Probabilistic Framework for Online Identification of Dynamic Behavior of Power Systems with Renewable Generation

Panagiotis N. Papadopoulos, *Member, IEEE*, Tingyan Guo, *Member, IEEE*, Jovica V. Milanović,
Fellow, IEEE

Abstract—**The paper introduces a probabilistic framework for online identification of post fault dynamic behavior of power systems with renewable generation. The framework is based on decision trees and hierarchical clustering and incorporates uncertainties associated with network operating conditions, topology changes, faults and renewable generation. In addition to identifying unstable generator groups, the developed clustering methodology also facilitates identification of the sequence in which the groups lose synchronism. The framework is illustrated on a modified version of the IEEE 68 bus test network incorporating significant portion of renewable generation.**

Index Terms—**clustering, data analytics, decision trees, phasor measurement units, probabilistic transient stability, renewable generation.**

I. INTRODUCTION

THE evolution of power systems driven by economic, social and environmental pressures and technology advancements is calling for robust techniques and algorithms for both, close to real time identification of system dynamic behavior and control of the system. With the advent of monitoring technologies and market driven investment in system development, there is an increasing trend of utilizing the assets more effectively. The load growth and more restrictive, market driven, investment in the network and generation at one hand and integration of stochastic and intermittent, hence less controllable, generation on the other are commonly identified as major reasons that might lead future power systems to operate closer to their stability margin. Under such conditions corrective control may become crucial to ensure secure and reliable operation of power systems. Phasor Measurement Units (PMUs) that are being installed in large numbers in power systems around the world can facilitate close-to-real-time identification of the dynamic behavior of power systems by using advanced data analytics techniques and subsequent development and deployment of suitable corrective control algorithms.

From the plethora of available data mining techniques, Decision Trees (DTs) [1]-[6], Ensemble Decision Trees (EDTs) [7], Support Vector Machine (SVM) [8], [9] and Artificial Neural Networks (ANNs) [10] have been most frequently used in online dynamic security assessment. In most of the cases, the prediction focuses on whether the system will remain stable or not (binary classification) [1]-[5], [7]-[9]. However, in [6], [10], [11] the grouping of unstable generators after the fault is cleared is also determined. This information is significant for assisting corrective control actions such as controlled islanding, fast valving, dynamic braking, generator tripping, load shedding, use of Flexible AC Transmission systems (FACTS), etc. [6], [12].

Most of the above methods have been almost exclusively demonstrated in systems with conventional synchronous generation whose operation and dynamic behavior is well understood. The increasing presence of Renewable Energy Resources (RES) in the network can affect the dynamic behavior of the power system due to various mechanisms, as illustrated in past research. The uncertainties associated with new market driven operation of the system, and new types of load aggravated by uncertainties associated with RES technologies are contributing significantly to the steady state and dynamic behavior of the system. The inclusion of uncertainties associated with the presence and operation of RESs is essential for understanding dynamic behavior of the system, as their pre-fault operating conditions and participation in the overall generation mix can change significantly depending on weather conditions, i.e., their contribution to overall system dynamics may vary temporally and spatially. This might lead to system operating scenarios that have not been previously identified and might be critical for system stability [13].

In addition to uncertainties associated with RES operation the impact of topology changes, which may be more frequent in systems with RES, may lead to difficulties in assessing system post fault behavior. The impact of topology changes on the prediction of system stability using DTs was investigated in [14], [15]. Using a number of regression trees [14] or retraining [15] is proposed to deal with topology changes. In both these cases the aim was to identify the system stability limits only. In a few other papers, the impact of topology change on DT was assessed [2] and it was suggested that DTs

should be updated when new operating condition arises [7]. The SVMs were shown to be more robust to network topology changes than DTs [8], [9] however they were only used for binary classification (stable/unstable).

This paper builds on previous work of the authors [5], [6] and develops a probabilistic framework for online identification of dynamic behavior of uncertain power systems with significant penetration of non-synchronous generation subject to frequent topology changes. The framework uses multiple pre-trained DTs to identify not only "binary status" (stable/unstable) of the uncertain power system but also the groupings of unstable generators (based on their predicted dynamic response) and the sequence in which the generator groups loose synchronism. Considering that this prediction is made within a fraction of a second, the information provided can be effectively used for corrective control to insure system stability or to prevent cascading failures.

In summary the main contributions of this paper with respect to the existing literature include: i) the uncertainties (and dynamics) of RES (both wind and Photo-Voltaic generation) are considered for the first time in the multiclass online identification of generator grouping patterns, ii) The poor performance of DTs when dealing with network topology changes is addressed by training multiple DTs, i.e. Decision Forest (DF), for different network topologies; iii) The order in which generator groups lose synchronism is identified as an addition to the method described in [6]. This provides valuable information when designing and applying corrective control measures; iv) Finally, the proposed framework deals with the problem of online identification of generator grouping patterns as a one-step multiclass classification problem rather than a two-step process as described in [6]. This improves the performance and reduces the complexity of the previously proposed method.

II. PROPOSED FRAMEWORK

A schematic of the proposed framework is shown in Fig. 1. The purpose of the framework is the online identification of the transient stability status of the system and the power system dynamic signature, after a disturbance. The power system dynamic signature here refers to post-fault responses of the synchronous generators in the system characterized by their rotor angle behaviors and corresponding generator groupings defined by the stability status of generators (stable/unstable) and the sequence in which instability occurs for each group. The identification of the sequence in which unstable groups lose synchronism and the ability to incorporate frequent topology changes in assessment are key contributions compared to previous work on power system dynamic signature identification [6].

The core of the framework is a Decision Forest (DF), consisting of several DTs that perform the online identification of the power system dynamic signature. It has been observed that changes in the network topology can have a significant impact on the performance of DTs [14], [15]. To address this issue, the DF concept is proposed in this paper, consisting of several DTs that are trained and used for different specific network topologies. Each DT is trained for a 24 hour period considering uncertainties resulting from the system load and the level of penetration of wind and Photo-Voltaic (PV) generation for a specific network topology. Since the topology of the transmission network at given time is known the corresponding DT from the DF is then used to predict the grouping of generators and their dynamic behavior following the disturbance.
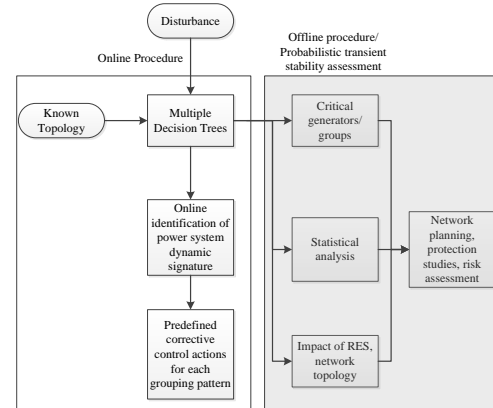

Fig. 1. Decision Forest framework.

