Prediction of Power System Post-Contingency Vulnerability Status by Mining Synchronized Phasor Measurements

Jaime C. Cepeda*, José L. Rueda**
István Erlich, Pawel Sowa*****

**Corporación Centro Nacional de Control de Energía -CENACE-, Quito, Ecuador*
(Tel: 593-984578259; e-mail: jcepeda@cenace.org.ec)
***Institute of Electrical Power Systems, University Duisburg-Essen, Duisburg, Germany*
(e-mail: jose.rueda@uni-due.de, istvan.erlich@uni-due.de)
****Institute of Power System & Control, Silesian University of Technology*
(e-mail: Pawel.Sowa@polsl.pl)

Abstract: Real-time vulnerability assessment (VA) is one of the essential tasks of the so called Smart Grid, since it has the function of detecting the necessity of performing global control actions. In view of this, the present paper will introduce a novel data-mining-based approach to map post-contingency Dynamic Vulnerability Regions (DVRs), taking into account three short-term instability phenomena. Based on probabilistic models of relevant inputs (e.g. nodal loads and occurrence of contingencies), the approach applies Monte Carlo (MC) simulation to recreate a wide variety of possible post-contingency dynamic data of some electric variables, which could be directly available from PMUs in a real system (e.g. voltage phasors or frequencies). From this information, a pattern decomposition method, based on empirical orthogonal functions (EOF), is used to approximately pinpoint the DVR spatial locations. The identified DVRs are then used to ascertain the actual dynamic state relative position with respect to their boundaries, which is accomplished by using a support vector classifier (SVC). The proposal is tested on the IEEE New England 39-bus test system. Results show the feasibility of the approach in finding hidden patterns in dynamic electric signals as well as in numerically mapping power system DVRs.

1. INTRODUCTION

Under high stressed system conditions, certain sudden perturbations might cause cascading events that may lead the system to blackouts, as stated by Amin (2000). It is crucial to ensure that these perturbations do not affect the system security, so that, the development of wide area protection systems, that allow guaranteeing the service continuity, is required. However, these protection systems are usually set to operate when specific pre-established operational conditions are reached, and they are unable to work under unconsidered contingencies that could begin cascading events. Under these considerations, the control of the system and the protection trigger should be adjusted depending on the real time event progress. This new context needs the development of a more intelligent system (*Smart Grid*) that could efficiently respond to the actual system conditions and provide autonomous control actions to enhance the system reliability (Moslehi and Kumar, 2010). This *Self-Healing Grid* has some specific requirements such as an adaptive control and protection system, adequate measurement equipment, sophisticated communication networks, and appropriate tools to analyze huge volumes of data in real time (such as appropriate *data mining* techniques). A fundamental task of this smart structure is the *vulnerability assessment* (VA), since it has the function of detecting the necessity of performing global control actions. Most VA methods are based on steady state (Static Security Assessment -SSA-) or dynamic (Dynamic Security Assessment -DSA-) simulations of N-x critical contingencies (Cepeda *et al.*, 2011). As exposed by Yuan Zeng *et al.* (2006) and Savulescu (2009), the aim of these methods is to determine whether the post-contingency states are within a “safe region”, and accordingly, to decide the most effective *preventive* control actions. In recent years, the achieved and expected breakthroughs in emerging technologies, such as Phasor Measurement Units (PMUs), and Wide Area Monitoring, Protection and Control Systems (WAMPAC), motivates new research endeavors to develop more sophisticated VA methods (Savulescu 2009; Cepeda *et al.*, 2011). Most of the current PMU-based approaches have been designed in order to perform preventive control actions, following the traditional practice. Nevertheless, as stated by Cepeda *et al.* (2011), the use of PMUs has a great potential to allow performing *post-contingency dynamic vulnerability assessment* (DVA) that could be used to trigger *corrective* control actions. In view of this, a novel data-mining-based approach to map post-contingency *Dynamic Vulnerability Regions* (DVRs), taking into account three short-term instability phenomena, was previously presented by Cepeda *et al.* (2012). Based on probabilistic models of relevant inputs (e.g. nodal loads and occurrence of contingencies), the approach applies Monte Carlo (MC) simulation to recreate a wide variety of possible post-contingency dynamic data of some electric variables, which could be directly available from PMUs in a real system (e.g. voltage phasors or frequencies). From this information, a pattern decomposition method, based on empirical orthogonal functions (EOF), is used to approximately pinpoint the DVR spatial locations. Along this research line, the work presented in this paper presents a comprehensive pattern recognition approach for predicting the power system's post-contingency dynamic vulnerability status (PCDVS) by considering the MC-based DVRs introduced by Cepeda *et al.* (2012) together with a support vector classifier (SVC) adequately adapted to this specific task. The ultimate goal is to perform early classification of the system status into "vulnerable" or "non-vulnerable" in real time, by mining the post-contingency dynamic data obtained directly from PMUs.

