Improved Observability In Distribution Grids using Correlational Measurements

YOHAN JOHN1, VENKATESH VENKATARAMANAN1, (Member, IEEE), ANURADHA M ANNASWAMY1, (Fellow, IEEE) ANAMIKA DUBEY2, (Member, IEEE) and Anurag K Srivastava.2 (Senior Member, IEEE)

1Active-Adaptive Control Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (email: [yjohn, vvenkata, aanna]@mit.edu)
2School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99163 USA (email: [anamika.dubey, anurag.k.srivastava]@wsu.edu)

Corresponding author: Venkatesh Venkataramanan (e-mail: vvenkata@mit.edu).

This work is supported by the US DOE “AGGREGATE: dAta-driven modelinG preservinG contRollable dEr for outaGe mAnagemenT and rEsiliency” project (CID: OE000878) under the Office of Electricity Delivery and Energy Reliability.

ABSTRACT This paper develops a novel pseudo-measurement construct termed a Correlational Measurement (CM) to improve observability and state estimation accuracy. CMs utilize the correlation between similar classes of loads and DERs. The modifications required to accommodate CMs in traditional node voltage or branch current based state estimators are derived. Additionally, an observability analysis framework for distribution systems with DERs is presented to seamlessly integrate CMs. The framework assembles an observability test and an observable island identification procedure from the literature along with a novel observability metric. Finally, an on-line CM parameter estimation procedure is presented. Simulations on the IEEE 123 bus test system and real field data from Pecan Street demonstrate the utility of the proposed CMs and the observability analysis framework in improving observability and state estimation accuracy.

INDEX TERMS Correlation coefficient, Distributed power generation, Observability, Parameter estimation, State estimation

I. INTRODUCTION

In the distribution grid of the future, the power injection from controllable DERs will be a powerful tool for outage management and restoration [1]. However, determining the optimal DER injections and restoration actions is dependent on knowledge of the system states. State estimation (SE), already the cornerstone of transmission energy management systems, will become commonplace in distribution management systems as more measurements become available from deployment of automated metering infrastructure (AMI) [2]. Observability analysis is the first step in SE, as it determines the sufficiency of the available measurements in accurately estimating the current system states [3].

Traditionally, the literature on power system observability analysis could be classified either as topological observability that relies on the existence of a spanning tree of full rank or numerical observability that is based on the rank of the Measurement Jacobian [4]. References [5]–[7] present topological observability analysis methods. Topological methods are at a disadvantage because certain combinations of line parameter values can prevent SE solvability despite apparent observability [8]. References [3], [9]–[11] present numerical observability analysis methods. Numerical methods are vulnerable to ill-conditioning and errors associated with floating-point calculations [6]. Newer hybrid methods were suggested in [12]–[14] to take advantage of the best aspects of both approaches. Recently, [15] introduced a graph-theoretic criterion for local observability, and [16] introduced a probabilistic assessment indicating the accuracy of the SE. This paper’s contribution focuses more on the development and integration of a novel pseudo-measurement so observability is measured using the traditional numerical method. This has the advantage of computational efficiency and often shared operations with SE [9].

In this paper, we apply the non-iterative Observability Test (OT) and Observable Island Identification Procedure (OIIIP) in [9] to the gain matrix from the distribution system state estimator. Because sensors are few in distribution...
systems, pseudo-measurements based on historical load profiles [17] and zero-injection virtual measurements [18] are often employed to ensure observability. There are also newer machine learning based methods used to generate virtual measurements to increase observability [19]. However, both the observability and the accuracy of the state estimates are uncertain when historical data is limited/unavailable. In such cases, machine learning approaches are inappropriate and alternative methods are needed.

To this end, the authors present a new type of pseudo-measurement called a Correlational Measurement (CM). CMs encapsulate knowledge of correlation between demand patterns for similar classes of loads as well as injection patterns for same-technology renewable DERs. To utilize CMs, an Observability Analysis Framework (OAF) for primary distribution systems is also presented. The framework includes an efficient OT and OIIP from the literature and a novel Observability Metric (OM) that quantifies the accuracy of the state estimates.

Related pseudo-measurements in the literature fall into three categories: 1) sensor correlation [20], 2) intra-node correlation [21], [22], and 3) inter-node correlation [22], [23]. References [22], [23] consider correlation due to geographic proximity for probabilistic power flow studies. In this paper, the authors represent inter-node correlation of medium voltage loads and DERs using a novel linear formulation. Load profiles exhibit similarities due to class (residential, commercial, etc.) [24]. DERs such as wind turbines located close together also exhibit similar generation profiles [25]. CMs capture both sources of correlation and interface naturally with the familiar SE algorithms. It is demonstrated that the proposed CMs improve system observability and SE accuracy. The traditional Node Voltage based State Estimation (NVSE) and Branch Current based State Estimation (BCSE) are employed in this paper due to their widespread use, but CMs could easily be incorporated into newer approaches such as Semi-Definite Programming (SDP) relaxation [26], Least Absolute Value (LAV) [19], matrix completion [27], etc.

