



HAL
open science

Dimensionality reduction of the resilience model of a critical infrastructure network by means of elementary effects sensitivity analysis

Pietro Turati, Nicola Pedroni, Enrico Zio

► To cite this version:

Pietro Turati, Nicola Pedroni, Enrico Zio. Dimensionality reduction of the resilience model of a critical infrastructure network by means of elementary effects sensitivity analysis. European Safety and RELiability Conference (ESREL) 2016, Sep 2016, Glasgow, United Kingdom. pp.2797-2804. hal-01436629

HAL Id: hal-01436629

<https://hal.science/hal-01436629>

Submitted on 16 Jan 2017

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.

Dimensionality reduction of the resilience model of a critical infrastructure network by means of elementary effects sensitivity analysis

P. Turati & N. Pedroni

Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), Laboratoire Genie Industriel, CentraleSupélec, Université Paris Saclay. Paris, France.

E. Zio

Energy Department, Politecnico di Milano, Milan, Italy.

Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), Laboratoire Genie Industriel, CentraleSupélec, Université Paris Saclay. Paris, France.

ABSTRACT: Modern Critical Infrastructures (CIs) are typically characterized by a large number of elements interconnected and interdependent. Their mathematical representation reflects these characteristics in models that typically turn out to be: 1) complex, since the relation between the variables can be nonlinear; 2) large, since a high number of variables is typically involved in the model; 3) dynamic, because the behavior of the system evolves in time. For this reason, the opportunities of exploring these models in order to extract information, such as identifying the most critical events, is conditioned by the computational cost of a simulation run and by the number of variables to explore. In this paper, we investigate the possibility of reducing the dimensionality of a model by identifying the variables that affect it most, by means of the Elementary Effects (EEs) method, which is a sensitivity analysis method capable of screening the input variables resorting to a limited number of model evaluations. Since the performance of the method relies on its settings, we analyze them proposing at the same time possible improvements. A hybrid network for gas and power distribution is considered as case study. The objective is to rank the importance of some uncertain parameters of the network (e.g., its failure and recovery characteristics) with respect to the system resilience properties (i.e., the capability of mitigating the effect of components failures and/or recovering its performance).

1 INTRODUCTION

Critical infrastructures (CIs) are essential in providing goods (such as energy and water) and services (such as banking and health care) for the welfare of modern society (Kröger and Zio, 2011). By means of a pervasive introduction of information technologies, CIs are becoming strongly automated and interconnected. Although these new technologies have increased the efficiency of CIs, they have also introduced new unexpected accident scenarios: a failure in one system can propagate and cause failure in a connected system, leading to cascade effects that can strike areas also very far from the impact zone.

Identifying and uncovering these emerging behaviors, especially the most critical, can increase the knowledge on the system, and help defining of effective preventive and mitigating actions for the CI (Bier et al., 1999, Paté-Cornell, 2012, Zio, 2016, Zio and Aven, 2013).

To this aim, mathematical models (and corresponding computer codes) are used for reproducing the response of the CI to various perturbations and stresses. Due to the complexity, and the level of interconnectedness and interaction of modern CIs, such

mathematical models can be: 1) complex, since the many relations between the variables can be nonlinear; 2) large, since a high number of variables and parameters are introduced in the model; 3) dynamic, because the structure and behavior of the system evolves in time. As a result, the computer codes that numerically translate the mathematical models are computationally demanding, requiring a significant amount of time to run a single simulation of CI response. In addition, since a large number of uncertain variables and parameters is contained in these models (e.g. parameters characterizing the CI structure, its failure and recovery dynamic characteristics, etc.), then a consistent amount of simulations is needed to explore the system state-space and obtain accurate and representative results of the system response behavior (Santner et al., 2003).

In this context of highly dimensional parameter and variable spaces, and associated computationally demanding codes, one possible approach for exploration is to resort to “guided” methods that intelligently search for conditions of interest (e.g. those that lead to the most severe outputs of the CI model) (Turati et al., 2016). However, the system response, including the emergence of critical conditions, is often most influenced by a limited subset of the model variables and parameters: identifying this subset and working

in the reduced state-space with the reduced model can significantly speed up the exploration process.

In general terms, dimensionality reduction includes a number of strategies for identifying a lower dimensional variables subspace where it is possible to build a reduced and simplified, yet representative and understandable, model of the system behavior (Pudil and Novovicova, 1998, Liu and Motoda, 2012, Fodor, 2002, Van der Maaten et al., 2009). Two main strategies have been proposed: *i*) feature selection, which aims at selecting a subset of the available variables and parameters, input to the model, and *ii*) feature extraction, which aims at identifying a subset of “new” features created by means of transformations of the initial ones. However, dimensionality reduction methods usually rely on a large set of input/output data examples, which is not typically available when the system model is computationally expensive to run.

In alternative, sensitivity analysis methods can be employed to achieve the same final objective as feature selection, by ranking the (input) variables and parameters according to their influence on the output of the model. Among several sensitivity analysis methods (Borgonovo and Plischke, 2016, Helton et al., 2006, Saltelli, 2008), we resort to the Elementary Effects (EEs) method (Morris, 1991), which is a qualitative screening method that has recently regained attention in light of *i*) its relatively low computational cost (comparable with local sensitivity analysis) and *ii*) its global interpretation of the results in identifying the inputs that most affect the model (Campolongo et al., 2007, Campolongo et al., 2011, Ge et al., 2014, Ge et al., 2015, Ge and Menendez, 2014, Ruano et al., 2011, Ruano et al., 2012, Saltelli et al., 2009, Sin et al., 2009, Janssen et al., 2012). However, in spite of the extensive research on the topic, a unified and general agreement on how to implement the method in practice has not been reached yet: *a*) different random sampling strategies have been proposed for the selection of the design points (i.e., the input configurations) on which the EEs have to be evaluated; *b*) different criteria for identifying the important inputs of the model have been defined.

