

MOX–Report No. 20/2009

**Bayesian hidden Markov models for
performance-based regulation of continuity of
electricity supply**

NADIA ACCOTO, TOBIAS RYDÉN, PIERCESARE SECCHI

MOX, Dipartimento di Matematica “F. Brioschi”
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it

<http://mox.polimi.it>

Bayesian hidden Markov models for performance-based regulation of continuity of electricity supply*

Nadia Accoto[◇], Tobias Rydén[△], Piercesare Secchi[‡]

July 13, 2009

◇ Dipartimento di Scienze Economiche e Metodi Quantitativi
Università del Piemonte Orientale
via Perrone 18, 28100 Novara, Italy
`nadia.accoto@eco.unipmn.it`

△ Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden
`tobias@maths.lth.se`

‡ MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica “F. Brioschi”
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
`piercesare.secchi@polimi.it`

Keywords: Reliability performance regulation, finite hidden Markov models, Phase-type distribution, cluster analysis

AMS Subject Classification: ...

Abstract

A fundamental aspect in the regulation of continuity of electricity supply is the identification of faults that could be caused by an exceptional event and therefore that are outside the utility control and responsibility. Different methods have been proposed during the years; the interpretation of the observed faults as a signal of an underlying system naturally leads to the analysis of the problem by means of an hidden Markov model. These models, in fact, are widely used for introducing dependence in data and/or for modelling observed phenomena depending on hidden processes. The application of this method shows that the model is able to identify exceptional events; moreover the study of the estimated model parameters gives rise to reality-linked considerations.

*We thank Ing. Luca Lo Schiavo and Autorità per l'Energia Elettrica e il Gas for data and support.

1 Introduction

Quality regulation in electricity distribution has received significant attention in recent years (CEER [5]; Fumagalli *et al.* [10]; Fumagalli *et al.* [11]).

One of the main problems in reliability performance regulation is how to identify events that are exceptional with respect to normal performance; regulators usually design incentive mechanisms specifically targeted at quality of supply, assuming the form of financial penalties and rewards to the distribution utility based on the expected performance. It is therefore crucial to understand when failure in meeting the regulatory targets is due to the utility behavior or to events that are outside the utility control and that could be considered exceptional. Moreover, even if some events, such as extreme weather conditions, are unavoidable, regulators have become more interested in controlling the efficiency and effectiveness of utility restoration schemes.

Traditional criteria for identifying data due to exceptional situations are based on different definitions of exceptional event, given in terms of number of customers interrupted, duration of the interruptions, weather conditions, extent of the mechanical damage to the distribution system and combinations of these factors. The application of these criteria in Italy from 2000 to 2003, however, resulted in some practical cases quite difficult and ambiguous. For these reasons the Italian Regulatory Authority for Energy and Gas - AEEG - suggested, in 2004, the introduction of statistical methodologies, that should reduce ambiguities and increase fairness. Nevertheless, statistical analyses of exceptional events can be performed in very different ways, depending on the choice of the quality indicator, on the spatial and temporal units of such measure, and, of course, on the statistical methods employed. We aim at contributing to the reliability performance regulation by means of a Hidden Markov Model (HMM - Rabiner [18]; MacDonald and Zucchini [14]; Cappé *et al.* [4]; Frühwirth-Schnatter [8]).

HMMs describe the relationship between an observed process $\{Y_t\}_{t>0}$ and an underlying and unobserved process $\{X_t\}_{t\geq 0}$; the hidden process is assumed to follow a Markov chain whose realization X_k governs the distribution of the corresponding Y_k . The interpretation of the number of electrical faults as a signal of an underlying and not observable process naturally leads to analyse the problem by means of an HMM. The proposed statistical method is based on the idea that the observed number of interruptions depends on the latent status of the “global” system controlling the electricity distribution; the latent system operating status along time is modeled by a Markov chain, whose realization controls the distribution of the corresponding number of observed faults.

The brief description of methods adopted by the AEEG during the years, in Section 2, will help us to clarify some aspects of the problem. In Section 3 we will present the *finite hidden Markov model* and how to make inference when a Bayesian approach is adopted. After briefly presenting the data in Section 4, we will report and widely discuss results in Section 5. The final considerations

in Section 6 complete the analysis. All data analysis are performed in R (R Development Core Team [17]).

2 The methods adopted by the Italian Regulatory Authority

Since the Italian Regulatory Authority introduced the first continuity of supply regulation in 2000, different methods have been proposed, applied and in some cases criticized and improved (see Fumagalli *et al.* [11] for a review).

In the first regulatory period (2000 - 2003), the AEEG required utilities to classify interruptions according to three categories: *Force Majeure*, external causes (*i.e.* third party responsibilities and interruptions originated on the transmission grid) and utility responsibilities. AEEG accepted a *Force Majeure* attribution only if the exceptional nature of the event could be proven by technical or administrative evidence. For instance, a formal declaration of calamity made by the government or measures of wind speed made by an independent weather center. In practical terms, this procedure turned out to be onerous both for the companies, that were collecting the data, and for the regulatory authority, that was controlling the documentation provided. In addition, a few controversial cases, where the exceptional nature of the event was claimed by the companies, but could not be formally proven, generated a large amount of disputes.

For these reasons, the AEEG began to consider a simpler procedure for identifying an exceptional event on the basis of the nature of the interruption it caused, compared to the characteristics of the interruptions caused by “normal events”. Then, in 2004, for the regulatory period 2004 - 2007, AEEG introduced a statistical methodology for identifying exceptional events, based on the idea that such events are characterized by extremely long restoration times (Christie [6]; Warren *et al.* [25]; Fumagalli *et al.* [9]). Even if a significant reduction in administrative work resulted, on both the regulator and the utilities side, the method showed some empirical problems.

During the consultation process for the third regulatory period additional elements emerged that led to a new statistical approach for the identification of exceptional events; the new proposed method was incorporated by AEEG, in the Regulatory Order 333/07 (AEEG [1]). With this method, given a utility observed performance, a 6 hour interval is deemed to be exceptional when the number of faults registered in that interval is larger than an *exceptionality threshold* determined by a four step procedure (see Fumagalli *et al.* [11] for details). Once the exceptional intervals have been identified, an exceptional period is defined as the time interval beginning 3 hours before the beginning of an exceptional interval and terminating 3 hours after the end of the same interval. The AEEG method is briefly presented in Appendix A.

The method adopted by the Authority analyses each utility separately from the others and uses only the observed performance in terms of number of faults, without considering any other type of information. Moreover this method incorporates the idea that an exceptional event causes several faults protracting in time - in fact it considers the number of interruptions greater than a threshold and the analysis is based on the sum of the interruptions occurred in a 6 hour time interval. Finally enlarging an exceptional interval to the 3 hours before the beginning and 3 hours after the end of the same interval, the AEEG method incorporates the idea that the exceptional event is preceded and followed by some instability conditions.

Taking in mind these considerations we now present a *finite hidden Markov model* for the identification of exceptional events; we will adopt a Bayesian approach for solving the inferential problems involved in the analysis.

3 Bayesian hidden Markov model

3.1 The model

A Hidden Markov Model (HMM) is a bivariate discrete time process $\{X_t, Y_t\}$, where $\{X_t\}_{t \geq 0}$ is a Markov chain and, conditional on $\{X_t\}$, $\{Y_t\}_{t > 0}$ is a sequence of independent random variables such that the conditional distribution of Y_k only depends on X_k .

The distribution of a variable X_{k+1} conditionally on the history of the process, X_0, \dots, X_k , is determined by the value taken by the preceding one, X_k (Markov property); likewise, the distribution of Y_k , conditionally on the past observations Y_1, \dots, Y_{k-1} and the past value of the state X_0, \dots, X_k , is determined by X_k only.

We consider a *finite hidden Markov model*, where the state space of the Markov chain and the set in which $\{Y_t\}$ takes its values are finite sets, say $\mathbf{X} = \{1, 2, \dots, K\}$ and $\mathbf{Y} = \{0, 1, \dots, q\}$. This model can be characterized by the initial state distribution $\pi = \{\pi_i\}$, with $\pi_i = P(X_0 = i)$, $i \in \mathbf{X}$, the transition matrix $A = \{a_{i,j}\}$, where $a_{i,j} = P(X_{k+1} = j | X_k = i)$, $i, j \in \mathbf{X}$ and the emission matrix $B = \{b_i(y)\}$, with the conditional probabilities $b_i(y) = P(Y_k = y | X_k = i)$, $i \in \mathbf{X}$, $y \in \mathbf{Y}$.

Let $\boldsymbol{\vartheta}$ be a vector containing the model parameters, arising from the transition mechanism (*i.e.* the transition matrix A and the initial state probability distribution π) and from the emission mechanism (*i.e.* the emission matrix B). Consider a sequence of length T and let (\mathbf{y}, \mathbf{X}) be the complete data, $(Y_1 = y_1, \dots, Y_T = y_T, X_0 = x_0, X_1 = x_1, \dots, X_T = x_T)$; then the complete data likelihood function $p(\mathbf{y}, \mathbf{X} | \boldsymbol{\vartheta})$ is given by

$$p(\mathbf{y}, \mathbf{X} | \boldsymbol{\vartheta}) = \pi_{x_0} \prod_{k=1}^K \left(\prod_{t: X_t=k} b_k(y_t) \right) \prod_{i=1}^K \prod_{j=1}^K a_{i,j}^{n_{ij}} \quad (1)$$

where $n_{ij} = \#\{1 \leq t \leq T : X_{t-1} = i, X_t = j\}$, $\forall i, j \in \mathbf{X}$.

Regarding the transition mechanism, we fix the initial state at 1, *i.e.* $X_0 = 1$ (for a discussion on the choices of π and related prior specification see Cappé *et al.* [4], Subsection 13.1.2 or Frühwirth-Schnatter [8], Subsection 10.3.4). We will consider an HMM with a four state Markov chain, *i.e.* $K = 4$. States have a physical meaning: state 1 indicates that the system is in a normal operating status while state 4 indicates an exceptional operating status; states 2 and 3 are transitional and refer to an increasing degree of perturbation of the system operating status, as yet non exceptional.

Regarding the emission mechanism, we will consider observations greater than 9 just as “many interruptions”, *i.e.* $Y = \{0, 1, \dots, 9, > 9\}$; because the finite HMM could be considered as a nonparametric model, the introduction of this threshold does not affect the estimating method and it makes possible to compare not only different utilities, but also the performance of the same utility in different years.

