# Overlapping Clustering Based Technique for Scalable Uncertainty Quantification in Physical Systems* 

Arpan Mukherjee ${ }^{\dagger}$, Rahul Rai ${ }^{\dagger}$, Puneet Singla ${ }^{\ddagger}$, Tarunraj Singh ${ }^{\dagger}$, and Abani Patra ${ }^{\dagger}$


#### Abstract

Forward propagation of uncertainty in physics-based model is nontrivial and a necessary undertaking. This paper provides a methodology for decomposing the state space of scalable dynamical systems with strong interstate coupling. The outlined approach intends to make rigorous Uncertainty Quantification (UQ) of the high-dimension problem feasible by partitioning the overall high-dimensional state space problem into multiple lower-dimensional state space problems. This approach will work quicker with a lesser memory space requirement than existing methods. To enable accelerated and scalable UQ in high-dimensional complex physical system models, the proposed decomposition process leverages an overlapping community detection to detect state variables participating in more than one subsystems (clusters). The final UQ solution is obtained by using the concept of Hadamard product of the state variables in a subsystem (cluster) and their association in the cluster. The developed approach has been tested to detect connected subsystems in coupled dynamical systems. The results analyzing spatio-temporal flow equation are also presented. It is also shown that proposed framework approach is faster and works with a lesser memory requirement to carry out UQ of high-dimensional physical system models.


Key words. uncertainty quantification, large scale systems, strongly coupled subsystems, state-space clustering, overlapping cluster detection, Hadamard product

AMS subject classifications. $65 \mathrm{C} 30,91 \mathrm{C} 20$, 65D32
DOI. 10.1137/18M1200567

1. Introduction. Physical processes are often modeled as scalable nonlinear systems. Additionally, the presence of uncertainty in the state variables or parameters underlying such a system is highly probable. The number of state variables involved in the complex mathematical models is often large. To get a better understanding of a system, Uncertainty Quantification (UQ) is used as a tool to enable rigorous prediction modeling. UQ by analytic methods of high-dimensional systems is computationally intractable due to the well-known phenomenon of curse of dimensionality. For a high-dimensional system, conventional UQ methods become ineffective due to significant error in the approximation with propagation. To make rigorous UQ of the high-dimension problem feasible, it is prudent to utilize the strategy of divide and conquer. We present a computational scheme for solving such high-dimensional physics-based

[^0]models by decomposing the overall system into connected subsystems. This is achieved by partitioning the set of physical states and the functional decomposition in the probability space. The UQ for the overall system is carried out by agglomerating results of UQ techniques on smaller subsystems. Such a decomposition facilitates a faster analysis by utilizing the power of parallel computation. In addition, the partitioning scheme when integrated with UQ techniques such as quadrature-based methods exhibit drastic reduction in the number of collocation points. Our prior work focused on using techniques of linearization and graph clustering to decompose such systems into a mutually exclusive and exhaustive set of subsystems [35]. The current work relaxes this assumption and focuses on the UQ of Strongly Coupled System (SCS). The concept of degree of participation or fuzzy association of a state variable in a particular subsystem has been employed to enable UQ of high-dimensional systems (Figure 1.1). The fuzzy association allows us to determine the number of subsystems and the state variables participating in each subsystem.


Figure 1.1. Use of overlapping clusters of state variable for propagation of individual clusters in parallel. The third step shows mapping of overlapping variables to nonoverlapping clusters.

Our proposed framework integrates the concept of variance and domain decomposition. Given a dynamical system with uncertainty information, the best possible linearization of the velocity function is approximated. The fuzzy association of the state variables in a particular subsystem is then obtained from an Overlapping Graph Clustering Algorithm. Such an algorithm treats a linearized matrix as a well-connected graph and provides the number of clusters and the degree of participation of each node in a cluster. The clusters are treated individually in parallel. Thereafter, the solution is obtained by the element-to-element or Hadamard product of the solution of the state variables in a cluster and their association in the cluster.

