This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

A Knowledge-based Framework for Power Flow and Optimal Power Flow Analyses

Alfredo Vaccaro, *Senior Member, IEEE* and Claudio A. Cañizares, *Fellow, IEEE*

***Abstract*—This paper proposes the application of formal methods for knowledge discovery from large quantity of data to reduce the complexity of Power Flow (PF) and Optimal Power Flow (OPF) problems. In particular, a knowledge-based paradigm for PF and OPF analyses is used to extract complex features, hidden relationships, and useful hypotheses potentially describing regularities in the problem solutions from operation data-sets. This is realized by designing a knowledge-extraction process based on Principal Components Analysis (PCA). The structural knowledge extracted by this process is then used to project the problem equations into a domain in which these equations can be solved more effectively. In this new domain, the cardinality of the PF and OPF problem is sensibly reduced and, consequently, the problem solutions can be obtained more efficiently. The effectiveness of the proposed framework is demonstrated with numerical results obtained for realistic power networks for several operating conditions.**

***Index Terms*—Power flow analysis, optimal power flow analysis, knowledge discovery, intelligent systems, system reduction, big data.**

NOMENCLATURE

Sets

$\mathcal{N}_{\mathcal{P}}$	Set of the buses in which the active power is specified.
$\mathcal{N}_{\mathcal{Q}}$	Set of the buses in which the reactive power is specified.
$\mathbf{F}$	Set of PF equations.

Indices

i_x	i_x^{th} component of the state vector.
i_u	i_u^{th} component of the control/decision vector.
k	k^{th} sample time.
i	i^{th} element of the set $\mathcal{N}_{\mathcal{P}}$.
i_g	i_g^{th} generator.
j	j^{th} element of the set $\mathcal{N}_{\mathcal{Q}}$.
q	q^{th} component of the objective function vector.
p	p^{th} component of the vector of equality constraint functions.
m	m^{th} component of the vector of inequality constraint functions.

Parameters

n_x	Number of components of the state vector.
n_u	Number of components of the control/decision vector.
$\overline{T}$	Integer sample time interval of observed variables.
n_p	Number of equality constraint functions.
n_m	Number of inequality constraint functions.
N	Total number of buses.
$Y_{ij}\angle\theta_{ij}$	Element of the bus admittance matrix.
L	Rank of the matrix $\mathbf{X}$.
α_{G_i}	Dispatch factors.
n_G	Number of generators.
P_i^{SP}	Real power injection specified.
Q_j^{SP}	Reactive power injection specified.
ω_q	Weighting coefficient.
N_x	Number of state variables.
N_{PC}	Number of principal components.

Variables

$V_i\angle\delta_i$	i^{th} bus voltage in polar coordinates.
$\mathbf{x}$	Vector of the dependent variables.
$\mathbf{u}$	Vector of the control/decision variables.
ζ	Components of the state vector in the transformed domain.
$\mathbf{r}$	Residual error vector.
$\mathbf{A}$	Matrix of left singular vectors.
$\mathbf{\Omega}$	Matrix of right singular vectors.
$\mathbf{\Sigma}$	Diagonal matrix of singular values.
$\mathbf{S}$	Matrix of the factor scores
$\mathbf{A}_{N_{PC}}$	First N_{PC} columns of the matrices $\mathbf{A}$.
$\mathbf{\Sigma}_{N_{PC}}$	First N_{PC} columns of the matrices $\mathbf{\Sigma}$.
$\mathbf{\Omega}_{N_{PC}}$	First N_{PC} columns of the matrices $\mathbf{\Omega}$.
$\mathbf{J}$	Jacobian matrix of the PF equations.
P_{D_j}	Active power demands.
$P_{G_{i_g}}$	Active power generated.
t_R	CPU time required to solve the PF problem by the traditional solution algorithm.
t_{PC}	CPU time required to solve the PF problem by the proposed approach.

Functions

$f(.)$	Objective function.
$\mathbf{g}(.)$	Vector representing the equality constraints.
$\mathbf{h}(.)$	Vector representing the inequality constraints.
Π	Continuous function describing the domain transformation mapping.

I. INTRODUCTION

Manuscript submitted May 2015; revised and resubmitted October 2015.

A. Vaccaro is with the Department of Engineering, University of Sannio, Benevento, Italy (email: vaccaro@unisannio.it).

C. A. Cañizares is with the Department of Electrical & Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (email: ccanizar@uwaterloo.ca).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

OPTIMAL power system operation requires intensive numerical analysis aimed at studying and improving system security and reliability. To this aim, the streams of data acquired by field sensors, like Phasor Measurement Units (PMUs), could be effectively processed in order to provide power system operators with the necessary information to better understand and reduce the impact of system uncertainties associated with load and generation variations, especially from solar and wind power sources. For large-scale networks, this process requires massive data processing and complex and NP-hard problem solutions, in computation times that should be fast enough for the information to be useful in a short-term operation horizon. For solving this challenging issue, the development of computing paradigms aimed at supporting rapid power systems analysis in a data rich, but information limited, environment, in the context of "big data" are necessary [1]–[3]. These paradigms should quickly convert the field data into actionable information by allowing the power system operator to have a full understanding of the available information [4].

The fundamental tools that could sensibly benefit from the conversion of data into information could be the Power Flow (PF) and Optimal Power Flow (OPF) analyses, since they are the most heavily used tools for solving many complex power system operation problems such as network reconfiguration, optimal power dispatch, voltage control, state estimation, etc. In fact, PF and OPF analyses have attracted a large amount of research efforts aimed at defining effective paradigms for reducing the complexities of the solution algorithms, which are typically based on expensive iterative numerical techniques. Thus, to deal with the intrinsic complexities of these analysis, alternative formalization of the problem equations, soft-computing based solution techniques, and distributed processing architectures have been proposed in the literature [5], [6]. Some of these approaches try to reduce the complexity and improve the convergence of the PF and OPF solution algorithms by defining more effective iterative schemes (e.g. modified Newton-Raphson and Trust-Region Interior-Point Methods [7], [8]), making some simplifying hypothesis on the PF equations (e.g. DC PF and DC OPF) or the Jacobian matrix (e.g. fast decoupled PF [9]), and/or applying advanced decomposition techniques (e.g. factorized load flow, Lagrangian relaxation, augmented Lagrangian decomposition [10], [11]). Moreover, there are several papers in the OPF literature that propose solution techniques based on convex relaxation, which formalize the optimization problem as a semi-definite program [12], convex quadratic program, or a second-order cone program [13], to solve the OPF using well-established and reliable conic solvers. In some conditions, these convex relaxation techniques allow to efficiently address several class of OPF problems, often providing useful information to compute the global optimal solution of the original non-convex problem [14], and offering several advantages compared to other traditional solution methods [15], [16]. In particular, it is shown in [15] that the computed OPF solution is globally optimal only when the convex relaxation is exact, while, in the general case, it is a lower bound of the minimum cost, which could be useful in evaluating the fitness of any feasible solution.

Other methods reported in the literature have proposed the employment of advanced frameworks based on computational intelligence techniques, which include evolutionary programming (e.g. Particle Swarm Optimization and Genetic Algorithms [17], [18]), Fuzzy logic programming [19], [20], and Neural Networks [21], [22]. A different approach to complexity burden reduction in PF and OPF analysis has been recently proposed in [23], [24], where the underlying principle is to employ domain decomposition techniques aimed at parallelizing the solution algorithm on pools of computers interconnected by commodity networks. Furthermore, the availability of massive datasets reinforced by the broad use of communication and information technologies in modern power systems is resulting in the application of "big data" and knowledge-based methods for OPF analysis, including distributed inferential paradigms [16], data analytic techniques [25], [26], and cooperative computing [27]. These novel algorithms are potentially suitable to deal with complex optimization problems in a big data setting, overcoming the intrinsic limitations of traditional solution methods, which cannot efficiently deal with large data sets.