2. RELEVANT DATA MINING TOOLS

The main mathematical tools employed in this work are the so-called data mining techniques, which have proven to be useful for extracting or mining knowledge (i.e. pattern recognition) from large amounts of data, as exposed by Han and Kamber (2006). In the following, a brief review on the rationale behind the use of empirical orthogonal functions and support vector classifiers is presented.

2.1 Empirical orthogonal functions

Empirical orthogonal functions (EOFs) are the result of applying singular value decomposition (SVD) to time series data. EOF is a time series data mining technique that allows decomposing a discrete function of time f(t) (such as voltage angle, voltage magnitude or frequency) into a sum of a set of discrete pattern functions, namely EOFs. Thus, EOF transformation is used in order to extract the most predominant individual components of a compound signal waveform (similarly to Fourier analysis), which allows revealing the main patterns immersed in the signal.

The main approaches related to EOFs have been developed for using in the analysis of spatio-temporal atmospheric science data, whereas their application in other scientific fields continues to be scarce. The data concerned consist of measurements of specific variables, (such as sea level pressure, temperature, etc.) at n spatial locations at p different times (Jollife, 2012). The present paper employs a variation of this definition, where the n spatial locations are replaced by n different post-contingency power system states (obtained from MC simulation), and the p different times consist of PMU instant values of post-disturbance dynamic variables (voltage phasors or frequencies) measured at m buses, at r different instants that depend on the selected time window (i.e. p = m×r).

Therefore, a (n × p) data matrix of discrete functions (**F**) is structured, where the post-contingency measurements at different power system states (n) are treated as observations, and the PMU samples belonging to a pre-specified time window (p time points) play the role of variables. Since the different power system states result from the application of MC-based simulations, n is conceptually greater than p (n > p), and so **F** is a rectangular matrix.

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_1(t) \\ \vdots \\ \mathbf{f}_n(t) \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \tag{1}$$

where $\mathbf{f}_k$ is the *k*-th discrete function of time obtained in the *k*-th MC-repetition that consists of p samples.

Formally, the SVD of the real rectangular matrix **F** of dimensions (n × p) is a factorization of the form (Peña, 2002):

$$\mathbf{F}_{np} = \mathbf{U}_{nn}\mathbf{\Lambda}_{np}^{1/2}\mathbf{V}'_{pp} \tag{2}$$

where **U** is an orthogonal matrix whose columns are the orthonormal eigenvectors of **FF′**, **V′** is the transpose of an orthogonal matrix whose columns are the orthonormal eigenvectors of **F′F**, and $\mathbf{\Lambda}^{1/2}$ is a diagonal matrix containing the square roots of eigenvalues from **U** or **V** in descending order, which are called the singular values of **F**.

Taking into account that n > p, this matrix decomposition can be written, as a finite summation, as follows:

$$\mathbf{F} = \sum_{i=1}^{p} \lambda_i^{1/2}\mathbf{u}_i\mathbf{v}'_i \tag{3}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the *i*-th column eigenvectors belonging to **U** and **V** respectively, and $\lambda_i^{1/2}$ is the *i*-th singular value of **F**.

