1

ANN Design for Fast Security Evaluation of Interconnected Systems with Large Wind Power Production

Helena Vasconcelos, and J. A. Peças Lopes, Senior Member, IEEE

Abstract--This paper presents the performed steps to design an Artificial Neural Network (ANN) tool, able to evaluate, within the framework of on-line security assessment, the dynamic security of interconnected power systems having an increased penetration of wind power production. This approach exploits functional knowledge generated off-line, the Linear Regression (LR) variable selection stepwise method to perform automatic Feature Subset Selection (FSS) and ANN to provide a way for fast evaluation of the system security degree. In order to choose the best input/output set of variables for the ANN tool, a comparative analysis is performed, regarding the obtained predicting error, by performing a statistical hypothesis test. The reduced error results confirm the feasibility and quality of the derived security structures.

Index Terms-- Artificial neural networks, Dynamic behavior, Feature selection, Interconnected systems, Linear regression, Security assessment, Wind generation.

I. INTRODUCTION

 \mathbf{I} is well know that, nowadays, the number of cross border power transactions is increasing due to the electricity market liberalization paradigm. At the same time, the need to decrease CO_2 emissions is leading to an increase of wind power penetration and of other renewable power sources in power systems. The conjunction of these two facts may lead to an increased use of the main transmission lines in interconnected systems that may provoke stressing operating conditions.

In fact, changes in wind power production may result from unexpected wind speed variations or from sudden disconnection of a large number of wind generators, as well as other Dispersed Generation (DG) units, due to the triggering of their protection relays following grid disturbances. Although Automatic Generation Control (AGC) takes care of interchange power flow deviations, it will take some time to eliminate these changes. Moreover, AGC will create a new dispatch solution and therefore will generate a new power flow solution inside the control area. Therefore, these wind power disturbances may lead to quasi-steady-state overloads in transmission lines that may provoke a set of undesired cascading events that afterwards may involve load curtailment or even system collapse.

Transmission System Operators (TSO) have been defining the levels of acceptance of wind generation and other DG on the basis of deterministic (n-1) steady-state security studies for worst case scenarios. More recently, TSO have started conducting also dynamic behavior and stability analysis studies following grid disturbances and subsequent operation of DG protection relays [1]. Again, these studies have been conducted for worst case scenarios, leading to severe limitations on system wind generation integration.

In order to increase acceptable wind penetration limits, interconnected systems with large wind power production require on-line system security assessment tools, able to make prediction of electrical current flow dynamic behavior in transmission lines following the occurrence of system disturbances or changes in wind power production, and, based on those predictions, provide preventive control measures if undesired line overload conditions are detected. Such prediction, for current or alternative operating conditions, may be obtained by performing full dynamic simulations of the interconnected system, namely, including AGC operation. However, this is not compatible with the time frame requirements for the management of secondary reserves or the acceptable time period for overload in transmission lines. In fact, in available power systems simulation tools, several minutes are required to obtain some seconds of a large interconnected system dynamic behavior. On the other hand, TSO usually accept no more than 20 minutes of 20% overload in transmission lines. This means that TSO dispatching centers require new tools able to provide fast and accurate forecasts of interconnected systems security.

To overcome these on-line time requirements, an ANN based approach was designed to emulate a set of security indices, characterizing the level of security of interconnected systems, following a disturbance that provokes the disconnection of a large share of wind power. For this purpose, a test system was created, modeling a two area control system. The main concerns of this approach were to obtain a security structure that can:

provide fast and accurate prediction of system security;

Helena Vasconcelos (e-mail: mhv@fe.up.pt) and J. A. Peças Lopes (e-mail: jpl@fe.up.pt) are with INESC Porto (Instituto de Engenharia de

Sistemas e Computadores do Porto) and FEUP (Faculty of Engineering of the University of Porto), Porto, Portugal.

 be exploited to provide preventive control measures, in order to reach system security if unsafe conditions are detected.

