Data-Driven Modeling of the Temporal Evolution of Breakers’ States in the French Electrical Transmission Grid
Mauricio Gonzalez, Antoine Girard

To cite this version:

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\textsuperscript{a}, Antoine Girard\textsuperscript{b}

\textsuperscript{a}Qivalio, Quivalio Analytics, R&D Team. 11 avenue Delcassé, 75008 Paris, France
\textsuperscript{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.
\[\ast\] This work has been funded by the RTE-CentraleSupélec Chair.

Email: mauricio.gonzalez@outlook.fr (Mauricio Gonzalez), antoine.girard@centralesupelec.fr (Antoine Girard)
“jump” times. We then propose a general multi-observation HMM with parameters fitted (that define the
transition rate matrices of the hidden and observed state processes) based on the standing assumption that
the finite-state spaces of all stochastic processes are known. Our continuous-time finite-state HMM is then
applied to our framework as a running example. Precisely, the HMM drives the grid and consists of a hidden
Markov process (the reference configuration) and multiple observed Markov processes (the breakers’ states).
This approach may be potentially applicable in a system monitoring scheme.

Some works focus on different assumed-types of malicious attacks to design system monitoring on elec-
trical grids. For instance, replay attack is considered in [1] [5] [6], in which an attacker hijack sensors, observe
and record the outputs. Denial-of-Service (DoS) attack models were assumed in [7] [8], in which an optimal
control problem under security constraints is solved. False data injection was considered in [9] [10] [11] [17], in
which some attacker can inject error measurements in the state estimation. For a study considering all these
types of attacks, see, e.g., [12]. However, it is difficult to know a priori the attack type on a system, and the
assumptions made by these works may not be close to reality. In [13] [17] [18], it had been analyzed the case
when an arbitrary error or unknown parameter is additively injected on the state model or measurement
to represent, e.g., malicious attacks. However, in these works, it is common to assume that all the model
parameters are known to estimate the system state. That is not the case in our problem because we need to
estimate the parameters. While our HMM is later intended to be included in online monitoring algorithms
to detect abnormal behaviors (e.g., as in [13]), the current paper mainly focuses on modeling and state and
parameter estimation problems.

We provide an iteratively filter-based Expectation-Maximization (EM) approach [19] [20] [21] [22] to esti-
mate the model parameters and the hidden state (e.g., the reference configuration). This approach aims to
maximize a log-likelihood function over parameter space. While our parameter estimation method is close
as obtained in previously cited papers and [3], we adapt the filter-based approach to handle multi-observed
processes. Each one has its transition rate matrix (i.e., parameters), and no average is considered over them
as in [23]. In detail, we suppose first that all state processes belong in a probability space representing the
“real world”. Then, we use a change of probability measure technique (Girsanov’s Theorem, see, e.g., [13] [24])
to define a new probability measure representing a “fictitious world”. In this new space, filters for estimat-
ing the hidden state and the model’s parameters are easy to obtain. They are linear Stochastic Differential
Equations (SDEs) modulated by counting processes. Then, instead of using a classical Euler-Maruyama
discretization (with small-time step) for all the SDEs obtained (e.g., as in [16] [25] [26]), we present a strong
scheme with no discretization error for numerical purposes. This scheme adapts to all state change times
generated by the temporal evolution of all observed processes.

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

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

Let us assume that the grid has finitely many reference configurations \( h_1, \ldots, h_N \) known \textit{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 \( h_1, \ldots, h_N \), 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 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 \( K^p \), \( p = 1, \ldots, P \).

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:

\[
H := \{h_1, h_2, \ldots, h_N\} \subseteq \prod_{p=1}^{P} \mathcal{M}_p,
\]

where, for each \( p = 1, \ldots, P \),

\[
\mathcal{M}_p := \{m_1^p, m_2^p, \ldots, m_{\mathcal{M}_p}^p\} \subseteq \mathbb{R}^2
\]

represents a finite set, with \( \mathcal{M}_p, R \in \mathbb{N}^* \).

\textbf{Example} (Electrical grid). The sets \( \mathcal{H} \) and \( \mathcal{M}_p \) 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 \( 1 = 2 \), then \( \mathcal{M}_p = \{m_1^p, m_2^p\} \). Thus, if e.g., we assume \( R = 1 \), then \( \mathcal{M}_p \) represents the breaker states “off” and “on”, i.e., we can let \( m_1^p = 0 \) and \( m_2^p = 1 \), resp.; and \( \mathcal{H} \) is therefore a set of binary vectors. On the contrary, if we assume \( R = 2 \), then \( \mathcal{M}_p \) 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.

In order to simplify the modeling of the hidden process, we introduce a transformation over \( H \). Let \( \mathbb{I}_n : \mathcal{H} \rightarrow \{0,1\} \) the indicator function defined for each \( n = 1, \ldots, N \) by \( \mathbb{I}_n(h) = 1 \) if \( n \in h \), and \( \mathbb{I}_n(h) = 0 \) otherwise. Then, the vector \( (\mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_N) \) is a bijection from \( \mathcal{H} \) to the set of unit vectors \( \mathbb{X} := \{e_1, e_2, \ldots, e_N\} \subset \mathbb{R}^N \), where \( e_n \in \mathbb{X} \) denotes a vector in \( \mathbb{R}^N \) with unity in the \( n \)-th position and zero elsewhere, \( n = 1, \ldots, N \). 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

\[
X_t := (\mathbb{I}_1(H_t), \mathbb{I}_2(H_t), \ldots, \mathbb{I}_N(H_t))
\]

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 \( X_t = \sum_{n=1}^{N} (X_t, e_n) e_n \), where \( (\cdot, \cdot) \) 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 \( h_n \in \mathcal{H} \). Mathematically, \( H \) can be computed as follows:

\[
H_t = \sum_{n=1}^{N} (X_t, e_n) h_n
\]
In this way, each state $h_n \in \mathbb{H}$ is identified with an unit vector $e_n \in \mathbb{X}$, $n = 1, \ldots, N$. Thus, instead of inferring the hidden evolution of $H$, we can estimate w.l.o.g. the temporal evolution of the underlying state process $X$.

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

$$\mathbb{Y}^p := \{f^p_1, f^p_2, \ldots, f^p_{K^p}\} \subset \mathbb{R}^{K^p},$$

where $f^p_k \in \mathbb{Y}^p$ denotes the vector in $\mathbb{R}^{K^p}$ with unity in the $k$-th position and zero elsewhere, $k = 1, \ldots, K^p$. Thus, w.l.o.g., the temporal evolution of the observable process $K^p$ can be represented by a state process $Y^p := \{Y^p_t\}_{t \geq 0}$. When $Y^p_t$ is in the state $f^p_k$, that means that the observable process $K^p_t$ is in state $k-1 \in K^p$ at time $t \geq 0$.

**Example** (Electrical grid). If we consider $K^p$ as the observable state process of the breaker $p$ at time $t \geq 0$, then $Y^p$ being in the state $f^p_k$ means that breaker $p$ is in state $k-1 \in K^p$ at time $t \geq 0$. If $K^p = 2$ for each $p = 1, \ldots, P$, then all breakers take binary values (representing the values "off" and "on"), and $f^p_1 = (1, 0)$ and $f^p_2 = (0, 1)$ represent the state values 0 $\in K^p$ and 1 $\in K^p$, resp. In addition, from eq. (2), if we consider $R = 1$, $M^p = K^p = 2$ and $M^p = K^p$, then $M^p$ is also the set of values "off" and "on" and $\mathbb{H}$ is a set of binary vectors.

2.2. The Dynamic of the Hidden Process

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

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

represents the transition probability rate of $X$ from state $e_j \in \mathbb{X}$ to the state $e_i \in \mathbb{X}$, for each $i, j = 1, \ldots, N$, $i \neq j$. In addition, the transpose of $A$ belongs to the $Q$-matrix class\(^3\) Thus, defining $p_{i,t} := \mathbb{P}[X_t = e_i]$

$i = 1, \ldots, N, t \geq 0$, the probability distribution vector $p_t := (p_1,t, p_2,t, \ldots, p_N,t)$ satisfies the forward equation

$$d p_t / d t = A p_t.$$

On the other hand, $X$ is adapted to the (complete) right-continuous increasing family of the natural $\sigma$-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

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

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

$$X_t = X_0 + \int_0^t AX_s ds + V_t.$$

2.3. The Dynamic of the Observed Processes

The observed state process $Y^p = \{Y^p_t\}_{t \geq 0}$ is directly related with $H = \{H_t\}_{t \geq 0}$. First, since $X = \{X_t\}_{t \geq 0}$

takes unit vectors in $\mathbb{X} \subset \mathbb{R}^N$, then we can express any matrix of real-valued functions with finite range, let say $C^p : \mathbb{X} \rightarrow \mathbb{R}^{K^p \times K^p}$ in function of $X_t$ as:

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

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

4
Second, it is assumed that the transition probability rate of $Y^p$ from state $f^p_k \in \mathcal{Y}_p$ to the state $f^p_l \in \mathcal{Y}_p$ also 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 eq. (1) and eq. (2).

