

Delimiting synchronous populations from monitoring data

Christophe Giraud · Romain Julliard ·
Emmanuelle Porcher

the date of receipt and acceptance should be inserted later

Abstract We propose to investigate spatial synchrony in population dynamics from monitoring data. We develop a statistical procedure to delineate populations of sites with synchronous dynamics from short time series. The procedure relies on a new norm, the Synchronous Total Variation norm, which promotes synchrony in the estimation of the sites dynamics. The method is tested on some synthetic data sets and is applied on data from the French breeding bird monitoring program.

Keywords Synchronous population · monitoring data · penalized log-likelihood · ℓ^1 -penalty · primal-dual optimization

1 Introduction

Spatial synchrony in population dynamics, the spatial autocorrelation of temporal variations in the abundance of geographically disjoint populations, is a widespread phenomenon that has received significant attention from ecologists since the 1990s (see Liebhold *et al.* 2004 [10] for a review). The primary mechanisms underlying spatial synchrony are thus generally well-known: (1) spatial synchrony can be an endogenous phenomenon attributable to dispersal among populations (Ranta *et al.*

C. Giraud
CMAP, UMR CNRS 7641, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France
Tel.: +33 1 69 33 45 92
Fax: +33 1 69 33 46 46
E-mail: christophe.giraud@polytechnique.edu

R. Julliard
CERSP, UMR CNRS 7204, Muséum National d'Histoire Naturelle, 61 rue Buffon, 75005 Paris, France
E-mail: julliard@mnhn.fr

E. Porcher
CERSP, UMR CNRS 7204, Muséum National d'Histoire Naturelle, 61 rue Buffon, 75005 Paris, France
E-mail: porcher@mnhn.fr

1995 [15]); (2) it may result from trophic interactions in which predator-prey dynamics are synchronized by predator mobility (Hassell 2000 [7]); and (3) it may be driven by biotic (e.g. resources) or abiotic (e.g. weather or climate) synchronous exogenous factors. The latter phenomenon of environmental forcing is known as the Moran effect (Moran 1953 [12]).

Distinguishing between these mechanisms causing spatial synchrony in population dynamics would be of paramount importance to understand the contribution of biotic and abiotic factors to the regulation of populations and to explore the impact of global change on population dynamics and probabilities of extinction (Heino *et al.* 1997 [8]), but this remains a difficult task (Liebhold *et al.* 2004 [10]). A promising approach, which has received little attention so far (but see Ranta & Kaitala 1997 [14]), would be to explore how spatial synchrony in population dynamics varies through space. Specifically, we expect that the identification of areas with synchronous dynamics within each of them, but with contrasting dynamics among them will provide critical information to assess the spatial scale of synchrony and improve our understanding of the relationship between climate and population dynamics.

Exploring the spatial heterogeneity of synchrony in population dynamics however calls for new statistical tools. Most methods that were developed to analyse spatial synchrony are directed toward long time-series data over a limited number of study sites, and often target populations with oscillatory dynamics (phase synchrony analysis, e.g. Grenfell *et al.* 2001 [6]). With such design, it is usually possible to identify with certainty the best grouping of sites for which sites within a group are more synchronic than sites from other groups, due to the limited number of combination of sites and the accuracy of the estimation of synchronization. Another situation occurs when one wants to delineate frontiers across a dense number of sites to separate asynchronous populations of sites. Such data are available through extensive biodiversity monitoring programmes that were generally launched at best after the 1992 Earth summit; there are thus characterized by a potentially large number of monitored sites but a limited time span (one or two decades, Schmeller *et al.* 2009 [18]). Then, there is way too many possibilities of grouping sites to explore them systematically, and quite much imprecision in assessing synchrony among sites due to short time series (and possibly sampling variance).

Here, we present a new statistical method to identify geographical regions with synchronous population dynamics from abundance monitoring data. The identification of the regions is achieved by minimizing a penalized log-likelihood criterion. This criterion is the sum of the negative log-likelihood and a convex penalty which promotes synchronous dynamics. We propose to penalize proportionally to a new norm, the Synchronous Total Variation norm, which is related to the popular Total-Variation norm used for spatial segmentation (Chambolle *et al.* 2010 [4]). The resulting criterion is convex and the minimization can be performed efficiently with a primal-dual optimization scheme.

The method is tested using two contrasting datasets: (1) a synthetic dataset where the number of regions and their dynamics are known a priori and (2) real data from the French breeding bird monitoring program, where the abundance of a widespread passerine species is monitored over 9 years in 361 randomly distributed sites.

2 Model and estimation procedure

2.1 Monitoring data

Typical population monitoring data consist of abundance counting over multiple sites s at different dates t . Ideally, observations at site s and date t should be repeated (Schmeller 2008 [17]). In the following, we assume that we have at disposal such a data set, although our model does not require the existence of such repetitions. We write x_s for the location of the site s and Z_{stk} for the k th observation at site s and date t . These count data are assumed to be distributed according to a Poisson distribution.

2.2 Statistical model

We have in mind that each site s has its own mean abundance, but there exists some synchrony among the population-size variations. More formally, we expect a partition of the sites s into regions R_1, R_2, \dots , so that every sites from the same region R share the same temporal dynamic $t \rightarrow \rho_R(t)$. Therefore, in order to identify synchronous population, we ideally would like to fit the following model

$$\begin{aligned} Z_{stk} &\sim \text{Poisson}(\exp(\theta_s + f(x_s, t))) \\ \text{with } f(x, t) &= \sum_R \rho_R(t) \mathbf{1}_R(x) \text{ and } f(., 1) = 0, \end{aligned} \quad (1)$$

the regions R and the parameters $\theta_s, \rho_R(t)$ being unknown. The site effect $\exp(\theta_s)$ is proportional to the mean abundance of the site s .