This work provides four key contributions: 1) We introduce a new class of measurement, i.e., CMs, to improve the system observability and SE accuracy. 2) To leverage CMs, we derive the modifications to the standard node voltage and branch current based SE algorithms and discuss the comparative benefits when using CMs. 3) We propose an on-line CM parameter estimation scheme using a least squares approach. 4) We present a novel OM to quantify the SE accuracy within an observable network. CMs and the proposed OM will provide the Distribution System Operator (DSO) better visibility into the network. Simulations on the IEEE 123 bus test system and real field data from Pecan Street demonstrate the utility of the proposed CMs and the observability analysis framework in improving observability and state estimation accuracy. Ultimately, this will facilitate enhanced grid resiliency through more effective utilization of controllable DERs.

II. BACKGROUND

A. OBSERVABILITY TEST & OBSERVABLE ISLAND IDENTIFICATION

This section briefly outlines the OT and OIIP presented in [9]. The following model relates measurements \( z \in \mathbb{R}^M \) to the state \( x = [\theta^T \ v^T]^T \in \mathbb{R}^N \) that is composed of voltage angles and magnitudes at all the nodes:

\[
    z = h(x) + r
\]

where, \( h(\cdot) \) is a nonlinear function that maps the state to the measurements and \( r \) refers to the measurement errors (assumed zero-mean Gaussian).

Observability of the power system can be stated using the gain matrix \( G \) as follows: A distribution system is observable if and only if \( G \) is full rank [3]. \( G \) is computed as follows:

\[
    G = H^T R^{-1} H
\]

where, \( H = \frac{\partial h(x)}{\partial x} \) is the measurement Jacobian and \( R = \mathbb{E}[rr^T] \) is the diagonal error covariance matrix. Further details can be found in [3], [9].

B. STATE ESTIMATION

Once observability is guaranteed, state estimation can be carried out [28]. One procedure is the Maximum Likelihood method, which estimates the state as \( \hat{x} \), the solution that minimizes the measurement error \( z - h(x) \). That is, \( \hat{x} \) is the solution of:

\[
    \min_x J(x) = \frac{1}{2} [z - h(x)]^T R^{-1} [z - h(x)]
\]

There are two popular methods for determining the solution: 1) NVSE and 2) BCSE.

1) NVSE

NVSE uses the nodal voltage magnitudes and phases as state variables. The optimality condition for Eq. (3) is as follows:

\[
    \frac{\partial J}{\partial x^{NV}} = \left[ H^{NV}(x^{NV}) \right]^T (R^{NV})^{-1} [z^{NV} - h^{NV}(x^{NV})] = 0
\]

where, the superscript NV refers to the choice of state variables and \( H^{NV} = \frac{\partial h^{NV}(x^{NV})}{\partial x^{NV}} \). The venerable Gauss-Newton method for computing the solution of Eq. (4) follows [29]:

\[
    G^{NV}(x_k^{NV}) \Delta x_k^{NV} = \left[ H^{NV}(x_k^{NV}) \right]^T (R^{NV})^{-1} [z^{NV} - h^{NV}(x_k^{NV})]
\]

\[
    G^{NV}(x_k^{NV}) = \left[ H^{NV}(x_k^{NV}) \right]^T (R^{NV})^{-1} H^{NV}(x_k^{NV})
\]

\[
    x_k^{NV+1} = x_k^{NV} + \Delta x_k^{NV}
\]

The converged state estimate from the above iterative method is denoted \( \hat{x}^{NV} \).
2) BCSE

BCSE uses the real and imaginary components of the branch currents as state variables: \( x_{BC} = [I_R \ I_I]^T \). This approach requires converting power measurements \( z \) into equivalent current measurements \( z_{BC} \) (see [28] for details). Using the equivalent current measurements \( z_{BC} \), the measurement function \( h(x) \) becomes linear and the objective function of Eq. (3) simplifies:

\[
J(x_{BC}) = \frac{1}{2}[z_{BC} - H_{BC} x_{BC}]^T (R_{BC})^{-1} [z_{BC} - H_{BC} x_{BC}] 
\]

(8)

The minima can be found directly from the optimality condition \( \frac{\partial}{\partial x} = 0 \):

\[
(H_{BC})^T (R_{BC})^{-1} H_{BC} x_{BC} = (H_{BC})^T (R_{BC})^{-1} z_{BC} 
\]

(9)

In order to find the state estimate \( \hat{x}_{BC} \), the following algorithm is used:

Step 1: Use the node voltage estimates \( u_k \) to convert power flow measurements into current measurements.

Step 2: Solve linear system in Eq. (9) to obtain an estimate of the line currents.

Step 3: Update the node voltage estimates \( u_k \) using the forward sweep procedure. See [30] for details.

Step 4: Check for convergence of node voltage estimates. If converged, stop. Else, return to Step 1.

III. PROPOSED METHOD

A. OBSERVABILITY ANALYSIS FRAMEWORK (OAF)

Fig. 1 shows a schematic of the proposed OAF. First, the vector \( z \) is constructed using all available measurements, pseudo-measurements, virtual measurements, and CMs. Then, the OT described in Section II-A is run. If the system is observable, then either NVSE or BCSE is performed using all the measurements. The converged gain matrix \( G \) from the state estimator is used to compute the OM. If the system is not observable, the observable subnetworks and corresponding measurement sets are identified. For each subnetwork, a slack bus with a voltage measurement is chosen. If a voltage measurement is present, that bus is a natural choice for the slack. Then NVSE or BCSE is performed for each subnetwork. The set of converged gain matrices \( G_i \) is used to compute a set of OMs.

