Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian model-based clustering of ecological time series

Nourhan Shafik

nourhan.shafik@helsinki.fi

University of Helsinki

Arthur V. Rodrigues

University of Helsinki

Pasi Sihvonen

Finnish Museum of Natural History

Jarno Vanhatalo

University of Helsinki

Research Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11525622/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bayesian model-based clustering of ecological time series

Nourhan Shafik[1,2*], Arthur V. Rodrigues[1], Pasi Sihvonen[3], Jarno Vanhatalo[1,2]

[1*]Department of Organismal and Evolutionary Biology, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.

[2]Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, Helsinki, Finland.

[3]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland.

*Corresponding author(s). E-mail(s): nourhan.shafik@helsinki.fi;
Contributing authors: arthur.rodrigues@helsinki.fi;
pasi.sihvonen@helsinki.fi; jarno.vanhatalo@helsinki.fi;

Abstract

Ecological time series provide opportunities for investigating temporal dynamics across different levels of biological organization and can reveal distinct cyclical frequency signatures. Time-series clustering allows the identification of shared fluctuation patterns among ecological systems and consequently help explaining these patterns. However, missing data, heterogeneous sampling, and uncertainty present major challenges for clustering ecological time series in practice. We present a novel model-based approach for clustering time series. Our framework uses frequency domain analysis (Fourier analysis) to identify temporal frequency signatures and assign time series to clusters. The model is implemented within a hierarchical Bayesian framework, enabling the identification of clusters while accounting for uncertainty and incorporating covariates that explain cluster membership. This approach enables ecologists to group communities or populations with similar fluctuation dynamics and investigate the potential drivers underlying differences among groups. We evaluated the proposed method using a simulation study, which showed that the framework can successfully recover distinct frequency profiles. We then applied the method to Finnish moth monitoring data collected between 1993 and 2022, including 162 moth species in a

species-level analysis and 49 moth monitoring sites in a community-level analysis. The method identified distinct temporal frequency profiles among species and communities and revealed associations between fluctuation dynamics and ecological characteristics. Our framework provides a flexible tool for studying temporal ecological dynamics, identifying groups with similar fluctuation patterns, and investigating the factors associated with those patterns in long-term monitoring datasets.

Keywords: Temporal dynamics, Frequency profiles, Flucutation patterns, Time-series clustering, Model-based clustering, Fourier analysis, Community stability.

1 Introduction

Ecological time-series can reveal temporal dynamics at populations, communities, and ecosystem levels, as well as the drivers of these dynamics. The dynamics in these systems often reveal periodic or cyclic events, and the mechanisms that produce these fluctuations have been a long subject of studies in ecology, especially at the population level (e.g. [1–3]). The fluctuation in population size, for example, can be explained by intraspecific density-dependent variation [1, 4], consumer-resource interactions, responses to environmental fluctuation and even random exogenous forces that can cause cycles [3]. As such, ecological systems can be viewed as an ensemble of oscillators, which present cycles in a variety of frequencies and amplitudes [2, 5, 6]. Spectral analysis is a powerful tool for examining patterns and periodicities in time series data because it decomposes the oscillatory dynamics at different frequencies and reveal cyclical patterns that may not be apparent in raw data [2, 7–10]. A spectral profile – describing the dominant frequencies and their relative importance – can be compared among populations, species, communities, or any ecological metric of interest. Ultimately, these time-series can be aggregated into groups of similar temporal dynamics.

Spectral analysis in ecology has been used to address spatiotemporal synchrony in the variation of species and communities [11–13], as well as to classify time series into similar periodicity groups based on their spectral profiles [7, 14, 15].

To identify these similarities, a common approach is to cluster ecological time series either using correlation-based methods, which capture similarities in temporal patterns regardless of scale [14, 16], or by their frequency properties, which emphasize periodic or oscillatory dynamics across species or sites [7]. The benefit from the latter approach is that frequency profiles can inform about potential shared drivers and dependencies among time-series.

Key challenges in spectral analyses of ecological data are, however, the scarcity of continuous long-term time series, restricting the number of systems available for these studies, as well as the heterogeneity in sampling efficiency and observation processes. Ecological time series commonly present temporal and spatial gaps in sampling, which need to be filled before applying common time series analysis tools [17–19]. For example, traditional Fourier and wavelet analyses require non-interrupted time series to

make the mapping from time to frequency domain. Hence, in the context of missing data, most existing methods require an imputation step before spectral or frequency-based analyses can be conducted. Missing values are typically imputed using linear interpolation, k-nearest neighbor (kNN) imputation, or regression-based techniques; however, these approaches can perform poorly for infrequently sampled or nonlinear ecological time series, motivating the use of dynamical approaches such as empirical dynamic modeling [17–19]. Moreover, sequential approaches to analysis and data standardization often fail to account for the uncertainties inherent in ecological datasets, potentially propagating errors through downstream analyses [18, 19].

In this work, we propose a model-based Bayesian framework for clustering ecological time-series data with particular attention to missing data, sampling error, and uncertainty estimation. Our method differs from existing approaches by automatically handling missing data without the need for imputation, providing uncertainty estimates, and incorporating additional covariates, such as functional traits and environmental variables, to explain (and predict) clusters of similar frequency profiles. Our approach builds on Fourier analysis, in which a time series is decomposed into a sum of sinusoidal components at different frequencies. The resulting representation, known as the periodogram or spectral density function [20], partitions the total variance of a time series across frequencies. Peaks in the periodogram indicate the frequencies that contribute most to the variance of the time series, allowing the identification of dominant oscillatory behaviors in the data. In this way, the periodogram serves as a tool for detecting underlying periodicities and provides information on the cyclical structure of the data [8].

However, unlike in classical Fourier analysis, we treat the Fourier expansion as a latent process that describes the variability of (log) abundance of an ecological entity (e.g., population or community). To cluster species or communities to similar latent processes, we construct a number of Fourier expansions each with different spectral density – to be called a spectral profile hereafter. Each species or community is then assigned weights measuring how likely their observed fluctuations are linked to the different spectral profiles. To account for the observation error, the observed abundances are linked to the underlying latent processes through an observation model accounting for variation in sampling effort. Finally, we implement all the steps within a single hierarchical Bayesian model and conduct inference using Markov chain Monte Carlo to sample from the posterior distribution for the model parameters.

We demonstrate and test our methods with a simulation study. Subsequently, we apply the method to a range of ecological questions to illustrate its utility in real-world scenarios. Using long term moth populations data from Finland, we ask: (i) what are the dominant fluctuation frequencies of moth species in Finland and are they related to life-history traits using community abundance data; (ii) What are the frequency profiles of moth community abundance fluctuations, how are these related to climate conditions, and do they show spatial patterns.

2 Motivating case study

To demonstrate and test our method, we used long-term macro moth monitoring data from Finland. Among insect taxa, moths are one of the few groups supported by extensive long-term monitoring programs. Previous studies have documented pronounced fluctuations in total moth abundance that occur at roughly decadal (∼10-year) intervals [21]. Hence, prior evidence indicates the presence of a 10-year cycle in moth total abundance dynamics. This periodical abundance cycle appears to occur simultaneously across several common species during peak years [22].

2.1 Finnish Moth monitoring data

The Finnish moth monitoring scheme has been carried out since 1993 by the Finnish Environmental Institute (Syke) [21]. The monitoring scheme covers all Finland and a total of 297 unique sites. The number of traps deployed differs each year and also varies from the southern to northern regions. The traps are typically emptied weekly from April to October, although the interval varies within and between sites depending on logistical reasons. Experienced voluntary lepidopterists identify the species and count the individuals collected. The moth dataset comprises 1,721 taxa (micro- and macro-moths). Of these, 82 are identified only to genus level, while the remaining taxa are identified to species level. We used these data to study moth fluctuations at two levels of organization: species and communities. First, to analyze similarities in the fluctuations of moth species in Finland, we compiled the moth data into species-specific population estimates over the country for years 1993-2022. We used the yearly species-specific aggregated numbers over all sites as population observations. The total observation effort was correspondingly the total number of observation days over all sites per year. We selected species for which functional trait values (see Section 2.2) were available. As a result, we analyzed the time series of 162 species. Second, to analyze community abundance fluctuations in moth communities across Finland, we aggregated the yearly abundance (measured as biomass) of all species present in a site for years 1993-2022. For this analysis, we selected 49 sampling sites shown in (Figure 1), which have at least 15 different years with data during the study period which is the minimum number of observed time points to be able to detect 15-year-cycles [23]. As covariates, we included climatic and habitat variables (see Section 2.3).