A natural estimation procedure is to minimize over θ and f the negative log-likelihood of the observations Z

$$\mathcal{L}_Z(\theta, f) = \sum_{s,t,k} [e^{\theta_s + f(x_s, t)} - Z_{stk}(\theta_s + f(x_s, t)) + \log(Z_{stk}!)],$$

with the constraint that f is of the form $f(x, t) = \sum_R \rho_R(t) \mathbf{1}_R(x)$ with $f(., 1) = 0$. Unfortunately, the minimization

$$(\hat{\theta}, \hat{f}) \in \operatorname{argmin} \left\{ \mathcal{L}_Z(\theta, f) : \theta_s \in \mathbf{R}, f(., 1) = 0 \text{ and } f(x, t) = \sum_{j=1}^J \rho_j(t) \mathbf{1}_{R_j}(x) \right\} \quad (2)$$

is intractable in practice : while we can estimate $\hat{\theta}_s$ and $\hat{\rho}_j(t)$ efficiently when the regions R_j are known, the exhaustive search over all possible regions R_j is computationally very intensive and cannot be performed in high-dimensions. In such a situation, a classical approach is to relax the minimization Problem [20, 16, 4, 2, 3] and solve instead a convex problem for which we have powerful optimization tools.

2.3 Convex relaxation of Problem (2)

A nice feature of Problem (2) is that the functional $(\theta, f) \rightarrow \mathcal{L}_Z(\theta, f)$ is convex. We can then relax the Problem (2) by solving the unconstrained problem

$$(\hat{\theta}, \hat{f}) \in \underset{\theta, f}{\operatorname{argmin}} \{ \mathcal{L}_Z(\theta, f) + \Omega(f) \},$$

where $\Omega(f)$ is a convex penalty which promotes solutions of the form $\hat{f}(x, t) = \sum_{j=1}^J \rho_j(t) \mathbf{1}_{R_j}(x)$. Such relaxation procedures have been successfully applied in many statistical settings, for example for variable selection (Tibshirani 1996 [20], Bickel *et al.* 2009 [1]), noisy matrix completion (Candès and Recht 2009 [2], Candès and Tao 2010 [3]), image denoising (Rudin *et al.* 1992 [16], Chambolle *et al.* 2010 [4]), etc. We design below a penalty $\Omega(f)$ suited for our problem. This penalty is closely related to the Total Variation norm (TV) that we briefly present in the next section.

2.3.1 Total Variation norm $\operatorname{TV}(F)$

Let \mathcal{D} be a bounded open set of \mathbf{R}^d and $\mathcal{L}(F)$ be a convex functional on real valued functions $F : \mathcal{D} \rightarrow \mathbf{R}$. Similarly to the Problem (2), the simpler problem of minimizing $\mathcal{L}(F)$ among the functions $F(x) = \sum_{j=1}^J a_j \mathbf{1}_{R_j}(x)$ is numerically intractable in general. A classical convex relaxation of this problem is to solve the unconstrained convex problem (Rudin *et al.* 1992 [16], Chambolle *et al.* 2010 [4])

$$\hat{F} \in \operatorname{argmin} \{ \mathcal{L}(F) + \alpha \operatorname{TV}(F) \}$$

with $\alpha > 0$, and $\operatorname{TV}(F)$ the Total Variation norm defined as follows. When F is C^1 , the Total Variation norm of F is defined by $\operatorname{TV}(F) = \int_{\mathcal{D}} \|\nabla F(x)\| dx$. This definition is extended to any function $F \in L^1_{\operatorname{loc}}(\mathcal{D})$ by

$$\operatorname{TV}(F) = \sup \left\{ - \int_{\mathcal{D}} F(x) \operatorname{div}(\phi(x)) dx : \phi \in C_c^\infty(\mathcal{D}, \mathbf{R}^d) \text{ and } \|\phi\|_\infty \leq 1 \right\}$$

where $C_c^\infty(\mathcal{D}, \mathbf{R}^d)$ is the space of C^∞ functions with compact support from \mathcal{D} to \mathbf{R}^d and $\|\phi\|_\infty = \sup_{x \in \mathcal{D}} |\phi(x)|$. For example, the Total Variation of a piecewise constant $F :]0, 1[\rightarrow \mathbf{R}$ is

$$\operatorname{TV}(F) = \sum_{j=1}^{J-1} |a_{j+1} - a_j|, \quad \text{for } F(x) = \sum_{j=1}^J a_j \mathbf{1}_{R_j}(x)$$

with $R_j = [x_{j-1}, x_j[\cap]0, 1[$ and $0 = x_0 < x_1 < \dots < x_J = 1$. In higher dimension, for any (measurable) subset $R \subset \mathbf{R}^d$, the total variation norm $\operatorname{TV}(\mathbf{1}_R)$ gives the *perimeter* of R . We refer to [4] for more details on TV and to [16, 4] for implementations on image denoising or segmentation.

2.3.2 Synchronous Total Variation norm $\text{STV}(f)$

For relaxing Problem (2), we introduce a new spatiotemporal Synchronous Total Variation norm suited to our problem. Similar to the problem described in the previous section, we seek a segmentation $f(x,t) = \sum_j \rho_j(t) \mathbf{1}_{R_j}(x)$ for each year t . The point is that we want the regions R_j to be the *same* for all years t (we say that the regions are synchronous).

