1 Estimation of colonization and extinction rates

The fundamental equation of the Equilibrium Theory of Island Biogeography (ETIB) can be applied to (and solved for) a single species (Alonso et al., 2015; Hanski, 2001). Two assumptions (species equivalence and independence) allow to recover the classic equation of the ETIB (Hanski, 2001). Independence considers species as if they did not interact. Equivalence assumes the same colonization and extinction rates across species, which means that all species have the same response to the different factors affecting their state. This may be the case for many horizontal communities (sensu Vellend 2010). Under these assumptions,
the same equation governs the probability of any species being present at time $t$ in the community, $P(1, t)$:

$$\frac{dP(1, t)}{dt} = c \left[ 1 - P(1, t) \right] - e \cdot P(1, t)$$  \hspace{1cm} (S1)

The temporal evolution of this probability is not fully defined without an initial condition. When we consider that the species was initially absent ($P(1, 0) = 0$), $P(1, t)$ can be regarded as a transition probability, this is, the probability that a particular species is present in the community at time $t$ given that it was initially absent, given the colonization and extinction rates. Accordingly, we can label this probability $T_{10}(t|c, e)$. Analogously, $T_{00}(t|c, e)$ is defined as the probability that the species was initially absent and remains so at time $t$. Note that, because of normalization, $T_{00} + T_{10} = 1$. The solution of Eq (S1) with these two particular initial conditions yields:

$$T_{10}(t|c, e) = \frac{c}{c + e} \left[ 1 - \exp\left(- (e + c) t \right) \right],$$

$$T_{00}(t|c, e) = 1 - \frac{c}{c + e} \left[ 1 - \exp\left(- (e + c) t \right) \right].$$  \hspace{1cm} (S2)

Similarly, the probability $T_{01}(t)$ that a species initially present is not observed at time $t$, together with the probability $T_{11}(t)$ that a species initially present remains in the community at time $t$, can be expressed as

$$T_{01}(t|c, e) = \frac{e}{c + e} \left[ 1 - \exp\left(- (e + c) t \right) \right],$$

$$T_{11}(t|c, e) = 1 - \frac{e}{c + e} \left[ 1 - \exp\left(- (e + c) t \right) \right].$$  \hspace{1cm} (S3)

These probabilities actually define a two-state (0 for absent and 1 for present) discrete-time Markov chain for each species that can be used to predict the community configuration vector $n(t) = (n_1(t), n_2(t), \ldots, n_{S_P}(t))$ of the community after a time $t$ has elapsed, where $n_i \in \{0, 1\}$. With these probabilities, we can estimate the likelihood of observing a sequence of presences and absences for a single species. Conversely, given a presence-absence matrix $M$ obtained by sampling the community at uniformly spaced times (i.e., a species \times samples matrix), thanks to species independence we can estimate community colonization and extinction rates by means of the following likelihood function:

$$P(M|c, e) = T_{00}^{N_{00}} T_{10}^{N_{10}} T_{01}^{N_{01}} T_{11}^{N_{11}},$$  \hspace{1cm} (S4)

$N_{01}$ representing the number of times that we have observed species changing from state 1 (present) to state 0 (absent), and so on.

We used different methods to find the maximum likelihood estimators (m.l.e.) for colonization and extinction rates. If we are dealing with regular sampling schemes, i.e. equally spaced temporal samples, $\Delta t$ being the time separation between samples, we used an analytically exact method that is based on the maximum likelihood estimators for transition probabilities, $T_{10} = N_{10}/(N_{10} + N_{00})$ and $T_{01} = N_{01}/(N_{01} + N_{11})$, as in Alonso et al. (2015): in this reference, it is shown that m.l.e. estimators for $c$ and $e$ can be calculated in terms of
\( N_{ij} \) and \( \Delta t \) by solving the system
\[
\frac{e}{c} = \frac{N_{01}(N_{10} + N_{00})}{N_{10}(N_{01} + N_{11})},
\]
\[
e + c = -\frac{1}{\Delta t} \log \left( \frac{N_{00}N_{11} - N_{01}N_{10}}{(N_{10} + N_{00})(N_{01} + N_{11})} \right).
\]  \( \text{(S5)} \)

