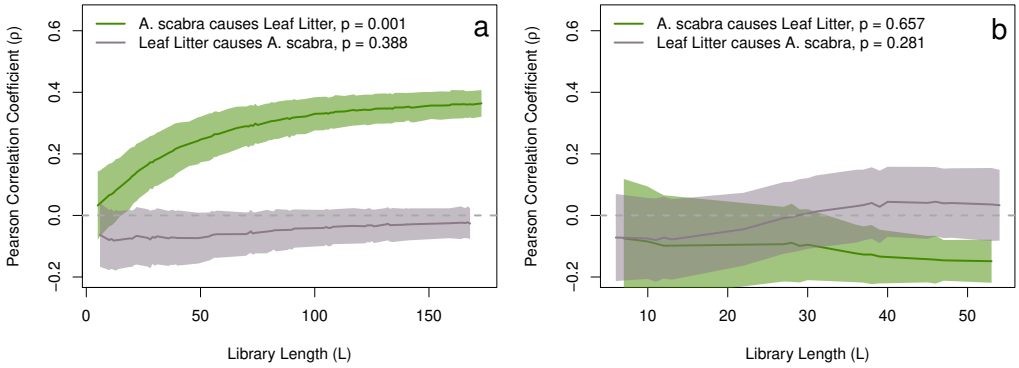
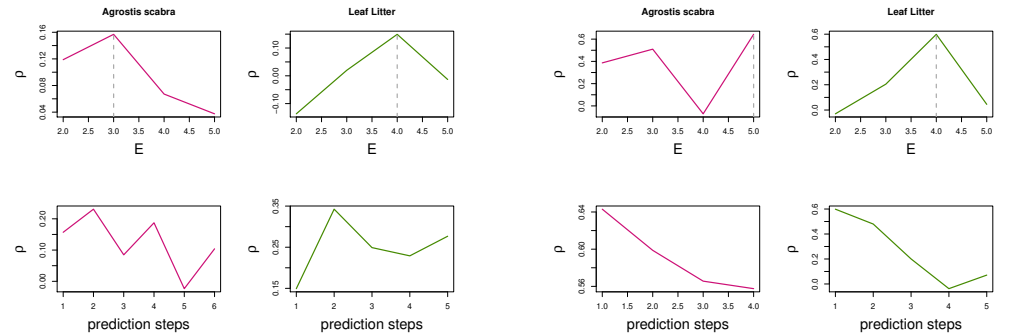


Figure A8: Test of causal forcing between *Agrostis scabra* above-ground biomass and Leaf Litter biomass dynamics in (a) fertile soil, and (b) fertile soil with added NH_4NO_3.

Based on the same diagnostic tests as we used above, we find two nitrogen treatments with sufficiently similar dynamics among plots for us to apply mutlispatial CCM. These tests suggest that in fertile soil, litter dynamics are forced by *A. scabra* dynamics (Figure A8a), but not the other way around. When additional fertilizer is added to plots, the causal direction appears to reverse, with *A. scabra* becoming forced by litter (Figure A8b). However, this second signal is not significant, likely because there are fewer plots with added fertilizer (note the shorter L).



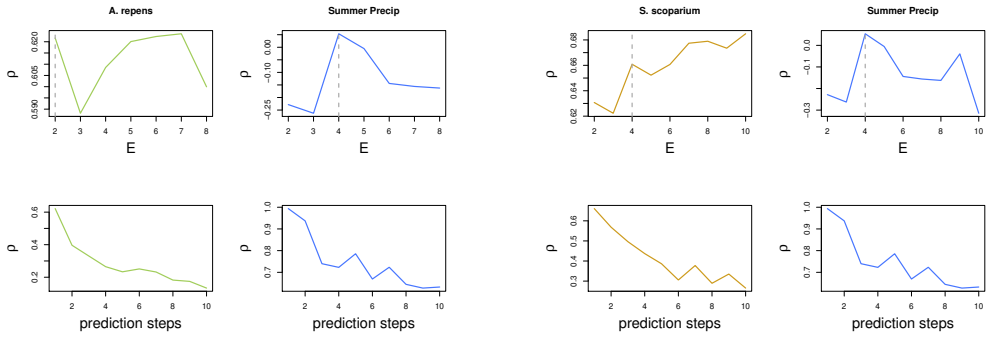
(a) fertile soil (b) fertile soil + fertilizer

Figure A9: Diagnostic plots for Figure A8.

