

Random walk simulations of fossil proxy data

The Holocene
1–5
© The Author(s) 2010
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0959683609355180
http://hol.sagepub.com
SAGE

Maarten Blaauw,¹ K.D. Bennett^{1,2} and J. Andrés Christen³

Abstract

A wealth of palaeoecological studies (e.g. pollen, diatoms, chironomids and macrofossils from deposits such as lakes or bogs) have revealed major as well as more subtle ecosystem changes over decadal to multimillennial timescales. Such ecosystem changes are usually assumed to have been forced by specific environmental changes. Here, we test if the observed changes in palaeoecological records may be reproduced by random simulations, and we find that simple procedures generate abrupt events, long-term trends, quasi-cyclic behaviour, extinctions and immigrations. Our results highlight the importance of replicated and multiproxy data for reliable reconstructions of past climate and environmental changes.

Keywords

environmental change, forcing factors, interpreting proxy data, quasi-cyclicities, random walk, simulations.

Introduction

Past ecosystem change is often used to infer past environmental change, e.g. temperature, precipitation, pH or forest openness (e.g. Willis and Birks, 2006). Because species are sensitive to environmental conditions, changing species spectra (extinctions, appearances, long-term trends, abrupt events and periodic cycles) may be responses to environmental changes. However, as most environments are complex and multidimensional, it can be problematic to single out specific environmental factors (Belyea, 2007). Further, chaotic processes (e.g. non-linear species interactions; May, 1976; Bennett, 1993; Ives *et al.*, 2008) and ecology-independent random drift (Hubbell, 2001) can cause large and unpredictable changes in species composition. Moreover, available records are undoubtedly subject to measurement errors.

Random walks are time series where values drift randomly over time (also known as ‘red noise’ series: Roe, 2009). A famous visualisation of a random walk is that of a drunkard stumbling home after leaving a pub, by proceeding in a random direction from each lamp pole or street corner. A Gaussian random walk can be constructed by sampling, for each step, a random value from a normal distribution with the mean taken from a variable’s value at the previous step; the standard deviation of the normal distribution prescribes how fast the time series is allowed to drift. Less common are Poisson random walks, where values are sampled from a Poisson distribution with the intensity set at the previous value of a variable (the Poisson distribution describes the amount of rare events within a period of time, such as the quantity of *Betula* grains ending up in a slice of lake sediment). See Appendix A for more details.

In this paper we simulate fossil proxy time series through Gaussian and Poisson random walks. We test whether these extremely simple simulations can produce results that resemble the complex features commonly identified in palaeoecological data, such as immigrations, extinctions, abrupt events, long-term trends and cyclic behaviour.

Methods

Here we simulate fossil proxy time series through random walks. In the first experiment, Gaussian random walks simulate the temporal evolution of several environmental factors (random environment), which then dictate the abundances of simulated proxies. In the second experiment we remove all environmental forcing, and let species drift randomly over time using Poisson random walks (random proxies). All simulations are performed in R (<http://www.r-project.org>; R Development Core Team, 2009) using code which is available as supplementary information. The simulations are intentionally kept at very basic levels, e.g., factors such as interdependence between environmental forcing factors or competition between species are not taken into account. Including these factors would likely increase the complexity and realism of the outcome; however our aim is to show how seemingly realistic results may be reproduced with obviously unrealistic models without bearing on an underlying environment.

Random environment

First we simulate the reactions of proxies forced by a randomly fluctuating environment. At the start of a time series, for a number of environmental variables we sample random values from a normal distribution. Then for each time step, the environmental variables are allowed to fluctuate from their previous values by sampling

¹Queen’s University Belfast, UK

²Uppsala University, Sweden

³Centro de Investigación en Matemáticas, Guanajuato, Mexico

Received 16 June 2009; revised manuscript accepted 1 October 2009

Corresponding author:

Maarten Blaauw, School of Geography, Archaeology and Palaeoecology, Queen’s University Belfast, Belfast BT7 1NN, UK
Email: maarten.blaauw@qub.ac.uk

