British Journal of Applied Science & Technology
11(1): 1-26, 2015, Article no.BJAST.19043
ISSN: 2231-0843

SCIENCEDOMAIN *international*
www.sciencedomain.org

Clusterization for Distributed Timely Detection of Changes in Smart Grids

Jose Jimenez David Lara[1], P. Papantoni-Kazakos[2*], A. T. Burrell[3] and Fernando Mancilla-David[2]

[1]CINVESTAV IPN, Guadalajara, Mexico.
[2]Department of Electrical Engineering, University of Colorado Denver, Denver, Colorado 80217, USA.
[3]Department of Computer Science, Oklahoma State University, Stillwater, Oklahoma 74078, USA.

Authors' contributions

This work was carried out in collaboration with all authors. Author PPK designed the study and carried the mathematical analysis, jointly with author ATB. Author ATB also completed the type-setting of the equations. Author JJDL performed the literature search in power systems and constructed the power system example used in the numerical evaluations. Author FMD oversaw the power system relevance. All authors checked the final results as well as read and approved the final manuscript.

Article Information

DOI: 10.9734/BJAST/2015/19043
Editor(s):
(1) Rodolfo Dufo Lopez, Electrical Engineering Department, University of Zaragoza, Spain.
Reviewers:
(1) Dominik Strzałka, Rzeszów University of Technology, Poland.
(2) Anonymous, Mohamed Premier University, Morocco.
(3) Anonymous, Annamalai University, India.
(4) Anonymous, St. John's University, USA.
Complete Peer review History: http://sciencedomain.org/review-history/10305

Original Research Article

Received 23rd May 2015
Accepted 8th July 2015
Published 24th July 2015

ABSTRACT

For real-time management of power distribution systems, when rapid operational adjustments are required to cope with intermittent power generation which is typical of renewable-based units, it is imperative that the optimization of the overall power distribution system be addressed in a distributed fashion. Then, the power distribution system may be partitioned into clusters whose size is determined by the delay constraints induced by the regular operations and the required operational adjustments.
In this paper, clusterization is considered as directly addressing the operational adjustment problem in the presence of operational changes. Then, such changes need to be identified timely and accurately before pertinent adjustments be performed. Clusterization may thus be dictated by the

**Corresponding author: E-mail: titsa.papantoni@ucdenver.edu;*

accuracy and delay constraints imposed on the detection and identification of such changes. In particular, we first consider the initially non-clusterized power distribution network and determine the bus voltage and/or current variations perceived as considerable changes. Then, we formulate a recursive maximum likelihood (ML) approach which naturally points to an initial network clusterization via incorporated sufficient identifiability conditions. We subsequently develop, analyze and evaluate a distributed sequential detection of change algorithm, implemented by the supporting data communication network, whose performance (including accuracy and decision delay) is controlled by a set of threshold parameters. Required algorithmic performance constraints may dictate final cluster architecture and dimensionality. This performance monitoring and clusterization approach has never been considered in power systems before.

Keywords: Smart grids; clusterization; distributed detection of changes; maximum likelihood; stochastic approximation.

1. INTRODUCTION

The Smart Grid concept evolved as a response to the ever increasing demand for highly reliable electric power, the increasing penetration of renewable sources and the need of a robust and resilient grid design for surmounting challenges of an aging infrastructure. Modernizing the grid encompasses both the distribution and the transmission levels, where due to its less advanced state, the distribution level is currently receiving most of the design attention. At the distribution level, main objectives to be attained include improved efficiency, reliability and power quality, high penetration of renewables, active load control, self–healing, and vulnerability monitoring. These objectives could in principle be achieved through the fast control of hundreds of individual distributed generators (DGs) and other devices linked to the grid through a power electronic interface. However, this would require real time information on each DG unit and key loads, leading to a daunting control problem. The control complexity and reliability of such a system may be greatly reduced if the distribution system is broken down into smaller partitions, named *clusters*, with each cluster containing a data communications supporting infrastructure and resembling the concept of *microgrids* [1].

The challenge addressed in this paper is the formalization of an optimization criterion for the appropriate formation of clusters at the distribution level. Our formalization is based on the following idea: *clusterization* evolves from the necessity for fast response in the presence of significant power flow (PF) or power loss (PL) changes, reflected by measurable bus voltage and/or current changes. Thus, the pertinent optimization criterion adopted is accurate and timely detection of such changes. The adopted criterion leads to the development of a distributed stochastic sequential algorithm, implemented by the supporting data communication network, whose distributed characteristics are initially defined by sufficient identifiability conditions and whose accuracy and speed performances are controlled by a set of threshold parameters. The algorithm is additionally asymptotically optimal in a mathematically precise sense.

In distribution networks, real-time monitoring is a critical issue, since the number of supervisory devices installed may be limited. To provide reasonable and meaningful estimates of the distribution network state from a very limited set of real measurements, State Estimation (SE) methods have been widely studied in the literature [2-12]. The network state is typically given by the voltages at all nodes [13], estimated from a few real measurements and complimented with pseudo-measurements, to ensure system observability; the pseudo-measurements need to be accurately modeled to improve the estimation quality [2]. In [3], the network is first split into smaller sections, before an SE process is deployed to estimates voltage magnitudes via artificial neural networks. In [4], SE based on a statistical technique is utilized to estimate the level, location and impact of voltage unbalance. In [5] and [6], the status of the distribution network is estimated using a distributed measurement system in a multi-area framework. The work in [7] utilizes synchrophasor measurements to study dynamic SE techniques in distribution grids. The SE approach in [8] is proposed for identification of network topology changes. With the integration of distributed generation, the evolution of the distribution grids may involve enhanced complex

behavior. As mentioned in [9], in the evaluation of network status, traditional SE methods may be insufficient and inaccurate. Hence, while robust SE techniques are required to support the secure operation of the network, such approaches may require in some cases [10,11] a significant number of measurements yielding to complex computational processing. In addition, inaccuracies of the deployed SE methodologies may affect confidence in their results [12].

In the present work we depart from SE methods in monitoring. Instead, we propose a sequential distributed algorithm which monitors failure probabilities of distribution lines (including the equipment they incorporate), to detect effectively alarming dysfunctionalities, while simultaneously imposing network clusterization. The organization of the paper is as follows. In Section 2, the system model and problem formulation are presented, while, in addition, a maximum likelihood (ML) estimation optimization criterion is formulated and identifiability analysis is performed. In Section 3, a stochastic approximation approach to the maximum likelihood solution is presented and analyzed. In Section 4, the distributed sequential detection of change algorithm is presented and analyzed. In Section 5, discussions on the structure of numerical evaluations and an example are included. In Section 6, conclusions are drawn.

2. SYSTEM MODEL, PROBLEM FORMULATION AND ML IDENTIFIABILITY

We view a power distribution network as a graph comprised of nodes (DGs and power consuming units) and lines. For each of the lines, we determine metrics of good condition or *acceptable functionality* which may be represented by acceptable ranges of bus voltages and currents. It is important to note here that the status of DGs and power consuming units may be incorporated into the acceptable line functionality concept; alternatively, dysfunctionality of lines may be caused by failing equipment at their ends. Thus, when a line is declared faulty, functionality investigation of all its components (including the equipment at its ends) should be next initiated. If a line meets the acceptable functionality conditions, it is given a success score 0; otherwise, it is given a failure score 1. Similarly, we determine acceptable functionality for power demands raised at some network node k and addressing another network node *l,* giving the same 0, 1 scores. The (*kl*) acceptable functionality is determined by metrics same as those for line functionality, only that the ordered pair (*kl*) is generally a route involving several lines, where a (*kl*) unacceptable functionality may be caused by the failure of any of the lines involved in the route. We will assume that the metrics of acceptable functionality are well defined. For given topology of the power distribution network, let us denote:

(*kl*): Ordered power source-to- demand pair.

i: Line index, where $1 \le i \le M$ and M is the total number of lines in the network.

r_{kl}: The relative load for (k*l*); equivalently, the probability that a random demand generated in the network is a (k*l*) demand, where $\sum_{(kl)} r_{kl} = 1$.

$q_{i\ ,\ (kl)}$: The probability that a (k*l*) demand uses line i. Equivalently, the fraction of a (k*l*) demand that is carried by line i. This represents a routing probability.

v_i: The probability that a demand going through line i fails.

p_i: The probability that a demand made somewhere in the network fails due to line i; $1 \le i \le M$.

p_0: The probability that a random demand generated somewhere in the network does not fail.