Note that in the diagnostic plot for leaf litter in fertile soil (Figure A9a),

ρ follows a somewhat ambiguous trend as a function of prediction distance. Similarly, it is not clear that we have found the "best" embedding dimension for *A. scabra* in fertile soils with added fertilizer (Figure A9b), as the best fit occurs at the highest embedding dimension that we can test. The predictive power is also somewhat low across all diagnostic tests. While these are not ideal results for the diagnostic tests, they are not as egregious as the results for E001, and the resulting CCM tests show sensible patterns.

E054: Plant biomass in old fields This test shows some of the most easily interpretable causal results that we present here. They also show the importance of choosing a sensible embedding dimension. Though it is usually prudent to choose E that maximizes predictive ability, for both *A. repens* and *S. scoparium* we chose somewhat smaller embedding dimensions with predictive powers that are comparable to, but slightly smaller than, the "best fitting" E (Figure A10, using $E = 2$ rather than $E = 7$ and $E = 4$ rather than $E = 10$ respectively). We do this both to prevent over-fitting the model, and to retain a longer time series, as increasing E necessarily reduces the maximum library length that we can test.



(a) *A. repens* (b) *S. scoparium*

Figure A10: Diagnostic plots for Figure A11.

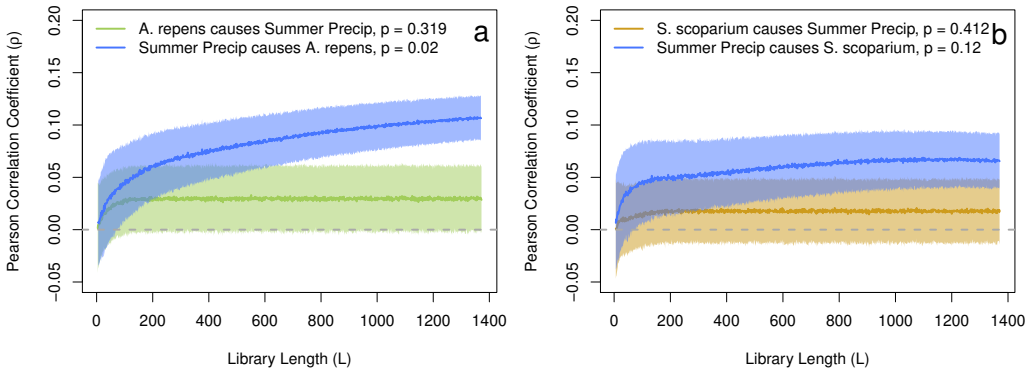


Figure A11: Test of causal forcing between total summer precipitation and population dynamics for (a) *Agropyron repens* and (b) *Schizachyrium scoparium*.

Sensibly, we find that plant population dynamics do not influence precipitation dynamics for either species. For *A. repens*, the more drought-sensitive of the two species, we find significant causal forcing by annual summer precipitation, as expected (Figure A11a). Note that this is the test presented in Figure 1c in the main text. For the more drought-tolerant *S. scoparium*, we find a trend that suggests forcing by summer precipitation, but it is not significant (Figure A11b). This is consistent with a weaker forcing effect, which would take somewhat more data to detect as significant.

E120: "Big Biodiversity" experiment Our results for insect abundance and richness partially match those for existing studies. For richness, we find a clear trend across all diversity treatments showing that above-ground plant biomass influences insect abundance dynamics, but not the other way around (Figure A12a). Insect richness and plant biomass do not appear to be significantly causally related, though the ρ vs. L relationship is somewhat strange (Figure A12b).

