

A methodology to evaluate the uncertainties used to perform security assessment for branch overloads



M.H. Vasconcelos^{a,b,*}, C. Gonçalves^{a,c}, J. Meirinhos^a, N. Omont^d, A. Pitto^e, G. Ceresa^e

^a INESC Technology and Science (INESC TEC), Porto, Portugal

^b Faculdade de Engenharia da Universidade do Porto (FEUP), Porto, Portugal

^c Faculdade de Ciências da Universidade do Porto (FCUP), Porto, Portugal

^d Réseau de Transport d'Électricité (RTE), Paris, France

^e Ricerca sul Sistema Energetico (RSE S.p.A.), Milano, Italy

ARTICLE INFO

Keywords:

Forecast uncertainty
Online security assessment
Performance evaluation
Probabilistic forecasting
Power system operation

ABSTRACT

This paper presents a generic framework to evaluate and compare the quality of the uncertainties provided by probabilistic forecasts of power system state when used to perform security assessment for branch overloads. Besides exploiting advanced univariate and multivariate metrics that are traditionally used in weather prediction, the evaluation is complemented by assessing the benefits from exploiting probabilistic forecasts over the current practices of using deterministic forecasts of the system operating conditions. Another important feature of this framework is the provision of parameters tuning when applying flow probabilistic forecasts to perform security assessment for branch overloads. The quality and scalability of this framework is demonstrated and validated on recent historical data of the French transmission system. Although being developed to address branch overload problems, with proper adaptations, this work can be extended to other power system security problems.

1. Introduction

Online security assessment is a continuous decision-making problem for power system operators. This involves checking ahead the impact of plausible contingencies on system operational limits and then, conditioned by the results of a risk-based assessment defined by the contingencies probability and by the potential impacts of these contingencies on the energy not supplied, operators decide on the necessary actions to keep the system in a secure state [1,2]. Usually, this analysis is performed over time horizons including two-days ahead, one-day ahead and intraday operation, based on deterministic forecasts of the grid state. However, these forecasts are nowadays being affected by an increased amount of uncertainty [3,4], which in turn may impact the forecasts of the disturbances' severity. Major sources of uncertainties in power systems operating conditions are the increased penetration of intermittent renewable energy sources, the liberalization of the electricity market, the arising opportunities of demand side

management and energy storage and, last but not least, the online control actions taken by operators to ensure system security (like changing the tap position of phase shifter transformers, topology measures, generation rescheduling and re-dispatch).

In the last years, this rationale triggered several research works aiming to include the uncertainty of system state forecasts, beyond the traditional consideration of deterministic forecasts (i.e. point estimation), in the algorithms of tools aimed to support operators in power system security assessment [4–11]. In particular, the R&D iTesla project¹, co-funded by the European Commission 7th Framework Program (EC FP7), targeted the development of an online dynamic security analysis platform for European-wide grid models, able to account for uncertainties in security margin evaluation and to handle curative remedial actions to face contingencies. The iTesla security assessment approach was successfully tested for overload situations on the French network, as described in [9]. The presented results show that considering forecasted uncertainties is of the utmost importance, since

Abbreviations: CRPS, Continuous Ranked Probability Score; DACF, Day-Ahead Congestion Forecasts; DF, Deterministic Forecast; ES, Energy Score; FA, False Alarm; MA, Missed Alarm; MCLA, Monte Carlo Like Approach; Q_p , marginal Quantile with probability p ; SM, Security Margin; SN, Snapshot; TSO, Transmission System Operator

* Corresponding author.

E-mail address: mhv@fe.up.pt (M.H. Vasconcelos).

¹ iTesla: Innovative Tools for Electrical System Security within Large Areas (<http://www.itesla-project.eu>).

from apparently secure deterministic forecasted network states, it is possible to arise insecure situations that need to be tackled in advance by the system operator. In [6], probabilistic wind power predictions are used to perform steady-state security assessment of the Portuguese transmission system with fuzzy power flow. By comparing them with the results provided by the ac deterministic power flow, the obtained results demonstrate that uncertainty forecasts can reduce the percentage of missed alarm violations detection at the cost of increasing false alarm situations. In [12], an algorithm is presented to perform on-line probabilistic transient stability assessment, claiming that probabilistic assessment is capable of providing a more comprehensive, rational and realistic measure of the system stability level. Although it is clear from these works that the use of uncertainty forecasts provides benefits for system security assessment, a methodology is missing to evaluate the overall quality of the employed probabilistic forecasts.

Aiming to fill this gap, the current work describes a generic framework developed to evaluate and compare the quality of the uncertainty provided by probabilistic forecasts of the power system state, when used to perform security assessment for branch overloading. Besides exploiting advanced univariate and multivariate metrics that traditionally are used in weather prediction, the evaluation is completed by assessing the benefits of exploiting probabilistic forecasts over the Transmission System Operator (TSO) current practices of using deterministic forecasts of the system operating conditions. To the best of authors' knowledge, the proposed validation framework is new, being of paramount importance in order to enable the electric power industry to compare the performance among alternative probabilistic uncertainty models, when used to perform security assessment of their power system. The quality and scalability of this framework is illustrated and validated by presenting the obtained results for the French transmission system. To this end, the proposed evaluation framework was applied for alternative probabilistic uncertainty models that were computed and used by an improved version of the advanced iTesla security assessment platform. The validation framework was implemented through the development of scripts written in the R programming language [13]. These scripts are open source and available on the iTesla Power System Tools (iPST) repository². Some preliminary results obtained with these scripts are described in [14]. The contributions of the current paper are the overall description of the developed methodology, including an explanation and justification for all the adopted steps of the methodology and describing, through the analyzed case study, how to perform a proper interpretation of all the results.

The general organization of the paper is as follows. First, Section 2 describes the univariate and multivariate metrics that are used to evaluate and compare the quality of branch flow uncertainties when provided by probabilistic forecasts. Then, Section 3 explains how the methodology is extended to evaluate the benefits from exploiting these uncertainties to assess security for the overload problem. Next, Section 4 presents the results obtained by applying the described validation framework for the analyzed case study and, finally, Section 5 presents a summary of the main conclusions obtained from this research.

2. Evaluation of flow uncertainty forecasts

The aim of this analysis is to measure the quality of probabilistic uncertainty models in estimating the flows in transmission lines. With this goal, the analyzed variables are the rms value of electric currents in transmission lines (I , in A). Since the evaluation uses metrics that are based on distance computations, in order to have equivalent measures between variables, these are previously normalized by dividing them by the line maximum allowable permanent limit (I_{max} , in A) associated to a single point in time (timestamp) of the evaluated time period. In fact,

although line limits may change over time due to weather conditions, in this analysis a unique base value is mandatory for each variable since it involves the computation of multi-temporal metrics. For each considered normalized electric current (I , in pu A of I_{max}), a distinct statistical analysis is performed for the cases with opposite directions of active power flow, since these two situations define very different operating conditions in the power system.