A case study of a CI consisting of a hybrid gas and electric power transmission network is considered, in order to: *i*) show the potential of the above mentioned strategies, *ii*) draw some conclusions regarding their performance, *iii*) provide some guidelines for setting the sensitivity analysis methods from an engineering viewpoint and, ultimately, *iv*) evaluate the importance of some network parameters (e.g. the failure and the recovery characteristics of the components) with respect to the resilience of the network (Liu et al., 2016), to highlight which variables should be chosen for a possible guided exploration method. For the purpose of having a countercheck, the results have been compared to those obtained by a total order sensitivity index computed by a large number of simulations.

The rest of the paper is organized as follows. In Section 2, the main idea of the EEs is given. In Section 3, a review of the settings regarding the EEs method is provided. In Section 4, some improvements are proposed. In Section 5, a case study and the corresponding results are presented. In Section 6, some conclusions are drawn and prospective research challenges are suggested.

2 THE EEs METHOD

The EEs method allows quantitatively identifying the effect that a given input variable or parameter (hereafter also called factors) $X(i)$, $i = 1, \dots, K$ has on the model output Y . The method is based on the calculation of a number of incremental ratios, called Elementary Effects, on the basis of which some statistics are evaluated and used to characterize the influence of the different input variables and parameters (Campolongo et al., 2007, Morris, 1991). Considering a given configuration of the input vector \mathbf{X}_0 , the EE for the i -th input $X(i)$ is defined as follows:

$$d_i(X_0) = \frac{Y(X_0(1), \dots, X_0(i) + \Delta, \dots, X_0(K)) - Y(X_0)}{\Delta} \quad (1)$$

where Δ is a given variation on the input variable or parameter whose effect on the model output is being evaluated. Intuitively, the higher the value of $d_i(\mathbf{X})$, the more important the i -th input is (see the following Section 3.2 for details). A sequence of One-At-a-Time (OAT) input variable EEs calculations on random configurations of the input vector \mathbf{X} provides the method with global sensitivity analysis features.

The main advantages of the EEs method are: *a*) it demands a limited number of evaluations of the model, as compared to other global sensitivity measures; *b*) it is defined on a very simple criterion (incremental ratio), which is easy to compute and communicate to decision makers. For these characteristics, it particularly suits the analysis of models with a large number of inputs and/or computationally expensive to run.

3 SETTING THE EEs METHOD

In this section, we recall the current state of the art regarding the EEs method, with a particular focus on the settings that most affect its performance. For a detailed description the reader is referred to the references hereafter given.

3.1 Design Of the Experiment (DOE)

The values of the EEs (1) necessarily depend on the configurations of the input vector \mathbf{X} with respect to

which they are evaluated. A number of proposals of design of experiments have been made in the literature for selecting such input configurations. They share the common goals of spanning as much as possible the input space in order to have a representative population of EEs (global analysis) and increasing, at the same time, the ratio between the number of EEs available and the number of evaluations of the model to limit CPU cost. In the Trajectory-Based (*TB*) design, trajectories of $K + 1$ steps are intelligently constructed in the state space by varying at each step only one input at the time, allowing the evaluation of an EE for each factor in each trajectory (Campolongo et al., 2007) and keeping the trajectories as far as possible from the other (Ruano et al., 2012, Ge et al., 2014). In Cell-Based (*CB*) design, a rectangular scheme for the system response evaluation allows the assessment of an additional index measuring the effects of the interaction between pairs of inputs (Saltelli et al., 2009). Finally, in the Radial-Based (*RB*) design, a set of well distributed base point is firstly designed and, then, the incremental ratio are computed radially around each point (see Appendix for details).

3.2 Selection of the important factors

A criterion for the identification of the important input variables and parameters needs to be given. Since the number of important factors is unknown, the criterion for the selection of the important factors must not only rank the factors according to their effects on the model output, but also identify the proper number of them. In practice, the selection is based on statistics evaluated on the population of EEs. Three main statistics are used: the mean of the EEs values $\mu_i = \frac{1}{R} \sum_{r=1}^R d_i(X_r)$, the standard deviation $\sigma_i = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (d_i(X_r) - \mu_i)^2}$ and the mean of the modulus values $\mu_i^* = \frac{1}{R} \sum_{r=1}^R |d_i(X_r)|$. In what follows, two ranking criteria are reported.

