

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 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. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

³ Sebastian Le Coz^{a,b,*}, Pierre-Olivier Cheptou^b, Nathalie Peyrard^a

⁴ ^aINRA UR 875 MIAT, Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

⁵ ^b CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valery

6 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,
- ⁹ Prediction, Expectation-Maximisation algorithm

March 3, 2019

^{*}Corresponding author Preprint submitted to Elsevier

10 1. Introduction

A species's distribution depends on its local and regional dynamics (Han-11 ski, 1998; Bullock et al., 2006). The metapopulation model, developed by 12 Levins et al. (1969), uses colonisation and extinction parameters to describe 13 a species dynamics across multiple patches. Levins' model considered coloni-14 sation to be dependent on the fraction of occupied patches. Multiple studies 15 have used the metapopulation concept on different types of species (Hanski, 16 1994; Dornier et al., 2011). However, for many organisms (plants or animals) 17 that can survive rough environmental conditions by stopping their develop-18 ment through a process called dormancy, the use of the original metapopu-19 lation model has been questioned due to the fact that the dormancy stage is 20 not modelled (Freckleton and Watkinson, 2002; Bullock et al., 2006). Fréville 21 et al. (2013) have shown that such models tend to overestimate colonisation 22 and extinction parameters for annual plants with seed banks. Understand-23 ing and modelling dormancy in a species dynamics is important in order to 24 better conserve and control the species. 25

Multiple models have been developed to represent the dynamics of species 26 that can undergo dormancy using information on the dormant and non-27 dormant populations of the species (Cohen, 1966; Levin et al., 1984; Jarry 28 et al., 1995; Amarasekare and Possingham, 2001; Mistro et al., 2005; Soubeyrand 29 et al., 2009; Han et al., 2014). However, in practice, data is often only col-30 lected on the species' non-dormant population since the dormant population 31 is hard to observe. This implies that the models mentioned above, which re-32 quire full knowledge about local dormant populations, would not be adapted 33 to estimating a species' local and regional dynamics from partial data alone. 34 A model with incomplete information where the dormancy stage is modelled 35 as a hidden variable and the non-dormant stage is the observable data is 36 more appropriate to estimate the dynamics of a species that can undergo 37 dormancy. 38

The local dynamics of species with dormancy stages has already been 39 studied using models with incomplete information (David et al., 2010; Quintana-40 Ascencio et al., 2011; Lamy et al., 2013; Fréville et al., 2013; Borgy et al., 41 2015; Manna et al., 2017). The models in Lamy et al. (2013), Fréville et al. 42 (2013), Borgy et al. (2015), and Manna et al. (2017) use a Markovian frame-43 work where the current state of the population only depends on the state 44 of the population in the previous time step. A classical extension of the 45 Markov model to deal with incomplete observation is the Hidden Markov 46

Model (HMM). In a HMM, a hidden Markov chain emits information at 47 each time step and the emitted information does not influence the states of 48 hidden chain. By considering the species non-dormant population as observ-49 able and the species dormant population as hidden, it would be tempting to 50 use HMM to study such species. However, the HMM does not include the 51 influence of a species' non-dormant populations on its dormant populations. 52 Therefore, Borgy et al. (2015) and Pluntz et al. (2018) extended the HMM 53 to include the influence of emitted observations on the hidden variables. The 54 model was used to analyse the dynamics of weed species in crop fields from 55 real data on weed abundance and weed occurrence. We will refer to this 56 model as a HMM with Data Feedback (HMM-DF). 57

The main limitation of the models developed by Pluntz et al. (2018), 58 Borgy et al. (2015), Fréville et al. (2013), Lamy et al. (2013), Quintana-59 Ascencio et al. (2011), and David et al. (2010) is the absence of an explicit 60 contribution of neighbouring patches in the colonisation process. Colonisa-61 tion, when modelled, is through seed rain, as the propagule rain effet de-62 scribed in Gotelli (1991). As a result, these authors consider patches to be 63 independent. To take colonisation between patches into account, the model 64 should include interactions between patches. Factorial HMM (Ghahramani 65 and Jordan, 1997) or Coupled HMM (Brand et al., 1997; Wainwright and 66 Jordan, 2008) are models that extend HMM to include interactions between 67 patches (see Appendix A for a graphical representation). However, Factorial 68 HMMs consider that the observation at time n depends on the state of the 69 hidden variables of all chains at time n and, more importantly, that this ob-70 servation does not influence the hidden states at n+1. Thus, the colonisation 71 process, from non-dormant populations to dormant ones, cannot be modelled 72 with a FHMM. In a Coupled HMM the hidden state of a patch at time n in-73 fluences the hidden states of other patches at time n+1. This implies that the 74 colonisation process occurs between dormant populations. This assumption 75 is often wrong since dormant populations usually cannot move. In addition 76 to having a dependency structure unsuited to model colonisation, the Ex-77 pectation Maximisation (EM) algorithm (Dempster et al., 1977), which is 78 often used for parameter estimation based on incomplete information, has 79 a time complexity exponential in terms of the number of patches, for both 80 frameworks (Ghahramani and Jordan, 1997; Brand et al., 1997). The model 81 proposed in Manna et al. (2017), which extends the model of Fréville et al. 82 (2013), considered patches to be organised on a circle with colonisation com-83 ing from the closest neighbouring patches. Their model makes it possible to 84

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 91 with Data Feedback (MHMM-DF) that includes local (within patch) and 92 regional (between patches) dynamics to describe the dynamics of species 93 that undergo dormancy. The model considers that dormant and non-dormant 94 populations depend stochastically on each other and the species non-dormant 95 and dormant populations are modelled with abundance classes. The state 96 of a dormant population in a given patch at time n is described as the 97 result of four processes : (i) dormancy, i.e., survival of a population in the 98 dormant stage between successive time steps where this survival is not limited 99 in time; (ii) locally newly produced dormant individuals; (iii) colonisation 100 from neighbouring patches; and (iv) exogenous colonisation. Processes (ii)101 and (*iii*) induce a dependency of the observed data on the hidden state. 102

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 ap-110 plying the EM algorithm on the reformulation of the MHMM-DF as a HMM 111 with a single multidimensional chain has an exponential time and space com-112 plexity in terms of the number of patches. However, parameter estimation 113 with exponential time and space complexity can only be done for a MHMM-114 DF with a small number of hidden states and a small number of patches. 115 Thus, we demonstrate that for the MHMM-DF structure, estimation using 116 the EM algorithm is achievable with a linear time and space complexity in 117 terms of the number of patches. This linear complexity is obtained thanks 118 to independence between dormant populations at time n and patch c and 119 dormant populations in other patches at previous times, conditional on ob-120 served non-dormant populations. The same independence property is used 121 for the Viterbi algorithm (Forney, 1973), which enables the recovery of the 122

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.

135 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.

140 2.1. Definition

Let us consider a set \mathcal{C} of C patches. At time $n \in \{1, \ldots, N\}$ on patch 141 $c \in \mathcal{C}$ two variables are defined: $X_{c,n}$ is the abundance class of the dormant 142 population (hidden) and $Y_{c,n}$ is the abundance class of the non-dormant 143 population (observed). Their domains are $\Omega_X = \{0, 1, \dots, |\Omega_X| - 1\}$ and 144 $\Omega_Y = \{0, 1, \dots, |\Omega_Y| - 1\}$, respectively. In order to simplify the expression of 145 probabilities, we will resort to extra notations summarised in Table 1. The 146 sequence of hidden variables of patch c forms the hidden chain c. A Multidi-147 mensional HMM with Data Feedback (MHMM-DF) of C dimensions models 148 the joint dynamics of the C chains when the following two assumptions are 149 fulfilled. First, for a given patch c at a given time n, the non-dormant popu-150 lation state $Y_{c,n+1}$ only depends on the dormant population state $X_{c,n}$. The 151 second assumption is that the dormant population state $X_{c,n+1}$ depends on 152 the dormant population state of the same chain at the previous time step 153 $X_{c,n}$, and on all non-dormant population states at time n+1, Y_{n+1}^{C} . This 154 direct dependence of hidden variables on observable variables is the data feed-155 back part of the model. The dependences in the MHMM-DF are represented 156 in Fig. 1 for two chains or patches. 157

Variable	Domain	Definition
$X_{c,n}$	$\Omega_X = \{0, 1, \dots \Omega_X - 1\}$	abundance class of dormant
		population in patch c at time n
$Y_{c,n}$	$\Omega_Y = \{0, 1, \dots \Omega_Y - 1\}$	abundance class of non-dormant
		population in patch c at time n
X_n^C	Ω_X^C	$\{X_{c,n}, 1 \le c \le C\}$
Y_n^C	Ω_X^C	$\{Y_{c,n}, 1 \le c \le C\}$
X_n^C	Ω^{nC}_X	$\{X_n^C, 1 \le n \le N\}$
Y_n^C	Ω_X^{nC}	$\{Y_n^C, 1 \le n \le N\}$

Table 1: Variables of the MHMM-DF.

The dynamics of a MHMM-DF is fully defined by three probabilities. 158 The first one is the initial probability of the hidden states, $\mathbb{P}(X_0^C = x_0^C)$. For the sake of simplicity we will assume here that $\mathbb{P}(X_0^C = x_0^C) = \prod_{c=1}^C \pi(x_{c,0})$. 159 160 The emission probability, defined as $\phi(x_{c,n-1}, y_{c,n}) = \mathbb{P}(Y_{c,n} = y_{c,n} | X_{c,n-1} =$ 161 $x_{c,n-1}$), models the awakening process, which corresponds to the generation 162 of the local non-dormant population from the local dormant population. Fi-163 nally, the transition probability of the hidden variable of chain c, defined as 164 $A(x_{c,n-1}, x_{c,n}, y_n^C) = \mathbb{P}(X_{c,n} = x_{c,n} | X_{c,n-1} = x_{c,n-1}, Y_n^C = y_n^C)$, models the 165 generation of the current dormant population at patch c, given the previous 166 dormant population state and the non-dormant population state both lo-167 cally and in the neighbourhood. Note that in ϕ and A, variable ordering first 168 corresponds to hidden variables and then to observed ones, and not to the 169 temporal ordering of the events. In Section 3, we describe these probabilities 170 in the case of annual plants. 171

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

$$\prod_{c=1}^{C} \pi(x_{c,0}) \prod_{n=1}^{N} A(x_{c,n-1}, x_{c,n}, y_n^C) \phi(x_{c,n-1}, y_{c,n}).$$
(1)

174

175 2.2. Model parameterisation

The estimation of ϕ 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 ϕ 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 ϕ 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 $x_{c,n-1}$, $\phi(x_{c,n-1}, y_{c,n})$ is modelled as a Binomial distribution with parameters $|\Omega_Y|$, and $p_{x_{c,n-1}}$. The probability $p_{x_{c,n-1}}$ is then modelled as a logistic regression: $p_{x_{c,n-1}} = \frac{1}{1+\exp(-(\mu_0+\mu_1x_{c,n-1}/|\Omega_X|))}$, where μ_0 and μ_1 are hyper parameters. Thus:

$$\phi(x_{c,n-1}, y_{c,n}) = \begin{pmatrix} |\Omega_Y| - 1 \\ y_{c,n} \end{pmatrix} \left[\frac{1}{1 + \exp(-(\mu_0 + \mu_1 \frac{x_{c,n-1}}{|\Omega_X|}))} \right]^{|\Omega_Y| - 1} \times \left[\exp(-(\mu_0 + \mu_1 \frac{x_{c,n-1}}{|\Omega_X|})) \right]^{|\Omega_Y| - y_{c,n} - 1}.$$
(2)

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 $(|\Omega_X|, p_{x_{c,n-1},y_n^C})$, 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 $y_n^{\mathcal{C}\setminus c}$, where $y_n^{\mathcal{C}\setminus c} = \{y_{c',n}, c' \in \mathcal{C}, c' \neq c\}$ is the set of observations of all the patches except c at time n.

