Non-parametric Observation Driven HMM

Hanna Bacave, Pierre-Olivier Cheptou, Nikolaos Limnios, Nathalie Peyrard

Non-parametric Observation Driven HMM

Hanna Bacave[1] & Pierre-Olivier Cheptou[2] & Nikolaos Limnios[3] & Nathalie Peyrard[1]

[1] *INRAE, UR MIAT, Université de Toulouse, Castanet-Tolosan, France. {hanna.bacave, nathalie.peyrard}@inrae.fr*

[2] *CEFE-CNRS, Montpellier, France. pierre-olivier.cheptou@cefe.cnrs.fr*

[3] *Sorbonne University Alliance, Unversité de Technologie de Compiègne, LMAC, France. nikolaos.limnios@utc.fr*

Abstract. The hidden Markov models (HMM) are used in many different fields, to study the dynamics of a process that cannot be directly observed. However, in some cases, the structure of dependencies of a HMM is too simple to describe the dynamics of the hidden process. In particular, in some applications in finance or in ecology, the transition probabilities of the hidden Markov chain can also depend on the current observation. In this work we are interested in extending the classical HMM to this situation. We define a new model, referred to as the Observation Driven - Hidden Markov Model (OD-HMM). We present a complete study of the general non-parametric OD-HMM with discrete and finite state spaces (hidden and observed variables). We study its identifiability. Then we study the consistency of the maximum likelihood estimators. We derive the associated forward-backward equations for the E-step of the EM algorithm. The quality of the procedure is tested on simulated data sets. Finally, we illustrate the use of the model on an application on the study of annual plants dynamics. This works sets theoretical and practical foundations for a new framework that could be further extended, on one hand to the non-parametric context to simplify estimation, and on the other hand to the hidden semi-Markov models for more realism.

Keywords. non homogeneous HMM, identifiability, consistency, EM algorithm

1 Introduction

The hidden Markov models (HMM, Cappé et al. 2005) are used in many different fields such as, for example, medicine (Le Strat and Carrat 1999) to analyse epidemiologic surveillance data, signal processing (Gales and Young 2008) for speech recognition, ecology (McClintock et al. 2020) to reconstruct hidden or partially observed ecological dynamics, bioinformatics (Yoon 2009) for the analysis of biological sequences, or finance (Engel and Hamilton 1990) to predict the regime of a monetary system thanks to the exchange rate. Their interest lies the fact that they allow to study the dynamic of a process that cannot be directly observed. Indeed, in the above domains, the only observation available is often an imperfect information of the process (e.g. symptoms of a disease), or another process driven by the hidden one (e.g. accelerometer data used to classify animal behavior (Leos-Barajas et al. 2017)). In the field of HMM results have been established on the model identifiability (Allman et al. 2009; Cappé et al. 2005) and on the asymptotic properties of the Maximum Likelihood Estimator (MLE, Cappé et al. 2005). In practice, the MLE is computed using the Baum-Welch algorithm, a special case of the Expectation-Maximisation algorithm for HMM (Cappé et al. 2005).

The HMM relies on two assumptions: the hidden process is modeled by a Markov chain, and the observations are independents given the hidden chain. In some cases, this structure of dependencies is too simple to describe the dynamics of the hidden system. In particular, the transition probabilities of the hidden Markov chain can also be dependent on the current observation, as we will see in some examples below. In this work we are interested in extending the classical HMM to this situation. We refer to the new model as the Observation Driven - Hidden Markov Model (OD-HMM) exemplified below, which is an adaptation of a HMM where the next hidden state depends not only on the current hidden state but also on the current observation. Theoretical results and inference algorithms for HMM do not apply directly to the OD-HMM, mainly because in the OD-HMM the transition probabilities are not constant across time, they depend on the current observation. For this new dependency structure, it is necessary to study conditions for model identifiability and properties of the MLE. Furthermore, the forward-backward equations of the Baum-Welch algorithm must be adapted.

To highlight the usefulness of the OD-HMM, we present here two applications where the hidden process dynamics are driven by the observations, one in ecology and one in finance. The structure of the classical HMM is not always suitable to study the development of a partially observable species, in particular when the hidden and the observed processes represent two different life stage of the species, one dormant and one non dormant. Let us consider the case of plants. In practice we can only observe the grown plants while the seeds into the soil are not reachable. The seeds remain in the soil for several years. This make relevant the use of a model with hidden dynamical state like a HMM. However, in addition new seeds produced by the grown plants enter the soil each year. A classical HMM does not model this latter event, the OD-HMM is more suited.

The OD-HMM can also be used in finance as an extension of the Hamilton's Markov-switching Model (Hamilton 1989), to study series of financial data that oscillates between two regimes. The characteristics of the observed financial data (e.g. mean, variance) depends on the hidden regime. In Engel and Hakkio (1996) the authors propose a study of the exchange rate of the European monetary system which can be in a 'stable' or 'volatile' regime, which is unknown, from observations of the latter. The dependence structure of this model is similar to that of the OD-HMM, since the exchange rate observed at time $t-1$ influences the regime it is in at time t.

Extension of HMM to the case where the current observation influences the transition of the hidden chain has already been considered, on one hand for particular applications and on the other hand from a theoretical point of view to study the properties of the MLE.

Indeed, the OD-HMM has been used in ecology to model the dynamics of annual plants (Pluntz et al. 2018; Le Coz et al. 2019). In Pluntz et al. (2018), the proposed model is a parametric model for presence/absence of seeds as hidden states and the presence/absence of grown plants as observations. The simplification assumption considers a fixed probability of seed production equal to 1. Parameter estimation is performed by reformulating the model into the framework of classical HMM and applying the Baum-Welch algorithm. In Le Coz et al. (2019), the model takes into account dispersal of seeds from one patch to another, using multiple interacting chains. The parametric model is still specific to plants dynamics and the associated estimation method based on EM is proposed. So, these articles present parameterized OD-HMM dedicated to annual plants dynamics.

From a theoretical point of view, there is a multitude of resources available about the consistency of MLE for the classical HMM (Baum and Petrie 1966; Leroux 1992; Gámiz et al. 2023), which we could have adapted to the OD-HMM case. However, there exists theoretical works on the properties of the MLE for more complex dependency structures between observations and hidden states than in the classical HMM (Ailliot and Pène 2015; Pouzo et al. 2022), including the OD-HMM case. In Ailliot and Pène (2015) the authors define the framework of non-homogeneous Markov-switching models and propose a study of consistency of MLE of a general model with dependency between the observations themselves at multiple times and between observations and hidden states and some applications in ecology. In this work the state spaces can either be continuous or discrete but the results are given in the continuous case. The dependency structure of the OD-HMM is a particular case of the general framework of Ailliot and Pène (2015), however they do not consider it in the different particular cases they study. In Pouzo et al. (2022), the author established sufficient conditions for the consistency and the local asymptotic normality of the MLE for a Markov regime switching model where transition probabilities can depend on covariates. Their results are established in the case where the observations state space is continuous and the hidden state space is discrete. In conclusion, the works related to the OD-HMM dependency structure are either based on parameter models dedicated to a particular application (plants dynamics or finance), or theoretical results for more general models or with continuous state space for the observed process.

We start in section 2 by defining this model and studying its identifiability. Then, in section 3, we study the consistency of the MLE. We also provide the necessary modifications, compared to the EM algorithm for HMM, in the Expectation step and all detailed calculations to obtain the associated EM algorithm. In section 4 we perform several tests on simulated data in order to evaluate the quality of the EM estimates. Finally, in section 5, we illustrate how the model can be used to obtain knowledge on the mechanisms driving the development of an annual plant, in particular on the survival of seed in the soil.

2 The non parametric OD-HMM

2.1 Model definition

Let us consider two sets of random variables, indexed by (discrete) time: Y_t, the observed system at time t, with state space $\Omega_Y = \{1, ..., D\}$ and Z_t the hidden state at time t, with state space $\Omega_Z = \{1, ..., S\}$. We denote the vector of the observations between $t = 0$ and $t = M$ by $Y_{0:M} = (Y_0, Y_1, Y_2, ..., Y_M)$. In the same way, the vector of the hidden states between $t = 0$ and $t = M$ is denoted $Z_{0:M} = (Z_0, Z_1, ..., Z_M)$. In the most common version of a HMM, presented in Figure 1a, (Z_t) is a Markov chain and $\mathbb{P}(Y_{0:M} = y_{0:M} \mid Z_{0:M} = z_{0:M}) = \prod_{t=0}^{M} \mathbb{P}(Y_t = y_t \mid Z_t = z_t)$[1].

Here we consider an extension of the classical HMM, where the observation Y_{t-1} has an influence on the next hidden variable Z_t. This corresponds to the graphical representation of conditional independencies shown on Figure 1b.

The joint distribution of $(Z_{0:M}, Y_{0:M})$ is fully determined by the following distributions. The initial probability is noted $\pi(z_0)$, where:

$$\forall z_0 \in \Omega_Z, \pi(z_0) = \mathbb{P}(Z_0 = z_0).$$

The emission probability is noted $R(z_t, y_t)$, where:

$$\forall z_t \in \Omega_Z, y_t \in \Omega_Y, R(z_t, y_t) = \mathbb{P}(Y_t = y_t | Z_t = z_t).$$

The transition matrix is noted $P_{y_{t-1}}(z_{t-1}, z_t)$, where:

$$\forall (z_t, z_{t-1}) \in \Omega_Z^2, y_{t-1} \in \Omega_Y, P_{y_{t-1}}(z_{t-1}, z_t) = \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, Y_{t-1} = y_{t-1}).$$

Note that due to the influence of observations on hidden states, the transition matrix depends of the observation y_{t-1}. Therefore a specificity of this model is that there are as many transition matrices as observed states, as opposed to the classical HMM. The OD-HMM is non-homogeneous.

Eventhough the model is non-parametric, for the sake of simplicity, we will refer to these distributions as the model parameters. Since they are all probabilities with a constraint to sum up to one, the set of parameters is $\theta = (P_y(z', z), y \in \Omega_Y, z' \in \Omega_Z, z \in \Omega_Z \backslash \{S\}) \cup (R(y, z), y \in \Omega_Y \backslash \{D\})$, in which there is no the initial distribution π because it is not estimated. It takes values in $\Theta = [0, 1]^{|\Omega_Z|(|\Omega_Z|-1)|\Omega_Y|+(|\Omega_Y|-1)|\Omega_Z|}$.

Definition 1 (OD-HMM)**.** *(Z_t, Y_t) is said to follow an Observation Driven HMM (OD-HMM) if the conditional independencies between observed variables and hidden states as described in Figure 1b. The OD-HMM with parameter θ is denoted $\mathcal{M}_\theta^{ODHMM}$.*

1. By convention, we use uppercase letters, Z_t or Y_t, for the random variables and lowercase letters, z_t or y_t, for realizations.

(a) for an HMM. (b) for an OD-HMM.

Figure 1: Graphical representation of conditional independencies in the chain (Z_t, Y_t).

2.2 Generic identifiability

As explained in Allman et al. (2009), the requirement of identifiability may be too strict when considering statistical parameter estimation. Indeed, for some models, only a subset of parameters of measure zero may not be identifiable and in practice estimation will perform well. So here we consider generic identifiability and we provide sufficient conditions to ensure the generic identifiability of the parameters of the OD-HMM. We recall first the definition of identifiability (referred to as strict identifiability in Allman et al. (2009)) and generic identifiability.

