An advanced platform for power system security assessment accounting for forecast uncertainties

Emanuele Ciapessoni, Diego Cirio and Andrea Pitto*

Ricerca sul Sistema Energetico RSE S.p.A., Italy
Email: emanuele.ciapessoni@rse-web.it
Email: diego.cirio@rse-web.it
Email: andrea.pitto@rse-web.it
*Corresponding author

Nicolas Omont

RTE – Réseau de Transport d'Electricité, France
Email: nicolas.omont@rte-france.com

Leonel M. Carvalho and Maria Helena Vasconcelos

Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias 4200-465 Porto, Portugal
Email: leonel.m.carvalho@inesctec.pt
Email: mhv@fe.up.pt

Abstract: Accounting for the increasing uncertainties related to forecast of renewables is becoming an essential requirement while assessing the security of future power system scenarios. Project iTesla in the Seventh Framework Program (FP7) of the European Union (EU) tackles these needs and reaches several major objectives, including the development of a security platform architecture. In particular, the platform implements a stochastic dependence model to simulate a reasonable cloud of plausible 'future' states – due to renewable forecast – around the expected state, and evaluates the security on relevant states after sampling the cloud of uncertainty. The paper focuses on the proposed model for the uncertainty and its exploitation in power system security assessment process and it reports the relevant validation results.

Keywords: Nataf transformation; pair copula decomposition; power systems; principal component analysis; renewables; security assessment; uncertainties.

Reference to this paper should be made as follows: Ciapessoni, E., Cirio, D., Pitto, A., Omont, N., Carvalho, L.M. and Vasconcelos, M.H. (2019) 'An advanced platform for power system security assessment accounting for forecast uncertainties', *Int. J. Management and Decision Making*, Vol. 18, No. 3, pp.257–281.

Biographical notes: Emanuele Ciapessoni received his MSc in Physics and PhD in Computer Science from the Milan University, Italy. He is working in the RSE S.p.A., where he is member of the Scientific Committee. His main interest is power system resilience.

Diego Cirio received his MSc and PhD in Electrical Engineering from the Genoa University, Italy. He is working in the RSE S.p.A., where he leads the Grid Development and Security Research Group. He was a WP Leader of the iTesla.

Andrea Pitto received her MSc in 2005 and PhD in 2009 in Electrical Engineering from the Genoa University, Italy. He joined in the RSE in 2011. His interests are probabilistic approaches to power system security, and HVDC modelling.

Nicolas Omont joined in the RTE as the French Transmission System Operator in 2012. He obtained his Engineering from the Ecole Polytechnique and a PhD in Computer Sciences in 2008. He develops tools based on statistics and mathematical optimisation techniques to improve power system operation.

Leonel M. Carvalho received his BSc in 2006, MSc in 2008, and PhD in 2013 in Electrical Engineering from the University of Porto (FEUP), Portugal. His current research interests include power system reliability assessment and the application of computational intelligence to power systems.

Maria Helena Vasconcelos obtained her MSc in 1999 and PhD in 2008 in Electrical Engineering from the Engineering Faculty of Porto University (FEUP), Portugal. She worked in the INESC TEC (former INESC Porto) since 1996 and at the FEUP since 2008. Her topics are dynamic studies, and security analysis studies

This paper is a revised and expanded version of a paper entitled ‘Managing forecast uncertainty in power system security assessment’ presented at CODIT (Conference on Control, Decision and Information Technologies) 2017, Barcelona, 5–7 April 2017.

1 Introduction

Security issues of the pan-European electricity transmission system are likely to become more and more challenging in the coming years due to the growing contribution of less predictable and intermittent RESs, the introduction of new controllable devices such as HVDC lines, a partially controllable electricity demand, the increasing difficulty to build new overhead transmission lines, and the progressive construction of a single European electricity market where transnational power exchanges will naturally increase.

These new constraints but also new opportunities will result in more complex system operation, a grid working closer to its operational limits and, therefore, the need for a major revision of operational rules and procedures. In this context, it is clear that current tools for security assessment will no longer be suitable for network operators to take the right decisions and that a new generation of tools is needed: in particular, the increasing penetration of Renewable Energy Sources (RES) and the constraints posed by pan-European market make more and more crucial the need to perform security assessment of the whole grid considering forecast uncertainties from operational planning to online environment (CIGRE WG C4-601, 2010; Panciatici et al., 2010). Furthermore, Regional Security Coordination Initiatives (RSCI) have already emerged for different regions of the pan-European transmission system (e.g., CORESO and TSC). Currently in

Europe, these initiatives perform security analyses over time horizons of two days ahead, one day ahead and close to operation. Generally, the analyses consist of the verification of the 'N-1' rule and are based on the deterministic forecasting of the grid state. The forecasts are prepared by the individual transmission system operators (TSO) and merged into one common grid forecast. However, these coordination initiatives will not be fully efficient without a new toolbox, allowing the different TSOs to increase coordination.

This rationale triggered the developments of the iTesla project (iTesla Consortium, 2014; Vasconcelos et al., 2016) co-funded by the European Commission Framework Program 7 (EC FP7). The project targeted the development of an online dynamic security analysis platform for European-wide grid models, able to account for uncertainty in the security margins evaluation and to handle curative remedial actions to face contingencies.

This paper reports some of the key aspects developed within the project to address uncertainties in the security assessment analyses and it describes the main results of the validation process aimed to check the performances of the proposed uncertainty model and of the security assessment functions.

In particular, the paper is organised as follows: Section 2 presents the iTesla platform architecture, the model adopted for load and renewable forecast uncertainties and the adopted security assessment functions. Sections 3 and 4 respectively describe the validation process for uncertainty model and for security rules. Section 5 briefly describes the case studies adopted. Sections 6 and 7 present some results of the validation of the platform on real world datasets. In Section 8 some conclusions are drawn.

2 Architecture of iTesla platform

The iTesla platform is an open source and interoperable toolset designed to support the security assessment of forecast situations from several hours ahead of operation up to near real-time. Recommendations are also provided in terms of efficiency of curative remedial actions. Evaluations account for the uncertainties affecting power injections, such as non-programmable RES and loads, and the dynamic behaviour of the grid. To this aim, computations are performed via two complementary workflows, namely the offline and the online workflows (see Figure 1), and a filtering approach that exploits machine learning techniques.

The offline workflow builds

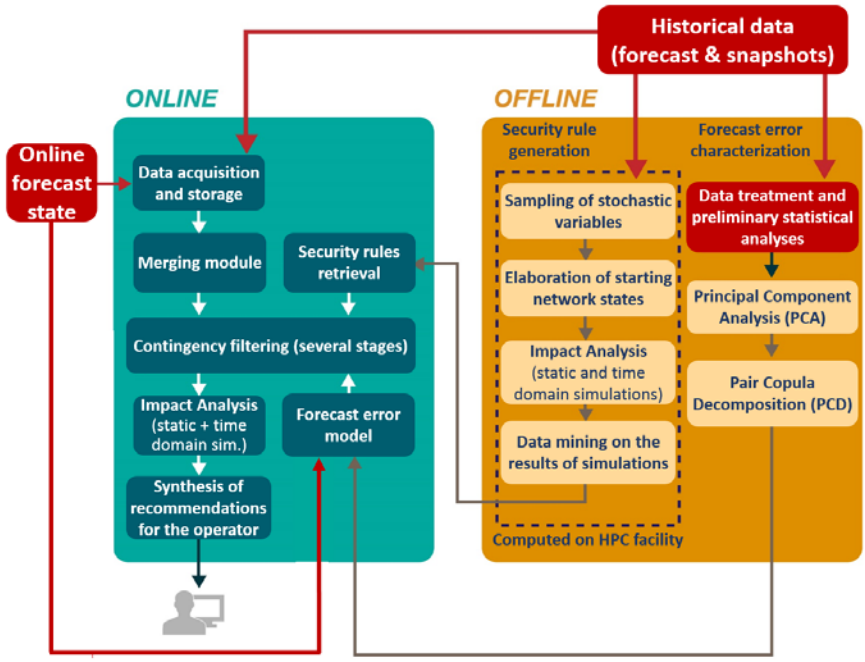
1 'Security rules'.

2 Uncertainty models for use in the online workflow: the security rules are simple logical expressions based on the pre-contingency quantities (active and reactive powers) that predict if the post-contingency state presents any violations of the operating constraints.

