Data-Driven Modeling of the Temporal Evolution of Breakers' States in the French Electrical Transmission Grid

Mauricio Gonzalez, Antoine Girard

Working paper

Data-Driven Modeling of the Temporal Evolution of Breakers' States in the French Electrical Transmission Grid ⋆

Mauricio Gonzalez[1] [a] and Antoine Girard[2]

[1] Qivalio, Quivalio Analytics, R&D Team
23 rue de Liège, 75008 Paris, France
[2] Université Paris-Saclay, CNRS, CentraleSupélec, L2S
3 rue Joliot Curie, 91190 Gif-sur-Yvette, France

Oct 25, 2021

Abstract. The electrical transmission grid is a complex network whose configuration can be adapted using a set of circuit breakers. As the grid is becoming ever more complex and prone to failures or attacks, being able to monitor the evolution of the breakers' states to detect abnormal behaviors is of the highest importance in order to ensure a safe and secure operation of the grid. In this paper, we propose a data-driven approach to the modeling of the temporal evolution of the breakers' states in an electrical transmission grid. The available data consists of a historical record of the breakers' states over a period during which the electrical transmission grid operated normally. A stochastic modeling framework is introduced where we make the assumption that the state of the grid (given by the states of all breakers, which are known at all time) is a stochastic process which varies around a reference configuration, chosen among a finite set of reference configurations. The current reference configuration of the grid is unknown and constitutes a hidden stochastic process. Then, the evolution of the breakers' states is driven by a continuous-time finite-state hidden Markov model. Based on this framework, we provide a filter-based expectation-maximization approach using a change of probability measure method to estimate the model parameters and the hidden reference configuration. Further, we present a strong scheme with no discretization error on the filter dynamics for numerical purposes. We also present a clustering approach to identify the set of reference configurations of the network.

Keywords: Data-driven modeling, Hidden Markov Models, EM algorithms, Electrical transmission grid.

1 Introduction

The electrical transmission grid is an interconnected network that permits the electrical energy movements from producers, such as nuclear plants, to electrical substations. Central components of these networks are circuit breakers, which are electrical switches designed to interrupt or continue the electrical energy flow. Hence, the configuration of the grid is determined by the states of all breakers, which can be used to adapt the grid to diverse operating conditions. As the electrical transmission grid can be prone to failures or malicious attacks [1,2], it is of the highest importance to be able to monitor the temporal evolution of the breakers' states to detect abnormal behaviors in order to ensure safe and secure operation. However, the electrical grid is an overly complex system and developing models that can be embedded in online monitoring algorithms is a challenging task. In this paper, we propose a data-driven approach that allows us to derive stochastic models for the temporal evolution of the breakers' states in the electrical transmission grid. The available data consists of the historical record of the evolution of the breakers' states over a period during which the grid operated normally.

We propose a modeling framework based on the standing assumption that the grid configuration (given by the states of all breakers) is a stochastic process which varies around a reference configuration. The current reference configuration is itself a stochastic process that takes values in a finite set of reference configurations. While the breakers' states are known at all time, the current reference configuration is assumed to be unknown. Hence, in our framework, the temporal evolution of the breakers' state is driven by a continuous-time finite-state Hidden Markov Model (HMM) [3] that consists of a hidden Markov process (the reference configuration) and multiple observed Markov

⋆ This work has been funded by the RTE-CentraleSupélec Chair.

[a] *Previous address:* Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France.

processes (the breakers' states). Given the data of the temporal evolution of the breakers' states, the problem is then to find an estimate of the model parameters and of the hidden reference configuration. While our HMM is later intended to be included in online monitoring algorithms for the detection of abnormal behaviors (as e.g. in [4]), the current work focuses on the modeling and estimation problems.

We provide an iteratively filter-based Expectation-Maximization (EM) approach [5,6,7,8] to estimate the model parameters and the temporal evolution of the hidden reference configuration. This approach aims to maximize a log-likelihood function over a parameter space, where the parameters define the transition rate matrices of the hidden and observed state processes. First, we suppose that all state processes belong in a probability space that represents the "real world", and then, we use a change of probability measure technique (Girsanov's Theorem, see, e.g., [3,9]) to define a new probability measure to represent a "fictitious world". In this new space, filters for estimating the temporal evolution of the reference configuration and the model's parameters are easy to obtain. They are linear Stochastic Differential Equations (SDEs) modulated by counting processes. For numerical purposes, we present a strong scheme with no discretization error on the SDEs solution. This scheme adapts to all state change times generated by the temporal evolution of all breaker states.

The proposed modeling approach is finally confronted with available data consisting of Boolean temporal sequences describing the breakers' states (off/on) of the network. The data has been collected during the normal behavior of the network over a given period. The reference configurations of the network are also obtained from this data using a weighted version of the well-known K-means method [10]. We then identify the normal behavior of the French electrical grid, which will be embedded in a monitoring and detection algorithm in the future.

The main contributions of this paper are the following. First, we propose a general Hidden Markov Model for the temporal evolution of networks. While the targeted application is the Electrical transmission grid, we believe the model is general enough to serve for other types of dynamic networks. Second, we generalize the results of [11,12] because the proposed HMM is based on multiple-observed counting processes. Third, instead of using the classical Euler-Maruyama discretization (with small-time step) for all the SDEs obtained for the filter estimates (e.g., as in [12, 13,14]), we present a strong scheme with no discretization error. We obtain an exact solution of all SDEs between two-state change times, and the state change effects are then added at the correct state change times. Finally, we also propose a weighted version of the well-known K-means method to obtain the reference configurations of the grid. This method can be used for another kind of temporal data.

The rest of this paper is organized as follows. In Section 2, we introduce the finite-state continuous-time hidden Markov model dynamics for the temporal evolution of the breakers' states and the temporal evolution of the grid over the reference configurations. In Section 3, we provide a general filtering approach to estimate the grid's hidden temporal evolution based on all observed state processes of the breaker states. We also briefly recall the EM algorithm and compute a filter-based EM algorithm for all model parameters. In Section 4, we present a general numerical method for all filter estimates. We also show how to obtain the initial estimation of all parameters and the reference configurations of the grid using available data of all breakers states. In Section 5, some approaches are analyzed for detecting abnormal behavior on the grid based on its filter estimate. In Section 6, we show the numerical results for the temporal evolution of the breakers' states in the French electrical transmission grid, which illustrates a real-world application of the HMM considered in this paper. The final Section 7 gives concluding remarks.

2 The Modeling

Let's assume that the grid has finitely many reference configurations $h_1, ..., h_N$ known *a priori*, $N \in \mathbb{N}^*$. Suppose that a state process $H := \{H_t\}_{t\geq 0}$ represents the evolution over time of the grid between $h_1, ..., h_N$, where H_t denotes the unknown reference configuration at time $t \geq 0$ and then it constitutes the hidden process, i.e., H is not directly observable. Let $P \in \mathbb{N}^*$ the number of breakers in the grid. Suppose that we observe each breaker state's temporal evolution and is represented by the state process $Y^p := \{Y_t^p\}_{t\geq 0}$. Thus, Y_t^p denotes the breaker state $p = 1, ..., P$ at time $t \geq 0$ and constitutes the observable information of the grid. We aim in this paper to estimate the hidden evolution of the grid in some optimal way based on the temporal evolution of the breaker states Y^p, $p = 1, ..., P$.

In this paper, we fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the probability measure of the "real-world", and we denote by $\mathbb{E}$ the expectation operator under $\mathbb{P}$. We suppose that all state processes are continuous-time finite-state Markov chains defined on the common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is also assumed that (almost) all sample functions are right-continuous with left limits.

2.1 The State Processes

Consider that the state space of $H = \{H_t\}_{t\geq 0}$ is the set of known reference configurations of the grid defined by:

$$\mathbb{H} \;\coloneqq\; \big\{h_1, h_2, ..., h_N\big\} \subseteq \prod_{p=1}^{P} \mathbb{M}_p \;,$$

where for each $p = 1, ..., P$,

$$\mathbb{M}_p \;\coloneqq\; \Big\{\overline{m}_1^p,\, \overline{m}_2^p,\, ...\,,\, \overline{m}_{M_p}^p\Big\} \;\subset\; \mathbb{R}^R$$

represents the set of different modes that the breaker p can eventually take, with $M_p \in \mathbb{N}^*$ the total number of modes and $R \in \mathbb{N}^*$ the length of these. For instance, if $M_p = 2$, then $\mathbb{M}_p = \big\{\overline{m}_1^p, \overline{m}_2^p\big\}$. Suppose $R = 1$, then $\mathbb{M}_p$ could represent the breaker state's modes *off* and *on*, i.e., we can let, e.g., $\overline{m}_1^p = 0$ and $\overline{m}_2^p = 1$, respectively (resp.); and $\mathbb{H}$ will be a set of binary vectors. Second, if we suppose $R = 2$, then $\mathbb{M}_p$ could represent breaker state's modes with information in two dimensions. For example, a state mode is a vector in $\mathbb{R}^2$ representing first the number of jumps between state values and second the visited state values, both on average in one hour.

Let $\mathbb{1}_n : \mathbb{H} \to \{0, 1\}$ the indicator function defined for each $n = 1, ..., N$ by $\mathbb{1}_n(h_m) = 1$ if $n = m$, and $\mathbb{1}_n(h_m) = 0$ otherwise. Then, the vector $(\mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_N)$ is a bijection from $\mathbb{H}$ to the set of unit vectors

$$\mathbb{X} \;\coloneqq\; \{e_1, e_2, ..., e_N\} \;\subset\; \mathbb{R}^N \;,$$

where $e_n \in \mathbb{X}$ denotes a vector in $\mathbb{R}^N$ with unity in the n-th position and zero elsewhere, $n = 1, ..., N$. Without loss of generality (w.l.o.g.), we shall consider an underlying state process $X \coloneqq \{X_t\}_{t\geq 0}$ with state space $\mathbb{X}$ defined by

$$X_t \;\coloneqq\; \big(\mathbb{1}_1(H_t)\,,\, \mathbb{1}_2(H_t)\,,\, ...\,,\, \mathbb{1}_N(H_t)\big)\;. \tag{1}$$

Note that at any time $t \geq 0$, just one component of X_t is one and the others are all zero. In this way, X_t can be represented as:

$$X_t \;=\; \sum_{n=1}^{N} \langle X_t, e_n\rangle e_n \;,$$

where $\langle\,\cdot\,,\,\cdot\,\rangle$ denotes the inner product in $\mathbb{R}^N$. For instance, if X_t is at state $e_n \in \mathbb{X}$, this means the grid is at reference configuration $h_n \in \mathbb{H}$. Mathematically, the state process representing at which reference configuration the grid is at time $t \geq 0$, it can be computed by:

$$H_t \;=\; \sum_{n=1}^{N} \langle X_t, e_n\rangle h_n \;. \tag{2}$$

In this way, each reference configuration $h_n \in \mathbb{H}$ is identified with an unit vector $e_n \in \mathbb{X}$, $n = 1, ..., N$. Thus, instead of inferring the hidden evolution of the grid represented by H, we can estimate w.l.o.g. the temporal evolution of the underlying state process X.

Similarly, the set of different state values of the breaker $p = 1, ..., P$ is defined by

$$\mathbb{K}_p \;\coloneqq\; \big\{0,\, 1\,,\, ...,\, K_p - 1\big\} \;\subset\; \mathbb{N}\,,$$

where $K_p \in \mathbb{N}^*$. For instance, if $K_p = 2$, then the breaker p takes binary values 0 and 1. This set is identified with the set of unit vectors

$$\mathbb{Y}_p \;\coloneqq\; \Big\{f_1^p, f_2^p, ..., f_{K_p}^p\Big\} \;\subset\; \mathbb{R}^{K_p}\;,$$

where $f_k^p \in \mathbb{Y}_p$ denotes the vector in $\mathbb{R}^{K_p}$ with unity in the k-th position and zero elsewhere, $k = 1, ..., K_p$. Thus, the temporal evolution of the breaker state of $p = 1, ..., P$ is represented by $Y^p = \{Y_t^p\}_{t\geq 0}$, and Y_t^p at state f_k^p means that breaker p is in state $k - 1 \in \mathbb{K}_p$ at time $t \geq 0$. For instance, if $K_p = 2$ for each $p = 1, ..., P$, then all breakers take binary values, and $f_1^p = (1, 0)$ and $f_2^p = (0, 1)$ represent the state values $0 \in \mathbb{K}_p$ and $1 \in \mathbb{K}_p$, resp.

2.2 The Dynamic of the Hidden Process of the Grid

Since $X = \{X_t\}_{t\geq 0}$ is a Markov chain by assumption, we shall suppose that X has a transition rate matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$, where

$$a_{ij} \;\coloneqq\; \frac{d}{dt}\mathbb{P}\big[X_t = e_i \mid X_0 = e_j\big]\bigg|_{t=0}$$

represents the transition probability rate of X from state $e_j \in \mathbb{X}$ to the state $e_i \in \mathbb{X}$, for each $i, j = 1, ..., N$, $i \neq j$. In addition, the transpose of A belongs to the Q-matrix class[1]. Thus, defining $p_{i,t} \coloneqq \mathbb{P}[X_t = e_i]$, $i = 1, ..., N$, $t \geq 0$, the probability distribution vector $p_t \coloneqq (p_{1,t}, p_{2,t}, ..., p_{N,t})$ satisfies the forward equation $dp_t/dt = Ap_t$.