3.2 Inference

When an HMM is considered, problems of interest are inference on the model parameters $\boldsymbol{\vartheta}$ and on the unobserved chain $\{X_t\}$. For many years HMMs have been implemented using recursive algorithms developed for parameter estimation (Baum and Petrie [2]; Baum *et al.* [3]) and for restoring the hidden Markov chain (Viterbi [24]). More recently, these models have been studied from a Bayesian point of view (among others Robert *et al.* [20]; Robert *et al.* [21]; Chib [7]).

In a Bayesian approach, model parameters are random quantities, on which a prior has to be assigned; a standard prior assumption is that the emission matrix is a priori independent of the transition matrix: $p(\boldsymbol{\vartheta}) = p(A)p(B)$.

Moreover, we assume the rows of the transition and emission matrices to be independent a priori, each following a conjugate Dirichlet prior; for instance $p(A) = \prod_{i=1}^K \text{Dir}(\mathbf{a}_i | \alpha_{i1}, \dots, \alpha_{iK})$, where $\mathbf{a}_i = (a_{i,1}, \dots, a_{i,K})$ is the i th row of the transition matrix A and $p(B) = \prod_{i=1}^K \text{Dir}(\mathbf{b}_i | \beta_{i0}, \dots, \beta_{iq})$, where $\mathbf{b}_i = (b_i(0), \dots, b_i(q))$ is the i th row of the emission matrix B .

Inference on the model parameters and on the hidden chain is based on the posterior distribution

$$\begin{aligned} p(\mathbf{X}, \boldsymbol{\vartheta} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{X}, \boldsymbol{\vartheta}) p(\mathbf{X} | \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}) \\ &\propto p(\mathbf{y}, \mathbf{X} | \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}). \end{aligned} \tag{2}$$

Sampling from the posterior (2) is commonly carried out by a Markov Chain Monte Carlo (MCMC) sampling scheme; an HMM’s missing-data structure naturally admits posterior samplers that alternate between simulating \mathbf{X} given $\boldsymbol{\vartheta}$ and \mathbf{y} , and simulating $\boldsymbol{\vartheta}$ given the complete data \mathbf{X} and \mathbf{y} (*i.e.* we can apply a particular MCMC sampling scheme that is the Gibbs sampling method).

Because the transition and the emission matrices are assumed independent a priori, they are (given \mathbf{X}) independent a posteriori. Given a trajectory \mathbf{X} , each

row of the transition matrix has, independently from the others $K - 1$ rows, a posterior Dirichlet distribution

$$\text{Dir}(\mathbf{a}_i | \alpha_{i1} + n_{i1}, \dots, \alpha_{ij} + n_{ij}, \dots, \alpha_{iK} + n_{iK}), \quad (3)$$

where $n_{ij} = \#\{0 \leq t \leq T - 1 : X_t = i, X_{t+1} = j\}$, $i, j \in \{1, \dots, K\}$. Moreover, given the complete data \mathbf{X} and \mathbf{y} , the rows of the emission matrix are independent and each row has a posterior Dirichlet distribution

$$\text{Dir}(\mathbf{b}_i | \beta_{i0} + e_{i0}, \dots, \beta_{iy} + e_{iy}, \dots, \beta_{iq} + e_{iq}), \quad (4)$$

where $e_{iy} = \#\{1 \leq t \leq T : X_t = i, Y_t = y\}$, with $i \in \{1, \dots, K\}$ and $y \in \mathcal{Y}$. The path \mathbf{X} of the hidden Markov chain is sampled by the *global updating scheme* (Cappé *et al.* [4]), where the trajectory of the hidden chain is updated as a whole from its conditional distribution given the data and the model parameters $\boldsymbol{\vartheta}$. The *global updating scheme* and the Gibbs sampling algorithm are described in Appendix B.

An HMM is a generalization of a Mixture model, where the components are not selected independently, but according to a Markov chain; with these models it shares an identifiability issue known as label-switching (Scott [23]). In fact, the HMM likelihood (1) is invariant under permutations of the state labels and, because also the introduced prior distributions are exchangeable (*i.e.* they are invariant under permutations of the components), the posterior will also be exchangeable. In order to avoid this problem it is common to assume constraints ordering the parameters; these constraints can be informative, because we construct a new prior that is zero in regions where the constraints do not hold; however, scientific insight about the chain may suggest an order for the parameter. In our case, given the physical interpretation of the states, it is natural to assume that the mean of the observed number of faults in the normal state 1 is smaller than the mean in the transitional state 2 and so on; in other words we will assume that $\mu_1 < \mu_2 < \mu_3 < \mu_4$, where $\mu_i = \sum_{j=0}^q j \cdot b_i(j)$.

Posterior draws produced by the MCMC sampler, provided that a sufficiently large number of initial draws (the burn-in period) are discarded, are used for statistical inference (for a discussion on different methods to estimate the hidden chain, see for example Scott [23]). Considering a quadratic loss function (*i.e.* $\text{loss}(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}}) = \|\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}\|^2$), the parameter estimate is the posterior expectation, that is approximated by averaging over the draws from the posterior distribution. Point estimation of the hidden Markov chain \mathbf{X} may be obtained by considering a 0/1 loss function, minimized by the mode of the marginal posterior probability, also called the *maximum a posteriori* (MAP) estimator.

4 Data presentation

First of all we define the statistical units that we are going to consider in the analysis. The Italian electricity distribution sector includes one very large utility

(that serves more than 80% of consumers and covers more or less the entire Italian territory) and a number of local companies (whose distribution territory is always confined within a single province). Each statistical unit is an ‘object’ identifiable by its geographical position (the reference province) and the serving utility (the large or the local one); in the following we will refer to these statistical units as *administrative units*.

The dataset provided by the AEEG includes the observed hourly number of faults, in each of the 113 administrative units for the year 2004. For illustrative purposes we first present data and results for one administrative unit; we then proceed to summarize the results for the whole dataset. The distribution of absolute frequencies of the number of faults observed during 2004 for the (randomly chosen) administrative unit is:

y	0	1	2	3	4	5	6	7	8	11	17	24
	8100	570	66	23	8	7	4	1	1	2	1	1

This situation is quite typical and the following comments hold for each administrative unit. The great majority of observations (more than 90%) are equal to 0, a quite large number of faults are equal to 1 and so on in a decreasing order. Note that just four observations are greater than 9, and thus the assumption to relabel observations greater than 9 with >9 (see last part of Section 3.1) involves a very small number of observations.

Moreover the performance of each administrative unit follows a common pattern; for example, consider the observed performance for the administrative unit previously considered during the whole 2004 (Figure 1) and during a week, from the beginning of July 23rd to the beginning of July 30th (Figure 2). We see that very often no interruptions occur, sometimes an intermediate number of faults are observed and there are periods with several faults lasting in time.

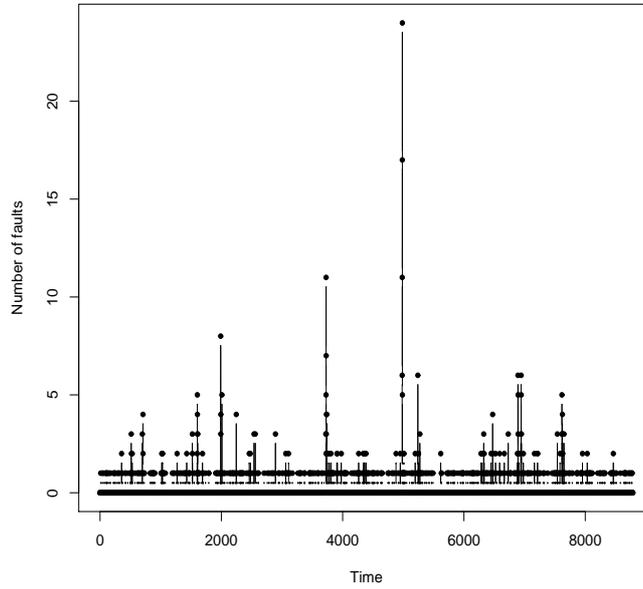


Figure 1: Observed performance for an administrative unit, year 2004.

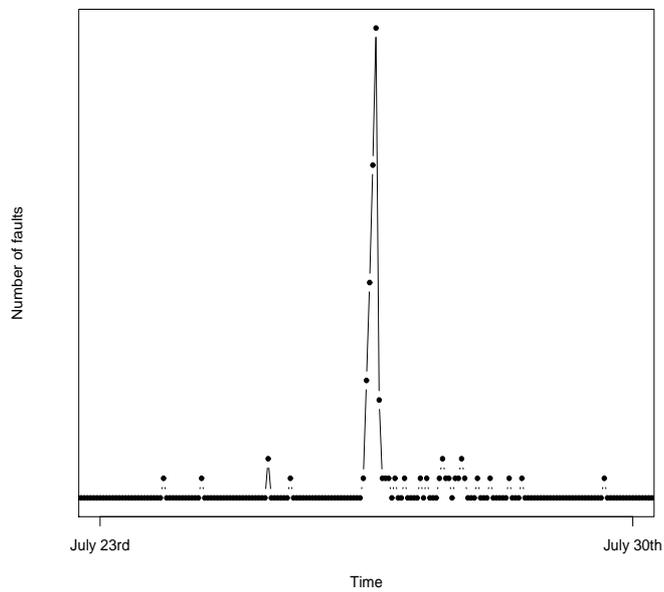


Figure 2: Observed performance for an administrative unit during a week, from the beginning of July 23rd to the beginning of July 30th.

5 Results

5.1 Identifying exceptional events

We analyse each administrative unit separately from the others, as an independent system.

Dirichlet prior parameters are set equal to 1, for the transition matrix, and equal to $4/11$, for the emission matrix. A Dirichlet distribution with parameters equal to 1 is a Uniform distribution; parameters of the Dirichlet distributions on the emission matrix are set to be smaller than 1 in order to have a large prior variance. Estimates are obtained running 30 000 iterations of the Gibbs sampler and discarding the first 2 000 generated values; convergence of the sampling algorithm is verified.

