



HAL
open science

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

Mauricio Gonzalez, Antoine Girard

► **To cite this version:**

Mauricio Gonzalez, Antoine Girard. Data-Driven Modeling of the Temporal Evolution of Breakers' States in the French Electrical Transmission Grid. *Nonlinear Analysis: Hybrid Systems*, 2022, 46, pp.101215. 10.1016/j.nahs.2022.101215 . hal-03402283v2

HAL Id: hal-03402283

<https://hal.science/hal-03402283v2>

Submitted on 29 Jun 2022

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

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

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

Mauricio Gonzalez^{1a}, Antoine Girard^b

^a*Quivalio, Quivalio Analytics, R&D Team. 11 avenue Delcassé, 75008 Paris, France*

^b*Université Paris-Saclay, CNRS, CentraleSupélec, L2S. 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France*

Abstract

In electrical transmission grids, it is common to observe the states of circuit breakers. While they are known at irregular times, system modeling and grid state estimation are of the highest importance to ensure secure operations. This paper proposes a richer method to estimate the grid state over its reference configurations based on the temporal evolution of its breakers' states. The first contribution consists in developing a general multi-observation continuous-time finite-state Hidden Markov Model with filter-based parameter estimation to infer the hidden state (e.g., the grid reference configuration) handling multiple observed processes with irregular “jump” times (e.g., the breakers' states). As a second contribution, we build a numerical scheme with no discretization error adapted to all state jumps generated by the observed processes. Finally, we apply our model to simulated and real data to illustrate the approach's performance. The available data consists of historical records of breakers' states during the electrical transmission grid operated normally. For this real-data-driven application, we also present a clustering approach to identify the set of grid reference configurations.

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

1. Introduction

An *electrical transmission grid* is an interconnected network that permits the electrical energy movements from producers, e.g., 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 grid configuration is determined by all breakers' states, which can be used to adapt the grid to diverse operating conditions. However, while the breakers' states are known at all irregular times, the current grid configuration is assumed to be unknown. As the grid can be prone to failures or malicious attacks, system modeling is highly important to ensure secure operations [1, 2]. This paper proposes an ad-hoc Hidden Markov Model (HMM) [3] to estimate the state temporal evolution of the grid over its reference configurations that maximize the likelihood of observing the different breakers' states.

The literature has generally investigated system state estimation in the discrete-time framework under the (most recurrent) assumption of partially observing the state in Gaussian noise [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. However, applying this standard modeling is not appropriate for our problem. In fact, we need to consider two specific characteristics of our framework. First, breakers' states switch between the modes “*off/on*,” which may happen at any time. Hence, our model must handle breakers' state changes at any time, which is a feature of the continuous-time modeling. Second, we observe several breakers. Hence, we need to extend standard HMMs [3, 15, 16] to handle multiple observed stochastic processes with irregular

¹*Previous address:* Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France.

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

Email: mauricio.gonzalez@outlook.fr (Mauricio Gonzalez), antoine.girard@centralesupelec.fr (Antoine Girard)

18 “jump” times. We then propose a general multi-observation HMM with parameters fitted (that define the
19 transition rate matrices of the hidden and observed state processes) based on the standing assumption that
20 the finite-state spaces of all stochastic processes are known. Our continuous-time finite-state HMM is then
21 applied to our framework as a running example. Precisely, the HMM drives the grid and consists of a hidden
22 Markov process (the reference configuration) and multiple observed Markov processes (the breakers’ states).
23 This approach may be potentially applicable in a system monitoring scheme.

24 Some works focus on different assumed-types of malicious attacks to design system monitoring on elec-
25 trical grids. For instance, *replay attack* is considered in [4, 5, 6], in which an attacker hijack sensors, observe
26 and record the outputs. *Denial-of-Service* (DoS) attack models were assumed in [7, 8], in which an optimal
27 control problem under security constraints is solved. *False data injection* was considered in [9, 10, 11, 17], in
28 which some attacker can inject error measurements in the state estimation. For a study considering all these
29 types of attacks, see, e.g., [12]. However, it is difficult to know a priori the attack type on a system, and the
30 assumptions made by these works may not be close to reality. In [14, 17, 18], it had been analyzed the case
31 when an arbitrary error or unknown parameter is additively injected on the state model or measurement
32 to represent, e.g., malicious attacks. However, in these works, it is common to assume that all the model
33 parameters are known to estimate the system state. That is not the case in our problem because we need to
34 estimate the parameters. While our HMM is later intended to be included in online monitoring algorithms
35 to detect abnormal behaviors (e.g., as in [13]), the current paper mainly focuses on modeling and state and
36 parameter estimation problems.

37 We provide an iteratively filter-based Expectation-Maximization (EM) approach [19, 20, 21, 22] to esti-
38 mate the model parameters and the hidden state (e.g., the reference configuration). This approach aims to
39 maximize a log-likelihood function over parameter space. While our parameter estimation method is close
40 as obtained in previously cited papers and [3], we adapt the filter-based approach to handle multi-observed
41 processes. Each one has its transition rate matrix (i.e., parameters), and no average is considered over them
42 as in [23]. In detail, we suppose first that all state processes belong in a probability space representing the
43 “real world”. Then, we use a change of probability measure technique (Girsanov’s Theorem, see, e.g., [3, 24])
44 to define a new probability measure representing a “fictitious world”. In this new space, filters for estimat-
45 ing the hidden state and the model’s parameters are easy to obtain. They are linear Stochastic Differential
46 Equations (SDEs) modulated by counting processes. Then, instead of using a classical Euler-Maruyama
47 discretization (with small-time step) for all the SDEs obtained (e.g., as in [16, 25, 26]), we present a strong
48 scheme with no discretization error for numerical purposes. This scheme adapts to all state change times
49 generated by the temporal evolution of all observed processes.

50 The proposed modeling approach is finally confronted with available real data provided by France’s
51 transmission system operator (RTE), consisting of Boolean temporal sequences describing a set of breakers’
52 states (*off/on*). The data has been collected during the normal behavior of the network over a given period.
53 The reference configurations are also obtained from this data using a version of the well-known K -means
54 method [27]. We then identify the normal behavior of the French electrical grid, which could be embedded
55 in a monitoring and detection algorithm in the future.

56 The rest of this paper is organized as follows. In Section 2, we introduce the general finite-state
57 continuous-time HMM and dynamics. Section 3 provides a filtering approach to estimate the hidden state
58 based on all observed processes. We also briefly recall the EM algorithm and compute a filter-based EM
59 algorithm for all model parameters. In Section 4, we present a numerical method for all filter estimates. We
60 also show how to obtain the initial estimation of all parameters. Section 5 shows the numerical results in
61 a simulated scenario and the application of the breakers’ states in the French electrical transmission grid,
62 which illustrates a real-world application of the HMM considered in this paper. The final Section 6, gives
63 concluding remarks.

2. The Modeling

Let us assume that the grid has finitely many reference configurations h_1, \dots, h_N known *a priori*, $N \in \mathbb{N}^*$. Each reference configuration is a vector of length $P \in \mathbb{N}^*$ (the number of breakers in the grid), and each of its components represents some information of a breaker (e.g., modes, states, etc.). Suppose that a state process $H := \{H_t\}_{t \geq 0}$ represents the evolution over time of the grid between the reference configurations h_1, \dots, h_N , where H_t denotes the unknown reference configuration at time $t \geq 0$. Then, H constitutes a hidden process, i.e., it is not directly observable. Suppose that we observe each breaker state over time represented by the state process $K^p := \{K_t^p\}_{t \geq 0}$. Then, K_t^p denotes the available information of the breaker $p = 1, \dots, P$ 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 H in some optimal way based on the temporal evolution of K^p , $p = 1, \dots, P$.

First, 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 (MCs) 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.

To be more general, we work with general hidden and observable processes. The case study considered of the electrical grid is a running example while explaining the general proposed approach.

2.1. The State Processes

Consider that the state space of the hidden process $H = \{H_t\}_{t \geq 0}$ is the finite set:

$$\mathbb{H} := \{h_1, h_2, \dots, h_N\} \subseteq \prod_{p=1}^P \mathbb{M}_p, \quad (1)$$

where, for each $p = 1, \dots, P$,

$$\mathbb{M}_p := \{\bar{m}_1^p, \bar{m}_2^p, \dots, \bar{m}_{M_p}^p\} \subset \mathbb{R}^R \quad (2)$$

represents a finite set, with $M_p, R \in \mathbb{N}^*$.

Example (Electrical grid). *The sets \mathbb{H} and \mathbb{M}_p can represent the set of known reference configurations of the grid and the set of different modes or states that the breaker p can eventually take, respectively (resp.). For instance, if we consider $M_p = 2$, then $\mathbb{M}_p = \{\bar{m}_1^p, \bar{m}_2^p\}$. Thus, if e.g., we assume $R = 1$, then \mathbb{M}_p represents the breaker states “off” and “on”, i.e., we can let $\bar{m}_1^p = 0$ and $\bar{m}_2^p = 1$, resp.; and \mathbb{H} is therefore a set of binary vectors. On the contrary, if we assume $R = 2$, then \mathbb{M}_p represents breaker 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 breaker’s state values (“off” and “on”) and second the time spent at each breaker’s state value (“off” and “on”), both on average in one hour.* \square

In order to simplify the modeling of the hidden process, we introduce a transformation over H . Let $\mathbb{1}_n : \mathbb{H} \rightarrow \{0, 1\}$ the indicator function defined for each $n = 1, \dots, N$ by $\mathbb{1}_n(h_m) = 1$ if $n = m$, and $\mathbb{1}_n(h_m) = 0$ otherwise. Then, the vector $(\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_N)$ is a bijection from \mathbb{H} to the set of unit vectors $\mathbb{X} := \{e_1, e_2, \dots, e_N\} \subset \mathbb{R}^N$, where $e_n \in \mathbb{X}$ denotes a vector in \mathbb{R}^N with unity in the n -th position and zero elsewhere, $n = 1, \dots, N$. Thus, without loss of generality (w.l.o.g.), we shall consider an underlying state process $X := \{X_t\}_{t \geq 0}$ with state space \mathbb{X} and defined by

$$X_t := (\mathbb{1}_1(H_t), \mathbb{1}_2(H_t), \dots, \mathbb{1}_N(H_t)). \quad (3)$$

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 that H_t is in the estate $h_n \in \mathbb{H}$. Mathematically, H can be computed as follows:

$$H_t = \sum_{n=1}^N \langle X_t, e_n \rangle h_n. \quad (4)$$

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

106 Similarly, the set of different state values of the observable process $K^p = \{K_t^p\}_{t \geq 0}$, $p = 1, \dots, P$, is defined
 107 by $\mathbb{K}_p := \{0, 1, \dots, K_p - 1\} \subset \mathbb{N}$, where $K_p \in \mathbb{N}^*$. Note in particular that we can consider in eq. (2), $R = 1$,
 108 $M_p = K_p$ and $\mathbb{M}_p = \mathbb{K}_p$. For instance, if $K_p = 2$, then K^p takes the binary values 0 and 1. This set is
 109 identified with the set of unit vectors

$$\mathbb{Y}_p := \{f_1^p, f_2^p, \dots, f_{K_p}^p\} \subset \mathbb{R}^{K_p},$$

110 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, $k = 1, \dots, K_p$.
 111 Thus, w.l.o.g., the temporal evolution of the observable process K^p can be represented by a state process
 112 $Y^p := \{Y_t^p\}_{t \geq 0}$. When Y_t^p is in the state f_k^p , that means that the observable process K_t^p is in state $k - 1 \in \mathbb{K}_p$
 113 at time $t \geq 0$.

114 **Example** (Electrical grid). *If we consider K_t^p as the observable state process of the breaker p at time $t \geq 0$,
 115 then Y_t^p being in the state f_k^p means that breaker p is in state $k - 1 \in \mathbb{K}_p$ at time $t \geq 0$. If $K_p = 2$ for each
 116 $p = 1, \dots, P$, then all breakers take binary values (representing the values “off” and “on”), and $f_1^p = (1, 0)$
 117 and $f_2^p = (0, 1)$ represent the state values $0 \in \mathbb{K}_p$ and $1 \in \mathbb{K}_p$, resp. In addition, from eq. (2), if we consider
 118 $R = 1$, $M_p = K_p = 2$ and $\mathbb{M}_p = \mathbb{K}_p$, then \mathbb{M}_p is also the set of values “off” and “on” and \mathbb{H} is a set of
 119 binary vectors. \square*

120 2.2. The Dynamic of the Hidden Process

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

$$a_{ij} := \left. \frac{d}{dt} \mathbb{P}[X_t = e_i \mid X_0 = e_j] \right|_{t=0}$$

123 represents the transition probability rate of X from state $e_j \in \mathbb{X}$ to the state $e_i \in \mathbb{X}$, for each $i, j = 1, \dots, N$,
 124 $i \neq j$. In addition, the transpose of A belongs to the Q -matrix class³. Thus, defining $p_{i,t} := \mathbb{P}[X_t = e_i]$,
 125 $i = 1, \dots, N$, $t \geq 0$, the probability distribution vector $p_t := (p_{1,t}, p_{2,t}, \dots, p_{N,t})$ satisfies the forward equation
 126 $dp_t/dt = Ap_t$.

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

$$V_t := X_t - X_0 - \int_0^t AX_s ds$$

129 is a $(\mathcal{F}_t, \mathbb{P})$ -Martingale [28, Lemma 2.6.18]. The semi-Martingale representation of X is therefore:

$$X_t = X_0 + \int_0^t AX_s ds + V_t. \quad (5)$$

130 2.3. The Dynamic of the Observed Processes

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

$$C^p(X_t) = \sum_{n=1}^N C^p(e_n) \langle X_t, e_n \rangle. \quad (6)$$

³That is, for each $j = 1, \dots, N$, $\sum_{i=1}^N a_{ij} = 0$ and $a_{ij} \geq 0$, $\forall i \neq j$.

134 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
 135 depends on the local p -th position of the value that H takes, i.e., the p -th position in the vector $h_n \in \mathbb{H}$, see
 136 eq. (1) and eq.(2).

137 Mathematically, let $\text{proj}_p : \mathbb{H} \rightarrow \mathbb{M}_p$ the p -th projection function. We denote by $H_t^p := \text{proj}_p(H_t)$ the p -th
 138 projection in H_t at time $t \geq 0$, i.e., the value of H_t^p is $\bar{m}_m^p \in \mathbb{M}_p$ for some $m = 1, \dots, M_p$. In line with [15],
 139 we can relate each observed state process Y^p with the underlying hidden process X by its transition rate
 140 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
 141 be expressed as the sum of the eq. (6). In our case, we can obtain for each $n = 1, \dots, N$ an expression of
 142 the matrix $C^p(e_n) = (c_{kl}^p(e_n))$ as a function of the p -th position of $h_n \in \mathbb{H}$. Indeed, let $e_n \in \mathbb{X}$. For each
 143 $p = 1, \dots, P$ and $k, l = 1, \dots, K_p$, $k \neq l$, the transition probability rate of Y^p from state $f_l^p \in \mathbb{Y}_p$ to the state
 144 $f_k^p \in \mathbb{Y}_p$ can be expressed by:

$$\begin{aligned} c_{kl}^p(e_n) &= \left. \frac{d}{dt} \mathbb{P}[Y_t^p = f_k^p \mid Y_0^p = f_l^p, X_0 = e_n] \right|_{t=0} \\ &= \left. \frac{d}{dt} \mathbb{P}[Y_t^p = f_k^p \mid Y_0^p = f_l^p, H_0 = h_n] \right|_{t=0} \\ &\quad (\text{by eq. (4), because } \langle X_0, e_n \rangle = 1 \text{ and } \langle X_0, e_{n'} \rangle = 0, \text{ for each } n' = 1, \dots, N, n' \neq n) \\ &= \left. \frac{d}{dt} \mathbb{P}[Y_t^p = f_k^p \mid Y_0^p = f_l^p, H_0^p = \bar{m}_m^p] \right|_{t=0}, \end{aligned}$$

145 where the last equality holds by the local assumption over the vectors $h_n \in \mathbb{H}$. In such a way, for each
 146 $\bar{m}_m^p \in \mathbb{M}_p$, $m = 1, \dots, M_p$, we define the transition rate matrix $C_m^p = (c_{kl}^{p,m})$ of Y^p (whose transpose belongs
 147 to the Q -matrix class), by:

$$c_{kl}^{p,m} := \left. \frac{d}{dt} \mathbb{P}[Y_t^p = f_k^p \mid Y_0^p = f_l^p, H_0^p = \bar{m}_m^p] \right|_{t=0}, \quad (7)$$

148 for each $k, l = 1, \dots, K_p$, $k \neq l$. In addition, based on the eq. (6), we express $C^p(X_t)$ in function of the
 149 number of elements in \mathbb{M}_p , by:

$$C^p(X_t) = \sum_{m=1}^{M_p} C_m^p \sum_{n \in I_m^p} \langle X_t, e_n \rangle, \quad (8)$$

150 where $C_m^p \in \mathbb{R}^{K_p \times K_p}$ is the matrix with components of the eq. (7), and $I_m^p \subseteq I := \{1, \dots, N\}$ is the subset
 151 of indices $n = 1, \dots, N$ for which the p -th position of the vector $h_n \in \mathbb{H}$ is $\bar{m}_m^p \in \mathbb{M}_p$, i.e.,

$$I_m^p := \{n \in I \mid \text{proj}_p(h_n) = \bar{m}_m^p, h_n \in \mathbb{H}, \bar{m}_m^p \in \mathbb{M}_p\}.$$

152 For convenience, we associate with each I_m^p a diagonal matrix $\text{diag}_m^p \in \mathbb{R}^{N \times N}$ defined by:

$$\text{diag}_m^p := \text{diag}\left(\mathbb{1}_{\{1 \in I_m^p\}}, \mathbb{1}_{\{2 \in I_m^p\}}, \dots, \mathbb{1}_{\{N \in I_m^p\}}\right), \quad (9)$$

153 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.