The rules are used in the online environment in order to speed up the analysis process. In particular, the rules are applied to plausible states in the 'uncertainty domain' of the forecast under analysis, to identify the contingencies for which control actions are needed while limiting the number of accurate network simulations to be performed online. Both workflows include different computation modules, each fulfilling a specific technical function such as power flow computation or time-domain simulation. The offline workflow is expected to be executed periodically in order to update the security rules

with the latest information concerning possible system topological changes, general weather patterns, etc. To achieve this goal, the offline platform is based on the following steps: anticipation, analysis, and classification.

Figure 1 The online and offline workflows of iTesla platform (see online version for colours)



The first step is to build a large population of operating points that are *anticipated* to occur in the next period, by modelling the variability related to RES and loads. Historical measurements from similar periods are used to inform a sophisticated statistical model that can generate a very large number of scenarios. The scenarios are consistent with what has been observed historically, but also capable of exploring marginal cases that have rarely occurred in the past.

The second step is to *analyse* the impact of each possible contingency across all anticipated operating points. This is done by carrying out a time-domain simulation and investigating the impact of a particular asset failure event.

In the final step, by applying a suitable machine learning algorithm, the results of millions of simulations are *compressed* into a set of rules that can subsequently classify unseen operating points as being acceptable or unacceptable. This way, a rule is obtained for each analysed contingency that will support the online iTesla platform by predicting in a very fast way whether a particular operating point is susceptible to violation of security margins in case of a contingency.

The offline rule generation workflow has been optimised to run on high performance computing (HPC) facilities i.e., computer farms with several thousands of cores. By leveraging the workflow's inherent parallelisability, millions of simulations were carried out, allowing operators to compute rules across thousands of contingencies in the span of a few days of intense processing.

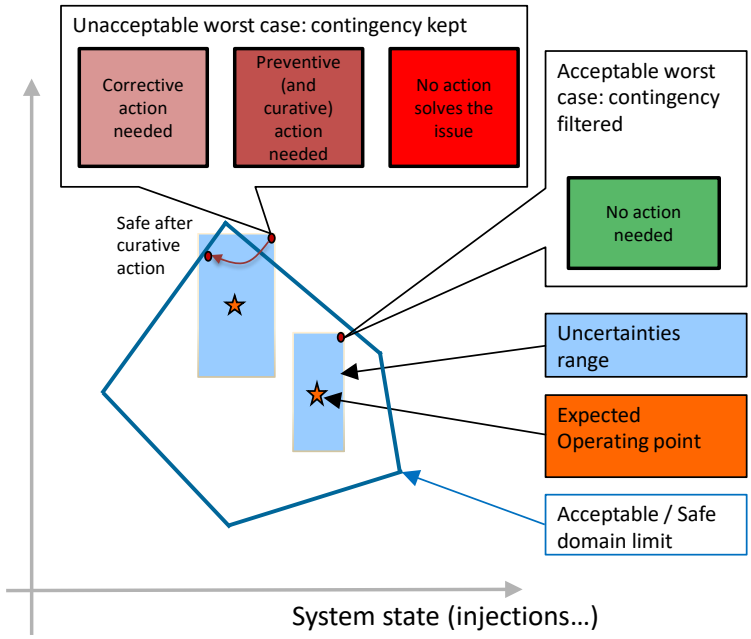
Decision trees (DTs), widely used in different scientific fields (Amrouch et al., 2016; Gavankar and Sawarkar, 2015; Lopez and Sigrist, 2017; Mogre et al., 2016; Nithya and Santhi, 2015), are trained on a set of plausible states of the power system under study in order to obtain these rules. In this way, generalisable security rules with high predictive capability can be constructed (He et al., 2013).

The offline workflow also elaborates the historical data to generate a complex model of the statistical dependence among RES and loads which is fully exploited in the on-line platform (discussed in the next section). Due to the high dimension of the problem and to issues in the real data under analysis (e.g., substation configuration, missing data etc.), historical data need to be pre-processed and undergo various compression steps such as clustering and principal component analysis (PCA), with some functions in common with the rule generation.

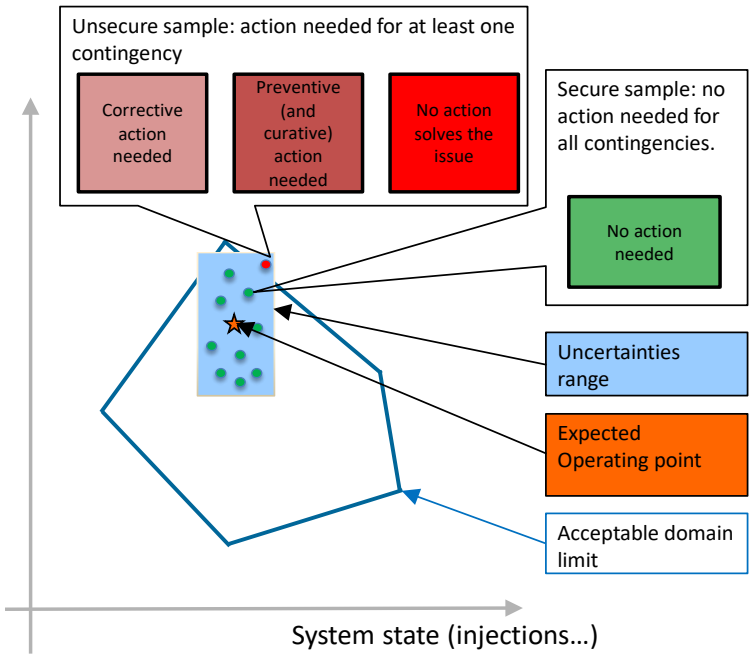
The online workflow contains several modules:

- A merging module in charge of reconciling data from the TSOs in order to obtain a consistent network state for the European electricity network. Data are affected by inconsistencies (e.g., measurement errors, time differences, etc.) and characterised by different reliability levels. A hierarchical merging procedure solves successive AC-optimal power flows (OPF) in order to build a consistent merged network state by minimising the deviations from the data provided by the different TSOs.
- A filtering procedure to reduce the computational burden of the online platform, by focusing on the contingencies which are likely to provoke security problems in the forecast system state subject to uncertainties. The procedure consists in two main stages:
  - a The worst state approach (WSA) aimed to discard contingencies which are highly stable within the whole uncertainty domain, considering in a simplified way the security limits, the uncertainties of the injections, and the available control actions [see Figure 2(a)]. The method addresses a very complex optimisation problem (bi-level optimisation with discrete variables). To solve this problem, the 'DC loadflow' approximation for grid equations is adopted and all aspects of the problem are dealt with accordingly. 'Security rules' for WSA account for security constraints (including stability) in terms of variables of 'DC loadflow', in order to be included as constraints in the optimisation problem. The domain of uncertainty of the stochastic variables is expressed as a convex domain.
  - b The Monte Carlo-like approach (MCLA), which makes a limited sampling of the uncertain injections around the forecast state, and checks the new sampled states against 'rich' security rules [see Figure 2(b)]. Unlike the rules defined for WSA, the ones for MCLA are defined on the whole range of AC quantities, hence they can more effectively take into account stability limits. The uncertainty model developed for MCLA and based on historical data analysis accounts for complex dependencies. MCLA is performed for each contingency classified as not secure by the previous filtering stage. For each contingency, the states that are classified as potentially insecure are moved on to the next detailed analysis.

Figure 2 (a) WSA (b) MCLA approaches (see online version for colours)



(a)



(b)

- A further module aimed to identify adequate corrective control actions is run for all the {state, contingency} couples resulting as potentially harmful from 'rich security rules' judgement. The objective is to check more accurately if the situation is actually critical, and to identify post-contingency actions that eliminate the violations. Security-constrained optimal power flow (SC-OPF) techniques are involved in this step. In order to model the operating rules adopted by control room operators, a pre-defined set of corrective actions is associated to specific violations and to specific contingencies. The optimisation module includes a solver dedicated to topological reconfiguration. Topology and the related corrective actions are modelled through binary variables in the nonlinear optimisation problem. It is well-known that the resulting MINLP is a NP-hard problem. Thanks to an innovative method, this module is able to find necessary corrective actions to avoid current limits violation. An important element in that approach is the fact that a limited number of corrective actions are considered. Indeed, the choice has to be made between corrective actions that have been selected by the dispatchers and which are dedicated to the specific couple {contingency, violated constraints}.

The remainder of the paper will focus on the uncertainty modelling for MCLA module in the online platform and on the offline procedure to build the security rules.

