Reconstructing Missing Data in State Estimation With Autoencoders

Vladimiro Miranda, *Fellow, IEEE*, Jakov Krstulovic, Hrvoje Keko, Cristiano Moreira, and Jorge Pereira

***Abstract*—This paper presents the proof of concept for a new solution to the problem of recomposing missing information at the SCADA of energy/distribution management systems (EMS/DMS), through the use of offline trained autoencoders. These are neural networks with a special architecture, which allows them to store knowledge about a system in a nonlinear manifold characterized by their weights. Suitable algorithms may then recompose missing inputs (measurements). The paper shows that, trained with adequate information, autoencoders perform well in recomposing missing voltage and power values, and focuses on the particularly important application of inferring the topology of the network when information about switch status is absent. Examples with the IEEE RTS 24-bus network are presented to illustrate the concept and technique.**

***Index Terms*—Autoencoders, distribution management systems, energy management systems, neural networks, state estimation.**

I. Introduction

THE problem of missing signals in power system state estimation (SE) is traditionally treated in the context of the generation of pseudo-measurements to replace the data that should have arrived at the SCADA but, for one reason or another, must be considered as not received. Those artificial values are classically derived from historical data on the missing measurements, usually power or voltage values.

Pseudo-measurements are normally established before the state estimation procedure is run and may serve to improve the estimation of the unknown values or to restore the observability of parts of the system. This concept is now extended to provide the reconstruction of the status of switching devices which determine the topology of the network, in the absence of signals.

This is a new approach to restoring missing data in the context of an energy or a distribution management system, or still of a micro-grid or smart grid. It is based on the properties of an autoassociative neural network (NN) or autoencoder: when properly trained, any input pattern coherent with the real system will produce a similar output with negligible error. However, an input with some data inconsistency will generate a significant error between the input and the output vector. This property has been used to reconstruct missing sensor signals and will be used in this paper to reconstruct switching device status, as well as voltage and active and reactive power values.

Reconstructing missing signals from switches may play an extremely important role in SE in networks where topology is variable and must be determined before a classical estimation procedure is run. This may be the case of distribution networks and also in some types of micro-grids and smart grids. This paper, therefore, presents the *proof of concept* of a new approach that may be suitable for all these cases. Exercises in topology reconstruction will be presented, including a demanding case of split bus, showing a remarkable robustness to growing missing data. Furthermore, the paper advances evidence suggesting that a mosaic of local autoencoders could well be a flexible and economic strategy, instead of attempting to emulate the whole system with a single neural network.

II. Pseudo-Measurements and Topology

Pseudo-measurements have been seen traditionally as values used to replace missing information; their calculation derives from external processes such as historical data, prediction procedures, load curve assessment, calculations on an external network, etc. The importance of having external pseudo-measurements available was already dealt with, for instance, in [1]. In [2], one has an example of the concern for pseudo-measurements based on forecasted load data in systems with distributed generation. The problem of the generation of pseudo-measurements in distribution systems, using probabilistic models on historical load data, is addressed in [3]. The generation of pseudo-measurements as an internal process, however, is not usual and has not been considered.

The determination of the actual network topology has been dealt with mostly under the perspective of detecting topology errors. This is an old concern and already in [4], one may find models to identify bad data related to topology information. More recently, in [5] as an example, switch status information led to the division of data in "true" and "suspect" subsets and then, from a Bayesian model, *a posteriori* conditional probabilities are assigned to the possible combinations of suspect switches. These examples are not meant to represent all efforts in modeling and detection of topology errors: however, we see that the specific problem of reconstructing missing switch status information is not directly addressed, even if it could be argued that some of the techniques could be adapted for such use. Reconstructing missing signals is the purpose of this paper, which is distinct from the problem of identifying bad or erroneous data or dealing with explicit gross errors.

Manuscript received June 28, 2010; revised November 03, 2010, June 10, 2011 and September 17, 2011; accepted October 31, 2011. Date of publication December 08, 2011; date of current version April 18, 2012. This work was supported by partially integrated projects LASCA PTDC/EEA-EEL/104278/2008 and GEMS PTDC/EEA-EEL/105261/2008, both financed by FCT, Portugal. Paper no. TPWRS-00509-2010.

V. Miranda, H. Keko, C. Moreira and J. Pereira are with INESC TEC (INESC Technology and Science, coordinated by INESC Porto), Porto, Portugal, and also with FEUP, Faculty of Engineering, University of Porto, Portugal (e-mail: vmiranda@inescporto.pt; hkeko@inescporto.pt; jpereira@inescport.pt).

J. Krstulovic is with INESC TEC (INESC Technology and Science, coordinated by INESC Porto), Porto, Portugal (e-mail: jopara@inescporto.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2011.2174810

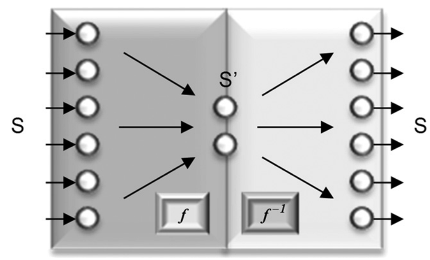

Fig. 1. Autoencoder neural network, with a *bottleneck* inner layer input and output layers of the same dimension, is trained to reproduce the input variables in the output. In the inner layer, one has a compressed set of values that encode, in a reduced dimension space S′, the values in S.