The paper is organized as follows: Section 2 reflects on some of the recent related works. Section 3 discusses the overall methodology for the UQ of a high-dimensional dynamical system. Specifically, section 3.1 outlines the problem formulation. Section 3.2 presents the overall framework. Section 3.3 defines the concept of Weakly Coupled Subsystem (WCS) and SCS. Section 3.5 provides theoretical details on the distribution of 27 SCS. Sections 3.4-3.8 provide details involving the method of linearizing, clustering, and UQ. The effectiveness of the developed framework is demonstrated on suitable numerical experiments in sections 4 and 5. Section 6 concludes with a summary of the framework, its effectiveness, and some possible future works.
2. Related work. The presented work follows the idea of stochastic domain decomposition $[12,42,48]$ and additive Schwarz method [7, 44, 53]. Although these works focus on the overlapping or nonoverlapping partitioning of a continuous space domain, the methods provide a computational intuition for partitioning a set of discrete physical states. The current section is focused on the literature survey of two main pertinent categories. The first category is related to overlapping cluster detection algorithms. These algorithms have been researched widely in the field of community detection problems in a connected network. The second category of related works discusses the useful UQ techniques.
2.1. Overlapping community detection. The study of overlapping community has been a recent addition to the world of scientific research. Instead of detecting disjoint community in a network, in overlapping community framework a node in a network can participate in multiple communities. In this context, Baumes et al. [4] proposed a two-step algorithm to identify disjoint clusters, followed by an iterative method of assigning nodes to clusters until a density function is improved. Palla et al. [37] have developed a clique-based method and have applied the method to a protein network comprising of 30,739 links. Algorithms such as LFM [26], MONC [21], CPMw [16], and SCP [25] have extended the earlier work to enhance the applicability of community detection algorithms. OSLOM [27] and several effective overlapping community detection algorithms identifying overlapping cluster structures in complex networks also exist $[8,36,43]$. Although, all the above mentioned works detect overlapping clusters, they do not quantify the degree of participation of a node in a given cluster. To address the shortcoming of quantifying the degree of participation, fuzzy based clustering methods have been developed. The fuzzy $k$-means algorithm in which, instead of optimizing the sum of the Euclidean distance between the nodes and the cluster centers, a weighted objective function is used to quantify the degree of participation [15]. Zhang, Wang, and Zhang [54] has combined the fuzzy k-means clustering algorithm with the well-established method of Spectral Clustering [51]. Distribution-based algorithms have been further developed, whereby the association of a node to a cluster is modeled as a probability distribution and is updated with the availability of the adjacency information [28, 31, 39]. In our current work, we have used the nonnegative matrix factorization based algorithm to detect the degree of association of a node to a cluster [39]. The algorithm iteratively minimizes the rank of the degree matrix along with finding the association of the nodes in each cluster.
2.2. Scalable uncertainty quantification. The UQ of a physical system refers to the forward propagation of the statistical properties of the variables of interest based on the random-
ness in its initial conditions or parameters. This involves repeated simulations of the system on a pool of realizations generated for the multidimensional state variable. Sampling-based methods $[5,6,17,18,22,32]$ requires a generation of a large number of samples from the initial probability distribution. While Monte Carlo (MC) methods suffer from slow convergence rates, the other sampling strategies to alleviate this problem (e.g., Importance Sampling [32]) cannot be paralleled effectively. For large dynamical systems, simulation-based approaches require a significant number of sample points for state space distribution approximation. A large number of sample points make simulation-based methods very computationally intensive.

An alternative to the random sampling is the Quadrature Scheme such as Gaussian Quadrature [47] and its variation such as Sparse-grid Collocation [3, 19], Minimal Cubatures $[2,14,46]$, which involve a deterministic scheme to generate points to reproduce exactly the integrals for polynomials, or the moments of the density function. These schemes require a relatively smaller number of points (quadrature, cubature, or sigma points) that can approximate the distribution of the large-dimensional random vectors. While Quadrature and Sparse-grid Quadrature based methods require exponential or high-dimensional polynomial order of realizations, minimal cubature methods exist only for few dimensional problems and for selected probability density functions (pdfs).

Scalable problems in UQ have been addressed by Domain Decomposition [30], partial observation [38], network-based method [20, 24], and Anova Decomposition [40]. The method presented in this paper is a combination of networks-based and Anova decomposition that uses an overlapping graph clustering algorithm. The method also aids in performing meaningful domain decomposition in spatio-temporal flow problems.
3. Overlapping cluster detection based method for uncertainty quantification of large dynamical system. This section discusses following components in details: (1) Overall framework for finding Strongly Coupled Subsystems (SCSs) and (2) propagation of uncertainty in a large dynamical system through the SCSs.
3.1. Problem formulation. Consider a $n$-dimensional coupled dynamical system defined by the following Stochastic Differential Equation (SDE):

$$
\begin{equation*}
\dot{\mathbf{x}}_{t}=f\left(\mathbf{x}_{t}\right), \quad \mathbf{x}_{t_{0}}=\mathbf{x}_{0}, \tag{3.1}
\end{equation*}
$$

where $f(x)$ is an $n$-dimensional vector of deterministic square integrable functions $f=$ $\left[f_{1}, f_{2}, \ldots, f_{n}\right]$, with $f: \mathbb{R}^{n} \times \mathbb{R} \rightarrow \mathbb{R}^{n}$. Here $\mathbf{x}_{t}=\left\{\mathbf{x}(t, \omega), t \in[0, \infty), \omega \in \Omega_{\mathbf{x}}\right\}$ is a stochastic process defined on the probability space $\left(\Omega_{\mathbf{x}}, \mathcal{F}_{\mathbf{x}}, P_{\mathbf{x}}\right)$ and