Extracting a subset of $y_n^{\mathcal{C}\setminus c}$ can be done in order to analyse a specific 196 colonisation process. This will be the case, for example, in a metapopula-197 tion process where only populations from geographically-close patches can 198 colonise a given patch. In other situations, it can be reasonable to assume 199 that colonisation is possible with the same intensity from all patches. Coloni-200 sation scenarios can be modelled in the MHMM-DF framework by building 201 an appropriate A function. Their suitability to data can be compared using 202 model selection techniques. Here, for illustration purposes, we focused on 203 models for which each patch contributes equally to colonisation. 204

In order to limit the number of parameters in the regression model for 205 $p_{x_{c,n},y_n^C}$, we propose to aggregate $y_n^{\mathcal{C}\setminus c}$ in a one dimension variable. There are 206 several options to summarise the information in $y_n^{\mathcal{C}\setminus c}$. For instance, we will 207 refer to *mean colonisation* as the colonisation process that considers that 208 only the mean value of the non-dormant population state in patches other 209 than c influences the dormant population state in patch c. The function that 210 summarises the information is called f_{mean} and associates the set of observed 211 states of non-dormant population $y_n^{\mathcal{C}\setminus c}$ to its mean value (rounded to the first 212 smaller integer to remain in Ω_Y). It is a measure of the mean colonisation 213 capacity in the neighbourhood. The output of f_{mean} is then used to define the 214 logistic regression model for $p_{x_{c,n},y_n^C}$ as follows: 215

$$p_{x_{c,n},y_n^C} = \frac{1}{1 + \exp(-(\nu_0 + \nu_1 \times \frac{x_{c,n-1}}{|\Omega_X|} + \nu_2 \times \frac{y_{c,n}}{|\Omega_Y|} + \nu_3 \times \frac{f_{\text{mean}}(y_n^{C \setminus c})}{|\Omega_Y|}))}.$$
 (3)

²¹⁶ 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 $|\Omega_X|$ and p_{τ} where:

$$p_{\tau} = \frac{1}{1 + \exp(-\tau)}.$$
 (4)

220

Hyperparameters	Interpretation for annual plants	
Initial distribution $\mathbb{P}(x_{c,0}) = \pi(x_{c,0})$		
au	Mean trend	
Awakening process $\mathbb{P}(y_{c,n} x_{c,n-1}) = \phi(x_{c,n-1}, y_{c,n})$		
μ_0	Exogenous colonisation of non-dormant population	
	Not relevant for plants	
μ_1	Influence of seed bank state	
	on next adult plant population	
Seed bank dynamics $\mathbb{P}(x_{c,n} x_{c,n-1}, y_n^C) = (x_{c,n-1}, x_{c,n}, y_n^C)$		
$ u_0$	Exogenous colonisation by seed rain	
$ u_1 $	Influence of the previous seed bank state	
	on the current one (dormancy)	
ν_2	Influence of local seed production	
	on the current seed bank state	
$ u_3$	Influence of neighbouring seed production	
	on the current seed bank state	

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 >2under the mild assumption that N must be greater than 7 (see Appendix B).

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

The MHMM-DF is well adapted to describe the regional and local dynam-227 ics of annual plants like weeds. In order to model a species dynamics with a 228 MHMM-DF, two assumptions must be made. The first assumption imposes 229 immobility on the species' dormant populations in a patch. This assump-230 tion is usually verified for annual plants since seeds are immobile once in the 231 seed bank. The second assumption is that the species' non-dormant popu-232 lation cannot directly influence another non-dormant population. A species' 233 non-dormant population can only influence the species' dormant populations. 234 Thus, a non-dormant population will either die, become sterile or transform 235 into a dormant population at the end of each time step. This assumption 236

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 241 species dormant populations $(X_{c,n})_{(c,n)\in\{1,\dots,C\}\times\{1,\dots,N\}}$ correspond to the seed 242 bank's state at time n in patch c. The observable variables or the species? 243 non-dormant populations $(Y_{c,n})_{(c,n)\in\{1,\dots,C\}\times\{1,\dots,N\}}$ correspond to the adult 244 plant's state at time n in patch c. The seed bank state is influenced by 245 dormancy, colonisation and the current state of adult flora in the patch. All 246 these processes are represented with red arrows in Fig. 1. The blue arrows 247 correspond to germination and survival up to adulthood. 248

For a parametric MHMM-DF with indistinguishable effects of neighbours, 249 each hyperparameter is associated with a specific process of a plant's dynam-250 ics, with the exception of τ , which is associated with the initial probability 251 of dormant populations. With a parametric MHMM-DF, the annual plant 252 dynamics is described by multiple processes: survival of seeds in the seed 253 bank corresponding to the dormancy process (ν_1) ; local production of seeds 254 entering the local seed bank corresponding to the influence of the local non-255 dormant population on the local dormant population (ν_2) ; and neighbouring 256 production of seeds entering the local seed bank corresponding to the spa-257 tial colonisation process (ν_3) and the germination process corresponding to 258 (μ_1) . The exogenous colonisation process has two hyperparameters, μ_0 and 259 ν_0 . The exogenous production of seeds entering the regional seed bank by 260 seed rain corresponds to the exogenous colonisation of dormant populations 261 (ν_0) . The process modelled with the hyperparameter μ_0 corresponds to ex-262 ogenous colonisation of non-dormant populations. This process is not present 263 with annual plants. Even if this process cannot be removed, the smaller μ_0 264 is, the smaller the probability of non-dormant populations being colonised 265 by the exterior will be. For instance, with $\mu_0 = -5.7$, the probability of 266 exogenous colonisation of a non-dormant population occurring when all past 267 populations are extinct is 0.004. 268

269 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 well-273 known Forward-Backward algorithm (Rabiner, 1989), which takes advantage 274 of the linear structure of a HMM (see Fig. A.9) and variable elimination 275 to efficiently compute all conditional probabilities of the hidden variables 276 given the observations. However, when the hidden variables are multidimen-277 sional, the size of the domains and the structure of interaction can make the 278 forward-backward algorithm intractable. Since chains only depend on one 279 another through observations in MHMM-DF, we have demonstrated that 280 independently running a forward-backward per chain is possible. While re-281 maining exact, it significantly reduces the computational complexity of the 282 E step from exponential in terms of the number of patches $O(|\Omega_X|^{2C}N)$ to 283 linear $O(|\Omega_X|^2 CN)$. In this section, the EM algorithm is described for one-284 dimensional HMM and then detailed for MHMM-DF. 285

286 4.1. EM algorithm for HMM

²⁸⁷ A HMM (see Fig. A.9) is defined by: the initial probability $\pi(x_0)$, the ²⁸⁸ emission probability $\phi(x_n, y_n)$ and the transition probability $A(x_{n-1}, x_n)$. As ²⁸⁹ can be observed the transition probability does not depend on observations, ²⁹⁰ and is $A(x_{n-1}, x_n) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1})$. Let us denote $\lambda = (\pi, \phi, A)$ ²⁹¹ corresponding to the model's parameters.

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

$$Q(\lambda \mid \lambda') = \sum_{n=0}^{N} \sum_{x_n \in \Omega_X} \gamma_n(x_n) \ln(\phi(x_n, y_n)) + \sum_{x_0 \in \Omega_X} \gamma_0(x_0) \ln(\pi(x_0)) + \sum_{n=1}^{N} \sum_{x_{n-1} \in \Omega_X} \sum_{x_n \in \Omega_X} \xi_n(x_{n-1}, x_n) \ln(A(x_{n-1}, x_n))$$

where $\gamma_n(x_n) = \mathbb{P}(X_n = x_n | Y^N = y^N, \lambda_t)$ and $\xi_n(x_{n-1}, x_n) = \mathbb{P}(X_{n-1} = x_{n-1}, X_n = x_n | Y^N = y^N, \lambda_t).$

We denote λ_t as the current estimator at iteration t of the algorithm. The E-steps and M-steps are defined as follows:

298

E-step: compute all the probabilities $\gamma_n(x_n)$ and $\xi_n(x_{n-1}, x_n)$.

³⁰⁰ **M-step**: update the parameter λ by maximising $Q(\lambda \mid \lambda_t)$.

301

The direct application of the EM algorithm to a HMM where the hidden variable is of dimension C (i.e., $x_n = (x_{1,n}, \ldots, x_{c,n})$) 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.

308 4.2. E step for MHMM-DF

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

$$\mathbb{P}(X^{C,N} = x^{C,N} | Y^{C,N} = y^{C,N}) = \prod_{c=1}^{C} \mathbb{P}(X_c^N = x_c^N | Y^{C,N} = y^{C,N}).$$

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' go through an observed variable. This implies that the function $Q(\lambda \mid \lambda')$ can be written in terms of the following probabilities :

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

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 :

$$\alpha_{c,n}(x_{c,n}) = \mathbb{P}(Y^{C,n} = y^{C,n}, X_{c,n} = x_{c,n} | \lambda_t)$$

$$\beta_{c,n}(x_{c,n}) = \mathbb{P}(Y^C_{n+1} = y^C_{n+1}, \dots, Y^C_N = y^C_N | X_{c,n} = x_{c,n}, \lambda_t)$$

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

$$\alpha_{c,n}(x_{c,n}) = \mathbb{P}(Y^{C,n} = y^{C,n}, X_{c,n} = x_{c,n} | \lambda_t),$$

327

$$\beta_{c,n}(x_{c,n}) = \mathbb{P}(Y_{n+1}^C = y_{n+1}^C, \dots, Y_N^C = y_N^C | Y^{C,n} = y^{C,n}, X_{c,n} = x_{c,n}, \lambda_t)$$

This way, the product $\alpha_{c,n}(x_{c,n})\beta_{c,n}(x_{c,n})$ is equal to $\mathbb{P}(X_{c,n}=x_{c,n},y^{C,N}|\lambda_t)$.

³²⁹ The auxiliary variables can be defined by their recursive equations :

$$\alpha_{c,n}(x_{c,n}) \propto \sum_{x_{c,n-1} \in \Omega_X} \alpha_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}) A_t(x_{c,n-1}, x_{c,n}, y_n^C).$$

A complete proof is given in Appendix C. Equality is obtained by multiplying the right part of the equation by the constant factor $K_{c,n} = \mathbb{P}(y_n^{C\setminus c}|y^{C,n-1})$. $K_{c,n}$ 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:

$$\tilde{\alpha}_{c,n}(x_{c,n}) = \sum_{x_{c,n-1} \in \Omega_X} \tilde{\alpha}_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}) A_t(x_{c,n-1}, x_{c,n}, y_n^C)$$

334 where $\tilde{\alpha}_{c,0}(x_{c,0}) = \alpha_{c,0}(x_{c,0})$.