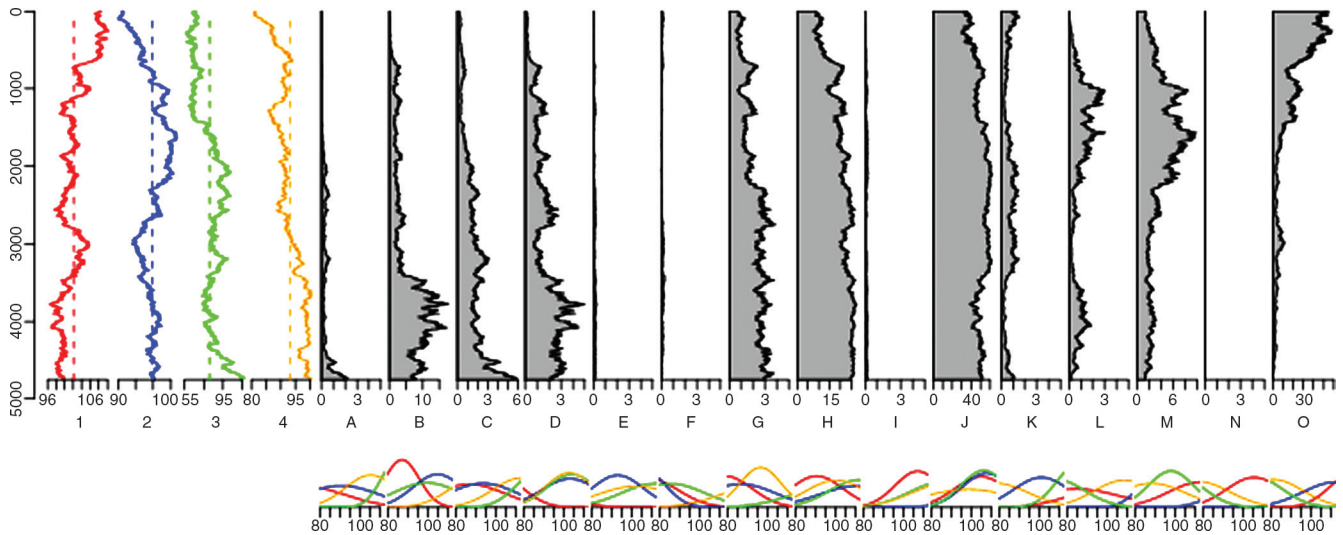


Figure 1. Representative result of random-walk simulation of proxy reactions to fluctuating environmental forcing factors (see text). Environmental factors are represented by the four coloured curves. Proxies (grey curves) are ordered according to the timing of their weighted mean. Proxy responses to the environmental variables are shown in bottom graphs (coloured distributions indicate preferences for the individual environmental factors). Vertical axes show time steps, horizontal axes show relative proxy abundances (%). This result was obtained using the code in Appendix A, with random number seed of 199

from normal distributions. We then simulate a number of proxies, whose abundances are forced entirely by the environmental variables. Each proxy has an optimum and range for each environmental factor. If the environmental factors obtain values close to a proxy's optimum, the proxy will be most abundant, whereas it will be rarer in less favourable conditions. The abundance of each proxy at a certain time is given by the product of its reactions to the individual variables. Proxy abundances are expressed as percentages. See Appendix A for details.

Random proxies

In the second simulation, initial simulated proxy abundances are sampled from a negative binomial distribution. For each following step, species abundances are allowed to fluctuate without any external forcing and independently of each other, depending on their preceding abundances through Poisson random walks. Abundances can increase slightly through immigration from the surrounding communities, by adding a small factor in the intensity of the Poisson distribution at each step. See Appendix A for details.

Results

Random environment

Each run of the program is unique, so a representative outcome is shown in Figure 1. Time series length was set at 5000. For four environmental factors, initial values were sampled from a normal distribution with mean 100 and standard deviation 0.2. For 15 proxies (grey filled curves A–O), responses to the environmental factors were normal distributions based on (i) means sampled from normal distributions with mean 100 and standard deviation 20; (ii) standard deviations sampled from gamma distributions with mean 15 (proxy responses to the environmental variables are shown in bottom graphs).

