IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE POWER & ENERGY SOCIETY SECTION

Received 10 September 2023, accepted 3 October 2023, date of publication 10 October 2023, date of current version 25 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3323604

RESEARCH ARTICLE

Multiview Spectral Clustering of High-Dimensional Observational Data

A. ROMÁN-MESSINA[1], (Life Fellow, IEEE), CLAUDIA M. CASTRO-ARVIZU[2], (Member, IEEE), ALEJANDRO CASTILLO-TAPIA[1], ERLAN R. MURILLO-AGUIRRE[1], AND O. RODRÍGUEZ-VILLALÓN[3]

[1]Center for Research and Advanced Studies (CINVESTAV), National Polytechnic Institute, Zapopan 45017, Mexico
[2]Department of Design, Science, and Technology, Autonomous University of Guadalajara, Zapopan 45129, Mexico
[3]DICIS, University of Guanajuato, Salamanca, Guanajuato 36885, Mexico

Corresponding author: A. Román-Messina (arturo.roman@cinvestav.mx)

ABSTRACT The joint analysis of large-scale datasets is crucial when studying complex processes involving diverse sensing sources and multiple variables. This paper proposes a multiview nonlinear manifold learning framework to fuse or combine data from different types of measurements. Spectral clustering techniques are employed to obtain a low-dimensional system representation, where the physical data are projected onto a low-dimensional Euclidean space that preserves the intrinsic geometry of the data. The theoretical properties of various multiview diffusion maps are examined, and algorithms for the efficient computation of multiview kernel representations are outlined. Measures of similarity are also derived, and the results are compared with other state-of-the-art methods for model reduction. Finally, multiple datasets obtained from transient stability simulations of a large-scale power system model are utilized to evaluate the effectiveness of the developed algorithms, thereby illustrating their superiority over other state-of-the-art multiview clustering approaches.

INDEX TERMS Alternating diffusion maps, multiview data fusion, spectral clustering.

I. INTRODUCTION

The study of high-dimensional datasets, which may be multimodal or heterogeneous, is becoming increasingly significant in the analysis of large-scale datasets collected by wide-area monitoring systems (WAMS) [1], [2]. With the continuous growth in the dimension and complexity of observed data, developing efficient techniques for combining or fusing large-scale sets of measurements from different types or modalities is essential.

Such approaches can aid in discovering key association patterns among different data modalities or uncover features or relations that may need to be evident from analyzing a single dataset. These approaches offer valuable insights by identifying spatial patterns and extracting relevant dynamical features using fewer variables [3].

Various tensor-based approximations have recently been employed to fuse and analyze multichannel data [3], [4], [5], [6], [7]. Integrating or combining multimodal information from an observed event captured by different sensors aids in eliminating redundant data and proves beneficial for tasks such as feature extraction, classification, state reconstruction, and data visualization.

While there has been relative success in developing and applying these formulations, the joint analysis of multiple datasets still needs to be solved due to the high dimensionality of the data and measurement redundancy. This challenge arises because multimodal data may exhibit high correlation, possess different units, or have varying structures [8], [9].

Tensor factorization-based models have emerged as effective methods for fusing and clustering large datasets. These techniques prove particularly beneficial in clustering and visualizing data across multiple high-dimensional sets, often encompassing different types or modalities. Noteworthy among these methods are Tucker decomposition [10], [11], sequentially truncated higher-order singular value decomposition (ST-HOSVD) [12], and hybrid approaches that combine dynamic mode decomposition (DMD) with graph Laplacian methods [13].

The associate editor coordinating the review of this manuscript and approving it for publication was Jethro Browell.

These techniques have been proposed and employed to extract meaningful insights from extensive datasets efficiently. Markov chain-based multiview clustering methods also efficiently merge multiple datasets and generate low-dimensional representations [14], [15], [16].

However, inherent limitations, such as computational complexity and scalability issues, persist in these approaches when modeling large-scale multimodal data with more than two data sets.

This paper introduces a multiview spectral clustering technique that achieves adaptive dimensionality reduction and identifies collective variables governing the long-term motion of the observed data. This framework extends previous work on multiview spectral clustering [7] in three primary directions. First, it generalizes the single-view analysis of measured data, allowing for the simultaneous examination of multiple single-type data from various fault scenarios or multimodal data collected using the same set of sensors. Second, it introduces a robust analytical framework that leverages cross-diffusion information across data modalities or sensor channels. Third, the paper presents graphical methods for visualizing and analyzing multi-view data.

Finally, in practical applications, numerical results obtained from large-scale simulations of a practical power grid are used to combine and extract feature patterns. Additionally, these results help determine and classify diffusion distances within the low-dimensional embedding space, a dimension that is often overlooked in power system literature.

II. BACKGROUND

Modern Synchrophasor Measurement Units (PMUs) produce data as tensors or multidimensional arrays. To establish a common analysis framework, consider a WAMS that collects data from M interconnected regional systems indexed $\{1,2,\ldots,M\}$; each assumed to include a phasor data concentrator (PDC).

Now suppose the data collected by the PDCs correspond to different faults or analysis scenarios, f, represented by a two-dimensional array of the form $\mathbf{X}_f^l \in \Re^{N\times m} = [\,\mathbf{x}_1^l\ \mathbf{x}_2^l\ \cdots\ \mathbf{x}_m^l\,]$, with $\mathbf{x}_j^l = [\,x_j(t_1)\ x_j(t_2)\ \cdots\ x_j(t_N)\,]^T \in \Re^N$, $j = 1,\ldots,m$, $l = 1,\ldots,F$, where $N$is the number of samples, and m is the total number of sensors. F is the number of fault scenarios or types of data measurement considered in the study.

As depicted in Fig. 1, two problems are of interest here: (a) the joint analysis of different data types, such as bus voltage magnitudes and phases, and active and reactive power obtained from a single event, and b) the simultaneous study of multiple datasets of the same type acquired from various events. The latter case includes historical records associated with numerous contingency scenarios or extensive single-type datasets acquired by regional PDCs, while the former includes multichannel data recorded by PMUs. In both

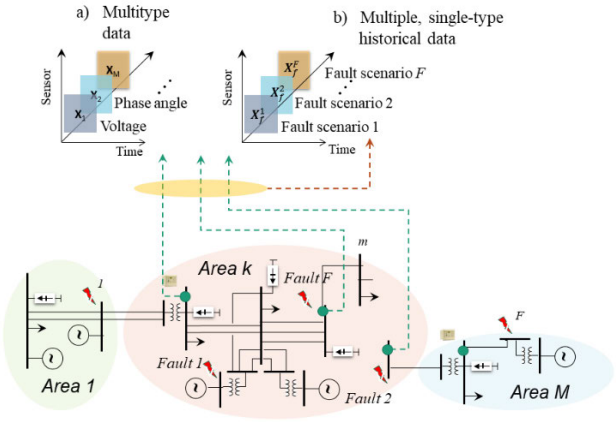


FIGURE 1. Illustration of tensor-based representation of multichannel data. a) Multimodal data, b) Historical data; f denotes the feature data selected for analysis, e.g., bus voltage magnitudes and phases.

instances, various datasets can be effectively represented as a data hypercube (see Fig. 1).

The first coordinate represents time, the second corresponds to sensor location, and each slice along the third coordinate corresponds to a specific data type or fault scenario. The first problem has been examined in [3], while the second is addressed here.