Investigating the performance of the online identification method including the impact of RES poses the following challenges: i) increased uncertainty in the pre-fault operating conditions due to the intermittent behavior of RES and their availability, both temporal and spatial, that can have a significant impact on transient stability, ii) the displacement of synchronous generation, either by de-loading or disconnection, due to RES and therefore change of the overall system inertia and consequently system dynamic response; iii) the different dynamic behavior of RES changes the system dynamic behavior (change in generator grouping patterns as presented in Section V A below).

The proposed framework also includes an offline procedure for probabilistic transient stability assessment based on which several indices are calculated and statistical analysis is carried out to determine critical generator behavior. The results of the offline analysis of the simulated data used to train the DTs can be further used for network planning and protection studies. The offline transient stability analysis however, is out of the scope of this paper.

III. METHODOLOGY

There are four main stages in the development of the DF highlighted in Fig. 2: i) the off-line generation of the training datasets; ii) the off-line development of clustering method to determine the generator grouping patterns; iii) the off-line training of DTs and iv) the online classification using the appropriate pre-trained DT. The rotor angles of generators, obtained from detailed dynamic simulations, are used to initially identify the patterns of unstable generator groupings using an unsupervised learning approach (i.e. hierarchical clustering) and a corresponding pattern number is assigned to each simulated contingency. The simulated contingencies are

obtained following a probabilistic Monte Carlo approach, to generate a realistic training dataset (with pre-defined accuracy of coverage of possible contingencies) for the DTs. Following this, the DTs are trained as multiclass classifiers for the purpose of identifying in real time the generator grouping patterns. The values (measured by PMUs or, as in this study, obtained by computer simulations) of generator rotor angles are used as predictors. In practical applications the rotor angles can be obtained from PMU measurements [16]-[18] or alternatively through dynamic state estimators which offer increased accuracy [19].

A. Generation of Training Database

The procedure for generating the required training dataset is shown in Fig. 2. A full power system dynamic model suitable for stability studies is required, taking into consideration the connected RESs with the respective controllers. After the dynamic model is developed, the uncertainties concerning system loading and wind/PV generation for a 24 hour period are considered. Moreover, the uncertainties of the fault location and duration are also accounted for. All random variables are sampled according to appropriate probability distributions describing the behavior of corresponding uncertain parameter. Further details can be found in [6], [20]. The reason for performing probabilistic studies ensures that the representative (most probable) contingencies and operating conditions, from the set of all possible contingencies and operating conditions (determined by setting realistic ranges of uncertainties considered) will be considered and used for the training of the DTs (for the given system within the time frame of the study) which will consequently enable them to perform well in a realistic environment.

Following random sampling of different parameter values including RES availability at given time, an Optimal Power Flow (OPF) problem is solved to determine the output of conventional generators. The dispatch obtained from OPF determines also the amount of disconnection of conventional generation and consequently system inertia reduction due to increased RES penetration.

A number of Monte Carlo dynamic simulations, N_s, are executed by sampling all the above mentioned probability distributions. The sampling of the respective distributions is performed separately for each load and each RES unit in the system to consider independent behavior of loads and RES units within the system. The rotor angles of each generator are stored in a vector δ_{ig} for a duration of t_s seconds, where $i=1…N_s$ is the case number and $g=1…N_g$ is the generator number. The vector δ_{ig} consists of $n=1…n_s$ samples for the duration of t_s with a sampling time of 1 cycle (0.0167s for a 60 Hz system). A database of a large number of cases is generated for a 24 hour period and used to train a single DT for the specific network topology.
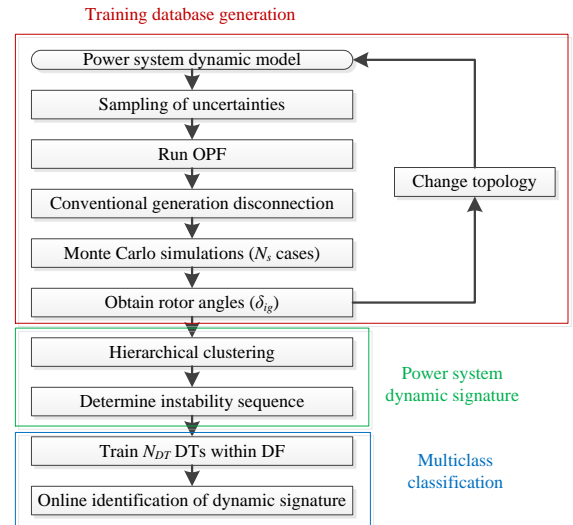

Fig. 2. Flowchart illustrating DF training and online identification procedure.

The whole procedure is repeated for all possible (viable) network topologies generating a number of N_{DT} datasets for different network topologies. In case, the number of possible transmission network configurations (i.e., line connections/disconnections) is too large and becomes impractical, the representative set of most critical topologies can be used by observing changes in generator grouping patterns caused by the topology change.

The proposed framework offers significant flexibility in creating the training database. The sampling of uncertain factors can be done according to any probability distribution based on historical data, prior knowledge or forecast [21]. Using probability distributions based on actual historical data (when available) is expected to provide more realistic results since some aspects of the uncertain behavior of power systems would be represented more accurately. Moreover, solving the OPF problem can include any number of additional security constraints associated with RESs as proposed in [22].

By using a selected set of topologies and operating conditions the number of required cases to be considered can be reduced. While the proposed framework can be used in this manner, sampling using probability distributions is proposed in this study in order to cover possible cases that may arise due to the inclusion of new uncertainties in the system (i.e. wind/PV generation). These new uncertainties could cause significant changes in power flows, as well as the dynamic behavior of the system. Certain contingencies and/or operating conditions can therefore become more or less important, possibly rendering a predefined database inefficient.

B. Power System Dynamic Signature

Since transient stability is the main focus of this paper, it is important that the grouping of generators (based on similarity of rotor responses) reflects *which generators* are unstable, *what group* they belong to and *in which order* the groups of, or individual generators, lose synchronism. There are several clustering algorithms available in the literature that have been applied to identify coherent groups of generators, such as fuzzy C-means, principal and independent component

analysis, support vector clustering and hierarchical clustering [23]-[26]. Hierarchical clustering is applied in this study, to determine the unstable generator groups, for each simulated contingency following the Monte Carlo probabilistic approach. This way the unstable generator groups that might appear in the specific system under study for a time frame of one day are obtained.