2.1 Uncertainty modelling in online platform

Loads and RES forecast uncertainties are taken into account by developing probabilistic models of the forecast error, i.e., of the difference between forecast values FO (e.g., evaluated on the day ahead) and values occurring in real operation (snapshots, SN, from the state estimator in the real-time operation environment). Forecast error models, based on historical data series, account for:

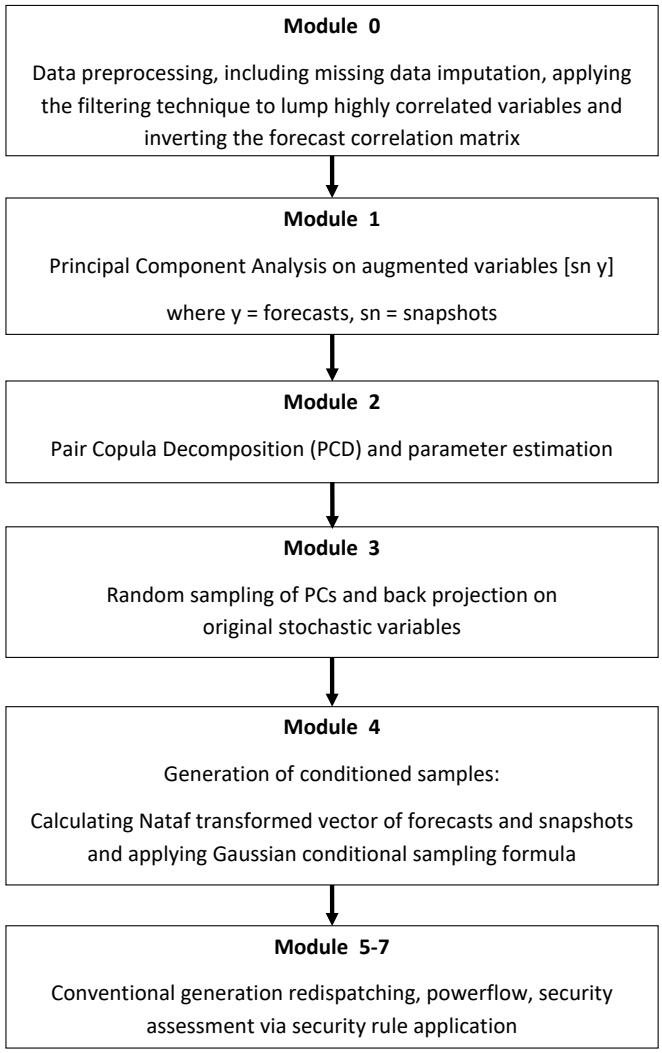
- Dependences between injections (e.g., similar errors in case of forecasts based on weather variables).
- Forecast values themselves (e.g., in case of very sunny or very cloudy weather, uncertainty on PV production will be small, and vice versa in case of partially clouded sky).

The uncertainty modelling (developed in the MCLA workflow illustrated in Figure 3) can be very complicated due to the high number of stochastic injections. Moreover, probability distributions of loads and RES forecast errors are non-Gaussian and their dependences are non-linear. To deal with these problems offline platform applies an approach based on k-means clustering, PCA (PCA, performed in module 1), a technique described in Chatfield (2018), Jolliffe (2002), Jolliffe and Cadima (2016) and Scott (2015) and used in different scientific fields (Han et al., 2015; Li et al., 2015; Reddy et al., 2017; Sharma and Singh, 2015; Tani et al., 2016; Yin et al., 2015), and pair copula decomposition (PCD, in module 2) described in Aas et al. (2009) and Aas (2016) and applied in Klein (2016) and Ruscone and Osmetti (2017), with the final aim of building a reasonable statistical model of forecast errors.

Again, some computationally time-intensive operations on large historical datasets (e.g., matrix inversions) are performed offline (in module 0) to pre-compute some quantities to be used in the on-line platform.

'MCLA' performs a sampling of the uncertainty domain around the forecast state, and then checks each new sampled state against the security rules. In particular, MCLA receives as input the forecast states with an associated uncertainty model computed offline, the list of contingencies and the security rules computed by the offline workflow. New plausible states are computed online from the forecast state by sampling the stochastic injections (loads and RES) and adjusting the conventional generation accordingly.

Figure 3 MCLA workflow with focus on the RES and load forecast uncertainty modelling



The sampling performed in module 4 is 'conditioned' by the forecast error model, i.e., the sampling process takes into account the current value of the forecast under analysis, as well as the dependence among conditioned variables. The adopted sampling technique limits the on-line computation burden by exploiting several pre-processing techniques

applied offline (in module 0), the Nataf transformation, described in Nataf (1962) and already applied in many scientific fields (Chen et al., 2015a, 2015b; Zhang et al., 2015; Zghal et al., 2015), as well as the properties of Gaussian conditional multivariates. This makes the technique suitable for on-line applications dealing with large sets of stochastic variables. The output of MCLA is the security assessment of each evaluated state with respect to the analysed contingencies.

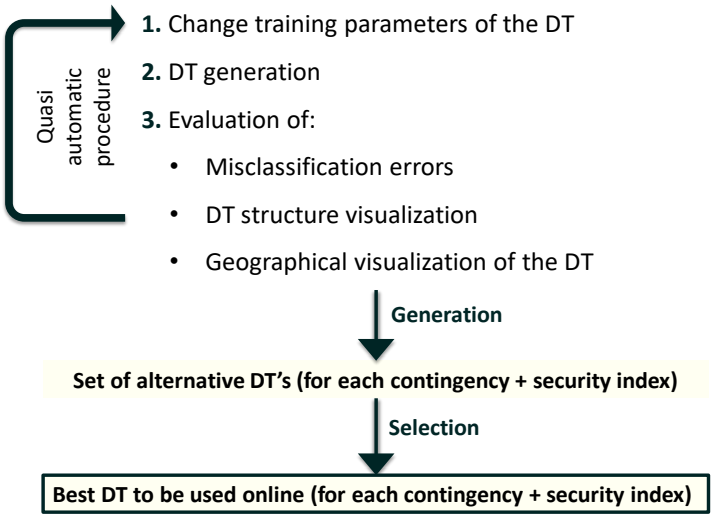
2.2 Online security assessment functions

As far as the last modules of the MCLA workflow are concerned, module 5 applies the redispatching to conventional generation to compensate the power imbalance due to forecast errors, thus getting a plausible system state, module 6 runs the loadflow on the generated power system state and finally module 7 evaluates the security of each system state by applying the security rules generated in the offline platform for each contingency and each security problem.

2.3 Offline DT training procedure

For the generation/selection of the best security rules, i.e., DT for each contingency, a quasi-automatic procedure is adopted in the iTesla offline platform, which is summarised in Figure 4.

Figure 4 Applied steps for DT generation/selection (see online version for colours)



This procedure comprises the following main steps:

Step 1 Change the training parameters of the DT (detailed in Section 4).

Step 2 DT generation.

Step 3 Evaluation of the DT performance results by analysing:

a Misclassification errors provided by the k-fold cross validation method (detailed in Section 4).

b DT tree structure visualisation.

c Geographical visualisation of the DT.

From this procedure, a set of alternative DTs is obtained for each contingency/security index pair, being each alternative DT associated with a different setting of the training parameters. A different set of alternative DT's is obtained for each one of the following online filtering methods:

- Worst case approach (WCA) method (the set of candidate attributes of the DT can only be related with active power operating conditions).
- MCLA method (the set of candidate attributes of the DT can also consider other type of operating conditions).

From the set of alternative DTs, the best DT is selected by analysing each DT performance results and by pursuing the following criteria:

- Criterion 1 (highest priority): To maximise the accuracy of the DT in detecting unsecure situations, being achieved by minimising the missed alarms (MA) situations (i.e., an unsecure state/contingency pair for which the DT provided a secure classification).
- Criterion 2: To maximise the efficiency of the DT in detecting secure situations, aiming to maximise the filtering capability of the contingency filtering process of the online part. This is achieved by minimising the false alarms (FA) situations (i.e., a secure state/contingency pair for which the DT provided an unsecure classification).
- Criterion 3: To avoid over-fitting to the learning states. In fact, if the computed DT is over-fitted to the learning network states, then the DT may lose generalisation capability to obtain an accurate estimation for unseen new states (i.e., states not used in the learning process).

After selecting the assumed best DT for each contingency, the accuracy of the DT was further evaluated with historical data, aiming to assess the generalisation capability of the DTs.

Next sections are dedicated to the procedures run to validate the uncertainty model used in the online MCLA workflow and the security assessment performance achieved via DTs in module 7 of MCLA.

3 Validation of uncertainty model

The objective of the validation process is to evaluate the quality of the samples of stochastic injections conditioned to the forecast state, generated in module 4 of MCLA workflow. To achieve this goal, it is necessary to introduce suitable metrics to assess the distance between the probability distribution of conditioned samples generated by module 4 and the 'reference' conditioned distribution $f(sn|y)$.