![Figure 1: Observability analysis framework flowchart](image)

The observability metric \( O \) is defined as the condition number of \( G \):

\[
O = \text{cond}(G) = \frac{\lambda_{\text{max}}(G)}{\lambda_{\text{min}}(G)}
\]

(10)

where \( \lambda_{\text{max}}(\cdot) \), \( \lambda_{\text{min}}(\cdot) \) return the largest and smallest eigenvalues of the matrix argument, respectively. \( G \) can be the converged gain matrix from either NVSE or BCSE. Larger values of the dimensionless \( O \) indicate less accurate state estimates. For reference, a condition number exceeding 10\(^{12} \) may yield an inaccurate state estimate [31]. Alternative metrics are provided in [32]. The condition number is preferable because it captures both the “distance” from singularity and the numerical conditioning of the gain matrix. Because sensors are not yet widely deployed in the distribution system, full observability is unlikely. Therefore, we introduce the two types of CMs—CLMs and CDMs—as a means of increasing observability.

B. CORRELATIONAL LOAD MEASUREMENTS (CLMS)

The authors of [33] have proposed a linear regression model for correlated low voltage residential loads. In this section, we propose a similar formulation for CLMs, but for correlated medium voltage loads from the same load class (residential, commercial, industrial, etc.). CLMs are designed for the primary distribution system because of the increased correlation that results from the aggregation of smaller, less predictable loads [34].

Consider a load (power injection) measurement \( z_{i,O}^p, z_{i,O}^q \) at node \( i \) [29]:

\[
\begin{align*}
z_{i,O}^p &= p_i + r_i^p = h_i^p(x) + r_i^p \\
z_{i,O}^q &= q_i + r_i^q = h_i^q(x) + r_i^q
\end{align*}
\]

(11)

where, \( p_i, q_i \) represent the true real and reactive load values, \( h_i^p(x), h_i^q(x) \) are the models of the real and reactive load, and \( r_i^p, r_i^q \sim N(0, \sigma_i^p), r_i^q, q_i^q \sim N(0, \sigma_i^q) \) are the sensor noise. The phase superscript has been dropped for clarity. Let the load at node \( j \) be linearly correlated to the (same phase) load at node \( i \) with constant parameters \( c_p, c_q, d_p, d_q \) as follows:

\[
\begin{align*}
p_j &= c_p p_i + d_p, \quad q_j = c_q q_i + d_q
\end{align*}
\]

(12)

Suppose that the parameters in Eq. (12) are composed of a deterministic part \( c_p^0, c_q^0, d_p^0, d_q^0 \) and a stochastic part \( r_p^c, r_q^c, r_p^d, r_q^d \):

\[
\begin{align*}
c_p &= c_p^0 + r_p^c, \quad c_q = c_q^0 + r_q^c \\
d_p &= d_p^0 + r_p^d, \quad d_q = d_q^0 + r_q^d
\end{align*}
\]

(13)

1) Deterministic part known

In this case, a CLM \( z_{i,O}^p, z_{i,O}^q \) can be constructed from the measurement \( z_{i,O}^p, z_{i,O}^q \) at node \( i \) as follows:

\[
\begin{align*}
z_{j,O}^p &= c_p^{0,p} p_i + d_p^{0,p}, \quad z_{j,O}^q = c_q^{0,q} q_i + d_q^{0,q}
\end{align*}
\]

(15)

In order to use the CLM for SE, it must be written in the form \( \bar{z}_j = h_j(x) + \bar{r}_j \) for some stochastic component \( \bar{r}_j \). It can be shown that \( \bar{r}_j \) has the following form:

\[
\begin{align*}
r_j^p &= c_p^{0,p} p_i - r_p^c p_i - r_p^d q_i - r_q^d q_i \\
r_j^q &= c_q^{0,q} q_i - r_p^c p_i - r_q^d q_i
\end{align*}
\]

(16)

The three sources of uncertainty in the CLM are measurement noise from \( z_i \) and the stochastic components of the constants \( c, d \).
In order to obtain the maximum likelihood state estimate using the CLM for NVSE, the error covariance matrix \( R^{NV} \) must be updated. Assuming that \( r_c \sim \mathcal{N}(0, \sigma_c^2) \) and \( r_d \sim \mathcal{N}(0, \sigma_d^2) \), the nonzero entries in the rows and columns of \( R^{NV} \) corresponding to the measurements \( z^p_j, z^q_j \) are given by the following:

\[
R_{ij}^{NV} = \begin{bmatrix} \mathbb{E}[r_i^p r_j^p] & \mathbb{E}[r_i^p r_j^q] \\ \mathbb{E}[r_i^q r_j^p] & \mathbb{E}[r_i^q r_j^q] \end{bmatrix} = \begin{bmatrix} \sigma_i^2 & \sigma_i \sigma_j \\ \sigma_j \sigma_i & \sigma_j^2 \end{bmatrix}
\]  

(17)

\[
\hat{\sigma}^2 = (c_0^p \sigma_i)^2 + [h_i^j(x)]^2 \sigma_j^2 + \sigma_d^2
\]  

(18)

where, the superscript NV on \( h_i^j(x) \) has been dropped for brevity. The zero-mean assumption for the \( r_c \) and \( r_d \) terms follows from the assumption of known deterministic components \( c_0, d_0 \). The entries in \( R^{NV} \) corresponding to reactive power measurements \( z^q_j, z^q_j \) are identical. See Section III-D below for the details on incorporating CLMs in NVSE.