Mathematically, let $\text{proj}_p : \mathbb{H} \to M_p$ the $p$-th projection function. We denote by $H^p_t := \text{proj}_p(H_t)$ the $p$-th projection in $H_t$ at time $t \geq 0$, i.e., the value of $H^p_t$ is $\mathbf{m}_m^p \in M_p$ for some $m = 1, ..., M_p$. In line with [13], we can relate each observed state process $Y^p$ with the underlying hidden process $X$ by its transition rate matrix $C^p(X_t) \in \mathbb{R}^{K_p \times K_p}$, where the transpose of $C^p(X_t)$ belongs to the $Q$-matrix class. This matrix can be expressed as the sum of the eq. (6). In our case, we can obtain for each $n = 1, ..., N$ an expression of the matrix $C^p(e_n) = (c^p_{kl}(e_n))$ as a function of the $p$-th position of $h_n \in \mathbb{H}$. Indeed, let $e_n \in \mathbb{X}$. For each $p = 1, ..., P$ and $k, l = 1, ..., K_p$, $k \neq l$, the transition probability rate of $Y^p$ from state $f^p_k \in \mathcal{Y}_p$ to the state $f^p_l \in \mathcal{Y}_p$ can be expressed by:

$$
c^p_{kl}(e_n) = \frac{d}{dt} \mathbb{P}[Y^p_t = f^p_k \mid Y^p_0 = f^p_l, \ X_0 = e_n] \bigg|_{t=0}
$$

(by eq. (4), because $\langle X_0, e_n \rangle = 1$ and $\langle X_0, e_{n'} \rangle = 0$, for each $n' = 1, ..., N, n' \neq n$)

where the last equality holds by the local assumption over the vectors $h_n \in \mathbb{H}$. In such a way, for each $\mathbf{m}_m^p \in M_p$, $m = 1, ..., M_p$, we define the transition rate matrix $C^p_m = (c^p_{kl}(m))$ of $Y^p$ (whose transpose belongs to the $Q$-matrix class), by:

$$
c^p_{kl}(m) := \frac{d}{dt} \mathbb{P}[Y^p_t = f^p_k \mid Y^p_0 = f^p_l, \ H^p_0 = \mathbf{m}_m^p] \bigg|_{t=0},
$$

(7)

for each $k, l = 1, ..., K_p$, $k \neq l$. In addition, based on the eq. (7), we express $C^p(X_t)$ in function of the number of elements in $M_p$, by:

$$
C^p(X_t) = \sum_{m=1}^{M_p} C^p_m \sum_{n \in I^p_m} \langle X_t, e_n \rangle,
$$

(8)

where $C^p_m \in \mathbb{R}^{K_p \times K_p}$ is the matrix with components of the eq. (7), and $I^p_m \subseteq I := \{1, ..., N\}$ is the subset of indices $n = 1, ..., N$ for which the $p$-th position of the vector $h_n \in \mathbb{H}$ is $\mathbf{m}_m^p \in M_p$, i.e.,

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

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

$$
\text{diag}^p_m := \text{diag} \left( I_{\{1 \in I^p_m\}}, I_{\{2 \in I^p_m\}}, ..., I_{\{N \in I^p_m\}} \right),
$$

(9)

where $I_{\{n \in I^p_m\}}$ is the indicator function for sets, i.e., $I_{\{n \in I^p_m\}} = 1$ if $n \in I^p_m$, and $I_{\{n \in I^p_m\}} = 0$ otherwise.

**Example (Electrical grid).** If for the breakers $p = 1, ..., P$, we consider $M_p = 2$, and each $M_p$ represents the set of modes of the breaker $p$, then we look for the index of the reference configurations of the grid $h_n \in \mathbb{H}$ in which the values in their $p$-th position are the modes $\mathbf{m}_m^p$ and $\mathbf{m}_m^p$ (which could represent the values “on”, “off”, resp.), and for each $m = 1, 2$, there is therefore a transition rate matrix $C^p_m$.

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

$$
W^p_t := Y^p_t - Y^p_0 - \int_0^t C^p(X_s)Y^p_s ds,
$$

5
is a \((\mathcal{G}_t, \mathbb{P})\)-Martingale \cite[Lemma 2.2]{15}, where \(\mathcal{G}_t := \sigma(X_s, Y^p_s; s \leq t, p = 1, \ldots, P)\) represents the right-continuous complete filtration generated by \(X\) and all observed processes \(Y^p, p = 1, \ldots, P\). The semi-Martingale representation of \(Y^p_t\) is therefore:

\[
Y^p_t = Y^p_0 + \int_0^t C^p(X_s)Y^p_ds + W^p_t. \tag{10}
\]

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

### 2.4. Summary

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

#### Example (Electrical grid)

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

![Figure 1](image1.png)

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

![Figure 2](image2.png)

**Figure 2:** Temporal evolution of the grid represented by \(H\).
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 $C_{p}^{m} = (c_{p,m}^{k,l})$, for each $p = 1, ..., P$ and $m = 1, ..., M_{p}$, 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 $Y_{t} = \sigma(Y_{p}s; s \leq t, p = 1, ..., P)$ generated by all observed processes $Y_{p}$. This is done through the estimation of $X_{t}$ under $P$ is the expectation operator $E$ over $X_{t}$ given $Y_{t}$. In fact, since $X_{t}$ is defined as an indicator function in eq. (3), the filtered estimate is a conditional probability distribution, i.e.,

$$E[X_{t} | Y_{t}] = (P[X_{t} = e_{1} | Y_{t}], ..., P[X_{t} = e_{N} | Y_{t}])$$

It can be shown that an explicit equation for $E[X_{t} | Y_{t}]$ can be obtained, but it will be nonlinear. In contrast, by using some change of the probability measure $P$, we can obtain filtered estimate that will be is linear, as it will be shown below. To obtain $E[X_{t} | Y_{t}]$ we can use a simple Bayes’ rule.

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

$$N_{kl,t}^{p} = \int_{0}^{t} \langle f_{k}^{p}, Y_{s-}^{p} \rangle \langle f_{l}^{p}, dY_{s}^{p} \rangle = \int_{0}^{t} \langle f_{k}^{p}, Y_{s-}^{p} \rangle \langle f_{l}^{p}, C^{p}(X_{s})Y_{s}^{p} \rangle ds + \int_{0}^{t} \langle f_{k}^{p}, Y_{s-}^{p} \rangle \langle f_{l}^{p}, dW_{s}^{p} \rangle ,$$

where, we have used the eq. (10) in differential form, and $Y_{s}^{p} := \lim_{s \uparrow t} Y_{s}^{p}$ is the left limit of the state process $Y_{t}^{p}$ at $t \geq 0$. Note that each $N_{kl,t}^{p}$ is $Y_{t}$-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:
\[ N_{kl,t}^p = \int_0^t (f_k^p, Y_{s-}^p) \sum_{m=1}^{M_p} (f_m^p, Y_{s-}^p) \sum_{n \in I_m^p} (X_s, e_n) ds + \int_0^t (f_k^p, Y_{s-}^p) (f_m^p, dW_s) \]
\[ = \int_0^t (f_k^p, Y_{s-}^p) \sum_{m=1}^{M_p} \sum_{n \in I_m^p} (X_s, e_n) ds + \int_0^t (f_k^p, Y_{s-}^p) (f_m^p, dW_s) \]
\[ = \int_0^t \lambda_{kl,s}^p ds + M_{kl,t}^p , \]
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 counting process \( N_{kl,t}^p \), defined by:
\[ \lambda_{kl,t}^p := (f_k^p, Y_{s-}^p) \sum_{m=1}^{M_p} \sum_{n \in I_m^p} (X_s, e_n). \] (12)

**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 values. In this context, the process \( N_{kl,t}^{p,12} \) (resp. \( N_{kl,t}^{p,21} \)) counts the number of jumps that \( Y^p \) does from state \( f_1^p = (1, 0) \) (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 representation of the number of jumps of the breaker \( p \) from state “off” to “on” (resp., from “on” to “off”).

From eq. (12), we find back the intuition that a higher transition rate \( c_{21}^p \) (resp. \( c_{12}^p \)) between state “off” to “on” (resp., from “on” to “off”) in the breaker \( p \) (which are entries in the matrix \( C_{in}^p \)), is related to a higher intensity in the counting process \( N_{kl,t}^{p,12} \) (resp. \( N_{kl,t}^{p,21} \)).

The idea is then to introduce a new probability measure \( \mathbb{P} \) for a “fictitious world” from the probability measure \( \mathbb{P} \) of the “real world” to change all intensities to one under \( \mathbb{P} \). This is described by means of the Radon-Nikodym derivative, see, e.g. [24, Ch. VI, Sec.2-3]. By using [24, Ch. VI, eq. (3.3)] but for multidimensional case[4] we define \( \mathbb{P} \) by putting:

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

which is a \((\mathcal{G}_t, \mathbb{P})\)-martingale. Using now Ito’s Lemma, see, e.g., [29], we have:
\[ \Lambda_t = 1 - \sum_{p=1}^{P} \sum_{k,l=1}^{K_p} \int_0^t \Lambda_s (\lambda_{kl,s}^p)^{-1} (\lambda_{kl,s}^p - 1) (dN_{kl,s}^p - \lambda_{kl,s}^p ds). \] (14)

We also define the reverse counterpart of (13) by putting:
\[ \overline{\Lambda}_t := \exp \left\{ \sum_{p=1}^{P} \sum_{k,l=1}^{K_p} \int_0^t \ln (\lambda_{kl,s}^p) dN_{kl,s}^p - \sum_{p=1}^{P} \sum_{k,l=1}^{K_p} \int_0^t (\lambda_{kl,s}^p - 1) ds \right\}, \] (15)
so that \( \overline{\Lambda}_t \Lambda_t = 1 \). Again by Ito’s Lemma, it holds:
\[ \overline{\Lambda}_t = 1 + \sum_{p=1}^{P} \sum_{k,l=1}^{K_p} \int_0^t \overline{\Lambda}_s (\lambda_{kl,s}^p - 1) (dN_{kl,s}^p - \lambda_{kl,s}^p ds) \] (16)

In this way, $\overline{X}_t$ and $(N_{kl,t}^p - t)$ are $(\mathcal{G}_t, \mathbb{F})$-martingale $\forall t \geq 0$. It can be also shown that, under $\mathbb{P}$, the dynamic for $X_t$ is still given by [(24), Ch. II, Theorem T6] and that they have fixed intensity one, see, e.g., [24, Lemma 1], [24, Ch. VI, Lemma L5] and [28, Lemma 4.7.1] resp., mutatis mutandis.

