Hidden Markov Models for wind farm power output

Citation for published version (APA):

DOI:
10.1109/TSTE.2018.2834475

Document status and date:
Published: 08/05/2018

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us:
openaccess@tue.nl
providing details. We will immediately remove access to the work pending the investigation of your claim.
I. Scatterplots of Wind Turbine Power Output

The power output between the various wind turbines located in the same wind park (or in geographic proximity) is highly dependent. As evidence to this dependency structure, we plot the scatterplots of the power outputs of two wind turbines. In Fig. [4], we present the scatterplots of the power output of wind turbine 6 and those of wind turbines 1-5, respectively.

II. General Description of HMM

A hidden Markov model can be viewed as a tool for representing probability distributions over sequences of observations. Let $Y_t$ denote the observation at time $t$, $t = 1, 2, \ldots$ (we assume here the discrete time case), with state-space $\mathcal{N}$. The HMM gets its name from two defining properties: First, it is assumed that the state of the hidden process satisfies the Markov property. More concretely, under the standard HMM formalism, for a given set of observations $(y_1, \ldots, y_T)$, the stochastic processes $\{X_t, t = 1, 2, \ldots\}$ and $\{Y_t, t = 1, 2, \ldots\}$ are governed by the one step transition probability matrix $A$ and the emission distribution $B$ (also known as output probabilities):

$$
\begin{align*}
A_{i,j} &= \mathbb{P}(X_{t+1} = j|X_t = i), \\
B_i(y_t) &= \mathbb{P}(Y_t = y_t|X_t = i),
\end{align*}
$$

$\forall i, j \in \mathcal{M}$. For $t = 1$, the initial distribution of $\{X_t, t = 1, 2, \ldots\}$ is given by $\pi_i = \mathbb{P}(X_1 = i)$. A schematic diagram showing the dependence of the random variables $\{(X_t, Y_t), t = 1, 2, \ldots\}$ of the HMM described by (1) is presented in Fig. [2]. However, the structure of the emission distribution $B$ varies depending on the problem under consideration. Note that HMMs can also be viewed as a particular kind of linear neural networks. [3] page 312.

III. Expectation-Maximization (EM) Algorithm

The EM algorithm is a general method for finding the maximum-likelihood parameter estimates of an underlying distribution given a data set which has incomplete values (see [22], [24], [26], [28] references of main text). Given the sequence of observations for all the wind turbine power outputs, $y = [y_1, \ldots, y_W]$ where $y_w = [y_1^w, \ldots, y_T^w] = [P_1^w, \ldots, P_T^w]$, our goal is to estimate the parameters of the model, $\lambda$. In order to do so we will maximize the joint likelihood function for the model given the complete data set of the output processes and the hidden process.

A. Likelihood Function

For a given model $\lambda$, the joint likelihood function for the model given the complete data set is the joint probability distribution for the observation and the hidden state sequences

$$
\mathcal{L}(\lambda|y, x) = \mathbb{P}(Y = y, X = x|\lambda)
$$

$$
= \pi_x \prod_{w=1}^{W} \rho_{y} \prod_{t=2}^{T} (A_{x_{t-1}, x_t} \prod_{w=1}^{W} P_{y_{t-1}, y_t}).
$$

The above equation can be factorized as a result of the Markov property, cf. (3) of the main text. The EM algorithm for parameter estimation of HMM given a set of observations is known as the Baum-Welch (BW) algorithm ([27] in the reference of main text). The EM algorithm first finds the expected value of the complete data set log-likelihood with respect to the hidden data set $X = x$ given the observed data $Y = y$ and the current parameter estimates, in the expectation-step, i.e.,

$$
Q(\lambda, \lambda_k) = \mathbb{E}[\mathbb{P}(Y = y, X = x|\lambda)|Y = y, \lambda_k]
$$

$$
= \sum_{x \in \mathcal{M}^T} \log(\mathbb{P}(Y = y, X = x|\lambda))|\mathbb{P}(Y = y, X = x|\lambda_k),
$$

where $\lambda_k$ is the current set of parameters estimates used to calculate the expectation $Q$. A detailed discussion of expression (3) is given in reference [28] in the main text.
Fig. 1: Scatter plots of power output of wind turbine 6 with wind turbines 1-5

The **maximization-step** determines the next iterate $\lambda_{k+1}$ by maximizing the expectation $Q$, i.e.

$$\lambda_{k+1} = \arg\max_{\lambda} Q(\lambda, \lambda_k). \quad (4)$$

The maximization step guarantees that $L(\lambda_{k+1}|y, x) \geq L(\lambda_k|y, x)$. The expectation and maximization steps are repeated until the desired convergence is reached. For literature on the convergence of the EM algorithm, see [29], [30] of main text.