However, if communities are sampled at not equally-spaced times (i.e., we have irregular sampling schemes), the likelihood function is a bit more complicated — but amenable for some analytical treatment, however. Let \( t_0, t_1, \ldots, t_\tau \) be the irregularly-sampled times, so the species \( \times \) times presence-absence matrix \( M \) is an \( S_P \times (\tau + 1) \) matrix. Now focus on an arbitrary species \( i \) \((1 \leq i \leq S_P)\). Let \( T(n_i(t_{j+1}) \leftarrow n_i(t_j)|c,e) \) denote the transition probability from state \( n_i(t_j) \) at time \( t_j \) to state \( n_i(t_{j+1}) \) at time \( t_{j+1} \) for that species, given the rates \( c, e \). As for the regular sampling schemes, the probabilities can be expressed as
\[
T_{00}(\Delta t_j|c,e) = 1 - \frac{e}{c+e} [1 - \exp(-(e + c)\Delta t_j)],
\]
\[
T_{10}(\Delta t_j|c,e) = \frac{e}{c+e} [1 - \exp(-(e + c)\Delta t_j)],
\]
\[
T_{01}(\Delta t_j|c,e) = \frac{e}{c+e} [1 - \exp(-(e + c)\Delta t_j)],
\]
\[
T_{11}(\Delta t_j|c,e) = 1 - \frac{e}{c+e} [1 - \exp(-(e + c)\Delta t_j)],
\]  \( \text{(S6)} \)

where \( \Delta t_j = t_{j+1} - t_j \). Note here that the temporal evolution of the model is determined by \( c + e \). The likelihood of the configurations for species \( i \) across sampling times can be written as
\[
P(n_i(t_0), \ldots, n_i(t_\tau)|c,e) = \prod_{j=0}^{\tau-1} T(n_i(t_{j+1}) \leftarrow n_i(t_j)|c,e),
\]  \( \text{(S7)} \)

and, because of species independence, the likelihood of matrix \( M \) is
\[
P(M|c,e) = \prod_{i=1}^{S_P} \prod_{j=0}^{\tau-1} T(n_i(t_{j+1}) \leftarrow n_i(t_j)|c,e).
\]  \( \text{(S8)} \)

We aim to optimize this likelihood in order to obtain estimates for colonization and extinction rates \( c \) and \( e \). For this purpose, we have devised two methods, a heuristic search or a semianalytical method based on the gradient of the likelihood function. Once we have obtained our estimates, we can also find their confidence intervals using two different methods: a stepwise procedure and a Hessian-based binary search. Both procedures find the values at which the likelihood function varies 1.96 units for each of the rates.

The semianalytical method to find m.l.e. for colonization and extinction rates is based on the gradient of the log-likelihood function. It can be expressed in a compact form as follows:
\[
\log P(M|c,e) = \sum_{j=0}^{\tau-1} \left[ N_{00}^j \log T_{00}(\Delta t_j|c,e) + N_{10}^j \log T_{10}(\Delta t_j|c,e) \right.
\]
\[\left. + N_{01}^j \log T_{01}(\Delta t_j|c,e) + N_{11}^j \log T_{11}(\Delta t_j|c,e) \right].
\]  \( \text{(S9)} \)
where \( N^j_{00} \) was defined as the number of \( 0 \leftrightarrow 0 \) transitions, across all species, when \( t_{j+1} \leftrightarrow t_j \), and likewise for \( N^j_{10}, N^j_{01} \) and \( N^j_{11} \). These four numbers, as well as the \( \tau \) time intervals \( \Delta t_j \), are known from the data. We simply calculated the critical points solving the nonlinear equations obtained by equating to zero the gradient of the log-likelihood. We reproduce here the two partial derivatives,

\[
\frac{\partial \log P}{\partial c} = -\frac{1}{e+c} \sum_{j=0}^{\tau-1} \left\{ \left( \frac{N^j_{00}}{T^j_{00}} - \frac{N^j_{10}}{T^j_{10}} \right) \left[ c \lambda_j \Delta t_j + \frac{e}{e+c} (1 - \lambda_j) \right] + \left( \frac{N^j_{11}}{T^j_{11}} - \frac{N^j_{01}}{T^j_{01}} \right) \left[ e \lambda_j \Delta t_j - \frac{e}{e+c} (1 - \lambda_j) \right] \right\}, \tag{S10}
\]

and

\[
\frac{\partial \log P}{\partial e} = -\frac{1}{e+c} \sum_{j=0}^{\tau-1} \left\{ \left( \frac{N^j_{00}}{T^j_{00}} - \frac{N^j_{10}}{T^j_{10}} \right) \left[ c \lambda_j \Delta t_j - \frac{e}{e+c} (1 - \lambda_j) \right] + \left( \frac{N^j_{11}}{T^j_{11}} - \frac{N^j_{01}}{T^j_{01}} \right) \left[ e \lambda_j \Delta t_j + \frac{c}{e+c} (1 - \lambda_j) \right] \right\}. \tag{S11}
\]

