Task Overview

EPOC task 2.2.6 aimed at developing models and methods suitable for assessing the security of the future electricity supply through the High Voltage (HV) electricity transmission network. This essentially amounts to assessing whether the electrical system can still function without any involuntary interruption of power supply to its end-users while potentially facing (i) "fatal" (i.e., inevitable) failures of transmission and generation components, associated with asset reliability and (ii) deviations in the amount of power generated by renewable generation resources and consumed by the system end-users, associated with forecasting errors.

It is useful at this stage to already clarify the complementarity between electric power system security (i.e., the scope of this task) and adequacy (i.e., the scope of EPOC task 2.2.5). A power system is considered to be adequate if it is of sufficient generation, transmission (and distribution) capacity so as to fulfill the end-user demand for electricity under a broad range of credible operational conditions. A power system is considered to be secure, if it can withstand the operational state transitions induced by credible exogenous disturbances (e.g., the fatal failure of generation/transmission components, the intermittency and variability of renewable power generation etc.) without any involuntary interruption in the supply of electricity towards end-users.

The precise scope for power system security assessment within the context of EPOC task 2.2.6. was defined as:

- **Security of Eelectricity Supply indicators**:
    - loss of load indicators which correspond to facing a situation where the electricity supply is disrupted;
    - load margin indicators which correspond to facing a situation where the electric power system can securely supply the end-user demand
- **Temporal horizon**: focus on the operational stages, i.e. given the day-ahead unit commitment decisions (in complementarity with EPOC task 2.2.5);
- **Exogenous uncertainties**: grid component outages/repairs, renewable power injections, extreme weather events;
- **System modeling**: quasi-steady-state, taking into account control room operator actions.

Figure 1 Power system security assessment components

Figure 1 illustrates the modeling components of a `standard' Monte Carlo approach for electric power system security assessment. As shown in this figure, the three main modeling components that are needed to compute electric power system security indicators are:

a) a so-called representative grid, corresponding to the anticipated future bulk power system, and,
b) a set of trajectories, describing in chronological order the operational conditions within which the power system is anticipated to operate (including the potential development of the exogenous disturbances of interest), and,

c) a *Quasi-steady-state* (QSS) simulator, modeling the effect of disturbances on the state of the electric power grid while taking into account the potential control actions to be applied by the power grid operator in response to such disturbances.

The following parts of this report detail the work done and contributions achieved within EPOC task 2.2.6 for these three sub-models. Let us however already acknowledge that the computational burden of the workflow illustrated in Figure 1 can be massive in the context of modern electric power system. Indeed, concerning sub-model a) the European interconnected power system size should not be neglected. Concerning sub-model b) the critical factor is the increasing penetration of renewable power generation which brings about additional variability and intermittency and calls for the consideration of an increasing number of trajectory samples to compute security indicators with acceptable accuracy. Concerning sub-model c) the need to explicitly take into account the potential control actions of the power grid operator implies the use of optimization models (e.g., Optimal Power Flow and Security Constrained Optimal Power Flow) instead of less complex power system steady-state analysis models (e.g., Power Flow approximations). A considerable part of the effort in EPOC task 2.2.6. focused explicitly in resolving the associated computational complexity challenge. More specifically we proposed and investigated the use of Machine-Learning approximations for the power system + operator simulator of sub-model c). This approach is illustrated in Figure 2 and also detailed in the following parts of this report.

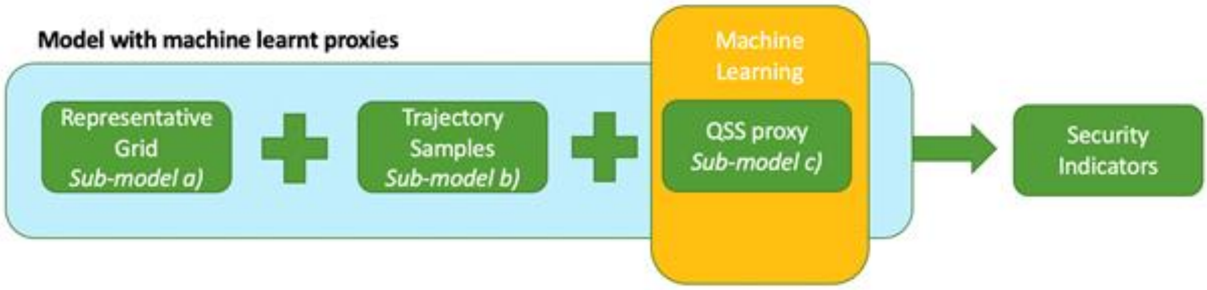


Figure 2 Security assessment using machine learnt proxies

Representative Grid Modeling

In order to define a representative grid model for the future European HV power system, we investigated the usefulness & suitability of several publicly available datasets for quasi-steady-state security assessment studies. After a brief literature survey, we considered in detail the following data sets:

- The RE-Europe dataset published by Jensen & Pinson in 2017,
- The PyPSA model,
- The ENTSOe grid datasets used in the context of the Ten Year Network Development Plan,
- The so-called Belderbos model for the future Belgian power system.

The RE-Europe dataset is a synthetic dataset including a European transmission grid model which consists of 1494 buses (24 of which are located in Belgium) and 2156 transmission lines. Notably, this dataset includes time series of weather driven forecasts and realizations for wind and solar power generation over a period of 3 years and at the nodal resolution. It also includes load realizations for a period of 3 years. However, it relies on a pre-existing model of the transmission grid for Mainland Europe which only includes the transmission capacity ratings for cross-border interconnectors. Not including the thermal ratings of intra-zonal transmission lines (e.g., the transmission capacity inside

Belgium) is a major limitation to the usability of this dataset for the purpose of the EPOC power system security assessment. Indeed, quasi-steady-state security assessment seeks to quantify whether or not power flow can satisfy the thermal ratings of transmission components. We have therefore only retained the wind, solar and RES time series data which we used to develop the uncertainty models described in the following parts of this report.