**B. Parameter estimation**

In this section we present expressions for the parameter estimates $\lambda$ given the observed data $y$ and the hidden sequence $x = [x_1, \ldots, x_T]$. For convenience we will denote $\lambda_k$, the old parameter set, as $\lambda'$ and the next iterate of the parameter set $\lambda_{k+1}$ as $\lambda$. First we find the expectation function using (2), the expectation function (3) can be expanded as

$$Q(\lambda, \lambda') = \sum_{x \in \mathcal{M}^T} (\log \pi_{x_1}) \mathbb{P}(Y = y, X = x|\lambda')$$

$$+ \sum_{w=1}^W \left[ \sum_{x \in \mathcal{M}^T} \left( \log \rho^{w,x}_{y_1,t} \right) \mathbb{P}(Y = y, X = x|\lambda') \right]$$

Fig. 2: Schematic diagram of the HMM described in [1]
\[\sum_{x \in \mathcal{M}^T} \left( \sum_{t=2}^{T} \log A_{x_{t-1},x_t} \right) \mathbb{P}(Y = y, X = x|\lambda') + \sum_{w=1}^{W} \left[ \sum_{x \in \mathcal{M}^T} \left( \sum_{t=2}^{T} \log L_{y_{t-1},y_t}^{w,x_t} \right) \right] \mathbb{P}(Y = y, X = x|\lambda').\]

Since the parameters we wish to optimize can be independently factorized into explicit terms as shown in the equation above, we can optimize each term individually using Lagrange multipliers.

**a) First term, \(\pi\):** The first term of (5) can be written as
\[
\sum_{x \in \mathcal{M}^T} \left( \sum_{i=1}^{M} \log \pi_i \right) \mathbb{P}(Y = y, X_1 = i|\lambda') = \sum_{i=1}^{M} \mathbb{P}(Y = y, X_1 = i|\lambda') - \gamma \sum_{i=1}^{M} \mathbb{P}(Y = y, X_1 = i|\lambda').
\]

Using the Lagrange multiplier \(\gamma\) and the constraint that \(\sum_{i=1}^{M} \pi_i = 1\), we set the derivative equal to zero, i.e.,
\[
\frac{\partial}{\partial \pi_i} \left( \sum_{i=1}^{M} \log \pi_i \right) \mathbb{P}(Y = y, X_1 = i|\lambda') - \gamma \left( \sum_{i=1}^{M} \mathbb{P}(Y = y, X_1 = i|\lambda') - 1 \right) = 0.
\]

Taking the derivative and using \(\sum_{i=1}^{M} \pi_i = 1\) we get \(\gamma = \mathbb{P}(Y = y|\lambda')\) and for \(\pi_i\) we get,
\[
\pi_i = \frac{\mathbb{P}(Y = y, X_1 = i|\lambda')}{\mathbb{P}(Y = y|\lambda')}.\]

**b) Second term, \(\rho^{w,i}\):** The second term of (5) has a sum in \(w\). We solve for the \(w^{th}\) term inside the sum,
\[
\sum_{x \in \mathcal{M}^T} \left( \log \rho_{y_1}^{w,x_1} \right) \mathbb{P}(Y = y, X = x|\lambda') = \sum_{i=1}^{M} \left( \log \rho_{y_1}^{w,i} \right) \mathbb{P}(Y = y, X_1 = i|\lambda').
\]