Definition 2 (Strict identifiability)**.** *Let $\mathcal{F}(\Theta) = \{\mathbb{P}_\theta, \theta \in \Theta\}$ be a family of probability distributions. We say that the model's parameters θ are strictly identifiable if $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$ implies that $\theta = \theta'$.*

Definition 3 (Generic identifiability)**.** *Let $\mathcal{F}(\Theta) = \{\mathbb{P}_\theta, \theta \in \Theta\}$ be a family of probability distributions. We say that the model's parameters θ are generically identifiable if the elements of Θ that do not satisfy $\mathbb{P}_\theta = \mathbb{P}_{\theta'} \Longrightarrow \theta = \theta'$ are of mesure zero in the parameter space.*

It means that any observed data set has probability one of being drawn from a distribution with identifiable parameters. In Allman et al. (2009) the authors have established the following proposition on generic identifiability for HMM.

Proposition A (Generic identifiability for HMM (Allman et al. 2009))**.** *The parameters of an HMM with r hidden states and k observable states are generically identifiable from the marginal distribution of $2L+1$ consecutive variables provided L satisfies:*

$$\binom{L+k-1}{k-1} \geqslant r.$$

In order to establish sufficient conditions for the generic identifiability of the OD-HMM parameters θ, we first reformulate the model $\mathcal{M}_\theta^{ODHMM}$ into an equivalent HMM, $\mathcal{M}_\theta^{HMM}$, with the dimension of the state space of the hidden variable being $r = |\Omega_Z| \times |\Omega_Y|$ and an observed state space of dimension $k = |\Omega_Y|$, and whose transition and emission probabilities are functions of θ. Using Proposition A we establish sufficient conditions for the generic identifiability of the parameters of $\mathcal{M}_\theta^{HMM}$. Then we establish that there is a one-to-one map between the two formulations, i.e. if $\theta \neq \theta'$ then $\mathcal{M}_\theta^{HMM}$ and $\mathcal{M}_{\theta'}^{HMM}$ are not the same model. Therefore the condition for generic identifiability holds also for the original OD-HMM model, $\mathcal{M}_\theta^{ODHMM}$.

Let us first present the reformulation of a OD-HMM model (Z_t, Y_t) as a HMM. The hidden variable is $H_t = (Z_t, Y_t) \in \Omega_Z \times \Omega_Y$ and the observed variable is a copy of Y_t i.e. $O_t = Y_t \in \Omega_Y$. The couple (H_t, O_t) satisfies the definition of a HMM since H_t is a Markov chain, and conditionally to (H_t) the O_ts are mutually independent and each O_t depends only on H_t. Now, we express P^{HMM} the transition matrix and R^{HMM} the emission matrix of $\mathcal{M}_\theta^{HMM}$, using the parameters of $\mathcal{M}_\theta^{ODHMM}$:

$$\begin{aligned} P_\theta^{\text{HMM}}(h_{t-1}, h_t) &= R(z_t, y_t) P_{y_{t-1}}(z_{t-1}, z_t), \\ R_\theta^{\text{HMM}}(h_t, o_t) &= \mathbb{1}_{(y_t = o_t)}. \end{aligned}$$

Then, using Proposition A, we know that the parameters of $\mathcal{M}_\theta^{HMM}$ (i.e. the elements of the transition P_θ^{HMM} and the emission matrices R_θ^{HMM}) are generically identifiable when the observed chain is longer than $2L+1$, where L satisfies

$$\binom{L+|\Omega_Y|-1}{|\Omega_Y|-1} \geqslant |\Omega_Z||\Omega_Y|.$$

It is easy to show (see Appendix A) that two different parameters θ and θ' of an OD-HMM that lead to the same transition matrix P^{HMM} and the same emission matrix R^{HMM}, are equal ($\theta = \theta'$). Therefore, generic identifiability also holds for $\mathcal{M}_\theta^{ODHMM}$

Proposition 1 (Generic identifiability for OD-HMM)**.** *The parameters of an ODHMM with $|\Omega_Z|$ hidden states and $|\Omega_Y|$ observable states are generically identifiable from the marginal distribution of $2L+1$ consecutive variables provided L satisfies:*

$$\binom{L+|\Omega_Y|-1}{|\Omega_Y|-1} \geqslant |\Omega_Z||\Omega_Y|.$$

Example 1. *For $\Omega_Z = 2$ and $\Omega_Y = 2$, the parameters θ are identifiable as soon as the chain has more than $2L+1 = 7$ observations. Indeed,*

$$\binom{L+2-1}{2-1} \geqslant 2 \times 2 \quad \Leftrightarrow \quad L \geqslant 3.$$

3 Maximum likelihood estimation

We are interested in the calculation of the Maximum Likelihood Estimator (MLE), noted $\hat{\theta}$, defined by the following formula:

$$\hat{\theta} = \arg\max_\theta L(\theta; y_{0:M}),$$

where $L(\theta; y_{0:M}) = \mathbb{P}_\theta(Y_{0:M} = y_{0:M})$ is the likelihood.

3.1 Consistency

We work on probability space $(\Omega_Z \times \Omega_Y, \mathcal{P}(\Omega_Z \times \Omega_Y), \mathbb{P}_\theta, \theta \in \Theta)$ and with a reference measure μ.

As shown in Ailliot and Pène (2015, theorem 2) we can obtain the consistency of the MLE of a more general model, named Non Homogeneous Markov-Switching Auto-Regressive model (NHMS-AR model) in which the state spaces can be either continuous or discrete, but the results are given in the continuous case. The NHMS-AR model is based on the two following assumptions:

- the distribution of Z_t, conditionally to $\{Z_{t'} = z_{t'}\}_{t'<t}$ and $\{Y_{t'} = y_{t'}\}_{t'<t}$, only depends on $y_{t-s:t-1}$ and z_{t-1}, where $s \in \{1, ..., t\}$.
- the distribution of Y_t, conditionally to $\{Z_{t'} = z_{t'}\}_{t'<t}$ and $\{Y_{t'} = y_{t'}\}_{t'<t}$, only depends on $y_{t-s:t-1}$ and z_t.

Thus, the OD-HMM is a particular case of the NHMS-AR model, where $s = 1$ and where y_t does not depend on y_{t-1}. Besides the chain (Z_t) is also non homogeneous with respect to the time, because P_y depends on y.

So, to adapt the work of Ailliot and Pène (2015) to the case of the OD-HMM MLE, we consider the transition matrix of the couple (Z_t, Y_t) noted:

$$\forall (i,a),(j,b) \in \Omega_Z \times \Omega_Y, \tilde{P}_\theta(i,a;j,b) = P_a(i,j;\theta)R(j,b;\theta).$$

Thus, the Markov chain of the couple (Z_t, Y_t) is homogeneous, with the transition matrix $\tilde{P}_\theta$ and the stationary distribution of the couple $\tilde{\pi}_\theta$. Besides, the marginals of $\tilde{\pi}_\theta$ are noted $\overline{\mathbb{P}}_\theta$ and $\overline{\mathbb{P}}_\theta^Y$ for the observed chain. Finally, $\overline{\mathbb{E}}_\theta$ is the expectation taken with respect to $\overline{\mathbb{P}}_\theta$.

The adaptation of the assumptions of the theorem 2 from Ailliot and Pène (2015) to the case of the OD-HMM MLE, with our notations, takes the following form :

(A1) Θ is a compact space ;

(A2) The chain (Z_t, Y_t) is ergodic with an invariant probability for each $\theta \in \Theta$ denoted $\tilde{\pi}_\theta$;

(A3) The elements of $\overline{\mathbb{P}}_\theta$ are absolutely continuous with respect to $\mathbb{P}_\theta$ for all $\theta \in \Theta$;

(A4) The elements of P_y and R are continuous in θ, for any y in Ω_Y.

Under theses assumptions, theorem 2 from Ailliot and Pène 2015 becomes the following proposition.

Proposition B (Consistency of the NHMS-AR model MLE (Ailliot and Pène 2015), in the particular case of the OD-HMM MLE)**.** *If we have :*

1. $0 < P_{y,-} := \min_{\theta,z_0,y_0,z_1} P_{y_0}(z_0,z_1;\theta) \leqslant P_{y,+} := \max_{\theta,z_0,y_0,z_1} P_{y_0}(z_0,z_1;\theta) < \infty$;
2. $B_- = \overline{\mathbb{E}}_{\theta*}[|\ln(\min_\theta \sum_{z_0 \in \Omega_Z} R(z_0, Y_0;\theta)|] < \infty$;
3. $B_+ = \overline{\mathbb{E}}_{\theta*}[|\ln(\max_\theta \sum_{z_0 \in \Omega_Z} R(z_0, Y_0;\theta)|] < \infty$;
4. $\forall \theta \in \Theta, \sum_{z \in \Omega_Z} R(z, Y_0) < \infty$, $\mathbb{P}_{\theta*}$*-a.s.* ;
5. $\forall \theta \in \Theta$, *for* μ*-a.e.*, $\lim_{k \longrightarrow \infty} ||Q^{*k}(.|(Z_0,Y_0)) - h_\theta|| = 0$, *where* $Q^{*k}(.|(Z_0,Y_0))$ *is the density of* (Z_k, Y_k^k) *with respect to the measure* μ *and* h_θ *is the limit density when* k *tends to* ∞.

So, for all $z_0 \in \Omega_Z$*, the limit values of the MLE* $\hat{\theta}$ *are* $\mathbb{P}_{\theta*}$*-a.s. contained in the space* $\{\theta \in \Theta; \overline{\mathbb{P}}_\theta^Y = \overline{\mathbb{P}}_{\theta*}^Y\}$*, where* $\theta*$ *is the true value of the parameters.*

Now, we focus on the assumptions made on the model to propose a fully adapted version of the proposition above. Indeed, unlike the general NHMS-AR model, in the OD-HMM we assume that the elements of $\tilde{P}_\theta$ are all strictly positive. So, this assumption leads us to the intermediate Proposition 2.

Proposition 2. *Let* (Z_t, Y_t) *follow a OD-HMM model with parameter* $\theta \in \Theta$*. There exists an invariant probability* $\tilde{\pi}_\theta$ *for the Markov chain* (Z_t, Y_t)*.*

The above proposition allows us to ensure the existence of a stationary distribution for the OD-HMM. In the following we will assume that the initial distribution is equal to $\tilde{\pi}_\theta$. Therefore, the process is stationary.

Now, to completely adapt Proposition B to the case of OD-HMM MLE estimators, we first consider assumptions (A1) to (A4) and we explain why assumptions (A1), (A2) and (A3) are always satisfied by the OD-HMM.

(A1) The space of the OD-HMM parameters θ is $\Theta = [0,1]^{|\Omega_Z|(|\Omega_Z|-1)|\Omega_Y|+(|\Omega_Y|-1)|\Omega_Z|}$, so it is compact.

(A2) Since we assume that the elements of $\tilde{P}_\theta$ are all strictly positive, this implies that the chain (Z,Y) is ergodic.

(A3) In the case of the OD-HMM, the state spaces are finite, so it is possible to free oneself from the assumptions on the continuity of the fact that the continuity is locally verified.

So only assumption $(A4)$ is necessary.

