The final publication is available at springerlink.com. DOI: 10.1007/s10651-012-0219-y

Determining patterns of variability in ecological communities: Time lag analysis revisited

Christian Kampichler[1,2] and Henk P. van der Jeugd[1]

Address:
[1] Vogeltrekstation – Dutch Centre for Avian Migration and Demography, NIOO-KNAW,
Postbus 50, 6700 AB Wageningen, The Netherlands
[2] Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias Biológicas,
Carretera Villahermosa-Cárdenas Km. 0.5 s/n, C.P. 86150 Villahermosa, Tabasco, Mexico

Corresponding author:
Christian Kampichler. Current address: Sovon Dutch Centre for Field Ornithology, Postbus 6521, 6503 GA Nijmegen, The Netherlands;
email: christian.kampichler@web.de; tel +31 24 7410 477

Author Contributions: HPJ and CK developed the idea, CK conceived and designed the simulations and analysed the data, HPJ and CK wrote the manuscript.

Abstract

All ecological communities experience change over time. One method to quantify temporal variation in the patterns of relative abundance of communities is time lag analysis (TLA). It uses a distance-based approach to study temporal community dynamics by regressing community dissimilarity over increasing time lags (one-unit lags, two-unit lags, three-unit lags). Here, we suggest some modifications to the method and revaluate its potential for detecting patterns of community change. We apply Hellinger distance based TLA to artificial data simulating communities with different levels of directional and stochastic dynamics and analyse their effects on the slope and its statistical significance. We conclude that statistical significance of the TLA slope (obtained by a Monte Carlo permutation procedure) is a valid criterion to discriminate between (i) communities with directional change in species composition, regardless whether it is caused by directional abundance change of the species or by stochastic change according to a Markov process, and (ii) communities that are composed of species with population sizes oscillating around a constant mean or communities whose species abundances are governed by a white noise process. TLA slopes range between 0.02 and 0.25, depending on the proportions of species with different dynamics; higher proportions of species with constant means imply shallower slopes; and higher proportions of species with stochastic dynamics or directional change imply steeper slopes. These values are broadly in line with TLA slopes from real world data. Caution must be exercised when TLA is used for the comparison of community time series with different lengths since the slope depends on time series length and tends to decrease non-linearly with it.

Key words

community change; Markov process; species composition; stochasticity; temporal dynamics

Introduction

All ecological communities are subjected to change over time (MacArthur and Wilson 1967; White et al. 2006; Magurran and Henderson 2010). Long-term datasets of ecological communities are the most important source of information on the temporal dynamics of species composition and patterns of relative abundance (Magurran et al. 2010). “Long-term”, however, is relative (Rull and Vegas-Vilarúbia 2011); while some studies analyse data collected over an exceptionally long period of time—for example, the Park Grass Experiment at Rothamsted, England (Silvertown et al. 2006)—the majority of datasets that are regarded as “long-term” by ecologists are comparably short and do not allow for the application of standard tools for time series analysis (Cowpertwait and Metcalfe 2009).

Time lag analysis (TLA) was introduced by Collins et al. (2000) as a method to quantify temporal variation in the patterns of relative abundance of communities. It applies a distance-based approach and is used to study temporal community dynamics by regressing community dissimilarity over increasing time lags (one-unit lags, two-unit lags, three-unit lags,...). To prevent the smaller number of data points of larger time lags from biasing the result, the time lags are square root transformed. Collins et al. (2000) mention three instances that can be distinguished by TLA. (1) When the slope of the regression line of dissimilarity of lag is significant and positive. In this case the community is undergoing directional change. (2) When the regression line is significant and negative. This indicates a convergent dynamics of the community, i.e., the community returns to an earlier state in the time series such as following perturbation or other cyclical behaviour. (3) When there is stochastic variation over time. This implies that the slope of the regression line is not significantly different from zero. Collins et al. (2000) further state that the slope of the regression and the coefficient of determination, R^2, can be used as a measure of signal versus noise. For example, a small but significant positive slope with a small R^2 would indicate slow directional change with high stochastic variation between sample intervals, whereas a steeper slope and a large R^2 would indicate a strong signal of directional change and less stochastic variation. Since its publication, TLA has gained popularity, and it has been applied to study the temporal dynamics of a variety of communities, for example, desert rodents (Thibault et al. 2004), soil microarthropods (Kampichler and Geissen 2005), and tide pool fishes (Pfister 2006).

The stochastic change as outlined above assumes that populations are governed by a white noise process, which means that the abundance of a species at a given time is completely independent of any previous state. Examples are communities in which species abundances oscillate around time-invariant means according to a normal distribution, or communities in which the abundance of each species corresponds to random values from a uniform distribution at every time step (which in fact also leads to stationary population means when the upper and lower limits of the uniform distribution are constant in time). In an analysis of 544 natural populations of 123 species, Inchausti and Halley (2002) showed, however, that in almost all cases population variability increased with observed timespan, which is not consistent with the assumption of a white noise population process. We therefore aimed to evaluate the behaviour of TLA with data based on the most simple stochastic model next to a white noise process which assumes that the abundance at any point in time is dependent only on its previous value (Williams et al. 2002, p. 188). Processes whose future process behaviour is influenced only by the present system state are known as Markov processes, and they are widely used for the modelling of biological populations (Meyn and Tweedie 1993, p. 5, Williams et al. 2002). We argue that it is highly improbable that a first-order Markov process, i.e. a state at time t+1 depends only on the state at time t, will move a community back to a position in variable space where it had been some time before. Thus, any stochastic change concerning the abundance of the species that constitute the community inevitably will veer it away from the original state and thus increase any distance measure. The terms “directional” and “stochastic” must be used with caution since they suggest that directional community change implies the action of an internal or external force that drives it from its original position in variable space. According to our reasoning, stochastic variation that can be described by a first-order Markov process also leads to “directional” change, for example, by ecological drift (Hubbell 2001), which is analogous to genetic drift caused by random mutation (Ricklefs 2003).

In this paper, we apply TLA to artificial data simulating communities with different levels of directional and stochastic dynamics and analyse their effects on the slope and its statistical significance. According to our hypothesis, we expect to detect significant community change when first-order Markov processes are involved. Finally, we draw conclusions on the potential of TLA for the study of real-world community time series.

Materials and Methods

Simulation of community time series

We simulated community change with communities having a species richness of 20 and a time-series length of 20 and 100 units. They were initialised according to a log-normal model of abundance distribution with a mean of log(N)=3 and a standard deviation of 2 and rounded to the closest integer. We assumed that all species within a community behaved independently from the others. We used species with three types of dynamics: species that fluctuate around a constant mean (“constant species“), species with stochastic dynamics (“stochastic species“), and species with a monotonously increasing or decreasing trend (“directional species“).