On the other hand, X is adapted to the (complete) right-continuous increasing family of the natural σ-fields generated by himself, i.e., to the natural filtration $\mathcal{F}_t \coloneqq \sigma(X_s;\, s \leq t) \subset \mathcal{F}$. Then, the following process

$$V_t \;\coloneqq\; X_t \;-\; X_0 \;-\; \int_0^t AX_s ds$$

is a $(\mathcal{F}_t, \mathbb{P})$-Martingale [15, Lemma 2.6.18]. The semi-Martingale representation of X is therefore:

$$X_t \;=\; X_0 \;+\; \int_0^t AX_s ds \;+\; V_t \;. \tag{3}$$

2.3 The Dynamic of the Observed Processes of the Breakers

The observed state process $Y^p = \{Y_t^p\}_{t\geq 0}$ related with the breaker $p = 1, ..., P$ is directly related with the state process of the grid $H = \{H_t\}_{t\geq 0}$. First, since $X = \{X_t\}_{t\geq 0}$ takes unit vectors in $\mathbb{X} \subset \mathbb{R}^N$, then we can express any matrix of real-valued functions with finite range, let say $C^p : \mathbb{X} \to \mathbb{R}^{K_p \times K_p}$ in function of X_t as:

$$C^p(X_t) \;=\; \sum_{n=1}^{N} C^p(e_n)\langle X_t, e_n\rangle \;. \tag{4}$$

Second, it is assumed that the transition probability rate of Y^p from state $f_l^p \in \mathbb{Y}_p$ to the state $f_k^p \in \mathbb{Y}_p$ also depends on the local p-th position of the reference configurations of the grid, i.e., the mode of p in the reference configurations. Mathematically, let $\text{proj}_p : \mathbb{H} \to \mathbb{M}_p$ the p-th projection function. We denote by $H_t^p \coloneqq \text{proj}_p(H_t)$ the mode of the breaker p in the unknown reference configuration H_t at time $t \geq 0$, i.e., H_t^p is at mode $\overline{m}_m^p \in \mathbb{M}_p$ for some $m = 1, ..., M_p$. Following [11], we can relate each observed state process Y^p with the underlying hidden process X by its transition rate matrix $C^p(X_t) \in \mathbb{R}^{K_p \times K_p}$, where the transpose of $C^p(X_t)$ belongs to the Q-matrix class. This matrix can be expressed as the sum in eq. (4). In our case, we can obtain for each $n = 1, ..., N$ an expression of the matrix $C^p(e_n) = \big(c_{kl}^p(e_n)\big)$ as a function of the p-th position of the reference configuration h_n. Indeed, let $e_n \in \mathbb{X}$. For each $p = 1, ..., P$ and $k, l = 1, ..., K_p$, $k \neq l$, the transition probability rate of Y^p from state $f_l^p \in \mathbb{Y}_p$ to the state $f_k^p \in \mathbb{Y}_p$ can be expressed by:

$$\begin{aligned}
c_{kl}^p(e_n) &= \frac{d}{dt}\mathbb{P}\big[\, Y_t^p = f_k^p \;\mid\; Y_0^p = f_l^p \,,\; X_0 = e_n \big]\Big|_{t=0} \\
&= \frac{d}{dt}\mathbb{P}\big[\, Y_t^p = f_k^p \;\mid\; Y_0^p = f_l^p \,,\; H_0 = h_n \big]\Big|_{t=0} \\
&\quad \big(\text{by eq. (2), because } \langle X_0, e_n\rangle = 1 \text{ and } \langle X_0, e_{n'}\rangle = 0, \text{ for each } n' = 1, ..., N,\, n' \neq n\big) \\
&= \frac{d}{dt}\mathbb{P}\big[\, Y_t^p = f_k^p \;\mid\; Y_0^p = f_l^p \,,\; H_0^p = \overline{m}_m^p \big]\Big|_{t=0} \,,
\end{aligned}$$

where the last equality holds by the local assumption over the reference configurations. In such a way, for each mode $\overline{m}_m^p \in \mathbb{M}_p$, $m = 1, ..., M_p$, of the breaker $p = 1, ..., P$, we define the transition rate matrix $C_m^p = \big(c_{kl}^{p,m}\big)$ of Y^p (whose transpose belongs to the Q-matrix class), by:

$$c_{kl}^{p,m} \;\coloneqq\; \frac{d}{dt}\mathbb{P}\Big[\, Y_t^p = f_k^p \;\mid\; Y_0^p = f_l^p \,,\; H_0^p = \overline{m}_m^p \Big]\Big|_{t=0} \,, \tag{5}$$

for each $k, l = 1, ..., K_p$, $k \neq l$. Finally, based on eq. (4), we express $C^p(X_t)$ in function of the number of modes of p, by:

$$C^p(X_t) \;=\; \sum_{m=1}^{M_p} C_m^p \sum_{n \in I_m^p} \langle X_t, e_n\rangle \;, \tag{6}$$

[1] That is, for each $j = 1, ..., N$, $\sum_{i=1}^N a_{ij} = 0$ and $a_{ij} \geq 0$, $\forall\, i \neq j$.

where $C_m^p \in \mathbb{R}^{K_p \times K_p}$ is the matrix with components of eq. (5), and $I_m^p \subseteq I \coloneqq \{1, ..., N\}$ is the subset of indices $n = 1, ..., N$ for which the mode of the breaker p is $\overline{m}_m^p \in \mathbb{M}_p$, $m = 1, ..., M_p$, in the p-th position of the reference configuration $h_n \in \mathbb{H}$, i.e.,

$$I_m^p \;\coloneqq\; \left\{ n \in I \;\middle|\; \mathrm{proj}_p(h_n) = \overline{m}_m^p \,,\; h_n \in \mathbb{H} \,,\; \overline{m}_m^p \in \mathbb{M}_p \right\} .$$

For instance, if for some breaker $p = 1, ..., P$, $M_p = 2$, then we look for the index of the reference configurations in which the values in their p-th position are the modes $\overline{m}_1^p$ and $\overline{m}_2^p$ (which could represent the modes *off* and *on*, resp.), and for each $m = 1, 2$, there is therefore a transition rate matrix C_m^p. For convenience, we associate with each I_m^p a diagonal matrix $\mathrm{diag}_m^p \in \mathbb{R}^{N \times N}$ defined by:

$$\mathrm{diag}_m^p \;\coloneqq\; \mathrm{diag}\Big(\mathbb{1}_{\{1 \in I_m^p\}},\; \mathbb{1}_{\{2 \in I_m^p\}},\; ...,\; \mathbb{1}_{\{N \in I_m^p\}} \Big) \,, \tag{7}$$

where $\mathbb{1}_{\{n \in I_m^p\}}$ is the indicator function for sets, i.e., $\mathbb{1}_{\{n \in I_m^p\}} = 1$ if $n \in I_m^p$, and $\mathbb{1}_{\{n \in I_m^p\}} = 0$ otherwise.

Now, for each state process $Y^p = \{Y_t^p\}_{t \geq 0}$, $p = 1, ..., P$, the following process:

$$W_t^p \;\coloneqq\; Y_t^p \;-\; Y_0^p \;-\; \int_0^t C^p(X_s) Y_s^p ds \,,$$

is a $(\mathcal{G}_t, \mathbb{P})$-Martingale [11, Lemma 2.2], where $\mathcal{G}_t \coloneqq \sigma(X_s, Y_s^p;\; s \leq t,\, p = 1, ..., P)$ represents the right-continuous complete filtration generated by X and all observed processes Y^p, $p = 1, ..., P$. The semi-Martingale representation of Y_t^p is therefore:

$$Y_t^p \;=\; Y_0^p \;+\; \int_0^t C^p(X_s) Y_s^p ds \;+\; W_t^p \,. \tag{8}$$

We denote by $\mathcal{Y}_t \coloneqq \sigma(Y_s^p;\; s \leq t,\, p = 1, ..., P)$ the corresponding right-continuous complete filtration generated by all observed processes Y^p, $p = 1, ..., P$.

In summary, H_t represents the hidden evolution of the grid at time $t \geq 0$ between the known reference configurations in $\mathbb{H}$. Each reference configuration is a vector of breakers' modes. Instead of estimating each H_t, we estimate w.l.o.g. the hidden underlying process X_t, whose state space is $\mathbb{X}$ of unit vectors of $\mathbb{R}^N$. The definition of X_t is given by eq. (1) and its semi-Martingale representation is given in eq. (3). The evolution at time $t \geq 0$ of the breaker state $p = 1, ..., P$ of the grid is represented by Y_t^p. This is the observable information to estimate X_t, and then H_t. The state space of all Y_t^p is the set $\mathbb{Y}_p$ of unit vectors of $\mathbb{R}^{K_p}$, and the semi-Martingale representation of Y_t^p is given by eq. (8). For instance, suppose that there are $P = 4$ breakers in the grid, the beaker's states are 0 and 1, and that the reference configurations are $h_1 = (1, 0, 1, 0)$, $h_2 = (0, 1, 0, 1)$, $h_3 = (0, 0, 0, 0)$, and $h_4 = (1, 1, 1, 0)$. Also suppose that temporal evolution of each breaker state is as in Figure 1. Then, under these observations over time, we want to estimate the temporal evolution of the grid represented by H, but equivalently, using the process X as it is shown in Figure 2.

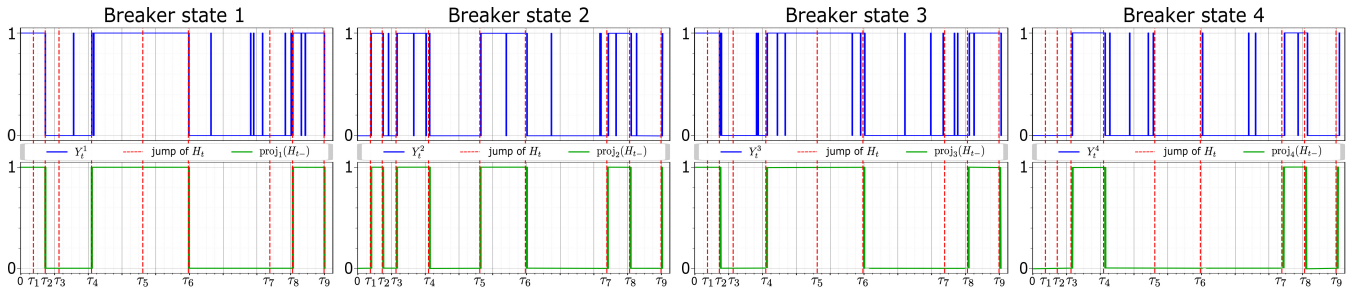

Fig. 1: Temporal evolution of breakers' states and its resp. component in the reference configuration of the grid.

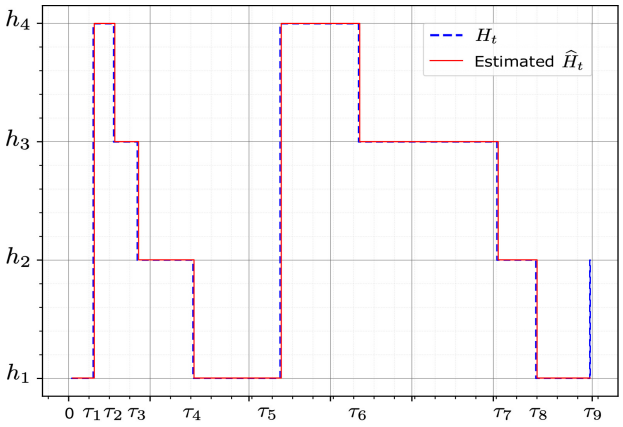

Fig. 2: Temporal evolution of the hidden process H and its estimator $\widehat{H}$ representing the grid over the reference configurations.

3 Finite-Dimensional Filter for the Temporal Evolution of the Grid

In this section, we provide filtering approaches to estimate the model parameters and the evolution of the reference configuration from all observed states processes of the breakers.

3.1 Moving to a "Fictitious World"

Recall that we are looking for an estimation of grid evolution over time by using the filtration $\mathcal{Y}_t = \sigma(Y_s^p;\ s \leq t,\ p = 1, ..., P)$ generated by all observed processes Y^p of the breakers $p = 1, ..., P$. This is done through the estimation of X_t at each time $t \geq 0$. The filtered estimate of X_t under $\mathbb{P}$ is the expectation operator $\overline{\mathbb{E}}$ over X_t given $\mathcal{Y}_t$. In fact, since X_t is defined as an indicator function in eq. (1), the filtered estimate is a conditional probability distribution, i.e., $\mathbb{E}[X_t \mid \mathcal{Y}_t] = \big(\mathbb{P}[X_t = e_1 \mid \mathcal{Y}_t], ..., \mathbb{P}[X_t = e_N \mid \mathcal{Y}_t]\big)$. It can be shown that an explicit equation for $\mathbb{E}[X_t \mid \mathcal{Y}_t]$ can be obtained, but it will be nonlinear. In contrast, by using some change of the probability measure $\mathbb{P}$, we can obtain filtered estimate that will be is linear, as it will be shown below. To obtain $\mathbb{E}[X_t \mid \mathcal{Y}_t]$ we can use a simple Bayes' rule.