To this aim, two tests are performed:

- A comparison between the conditioned distribution sampled in module 4 and the 'benchmark' conditioned distribution obtained from the so called 'nearest neighbour (NN) method' for a specific forecast state y_0: this validation is suitable for small sets of variables.
- A cross-validation scheme to assess the general adequacy of the conditional density model $f(sn|y)$, for any forecast state y: the proposed conditional density estimator is tested against a well-established alternative conditional density estimator, the kernel-based estimator based on Gaussian kernels, often used in literature (Arora and Taylor, 2016; Jeon and Taylor, 2016; O'Brien et al., 2016; Taylor and Jeon, 2015; Zolfaghari et al., 2015).

The metrics used for comparisons are:

- Average root mean square error (ARMS) of the CDFs for the marginal distributions of conditioned variables.
- The Kendall's rank correlation coefficients among each pair of conditioned variables.
- The Kullback-Leibler (KL) divergence (only for cross-validation) among the conditioned densities.

3.1 Benchmarking against NN validation method

This test consists in comparing the first two statistical moments of the marginal distributions (and the Kendall's rank pairwise correlation coefficients) of the conditioned variables sampled respectively by module 4 of MCLA workflow and by NN method.

The NN method consists in the following steps:

- Run modules 1, 2 and 3 on an augmented matrix $X = [sn\ y]$ which includes both snapshots sn's and forecast y's
- For any sample h $[sn\ y_0]$ to be generated with given vector y_0 of forecasts, perform the following sub-steps:
  - a Generate M trial samples from module 3.
  - b Any trial sample j has a metrics $d_j = \max_{i \in Y}(|\ y_i^{(j)} - y_i^0\ |)$.
  - c Select the trial sample (*) s.t. $d_{(*)} = \min_j(d_j)$.

The NN method can be used for validation purposes, even though it may have some criticalities especially in case of large sets of variables: in fact, when the number of stochastic variables increases, the present method leads to increasing inaccuracy because metrics $d_{(*)}$ may get unacceptably high.

3.2 Cross-validation against kernel-based estimator

For large sets of variables, benchmarking against the NN is impractical: the most convenient validation scheme consists in performing a cross-validation between the

proposed method and a well-established conditional density estimator: a kernel-based estimator with Gaussian kernels.

The metrics used to compare the two methods is the KL divergence (Van Erven and Harremoes, 2014) representing the distance between true and estimated functions and given by

$$D\left(h\|\hat{h}\right) = -\int h(x)\log\hat{h}(x)dx + \int h(x)\log h(x)dx \tag{1}$$

where function $h(x) = f(sn|y)$ of the vector of augmented variables $x = [sn\ y]$ is the true conditional distribution density while $\hat{h}$ is the density estimate.

In particular, the first term $-\int h(x)\log\hat{h}(x)dx$ is called 'loss function' R and depends on the density estimate, while the second term, called 'entropy', does not depend on estimate: thus, minimising the KL divergence means minimising the related loss function. An interesting property of the KL divergence that arises from Jensen's inequality (Jensen, 1906) is that $D(h\,\|\,\hat{h})$ is always not negative and it is zero only if and only if $h \equiv \hat{h}$. Thus, the loss function is bounded below by the negative of the entropy of h. The goal is to calculate the loss function, which can be approximated in a leave-one-out cross-validation with:

$$-\int h(x)\log\hat{h}(x)dx \approx -\frac{1}{n}\sum_{i=1}^{n}\log\hat{h}\left(x_i\right) \tag{2}$$

where n is the number of samples, x_i is the i^{th} test sample and $\hat{h}(x) = \hat{f}(sn\,|\,y)$ is the density estimate obtained from the remaining $n - 1$ samples (training set).

The paper adopts the k-fold cross-validation scheme due to its relatively small sensitiveness to the way how data are partitioned. The method works according to the next steps:

1 Split the historical data into $N = 10$ sets; at each iteration of the procedure each subset is alternatively used as test set and the other subsets compose the training set.

2 Calculate the estimate of loss function R_p for p^{th} training/test set combination ($p = 1, \ldots N$) in equation (3) where in this case density estimate is obtained from p^{th} training set Λ_p and x_i are the samples of p^{th} test set, and $n_{test,p}$ is the number of samples of p^{th} test set Ω_p

$$R_p \approx -\frac{1}{n_{test,p}}\sum_{jtest\in\Omega_p}\log\hat{h}_p\left(x_{jtest}\right) \tag{3}$$

3 Sum up all the terms in step 2 over all the N combinations and divide it by N, thus getting R estimate: $R = \sum_{p=1}^{N} R_p \Big/ N$.

4 The estimate with the lowest loss function $-\int h(x)\log\hat{h}(x)dx$ is the best estimate of density f.

Step 2 requires the estimation of the conditional density which can be performed via two methods:

- For kernel-based estimator, the estimate (Silverman, 1986) is given by

$$\hat{f}\left(sn_{itest} \mid y_{itest}\right) = \frac{\sum_{jtrain}^{Ntrain} K\left(\left|sn_{jtrain} - sn_{itest}\right|\right) \times K\left(\left|y_{jtrain} - y_{itest}\right|\right)}{\sum_{jtrain}^{Ntrain} K\left(\left|y_{jtrain} - y_{itest}\right|\right)} \tag{4}$$

  where K is the Gaussian kernel with bandwidth vector b, i.e., $K(t) = e^{-\frac{t \cdot diag(b^{-2}) \cdot t^T}{2}} \Big/ \left(\left(\sqrt{2\pi}\right)^d \prod_{i=1}^{d} b_i \right)$. This method requires the solution of an optimisation problem to find the optimal bandwidths of the conditional density kernel estimator with Gaussian kernels in order to minimise the divergence term over all the alternative test sets: to do this, the *fmincon* function in MATLAB is used.

- For the NT-based method, the estimate of the multivariate conditional density $\hat{f}$ (Silverman, 1986) is given by equation (5)

$$\hat{f}\left(sn_{itest} \middle| y_{itest}\right) = \Phi\left(sn'_{itest} \middle| y'_{itest}\right) = mvnpdf\left(sn'_{itest}, \mu_{train}\left(y'_{itest}\right), \Sigma_{train}\right) \tag{5}$$

  where *mvnpdf* is the multivariate normal pdf function in MATLAB, Φ is the multivariate normal density, sn'_{itest} and y'_{itest} are the Nataf transformed variables corresponding to snapshot sn and forecast y of i^{th} sample of the test set, and are given by φ-1($F_{sn\ itest}$) and φ-1($F_{y\ itest}$), where F_x are the CDF value of variable x, $\sum_{train}$ and μtrain are the covariance matrix and the mean vector derived from training set (μtrain also depends on y'_{itest}).

The output metrics to evaluate the cross-validation results are the loss function R estimate and the standard deviation (SD) of the 'log' estimates for any test/training combination, i.e.,

- loss function estimate R
- SD of the log estimates $SD = \underset{p=1\ldots N}{std}(R_p)$.

4 Validation of the security assessment performance

The MCLA security assessment relies on the security rules generated in the iTesla offline platform. Hence, validating MCLA is in some extent equivalent to validating the rules.

The validation consists in running detailed simulations to assess the true classification 'secure/unsecure' for each state belonging to the uncertainty cloud around the forecast state. After that, the same states are evaluated using a security rule for each contingency and each security problem. The comparison of the results from rule application and detailed simulations allows to compute some performance indicators.

The performance indices defined for validation purposes consist in accuracy and efficiency, and they are defined for each phenomenon (described by the rules trained on specific security indices). The relevant indices can be defined as:

$$Accuracy = \left(1 - \#UCS_{specific\_phenomenon} / \#U_{specific\_phenomenon}\right) \times 100$$

$$Efficiency = \left(1 - \#SCU_{specific\_phenomenon} / \#S_{specific\_phenomenon}\right) \times 100$$

where $\#UCS_{specific\_phenomenon} = nr$ of unsecure states classified as secure, and $\#SCU_{specific\_phenomenon} = nr$ of secure states classified as unsecure, $\#U_{specific\_phenomenon} = nr$ of unsecure states, $\#S_{specific\_phenomenon} = nr$ of secure states for the specific phenomenon.