We now present results obtained by analysing, by the proposed HMM model, the performance of the administrative unit described in Section 4. In Figure 3, for each hour t , the observation y_t is represented with a different symbol and color according to the state x_t of the estimated Markov chain; to better understand the results, the values actually observed are plotted (in fact, recall that, during the estimation process, observations greater than 9 were relabeled as >9). Red marks at the bottom of the plot refer to periods deemed as exceptional by the AEEG method; in order to obtain these periods, we compute the *exceptionality threshold* using the 6 hours time-interval data, and we declare (label) the hours belonging to the obtained exceptional periods as AEEG exceptional (see Appendix A for details on the AEEG method). There is a good concordance between the HMM exceptional observations (dark red triangles in Figure 3) and the AEEG exceptional periods. Given the large number of observations ($8\,784 = 366 \times 24$, in fact 2004 was a leap year) and the differences, also conceptual, in the identification methods, it is almost impossible that the exceptional events perfectly match.

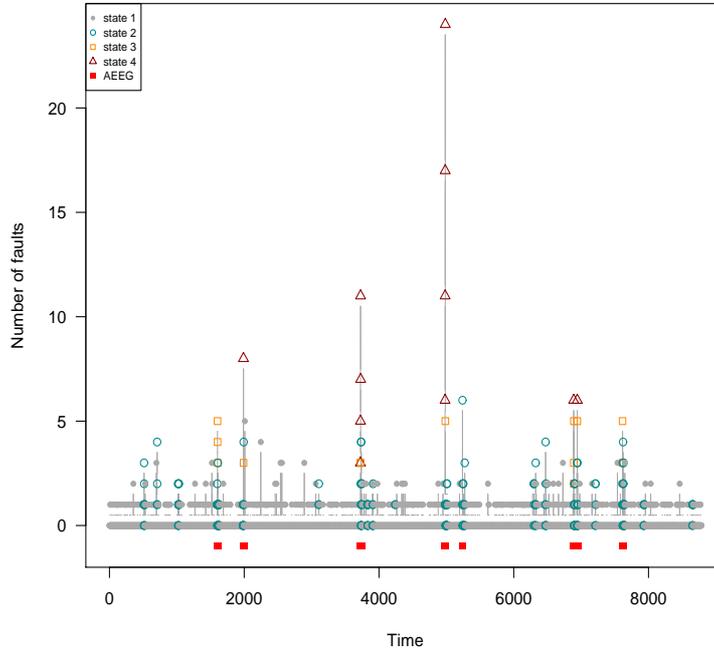


Figure 3: Observed performance for an administrative unit, year 2004; observations with different colors and symbols based on the estimated underlying Markov chain. Red marks at the bottom represent the AEEG exceptional periods.

Moreover, by inspecting Figure 3, we see that, unlike the AEEG method, the HMM method does not consider as exceptional all the observations greater than a threshold, but it identifies more complex patterns. Because of the underlying Markov structure, the same number of faults is exceptional or not depending on the system performance close to that observation. To better understand this consideration, we report the estimated transition matrix and (transpose) emission matrix in Figure 4. We may notice some general features (that appear in the estimated transition and emission matrices for all administrative units). The probability of staying in state 1, $a_{1,1}$, is almost equal to 1: when the global system is in state 1, it means that it is in the normal operating status and this is the most common situation. Also, the probability of staying in the exceptional state, $a_{4,4}$, is large: this means that, when the system enters in an exceptional operating status, it is highly probable that it remains in it for some time steps (hours), before visiting other states. Regarding the emission mechanism, by inspecting Figure 4 it emerges that: states 1 and 2 mainly “emit” observations equal to 0 and 1, when the chain is in state 3 it is highly probable to have from

3 to 5 faults and when the underlying system is in state 4, it is likely to observe higher values.

$$\hat{A} = \begin{pmatrix} 0.994 & 0.005 & 0.001 & \approx 0 \\ 0.006 & \approx 0 & 0.979 & 0.014 \\ \approx 0 & 0.711 & 0.077 & 0.212 \\ 0.001 & 0.019 & 0.392 & 0.588 \end{pmatrix} \quad \hat{B}^T = \begin{pmatrix} 0.943 & 0.477 & 0.085 & 0.047 \\ 0.052 & 0.378 & 0.103 & 0.055 \\ 0.004 & 0.095 & 0.110 & 0.073 \\ 0.001 & 0.024 & 0.220 & 0.180 \\ \approx 0 & 0.016 & 0.080 & 0.067 \\ \approx 0 & 0.003 & 0.156 & 0.155 \\ \approx 0 & 0.003 & 0.097 & 0.117 \\ \approx 0 & 0.001 & 0.036 & 0.055 \\ \approx 0 & 0.002 & 0.037 & 0.044 \\ \approx 0 & 0.001 & 0.018 & 0.021 \\ \approx 0 & 0.001 & 0.058 & 0.185 \end{pmatrix}$$

Figure 4: Estimated transition and (transpose) emission matrices.

Let us focus on time periods containing observations deemed, by the HMM method, as due to an exceptional event. Given that the estimated Markov chain visited an exceptional state 4, we can go backward and forward in time until the estimated chain reaches again the normal state 1; for this administrative unit and for the year 2004, we thus obtain 5 intervals that are plotted in Figure 5, where, for each plot, the title indicates the month and the abscissa labels the day and the hour when each interval starts and finishes.

By inspecting the plots in Figure 5, we see that each exceptional event is preceded and followed by an instability situation; indeed, this is also incorporated in the AEEG method. When an exceptional event occurs (*i.e.* the underlying chain is in state 4), it is interesting to consider the distribution of the random time requested for the system to reestablish the normal operating status (*i.e.* the chain comes back to state 1); this distribution can be gathered by the analysis of the estimated transition matrix of the hidden Markov chain.

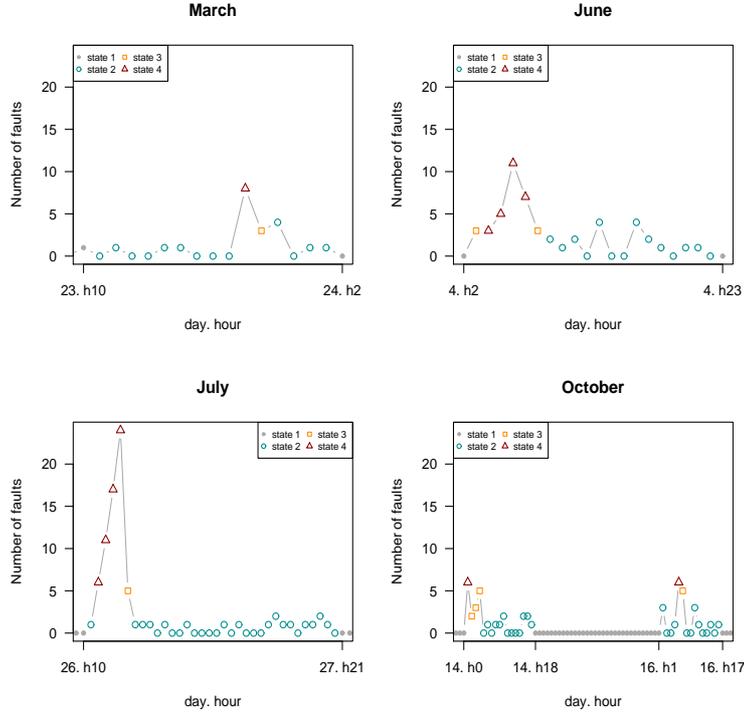


Figure 5: Intervals containing at least one exceptional event. Given that an exceptional event occurs, we can go backward and forward until the estimated chain is in the normal state 1; the title indicates the month and the abscissa labels the day and the hour when each interval starts and finishes.

5.2 Exceptional excursions

The study of the time needed to the chain to come back in the normal state, given that an exceptional event occurred, starts with the definition of an *exceptional excursion*. An exceptional excursion is a sequence of states beginning in the first state after a state 1 (it could be state 2, 3 or 4), ending in state 1 and containing at least one state 4; in Figure 5 are represented five exceptional excursions. It can be shown that the length of an exceptional excursion is discretely Phase-type distributed, where a Phase-type distribution (Neuts [16]) is the distribution of the number of steps until absorption for a finite Markov chain with all states that are transient but one that is absorbing. We refer to Appendix C for details relative to the definition of the Phase-type distribution, the proof that an exceptional excursion is discretely Phase-type distributed and how to obtain the estimated Phase-type distribution.

The estimated Phase-type distribution for the administrative unit so far analysed is reported in Figure 6. The length of an exceptional excursion represents the time spent by the chain outside the normal state 1, when an exceptional

event occurs and thereby the Phase-type distribution represents the distribution of the time needed to the electricity distribution utility to reestablish the normal operating status. Therefore we can use the estimated (length of) exceptional excursion distributions as a tool for evaluating the efficiency and the effectiveness of different utility restoration schemes. Consider for example Figure 7 showing the estimated Phase-type distribution of two administrative units (labeled as P1 and P2, where P2 is the administrative unit so far considered), in 2004. We see that the Phase-type distribution associated to P2 concentrates more mass on larger values and then it is more probable that P2 needs more time, compared to administrative unit P1, to reestablish the normal operating status.

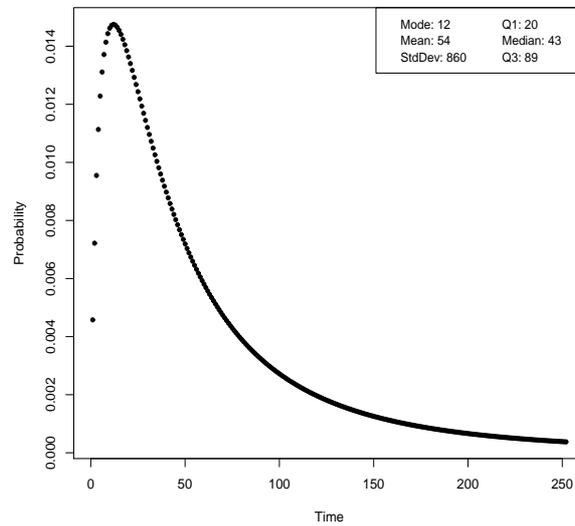


Figure 6: Estimated Phase-type distribution, that is the estimated distribution of the time needed to the system to reestablish the normal operating status. In the legend the basic indicators.

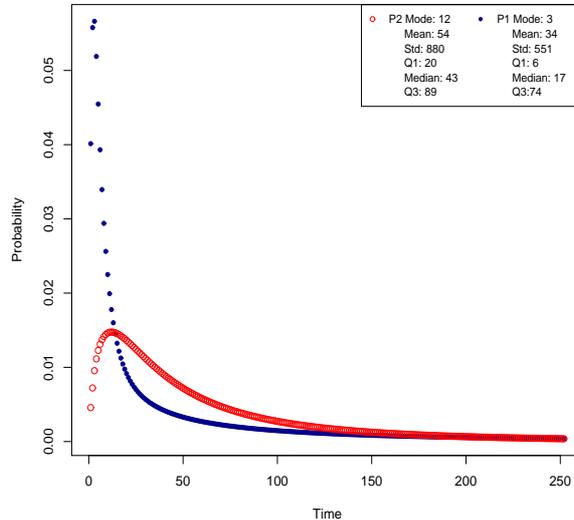


Figure 7: Estimated Phase-type distribution for administrative units labeled as P1 and P2; in the legend the basic indicators.