In a second step we consider assumptions (1) to (5) and we explain why there are all satisfied by the OD-HMM :

1. Under assumption (A2), $\min_{\theta,z_0,y_0,z_1} P_{y_0}(z_0,z_1;\theta) > 0$ is satisfied because the elements of P_y are all stricly positives. Besides, since Ω_Z and Ω_Y are finite, $\max_{\theta,z_0,y_0,z_1} P_{y_0}(z_0,z_1;\theta) < \infty$ is also satisfied.

2. The terms of the matrix R are strictly positives, so $|\ln R(z,y;\theta)| < \infty$. Also, we have $\sum_{z_0\in\Omega_Z} R(z_0,Y_0;\theta) < \infty$. Finally, we obtain $B_- < \infty$.

3. By the same reasoning, we have $B_+ < \infty$.

4. Since Ω_Z is finite, $\sum_{z\in\Omega_Z} R(z,Y_0;\theta)$ is finite.

5. Since Ω_Z and Ω_Y are finite, this assumption is verified when the chain (Z_t,Y_t) is ergodic.

We finally obtain the following proposition concerning the consistency of the maximum likelihood estimators of the OD-HMM.

Proposition 3 (Consistency of the OD-HMM MLE)**.** *Under the assumption (A4), for all $z_0 \in \Omega_Z$, the limit values of the MLE $\hat{\theta}$ are $\mathbb{P}_{\theta*}$-a.s. contained in the space $\{\theta \in \Theta; \overline{\mathbb{P}}_\theta^Y = \overline{\mathbb{P}}_{\theta*}^Y\}$, where $\theta*$ is the true value of the parameters.*

Note that, since the states spaces are finite, in the non-parametric case, the assumption $(A4)$ is satisfied and therefore the consistency of the OD-HMM MLE estimator is guaranteed.

3.2 EM algorithm

We compute the MLE of θ using the Expectation-Maximisation (EM) algorithm. To take into account the dependence between the observations (Y_t) and the hidden states (Z_t), as shown in Figure 1b, we propose an adaptation of the EM for HMM (Cappé et al. 2005). We could consider estimating the $\mathcal{M}_\theta^{HMM}$ model, which would allow us to use the EM algorithm for HMM without modification. However, the transition matrix describing the hidden states $H_t = (Z_t, Y_t)$ is too large to be well estimated.

We consider the situation where we have C realizations $(y_{c,t})$ of C independent identically distributed OD-HMM $(Z_{c,t}, Y_{c,t})$, where $c \in \{1,...,C\}$. In the following, we denote the vector of hidden state at time t for chain 1 to chain C $Z_{1:C,t} = (Z_{1,t}, Z_{2,t}, ..., Z_{C,t})$. In the same way we denote $Y_{1:C,t} = (Y_{1,t}, Y_{2,t}, ..., Y_{C,t})$ the vector of the observations at time t for the C chains. Besides, we denote $\pi(z_{c,0};\theta), P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t};\theta)$ and $R(z_{c,t}, y_{c,t};\theta)$ the probabilities taken conditionally to θ.

With these notations, the complete likelihood is equal to (see Appendix B.1) :

$$\begin{aligned} L(\theta; z_{1:C,0:M}, y_{1:C,0:M}) &= \mathbb{P}(Z_{1:C,0:M} = z_{1:C,0:M}, Y_{1:C,0:M} = y_{1:C,0:M}|\theta) \\ &= \prod_{c=1}^{C} \left\{ \pi(z_{c,0};\theta) R(z_{c,0}, y_{c,0};\theta) \prod_{t=1}^{M} P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t};\theta) R(z_{c,t}, y_{c,t};\theta) \right\}. \end{aligned}$$

The EM algorithm for OD-HMM is an iterative algorithm and each iteration is composed of two steps. Let us consider $\theta^{(m)}$ the parameter estimates at iteration m and define $Q(\theta|\theta^{(m)})$ the intermediate quantity. The two steps in each iteration are the following :

1. Expectation Step (E step): we calculate the marginal distributions involved in the expression of the intermediate quantity $Q(\theta|\theta^{(m)})$. It relies on an adaptation of the Forward-Backward algorithm.

2. Maximisation step (M step): we update the set of parameters θ thanks to the quantities found in the E Step, by resolving $\theta^{(m+1)} = \arg\max_{\theta} Q(\theta|\theta^{(m)})$.

These two steps are repeated until the algorithm converges.

The intermediate quantity $Q(\theta|\theta^{(m)})$, can be decomposed into three terms, one depending on the initial distribution, another depending on the transition matrix and the last depending on the emission distribution (see Appendix B.2).

$$
\begin{aligned}
Q(\theta|\theta^{(m)}) &= \mathbb{E}\left[\ln \mathbb{P}(Y_{1:C,0:M}, Z_{1:C,0:M}|\theta)|Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] \\
&= \sum_{c=1}^{C} \sum_{z_0 \in \Omega_Z} \ln\left(\pi(z_0;\theta)\right) \mathbb{P}(Z_{c,0} = z_0|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \\
&+ \sum_{c=1}^{C} \sum_{t=1}^{M} \sum_{(z,z') \in \Omega_Z^2} \ln\left(P_{y_{c,t-1}}(z, z';\theta)\right) \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z'|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \\
&+ \sum_{c=1}^{C} \sum_{t=0}^{M} \sum_{z' \in \Omega_Z} \ln\left(R(z', y_{c,t};\theta)\right) \mathbb{P}(Z_{c,t} = z'|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}).
\end{aligned}
$$

3.2.1 E step

The EM algorithm for OD-HMM is very similar to the EM algorithm for HMM. However, the Backward algorithm in step E has been modified to take into account the fact that the transition matrix depends on the observation.

The E step consist in computing the marginal probabilities of interest appeared in the expression of $Q(\theta|\theta^{(m)})$. They are :

- $\forall 0 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z$,
$$\rho_{c,t}^{(m)}(z_t) = \mathbb{P}(Z_{c,t} = z_t|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)});$$

- $\forall 1 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall (z_{t-1}, z_t) \in \Omega_Z^2$,
$$\xi_{c,t}^{(m)}(z_{t-1}, z_t) = \mathbb{P}(Z_{c,t-1} = z_{t-1}, Z_{c,t} = z_t|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}).$$

To obtain $\rho_{c,t}^{(m)}(z_t)$ and $\xi_{c,t}^{(m)}(z_{t-1}, z_t)$, we introduce the following variables :

- $\alpha_{c,t}^{(m)}(z_t)$, such as, $\forall 0 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z$,
$$\alpha_{c,t}^{(m)}(z_t) = \mathbb{P}(Y_{c,0:t} = y_{c,0:t}, Z_{c,t} = z_t|\theta^{(m)});$$

- $\beta_{c,t}^{(m)}(z_t)$, such as, $\forall 0 \leqslant t < M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z$,
$$\beta_{c,t}^{(m)}(z_t) = \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M}|Z_{c,t} = z_t, Y_{c,t} = y_{c,t}, \theta^{(m)}).$$

The E step works by using the Forward-Backward algorithm. The specificity of the Forward-Backward algorithm for the OD-HMM compared to that of the HMM is in the expression of $\beta_{c,t}^{(m)}(z_t)$, because it is taken conditionally to the current observations.

In the Forward algorithm, we express $\alpha_{c,t}^{(m)}(z_t)$ using the following recurrence formula (see Appendix B.3.1 for the calculations):

$$\forall 1 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z,$$
$$\alpha_{c,t}^{(m)}(z_t) = R^{(m)}(z_t, y_{c,t}) \sum_{z_{t-1} \in \Omega_Z} \alpha_{c,t-1}^{(m)}(z_{t-1}) P_{y_{c,t-1}}^{(m)}(z_{t-1}, z_t),$$

where $\alpha_{c,0}^{(m)}(z_0) = R^{(m)}(z_0, y_{c,0})\pi_(z_0)^{(m)}$.

In the Backward algorithm, we compute $\beta_{c,t}^{(m)}(z_t)$ using the following recurrence formula (see Appendix B.3.2 for the calculations):

$$\forall 0 \leqslant t < M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z,$$
$$\beta_{c,t}^{(m)}(z_t) = \sum_{z_{t+1} \in \Omega_Z} R^{(m)}(z_{t+1}, y_{c,t+1}) \beta_{c,t+1}^{(m)}(z_{t+1}) P_{y_{c,t}}^{(m)}(z_t, z_{t+1}),$$

where $\beta_{c,M}^{(m)}(z_M) = 1$.

The quantities $\alpha_{c,t}^{(m)}(z_t)$ and $\beta_{c,t}^{(m)}(z_t)$ are used to compute $\rho_{c,t}^{(m)}(z_t)$ and $\xi_{c,t}^{(m)}(z_{t-1}, z_t)$, as follows (see Appendix B.3.3 for the calculations):

- $\forall 0 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall z_t \in \Omega_Z,$

$$\rho_{c,t}^{(m)}(z_t) = \frac{\alpha_{c,t}^{(m)}(z_t)\beta_{c,t}^{(m)}(z_t)}{\sum\limits_{z_t \in \Omega_Z} \alpha_{c,t}^{(m)}(z_t)\beta_{c,t}^{(m)}(z_t)};$$

- $\forall 1 \leqslant t \leqslant M, \forall c \in \{1, ..., C\}, \forall (z_{t-1}, z_t) \in \Omega_Z^2,$

$$\xi_{c,t}^{(m)}(z_{t-1}, z_t) = \frac{\alpha_{c,t-1}^{(m)}(z_{t-1}) P_{y_{c,t-1}}^{(m)}(z_{t-1}, z_t) R^{(m)}(z_t, y_{c,t}) \beta_{c,t}^{(m)}(z_t)}{\sum\limits_{z_t \in \Omega_Z} \alpha_{c,t}^{(m)}(z_t)\beta_{c,t}^{(m)}(z_t)}.$$

3.2.2 M step

During the M step we resolve the maximisation problem to obtain the expression of updated parameters. We express these parameters in terms of $\rho_{c,t}^{(m)}(z_t)$ and $\xi_{c,t}^{(m)}(z_{t-1}, z_t)$ as follows (see Appendix B.4):

- $\forall y \in \Omega_Y, \forall z_t \in \Omega_Z, \forall z_{t-1} \in \Omega_Z,$

$$P_y^{(m+1)}(z_{t-1}, z_t) = \frac{\sum\limits_{c=1}^{C} \sum\limits_{t=1}^{M} \xi_{c,t}^{(m)}(z_{t-1}, z_t) \mathbb{1}_{(y_{c,t-1}=y)}}{\sum\limits_{c=1}^{C} \sum\limits_{t=1}^{M} \sum\limits_{z'_t \in \Omega_Z} \xi_{c,t}^{(m)}(z_{t-1}, z'_t) \mathbb{1}_{(y_{c,t-1}=y)}};$$

- $\forall z_t \in \Omega_Z, \forall y \in \Omega_Y,$

$$R^{(m+1)}(z_t, y) = \frac{\sum\limits_{c=1}^{C} \sum\limits_{t=0}^{M} \rho_{c,t}^{(m)}(z_t) \mathbb{1}_{(y_{c,t}=y)}}{\sum\limits_{c=1}^{C} \sum\limits_{t=0}^{M} \rho_{c,t}^{(m)}(z_t)}.$$

3.2.3 Initialization and stopping criterion

To initialize the EM algorithm for OD-HMM, we simulate N_{init} sets of initial parameters values and we run the algorithm for each of them.