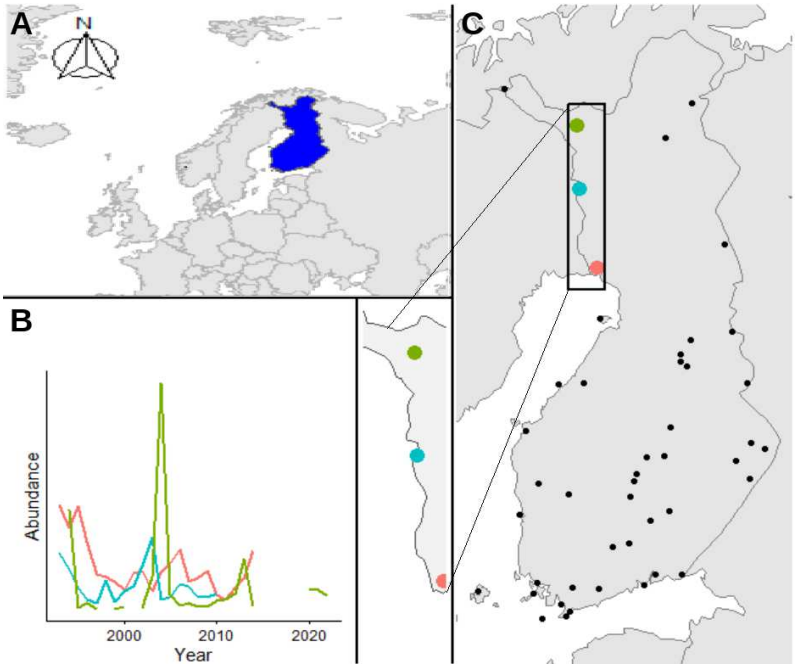


Fig. 1: The study area (A), Finland in the northern part of Europe, included 49 moth sampling sites (black dots in B) selected to have at least 15 observations in their corresponding time series. The rectangle in B indicates the north-western region, which is enlarged in inset C, illustrating the temporal fluctuations in abundance observations for three different moth communities from 1993 to 2022. A clear gap is observed in the time series between 2014 and 2020, resulting from missing data.

2.2 Moth trait data

Divergent life histories and ecological strategies, and external factors like parasitism, lead to variation in species abundances and distribution. Ecological and life-history traits, such as phenology, habitat affinity, and feeding guild, can predict temporal trends in moth biomass and species-specific responses to climatic variation [24]. Life-history traits have also been shown to influence moth responses to landscape fragmentation and environmental heterogeneity, illustrating that taxa with different strategies exhibit different ecological dynamics [25]. These patterns reflect the broad ecological diversity and morphological and developmental variation observed across moth species [26, 27]. Since fluctuations in population abundances are influenced by life-history traits we investigated how moth species fluctuation profiles are affected by life-history traits and niche characteristics. We expected to find a clear relationship of species groups and species traits. We used a set of species-level traits that might be related to abundance fluctuation: female body size (average wingspan), overwintering stage (adult, catepillar, egg, and pupa), and larval diet breadth within host type (monophagous, oligophagous, polyphagous, and mushroom) [24]. We also included

niche characteristics, as they can influence species responses to environment, defined as the species temperature index (STI-range) and the species precipitation index (SPI-mean and SPI-range). The species temperature and precipitation indexes summarize the mean (SPI-mean) and the range (STI-range and SPI-range) of the average annual temperature and precipitation over the distribution area of the species in Europe [28, 29].

2.3 Climatic and habitat covariates for moth communities

When classifying community-level fluctuation profiles into clusters, we used, as environmental covariates, climatic and habitat variables that have been described as important for moth communities [30–32]. To describe climatic variation, we used growing degree-days, average winter snow depth, and precipitation. Growing degree-days were calculated as the cumulative sum of daily mean temperatures exceeding 5°C during the sampling season (from 1 April to mid-October). Cumulative summer precipitation was computed over the same period using daily precipitation records. Average winter snow depth was derived from daily snow depth measurements taken during preceding winter period (15 October to 1 April). To describe habitat characteristic around the sites location we used the proportional cover of four major habitat types: broad-leaved forest, coniferous forest, mixed forest, and semi-natural or herbaceous habitats. In addition to the proportion of habitat amount, we computed two landscape indices, marginal entropy index and clumpiness index, using *landscapemetric* R package. Marginal entropy index quantifies the habitat diversity, while clumpiness quantifies fragmentation of forest habitats. Habitat information was extracted from CORINE Land-cover maps [33] for the years 2000, 2006, 2012 and 2018, in buffers of 500×500 m^2 around sampling sites. All environmental variables were summarized over the time series as temporal mean and temporal standard deviation.

3 Methods

3.1 Frequency Analysis

According to the fundamental principles of Fourier analysis, any arbitrary analytical function defined over a finite interval of the real line can be approximated to any desired degree of accuracy by a weighted sum of sine and cosine functions with harmonically increasing frequencies [23]. Therefore, for any time series $x_1, x_2, \cdots x_T$ where T is even, we can express the function as follows:

$$x_t = a_0 + \sum_{j=1}^{T/2} \left[a_j \cos\left(2\pi f_j t\right) + b_j \sin\left(2\pi f_j t\right) \right], \tag{1}$$

where a_0 is the average of the time series, while a_j and b_j represent the amplitudes for the sine and cosine components, respectively, with frequencies given by $f_j = \frac{j}{T}$. These frequencies are referred to as Fourier frequencies or fundamental frequencies. The j^{th} frequency corresponds to a period of $1/f_j = T/j$ indicating that a sinusoid with this

frequency completes j cycles over the duration of the data. Consequently, few periods will be integers, although all are rational.

The Fourier series can also be rewritten as

$$x_t = a_0 + \sum_{j=1}^{T/2} c_j \sin(2\pi tj/T + \phi_j) \tag{2}$$

where $c_j = \sqrt{a_j^2 + b_j^2}$ is the total amplitude of the j^{th} frequency component and $\phi_j = \tan^{-1}\left(\frac{a_j}{b_j}\right)$ is the phase shift for this basis function. Large values of c_j indicate predominant frequencies in the series, while small values of c_j may be associated with noise [23].

3.2 Model-based clustering of time series/ Model specification

Time-series clustering methods aim to partition observed time series into K groups that are more similar to each other than they are to observations in different groups. In our model-based time series clustering, the clusters are found by assigning each time series a latent group label. Formally, we define the latent group label for the ith observed time series as a K vector $z_i = (z_{i1}, z_{i2}, \cdots, z_{ik})$ with the k^{th} element equal to "1" if the observation belongs to group k and "0" otherwise. The groups are mutually exclusive such that the observation is assigned to only one of the K groups. The latent group label, z_i, is then modeled with a categorical distribution with mean $\theta_i = [\theta_{i,1}, \ldots, \theta_{i,K}]^T$, where $\theta_{i,k} > 0$ for $k = 1, 2, \ldots, K$, $i = 1, \ldots, n$, and $\sum_{k=1}^{K} \theta_{i,k} = 1$.