The four simulated environmental variables, although entirely independent, appear to show some simultaneous reactions such as

inflections around step 3000, and general downward trends for curves 2–4. Several proxy curves show comparable reactions (e.g. B–D, L–M), even if their underlying environmental preferences differ. Some peaks in the proxies are more abrupt than the fluctuations in the environmental forcing factors, suggesting non-linear reactions to combined environmental changes. The heights of the coloured curves at the bottom of Figure 1 indicate the importance of specific environmental factors for the individual proxies; broad curves indicate generalists while peaked curves indicate specialists for the environmental factor. Some proxies are very rare or absent (A, E, F, I, N), because their preferences for the environmental factors do not coincide with the realised environment.

Random proxies

Figure 2 shows a typical outcome, where proxy values are allowed to fluctuate freely without any forcing following a Poisson random walk. Time series length was set at 500. For 15 proxies, initial populations were sampled from a negative binomial distribution with population size 2000 and mean 100. Sporadic immigration events were set at mean size of 5, and with immigration probability of 0.5%. Some proxy fluctuations appear simultaneous, e.g. several proxies re-appear after near absence around step 350. Several proxies show long-term increases or decreases, as well as abrupt events and some indications of quasi-cyclic fluctuations (e.g. the red curve). Curves reminiscent of locally expressed proxies such as microfossils, fungal spores, diatoms or chironomids can be obtained by setting smaller population levels (Figure 3). As for the previous simulations, covariances between proxy curves remain present even if absolute proxy values are plotted instead of percentages.

Discussion

The similarities between simulated random walks and real time series in nature have been noted before (e.g. Yule, 1926; Raup *et al.*, 1973; Roe, 2009). To our knowledge however, this is the first time that

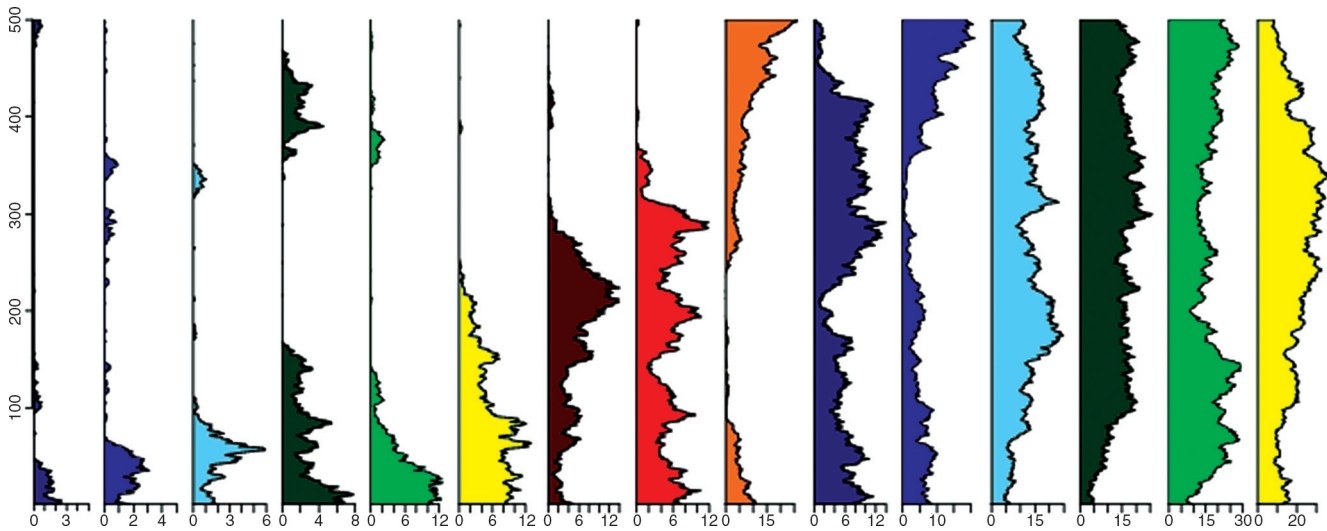


Figure 2. Representative result of random-walk simulation of proxy abundances through time (see text). The 15 proxies are ordered according to the timing of their weighted mean. Vertical axes show time steps, horizontal axes show relative proxy abundances (%). This result was obtained using the code in Appendix A, with random number seed of 1085