Traditional multiblock analysis techniques, such as multi-block principal component analysis and other conventional approaches, require concatenating the individual datasets (blocks) $\mathbf{X}_f^l$ into a single matrix or tensor. This combined data structure is then used to analyze the overall model using single-view analysis techniques [3], [7]. However, the simple aggregation of datasets fails to capture three-way interactions – it does not leverage the complementary nature of multichannel, multimodal data and may be impractical for the joint analysis of high-dimensional datasets.

Other, more advanced techniques include multiblock canonical correlation analysis and Tucker-based tensor decomposition, to mention just a few approaches [17], [18].

III. TENSOR-BASED MULTIBLOCK REPRESENTATIONS

In many existing data fusion approaches, multiview data can be effectively encoded in an $N\times m\times F$ tensor $\boldsymbol{\mathcal{X}}^l$ (a third-order array) in the following form,

$$\boldsymbol{\mathcal{X}}^l(:,:,l) \in \Re^{N\times m\times F} = \mathbf{X}_f^l\ ,\ l = 1,\ldots,F \qquad (1)$$

Here, the feature matrices $\mathbf{X}_f^l$ are regarded as slices of the data hypercube, as shown in Fig. 1. As previously mentioned, N represents the number of collected samples, m represents the number of sensors or measurement points, and F might denote multimodal or historical data. Tensor-based methods offer greater flexibility in storing and analyzing multi-block data, resulting in concise spatiotemporal representations. Building on the concepts introduced in the ST-HOSVD in [7], the third-order tensor $\boldsymbol{\mathcal{X}} \in \Re^{I_1\times I_2\times I_3}$ is decomposed into three-factor matrices and a core tensor,

as detailed in [7] and [19]:

$$\hat{\mathcal{X}} = \boldsymbol{S} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)} \tag{2}$$

where $\times_n$ indicates the n-mode tensor multiplication $\boldsymbol{U}^{(i)} = \left[\boldsymbol{u}_1^{(i)}\ \boldsymbol{u}_2^{(i)}\ \ldots\ \boldsymbol{u}_{I_i}^{(i)}\right] \in \Re^{I_i \times I_i}$, i=1,2,3 represents the orthogonal (factor) matrices that form the Tucker factors $(\boldsymbol{U}^{(i)})^{\mathrm{T}}\boldsymbol{U}^{(i)} = \boldsymbol{I}_i$, and $\boldsymbol{S}$ is an all-orthogonal core tensor.

This core tensor contains the ith mode singular values of $\hat{\mathcal{X}}$ indicating the extent of interaction among the $\boldsymbol{U}^{(i)}$ components. It is straightforward to confirm that (2) allows for a spatiotemporal representation in the form [7]:

$$\hat{\mathcal{X}} = \underbrace{(\hat{\boldsymbol{S}} \times_2 \hat{\boldsymbol{U}}^{(2)} \times_3 \hat{\boldsymbol{U}}^{(3)})}_{\boldsymbol{G}} \times_1 \hat{\boldsymbol{U}}^{(1)} = \boldsymbol{G} \times_1 \hat{\boldsymbol{U}}^{(1)} \tag{3}$$

In this equation, $\hat{\boldsymbol{U}}^{(1)} = \left[\hat{\boldsymbol{u}}_1^{(1)}\ \hat{\boldsymbol{u}}_2^{(1)}\ \ldots\ \hat{\boldsymbol{u}}_{r_1}^{(1)}\right]$ provides a common basis for the subspace of measurements linked with (1). The tensor $\boldsymbol{G} = \hat{\boldsymbol{S}} \times_2 \hat{\boldsymbol{U}}^{(2)} \times_3 \hat{\boldsymbol{U}}^{(3)} \in \Re^{r_1 \times N \times m_k}$ contains the temporal patterns or modes of each subspace of measurements.

Physically, $\boldsymbol{U}^{(i)}$ is the principal component of each tensor mode, and $\boldsymbol{U}^{(1)}$ represents the spatial shape or pattern. This framework facilitates the derivation of measures similar to those introduced in conventional small-signal or Proper Orthogonal Decomposition (POD) analysis [20].

IV. MARKOV CHAIN MULTIVIEW SPECTRAL CLUSTERING

Spectral analysis of graph-based multiview kernel matrices offers a powerful analytical tool for efficiently computing higher-order models of high-dimensional data sets. The core concept of such approaches involves the spectral decomposition of a squared symmetric feature matrix and the determination of a diffusion operator.

A. ALTERNATING DIFFUSION MAPS

Alternating diffusion (AD) maps have emerged as a technique for non-linearly reducing large datasets [21], [22].

Consider multiple sets of observational data $\mathbf{X}_f^l \in \Re^{N \times m} = [\,\mathbf{x}_1^l\ \mathbf{x}_2^l \cdots \mathbf{x}_m^l\,]$, l=1,...,F, each represented as a slice in the data hypercube in Fig. 1. In this context, each column signifies the time evolution of a point or sensor, denoted as x_i with subscript f associated with the fault scenario or data type. Generally, it is assumed that the $\mathbf{x}_f^l$ can vary in type or modality. Multiview DMs aim to derive a low-dimensional representation from which the dominant variables or states governing the system's slow response can be identified. This section presents a generalization of the DM framework introduced in [17].

Now, assume that each dataset $\mathbf{X}_f^l$ is modeled as an (undirected) weighted graph $G = (V, E)$. The vertex set $\{V_i\}_{i=1}^m$ comprises measurement points (nodes) or sensor locations $\{x_1\ x_2 \cdots x_m\}$, and the set of graph edges, denoted as E, represents pairwise distances or transitions [3], [23].

These graph nodes can also be perceived as data points in the low-dimensional space, as suggested in Fig. 2. The edges

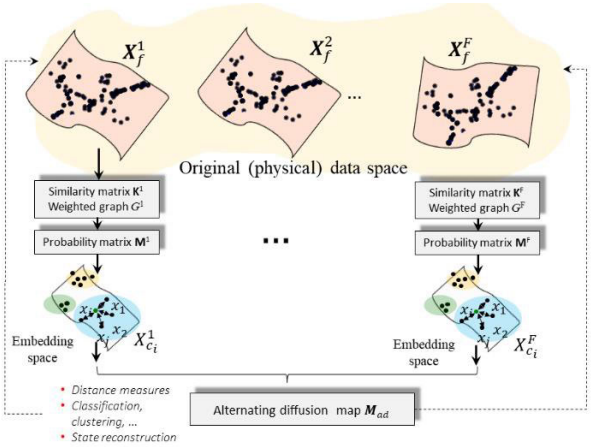


FIGURE 2. Diagram illustrating multiview analysis for combining high-dimensional data. Each data set, represented as $\mathbf{X}_f^l$, produces a distinct probability matrix, $\mathbf{M}_{ad}$ and undergoes spectral decomposition.

of the graph are determined by an affinity or similarity measure, which signifies how closely two observations are related. These concepts are now clarified.

For each data set $\mathbf{X}_f^l$, a positive-definite kernel that characterizes the proximity between two observations $\mathbf{x}_i^l$ and $\mathbf{x}_j^l$, can be expressed as [24] and [25]

$$k_{ij}^l = \exp\left(\frac{-(d_{ij}^l)^2}{\varepsilon_i^l \varepsilon_j^l}\right) i, j = 1, \ldots, m \tag{4}$$

where $d_{ij}^l = \left\|\mathbf{x}_i^i - \mathbf{x}_j^i\right\|_2$ represents the Euclidean norm or a similarity distance measure of the time trajectories (measurements) in the original physical space, and the $\varepsilon_{ij}^l = \varepsilon_i^l \varepsilon_j^l$ correspond to the bandwidth or kernel scale that represent the affinity between data points (spatial distance).