We then assume that, conditional on the latent group label z_i, the relative changes in the expectation for the observed data over time are the same among the time series. That is, suppose that we have n time series $y_1, \ldots, y_n$ of observed numbers of individuals where the ith time-series in the time interval $[1, T]$ is $y_i = [y_{i,1} \ldots y_{i,T}]^T$ and in case of missing observation at time t the observation is encoded as $y_{i,t} = \emptyset$ (that is "NA" in data tables) so that for all t and i the observation $y_{i,t} \in N_0 \cup \emptyset$, then the counts of individuals in a time series over the time interval $[1, T]$ have conditional element-wise expectation $\mathrm{E}[y_{i,t}|z_{ik} = 1] = \lambda_{0,i} \times \lambda_{k,t}$, where $\lambda_{k,t}$ remains the same over time series belonging to the same cluster but $\lambda_{0,i}$, controlling the absolute magnitude of variation, can vary between time series. The actual distribution of the counts is modeled with Poisson (or Negative-Binomial) distribution

$$y_{i,t}|z_{ik} = 1 \sim \text{Poisson}(\lambda_{0,i} \times \lambda_{k,t}). \tag{3}$$

We then model the logarithm of the conditional expectation with Fourier basis expansion, such that, assuming T is even,

$$\log \lambda_k(t) = \sum_{j=1}^{T/2} c_{jk} \sin(2\pi tj/T + \phi_{jk}). \tag{4}$$

We will call the vector of frequency amplitudes $c_k = [c_{k,1}, \ldots, c_{k,T/2}]^T$ *a frequency profile* of cluster k. Since all observations within a cluster have the same time-wise conditional expectation, they satisfy the requirement that observations within a cluster are more similar than observations in a different cluster.

Marginalizing over the group labels gives the likelihood of abundance observations as:

$$P(y_i|\theta) = \sum_{k=1}^{K} \theta_{k,i} \prod_{t=1}^{T} \text{Poisson}(\lambda_{0,i} \times \lambda_{k,t}). \tag{5}$$

We include covariates into the model, to explain cluster identity of a time series, by modeling the cluster probabilities with logit regression

$$\text{logit } \theta_{i,k} = \alpha_k + \beta_k^T x_i \tag{6}$$

where α_k, β_k are the cluster specific intercept and regression coefficients and x_i is the vector of covariates associated with time series i.

The model can also take into account the effect of varying sampling effort by including it into the observation model as an offset so that

$$y_{i,t}|z_{ik} = 1 \sim \text{Poisson}(V_{i,t} \times \lambda_{0,i} \times \lambda_{k,t}), \tag{7}$$

where $V_{i,t}$ is the sampling effort of an observation from time series i in year t.

3.3 Prior specification

To promote sparsity in the frequency profiles, and to mitigate label switching, we employed weakly informative shrinkage priors on frequency amplitude components. In particular, the group inverse-gamma–gamma [GIGG; 34] prior provides adaptive regularization through a hierarchical structure that shrinks small coefficients strongly toward zero while allowing large signals to remain relatively unshrunk. Hence, we gave the amplitude components $c_{j,k}$ GIGG priors that shrink unnecessary frequency components towards zero so that

$$c_{j,k} \sim N(0, \tau^2 \gamma_k^2 \eta_{kj}^2)$$

where $\gamma_k \sim \text{Gamma}(0.5, 1)$, $\eta_{kj} \sim \text{Inv-Gamma}(0.5, 1)$, $\tau \sim \text{Cauchy}(0, \sigma_\tau)$ and $\sigma_\tau \sim \text{Student-}t(4, 0, 1)$. Here, τ is a global shrinkage parameter pulling the variability in log intensity towards zero while γ_k pulls the variability explained by the k'th cluster towards zero and η_{kj} pulls the weight of each frequency component towards zero.

We then specified weakly informative priors for hyperparameters. Prior for the phase was chosen to be $\phi \sim \mathbf{N}(0, \frac{\pi}{2})$ with lower bound of $-\pi$ and upper bound of π. Time series wise intercepts were given Gaussian priors, $\lambda_{0,i} \sim \mathbf{N}(0, 2)$.

We implemented two types of models for the group membership probability: a covariate informed model (6) and model where group membership probabilities do not depend on covariates. In the latter case, prior for time series wise cluster probabilities follow Dirichlet distribution with $\theta_i \sim \text{dirichlet}(\alpha_1, \ldots, \alpha_K)$ where $\alpha_1 = \cdots = \alpha_K =$

0.01, so that the prior enforces sparsity of group membership probabilities [34]. In the model with covariate informed cluster probabilities, we used a hierarchical shrinkage prior similar to the prior for frequency profiles. Let s denote a covariate, and α_k and $\mathbf{B}_k$ the cluster-specific intercept and regression coefficients, respectively. Their priors are defined as $\alpha_k \sim N(0, 0.5)$ and $\beta_k \sim \text{Normal}(0, \tau_s\gamma_{sk})$ where $\tau_s \sim \text{Cauchy}(0, \sigma_s)$, $\gamma_{sk} \sim \text{Gamma}(0.5, 1)$, and $\sigma_s \sim \text{Student-}t(4, 0, 1)$.

3.4 Posterior sampling

Posterior inference was performed using Markov chain Monte Carlo (MCMC) sampling implemented using Stan via the `rstan` interface in R. The sampled parameters included the cluster-specific parameters governing the temporal dynamics, the mixture weights determining the relative contribution of each cluster, and the parameters associated with the observation models. Posterior samples of the latent cluster allocation variables were also obtained, allowing posterior probabilities of cluster membership to be calculated.

Due to the label-switching invariance inherent in model-based clustering, we ran for each model fit a single Markov chain consisting of sufficiently large number of iterations (at least 4000 iterations). The first 2000 iterations of each run were discarded as warm-up, and the remaining samples were retained for posterior inference. Because the mixture components are exchangeable, label switching was addressed using the relabeling procedure proposed by [35]. This procedure identifies a common labeling of the mixture components across MCMC iterations by finding the permutation of component labels that minimizes the posterior expected Kullback–Leibler divergence between the estimated allocation probabilities and the allocation probabilities obtained from the individual MCMC samples. In this way, the procedure provides a consistent labeling of the components across MCMC iterations while retaining uncertainty in cluster membership [35]. The resulting relabeled posterior allocation probabilities were subsequently used to characterize the clustering structure. Posterior inference was then based on the retained, and relabeld, MCMC samples. Posterior means and posterior credible intervals were used to summarize parameter uncertainty.

3.5 Estimation of the effective number of clusters

To obtain an estimate of the effective number of clusters, we applied the Merge–Truncate–Merge (MTM) procedure of Guha and Nguyen [36] to the posterior samples of the mixing measure. The MTM procedure accounts for closely located mixture components and components with negligible mixing weights, thereby providing a reduced and interpretable representation of the posterior clustering structure. For each mixture component, the algorithm determines its associated mixing weight and component center. Components with mixing weights exceeding a prespecified threshold are retained as substantive components. Components whose weights fall below the threshold are considered negligible and are merged with the nearest substantive component, where proximity is determined by the minimum distance between their component centers. The mixing weight of each negligible component is then added to that of the corresponding nearest substantive component [36, 37].

3.6 Assessment of the model and methods using simulation studies

To test the method and examine the clustering performance of the model for a variety of different models and data scenarios, we conducted a set of simulation studies. In the first simulation (*No covariates* scenario), we tested whether the model can cluster time series and recover the true estimated values for the amplitudes of the sine and cosine functions. There were no covariates associated to cluster probabilities in this scenario but we clustered the time series based on their temporal characteristics only. We generated 100, 200, and 300 abundance time series of length of 20 years, and yearly sampling frequency, by first generating representative log densities for three clusters. Each representative time series was generated from a sum of amplitudes of different sine and cosine functions. The frequency profiles were made clearly different, and the time series intercepts were set to a constant value of 1.7, to generate an easy clustering task. For each time series, the observed counts were drawn from a Poisson distribution conditioned on the density. To assess whether the model can reproduce the representative time series and the cluster profiles, we fitted the clustering model to the simulated data and compared the posterior distribution of the parameter estimates with their true values. To account for randomness in the data generating process, we repeated this procedure 100 times.

Next, we tested whether the model is able to identify the relationship between covariates and the clustering (*With covariates* scenario). Again we simulated 100, 200, and 300 abundance time series with three clusters but now we generated the cluster probabilities using three covariates. For each time series we generated the covariates from a standard normal distribution. We then drew the intercept (α) and coefficients β for each cluster from standard normal distribution. Next, we calculated the cluster probabilities (θ) for each time series and used them to sample the cluster identity from a categorical distribution to induce uncertainty into how well covariates explain cluster membership. For each cluster, we generated a representative log density time series to which we added the site wise intercept, but now we made the cluster specific frequency profiles less distinct than in the first scenario to generate hard clustering task. Finally, for each site, we sampled counts for time points from 1 to 20 from the Poisson distribution with cluster specific density function. We then fitted our model to the data and compared the posterior distributions of the parameter estimates with their true values.