Abundance values of the constant species time series were drawn from a normal distribution with the initial abundance N_0 as the mean and the standard deviation s calculated as $N_{t-1} * v$ where v is a scaling factor ranging from 0.001 to 0.5 thus fixing the standard deviation in the range from 0.1% (species with small fluctuations) to 50% (species with large fluctuations) of the mean of the preceding time step (Fig.1a). For the stochastic species, we used the general model $N_t = f(N_{t-1}, \varepsilon_t)$ as a starting point where N is the population density at time t, ε_t represents environmental stochasticity, and f is a function that relates the density and environmental stochasticity to a population size at time $t + 1$ (Lundberg et al. 2000). We drew the abundance values N_t from a normal distribution with mean N_{t-1} and standard deviation calculated as above as $N_{t-1} * v$ with v ranging from

0.001 to 0.5 (Fig.1b). Thus the trajectories of the stochastic species through time form first-order Markov chains since the transition probabilities from N_{t-1} to N_t depend only on N_{t-1}, not on how N_{t-1} was reached (e.g., by an increase or a decrease from N_{t-2} to N_{t-1}) like in a correlated random walk (Meyn and Tweedie 1993; Williams et al. 2002). Stochastic species were allowed to go extinct and to re-enter the community. We applied a procedure similar to the random walk on a half-line (Meyn and Tweedie 1993) and permitted the species trajectory to include negative abundances during data generation. Prior to TLA these data were set to zero, i.e. the species were "absent" from the community at the corresponding points in time. Time series of the directional species were constructed in the same manner as for stochastic species, but for each species the changes were forced to be always either positive or negative (Fig.1c). Due to their directional character, species that went extinct in the time series could not re-enter the community. The minimum change between time steps was set to one. Finally, all of the time series values were rounded to the closest integer. For the sake of clarity in this paper we replace the term stochasticity, which is introduced into the time series by v, with *temporal variability* and thus avoid confusing it with the terms referring to the three different types of dynamics (constant, stochastic, directional).

We constructed communities that were exclusively composed of constant species (const100), stochastic species (stoch100) and directional species (dir100), as well as communities that were composed of 25 and 75%, 50 and 50% and 75 and 25% species of two given types. For example, community const50stoch50 was composed of 50% constant and 50% stochastic species. Among the directional species in a given community, one half was assigned an increasing trend, and the other half a decreasing one. For all species in a given community *s* had the identical value. Thus, we constructed communities with a range from very low to very high temporal variability. No attempt was made to simulate communities with a converging dynamic, that is, communities that return to a state of one of the early sample dates, which should yield a significant negative slope according to Collins et al. (2000).

Data transformation

In using Euclidean distance based on the absolute abundances (ED_{abs}) as a distance measure, any changes in species abundances in the same direction that do not change relative abundance patterns—e.g., all species increase their population size by 10%—will lead to increasing dissimilarity over time and yield significant TLA slopes. It is, thus, difficult to disentangle the abundance component and the compositional component. We assume that in most cases it is desirable to model changes in abundance and changes in composition separately, and in these cases distance measures other than ED_{abs} should be used. Furthermore, the comparison of temporal trends of communities with different numbers of individuals is hampered since higher numerical abundance leads to larger ED_{abs} between years and, thus, to steeper TLA slopes. Last but not least, ED_{abs} can cause the well-known species-abundance paradox: two sites having no species in common may be more similar than two sites sharing species but with different abundances (Legendre and Legendre 1998). Based on a preliminary evaluation of different data transformations to circumvent the undesired properties of ED_{abs} (Online Resource 1) we applied the Hellinger transformation $N'_{ij} = \sqrt{(N_{ij}/\sum N_{ij})}$ where N_{ij} is the population size of species *i* in year *j*, and ΣN_{ij} is the sum of individuals across all species in year *j* (Legendre and Gallagher 2001).

TLA based on Hellinger distance (HD), i.e., Euclidean distance of Hellinger transformed data, has the properties of (i) not being sensitive to changes in absolute abundance while patterns of relative abundance stay constant, (ii) making assemblages directly comparable independent of their species richness, and (iii) being sensitive also to rare species (Online Resource 1). Hellinger distance is not mandatory for TLA; depending on the research question other distance measures might be preferred for example when abundance effects are to be included or when more emphasis is to be given to dominant species (Online Resource 1).

Time lag analysis

Community change was replicated 1000 times for each combination of composition (const 100, stoch100, dir100, const25stoch75, const50stoch50, const75stoch25, const25dir75, const50dir50, const75dir25, stoch75dir25, stoch50dir50, stoch75dir25), temporal variability (v = 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5) and time series length (20 and 100), yielding a total of 12*13*2 = 312 000 simulation runs. For each simulated community we fitted the linear model HD = a + b*sqrt(lag). For a time series of length n there are $(n^2 – n)/2$ possible distance values. The time series of length 20 and 100 thus produce 190 (19 values for lag 1, 18 values for lag 2, …, 1 value for lag 19) and 4950 distance values, respectively. The inflated number of degrees of freedom and the lack of independence between the data points are problematic for the determination of the statistical significance of the slope. Following Thibault et al. (2004) we applied a Monte Carlo permutation procedure and (i) permuted the order of the year columns in the data matrix, (2) calculated the slope b for each permutation, and (3) compared the resulting distribution of slopes with the slope for the original data matrix by dividing the number of random slopes greater than the original TLA slope by the number of permutations. Doing this for all 312 000 simulated communities, however, would have increased the number of calculations to an unfeasible amount. We therefore limited significance testing to a 100-fold randomisation of 100 simulated communities for each combination of community composition, temporal variability and time series length, which still added up to 3 120 000 permutation runs. Slopes were regarded significant when the error probability P was equal to or < 0.05. The highest level of significance attainable with the applied Monte Carlo permutation procedure was $P < 0.01$, when all random slopes were lower than the original TLA slope. All simulations were performed with the R language and environment for statistical computing (R Development Core Team 2010).

Results

Community const100 did not show significant slopes (Figs. 2 and 3) whereas the slopes of stoch100 (Figs. 2 and 4) and directional100 (Figs. 3 and 4) were highly significant ($P < 0.01$) at any given level of temporal variability. Even if a small proportion of species in the communities containing constant species was stochastic or directional, p decreased rapidly and the slopes attained high significance ($P < 0.01$) in almost all cases (Figs. 2 and 3). For the mixture of constant and stochastic species with a temporal variability of $v \leq 0.05$ these slopes were very low ($b < 0.02$) but still highly significant ($P < 0.01$) (Fig. 2a-d). The only exception was community constant75stoch25 (composed of many constant species and few stochastic species) whose p varied between 0.25 and 0.45 (Fig. 2c). All communities composed of stochastic and directional species had highly significant

($P < 0.01$) slopes (Fig. 4).

Slopes became steeper with increasing temporal variability; at the highest levels of variability, however, slopes tended to decrease. This pattern was more pronounced for the longer (Fig. 2b, 3b, 4b) than for the shorter time series (Fig. 2a, 3a, 4a). Slopes were clearly dependent on time series length and were generally higher in the shorter time series, particularly in the communities with higher temporal variability (Figs. 2a-b, 3a-b and 4a-b).