We have \(M\) constraint equations \(\sum_{\mu=1}^{\mu} \rho_{\mu}^{w,i} = 1\), as \(i \in \mathcal{M}\). Hence we have \(M\) Lagrange multipliers. Setting the derivative to zero and using \(\sum_{\mu=1}^{\mu} \rho_{\mu}^{w,i} = 1\), we get
\[
\rho_{\mu}^{w,i} = \frac{1(\mu = \mu)}{\mathbb{P}(Y = y, X_1 = i|\lambda')} \mathbb{P}(Y = y, X_1 = i|\lambda') = 1(\mu = \mu).
\]

**c) Third term, \(A\):** The third term of (5) can be written as
\[
\sum_{x \in \mathcal{M}^T} \left( \sum_{t=2}^{T} \log A_{x_{t-1},x_t} \right) \mathbb{P}(Y = y, X = x|\lambda') = \sum_{t=2}^{T} \sum_{i=1}^{M} \mathbb{P}(Y = y, X_{t-1} = i, X_t = j|\lambda').
\]

We have \(M\) constraint equations \(\sum_{k=1}^{M} a_{ik} = 1\), as \(l \in \mathcal{M}\). Hence we need \(M\) Lagrange multipliers. Setting the derivative to zero and using \(\sum_{k=1}^{M} a_{ik} = 1\), we get
\[
A_{ij} = \frac{\sum_{t=2}^{T} \mathbb{P}(Y = y, X_{t-1} = i, X_t = j|\lambda')}{\sum_{t=2}^{T} \mathbb{P}(Y = y, X_{t-1} = i|\lambda')}.
\]

**d) Fourth term, \(L^{w,i}\):** We solve for the \(w^{th}\) term inside the sum of the fourth term of (5),
\[
\sum_{x \in \mathcal{M}^T} \left( \sum_{t=2}^{T} \log L_{y_{t-1},y_t}^{w,x_t} \right) \mathbb{P}(Y = y, X = x|\lambda') = \sum_{t=2}^{T} \sum_{i=1}^{M} \left( \log L_{y_{t-1},y_t}^{w,i} \right) \mathbb{P}(Y = y, X_t = i|\lambda').
\]

Solving for \(L_{\mu}^{w,i}\) using the Lagrange multipliers we get,
\[
L_{\mu}^{w,i} = \sum_{t=2}^{T} \mathbb{P}(Y_{t-1}^{w} = \mu) \mathbb{P}(Y_t^{w} = \nu) \mathbb{P}(Y = y, X_t = i|\lambda').
\]

Hence, the parameter estimates are given by
\[
\left\{ \begin{array}{l}
\hat{\pi}_i = \frac{\mathbb{P}(Y = y, X_1 = i|\lambda')}{\mathbb{P}(Y = y|\lambda')}, \\
\hat{A}_{ij} = \frac{\sum_{t=2}^{T} \mathbb{P}(Y = y, X_{t-1} = i, X_t = j|\lambda')}{\sum_{t=2}^{T} \mathbb{P}(Y = y, X_{t-1} = i|\lambda')}, \\
\hat{\rho}_{\mu}^{w,i} = 1(\mu = \mu), \\
\hat{L}_{\mu}^{w,i} = \frac{\sum_{t=2}^{T} \mathbb{P}(Y_{t-1}^{w} = \mu, Y_t^{w} = \nu) \mathbb{P}(Y = y, X_t = i|\lambda')}{\sum_{t=2}^{T} \mathbb{P}(Y_{t-1}^{w} = \mu) \mathbb{P}(Y = y, X_t = i|\lambda')}.
\end{array} \right.
\]

**1) Forward-backward variables:** For calculating the estimates in (12) it is convenient to introduce the so-called forward backward variables \(\alpha_{i}(t), \beta_{i}(t), \forall i \in \mathcal{M}\) ([24], [30]). For lack of space we have dropped the random variable \(Y\) terms in front of the \(y\) terms,
\[
\begin{align*}
\alpha_i(t) &= \mathbb{P}(y_1, \ldots, y_{t-1}, y_t^W, X_t = i | Y_{t-1}), \\
\beta_i(t) &= \mathbb{P}(y_1, \ldots, y_{t-1}, y_t^W | Y_{t-1}, X_t = i).
\end{align*}
\]

These variables are computed recursively and the numerical effort grows linearly in \( T \).