The agglomerative (bottom up) method is used to create a hierarchical cluster tree in this paper. The clustering begins by assuming that each object (generator in this case) forms a cluster of its own and then iteratively merges appropriate clusters until all the objects are in a single cluster. To measure the distance between two clusters the maximum distance linkage measure is used (complete linkage criterion) [27], i.e., the distance between two clusters is determined as the distance between the elements of each cluster (generator rotor angles) that are the farthest away from each other. The similarity between each pair of objects according to the linkage criterion, is measured by a similarity measure (Euclidean distance is used in this paper [6]). Finally, the obtained hierarchical cluster tree is cut using a predefined threshold for the linkage criterion to form the final clusters [27]. The important advantage of hierarchical clustering is that it facilitates setting a threshold with a physical meaning (when cutting the hierarchical tree), based on which the number of clusters can be automatically determined without having to provide it beforehand, which is an important feature of the developed methodology. Moreover, the obtained clusters are directly related to transient stability rather than to slow coherency as it has been done in the majority of past work. More information on the application of hierarchical clustering to determine unstable generator groups are also given in [6].

A generic case of generator rotor angle responses after the disturbance is cleared, is shown in Fig. 3a. The values of the vectors $\delta_{ig}(n_c)$ after a specified period of time t_c for each of the N_g generators, are used as the observations for the hierarchical clustering process. The period of time t_c is defined by the time $t_m < t_c$ after which the system operator will start applying the corrective measures. The value of t_m should be smaller than t_c to allow enough time to apply corrective measures following the identification of characteristic generator groups. Therefore, N_g values of generator rotor angles at the instance t_c, i.e. $\delta_{ig}(n_c)$, are used as input in the hierarchical clustering method.

As mentioned earlier, the agglomerative (bottom up) method is applied with a cut-off value of 360 degrees, since this is considered to be the transient stability limit. Euclidean distance between the data points is used as the similarity measure and complete linkage is chosen as the linkage criterion [6]. This results in generator groupings where generators belonging to one group have less than 360 degrees difference in rotor angles. In the case where only one group exists, the difference between the rotor angles of all generators is below 360 degrees and consequently the case is stable, as shown in Fig. 3a. When two or more groups appear in a pattern, this means that at least one generator of the additional groups has exhibited instability and therefore the case is

unstable. The hierarchical clustering method is applied in such a manner (use of complete linkage criterion, bottom-up strategy, 360 degrees threshold, etc.) that it is certain that the obtained generator groups always follow the rules specified above.

After the generator groupings are obtained for the time instance t_c, the sequence in which groups lose synchronism is determined. The Euclidean distances and complete linkage criterion are calculated sequentially for each set of samples $\delta_{ig}(n)$ for $n=1\ldots n_c$ (i.e. the above described methodology is applied for each time instance), for the cases that there is more than one group. Assuming there are m unstable groups in one case, the time instances $t_{u1}, t_{u2}\ldots t_{um}$ when the linkage criterion exceeds 360 degrees for generators belonging to groups 1, 2$\ldots m$, respectively, is recorded. Time instances $t_{u1}, t_{u2}\ldots t_{um}$ correspond to sample (and therefore cycle) numbers $n_{u1}, n_{u2}\ldots n_{um}$. In this way the order and the time at which groups lose synchronism is identified. The first group of generators (at least one generator of the group) to exceed the threshold of 360 degrees is considered to be the critical group. The rest of the groups are presented in the order the threshold is exceeded.

In Fig. 3b the rotor angles of 16 generators (G1-G16) clustered within three groups is shown. Generators G4 and G5 form the critical group and G6 and G7 belong to the second unstable group. The cycles where the linkage criterion between groups exceeds 360 degrees are also marked in the figure. For the critical group the $n_{u1}=36$ cycles is the sample when the linkage between G4 and G1 exceeds 360 degrees while for the second unstable group the linkage criterion is exceeded at $n_{u2}=74$ cycles for G6 and G1.

Following the proposed methodology of identifying grouping patterns, there is no need for separate binary detection of instability as in [6]. Multiclass classification can therefore, be directly applied in the following step, distinguishing simultaneously between stable and unstable cases using the assigned patterns from hierarchical clustering. The identification of unstable generator groups and the order in which those groups exhibit instability is the aim of applying the hierarchical clustering methodology. A more detailed reflection of the similarity of observed generator responses in the obtained clusters is not considered important as part of the proposed methodology, which aims at applying corrective control measures. For this reason, further evaluation of the hierarchical clustering procedure (e.g. calculation of cophenetic correlation) is not significant in terms of the aim of the methodology proposed in this paper.

C. DF Training Procedure

The identified generator grouping patterns for a large number of simulated contingencies are used as the targets to train DTs as multiclass classifiers. Generator rotor angles δ_{ig} for a duration t_{DT} are used as predictors. All values of the vector δ_{ig} between the time when the fault is cleared and t_{DT} (i.e. n_{DT} samples) are used for both training the DTs and for the online identification of grouping patterns. The value of t_{DT} is defined as a trade-off between the accuracy of the prediction and the time delay in corrective control actions. t_{DT} should be

less than or equal to t_m to ensure the prediction is available to the operator when needed in order to successfully apply the corrective measures. Ensemble Decision Tree (EDT) methods are used following the finding that they perform better for this class of problems as demonstrated in [6]. From the total of N_s simulated cases, a sub-set is used for training and the rest for testing purposes. A total of N_{DT} DTs are trained for different network topologies.
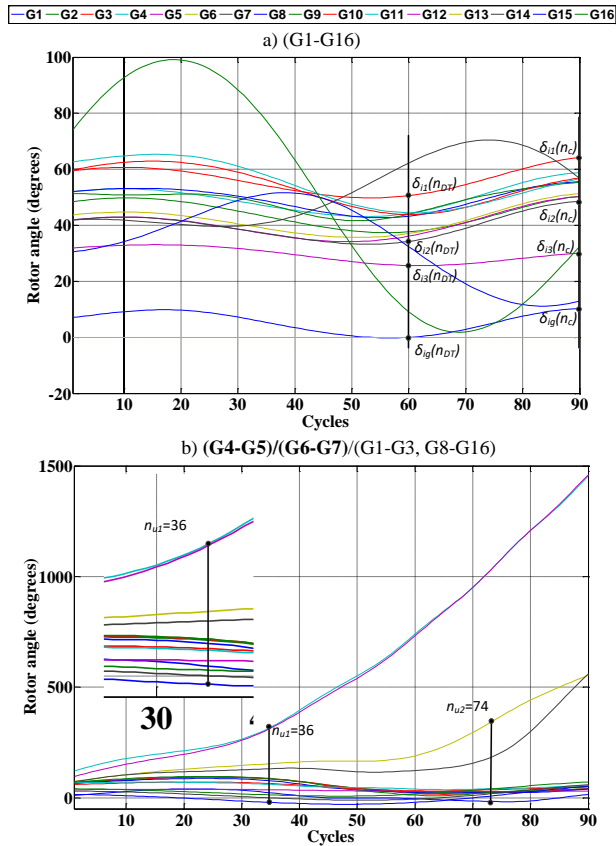

Fig. 3. Representative cases of stable and unstable grouping patterns.