The sample points x_i, and their associated coefficients k_{ij}^l constitute a weighted undirected graph characterized by the weight matrix $\mathbf{K}^l = [k_{ij}^l]$, with $k_{ij}^l = k_{ji}^l$. Following that, a set of symmetric and non-negative affinity matrices (m-by-$m$pairwise distance matrices) can be defined as follows [7],

$$\mathbf{K}^l = [k_{ij}^l] = \begin{bmatrix} k_{11}^l & k_{12}^l & \cdots & k_{1m}^l \\ k_{21}^l & k_{22}^l & \cdots & k_{2m}^l \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1}^l & k_{m2}^l & \cdots & k_{mm}^l \end{bmatrix}, l = 1, \ldots, F \tag{5}$$

In a physical sense, the affinity matrices, $\mathbf{K}^l$, discard high values ($d_{ij}^l \gg \varepsilon_{ij}^l$) through thresholding. This process preserves local information by ensuring that high values of ε_{ij}^l correspond to rapid diffusion on the manifold.

Each fault scenario's data is subsequently treated as a high-dimensional nonlinear process, from which a low-dimensional representation can be extracted using a single view.

Figure 2 presents a schematic representation of the employed model.

Conceptually, the multiview representation effectively amalgamates the dynamics of individual trajectories into a single mathematical operator $\mathbf{M}_{ad}$, which approximates the exact dynamics of the observed measurements.

After computing the kernel, the diagonal matrix of pairwise distances can be calculated from

$$\mathbf{D}^l = [d_{ii}^l] = diag\left[k_{jj}^l\right] \tag{6}$$

Subsequently, a normalized Markov transition matrix denoted as $\mathbf{M}^l$ can be defined as follows:

$$\mathbf{M}^l = [m_{ij}^l] = (\mathbf{D}^l)^{-1}\mathbf{K}^l \tag{7}$$

with $m_{ij}^l = [k_{ij}^l / k_{ii}^l] = [k_{ij}^l / q_i^l(\mathbf{x}_i, \mathbf{x}_j)]$, $i, j = 1, \ldots, m$, $l = 1, \ldots, F$ where the term $q_i^l(\mathbf{x}_i, \mathbf{x}_j) = \sum_{j=1}^m \exp(-\left\|(\mathbf{x}_i - \mathbf{x}_j)\right\|^2 / \varepsilon_{ij}^l)$ is proportional to the degree of x_i in the graph and thus serves as a measure of the density of points of the vertices, and

$$\mathbf{M}^l = [m_{ij}^l] = \begin{bmatrix} m_{11}^l & m_{12}^l & \cdots & m_{1m}^l \\ m_{21}^l & m_{22}^l & \cdots & m_{2m}^l \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1}^l & m_{m2}^l & \cdots & m_{mm}^l \end{bmatrix}, l = 1, \ldots, F$$

Note that, for each Markov matrix, $\mathbf{M}^l$, $\sum_{j=1, j\neq i}^m m_{ij}^l = 1$, $0 \leq m_{ij}^l \leq 1$. Formally, the elements m_{ij}^l represent the conditional probability of transitioning from state i to state j, describing a discrete random walk on the data graph.

Several interpretations of this model are of interest. Like [17] and [26], let $X_{c_i}^l$ represent a cloud of points around x_i^l that encompasses the m/d nearest neighbors as depicted in Fig. 2. To capture the local geometry around each data point, a weight function, $k_{\varepsilon_{ij}}^l = \sum_{x_i^l \in X_c^l} \exp(-\left\|\mathbf{x}_i^l - \mathbf{x}_2^l\right\|^2 / \varepsilon_{ij}^l)$ is introduced – Refer to (4). Using this representation, the variance of the distance, i.e., a measure of spatial distance, between each data or measurement point, x_i^l and all other points $x_i^l \in X_c^l$ can be calculated from

$$\varepsilon_i^l = \sum_{x_j^l \in X_{c_i}^l} \frac{\left\| \left\|\mathbf{x}_i^l - \mathbf{x}_j^l\right\| - \hat{X}_{c_i}^l \right\|^2}{\left\|\hat{X}_{c_i}^l\right\|} \tag{8}$$

where $\hat{X}_{c_i}^l = \sum_{x_j^l \in X_{c_i}^l} \left\{ \left\|\mathbf{x}_i^l - \mathbf{x}_j^l\right\| / \hat{X}_{c_i}^l \right\}$. This approach makes it possible to compute the kernel scale ε_i^l automatically. Furthermore, the denominator of (7), $q_i^l(\mathbf{x}_i, \mathbf{x}_j)$, is associated with the graph's efficiency and the relevant AD transition matrix necessary for (7) is defined as follows:

$$\mathbf{M}_{ad} = \mathbf{M}^1\mathbf{M}^2 \cdots \mathbf{M}^m \in \Re^{m\times m} \tag{9}$$

Here, the individual Markov matrices $\mathbf{M}^l$ follow the form given in (7). However, in contrast to typical tensor-based applications, this approach is not restricted to third-order tensors.

Two distinct advantages of the model (9) are that the dimension of the matrix $\mathbf{M}_{ad}$ is equal to the number of sensors m, and it offers the possibility to define diffusion distances based on a (single) diffusion operator.

It is important to note that the Markov matrix $\mathbf{M}_{ad}$ is inherently real, nonsymmetric, and row stochastic. As a result, it lacks diagonalizability, and its eigenvectors do not form an orthonormal basis within the space of system measurements.

Symmetry is introduced here by noting that $\mathbf{M}_{ad}$ is the adjoint of a symmetric matrix $\mathbf{M}_{sad}$, which satisfies $\mathbf{M}_{sad} = \mathbf{D}^{1/2}\mathbf{M}_{ad}\mathbf{D}^{-1/2}$ with corresponding eigenvectors $\boldsymbol{\Psi}$ and $\boldsymbol{\varphi}$, respectively. This property enables the extraction of a normal basis from the reduced representation: the eigenvalues for the AD maps operators are given by

$$\begin{cases} \mathbf{M}_{sad}\psi_j = \lambda_j\psi_j \\ \mathbf{M}_{ad}\varphi_j = \lambda_j\varphi_j \end{cases} \quad j = 1, \ldots, m \tag{10}$$

Since $\mathbf{M}_{sad}$ is a Markov matrix, the eigenvalues are bounded in the range [0, 1], and therefore $1 = |\lambda_o| \geq |\lambda_1| \geq \cdots \geq |\lambda_{m-1}|$. This defines the embedding $\boldsymbol{\Psi}_t : x_i \rightarrow \{\lambda_1^t\boldsymbol{\Psi}_1(x_i)\ \lambda_1^t\boldsymbol{\Psi}_2(x_i) \cdots \lambda_1^t\boldsymbol{\Psi}_d(x_i)\}$. The eigenvectors serve as the diffusion coordinates, and the AD general matrix $\mathbf{M}_{sad}$ is assumed to exhibit only a few d dominant coordinates or non-trivial eigenvalues ($d \ll m$). It is easy to demonstrate that the right eigenvectors of $\mathbf{M}_{ad}$ are related to those of'$\mathbf{M}_{sad}$ as $\boldsymbol{\Psi}_j = \mathbf{D}^{-1/2}\boldsymbol{\varphi}_j$. The natural basis can also be obtained from the singular value decomposition of the data, as $\mathbf{M}_{sad}\psi_j = \sigma_j\boldsymbol{\Psi}_j$. This approach is adopted here to enable comparison with other analysis techniques in Section II.