154 **Example** (Electrical grid). *If for the breakers $p = 1, \dots, P$, we consider $M_p = 2$, and each \mathbb{M}_p represents the*
 155 *set of modes of the breaker p , then we look for the index of the reference configurations of the grid $h_n \in \mathbb{H}$*
 156 *in which the values in their p -th position are the modes \bar{m}_1^p and \bar{m}_2^p (which could represent the values “off”*
 157 *and “on”, resp.), and for each $m = 1, 2$, there is therefore a transition rate matrix C_m^p . \square*

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

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

159 is a $(\mathcal{G}_t, \mathbb{P})$ -Martingale [15, Lemma 2.2], where $\mathcal{G}_t := \sigma(X_s, Y_s^p; s \leq t, p = 1, \dots, P)$ represents the right-
 160 continuous complete filtration generated by X and all observed processes Y^p , $p = 1, \dots, P$. The semi-
 161 Martingale representation of Y_t^p is therefore:

$$Y_t^p = Y_0^p + \int_0^t C^p(X_s) Y_s^p ds + W_t^p. \quad (10)$$

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

164 2.4. Summary

165 In summary, H_t represents a hidden MC (e.g., the temporal evolution of the electrical grid) that takes
 166 values in the known set \mathbb{H} of eq. (1) (e.g., a set of reference configurations in the grid). Each element in \mathbb{H}
 167 is a vector constructed from the sets \mathbb{M}_p of eq. (2) (e.g., the set of different modes or states of the breakers
 168 $p = 1, \dots, P$). Instead of estimating H_t over time, we estimate w.l.o.g. the hidden underlying process X_t ,
 169 whose state space is \mathbb{X} of unit vectors of \mathbb{R}^N . This quantities are related by the eq. (4). The definition of X_t
 170 is given by eq. (3) and its semi-Martingale representation is given in eq. (5). The evolution at time $t \geq 0$
 171 of the observable process $p = 1, \dots, P$ is represented by K_t^p (e.g., the available information of the breaker p
 172 in the grid). This process takes numerical values in the set \mathbb{K}_p (e.g., the set of binary values representing the
 173 values “off” and “on” of breaker p). In the same way, we identify this set with the set \mathbb{Y}_p of unit vectors
 174 of \mathbb{R}^{K_p} and we work with the observable underlying process Y_t^p . This is the observable information to
 175 estimate X_t , and then H_t . The semi-Martingale representation of Y_t^p is given by eq. (10).

176 **Example** (Electrical grid). For instance, suppose that there are $P = 4$ breakers in the grid, the beaker’s
 177 states are 0 and 1, and that the reference configurations are $h_1 = (1, 0, 1, 0)$, $h_2 = (0, 1, 0, 1)$, $h_3 = (0, 0, 0, 0)$,
 178 and $h_4 = (1, 1, 1, 0)$. Also suppose that the temporal evolution of each breaker state is as in Figure 1. Then,
 179 under these observations over time, we want to estimate the temporal evolution of the grid represented by
 180 H , but equivalently, using the process X as it is shown in Figure 2. \square

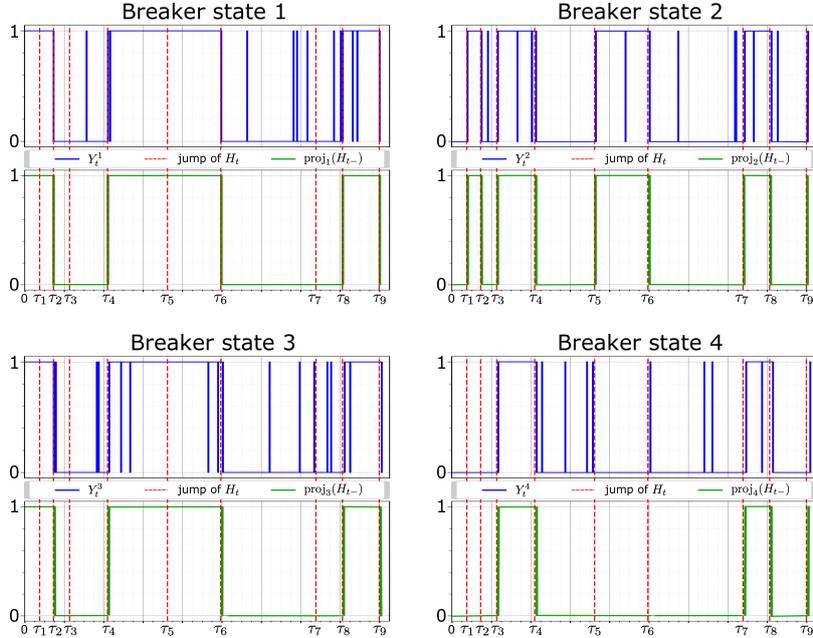


Figure 1: Temporal evolution of breakers' states and its resp. component in the reference configuration of the grid.

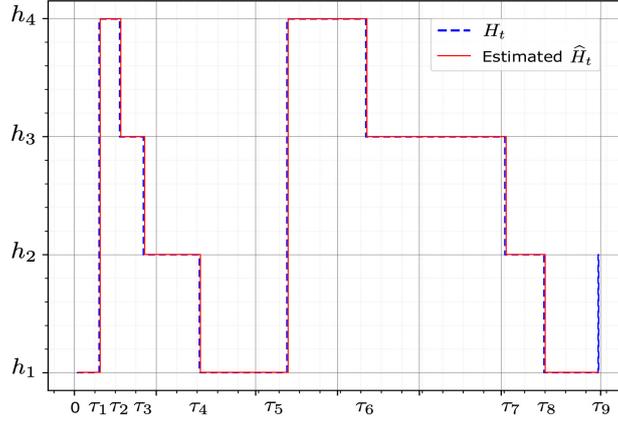


Figure 2: Temporal evolution of the hidden process H and its estimator \hat{H} representing the grid over the reference configurations.

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

In this section, we provide a filtering approach to estimate the hidden process from all observed state processes. To do that, we also need filter estimates for the parameters involved in our model, i.e., the matrices $A = (a_{ij})$ and $C_m^p = (c_{kl}^{p,m})$, for each $p = 1, \dots, P$ and $m = 1, \dots, M_p$, that define resp. the transition rates of X and Y^p .

3.1. Moving to a “Fictitious World”

We are looking for an estimation of the hidden MC over time by using the filtration $\mathcal{Y}_t = \sigma(Y_s^p; s \leq t, p = 1, \dots, P)$ generated by all observed processes Y^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 \mathbb{E} over X_t given \mathcal{Y}_t . In fact, since X_t is defined as an indicator function in eq. (3), the filtered estimate is a conditional probability distribution, i.e., $\mathbb{E}[X_t | \mathcal{Y}_t] = (\mathbb{P}[X_t = e_1 | \mathcal{Y}_t], \dots, \mathbb{P}[X_t = e_N | \mathcal{Y}_t])$. It can be shown that an explicit equation for $\mathbb{E}[X_t | \mathcal{Y}_t]$ 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 linear, as it will be shown below. To obtain $\mathbb{E}[X_t | \mathcal{Y}_t]$ we can use a simple Bayes’ rule.

Suppose that on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there is for each $p = 1, \dots, P$ 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 $[0, t]$, $k, l = 1, \dots, K_p$, $k \neq l$. The semi-Martingale representation of $N_{kl,t}^p$ can be obtained via the following decomposition:

$$\begin{aligned} N_{kl,t}^p &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dY_s^p \rangle \\ &= \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, C^p(X_s) Y_s^p \rangle ds + \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle f_l^p, dW_s^p \rangle, \end{aligned} \quad (11)$$

where, we have used the eq. (10) in differential form, and $Y_{t-}^p := \lim_{s \uparrow t} Y_s^p$ 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 (8), the semi-Martingale representation of $N_{kl,t}^p$ is given by:

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

202 where $M_{kl,t}^p := N_{kl,t}^p - \int_0^t \lambda_{kl,s}^p ds$ is a $(\mathcal{G}_t, \mathbb{P})$ -Martingale [15], and $\lambda_{kl,t}^p$ represents the “ \mathbb{P} -intensity” of the
203 counting process $N_{kl,t}^p$, defined by:

$$\lambda_{kl,t}^p := \langle f_k^p, Y_{t-}^p \rangle \sum_{m=1}^{M_p} c_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_t, e_n \rangle. \quad (12)$$

204 **Example** (Electrical grid). Suppose that $R = 1$, $M_p = K_p = 2$ and $\mathbb{M}_p = \mathbb{K}_p$, then \mathbb{M}_p is a set of binary
205 values. In this context, the process $N_{12,t}^p$ (resp. $N_{21,t}^p$) counts the number of jumps that Y^p does from state
206 $f_1^p = (1, 0)$ (resp., $f_2^p = (0, 1)$), to the state $f_2^p = (0, 1)$ (resp., $f_1^p = (1, 0)$) in the time interval $[0, t]$, i.e., the
207 representation of the number of jumps of the breaker p from state “off” to “on” (resp., from “on” to “off”).
208 From eq. (12), we find back the intuition that a higher transition rate $c_{21}^{p,m}$ (resp. $c_{12}^{p,m}$) between state “off”
209 to “on” (resp., from “on” to “off”) in the breaker p (which are entries in the matrix C_m^p), is related to a
210 higher intensity in the counting process $N_{12,t}^p$ (resp. $N_{21,t}^p$). \square

211 The idea is then to introduce a new probability measure $\bar{\mathbb{P}}$ for a “fictitious world” from the probability
212 measure \mathbb{P} of the “real world” to change all intensities to one under $\bar{\mathbb{P}}$. This is described by means of
213 the Radon-Nikodym derivative, see, e.g. [24, Ch. VI, Sec.2-3]. By using [24, Ch. VI, eq. (3.3)] but for
214 multidimensional case⁴, we define $\bar{\mathbb{P}}$ by putting:

$$\left. \frac{d\bar{\mathbb{P}}}{d\mathbb{P}} \right|_{\mathcal{G}_t} = \Lambda_t := \exp \left\{ - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln(\lambda_{kl,s}^p) dN_{kl,s}^p + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t (\lambda_{kl,s}^p - 1) ds \right\}, \quad (13)$$

215 which is a $(\mathcal{G}_t, \mathbb{P})$ -martingale. Using now Ito’s Lemma, see, e.g., [29], we have:

$$\Lambda_t = 1 - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \Lambda_{s-} (\lambda_{kl,s}^p)^{-1} (\lambda_{kl,s}^p - 1) (dN_{kl,s}^p - \lambda_{kl,s}^p ds). \quad (14)$$

216 We also define the reverse counterpart of (13) by putting:

$$\bar{\Lambda}_t := \exp \left\{ \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln(\lambda_{kl,s}^p) dN_{kl,s}^p - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t (\lambda_{kl,s}^p - 1) ds \right\}, \quad (15)$$

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

$$\bar{\Lambda}_t = 1 + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} (\lambda_{kl,s}^p - 1) d(N_{kl,s}^p - s). \quad (16)$$

⁴See, e.g., [24, Ch. VI, Theorem T2] or [15, eq. (14)] for a general case.

218 In this way, $\bar{\Lambda}_t$ and $(N_{kl,t}^p - t)$ are $(\mathcal{G}_t, \bar{\mathbb{P}})$ -martingale $\forall t \geq 0$. It can be also shown that, under $\bar{\mathbb{P}}$, the dynamic
 219 for X_t is still given by (5), $N_{kl,t}^p$ are independent Poisson processes, and that they have fixed intensity one,
 220 see, e.g., [16, Lemma 1], [24, Ch. II, Theorem T6] and [28, Lemma 4.7.1] resp., *mutatis mutandi*.

221 3.2. Filter Estimate for the Grid States

222 The idea is to use $\bar{\Lambda}_t$ to compute the estimator $\sigma_t(X_t) := \mathbb{E}[X_t | \mathcal{Y}_t]$ by means of a version of Bayes' rule,
 223 see, e.g., [24, Ch. VI, Lemma L5]. More precisely, for any \mathcal{G}_t -adapted and integrable process F_t , the filtered
 224 estimate of F_t can be computed via:

$$\mathbb{E}[F_t | \mathcal{Y}_t] = \frac{\bar{\mathbb{E}}[\bar{\Lambda}_t F_t | \mathcal{Y}_t]}{\bar{\mathbb{E}}[\bar{\Lambda}_t | \mathcal{Y}_t]}, \quad (17)$$

225 where $\bar{\mathbb{E}}$ denotes the expectation operator under the probability measure $\bar{\mathbb{P}}$. We denote by $\bar{\sigma}_t(F_t)$ the
 226 expectation $\bar{\mathbb{E}}[\bar{\Lambda}_t F_t | \mathcal{Y}_t]$. Consequently $\bar{\sigma}_t(1) = \bar{\mathbb{E}}[\bar{\Lambda}_t | \mathcal{Y}_t]$. Note that $\bar{\sigma}_t(1)$ can be computed as the sum
 227 of the components of $\bar{\sigma}_t(X_t)$. Indeed, since X_t takes values in the space \mathbb{X} of unit vectors of \mathbb{R}^N , then
 228 $\langle X_t, \mathbf{1}_N \rangle = 1$ for all $t \geq 0$, where $\mathbf{1}_N := \sum_{n=1}^N e_n$, and therefore $\bar{\sigma}_t(F_t) = \bar{\sigma}_t(F_t \langle X_t, \mathbf{1}_N \rangle) = \langle \bar{\sigma}_t(F_t X_t), \mathbf{1}_N \rangle$.
 229 Thus, in particular taking $F_t \equiv 1$ we have $\bar{\sigma}_t(1) = \langle \bar{\sigma}_t(X_t), \mathbf{1}_N \rangle$. The linear filtered estimate of X_t is given
 230 in the next Proposition 3.1.

231 **Proposition 3.1.** *The finite-dimensional (unnormalized) estimator for the states of X_t is of the form:*

$$\begin{aligned} \bar{\sigma}_t(X_t) = & \bar{\sigma}_0(X_0) + A \int_0^t \bar{\sigma}_s(X_s) ds - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\sigma}_s(X_{s-}) d(N_{kl,s}^p - s) \\ & + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_s(X_{s-}) d(N_{kl,s}^p - s), \end{aligned} \quad (18)$$

232 where diag_m^p is the diagonal matrix defined in eq. (9).

233 **Proof.**

The proof is postponed after that of the Theorem 3.3. ■

235 To obtain $\bar{\sigma}_t(X_t)$, we need the estimation of all parameters involved in eq. (18), i.e., the matrices $A = (a_{ij})$
 236 and $C_m^p = (c_{kl}^{p,m})$, for each $p = 1, \dots, P$ and $m = 1, \dots, M_p$. This is the purpose of the next section.

237 3.3. Parameter Estimation

238 To estimate the parameters that define the transition rate matrices of X and Y^p , $p = 1, \dots, P$, we focus
 239 on the EM algorithm for continuous-time stochastic processes, see, e.g., [21, 30, 31]. The idea is to maximize
 240 a likelihood function in an iterative form. Let $\{\mathbb{P}_\theta, \theta \in \Theta\}$ be a family of probability measures on the
 241 measurable space (Ω, \mathcal{F}) , all absolutely continuous with respect to the (initial) fixed probability measure \mathbb{P} ,
 242 wherein our case,

$$\Theta := \bigcup \{a_{ij}, c_{kl}^{p,m}; 1 \leq i, j \leq N, i \neq j, 1 \leq k, l \leq K_p, k \neq l, 1 \leq m \leq M_p, 1 \leq p \leq P\}. \quad (19)$$

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

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

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

245 In general, computing directly the MLE is challenging. The Expectation–Maximization (EM) algorithm
 246 provides an iterative approximation method starting from an initial estimation θ_0 , see Section 4.4. This
 247 algorithm is based on the following straightforward application of the well-known Jensen's inequality:

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

248 This gives a global minoration for the log-likelihood mapping $\theta \mapsto \mathcal{L}(\theta)$ by means of the auxiliary mapping
 249 $\theta \mapsto \mathcal{L}(\hat{\theta}) + \mathcal{Q}(\theta, \hat{\theta})$. At each iteration $r \in \mathbb{N}_0$, the EM algorithm consists of two main steps:

250 **(1) E-step:** set $\hat{\theta} = \hat{\theta}_r$ and compute $\mathcal{Q}(\cdot, \hat{\theta})$,

251 **(2) M-step:** find $\hat{\theta}_{r+1} \in \arg \max_{\theta \in \Theta} \mathcal{Q}(\theta, \hat{\theta})$.

252 This algorithm can be stopped when a stopping test is satisfied, see Section 4.2. The generated sequence
 253 $\{\hat{\theta}_r\}_{r \in \mathbb{N}_0}$ gives nondecreasing values of the likelihood function, i.e., $\mathcal{L}(\hat{\theta}_{r+1}) > \mathcal{L}(\hat{\theta}_r)$ unless $\hat{\theta}_{r+1} = \hat{\theta}_r$. For
 254 convergence issues, see, e.g., [30, 31, 32].

255 In our context, suppose our model is determined by some parameters $\theta \in \Theta$, i.e., we have computed
 256 already the *E-step* under θ . To compute the new parameters $\hat{\theta} \in \Theta$ that maximize the log-likelihood, i.e.,
 257 the *M-step*, we have the following Theorem 3.2.