$$
\begin{equation*}
\mathbf{x}_{t}:\left([0, \infty) \times \Omega_{\mathbf{x}}, \mathcal{B}([0, \infty)), \mathcal{F}_{\mathbf{x}}\right) \rightarrow\left(\mathbb{R}^{n}, \mathcal{B}\left(\mathbb{R}^{n}\right)\right) \tag{3.2}
\end{equation*}
$$

Equation (3.1) admits the solution $\mathbf{x}_{t}=\phi^{t}\left(\mathbf{x}_{0}\right)$, where the deterministic flow is given by

$$
\begin{equation*}
\phi: \mathbb{R} \times \Omega_{\mathrm{x}} \rightarrow \mathbb{R}^{n} . \tag{3.3}
\end{equation*}
$$

The probability measure $P_{\mathbf{x}_{t}}$ is characterized by the density function $p\left(\mathbf{x}_{t}, t\right): \mathbb{R}^{n} \times \mathbb{R}$ $\rightarrow[0,1]$, which is a function of both $\mathbf{x}_{t}$ and $t$, where

$$
\begin{equation*}
P_{\mathbf{x}_{t}}=P\left(\mathbf{x}_{t} \leq \mathbf{z}_{t}\right)=\iint \cdots \int_{-\infty}^{\mathbf{z}_{t}} p(\tau, t) d \tau \tag{3.4}
\end{equation*}
$$

Our objective is to efficiently compute the statistical properties (say, first few moments) of state variable $\mathbf{x}_{t}=\phi^{t}\left(\mathbf{x}_{0}\right)$, given as

$$
\begin{align*}
& \mu_{t_{k}}=E\left[\mathbf{x}_{t_{k}}\right],  \tag{3.5}\\
& \Sigma_{t_{k}}=E\left[\left(\mathbf{x}_{t_{k}}-\mu_{t_{k}}\right)\left(\mathbf{x}_{t_{k}}-\mu_{t_{k}}\right)^{\prime}\right] .
\end{align*}
$$

The main challenge lies in computing statistical moments when $n$ is very large.
3.2. Proposed framework. We brief on the main components of our proposed framework. The key idea is to show how a large-dimensional dynamical system can be decomposed into small interconnected subsystems to be solved in parallel. Figure 3.1 depicts the computational pipeline of the underlying methodology. Similar to our previous methodology [33, 35], a graphtheoretic representation is adopted for a given dynamical system with involved uncertainty. Unlike the previous approach, the graph is hypothesized to comprise both weak and strong edges or couplings. The statistical linearization method is used to linearize the dynamical system in the domain of interest represented by the initial state density function (Figure 3.1a). Next, a suitable overlapping community detection algorithm is applied to detect the SCSs (Figure 3.1b). These SCSs determine the assignment of each node or variable to a cluster and their participation (Figure 3.1c). The stochastic dynamical system corresponding to each SCS of reduced order is then propagated by a UQ method (Figure 3.1d). An element-to-element product based step is applied to estimate the statistical properties of the state vector from the properties of the SCSs. These steps are followed to approximate the statistical properties, that is being propagated through the dynamics of the overall system. The overall procedure is continued until a measurement data is available. The measurement is used to filter out noise from the state variable and the statistical properties are recalibrated using a Filtering technique (Figure 3.1e). The cluster structure or the SCSs is/are then recomputed based on the updated statistical properties. The whole process is followed for the required time of analysis. In the subsequent subsections, the theoretical and algorithmic details of each of these steps are discussed.
3.3. Definition of weakly coupled subsystems and overlapping clusters. Given the system of SDE in (3.1), the solution for an initial condition $\mathbf{x}_{0}$ is given by the vector of functions $\phi^{t}\left(\mathbf{x}_{0}\right)=\left[\phi^{t}\left(\mathbf{x}_{0}\right)_{1}, \phi^{t}\left(\mathbf{x}_{0}\right)_{2}, \ldots, \phi^{t}\left(\mathbf{x}_{0}\right)_{n}\right]$. The function $\phi^{t}\left(\mathbf{x}_{0}\right)$ is computed by the following stochastic integral:

$$
\begin{equation*}
\mathbf{x}_{\mathbf{t}}=\phi^{t}\left(\mathbf{x}_{0}\right)=\mathbf{x}_{0}+\int_{0}^{t} f\left(x_{\tau 1}, x_{\tau 2}, \ldots, x_{\tau n}\right) d \tau \tag{3.6}
\end{equation*}
$$