When the training datasets for different network topologies are clustered, different grouping patterns tend to appear for different network topologies. A number of common patterns, usually the most critical ones, appear between all datasets. However, the frequency of the appearance of those patterns can change and more importantly new patterns can also appear. This means that a DT trained with a dataset corresponding to only one specific topology would be less accurate, since it would not be able to recognize at all the new patterns. Moreover, even for patterns appearing in both datasets, there can be slightly different rotor angle responses that lead to an increase in the prediction error. The DF concept is suitable to deal with this problem. Furthermore, different DT rules that may appear for different DTs can provide an insight considering parameters that affect the dynamic behavior of the system.

Therefore, it is proposed that specific DTs are trained and used for different network topologies. Within the 24 hour time frame considered, the correct DT according to the current network topology is chosen from a database (i.e. the DF) to

perform the classification. When a disturbance occurs, the corresponding DT is fed with the generator rotor angles δ_{ig} for duration t_{DT}. The output of the DT is the prediction of the generator grouping pattern about to happen.

IV. SYSTEM UNDER STUDY

The test network used, is a modified version of the IEEE 68 bus, 16 machine reduced order equivalent model of the New England Test System and the New York Power System (NETS – NYPS). The conventional part of the test network is adopted from [28], [29] and RESs are added at the buses shown in Fig. 4. Two types of RES units are connected to each bus: Doubly Fed Induction Generators (DFIGs), representing wind generators and Full Converter Connected (FCC) units, representing both wind generators and Photo-Voltaic (PV) units.
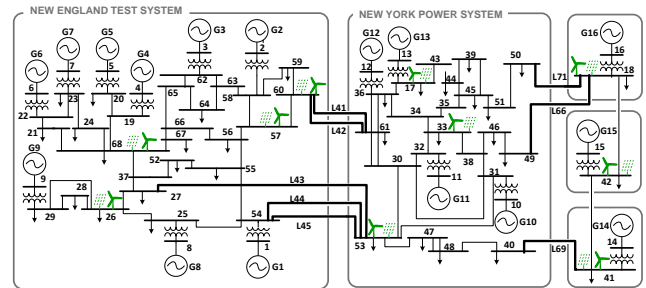

Fig. 4. Modified IEEE 68 bus test network.

A. Components modelling

The test network consists of 16 generators (G1-G16) in five interconnected areas. NETS consists of G1 to G9, NYPS of G10 to G13 and the three areas are represented by equivalent generators G14, G15 and G16, respectively. Standard 6th order models are used for all synchronous generators. G1-G16 are equipped with either slow IEEE DC1A dc exciters or fast acting static exciters type IEEE ST1A and G9 is equipped with a Power System Stabilizer (PSS). All generators are also equipped with generic governors, representing gas, steam and hydro turbines.

A Generic type 3 model, suitable for large scale stability studies is used in this paper to represent DFIGs. The model has a structure similar to the one proposed by WECC [30] and IEC [31], as shown in Fig. 5 and is available in DIgSILENT – PowerFactory [32]. It takes into consideration the aerodynamic part and the shaft of the wind turbine/generator as well as the pitch control of the blades. The rotor side converter controller is also modeled including relevant limitations, ramp rates and protection mechanisms, such as the crowbar. The DFIG is represented by a typical 2nd order model of an induction machine neglecting the stator transients and including the mechanical equation [33]. The rotor side converter is controlling the voltage in the rotor as in [34]. Therefore, the model represents all the relevant parts that influence the dynamic behavior of DFIGs.

Similarly, a type 4 wind generator model is used to represent all FCC units. Both wind generators and PV units can be represented by a type 4 model in stability studies, since the converter can be considered to decouple the dynamics of

the source on the dc part. This is also suggested by the WECC Renewable Energy Modeling Task Force [35], which develops a PV model by slightly modifying the type 4 wind generator model. The FCC model used in this paper and shown in Fig. 6 has a similar structure to [30], [31] and is available in the DIgSILENT – PowerFactory software [32].

Both DFIGs and FCC units are treated as aggregate units. Each RES unit model has a 2 MW power output and the number of connected units is varied to determine the output of the aggregate unit. Furthermore, all RES units are considered to have Fault Ride Through (FRT) capability and remain connected during the fault.
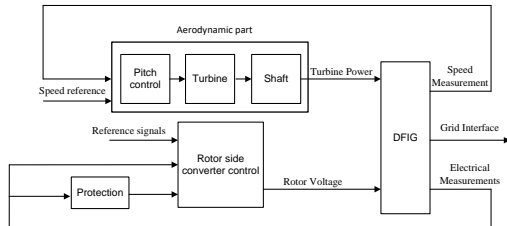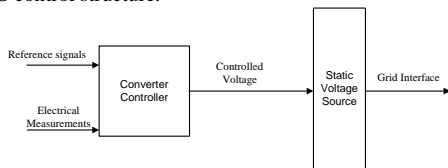

Fig. 5. DFIG control structure.

Fig. 6. FCC unit control structure.

The nominal apparent power of synchronous generators is derived considering 15% spare generation capacity available based on the nominal power flow data from [28]. A power factor of 0.85 is also assumed as nominal value. For the base case study, 20% of RES penetration based on conventional synchronous generation nominal values is considered for each of the five areas of the network. This means that the nominal capacity of connected RES generation is 20% of the nominal synchronous generation available (i.e. 16.67% of total generation). Out of this 20%, around 66.67% (i.e., approximately 11.11% of total generation) are considered to be DFIG wind generators and the remaining 33.33% (approx. 5.56% of the total generation) are FCCs. FCCs are further considered to be 30% wind generators (approx. 1.67% of total generation) and 70% PV units (approx. 3.89% of total generation).

B. Modelling of uncertainties

The daily loading and PV curves are initially used as shown in Fig. 7. The curves shown are typical pu curves obtained from National Grid data [36] and from the literature [37], respectively. First, the hour of the day is sampled randomly following a uniform distribution to determine the pu values for all the loads and all PV units according to the respective curves. For every hour within the day, the corresponding uncertainties are also modeled using a normal distribution for the system load [5] and a beta distribution for the PV generation [38]. Therefore, an extra uncertainty scaling factor for loads and PVs is introduced which is eventually multiplied

with the corresponding value from the daily loading or PV curve, respectively. The normal distribution for the system loading uncertainty has mean value 1 pu and standard deviation 3.33% and the beta distribution a and b parameters are 13.7 and 1.3 respectively [39]. For wind generation, the mean value of the wind speed within one day is considered constant [40], and the uncertainty of the wind speed is modelled using a Weibull distribution [41]. After considering the wind speed uncertainty, the power curve of a typical wind generator is used [42] to derive the power output. The Weibull distribution parameters used are φ =11.1 and k=2.2 [41].