258 **Theorem 3.2.** *The estimation $\hat{A} = (\hat{a}_{ij})$ of $A = (a_{ij})$, and $\hat{C}_m^p = (\hat{c}_{kl}^{p,m})$ of $C_m^p = (c_{kl}^{p,m})$, for each
 259 $p = 1, \dots, P$ and $m = 1, \dots, M_p$; are given for $i \neq j$ and $k \neq l$, by:*

$$\hat{a}_{ji} = \frac{\mathbb{E}[J_{ij,t} \mid \mathcal{Y}_t]}{\mathbb{E}[O_{i,t} \mid \mathcal{Y}_t]} \quad , \quad \hat{c}_{lk}^{p,m} = \frac{\sum_{n \in I_m^p} \mathbb{E}[L_{kl,t}^{p,n} \mid \mathcal{Y}_t]}{\sum_{n \in I_m^p} \mathbb{E}[S_{k,t}^{p,n} \mid \mathcal{Y}_t]} \quad ,$$

260 where, for $i, j = 1, \dots, N$, $i \neq j$,

$$J_{ij,t} := \int_0^t \langle e_i, X_{s-} \rangle \langle e_j, dX_s \rangle \quad , \quad O_{i,t} := \int_0^t \langle e_i, X_s \rangle ds \quad , \quad (20)$$

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

$$L_{kl,t}^{p,n} := \int_0^t \langle e_n, X_{s-} \rangle dN_{kl,s}^p \quad , \quad S_{k,t}^{p,n} := \int_0^t \langle f_k^p, Y_s^p \rangle \langle e_n, X_s \rangle ds \quad . \quad (21)$$

262 **Proof.**

See Proof 1 in Appendix A. ■

263

264 In Theorem 3.2, note that $J_{ij,t}$ represents a counting process of the number of jumps of X from state
 265 $e_i \in \mathbb{X}$ 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 $e_i \in \mathbb{X}$
 266 within $[0, t]$, $L_{kl,t}^{p,n}$ represents the process that increases only when Y^p jumps from state $f_k^p \in \mathbb{Y}_p$ to state
 267 $f_l^p \in \mathbb{Y}_p$ and X is in state $e_n \in \mathbb{X}$, $k \neq l$; and $S_{k,t}^{p,n}$ stands for the total time up to $t \geq 0$ for which X is in
 268 state $e_n \in \mathbb{X}$ and simultaneously Y^p is in state $f_k^p \in \mathbb{Y}_p$.

269 **Example** (Electrical grid). *For instance, consider $R = 1$, $M_p = K_p = 2$ and $\mathbb{M}_p = \mathbb{K}_p$. First, $J_{ij,t}$ counts
 270 the number of jumps that the grid does from the reference configuration $h_i \in \mathbb{H}$ to the reference configuration
 271 $h_j \in \mathbb{H}$ within $[0, t]$. Second, $O_{i,t}$ is the time that the grid spends on the reference configuration $h_i \in \mathbb{H}$ up
 272 to time $t \geq 0$. Third, $L_{12,t}^{p,n}$ (resp. $L_{21,t}^{p,n}$) increases only when the breaker p changes resp. from state “off” to
 273 “on” (resp. from state “on” to “off”), and simultaneously the grid is in the reference configuration $h_n \in \mathbb{H}$
 274 within the time interval $[0, t]$. Fourth, $S_{1,t}^{p,n}$ (resp. $S_{2,t}^{p,n}$) is the time that the grid spends on the reference
 275 configuration $h_n \in \mathbb{H}$ up to time $t \geq 0$ and simultaneously the breaker p is in the state “on” (resp. the state
 276 “off”). □*

277 From eq. (17), the estimation \hat{a}_{ij} and $\hat{c}_{kl}^{p,m}$ can be obtained via the probability measure $\bar{\mathbb{P}}$ by:

$$\hat{a}_{ji} = \frac{\langle \bar{\sigma}_t(J_{ij,t}X_t), \mathbf{1}_N \rangle}{\langle \bar{\sigma}_t(O_{i,t}X_t), \mathbf{1}_N \rangle}, \quad \hat{c}_{lk}^{p,m} = \frac{\sum_{n \in I_m^p} \langle \bar{\sigma}_t(L_{kl,t}^{p,n}X_t), \mathbf{1}_N \rangle}{\sum_{n \in I_m^p} \langle \bar{\sigma}_t(S_{k,t}^{p,n}X_t), \mathbf{1}_N \rangle}. \quad (22)$$

278 In this way, it is sufficient to compute the estimators $\bar{\sigma}_t(J_{ij,t}X_t)$, $\bar{\sigma}_t(O_{i,t}X_t)$, $\bar{\sigma}_t(S_{k,t}^{p,n}X_t)$ and $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$.
279 Now, if we consider the process:

$$F_t = F_0 + \int_0^t \alpha(X_s)ds + \int_0^t \langle \beta(X_s), dV_s \rangle \quad (23)$$

280 then, for each $i, j = 1, \dots, N$, $i \neq j$, $p = 1, \dots, P$, $k = 1, \dots, K_p$, $n = 1, \dots, N$, the processes $J_{ij,t}$, $O_{i,t}$ and $S_{k,t}^{p,n}$
281 are considered into F_t , where $F_0 \in \mathbb{R}$ is known, and $\alpha : \mathbb{X} \rightarrow \mathbb{R}$ and $\beta : \mathbb{X} \rightarrow \mathbb{R}^N$ are known functions with
282 finite range, \mathcal{G}_t -adapted and integrable for each $t \geq 0$. Indeed, by using eq. (5), and taking

- 283 (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 $F_t = J_{ij,t}$,
284 (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 $F_t = O_{i,t}$,
285 (iii) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = \langle f_k^p, Y_t^p \rangle \langle e_n, X_t \rangle$, and $\beta(X_t) = \mathbf{0}_N \in \mathbb{R}^N$, we obtain $F_t = S_{k,t}^{p,n}$.

286 Therefore, to compute $\bar{\sigma}_t(J_{ij,t}X_t)$, $\bar{\sigma}_t(O_{i,t}X_t)$, and $\bar{\sigma}_t(S_{k,t}^{p,n}X_t)$, we can compute once $\bar{\sigma}_t(F_tX_t)$ and re-
287 strict afterwards to the particular cases of $\alpha(X_t)$ and $\beta(X_t)$. On the other hand, we know that $\bar{\sigma}_t(F_t) =$
288 $\langle \bar{\sigma}_t(F_tX_t), \mathbf{1}_N \rangle$, so that, we make the inner product between $\bar{\sigma}_t(F_tX_t)$ and $\mathbf{1}_N$ to have the estimation for
289 all parameters, see eq. (22). The following Theorem 3.3 gives the linear filter estimate $\bar{\sigma}_t(F_tX_t)$. The filter
290 estimate $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$ is given in Theorem 3.6.

291 **Theorem 3.3.** *The finite-dimensional (unnormalized) estimator for F_tX_t is of the form:*

$$\begin{aligned} \bar{\sigma}_t(F_tX_t) = & \bar{\sigma}_0(F_0X_0) + A \int_0^t \bar{\sigma}_s(F_sX_s)ds - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\sigma}_s(F_sX_{s-})d(N_{kl,s}^p - s) \\ & + \int_0^t \bar{\sigma}_s(X_s\alpha_s)ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\sigma}_s((\beta_{j,s} - \beta_{i,s})X_s), e_i \rangle a_{ji}(e_j - e_i)ds \\ & + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_s(F_sX_{s-})d(N_{kl,s}^p - s), \end{aligned} \quad (24)$$

292 where $\alpha_t := \alpha(X_t) \in \mathbb{R}$ and $\beta_t := \beta(X_t) \in \mathbb{R}^N$ are known functions with finite range, \mathcal{G}_t -adapted and
293 integrable for each $t \geq 0$, F_t is given in eq. (23), and diag_m^p is the diagonal matrix of eq. (9).

294 **Proof.**

See Proof 2 in Appendix A.

295
296 ■

297 **Remark 3.4.** *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. (23), then*
298 *the Proposition 3.1 is a particular case of Theorem 3.3. \square*

299 In this way, the filter estimates for the parameter estimation are given in the next Corollary 3.5 by taking
300 the particular cases of α_t and β_t within F_t , see eq. (23) and the cases (i), (ii) and (iii).

301 **Corollary 3.5.** *The finite-dimensional (unnormalized) estimator for $J_{ij,t}X_t$, $O_{i,t}X_t$ and $S_{k,t}^{p,n}X_t$ are resp.*
 302 *of the form:*

$$\begin{aligned}
 \bar{\sigma}_t(J_{ij,t}X_t) &= A \int_0^t \bar{\sigma}_s(J_{ij,s}X_s)ds + \int_0^t \langle \bar{\sigma}_s(X_s), e_i \rangle e_j a_{ji} ds - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\sigma}_s(J_{ij,s-}X_{s-})d(N_{kl,s}^p - s) \\
 &\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_s(J_{ij,s-}X_{s-})d(N_{kl,s}^p - s), \\
 \bar{\sigma}_t(O_{i,t}X_t) &= A \int_0^t \bar{\sigma}_s(O_{i,s}X_s)ds + \int_0^t \langle \bar{\sigma}_s(X_s), e_i \rangle e_i ds - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\sigma}_s(O_{i,s-}X_{s-})d(N_{kl,s}^p - s) \\
 &\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_s(O_{i,s-}X_{s-})d(N_{kl,s}^p - s), \\
 \bar{\sigma}_t(S_{k,t}^{p,n}X_t) &= A \int_0^t \bar{\sigma}_s(S_{k,s}^{p,n}X_s)ds + \int_0^t \langle f_k^p, Y_{s-}^p \rangle \langle \bar{\sigma}_s(X_s), e_n \rangle e_n ds - \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \int_0^t \bar{\sigma}_s(S_{k,s-}^{p,n}X_{s-})d(N_{uv,s}^q - s) \\
 &\quad + \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \sum_{m=1}^{M_q} \int_0^t \langle f_u^q, Y_{s-}^q \rangle c_{vu}^{q,m} \text{diag}_m^q \bar{\sigma}_s(S_{k,s-}^{p,n}X_{s-})d(N_{uv,s}^q - s),
 \end{aligned}$$

303 where diag_m^p is the matrix of eq. (9). ■

304 Finally, the filter estimate we need to complete the estimation of all parameters is $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$. For this,
 305 we write the semi-Martingale representation of $L_{kl,t}^{p,n}$ from eq. (21), $k \neq l$, by:

$$L_{kl,t}^{p,n} = \int_0^t \langle e_n, X_{s-} \rangle d(N_{kl,s}^p - s) + \int_0^t \langle e_n, X_{s-} \rangle ds. \quad (25)$$

306

307 **Theorem 3.6.** *The finite-dimensional (unnormalized) estimator for $L_{kl,t}^{p,n}X_t$, $k \neq l$, is of the form:*

$$\begin{aligned}
 \bar{\sigma}_t(L_{kl,t}^{p,n}X_t) &= A \int_0^t \bar{\sigma}_s(L_{kl,s}^{p,n}X_s)ds - \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \int_0^t \bar{\sigma}_s(L_{kl,s-}^{p,n}X_{s-})d(N_{uv,s}^q - s) \\
 &\quad + \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \langle e_n, \bar{\sigma}_s(X_{s-}) \rangle e_n dN_{kl,s}^p \\
 &\quad + \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \sum_{m=1}^{M_q} \int_0^t \langle f_u^q, Y_{s-}^q \rangle c_{vu}^{q,m} \text{diag}_m^q \bar{\sigma}_s(L_{kl,s-}^{p,n}X_{s-})d(N_{uv,s}^q - s),
 \end{aligned} \quad (26)$$

308 where diag_m^p is the diagonal matrix defined in eq. (9).

309 **Proof.**

See Proof 3 in Appendix A. ■

310

311 4. Numerical Methods

312 This section presents a general numerical method for all SDEs involved in our model, i.e., a general scheme
 313 involving each SDE of the filter estimates $\bar{\sigma}_t(X_t)$, $\bar{\sigma}_t(J_{ij,t}X_t)$, $\bar{\sigma}_t(O_{i,t}X_t)$, $\bar{\sigma}_t(S_{k,t}^{p,n}X_t)$ and $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$,
 314 obtained resp. in Proposition 3.1 Corollary 3.5 and Theorem 3.6. We also show how to obtain the initial
 315 estimation of all parameters and the set \mathbb{H} when $\mathbb{M}_p = \mathbb{K}_p$ for each $p = 1, \dots, P$. This is a particular case
 316 that applies in the case study considered in this paper of the electrical grid by using available data of all
 317 breakers states.

318 First, we write $t_0 < t_1 < \dots < t_{W+1}$ the increasingly ordered instances of all “jump times” of the observed
 319 processes. This is shown in Figure 3 and is obtained by superposing all state change times $\{\tau_1^p, \tau_2^p, \dots\}$ of the
 320 observed processes, $p = 1, \dots, P$. We denote by t_w , $w = 0, \dots, W$, the instant where at least one observed
 321 process 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
 322 in which the observed processes remain constant between the time interval $[t_w, t_{w+1})$.

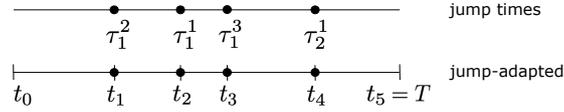


Figure 3: Representation of the jump-adapted scheme given by the jump times of all observed processes.

323 4.1. Jump-Adapted Scheme for Filters

324 The scheme presented here is a strong approximation with no discretization error on the SDEs solution,
 325 in which a jump-adapted scheme is given by a superposition of all “jump times”, i.e., all state change times
 326 generated by the temporal evolution of the observed processes, see Figure 3. The jump effects are then
 327 added at the correct jump times. To use this kind of method, one has to check whether the SDE concerned
 328 belongs to the particular subclass of SDEs for which the corresponding non-jump part admits an exact
 329 solution, see, e.g., [33, Ch. II]. In our case, all SDEs admit an explicit solution in the non-jump parts as we
 330 see in next.

331 Instead of building a scheme for each filter estimate $\bar{\sigma}_t(X_t)$, $\bar{\sigma}_t(J_{ij,t}X_t)$, $\bar{\sigma}_t(O_{i,t}X_t)$, $\bar{\sigma}_t(S_{k,t}^{p,n}X_t)$ and
 332 $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$, we present a generalized scheme for the following SDEs system:

$$dK_t = \Xi_t K_t dt + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} Q_{kl,t-}^p K_{t-} dN_{kl,t}^p, \quad (27)$$

$$dG_t = (\Upsilon_t G_t + \Gamma_t K_t) dt + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} (\Pi_{kl,t-}^p G_{t-} + \Lambda_{kl,t-}^p K_{t-}) dN_{kl,t}^p, \quad (28)$$

333 where $K_t, G_t \in \mathbb{R}^N$ are \mathcal{G}_t -adapted and integrable for any $t \geq 0$, and $\Xi_t, Q_{kl,t}^p, \Upsilon_t, \Gamma_t, \Pi_{kl,t}^p, \Lambda_{kl,t}^p \in \mathbb{R}^{N \times N}$
 334 are constant matrices between “jumps” for each $k, l = 1, \dots, K_p$, $p = 1, \dots, P$. They are defined by $\Xi_t = \Xi_{t_w}$,
 335 $\Upsilon_t = \Upsilon_{t_w}$, $\Gamma_t = \Gamma_{t_w}$, $Q_{kl,t}^p = Q_{kl,t_w}^p$, $\Pi_{kl,t}^p = \Pi_{kl,t_w}^p$, $\Lambda_{kl,t}^p = \Lambda_{kl,t_w}^p$, for any $t \in [t_w, t_{w+1})$, $w = 0, \dots, W$.

336 **Theorem 4.1.** *The jump-adapted exact solution scheme for K_t and G_t with initial conditions $K_0, G_0 \in \mathbb{R}^N$*
 337 *is written by:*

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

338 where $\Delta N_{kl,t_{w+1}}^p = N_{kl,t_{w+1}}^p - N_{kl,t_{w+1}^-}^p$ is defined by $\Delta N_{kl,t_{w+1}}^p = 1$ if Y_t^p jumps from state $f_k^p \in \mathbb{Y}_p$ to state
 339 $f_l^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, and both, $G_{t_{w+1}^-} := \lim_{s \uparrow t_{w+1}} G_s$ and $K_{t_{w+1}^-} := \lim_{s \uparrow t_{w+1}} K_s$ are the respective values
 340 “before” the jump at time t_{w+1} .

341 **Proof.**

See Proof 4 in Appendix A. ■

343 As a consequence, the jump-adapted scheme for each filter estimate $\bar{\sigma}_t(X_t)$, $\bar{\sigma}_t(J_{ij,t} X_t)$, $\bar{\sigma}_t(O_{i,t} X_t)$,
 344 $\bar{\sigma}_t(S_{k,t}^{p,n} X_t)$ and $\bar{\sigma}_t(L_{kl,t}^{p,n} X_t)$ can be easily obtained from the Theorem 4.1 as follows. First, to relax the
 345 notation of all equations, we define, for each $k, l = 1, \dots, K_p$, $p = 1, \dots, P$, the matrices Φ_t and $\Psi_{kl,t}^p$ for each
 346 $t \geq 0$, by:

$$\Phi_t := A - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t}^p, \quad (29)$$

$$\Psi_{kl,t}^p := \sum_{m=1}^{M_p} \langle f_k^p, Y_t^p \rangle c_{lk}^{p,m} \text{diag}_m^p - \text{Id}_N, \quad (30)$$

347 where Id_N is the identity matrix in $\mathbb{R}^{N \times N}$, and diag_m^p is the diagonal matrix defined in eq. (9). Note that
 348 Φ_t and $\Psi_{kl,t}^p$ are constant between “jumps”. That is, $\Phi_t = \Phi_{t_w}$ and $\Psi_{kl,t}^p = \Psi_{kl,t_w}^p$ for any $t \in [t_w, t_{w+1})$,
 349 $w = 0, \dots, W$.

350 Note that we can obtain the SDE of $\bar{\sigma}_t(X_t)$ (see eq. (18)) from eq. (27) by considering $K_t = \bar{\sigma}_t(X_t)$,
 351 $\Xi_t = \Phi_t$, and for each $p = 1, \dots, P$, $k, l = 1, \dots, K_p$, $Q_{kl,t}^p = \Psi_{kl,t}^p$. Thus, from Theorem 4.1, the following
 352 scheme holds for the filter estimate $\bar{\sigma}_t(X_t)$ of the grid state.