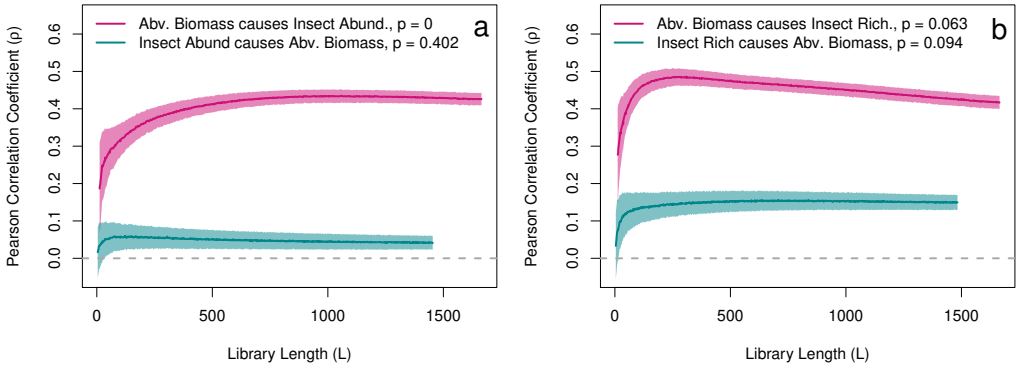
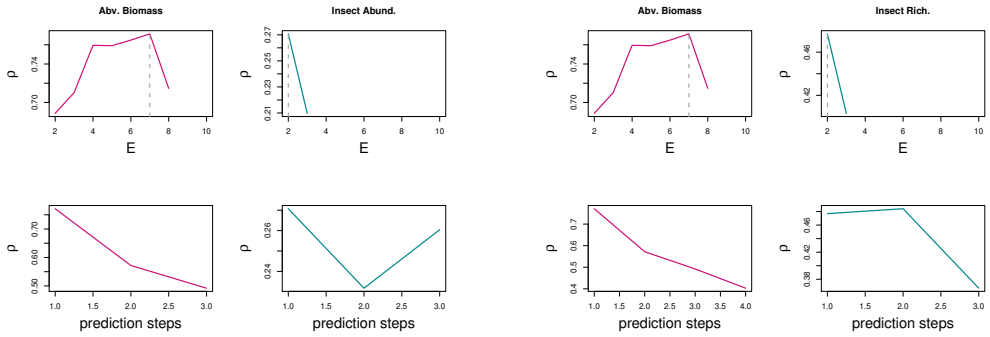


Figure A12: Test of causal forcing between total above-ground biomass and (a) insect abundance and (b) insect species richness.

Diagnostics are somewhat less useful in this case, because we don't have sufficiently long time series within each plot to show how predictive power varies with prediction interval for insect abundance or richness (Figure A13). Since we only have five sequential observations and an embedding dimension of 2, we can only predict three time steps into the future. While diagnostics for above-ground biomass seem okay, we cannot really determine whether the insect dynamics are appropriate for CCM analysis.

The CCM plot itself (Figure A12b) offers some possible information. Because ρ increases, but then decreases, with library length, our analysis shows that increasing the amount of information we have about the system (by adding more plots) decreases our ability to predict its dynamics. This suggests that the plots that we have combined for the analysis are not all well-predicted by the manifold we have constructed. However, separating plots by diversity treatment does not yield meaningful results either (analyses not shown). This suggests that insect richness dynamics differ among plots for some other reason. In the case of insect abundance, however, the CCM plot shows a more or less monotonic increase with library length, suggesting that the plots can be well-described by our estimated manifold. Thus, our results for insect abundance are probably believable, whereas our results for richness are likely confounded.



(a) Insect abundance

(b) Insect richness

Figure A13: Diagnostic plots for Figure A12.

Next, we consider causal relationships between soil nitrate dynamics and above-ground biomass (Figure A14a). There is a clear signal that above-ground biomass forces nitrate dynamics, matching our expectations. However, results for the effects of soil nitrate on above-ground biomass are somewhat more ambiguous. The p-value suggests that soil nitrate does not force biomass dynamics, but there is again an increase in ρ, followed by a decrease. Since the nitrate time series is relatively short (at most 7 sequential observations per plot) while the best embedding dimension is rather large ($E = 5$, or potentially higher), we cannot glean much information from the relationship between predictive power and number of prediction steps. Separating plots by diversity treatment does not provide a clearer pattern either (analysis not shown). Possibly, species differ sufficiently in their effect on soil nitrate that plots cannot be combined using our method.