The practical application of the AD framework involves solving two interconnected problems. Firstly, a bandwidth kernel ε_{ij}^l must be computed for each set of measurements, $\mathbf{X}_f^l$. Additionally, given that multiple sets of observations are gathered from the same number of sensors for different fault scenarios, the data exhibit a high correlation. Criteria for extending this approach to more intricate data representations are discussed below.

B. APPROXIMATE SOLUTIONS IN PHYSICAL SPACE

Drawing an analogy with the POD method in [20], an approximation to the data set $\mathbf{X}_f$ is derived as follows:

$$\widehat{\mathbf{X}}_f = \left[\mathbf{a}_1(t)\ \mathbf{a}_2(t) \cdots \mathbf{a}_d(t)\right]\left[\boldsymbol{\Psi}_1\ \boldsymbol{\Psi}_2 \cdots \boldsymbol{\Psi}_d\right] \tag{11}$$

where the $\mathbf{a}_j(t) = \mathbf{X}_j\boldsymbol{\Psi}_j = [a_j(t_1)\ a_j(t_2) \cdots a_j(t_N)]$ are the temporal modes. Here, $\mathbf{a}_o$ denotes the temporal mean or average system behavior and the $\mathbf{a}_d$, $d \geq 1$ capture the significant oscillatory behavior.

The super index l has been dropped for notational convenience.

C. DIFFUSION DISTANCES

Consider two vertexes or measuring points, i,and j, in G. The alternating diffusion distance, denoted as $d_{ad}(i, j)$ between these two points i,j, is defined analogously to the single-view

case [27] as

$$\mathbf{D}_{ad}^2 \in \Re^{(m\times m)} = [d_{ad_{ij}}^2] = \sum_{k>1}^{d} \lambda_k^{-2t} \left|\boldsymbol{\psi}_k(x_i) - \boldsymbol{\psi}_k(x_j)\right|^2 \tag{12}$$

where $\boldsymbol{\psi}_k(x_i)$ denotes the ith element of the eigenvector $\boldsymbol{\psi}_k$ with associated eigenvalues λ_k for matrix $\mathbf{M}_{sad}$ in (10), and t indicates time. The diffusion distance reveals the intrinsic geometric relationship among data points and is robust to noise.

D. OTHER APPROACHES TO MULTIVIEW DIFFUSION MAPS

An alternate and useful formulation, utilizing a multiview approach, can be derived through cross-diffusion and the associated random walk process, which is based on the individual datasets. The multiview kernel or diffusion operator, denoted as $\mathbf{M}_{mv}$ can be expressed in the form [14]:

$$\mathbf{M}_{mv} = \begin{bmatrix} \mathbf{O}_{m\times m} & \mathbf{M}^1\mathbf{M}^2 & \cdots & \mathbf{M}^1\mathbf{M}^m \\ \mathbf{M}^2\mathbf{M}^1 & \mathbf{O}_{m\times m} & \cdots & \mathbf{M}^2\mathbf{M}^m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}^m\mathbf{M}^1 & \mathbf{M}^m\mathbf{M}^2 & \cdots & \mathbf{O}_{m\times m} \end{bmatrix} \in \Re^{(F\times m)\times(F\times m)} \tag{13}$$

where $\mathbf{O}_{m\times m}$ represents a matrix of zeros with dimension $m\times m$, and the terms $\mathbf{M}^i$ and $\mathbf{M}^j$ refer to cross-diffusion (multiview) matrices.

In interpreting this model, one should note the following:

- The multiview diffusion matrix $\mathbf{M}_{mv}$, has dimensions $(F\times m)\times(F\times m)$.3-D representations can be obtained from this matrix, illustrating how clusters evolve from one contingency scenario to another.
- Based on this notion, a (multiview) diffusion distance matrix, $\mathbf{D}_{mv}$ is defined in this study. It shares a similar structure with the multiview matrix $\hat{\mathbf{M}}_{mv}$ in (13) and is defined as

$$\mathbf{D}_{mv} = \begin{bmatrix} \mathbf{D}_{mv}^{11} & \mathbf{D}_{mv}^{12} & \cdots & \mathbf{D}_{mv}^{1F} \\ \mathbf{D}_{mv}^{21} & \mathbf{D}_{mv}^{22} & \cdots & \mathbf{D}_{mv}^{2F} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{mv}^{F1} & \mathbf{D}_{mv}^{F2} & \cdots & \mathbf{D}_{mv}^{FF} \end{bmatrix} \in \Re^{(F\times m)\times(F\times m)} \tag{14}$$

In this equation, the diagonal block matrices $\mathbf{D}_{mv}^{ii}$, $i = 1,2,\ldots,F$, represent the intra-event distances, while off-diagonal blocks, $\mathbf{D}_{mv}^{ij}$, denote distances or correlations across events or fault scenarios.

The approximations (13) and (14) have two advantages. First, they preserve the individual characteristics of each data set $\mathbf{X}^l$. Additionally, individual inverse transformations of the form (11) can be used to determine the appropriate system behavior in physical space. A disadvantage is the large size of the resulting model of dimension $Fm\times Fm$.

Compared with the AD maps technique in (9), the multiview spectral decomposition offers a global view of system behavior, as discussed in the numerical simulations. However, it results in a large modal representation of dimension $(F \times m) \times F\times m)$.

Similar to (12), the multiview diffusion distance matrix can be defined as $(\mathbf{D}_{mv}^{ii})^2 = [d_{mv}^{ii\ 2}] = \sum_{k\geq 1}^{m} \lambda_k^{-2t} \left|\boldsymbol{\Psi}_k(\mathbf{x}_i) - \boldsymbol{\Psi}_k(\mathbf{x}_j)\right|^2$. Additionally, cross-diffusion distances, d_{mv}^{ij} or cross-correlation measures $\mathbf{C}_{mv}^{ij}$ can be computed using the off-diagonal blocks of the diffusion distance matrix in (14).

E. NUMERICAL IMPLEMENTATION

As mentioned earlier, the computation of multiview diffusion kernels requires solving two interconnected problems: the calculation of the individual local bandwidth kernels, ε_i^l, ε_j^l and the computation of diffusion operators, $\hat{\mathbf{M}}_{mv}$.

The following is a summary of the proposed procedure:

Given datasets $\mathbf{X}^l$, $l = 1,\ldots, F$, corresponding to multiple fault events:

Step 1. Calculate the bandwidth kernels ε_i^l, ε_j^l for each data set $\mathbf{X}^l$ using the methodology in Section IV-A.

Step 2. Construct a graph-based Markov chain for each dataset and compute the decomposition, $\hat{\mathbf{M}}_{sad}\boldsymbol{\psi}_j = \lambda_j\boldsymbol{\psi}_j$ or $\hat{\mathbf{M}}_{mv}\boldsymbol{\psi}_j = \lambda_j\boldsymbol{\psi}_j$ using (9) or (14)

Step 3. Generate the embedding $\boldsymbol{\Psi}_t$ that captures the spatial and temporal structure of critical modes and identify features important for wide-area inter-area oscillation monitoring.