353 **Corollary 4.2.** *A jump-adapted exact solution scheme for $\bar{\sigma}_t(X_t)$ is written by:*

$$\begin{aligned}
 \bar{\sigma}_{t_0}(X_{t_0}) &= \bar{\sigma}_0(X_0), \\
 \bar{\sigma}_{t_{w+1}^-}(X_{t_{w+1}^-}) &= \exp\{\Phi_{t_w} \Delta t_{w+1}\} \bar{\sigma}_{t_w}(X_{t_w}), \\
 \bar{\sigma}_{t_{w+1}}(X_{t_{w+1}}) &= \bar{\sigma}_{t_{w+1}^-}(X_{t_{w+1}^-}) + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t_{w+1}^-}^p \bar{\sigma}_{t_{w+1}^-}(X_{t_{w+1}^-}) \Delta N_{kl,t_{w+1}}^p,
 \end{aligned}$$

354 where $\Delta t_{w+1} = t_{w+1} - t_w$, $\Delta N_{kl,t_{w+1}}^p = N_{kl,t_{w+1}}^p - N_{kl,t_{w+1}^-}^p$ is defined by $\Delta N_{kl,t_{w+1}}^p = 1$ if Y_t^p jumps from
355 state $f_k^p \in \mathbb{Y}_p$ to state $f_l^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, and $\bar{\sigma}_{t_{w+1}}(X_{t_{w+1}^-}) := \lim_{s \uparrow t_{w+1}} \bar{\sigma}_{t_{w+1}}(X_s)$ is the values “before”
356 the jump at time t_{w+1} . ■

357 For the other filter estimates, $\bar{\sigma}_t(J_{ij,t}X_t)$, $\bar{\sigma}_t(O_{i,t}X_t)$, and $\bar{\sigma}_t(S_{k,t}^{p,n}X_t)$, we can express the scheme for
358 $\bar{\sigma}_t(F_tX_t)$ from the Theorem 3.3, see the cases (i), (ii) and (iii) after eq. (23). The following scheme holds
359 for the filter estimate $\bar{\sigma}_t(F_tX_t)$.

360 **Corollary 4.3.** *The jump-adapted exact solution scheme for $\bar{\sigma}_t(F_tX_t)$ with initial condition $\bar{\sigma}_0(F_0X_0) \in \mathbb{R}^N$*
361 *is written by:*

$$\begin{aligned} \bar{\sigma}_{t_0}(F_{t_0}X_{t_0}) &= \bar{\sigma}_0(F_0X_0), \\ \bar{\sigma}_{t_{w+1}}(F_{t_{w+1}}X_{t_{w+1}}) &= [\text{Id}_N \quad \mathbf{0}_{N \times N}] \exp \left\{ \begin{bmatrix} \Phi_{t_w} & \Gamma_{t_w} \\ \mathbf{0}_{N \times N} & \Phi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} \bar{\sigma}_{t_w}(F_{t_w}X_{t_w}) \\ \bar{\sigma}_{t_w}(X_{t_w}) \end{bmatrix}, \\ \bar{\sigma}_{t_{w+1}}(F_{t_{w+1}}X_{t_{w+1}}) &= \bar{\sigma}_{t_{w+1}}(F_{t_{w+1}^-}X_{t_{w+1}^-}) + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t_{w+1}}^p \bar{\sigma}_{t_{w+1}}(F_{t_{w+1}^-}X_{t_{w+1}^-}) \Delta N_{kl,t_{w+1}}^p, \end{aligned}$$

362 where $\Delta N_{kl,t_{w+1}}^p = N_{kl,t_{w+1}}^p - N_{kl,t_{w+1}^-}^p$ is defined by $\Delta N_{kl,t_{w+1}}^p = 1$ if Y_t^p jumps from state $f_k^p \in \mathbb{Y}_p$ to state
363 $f_l^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, $\bar{\sigma}_{t_{w+1}}(F_{t_{w+1}^-}X_{t_{w+1}^-}) := \lim_{s \uparrow t_{w+1}} \bar{\sigma}_{t_{w+1}}(F_sX_s)$ is the respective values “before” the
364 jump at time t_{w+1} , and

$$\Gamma_{t_w} = \text{diag}(\alpha(e_1), \dots, \alpha(e_N)) + \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^\top,$$

365 where $\text{diag}(\alpha(e_1), \dots, \alpha(e_N))$ is a diagonal matrix in $\mathbb{R}^{N \times N}$ with diagonal $(\alpha(e_1), \dots, \alpha(e_N))$, and $\alpha : \mathbb{X} \rightarrow \mathbb{R}$
366 and $\beta : \mathbb{X} \rightarrow \mathbb{R}^N$ are the known functions in F_t , see eq. (23).

367 **Proof.**

See Proof 5 in Appendix A. ■

369 Now, for each $u, v = 1, \dots, K_q$, $u \neq v$, $q = 1, \dots, P$, and $n = 1, \dots, N$, the SDE of $\bar{\sigma}_t(L_{uv,t}^{q,n}X_t)$ (see eq. (26))
370 can be obtained from eq. (28) by considering $G_t = \bar{\sigma}_t(L_{uv,t}^{q,n}X_t)$, $K_t = \bar{\sigma}_t(X_t)$, $\Upsilon_t = \Xi_t = \Phi_t$, $\Gamma_t = \mathbf{0}_{N \times N} \in$
371 $\mathbb{R}^{N \times N}$, and for each $p = 1, \dots, P$, $k, l = 1, \dots, K_p$, $\Pi_{kl,t}^p = Q_{kl,t}^p = \Psi_{kl,t}^p$ and $\Lambda_{kl,t}^p = (\Psi_{uv,t}^q + \text{Id}_N) \text{diag}(e_n)$
372 for $k = u$, $l = v$, $p = q$, and $\Lambda_{kl,t}^p = \mathbf{0}_{N \times N} \in \mathbb{R}^{N \times N}$ otherwise, where $\text{diag}(e_n)$ is a diagonal matrix in
373 $\mathbb{R}^{N \times N}$ with diagonal $e_n \in \mathbb{X}$. Thus, from Theorem 4.1, the following scheme holds for the filter estimate
374 $\bar{\sigma}_t(L_{uv,t}^{q,n}X_t)$.

375 **Corollary 4.4.** *A jump-adapted exact solution scheme for $\bar{\sigma}_t(L_{uv,t}^{q,n}X_t)$, $u, v = 1, \dots, K_q$, $u \neq v$, $q = 1, \dots, P$,*
376 *and $n = 1, \dots, N$, is written by:*

$$\begin{aligned} \bar{\sigma}_{t_0}(L_{uv,t_0}^{q,n}X_{t_0}) &= \mathbf{0}_N, \\ \bar{\sigma}_{t_{w+1}}(L_{uv,t_{w+1}}^{q,n}X_{t_{w+1}}) &= [\text{Id}_N \quad \mathbf{0}_{N \times N}] \exp \left\{ \begin{bmatrix} \Phi_{t_w} & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \Phi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} \bar{\sigma}_{t_w}(L_{uv,t_w}^{q,n}X_{t_w}) \\ \bar{\sigma}_{t_w}(X_{t_w}) \end{bmatrix}, \\ \bar{\sigma}_{t_{w+1}}(L_{uv,t_{w+1}}^{q,n}X_{t_{w+1}}) &= \bar{\sigma}_{t_{w+1}}(L_{uv,t_{w+1}^-}^{q,n}X_{t_{w+1}^-}) + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t_{w+1}}^p \bar{\sigma}_{t_{w+1}}(L_{uv,t_{w+1}^-}^{q,n}X_{t_{w+1}^-}) \Delta N_{kl,t_{w+1}}^p \\ &\quad + \left(\Psi_{uv,t_{w+1}}^q + \text{Id}_N \right) \text{diag}(e_n) \bar{\sigma}_{t_{w+1}}(X_{t_{w+1}^-}) \Delta N_{uv,t_{w+1}}^q, \end{aligned}$$

377 where $\Delta t_{w+1} = t_{w+1} - t_w$, $\Delta N_{kl,t_{w+1}}^p = N_{kl,t_{w+1}}^p - N_{kl,t_w}^p$ is defined by $\Delta N_{kl,t_{w+1}}^p = 1$ if Y_t^p jumps from
378 state $f_k^p \in \mathbb{Y}_p$ to state $f_l^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, and $\bar{\sigma}_{t_{w+1}}(L_{uv,t_{w+1}}^{q,n} X_{t_{w+1}}^-) := \lim_{s \uparrow t_{w+1}} \bar{\sigma}_{t_{w+1}}(L_{uv,s}^{q,n} X_s)$ and
379 $\bar{\sigma}_{t_{w+1}}(X_{t_{w+1}}^-) := \lim_{s \uparrow t_{w+1}} \bar{\sigma}_{t_{w+1}}(X_s)$ are the values “before” the jump at time t_{w+1} . ■

380 4.2. An Stopping Criteria for the EM Algorithm

381 Instead of using the strict stopping criteria for the EM algorithm⁵ $\hat{\theta}_{r+1} = \hat{\theta}_r$ for some $r \in \mathbb{N}$, we define
382 the following stopping test for numerical purposes:

$$\frac{\left\| \bar{\sigma}_t^{(r)}(X_t) - \bar{\sigma}_t^{(r-1)}(X_t) \right\| + \left\| \hat{A}^{(r)} - \hat{A}^{(r-1)} \right\| + \sum_{p=1}^P \sum_{m=1}^{M_p} \left\| \hat{C}_m^{p,(r)} - \hat{C}_m^{p,(r-1)} \right\|}{\left\| \bar{\sigma}_t^{(r-1)}(X_t) \right\| + \left\| \hat{A}^{(r-1)} \right\| + \sum_{p=1}^P \sum_{m=1}^{M_p} \left\| \hat{C}_m^{p,(r-1)} \right\|} \leq \varepsilon, \quad (31)$$

383 where $\varepsilon > 0$ is a given stopping parameter. We note that the normalization by the sum of parameter norms
384 limits the dependency of ε to the magnitude of the parameters.

385 4.3. Uncertainty and Final Hidden State

386 Note that in Proposition 3.1 we obtain the estimator $\bar{\sigma}_t(X_t) = \mathbb{E}[\bar{\Lambda}_t X_t \mid \mathcal{Y}_t]$ and then, by applying
387 the Bayes’ rule of eq. (17), we obtain the filtered estimate $\sigma_t(X_t) = \mathbb{E}[X_t \mid \mathcal{Y}_t]$ which is the probability
388 distribution $(\mathbb{P}[X_t = e_1 \mid \mathcal{Y}_t], \dots, \mathbb{P}[X_t = e_N \mid \mathcal{Y}_t])$. In order to know the exact estate of the hidden process
389 over the states h_1, \dots, h_N , we then take for each time $t \geq 0$, the value:

$$\hat{X}_t = e_n \quad , \quad n \in \arg \max_{n'=1, \dots, N} \langle \sigma_t(X_t), e_{n'} \rangle. \quad (32)$$

390 Considering this choice, at fixed time $t \geq 0$, $\langle \hat{X}_t, e_n \rangle = 1$ and $\langle \hat{X}_t, e_{n'} \rangle = 0$ for each $n' = 1, \dots, N$, $n' \neq n$,
391 and by eq. (4), the estimated hidden state is therefore h_n at such time $t \geq 0$.

392 Concerning the uncertainty of the choice in eq. (32), we can use the estimator $\sigma_t(X_t)$ for each $t \geq 0$ in
393 the following way. First, for each $t \geq 0$, let

$$\epsilon_t := 1 - \max_{n=1, \dots, N} \langle \sigma_t(X_t), e_n \rangle \quad (33)$$

394 be the function that represents how far the highest $\sigma_t(X_t)$ is from the value one. Recall that this estimator
395 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
396 the estimated hidden state by using the eq. (32), is an almost-sure choice. Thus, for any time $t \geq 0$, the
397 function ϵ_t represents the uncertainty on the hidden states. For the Section 5 of numerical results, we also
398 define the mean

$$\delta_t := \frac{1}{t} \int_0^t \epsilon_s ds \quad (34)$$

399 to see how far is ϵ_t of the mean δ_t , in particular at the jump times.

400 4.4. An Estimation Procedure for the Initial Parameters when $\mathbb{M}_p = \mathbb{K}_p$

401 In this section, we consider the particular case when $\mathbb{M}_p = \mathbb{K}_p$, $R = 1$, $M_p = K_p$ for each $p = 1, \dots, P$.
402 For the EM algorithm, a $\theta_0 \in \Theta$ must be initialized, i.e., we need to choose initial values for the matrices
403 $\hat{A}^{(0)} = (\hat{a}_{ij}(0))$ and $\hat{C}_m^{p(0)} = (\hat{c}_{kl}^{p,m}(0))$, for each observed process $p = 1, \dots, P$ and index $m = 1, \dots, M_p$.
404 At first glance, we can use Theorem 3.2 empirically, i.e., we can discretize all the involved integrals in the

⁵Because, e.g., it could take several iterations to have the equality in all the parameters.

parameter estimation (those in (20) and eq. (21)), and use all the information of the state processes to estimate θ_0 .

First, since we know the change of each observed process, then we know the values of each Y_t^p , $p = 1, \dots, P$, 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} , then we can compute an empirical estimation of H , and so that, an empirical estimation of the underlying state process X_t by eq. (3) and eq. (4). This can be computed through a distance measure by finding the closest 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 joint state of all observed processes can be represented by a piecewise constant state process $Y := \{Y_t\}_{t \geq 0}$, where Y_t is the joint information of observed processes at time $t \geq 0$, defined by:

$$Y_t := \sum_{p=1}^P \sum_{k=1}^{K_p} \langle f_k^p, Y_t^p \rangle (k-1) g_p \in \prod_{p=1}^P \mathbb{K}_p, \quad (35)$$

where for each $p = 1, \dots, P$, $k = 1, \dots, K_p$, the k -th unit vector $f_k^p \in \mathbb{K}_p$, 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 state process Y^p changes of value at time $t \geq 0$, then Y_t changes too. This occurs at the times $t_0 < t_1 < \dots < t_W$, see Figure 3.

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

$$\hat{H}_t \in \arg \min_{\substack{h_n \in \mathbb{H} \\ n=1, \dots, N}} d(Y_t, h_n). \quad (36)$$

This approach is a classification procedure over the joint information of all observable processes at time $t \geq 0$, which is represented by Y_t in eq. (35).

Now, with this empirical estimation \hat{H}_t , we can compute the empirical values of X_t by means of its definition in eq. (3). We denote this estimation by $\hat{X} := \{\hat{X}_t\}_{t \geq 0}$. Note that \hat{H} and \hat{X} are also piecewise right-continuous with left limits. Since we know now the values of \hat{X}_t at the state change times of Y , t_0, \dots, t_W , we discretize all the involved integrals in Theorem 3.2 to estimate empirically all the parameters of our model. For each $i, j = 1, \dots, N$, $i \neq j$, the initial estimation of $A = (a_{ij})$, i.e., the matrix $\hat{A}^{(0)} = (\hat{a}_{ij}(0))$, and of $C_m^p = (c_{kl}^{p,m})$, i.e, the matrix $\hat{C}_m^{p(0)} = (\hat{c}_{kl}^{p,m}(0))$, for each $p = 1, \dots, P$ and $m = 1, \dots, M_p$, are obtained as follows:

$$\hat{a}_{ij}(0) := \frac{\sum_{w=0}^{W-1} \langle e_i, \hat{X}_{t_w} \rangle \langle e_j, \hat{X}_{t_{w+1}} \rangle}{\sum_{w=0}^{W-1} \langle e_i, \hat{X}_{t_w} \rangle \Delta t_{w+1}}, \quad \hat{c}_{lk}^{p,m}(0) := \frac{\sum_{w=0}^{W-1} \langle f_k^p, Y_{t_w}^p \rangle \langle f_l^p, Y_{t_{w+1}}^p \rangle \sum_{n \in I_m^p} \langle e_n, \hat{X}_{t_w} \rangle}{\sum_{w=0}^{W-1} \langle f_k^p, Y_{t_w}^p \rangle \sum_{n \in I_m^p} \langle e_n, \hat{X}_{t_w} \rangle \Delta t_{w+1}}. \quad (37)$$

On the other hand, the initial estimation $\bar{\sigma}_0(X_0)$ of the process X at time $t = 0$ that we need in eq. (18), is given by the empirical estimation $\bar{\sigma}_0(\hat{X}_0)$ defined for each $n = 1, \dots, N$, by:

$$\langle \bar{\sigma}_0(\hat{X}_0), e_n \rangle := \frac{\sum_{w=0}^{W-1} \langle e_n, \hat{X}_{t_w} \rangle \Delta t_{w+1}}{\sum_{n=1}^N \sum_{w=0}^{W-1} \langle e_n, \hat{X}_{t_w} \rangle \Delta t_{w+1}}. \quad (38)$$

4.5. A Method to Find the Reference Configurations of the Grid

We want to build the set \mathbb{H} by using the available data which consists of the temporal evolution of the breakers states in the grid, by using $Y = \{Y_t\}_{t \geq 0}$ of eq. (35). Here, we also consider the particular case when $\mathbb{M}_p = \mathbb{K}_p$, $R = 1$, $M_p = K_p$, for each $p = 1, \dots, P$.

435 To obtain the reference configurations of the grid, we construct clusters from data by partitioning it into
 436 N subsets. Each subset U_1, \dots, U_N , called cluster, is represented by its representative state μ_1, \dots, μ_N , resp.
 437 To obtain optimal clusters, we use K -means method, see, e.g., [27]. The extension of K -means in continuous
 438 time is given by the minimization of the following cost function in the horizon time $T > 0$:

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

439 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.
 440 Since $Y = \{Y_t\}_{t \geq 0}$ is piecewise right-continuous on $0 = t_0 < t_1 < \dots < t_W = T$, see eq. (35), we have

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

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

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

443 We use the traditional approach to construct the clusters by classification. For each $t \geq 0$, the cluster U_n ,
 444 $n = 1, \dots, N$, is obtained as the set:

$$U_n = \bigcup_{w=0}^{W-1} \left\{ Y_{t_w} \in \prod_{p=1}^P \mathbb{K}_p \mid d(Y_{t_w}, \mu_n) \leq d(Y_{t_w}, \mu_{n'}) , \text{ for each } n' = 1, \dots, N \right\} . \quad (39)$$

445 Note that the empirical estimation of the grid states in eq. (36) also means that for any time $t \geq 0$, there is
 446 a $n \in \{1, \dots, N\}$ such that $\widehat{H}_t \in U_n$.