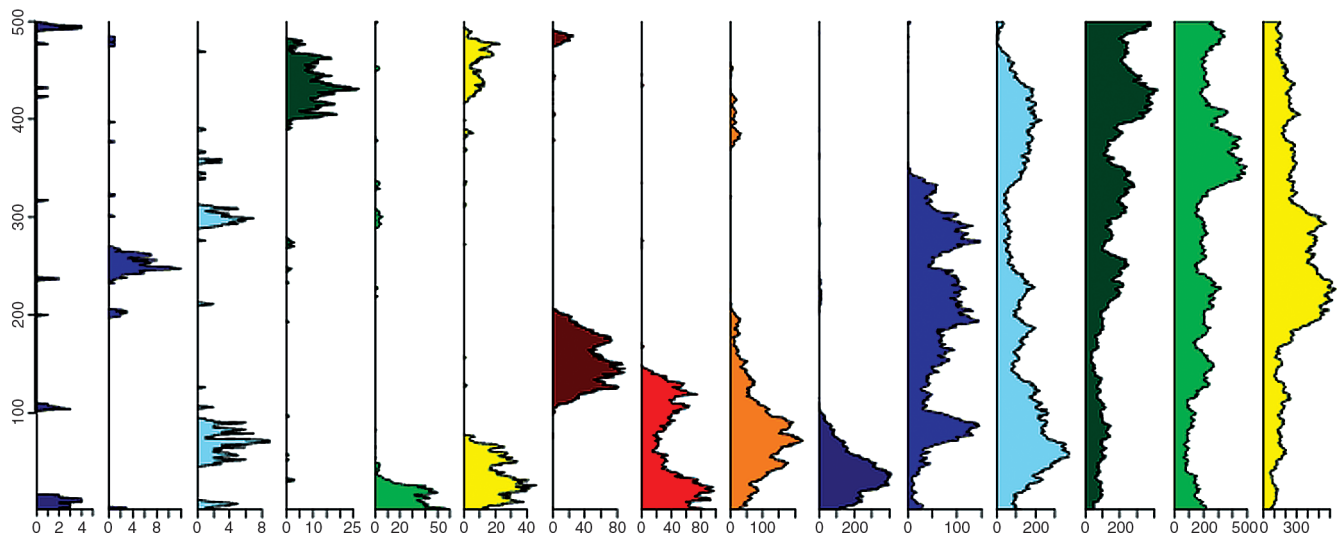


Figure 3. As Figure 2, but using a lower population size of 500, and plotting absolute proxy abundances instead of percentages. The random number seed was 993

long-term trends, abrupt events and quasi-cyclicities encountered in fossil proxy data have been re-created using simple random processes. This topic was partly touched upon by Clark and McLachlan (2003), but they did not discuss if and how these peculiar proxy fluctuations can be simulated with random walks. More complex and ecologically realistic simulations could have been set up, e.g. introducing interdependences between environmental forcing factors, or including species interactions such as competition (Bennett, 1993). However, even without including such arguably vital factors, proxy curves can be simulated which look surprisingly like real-life proxy diagrams.

The ‘environment’ random walk demonstrates several important features of fossil proxy data and their use for reconstructing past environments. First, whereas in our simulations the environmental conditions and environmental requirements of the proxies are known exactly, this is obviously not the case for most real-life fossil proxy data, where only the observations are known (with error;

Maher, 1972), and limited quantitative information on past environmental changes and the proxies’ ecological requirements is available. Second, most likely there is no single environmental factor that can explain the observed proxy changes, and only with multiple proxies can one discern which of the environmental factors forces changes in a particular direction. The system is underdetermined as several environmental configurations can lead to similar proxy reactions (especially if using few proxies). Third, generalist as well as specialist proxies occur in the simulation (low versus high peaks at bottom Figure 1), and both are helpful for environmental reconstructions. Finally, small and gradual combined changes in forcing factors can cause abrupt proxy changes.