Step 4. Compute diffusion distances using (12). In practice, only two or three dominant coordinates are typically sufficient to accurately characterize system dynamics, even though the definition allows for an arbitrary number of relevant coordinates, d.

The AD methods were implemented within the MATLAB programming environment.

V. APPLICATION

The developed procedures have been tested on a realistic 5449-bus 635-generator equivalent model of a large power system [3], [20]. Furthermore, detailed dynamic models are incorporated for system generators and static var systems.

Figure 3 illustrates a section of the study region displaying the locations of significant transmission buses and generators chosen for analysis. The study region comprises 896 buses, 26 wind farms, and three major static VAR compensators.

A. SIMULATION DETAILS

In the study, the diffusion framework was utilized to jointly analyze independent voltage datasets obtained from transient stability simulations. Table 1 summarizes the fault scenarios considered in this study to evaluate the applicability of multiview diffusion techniques in analyzing spatiotemporal datasets. For all cases, 27 major transmission buses were selected for analysis.

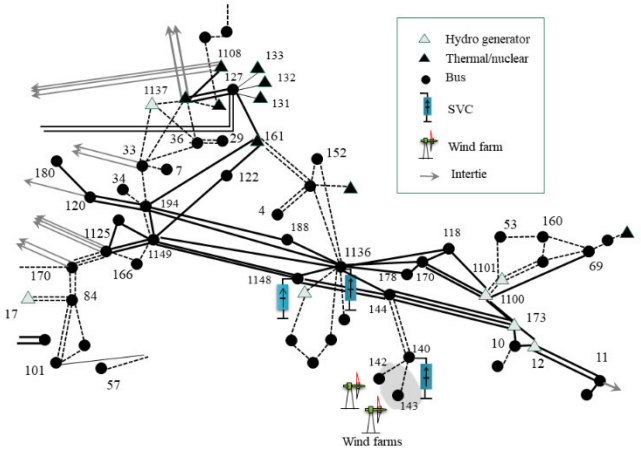


FIGURE 3. Simplified diagram of the study region showing major 400/230 kV transmission and generating resources.

TABLE 1. Bus voltage data sets selected for study.

Contingency Scenario	Description	Data description
CTYS1	Outage without fault of line 1136-144	$\mathbf{X}_v^1 \in \Re^{3600\times 27}$
CTYS2	Outage of unit # 4 at bus 12	$\mathbf{X}_v^2 \in \Re^{3600\times 27}$
CTYS3	Outage of unit # 1 at bus 161	$\mathbf{X}_v^3 \in \Re^{3600\times 27}$

Figure 4 displays the system responses obtained from the transient stability simulation of the entire system model (5449 buses) for the contingency scenarios outlined in Table 1. Each time history spans 30 seconds and consists of 3600 data points ($N = 3600$). To aid interpretation, signals are detrended. Prior research has indicated that these contingencies stimulate the slowest interarea electromechanical mode of the system at about 0.396 Hz [3], [7].

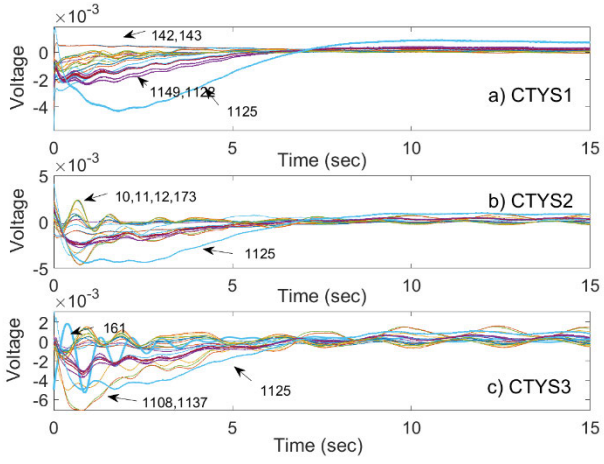


FIGURE 4. Bus voltage magnitudes for the contingencies listed in Table 1. The signals have been detrended for enhanced clarity.

The frequency (and corresponding damping ratios) extracted using multichannel Prony analysis of the signals in Fig. 4 are as follows: 0.397 Hz ($\xi = 0.540$ %) for CTYS1, 0.396 Hz ($\xi = -0.550$ %) for CTSY2, and 0.396 Hz ($\xi = -0.565$ %) for CTYS 3. In the subsequent studies outlined below, three different clustering methods are applied to the data sets listed in Table 1: a) Alternating diffusion analysis, Multiview (MV) diffusion analysis, and ST-HOSVD.

In all three cases, the voltage data tensor is expressed as

$$\boldsymbol{\mathcal{X}}^l(:,:,l) \in \Re^{N\times m\times F} = \mathbf{X}_v^l\ ,\ l = 1,\ldots,3 \qquad (15)$$

where the matrices $\mathbf{X}_v^i$ are defined in Table 1.

B. ALTERNATING DIFFUSION ANALYSIS

In order to identify voltage clusters and calculate diffusion distances, alternating diffusion analysis is employed. In the initial step, the kernel views are constructed utilizing the methods described in Section III.A.

The bandwidth kernels determined using the algorithm in [17] are as follows: $\varepsilon_1 = 0.0420$, $\varepsilon_2 = 0.0651$, and $\varepsilon_3 = 0.0360$ for the three contingency scenarios listed in Table 1.

Figure 5a illustrates the spatial distribution of the dominant eigenvector obtained from the spectral model $\hat{\mathbf{M}}_{sad}\boldsymbol{\Psi}_{v_j} = \lambda_1\boldsymbol{\Psi}_{v_j}$ of voltage datasets, which represents the eigenvalues, and $\boldsymbol{\Psi}_1 = [\,\boldsymbol{\Psi}_v^1\ \boldsymbol{\Psi}_v^2\ \boldsymbol{\Psi}_v^3\,]^T$. On the other hand, Fig. 5b provides a spatial interpretation of this model.

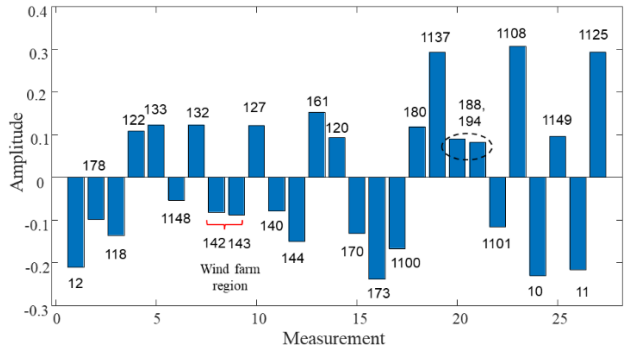
In this context, the red bars represent positive entries while the blue bars represent negative entries in ψ_j. Understanding the interpretation of the diffusion coordinates requires clarification. Components with similar amplitude (and sign) signify highly coherent states or points, whereas the sign change from one eigenvector to another defines coherency. Based on these results, buses 142, 143, bus 140, connected with the wind farm's location and the point of interconnection, and to a lesser extent, bus 144, exhibit high coherency.

Another approach to gain insight into system dynamics is by visualizing the dominant diffusion coordinates in a three-dimensional plot, as shown in Fig. 6. However, this method is primarily visual. Alternatively, coherency can be determined by employing clustering and pattern recognition techniques on the low-dimensional diffusion map embedding.