From (3), and after some computations, each element of **F** (each discrete function) can be written as follows:

$$\mathbf{f}_k = \lambda_1^{1/2}u_{k1}\mathbf{v}_1 + \lambda_2^{1/2}u_{k2}\mathbf{v}_2 + \ldots + \lambda_p^{1/2}u_{kp}\mathbf{v}_p \tag{4}$$

It is worth mentioning that the expression shown by (4) actually represents the decomposition of the discrete function of time $\mathbf{f}_k$ into a sum of a set of discrete functions ($\mathbf{v}_j$) which are orthogonal in nature (since they are the orthonormal eigenvectors of **F′F**), weighted by real coefficients resulting from the product of the *j*-th singular value of **F** by the *j*-th element of the eigenvector $\mathbf{u}_k$. Thus, $\mathbf{v}_j$ represents the *j*-th EOF and its coefficient $a_{kj} = \lambda_j^{1/2}u_{kj}$ is called the EOF score.

The sum of the singular values of **F** ($\lambda_i^{1/2}$) is equivalent to the total variance of the data matrix, and each *i*-th singular value offers a measurement of the explained variability (EV_i) given by EOF_i as defined by (5). Thus, the number of the chosen EOFs depends on the desired explained variability.

$$EV_i = \frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i} \times 100 \tag{5}$$

It is worth to mention that the main advantage of EOFs is their ability to determine the orthogonal functions that better adapt to the set of dynamic functions. This feature enables the mining of the signal immersed patterns, and allows EOF to overcome other signal processing tools, such as Fourier analysis, which always employ the same pre-defined pattern functions that are not always suited to represent specific dynamic behavior.

2.2 Support vector classifier

A support vector classifier (SVC) acquires decision functions that classify an input into one of the given classes through training using input–output (features-label) pair data. The optimal decision function is called the Optimal Hyper-plane (OH), and it is determined by a small subset of the training set which are called the Support Vectors (SV), using the concept of VC (Vapnik-Chervonenskis) dimension as the theoretical basis (Abe, 2010).

SVC needs a priori an off-line learning stage, in which the classifier has to be trained using a training set of data. Hence, the data have to be split into training and testing sets. Each element in the training set contains one "target value" (class labels) and several "attributes" (features). The objective of SVC is to yield a training data based model, which predicts the target values of the test data given only the test data features (Hsu *et al.*, 2010). Given a training set of features-label pairs ($\mathbf{x}_i, y_i$), $i = 1,..., l,$ where $\mathbf{x}_i \in R^n$ and $y \in \{1, -1\}^l$, for a two-class classification problem, the support vector classifier requires the solution of the optimization problem formulated in (6) (Hsu *et al.*, 2010).

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}+C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i\left(\mathbf{w}^{\mathrm{T}}\phi\left(\mathbf{x}_i\right)+b\right)\geq 1-\xi_i, \\ & \xi_i \geq 0 \end{aligned} \tag{6}$$

where $\mathbf{w}$ is an n-dimensional normal vector to the hyper-plane, b is a bias term, ξ_i is a slack variable associated with $\mathbf{x}_i$, C is the margin parameter, and $\phi(\mathbf{x}_i)$ is the mapping function from $\mathbf{x}$ to the feature space (Abe, 2010).

The mapping function $\phi(\mathbf{x}_i)$ is usually defined as the so called "kernel function" $\mathrm{K}(\mathbf{x}_i, \mathbf{x}_j)$, as shown in (7) (Hsu *et al.*, 2010).

$$\mathrm{K}\left(\mathbf{x}_i,\mathbf{x}_j\right)\equiv\phi\left(\mathbf{x}_i\right)^{\mathrm{T}}\phi\left(\mathbf{x}_j\right) \tag{7}$$

There are several kernel functions such as linear, polynomial, radial basis function (RBF), among others. In this work, RBF kernel is used because this function is capable of handling possible nonlinear relations between labels and features (Hsu *et al.*, 2010). This type of kernel is shown in (8).

$$\mathrm{K}\left(\mathbf{x}_i,\mathbf{x}_j\right)=e^{-\gamma\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2},\gamma>0 \tag{8}$$

Before training the SVC, it is necessary to identify the best parameters C of (6) and γ of (8) (Hsu *et al.*, 2010), as well as W_m that represents a weight factor used to change the penalty of class m (implicit into the optimization formulation), which is useful for training classifiers using unbalanced input data. For this purpose, an optimization problem is posed and solved, in this paper, via the swarm version of the mean-variance mapping optimization (MVMO$^{\mathrm{S}}$), firstly presented by Rueda and Erlich (2013).