In order to gather enough information about the power system behavior, an algorithm able to automatically generate all possible operating conditions for this kind of power systems was developed. This includes a full dynamic simulation stage where system dynamic behavior is computed and the variables of interest are kept in a knowledge Data Set (DS). In a next stage, this data was analyzed in order to understand and characterize the security problem. From this stage, the candidate features to be used as ANN inputs and the security indices to be predicted were selected based on engineering judgments. In this procedure, the ANN preventive control purposes introduced some restrictions to the type of input feature to be considered. Namely, controllable variables (like dispatch conditions) were preferred, and non-directly controllable variables (like pre-fault power solution) were avoided. An automatic feature subset selection procedure was also conducted, based on the Linear Regression (LR) variable selection stepwise method, in order to eliminate irrelevant information and multicollinearity problems from the ANN input set, and therefore aiming to improve ANN predicting capability. For the ANN output set, two hypotheses were considered: 1) to use only one ANN to predict security for all the critical lines; 2) to use a different ANN to predict security for each critical line.

In order to choose the best input/output set solution for the ANN security tool, a comparative analysis was performed regarding the predicting error provided from each considered solution.

II. POWER SYSTEM MODELING

In this research, a test system was created based on the Portuguese – Spanish interconnected system. Fig. 1 presents the considered values of installed capacity, for each type of generating power, and the minimum and maximum load for the studied power system. The single-line diagram of the created system is presented in Fig. 2.



Fig. 1. Installed capacity and minimum and maximum load considered for the power system presented in Fig. 2 $\,$

Control area 1 corresponds to an approximation of the Portuguese transmission system, and control area 2 models an equivalent of the Spanish/European UCTE system. In order to reduce power system dimension, without losing relevant information, all the generating units of control area 1 are equivalent machines modeling similar generators operating in parallel in the same power plant. Regarding that this research was focused on the security of control area 1, the neighboring system was modeled by one busbar with equivalent generation units. In each control area, thermal, hydro and wind generating units were considered. In control area 2, nuclear units were considered as must run units and not participating in secondary frequency control. All the hydro and thermal units were considered to participate in primary and secondary frequency control.

In order to obtain the dynamic behavior of thermal, hydro and nuclear units, the usual corresponding local frequency regulator models as described in [2] were used, including also the voltage regulator behavior adopting an IEEE1 model type. Wind generators were modeled by a classical third order asynchronous machine model. The AGC system response is also modeled, adopting the traditional integral control approach and using the configuration described in [3]. According to this configuration, besides keeping the system frequency and interchange power between control areas in the specified value, changes in power production are distributed among generators through participation factors in order to maintain generating units at the most economic operating conditions. For all the parameters of the power system model, typical values were considered, and extracted from the full Portuguese - Spanish interconnected system.

III. DATA SET GENERATION

A special care was devoted to the Data Set (DS) generation step, in order to cover all possible operating conditions of the power system. In addition, several operating restrictions were mandatory to be included in order to filter out unrealistic scenarios, and therefore decrease computational time without compromising the knowledge data quality. Besides, this also avoids load flow convergence problems, in face of unfeasible conditions sampling. The developed algorithm includes the following main steps:

A. Sampling

Based on typical operating conditions, a structured Monte Carlo sampling method [4] is applied in order to produce all possible operating scenarios of the system, characterized by different settings of: system load level, wind power production levels, and import levels. These conditions were defined regarding their potential influence on the transmission lines power flow values during pre-fault and post-fault periods.

Based on available statistical data about wind power production of the Portuguese power system, the following dependencies were considered among wind parks of area 1:

```
WCF(Wind Park) = b + WCF(Control Area 1) \times m (1)
where WCF is the wind capacity factor, being defined has the
wind power production divided by the installed capacity in
operation; m and b are the slope and the y-intercept of the
best linear regression between WCF(Wind Park) and
WCF(Control Area 1). In order to introduce diversity among
wind parks, in each sampled scenario, m and b are randomly
sampled between \pm 3 \times their standard deviation values.
```



Fig. 2. Single-line diagram of the interconnected transmission system

B. Scheduling of Conventional Units

For each sampled scenario, the identification of several unit commitment solutions is performed, based on specified scheduling settings. To obtain a solution, in each control area the units are sequentially connected until the load supply being satisfied, constrained by:

- a pre-defined connecting order among power plants of each control area;
- the minimum and maximum number of available units in each power plant, being the availability of several units sampled in order to provide some scheduling diversity;
- the minimum and maximum technical limits of each unit;
- a primary control reserve (*PR*) criteria for each control area, namely:

$$PR$$
 > capacity of the largest unit in operation (2)

• a secondary control reserve (*SR*) criteria for each control area, namely the one presented in [5]:

$$SR > \sqrt{a \times L \max + b^2} - b \tag{3}$$

being Lmax = maximum estimated load in that time period; a = 10 MW and b = 150 MW.