$f_{(kl)}(x)$: The probability that a random demand is generated in the network, it is a (k*l*) demand and the outcome observed is x, where,

$$x = \begin{cases} 1, \text{ if demand fails} \\ 0, \text{ if demand succeeds} \end{cases}$$

In our model, changes are reflected by the probabilities $\{v_i\}$ and $\{p_i\}$ of line failures, where failure actually means non conforming with a priori defined acceptable operational conditions; acceptable operational conditions translate then to these probabilities maintaining values below predetermined upper bounds. Our problem formalization thus consists of monitoring these probabilities and declaring changes when their values are estimated to be above the latter upper bounds. For given network topology, we assume that relative loads and routing probabilities are design quantities and remain unchanged during the monitoring process (before increase of their values above the upper bounds is

estimated). We use (k*l*) outcomes as observations. We first formulate a maximum likelihood (ML) estimation approach [14] for the probabilities $\{p_i\}$ which encompasses identifiability conditions. We then transform the result of the approach to a sequential stochastic approximation [15] format, which we finally use to develop a sequential detection of change algorithm for the set $\{p_i\}$. Similar approach was first taken in [16] for the distributed monitoring of the telephone network, where in this paper, the delay characteristics induced by the distributed sequential detection of change algorithm are used as design guidelines for the clusterization of the power distribution network.

To avoid formulating a cumbersome problem, we make the following simplified assumption: For each demand failure, there is a major contributor: a single line which is its initial or main cause. This assumption is consistent with our network maintenance objective: we wish to monitor possible deteriorating network conditions, characterized as "soft faults", rather than recognize "obvious" catastrophic events. Under this assumption we have:

$$\sum_{0 \le i \le M} p_i = 1 \tag{1}$$

Without much effort, we can also derive the expression below, which expresses the probability $f_{(kl)}(x)$ as a function of load, routing and failure probabilities.

$$f_{(kl)}(x) = r_{kl}\left[\sum_{1\le i\le M} v_i\, q_{i,(kl)}\right]^{x} \quad \left[1-\sum_{1\le i\le M} v_i\, q_{i,(kl)}\right]^{1-x} \tag{2}$$

where,

$$v_i\, q_{i,(kl)} = p_i\, q_{i,(kl)}\left[\sum_{mp} r_{mp} q_{i,(mp)}\right]^{-1} \tag{3}$$

The overall failure probability for pairs (*kl*) is obtained by summing up the above expression over all the network lines. Finally, we consider a sequence of pair observations, where $x_{j,\,kl}$ denotes the j-th outcome of a (*kl*) pair demand and N_{kl} is the number of (*kl*) pair demands made. Then, assuming that demands are independent from each other and summing up over all (k*l*) pair demands, we form the following ML function for the probabilities $\{p_i\}$:

$$f(\{p_i\}) = \sum_{kl}\sum_{j=1}^{N_{kl}}\left[x_{j,kl}\log\left(\sum_{i=1}^{M} p_i \frac{q_{i,(k,l)}}{\sum_{mp} r_{mp} q_{i,(mp)}}\right)+\right.$$

$$\left.+(1-x_{j,kl})\log\left(1-\sum_{i=1}^{M} p_i \frac{q_{i,(k,l)}}{\sum_{mp} r_{mp} q_{i,(mp)}}\right)\right]+ \quad \sum_{kl}\sum_{j=1}^{N_{kl}}\log r_{kl} \tag{4}$$

We note that the independence assumption made in the formulation of the ML function in (4) represents a worst case scenario: in the presence of dependent demands, the probability of error induced by the then optimal ML function is bounded from above by that induced by the ML function in (4), [14].

If the ML function in (4) is strictly concave, the probabilities $\{p_i\}$ may be estimated as the unique values which set the gradient of $f(\{p_i\})$ equal to zero. That is, the ML estimates of the probabilities $\{p_i\}$ are then given by the unique solution of the following set of equations:

$$\frac{\partial f(\{p_i\})}{\partial p_i} = \sum_{kl}\sum_{j=1}^{N_{kl}} \left[x_{j,kl} \cdot \frac{\dfrac{q_{i,(k,l)}}{\sum_{mp} r_{mp} q_{i,(mp)}}}{\sum_{s=1}^{M} p_s \dfrac{q_{s,(k,l)}}{\sum_{mp} r_{mp} q_{s,(mp)}}} - \right.$$

$$\left. -(1-x_{j,kl}) \cdot \frac{\dfrac{q_{i,(k,l)}}{\sum_{mp} r_{mp} q_{i,(mp)}}}{1-\sum_{s=1}^{M} p_s \dfrac{q_{s(k,l)}}{\sum_{mp} r_{mp} q_{s,(mp)}}} \right] = 0 \qquad ; \text{i = 1, .. , M} \tag{5}$$

The existence of unique solutions of the system in (5) coins the term *identifiable* for the probabilities $\{p_i\}$, where identifiability is guaranteed if and only if the matrix $M(\{p_i\})$ defined in (6) below is strictly negative definite. By definition, $M(\{p_i\})$ is strictly negative definite if and only if , given any nonzero column vector $A = \{a_i ; i = 1,\ldots,M\}$, the quadratic scalar $A^T\ M(\{p_i\})\ A$ is strictly negative.

$$M(\{p_i\}) = \left[\frac{\partial^2 f(\{p_i\})}{\partial p_i \partial p_n}; \quad i,n = 1,...,M \right] \tag{6}$$

where,

$$\frac{\partial^2 f(\{p_i\})}{\partial p_i \partial p_n} = \sum_{kl}\sum_{j=1}^{N_{kl}} \frac{q_{i,(k,l)}}{\sum_{mp} r_{mp}\, q_{s,(mp)}} \cdot \frac{q_{n,(k,l)}}{\sum_{mp} r_{mp}\, q_{n,(mp)}} \cdot$$

$$\cdot \left[x_{j,kl} \left(\sum_{s=1}^{M} p_s \frac{q_{s,(k,l)}}{\sum_{mp} r_{mp}\, q_{s,(mp)}} \right)^{-2} + (1-x_{j,kl}) \left(1-\sum_{s=1}^{M} p_s \frac{q_{s,(k,l)}}{\sum_{mp} r_{mp}\, q_{s,(mp)}} \right)^{-2} \right] \tag{7}$$

while,

$$\overline{G}^T(x,(kl)_n,\overline{p}) = \left\{ h(a_s,(kl)_n) \left[x_{(kl)_n} \left(\sum_{s=1}^{M_a} p_{a_s} \cdot \right. \right. \right.$$

$$\cdot \sum_{j=1}^{N_{kl}} \left[x_{j,kl} \left(\sum_{s=1}^{M} p_s \frac{q_{s,(k,l)}}{\sum_{mp} r_{mp}\, q_{s,(mp)}} \right)^{-2} + (1 - x_{j,kl}) \left(1 - \sum_{s=1}^{M} p_s \frac{q_{s,(k,l)}}{\sum_{mp} r_{mp}\, q_{s,(mp)}} \right)^{-2} \right] \tag{8}$$

Let us index all the power demand-to-power source ordered pairs in the network as $(k\ l)_j$; j = 1,…, K. Let us then define the column vectors in (9) below.

$$\begin{matrix} q_{i,((kl)_1)} \\ . \\ . \quad ; i = 1,...,M \\ . \\ q_{i,((kl)_K)} \end{matrix} \tag{9}$$

From the above derivations and discussions, we conclude then that $\{p_i\}$ identifiability is represented by the sign of the quadratic expression in (8), where this sign is negative if and only if the vectors in (9) are all linearly independent. Thus, the $\{p_i\}$ is identifiable if and only if the vectors in (9) are linearly independent. Alternatively, subsets of components in $\{p_i\}$ which are identifiable correspond to linearly independent routing vectors in (9). Such subsets determine an initial clusterization of the network, where the cluster sizes arising from them are generally larger than those which will arise from delay constraints imposed on the reliable detection of changes.

Let us now assume that we have identified M_a linearly independent routing vectors and matching a_s ;s = 1,…,M_a identifiable lines in the network. For such vectors, let us define:

$$h(a_s,(kl)_n) = \frac{q_{s,(kl)_n}}{\sum_{mp} r_{mp} q_{s,(mp)}} \tag{10}$$

Then, the ML system in (5) can be written as:

$$\sum_{n=1}^{M_a} \sum_{j=1}^{N_{(kl)_n}} h(a_s,(kl)_n) \left[\frac{x_{j,(kl)_n}}{\sum_{s=1}^{M_a} p_{a_s} \cdot h(a_s,(kl)_n)} - \frac{1 - x_{j,(kl)_n}}{1 - \sum_{s=1}^{M_a} p_{a_s} \cdot h(a_s,(kl)_n)} \right] = 0 \tag{11}$$

$$; i = 1,..., M_a$$

Due to the relationships in (3), the ML estimates of the probabilities $\{v_i\}$ are given by the system of equations in (11), if $\{p_i\}$ is substituted by $\{v_i\}$ and $h(a_s,(kl)_n)$ is substituted by the routing probability $q_{s,(kl)}$. The resulting ML system depends then only on observations and routing probabilities; not on relative loads.