Finally, to assess that our model is able to identify cluster also in the presence of gaps in time series, we run a simulation study with sparse data (i.e., a *Sparse* scenario). For each ensemble size (100, 200, and 300 time series) we ran 100 simulation replicates so that in each replicate, 25% of the series were made sparse by randomly removing up to 10 of the 20 annual observations (i.e., the missing time points were selected at random for each sparse series). After this we again compared the posterior distributions of the parameter estimates with their true values.

4 Results

4.1 Simulation studies

In all three simulation scenarios, the coverage of the 95% posterior credible intervals matched well the true data generating parameters because most of them contained the true parameter value in close to corresponding proportion of the simulations (Table 1), indicating good inference performance of the model and parameter identifiability. The cluster membership was better estimated than the parameters governing the amplitude of the frequency components and the effects of the covariates on cluster identity. On average, the frequency amplitude estimates were very good, with coverage ranging from $89 - 95.1\%$ indicating strong estimation performance.

As the simulated clusters were well separated in the frequency space in the *No covariates* and *Sparse* scenarios, this indicates that the model is able to reliably identify the clusters at least when the clusters have clearly distinct frequency profiles. The posterior coverage for the amplitude parameters was good in all models and simulation scenarios as it was close to the theoretical optimum on 95%. However, the coverage for the covariate effects (β) was less than the optimum. At the same time, while the clustering accuracy was good in all models, it was lower in the *With covariates* models than in the other models (81-82% compared to $> 98\%$). These results reflect the fact that the simulated data were more challenging in the *With covariates* models.

Table 1: Performance of the models in simulated datasets. The simulation scenarios included N time series from data generating processes with full observed time series and no covariates to explain cluster membership (*No covariates*), full observed time series and covariates to explain cluster membership (*With covariates*), and sparsely observed time series (i.e., gaps in observations in some time series) and no covariates to explain cluster membership (*Sparse*). We report the average (and std) clustering accuracy (*Cluster accuracy*) and average coverage of the 95% posterior credible interval for frequency amplitudes (*Coverage c*) and fixed effects (*Coverage* β) over 100 simulation replicates.

		Accuracy%		Coverage (95%)	
Scenario	N	mean	std σ	c	β
No covariates	100	98.0	8.9	89.0%	–
No covariates	200	98.5	8	91.6%	–
No covariates	300	99.9	1.8	94.8%	–
With covariates	100	82.0	3.9	91.6%	79.4%
With covariates	200	81.4	2.5	93.9%	82.5 %
With covariates	300	81.5	2.3	93.0%	82.0%
Sparse	100	99.8	0.5	94.0%	–
Sparse	200	99.8	0.3	93.0%	–
Sparse	300	99.5	1.0	94.0%	–

Note: The generated frequency amplitudes in scenario (*No covariates*) are not similar to the (*Sparse*) case.

4.2 Fluctuations of moth species population abundances in Finland and their associations to species traits

To characterize the dominant fluctuation frequencies of moth species in Finland and relate them to life-history traits, we applied our model-based clustering with cluster-specific latent intensity functions defined in Equation (4). Species-specific cluster probabilities were expressed as in Equation (6) where x_i is the vector of trait values of species i (Section 2.2). We fitted models using three to five clusters. All models returned three clusters, meaning that models using four and five clusters returned one and two empty clusters respectively. Therefore, we show results for the model using three clusters.

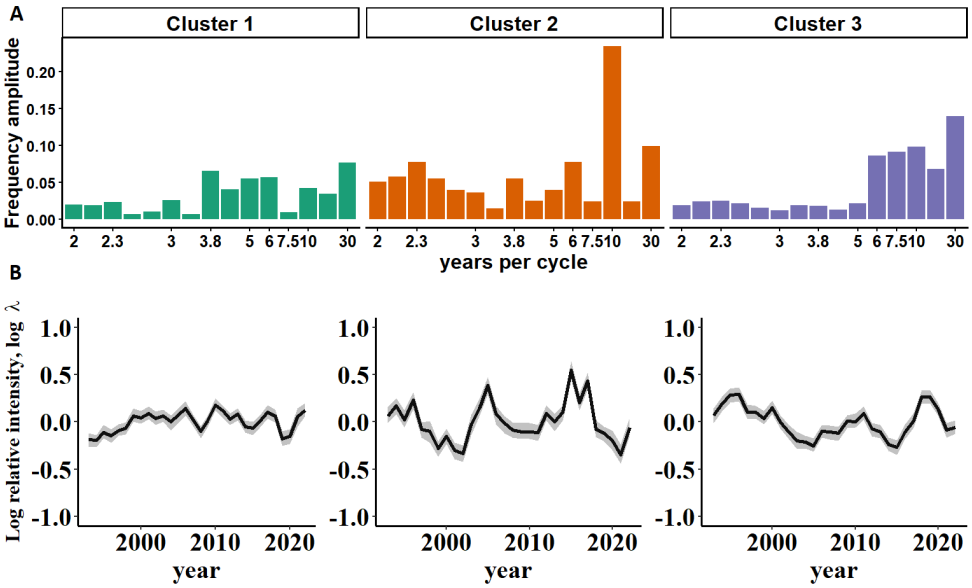


Fig. 2: (A) Frequency profiles of the three different groups of species identified by the clustering analysis. (B) The corresponding representative time series of each group.

Clusters 1 to 3 contained 62, 61, and 39 moth species, respectively. Probability of belonging to cluster 1 was on average 0.64 (sd 0.0707), to cluster 2 it was 0.617 (0.0693), and to cluster 3 it was 0.757 (0.0786) (Figures A1A–C).

Cluster 2 had the most dominant frequency component at 10 years per cycle. This most dominant component was the highest among all the clusters resulting in high variability in the representative time series for this cluster. In addition, cluster 2 presented subdominant frequency components at 2.3, 3.8, 6 years per cycle. Cluster 1 presented multiple frequency components with similar amplitude, from 3.8 to 6 and 10 years per cycle (Figure 2A). Cluster 3 had the dominant frequencies at longer cycles at six years per cycle and above. The representative time series of the three clusters

(Figure 2B), derived from the frequency profiles, show that clusters 1 and 3 have more stable fluctuations compared to the cluster 2 fluctuations.

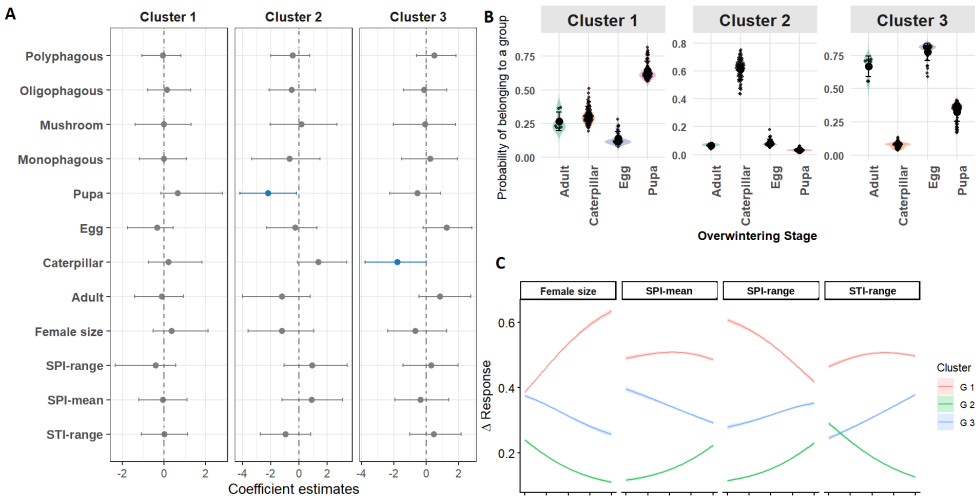


Fig. 3: (A) Estimated effects (posterior mean and 90% CI) of covariates on log odds of cluster probability: Larval diet within host type: (Monophagous, Oligophagous, Polyphagous, and Mushroom), Overwintering stage (adult, catepillar, egg, and pupa), Female size, species precipitation index (SPI-mean and SPI-range) and species temperature index (STI-range). (B) Species specific probabilities of belonging to a cluster (θ; black dots) among species with different stages of overwintering. (C) Conditional responses of species specific cluster probabilities along the continuous covariates: Female size, SPI-mean, SPI-range, and STI-range.