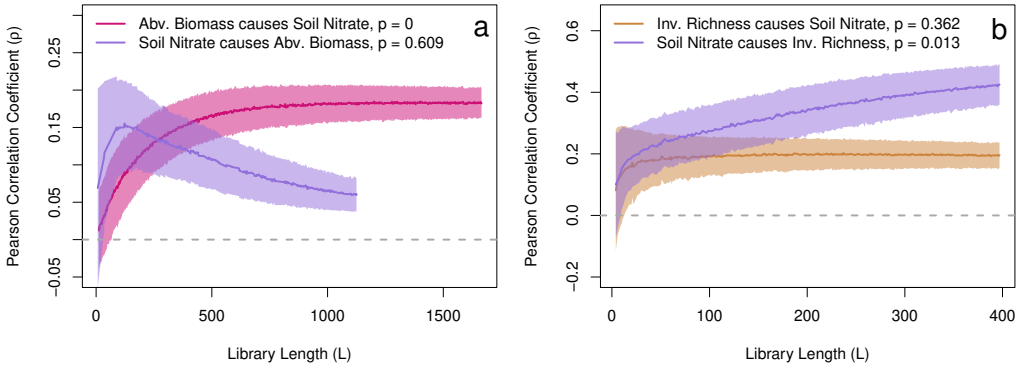
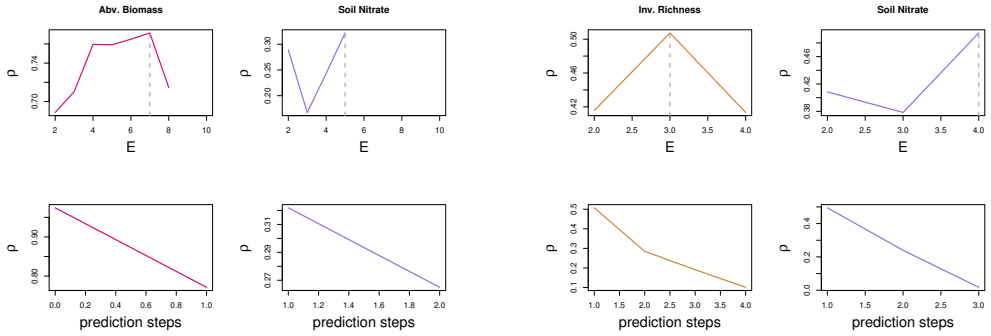


Figure A14: Test of causal forcing between soil nitrate and (a) total biomass or (b) species richness of invading (non-planted) plant species.



(a) *Above-ground biomass* (b) *Invader richness*

Figure A15: Diagnostic plots for Figure A14.

Results for the relationship between invader richness and soil nitrate are somewhat clearer. Here, we are able to find a subset of diversity treatments that appeared to share a single low-dimensional manifold (4 and 8 species treatments). While predictive power vs. prediction steps is still unclear in this case because of short time series (Figure A15b), the CCM analysis itself shows relatively clear signals (Figure A14b). In this case, the results suggest that invader richness does not force nitrate dynamics, whereas nitrate dynamics do influence plant invader species richness, matching our expectations from previous research. Note that this is the same as the analysis presented in Figure 1b in the main text.

4 Bi-directional causality

In order to test algorithm performance when bi-directional forcing exists, we repeated the simulations described in the manuscript, except with $\beta = 1.25$ (i.e. effect of process Y on process X, which was 0 for all other analyses). We include results from 160 simulations of the parameter ranges discussed in the main manuscript. Results show similar performance as reported in the uni-directional case for both directions of forcing (Figure A16).