We will complete this section by stating a theorem whose proof is in the appendix.

Theorem 1

The ML estimate in (11) is asymptotically consistent and efficient if each true value of the identifiable components in $\{p_i\}$ is larger than some $e_j > 0$, where

$$\sum_j \mathrm{e_j} < 1$$

Theorem 1 states an important algorithmic guideline. It basically implies that perfectly operating or totally disconnected (k*l*) power demand-to-power source pairs should be excluded from the ML algorithm, because they may dominate the estimation scheme and lead to false overall estimates.

3. A STOCHASTIC APPROXIMATION ESTIMATION ALGORITHM

In this section, we are seeking sequential ML algorithms for the identifiable lines in the network. Drawing from the notation in Section 2 and using the vector notation $\bar{\mathrm{p}}$ for the set $\{p_i\}$ of probabilities, we first define a vector for given identifiable pair $(kl)_n$ and observed outcome x from the pair:

$$\overline{G}^T(x,(kl)_n,\bar{p}) = \{\, h(a_s,(kl)_n)\,[\, x_{(kl)_n}\left(\sum_{s=1}^{M_a} p_{a_s} h(a_s,(kl)_n)\right)^{-1} -$$

$$-(1-x_{(kl)_n})\left(1-\sum_{s=1}^{M_a} p_{a_s} h(a_s,(kl)_n)\right)^{-1}\,]\;;i=1,...,M_a\,\} \tag{12}$$

We then express the following recursive approximation algorithm, where $x(t+1)$ is the outcome of the $(t+1)^{th}$ pair observation:

$$\hat{\bar{p}}_a(t+1) = \hat{\bar{p}}_a(t) + (t+1)^{-1} L(\hat{\bar{p}}_a(t))\,\overline{G}(x(t+1),\hat{\bar{p}}_a(t)) \tag{13}$$

In (13), sequential estimates of all the identifiable components of the $\{p_i\}$ probabilities are computed at successive observation instances and the evolution of the scalar $L(\,)$ in time is termed *gain sequence*. If $x(t+1)$ is a pair (kl) observation, then:

$$\overline{G}^T(x,\bar{p}) = \overline{G}^T(x,(kl)_n,\bar{p}) \tag{14}$$

In the appendix, we prove that the error of the stochastic approximation estimate in (13) converges asymptotically to a Gaussian variable, if the gain sequence and its first order derivatives are all bounded in the region of probability values where the conditions in Theorem 1 are satisfied. The specific choice of the gain sequence fine-tunes the convergence rate. From the above, as well as from results in [17] and [18], as discussed in the appendix, we conclude the following refinement of the algorithm in (13):

$$\hat{\bar{p}}_a(t+1) = U(\hat{\bar{p}}_a(t) + (t+1)^{-1} L(\hat{\bar{p}}_a(t)) \cdot \overline{G}(x(t+1),\hat{\bar{p}}_a(t)) \tag{15}$$

where,

$$U(X) = \begin{cases} X\;; & \text{if} \quad \mathrm{x_i} > \mathrm{e_i}\,; \text{where } X = \{x_i, i = 1,...,M_a\} \\ & \text{truncation of } x_i \text{ to } \mathrm{e_i} \text{ otherwise} \end{cases} \tag{16}$$

where

$$L_0(\overline{p}_a) = 2^{-1}(d_1^{-1} + d_{M_a}^{-1})\ ;\ \text{if}\ d_1, d_{M_a} \neq 0. \tag{17}$$

for the two terms in (17) being the smallest and largest eigenvalues of the ML covariance matrix below.

$$D(\overline{p}_a) = \sum_{n=1}^{M_a} D_{(kl)_n}\left(\overline{p}_a\right) \tag{18}$$

$$D_{(kl)_n}(\overline{p}_a) = \{\, h(a_i,(kl)_n\ h(a_j,(kl)_n)\left[\sum_{s=1}^{M_a} \overline{p}_{a_s} h(a_s,(kl)_n)\right]^{-1} \cdot$$

$$\cdot\left[1 - \sum_{s=1}^{M_a} \overline{p}_{a_s} h(a_s,(kl)_n)\right]^{-1};\qquad i,j = 1,...,M_a\} \tag{19}$$

The stochastic approximation algorithm for the set $\{v_i\}$ of probabilities is identical to that for the set $\{p_i\}$, as presented in this section.

4. THE DISTRIBUTED DETECTION OF CHANGE ALGORITHM (DDCA)

In Section 2, we determined an initial network clusterization by isolating network lines which are identifiable by an ML estimation algorithm. In Section 3, we developed a sequential stochastic approximation format of the latter ML algorithm, as implemented within a set of identifiable network lines. In this section, we develop a sequential Distributed Detection of Change Algorithm (DDCA), still maintaining operations within an identifiable set of lines. In particular, the algorithm uses the sequential steps in the stochastic approximation algorithm of Section 3, in conjunction with the sequential evolution of the algorithms in [19,20] and [21], to detect changes from a satisfactory to an unsatisfactory regions of $\{p_i\}$ values.

The algorithm in [19,20] and [21] is developed to detect a change from one distribution to another in an optimal fashion and involves a threshold parameter $\delta > 0$. In particular, the algorithm is asymptotically optimal in the sense that, for $\delta \rightarrow \infty$, the expected time for detecting a correct change is of order *l*og δ, while the expected time for an incorrect decision is of order δ, while there exists no algorithm attaining faster correct decision subject to order δ speed of incorrect decision. Here, we transfer the concept of sequential stochastic approximation estimation to the concept of asymptotically optimal sequential detection of change via the following logical steps:

(i) Let us assume that a satisfactory functionality of each line *i* is reflected by a probability v_i which is bounded from above by a given value ρ_i. Let us assume that unsatisfactory functionality is then reflected by a v_i value which is bounded from below by $\rho_i + \eta_i$, where $\eta_i > 0$ and $\rho_i + \eta_i < 1$.

(ii) Let us then require that we detect a ρ_i to $\rho_i + \eta_i$ change per v_i by detecting rapidly a change from a likelihood function $f(\{v_i = \rho_i\})$ to a likelihood function $f(\{v_i = \rho_i + \eta_i\})$, where $f(\{v_i\})$ is given by (4), when $\{p_i\}$ is substituted by $\{v_i\}$. Via the logic presented in [20], this requirement may be transformed to performing a sequential stochastic approximation estimate of the probabilities $\{p_i\}$ using the difference likelihood function $f(\{v_i = \rho_i + \eta_i\}) - f(\{v_i = \rho_i\})$ and applying a modified version of the algorithm in [20] to it.

Considering the probability $q_{s,(kl)_n}$ let us define,

$$y(\{\rho_1\},(kl)_n) = \sum_{1 \le i \le M} \rho_i\ q_{i,(kl)_n} \tag{20}$$

$$y(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n) = \sum_{1 \le i \le M} \rho_i \; q_{i,(kl)_n} + \eta_s \, q_{s,(kl)_n} \tag{21}$$

$$\gamma(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n) = \left(\log \frac{1 - y(\{\rho_i\}, (kl)_n)}{1 - y(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n)} \right) /$$

$$/ \left(\log \frac{[1 - y(\{\rho_i\}, (kl)_n)] \cdot y(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n)}{y(\{\rho_i\}, (kl)_n) \cdot [1 - y(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n)]} \right) \tag{22}$$

By substitution in the modified expression (4), as applying to the probabilities $\{v_i\}$, we conclude that the likelihood difference $f(\{v_i=\rho_i\};\ i \neq s,\ v_s = \rho_s + \eta_s\ ) - f(\{v_i = \rho_i\})$ equals a positive constant, times the expression below:

$$F(\{\rho_1\}, \rho_s + \eta_s) = \sum_{kl} \left(x_{j,kl} - \gamma(\{\rho_i\},\ \rho_s + \eta_s,\ a_s,\ (kl)_n) \right) \tag{23}$$