Suppose that on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there is for each breaker $p = 1, ..., P$ a counting process $N_{kl,t}^p$ of the number of jumps of the state process Y^p from state $f_k^p \in \mathbb{Y}_p$ to state $f_l^p \in \mathbb{Y}_p$ within the time interval $[0, t]$, $k, l = 1, ..., K_p$, $k \neq l$. The semi-Martingale representation of $N_{kl,t}^p$ can be obtained via the following decomposition:

$$\begin{aligned} N_{kl,t}^p &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dY_s^p \rangle \\ &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, C^p(X_s) Y_s^p \rangle ds \ + \ \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dW_s^p \rangle \ , \end{aligned} \tag{9}$$

where, we have used the eq. (8) in differential form, and $Y_{t^-}^p \coloneqq \lim_{s \uparrow t} Y_s^p$ is the left limit of the state process Y_t^p at $t \geq 0$. Note that each $N_{kl,t}^p$ is $\mathcal{Y}_t$-measurable for each $t \geq 0$ and have no common jumps for indices $(k', l') \neq (k, l)$. Now, since $C^p(X_t)$ is given by (6), the semi-Martingale representation of $N_{kl,t}^p$ is given by:

$$\begin{aligned} N_{kl,t}^p &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \sum_{m=1}^{M_p} \langle f_l^p, C_m^p Y_s^p \rangle \sum_{n \in I_m^p} \langle X_s, e_n \rangle ds \ + \ \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dW_s^p \rangle \\ &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \sum_{m=1}^{M_p} c_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_s, e_n \rangle ds \ + \ \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dW_s^p \rangle \\ &= \int_0^t \lambda_{kl,s}^p ds \ + \ M_{kl,t}^p \ , \end{aligned}$$

where $M^p_{kl,t} \coloneqq N^p_{kl,t} - \int_0^t \lambda^p_{kl,s} ds$ is a $(\mathcal{G}_t, \mathbb{P})$-Martingale [11], and $\lambda^p_{kl,t}$ represents the "$\mathbb{P}$-intensity" of the counting process $N^p_{kl,t}$ and is defined by:

$$\lambda^p_{kl,t} \coloneqq \langle f^p_k, Y^p_{t-} \rangle \sum_{m=1}^{M_p} c^{p,m}_{lk} \sum_{n \in I^p_m} \langle X_t, e_n \rangle \, . \tag{10}$$

The idea is then to introduce a new probability measure $\overline{\mathbb{P}}$ for a "fictitious world" from the probability measure $\mathbb{P}$ of the "real world" to change all intensities to one under $\overline{\mathbb{P}}$. This is described by means of the Radon-Nikodym derivative, see, e.g. [9, Ch. VI, Sec.2-3]. By using [9, Ch. VI, eq. (3.3)] but for multidimensional case[2], we define $\overline{\mathbb{P}}$ by putting:

$$\left.\frac{d\overline{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{G}_t} \;=\; \Lambda_t \;\coloneqq\; \exp\left\{ -\sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln\big(\lambda^p_{kl,s}\big) dN^p_{kl,s} \;+\; \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \big(\lambda^p_{kl,s} - 1\big) ds \right\} , \tag{11}$$

which is a $(\mathcal{G}_t, \mathbb{P})$-martingale. Using now Ito's Lemma, see, e.g., [16], we have:

$$\Lambda_t \;=\; 1 \;-\; \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \Lambda_{s-} \big(\lambda^p_{kl,s}\big)^{-1} \big(\lambda^p_{kl,s} - 1\big) \big(dN^p_{kl,s} - \lambda^p_{kl,s} ds\big) \, . \tag{12}$$

We also define the reverse counterpart of (11) by putting:

$$\overline{\Lambda}_t \;\coloneqq\; \exp\left\{ \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln\big(\lambda^p_{kl,s}\big) dN^p_{kl,s} \;-\; \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \big(\lambda^p_{kl,s} - 1\big) ds \right\} , \tag{13}$$

so that $\overline{\Lambda}_t \Lambda_t = 1$. Again by Ito's Lemma, it holds:

$$\overline{\Lambda}_t \;=\; 1 \;+\; \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \overline{\Lambda}_{s-} \big(\lambda^p_{kl,s} - 1\big) d\big(N^p_{kl,s} - s\big) \, . \tag{14}$$

In this way, $\overline{\Lambda}_t$ and $\big(N^p_{kl,t} - t\big)$ are $(\mathcal{G}_t, \overline{\mathbb{P}})$-martingale $\forall\, t \geq 0$. It can be also shown that, under $\overline{\mathbb{P}}$, the dynamic for X_t is still given by (3), $N^p_{kl,t}$ are independent Poisson processes, and that they have fixed intensity one, see, e.g., [12, Lemma 1], [9, Ch. II, Theorem T6] and [15, Lemma 4.7.1] resp., *mutatis mutandi*.

3.2 Filter Estimate for the Grid States

The idea is to use $\overline{\Lambda}_t$ to compute the estimator $\sigma_t(X_t) \coloneqq \mathbb{E}[X_t \mid \mathcal{Y}_t]$ by means of a version of Bayes' rule, see, e.g., [9, Ch. VI, Lemma L5]. More precisely, for any $\mathcal{G}_t$-adapted and integrable process F_t, the filtered estimate of F_t can be computed via:

$$\mathbb{E}[F_t \mid \mathcal{Y}_t] \;=\; \frac{\overline{\mathbb{E}}[\overline{\Lambda}_t F_t \mid \mathcal{Y}_t]}{\overline{\mathbb{E}}[\overline{\Lambda}_t \mid \mathcal{Y}_t]} \, , \tag{15}$$

where $\overline{\mathbb{E}}$ denotes the expectation operator under the probability measure $\overline{\mathbb{P}}$. We denote by $\overline{\sigma}_t(F_t)$ the expectation $\overline{\mathbb{E}}[\overline{\Lambda}_t F_t \mid \mathcal{Y}_t]$. Consequently $\overline{\sigma}_t(1) = \overline{\mathbb{E}}[\overline{\Lambda}_t \mid \mathcal{Y}_t]$. Note that $\overline{\sigma}_t(1)$ can be computed as the sum of the components of $\overline{\sigma}_t(X_t)$. Indeed, since X_t takes values in the space $\mathbb{X}$ of unit vectors of $\mathbb{R}^N$, then $\langle X_t, \mathbf{1}_N \rangle = 1$ for all $t \geq 0$, where $\mathbf{1}_N \coloneqq \sum_{n=1}^{N} e_n$, and therefore $\overline{\sigma}_t(F_t) = \overline{\sigma}_t(F_t \langle X_t, \mathbf{1}_N \rangle) = \langle \overline{\sigma}_t(F_t X_t), \mathbf{1}_N \rangle$. Thus, in particular taking $F_t \equiv 1$ we have $\overline{\sigma}_t(1) = \langle \overline{\sigma}_t(X_t), \mathbf{1}_N \rangle$. The linear filtered estimate of X_t is given in the next Proposition 1.

[2] See, e.g., [9, Ch. VI, Theorem T2] or [11, eq. (14)] for a general case.

Proposition 1 *The finite-dimensional (unnormalized) estimator for the states of X_t is of the form:*

$$\begin{aligned}\overline{\sigma}_t(X_t) \;=\; \overline{\sigma}_0(X_0) \;+\; A\int_0^t \overline{\sigma}_s(X_s)ds \;-\; \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\int_0^t \overline{\sigma}_s(X_{s-})d\big(N^p_{kl,s}-s\big)\\ +\;\sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\sum_{m=1}^{M_p}\int_0^t \langle f^p_k, Y^p_{s-}\rangle c^{p,m}_{lk}\,\mathrm{diag}^p_m\,\overline{\sigma}_s(X_{s-})d\big(N^p_{kl,s}-s\big)\;,\end{aligned} \tag{16}$$

where diag^p_m *is the diagonal matrix defined in eq.* (7).

Proof ■

To obtain $\overline{\sigma}_t(X_t)$, we need the estimation of all parameters involved in eq. (16), i.e., the matrices $A=(a_{ij})$ and $C^p_m=\big(c^{p,m}_{kl}\big)$, for each $p=1,...,P$ and $m=1,...,M_p$. This is the purpose of the next section.

3.3 Parameter Estimation

To estimate the parameters that define the transition rate matrices of X and Y^p, $p=1,...,P$, we focus on the EM algorithm for continuous-time stochastic processes, see, e.g., [17,18,7]. The idea is to maximize a likelihood function in an iterative form. Let $\{\mathbb{P}_\theta,\,\theta\in\Theta\}$ be a family of probability measures on the measurable space $(\Omega,\mathcal{F})$, all absolutely continuous with respect to the (initial) fixed probability measure $\mathbb{P}$, wherein our case,

$$\Theta \;:=\; \bigcup\big\{a_{ij}\,,\,c^{p,m}_{kl}\;;\;1\leq i,j\leq N\,,\;i\neq j\,,\;1\leq k,l\leq K_p\,,\;k\neq l\,,\;1\leq m\leq M_p\,,\;1\leq p\leq P\big\}\,.$$

The log-likelihood for an estimation of a $\theta\in\Theta$ can be defined by:

$$\mathcal{L}(\theta) \;:=\; \ln\left(\mathbb{E}\left[\frac{d\mathbb{P}_\theta}{d\mathbb{P}}\;\middle|\;\mathcal{Y}\right]\right)\,,$$

where $\mathcal{Y}\subset\mathcal{F}$, and then, the Maximum Likelihood Estimator (MLE) is defined by $\theta^*\in\arg\max_{\theta\in\Theta}\mathcal{L}(\theta)$.

In general, computing directly the MLE is challenging. The Expectation–Maximization (EM) algorithm provides an iterative approximation method starting from an initial estimation θ_0, see Section 4.2. This algorithm is based on the following straightforward application of the well-known Jensen's inequality:

$$\mathcal{L}(\theta)\;-\;\mathcal{L}(\widehat{\theta}) \;=\; \ln\left(\mathbb{E}_{\widehat{\theta}}\left[\frac{d\mathbb{P}_\theta}{d\mathbb{P}_{\widehat{\theta}}}\;\middle|\;\mathcal{Y}\right]\right) \;\geq\; \mathbb{E}_{\widehat{\theta}}\left[\ln\left(\frac{d\mathbb{P}_\theta}{d\mathbb{P}_{\widehat{\theta}}}\right)\;\middle|\;\mathcal{Y}\right] \;=:\; \mathcal{Q}\big(\theta,\widehat{\theta}\big)\,.$$

This gives a global minoration for the log-likelihood mapping $\theta\mapsto\mathcal{L}(\theta)$ by means of the auxiliary mapping $\theta\mapsto\mathcal{L}(\widehat{\theta})+\mathcal{Q}\big(\theta,\widehat{\theta}\big)$. At each iteration $r\in\mathbb{N}_0$, the EM algorithm consists of two main steps:

(1) **E-step:** set $\widehat{\theta}=\widehat{\theta}_r$ and compute $\mathcal{Q}\big(\,\cdot\,,\widehat{\theta}\big)$,
(2) **M-step:** find $\widehat{\theta}_{r+1}\in\arg\max\limits_{\theta\in\Theta}\,\mathcal{Q}\big(\theta,\widehat{\theta}\big)$.

This algorithm can be stopped when a stopping test is satisfied, see Section 4.3. The generated sequence $\big\{\widehat{\theta}_r\big\}_{r\in\mathbb{N}_0}$ gives nondecreasing values of the likelihood function, i.e., $\mathcal{L}\big(\widehat{\theta}_{r+1}\big)>\mathcal{L}\big(\widehat{\theta}_r\big)$ unless $\widehat{\theta}_{r+1}=\widehat{\theta}_r$. For convergence issues, see, e.g., [17,18,19].

In our context, suppose our model is determined by the parameters:

$$\theta \;=\; \big\{a_{ij}\,,\,c^{p,m}_{kl}\;\;;\;\;1\leq i,j\leq N\,,\;i\neq j\,,\;1\leq k,l\leq K_p\,,\;k\neq l\,,\;1\leq m\leq M_p\,,\;1\leq p\leq P\big\}\,, \tag{17}$$

i.e., we have computed already the *E-step* under θ. To compute the new parameters

$$\widehat{\theta} \;=\; \big\{\widehat{a}_{ij}\,,\,\widehat{c}^{p,m}_{kl}\;\;;\;\;1\leq i,j\leq N\,,\;i\neq j\,,\;1\leq k,l\leq K_p\,,\;k\neq l\,,\;1\leq m\leq M_p\,,\;1\leq p\leq P\big\} \tag{18}$$

that maximize the log-likelihood, i.e., the *M-step*, we have the following Theorem 1.