In these two expressions above we have defined \( \lambda_j = \exp(-(e+c)\Delta t_j) \) and denoted \( T^j_{00} = T_{00}(\Delta t_j | c, e) \), and so on. For each dataset (matrix \( M \)), we found the roots of the gradient using function \texttt{multiroot} of the R package \texttt{rootSolve}. These are the critical points of the likelihood function. Thus, we can encounter a local maximum, local minimum, or a saddle point, and in the case of a local maximum, these values are the m.l.e. of colonization and extinction rates, \( c \) and \( e \).

To obtain confidence intervals for the estimates of \( c \) and \( e \), we developed two methods, a stepwise procedure and a Hessian-based binary search. We defined the confidence intervals as the values of \( c \) or \( e \) at which the likelihood function varies 1.96 log-units. The stepwise procedure starts at the m.l.e. for each rate and proceeds stepwise, adding (or subtracting) a small, user-defined quantity to them. When the likelihood function reaches a difference of 1.96 log-units, the algorithm has found the confidence interval. The Hessian-based procedure uses that, for a multivariate Gaussian random vector, the Hessian of the log-likelihood exactly coincides with minus the inverse of the covariance matrix, so the amplitudes of the confidence intervals (standard deviations) can be calculated as the square roots of the diagonal elements of the inverse negative Hessian matrix. In general (for arbitrarily distributed random variables), standard deviations for each m.l.e. can be approximated this way and yield good approximations to true deviations when their values are small. We used the standard deviations so estimated as an initial guess, and then refined the estimates with a binary search to find the exact value at which the likelihood function has a difference with the m.l.e. of 1.96 log-units. Explicit expressions for the entries of the Hessian matrix can be computed in terms of the number of transitions \( N^j \) of each class, the time intervals \( \Delta t_j \), the transition probabilities \( T^j \), and the rates \( c \) and \( e \). Expressions are too convoluted to be reproduced here, in any case.

We end this section showing the simulated species richness as a function of time for three classic datasets, namely: Farne Island birds, Neotoma deciduous forest birds, and arthropods.
in the Florida Keys (island E3), see Figure S1. The observed richness lies within the 95% confidence interval of the simulations.

2 Community composition patterns

Jaccard index allowed us to compare sites using presence-absence data, and it is defined as follows:

\[ J_{ij} = \frac{C}{A + B + C}, \]  \hspace{1cm} (S12)

where \( A \) is the number of species present at time \( i \) and not at time \( j \), \( B \) the number of species present at time \( j \) and not at \( i \), and \( C \) the number of species present at both times. Jaccard’s index also lets us compare the same site at different times, and its expected value can be estimated with our colonization and extinction dynamics and its associated Markov chain.

Then, if we considered \( i \) and \( j \) different samples taken at the same site at times \( t_i = 0 \) and \( t_j = t \), the expected number of species present at both times, \( C \), can be calculated as the product of \( T_{11} \), the transition probability for a species being present in both \( i \) and \( j \), \( S_P \), the number of species in the pool, and \( p_i \), the proportion of species present at \( t_i \). Besides, the average number of species present at time \( i \) but absent at time \( j \), \( A \), can be obtained as the product of \( T_{01} \), the transition probability for species being present in \( i \) and not \( j \), \( S_P \) and \( p_i \), whereas \( B \) can be estimated as the product of \( T_{10} \), the transition probability of a species being absent from \( i \) and present in \( j \), and \( S_P \) minus the product of \( S_P \) and \( p_i \). If the random variables \( C \) and \( A + B + C \) were independent, we could calculate the expected value of the ratio \( C/(A + B + C) \) as

\[ E \left[ \frac{C}{A + B + C} \right] = E[C]E \left[ \frac{1}{A + B + C} \right]. \] \hspace{1cm} (S13)