5.3 Analysing the Italian electricity distribution utilities

We have presented and discussed results related to a single administrative unit, but the same analysis has been performed for the whole dataset. Let us now summarize results related to the whole dataset. Considering together the 113 administrative units for the year 2004 (*i.e.* a total of $113 \times 8784 = 992592$ observations), we have that the great majority of observations - more than 97% - are classified in the normal state (*i.e.* the estimated underlying Markov chain is in state 1), 2% and 0.1% of the interruptions are classified in the transitional states 2 and 3 respectively, while just about the 0.09% of the faults - 881 observations - are classified as due to an exceptional event (*i.e.* the estimated hidden chain is in state 4). This is not an astonishing result, given that for each administrative unit, about 90% of observations are equal to 0.

For each administrative unit, because they are analysed separately, the model evaluates as exceptional, interruptions of different magnitude. The presumption is that the model indirectly incorporates all relevant exogenous information, even though it does not explicitly takes into account any covariate different from the hourly number of faults, like, for instance, the geographical size of the administrative unit or the number of served costumers. Note that the same adaptive mechanism characterizes the AEEG method; indeed the AEEG exceptionality threshold, based on the observed performance, is typically different for each administrative unit.

In order to understand if, at the same hour, the exceptionality involves more

than one administrative unit, consider the number of administrative units that are simultaneously experiencing an exceptional situation. We previously said that the exceptionality involves a larger time period, described by the exceptional excursion; then given that an exceptional event occurs, we enlarge the interval in order to also consider the instability condition preceding and following an exceptional event. Note that this relabelling process is conceptually similar to the AEEG procedure, that considers as exceptional the three hours preceding and following an exceptional interval.

For each hour when at least one administrative unit is experiencing an exceptional excursion, we count the number of administrative units that are simultaneously experiencing an exceptional excursion. Table 1 reports the number of times one, two, . . . , different administrative units experienced a simultaneous exceptional instability situation in 2004; in particular during a single hour (January 30th, 5 AM), 15 different administrative units were managing an exceptional situation. Let us focus on the administrative units that generated the 796 “single exceptional excursions”.

Without considering any technical or morphological information, we expect that an exceptional situation (e.g. due to a meteorological phenomenon) involves more than one administrative unit; consequently administrative units that more often are alone in the exceptional excursions are likely to be particularly sensible to changes in underlying conditions. In Figure 8, the province relative to each administrative unit is colored with a different intensity on the basis of the number of hours the administrative unit has been the only one in an exceptional excursion. Note that administrative units with “single exceptional excursions” are mainly placed in the South part of Italy.

# of prov.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Hours	796	291	170	111	85	81	57	24	11	5	10	8	7	2	1

Table 1: Number of administrative units simultaneously experiencing an exceptional excursion in 2004. For 796 times only one administrative unit was in an exceptional instability situation and so on.

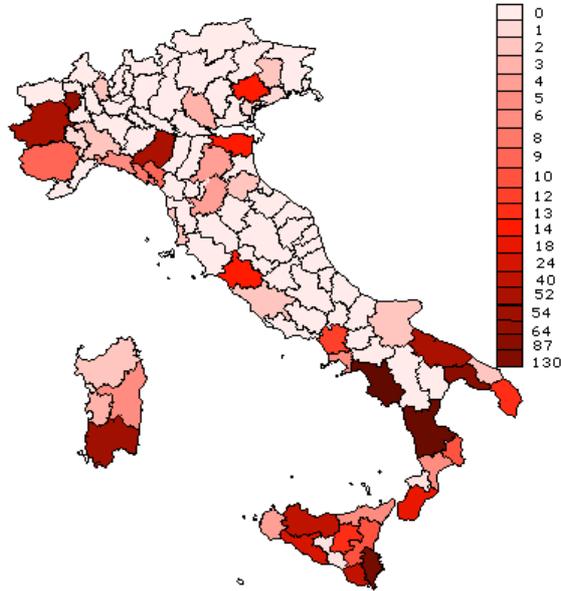
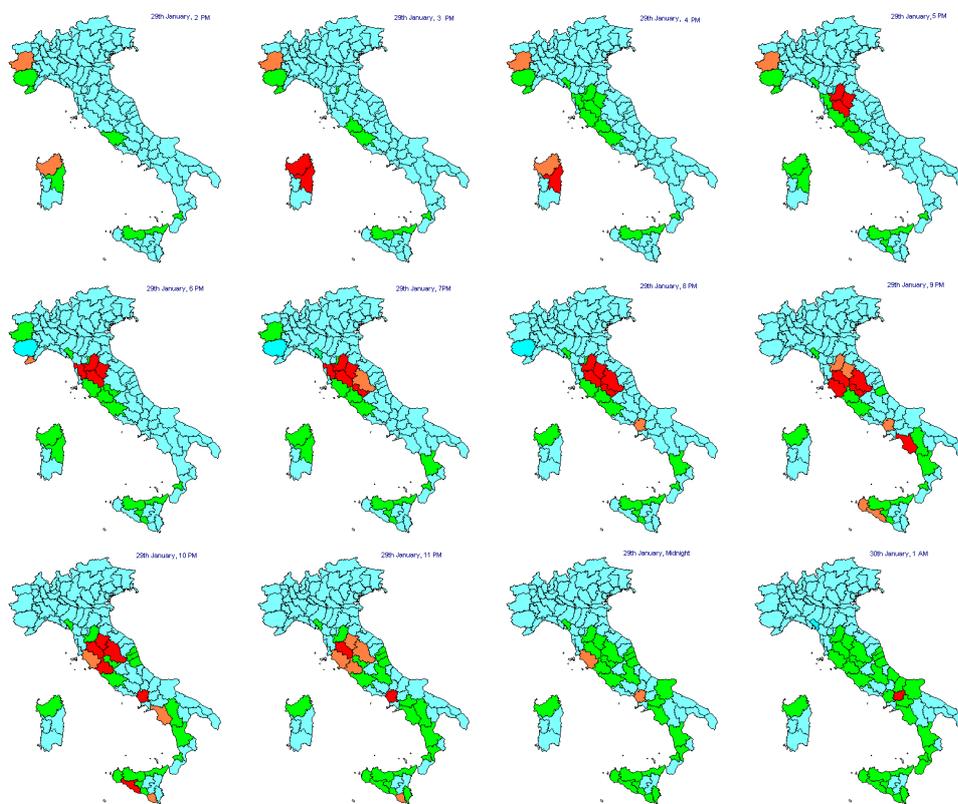


Figure 8: Geographical distribution of the number of times the administrative units have been the only one experiencing an exceptional excursion in 2004: the province relative to each administrative unit is colored with different intensity proportional to the number of times the administrative unit has been the only one experiencing the exceptional excursion.

For each hour t in 2004 and for each administrative unit, the estimated Markov chain indicates the system working operating status; moreover the geographical position of each administrative unit is known. Therefore, for each hour t , we can picture the simultaneous system operating statuses by coloring the province relative to each administrative unit with a different color on the basis of the state of the underlying Markov chain at that time. Then, by letting t run, we may appreciate the evolution in space and time of the exceptional events. Consider, for example, Figure 9; each plot shows the provinces operating status at a certain time: if the underlying system is in state 1 the area is colored in pale blue, if it is in state 2 in green, in state 3 in orange and in state 4 in red. The Figure shows the temporal evolution of the situation since January 29th 2 PM to January 30th 1 PM; focussing on red areas that indicate an exceptional operating status, we notice that the instability situation started from the North/West and, as time goes by, it involves the Center and finally the South-East part of Italy. Therefore it seems that with the model (even if it was estimated separately for each administrative unit) we are gathering an unsettled situation that plausibly was a meteorological phenomenon.



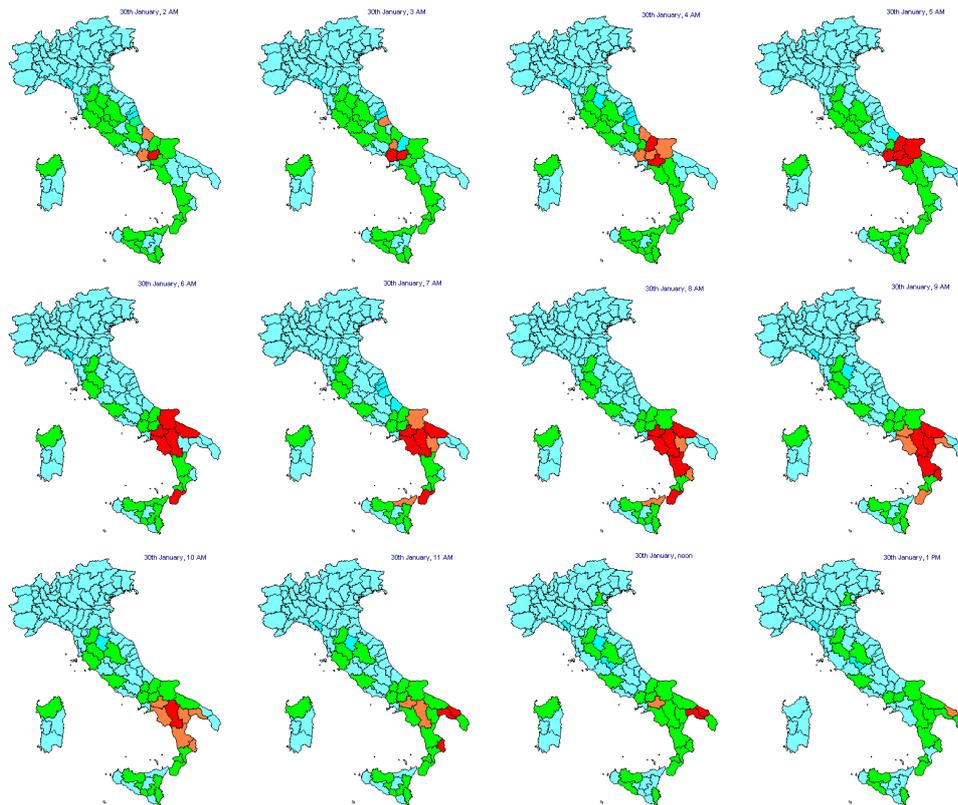


Figure 9: Spatio-temporal evolution of the exceptional events: the province relative to each administrative unit has a different color on the basis of the state of the estimated underlying Markov chain; if the system is in state 1 the area is colored in pale blue, if it is in state 2 in green, in state 3 in orange and in state 4 in red. The plots show the situation from January 29th 2 PM to January 30th 1PM; time evolves from top to bottom and from left to right.

