# A spatial Markovian framework for estimating regional and local dynamics of annual plants with dormancy 

Sébastian Le Coz, Pierre-Olivier Cheptou, Nathalie Peyrard

## - To cite this version:

Sébastian Le Coz, Pierre-Olivier Cheptou, Nathalie Peyrard. A spatial Markovian framework for estimating regional and local dynamics of annual plants with dormancy. Theoretical Population Biology, 2019, 127, pp.120-132. 10.1016/j.tpb.2019.03.002 . hal-02939149

## HAL Id: hal-02939149 <br> https://hal.inrae.fr/hal-02939149

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

©®®

# A spatial Markovian framework for estimating regional and local dynamics of annual plants with dormancy 

Sebastian Le Coz ${ }^{\text {a,b,* }}$, Pierre-Olivier Cheptou ${ }^{\text {b }}$, Nathalie Peyrard ${ }^{\text {a }}$<br>${ }^{a}$ INRA UR 875 MIAT, Chemin de Borde Rouge, 31326 Castanet-Tolosan, France<br>${ }^{b}$ CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valery<br>Montpellier, EPHE - 1919, route de Mende - 34293 Montpellier cedex 05, France

7 Abstract
Many species have a dormant stage in their life cycle, including seeds for plants. The dormancy stage influences the species dynamics but is often undetectable. One way to include dormancy is to model it as a hidden dynamical state within a Markovian framework. Models within this framework have already been proposed but with different limitations: only presence/absence observations are modelled, the dormancy stage is limited to one year, or colonisation from neighbouring patches is not taken into account. We propose a hidden Markov model that describes the local and regional dynamics of a species that can undergo dormancy with a potentially infinite dormancy time. Populations are modelled with abundance classes. Our model considers the colonisation process as the indistinguishable influence of neighbour non-dormant population states on a dormant population state in a patch. It would be expected that parameter estimation, hidden state estimation and prediction of the next non-dormant populations would have an exponential computational time in terms of the number of patches. However, we demonstrate that estimation, hidden state estimation and prediction are all achievable in a linear computational time. Numerical experiments on simulated data show that the state of dormant populations can easily be retrieved, as well as the state of future non-dormant populations. Our framework provides a simple and efficient tool that could be further used to analyse and compare annual plants dynamics like weed species survival strategies in crop fields.
8 Keywords: Metapopulation, Seed bank, Hidden life stage, Colonisation, Prediction, Expectation-Maximisation algorithm

[^0]March 3, 2019

## 1. Introduction

A species's distribution depends on its local and regional dynamics (Hanski, 1998; Bullock et al., 2006). The metapopulation model, developed by Levins et al. (1969), uses colonisation and extinction parameters to describe a species dynamics across multiple patches. Levins' model considered colonisation to be dependent on the fraction of occupied patches. Multiple studies have used the metapopulation concept on different types of species (Hanski, 1994; Dornier et al., 2011). However, for many organisms (plants or animals) that can survive rough environmental conditions by stopping their development through a process called dormancy, the use of the original metapopulation model has been questioned due to the fact that the dormancy stage is not modelled (Freckleton and Watkinson, 2002; Bullock et al., 2006). Fréville et al. (2013) have shown that such models tend to overestimate colonisation and extinction parameters for annual plants with seed banks. Understanding and modelling dormancy in a species dynamics is important in order to better conserve and control the species.

Multiple models have been developed to represent the dynamics of species that can undergo dormancy using information on the dormant and nondormant populations of the species (Cohen, 1966; Levin et al., 1984; Jarry et al., 1995; Amarasekare and Possingham, 2001; Mistro et al., 2005; Soubeyrand et al., 2009; Han et al., 2014). However, in practice, data is often only collected on the species' non-dormant population since the dormant population is hard to observe. This implies that the models mentioned above, which require full knowledge about local dormant populations, would not be adapted to estimating a species' local and regional dynamics from partial data alone. A model with incomplete information where the dormancy stage is modelled as a hidden variable and the non-dormant stage is the observable data is more appropriate to estimate the dynamics of a species that can undergo dormancy.

The local dynamics of species with dormancy stages has already been studied using models with incomplete information (David et al., 2010; QuintanaAscencio et al., 2011; Lamy et al., 2013; Fréville et al., 2013; Borgy et al., 2015; Manna et al., 2017). The models in Lamy et al. (2013), Fréville et al. (2013), Borgy et al. (2015), and Manna et al. (2017) use a Markovian framework where the current state of the population only depends on the state of the population in the previous time step. A classical extension of the Markov model to deal with incomplete observation is the Hidden Markov

Model (HMM). In a HMM, a hidden Markov chain emits information at each time step and the emitted information does not influence the states of hidden chain. By considering the species non-dormant population as observable and the species dormant population as hidden, it would be tempting to use HMM to study such species. However, the HMM does not include the influence of a species' non-dormant populations on its dormant populations. Therefore, Borgy et al. (2015) and Pluntz et al. (2018) extended the HMM to include the influence of emitted observations on the hidden variables. The model was used to analyse the dynamics of weed species in crop fields from real data on weed abundance and weed occurrence. We will refer to this model as a HMM with Data Feedback (HMM-DF).

The main limitation of the models developed by Pluntz et al. (2018), Borgy et al. (2015), Fréville et al. (2013), Lamy et al. (2013), QuintanaAscencio et al. (2011), and David et al. (2010) is the absence of an explicit contribution of neighbouring patches in the colonisation process. Colonisation, when modelled, is through seed rain, as the propagule rain effet described in Gotelli (1991). As a result, these authors consider patches to be independent. To take colonisation between patches into account, the model should include interactions between patches. Factorial HMM (Ghahramani and Jordan, 1997) or Coupled HMM (Brand et al., 1997; Wainwright and Jordan, 2008) are models that extend HMM to include interactions between patches (see Appendix A for a graphical representation). However, Factorial HMMs consider that the observation at time $n$ depends on the state of the hidden variables of all chains at time $n$ and, more importantly, that this observation does not influence the hidden states at $n+1$. Thus, the colonisation process, from non-dormant populations to dormant ones, cannot be modelled with a FHMM. In a Coupled HMM the hidden state of a patch at time $n$ influences the hidden states of other patches at time $n+1$. This implies that the colonisation process occurs between dormant populations. This assumption is often wrong since dormant populations usually cannot move. In addition to having a dependency structure unsuited to model colonisation, the Expectation Maximisation (EM) algorithm (Dempster et al., 1977), which is often used for parameter estimation based on incomplete information, has a time complexity exponential in terms of the number of patches, for both frameworks (Ghahramani and Jordan, 1997; Brand et al., 1997). The model proposed in Manna et al. (2017), which extends the model of Fréville et al. (2013), considered patches to be organised on a circle with colonisation coming from the closest neighbouring patches. Their model makes it possible to
compare the relative impact of seed dormancy and limited dispersal on the species dynamics. However, they conserved the limiting assumptions of the model of Fréville et al. (2013), such as: dormant and non-dormant populations modelled by presence/absence; seed bank survival limited to one year; and the automatic presence of seeds in the seed bank when plants were found in the same patch at the previous time step.

In order to avoid these assumptions, we propose a Multidimensional HMM with Data Feedback (MHMM-DF) that includes local (within patch) and regional (between patches) dynamics to describe the dynamics of species that undergo dormancy. The model considers that dormant and non-dormant populations depend stochastically on each other and the species non-dormant and dormant populations are modelled with abundance classes. The state of a dormant population in a given patch at time $n$ is described as the result of four processes : (i) dormancy, i.e., survival of a population in the dormant stage between successive time steps where this survival is not limited in time; (ii) locally newly produced dormant individuals; (iii) colonisation from neighbouring patches; and (iv) exogenous colonisation. Processes (ii) and (iii) induce a dependency of the observed data on the hidden state.

A MHMM-DF model is fully defined by the framework describing the interactions between populations. Since data on non-dormant populations are rarely collected over a long period of time, we propose a parametric version of the MHMM-DF. This version relies on simple probability distributions and functional forms for the sake of parsimony. The resulting parametric MHMM-DF has seven parameters and can characterise the dynamics of annual plants with seed banks.

Regarding parameter estimation, the naive procedure that consists of applying the EM algorithm on the reformulation of the MHMM-DF as a HMM with a single multidimensional chain has an exponential time and space complexity in terms of the number of patches. However, parameter estimation with exponential time and space complexity can only be done for a MHMMDF with a small number of hidden states and a small number of patches. Thus, we demonstrate that for the MHMM-DF structure, estimation using the EM algorithm is achievable with a linear time and space complexity in terms of the number of patches. This linear complexity is obtained thanks to independence between dormant populations at time $n$ and patch $c$ and dormant populations in other patches at previous times, conditional on observed non-dormant populations. The same independence property is used for the Viterbi algorithm (Forney, 1973), which enables the recovery of the
state of the species' dormant populations for all time steps with the same complexity. The Viterbi algorithm unlike the EM algorithm finds the most probable sequence of hidden states using a set of parameters, whereas the EM finds the set of parameters which is the most probable to have generated the observable data. Additionally, the model can be used to predict the subsequent states of the species' non-dormant population.

We used simulated data to illustrate the quality of estimations provided by the EM algorithm for MHMM-DF, and of hidden state estimations and predictions provided by the Viterbi algorithm. Finally, using model selection techniques, we demonstrate how the framework can be used to discriminate between dynamics with and without dormancy or with and without colonisation.

## 2. Multidimensional HMM with data feedback

The MHMM-DF is a Dynamic Bayesian Network (Ghahramani, 1998). In the following sections we define the dependency structure of a MHMM-DF and model the interactions between dormant and non-dormant populations with a parametric version of the MHMM-DF.

### 2.1. Definition

Let us consider a set $\mathcal{C}$ of $C$ patches. At time $n \in\{1, \ldots, N\}$ on patch $c \in \mathcal{C}$ two variables are defined: $X_{c, n}$ is the abundance class of the dormant population (hidden) and $Y_{c, n}$ is the abundance class of the non-dormant population (observed). Their domains are $\Omega_{X}=\left\{0,1, \ldots,\left|\Omega_{X}\right|-1\right\}$ and $\Omega_{Y}=\left\{0,1, \ldots,\left|\Omega_{Y}\right|-1\right\}$, respectively. In order to simplify the expression of probabilities, we will resort to extra notations summarised in Table 1. The sequence of hidden variables of patch $c$ forms the hidden chain $c$. A Multidimensional HMM with Data Feedback (MHMM-DF) of $C$ dimensions models the joint dynamics of the $C$ chains when the following two assumptions are fulfilled. First, for a given patch $c$ at a given time $n$, the non-dormant population state $Y_{c, n+1}$ only depends on the dormant population state $X_{c, n}$. The second assumption is that the dormant population state $X_{c, n+1}$ depends on the dormant population state of the same chain at the previous time step $X_{c, n}$, and on all non-dormant population states at time $n+1, Y_{n+1}^{C}$. This direct dependence of hidden variables on observable variables is the data feedback part of the model. The dependences in the MHMM-DF are represented in Fig. 1 for two chains or patches.

| Variable | Domain | Definition |
| :--- | :--- | :--- |
| $X_{c, n}$ | $\Omega_{X}=\left\{0,1, \ldots\left\|\Omega_{X}\right\|-1\right\}$ | abundance class of dormant |
|  |  | population in patch $c$ at time $n$ |
| $Y_{c, n}$ | $\Omega_{Y}=\left\{0,1, \ldots\left\|\Omega_{Y}\right\|-1\right\}$ | abundance class of non-dormant |
| $X_{n}^{C}$ | $\Omega_{X}^{C}$ | population in patch $c$ at time $n$ |
| $Y_{n}^{C}$ | $\Omega_{X}^{C}$ | $\left\{X_{c, n}, 1 \leq c \leq C\right\}$ |
| $X_{n}^{C}$ | $\Omega_{X}^{n}$ | $\left\{Y_{c, n}, 1 \leq c \leq C\right\}$ |
| $Y_{n}^{C}$ | $\Omega_{X}^{n C}$ | $\left\{X_{n}^{C}, 1 \leq n \leq N\right\}$ |

Table 1: Variables of the MHMM-DF.

The dynamics of a MHMM-DF is fully defined by three probabilities. The first one is the initial probability of the hidden states, $\mathbb{P}\left(X_{0}^{C}=x_{0}^{C}\right)$. For the sake of simplicity we will assume here that $\mathbb{P}\left(X_{0}^{C}=x_{0}^{C}\right)=\prod_{c=1}^{C} \pi\left(x_{c, 0}\right)$. The emission probability, defined as $\phi\left(x_{c, n-1}, y_{c, n}\right)=\mathbb{P}\left(Y_{c, n}=y_{c, n} \mid X_{c, n-1}=\right.$ $x_{c, n-1}$ ), models the awakening process, which corresponds to the generation of the local non-dormant population from the local dormant population. Finally, the transition probability of the hidden variable of chain $c$, defined as $A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)=\mathbb{P}\left(X_{c, n}=x_{c, n} \mid X_{c, n-1}=x_{c, n-1}, Y_{n}^{C}=y_{n}^{C}\right)$, models the generation of the current dormant population at patch $c$, given the previous dormant population state and the non-dormant population state both locally and in the neighbourhood. Note that in $\phi$ and $A$, variable ordering first corresponds to hidden variables and then to observed ones, and not to the temporal ordering of the events. In Section 3, we describe these probabilities in the case of annual plants.

With these notations, the joint probability distribution of the hidden and observed variables of the $C$ chains is:

$$
\begin{equation*}
\prod_{c=1}^{C} \pi\left(x_{c, 0}\right) \prod_{n=1}^{N} A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) \tag{1}
\end{equation*}
$$

### 2.2. Model parameterisation

The estimation of $\phi$ and $A$ functions in a non-parametric approach could be difficult, in particular, when the amount of available data is low. We present here a possible parameterisation for $\phi$ and $A$ for the case where hidden and observed variables are abundance classes.


Figure 1: Dependency graph in a Multidimensional HMM with Data Feedback. Case of two hidden chains (grey nodes). Blue edges are involved in the emission probability $\phi$ and red ones in the transition probability $A$. Each chain emits its own observations and then observations from all chains influence the next hidden state of a chain.

For a given value $x_{c, n-1}, \phi\left(x_{c, n-1}, y_{c, n}\right)$ is modelled as a Binomial distribution with parameters $\left|\Omega_{Y}\right|$, and $p_{x_{c, n-1}}$. The probability $p_{x_{c, n-1}}$ is then modelled as a logistic regression: $p_{x_{c, n-1}}=\frac{1}{1+\exp \left(-\left(\mu_{0}+\mu_{1} x_{c, n-1} /\left|\Omega_{X}\right|\right)\right)}$, where $\mu_{0}$ and $\mu_{1}$ are hyper parameters. Thus:

$$
\begin{align*}
\phi\left(x_{c, n-1}, y_{c, n}\right)= & \binom{\left|\Omega_{Y}\right|-1}{y_{c, n}}\left[\frac{1}{1+\exp \left(-\left(\mu_{0}+\mu_{1} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}\right)\right)}\right]^{\left|\Omega_{Y}\right|-1} \\
& \times\left[\exp \left(-\left(\mu_{0}+\mu_{1} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}\right)\right)\right]^{\left|\Omega_{Y}\right|-y_{c, n}-1} \tag{2}
\end{align*}
$$

The choice of a Binomial distribution combined with a logistic regression is motivated by the need for a parsimonious model (i.e., a model with few parameters). Additionally, a Binomial distribution combined with a logistic regression enables us to have a mean that increases when the class of the explanatory variables increases.