For \mathcal{D} a bounded open set of \mathbf{R}^d and $f : \mathcal{D} \times \{1, \dots, T\} \rightarrow \mathbf{R}$ which is C^1 in the variable x for each t , we define the STV norm as

$$\text{STV}(f) = \int_{\mathcal{D}} \max_t \|\nabla_x f(x,t)\| dx,$$

where ∇_x is the gradient in the variable x . We extend this definition to functions f such that $f(\cdot, t) \in L^1_{\text{loc}}(\mathcal{D})$ for each $t \in \{1, \dots, T\}$, as follows :

$$\text{STV}(f) = \sup \left\{ -\sum_t \int_{\mathcal{D}} f(x,t) \text{div}_x(\phi(x,t)) dx : \begin{array}{l} (i) \phi(\cdot, t) \in C_c^\infty(\mathcal{D}, \mathbf{R}^d) \\ (ii) \|\sum_t \|\phi(\cdot, t)\|\|_\infty \leq 1 \end{array} \right\},$$

where div_x is the divergence operator acting on the variable x . We refer to the Appendix for the proof that these two definitions of $\text{STV}(f)$ coincide when f is regular. To fix up ideas, let us give the STV-norm of a function which is "synchronous" piecewise constant in dimension $d = 1$. In the domain $\mathcal{D} =]0, 1[$, we have

$$\text{STV}(f) = \sum_{j=1}^{J-1} \max_t |\rho_{j+1}(t) - \rho_j(t)|, \quad \text{when } f(x,t) = \sum_{j=1}^J \rho_j(t) \mathbf{1}_{R_j}(x)$$

with $R_j = [x_{j-1}, x_j[\cap \mathcal{D}$ and $0 = x_0 < x_1 < \dots < x_J = 1$. Let us explain from this formula the purpose of introducing the norm STV. Assume that at some time t_* , the function $f(\cdot, t_*)$ has a jump of size Δ_* at location x_j . Then, we see that in terms of the STV-norm, there is no cost for f to have $|\rho_{j+1}(t) - \rho_j(t)|$ as high as Δ_* for all time $t \in \{1, \dots, T\}$. Loosely speaking, if f has a jump at x_* at *some* time t_* , then there is no cost (in terms of the STV-norm) for f to have a jump at x_* for *each* time t . The STV-norm then promotes segmentation in the spatial variable x which is synchronous across the times t .

Let us see now how to relax the Problem (2). Since the STV-norm of f is the supremum of the linear functional

$$L_\phi : f \rightarrow -\sum_t \int_{\mathcal{D}} f(x,t) \text{div}_x(\phi(x,t)) dx$$

the functional $f \rightarrow \text{STV}(f)$ is convex. In particular, we can relax the Problem (2) by solving the unconstrained convex problem

$$(\hat{\theta}, \hat{f}) \in \underset{\theta, f}{\text{argmin}} \{ \mathcal{L}_Z(\theta, f) + \alpha \text{STV}(f) \}, \quad (3)$$

with boundary condition $f(\cdot, 1) = 0$, for some $\alpha > 0$.

2.4 The estimation procedure

Let us write f_{st} for $f(x_s, t)$. A natural discretization of $\text{STV}(f)$ is

$$\sum_s \max_t \|\delta_s f_{st}\|$$

where $\delta_s f_{st}$ is a discrete approximation of $\nabla_x f(x_s, t)$. When the sites s are spread on a regular grid in \mathbf{R}^2 , the norm $\|\nabla_x f(x_s, t)\|$ can be conveniently approximated by the norm of the increments of f in two orthogonal directions. When, the locations x_s are irregularly spread, we can no more use this approximation. We approximate here the norm $\|\nabla_x f(x_s, t)\|$ by

$$\max_{u \in V(s)} |f_{st} - f_{ut}|$$

where $V(s)$ is a neighborhood of s . For simplicity, we will assume that the neighborhood V is reflexive : $u \in V(s)$ if and only if $s \in V(u)$.

We then propose to estimate θ and f by minimizing the convex problem

$$(\hat{\theta}, \hat{f}) \in \operatorname{argmin}_{\theta, f} \left\{ \sum_{s,t,k} \left[e^{\theta_s + f_{st}} - Z_{stk}(\theta_s + f_{st}) \right] + \alpha \sum_s \max_t \max_{u \in V(s)} |f_{st} - f_{ut}| \right\} \quad (4)$$

with boundary condition $f_{s1} = 0$ for all s . This minimization can be done by using a primal-dual (or "Arrow-Hurwicz") optimization scheme. Let us write the discretized STV norm in a dual form

$$\sum_s \max_t \max_{u \in V(s)} |f_{st} - f_{ut}| = \sup \left\{ \sum_s \sum_t \sum_{u \in V(s)} (f_{st} - f_{ut}) \phi_s(u, t) : |\phi_s|_{\ell^1} \leq 1 \text{ for all } s \right\},$$

where $|\phi_s|_{\ell^1} = \sum_{u,t} |\phi_s(u, t)|$ and introduce the function

$$\mathcal{F}(\theta, f, \phi) = \sum_{s,t,k} \left[e^{\theta_s + f_{st}} - Z_{stk}(\theta_s + f_{st}) \right] + \alpha \sum_s \sum_t \sum_{u \in V(s)} (f_{st} - f_{ut}) \phi_s(u, t).$$

The Problem (4) can then be written as

$$(\hat{\theta}, \hat{f}) \in \operatorname{argmin}_{\theta, f} \sup_{\phi} \left\{ \mathcal{F}(\theta, f, \phi) : |\phi_s|_{\ell^1} \leq 1 \text{ for all } s \right\}.$$

The idea of the primal-dual optimization, is to alternate gradient descents of $\mathcal{F}(\theta, f, \phi)$ in the variables f and θ , with gradient ascents of $\mathcal{F}(\theta, f, \phi)$ in the variable ϕ plus a projection of ϕ_s onto the unit ℓ^1 -ball. We refer to the Appendix for a detailed description of the algorithm. When the gradient steps are small enough, Theorem 3.14 in [4] ensures that the algorithm converges to the solution of (4).