We conclude this Section comparing results obtained with the HMM method and those obtained with the method adopted by the AEEG. The two methods are quite different and, in particular, the AEEG method is “rigid” in the sense that it works with 6 hours intervals and it declares exceptional the three hours before and after an exceptional interval. It is almost impossible that periods deemed as exceptional by the two methods perfectly overlap, but, when there is an agreement, those periods intersect.

Let us introduce a simple notation: we indicate with \mathbf{H} the set of HMM exceptional periods H , i.e. the intervals of time spent by the estimated Markov chain in state 4, and with \mathbf{A} the set of time intervals A declared exceptional by the AEEG method. We then define the set of “stretched intervals” as

$$S = \{H \cup A : H \in \mathbf{H}, A \in \mathbf{A}, H \cap A \neq \emptyset\},$$

and set $\mathbf{H}' = \mathbf{H} \cup S$ and $\mathbf{A}' = \mathbf{A} \cup S$. If an hour t belongs to a time interval in \mathbf{H}' it is labeled as “ H -exceptional” otherwise it is labeled as “ H -normal”. Likewise, if an hour t belongs to a time interval in \mathbf{A}' it is labeled as “ A -exceptional” otherwise it is labeled “ A -normal”. For each administrative unit we may then compute the confusion matrix:

	HMM	H -exceptional	H -normal
AEEG			
A -exceptional		n_{AH}	n_{AHC}
A -normal		n_{ACH}	n_{ACHC}

For example, n_{AH} is the number of hours simultaneously declared exceptional by the two methods. The (A -normal, H -normal) situation is by far the most frequent. Then, as a measure of similarity, instead of the simple matching coefficient that could overestimate similarity, we use the Dice coefficient (it is similar to the Jaccard similarity coefficient but gives twice the weight to agreements); that is, for each administrative unit, we evaluate the similarity (the agreement) between the two methods by

$$D = \frac{2n_{AH}}{2n_{AH} + n_{AHC} + n_{ACH}}.$$

In 21 administrative units the HMM method does not identify exceptional events, whilst the AEEG method declares exceptional a few hours; for 3 administrative units no observations are considered as due to an exceptional event by any method. For the other administrative units the absolute frequencies distribution of the Dice coefficient is reported in the next table, that shows a satisfactory agreement between the two methods.

D	(0; 0.25]	(0.25; 0.5]	(0.5; 0.75]	(0.75; 1]
	4	13	31	41

5.4 A Cluster analysis

Even if each administrative unit was analysed separately from the others it is of interest to understand if administrative units are in some sense similar. Exceptional events are the central point of our study and, given the interpretation of the model, the hidden process is the mechanism that manages the occurrence of the exceptional operating status experienced by the system. For this reason we will investigate, by means of a Cluster analysis, if administrative units are similar with respect to the underlying process.

As a by product of the application of an HMM we have, for each administrative unit, estimated values for the transition matrix A , the emission matrix B , the hidden Markov chain and the Phase-type distribution. In order to reach the set goal we will consider estimated parameters related to the hidden Markov chain. The Phase-type distribution has a “technical interpretation” (in fact it is the distribution of the time needed for each system to reestablish the normal situation) and it also permits a more visible comparison between administrative units; nevertheless it is obtained by a transformation of the transition probabilities and then it “contains less information” than the transition matrix. Therefore the transition matrix seems to be more adequate for our purpose. Of course, given a transition matrix, generated paths can be different; in other words, even if two transition matrices are quite similar the generated chains might be different. Note that considering the transition matrices and the path of the chains, we want to gather two different aspects of the system: the *potential* transition dynamic between states and the operating status *actually* experienced by the systems.

A measure of the distance between probability distributions and a dissimilarity measure between the estimated Markov chains, able to underline if there are periods with a similar behavior, need to be introduced.

Because each transition matrix is a collection of K probability distributions (where we recall $K = 4$ is the number of possible states) and rows with the same index are probability distributions conditional on the same event (that is “the chain is in state i ”), we consider as a measure of dissimilarity the average of the symmetrized Kullback-Liebler distance between corresponding rows (Ramoni *et al.* [19]). Let $a_{i,j}^q$ and $a_{i,j}^r$ be the transition probabilities from i to j in two transition matrices A^q and A^r (corresponding to administrative units labeled q and r). The Kullback-Liebler divergence between rows i , \mathbf{a}_i^q and \mathbf{a}_i^r , of these matrices is

$$d_p(\mathbf{a}_i^q, \mathbf{a}_i^r) = \sum_{j=1}^K a_{i,j}^q \log \frac{a_{i,j}^q}{a_{i,j}^r}, \quad (5)$$

where the subscript p is for “potential”. The distance in equation (5) is not symmetric because $d_p(\mathbf{a}_i^q, \mathbf{a}_i^r) \neq d_p(\mathbf{a}_i^r, \mathbf{a}_i^q)$; the symmetric version of it is defined as $D_p(\mathbf{a}_i^q, \mathbf{a}_i^r) = [d_p(\mathbf{a}_i^q, \mathbf{a}_i^r) + d_p(\mathbf{a}_i^r, \mathbf{a}_i^q)]/2$. Then the average distance between administrative units labeled with q and r , with respect to the transition matrices

A^q and A^r is

$$D_p(q, r) = \frac{1}{K} \sum_{i=1}^K D_p(\mathbf{a}_i^q, \mathbf{a}_i^r). \quad (6)$$

Consider now dissimilarity measures between the estimated Markov chains. Given the physical interpretation of the states of the hidden Markov chain, as increasing degree of system perturbation, we could consider a path of the chain as a sequence of ordinal values.

Indicate by \mathbf{Z} the estimated hidden Markov chain with observations in an exceptional excursion relabelled as exceptional. Given the relabelled paths \mathbf{Z}^q and \mathbf{Z}^r , relative to administrative units q and r , consider the Spearman rank correlation coefficient $\rho(\mathbf{Z}^q, \mathbf{Z}^r)$ (see Lehmann [13]) and set the dissimilarity between q and r with respect to the estimated chain as

$$D_a(q, r) = \frac{1 - \rho(\mathbf{Z}^q, \mathbf{Z}^r)}{2}, \quad (7)$$

where in a similar way as before the subscript a is for “actual”.

In order to reach our goal, we will apply the *Partitioning Around Medoids - PAM* algorithm (also called *k-medoids method*, Kaufman and Rousseeuw [12]). The number of groups to consider is suggested by an *agglomerative hierarchical method* (with the distance between clusters calculated by the *complete linkage method*).

Output concerning each cluster obtained by the PAM method can be graphically represented by the Silhouettes introduced by Rousseeuw [22]; briefly, Silhouettes provides a measure of how well a data point was classified when it was assigned to a cluster by according to both the tightness of the clusters and the separation between them. Consider an object i and denote by A the cluster to which it has been assigned; calculate

$$a(i) = \text{average dissimilarity of } i \text{ to all other objects of } A;$$

consider any cluster C different from A and define

$$\delta(i, C) = \text{average dissimilarity of } i \text{ to all other objects of } C.$$

After computing $\delta(i, C)$ for all clusters $C \neq A$, let

$$b(i) = \min_{C \neq A} d(i, C).$$

The cluster B such that $\delta(i, B) = b(i)$ is called the *neighbor* of object i and it is like the second-best choice for object i . The silhouette $s(i)$ is obtained as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (8)$$

and then

$$-1 \leq s(i) \leq 1.$$

Moreover by the definition of $s(i)$ we can deduce that observations with a large $s(i)$ (almost 1) are very well clustered, a small $s(i)$ (around 0) means that the observation lies between two clusters, and observations with a negative $s(i)$ are probably placed in the wrong cluster.

k-medoids methods are implemented in the R package *cluster* (Maechler *et al.* [15], that also provides the silhouette plot and other diagnostic tools.

We first consider clustering by means of the estimated transition matrices, with $k = 3$ possible clusters.

Figure 10 shows the silhouette plot, where, for each cluster C_j , $j = 1, \dots, k$, the silhouette $s(i)$, $i \in C_j$ is plotted by a bar in decreasing order; then, because values of $s(i)$ are positive and quite close to 1, we can deduce that administrative units are well classified.

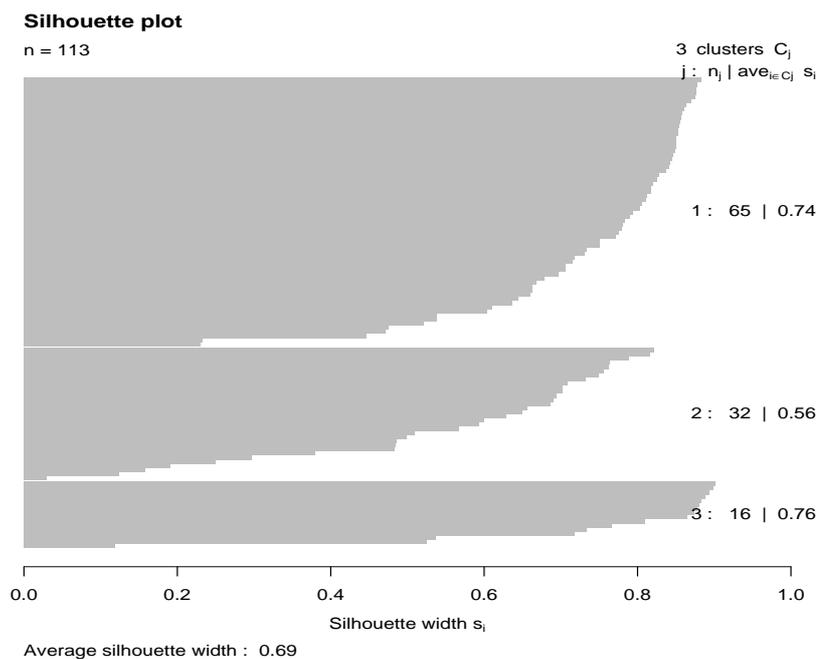


Figure 10: Silhouette plot for the clustering by the transition matrices. For each cluster the silhouette is plotted by a bar in decreasing order. Large values indicate that the objects are very well clustered.