It can be seen from Figure 5 that different levels of output are obtained, for each analysed contingency, namely:

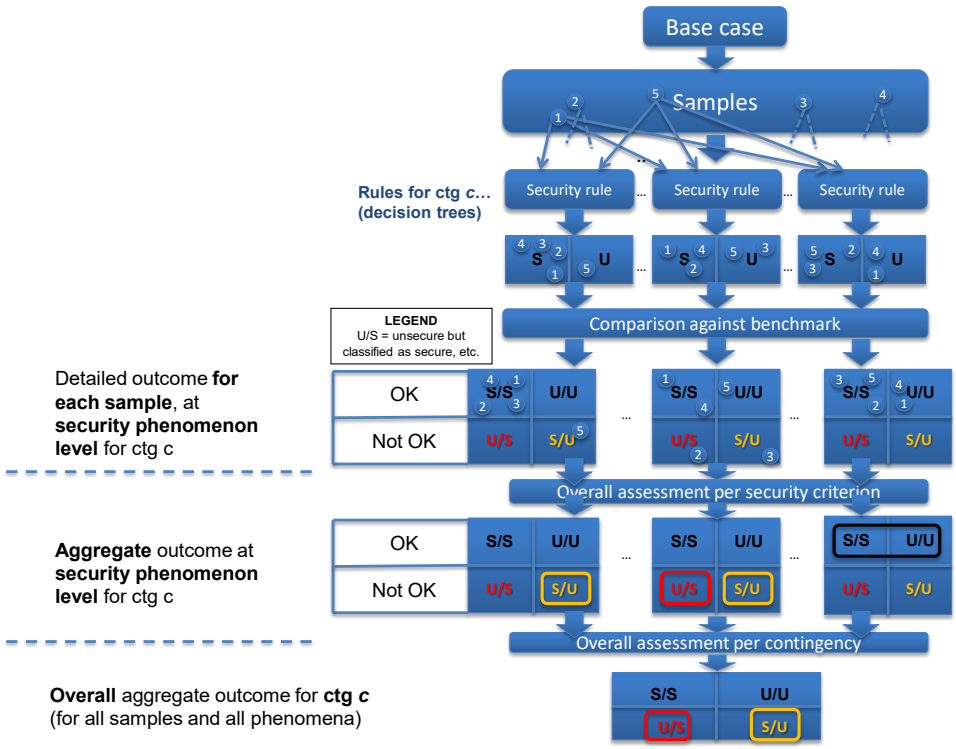
1 sample by sample, per phenomenon

2 aggregated for all samples, per phenomenon

3 aggregated for all samples and all phenomena.

In the validation phase, all contingencies must be evaluated, including the ones that would be discarded as harmless (cluster 1) by the previous filtering module WCA and the 'null contingency'.

Figure 5 Performance index computation for MCLA (see online version for colours)



Note: This process also validates the security rules

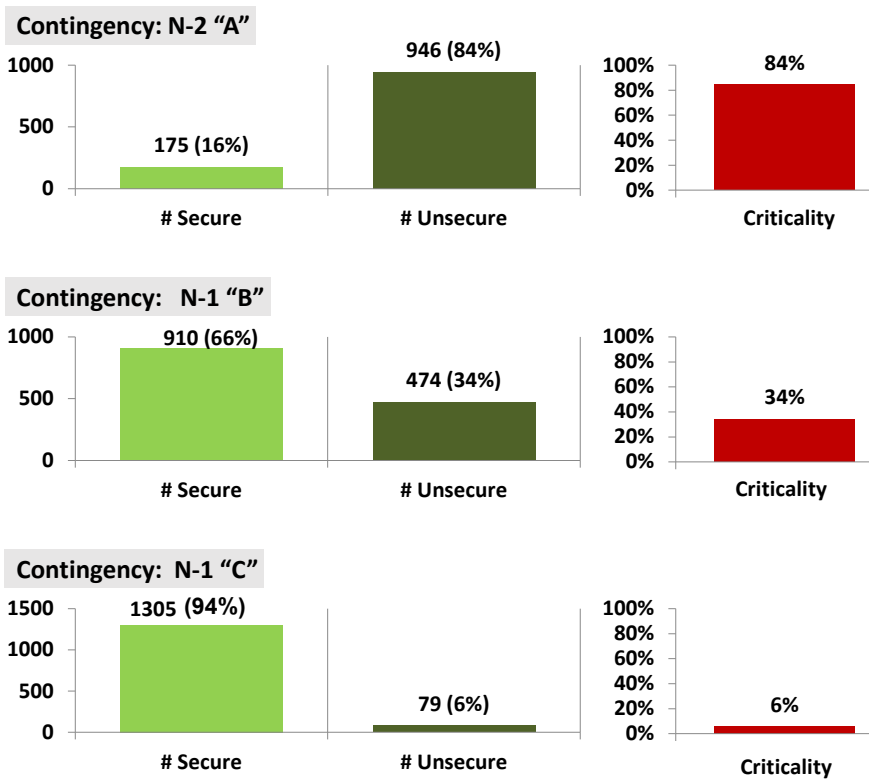
5 Case studies

This section illustrates some preliminary results on the validation of the RES forecast uncertainty modelling, and of the security assessment performance of the online iTesla platform. The datasets used for this study case are:

- Dataset 1 consisting in a set of snapshots (SN) and forecasts (FO) for about 3,800 stochastic variables (active/reactive powers from renewable and loads in the French EHV grid) related to two months (January and February) in 2013.
- Dataset 2 consisting of SNs and FOs of March 2013.
- Dataset 3 consisting in 2736 samples of forecasts and snapshots of four stochastic variables representing active powers of aggregated loads in a simplified 'French system' with seven buses.

The historical data used were French day-ahead congestion forecast (DACF) and SN files, in CIM profile 1 format, with persistent identifiers for equipment regardless of the time and type of file (DACF and SN).

Figure 6 Criticality of the analysed contingencies (see online version for colours)



To validate the security assessment performance, the present study considers three contingencies involving overloaded lines:

- Two 400 kV tripping lines in Normandy area leading to an overloaded 225 kV line (henceforth called contingency N-2 ‘A’).
- One 400 kV tripping line in Languedoc area leading to an overloaded 225 kV line (henceforth called contingency N-1 ‘B’).
- One 400 kV tripping line in the Charente area leading to an overloaded 225 kV line (contingency N-1 ‘C’).

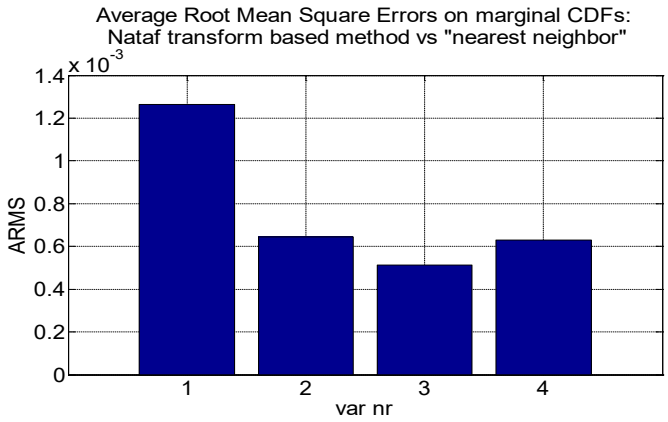
The first case refers to a very severe contingency which is monitored in case of exceptional weather conditions whereas the second one and third one are continually monitored.

From an operational point of view, the three situations might seem very close but when processed in the offline part they lead to various levels of criticality which has a strong impact on the generation of the DTs. To confirm the previous statement, Figure 6 describes the number of secure and unsecure states that are included in the generated dataset of pre-analysed network states. From this figure, it is possible to observe that for contingency ‘A’, the number of generated pre-analysed states includes 175 secure states (i.e., #secure = 175) and 946 unsecure states (i.e., #unsecure = 946), comprising a total of 1,121 states (i.e., #states = #secure + #unsecure). Therefore, in this dataset, the criticality of the contingency is around 84% (i.e., #unsecure/#states).

6 Validation results for the uncertainty model

The accuracy of the uncertainty model used in the online MCLA workflow is evaluated in the present section.

Figure 7 ARMS between the CDFs of the four variables evaluated by the two methods (Nataf and NN) (see online version for colours)



6.1 Comparison against the NN method

The benchmarking of the MCLA conditional sampling against NN method is performed on the seven bus grid, given the specific set of forecasts y_0 = [240 240 480 480] MW.

Figure 7 reports the ARMS of the CDFs of the four variables evaluated by NT-based and NN methods: the small ARMS values (well below 0.5%) confirm a good matching of the marginals given by the two methods.

6.2 Cross validation against kernel estimator

The cross-validation has been applied to both dataset 3 (consisting in active power absorptions of four equivalent loads over one year interval) and a subset of 40 variables (RES and load power exchanges) from dataset 2 related to March 2013. From the original large sets of samples, the present simulation extracts 300 samples uniformly spread over the whole set (in order to avoid any possible polarisation, especially for dataset 3 subject to seasonal effects). These samples are then clustered into ten sets with 30 samples each: the number of samples for each set has been chosen as a trade-off between a fair representation of the statistical properties of the data and a sufficiently fast solution of the bandwidth optimisation for the kernel model.