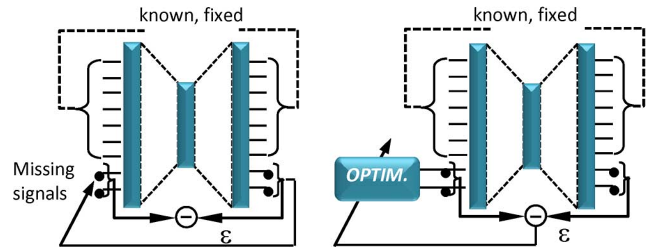

Fig. 2. Illustration of the algorithms. Left: POCS—the missing input is iteratively fed from the corresponding output until convergence is achieved. Right: constrained search—an optimization algorithm searches for the input values that minimize the input/output error on the missing signals.

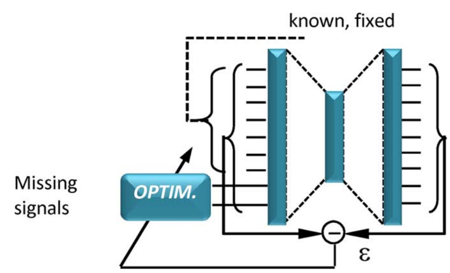

Fig. 3. Illustration of the algorithm. An optimization algorithm searches for the input values that minimize the input/output error on all the signals.

III. Autoencoders

A one-to-one mapping between points of a space of dimension m and a space of dimension n (with $n < m$ without loss of generality) is not a trivial problem. An approximation may however be achievable through the use of auto-associative neural networks or *autoencoders*, which are feedforward networks that are trained to mirror the input space S in their output.

In an autoencoder, the first half of the neural network approximates the function f that maps the input space to the space of compressed encoding S′ while the second half approximates the inverse function f^{-1}—the autoencoder is trained to display an output equal to its input. A trained autoencoder stores in its weights information about the data manifold.

Fig. 1 suggests the architecture of a simple NN with only one middle layer; this is just a pictorial representation and a higher number of layers may be adopted, although at the cost of heavier training [6], while obtaining increasing accuracy. There is no requirement though that the two NN halves should be built with a symmetrical architecture.

There is no *a priori* indication of an adequate reduction rate (measured as the ratio between the number of neurons in the smallest middle layer and the number of neurons in the inputs/output layers) to be adopted. This decision on the space reduction rate is dictated in present-day practice by trial and error and by the characteristics of the problem.

Autoencoders with thousands of inputs have been proposed [7], [8] with the purpose of using the reduced encoded values in S′ to represent compressed images, so that this representation would be subject then to distinct processing techniques such as identification and pattern recognition. For instance, face images could be identified and clustered according to sex, distinguished from non-faces, etc. [9]. A different application has been given in [10], to the reconstruction of missing sensor signals. This belongs to the type of "missing data" problems, where autoencoder properties have been used to reconstruct some missing input data in such a way that the reconstruction appears reproduced in the output, through minimizing a function of the input-output error. Publication [11] describes some useful properties of autoencoders in restoring missing sensor signals. Once the autoencoder is trained, its use may follow three basic approaches:

POCS—Projection Onto Convex Sets [12]: This model uses alternating linear projections on the input and output space to converge to the assumed missing value. If an input signal is missing, its variable can be set to a random value and this will produce a mismatch between input and output. Iteratively reintroducing the output value in the input will converge to a value that minimizes the input-output error—see Fig. 2.

Unconstrained or Constrained Search: The second and third models require an optimization algorithm to minimize the input-output error. The unconstrained search controls convergence by the error on the missing signals—see Fig. 2; the constrained search controls the error on all the outputs of the autoencoder—see Fig. 3. In [13], for instance, a genetic algorithm is used to optimize the error function.

An autoenconder with one hidden layer and linear activation functions performs the same basic information compression from space S to space S′ as principal component analysis [14]. With nonlinear activation functions and multiple layers, autoencoders chart the input space on a nonlinear manifold in such a way that an approximate reconstruction is possible with less error [15]. Plus, PCA does not easily show how to do the inverse reconstruction, which is straightforward with autoencoders.

Not many applications in power systems can be associated to autoencoders. Publication [16] is worth mentioning, where models for missing sensor signal restoration are discussed in the context of nonlinear power plant operation. Also in [17], one can find the proposal of autoencoders for restoring missing signals coupled with a wide area state predictor, as a part of a wide area controller.

IV. Autoencoder Preparation and Use

A. Creation of a Database for Training Purposes

In order to perform studies necessary to assess the value of autoencoders in power systems state estimation, a database of correct results had to be built. As test system, the IEEE RTS 24-bus network [18] was selected; then, minor modifications were introduced by defining a set of hypothetical measuring devices providing information on active and reactive power, as well as on voltage and angle, in buses and branches.

For all experiments to be conducted, the preparation of the load scenarios was done as follows:

1) A cumulative load curve was derived from data in [18]; then load levels were randomly sampled. So, scenarios from valley to peak of the load curve were created.
2) Each node had a proportional value allocated, according to the original load distribution.
3) Load variation was simulated by adding to all nodes perturbations generated by a Gaussian distribution with $\sigma = 5\%$.

The preparation of generation scenarios was as follows:

4) Locations for switches were defined (see Section VI).
5) A topological alternative is generated by randomly selecting which switches are closed and open.
6) A selection of a load scenario and topological alternative forms a study scenario, and an AC optimal power flow (OPF) defines the generation scenario.
7) A set of 20 000 study scenarios formed the data base.