Theorem 1 *The estimation $\widehat{A} = (\widehat{a}_{ij})$ of $A = (a_{ij})$, and $\widehat{C}^p_m = \big(\widehat{c}^{\,p,m}_{kl}\big)$ of $C^p_m = \big(c^{p,m}_{kl}\big)$, for each $p = 1, ..., P$ and $m = 1, ..., M_p$; are given for $i \neq j$ and $k \neq l$, by:*

$$\widehat{a}_{ji} \;=\; \frac{\mathbb{E}[J_{ij,t} \mid \mathcal{Y}_t]}{\mathbb{E}[O_{i,t} \mid \mathcal{Y}_t]} \qquad , \qquad \widehat{c}^{\,p,m}_{lk} \;=\; \frac{\sum\limits_{n \in I^p_m} \mathbb{E}[L^{p,n}_{kl,t} \mid \mathcal{Y}_t]}{\sum\limits_{n \in I^p_m} \mathbb{E}[S^{p,n}_{k,t} \mid \mathcal{Y}_t]} \;,$$

where, for $i, j = 1, ..., N$, $i \neq j$,

$$J_{ij,t} \;\coloneqq\; \int_0^t \langle e_i, X_{s-}\rangle \langle e_j, dX_s\rangle \qquad , \qquad O_{i,t} \;\coloneqq\; \int_0^t \langle e_i, X_s\rangle ds \;, \tag{19}$$

and, for $p = 1, ..., P$, $k, l = 1, ..., K_p$, $k \neq l$, and $n = 1, ..., N$,

$$L^{p,n}_{kl,t} \;\coloneqq\; \int_0^t \langle e_n, X_{s-}\rangle dN^p_{kl,s} \qquad , \qquad S^{p,n}_{k,t} \;\coloneqq\; \int_0^t \langle f^p_k, Y^p_s\rangle \langle e_n, X_s\rangle ds \;. \tag{20}$$

Proof ■

In Theorem 1, note that $J_{ij,t}$ represents a counting process of the number of jumps of X from state $e_i \in \mathbb{X}$ to state $e_j \in \mathbb{X}$ within $[0, t]$, $i \neq j$, $O_{i,t}$ stands for the occupation time by X on the state $e_i \in \mathbb{X}$ within $[0, t]$, $L^{p,n}_{kl,t}$ represents the process that increases only when Y^p jumps from state $f^p_k \in \mathbb{Y}_p$ to state $f^p_l \in \mathbb{Y}_p$ and X is in state $e_n \in \mathbb{X}$, $k \neq l$; and $S^{p,n}_{k,t}$ stands for the total time up to $t \geq 0$ for which X is in state $e_n \in \mathbb{X}$ and simultaneously Y^p is in state $f^p_k \in \mathbb{Y}_p$.

From eq. (15), the estimation $\widehat{a}_{ij}$ and $\widehat{c}^{\,p,m}_{kl}$ can be obtained via the probability measure $\overline{\mathbb{P}}$ by:

$$\widehat{a}_{ji} \;=\; \frac{\big\langle \overline{\sigma}_t(J_{ij,t}X_t), \mathbf{1}_N \big\rangle}{\big\langle \overline{\sigma}_t(O_{i,t}X_t), \mathbf{1}_N \big\rangle} \qquad , \qquad \widehat{c}^{\,p,m}_{lk} \;=\; \frac{\sum\limits_{n \in I^p_m} \big\langle \overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big), \mathbf{1}_N \big\rangle}{\sum\limits_{n \in I^p_m} \big\langle \overline{\sigma}_t\big(S^{p,n}_{k,t}X_t\big), \mathbf{1}_N \big\rangle} \;. \tag{21}$$

In this way, it is sufficient to compute the estimators $\overline{\sigma}_t(J_{ij,t}X_t)$, $\overline{\sigma}_t(O_{i,t}X_t)$, $\overline{\sigma}_t\big(S^{p,n}_{k,t}X_t\big)$ and $\overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big)$. Now, if we consider the process:

$$F_t \;=\; F_0 \;+\; \int_0^t \alpha(X_s)ds \;+\; \int_0^t \langle \beta(X_s), dV_s\rangle \tag{22}$$

then, for each $i, j = 1, ..., N$, $i \neq j$, $p = 1, ..., P$, $k = 1, ..., K_p$, $n = 1, ..., N$, the processes $J_{ij,t}$, $O_{i,t}$ and $S^{p,n}_{k,t}$ are considered into F_t, where $F_0 \in \mathbb{R}$ is known, and $\alpha : \mathbb{X} \to \mathbb{R}$ and $\beta : \mathbb{X} \to \mathbb{R}^N$ are known functions with finite range, $\mathcal{G}_t$-adapted and integrable for each $t \geq 0$. Indeed, by using eq. (3), and taking

(i) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = \langle e_i, X_t\rangle a_{ji}$, and $\beta(X_t) = \langle e_i, X_t\rangle e_j$, we obtain $F_t = J_{ij,t}$,
(ii) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = \langle e_i, X_t\rangle$, and $\beta(X_t) = \mathbf{0}_N \in \mathbb{R}^N$, we obtain $F_t = O_{i,t}$,
(iii)$F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = \langle f^p_k, Y^p_t\rangle \langle e_n, X_t\rangle$, and $\beta(X_t) = \mathbf{0}_N \in \mathbb{R}^N$, we obtain $F_t = S^{p,n}_{k,t}$.

Therefore, to compute $\overline{\sigma}_t(J_{ij,t}X_t)$, $\overline{\sigma}_t(O_{i,t}X_t)$, and $\overline{\sigma}_t\big(S^{p,n}_{k,t}X_t\big)$, we can compute once $\overline{\sigma}_t(F_tX_t)$ and restrict afterwards to the particular cases of $\alpha(X_t)$ and $\beta(X_t)$. On the other hand, we know that $\overline{\sigma}_t(F_t) = \langle \overline{\sigma}_t(F_tX_t), \mathbf{1}_N\rangle$, so that, we make the inner product between $\overline{\sigma}_t(F_tX_t)$ and $\mathbf{1}_N$ to have the estimation for all parameters, see eq. (21). The following Theorem 2 gives the linear filter estimate $\overline{\sigma}_t(F_tX_t)$. The filter estimate $\overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big)$ is given in Theorem 3.

Theorem 2 *The finite-dimensional (unnormalized) estimator for F_tX_t for is of the form:*

$$\begin{aligned}
\overline{\sigma}_t(F_tX_t) = \overline{\sigma}_0(F_0X_0) \;+\; A\int_0^t \overline{\sigma}_s(F_sX_s)ds \;-\; \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k\neq l}}^{K_p} \int_0^t \overline{\sigma}_s(F_{s-}X_{s-})d\big(N^p_{kl,s} - s\big) \\
+ \int_0^t \overline{\sigma}_s(X_s\alpha_s)ds \;+\; \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \int_0^t \big\langle \overline{\sigma}_s\big((\beta_{j,s} - \beta_{i,s})X_s\big), e_i\big\rangle a_{ji}(e_j - e_i)ds \\
+ \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k\neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f^p_k, Y^p_{s-}\rangle c^{p,m}_{lk} \operatorname{diag}^p_m \overline{\sigma}_s(F_{s-}X_{s-})d\big(N^p_{kl,s} - s\big) \;,
\end{aligned} \tag{23}$$

where $\alpha_t \coloneqq \alpha(X_t) \in \mathbb{R}$ and $\beta_t \coloneqq \beta(X_t) \in \mathbb{R}^N$ are known functions with finite range, $\mathcal{G}_t$-adapted and integrable for each $t \geq 0$, F_t is given in eq. (22), and diag^p_m is the diagonal matrix of eq. (7).

Proof ■

Remark 1 Note that if we consider $F_t = F_0 = 1$, $\alpha_t = 0 \in \mathbb{R}$, and $\beta_t = \mathbf{0}_N \in \mathbb{R}^N$ in the eq. (22), then the Proposition 1 is a particular case of Theorem 2. □

In this way, the filter estimates for the parameter estimation are given in the next Corollary 1 by taking the particular cases of α_t and β_t within F_t, see eq. (22) and the cases (i), (ii) and (iii).

Corollary 1 *The finite-dimensional (unnormalized) estimator for $J_{ij,t}X_t$, $O_{i,t}X_t$ and $S^{p,n}_{k,t}X_t$ are resp. of the form:*

$$\begin{aligned}
\overline{\sigma}_t(J_{ij,t}X_t) =& A\int_0^t \overline{\sigma}_s(J_{ij,s}X_s)ds + \int_0^t \langle \overline{\sigma}_s(X_s), e_i\rangle e_j a_{ji} ds - \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\int_0^t \overline{\sigma}_s(J_{ij,s-}X_{s-})d\big(N^p_{kl,s} - s\big) \\
&+ \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\sum_{m=1}^{M_p}\int_0^t \langle f^p_k, Y^p_{s-}\rangle c^{p,m}_{lk}\,\mathrm{diag}^p_m\,\overline{\sigma}_s(J_{ij,s-}X_{s-})d\big(N^p_{kl,s} - s\big)\ , \\
\overline{\sigma}_t(O_{i,t}X_t) =& A\int_0^t \overline{\sigma}_s(O_{i,s}X_s)ds + \int_0^t \langle \overline{\sigma}_s(X_s), e_i\rangle e_i ds\ - \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\int_0^t \overline{\sigma}_s(O_{i,s-}X_{s-})d\big(N^p_{kl,s} - s\big) \\
&+ \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p}\sum_{m=1}^{M_p}\int_0^t \langle f^p_k, Y^p_{s-}\rangle c^{p,m}_{lk}\,\mathrm{diag}^p_m\,\overline{\sigma}_s(O_{i,s-}X_{s-})d\big(N^p_{kl,s} - s\big)\ , \\
\overline{\sigma}_t(S^{p,n}_{k,t}X_t) =& A\int_0^t \overline{\sigma}_s(S^{p,n}_{k,s}X_s)ds + \int_0^t \langle f^p_k, Y^p_{s-}\rangle\langle \overline{\sigma}_s(X_s), e_n\rangle e_n ds - \sum_{q=1}^{P}\sum_{\substack{u,v=1\\u\neq v}}^{K_q}\int_0^t \overline{\sigma}_s(S^{p,n}_{k,s-}X_{s-})d\big(N^q_{uv,s} - s\big) \\
&+ \sum_{q=1}^{P}\sum_{\substack{u,v=1\\u\neq v}}^{K_q}\sum_{m=1}^{M_q}\int_0^t \langle f^q_u, Y^q_{s-}\rangle c^{q,m}_{vu}\,\mathrm{diag}^q_m\,\overline{\sigma}_s(S^{p,n}_{k,s-}X_{s-})d\big(N^q_{uv,s} - s\big)\ ,
\end{aligned}$$

where diag^p_m *is the matrix of eq.* (7). ■

Finally, the filter estimate we need to complete the estimation of all parameters is $\overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big)$. For this, we write the semi-Martingale representation of $L^{p,n}_{kl,t}$ from eq. (20), $k \neq l$, by:

$$L^{p,n}_{kl,t} \ = \ \int_0^t \langle e_n, X_{s-}\rangle d\big(N^p_{kl,s} - s\big) \ + \ \int_0^t \langle e_n, X_{s-}\rangle ds\ . \tag{24}$$

Theorem 3 *The finite-dimensional (unnormalized) estimator for $L^{p,n}_{kl,t}X_t$, $k \neq l$, is of the form:*

$$\begin{aligned}
\overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big) =\ & A\int_0^t \overline{\sigma}_s\big(L^{p,n}_{kl,s}X_s\big)ds\ - \sum_{q=1}^{P}\sum_{\substack{u,v=1\\u\neq v}}^{K_q}\int_0^t \overline{\sigma}_s\big(L^{p,n}_{kl,s-}X_{s-}\big)d\big(N^q_{uv,s} - s\big) \\
&+ \sum_{m=1}^{M_p}\int_0^t \langle f^p_k, Y^p_{s-}\rangle c^{p,m}_{lk}\,\mathrm{diag}^p_m\,\langle e_n, \overline{\sigma}_s(X_{s-})\rangle e_n dN^p_{kl,s} \\
&+ \sum_{q=1}^{P}\sum_{\substack{u,v=1\\u\neq v}}^{K_q}\sum_{m=1}^{M_q}\int_0^t \langle f^q_u, Y^q_{s-}\rangle c^{q,m}_{vu}\,\mathrm{diag}^q_m\,\overline{\sigma}_s\big(L^{p,n}_{kl,s-}X_{s-}\big)d\big(N^q_{uv,s} - s\big)\ ,
\end{aligned} \tag{25}$$

where diag^p_m *is the diagonal matrix defined in eq.* (7).

Proof ■

4 Numerical Methods

This section presents a numerical method for all SDEs involved in our model. We also show how to obtain the initial estimation of all parameters using available data of all breakers states of the grid.

First, we write $t_0 < t_1 < ... < t_{W+1}$ the increasingly ordered instances of all "jump times" of breakers. This is shown in Figure 3 and is obtained by superposing all state change times $\{\tau_1^p, \tau_2^p, ...\}$ of the breakers $p = 1, ..., P$. We denote by t_w, $w = 0, \ldots, W$, the instant where at least one breaker changes of state. By convention, $t_0 = 0$ and $t_{W+1} = T$. Let $\Delta t_{w+1} = t_{w+1} - t_w$ the length of time in which the breakers states remain constant between the time interval $[t_w, t_{w+1})$.