The algorithm is stopped when the first of the two criteria are met: either a number N_{iter} is reached or the algorithm has converged. To check for convergence, at each iteration m, we compute a distance between the estimation at iteration m and the one at iteration $m+1$. Since the st of parameters θ is composed of matrices, and the rows of these matrices all sum up to one, we consider the following distance between the i-th row of $\theta^{(m)}$, denoted $\theta_i^{(m)}$, and $\theta_i^{(m+1)}$:

$$dist(\theta_i^{(m)}, \theta_i^{(m+1)}) = \frac{1}{K} \sum_{k=1}^{K} \frac{|\theta_{i,k}^{(m)} - \theta_{i,k}^{(m+1)}|}{\theta_{i,k}^{(m)}},$$

where K is the length of θ_i and $\theta_{i,k}$ is the k-th element of the θ i-th row. When the average distance between all rows of θ is smaller than a given threshold ϵ we consider that the algorithm has converged.

When the N_{init} algorithms have converged or stopped, we keep the best estimate, i.e. the one that is associated to the larger likelihood. The likelihood is easily computed using the αs and βs (see Appendix B.3.3).

4 Validation of the estimation procedure on simulated data

In order to validate the estimation procedure, we realised three tests of data sets simulated from the OD-HMM with different true parameters corresponding to problems of increasing difficulty. With the first test, the transition and emission matrices are very contrasted. In the second test, the rows of the transition matrices are chosen close to $(\frac{1}{2}, \frac{1}{2})$. In the last test, the rows of the emission matrix are similar, so the observations bring little information on the hidden process. The true values of the parameters for each of the three tests are presented in Table 2. In each case, Ω_Z and Ω_Y are of size 2. So, by Proposition 1, the OD-HMM are generically identifiable as soon as the chain has more than 7 observations. We describe below the protocol common to all the tests and then we provide the results.

4.1 Protocol

For a given test, associated to the true parameters θ^*, and a given C, we set the initial distribution to $\pi = (1, 0)$ to simulate C chains (Z_t) and (Y_t) of length $M = 500$. Then, we used the EM algorithm for OD-HMM to obtain the parameters estimator $\hat{\theta}$ from the C observed chains. There are 6 parameters to estimate. We ran the EM with $N_{init} = 10$, $N_{iter} = 500$, and $\epsilon = 0.001$. To the same distance than for testing for convergence.

In order to capture the variability in the EM estimates we ran these steps 30 times. Finally we repeated the whole procedure for increasing numbers of chains: $C = 10, 50, 100$.

Table 1: True transition and emission matrices for the three tests.

	True parameters
Test 1	$P_0^* = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}, P_1^* = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, R^* = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$
Test 2	$P_0^* = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}, P_1^* = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}, R^* = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}$
Test 3	$P_0^* = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}, P_1^* = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, R^* = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$

4.2 Results

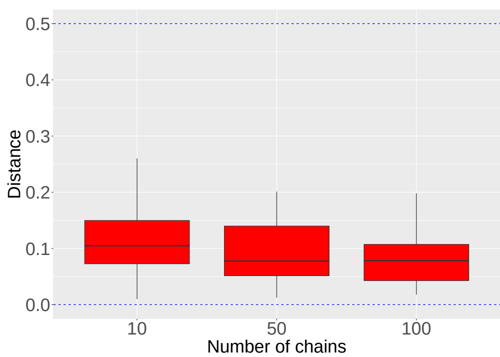
4.2.1 First test

As expected, we noticed that the average distance between the estimation $\hat{\theta}$ and the true parameters θ^* is decreasing with the number of chains. Whatever the number of chains, it is lower than 0.1 (i.e a 10% error) which is of very good quality. The convergence rate is increasing with the number of chains. Indeed, we have a convergence rate of 43.4% for $C = 10$ chains, of 56.7% for $C = 50$ chains and finally of 83.4% for $C = 100$ chains.

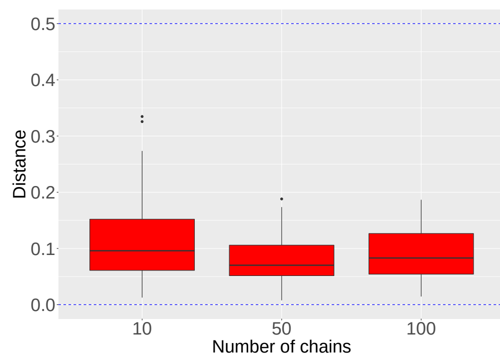
If we consider now specifically the estimation of the transition matrices (see Figure 2), we observed that the median distance between $\hat{P}_0$ and P_0^* is lower than 0.1 and decreasing with C. In the worst case, the distance does not exceed 0.25 of the value of the real parameter. The median distance between $\hat{P}_1$ and P_1^* is also lower than 0.1 and its ninth is lower than 0.25, but there are some extreme values for $C = 10$ and $C = 50$ and the distance does not decrease with C.

The estimation of the emission matrix is of very good quality (see Figure 3): the median value is much lower than 0.1 and is decreasing with the number of chains.

To conclude, in the case where the matrices are contrasted, i.e. the weights on the rows of P are distant from $(\frac{1}{2}, \frac{1}{2})$ and the rows of R are very different from each other, the estimation by the EM algorithm for OD-HMM is satisfactory.



(a) distance between $\hat{P}_0$ and P_0^*



(b) distance between $\hat{P}_1$ and P_1^*

Figure 2: Boxplot of the distance between the estimated transition matrices and the true ones depending on the number of chains, and for the first test.

4.2.2 Second test

Now, the weights on the rows of P are closer to $(\frac{1}{2}, \frac{1}{2})$ than in the first test. So, in theory estimating the transition matrices should be more difficult. This is the case, however, the average distance between the estimated and the true parameters remains lower than 0.15, it is, on average, of 0.143 for $C = 10$ chains, 0.138 for $C = 50$ chains and of 0.127 for $C = 100$ chains. We notice that it is decreasing with the number of chains. The distance for $C = 10$ is greater than the others but we can put it in perspective with the fact that the rate of convergence is only 40%. For the other values of C the convergence rate is increasing until 100% for $C = 100$.

The distances for the transition matrices and for the emission matrices are displayed separately on Figure 4 and Figure 5. There is no more systematic decrease of the distance with C. We can see that the distance between $\hat{P}_0$ and P_0^* remains low while the distance between $\hat{P}_1$ and P_1^* has increased compared to test 1. The median distance between the estimated and the true emission matrix remains of very good quality eventhough with more variability than for test 1.

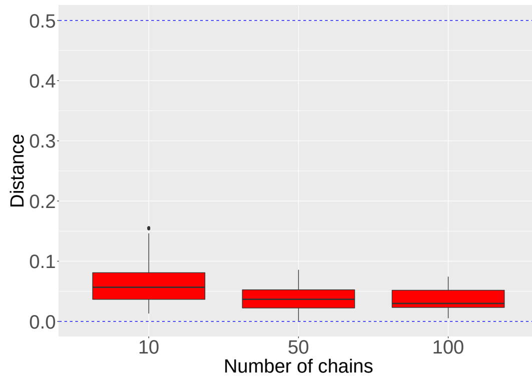
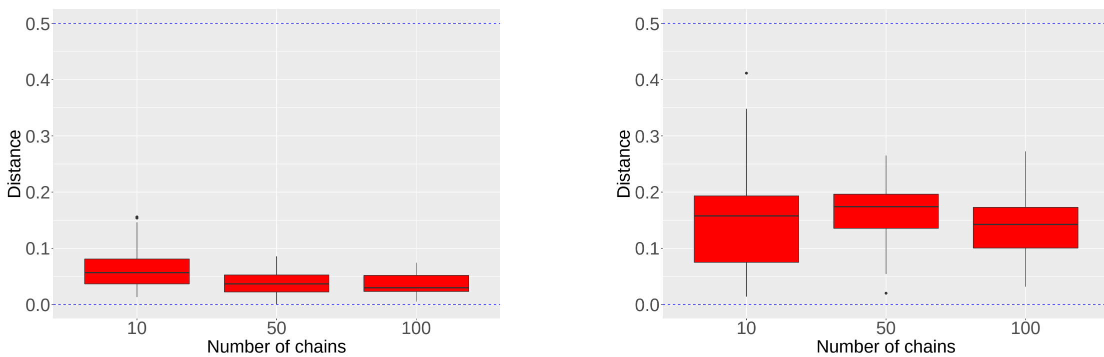


Figure 3: Boxplot of the distance between the estimated emission matrix and the true one depending on the number of chains, and for the first test.



(a) distance between $\hat{P}_0$ and P_0^*

(b) distance between $\hat{P}_1$ and P_1^*

Figure 4: Boxplot of the distance between the estimated transition matrices and the true ones depending on the number of chains, and for the second test.

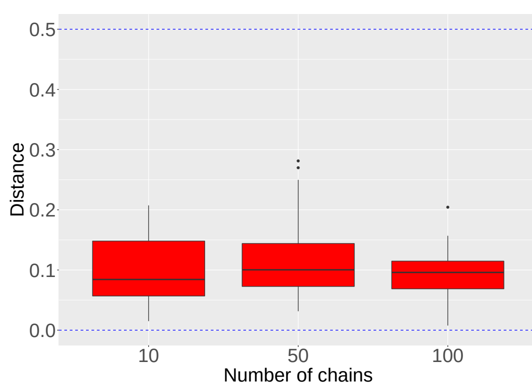
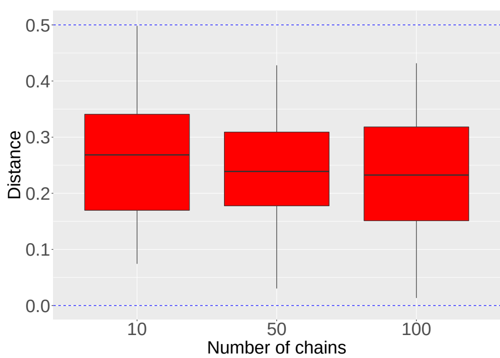


Figure 5: Boxplot of the distance between the estimated emission matrix and the true one depending on the number of chains, and for the second test.

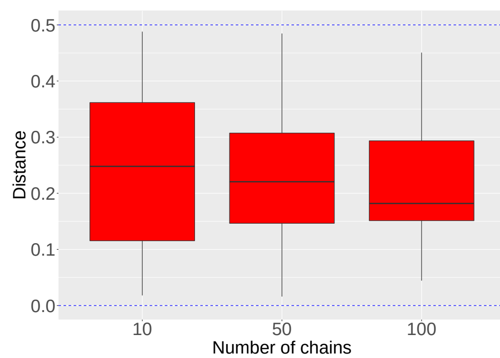
4.2.3 Third test

The third test is the more difficult: since the rows of the emission matrix are similar it is more difficult than for tests 1 and 2 to know which hidden state has generated the observation. One consequence is that for test 3 we faced a problem of label switching, i.e. the roles of the 2 hidden states in the matrices were switched during the estimation. We considered that label switching occurs when the average distance between the rows of $\theta*$ and of $\hat{\theta}$ is smaller when we permute the hidden state than without. We observed 23.3% of label switching for $C = 10$, 20% for $C = 50$, and 0% for $C = 100$. The distances we discuss know are obtained after correction of label switching when present. We observed that the quality of the estimation is really less good than in the two first tests. Indeed, the average distance between the estimated parameters and the true ones is decreasing from 0.22 to 0.15 when C increases, which is larger than before. The poor estimation quality for $C = 10$ can be explained by the fact that 33% of the estimates did not converge. For the other values of C the convergence is 97% then 100%.