Species that overwinter as pupae were primarily associated with Cluster 1, those that overwinter as caterpillars with Cluster 2, and species that overwinter as eggs or adults had a high probability of belonging to Cluster 3 (Figure 3B). Overall, overwintering stage had the strongest effect on cluster membership (Figures 3A,B). Larval diet showed no significant effect, while female size and niche characteristics exhibited only marginal effects. In particular, species with larger females were more likely to belong to Cluster 1 and less likely to belong to Cluster 2. Additionally, species experiencing a narrower annual precipitation range had a lower probability of belonging to Cluster 2 (Figure 3C).

4.3 Fluctuations of moth communities in Finland and their associations to environmental covariates

To study the frequency profiles of moth community abundance fluctuations and how they are related to climate conditions, we applied our model defined in Equation (4). Site specific cluster probabilities are expressed as in Equation (6) where x_i is the vector of climatic and habitat covariates associated with site i (Section 2.3).

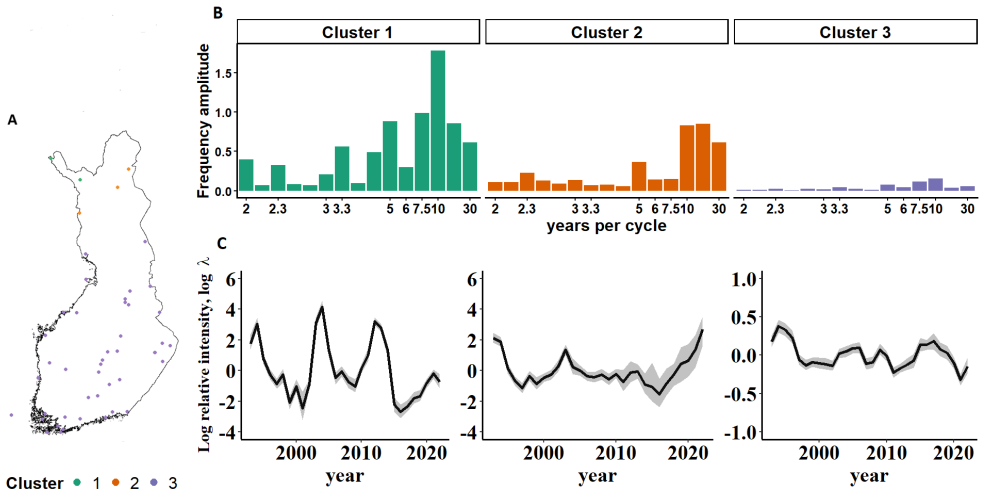


Fig. 4: (A) The study region (49 moth sites in Finland). (B) Frequency profiles of the three different groups of communities identified by the clustering analysis. (C) The corresponding representative time series of each group. Note that the y-axis limits differ between the figures in panel C.

The fluctuations of abundances in moth communities were classified into three clusters. Most of communities were classified as belonging to cluster 3 (44 sites), while cluster 1 and 2 presented only 2 and 3 sites, respectively. Probability of belonging to a group was high across groups (cluster 1 mean 0.705, sd 0.236, cluster 2 mean 0.575, sd 0.0731, cluster 3 mean 0.968, sd 0.0704). Across the clusters, the most dominant frequencies were related to longer cycles, but the amplitudes of the cycles were particularly large in clusters 1 and 2 while cluster 3 included sites with relatively stable communities presenting low fluctuations (Figures 4B,C). Cluster 1 had the largest amplitude in frequency of 10 years per cycle, followed by 7.5, 5, and 15 years per cycle while cluster 2 had the highest amplitudes at frequencies 15 and 10 years per cycle followed by 5 and 30 years per cycle (Figure 4B). Cluster 3 also had the highest amplitude at 10 years per cycle followed by 7.5 and 5 years per cycle. Spatially, cluster 3 represents sites from south to central boreal regions, while cluster 1 and 2 sites were located in the northern Finland (Figure 4A). This result could be partly explained by mass occurrences of *Epirrita autumnata* in northern Finland.

Regarding cluster association to environmental variables, belonging to cluster 2 had positive association with Broad-leaved forest and mean of Coniferous forest. For cluster 3, probability of belonging increases with mean precipitation and decreases with mean snow cover. The distinction between cluster 1 and 3 can be attributed to mean climatic differences (Figures 5A and 5B). Sites in cluster 3 have higher precipitation and lower mean snow cover, as well as less variability in broadleaf forest fragmentation (Clumpiness sd). Overall, the distinction in the fluctuation profiles among clusters seems to be more related to the temporal mean in the covariates than to the temporal standard deviation.

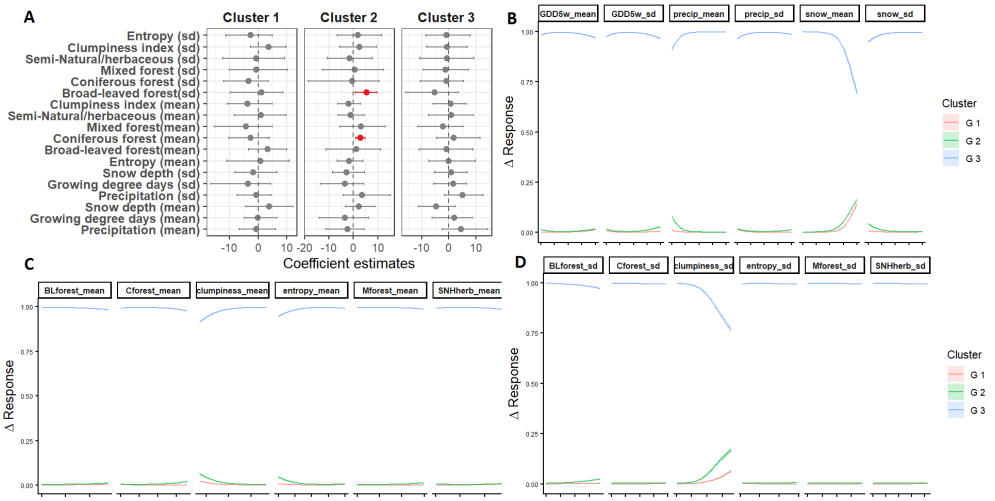


Fig. 5: (A) Estimated effect sizes (posterior mean and 95% CI) of climatic and habitat covariates on log odds of cluster probability. (B-D) Conditional responses of cluster probabilities along continuous covariates: mean and standard deviation of preciptation, growing degree days, and snow depth (B); mean proportion of Broad-leaved forest, Coniferous forest, mixed forest, and Semi-natural herbaceous forest as well as clumpiness index, and entropy (C); standard deviation in proportion of broad-leaved forest, Coniferous forest, Mixed forest, and Semi-natural herbaceous forest as well as clumpiness index, and entropy.

5 Discussion

Here, we presented a new model-based method for clustering ecological abundance time series that can be used in datasets that contain missing observations, accounts for uncertainty, and allows the inclusion of covariates to explain cluster probabilities. Our simulation study showed that the method is capable of recovering the cluster structure with high accuracy, demonstrating that the framework can identify distinct spectral profiles with and without the inclusion of covariates in the model. It therefore can be used to identify clusters of similar spectral profiles in time-series of count data, revealing temporal cyclic ecological patterns and consequently identifying drivers of periodic fluctuations, as our moth study case demonstrates.

A key advantage of the proposed approach is that it allows spectral analyses to be conducted in situations where traditional methods are difficult to apply. Ecological monitoring data often contain missing observations, uneven sampling effort, and observation error [18]. Common workflows typically require missing values to be imputed before spectral analyses can be performed, which may introduce additional uncertainty that is not accounted for in subsequent analyses [18, 19]. In addition, variables that may affect the difference between clusters are often assessed *a posteriori*, after clustering step. Our framework estimates the underlying temporal dynamics, cluster membership, and the effect of covariates on cluster membership while accounting

for uncertainty in a single model. This method is particularly suitable for long-term ecological monitoring datasets, where missing observations are common.