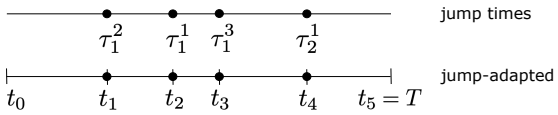

Fig. 3: Representation of the jump-adapted discretization given by the jump times of all breakers.

4.1 Jump-Adapted Scheme for Filters

The scheme presented here is a strong approximation with no discretization error on the SDEs solution, in which a jump-adapted time discretization is given by a superposition of all "jump times", i.e., all state change times generated by the temporal evolution of the breakers states, see Figure 3. The jump effects are then added at the correct jump times. To use this kind of method, one has to check whether the SDE concerned belongs to the particular subclass of SDEs for which the corresponding non-jump part admits an exact solution, see, e.g., [20, Ch. II]. In our case, all SDEs admit an explicit solution in the non-jump parts as we see in next.

Instead of building a scheme for each filter estimate $\overline{\sigma}_t(X_t)$, $\overline{\sigma}_t(J_{ij,t}X_t)$, $\overline{\sigma}_t(O_{i,t}X_t)$, $\overline{\sigma}_t\big(S_{k,t}^{p,n}X_t\big)$ and $\overline{\sigma}_t\big(L_{kl,t}^{p,n}X_t\big)$, see Corollary 1 and Theorem 3 resp.; we present a generalized scheme for the following SDEs system:

$$d\,K_t \;=\; \Xi_t K_t\,dt \;+\; \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p} Q_{kl,t-}^{p} K_{t-} dN_{kl,t}^{p}\,, \tag{26}$$

$$d\,G_t \;=\; \big(\Upsilon_t G_t \;+\; \Gamma_t K_t\big)dt \;+\; \sum_{p=1}^{P}\sum_{\substack{k,l=1\\k\neq l}}^{K_p} \big(\Pi_{kl,t-}^{p} G_{t-} \;+\; \Lambda_{kl,t-}^{p} K_{t-}\big) dN_{kl,t}^{p}\,, \tag{27}$$

where $K_t, G_t \in \mathbb{R}^N$ are $\mathcal{G}_t$-adapted and integrable for any $t \geq 0$, and $\Xi_t, Q_{kl,t}^p, \Upsilon_t, \Gamma_t, \Pi_{kl,t}^p, \Lambda_{kl,t}^p \in \mathbb{R}^{N\times N}$ are constant matrices between "jumps" for each $k, l = 1, ..., K_p$, $p = 1, ..., P$. They are defined by $\Xi_t = \Xi_{t_w}$, $\Upsilon_t = \Upsilon_{t_w}$, $\Gamma_t = \Gamma_{t_w}$, $Q_{kl,t}^p = Q_{kl,t_w}^p$, $\Pi_{kl,t}^p = \Pi_{kl,t_w}^p$, $\Lambda_{kl,t}^p = \Lambda_{kl,t_w}^p$, for any $t \in [t_w, t_{w+1})$, $w = 0, ..., W$.

Theorem 4 *The jump-adapted exact solution scheme for Z_t and G_t with initial conditions $K_0, G_0 \in \mathbb{R}^N$ is written by:*

$$\begin{aligned}
K_{t_0} &= K_0 \,, \\
G_{t_0} &= G_0 \,, \\
K_{t_{w+1}^-} &= \exp\left\{\Xi_{t_w} \Delta t_{w+1}\right\} K_{t_w} \,, \\
G_{t_{w+1}^-} &= \left[\mathrm{Id}_N \ \mathbf{0}_{N\times N}\right] \exp\left\{ \begin{bmatrix} \Upsilon_{t_w} & \Gamma_{t_w} \\ \mathbf{0}_{N\times N} & \Xi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} G_{t_w} \\ K_{t_w} \end{bmatrix} \,, \\
K_{t_{w+1}} &= K_{t_{w+1}^-} + \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k\neq l}}^{K_p} Q^p_{kl,t_{w+1}^-} K_{t_{w+1}^-} \Delta N^p_{kl,t_{w+1}} \,, \\
G_{t_{w+1}} &= G_{t_{w+1}^-} + \sum_{p=1}^{P} \sum_{\substack{k,l=1 \\ k\neq l}}^{K_p} \left(\Pi^p_{kl,t_{w+1}^-} G_{t_{w+1}^-} + \Lambda^p_{kl,t_{w+1}^-} K_{t_{w+1}^-} \right) \Delta N^p_{kl,t_{w+1}} \,,
\end{aligned}$$

where $\Delta N^p_{kl,t_{w+1}} = N^p_{kl,t_{w+1}} - N^p_{kl,t_{w+1}^-}$ is defined by $\Delta N^p_{kl,t_{w+1}} = 1$ if Y^p_t jumps from state $f^p_k \in \mathbb{Y}_p$ to state $f^p_l \in \mathbb{Y}_p$ at time $t = t_{w+1}$, and both, $G_{t_{w+1}^-} \coloneqq \lim_{s\uparrow t_{w+1}} G_s$ and $K_{t_{w+1}^-} \coloneqq \lim_{s\uparrow t_{w+1}} K_s$ are the respective values "before" the jump at time t_{w+1}.

Proof ■

As a consequence, the jump-adapted scheme for each filter estimate $\overline{\sigma}_t(X_t)$, $\overline{\sigma}_t(J_{ij,t}X_t)$, $\overline{\sigma}_t(O_{i,t}X_t)$, $\overline{\sigma}_t\big(S^{p,n}_{k,t}X_t\big)$ and $\overline{\sigma}_t\big(L^{p,n}_{kl,t}X_t\big)$ can be easily obtained from the Theorem 4.

4.2 Estimation of the Initial Parameters

For the EM algorithm, a $\theta_0 \in \Theta$ must be initialized, i.e., we need to choose initial values for the matrices $\widehat{A}^{(0)} = (\widehat{a}_{ij}(0))$ and $\widehat{C}^{p(0)}_m = \big(\widehat{c}^{p,m}_{kl}(0)\big)$, for each breaker $p = 1, ..., P$ and mode index $m = 1, ..., M_p$. At first glance, we can use Theorem 1 empirically, i.e., we can discretize all the involved integrals in the parameter estimation (those in (19) and eq. (20)), and use all the information of the breakers state to estimate θ_0.

First, since we know the change of breakers states, then we know the values of each Y^p_t, $p = 1, ..., P$, at any time $t \geq 0$, since it remains constant between "jumps". Second, because it is assumed that we know a priori the space $\mathbb{H}$ of the reference configuration of the grid, then we can compute an empirical estimation of the entire grid, and so that, an empirical estimation of the underlying state process X_t by eq. (1) and eq. (2). This can be computed through a distance measure by finding the closest reference state to the information vector of all observed processes at each jump time, i.e., by clustering and classification method. We show that in the following.

The state of the grid given by the states of all breakers can be represented by a piecewise constant state process $Y \coloneqq \{Y_t\}_{t\geq 0}$, where Y_t is the joint information of all breakers states at time $t \geq 0$ defined by:

$$Y_t \ \coloneqq \ \sum_{p=1}^{P} \sum_{k=1}^{K_p} \langle f^p_k, Y^p_t \rangle (k-1) g_p \ \in \ \prod_{p=1}^{P} \mathbb{K}_p \,, \tag{28}$$

where for each $p = 1, ..., P$, $k = 1, ..., K_p$, the k-th unit vector $f^p_k \in \mathbb{K}_p$, and $g_p \in \mathbb{R}^P$ is the p-th unit vector of $\mathbb{R}^P$. Note that Y_t is piecewise right-continuous with left limits. In this way, if a breaker state $p = 1, ..., P$ changes of value at time $t \geq 0$, then Y_t changes too. This occurs at the times $t_0 < t_1 < ... < t_W$, see Figure 3.

Let $d : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{R}_+$ a distance measure. Under the knowledge of the set of reference configurations $\mathbb{H} = \{h_1, ..., h_N\}$, we can compute an empirical estimation $\widehat{H} \coloneqq \{\widehat{H}_t\}_{t\geq 0}$ of the state of the grid over the reference configurations $h_1, ..., h_N$, by:

$$\widehat{H}_t \ \in \ \underset{\substack{h_n \in \mathbb{H} \\ n=1,...,N}}{\arg\min} \ d(Y_t, h_n) \,. \tag{29}$$

This approach is a classification procedure over the joint information of all breakers states at time $t \geq 0$, which is represented by Y_t in eq. (28).

Now, with this empirical estimation $\widehat{H}_t$, we can compute the empirical values of X_t by means of its definition in eq. (1). We denote this estimation by $\widehat{X} \coloneqq \{\widehat{X}_t\}_{t\geq 0}$. Note that $\widehat{H}$ and $\widehat{X}$ are also piecewise right-continuous with left

limits. Since we know now the values of $\widehat{X}_t$ at the state change times of Y, $t_0, ..., t_W$, we discretize all the involved integrals in Theorem 1 to estimate empirically all the parameters of our model. For each $i, j = 1, ..., N$, $i \neq j$, the initial estimation of $A = (a_{ij})$, i.e., the matrix $\widehat{A}^{(0)} = \big(\widehat{a}_{ij}(0)\big)$, and of $C_m^p = \big(c_{kl}^{p,m}\big)$, i.e, the matrix $\widehat{C}_m^{p(0)} = \big(\widehat{c}_{kl}^{p,m}(0)\big)$, for each $p = 1, ..., P$ and $m = 1, ..., M_p$, are obtained as follows:

$$\widehat{a}_{ij}(0) \;\coloneqq\; \frac{\sum\limits_{w=0}^{W-1} \langle e_i, \widehat{X}_{t_w}\rangle \langle e_j, \widehat{X}_{t_{w+1}}\rangle}{\sum\limits_{w=0}^{W-1} \langle e_i, \widehat{X}_{t_w}\rangle \Delta t_{w+1}} \quad , \quad \widehat{c}_{lk}^{p,m}(0) \;\coloneqq\; \frac{\sum\limits_{w=0}^{W-1} \langle f_k^p, Y_{t_w}^p\rangle \langle f_l^p, Y_{t_{w+1}}^p\rangle \sum\limits_{n\in I_m^p} \langle e_n, \widehat{X}_{t_w}\rangle}{\sum\limits_{w=0}^{W-1} \langle f_k^p, Y_{t_w}^p\rangle \sum\limits_{n\in I_m^p} \langle e_n, \widehat{X}_{t_w}\rangle \Delta t_{w+1}} \,. \tag{30}$$

On the other hand, the initial estimation $\overline{\sigma}_0(X_0)$ of the process X at time $t = 0$ that we need in eq. (16), is given by the empirical estimation $\overline{\sigma}_0(\widehat{X}_0) = \big(\overline{\sigma}_0(\widehat{X}_0)_1, ..., \overline{\sigma}_0(\widehat{X}_0)_N\big)$ defined for each $n = 1, ..., N$, by:

$$\overline{\sigma}_0(\widehat{X}_0)_n \;\coloneqq\; \frac{\sum\limits_{w=0}^{W-1} \langle e_n, \widehat{X}_{t_w}\rangle \Delta t_{w+1}}{\sum\limits_{n=1}^{N}\sum\limits_{w=0}^{W-1} \langle e_n, \widehat{X}_{t_w}\rangle \Delta t_{w+1}} \,. \tag{31}$$

4.3 Stopping Criteria for the EM Algorithm

Instead of using the strict stopping criteria for the EM algorithm[3] $\widehat{\theta}_{r+1} = \widehat{\theta}_r$ for some $r \in \mathbb{N}$, we define the following stopping test for numerical purposes:

$$\frac{\Big\|\overline{\sigma}_t^{(r)}(X_t) \,-\, \overline{\sigma}_t^{(r-1)}(X_t)\Big\| \;+\; \Big\|\widehat{A}^{(r)} \,-\, \widehat{A}^{(r-1)}\Big\| \;+\; \sum\limits_{p=1}^{P}\sum\limits_{m=1}^{M_p} \Big\|\widehat{C}_m^{p,(r)} \,-\, \widehat{C}_m^{p,(r-1)}\Big\|}{\Big\|\overline{\sigma}_t^{(r-1)}(X_t)\Big\| \;+\; \Big\|\widehat{A}^{(r-1)}\Big\| \;+\; \sum\limits_{p=1}^{P}\sum\limits_{m=1}^{M_p} \Big\|\widehat{C}_m^{p,(r-1)}\Big\|} \;\leq\; \varepsilon \,, \tag{32}$$

where $\varepsilon > 0$ is a given stopping parameter.