C. Dispatch of Conventional Units

For each created scheduling scheme, a dispatch procedure randomly distributes the insufficiency of power production by the conventional units that were defined to be in operation, considering again their production limits.

D. Power Flow

For each dispatch solution, a load flow is solved in order to identify all the system pre-fault operating conditions.

E. Dynamic Simulation

For each feasible steady-state operating solution, the time simulation of system dynamic behavior is computed in order to characterize system security following a specified disturbance that affects wind power generation.

F. Data Set Recording

After each dynamic simulation computation, a pattern is added to the DS, being characterized by all the variables needed to describe the system pre-fault operating conditions and the dynamic behavior of power flows in transmission lines after the disturbance. From this set, the most relevant pre-fault features will be used as ANN inputs and the violated post-fault operating conditions will be selected for the ANN output set. Is this work, the following post-fault condition was analyzed: quasi-steady-state post-fault load in transmission lines, *Ifim* (120 s after the disturbance, involving therefore AGC operation).

Besides being used for ANN training, recorded variables must also enable the analysis of the generated patterns quality regarding the feasibility of the generated operating conditions. Moreover, recorded variables must also include the necessary input data in order to enable the dynamic simulation analysis of the generated scenarios.

G. Learning and Testing Set

Finally, the DS gathered in this way is afterwards randomly divided in two sets, creating a learning set, with 3/5 of the patterns used for training, and a testing set, with the remaining patterns used for performance evaluation.

IV. AUTOMATIC FEATURE SUBSET SELECTION

An automatic feature subset selection procedure was conducted through the results obtained from the Linear Regression (LR) method available in SPSS software tool [6]. In this procedure, the stepwise variable selection method was chosen in order to obtain a LR structure for the prediction of each security index (in our problem, the *Ifim* value in transmission lines). The stepwise method is an expedite procedure of identifying the best combination of input features (independent variables), among an initial set of candidate ones, to perform LR forecasting. In this procedure, features are sequentially entered into the LR model, by choosing the one that most explains, through the LR model, the total variation of the security index to predict. The advantage of this method, regarding others available expedite methods in LR, is that during the iterative procedure of including one more feature, the exclusion of the variables already in the model is tested (i.e., inclusion is not definitive).

This method aims to remove multicollinearity problems (i.e., high correlation among features) and irrelevant features from the LR input set. Therefore, the features selected this way were considered to be, among the initial set of candidate ones, a possible better input set for the ANN structure.

V. DESIGN OF THE ANN STRUCTURE

In this research, an ANN based tool was chosen, since it performs generally better than other tools in the fast dynamic security evaluation of power systems [7]. For ANN training, the MATLAB Neural Network Toolbox tool was used [8]. ANN parameters were found through the Levenberg-Marquardt backpropagation algorithm. Before starting the training stage, all the inputs and outputs variables are normalized to have zero mean and a standard deviation of one and ANN parameters are randomly initialized between -1 and 1. To perform overfitting control, besides considering a maximum epochs number, when a validation error increases for a specified number of iterations, the training is stopped. The used ANN was a two-layer feedforward network, with a tan-sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. To choose the number of hidden units, a rule described in [9] was used. According to this rule, the number of hidden units of a single hidden layer network is given by:

$$\frac{N_{LS}}{K} \times \frac{1}{Ni + No + 1} \tag{4}$$

where N_{LS} is the number of learning patterns; K is a constant factor $\in [5;10]$; Ni and No are the number of input and output variables.