The obtained clusters are described in Figure 11, where the province relative to each administrative unit is colored with a different color on the basis of the cluster to which it has been assigned. There are administrative units with the same geographical position; when there is a disagreement in terms of assigned

cluster the corresponding area is striped and colored with both colors, otherwise the number “2” indicates that the two utilities have been classified in the same cluster.

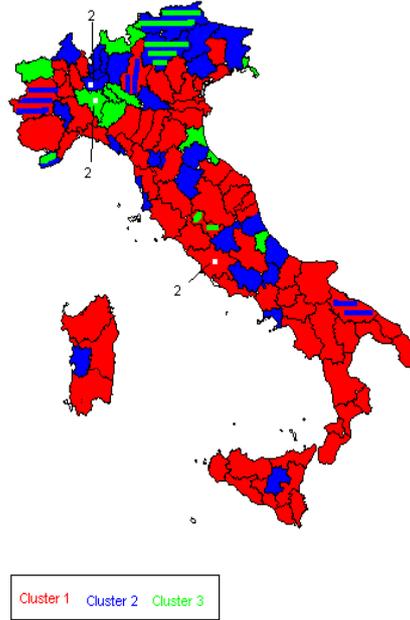


Figure 11: Clustering results by transitional dynamics. The area relative to each administrative unit is colored with a different color on the basis of the cluster to which it has been assigned; striped areas are referred to disagreements in terms of assigned clusters, while the number “2” indicates an agreement. White little points indicate the indicator starting point.

In order to understand which administrative units are classified in each cluster we consider the Phase-type distribution. Figure 12 shows Phase type distributions for medoid of each cluster; considering the graph where the distributions are plotted together (bottom-right) we can see that for administrative units in cluster 1 (red points) the Phase-type distribution concentrates more mass on larger values (that is they need more time to reestablish the normal situation), followed by administrative units in cluster 2 (blue circles) and in cluster 3 (green triangles).

Therefore a general classification could be:

- administrative units in cluster 1 → “exceptional persistent”
- administrative units in cluster 2 → “exceptional transitional”
- administrative units in cluster 3 → “fast recovering”

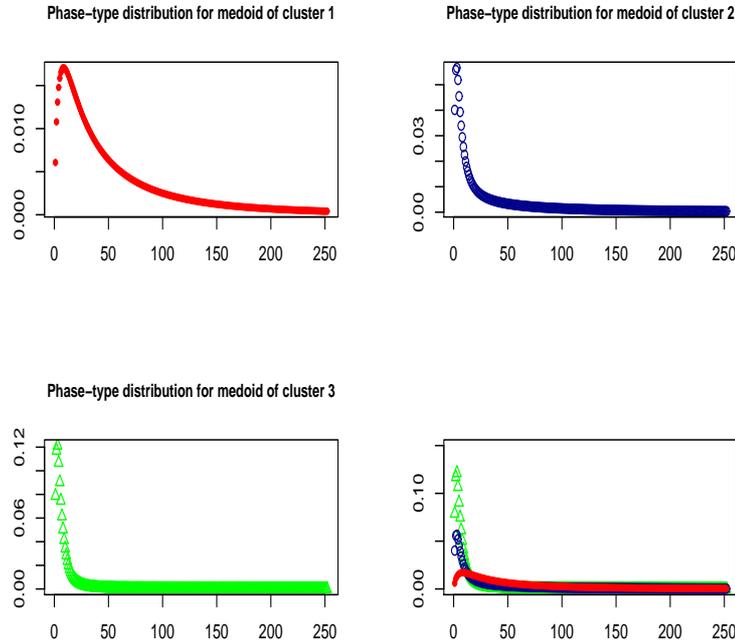


Figure 12: Phase-type distributions for medoids of each cluster; in the plot on bottom-right the three distributions are plotted together.

We now consider the clustering by means of the estimated hidden Markov chains. In order to calculate the Spearman correlation coefficient we will not consider administrative units whose estimated Markov chain was always in the normal state 1.

Silhouette plots, obtained by the PAM method with $k = 3$ and shown in Figure 13, indicate that clusters are not strongly defined and that there are administrative units in each cluster that have been classified in the wrong cluster by the PAM algorithm. Administrative units in each cluster are represented in Figure 14, plot on the left, with the same logic as in Figure 11. As we said the algorithm provides the neighbor cluster; then if we move administrative units with a negative silhouette to the neighbor we obtain a more uniform situation, shown in Figure 14, plot on the right. We underline that moving objects with negative silhouette value in the neighbor cluster is not a general rule (in fact, the neighbor cluster is the second-best choice); however we have information coming from the problem under analysis and this helps us for having more interpretable results.

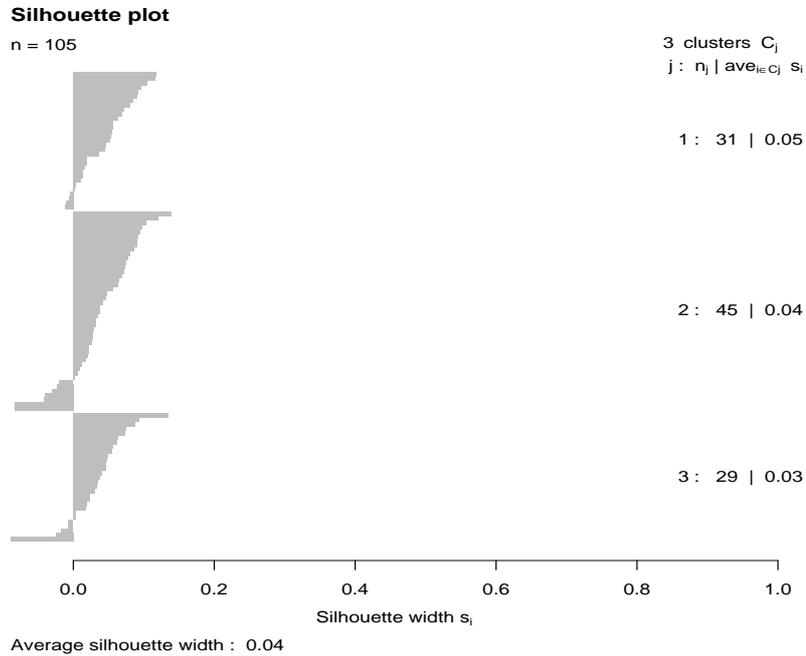


Figure 13: Silhouette plot for the clustering by the estimated hidden Markov chains. For each cluster the silhouette is plotted by a bar in decreasing order. Negative values indicate a misclassification.

Regarding the final interpretation, we recall that the dissimilarity measure used in this clustering analysis, $D_a(q, r)$, defined in equation (7), quantifies the dissimilarity between administrative units related to operating status actually experienced by the administrative units. Then by Figure 14, plot on the right, we could conclude that administrative units are affected by some “spatial dependence”; in fact Cluster 1 mainly contains administrative units in the South, Cluster 2 administrative units in the North and Cluster 3 administrative units in the Center part of Italy. Considering this clustering evidence along with what caught by analysing Figure 9, it seems that exceptional events spread geographically.

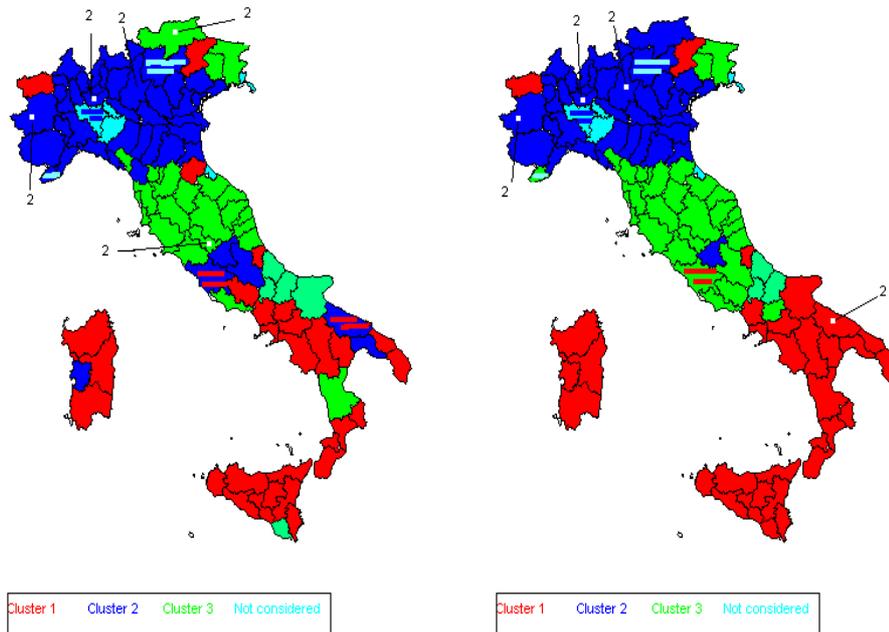


Figure 14: Clustering results by estimated Markov chains. The area supplied by each administrative unit is colored with a different color on the basis of the cluster to which it has been assigned; striped areas are referred to disagreements in terms of assigned clusters, while the number “2” indicates an agreement. Administrative units not considered in the cluster analysis are colored in pale blue. White little squares just indicate the indicators starting point. Left: clusters before the reallocation of misclassified administrative units to the neighbor cluster. Right: clusters after the reallocation of the misclassified administrative units.

6 Final considerations

We have presented a finite hidden Markov model for identifying exceptional events in electricity distribution. The application of this method showed that the model fits well the data and the estimated hidden chain is able to identify exceptional events as those where a large number of faults protracting in time occur. The HMM method supports the AEEG method, even if the applied statistical methods are completely different.

Moreover, given the model structure, by the inspection of the posterior distribution of the transition matrix, we can gather information relative to the behavior of the system when it is experiencing an exceptional operating status. We introduced, in fact, the exceptional excursions and the Phase-type distribution. This two concepts (and in particular the Phase-type distribution) could be applied as a tool for evaluating the efficiency and the effectiveness of the utility restoration schemes.

The cluster analysis identified administrative units potentially and actually similar. Grounding the clustering on the estimated transition matrices we gathered information related to the potential transition dynamic between different operating status, while considering the underlying Markov chain we compared the working situation actually experienced by the systems.