In this evaluation, probabilistic forecasts are assumed to be formed by a finite number of plausible future states (i.e. random vectors) independent and identically distributed like the ones generated by a Monte Carlo approach. These forecasts are henceforth named *ensembles*. Being the real observed behavior of a testing period provided by recorded snapshots (SN's), the quality of probabilistic forecasts is checked through the computation of advanced metrics widely used and tested in weather prediction [15,16]. In the developed methodology, these metrics were adapted and explored for evaluating branch flow estimations, namely by: (a) analyzing a distinct type of system condition – the normalized electric currents in transmission lines associated with a constant direction of active power flow (i.e. the previously described I in pu A of I_{max}); (b) given special attention for underestimated flow situations, since these are prone to provide missed alarm of overload situations in case of branch flows closer to their limits. These applied metrics are described next.

2.1. Univariate metrics

The goal of probabilistic forecasting is to maximize the *sharpness* of the predictive distributions subjected to *calibration* [16]. Calibration is the statistical consistency between the probabilistic forecasts and observations, diagnosing if they result from the same probability distribution function. Sharpness measures the concentration of the predictive distributions and, therefore, characterizes the uncertainty of probabilistic forecasts by estimating the range of forecast errors.

2.1.1. Univariate rank histogram

The univariate rank histogram, also known as *Talagrand diagram* [17], is a powerful tool for calibration check, since it summarizes the rank position of the verified observations with respect to its ensemble for a testing time period. An ensemble is assumed calibrated if the verifying observation is equally likely to fall into any of the bins. Assuming an ensemble with m members, for each timestamp the member values are ordered and the position (i.e. the rank) of the SN in this ordering is recorded. For instance, the rank will be 1 if the SN is below all the ensemble members and will be $m + 1$ if the SN is above all the m ensemble members. The height of each bin presents the relative number of times when the verified observations fall in the associated rank position over the analyzed time period. For optimal result, each bin of the rank histogram should have the same frequency of observations. Therefore, the rank histogram measures how well the spread of the ensembles represents the true uncertainty of the observations (SN's) and checks the bias of the ensembles. Fig. 1 illustrates some possible shapes of the rank histogram as well as its interpretation for the analyzed variable (I , in pu A).

2.1.2. Discrepancy index (Δ)

Together with each rank histogram, a discrepancy index is calculated to measure the deviation from uniformity in the rank histogram, named Δ index and given by Eq. (1):

$$\Delta = \sum_{i=1}^{m+1} |f_i - 1/(m+1)| \times (m+1) / (2m) \quad (1)$$

where f_i is the observed relative frequency of each bin i . It computes the sum of the deviations from uniformity between the $m + 1$ bins and the optimal result for the relative frequency in each bin (i.e. $1/(m + 1)$). By knowing that the worst result leads to $2m/(m + 1)$, this sum is normalized to obtain $\Delta \in [0,1]$. Lower Δ values means that the bins are

² <https://github.com/itesla/ipst/tree/master/mcla-evaluation>.

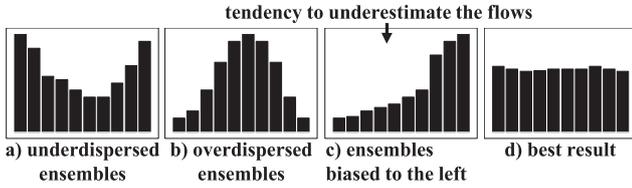


Fig. 1. Interpretation of rank histograms for branch flow estimation.

closer to represent frequencies associated with a perfectly reliable forecast.

It is important to remark that the Δ index value measures the degree of closeness of a Talagrand diagram to its ideal flat shape, without distinguishing between ensemble bias and under/over dispersion, therefore it does not replace the rank histogram.

Because a finite number of observations must be used, rank histograms cannot be expected to be exactly flat. To overcome this, statistical tests for “having a uniform histogram” can be applied, like the ones described in [18]. However, these tests require care in interpretation since they may lose power for a small number of observations and may also fail to detect rank histograms that clearly display a problem (like U-shape or peaked rank histograms).

2.1.3. Frequency of observations falling outside the forecast

The number of situations where the SN falls outside the uncertainty forecast defined by the ensemble can be obtained from the rank histogram, namely from the frequency of observations in bin 1 and $m + 1$. However, these values do not remove the influence of outliers. To eliminate this influence, the ensemble range of values is firstly defined to be inside quantiles p and $1 - p$ (Q_p and Q_{1-p} , respectively). These quantiles are calculated individually for each point in time, being therefore named *marginal quantiles* [19]. Having this in mind, by comparing the observed values with their marginal quantiles in the multi-temporal analysis of each variable, the relative frequency is computed for the following situations: (a) SN outside $[Q_p; Q_{1-p}]$, which for an optimal probabilistic uncertainty model should result in $2p$; (b) SN exceeding Q_{1-p} , quantifying therefore underestimated flow situations.

2.1.4. Continuous Ranked Probability Score (CRPS)

The Continuous Ranked Probability Score (CRPS) can be used to compare probabilistic forecasts of a scalar variable, using an omnibus scoring function that simultaneously addresses calibration and sharpness. In the specific case of forecasts formed by ensembles, the minimal value of this metric aims to identify the model that maximizes ensemble sharpness without deteriorating calibration, being calibration measured by the distance between the observation and the ensemble [16]. Namely, for each timestamp, the CRPS formula in kernel score representation is given by Eq. (2):

$$CRPS = \frac{1}{m} \sum_{i=1}^m |x_i - y| - \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m |x_i - x_j| \quad (2)$$

distance of the SN to ensemble members
distance between ensemble members

where y is the observed value (i.e. the SN) and x_i is the value for member i of the ensemble with m members. Considering the ensemble empirical cumulative distribution, (2) is equivalent to the integral of the squared heights of the shaded region that is illustrated in Fig. 2 for three distinct situations [20]. Lower CRPS values are preferred since it means that the ensembles present lower dispersion and, at the same time, are mostly concentrated around the SN value (as illustrated in the last situation of Fig. 2). The CRPS generalizes the mean absolute error. In fact, it reduces to the mean absolute error if the forecast is deterministic. For the multi-temporal analysis, the mean value of the CRPS metric is computed for each analyzed variable, here named \bar{CRPS} . This metric results from the combination of the following two metrics

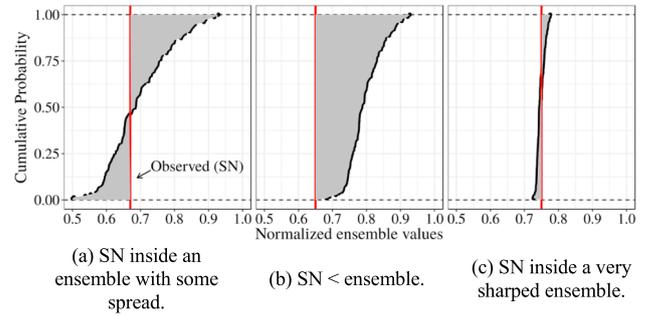


Fig. 2. Illustration of the CRPS: the integral of the squared heights in the shaded region. For convenience, the presented figures shade the absolute values, not the squared heights.