The aforementioned techniques offer improvements to the solution of PF and OPF problems by utilizing different numerical approaches and tools to solve these problems according to traditional mathematical formulations. The current paper, on the other hand, proposes exploiting the potential actionable information that could be extracted from historical operation data-sets, which are expected to sensibly grow over time due to the pervasive deployment of sensors and measurement systems [28], to speed up the PF and OPF solution process as discussed in detail here. Hence, formal methods are proposed for knowledge discovery from a large quantity of data as an enabling methodology for reducing the complexity of the PF and OPF problems. A knowledge-based paradigm for PF and OPF analyses is used to extract from operation data-sets complex features, hidden relationships, and useful hypotheses potentially describing regularities in PF and OPF problems. This is realized by designing a knowledge-extraction process based on Principal Components Analysis (PCA) [29]. The structural knowledge extracted by this process is then used to transform the PF equations into a domain in which these equations can be solved more effectively, by reducing the size of the PF and OPF problems, and thus more efficiently obtain PF and OPF solutions. Numerical results obtained for realistic power networks, namely, the IEEE 118-bus test system, and the 2383-bus Polish system, for various operating conditions, show the benefits of the proposed approach.

The rest of the paper is organized as follows: Section II presents a brief review of the classical PF and OPF problems. In Section III, the theoretical foundations and the main features of the proposed framework are described in detail. Section IV describes the numerical studies and discusses the associated results. Finally, Section V summarizes the main conclusions and contributions of the paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

II. Problem Formulation

A. Power Flow Analysis [5]

PF analysis deals mainly with the calculation of the steady-state voltage phasor angle and magnitude for each network bus, for a given set of variables such as load demand and real power generation, under certain assumptions such as balanced system operation. Based on this information, the network operating conditions, in particular, real and reactive power flows on each branch, power losses, and generator reactive power outputs, can be determined. Thus, the input (output) variables of the PF problem are typically:

- the real and reactive power (voltage magnitude and angle) at each load bus, i.e. PQ buses;
- the real power generated and the voltage magnitude (reactive power generated and voltage angle) at each generation bus, i.e. PV buses;
- the voltage magnitude and angle (the real and reactive power generated) at the reference or slack bus.

The equations typically used to solve the PF problem are the following real power balance equations at the generation and load buses, and the reactive power balance at the load buses:

$$\begin{array}{ll} P_i^{SP} = V_i \sum_{j=1}^{N} V_j Y_{ij} \cos\left(\delta_i - \delta_j - \theta_{ij}\right) & \forall i \in \mathcal{N_P} \\ Q_j^{SP} = V_j \sum_{i=1}^{N} V_i Y_{ji} \sin\left(\delta_j - \delta_i - \theta_{ji}\right) & \forall j \in \mathcal{N_Q} \end{array} \tag{1}$$

where N is the total number of buses; $\mathcal{N_P}$ is the set of the buses in which the active power is specified; $\mathcal{N_Q}$ is the set of the buses in which the reactive power is specified; P_i^{SP} and Q_j^{SP} are the real and reactive power injections specified at i^{th} and j^{th} bus; $V_i \angle \delta_i$ is the unknown i^{th} bus voltage in polar coordinates; and $Y_{ij} \angle \theta_{ij}$ is the ij^{th} element of the bus admittance matrix. Due to the nonlinear nature of these equations, the solution is not unique, and numerical algorithms, mainly based on Newton-Raphson or fast-decoupled methods, are employed to obtain a solution that is within an acceptable tolerance. These algorithms aim at approximating the nonlinear PF equations by linearized Jacobian-matrix equations, which are solved by means of numerical iteration algorithms and sparse factorization techniques.

The PF solution should take into account the limits on certain variables, in particular max/min values of the reactive power at generation buses, to properly model the generator voltage controls. To address this particular issue, the typical solution strategy is to use a bus-type "switching", which consists on converting a PV-bus into a PQ-bus with the reactive power set at the limiting value, if the corresponding limits are violated. If at any consequent iteration, the voltage magnitude at that bus is below or above its original set point, depending on whether the generator is respectively underexcited or overexcited, the bus is then reverted back to a PV-bus. An alternative and more effective strategy to represent generator bus voltage controls, including reactive power limits and voltage recovery processes, has been proposed in [30], based on a novel OPF-based model of the PF problem with complementarity constraints to represent reactive power limits.

B. Optimal Power Flow Analysis [5]

Optimal Power Flows (OPFs) aim at computing the power system operation state based on, for example, cost, planning, or reliability criteria without violating system and equipment operating limits. The solution of this problem yields the optimal value of the control/decision variables $\mathbf{u}$ that minimizes an objective function f, subject to a number of nonlinear equality g_p and inequality constraints h_m, where all these functions are continuous and differentiable. Hence, this problem can be formalized in general using the following constrained, nonlinear multi-objective programming problem:

$$\begin{array}{lll} \min\limits_{(\mathbf{x},\mathbf{u})} & f(\mathbf{x},\mathbf{u}) & \\ \text{s.t.} & g_p(\mathbf{x},\mathbf{u}) = 0 & \forall p \in [1, n_p] \\ & h_m(\mathbf{x},\mathbf{u}) \leq 0 & \forall m \in [1, n_m] \end{array} \tag{2}$$

where $\mathbf{x}$ is the vector of dependent variables, n_p is the number of equality constraints, and n_m is the number of inequality constraints. These equations can be expressed in a more compact vectorial form as follows:

$$\begin{array}{ll} \min\limits_{(\mathbf{x},\mathbf{u})} & f(\mathbf{x},\mathbf{u}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x},\mathbf{u}) = 0 \\ & \mathbf{h}(\mathbf{x},\mathbf{u}) \leq 0 \end{array} \tag{3}$$

where $\mathbf{g}(.)$ and $\mathbf{h}(.)$ are the n_p-dimensional and n_m-dimensional vectors representing the equality and inequality constraints, respectively.

The control/decision variables in (3) depend on the the specific application domain. These can include both real-valued variables, such as the active power generated by the available generators (i.e. optimal power dispatch), the set points of the primary voltage controllers (i.e. secondary voltage regulation), the optimal location of control/generator resources (i.e. planning studies), the maximum loading factor (i.e. voltage stability analysis), and integer variables, such as the set of the available generators (i.e. unit commitment). As a consequence, the OPF can be in general classified as a non-convex mixed integer/non-linear programming (MINLP) problem.

The equality constraints in (2) and (3) correspond to the PF equations (1). The dependent variables include the voltage magnitude and phase angle at PQ buses, the voltage phase angle and the reactive power generated at the PV buses, and the active and reactive power generated at the slack bus. Finally, the inequality constraints include the maximum allowable power flows for the power lines, the minimum and maximum allowable limits for most control/decision variables, i.e. $u_{min,i_u} \leq u_{i_u} \leq u_{max,i_u}$, $\forall i_u \in [1, n_u]$, such as generator voltages, and for some dependent variables, i.e. $x_{min,i_x} \leq x_{i_x} \leq x_{max,i_x}$, $\forall i_x \in [1, n_x]$, such as bus voltage limits. Many classes of programming algorithms, such as nonlinear programming [31], quadratic programming [32], [33], and linear programming [34], have been proposed to solve the OPF problem.