a) Forward variable recursion: For simplicity we drop the \( X' \) term for now. From (13) we have for \( t = 1 \),
\[
\alpha_i(1) = \mathbb{P}(y_1, y_2^W, X_1 = i) = \mathbb{P}(y_1 | y_2^W, X_1 = i) \cdot \mathbb{P}(y_2^W, X_1 = i) = \rho_{y_1} \cdot \rho_{y_2^W} \cdot \pi_i.
\]

For deriving the recursion relation, we have
\[
\alpha_i(t) = \mathbb{P}(y_1, \ldots, y_t^W | X_t = i) = \sum_{j=1}^{M} \mathbb{P}(y_1, \ldots, y_t^W | X_t = j, X_{t-1} = i) = \sum_{j=1}^{M} \mathbb{P}(y_1, y_i, \ldots, y_{t-1}^W, y_t^W | X_t = j, X_{t-1} = i).
\]

b) Backward equation: From (13) we have, \( \beta_i(T) = 1 \). We have,
\[
\beta_i(t) = \mathbb{P}(y_{t+1}^W, \ldots, y_T^W | X_t = i, y_1^W, \ldots, y_t^W) = \sum_{\{x_{t+1}, \ldots, x_T\} \in \mathcal{MT}^{-t}} \mathbb{P}(y_{t+1}^W, \ldots, y_T^W | x_{t+1}, \ldots, x_T = x_T | X_t = i, y_1^W, \ldots, y_t^W).
\]

2) Some expressions in terms of forward-backward variables: In order to re-write (12) in terms of forward-backward equations, we first write the following in terms of the forward-backwards equations. Note for lack of space we are dropping the random variable \( Y' \) term in front of the \( y \) terms.
\[
\begin{align*}
\mathbb{P}(Y = y, X_t = i) &= \mathbb{P}(y_1, \ldots, y_T, y_t^W, y_1^W, \ldots, y_T^W | X_t = i) = \alpha_i(t) \beta_i(t) \cdot \mathbb{P}(X_t = j | Y_t = y) \cdot \mathbb{P}(Y_t = y | X_t = i).
\end{align*}
\]

3) Parameter estimates in terms of forward-backward variables: Using (16) and (17), the expressions in (12) in terms of forward-backward equations becomes
\[
\begin{align*}
\hat{c}_i &= \sum_{t=1}^{T} \alpha_i(1) \beta_i(t), \\
\hat{A}_{ij} &= \sum_{t=1}^{T} \alpha_i(t-1) \hat{A}_{ij} \sum_{u=1}^{W} \hat{L}_{ij}^{uw} y_{t-1}^w \beta_i(t), \\
\hat{L}_{ij}^{uw} &= \sum_{t=1}^{T} \alpha_i(t-1) \beta_i(t) (Y_{t-1}^{uw} = \mu) (Y_t^{uw} = \nu).
\end{align*}
\]

Re-scaling forward backward equations: From (13) we see that as \( t \) increases the values of \( \alpha_i(t) \) and \( \beta_i(t) \) become very small. Hence, the terms in (18) diverge when computed numerically. To avoid this numerical problem we normalize the values of the forward and backward equations.
two simultaneous stopping criteria

4) Stopping criterion for the EM algorithm: We enforce two simultaneous stopping criteria for the EM algorithm:

1) The number of iterations, $n_t$, exceeds a predefined threshold value, $n_{\text{max}}$, i.e., $n_t \geq n_{\text{max}}$.

2) The improvements in $\lambda$, i.e., $\Delta \lambda_{\text{min}} = \max_{n_m} |\lambda^{n_t} - \lambda^{n_{t-1}}| \leq \epsilon$, where $\lambda^{n_t}$ is $\lambda$ for the iteration $n_t$.

If any one of the stopping criteria is true the algorithm stops. We take $n_{\text{max}} = 10^4$, $n_m = 100$ and $\epsilon = 10^{-6}$ for our simulations.