[21]: (a) $\bar{d}_{SN \leftrightarrow ensemble}$: mean of the Euclidean distance of the SN to ensemble members; (b) $\bar{std}_{ensemble}$: mean of the ensemble standard deviation (measuring ensembles spread). The CRPS is useful to rank and compare competing forecasting techniques, but not for the evaluation of a single model since it does not provide a proper interpretation for single evaluation.

2.1.5. Distance of observations falling outside the forecast

Besides the two earlier described distance metrics embraced by the CRPS, it is also important to measure how distant the observed values are when falling outside the ensemble. In particular, security assessment for branch overloads requires to evaluate the mean magnitude of the obtained underestimated flows, because these may provoke missed alarm of overload situations. Given that, and assuming that during online operation the transmission line flows are estimated by the upper value of their marginal quantiles (i.e. by Q_{1-p}), the following metrics (illustrated in Fig. 3) are also computed for the multi-temporal analysis of each variable: (a) $\bar{d}(Q_p)$: mean value of the distances between the SN and quantile p only for situations where the SN does not reach this quantile; (b) $\bar{d}(Q_{1-p}) = \bar{d}_{underestimate}$: mean value of underestimated distances, namely, of the distances between the SN and quantile $1 - p$ only for situations where the SN exceeds this quantile.

2.1.6. Metrics time evolution

Beyond multi-temporal metrics, the time evolution of metrics is also obtained. In fact, their visual inspection is valuable to provide a detailed analysis of the multi-temporal calculated metrics, namely to detect seasonal effects on model performance. In particular, the fan chart [22], like the one illustrated in Fig. 4 for the flow in a transmission line, has become a standard method to visualize forecasts with uncertainty.

2.2. Multivariate metrics

To evaluate the uncertainty quality for the complete network state with a single metric, a multivariate scoring function is adopted. Since the CRPS and the univariate rank histogram measure different aspects of probabilistic forecasts calibration, a multivariate generalization is performed for both metrics. In this work, we followed the multivariate metrics proposed in [16] for ensemble predictions of surface winds which are described next.

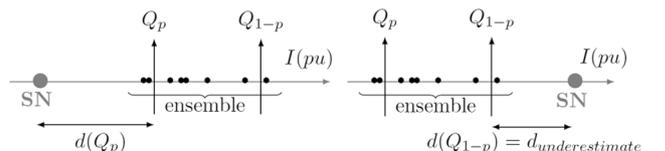


Fig. 3. Graphical representation of distance metrics.

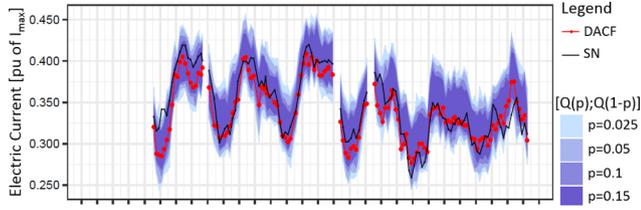


Fig. 4. Illustration of a fan chart for the flow in a transmission line. Legend: SN: observed value; DACF: deterministic forecast; $[Q(p); Q(1-p)]$: marginal forecast intervals.

2.2.1. Multivariate rank histogram

The multivariate rank histogram is a generalization of the earlier described Talagrand diagram. Namely, the same interpretation is applied for calibration check (as illustrated in Fig. 1) and a discrepancy Δ index is also computed to measure the deviation from uniformity, as described in (1).

For rank computation of each verified multivariate observation we consider: (a) vector \mathbf{y} with dimension d and containing the SN operating conditions that take values in \mathbb{R} ; (b) the ensemble with m members $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, where each member \mathbf{x}_i is also a vector taking values in \mathbb{R}^d . First, all the vectors (i.e. the SN and the ensemble members) are pre-ranked using a multivariable procedure: each vector is checked to see if any other vector lies in the hypercube below itself. If none exists, then the corresponding pre-rank is 1. Otherwise, the pre-rank will be the number of vectors contained in the hypercube including itself. Then, the rank position of the SN will be set by counting the number of vectors having a smaller pre-rank. Pre-rank ties between the SN and other vectors are solved at random.

2.2.2. Energy Score (ES)

The Energy Score (ES) is a multivariate generalization of the CRPS described in (2). Namely, for each timestamp, its formula is given by:

$$ES = \frac{1}{m} \sum_{i=1}^m \|x_i - y\| - \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m \|x_i - x_j\| \quad (3)$$

distance of the SN to ensemble members distance between ensemble members

where $\|\cdot\|$ denotes the Euclidean norm and \mathbf{y} and \mathbf{x}_i have the meaning also considered for the multivariate rank histogram. For the multi-temporal analysis, the mean value of the ES is computed, here named \bar{ES} . The \bar{ES} metric results from the combination of the following two metrics: (a) $\bar{d}_{SN \leftrightarrow ens.}$: mean distance of the SN to ensemble members; (b) $\bar{d}_{ens.i \leftrightarrow ens.j}$: mean distance between ensemble members.

3. Evaluation of uncertainty forecasts in assessing security for the overload problem

The ultimate goal of the evaluated probabilistic uncertainty models is to perform security assessment for the overload problem. Therefore, the quality evaluation of each model is only completed after assessing the benefits of exploiting the uncertainty provided by the probabilistic forecasts over the TSO current practices of using deterministic forecasts (DF) of the system operating conditions, when used to perform security assessment for branch overloads. As demonstrated in [6] with fuzzy power flows, using the uncertainty from forecasts can reduce the percentage of missed alarm (MA) violations detection at the cost of increasing false alarm (FA) situations. The idea of the developed evaluation methodology is precisely to check this situation for probabilistic power flows and, if this is true, to measure the obtained *trade-off of FA increase for each 1% decrease in MA* when compared to deterministic power flows. Another contribution of this trade-off analysis is the provision of a proper parameters tuning when using the probabilistic forecasts to perform security assessment for branch overloads.

In order to compute this trade-off, the true security classification

provided by the observed flows (i.e. by $I_{SN}(t)$) for a testing period is compared with the classifications provided by the deterministic and probabilistic forecasts. Then, the following situations are identified for each flow and also for the transmission system (with a distinct analysis for deterministic and probabilistic forecasts): (a) a true insecure situation wrongly classified secure, i.e. a MA (missed alarm); (b) a true secure situation wrongly classified insecure, i.e. a FA (false alarm). The bootstrap method [23] is applied to provide confidence intervals describing the uncertainty for the calculated relative frequencies of MA and FA. Besides, to complement the trade-off results, the computation of more detailed metrics is performed, namely to: (a) provide interpretation for the computed multi-temporal classification errors; (b) evaluate the feasibility of system operators to tackle the forecasted MVA overload situations. This proposed trade-off approach is detailed next.

3.1. Analyzed variables

Like in the methodology described in Section 2, the examined variables are the rms value of electric currents in transmission lines (I , in A). However, since now the computed metrics mainly aim to check if the steady-state flows in each transmission line are not exceeding the maximum permanent limit associated to each specific timestamp ($I_{max}(t)$, in A), a variable normalization is not mandatory. Besides, no distinct analysis is performed for flows with opposite directions of active power.