In case forecasted daily curves are used, the above mentioned PDFs are considered to include both the forecast error as well as the inherent uncertain behavior of the respective parameters (loads, wind/PV generation). As mentioned in Section III A, the proposed framework offers flexibility in using more realistic PDFs without compromising the performance of the method since the DTs will eventually be trained with the resulting simulations.
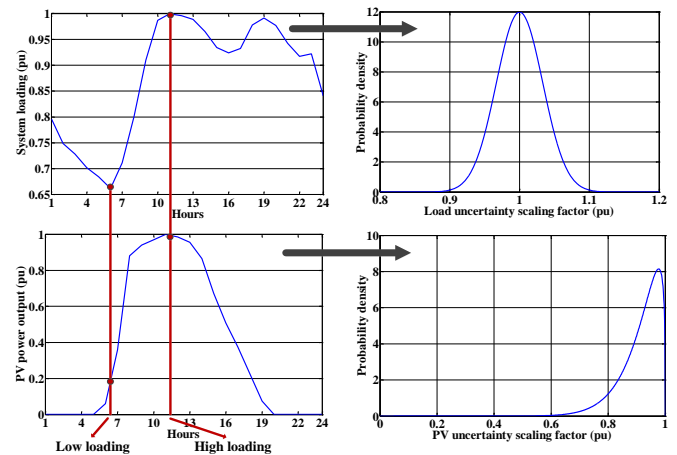

Fig. 7. Load and PV uncertainties modeling.

The above mentioned normal, beta and Weibull distributions are sampled separately for each load and RES unit in the system. Therefore, independent random variables are used for each specific load and RES to represent the variability of the uncertainties in a more realistic manner.

After considering the uncertainties, OPF is solved to determine the conventional generators dispatch P_{ig}, where g=1…16 for the specific test network and i=1…N_s. The cost functions for OPF are taken from [5]. The nominal capacity of each generator S_{ig} is then adjusted by adding 15% spare capacity according to $S_{ig} = P_{ig} / 0.85$. In case the resulting S_{ig} is larger than the nominal apparent power of the generators it is set to the nominal value. This means that there is no room for conventional generation disconnection in this case. The disconnection of conventional generation due to both load variations and RES penetration is considered in the following way. Since the generators are considered equivalent generators, reducing the nominal power, is equivalent to a reduction in the moment of inertia of the power plant and an increase in the generator reactance.

Only three phase faults are considered in this study due to

their strongest impact on system dynamic behavior. However, the simulation database could be extended to include other contingencies as well. A uniform distribution is used to model the fault location which means that the fault may happen with equal probability at any line of the test network and at any point along the line. A normal distribution with mean value of 13 cycles and standard deviation 6.67% is used to model the fault duration [6].

C. Monte Carlo Simulations

After considering all the uncertainties for a 24-hour time frame, 6000 cases (N_s) are simulated to construct the training and testing database for each one of the DTs within the DF. 70% (4200) of the cases is used as the training set and 30% (1800) as a testing set for each DT. The number of simulations N_s is chosen by keeping the error of the sample mean up to 5%, for 99% confidence interval, considering the TSI as the random variable. The error of the sample mean is calculated using (1), where Φ^{-1} is the inverse Gaussian CDF with a mean of zero and standard deviation one, σ^2 is the variance of the sampled random variable, δ is the confidence level (i.e. 0.01 for this study) and X_N is the sampled random variable with N samples [43]. In most cases increasing the number from 5000 to 6000 simulations, produced none, or only one new grouping pattern with no more than one corresponding case. The number of required Monte Carlo simulations might be affected by the shape of probability distributions used to represent the system uncertainties. Using the error threshold to define the number of required simulations ensures there is no loss of generalization of the proposed method.

$$e_{\overline{X}_N} = \frac{\Phi^{-1}(1-\frac{\delta}{2})\sqrt{\frac{\sigma^2(X_N)}{N}}}{\overline{X}_N} \qquad (1)$$

To illustrate the concept of DF, four different topologies are considered with a total of 24000 simulations performed. The DF therefore consists of four DTs, each one trained for the respective Test Case (TC). In the base case (TC1) no line is disconnected and the network is as described in Section IV. In TC2, all the RES units are disconnected and therefore the associated uncertainties are also not considered. TC3 and TC4 are the same as TC1 but lines 1 (between bus 21 and 68) and 2 (between bus 33 and 38) of NETS and NYPS are out of service, respectively. TC2 is used to compare the performance of the online identification method with and without RESs. The error of the sample mean ranges from 3.7% for TC1 up to 5% for TC3.

V. ONLINE IDENTIFICATION OF POWER SYSTEM DYNAMIC SIGNATURE

A. Hierarchical clustering results

The total number of unstable cases and different patterns observed in each TC are shown in Table I. Tripping of line 1 (TC3), which is a critical line within NETS, leads to the highest number of unstable cases. However, there are only few different patterns observed for this TC. For the case without considering RESs (TC2), there is the highest number of new

patterns observed, leading to the conclusion that the introduction of RESs changes the power system dynamic signature, and hence invalidates the DT training that might have been done with the network without RES. This difference is caused by the uncertainties introduced by RESs as well as the dynamic behavior of RES units. Finally, tripping line 2 does not cause significant changes. However, a few new patterns appear and the number of unstable cases slightly increases.

TABLE I
NUMBER OF UNSTABLE CASES AND PATTERNS OBSERVED FOR TCs

TCs	% of unstable cases	No of patterns	New patterns	Common patterns
TC1	9.45%	30	-	-
TC2	11.13%	32	13	19
TC3	13.48%	20	4	16
TC4	9.82%	29	5	24

In Table II the most significant identified unstable patterns are presented. The unstable groups are shown in bold letters and the order in which they are presented is the order they lose synchronism as described in Section III B. The percentages shown refer to the total number of unstable cases for each TC. New grouping patterns appear for different TCs and the frequency of appearance of common patterns is also changing. Pattern 1, in which G9 alone goes unstable, is the most common pattern in all cases. The frequency of appearance, however changes between TCs. The same applies for other patterns as well, such as pattern 2 and 3. Considering TC2, there are some patterns such as 1, 2, 8, etc. that appear less frequently, while others such as 3, 5, 12, 13 appear more frequently. This fact along with the fact that 13 new patterns appear in TC2 leads to the conclusion that RES units cause an overall change in the power system dynamic signature. In TC3 pattern 14 is introduced with a very significant number of cases. In Fig. 4, it can be seen that line 1 in NETS area is close to G6 and G7. Tripping line 1 therefore, weakens the network at that point and causes G6 and G7 to become more frequently unstable and resulting in a new distinct generator grouping pattern. In TC4 there are no significant changes in the most important groups leading to the conclusion that tripping line 2 in NYPS area is not so critical.