2.5 "Adaptive" version of the procedure

Similarly to the so-called Adaptive-Lasso (Zou 2006 [21]), we propose an "adaptive" version of our procedure. The idea is to replace the norm $\text{STV}(f)$ in (3) by

$$\int_{\mathcal{D}} \max_t \frac{\|\nabla_x f(x, t)\|}{\|\nabla_x \hat{f}^{\text{init}}(x, t)\|} dx,$$

where \hat{f}^{init} is a preliminary estimator of f . The motivation for replacing $\text{STV}(f)$ by this rescaled norm is to avoid to over-penalize large discontinuities compared to small discontinuities. It is expected that we have more accurate estimations with such schemes. For example, in the context of linear regression, it has been shown that contrary to the Lasso estimator, the Adaptive-Lasso estimator is consistent in variable selection whatever the design, see [21].

As a preliminary estimator \hat{f}^{init} , we use a spatially-regularized version of the maximum likelihood of estimator of f , obtained by solving for some $\beta > 0$

$$(\hat{\theta}^{\text{init}}, \hat{f}^{\text{init}}) \in \underset{\theta, f}{\text{argmin}} \left\{ \mathcal{L}_Z(\theta, f) + \frac{\beta}{2} \sum_t \int_{\mathcal{D}} \|\nabla_x f(x, t)\|^2 dx \right\},$$

with boundary condition $f(\cdot, 1) = 0$. We penalize the negative log-likelihood $\mathcal{L}_Z(\theta, f)$ by the L^2 -norm of the gradient of f in order to get a smooth solution \hat{f}^{init} . As before, the L^2 -norm $\sum_t \int_{\mathcal{D}} \|\nabla_x f(x, t)\|^2 dx$ could be discretized by $\sum_s \sum_t |\delta_s f_{st}|^2$ where $\delta_s f_{st}$ is a discrete approximation of $\nabla_x f(x_s, t)$. For computational efficiency, we will rather work with a different version and compute $(\hat{\theta}^{\text{init}}, \hat{f}^{\text{init}})$ by minimizing for some $\beta > 0$

$$(\hat{\theta}^{\text{init}}, \hat{f}^{\text{init}}) \in \underset{\theta, f}{\text{argmin}} \left\{ \sum_{s,t,k} \left[e^{\theta_s + f_{st}} - Z_{stk}(\theta_s + f_{st}) \right] + \frac{\beta}{2} \sum_{s,t} \sum_{u \in V(s)} \frac{(f_{st} - f_{ut})^2}{\text{card}(V(s))} \right\} \quad (5)$$

with boundary condition $f(\cdot, 1) = 0$. Finally, the adaptive estimator $(\hat{\theta}, \hat{f})$ is obtained by minimizing for some $\alpha > 0$

$$(\hat{\theta}, \hat{f}) \in \underset{\theta, f}{\text{argmin}} \left\{ \sum_{s,t,k} \left[e^{\theta_s + f_{st}} - Z_{stk}(\theta_s + f_{st}) \right] + \alpha \sum_s \max_t \max_{u \in V(s)} \frac{|f_{st} - f_{ut}|}{\max_t |\hat{f}_{st}^{\text{init}} - \hat{f}_{ut}^{\text{init}}|} \right\}$$

with boundary condition $f(\cdot, 1) = 0$. This optimization is performed with the primal-dual optimization algorithm described in the Appendix.

3 Numerical experiments

We implement the estimation procedure both on real and synthetic data. We first investigate the patterns obtained when applying the procedure on bird monitoring data. Then, we analyze the accuracy and the limits of the procedure on three synthetic data sets.

3.1 Analysis of bird monitoring data

We use the 2001-2009 data from the French Breeding Bird Survey (FBBS), a standardized monitoring program in which skilled volunteer ornithologists identify breeding birds by song or visual contact twice each spring in ca. 1700 randomly selected 2x2 km plots (Jiguet *et al.* 2011 [9]). In each plot, observers record all individuals during a five minutes period in 10 evenly distributed point counts, which are located at least 300 m apart. For a given point count, we retain the maximum count of the two annual visits. For this first implementation of the estimation procedure, we focus on one of the most abundant species in France, the blackcap (*Sylvia atricapilla* L.), a generalist passerine species that occurs in many habitat types. The population size of this species has been increasing at a national scale over the past 10 years. We retain only sites (2x2 km plots) in which the species was seen at least once between 2001 and 2009 and for which we have at least seven years of data, which left us with a total of 361 sites (right-hand side of Figure 1).

The neighborhood V is obtained by the following procedure :

- link each s to its 4 nearest neighbors,
- link sites s, u with distance less than 100 km.