If we look separately at the transition matrices (Figure 6) and the emission matrix (Figure 7), we observed that the quality of the estimation of the former is strongly degraded while the estimation of the latter is of quality similar to that of the other tests.



(a) distance between $\hat{P}_0$ and P_0^*



(b) distance between $\hat{P}_1$ and P_1^*

Figure 6: Boxplot of the distance between the estimated transition matrices and the true ones depending on the number of chains, and for the third test.

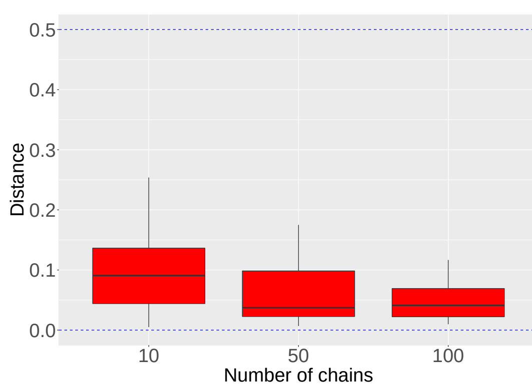


Figure 7: Boxplot of the distance between the estimated emission matrix and the true one depending on the number of chains, and for the third test.

5 Illustration on simulated dynamics of a real ecological system

In this section, we present the use of the OD-HMM for modeling annual plants dynamics. Based on a real use case and using simulated data, we propose two studies. First we illustrate the discriminating power of the model. Second we illustrate how a key statistics of the dynamics can be estimated from data using the model and the EM algorithm.

5.1 The model for annual plants dynamics

The model we use is that of Pluntz et al. (2018) who described a parametric OD-HMM for annual plants. The model take into account four mains processes of plants dynamics. Germination is the transition from seed to adult plant. The probability of germination is denoted g. Seed production characterizes the ability of plants to produce and disperse seeds into the soil. The probability of seed production is denoted d. Seed survival designates the fact that the seeds stay in the soil from one year to the next. The probability of seed survival is s. Finally, colonization represents the arrival of exogenous seed, for instance by wind. The probability of this event is c. The time step is the year The observation Y_t represents presence or absence of standing flora, while the hidden state Z_t represents presence or absence of seeds in the soil at year t.

From probabilities g, c, d and s we can construct the matrices P_0^*, P_1^* and R^* as shown in Pluntz et al. (2018).

$$
\begin{aligned}
P_0 &= \begin{pmatrix} 1-c & c \\ (1-c)(1-s) & 1-(1-c)(1-s) \end{pmatrix}, \\
P_1 &= \begin{pmatrix} (1-c)(1-d) & 1-(1-c)(1-d) \\ (1-c)(1-d)(1-s) & 1-(1-c)(1-d)(1-s) \end{pmatrix}, \\
R &= \begin{pmatrix} 1 & 0 \\ 1-g & g \end{pmatrix}.
\end{aligned}
$$

To understand how to derive this expressions, let us describe how to obtain $P_1(1,0) = \mathbb{P}((Z_{c,t} = 0 \mid Z_{c,t-1} = 1, Y_{c,t-1} = 1)$. It is the probability that that there are no seed in the soil at time t, knowing that there were seeds in the soil and standing flora at $t-1$. Thus, the colonization, the production of seed and the survival of seeds has not succeed, which occurs with probability $P_1(1,0) = (1-c)(1-d)(1-s)$.

Note that P_0, P_1 and R can be computed from g, c, d and s but the opposite operation is not possible. Furthermore, in Pluntz et al. (2018), d is set to 1 and not estimated. So in the following we will use the values of colonization, dispersion and survival probabilities obtained by Pluntz et al. (2018) for two species, and we will choose an arbitrary value of d to to construct the matrices defining the OD-HMM and simulate data. Then we will use EM to estimate these matrices, not the four parameters.

5.2 Possibility to detect the absence of seed survival

Here, the objective is to test if the above EM OD-HMM is able to detect that a plant does not use seed survival (meaning that it will instead use colonization to avoid extinction), by using a model selection approach and the Bayesian Information Criterion (BIC, (Raftery 1995)).

In the case where $s = 0$, the two transition matrices are expressed as follows:

$$
P_0 = \begin{pmatrix} 1-c & c \\ (1-c) & c \end{pmatrix}, \qquad P_1 = \begin{pmatrix} (1-c)(1-d) & 1-(1-c)(1-d) \\ (1-c)(1-d) & 1-(1-c)(1-d) \end{pmatrix}.
$$

The two rows of each transition matrix are equal. This corresponds to a constrained OD-HMM with less parameters to estimate (4 instead of 6). We modified the M-step of EM to take into account this constraint. The corresponding formula for updating the P_y matrices is: $\forall y, z_t, z_{t-1} \in \Omega_Y \times \Omega_Z^2$,

$$P_y^{(m+1)}(z_{t-1}, z_t) = \frac{\sum_{c=1}^{C} \sum_{t=1}^{M} \sum_{z_{t-1} \in \{0,1\}} \xi_{c,t}^{(m)}(z_{t-1}, z_t) \mathbb{1}_{(y_{c,t-1}=y)}}{\sum_{c=1}^{C} \sum_{t=1}^{M} \mathbb{1}_{(y_{c,t-1}=y)}}.$$

For this test, we used the parameter values estimated in Pluntz et al. (2018) for the species *Taraxacum officinale*, i.e. $c = 0.22$, $g = 0.3$, and we fixed $s = 0$ and $d = 0.5$. We obtained the following transition and emission matrices:

$$P_0^* = \begin{pmatrix} 0.88 & 0.22 \\ 0.88 & 0.22 \end{pmatrix}; P_1^* = \begin{pmatrix} 0.44 & 0.56 \\ 0.44 & 0.56 \end{pmatrix}; R^* = \begin{pmatrix} 1 & 0 \\ 0.7 & 0.3 \end{pmatrix}.$$

We considered them as the true parameters θ^* and we used them to simulate $C = 100$ observations sequences of length $M = 500$ which form the data. Based on these data, we estimated two models with EM: a full OD-HMM without constraint, $\mathcal{M}_f$, and a constrained OD-HMM, $\mathcal{M}_c$, where we impose that the 2 rows of each transition matrices are equal. Then we computed BIC for the two models. The formula of BIC for a model $\mathcal{M}$, with associated MLE $\theta_{\mathcal{M}}$ is:

$$BIC(\mathcal{M}) = -2\ln(L(\theta_{\mathcal{M}})) + K_{\mathcal{M}} \ln(C \times M),$$

where L is the likelihood and $K_{\mathcal{M}}$ is the number of free parameters. For the full OD-HMM there are $K_{\mathcal{M}_f} = 6$ parameters. For the constrained OD-HMM there are $K_{\mathcal{M}_c} = 4$ parameters.

We repeated this experimentation 50 times by simulating 50 data set from the same true parameters θ^*. We obtained that $BIC(\mathcal{M}_c)$ is always lower than $BIC(\mathcal{M}_f)$, which means that the constrained model $\mathcal{M}_c$ is better than the other. Besides, the difference between these two BIC criteria is, on average, of 20.53, and its minimal value is 15.14. As presented in Raftery (1995), a difference greater than 10 can be interpreted as a strong evidence for the selected model. So, we can conclude that the OD-HMM model is able to detect the absence of survival with strong evidence.

5.3 Estimation of the average time of seeds stock persistence

Now we illustrate how the OD-HMM can be used to estimate how long can there be seeds present in the soil. For that, we used the parameters estimated by Pluntz et al. (2018) for the species *Alopecurus Myosoroides* to generate a pseudo true data set, with $C = 100$ chains of length $M = 500$. For this species, $g = 0.59$, $s = 0.51$, $c = 0.09$ and we choose $d = 0.5$, which leads to the following true transition and emission matrices:

$$P_0^* = \begin{pmatrix} 0.91 & 0.09 \\ 0.4459 & 0.5541 \end{pmatrix}; P_1^* = \begin{pmatrix} 0.455 & 0.545 \\ 0.223 & 0.777 \end{pmatrix}; R^* = \begin{pmatrix} 1 & 0 \\ 0.41 & 0.59 \end{pmatrix}.$$

From these data we estimated $\hat{P}_0$, $\hat{P}_1$ and $\hat{R}$, with the EM algorithm for OD-HMM. Then, we simulated the trajectories of 100 chains of length 500 from which we estimated the average time of continuous presence of seeds in the soil (average duration of seed stock persistence). For sake of comparison, we realized the same estimation but estimating an HMM, instead of a OD-HMM, from the data. The estimation was obtained by modifying the EM of OD-HMM by forcing the transition probabilities to be equal. We compared the value of these average durations with the empirical estimate computed on the pseudo-real data set, as shown in the Table 2.

As results, we obtained that the duration computed using the EM for OD-HMM is very similar to the empirical estimate computed on data generated by the true model. On contrary, the duration estimated under the assumption of a HMM model is poorer, which illustrates the risk in ignoring the influence of the observation on the hidden state.

Table 2: Comparison of the duration of seeds stock persistence estimated by EM algorithm for HMM or for OD-HMM with the empirical estimates obtained from the data.

	Min.	**1st Decile**	**Median**	**Mean**	**9th Decile**	**Max.**
Empirical estimate from data	1	1	2	2.973	6	12
EM for OD-HMM	1	1	2	3.155	7	22
EM for HMM	1	2	9	12.166	27	113

6 Conclusion

In this paper, we have developed a new framework, extension of the classical HMM, originally motivated by applications in ecology and the modeling of populations dynamics when part of the individuals are hidden. The interest of the OD-HMM goes beyond ecology, and has also application in finance, for example to study the exchange rate regime of a money.

In these applications, the OD-HMM framework provides tools to estimate the parameters involved in the dynamics of the systems under study. For this purpose, we have chosen to focus on the estimation of the parameters by the maximum likelihood principle.

On the theoretical aspect, we studied the consistency of the MLE of the OD-HMM by adapting the work of Ailliot and Pène (2015) to the case of the OD-HMM. The next step would be to prove the asymptotic normality of the MLE of the OD-HMM, either by extending the work of Bickel et al. (1998) and Gámiz et al. (2023) proposed for a classical HMM, or by adapting the work of Pouzo et al. (2022) proposed for more complex structures including the case of the OD-HMM. In the same way, measuring the reliability of the system modeled by the OD-HMM (Durand and Gaudoin 2016; Gámiz et al. 2023) may be an interesting perspective.

On the practical side, we derived the EM algorithm for OD-HMM to estimate the parameters of the model. By performing three experiments, we obtained that when the values of the true parameters P^* and R^*, or just R^*, are contrasted, the estimation provides good results. However, it is not recommended to use the EM algorithm for OD-HMM when the true emission matrix R^* is not contrasted. These experiments were conducted on observation trajectories of length at most 500 to avoid the appearance of possible numerical problems. Moreover, we have carried out the experiments for C varying from 10 to 100 chains in order to be able to run more tests in a faster way, although running the algorithm on a larger number of chains would certainly improve the quality of the estimate. In the same way, although the EM algorithm for OD-HMM presented here is built for any size of discrete and finite state spaces, we only tested the 2×2 case to perform multiple experiments and to be able to detect the possible appearance of label switching, as in the third proposed experiment. To work with more complex systems (more trajectories and states) it would be wise to consider a parametric model, as underlined in Bazzi et al. (2022) : "*A key challenge is to specify an appropriate and parsimonious function that links the lagged dependent variables to future transition probabilities*".