The objective function $f(.)$ considers both technical and economic criteria, including the minimization of the production costs, transmission line losses, voltage deviations, etc. Due

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

to its non-convexity and non-linear characteristics, the OPF problem has multiple solutions.

III. Power Flow and Optimal Power Flow Problems in the Principal Components Domain

This paper advocates the use of knowledge-discovery paradigms to effectively solve the PF and OPF problems formalized in (1) and (2), respectively. The rationale is that, in practical applications, large datasets of historical operation data are available and could be processed in order to extract complex features, hidden relationships, and useful hypotheses describing potential regularities in the PF and OPF solutions. The idea is then to apply advanced information paradigms to extract knowledge from a database of historical power system data and measurements, and then use this structural knowledge to define a new domain in which the PF and OPF problems can be solved more efficiently by reducing the cardinality of the original problem. The proposed technique for knowledge extraction is based on the PCA, which extracts relevant information characterizing patterns in the PF and OPF solutions from historical data to transforms the problem equations.

The idea of applying PCA to PF analysis was originally explored in [35], which treats the PF problem as a Multiple Input Multiple Output system, where the control variables, e.g. the fixed active and reactive bus powers, are considered as input variables, while the state variables, e.g. voltage angles and magnitudes, are considered as output variables. In this context, PCA is mainly adopted to infer, from historical input/output data, the correlations between the control and the state variables, thus avoiding the need for repetitive power flow solutions to do sensitivity analyses. Conceptually, the framework proposed in [35] could be classified as a particular instance of the supervised machine-learning paradigm, since it aims at extracting semantic information from historical data by training a data-mining algorithm based on PCA. In particular, it could be considered as an improvement of the non-linear learning systems (e.g. fuzzy inference systems or neural networks) proposed in the literature for solving the PF problem. The solution approach proposed in the present paper is different, since it aims at processing historical information registered during power system operation in order to discover the correlations between the state variables. This is obtained by identifying a limited number of principal components, which could be considered as the basis of a new reference system, that allow representing all the state variables. This allows to reduce the cardinality of the PF problem, since only a reduced number of variables needs to be identified, namely the principal components, which satisfy the PF equations and the problem constraints. Once these variables have been identified, the corresponding state variables can be identified by a simple linear transformation. Thus, compared to [35] the benefits deriving by the application of the proposed approach are:

1) It doesn't identify a correlation between the control and the state variables of the power flow problem, which is a very complex task due to the non-linearity of the input/output mapping. Instead, it tries to discover the correlation between the state variables, which are highly correlated during normal power system operation, by defining a new and reduced set of "transformed" state variables.
2) The solution obtained by applying the proposed method, if it exists, is obtained very quickly, since the algorithm needs to identify a reduced number of state variables, and represents a rigorous PF solution, as it has been obtained by explicitly solving the PF equations, considering all the problem constraints. In other words, the proposed algorithm, rather than approximating the input/output mapping between the control and state variables of the PF problem, it solves the PF problem in a different reference system.
3) The proposed algorithm is applied to both PF and OPF problems.

The main modules of the proposed knowledge-based framework are explained next.

A. PCA-based Knowledge Extraction

PCA aims at discovering the potential relationships among a set of state variables $x_i \quad \forall i \in [1, N_x]$, from the following set of historical observations (usually referred to as the knowledge base):

$$\mathbf{x}(k) = [x_1(k)...x_{N_x}(k)]^T \quad \forall k \in [0, \overline{T}] \tag{4}$$

where $[0, \overline{T}]$ defines the integer sample time interval of available data.

This is accomplished by identifying a suitable domain transformation such that the elements of the knowledge base can be accurately represented by an inverse model of the form:

$$\mathbf{x}(k) = \Pi^{-1}(\zeta(k)) + \mathbf{r}(k) \quad \forall k \in [0, \overline{T}] \tag{5}$$

where $\Pi : \Re^{N_x} \to \Re^{N_\zeta}$ is a continuous function describing the domain transformation mapping; $\zeta(k) = [\zeta_1(k)...\zeta_{N_\zeta}(k)]^T$ are the components of the state vector $\mathbf{x}(k)$ in the transformed domain; and $\mathbf{r}(k)$ represents the residual error vector.

In standard PCA, the knowledge-base to be analyzed is represented by the $\overline{T} \times N_x$ matrix $\mathbf{X}$ defined as:

$$\mathbf{X} = \left[\mathbf{x}(0) - \frac{1}{\overline{T}}\sum_{k=0}^{\overline{T}}\mathbf{x}(k), ..., \mathbf{x}(\overline{T}) - \frac{1}{\overline{T}}\sum_{k=0}^{\overline{T}}\mathbf{x}(k)\right]^T \tag{6}$$

which has rank $L \leq \min(\overline{T}, Nx)$, and could be expressed by the following singular value decomposition [36]:

$$\mathbf{X} = \mathbf{A\Sigma\Omega}^T \tag{7}$$

where $\mathbf{A}$ is the $\overline{T} \times L$ matrix of left singular vectors, $\mathbf{\Omega}$ is the $N_x \times L$ matrix of right singular vectors, and $\mathbf{\Sigma}$ is the diagonal matrix of singular values. Note that $\mathbf{\Sigma}^2$ is a diagonal matrix of the eigenvalues of $\mathbf{XX}^T$ and $\mathbf{X}^T\mathbf{X}$ [36]. Starting from this decomposition, the $\overline{T} \times L$ matrix of the factor scores can be obtained as:

$$\mathbf{S} = \mathbf{A\Sigma} = \mathbf{A\Sigma\Omega}^T\mathbf{\Omega} = \mathbf{X\Omega} \tag{8}$$

where the matrix $\mathbf{\Omega}$ can be considered as a projection matrix, since the product $\mathbf{X\Omega}$ represents the projections of the observations on the principal components. The only information

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

available from the principal components computed by (8) is a measure of the relative importance of the observed variables. In this context, the $\overline{T}$ observations are considered to be the population of interest, and conclusions are limited to these specific observations.

An attractive feature of PCA is its ability to represent the observed data by using a fixed regressive model, which is identified by using only the first N_{PC} principal components of the knowledge-base $\mathbf{X}$. In particular, let $\hat{\mathbf{X}}_{N_{PC}}$ the estimation of the matrix $\mathbf{X}$ computed by using only N_{PC} principal components, then it follows that:

$$\hat{\mathbf{X}}_{N_{PC}} = \mathbf{A}_{N_{PC}} \mathbf{\Sigma}_{N_{PC}} \mathbf{\Omega}_{N_{PC}}^T = \mathbf{S}_{N_{PC}} \mathbf{\Omega}_{N_{PC}}^T \tag{9}$$

where $\mathbf{A}_{N_{PC}}$, $\mathbf{\Sigma}_{N_{PC}}$, and $\mathbf{\Omega}_{N_{PC}}$ correspond to the first N_{PC} columns of the corresponding matrices $\mathbf{A}$, $\mathbf{\Sigma}$, and $\mathbf{\Omega}$. A different way to represent (9), is to define the error matrix $\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}_{N_{PC}}$, and to express the matrix $\mathbf{X}$ as follows:

$$\mathbf{X} = \hat{\mathbf{X}}_{N_{PC}} + \mathbf{E} = \mathbf{S}_{N_{PC}} \mathbf{\Omega}_{N_{PC}}^T + \mathbf{E} \tag{10}$$