3.2.1 Interval (Morris)

The idea is to consider a factor as important if either μ_i or σ_i has a large value (Morris, 1991). To do so, the respective values of μ_i and σ_i are plotted on the same graph. In addition, the interval defined by two times the estimated Standard Error of the Mean $SEM_i = \sigma_i / \sqrt{nR}$ is plotted. According to this plot, three regions can be identified (Figure 1): *i*) μ_i large and outside $\pm 2 \cdot SEM_i$, i.e. the *i*-th factor is important and probably presents a linear relation with the output (dark grey); *ii*) σ_i large and $\mu_i \in \pm 2 \cdot SEM_i$, i.e. the *i*-th input is important, but probably presents a nonlinear relation with the output, or it has some interactions

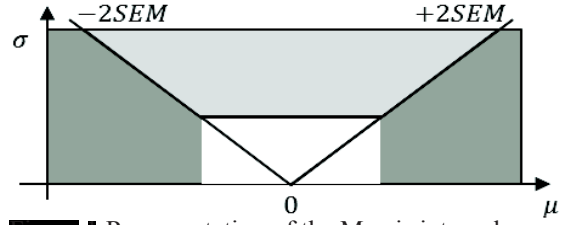


Figure 1 Representation of the Morris intervals.

with the other inputs (light grey). It could be very promising during the exploration process to span more in detail this variable, since it can have both a positive and a negative effect on the output. *iii*) μ_i and σ_i are small, i.e. the *i*-th input does not affect the model (white) and, thus, they may be of no interest for the exploration method because, by assigning to them a fixed value, we are not going to lose a lot of variability in the model output.

3.2.2 Average Of The Modulus (μ^*)

In (Campolongo et al., 2007), the authors show empirically that the ranking derived by using the mean of the modulus values of the Elementary Effects μ_i^* agrees with the ranking based on the robust variance-based total effect measure S_T (Homma and Saltelli, 1996). Thus, as a possible rule-of-thumb, the factor *i* has an important influence on the model if μ_i^* is larger than the average of μ^* , i.e. if $\mu_i^* > \frac{1}{K} \sum_{k=1}^K \mu_k^*$. However, the identification of the set of important factors can be conditioned by the presence of extremely important inputs, which can result in a misidentification of inputs that are important as well: as a consequence, a clustering method should be used for identifying different groups of inputs having possibly similar importance.

Finally, it must be mentioned that in some practical cases, the number of inputs to be selected as important is set a priori, e.g. because they are the inputs for another model or because of some specific constraints. In this case, the ranking based on the above explained statistics is enough for selecting the desired number of important inputs.

4 IMPROVEMENTS

Regarding the Radial-Based strategy, we have observed that if the approach is applied as described in Appendix (Campolongo et al., 2011), some of the EEs may be evaluated with a null step, since the skip factor *s* does not guarantee that Sobol' sequences associated to different inputs always get different values (Sobol, 1967, Sobol, 1976). Table 1 reports the maximum percent number of points \mathbf{a}_r involving at least one equal value in the corresponding point \mathbf{b}_{r+s} lead-

ing, thus, to a useless EE evaluation, for 1000 different starting points of Sobol' sequences of 500 numbers and 20 inputs. Different values of the skip factor $s = \{4, 16, 64, 256, 1024\}$ have been tested recurring to different matrices of dimensions $[500+s, 40]$. A high percentage value means a large inefficiency in the selection of the input configurations to be evaluated: the ideal result is zero, i.e. no repetitions.

Table 1 Maximum percentage of degenerate EEs due to null steps occurring in the Radial-Based strategy for different values of the skip factor s and for different starting points of the Sobol' sequences.

s	4	10	16	64	256	1024	2048
% of overlapping	15,2	12,4	16,4	6,4	6,4	3,2	1,6

In order to tackle this problem, two novel strategies are proposed in this paper: the first proposal (named, strategy 1) is to look for a Sobol' sequence (initial point) and a proper value of the skip factor s that guarantee no degenerate EEs, i.e. no overlap between \mathbf{a}_r and \mathbf{b}_{r+s} for a desired number of samples R and dimension K . This can be achieved through a trial-and-error strategy that tests several initial points for the Sobol' sequences and various skip factors. This verification is conducted without running the computationally demanding model, allowing the exploration of many configurations at a limited computational cost.

In alternative (named, strategy 2), a matrix of Quasi-Monte Carlo QMC (e.g. Sobol' sequences) values having K columns and $2R$ rows can be sampled. Indeed, the first R rows guarantee a uniform selection of the input points around which the screening analysis is developed (i.e., assuring a global sensitivity analysis); the last R rows guarantee the selection of other R values for each input, avoiding repetition. In order to guarantee that the selection of the design points is uniform in the input space (efficient global exploration), they have to be taken consecutively in the Sobol' sequences. Thus, if the number R of EEs is not known a priori a practical solution is to overestimate the requested number R_{over} of EEs and/or to set a maximum number R_{max} of EEs to be evaluated, given the computational resources available. In this case, the matrix of QMC samples has dimensions $(2 \cdot \min(R_{over}, R_{max}), K)$.

5 ANALYSIS OF THE RESILIENCE OF A HYBRID GAS AND ELECTRICITY DISTRIBUTION NETWORK

We consider a case study proposed in (Nozick et al., 2004) and further developed in (Ferrario et al., 2015, Liu et al., 2016). It concerns a hybrid network Figure 2 involving gas distribution (solid lines) and a power grid (dash-dotted lines), with real time flux optimization provided by a Model Predictive Control.