The transition distribution $A$ can also be modelled as a Binomial distribution. The parameters are $\left(\left|\Omega_{X}\right|, p_{x_{c, n-1}, y_{n}^{C}}\right)$, and the second parameter is
again modelled as a logistic regression. It is natural to distinguish the influence of $y_{c, n}$ from the influence of the observations on the chains other than chain $c$, so we propose to model $p_{x_{c, n-1}, y_{n}^{C}}$ as a function of $x_{c, n-1}, y_{c, n}$ and $y_{n}^{\mathcal{C} \backslash c}$, where $y_{n}^{\mathcal{C} \backslash c}=\left\{y_{c^{\prime}, n}, c^{\prime} \in \mathcal{C}, c^{\prime} \neq c\right\}$ is the set of observations of all the patches except $c$ at time $n$.

Extracting a subset of $y_{n}^{\mathcal{C} \backslash c}$ can be done in order to analyse a specific colonisation process. This will be the case, for example, in a metapopulation process where only populations from geographically-close patches can colonise a given patch. In other situations, it can be reasonable to assume that colonisation is possible with the same intensity from all patches. Colonisation scenarios can be modelled in the MHMM-DF framework by building an appropriate $A$ function. Their suitability to data can be compared using model selection techniques. Here, for illustration purposes, we focused on models for which each patch contributes equally to colonisation.

In order to limit the number of parameters in the regression model for $p_{x_{c, n}, y_{n}^{C}}$, we propose to aggregate $y_{n}^{\mathcal{C} \backslash c}$ in a one dimension variable. There are several options to summarise the information in $y_{n}^{\mathcal{C} \backslash c}$. For instance, we will refer to mean colonisation as the colonisation process that considers that only the mean value of the non-dormant population state in patches other than $c$ influences the dormant population state in patch $c$. The function that summarises the information is called $f_{\text {mean }}$ and associates the set of observed states of non-dormant population $y_{n}^{\mathcal{C} \backslash c}$ to its mean value (rounded to the first smaller integer to remain in $\Omega_{Y}$ ). It is a measure of the mean colonisation capacity in the neigbourhood. The output of $f_{\text {mean }}$ is then used to define the logistic regression model for $p_{x_{c, n}, y_{n}^{C}}$ as follows:

$$
\begin{equation*}
p_{x_{c, n}, y_{n}^{C}}=\frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \times \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}+\nu_{2} \times \frac{y_{c, n}}{\left|\Omega_{Y}\right|}+\nu_{3} \times \frac{f_{\text {mean }}\left(y_{n}^{C c}\right)}{\left|\Omega_{Y}\right|}\right)\right)} . \tag{3}
\end{equation*}
$$

Note that each explanatory variable was normalised.
Finally, we modelled the initial distribution of the species dormant population in patch $c$ by using a parameterised Binomial distribution with parameters $\left|\Omega_{X}\right|$ and $p_{\tau}$ where:

$$
\begin{equation*}
p_{\tau}=\frac{1}{1+\exp (-\tau)} \tag{4}
\end{equation*}
$$

| Hyperparameters | Interpretation for annual plants |
| :--- | :--- |
| Initial distribution $\mathbb{P}\left(x_{c, 0}\right)=\pi\left(x_{c, 0}\right)$ |  |
| $\tau$ | Mean trend |

Table 2: Parameterisation of a MHMM-DF using Binomial Logistic Regressions. The hyperparameters are the parameters of the logistic regression.

This parameterisation of the MHMM-DF requires seven hyper parameters that are listed in Table 2. Using classical results of identifiability for HMM, we established that the MHMM-DF is generically identifiable as soon as $C>$ 2 under the mild assumption that $N$ must be greater than 7 (see Appendix B).

## 3. Modelling the dynamics of annual plants with a MHMM-DF

The MHMM-DF is well adapted to describe the regional and local dynamics of annual plants like weeds. In order to model a species dynamics with a MHMM-DF, two assumptions must be made. The first assumption imposes immobility on the species' dormant populations in a patch. This assumption is usually verified for annual plants since seeds are immobile once in the seed bank. The second assumption is that the species' non-dormant population cannot directly influence another non-dormant population. A species' non-dormant population can only influence the species' dormant populations. Thus, a non-dormant population will either die, become sterile or transform into a dormant population at the end of each time step. This assumption
is verified for annual plants when the time step is at least a year since the species' non-dormant population does not live longer than a year. This implies a complete renewal of the plant's non-dormant population between each time step.

In a MHMM-DF for annual plant dynamics, the hidden variables or the species dormant populations $\left(X_{c, n}\right)_{(c, n) \in\{1, \ldots, C\} \times\{1, \ldots, N\}}$ correspond to the seed bank's state at time $n$ in patch $c$. The observable variables or the species' non-dormant populations $\left(Y_{c, n}\right)_{(c, n) \in\{1, \ldots, C\} \times\{1, \ldots, N\}}$ correspond to the adult plant's state at time $n$ in patch $c$. The seed bank state is influenced by dormancy, colonisation and the current state of adult flora in the patch. All these processes are represented with red arrows in Fig. 1. The blue arrows correspond to germination and survival up to adulthood.

For a parametric MHMM-DF with indistinguishable effects of neighbours, each hyperparameter is associated with a specific process of a plant's dynamics, with the exception of $\tau$, which is associated with the initial probability of dormant populations. With a parametric MHMM-DF, the annual plant dynamics is described by multiple processes: survival of seeds in the seed bank corresponding to the dormancy process $\left(\nu_{1}\right)$; local production of seeds entering the local seed bank corresponding to the influence of the local nondormant population on the local dormant population $\left(\nu_{2}\right)$; and neighbouring production of seeds entering the local seed bank corresponding to the spatial colonisation process $\left(\nu_{3}\right)$ and the germination process corresponding to $\left(\mu_{1}\right)$. The exogenous colonisation process has two hyperparameters, $\mu_{0}$ and $\nu_{0}$. The exogenous production of seeds entering the regional seed bank by seed rain corresponds to the exogenous colonisation of dormant populations $\left(\nu_{0}\right)$. The process modelled with the hyperparameter $\mu_{0}$ corresponds to exogenous colonisation of non-dormant populations. This process is not present with annual plants. Even if this process cannot be removed, the smaller $\mu_{0}$ is, the smaller the probability of non-dormant populations being colonised by the exterior will be. For instance, with $\mu_{0}=-5.7$, the probability of exogenous colonisation of a non-dormant population occurring when all past populations are extinct is 0.004 .

## 4. Estimation

A classical algorithm to estimate the parameters of a HMM model is the EM algorithm (Dempster et al., 1977). This iterative algorithm alternates an Expectation step and a Maximisation step, and converges towards
a local maximum of the log-likelihood. The E-step for HMM is the wellknown Forward-Backward algorithm (Rabiner, 1989), which takes advantage of the linear structure of a HMM (see Fig. A.9) and variable elimination to efficiently compute all conditional probabilities of the hidden variables given the observations. However, when the hidden variables are multidimensional, the size of the domains and the structure of interaction can make the forward-backward algorithm intractable. Since chains only depend on one another through observations in MHMM-DF, we have demonstrated that independently running a forward-backward per chain is possible. While remaining exact, it significantly reduces the computational complexity of the E step from exponential in terms of the number of patches $O\left(\left|\Omega_{X}\right|{ }^{2 C} N\right)$ to linear $O\left(\left|\Omega_{X}\right|^{2} C N\right)$. In this section, the EM algorithm is described for onedimensional HMM and then detailed for MHMM-DF.

### 4.1. EM algorithm for HMM

A HMM (see Fig. A.9) is defined by: the initial probability $\pi\left(x_{0}\right)$, the emission probability $\phi\left(x_{n}, y_{n}\right)$ and the transition probability $A\left(x_{n-1}, x_{n}\right)$. As can be observed the transition probability does not depend on observations, and is $A\left(x_{n-1}, x_{n}\right)=\mathbb{P}\left(X_{n}=x_{n} \mid X_{n-1}=x_{n-1}\right)$. Let us denote $\lambda=(\pi, \phi, A)$ corresponding to the model's parameters.

Let us define $Q\left(\lambda \mid \lambda^{\prime}\right)=E\left[\ln \left(\mathbb{P}\left(X^{N}, Y^{N} \mid \lambda\right) \mid Y^{N}=y^{N}, \lambda^{\prime}\right)\right]$. The function $Q$ can be expressed in terms of $\pi, \phi$ and $A$ :

$$
\begin{aligned}
Q\left(\lambda \mid \lambda^{\prime}\right) & =\sum_{n=0}^{N} \sum_{x_{n} \in \Omega_{X}} \gamma_{n}\left(x_{n}\right) \ln \left(\phi\left(x_{n}, y_{n}\right)\right) \\
& +\sum_{x_{0} \in \Omega_{X}} \gamma_{0}\left(x_{0}\right) \ln \left(\pi\left(x_{0}\right)\right) \\
& +\sum_{n=1}^{N} \sum_{x_{n-1} \in \Omega_{X}} \sum_{x_{n} \in \Omega_{X}} \xi_{n}\left(x_{n-1}, x_{n}\right) \ln \left(A\left(x_{n-1}, x_{n}\right)\right)
\end{aligned}
$$

where $\gamma_{n}\left(x_{n}\right)=\mathbb{P}\left(X_{n}=x_{n} \mid Y^{N}=y^{N}, \lambda_{t}\right)$ and $\xi_{n}\left(x_{n-1}, x_{n}\right)=\mathbb{P}\left(X_{n-1}=\right.$ $\left.x_{n-1}, X_{n}=x_{n} \mid Y^{N}=y^{N}, \lambda_{t}\right)$.

We denote $\lambda_{t}$ as the current estimator at iteration $t$ of the algorithm. The E-steps and M-steps are defined as follows:

E-step: compute all the probabilities $\gamma_{n}\left(x_{n}\right)$ and $\xi_{n}\left(x_{n-1}, x_{n}\right)$.
M-step: update the parameter $\lambda$ by maximising $Q\left(\lambda \mid \lambda_{t}\right)$.

The direct application of the EM algorithm to a HMM where the hidden variable is of dimension $C$ (i.e., $\left.x_{n}=\left(x_{1, n}, \ldots, x_{c, n}\right)\right)$ requires a number of computations that is exponential in $C$. Therefore, for multidimensional hidden states HMMs , the E step is generally intractable. However, for the structure of a MHMM-DF, we can derive an exact E-step that is only linear in terms of $C$.

### 4.2. E step for MHMM-DF

In a MHMM-DF, the $C$ hidden chains are independent conditionally to all the observations:

$$
\mathbb{P}\left(X^{C, N}=x^{C, N} \mid Y^{C, N}=y^{C, N}\right)=\prod_{c=1}^{C} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N} \mid Y^{C, N}=y^{C, N}\right)
$$

This can be graphically seen in Fig. 1 where all the paths going from a hidden variable of chain $c$ to a hidden variable of chain $c^{\prime}$ go through an observed variable. This implies that the function $Q\left(\lambda \mid \lambda^{\prime}\right)$ can be written in terms of the following probabilities :

$$
\begin{aligned}
\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right) & =\mathbb{P}\left(X_{c, n}=x_{c, n}, X_{c, n-1}=x_{c, n-1} \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right), \\
\rho_{c, n-1}\left(x_{c, n-1}\right) & =\mathbb{P}\left(X_{c, n-1}=x_{c, n-1} \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right) .
\end{aligned}
$$

The proof is provided in Appendix C. Consequently, we could consider applying one forward-backward algorithm per chain. A direct mimicking of the forward-backward algorithm for a one-dimensional HMM would suggest that we define the auxiliary variable for chain $c$ as :

$$
\begin{gathered}
\alpha_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(Y^{C, n}=y^{C, n}, X_{c, n}=x_{c, n} \mid \lambda_{t}\right) \\
\beta_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(Y_{n+1}^{C}=y_{n+1}^{C}, \ldots, Y_{N}^{C}=y_{N}^{C} \mid X_{c, n}=x_{c, n}, \lambda_{t}\right)
\end{gathered}
$$

where $y^{C, n}=\left\{y_{c, n^{\prime}}\right\}_{1 \leq c \leq C, 1 \leq n^{\prime} \leq n}$ corresponds to the observations in all patches from the initial time step until step $n$. As opposed to the case $C=1$, the variables $Y_{n+1}^{C}, \ldots, Y_{N}^{C}$ are not independent of $Y^{C, n}$ conditionally on $X_{c, n}$. As a consequence, the property $\alpha_{c, n}\left(x_{c, n}\right) \beta_{c, n}\left(x_{c, n}\right) \propto \mathbb{P}\left(X_{c, n}=x_{c, n}, y^{C, N} \mid \lambda_{t}\right)$, which is crucial to calculate $\rho$ and $\xi$ in a HMM, does no longer hold. However, the property can be recovered by defining the auxiliary variables of the E-step as:

$$
\alpha_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(Y^{C, n}=y^{C, n}, X_{c, n}=x_{c, n} \mid \lambda_{t}\right)
$$

$$
\beta_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(Y_{n+1}^{C}=y_{n+1}^{C}, \ldots, Y_{N}^{C}=y_{N}^{C} \mid Y^{C, n}=y^{C, n}, X_{c, n}=x_{c, n}, \lambda_{t}\right) .
$$

This way, the product $\alpha_{c, n}\left(x_{c, n}\right) \beta_{c, n}\left(x_{c, n}\right)$ is equal to $\mathbb{P}\left(X_{c, n}=x_{c, n}, y^{C, N} \mid \lambda_{t}\right)$.
The auxiliary variables can be defined by their recursive equations :

$$
\alpha_{c, n}\left(x_{c, n}\right) \propto \sum_{x_{c, n-1} \in \Omega_{X}} \alpha_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

A complete proof is given in Appendix C. Equality is obtained by multiplying the right part of the equation by the constant factor $K_{c, n}=\mathbb{P}\left(y_{n}^{\mathcal{C} \backslash c} \mid y^{C, n-1}\right)$. $K_{c, n}$ is the probability of the observations in all patches except $c$ at time $n$, conditionally on all the observations up to time $n-1$. Let us define:

$$
\tilde{\alpha}_{c, n}\left(x_{c, n}\right)=\sum_{x_{c, n-1} \in \Omega_{X}} \tilde{\alpha}_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

where $\tilde{\alpha}_{c, 0}\left(x_{c, 0}\right)=\alpha_{c, 0}\left(x_{c, 0}\right)$.
Similarly we can show that

$$
\beta_{c, n}\left(x_{c, n}\right) \propto \sum_{x_{c, n+1} \in \Omega_{X}} \beta_{c, n+1}\left(x_{c, n+1}\right) A_{t}\left(x_{c, n}, x_{c, n+1}, y_{n+1}^{C}\right) \phi_{t}\left(x_{c, n}, y_{c, n+1}\right)
$$

Let us define:

$$
\tilde{\beta}_{c, n}\left(x_{c, n}\right)=\sum_{x_{c, n+1} \in \Omega_{X}} \tilde{\beta}_{c, n+1}\left(x_{c, n+1}\right) A_{t}\left(x_{c, n}, x_{c, n+1}, y_{n+1}^{C}\right) \phi_{t}\left(x_{c, n}, y_{c, n+1}\right)
$$

with $\tilde{\beta}_{c, N}\left(x_{c, N}\right)={\underset{\tilde{\beta}}{c, N}}\left(x_{c, N}\right)$.
All the $\tilde{\alpha}$ and $\tilde{\beta}$ can be calculated recursively. Thus, the probabilities of interest for the E-step are calculated using the following equalities:

$$
\begin{gathered}
\rho_{c, n}\left(x_{c, n}\right)=\frac{\tilde{\beta}_{c, n}\left(x_{c, n}\right) \tilde{\alpha}_{c, n}\left(x_{c, n}\right)}{\sum_{x \in \Omega_{X}} \tilde{\beta}_{c, n}(x) \tilde{\alpha}_{c, n}(x)}, \\
\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)=\frac{A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right) \tilde{\alpha}_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right)}{\sum_{\left(x, x^{\prime}\right) \in \Omega_{X}^{2}} A\left(x^{\prime}, x, y_{n}^{C}\right) \tilde{\beta}_{c, n}(x) \tilde{\alpha}_{c, n-1}\left(x^{\prime}\right) \phi_{t}\left(x^{\prime}, y_{c, n}\right)} .
\end{gathered}
$$