For BCSE, there is an alternative approach: the CLM can be represented as a linear equality constraint using the node voltage estimates. Under the assumption that \( c_0^p = c_0^q \), it can be shown that Eq. (12) can be rewritten as follows:

\[
I^p_j = c_0 \left[ I^p_j (u_j^p u_j^p + u_j^q u_j^q) + I^q_j (u_j^p u_j^q - u_j^q u_j^p) \right] / (u_j^p)^2 + (u_j^q)^2 + d_0^p u_j^p + d_0^q u_j^q
\]  

(19)

\[
I^q_j = c_0 \left[ I^p_j (u_j^p u_j^p - u_j^q u_j^q) + I^q_j (u_j^p u_j^q + u_j^q u_j^p) \right] / (u_j^p)^2 + (u_j^q)^2 + d_0^p u_j^p - d_0^q u_j^q
\]  

(20)

This assumption can be regarded as the loads at nodes \( i \) and \( j \) having the same power factor. This is reasonable considering that the loads belong to the same class, and the power factor is maintained by various control mechanisms such as volt-var control. See Section III-D below for the details on incorporating the above equality constraints in BCSE.

2) Deterministic part unknown

In this more realistic case, we use the estimates of the parameters \( \hat{c}^p, \hat{c}^q, \hat{d}^p, \hat{d}^q \) to construct the CLM \( z_i^p, z_i^q \):

\[
z_i^p = \hat{c}^p z_i^p + \hat{d}^p, \quad z_i^q = \hat{c}^q z_i^q + \hat{d}^q
\]  

(21)

For NVSE, the \( R^{NV} \) matrix should be updated according to Eq. (17) replacing \( c_0, d_0 \) with \( \hat{c}, \hat{d} \) respectively. Similarly for BCSE, the equality constraints shown in Eqs. (19,20) should use the estimates \( \hat{c}, \hat{d} \).

If we have access to even sporadic measurements \( z_i^p \) of the load at node \( j \), we can use them to improve our parameter estimates for CLM on-line. These sporadic measurements could be the result of infrequent manual monitoring by technicians. Using vector notation and omitting the identical equations for reactive power, we have:

\[
z_i^p = (w^p)^T \eta^p + r_i^p, \quad z_i^q = (w^q)^T \eta^p
\]  

(22)

\[
w^p = [p_i, 1]^T, \quad w^q = [z_i^p, 1]^T
\]  

(23)

\[
\eta^p = [c^p, d^p]^T, \quad \eta^q = [c^q, d^q]^T
\]  

(24)

Under the assumption that \( r_c \sim \mathcal{N}(0, \sigma_c^2) \) and \( r_d \sim \mathcal{N}(0, \sigma_d^2) \), we propose to improve our parameter estimates \( \hat{\eta} \) using a least squares approach [35]:

\[
\hat{\eta}(k) = \left\{ \left[ \hat{W}^p(k) \right]^T \hat{W}^p(k) \right\}^{-1} \left[ \hat{W}^p(k) \right]^T z_i^p(k)
\]  

(25)

\[
\hat{W}^p(k) = \left[ \hat{w}^p(1) \cdots \hat{w}^p(N) \right]^T
\]  

(26)

where, \( k \) is the number of available measurements in \( z_i^p \). As each new measurement \( z_i^p(k + 1) \) comes in, Eq. (25) can be used as a fixed-memory estimator to update the parameter estimates. Because the sensor noise \( r_j \) is zero mean and \( \hat{W}, r_j \) are statistically independent, the above estimator is unbiased [35].

C. CORRELATIONAL DER MEASUREMENTS (CDMS)

This section presents the formulation of CDMs, the second type of CMs proposed in this work. CDMs could be used to represent nearby PV units that are correlated based on orientation and tilt angle or storage units that are correlated based on state of charge. In this paper, we utilize CDMs to capture the correlation between wind turbines that are located close together geographically. The correlation coefficient \( \rho_{ij} \) between the generation \( p_i, p_j \) of two wind turbines can be modeled as an exponential function of the distance \( \delta_{ij} \) between them [25]:

\[
\rho_{ij} = \exp \left[ -\frac{\delta_{ij}}{\zeta_1} \right]
\]  