Moreover, the number of subsets ($N = 10$) assures a good trade-off between the bias (mean value) and the variance of loss function values R_p computed for any test/training combination. Moreover, 50 runs of the cross-validation procedure have been done by randomly associating the samples to each set.

Table 1 reports the median, minimum and maximum value of the loss function R of KL divergence (over the 50 cross-validation runs) for the two estimation methods, for datasets 2 and 3. The close values between median, maximum and minimum values for R estimates show that the cross-validation is poorly sensitive to the specific association of the samples to the sets.

Table 1 Estimated loss function R of KL DIV

Data set #	*3*			*2 (subset of 40 variables)*		
R value / *Estimate*	*Min*	*Median*	*Max*	*Min*	*Median*	*Max*
Kernel-based	+3.52	+3.58	+3.66	+37.47	+41.17	+42.19
NT-based.	–2.89	–2.88	–2.87	+6.75	+6.97	+7.52

Note: For Datasets 2 and 3.

In both datasets NT-based method has a lower loss function R, which means that it provides the best estimates of the true conditional distributions $f(sn|y)$. Changing the number and the sample size of the sets does not impair the validity of the previous statement even though it affects the individual values of R estimate for the two methods, as demonstrated in the next subsection.

6.3 Assessing the effect of different cross-validation schemes

The impact of the cross-validation schemes on the bias and the variance of the loss function (defined in Section 3) is analysed in detail on dataset 3. To this aim, we consider the same total number of samples for the simulation (200), but two different cross-validation schemes: 4/1 and 9/1. From the original large sets of samples, the present simulation extracts 200 samples uniformly spread over the whole set (in order to avoid any possible polarisation). These samples are then clustered into 5 (or 10) sets with 40 (or

20) samples each: the number of samples for each set has been chosen as a trade-off between a fair representation of the statistical properties of the data and a sufficiently fast solution of the bandwidth optimisation for the kernel model. Moreover, 11 runs of the cross-validation procedure have been done by randomly associating the samples to each set.

Table 2 reports compare the mean, median and maximum value of the loss function R for the two cross-validation schemes (over the 11 cross-validation runs). The close values between median, maximum and minimum values for R estimates show that the cross-validation is poorly sensitive to the specific association of the samples to the sets. Passing from 4/1 to 9/1 the loss function decreases: in fact, at each combination of training/test sets, a larger number of samples represent the training set to build the density model, which improves its accuracy.

Table 2 Loss function R estimate for cross-validation schemes 4/1 and 9/1

Loss function R	*4/1 scheme*			*9/1 scheme*		
	Mean	*Median*	*Max*	*Mean*	*Median*	*Max*
Nataf method	–2.7617	–2.7647	–2.7454	–2.7693	–2.7711	–2.7560
Kernel estimator	3.7439	3.7091	3.8810	3.6339	3.6389	3.6729

Table 3 instead reports the SD value of the loss function R for Nataf method over the N test/training combinations for three cross-validation schemes 4/1, 9/1 and 19/1 (i.e., $k = 5$, 10 and 20) for a specific random assignment of samples to the sets. It can be noticed that in the large majority of cases SD increases as k increases: in fact, fewer samples in the test set leads to a larger volatility in the log estimates R_p.

Table 3 SD value for loss function R for both methods for three cross-correlation schemes

	4/1 scheme	*9/1 scheme*	*19/1 scheme*
Nataf method	0.1030	0.2142	0.3553
Kernel estimator	0.2740	0.2550	0.3192

These tests are repeated for the three cross-correlation schemes and for Nataf method (with one random assignment of the samples to the sets) for the same subset of 40 injections mentioned in Table 1. From Table 4 it can be noticed that also for this case the increase of the number of folds k causes a reduced loss function R but a larger variability of log estimates among the training/test combinations.

Table 4 Loss function R estimate and SD for R for both methods and three cross-validation schemes

	Nataf method			*Kernel estimator*		
	4/1 scheme	*9/1 scheme*	*19/1 scheme*	*4/1 scheme*	*9/1 scheme*	*19/1 scheme*
Loss of function R	6.3028	5.9099	5.8217	17.3623	17.2348	17.1201
SD for R	0.9024	1.2867	1.5959	0.8920	1.2847	1.7609

The increase of k from 10 to 20 is not justified by the slight decrease of loss function (from 5.90 to 5.82); thus, simulations seem to confirm that $k = 10$ is a good trade-off between a reduced loss function and a limited variability of log estimates R_p. Anyway,

any choice of k confirms that the conditioned density structure computed by Nataf method works better than kernel estimator at least for the sets of variables under test.

To summarise, it is worth highlighting that the simulated cases do not demonstrate that NT-based method always works better than kernel-based estimator, but the extensive campaign of simulations run up to now confirms the abovementioned remark. This is due to the fact that kernel estimator considers independent perturbations in the Gaussian kernel formulation, i.e., the covariance matrix in the kernel is diagonal: moreover, many historical samples are required around the specific forecast vector to achieve a good estimate of local correlation.

7 Validation results for the security assessment performance

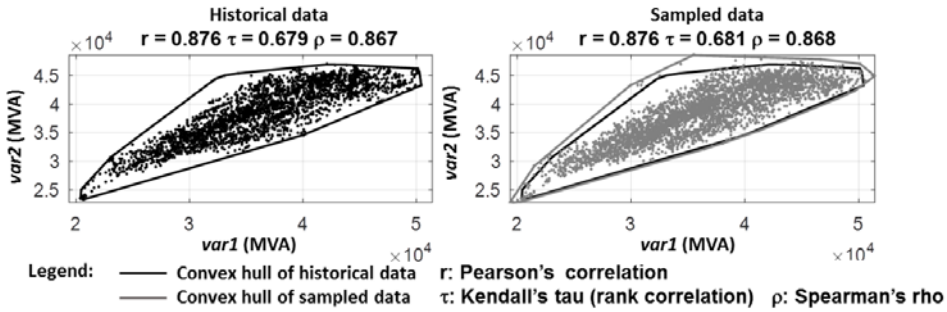
This section presents some results related to the offline training of the DTs and to the validation of the security assessment capability of the online platform for two of the contingencies in Section 5.

7.1 Training of the DTs and selection of the best choice: some results

The iTesla platform detects unacceptable situations due to overload problems by comparing the post-contingency steady-state current in all the lines and transformers with its operational limits available in the CIM file: if at least one limit is violated, then the situation is considered unacceptable. As already pointed out, firstly the uncertainty modelling is used to extract the samples of stochastic variables to generate plausible states of the network on which extensive security analyses are performed to get the information to train DTs. The iTesla platform allows a visual comparison between the historical SN used for sampling and the sampled states in the form of plots on relevant sets of the stochastic variables (RES injections and loads).

Figure 8 presents some of the plots obtained for aggregated stochastic variables. The variable in x-axis of this figure, var1, represents the summation of the first half of the stochastic variables (from 1 to 1,899), whereas the variable in the y-axis, var2, comprises the summation of the second half of stochastic variables (from 1,900 to 3,798). Note that the left plot represents the historical data and the right plot was obtained for the sampled data.

Figure 8 Visual comparison between historical and sampled stochastic data



For easier detection of outliers, these plots also present the convex hull of each dataset (in black for historical and in gray for sampled data). Pearson's correlation (r), Kendall's tau rank correlation (τ) and Spearman's rho (ρ) are also provided to assist the visual comparison. These plots show that between historical and sampled data:

a the stochastic aggregated variables occupy roughly the same area of the state space

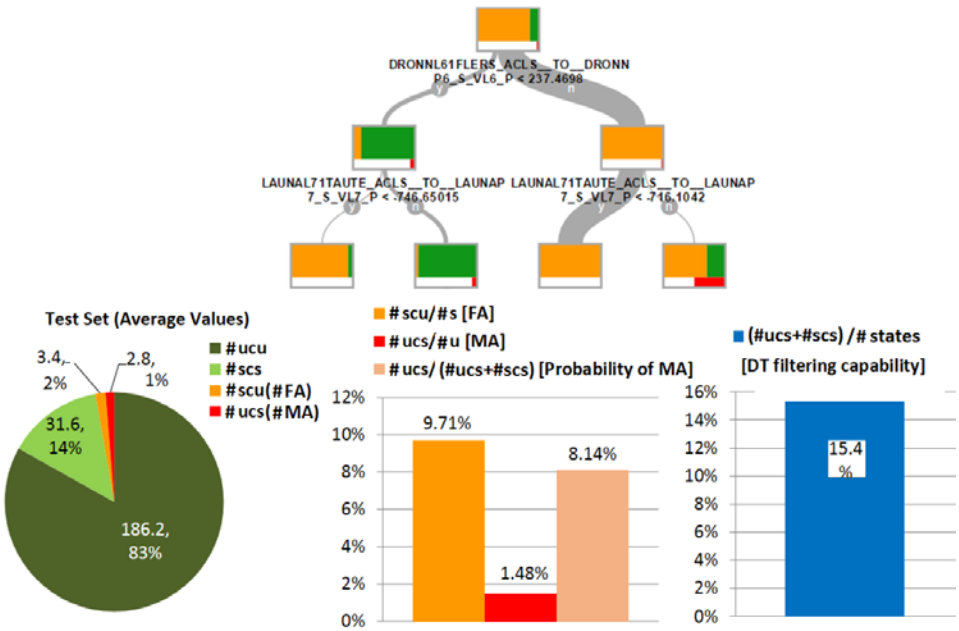
b the dependence structure is largely the same.