Equivalently, let us consider the case, where subject to $\{v_i = \rho_i\ ;\ i \neq s,\ v_s = \rho_s + \eta_s\}$ versus $\{v_i = \rho_i$; all i$\}$, the stochastic approximation algorithms in (13) to (19), as modified for the probability set $\{v_i\}$, continue operating after they have converged to their $\{ v_i = \rho_i\ ;\ i \neq s,\ v_s = \rho_s + \eta_s \}$ versus $\{v_i = \rho_i$; all i$\}$ values. Then, via substitution of the pertinent variables in (12), we conclude without much effort that the difference between the two stochastic approximation estimates – each corresponding to $\{ v_i = \rho_i\ ;\ i \neq s,\ v_s = \rho_s + \eta_s \}$ versus $\{v_i = \rho_i$; all $i\}$ true convergent values- is expressed as functions of $x_{(kl)} - \gamma(\{\rho_i\},\ \rho_s + \eta_s,\ a_s,\ (kl)_n)$, where $\gamma(\{\rho_i\},\ \rho_s + \eta_s,\ a_s,\ (kl)_n)$ is given by (22). Extending the logic explained in [20] and modifying the gain in the stochastic approximation algorithm, we thus define the following sequential detection of change algorithm:

Algorithm (DDCA)

Consider the identifiable lines. Let us denote by m(s) the m-th observation involving line s and let this observation be a pair (kl) observation; let us denote this observation $x_{m(s)}$. Then, the sequential detection of change $\rho_s \rightarrow \rho_s + \eta_s$ algorithm $\{ W(m(s)) \}$ operates as follows:

Given design threshold parameter $\delta_s > 0$:

$$W(0) \equiv 0$$

$$W(m(s)+1) = \max\ (0,\ W(m(s)) + \left(x_{(m+1)(i)} - \gamma(\{\rho_i\},\ \rho_s + \eta_s,\ s,\ (kl)) \right)) \tag{24}$$

Stop the first time n(s), such that $W(n(s)) \geq \delta_s$ and decide that the change $\rho_s \rightarrow \rho_s + \eta_s$ has occurred.

The above algorithm uses only observations and routing probabilities and basically detects a change from a Bernoulli random variable with parameter ρ_s to another Bernoulli random variable with parameter $\rho_s + \eta_s$, where the Kullback-Leibler number between these two Bernoulli variables is:

$$I(\rho_s, \rho_s + \eta_s) = (\rho_s + \eta_s) \log \left(\frac{\rho_s + \eta_s}{\rho_s} \right) + (1 - \rho_s + \eta_s) \log \left(\frac{1 - \rho_s + \eta_s}{1 - \rho_s} \right) \tag{25}$$

Directly from [14], [20] and [21], we can express the following theorem:

Theorem 2

Let N_s denote the extended stopping variable induced by the algorithm in (24); that is, $N_s \triangleq \inf\{n : W(n) \geq \delta_s\}$. Then,

(a) Given that no $\rho_s \rightarrow \rho_s + \eta_s$ change occurs throughout all observations,

$$As\ \delta_s \rightarrow \infty,\ E\{N_s\} \geq \delta_s / 2 \tag{26}$$

(b) Given that the $\rho_s \rightarrow \rho_s + \eta_s$ change occurred before the beginning of observations,

$$As\ \delta_s \rightarrow \infty,\quad E\{N_s\} \approx I^{-1}(\rho_s, \rho_s + \eta_s) \log \delta_s \tag{27}$$

(c) Subject to the constraint in (a), there is no algorithm which attains asymptotic (for $\delta_s \rightarrow \infty$) expected stopping time E { N_s } less than that in (27) when all observations originate from the Bernoulli variable, with parameter $\rho_s + \eta_s$.

Theorem 2 establishes the asymptotic optimality of the algorithm in (24), in terms of expected stopping time. Non-asymptotically, the performance of the algorithm is determined by the power and false alarm probabilities it induces, as functions of time. Given finite threshold value δ_s , the false alarm probability $\alpha(\delta_s, n)$ is the probability that the algorithm crosses the threshold at the nth observation for the first time, given that no change $\rho_s \rightarrow \rho_s + \eta_s$ has occurred. Given the same threshold, the power probability $\beta(\delta_s, n)$ is the probability that the algorithm crosses the threshold at the nth observation for the first time, given that the change $\rho_s \rightarrow \rho_s + \eta_s$ occurred before the collection of observations began. The methodology for the recursive in time computations of the probabilities $\alpha(\delta_s, n)$ and $\beta(\delta_s, n)$ in the presence of the Bernoulli model represented by the algorithm in (24) is an extension of that found in [14] and [19]. The design decision regarding the operating threshold δ_s is based on the required power versus false alarm tradeoff. The specific requirement is then that at a given time n, the power probability exceeds a given lower bound, while the false alarm probability remains below another given upper bound. Extensive discussion on the selection of operating decision thresholds for the non-distributed simpler version of the algorithm can be found in [22].

5. NUMERICAL EVALUATIONS – AN EXAMPLE

Given a power distribution system, the DDCA is implemented stepwise, as follows:

(i) The global system matrix with columns as those in (9) is first formed and the M_a lines to be monitored are identified. From this matrix, it is first investigated if the M_a lines are identifiable within the global system matrix: if yes, no further clusterization of the system is required; if not, appropriate clusterization is imposed, to warrant identifiability of all M_a lines. Each such cluster comprises then an identifiable system.

(ii) Given a cluster/ identifiable system in (i), the probabilities $\{\rho_s\}$ and $\{\rho_s + \eta_s\}$ are selected to reflect normal versus abnormal line conditions, where these probabilities may be reflected by the percentage of time during which abnormalities (current, voltage, power, etc.) may be tolerated versus non-tolerated. Subsequently, the quantities $\gamma(\{\rho_s\}, \{\rho_s + \eta_s\}, a_s, (kl)_n)$ in (22) are computed, named *updating steps*.

(iii) Thresholds $\{\delta_i\}$ are selected based on the power and false alarm curves they induce, as explained in Section 4 above and as studied in detail in [23].

(iv) The quantities in (ii) and (iii) are used to implement the DDCA in (24).

We note that in the case where $\rho_s = \rho$ and $\eta_s = \eta$; for all s, then the *updating steps* in (22) take the following simplified form:

$$\gamma(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n) = \left(\log\frac{1-\rho}{1-\rho-\eta q_{s,(kl)_n}}\right) /$$

$$/\left(\log\frac{[\rho+\eta q_{s,(kl)_n}]\cdot(1-\rho)}{[1-\rho-\eta q_{s,(kl)_n}]\rho}\right) \quad (28)$$

In this section, we consider an example of a simple hypothetical 12.66 kV distribution system with 33 buses, 37 lines, and a looping (auxiliary) branch [23], exhibited in Fig. 1, where it is assumed that the power sources in the system are located at buses 1, 6 and 12 and the loads are located at buses 8, 14, 18, 19, 23 and 26. We also consider the possibility that the auxiliary line 35 may be utilized 20% of the time, when line 18 is disconnected (80% of the time), e.g. for minimizing real power losses and improving the voltage profile [24] or optimal day-ahead operational scheduling [25] via network topology reconfiguration. Subsequently, two different scenarios arise: one referring to the above utilization of line 35 and one when line 35 is absent. In Tables 1 and 5, we exhibit the routing probabilities induced by the two different scenarios, where we indicate in bold, the vulnerable lines which need to be monitored: lines 1, 4, 6, 17, 22, 25 and 35 (35 is absent in scenario 1). In view of the later lines of interest, the matrix in Table 1 has a maximum of three independent columns; corresponding to lines 17, 22 and 25, while Table 5 has a maximum of four independent columns; corresponding to lines 17, 22, 25 and 35. Thus, without clusterization, not all lines of interest can be monitored via source- to- demand/load measurements, in both scenarios. Non unique, clusterization approaches are exhibited for each scenario, by Tables 2, 3, 4; 6, 7, 8, where a final clusterization decision may be dictated by possible additional limiting physical factors of the network, such as line ratings. We note that, in both scenarios, lines 17, 22 and 25 may be monitored by the global non-clustered system, in which case measurements from all three sources, 1, 6 and 12 contribute to their monitoring; then, as compared to the clusterized approaches, their monitoring is accelerated.