3. PROPOSED APPROACH

This section depicts a novel approach to estimate post-contingency dynamic vulnerability regions (DVR) considering three phenomena regarding short-term stability (transient stability, voltage stability and frequency stability -TVFS-). The DVRs are composed by two areas: the *vulnerable* region and the *non- vulnerable* region, delimited by a hyper-plane.

A pattern recognition method based on empirical orthogonal functions (EOF) is used to determine the approximate spatial distribution of DVR. Then a support vector classifier (SVC) is used to estimate the post-contingency dynamic vulnerability status (PCDVS). The objective is to determine if the current operating state of the system is or not within the vulnerable region, which is achieved by evaluating the tendency of the system to change its conditions to a critical state as regards TVFS phenomena. Fig. 1 schematizes the proposed approach, highlighting the coupling between the off-line stage regarding the mapping of DVRs and the training of the SVCs, and its application for real-time prediction of vulnerability status.

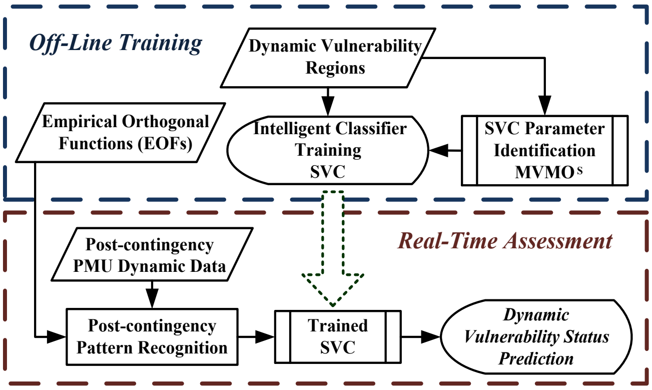
Off-Line Training
Dynamic Vulnerability Regions
Empirical Orthogonal Functions (EOFs)
Intelligent Classifier Training SVC
SVC Parameter Identification MVMO^{S}
Real-Time Assessment
Post-contingency PMU Dynamic Data
Post-contingency Pattern Recognition
Trained SVC
Dynamic Vulnerability Status Prediction

Fig. 1. Post-contingency vulnerability status prediction

3.1 Recognition of Post-contingency Dynamic Vulnerability Regions

The DVRs are empirically determined through mining post-contingency data recorded by PMUs. As an alternative to the limited availability of statistics on dynamic measurements, probabilistic Monte Carlo (MC) N-1 contingency simulations allow obtaining the dynamic responses representing those data registered by PMUs in real time, through evaluating the time domain system response. This procedure is widely explained by Cepeda *et al.* (2012).

Once the dynamic database has been structured, it is analysed using the EOFs with the aim of mapping the system DVRs, based on the patterns associated with the three TVFS phenomena. The corresponding *EOF scores* make vectors of real numbers that represent the system post-contingency dynamic behaviour patterns. These patterns vectors allow mapping the spatial DVRs on the coordinate system formed by the EOFs. Additionally, in order to prevent large numeric values give an erroneous interpretation of the vulnerability regions, it is suggested to normalize the pattern vectors before mapping the DVRs. In this research work, a linear normalization in the range [0, 1] has been adopted.