The ‘proxy’ random walk reveals that using a very simple random model, surprisingly realistic ecosystem changes occur that resemble decadal- to millennial-scale features commonly registered in fossil proxy archives (e.g. Clark and McLachlan, 2003;

Willis and Birks, 2006). Long-term trends and abrupt events lead to considerable changes in dominance and absence of species. Quasi-cyclic behaviour occurs at a range of timescales. Thus when interpreting fossil proxy events, random changes should be considered as an alternative explanation, especially during stable and favourable environmental conditions (e.g. Chase, 2007).

The term ‘random’ can be interpreted in a number of ways. It is hard to believe that fossil proxy values fluctuated truly randomly without any outside forcing, except for noise introduced by sampling errors (Maher, 1972). Hubbell (2001) proposed that biodiversity phenomena can be explained by neutral dynamics, where births, deaths and immigration of individuals within an area are entirely random without any underlying ecological causative factors. Although controversial (Clark and McLachlan, 2003; Volkov *et al.*, 2004; Clark *et al.*, 2007), it can be seen as a useful null hypothesis (Alonso *et al.*, 2006). Here we interpret random fluctuations to be deterministic but complex and non-linear, such as in Brownian motion of molecules in a gas, the tossing of a coin, the precise time at which a radioactive atom disintegrates, or chaotic ecosystems where multiple factors interact in such a way that species spectra can only be predicted a few seasons ahead (May, 1976; Ives *et al.*, 2008). Although these phenomena are forced by physical processes, multiple interacting factors create complex non-linear reactions, preventing us from knowing the precise processes that caused particular events (e.g. Raup *et al.*, 1973; Bennett, 1993). Similarly, we can only obtain a vague picture of past environmental changes from reconstructing proxy fluctuations (with error) and making assumptions about imperfectly known relationships between the proxies and their forcing factors.

In both simulations, there are times of seemingly simultaneous reactions between proxy curves (co-occurring events of extinctions, immigrations, decreases and increases). In many past ecosystem studies, such proxy changes are attributed to a specific environmental change, e.g. an increase in temperature. Our results warn that ecosystem changes can only be assumed to be environmentally forced if supported by (i) a plausible cause–effect mechanism, (ii) independent sources of well-dated replicate data (e.g. events noticeable in multiproxy records from multiple archives and sites) and (iii) reliably dated environmental data for the inferred change. Further, the response rate of a proxy appears to pay little relation to the rate of forcing environmental change. These are not trivial aspects, because of difficulties with identification of cause–effect mechanisms, and dating uncertainties confound attempts to ascertain whether events were truly synchronous between forcing factors and proxy archives (von Post, 1946; Smith and Pilcher, 1973; Baillie, 1991; Blaauw *et al.*, 2007, 2010; Parnell *et al.*, 2008).

What implications do our simulations have for the interpretation of fossil proxy data? Proxies reflecting a local signal often show rather erratic proxy curves that are hard to reproduce within and between sites. In such cases, even pronounced proxy fluctuations could have been caused by neutral dynamics, and could thus be simulated by our random walks. The large-scale patterns of Holocene pollen diagrams are based on the pool of available species and their migration order, and as such are highly reproducible within regions even if the causative factors can be surprisingly hard to decipher (Bennett, 1993). Some of the variation at fine scales (e.g. between neighbouring cores or pollen slices) might be due to sampling error (Maher, 1972). For pollen, our simulations are probably nested within these two

extremes, i.e. reflecting fluctuations at decadal to millennial scales. The null hypothesis of neutral dynamics needs to be considered when interpreting proxy curves.

Acknowledgements

We thank Kathy Willis, John Birks, Donatella Magri and Enrique Lemus-Rodríguez for helpful comments. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Appendix A

There are many types of random walks. A Gaussian random walk for variable X can be constructed by sampling, for each step $i > 1$, a random value from a normal distribution with the mean taken from the variable’s value at the previous step ($X_i = N(x_{i-1}, \sigma)$, where the size of σ prescribes how fast the time series is allowed to drift and is set at the beginning). Poisson random walks can be constructed with the formula $X_i \sim \text{Pois}(X_{i-1})$.