TABLE II
MOST SIGNIFICANT UNSTABLE PATTERNS

Pat.	Grouping	TC1	TC2	TC3	TC4
1	**(G9)**/(G1-G8,G10-G16)	49.38	42.96	35.19	49.38
2	**(G11)**/(G1-G10,G12-G16)	22.12	19.91	15.49	22.12
3	**(G2-G9)**/(G1,G10-G16)	2.65	7.63	0.12	2.65
4	**(G4-G5)**/**(G6-G7)**/(G1-G3,G10-G16)	1.95	1.35	3.84	1.95
5	**(G3)**/(G1-G2,G4-G16)	3.01	4.04	2.73	3.01
6	**(G4-G7,G9)**/**(G3)**/(G1-G2,G8,G10-G16)	0.88	-	0.12	0.88
7	**(G4-G7)**/(G1-G3,G8-G16)	4.78	5.69	8.67	4.78
8	**(G4-G5)**/(G1-G3,G6-G16)	4.60	1.20	1.73	4.60
9	**(G1-G9)**/(G10-G16)	1.59	2.40	-	1.59
10	**(G8)**/(G1-G7,G9-G16)	2.12	1.80	1.61	2.12
11	**(G5)**/(G1-G4,G6-G16)	0.88	0.15	0.25	1.27
12	**(G1-G10)**/(G11-G16)	0.35	2.69	-	0.71
13	**(G10)**/(G1-G9,G11-G16)	0.18	2.25	0.12	0.18
14	**(G1-G9)**/**(G10)**/(G11-G16)	-	2.25	-	-

B. Decision Tree Performance

The DF for the studied system consists of four DTs for the four TCs studied. t_{DT} is chosen as 60 cycles, which corresponds to 1 second duration for a 60 Hz system. In Fig. 8, the performance results of all the DTs within the DF are presented. Each DT trained with the simulations from a specific TC is tested against all TCs. The DTs trained and tested with simulations from the same TC are the ones chosen for the DF. In Fig. 8a the overall performance of the DTs in identifying both stable and unstable cases is shown, while in Fig. 8b the performance related only to unstable cases (i.e. identify correctly the specific unstable grouping pattern) is presented. The performance considering both stable and unstable cases is always higher. More specifically for the DTs within the DF, the accuracy is around 99% for all TCs. In general, stable cases are more easily identified by the DTs. Stable cases appear in larger numbers in both the training and testing dataset and only very few are misclassified as unstable. Therefore, the accuracy of the DTs in identifying between stable and unstable cases is very high. However, it is also critical to identify the correct generator grouping pattern. For this reason, the performance in identifying the correct grouping pattern considering only the unstable cases is also provided as a performance measure.
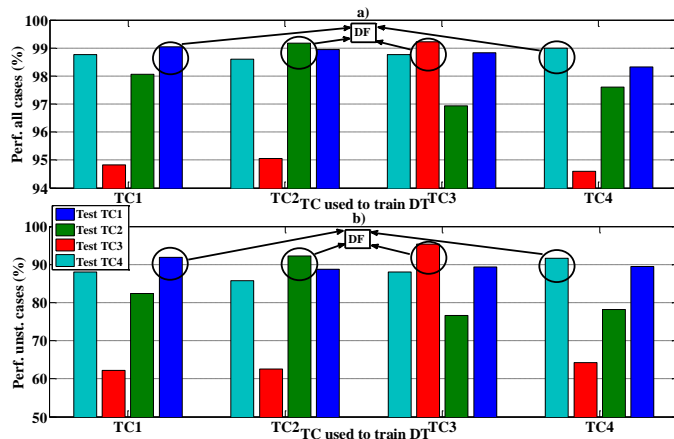

Fig. 8. Performance of DTs for a) all simulated cases (stable and unstable) and b) only for unstable cases.

The accuracy of DTs drops as low as 62% when they are trained with a different topology than the one used to test them. The drop in accuracy can be either attributed to new patterns with small occurrence frequency (as in TC4) or to a new pattern with significant number of cases (as in TC3). This fact supports the use of the DF to increase the overall performance and is essentially a comparison with methods based on DTs that do not address changes in network topology (e.g. [6]). The overall performance of the DF is calculated by summing all the misclassified cases of each DT within the DF and dividing by the total number of test cases for all the DTs within the DF. The DF can identify generator grouping patterns for the unstable cases with 93% accuracy, accounting for different network topologies and connection of RES. This

is a significant improvement, e.g. approximately 30% for critical topology changes and 10% with or without RES. It should be mentioned, that in some cases new unstable generator grouping patterns could appear in the test dataset. This means that the DTs have not been trained to identify those patterns and are therefore guaranteed to misclassify them. The errors presented in this paper include this type of errors to reflect the possible accuracy drop in a realistic environment. The number of these new generator grouping patterns can be minimized by increasing the training data set, which can be achieved during the off-line training of DTs.

Considering the accuracy of the proposed methodology with respect to the network size; the possible impact of the network size on the performance of the DF is related to the number of generator grouping patterns appearing and therefore the number of classes and features of the multiclass classification problem. However, the total number and the frequency of appearing patterns depends also on the specific system dynamics. For example, for the given test network, the number of patterns varies from 20 to 32 for different network topologies. Comparing between TC1 and TC3 there are 30 and 20 patterns respectively. The performance of the respective DTs is 91.88% and 95.38%, respectively (i.e. higher for the case with reduced number of patterns). However, comparing between TC1 and TC2 with 30 and 32 patterns the performance of the DTs is 91.88% and 92.23%, respectively (i.e. slightly higher for the case with increased number of patterns). In general it can be concluded that the performance of DTs might be affected by both the number and the frequency of appearance of the patterns.