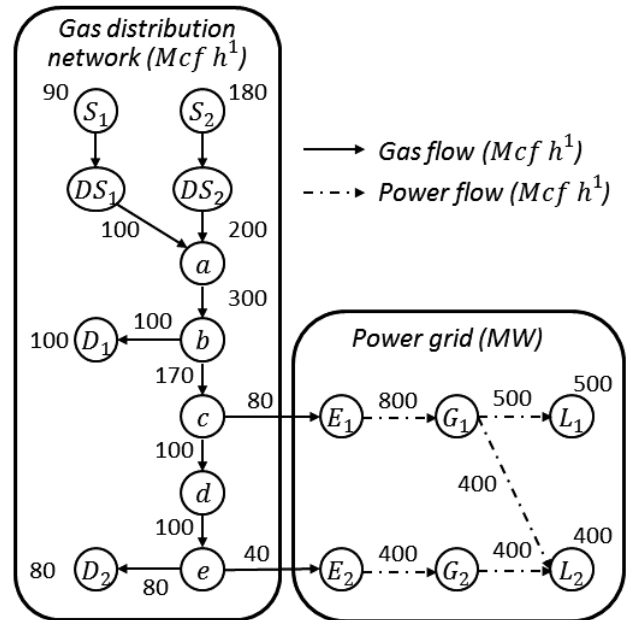


Figure 2 Integrated gas distribution and power network.

The gas network is composed of two gas suppliers, S_1 and S_2 ; two buffers (gas reservoirs), DS_1 and DS_2 ; five transporters a , b , c , d and e ; and two users D_1 and D_2 . The electric power network has two converters (electric power generators), E_1 and E_2 , with a constant conversion coefficient $\beta = 10\text{MWh}/1\text{Mcf}$; two transporters, G_1 and G_2 , and two users L_1 and L_2 . The objective of the system is to provide the necessary amount of gas, D_{D1} and D_{D2} , and electricity, D_{L1} and D_{L2} , to the demand nodes. We refer the readers to (Liu et al., 2016) for further details concerning the model.

The main objective of the analysis is to study the resilience of the system, i.e. the capability of mitigating the consequences of failures (hereafter called mitigation resilience) and of recovering system performance upon having taken proper (repair) actions (hereafter called recovery resilience). We refer here to the definitions given in (Liu et al., 2016). These can enable us to understand which factors most characterize the system behavior in failure scenarios and that, thus, should be the objects of a thorough exploration. However, for the sake of brevity, we report in this paper only the analysis concerning the recovery resilience.

Table 2 Set of the factors consider in the network failure.

Factor	Interval	Factor	Interval
H_r	[0 30]h	H_h	[50 120]h
X_{b1}	[0 3000] Mcf	X_{b2}	[0 6000] Mcf
$F_1: S_1$	[0 90] Mcf h-1	m_1	[0 1.8] Mcf h-2
$F_2: S_2$	[0 180] Mcf h-1	m_2	[0 3.6] Mcf h-2
$F_3: a \rightarrow b$	[0 300] Mcf h-1	m_3	[0 6] Mcf h-2
$F_4: b \rightarrow c$	[0 170] Mcf h-1	m_4	[0 3.4] Mcf h-2
$F_5: c \rightarrow d$	[0 100] Mcf h-1	m_5	[0 2] Mcf h-2
$F_6: d \rightarrow e$	[0 100] Mcf h-1	m_6	[0 2] Mcf h-2
$F_7: E_1 \rightarrow G_1$	[0 800] MW	m_7	[0 8] MW h-1
$F_8: E_2 \rightarrow G_2$	[0 400] MW	m_8	[0 4] MW h-1

The factors affecting system resilience are both system parameters and the variables representing the magnitudes of the component failures characterizing the failure scenarios. In details, they are: the response time H_r , i.e. the time needed to discover the failure and deploy the proper repair actions; the time horizon H_h : from (Ouyang et al., 2012), we assume that a CI can recover to its nominal operational state after 2-5 days from the occurrence of a failure; the initial storage of the buffer X_{b1} and X_{b2} : we assume that the buffers with initial storage at their highest level are capable of substituting the suppliers in servicing the system during the highest response time (i.e., 30 hours); the magnitude of failure of a generic node or link i , F_i , which is limited by its maximum capacity; the recovery rate of a failed component m_i . The ranges of variability of these factors are synthesized in Table 2.

5.1 Recovery resilience

Recovery Resilience, R_r , measures the rapidity of the system in recovering once the failure has been identified and proper repair actions are taken. The analytical expression taken from (Liu et al., 2016) is reported in what follows:

$$R_r = \frac{\sum_{i \in D} \int_{H_r}^{H_h} \omega_{conv}^i \cdot X_i(\tau) d\tau}{\sum_{i \in D} \int_{H_r}^{H_h} \omega_{conv}^i \cdot D_i(\tau) d\tau} \quad (2)$$

Since the case study tackles a hybrid network, the two flows (gas and electricity) have to be converted to a unique unit of measure, to allow comparison. To this aim, a conversion coefficient ω_{conv}^i (1 if X_i refers to gas flow, 0.1 if X_i refers to power flow) has been introduced for each user. The computational time for running a simulation of the model on a personal computer (processor Intel(R) Core(TM) i5-4300 CPU @ 1.90GHz 2.50 GHz) is around 8 sec.