3.2. Filter Estimate for the Grid States

The idea is to use $\overline{X}_t$ to compute the estimator $\sigma_t(X_t) := \mathbb{E}[X_t | \mathcal{Y}_t]$ by means of a version of Bayes’ rule, see, e.g., [16, Lemma 1], [24, Ch. II, Theorem T6] and [28, Lemma 4.7.1] resp., for any $\mathcal{G}_t$-adapted and integrable process $F_t$, the filtered estimate of $F_t$ can be computed via:

$$
\mathbb{E}[F_t | \mathcal{Y}_t] = \frac{\mathbb{E}[\overline{X}_t F_t | \mathcal{Y}_t]}{\mathbb{E}[\overline{X}_t | \mathcal{Y}_t]},
$$

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

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

$$
\overline{\sigma}_t(X_t) = \overline{\sigma}_0(X_0) + \int_0^t \overline{\sigma}_s(X_s) ds - \sum_{p=1}^P \sum_{k,l=1}^{K_p} \int_0^t \overline{\sigma}_s(N_{kl,s}^p - s) d(N_{kl,s}^p)
$$

$$
+ \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^{M_p} \int_0^t f_k^p, Y_{s-}^p c_{kl}^p \overline{\sigma}_s d(N_{kl,s}^p - s),
$$

where $\text{diag}_{m}^p$ is the diagonal matrix defined in eq. [9].

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

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

3.3. Parameter Estimation

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

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

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

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

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

In general, computing directly the MLE is challenging. The Expectation–Maximization (EM) algorithm provides an iterative approximation method starting from an initial estimation $\theta_0$, see Section 4.4. This algorithm is based on the following straightforward application of the well-known Jensen’s inequality:
\[ \mathcal{L}(\theta) - \mathcal{L}(\hat{\theta}) = \ln \left( \mathbb{E}_{\tilde{\theta}} \left[ \frac{d\mathbb{P}_{\tilde{\theta}}}{d\mathbb{P}_{\hat{\theta}}} \mid \mathcal{Y} \right] \right) \geq \mathbb{E}_{\tilde{\theta}} \left[ \ln \left( \frac{d\mathbb{P}_{\tilde{\theta}}}{d\mathbb{P}_{\hat{\theta}}} \mid \mathcal{Y} \right) \right] =: Q(\hat{\theta}, \hat{\theta}). \]

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

1. **E-step**: set \( \hat{\theta} = \hat{\theta}_r \) and compute \( Q(\cdot, \hat{\theta}) \).
2. **M-step**: find \( \hat{\theta}_{r+1} \in \arg\max_{\theta \in \Theta} Q(\hat{\theta}, \theta) \).

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

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

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

\[
\hat{a}_{ji} = \frac{\mathbb{E}[J_{ij,t} \mid \mathcal{Y}_t]}{\mathbb{E}[O_{i,t} \mid \mathcal{Y}_t]}, \quad \hat{c}^n_{kl,m} = \sum_{n \in I^n_m} \frac{\mathbb{E}[L^n_{kl,t} \mid \mathcal{Y}_t]}{\mathbb{E}[S^n_{k,t} \mid \mathcal{Y}_t]},
\]

where, for \( i, j = 1, \ldots, N, i \neq j \),

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

and, for \( p = 1, \ldots, P, k, l = 1, \ldots, K_p, k \neq l \), and \( n = 1, \ldots, N \),

\[
L^n_{kl,t} := \int_0^t \langle e_n, X_{s-} \rangle dN^n_{kl,s}, \quad S^n_{k,t} := \int_0^t \langle f^n_k, Y_s \rangle \langle e_n, X_s \rangle ds.
\]

**Proof.** See [Proof 1] in Appendix A.

In Theorem 3.2 note that \( J_{ij,t} \) represents a counting process of the number of jumps of \( X \) from state \( e_i \in \mathbb{X} \) to state \( e_j \in \mathbb{X} \) within \([0, t]\), \( i \neq j \), \( O_{i,t} \) stands for the occupation time by \( X \) on the state \( e_i \in \mathbb{X} \) within \([0, t]\). \( L^n_{kl,t} \) represents the process that increases only when \( Y^p \) jumps from state \( f^n_k \in \mathbb{Y}_P \) to state \( f^n_k \in \mathbb{Y}_P \) and \( X \) is in state \( e_n \in \mathbb{X} \), \( k \neq l \); and \( S^n_{k,t} \) stands for the total time up to \( t \geq 0 \) for which \( X \) is in state \( e_n \in \mathbb{X} \) and simultaneously \( Y^p \) is in state \( f^n_k \in \mathbb{Y}_P \).

**Example (Electrical grid).** For instance, consider \( R = 1, M_p = K_p = 2 \) and \( M_p = \mathbb{K}_p \). First, \( J_{ij,t} \) counts the number of jumps that the grid does from the reference configuration \( h_i \in \mathbb{H} \) to the reference configuration \( h_j \in \mathbb{H} \) within \([0, t]\). Second, \( O_{i,t} \) is the time that the grid spends on the reference configuration \( h_i \in \mathbb{H} \) up to time \( t \geq 0 \). Third, \( L^n_{kl,t} \) (resp. \( L^n_{21,t} \)) increases only when the breaker \( p \) changes resp. from state “off” to “on” (resp. from state “on” to “off”). And simultaneously the grid is in the reference configuration \( h_n \in \mathbb{H} \) within the time interval \([0, t]\). Fourth, \( S^n_{1,t} \) (resp. \( S^n_{2,t} \)) is the time that the grid spends on the reference 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 “off”).
From eq. (17), the estimation $\hat{a}_{ij}$ and $c_{ik}^{p,m}$ can be obtained via the probability measure $\mathbb{P}$ by:

$$\hat{a}_{ij} = \frac{\langle \tau_i(J_{ij,t}X_t), 1_N \rangle}{\langle \tau_i(O_{ij,t}X_t), 1_N \rangle}, \quad c_{ik}^{p,m} = \sum_{n \in K} \frac{\langle \tau_n(L^{p,m}_{kl,t}X_t), 1_N \rangle}{\langle \tau_n(S^{p,n}_{k,l,t}X_t), 1_N \rangle}.$$  \hfill (22)

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

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

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

(i) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = (e_i, X_t) a_{ji}$, and $\beta(X_t) = (e_i, X_t) \epsilon_j$, we obtain $F_t = J_{ij,t}$,

(ii) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = (e_i, X_t)$, and $\beta(X_t) = 0_N \in \mathbb{R}^N$, we obtain $F_t = O_{ij,t}$,

(iii) $F_0 = 0 \in \mathbb{R}$, $\alpha(X_t) = (f^p_k, X_t^p) (e_i, X_t)$, and $\beta(X_t) = 0_N \in \mathbb{R}^N$, we obtain $F_t = S^{p,n}_{k,t}$.

Therefore, to compute $\tau_i(J_{ij,t}X_t)$, $\tau_i(O_{ij,t}X_t)$, and $\tau_n(S^{p,n}_{k,t}X_t)$, we can compute once $\tau_i(F_tX_t)$ and restrict afterwards to the particular cases of $\alpha(X_t)$ and $\beta(X_t)$. On the other hand, we know that $\tau_i(F_t) = \langle \tau_i(F_tX_t), 1_N \rangle$, so that, we make the inner product between $\tau_i(F_tX_t)$ and $1_N$ to have the estimation for all parameters, see eq. (22). The following Theorem 3.3 gives the linear filter estimate $\tau_i(F_tX_t)$. The filter estimate $\tau_i(L^{p,m}_{kl,t}X_t)$ is given in Theorem 3.6.

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

$$\tau_i(F_tX_t) = \tau_0(F_0X_0) + A \int_0^t \tau_s(F_sX_s)ds - \sum_{p=1}^{K_p} \sum_{k=1}^{N} \sum_{k \neq l} \int_0^t \tau_s(F_{s-}X_{s-})d(N^{p}_{kl,s} - s)$$

$$+ \int_0^t \tau_s(X_s \alpha_s)ds + \sum_{i,j=1}^{N} \int_0^t \langle \tau_s((\beta_{j,s} - \hat{\beta}_{s,s})X_s), e_i \rangle a_{ji}(e_j - e_i)ds$$

$$+ \sum_{p=1}^{K_p} \sum_{k=1}^{M_p} \sum_{k \neq l} \int_0^t \langle f^p_k, Y^p_{k,s} \rangle c^{p,m}_{ik} \tau_s(F_{s-}X_{s-})d(N^{p}_{kl,s} - s),$$  \hfill (24)

where $\alpha_i := \alpha(X_t) \in \mathbb{R}$ and $\beta_i := \beta(X_t) \in \mathbb{R}^N$ are known functions with finite range, $G_t$-adapted and integrable for each $t \geq 0$, $F_t$ is given in eq. (23), and $\tau_i(F_tX_t)$ is the diagonal matrix of eq. (9).

**Proof.** See Proof 2 in Appendix A.

**Remark 3.4.** Note that if we consider $F_t = F_0 = 1$, $\alpha_t = 0 \in \mathbb{R}$, and $\beta_t = 0_N \in \mathbb{R}^N$ in the eq. (23), then the Proposition 3.1 is a particular case of Theorem 3.3.