This expression allows to approximate the primitive variables by a linear combination of a proper number of orthogonal and uncorrelated principal components with decreasing variance, namely [29]:

$$\begin{aligned} \mathbf{x}(k) = \mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{x_{med}} \quad \forall k \in [0, \overline{T}] \\ \mathbf{x_{med}} = \tfrac{1}{\overline{T}} \textstyle\sum_{k=0}^{\overline{T}} \mathbf{x}(k) \end{aligned} \tag{11}$$

where $\mathbf{s}(k)$ is the principal component vector. This domain transformation mainly consists of translating and rotating the original coordinate axes, in such a way that the first principal component is characterized by the largest variance, and each following component by the highest variance that is orthogonal and uncorrelated with the previous components. As a consequence, each principal component carries different and uncorrelated information to other components, and only a limited number of them are necessary to accurately compute the state variables for highly correlated datasets ($N_{PC} \ll N_x$) [35]. Thanks to this feature, the $N_x\overline{T}$ historical data can be approximated by storing and processing a limited number of variables, namely, the principal components profiles, the static matrix $\mathbf{\Omega}_{N_{PC}}$, and the static vector $\mathbf{x}_{med}$, for a total of $N_{PC}\overline{T} + N_x N_{PC} + N_x$ elements. The ratio between these quantities provides a rough estimation of the data compression capability of the PCA-based knowledge extraction process, which, for a large number of observations, tends to the following value:

$$\begin{aligned} C_R^{\infty} &= \lim_{\overline{T}\to\infty} C_R(\overline{T}) = \lim_{\overline{T}\to\infty} \tfrac{N_x\overline{T}}{N_{PC}\overline{T} + N_x N_{PC} + N_x} = \\ &= \tfrac{N_x}{N_{PC}} \end{aligned} \tag{12}$$

This result demonstrates the effectiveness of PCA in compressing the knowledge base by extracting only the more relevant information, which mainly depends on the number of principal components N_{PC} assumed in the computation. The latter can be determined by adopting various statistical methods, including:

- Kaiser criterion: it selects principal components with eigenvalues greater than 1.
- Scree test: it is based on the analysis of the scree plot of the available data.
- Cumulative percentage method: it selects the components that cumulatively explain a certain percentage of variation.
- Binary search approach: it selects the components by identifying a proper trade-off between statistical fidelity, i.e. maximizing the variance in the data, and interpretability, i.e. minimizing the coordinate axes.

More details about these methods can be found in [37].

As recently discussed in several papers [38]–[40], PCA could be useful in data management for smart grids, where a massive increase of data exchanging and processing is expected in the short/medium term. Furthermore, as described next, the described PCA-based knowledge-extraction process, codified in the matrix $\mathbf{\Omega}_{N_{PC}}$, could also be used for PF and OPF analyses.

B. PCA-based Power Flow Analysis

The main idea of the PCA-based PF is to generalize the mathematical formulation defined in (11) by extrapolating the linear mapping between the power system state variables and the principal components as follows:

$$\mathbf{x}(k) = \mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{x}_{med} \quad \forall k > \overline{T} \tag{13}$$

This linear extrapolation allows to solve the PF problem for each $k > \overline{T}$, by identifying the unknown principal components $\mathbf{s}(k) = [s_1(K) ... s_{N_{PC}}(k)]^T$, such that:

$$\begin{aligned} P_i^{SP}(k) = P_i(\mathbf{x}(k)) = P_i(\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{x}_{med}) \quad \forall i \in \mathcal{N_P} \\ Q_j^{SP}(k) = Q_j(\mathbf{x}(k)) = Q_j(\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{x}_{med}) \quad \forall j \in \mathcal{N_Q} \end{aligned} \tag{14}$$

A noticeable benefit deriving from this mathematical formulation is the drastic reduction of the problem cardinality, since the number of design variables that should be identified at each time step is reduced from N_x to N_{PC}. This important feature should improve the convergence properties of the solution algorithm and lower its complexity and computational burden, based on the reduction of the asymptotic complexity of the solution algorithm, which is $O(N_x N_{PC}^2)$, due to the pseudo inverse of the Jacobian matrix of dimension $N_x \times N_{PC}$ of the PF equations in the principal component domain. However, the sparsity of the Jacobian is reduced with respect to the "standard" F Jacobian, since the latter roughly depends on the number of power system elements, whereas the former would have more intertwining variables. Nevertheless, the complexity reduction of the solution algorithm could be noticeable, given the significant Jacobian size reduction, especially in the presence of variable load/generation patterns, which may require multiple PF solutions (e.g. Monte Carlo simulations).

Observe that the integration of the proposed solution paradigm on existing power systems analysis toolboxes is straightforward, since the Jacobian of the PF equations in the principal components domain can be easily computed as:

$$\mathbf{J}_{PC} = \frac{\partial \mathbf{F}}{\partial \mathbf{s}} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{s}} = \mathbf{J}\ \mathbf{\Omega}_{N_{PC}} \tag{15}$$

where $\mathbf{F}$ is the set of PF equations and $\mathbf{J}$ is the corresponding Jacobian matrix. Observe as well that the reactive power generation limits in conventional PF programs can be readily integrated in the proposed framework by properly redimensioning the matrix $\mathbf{\Omega}_{N_{PC}}$ when a PV to PQ bus switch, or vice-versa, takes place.

C. PCA-based Optimal Power Flow Analysis

The benefits deriving by the formalization of the PF equations in the principal components domain, can be easily extended to OPF analysis. In this context, the main idea is to extrapolate a linear mapping between the variables $\mathbf{z} = [\mathbf{x}, \mathbf{u}]$ of the OPF problem (3) and the principal components $\mathbf{s}(k)$ as follows:

$$\mathbf{z}(k) = \mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{z}_{med} \quad \forall k > \overline{T} \tag{16}$$

This linear extrapolation allows to solve the OPF problem for each $k > \overline{T}$, by identifying the unknown principal components $\mathbf{s}(k)$ such that:

$$\begin{aligned} \min_{\mathbf{s}(k)} \quad & f(\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{z}_{med}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{z}_{med}) = 0 \\ & \mathbf{h}(\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{z}_{med}) \leq 0 \end{aligned} \tag{17}$$

Observe again the drastic reduction of the problem cardinality, since the number of variables has been reduced from N_z to N_{PC}. Moreover, the gradient of the cost and constraints functions in the principal components domain can be easily computed based on (15).

IV. Numerical Results

This section describes the results obtained by applying the proposed framework to solve PF and/or OPF problems for the 2383-bus Polish power system, and the IEEE 118 bus test system, respectively, for varying realistic operating conditions.