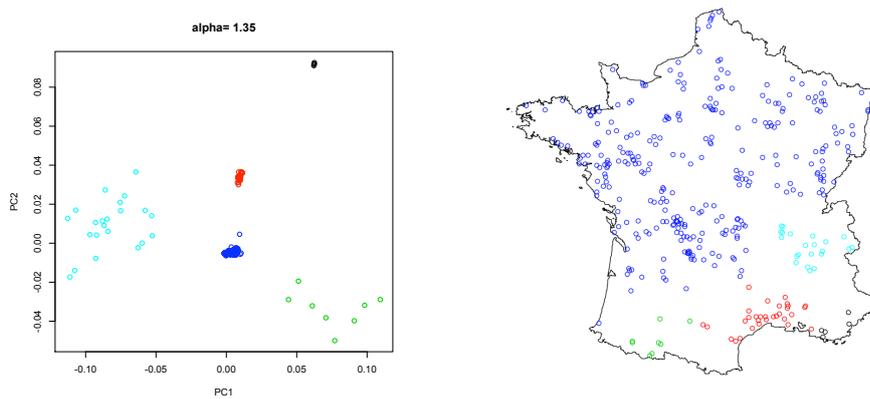


Fig. 1 Left : PCA of (\hat{f}) and kmeans clustering. Right : regions selected by kmeans.

With the choice $\alpha = 1.35$ we obtain the results displayed in Figure 1. The left-hand side of Figure 1 shows for each site s the projection of $\hat{f}[s, \cdot]$ on the plan spanned by the two first principal axes of $\text{PCA}(\hat{f})$. The right-hand side of Figure 1 shows the regions selected when clustering \hat{f} with kmeans. Figure 2 displays the same results for the adaptive procedure. The regions delimited by the simple procedure and the adaptive procedure are mostly the same, except that the alpine region is split into 3 regions by the adaptive procedure.

Finally, the temporal dynamic for each group of Figure 1 is given in Figure 3. The temporal dynamic is obtained by fitting (1) with the regions R of Figure 1. The dotted

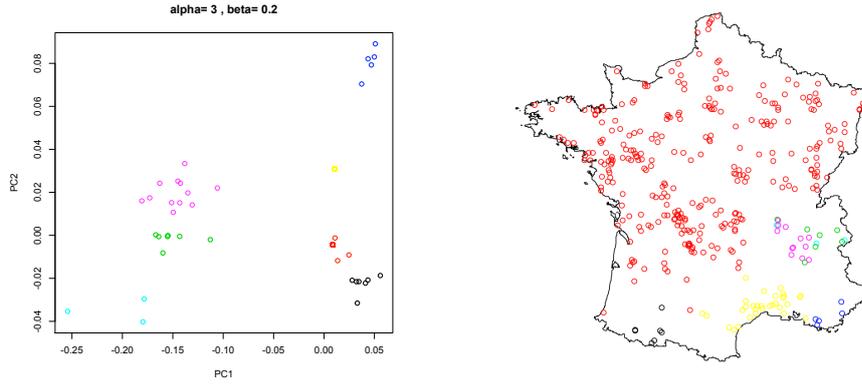


Fig. 2 Left : PCA of \hat{f}^{adapt} and kmeans clustering. Right : regions selected by kmeans.

lines give $mean(\hat{f}[R == r,])$. We notice a strong shrinkage of $mean(\hat{f}[R == r,])$ towards the global behavior. We emphasize that due to the bias towards the mean created by the STV norm, the quantity $mean(\hat{f}[R == r,])$ should not be used for estimating trends.

Results are consistent with what we know on biogeographic regions in France. They separate mountain areas and mediterranean areas from the atlantic and continental regions. We observe a decline in abundance in warm and south regions, whereas population sizes are increasing in temperate and northern areas.

3.2 Synthetic experiments

We investigate the ability of the procedure to detect synchronous populations on three synthetic data sets. We generate 300 sites at random in the square. We divide the sites in three groups with different temporal dynamics given in Figure 4. For each site s and time t we generate observations according to (1) where the θ_s have been sampled according to a gaussian law. Finally, 10% of the data is erased (at random) giving incomplete data.

The accuracy of the results depends on the geometry of the regions. For simple geometries, the estimator (4) with $\alpha = 2.2$ recovers the groups, see Figure 5, Model 1 and Model 2. For more complex geometries, where there is a "central" region, the estimator fails to clearly identify the central region, see Figure 5, Model 3. This can be explained by the fact that it is much more costly in term of the STV-norm to identify a region in the center than in the border. Actually, the STV-norm penalizes the frontier of a region proportionally to its length times the size of the "jump" at the frontier. For border regions, only the "inside" frontier enters into the STV-norm, whereas for central regions, the whole frontier enters into the STV-norm. This difference can explain that central regions are harder to identify.

In these experiments, the adaptive scheme does not significantly change the results, see Figure 5, column 3. We mainly observe that the clusters are more tight.