In this way, the filter estimates for the parameter estimation are given in the next Corollary 3.5 by taking the particular cases of $\alpha_t$ and $\beta_t$ within $F_t$, see eq. (23) and the cases (i), (ii) and (iii).
Corollary 3.5. The finite-dimensional (unnormalized) estimator for $J_{i,j,t}X_t$, $O_{i,t}X_t$ and $S_{k,t}^{p,n}X_t$ are resp. of the form:

\[
\hat{\sigma}_t(J_{i,j,t}X_t) = A \int_0^t \hat{\sigma}_s(J_{i,j,s}X_s)ds + \int_0^t (\hat{\sigma}_s(X_s), e_i) c_{ij}ad_s - \sum_{p=1}^{P} \sum_{l=1}^{K_p} \int_0^t \hat{\sigma}_s(J_{i,j,s-X_s})d(N_{s}^{p,l}-s) \\
+ \sum_{p=1}^{P} \sum_{l=1}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_{s}^{P}, Y_{s}^{P} \rangle c_{s,t}^{p,m} \text{diag}_{m}^{P} \hat{\sigma}_s(J_{i,j,s-X_s})d(N_{s}^{p,l}-s) \\
\]

\[
\hat{\sigma}_t(O_{i,t}X_t) = A \int_0^t \hat{\sigma}_s(O_{i,s}X_s)ds + \int_0^t (\hat{\sigma}_s(X_s), e_i) c_{ij}ad_s - \sum_{p=1}^{P} \sum_{l=1}^{K_p} \int_0^t \hat{\sigma}_s(O_{i,s-X_s})d(N_{s}^{p,l}-s) \\
+ \sum_{p=1}^{P} \sum_{l=1}^{K_p} \sum_{m=1}^{M_p} \int_0^t \langle f_{s}^{P}, Y_{s}^{P} \rangle c_{s,t}^{p,m} \text{diag}_{m}^{P} \hat{\sigma}_s(O_{i,s-X_s})d(N_{s}^{p,l}-s) \\
\]

\[
\hat{\sigma}_t(S_{k,t}^{p,n}X_t) = A \int_0^t \hat{\sigma}_s(S_{k,s}^{p,n}X_s)ds + \int_0^t (\hat{\sigma}_s(X_s), e_i) c_{ij}ad_s - \sum_{q=1}^{Q} \sum_{u,v} \sum_{m=1}^{M_q} \int_0^t \hat{\sigma}_s(S_{k,s}^{p,n-X_s})d(N_{s}^{q,u,v}-s) \\
+ \sum_{q=1}^{Q} \sum_{u,v} \sum_{m=1}^{M_q} \int_0^t \langle f_{s}^{q}, Y_{s}^{q} \rangle c_{s,t}^{u,v} \text{diag}_{m}^{q} \hat{\sigma}_s(S_{k,s}^{p,n-X_s})d(N_{s}^{q,u,v}-s) \\
\]

where $\text{diag}_{m}^{P}$ is the matrix of eq. (9).

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

\[
L_{k,l,t}^{p,n} = \int_0^t (e_n, X_{s-})d(N_{s}^{p,l}-s) + \int_0^t (e_n, X_{s-})ds .
\]

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

\[
\hat{\sigma}_t(L_{k,l,t}^{p,n}X_t) = A \int_0^t \hat{\sigma}_s(L_{k,l,s}^{p,n}X_s)ds - \sum_{q=1}^{Q} \sum_{u,v} \sum_{m=1}^{M_q} \int_0^t \hat{\sigma}_s(L_{k,l,s}^{p,n-X_s})d(N_{s}^{q,u,v}-s) \\
+ \sum_{m=1}^{M_p} \int_0^t \langle f_{s}^{P}, Y_{s}^{P} \rangle c_{s,t}^{p,m} \text{diag}_{m}^{P} (e_n, \hat{\sigma}_s(X_s-))e_n dN_{s}^{p,l} \\
+ \sum_{q=1}^{Q} \sum_{u,v} \sum_{m=1}^{M_q} \int_0^t \langle f_{s}^{q}, Y_{s}^{q} \rangle c_{s,t}^{u,v} \text{diag}_{m}^{q} \hat{\sigma}_s(L_{k,l,s}^{p,n-X_s})d(N_{s}^{q,u,v}-s) \\
\]

where $\text{diag}_{m}^{P}$ is the diagonal matrix defined in eq. (9).

Proof.

See Proof 3 in Appendix A

12
4. Numerical Methods

This section presents a general numerical method for all SDEs involved in our model, i.e., a general scheme involving each SDE of the filter estimates \( \sigma_t(X_t), \sigma_t(J_{j,t}X_t), \sigma_t(O_{i,t}X_t), \sigma_t(S_{k,t}^{p,n}X_t) \) and \( \sigma_t(L_{kl,t}^{p,n}X_t) \), obtained resp. in Proposition 3.1, Corollary 3.5 and Theorem 3.6. We also show how to obtain the initial estimation of all parameters and the set \( H \) when \( M_p = K_p \) for each \( p = 1, \ldots, P \). This is a particular case that applies in the case study considered in this paper of the electrical grid by using available data of all breakers states.

First, we write \( t_0 < t_1 < \ldots < t_{W+1} \) the increasingly ordered instances of all “jump times” of the observed processes. This is shown in Figure 3 and is obtained by superposing all state change times \( \{ \tau_1, \tau_2, \ldots \} \) of the observed processes, \( p = 1, \ldots, P \). We denote by \( t_w, w = 0, \ldots, W \), the instant where at least one observed 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 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.](image)

4.1. Jump-Adapted Scheme for Filters

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

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

\[
d K_t = \Xi_t K_t dt + \sum_{p=1}^{P} \sum_{k \neq l}^{K_p} Q_{kl,t}^{p} N_{kl,t}^{p},
\]

\[
d G_t = (\Upsilon_t G_t + \Gamma_t K_t) dt + \sum_{p=1}^{P} \sum_{k \neq l}^{K_p} \left( \Pi_{kl,t}^{p} G_{kl,t}^{p} + N_{kl,t}^{p} K_{kl,t}^{p} \right) dN_{kl,t}^{p},
\]

where \( K_t, G_t \in \mathbb{R}^N \) are \( G_t \)-adapted and integrable for any \( t \geq 0 \), and \( \Xi_t, Q_{kl,t}^{p}, \Upsilon_t, \Gamma_t, \Pi_{kl,t}^{p}, N_{kl,t}^{p} \in \mathbb{R}^{N \times N} \) are constant matrices between “jumps” for each \( k, l = 1, \ldots, K_p, p = 1, \ldots, P \). They are defined by \( \Xi_t = \Xi_{t_w}, \Upsilon_t = \Upsilon_{t_w}, \Gamma_t = \Gamma_{t_w}, Q_{kl,t}^{p} = Q_{kl,t_{w}}^{p}, \Pi_{kl,t}^{p} = \Pi_{kl,t_{w}}^{p}, N_{kl,t}^{p} = N_{kl,t_{w}}^{p} \), for any \( t \in [t_w, t_{w+1}), w = 0, \ldots, W \).
Theorem 4.1. The jump-adapted exact solution scheme for $K_t$ and $G_t$ with initial conditions $K_0, G_0 \in \mathbb{R}^N$ is written by:

$$K_{t_0} = K_0,$$
$$G_{t_0} = G_0,$$
$$K_{t_w} = \exp \{ \Xi_{t_w} \Delta t_{w+1} \} K_{t_w},$$
$$G_{t_w} = [I_{N} \ 0_{N \times N}] \exp \left\{ \begin{bmatrix} \Upsilon_{t_w} & \Gamma_{t_w} \\ 0_{N \times N} & \Xi_{t_w} \end{bmatrix} \Delta t_{w+1} \right\} \begin{bmatrix} G_{t_w} \end{bmatrix},$$
$$K_{t_{w+1}} = K_{t_w} + \sum_{p=1}^{P} \sum_{k,l=1 \atop k \neq l}^{K_p} Q_{kl,t_{w+1}}^{p} K_{t_{w+1}} \Delta N_{kl,t_{w+1}}^{p},$$
$$G_{t_{w+1}} = G_{t_{w+1}} + \sum_{p=1}^{P} \sum_{k,l=1 \atop k \neq l}^{K_p} \begin{bmatrix} \Pi_{kl,t_{w+1}}^{p} & 0_{N \times N} \\ 0_{N \times N} & 1 \end{bmatrix} G_{t_{w+1}}^{p} + \begin{bmatrix} 0_{N \times N} \\ \Delta N_{kl,t_{w+1}}^{p} \end{bmatrix} K_{t_{w+1}}^{p} \Delta N_{kl,t_{w+1}}^{p},$$

where $\Delta N_{kl,t_{w+1}}^{p} = N_{kl,t_{w+1}+1}^{p} - N_{kl,t_{w+1}}^{p}$ is defined by $\Delta N_{kl,t_{w+1}}^{p} = 1$ if $Y_{t_{w+1}}^{p}$ jumps from state $f_{t}^{p} \in \mathcal{Y}_{p}$ to state $f_{t}^{p} \in \mathcal{Y}_{p}$ 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 “before” the jump at time $t_{w+1}$.

Proof. See Proof 4 in Appendix A.

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

$$\Phi_t := A - \sum_{p=1}^{P} \sum_{k,l=1 \atop k \neq l}^{K_p} \Psi_{kl,t}^{p},$$

$$\Psi_{kl,t} := \sum_{m=1}^{M_p} (r_{kl}^{p} Y_{t}^{p}) c_{lk}^{p,m} \text{diag}_m - I_{N},$$

where $I_{N}$ is the identity matrix in $\mathbb{R}^{N \times N}$, and $\text{diag}_m$ is the diagonal matrix defined in eq. (9). Note that $\Phi_t$ and $\Psi_{kl,t}$ are constant between “jumps”. That is, $\Phi_t = \Phi_{t_w}$ and $\Psi_{kl,t} = \Psi_{kl,t_w}$ for any $t \in [t_w, t_{w+1})$, $w = 0, \ldots, W$.