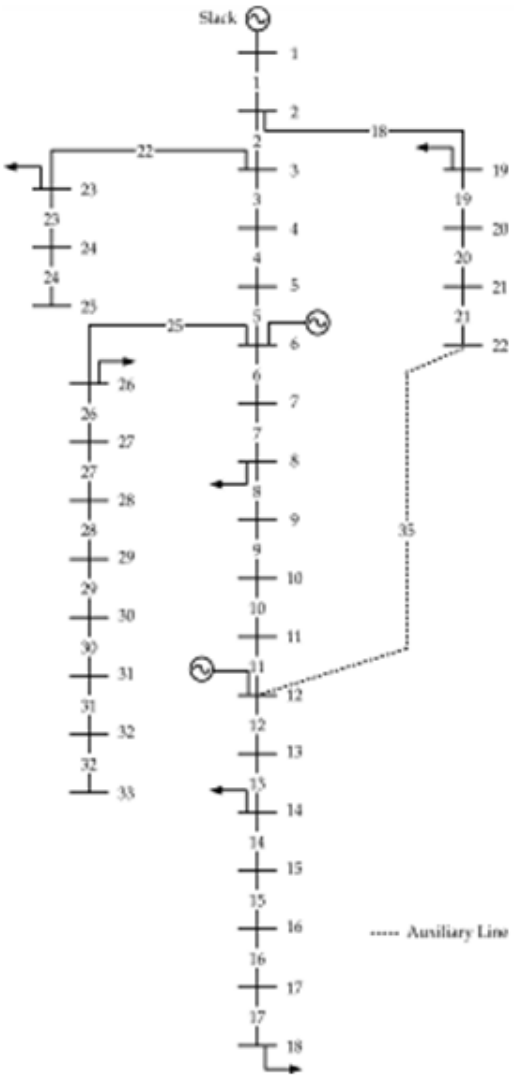

Fig. 1

An example of a power distribution system

Scenario 1

Sources: Buses 1, 6, 12
Loads: Buses 8, 14, 18 19 23, 26
Involved lines: 1, 2, 3 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 25.

Auxiliary line absent.

Lines to be monitored: 1, 4, 6, 17, 22, 25

Table 1

Global system matrix
{ $q_{i,(kl)}$ } Identifiability probabilities for the power distribution system in Fig. 1

Pair (kl) Source k	**Line**																			
Load *l*	**1**	2	3	**4**	5	**6**	7	8	9	10	11	12	13	14	15	16	**17**	18	**22**	**25**
(1,8)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,14)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0
(1,18)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0
(1,19)	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
(1,23)	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
(1,26)	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
(6,8)	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(6,14)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0
(6,18)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0
(6,19)	0	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
(6,23)	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
(6,26)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
(12,8)	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0
(12,14)	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0
(12,18)	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0
(12,19)	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	**1**	0	0
(12,23)	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	**1**	0
(12,26)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	**1**

Without clusterization, only lines 17, 22 and 25 can be monitored. A clusterization choice is shown below

Table 2

Cluster 1: Source 1 with loads 8, 14, 18, 19, 23, 26
It monitors line 1.

Pair (k*l*) Source k Load *l*	Line
	1
(1,8)	**1**
(1,14)	**1**
(1,18)	**1**
(1,19)	**1**
(1,23)	**1**
(1,26)	**1**

Table 3

Cluster 2: Source 6 with loads 8, 14, 18, 19, 23, 26
It monitors lines 4, 6 and 25.

Pair (k*l*) Source k	Line		
Load *l*	**4**	**6**	**25**
(6,8)	0	**1**	0
(6,14)	0	**1**	0
(6,18)	0	**1**	0
(6,19)	**1**	0	0
(6,23)	**1**	0	0
(6,26)	0	0	**1**

Table 4

Cluster 3: Source 12 with loads 18, 23.
It monitors lines 17 and 22.

Pair (k*l*) Source *k*	Line	
Load *l*	**17**	**22**
(12,18)	**1**	0
(12,23)	0	**1**

Scenario 2

Sources: Buses 1, 6, 12
Loads: Busses 8, 14, 18 19 23, 26
Involved lines: 1, 2, 3 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 35.
Auxiliary line 35 is connected 20% of the time while line 18 is disconnected.
Lines to be monitored: 1, 4, 6, 17, 22, 25, 35

Table 5

Global System Matrix
{ $q_{i,(k/)}$ } Identifiability Probabilities for the Power Distribution System in Fig. 1

Pair (k/) Source k	**Line**																						
Load *l*	**1**	2	3	**4**	5	**6**	7	8	9	10	11	12	13	14	15	16	**17**	18	19	20	21	**22**	**25**	**35**
(1,8)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,14)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0
(1,18)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0
(1,19)	**1**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	0	0	0	0	0	0	**.8**	**.2**	**.2**	**.2**	0	0	**.2**
(1,23)	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
(1,26)	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
(6,8)	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(6,14)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0
(6,18)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0
(6,19)	0	**.8**	**.8**	**.8**	**.8**	**.2**	**.2**	**.2**	**.2**	**.2**	**.2**	0	0	0	0	0	0	**.8**	**.2**	**.2**	**.2**	0	0	**.2**
(6,23)	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
(6,26)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
(12,8)	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0
(12,14)	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0
(12,18)	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0
(12,19)	0	**.8**	**.8**	**.8**	**.8**	**.8**	**.8**	**.8**	**.8**	**.8**	**.8**	0	0	0	0	0	0	**.8**	**.2**	**.2**	**.2**	0	0	**.2**
(12,23)	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	**1**	0	0
(12,26)	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	**1**	0

Without clusterization, only lines 17, 22, 25 and 35 can be monitored. A clusterization choice is shown below

Table 6

Cluster 1: Source 1 with loads 8, 14, 18, 19, 23, 26
It monitors line 1.

Pair (k*l*) Source k Load *l*	Line
	1
(1,8)	**1**
(1,14)	**1**
(1,18)	**1**
(1,19)	**1**
(1,23)	**1**
(1,26)	**1**

Table 7

Cluster 2: Source 6 with loads 8, 14, 18, 23, 26
It monitors lines 4, 6 and 25.

Pair (k*l*) Source k Load *l*	Line		
	4	**6**	**25**
(6,8)	0	**1**	0
(6,14)	0	**1**	0
(6,18)	0	**1**	0
(6,23)	**1**	0	0
(6,26)	0	0	**1**

Table 8

Cluster 3: Source 12 with loads 18, 19, 23.
It monitors lines 17, 22, and 35.

Pair (k*l*) Source k Load *l*	Line		
	4	**6**	**25**
(12,18)	**1**	0	0
(12,19)	0	0	**.2**
(12,23)	0	**1**	0

Regarding the monitoring of the vulnerable lines, let us also assume that $\rho_s = \rho$ and $\eta_s = \eta$; for all s, in which case the simplified form of the *updating step* in (28) is used in the implementation of the DDCA algorithm in (24). Finally, let us assume that the tolerable versus non-tolerable power system conditions are reflected by the choices $\rho = 0.01$ and $\eta = 0.04$; for non-auxiliary lines, while $\rho = 0.01$ and $\eta = 0.2$; for auxiliary lines. The latter choices, in conjunction with the probabilities in Tables 2, 3, 4; 6, 7, 8 induce then the following specific steps in the implementation of the monitoring algorithm in (24):

(a) For line 1

In both scenarios, update the algorithm each time a (1,8), (1,14), (1,18), (1,19), (1,23) or (1,26) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29) below.

$$\gamma(\{\rho_i\}, \rho_s + \eta_s, a_s, (kl)_n) = \left(\log \frac{1-0.01}{1-0.01-0.04}\right) / \left(\log \frac{[0.01+0.04]\cdot(1-0.01)}{[1-0.01-0.04]0.01}\right) \quad (29)$$

(b) For line 4

In scenario 1, update the algorithm each time a (6,19) or (6,23) source-to-demand/load pair measurement is collected. In scenario 2, update each time a (6,23) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

(c) For line 6

In both scenarios, update the algorithm each time a (6,8), (6,14) or (6,18) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

(d) For line 25

In both scenarios, update the algorithm each time a (6,26) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

(e) For line 17

In all both scenarios, update the algorithm each time a (12,18) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

(f) For line 22

In both scenarios, update the algorithm each time a (12,23) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

(g) For line 35

In scenario 2, update the algorithm each time a (12,19) source-to-demand/load pair measurement is collected. For each source-to-demand pair measurement, use the *updating step* in (29).

Selecting a threshold value $\delta_s = 4.47$ for the algorithm with *updating step* as in (29), we attain superior values of the false alarm and power probabilities, α (δ_s , n) and β (δ_s , n). Specifically, in 100 measurements, the power equals then 0.98, while the false alarm is practically zero (see [14], Figure 8.5.1). The specific tolerable upper limit on the number of measurements is determined by the time required for their collection, in conjunction with the time-limit demanded for the detection of substantial changes.

6. CONCLUSIONS

Our overall objective has been the performance monitoring of a power distribution system, for real-time dynamic operational adjustment in the presence of substantial operational changes. Then, such changes need to be identified timely and accurately before pertinent adjustments be performed, necessitating a distributed approach. This approach may be implemented by network clusterization whose design should be dictated by the accuracy and delay constraints imposed on the detection and identification of operational changes.