4.4 Finding the Reference Configurations of the Grid

In this section, we want to build the set of reference configurations $\mathbb{H}$ using available data which consists of the temporal evolution of the breakers states in the grid, i.e., by using $Y = \{Y_t\}_{t\geq 0}$ of eq. (28). For instance, if $K_p = 2$ for each $p = 1, ..., P$, the available data is a boolean matrix of dimension $T \times P$. We proceed by a clustering-classification method. To obtain the reference configurations of the grid, we construct clusters from data by partitioning it into N subsets. Each subset $U_1, ..., U_N$, called cluster, is represented by its representative state $\mu_1, ..., \mu_N$, resp. To obtain optimal clusters, we use K-means method, see, e.g., [10]. The extension of K-means in continuous time is given by the minimization of the following cost function in the horizon time $T > 0$:

$$J \;=\; \sum_{n=1}^{N} \int_0^T \mathbb{1}_{\{Y_t \in U_n\}} d(Y_t, \mu_n) dt \,,$$

where $\mathbb{1}_{\{Y_t\in U_n\}}$ is the indicator function for sets, i.e., $\mathbb{1}_{\{Y_t\in U_n\}} = 1$ if $Y_t \in U_n$, and $\mathbb{1}_{\{Y_t\in U_n\}} = 0$ otherwise. Since $Y = \{Y_t\}_{t\geq 0}$ is piecewise right-continuous on $0 = t_0 < t_1 < ... < t_W = T$, see eq. (28), we have

$$J \;=\; \sum_{n=1}^{N}\sum_{w=0}^{W-1} \mathbb{1}_{\{Y_{t_w}\in U_n\}} d(Y_{t_w}, \mu_n) \Delta t_{w+1} \,,$$

where $\Delta t_{w+1} = t_{w+1} - t_w$. This corresponds to the classical discrete K-means approach with a weighted cost. Here, the representative state of a cluster U_n, $n = 1, ..., N$, is:

$$\mu_n \;\in\; \operatorname*{arg\,min}_{\eta\in U_n} \left\{ \sum_{w=0}^{W-1} \mathbb{1}_{\{Y_{t_w}\in U_n\}} d(Y_{t_w}, \eta) \right\} .$$

[3] Because, e.g., it could take several iterations to have the equality in all the parameters.

For instance, if $M = 2$, it is a boolean version. We use the traditional approach to construct the clusters by classification. For each $t \geq 0$, the cluster U_n, $n = 1, ..., N$, is obtained as the set:

$$U_n \;=\; \bigcup_{w=0}^{W-1} \left\{ Y_{t_w} \in \prod_{p=1}^{P} \mathbb{K}_p \;\middle|\; d(Y_{t_w}, \mu_n) \leq d(Y_{t_w}, \mu_{n'}) \,, \;\text{ for each }\; n' = 1, ..., N \right\} \tag{33}$$

Note that the empirical estimation of the grid states in eq. (29) also means that for any time $t \geq 0$, there is a $n \in \{1, ..., N\}$ such that $\widehat{H}_t \in U_n$.

This method generates a sequence $\{\mu_r\}_{r \in \mathbb{N}_0}$, where $\mu_r \coloneqq \left(\mu_1^{(r)}, ..., \mu_N^{(r)}\right)$ is the vector of all representative states of the clusters at iteration $r \in \mathbb{N}_0$. This can be initialized from the available data, e.g., randomly, heuristically, or by K-means++ approach [21]. It should be noted that the performance of an iterative clustering algorithm may converges to numerous local minima and depends highly on initial cluster centers [22]. Finally, the reference configurations of the grid are given at the last iteration of the method, that is when $\mu_{r+1} = \mu_r$ for some $r \in \mathbb{N}_0$. In such a way, the reference configuration h_n is $\mu_n^{(r)}$ for each $n = 1, ..., N$.

5 Towards Detecting Abnormal Behavior on the Network

Note that in Proposition 1 we obtain the estimator $\overline{\sigma}_t(X_t) = \overline{\mathbb{E}}[\overline{\Lambda}_t X_t \mid \mathcal{Y}_t]$ and then, by applying the Bayes' rule of eq. (15), we obtain the filtered estimate $\sigma_t(X_t) = \mathbb{E}[X_t \mid \mathcal{Y}_t]$ which is the probability distribution $\big(\mathbb{P}[X_t = e_1 \mid \mathcal{Y}_t], ..., \mathbb{P}[X_t = e_N \mid \mathcal{Y}_t]\big)$. In order to know the exact estate of the temporal evolution of the grid over the reference configurations $h_1, ..., h_N$, we then take for each time $t \geq 0$, the value

$$\widehat{X}_t = e_n \qquad , \qquad n \in \underset{n'=1,...,N}{\arg\max} \langle \sigma_t(X_t), e_{n'} \rangle \,. \tag{34}$$

Considering this choice, at fixed time $t \geq 0$, $\langle \widehat{X}_t, e_n \rangle = 1$ and $\langle \widehat{X}_t, e_{n'} \rangle = 0$ for each $n' = 1, ..., N$, $n' \neq n$, then by eq. (2), the estimated grid state $\widehat{H}_t$ is at reference configuration h_n at such time $t \geq 0$.

Concerning the detection of abnormal behaviors on the grid, we can use the estimator $\sigma_t(X_t)$ for each $t \geq 0$ in the following way. First, for each $t \geq 0$, let

$$\epsilon_t \;\coloneqq\; 1 \;-\; \max_{n=1,...,N} \langle \sigma_t(X_t), e_n \rangle \tag{35}$$

be the function that represents how far $\sigma_t(X_t)$ is from the value one. Recall that this estimator is a probability distribution. So, for a fixed $t \geq 0$, if one component of $\sigma_t(X_t)$ is near to one, then selecting the reference configuration at which the grid is (by using eq. (34)), is an almost-sure choice. Thus, for any time $t \geq 0$, the function ϵ_t represents the temporal evolution's uncertainty on the reference configurations. This can be, therefore, embedded in a monitoring algorithm for detecting abnormal behavior.

Two approaches are analyzed. First, we define a threshold $\delta_t \in [0, 1]$ up to time $t \geq 0$,

$$\delta_t \;\coloneqq\; \frac{1}{t} \int_0^t \epsilon_s \, ds \,. \tag{36}$$

Then if $\epsilon_t > \delta_t$ for a considerable amount of time, this will be an abnormal behavior of the grid because there is an uncertainty concerning which reference configuration the grid is. Second, we look for the function $\gamma : [0, 1] \to [0, 1]$ defined for each $\xi \in [0, 1]$, by

$$\gamma(\xi) \;\coloneqq\; \frac{1}{t} \int_0^t \mathbb{1}_{\{\epsilon_s \geq \xi\}} ds \,. \tag{37}$$

Thus, for each fixed threshold $\xi \in [0, 1]$, $\gamma(\xi)$ represents the percentage of time that ϵ_t exceeds ξ. Obviously, if $\xi = 0$, then $\epsilon_t >= \xi$ for any time $t \geq 0$. Thus, $\gamma(\xi) = 1$ quantifies this case from $[0, t]$. The main idea is then to fix a "good" threshold ξ near to zero such that $\gamma(\xi)$ is also near to zero. In such a way, ϵ_s is small enough for each $s \in [0, t]$ which represents the certitude at which reference configuration the grid is.

6 Numerical Results

In this section, we present our model's numerical results to estimate the hidden temporal evolution of the electrical transmission grid. Basically, we consider boolean temporal sequences describing the breakers' states (*off*/*on*) of the network. We evaluate a simulated scenario in which the Markov state process that represents the network, is known. Recall that only the observed processes are used to infer the hidden process. After the temporal evolution of the hidden process is estimated, we compare it with the "real" state process of the network for validation.

6.1 Simulated Data

The simulated scenario is performed by the initial parameters shown in next.

6.1.1 Initial Parameters

We fix first the space of $N = 4$ reference states $\mathbb{H} = \{h_1, h_2, h_3, h_4\}$, by choosing the states for the $P = 6$ breakers as *off*/*on*, i.e., $M = 2$. In this way, the space of the observed states is $\mathbb{Y} = \{f_1, f_2\}$, where $f_1 = (1, 0)$ and $f_2 = (0, 1)$. The space $\mathbb{H}$ for the simulated scenario is fixed to be:

$$\mathbb{H} = \big\{(1,1,1,1,0,1),\ (0,1,0,0,0,1),\ (1,0,1,1,1,1),\ (0,1,0,1,0,0)\big\}\ . \tag{38}$$

Thus, the space of $\mathbb{X} = \{e_1, e_2, e_3, e_4\}$ stands for the space of canonical vectors of $\mathbb{R}^4$, where $e_1 = (1, 0, 0, 0)$, $e_2 = (0, 1, 0, 0)$, $e_3 = (0, 0, 1, 0)$, $e_4 = (0, 0, 0, 1)$. The simulation of $X = \{X_t\}_{t\geq 0}$ and $Y^p = \{Y_t^p\}_{t\geq 0}$, $p = 1, ..., 6$, are obtained by the classical simulation procedure of jump chains and holding times with exponential distribution, see, e.g., [23, Section 2.6]. The Markov processes are performed under the fixed matrices of eq. (39) and Table 1 in A.1.1. The sample path for X is shown on the left side in Figure 5. This simulation was stopped at 50 jump-events, giving a horizon time of two years from 2018 to 2020. The total number of jump-events of the observed breakers is 1964. Each simulated temporal evolution of the breakers' states is shown in Figure 4.

Concerning the parameters of our model, we compute the initial estimation of the matrices $\widehat{A}^{(0)}$ and $\widehat{C}_m^{p(0)}$, $p = 1, ..., 6$, $m = 0, 1$, by the empirical estimation of eq. (30). These values are shown in eq. (40) and Table 2 in A.1.2, resp. The initial filter estimate of X is obtained from eq. (31). Under the simulated data, we obtain $\overline{\sigma}_0(\widehat{X}_0) = (0.307, 0.196, 0.388, 0.109)$. The initial state for X is therefore chosen to be $\widehat{X}_0 = e_3$ by eq. (34). Thus, by eq. (2), the initial state of the grid is $h_3 = (1, 0, 1, 1, 1, 1)$. Finally, for the stopping criteria, we choose $\varepsilon = 10^{-5}$, see eq. (32).

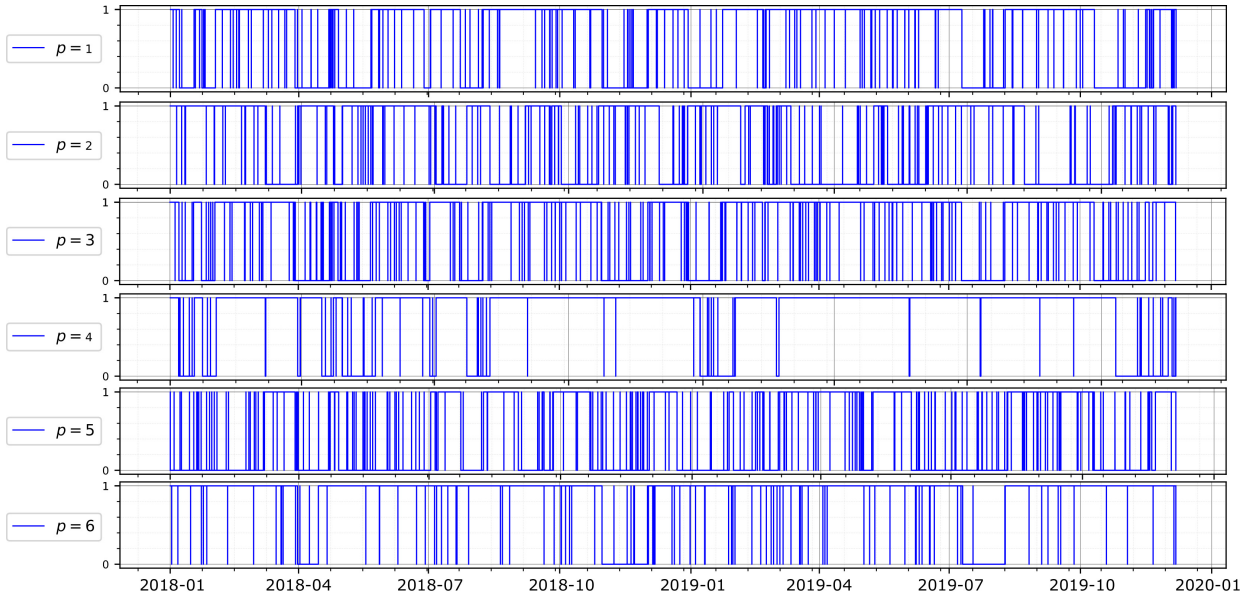

Fig. 4: Simulated breakers state's temporal evolution.