(27)

where \( \zeta_1 = 20 \) miles and \( \zeta_2 = 2 \) are constant parameters. \( \zeta_1 \) is the average distances between the DER units and the measurement point and \( \zeta_2 \) is an exponential term. These selections can be changed if the generated CDN are not accurate enough. The correlation coefficient takes values between 0 and 1 and measures the linearity of the relationship between the generation of the two DERs. The authors propose the following CDN formulation combining \( \rho_{ij} \) with the CLM formulation in Section III-B. Suppose the generation \( p_j, q_j \) of an unmeasured DER is correlated to the generation \( p_i, q_i \) of a measured DER with an uncorrelated component \( \tilde{p}_j, \tilde{q}_j \):

\[
p_j = \rho_{ij}(c^p p_i + d^p) + (1 - \rho_{ij}) \tilde{p}_j
\]  

\[
q_j = \rho_{ij}(c^q q_i + d^q) + (1 - \rho_{ij}) \tilde{q}_j
\]  

(28)

where, as before \( c^p, d^p, c^q, d^q \) are composed of a deterministic part and a stochastic part (see Eqs. (13,14)). Let us also suppose that the uncorrelated component \( \tilde{p}_j, \tilde{q}_j \) is a parameter composed of a deterministic part \( \tilde{p}_{j0}, \tilde{q}_{j0} \) and a stochastic part \( \tilde{p}_{j}^p, \tilde{p}_{j}^q \):

\[
\tilde{p}_j = \tilde{p}_{j0} + \tilde{p}_{j}^p, \quad \tilde{q}_j = \tilde{q}_{j0} + \tilde{q}_{j}^p
\]  

(29)

where \( \tilde{p}_{j}^p \sim \mathcal{N}(0, \delta_{ij}^2), \tilde{p}_{j}^q \sim \mathcal{N}(0, \delta_{ij}^2) \). We consider two cases below.
1) Deterministic part known

If the deterministic parts of all the parameters in Eq. (28) are known, then we can construct a CDM \( z_P^j, z_Q^j \) as follows:

\[
\begin{align*}
    z_P^j &= \rho_j(c_0^j z_P^j + d_0^j) + (1 - \rho_j)\hat{p}_j \\
    z_Q^j &= \rho_j(c_0^j z_Q^j + d_0^j) + (1 - \rho_j)\hat{q}_j
\end{align*}
\]  

where we have set the \( d \) parameter in Eq. (30) to zero and replaced the remaining parameters’ known deterministic parts (subscripted 0) with estimates (denoted by the hat notation).

For NVSE, the \( R_N^V \) matrix should also be updated according to Eq. (32) replacing the parameters with subscript 0 with their estimates and eliminating the \( \sigma_d^2 \) term. Similarly for BCSE, the equality constraints shown in Eqs. (34,35) should be modified.

Writing Eq. (28) with \( d \) set to zero in vector notation and the CDM in Eq. (36) in vector notation yields the same result as before (Eq. (22)). However, the vectors \( w, \eta \) have different contents. Omitting the identical equations for reactive power, we have:

\[
\begin{align*}
    \hat{w}^p &= [\rho_j p_i (1 - \rho_j)]^T, \\
    \hat{w}^q &= [\rho_j z_P^p (1 - \rho_j)]^T \\
    \hat{\eta}^p &= [c^p \hat{p}_j]^T, \quad \hat{\eta}^q = [c^p \hat{q}_j]^T
\end{align*}
\]

Under the zero-mean Gaussian assumptions on the stochastic components of the parameters, we can improve our parameter estimates \( \eta \) using the same least squares approach in Eqs. (25,26).

D. STATE ESTIMATION WITH CMS

For NVSE, \( R_N^V \) is no longer diagonal and is now a function of the state \( x \) as a result of the CMs. The NVSE problem now has the following form:

\[
\min_x J(x) = \frac{1}{2} [x - h(x)]^T R^{-1} (x) [x - h(x)]
\]  

where the superscript NV has been omitted for brevity. The entries in \( R_N^V \) corresponding to reactive power measurements \( z_Q^j, z_Q^j \) are identical.

For BCSE, the CDM can be represented as a linear equality constraint using the node voltage estimates. Under the same power factor assumption as for a CLM \( (c_0^j = c_0^q = c_0) \), it can be shown that Eq. (28) can be rewritten as follows:

\[
\begin{align*}
    I_1^j &= \frac{c_0 \rho_j}{2} \left[ I_1^j (u_1^j u_2^j + u_2^j u_1^j) + I_1^j (u_2^j u_1^j - u_1^j u_2^j) \right] \\
    &\quad + \frac{\rho_j}{2} \left[ (d_0^j u_2^j + d_0^j u_1^j) + (1 - \rho_j) (\hat{p}_j u_2^j + \hat{q}_j u_3^j) \right] \\
    &\quad + \frac{\rho_j}{2} \left[ (d_0^j u_2^j - d_0^j u_1^j) + (1 - \rho_j) (\hat{p}_j u_2^j - \hat{q}_j u_3^j) \right]
\end{align*}
\]

The assumption is reasonable for CDMs because most DERs are interfaced to the grid via inverters using a constant power factor mode [36]. See Section III-D below for the details on incorporating CDMs into either NVSE or BCSE.