VI. PREDICTING ERROR COMPARATIVE ANALYSIS

For the security problem under analysis, the predicting capability of trained ANN showed to be very sensitive to the initial values randomly considered for the ANN parameters. Therefore, the comparative analysis regarding the predicting capability provided by two different ANN input/output solutions can not be performed by simply comparing the regression errors of two trained ANN structures. In the presented work, this comparative analysis was carry out by performing a statistical hypothesis test, namely by comparing the mean of two unknown populations. In this procedure, the following steps were performed:

Step 1: For each input/output set solution, A and B, a large sample is generated by training 50 ANN with random initial parameters. For each trained ANN, a regression error, RE, is calculated with the testing set, namely the mean squared of predicting errors divided by the output variance.

<u>Step 2</u>: The sample mean (*RE*) and sample standard deviation (S_{RE}) of the 50 errors is computed for each generated sample, namely \overline{RE}_A , \overline{RE}_B , $S_{RE,A}$ and $S_{RE,B}$.

<u>Step 3:</u> Next, the following null hypothesis is tested:

$$H_0: \mu_{RE,A} = \mu_{RE,B}$$

(i.e., the 2 sets provide the same error), being μ_{RE} the mean of the unknown distribution of RE. If H_0 is proven to be wrong, then the following alternative hypothesis is accepted:

$$H_1: \mu_{RE,A} > \mu_{RE,B}$$

(i.e., among the two sets, set A provides worst predicting capacities). Under the assumption that H_0 is true, these inferences are made by calculating the following z value:

$$z = \frac{\overline{RE}_A - \overline{RE}_B}{\sqrt{S_{RE,A}^2 / 50 + S_{RE,B}^2 / 50}}$$
(5)

Assuming that z has a standardized normal distribution, a critical value c is obtained from this distribution, has being the value for which z has a predefined α probability of having a higher value (being α the level of significance). If z > c, H_0 is rejected with a confidence level of $(1-\alpha) \times 100\%$. The lower the α value for which H_0 is rejected, the stronger is the assumption that an ANN obtained from set A is less accurate when compared with an ANN obtained from set B.

VII. NUMERICAL RESULTS

A. Data Set Results

For the DS generation, the load was considered to change from light load scenarios to peak load scenarios. In the wind power sampling procedure, a 0.8 value was considered for the probability of each wind park being in operation, the number of connected units in each wind park was sampled between 80% and 100% of the total installed number, and the capacitor factor was considered to change from 10% to 90%. The import level provided from control area 2 was considered to change from 0 to 1700 MW. For the scheduling and dispatch of the conventional units, 2 different situations were taken into account – a thermal based and a hydro based dispatch scenario – which defined two different connecting order solutions among conventional units inside the scheduling procedure. Also, special care was taken in order to generate scenarios with higher spinning reserve mainly provided by hydro power plants. The idea was to include, in ANN structure, the system behavior regarding to likely preventive control solutions, having in mind the results presented in [10], where it was concluded that in systems with very high wind power production where secondary control is mainly provided by thermal power plants, the operation of additional or faster secondary control (like pumped hydro production) is required in order not to compromise the quality of generation control.

System security was evaluated regarding to a short-circuit that takes place in a sensitive line of the transmission system (one of the two parallel lines installed between buses 15 and 16 of Fig. 2). A duration of 300 ms was considered before the disconnection of the faulty line, leading to the lost of the nearest wind generating units due to the triggering of their under-voltage protection relays (which operate if the voltage drops bellow 0.9 p.u.). System was considered to be insecure if, 2 minutes after the disturbance, any transmission line current is 20% above the maximum current technical limit.

After applying the data set generation procedure earlier described, 4596 patterns were generated where the 7 transmission lines mentioned in Fig. 3 were identified has losing security for some of the generated patterns. A total number of 983 patterns were classified as insecure. Obviously, the connection between bus 15 and 16 is the one with the major number of insecure scenarios, since it loses one of the two parallel installed lines.



B. ANN Output Set

For the ANN output set, the following security indices were selected: *Ifim* of the 7 critical transmission lines presented in Fig. 3, where two hypotheses were considered:

- to use only one ANN to predict security for all the critical lines;
- to use a different ANN to predict security for each critical line.

C. ANN Input Set

Before ANN training, the generated data was analyzed in order to understand the security problem under analysis. From this procedure, 82 features were selected as candidate features for ANN inputs, describing:

- Total active load;
- Dispatch conditions in each conventional power plant;
- Number of connected units and production level in each wind park of control area 1;
- Initial pre-fault value of the electrical current in the 7 critical transmission lines.