Each pattern vector has a specific associated *class label* based on the simulation resulting vulnerability status, as depicted by Cepeda *et al.* (2012). These class labels might correspond to a *non-vulnerable* case (label 0) or a *vulnerable* case (label 1), depending on whether one or more of the local protection relays associated with the phenomena under study (out-of-step relay -OSR-, low and high frequency relays -FR-, and low and high voltage relays -VR-) have triggered during the event progress. Using the resulting vector patterns and their corresponding vulnerable state class labels, the DVRs can be numerically mapped on the coordinate system formed by the main EOFs (Cepeda *et al.*, 2012).

In order to adequately capture the system response for different TVFS stability phenomena, several time windows (TW) have to be defined. These time windows are established according to the statistics of the triggering times of the relays, resulting from the MC simulation, influenced by the WAMPAC communication time delay (t_{delay}).

$$t_{\min} = \min_{i=1\ldots n}\left\{t_{OSR_i}, t_{VR_i}, t_{FR_i}\right\} - t_{delay} \tag{9}$$

where t_{OSRi}, t_{VRi} and t_{FRi} are the tripping times of out-of-step, voltage and frequency relays recorded in *n* MC repetitions, respectively. Typically, the out-of-step relay presents the fastest tripping time due to the fast time frame evolution of transient instability.

Since the post-contingency data comprise the samples taken immediately after the fault is cleared, the first time window (TW_1) is defined by the difference between t_{min} and the clearing time (t_{cl}).

$$TW_1 \le t_{\min} - t_{cl} \tag{10}$$

The rest of the time windows are defined based on the statistical concept of confidence interval related to Chebyshev's inequality (Han and Kamber, 2006), which settles that at least 89% of the data lie within three standard deviations (3σ).

$$TW_k \approx 3 \cdot std\left\{t_{OSR/VR/FR}\right\} + TW_{k-1} \tag{11}$$

where std{·} represents the standard deviation (σ) of the relay tripping time that most intersects the corresponding time window TW_k.

3.2 Prediction of Post-contingency Dynamic Vulnerability Status

DVRs are used to specify the relative position of the actual system dynamic state regarding its hyper-plane limit, which can be achieved by using a smart classifier. This paper employs a support vector classifier (SVC) with this aim.

The SVC needs a preliminary off-line learning stage. This task is performed by using the post-contingency database obtained via MC simulations and the corresponding associated DVRs. The data is divided into training and testing. Each element of the training and testing sets contains a *target value* (class label) and several *attributes* (pattern vectors that best represent each DVR for every TW).

Based on the two DVR associated regions (vulnerable and non-vulnerable), a two-class classifier is adopted in order to specify the system vulnerability status. It is worth mentioning that a SVC has to be trained for each TW.

There are two essential aspects to be considered before training the SVC:

- Choice of appropriate pattern vectors, showing the evolution of specific phenomena (TVFS). In this sense, a procedure for extracting and selecting the most relevant features is required. Thus, a method that maximizes the classification accuracy (CA) using decision trees (DT), originally introduced by Teeuwsen (2005), in combination with a certain explained variability (above 97%) has been employed to solve this problem.
- Identification of the best parameters of SVC. To this end, a parameter identification problem has been defined and solved in this paper by optimizing an appropriated objective function based on the maximization of the classification accuracy. This optimization problem is solved by means of MVMO^S.

3.2 Implementation for real-time assessment

In real-time application, the off-line trained SVCs will be in charge of classifying the PCDVS of the power system, using the actual post-contingency PMU voltage phasors and frequencies as relevant data. Firstly, the dynamic signals must be transformed to the corresponding EOF scores. For this purpose, the measured data have to be multiplied by the EOFs determined in the off-line training and stored in the control-center processor. Then, the obtained EOF scores will be the input data for the trained SVCs, which will be adequately selected depending on the corresponding TW. These SVCs will automatically indicate whether the system is within *vulnerable* or *non-vulnerable* regions associated to immediate post-disturbance short-term instability phenomena.

This vulnerability status prediction might be then used by a *real-time vulnerability assessment* module, in which the predicted status might be considered along with estimated vulnerability indices (i.e. measures of the actual system security level like those suggested Cepeda and Colome, 2012) to arrive at a more conclusive indicator (i.e. diagnosis) of the system vulnerability condition. The calculation of vulnerability indices and the coupling with the approach presented in this work will be thoroughly discussed in a future issue.