Discussion

The results confirm our hypothesis that stochastic change other than a white noise process would lead to significant slopes when regressing community dissimilarity over increasing time lags. Collins et al. (2000) also presented simulated stochastic data and concluded that the resulting slope would be non-significant (Collins et al. 2000, Fig. 4). Their trajectories through time, however, were constructed in a different way, randomly choosing the abundance for each species at each point in time. For natural populations this rarely seems to be the case as has been shown by Inchausti and Halley (2002). The Markov chains of abundance of “stochastic species” used in our study represent a more realistic realisation of the time series of animal populations that have neither a stable mean (“constant species”) nor tend to approach an attractor of high or low abundance (“directional species”) but are governed only by random fluctuations (Williams et al. 2002). Consequentially the significance level of the slope does not allow the discrimination between communities with directional and stochastic change. Both processes—directional change of constituent species and stochastic change according to a Markov process—lead to directional changes in species composition and these are identified as such by TLA even when the changes between sampling dates are very small. Applied to real community data, TLA will yield significant slopes for communities characterised by directional and stochastic dynamics, and will discriminate them from the communities that are almost entirely composed of species with constant population sizes (Figs. 2 and 3) and communities whose species abundances are governed by a white noise process, such as in the simulations of stochastic dynamics by Collins et al. (2000).

Since significance will be achieved in almost any case, the slope itself remains as the most important measure to judge the dynamic of a community undergoing changes in species composition. When temporal variability is low, the slope rarely exceeds values of 0.05; in the communities composed of constant and stochastic species it even remains below 0.02. The maximum values attained for higher temporal variability fall between 0.15 and 0.25 (Fig. 2a, 3a, 4a); only community stoch75dir25 yielded a slope even larger than 0.25 (Fig. 4a). The range of slopes observed in the simulations correspond quite well to the empirical Hellinger-distance based TLA slopes of bird communities (Kampichler et al., in preparation). Bird communities from pristine ecosystems have been described as being remarkably stable (primeval temperate forests: Enemar et al. (2004); Wesołowski et al. (2010); Scandinavian alpine vegetation: Svensson (2006)). Their slopes range from 0.02 to 0.04 which would be consistent with the assumption that these communities are mainly composed of a mixture of constant species and stochastic or directional species with low temporal variability. Slopes from successional forests (deciduous forest on abandoned fields: Kendeigh (1982); spruce regrowth after clearcutting: Hall (1984)) with a considerable number of directional bird species (early species that later become locally extinct, species that enter the community later and continually increase their abundance) show slopes steeper than 0.1. A closer interpretation of the slopes seems to be hampered by the lack of a monotonous relationship between temporal variability and slope, as shown by the undulating curves in Figs. 3b and 4b. This is, however, due to the fact that in this simulation the directional species continue to increase or decrease their abundance throughout the entire time series which might be the case in short time series but cannot be assumed to be very realistic for long real-world time series. At high levels of temporal variability the decreasing directional species very rapidly become extinct while the remaining species continue increasing; relative abundance patterns change only slightly for the rest of the time series and thus lead to a lower TLA slope.

A more serious complication is that longer time series yield lower slopes than shorter ones, particularly when temporal variability is high (Fig. 2b, 3b, 4b), although the generation process of the time series of the constituent species is identical. To explore this hitherto unreported behaviour of TLA we simulated another 6000 communities—2000 each for the compositions const50stoch50, const50dir50 and stoch50dir50—for time series lengths (TSL) increasing from 10 to 200 in steps of 10 (ten replicates for each TSL, temporal variability set at $v = 0.1$) and regressed their slopes on TSL. TLA slopes vary considerably with the TSL; their relationship is quite complicated and cannot be approximated by a linear model but only with a segmented regression approach (Muggeo 2003), fitting separate line segments to different TSL intervals (Fig. 5a, 5c, 5e). Standardising TSL and thus regressing Hellinger distance on sqrt(lag)/max(sqrt(lag)) eliminates much of the nonlinearity of the relationship between slope and TSL (no segmented regression could be fitted). On the one hand, this would permit the direct comparison of community data with different TSL; on the other hand, comparability with conventional TLA studies becomes lost due to the changed slope (compare the slopes on the y-axes of Fig. 5a, c and e with Fig. 5b, d and f). Consequentially, caution has to be exercised when TLA is used for the comparison of communities where time series length differs.

There are a number of alternatives to distance-based time lag analysis, such as redundancy analysis based on principal coordinates of neighbourhood matrices (Borcard et al. 2004) or asymmetric eigenvector maps (Blanchet et al. 2011). These methods were developed for the analysis of spatial patterns but can easily be adopted for time series analysis (Angeler et al. 2009). These direct canonical ordination approaches conserve the taxonomic identity of species during the calculation of distance metrics and allow identification of the species' contributions to the patterns of temporal change. Thus it has been argued that they are superior to distance based methods (Angeler et al. 2009). We suspect that distance and ordination methods do not exclude each other but mutually complement the other. The merits of TLA compared to these powerful methods are its computational ease, its easy comprehensibility for an audience not experienced in interpreting ordination results, and the possibility of characterising and comparing the temporal dynamics of large numbers of communities with a single measure (the slope along with its significance level) without being drowned in masses of detailed information. We thus feel that further methodological improvements of TLA are desirable and necessary.

Acknowledgements

We owe thanks to John Plant for correcting the English. This is NIOO publication xxxx.

References

Angeler DG, Viedma O, Moreno JM (2009) Statistical performance and information content of time lag analysis and redundancy analysis in time series modelling. Ecology 90:3245-3257 doi: 10.1890/07-0391.1

Blanchet FG, Legendre P, Maranger R, Monti D, Pepin P (2011) Modelling the effect of directional spatial ecological processes at different scales. Oecologia 166:357-368 doi: 10.1007/s00442-010-1867-y

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826-1832 doi: 10.1890/03-3111

Collins SL, Micheli F, Hartt L (2000) A method to determine rates and patterns of variability in ecological communities. Oikos 91:285–293 doi: 10.1034/j.1600-0706.2000.910209.x

Cowpertwait PSP, Metcalfe AV (2009) Introductory Time Series with R. Springer, Heidelberg doi: 10.1007/978-0-387-88698-5

Enemar A, Sjöstrand B, Andersson G, von Proschwitz T (2004) The 37-year dynamics of a subalpine passerine bird community, with special emphasis on the influence of environmental temperature and *Epirrita autumnata* cycles. Ornis Svecica 14:63–106

Hall GA (1984) A long-term bird population study in an Appalachian spruce forest. Wilson Bulletin 96:228-240

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton

Inchausti P, Halley J (2002) The long-term temporal variability and spectral colour of animal populations. Evol Ecol Res 4:1033–1048