IV. EDF COMPARISON FIGURE

![Fig. 3: Comparing EDFs of the total power produced by the wind park for data sets $P_w^w$, $\tilde{P}_w^w$ and $\tilde{P}_{HMM}^w$, $\forall w \in W$ for $N = 5, M = 9$.](image)

V. HMM COMPLEXITY AND IMPROVEMENTS

The problem of inferring the probability of a state sequence, i.e., finding the stochastic state sequence (based on a MC or a HMM), is most likely to explain a given observation sequence (time series of data), can be solved by using the well-known Viterbi algorithm based on dynamic programming, see [1] and the references therein. The problem of learning the underlying model (MC or HMM), i.e., the corresponding one step probability transition matrix, which are referred to as the model parameters, is solved by iteratively adjusting the model parameters to optimize the model fitting according to some criterion, e.g., the maximum-likelihood (ML). For HMMs, the iterative parameter estimation is typically performed using an expectation-maximization (EM) algorithm with the ML criterion. In the context of HMM, the EM algorithm becomes the Baum-Welch algorithm, [5]. Given a training data set, the EM algorithm iteratively estimates the HMM parameters in two stages; an expectation step (E-step) followed by a maximization step (M-step). In order to reduce the number of computations required for this process, one of the most common approaches proposed uses Viterbi training [2], [3].

The EM algorithm is an iterative scheme that is well-defined and is numerically stable, but convergence may require a large number of iterations, and may very well lead to local optima. Therefore, starting with a (any) set of parameters and iteratively re-estimating the parameters, one can improve the likelihood function until some limiting point is reached. For this reason, the choice of the starting set of parameters can be of crucial importance. Furthermore, the number of states should be specified in advance. In fact, learning the model topology is a difficult task, see [1] and the references therein. In [5], the authors describe the EM algorithm and they compare it with other relevant algorithmic approaches. Furthermore, they suggest some improvements and variations to deal with both the off-line and the on-line case. Finally, in the mathematical appendix of [5], the authors sketch the mathematical details of the algorithms.

A. HMMs with second order Markov properties

HMMs have been generalized to also cover “higher order Markov properties”, see, e.g., [4] and the references therein. Thus, one could possibly extend the first order Markov models used in the paper to capture second or higher orders Markov properties. However, this would be out of the scope of the study we performed as the tail distribution is already very accurately captured and an extension to a higher order Markov property would result in an increase to the state-space HMM description and thus an increase in the computational complexity of the model.

VI. PARAMETER ESTIMATES FOR SAME ONE-STEP TRANSITION MATRIX MODEL

In this case we assume that the individual wind turbine power output processes $Y^w$, are governed by matrices $(L^v)$ and $A$ as follows:

$$
\left\{
\begin{array}{l}
P_{\mu \nu}^v = \mathbb{P}(Y^w_{t+1} = \nu|Y^w_t = \mu, X_{t+1} = i) \\
A_{ij} = \mathbb{P}(X_{t+1} = j|X_t = i),
\end{array}
\right.
$$

The initial distribution of $Y^w_t$ is given by $\rho^v_\mu = \mathbb{P}(Y^w_1 = \mu|X_1 = i)$. The parameter estimates for this model are given by
with initialization $\hat{\pi}_i = \frac{\alpha_i(1)\beta_i(1)}{\sum_{i=1}^{M} \alpha_i(1)\beta_i(1)}, $
\[ \hat{A}_{ij} = \frac{\sum_{t=2}^{T} \alpha_i(t-1)\hat{A}_{ij}\left[ \prod_{w=1}^{W} \tilde{L}_{y_{t-w}}^{i} y_{t-w}^{w} \right] \beta_j(t)}{\sum_{t=2}^{T} \alpha_i(t-1)\beta_j(t-1)}, \]
\[ \hat{\rho}_{\mu} = \frac{1}{W} \sum_{w=1}^{W} \mathbb{1}(Y_{1}^{w} = \mu), \] and \[ \hat{L}_{\mu}^{i} = \frac{\sum_{t=2}^{T} \alpha_i(t-1)\beta_i(t) \sum_{w=1}^{W} \mathbb{1}(Y_{t-w}^{w} = \mu) \mathbb{1}(Y_{t-w}^{w} = \mu)}{\sum_{t=2}^{T} \alpha_i(t-1)\beta_i(t) \sum_{w=1}^{W} \mathbb{1}(Y_{t-w}^{w} = \mu)}. \] 