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

Corollary 4.2. A jump-adapted exact solution scheme for $\mathcal{F}_t(X_t)$ is written by:

$$\mathcal{F}_{t_0}(X_{t_0}) = \mathcal{F}_0(X_0),$$
$$\mathcal{F}_{t_{w+1}}(X_{t_{w+1}}) = \exp \{ \Phi_{t_{w}} \Delta t_{w+1} \} \mathcal{F}_{t_{w}}(X_{t_{w}}),$$
$$\mathcal{F}_{t_{w+1}}(X_{t_{w+1}}) = \mathcal{F}_{t_{w+1}}(X_{t_{w+1}}) + \sum_{p=1}^{P} \sum_{k,l=1 \atop k \neq l}^{K_p} \Psi_{kl,t_{w+1}}^{p} \mathcal{F}_{t_{w+1}}(X_{t_{w+1}}) \Delta N_{kl,t_{w+1}}^{p},$$

Proof. See Proof 4 in Appendix A.
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 state $f_k^p \in \mathbb{Y}_p$ to state $f_k^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, and $\tau_{tw+1}(X_{t_{w+1}}) := \lim_{s \uparrow t_{w+1}} \tau_{tw+1}(X_s)$ is the values “before” the jump at time $t_{w+1}$.

For the other filter estimates, $\tau_t(J_{ij,t}X_t)$, $\tau_t(O_{ij,t}X_t)$, and $\tau_t(S_{k,i}^N X_t)$, we can express the scheme for $\tau_t(F_t X_t)$ from the Theorem 3.3, see the cases (i), (ii) and (iii) after eq. (23). The following scheme holds for the filter estimate $\tau_t(F_t X_t)$.

**Corollary 4.3.** The jump-adapted exact solution scheme for $\tau_t(F_t X_t)$ with initial condition $\tau_0(F_0 X_0) \in \mathbb{R}^N$ is written by:

\[
\tau_{tw+1}(F_{tw+1} X_{tw+1}) = \begin{bmatrix} \Phi_{tw} & \Gamma_{tw} \\ \Phi_{tw} & \Phi_{tw} \end{bmatrix} \Delta t_{w+1} \begin{bmatrix} \Phi_{tw} & \Gamma_{tw} \\ \Phi_{tw} & \Phi_{tw} \end{bmatrix} \tau_{tw}(F_{tw} X_{tw}) + \sum_{p=1}^{K_p} \sum_{k,l=1}^{K_p} \psi_{kl} \Delta N_{kl,t_{w+1}}^p \tau_{tw+1}(F_{tw+1} X_{tw+1})
\]

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 $f_k^p \in \mathbb{Y}_p$ at time $t = t_{w+1}$, $\tau_{tw+1}(F_{tw+1} X_{tw+1}) := \lim_{s \uparrow t_{w+1}} \tau_{tw+1}(F_s X_s)$ is the respective values “before” the jump at time $t_{w+1}$, and

\[
\Gamma_{tw} = \operatorname{diag}(\alpha(e_1), ..., \alpha(e_N)) + \sum_{i,j=1}^{N} \beta(e_i, e_j - e_i) a_{ji} (e_j - e_i) e_i^T
\]

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

**Proof.** See Proof 5 in Appendix A.
where $\Delta t_{w+1} = t_{w+1} - t_w$, $\Delta N_{kl,t_{w+1}} = N_{kl,t_{w+1}} - N_{kl,t_w}$ is defined by $\Delta N_{kl,t_{w+1}} = 1$ if $Y^p_t$ jumps from state $f^p_k \in \mathcal{V}_p$ to state $f^p_k \in \mathcal{V}_p$ at time $t = t_{w+1}$, and $\mathfrak{S}_{t_{w+1}}(L_{u,v,t_{w+1}}^{k,n} X_{t_{w+1}}) := \lim_{s \uparrow t_{w+1}} \mathfrak{S}_{t_{w+1}}(L_{u,v,s}^{k,n} X_s)$ and $\mathfrak{S}_{t_{w+1}}(X_{t_{w+1}}) := \lim_{s \uparrow t_{w+1}} \mathfrak{S}_{t_{w+1}}(X_s)$ are the values “before” the jump at time $t_{w+1}$.

4.2. An Stopping Criteria for the EM Algorithm

Instead of using the strict stopping criteria for the EM algorithm, we define the following stopping test for numerical purposes:

$$
\left\| \sigma_t^{(r)}(X_t) - \sigma_t^{(r-1)}(X_t) \right\| + \left\| \hat{A}^{(r)} - \hat{A}^{(r-1)} \right\| + \sum_{p=1}^{P} \sum_{m=1}^{M} \left\| \hat{C}^{p,(r)}_m - \hat{C}^{p,(r-1)}_m \right\| \leq \varepsilon,
$$

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

4.3. Uncertainty and Final Hidden State

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

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

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 $n' = 1, ..., N$, $n' \neq n$, and by eq. (4), the estimated hidden state is therefore $h_n$ at such time $t \geq 0$.

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

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

be the function that represents how far the highest $\sigma_t(X_t)$ is from the value one. Recall that this estimator is a probability distribution. So, for a fixed $t \geq 0$, if one component of $\sigma_t(X_t)$ is near to one, then selecting the estimated hidden state by using the eq. (32), is an almost-sure choice. Thus, for any time $t \geq 0$, the function $\epsilon_t$ represents the uncertainty on the hidden states. For the Section 5 of numerical results, we also define the mean

$$
\delta_t := \frac{1}{t} \int_0^t \epsilon_s \, ds
$$

to see how far is $\epsilon_t$ of the mean $\delta_t$, in particular at the jump times.

4.4. An Estimation Procedure for the Initial Parameters when $M_p = K_p$

In this section, we consider the particular case when $M_p = K_p$, $R = 1$, $M_p = K_p$ for each $p = 1, ..., P$.

For the EM algorithm, a $\theta_0 \in \Theta$ must be initialized, i.e., we need to choose initial values for the matrices $\hat{A}^{(0)} = (\hat{a}_{ij}(0))$ and $\hat{C}^{p,(0)}_m = (\hat{c}_{kl}^{p,m}(0))$, for each observed process $p = 1, ..., P$ and index $m = 1, ..., M_p$.

At first glance, we can use Theorem 3.2 empirically, i.e., we can discretize all the involved integrals in the

\[\footnotesize\text{Because, e.g., it could take several iterations to have the equality in all the parameters.}\]

16
parameter estimation (those in \eqref{eq:theta_est} and eq. \ref{eq:theta_est}), and use all the information of the state processes to estimate \( \theta_0 \).

First, since we know the change of each observed process, then we know the values of each \( Y^p_t, p = 1, ..., P \), at any time \( t \geq 0 \), since it remains constant between “jumps”. Second, because it is assumed that we know a priori the space \( \mathbb{H} \), then we can compute an empirical estimation of \( H \), and so that, an empirical estimation of the underlying state process \( X_t \) by eq. \ref{eq:H_est} and eq. \ref{eq:H_est}. 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:

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

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

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

\[
\hat{H}_t \in \arg \min_{h_t \in \mathbb{H}} d(Y_t, h_t) .
\]

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. \ref{eq:Y_t}.

Now, with this empirical estimation \( \hat{H}_t \), we can compute the empirical values of \( X_t \) by means of its definition in eq. \ref{eq:process}. 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 \( Y \), \( t_0, ..., t_W \), we discretize all the involved integrals in Theorem \ref{thm:3.2} to estimate empirically all the parameters of our model. For each \( i, j = 1, ..., N, i \neq j \), the initial estimation of \( A = (a_{ij}) \), i.e., the matrix \( \hat{A}^{(0)} = (\hat{a}_{ij}(0)) \), and of \( C_{m}^{p} = (c_{kl}^{p,m}) \), i.e, the matrix \( \hat{C}_{m}^{p} = (\hat{c}_{kl}^{p,m}(0)) \), for each \( p = 1, ..., P \) and \( m = 1, ..., M_p \), are obtained as follows:

\[
\hat{a}_{ij}(0) := \frac{\sum_{w=0}^{W-1} \langle e_i, \hat{X}_{t_w} \rangle \langle e_j, \hat{X}_{t_{w+1}} \rangle}{\sum_{w=0}^{W-1} \langle e_i, \hat{X}_{t_w} \rangle \Delta t_{w+1}} , \quad \hat{c}_{kl}^{p,m}(0) := \frac{\sum_{w=0}^{W-1} \langle f^p_k, Y^p_{t_w} \rangle (f^p_l, Y^p_{t_{w+1}}) \sum_{n \in I^p_m} \langle e_n, \hat{X}_{t_w} \rangle}{\sum_{w=0}^{W-1} \langle f^p_k, Y^p_{t_w} \rangle \sum_{n \in I^p_m} \langle e_n, \hat{X}_{t_w} \rangle \Delta t_{w+1}} .
\]

On the other hand, the initial estimation \( \sigma_0(X_0) \) of the process \( X \) at time \( t = 0 \) that we need in eq. \ref{eq:sigma_est}, is given by the empirical estimation \( \hat{\sigma}_0(\hat{X}_0) \) defined for each \( n = 1, ..., N \), by:

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

\[\text{4.5. A Method to Find the Reference Configurations of the Grid}\]

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. \ref{eq:Y_t}. Here, we also consider the particular case when \( M_p = \mathbb{K}_p, R = 1, M_p = K_p \), for each \( p = 1, ..., P \).
To obtain the reference configurations of the grid, we construct clusters from data by partitioning it into \( N \) subsets. Each subset \( U_1, ..., U_N \), called cluster, is represented by its representative state \( \mu_1, ..., \mu_N \), resp. To obtain optimal clusters, we use \( K \)-means method, see, e.g., [35]. The extension of \( K \)-means in continuous time is given by the minimization of the following cost function in the horizon time \( T > 0 \):

\[
J = \sum_{n=1}^{N} \int_0^T I_{\{Y_t \in U_n\}} d(Y_t, \mu_n) dt,
\]

where \( I_{\{Y_t \in U_n\}} \) is the indicator function for sets, i.e., \( I_{\{Y_t \in U_n\}} = 1 \) if \( Y_t \in U_n \), and \( I_{\{Y_t \in U_n\}} = 0 \) otherwise.