3.2. Security classification

For each timestamp, each observed flow ($I_{SN}(t)$) is assumed overloaded if exceeding its associated $I_{max}(t)$ value.

A security margin (SM) is assumed for deterministic forecasting (i.e. for $I_{DF}(t)$). Namely, an overload is identified if $I_{DF}(t) \cdot (1 + SM)$ exceeds $I_{max}(t)$. The adoption of this approach was inspired by the market procedures described in [24], where a flow reliability margin (named FRM) is assumed for each critical branch flow to reduce the transmission line available capacity margin (i.e. the difference between the deterministic forecasted flows in the line and its load limit), aiming to take into account the uncertainties involved in the branch flow forecasting process. The approach described in [24] adopts a constant flow margin (FRM) for each branch flow, resulting from the computation of a multi-period quantile of the branch flow historical forecasted errors for a testing period. In contrast, in the proposed SM approach, the adopted flow margin ($I_{DF}(t) \cdot SM$) is a function of the branch flow deterministic forecast (i.e. of $I_{DF}(t)$), providing therefore a sharper model for the uncertainty of deterministic forecasts.

For probabilistic forecasting, assuming that transmission line flows are estimated by the upper value of their marginal quantiles (by Q_{1-p}), an overload situation is identified if Q_{1-p} exceeds $I_{max}(t)$. With this approach, the adopted flow margin ($Q_{1-p} - I_{DF}(t)$) is a function of the branch flow deterministic and probabilistic forecasts.

For each instant in time, the transmission system is assumed insecure if at least one transmission line flow presents overload problems.

3.3. Trade-off analysis

The purpose of the trade-off analysis is to: (a) check if probabilistic forecasts can reduce the transmission system MA rate provided by deterministic forecasts (i.e. by $I_{DF}(t)$ with alternative values of SM); (b) if this is true, to identify the cutoff value (of $1-p$) for probabilistic forecasts being able to decrease the last described MA rate without an excessive increase of FA situations.

In this trade-off analysis, the number of transmission system MA and FA situations is obtained for a testing period. This procedure is repeated for: (a) the traditionally used deterministic forecasts (i.e. for $I_{DF}(t)$ with $SM = 0$); (b) deterministic forecasts with alternative SM values; (c)

probabilistic forecasts with alternative $1-p$ values. For each computed relative frequency of MA and FA, its uncertainty is characterized by applying the bootstrap method with the following approach: (a) random sampling with replacement (by assuming no seasonality dependency in the time series of classification errors for each flow); (b) with the studentized bootstrap [23] to infer confidence intervals for the percentage of MA or FA (i.e. for parameter θ of the bootstrap method). In this approach, an extra challenge lies on the estimation of se_{θ_i} (the standard error for each sample i , in a total of B bootstrap samples, being B selected by the user). This was addressed by considering that the percentage of MA (or FA) in a time series sample is seen as the mean value of a binary sample, where values equal to 1 represent the timestamps where a MA (or FA) occurred. For these cases, a good estimator of se_{θ_i} is provided by the quotient s_{θ_i}/\sqrt{n} , where s_{θ_i} and n are the standard deviation and size of the i -th sample, respectively.

3.4. Computation of detailed metrics

To detect seasonal effects for the computed multi-temporal classification errors, the time evolution of these classification errors is also obtained. This is complemented by the visualization of fan charts for the flows presenting overload problems.

To evaluate the feasibility of system operators to tackle the forecasted overload situations, the transmission lines with forecasted overload problems are identified and the overloads severity for these lines is characterized with location and dispersion metrics. Besides, an overall overload severity metric is computed for the multi-temporal analysis of the transmission system. Namely, assuming that: (a) overload situations detected online by the system operator will provoke the setting of re-dispatch actions; (b) the amount of re-dispatched MW is similar to the total amount of detected MVA overload situations; the following metric is computed to estimate the amount of required MW re-dispatch to solve forecasted overload situations for the analyzed test period:

$$re - dispatch = \sum_t \sum_f (overload_f(t) \cdot Un_f) \cdot \sqrt{3} / 1000 \quad (4)$$

where for each analyzed timestamp t :

$$\begin{cases} overload_f(t) = I_f(t) - I_{max_f}(t) & (\text{if } I_f(t) > I_{max_f}(t)) \\ overload_f(t) = 0 & (\text{otherwise}) \end{cases} \quad (5)$$

and f refers to each analyzed flow in transmission lines; I is the line flow forecast (rms value of electric current, in A); I_{max} is the line maximum permanent limit (in A); Un is the line phase to phase nominal voltage (in kV).

4. Case study

The quality and scalability of the proposed validation framework is here illustrated on recent historical data of the French transmission system. All the results were obtained through the scripts written in the R programming language that are available in the iPST repository.

4.1. Network data

In the used network model, the boundary nodes of the French transmission system to lower voltage levels (20 kV and below) are assumed to be connected to fictitious loads. Foreign grids are represented by equivalents. The stochastic variables comprise the active and reactive power injection of loads and renewable energy sources, totaling thousands of variables (around 8000). The used historical data include Day-Ahead Congestion Forecasts (DACF's) for deterministic forecasts and SN files for the observed network states. In this study, the DACF's were previously modified by assuming the topology of the SN's with the injections of the forecasts. This procedure eliminated the topological

discrepancies between DACF and SN network states from the analysis: a major source of inconsistency between SN's and DACF's of the French transmission system that was not treated by the evaluated probabilistic uncertainty models. To avoid insufficient statistical data, a minimum number of 100 timestamps was assumed to include each variable in the analysis. In all the studies, probabilistic forecasts are formed by ensembles with 50 members (including the DACF).

4.2. Evaluated uncertainty models

The evaluated probabilistic uncertainty models were generated by an upgraded version of the advanced security assessment platform that was created in the iTesla project [9]. The platform builds a model for the forecast errors of the stochastic variables conditioned to their expected values (deterministic forecasts). This task is carried out by a Monte Carlo Like Approach (MCLA) module fully described in [25] and available in the iPST repository³. An overview of the used uncertainty models is described next.

4.2.1. The “original” uncertainty model

In an offline workflow, the raw data (snapshots and deterministic forecasts of the stochastic variables) are pre-treated and clustered by the k-means clustering technique. After applying the Principal Component Analysis (PCA) for dimensionality reduction, and a pair copula decomposition with C-vines to simulate higher order dependencies among Principal Components (PC), the PC's are sampled and back-projected onto the original variable space, getting the unconditioned samples of snapshots (SN) and forecasts (FO). In the online platform, these samples are conditioned to the specific forecast power system state using a conditional sampling based on Nataf transformation, as detailed in [25].