A. 2383-bus Polish Power System PF

In order to test the proposed technique, the power flow analysis of a large scale power system was studied. Thus, 735 bi-weekly 15 min load profile, whose ranges are shown Fig. 1, which were assumed for the 2383-bus Polish test system; these real demand patterns were obtained from the Australian Energy Market Operator database [41]. To obtain a realistic value of the generated active power, dispatch factors for each generator were computed for the base case, and then these factors were scaled up proportionally to the load demand, perturbing them with uniform noise signals ranging in the interval [-10%,+10%]. Thus, the corresponding generation profiles were defined for each time sample k as follows:

$$P_{G_{i_g}}(k) = \alpha_{G_{i_g}}\ r \sum_{j=1}^{N} P_{D_j}(k) \quad \forall i_g \in [1, n_G] \tag{18}$$

where r is a random noise uniformly distributed in the range 0.9-1.1, and $\alpha_{G_{i_g}}$ are dispatch factors for the base case, i.e.

$$\alpha_{G_{i_g}} = \frac{P_{G_{i_g}}(0)}{\sum_{j=1}^{N} P_{D_j}(0)} \quad \forall i_g \in [1, n_G] \tag{19}$$

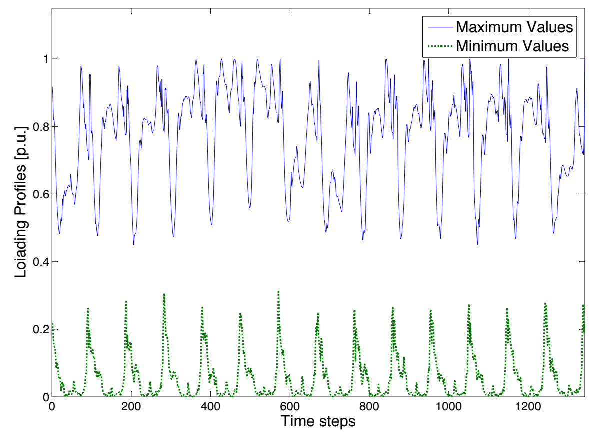

Fig. 1. Range of loading profiles used for the power flow analysis of the 2382-bus Polish test system.

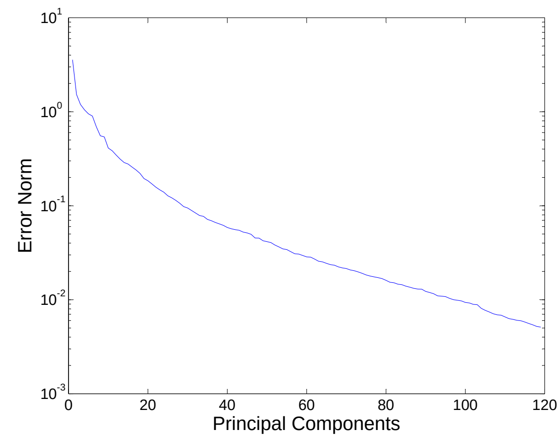

Fig. 2. Norm of the approximation error, in semi-logarithmic scale, versus the number of principal components for the 2382-bus Polish test system PF.

n_G is the number of generators; and $P_{D_j}(0)$ and $P_{G_{i_g}}(0)$ are the active power demands at the j^{th} bus and the active power generated by the i_g^{th} generator, respectively. The bus voltages phasors were then computed and the results were arranged in two sets, namely the knowledge base (first 500 sample points) and the Validation Set (remaining 843 sample points). The knowledge extraction process was then implemented for various numbers of principal components in the interval $N_{PC} = [1, 120]$, obtaining the results summarized in Fig. 2, where the approximation error is defined as:

$$\begin{aligned} \mathbf{e}_{app}^{(k)}(N_{PC}) &= \mathbf{x}(k) - \mathbf{x}_{PC}(N_{PC}, k) = \\ &= \mathbf{x}(k) - [\mathbf{\Omega}_{N_{PC}}\ \mathbf{s}(k) + \mathbf{x}_{med}(k)] \\ \Rightarrow \mathbf{e}_{app}(N_{PC}) &= [\mathbf{e}_{app}^{(1)}(N_{PC}), ..., \mathbf{e}_{app}^{(500)}(N_{PC})] \\ & \forall N_{PC} \in [1, 120],\ k \in [0, 500] \end{aligned} \tag{20}$$

Thus, for an approximation error tolerance of 0.05, 40 principal components can be extracted from the knowledge base, corresponding to a compression ratio of 10.98.

To assess the extrapolation capabilities of the proposed methodology, the PF problem was solved in the principal components domain $\forall k \in [500, 1343]$. The resulting solutions

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

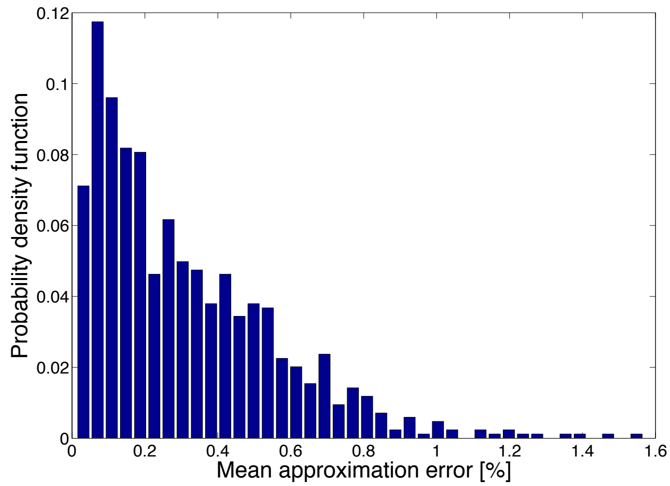

Fig. 3. Statistical characterization of the approximation accuracy for the 2382-bus Polish test system PF.

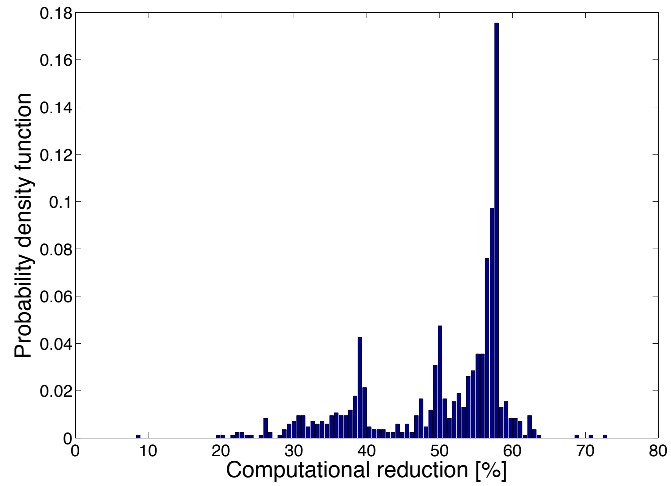

Fig. 4. Statistical characterization of the complexity reduction factor for the 2382-bus Polish test system PF.

were compared to those obtained by applying the built-in Matpower PF program. Figures 3 and 4 depict the error and complexity reduction indices defined as follows, respectively:

$$e(k) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{V}_{i,PCA}(K) - \mathbf{V}_{i,R}(k)\| \quad \forall k \in [500, 1343] \tag{21}$$

where $\mathbf{V}_{i,PCA}(k)$ and $\mathbf{V}_{i,R}(k)$ are the phasors of the i^{th} bus voltage computed with the proposed approach and the traditional solution algorithm, respectively; and:

$$C_r(k) = \frac{t_R(k) - t_{PC}(k)}{t_R(k)} * 100 \quad \forall k \in [500, 1343] \tag{22}$$

where $t_R(k)$ and $t_{PC}(k)$ are the CPU times required to solve the PF problem at the k^{th} time sample by the traditional solution algorithm and by the proposed approach, respectively.

Observe that Fig. 3 confirms the good accuracy of the solutions computed with the proposed paradigm, and Fig. 4 shows that the CPU times of the proposed approach, even for non-optimal software routines, are on average 58% faster with respect to those of the traditional solution algorithm. It should be mentioned that both algorithms were tested on the same computer and the same software (Matlab 2013b), and the reactive power generation limits were assumed to be the same in all cases. Furthermore, to solve the PF problem in the principal components domain, the standard pseudo-inverse operator available in the Matlab suite was used; a more effective pseudo-inverse algorithm may further reduce the computational burden of the proposed method [42].