PyPSA is an open-source network that approximates the European transmission network. The static grid has been derived using GridKit from the ENTSO-E interactive map by approximating location of every component of the network. It results in a static grid model with a localization of the substations. However, this dataset does not include the exact values for line impedances but rather approximations based on line lengths and standard line parameters (ignoring specific conductoring choices for particular lines). Missing the exact line impedance values was considered as a major limitation to the suitability of this dataset for the purpose of the EPOC power system security assessment. Indeed, the grid topology and the line impedances are two main determinants of power flow (given a set of power injections and demands). We have therefore only retained information on the localization of the ENTSO-E grid substations out of this dataset.

The ENTSO-E Ten Year Network Development Plan grid dataset is a detailed snapshot of the anticipated future operation of the European HV grid. Notice that this dataset is not entirely public, but rather made available upon request and under certain confidentiality restrictions. The grid data is provided in the CIM/CGMES format which is the standard format for TSO data exchange. This format is supported by certain commercial power system simulation computational tools, however, at the start of the EPOC project execution there was no available free/open source software to process the available data. We have first developed a set of scripts to import the data from the CIM/CGMES format, build the respective snapshot of the European HV grid and perform power flow computations in the Julia modeling environment.

Due to the realism of these grid data, we have also invested considerable efforts to expand the available single snapshot into a detailed data-set suitable for power system security assessment. In particular, we developed a complete workflow to associate every substation with wind/solar power forecast error models, while preserving spatial-temporal correlations. In brief, the workflow consists of localizing all TYNDP substations (using the information from the PyPSA dataset, automatic identification through an on-line search engine) and attaching location-specific forecast and error models. The location-specific attributes are associated by approximating the forecast error marginal distribution of any TYNDP substation by the distribution of the closest RE-Europe substation and estimating the spatio-temporal covariance matrix using a tree-based model. Figure 3 shows the location of the Belgian grid from the ENTSO-E data set (left) and the association of these substations to a location of the RE-Europe dataset. It can be observed that locations appear to be consistent with respect to the lines between substations.

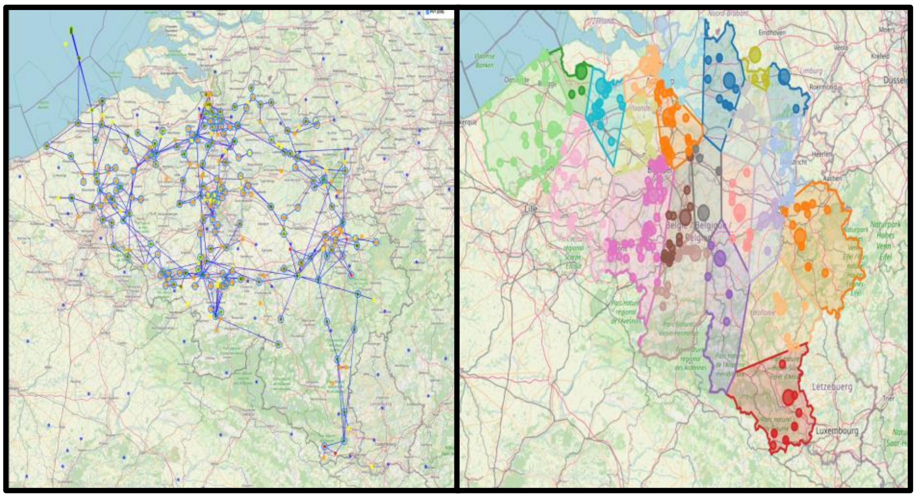


Figure 3 TYNDP substations in Belgium

We have also associated every component of this grid with a reliability model, allowing to simulate the occurrence of fatal outages. While we have collected and/or developed the data/models necessary to perform quasi-steady-state power system assessment on top of the snapshots made available by ENTSO-E, unfortunately we found no source for the economic data of the generation subsystem that would be needed in the task 2.2.5 adequacy assessment methodology. It follows that without the day-ahead unit commitment decisions from task 2.2.5, we could not apply our full assessment methodology on this dataset.

The so-called Belderbos grid is a simplified model of the Belgian HV transmission grid that has been used in a number of scientific publications related to the energy policy. Since this grid was relevant to our colleagues working on EPOC task 2.2.5, we have also applied our complete methodology on this test case.

Exogenous uncertainty modeling

In order to model the relevant uncertainties for probabilistic power system security assessment we have developed trajectory (i.e., sequence of realizations over a day at an hourly resolution) samples. Each trajectory sample specifies, per hour, the random realization of:

- Forced outages and repairs per transmission system component.
- Wind and solar power generation forecast errors per node.
- Load demand forecast errors per node.

Concerning the occurrence of forced outages and respective repairs, we used the classical two-state transition diagram from power system reliability theory shown in Figure 4, along with the most standard assumption that both failures and repairs are Poisson distributed. The gap that we had to fill here concerned the lack of reliability data for all the candidate representative grid models discussed in the previous section. To fill this gap, we collected component reliability data (specifically, forced outage rates and repair rates) from the website of the Belgian Transmission System Operator (ELIA) as well as from well-known and widely used academic benchmarks. We used these data to compute failure and repair probabilities per component and subsequently implemented a Monte Carlo approach to generate a set of hourly availability values per component. Notice that the resulting set of operational situations for the power grid may well include situations where more than a single component is unavailable. In other words, the component availability scenarios included in our trajectory samples for power system security assessment may go beyond the deterministic standard N-1 set of events.

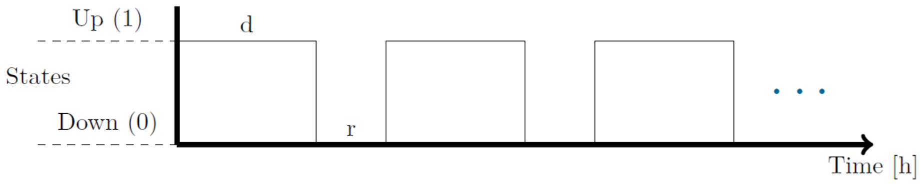


Figure 4 Two-state component availability cycle

Concerning the uncertainty of wind/solar generation and load demand, our goal was to capture the error between day-ahead forecast and real-time realization[1]. Doing so is not trivial especially in the context of studying the future power grid, which would include many additional new renewable power generation resources in new locations (i.e. connected at new substations) for which no historical day-ahead forecast and real-time realization data points are presently available. In order to perform a meaningful power system security assessment, it was important to capture both the spatial and the temporal correlation of the power generation/demand forecast errors so as to be able to identify possible security bottlenecks in the transmission grid and/or the flexibility of the generation subsystem.

We have implemented two alternative techniques, namely the Gaussian Copulas and Normalizing Flows, to generate spatially and temporally correlated realizations of renewable power generation and demand values. In brief, the Gaussian Copula approach models all variables simultaneously by a multivariate Gaussian distribution that generates correlated values. It models the interdependence of N variables by means of N marginal distributions and an NxN covariance matrix. Normalizing Flows belong to the class of deep generative models, that have recently gained considerable interest. Normalizing Flows learn a sequence of invertible transformations (a.k.a. a “flow”) from an analytically known density (for instance, the Normal distribution) to a complex target distribution through maximum likelihood estimation. To the best of our knowledge/understanding, the work we performed on Normalizing Flows in the context of the EPOC project is one of the first power systems applications of this interesting methodology.

The two approaches showed similar performance in our analysis. We have concluded to relying on the Gaussian Copulas as the preferred option for the modelling effort in EPOC task 2.2.6, mostly because of the fact that it was already more widely used in power system applications. We note however that the Normalizing Flows approach models scale differently with the number of locations, and this feature may be interesting to exploit in a different study. For illustration, Figure 5 plots the daily trajectories of wind power forecast values and respective scenarios, generated with the Gaussian Copula approach, for two regions in Belgium and according to publicly available data sourced from the website of the Transmission System Operator (ELIA). Visual inspection reveals that the technique preserve both the temporal correlation and the spatial correlation of forecast errors.

[1] Notice that this is also in line with the complementary relationship of tasks 2.2.5 and 2.2.6. of EPOC. Indeed, while task 2.2.5 focuses on the day-ahead power system operation planning context, task 2.2.6 completes the picture by assessing the impact of the uncertainties between planning and operation.

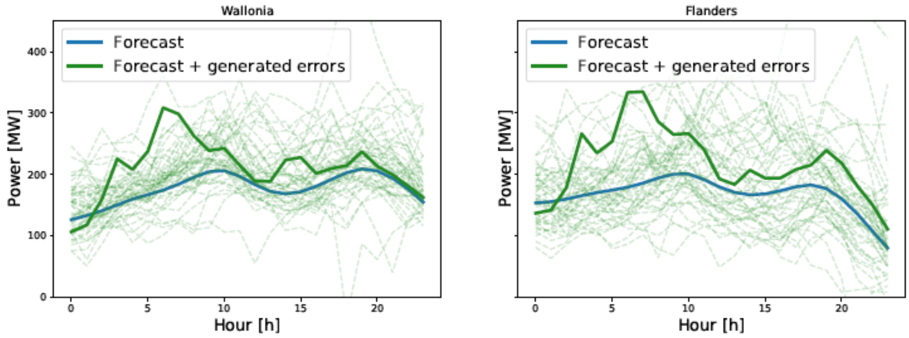


Figure 5 Wind Power Generation Scenarios -- Gaussian copula approach

While the Gaussian Copula approach already has several power system applications, we faced the issue of not having access to any historical data that would allow to estimate the marginal distributions and covariance structure. As mentioned, this is a particular challenge for power system security assessment studies with a long-term horizon. We developed a novel solution approach to this challenge, that is also generally applicable in other power system study context with a long-term horizon (e.g, transmission planning). Our idea was to learn the spatio-temporal covariance between any two different locations from the RE-Europe dataset[2] and apply the learnt model on any other power system data set. In that end, we relied on a tree-based model to learn this relationship using the geographical coordinates, the absolute time and the time difference for a pair of locations and points in time as features of the model. Indicatively, Figure 6 plots the evolution of the performance score (R2 score) with an increasing number of locations with wind measurements on four alternative data subsets. Train is made of pairs of known locations, test is made of pair of unknown locations and val1 and val2 are for sets with only one known locations. Known locations are for locations with past measurements and unknown for those without. Color scale shows the temporal difference between locations. A blue curve corresponds to a model learnt only on data with the same temporal difference. Green curve shows when data with all temporal difference are used simultaneously.

We wish to finally note that we have applied the uncertainty modeling workflow described here in two different power system applications, namely the Belgian instance of the ENTSO-E TYNDP data set as well as the so-called Belderbos grid. For the latter case, we have provided the outputs of our uncertainty workflow (i.e., a set of trajectory samples) along with the Python implementation of the workflow to our colleagues in EPOC task 2.2.5. The trajectory samples concerning the Belderbos grid have also been shared with our colleagues in EPOC task 2.2.7 (market models).

[2] RE-Europe does not provide true past realizations and corresponding forecasts and is therefore not possible to compute the true past forecast errors. Therefore, they are several ways of deriving forecast errors from RE-Europe. One possibility is to compute the difference between COSMO and ECWMF realizations. Doing so, we would probably model the difference between two forecasting tools rather than the error of a single forecasting tool. Since ECMWF and COSMO may not be consistent with each other, this error may be larger than desired. Another possibility is to compute the forecast error at a given time step *t* by taking the difference of the realization (*i.e.*, the most recent forecast for *t*) and the forecast in day-ahead for time *t* (*i.e.*, the forecast made the previous day at 12:00).

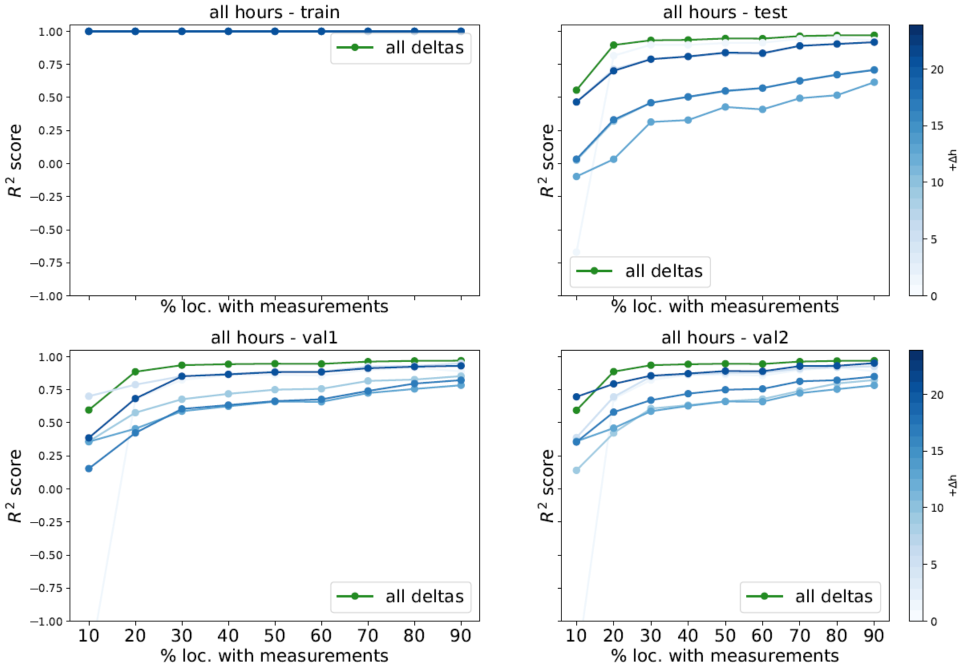


Figure 6 Evaluation of the spatio-temporal approach for covariance learning

Quasi-steady-state simulator implementation

In order to assess any representative grid & set of trajectory samples combination, we have first developed an analytical computational tool. The core module of this tool is the sequential resolution per snapshot of a Security Constrained Optimal Power Flow (SCOPF) problem formulation. The SCOPF formulation models the reaction of the system operator to the realization of component outages & forecast errors in seeking to avoid load shedding actions, while also taking into account the possible occurrence of N-1 contingency events.

The detailed formulation of the SCOPF for security assessment is shown in (1-17). Objective function (1) seeks to minimize the use of load shedding (first term) as well as modifying the charge/discharge schedule of storage units (second term). Coefficients are used as needed to ensure that load-shedding is less preferable than modifying the schedule of storage units and that, if load shedding is used, it is more preferable to do so only after a contingency happens. The problem constraints are the standard constraints under DC power flow approximation and include the nodal power balance, transmission capacities, generation and storage bounds which are set according to the day-ahead operational planning of the system (input data from task 2.2.5).

$$\min_{\mathbf{u},\mathbf{e},\mathbf{w},\mathbf{g},\mathbf{f},\theta}\left\{\sum_{n\in\mathcal{N}}\left[u_{n,0}+\sum_{c\in\mathcal{C}\setminus 0}\frac{u_{n,c}}{|\mathcal{C}|-1}\right]+\frac{0.1}{|\mathcal{C}|}\cdot\sum_{s\in\mathcal{S}}\left[(e^{+}_{s,0}+e^{-}_{s,0})+\sum_{c\in\mathcal{C}\setminus 0}\frac{(e^{+}_{s,c}+e^{-}_{s,c})}{|\mathcal{C}|-1}\right]\right\}, \tag{1}$$

for all nodes $n\in\mathcal{N}$:

$$\sum_{\ell\in\mathcal{L}_n} f_{\ell,0}=\sum_{g\in\mathcal{G}_n}\left[\hat{p}_g+\left(p^{+}_{g,0}-p^{-}_{g,0}\right)\right] + (r_n-w_{n,0})-(d_n-u_{n,0})-\sum_{s\in\mathcal{S}_n}\left[\hat{e}_s-\left(e^{+}_{s,0}-e^{-}_{s,0}\right)\right], \tag{2}$$

for all nodes $n\in\mathcal{N}$ *and contingencies* $c\in\mathcal{C}\setminus 0$:

$$\sum_{\ell\in\mathcal{L}_n} f_{\ell,c}=\sum_{g\in\mathcal{G}_n}\left[\hat{p}_g+\left(p^{+}_{g,c}-p^{-}_{g,c}\right)\right]+(r_n-w_{n,c}) -(d_n-u_{n,0}-u_{n,c})-\sum_{s\in\mathcal{S}_n}\left[\hat{e}_s-\left(e^{+}_{s,0}-e^{-}_{s,0}\right)-\left(e^{+}_{s,c}-e^{-}_{s,c}\right)\right], \tag{3}$$

for all nodes $n\in\mathcal{N}$ *and contingencies* $c\in\mathcal{C}$:

$$f_{\ell,c}=(a_{\ell,c}/X_\ell)\cdot\sum_{n\in\mathcal{N}}\lambda_{\ell,n}\cdot\theta_{n,c} \tag{4}$$

$$-\overline{f}_\ell\le f_{\ell,c}\le\overline{f}_\ell, \tag{5}$$

for all hydro/thermal generators $g\in\mathcal{G}$ *and contingencies* $c\in\mathcal{C}$:

$$v_g\cdot\underline{P}_g\le\hat{p}_g+\left(p^{+}_{g,c}-p^{-}_{g,c}\right)\le v_g\cdot\bar{P}_g, \tag{6}$$

for all RES generators at nodes $n\in\mathcal{N}$ *and contingencies* $c\in\mathcal{C}$:

$$0\le w_{n,c}\le r_n, \tag{7}$$

*f*or all nodes $n\in\mathcal{N}$ *and contingencies* $c\in\mathcal{C}\setminus 0$:

$$0\le u_{n,0}+u_{n,c}\le d_n, \tag{8}$$

$$0\le u_{n,c} \quad , \tag{9}$$

*f*or all storages $s\in\mathcal{S}$ *and contingencies* $c\in\mathcal{C}\setminus 0$,

charging mode – if $\hat{e}_s\ge 0$:

$$0\le {e_{s,0}}^{+}+e^{+}_{s,c}\le\hat{e}_s, \tag{10}$$

$$0\le e^{+}_{s,c}, \tag{11}$$

$$0\le {e_{s,0}}^{-}+e^{-}_{s,c}\le 0, \tag{12}$$

$$0\le e^{-}_{s,c}\le 0, \tag{13}$$

discharging mode – else:

$$0\le {e_{s,0}}^{+}+e^{+}_{s,c}\le 0, \tag{14}$$

$$0\le e^{+}_{s,c},\le 0, \tag{15}$$

$$0\le {e_{s,0}}^{-}+e^{-}_{s,c}\le|\hat{e}_s|, \tag{16}$$

$$0\le e^{-}_{s,c}. \tag{17}$$

$\hat{p}_g$: hydro/thermal power generation dispatch (T2.2.5 inputs);

$\hat{e}_s$: energy storage dispatch (T2.2.5 inputs);

r_n: renewable power generation;

d_n: load demand;

$p_{g,0}^{+/-} \geq 0$: hydro/thermal power generation redispatch;

$e_{s,0}^{+/-} \geq 0$: energy storage redispatch;

$w_{n,0} \geq 0$: renewable power generation curtailment;

$u_{n,0} \geq 0$: load demand shedding;

$f_{\ell,0}$: branch power flow.

$\lambda_{\ell,n}$: branch incidence;

$a_{\ell,c}$: branch availability;

- 0, if $\ell = c$ (N-1 outage);
- 1, else;

X_ℓ: reactance;

$\bar{f}_\ell$: capacity;

$\theta_{n,c}$: voltage angle;

v_g: generator on/off status indicator (T2.2.5 inputs);

$\hat{p}_g$: generator day-ahead dispatch (T2.2.5 inputs);

$\underline{P}_g/\bar{P}_g$: generator min./max. power output bounds;

r_n: realized RES power generation.

In the event that the snapshot is deemed to be secure, i.e. there is no need to resort to load shedding in the base-case or after the occurrence of any N-1 contingency, we also solve a variant of the SCOPF seeking to compute the load margin to insecurity. More specifically, we replace the objective function and modify the nodal power balance constraint so as to maximize the additional system load that is supplied at every node of the system:

$$\max_{\mathbf{m},\mathbf{w},\mathbf{g},\mathbf{f},\theta} \sum_{n \in \mathcal{N}} m_n$$

for all nodes $n \in \mathcal{N}$ *and contingencies* $c \in \mathcal{C}$:

$$\sum_{\ell \in \mathcal{L}_n} f_{\ell,c} = \sum_{g \in \mathcal{G}_n} \left[\hat{p}_g + \left(p_{g,c}^{+} - p_{g,c}^{-}\right)\right] + (r_n - w_{n,0}) - (d_n + m_n) - \sum_{s \in \mathcal{S}_n} \hat{e}_s.$$

We have implemented this simulator in the Julia JuMP environment and while using the CPLEX solver to solve the respective optimization problems. Indicatively, Figure 7 presents the results of our security analysis for the case of the Belderbos system and a selected representative day (for which a Unit Commitment and storage dispatch solution was provided by the Task 2.2.5 team). It can be seen that the system is generally secure. While there should be no problem in the middle of the day (when ample wind and solar power production is available) security issues arise in few scenarios from the early evening onwards.

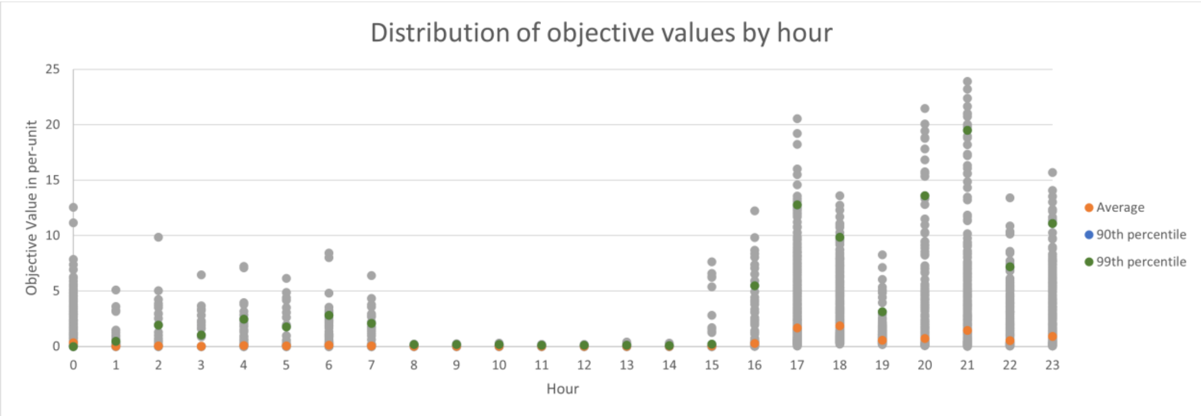


Figure 7 Security Assessment Output

It should be noted here that we processed with our analytical QSS implementation all the test cases that were designed by our colleagues working on the EPOC task 2.2.5, and for which our uncertainty modelling workflow served to prepare sets of trajectory samples. We would like to exemplify the benefits of this integrated working approach, as well as the point of assessing both power system adequacy and power system security, by discussing a particular case of a so-called "representative day" from the Belderbos grid simulations. Figure 8 displays the security assessment output for the particular test case. The large spike concerning period 23 is apparent, especially in contrast to all other periods of the day wherein we spotted practically no security problem. Closer inspection revealed that the Unit Commitment provided by our colleagues in task 2.2.5, while ensuring an adequate power system, did not retain any dispatchable generator online for period 23. This was suspected to be the reason for the identified high risk of insecurity.

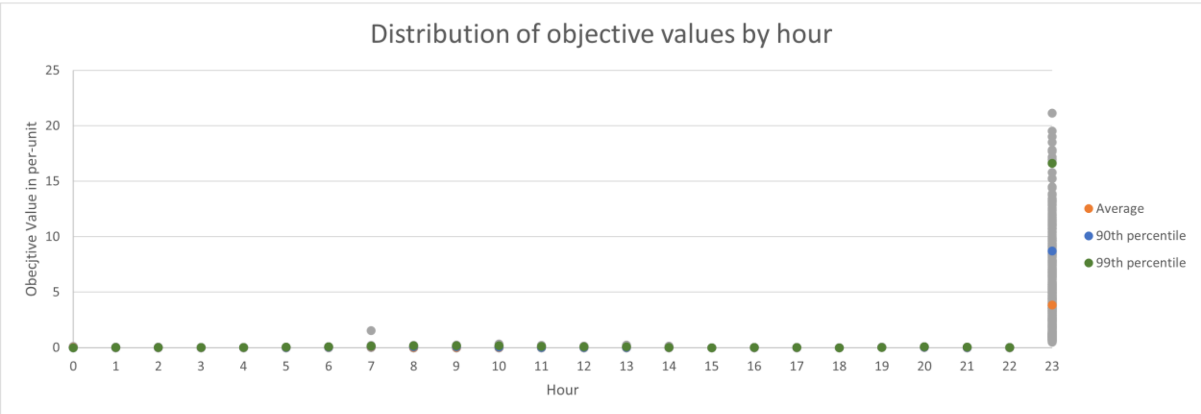


Figure 8 Security Assessment Output - Insecure instance

Machine Learning Approximations of the QSS implementation

In the next stage of this research, we investigated whether it is possible to replace the analytical SCOPF solution process (which can be computationally expensive for large scale power systems) with machine learnt approximations. The potential advantage of doing so is obvious, taking into consideration that a larger & larger number of trajectory samples needs to be processed in order to accurately reflect the increasing contribution of renewable power generation resources in the future energy mix.

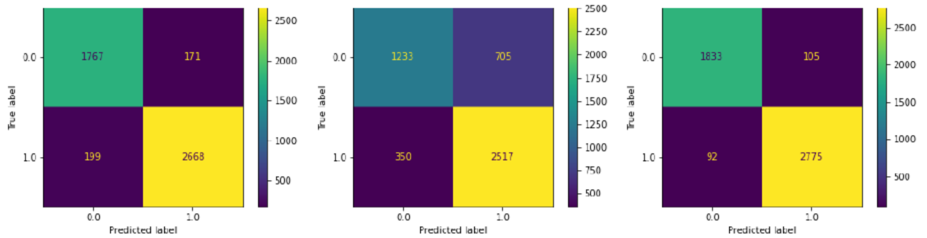
We tested both whether machine learning algorithms can quickly classify a power system snapshot as "secure/insecure" (the latter can be identified with a non-zero value for objective function (1) indicating the necessary use of load shedding and/or modification of the energy storage dispatch) as well as whether machine learning algorithms can predict the numerical value for the security indicators we considered with reasonable accuracy. We further performed a so-called "feature importance" analysis seeking to identify which input data (a.k.a. features) are most influential towards the predicted output. This latter analysis can be very useful in power system security assessment applications as it offers insight into the factors that determine the security status of a given power system situation. The Machine learning algorithms that we tested are summarized in Table 1. We used two different approaches to organizing the training data, by using either the forecast errors for power generation and demand (i.e., deviations from the forecast values used to derive the unit commitment and base-case economic dispatch) or the absolute quantities of realized power generation and demand.

Table 1 Machine Learning Algorithms

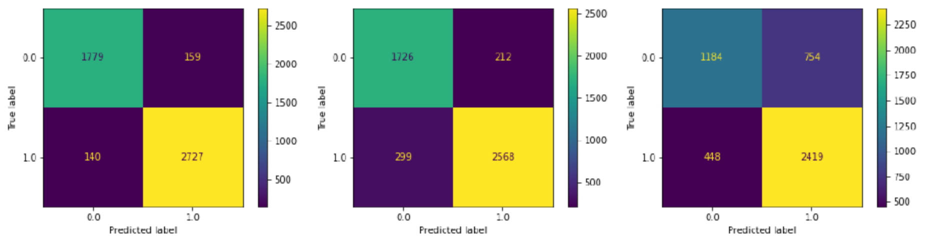
Random Forests	Random forest is an ensemble method that builds a series of trees. It uses bootstrap sampling to associate each tree to a unique learning sample. Each tree is then built by using random feature subset selection. Instead of choosing the best split among all features, like a classic decision tree, the split is chosen from a subset of k features. The value of k needs to be set as a hyperparameter. A smaller value of k leads to a bigger variance decrease at the cost of an increase in bias.
Extra Trees	Extra Trees is a method similar to random forest as it also selects a random subset of features to perform the split on. However, instead of choosing the split point that decreases the most the impurity, the split point is chosen randomly for each feature, and the best split among all features is chosen. This increases the randomness compared to random forests, which usually leads to a slight increase in bias but a decrease in variance.
Linear Regression	Linear regression is a simple weighted linear model that minimizes the sum of squares between the observed and the predicted values while logistic regression computes a linear classification threshold. The output is expressed as a weighted sum of the input variables. As the name implies, linear regression is limited to produce linear models. Due to the limited expressive capabilities of such a model, state of the art performance is not expected, but it should be interesting to compare how this simple model compares to more complicated ones, like ensemble of trees and neural networks.
Multilayer Perceptron	The multilayer perceptron is an architecture composed of layers where each layer is comprised of nodes. Each node in a layer is the result of a weighted sum of nodes in the previous layer. We talk of fully connected feedforward networks when all nodes in a layer are connected to all nodes in the next layer. Training a neural network consists of performing gradient descent to learn

	good weight values between nodes of consecutive layers. Usually, activation functions are used to break the linearity and allow the MLP to express nonlinear dependencies.

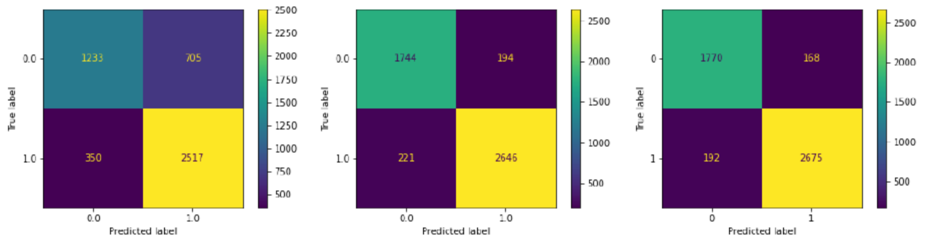
Once again, we tested this approach on all test cases that were designed by our colleagues working on the EPOC task 2.2.5, including the insecure instance shown in Figure 8. We first present an indicative result for the classification task, by means of a so-called “confusion matrix” in Figure 9. The caption of sub-figures describes the corresponding machine learning algorithm. The additional keyword “time” indicates that the specific period of the day was presented as an additional input to the algorithms while the additional keyword “features” indicates that additional variables (e.g., the sum of the nodal load demand, etc.) were also presented as inputs to the algorithms. It can be seen that all tested algorithms achieve good performance on the classification tasks with relatively low misclassification instances.



(a) Forest Basic Absolute (b) Forest Basic Deviations (c) Forest Time Features Absolute



(d) Forest Time Features Deviations (e) Extra Basic Absolute (f) Extra Basic Deviations



(g) Extra Time Features Absolute (h) Extra Time Features Deviations (i) MLP basic absolute

Figure 9 Confusion matrix

Feature importance analysis also revealed valuable insight both regarding the performance of the algorithms as well as the security of the electric power grid. It showcased that the time period is a most important input variable for all alternative machine learning algorithms. The example introduced in Figure 8 is useful to understand why this is the case. Blending snapshots from period 23 with snapshots from other periods would make the task of predicting the security assessment more challenging. Specifying the period as an additional feature in the data set makes the prediction task easier. The finding is also sensible from a power system perspective, since as explained a different set of dispatchable generating units is available in different periods of the day. Therefore, the time period is a useful piece of information to predict whether or not the power system will be secure. Finally, Table 2 presents an indicative result on the performance of the different machine learning methods. It can be seen that the Extra Trees algorithm with the additional input variables slightly outperforms all other methods.

Table 2 Prediction Performance Analysis

Model	Data	Values	Mean Squared Error	R2 Score
Random Forest	Basic	Absolute	0.379733811	0.8167131
		Deviation	0.581032674	0.719551764
	Features	Absolute	0.276349765	0.866613691
		Deviation	0.571131357	0.724330853
	Time	Absolute	0.144980706	0.930021864
		Deviation	0.129463471	0.9375116
	Time / Features	Absolute	0.13114375	0.936700576
		Deviation	0.131535028	0.936511717
ExtraTrees	Basic	Absolute	0.234034669	0.887038005
		Deviation	0.5868895	0.716724837
	Features	Absolute	**0.215895248**	**0.895793397**
		Deviation	0.576514586	0.721732519
	Time	Absolute	**0.089706922**	**0.956700975**
		Deviation	0.143872871	0.930556585
	Time / Features	Absolute	**0.079740149**	**0.961511658**
		Deviation	0.142102105	0.931411284
Linear Regression	Basic	Absolute	1.766060799	0.1475718
		Deviation	1.742061735	0.159155478
	Features	Absolute	1.764935974	0.148114722
		Deviation	1.742808771	0.158794905
	Time	Absolute	1.71072047	0.174283031
		Deviation	1.632839843	0.211873833
	Time / Features	Absolute	1.710137572	0.17456438
		Deviation	1.633427139	0.211590362
Multi Layer Perceptron	Basic	Absolute	**0.21204334**	**0.897652595**
		Deviation	0.7531615	0.636470007
	Features	Absolute	0.24318098	0.882623331
		Deviation	0.6935733	0.665231567
	Time	Absolute	0.19361906	0.906545479
		Deviation	0.4166859	0.798877343
	Time / Features	Absolute	0.24176078	0.883308823
		Deviation	0.6461919	0.688101206

Publications and documentation

The work of the ULiege team on the representative grid development and uncertainty modeling is documented in an internal report that was shared between the EPOC project partners. While this report contains all details that are necessary to reproduce the ULiege uncertainty model, additional documentation can be found in the comments of the code that has also been shared with EPOC colleagues. The contributions of Antonio Sutera related to uncertainty modeling also lead to the following two publications, which should be cited when using the ULiege scenario samples.

[1] Dumas, J., Lanaspeze, D., Wehenkel, A., Cornélusse, B., & Sutera, A. (2021). Diepe generatieve modellering voor probabilistische voorspellingen in energiesystemen. Geaccepteerd in Toegepaste Energie.
[2] Dumas, J., Cointe, C., Wehenkel, A., Sutera, A., Fettweis, X., & Cornélusse, B. (2021). Een op probabilistische voorspellingen gebaseerde strategie voor een risicobewuste deelname aan de markt voor capaciteitsversterking. arXiv preprint arXiv:2105.13801. Ingediend bij IEEE Transactions on Sustainable Energy

The investigation on the use of Machine Learning algorithms for power system security assessment, along with a detailed set of results not included here for the sake of conciseness, is documented in the MSc thesis of Mr Tristan Catteeuw, titled "Machine learning applied to evaluation of power grid reliability", accepted by the University of Liege in the academic year 2021-2022.

Tamás Borbáth (KULeuven) additional contributions

Electricity shortfalls in Belgium are strongly correlated to neighbouring countries; in this work package, we did fundamental research to understand the interactions better and, more broadly, the role of a strong transmission backbone in lowering the magnitude and frequency of scarcity events.

Firstly we investigated different mathematical models to represent the transmission grid for power system dispatch simulations. Most importantly, system operators prepare forward-looking studies to estimate the likelihood of such events. When performing adequacy assessments, they simulate many potential future operational scenarios, sampling the uncertainties of weather and asset outages in a Monte Carlo simulation. Many scenarios must be simulated to gain statistically significant results, enabling their outcomes to be compiled into adequacy indicators. These indicators are then compared to societal expectations expressed via reliability standards to estimate future investment needs. There is a strong incentive to make these simulations computationally efficient (fast) to enable quick results. However, the flows on the AC transmission grid are best described by hard-to-compute non-linear equations. Current practices use mathematical relaxations and approximations to model these flows and ultimately understand the limits for transmission (for example, imports and exports to Belgium). We investigated how these more simple grid models ultimately affect the results of these studies, the adequacy indicators. The results demonstrate that the representation of cross-border transmission constraints can significantly impact the perceived level of resource adequacy. Using a simple model fails to portray the true capabilities of the grid and can potentially mislead practitioners on the actual investment needs of the power system. The results of this work have been published in the paper "The Impact of Cross-Border Transmission Constraints on Resource Adequacy Assessment." presented at the 2021 PowerTech conference. A more detailed publication is under preparation to be published later this year in a scientific journal.

As previously discussed, the flows on the AC network can be described by the injections and offtakes on the grid. For DC components, this is not the case. HVDC interconnectors operate based on a set schedule. Belgium is currently connected to the United Kingdom and Germany via such connections.

Still, there are plans to expand further the use of such devices (for example, to Denmark via Princess Elisabeth island). While these devices have a nominal power transfer capability, in practice, they can be slightly overloaded. The extra capacity offered on top of the nominal value is typically only available for a few hours daily. We investigated the value of using these dynamic capacities in the paper “Dynamic Rating of HVDC Interconnectors” by Tamás Borbáth and Dirk Van Hertem presented at the 19th International Conference on AC and DC Power Transmission in March 2023 in Glasgow.

We expect shortages would often simultaneously affect more than a single European country. Prices in the affected countries would reach the technical price limit eliminating any price-based competition between countries. In this case, alternative rules dictate how the available imports from third countries are allocated between the affected regions. In the “Model-based Approaches to Demand Curtailment Allocation” paper by Tamás Borbáth and Dirk Van Hertem, presented at the International Conference on the European Energy Market (EEM) in September 2022 in Ljubljana, we investigated the design of such rules. These allocation rules significantly affect the zonal adequacy indicators and, ultimately, the investment needs of each country. Using the flow-based approach to cross-zonal capacity calculation in Europe, some of the complexities of the grid are exposed to the allocation algorithm, and it could use demand curtailment for congestion management. These actions would indeed lower the overall need for curtailment in Europe but can have detrimental effects on mostly smaller countries such as Belgium.