4.2.2. The “adapted” uncertainty model

Applying the conditional sampling techniques in the online environment has highlighted two issues [26]: (1) the occurrence of overfitting of the conditioning technique (i.e. the consistency of conditioned forecast intervals with observations depends a lot on the specific set of historical data used to build the uncertainty model); (2) the original variables may be multimodal, with a non-Gaussian probability distribution, which reduces the consistency of the conditioned samples – extracted using Nataf transformation – with actual observations. Therefore, as detailed in [26], an “adapted” uncertainty model was developed by performing some upgrades to the earlier described “original” uncertainty model, to increase the algorithm prediction capability at the expense of the spread of the relevant samples, namely by:

- (1) reducing the complexity of the covariance matrices used in the Nataf transformation;
- (2) sampling separately the multimodal and the unimodal variables. The pair (SN, FO) of each multimodal injection is fitted using a Gaussian mixture with the lowest AIC (Akaike Information Criterion), while the unimodal variable set is treated as a whole with the Nataf transformation-based conditional sampling method, thus neglecting correlations between multimodal and unimodal variables, and among multimodal variables.

4.3. Evaluation of the “original” uncertainty model

In order to measure the quality of the “original” uncertainty model in estimating the flows in transmission lines, this model was trained by the offline workflow referred in Section 4.2 with the French historical data from February to March 2017 (with 890 hourly timestamps, being the others unavailable). This model was then used by the MCLA (also

³ <https://github.com/itesla/ipst/tree/master/mcla>.

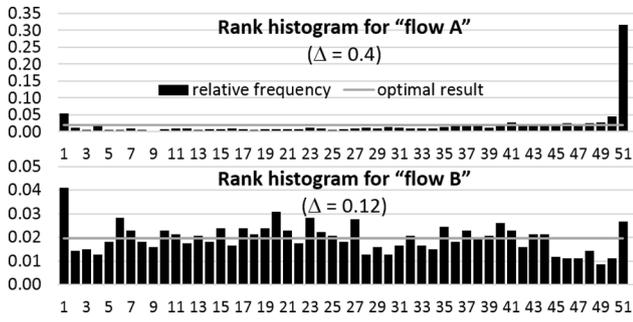


Fig. 5. Rank histograms for “flow A” and “B” with the Δ index value.

referred in Section 4.2), to estimate the uncertainty of pre-contingency power flows from February to April 2017 (1354 timestamps). Finally, these probabilistic forecasts were used to compute the univariate metrics earlier described in Section 2.1 to evaluate the quality of flow uncertainty forecasts for all the 380 kV French lines (for 695 lines providing 1241 flow variables). For this analysis, the ensembles range of values was defined by marginal quantiles $Q_{0.05}$ and $Q_{0.95}$ (i.e. by setting $p = 0.05$). The training period was here intentionally included with the purpose of having a large amount of timestamps to illustrate the quality of the proposed metrics. The most relevant results are described next.

Fig. 5 presents the obtained rank histogram for two analyzed flows (i.e. two electric currents associated with a constant direction of active power flow). The rank histogram for “flow A” reveals that the ensembles characterizing the uncertainty for this flow are underdispersed and biased to the left (i.e. with tendency to underestimate the flows), presenting a large Δ index of 0.4. The rank histogram for “flow B” is much more calibrated (i.e. with more consistency between the position of the observed values and the ensemble members), having a smaller Δ index of 0.12. These behaviors can be confirmed by the fan charts presented in Fig. 6. This example illustrates how well the Δ index accurately summarizes the quality of the rank histogram.

The scatter plots of Fig. 7 present the obtained values of univariate metrics for all the analyzed flows. As expected, the first scatter plot describes a strong relationship between the quality of the rank histogram (summarized by Δ index) and the relative frequency of SN’s falling outside the ensemble. It also shows that the aspects of calibration measured by the rank histogram differs a lot between transmission lines, going from a good result with $\Delta = 0.1$ to the worst possible result of $\Delta = 1$

The second plot indicates that no relevant relationship exists between the quality of the rank histogram and distance metrics (here illustrated for $\bar{d}_{SN \leftrightarrow ensemble}$, the mean Euclidean distance of the SN to ensemble members). On the contrary, the third plot describes a strong relationship between $CRPS$ and distance metrics (as expected from

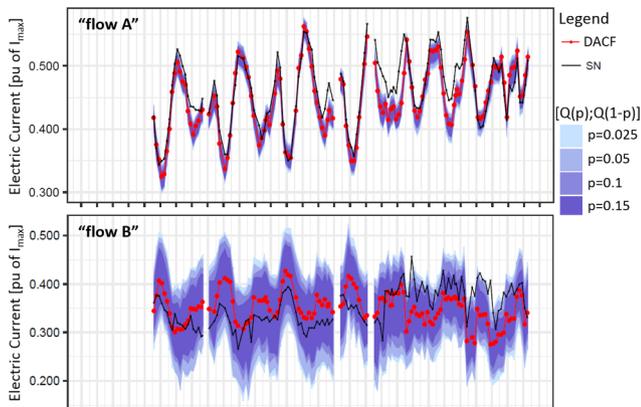


Fig. 6. Fan chart for “flow A” and “B” at week 15 of 2017.

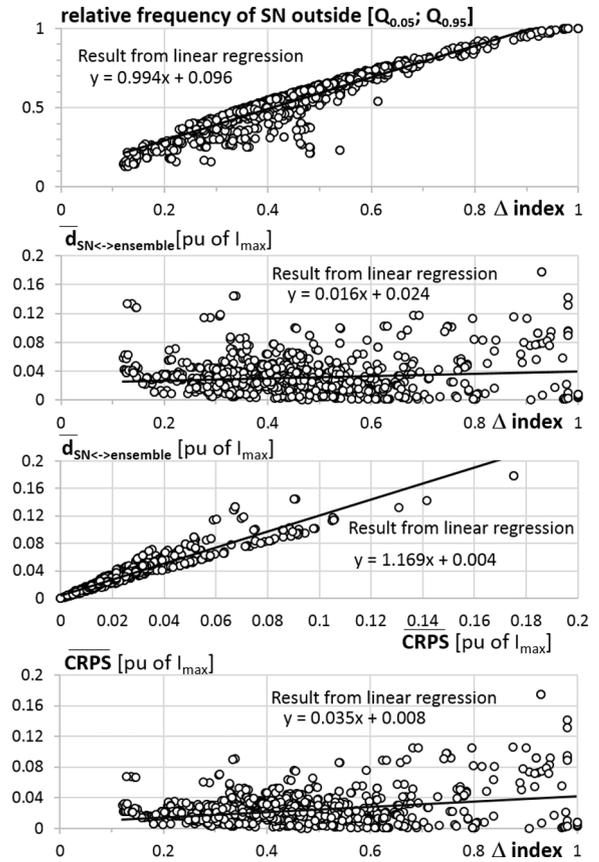


Fig. 7. Univariate metrics for the “original” uncertainty model.

$CRPS$ definition). Finally, the last plot indicates that the rank histogram has no relevant relationship with $CRPS$. By knowing that these two metrics evaluate two distinct aspects of probabilistic forecasts calibration, the obtained results illustrate that to obtain a clear picture of the quality of probabilistic forecasts, we must combine the evaluation provided by the rank histogram and the $CRPS$ metrics, complemented by each associated metrics.