6.1.2 Grid State Estimation by Clustering Method

On the left side in Figure 5, we observe the "real" temporal evolution of the hidden Markov process X. This graphic represents the grid state behavior over time. The first estimation that we do is the empirical estimation by clustering method by using eq. (29). This approach is a classification procedure that takes the joint information of all breakers states at each time and computes the arg min set to know at which cluster the grid is. A cluster is obtained by finding

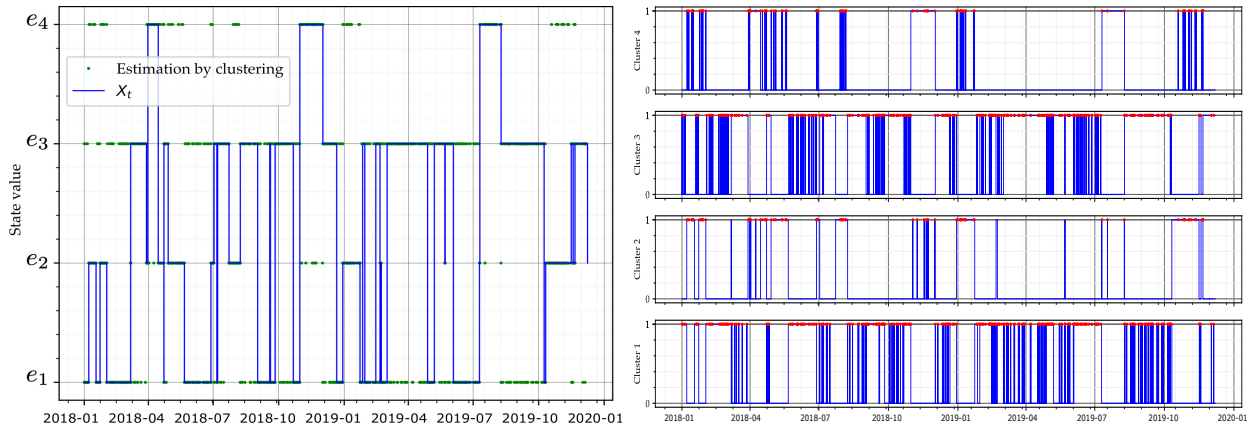

Fig. 5: On the left side, a sample path of the hidden Markov state process X (blue line) representing the grid state over time, and the estimation by clustering method (green points). At several times, the arg min set for clustering could not be a singleton. On the right side, the temporal evolution of the active clusters (blue lines) representing when the arg min set has more than one point (red points).

the points of the breakers' joint information that have the minimum distance to a cluster center, see eq. (33), where the centers are the reference configurations of eq. (38). At several times, the arg min set might not be a single point because there are points of the joint information of all breakers states at the same distance of different reference configurations of the grid. In such cases, the clustering method is not exact because we cannot know at which configuration the grid is. This is represented with green points on the left side in Figure 5. When a cluster is active, i.e., when the reference configurations can be chosen by clustering, it is represented on the right side in Figure 5. This is shown with a red point in the same picture when there is more than one choice. The metric distance used is the euclidean distance.

6.1.3 **Grid State Estimation by HMM**

Figure 6 shows the filter estimate of the grid's temporal evolution over the reference configurations. The hidden process X represents this over the canonical vectors of $\mathbb{R}^4$. On the left side in Figure 6, the filtering is done using the (empirical) initial parameter estimation of eq. (30). At first glance, the filter estimate "jumps" several times when the "real" temporal evolution of the grid does not. This is because the first parameter estimation is not exact. However, when the parameter estimation is computed through the Theorem 1, the filter estimate fits better as the number of iterations increases. This is confirmed by the Mean Squared Error (MSE) between the "real" values of X, and the filter estimate $\widehat{X}$. The MSE values over the number of iterations are plotted in Figure 7 on the left side, showing the accuracy of our model over the simulated scenario. The filter estimate for the last iteration is shown in Figure 6 on the right side.

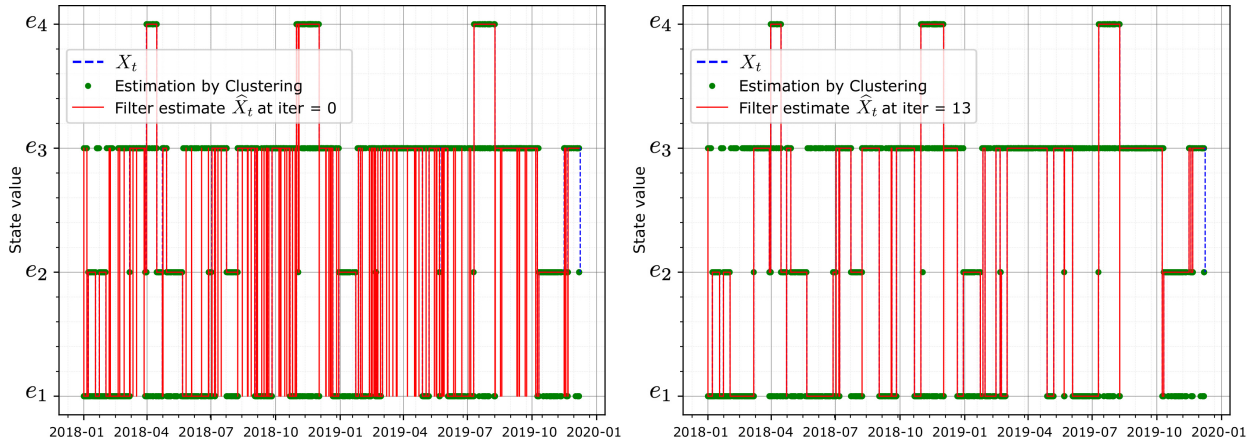

Fig. 6: Filter estimate for the temporal evolution of the grid represented by the hidden process X. The "real" evolution is shown by the blue dashed line. The filter is obtained by the parameter estimation at the first iteration (on the left side) and the final one (on the right side).

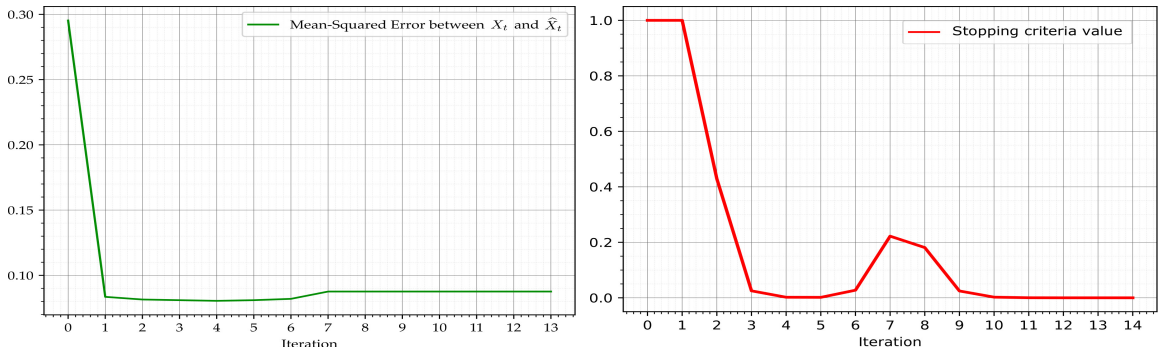

Fig. 7: Mean Squared Error (MSE) between X and the filter estimate $\widehat{X}$ (on the left side) and the stopping criteria values of the eq. (32) (on the right side).

6.1.4 **Parameter Estimation**

Figure 8 and Figure 9 show the values of the parameters of our model over the number of iterations, i.e., the values of the matrices $\widehat{A}^{(r)}$ and $\widehat{C}_m^{p(r)}$, $p = 1, ..., 6$, $m = 0, 1$, for each iteration $r = 0, ..., 13$. The first estimation is made by eq. (30) and the last one is obtained when the stopping criteria of eq. (32) is verified with $\varepsilon = 10^{-5}$. These latter values are shown in Figure 7.

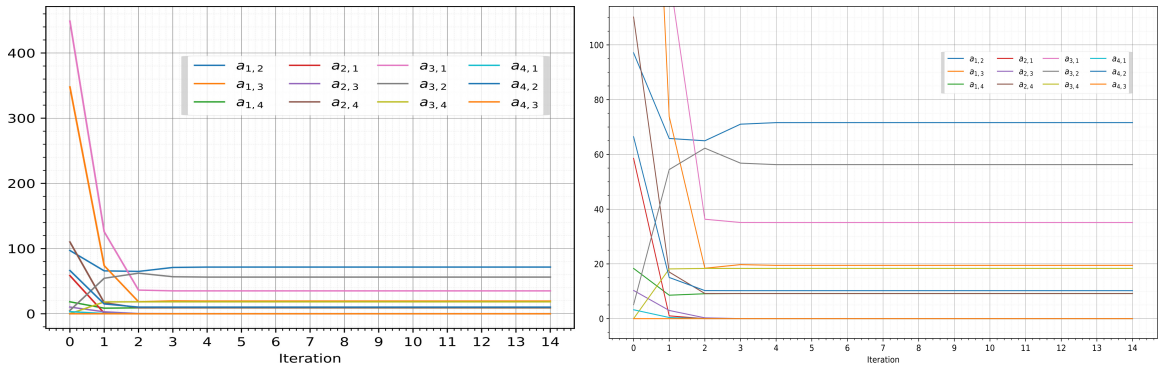

Fig. 8: Estimated values for the matrix A on the left side, and a zoom on these values on the right side.

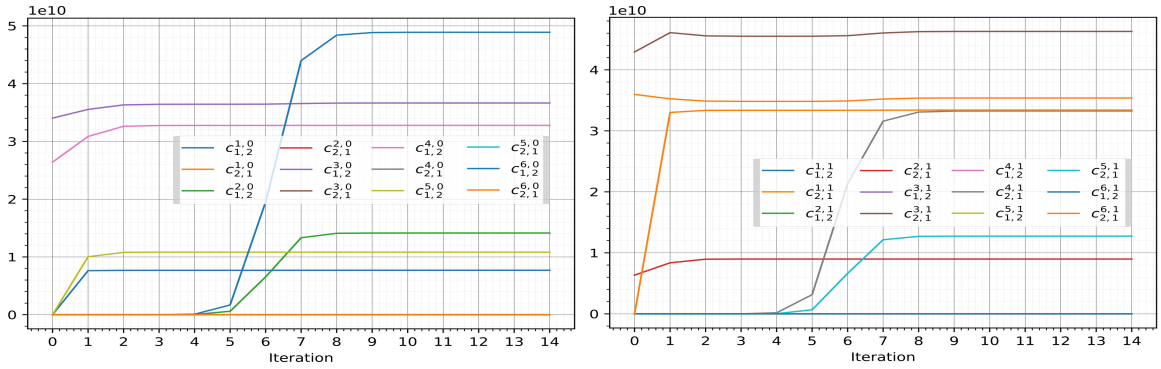

Fig. 9: Estimated values for the matrices C_0^p (on the left side) and C_1^p (on the right side) for each breaker $p = 1, ..., 6$.

6.1.5 **Towards Detecting Abnormal Behaviors in the Grid**

Using the filter estimate for the grid state, we can compute the function ϵ_t of eq. (35), that represents a temporal evolution's uncertainty signal on the reference configurations. Figure 10 shows the values of this uncertainty over time for the first and the last iterations. Comparing the two pictures, we observe that there are fewer peaks at the last iteration compared to the first iteration. This is because finding the optimal parameters to fit our model, the uncertainty decreases as the number of iterations increases. This is reflected in Figure 11 that shows the threshold δ_t values of eq. (36) over the number of iterations. Such a threshold represents the amount of time for which the uncertainty is great. Observing Figure 11, we conclude that from the iteration 1 (at the first estimation of all parameters by Theorem 1), less than 1% of the time, ϵ_t is over the threshold δ_t. When ϵ_t is over δ_t means that the state at which the grid "jumps" is known with some uncertainty. However, such uncertainty is almost instantaneous because it is not remaining in time, as reflected in Figure 10 and Figure 11.

On the other hand, Figure 12 shows γ in function of $\xi \in [0, 1]$, see eq. (37), which represents the percentage of time that ϵ_t exceeds a fixed ξ. The main difference between the threshold ξ and δ_t is that the latter is computed in function of ϵ_t. We observe in Figure 12 that at iteration 0 (with the empirical estimation of the parameters), almost 20% of the uncertainty signal is over the threshold $\xi = 0.15$. Fixing $\xi = 0.6$, almost all the uncertainty signal is below

ξ, as it is shown on the left side in Figure 10. On the other hand, fixing ξ near to zero such that $\gamma(\xi)$ is also near to zero means that ϵ_t is small enough at each time so that the uncertainty is almost null. This is reflected on the right side of Figure 10. So that, for a small threshold ξ, ϵ_t is almost 0 the most time.

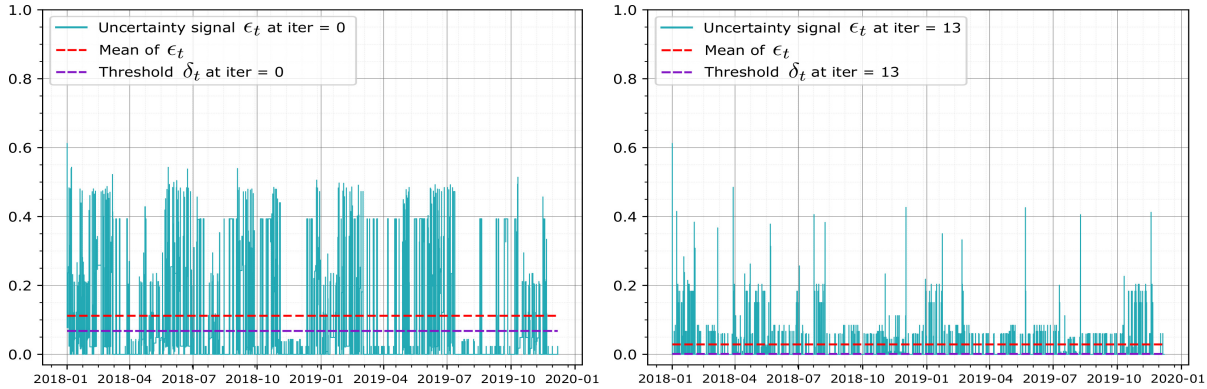

Fig. 10: Temporal evolution's uncertainty signal on the reference configurations of the grid at the first iteration (on the left side) and on the last iteration 13 (on the right side).