Then, for the training of autoencoders:

8) A set of locations and measured values was defined in each experiment.
9) Training and test sets were defined, by extracting from the database vectors corresponding to the measurements, assumed exact.
10) An autoencoder with a convenient architecture was defined for each experiment and trained to learn the training set, using a current backpropagation approach, and was verified in the test set. Each autoencoder was composed of a single hidden layer, input neurons had linear activation functions, the remainder neurons had activation functions of the type $2\arctan()/\pi$ and $\varphi(\varepsilon)$, the performance criterion of the input-output errors ε, was the minimum absolute error (MAE).

B. Algorithmic Strategy in Restoring Missing Signals

Before addressing the general problem of the quality of the reconstruction of missing signals, a suitable algorithm had to be adopted. We tested the three methods from [11] referred in Section III and applied them to a set of sampled cases extracted from the database; some components of the input vector were specified as missing and initialized at a random value. The POCS algorithm is straightforward and simple. For the other two methods, we selected as the optimization algorithm a meta-heuristic denoted EPSO, for evolutionary particle swarm optimization, whose details may be found in [19]–[21]. We have concluded that with EPSO as the optimizer, the constrained search method is much more efficient than the other two methods in recomposing missing signals. The results in the following sections are, therefore, based on the application of an EPSO algorithm within the constrained search approach.

C. Dealing With Binary Missing Signals

A further novelty is introduced in this paper, associated with the fact that the missing signals, when related to switch status, are binary (denoting open or closed states). Previous publications (such as [10]) refer only to continuous signals.

For convenience, we defined the open state as -1 and the closed state as 1. With this definition, we added to the performance criterion $\varphi(\varepsilon)$ of the constrained search algorithm a penalty term leading to a new function $\Psi(\varepsilon)$ to be minimized in the form

$$\psi(\boldsymbol{\varepsilon}) = \varphi(\boldsymbol{\varepsilon}) + \mathrm{k}\sum_{\mathrm{i}=1}^{\mathrm{S}}(1 - \mathrm{x}_\mathrm{i}^2) \quad (1)$$

where ε is the vector of errors between input and output of the autoencoder, S is the number of switch status signals and $\mathrm{x_i}$ is switch status signal i (a real-valued signal output from the autoencoder); k is a scaling real parameter to be discussed in Section VI-D. This penalty factor k, for each signal, has the effect of pushing the signals to -1 or 1, solutions of the quadratic equation, and away from 0 (center of the interval [-1,1] when $\Psi(\varepsilon)$ is minimized. This avoids ambiguous solutions to be produced by the model.

V. Voltage and Power Pseudo-Measurements

This section is meant to show that defining pseudo-measurements with autoencoders is possible and that the reconstruction of the actual missing values may be done with a high degree of accuracy.

Two experiments are presented: case A, based on voltages and angles, and case B, based on active and reactive powers.

All experiments implied:

1) sampling states from the data base;
2) randomly considering some values to be missing;
3) submitting the sampled state vector with some components missing as input to the autoencoder;
4) recomposing the missing signal(s);
5) registering the error of the recomposed signal relatively to the value stored in the database.

A. Case A: Voltage Reconstruction

This case illustrates the capability of autoencoders for capturing implicit information and is not meant as an example of practical state estimation application.

In case A, an autoencoder has been trained using 10 000 patterns produced as described in Section IV-A; the input vector was composed of 24 nodal voltages and 24 nodal angles. The middle layer had 40 neurons. The error in the test set, after training, had an absolute average of 1.25×10^{-2}. Then, 400 new vectors were generated with random missing signals (between 1 and 10 missing values), voltage, and/or angle, and submitted to the autoencoder. The EPSO algorithm was used to recover the missing signals, using 20 particles and 20 generations.

The autoencoder rebuilt the missing signals close to the real values: in voltage signals, the average absolute error was of 1.95×10^{-3} p.u. relative to the missing values. Fig. 4 represents the error distribution in voltage signal reconstruction.

Given that voltages are the state variables and that these are independent, how is it possible that the autoencoder could reconstruct missing values with reasonable accuracy, if no redundancy is present in the input vector? The answer to this question

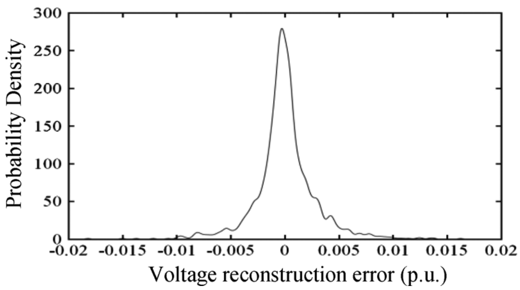

Fig. 4. Probability density function estimate of the error distribution for voltage reconstruction errors in all the 400 scenarios.

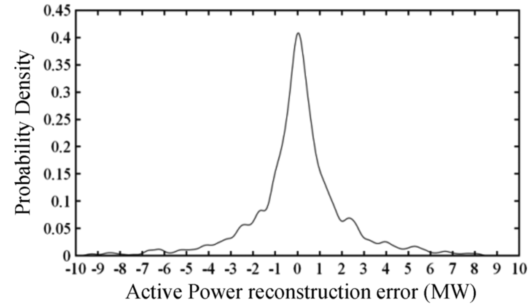

Fig. 5. Probability density function estimate of the error distribution for active power reconstruction errors in all the 400 scenarios.