For the specific tested case, the distance metric $\bar{d}_{SN \leftrightarrow ensemble}$ suggests higher calibration than the one associated to the rank histogram. In Fig. 6, this is also visible in the fan chart of “flow A”. By identifying the flows with the highest Δ values (i.e. worst rank histograms), the least calibrated transmission lines were detected to be evacuation lines of conventional power plants, suggesting that their flow probabilistic forecasts may be improved by including the unscheduled operating conditions of conventional power plants in the uncertainty model.

4.4. Comparing uncertainty models

Aiming to compare the “adapted” and “original” uncertainty models (both characterized in Section 4.2), each model was trained with the French historical data from February to March 2017. Both models were then used by the online MCLA to estimate the uncertainty of pre-contingency flows in all the 380 kV French lines for April 2017 (for the available 464 hourly timestamps). Here, it is necessary to remark that, in order to increase the accuracy of the comparison, more months should be used for the testing set besides April 2017. This was not performed due to limitations of data availability. Finally, the univariate and multivariate metrics described in Section 2 were computed assuming that ensembles range of values is defined by $Q_{0.05}$ and $Q_{0.95}$.

The most relevant results obtained with univariate metrics are presented in Fig. 8, showing the statistical difference between the “adapted” and “original” model metrics values for each analyzed flow

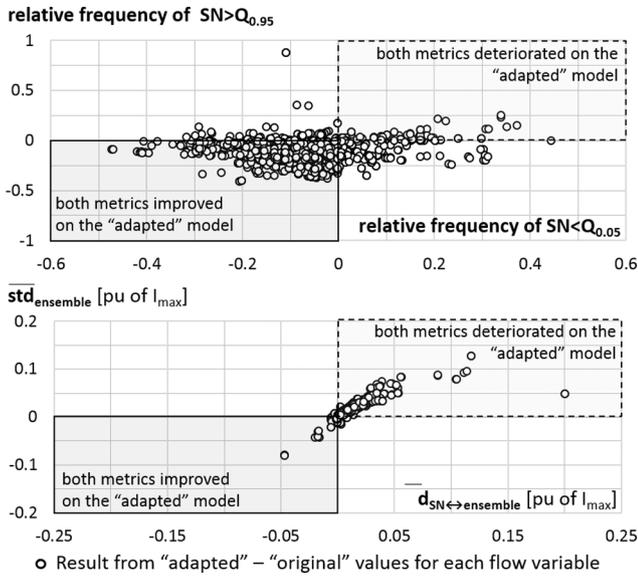


Fig. 8. Comparing uncertainty models using univariate metrics.

in transmission lines. Since the best result aims to minimize all the metrics, in the scatter plots of Fig. 8, a negative value means that the metric is being improved by the “adapted” approach (and vice-versa). The first plot of Fig. 8 indicates that, for most of the flows, the quality of the uncertainty model was improved (from the “original” to the “adapted” approach) in reducing the number of observed flows falling outside the uncertainty forecast. From the second plot, it is visible that this improvement was obtained at the expense of increasing the ensembles spread (measured by $\bar{std}_{ensemble}$) and the distance between the observed values in the SN’s and the ensemble members (measured by $\bar{d}_{SN \leftrightarrow ensemble}$).

The obtained results with multivariate metrics are presented in Fig. 9, where the relevant conclusions are similar to the ones obtained with the univariate metrics. In particular, in Fig. 9 we can see that the calibration aspects measured by the rank histogram are improved in the “adapted” model (from Δ index values and by visual inspection of the rank histograms), at the cost of deteriorating the calibration aspects evaluated by the energy score (from the values of \bar{ES} , $\bar{d}_{SN \leftrightarrow ens.}$ and $\bar{d}_{ens.i \leftrightarrow ens.j}$). This test case illustrates the discriminative capabilities of the used multivariate metrics.

4.5. Evaluation of the “adapted” uncertainty model in assessing security for the overload problem

In order to evaluate the benefits from exploiting the “adapted”

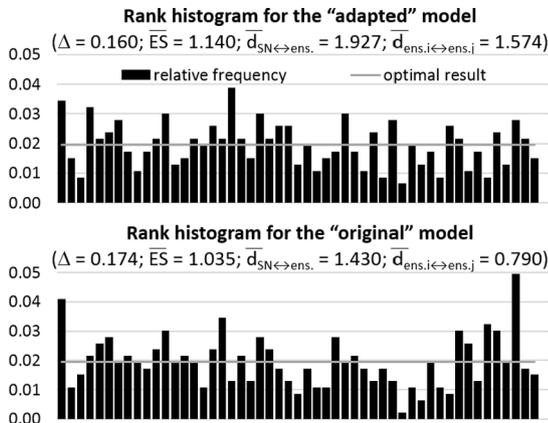


Fig. 9. Comparing uncertainty models using multivariate metrics.

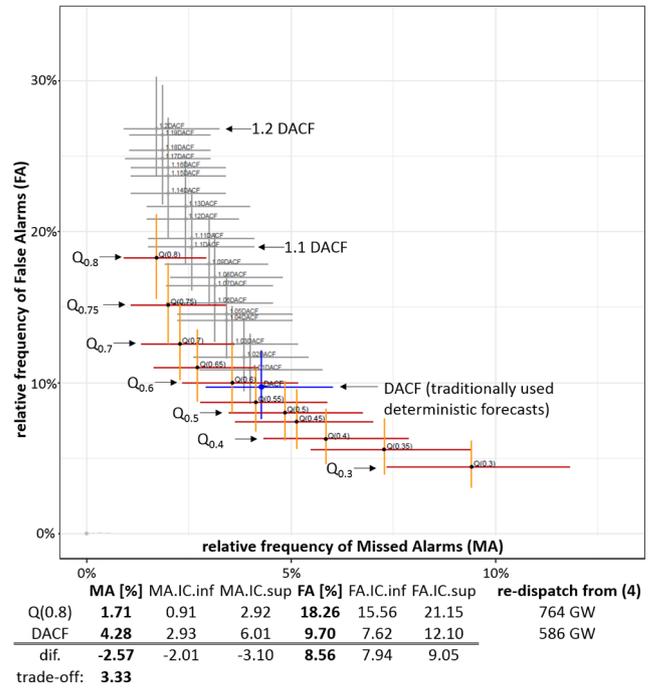


Fig. 10. Trade-off analysis for “contingency 1”. Legend: (1 + SM) DACF: results provided by deterministic forecasting with a SM security margin; Q_{1-p} : results provided by probabilistic forecasting with a $1 - p$ marginal quantile; trade-off = (FA increase)/(MA decrease).

probabilistic uncertainty model to assess security for the overload problem, the trade-off methodology described in Section 3 was applied for two severe N-2 contingencies. One, here named “contingency 1”, impacts an electric peninsula with very low level of conventional generation in the analyzed time period and, therefore, provides a test case adapted to the capability limits of the used uncertainty model. The other situation, named “contingency 2”, has major impacts on the security of transmission lines close to cross border flows. Since cross border flows (modeled as fictitious loads) were not included on the set of stochastic variables of the uncertainty model, a worse uncertainty characterization is expected for “contingency 2”. In this analysis, the “adapted” uncertainty model was trained with historical data from January to February 2018. This model was then used, by the MCLA, to estimate the uncertainty of post-contingency flows in all the 380 kV and 225 kV French lines (around 1000 lines leading to 4000 flows) for March 2018 (with 701 available hourly timestamps). Finally, these results were used to compute the metrics described in Section 3, namely to produce the trade-off analyses presented in Figs. 10 and 11 for “contingency 1” and “2”, respectively.