Kampichler C, Geissen V (2005) Temporal predictability of soil microarthropod communities in temperate forests. Pedobiologia 49:41–50 doi: 10.1016/j.pedobi.2004.07.011

Kendeigh SC (1982) Bird Populations in East Central Illinois: Fluctuations, Variations, and Development over a Half-Century. Illinois Biological Monographs 52, University of Illinois Press, Champaign.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271-280 doi: 10.1007/s004420100716

Legendre P, Legendre L (1998) Numerical Ecology, 2nd edn. Elsevier, Amsterdam

Lundberg P, Ranta E, Ripa J, Kaitala V (2000) Population variability in space and time. Trends Ecol Evol 15:460-464 doi: 10.1016/S0169-5347(00)01981-9

Magurran AE, Baillie SR, Buckland ST, Dick JMcP, Elston DA, Scott EM, Smith RI, Somerfield PJ, Watt AD (2010) Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends Ecol Evol 25:574-582 doi: 10.1016/j.tree.2010.06.016

Magurran AE, Henderson PA (2010) Temporal turnover and the maintenance of diversity in ecological assemblages. Phil Trans R Soc B 365:3611–3620 doi: 10.1098/rstb.2010.0285

McArthur RH, Wilson EO (1967) The Theory of Island Biogeography. Princeton University Press, Princeton

Meyn SP, Tweedie RL (1993) Markov Chains and Stochastic Stability. Springer, London

Muggeo VMR (2003) Estimating regression models with unknown break-points. Stat Med 22:3055–3071

Muggeo VMR (2008) segmented: an R package to fit regression models with broken-line relationships. R News 8 (1):20-25 URL http://cran.r-project.org/doc/Rnews/

Pfister CA (2006) Concordance between short–term experiments and long–term censuses in tide pool fishes. Ecology 87: 2905–2914 doi: 10.1890/0012-9658(2006)87[2905:CBSEAL]2.0.CO;2

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org

Ricklefs RE (2003) A comment on Hubbell's zero-sum ecological drift model. Oikos 100:185-192 doi: 10.1034/j.1600-0706.2003.12018.x

Rull V, Vegas-Vilarrúbia T (2011) What is long-term in ecology? Trends Ecol Evol 26:3-4 doi: 10.1016/j.tree.2010.10.002

Silvertown J, Poulton P, Johnston E, Edwards G, Heard M, Biss PM (2006) The Park Grass Experiment 1856-2006: its contribution to ecology. J Ecol 94:801–814 doi: 10.1111/j.1365-2745.2006.01145.x

Svensson S (2006) Species composition and population fluctuations of alpine bird communities during 38 years in the Scandinavian mountain range. Ornis Svecica 16:183–210

Thibault KM, White EP, Ernest SKM (2004) Temporal dynamics in the structure and composition of a desert rodent community. Ecology 85: 2649–2655. doi: 10.1890/04-0321

Wesołowski T, Mitrus C, Czeszczewik D, Rowiński P (2010) Breeding bird dynamics in a primeval temperate forest over thirty-five years: variation and stability in the changing world. Acta Ornithologica 45:209–232 doi: 10.3161/000164510X551354

White EP, Adler PB, Lauenroth WK, Gill RA, Greenberg D, Kaufman DM, Rassweiler A, Rusak JA, Smith MD, Steinbeck JR, Waide RB, Yao J (2006) A comparison of the species/time relationship across ecosystems and taxonomic groups. Oikos 112:185-195 doi: 10.1111/j.0030-1299.2006.14223.x

Williams BK, Nichols J, Conroy M (2002) Analysis and Management of Animal Populations. Academic Press, San Diego

Figure captions

Fig. 1. Sample time series of constant (A), stochastic (B) and directional (C) species with varying temporal variability, characterised by the scaling factor *v*. For ease of comparison all sample time series share the same initial value $N_0 = 50$. The scaling factor *v* used for constructing the time series ranges from 0.001 to 0.5 (see text for details). Note the different scaling of y-axes.

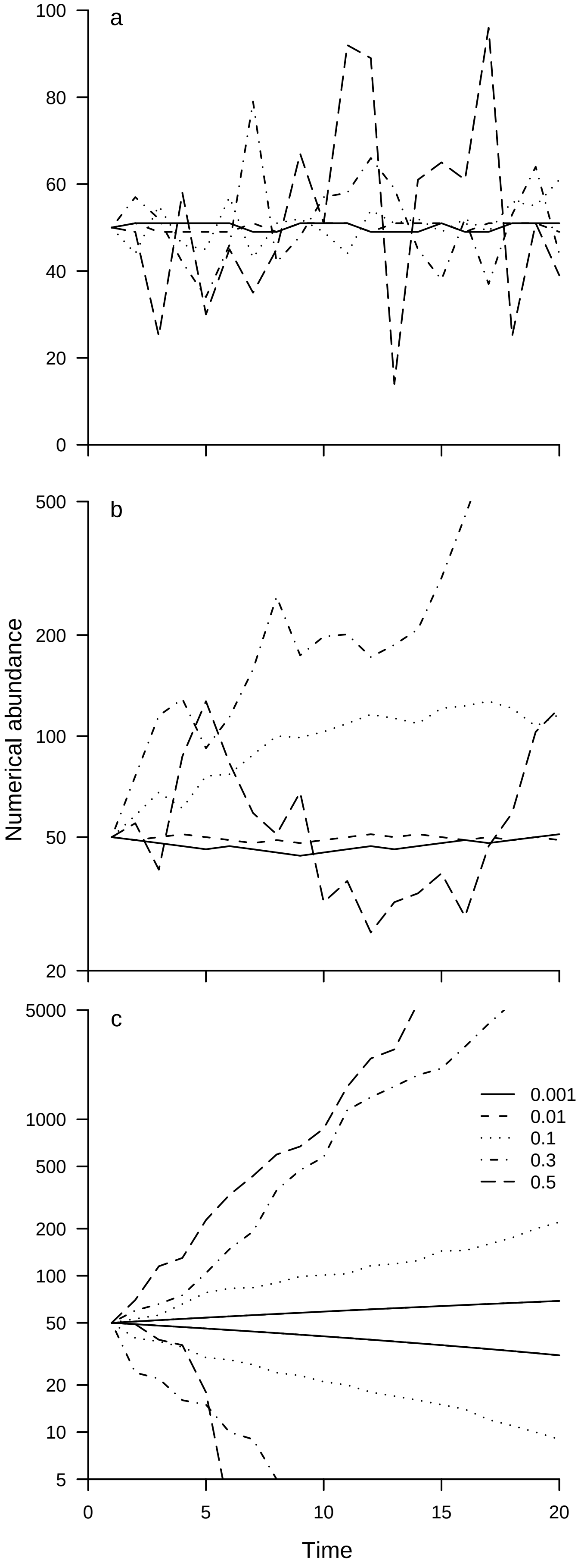
Fig. 2. Time lag analysis of simulated communities composed of constant and stochastic species in time series of a, c) length 20 and b, d) length 100 with different temporal variability, determined by the scaling factor *v* used in the generation of the species time series (see text for description). Reported are a, b) the slopes of the regression lines of Hellinger distance on square root of time lag and c, d) the error probability, *P*, as determined by a Monte Carlo permutation procedure.