The recursions of the forward-backward equations are given by $\forall w \in \mathcal{W}, \forall i \in \mathcal{M}$
\[ \alpha_i(t) = \sum_{j=1}^{M} \alpha_j(t-1) A_{ij} \left[ \prod_{w=1}^{W} \tilde{L}_{y_{t-w}}^{i} y_{t-w}^{w} \right], \] and \[ \beta_i(t) = \sum_{j=1}^{M} A_{ij} \left[ \prod_{w=1}^{W} \tilde{L}_{y_{t-w}}^{i} y_{t-w}^{w} \right] \beta_j(t+1), \] with initialization $\alpha_i(1) = \hat{\pi}_i \cdot \prod_{w=1}^{W} \rho_{\mu w}^{i}$ and $\beta_i(T) = 1.$

\[ \text{VII. COMPARISON TABLES} \]
In Table I we compare $\gamma_{P}^{(N,M)}$ for different quantile fractions, $G^*$, with $\gamma_P$ for $N = 5$ and different number of hidden states $M$ for the model described in (3) of the main text. We tabulate the relative error of $\gamma_{P}^{(N,M)}$ and $\gamma_{P}$, $\text{RE} = \frac{\gamma_{P} - \gamma_{P}^{(N,M)}}{\gamma_{P}} \times 100$ (in %). We also tabulate $-\log \hat{L}$, AIC values, BIC values and the number of parameters in the model $p$ in the table in order to find the best fitted model.

In Table II we compare $\gamma_{P}^{(N,M)}$ for different quantile thresholds, $G^*$, with $\gamma_P$ for $N = 5$ and different number of hidden states $M$ for the model described in Section IV-B of the main text and by (20). We tabulate the relative error $\text{RE} = \frac{\gamma_{P} - \gamma_{P}^{(N,M)}}{\gamma_{P}} \times 100$ (in %), $-\log \hat{L}$, AIC & BIC values and the number of parameters in the model $p$ in the table.

In Table III we compare the quantile fraction values for different number of hidden states $M$ for the macroscopic approach described in Section V of main text. Given the observation of the binned measurement total power produced by the wind farm data $\hat{G} = \{\hat{G}_{1}, \ldots, \hat{G}_{T}\}$ we first estimate the HMM model parameters. We generate surrogate data series of the total power produced by the wind farm, $\tilde{G}$ from the model parameters. $\gamma_{\tilde{G}}^{(N,M)} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(\hat{G}_{t} > G^*)$ is the fraction of time the total power produced by the wind farm obtained from HMM is greater than $G^*$ threshold. We generate 100 realizations of $\hat{G}$ of length $T = 10^4$. Then we calculate the mean and the standard deviations of $\gamma_{\tilde{G}}^{(N,M)}$ from the 100 realizations, $\gamma_{\tilde{G}}^{(N,M)}$ and $\sigma(\gamma_{\tilde{G}}^{(N,M)})$ respectively.

We also tabulate the relative error $\text{RE} = \frac{\left| \gamma_{\tilde{G}}^{(N,M)} - \gamma_{\hat{G}}^{(N,M)} \right|}{\gamma_{\hat{G}}^{(N,M)}} \times 100$ (in %). In order to find the best model fit we also compare the values of $-\log \hat{L}$, AIC and BIC. Note that the simplest discrete time Markov model, i.e. $M = 1$, can capture the quantile fractions very well, however the values of $-\log \hat{L}$, AIC and BIC are much higher compared to the $M \geq 2$ cases.

\[ \text{REFERENCES} \]
TABLE I: Comparing $\gamma_P^{(N,M)}$ for different quantile thresholds, $G^*$, with $\gamma_P$ for $N = 5$ for different number of hidden states $M$. 