It is important to note that the approximation accuracy and the CPU time of the proposed algorithm are strictly influenced by the number of principal components assumed for the inverse domain reconstruction (see Figs. 6 and 2). Hence, a proper selection criteria is needed to obtain a suitable tradeoff between solutions accuracy and algorithm complexity. The latter mainly depends on the statistical characteristics of the historical data set adopted for knowledge extraction.

B. IEEE 118-bus Test System OPF

The proposed technique was also applied to solve the optimal active power dispatch problem for the IEEE 118-bus test system, which represents a portion of the Midwestern American Electric Power System, composed by 118 bus, 54 generators, 64 loads, and 186 lines [43]. The control variables of the optimization problem in this case are the active power generated by the 54 dispatchable generators, while the dependent variables are the voltage magnitude at the load buses, and the voltage phase angle at all buses except the slack bus. The objective is to minimize the total generation cost satisfying both the equality constraints, i.e. the power flow equations, and the inequality constraints, corresponding to the limits in the voltage magnitudes at all buses and the reactive power limits at the generator buses.

The load buses were clustered in 10 different classes characterized by 10 bi-weekly 15 min load profile, whose ranges are depicted in Fig. 5, based on similar load profiles proposed in [44]. The OPF problem was then solved and the corresponding results were arranged in two sets, namely, the knowledge base (first 500 sample points) and the validation set (remaining 843 sample points). The knowledge extraction process was then implemented for various numbers of principal components in the interval $N_{PC} \in [5, 50]$. The obtained results are summarized in Fig. 6 using a semi-logarithmic scale to plot the norm of the approximation error e_{app} versus N_{PC}, obtained from (20). This figure shows that when the principal components number N_{PC} increases, the approximation error drastically decreases approaching a saturation threshold. This allows to identify the adequate number of principal components needed to properly solve the power flow problem for a given approximation error tolerance. Thus, for an approximation error tolerance of 10^{-3}, 24 principal components can be extracted from the knowledge base.

To assess the extrapolation capabilities of the proposed methodology, the OPF problem was solved in the principal components domain $\forall k \in [500, 1343]$. The resulting solutions

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2549560, IEEE Transactions on Smart Grid

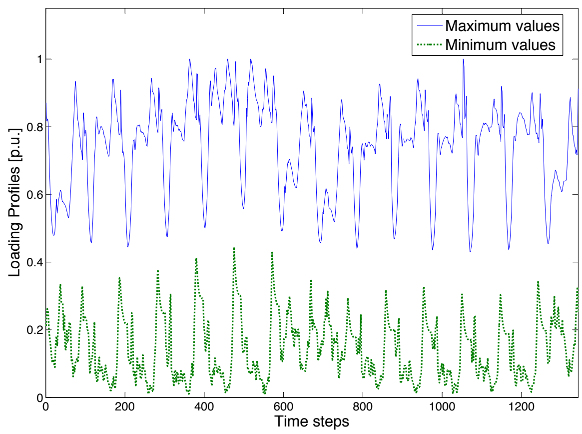

Fig. 5. Range of loading profiles used for the OPF analysis of the 118-bus IEEE test system.

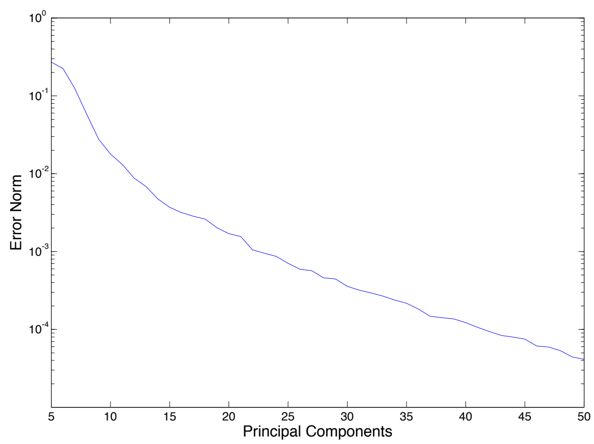

Fig. 6. Norm of the approximation error, in semi-logarithmic scale, versus the number of principal components for the OPF analysis of the 118-bus IEEE test system.

were then compared to those obtained by applying an open-source optimal power flow program (Matpower 4.11 [45]). The corresponding error surfaces are reported in Figs. 7 and 8, where it is possible to observe the high accuracy of the solutions computed by the proposed solution framework. This is also confirmed by analyzing Figs. 9 and 10, which depict the error and complexity reduction indexes obtained from (21) and (22), respectively. Observe that Fig. 9 confirms the good accuracy of the solutions computed with the proposed paradigm, and Fig. 10 shows that the CPU times of the proposed approach, even for non-optimal software routines, are on average 77% faster with respect to those of the traditional solution algorithm.

C. Discussion

Based on the aforementioned results it could be argued that, as expected, the projection of the state variables into the principal component domain is approximate, with the approximation error decreasing when the number of principal components increases. For this reason, the adoption of PCA for data set compression is typically classified as a compression

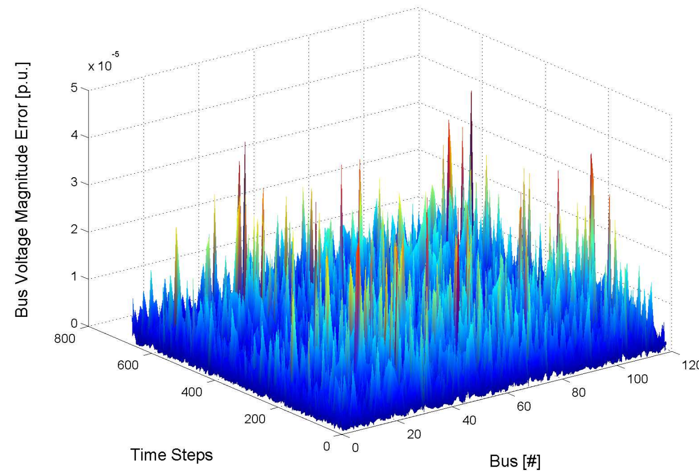

Fig. 7. Approximation errors of the proposed technique versus the true OPF solution for the IEEE 118-bus test system: bus voltage magnitude error.

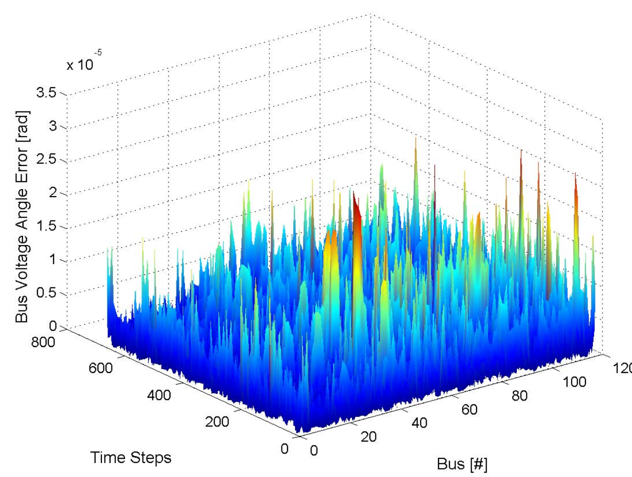

Fig. 8. Approximation errors of the proposed technique versus the true OPF solution for the IEEE 118-bus test system: bus voltage angle error.