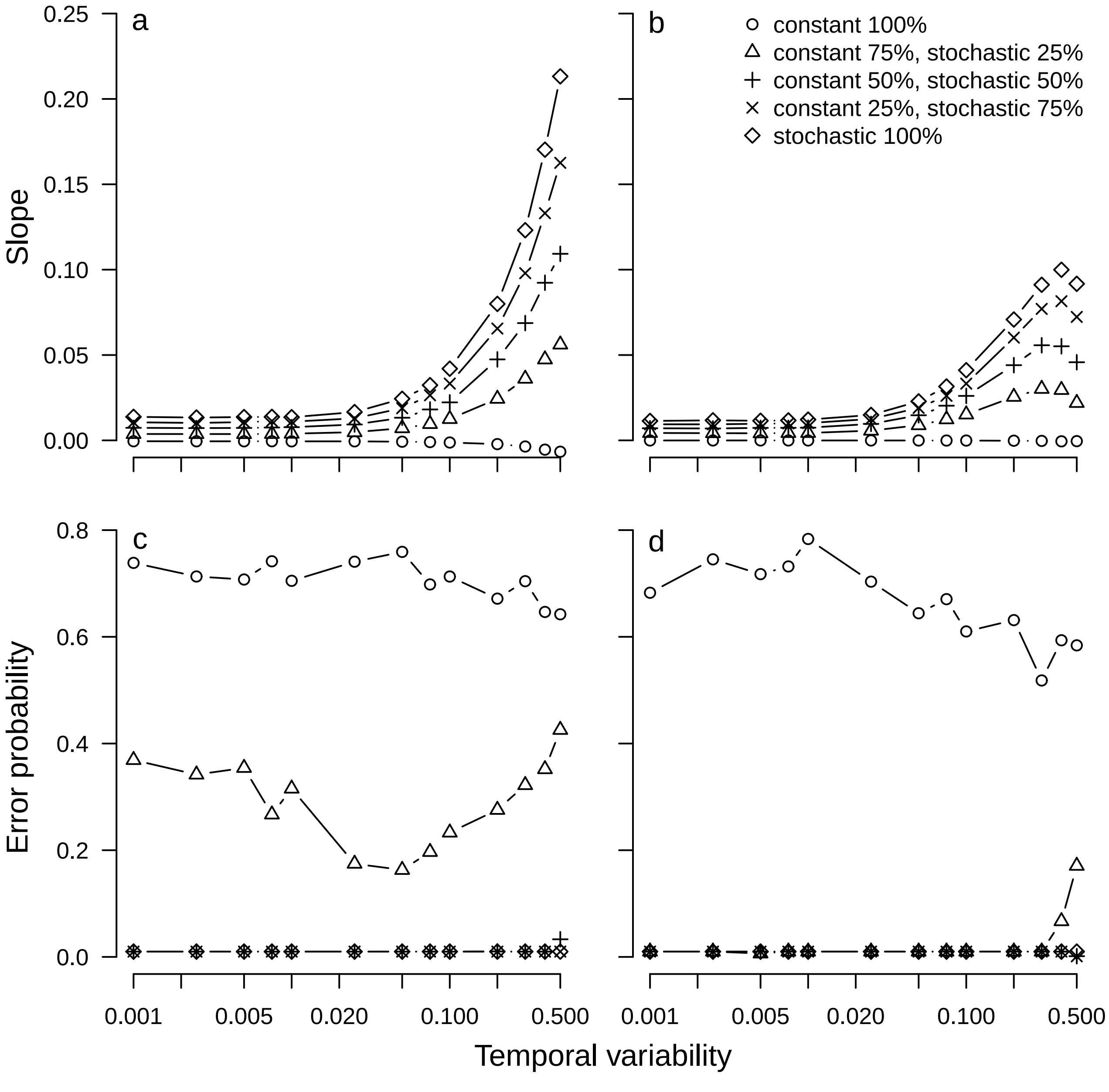
Fig. 3. Time lag analysis of simulated communities composed of constant and directional species in time series of a, c) length 20 and b, d) length 100 with different temporal variability, determined by the scaling factor *v* used in the generation of the species time series (see text for description). Reported are a, b) the slopes of the regression lines of Hellinger distance on the square root of time lag and c, d) the error probability, *P*, as determined by a Monte Carlo permutation procedure.

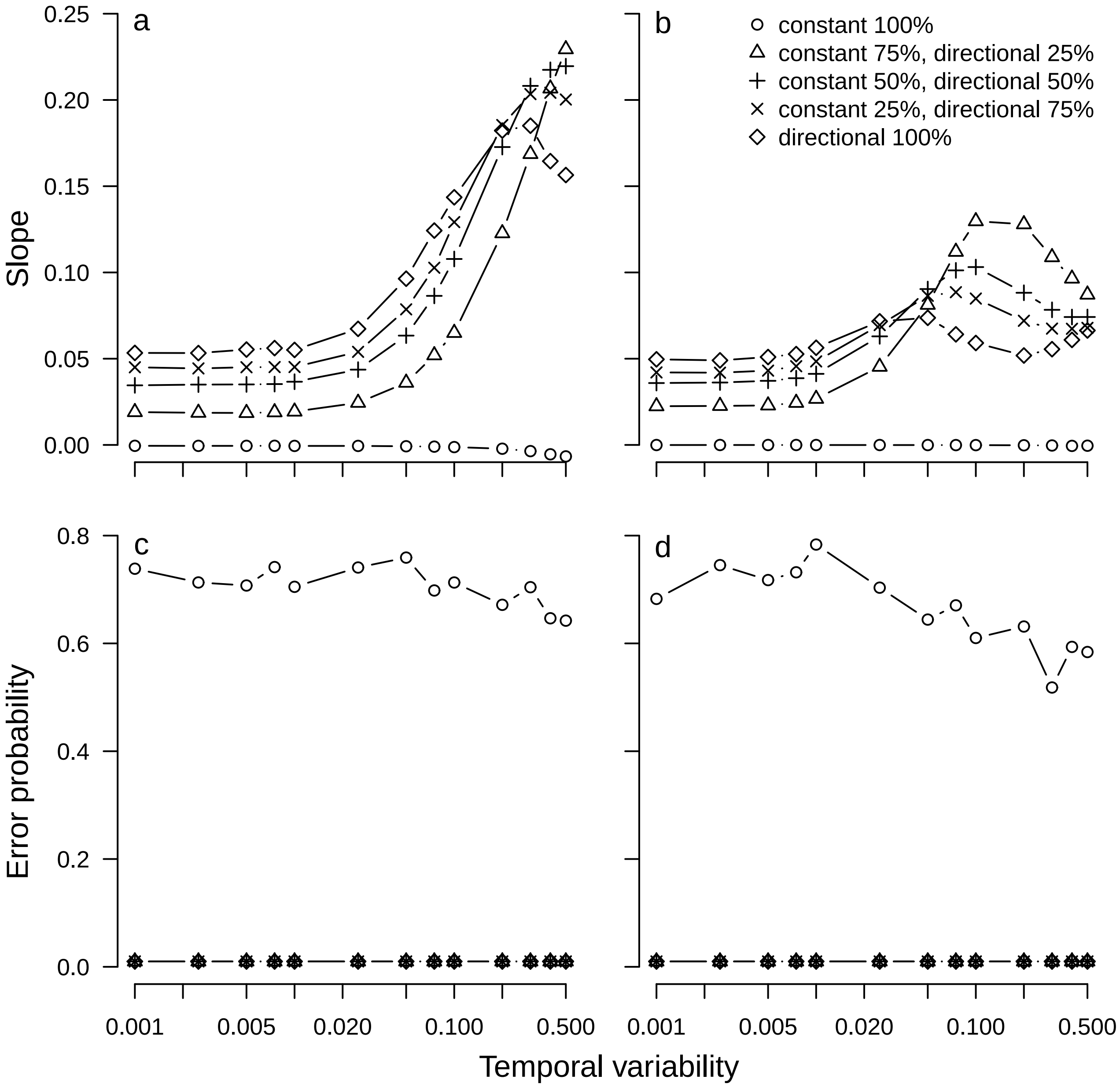
Fig. 4. Time lag analysis of simulated communities composed of stochastic and directional species in time series of a, c) length 20 and b, d) length 100 with different temporal