The general approach taken in this paper involves the following steps: (i) We first consider the initially non-clusterized power distribution system and determine the -current, voltage, power- variations perceived as considerable changes; we also determine the vulnerable lines which need monitoring. (ii) We secondly formulate a recursive maximum likelihood (ML) approach which naturally points to an initial network clusterization via incorporated sufficient identifiability conditions. (iii) We subsequently develop, analyze and evaluate a distributed sequential detection of change algorithm, implemented by the supporting data computer-communication network, whose performance (including accuracy and decision delay) is controlled by a set of threshold parameters and the architecture and dimensionality of each cluster.

Specifically, we have proposed a distributed algorithm for monitoring the quality of power lines (and their incorporated equipment) in power distribution systems. The algorithm utilizes sequentially processed power source-to-demand measurements within an identifiable system, to generate alerts about faulting lines, rapidly and with a high level of accuracy. System identifiability, in conjunction with constraints on the speed of correct decisions, provide system clusterization guidelines.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Hatziargyriou N, Asano H, Iravani R Marnay C. Microgrids-An overview of ongoing research, development and demonstration projects. IEEE Power Energy Mag. 2007;5(4):78-94.
2. Manitsas E, Singh R, Pal BC, Strbac G. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. IEEE Trans. Power Syst. 2012;27;(4):1888-1896.
3. Ferdowsi M, Benigni A, Lowen A, Zargar B, Monti A, Ponci F. A Scalable Data-Driven Monitoring Approach for Distribution Systems. Instrumentation and Measurement, IEEE Transactions on, On page(s): 2015;64(5):1300-1313.
4. Woolley NC, Milanovic JV. Statistical estimation of the source and level of voltage unbalance in distribution networks. IEEE Trans. Power Del. 2012;27(3):1450-1460.
5. Nusrat N, Irving M, Taylor G. Development of distributed state estimation methods to enable smart distribution management systems. In Proc. IEEE Int. Symp. Ind. Electron. (ISIE). 2011;1691–1696.
6. Muscas C, Pau M, Pegoraro PA, Sulis S, Ponci F. Monti A. Multiarea distribution system state estimation. Instrumentation and Measurement, IEEE Transactions on, on page(s). 2015;64(5):1140-1148.
7. Castello P, Liu J, Muscas C, Pegoraro PA, Ponci F, Monti A. A fast and accurate PMU algorithm for P+M class measurement of synchrophasor and frequency. IEEE Trans. Instrum. Meas. 2014;63(12):2837–2845.

8. Singh R, Manitsas E, Pal BC, Strbac G. A recursive Bayesian approach for identification of network configuration changes in distribution system state estimation. IEEE Trans. Power Syst. 2010;25(3):1329-1336.
9. Pau M, Pegoraro PA, Sulis S. Efficient branch-current-based distribution system state estimation including synchronized measurements. IEEE Trans. Instrum. Meas. 2013;62(9):2419-2429.
10. Bernieri A, Betta G, Liguori C, Losi A. Neural networks and pseudo-measurements for real-time monitoring of distribution systems. IEEE Trans. Instrum. Meas.1996;45(2):645–650.
11. Muscas C, Sulis S, Angioni A, Ponci F, and Monti A. Impact of different uncertainty sources on a three-phase state estimator for distribution networks. IEEE Trans. Instrum. Meas. 2014;63(9):2200–2209.
12. Pegoraro PA , Sulis S. On the robustness in distribution system state estimation," in Proc. IEEE Int. Instrum. Meas. Technol. Conf. 2012;776–780.
13. Abur A, Exposito AG. Power System State Estimation: Theory and Implementation. New York, NY, USA: Marcel Dekker; 2004.
14. Kazakos D, Papantoni-Kazakos P. Detection and Estimation, Computer Science Press, New York; 1990.
15. Sakrison D. Stochastic approximation: A recursive method for solving regression problems. Advances in Communication Systems, (Ed. A.V. Balakrishnan), Academic Press, New York. 1966;2:51-106.
16. Papantoni-Kazakos P. The potential of end-to-end observations in trouble localization and quality control pf network links. IEEE Trans. Commun. 1979;27 (1):16-28.
17. Sacks J. Asymptotic distribution of stochastic approximation methods. Ann. Math. Stat. 1958;29:373-405.
18. D. Kazakos. Recursive estimation of prior probabilities using a mixture", IEEE Trans. Inf. Th. 1977;IT-23(2):203-211.
19. Page ES. Continuous inspection schemes. Biometrika, 1954;41:100-115.
20. Bansal RK, Papantoni-Kazakos P. An algorithm for detecting a change in a stochastic process. IEEE Trans. Inf. Th. 1986; IT-32:227-235.
21. Burrell AT, P. Papantoni-Kazakos. Robust sequential algorithms for the detection of changes in data generating processes. Journal of Intelligent and Robotic Systems. 2010;60(1):3-17.
22. Etellisi EA. Burrell AT. Papantoni-Kazakos P. A Core Algorithm in Object Tracking and Monitoring Distributed Wireless Sensor Networks. International Journal of Sensor Networks and Data Communications (IJSNDC). 2012;1:15. Article IDX110601, DOI:10.4303/ijsndc/x110601.
23. Baran ME, Wu FF. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Transactions on Power Delivery. 1989;9(4):101–102.
24. Rao RS, Ravindra K, Satish K,Narasimham SVL. Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation. Power Systems, IEEE Transactions on, on page (s): 2013;317-325.
25. Golshannavaz S, Afsharnia S, Aminifar F. Smart distribution grid: optimal day-ahead scheduling with reconfigurable topology. Smart Grid, IEEE Transactions on, On page(s). 2014;5(5):2402–2411.
26. Cramer H. Mathematical methods of statistics. Princeton Univ., Princeton NJ; 1946.

APPENDIX

Proof of Theorem 1

Define:

$$g_{M_a}(x, p_a, (kl)_n) = x \log\left(\sum_{s=1}^{M_a} \hat{p}_{a_s}\ h(a_s, (kl)_n)\right) +$$

$$+ (1-x) \log\left(1 - \sum_{s=1}^{M_a} \hat{p}_{a_s}\ h(a_s, (kl)_n)\right).$$

It is easy to verify that Cramer's conditions [26] for asymptotic weak consistency and efficiency are satisfied. Specifically,

i) The derivatives $\dfrac{\partial^{i+j}}{\partial^i \hat{p}_{a_k}\, \partial^j \hat{p}_{a_q}} g_{M_a}(x, p_a, (kl)_n)$ for i,j= 1,2,3 exist.

ii) $E\left\{\dfrac{\partial}{\partial \bar{p}_{a_k}} g_{M_a}(x, \bar{p}_a, (kl)_n)\right\} = 0 \ ; \quad k = 1, \cdots, M_a \ ; \bar{p}$

where the expectation is taken over all $x_{j,(kl)_n}$ outcomes and over all considered $(kl)_n$ pairs.

iii) $E\left\{\dfrac{\partial}{\partial \hat{p}_{a_k}} g(x, \bar{p}_a, (kl)_n) \dfrac{\partial}{\partial \hat{p}_{a_g}} g(x, \bar{p}_a, (kl)_n)\right\} =$

$$= -E\left\{\frac{\partial^2}{\partial \hat{p}_{a_k}\, \partial \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n)\right\} < \infty .$$

It is easily seen that here:

$$E\left\{\frac{\partial}{\partial \hat{p}_{a_k}} g(x, \bar{p}_a, (kl)_n) \frac{\partial}{\partial \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n)\right\} =$$

$$= - E\left\{ \frac{\partial^2}{\partial \hat{p}_{a_k} \partial \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n) \right\} .$$

$$= \sum_{n=1}^{M_a} h(a_k, (kl)_n) \cdot h(a_q, (kl)_n)$$

$$\frac{1}{\left[\sum_{s=1}^{M_a} \hat{p}_{a_s} h(a_s, (kl)_n) \right] \left[1 - \sum_{s=1}^{M_a} \hat{p}_{a_s} h(a_s, (kl)_n) \right]} < \infty$$

(because all $p_{a_i} > 0$), where the expectation is again over all $x_{j,(kl)_n}$ outcomes and all $(kl)_n$ pairs considered.

iv) $\left| \frac{\partial^{i+j}}{\partial^i \hat{p}_{a_k} \partial^j \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n) \right| < m(x)$, where $m(x)$ are some almost everywhere finite functions.

$i, j = 1, 2, 3;$ $k, q = 1, \ldots M_a$

But here

$$\left| \frac{\partial^{i+j}}{\partial^i \hat{p}_{a_k} \partial^j \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n) \right|$$

is itself finite (x = 0, 1) and this condition is trivially satisfied.