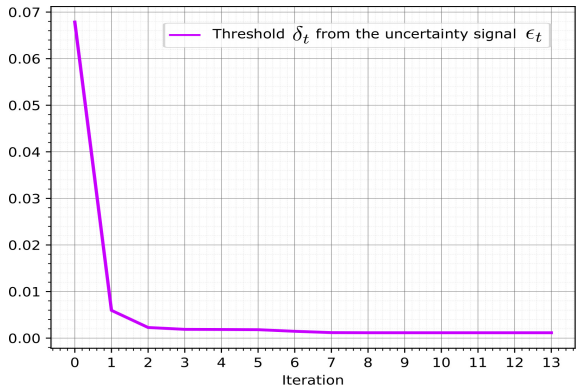

Fig. 11: Values of the threshold δ_t of eq. (36) for anomaly detection.

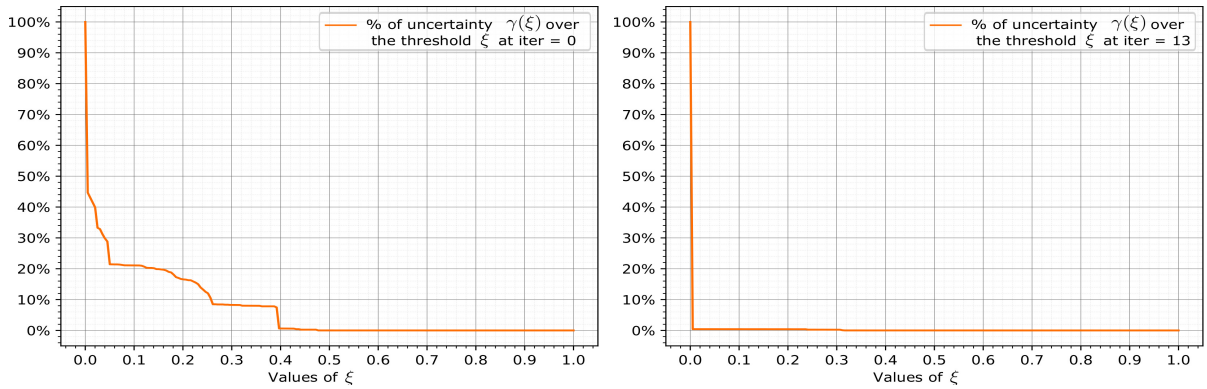

Fig. 12: Values of function γ of eq. (37) over the threshold ξ for anomaly detection.

7 Conclusion and Remarks

In this paper, we have proposed a general data-driven approach for the temporal evolution modeling of networks. While the application is based on the breakers' states in an electrical transmission grid, we believe the model is general enough to serve other types of dynamic networks. Our framework was based on a continuous-time finite-state Hidden Markov Model (HMM) driven by multiple-observed counting processes. The central assumption in the application was that the grid's state varies around a finite set of reference configurations. The grid's current reference configuration is unknown and constitutes the hidden state, while each breakers' state is an observable process. We have provided a filter-based expectation-maximization approach using a change of probability measure method to estimate recursively the model parameters and the hidden reference configuration of the grid. Filter estimates are also given for various processes related to the Markov state processes.

Further, we have presented a strong scheme with no discretization error for a general filter dynamic for numerical purposes. The state change effects of the breakers are then added at the correct "jump" times. In addition, a clustering approach was also presented to identify the set of reference configurations of the grid. Using our theoretical results, we have then shown the performance of the framework by considering boolean temporal sequences describing the breakers' states (off/on) of the grid. We have evaluated a simulated scenario, showing the advantages of the HMM approach with the proposed strong scheme. We finally identify the normal behavior of the French electrical grid, which will be embedded in a monitoring and detection algorithm in the future.

A Parameter Values

In this section, we present the matrices used in the Section 6 of our model's numerical results.

A.1 Simulated Data

A.1.1 Matrices for Markov Process Simulation

This section presents the matrices A and C_m^p, $p = 1, ..., 6$, $m = 0, 1$, used for the simulations of the Markov states processes X and Y^p, resp. in Section 6.1.

$$A = 10^{-7} \begin{bmatrix} -8.207 & 3.346 & 3.263 & 1.598 \\ 2.821 & -9.286 & 4.657 & 1.808 \\ 3.047 & 3.930 & -7.057 & 0.080 \\ 2.687 & 1.049 & 2.266 & -6.002 \end{bmatrix} \tag{39}$$

C_m^p	$m=0$	$m=1$
$p=1$	$\begin{bmatrix} -2.55\times10^{-6} & 2.55\times10^{-6} \\ 1.53\times10^{2} & -1.53\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -5.10\times10^{2} & 5.10\times10^{2} \\ 2.45\times10^{-6} & -2.45\times10^{-6} \end{bmatrix}$
$p=2$	$\begin{bmatrix} -3.15\times10^{-6} & 3.15\times10^{-6} \\ 2.62\times10^{2} & -2.62\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -1.55\times10^{2} & 1.55\times10^{2} \\ 3.30\times10^{-6} & -3.30\times10^{-6} \end{bmatrix}$
$p=3$	$\begin{bmatrix} -3.60\times10^{-6} & 3.60\times10^{-6} \\ 5.42\times10^{2} & -5.42\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -8.58\times10^{2} & 8.58\times10^{2} \\ 3.15\times10^{-6} & -3.15\times10^{-6} \end{bmatrix}$

C_m^p	$m=0$	$m=1$
$p=4$	$\begin{bmatrix} -3.80\times10^{-6} & 3.80\times10^{-6} \\ 4.76\times10^{2} & -4.76\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -5.49\times10^{2} & 5.49\times10^{2} \\ 1.50\times10^{-7} & -1.50\times10^{-7} \end{bmatrix}$
$p=5$	$\begin{bmatrix} -3.85\times10^{-6} & 3.85\times10^{-6} \\ 1.78\times10^{2} & -1.78\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -1.86\times10^{2} & 1.86\times10^{2} \\ 2.00\times10^{-6} & -2.00\times10^{-6} \end{bmatrix}$
$p=6$	$\begin{bmatrix} -1.60\times10^{-6} & 1.60\times10^{-6} \\ 7.52\times10^{2} & -7.52\times10^{2} \end{bmatrix}$	$\begin{bmatrix} -5.82\times10^{2} & 5.82\times10^{2} \\ 1.65\times10^{-6} & -1.65\times10^{-6} \end{bmatrix}$

Table 1: Simulated C_m^p matrix for the observed breakers.

A.1.2 Empirical Initial Estimation for HMM

In this section, we present the empirical initial estimation of the matrices $\widehat{A}^{(0)}$ and $\widehat{C}_m^{p(0)}$, $p = 1, ..., 6$, $m = 0, 1$, used for the first iteration of the filter estimate by HMM, in the simulated case.

$$\widehat{A}^{(0)} = 10^{-7} \begin{bmatrix} -5.107\times10^{2} & 9.710\times10 & 3.480\times10^{2} & 1.836\times10 \\ 5.855\times10 & -1.686\times10^{2} & 1.031\times10 & 1.102\times10^{2} \\ 4.489\times10^{2} & 5.11 & -3.583\times10^{2} & 1.000\times10^{-3} \\ 3.253 & 6.643\times10 & 1.000\times10^{-3} & -1.285\times10^{2} \end{bmatrix} \tag{40}$$

$\widehat{C}_m^{p(0)}$	$m=0$	$m=1$
$p=1$	$\begin{bmatrix} -1.84\times10^{2} & 2.88\times10^{5} \\ 1.84\times10^{2} & -2.88\times10^{5} \end{bmatrix}$	$\begin{bmatrix} -3.60\times10^{10} & 1.68\times10^{2} \\ 3.60\times10^{10} & -1.68\times10^{2} \end{bmatrix}$
$p=2$	$\begin{bmatrix} -1.68\times10^{2} & 1.00\times10^{3} \\ 1.68\times10^{2} & -1.00\times10^{3} \end{bmatrix}$	$\begin{bmatrix} -6.36\times10^{9} & 3.12\times10^{2} \\ 6.36\times10^{9} & -3.12\times10^{2} \end{bmatrix}$
$p=3$	$\begin{bmatrix} -2.59\times10^{2} & 3.40\times10^{10} \\ 2.59\times10^{2} & -3.40\times10^{10} \end{bmatrix}$	$\begin{bmatrix} -4.29\times10^{10} & 2.22\times10^{2} \\ 4.29\times10^{10} & -2.22\times10^{2} \end{bmatrix}$

$\widehat{C}_m^{p(0)}$	$m=0$	$m=1$
$p=4$	$\begin{bmatrix} -2.61\times10^{2} & 2.64\times10^{10} \\ 2.61\times10^{2} & -2.64\times10^{10} \end{bmatrix}$	$\begin{bmatrix} -7.06\times10^{2} & 3.20\times10 \\ 7.06\times10^{2} & -3.20\times10 \end{bmatrix}$
$p=5$	$\begin{bmatrix} -3.72\times10^{2} & 7.26\times10^{3} \\ 3.72\times10^{2} & -7.26\times10^{3} \end{bmatrix}$	$\begin{bmatrix} -1.00\times10^{3} & 1.88\times10^{2} \\ 1.00\times10^{3} & -1.88\times10^{2} \end{bmatrix}$
$p=6$	$\begin{bmatrix} -1.01\times10^{2} & 1.00\times10^{3} \\ 1.01\times10^{2} & -1.00\times10^{3} \end{bmatrix}$	$\begin{bmatrix} -2.74\times10^{5} & 1.24\times10^{2} \\ 2.74\times10^{5} & -1.24\times10^{2} \end{bmatrix}$

Table 2: Initial estimation of the matrix $\widehat{C}_m^{p(0)}$, $m = 0, 1$, for each breaker $p = 1, ..., 6$.

References

1. W. Wang, Z. Lu, Cyber security in the smart grid: Survey and challenges, Computer networks 57 (5) (2013) 1344–1371.
2. C.-C. Sun, A. Hahn, C.-C. Liu, Cyber security of a power grid: State-of-the-art, International Journal of Electrical Power & Energy Systems 99 (2018) 45–56.
3. R. J. Elliott, L. Aggoun, J. B. Moore, Hidden Markov models: estimation and control, Vol. 29, Springer Science & Business Media, 2008.
4. D. Shi, R. J. Elliott, T. Chen, On finite-state stochastic modeling and secure estimation of cyber-physical systems, IEEE Transactions on Automatic Control 62 (1) (2016) 65–80.
5. W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, Journal of the Society for Industrial and Applied Mathematics, Series A: Control 2 (3) (1964) 347–369.
6. M. Zakai, On the optimal filtering of diffusion processes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 11 (3) (1969) 230–243.
7. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22.
8. G. J. McLachlan, T. Krishnan, The EM algorithm and extensions, Vol. 382, John Wiley & Sons, 2007.
9. P. Brémaud, Point processes and queues: martingale dynamics, Vol. 50, Springer, 1981.
10. H.-H. Bock, Clustering methods: a history of K-means algorithms, in: Selected contributions in data analysis and classification, Springer, 2007, pp. 161–172.
11. L. Aggoun, R. J. Elliott, Finite-dimensional models for hidden Markov chains, Advances in applied probability 27 (1) (1995) 146–160.
12. R. J. Elliott, W. P. Malcolm, Discrete-time expectation maximization algorithms for Markov-modulated poisson processes, IEEE Transactions on Automatic Control 53 (1) (2008) 247–256.
13. M. R. James, V. Krishnamurthy, F. Le Gland, Time discretization of continuous-time filters and smoothers for HMM parameter estimation, IEEE Transactions on Information Theory 42 (2) (1996) 593–605.
14. V. Krishnamurthy, J. Evans, Finite-dimensional filters for passive tracking of Markov jump linear systems, Automatica 34 (6) (1998) 765–770.
15. L. Aggoun, R. J. Elliott, Measure theory and filtering: Introduction and applications, Vol. 15, Cambridge University Press, 2004.
16. K. Sennewald, K. Wälde, "Itô's lemma" and the bellman equation for poisson processes: An applied view, Journal of Economics 89 (1) (2006) 1–36.
17. F. Campillo, F. Le Gland, Mle for partially observed diffusions: direct maximization vs. the EM algorithm (1988).
18. A. Dembo, O. Zeitouni, Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm, Stochastic Processes and their Applications 23 (1) (1986) 91–113.
19. C. J. Wu, On the convergence properties of the EM algorithm, The Annals of statistics (1983) 95–103.
20. E. Platen, N. Bruti-Liberati, Numerical solution of stochastic differential equations with jumps in finance, Vol. 64, Springer Science & Business Media, 2010.
21. D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, Tech. rep., Stanford (2006).
22. S. S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering, Pattern recognition letters 25 (11) (2004) 1293–1302.
23. J. R. Norris, J. R. Norris, Markov chains, no. 2, Cambridge university press, 1998.