To determine the robustness of the DTs, results for 30 and 10 cycles t_{DT} duration (0.5 and 0.167 seconds respectively in a 60 Hz system) are also provided in Fig. 9. In all cases, the performance of the DTs does not fall below 80%. If both stable and unstable cases are considered the performance is still high (above 97%) for each specific DT as well as for the whole DF. Considering the accuracy of identifying generator grouping patterns for unstable cases, there is a general performance drop as the prediction time is reduced. From 60 to 30 cycles, the drop in accuracy is not very significant. It ranges from less than 1% for TC3 up to almost 6% for TC2. A more significant drop is noticed when only 10 cycles are used for the prediction. However, TC3 still exhibits high performance (91.6% for unstable cases). This is attributed to the fact that patterns appearing frequently can be identified more easily, even with smaller duration of prediction time. In general, different TCs are affected in a different way by the reduction of t_{DT} due to different patterns that can be either identified more or less easily with shorter predictor time. The overall DF performance remains acceptable at 89.86% and 84.32% for 30 and 10 cycles respectively.
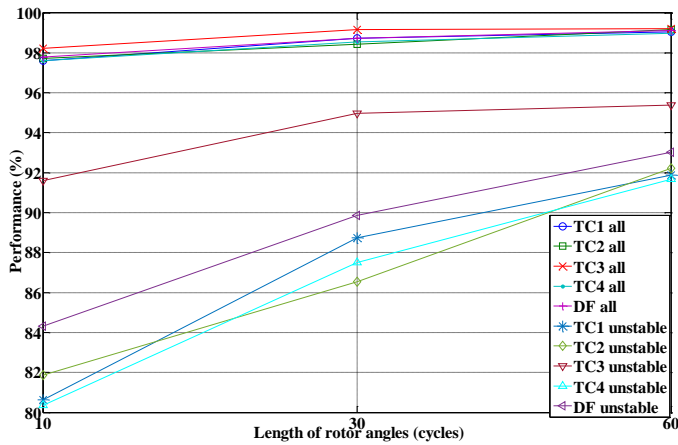
Fig. 9. Performance of DTs for different length of predictors.

C. Computational speed and practical considerations

Considering the online part of the method, the computational time required for the respective DT to perform the prediction is in the order of 10^{-4} seconds. Therefore, the time delay which is added to t_{DT} until the actual prediction is available can be considered negligible.

Considering the offline part, the most computation intensive task is the dynamic simulations used to train and test the DTs. The simulations are performed in DigSILENT/PowerFactory software as mentioned before. For the TC without RES, approximately 18 hours are required to perform 6000 Monte Carlo simulations. For the TCs that include RES, approximately 60 hours are required for 6000 dynamic simulations. After the dynamic simulations are obtained, approximately 13 seconds are required to obtain the generator grouping patterns for 6000 simulated responses, by applying the hierarchical clustering approach described in Section III A. Finally, the training of one DT requires approximately 10 to 20 seconds using the C5.0 boosting algorithm. An Intel Core i7 3.4 GHz with 16 GB of RAM is used for all the computations. In case there is need to decrease the computational time required when performing the offline simulations, efficient sampling techniques can be applied to reduce the required number of simulations [43]. Alternatively, since the offline Monte Carlo simulations can be performed in parallel, the computational time can be reduced by utilizing multi-core processors or even multiple computers.

Another possible practical issue that might arise from the proposed method is related to communication error and signal loss of PMU measurements. This has been investigated by the authors in [16] and [18]. In [16] the effect of PMU measurement errors is studied and the importance of measured signals from each generator is also investigated. It has been concluded that measurement signals from specific generators might be more important in order to identify the unstable generator grouping pattern. This means that even with fewer available signals (only from certain identified important generators), the prediction error can be maintained low. In [18], the effect of signal loss is investigated (due to possible disruption of the communication channel) and it is concluded that ensemble decision tree methods are more robust under measurement signal loss. This is one of the reasons for which C5.0 with boosting algorithm is chosen in this paper.

VI. Conclusions

A probabilistic framework for online identification of the power system dynamic signature in systems with RES is presented in this paper. There are three main improvements with respect to previous methodologies: i) it considers the impact and uncertainties related to RESs; ii) it is more robust and capable of handling topology changes in the network by training multiple DTs; iii) the prediction of the unstable generator grouping pattern includes the sequence in which the groups lose synchronism.

While the connection of RES as well as network topology changes alter the dynamic behavior of the system, DTs trained for specific topologies can still achieve high performance. The proposed DF can achieve very high accuracy (close to 99%) when distinguishing between stable and unstable cases. When it comes to specific unstable grouping pattern identification, the accuracy is lower but still high (93%). The best improvement in the accuracy of identification of system status resulting from the application of the DF when dealing with topology changes was from 62.18% to 95.38%. For very fast prediction, the overall DF performance is still very high with 89.86% and 84.32% accuracy for 0.5 and 0.167 seconds respectively, when it comes to identification of specific generator grouping pattern.

Finally, the proposed framework has been also applied to a modified version of the IEEE 118 bus test system, with 54 generators, 19 of which are modelled as synchronous generators (using full detailed 6th order models) and the rest as synchronous compensators [45] and the results obtained were fully constraint (speed, accuracy, etc.) with those presented in the paper. This confirms that the methodology can be easily scaled up to larger and/or different systems without any problem, and therefore is generally applicable.

References

[1] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1417–1426, 1994.

[2] V. Vittal, N. Senroy, and G. T. Heydt, "Decision Tree Assisted Controlled Islanding," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1790–1797, 2006.

[3] Q. Gao and S. M. Rovnyak, "Decision trees using synchronized phasor measurements for wide-area response-based control," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 855–861, 2011.

[4] K. Mei and S. M. Rovnyak, "Response-Based Decision Trees to Trigger One-Shot Stabilizing Control," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 531–537, 2004.

[5] T. Guo and J. V Milanovic, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 377–385, 2014.

[6] T. Guo and J. V Milanović, "Online Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining," *IEEE Trans. Power Syst.*, 2015, available online.

[7] M. He, J. Zhang, and V. Vittal, "Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4089–4098, 2013.

[8] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient

stability status prediction using synchronized measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1474–1483, 2011.

[9] A. D. Rajapakse, F. Gomez, K. Nanayakkara, P. A. Crossley, V. V. Terzija, "Rotor angle instability prediction using post-disturbance voltage trajectories," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 947–956, 2010.

[10] N. Amjady, S. F. Majedi, "Transient stability prediction by a hybrid intelligent system," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1275-1283, 2007.

[11] S. Kretsinger, S. Rovnyak, D. Brown, and J. Thorp, "Parallel decision trees for predicting groups of unstable generators from synchronized phasor measurements," in *Proc. Precise Measurements in Power Systems Conf.*,Washington, DC, USA, 1993.

[12] R. Moreno, D. Pudjianto, and G. Strbac, "Transmission network investment with probabilistic security and corrective control," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3935–3944, 2013.

[13] R. Preece and J. V Milanović, "Assessing the Applicability of Uncertainty Importance Measures for Power System Studies*," IEEE Trans. Power Syst.*, 2015, available online.

[14] J. a. Huang, G. Vanier, L. Loud, S. Guillon, J. C. Rizzi, and F. Guillemette, "Topology information based decision trees to predict dynamic transfer limits and their sensitivities for Hydro-Quebec's network," *IEEE PES Gen. Meet. PES 2010*, pp. 1–7, 2010.

[15] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1978–1987, 2013.