447 This method generates a sequence $\{\mu_r\}_{r \in \mathbb{N}_0}$, where $\mu_r := (\mu_1^{(r)}, \dots, \mu_N^{(r)})$ is the vector of all representative
 448 states of the clusters at iteration $r \in \mathbb{N}_0$. This can be initialized from the available data, e.g., randomly,
 449 heuristically, or by K -means++ approach [34]. It should be noted that the performance of an iterative cluster-
 450 ing algorithm may converge to numerous local minima and depends highly on initial cluster centers [35].
 451 Finally, the reference configurations of the grid are given at the last iteration of the method, that is when
 452 $\mu_{r+1} = \mu_r$ for some $r \in \mathbb{N}_0$. In such a way, the reference configuration h_n is $\mu_n^{(r)}$ for each $n = 1, \dots, N$.

453 5. Numerical Results

454 In this section, we present our model's numerical results. Because we focus on the application for
 455 an electrical grid, we consider Boolean temporal sequences describing the breakers' states (*off/on*) of the
 456 network. First, we evaluate a simulated scenario in which the Markov state process, that represents the
 457 hidden process, is known. Recall that only the observed processes are used to infer the hidden process. After
 458 the temporal evolution of the hidden process is estimated, we compare it with the "real" state process for
 459 validation. Second, the proposed modeling is confronted with real data provided by the France's transmission
 460 system operator (RTE). In both cases, we fix the states for the observable processes as binary values, i.e.,
 461 $\mathbb{M}_p = \mathbb{K}_p = \{0, 1\}$, $R = 1$, and $M_p = K_p = 2$ for each $p = 1, \dots, P$, see eq. (2). In this way, the underlying
 462 space of the observed states is $\mathbb{Y}_p = \{f_1^p, f_2^p\}$, where $f_1^p = (1, 0)$ and $f_2^p = (0, 1)$.

463 5.1. Simulated Data

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

465 **5.1.1. Initial Parameters**

466 We fix first the space of $N = 4$ hidden states, belonging to the space \mathbb{H} considered here as:

$$\mathbb{H} = \{(1, 1, 1, 1, 0, 1), (0, 1, 0, 0, 0, 1), (1, 0, 1, 1, 1, 1), (0, 1, 0, 1, 0, 0)\}. \quad (40)$$

467 Thus, the space $\mathbb{X} = \{e_1, e_2, e_3, e_4\}$ stands for the space of canonical vectors in \mathbb{R}^4 , where $e_1 = (1, 0, 0, 0)$,
 468 $e_2 = (0, 1, 0, 0)$, $e_3 = (0, 0, 1, 0)$, $e_4 = (0, 0, 0, 1)$. The simulation of $X = \{X_t\}_{t \geq 0}$ and $Y^p = \{Y_t^p\}_{t \geq 0}$,
 469 $p = 1, \dots, 6$, are obtained by the classical simulation procedure of jump chains and holding times with
 470 exponential distribution, see, e.g., [36, Section 2.6]. The Markov processes are performed under the fixed
 471 matrices of eq. (B.1) and Table B.1 in Appendix B.1.1. The sample path for X is shown on the left side
 472 in Figure 5. This simulation was stopped at 50 jump-events, giving a horizon time of two years from 2018
 473 to 2020. The total number of jump-events of the observed states is 1964. Each simulated observed process
 474 is shown in Figure 4.

475 Concerning the parameters of our model, we compute the initial estimation of the matrices $\hat{A}^{(0)}$ and
 476 $\hat{C}_m^{p(0)}$, $p = 1, \dots, 6$, $m = 1, 2$, by the empirical estimation of eq. (37). These values are shown in eq. (B.2) and
 477 Table B.2 in Appendix B.1.2, resp. The initial filter estimate of X is obtained from eq. (38). Under the
 478 simulated data, we obtain $\bar{\sigma}_0(\hat{X}_0) = (0.307, 0.196, 0.388, 0.109)$. The initial state for X is therefore chosen
 479 to be $\hat{X}_0 = e_3$ by eq. (32). Thus, by eq. (4), the initial hidden state is $h_3 = (1, 0, 1, 1, 1, 1)$. Finally, for the
 480 stopping criteria of eq. (31), we choose $\varepsilon = 10^{-5}$.

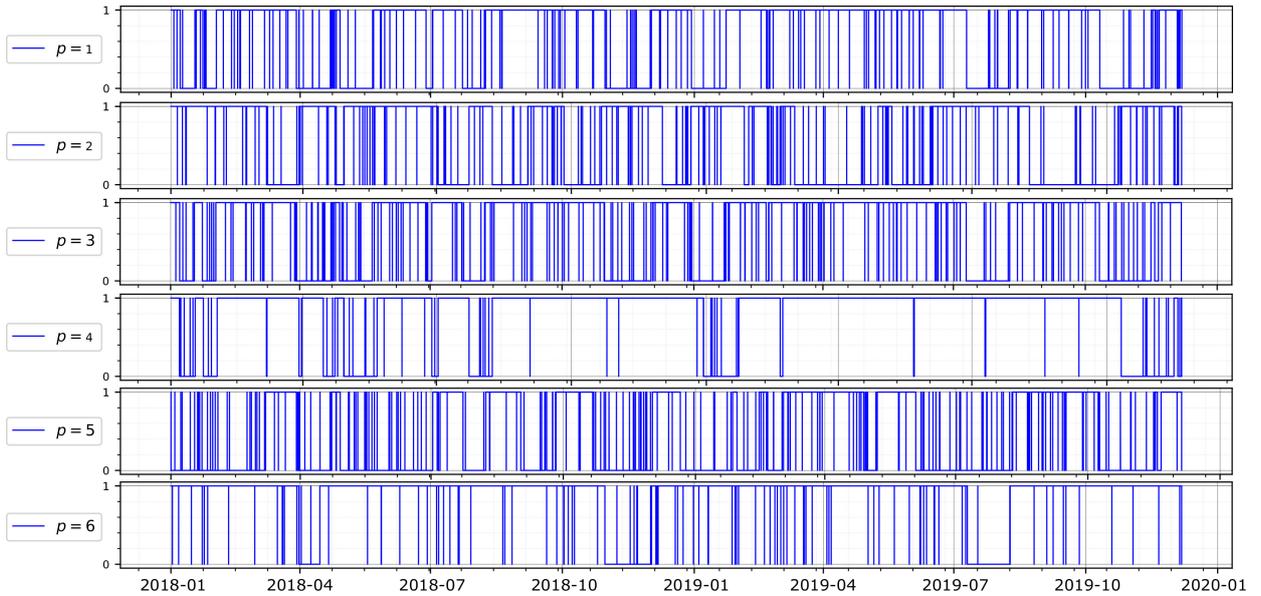


Figure 4: Simulated temporal evolution of the observable processes.

481 **5.1.2. Hidden State Estimation by Clustering Method**

482 On the left side in Figure 5, we observe the “real” temporal evolution of the hidden Markov process X .
 483 The first estimation that we do is the empirical estimation by clustering method by using eq. (36). This
 484 approach is a classification procedure that takes the joint information of all observed processes at each time
 485 and computes the arg min set to know at which cluster the hidden process is. A cluster is obtained by finding
 486 the points of the observed processes’ joint information that have the minimum distance to a cluster center,
 487 see eq. (39), where the centers are the states of eq. (40). The metric distance that we use is the euclidean
 488 one. The purpose of using this method is only for comparison. This paper does not aim to find the best
 489 clustering method with the appropriate distance measure. A study about this subject will be addressed in
 490 the future, based, e.g., on [37].

491 Under the clustering method with the euclidean distance, the arg min set might not be a single point
492 several times because there are points of the joint information of all observable processes at the same distance
493 of different vectors of \mathbb{H} . In such cases, the clustering method is not exact because we cannot know which
494 state the hidden process is. This is represented with green points on the left side in Figure 5. When a cluster
495 is active, i.e., when the hidden state can be chosen by the clustering method, 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.

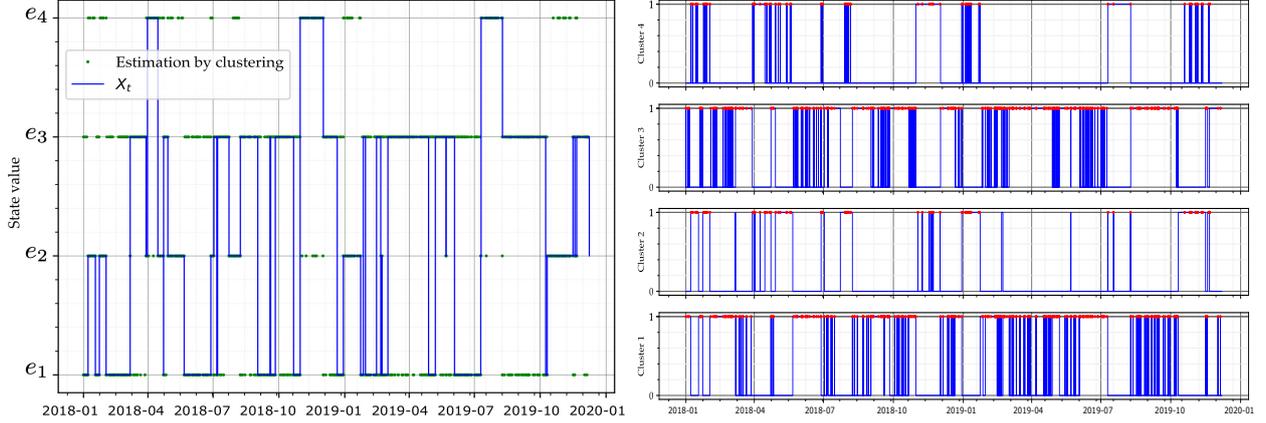


Figure 5: On the left side, a sample path of the hidden Markov state process X (blue line) representing the hidden state over time, and the estimation by clustering method (green points). At several times, the arg min set for clustering could not be a singleton. On the right side, the temporal evolution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

496 5.1.3. Hidden State Estimation by HMM

497 Figure 6 shows the filter estimate of the hidden state over the set \mathbb{H} of eq. (40). The hidden process
498 X represents this over the canonical vectors of \mathbb{R}^4 . On the left side in Figure 6, the filtering is done using
499 the (empirical) initial parameter estimation of eq. (37). At first glance, the filter estimate “jumps” several
500 times when the “real” temporal evolution of the hidden process does not. This is because the first parameter
501 estimation is not exact. However, when the parameter estimation is computed through the Theorem 3.2,
502 the filter estimate fits better as the number of iterations increases. This is confirmed by the Mean Squared
503 Error (MSE) between the “real” values of X , and the filter estimate \hat{X} . The MSE value obtained at the
504 last iteration is less than 0.09, showing the accuracy of our model over the simulated scenario. The filter
505 estimate for the last iteration is shown in Figure 6 on the right side.
506

507 5.1.4. Parameter Estimation

508 Figure 7 shows the values of the parameters of our model over the number of iterations, i.e., the values
509 of the matrices $\hat{A}^{(r)}$ and $\hat{C}_m^{p(r)}$, $p = 1, \dots, 6$, $m = 1, 2$, for each iteration $r = 0, \dots, 14$. The first estimation
510 is made by eq. (37) and the last one is obtained when the stopping criteria of eq. (31) is verified with $\varepsilon = 10^{-5}$.
511

512 5.1.5. Uncertainty and Final Hidden State

513 Using the filter estimate of the hidden state, we can compute the function ϵ_t of eq. (33), that represents
514 a temporal evolution’s uncertainty signal on the vectors in \mathbb{H} . Figure 8 shows the values of this uncertainty
515 over time for the first and the last iterations. Comparing the two pictures, we observe that there are fewer
516 peaks at the last iteration compared to the first iteration. This is because finding the optimal parameters to
517 fit our model, the uncertainty decreases as the number of iterations increases. In the Figure 8, we also see
518 the mean δ_t of eq. (34). At the last iteration when the parameters are fitted, less than 1% of the time, ϵ_t
519 is over δ_t . When ϵ_t is over δ_t could mean that the state at which the hidden process “jumps” is known with
520 some uncertainty. However, such uncertainty is almost instantaneous because it is not remaining in time,
521 as reflected in Figure 8.

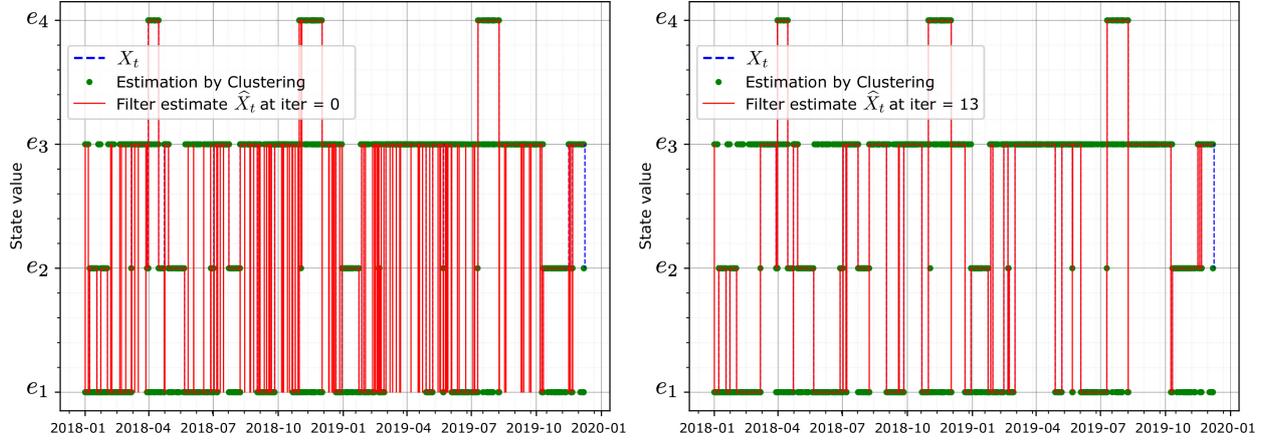


Figure 6: Filter estimate of the hidden process X . The “real” state over time is shown by the blue dashed line. The filter is obtained by the parameter estimation at the first iteration (on the left side) and the final one (on the right side).

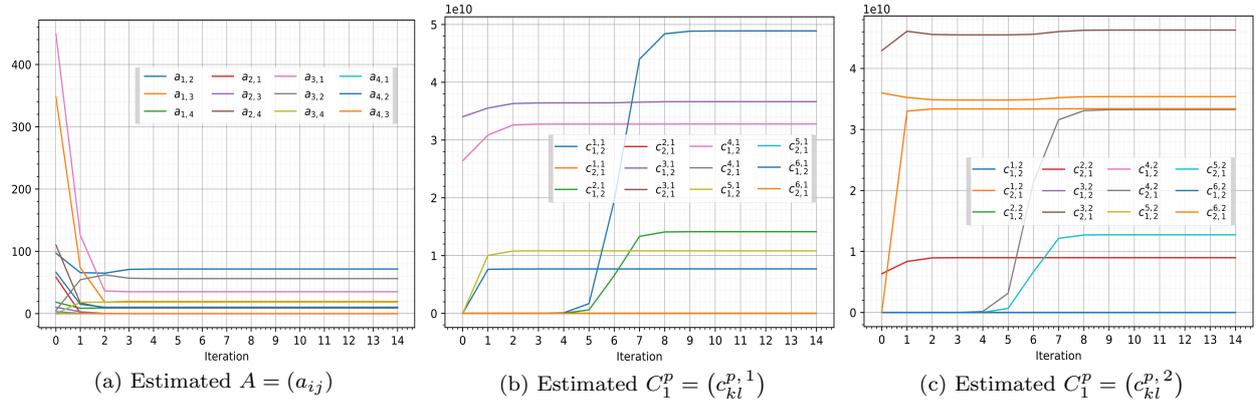


Figure 7: Estimated values for the matrices A , C_1^p , and C_2^p , resp., for each $p = 1, \dots, 6$, over the number of iterations.

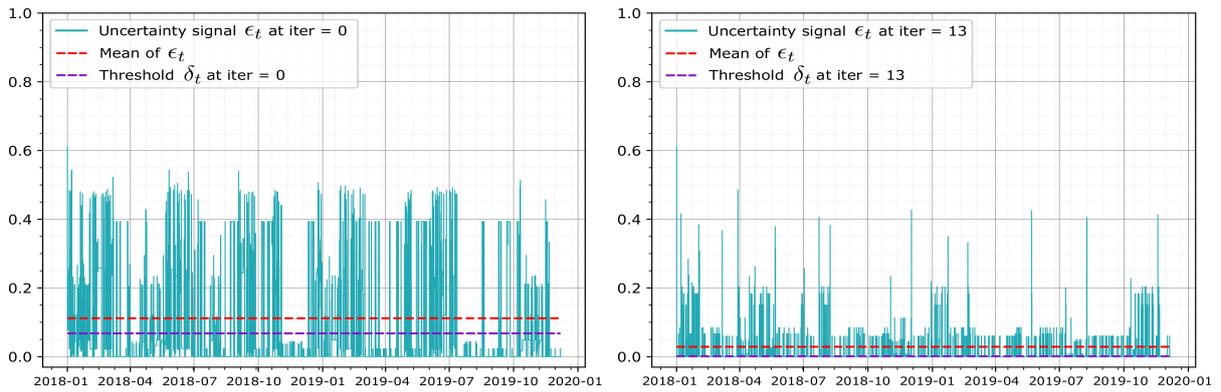


Figure 8: Temporal evolution's uncertainty signal on the states of \mathbb{H} at the first iteration (on the left side) and on the last iteration 14 (on the right side).

522 **5.2. Real Data**

523 In this section, our model is confronted with real data provided by France’s transmission system operator
 524 (RTE). The available data consists of historical records of the evolution of the breakers’ states over a period
 525 during which the grid operated normally.

526 **5.2.1. Initial Parameters**

527 We take from the data an electrical transmission grid with $P = 6$ breakers, with states *off/on* for each
 528 one. The observations’ horizon time takes place on three months from 2016 – 01 to 2019 – 01, giving a total
 number of jump-events of 1264. The breakers’ states over time are shown in Figure 9.