Appendices

A The AEEG method

In the method applied by the AEEG for the third Regulatory period (Fumagalli *et al.* [11] and AEEG [1]), a 6 hour interval is deemed to be exceptional when the number of faults registered in that interval is larger than an *exceptionality threshold*; the procedure for the computation of these thresholds is briefly presented in the following.

First of all, for each administrative unit, the distribution of the number X of faults in 6 hours time-interval is clustered in two groups: that for Ordinary Intervals (OIs) and that for Exceptional Intervals (EIs). The threshold separating the two clusters is computed using a k -means algorithm with $k = 2$.

Hence the distribution for the OI number of faults is modeled with a geometric distribution with parameter $(1 - e^{-\lambda})$,

$$P(X = h) = (1 - e^{-\lambda})e^{-\lambda h}$$

where $h \in \{0, 1, \dots\}$. The parameter λ is estimated with the maximum likelihood estimator

$$\hat{\lambda} = -\log\left(\frac{\bar{\mu}_1}{1 + \bar{\mu}_1}\right),$$

where $\bar{\mu}_1$ is the mean of the observations of X smaller than the threshold separating the two groups (*i.e.* it is the mean of the observations in the first OI-cluster).

Finally a quantile q_α of the fitted geometric distribution for the OI number of faults is computed and an interval is declared exceptional for the utility under exam when the number of faults observed in the period is larger than q_α , which is therefore called the *exceptionality threshold*. In particular α is set so that, according to the fitted distribution for the OI number of faults, a value of X greater than q_α would be seen once every t years, with t large; for example suppose that an event is considered exceptional if it happens once every 20 years (*i.e.* $t = 20$), then $\alpha = 1 - \frac{1}{365 \times 4 \times 20} = 0.9999658$. Therefore, by the definition of quantile, the *exceptionality threshold* q_α is the minimum real value such that $P(X \leq q_\alpha) = \alpha$: for each utility

$$q_\alpha = \left\lceil -1 - \frac{\log(1 - \alpha)}{\hat{\lambda}} \right\rceil,$$

where $\lceil x \rceil$ indicates the integer part of x .

For each administrative unit we have the exceptionality threshold q_α and the average number m of faults per 6 hours period computed over the years 2004–2006; then, for ease of elicitation by the AEEG (see Fumagalli *et al.*, 2008 for details), q_α is expressed as a linear function of m

$$q_\alpha^* = \beta_0 + \beta_1 m$$

with $\beta_0 \neq 0$.

Given the EIs (where the number of faults occurred is greater than q_α^*) an exceptional period is defined considering 3 hours before the beginning of an EI and 3 hours after the end of the same interval. Then the methodology labels as an exceptional period a larger time span and in the simplest case (where there are no contiguous EIs) the exceptional period covers a period of 12 hours.

Note that in order to compare results obtained by the AEEG method and with the HMM, we compute the exceptionality threshold using the 6 hours time intervals data, but without considering the regression function; in fact estimating the HMM we did not consider information relative to the years 2005 and 2006.

B The MCMC algorithm

In Section 3.2 we introduced inference on the model parameters mentioning that sampling of the hidden chain is carried out by the the global updating scheme (Cappé *et al.* [4]). Global updating of the hidden chain is based on writing the joint posterior distribution $p(\mathbf{X}|\mathbf{y}, \boldsymbol{\vartheta})$ as

$$p(\mathbf{X}|\mathbf{y}, \boldsymbol{\vartheta}) = \prod_{t=1}^T p(X_t|X_{t-1}, \mathbf{y}, \boldsymbol{\vartheta}). \quad (9)$$

Now, let $\mathbf{y}_{h:k} = (Y_h = y_h, \dots, Y_k = y_k)$ with $h \leq k$; so

$$p(X_t = j|X_{t-1} = i, \mathbf{y}, \boldsymbol{\vartheta}) \propto p(y_t|X_t = j, \boldsymbol{\vartheta}) a_{i,j} p(\mathbf{y}_{t+1:T}|X_t = j, \boldsymbol{\vartheta})$$

where $p(\mathbf{y}_{t+1:T}|X_t = j, \boldsymbol{\vartheta})$ is the so called *backward variable* and it is the probability of the partial observation sequence from $t + 1$ to T , given state $X_t = j$ and the model parameters $\boldsymbol{\vartheta}$.

Let $p(\mathbf{y}_{t+1:T}|X_t = j, \boldsymbol{\vartheta}) = \beta_t(j)$ and solve inductively as follows:

a) Initialize with

$$\beta_T(j) = 1 \quad 1 \leq j \leq K$$

b) and for $t = T - 1, T - 2, \dots, 1, 1 \leq i \leq K$

$$\beta_t(i) = \sum_{j=1}^K a_{i,j} p(y_{t+1}|X_{t+1} = j) \beta_{t+1}(j). \quad (10)$$

When $p(y_{t+1}|X_{t+1} = j)$ is a density distribution, it is not necessarily bounded by 1. Then the backward variables may converge, at a geometric rate, to either zero or infinity. For this reason the introduction of a *scaling* factor is needed; then we scale $\beta_t(i)$ by multiplying each variable by a scale coefficient $\frac{1}{\sum_j \beta_t(j)}$, that depends only on t ; each scale factor effectively restores the magnitude of the $\beta_t(i)$ terms to 1.

The conditional distribution (9) becomes

$$p(\mathbf{X}|\mathbf{y}, \boldsymbol{\vartheta}) \propto \prod_{t=1}^T p(y_t|X_t = x_t, \boldsymbol{\vartheta}) a_{x_{t-1}, x_t} \beta_t(x_t) \quad (11)$$

and X_t , for $1 \leq t \leq T$, can be sampled from

$$Pr(X_t = j|X_{t-1} = x_{t-1}, \mathbf{y}, \boldsymbol{\vartheta}) = \frac{p(y_t|X_t = j, \boldsymbol{\vartheta}) a_{x_{t-1}, j} \beta_t(j)}{\sum_{i=1}^K p(y_t|X_t = i, \boldsymbol{\vartheta}) a_{x_{t-1}, i} \beta_t(i)} \quad (12)$$

where we recall that $X_0 = 1$.

Summarizing we can state the Gibbs sampling algorithm.

Algorithm 1 Gibbs sampling

Start with some state sequence $\mathbf{X}^{(0)}$ and repeat the following steps for $l = 1, \dots, L_0, \dots, L$.

1. Sample each row of A from the complete-data posterior distribution $p(A|\mathbf{X}^{(l-1)})$ in equation (3), and store the values.
 2. Sample the emission parameter from the complete-data posterior in equation (4), $p(B|\mathbf{y}, \mathbf{X}^{(l-1)})$ and store the values.
 3. Conditional of knowing the model parameters $\boldsymbol{\vartheta}^{(l)}$
 - a) Compute the means of the observable values in each state, $\boldsymbol{\mu}^{(l)} = (\mu_1^{(l)}, \mu_2^{(l)}, \dots, \mu_K^{(l)})$, with $\mu_i^{(l)} = \sum_{j=0}^q j \cdot b_i^{(l)}(j)$, and
 - if $\mu_1^{(l)} < \mu_2^{(l)} < \dots < \mu_K^{(l)}$ go to step b)
 - else order the mean vector $\rho(\boldsymbol{\mu})$ and $\rho(A)$, $\rho(B)$
 - b) Compute the scaled backward variables, like in equation (10)
 - c) Sample a path \mathbf{X} of the hidden Markov chain from the conditional posterior $p(\mathbf{X}|\boldsymbol{\vartheta}^{(l)}, \mathbf{y})$ in equation (12), through the global updating sampler and store all states.
 4. Increase l and return to step 1.
-

L is the number of samples and L_0 is the number of burn-in samples to be

discarded from the estimate. Step 3.a) is introduced for controlling for the label switching problem.

C The Phase-type distribution and the exceptional excursions

In order to analyse the length of an exceptional excursion, we consider the *Phase-type distribution*, that is the distribution of the number of steps from a Markov chain starting until absorption into absorbing state (for a complete and clear explanation see Neuts [16]).

Definitions C.1 Consider a discrete-time Markov chain with $m + 1$ states, where $m \geq 1$. The states $1, \dots, m$ are transient and state $m + 1$ is an absorbing state. The process has an initial probability of starting in any of the $m + 1$ phases given by the probability vector $(\boldsymbol{\alpha}, \alpha_{m+1})$.

This process can be written in the form of a transition probability matrix

$$P = \begin{pmatrix} T & T^o \\ 0 & 1 \end{pmatrix}$$

where T is an $m \times m$ substochastic matrix, $T^o + T\mathbf{e} = \mathbf{1}$ and \mathbf{e} is the column vector with all its components equal to one.

The distribution of the number of steps S until the process reaches the absorbing state is said to be discretely Phase-type distributed and it is represented by the pair $(\boldsymbol{\alpha}, T)$.

Moreover, the probability density $\{p_s\}$ of the Phase-type distribution is given by

$$\begin{aligned} p_0 &= \alpha_{m+1} \\ p_s &= \boldsymbol{\alpha} T^{s-1} T^o, \quad \text{for } s \geq 1. \end{aligned} \tag{13}$$

Proposition C.1 The length of an exceptional excursion is discretely Phase-type distributed.

Proof. Let the normal state 1 be the absorbing state and, without loss of generality, fix at 1 the starting time of the exceptional excursion; consider the two stopping times $E = \min\{t > 1 : X_t = 1\}$, that is the instant the chain reaches the absorbing state 1 and $H = \min\{t \geq 1 : X_t = 4\}$, that is the first time the chain enters in state 4. In order to prove the statement we need to verify that

$$P(X_1 = x_1, \dots, X_E = 1 | X_0 = 1, X_1 \neq 1, X_h = 4 \text{ for at least one } 0 < h < E) \tag{14}$$

has a Markov structure.

In the following we will write “ $X_h = 4, 0 < h < E$ ” to indicate “ $X_h = 4$ for at least one $0 < h < E$ ”. We have

$$\begin{aligned}
& P(X_1 = x_1, \dots, X_E = 1 | X_0 = 1, X_1 \neq 1, X_h = 4, 0 < h < E) \\
& \propto P(X_1 = x_1 | X_0 = 1, X_1 \neq 1) \\
& \cdot P(\mathbf{X}_{2:E} = \mathbf{x}_{2:E}, X_h = 4, 0 < h < E | X_1 = x_1, X_0 = 1, X_1 \neq 1) \\
& \propto \frac{a_{1,x_1}}{\sum_{i \in X \setminus \{1\}} a_{1,i}} \mathbf{1}_{X \setminus \{1\}}(x_1) \\
& \cdot P(\mathbf{X}_{2:E} = \mathbf{x}_{2:E}, X_h = 4, 0 < h < E | X_1 = x_1, X_0 = 1, X_1 \neq 1)
\end{aligned} \tag{15}$$

where $\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$ and $\mathbf{X}_{t:s} = (X_t, X_{t+1}, \dots, X_s)$, with $t < s$.