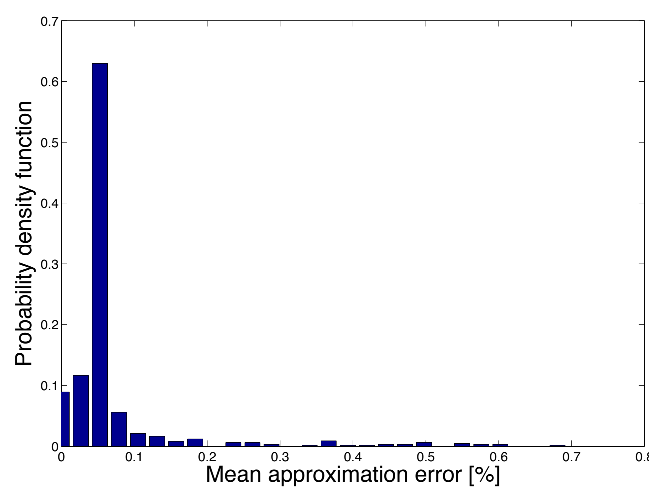

Fig. 9. Statistical characterization of the approximation accuracy for the OPF analysis of the IEEE 118-bus test system.

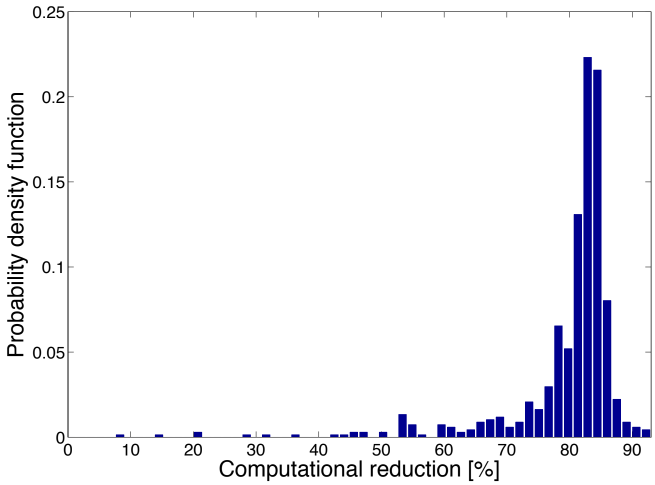

Fig. 10. Statistical characterization of the complexity reduction factor for the OPF analysis of the IEEE 118-bus test system.

technique with information loss, since the process of building the original data set by combining the principal components is approximate. Nevertheless, this approximation error is forced to be low, since the number of principal components that yield a reasonably low error is chosen a priori (see Figs. 2 and 6).

A further, and certainly more relevant, approximation error derives from the projection of new observations onto the principal components. During this process, the matrix $\mathbf{\Omega}$ is used to compute the principal components for observations that were not included in the original data set, which are typically called supplementary or illustrative observations, in order to emphasize the difference with the active observations. This kind of extrapolation process is also approximate, with the corresponding approximation error not being rigorously characterized a priory, but rather characterized by analyzing the results obtained for a particular case study. In the experimental studies discussed in Section IV, it has been shown that this approximation error is low by comparing the actual solutions with the PCA-based solutions for many system conditions.

From the the reported results, it could be argued that the algorithm works very well for the considered load profiles, since it allows discovering the correlations between the state variables on the basis of historical information. In this context, it is important to outline that PCA is useful in analyzing correlated data, since the level of correlation characterizing the input data sensibly affects the performance of the proposed algorithm, in terms of solution accuracy and convergence. Based on the authors experience, one can expect that the level of correlation characterizing the state variables during normal power system operation would be very high and represented by ellipsoids, which is consistent with the conclusions in [46]. This feature makes the application of PCA to PF and OPF analysis promising, since this technique is particularly suitable to represent multidimensional random vectors whose uncertainty region is represented by an ellipsoid. Hence, the PCA domain transformation allows discovering the internal structure of the data set in a way that best explains the variance and the correlation of the data using only a limited number of principal components, so that the dimensionality of the transformed problem is reduced.

It is important to note that if the approximation error of the extrapolation process is significant, the projection process does not accurately model the state-variable correlation and thus the algorithm fails to converge to a feasible solution; in this case, an exact solution algorithm should be adopted to find the required PF/OPF solution. This could happen due to the lack of an exhaustive knowledge base that considers critical power system operating states (e.g. close to the maximum system loading condition). These conditions have not been analyzed in the presented simulations, since the realistic data sets used did not push the test systems into critical operating states.

Finally, it is worth observing that the adoption of PCA to model integer variables is not trivial, and requires the use of more sophisticated mathematical methodologies. The authors are currently working in addressing this complex issue in the context of Unit Commitment problems, which is outside of the scope of the present paper.

V. Conclusion

In this paper, a novel framework aimed at identifying potentially regularities of the PF and OPF solutions has been proposed. Based on structural knowledge, and assuming the availability of large data sets from widespread sensors and measurement units, a mathematical method projecting the PF and OPF equations to a new domain has been defined, thus reducing the complexity and computational burden of the PF and OPF problem. The proposed approach can be considered as part of a "big data" processing engine for power systems analysis, in which actionable intelligence is effectively extracted from large volume of historical operation data, and delivered at the right time, and to the right power system operator.

The numerical results obtained for realistic power systems under various operating scenarios demonstrate that the overall complexity of the PF and OPF problems in the proposed transformed domain could be sensibly reduced, especially in the presence of correlated variables. In particular, it was observed that the approximation accuracy and the computational burdens observed during the experiments were strictly influenced by the number of principal components selected to decompose the power system state variables. Therefore, formal methods aimed at defining a proper tradeoff between the solutions accuracy and the algorithm complexity would be necessary for a comprehensive deployment of the proposed approach. This topic is currently under investigation.

References

[1] V. Madani and R. L. King, "Strategies and applications to meet grid challenges and enhance power system performance," in *Proceedings of the IEEE Power Engineering Society General Meeting*, pp. 1–9, 2007.

[2] J. Baek, Q. Vu, J. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE Transactions on Cloud Computing*, vol. to appear, no. 99, pp. 1–12, 2015.

[3] M. Kezunovic, L. Xie, and S. Grijalva, "The role of big data in improving power system operation and protection," in *Proceedings of the 2013 IREP Symposium on Bulk Power System Dynamics and Control*, pp. 1–9, 2013.

[4] R. King, "Information services for smart grids," in *Proceedings of the IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–5, 2008.

[5] A. Gomez-Exposito, A. J. Conejo, and C. A. Cañizares, *Electric Energy Systems: Analysis and Operation*. CRC Press, 2009.

[6] P. J. Lagace, "Power flow methods for improving convergence," in *Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society*, pp. 1387–1392, 2012.

[7] A. Sousa, G. L. Torres, and C. A. Cañizares, "Robust optimal power flow solution using trust region and interior-point methods," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 487–499, 2011.

[8] A. A. Sousa and G. L. Torres, "Globally convergent optimal power flow by trust-region interior-point methods," in *Proceedings of the IEEE Lausanne Power Tech*, pp. 1386–1391, 2007.

[9] A. Monticelli, A. Garcia, and O. Saavedra, "Fast decoupled load flow: Hypothesis, derivations, and testing," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1425–1431, 1990.

[10] A. Gomez-Exposito and C. Gmez-Quiles, "Factorized load flow," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4607–4614, 2013.

[11] F. Nogales, F. Prieto, and A. Conejo, "A decomposition methodology applied to the multi-area optimal power flow problem," *Annals of Operations Research*, vol. 20, no. 4, pp. 99–116, 2003.