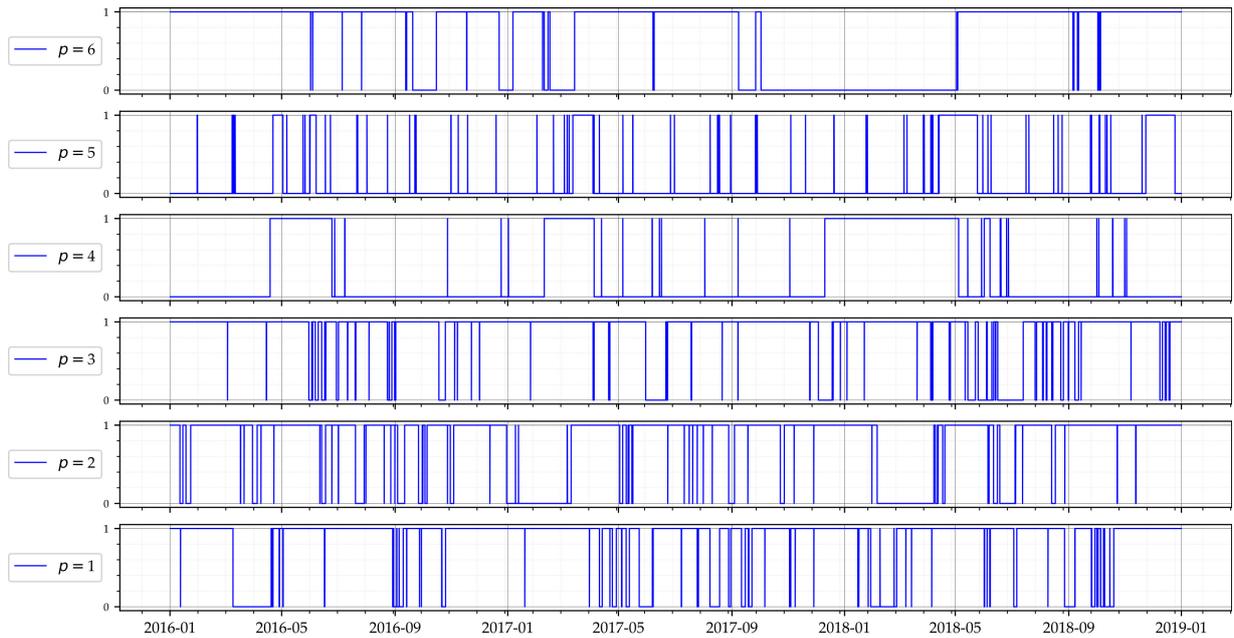


Figure 9: Breakers state’s temporal evolution obtained from real data provided by RTE.

529 Using the clustering method presented in Section 4.5, we choose $N = 3$ reference states defining the
 530 space $\mathbb{H} = \{h_1, h_2, h_3\}$ by:
 531

$$\mathbb{H} = \{(1, 1, 0, 1, 0, 1), (0, 0, 1, 0, 0, 0), (1, 1, 1, 1, 1, 1)\} . \quad (41)$$

532 Thus, the space of $\mathbb{X} = \{e_1, e_2, e_3\}$ stands for the space of canonical vectors of \mathbb{R}^3 , where $e_1 = (1, 0, 0, 0)$,
 533 $e_2 = (0, 1, 0, 0)$, $e_3 = (0, 0, 1, 0)$. The initial filter estimate of the grid (represented equivalently by X) is
 534 obtained from eq. (38). Under the real data, we obtain $\bar{\sigma}_0(\hat{X}_0) = (0.111, 0.785, 0.104)$. The initial state for
 535 X is therefore chosen to be $\hat{X}_0 = e_2$ by eq. (32). Thus, by eq. (4), the initial state of the grid is h_2 , see
 536 eq. (41). The initial estimation of the matrices $\hat{A}^{(0)}$ and $\hat{C}_m^{p(0)}$, $p = 1, \dots, 6$, $m = 1, 2$, is computed by the
 537 empirical estimation of eq. (37). Finally, for the stopping criteria of eq. (31), we fix $\varepsilon = 10^{-3}$.

538 **5.2.2. Grid State Estimation by Clustering Method**

539 In Figure 10, we observe on the left side the temporal evolution of the empirical estimation of the
 540 grid by clustering method, see eq. (36). At any time, all breakers states’ joint information is classified
 541 with the minimum distance to the cluster centers, see eq. (39). Here, the cluster centers are the reference
 542 configurations of eq. (41). However, the *argmin* set might not be a single point because of the nature of the
 543 dataset. In such cases, the clustering method is not exact because we cannot know at which configuration

544 the grid is. The green dots show all the points of the arg min set. Because of the several choices, we represent
545 the first choice of the arg min set by the blue line in the same picture. Also, we show on the right side in
546 Figure 10 when a cluster is active, i.e., when the reference configurations can be chosen by clustering. The
547 red dots represent when there is more than one choice. The metric distance used is the euclidean distance.
548 This is a crucial aspect in our estimation since the first estimation of the matrices $\hat{A}^{(0)}$ and $\hat{C}_m^{p(0)}$, $p = 1, \dots, 6$,
549 $m = 1, 2$, may be affected by this issue. So, by taking the first point of the arg min set at each time, we
550 have estimated using eq. (37) these matrices. Thus, this estimation might not be correct and could affect
the estimation by HMM.

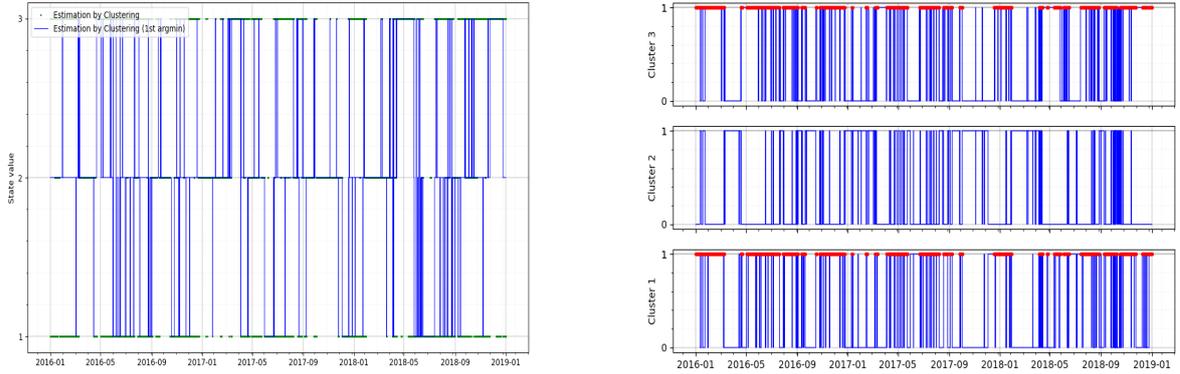


Figure 10: On the left side, the estimation by clustering method (green points) from all joint information of the breaker states over time. The blue line is the first element taken from the arg min set when classification is made, since at several times the arg min set for clustering could not be a singleton. 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).

551

552 5.2.3. Grid State Estimation by HMM

553 We show the filter estimate of the grid's temporal evolution over the reference configurations in Figure 11.
554 Recall that the estimation is represented (equivalently) by the process X over the canonical vectors of \mathbb{R}^3 .
555 On the left side in Figure 11, the first filter estimation is done using the (empirical) initial parameter
556 estimation by clustering method that could not be correct. So, we iterate the estimation of the parameters
557 using Theorem 3.2. In the simulated case shown in Figure 6, it shows that the filter estimate fits better as
558 the number of iterations increases. However, we can not compute the MSE between the real values of X
559 and the filter estimate \hat{X} , because the real grid's temporal evolution is not known. The filter estimate for
560 the last iteration is shown in Figure 11 on the right side.

561 5.2.4. Uncertainty and Final Hidden State of the Grid

562 Now, we focus on the uncertainty in the grid by using ϵ_t , see eq. (33). We show in Figure 12 the
563 values of the uncertainty signal ϵ_t that represents a temporal evolution's uncertainty signal on the reference
564 configurations. We notice that there are fewer peaks in ϵ_t at the last iteration compared to the first iteration.
565 This is because finding the optimal parameters to fit our model, the uncertainty decreases as the number
566 of iterations increases. Recall that when ϵ_t is over δ_t , the state at which the grid "jumps" is known with
567 some uncertainty. However, such uncertainty is almost instantaneous because it does not remain in time, as
568 reflected in Figure 12.

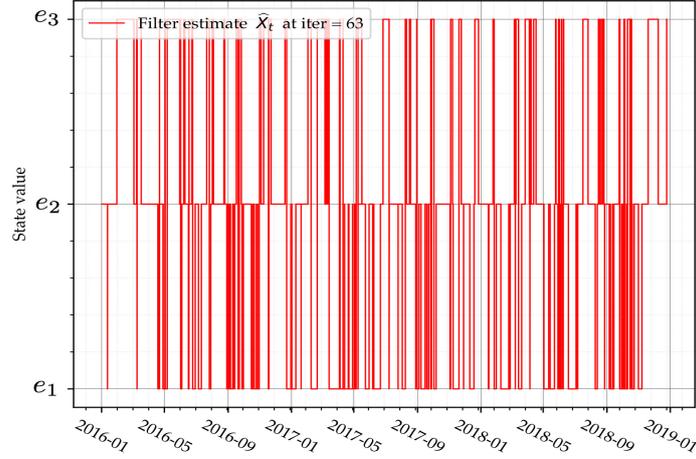


Figure 11: Final filter estimate \hat{X} for the temporal evolution of the grid.

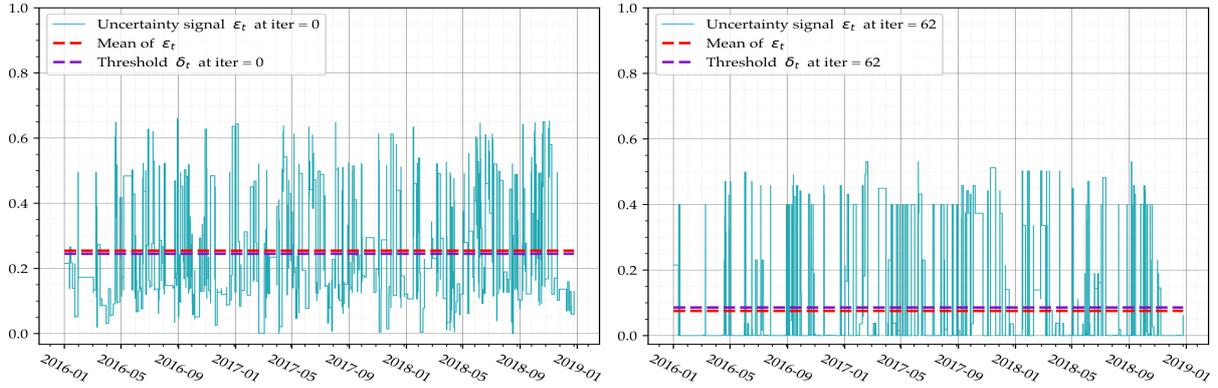


Figure 12: Temporal evolution's uncertainty signal on the reference configurations of the grid at the first iteration (on the left side) and on the last iteration 62 (on the right side).

6. Conclusion and Remarks

In this paper, we have proposed a general data-driven approach for the hidden MC with several observed process. 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. A future work will be to identify the "best" number of hidden states, based for example on [38].

584 Using our theoretical results, we have then shown the performance of the framework by considering
585 Boolean temporal sequences describing the breakers' states (*off/on*) of the grid. First, we have evaluated
586 a simulated scenario, showing the advantages of the HMM approach with the proposed strong scheme.
587 Second, we have confronted our model with real data provided by the France's transmission system operator
588 (RTE), showing stability in the filter estimates for the parameters and the hidden state of the grid. We
589 finally identify the normal behavior of the French electrical grid, which could be embedded in monitoring
590 and detection algorithm in the future.

591 **Appendix A. Mathematical Proofs**

592 **Proof 1** [THEOREM 3.2]. The idea is to change the parameters that define the “intensities” of the counting
 593 processes $J_{ij,t}$ and $N_{kl,t}^p$ of X and Y^p resp., i.e., to modify a_{ij} and $c_{kl}^{p,m}$ to \hat{a}_{ij} and $\hat{c}_{kl}^{p,m}$ resp. Next, we
 594 proceed to the maximization step. More precisely, for each $p = 1, \dots, P$, the parameters $c_{kl}^{p,m}$, $m = 1, \dots, M_p$,
 595 define $\lambda_{kl,t}^p$ of the process $N_{kl,t}^p$, see eq. (12), and a_{ij} appears in $J_{ij,t}$ since its semi-Martingale decomposition
 596 is:

$$J_{ij,t} = \int_0^t \langle e_i, X_{s-} \rangle \langle e_j, dX_s \rangle = \langle e_i, X_{s-} \rangle a_{ji} dt + Q_{ij,t},$$

597 from eq. (5), where $Q_{ij,t} := J_{ij,t} - \int_0^t \langle e_i, X_{s-} \rangle a_{ji} ds$ is a $(\mathcal{G}_t, \mathbb{P})$ -Martingale [15].

598 To estimate all the parameters, we define a new probability measure⁶ $\hat{\mathbb{P}} = \mathbb{P}_{\hat{\theta}}$, with $\hat{\theta} \in \Theta$, see eq. (19),
 599 for a “fictitious world” from the probability measure $\mathbb{P} = \mathbb{P}_{\theta}$, with $\theta \in \Theta$, of the “real world”, by putting:

$$\begin{aligned} \left. \frac{d\hat{\mathbb{P}}}{d\mathbb{P}} \right|_{\mathcal{G}_t} = \hat{\Lambda}_t := \exp \left\{ - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln \left(\frac{\lambda_{kl,s-}^p}{\hat{\lambda}_{kl,s-}^p} \right) dN_{kl,s}^p - \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \ln \left(\frac{a_{ji}}{\hat{a}_{ji}} \right) dJ_{ij,s} \right. \\ \left. + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t (\lambda_{kl,s}^p - \hat{\lambda}_{kl,s}^p) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle X_s, e_i \rangle (a_{ji} - \hat{a}_{ji}) ds \right\}, \end{aligned} \quad (\text{A.1})$$

600 where $\hat{\lambda}_{kl,t}^p = \langle f_k^p, Y_t^p \rangle \sum_{m=1}^{M_p} \hat{c}_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_t, e_n \rangle$. The log-likelihood is, therefore:

$$\begin{aligned} \ln(\hat{\Lambda}_t) = & - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \ln \left(\frac{\lambda_{kl,s-}^p}{\hat{\lambda}_{kl,s-}^p} \right) dN_{kl,s}^p - \sum_{\substack{i,j=1 \\ i \neq j}}^N \ln \left(\frac{a_{ji}}{\hat{a}_{ji}} \right) J_{ij,t} \\ & + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t (\lambda_{kl,s}^p - \hat{\lambda}_{kl,s}^p) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N (a_{ji} - \hat{a}_{ji}) O_{i,t}, \end{aligned} \quad (\text{A.2})$$

601 where we have used the definition of the process $O_{i,t}$ in eq. (20). Now, noting that

$$\lambda_{kl,t}^p - \hat{\lambda}_{kl,t}^p = \langle f_k^p, Y_t^p \rangle \sum_{m=1}^{M_p} (c_{lk}^{p,m} - \hat{c}_{lk}^{p,m}) \sum_{n \in I_m^p} \langle X_t, e_n \rangle,$$

602 and that, since X_t takes values in the space \mathbb{X} of unit vectors of \mathbb{R}^N ,

$$\ln \left(\frac{\lambda_{kl,t}^p}{\hat{\lambda}_{kl,t}^p} \right) = \ln \left(\frac{\sum_{m=1}^{M_p} c_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_t, e_n \rangle}{\sum_{m=1}^{M_p} \hat{c}_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_t, e_n \rangle} \right) = \sum_{m=1}^{M_p} \ln \left(\frac{c_{lk}^{p,m}}{\hat{c}_{lk}^{p,m}} \right) \sum_{n \in I_m^p} \langle X_t, e_n \rangle.$$

603 Thus, by rearranging terms in eq. (A.2) and using the definition of $L_{kl,t}^{p,n}$ and $S_{k,t}^{p,n}$ see eq. (21), we have:

⁶See, e.g., [24, Ch. VI, Eq. (2.17)] or [28, Lemma 4.7.3].

$$\begin{aligned} \ln(\widehat{\Lambda}_t) &= - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \ln \left(\frac{c_{lk}^{p,m}}{\widehat{c}_{lk}^{p,m}} \right) \sum_{n \in I_m^p} L_{kl,t}^{p,n} - \sum_{\substack{i,j=1 \\ i \neq j}}^N \ln \left(\frac{a_{ji}}{\widehat{a}_{ji}} \right) J_{ij,t} \\ &\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} (c_{lk}^{p,m} - \widehat{c}_{lk}^{p,m}) \sum_{n \in I_m^p} S_{k,t}^{p,n} + \sum_{\substack{i,j=1 \\ i \neq j}}^N (a_{ji} - \widehat{a}_{ji}) O_{i,t}. \end{aligned}$$

604 Taking now $\mathbb{E}[\cdot | \mathcal{Y}_t]$ above and using notation of eq. (A.1),

$$\begin{aligned} \mathbb{E} \left[\ln \left(\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} \right) \middle| \mathcal{Y}_t \right] &= - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \ln \left(\frac{c_{lk}^{p,m}}{\widehat{c}_{lk}^{p,m}} \right) \sum_{n \in I_m^p} \mathbb{E}[L_{kl,t}^{p,n} | \mathcal{Y}_t] - \sum_{\substack{i,j=1 \\ i \neq j}}^N \ln \left(\frac{a_{ji}}{\widehat{a}_{ji}} \right) \mathbb{E}[J_{ij,t} | \mathcal{Y}_t] \\ &\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} (c_{lk}^{p,m} - \widehat{c}_{lk}^{p,m}) \sum_{n \in I_m^p} \mathbb{E}[S_{k,t}^{p,n} | \mathcal{Y}_t] + \sum_{\substack{i,j=1 \\ i \neq j}}^N (a_{ji} - \widehat{a}_{ji}) \mathbb{E}[O_{i,t} | \mathcal{Y}_t]. \end{aligned}$$