5.2 Sensitivity analysis

A sensitivity analysis based on the EEs has been conducted for identifying the variables and parameters that most affect R_r . All three strategies for imple-

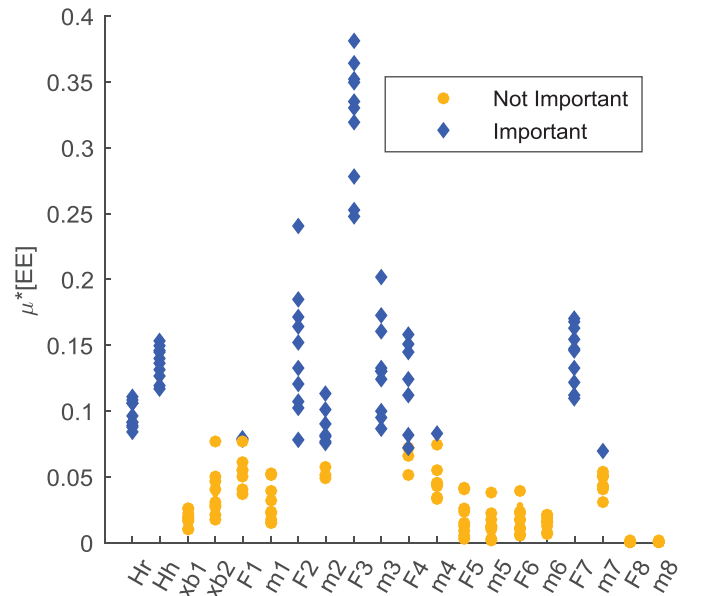


Figure 3 Values of μ^* for all inputs for 10 experiments. The dark marks identify the inputs having μ^* larger than the average within the experiment.

menting the EEs method (Section 3.1) have been applied. Only the plots of the results by the *RB* strategy are reported, for the sake of brevity. For each method both selection criteria of Section 3.2 have been utilized and compared. The computational resources have been set for the different EEs strategies as similar as possible by resorting to similar numbers of model simulations, in order to allow a comparison between them. A fixed number of EEs is evaluated for the purpose of controlling the computational cost and time. In this respect, the number of EEs has been fixed according to the optimal criterion proposed for the *CB* approach in (Saltelli et al., 2009), leading to 376 model runs and 16 EEs evaluations. For the other DOEs a similar number of model runs (i.e., 378) has been considered, leading to the evaluation of 18 EEs for each input. The experiments and the corresponding assessments of the indicators for the selection of the important factors have been repeated 10 times, in order to account for the variability of the DOEs and explore the robustness of the results.

By looking at Figure 3, it can be noticed that the factor that most affects R_r is the failure F_3 of link $a \rightarrow b$, followed by the factors $H_r, H_h, F_2, m_3, F_4, F_7$. Finally, the factors F_1, m_2, m_4, m_7 are identified as important only in some experiments: this means that

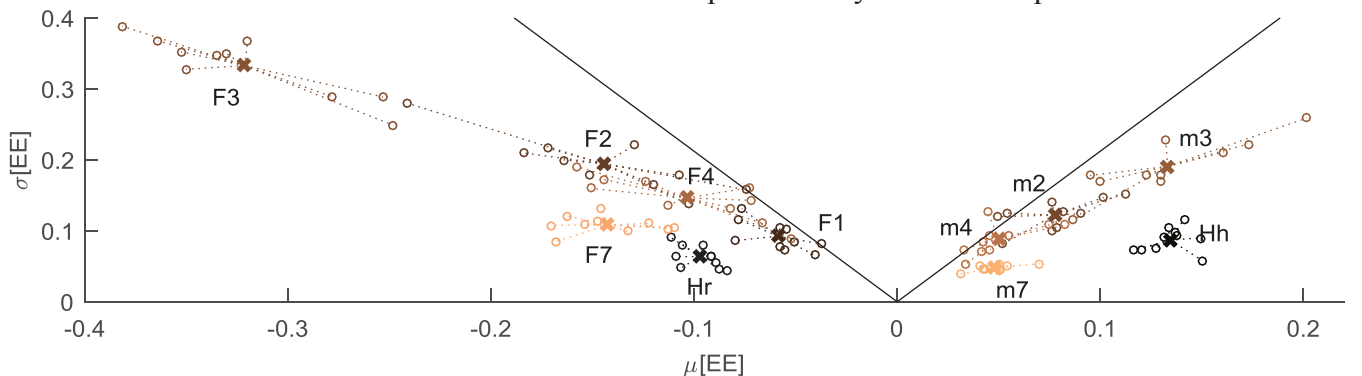


Figure 4 Morris representation of the statistics μ and σ . Bold crosses represent the average value of the two statistics over 10 experiments.

they have a lower impact on R_r than the parameters listed above, but they can still be taken into account in the modeling and in the corresponding exploration methods, according to the level of detail required by the analysis and depending on the associated computational constraint. Moreover, it can be observed that some recovery rates are important, especially those associated to important failures, suggesting that in the exploration of possible critical scenarios, the recovery rate of the most critical links must be considered as an influential variable to furtherly check.

Figure 4 shows the values of the statistics μ and σ as for the Morris criterion. According to the values taken by μ , it can be seen that all the failures F_i and the response time H_r have a negative impact on R_r , being $\mu_i < 0$; indeed, the more severe the failures and the later in time their detections, the larger the losses. On the other side, the recovery rates m_i as well as the time horizon H_h have a positive impact on R_r . Indeed, on one side, the larger the recovery rate, the faster the system returns to a normal condition; on the other side, the longer the time horizon, the longer the time available to recover.