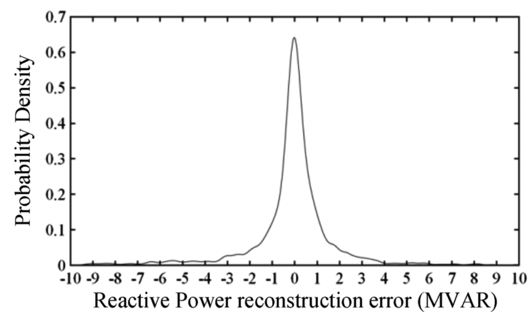

Fig. 6. Probability density function estimate of the error distribution for reactive power reconstruction errors in all the 400 scenarios.

reveals one important feature of an autoencoder: it learns *information*. In Case A, the autoencoder implicitly learned the underlying structure of nodal power injection or, in other words, it learned the multivariate probability density function of the geographical distribution of voltage profiles used in the training phase, associated with the load patterns used as basic data to generate voltage values.

Therefore, when presented with an incomplete pattern of voltages, it reconstructs the missing values over the nonlinear manifold that it learned (which is linked to the load distribution pattern used in the training phase). As a result, it produces a point over that manifold that is the closest to the known values of the incomplete input vector. One could therefore say that the autoencoder produces the "best guess" on the missing voltages using the knowledge it learned on system behavior.

This means that, in a way, the autoencoder is also taking in account the past history of the system to produce its proposal for the missing values, and underlines the importance of training an autoencoder with data from the real system and not just randomly generated.

B. Case B: Power Reconstruction

This case illustrates the reconstruction ability of power values when redundancy is clearly present in the input data.

In case B, the autoencoder has been trained using 10 000 patterns and the input vector composed of 58 active power and 58 reactive power values, with 20 nodal injections and 38 line flows. The middle layer had 85 neurons. The error in the test set, after training, had an absolute average of 1.83×10^{-2}.

Then, 400 new vectors were generated (using the same basic pattern as described before) with random missing power signals (between 1 and 10 missing values), active and/or reactive power, and submitted to the autoencoder.

Again the results are encouraging: the autoencoder could reconstruct the missing signals in all cases and with high accuracy. In active power signals, experiment B led to an average absolute error of 1.53 MW; in reactive power signals, the mean absolute error was 1.17 MVAR.

Figs. 5 and 6 represent the error distributions in active and reactive power signal reconstruction. These probability density function estimates were calculated from the results using the Parzen windows method [22] with Gaussian kernels.

They all display a high peak centered on error 0. These results confirm the hypothesis of the feasibility of using autoencoders to reconstruct or infer missing signals of continuous nature (real valued measurements) in the context of power system state estimation.

VI. Network Topology Identification

A. Challenges

In most SE procedures, topology determination precedes the estimation of voltage or power values consistent with the Kirchhoff laws. Topology is defined by the status of switching devices and this information is also present at the SCADA of the control center: it can be introduced by hand or from sensors in the field. Nevertheless, not only errors may be present but also missing information may happen. This is particularly sensitive in distribution systems or sub-transmission systems: experience has shown that a missing switch status signal is not unusual.

The problems researched and reported in this section are the following:

- Is an autoencoder able to reconstruct the network topology from a corrupt database with missing switch status data over the whole system?
- Is an autoencoder able to deal with the assumedly difficult problem of the *split bus*?
- Is a local autoencoder (trained with only data from a segment of the system) able to reconstruct locally the unknown topology with a performance comparable with a global autoencoder (trained with data from the whole system)?
- Is the autoencoder model a feasible alternative to reconstructing missing switch status signals?

To answer these questions, a series of experiments were performed. The test bed was the IEEE RTS 24-bus system [18]. To make the tests more significant, Gaussian noise with $3\sigma = 1\%$ of the largest measurement value was added to power and

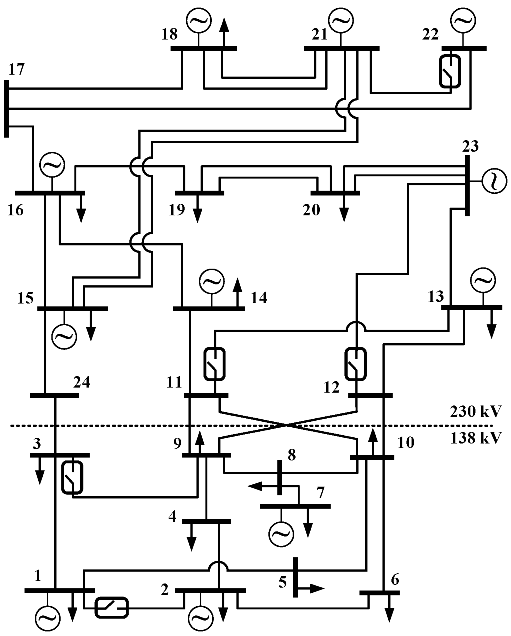

Fig. 7. IEEE 24 RTS with indication of the branches where switches have been introduced: 1–2, 3–9, 11–13, 12–23, and 21–22.

voltage data, simulating measurement errors (but not gross errors)—a large value if one takes in account only the precision class of the measuring devices.

B. Global Autoencoder and Global Data

To simulate missing signals over the whole network, new switches were added to the test system, as represented in Fig. 7. This topology information is of discrete nature and must be added to training and test sets of an autoencoder. The database referred to in Section IV-A was enlarged with fields related to switches and enriched with the results from OPF analysis over different system topologies randomly selected (including the case of all switches closed).

Switch status was defined as: closed: 1; open: -1. This is a convenient arrangement given that the output neuron activation functions of the autoencoder are of the arctan() type and saturate at those values. Then, an autoencoder with 121 inputs and a middle layer of size 60 has been trained with a training set of 3000 patterns.