605 The unique maximum above over \widehat{a}_{ji} and $\widehat{c}_{lk}^{p,m}$ occurs at such values obtained by equaling to zero the partial
606 derivatives:

$$\frac{\partial}{\partial \widehat{a}_{ji}} \mathbb{E} \left[\ln \left(\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} \right) \middle| \mathcal{Y}_t \right] = 0 \quad \text{and} \quad \frac{\partial}{\partial \widehat{c}_{lk}^{p,m}} \mathbb{E} \left[\ln \left(\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} \right) \middle| \mathcal{Y}_t \right] = 0.$$

607 Thus, we have $\widehat{a}_{ji} = \frac{\mathbb{E}[J_{ij,t} | \mathcal{Y}_t]}{\mathbb{E}[O_{i,t} | \mathcal{Y}_t]}$ and $\widehat{c}_{lk}^{p,m} = \frac{\sum_{n \in I_m^p} \mathbb{E}[L_{kl,t}^{p,n} | \mathcal{Y}_t]}{\sum_{n \in I_m^p} \mathbb{E}[S_{k,t}^{p,n} | \mathcal{Y}_t]}$. ■

608 **Proof 2** [THEOREM 3.3]. We proceed as in [15, Theorem 3.2] but in our context. To derive an equation
609 for $\bar{\sigma}_t(F_t X_t)$, we use first Ito's product rule on $F_t X_t$. Let $\Delta X_s := X_s - X_{s-}$. The following holds:

$$\begin{aligned} F_t X_t &= F_0 X_0 + \int_0^t X_{s-} dF_s + \int_0^t F_{s-} dX_s + [F, X]_t \\ &= F_0 X_0 + \int_0^t X_{s-} \alpha_s ds + \int_0^t X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t F_{s-} X_s ds + \int_0^t F_{s-} dV_s + \sum_{0 < s \leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s \\ &\quad \left(\text{by using eqs. (5) and (23), and since } [F, X]_t = \sum_{0 < s \leq t} \Delta F_s \Delta X_s = \sum_{0 < s \leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s \right) \\ &= F_0 X_0 + \int_0^t X_{s-} \alpha_s ds + \int_0^t X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t F_{s-} X_s ds + \int_0^t F_{s-} dV_s \tag{A.3} \\ &\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{ji} (e_j - e_i) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle (\beta_{j,s} - \beta_{i,s}) X_{s-}, e_i \rangle \langle e_j, dV_s \rangle (e_j - e_i) \\ &\quad \left(\text{since } \sum_{0 < s \leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s = \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle (\beta_{j,s} - \beta_{i,s}) X_{s-}, e_i \rangle \langle e_j, dX_s \rangle (e_j - e_i) \text{ and using eq. (5)} \right). \end{aligned}$$

610 Taking now Ito's product rule on $\bar{\Lambda}_t F_t X_t$, we have that:

$$\begin{aligned}
\bar{\Lambda}_t F_t X_t &= \bar{\Lambda}_0 F_0 X_0 + \int_0^t F_{s-} X_{s-} d\bar{\Lambda}_s + \int_0^t \bar{\Lambda}_{s-} dF_s X_s + [\bar{\Lambda}, FX]_t \\
&= F_0 X_0 + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} F_{s-} X_{s-} \lambda_{kl,s-}^p d(N_{kl,s}^p - s) - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} F_{s-} X_{s-} d(N_{kl,s}^p - s) \\
&\quad + \int_0^t \bar{\Lambda}_{s-} X_{s-} \alpha_s ds + \int_0^t \bar{\Lambda}_{s-} X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t \bar{\Lambda}_{s-} F_{s-} X_s ds + \int_0^t \bar{\Lambda}_{s-} F_{s-} dV_s \\
&\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{ji} (e_j - e_i) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_{s-}, e_i \rangle \langle e_j, dV_s \rangle (e_j - e_i) \\
&\quad (\text{since } \bar{\Lambda}_0 = 1, \text{ using eq. (A.3), and because there is no jump at the sametime a.s. } [\bar{\Lambda}, FX]_t = 0) \\
&= F_0 X_0 + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle \bar{\Lambda}_{s-} F_{s-} \left(\sum_{m=1}^{M_p} c_{lk}^{p,m} \sum_{n \in I_m^p} \langle X_{s-}, e_n \rangle e_n \right) d(N_{kl,s}^p - s) \\
&\quad - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} F_{s-} X_{s-} d(N_{kl,s}^p - s) \\
&\quad + \int_0^t \bar{\Lambda}_{s-} X_{s-} \alpha_s ds + \int_0^t \bar{\Lambda}_{s-} X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t \bar{\Lambda}_{s-} F_{s-} X_s ds + \int_0^t \bar{\Lambda}_{s-} F_{s-} dV_s \\
&\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{ji} (e_j - e_i) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_{s-}, e_i \rangle \langle e_j, dV_s \rangle (e_j - e_i) \\
&\quad (\text{by using eq. (12) and because } X_t \langle X_t, e_n \rangle = \langle X_t, e_n \rangle e_n, \forall t \geq 0) \\
&= F_0 X_0 - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} F_{s-} X_{s-} d(N_{kl,s}^p - s) \\
&\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle \bar{\Lambda}_{s-} F_{s-} \left(\sum_{m=1}^{M_p} c_{lk}^{p,m} \sum_{n=1}^N \text{diag}_m^p \langle X_{s-}, e_n \rangle e_n \right) d(N_{kl,s}^p - s) \\
&\quad + \int_0^t \bar{\Lambda}_{s-} X_{s-} \alpha_s ds + \int_0^t \bar{\Lambda}_{s-} X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t \bar{\Lambda}_{s-} F_{s-} X_s ds + \int_0^t \bar{\Lambda}_{s-} F_{s-} dV_s \\
&\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{ji} (e_j - e_i) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-} (\beta_{j,s} - \beta_{i,s}) X_{s-}, e_i \rangle \langle e_j, dV_s \rangle (e_j - e_i) \\
&\quad (\text{by using the definition of the diagonal matrix } \text{diag}^p \text{ see eq. (9)}) \\
&= F_0 X_0 - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \int_0^t \bar{\Lambda}_{s-} F_{s-} X_{s-} d(N_{kl,s}^p - s) + \int_0^t \bar{\Lambda}_{s-} X_{s-} \langle \beta_s, dV_s \rangle + A \int_0^t \bar{\Lambda}_{s-} F_{s-} X_s ds \\
&\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\Lambda}_{s-} F_{s-} X_{s-} d(N_{kl,s}^p - s) + \int_0^t \bar{\Lambda}_{s-} X_{s-} \alpha_s ds + \int_0^t \bar{\Lambda}_{s-} F_{s-} dV_s
\end{aligned}$$

$$+ \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-}(\beta_{j,s} - \beta_{i,s})X_s, e_i \rangle a_{ji}(e_j - e_i) ds + \sum_{\substack{i,j=1 \\ i \neq j}}^N \int_0^t \langle \bar{\Lambda}_{s-}(\beta_{j,s} - \beta_{i,s})X_{s-}, e_i \rangle \langle e_j, dV_s \rangle (e_j - e_i),$$

611 where the last equality holds because $X_t = \sum_{n=1}^N \langle X_t, e_n \rangle e_n, \forall t \geq 0$. Taking now $\bar{\mathbb{E}}[\cdot | \mathcal{Y}]$ above and
 612 interchanging expectation and integration [39], we get the result. \blacksquare

613 **Proof 3** [THEOREM 3.6]. We proceed as in Theorem 3.3. To derive an equation for $\bar{\sigma}_t(L_{kl,t}^{p,n}X_t)$, we use
 614 first Ito's product rule on $L_{kl,t}^{p,n}X_t$. The following holds:

$$\begin{aligned} L_{kl,t}^{p,n}X_t &= L_{kl,0}^{p,n}X_0 + \int_0^t X_{s-} dL_{kl,s}^{p,n} + \int_0^t L_{kl,s-}^{p,n} dX_s + [L_{kl}^{p,n}, X]_t \\ &= \int_0^t X_s \langle e_n, X_s \rangle ds + \int_0^t X_{s-} \langle e_n, X_{s-} \rangle d(N_{kl,s}^p - s) + A \int_0^t L_{kl,s}^{p,n} X_s ds + \int_0^t L_{kl,s}^{p,n} dV_s, \end{aligned} \quad (\text{A.4})$$

615 where we have used the eqs. (5) and (25), and since there is no jump at the same time a.s, $[L_{kl}^{p,n}, X]_t = 0$.
 616 Taking now Ito's product rule on $\bar{\Lambda}_t F_t X_t$, and considering the eq. (16) and the eq. (A.4), we have:

$$\begin{aligned} \bar{\Lambda}_t L_{kl,t}^{p,n}X_t &= \int_0^t L_{kl,s-}^{p,n} X_{s-} d\bar{\Lambda}_s + \int_0^t \bar{\Lambda}_{s-} dL_{kl,s}^{p,n} X_s + [\bar{\Lambda}, L_{kl}^{p,n} X]_t \\ &= \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \int_0^t \bar{\Lambda}_{s-} L_{kl,s-}^{p,n} X_{s-} \lambda_{uv,s-}^q d(N_{uv,s}^q - s) - \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \int_0^t \bar{\Lambda}_{s-} L_{kl,s-}^{p,n} X_{s-} d(N_{uv,s}^q - s) \\ &\quad + A \int_0^t \bar{\Lambda}_{s-} L_{kl,s}^{p,n} X_s ds + \int_0^t \bar{\Lambda}_{s-} L_{kl,s}^{p,n} dV_s + \int_0^t \bar{\Lambda}_{s-} \langle e_n, X_{s-} \rangle X_{s-} \lambda_{kl,s-}^p dN_{kl,s}^p \\ &\quad \left(\text{since there are common jumps, } [\bar{\Lambda}, L_{kl}^{p,n} X]_t = \int_0^t \bar{\Lambda}_{s-} (\lambda_{kl,s-}^p - 1) \langle e_n, X_{s-} \rangle X_{s-} dN_{kl,s}^p \right) \\ &= \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \sum_{m=1}^{M_q} \int_0^t \langle f_u^q, Y_{s-}^q \rangle c_{vu}^{q,m} \text{diag}_m^q \bar{\Lambda}_{s-} L_{kl,s-}^{p,n} X_{s-} d(N_{uv,s}^q - s) \\ &\quad - \sum_{q=1}^P \sum_{\substack{u,v=1 \\ u \neq v}}^{K_q} \int_0^t \bar{\Lambda}_{s-} L_{kl,s-}^{p,n} X_{s-} d(N_{uv,s}^q - s) + A \int_0^t \bar{\Lambda}_{s-} L_{kl,s}^{p,n} X_s ds + \int_0^t \bar{\Lambda}_{s-} L_{kl,s}^{p,n} dV_s \\ &\quad + \sum_{m=1}^{M_p} \int_0^t \langle f_k^p, Y_{s-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \langle e_n, \bar{\Lambda}_{s-} X_{s-} \rangle e_n dN_{kl,s}^p \\ &\quad \left(\text{because } X_t \lambda_{kl,t}^p = \sum_{m=1}^{M_p} \langle f_k^p, Y_t^p \rangle c_{lk}^{p,m} \text{diag}_m^p X_t \text{ and } X_t \langle e_n, X_t \rangle = \langle e_n, X_t \rangle e_n, \forall t \geq 0 \right). \end{aligned}$$

617 Taking $\bar{\mathbb{E}}[\cdot | \mathcal{Y}]$ above and interchanging expectation and integration [39], we get the result. \blacksquare

618 **Proof 4** [THEOREM 4.1]. First, let's focus on the no-jump part of the SDE in eq. (27) and eq. (28), i.e, for
 619 t not be jump time, let say $t \in (t_w, t_{w+1})$, $w = 0, \dots, W$, we consider

$$\begin{aligned} dG_t &= (\Upsilon_t G_t + \Gamma_t K_t) dt \\ dK_t &= \Xi_t K_t dt, \end{aligned}$$

620 which is an ordinary linear differential equation system. Recall $\Upsilon_t, \Gamma_t, \Xi_t \in \mathbb{R}^{N \times N}$ are constant matrices
 621 matrices between jumps. Now, let

$$Z_t := \begin{bmatrix} G_t \\ K_t \end{bmatrix} \quad (\text{A.5})$$

622 the column vector in \mathbb{R}^{2N} . Thus we can redefine the ordinary differential equation system by:

$$dZ_t = \begin{bmatrix} \Upsilon_t & \Gamma_t \\ \mathbf{0}_{N \times N} & \Xi_t \end{bmatrix} Z_t dt \quad , \quad Z_{t_w} = \begin{bmatrix} G_{t_w} \\ K_{t_w} \end{bmatrix} ,$$

623 where $\mathbf{0}_{N \times N}$ is the zero matrix in $\mathbb{R}^{N \times N}$ and Z_{t_w} is the initial condition $\forall t \in (t_w, t_{w+1})$. The solution of
 624 the latter differential equation with initial condition at t_w is:

$$Z_t = \exp \left\{ \begin{bmatrix} \Upsilon_{t_w} & \Gamma_{t_w} \\ \mathbf{0}_{N \times N} & \Xi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} Z_{t_w} , \quad (\text{A.6})$$

625 where $\Delta t = t - t_w$ is the length of $[t_w, t]$. Consider now the definition of Z_t in eq. (A.5) but for $t = t_{w+1}^-$
 626 (i.e., before the next jump-time t_{w+1}), and the initial condition Z_{t_w} described above. First, we multiply by
 627 the matrix $[\text{Id}_N \quad \mathbf{0}_{N \times N}]$ on both sides of the eq. (A.6), where Id_N is the identity matrix in $\mathbb{R}^{N \times N}$. We
 628 have:

$$G_{t_{w+1}^-} = [\text{Id}_N \quad \mathbf{0}_{N \times N}] \exp \left\{ \begin{bmatrix} \Upsilon_{t_w} & \Gamma_{t_w} \\ \mathbf{0}_{N \times N} & \Xi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} G_{t_w} \\ K_{t_w} \end{bmatrix} .$$

629 Second, we multiply by the matrix $[\mathbf{0}_{N \times N} \quad \text{Id}_N]$ on both sides of the eq. (A.6), and we obtain:

$$\begin{aligned} K_{t_{w+1}^-} &= [\mathbf{0}_{N \times N} \quad \text{Id}_N] \exp \left\{ \begin{bmatrix} \Upsilon_{t_w} & \Gamma_{t_w} \\ \mathbf{0}_{N \times N} & \Xi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} G_{t_w} \\ K_{t_w} \end{bmatrix} \\ &= [\mathbf{0}_{N \times N} \quad \text{Id}_N] \begin{bmatrix} \exp \{ \Upsilon_{t_w} \Delta t_{w+1} \} & \int_0^{\Delta t_{w+1}} \exp \{ \Upsilon_{t_w} (\Delta t_{w+1} - s) \} \Gamma_{t_w} \exp \{ \Xi_{t_w} s \} ds \\ \mathbf{0}_{N \times N} & \exp \{ \Xi_{t_w} \Delta t_{w+1} \} \end{bmatrix} \begin{bmatrix} G_{t_w} \\ K_{t_w} \end{bmatrix} \\ &= \exp \{ \Xi_{t_w} \Delta t_{w+1} \} K_{t_w} , \end{aligned}$$

630 where we have used the fact that the matrix within the exponential is block triangular, see for instance [40].

631 The jump-adapted almost exact solution scheme for K_t and G_t can be easily obtained from eq. (27) and
 632 eq. (28) by adding the effect of jumps, resp. \blacksquare

633 **Proof 5** [COROLLARY 4.3]. First, note that $\bar{\sigma}_t(F_t X_t)$ in eq. (24) can be rewritten by splitting it as:

$$\begin{aligned} d\bar{\sigma}_t(F_t X_t) &= A \bar{\sigma}_t(F_t X_t) dt + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \bar{\sigma}_t(F_t X_t) dt - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \bar{\sigma}_s(F_{t-} X_{t-}) dN_{kl,t}^p + \bar{\sigma}_t(X_t \alpha_t) dt \\ &+ \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \bar{\sigma}_t((\beta_{j,t} - \beta_{i,t}) X_t), e_i \rangle a_{ji} (e_j - e_i) dt - \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \langle f_k^p, Y_t^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_t(F_t X_t) dt \\ &+ \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \sum_{m=1}^{M_p} \langle f_k^p, Y_{t-}^p \rangle c_{lk}^{p,m} \text{diag}_m^p \bar{\sigma}_t(F_{t-} X_{t-}) dN_{kl,t}^p , \end{aligned}$$

634 wherein the integral w.r.t. Lebesgue measure dt , we are allowed to evaluate $\bar{\sigma}_t(F_{t-}X_{t-})$ as $\bar{\sigma}_t(F_tX_t)$.
635 Regrouping the terms and using the definition of Φ_t and Ψ_t in eq. (29) and eq. (30) resp., the following SDE
636 holds:

$$\begin{aligned} d\bar{\sigma}_t(F_tX_t) &= \Phi_t \bar{\sigma}_t(F_tX_t)dt + \left(\bar{\sigma}_t(X_t\alpha_t) + \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \bar{\sigma}_t((\beta_{j,t} - \beta_{i,t})X_t), e_i \rangle a_{ji} (e_j - e_i) \right) dt \\ &+ \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t-}^p \bar{\sigma}_t(F_{t-}X_{t-}) dN_{kl,t}^p, \end{aligned} \quad (\text{A.7})$$

637 Now, let us focus on the second term in the no-jump part of the SDE of eq. (A.7). First, because by
638 definition $\alpha_t = \alpha(X_t)$ and $\beta_t = \beta(X_t)$, with $\alpha : \mathbb{X} \rightarrow \mathbb{R}$ and $\beta : \mathbb{X} \rightarrow \mathbb{R}^N$ known function with finite range,
639 we can use the eq.(6) under this functions. Thus, we have for α_t :

$$\begin{aligned} X_t\alpha_t &= \sum_{n=1}^N \alpha(e_n) \langle X_t, e_n \rangle e_n \\ &= \sum_{n=1}^N \text{diag}(\alpha(e_1), \dots, \alpha(e_N)) \langle X_t, e_n \rangle e_n \\ &= \text{diag}(\bar{\alpha}) X_t, \end{aligned}$$

640 where $\bar{\alpha} := (\alpha(e_1), \dots, \alpha(e_N)) \in \mathbb{R}^N$, $\text{diag}(\bar{\alpha})$ is a diagonal matrix with $\bar{\alpha}$ in the diagonal, and because we
641 have used the fact that $X_t \langle X_t, e_n \rangle = \langle X_t, e_n \rangle e_n$ for each $n = 1, \dots, N$, and that $X_t = \sum_{n=1}^N \langle X_t, e_n \rangle e_n$. Thus,
642 applying $\bar{\sigma}_t(\cdot)$ and by linearity, we have

$$\bar{\sigma}_t(X_t\alpha_t) = \text{diag}(\bar{\alpha}) \bar{\sigma}_t(X_t). \quad (\text{A.8})$$

In the same way, using eq.(6) under β_t , and taking $\bar{\sigma}_t(\cdot)$, we have:

$$\begin{aligned} \bar{\sigma}_t((\beta_{j,t} - \beta_{i,t})X_t) &= \sum_{n=1}^N (\beta_{j,t}(e_n) - \beta_{i,t}(e_n)) \langle \bar{\sigma}_t(X_t), e_n \rangle e_n \\ &= \sum_{n=1}^N \langle \beta(e_n), e_j - e_i \rangle \langle \bar{\sigma}_t(X_t), e_n \rangle e_n, \end{aligned}$$

643 but, because in the second term in the no-jump part of the SDE of eq. (A.7), we make inner product with
644 $e_i \in \mathbb{X}$, $i = 1, \dots, N$, it holds $\langle \bar{\sigma}_t((\beta_{j,t} - \beta_{i,t})X_t), e_i \rangle = \langle \beta(e_i), e_j - e_i \rangle \langle \bar{\sigma}_t(X_t), e_i \rangle$. Computing all the sum
645 involved in eq. (A.7), we have that:

$$\begin{aligned} \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \bar{\sigma}_t((\beta_{j,t} - \beta_{i,t})X_t), e_i \rangle a_{ji} (e_j - e_i) &= \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \beta(e_i), e_j - e_i \rangle \langle \bar{\sigma}_t(X_t), e_i \rangle a_{ji} (e_j - e_i) \\ &= \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^\top \bar{\sigma}_t(X_t), \end{aligned} \quad (\text{A.9})$$

646 since $\langle \bar{\sigma}_t(X_t), e_i \rangle = e_i^\top \bar{\sigma}_t(X_t) \in \mathbb{R}$ for any $i = 1, \dots, N$.