Random environment

At the start of a time series of $i=n$ steps (t_0, \dots, t_n), for $j=m$ environmental variables we sample random values from a normal distribution with mean μ_{env} and standard deviation σ_{env} ($X_j \sim N(\mu_{env}, \sigma_{env})$, where X_j denotes environmental variable number j). Then for each time step t_i , the environmental variables are allowed to fluctuate from their previous values by sampling from normal distributions: $X_{i,j} \sim N(X_{i-1,j}, \sigma_{env})$.

We then simulate $k=l$ proxies (Y_1, \dots, Y_l), whose abundances are forced by the environmental variables: $Y_k \sim f(X_1, \dots, X_l)$. Each proxy has an optimum and range for each environmental factor; $\mu_{k,j} \sim N(\mu_{env}, var)$ where var determines the spread of the optima across the environment gradients, and $\sigma_{k,j} \sim \text{gamma}(range, 1)$ where $range$ gives the environmental tolerances of the proxies (the gamma distribution with shape 1 is a continuous approximation of the discrete Poisson distribution, and ensures that sampled σ values are positive). If the environmental factors obtain values close to a proxy’s optimum, the proxy will be most abundant, whereas it will be rarer in less favourable conditions. The abundance of each proxy at time t_i is given by the product of its reactions to the individual variables, $Y_{ik} = \prod_{j=1}^m N(X_{i,j}, \mu_{j,k}, \sigma_{j,k})$. By default, proxy abundances are expressed as percentages.

Random proxies

In the second simulation, species abundances are allowed to fluctuate without any external forcing and independently of each other, depending on their preceding abundances through Poisson sampling. Additionally, abundances can increase slightly through immigration from the surrounding communities. At the start of a time series of $i=n$ steps (t_1, \dots, t_n), for $j=m$ simulated species abundances $X_{j,0}$ are sampled from a negative binomial distribution (two parameters, *population* and *mean*). Then for each step t_i , species abundances are sampled from Poisson distributions based on the species’ previous abundances, $X_{i,j} \sim \text{Pois}(X_{i-1,j})$. To simulate sporadic immigration events (mean size *seed*, probability *imm*), an independent Poisson random variable is added, obtaining $X_{i,j} \sim \text{Pois}(X_{i-1,j} + \text{seed} * \text{imm})$. This process may be described as

$X_{i,j} = Z_{i,j} + M_{i,j}$, where $Z_{i,j} \sim \text{Pois}(X_{i-1,j})$ and $M_{i,j} \sim \text{Bi}(n=\text{seed}, p=\text{imm})$ are independent. $M_{i,j}$ models the sporadic immigration events. Since $M_{i,j}$ is approximately distributed as $\text{Pois}(\text{seed}*\text{imm})$ and adding two independent Poisson random variables results in a single Poisson random variable, we obtain $X_{i,j} \sim \text{Pois}(X_{i-1,j} + \text{seed}*\text{imm})$.

Using fairly advanced results in Markov chains (Bremaud, 1999), with the superharmonic function $h(k)=k-2e$, $F = \{0\}$, and with $0 < \varepsilon < e = \text{seed}*\text{imm}$, we prove that the Poisson random walk is stationary (since all states are connected with period 1). We have not yet been able to find the stationary (equilibrium) distribution, but numerical experiments suggest it is highly peaked at 0, similar to a negative binomial with parameters $r=0.2$ and $p=0.01$, but with a heavier and longer tail. Over very long timescales, our modelled species will thus be absent most of the time, while never becoming permanently extinct. It is also well known that the Gaussian random walk is ergodic (values that are very far away in time from a sample cannot be predicted, irrespective of initial conditions). Initial values are a Poisson and a Gaussian (with suitable parameters), respectively, but these are not crucial given the stationary properties of the chains.