335 Similarly we can show that

$$\beta_{c,n}(x_{c,n}) \propto \sum_{x_{c,n+1} \in \Omega_X} \beta_{c,n+1}(x_{c,n+1}) A_t(x_{c,n}, x_{c,n+1}, y_{n+1}^C) \phi_t(x_{c,n}, y_{c,n+1}).$$

336 Let us define:

$$\tilde{\beta}_{c,n}(x_{c,n}) = \sum_{x_{c,n+1} \in \Omega_X} \tilde{\beta}_{c,n+1}(x_{c,n+1}) A_t(x_{c,n}, x_{c,n+1}, y_{n+1}^C) \phi_t(x_{c,n}, y_{c,n+1})$$

337 with $\tilde{\beta}_{c,N}(x_{c,N}) = \underset{\sim}{\beta}_{c,N}(x_{c,N}).$

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:

$$\rho_{c,n}(x_{c,n}) = \frac{\tilde{\beta}_{c,n}(x_{c,n})\tilde{\alpha}_{c,n}(x_{c,n})}{\sum\limits_{x \in \Omega_X} \tilde{\beta}_{c,n}(x)\tilde{\alpha}_{c,n}(x)},$$

$$\xi_{c,n}(x_{c,n-1}, x_{c,n}) = \frac{A_t(x_{c,n-1}, x_{c,n}, y_n^C)\tilde{\beta}_{c,n}(x_{c,n})\tilde{\alpha}_{c,n-1}(x_{c,n-1})\phi_t(x_{c,n-1}, y_{c,n})}{\sum\limits_{(x,x')\in\Omega_X^2} A(x', x, y_n^C)\tilde{\beta}_{c,n}(x)\tilde{\alpha}_{c,n-1}(x')\phi_t(x', y_{c,n})}.$$

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

$$\mathbb{P}(y^{C,N}) = \left[\sum_{x_{c,n} \in \Omega_X} \tilde{\alpha}_{c,n}(x_{c,n}) \tilde{\beta}_{c,n}(x_{c,n}) \right] \\ \times \left[\prod_{j=1}^N \prod_{l \in \{1,\dots,C\} \setminus c} \sum_{x_{l,j-1} \in \Omega_X} \phi(x_{l,j-1}, y_{l,j}) \frac{\tilde{\alpha}_{l,j-1}(x_{l,j-1})}{\sum_{x \in \Omega_X} \tilde{\alpha}_{l,j-1}(x)} \right].$$
(5)

342 See Appendix D for a proof.

343 4.3. M step for MHMM-DF

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 τ , $\mu = (\mu_0, \mu_1)$ and $\nu = (\nu_0, \dots, \nu_3)$ associated with $(\pi_{\tau}, \phi_{\mu}, A_{\nu})$. If (τ_t, ν_t, μ_t) are the current parameter estimators, and $\rho_{c,n}$ and $\xi_{c,n}$ are the corresponding functions computed during the E step, then $(\tau_{t+1}, \nu_{t+1}, \mu_{t+1})$ are obtained by maximizing $Q(\lambda|\lambda_t)$:

$$Q(\lambda|\lambda_t) = \sum_{c=1}^C \sum_{x \in \Omega_X} \ln(\pi(x)) \rho_{c,0}(x) + \sum_{c=1}^C \sum_{n=1}^N \sum_{(x_n, x_{n-1}) \in \Omega_X^2} \ln(A(x_{n-1}, x_n, y_n^C)) \xi_{c,n}(x_{n-1}, x_n) + \sum_{x_{n-1} \in \Omega_X} \ln(\phi(x_{n-1}, y_{c,n})) \rho_{c,n-1}(x_{n-1}).$$

The solution λ that maximizes $Q(\lambda|\lambda_t)$ is a zero of the partial derivatives of Q_{22} . Thus:

$$\frac{\partial Q}{\partial \tau} = \sum_{c=1}^{C} \sum_{x_{c,0} \in \Omega_X} [(x_{c,0}) + \frac{(1 - |\Omega_X|)}{1 + \exp(-\tau)}] \rho_{c,0}(x_{c,0})$$
$$= \frac{C(1 - |\Omega_X|)}{1 + \exp(-\tau)} - C + \sum_{c=1}^{C} \sum_{x_{c,0} \in |\Omega_X|} x_{c,0} \rho_{c,0}(x_{c,0}),$$

$$\frac{\partial Q}{\partial \mu_0} = \sum_{c=1}^C \sum_{n=1}^N \sum_{x_{c,n-1} \in \Omega_X} \left[(y_{c,n}) + (1 - |\Omega_Y|) p_{x_{c,n-1}} \right] \rho_{c,n-1}(x_{c,n-1}),$$

$$\frac{\partial Q}{\partial \mu_1} = \sum_{c=1}^C \sum_{n=1}^N \sum_{x_{c,n-1} \in \Omega_X} \left[(y_{c,n}) + (1 - |\Omega_Y|) p_{x_{c,n-1}} \right] \frac{x_{c,n-1}}{|\Omega_X|} \rho_{c,n-1}(x_{c,n-1}),$$

353 where we recall that $p_{x_{c,n-1}} = \frac{1}{1 + \exp(-(\mu_1 \frac{x_{c,n-1}}{|\Omega_X|} + \mu_0))}$.

$$\begin{aligned} \frac{\partial Q}{\partial \nu_0} &= \sum_{c=1}^C \sum_{n=1}^N \sum_{(x_n, x_{n-1}) \in \Omega_X^2} q_{x_{c,n}, y_n^C} \,\xi_{c,n}(x_{c,n-1}, x_{c,n}), \\ \frac{\partial Q}{\partial \nu_1} &= \sum_{c=1}^C \sum_{n=1}^N \sum_{(x_n, x_{n-1}) \in \Omega_X^2} \frac{x_{c,n-1}}{|\Omega_X|} q_{x_{c,n}, y_n^C} \,\xi_{c,n}(x_{c,n-1}, x_{c,n}), \\ \frac{\partial Q}{\partial \nu_2} &= \sum_{c=1}^C \sum_{n=1}^N \sum_{(x_n, x_{n-1}) \in \Omega_X^2} \frac{y_{c,n}}{|\Omega_Y|} q_{x_{c,n}, y_n^C} \,\xi_{c,n}(x_{c,n-1}, x_{c,n}), \\ \frac{\partial Q}{\partial \nu_3} &= \sum_{c=1}^C \sum_{n=1}^N \sum_{(x_n, x_{n-1}) \in \Omega_X^2} \frac{f_{\text{mean}}(y_n^{C\setminus C})}{|\Omega_Y|} q_{x_{c,n}, y_n^C} \,\xi_{c,n}(x_{c,n-1}, x_{c,n}). \end{aligned}$$

354 where $q_{x_{c,n},y_n^C} = [(x_{c,n}) + (1 - |\Omega_X|) \frac{1}{1 + \exp(-(\nu_0 + \nu_1 \frac{x_{c,n-1}}{|\Omega_X|} + \nu_2 \frac{y_{c,n}}{|\Omega_Y|} + \nu_3 \frac{f_{\text{mean}}(y_n^{C\setminus C})}{|\Omega_Y|})]}$.

Note that each group of hyperparameters τ , ν and μ 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).

³⁶⁰ 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.

372 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:

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

³⁸⁰ which is equivalent to computing:

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

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:

$$\max_{x^{C,N} \in \Omega_X^{C \times N}} \mathbb{P}(X^{C,N} = x^{C,N}, Y^{C,N} = y^{C,N}) = \prod_{c=1}^C \max_{x_c^N \in \Omega_X^N} \mathbb{P}(X_c^N = x_c^N, Y^{C,N} = y^{C,N}).$$

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:

$$\delta_{c,n}(x_{c,n}) = \max_{x_c^{n-1} \in \Omega_X^{n-1}} \mathbb{P}(X_c^{n-1} = x_c^{n-1}, X_{c,n} = x_{c,n}, Y^{C,n} = y^{C,n}),$$

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

$$\delta_{c,n}(x_{c,n}) = \max_{x_{c,n-1} \in \Omega_X} K_{c,n} \delta_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A(x_{c,n-1}, x_{c,n}, y_n^C)$$

³⁸⁷ where $K_{c,n}$ is the same constant as defined in the E-step for MHMM-DF. A ³⁸⁸ detailed proof can be found in Appendix E. In order to avoid the constant, ³⁸⁹ we define, for all c and all n, $\tilde{\delta}_{c,n}$ as:

$$\tilde{\delta}_{c,n}(x_{c,n}) = \max_{x_{c,n-1} \in \Omega_X} \tilde{\delta}_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A(x_{c,n-1}, x_{c,n}, y_n^C)$$

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:

$$\psi_{c,n}(x_{c,n}) = \operatorname*{argmax}_{x_{c,n-1} \in \Omega_X} \tilde{\delta}_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A(x_{c,n-1}, x_{c,n}, y_n^C).$$

³⁹² The Viterbi procedure is then as follows for each patch c:

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

$$\tilde{\delta}_{c,0}(x_{c,0}) = \pi(x_{c,0}),$$

$$\psi_{c,0}(x_{c,0}) = 0.$$

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

$$\tilde{\delta}_{c,n}(x_{c,n}) = \max_{x_{c,n-1} \in \Omega_X} \tilde{\delta}_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A(x_{c,n-1}, x_{c,n}, y_n^C),$$

$$\psi_{c,n}(x_{c,n}) = \operatorname*{argmax}_{x_{c,n-1} \in \Omega_X} \tilde{\delta}_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A(x_{c,n-1}, x_{c,n}, y_n^C).$$

396

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:

402 1. Initialization :

$$\hat{x}_{c,N} = \underset{x_{c,N} \in \Omega_X}{\operatorname{argmax}} \tilde{\delta}_{c,N}(x_{c,N}) = \underset{x_{c,N} \in \Omega_X}{\operatorname{argmax}} \underset{x_c}{\operatorname{max}} \underset{x_c}{\operatorname{max}} \mathbb{P}(X_c^N = x_c^N, Y^{C,N} = y^{C,N}).$$

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

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

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

405 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:

$$\hat{y}_{c,N+1} = \operatorname*{argmax}_{y_{c,N+1} \in \Omega_Y} \mathbb{P}(Y_{c,N+1} = y_{c,N+1} | Y^{C,N} = y^{C,N}).$$

⁴⁰⁹ This predictor is easily obtained using the quantity $\rho_{c,N}$ computed in the E ⁴¹⁰ step of EM for MHMM-DF:

$$\hat{y}_{c,N+1} = \underset{y_{c,N+1} \in \Omega_Y}{\operatorname{argmax}} \sum_{x_{c,N} \in \Omega_X} \rho_{c,N}(x_{c,N}) \phi(x_{c,N}, y_{c,N+1})$$

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.

413 6. Numerical experiments

Three types of experiments are done using simulated data. The first ex-414 periment evaluates the quality of the estimators and characterises situations 415 for which estimation may be difficult. The second experiment is model selec-416 tion. It illustrates how MHMM-DFs with the EM algorithm can be used to 417 discriminate between different types of dynamics from observations. The last 418 experiment evaluates the quality of predictions and hidden state estimation. 419 The code used is available on Figshare and the digital object identifier is 420 10.6084/m9.figshare.7796612. 421

All the following simulations are done with C = 10 patches and N = 100422 time steps, where $|\Omega_X| = |\Omega_Y| = 5$. For a given value of the vector (μ, ν) , 423 we simulated the states of dormant and non-dormant populations with mean 424 colonisation and only used observations of non-dormant populations for pa-425 rameter estimation, hidden state estimation and prediction. The follow-426 ing numerical experiments were programmed with R. The M step of the 427 EM algorithm uses the R function optimr with the limited-memory Broy-428 den–Fletcher–Goldfarb–Shanno algorithm. Since the EM algorithm finds a 420 local maximum, we ran eight EM algorithms initialised with randomly gen-430 erated vectors of hyperparameters and selected the resulting vector of esti-431 mators with the highest likelihood. Most of the 8 starting point converged to 432 the same result. The EM algorithm stops if the number of iterations exceeds 433 100 or if $\max(\max(|\nu_t - \nu_{t-1}|), \max(|\mu_t - \mu_{t-1}|), |\tau_t - \tau_{t-1}|) < 0.01.$ 434



Figure 2: Evolution of the root mean squared error of the hyperparameters estimators when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_2, \nu_3) = (-1, -3.7, 6.5, -3, 4, 2)$ and ν_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 A436 and ϕ . We used 14 vectors of hyperparameters (see Table 2 for definitions of 437 parameters) for which we only had ν_1 , associated with the dormancy process 438 (i.e., influence of the previous state of the dormant population on the cur-439 rent one), which varies from 0 to 6.5. The hyperparameters of the awakening 440 process is fixed to $\mu = (-3.7, 6.5)$. The parameter μ_0 was chosen to have a 441 small probability of generating non-dormant population when the dormant 442 population is extinct (abundance class 0). We then chose $\mu_1 = 6.5$ to simu-443 late situations where the influence of dormant populations on non-dormant 444 populations is strong. The hyperparameter associated with the initial proba-445 bility is fixed at $\tau = -1$. The hyperparameters associated with A are fixed at 446 $\nu = (-3, \nu_1, 4, 2)$: low probability of external colonisation by propagule rain, 447