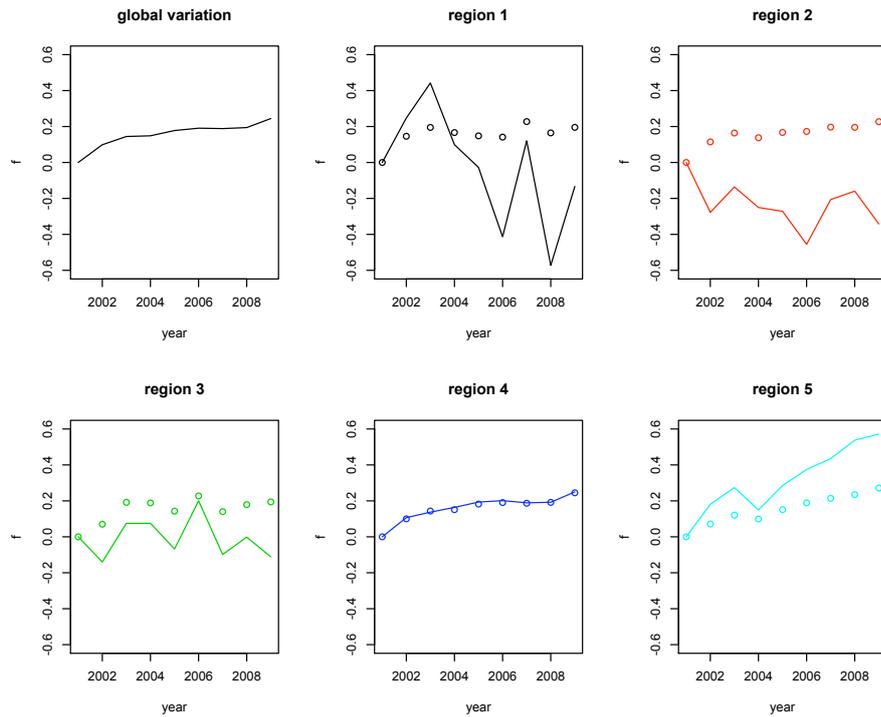


Fig. 3 Temporal dynamics : global and for each region. Plain : model (1) fitted with the regions R estimated Figure 3. Dots : $mean(\hat{f}[R == r_i])$.

4 Discussion

The monitoring data from the French Breeding Bird Survey (FBBS) presents a strong spatial heterogeneity. In Figure 1 and Figure 2, we notice that the frontiers of the delimited regions tend to go through some areas with low-density of sampling points. Therefore, it is unclear whether the estimated frontiers reflect the distribution of the sites or the boundaries of some synchronous populations. The experiments on synthetic data suggest that the estimation procedure can recover simple peripheral regions, whereas the detection of central regions can be more challenging. We also note in Figure 1 and Figure 2 that some regions crossed by areas of low-density of sampling points are not split into several regions. These features suggest that the estimated regions do not only reflect the distribution of the sites.

It is interesting to observe the patterns produced by \hat{f} for different values of $\alpha > 0$. The problem of the optimal choice of α remains tricky. We cannot rely on standard asymptotic criteria such as BIC due to the high-dimensional nature of the data. A natural idea is to apply V -fold Cross-Validation to select α . Nevertheless, an optimal choice of α in terms of the prediction error can lead to inconsistency in terms of the selected regions due to the shrinking effect of the STV penalty (see Meinshausen &

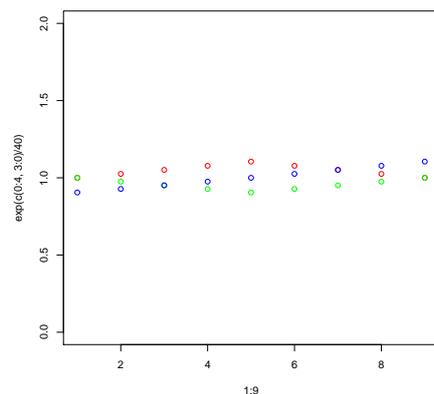


Fig. 4 Synthetic temporal dynamic

Bühlmann 2006 [11] for a similar discussion for the Lasso). When the interest is on the selected regions, we suggest to first apply our procedure with different values of α to obtain different candidate partitions, and then to apply cross-validation to select among the different partitions. For a given partition R , the model (1) can be fitted on the learning set and then test on the validation set.

Overall, our analyses of the population dynamics of a single species, the blackcap, yielded a partition of France that suggests a predominant role of climate forcing in spatial synchrony of bird population dynamics, together with a possible role of mountainous area in isolating populations with different dynamic. Our two approaches identified five to seven regions, with temporal declines in abundance in southernmost areas (especially in Mediterranean regions) and stable or increasing abundances elsewhere, which is consistent with a negative impact of climate warming in warm areas and a general tendency of European bird populations and communities to shift northward (Devictor et al. 2008 [5]). This approach now needs to be extended to other species and to broader spatial scales, so that we can compare the number and identity of synchronous regions, as well as the spatial scale of temporal synchrony in population dynamics across taxa with contrasting ecological and life-history traits (e.g. dispersal ability, latitudinal breeding distribution, ecological specialization, etc.) or between preys and predators. Several studies suggest for example that more mobile species tend to be more synchronized, although this may depend on spatial scale (Sutcliffe et al. 1996 [19]; Paradis et al. 1999 [13]). Comparing the spatial heterogeneity in synchrony of population dynamics across species could ultimately provide critical information for understanding the drivers of the dynamics of populations and communities facing global change.