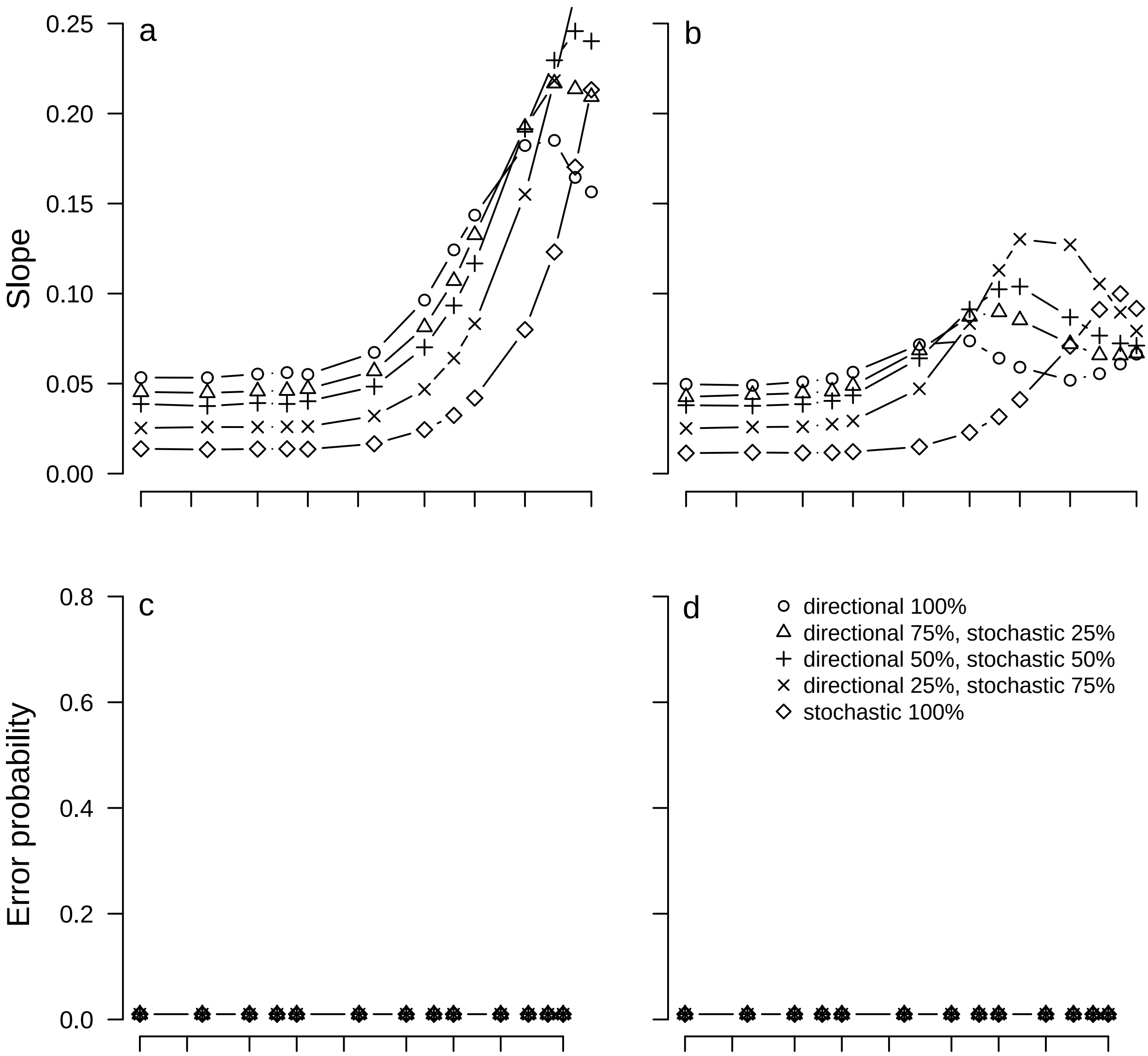
variability, determined by the scaling factor *v* used in the generation of the species time series (see text for description). Reported are a, b) the slopes of the regression lines of Hellinger distance on the square root of time lag and c, d) the error probability, *P*, as determined by a Monte Carlo permutation procedure.