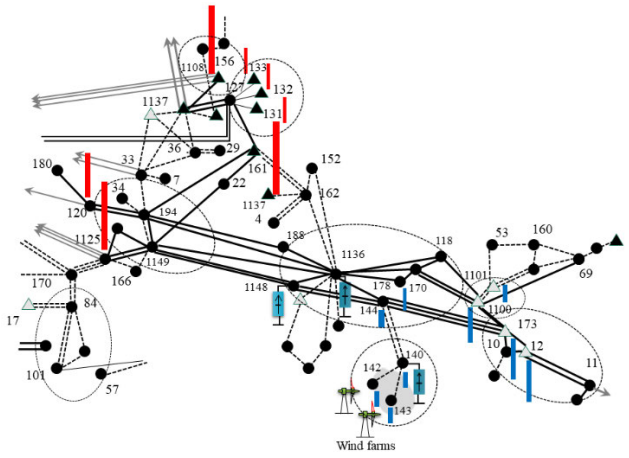
As a subsequent step, the centroid of the data is calculated and a dominant bus (point) is chosen by determining the smallest distance d_{iC} between the centroids and the buses belonging to a specific cluster, as described in the following subsection.

Buses near the geometrical centroid are selected as critical buses for monitoring. This intuition aligns with the anticipated physical behavior using other well-established techniques [28].

Furthermore, the diffusion distances for bus # 142 (measurement # 8) computed using (12) with $d = 2$ are shown in Fig. 7. For completeness, results are compared with the pairwise distances obtained using the Euclidean norm



a). Shape of dominant mode computed using Ψ_1



b). Physical distribution of dominant coordinates

FIGURE 5. Illustration depicting the shape of voltage-based dominant eigenvector Ψ_1 and its physical distribution. The dotted ellipses represent the voltage control areas determined using the approach outlined in [28].

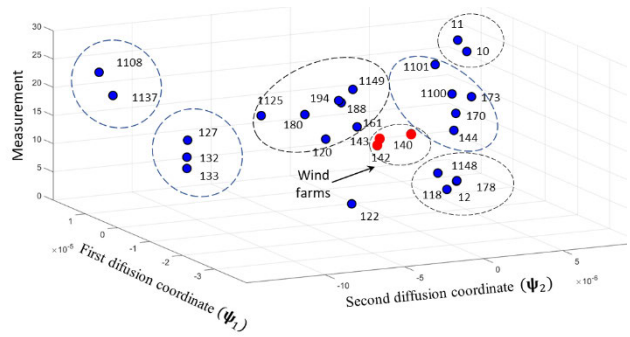


FIGURE 6. Scatter plot of eigenvector Ψ_1 versus Ψ_2.

$d_{ij}^3 = \left\| \mathbf{x}_8^3 - \mathbf{x}_j^3 \right\|, j = 1,\ldots,27$, in the original physical space for contingency scenario CTYS3 in Table 1 and Fig. 4.

Consistency in the results is evident, indicating that physical distance is preserved in the diffusion maps space. Analyzing this plot reveals that the shortest distance from bus # 142 is observed with nearby buses 1148, 143, 140, 178, 1101, and 118. Taken together, this approach improves both the visualization and characterization of the system dynamics, building upon the results depicted in Fig. 5a.

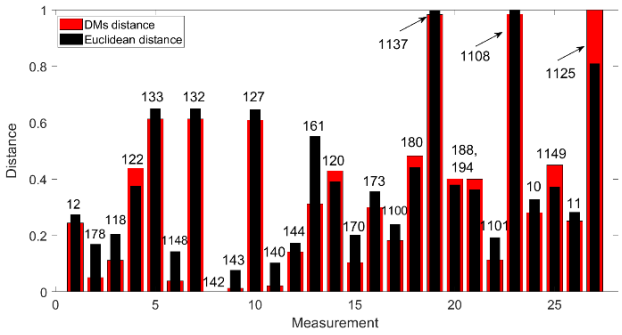


FIGURE 7. Comparison of (normalized) AD distances for bus # 142 with the Euclidean norm in the physical space for contingency scenario CTYS3.

C. TRANSITION PROBABILITY HEATMAPS

An alternative interpretation of the transition probability matrix is obtained by examining the transition probability for the matrix $\mathbf{M}_{sad}$, as demonstrated in Fig. 8. In this representation, the color intensity of each cell represents the strength of the probability of transition from the state represented by the row to state represented by the column.

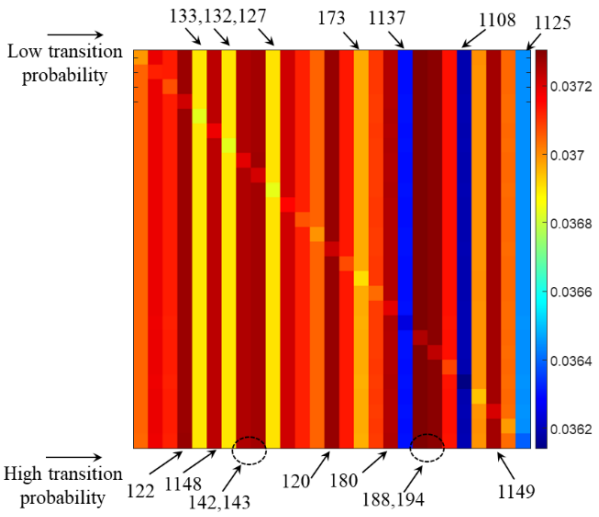


FIGURE 8. Heat map of the transition matrix.

The heatmap color map illustrates the measurements of buses with the highest or lowest transition probability, offering a general overview of the significant spatial patterns of the dominant coordinates.

While diffusion distances can be utilized to define microstates or states nearby, transition probability heatmaps provide a comprehensive visualization of system dynamics.

The plot reveals a significant probability of staying within one group mainly of states for measurements 25 (bus 1149), 20,21 (buses 188,194), 8,9 (buses 142,143), 4 (122), 14 (bus 120) as well as a low probability of leaving that group of measurements 23 (bus 1108), 19 (bus 1137), and 27 (bus 1125).

D. MULTIVIEW SPECTRAL CLUSTERING

To highlight the improved system characterization achieved through the adoption of a multiview approach, the datasets presented in Table 1 are analyzed using the method in Section IV.

In this instance, the diffusion operator $\mathbf{M}_{sad}$ has dimensions $Fm \times Fm$ with $F = 3$ and $m = 27$ corresponding to three fault scenarios outlined in Figure 2 and Table 1.

Multiview representations offer a more intricate interpretation compared to their single-view counterparts. In accordance with the principles of diffusion map theory, the diffusion coordinates are computed from the following $\hat{\mathbf{M}}_{mv}\boldsymbol{\Psi}_j = \lambda_j \boldsymbol{\Psi}_j$, $j = 1, \ldots, d$.

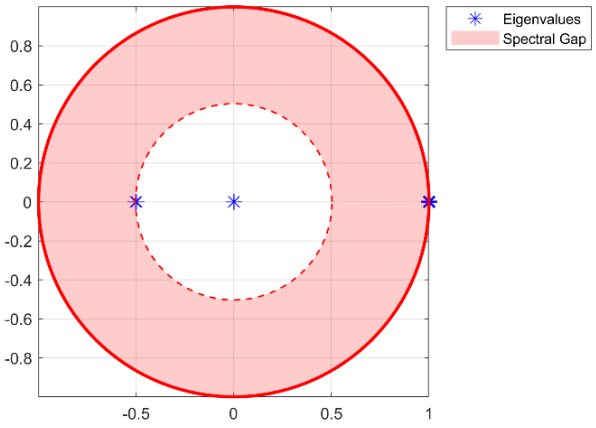


FIGURE 9. Spectral gap of the multiview representation. The blue asterisks indicate the presence of two dominant eigenvalues at approximately $\lambda_2 = -0.5032$ and $\lambda_3 = -0.4968$ (almost identical) and a trivial, leading eigenvalue at $\lambda_1 = 1.0$. All other values are negligible.