We were also able to derive the expression of the likelihood in terms of the quantities $\tilde{\alpha}_{c, n}$ and $\tilde{\beta}_{c, n}$.

$$
\begin{align*}
\mathbb{P}\left(y^{C, N}\right)= & {\left[\sum_{x_{c, n} \in \Omega_{X}} \tilde{\alpha}_{c, n}\left(x_{c, n}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right)\right] } \\
& \times\left[\prod_{j=1}^{N} \prod_{l \in\{1, \ldots, C\} \backslash c} \sum_{x_{l, j-1} \in \Omega_{X}} \phi\left(x_{l, j-1}, y_{l, j}\right) \frac{\tilde{\alpha}_{l, j-1}\left(x_{l, j-1}\right)}{\sum_{x \in \Omega_{X}} \tilde{\alpha}_{l, j-1}(x)}\right] . \tag{5}
\end{align*}
$$

See Appendix D for a proof.
4.3. $M$ step for $M H M M-D F$

We present here the equations for the M step for a MHMM-DF model with indistinguishable effects of the neighbours, as defined in Section 2.2. In this case, the M-step corresponds to an update of the current estimators of the hyperparameters $\tau, \mu=\left(\mu_{0}, \mu_{1}\right)$ and $\nu=\left(\nu_{0}, \ldots, \nu_{3}\right)$ associated with $\left(\pi_{\tau}, \phi_{\mu}, A_{\nu}\right)$. If $\left(\tau_{t}, \nu_{t}, \mu_{t}\right)$ are the current parameter estimators, and $\rho_{c, n}$ and $\xi_{c, n}$ are the corresponding functions computed during the E step, then $\left(\tau_{t+1}, \nu_{t+1}, \mu_{t+1}\right)$ are obtained by maximizing $Q\left(\lambda \mid \lambda_{t}\right)$ :

$$
\begin{aligned}
Q\left(\lambda \mid \lambda_{t}\right)= & \sum_{c=1}^{C} \sum_{x \in \Omega_{X}} \ln (\pi(x)) \rho_{c, 0}(x) \\
& +\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \ln \left(A\left(x_{n-1}, x_{n}, y_{n}^{C}\right)\right) \xi_{c, n}\left(x_{n-1}, x_{n}\right) \\
& +\sum_{x_{n-1} \in \Omega_{X}} \ln \left(\phi\left(x_{n-1}, y_{c, n}\right)\right) \rho_{c, n-1}\left(x_{n-1}\right)
\end{aligned}
$$

The solution $\lambda$ that maximizes $Q\left(\lambda \mid \lambda_{t}\right)$ is a zero of the partial derivatives of $Q$. Thus:

$$
\begin{aligned}
\frac{\partial Q}{\partial \tau} & =\sum_{c=1}^{C} \sum_{x_{c, 0} \in \Omega_{X}}\left[\left(x_{c, 0}\right)+\frac{\left(1-\left|\Omega_{X}\right|\right)}{1+\exp (-\tau)}\right] \rho_{c, 0}\left(x_{c, 0}\right) \\
& =\frac{C\left(1-\left|\Omega_{X}\right|\right)}{1+\exp (-\tau)}-C+\sum_{c=1}^{C} \sum_{x_{c, 0} \in\left|\Omega_{X}\right|} x_{c, 0} \rho_{c, 0}\left(x_{c, 0}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \mu_{0}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{x_{c, n-1} \in \Omega_{X}}\left[\left(y_{c, n}\right)+\left(1-\left|\Omega_{Y}\right|\right) p_{x_{c, n-1}}\right] \rho_{c, n-1}\left(x_{c, n-1}\right), \\
\frac{\partial Q}{\partial \mu_{1}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{x_{c, n-1} \in \Omega_{X}}\left[\left(y_{c, n}\right)+\left(1-\left|\Omega_{Y}\right|\right) p_{x_{c, n-1}}\right] \frac{x_{c, n-1}}{\left|\Omega_{X}\right|} \rho_{c, n-1}\left(x_{c, n-1}\right)
\end{aligned}
$$

where we recall that $p_{x_{c, n-1}}=\frac{1}{1+\exp \left(-\left(\mu_{1} \frac{x_{C, n-1}}{\left|\Omega_{X}\right|}+\mu_{0}\right)\right)}$.

$$
\begin{aligned}
\frac{\partial Q}{\partial \nu_{0}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} q_{x_{c, n}, y_{n}^{C}} \xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right), \\
\frac{\partial Q}{\partial \nu_{1}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|} q_{x_{c, n}, y_{n}^{C}} \xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right), \\
\frac{\partial Q}{\partial \nu_{2}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \frac{y_{c, n}}{\left|\Omega_{Y}\right|} q_{x_{c, n}, y_{n}^{C}} \xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right), \\
\frac{\partial Q}{\partial \nu_{3}} & =\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \frac{f_{\operatorname{mean}}\left(y_{n}^{C} \backslash c\right.}{\left|\Omega_{Y}\right|} q_{x_{c, n}, y_{n}^{C}} \xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)
\end{aligned}
$$

where $q_{x_{c, n}, y_{n}^{C}}=\left[\left(x_{c, n}\right)+\left(1-\left|\Omega_{X}\right|\right) \frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}+\nu_{2} \frac{y_{c, n}}{\Omega_{Y} \mid}+\nu_{3} \frac{f_{\text {mean }}\left(y_{n}^{C \backslash c}\right)}{\left|\Omega_{Y}\right|}\right)\right)}\right]$.
Note that each group of hyperparameters $\tau, \nu$ and $\mu$ can be updated independently of one another, since the group of partial derivatives associated with one of them does not involve the others. The numerical method used to search for the zeros is the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm of Byrd et al. (1995).

## 5. Hidden state estimation and prediction

When managing annual plants, knowing the past history of the seed bank is valuable information that can be used to target control or conservation actions. In a MHMM-DF model, the species' dormant population of patch $c$ at all time steps up to $N$ can be recovered by computing the most likely sequence given the observed sequences of non-dormant populations in all
patches. In this section, we show how to efficiently compute this sequence by deriving the Viterbi algorithm for MHMM-DF chain-by-chain. Prediction of the state of the adult flora the next season is also valuable information for management. We also describe how to predict the most likely non-dormant state of the species at time $N+1$ based on the observations of the previous years.

### 5.1. Hidden state estimation

The Viterbi algorithm (Forney, 1973) finds the most likely sequence of hidden states. When applied to the MHMM-DF model, it makes it possible to recover the state of a species' dormant population for any patch at any time. The most likely sequence of hidden states is the sequence of hidden states that maximises the joint probability of all the hidden variables conditional on the observations (i.e. non-dormant population). Thus, the following must be computed:

$$
\underset{x^{C, N \in \Omega_{X}^{C \times N}}}{\operatorname{argmax}} \mathbb{P}\left(X^{C, N}=x^{C, N} \mid Y^{C, N}=y^{C, N}\right)
$$

which is equivalent to computing:

$$
\underset{x^{C, N} \in \Omega_{X}^{C \times N}}{\operatorname{argmax}} \mathbb{P}\left(X^{C, N}=x^{C, N}, Y^{C, N}=y^{C, N}\right) .
$$

As for the forward-backward algorithm, a direct application of the Viterbi algorithm is not possible due to the size of $\Omega_{X}^{C \times N}$. However, here again we show that we can find the most likely sequence of hidden states independently, chain-by-chain. Independent recovery can be done due to the following property of MHMM-DF:
$\max _{x^{C, N \in \Omega_{X}^{C \times N}}} \mathbb{P}\left(X^{C, N}=x^{C, N}, Y^{C, N}=y^{C, N}\right)=\prod_{c=1}^{C} \max _{x_{c}^{N} \in \Omega_{X}^{N}} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right)$.
Proof of this property can be found in Appendix E. Let us now consider a single chain, $c$, and see how we can recursively compute the most likely state of the hidden sequence for this chain. We first define:

$$
\delta_{c, n}\left(x_{c, n}\right)=\max _{x_{c}^{n-1} \in \Omega_{X}^{n-1}} \mathbb{P}\left(X_{c}^{n-1}=x_{c}^{n-1}, X_{c, n}=x_{c, n}, Y^{C, n}=y^{C, n}\right),
$$

where $x_{c}^{n-1}=\left(x_{c, 1}, \ldots, x_{c, n-1}\right) . \delta_{c, n}$ can be reformulated as:

$$
\delta_{c, n}\left(x_{c, n}\right)=\max _{x_{c, n-1} \in \Omega_{X}} K_{c, n} \delta_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

where $K_{c, n}$ is the same constant as defined in the E-step for MHMM-DF. A detailed proof can be found in Appendix E. In order to avoid the constant, we define, for all $c$ and all $n, \tilde{\delta}_{c, n}$ as:

$$
\tilde{\delta}_{c, n}\left(x_{c, n}\right)=\max _{x_{c, n-1} \in \Omega_{X}} \tilde{\delta}_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

where $\tilde{\delta}_{c, 0}=\delta_{c, 0}$. Then, in order to compute the most likely state of the hidden chain at patch $c$, we used the following intermediate quantity:

$$
\psi_{c, n}\left(x_{c, n}\right)=\underset{x_{c, n-1} \in \Omega_{X}}{\operatorname{argmax}} \tilde{\delta}_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

The Viterbi procedure is then as follows for each patch $c$ :

1. Initialization for all $x_{c, 0} \in \Omega_{X}$ :

$$
\begin{gathered}
\tilde{\delta}_{c, 0}\left(x_{c, 0}\right)=\pi\left(x_{c, 0}\right), \\
\psi_{c, 0}\left(x_{c, 0}\right)=0 .
\end{gathered}
$$

2. Forward recursion for $n \in\{1, \ldots, N\}$ :

$$
\begin{aligned}
& \tilde{\delta}_{c, n}\left(x_{c, n}\right)=\max _{x_{c, n-1} \in \Omega_{X}} \tilde{\delta}_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right), \\
& \psi_{c, n}\left(x_{c, n}\right)=\underset{x_{c, n-1} \in \Omega_{X}}{\operatorname{argmax}} \tilde{\delta}_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) .
\end{aligned}
$$

Once the recursion is done, we can collect the hidden state of patch $c$ at time $N$ by looking at the state that maximises $\tilde{\delta}_{c, N}$. In a recursive manner, we can then collect the hidden state of patch $c$ at times $N-1$ to 1 . Thus, the species' dormant populations state can be recovered with the following procedure:

1. Initialization :

$$
\hat{x}_{c, N}=\underset{x_{c, N} \in \Omega_{X}}{\operatorname{argmax}} \tilde{\delta}_{c, N}\left(x_{c, N}\right)=\underset{x_{c, N} \in \Omega_{X} x_{c}^{N-1} \in \Omega_{X}^{N-1}}{\operatorname{argmax}} \max _{c} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right) .
$$

2. Backward recursion for $n \in\{N-1, \ldots, 0\}$ :

$$
\hat{x}_{c, n}=\psi_{c, n}\left(\hat{x}_{c, n+1}\right)
$$

where $\hat{x}_{c, n}$ is the most likely state for $X_{c, n}$.

### 5.2. Prediction

We predict the species non-dormant population state in patch $c$ at time $N+1$ as the state with maximum probability given the observed past nondormant populations:

$$
\hat{y}_{c, N+1}=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \mathbb{P}\left(Y_{c, N+1}=y_{c, N+1} \mid Y^{C, N}=y^{C, N}\right)
$$

This predictor is easily obtained using the quantity $\rho_{c, N}$ computed in the E step of EM for MHMM-DF:

$$
\hat{y}_{c, N+1}=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \sum_{x_{c, N} \in \Omega_{X}} \rho_{c, N}\left(x_{c, N}\right) \phi\left(x_{c, N}, y_{c, N+1}\right)
$$

where $\hat{y}_{c, N+1}$ is the most likely state for $Y_{c, N+1}$. A detailed proof can be found in Appendix F.

## 6. Numerical experiments

Three types of experiments are done using simulated data. The first experiment evaluates the quality of the estimators and characterises situations for which estimation may be difficult. The second experiment is model selection. It illustrates how MHMM-DFs with the EM algorithm can be used to discriminate between different types of dynamics from observations. The last experiment evaluates the quality of predictions and hidden state estimation. The code used is available on Figshare and the digital object identifier is 10.6084/m9.figshare. 7796612.

All the following simulations are done with $C=10$ patches and $N=100$ time steps, where $\left|\Omega_{X}\right|=\left|\Omega_{Y}\right|=5$. For a given value of the vector $(\mu, \nu)$, we simulated the states of dormant and non-dormant populations with mean colonisation and only used observations of non-dormant populations for parameter estimation, hidden state estimation and prediction. The following numerical experiments were programmed with $R$. The M step of the EM algorithm uses the R function optimr with the limited-memory Broy-den-Fletcher-Goldfarb-Shanno algorithm. Since the EM algorithm finds a local maximum, we ran eight EM algorithms initialised with randomly generated vectors of hyperparameters and selected the resulting vector of estimators with the highest likelihood. Most of the 8 starting point converged to the same result. The EM algorithm stops if the number of iterations exceeds 100 or if $\max \left(\max \left(\left|\nu_{t}-\nu_{t-1}\right|\right), \max \left(\left|\mu_{t}-\mu_{t-1}\right|\right),\left|\tau_{t}-\tau_{t-1}\right|\right)<0.01$.


Figure 2: Evolution of the root mean squared error of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{2}, \nu_{3}\right)=(-1,-3.7,6.5,-3,4,2)$ and $\nu_{1}$ varies from 0 to 6.5 with a 0.5 step.See Table 2 for definitions of parameters.

The experiments are focused on the hyperparameters associated with $A$ and $\phi$. We used 14 vectors of hyperparameters (see Table 2 for definitions of parameters) for which we only had $\nu_{1}$, associated with the dormancy process (i.e., influence of the previous state of the dormant population on the current one), which varies from 0 to 6.5. The hyperparameters of the awakening process is fixed to $\mu=(-3.7,6.5)$. The parameter $\mu_{0}$ was chosen to have a small probability of generating non-dormant population when the dormant population is extinct (abundance class 0 ). We then chose $\mu_{1}=6.5$ to simulate situations where the influence of dormant populations on non-dormant populations is strong. The hyperparameter associated with the initial probability is fixed at $\tau=-1$. The hyperparameters associated with $A$ are fixed at $\nu=\left(-3, \nu_{1}, 4,2\right)$ : low probability of external colonisation by propagule rain,


Figure 3: Evolution of the coefficient of variation of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{2}, \nu_{3}\right)=(-1,-3.7,6.5,-3,4,2)$ and $\nu_{1}$ varies from 0 to 6.5 with a 0.5 step.See Table 2 for definitions of parameters.
and intermediate strengths for the influence of the local and neighbouring non-dormant plant populations on the dormant populations. For each of the 14 vectors, ten population trajectories were simulated, and for each trajectory the hyperparameters estimators were computed. The ten simulations were used to calculate the variance of the hyperparameters estimators, as well as the bias,the root mean squared error and the coefficient of variance.