Figure 3: Evolution of the coefficient of variation of the hyperparameters estimators when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_2, \nu_3) = (-1, -3.7, 6.5, -3, 4, 2)$ and ν_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 ν_1 associated with dormancy takes intermediate values all parameters are correctly estimated as seen from the root mean squared error (Fig. 2). However, when ν_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 ν_0 and μ_0 (constant terms of the logistic regressions) which are always well estimated. When simulating data with a small ν_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 $(\tau, \mu_0, \mu_1, \nu_0, \nu_2, \nu_3) = (-1, -3.7, 6.5, -3, 4, 2)$ and ν_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 461 extinct. Multiple simulations with small ν_1 generated data with neighbours 462 only in the extinction state thus making ν_3 unidentifiable. Logically the 463 unidentifiable parameter ν_3 has a high coefficient of variance as shown in 464 Fig. 3 and a high root mean squared error shown in Fig. 2. On the other 465 hand when ν_1 is large, the dormant population is predominantly in the max-466 imum state because all non-dormant populations will either maintain their 467 previous state, though survival, or grow. This tendency was confirmed by 468 computing state frequencies from simulations (see Fig. 4). Therefore, when 469 ν_1 is low, the poor quality of the estimator of μ_1 (influence of the dormant 470 population state on the non-dormant population state) could be explained 471 by the fact that we seldom observe the awakening process in the data (i.e., 472 $x_{c,n} > 0$ and $y_{c,n+1} > 0$, making estimation harder. When ν_1 is high, (high 473



Figure 5: Evolution of the root mean squared error of the hyperparameters estimators when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2) = (-1, -3.7, 6.5, -3, 4, 4)$ and ν_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 474 visit the maximum state, it becomes difficult to distinguish which process 475 (dormancy, colonisation, local generation of dormant population) drives the 476 dynamics since many different vectors of hyperparameters may generate the 477 same data. When ν_1 takes intermediate values, the species is not predomi-478 nantly visiting one extreme state, and the species survival strategy can easily 479 be estimated. We observed the same phenomena for ν_3 , the hyperparame-480 ter associated with mean colonisation (see Figs. G.15, 6 and 5). However, 481 in this case, variance remains low for low values of ν_3 because experiments 482 were run for $\nu_1 = 4$, a value that ensures that the seed survival process 483 occurs. Appendix H shows an example of a confidence interval for varying 484 ν_3 and additionally explains how to calculate confidence intervals for each 485 hyperparameter. 486



Figure 6: Evolution of the Coefficient of variation of the hyperparameters estimators when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2) = (-1, -3.7, 6.5, -3, 4, 4)$ and ν_3 varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.

487 6.2. Model selection

Model selection compares different models and chooses the one that best 488 explains the data. Using simulated data, we used model selection techniques 489 to compare a model with mean spatial colonisation against a model without 490 spatial colonisation. Additionally, we repeated the experiment with a model 491 without dormancy against a model with dormancy. Using the Akaike Infor-492 mation Criterion (AIC, Akaike 2011, Akaike 1981), we tried to determine 493 whether a specific process is negligible in the species dynamics. When the 494 AIC is calculated for each model, the model with the smallest AIC will be 495 selected. However, if the difference in AIC between the models is not greater 496 than 2, we then assume that we cannot determine which model is best suited 497 for the data (Burnham and Anderson, 2004). In addition to the AIC, we 498 looked at the sum of the Kullback-Leibler distance between the estimated 499

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 ($\nu_3 = 0$) and the model without dormancy ($\nu_1 = 0$) both have six hyperparameters.

To select between the full model and the model without dormancy, we 506 simulated 30 trajectories of population states, with 100 time steps and ten 507 patches, where the hyperparameters were: $\mu = (-3.7, 6.5), \tau = -1$ and $\nu =$ 508 $(-3, \nu_1, 4, 2)$, where $\nu_1 \in \{0, 1, 2, 3\}$. Since the AIC penalizes models with 509 more parameters, for $\nu_1 = 0, 1, 2, 3$, the full model was chosen respectively 510 0 times, 0 times, 5 times and 22 times out of 30. On the other hand, the 511 model without dormancy was chosen 20 times, 21 times, 4 times and 0 times 512 out of 30, respectively. In the remaining cases, the difference in AIC was 513 smaller than 2. The Kullback-Leibler based selection method chose the full 514 model 19 times, 20 times, 22 times and 30 times out of 30, for $\nu_1 = 0, 1, 2, 3$ 515 respectively. As seen in Fig. G.13, when $\nu_1 = 0, 1$ or 2 estimators have a 516 high variance. Additionally, the most visited state for dormant populations 517 is the extinction state, as seen in Fig. 4. In our experiment, the dynamics 518 considered are highly similar since simulating under one of the dynamics 519 will give similar observations. However, when the hyperparameters are $\mu =$ 520 $(-3.7, 6.5), \tau = -1 \text{ and } \nu = (-3, \nu_1, 8, 2), \text{ for } \nu_1 \in \{0, 1, 2, 3\}, \text{ the full model}$ 521 was chosen 0 times, 11 times, 25 times and 22 times out of 30, respectively. 522 As can be seen the results are more precise with a higher ν_2 (influence of the 523 local non-dormant population on the dormant population). This is due to 524 the variance of the estimators being lower when $\nu_2 = 8$ than when $\nu_2 = 4$. 525

To select a model between the full model and the model with spatial 526 colonisation, we simulated 30 sets of data with hyperparameters : $\mu =$ 527 $(-3.7, 6.5), \tau = -1 \text{ and } \nu = (-3, 4, 4, \nu_3), \text{ where } \nu_3 \in \{0, 1, 2, 3\}.$ When sim-528 ulating with ν_3 fixed to 0, the species dynamics does not have spatial coloni-529 sation. For $\nu_3 = 0, 1, 2, 3$, the full model was chosen one time, three times, 530 24 times and 28 times out of 30, respectively. However, for $\nu_3 = 0, 1, 2, 3$ the 531 difference in AIC was smaller than 2 respectively 19 times, 22 times, 6 times 532 and 2 times out of 30. The Kullback-Leibler model selection chose the full 533 model respectively 7 times, 25 times, 30 times and 23 times out of 30. The 534 Kullback-Leibler method often favors the full model as opposed to the AIC 535 method. This is due to the AIC penalizing models with more parameters 536 and thus being less likely to chose a model which overfits the data. 537



Figure 7: Mean percentage of correctly recovered and predicted populations for ν_1 (influence of the previous dormant population state) varying from 0 to 6.5.

We looked at the mean percentage of correctly recovered dormant popula-539 tion states and the mean percentage of correctly predicted next non-dormant 540 population states. For a vector of hyperparameter, we simulated population 541 trajectories in ten patches over 100 time steps with $|\Omega_X| = |\Omega_Y| = 5$. We 542 then simulated the non-dormant populations in the ten patches at the 101st 543 time step, 100 times. We used the observed non-dormant population states 544 of the first 100 time steps to estimate the hyperparameters using the EM al-545 gorithm. The estimators were then used to recover the dormant populations 546 in the first 100 time steps over all the patches. Additionally, the estimators 547 were used to predict the non-dormant populations in all patches for the 101st 548 time step. The predictions were then evaluated using the 100 simulations of 549 the 101st time step. Twenty-eight different vectors of hyperparameters, for 550 which there were ten trajectories of populations, were simulated. For the first 551



Figure 8: Mean percentage of correctly recovered and predicted populations for ν_0 (influence of external colonisation) varying from -6.5 to 0.

⁵⁵² 14 vectors of hyperparameters, all hyperparameters were fixed except for the ⁵⁵³ hyperparameter associated with dormancy, which varied from 0 to 6.5. The ⁵⁵⁴ hyperparameters were : $\mu = (-3.7, 6.5), \tau = -1$ and $\nu = (-3, \nu_1, 4, 2)$. The ⁵⁵⁵ next 14 vectors of hyperparameters all had fixed values, with $\nu_1 = 4$, except ⁵⁵⁶ for ν_0 , which varied from -6.5 to 0.

557

We observed that the quality of hidden state estimation is rarely below 558 70% (see Figs. 7 and 8) and is better with high values of ν_1 (Fig. 7), i.e. 559 when dormant population survival is high. On the other hand, the quality of 560 prediction is low for high values of ν_1 . In this case, the dormant populations 561 frequently visit the maximum state. There is almost no stochasticity in the 562 dynamics of dormant populations (the variance of A is low) thus making 563 their recovery easier. On the other hand, the percentage of correct predic-564 tion dropped to 40% when ν_1 was large. The poor quality of prediction is 565

directly linked to the high variance in the non-dormant population distribu-566 tion. Since the dormant population is predominantly in the maximum state, 567 the non-dormant population distribution, conditional on the dormant popu-568 lation being in the maximum state, is a binomial distribution with parameters 569 $(|\Omega_Y|, \frac{1}{1+\exp(-(\mu_0-\mu_1(|\Omega_X|-1)/|\Omega_X|))})$ and its state distribution is as follows: the 570 non-dormant population has a 44% chance to be in the maximum state, a 571 39% chance to be in state 3, a 13% chance to be in state 2, a 2% chance to be 572 in state 1 and a 0.1% chance to be extinct. As the model predicts the most 573 probable state, it is not surprising to only have 40% of correct predictions. 574

For low ν_1 the situation is easier since the simulated dormant and nondormant populations frequently visit the extinction state. In addition, the effect of the hyperparameter ν_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 exper-581 iment was reproduced with ten time steps with 100 patches with 30 simula-582 tions. The non-dormant population at the 11th time step was then simulated 583 100 times. The exact values of the hyperparameters were : $\mu = (-3.7, 6.5)$, 584 $\tau = -1$ and $\nu = (-3, 3, 4, 2)$. We obtained 80% of correct prediction of 585 the non-dormant population states and 68.9% of correct recovery of dormant 586 population states. This is slightly less that when estimating from 10 patches 587 and 100 time steps, maybe because estimation is easier from a small number 588 of patched but long trajectories of observations than for the opposite. 589

⁵⁹⁰ 7. Discussion

We propose the MHMM-DF framework that can describe, from abun-591 dance classes of non-dormant populations, the local and regional dynamics 592 of annual plants that can undergo a potentially infinite dormancy. In a 593 non-parametric approach, the MHMM-DF can be defined using only three 594 probabilities: the initial probability, the transition probability from past pop-595 ulations to current dormant populations and the emission probability from 596 past dormant populations to current non-dormant populations. For the sake 597 of sparsity, we have proposed to model these probabilities as Binomial dis-598 tributions with the probability parameter modelled as a logistic function de-599 pendent on hyperparameters. Each hyperparameter is associated with a spe-600 cific process in the species dynamics. They describe the colonisation process 601

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 asso-606 ciated with one hyperparameter and is described as the indistinguishable 607 influence of neighbour non-dormant population states on a dormant popu-608 lation state in a patch. Several options exist to aggregate the neighbour 609 states into a single variable and we proposed the mean colonisation. How-610 ever, this choice has drawbacks. The mean colonisation can assign the same 611 value in Ω_Y for very contrasted configurations of neighbours states. Addi-612 tionally, when the number of patches increases, it increases the tendency for 613 the mean colonisation process to be identical for every patch. To circumvent 614 the first limit, we proposed the alphabetic colonisation (detailed in Appendix 615 I), which only assigns the same aggregated value to two vectors of neighbour 616 states if both vectors are equal though patch permutation. The influence of 617 the neighbour patches are still indistinguishable. However contrasted distri-618 butions of the abundance classes of the neighbourhood will lead to different 619 aggregated values. In practice, to create an ordinal variable from the vector 620 of neighbours, we ranked the distributions of the abundance classes according 621 to alphabetic (or lexicographic) order. We demonstrated that the parametric 622 MHMM-DF with mean or alphabetic colonisation is generically identifiable 623 for a least three patches. 624

Different colonisation processes can be implemented preferably if identifi-625 ability is verified. Instead of considering one hyperparameter for the spatial 626 colonisation process, the colonisation process could include a hyperparam-627 eter per number of non-dormant populations in each state as modelled in 628 Gyllenberg and Hanski (1997). Each hyperparameter associated with the 629 colonisation process would measure the influence of the number of neigh-630 bour non-dormant populations in a specific state. However, establishing the 631 model's identifiability is not straightforward, and moreover the model would 632 require more data for estimation. 633