<table>
<thead>
<tr>
<th>$M$</th>
<th>Quanlize</th>
<th>$\gamma_P^{(5,M)} \pm \sigma(\gamma_P^{(5,M)})$</th>
<th>RE (%)</th>
<th>$- \log \hat{L} \times 10^{-4}$</th>
<th>AIC $\times 10^{-4}$</th>
<th>BIC $\times 10^{-4}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90%</td>
<td>$6.67 \times 10^{-3} \pm 4.23 \times 10^{-5}$</td>
<td>93.33</td>
<td>3.4465</td>
<td>6.9218</td>
<td>7.0236</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$3.01 \times 10^{-4} \pm 7.12 \times 10^{-6}$</td>
<td>99.39</td>
<td>2.5176</td>
<td>5.1592</td>
<td>5.5974</td>
<td>620</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$1.70 \times 10^{-8} \pm 2.14 \times 10^{-8}$</td>
<td>99.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>90%</td>
<td>$0.1248 \pm 0.0048$</td>
<td>24.856</td>
<td>6.1418</td>
<td>5.0822</td>
<td>5.6123</td>
<td>750</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0573 \pm 0.0028$</td>
<td>14.465</td>
<td>30.850</td>
<td>5.1592</td>
<td>5.5974</td>
<td>620</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0069 \pm 0.0004$</td>
<td>9.528</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90%</td>
<td>$0.1022 \pm 0.0031$</td>
<td>2.2513</td>
<td>65.270</td>
<td>5.1592</td>
<td>5.6123</td>
<td>750</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0470 \pm 0.0018$</td>
<td>6.1418</td>
<td>2.4661</td>
<td>5.0822</td>
<td>5.6123</td>
<td>750</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0035 \pm 0.0002$</td>
<td>30.850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>90%</td>
<td>$0.0993 \pm 0.0030$</td>
<td>6.857</td>
<td>55.421</td>
<td>5.1592</td>
<td>5.6123</td>
<td>750</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0468 \pm 0.0020$</td>
<td>6.5061</td>
<td>2.4534</td>
<td>5.1592</td>
<td>5.6123</td>
<td>750</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0045 \pm 0.0004$</td>
<td>9.528</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>90%</td>
<td>$0.1013 \pm 0.0036$</td>
<td>1.3336</td>
<td>18.232</td>
<td>5.8016</td>
<td>5.9404</td>
<td>1215</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0547 \pm 0.0023$</td>
<td>9.3667</td>
<td>2.4193</td>
<td>5.8016</td>
<td>5.9404</td>
<td>1215</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0082 \pm 0.0005$</td>
<td>1.3336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>90%</td>
<td>$0.1011 \pm 0.0032$</td>
<td>1.1512</td>
<td>13.127</td>
<td>5.0776</td>
<td>6.0500</td>
<td>1376</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0517 \pm 0.0020$</td>
<td>3.2605</td>
<td>2.4011</td>
<td>5.0776</td>
<td>6.0500</td>
<td>1376</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0114 \pm 0.0060$</td>
<td>13.127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>90%</td>
<td>$0.1268 \pm 0.0050$</td>
<td>26.887</td>
<td>24.684</td>
<td>5.1312</td>
<td>6.2190</td>
<td>1539</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0633 \pm 0.0028$</td>
<td>26.496</td>
<td>2.4117</td>
<td>5.1312</td>
<td>6.2190</td>
<td>1539</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0076 \pm 0.0005$</td>
<td>24.684</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II: Comparing $\gamma_P^{(N,M)}$ for different quantile thresholds, $G^*$, with $\gamma_P$ for $N = 5$ for different number of hidden states $M$ along with $- \log \hat{L}$, AIC and BIC values. 