When the hyperparameter $\nu_{1}$ associated with dormancy takes intermediate values all parameters are correctly estimated as seen from the root mean squared error (Fig. 2). However, when $\nu_{1}$ is either high or low, the variance, root mean squared error and bias of the estimators are large (see Figs. G.13, 2 and G.14), except for $\nu_{0}$ and $\mu_{0}$ (constant terms of the logistic regressions) which are always well estimated. When simulating data with a small $\nu_{1}$ (low survival of the dormant population), extinction of dormant


Figure 4: Evolution of the dormant population visited states in relation to $A$ 's hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{2}, \nu_{3}\right)=(-1,-3.7,6.5,-3,4,2)$ and $\nu_{1}$ varies from 0 to 6.5 with a 0.5 step.See Table 2 for definitions of parameters.
populations is the most visited state and the observable neighbours are often extinct. Multiple simulations with small $\nu_{1}$ generated data with neighbours only in the extinction state thus making $\nu_{3}$ unidentifiable. Logically the unidentifiable parameter $\nu_{3}$ has a high coefficient of variance as shown in Fig. 3 and a high root mean squared error shown in Fig. 2. On the other hand when $\nu_{1}$ is large, the dormant population is predominantly in the maximum state because all non-dormant populations will either maintain their previous state, though survival, or grow. This tendency was confirmed by computing state frequencies from simulations (see Fig. 4). Therefore, when $\nu_{1}$ is low, the poor quality of the estimator of $\mu_{1}$ (influence of the dormant population state on the non-dormant population state) could be explained by the fact that we seldom observe the awakening process in the data (i.e., $x_{c, n}>0$ and $y_{c, n+1}>0$ ), making estimation harder. When $\nu_{1}$ is high, (high


Figure 5: Evolution of the root mean squared error of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}\right)=(-1,-3.7,6.5,-3,4,4)$ and $\nu_{3}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.
survival of the dormant population) and all dormant populations frequently visit the maximum state, it becomes difficult to distinguish which process (dormancy, colonisation, local generation of dormant population) drives the dynamics since many different vectors of hyperparameters may generate the same data. When $\nu_{1}$ takes intermediate values, the species is not predominantly visiting one extreme state, and the species survival strategy can easily be estimated. We observed the same phenomena for $\nu_{3}$, the hyperparameter associated with mean colonisation (see Figs. G.15, 6 and 5). However, in this case, variance remains low for low values of $\nu_{3}$ because experiments were run for $\nu_{1}=4$, a value that ensures that the seed survival process occurs. Appendix H shows an example of a confidence interval for varying $\nu_{3}$ and additionally explains how to calculate confidence intervals for each hyperparameter.


Figure 6: Evolution of the Coefficient of variation of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}\right)=(-1,-3.7,6.5,-3,4,4)$ and $\nu_{3}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.

### 6.2. Model selection

Model selection compares different models and chooses the one that best explains the data. Using simulated data, we used model selection techniques to compare a model with mean spatial colonisation against a model without spatial colonisation. Additionally, we repeated the experiment with a model without dormancy against a model with dormancy. Using the Akaike Information Criterion (AIC, Akaike 2011, Akaike 1981), we tried to determine whether a specific process is negligible in the species dynamics. When the AIC is calculated for each model, the model with the smallest AIC will be selected. However, if the difference in AIC between the models is not greater than 2, we then assume that we cannot determine which model is best suited for the data (Burnham and Anderson, 2004). In addition to the AIC, we looked at the sum of the Kullback-Leibler distance between the estimated
transition probability and the exact transition probability for each model. The model with the smallest value for the sum of the Kullback-Leibler is the best suited for the data. The parametric MHMM-DF, which will be referred to as the full model, has seven hyperparameters, whereas the model without spatial colonisation $\left(\nu_{3}=0\right)$ and the model without dormancy $\left(\nu_{1}=0\right)$ both have six hyperparameters.

To select between the full model and the model without dormancy, we simulated 30 trajectories of population states, with 100 time steps and ten patches, where the hyperparameters were: $\mu=(-3.7,6.5), \tau=-1$ and $\nu=$ $\left(-3, \nu_{1}, 4,2\right)$, where $\nu_{1} \in\{0,1,2,3\}$. Since the AIC penalizes models with more parameters, for $\nu_{1}=0,1,2,3$, the full model was chosen respectively 0 times, 0 times, 5 times and 22 times out of 30 . On the other hand, the model without dormancy was chosen 20 times, 21 times, 4 times and 0 times out of 30 , respectively. In the remaining cases, the difference in AIC was smaller than 2. The Kullback-Leibler based selection method chose the full model 19 times, 20 times, 22 times and 30 times out of 30 , for $\nu_{1}=0,1,2,3$ respectively. As seen in Fig. G.13, when $\nu_{1}=0,1$ or 2 estimators have a high variance. Additionally, the most visited state for dormant populations is the extinction state, as seen in Fig. 4. In our experiment, the dynamics considered are highly similar since simulating under one of the dynamics will give similar observations. However, when the hyperparameters are $\mu=$ $(-3.7,6.5), \tau=-1$ and $\nu=\left(-3, \nu_{1}, 8,2\right)$, for $\nu_{1} \in\{0,1,2,3\}$, the full model was chosen 0 times, 11 times, 25 times and 22 times out of 30 , respectively. As can be seen the results are more precise with a higher $\nu_{2}$ (influence of the local non-dormant population on the dormant population). This is due to the variance of the estimators being lower when $\nu_{2}=8$ than when $\nu_{2}=4$.

To select a model between the full model and the model with spatial colonisation, we simulated 30 sets of data with hyperparameters : $\mu=$ $(-3.7,6.5), \tau=-1$ and $\nu=\left(-3,4,4, \nu_{3}\right)$, where $\nu_{3} \in\{0,1,2,3\}$. When simulating with $\nu_{3}$ fixed to 0 , the species dynamics does not have spatial colonisation. For $\nu_{3}=0,1,2,3$, the full model was chosen one time, three times, 24 times and 28 times out of 30 , respectively. However, for $\nu_{3}=0,1,2,3$ the difference in AIC was smaller than 2 respectively 19 times, 22 times, 6 times and 2 times out of 30 . The Kullback-Leibler model selection chose the full model respectively 7 times, 25 times, 30 times and 23 times out of 30 . The Kullback-Leibler method often favors the full model as opposed to the AIC method. This is due to the AIC penalizing models with more parameters and thus being less likely to chose a model which overfits the data.

### 6.3. Hidden state estimation and prediction



Figure 7: Mean percentage of correctly recovered and predicted populations for $\nu_{1}$ (influence of the previous dormant population state) varying from 0 to 6.5 .

We looked at the mean percentage of correctly recovered dormant population states and the mean percentage of correctly predicted next non-dormant population states. For a vector of hyperparameter, we simulated population trajectories in ten patches over 100 time steps with $\left|\Omega_{X}\right|=\left|\Omega_{Y}\right|=5$. We then simulated the non-dormant populations in the ten patches at the 101st time step, 100 times. We used the observed non-dormant population states of the first 100 time steps to estimate the hyperparameters using the EM algorithm. The estimators were then used to recover the dormant populations in the first 100 time steps over all the patches. Additionally, the estimators were used to predict the non-dormant populations in all patches for the 101st time step. The predictions were then evaluated using the 100 simulations of the 101st time step. Twenty-eight different vectors of hyperparameters, for which there were ten trajectories of populations, were simulated. For the first


Figure 8: Mean percentage of correctly recovered and predicted populations for $\nu_{0}$ (influence of external colonisation) varying from -6.5 to 0 .

14 vectors of hyperparameters, all hyperparameters were fixed except for the hyperparameter associated with dormancy, which varied from 0 to 6.5 . The hyperparameters were : $\mu=(-3.7,6.5), \tau=-1$ and $\nu=\left(-3, \nu_{1}, 4,2\right)$. The next 14 vectors of hyperparameters all had fixed values, with $\nu_{1}=4$, except for $\nu_{0}$, which varied from -6.5 to 0 .

We observed that the quality of hidden state estimation is rarely below $70 \%$ (see Figs. 7 and 8) and is better with high values of $\nu_{1}$ (Fig. 7), i.e. when dormant population survival is high. On the other hand, the quality of prediction is low for high values of $\nu_{1}$. In this case, the dormant populations frequently visit the maximum state. There is almost no stochasticity in the dynamics of dormant populations (the variance of $A$ is low) thus making their recovery easier. On the other hand, the percentage of correct prediction dropped to $40 \%$ when $\nu_{1}$ was large. The poor quality of prediction is
directly linked to the high variance in the non-dormant population distribution. Since the dormant population is predominantly in the maximum state, the non-dormant population distribution, conditional on the dormant population being in the maximum state, is a binomial distribution with parameters $\left(\left|\Omega_{Y}\right|, \frac{1}{1+\exp \left(-\left(\mu_{0}-\mu_{1}\left(\left|\Omega_{X}\right|-1\right) /\left|\Omega_{X}\right|\right)\right)}\right)$ and its state distribution is as follows: the non-dormant population has a $44 \%$ chance to be in the maximum state, a $39 \%$ chance to be in state 3 , a $13 \%$ chance to be in state 2 , a $2 \%$ chance to be in state 1 and a $0.1 \%$ chance to be extinct. As the model predicts the most probable state, it is not surprising to only have $40 \%$ of correct predictions.

For low $\nu_{1}$ the situation is easier since the simulated dormant and nondormant populations frequently visit the extinction state. In addition, the effect of the hyperparameter $\nu_{0}$, which corresponds to exogenous colonisation of the dormant population by propagule rain, contribute to adding extra noise in the species dynamics: hidden state estimation and next non-dormant population predictions are easier for low values of this parameter.

Finally, to illustrate the scalability of the EM algorithm, the same experiment was reproduced with ten time steps with 100 patches with 30 simulations. The non-dormant population at the 11th time step was then simulated 100 times. The exact values of the hyperparameters were : $\mu=(-3.7,6.5)$, $\tau=-1$ and $\nu=(-3,3,4,2)$. We obtained $80 \%$ of correct prediction of the non-dormant population states and $68.9 \%$ of correct recovery of dormant population states. This is slightly less that when estimating from 10 patches and 100 time steps, maybe because estimation is easier from a small number of patched but long trajectories of observations than for the opposite.

## 7. Discussion

We propose the MHMM-DF framework that can describe, from abundance classes of non-dormant populations, the local and regional dynamics of annual plants that can undergo a potentially infinite dormancy. In a non-parametric approach, the MHMM-DF can be defined using only three probabilities: the initial probability, the transition probability from past populations to current dormant populations and the emission probability from past dormant populations to current non-dormant populations. For the sake of sparsity, we have proposed to model these probabilities as Binomial distributions with the probability parameter modelled as a logistic function dependent on hyperparameters. Each hyperparameter is associated with a specific process in the species dynamics. They describe the colonisation process
from neighbouring patches, the exogenous colonisation process from propagule rain, the dormancy process, the influence of the locally newly produced dormant individuals and the awakening process from dormant populations to non-dormant populations.

The colonisation process from neighbouring patches for plants is associated with one hyperparameter and is described as the indistinguishable influence of neighbour non-dormant population states on a dormant population state in a patch. Several options exist to aggregate the neighbour states into a single variable and we proposed the mean colonisation. However, this choice has drawbacks. The mean colonisation can assign the same value in $\Omega_{Y}$ for very contrasted configurations of neighbours states. Additionally, when the number of patches increases, it increases the tendency for the mean colonisation process to be identical for every patch. To circumvent the first limit, we proposed the alphabetic colonisation (detailed in Appendix I), which only assigns the same aggregated value to two vectors of neighbour states if both vectors are equal though patch permutation. The influence of the neighbour patches are still indistinguishable. However contrasted distributions of the abundance classes of the neighbourhood will lead to different aggregated values. In practice, to create an ordinal variable from the vector of neighbours, we ranked the distributions of the abundance classes according to alphabetic (or lexicographic) order. We demonstrated that the parametric MHMM-DF with mean or alphabetic colonisation is generically identifiable for a least three patches.

Different colonisation processes can be implemented preferably if identifiability is verified. Instead of considering one hyperparameter for the spatial colonisation process, the colonisation process could include a hyperparameter per number of non-dormant populations in each state as modelled in Gyllenberg and Hanski (1997). Each hyperparameter associated with the colonisation process would measure the influence of the number of neighbour non-dormant populations in a specific state. However, establishing the model's identifiability is not straightforward, and moreover the model would require more data for estimation.

Modelling the awakening process with a logistic regression for the probability parameter of the binomial distribution implies that, when the dormant population is extinct, the probability of the non-dormant population being extinct is smaller than 1 . In other words, non-dormant populations may be colonised by exterior non-dormant populations. However, since plants are sessile, colonisation of non-dormant populations by exogenous factors should
not be possible. Thus the use of the logit function can be questioned when modelling the awakening process. One option is to set the probability of exterior colonisation of non-dormant populations to a very low value instead of estimating it. Alternatively, we could model the awakening probability as a zero-inflated Binomial distribution. The zero-inflated Binomial distribution would force the non-dormant population to be extinct when the dormant one is extinct, and in the other situations, the awakening process would still be a Binomial distribution with logit regression. This is obtained by setting the mixture parameter of this zero-inflated model to 1 in the case of absence of a dormant population and to 0 otherwise. The generic identifiability is still satisfied for this model.

The MHMM-DF framework is built under biological assumptions, where the dependencies between the species populations represent known biological processes. These biological processes are often modelled using precise information about the number of individuals (David et al., 2010). Even though, this type of modelling is rewarding, since predictions are more accurate, they are often computationally intensive when dealing with hidden dynamics and may have unidentifiable parameters. Modelling with abundance classes helps reduce the computational complexity. The parametric MHMM-DF is used to infer the input of each process on the species dynamics by learning statistical relationships, as opposed to mechanistic approaches that would require precise knowledge beforehand about mechanisms and environmental factors that influence the species dynamics.

The MHMM-DF relies on two main assumptions about the species dynamics. The model assumes dormant populations to be sessile and assumes that non-dormant populations do not influence one another directly. The first assumption implies that dormant populations can not travel between patches, which reduces the time and space complexity of the E step from exponential to linear in terms of patches, $O\left(\left|\Omega_{X}\right|^{2} N C\right)$, where $C$ is the number of patches, $N$ is the number of time steps, and $\left|\Omega_{X}\right|$ is the number of states of a dormant population. Reducing the computational complexity of the E step implies that dealing with a large number of patches is not an issue with a MHMM-DF. The second assumption implies that once a nondormant population is observed, the non-dormant population will be renewed in the next time step. These two assumptions are verified for annual plants. Even though the MHMM-DF was created to model the dynamics of plants in patches, the model is general and could be applied to different organisms.

In specific ecosystems, the assumption of seed immobility in the seed bank
is questionable. If it is relevant for most natural populations, where seeds are only mobile during dispersal, seeds in the seed bank of riparian ecosystems may move because of floods. Furthermore, in agricultural systems, tools or machines may transfer seeds from one seed bank to another. The second assumption, which assumes no interaction between past and current non-dormant populations, can easily be removed without hindering the complexity of the E step. Removing the assumption enables us to model the local non-dormant population's survival between two time steps and the migration of non-dormant populations between patches. The associated parametric MHMM-DF with both survival and migration of non-dormant populations would have nine hyperparameters and is identifiable. Adding the survival process of non-dormant populations is relevant for plants since it would enable us to study the dynamics of perennial plants. Even though, including the migration process is not relevant to plant dynamics, a model with migration could be applied to many more species such as fresh water snails whose dormancy process would be seen as aestivation during dry seasons (Lamy et al., 2013).

MHMM-DF can have multiple uses such as estimating a species' dynamics via hyperparameters, discriminating between two different types of dynamics, predicting the state of the next non-dormant populations and even recovering all past states of dormant populations. The results show that the estimators, obtained using the EM algorithm, have a low variance and low bias when the species' populations are not predominantly visiting either the extinction state or the maximum state. On the other hand, distinguishing the input resulting from each process is difficult when dormant populations frequently visit an extreme state since many different vectors of hyperparameters have a similar likelihood. When only observing the non-dormant populations, the predictions and hidden state estimations obtained were globally of good quality. When a population is inferred to be or to become extinct, errors in predictions and hidden state estimations seem to be dependent on exogenous colonisation. Additionally, errors in predictions and hidden state estimations depend on the variance of the Binomial distribution. The model selection technique is able to discriminate between dynamics with dormancy and dynamics without dormancy, or even dynamics with colonisation and dynamics without colonisation. In order to fully validate the parameterisation, the model should now be tested on real data. For instance, studying the dynamics of weed species with the model could be a subsequent extension.