Consider the second term in equation (15),

$$\begin{aligned}
& P(\mathbf{X}_{2:E} = \mathbf{x}_{2:E}, X_h = 4, 0 < h < E | X_1 = x_1, X_0 = 1, X_1 \neq 1) = \\
& = P(\mathbf{X}_{2:H} = \mathbf{x}_{2:H}, \mathbf{X}_{H+1:E} = \mathbf{x}_{H+1:E}, H < E | X_1 = x_1, X_1 \neq 1) \\
& = P(\mathbf{X}_{2:H} = \mathbf{x}_{2:H}, H < E | X_1 = x_1, X_1 \neq 1) \\
& \cdot P(\mathbf{X}_{H+1:E} = \mathbf{x}_{H+1:E}, H < E | X_1 \neq 1, \mathbf{X}_{1:H} = \mathbf{x}_{1:H}) \\
& = P(\mathbf{X}_{2:H} = \mathbf{x}_{2:H}, H < E | X_1 = x_1, X_1 \neq 1) \prod_{i=H+1}^E a_{i-1,i}.
\end{aligned} \tag{16}$$

Consider the first term in equation (16); being $H < E$ and $H = \min\{t \geq 1 : X_t = 4\}$, we know that from time 1 to $H - 1$, in the exceptional excursion there are no 1 and no 4. Therefore equation (14) becomes

$$\begin{aligned}
& P(X_1 = x_1, \dots, X_E = 1 | X_0 = 1, X_1 \neq 1, X_h = 4 \text{ for at least one } 0 < h < E) \\
& = \frac{a_{1,x_1}}{\sum_{i \in X \setminus \{1\}} a_{1,i}} \mathbf{1}_{X \setminus \{1\}}(x_1) \prod_{l=2}^H \frac{a_{x_{l-1},x_l}}{\sum_{j \in X \setminus \{1\}} a_{x_{l-1},j}} \mathbf{1}_{X \setminus \{1\}}(x_l) \prod_{i=H+1}^E a_{i-1,i}
\end{aligned}$$

and this complete the proof. \square

We can now obtain the pair, say (α^*, T^*) , representing the Phase-type distribution, that describes the time needed to the system to reestablish the normal operating status. Consider a Markov chain with absorbing state 1 and transient states $\{2^*, 3^*, 2, 3, 4\}$, where 2^* and 3^* are states 2 and 3 conditioning on the fact that in following there is at least a 4; indicate the corresponding state space with X^* . Then the initial probability vector is

$$\alpha^* = \left(\frac{a_{1,2}}{a_{1,2} + a_{1,3} + a_{1,4}}, \frac{a_{1,3}}{a_{1,2} + a_{1,3} + a_{1,4}}, 0, 0, \frac{a_{1,4}}{a_{1,2} + a_{1,3} + a_{1,4}} \right)$$

and $\alpha_1 = 0$. The first element in α^* is, for example, the probability to have $X_1 = 2^*$, given that $X_0 = 1$ and $X_1 \neq 1$; null values are relative to states 2 and

3, in fact, given the definition of an exceptional excursion, it is not possible to have $X_1 = 2$ or $X_1 = 3$ (we can have $X_1 = 2^*$, $X_1 = 3^*$ or $X_1 = 4$).

Using similar considerations we obtain the following transition probability matrix

$$P^* = \left(\begin{array}{ccc|cc|c} \frac{a_{2,2}}{a_{2,2}+a_{2,3}+a_{2,4}} & \frac{a_{2,3}}{a_{2,2}+a_{2,3}+a_{2,4}} & 0 & 0 & \frac{a_{2,4}}{a_{2,2}+a_{2,3}+a_{2,4}} & 0 \\ \frac{a_{3,2}}{a_{3,2}+a_{3,3}+a_{3,4}} & \frac{a_{3,3}}{a_{3,2}+a_{3,3}+a_{3,4}} & 0 & 0 & \frac{a_{3,4}}{a_{3,2}+a_{3,3}+a_{3,4}} & 0 \\ 0 & 0 & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,1} \\ 0 & 0 & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,1} \\ 0 & 0 & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,1} \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

In order to obtain the estimated Phase-type distribution, shown in Figures 6 and 7, we compute, for any MCMC iteration, $l = 1, \dots, L$, $\{p_s^{(l)}\}$, $s \geq 1$, using (13), with $\alpha = \alpha^*$ and $T = T^*$; then the estimated probability density is $\{\hat{p}_s\}$, where $\forall s \geq 1$

$$\hat{p}_s = \frac{1}{L - L_0} \sum_{l=L_0}^L p_s^{(l)}.$$

References

- [1] AEEG, Autorità per l'energia elettrica e il gas, *Regulatory Order n. 333/07*, Available (in Italian) on the web site www.autorita.energia.it.
- [2] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Ann. of Math. Stat.*, vol. 37(6), pp. 1554–1563, 1966.
- [3] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Ann. of Math. Stat.*, vol. 41, pp. 164–171, 1970.
- [4] O. Cappé, E. Moulines and T. Rydén, *Inference in Hidden Markov Models*, Springer Series in Statistics, 2005.
- [5] CEER, *Third benchmarking report on quality of electricity supply*, Council of European Energy Regulators, Brussels, 2005.
- [6] R.D. Christie, "Statistical classification of major event days in distribution system reliability," *IEEE Trans. Power Del.*, vol. 18, n. 4, pp. 1336–1341, 2003.
- [7] S. Chib., "Calculating posterior distributions and modal estimates in Markov mixture models," *J. of Econometrics*, vol. 75, pp. 79–97, 1993.

- [8] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer Series in Statistics, 2006.
- [9] E. Fumagalli, L. Lo Schiavo, S. Salvati and P. Secchi, “Statistical identification of major event days: an application to continuity of supply regulation in Italy,” *IEEE Trans. Power Del.*, vol. 21, n. 2, 761–767, 2006.
- [10] E. Fumagalli, L. Lo Schiavo, and F. Delestre, *Service quality regulation in electricity distribution and retail*, Springer-Verlag, 2007.
- [11] E. Fumagalli, L. Lo Schiavo, A. Paganoni, and P. Secchi, “Statistical analyses of exceptional events: the Italian Regulatory experiences,” *IEEE Trans. on Power Del.*, 2008, forthcoming.
- [12] L. Kaufman and P.J. Rousseeuw, *Finding groups in data: an introduction to Cluster Analysis*. Wiley series in probability and mathematical statistics. John Wiley and Sons Inc, 1990.
- [13] E.L. Lehmann, *Nonparametrics: statistical methods based on ranks*, Originally published by Prentice-Hall, 1st ed. 1975. Revised edition 2006, Springer-Verlag.
- [14] I.L. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman&Hall, London, 1997.
- [15] M. Maechler, P.J. Rousseeuw, A. Struyf and M. Hubert, Cluster Analysis Basics and Extensions; unpublished
- [16] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*, Corrected reprint of the 1981 original, Dover Publications, Inc., New York, 1994.
- [17] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [18] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77(2), pp. 257–286, 1989.
- [19] M. Ramoni, P. Sebastiani, and P. Cohen, “Bayesian clustering by dynamics,” *Machine Learning*, vol. 47, pp. 91–121, 2002.
- [20] C.P. Robert, G. Celeux, and J. Diebold, “Bayesian estimation of hidden Markov chains: a stochastic implementation,” *Stat. and Probab. Lett.*, vol. 16, pp. 77-83, 1993.
- [21] C.P. Robert, T. Rydén, and M. Titterton, “Bayesian inference in hidden Markov models through reversible jump Markov chain Monte Carlo,” *J. of the Roy. Stat. Soc. Ser. B*, vol. 62, pp. 57–75, 2000.

- [22] P.J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. of Computational and Appl. Math.*, vol. 20, pp. 53–65, 1987.
- [23] S.L. Scott, "Bayesian methods for Hidden Markov Models: Recursive computing in the 21st century," *J. of the Amer. Stat. Assoc.*, 97, 337– 351, 2002.
- [24] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13(2), pp. 260–269, 1967.
- [25] C.A. Warren, J.D. Bouford, R.D. Christie, D. Kowalewski, J. McDaniel, R. Robinson, D.J. Schepers, J. Viglietta, and C. Williams, Classification of major event days. In proceeding of *Power Eng. Soc. General Meeting*, IEEE, vol. 1, pp. 466–471, 2003.

MOX Technical Reports, last issues

Dipartimento di Matematica “F. Brioschi”,
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 20/2009** N. ACCOTO, T. RYDÉN, P. SECCHI:
Bayesian hidden Markov models for performance-based regulation of continuity of electricity supply
- 19/2009** L. FORMAGGIA, A. MOLA, N. PAROLINI, M. PISCHIUTTA:
A three-dimensional model for the dynamics and hydrodynamics of rowing boats
- 18/2009** J. SILVA SOARES, P. ZUNINO:
A mathematical model for water uptake, degradation, erosion, and drug release from degradable polydisperse polymeric networks
- 17/2009** M. LONGONI, C. MAGISTRONI, P. RUFFO, G. SCROFANI:
3D Inverse and Direct Structural Modeling Workflow
- 16/2009** G. ALETTI, C. MAY, P. SECCHI:
A functional equation whose unknown is $P([0; 1])$ valued
- 15/2009** S. PEROTTO, A. ERN, A. VENEZIANI:
Hierarchical local model reduction for elliptic problems I: a domain decomposition approach
- 14/2009** L. BEIRÃO DA VEIGA, M. VERANI:
A posteriori boundary control for FEM approximation of elliptic eigenvalue problems
- 13/2009** E. MIGLIO, A. VILLA:
A mathematical derivation of compaction and basin models
- 12/2009** S. BADIA, A. QUAINI, A. QUARTERONI:
Coupling Biot and Navier-Stokes equations for modelling fluid-poroelastic media interaction
- 11/2009** L. FORMAGGIA, A. VILLA:
Implicit tracking for multi-fluid simulations