As this work stems from an application needs, in order to move towards more realism, we plan to extend the OD-HMM to the case where the hidden chain is a semi-Markov chain (Barbu and Limnios 2008; Abdullah and Hoek 2022; Yu 2016). The sojourn time distribution will then be generalized to any probability distribution and not only by a geometric one as in a HMM. Then, it will be necessary to adapt the framework of Hidden Semi-Markov Models (HSMM) to the case of OD-HSMM by modelling the impact of observations on sojourn time distribution. Results on identifiability of HSMM and properties of the MLE (Barbu and Limnios 2008) will also have to be adapted, as well as the EM algorithm for HSMM to the OD-HSMM (Bulla 2006; Barbu and Limnios 2008; Yu 2016).

Acknowledgement

HB, POC, NL and NP, acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-21-CE40-005 (project HSMM-INCA).

A Generic identifiability

Let us consider two ODHMM, the first one with parameter θ and the second with parameter θ'. They transition matrix and emission matrix are respectively P_y and R, and P'_y and R'. We assume that their reformulations as a HMM lead to the same model, i.e. $\mathcal{M}_\theta^{HMM} = \mathcal{M}_{\theta'}^{HMM}$ and we show that this implies that $\theta = \theta'$.

If the two transition matrices P_θ^{HMM} and $P_{\theta'}^{\text{HMM}}$ are identical we have that

$$\begin{cases} P_\theta^{\text{HMM}}((0,0),(0,0)) & = P_{\theta'}^{\text{HMM}}((0,0),(0,0)) \\ P_\theta^{\text{HMM}}((0,0),(0,1)) & = P_{\theta'}^{\text{HMM}}((0,0),(0,1)) \\ ... \\ P_\theta^{\text{HMM}}((0,0),(0,D)) & = P_{\theta'}^{\text{HMM}}((0,0),(0,D)) \end{cases}$$

Using the definition of P_θ^{HMM} in terms of the original OD-HMM, we obtain

$$\begin{cases} P_0(0,0)R(0,0) & = P'_0(0,0)R'(0,0) \\ P_0(0,0)R(0,1) & = P'_0(0,0)R'(0,1) \\ ... \\ P_0(0,0)R(0,D) & = P'_0(0,0)R'(0,D) \end{cases}$$

Adding all the lines leads to

$$P_0(0,0)[\sum_{d=0}^{D} R(0,d)] = P'_0(0,0)[\sum_{d=0}^{D} R'(0,d)].$$

Since $\sum_{d=0}^{D} R(0,d) = 1$ et $\sum_{d=0}^{D} R'(0,d) = 1$ we obtain $P_0(0,0) = P'_0(0,0)$. If we replace $P'_0(0,0)$ by $P_0(0,0)$ in the above system we obtain also that $\forall d \in \{0,...,D\},\ R(0,d) = R'(0,d)$.

We can perform the same calculs with all possible values for (z_{t-1}, y_{t-1}, z_t) and we will obtain that $P_{y_{t-1}}(z_{t-1}, z_t) = P'_{y_{t-1}}(z_{t-1}, z_t)$ and that $\forall y_t \in \{0,...,D\},\ R(z_t, y_t) = R'(z_t, y_t)$. This establishes that if $\mathcal{M}_\theta^{HMM} = \mathcal{M}_{\theta'}^{HMM}$, then $\mathcal{M}_\theta^{ODHMM} = \mathcal{M}_{\theta'}^{ODHMM}$.

B Calculations of EM algorithm

B.1 Complete likelihood expression

We recall the definition of the complete likelihood

$$L(\theta; z_{1:C,0:M}, y_{1:C,0:M}) = \mathbb{P}(Y_{1:C,0:M} = y_{1:C,0:M}, Z_{1:C,0:M} = z_{1:C,0:M}|\theta).$$

Since the C chains are i.i.d we have that

$$\begin{aligned}
L(\theta; z_{1:C,0:M}, y_{1:C,0:M}) &= \prod_{c=1}^{C} \mathbb{P}(Z_{c,0:M} = z_{c,0:M}, Y_{c,0:M} = y_{c,0:M}|\theta) \\
&= \prod_{c=1}^{C} \mathbb{P}(Z_{c,M} = z_{c,M}, Y_{c,M} = y_{c,M}|Z_{c,0:M-1} = z_{c,0:M-1}, Y_{c,0:M-1} = y_{c,0:M-1}, \theta) \\
&\quad \times \mathbb{P}(Z_{c,0:M-1} = z_{c,0:M-1}, Y_{c,0:M-1} = y_{c,0:M}|\theta) \\
&= \dots \\
&= \prod_{c=1}^{C} \prod_{t=1}^{M} \mathbb{P}(Z_{c,t} = z_{c,t}, Y_{c,t} = y_{c,t}|Z_{c,t-1} = z_{c,t-1}, Y_{c,t-1} = y_{c,t-1}, \theta) \\
&\quad \times \mathbb{P}(Z_{c,0} = z_{c,0}, Y_{c,0} = y_{c,0}|\theta) \\
&= \prod_{c=1}^{C} \prod_{t=1}^{M} \mathbb{P}(Y_{c,t} = y_{c,t}|Z_{c,t} = z_{c,t}, \theta)\mathbb{P}(Z_{c,t} = z_{c,t}|Z_{c,t-1} = z_{c,t-1}, Y_{c,t-1} = y_{c,t-1}, \theta) \\
&\quad \times \mathbb{P}(Y_{c,0} = y_{c,0}|Z_{c,0} = z_{c,0}, \theta)\mathbb{P}(Z_{c,0} = z_{c,0}, \theta)
\end{aligned}$$

Finally,

$$L(y_{1:C,0:M}, z_{1:C,0:M}|\theta) = \prod_{c=1}^{C} \left\{ \pi(z_{c,0}) R(z_{c,0}, y_{c,0}) \prod_{c,t=1}^{M} P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t}) R(z_{c,t}, y_{c,t}) \right\}.$$

B.2 Intermediate quantity calculation expression

We recall the definition of the intermediate quantity, where $\theta^{(m)}$ is the value of the parameters at iteration m.

$$Q(\theta|\theta^{(m)}) = \mathbb{E}\left[\ln(\mathbb{P}(Y_{1:C,0:M}, Z_{1:C,0:M}|\theta) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right].$$

We replace the complete likelihood by its expression and we take its logarithm:

$$\begin{aligned}
\ln(\mathbb{P}(Y_{1:C,0:M} = y_{1:C,0:M}, Z_{1:C,0:M} = z_{1:C,0:M}|\theta)) &= \ln\left(\prod_{c=1}^{C} \left\{ \pi(z_{c,0}) R(z_{c,0}, y_{c,0}) \prod_{t=1}^{M} P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t}) R(z_{c,t}, y_{c,t}) \right\}\right) \\
&= \sum_{c=1}^{C} \ln(\pi(z_{c,0})) + \ln(\prod_{t=0}^{M} R(z_{c,t}, y_{c,t})) + \ln(\prod_{t=1}^{M} P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t})) \\
&= \sum_{c=1}^{C} \ln(\pi(z_{c,0})) + \sum_{t=0}^{M} \ln(R(z_{c,t}, y_{c,t})) + \sum_{t=1}^{M} \ln(P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t}))
\end{aligned}$$

So

$$
\begin{aligned}
Q(\theta|\theta^{(m)}) &= \mathbb{E}\left[\sum_{c=1}^{C}\left\{\ln(\pi(Z_{c,0}) + \sum_{t=0}^{M}\ln(R(Z_{c,t},Y_{c,t})) + \sum_{t=1}^{M}\ln(P_{Y_{c,t}}(Z_{c,t-1},Z_{c,t}))\right\} \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] \\
&= \sum_{c=1}^{C}\left\{\mathbb{E}\left[\ln(\pi(Z_{c,0})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] + \mathbb{E}\left[\sum_{t=0}^{M}\ln(R(Z_{c,t},Y_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right]\right. \\
&\quad \left. + \ \mathbb{E}\left[\sum_{t=1}^{M}\ln(P_{Y_{c,t-1}}(Z_{c,t-1},Z_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right]\right\} \\
&= \sum_{c=1}^{C}\left\{\mathbb{E}\left[\ln(\pi(Z_{c,0})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] + \sum_{t=0}^{M}\mathbb{E}\left[\ln(R(Z_{c,t},Y_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right]\right. \\
&\quad \left. + \ \sum_{t=1}^{M}\mathbb{E}\left[\ln(P_{Y_{c,t-1}}(Z_{c,t-1},Z_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right]\right\}
\end{aligned}
$$

where

$$
\mathbb{E}\left[\ln(\pi(Z_{c,0})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] = \sum_{z_{c,0}\in\Omega_Z}\ln(\pi(z_{c,0})) \times \mathbb{P}(Z_{c,0} = z_{c,0}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}),
$$

and

$$
\mathbb{E}\left[\ln(R(Z_{c,t},Y_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right] = \sum_{z_{c,t}\in\Omega_Z}\ln(R(z_{c,t},y_{c,t})) \times \mathbb{P}(Z_{c,t} = z_{c,t}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})
$$

In the same way, we have :

$$
A = \mathbb{E}\left[\ln(P_{Y_{c,t-1}}(Z_{c,t-1},Z_{c,t})) \mid Y_{1:C,0:M} = y_{1:C,0:M}, \theta^{(m)}\right]
$$

where

$$
\begin{aligned}
A = &\sum_{(z_{c,t},z_{c,t-1})\in\Omega_Z^2}\ln(P_{y_{c,t-1}}(z_{c,t-1},z_{c,t})) \\
&\times \mathbb{P}(Z_{c,t} = z_{c,t}, Z_{c,t-1} = z_{c,t-1}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})
\end{aligned}
$$

So, we finally obtain:

$$
\begin{aligned}
Q(\theta|\theta^{(m)}) &= \sum_{c=1}^{C}\sum_{z_{c,0}\in\Omega_Z}\ln\left(\pi(z_{c,0})\right)\mathbb{P}(Z_{c,0} = z_{c,0}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \\
&+ \sum_{c=1}^{C}\sum_{t=1}^{M}\sum_{(z_{c,t},z_{c,t-1})\in\Omega_Z^2}\ln\left(P_{y_{c,t-1}}(z_{c,t-1},z_{c,t})\right)\mathbb{P}(Z_{c,t-1} = z_{c,t-1}, Z_{c,t} = z_{c,t}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \\
&+ \sum_{c=1}^{C}\sum_{t=0}^{M}\sum_{z_{c,t}\in\Omega_Z}\ln\left(R(z_{c,t},y_{c,t})\right)\mathbb{P}(Z_{c,t} = z_{c,t}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})
\end{aligned}
$$

B.3 E Step calculations

During the E Step, we want to calculate the marginal probabilities of interest appeared in the expression of $Q(\theta|\theta^{(m)})$. We denote:

- $\rho_{c,t}^{(m)}(z_t) = \mathbb{P}(Z_{c,t} = z_t|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})$,
- $\xi_{c,t}^{(m)}(z_{t-1}, z_t) = \mathbb{P}(Z_{c,t-1} = z_{t-1}, Z_{c,t} = z_t|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})$.