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 640 modelling the awakening process. One option is to set the probability of ex-641 terior colonisation of non-dormant populations to a very low value instead of 642 estimating it. Alternatively, we could model the awakening probability as a 643 zero-inflated Binomial distribution. The zero-inflated Binomial distribution 644 would force the non-dormant population to be extinct when the dormant one 645 is extinct, and in the other situations, the awakening process would still be 646 a Binomial distribution with logit regression. This is obtained by setting the 647 mixture parameter of this zero-inflated model to 1 in the case of absence of 648 a dormant population and to 0 otherwise. The generic identifiability is still 649 satisfied for this model. 650

The MHMM-DF framework is built under biological assumptions, where 651 the dependencies between the species populations represent known biological 652 processes. These biological processes are often modelled using precise infor-653 mation about the number of individuals (David et al., 2010). Even though, 654 this type of modelling is rewarding, since predictions are more accurate, they 655 are often computationally intensive when dealing with hidden dynamics and 656 may have unidentifiable parameters. Modelling with abundance classes helps 657 reduce the computational complexity. The parametric MHMM-DF is used 658 to infer the input of each process on the species dynamics by learning statis-659 tical relationships, as opposed to mechanistic approaches that would require 660 precise knowledge beforehand about mechanisms and environmental factors 661 that influence the species dynamics. 662

The MHMM-DF relies on two main assumptions about the species dy-663 namics. The model assumes dormant populations to be sessile and assumes 664 that non-dormant populations do not influence one another directly. The 665 first assumption implies that dormant populations can not travel between 666 patches, which reduces the time and space complexity of the E step from 667 exponential to linear in terms of patches, $O(|\Omega_X|^2 NC)$, where C is the num-668 ber of patches, N is the number of time steps, and $|\Omega_X|$ is the number of 669 states of a dormant population. Reducing the computational complexity of 670 the E step implies that dealing with a large number of patches is not an 671 issue with a MHMM-DF. The second assumption implies that once a non-672 dormant population is observed, the non-dormant population will be renewed 673 in the next time step. These two assumptions are verified for annual plants. 674 Even though the MHMM-DF was created to model the dynamics of plants 675 in patches, the model is general and could be applied to different organisms. 676 In specific ecosystems, the assumption of seed immobility in the seed bank 677

is questionable. If it is relevant for most natural populations, where seeds 678 are only mobile during dispersal, seeds in the seed bank of riparian ecosys-679 tems may move because of floods. Furthermore, in agricultural systems, 680 tools or machines may transfer seeds from one seed bank to another. The 681 second assumption, which assumes no interaction between past and current 682 non-dormant populations, can easily be removed without hindering the com-683 plexity of the E step. Removing the assumption enables us to model the local 684 non-dormant population's survival between two time steps and the migration 685 of non-dormant populations between patches. The associated parametric 686 MHMM-DF with both survival and migration of non-dormant populations 687 would have nine hyperparameters and is identifiable. Adding the survival 688 process of non-dormant populations is relevant for plants since it would en-689 able us to study the dynamics of perennial plants. Even though, including 690 the migration process is not relevant to plant dynamics, a model with migra-691 tion could be applied to many more species such as fresh water snails whose 692 dormancy process would be seen as aestivation during dry seasons (Lamy 693 et al., 2013). 694

MHMM-DF can have multiple uses such as estimating a species' dynam-695 ics via hyperparameters, discriminating between two different types of dy-696 namics, predicting the state of the next non-dormant populations and even 697 recovering all past states of dormant populations. The results show that the 698 estimators, obtained using the EM algorithm, have a low variance and low 699 bias when the species' populations are not predominantly visiting either the 700 extinction state or the maximum state. On the other hand, distinguishing 701 the input resulting from each process is difficult when dormant populations 702 frequently visit an extreme state since many different vectors of hyperparam-703 eters have a similar likelihood. When only observing the non-dormant popu-704 lations, the predictions and hidden state estimations obtained were globally 705 of good quality. When a population is inferred to be or to become extinct, 706 errors in predictions and hidden state estimations seem to be dependent on 707 exogenous colonisation. Additionally, errors in predictions and hidden state 708 estimations depend on the variance of the Binomial distribution. The model 709 selection technique is able to discriminate between dynamics with dormancy 710 and dynamics without dormancy, or even dynamics with colonisation and 711 dynamics without colonisation. In order to fully validate the parameterisa-712 tion, the model should now be tested on real data. For instance, studying the 713 dynamics of weed species with the model could be a subsequent extension. 714 Since multiple exterior factors can influence species dynamics, such as 715

temperature for plant germination (Taab, 2009; Seglias et al., 2018; El-716 Keblawy et al., 2017; Ueno, 2002; Gómez-González et al., 2018; Zhang et al., 717 2017), the logit function, associated with the awakening process, can include 718 covariables to increase the model's realism. For weed species in agricultural 719 fields, the species dynamics could be made to depend on crop type (Borgy 720 et al., 2015). It would be also relevant to model different sizes or types of 721 neighbourhood, since the distance of seed dispersal or even the paths vary be-722 tween species (Nathan, 2000). Indeed colonisation may be related to specific 723 dissemination agents, such as rivers for riparian ecosystems or agricultural 724 machines for weed species. In the latter case, dispersal is not defined by a 725 distance from the source but by the path formed by the successive patches in 726 which the agricultural machines travelled before entering the current patch. 727 In practice, the size or the paths for dispersal are not known. Using model 728 selection and several MHMM-DF models associated to different neighbour-720 hoods, it would be possible to determine which model is best suited for a 730 specific species. 731

The framework presented is the first step towards modelling the local and regional dynamics of species with an undetectable life form. The MHMM-DF 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.

739 8. Acknowledgements

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

- ⁷⁴² Akaike, H. (1981). Likelihood of a model and information criteria. *Journal* ⁷⁴³ 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.
- ⁷⁵² Borgy, B., X. Reboud, N. Peyrard, R. Sabbadin, and S. Gaba (2015). Dy⁷⁵³ namics of weeds in the soil seed bank : A hidden Markov model to estimate
 ⁷⁵⁴ 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: un derstanding AIC and BIC in model selection. Sociological methods & re search 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.
- ⁷⁶⁹ David, O., A. Garnier, C. Larédo, and J. Lecomte (2010). Estimation of ⁷⁷⁰ plant demographic parameters from stage-structured censuses. *Biomet-*⁷⁷¹ $rics \ 66(3), 875-882.$
- ⁷⁷² Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from ⁷⁷³ incomplete data via the EM algorithm. *Journal of the Royal Statistical* ⁷⁷⁴ 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.
- ⁷⁸² 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. *Jour- nal 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.
- ⁷⁹³ Gotelli, N. J. (1991). Metapopulation models: The rescue effect, the ⁷⁹⁴ propagule rain, and the core-satellite hypothesis. *The American Natu-*⁷⁹⁵ *ralist* 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*, 67–73.
- 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.
- ⁸⁵² Ueno, K. (2002). Effects of desiccation and a change in temperature on ⁸⁵³ germination of immature grains of wheat (triticum aestivum l.). *Euphyt*-⁸⁵⁴ *ica* 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.

862 Appendices

⁸⁶³ 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 θ such that $\mathbb{P}(Y^N = y^N | \theta) = \mathbb{P}(Y^N = y^N | \theta')$ 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 2L + 1 consecutive variables provided L satisfies



Figure A.11: Coupled Hidden Markov Model.

$$\binom{L+s-1}{s-1} \ge r.$$

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, $H_n = (X_n^C, Y_{n+1}^C)$, and by setting $O_n = Y_{n+1}^C$. Note that the observed variables are duplicated. We will denote ϕ^{hmm} and A^{hmm} as respectively the emission and the transition probabilities of the HMM $(H_n, O_n)_{n \in \{0, \dots, N-1\}}$. ϕ^{hmm} is deterministic, and A^{hmm} depends on the probabilities ϕ and A of the MHMM-DF.

⁸⁸³ When applying Theorem 1 to the HMM defined by (H_n, O_n) we obtain ⁸⁸⁴ the following theorem

Theorem 2. The parameters (ϕ^{hmm}, A^{hmm}) of the HMM model corresponding to the conversion of a MHMM-DF are generically identifiable from seven consecutive observations, if $|\Omega_X| \leq |\Omega_Y|$, and if C > 2.

888 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 $r = |\Omega_X|^C |\Omega_Y|^C$, the number of possible observable states is $s = |\Omega_Y|^C$ and L = 3. We must show that the following inequality is satisfied:

$$\begin{pmatrix} |\Omega_Y|^C + 2\\ |\Omega_Y|^C - 1 \end{pmatrix} \ge |\Omega_X|^C |\Omega_Y|^C.$$

⁸⁹² The inequality can be reformulated as:

$$(|\Omega_Y|^C + 2)(|\Omega_Y|^C + 1) \ge |\Omega_X|^C.$$

Since we assume that $|\Omega_X| \leq |\Omega_Y|$, the inequality is true.

894 895

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

In the HMM representation of a MHMM-DF, the emission matrix ϕ_{hmm} is deterministic and is independent of μ, ν . Therefore, we need to show that $A^{hmm}_{\mu,\nu} = A^{hmm}_{\mu',\nu'}$ implies $(\mu, \nu) = (\mu', \nu')$. We recall that $A^{hmm}_{\mu,\nu}(h_n, h_{n-1}) = \mathbb{P}(h_n|h_{n-1}, \nu, \mu)$. The HMM's transition probability is:

$$\begin{split} \mathbb{P}(h_{n}|h_{n-1},\nu,\mu) &= \mathbb{P}(x_{n}^{C},y_{n+1}^{C}|x_{n-1}^{C},y_{n}^{C}) \\ &= \mathbb{P}(y_{n+1}^{C}|x_{n}^{C})\mathbb{P}(x_{n}^{C}|x_{n-1}^{C},y_{n}^{C}) \\ &= \prod_{c=1}^{C} \mathbb{P}(y_{c,n+1}|x_{c,n})\mathbb{P}(x_{c,n}|x_{c,n-1},y_{n}^{C}) \\ &= \prod_{c=1}^{C} \left(|\Omega_{Y}| - 1 \right) \\ &\times \left[\frac{1}{1 + \exp(-(\mu_{0} + \mu_{1}\frac{x_{c,n}}{|\Omega_{X}|}))} \right]^{|\Omega_{Y}|-1} \left[\exp(-(\mu_{0} + \mu_{1}\frac{x_{c,n}}{|\Omega_{X}|})) \right]^{|\Omega_{Y}|-y_{c,n+1}-1} \\ &\times \left(|\Omega_{X}| - 1 \right) \left[\frac{1}{1 + \exp(-(\nu_{0} + \nu_{1}\frac{x_{c,n-1}}{|\Omega_{X}|} + \nu_{2}\frac{y_{c,n}}{|\Omega_{Y}|} + \nu_{3}(\frac{f_{\text{mean}}(y_{n}^{C})}{|f_{\text{mean}}(\Omega_{Y}^{C-1})|}))} \right]^{|\Omega_{X}|-x_{c,n}-1} \\ &\times \left[\exp(-(\nu_{0} + \nu_{1}\frac{x_{c,n-1}}{|\Omega_{X}|} + \nu_{2}\frac{y_{c,n}}{|\Omega_{Y}|} + \nu_{3}(\frac{f_{\text{mean}}(y_{n}^{C})}{|f_{\text{mean}}(\Omega_{Y}^{C-1})|})) \right]^{|\Omega_{X}|-x_{c,n}-1} \end{split}$$

Theorem 3. The hyperparameter $(\mu, \nu) = (\mu_0, \mu_1, \nu_0, \nu_1, \nu_2, \nu_3)$ of a MHMM-DF with the mean colonisation process is generically identifiable from seven consecutive observations if the following conditions hold: if $|\Omega_X| \leq |\Omega_Y|$, $|\Omega_Y| > 1$ and C > 2.