Ignoring correlations and approximating \( E \left[ \frac{1}{A + B + C} \right] \approx E [A + B + C]^{-1} \) —which is not necessarily true in general—we obtain the following approximation for the Jaccard index in terms of \( p_i \) and transition probabilities,

\[ J_i(t) \approx \frac{p_i T_{11}}{p_i T_{01} + (1 - p_i) T_{10} + p_i T_{11}}. \] \hspace{1cm} (S14)

Using that \( T_{11} + T_{01} = 1 \), Eq. (S14) can be expressed as

\[ J_i(t) \approx \frac{p_i T_{11}}{p_i + (1 - p_i) T_{10}}. \] \hspace{1cm} (S15)

However, note that (S14) is not the expectation of the Jaccard index, as such expected value would imply summing over all values of the Jaccard index given the probability of each possible configuration of the community. However, the law of large numbers probably reduce that hypothetical expression to equation (S14) and substituting the rates (S2) and (S3) leads into the expressions for \( J_i(t) \) reproduced in the main text (Box 1). If we compare any given
community with the same community when time goes to infinity, we can calculate the Jaccard Index as follows:

\[ J_i = \lim_{\Delta t \to \infty} J_i \approx \frac{p_i c}{p_i e + c}. \]  \hspace{1cm} (S16)

If the initial proportion of species coincides with that of equilibrium, \( p_i = \frac{c}{e + c} \), our approximation for the Jaccard index simplifies to

\[ J^*(t) \approx \frac{c + e \exp(-(e + c)t)}{c + 2e - e \exp(-(e + c)t)}, \]  \hspace{1cm} (S17)

which converges to \( \tilde{J}^* = \frac{c}{c + 2e} \) as \( t \to \infty \). We now estimate the time \( t \), in units of the characteristic time \( T_c = \frac{1}{e + c} \), that takes the system to reach the asymptotic Jaccard index value. Let \( \tau = t/T_c = (e + c)t \) and let us fix the relative error \( |J^*(t) - \tilde{J}^*|/\tilde{J}^* \) equal to a certain amount \( \chi \). From

\[ \frac{|J^*(t) - \tilde{J}^*|}{\tilde{J}^*} = \chi \]  \hspace{1cm} (S18)
we can solve for \( x = \exp(-(e + c)t) = e^{-\tau} \) to get

\[ e^{-\tau} = \frac{c(c + 2e)\chi}{2e(c + e) + c\exp(\chi)} \approx \frac{\chi(1 + 2z)}{2z(1 + z)} \]  \hspace{1cm} (S19)

where we used that \( \chi \ll 1 \) and defined \( z = e/c \). Taking logarithms at both sides, we get

\[ \tau = -\log \chi - \log \frac{1 + 2z}{2z(1 + z)} = 4.61 - \log \frac{1 + 2z}{2z(1 + z)} \]  \hspace{1cm} (S20)

for a small relative error \( \chi = 0.01 \). Now we solve for the values of \( z \) that yield \( \log \frac{1 + 2z}{2z(1 + z)} = \pm 1 \), which are \( z = 0.217 \) and \( z = 2.31 \), respectively. This means that, if \( 0.22 \leq \frac{e}{c} \leq 2.31 \), then the ratio \( \frac{T_r}{T_c} \) is bounded by 3.61 and 5.61. Therefore, for a wide range of extinction to colonization ratios, we expect that the time \( t \) the system takes to reach the asymptotic Jaccard index is about 4 or 5 characteristic times, as stated in the main text.

We have defined an alternative measure to the characteristic time \( T_c \), the half-relaxation time, which corresponds to the time that takes the system to have a decrease in similarity, measured with the Jaccard index, that is half the decrease that it encounters as time tends to infinity. To find an expression for the half-relaxation time, we have to solve for \( \Delta t = T_r \) in the following equation:

\[ \frac{1}{2} \left( 1 - \frac{p_i c}{p_i e + c} \right) = \frac{p_i [c + e \exp(-(e + c)T_r)]}{p_i [e + c \exp(-(e + c)T_r)] + c [1 - \exp(-(e + c)T_r)]}. \]  \hspace{1cm} (S21)

After some algebra, it reduces to the following expressions, starting or not with an initial \( p_i \) equal to that of the equilibrium:

\[ T_r \approx T_c \log \frac{c + (c + 2e)p_i}{c + ep_i}, \]  \hspace{1cm} (S22)

\[ T_r^* \approx T_c \log \frac{2c + 3e}{c + 2e}. \]  \hspace{1cm} (S23)
In Figure S2, we simulated colonization and extinction dynamics for a hundred simulations of a model with $T_c = 20$ and occupancy at equilibrium equal to 0.1. Starting from an initial occupancy of 0.5, we have calculated the Jaccard index over a hundred timesteps for each of these realizations. Besides, we calculated our theoretical approximation (blue line) for the expected Jaccard index [cf. Eq. (S15)], and the half-saturation time for this set of model parameters.