To obtain $\rho_{c,t}^{(m)}(z_t)$ and $\xi_{c,t}^{(m)}(z_{t-1}, z_t)$, we introduce the following variables :

- $\alpha_{c,t}^{(m)}(z_t)$, such as $\alpha_{c,t}(z_t) = \mathbb{P}(Y_{c,0:t} = y_{c,0:t}, Z_{c,t} = z_t|\theta^{(m)})$;
- $\beta_{c,t}^{(m)}(z_t)$, such as $\beta_{c,t}(z_t) = \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M}|Z_{c,t} = z_t, Y_{c,t} = y_{c,t}, \theta^{(m)})$.

B.3.1 Forward algorithm for computing the αs

We begin by expressing $\alpha_0(z_{c,0})$:

$$\begin{aligned}
\alpha_{c,0}^{(m)}(z_{c,0}) &= \mathbb{P}(Y_{c,0} = y_{c,0}, Z_{c,0} = z_{c,0}|\theta^{(m)}) \\
&= \mathbb{P}(Y_{c,0} = y_{c,0}|Z_{c,0} = z_{c,0}, \theta^{(m)})\mathbb{P}(Z_{c,0} = z_{c,0}|\theta^{(m)}) \\
&= R(z_{c,0}, y_{c,0})^{(m)}\pi(z_{c,0})^{(m)}
\end{aligned}$$

Now, $\forall 1 \leqslant t \leqslant M$, we have :

$$\begin{aligned}
\alpha_{c,t}^{(m)}(z) &= \mathbb{P}(Y_{c,0:t} = y_{c,0:t}, Z_{c,t} = z|\theta^{(m)}) \\
&= \mathbb{P}(Y_{c,0:t} = y_{c,0:t}|Z_{c,t} = z, \theta^{(m)})\mathbb{P}(Z_{c,t} = z|\theta^{(m)}) \\
&= \mathbb{P}(Y_{c,t} = y_{c,t}|Z_{c,t} = z, \theta^{(m)})\mathbb{P}(Y_{c,0:t-1} = y_{c,0:t-1}|Z_{c,t} = z, \theta^{(m)})\mathbb{P}(Z_{c,t} = z|\theta^{(m)}) \\
&= \mathbb{P}(Y_{c,t} = y_{c,t}|Z_{c,t} = z\theta^{(m)})\mathbb{P}(Y_{c,0:t-1} = y_{c,0:t-1}, Z_{c,t} = z|\theta^{(m)}) \\
&= R^{(m)}(z, y_{c,t}) \times \sum_{z_{c,t-1}\in\Omega_Z} \mathbb{P}(Y_{c,0:t-1} = y_{c,0:t-1}, Z_{c,t} = z, Z_{c,t-1} = z_{c,t-1}|\theta^{(m)}) \\
&= R^{(m)}(z, y_{c,t}) \times \\
&\quad \sum_{z_{c,t-1}\in\Omega_Z} \mathbb{P}(Z_{c,t} = z|Y_{c,0:t-1} = y_{c,0:t-1}, Z_{c,t-1} = z_{c,t-1}, \theta^{(m)})\mathbb{P}(Y_{c,0:t-1} = y_{c,0:t-1}, Z_{c,t-1} = z_{c,t-1}|\theta^{(m)}) \\
&= R^{(m)}(z, y_{c,t}) \times \sum_{z_{c,t-1}\in\Omega_Z} \mathbb{P}(Z_{c,t} = z|Y_{c,t-1} = y_{c,t-1}, Z_{c,t-1} = z_{c,t-1}, \theta^{(m)}) \times \alpha_{c,t-1}^{(m)}(z_{c,t-1}) \\
&= R^{(m)}(z, y_{c,t}) \times \sum_{z_{c,t-1}\in\Omega_Z} P_{y_{c,t-1}}^{(m)}(z_{c,t-1}, z) \times \alpha_{c,t-1}^{(m)}(z_{c,t-1}).
\end{aligned}$$

So, we obtain the forward recursive expression

$$\forall 1 \leqslant t \leqslant M, \alpha_{c,t}^{(m)}(z) = R^{(m)}(z, y_{c,t}) \sum_{z_{c,t-1}\in\Omega_Z} \alpha_{c,t-1}^{(m)}(z_{c,t-1})P_{y_{c,t-1}}^{(m)}(z_{c,t-1}, z).$$

B.3.2 Backward algorithm for computing the βs

We start by initializing $\beta_M^{(m)}(z)$. By convention we set $\beta_M^{(m)}(z) = 1$.

Now, $\forall 1 \leqslant t \leqslant M - 1$, we have :

$$\begin{aligned}
\beta_{c,t}^{(m)}(z) &= \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M}|Z_{c,t} = z, Y_{c,t} = y_{c,t}, \theta^{(m)}) \\
&= \sum_{z_{c,t+1}\in\Omega_Z} \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M}, Z_{c,t+1} = z_{c,t+1}|Z_{c,t} = z, Y_{c,t} = y_{c,t}, \theta^{(m)}) \\
&= \sum_{z_{c,t+1}\in\Omega_Z} \mathbb{P}(Y_{c,t+2:M} = y_{c,t+2:M}|Y_{c,t+1} = y_{c,t+1}, Z_{c,t+1} = z_{c,t+1}, \theta^{(m)}) \\
&\quad \times \mathbb{P}(Y_{c,t+1} = y_{c,t+1}, Z_{c,t+1} = z_{c,t+1}|Y_{c,t} = y_{c,t}, Z_{c,t} = z, \theta^{(m)}) \\
&= \sum_{z_{c,t+1}\in\Omega_Z} \beta_{c,t+1}^{(m)}(z_{c,t+1})\mathbb{P}(Z_{c,t+1} = z_{c,t+1}|Y_{c,t} = y_{c,t}, Z_{c,t} = z, \theta^{(m)}) \\
&\quad \times \mathbb{P}(Y_{c,t+1} = y_{c,t+1}|Z_{c,t+1} = z_{c,t+1}, \theta^{(m)}) \\
&= \sum_{z_{c,t+1}\in\Omega_Z} \beta_{c,t+1}^{(m)}(z_{c,t+1}) \times P_{y_{c,t}}^{(m)}(z, z_{c,t+1}) \times R^{(m)}(z_{c,t+1}, y_{c,t+1}).
\end{aligned}$$

Finally we obtain the backward recursive expression

$$\forall 1 \leqslant t \leqslant M-1, \beta_{c,t}^{(m)}(z) = \sum_{z_{c,t+1} \in \Omega_Z} R^{(m)}(z_{c,t+1}, y_{c,t+1}) \beta_{c,t+1}^{(m)}(z_{c,t+1}) P_{y_{c,t}}^{(m)}(z, z_{c,t+1}).$$

B.3.3 Expression of quantities ρ and ξ in terms of α and β :

We have

$$\rho_{c,t}^{(m)}(z) = \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}).$$

and we can express it as follows :

$$\rho_{c,t}^{(m)}(z) = \frac{\mathbb{P}(Z_{c,t} = z, Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})}{\mathbb{P}(Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})}.$$

In the same way, we have :

$$\xi_{c,t}^{(m)}(z', z') = \mathbb{P}(Z_{c,t-1} = z', Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}),$$

which we can transform as follows :

$$\xi_{c,t}^{(m)}(z', z') = \frac{\mathbb{P}(Z_{c,t-1} = z', Z_{c,t} = z', Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})}{\mathbb{P}(Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})}.$$

First, we express the quantity $\mathbb{P}(Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})$, which is the likelihood, in terms of α and β. We have :

$$\begin{aligned}
\mathbb{P}(Y_{c,0:M} = y_{c,0:M} | \theta^{(m)}) &= \sum_{z_{c,t} \in \Omega_Z} \mathbb{P}(Y_{c,0:M} = y_{c,0:M}, Z_{c,t} = z_{c,t} | \theta^{(m)}) \\
&= \sum_{z_{c,t} \in \Omega_Z} \mathbb{P}(Y_{c,0:M} = y_{c,0:M} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Z_{c,t} = z_{c,t} | \theta^{(m)}) \\
&= \sum_{z_{c,t} \in \Omega_Z} \mathbb{P}(Y_{c,0:t} = y_{c,0:t} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&\times \mathbb{P}(Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&= \sum_{z_{c,t} \in \Omega_Z} \beta_{c,t}^{(m)}(z_{c,t}) \times \mathbb{P}(Y_{c,0:t} = y_{c,0:t} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&= \sum_{z_{c,t} \in \Omega_Z} \beta_{c,t}^{(m)}(z_{c,t}) \times \mathbb{P}(Y_{c,0:t} = y_{c,0:t}, Z_{c,t} = z_{c,t} | \theta^{(m)}) \\
&= \sum_{z_{c,t} \in \Omega_Z} \beta_{c,t}^{(m)}(z_{c,t}) \times \alpha_{c,t}^{(m)}(z_{c,t}).
\end{aligned}$$

So, we have

$$\mathbb{P}(Y_{c,0:M} = y_{c,0:M} | \theta^{(m)}) = \sum_{z_{c,t} \in \Omega_Z} \alpha_{c,t}^{(m)}(z_{c,t}) \beta_{c,t}^{(m)}(z_{c,t}).$$

- **Expression of ρ:**

  Let us denote $A = \mathbb{P}(Z_{c,t} = z_{c,t}, Y_{c,0:M} = y_{c,0:M} | \theta^{(m)})$. We can decompose A as follows:

$$\begin{aligned}
A &= \mathbb{P}(Y_{c,0:M} = y_{c,0:M} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Z_{c,t} = z_{c,t} | \theta^{(m)}) \\
&= \mathbb{P}(Y_{c,0:t} = y_{c,0:t} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M} | Z_{c,t} = z_{c,t}, Y_{c,t} = y_{c,t}, \theta^{(m)}) \mathbb{P}(Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&= \beta_{c,t}^{(m)}(z_{c,t}) \times \mathbb{P}(Y_{c,0:t} = y_{c,0:t} | Z_{c,t} = z_{c,t}, \theta^{(m)}) \mathbb{P}(Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&= \beta_{c,t}^{(m)}(z_{c,t}) \times \mathbb{P}(Y_{c,0:t} = y_{c,0:t}, Z_{c,t} = z_{c,t} | \theta^{(m)}) \\
&= \beta_{c,t}^{(m)}(z_{c,t}) \alpha_{c,t}^{(m)}(z_{c,t}).
\end{aligned}$$

So, we deduce that :

$$\rho_{c,t}^{(m)}(z) = \frac{\alpha_{c,t}^{(m)}(z)\beta_{c,t}^{(m)}(z)}{\sum\limits_{z\in\Omega_Z} \alpha_{c,t}^{(m)}(z)\beta_{c,t}^{(m)}(z)}.$$