910 Proof.

As explained above, we need to find conditions under which $\mathbb{P}(h_n|h_{n-1},\nu,\mu) = \mathbb{P}(h_n|h_{n-1},\nu',\mu')$ implies $(\mu,\nu) = (\mu',\nu')$. Let us show one-by-one the equality for each hyperparameter. We start with μ_0 . Let us define L1 as the transition probability from $(x_{n-1}^C, y_n^C) = (\begin{pmatrix} 0\\ \dots\\ 0 \end{pmatrix}, \begin{pmatrix} 0\\ \dots\\ 0 \end{pmatrix})$ to $(x_n^C, y_{n+1}^C) =$

916

We now define L2 as the transition probability from $(x_{n-1}^C, y_n^C) = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}$

⁹¹⁷ to
$$(x_n^C, y_{n+1}^C) = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} |\Omega_Y| - 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \end{pmatrix}$$
:

$$L2(\mu_{0},\nu_{0}) = \binom{|\Omega_{Y}|-1}{0}^{C-1} \binom{|\Omega_{Y}|-1}{|\Omega_{Y}|-1} \left[\frac{1}{1+\exp(-\mu_{0})}\right]^{(|\Omega_{Y}|-1)C} \left[\exp(-\mu_{0})\right]^{(|\Omega_{Y}|-1)(C-1)} \\ \times \binom{|\Omega_{X}|-1}{0}^{C} \left[\frac{1}{1+\exp(-\nu_{0})}\right]^{(|\Omega_{X}|-1)C} \left[\exp(-\nu_{0})\right]^{(|\Omega_{X}|-1)C}.$$

If $L1(\mu_0, \nu_0) = L1(\mu'_0, \nu'_0)$ and $L2(\mu_0, \nu_0) = L2(\mu'_0, \nu'_0)$ then $L1(\mu_0, \nu_0)/L2(\mu_0, \nu_0) = L1(\mu'_0, \nu'_0)/L2(\mu'_0, \nu'_0)$, which implies $e^{-(|\Omega_Y|-1)\mu_0} = e^{-(|\Omega_Y|-1)\mu'_0}$ Thus $\mu'_0 = \mu_0$. Since $\mu'_0 = \mu_0$, we can now identify ν_0 using L1:

$$L1(\mu_{0},\nu_{0}) = L1(\mu_{0},\nu'_{0})$$

$$<=>$$

$$\left(\left(\frac{1}{1+\exp(-\nu_{0})}\right)\exp(-\nu_{0})\right)^{(|\Omega_{X}|-1)C} = \left(\left(\frac{1}{1+\exp(-\nu'_{0})}\right)\exp(-\nu'_{0})\right)^{(|\Omega_{X}|-1)C}$$

$$<=>$$

$$(1+\exp(-\nu_{0}))\exp(\nu_{0}) = (1+\exp(-\nu'_{0}))\exp(\nu'_{0})$$

$$<=>$$

$$\nu_{0} = \nu'_{0}.$$

Let us now show the identifiability of μ_1 . We consider L3, the expression defined by the transition probability from $(x_{n-1}^C, y_n^C) = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}$) to

 ${}_{923} (x_n^C, y_{n+1}^C) = \begin{pmatrix} |\Omega_X| - 1 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix} \end{pmatrix}.$ Note that from now on, in the expression used to derive identificability, we exit the terms that are known to be

sions 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

$$L3(\nu_{0},\mu_{0},\mu_{1}) \propto \left[\frac{1}{1+\exp(-\mu_{0})}\right]^{(|\Omega_{Y}|-1)(C-1)} \left[\exp(-\mu_{0})\right]^{(|\Omega_{Y}|-1)(C-1)} \\ \times \left[\frac{1}{1+\exp(-(\mu_{0}+\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}))}\right]^{(|\Omega_{Y}|-1)} \\ \times \left[\exp(-(\mu_{0}+\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}))\right]^{(|\Omega_{Y}|-1)} \\ \times \left[\frac{1}{1+\exp(-\nu_{0})}\right]^{(|\Omega_{X}|-1)C} \left[\exp(-(\nu_{0}))\right]^{(|\Omega_{X}|-1)(C-1)}.$$

927 Thus

$$\begin{array}{rcl} L3(\nu_{0},\mu_{0},\nu_{1}') &=& L3(\nu_{0},\mu_{0},\nu_{1}) \\ &<=>\\ \left[\frac{1}{1+\exp(-(\mu_{0}+\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}'))}\right] \left[\exp(-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}')\right] &=& \left[\frac{1}{1+\exp(-(\mu_{0}+\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}))}\right] \left[\exp(-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1})\right] \\ &<=>\\ \exp(-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1})+\\ \exp(-\mu_{0}-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1}') &=& \exp(-\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}\mu_{1})+\\ &<=>\\ \mu_{1} &=& \mu_{1}'. \end{array}$$

We establish the identifiability of
$$\nu_{1}$$
 by using $L4$, the transition probabil-
ity from $(x_{n-1}^{C}, y_{n}^{C}) = (\begin{pmatrix} |\Omega_{X}| - 1 \\ ... \\ |\Omega_{X}| - 1 \end{pmatrix}, \begin{pmatrix} 0 \\ ... \\ 0 \end{pmatrix})$ to $(x_{n}^{C}, y_{n+1}^{C}) = (\begin{pmatrix} 0 \\ ... \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ ... \\ 0 \end{pmatrix})$.
 $L4(\mu_{0}, \nu_{0}, \nu_{1}) \propto \left[\frac{1}{1 + \exp(-\mu_{0})}\right]^{(|\Omega_{Y}| - 1)C} [\exp(-\mu_{0})]^{(|\Omega_{Y}| - 1)C}$
 $\times \left[\frac{1}{1 + \exp(-(\nu_{0} + \nu_{1}\frac{(|\Omega_{X}| - 1)}{|\Omega_{X}|}))}\right]^{(|\Omega_{X}| - 1)C}$
 $\times \left[\exp(-(\nu_{0} + \nu_{1}\frac{(|\Omega_{X}| - 1)}{|\Omega_{X}|}))\right]^{(|\Omega_{X}| - 1)C}$.

930

Thus :

$$\begin{array}{rcl} L4(\mu_{0},\nu_{0},\nu_{1}^{'}) & = & L4(\mu_{0},\nu_{0},\nu_{1}) \\ & <=> \\ \left[\frac{1}{1+\exp(-(\nu_{0}+\nu_{1}^{'}\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}))}\right] \left[\exp(-\nu_{1}^{'}\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|})\right] & = & \left[\frac{1}{1+\exp(-(\nu_{0}+\nu_{1}\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|}))}\right] \left[\exp(-\nu_{1}\frac{(|\Omega_{X}|-1)}{|\Omega_{X}|})\right] \\ & <=> \\ & \nu_{1} & = & \nu_{1}^{'}. \end{array}$$

We have established identifiability for all hyperparameters except ν_2 and ν_3 . When C = 2 it can be seen that $f_{\text{mean}}(Y_n^{C \setminus c}) = Y_{c',n}$ with $c \neq c'$. By using the following states defined below one can not show identifiability of ν_2 and ν_3 . Let us assume C > 2. For these two hyperparameters, we consider the

⁹³⁵ following two transition probabilities: L5 from $(x_{n-1}^C, y_n^C) = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}$

⁹³⁶ to
$$(x_n^C, y_{n+1}^C) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}$$
) and L6 from $(x_{n-1}^C, y_n^C) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix}$) to
⁹³⁷ $(x_n^C, y_{n+1}^C) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}$).

$$L5(\nu_0,\nu_2) \propto \left[\frac{1}{1+\exp(-(\nu_0+\frac{1}{|\Omega_Y|}\nu_2))}\right]^{(|\Omega_X|-1)} \left[\frac{1}{1+\exp(-\nu_0)}\right]^{(|\Omega_X|-1)(C-1)} \times \left[\exp(-(\nu_0+\frac{1}{|\Omega_Y|}\nu_2))\right]^{(|\Omega_X|-1)} \left[\exp(-\nu_0)\right]^{(|\Omega_X|-1)(C-1)}.$$

938 Therefore:

$$L5(\nu_0,\nu'_2) = L5(\nu_0,\nu_2)$$

$$<=>$$

$$\exp(\frac{1}{|\Omega_Y|}\nu'_2) = \exp(\frac{1}{|\Omega_Y|}\nu_2)$$

$$<=>$$

$$\nu'_2 = \nu_2.$$

$$L6(\nu_{0},\nu_{2},\nu_{3}) \propto \left[\frac{1}{1+\exp(-(\nu_{0}+\frac{1}{|\Omega_{Y}|}\nu_{2}+\frac{1}{|\Omega_{Y}|}\nu_{3}))}\right]^{(|\Omega_{X}|-1)C} \times \left[\exp(-(\nu_{0}+\frac{1}{|\Omega_{Y}|}\nu_{2}+\frac{1}{|\Omega_{Y}|}\nu_{3}))\right]^{(|\Omega_{X}|-1)C}$$

939 Therefore

$$L6(\nu_{0}, \nu_{2}, \nu_{3}') = L6(\nu_{0}, \nu_{2}, \nu_{3})$$

$$< =>$$

$$\exp(\frac{1}{|\Omega_{Y}|}\nu_{3}') = \exp(\frac{1}{|\Omega_{Y}|}\nu_{3})$$

$$< =>$$

$$\nu_{3}' = \nu_{3}.$$

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.

945 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 $\rho_{c,n}(x_{c,n})$ and $\xi_{c,n}(x_{c,n-1}, x_{c,n})$ defined as

$$\begin{array}{lll} \xi_{c,n}(x_{n-1},x_n) &=& \mathbb{P}(X_{c,n}=x_n,X_{c,n-1}=x_{n-1}|Y^{C,N}=y^{C,N},\lambda_t),\\ \rho_{c,n}(x_n) &=& \mathbb{P}(X_{c,n}=x_n|Y^{C,N}=y^{C,N},\lambda_t) \end{array}$$

where $\lambda_t = (\pi_t, \phi_t, A_t)$ is the current estimator. We describe here how these probabilities can be computed from auxiliary quantities defined recursively and independently for each chain.

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

Let us define $\alpha_{c,n}(x_{c,n}) = \mathbb{P}(y^{C,n}, x_{c,n}|\lambda_t)$ and $\beta_{c,n}(x_{c,n}) = \mathbb{P}(y_{n+1}^C, ..., y_N^C|y^{C,n}, x_{c,n}, \lambda_t)$. For a given chain c, the $\alpha_{c,n}$ can be computed recursively from $\alpha_{c,n-1}$ (independently of $\alpha_{c',n}$ for $c' \neq c$):

$$\begin{aligned} \alpha_{c,n}(x_{c,n}) &= & \mathbb{P}(y^{C,n}, X_{c,n} = x_{c,n} | \lambda_t) \\ &= & \sum_{x_{c,n-1} \in \Omega_X} \mathbb{P}(y^{C,n}, X_{c,n} = x_{c,n}, X_{c,n-1} = x_{c,n-1} | \lambda_t) \\ &= & \sum_{x_{c,n-1} \in \Omega_X} \mathbb{P}(y^{C,n-1}, X_{c,n-1} = x_{c,n-1} | \lambda_t) \\ &= & \sum_{x_{c,n-1} \in \Omega_X} \mathbb{P}(y_n^C, X_{c,n} = x_{c,n} | y^{C,n-1}, x_{c,n-1}, \lambda_t) \\ &= & \sum_{x_{c,n-1} \in \Omega_X} \alpha_{c,n-1}(x_{c,n-1}) \mathbb{P}(y_n^C | y^{C,n-1}, x_{c,n-1}, \lambda_t) A_t(x_{c,n-1}, x_{c,n}, y_n^C) \\ &= & \sum_{x_{c,n-1} \in \Omega_X} \alpha_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}) A_t(x_{c,n-1}, x_{c,n}, y_n^C) \\ &\quad \times \mathbb{P}(y_n^{C \setminus c} | y^{C,n-1}, y_{c,n}, x_{c,n-1}, \lambda_t). \end{aligned}$$

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

$$\alpha_{c,n}(x_{c,n}) \propto \sum_{x_{c,n-1} \in \Omega_X} \alpha_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}) A_t(x_{c,n-1}, x_{c,n}, y_n^C).$$

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

$$\tilde{\alpha}_{c,n}(x_{c,n}) = \sum_{x_{c,n-1} \in \Omega_X} \tilde{\alpha}_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) A_t(x_{c,n-1}, x_{c,n}, y_n^C)$$

961 with $\tilde{\alpha}_{c,0}(x_{c,0}) = \alpha_{c,0}(x_{c,0})$.