One of the benefits of clustering spectral profiles is that it can inform about shared drivers and dependencies among time-series. In our study case, using Finnish moth abundance data, we found that moth species can be grouped according to spectral profiles that are associated with different overwintering strategies. Overwintering stage can affect phenological timing (the peak of abundance of species over the summer season) across moths, in which abundance peaks corresponds to different overwintering stages [38, 39], as well as may produce distinct temporal fluctuation dynamics [39]. Our results indicate that overwintering may affect the periodicity in species abundance fluctuation. Moth species overwintering as caterpillar tend to have short abundance cycles of about 2.3 years, and bursts of abundance in cycles of 10 years. Species overwintering as pupa tend to have medium term cycles (3.8 and 6 years per cycle), while species overwintering as egg or adult tend to have longer cycles (7.5 and 10 years per cycles). This indicates that life-history trait may be related to temporal population fluctuation and highlights how traits can be used to explain cluster membership.

When focusing on community time-series our results captured difference in frequency profiles related to the latitudinal gradient. Communities had similar length of fluctuation cycles, since all clusters presented 10 years cycles as the most dominant frequency, and the second most dominant frequency was 7.5 for cluster 1 and 2. This is consistent with previous documentation of 10 years cycles in total moth abundance [21], [22] and was also found at the species-level clusters, where cluster 1 and 3 presented dominant frequencies at 10 year cycles. Despite the common 10 year cycle across moth communities, the key distinction between cluster is the difference in the amplitude of the higher amplitudes in cluster 1, 2 than in cluster 3. Northern moth communities have higher temporal variability and more pronounced abundance outbreaks than Central and Southern communities [see also 40], which might be associated to 10 years cycle outbreaks of *Epirrita autumnata* and *Operophtera brumata* in the Northern region [41–45]. We also found different fluctuation patterns between northern communities, which separated them into two distinct spectral profiles. In general, this example shows how the method can be used to identify large-scale spatial patterns in ecological dynamics.

The selection of the number of clusters is a common challenge in clustering methods. Traditionally the selection of the number of clusters in model-based clustering is guided by information-theoretic or likelihood-based criteria such as AIC or BIC [46, 47]. An alternative has been to use predictive performance, for example via leave-one-out cross-validation and the log-predictive density, which evaluate how well the model generalizes to unseen data [48]. These methods typically assume well-separated mixture components and may perform poorly when clusters overlap, leading to unstable estimates. Cluster models are also known to be *a posteriori* inconsistent for the number of clusters [49]. In particular, the posterior tends to favor small extra clusters even in large samples [49]. To address this issue, several approaches have been proposed. Redundant components can be shrunk with suitable priors [50], while post-processing methods, such as the Merge–Truncate–Merge (MTM) algorithm, remove or combine negligible clusters [36]. To obtain an estimate of the effective number of

clusters from the posterior distribution, we applied the MTM algorithm of [36]. The MTM algorithm is a post-processing procedure for posterior samples of the mixing measure. It is designed to address the presence of multiple closely located mixture components and components with negligible posterior mass, which may otherwise lead to an overestimation of the number of clusters. By merging redundant components and removing components with negligible mass, the MTM procedure provides a more stable and interpretable representation of the underlying clustering structure. In our study case, we fit models with the number of clusters ranging from three to five. In all models, after applying the MTM algorithm the number of clusters was found to be three, indicating that the MTM algorithm is a robust to selecting the number of clusters.

As future development, the model structure could be adapted to address different ecological questions beyond what was demonstrated here – for example, to evaluate the degree of synchrony between time-series in the same cluster. For this, the phase component in the model could be attributed per time-series and the estimates of the phase for each frequency can be used to evaluate the degree of synchrony between time-series. This would allow to evaluate compensatory dynamics from a frequency profile perspective (e.g. [51]). Another possibility is to cluster time-series of abundance and environmental fluctuations in the same model. Clusters grouping together time-series and environmental covariates could inform association of environmental fluctuation with abundance dynamics, or, in contrast, show independence of abundance dynamics from environmental factors.

Here, we propose a novel method to cluster ecological time series based on similarity in their frequency profiles and explanatory covariates that are linked to potential explanations of these frequency patterns. We employed Fourier analysis within hierarchical model based Bayesian approach to allow for uncertainty propagation and analysis of sparse data. The framework is flexible and can be used to identify common temporal dynamic patterns at population, species, or community levels.

Acknowledgements. The work was funded by the Jane and Aatos Erkko Foundation, Ministry of Environment, Finland (through the BIOMON research programme; N.S., A.R., and J.V.), the European Union via ERC Consolidator Grant (BEFPREDICT, 101087409, J.V.), the Research Council of Finland (decision 375360, A.R) and The Finnish Moth Monitoring scheme (Nocturna) which was supported by the Ministry of the Environment (Finland). We are grateful to all the volunteers and researchers who have collected and curated the data over several decades, and particularly Ida-Maria Huikkonen, Reima Leinonen, Guy Söderman and Liisa Tuominen-Roto.

Declarations

- Funding: The work was funded by the Jane and Aatos Erkko Foundation, Ministry of Environment, Finland (through the BIOMON research programme; N.S., A.R., and J.V.), the European Union via ERC Consolidator Grant (BEFPREDICT, 101087409, J.V.), the Research Council of Finland (decision 375360, A.R) and The

Finnish Moth Monitoring scheme (Nocturna) which was supported by the Ministry of the Environment (Finland).

- Conflict of interest/Competing interests: The authors declare no competing interests.
- Ethics approval and consent to participate: Not applicable.
- Consent for publication: Not applicable.
- Data availability: All data used in the work has been collected and published in earlier studies. Data are available at https://github.com/EnvStat/TSclustering.
- Code availability: Code are available at https://github.com/EnvStat/TSclustering.
- Author contribution: Conceptualization: Nourhan Shafik, Arthur Rodiguez, Jarno Vanhatalo. Methodology: Nourhan Shafik, Jarno Vanhatalo. Software: Nourhan Shafik. Data curation: Nourhan Shafik. Writing – original draft: Nourhan Shafik, Arthur Rodiguez, Jarno Vanhatalo. Writing – review and editing: Nourhan Shafik, Arthur Rodiguez, Pasi Sihvonen, Jarno Vanhatalo. Supervision: Jarno Vanhatalo. All authors read and approved the final manuscript.