Since multiple exterior factors can influence species dynamics, such as
temperature for plant germination (Taab, 2009; Seglias et al., 2018; ElKeblawy et al., 2017; Ueno, 2002; Gómez-González et al., 2018; Zhang et al., 2017), the logit function, associated with the awakening process, can include covariables to increase the model's realism. For weed species in agricultural fields, the species dynamics could be made to depend on crop type (Borgy et al., 2015). It would be also relevant to model different sizes or types of neighbourhood, since the distance of seed dispersal or even the paths vary between species (Nathan, 2000). Indeed colonisation may be related to specific dissemination agents, such as rivers for riparian ecosystems or agricultural machines for weed species. In the latter case, dispersal is not defined by a distance from the source but by the path formed by the successive patches in which the agricultural machines travelled before entering the current patch. In practice, the size or the paths for dispersal are not known. Using model selection and several MHMM-DF models associated to different neighbourhoods, it would be possible to determine which model is best suited for a specific species.

The framework presented is the first step towards modelling the local and regional dynamics of species with an undetectable life form. The MHMMDF can be used to describe the dynamics of annual plants by only observing standing plants. Estimating the hyperparameters associated with an annual plant will help to identify which process has the greatest influence on the seed bank's state. Moreover, groups of species with similar survival strategies can be identified using a classification based on their estimators.

## 8. Acknowledgements

This work was partially funded by ANR AGROBIOSE (ANR-2013-0001) and Région Occitanie.

Akaike, H. (1981). Likelihood of a model and information criteria. Journal of Econometrics 16(1), 3-14.

Akaike, H. (2011). Akaike's Information Criterion, pp. 25-25. Berlin, Heidelberg: Springer Berlin Heidelberg.

Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. The Annals of Statistics 37(6A), 3099-3132.

Amarasekare, P. and H. Possingham (2001). Patch dynamics and metapopulation theory: the case of successional species. Journal of Theoretical Biology 209(3), 333-344.

Borgy, B., X. Reboud, N. Peyrard, R. Sabbadin, and S. Gaba (2015). Dynamics of weeds in the soil seed bank : A hidden Markov model to estimate life history traits from standing plant. PLoS ONE 10, e0139278.

Brand, M., N. Oliver, and A. Pentland (1997). Coupled hidden Markov models for complex action recognition. Computer Vision and Pattern Recognition, $994-999$.

Bullock, J., K. Shea, and O. Skarpaas (2006). Measuring plant dispersal: an introduction to field methods and experimental design. Plant Ecology 186, 217-234.

Burnham, K. P. and D. R. Anderson (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociological methods $\mathfrak{E}$ research 33(2), 261-304.

Byrd, R., P. Lu, J. Nocedal, and C. Zhu (1995, 9). A limited memory algorithm for bound constrained optimization. SIAM Journal of Scientific Computing 16, 1190-1208.

Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. Journal of Theoretical Biology 12(1), 119-129.

David, O., A. Garnier, C. Larédo, and J. Lecomte (2010). Estimation of plant demographic parameters from stage-structured censuses. Biometrics $66(3), 875-882$.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39(1), 1-38.

Dornier, A., V. Pons, and C. PO. (2011). Colonization and extinction dynamics of an annual plant metapopulation in an urban environment. Oikos 120(8), 1240-1246.

El-Keblawy, A., H. Shabana, T. Navarro, and S. Soliman (2017). Effect of maturation time on dormancy and germination of citrullus colocynthis
(cucurbitaceae) seeds from the arabian hyper-arid deserts. BMC Plant Biology 17(1), 263.

Forney, G. (1973). The Viterbi algorithm. IEEE 63(1), 268-278.
Freckleton, R. and A. Watkinson (2002). Large-scale spatial dynamics of plants: metapopulation regional ensembles and patchy populations. Journal of Ecology 90, 419-434.

Fréville, H., R. Choquet, R. Pradel, and P. Cheptou (2013). Inferring seed bank from hidden Markov models: new insights into metapopulation dynamics in plants. Journal of Ecology, 1572-1580.

Ghahramani, Z. (1998). Learning dynamic Bayesian networks, pp. 168-197. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ghahramani, Z. and M. Jordan (1997). Factorial hidden Markov models. Mach. Learn. 29(2-3), 245-273.

Gotelli, N. J. (1991). Metapopulation models: The rescue effect, the propagule rain, and the core-satellite hypothesis. The American Naturalist 138(3), 768-776.

Gyllenberg, M. Hastings, A. and I. Hanski (1997). 5 - Structured metapopulation models. In I. Hanski and M. E. Gilpin (Eds.), Metapopulation Biology, pp. 93 - 122. San Diego: Academic Press.

Gómez-González, S., M. Paniw, K. Antunes, and F. Ojeda (2018). Heat shock and plant leachates regulate seed germination of the endangered carnivorous plant drosophyllum lusitanicum. Web Ecology 18(1), 7-13.

Han, Z., T. Lui, Q. Sun, R. Li, J. Xie, and B. Li (2014). Application of compound interest laws in biology: Reunification of existing models to develop seed bank dynamics model of annual plants. Ecological Modelling 278, 6773.

Hanski, I. (1994). A practical model of metapopulation dynamics. Journal of Animal Ecology 63(1), 151-162.

Hanski, I. (1998). Metapopulation dynamics. Nature 396, 41-49.

Jarry, M., M. Khaladi, M. Hossaert-McKey, and D. McKey (1995). Modelling the population dynamics of annual plants with seed bank and density dependent effects. Acta Biotheoretica 43(1-2), 53-65.

Lamy, T., O. Gimenez, J. Pointier, P. Jarne, and P. David (2013). Metapopulation dynamics of species with cryptic life stages. the American Naturalist 181(4), 479-491.

Levin, S., D. Cohen, and A. Hastings (1984). Dispersal strategies in patchy environments. Theoretical Population Biology 26(2), 165-191.

Levins, R., D. Vagaggini, P. Zarattini, and G. Mura (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America 15(3), 237-240.

Manna, F., R. Pradel, R. Choquet, H. Fréville, and P. Cheptou (2017). Disentangling the role of seed bank and dispersal in plant metapopulation dynamics using patch occupancy surveys. Ecology 98(10), 2662-2672.

Mistro, D., L. Rodrigues, and A. Schmid (2005). A mathematical model for dispersal of an annual plant population with a seed bank. Ecological Modelling 188, 52-61.

Nathan, R. Muller-Landau, H. (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. Trends in Ecology \& Evolution 15(7), 278-285.

Oakes, D. (1999). Direct calculation of the information matrix via the EM. Journal of the Royal Statistical Society: Series B (Statistical Methodology) $61(2), 479-482$.

Pluntz, M., S. L. Coz, N. Peyrard, R. Pradel, R. Choquet, and P.-O. Cheptou (2018). A general method for estimating seed dormancy and colonisation in annual plants from the observation of existing flora. Ecology Letters 21 (9), 1311-1318.

Quintana-Ascencio, P. F., E. S. Menges, C. W. Weekley, M. I. Kelrick, and B. Pace-Aldana (2011). Biennial cycling caused by demographic delays in a fire-adapted annual plant. Population Ecology 53(1), 131-142.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In Proceedings of the IEEE, Volume 77, pp. 257-286.

Seglias, A.and Williams, E., A. Bilge, and A. Kramer (2018). Phylogeny and source climate impact seed dormancy and germination of restorationrelevant forb species. Plos one 13(2), e019193.

Soubeyrand, S., A. Laine, I. Hanski, and A. Penttinen (2009). Spatiotemporal structure of host-pathogen interactions in a metapopulation. American Naturalist 174 (3), 308-320.

Taab, A. (2009). Seed Dormancy and Germination in Solanum nigrum and S. physalifolium as Influenced by Temperature Conditions. Ph. D. thesis, Swedish University of Agricultural Sciences Uppsala.

Ueno, K. (2002). Effects of desiccation and a change in temperature on germination of immature grains of wheat (triticum aestivum l.). Euphytica 126(1), $107-113$.

Wainwright, M. and M. Jordan (2008). Graphical models, exponential families, and variational inference. In Foundations and Trends in Machine Learning, Volume 1, pp. 1-305.

Zhang, R., J. Baskin, C. Baskin, Q. Mo, L. Chen, X. Hu, and Y. Wang (2017). Effect of population, collection year, after-ripening and incubation condition on seed germination of stipa bungeana. Scientific reports 7(1), 13893.

## Appendices

## A. Graphical structure of HMM, CHMM and FHMM



Figure A.9: Hidden Markov Chain. $Y_{n}$ variables are observed at each time step $n$ while $X_{n}$ variables (grey nodes) are not. The dynamics of the hidden variables is not influenced by the observed variables.


Figure A.10: Factorial Hidden Markov Model.

## B. Proof of generic identifiability for a MHMM-DF with the mean colonisation process

A HMM model with $N$ time steps is generically identifiable if the set of values of $\theta$ such that $\mathbb{P}\left(Y^{N}=y^{N} \mid \theta\right)=\mathbb{P}\left(Y^{N}=y^{N} \mid \theta^{\prime}\right)$ has null Lebesgue measure. We use the results for HMM here to demonstrate the generic identifiability of the MHMM-DF models with the mean colonisation process.The proof relies on Theorem 6 from Allman et al. 2009:

Theorem 1 (Theorem 6 of Allman et al. (2009)). The parameters of a HMM with $r$ hidden states and $s$ observable states are generically identifiable from the marginal distribution of $2 L+1$ consecutive variables provided $L$ satisfies


Figure A.11: Coupled Hidden Markov Model.

$$
\binom{L+s-1}{s-1} \geq r
$$

This results holds for a HMM, with a dependency structure identical to Fig. A.9. We can convert a MHMM-DF model into a HMM with hidden variable $H_{n}$ and observed variables $O_{n}$ (see Fig. B.12) by grouping the observed and hidden variables of all patches, $H_{n}=\left(X_{n}^{C}, Y_{n+1}^{C}\right)$, and by setting $O_{n}=Y_{n+1}^{C}$. Note that the observed variables are duplicated. We will denote $\phi^{h m m}$ and $A^{h m m}$ as respectively the emission and the transition probabilities of the $\operatorname{HMM}\left(H_{n}, O_{n}\right)_{n \in\{0, \ldots, N-1\}} . \phi^{h m m}$ is deterministic, and $A^{h m m}$ depends on the probabilities $\phi$ and $A$ of the MHMM-DF.

When applying Theorem 1 to the HMM defined by $\left(H_{n}, O_{n}\right)$ we obtain the following theorem

Theorem 2. The parameters ( $\phi^{h m m}, A^{h m m}$ ) of the HMM model corresponding to the conversion of a MHMM-DF are generically identifiable from seven consecutive observations, if $\left|\Omega_{X}\right| \leq\left|\Omega_{Y}\right|$, and if $C>2$.

Proof.


Figure B.12: Conversion of a MHMM-DF into a HMM. When grouping the observed and hidden variables of the MHMM-DF and duplicating the observed variables, the dependency structure is identical to a classical HMM.

The number of possible hidden states is $r=\left|\Omega_{X}\right|^{C}\left|\Omega_{Y}\right|^{C}$, the number of possible observable states is $s=\left|\Omega_{Y}\right|^{C}$ and $L=3$. We must show that the following inequality is satisfied:

$$
\binom{\left|\Omega_{Y}\right|^{C}+2}{\left|\Omega_{Y}\right|^{C}-1} \geq\left|\Omega_{X}\right|^{C}\left|\Omega_{Y}\right|^{C}
$$

The inequality can be reformulated as:

$$
\left(\left|\Omega_{Y}\right|^{C}+2\right)\left(\left|\Omega_{Y}\right|^{C}+1\right) \geq\left|\Omega_{X}\right|^{C} .
$$

Since we assume that $\left|\Omega_{X}\right| \leq\left|\Omega_{Y}\right|$, the inequality is true.

This theorem demonstrates the generic identifiability of $A_{h m m}$ and $\phi_{h m m}$, considered as non-parameterised, but not yet the generic identifiability of $\phi$ and $A$. We now consider the model with a non-distinguishable effect of the neighbours, defined by $\phi_{\mu}$ and $A_{\nu}$ and provide sufficient conditions for their generic identifiability. We need to prove that $\left(A_{\mu, \nu}^{h m m}, \phi_{\mu, \nu}^{h m m}\right)=\left(A_{\mu^{\prime}, \nu^{\prime}}^{h m m}, \phi_{\mu^{\prime}, \nu^{\prime}}^{h m m}\right)$ implies $(\mu, \nu)=\left(\mu^{\prime}, \nu^{\prime}\right)$.

In the HMM representation of a MHMM-DF, the emission matrix $\phi_{h m m}$ is deterministic and is independent of $\mu, \nu$. Therefore, we need to show that $A_{\mu, \nu}^{h m m}=A_{\mu^{\prime}, \nu^{\prime}}^{h m m}$ implies $(\mu, \nu)=\left(\mu^{\prime}, \nu^{\prime}\right)$. We recall that $A_{\mu, \nu}^{h m m}\left(h_{n}, h_{n-1}\right)=$ $\mathbb{P}\left(h_{n} \mid h_{n-1}, \nu, \mu\right)$. The HMM's transition probability is:

$$
\begin{aligned}
& \mathbb{P}\left(h_{n} \mid h_{n-1}, \nu, \mu\right)=\mathbb{P}\left(x_{n}^{C}, y_{n+1}^{C} \mid x_{n-1}^{C}, y_{n}^{C}\right) \\
& =\mathbb{P}\left(y_{n+1}^{C} \mid x_{n}^{C}\right) \mathbb{P}\left(x_{n}^{C} \mid x_{n-1}^{C}, y_{n}^{C}\right) \\
& =\prod_{c=1}^{C} \mathbb{P}\left(y_{c, n+1} \mid x_{c, n}\right) \mathbb{P}\left(x_{c, n} \mid x_{c, n-1}, y_{n}^{C}\right) \\
& =\prod_{c=1}^{C}\binom{\left|\Omega_{Y}\right|-1}{y_{c, n+1}} \\
& \times\left[\frac{1}{1+\exp \left(-\left(\mu_{0}+\mu_{1} \frac{x_{c, n}}{\mid \Omega_{X} X}\right)\right)}\right]^{\left|\Omega_{Y}\right|-1}\left[\exp \left(-\left(\mu_{0}+\mu_{1} \frac{x_{c, n}}{\Omega_{X} \mid}\right)\right)\right]^{\left|\Omega_{Y}\right|-y_{c, n+1}-1} \\
& \times\binom{\left|\Omega_{X}\right|-1}{x_{c, n}}\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}+\nu_{2} \frac{y_{c, n}}{\left|\Omega_{Y}\right|}+\nu_{3}\left(\frac{f_{\text {mean }}\left(y_{n}^{C l c}\right)}{\left.\mid f_{\text {mean }}\left(\Omega_{Y}^{C-1}\right)\right)}\right)\right)\right.}\right]^{\left|\Omega_{X}\right|-1} \\
& \times\left[\operatorname { e x p } \left(-\left(\nu_{0}+\nu_{1} \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}+\nu_{2} \frac{y_{c, n}}{\left.\left.\frac{\Omega_{Y} \mid}{}+\nu_{3}\left(\frac{f_{\text {mean }}\left(y_{n}^{C \backslash c}\right)}{\left|f_{\text {mean }}\left(\Omega^{C-1}\right)\right|}\right)\right)\right]^{\left|\Omega_{X}\right|-x_{c, n}-1}}\right.\right.\right.
\end{aligned}
$$

Theorem 3. The hyperparameter $(\mu, \nu)=\left(\mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}, \nu_{3}\right)$ of a MHMM-
$D F$ with the mean colonisation process is generically identifiable from seven consecutive observations if the following conditions hold: if $\left|\Omega_{X}\right| \leq\left|\Omega_{Y}\right|$, $\left|\Omega_{Y}\right|>1$ and $C>2$.