<table>
<thead>
<tr>
<th>$M$</th>
<th>Quanlize</th>
<th>$\gamma_P^{(5,M)} \pm \sigma(\gamma_P^{(5,M)})$</th>
<th>RE (%)</th>
<th>$- \log \hat{L} \times 10^{-4}$</th>
<th>AIC $\times 10^{-4}$</th>
<th>BIC $\times 10^{-4}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>90%</td>
<td>$0.0924 \pm 0.0027$</td>
<td>7.532</td>
<td>2.5421</td>
<td>5.1081</td>
<td>5.1929</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0477 \pm 0.0017$</td>
<td>4.613</td>
<td>52.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0048 \pm 0.0003$</td>
<td>52.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90%</td>
<td>$0.0983 \pm 0.0030$</td>
<td>1.707</td>
<td>2.5087</td>
<td>5.0975</td>
<td>5.1535</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0494 \pm 0.0019$</td>
<td>1.290</td>
<td>52.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0048 \pm 0.0004$</td>
<td>52.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>90%</td>
<td>$0.1099 \pm 0.0031$</td>
<td>0.940</td>
<td>0.689</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0497 \pm 0.0018$</td>
<td>0.689</td>
<td>25.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0075 \pm 0.0006$</td>
<td>25.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>90%</td>
<td>$0.1025 \pm 0.0029$</td>
<td>2.332</td>
<td>2.4766</td>
<td>5.0041</td>
<td>5.1844</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0512 \pm 0.0019$</td>
<td>6.369</td>
<td>25.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0120 \pm 0.0009$</td>
<td>6.369</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>90%</td>
<td>$0.0987 \pm 0.0032$</td>
<td>1.243</td>
<td>3.391</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0517 \pm 0.0020$</td>
<td>3.391</td>
<td>2.4724</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0108 \pm 0.0060$</td>
<td>8.116</td>
<td>5.2132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>90%</td>
<td>$0.0984 \pm 0.0043$</td>
<td>1.562</td>
<td>2.4729</td>
<td>5.0038</td>
<td>5.2087</td>
<td>339</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>$0.0522 \pm 0.0027$</td>
<td>4.313</td>
<td>5.2087</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>$0.0076 \pm 0.0008$</td>
<td>1.728</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE III: Comparing $\gamma_G^{(N,M)}$ for different quantile thresholds, $G^*$, with $\gamma_P$ for $N = 15$ for different number of hidden states $M$ along with $-\log \hat{L}$, AIC, BIC and number of parameters of the model, $p$.

<table>
<thead>
<tr>
<th>$M$</th>
<th>Quantile</th>
<th>$\gamma_G^{(15,M)}$ $\pm \sigma(\gamma_G^{(15,M)})$</th>
<th>RE (%)</th>
<th>$-\log \hat{L} \times 10^{-4}$</th>
<th>AIC $\times 10^{-5}$</th>
<th>BIC $\times 10^{-5}$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90%</td>
<td>0.0984 $\pm$ 0.0042</td>
<td>1.569</td>
<td>6.2738</td>
<td>1.2590</td>
<td>1.2738</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0469 $\pm$ 0.0025</td>
<td>6.302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0096 $\pm$ 0.0008</td>
<td>4.216</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>90%</td>
<td>0.1067 $\pm$ 0.0101</td>
<td>6.717</td>
<td>1.0129</td>
<td>0.2116</td>
<td>0.2435</td>
<td>451</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0509 $\pm$ 0.0060</td>
<td>1.788</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0110 $\pm$ 0.0024</td>
<td>9.918</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>90%</td>
<td>0.1055 $\pm$ 0.0111</td>
<td>5.572</td>
<td>0.9983</td>
<td>0.2132</td>
<td>0.2613</td>
<td>680</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0512 $\pm$ 0.0073</td>
<td>2.409</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0107 $\pm$ 0.0021</td>
<td>6.488</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>90%</td>
<td>0.1173 $\pm$ 0.0128</td>
<td>17.37</td>
<td>0.9852</td>
<td>0.2152</td>
<td>0.2796</td>
<td>911</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0567 $\pm$ 0.0076</td>
<td>13.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0101 $\pm$ 0.0021</td>
<td>0.487</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>90%</td>
<td>0.0988 $\pm$ 0.0102</td>
<td>1.154</td>
<td>0.9761</td>
<td>0.2181</td>
<td>0.2989</td>
<td>1144</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0475 $\pm$ 0.0057</td>
<td>5.095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0101 $\pm$ 0.0022</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90%</td>
<td>0.1004 $\pm$ 0.0113</td>
<td>0.387</td>
<td>0.9793</td>
<td>0.2234</td>
<td>0.3209</td>
<td>1379</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0471 $\pm$ 0.0065</td>
<td>5.850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0096 $\pm$ 0.0020</td>
<td>4.597</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>90%</td>
<td>0.1026 $\pm$ 0.0099</td>
<td>2.594</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.0505 $\pm$ 0.0063</td>
<td>0.866</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>0.0113 $\pm$ 0.0025</td>
<td>13.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>