Hence, the results presented illustrate that the generated sampled stochastic variables seem consistent with the input historical data (SN).

After building the DT related to the contingency under study, in the on-line workflow this DT is used to perform online security assessment of the analysed contingency and security problem, including the uncertainties provided by the MCLA. For validation purposes, this procedure was applied to the DACFs of the 25 and 27 of February 2013 since these were the most unsecure days forecasted for this month. The forecast error model of stochastic variables was calculated offline using the DACFs and SNs of January 2013 as historical dataset.

As an example, Figure 9 shows the best DT for the N-2 contingency 'A' and for the overloading problem to be applied for the MCLA filtering process.

Figure 9 Best DT for contingency N-2 'A' and for overloading problem (see online version for colours)



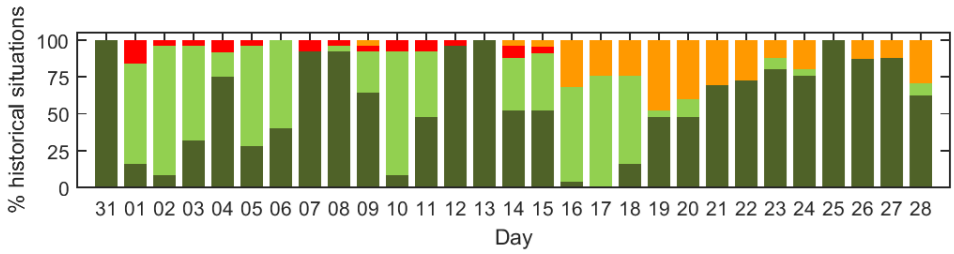
In the pie chart, light green colour corresponds to secure states in the test set classified as secure, dark green colour to the unsecure states classified as unsecure, orange colour to the secure states classified as unsecure (representing FAs) and the red colour to unsecure states classified as secure (i.e., MAs). This DT minimises the overall error rate

([#usc + #scu]/#states) for the MCLA, presenting a very small MA error (1.48%) and, therefore, also providing a good accuracy.

7.2 Security assessment on contingency N-2 'A'

From DT evaluation performed in the previous section, it is expected that the DT selected for the N-2 contingency 'A' provides some misclassifications for the days after the 15th of February 2013, because of the disconnection, on that day, of a nuclear power plant that remained off during that month (topology not considered in DT training). So, also for the two days selected to validate the MCLA, some amount of misclassifications are expected to be provided for the states generated by the MCLA. Figure 10 shows the daily summary of DT performance for all the DACF of February 2013.

Figure 10 Daily performance for the DACF of February 2013 (N-2 'A' (see online version for colours)



From Figure 10 one can find that:

- A considerable amount of FA were obtained for 19–20 February 2013.
- 100% of unsecure situations were correctly identified for 25 February 2013.
- A large amount of unsecure situations, including some FAs, was identified for 26–27 February 2013.

7.3 Security assessment on contingency N-1 'B'

The present analysis consists in applying the validation scheme illustrated in section 4 to contingency N-1 'B', i.e., a short-circuit in a single 400 kV line that leads to the line disconnection, which may create overload problems in a neighbouring 225 kV line. Given the 100 samples generated by the MCLA and the DACF in the security domain of the DT, the 2D four-colour plots in Figure 11 present four relevant situations in which:

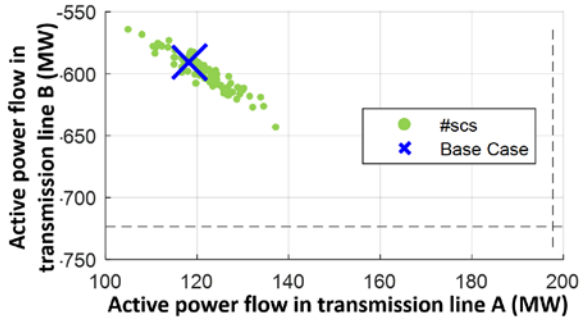
a the blue cross defines the location of the DACF (base case)

b the dots describe the location of the MCLA samples (uncertainty).

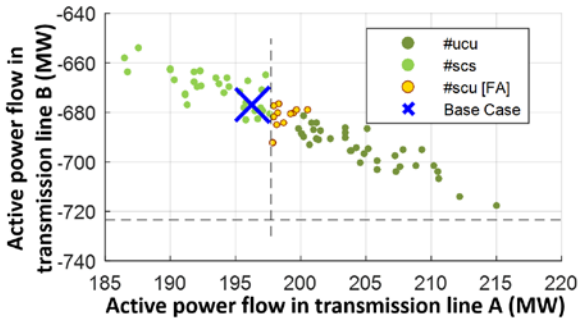
The results of Figure 11 show that there is only a small number of states that are FA (FA, namely, for the DACF at 12:30 and 18:30) when considering the forecast uncertainty. As expected, if the DACF is not located near the DT security boundary (like in the DACF detailed in Figure 11 at 2:30 and 19:30), all states sampled by MCLA share the same classification as the originating DACF. However, there are two DACFs in Figure 11

(namely, at 12:30 and 18:30), in which it was possible to observe the interesting situation where from an apparently secure situation, i.e., from a DACF with no overload problems, there is a considerable amount of unsecure situations when the forecast uncertainty is taken into account.

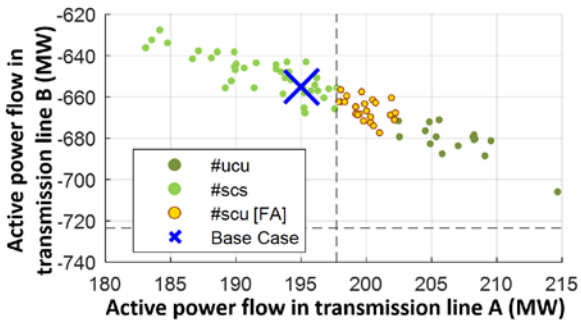
Figure 11 Plots of MCLA samples in the DT variable plane for some DACFs of February 2013, (a) 2:30, 27 February 2013 (b) 12:30, 27 February 2013 (c) 18:30, 27 February 2013 (d) 19:30, 27 February 2013 (see online version for colours)



(a)



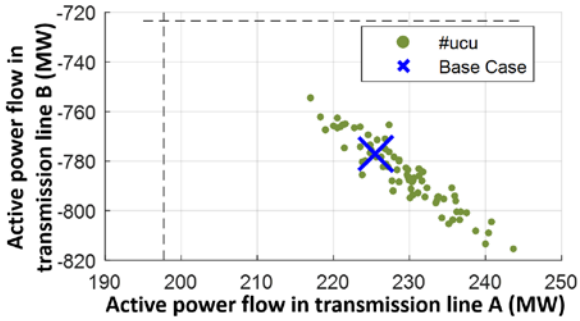
(b)



(c)

Notes: ucu – states unsecure classified unsecure; scs – states secure classified as secure; scu –states secure classified as unsecure (false alarms = FA).

Figure 11 Plots of MCLA samples in the DT variable plane for some DACFs of February 2013, (a) 2:30, 27 February 2013 (b) 12:30, 27 February 2013 (c) 18:30, 27 February 2013 (d) 19:30, 27 February 2013 (continued) (see online version for colours)



(d)

Notes: ucu – states unsecure classified unsecure; scs – states secure classified as secure; scu –states secure classified as unsecure (false alarms = FA).