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

\[
J = \sum_{n=1}^{N} \sum_{w=0}^{W-1} I_{\{Y_{tw} \in U_n\}} d(Y_{tw}, \mu_n) \Delta t_{w+1},
\]

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

\[
\mu_n \in \arg \min_{\eta \in U_n} \left\{ \sum_{w=0}^{W-1} I_{\{Y_{tw} \in U_n\}} d(Y_{tw}, \eta) \right\}.
\]

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

\[
U_n = \bigcup_{u=0}^{W-1} \left\{ Y_{tw} \in \bigcap_{p=1}^{P} \mathbb{K}_p : d(Y_{tw}, \mu_n) \leq d(Y_{tw}, \mu_{n'}) \text{, for each } n' = 1, ..., N \right\}.
\]

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

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

5. Numerical Results

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

5.1. Simulated Data

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

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

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

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

Concerning the parameters of our model, we compute the initial estimation of the matrices $\hat{A}^{(0)}$ and $\hat{C}_{p(0)}$, $p = 1, \ldots, 6$, $m = 1, 2$, by the empirical estimation of eq. (37). These values are shown in eq. (B.2) and Table B.2 in Appendix B.1.2 resp. The initial filter estimate of $X$ is obtained from eq. (38). Under the simulated data, we obtain $\pi_0(X_0) = (0.307, 0.196, 0.388, 0.109)$. The initial state for $X$ is therefore chosen to be $\hat{X}_0 = e_3$ by eq. (32). Thus, by eq. (41), the initial hidden state is $h_3 = (1,0,1,1,1,1)$. Finally, for the stopping criteria of eq. (31), we choose $\varepsilon = 10^{-5}$.

5.1.2. Hidden State Estimation by Clustering Method

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

Figure 4: Simulated temporal evolution of the observable processes.

Figure 5: Simulated temporal evolution of the observable processes.
Under the clustering method with the euclidean distance, the arg min set might not be a single point several times because there are points of the joint information of all observable processes at the same distance of different vectors of $\mathbb{H}$. In such cases, the clustering method is not exact because we cannot know which state the hidden process is. This is represented with green points on the left side in Figure 5. When a cluster 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.

5.1.3. Hidden State Estimation by HMM

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

5.1.4. Parameter Estimation

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

5.1.5. Uncertainty and Final Hidden State

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

(a) Estimated $A = (a_{ij})$
(b) Estimated $C_{p1} = (c_{p1})$
(c) Estimated $C_{p2} = (c_{p2})$

Figure 7: Estimated values for the matrices $A$, $C_{p1}$, and $C_{p2}$, resp., for each $p = 1, ..., 6$, over the number of iterations.

Figure 8: Temporal evolution’s uncertainty signal on the states of $H$ at the first iteration (on the left side) and on the last iteration 14 (on the right side).
5.2. Real Data

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

5.2.1. Initial Parameters

We take from the data an electrical transmission grid with $P = 6$ breakers, with states off/on for each 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.](image)

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

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

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

5.2.2. Grid State Estimation by Clustering Method

In Figure 10, we observe on the left side the temporal evolution of the empirical estimation of the grid by clustering method, see eq. (36). At any time, all breakers states’ joint information is classified with the minimum distance to the cluster centers, see eq. (39). Here, the cluster centers are the reference configurations of eq. (41). However, the $\arg\min$ set might not be a single point because of the nature of the dataset. In such cases, the clustering method is not exact because we cannot know at which configuration
the grid is. The green dots show all the points of the arg min set. Because of the several choices, we represent
the first choice of the arg min set by the blue line in the same picture. Also, we show on the right side in
Figure 10 when a cluster is active, i.e., when the reference configurations can be chosen by clustering. The
red dots represent when there is more than one choice. The metric distance used is the euclidean distance.
This is a crucial aspect in our estimation since the first estimation of the matrices \( \hat{A}^{(0)} \) and \( \hat{C}_p^{(0)} \),
\( p = 1, \ldots, 6, m = 1, 2 \), may be affected by this issue. So, by taking the first point of the arg min set at each time, we
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
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).](image)

5.2.3. Grid State Estimation by HMM

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

5.2.4. Uncertainty and Final Hidden State of the Grid

Now, we focus on the uncertainty in the grid by using \( \epsilon_t \), see eq. (33). We show in Figure 12 the
values of the uncertainty signal \( \epsilon_t \) that represents a temporal evolution’s uncertainty signal on the reference
configurations. We notice that there are fewer peaks in \( \epsilon_t \) at the last iteration compared to the first iteration.
This is because finding the optimal parameters to fit our model, the uncertainty decreases as the number
of iterations increases. Recall that when \( \epsilon_t \) is over \( \delta_t \), the state at which the grid “jumps” is known with
some uncertainty. However, such uncertainty is almost instantaneous because it does not remain in time, as
reflected in Figure 12.
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].
Using our theoretical results, we have then shown the performance of the framework by considering Boolean temporal sequences describing the breakers’ states (off/on) of the grid. First, we have evaluated a simulated scenario, showing the advantages of the HMM approach with the proposed strong scheme. Second, we have confronted our model with real data provided by the France’s transmission system operator (RTE), showing stability in the filter estimates for the parameters and the hidden state of the grid. We finally identify the normal behavior of the French electrical grid, which could be embedded in monitoring and detection algorithm in the future.
Appendix A. Mathematical Proofs

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

$$J_{ij,t} = \int_{0}^{t} (e_{i}, X_{s-}) (e_{j}, dX_{s}) = (e_{i}, X_{s-}) a_{ij} dt + Q_{ij,t},$$

from eq. (5), where $Q_{ij,t} := J_{ij,t} - \int_{0}^{t} (e_{i}, X_{s-}) a_{ij} ds$ is a $(\mathcal{G}_{t}, \mathbb{P})$-Martingale [15].

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

$$\left. \frac{d\mathbb{P}}{d\mathbb{P}_{\tilde{\theta}}} \right|_{\mathcal{G}_{t}} := \tilde{\Lambda}_{t} := \exp \left\{ - \sum_{p=1}^{P} \sum_{k,l=1}^{K_{p}} \int_{0}^{t} \log \left( \frac{\lambda^{p}_{kl,s}}{\lambda^{\tilde{\theta}}_{kl,s}} \right) dN^{p}_{kl,s} - \sum_{i,j=1}^{N} \int_{0}^{t} \log \left( \frac{a_{ij}}{\hat{a}_{ij}} \right) dJ_{ij,s} \right. $$

$$+ \sum_{p=1}^{P} \sum_{k,l=1}^{K_{p}} \int_{0}^{t} (\lambda^{p}_{kl,s} - \tilde{\lambda}_{kl,s}^{p}) ds + \sum_{i,j=1}^{N} \int_{0}^{t} (X_{s-}, e_{i}) (a_{ij} - \hat{a}_{ij}) ds \right\} ,$$

(A.1)

where $\tilde{\lambda}_{kl,t}^{p} = \langle f_{k}^{p}, Y_{t-}^{p} \rangle \sum_{m=1}^{M_{p}} \hat{c}_{k_{l},m}^{p} \sum_{n \in I_{0}^{m}} \langle X_{t}, e_{n} \rangle$. The log-likelihood is, therefore:

$$\ln (\tilde{\Lambda}_{t}) = - \sum_{p=1}^{P} \sum_{k,l=1}^{K_{p}} \int_{0}^{t} \log \left( \frac{\lambda^{p}_{kl,s}}{\lambda^{\tilde{\theta}}_{kl,s}} \right) dN^{p}_{kl,s} - \sum_{i,j=1}^{N} \int_{0}^{t} \log \left( \frac{a_{ij}}{\hat{a}_{ij}} \right) dJ_{ij,s} \right. $$

$$+ \sum_{p=1}^{P} \sum_{k,l=1}^{K_{p}} \int_{0}^{t} (\lambda^{p}_{kl,s} - \tilde{\lambda}_{kl,s}^{p}) ds + \sum_{i,j=1}^{N} (a_{ij} - \hat{a}_{ij}) O_{i,t} ,$$

(A.2)

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

$$\lambda^{p}_{kl,t} - \tilde{\lambda}_{kl,t}^{p} = \langle f_{k}^{p}, Y_{t}^{p} \rangle \sum_{m=1}^{M_{p}} (c_{k_{l},m}^{p} - \hat{c}_{k_{l},m}^{p}) \sum_{n \in I_{0}^{m}} \langle X_{t}, e_{n} \rangle ,$$

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

$$\ln \left( \frac{\lambda^{p}_{kl,t}}{\lambda^{\tilde{\theta}}_{kl,t}} \right) = \ln \left( \sum_{m=1}^{M_{p}} c_{k_{l},m}^{p} \sum_{n \in I_{0}^{m}} \langle X_{t}, e_{n} \rangle \right) = \sum_{m=1}^{M_{p}} \ln \left( \frac{c_{k_{l},m}^{p}}{\hat{c}_{k_{l},m}^{p}} \right) \sum_{n \in I_{0}^{m}} \langle X_{t}, e_{n} \rangle ,$$

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

\[ \text{[See, e.g., [2] Ch. VI, Eq. (2.17)] or [25] Lemma 4.7.3].}
Thus, we have \(^\hat{a}_{ji}\) and \(\hat{c}^{pm}_{lk}\) occur at such values obtained by equaling to zero the partial derivatives:

\[
\frac{\partial}{\partial \hat{a}_{ji}} \mathbb{E}\left[ \ln \left( \frac{d\hat{p}}{dp} \right) \mid \mathcal{Y}_t \right] = 0 \quad \text{and} \quad \frac{\partial}{\partial \hat{c}^{pm}_{lk}} \mathbb{E}\left[ \ln \left( \frac{d\hat{p}}{dp} \right) \mid \mathcal{Y}_t \right] = 0.
\]

Thus, we have \(\hat{a}_{ji} = \frac{\mathbb{E}[L_{ji,t} \mid \mathcal{Y}_t]}{\mathbb{E}[O_{ji,t} \mid \mathcal{Y}_t]}\) and \(\hat{c}^{pm}_{lk} = \frac{\mathbb{E}[S_{lk,t}^{pm} \mid \mathcal{Y}_t]}{\mathbb{E}[O_{ji,t} \mid \mathcal{Y}_t]}\).