Similarly, there is a recursive expression of $\beta_{c,n}$ involving $\beta_{c,n-1}$:

$$\begin{split} \beta_{c,n}(x_{c,n}) &= & \mathbb{P}(y_{n+1}^C, ..., y_N^C | y^{C,n}, x_{c,n}, \lambda_t) \\ &= & \sum_{x_{c,n+1} \in \Omega_X} \mathbb{P}(y_{n+1}^C, ..., y_N^C, x_{c,n+1} | y^{C,n}, x_{c,n}, \lambda_t) \\ &= & \sum_{x_{c,n+1} \in \Omega_X} \mathbb{P}(y_{n+2}^C, ..., y_N^C, x_{c,n+1} | y^{C,n+1}, x_{c,n}, \lambda_t) \mathbb{P}(y_{n+1}^C | y^{C,n}, x_{c,n}, \lambda_t) \\ &= & \sum_{x_{c,n+1} \in \Omega_X} \mathbb{P}(y_{n+2}^C, ..., y_N^C | y^{C,n+1}, x_{c,n}, x_{c,n+1}, \lambda_t) A_t(x_{c,n}, x_{c,n+1}, y_{n+1}^C) \\ &= & \times \phi_t(x_{c,n}, y_{c,n+1}) \mathbb{P}(y_{n+1}^{C,c} | y^{C,n}, y_{c,n+1}, x_{c,n}, \lambda_t) \\ &= & K_{c,n+1} \sum_{x_{c,n+1} \in \Omega_X} \mathbb{P}(y_{n+2}^C, ..., Y_N^C | y^{C,n+1}, x_{c,n}, x_{c,n+1}, \lambda_t) \\ &\times A_t(x_{c,n}, x_{c,n+1}, y_{n+1}^C) \phi(x_{c,n}, y_{c,n+1}). \end{split}$$

Since the vector $(Y_{n+2}^C, ..., Y_N^C)$ is independent of $X_{c,n}$ given $y^{C,n+1}, x_{c,n+1}$, we have:

$$\beta_{c,n}(x_{c,n}) \propto \sum_{x_{c,n+1} \in \Omega_X} \beta_{c,n+1}(x_{c,n+1}) A_t(x_{c,n}, x_{c,n+1}, Y_{n+1}^C) \phi_t(x_{c,n}, y_{c,n+1}).$$

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

$$\tilde{\beta}_{c,n}(x_{c,n}) = \sum_{x_{c,n+1} \in \Omega_X} \tilde{\beta}_{c,n+1}(x_{c,n+1}) A_t(x_{c,n}, x_{c,n+1}, y_{n+1}^C) \phi_t(x_{c,n}, y_{c,n+1})$$

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

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

$$\rho_{c,n}(x_{c,n}) = \frac{\mathbb{P}(X_{c,n} = x_{c,n}, y^{C,N} | \lambda_t)}{\mathbb{P}(y^{C,N} | \lambda_t)} = \frac{\beta_{c,n}(x_{c,n})\alpha_{c,n}(x_{c,n})}{\sum_{x \in \Omega_X} \beta_{c,n}(x)\alpha_{c,n}(x)} = \frac{\tilde{\beta}_{c,n}(x_{c,n})\tilde{\alpha}_{c,n}(x_{c,n})}{\sum_{x \in \Omega_X} \tilde{\beta}_{c,n}(x)\tilde{\alpha}_{c,n}(x)},$$
$$\xi_{c,n}(x_{c,n-1}, x_{c,n}) = \frac{\mathbb{P}(X_{c,n} = x_{c,n}, X_{c,n-1} = x_{c,n-1}, y^{C,N} | \lambda_t)}{\mathbb{P}(y^{C,N} | \lambda_t)}.$$

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

$$\begin{aligned} \mathbb{P}(x_{c,n}, x_{c,n-1}, y^{C,N}) &= \mathbb{P}(X_{c,n} = x_{c,n}, X_{c,n-1} = x_{c,n-1}, y_n^C | \lambda_t) \\ &\times \mathbb{P}(\{y_l^C\}_{l \neq n} | x_{c,n}, x_{c,n-1}, y_n^C, \lambda_t) \\ &= A_t(x_{c,n-1}, x_{c,n}, y_n^C) \mathbb{P}(X_{c,n-1} = x_{c,n-1}, y_n^C | \lambda_t) \\ &\times \mathbb{P}(y^{C,n-1} | x_{c,n}, x_{c,n-1}, y_n^C, \lambda_t) \\ &= A_t(x_{c,n-1}, x_{c,n}, y_n^C) \mathbb{P}(X_{c,n-1} = x_{c,n-1}, y_n^C | \lambda_t) \\ &\times \mathbb{P}(y^{C,n-1} | x_{c,n-1}, y_n^C, \lambda_t) \beta_{c,n}(x_{c,n}) \\ &= A_t(x_{c,n-1}, x_{c,n}, y_n^C) \beta_{c,n}(x_{c,n}) \mathbb{P}(X_{c,n-1} = x_{c,n-1}, y^{C,n} | \lambda_t) \\ &= A_t(x_{c,n-1}, x_{c,n}, y_n^C) \beta_{c,n}(x_{c,n}) \mathbb{P}(X_{c,n-1} = x_{c,n-1}, y^{C,n-1} | \lambda_t) \\ &\times \mathbb{P}(y_n^C | x_{c,n-1}, y^{C,n-1}, \lambda_t) \\ &= A_t(x_{c,n-1}, x_{c,n}, y_n^C) \beta_{c,n}(x_{c,n}) \alpha_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}) K_{c,n} \\ &\propto A_t(x_{c,n-1}, x_{c,n}, Y_n^C) \tilde{\beta}_{c,n}(x_{c,n}) \tilde{\alpha}_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n}). \end{aligned}$$

972 This leads to

$$\xi_{c,n}(x_{c,n-1}, x_{c,n}) = \frac{A_t(x_{c,n-1}, x_{c,n}, y_n^C) \hat{\beta}_{c,n}(x_{c,n}) \tilde{\alpha}_{c,n-1}(x_{c,n-1}) \phi_t(x_{c,n-1}, y_{c,n})}{\sum_{(x,x') \in \Omega_X^2} A_t(x', x, y_n^C) \tilde{\beta}_{c,n}(x) \tilde{\alpha}_{c,n-1}(x') \phi_t(x', y_{c,n})}.$$

973 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. ⁹⁷⁶ However, there are situations where we are interested in computing the like-⁹⁷⁷ lihood, such as for model selection. We derive here the likelihood expression ⁹⁷⁸ in terms of $\tilde{\alpha}_{c,n}$ and $\tilde{\beta}_{c,n}$.

$$\mathbb{P}(y^{C,N}) = \sum_{\substack{x_{c,n} \in \Omega_X \\ x_{c,n} \in \Omega_X}} \mathbb{P}(X_{c,n} = x_{c,n}, y^{C,N}) \\
= \sum_{\substack{x_{c,n} \in \Omega_X \\ x_{c,n} \in \Omega_X}} [\prod_{j=1}^n K_{c,j}] \tilde{\alpha}_{c,n}(x_{c,n}) [\prod_{j'=n+1}^N K_{c,j'}] \tilde{\beta}_{c,n}(x_{c,n}) \\
= [\prod_{j=1}^N K_{c,j}] \sum_{\substack{x_{c,n} \in \Omega_X \\ x_{c,n} \in \Omega_X}} \tilde{\alpha}_{c,n}(x_{c,n}) \tilde{\beta}_{c,n}(x_{c,n}).$$

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

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:

$$\prod_{j=1}^{N} K_{c,j} = \prod_{l=1}^{N} \prod_{l \in \{1, \dots, C\} \setminus \{c\}} \sum_{x_{l,j-1} \in \Omega_X} \phi(x_{l,j-1}, y_{l,j}) \frac{\tilde{\alpha}_{l,j-1}(x_{l,j-1})}{\sum_{x \in \Omega_X} \tilde{\alpha}_{l,j-1}(x)}.$$

986 Finally

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

987 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 $\mathbb{P}(X^{C,N} = x^{C,N} | Y^{C,N} = y^{C,N})$ which is equivalent to maximising $\mathbb{P}(X^{C,N} = x^{C,N}, Y^{C,N} = y^{C,N})$. The latter can be factored over the *C* chains as follows:

$$\begin{split} \mathbb{P}(X^{C,N} = x^{C,N}, Y^{C,N} = y^{C,N}) &= \mathbb{P}(X^{C,N} = x^{C,N} | Y^{C,N} = y^{C,N}) \times \mathbb{P}(Y^{C,N} = y^{C,N}) \\ &= \mathbb{P}(Y^{C,N} = y^{C,N}) \prod_{c=1}^{C} p(X_c^N = x_c^N | Y^{C,N} = y^{C,N}) \\ &= \mathbb{P}(Y^{C,N} = y^{C,N}) \prod_{c=1}^{C} \frac{p(X_c^N = x_c^N, Y^{C,N} = y^{C,N})}{\mathbb{P}(Y^{C,N} = y^{C,N})} \\ &= \frac{1}{(\mathbb{P}(Y^{C,N} = y^{C,N}))^{C-1}} \prod_{c=1}^{C} p(X_c^N = x_c^N, Y^{C,N} = y^{C,N}) \\ &\propto \prod_{c=1}^{C} \mathbb{P}(X_c^N = x_c^N, Y^{C,N} = y^{C,N}). \end{split}$$

⁹⁹³ This implies that:

$$\max_{x^{C,N} \in \Omega_X^{C \times N}} \mathbb{P}(X^{C,N} = x^{C,N}, Y^{C,N} = y^{C,N}) = \max_{x^{C,N} \in \Omega_X^{C \times N}} \prod_{c=1}^C \mathbb{P}(X_c^N = x_c^N, Y^{C,N} = y^{C,N})$$
$$= \prod_{c=1}^C \max_{x_c^N \in \Omega_X^N} \mathbb{P}(X_c^N = x_c^N, Y^{C,N} = y^{C,N}).$$

⁹⁹⁴ Thus we can perform recovery separately per patch. Let us note $\delta_{c,n}(x_{c,n}) = \max_{\substack{x_c^{n-1} \in \Omega_X^{n-1}}} \mathbb{P}(X_c^n = x_c^n, Y^{C,n} = y^{C,n})$. Here is the recursive definition of $\delta_{c,n}$ in ⁹⁹⁶ terms of $\delta_{c,n-1}$:

$$\begin{split} \delta_{c,n}(x_{c,n}) &= \max_{x_c^{n-1} \in \Omega_X^{n-1}} \mathbb{P}(x_c^{n-1}, y^{C,n-1}) \mathbb{P}(x_{c,n}, y_n^C | x_c^{n-1}, y^{C,n-1}) \\ &= \max_{x_{c,n-1} \in \Omega_X} \delta_{c,n-1}(x_{c,n-1}) \mathbb{P}(x_{c,n}, y_n^C | x_{c,n-1}, y^{C,n-1}) \\ &= \max_{x_{c,n-1} \in \Omega_X} \delta_{c,n-1}(x_{c,n-1}) \mathbb{P}(y_n^C | x_{c,n-1}, y^{C,n-1}) A(x_{c,n-1}, x_{c,n}, y_n^C) \\ &= \max_{x_{c,n-1} \in \Omega_X} \delta_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) \mathbb{P}(y_n^{C \setminus c} | x_{c,n-1}, y^{C,n-1}) A(x_{c,n-1}, x_{c,n}, y_n^C) \\ &= \max_{x_{c,n-1} \in \Omega_X} \delta_{c,n-1}(x_{c,n-1}) \phi(x_{c,n-1}, y_{c,n}) \mathbb{P}(y_n^{C \setminus c} | x_{c,n-1}, x_{c,n}, y_n^C). \end{split}$$

⁹⁹⁷ F. Prediction of the next non-dormant population state for MHMM ⁹⁹⁸ DF

⁹⁹⁹ The prediction of a species population state in patch c at time N + 1, ¹⁰⁰⁰ given the sequences of observations of the non-dormant states in all patches ¹⁰⁰¹ up to time N, is defined as:

$$\hat{y}_{c,N+1} = \underset{\substack{y_{c,N+1} \in \Omega_{Y} \\ y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{Y}}}{\operatorname{argmax}} \sum_{\substack{y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{X}}} \mathbb{P}(Y_{c,N+1} = y_{c,N+1}, X_{c,N} = x_{c,N} | Y^{C,N} = y^{C,N})$$

$$= \underset{\substack{y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{X}}}{\operatorname{argmax}} \sum_{\substack{y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{X}}} \mathbb{P}(x_{c,N} | y^{C,N}) \mathbb{P}(y_{c,N+1} | x_{c,N}, y^{C,N})$$

$$= \underset{\substack{y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{X}}}{\operatorname{argmax}} \sum_{\substack{y_{c,N+1} \in \Omega_{Y} \\ x_{c,N} \in \Omega_{X}}} \rho_{c,N}(x_{c,N}) \phi(x_{c,N}, y_{c,N+1}).$$





Figure G.13: Evolution of the variance of the hyperparameters estimators when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_2, \nu_3) = (-1, -3.7, 6.5, -3, 4, 2)$ and ν_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 $(\tau, \mu_0, \mu_1, \nu_0, \nu_2, \nu_3) = (-1, -3.7, 6.5, -3, 4, 2)$ and ν_1 varies from 0 to 6.5 with a 0.5 step.See Table 2 for definitions of parameters.