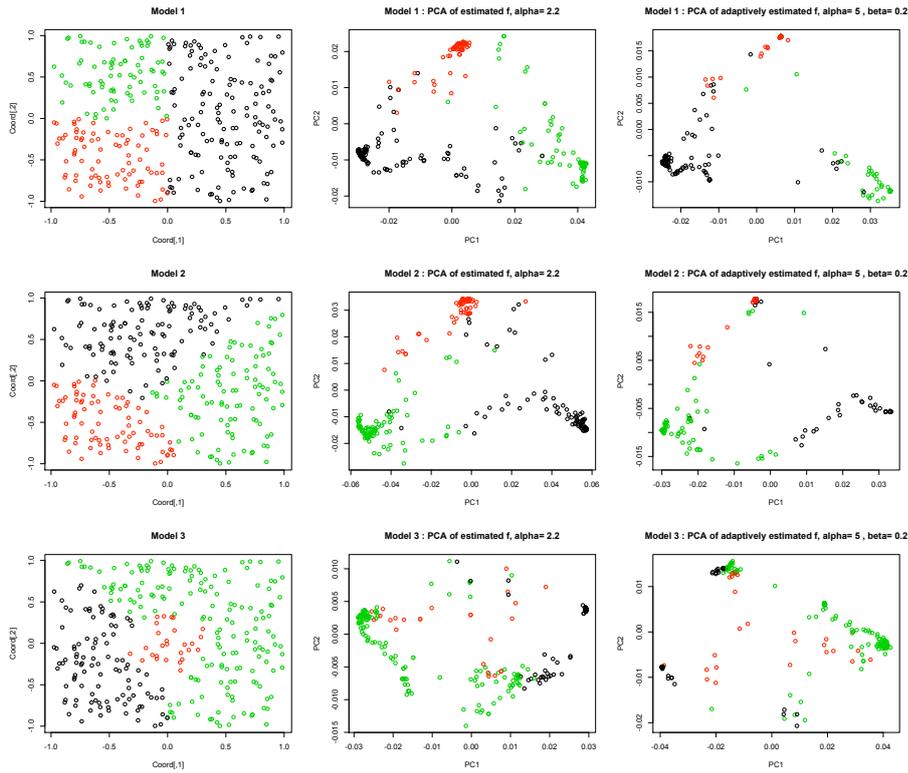


Fig. 5 Synthetic experiments. First column: "true" regions. Second column: projection of $\{\hat{f}[x_s, \cdot], s \in \mathcal{S}\}$ onto the the 2 first axes given by the PCA. Third column: projection of the adaptive estimator $\{\hat{f}^{\text{adapt}}[x_s, \cdot], s \in \mathcal{S}\}$ onto the the 2 first axes given by the PCA.

Acknowledgment

This work was partially funded by the Chaire Modélisation Mathématique et Biodiversité VEOLIA-Ecole Polytechnique-MNHN-F.X. and by the FP7 Large-scale Integrating EU project SCALES (contract number 226852) to R.J. and E.P.

References

1. Bickel, P., Ritov, Y. and Tsybakov, A. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* 37, no. 4, 1705–1732.
2. Candès, E. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* 9, no. 6, 717–772.
3. Candès, E. and Tao, T. (2010). The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory* 56, no. 5, 2053–2080.
4. Chambolle, A., Caselles, V., Novaga, M., Cremers, D. and Pock, T. (2010). An introduction to Total Variation for Image Analysis. *Radon Series Comp. Appl. Math.* 9, 1–78.
5. Devictor, V., Julliard, R., Couvet, D. and Jiguet, F. (2008). Birds are tracking climate warming, but not fast enough. *Proceedings of the Royal Society B Biological Sciences* 275:2743–2748.

6. Grenfell, B., Bjornstad, O. and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414, 716–723.
7. Hassell, M. (2000). The spatial and temporal dynamics of host-parasitoid interactions. Oxford University Press.
8. Heino, M., Kaitala, V., Ranta, E. and Lindstrom, J. (1997). Synchronous dynamics and rates of extinction in spatially structured populations. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 264, 481–486.
9. Jiguet, F., Devictor, V., Julliard, R. and Couvet, D. (2011) French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*. doi:10.1016/j.actao.2011.05.003
10. Liebhold, A., Koenig, W. and Bjornstad, O. (2004). Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution and Systematics*, 35, 467–490.
11. Meinshausen, N. and Bühlmann, P. High dimensional graphs and variable selection with the lasso. *Annals of Statistics* 34 (2006), 1436–1462.
12. Moran, P. (1953). The statistical analysis of the Canadian lynx cycle. 2. Synchronization and meteorology. *Australian Journal of Zoology*, 1, 291–298.
13. Paradis, E., Baillie, S.R., Sutherland, W. J., and Gregory, R. D. (1999). Dispersal and spatial scale affect synchrony in spatial population dynamics. *Ecology Letters* 2:114–120.
14. Ranta, E. and Kaitala, V. (1997). Travelling waves in vole population dynamics. *Nature*, 390, 456–456.
15. Ranta, E., Kaitala, V., Lindstrom, J. and Linden, H. (1995). Synchrony in population dynamics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 262, 113–118.
16. Rudin, L., Osher, S.J. and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms, *Physica D*. 60, 259–268.
17. Schmeller, D.S. (2008). European species and habitat monitoring: where are we now? *Biodiversity and Conservation* 17:3321–3326.
18. Schmeller, D., Henry, P., Julliard, R., Gruber, B., Clobert, J., Dziock, F., et al. (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conservation Biology*, 23, 307–316.
19. Sutcliffe, O.L., Thomas, C.D. and Moss, D. 1996. Spatial synchrony and asynchrony in butterfly population dynamics. *Journal of Animal Ecology* 65:85-95.
20. Tibshirani, R. Regression shrinkage and selection via the lasso (1996). *J. Roy. Statist. Soc. Ser. B* 58, no. 1, 267–288.
21. Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101, no. 476, 1418–1429.

A Consistency of the definitions of STV-norm

We prove in this section that the supremum

$$\sup \left\{ - \sum_t \int_{\mathcal{D}} f(x,t) \operatorname{div}_x(\phi(x,t)) dx : \phi(\cdot, t) \in C_c^\infty(\mathcal{D}, \mathbf{R}^d) \text{ and } \left\| \sum_t \|\phi(\cdot, t)\| \right\|_\infty \leq 1 \right\} \quad (6)$$

equals $\int_{\mathcal{D}} \max_t \|\nabla_x f(x,t)\| dx$ for any function f such that $f(\cdot, t)$ is C^1 for all time $t \in \{1, \dots, T\}$.