These figures include a plot presenting the impacts, on the relative rate of MA and FA for the French EHV/HV transmission system, of using alternative cutoff values (of $1 - p$) on the probabilistic forecasts used for the testing period (March 2018). For comparative purposes, the plot also presents the misclassification results obtained from the traditionally used deterministic forecasts (DACF with $SM = 0$) and with DACF’s with alternative SM values. A 95% confidence level is assumed for the bootstrap confidence intervals (presented on the magnitude of the cross lines in the plots and on $CI.inf$ and $CI.sup$ values in the tables). A maximum value of 20% load was assumed for the DACF’s security margin (i.e. a maximum of $SM = 0.2$ pu). The plot in Fig. 10 shows that, at a 95% confidence level, for “contingency 1” there is no SM value that enables deterministic forecasting to decrease the MA rate provided by the traditionally used DACF’s (with $SM = 0$). On the contrary, this is achieved by probabilistic forecasting if using a proper quantile value, namely a minimum of $Q_{0.8}$. This presents an added value of using probabilistic forecasts over deterministic forecasts. In particular, as

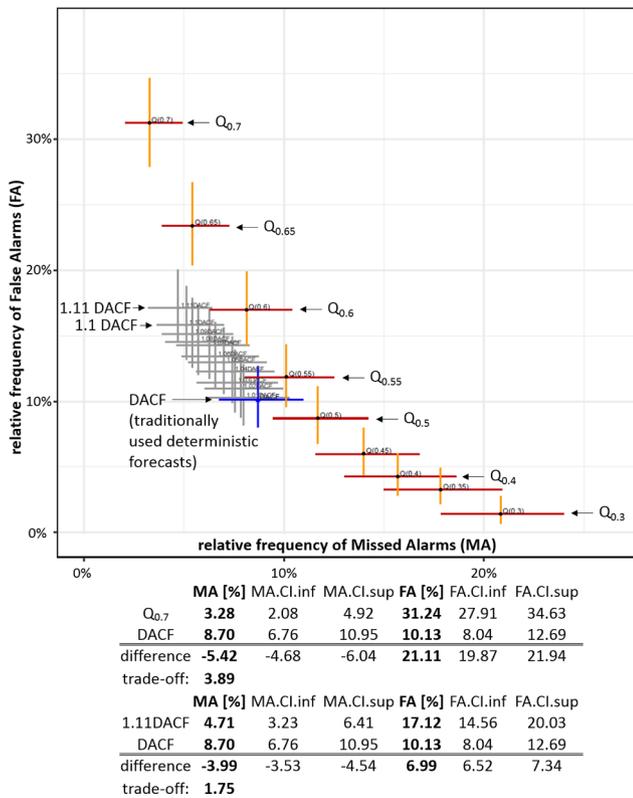


Fig. 11. Trade-off analysis for “contingency 2”.

described in the table of Fig. 10, if one uses probabilistic forecasts with $Q_{0.8}$, false alarms only increase by 3.3% for each 1% decrease in missed alarms. The table in Fig. 10 also presents the obtained values of re-dispatch from (4), namely, the estimate amount of required MW re-dispatch to solve the forecasted overload situations for the analyzed test period (with 701 hourly timestamps). For “contingency 1”, we can see that these re-dispatch requirements are increased if probabilistic forecasts are used (from 586 GW to 764 GW). However, this gives around 1.1 GW of required re-dispatch for each analyzed timestamp, which is in fact feasible for the French power system. Besides, most of the overloads appear in transmission lines which flows can be controlled by phase-shift transformers. Therefore, these results show that using the “adapted” uncertainty model for flow security assessment of “contingency 1” will create forecasted MVA situations that can be tackled by system operators. Therefore, starting to use this probabilistic uncertainty model over the traditionally used deterministic forecasts for flow security assessment of “contingency 1” is a choice of the TSO, according to their adopted risk criteria.

The obtained results for “contingency 2” leads to a different conclusion about the use of this probabilistic uncertainty model. In fact, as presented in Fig. 11, deterministic forecasts show to provide the best trade-off between MA and FA classification errors. In particular, if probabilistic forecasts with $Q_{0.7}$ are adopted instead of the traditionally used deterministic forecasts, the MA rate decreases (at a 95% confidence level) with the following trade-off: for each 1% decrease in missed alarms, false alarms increase by around 3.89%. However, a better trade-off of 1.75% is obtained by using deterministic forecasts with $SM = 0.11$. These results indicate that the used probabilistic uncertainty model still requires some improvements, to increase the accuracy of flow branch forecast for “contingency 2”. This conclusion is supported by the plot presented in Fig. 12.

Fig. 12 presents a plot with the values obtained for the main univariate metrics (described in Section 2) to measure the quality of the “adapted” uncertainty model, when used to estimate pre-contingency flows for the transmission lines that most contribute for the

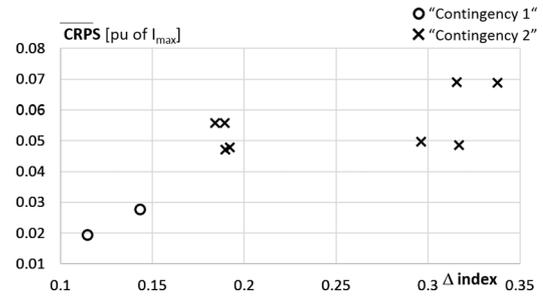


Fig. 12. Univariate metrics for the “adapted” probabilistic uncertainty model (variables: pre-contingency flows of the transmission lines that most contribute for the misclassification errors of the analyzed contingency; testing period: March 2018).

misclassification errors of “contingency 1” and of “contingency 2”, during March 2018. From this plot, it is evident that the flow uncertainty forecasts of the transmission lines associated with “contingency 1” are clearly the most calibrated ones.

5. Conclusions

In this work a generic framework methodology is proposed to evaluate and compare the quality of the uncertainties provided by probabilistic forecasts of the power system state, when used to perform security assessment for branch overloads. Besides evaluating the quality of the generated branch flow uncertainties with advanced univariate and multivariate metrics that traditionally are used in weather prediction, the evaluation is completed by assessing the benefits from exploiting probabilistic forecasts over the TSO current practices of using deterministic forecasts of the system operating conditions.

Another important feature of the proposed framework is the tuning of the marginal quantiles that provide the best misclassification errors trade-off, when using probabilistic forecasts to perform security assessment for branch overloads.