C. Discovering the Network Topology

Experiment C was conducted by sampling a set of 50 000 scenarios with diverse active and reactive power measurements and switch status data. Some of the latter were then randomly removed. Five sets of trials have then been defined, each with 10 000 scenarios; in each set, a fixed number of switches (1 to 5) with status unknown have been tested. The training time, on a normal desktop, took about 10 min. The autoencoder was applied to these sets of scenarios with missing switch status information, using the same EPSO algorithm, using 30 particles and with random initialization of a switch status signal in $[-1, 1]$. The algorithm was allowed to run for 40 generations.

The results from experiment C are illustrated inTable I, which also includes information about the effect of the penalty term

TABLE I
RESULTS FROM APPLYING AN AUTOENCODER TO THE IEEE RTS IN RECOMPOSING MISSING SWITCH STATUS SIGNALS—10 × 1000 RANDOM SAMPLED SCENARIOS TESTED FOR EACH CASE

Miss. Signals	**Signal reconstructions**				**Topologies failed to recover**		**Efficiency in topology recovery**
	no penalty		**w/ penalty**				
	Corr.	*Wrong*	**Corr.**	**Wrong**	*no pen.*	**w/ pen**	
1	*9980*	*20*	**9980**	**8**	*20*	**8**	99.92%
2	*19996*	*34*	**19990**	**10**	*32*	**10**	99,90%
3	*29995*	*45*	**29983**	**17**	*45*	**16**	99,84%
4	*39974*	*26*	**39984**	**16**	*26*	**16**	99,84%
5	*49910*	*90*	**49971**	**29**	*81*	**29**	99,71%
Total	*149785*	*215*	**149920**	**80**	*204*	**79**	99,84%

in the optimization constrained search—to be discussed in the following section. The values in bold in Table I are the ones that are taken as final.

This table clearly shows that the autoencoder is efficient in recomposing the missing switch status signals and therefore in contributing to define the network topology. As expected, this efficiency increases when the number of switches with missing information decreases.

In terms of correct topologies identified, the efficiency obtained is a remarkable value of 99.90% or above when one or two signals are missing. But this efficiency is still above 99.71% when the information is missing about 5 switches simultaneously. Albeit this massive absence of information the autoencoder is able to reconstruct it.

In terms of percentage of correct signal reconstructions, the results are still more impressive: above 99.95% in all cases, or an error rate of 0.05% on average. Furthermore, this rate is quite constant in all sets: namely, for the set of scenarios with 5 simultaneous missing signals, the error rate is only of 0.06% which is extremely low in fact, considering the amount of missing information in this case.

D. Benefit From the Penalty Term

The addition of the penalty term changes the shape of the objective function to favor solutions for a missing signal x in the form $\mathrm{x} = -1$ or $\mathrm{x} = 1$. Fig. 8 illustrates the shape of the objective function for a selected case of 1 sensor missing. In this figure, the x axis represents the value of x between -1 and 1 that is produced as output of the autoencoder (a real value), and the y axis represents the cost function $\Psi(\varepsilon)$ in the last iteration, as a function of the value of x—see (1).

In the case represented, the penalty term has been useful in defining the final solution. An optimization to recover the missing signal, without the penalty term—by setting $\mathrm{k} = 0$ in (1)—would give an optimum close to the center (lower dashed line) but slightly leaning to the left. Rounding this value to the closest extreme would give a reconstructed signal of -1. With $\mathrm{k} = 0.5$ and $\mathrm{k} = 1$, one already has the definition of $\mathrm{x} = 1$ as the optimum (which was the correct reconstruction).

All experiments led to the conclusion that a moderate coefficient k is enough to assure good corrective effects induced by the penalty term. The solution has shown to be rather insensitive to a range of values of k—the results below are taken with $\mathrm{k} = 1$. This did not work in all cases but a net benefit has nevertheless been observed.

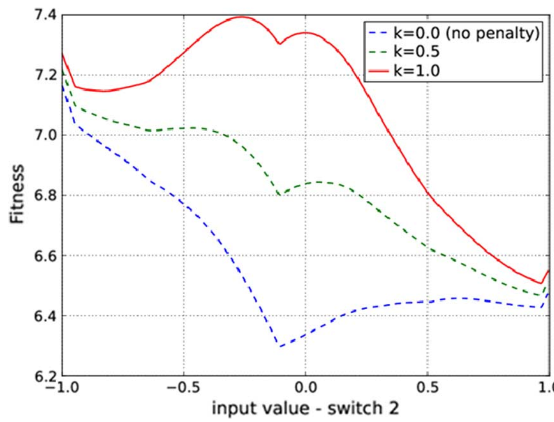

Fig. 8. Example of the effect of the penalty term k in the shape of the objective function to be minimized: a higher k raises the value in the center of the interval $[-1, 1]$ and displaces the global minimum from this region to the correct extreme $(\text{correct value} = 1)$; without penalty, the minimum is at a negative value and rounding leads to an incorrect answer.

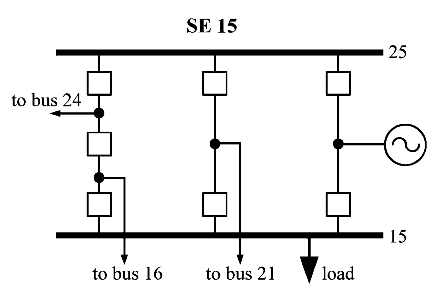

Fig. 9. Connection scheme in bus 15 with 7 switches.

E. Split Bus

The identification of switch status in the case known as *split bus* is considered difficult and therefore a study on this case is included, focusing on the substation corresponding to bus 15. The switch configuration is adapted from [23] and is represented in Fig. 9. In this experiment, cases leading to total generator or line disconnection were not considered because they have the same effect as tested in the previous experiment.