2) Deterministic part unknown

A CDM contains three parameters, but the coefficients of \( d, \hat{p}, \hat{q} \) are constants. Only the coefficient of \( c \) is time-varying. As a result, only two parameters can be estimated on-line using sporadic measurements. Therefore, we build a CDM of the following form:

\[
\begin{align*}
    z_P^j &= \rho_j c^p z_P^j + (1 - \rho_j)\hat{p}_j \\
    z_Q^j &= \rho_j c^q z_Q^j + (1 - \rho_j)\hat{q}_j
\end{align*}
\]  

where \( \rho_j \) is the correlation coefficient between the state \( x \) and the measurement \( z \) and \( c^p, c^q \) are the deterministic parts of the parameters.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
FIGURE 2: Modified IEEE 123 node test feeder topology. Black line segments correspond to \(3\phi\) lines, whereas colored segments represent \(1\phi\) or \(2\phi\). Green boxes represent \(3\phi\) power flow meters, green circles represent \(3\phi\) power injection meters, and blue x’s represent a lack of historical data. DERs are represented as large black circles with a tilde inside.

A. OBSERVABILITY ANALYSIS AND STATE ESTIMATION WITHOUT CMS

Without the historical data at the indicated nodes, the two islands are not fully observable. As a result, the OIIP removes the unobservable nodes and branches to perform SE of the remaining networks. Pruning the network to remove unobservable branches allows SE to proceed; however, it introduces topology error compared to the true system. Fig. 3 displays the NVSE and BCSE results for the observable nodes in the system.

Table 1 contains the OM values and corresponding SE error. As mentioned previously, OM values exceeding \(10^{12}\) could indicate poor state estimation accuracy. It is known that NVSE is susceptible to ill-conditioning as system size increases; this is shown in the OM values of Island 2. However, the OM values from NVSE should not be compared to BCSE because the gain matrices have completely different forms. For both estimators, Island 2 has a higher metric value than Island 1, and the SE error is likewise higher. This can be attributed to the lower ratio of measurements to nodes in Island 2.

B. OBSERVABILITY ANALYSIS AND STATE ESTIMATION WITH CMS

In this section, we achieve full observability of both islands in Fig. 2 by leveraging CLMs and CDMs. Each load node without historical data is given a CLM based on the nearest (same-phase) load node with a pseudo-measurement. The unmeasured DERs at nodes 16 and 64 are given CDMs based on the measured DERs at nodes 29 and 63, respectively. The DERs in Island 1 are assumed to be 10 miles apart geographically; the DERs in Island 2 are assumed to be 1 mile apart. The error associated with the CMs’ parameters is zero mean with a standard deviation of 10% of the true value. Fig. 4 displays the NVSE and BCSE results. Note that with CMs, both NVSE and BCSE algorithms are able to estimate the states for all nodes in the test system. This is because, the proposed CMs have improved the observability compared to the base case. Regarding the implementation, occasionally the error covariance matrix \(R\) can become singular for NVSE. This was handled by replacing \(R\) with a “damped” version: \(R' = R + \lambda \text{diag}(R)\) where \(\lambda\) is chosen to be as small as possible but large enough to ensure invertibility. This approach is based on the Levenberg-Marquardt algorithm [38].

Table 1 contains the OM values and corresponding SE error. Island 2 once again has a higher metric value than Island 1, and the SE error is likewise higher. As before, this can be attributed to the lower ratio of measurements to nodes in Island 2. Compared to the results without CMs, it can be seen that the NVSE error increases due to the relatively high uncertainty associated with the CMs. On the other hand, BCSE’s error decreases because the truncated system topology resulted in greater error than the CM constraints. The OM does not reflect this improved accuracy because the OM does not capture the topology error introduced by...
truncating the system. If the baseline case did not have CMs and used the correct topology, then the OM would remain at $3 \times 10^7$, and the SE error would decrease to less than 0.6%.

C. ON-LINE PARAMETER AND STATE ESTIMATION

This section presents the simulation results for state and parameter estimation with CMs. This is how the authors actually envision CMs being used because the DSO is unlikely to have accurate estimates of the CM parameters. The same modified IEEE 123 Node Test Feeder in Fig. 2 was utilized. Random load data was simulated to generate time series data. Each snapshot was used to perform NVSE and BCSE. The CMs for both estimators were initialized with gross parameter errors. However, it is assumed that sporadic measurements of the nodes with CMs are available. The measurements are used to update the parameter estimates online using the least squares approach presented in Sections

| TABLE 1: Observability metric and state estimation error of NVSE and BCSE with and without CMs |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | NVSE            | BCSE            |                |                |
|                | OM | SE Err. | OM | SE Err. |                |                |
| Without CMs   | Island 1 | $1 \times 10^7$ | 0.1% | $3 \times 10^7$ | 0.2% |                |                |
|                | Island 2 | $4 \times 10^6$ | 1.0% | $3 \times 10^6$ | 1.2% |                |                |
| With CMs      | Island 1 | $2 \times 10^7$ | 0.1% | $3 \times 10^7$ | 0.2% |                |                |
|                | Island 2 | $3 \times 10^8$ | 2.0% | $3 \times 10^8$ | 0.6% |                |                |
III-B2 and III-C2.