- **Expression of ξ:**
  Let us denote $A' = \mathbb{P}(Z_{c,t-1} = z_{c,t-1}, Z_{c,t} = z_{c,t}, Y_{c,0:M} = y_{c,0:M}|\theta^{(m)})$. We can decompose A' as follows:

$$\begin{aligned}
A' &= \mathbb{P}(Y_{c,0:t-2} = y_{c,0:t-2}, Y_{c,t:M} = y_{c,t:M}|Z_{c,t-1} = z_{c,t-1}, Z_{c,t} = z_{c,t}, Y_{c,t-1} = y_{c,t-1}, \theta^{(m)}) \\
&\quad \times \mathbb{P}(Z_{c,t-1} = z_{c,t-1}, Z_{c,t} = z_{c,t}, Y_{c,t-1} = y_{c,t-1}|\theta^{(m)}) \\
&= \mathbb{P}(Y_{c,0:t-2} = y_{c,0:t-2}|Y_{c,t-1} = y_{c,t-1}, Z_{c,t-1} = z_{c,t-1}, \theta^{(m)})\mathbb{P}(Z_{c,t-1} = z_{c,t-1}, Y_{c,t-1} = y_{c,t-1}|\theta^{(m)}) \\
&\quad \times \mathbb{P}(Y_{c,t:M} = y_{c,t:M}|Z_{c,t} = z_{c,t}, \theta^{(m)})\mathbb{P}(Z_{c,t} = z_{c,t}|Z_{c,t-1} = z_{c,t-1}, Y_{c,t-1} = y_{c,t-1}, \theta^{(m)}) \\
&= \mathbb{P}(Y_{c,0:t-1} = y_{c,0:t-1}, Z_{c,t-1} = z_{c,t-1}|\theta^{(m)})\mathbb{P}(Y_{c,t:M} = y_{c,t:M}|Z_{c,t} = z_{c,t}, \theta^{(m)}) \times P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t}) \\
&= \alpha_{c,t-1}^{(m)}(z_{c,t-1})P_{y_{c,t-1}}^{(m)}(z_{c,t-1}, z_{c,t})\mathbb{P}(Y_{c,t} = y_{c,t}|Z_{c,t} = z_{c,t}, \theta^{(m)}) \\
&\quad \times\mathbb{P}(Y_{c,t+1:M} = y_{c,t+1:M}|Z_{c,t} = z_{c,t}, Y_{c,t} = y_{c,t}, \theta^{(m)}) \\
&= \alpha_{c,t-1}^{(m)}(z_{c,t-1})P_{y_{c,t-1}}^{(m)}(z_{c,t-1}, z_{c,t})R^{(m)}(z_{c,t}, y_{c,t})\beta_{c,t}^{(m)}(z_{c,t}).
\end{aligned}$$

  Finally, we obtain :

$$\xi_{c,t}^{(m)}(z_{c,t-1}, z) = \frac{\alpha_{c,t-1}^{(m)}(z_{c,t-1})P_{y_{c,t-1}}^{(m)}(z_{c,t-1}, z)R^{(m)}(z, y_{c,t})\beta_{c,t}^{(m)}(z)}{\sum\limits_{z\in\Omega_Z} \alpha_{c,t}^{(m)}(z)\beta_{c,t}^{(m)}(z)}.$$

B.4 Step M - Solving the maximization problem

We want to solve the following problem of maximization :

$$\theta^{(m+1)} = \arg\max_{\theta} Q(\theta|\theta^{(m)}),$$

under the following constraints:

- $\sum\limits_{z_t\in\Omega_Z} P_{y_{c,t-1}}(z_{t-1}, z_t) = 1$;
- $\sum\limits_{y_t\in\Omega_Y} R(z_{c,t}, y_t) = 1$.

(We recall that we do not estimate $\pi()$.)

B.4.1 Writing the Lagrangian of the problem

We write the Lagrangian of the problem :

$$\begin{aligned}
\mathcal{L} &= \sum_{c=1}^{C} \sum_{z_{c,0}\in\Omega_Z} \ln(\pi(z_{c,0}))\,\mathbb{P}(Z_{c,0} = z_{c,0}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) - \eta_1(\sum_{z_0\in\Omega_Z} \pi(z_0) - 1) \\
&+ \sum_{c=1}^{C}\sum_{t=1}^{M} \sum_{(z_{c,t}, z_{c,t-1})\in\Omega_Z^2} \ln\left(P_{y_{c,t-1}}(z_{c,t-1}, z_{c,t})\right)\mathbb{P}(Z_{c,t-1} = z_{c,t-1}, Z_{c,t} = z_{c,t}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \\
&- \eta_2(y_{c,t-1}, z_{c,t-1})(\sum_{z'\in\Omega_Z} P_{y_{c,t-1}}(z_{c,t-1}, z') - 1) \\
&+ \sum_{c=1}^{C}\sum_{t=0}^{M} \sum_{z_{c,t}\in\Omega_Z} \ln(R(z_{c,t}, y_{c,t}))\,\mathbb{P}(Z_{c,t} = z_{c,t}|Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) - \eta_3(z_{c,t})(\sum_{y\in\Omega_Y} R(z_{c,t}, y) - 1)
\end{aligned}$$

B.4.2 Resolution

- **For $P_y^{(m+1)}(z, z')$:**

$$\frac{\partial \mathcal{L}}{\partial P_y(z,z')} = 0 \Leftrightarrow \frac{\sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y=y_{c,t-1})}}{P_y(z,z')} = \eta_2(y,z)$$

$$\Leftrightarrow \sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t-1}=y)} = \eta_2(y,z) \times P_y(z,z')$$

$$\Rightarrow \sum_{c=1}^{C}\sum_{t=1}^{M}\sum_{z'\in\Omega_Z} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t-1}=y)} = \eta_2(y,z) \times \sum_{z'\in\Omega_Z} P_y(z,z')$$

$$\Leftrightarrow \sum_{c=1}^{C}\sum_{t=1}^{M}\sum_{z'\in\Omega_Z} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t-1}=y)} = \eta_2(y,z)$$

So, we obtain :

$$P_y^{(m+1)}(z,z') = \frac{\sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t-1}=y)}}{\sum_{c=1}^{C}\sum_{t=1}^{M}\sum_{z'\in\Omega_Z} \mathbb{P}(Z_{c,t-1} = z, Z_{c,t} = z' | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t-1}=y)}}.$$

- **For $R(z,y)^{(m+1)}$:**

$$\frac{\partial \mathcal{L}}{\partial R(z,y)} = 0 \Leftrightarrow \frac{\sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t}=y)}}{R(z,y)} = \eta_3(z)$$

$$\Leftrightarrow \sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t}=y)} = \eta_3(z) R(z,y)$$

$$\Rightarrow \sum_{c=1}^{C}\sum_{t=1}^{M}\sum_{y\in\Omega_Y} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t}=y)} = \eta_3(z) \sum_{y\in\Omega_Y} R(z,y)$$

$$\Leftrightarrow \sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \sum_{y\in\Omega_Y} \mathbb{1}_{(y_{c,t}=y)} = \eta_3(z)$$

$$\Leftrightarrow \sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) = \eta_3(z)$$

So, we have :

$$R(z,y)^{(m+1)} = \frac{\sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | y_{c,0:M} = y_{c,0:M}, \theta^{(m)}) \mathbb{1}_{(y_{c,t}=y)}}{\sum_{c=1}^{C}\sum_{t=1}^{M} \mathbb{P}(Z_{c,t} = z | Y_{c,0:M} = y_{c,0:M}, \theta^{(m)})}.$$

References

Abdullah, K., and J.-V.-D. Hoek. 2022. “The Construction and Estimation of Hidden Semi-Markov Models.” *Journal of stochastic analysis* 3 (2).

Ailliot, P., and F. Pène. 2015. “Consistency of the maximum likelihood estimate for non-homogeneous Markov-switching models.” *ESAIM: PS* 19:268–292.

Allman, E., C. Matias, and J. Rhodes. 2009. “Identifiability of parameters in latent structure models with many observed variables.” *Annals of Statistics* 37 (6A):3099–3132.

Barbu, V-S., and N. Limnios. 2008. *Semi-Markov Chains and Hidden Semi-Markov Models towards Applications.* Springer.

Baum, L.-E., and T. Petrie. 1966. “Statistical Inference for Probabilistic Functions of Finite State Markov Chains.” *The Annals of Mathematical Statistics* 37 (6): 1554–1563.

Bazzi, M., F. Blasques, S.-J. Koopman, and A. Lucas. 2022. “Time-Varying Transition Probabilities for Markov Regime Switching Models.” *Journal of Time Series Analysis* 38 (3): 458–478.

Bickel, P.-J., Y. Ritov, and T. Rydén. 1998. “Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models.” *The Annals of Statistics* 26 (4): 21614–1635.

Bulla, J. 2006. *Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series.*

Cappé, O., E. Moulines, and T. Rydén. 2005. *Inference in Hidden Markov Models.* Springer.

Durand, J.-B., and O. Gaudoin. 2016. “Software reliability modelling and prediction with hidden Markov chains.” *Statiscal Modelling* 5 (1): 75–93.

Engel, C., and S. C. Hakkio. 1996. “The Distribution of Exchange Rates in the EMS.” *International Journal of Finance and Economics* 1:55–67.

Engel, C., and J.-D. Hamilton. 1990. “Long Swings in the Dollar: Are They in the Data and Do Markets Know It?” *The American Economic Review* 80 (4): 689–713.

Gales, M., and S. Young. 2008. *The application of Hidden Markov Models in Speech Recognition.* Now Foundations / Trends.

Gámiz, M.-L., N. Limnios, and M.-D.-C. Segovia-García. 2023. “Hidden Markov models in reliability and maintenance.” *European Journal of Operational Research* 304 (3): 1242–1255.

Hamilton, J.-D. 1989. “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.” *Econometrica* 57 (2): 357–384.

Le Coz, S., P. O. Cheptou, and N. Peyrard. 2019. “A spatial Markovian framework for estimating regional and local dynamics of annual plants with dormancy.” *Theoretical Population Biology* 127:120–132.

Le Strat, Y., and F. Carrat. 1999. “Monitoring epidemiologic surveillance data using hidden Markov models.” *Statistics in Medicine* 18 (24): 3377–3513.

Leos-Barajas, V., T. Photopoulou, R. Langrock, T.-A. Patterson, Y.-Y. Watanabe, M. Murgatroyd, and Y.-P. Papastamatiou. 2017. “Analysis of animal accelerometer data using hidden Markov models.” *Methods in Ecology and Evolution* 8 (2): 161–173.

Leroux, B.-G. 1992. “Maximum-likelihood estimation for hidden Markov models.” *Stochastic Processes and their applications* 40 (1): 127–143.

McClintock, B.-T., R. Langrock, O. Gimenez, E. Cam, D.-L. Borchers, R. Glennie, and T.-A. Patterson. 2020. "Uncovering ecological state dynamics with hidden Markov models." *Ecology Letters* 23 (12): 1878–1903.

Pluntz, M., S. Le Coz, N. Peyrard, R. Pradel, R. Coquet, 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* 21:1311–1318.

Pouzo, D., Z. Psaradakis, and M. Sola. 2022. "Maximum Likelihood Estimation in Markov Regime-Switching Models with Covariate-Dependent Transition Probabilities." *Econometrica : Journal of the econometric society* 90:1681–1710.

Raftery, A.-E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–163.

Yoon, B.-J. 2009. "Hidden Markov Models and their Applications in Biological Sequence." *Current Genomics* 10 (6): 402–415.

Yu, S-Z. 2016. *Hidden Semi-Markov Models Theory, Algorithms and Applications.* Elsevier.