A 125-60-125 autoencoder was trained in the general fashion described above, and the experiment conducted assuming 1 to 7 missing signals. Table II summarizes the results for 1000 random trials for each case of a fixed number of missed signals. It is evident that up to a reasonable number of randomly missing information (5 out of 7), the autoencoder reconstructs the correct topology with remarkable efficiency.

It is even extraordinary that, in the total absence of information (all 7 switches with status unknown), a correct recovery is still achieved in about 80% of the cases.

F. Local Autoencoder versus Global Autoencoder

This last experiment aimed at testing the hypothesis that a small autoencoder, collecting only local limited information, can perform as efficiently as an autoencoder trained with complete system measurement data. The split bus problem was again addressed, but now with a 31-25-31 autoencoder collecting information from bus 15 and the region defined by the neighboring buses (see Fig. 10). The training time now took only around 1 min on average.

TABLE II
RECONSTRUCTION OF MISSING SWITCH SIGNALS IN BUS 15

Miss. Signals	**Signal reconstructions**				**Topologies failed to recover**		**Efficiency in topology recovery**
	no penalty		**w/ penalty**				
	Corr.	*Wrong*	**Corr.**	**Wrong**	*no pen.*	**w/ pen**	
1	*1000*	*0*	**1000**	**0**	*0*	**0**	**100,00%**
2	*1999*	*2*	**1999**	**1**	*1*	**1**	**99,90%**
3	*2995*	*5*	**2997**	**3**	*2*	**2**	**99,80%**
4	*3992*	*8*	**3993**	**7**	*4*	**4**	**99,60%**
5	*4975*	*25*	**4973**	**27**	*9*	**9**	**99,10%**
6	*5762*	*238*	**5756**	**244**	*59*	**56**	**94,40%**
7	*6024*	*976*	**5999**	**1001**	*188*	**182**	**81,80%**
Total	*26747*	*1254*	**26717**	**1283**	*263*	**254**	**96,37%**

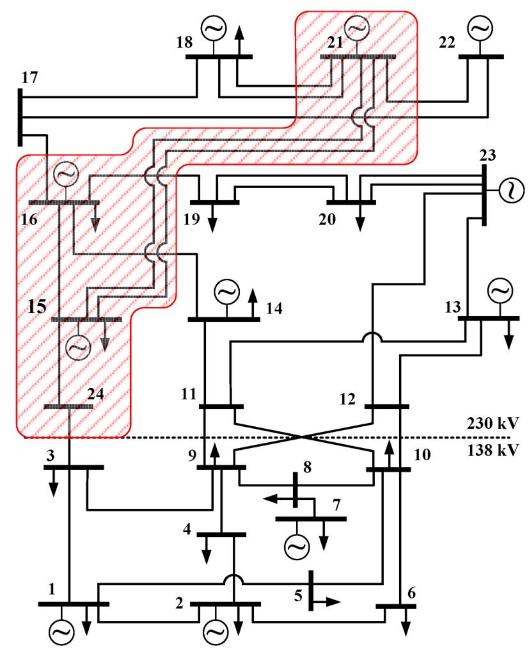

Fig. 10. Area with measurements feeding the local autoencoder for bus 15.

TABLE III
RECONSTRUCTION OF MISSING SWITCH SIGNALS IN BUS 15 USING A LOCAL AUTOENCODER

Miss. Signals	**Signal reconstructions**				**Topologies failed to recover**		**Efficiency in topology recovery**
	no penalty		**w/ penalty**				
	Corr.	*Wrong*	**Corr.**	**Wrong**	*no pen.*	**w/ pen**	
1	*1000*	*0*	**1000**	**0**	*0*	**0**	**100,00%**
2	*1997*	*3*	**1998**	**2**	*2*	**1**	**99,90%**
3	*2991*	*9*	**2993**	**7**	*5*	**4**	**99,60%**
4	*3975*	*25*	**3983**	**17**	*11*	**7**	**99,30%**
5	*4834*	*166*	**4833**	**167**	*67*	**63**	**93,70%**
6	*5443*	*557*	**5449**	**551**	*178*	**172**	**82,80%**
7	*5682*	*1318*	**5692**	**1308**	*332*	**321**	**67,90%**
Total	*25922*	*2078*	**25948**	**2052**	*595*	**568**	**91,89%**

The results for this experiment are in Table III and show that the performance of the local autoencoder, up to 4 in 7 missing switch signals in bus 15, is almost as good as the performance of the global autoencoder, working with the full set of measurements for the whole system.

The performance degrades somewhat when the amount of missing signals approaches the totality of the 7 switches, which is not surprising given that this becomes an important percentage of the signals admitted to the autoencoder. Nevertheless, the difference is not large compared to the performance of the global autoencoder given in Table II.

This experiment also answers one possible concern, about the time and effort to train the neural networks. Although the training is done offline, it is known that it becomes costly when the size of neural networks increase. On a regular desktop computer, training the global autoencoder took on average about 14 min, while training the local autoencoder required only about 1 min.

Given that the impact of a variable is usually mostly local and that the effects in remote areas of a system may be negligible, one may envision using, in practice, a coordinated set of local autoassociators instead of a single global one. This eliminates worries about the relation between the system size and the need for large scale neural networks. Plus, this modular architecture would allow easy adaptation for network growth, because only a minor number of local autoencoders would require retraining, in case of network changes.