The ML scheme described by the system in (11), being asymptotically efficient (as proved by Theorem 1) results in the asymptotic covariance matrix $D(\bar{p}_a)$ of the estimate $\bar{p}_a$ satisfying the Cramer bound. That is :

$$D(\bar{p}_a) = E\left\{ (\hat{\bar{p}}_a - \bar{p}_a)(\hat{\bar{p}}_a - \bar{p}_a)^{\mathrm{T}}{}_{\bar{p}_a} \right\} =$$

$$= \left\{ d_{iq}(\bar{p}_a); \quad i, q = 1, \cdots, M_a \right\}$$

where,

$$d_{iq}(\bar{p}_a) = -E\left\{\frac{\partial^2}{\partial \hat{p}_{a_i}\,\partial \hat{p}_{a_q}} g(x, \bar{p}_a, (kl)_n)\right\}$$

$$= -\sum_{n=1}^{M_a} h(a_i, (kl)_n)\cdot h(a_q, (kl)_n)$$

$$\cdot \frac{1}{\left[\sum_{s=1}^{M_a} p_{a_s}\, h(a_s, (kl)_n)\right]\left[1-\sum_{s=1}^{M_a} p_{a_s}\, h(a_s, (kl)_n)\right]} \cdot$$

Convergence of the Stochastic Approximation Algorithm in (11)

Let us define the vector regression function below.

$$R(\hat{\bar{p}}_a) = -E\left\{L(\hat{\bar{p}}_a)\cdot G(x, \hat{\bar{p}})_{/\bar{p}_a, \hat{\bar{p}}_a}\right\} = -\sum_{n=1} L(\hat{\bar{p}}_a, (kl)_n)$$

$$E\left\{\bar{G}(x, (kl)_n, \hat{\bar{p}})_{/\bar{p}_a, \hat{\bar{p}}_a}\right\} = -\sum_{n=1}^{M_a} L(\hat{\bar{p}}_a, (kl)_n)$$

$$\cdot h(a_i, (kl)_n)\left[\frac{\sum_{s=1}^{M_a} p_{a_s} h(a_s, (kl)_n)}{\sum_{s=1}^{M_a} \hat{p}_{a_s}\, h(a_s, (kl)_n)} - \right.$$

$$\left. - \frac{1-\sum_{s=1}^{M_a} p_{a_s} h(a_s, (kl)_n)}{1-\sum_{s=1}^{M_a} \hat{p}_{a_s}\, h(a_s, (kl)_n)}\right]; i = 1, \cdots, M_a \qquad \text{(A.1)}$$

and we use Sacks' conditions [17], as in [18]. The required conditions for convergence, as given by Sacks' [17] are:

1. $R(\bar{p}_a) = 0$ and $\inf_{\hat{\bar{p}}_a} (\hat{\bar{p}}_a - \bar{p}_a) R^T(\hat{\bar{p}}_a) > 0$ for $\hat{\bar{p}}_a$ in the region of interest and for $\left\| \hat{\bar{p}}_a - \bar{p}_a \right\| > s > 0$ for some s > 0.

2. There exists some positive $K_1 < \infty$ such that $\| R(\bar{p}_a) \| \leq K_1 \left\| \hat{\bar{p}}_a - \bar{p}_a \right\|$ for all $\hat{\bar{p}}_a$ in the region of interest.

3. $R(\hat{\bar{p}}_a) = (\hat{\bar{p}}_a - \bar{p}_a) B + \delta(\hat{\bar{p}}_a, \bar{p}_a)$, where *B* positive definite matrix and

$$\left\| \hat{\bar{p}}_a - \bar{p}_a \right\|^{-1} \left\| \delta(\hat{\bar{p}}_a, \bar{p}_a) \right\| \to 0 \quad \text{as} \quad \left\| \hat{\bar{p}}_a - \bar{p}_a \right\| \to 0 \quad .$$

4. Define $Z(x, \hat{\bar{p}}_a) = -L(\hat{\bar{p}}_a) \bar{G}^T(x, \hat{\bar{p}}_a) - R^T(\hat{\bar{p}}_a)$. Then we should have:

$$\sup E\left\{ \left\| Z(x, \hat{\bar{p}}_a) \right\|^2 / \hat{\bar{p}}_a \right\} < \infty$$

$$\lim_{\hat{\bar{p}}_a \to \bar{p}_a} E\left\{ Z(x, \hat{\bar{p}}_a) Z^T(x, \hat{\bar{p}}_a) / \hat{\bar{p}}_a \right\} = S(\bar{p}_a) < \infty$$

where $S(\bar{p}_a)$ is a non-negative definite matrix.

5. The vectors $Z(x(t), \hat{\bar{p}}_a)$ are (for different t's) identically distributed when conditioned on $\bar{p}_a$.

Now let $b_1 \geq b_2 \geq \cdots \geq b_{M_a}$ be the ranked eigenvalues of the matrix *B* in condition 3, and let P be the matrix of the *B* eigenvectors (*P* is orthogonal). Define then:

$$S^*(\bar{p}_a) = P^{-1} S(\bar{p}_a) P \tag{A.2}$$

where $S(\bar{p}_a)$ is the nonnegative definite matrix condition in 4. If $s_{i,j}^*(\bar{p}_a)$ is the *i, j* element of the matrix $S^*(\bar{p}_a)$, Sacks's theorem is expressed as follows.

Theorem (*Sacks*) : If conditions 1-5 are satisfied by the regression function $R(\hat{\bar{p}}_a)$ and if $b_{M_a} > \frac{1}{2}$, then the random vector $t^{1/2}(\hat{\bar{p}}_a(t) - \bar{p}_a)$ is asymptotically normal, with zero mean and covariance matrix PWP^{-1}. *P* is the eigenvector matrix of *B* (condition 3) and *W* is a matrix with *i* , *j* element equal to $(b_i + b_j - 1)^{-1} s_{ij}^*(\bar{p}_a)$, where $S^*(\bar{p}_a)$ is given by (A.2) and $S(\bar{p}_a)$ is defined in condition 4.

In conditions 1-4 and theorem 2, $\bar{p}_a$ is the true value of the vector consisting of the p_is of the identifiable lines.

We now prove that Sacks' conditions are satisfied by the expression in (11).

1. The regression function $R(\hat{\bar{p}}_a)$ in (A.1) obviously satisfies $R(\hat{\bar{p}}_a) = 0$. Also,

$$(\hat{\bar{p}}_a - \bar{p}_a)\, R^T(\hat{\bar{p}}_a) =$$

$$= \sum_{n=1}^{M_a} L(\hat{\bar{p}}_a, (kl)_n) \cdot$$

$$\cdot \frac{\left[\sum_{i=1}^{M_a} (p_{a_i} - \hat{p}_{a_i}) h(a_i, (kl)_n)\right]^2}{\left[\sum_{s=1}^{M_a} \hat{p}_{a_s} h(a_s, (kl)_n) + \xi(\hat{\bar{p}}_{(na)}, (kl)_n)\right]\left[1 - \sum_{s=1}^{M_a} \hat{p}_{a_s} h(a_s, (kl)_n) - \xi(\hat{\bar{p}}_{(na)}, (kl)_n)\right]}$$

which, due to the determinancy of the system is positive for $\left\| \hat{\bar{p}}_a - \bar{p}_a \right\| > 0$

2. Use a Taylor expansion and the middle value theorem to write :

$$r_J(\hat{\bar{p}}_a) = r_J(\bar{p}_a) + \sum_{k=1}^{M_a} (\bar{p}_{a_k} - p_{a_k}) \frac{\partial}{\partial \bar{p}_{a_k}}$$

$$\cdot r_J\left(\bar{p}_a + s_J\left[\hat{\bar{p}}_a - \bar{p}_a\right]\right) \qquad \text{(A.3)}$$

where, $r_J(\cdot)$ is the J^{th} component of the regression function $R(\cdot)$ and $s_J \in [0,1]$. Due to $R(\bar{p}_a) = 0$, we can write using (19) and (A.3):

$$r_J(\hat{\bar{p}}_a) =$$

$$= \sum_{k=1}^{M_a} (\hat{p}_{a_k} - p_{a_k}) \cdot \sum_{n=1}^{M_a} \{L(\bar{p}_a + s_J\,[\hat{\bar{p}}_a - \bar{p}_a],$$

$$, (kl)_n \cdot h(a_J, (kl)_n) \cdot h(a_k, (kl)_n)$$

$$\cdot\left[\frac{\left[\sum_{s=1}^{M_a}\bar{p}_{a_s}h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)\right]}{\left(\sum_{s=1}^{M_a}\left[\hat{p}_{a_s}+s_J(p_{a_s}-\hat{p}_{a_s})\right]h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)\right)^2}+\right.$$

$$\left.+\frac{1-\left[\sum_{s=1}^{M_a}\bar{p}_{a_s}h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)\right]}{\left(1-\sum_{s=1}^{M_a}\left[\hat{p}_{a_s}+s_J(p_{a_s}-\hat{p}_{a_s})\right]h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)\right)^2}\right]-$$

$$-\frac{\partial}{\partial\hat{p}_{a_k}}L(\bar{p}_a+s_J[\hat{\bar{p}}_a-\bar{p}_a],(kl)_n\cdot h(a_J,(kl)_n)$$

$$\cdot\left[\frac{\left[\sum_{s=1}^{M_a}p_{a_s}h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)\right]}{\sum_{s=1}^{M_a}\left[p_{a_s}+s_J(\hat{p}_{a_s}-p_{a_s})\right]h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)}-\right.$$

$$\left.-\frac{1-\sum_{s=1}^{M_a}p_{a_s}h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)}{1-\sum_{s=1}^{M_a}\left[p_{a_s}+s_J(\hat{p}_{a_s}-p_{a_s})\right]h(a_s,(kl)_n)+\xi(\hat{\bar{p}}_{(na)},(kl)_n)}\right]\}\qquad(A.4)$$