The results obtained by the EEs method have been compared with the results obtained from the evaluation of the total order sensitivity S_T values, which account for the total effects of a factor, including all its possible interactions with other factors (Homma and Saltelli, 1996).

$$S_{Ti} = \frac{E_{X_{-i}} [V_{X_i} (Y|X_{-i})]}{V(Y)} \quad (3)$$

where the numerator is estimated as in (Campolongo et al. 2011) with the same design matrices used for the design of experiment as in Appendix (Saltelli et al., 2010). In order to achieve a satisfactory conver-

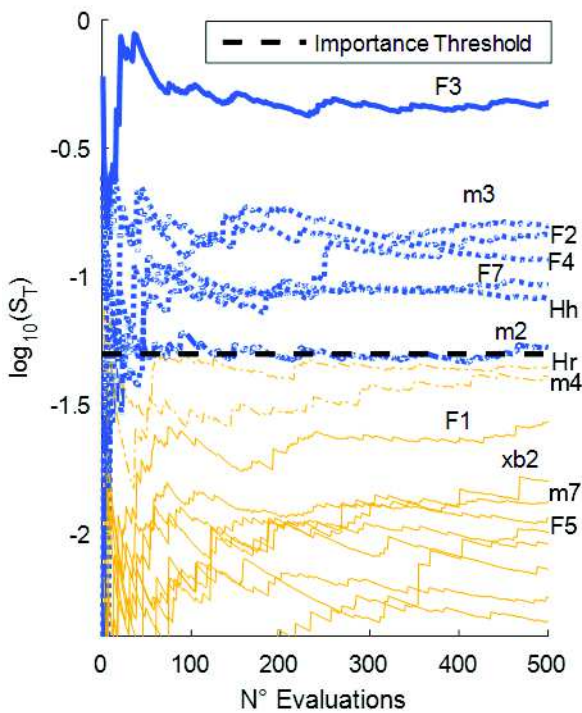


Figure 6 Convergence plots for the total order sensitivity indices. Factors having $S_{Ti} > 1/K$ are represented in bold.

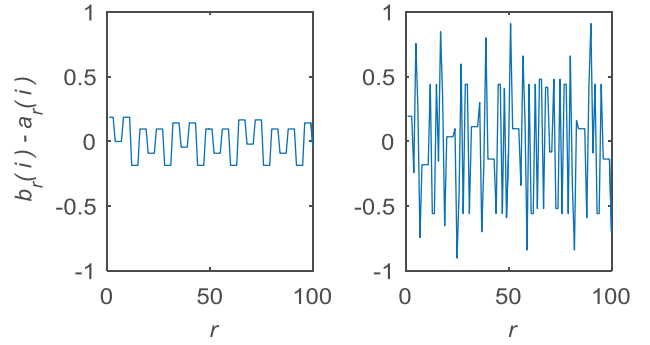


Figure 5 Length of the r steps utilized to evaluate S_{Ti} , according to strategy 2 for factors H_r and H_h .

gence of the estimated S_{Ti} , we set $R = 500$, considering as important factor having $S_{Ti} > 1/K$, similarly to the criterion described in Section 3.2.2.

Regarding the improved sampling strategies proposed in Section 4 for the selection of a_r and b_r , both have been implemented. In Figure 5, the convergence of the total order sensitivity indices is verified (at least for the most important factors), and the results are plotted in Log_{10} scale. It can be stated that apart from F_3 , three “clusters” can be identified: *i*) m_3, F_2, F_4, F_7, H_h ; *ii*) m_2, H_r, m_4 ; *iii*) the remaining factors. According to the criterion utilized, all the factors down to m_2 are identified as important (bold dark lines). However, by looking at Figure 5, it is more reasonable either to consider only F_3 and the factors of cluster *i*) or to add also all the factors of cluster *ii*), since they have similar importance among them.

The results regarding strategy 2, here not reported, are affected by a bias in the estimate of one of the total order sensitivity indices S_{Ti} . An accurate analysis of the bias has shown that it is due to possible dependences between a_r and b_r that are introduced in the sampling strategy when the sequel of the same Sobol’ sequence of $a_r(i)$ is used for the selection of $b_r(i)$. This dependence implies a limited extension of the R steps defined by the difference $b_r(i) - a_r(i)$, that, in turn, could lead to a biased estimate of S_{Ti} . Figure 6 reports the length of a fraction of the steps proposed by strategy 2 for H_r (left), which has a limited extension, and for H_h (right) whose steps, instead, are correctly distributed over the entire domain. Nevertheless, it must be pointed out that this drawback is relevant only if we are interested in evaluating the total order sensitivity index, while it does not affect the estimate of the EEs, since they are normalized by the length of the step during the evaluation (Eq. 1).

To conclude, we compare the factors that have been identified as important by the EEs method according to the criterion proposed in Section 3.2.2 to those that have been selected by the total order sensitivity indices. Table 3 reports the results obtained in 10 runs with the all three designs of experiments discussed in Section 3.1 and those produced by the total order sensitivity analysis. The factors have been ranked according to the value of S_T , which is taken as

the “reference index” for the importance of the factors. All the EEs designs are capable of identifying the most influential factors. However, it must be pointed out that in almost all the experiments, all the EEs methods identify H_r as important, even if according to S_{Ti} it is less important than other factors. This might be related to the fact that according to Figure 4, H_r has a lower value of σ , i.e. the EEs of H_r are less dispersed than those of F_4 and m_2 ; in other words, the EEs of H_r depend less on the configurations selected by the design of the experiments. This could justify why H_r has been identified in all the experiments as important, whereas F_4 and m_2 have not.