## Proof.

As explained above, we need to find conditions under which $\mathbb{P}\left(h_{n} \mid h_{n-1}, \nu, \mu\right)=$ $\mathbb{P}\left(h_{n} \mid h_{n-1}, \nu^{\prime}, \mu^{\prime}\right)$ implies $(\mu, \nu)=\left(\mu^{\prime}, \nu^{\prime}\right)$. Let us show one-by-one the equality for each hyperparameter. We start with $\mu_{0}$. Let us define $L 1$ as the transition probability from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$ to $\left(x_{n}^{C}, y_{n+1}^{C}\right)=$ $915\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$.

$$
\begin{aligned}
L 1\left(\mu_{0}, \nu_{0}\right)= & \binom{\left|\Omega_{Y}\right|-1}{0}^{C}\left[\frac{1}{1+\exp \left(-\mu_{0}\right)}\right]^{\left(\left|\Omega_{Y}\right|-1\right) C}\left[\exp \left(-\mu_{0}\right)\right]^{\left(\left|\Omega_{Y}\right|-1\right) C} \\
& \times\binom{\left|\Omega_{X}\right|-1}{0}^{C}\left[\frac{1}{1+\exp \left(-\nu_{0}\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right) C}\left[\exp \left(-\nu_{0}\right)\right]^{\left(\left|\Omega_{X}\right|-1\right) C} .
\end{aligned}
$$

${ }_{923}\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}\left|\Omega_{X}\right|-1 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$. Note that from now on, in the expressions used to derive identifiability, we omit the terms that are known to be equal for two distinct values of the current hyperparameter considered for identifiability. Therefore

$$
\begin{aligned}
L 3\left(\nu_{0}, \mu_{0}, \mu_{1}\right) \propto & {\left[\frac{1}{1+\exp \left(-\mu_{0}\right)}\right]^{\left(\left|\Omega_{Y}\right|-1\right)(C-1)}\left[\exp \left(-\mu_{0}\right)\right]^{\left(\left|\Omega_{Y}\right|-1\right)(C-1)} } \\
& \times\left[\frac{1}{1+\exp \left(-\left(\mu_{0}+\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}\right)\right)}\right]^{\left(\left|\Omega_{Y}\right|-1\right)} \\
& \times\left[\exp \left(-\left(\mu_{0}+\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}\right)\right)\right]^{\left(\left|\Omega_{Y}\right|-1\right)} \\
& \times\left[\frac{1}{1+\exp \left(-\nu_{0}\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right) C}\left[\exp \left(-\left(\nu_{0}\right)\right)\right]^{\left(\left|\Omega_{X}\right|-1\right)(C-1)}
\end{aligned}
$$

Thus

$$
\begin{array}{ccc}
L 3\left(\nu_{0}, \mu_{0}, \nu_{1}^{\prime}\right) & & L \\
& <=> & L 3\left(\nu_{0}, \mu_{0}, \nu_{1}\right) \\
{\left[\frac{1}{1+\exp \left(-\left(\mu_{0}+\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}^{\prime}\right)\right)}\right]\left[\exp \left(-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}^{\prime}\right)\right]} & = & {\left[\frac{1}{1+\exp \left(-\left(\mu_{0}+\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}\right)\right)}\right]\left[\exp \left(-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}\right)\right]} \\
\exp \left(-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}\right)+ & <> & \exp \left(-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}^{\prime}\right)+ \\
\exp \left(-\mu_{0}-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}^{\prime}\right) & = & \exp \left(-\mu_{0}-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}-\frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|} \mu_{1}^{\prime}\right) \\
\mu_{1} & <=> & \mu_{1}^{\prime}
\end{array}
$$

We establish the identifiability of $\nu_{1}$ by using $L 4$, the transition probabil-
ity from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}\left|\Omega_{X}\right|-1 \\ \ldots \\ \left|\Omega_{X}\right|-1\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$ to $\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right.$.

$$
\begin{aligned}
L 4\left(\mu_{0}, \nu_{0}, \nu_{1}\right) \propto & {\left[\frac{1}{1+\exp \left(-\mu_{0}\right)}\right]^{\left(\left|\Omega_{Y}\right|-1\right) C}\left[\exp \left(-\mu_{0}\right)\right]^{\left(\left|\Omega_{Y}\right|-1\right) C} } \\
& \times\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right) C} \\
& \times\left[\exp \left(-\left(\nu_{0}+\nu_{1} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right)\right]^{\left(\left|\Omega_{X}\right|-1\right) C}
\end{aligned}
$$

$$
\begin{array}{ccc}
L 4\left(\mu_{0}, \nu_{0}, \nu_{1}^{\prime}\right) & = & L 4\left(\mu_{0}, \nu_{0}, \nu_{1}\right) \\
{\left[\frac{<}{} \frac{=>}{} \frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1}^{\prime} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right)}\right]\left[\exp \left(-\nu_{1}^{\prime} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right]} & = & {\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right)}\right]\left[\exp \left(-\nu_{1} \frac{\left(\left|\Omega_{X}\right|-1\right)}{\left|\Omega_{X}\right|}\right)\right]} \\
<=> & \nu_{1}^{\prime} .
\end{array}
$$

We have established identifiability for all hyperparameters except $\nu_{2}$ and $\nu_{3}$. When $C=2$ it can be seen that $f_{\text {mean }}\left(Y_{n}^{\mathcal{C} \backslash c}\right)=Y_{c^{\prime}, n}$ with $c \neq c^{\prime}$. By using the following states defined below one can not show identifiability of $\nu_{2}$ and $\nu_{3}$. Let us assume $C>2$. For these two hyperparameters, we consider the
following twotransition probabilities: $L 5$ from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}1 \\ 0 \\ \ldots \\ 0\end{array}\right)\right)$
936 to $\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$ and $L 6$ from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}1 \\ \ldots \\ 1\end{array}\right)\right)$ to
${ }_{937}\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$.

$$
\begin{aligned}
L 5\left(\nu_{0}, \nu_{2}\right) \propto & {\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}\right)\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right)}\left[\frac{1}{1+\exp \left(-\nu_{0}\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right)(C-1)} } \\
& \times\left[\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}\right)\right)\right]^{\left(\left|\Omega_{X}\right|-1\right)}\left[\exp \left(-\nu_{0}\right)\right]^{\left(\left|\Omega_{X}\right|-1\right)(C-1)} .
\end{aligned}
$$

938
Therefore:

$$
\begin{array}{rcc}
L 5\left(\nu_{0}, \nu_{2}^{\prime}\right) & = & L 5\left(\nu_{0}, \nu_{2}\right) \\
& <= & \\
\exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}^{\prime}\right) & = & \exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}\right) \\
& <=> & \\
\nu_{2}^{\prime} & = & \nu_{2} .
\end{array}
$$

$$
\begin{aligned}
L 6\left(\nu_{0}, \nu_{2}, \nu_{3}\right) \propto & {\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{3}\right)\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right) C} } \\
& \times\left[\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{3}\right)\right)\right]^{\left(\left|\Omega_{X}\right|-1\right) C} .
\end{aligned}
$$

Therefore

$$
\begin{array}{rcc}
L 6\left(\nu_{0}, \nu_{2}, \nu_{3}^{\prime}\right) & = & L 6\left(\nu_{0}, \nu_{2}, \nu_{3}\right) \\
& <= & \\
\exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{3}^{\prime}\right) & = & \exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{3}\right) \\
& <=> & \nu_{3} .
\end{array}
$$

The MHMM-DF with mean colonisation is identifiable with at least two hidden states and at least two observable states with at least three patches and with the number of hidden states smaller or equal to the number of observable states.

## C. Expression of E step for MHMM-DF

We derive the recursive expression of the forward-backward algorithm for a chain $c$ used in the E-step of EM for MHMM-DF. First, we show how the function $Q$ can be expressed in terms of the probabilities $\rho_{c, n}\left(x_{c, n}\right)$ and $\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)$ defined as

$$
\begin{aligned}
\xi_{c, n}\left(x_{n-1}, x_{n}\right) & =\mathbb{P}\left(X_{c, n}=x_{n}, X_{c, n-1}=x_{n-1} \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right), \\
\rho_{c, n}\left(x_{n}\right) & =\mathbb{P}\left(X_{c, n}=x_{n} \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right)
\end{aligned}
$$

where $\lambda_{t}=\left(\pi_{t}, \phi_{t}, A_{t}\right)$ is the current estimator. We describe here how these probabilities can be computed from auxiliary quantities defined recursively and independently for each chain.

$$
\begin{aligned}
Q\left(\lambda \mid \lambda_{t}\right) & =E\left[\ln \left(\mathbb{P}\left(Y^{C, N}, X^{C, N} \mid \lambda\right)\right) \mid y^{C, N}, \lambda_{t}\right] \\
& =E\left[\ln \left(\prod_{c=1}^{C} \mathbb{P}\left(X_{c, 0} \mid \lambda\right) \prod_{n=1}^{N} \mathbb{P}\left(X_{c, n} \mid X_{c, n-1}, Y_{n}^{C}, \lambda\right) \mathbb{P}\left(Y_{c, n} \mid X_{c, n-1}, \lambda\right)\right) \mid y^{C, N}, \lambda_{t}\right] \\
& =\sum_{c=1}^{C} \sum_{x \in \Omega_{X}} \ln (\pi(x)) \mathbb{P}\left(X_{c, 0}=x \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right) \\
& +\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \ln \left(A\left(x_{n-1}, x_{n}, y_{n}^{C}\right)\right) \mathbb{P}\left(x_{n}, x_{n-1} \mid y^{C, N}, \lambda_{t}\right) \\
& +\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{x_{n-1} \in \Omega_{X}} \ln \left(\phi\left(x_{n-1}, y_{c, n}\right)\right) \mathbb{P}\left(X_{c, n-1}=x_{n-1} \mid Y^{C, N}=y^{C, N}, \lambda_{t}\right) \\
& =\sum_{c=1}^{C} \sum_{x \in \Omega_{X}} \ln (\pi(x)) \rho_{c, 0}(x) \\
& +\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{\left(x_{n}, x_{n-1}\right) \in \Omega_{X}^{2}} \ln \left(A\left(x_{n-1}, x_{n}, y_{n}^{C}\right)\right) \xi_{c, n}\left(x_{n-1}, x_{n}\right) \\
& +\sum_{c=1}^{C} \sum_{n=1}^{N} \sum_{x_{n-1} \in \Omega_{X}} \ln \left(\phi\left(x_{n-1}, y_{c, n}\right)\right) \rho_{c, n-1}\left(x_{n-1}\right) .
\end{aligned}
$$

Let us define $\alpha_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(y^{C, n}, x_{c, n} \mid \lambda_{t}\right)$ and $\beta_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(y_{n+1}^{C}, \ldots, y_{N}^{C} \mid y^{C, n}, x_{c, n}, \lambda_{t}\right)$. For a given chain $c$, the $\alpha_{c, n}$ can be computed recursively from $\alpha_{c, n-1}$ (independently of $\alpha_{c^{\prime}, n}$ for $c^{\prime} \neq c$ ):

$$
\begin{aligned}
\alpha_{c, n}\left(x_{c, n}\right)= & \mathbb{P}\left(y^{C, n}, X_{c, n}=x_{c, n} \mid \lambda_{t}\right) \\
= & \sum_{x_{c, n-1} \in \Omega_{X}} \mathbb{P}\left(y^{C, n}, X_{c, n}=x_{c, n}, X_{c, n-1}=x_{c, n-1} \mid \lambda_{t}\right) \\
= & \sum_{x_{c, n-1} \in \Omega_{X}} \mathbb{P}\left(y^{C, n-1}, X_{c, n-1}=x_{c, n-1} \mid \lambda_{t}\right) \\
= & \quad \sum_{x_{c, n}-1 \in \Omega_{X}} \alpha_{n}\left(y_{c, n-1}^{C}, X_{c, n}=x_{c, n} \mid y_{c, n-1}^{C, n-1}\right) \mathbb{P}\left(y_{n}^{C} \mid y^{C, n-1}, x_{c, n}, \lambda_{t}\right) \\
= & \left.\sum_{c, n-1}, \lambda_{t}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \\
& x_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \\
& \quad \times \mathbb{P}\left(y_{n}^{\mathcal{C} \backslash c} \mid y^{C, n-1}, y_{c, n}, x_{c, n-1}, \lambda_{t}\right) .
\end{aligned}
$$

Since $Y_{n}^{\mathcal{C} \backslash c}$ is independent of $X_{c, n-1}$ and $Y_{c, n}$, conditional on $y^{C, n-1}$, we have $\mathbb{P}\left(y_{n}^{\mathcal{C} \backslash c} \mid y^{C, n-1}, y_{c, n}, x_{c, n-1}\right)=\mathbb{P}\left(y_{n}^{\mathcal{C} \backslash c} \mid y^{C, n-1}\right)=K_{c, n}$, where $K_{c, n}$ is a constant in $x_{c, n-1}$ and $x_{c, n}$. Therefore,

$$
\alpha_{c, n}\left(x_{c, n}\right) \propto \sum_{x_{c, n-1} \in \Omega_{X}} \alpha_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

In practice, instead of computing $\alpha_{c, n}\left(x_{c, n}\right)$, we can compute $\tilde{\alpha}_{c, n}\left(x_{c, n}\right)$ defined as:

$$
\tilde{\alpha}_{c, n}\left(x_{c, n}\right)=\sum_{x_{c, n-1} \in \Omega_{X}} \tilde{\alpha}_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right)
$$

with $\tilde{\alpha}_{c, 0}\left(x_{c, 0}\right)=\alpha_{c, 0}\left(x_{c, 0}\right)$.
Similarly, there is a recursive expression of $\beta_{c, n}$ involving $\beta_{c, n-1}$ :

$$
\begin{aligned}
\beta_{c, n}\left(x_{c, n}\right)= & \mathbb{P}\left(y_{n+1}^{C}, \ldots, y_{N}^{C} \mid y^{C, n}, x_{c, n}, \lambda_{t}\right) \\
= & \sum_{x_{c, n+1} \in \Omega_{X}} \mathbb{P}\left(y_{n+1}^{C}, \ldots, y_{N}^{C}, x_{c, n+1} \mid y^{C, n}, x_{c, n}, \lambda_{t}\right) \\
= & \sum_{x_{c, n+1} \in \Omega_{X}} \mathbb{P}\left(y_{n+2}^{C}, \ldots, y_{N}^{C}, x_{c, n+1} \mid y^{C, n+1}, x_{c, n}, \lambda_{t}\right) \mathbb{P}\left(y_{n+1}^{C} \mid y^{C, n}, x_{c, n}, \lambda_{t}\right) \\
= & \sum_{x_{c, n+1} \in \Omega_{X}} \mathbb{P}\left(y_{n+2}^{C}, \ldots, y_{N}^{C} \mid y^{C, n+1}, x_{c, n}, x_{c, n+1}, \lambda_{t}\right) A_{t}\left(x_{c, n}, x_{c, n+1}, y_{n+1}^{C}\right) \\
= & \quad \times \phi_{t}\left(x_{c, n}, y_{c, n+1}\right) \mathbb{P}\left(y_{n+1}^{\mathcal{C} \backslash c} \mid y^{C, n}, y_{c, n+1}, x_{c, n}, \lambda_{t}\right) \\
& K_{c, n+1} \sum_{x_{c, n+1} \in \Omega_{X}} \mathbb{P}\left(y_{n+2}^{C}, \ldots, Y_{N}^{C} \mid y^{C, n+1}, x_{c, n}, x_{c, n+1}, \lambda_{t}\right) \\
& \quad \times A_{t}\left(x_{c, n}, x_{c, n+1}, y_{n+1}^{C}\right) \phi\left(x_{c, n}, y_{c, n+1}\right) .
\end{aligned}
$$

Since the vector $\left(Y_{n+2}^{C}, \ldots, Y_{N}^{C}\right)$ is independent of $X_{c, n}$ given $y^{C, n+1}, x_{c, n+1}$, we have:

$$
\beta_{c, n}\left(x_{c, n}\right) \propto \sum_{x_{c, n+1} \in \Omega_{X}} \beta_{c, n+1}\left(x_{c, n+1}\right) A_{t}\left(x_{c, n}, x_{c, n+1}, Y_{n+1}^{C}\right) \phi_{t}\left(x_{c, n}, y_{c, n+1}\right)
$$

Similarly to the forward part of the algorithm, we compute the backward part with $\tilde{\beta}_{c, n}$ defined as:

$$
\tilde{\beta}_{c, n}\left(x_{c, n}\right)=\sum_{x_{c, n+1} \in \Omega_{X}} \tilde{\beta}_{c, n+1}\left(x_{c, n+1}\right) A_{t}\left(x_{c, n}, x_{c, n+1}, y_{n+1}^{C}\right) \phi_{t}\left(x_{c, n}, y_{c, n+1}\right)
$$

with $\tilde{\beta}_{c, N}\left(x_{c, N}\right)=\beta_{c, N}\left(x_{c, N}\right)$.