Fig. 5 displays the time evolution of the OM for both estimators during the same simulation. At integer values of time, OM values are the smallest due to the availability of real measurements of the nodes with CMs. Island 2 has higher OM values than Island 1 for both estimators corresponding to the lower state estimation accuracy shown in Fig. 5.

D. VALIDATION OF PROPOSED CMS USING PECAN STREET DATA

As demonstrated in the results above, the use of CMs has the potential to significantly increase the observability of distribution systems. To validate that the model of CMs proposed in Section III, we use the Pecan street dataset. The Pecan street dataset is the “the largest source of disaggregated customer energy data”, as described on their website. Consumer load data and solar generation for a distribution system feeder in Austin, TX are used to derive the correlational measurements. The Pecan street dataset does not provide GIS coordinates or any other geographical identifiers, so it is assumed that sequential node numbers are adjacent. This is inline with the description provided in the dataset. To demonstrate the validity of the CM model, a “source” node \( i \) is used to estimate the measurements for the adjacent node \( j \). This exercise is carried for measurements over one week. The Fig. 6 shows the measurements for node \( i \), node \( j \), and CM for CLM and CDM cases. It can be seen that the CM model closely follows the actual node \( j \) measurement, and the RMSE was 0.8418 and 0.1976 for the CLM and CDM cases respectively. It is important to note that the notion of geographical distance between the DER units modeled in Eqn. 29 has not been used for this validation as the relevant data is not available in the dataset. The RMSE can be reduced further if the correlation model can be enhanced with infrequent measurements from the load, and geographical correlation for the DERs.

E. CONFIDENCE INTERVALS FOR STATE ESTIMATION

To ensure that the state estimation augmented with CMs can be used by the operators, confidence intervals are constructed for the states estimated using CMs. The confidence intervals are constructed using well established methods in statistics, and following the work of authors in [39]. As detailed in Section II-B, the state estimation can be stated as the MLE that estimates the state \( \hat{x} \), the solution that minimizes the measurement error \( z - h(x) \). The residual vector from the MLE \( r_i \) is used to calculate the mean and variance as follows:

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} r_i; \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (r_i - \mu)^2
\]

The value of the objective function follows the \( \chi^2 \)-squared distribution, and the \( \chi^2 \)-squared table is used to obtain the confidence interval for the parameters estimated. The confidence interval is shown in Eqn. 43.

\[
\hat{x}_j - \Delta \sqrt{\sigma^2 C_{jj}} \leq x_j \leq \hat{x}_j + \Delta \sqrt{\sigma^2 C_{jj}}
\]

In Eqn. 43, \( x \) is the vector of states, \( C_{jj} \) is the \( j^{th} \) diagonal element of \( (H^T H)^{-1} \), and \( \Delta \) is defined as follows -

\[
\Delta = \chi^2_{p, n-p}
\]

where, \( \chi^2 \) refers to the Chi-squared distribution, \( p \) is the number of estimated states, and \( n \) is the total number of measurements used. The confidence intervals for the state estimated using CMs is shown in Fig. 7. The confidence interval allows the system operator to quickly decide if the state estimates can be used for further analysis. The measurements that do not pass the chi-squared test are removed from the state estimation.

V. CONCLUSIONS

In response to the lack of observability of distribution systems, this paper presents the concept of correlational measurements (CMs) and an observability metric (OM). CMs can be used by system operators to leverage known correlation between loads and between DERs in the system. In order to use CMs to obtain a maximum likelihood state estimate, the required modifications to the classical NVSE and BCSE algorithms are derived. The proposed OM quantifies state estimation accuracy. Simulation results evidence the value
of CMs in improving both observability of the system and the OM/state estimation accuracy. These improvements have the potential to improve distribution system resiliency by enabling DER-based control.

REFERENCES

[1] C. Chen, J. Wang, and D. Ton, “Modernizing distribution system restoration to achieve grid resiliency against extreme weather events: An integrated solution,” Proc. of the IEEE, July 2017.

[2] H. Gharavi and R. Ghaifurian, “Smart grid: The electric energy system of the future.” Proc. of the IEEE, June 2011.

[3] F. Wu and A. Monticelli, “Network observability: Theory,” IEEE Trans. Power Syst., 1985.

[4] F. Wu, “Power system state estimation: a survey,” Intl. Jour. Electr. Power & Energy Syst., 1990.

[5] M. Shahraeini and M. H. Javidi, “A survey on topological observability of power systems,” in 2011 IEEE Power Engineering and Automation Conference, Sep. 2011.

[6] R. R. Nucera and M. L. Gilles, “Observability analysis: a new topological algorithm,” IEEE Trans. Power Syst., May 1991.

[7] G. Tran, A. Kiani, A. Annaswamy, Y. Sharon, A. L. Motto, and A. Chakraborty, “Necessary and sufficient conditions for observability in power systems,” in 2012 IEEE PES Innov. Smart Grid Tech., Jan 2012.

[8] A. Monticelli, “Electric power system state estimation,” Proc. of the IEEE, Feb 2000.

[9] B. Gou and A. Abur, “A direct numerical method for observability analysis,” IEEE Trans. Power Syst., 2000.