**Proof 2 (Theorem 3.3).** We proceed as in [15, Theorem 3.2] but in our context. To derive an equation for \(\sigma_t(F_tX_t)\), we use first Ito’s product rule on \(F_tX_t\). Let \(\Delta X_s := X_s - X_{s-}\). The following holds:

\[
F_tX_t = F_0X_0 + \int_0^t X_s - dF_s + \int_0^t F_s - dX_s + [F, X]_t
\]

(by using eqs. [4] and [23], and since \([F, X]_t = \sum_{0<s\leq t} \Delta F_s \Delta X_s = \sum_{0<s\leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s\))

\[
= F_0X_0 + \int_0^t X_s - \alpha d\xi_s + \int_0^t X_s - \langle \beta_s, dV_s \rangle + A \int_0^t F_s - dX_s + \int_0^t F_s - dV_s + \sum_{0<s\leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s
\]

\[
= F_0X_0 + \int_0^t X_s - \alpha d\xi_s + \int_0^t X_s - \langle \beta_s, dV_s \rangle + A \int_0^t F_s - dX_s + \int_0^t F_s - dV_s + \sum_{0<s\leq t} \langle \beta_s, \Delta X_s \rangle \Delta X_s \quad \text{(A.3)}
\]

Taking now Ito’s product rule on \(X_tF_tX_t\), we have that:

\[
\ln (\hat{A}_t) = - \sum_{p=1}^P \sum_{k, l = 1}^K \sum_{m, n = 1}^M \ln \left( \frac{c^{pm}_{lk}}{c^{pm}_{lk}} \right) \sum_{n \in I_{lk}} L^{pm}_{kl,t} - \sum_{i, j = 1}^N \ln \left( \frac{a_{ji}}{\hat{a}_{ji}} \right) J_{ij,t}
\]

+ \sum_{p=1}^P \sum_{k, l = 1}^K \sum_{m, n = 1}^M (\hat{c}^{pm}_{lk} - c^{pm}_{lk}) \sum_{n \in I_{lk}} S^{pm}_{lk,t} + \sum_{i, j = 1}^N (a_{ji} - \hat{a}_{ji}) O_{ij,t}.
\]
\[\bar{X}_t F_t X_t = \bar{X}_0 F_0 X_0 + \int_0^t F_s - X_s - d\bar{X}_s + \int_0^t \bar{X}_s - dF_s X_s + [\bar{X}, F X]_t\]

\[= F_0 X_0 + \sum_{p=1}^P \sum_{k,l=1}^{K_p} \int_0^t F_{k,l} X_{k,l} - X_s - \sum_{m=1}^M \sum_{n=1}^N \left( \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^M \sum_{n=1}^N \bar{X}_s - dN_{k,l,s} - s \right) \]

\[+ \int_0^t \bar{X}_s - \alpha_s ds + \int_0^t \bar{X}_s - \langle \beta_s, dV_s \rangle + \int_0^t \bar{X}_s - F_s - X_s ds + \int_0^t \bar{X}_s - dV_s \]

\[+ \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{j} d(e_j - e_i) ds + \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle e_j ds + \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle (e_j - e_i) ds \]

(by using eq. [12] and because \(X_t, e_n = (X_t, e_n), \forall t \geq 0\))

\[= F_0 X_0 - \sum_{p=1}^P \sum_{k,l=1}^{K_p} \int_0^t \bar{X}_s - F_s - X_s - dN_{k,l,s} - s \]

\[+ \sum_{p=1}^P \sum_{k,l=1}^{K_p} \int_0^t \langle f_k, Y_{s-}^p \rangle \bar{X}_s - F_s - \left( \sum_{m=1}^M \sum_{n=1}^N \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^M \sum_{n=1}^N \bar{X}_s - dN_{k,l,s} - s \right) \]

\[+ \int_0^t \bar{X}_s - \alpha_s ds + \int_0^t \bar{X}_s - \langle \beta_s, dV_s \rangle + \int_0^t \bar{X}_s - F_s - X_s ds + \int_0^t \bar{X}_s - dV_s \]

\[+ \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle a_{j} d(e_j - e_i) ds + \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle e_j ds + \sum_{i,j=1}^N \int_0^t \langle \bar{X}_s - (\beta_{j,s} - \beta_{i,s}) X_s, e_i \rangle (e_j - e_i) ds \]

(by using the definition of the diagonal matrix \(\text{diag}^p\) see eq. [9])

\[= F_0 X_0 - \sum_{p=1}^P \sum_{k,l=1}^{K_p} \int_0^t \bar{X}_s - F_s - X_s - dN_{k,l,s} - s + \int_0^t \bar{X}_s - \langle \beta_s, dV_s \rangle + A \int_0^t \bar{X}_s - F_s - X_s ds \]

\[+ \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^M \int_0^t \langle f_k, Y_{s-}^p \rangle c_{ik} \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^M \sum_{n=1}^N \bar{X}_s - dN_{k,l,s} - s \]

\[+ \int_0^t \bar{X}_s - \alpha_s ds + \int_0^t \bar{X}_s - F_s - dV_s \]
\[ + \sum_{i,j=1\atop i\neq j}^{N} \int_{0}^{t} \langle \Lambda_{s} - (\beta_{j,s} - \beta_{i,s})X_{s}, e_{i} \rangle a_{ji}(e_{j} - e_{i}) \, ds + \sum_{i,j=1}^{N} \int_{0}^{t} \langle \Lambda_{s} - (\beta_{j,s} - \beta_{i,s})X_{s}, e_{i} \rangle (e_{j}, dV_{s})(e_{j} - e_{i}) \, , \]

where the last equality holds because \( X_{t} = \sum_{n=0}^{N} \langle X_{t}, e_{n} \rangle e_{n}, \forall t \geq 0 \). Taking now \( \mathbb{E}[\cdot | \mathcal{Y}] \) above and interchanging expectation and integration \([39], \) we get the result.

**Proof 3** (Theorem 3.6). We proceed as in Theorem 3.3. To derive an equation for \( \widehat{\pi}_t(L^{p,n}_{kl,t} X_t) \), we use first Ito’s product rule on \( L^{p,n}_{kl,t} X_t \). The following holds:

\[
L^{p,n}_{kl,t} X_{t} = L^{p,n}_{kl,0} X_{0} + \int_{0}^{t} X_{s} dL^{p,n}_{kl,s} + \int_{0}^{t} L^{p,n}_{kl,s} dX_{s} + [L^{p,n}_{kl}, X]_{t} \quad (A.4)
\]

where we have used the eqs. \( 6 \) and \( 25 \), and since there is no jump at the same time a.s. \( [L^{p,n}_{kl}, X]_{t} = 0 \).

Taking now Ito’s product rule on \( X_{t} F_{t} X_{t} \), and considering the eq. \( 10 \) and the eq. \( (A.4) \), we have:

\[
\mathbb{E} \, L^{p,n}_{kl,t} X_{t} = \int_{0}^{t} L^{p,n}_{kl,s} X_{s} d\Lambda_{s} + \int_{0}^{t} L^{p,n}_{kl,s} dX_{s} + \mathbb{E} [ L^{p,n}_{kl} X]_{t} \\
= \sum_{p=1}^{P} \sum_{q=1}^{K_{q}} \int_{0}^{t} \Lambda_{s} - X_{s} - \lambda_{q,s} - d(N_{q,s}^{p} - s) - \sum_{p=1}^{P} \sum_{q=1}^{K_{q}} \int_{0}^{t} \langle \Lambda_{s} - X_{s} - \lambda_{q,s} \rangle d(N_{q,s}^{p} - s) \\
+ A \int_{0}^{t} \Lambda_{s} - X_{s} - dV_{s} + \int_{0}^{t} \Lambda_{s} - (e_{n}, X_{s}) X_{s} - \lambda_{q,s} - dN_{q,s}^{p} \\
\text{(since there are common jumps,} \quad \mathbb{E} [ L^{p,n}_{kl} X]_{t} = \int_{0}^{t} \Lambda_{s} - \lambda_{q,s} - 1 (e_{n}, X_{s}) X_{s} - dN_{q,s}^{p} \text{)}
\]

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

**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 \( t \) not be jump time, let say \( t \in (t_{w}, t_{w+1}), w = 0, ..., W \), we consider

\[
dG_{t} = (\Gamma_{t} G_{t} + \Gamma_{t} K_{t}) dt \\
dK_{t} = \Xi_{t} K_{t} dt,
\]

29
which is an ordinary linear differential equation system. Recall $\Upsilon_t, \Gamma_t, \Xi_t \in \mathbb{R}^{N \times N}$ are constant matrices between jumps. Now, let

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

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

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

where $0_{N \times N}$ 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 the latter differential equation with initial condition at $t_w$ is:

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

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

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

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

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

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

$$= \exp \{ \Xi_{t_w} \Delta t_{w+1} \} K_{t_w},$$

where we have used the fact that the matrix within the exponential is block triangular, see for instance [10].