647 Using then eq. (A.8) and eq. (A.9) into the SDE in eq. (A.7), the following holds:

$$\begin{aligned}
d\bar{\sigma}_t(F_t X_t) &= \left(\Phi_t \bar{\sigma}_t(F_t X_t) + \left(\text{diag}(\bar{\alpha}) + \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^\top \right) \bar{\sigma}_t(X_t) \right) dt \\
&\quad + \sum_{p=1}^P \sum_{\substack{k,l=1 \\ k \neq l}}^{K_p} \Psi_{kl,t}^p \bar{\sigma}_t(F_t - X_{t-}) dN_{kl,t}^p .
\end{aligned}$$

648 The results follows by considering in Theorem 4.1, $G_t = \bar{\sigma}_t(F_t X_t)$, $K_t = \bar{\sigma}_t(X_t)$, $\Upsilon_t = \Xi_t = \Phi_t$, and for
649 each $p = 1, \dots, P$, $k, l = 1, \dots, K_p$, $\Pi_{kl,t}^p = Q_{kl,t}^p = \Psi_{kl,t}^p$ and $\Lambda_{kl,t}^p = \mathbf{0}_{N \times N}$, and

$$\Gamma_t = \text{diag}(\bar{\alpha}) + \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^\top .$$

650

■

651 **Appendix B. Parameter Values**

652 In this section, we present the matrices used in the Section 5 of our model's numerical results.

653 *Appendix B.1. Simulated Data*

654 *Appendix B.1.1. Matrices for Markov Process Simulation*

655 This section presents the matrices A and C_m^p , $p = 1, \dots, 6$, $m = 0, 1$, used for the simulations of the
 656 Markov states processes X and Y^p , resp. in Section 5.1.

$$A = 10^{-7} \begin{bmatrix} -8.207 & 3.346 & 3.263 & 1.598 \\ 2.821 & -9.286 & 4.657 & 1.808 \\ 3.047 & 3.930 & -7.057 & 0.080 \\ 2.687 & 1.049 & 2.266 & -6.002 \end{bmatrix} \quad (\text{B.1})$$

| C_m^p | $m = 0$ | $m = 1$ | C_m^p | $m = 0$ | $m = 1$ |
|---------|--|--|---------|--|--|
| $p = 1$ | $\begin{bmatrix} -2.55 \times 10^{-6} & 2.55 \times 10^{-6} \\ 1.53 \times 10^2 & -1.53 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -5.10 \times 10^2 & 5.10 \times 10^2 \\ 2.45 \times 10^{-6} & -2.45 \times 10^{-6} \end{bmatrix}$ | $p = 4$ | $\begin{bmatrix} -3.80 \times 10^{-6} & 3.80 \times 10^{-6} \\ 4.76 \times 10^2 & -4.76 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -5.49 \times 10^2 & 5.49 \times 10^2 \\ 1.50 \times 10^{-7} & -1.50 \times 10^{-7} \end{bmatrix}$ |
| $p = 2$ | $\begin{bmatrix} -3.15 \times 10^{-6} & 3.15 \times 10^{-6} \\ 2.62 \times 10^2 & -2.62 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -1.55 \times 10^2 & 1.55 \times 10^2 \\ 3.30 \times 10^{-6} & -3.30 \times 10^{-6} \end{bmatrix}$ | $p = 5$ | $\begin{bmatrix} -3.85 \times 10^{-6} & 3.85 \times 10^{-6} \\ 1.78 \times 10^2 & -1.78 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -1.86 \times 10^2 & 1.86 \times 10^2 \\ 2.00 \times 10^{-6} & -2.00 \times 10^{-6} \end{bmatrix}$ |
| $p = 3$ | $\begin{bmatrix} -3.60 \times 10^{-6} & 3.60 \times 10^{-6} \\ 5.42 \times 10^2 & -5.42 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -8.58 \times 10^2 & 8.58 \times 10^2 \\ 3.15 \times 10^{-6} & -3.15 \times 10^{-6} \end{bmatrix}$ | $p = 6$ | $\begin{bmatrix} -1.60 \times 10^{-6} & 1.60 \times 10^{-6} \\ 7.52 \times 10^2 & -7.52 \times 10^2 \end{bmatrix}$ | $\begin{bmatrix} -5.82 \times 10^2 & 5.82 \times 10^2 \\ 1.65 \times 10^{-6} & -1.65 \times 10^{-6} \end{bmatrix}$ |

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

657

658 *Appendix B.1.2. Empirical Initial Estimation for HMM*

659 In this section, we present the empirical initial estimation of the matrices $\hat{A}^{(0)}$ and $\hat{C}_m^{p(0)}$, $p = 1, \dots, 6$,
 660 $m = 0, 1$, used for the first iteration of the filter estimate by HMM, in the simulated case.

$$\hat{A}^{(0)} = 10^{-7} \begin{bmatrix} -5.107 \times 10^2 & 9.710 \times 10 & 3.480 \times 10^2 & 1.836 \times 10 \\ 5.855 \times 10 & -1.686 \times 10^2 & 1.031 \times 10 & 1.102 \times 10^2 \\ 4.489 \times 10^2 & 5.11 & -3.583 \times 10^2 & 1.000 \times 10^{-3} \\ 3.253 & 6.643 \times 10 & 1.000 \times 10^{-3} & -1.285 \times 10^2 \end{bmatrix} \quad (\text{B.2})$$

| $\hat{C}_m^{p(0)}$ | $m = 0$ | $m = 1$ | $\hat{C}_m^{p(0)}$ | $m = 0$ | $m = 1$ |
|--------------------|--|--|--------------------|--|--|
| $p = 1$ | $\begin{bmatrix} -1.84 \times 10^2 & 2.88 \times 10^5 \\ 1.84 \times 10^2 & -2.88 \times 10^5 \end{bmatrix}$ | $\begin{bmatrix} -3.60 \times 10^{10} & 1.68 \times 10^2 \\ 3.60 \times 10^{10} & -1.68 \times 10^2 \end{bmatrix}$ | $p = 4$ | $\begin{bmatrix} -2.61 \times 10^2 & 2.64 \times 10^{10} \\ 2.61 \times 10^2 & -2.64 \times 10^{10} \end{bmatrix}$ | $\begin{bmatrix} -7.06 \times 10^2 & 3.20 \times 10 \\ 7.06 \times 10^2 & -3.20 \times 10 \end{bmatrix}$ |
| $p = 2$ | $\begin{bmatrix} -1.68 \times 10^2 & 1.00 \times 10^3 \\ 1.68 \times 10^2 & -1.00 \times 10^3 \end{bmatrix}$ | $\begin{bmatrix} -6.36 \times 10^9 & 3.12 \times 10^2 \\ 6.36 \times 10^9 & -3.12 \times 10^2 \end{bmatrix}$ | $p = 5$ | $\begin{bmatrix} -3.72 \times 10^2 & 7.26 \times 10^3 \\ 3.72 \times 10^2 & -7.26 \times 10^3 \end{bmatrix}$ | $\begin{bmatrix} -1.00 \times 10^3 & 1.88 \times 10^2 \\ 1.00 \times 10^3 & -1.88 \times 10^2 \end{bmatrix}$ |
| $p = 3$ | $\begin{bmatrix} -2.59 \times 10^2 & 3.40 \times 10^{10} \\ 2.59 \times 10^2 & -3.40 \times 10^{10} \end{bmatrix}$ | $\begin{bmatrix} -4.29 \times 10^{10} & 2.22 \times 10^2 \\ 4.29 \times 10^{10} & -2.22 \times 10^2 \end{bmatrix}$ | $p = 6$ | $\begin{bmatrix} -1.01 \times 10^2 & 1.00 \times 10^3 \\ 1.01 \times 10^2 & -1.00 \times 10^3 \end{bmatrix}$ | $\begin{bmatrix} -2.74 \times 10^5 & 1.24 \times 10^2 \\ 2.74 \times 10^5 & -1.24 \times 10^2 \end{bmatrix}$ |

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

661

662 **References**

- 663 [1] W. Wang, Z. Lu, Cyber security in the smart grid: Survey and challenges, *Computer networks* 57 (5) (2013) 1344–1371.
- 664 [2] C.-C. Sun, A. Hahn, C.-C. Liu, Cyber security of a power grid: State-of-the-art, *International Journal of Electrical Power*
- 665 *& Energy Systems* 99 (2018) 45–56.
- 666 [3] R. J. Elliott, L. Aggoun, J. B. Moore, *Hidden Markov models: estimation and control*, Vol. 29, Springer Science & Business
- 667 *Media*, 2008.
- 668 [4] Y. Mo, R. Chabukswar, B. Sinopoli, Detecting integrity attacks on scada systems, *IEEE Transactions on Control Systems*
- 669 *Technology* 22 (4) (2013) 1396–1407.
- 670 [5] Y. Mo, B. Sinopoli, Secure control against replay attacks, in: 2009 47th annual Allerton conference on communication,
- 671 *control, and computing (Allerton)*, IEEE, 2009, pp. 911–918.
- 672 [6] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, P. Tabuada, Secure state estimation against sensor attacks in
- 673 *the presence of noise*, *IEEE Transactions on Control of Network Systems* 4 (1) (2016) 49–59.
- 674 [7] S. Amin, A. A. Cárdenas, S. S. Sastry, Safe and secure networked control systems under denial-of-service attacks, in:
- 675 *International Workshop on Hybrid Systems: Computation and Control*, Springer, 2009, pp. 31–45.
- 676 [8] G. K. Befekadu, V. Gupta, P. J. Antsaklis, Risk-sensitive control under markov modulated denial-of-service (dos) attack
- 677 *strategies*, *IEEE Transactions on Automatic Control* 60 (12) (2015) 3299–3304.
- 678 [9] N. Forti, G. Battistelli, L. Chisci, B. Sinopoli, Joint attack detection and secure state estimation of cyber-physical systems,
- 679 *International Journal of Robust and Nonlinear Control* 30 (11) (2020) 4303–4330.
- 680 [10] Y. Mo, E. Garone, A. Casavola, B. Sinopoli, False data injection attacks against state estimation in wireless sensor
- 681 *networks*, in: 49th IEEE Conference on Decision and Control (CDC), IEEE, 2010, pp. 5967–5972.
- 682 [11] Z. Guo, D. Shi, K. H. Johansson, L. Shi, Optimal linear cyber-attack on remote state estimation, *IEEE Transactions on*
- 683 *Control of Network Systems* 4 (1) (2016) 4–13.
- 684 [12] D. Shi, Z. Guo, K. H. Johansson, L. Shi, Causality countermeasures for anomaly detection in cyber-physical systems,
- 685 *IEEE Transactions on Automatic Control* 63 (2) (2017) 386–401.
- 686 [13] D. Shi, R. J. Elliott, T. Chen, On finite-state stochastic modeling and secure estimation of cyber-physical systems, *IEEE*
- 687 *Transactions on Automatic Control* 62 (1) (2016) 65–80.
- 688 [14] D. Shi, T. Chen, M. Darouach, Event-based state estimation of linear dynamic systems with unknown exogenous inputs,
- 689 *Automatica* 69 (2016) 275–288.
- 690 [15] L. Aggoun, R. J. Elliott, Finite-dimensional models for hidden Markov chains, *Advances in applied probability* 27 (1)
- 691 (1995) 146–160.
- 692 [16] R. J. Elliott, W. P. Malcolm, Discrete-time expectation maximization algorithms for Markov-modulated poisson processes,
- 693 *IEEE Transactions on Automatic Control* 53 (1) (2008) 247–256.
- 694 [17] Y. Liu, P. Ning, M. K. Reiter, False data injection attacks against state estimation in electric power grids, *ACM Trans-*
- 695 *actions on Information and System Security (TISSEC)* 14 (1) (2011) 1–33.
- 696 [18] H. Fawzi, P. Tabuada, S. Diggavi, Secure state-estimation for dynamical systems under active adversaries, in: 2011 49th
- 697 *annual allerton conference on communication, control, and computing (allerton)*, IEEE, 2011, pp. 337–344.
- 698 [19] W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, *Journal of the Society*
- 699 *for Industrial and Applied Mathematics, Series A: Control* 2 (3) (1964) 347–369.
- 700 [20] M. Zakai, On the optimal filtering of diffusion processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*
- 701 *11* (3) (1969) 230–243.
- 702 [21] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of*
- 703 *the Royal Statistical Society: Series B (Methodological)* 39 (1) (1977) 1–22.
- 704 [22] G. J. McLachlan, T. Krishnan, *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons, 2007.
- 705 [23] M. W. Korolkiewicz, R. J. Elliott, A hidden markov model of credit quality, *Journal of Economic Dynamics and Control*
- 706 *32* (12) (2008) 3807–3819.
- 707 [24] P. Brémaud, *Point processes and queues: martingale dynamics*, Vol. 50, Springer, 1981.
- 708 [25] M. R. James, V. Krishnamurthy, F. Le Gland, Time discretization of continuous-time filters and smoothers for HMM
- 709 *parameter estimation*, *IEEE Transactions on Information Theory* 42 (2) (1996) 593–605.
- 710 [26] V. Krishnamurthy, J. Evans, Finite-dimensional filters for passive tracking of Markov jump linear systems, *Automatica*
- 711 *34* (6) (1998) 765–770.
- 712 [27] H.-H. Bock, Clustering methods: a history of K-means algorithms, in: *Selected contributions in data analysis and classi-*
- 713 *fication*, Springer, 2007, pp. 161–172.
- 714 [28] L. Aggoun, R. J. Elliott, *Measure theory and filtering: Introduction and applications*, Vol. 15, Cambridge University
- 715 *Press*, 2004.
- 716 [29] K. Sennewald, K. Wälde, “Itô’s lemma” and the bellman equation for poisson processes: An applied view, *Journal of*
- 717 *Economics* 89 (1) (2006) 1–36.
- 718 [30] F. Campillo, F. Le Gland, MLE for partially observed diffusions: direct maximization vs. the EM algorithm (1988).
- 719 [31] A. Dembo, O. Zeitouni, Parameter estimation of partially observed continuous time stochastic processes via the EM
- 720 *algorithm*, *Stochastic Processes and their Applications* 23 (1) (1986) 91–113.
- 721 [32] C. J. Wu, On the convergence properties of the EM algorithm, *The Annals of statistics* (1983) 95–103.
- 722 [33] E. Platen, N. Bruti-Liberati, Numerical solution of stochastic differential equations with jumps in finance, Vol. 64, Springer
- 723 *Science & Business Media*, 2010.
- 724 [34] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, Tech. rep., Stanford (2006).

- 725 [35] S. S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering, *Pattern recognition letters* 25 (11)
726 (2004) 1293–1302.
- 727 [36] J. R. Norris, J. R. Norris, *Markov chains*, no. 2, Cambridge university press, 1998.
- 728 [37] S.-S. Choi, S.-H. Cha, C. C. Tappert, A survey of binary similarity and distance measures, *Journal of systemics, cybernetics*
729 *and informatics* 8 (1) (2010) 43–48.
- 730 [38] J. Pohle, R. Langrock, F. M. van Beest, N. M. Schmidt, Selecting the number of states in hidden markov models: pragmatic
731 solutions illustrated using animal movement, *Journal of Agricultural, Biological and Environmental Statistics* 22 (3) (2017)
732 270–293.
- 733 [39] E. Wong, B. Hajek, *Stochastic processes in engineering systems*, Springer Science & Business Media, 2012.
- 734 [40] C. Van Loan, Computing integrals involving the matrix exponential, *IEEE transactions on automatic control* 23 (3) (1978)
735 395–404.