In order to analyze the influence of removing non-directly controllable variables from the ANN input set, a smaller set with 63 features was selected from the previous one by removing the following variables:

- Dispatch conditions in each conventional power plant of control area 2;
- Initial pre-fault value of the electrical current in transmission lines.

In the hypothesis of designing a different ANN to predict security for each critical line, the feature subset selection procedure described before was also applied.

D. Predicting Error Comparative Analysis

Fig. 4 to Fig. 7 present the samples of 50 testing set regression error results, obtained by considering the following different ANN input/output sets to predict *Ifim* for line 15-16 (*Ifim*₁₅₋₁₆):

- Fig. 4: prediction of $Ifim_{15-16}$ with all the 82 candidate features (a 82-7-1 structure);
- Fig. 5: prediction of *Ifim*₁₅₋₁₆ with the 63 directly controllable features (a 63-9-1 structure);
- Fig. 6: prediction of *Ifim*₁₅₋₁₆ with the 41 features selected from applying the automatic FSS to the set of controllable variables (a 41-13-1 structure);
- Fig. 7: prediction of *Ifim* for all the 7 critical lines, with the 63 directly controllable features (a 63-8-7 structure).

From the application of the hypothesis test described before, when comparing the results described in Fig. 4 with the ones of Fig. 5, the hypothesis of improving prediction by removing non-directly controllable variables from the ANN input set is accepted with a confidence level of 99.9%. By making the comparison between the results of Fig. 5 with the ones of Fig. 6, the hypothesis of improving prediction by applying the automatic FSS method to the set of controllable features is also accepted with a confidence level of 95.5%. Finally, by comparing the results presented in Fig. 7 with the ones presented in Figures 4 to 6, the hypothesis of improving prediction by considering a separate ANN to evaluate the security for this critical line is accepted with a confidence level of almost 100%.



For illustration purposes, Fig. 8 depicts the predicting

errors provided by the most accurate ANN selected from the sample of Fig. 6 and describes the linear regression between the predicted and observed values.



Fig. 8. Predicted and observed values, by considering the most accurate ANN from the sample of Fig. 6 and the linear regression model results

The obtained results for the sample mean (RE) and standard deviation (S_{RE}) of the 50 regression errors, for this and the other 6 critical transmission lines, from considering the different input/output sets earlier described are presented in TABLE I. These results show that, for 3 of the transmission lines (lines 15-16, 14-18 and 16-18), removing non-directly controllable variables reduces prediction error with a confidence level higher than 99%. For 3 of the other lines (15-17, 16-4 and 20-21), the confidence level to accept an error improvement is less than 82%, and therefore H_0 can not be rejected. For line 16-6, the prediction error increases with a confidence level of 97.9% (however the error increase is very small). From these results, we may conclude that when removing non-directly controllable variables, there is no risk to increase considerably the prediction error. This allows the simplification of the preventive control algorithm that is under development.

TABLE I Obtained Sample Mean and Standard Deviation of the Regression Error for the Tested ANN Structures

	ANN structure									
	82-7-1		63-9-1		Ni-Nh-1				63-8-7	
	(all features)		(controllable		(features from FSS)				(controllable	
			features)						features)	
Line	\overline{RE}	S_{RE}	RE	S_{RE}	Ni	Nh	\overline{RE}	S_{RE}	\overline{RE}	S_{RE}
15-16	0.0025	0.0010	0.0020	0.0006	41	13	0.0019	0.0003	0.0074	0.0013
15-17	0.0017	0.0006	0.0016	0.0004	41	13	0.0013	0.0002	0.0053	0.0012
16-4	0.0026	0.0010	0.0026	0.0009	43	12	0.0025	0.0003	0.0067	0.0021
16-6	0.0030	0.0014	0.0035	0.0010	40	13	0.0032	0.0007	0.0070	0.0018
14-18	0.0023	0.0007	0.0018	0.0004	39	14	0.0018	0.0003	0.0076	0.0009
16-18	0.0024	0.0006	0.0020	0.0003	42	13	0.0018	0.0003	0.0099	0.0022
20-21	0.0021	0.0008	0.0020	0.0004	40	13	0.0017	0.0003	0.0057	0.0019
Ni Number of input features: Nh Number of hidden units										