4. SIMULATION RESULTS

The proposed approach is tested on the IEEE New England 39 bus test system (Pai, 1989), slightly modified in order to satisfy the N-1 security criterion. The functionalities of MATPOWER (Zimmerman, 2013), and DIgSILENT Power Factory (DIgSILENT GmbH Web Page) were employed to perform MC simulations in order to create the data base of system dynamic performance. At every MC simulation, an operating condition (load and corresponding dispatch) and a contingency are randomly generated, such that the causes of system vulnerability could be transient instability, short-term voltage instability, or short-term frequency instability. Two types of events are considered: three phase short circuits and generation outage. The short circuits are applied at different random locations of the transmission lines at 0.12 s, followed by opening of the corresponding transmission line at 0.2 s (i.e. fault clearing time t_{cl}). Generation outages are applied at 0.2 s.

Both voltage components (magnitude and angle), and bus frequencies are considered as the potential input variables. A total number of 10,000 cases have been simulated, from which 7,600 are stable or *non-vulnerable* and 2,400 are unstable or *vulnerable*: 1,308 are transient unstable, 682 are frequency unstable, and 410 are voltage unstable.

4.1 *Mapping the DVRs*

First, several time windows are defined based on the MC statistics of the relay tripping times, and the procedure presented in Section 3.1. It is assumed that t_{delay} = 250 ms. The out-of-step relay time has a mean of 1.2252 s, a standard deviation of 0.3746 s, and a minimum value of 0.8342 s. Thus, vulnerability assessment has to be done in less than 0.5842 s (t_{min} = 0.8342 s - t_{delay}). For this reason, an adequate data window (TW_1) for TS phenomenon can be 300 ms (t_{min} - t_{cl} = 0.3842 s) starting from the fault clearing. In this test system, the tripping of voltage relays presents a mean of 4.1275 s, a standard deviation of 1.6872 s, and a minimum value of 3.22 s, whereas the frequency relay tripping time has a mean of 10.6829 s, a standard deviation of 2.4921 s, and a minimum value of 5.987 s. Using these values and (11), the rest of time windows are determined. This time-window definition is summarized in Table 1.

Table 1. Time window definition

Time Window	*std{$t_{OSR/VR/FR}$} (s)*	*$3\times std\{t_{OSR/VR/FR}\}+TW_{k-1}$ (s)*	*TW (s)*
TW_1	-	-	0.30
TW_2	0.3746	1.4238	1.50
TW_3	0.3746	2.6238	2.70
TW_4	0.3746	3.8238	3.90
TW_5	1.6872	8.9616	9.00

After TW definition, the corresponding EOFs are determined using the resulting MC dynamic data. Afterwards, the EOF scores (i.e. pattern vectors) are computed, and their related DVRs can be then mapped, using also the vulnerability status indicators. For instance, Fig. 2 presents the three dimensional distribution of the pattern vectors obtained from the voltage angles corresponding to TW_1. In the figure, the blue areas (enclosing the "vulnerable" pattern vectors represented by blue surfaces) represent the vulnerable regions; whereas the white area (behooving to the "non-vulnerable" green-diamond pattern vectors) corresponds to the non-vulnerable region. These areas have been empirically delimited, bordering the obtained pattern shapes which depend on the pattern vector spatial locations.