The quality and scalability of the proposed evaluation framework is illustrated on recent historical data of the French transmission system. In particular, it is shown that in order to obtain a clear picture of the quality of probabilistic forecasts, the evaluations provided by the univariate rank histogram and the CRPS (Continuous Ranked Probability Score) advanced metrics must be combined. The obtained results also illustrate how the interpretability of these advanced metrics can be complemented by using simpler associated univariate metrics. Besides, the presented case study also shows the discriminative capabilities of the used multivariate metrics, namely of the multivariate rank histogram and of the ES (Energy Score). Furthermore, the added value of using probabilistic forecasts over deterministic forecasts is demonstrated for a severe “N-2” contingency situation.

The proposed validation framework was implemented through the development of open source scripts written in the R programming language that are available on the iPST repository. Although being developed to address branch overload problems, this work with proper adaptations can be extended to other power system security problems.

Acknowledgments

The authors gratefully acknowledge TechRain S.p.A. for the technical support in using the advanced platform that was further developed after the iTesla project.

This work was supported by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation – COMPETE 2020 Programme within project «POCI-01-0145-FEDER-006961», and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/

50014/2019.

References

- [1] McCalley J et al. Probabilistic security assessment for power system operations. In: Proc of 2004 IEEE power engineering society general meeting, vol. 1; 2004. p. 212–20. <http://doi.org/10.1109/PES.2004.1372788>.
- [2] ENTSO-E. Continental Europe operation handbook – Policy 3 operational security; March 2009.
- [3] Heylen E, Labeuw W, Deconinck G, Van Hertem D. Framework for evaluating and comparing performance of power system reliability criteria. IEEE Trans Power Syst 2016;31(6). <https://doi.org/10.1109/TPWRS.2016.2533158>.
- [4] Duan Y, Zhang B. Security risk assessment using fast probabilistic power flow considering static power-frequency characteristics of power systems. Int J Electr Power Energy Syst 2014;60:53–8. <https://doi.org/10.1016/j.ijepes.2014.02.030>.
- [5] Panciatici P et al. Security management under uncertainty: from day-ahead planning to intraday operation. In: Proc 2010 IREP symposium-bulk power system dynamics and control – VIII, Buzios, Brazil; 2010. <http://doi.org/10.1109/IREP.2010.5563278>.
- [6] Bessa RJ, et al. Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study. IEEE Trans Sustain Energy 2012;3(4). <https://doi.org/10.1109/TSTE.2012.2199340>.
- [7] Ciapessoni E, Cirio D, Grillo S, Massucco S, Pitto A, Silvestro F. An integrated platform for power system security assessment implementing probabilistic and deterministic methodologies. IEEE Syst J 2013;7(4):845–53. <https://doi.org/10.1109/JSYST.2012.2190697>.
- [8] Fliscounakis S, Panciatici P, Capitanescu F, Wehenkel L. Contingency ranking with respect to overloads in very large power systems taking into account uncertainty, preventive and corrective actions. IEEE Trans Power Syst 2013;28(4):4909–17. <https://doi.org/10.1109/TPWRS.2013.2251015>.
- [9] Vasconcelos MH et al., Online security assessment with load and renewable generation uncertainty: the iTesla project approach. In: Proc PMAPS 2016 conference, Beijing, China; 2016. <http://doi.org/10.1109/PMAPS.2016.7764166>.
- [10] Fliscounakis S, Djelassi H, Mitsos A, Panciatici P. Robust optimization taking into account forecasting errors and corrective actions. In: Proc 10th bulk power systems dynamics and control symposium (IREP 2017), Espinho, Portugal; 2017.
- [11] Pitto A, Ciapessoni E, Cirio D, Omont N, Vasconcelos H, Carvalho L. An advanced platform for power system security assessment accounting for forecast uncertainties. Int J Manage Decis. Making (IJMDM) 2019. <https://doi.org/10.1504/IJMDM.2019.10015533>. (in press).
- [12] Abapour M, Haghifam MR. Probabilistic transient stability assessment for on-line applications. Int J Electr Power Energy Syst 2012;42(1):627–34. <https://doi.org/10.1016/j.ijepes.2012.03.025>.
- [13] Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
- [14] Vasconcelos MH, Gonçalves C, Meirinhos J, Omont N, Pitto A, Ceresa G. Evaluation of the uncertainties used to perform flow security assessment: a real case study. Submitted to 13th IEEE PES PowerTech conference; 23–27 June 2019. [unpublished].
- [15] Candille G, Talagrand O. Evaluation of probabilistic prediction systems for a scalar variable. Q J R Meteorolog Soc 2005;131:2131–50. <https://doi.org/10.1256/qj.04.71>.
- [16] Gneiting T, Stanberry LI, Grit EP, Held L, Johnson NA. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. TEST 2008;17:211–35. <https://doi.org/10.1007/s11749-008-0114-x>.
- [17] Talagrand O, Vautard R, Strauss B. Evaluation of probabilistic prediction systems. Proc. of workshop on predictability held at the European Centre for Medium-Range Weather Forecasts (ECMWF). 1997. p. 1–25.
- [18] Elmore KL. Alternatives to the chi-square test for evaluating rank histograms from ensemble forecasts. Weather Forecast 2005;20(5):789–95. <https://doi.org/10.1175/WAF884.1>.
- [19] Bessa RJ, et al. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. Energies. Energies – Electr Power Energy Syst Sec 2017;10(9). <https://doi.org/10.3390/en10091402>.
- [20] Bjørnland HC, Gerdrup K, Jore AS, Smith C, Thorsrud LA. Weights and pools for a Norwegian density combination. N Am J Econ Finance 2011;22:61–76. <https://doi.org/10.1016/j.najef.2010.09.001>.
- [21] Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast 2000;15(5):559–70. [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- [22] Pinson P, Nielsen HAA, Møller JK, Madsen H, Kariniotakis GN. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy 2007;10(6):497–516. <https://doi.org/10.1002/we.230>.
- [23] Davison AC, Hinkley DV. Bootstrap methods and their application. In: Cambridge series in statistical and probabilistic mathematics, vol. 1. Cambridge University Press; 1997. <http://doi.org/10.1017/CBO9780511802843>.
- [24] Amprion, APG, CREOS, ELIA, 50Hertz, EPEXSPOT, RTE, TENNET, TransnetBW. Documentation of the CWE FB MC solution; September 2017. Available: <http://www.jao.eu>.
- [25] Ciapessoni E, Cirio D, Pitto A, Omont N. Forecast uncertainty modeling and Data Management for a cutting-edge security assessment platform. In: Proc PMAPS 16 conference, Beijing, China; 2016. <http://doi.org/10.1109/PMAPS.2016.7764164>.
- [26] Ceresa G, Ciapessoni E, Cirio D, Pitto A, Omont N. Verification and upgrades of an advanced technique to model forecast uncertainties in large power systems. In: Proc PMAPS 18 conference, Boise, Idaho; June 2018. <http://doi.org/10.1109/PMAPS.2018.8440482>.