Finally, we have simulated a sequence of models with $T_c$ ranging from 10 to 100 and three different occupancies at the equilibrium, 0.1, 0.5, and 0.9. For each of these models, we obtained 200 realizations and averaged their evolution for the Jaccard index with time. We show in Figure S3 the relation between $T_c$ and half-relaxation time, that we found theoretically [colored lines, Eqs. (S22) and (S23)] and empirically (black points) the half-relaxation time from the averaged dynamics for each of these models.

3 Properties of $c$, $e$ and $T_c$ as estimators

The main properties that determine a good estimator are bias, consistency, and efficiency. Bias is the difference between the true parameter and the mean of the distribution of the estimator. An estimator with no difference is considered unbiased. Consistency is the property of an estimator of being near the value of the parameter as sample size increases. Finally, efficiency refers to the size of the distribution of the estimator, being efficient an estimator with small support.

To study bias, we simulated a 1000 species $\times$ 1001 samples dataset with equal colonization and extinction rates ($c = e = 0.1$) and a characteristic time $T_c = 5$. From this matrix, we obtained true colonization and extinction rates, because these can vary from true ones as it is a stochastic process. We subsampled 1000 times this matrix to have only 100 transitions each time and recalculated colonization and extinction rates. We used differences of true and estimated $c$, $e$, and $T_c$ to quantify bias using skewness and the D’Agostino test, which yielded non-significant biases for each of the estimators ($c$ skew = 0.0270, p-value n.s.; $e$ skew = $-0.0101$, p-value n.s.; $T_c$ skew = 0.0642, p-value n.s.). Following the same procedure for unbalanced values of the rates ($c = 0.02$, $e = 0.18$ or $c = 0.18$, $e = 0.02$) yielded the same absence of bias.

We also studied the consistency of the estimators. We simulated 1000 replicates of a 1000 species $\times$ 1001 samples matrix following a specified colonization and extinction dynamics with $T_c = 5$. We calculated the true parameters for each one, and then we subsampled them to obtain 10, 20, 50, 100, 200, and 500 transitions and estimate $c$, $e$, and $T_c$. Then we calculated the bias between true and estimated parameters along this sequence. Figure S4 shows that, as we increase the number of transitions, which can be interpreted as increasing the sample size, the estimated parameters get closer to the true values, decreasing their dispersion. This is also a sign of an increased efficiency.

4 Sources of error in $T_c$

We have identified three possible sources of error when estimating $c$, $e$, and $T_c$. The first one and probably the most evident is the lack of conformity of the studied communities to the
assumptions of the model, namely, *independence* and *equivalence*. However, we want to stress that our model should be regarded as a null model, acting as a mean-field approximation to the actual dynamics (that may include many other processes).

The second one is related to how thoroughly we have sampled our community. We have found that the estimates may not be precise when the *relative sampling frequency* is below \( \sim 0.3 \). Figure S5 shows the effect of low sampling frequencies on \( T_c \). We simulated 30 different matrices of 1000 species \( \times \) 1000 samples, each one with a different pair of colonization and extinction but a characteristic time of 10, and we subsampled to achieve relative sampling frequencies from 0.1 to 10. From the subsampled matrices we estimated their characteristic time, that seemed to vary wildly under a relative sampling frequency \( \nu_s \approx 0.3 \). A procedure to discern if our estimates are reliable would be to undersample our data to recalculate the estimates; if those are consistent with the previous estimates, these may be acceptable. If the estimates change, we might be sampling below \( T_c \).