[10] E. Castillo, A. J. Conejo, R. E. Pruneda, and C. Solares, “Observability analysis in state estimation: a unified numerical approach,” IEEE Trans. Power Syst., May 2006.

[11] R. Gelagaev, P. Vermeyen, J. Vandewalle, and J. Driesen, “Numerical observability analysis of distribution systems,” in Proc. 14th Intl. Conf. Harm. Qual. Power, Sep. 2010.

[12] O. Alsac, N. Vempati, B. Stott, and A. Monticelli, “Generalized state estimation [power systems],” in Proc. 20th Intl. Conf. Power Ind. Comp. Apps., May 1997.

[13] G. N. Korres, P. J. Katsikas, K. A. Clements, and P. W. Davis, “Numerical observability analysis based on network graph theory,” IEEE Trans. Power Syst., Aug 2003.

[14] G. N. Korres and P. J. Katsikas, “A hybrid method for observability analysis using a reduced network graph theory,” IEEE Trans. Power Syst., Feb 2003.

[15] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Enhancing observability in distribution grids using smart meter data,” 2016.

[16] B. Brinkmann and M. Nagnevitsky, “A probabilistic approach to observability of distribution networks,” IEEE Transactions on Power Systems, March 2017.

[17] A. Angioni, T. Schlösser, F. Ponci, and A. Monti, “Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids,” IEEE Transactions on Instrumentation and Measurement, Jan 2016.

[18] K. Clements, “Observability methods and optimal meter placement,” Int. Jour. Electr. Power & Energy Syst., 1990.

[19] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, “A survey on state estimation techniques and challenges in smart distribution systems,” IEEE Transactions on Smart Grid, March 2019.

[20] E. Caro, A. J. Conejo, and R. Minguiz, “Power system state estimation considering measurement dependencies,” IEEE Trans. Power Syst., Nov 2009.

[21] Q. Chen, D. Kaleshi, Z. Fan, and S. Armour, “Impact of smart metering data aggregation on distribution system state estimation,” IEEE Trans. Ind. Informat., Aug 2016.

[22] G. Valverde, A. T. Saric, and V. Terzija, “Stochastic monitoring of distribution networks including correlated input variables,” IEEE Trans. Power Syst., Feb 2013.

[23] J. M. Morales, L. Baringo, A. J. Conejo, and R. Minguiz, “Probabilistic power flow with correlated wind sources,” IET Gen., Trans. Dist., May 2010.

[24] A. K. Ghosh, D. L. Lubbkeman, and R. H. Jones, “Load modeling for distribution circuit state estimation,” IEEE Trans. Power Deliv., April 1997.

[25] S. M. S. Alam, B. Natarajan, and A. Pahwa, “Distribution grid state estimation from compressed measurements,” IEEE Trans. Smart Grid, 2014.

[26] G. Wang, G. B. Giannakis, J. Chen, and J. Sun, “Distribution system state estimation: an overview of recent developments,” Frontiers of Information Technology & Electronic Engineering, Jan 2019.

[27] Y. Zhang, A. Bernstein, A. Schmitt, and R. Yang, “State estimation in low-observable distribution systems using matrix completion,” National Renewable Energy Lab. (NREL), Golden, CO (United States), Tech. Rep., 2019.

[28] M. E. Baran and A. W. Kelley, “A branch-current-based state estimation method for distribution systems,” IEEE Trans. Power Syst., 1995.

[29] F. C. Schweppe and J. Wildes, “Power system static-state estimation, part I: Exact model,” IEEE Trans. Power Syst., 1970.

[30] C. S. Cheng and D. Shirnhammad, “A three-phase power flow method for real-time distribution system analysis,” IEEE Trans. Power Syst., May 1995.

[31] R. Ebrahimian and R. Balick, “State estimator condition number analysis,” IEEE Trans. Power Syst., May 2001.

[32] R. Pasqualetti, S. Zampieri, and F. Bullo, “Controllability metrics, limitations and algorithms for complex networks,” IEEE Trans. Contr. Netw. Syst., March 2014.

[33] Y. R. Gahrooei, A. Khodabakhshian, and R. Hooshmand, “A new pseudo load profile determination approach in low voltage distribution networks,” IEEE Trans. Power Syst., Jan 2018.

[34] M. Shafiei, G. Nourbaksh, A. Arefi, G. Ledwich, and H. Pezeshki, “Single iteration conditional based dsse considering spatial and temporal correlation,” Intl. Jour. Electr. Power and Energy Syst., vol. 107, 10 2017.

[35] J. M. Mendel, Lessons in Digital Estimation Theory. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.

[36] M. H. J. Bollen and A. Mannino, “Voltage control with inverter-based distributed generation,” IEEE Trans. Power Deliv., Jan 2005.

[37] G. N. Korres and N. M. Manousakis, “State estimation and bad data processing for systems including pmu and scada measurements,” Electric Power Systems Research, vol. 81, no. 7, pp. 1514–1524, 2011.

[38] J. Nocedal and S. Wright, Numerical Optimization. Springer Science & Business Media, 2006.

[39] E. Kyriakides and G. T. Heydt, “Calculating confidence intervals in parameter estimation: a case study,” IEEE Trans. on Pow. Del., 2006.