Appendix A Additional results

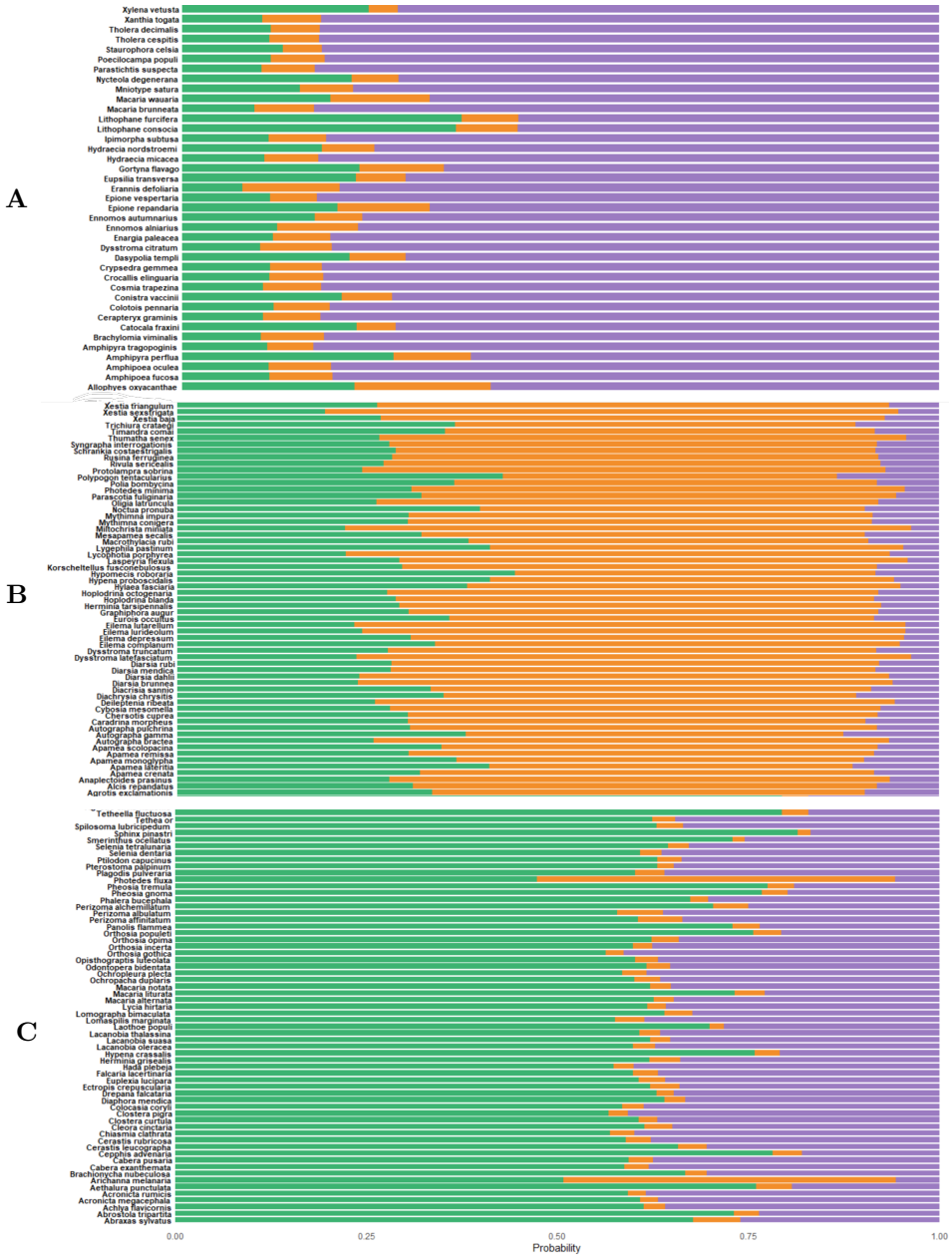


Fig. A1: Probability of each species to belonging to the identified clusters. Panels show species with high membership probability in each cluster: (A) species primarily assigned to Cluster 3; (B) species primarily assigned to Cluster 2; (C) species primarily assigned to Cluster 1.

References

[1] Moran, P.A.P.: The statistical analysis of the sunspot and lynx cycles. The Journal of Animal Ecology **18**(1), 115 (1949) https://doi.org/10.2307/1585

[2] Platt, T., Denman, K.L.: Spectral analysis in ecology. Annual Review of Ecology and Systematics **6**(1), 189–210 (1975) https://doi.org/10.1146/annurev.es.06.110175.001201

[3] Kendall, B.E., Briggs, C.J., Murdoch, W.W., Turchin, P., Ellner, S.P., Mccauley, E., Nisbet, R.M., Wood, S.N.: Why do populations cycle? a synthesis of statistical and mechanistic modeling approaches. Ecology **80**(6), 1789–1805 (1999)

[4] Lundberg, P., Ranta, E., Ripa, J., Kaitala, V.: Population variability in space and time. Trends in Ecology & Evolution **15**(11), 460–464 (2000) https://doi.org/10.1016/S0169-5347(00)01981-9

[5] Hajian-Forooshani, Z., Vandermeer, J.: Viewing communities as coupled oscillators: Elementary forms from lotka and volterra to kuramoto. Theoretical Ecology **14**(2), 247–254 (2021) https://doi.org/10.1007/s12080-020-00493-4

[6] Turchin, P., Taylor, A.D.: Complex dynamics in ecological time series. Ecology **73**(1), 289–305 (1992) https://doi.org/10.2307/1938740

[7] Benincà, E., Pinto, S., Cazelles, B., Fuentes, S., Shetty, S., Bogaards, J.A.: Wavelet clustering analysis as a tool for characterizing community structure in the human microbiome. Scientific Reports **13**, 8402 (2023) https://doi.org/10.1038/s41598-023-34713-8

[8] Cazelles, M. B.and Chavez, Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S., Stenseth, N.C.: Wavelet analysis of ecological time series. Oecologia **156**(2), 287–304 (2008) https://doi.org/10.1007/s00442-008-0993-2

[9] Reuman, D.C., Walter, J.A., Sheppard, L.W., Karatayev, V.A., Kadiyala, E.S., Lohmann, A.C., Anderson, T.L., Coombs, N.J., Haynes, K.J., Hallett, L.M., Castorani, M.C.N.: Insights into spatial synchrony enabled by long-term data. Ecology Letters **28**(4), 70112 (2025) https://doi.org/10.1111/ele.70112

[10] Yoh, N., Haley, C.L., Burivalova, Z.: Time series methods for the analysis of soundscapes and other cyclical ecological data. Methods in Ecology and Evolution **15**(7), 1158–1176 (2024) https://doi.org/10.1111/2041-210X.14361

[11] Ghosh, S., Cottingham, K.L., Reuman, D.C.: Species relationships in the extremes and their influence on community stability. Philosophical Transactions of the Royal Society B: Biological Sciences **376**(1835), 20200343 (2021) https://doi.org/10.1098/rstb.2020.0343

[12] Reuman, D.C., Castorani, M.C.N., Cavanaugh, K.C., Sheppard, L.W., Walter, J.A., Bell, T.W.: How environmental drivers of spatial synchrony interact. Ecography **2023**(10), 06795 (2023) https://doi.org/10.1111/ecog.06795

[13] Shoemaker, L.G., Hallett, L.M., Zhao, L., Reuman, D.C., Wang, S., Cottingham, K.L., Hobbs, R.J., Castorani, M.C.N., Downing, A.L., Dudney, J.C., Fey, S.B., Gherardi, L.A., Lany, N., Portales-Reyes, C., Rypel, A.L., Sheppard, L.W., Walter, J.A., Suding, K.N.: The long and the short of it: Mechanisms of synchronous and compensatory dynamics across temporal scales. Ecology **103**(4), 3650 (2022) https://doi.org/10.1002/ecy.3650

[14] Dove, S., Böhm, M., Freeman, R., Jellesmark, S., Murrell, D.J.: A user-friendly guide to using distance measures to compare time series in ecology. Ecology and Evolution **13**(10), 10520 (2023) https://doi.org/10.1002/ece3.10520

[15] Marques, A.R., Forde, H., Revie, C.W.: Time-series clustering of cage-level sea lice data. PLOS ONE **13**(9), 0204319 (2018) https://doi.org/10.1371/journal.pone.0204319

[16] Berthold, M.R., Höppner, F.: On clustering time series using euclidean distance and pearson correlation (2016). arXiv:1601.02213

[17] Johnson, B., Munch, S.B.: An empirical dynamic modeling framework for missing or irregular samples. Ecological Modelling **468**, 109948 (2022) https://doi.org/10.1016/j.ecolmodel.2022.109948

[18] Bowler, D.E., Boyd, R.J., Callaghan, C.T., Robinson, R.A., Isaac, N.J.B., Pocock, M.J.O.: Treating gaps and biases in biodiversity data as a missing data problem. Biological Reviews **100**, 50–67 (2024) https://doi.org/10.1111/brv.13127

[19] Carpenter, D., Deyle, E., Park, J., Saberski, E., Sugihara, G.: Repairing gaps in ecological time series. Methods in Ecology and Evolution **16**(3), 573–583 (2025) https://doi.org/10.1111/2041-210x.14491

[20] Chatfield, C.: The Analysis of Time Series: An Introduction, 6th edn. Chapman and Hall/CRC, Boca Raton, FL (2003)

[21] Huikkonen, I.-M., Korhonen, P., Leinonen, R., Suuronen, A., Pöyry, J.: Valtakunnallinen yöperhosseuranta 30 vuotta (1993–2022). Technical Report 26, Suomen ympäristökeskus (2024). https://helda.helsinki.fi/items/beca6891-f1d7-4e21-999d-98bce51025b1

[22] Kohonen, I.: No long-term change in total macro-moth abundance and biomass in Finland. Faculty of Biological and Environmental Sciences. https://helda.helsinki.fi/items/52e9a713-7a46-4b48-9f0e-80387648a32b