[12] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 6-7, pp. 383–392, 2008.

[13] R. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458–1459, 2006.

[14] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2012.

[15] S. H. Low, "Convex relaxation of optimal power flow -part i: Formulations and relaxations," *IEEE Transactions on Control Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[16] L. Liu and H. Z., "Multi-block admm for big data optimization in smart grid," in *Proceedings of the 2015 International Conference on Computing, Networking and Communications (ICNC)*, pp. 556–561, 2015.

[17] K. Wong, J. Yuryevich, and A. Li, "Evolutionary-programming-based load flow algorithm for systems containing unified power flow controllers," *IET Generation, Transmission & Distribution,*, vol. 150, no. 4, pp. 441–446, 2003.

[18] L. Lai, J. Ma, R. Yokoyama, and M. Zhao, "Improved genetic algorithms for optimal power flow under both normal and contingent operation states," *International Journal of Electrical Power & Energy Systems,*, vol. 19, no. 5, pp. 287–292, 1997.

[19] P. A. Pajan and V. L. Paucar, "Fuzzy power flow: considerations and application to the planning and operation of a real power system," in *Proceedings of the International Conference on Power System Technology*, pp. 433–437, 2002.

[20] S. Kumar and D. Chaturvedi, "Optimal power flow solution using fuzzy evolutionary and swarm optimization," *International Journal of Electrical Power & Energy Systems,*, vol. 47, no. 1, pp. 416–423, 2013.

[21] H. H. Muller, M. J. Rider, C. A. Castro, and V. L. Paucar, "Power flow model based on artificial neural networks," in *Proceedings of the IEEE Power Tech*, pp. 1–6, 2005.

[22] V. Gutierrez-Martinez, C. Cañizares, C. Fuerte-Esquivel, A. Pizano-Martinez, and G. Xueping, "Neural-network security-boundary constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 63–72, 2011.

[23] Q. Morante, N. Ranaldo, A. Vaccaro, and E. Zimeo, "Pervasive grid for large-scale power systems contingency analysis," *IEEE Transactions on Industrial Informatics*, vol. 2, no. 3, pp. 165–175, 2006.

[24] S.-Y. Lin and J.-F. Chen, "Distributed optimal power flow for smart grid transmission system with renewable energy sources," *Energy*, vol. 56, no. 1, pp. 184–192, 2013.

[25] C. Stimmel, *Big Data Analytics Strategies for the Smart Grid*. CRC Press, 2015.

[26] P. Devie and S. Kalyani, "An optimization framework for cloud-based data management model in smart grid," *International of Journal Research in Engineering and Technology*, vol. 4, no. 2, pp. 751–758, 2015.

[27] E. DallAnese, H. Zhu, and G. B. Giannakis, "An optimization framework for cloud-based data management model in smart grid," *IEEE Transactions on Smart Grids*, vol. 4, no. 3, pp. 1464–1475, 2013.

[28] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, S. Fitch, J. Skok, M. M. Begovic, and A. Phadke, "Wide-area monitoring, protection and control of future electric power networks," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 80–93, 2011.

[29] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *Proceedings of the International Conference on System Science, Engineering Design and Manufacturing Informatization*, vol. 1, pp. 27–30, 2010.

[30] M. Pirnia, C. A. Cañizares, and K. Bhattacharya, "Revisiting the power flow problem based on a mixed complementarity formulation approach," *IET Generation, Transmission & Distribution*, vol. 7, no. 11, pp. 1194–1201, 2013.

[31] R. R. Shoults and D. Sun, "Optimal power flow based upon pq decomposition," *IEEE Transactions on Power Apparatus and Systems*, no. 2, pp. 397–405, 1982.

[32] J. A. Momoh, "A generalized quadratic-based model for optimal power flow," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 261–271, 1989.

[33] R. Burchett, H. Happ, and K. Wirgau, "Large scale optimal power flow," *IEEE Transactions on Power Apparatus and Systems*, no. 10, pp. 3722–3732, 1982.

[34] K. Pandya and S. Joshi, "A survey of optimal power flow methods," *Journal of Theoretical & Applied Information Technology*, vol. 4, no. 5, 2008.

[35] R. Bo and F. Li, "Power flow studies using principal component analysis," in *Proceedings of the 40th North American Power Symposium*, pp. 1–6, 2008.

[36] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 433–459, 2010.

[37] P. R. Peres-Neto, D. A. Jackson, and K. M. Somers, "How many principal components? stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics & Data Analysis*, vol. 49, no. 1, pp. 974–997, 2005.

[38] D. J. Burke and M. J. O'Malley, "A study of principal component analysis applied to spatially distributed wind power," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2084–2092, 2011.

[39] J. Yin, I. Gorton, and S. Poorva, "Toward real time data analysis for smart grids," in *Proceedings of the SC Companion: High Performance Computing, Networking, Storage and Analysis*, pp. 827–832, 2012.

[40] S. Das and P. N. Rao, "Principal component analysis based compression scheme for power system steady state operational data," in *Proceedings of the IEEE PES Innovative Smart Grid Technologies-India*, pp. 95–100, 2011.

[41] A. E. M. Operator, "Energy market data," in *available on line at http://www.aemo.com.au/Electricity/Data*.

[42] P. Horata, S. Chiewchanwattana, and K. Sunat, "A comparative study of pseudo-inverse computing for the extreme learning machine classifier," in *Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, pp. 40–45, 2011.

[43] R. D. Christie, "Power systems test case archive," in *available on line at http://www.ee.washington.edu/research/pstca*.

[44] A. Mutanen, M. Ruska, S. Repo, and P. Järventausta, "Customer classification and load profiling method for distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 1, pp. 1755–1763, 2011.

[45] R. Zimmerman, C. , Murillo-Sanchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[46] A. Saric and A. Stankovic, "Ellipsoidal approximation to uncertainty propagation in boundary power flow," in *Proceedings of the IEEE PES Power Systems Conference and Exposition*, pp. 1722–1727, 2006.

Alfredo Vaccaro (M'01, SM'09) received the M.Sc. degree with honors in electronic engineering from the University of Salerno, Salerno, Italy, and the Ph.D degree in electrical and computer engineering from the University of Waterloo, Ontario, Canada. From 1999 to 2002, he was an Assistant Researcher at the University of Salerno, Department of Electrical and Electronic Engineering. From March 2002 to October 2015, he was an Assistant Professor in electric power systems at the Department of Engineering of the

University of Sannio, Benevento, Italy, where he is currently an Associate Professor of electrical Power System. His special fields of interest include soft computing and interval-based method applied to power system analysis, and advanced control architectures for diagnostic and protection of distribution networks. Prof. Vaccaro is a member of the Editorial Boards of IET Renewable Power Generation, and the International Journal of Reliability and Safety, and he is the Executive Editor of the International Journal of Renewable Energy Technology.

Claudio A. Cañizares (S'85, M'91, SM'00, F'07) received the Electrical Engineer Diploma from the Escuela Politecnica Nacional (EPN), Quito, Ecuador, in April 1984, the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, in 1988 and 1991, respectively. He had held various academic and administrative positions at the E&CE Department of the University of Waterloo, Waterloo, ON, Canada, since 1993, where he is currently a full Professor, the Hydro One Endowed Chair, and an Associate Director of the Waterloo Institute for Sustainable Energy (WISE). His research interests are in modeling, simulation, control, stability, computational and dispatch issues in sustainable power and energy systems in the context of competitive markets and smart grids.