The two probabilities $\rho_{c, n}\left(x_{c, n}\right)=\mathbb{P}\left(X_{c, n}=x_{c, n} \mid y^{C, N}, \lambda_{t}\right)$ and $\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)=$ $\mathbb{P}\left(X_{c, n}=x_{c, n}, X_{c, n-1}=x_{c, n-1} \mid y^{C, N}, \lambda_{t}\right)$ necessary to compute $Q$ are obtained from $\tilde{\alpha}_{c, n}$ and $\tilde{\beta}_{c, n}$ as follows:

$$
\begin{gathered}
\rho_{c, n}\left(x_{c, n}\right)=\frac{\mathbb{P}\left(X_{c, n}=x_{c, n}, y^{C, N} \mid \lambda_{t}\right)}{\mathbb{P}\left(y^{C, N} \mid \lambda_{t}\right)}=\frac{\beta_{c, n}\left(x_{c, n}\right) \alpha_{c, n}\left(x_{c, n}\right)}{\sum_{x \in \Omega_{X}} \beta_{c, n}(x) \alpha_{c, n}(x)}=\frac{\tilde{\beta}_{c, n}\left(x_{c, n}\right) \tilde{\alpha}_{c, n}\left(x_{c, n}\right)}{\sum_{x \in \Omega_{X}} \tilde{\beta}_{c, n}(x) \tilde{\alpha}_{c, n}(x)}, \\
\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)=\frac{\mathbb{P}\left(X_{c, n}=x_{c, n}, X_{c, n-1}=x_{c, n-1}, y^{C, N} \mid \lambda_{t}\right)}{\mathbb{P}\left(y^{C, N} \mid \lambda_{t}\right)} .
\end{gathered}
$$

It is possible to express the numerator in terms of $\tilde{\alpha}_{c, n}$ and $\tilde{\beta}_{c, n}$ :

$$
\begin{aligned}
\mathbb{P}\left(x_{c, n}, x_{c, n-1}, y^{C, N}\right)= & \mathbb{P}\left(X_{c, n}=x_{c, n}, X_{c, n-1}=x_{c, n-1}, y_{n}^{C} \mid \lambda_{t}\right) \\
& \times \mathbb{P}\left(\left\{y_{l}^{C}\right\}_{l \neq n} \mid x_{c, n}, x_{c, n-1}, y_{n}^{C}, \lambda_{t}\right) \\
= & A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \mathbb{P}\left(X_{c, n-1}=x_{c, n-1}, y_{n}^{C} \mid \lambda_{t}\right) \\
& \times \mathbb{P}\left(y^{C, n-1} \mid x_{c, n}, x_{c, n-1}, y_{n}^{C}, \lambda_{t}\right) \\
& \times \mathbb{P}\left(y_{n+1}^{C}, \ldots, y_{N}^{C} \mid x_{c, n}, x_{c, n-1}, y^{C, n}, \lambda_{t}\right) \\
= & A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \mathbb{P}\left(X_{c, n-1}=x_{c, n-1}, y_{n}^{C} \mid \lambda_{t}\right) \\
& \times \mathbb{P}\left(y^{C, n-1} \mid x_{c, n-1}, y_{n}^{C}, \lambda_{t}\right) \beta_{c, n}\left(x_{c, n}\right) \\
= & A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \beta_{c, n}\left(x_{c, n}\right) \mathbb{P}\left(X_{c, n-1}=x_{c, n-1}, y^{C, n} \mid \lambda_{t}\right) \\
= & A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \beta_{c, n}\left(x_{c, n}\right) \mathbb{P}\left(X_{c, n-1}=x_{c, n-1}, y^{C, n-1} \mid \lambda_{t}\right) \\
& \times \mathbb{P}\left(y_{n}^{C} \mid x_{c, n-1}, y^{C, n-1}, \lambda_{t}\right) \\
= & A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \beta_{c, n}\left(x_{c, n}\right) \alpha_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) K_{c, n} \\
\propto & A_{t}\left(x_{c, n-1}, x_{c, n}, Y_{n}^{C}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right) \tilde{\alpha}_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right) .
\end{aligned}
$$

This leads to

$$
\xi_{c, n}\left(x_{c, n-1}, x_{c, n}\right)=\frac{A_{t}\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right) \tilde{\alpha}_{c, n-1}\left(x_{c, n-1}\right) \phi_{t}\left(x_{c, n-1}, y_{c, n}\right)}{\sum_{\left(x, x^{\prime}\right) \in \Omega_{X}^{2}} A_{t}\left(x^{\prime}, x, y_{n}^{C}\right) \tilde{\beta}_{c, n}(x) \tilde{\alpha}_{c, n-1}\left(x^{\prime}\right) \phi_{t}\left(x^{\prime}, y_{c, n}\right)} .
$$

## D. Expression of the likelihood of a MHMM-DF

Since the forward-backward algorithm only requires quantity proportional to the likelihood, we never compute the likelihood of the data during EM.

However, there are situations where we are interested in computing the likelihood, such as for model selection. We derive here the likelihood expression in terms of $\tilde{\alpha}_{c, n}$ and $\tilde{\beta}_{c, n}$.

$$
\begin{aligned}
\mathbb{P}\left(y^{C, N}\right) & =\sum_{x_{c, n} \in \Omega_{X}} \mathbb{P}\left(X_{c, n}=x_{c, n}, y^{C, N}\right) \\
& =\sum_{x_{c, n} \in \Omega_{X}} \alpha_{c, n}\left(x_{c, n}\right) \beta_{c, n}\left(x_{c, n}\right) \\
& =\sum_{x_{c, n} \in \Omega_{X}}\left[\prod_{j=1}^{n} K_{c, j}\right] \tilde{\alpha}_{c, n}\left(x_{c, n}\right)\left[\prod_{j^{\prime}=n+1}^{N} K_{c, j^{\prime}}\right] \tilde{\beta}_{c, n}\left(x_{c, n}\right) \\
& =\left[\prod_{j=1}^{N} K_{c, j}\right] \sum_{x_{c, n} \in \Omega_{X}} \tilde{\alpha}_{c, n}\left(x_{c, n}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right) .
\end{aligned}
$$

One must calculate $\prod_{j=1}^{N} K_{c, j}$ in order to calculate the likelihood. First, let us show that: $\sum_{x \in \Omega_{X}}^{\sum_{\tilde{\alpha}_{c, n}}\left(x_{c, n}\right)} \tilde{\tilde{c}}_{c, n}(x)=\mathbb{P}\left(X_{c, n}=x_{c, n} \mid y^{C, n}\right)$. We know that $\alpha_{c, n}(x) \propto$ $\tilde{\alpha}_{c, n}(x)$ and that $\alpha_{c, n}(x)=\mathbb{P}\left(X_{c, n}=x_{c, n}, y^{C, n}\right)=\mathbb{P}\left(X_{c, n}=x_{c, n} \mid y^{C, n}\right) \mathbb{P}\left(y^{C, n}\right)$. Therefore, $\tilde{\alpha}_{c, n}(x) \propto \mathbb{P}\left(X_{c, n}=x_{c, n} \mid y^{C, n}\right)$. Hence, if we normalise we obtain $\underset{\sum_{x \in \Omega_{X}}^{\tilde{\alpha}_{c, n}\left(x_{c, n}\right)}}{\tilde{\alpha}_{c, n}(x)}=\mathbb{P}\left(X_{c, n}=x_{c, n} \mid y^{C, n}\right)$. We are left with computing $\prod_{j=1}^{N} K_{c, j}$.

$$
\begin{aligned}
& \prod_{j=1}^{N} K_{c, j}=\prod_{j=1}^{N} \mathbb{P}\left(y_{j}^{\mathcal{C} \backslash c} \mid y^{C, j-1}\right) \\
& =\prod_{j=1}^{N} \sum_{x_{j-1}^{\mathcal{C} \backslash} \in \Omega_{X}^{C-1}} \mathbb{P}\left(y_{j}^{\mathcal{C} \backslash c}, X_{j-1}^{\mathcal{C} \backslash c}=x_{j-1}^{\mathcal{C} \backslash c} \mid y^{C, j-1}\right) \\
& =\prod_{j=1}^{N} \sum_{x_{j-1}^{\mathcal{C} \backslash c} \in \Omega_{X}^{C-1}} \mathbb{P}\left(y_{j}^{\mathcal{C} \backslash c} \mid X_{j-1}^{\mathcal{C} \backslash c}=x_{j-1}^{\mathcal{C} \backslash c}, y^{C, j-1}\right) \mathbb{P}\left(X_{j-1}^{\mathcal{C} \backslash c}=x_{j-1}^{\mathcal{C} \backslash c} \mid y^{C, j-1}\right) \\
& =\prod_{j=1}^{N} \sum_{x_{j<1}^{c \backslash c} \in \Omega_{X}^{C-1}} \prod_{\{1, \ldots, C\} \backslash\{c\}} \mathbb{P}\left(y_{l, j} \mid X_{l, j-1}=x_{l, j-1}\right) \mathbb{P}\left(X_{l, j-1}=x_{l, j-1} \mid y^{C, j-1}\right) \\
& =\prod_{j=1}^{N} \sum_{x_{j-1}^{c \backslash c} \in \Omega_{X}^{C-1 l \in\{1, \ldots, C\} \backslash\{c\}}} \prod_{l} \phi\left(x_{l, j-1}, y_{l, j}\right) \frac{\tilde{\alpha}_{l, j-1}\left(x_{l, j-1}\right)}{\sum_{x \in \Omega_{X}}^{\tilde{\alpha}_{l, j-1}(x)}} .
\end{aligned}
$$

One can notice that each term in the product over $l$ involves only $x_{l, j-1}$ and not all the elements in $x_{j-1}^{\mathcal{C} \backslash c}$. We can therefore rewrite the last equality as:

$$
\prod_{j=1}^{N} K_{c, j}=\prod_{j=1}^{N} \prod_{l \in\{1, \ldots, C\} \backslash\{c\} x_{l, j-1} \in \Omega_{X}} \phi\left(x_{l, j-1}, y_{l, j}\right) \frac{\tilde{\alpha}_{l, j-1}\left(x_{l, j-1}\right)}{\sum_{x \in \Omega_{X}} \tilde{\alpha}_{l, j-j 1}(x)} .
$$

Finally
$\mathbb{P}\left(y^{C, N}\right)=\left[\prod_{j=1}^{N} \prod_{l \in\{1, \ldots, C\} \backslash\{c\} x_{l, j-1} \in \Omega_{X}} \phi\left(x_{l, j-1}, y_{l, j}\right) \frac{\tilde{\alpha}_{l, j-1}\left(x_{l, j-1}\right)}{\sum_{x \in \Omega_{X}} \tilde{\alpha}_{l, j-1}(x)}\right]\left[\sum_{x_{c, n} \in \Omega_{X}} \tilde{\alpha}_{c, n}\left(x_{c, n}\right) \tilde{\beta}_{c, n}\left(x_{c, n}\right)\right]$.

## E. The Viterbi algorithm for MHMM-DF

In this section, we show how to compute the most likely sequences of hidden states of all of the patches, independently for each patch. We want to maximise $\mathbb{P}\left(X^{C, N}=x^{C, N} \mid Y^{C, N}=y^{C, N}\right)$ which is equivalent to maximising $\mathbb{P}\left(X^{C, N}=x^{C, N}, Y^{C, N}=y^{C, N}\right)$. The latter can be factored over the $C$ chains as follows:

$$
\begin{aligned}
\mathbb{P}\left(X^{C, N}=x^{C, N}, Y^{C, N}=y^{C, N}\right) & =\mathbb{P}\left(X^{C, N}=x^{C, N} \mid Y^{C, N}=y^{C, N}\right) \times \mathbb{P}\left(Y^{C, N}=y^{C, N}\right) \\
& =\mathbb{P}\left(Y^{C, N}=y^{C, N}\right) \prod_{c=1}^{C} p\left(X_{c}^{N}=x_{c}^{N} \mid Y^{C, N}=y^{C, N}\right) \\
& =\mathbb{P}\left(Y^{C, N}=y^{C, N}\right) \prod_{c=1}^{C} \frac{p\left(X_{c}^{N}=x^{N}, Y^{C, N}=y^{C, N}\right.}{\mathbb{P}\left(Y^{C, N}=y^{C, N}\right)} \\
& =\frac{1}{\left(\mathbb{P}\left(Y^{C, N}=y^{C, N}\right)\right)^{C-1}} \prod_{c=1}^{C} p\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right) \\
& \propto \prod_{c=1}^{C} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right) .
\end{aligned}
$$

This implies that:

$$
\begin{aligned}
\max _{x^{C, N} \in \Omega_{X}^{C \times N}} \mathbb{P}\left(X^{C, N}=x^{C, N}, Y^{C, N}=y^{C, N}\right) & =\max _{x^{C, N \in \Omega_{X}^{C \times N}}} \prod_{c=1}^{C} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right) \\
& =\prod_{c=1}^{C} \max _{x_{c}^{N} \in \Omega_{X}^{N}} \mathbb{P}\left(X_{c}^{N}=x_{c}^{N}, Y^{C, N}=y^{C, N}\right) .
\end{aligned}
$$

Thus we can perform recovery separately per patch. Let us note $\delta_{c, n}\left(x_{c, n}\right)=$ $\max _{x_{c}^{n-1} \in \Omega_{X}^{n-1}} \mathbb{P}\left(X_{c}^{n}=x_{c}^{n}, Y^{C, n}=y^{C, n}\right)$. Here is the recursive definition of $\delta_{c, n}$ in terms of $\delta_{c, n-1}$ :

$$
\begin{aligned}
\delta_{c, n}\left(x_{c, n}\right) & =\max _{x_{c}^{n-1} \in \Omega_{X}^{n-1}} \mathbb{P}\left(x_{c}^{n-1}, y^{C, n-1}\right) \mathbb{P}\left(x_{c, n}, y_{n}^{C} \mid x_{c}^{n-1}, y^{C, n-1}\right) \\
& =\max _{x_{c, n-1} \in \Omega_{X}} \delta_{c, n-1}\left(x_{c, n-1}\right) \mathbb{P}\left(x_{c, n}, y_{n}^{C} \mid x_{c, n-1}, y^{C, n-1}\right) \\
& =\max _{x_{c, n-1} \in \Omega_{X}} \delta_{c, n-1}\left(x_{c, n-1}\right) \mathbb{P}\left(y_{n}^{C} \mid x_{c, n-1}, y^{C, n-1}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \\
& =\max _{x_{c, n-1} \in \Omega_{X}} \delta_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) \mathbb{P}\left(y_{n}^{\mathcal{C} \backslash c} \mid x_{c, n-1}, y^{C, n-1}\right) A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) \\
& =\max _{x_{c, n-1} \in \Omega_{X}} \delta_{c, n-1}\left(x_{c, n-1}\right) \phi\left(x_{c, n-1}, y_{c, n}\right) K_{c, n} A\left(x_{c, n-1}, x_{c, n}, y_{n}^{C}\right) .
\end{aligned}
$$

## F. Prediction of the next non-dormant population state for MHMMDF

The prediction of a species population state in patch $c$ at time $N+1$, given the sequences of observations of the non-dormant states in all patches up to time $N$, is defined as:

$$
\begin{aligned}
& \hat{y}_{c, N+1}=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \mathbb{P}\left(Y_{c, N+1}=y_{c, N+1} \mid Y^{C, N}=y^{C, N}\right) \\
&=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \sum_{x_{c, N} \in \Omega_{X}} \mathbb{P}\left(Y_{c, N+1}=y_{c, N+1}, X_{c, N}=x_{c, N} \mid Y^{C, N}=y^{C, N}\right) \\
&=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \sum_{c, N} \in \Omega_{X} \\
& \mathbb{P}\left(x_{c, N} \mid y^{C, N}\right) \mathbb{P}\left(y_{c, N+1} \mid x_{c, N}, y^{C, N}\right) \\
&=\underset{y_{c, N+1} \in \Omega_{Y}}{\operatorname{argmax}} \sum_{c, N} \in \Omega_{X}
\end{aligned} \rho_{c, N}\left(x_{c, N}\right) \phi\left(x_{c, N}, y_{c, N+1}\right) .
$$

## G. Variance and Bias



Figure G.13: Evolution of the variance of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{2}, \nu_{3}\right)=(-1,-3.7,6.5,-3,4,2)$ and $\nu_{1}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.