Table 3 Synthesis results produced by *TB*, *CB* and *RB* EEs methods for 10 experiments. The fraction of times that a factor is identified as important is reported. The final row represents the total order sensitivity indices S_{Ti} for the same factors.

	F_3	m_3	F_2	F_4	F_7	H_h	m_2	H_r	m_4	F_1	x_{b2}	m_7	F_5
<i>TB</i>	1	0.8	0.8	0.8	1	1	0.4	0.9	0.3	0.5	0	0	0
<i>CB</i>	1	0.9	0.9	0.8	1	1	0.5	1	0.2	0.2	0	0	0
<i>RB</i>	1	1	1	0.7	1	1	0.7	1	0.1	0.2	0	0.1	0
$ST[10^{-2}]$	47	16	15	12	9	8	5	4	4	3	2	1	1

In accordance with the analysis performed, we might claim that the EEs method is a practicable means for retrieving knowledge and information concerning the model under analysis. In particular, it provides insight on the model factors that must be analyzed with higher attention. In contrast to the total order sensitivity analysis, the EEs method not only identifies the factors that most affect the model, but it also provides insights on the relation between each factor and the output. This information can be extremely useful in the context of accident scenario exploration. For example, if we are looking for the worst conditions the system could undergo in terms of resilience, it is reasonable to focus the exploration on those factors having a large negative effect on R_r (e.g. the failure magnitudes F_3, F_2, F_7, F_4 and the response time H_r) and to set to small values those that have a positive effect (e.g. the recovery rates m_3, m_2, m_4, m_7 and the time horizon H_h).

6 CONCLUSIONS

In the present paper, we have addressed the problem of identifying, by means of sensitivity analysis, the factors that most affect the output of a CI model, providing insights for a proper dimensionality reduction and for a further exploration of the accident scenarios. In particular, we have analyzed the performance of the EEs method with respect to its parameters settings, by application to a CI model of a hybrid network involving a gas distribution and a power grid.

We suggest that the *RB* design of the experiments be employed, since it provides more robust results

than the other designs and, at the same time, it easily allows incrementing the number of EEs evaluations, whenever more accurate results are needed. In this respect, the strategies proposed in this paper aid the *RB* EEs method to avoid wasting computational resources in the simulation of configurations that do not provide any additional information. Also, while μ^* provides a synthetic and intuitive result of the importance of a factor, the Morris representation provides some additional information regarding the possible input-output relations (e.g. linear, nonlinear, dependencies, etc.), and about the positive or negative effects of a factor on the output.

Finally, with respect to the CI analyzed we observed that half of the factors have been identified as important and among them a number of factors seem to have a nonlinear relation with the output, which reinforces the interest in further exploring their role in the model in order to find the most critical system configurations.

7 REFERENCE

- BIER, V. M., HAIMES, Y. Y., LAMBERT, J. H., MATALAS, N. C. & ZIMMERMAN, R. 1999. A survey of approaches for assessing and managing the risk of extremes. *Risk analysis*, 19, 83-94.
- BORGONOVO, E. & PLISCHKE, E. 2016. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248, 869-887.
- CAMPOLONGO, F., CARIBONI, J. & SALTELLI, A. 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22, 1509-1518.
- CAMPOLONGO, F., SALTELLI, A. & CARIBONI, J. 2011. From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182, 978-988.
- FERRARIO, E., PEDRONI, N. & ZIO, E. 2015. Analysis of the robustness and recovery of critical infrastructures by Goal Tree Success Tree-Dynamic Master Logic Diagram, within a multi-state system-of-systems framework, in the presence of epistemic uncertainty. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 1, 14.
- FODOR, I. 2002. A Survey of Dimension Reduction Techniques.
- GE, Q., CIUFFO, B. & MENENDEZ, M. 2014. An Exploratory Study of Two Efficient Approaches for the Sensitivity Analysis of Computationally Expensive Traffic Simulation Models. *Ieee Transactions on Intelligent Transportation Systems*, 15, 1288-1297.
- GE, Q., CIUFFO, B. & MENENDEZ, M. 2015. Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. *Reliability Engineering & System Safety*, 134, 334-344.
- GE, Q. & MENENDEZ, M. 2014. An efficient sensitivity analysis approach for computationally expensive microscopic traffic simulation models. *International Journal of Transportation*, 2, 49-64.
- HELTON, J. C., JOHNSON, J. D., SALLABERRY, C. J. & STORLIE, C. B. 2006. Survey of sampling-based methods