[23] Bloomfield, P.: Fourier Analysis of Time Series, 2nd edn. Wiley, New York (2000)

[24] Yazdanian, M., Kankaanpää, T., Itämies, J., Leinonen, R., Merckx, T., Pöyry, J., Sihvonen, P., Suuronen, A., Välimäki, P., Kivelä, S.M.: Ecological and life-history traits predict temporal trends in biomass of boreal moths. Insect Conservation and Diversity **16**(5), 600–615 (2023) https://doi.org/10.1111/icad.12657

[25] Slade, E.M., Merckx, T., Riutta, T., Bebber, D.P., Redhead, D., Riordan, P., Macdonald, D.W.: Life-history traits and landscape characteristics predict macro-moth responses to forest fragmentation. Ecology **94**(7), 1519–1530 (2013) https://doi.org/10.1890/12-1366.1

[26] Scoble, M.J.: The Lepidoptera: Form, Function, and Diversity. Oxford University Press, Oxford, UK (1995)

[27] Li, K.: Morphology, development, diversity, ecological functions, and threats to moths. International Journal of Biology and Life Sciences **13**(2), 35–41 (2026) https://doi.org/10.54097/wtqxhj88

[28] Ellis, E.E., Antão, L., Davrinche, A., Mäkinen, J., Rees, M., Conenna, I., Huikkonen, I.-M., Leinonen, R., Pöyry, J., Suuronen, A., Laine, A.-L., Saastamoinen, M., Vanhatalo, J., Roslin, T.: Recent community warming of moths in finland is driven by extinction in the north and colonisation in the south **16** (2025)

[29] Mäkinen, J., Ellis, E., Antão, L., Davrinche, A., Laine, A.-L., Saastamoinen, M., Conenna, I., Hällfors, M.H., Santangeli, A., Kaarlejärvi, E., Heliölä, J., Huikkonen, I.-M., Kuussaari, M., Leinonen, R., Lehikoinen, A., Pöyry, J., Suuronen, A., Salemaa, M., Tonteri, T., Vuorio, K.M., Skjelbred, B., Järvinen, M., Drakare, S., Carvalho, L., Welk, E., Seidler, G., Vangansbeke, P., Malis, F., Hédl, R., Auffret, A.G., Plue, J., De Frenne, P., Kalwij, J.M., Vanhatalo, J., Roslin, T.: Thermal homogenization of boreal communities in response to climate warming. PNAS **122**(17), 2415260122 (2025)

[30] Guilbault, E.P., Sihvonen, P., Suuronen, A., Huikkonen, I.-M., Pöyry, J., Laine, A.-L., Roslin, T., Saastamoinen, M., Vanhatalo, J.: Strong context dependence in the relative importance of climate and habitat on nation-wide macro-moth community changes. Journal of Animal Ecology **94**(10), 1948–1961 (2025) https://doi.org/10.1111/1365-2656.70107

[31] Pöyry, J., Leinonen, R., Söderman, G., Nieminen, M., Heikkinen, R.K., Carter, T.R.: Climate-induced increase of moth multivoltinism in boreal regions. Global Ecology and Biogeography **20**(2), 289–298 (2011) https://doi.org/10.1111/j.1466-8238.2010.00597.x

[32] Pöyry, J., et al.: Climate and habitat as drivers of moth community dynamics. Ecology Letters (2018)

[33] Feranec, J., Soukup, T., Hazeu, G., Jaffrain, G. (eds.): European Landscape Dynamics: CORINE Land Cover Data. CRC Press, Boca Raton (2016). https:

//doi.org/10.1201/9781315372860

[34] Boss, J., Datta, J., Wang, X., Park, S.K., Kang, J., Mukherjee, B.: Group inverse-gamma gamma shrinkage for sparse linear models with block-correlated regressors. Bayesian Analysis **19**(3), 785–814 (2024) https://doi.org/10.1214/23-BA1371

[35] Stephens, M.: Dealing with label switching in mixture models. Journal of the Royal Statistical Society Series B (Statistical Methodology) **62**(4), 795–809 (2000)

[36] Guha, S., Nguyen, X.: On posterior contraction of parameters and interpretability in bayesian mixture modeling. Bernoulli **27**(4), 2159–2188 (2021)

[37] Alamichel, L., Bystrova, D., Arbel, J., Kon Kam King, G.: Bayesian mixture models (in)consistency for the number of clusters. Scandinavian Journal of Statistics. **51**(4), 1619–1660 (2024) https://doi.org/10.1111/sjos.12739

[38] Neff, F., Chittaro, Y., Korner-Nievergelt, F., Litsios, G., Rey, E., Knop, E.: Moth communities are shaped by season, weather, elevation, and landscape composition. Insect Conservation and Diversity **18**(4), 670–680 (2025) https://doi.org/10.1111/icad.12835

[39] Yazdanian, M., T., K., Merckx, T., Huikkonen, I.-M., Itämies, J., Jokimäki, J., al: Evidence for bottom-up effects of moth abundance on forest birds in the north-boreal zone alone. Ecology Letters **27** (2024) https://doi.org/10.1111/ele.14467

[40] Antão, L., Pöyry, J., Leinonen, R., Roslin, T.: Contrasting latitudinal patterns in diversity and stability in a high-latitude species-rich moth community. Global Ecol Biogeogr. **29** (2020) https://doi.org/10.1111/geb.13073

[41] Tenow, O.: The outbreaks of *oporinia autumnata* bkh. and *operophtera* spp. (lep., geometridae) in the scandinavian mountain chain and northern finland 1862–1968. Zoologiska Bidrag från Uppsala (Suppl. 2), 1–107 (1972)

[42] Ruohomäki, K., Tanhuanpää, M., Ayres, M.P., Kaitaniemi, P., Tammaru, T., Haukioja, E.: Causes of cyclicity of *epirrita autumnata* (lepidoptera, geometridae): grandiose theory and tedious practice. Population Ecology **42**(3), 211–223 (2000)

[43] Jepsen, J.U., Hagen, S.B., Ims, R.A., Yoccoz, N.G.: Climate change and outbreaks of the geometrids *operophtera brumata* and *epirrita autumnata* in subarctic birch forest: evidence of a recent outbreak range expansion. Journal of Animal Ecology **77**(2), 257–264 (2008)

[44] Tenow, O., Nilssen, A.C., Bylund, H., Hogstad, O.: Waves and synchrony in *epirrita autumnata/operophtera brumata* outbreaks. I. lagged synchrony: regionally, locally and among species. Journal of Animal Ecology **76**(2), 258–268 (2007)

[45] Bjørnstad, O.N., Ims, R.A., Lambin, X.: Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends in Ecology & Evolution **14**(11), 427–432 (1999)

[46] Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6), 716–723 (1974)

[47] Schwarz, G.: Estimating the dimension of a model. Annals of Statistics **6**(2), 461–464 (1978)

[48] Vanhatalo, J., Foster, S.D., Hosack, G.R.: Spatiotemporal clustering using gaussian processes embedded in a mixture model. Environmetrics **32**(7), 2681 (2021) https://doi.org/10.1002/env.2681

[49] Miller, J.W., Harrison, M.T.: Inconsistency of pitman–yor process mixtures for the number of components. Journal of Machine Learning Research **15**, 3333–3370 (2014)

[50] Rousseau, J., Mengersen, K.: Asymptotic behaviour of the posterior distribution in overfitted mixture models. Journal of the Royal Statistical Society: Series B **73**(5), 689–710 (2011)

[51] Vasseur, D.A., Gaedke, U.: Spectral analysis unmasks synchronous and compensatory dynamics in plankton communities. Ecology **88**(8), 2058–2071 (2007) https://doi.org/10.1890/06-1899.1

[52] Leinonen, R., Pöyry, J., Söderman, G., Tuominen-Roto, L.: Suomen yöperhosseuranta (nocturna) 1993–2012. Technical Report 15/2016, Suomen ympäristökeskuksen raportteja, Helsinki, Finland (2016)

[53] Vehtari, A., Gelman, A., Gabry, J.: Practical bayesian model evaluation using leave-one-out cross-validation and waic. Statistics and Computing **27**, 1413–1432 (2017)