First, for any ϕ such that $\phi(\cdot, t) \in C_c^\infty(\mathcal{D}, \mathbf{R}^d)$ for all $t \in \{1, \dots, T\}$, we have by integration by parts and Hölder inequality

$$\begin{aligned} - \sum_t \int_{\mathcal{D}} f(x,t) \operatorname{div}_x(\phi(x,t)) dx &= \sum_t \int_{\mathcal{D}} \nabla_x f(x,t) \cdot \phi(x,t) dx \\ &\leq \int_{\mathcal{D}} \max_t \|\nabla_x f(x,t)\| \sum_t \|\phi(x,t)\| dx \\ &\leq \left\| \sum_t \|\phi(\cdot, t)\| \right\|_\infty \int_{\mathcal{D}} \max_t \|\nabla_x f(x,t)\| dx. \end{aligned}$$

It then follows that the supremum (6) is not larger than $\int_{\mathcal{D}} \max_t \|\nabla_x f(x,t)\| dx$.

Conversely, write $\mathcal{T}(x) = \inf \{t : \|\nabla_x f(x,t)\| = \max_t \|\nabla_x f(x,t)\|\}$ and define \mathcal{D}_t by

$$\mathcal{D}_t = \{x \in \mathcal{D} : \mathcal{T}(x) = t \text{ and } \nabla_x f(x,t) \neq 0\}.$$

The function

$$\psi(x, t) = \frac{\nabla_x f(x, t)}{\|\nabla_x f(x, t)\|} \mathbf{1}_{\mathcal{D}_t}(x)$$

fulfills

$$\sum_t \int_{\mathcal{D}} \nabla_x f(x, t) \cdot \psi(x, t) dx = \int_{\mathcal{D}} \max_t \|\nabla_x f(x, t)\| dx \quad \text{and} \quad \left\| \sum_t \|\psi(\cdot, t)\| \right\|_{\infty} \leq 1.$$

For any $\varepsilon > 0$, let us consider ϕ fulfilling $\phi(\cdot, t) \in C_c^\infty(\mathcal{D}, \mathbf{R}^d)$, $\left\| \sum_t \|\phi(\cdot, t)\| \right\|_{\infty} \leq 1$ and

$$\sum_t \int_{\mathcal{D}} \|\psi(x, t) - \phi(x, t)\|^2 dx < \varepsilon^2.$$

An integration by parts together with Cauchy-Schwartz inequality ensure that

$$\begin{aligned} & - \sum_t \int_{\mathcal{D}} f(x, t) \operatorname{div}_x(\phi(x, t)) dx \\ &= \sum_t \int_{\mathcal{D}} \nabla_x f(x, t) \cdot \phi(x, t) dx \\ &= \sum_t \int_{\mathcal{D}} \nabla_x f(x, t) \cdot \psi(x, t) dx + \sum_t \int_{\mathcal{D}} \nabla_x f(x, t) \cdot (\phi(x, t) - \psi(x, t)) dx \\ &\geq \int_{\mathcal{D}} \max_t \|\nabla_x f(x, t)\| dx - \varepsilon \left(\sum_t \int_{\mathcal{D}} \|\nabla_x f(x, t)\|^2 dx \right)^{1/2}. \end{aligned}$$

As a consequence, the supremum (6) is not smaller than $\int_{\mathcal{D}} \max_t \|\nabla_x f(x, t)\| dx$.

B Primal-dual optimization algorithm

We describe below the primal-dual optimization algorithm we use for solving

$$(\hat{\theta}, \hat{f}) \in \operatorname{argmin}_{\theta, f} \left\{ \sum_{s, t, k} \left[e^{\theta_s + f_{st}} - Z_{stk}(\theta_s + f_{st}) \right] + \alpha \sum_s \max_t \max_{u \in V(s)} \frac{|f_{st} - f_{ut}|}{D_{su}} \right\}$$

where $D_{su} = 1$ in (4) and $D_{su} = \max_t |f_{st}^{\text{init}} - f_{ut}^{\text{init}}|$ for the adaptive estimator.

For a small positive h , the primal-dual optimization algorithm iterates until convergence the followings steps

1. gradient descent in θ : for each s

$$\theta_s \leftarrow \theta_s - h \left(e^{\theta_s} \sum_{t, k} e^{f_{st}} - \sum_{t, k} Z_{stk} \right)$$

2. gradient descent in f : for each s and each $t \in \{2, \dots, T\}$

$$f_{st} \leftarrow f_{st} - h \left(\sum_k (e^{\theta_s} e^{f_{st}} - Z_{stk}) + \alpha \sum_{u \in V(s)} (\phi_s(u, t) - \phi_u(s, t)) / D_{su} \right)$$

3. gradient ascent in ϕ

$$\phi_s(u, t) \leftarrow \phi_s(u, t) - h \alpha (f_{st} - f_{ut}) / D_{su}$$

4. projection of each ϕ_s on the unit ℓ^1 -ball.

Projection on the unit ℓ^1 -ball

The projection $\hat{\beta}$ of a vector β on the unit ℓ^1 ball is computed efficiently as follows:

- Sort the coordinates of β . Write $\beta_{(j)}$ for the j -th largest coordinate.
- Set

$$\hat{j} = \operatorname{argmax}_j \left\{ \sum_{j \leq J} |\beta_{(j)}| - j|\beta_{(j)}| \leq 1 \right\} \quad \text{and} \quad \hat{\lambda} = \left(\sum_{j \leq \hat{j}} |\beta_{(j)}| - 1 \right) / \hat{j}$$

- Return $\hat{\beta}$ with coordinates $\hat{\beta}_j = \beta_j \left(1 - \hat{\lambda} / |\beta_j| \right)_+$.