As depicted in Fig. 9, the embedding process pinpoints a gap in the eigenvalue spectrum following the third eigenvalue ($d = 3$). More precisely, the spectrum reveals the presence of two significant (non-trivial) eigenvalues, λ_2 and λ_3, indicating that only two eigenvectors are required for characterizing the nonlinear embedding, $\boldsymbol{\Psi}_t$. Using this model, the multiview diffusion distances were computed utilizing (12) for each diagonal block $\mathbf{D}_{mv}^{ii}$ in (14). It is worth noting that the utilization of off-diagonal blocks to correlate and classify fault events remains a topic for further exploration.

For the purpose of illustration, Figs 10(a) through 10(c) display the multiview diffusion distances between measurement # 8 (Bus 142) and the other buses in the system. These distances were obtained from the diagonal distance blocks specified in (14).

E. SEQUENTIALLY TRUNCATED SINGULAR VALUE DECOMPOSITION

The results were compared with tensor factorization-based techniques to further verify and quantify the accuracy of the proposed framework in capturing slow system motion.

In this study, two modeling techniques were considered.

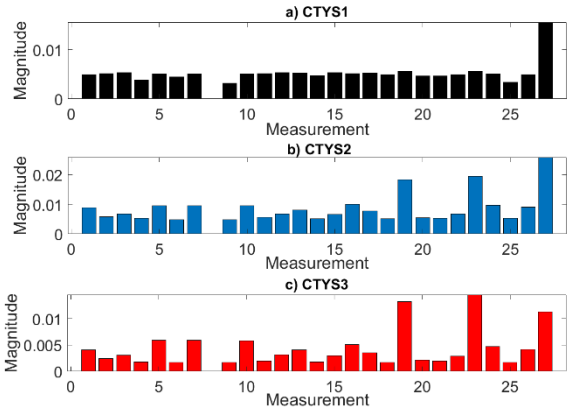
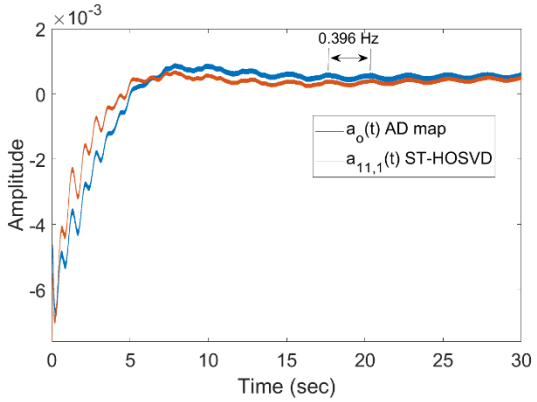
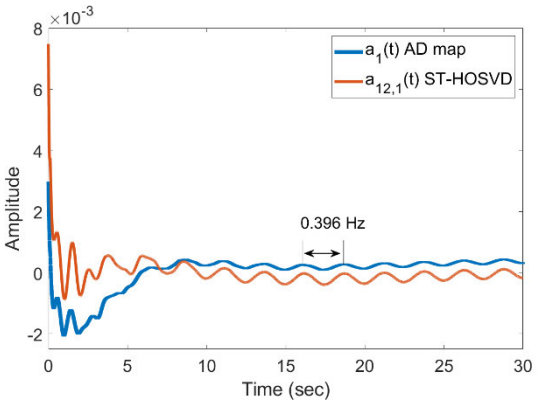


FIGURE 10. Multiview diffusion distances for bus 142 (WF #1) and three contingency scenarios. a) Contingency scenario CTYS1, b) Contingency scenario CTYS2, and c) Contingency scenario CTYS3.



a). Temporal modes $a_{11,1}$ in (3) and a_0 (11)



b). Temporal modes $a_{12,1}$ in (3) and a_1 (11)

FIGURE 11. Time evolution of the extracted temporal modes using AD maps and ST-HOSVD.

1) The ST-HOSVD method as presented in [7] and [12]
2) The AD map, as described in Section IV-A

In the first case, the three-way data presented in Table 1 is organized into a tensor $\hat{\boldsymbol{\mathcal{X}}} = \in \Re^{nt \times N \times p}$ with $n_t = 27$, $N = 3600$, and $p = 3$. The application of ST-HOSVD allows the three-way observational data to be expressed in

terms of a core matrix $\boldsymbol{G}$ and a reduced matrix $\boldsymbol{U}^{(1)}$ in (3) as $\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{G} \times_1 \boldsymbol{U}^{(1)}$. As emphasized before, the tensor $\boldsymbol{G}$ contains the temporal patterns of the three-way representation. When considering only two data sets, namely CTYS1 and CTYS2, ST-HOSVD analysis results in two sets of dominant temporal modes denoted as $a_{11,1}(t)$, $a_{12,1}(t)$ and $a_{11,2}(t)$, $a_{12,2}(t)$, each corresponding to one of the data matrices $\mathbf{X}_v^1$ and $\mathbf{X}_v^!$, respectively. For the given example, $\boldsymbol{G} \in \Re^{61\times1604\times2}$, $\boldsymbol{U}^{(1)} \in \Re^{66\times61}$, $\boldsymbol{U}^{(2)} \in \Re^{1604\times121}$, and $\boldsymbol{U}^{(3)} \in \Re^{2}\times2$. In the second case, the temporal amplitudes $a_j(t)$ are computed from (11) as $\mathbf{a}_j(t) = \mathbf{X}_j\boldsymbol{\Psi}_j$, $j = 0, 1$.

A comparison of simulation results in Fig. 11 shows that both techniques successfully capture the dominant participation of the slowest inter-area mode at about 0.396 Hz in the observed oscillations, validating the model. The results show overall agreement, although some discrepancies can be observed, particularly in the initial phase of the transient response (0-5 seconds).

Furthermore, Table 2 provides a summary of performance comparisons. As seen, AD maps demonstrate favorable results when compared with other data compression techniques in the literature regarding computational efficiency.

TABLE 2. Cpu effort comparison.

Simulation technique	CPU time (seconds)
AD maps (9)	0.1231
Multiview diffusion maps [13]	1.18
ST-HOSVD (3), [7]	1.66

VI. CONCLUSION

In this paper, nonlinear manifold learning technique for compressing, clustering, and classifying high-dimensional observational data has been presented. The resulting low-dimensional embedding accurately captures spatiotemporal structures, allowing for the identification of dominant variables that govern long-term system behavior.

Multiview data fusion leverages cross-information to enhance dynamic characterization, reduce redundancy, and extract informative features from measurements. It can also be applicable to event classification and prediction and improve the understanding of global dynamic phenomena.

Future research will aim to evaluate cross-information between PMU channels, investigate the interactions and relationships between different data modalities, and develop clustering techniques based on distance metrics. Additionally, there is a need for methods to determine the optimal combination of data modalities for better characterizing the dynamics of the global system.