Representing the whole expression in the brackets { } above (expression (A.4)) and applying Schwarz's inequality on (A.4) we obtain

$$[r_J(\hat{\bar{p}}_a)]^2\le\left\|\bar{p}_a-\hat{\bar{p}}_a\right\|^2\cdot\sum_{n=1}^{M_a}\left[\sum_{n=1}^{M_a}\{\}\right]^2\qquad(A.5)$$

And summing (A.4) w.r.t *J* we have :

$$\left\|R_J(\hat{\bar{p}}_a)\right\|^2\le\left\|\bar{p}_a-\hat{\bar{p}}_a\right\|^2\cdot\sum_{J=1}^{M_a}\sum_{k=1}^{M_a}\left[\sum_{n=1}^{M_a}\{\}\right]^2\qquad(A.6)$$

If $\bar{p}_a$ and $\hat{\bar{p}}_a$ are both such that for every $1 \le k \le M_a$ there is some positive number e_k such that p_{ak} , $\hat{p}_{ak} > e_k$ and if $L(\hat{\bar{p}}_a), (\partial / \partial \hat{p}_{ak}) L(\hat{\bar{p}}_a)$ are bounded for the $\hat{\bar{p}}_a$'s satisfying the condition $\hat{p}_{ak} > e_k$, then k= $\sum_{J=1}^{M_a} \sum_{k=1}^{M_a} \left[\sum_{n=1}^{M_a} \{\} \right]^2$ is bounded and condition 2 is satisfied.

3. Using again the mean value expansion, but including a second order term we can write (remembering that $R(\bar{p}_a) = 0$) :

$$r_K(\hat{\bar{p}}_a) = \sum_{J=1}^{M_a} (\hat{\bar{p}}_{a_J} - p_{a_J}) \frac{\partial}{\partial \hat{p}_{a_J}} r_k(\bar{p}_a)$$

$$+ (\hat{\bar{p}}_a - \bar{p}_a)^T W_k (\hat{\bar{p}}_a - \bar{p}_a) \qquad \text{(A.7)}$$

where W_k is an $M_a \text{ X } M_a$ matrix, such that

$$W_k = \left\{ w_{k/J} = \frac{\partial^2}{\partial \hat{p}_{a_i} \partial \hat{p}_{a_J}} r_k(\bar{p}_a + s_k[\hat{\bar{p}}_a - \bar{p}_a]) ; \right.$$

$$\left. i, J = 1, \cdots, M_a ; \quad s_k \in [0,1] \right\} \qquad \text{(A.8)}$$

From (A.7) and (A.8) we see that we can write:

$$R(\hat{\bar{p}}_a) = B(\hat{\bar{p}}_a - \bar{p}_a) + \delta(\hat{\bar{p}}_a, \bar{p}_a)$$

where B is an $M_a \text{ X } M_a$ matrix, such that :

$$B = \left\{ b_{kJ} = \sum_{n=1}^{M_a} L(\bar{p}_a, (kl)_m \cdot h(a_k, (kl)_n) \cdot h(a_J, (kl)_n) \right. \quad \cdot$$

$$\cdot \left[\left(\sum_{s=1}^{M_a} p_{a_s} h(a_s, (kl)_n) + \xi(\hat{\bar{p}}_{(na)}, (kl)_n) \right)^{-1} + \right.$$

$$\left. + \left(1 - \sum_{s=1}^{M_a} p_{a_s} h(a_s, (kl)_n) - \xi(\hat{\bar{p}}_{(na)}, (kl)_n) \right)^{-1} \right] ;$$

$$k, J = 1, \cdots, M_a \quad \Big\}. \tag{A.9}$$

It can be easily seen that due to the identifiability of the system, *B* is positive definite. It is obviously also symmetric.

We may further observe that if $L(\bar{p}_a, (kl)_n) = L(\bar{p}_a)\, L(\bar{p}_a, (kl)_n)$ independent of $(kl)_n$, the matrix *B* is then equal to the covariance matrix $D(\bar{p}_a)$ (in (18)) of the ML estimator, times $L(\bar{p}_a)$.

Denote now

$$\delta^T(\bar{p}_a, \hat{\bar{p}}_a) = \{\ (\hat{\bar{p}}_a - \bar{p}_a)^T W_k (\hat{\bar{p}}_a - \bar{p}_a)\ ;$$

$$k = 1, \cdots, M_a\ \}. \tag{A.10}$$

As in condition 2, we can easily see that here $\left\| \delta(\bar{p}_a, \hat{\bar{p}}_a) \right\|$ is also bounded for $\bar{p}_a$, $\hat{\bar{p}}_a$ such that $\forall\, 1 \le k \le M_a\ \exists\, e_k > 0$: p_{a_k} , $\hat{p}_{a_k} > e_k$ and for $L(\hat{\bar{p}}_a, (kP)_n)$, $(\partial / \partial \hat{p}_{a_k}) L(\hat{\bar{p}}_a, (kl)_n)$; $n = 1, \cdots, M_a$ bounded. Then $\left\| \hat{\bar{p}}_a - \bar{p}_a \right\|^{-1} \cdot \left\| \delta(\bar{p}_a, \hat{\bar{p}}_a) \right\| \to 0$ as $\left\| \hat{\bar{p}}_a - \bar{p}_a \right\|^{-1} \to 0$.

4. It is easy to show that, again for $L(\hat{\bar{p}}_a, (kl)_n)$, $(\partial / \partial \hat{p}_{a_k}) L(\bar{p}_a, (kl)_n)$; $n = 1, \cdots, M_a$ bounded and p_{a_k} , $\hat{p}_{a_k} > e_k > 0$, this condition is satisfied.

5. Obviously satisfied. Let us now calculate the matrix $S(\bar{p}_a)$ as defined by condition 4 of Sacks.

$$S(\bar{p}_a) = \{ s_{i,j}(\bar{p}_a)\ ;\ \ i, j = 1, \cdots, M_a\ \}$$

where,

$$s_{i,j}(\bar{p}_a) = \sum_{n=1}^{M_a} L^2(\bar{p}_a, (kl)_m \cdot h(a_i, (kl)_n) \quad \cdot$$

$$\cdot\ h(a_s, (kl)_n) \cdot \left[\sum_{s=1}^{M_a} p_{a_s} h(a_s, (kl)_n) \right.$$

$$+\xi(\hat{\overline{p}}_{(na)},(kl)_n)\Big]^{-1}\Big[1-\sum_{s=1}^{M_a}p_{a_s}$$

$$\cdot\ h(a_s,(kl)_n)-\xi(\hat{\overline{p}}_{(na)},(kl)_n)\Big]^{-1}. \qquad \text{(A.11)}$$

We observe that $S(\overline{p}_a)$ is a weighted ($L^2(\overline{p}_a,(kl)_m)$ weights) version of the ML covariance matrix in (18).

From (A.9), (A.2) and the fact that the P eigenvector matrix is orthogonal, we conclude [17] :

$$S^*(\overline{p}_a)=L^2(\overline{p}_a)\cdot\mathrm{diag}\,(d_1,\cdots,d_{M_a})$$

where $d_1\geq d_2\geq\cdots\geq d_{M_a}$ are the ordered eigenvalues of the ML covariance matrix

$$D(\overline{p}_a)=\sum_{n=1}^{M_a}D_{(kl)_n}(\overline{p}_a).$$

The analysis in [18] for the proper $L(\overline{p}_a)$ choice holds and we may choose :

$$L_0(\overline{p}_a)=2^{-1}(d_1^{-1}+d_{M_a}^{-1});\quad \text{if } d_1,d_{M_a}\neq 0.$$

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/10305