[16] P. N. Papadopoulos, T. Guo, X. Wang, J. V. Milanovic, "Impact of Measurement Signals on the Accuracy of On-line Identification of Power System Dynamic Signature," *PowerTech 2015*, Eindhoven, June 2015.

[17] *IEEE Standard for Synchrophasor Measurements for Power Systems*, *IEEE Std C37.118.1-2011*.

[18] T. Guo, P. Papadopoulos, P. Mohammed, J. V Milanovic, "Comparison of Ensemble Decision Tree Methods for On-line Identification of Power System Dynamic Signature Considering Availability of PMU Measurements," *PowerTech 2015*, Eindhoven, June-July 2015.

[19] M. A. M. Ariff, B. C. Pal, and A. K. Singh, "Estimating Dynamic Model Parameters for Adaptive Protection and Control in Power System," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 829–839, 2014.

[20] P. N. Papadopoulos, A. Adrees, J. V. Milanović, "Probabilistic Assessment of Transient Stability in Reduced Inertia Systems," *PESGM 2016*, Boston, 17-21 July 2016.

[21] T. Guo and J. V Milanovic, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 377–385, 2014.

[22] S. Eftekharnejad, G. T. Heydt, V. Vittal, "Optimal Generation Dispatch with High Penetration of Photovoltaic Generation," in *IEEE Power & Energy Society General Meeting*, 26-30 July 2015.

[23] I. Kamwa, A. K. Pradhan, and G. Joos, "Automatic segmentation of large power systems into Fuzzy coherent areas for dynamic vulnerability assessment," *IEEE Trans. Power Syst.*, vol. 22, pp. 1974–1985, 2007.

[24] K. K. Anaparthi, B. Chaudhuri, N. F. Thornhill, B. C. Pal, "Coherency identification in power systems through principal Component analysis," IEEE Trans. Power Syst., vol. 20, pp. 1658–1660, 2005.

[25] M. A. M. Ariff and B. C. Pal, "Coherency identification in interconnected power system – an Independent Component analysis approach," IEEE Trans. Power Syst., vol. 28, pp. 1747–1755, 2013.

[26] R. Agrawal, D. Thukaram, "Identification of coherent synchronous generators in a multi-machine power system using support vector clustering," in Proc. *Int.Conf. Power and Energy Syst.(ICPS)*, India, 22-24 Dec 2011.

[27] J. Han, M. Kamber, and J. Pei*, Data Mining : Concepts and Techniques*.San Mateo, CA, USA: Morgan Kaufmann, 2012.

[28] G. Rogers, *Power System Oscillations*. Kluwer Academic, 2000.

[29] B. Pal and B. Chaudhuri, *Robust Control in Power Systems*. New York, Springer Inc., 2005.

[30] WECC Wind Power Plant Dynamic Modeling Guide, WECC Renewable Energy Modeling Task Force, January 2014.

[31] *Wind turbines - Part 27-1: Electrical simulation models - Wind turbines*, IEC 61400-27-1, 2015.

[32] DIgSILENT-PowerFactory User Manual, DIgSILENT GmbH, 2014.

[33] P. Kundur, *Power System Stability and Control*. New York, USA: McGraw-Hill, 1994.

[34] J. Fortmann, S. Engelhardt, J. Kretschmann, C. Feltes, I. Erlich, "New Generic Model of DFG-Based Wind Turbines for RMS-Type Simulation," *IEEE Trans. Energy Conv.*, vol.29, no.1, pp.110-118, 2014.

[35] WECC PV Power Plant Dynamic Modeling Guide, WECC Renewable Energy Modeling Task Force, May 2014.

[36] Daily load curve data, National Grid, http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-Explorer/.

[37] M. Fan, V. Vittal, G. T. Heydt, R. Ayyanar, "Probabilistic power flow analysis with generation dispatch including photovoltaic resources," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1797–1805, 2013.

[38] M. Fan, V. Vittal, G. Heydt, R. Ayyanar, "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants," *IEEE Trans. Power Syst.*, vol.27, no.4, pp.2251-2261, 2012.

[39] Shi Tao, Yu Ruoying, Zhu Lingzhi, Gao Shan, "Power system probabilistic production simulation containing large-scale wind power and photovoltaic power," in *Power and Energy Engineering Conference (APPEEC), 2013 IEEE PES Asia-Pacific* , 8-11 Dec. 2013

[40] J. P. Coelingh, a. J. M. Van Wijk, and A. A. M. Holtslag, "Analysis of wind speed observations over the North Sea," *J. Wind Eng. Ind. Aerodyn.*, vol. 61, no. 1, pp. 51–69, 1996.

[41] R. Preece and J. V. Milanovic, "Tuning of a damping controller for multiterminal VSC-HVDC grids using the probabilistic collocation Method," *IEEE Trans. Power Deliv.*, vol. 29, no. 1, pp. 318–326, 2014.

[42] Vestas Americas Inc., "V80-2.0 MW: Unsurpassed reliability and performance at high-wind sites in North America," [Online]. Available: www.vestas.com.

[43] R. Preece, J. V. Milanovic, "Efficient Estimation of the Probability of Small-Disturbance Instability of Large Uncertain Power Systems," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1063–1072, 2015.

[44] P. Demetriou; M. Asprou; J. Quiros-Tortos; E. Kyriakides, "Dynamic IEEE Test Systems for Transient Analysis," in *IEEE Systems Journal* , available online (early access http://ieeexplore.ieee.org/document/7151773/).

Panagiotis N. Papadopoulos (S'05-M'14) received the Dipl. Eng. and Ph.D. degrees from the Department of Electrical and Computer Engineering at the Aristotle University of Thessaloniki, in 2007 and 2014, respectively. Since 2014 he has been postdoctoral Research Associate at the University of Manchester. His special interests are in the field of power system modeling, simulation and investigation of dynamic behavior of power systems with increased penetration of non-synchronous machines.

Tingyan Guo (S'11 M'16) received the B.Eng degree in Electrical and Electronic Engineering from the University of Manchester, UK in 2011, and the Ph.D. degree in Power System Dynamics at the same institution in 2016. She is currently a power system engineer with the network strategy department of National Grid, UK.

Jovica V. Milanović (M'95, SM'98, F'10) received the Dipl.Ing. and M.Sc. degrees from the University of Belgrade, Belgrade, Yugoslavia, the Ph.D. degree from the University of Newcastle, Newcastle, Australia, and the Higher Doctorate (D.Sc. degree) from The University of Manchester, U.K., all in electrical engineering. Currently, he is a Professor of Electrical Power Engineering, Deputy Head of School and Director of External Affairs in the School of Electrical and Electronic Engineering at the University of Manchester, U.K., Visiting Professor at the University of Novi Sad, Serbia, University of Belgrade, Serbia and Conjoint Professor at the University of Newcastle, Australia.