The last possible source of error that we have identified occurs when we subsample our initial data, and we find a linear relationship between mean sampling time and \( T_c \). As can be shown trivially from (S2) and (S3), in the regular sampling scheme \( T_c \) depends on the transition probabilities \( T_{01} \) and \( T_{10} \). Together they determine the degree of lag-1 autocorrelation, \( \rho \), in our data (following Tuljapurkar 1997):

\[
T_c = \frac{1}{c + e} = - \frac{\Delta t}{\log(1 - T_{01} - T_{10})} = - \frac{\Delta t}{\log \rho}.
\]  

(S24)

Thus, in order to find a linear relationship between sampling time and \( T_c \) we need a \( \rho \) constant at different sampling times. We have examined data from the left hand of individual M3 (Caporaso et al., 2011). We found an estimate of \( T_c = 2.05 \) days, with a mean \( \Delta t \) of 1.23 days. In order to examine the robustness of the estimates, we subsampled the data, increasing \( \Delta t \) to 1.5, 1.8, and 2.5. Lag-1 autocorrelation was measured row by row, adopting the convention of considering the correlation of a series formed only by ones or zeros as having a value of 1. We found that autocorrelation remained constant, thus making \( T_c \) to follow a linear relationship with \( \Delta t \). Also, we explored the autocorrelation of random strings of ones and zeros and the sequences produced by stochastic realizations of the dynamics corresponding to the estimated parameters for that dataset (Figure S6). The same procedure was followed to examine data of island E1 from the arthropods dataset (Simberloff and Wilson, 1969). The subsampled data showed less autocorrelation than the original data, and the model simulations were congruent with it. The difference between the observed autocorrelation and the stochastic realizations allow us to conclude that when this difference is high, the estimates should be considered with caution. The reason behind this mismatch may correspond to the subsampled data being close to a random matrix with the proportion of presence/absence of the original data, and it might indicate that we need to sample more to find correct estimates of the dynamics.

References

Alonso, D., A. Pinyol-Gallemí, T. Alcoverro, and R. Arthur. 2015. Fish community reassembly after a coral mass mortality: higher trophic groups are subject to increased rates of extinction. Ecology letters 18:451–61.
Caporaso, J. G., et al. 2011. Moving pictures of the human microbiome. Genome Biology 12:R50.

Hanski, I. 2001. Spatially realistic theory of metapopulation ecology. Naturwissenschaften 88:372–381.

Simberloff, D., and E. O. Wilson. 1969. Experimental zoogeography on islands: The coloni- zation of empty islands. Ecology 50:278–296.

Tuljapurkar, S., 1997. Stochastic Matrix Models. Pages 59–87 in Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Springer US, Boston, MA.

Vellend, M. 2010. Conceptual synthesis in community ecology. The Quarterly Review of Biology 85:183–206.
Figure S1: **Temporal dynamics of three classic communities.** A) Deciduous forest birds at Neotoma, B) island birds in the Farne Islands, C) arthropods community in the Florida Keys (island E3). Shaded areas correspond to the 95% confidence interval obtained from model simulations, whereas the dashed line refers to the median of the simulations.
Figure S2: **Evolution of the Jaccard index with time under simulations.** The blue line shows the expected value for the Jaccard index using our theoretical approximation. The magenta diamond indicates the characteristic Jaccard index, that is, the value of the Jaccard index between the initial community and the one present after a characteristic time (20 units in this case). Approximately after 4 times $T_c$, the Jaccard index stabilizes.
Figure S3: Relation of the $T_r$ with $T_c$. For the same characteristic time, the systems that have a higher occupancy at the equilibrium had a lower half-relaxation time, while the systems with a lower occupancy took the most to reach it. As we can see, there is a linear relation among $T_c$ and $T_r$ for equal levels of occupancy.
Figure S4: **Properties of $c$, $e$ and $T_c$ as estimators.** We found no bias. Consistency and efficiency increased as sample size (i.e., the number of transitions) increased.
Figure S5: **Influence of sampling frequency over estimates of characteristic time.** When we sample under a relative sampling frequency of $\approx 0.3$, we can observe that the estimates of the characteristic time, $T_c$, start to deviate from the true value, in this case $T_c = 10$ (white dashed line). Subsampling is needed in those cases to increase certainty.
Figure S6: **Autocorrelation is associated with characteristic time estimation.** Autocorrelation in the case of the data of the left hand of individual M3 seems stable after subsampling, and close to the autocorrelation of random sequences of 0s and 1s. However, in the case of the arthropods of island E1, each subsampling decreased autocorrelation, thus maintaining the estimate of characteristic time. In this last case, we also see that our simulations closely resemble the autocorrelation found for the original data.