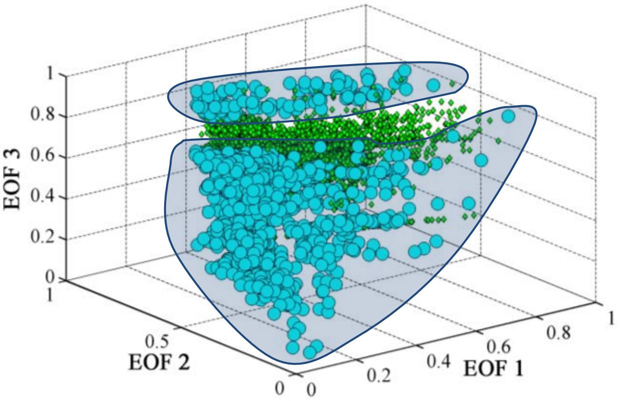

Fig. 2. TW_1 voltage-angle-based DVRs

4.2 *Building the SVC*

Next, it is necessary to choose the adequate number of EOFs that allows maintaining as much as possible of the variation presented in the original variables. For this purpose, the DT feature extraction and selection stage is applied, taking into account that EVi has to be at least 97%. As an illustration, Table 2 presents the number of chosen EOFs and their resulting EVis for TW_1 and TW_2.

Table 2. Feature extraction summary

Time Window	*Variable*	*Number of EOFs*	*EV_i (%)*
TW_1	V_{ang}	11	97.0325
	V_{mag}	5	97.2259
	Freq	2	99.7005
TW_2	V_{ang}	24	97.2816
	V_{mag}	10	97.1783
	Freq	2	99.6809

For instance, a summary of the feature selection results for TW_1 and TW_2 are presented in Table 3, where the selected features (those that offer the maximum CA) are additionally highlighted. After the selection of features for each TW, the MVMO^S-based SVC-parameter identification methodology is applied. For this purpose, LIBSVM (Chang and Lin, 2011) is used for running the SVC.

Table 3. Feature selection summary

Time Window	*Option*	*Feature*	*CA (%)*
TW_1	1	[V_{ang}]	97.477
	2	[V_{mag}]	96.779
	3	[Freq]	89.894
	4	[V_{ang}, V_{mag}]	98.030
	5	[V_{ang}, Freq]	98.103
	6	[V_{mag}, Freq]	97.773
	7	[V_{ang}, V_{mag}, Freq]	98.140
TW_2	1	[V_{ang}]	99.927
	2	[V_{mag}]	99.885
	3	[Freq]	99.832
	4	[V_{ang}, V_{mag}]	99.911
	5	[V_{ang}, Freq]	99.917
	6	[V_{mag}, Freq]	99.885
	7	[V_{ang}, V_{mag}, Freq]	99.906

The next step is to train the SVC. Table 4 presents the performance of the trained SVC for each TW, including, additionally, a performance comparison with other classifiers: decision tree classifier (DTC), pattern recognition network (PRN: a type of feed-forward networks), discriminant analysis (DA), and probabilistic neural networks (PNN: a type of radial basis networks). The performance of the classification is evaluated, for each TW, by using the mean of all iterations of a 10-fold CV classification accuracy (CA_i). Note that SVC outperforms all other classifiers in terms of classification accuracy.

Table 4. Classification performance

Classifier	*mean{CA_i} for Time Window (%)*				
	TW_1	*TW_2*	*TW_3*	*TW_4*	*TW_5*
DA	97.440	99.966	99.494	98.034	97.178
DTC	98.200	99.931	99.736	99.436	99.291
PRN	98.760	99.897	99.770	99.029	98.993
PNN	98.930	99.977	99.770	99.137	99.055
SVC	99.290	100.00	99.885	99.880	99.727

In order to validate the complete TVFS vulnerability status prediction, the percentage of the complete classification accuracy (N° samples correctly classified/Total N° samples), security (N° non-vulnerable samples correctly classified / Total N° non-vulnerable samples) and dependability (N° vulnerable samples correctly classified / Total N° vulnerable samples) are computed. These results are shown in Table 5, which also includes the number of cases correctly classified as well as those where the classification fails. Both security and dependability present more than 99% confidence level, which supports the excellent performance of the proposed approach in alerting about critical changes in the system condition.