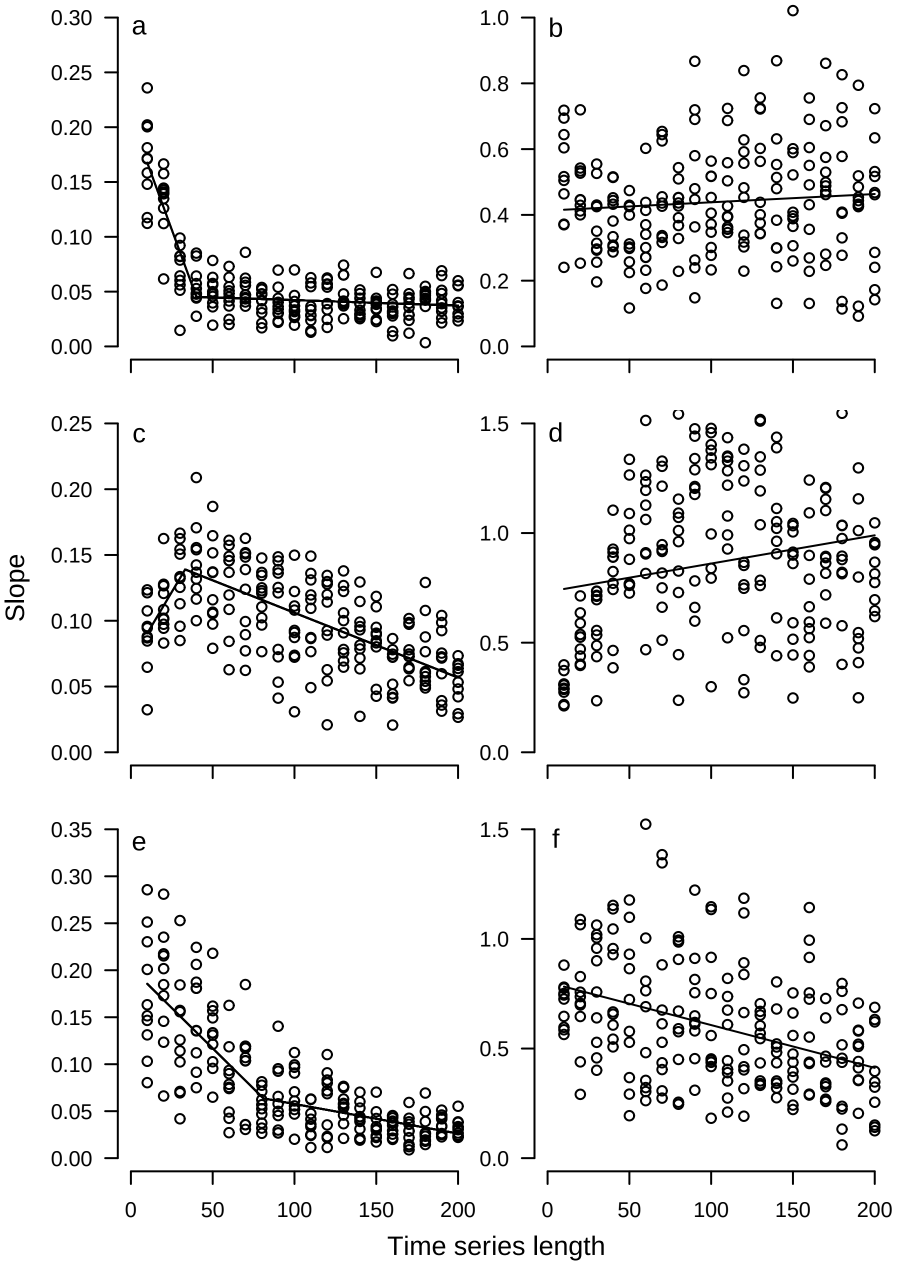
Fig. 5. Relationship between TLA slope for different time series lengths for communities composed of equal proportions of a, b) constant and stochastic species , c, d) directional and stochastic species, and e, f) constant and directional species (E, F). The scaling factor *v* used for determining temporal variability in the generation of the time series was set at 0.1 (see text for description). Panels on the left (a, c, e) are based on conventional TLA, panels on the right (b, d, f) are based on TLA with Hellinger distance regressed on sqrt(lag)/max(sqrt(lag)) instead on sqrt(lag). Break-points and slopes in panels a), c) and e) were determined by segmented regression (Muggeo 2003) using the package `segmented` (Muggeo 2008) for the R language and environment for statistical computing (R Development Core Team 2010). Regression lines in panels b), d) and f) are not significant ($P > 0.05$), significant at $P < 0.01$ and significant at $P < 0.001$, respectively.


a
b
c
Numerical abundance
Time
100
80
60
40
20
0
500
200
100
50
20
5000
1000
500
200
100
50
20
10
5
0
5
10
15
20
0.001
0.01
0.1
0.3
0.5


a
b
c
d
○ constant 100%
△ constant 75%, stochastic 25%
+ constant 50%, stochastic 50%
× constant 25%, stochastic 75%
◇ stochastic 100%
Slope
0.25
0.20
0.15
0.10
0.05
0.00
Error probability
0.8
0.6
0.4
0.2
0.0
0.001
0.005
0.020
0.100
0.500
Temporal variability


a
b
c
d
constant 100%
constant 75%, directional 25%
constant 50%, directional 50%
constant 25%, directional 75%
directional 100%
Slope
0.25
0.20
0.15
0.10
0.05
0.00
Error probability
0.8
0.6
0.4
0.2
0.0
0.001
0.005
0.020
0.100
0.500
Temporal variability


a
b
c
d
Slope
0.25
0.20
0.15
0.10
0.05
0.00
Error probability
0.8
0.6
0.4
0.2
0.0
0.001
0.005
0.020
0.100
0.500
Temporal variability
○ directional 100%
△ directional 75%, stochastic 25%
+ directional 50%, stochastic 50%
× directional 25%, stochastic 75%
◇ stochastic 100%


a
b
c
d
e
f
Slope
Time series length
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
0.8
0.6
0.4
0.2
0.0
0.25
0.20
0.15
0.10
0.05
0.00
1.5
1.0
0.5
0.0
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.5
1.0
0.5
0.0
0
50
100
150
200
0
50
100
150
200

Determining patterns of variability in ecological communities: Time lag analysis revisited
Environmental and Ecological Statistics
Christian Kampichler* and Henk van der Jeugd
* Sovon Dutch Centre for Field Ornithology, Postbus 6521, 6503 GA Nijmegen, The Netherlands; email: christian.kampichler@web.de

Electronic Supplemental Material 1

Evaluation of distance measures for time lag analysis

To illustrate the behaviour of time lag analysis (TLA) using different distance measures, we created nine simple artificial community datasets, each consisting of a time-series with length 10 on 10 (poor) or 20 (rich) species, respectively (Table S1). Datasets A and B are composed of species with increasing abundance (in A, linearly; in B, exponentially) but without changes in relative abundance patterns. In datasets C and D, one half of the species increases, the other half decreases (in C, linearly; in D, exponentially). Dataset E consists of four increasing and four decreasing species with low abundance and two highly abundant species that fluctuate randomly around a stable mean.

Table S1. Nine artificial datasets for the evaluation of the characteristics of TLA with different distance measures. Each dataset consists of a time-series of 10 observations on 10 (A-poor, B-poor, C-poor, D-poor, E) or 20 species (A-rich, B-rich, C-rich, D-rich), respectively. Column Time shows the abundance vector for each species within a given dataset. For example, dataset A-poor consists of 10 species, each of them increases in time from 1 to 10 individuals; dataset D-rich consists of 20 species, 10 of them increase from 1 to 512, 10 decrease from 512 to 1.