All modelling was coded in the functions *RandomEnv* and *RandomProx* in file *Random.R* (supplementary information), which can be loaded or pasted into R (R Development Core Team, 2009). Comments are preceded by a # symbol. Simulations may be produced by typing *RandomProx()* or *RandomEnv()*, while default settings can be changed (e.g. *time*: steps in time, *nforc*: number of environmental forcing factors, *setseed*: control on random seed, *perc*: calculate relative abundances).

References

- Alonso, D., Etienne, R.S. and McKane, A.J. 2006: The merits of neutral theory. *Trends in Ecology and Evolution* 21, 451–57.
- Baillie, M.G.L. 1991: Suck in and smear: two related chronological problems for the 90s. *Journal of Theoretical Archaeology* 2, 12–16.
- Belyea, L.R. 2007: Revealing the Emperor's new clothes: niche-based palaeoenvironmental reconstruction in the light of recent ecological theory. *The Holocene* 17, 683–88.
- Bennett, K.D. 1993: Holocene forest dynamics with respect to southern Ontario. *Review of Palaeobotany and Palynology* 79, 69–81.
- Blaauw, M., Christen, J.A., Mauquoy, D., van der Plicht, J. and Bennett, K.D. 2007: Testing the timing of radiocarbon-dated events between proxy archives. *The Holocene* 17, 283–88.
- Blaauw, M., Wohlfarth, B., Christen, J.A., Ampel, L., Veres, D., Hughen, K.A., Preusser, F. and Svensson, A. 2010: Were last glacial climate events simultaneous between Greenland and France? A quantitative analysis of non-tuned chronologies. *Journal of Quaternary Science* in press.
- Bremaud, P. 1999: *Markov chains: Gibbs fields, Monte Carlo simulation and queues*. Texts in Applied Mathematics, 31, Springer.
- Chase, J.M. 2007: Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences* 104, 17 430–34.
- Clark, J.S. and McLachlan, J.S. 2003: Stability of forest diversity. *Nature* 423, 635–38.
- Clark, J.S., Dietze, M., Chakraborty, S., Agarwal, P.K., Ibanez, I., LaDeau, S. and Wolosin, M. 2007: Resolving the biodiversity paradox. *Ecology Letters* 10, 647–62.
- Hubbell, S.P. 2001: *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Ives, A.R., Einarsson, Á., Jansen, V.A.A. and Gardarsson, A. 2008: High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. *Nature* 452, 84–87.
- Maher, L.J., Jr 1972: Nomograms for computing 0.95 confidence limits of pollen data. *Review of Palaeobotany and Palynology* 13, 85–93.
- May, R.M. 1976: Simple mathematical models with very complicated dynamics. *Nature* 261, 459–67.
- Parnell, A.C., Haslett, J., Allen, J.R.M., Buck, C.E. and Huntley, B. 2008: A flexible approach to assessing synchronicity of past events using Bayesian reconstructions of sedimentation history. *Quaternary Science Reviews* 27, 1872–85.
- Raup, D.M., Gould, S.J., Schopf, T.J.M. and Simberloff, D.S. 1973: Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology* 81, 525–42.
- R Development Core Team, 2009: R: A language and environment for statistical computing. <http://www.R-project.org>
- Roe, G. 2009: Feedbacks, timescales, and seeing red. *Annual Review of Earth and Planetary Sciences* 37, 93–115.
- Smith, A.G. and Pilcher, J.R. 1973: Radiocarbon dates and vegetational history of the British Isles. *New Phytologist* 72, 903–14.
- Volkov, I., Banavar, J.R., Maritan, A. and Hubbell, S.P. 2004: The stability of forest biodiversity. *Nature* 427, 696–97.
- von Post, L. 1946: The prospect for pollen analysis in the study of the Earth's climatic history. *New Phytologist* 45, 193–217.
- Willis, K.J. and Birks, H.J.B. 2006: What is natural? The need for a long-term perspective in biodiversity conservation. *Science* 314, 1261–65.
- Yule, G.U. 1926: Why do we sometimes get nonsense-correlations between time series? *Journal of the Royal Statistical Society* 89, 1–69.