- for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91, 1175-1209.
- HOMMA, T. & SALTELLI, A. 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52, 1-17.
- JANSSEN, P., BARKMEIJER, J., ALDBERG, T. & BEUSEN, A. H. W. 2012. Methods for sensitivity and uncertainty analysis of computer intensive simulation models. 7 ed.: NMDC.
- KRÖGER, W. & ZIO, E. 2011. *Vulnerable systems*, Springer Science & Business Media.
- LIU, H. & MOTODA, H. 2012. *Feature selection for knowledge discovery and data mining*, Springer Science & Business Media.
- LIU, X., FERRARIO, E. & ZIO, E. 2016. Resilience Assessment of Interconnected Critical Infrastructures within a Control-Based Modeling Framework. *IEEE Transactions on Smart Grid, Special Issue on Resilience*.
- MORRIS, M. D. 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33, 161-174.
- NOZICK, L. K., TURNQUIST, M., JONES, D., DAVIS, J. R. & LAWTON, C. R. Assessing the performance of interdependent infrastructures and optimizing investments. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, 2004. IEEE*, 7 pp.
- OUYANG, M., DUENAS-OSORIO, L. & MIN, X. 2012. A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety*, 36-37, 23-31.
- PATÉ-CORNELL, E. 2012. On "Black Swans"• and "Perfect Storms": risk analysis and management when statistics are not enough. *Risk Analysis*, 32, 1823-1833.
- PUDIL, P. & NOVOMICOVA, J. 1998. Novel methods for feature subset selection with respect to problem knowledge. *Feature Extraction, Construction and Selection*. Springer.
- RUANO, M. V., RIBES, J., FERRER, J. & SIN, G. 2011. Application of the Morris method for screening the influential parameters of fuzzy controllers applied to wastewater treatment plants. *Water Science and Technology*, 63, 2199-2206.
- RUANO, M. V., RIBES, J., SECO, A. & FERRER, J. 2012. An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors. *Environmental Modelling & Software*, 37, 103-109.
- SALTELLI, A. 2008. *Global sensitivity analysis : the primer*, Chichester, England ; Hoboken, NJ, John Wiley.
- SALTELLI, A., ANNONI, P., AZZINI, I., CAMPOLONGO, F., RATTO, M. & TARANTOLA, S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181, 259-270.
- SALTELLI, A., CAMPOLONGO, F. & CARIBONI, J. 2009. Screening important inputs in models with strong interaction properties. *Reliability Engineering & System Safety*, 94, 1149-1155.
- SANTNER, T. J., WILLIAMS, B. J. & NOTZ, W. 2003. *The Design and analysis of computer experiments*, New York, Springer.
- SIN, G., GERNAEY, K. V. & LANTZ, A. E. 2009. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. *Biotechnology Progress*, 25, 1043-1053.
- SOBOL, I. M. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational mathematics and mathematical physics*, 8(6)112.
- SOBOL, I. M. 1976. Uniformly distributed sequences with an additional uniform property. *USSR Computational mathematics and mathematical physics*, 16, 236-242.
- TURATI, P., PEDRONI, N. & ZIO, E. 2016. An Adaptive Simulation Framework for the Exploration of Extreme and Unexpected Events in Dynamic Engineered Systems. *Risk Anal.*
- VAN DER MAATEN, L. J., POSTMA, E. O. & VAN DEN HERIK, H. J. 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10, 66-71.
- ZIO, E. 2016. Critical Infrastructure Vulnerability and Risk Analysis. *European Journal for Security Research*.
- ZIO, E. & AVEN, T. 2013. Industrial disasters: Extreme events, extremely rare. Some reflections on the treatment of uncertainties in the assessment of the associated risks. *Process Safety and Environmental Protection*, 91, 31-45.

APPENDIX

The radial-based strategy takes this name from the fact that the incremental ratios are evaluated radially, one input at the time, around a chosen base point (Campolongo et al., 2011). For each base point, one EE is assessed for each input. The desired cardinality of EEs, R , is achieved by choosing an equivalent number of base points. Quasi-Monte Carlo sequences (e.g. quasi-random Sobol' numbers (Sobol, 1967, Sobol, 1976)) have been used for sampling the base points (Campolongo et al., 2011) and the corresponding radial movements necessary to assess the incremental ratios, because of their ability in spanning the input state space (K -dimensional hypercube). In details, the sampling strategy utilizes a matrix of $(R + s, 2K)$ quasi-random Sobol numbers, where the left half \mathbf{a} of the matrix represents the base points and the right half \mathbf{b} represents the corresponding radial movements. In order to guarantee that the movements are not null around the base points, a skip factor s is introduced (e.g. $s = 4$ in the referenced paper), see Figure 7 for a visual explanation.

Given the r -th base point \mathbf{a}_r , the i -th EE is evaluated as follows:

$$d_i(\mathbf{a}_r) = \frac{Y(\mathbf{a}_r(1), \dots, \mathbf{a}_r(i-1), \mathbf{b}_{r+s}(i), \mathbf{a}_r(i+1), \dots, \mathbf{a}_r(K)) - Y(\mathbf{a}_r)}{\mathbf{b}_{r+s}(i) - \mathbf{a}_r(i)} \quad (I)$$

where the movement is obtained by changing the i -th component of the base point with the i -th component of the corresponding row \mathbf{b}_{r+s} of the matrix \mathbf{b} . It must be noticed that, in contrast with other design strategies, the radial-based one uses non-fixed variations of the inputs to evaluate the incremental ratios, being $\mathbf{b}_{r+s}(i) - \mathbf{a}_r(i)$ not constant.

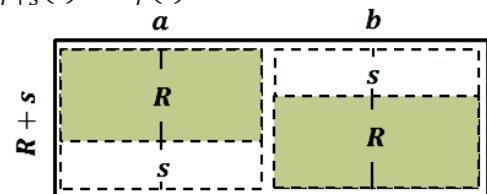


Figure 7 Visual representation of the design matrices.