VII. Conclusions

A. Acquired Knowledge

The reconstruction of missing signals, either voltage and power values or switch status information, is a way of generating pseudo-measurements necessary to SE. The research reported in this paper and the results obtained from a large number of tests illustrate that autoencoders do allow the correct reconstruction of the missing signals, either in voltage or power. Furthermore, the underlying network topology was discovered in a vast majority of cases, even if a considerable number of switch status information was missing.

The autoencoders tested had a moderate dimension, when compared with the size of thousands of inputs referred to in some publications, devoted to image compression [6]. Therefore the storage of the weight matrix for a local autoencoder is not a critical factor. Although training may be somewhat demanding, it is only required offline; online, the EPSO algorithm is quite fast and will cause no time burden in a DMS/EMS environment. The training time was of moderate length for 125-60-125 autoassociative neural network and only about 1 min for the 31-25-31 autoencoder.

We speculated that an autoencoder based on local information would perform as well as a global autoencoder, for local values. This hypothesis has been tested with success, especially in the usually considered difficult problem of the split bus: remarkable recovery rates have been observed even in cases with an impressive loss of information.

We found no specific limitation to the use of the technique. A "minimum requirement" for an autoencoder to give good results does not depend on the power system and cannot be defined quantitatively: it is related with the data available and naturally there is a degradation of performance with poor quality data, as in any SE procedure. Therefore, a local autoencoder must be designed to fit the characteristics of the local network—including available data and redundancy to consider.

The only significant limitation of the use of autoassociative neural networks would be on size—but we have shown that autoencoders with 100 inputs are feasible and this is more than enough to build efficient local autoencoders.

The autoassociator model may provide an answer to this question: what is the most likely topology identifiable as coherent with the electric data available. Of course, insufficient data or poor redundancy will generate a poorer quality response—but this is common to any SE procedure. Also, if data are filtered and gross errors in electrical values are eliminated or corrected before submitting data to the autoencoder, a better quality result is to be expected. We have not investigated the effect of gross errors in electrical data on the performance of the autoencoders—however, missing switch signals are gross errors themselves and we have tested autoencoders in stressing conditions (see the split bus case).

From our experience, this is a tool of general application.

B. Visions for the Future

The autoencoder output may be generated upstream of the SE process, thus not biased by assumptions on topology or assumptions on gross errors that may contaminate the SE. It has the potential to be combined with existing methods to reinforce the credibility of results that presently are obtained based on a heuristic approach, on the application of rules or even on resorting to past data when data are missing for the present situation. It is a systematic model with a mathematical background and not an empirical method, as most present-day practices may be argued to be. Plus, the error in the autoencoder output may become an indicator of the reliability of the reconstruction, which is a further advantage.

The results achieved suggest that local autoencoders can perform as well as global mappers for the whole power system. So, instead of attempting to build a huge neural network, a mosaic of small autoassociative networks may be envisioned to perform the reconstruction of the topology in specific zones of the power network.

Also, because an autoencoder is able to discover the underlying topology in electrical data, one may also use the technique as confirmation to other methods and to detect gross errors in switch status information. One simply has to compare the response of the autoencoder (assuming no knowledge of switch status) to the switch status information gathered at the SCADA and trigger an alarm if there is a mismatch.

Aspects related with the architecture of the autoencoders and the training functions have not been addressed, but one may envision that further accuracy may be obtained if novel strategies for autoencoder building and training are adopted.

The capacity to reconstruct topology may promote autoencoders to a useful tool in distribution management systems, especially for lightly meshed distribution networks, where topology variables are, in a way, dominant. Last, autoencoders may be a promising solution for local estimation in micro-grids, done by a micro-grid central controller, in the context of the smart grid paradigm, whenever no SE exercise is possible and yet an estimation of the state of the system is necessary locally.

This paper presented what may be called the *proof of concept* and, given the positive results obtained, a larger challenge is ahead: to integrate the approach into practical applications, namely designing a successful hybrid with existing methods.

References

[1] A. J. A. Simões Costa and M. Tardio Arze, "Critical pseudo-measurement selection for unreduced external system modeling," *Int. J. Elect. Power Energy Syst.*, vol. 18, no. 2, pp. 73–80, Feb. 1996.

[2] A. S. Costa and M. C. dos Santos, "Real-time monitoring of distributed generation based on state estimation and hypothesis testing," in *Proc. IEEE Lausanne Power Tech 2007*, Jul. 2007, pp. 538–543.

[3] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Modeling of pseudo-measurements for distribution system state estimation," in *Proc. IET-CIRED Seminar: SmartGrids for Distribution*, Frankfurt, Germany, Jun. 2008, pp. 1–4.

[4] F. F. Wu and W.-H. E Liu, "Detection of topology errors by state estimation," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 176–183, Feb. 1989.

[5] E. M. Lourenco, A. S. Costa, and K. A. Clements, "Bayesian-based hypothesis testing for topology error identification generalized state estimation," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1206–1215, May 2004.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[7] G. W. Cottrell, P. Munro, and D. Zipser, "Learning internal representations from gray-scale images: An example of extensional programming," in *Proc. 9th Annu. Conf. Cognitive Science Society*, Seattle, WA, 1987.

[8] M. K. Fleming and G. W. Cottrell, "Categorization of faces using unsupervised feature extraction," in *Proc. IJCNN—Int. Joint Conf. Neural Networks*, San Diego, CA, Jun. 17–21, 1990, vol. 2, pp. 65–70.

[9] B. Golomb and T. Sejnowski, "Sex recognition from faces using neural networks," in *Applications of Neural Networks*, A. Murray, Ed. Norwell, MA: Kluwer, 1995, pp. 71–92.