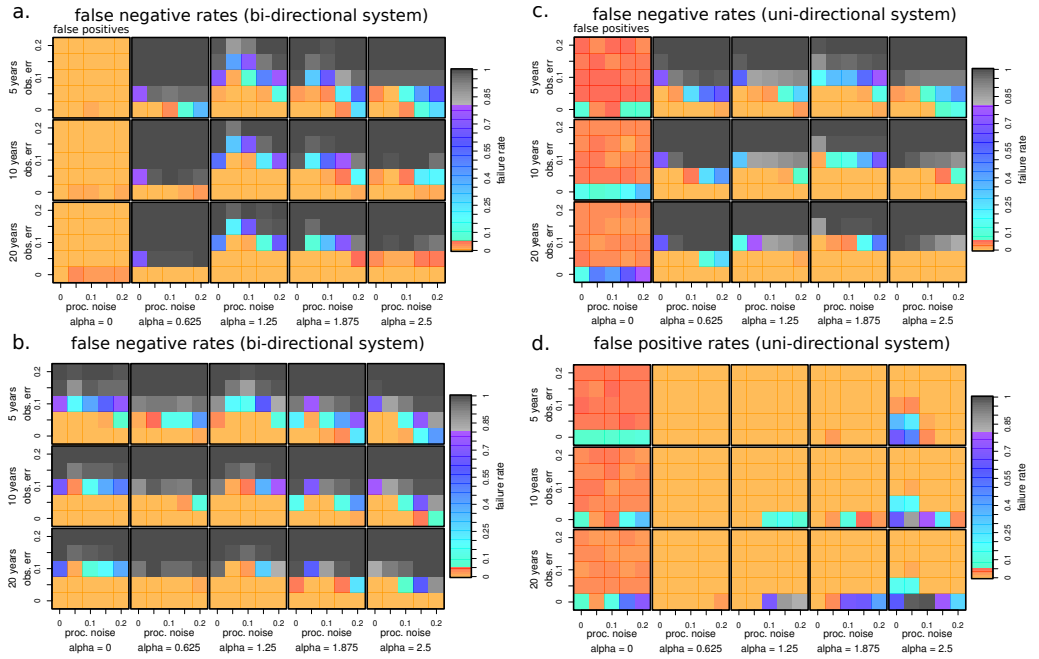


Figure A16: Multispatial CCM performance for bi-directional causal simulations for forcing of Y by X (a), and forcing of X by Y (b). We also include performance from the uni-directional test for forcing of Y by X (c), and forcing of X by Y (d) in the main text (where Y does not influece X) for comparison (Figure 2 in the main text). See Figure 2 caption in the main text for details.

References

[1] J. E. Fargione and D. Tilman. Diversity decreases invasion via both sampling and complementarity effects. *Ecology Letters*, 8(6), 2005. Times Cited: 105.

[2] R. S. Inouye, N. J. Huntly, D. Tilman, J. R. Tester, M. Stillwell, and K. C. Zinnel. OLD-FIELD SUCCESSION ON a MINNESOTA SAND PLAIN. *Ecology*, 68(1):12–26, 1987. Times Cited: 206.

[3] F Isbell, D Tilman, S Polasky, S Binder, and P Hawthorne. Low biodiversity state persists two decades after cessation of nutrient enrichment. *Ecology Letteres*, 2013.

[4] J.M.H. Knops, D. Tilman, N.M. Haddad, S. Naeem, C.E. Mitchell, J. Haarstad, M.E. Ritchie, K.M. Howe, P.B. Reich, E. Siemann, and J. Groth. Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecology Letters*, 2(5):286–293, September 1999.

[5] Evan Siemann, David Tilman, John Haarstad, and Mark Ritchie. Experimental tests of the dependence of arthropod diversity on plant diversity. *The American Naturalist*, 152(5):738–750, 1998.

[6] D. Tilman. RESOURCE COMPETITION BETWEEN PLANKTONIC ALGAE - EXPERIMENTAL AND THEORETICAL APPROACH. *Ecology*, 58(2):338–348, 1977. Times Cited: 538.

[7] D Tilman. *Resource Competition and Community Structure*. Princeton University Press, Princeton, NJ, 1982.

[8] D Tilman. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecological Monographs*, 57(3):189–214, 1987. Cited References Count:40|ECOLOGICAL SOC AMER|2010 MASSACHUSETTS AVE, NW, STE 400, WASHINGTON, DC 20036.

[9] D Tilman. Competition and biodiversity in spatially structured habitats. *Ecology*, 75(1):2–16, 1994. Cited References Count:77|ECOLOGICAL SOC AMER|2010 MASSACHUSETTS AVE, NW, STE 400, WASHINGTON, DC 20036.

[10] D Tilman. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10854–10861, 2004. Cited References Count:51|NATL ACAD SCIENCES|2101 CONSTITUTION AVE NW, WASHINGTON, DC 20418 USA.

[11] D. Tilman and John A. Downing. Biodiversity and stability in grasslands. *Nature*, 367(6461):363–365, 1994.

[12] D. Tilman, J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. The influence of functional diversity and composition on ecosystem processes. *Science*, 277(5330):1300–1302, 1997. Tilman, D Knops, J Wedin, D Reich, P Ritchie, M Siemann, E.

[13] D. Tilman, C. L. Lehman, and K. T. Thomson. Plant diversity and ecosystem productivity: Theoretical considerations. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5):1857–1861, 1997. Times Cited: 467.

[14] D. Tilman and D. Wedin. OSCILLATIONS AND CHAOS IN THE DYNAMICS OF a PERENNIAL GRASS. *Nature*, 353(6345):653–655, 1991. Times Cited: 105.

[15] David Tilman and David Wedin. Plant traits and resource reduction for five grasses growing on a nitrogen gradient. *Ecology*, 72(2):685, April 1991.

[16] David Wedin and David Tilman. Competition among grasses along a nitrogen gradient: Initial conditions and mechanisms of competition. *Ecological Monographs*, 63(2):199, May 1993.