Ni – Number of input features; Nh – Number of hidden units

Regarding the applied automatic FSS method to the set of controllable features, the obtained results showed that predicting errors are improved for five of the critical transmission lines (lines 15-16, 15-17, 16-6, 16-18 and 20-21) with a confidence level higher than 95%. For the remaining 2 lines, the confidence level to accept an error improvement is less than 85%, thus, H_0 can not be rejected. Therefore, the application of the automatic FSS method showed to be, in most of the times, able to remove non-desired features from the ANN input set.

Finally, for all the lines, an error reduction was observed when considering a separate ANN to predict security for each critical line, with an almost 100% confidence level.

VIII. CONCLUSIONS

The approach described in this paper presents a new dynamic security assessment concept and provides a new online security assessment tool able to deal with the impact of the presence, in multi control area systems, of large shares of wind and other DG generation, following system disturbances. The obtained reduced testing errors confirm the feasibility and quality of the approach used as a security assessment tool. Further research is being developed in order to exploit the ANN structure for the derivation of preventive control measures.

IX. ACKNOWLEDGMENT

The authors would like to thank the financial support of FCT within the "Programa Operacional Sociedade do Conhecimento (POS_C) do III Quadro Comunitário de Apoio".

X. REFERENCES

- [1] J.P. Sucena Paiva, J.M. Ferreira De Jesus, Rui M. Castro, Pedro F. Correia, João Ricardo, A. Reis Rodrigues, João Moreira, Bruno Nunes, "Transient Stability Study in the Portuguese Transmission Network with High Penetration of Wind Power" (translated from Portuguese), Presented at the CIGRE Symposium ERIAC, Paraguay, May 2004.
- [2] P. Kundur, "Power System Stability and Control", New York: McGraw-Hill, 1994.
- [3] A.J. Wood, B.F. Wollenberg, "Power Generation Operation and Control", ISBN 0-471-58699-4, John Wiley & Sons, Second Edition, 1996.
- [4] Van Acker, Vincent, et. al, "Data Generation using Automated Security Assessment for Neural Network Training," presented at North American Power Symposium, Laramie, WY., Oct. 1997.
- [5] UCTE Operation Handbook Policy 1 Load-Frequency control and Performance, Prelim v1.2, 3/10/2002.
- [6] SPSS Inc., "SPSS® Base 12.0 User's Guide", 2003.
- [7] J. N. Fidalgo, J. A. Peças Lopes, V. Miranda, "Neural Networks Applied To Preventive Control Measures For The Dynamic Security Of Isolated Power Systems With Renewables", IEEE Transactions on Power Systems, Vol. 11, No. 4, November 1996.
- [8] MathWorks, Inc., "MATLAB® Neural Network Toolbox User's Guide", Version 3.0, 1998.
- [9] Louis A. Wehenkel, "Automatic Learning Techniques in Power Systems", Kluwer Academic Publishers, ISBN 0-7923-8068-1, 1998.
- [10] G. Dany, "Power Reserve in Interconnected Systems with High Wind Power Production", Proc. Of the IEEE Porto Power Tech., Porto, September 2001.

XI. BIOGRAPHIES

Maria Helena Vasconcelos was born in Porto, Portugal, on March 19, 1973. She graduated in Electrical Engineering from the Engineering Faculty of Porto University (FEUP) concluded in July 1996, and obtained the MSc. from FEUP in October 1999. Since September 1996 she works as a researcher in the Power System Unit of INESC Porto. Currently, she is a Ph.D. student and Teaching Assistant at FEUP.

João A. Peças Lopes was born in Portugal in May 1958. He graduated in Electrical Engineering from the Engineering Faculty of Porto University (FEUP) in July 1981, and obtained the Ph.D. and Aggregation degrees also from FEUP in October 1988 and November 1996 respectively. Dr. Peças Lopes is presently Associate Professor with Aggregation at FEUP (http://www.fe.up.pt/) and Coordinator of the Power Systems unit at INESC Porto (http://power.inescn.pt/).