Figure G.14: Evolution of the bias of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{2}, \nu_{3}\right)=(-1,-3.7,6.5,-3,4,2)$ and $\nu_{1}$ varies from 0 to 6.5 with a 0.5 step.See Table 2 for definitions of parameters.

## H. Confidence Interval computation

The confidence interval of a hyperparameter is calculated using the Fisher information matrix $I(\lambda)$. In our case, $I(\lambda)$ is the Fisher information matrix of the transition probability and it can be calculated with $Q$ the expectation of the log likelihood of the model used in the EM algorithm (Oakes, 1999). According to Oakes (1999):

$$
I(\lambda)=-\left.\frac{\partial Q(\lambda \mid \hat{\lambda})}{\partial \lambda^{2}}\right|_{\lambda=\hat{\lambda}}
$$

The partial derivative according to one group of hyperparameters ( $\tau, \nu$ and $\mu$ ) does not involve the two other groups. Thus, $I(\nu), I(\mu), I(\tau)$ can be


Figure G.15: Evolution of the variance of the hyperparameters estimators when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}\right)=(-1,-3.7,6.5,-3,4,4)$ and $\nu_{3}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.
calculated separately. Consequently:

$$
I(\nu)=-\left.\frac{\partial Q(\nu \mid \hat{\nu})}{\partial \nu^{2}}\right|_{\nu=\hat{\nu}}
$$

Once the Fisher information matrix is calculated, we can calculate a (100(1$\epsilon)$ ) confidence interval for any hyperparameter. The confidence interval for $\nu_{1}$ is

$$
\left[\hat{\nu_{1}}-Z_{\epsilon / 2}\left(\frac{1}{\sqrt{I(\nu)}}\right)_{\nu_{1}}, \hat{\nu_{1}}+Z_{\epsilon / 2}\left(\frac{1}{\sqrt{I(\nu)}}\right)_{\nu_{1}}\right]
$$

where $\left(\frac{1}{\sqrt{I(\nu)}}\right)_{\nu_{1}}$ corresponds to the $\left(\nu_{1}, \nu_{1}\right)^{\text {th }}$ entry of the root of the inverse of the Fisher information. Additionally, $Z_{\epsilon / 2}$ corresponds to $Z_{\epsilon / 2}=-\varphi^{-1}(\epsilon / 2)$ where $\varphi$ is the cumulative distribution function of the Normal distribution.

Figure H.16: Evolution of the confidence interval of the hyperparameter $\nu_{1}$ when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}\right)=(-1,-3.7,6.5,-3,4,4)$ and $\nu_{3}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.

Figs. H. 16 and H. 17 show the confidence intervals for $\nu_{1}$ and $\nu_{3}$ when $\nu_{3}$ is varying. The red dots correspond to the lower bounds of the confidence interval and the blue dots correspond to the upper bound of the confidence intervals. The black lines correspond to the real value of the hyper parameter.


## I. Alphabetic colonisation

## I.1. Definition

We propose here a colonisation process, referred to as alphabetic colonisation, which assumes that the impact of colonisation in field $c$ depends only on the number of neighbours in each state of $\Omega_{Y}$. In other words it depends only on the distribution of the non-dormant population states in $y_{n}^{\mathcal{C} \backslash c}$. For a given patch $c$, if $m_{i}$ is the number of patches among the neighbours in


Figure H.17: Evolution of the confidence interval of the hyperparameter $\nu_{3}$ when the true values are $\left(\tau, \mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}\right)=(-1,-3.7,6.5,-3,4,4)$ and $\nu_{3}$ varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.
state $i$, then the impact of colonisation in field $c$ is determined by the vector $\left(m_{0}, \ldots, m_{\left|\Omega_{Y}\right|-1}\right)$. Each possible vector is associated, through alphabetic (or lexicographic) order, with a unique rank. This rank is then used to define the probability of success $p_{x_{c, n}, y_{n}^{C}}$ in the binomial distribution. If we denote the rank associated with $y_{n}^{\mathcal{C} \backslash c}$ by $f_{\text {alpha }}\left(y_{n}^{\mathcal{C} \backslash c}\right)$, then we have:

$$
p_{x_{c, n}, y_{n}^{C}}=\frac{1}{1+\exp \left(-\left(\nu_{0}+\nu_{1} \times \frac{x_{c, n-1}}{\left|\Omega_{X}\right|}+\nu_{2} \times \frac{y_{c, n}}{\left|\Omega_{Y}\right|}+\nu_{3} \times \frac{f_{a l p h a}\left(y_{n}^{C l c}\right)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|}\right)\right)} .
$$

where $\left(\nu_{0}, \nu_{1}, \nu_{2}, \nu_{3}\right)$ are hyper parameters and where $\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|$ is the number of possible outputs of the function $f_{\text {mean }}$.

In order to derive the expression of $f_{\text {alpha }}$, we first define $g$ a function,
from $\left|\Omega_{Y}\right|^{C-1}$ to $\left|\Omega_{Y}\right|^{C}$, which sorts the elements of $y_{n}^{\mathcal{C} \backslash c}$ in descending order and adds a zero at the last component of the vector. For example, if $C=$ 4 and $\Omega_{Y}=5, g((1,4,1))=(4,1,1,0)$. Multiple total orders for vectors exist; thus, multiple ways of associating a number with a vector of ordered observations exist. We chose alphabetical order. For example, thisimplies $(3,3,3,0)<(3,3,4,0)<(4,1,1,0)$. Let us assume $c=C$, without loss of generality, then the expression of $f_{\text {alpha }}$ is:
$f_{a l p h a}\left(y_{n}^{C-1}\right)=\sum_{c^{\prime}=C-1}^{1} \mathbb{1}_{\left\{g_{c^{\prime}}\left(y_{n}^{C-1}\right)>g_{c^{\prime}+1}\left(y_{n}^{C-1}\right)\right\}} \sum_{j=g_{c^{\prime}+1}\left(y_{n}^{C-1}\right)+1}^{g_{c^{\prime}}\left(y_{n}^{C-1}\right)}\binom{\left|\Omega_{Y}\right|-j+c^{\prime}-1}{c^{\prime}-1}$,
where $g_{l}$ is $l$ th element of $g$, and the $\binom{\left|\Omega_{Y}\right|-j+l-1}{l-1}$ is the number of unordered samplings with replacement of $l-1$ elements among $\left|\Omega_{Y}\right|-j+1$. The alphabetic colonisation considers more states that the mean colonisation: the function $f_{\text {alpha }}$ has $\binom{C+\left|\Omega_{Y}\right|-2}{C-1}$ different states, whereas $f_{\text {mean }}$ has only $\Omega_{Y}$ states. However, the computational complexity of calculating the state associated with $y_{n}^{\mathcal{C} \backslash c}$ with the alphabetic colonisation is $O(N C((C-$ 1) $\left.\left.\log (C-1)+\left|\Omega_{Y}\right|^{2}(C-1)\right)\right)$.

## I.2. Identifiability

We establish the following identifiability theorem for the alphabetic colonisation process.
Theorem 4. The hyperparameter $(\mu, \nu)=\left(\mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}, \nu_{2}, \nu_{3}\right)$ of a MHMM$D F$ with the alphabetic colonisation process is generically identifiable from seven consecutive observations if the following conditions hold: $\left|\Omega_{X}\right| \leq\left|\Omega_{Y}\right|$, $\left|\Omega_{Y}\right|>2, C>2$.

## Proof.

Identifiability of ( $\mu_{0}, \mu_{1}, \nu_{0}, \nu_{1}$ ) can be proven in the same manner as for the model with the mean colonisation process. To prove identifiability $\nu_{2}$ and $\nu_{3}$, we consider the following two transition probabilities: $L 7$ from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}1 \\ \ldots \\ 1\end{array}\right)\right)$ to $\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$ and $L 8$ from $\left(x_{n-1}^{C}, y_{n}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}2 \\ \ldots \\ 2\end{array}\right)\right)$ to $\left(x_{n}^{C}, y_{n+1}^{C}\right)=\left(\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right),\left(\begin{array}{c}0 \\ \ldots \\ 0\end{array}\right)\right)$ :

$$
\begin{aligned}
L 7\left(\nu_{0}, \nu_{2}, \nu_{3}\right) \propto & {\left[\frac{1}{1+\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{f_{\text {alpha }}(1, \ldots, 1)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}\right)\right)}\right]^{\left(\left|\Omega_{X}\right|-1\right) C} } \\
& \times\left[\exp \left(-\left(\nu_{0}+\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{f_{\text {alpha }}(1, \ldots, 1)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}\right)\right)\right]^{\left(\left|\Omega_{X}\right|-1\right) C} .
\end{aligned}
$$

$$
\begin{aligned}
L 8\left(\nu_{0}, \nu_{2}, \nu_{3}\right)-2 L 7\left(\nu_{0}, \nu_{2}, \nu_{3}\right) & = \\
= & L 8\left(\nu_{0}, \nu_{2}^{\prime}, \nu_{3}^{\prime}\right)-2 L 7\left(\nu_{0}, \nu_{2}^{\prime}, \nu_{3}^{\prime}\right) \\
\left(\frac{f_{a l p h a}(2, \ldots, 2)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|}-\frac{2 f_{\text {alpha }}(1, \ldots, 1)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|}\right) \nu_{3} & = \\
= & \left(\frac{f_{\text {alpha }}(2, \ldots, 2)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|}-\frac{2 f_{\text {alpha }}(1, \ldots, 1)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|}\right) \nu_{3}^{\prime} .
\end{aligned}
$$

Therefore:

$$
\begin{array}{rcc}
L 7\left(\nu_{0}, \nu_{2}^{\prime}, \nu_{3}^{\prime}\right) & = & L 7\left(\nu_{0}, \nu_{2}, \nu_{3}\right) \\
\exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}^{\prime}+\frac{f_{a l p h a}(1, \ldots, 1)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}^{\prime}\right) & = & \\
& < & \exp \left(\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{f_{a l p h a}(1, \ldots, 1)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}\right) \\
\frac{1}{\left|\Omega_{Y}\right|} \nu_{2}^{\prime}+\frac{f_{a l p h a}(1, \ldots, 1)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}^{\prime} & = & \frac{1}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{f_{a l p h a}(1, \ldots, 1)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3} .
\end{array}
$$

Similarly:

$$
\begin{array}{ccc}
L 8\left(\nu_{0}, \nu_{2}^{\prime}, \nu_{3}^{\prime}\right) & = & L 8\left(\nu_{0}, \nu_{2}, \nu_{3}\right) \\
\frac{2}{\left|\Omega_{Y}\right|} \nu_{2}^{\prime}+\frac{f_{a l p h a}(2, \ldots, 2)}{\left|f_{a l p h a}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3}^{\prime} & = & \frac{2}{\left|\Omega_{Y}\right|} \nu_{2}+\frac{f_{a l p h a}(2, \ldots, 2)}{\left|f_{\text {alpha }}\left(\Omega_{Y}^{C-1}\right)\right|} \nu_{3} .
\end{array}
$$

By subtracting $L 8-2 L 7$, we obtain:

Thus if $f_{\text {alpha }}(2, \ldots, 2)-2 f_{\text {alpha }}(1, \ldots, 1) \neq 0$, identifiability of $\nu_{3}$ is established. From the definition of $f_{\text {alpha }}$ we have:

$$
f_{a l p h a}(1, \ldots, 1)=\binom{\left|\Omega_{Y}\right|+C-3}{C-2}
$$

$$
f_{a l p h a}(2, \ldots, 2)=\sum_{j=1}^{2}\binom{\left|\Omega_{Y}\right|-j+C-2}{C-2} .
$$

Thus:

$$
\left.\begin{array}{rl}
f_{\text {alpha }}(2, \ldots, 2)-2 f_{\text {alpha }}(1, \ldots, 1) & =-2\binom{\left|\Omega_{Y}\right|+C-3}{C-2}+\sum_{j=1}^{2}\binom{\left|\Omega_{Y}\right|-j+C-2}{C-2} \\
& =\binom{\left|\Omega_{Y}\right|+C-4}{C-2}-\left(\left|\Omega_{Y}\right|+C-3\right.
\end{array}\right)
$$

Let us show that $f_{\text {alpha }}(2, \ldots, 2)-2 f_{\text {alpha }}(1, \ldots, 1)=0$ implies $C=2$. As $\frac{\left(\left|\Omega_{Y}\right|+C-4\right)!}{\left|\left|\Omega_{Y}\right|-2\right)!(C-2)!}>0$, if $f_{\text {alpha }}(2, \ldots, 2)-2 f_{\text {alpha }}(1, \ldots, 1)=0$ then $1-\frac{\left|\Omega_{Y}\right|+C-3}{\left|\Omega_{Y}\right|-1}=0$. 1064 Thus:

$$
\begin{array}{ccc}
1 & = & \frac{\left|\Omega_{Y}\right|+C-3}{\left|\Omega_{Y}\right|-1} \\
\left|\Omega_{Y}\right|-1 & = & \left|\Omega_{Y}\right|+C-3 \\
C & = & 2 .
\end{array}
$$

Since $f_{\text {alpha }}(2, \ldots, 2)-2 f_{\text {alpha }}(1, \ldots, 1)=0$ only when $C=2$, then when $3 \leq C$, we have $\nu_{3}=\nu_{3}^{\prime}$. Finally, by using $L 8$ or $L 7$, we easily obtain that if $\nu_{3}=\nu_{3}^{\prime}$, then $\nu_{2}=\nu_{2}^{\prime}$.

Thus, we were able to establish that if $3 \leq C$, then $\mathbb{P}\left(h_{n} \mid h_{n-1}, \nu, \mu\right)=$ $\mathbb{P}\left(h_{n} \mid h_{n-1}, \nu^{\prime}, \mu^{\prime}\right)$ implies $(\mu, \nu)=\left(\mu^{\prime}, \nu^{\prime}\right)$. To establish generic identifiability, conditions from Theorem 2 are required: $\left|\Omega_{X}\right| \leq\left|\Omega_{Y}\right|$, and $C>2$. Additionally, the number of observable states $\left|\Omega_{Y}\right|$ must be greater than 2 since three states were used to prove generic identifiability. The MHMM-DF with alphabetic colonisation is generically identifiable with at least two hidden states and at least three observable states with at least three patches and with the number of hidden states smaller or equal to the number of observable states.


[^0]:    * Corresponding author

    Preprint submittea to Elsevier