These results demonstrate the utmost need of including uncertainty when performing online security assessment of power systems. The preliminary tests of the conditional sampling algorithm applied to 'full' dataset 1 show that the CPU time to generate a representative set of conditioned samples is about 40 s, compliant with the 15-minute time limit to perform security analyses in on-line operation.

8 Conclusions

This paper has presented the overall architecture of the platform developed within the iTesla project to improve power system security assessment under load and renewable generation uncertainty. The ultimate goal is to support the decision making process during network operation from two-days ahead to real-time based on three main features:

a To provide a risk-based assessment taking into account different sources of uncertainties (e.g., load and renewable power generation) and contingency probability.

b To perform accurate and fast online security assessment using time-domain simulations and pattern-recognition techniques.

c To help operators with relevant proposals of preventive and curative actions to keep the system in a secure state.

Special attention is devoted to the description and the validation of the forecast uncertainty modelling approach uncertainty and of the security assessment procedure implemented in the iTesla platform. The results obtained from validation demonstrate a good accuracy of the iTesla modelling approach in predicting the uncertainty cloud around a specific forecast state, and good performances in detecting power system security via DTs.

The iTesla platform, whose code has been released open source on Github (MPL 2.0 license available at https://github.com/itesla/), will help TSOs to address security

assessment of their own system, of coordinated regional systems or of continental systems such as the whole Pan-European network.

The iTesla platform was validated for a well-studied security issue in the French network: overload situations in transmission circuits. The obtained results have demonstrated the need to capture forecast uncertainty since unsecure operating situations can arise from apparently secure forecast network states.

References

Aas, K. (2016) 'Pair-copula constructions for financial applications: a review', *Econometrics*, Vol. 4, No. 4, pp.1–15.

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009) 'Pair-copula constructions of multiple dependence', in *Insurance: Mathematics and Economics*, Vol. 44, No. 2, pp.182–198.

Amrouch, S., Mostefai, S. and Fahad, M., (2016) 'Decision trees in automatic ontology matching', *International Journal of Metadata, Semantics and Ontologies*, Vol. 11, No. 3, pp.180–190.

Arora, S. and Taylor, J.W. (2016) 'Forecasting electricity smart meter data using conditional kernel density estimation', *Omega*, Vol. 59, pp.47–59.

Chatfield, C. (2018) *Introduction to Multivariate Analysis*, Routledge, New York.

Chen, C., Wu, W., Zhang, B. and Sun, H. (2015a) 'Correlated probabilistic load flow using a point estimate method with Nataf transformation', *International Journal of Electrical Power & Energy System*, Vol. 65, pp.325–333.

Chen, F., Liu, H., Li, J. and Zhang, X. (2015b) 'Comparison of simulation methods of spatially correlated wind speeds', *Proc. of 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, pp.255–261.

CIGRE WG C4-601 (2010) *Review of On-line Dynamic Security Assessment Tools and Techniques*, Technical Brochure.

Gavankar, S. and Sawarkar, S. (2015) 'Decision tree: review of techniques for missing values at training, testing and compatibility', Paper presented at the *International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, Kota Kinabalu, Malaysia.

Han, L., Romero, C.E. and Yao, Z. (2015) 'Wind power forecasting based on principle component phase space reconstruction', *Renewable Energy*, Vol. 81, pp.737–744.

He, M., Zhang, J. and Vittal, V. (2013) 'Robust online dynamic security assessment using adaptive ensemble decision-tree learning', *IEEE Transactions on Power Systems*, Vol. 28, No. 4, pp.4089–4098.

iTesla Consortium (2014) *iTesla – Innovative Tools for Electrical System Security within Large Areas*, Annex I, 23 July.

Jensen, J.L. (1906) 'Sur les fonctions convexes et les inégalités entre les valeurs moyennes', *Acta Math.*, Vol. 30, No. 1, pp.175–193.

Jeon, J. and Taylor, J.W. (2016) 'Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation', *International Journal of Forecasting*, Vol. 32, No. 3, pp.991–1004.

Jolliffe, I.T. (2002) *Principal Component Analysis*, Springer, New York.

Jolliffe, I.T. and Cadima, J. (2016) 'Principal component analysis: a review and recent developments', *Phil. Trans. R. Soc. A*, Vol. 374, No. 2065, p.20150202.

Klein, B., Meissner, D., Kobialka, H.U. and Reggiani, P. (2016) 'Predictive uncertainty estimation of hydrological multi-model ensembles using pair-copula construction', *Water, Special Issue 'Uncertainty Analysis and Modeling in Hydrological Forecasting'*, Vol. 8, No. 4, pp.1–22.

Li, K., Hu, C., Liu, G. and Xue, W., (2015) 'Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis', *Energy and Buildings*, Vol. 108, pp.106–113.

Lopez, D. and Sigrist, L. (2017) 'A centralized UFLS scheme using decision trees for small isolated power systems', *IEEE Latin America Transactions*, Vol. 15, No. 10, pp.1888–1893.

Mogre, R., Talluri S. and D'Amico, F. (2016) 'A decision framework to mitigate supply chain risks: an application in the offshore-wind industry', *IEEE Transactions on Engineering Management*, Vol. 63, No. 3, pp.316–325.

Nataf, A. (1962) 'Determination des distributions dont les marges sont données', *Compt. rend. de l'acad. des sc.*, Vol 225, pp.42–43.

Nithya, R. and Santhi, B. (2015) 'Decision tree classifiers for mass classification', *International Journal of Signal and Imaging Systems Engineering*, Vol. 8, Nos. 1/2, pp.39–45.

O'Brien, T.A., Kashinath, K., Cavanaugh, N.R., Collins, W.D. and O'Brien, J.P. (2016) 'A fast and objective multidimensional kernel density estimation method: fast KDE', *Computational Statistics & Data Analysis*, Vol. 101, pp.148–160.

Panciatici, P. et al. (2010) 'Security management under uncertainty: from day-ahead planning to intraday operation', Paper presented at the *2010 IREP Symposium Bulk Power System Dynamics and Control – VIII (IREP)*, Rio de Janeiro.

Reddy, A.S., Agarwal, P.K. and Chand, S. (2017) 'Application of principal component analysis for the fault detection and diagnosis of active magnetic bearings', *International Journal of Advanced Mechatronic Systems*, Vol. 7, No. 4, pp.245–255.

Ruscone, M.N. and Osmetti, S.A. (2017) 'Modelling the dependence in multivariate longitudinal data by pair copula decomposition', in Ferraro, M. et al. (Eds.): *Proceedings of 'Soft Methods for Data Science'*, Advances in Intelligent Systems and Computing, Springer, Rome, Vol. 456.

Scott, D.W. (2015) *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, Hoboken, New Jersey.

Sharma, P. and Singh, J. (2015) 'Determinants of tax-revenue in India: a principal component analysis approach', *International Journal of Economics and Business Research*, Vol. 10, No. 1, pp.18–29.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis, Mon. on Statistics and Probability*, Chapman & Hall, London.

Tani, S.M., Hacene, I.B. and Bessaid, A. (2016) 'A new algorithm for medical images indexing based on wavelet transform and principal component analysis', *International Journal of Biomedical Engineering and Technology*, Vol. 22, No. 2, pp.126–136.

Taylor, J.W. and Jeon, J. (2015) 'Forecasting wind power quantiles using conditional kernel estimation', *Renewable Energy*, Vol. 80, pp.370–379.

Van Erven, T. and Harremoes, P. (2014) 'Rényi divergence and Kullback-Leibler divergence', *IEEE Transactions on Information Theory*, Vol. 60, No. 7, pp.3797–3820.

Vasconcelos, M.H. et al. (2016) 'Online security assessment with load and renewable generation uncertainty: the iTesla project approach', Paper presented at the *PMAPS Conference*, Beijing.

Yin, S., Li, X., Gao, H. and Kaynak, O. (2015) 'Data-based techniques focused on modern industry: an overview', *IEEE Transactions on Industrial Electronics*, Vol. 62, No. 1, pp.657–667.

Zghal, R., Ghorbel, A. and Triki, M. (2015) 'The conditional dependence structure of banking sector credit default swap indices', *International Journal of Financial Markets and Derivatives*, Vol. 4, Nos. 3/4, pp.273–298.

Zhang, L., Liu, J., Liu, Y., Yan, Z., Xu, L. and Wu, Y. (2015) 'Static security risk assessment of power system considering wind speed correlation', *Electric Power Automation Equipment*, Vol. 35, No. 4, pp.84–89.

Zolfaghari, S., Riahy, G.H. and Abedi, M. (2015) 'A new method to adequate assessment of wind farms' power output', *Energy Conversion and Management*, Vol. 103, pp.585–604.