REFERENCES

[1] J. F. Hauer, W. A. Mittelstadt, K. E. Martin, J. W. Burns, and H. Lee, "Integrated dynamic information for the western power system: WAMS analysis in 2005," in *Power System Stability and Control: The Electric Power Engineering Handbook*, L. L. Grigsby, Ed. Boca Raton, FL, USA, 2007, pp. 14-1–14-52. [Online]. Available: ftp://ftp.gov/pub/WAMS%20Information

[2] J. J. Sánchez-Gasca and D. J. Trudnowski, "Identification of electromechanical modes in power systems," Power Syst. Stability Subcommittee, Task Force Identificat. Electromech. Modes, IEEE Tech. Rep. PESTR15, Jun. 2012.

[3] A. R. Messina, *Data Fusion and Data Mining for Power System Monitoring*. Boca Raton, FL, USA: CRC Press, 2020.

[4] B. Sandoval, E. Barocio, P. Korba, and F. R. S. Sevilla, "Three-way unsupervised data mining for power system applications based on tensor decomposition," *Electric Power Syst. Res.*, vol. 187, Oct. 2020, Art. no. 106431.

[5] D. Osipov and J. H. Chow, "PMU missing data recovery using tensor decomposition," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4554–4563, Nov. 2020.

[6] H. Zhao and L. Ma, "Power distribution system stream data compression based on incremental tensor decomposition," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2469–2476, Apr. 2020.

[7] A. Román-Messina, A. Castillo-Tapia, D. A. Román-García, M. A. Hernández-Ortega, C. A. Morales-Rergis, and C. M. Castro-Arvizu, "Distributed monitoring of power system oscillations using multiblock principal component analysis and higher-order singular value decomposition," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 4, pp. 818–828, Jul. 2022.

[8] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[9] P. Ghamisi, B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. K. C. Anders, and R. Gloaguen, "Multisource and multitemporal data fusion in remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[10] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.

[11] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[12] Z. Fang, X. Yang, L. Han, and X. Liu, "A sequentially truncated higher order singular value decomposition-based algorithm for tensor completion," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1956–1967, May 2019.

[13] H. Zhu, S. Klus, and T. Sahai, "A dynamic mode decomposition approach for decentralized spectral clustering of graphs," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Trieste, Italy, Aug. 2022, pp. 1202–1207.

[14] O. Lindenbaum, Y. Bregman, N. Rabin, and A. Averbuch, "Multiview kernels for low-dimensional modeling of seismic events," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3300–3310, Jun. 2018.

[15] G. Pan, L. Xiao, Y. Bai, T. W. Wilson, J. M. Stephen, V. D. Calhoun, and Y.-P. Wang, "Multiview diffusion map improves prediction of fluid intelligence with two paradigms of fMRI analysis," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 8, pp. 2529–2539, Aug. 2021.

[16] R. R. Lederman and R. Talmon, "Learning the geometry of common latent variables using alternating-diffusion," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 3, pp. 509–536, 2015.

[17] C. M. C. Arvizu and A. R. Messina, "Dimensionality reduction in transient simulations: A diffusion maps approach," *IEEE Trans. Power Del.*, vol. 31, no. 5, pp. 2379–2389, Oct. 2016.

[18] E. J. Min, E. C. Chi, and H. Zhou, "Tensor canonical correlation analysis," *Stat*, vol. 8, no. 1, pp. 1–11, Jan. 2019.

[19] R. Badeau and R. Boyer, "Fast multilinear singular value decomposition for structured tensors," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1008–1021, Jan. 2008.

[20] A. R. Messina and V. Vittal, "Extraction of dynamic patterns from wide-area measurements using empirical orthogonal functions," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 682–692, May 2007.

[21] O. Lindenbaum, A. Yeredor, M. Salhov, and A. Averbuch, "Multi-view diffusion maps," *Inf. Fus.*, vol. 56, pp. 1729–1749, Mar. 2020.

[22] O. Katz, R. Talmon, Y.-L. Lo, and H.-T. Wu, "Alternating diffusion maps for multimodal data fusion," *Inf. Fusion*, vol. 45, pp. 346–360, Jan. 2019.

[23] F. R. K. Chung, "Spectral graph theory," Presented at the *Regional Conf. Syst. Math., CBMS Conf. Recent Adv. Spectral Graph Theory*, Fresno, CA, USA, 1994.

[24] R. R. Coifman and S. W. Zucker, "Geometric diffusion as a tool for harmonic analysis and structure definition of data: Diffusion analysis," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 21, pp. 7426–7431, May 2005.
[25] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, "Nonlinear dimensionality reduction in molecular simulations: The diffusion maps approach," *Chem. Phys. Lett.*, vol. 509, nos. 1–3, pp. 1–11, Jun. 2011.
[26] G. David and A. Averbuch, "Hierarchical data organization, clustering and denoising via localized diffusion folders," *Appl. Comput. Harm. Anal.*, vol. 33, no. 1, 2012.
[27] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems," *Multiscale Model. Simul.*, vol. 7, no. 2, pp. 842–864, Jan. 2008.
[28] P. Lagonotte, J. C. Sabonnadiere, J. Y. Leost, and J. P. Paul, "Structural analysis of the electrical system: Application to secondary voltage control in France," *IEEE Power Eng. Rev.*, vol. 9, no. 5, pp. 479–486, May 1989.

A. ROMÁN-MESSINA (Life Fellow, IEEE) received the M.Sc. degree (Hons.) in electrical engineering from the National Polytechnic Institute of Mexico, Mexico City, Mexico, in 1987, and the Ph.D. degree from Imperial College London, U.K., in 1991. Since 1997, he has been a Professor with the Center for Research and Advanced Studies (CINVESTAV), National Polytechnic Institute, Guadalajara, Mexico. His current research interests include power system stability analysis and control and the development and application of measurement-based signal processing techniques to study transient processes in power systems.

CLAUDIA M. CASTRO-ARVIZU (Member, IEEE) received the B.E. degree from the University of Guanajuato, Guanajuato, Mexico, in 2013, and the M.Sc. and Ph.D. degrees in electrical engineering from the Center for Research and Advanced Studies (CINVESTAV), Guadalajara, Mexico, in 2015 and 2020, respectively. She is currently affiliated with the Department of Design, Science, and Technology, Universidad Autónoma de Guadalajara (UAG), Mexico. Her current research interests include power system dynamics, computational intelligence, and data mining.

ALEJANDRO CASTILLO-TAPIA received the B.E. degree in communications and electronics engineering from the National Polytechnic Institute, Mexico City, Mexico, in 2000, and the M.Sc. and Ph.D. degrees in electrical engineering from the Center for Research and Advanced Studies (CINVESTAV), Mexico, in 2003. His current research interests include power system state estimation and optimal sensor placement.

ERLAN R. MURILLO-AGUIRRE is currently pursuing the Ph.D. degree in electrical engineering with the Center for Research and Advanced Studies (CINVESTAV), Guadalajara, Mexico.

O. RODRÍGUEZ-VILLALÓN received the Sc.D. degree from Universidad Michoacana de San Nicolás de Hidalgo, Mexico, in 2012. Since 2000, he has been a Professor and a Researcher with the DICIS, University of Guanajuato, Mexico. From 2014 to 2016, he was a Postdoctoral Researcher with PSL-ETH Zurich, Switzerland. His current research interests include nonlinear analysis of power systems, the analysis of distribution grids based on measurements, and hydrogen production from renewable sources.