1003 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):

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

The partial derivative according to one group of hyperparameters (τ, ν) and μ 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 $(\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2) = (-1, -3.7, 6.5, -3, 4, 4)$ and ν_3 varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.

¹⁰¹¹ calculated separately. Consequently:

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

¹⁰¹² Once the Fisher information matrix is calculated, we can calculate a $(100(1 - \epsilon))$ confidence interval for any hyperparameter. The confidence interval for ¹⁰¹⁴ ν_1 is

$$\left[\hat{\nu}_1 - Z_{\epsilon/2}(\frac{1}{\sqrt{I(\nu)}})_{\nu_1}, \hat{\nu}_1 + Z_{\epsilon/2}(\frac{1}{\sqrt{I(\nu)}})_{\nu_1}\right]$$

where $(\frac{1}{\sqrt{I(\nu)}})_{\nu_1}$ corresponds to the $(\nu_1, \nu_1)^{\text{th}}$ 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 φ is the cumulative distribution function of the Normal distribution. Figs. H.16 and H.17 show the confidence intervals for ν_1 and ν_3 when ν_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.



Figure H.16: Evolution of the confidence interval of the hyperparameter ν_1 when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2) = (-1, -3.7, 6.5, -3, 4, 4)$ and ν_3 varies from 0 to 6.5 with a 0.5 step. See Table 2 for definitions of parameters.

¹⁰²² I. Alphabetic colonisation

1023 I.1. Definition

We propose here a colonisation process, referred to as *alphabetic coloni*sation, which assumes that the impact of colonisation in field c depends only on the number of neighbours in each state of Ω_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 ν_3 when the true values are $(\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2) = (-1, -3.7, 6.5, -3, 4, 4)$ and ν_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 $(m_0, \ldots, m_{|\Omega_Y|-1})$. 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 $y_n^{C\setminus c}$ by $f_{alpha}(y_n^{C\setminus c})$, then we have:

$$p_{x_{c,n},y_n^C} = \frac{1}{1 + \exp(-(\nu_0 + \nu_1 \times \frac{x_{c,n-1}}{|\Omega_X|} + \nu_2 \times \frac{y_{c,n}}{|\Omega_Y|} + \nu_3 \times \frac{f_{alpha}(y_n^{C \setminus c})}{|f_{alpha}(\Omega_Y^{C-1})|}))}$$

where $(\nu_0, \nu_1, \nu_2, \nu_3)$ are hyper parameters and where $|f_{alpha}(\Omega_Y^{C-1})|$ is the number of possible outputs of the function f_{mean} .

1027 In order to derive the expression of f_{alpha} , we first define g a function,

from $|\Omega_Y|^{C-1}$ to $|\Omega_Y|^C$, which sorts the elements of $y_n^{C\setminus c}$ 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, this implies (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_{alpha} is:

$$f_{alpha}(y_n^{C-1}) = \sum_{c'=C-1}^{1} \mathbb{1}_{\{g_{c'}(y_n^{C-1}) > g_{c'+1}(y_n^{C-1})\}} \sum_{j=g_{c'+1}(y_n^{C-1})+1}^{g_{c'}(y_n^{C-1})} \binom{|\Omega_Y| - j + c' - 1}{c' - 1},$$

where g_l is *l*th element of g, and the $\binom{|\Omega_Y|-j+l-1}{l-1}$ is the number of unordered samplings with replacement of l-1 elements among $|\Omega_Y| - j + 1$. The alphabetic colonisation considers more states that the mean colonisation: the function f_{alpha} has $\binom{C+|\Omega_Y|-2}{C-1}$ different states, whereas f_{mean} has only Ω_Y states. However, the computational complexity of calculating the state associated with $y_n^{C\backslash c}$ with the alphabetic colonisation is O(NC((C-1))).

1042 I.2. Identifiability

¹⁰⁴³ We establish the following identifiability theorem for the alphabetic coloni-¹⁰⁴⁴ sation process.

Theorem 4. The hyperparameter $(\mu, \nu) = (\mu_0, \mu_1, \nu_0, \nu_1, \nu_2, \nu_3)$ of a MHMM-DF with the alphabetic colonisation process is generically identifiable from seven consecutive observations if the following conditions hold: $|\Omega_X| \leq |\Omega_Y|$, $|\Omega_Y| > 2, C > 2.$

1049 $\mathbf{Proof.}$

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 ν_2 and ν_3 , we consider the following two transition probabilities: L7 from

$$(x_{n-1}^{C}, y_{n}^{C}) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix}) \text{ to } (x_{n}^{C}, y_{n+1}^{C}) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}) \text{ and } L8 \text{ from}$$

$$(x_{n-1}^{C}, y_{n}^{C}) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ \cdots \\ 2 \end{pmatrix}) \text{ to } (x_{n}^{C}, y_{n+1}^{C}) = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}) :$$

$$\begin{split} L7(\nu_0,\nu_2,\nu_3) &\propto \left[\frac{1}{1 + \exp(-(\nu_0 + \frac{1}{|\Omega_Y|}\nu_2 + \frac{f_{alpha}(1,\dots,1)}{|f_{alpha}(\Omega_Y^{C-1})|}\nu_3))} \right]^{(|\Omega_X|-1)C} \\ &\times \left[\exp(-(\nu_0 + \frac{1}{|\Omega_Y|}\nu_2 + \frac{f_{alpha}(1,\dots,1)}{|f_{alpha}(\Omega_Y^{C-1})|}\nu_3)) \right]^{(|\Omega_X|-1)C}. \end{split}$$

Therefore:

$$L7(\nu_{0},\nu'_{2},\nu'_{3}) = L7(\nu_{0},\nu_{2},\nu_{3})$$

$$<=> \\ \exp(\frac{1}{|\Omega_{Y}|}\nu'_{2} + \frac{f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C-1})|}\nu'_{3}) = \exp(\frac{1}{|\Omega_{Y}|}\nu_{2} + \frac{f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C-1})|}\nu_{3})$$

$$<=> \\ \frac{1}{|\Omega_{Y}|}\nu'_{2} + \frac{f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C-1})|}\nu'_{3} = \frac{1}{|\Omega_{Y}|}\nu_{2} + \frac{f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C-1})|}\nu_{3}.$$

$$\begin{split} L8(\nu_0,\nu'_2,\nu'_3) &= L8(\nu_0,\nu_2,\nu_3) \\ <=> \\ \frac{2}{|\Omega_Y|}\nu'_2 + \frac{f_{alpha}(2,\dots,2)}{|f_{alpha}(\Omega_Y^{C-1})|}\nu'_3 &= \frac{2}{|\Omega_Y|}\nu_2 + \frac{f_{alpha}(2,\dots,2)}{|f_{alpha}(\Omega_Y^{C-1})|}\nu_3. \end{split}$$

By subtracting L8 - 2L7, we obtain:

$$L8(\nu_{0},\nu_{2},\nu_{3}) - 2L7(\nu_{0},\nu_{2},\nu_{3}) = L8(\nu_{0},\nu_{2}',\nu_{3}') - 2L7(\nu_{0},\nu_{2}',\nu_{3}')$$

$$< = >$$

$$\left(\frac{f_{alpha}(2,...,2)}{|f_{alpha}(\Omega_{Y}^{C^{-1}})|} - \frac{2f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C^{-1}})|}\right)\nu_{3} = \left(\frac{f_{alpha}(2,...,2)}{|f_{alpha}(\Omega_{Y}^{C^{-1}})|} - \frac{2f_{alpha}(1,...,1)}{|f_{alpha}(\Omega_{Y}^{C^{-1}})|}\right)\nu_{3}'.$$

Thus if $f_{alpha}(2,...,2) - 2f_{alpha}(1,...,1) \neq 0$, identifiability of ν_3 is established. From the definition of f_{alpha} we have:

$$f_{alpha}(1,...,1) = \binom{|\Omega_Y| + C - 3}{C - 2},$$

$$^2 \quad (|\Omega_Y| - i + C - 2)$$

$$f_{alpha}(2,...,2) = \sum_{j=1}^{2} \binom{|\Omega_Y| - j + C - 2}{C - 2}.$$

1061 Thus:

$$f_{alpha}(2,...,2) - 2f_{alpha}(1,...,1) = -2\binom{|\Omega_Y| + C - 3}{C - 2} + \sum_{j=1}^{2} \binom{|\Omega_Y| - j + C - 2}{C - 2}$$
$$= \binom{|\Omega_Y| + C - 4}{C - 2} - \binom{|\Omega_Y| + C - 3}{C - 2}$$
$$= \frac{(|\Omega_Y| + C - 4)!}{(|\Omega_Y| - 2)!(C - 2)!} - \frac{(|\Omega_Y| + C - 3)!}{(|\Omega_Y| - 1)!(C - 2)!}$$
$$= \frac{(|\Omega_Y| + C - 4)!}{(|\Omega_Y| - 2)!(C - 2)!} (1 - \frac{|\Omega_Y| + C - 3}{|\Omega_Y| - 1}).$$

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

$$\begin{array}{rcl}
1 & = & \frac{|\Omega_Y| + C - 3}{|\Omega_Y| - 1} \\
|\Omega_Y| - 1 & = & |\Omega_Y| + C - 3 \\
C & = & 2.
\end{array}$$

Since $f_{alpha}(2, ..., 2) - 2f_{alpha}(1, ..., 1) = 0$ only when C = 2, then when $3 \leq C$, we have $\nu_3 = \nu'_3$. Finally, by using L8 or L7, we easily obtain that if $\nu_3 = \nu'_3$, then $\nu_2 = \nu'_2$.

Thus, we were able to establish that if $3 \leq C$, then $\mathbb{P}(h_n | h_{n-1}, \nu, \mu) =$ 1068 $\mathbb{P}(h_n|h_{n-1},\nu',\mu')$ implies $(\mu,\nu)=(\mu',\nu')$. To establish generic identifiability, 1069 conditions from Theorem 2 are required: $|\Omega_X| \leq |\Omega_Y|$, and C > 2. Addition-1070 ally, the number of observable states $|\Omega_Y|$ must be greater than 2 since three 1071 states were used to prove generic identifiability. The MHMM-DF with alpha-1072 betic colonisation is generically identifiable with at least two hidden states 1073 and at least three observable states with at least three patches and with the 1074 number of hidden states smaller or equal to the number of observable states. 1075 1076