Table 5. Vulnerability status estimation performance

Feature	Value
Non-vulnerable cases correctly classified	7,587
Vulnerable cases correctly classified	2,390
Non-vulnerable cases classified as vulnerable	13
Vulnerable cases classified as non-vulnerable	10
Complete classification accuracy (%)	99.770
Security (%)	99.829
Dependability (%)	99.583

6. CONCLUSIONS

This paper summarizes a novel post-contingency pattern recognition method for predicting the dynamic vulnerability status of an Electric Power System. The methodology begins with the determination of post-contingency dynamic vulnerability regions (DVRs) using Monte Carlo simulation and empirical orthogonal functions, that allow finding the best pattern functions for representing the particular signal dynamic behaviour. This proposal considers three different short-term instability phenomena as the potential causes of vulnerability (TVFS), for which several time windows have been defined. These DVRs are then used to specify the actual dynamic state relative position with respect to their boundaries, which is established using an intelligent classifier together with an adequate feature extraction and selection scheme. In this connection, SVC is used due to its property of being more robust to over-fitting problems when adequate parameters are selected. To overcome this drawback, a MVMO^S-based parameter identification method has also been applied. The proposed data-mining based pattern recognition method has shown excellent performance due to its high classification accuracy.

REFERENCES

Abe, S. (2010). *Support Vector Machines for Pattern Classification*, Second Edition. London: Springer.

Amin, M. (2000). Toward Self-Healing Infrastructure Systems. *Electric Power Research Institute (EPRI)*, IEEE.

Cepeda, J.C., Colomé, D.G. and Castrillón, N.J. (2011). Dynamic Vulnerability Assessment due to Transient Instability based on Data Mining Analysis for Smart Grid Applications. *IEEE PES ISGT-LA Conference*, Medellín, Colombia.

Cepeda, J.C., Rueda J.L., Erlich I. and Colomé, D.G. (2012). Recognition of Post-contingency Dynamic Vulnerability Regions: Towards Smart Grids. *IEEE PES General Meeting*, San Diego, USA.

Cepeda, J.C. and Colomé, D.G. (2012). Vulnerability Assessment of Electric Power Systems through identification and ranking of Vulnerable Areas. *International Journal of Emerging Electric Power Systems*, **13**, Issue 1.

Chang, C.C and Lin. C.J. (2011). LIBSVM: a library for support vector machines. [Online]. Available: *http://www.csie.ntu.edu.tw/~cjlin/libsvm*.

DIgSILENT GmbH Web Page. [Online]. Available: *http://www.digsilent.com.*

Han, J. and Kamber,M. (2006). *Data Mining: Concepts and Techniques*, (2nd edition), Elsevier, Morgan Kaufmann Publishers.

Hsu, C.W., Chang, C.C. and Lin, C.J. (2010). A Practical Guide to Support Vector Classification, Sunnyvale. [Online]. Available: *http://www.csie.ntu.edu.tw/~cjlin*.

Jollife I. (2012). *Principal Component Analysis*, 2nd. Edition, Springer.

Moslehi, K. and Kumar, R. (2010). Smart Grid - A Reliability Perspective. *IEEE PES Conference on Innovative Smart Grid Technologies*, Washington, DC.

Pai M. A. (1989). *Energy Function Analysis for Power System Stability*, Kluwer Academic Publishers.

Peña, D. (2002). *Análisis de Datos Multivariantes*, Editorial McGraw-Hill, España, chapter 1 – 8.

Rueda J.L., and Erlich I. (2013). Optimal Dispatch of Reactive Power Sources Using Swarmed Mean-Variance Mapping Optimization. *IEEE Symposium Series on Computational Intelligence*, pp. 29-36.

Savulescu, S. C. *et al*. (2009). *Real-Time Stability Assessment in Modern Power System Control Centers*, IEEE Press Series on Power Engineering 2009.

Teeuwsen S. P. (2005). *Oscillatory Stability Assessment of Power Systems using Computational Intelligence*. Doctoral Thesis, Universität Duisburg-Essen, Germany.

Yuan Zeng, Pei Zhang, Meihong Wang *et. al.* (2006). Development of a New Tool for Dynamic Security Assessment Using Dynamic Security Region. *International Conference on Power System Technology*.

Zimmerman D. (2013), MATPOWER, *PSERC*. [Online]. Software Available at: *http://www.pserc.cornell.edu/matpower*.