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

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

$$d \tilde{\sigma}_t(F_t X_t) = A \tilde{\sigma}_t(F_t X_t)dt + \sum_{p=1}^P K_p \sum_{k,l=1}^{K_p} \tilde{\sigma}_t(F_t X_t)dt - \sum_{p=1}^P K_p \sum_{k,l=1}^{K_p} \tilde{\sigma}_t(F_t X_t) \frac{dN_p}{dt} + \tilde{\sigma}_t(X_t \alpha_t)dt$$

$$+ \sum_{i,j=1}^N \langle \tilde{\sigma}_t((\beta_{i,t} - \beta_{j,t})X_t), e_i a_{ji} (e_j - e_i) \rangle dt - \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^{M_p} \frac{f^p_{kl}}{\big| e_k \big|} \bigg| Y^p_{l,t} \bigg| c^p_{m} \bigg|_{ik} \text{diag}_m \tilde{\sigma}_t(F_t X_t)dt$$

$$+ \sum_{p=1}^P \sum_{k,l=1}^{K_p} \sum_{m=1}^{M_p} \frac{f^p_{kl}}{\big| e_k \big|} \bigg| Y^p_{l,t} \bigg| c^p_{m} \bigg|_{ik} \text{diag}_m \tilde{\sigma}_t(F_t X_t) \frac{dN_p}{dt},$$
wherein the integral w.r.t. Lebesgue measure $dt$, we are allowed to evaluate $\sigma_t(F_t - X_{t-})$ as $\sigma_t(F_t X_t)$. Regrouping the terms and using the definition of $\Phi_t$ and $\Psi_t$ in eq. (29) and eq. (30) resp., the following SDE holds:

$$
\begin{align*}
\sigma_t(F_t X_t) &= \Phi_t \sigma_t(F_t X_t)dt + \left( \sigma_t(X_t \alpha_t) + \sum_{i,j=1}^{N} \sigma_t((\beta_{j,t} - \beta_{i,t}) X_t) e_i \right) a_{ji} (e_j - e_i) dt \\
&\quad + \sum_{p=1}^{P} \sum_{k,l=1}^{K_p} \Psi_{kl,t}^{P} \sigma_t(F_t - X_{t-})dN_{kl,t}^{P},
\end{align*}
$$

(A.7)

Now, let us focus on the second term in the no-jump part of the SDE of eq. (A.7). First, because by 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, we can use the eq.(6) under this functions. Thus, we have for $\alpha_t$:

$$
X_t \alpha_t = \sum_{n=1}^{N} \alpha(e_n)(X_t, e_n)e_n
= \sum_{n=1}^{N} \text{diag}(\alpha(e_1), ..., \alpha(e_N))(X_t, e_n)e_n
= \text{diag}(\overline{\alpha})X_t,
$$

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

$$
\sigma_t(X_t \alpha_t) = \text{diag}(\overline{\alpha}) \sigma_t(X_t).
$$

(A.8)

In the same way, using eq.(6) under $\beta_t$, and taking $\sigma_t(\cdot)$, we have:

$$
\begin{align*}
\sigma_t((\beta_{j,t} - \beta_{i,t}) X_t) &= \sum_{n=1}^{N} (\beta_{j,t}(e_n) - \beta_{i,t}(e_n)) \sigma_t(X_t, e_n)e_n \\
&= \sum_{n=1}^{N} (\beta(e_n), e_j - e_i) \sigma_t(X_t, e_n)e_n,
\end{align*}
$$

but, because in the second term in the no-jump part of the SDE of eq. (A.7), we make inner product with $e_i \in \mathbb{X}, i = 1, ..., N$, it holds $\langle \sigma_t((\beta_{j,t} - \beta_{i,t}) X_t), e_i \rangle = (\beta(e_i), e_j - e_i) \sigma_t(X_t, e_i)$.

Computing all the sum involved in eq. (A.7), we have that:

$$
\sum_{\substack{i,j=1 \ i \neq j}}^{N} \sigma_t((\beta_{j,t} - \beta_{i,t}) X_t, e_i) a_{ji} (e_j - e_i) = \sum_{\substack{i,j=1 \ i \neq j}}^{N} (\beta(e_i), e_j - e_i) \sigma_t(X_t, e_i) a_{ji} (e_j - e_i)
= \sum_{\substack{i,j=1 \ i \neq j}}^{N} (\beta(e_i), e_j - e_i) a_{ji} (e_j - e_i) e_i^T \sigma_t(X_t),
$$

(A.9)

since $\langle \sigma_t(X_t), e_i \rangle = e_i^T \sigma_t(X_t) \in \mathbb{R}$ for any $i = 1, ..., N$.

Using then eq. (A.8) and eq. (A.9) into the SDE in eq. (A.7), the following holds:
\[
\sigma_t(F_t X_t) = \left( \Phi_t \sigma_t(F_t X_t) + \text{diag}(\sigma_t) + \sum_{i,j=1}^{N} \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^T \right) \sigma_t(X_t) dt \\
+ \sum_{p=1}^{P} K_p \sum_{k,l=1}^{K_p} \Psi_{kl,t} \sigma_t(F_t - X_t) dN_{kl,t}.
\]

The results follows by considering in Theorem 4.1, \( G_t = \sigma_t(F_t X_t) \), \( K_t = \sigma_t(X_t) \), \( \Upsilon_t = \Xi_t = \Phi_t \), and for each \( p = 1, ..., P \), \( k, l = 1, ..., K_p \), \( \Pi_{kl,t}^p = \Psi_{kl,t}^p \) and \( A_{kl,t}^p = 0_{N \times N} \), and

\[
\Gamma_t = \text{diag}(\alpha) + \sum_{i,j=1}^{N} \langle \beta(e_i), e_j - e_i \rangle a_{ji} (e_j - e_i) e_i^T.
\]

\[\blacksquare\]
Appendix B. Parameter Values

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

Appendix B.1. Simulated Data

Appendix B.1.1. Matrices for Markov Process Simulation

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

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

<table>
<thead>
<tr>
<th>$C_m^p$</th>
<th>$m = 0$</th>
<th>$m = 1$</th>
<th>$C_m^p$</th>
<th>$m = 0$</th>
<th>$m = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>$[-2.55 \times 10^{-6} \quad 2.55 \times 10^{-6}]$</td>
<td>$[1.53 \times 10^3 \quad -1.53 \times 10^3]$</td>
<td>$[-5.10 \times 10^2 \quad 5.10 \times 10^2]$</td>
<td>$[2.45 \times 10^{-6} \quad -2.45 \times 10^{-6}]$</td>
<td>$[1.53 \times 10^3 \quad -1.53 \times 10^3]$</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>$[-3.15 \times 10^{-6} \quad 3.15 \times 10^{-6}]$</td>
<td>$[2.62 \times 10^2 \quad -2.62 \times 10^2]$</td>
<td>$[-1.55 \times 10^2 \quad 1.55 \times 10^2]$</td>
<td>$[3.30 \times 10^{-6} \quad -3.30 \times 10^{-6}]$</td>
<td>$[2.62 \times 10^2 \quad -2.62 \times 10^2]$</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>$[-3.60 \times 10^{-6} \quad 3.60 \times 10^{-6}]$</td>
<td>$[5.42 \times 10^2 \quad -5.42 \times 10^2]$</td>
<td>$[-8.58 \times 10^2 \quad 8.58 \times 10^2]$</td>
<td>$[3.15 \times 10^{-6} \quad -3.15 \times 10^{-6}]$</td>
<td>$[5.42 \times 10^2 \quad -5.42 \times 10^2]$</td>
</tr>
</tbody>
</table>

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

Appendix B.1.2. Empirical Initial Estimation for HMM

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

$$\hat{A}^{(0)} = 10^{-7} \begin{bmatrix} -5.107 \times 10^2 & 9.710 \times 10 & 3.480 \times 10^2 & 1.836 \times 10 \\
5.855 \times 10 & -1.686 \times 10^2 & 1.031 \times 10 & 1.102 \times 10^2 \\
4.489 \times 10^2 & 5.11 & -3.583 \times 10^2 & 1.000 \times 10^{-3} \\
3.253 & 6.643 \times 10 & 1.000 \times 10^{-3} & -1.285 \times 10^2 \end{bmatrix}$$  \hspace{1cm} (B.2)

<table>
<thead>
<tr>
<th>$\hat{C}^{p(0)}$</th>
<th>$m = 0$</th>
<th>$m = 1$</th>
<th>$\hat{C}^{p(0)}$</th>
<th>$m = 0$</th>
<th>$m = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>$[-1.84 \times 10^2 \quad 2.88 \times 10^2]$</td>
<td>$[1.84 \times 10^2 \quad -2.88 \times 10^2]$</td>
<td>$[-3.60 \times 10^0 \quad 1.68 \times 10^2]$</td>
<td>$[3.60 \times 10^0 \quad -1.68 \times 10^2]$</td>
<td></td>
</tr>
<tr>
<td>$p = 2$</td>
<td>$[-1.68 \times 10^2 \quad 1.00 \times 10^2]$</td>
<td>$[1.68 \times 10^2 \quad -1.00 \times 10^2]$</td>
<td>$[-6.36 \times 10^0 \quad 3.12 \times 10^2]$</td>
<td>$[6.36 \times 10^0 \quad -3.12 \times 10^2]$</td>
<td></td>
</tr>
<tr>
<td>$p = 3$</td>
<td>$[-2.59 \times 10^2 \quad 3.40 \times 10^2]$</td>
<td>$[2.59 \times 10^2 \quad -3.40 \times 10^2]$</td>
<td>$[-4.29 \times 10^2 \quad 2.22 \times 10^2]$</td>
<td>$[4.29 \times 10^2 \quad -2.22 \times 10^2]$</td>
<td></td>
</tr>
</tbody>
</table>

Table B.2: Initial estimation of the matrix $\hat{C}_m^{p(0)}$, $m = 0, 1$, for each breaker $p = 1, ..., 6$. 33
References


34