		Time									
Dataset	Species	1	2	3	4	5	6	7	8	9	10
A-poor	1-10	1	2	3	4	5	6	7	8	9	10
A-rich	1-20	1	2	3	4	5	6	7	8	9	10
B-poor	1-10	1	2	4	8	16	32	64	128	256	512
B-rich	1-20	1	2	4	8	16	32	64	128	256	512
C-poor	1-5	1	2	3	4	5	6	7	8	9	10
	6-10	10	9	8	7	6	5	4	3	2	1
C-rich	1-10	1	2	3	4	5	6	7	8	9	10
	11-20	10	9	8	7	6	5	4	3	2	1
D-poor	1-5	1	2	4	8	16	32	64	128	256	512
	6-10	512	256	128	64	32	16	8	4	2	1
D-rich	1-10	1	2	4	8	16	32	64	128	256	512
	11-20	512	256	128	64	32	16	8	4	2	1
E	1-4	1	2	4	8	16	32	64	128	256	512
	5-8	512	256	128	64	32	16	8	4	2	1
	9	47212	47513	51414	58865	51192	39076	48365	51522	46215	55409
	10	48836	52204	50082	48589	58390	46327	48560	43324	42555	52100

We calculated TLA using five different distance measures:

1. *Euclidean distance based on absolute abundances* (*EDabs*)

$$ED_{abs}=\sqrt{\sum_{i=1}^{S}\left(N_{ij}-N_{ik}\right)^2} \tag{1}$$

where N_{ij} and N_{ik} represent the abundance of species i in years j and k, and S is the number of species in the community.

2. *Euclidean distance based on relative abundances* (*EDrel*)

Here, before equation (1) is applied, the data are transformed according to

$$N_{ij}'=\frac{N_{ij}}{\sum_{i=1}^{S} N_{ij}} \tag{2}$$

where ΣN_{ij} is the total of individuals in year j across all species.

3. *Hellinger distance (HD)*

HD is Euclidean distance based on Hellinger-transformed abundance data (Legendre and Gallagher 2001). Thus, before applying equation (1), the data are transformed according to

$$N_{ij}'=\sqrt{\frac{N_{ij}}{\sum_{i=1}^{S} N_{ij}}} \tag{3}$$

which is simply the square root of the relative abundance and thus assigns more weight to the rare species in the community.

4. *Chord distance (CD)*

CD is the Euclidean distance after scaling the time vectors to length 1 (Legendre and Gallagher 2001). After normalization, the Euclidean distance calculated between two steps in the time series is equivalent to the length of a chord joining two points within a segment of a hypersphere of radius 1. Before applying equation (1), the data are transformed according to

$$N_{ij}'=\frac{N_{ij}}{\sqrt{\sum_{i=1}^{S} N_{ij}^2}} \tag{4}$$

5. *Bray-Curtis dis*similarity *(B-C)*

The distances described above can all be obtained by transforming the species abundance data followed by computation of Euclidean distances between the transformed data. B-C cannot be calculated in that way. Moreover, it is not a distance measure *sensu strictu* since it does not match the triangle inequality axiom which demands that $D(a,b) + D(b,c) \geq D(a,c)$, where D is the distance between two objects a and b (Legendre and Legendre 1998). However, it is very popular and widely-used among ecologists. It is calculated according to

$$\text{B-C} = 1 - \frac{2 * \sum_{i=1}^{S} min\left(N_{ij}, N_{ik}\right)}{\sum_{i=1}^{S} N_{ij} + \sum_{i=1}^{S} N_{ik}} \tag{5}$$

where $min(N_{ij}, N_{ik})$ is the lower abundance of species i in the years j and k.

The performance all distance measures was evaluated with respect to (1) their response to abundance change without change in relative abundance by analysing datasets A and B; (2) differences in species richness by comparing the poor and rich versions of datasets A, B, C and D; and (3) their proneness to be affected by the dynamics of dominant species by analysing dataset E.

TLA performed with ED_{rel}, *HD* and *CD* did not indicate any change in datasets A and B where absolute abundance changed but relative abundances stayed constant. TLA based on ED_{abs} and *B-C*, in contrast, yielded significant slopes and is not appropriate for separating the abundance and the compositional component of community change (Table S2). TLA with ED_{abs} and ED_{rel} were sensitive to species richness, misleadingly indicating a faster directional change for datasets A, B, C and D when species richness was increased from 10 to 20. In contrast, TLA based on *HD, CD* and *B-C* yielded identical slopes for the datasets C-poor and C-rich, as well as for D-poor and D-rich (Table S2). Only TLA based on *HD* detected a community change when directional change of rare species was masked by the random dynamic of a few abundant species (Table S2). Table S3 sums up the properties of the distance measures and should help decide which measure to use for a given question.

Table S2. Time lag analysis of artificial datasets (see text and Table S1 for description) using different distance measures. b, slope of the regression of distance on sqrt(time lag); p, significance of the slope as determined by a permutation test; R^2, coefficient of determination; n.s., p > 0.05; ***, p < 0.001.

	Distance measure														
	ED (absolute abundance)			*ED (relative abundance)*			*Hellinger distance*			*Chord distance*			*Bray-Curtis dissimilarity*		
Dataset	*b*	*p*	R^2	*b*	*p*	R^2	*b*	*p*	R^2	*b*	*p*	R^2	*b*	*p*	R^2
A-poor	11.76	***	0.98	0.000	-	-	0.000	-	-	0.000	-	-	0.324	***	0.71
A-rich	16.63	***	0.98	0.000	-	-	0.000	-	-	0.000	-	-	0.324	***	0.71
B-poor	498.2	***	0.30	0.000	-	-	0.000	-	-	0.000	-	-	0.386	***	0.90
B-rich	704.5	***	0.30	0.000	-	-	0.000	-	-	0.000	-	-	0.386	***	0.90
C-poor	11.76	***	0.98	0.214	***	0.98	0.365	***	0.96	0.559	***	0.98	0.338	***	0.98
C-rich	16.63	***	0.98	0.151	***	0.98	0.365	***	0.96	0.559	***	0.98	0.338	***	0.98
D-poor	517.0	***	0.55	0.339	***	0.77	0.676	***	0.87	0.768	***	0.71	0.386	***	0.90
D-rich	731.1	***	0.55	0.239	***	0.77	0.676	***	0.87	0.768	***	0.71	0.386	***	0.90
E	-1193	n.s.	0.03	-0.005	n.s.	0.01	0.060	***	0.81	-0.007	n.s.	0.01	-0.002	n.s.	0.00

Table S3. Synopsis of the properties of different distance measures for time lag analysis.

Distance measure	*Sensitive to absolute abundance change*	*Sensitive to species richness*	*Sensitive to dominant species*
ED (absolute abundance)	yes	yes	yes
ED (relative abundance)	no	yes	yes
Hellinger distance	no	no	no
Chord distance	no	no	yes
Bray-Curtis dissimilarity	yes	no	yes

References

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271-280 doi: 10.1007/S004420100716

Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam