A Mixed Method Approach for the Analysis of Variable Renewable Energy Integration in Developing and Fragile States

Luca Pizzimbone, Member, IEEE

Abstract—The paper presents a mixed method approach for the analysis of power systems in augmented uncertainty scenarios, related to the increasing penetration of variable renewable energy and country specific constraints to be found in fragile states.

In the formulated methodology, both deterministic and probabilistic load flow have their own specific, necessary and interactive role. To establish the soundness of the methodology, the analysis is conducted for a real case study, along with wind speed measurements (eleven-month duration), visual model validations, statistical and load flow analysis.

The probabilistic simulations are based on Monte Carlo (MC) analysis. Synthetic data are created from probabilistic distribution functions (PDF) calculated on original measured samples, operational constraints, and load uncertainties.

To facilitate the implementation of the proposed methods, scripts developed in Python programming language have been created for the analysis of statistical data, sample generation, post processing, data visualization and the interaction with conventional software for load flow analysis. The scripts are made public and available for download.

The proposed methodology of analysis, conceptualized for developing and fragile states, may also be used as a basis for all power system planning where the number of uncertainties is no longer negligible, and the use of deterministic methods alone would provide inadequate results.

Index Terms—developing countries, fragile states, load flow, Monte Carlo, probabilistic analysis.

I. INTRODUCTION

THE development of fragile states [1] using investment in renewable energy is a central focus of governments and financial institutions, such as development banks, as a key tool for promoting growth while meeting sustainable development goals at the same time.

Implementation of renewables can attract and retain investment and also has the potential to contribute to the mitigation of immediate risks, promote local business through the creation of jobs and stability for displaced populations [2].

However, electrical grids may not be ready to receive the targeted volume of renewables, unless additional investment were to be allocated to the reinforcement of either transmission or distribution facilities. This investment is often based on least-cost plans, which need further analysis and endorsement by the system operators (SO) in terms of security standards.

Conventionally, the above assessment is conducted with the aid of deterministic load flow analysis. Load flow calculations provide an effective means of analyzing the electrical system for the operating status of a given point in time. It represents the daily tool of SO, dealing with short-, medium- and long-term planning. Load flow is also used to find immediate solutions to contingencies which may occur in the electrical system.

However, in the energy transition process towards renewable energies, smart grids and e-mobility, the deterministic nature of the load flow may find its limits due to the ever-growing uncertainties associated with stochastic variations in load demand and power generation.

This is even more relevant in developing and fragile states, where the uncertainty problem is much greater, due to additional risks of outages, delays in project implementation, volatility of investment and unpredictable behavior of consumers during emergency states.

Other research papers clearly demonstrate the maturity of probabilistic load flow analysis and the interest of the SO in this numerical calculation method [3]. However, despite this interest, SO may encounter obstacles in the implementation of the new tools, due to a lack of methodologies for integrating the new methods of analysis in existing planning strategies.

This work attempts to provide a solution to this, proposing a mixed method approach of analysis, in which both deterministic and probabilistic calculations would have their own space, position and interactive role with the other planning elements.

Based on our efforts to solve theoretical and practical problems of a real case study, this paper combines the diverse study conceptions into a coherent overall methodology of analysis which can be practically implemented in the daily tasks of SO.

The academic relevance of this work can be summarized with the following contributions:

- Formulation of a mixed method approach of analysis in which both deterministic and probabilistic calculation features are used.
- Provision of a use case for the application of the proposed methodology of analysis.
- Making of a set of Python scripts for the public domain developed for the execution of MC load flow and data analysis [4].
- Consolidation of the proposed methodology in a real case study.

L. Pizzimbone is with the Department of Energy Economics and Planning, Tractebel Engineering GmbH, Bad Vilbel, Hessen, 61118 Germany e-mail: (see <https://orcid.org/0000-0002-1793-6791>).

- Verification and endorsement of common choices of PDFs for wind and PV on the basis of Bayesian Information Criteria (BIC) and visual tests.

II. METHODOLOGY OF ANALYSIS

In general, the analysis of electrical networks can be seen as a process that moves from coarse to fine levels. Each level is characterized by different objectives, input and output data, resources and constraints. Steady-states and dynamic analysis, electromagnetic transients (EMT) simulations, power quality assessments, protection coordination, etc. are progressively introduced in the analysis process depending on data availability, objectives and specific problem space which may arise.

In this process, load flow analysis tools have a fundamental role in the grid security assessment for short-, medium- and long-term planning.

The most general formulation of the load flow problem is given by a set of differential-algebraic equations, which can be reduced to the following equations:

$$g(x) = 0 \quad (1)$$

$$[g^P(x)g^Q(x)]^T = 0 \quad (2)$$

where g is a set of algebraic equations defining the power balance at nodes and x the state vector.

In its very primitive form, starting from the known values of injected active power at all nodes, with the exception of the slack node and the known values of reactive power or voltage magnitude, respectively at load and generation nodes, the solution of a load flow problem provides the active and reactive power that flow in all branch elements, and the amplitude and angle of voltages at all grid nodes, which satisfy the problem equations.

The probabilistic MC load flow is based on the same theoretical foundation. However, it is comprised of a large sequence of deterministic load flows (in most cases between ten to fifty thousand), for which the initial conditions of a subset of modelled elements are modified in each simulation run, according to predefined PDF, representing the physical stochastic behavior of the elements themselves.

The proposed methodology of analysis is based on a combination of both deterministic and probabilistic MC load flow analysis.

The deterministic analysis has a twofold purpose: to reduce the number of simulation cases to be processed by the MC analysis and to successively decide on the relevance of specific outliers of the MC simulation. Therefore, the roles of the deterministic load flow are represented by the input and output filter functions of Fig. 1.

The input filter is comprised of load flow and contingency analysis. It aims to verify the security criteria (for example N-1) with a deterministic approach, based on tentative worst-case scenarios. The results of the input filter should be a set of a small number of optional scenarios, which need to be analyzed through the successive probabilistic MC load flow.

The MC load flow is represented in the central part of Fig. 1, between input and output filters, and it is comprised of three stages:

- 1) Identification of Best Fitting PDF
- 2) Execution of MC Load Flow
- 3) Data Analysis

The above stages are discussed in detail in the following sections of this paper.

As the MC load flow produces, in autonomy, a relatively large number of data results, in the order of ten to fifty thousand samples, the controls over load flow constraints, such as active and reactive power, tap-changer adjustments, shunt switching or inter-area power exchange, are in most cases limited to a set of instructions which will be common for all the simulated cases. For this reason, some of the outlier results need to be further analyzed to confirm or discard their validity. This is the role of the output filters, which is again comprised of deterministic load flow analysis.

For the central part of the methodology (MC load flow analysis), scripts interacting in GUI-less mode with the base load flow module of PowerFactory have been implemented in Python.

III. STOCHASTIC MODELS

The MC analysis adopts statistical sampling methods with random parameters to represent the random physical process of the original problem.

The random parameters can be represented by a probability distribution function (PDF), also called probability density function. The relationship of PDFs to field data, such as wind speed and solar radiation, probabilities of outages, load forecast errors and in general any uncertainty of a parameter pertinent to the elements of the model, allows the generation of synthetic samples that can be used for the MC load flow simulations. Synthetic samples shall approximate with sufficient accuracy the same PDF of the original data. Different methods can be used to evaluate the accuracy of the PDFs. The one we adopted in this paper is the Bayesian Information Criterion (BIC) [5].

In statistics, several types of PDFs have been thought to best approximate the original data samples. We found that in power system analysis, the PDFs in Table I are the most widely used in power system analysis.

TABLE I
MOST COMMON PDFs UTILIZED IN POWER SYSTEM ANALYSIS

<i>Parameter</i>	<i>PDF</i>
Wind Speed	Weibull ^a
Solar Radiation	Beta ^b
	Kernel function ^c
Load	Normal ^d
Outages	Uniform ^e

^a [6], [7], [8], [9], [10].

^b [11], [12].

^c [3].

^d [13], [14], [10], [3].

^e [15].

With the exception of the Kernel Smoothing (KS) function, which is used to estimate a non-parametric distribution, the other PDFs listed in Table I are parametric distributions.

Fig. 1. Overall process of the proposed load flow analysis. The central part is relevant to the probabilistic analysis, while for the input and output filters deterministic analyses are used.

Parametric distributions are functions that can be defined by a set of parameters.

When a sample of data $\{x_1, x_2, \dots, x_N\}$ is available, such as in the case of field measurements of wind speed or solar radiation, the best fitting of the parametric distributions can be estimated. For this estimation we used the Bayesian Information Criterion (BIC). This method deals with the modelling of a probability distribution of a random vector $X = (X^1, \dots, X^{n_x})$, trying to rank variable candidate distributions by using the sample of data.

The estimation of the parameters of the candidate distributions is derived from either the Maximum Likelihood method or from the method of moments, except for the bound parameters that are systematically modified to strictly include the extreme realizations of the underlying sample $\{x_1, \dots, x_n\}$ [16]. The higher the likelihood L_i , the better the model describes the sample. Therefore, the risk of ranking parametric distribution based only on the highest likelihood is that complex models with many parameters may often score the best results. However, it should also be considered that complex models may be less robust than simpler models with less parameters.

The BIC (criterion avoids the above problem).

With $\{M_1, \dots, M_K\}$ as a list of possible parametric models, the BIC method ranks the models according to the following quantity:

$$\text{BIC}_i = -2 \frac{\log(L_i)}{n} + \frac{p_i \log(n)}{n} \quad (3)$$

where p_i represents the number of parameters to be adjusted for the model M_i . The highest rank is given to the model scoring the smallest BIC.

As demonstrated in (3), the selection of models with higher amounts of parameters is penalized and it will be the preferred solution only where the likelihood is high enough to justify the complexity of the model.

A. Copulas

The problem of the identification of the PDF for the stochastic models can also be resolved when dependencies between sample data are taken into account. This is the case, for example, where two or more wind farms are located within close proximity to each other and which can be affected by similar variations in wind speed.

The dependency structure of the variables is mathematically described through copulas and this can be used for any dependent multivariate set of samples.

A copula allows the representation of the part of the joined cumulative density function which is not described by the marginal laws. It is a special cumulative density function defined by $[0, 1]^{n_x}$, the marginal distributions of which are uniform on $[0, 1]$. The choice of the dependence structure is disconnected from the choice of the marginal distributions. A copula, restricted to $[0, 1]^{n_x}$ is a n_U -dimensional cumulative density function with uniform marginals.

$$C(\underline{u}) \geq 0, \forall \underline{u} \in [0, 1]^{n_U} \quad (4)$$

$$C(\underline{u}) = u_i, \forall \underline{u} = (1, \dots, 1, u_i, 1, \dots, 1) \quad (5)$$

For all N-box $B = [a_1, b_1] \times \dots \times [a_{n_U}, b_{n_U}] \in [0, 1]^{n_U}$, we have $\nu_C(B) \geq 0$, where:

$$\nu_C(B) = \sum_{i=1, \dots, 2^{n_U}} \text{sign}(v_i) \times C(v_i) \quad (6)$$

The summation is made over the 2^{n_U} vertices $\underline{v}_i$ of B .

$$\text{sign}(\underline{v}_i) = +1 \text{ if } v_i^k = a_k \quad (7)$$

for an even number of k 's, $\text{sign}(\underline{v}_i) = -1$ otherwise.

After the estimation of the marginals of each 1-dimensional vector of the samples (for example with BIC), the estimation of the copula for multivariate samples can be completed in Python with the use of the OpenTURNS libraries [17], with the following steps:

- 1) Find the dependent components of the sample. For example by:
 - (i) Calculating Spearman correlation coefficients.
 - (ii) Filtering near zero values.
 - (ii) Associating a True value to the Boolean of the remaining correlation coefficients.
- 2) Transform the sample in a way that follows a uniform cumulative distribution in the interval $[0,1]$, maintaining the copula intact.
- 3) Select the best fitting copula with BIC [18].

In the following section, the numerical results of a real case study are presented. In the subsection IV-B, the best fitting for PDF and copula are discussed again and applied to the simulated case.

IV. CASE STUDY

In this section, the proposed methodology is applied to a real case study. The results of deterministic and probabilistic simulations are compared, in order to substantiate the roles of both methods (deterministic and MC load flow) within the proposed methodology of analysis.

A. Description of the Grid

The electrical grid considered in this case study is a transmission system with a maximum voltage level of 132 kV. The total installed power generation is close to 2 GW (see Table II) and it is mainly comprised of thermal power plants, with the exception of a PV system, which has a rated power level of 100 MW.

A preliminary technical/economic analysis conducted in the geographical region of the project provided the basis for the interest of the SO in two alternative renewable expansion plans: a base renewable plan, involving an installed wind power capacity of 215 MW, in which grid reinforcements are avoided, and an optional plan involving about 130 km of overhead line reinforcements to accommodate an additional 90 MW of wind power, for a total installed capacity of 305 MW.

The basic structure of the transmission system is depicted in a schematic diagram in Fig. 2 and the main data summarized in Table II.

Fig. 2. Schematic diagram of the grid. Nodes connecting VRE are identified with the ID codes of the respective generation systems (W-01 to W-05, and PV-01).

B. PDF and Copula Calculation

The MC simulation uses synthetic sample data generated on the basis of best fitting PDF (parametric or non-parametric). In this case study the stochastic models are implemented for renewable generation (wind and PV) and load. The analysis considers the uncertainties related to the variability of renewable energy, errors in load forecast and possible delays in project implementation, which may be more significant in developing and fragile states. The numerical results of this analysis are discussed in the following points.

- 1) *Wind Stochastic Model:* Wind speed measurements were conducted for a period of 11 months at six sites in the

TABLE II
GRID DATA SUMMARY

Parameter	Value
No. of 132 kV substations	47
No. of lines	133
No. of synchronous generators	65
Total load	1525 MW
Total losses	32.2 MW
Total installed dispatchable generation	1978 MW
Total Non-dispatchable generation	100 MW
Target for additional non-dispatchable generation	215 MW ^a
	305 MW ^b

^aBase plan without additional grid reinforcement.

^bOptional plan With additional grid reinforcement.

geographical region of the study. Sample data of the measurements, on an hourly basis, were used to estimate the PDF of the wind speed and the copulas in five of the six candidate sites (W-01, W-02, W-03, W-04 and W-05). BIC tests were used to fit the PDF and copulas.

Once the PDFs had been estimated, 10,000 synthetic data of wind speeds were generated for each site. The results are visualized in the diagram in Fig. 3, which is a combination of scatter plots and probability distribution curves. Each scatter plot is related to a pair of sample data arrays, corresponding to the relevant x and y axes, while on the diagonal, the calculated PDF for the wind speed of each corresponding site is represented. The relation matrix of the copula is indicated in (15).

The best fitting of the PDF for the wind speed measurement data is executed according to BIC with the PDF library of OPENTurns [17]. The following three PDF scored the lowest

Fig. 3. PDF for wind speed at the five wind farm sites. The influence of the copulas is evident in the scatter plots, which present a constrained, narrow shape. The narrower the shapes, the higher is the correlation between the samples.

BIC: Weibull Max, Weibull Min and Truncated Normal. While the Weibull is a well-known PDF for wind analysis, the Truncated Normal is not a common choice. Therefore, to confirm the validity of the results, a visual test was also conducted, showing the histogram of the original samples and PDF of the generated synthetic data for the Weibull and the Truncated Normal. The diagram is presented in Fig. 4

Fig. 4. Comparison between Weibull and Truncated Normal for the W-04 site. The visual results appear to confirm the validity of the Truncated Normal, which is therefore used for the W-04 site.

The results presented in Fig. 4 appear to confirm the validity of the Truncated Normal, which is therefore adopted in the model.

The selected functions are represented by the following equations, where $F_X(x)$ is the cumulative density function (CDF) and $f_X(x)$ the PDF.

WeibullMax function (parameters are α , β , and γ)

$$F_X(x) = \exp\left(-\left(-\frac{x-\gamma}{\beta}\right)^\alpha\right), \quad \forall x \in (-\infty; \gamma] \quad (8)$$

$$f_X(x) = \frac{\alpha}{\beta} \left(-\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(-\frac{x-\gamma}{\beta}\right)^\alpha\right), \quad x \in (-\infty; \gamma] \quad (9)$$

with $\beta > 0$ and $\alpha > 0$.

WeibullMin (parameters are α , β , and γ)

$$F_X(x) = 1 - \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right), \quad x \in [\gamma; +\infty) \quad (10)$$

$$f_X(x) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right), \quad x \in [\gamma; +\infty) \quad (11)$$

with $\beta > 0$ and $\alpha > 0$.

TruncatedNormal (parameters are μ , σ , a , b):

Fig. 5. Comparison between PDF estimation with KS-Silverman and Kolmogorov-Smirnov (Beta).

$$f_X(x) = \frac{1}{\sigma} \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad x \in [a; b] \quad (12)$$

where ϕ and Φ are, respectively, the probability density distribution function and the cumulative distribution function of a standard normal distribution.

The calculated parameters for the above PDFs are summarized in Table. III. For the generation of the synthetic data, the PDFs have been used in combination with the calculated copula, which in this case is a Gaussian copula, with the correlation matrix of (15).

2) *PV Stochastic Model*: For the PV stochastic model, we compared a non-parametric and parametric PDF and after a visual test of the provided approximations we decided to use the Beta function, which represents a robust and simple solution. The procedure used is indicated here.

The Beta PDF is calculated using the following equation (parameters are α, β, a , and b).

$$f_X(x) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}B(\alpha, \beta)}, \quad x \in [a, b] \quad (13)$$

with $\alpha, \beta > 0$ and $a < b$ and where B denotes Euler's beta function.

As mentioned above, a visual test between the Beta parametric function and a non-parametric function, such as the Kernel Silverman (KS), was conducted to gain more confidence in the selected PDF.

The comparison between the Beta and the KS PDF was calculated for the set of synthetic samples of PV power output and is presented in Fig. 5, together with the histogram of the original samples.

The Beta PDF appears to provide a better approximation of the left boundaries, but significantly overestimates the right ones. The KS provides a better approximation of the central

part of the histogram, but it has also some deficiencies at the boundaries.

We found that both functions could have their own reasons to be chosen. However, for this work we decided to use the Beta function, which is a common choice for PV applications [11], [12]. Based on this experience, we consider that the performance of a visual test on the original and synthetic samples represents good practice, providing an effective aid in the identification of disadvantages and contributing to the final decisions on a suitable PDF.

3) *Load Stochastic Model*: The stochastic model for load demand is very important. When incorporated in the majority of the load elements, it may significantly affect the simulation results.

In developing countries, the annual growth rate of load demand is often quite significant and in the event of severe project delays the actual load at the time of commissioning of VRE could be different to what was expected.

The accuracy of the load forecast is another factor to be accounted for in terms of uncertainties, especially when the tools used for annual forecasts are unsuitable for complex development scenarios.

Both uncertainty factors (project delays and load forecast accuracy) have been considered in the stochastic model of the load, by means of a standard deviation from a normal PDF. In particular, a total standard deviation of 9% has been assumed, as a composition of 4.25%, for project delays (deduced from the annual growth demand) and 4.75% for load forecast error.

The PDF adopted for the stochastic model of loads is the normal distribution, recommended for real-valued random variables (such as quantities with measurement errors), for which the distribution is not known. [15].

The general form of the normal probability density function $f(x)$ is represented by the following equation (parameters are the mean μ and the standard deviation σ):

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right) \quad (14)$$

with $\Sigma = \Lambda(\sigma)R\Lambda(\sigma)$, $\Lambda(\sigma) = \text{diag}(\sigma)$, $\sigma_i > 0$. If $\nu > 2$, R is the symmetric correlation matrix, definite and positive.

If $\nu \leq 2$: R is the generalised correlation matrix (i.e. $R = I_d \in \mathbb{R}^d \times \mathbb{R}^d$).

4) *Summary of Stochastic Model Parameters*: The summary of calculated parameters for the representation of the PDFs described in the previous points are shown in Table III.

For the five wind farm sites named W-01, W-02, W-03, W-04 and W-05, the correlation matrix of (15) was also computed.

$$R = \begin{bmatrix} 1 & 0.958468 & 0.955151 & 0.923061 & 0.929419 \\ 0.958468 & 1 & 0.925727 & 0.911453 & 0.974772 \\ 0.955151 & 0.925727 & 1 & 0.967778 & 0.888889 \\ 0.923061 & 0.911453 & 0.967778 & 1 & 0.877746 \\ 0.929419 & 0.974772 & 0.888889 & 0.877746 & 1 \end{bmatrix} \quad (15)$$

TABLE III
SUMMARY OF PDF PARAMETERS

VRE ID	PDF parameters
W-01	Weibull Min $\beta = 9.37251, \alpha = 2.5463, \gamma = -0.116207$
W-02	Weibull Max $\beta = 17.2285, \alpha = 5.56578, \gamma = 23.8581$
W-03	Weibull Min $\beta = 8.89613, \alpha = 2.52253, \gamma = 0.025128$
W-04	Truncated Normal $\mu = 7.62826, \sigma = 3.39656,$ $a=-0.000196298, b=20.9742$
W-05	Weibull Min $\beta = 8.78902, \alpha = 2.5815, \gamma = -0.00816791$
PV-01	Beta $\alpha = 0.146319, \beta = 0.418275, a=0.0, b=1.0$
Loads	Normal $\sigma = 0.1$

C. Analysis of MC Load Flow

In this section of the paper, the results of the MC load flow are analyzed. The results are firstly compared with the deterministic load flow, in order to evaluate the consistency of the results and to assess the quality of information that each of the two methods of analysis (deterministic and probabilistic) can provide.

The section continues with the analysis of correlations. In this paper we present the correlations for overhead line loading vs VRE generated power. However, in a similar way, correlations have been calculated for transformer loading and voltage at nodes vs VRE generated power as well.

The analysis is conducted for the base and the optional grid plan scenario.

In Fig. 6, the results of the deterministic and probabilistic load flow relevant to the loading of overhead lines is presented. The overhead lines in the diagrams are a subset of elements, having a maximum loading (calculated on the base of 0.95 quantile) higher than 75%.

The loading of the deterministic analysis is represented by the bar diagram (in grey), while the MC load flow results are presented in a set of colored curves, representing minimum and maximum values, mean and the 0.95 and 0.05 quantiles.

According to the results of Fig. 6, the deterministic load matches the probabilistic load flow for Line 050 only, while for the remaining elements the deterministic load flows underestimates the loading. The difference is quite significant, especially because in the deterministic load flow the results are below 100%, while in the probabilistic analysis they are above 100%. This clearly can make a difference when it comes to making planning decisions.

In particular, considering the uncertainties of load demands, the probabilistic load flow provides more relevant values. These values are, in at least three cases, above the maximum line loading capacity. The same is not apparent in the deterministic load flows, which rely on the forecast load. In this case, the deterministic load flow would provide positive feedback for the planning proposals, which would not apply in case

Fig. 6. Simulation results for line loading of the base plan. Comparison between deterministic and probabilistic results for lines with loading values higher than 75 percent

of load forecast errors or delays in project implementation.

The probabilistic analysis also provides additional information on the line loading, such as mean, percentile probabilities, minimum and maximum values. These are represented by curves in different colours in Fig. 6. This information may contribute towards gaining a greater awareness of the severity of overloads in the lines, to check for remedial action or, ultimately, for the need for grid reinforcements.

In the diagram in Fig. 7, the same types of results for the optional plan are shown.

In a similar way to the base plan, Fig. 7 shows that the deterministic load flow results for the line loading are generally underestimated. However, Fig. 7 is also helpful for identifying another limitation of the deterministic load flow. In the deterministic load flow the loading of Line 133 and Line 038 are very similar, but in the probabilistic analysis Line 038 has a maximum loading significantly higher than the one of Line 133.

This apparent inconsistency called for further verification, which indicated that the maximum loading for Line 038 occurs concurrently with in a low power production value at W-04 and maximum power production at the remaining sites (i.e., W-01, W-02, W-03 and W-05). This specific condition was only captured by the MC simulation, but not by the deterministic analysis which was conducted only for the case with all VRE operating at maximum power.

We also used the base plan scenario to calculate and show the Cumulative Distribution Function (CDF) and PDF (histograms) for line loading. The diagrams for the subset of high loaded lines is presented in Fig. 8.

These diagrams are used to calculate the cumulative probability of overload for a specific value of VRE penetration, which in turn provides guidance for the estimation of possible

Fig. 7. Simulation results for line loading of the optional plan. Comparison between deterministic and probabilistic results for lines with loading values higher than 75 percent.

VRE curtailments, for example in case the planned grid reinforcements of the grid were not to be implemented in time.

Based on the discussed results, we can say that deterministic load flow, despite the advantages of speedy execution time and the provision of an effective aid in the selection of critical scenarios, cannot provide information on the probability of occurrence of the calculated electrical parameters which can be essential for final planning decisions. This information can, however, be provided by MC load flow, which provides important statistical data and also takes the impact of uncertainties into consideration as well. The drawback of MC load flow is that the time required for the MC simulations is relatively long, making it not particularly suitable for the analysis of large numbers of cases. On this basis, a combination of deterministic and probabilistic load flow would appear to be the most logical and effective approach for such types of analysis.

Another important feature of probabilistic load flow is of relevance for the calculation of correlations, as analyzed in the following section.

D. Correlations of variables in MC load flow results

At the present time the planning of renewable energy in developing countries is a topic of considerable importance. Renewable projects are often financed by development banks, under energy efficiency programs. The viability of these programs is mainly based on energy efficiency indicators, such as greenhouse gas reduction. Whereas correlating VRE projects with greenhouse gas reduction is relatively straightforward, the same does not apply to transmission or distribution reinforcement projects, as they are used by both renewable and fossil fuel-based power generation.

The fair identification of the correlations between VRE and grid reinforcements could facilitate a means of overcoming

this impasse. This can be achieved by means of MC load flow analysis, as illustrated in this section for line loading. A similar approach can be applied to transformer loading, voltage at nodes and, in principle, for any calculated parameter.

Fig. 9 shows the scatter diagrams for the same five lines examined in the base plan. Correlations are calculated by means of least squares linear regression. On the bottom-right corner of the figure, a bar type subplot summarizes correlations r and slopes m .

From the calculated results, it can be observed that all line loadings have a relatively high correlation with VRE power generation, with the exception of Line 057. The low correlation of Line 057 is also confirmed by a p-value not equal to zero. The p-value is defined as the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct. This means that p-values is zero in case the null hypothesis is discarded, which is in a way confirming the existence of a correlation.

The identification of correlations, such as in the case of Line 057, could be helpful for planning activities, as it suggests different priorities in terms of the items of the investment. Items with low priority, for example, could be moved to other financing programs, better tailored to their project objectives.

E. Outliers

The final function of the proposed method of analysis illustrated in Fig. 1 is the output filter, which deals with outliers.

Outliers are values which differ from the other observation points to a significant extent. In statistics, a data point is considered an outlier if its value is below $Q1 - 1.5IQR$ or above $Q3 + 1.5IQR$, with $Q1$ and $Q3$ being respectively the first and third quartile, and IQR the inter quartile range, equal to $Q3 - Q1$.

The general problem with outliers is to determine whether they should be retained or discarded in the analysis results.

In the specific case of power system analysis, it is necessary to understand the cause of the outliers and its impact on the results. Outliers could be simply generated by low probability combinations of stochastic variables used in the analysis, or by real problems, in most cases related to an inadequate control of active and reactive power, automatic tap-changer, or inter-area power exchange.

It could be difficult to detect such problems prior the execution of the MC load flow, as they may not be evident in the deterministic load flow, which is typically conducted for a limited number of scenarios. In most cases, the problems would emerge only when the operating conditions significantly diverge from the reference operating scenario and the initial setup of the control conditions is no longer valid. This is the case of MC load flow, where a large number of simulations, with variable data, are executed without supervision.

The first step required to deal with outliers is to identify them. An effective approach is through data visualization involving the use of box plots. An example is presented in Fig. 10 for the nodes with voltage below 0.95 p.u. Outliers are marked with small circles.

Fig. 8. Cumulative and probabilistic distribution (histogram) for line loading. The results allow an estimation to be made of the probability of occurrence of each loading value of interest (i.e. near 100%).

The results presented in Fig. 10, for Node 066 and Node 036 are quite suspect, therefore, further investigations have been conducted with the aid of deterministic load flow simulations, replicating the same conditions leading to some of the data points of the outliers. The analysis pointed out that the problem was related to the inappropriate control of reactive power from an external grid. The injection of an uncontrolled large amount of reactive power was leading to excessive voltage drops in the lines connecting the two nodes to the grid. In this case the outliers are discarded and the control conditions adjusted.

The above case is an example of outlier filtering.

V. CONCLUSION

Probabilistic analysis in power system applications, based on MC or derived methods, have been the subject of research for more than forty years now [19]. However, it is now, with the energy transition towards renewable energies, that we discover that the time is ripe for the large-scale dissemination of this method among the SO.

The results presented in this paper demonstrate that the analysis of power systems in augmented uncertainty scenarios needs to rely on a combination of both deterministic and probabilistic analysis.

A mixed method approach has been formulated, presented in a functional diagram, which includes features related to renewable energies, load demand forecast errors and uncertainty factors related to the specific context of development and fragile states.

The effectiveness and the potential of the proposed method has been illustrated and consolidated through numerical calculations for a real case study.

Since developing and fragile states may encounter certain constraints in the acquisition of additional software, a set of Python scripts for data analysis have been prepared and will be made public and available for download.

ACKNOWLEDGMENT

The author would like to thank: A. Mirakyan, my colleague at Tractebel Engineering GmbH, for providing valuable suggestions for the preparation of UML diagrams; E. Salazar, at Kreditanstalt für Wiederaufbau (KfW), for the deep insights on the topics of fragile states and project finance; A. Multze, S. Wulf, and V. Scheithauer for providing me with motivation in the writing of this article and for looking over my transcriptions and language; T. Brandstätt, I. Hoffmann, S. Drenkard, and all management and colleagues at Tractebel Engineering GmbH for supporting this initiative.

REFERENCES

- [1] "Fragile States Index," <https://fragilestatesindex.org/>, accessed: 2021-04-27.
- [2] S. Enrique, "Volatile renewable energy in fragile and conflict states - a conflict adjusted approach," in *Energy Week 2021, Virtual Sector Retreat March 23-26 2021*. Frankfurt am Main, Germany: Kreditanstalt Für Wiederaufbau (KfW), 2021.
- [3] S. Song, C. Han, S. Jung, M. Yoon, and G. Jang, "Probabilistic power flow analysis of bulk power system for practical grid planning application," *IEEE Access*, vol. 7, pp. 45 494–45 503, 2019.
- [4] "MCLoadFlow, Open Source Code," <https://github.com/lucapizzibone/MCLoadFlow.git>, accessed: 2021-04-29.
- [5] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461 – 464, 1978. [Online]. Available: <https://doi.org/10.1214/aos/1176344136>
- [6] X. Li, J. Cao, and D. Du, "Probabilistic optimal power flow for power systems considering wind uncertainty and load correlation," *Neurocomputing*, vol. 148, pp. 240–247, 2015.

Fig. 9. Scatter plots of loading of lines vs total VRE power output. Correlations are calculated by means of least squares linear regression. Values for slopes, correlations and p-values are indicated. The p-values = 0 is used to confirm the existence of correlations.

Fig. 10. Box plot for nodes with minimum voltage. The upper border of each rectangle represents the 3Q and the lower border the 1Q. The horizontal line traversing the rectangle is the mean value. The vertical lines starting above and below each rectangle show the range of the additional 1.5IQR. Above and below these limits, the data points are considered outliers and marked with small circles.

- [7] X. Xu and Z. Yan, "Probabilistic load flow evaluation considering correlated input random variables," *International Transactions on Electrical Energy Systems*, vol. 26, no. 3, pp. 555–572, 2016.
- [8] J. Usaola, "Probabilistic load flow with correlated wind power injections," *Electric Power Systems Research*, vol. 80, no. 5, pp. 528–536, 2010.
- [9] —, "Probabilistic load flow with wind production uncertainty using cumulants and cornish-fisher expansion," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 9, pp. 474–481, 2009.
- [10] J. Zhu, Y. Zhang, and H. Chen, "Probabilistic load flow with wind farms using a frequency and duration method," *IEEE Access*, vol. 6, pp. 74 812–74 827, 2018.
- [11] M. K. N and E. A. Jasmin, "A generalised three phase power flow algorithm incorporating the uncertainty of photo voltaic(pv) source for unbalanced distribution network," in *2015 International Conference on Technological Advancements in Power and Energy (TAP Energy)*, 2015, pp. 29–34.

- [12] A. Kumar and B. Bag, "Incorporation of probabilistic solar irradiance and normally distributed load for the assesment of atc," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017, pp. 1–6.
- [13] A. L. Da Silva, V. Arienti, and R. Allan, "Probabilistic load flow considering dependence between input nodal powers," *IEEE Transactions on Power Apparatus and Systems*, no. 6, pp. 1524–1530, 1984.
- [14] Z. Hu and X. Wang, "A probabilistic load flow method considering branch outages," *IEEE Transactions on power systems*, vol. 21, no. 2, pp. 507–514, 2006.
- [15] DigSilent Team, *DigSilent PowerFactory User Manual*, DigSilent GmbH, Gomaringen, Germany, 2020.
- [16] "OpenTURNS Documentation," <https://openturns.github.io/openturns/latest/index.html>, accessed: 2021-05-11.
- [17] M. Baudin, A. Dutfoy, B. Iooss, and A.-L. Popelin, "Open turns: An industrial software for uncertainty quantification in simulation," 2015.
- [18] D. Huard, G. Ivin, and A.-C. Favre, "Bayesian copula selection," *Comput. Stat. Data Anal.*, vol. 51, no. 2, p. 809–822, Nov. 2006. [Online]. Available: <https://doi.org/10.1016/j.csda.2005.08.010>
- [19] B. Borkowska, "Probabilistic load flow," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 752–759, 1974.

Luca Pizzimbone (Member, IEEE) was awarded the Diploma degree in Industrial Engineering with Electrical Specialization from the C.I.T.I. G. Galilei, Genoa, Italy, in 1991, and the Dring. degree in Electrical Engineering from the University of Genoa, Genoa, Italy, in 2002. He is currently the technical lead for Advanced Power System Analysis at Tractebel Engineering GmbH, Bad Vilbel, Germany. He is currently conducting international studies for the planning and rehabilitation of power grids, integration of renewable energies and the analysis of new technologies for smart grids. His current research interests include advance power system analysis aided by computer data science, supporting the sustainable technological and economic growth of developing states.