[10] S. Narayanan, R. J. Marks I1, II. L. Vian, 1.1. Choi, M. A. El-Sharkawi, and B. B. Thompson, "Set constraint discovery: Missing sensor data restoration using auto-associative regression machines," in *Proc. 2002 Int. Joint Conf. Neural Networks, 2002 IEEE World Congr. Computational Intelligence*, Honolulu, HI, May 12–17, 2002, pp. 2872–2877.

[11] B. B. Thompson, R. J. Marks, and M. A. El-Sharkawi, "On the contractive nature of autoencoders: Application to missing sensor restoration," in *Proc. Int. Joint Conf. Neural Networks*, Jul. 2003, vol. 4, pp. 3011–3016.

[12] S. Narayanan, R. J. Marks I1, II. L. Vian, 1.1. Choi, M. A. El-Sharkawi, and B. B. Thompson, "Set constraint discovery: Missing sensor data restoration using auto-associative regression machines," in *Proc. Int. Joint Conf. Neural Networks, IEEE World Congr. Computational Intelligence*, Honolulu, HI, May 12–17, 2002, pp. 2872–2877.

[13] M. Abdella and T. Marwala, "The use of genetic algorithms and neural networks to approximate missing data databases," *Comput. Inf.*, no. 24, pp. 577–589, 2005.

[14] I. T. Jolliffe, "Principal component analysis," in *Springer Series Statistics*. New York: Springer, 2002.

[15] N. Japkowitz, S. J. Hanson, and M. A. Gluck, "Nonlinear autoassociation is not equivalent to PCA," *Neural Comput.*, vol. 12, no. 3, pp. 531–545, Mar. 2000.

[16] W. Qiaoa, Z. Gao, R. G. Harley, and G. K. Venayagamoorthy, "Robust neuro-identification of nonlinear plants electric power systems with missing sensor measurements," *Eng. Appl. Artif. Intell.*, no. 21, pp. 604–618, 2008.

[17] S. Mohagheghi, G. K. Venayagamoorthy, and R. G. Harley, "Optimal wide area controller and state predictor for a power system," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 693–705, May 2007.

[18] IEEE RTS Task Force of APM Subcommittee, "IEEE reliability test system," *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 6, pp. 2047–2054, Nov/Dec. 1979.

[19] V. Miranda and N. Fonseca, "EPSO—Best-of-two-worlds meta-heuristic applied to power system problems," in *Proc. WCCI 2002—World Congr. Computational Intelligence—CEC—Conf. Evolutionary Computing*, Honolulu, HI, May 2002.

[20] V. Miranda, H. Keko, and A. Jaramillo, "EPSO: Evolutionary particle swarms," in *Advances Evolutionary Computing for System Design*, ser. Computational Intelligence, L. Ja, Ed. New York: Springer, 2007, vol. 66, ch. 6, pp. 139–168.

[21] V. Miranda, H. Keko, and Á. J. Duque, "Stochastic star communication topology evolutionary particle swarms (EPSO)," *Int. J. Comput. Intell. Res.*, vol. 4, no. 2, 2008.

[22] E. Parzen, "On the estimation of a probability density function and the mode," *Ann. Math Statist.*, vol. 33, 1962.

[23] R. Billinton, P. K. Vohra, and S. Kumar, "Effect of station originated outages a composite system adequacy evaluation of the IEEE reliability test system," *IEEE Power App. Syst.*, vol. PAS-104, pp. 2249–2656, Oct. 1985.

Vladimiro Miranda (M'90–SM'04–F'05) received the B.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal, in 1977 and 1982, respectively.

In 1981, he joined FEUP and currently holds the position of Full Professor. He also has been a researcher at INESC since 1985 and is currently Director at INESC Porto, the leading institution of INESC TEC—INESC Technology and Science, an advanced research network in Portugal. He has authored many papers and been responsible for many projects in areas related with the application of computational intelligence to power systems.

Jakov Krstulovic received the B.Sc. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2008. He is currently pursuing the Ph.D. degree at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Croatia.

He also worked as a researcher and teaching assistant at the University of Split. Since October 2010, he has been a Junior Researcher in Power Systems Unit of INESC TEC, Porto, Portugal. His primary research interests are power system state estimation and large scale integration of the wind power.

Hrvoje Keko received the B.Sc. degree in electrical power systems engineering from the Faculty of Electrical Engineering and Computing of the University of Zagreb, Zagreb, Croatia. He is presently pursuing the Ph.D. degree enrolled in the Sustainable Energy Systems Program at FEUP—Faculty of Engineering of the University of Porto, Porto, Portugal. He is also a researcher at INESC TEC in its Power Systems Unit. His interests include computational intelligence tools, wind power forecasting, and impact of electrical transportation in power system planning and operation.

Cristiano Moreira is pursuing the M.Sc. degree in electrical and computer engineering at the Faculty of Engineering of the University of Porto, Porto, Portugal, and a young researcher at INESC TEC in its Power Systems Unit.

Jorge Pereira received the B.Sc. degree in applied mathematics from the Faculty of Sciences of the University of Porto, Porto, Portugal, in 1991, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from FEUP in 1995 and 2002, respectively.

In 1991, he joined INESC Porto and is presently a research manager at INESC TEC. Since 1995, he has been with the Faculty of Economics of the University of Porto, where he is currently an Assistant Professor. He has collaborated in several projects related to the development of DMS and the application of soft computing techniques to power systems, namely to the state estimation problem.