The expected value of a measurable function $\mathcal{G}$, square integrable function in the probability space of $\mathbf{x}_{t}$, is given by the deterministic multidimensional integral

$$
\begin{equation*}
E\left[\mathcal{G}\left(\mathbf{x}_{t}\right)\right]=\iint \ldots \int_{\Omega_{\mathbf{x}}} \mathcal{G}\left(x_{t 1}, x_{t 2}, \ldots, x_{t n}\right) d P_{\mathbf{x}} . \tag{3.7}
\end{equation*}
$$

For any high value of $n$, both of the integrals (see (3.6) and (3.7)) are computationally expensive. To facilitate a faster computation, the definitions of Weakly and Strongly Connected Subsystem (WCS and SCS) are introduced next.


Figure 3.1. Overall framework of the methodology.
3.3.1. Weakly coupled subsystem. The definition of WCSs follows from our earlier works [33, 35]. WCSs $\mathbf{y}_{t_{j}} \in \mathbb{R}^{n_{j}}$ are countable and nonoverlapping partitions of the state space $\mathbf{x}_{t} \in \mathbb{R}^{n}$ such that they are decoupled with each other, and their ensemble can approximate the propagation of the whole system. For the dynamical system in (3.1), the state space $\mathbf{x}_{t}$ is clustered into $m$ subsets as

$$
\begin{align*}
& \mathbf{y}_{t_{1}}=\left\{x_{t_{11}}, x_{t_{12}}, \ldots, x_{t_{1_{1}}}\right\}, \\
& \mathbf{y}_{t_{2}}=\left\{x_{t_{21}}, x_{t_{22}}, \ldots, x_{t_{2 n_{1}}}\right\},  \tag{3.8}\\
& \vdots \\
& \mathbf{y}_{t_{m}}=\left\{x_{t_{m 1}}, x_{t_{m 2}}, \ldots, x_{t_{m n_{1}}}\right\},
\end{align*}
$$

where $\sum_{j=1}^{m} n_{j}=n$. Each WCS $\mathbf{y}_{t_{j}}, j=1, \ldots, m$, is a reduced order stochastic process defined on the probability space $\left(\Omega_{j \mathbf{x}}, \mathcal{F}_{j \mathbf{x}}, P_{j}\right)$, such that $\Omega_{\mathbf{x}}$ is the disjoint union of the countable partitions $\Omega_{j \mathbf{x}}$ 's. The submanifold $\mathbf{y}_{t_{j}}$ is defined by the following $\mathbb{R}^{n_{j}}$-dimensional SDE:

$$
\begin{equation*}
\dot{\mathbf{y}}_{t_{j}}=f_{j}\left(\mathbf{y}_{t_{j}}\right), \quad j=1,2, \ldots, m, \tag{3.9}
\end{equation*}
$$

where $f_{j}\left(\mathbf{y}_{t_{j}}\right)=\left\{f_{j_{1}}, f_{j_{2}}, \ldots, f_{n_{j}}\right\}$ is a subset of the vector of functions $f(\cdot)$. The solution is
approximated as

$$
\begin{equation*}
\mathbf{y}_{t_{j}}=\mathbf{y}_{t_{j}}+\int_{0}^{t} f_{j}\left(x_{\tau_{j 1}}, x_{\tau_{j 2}}, \ldots, x_{\tau_{j n_{j}}}\right) d \tau \tag{3.10}
\end{equation*}
$$

and the expression for mathematical expectation is given as

$$
\begin{equation*}
E\left[\mathcal{G}\left(\mathbf{y}_{t_{j}}\right)\right]=\iint \ldots \int_{\Omega_{j \mathbf{x}}} \mathcal{G}\left(x_{t_{j 1}}, x_{t_{j 2}}, \ldots, x_{t_{j n_{j}}}\right) d P_{j} . \tag{3.11}
\end{equation*}
$$

At any instance, the mathematical for $\mathbf{x}_{t}$ is written as

$$
\begin{equation*}
E\left[\mathcal{G}\left(\mathbf{x}_{t}\right)\right]=\bigoplus_{j=1}^{m} E\left[\mathcal{G}\left(\mathbf{y}_{t_{j}}\right)\right] \tag{3.12}
\end{equation*}
$$

Under this decomposition scheme a particular state is allowed to participate in a single cluster only. Next, this concept of WCS is extended to formulate the propagation and moment equation for state variable $\mathbf{x}_{t} \in \mathbb{R}^{n}$ for dynamic systems exhibiting strong coupling.
3.3.2. Strongly coupled subsystems. To define an SCS, a new state-space decomposition scheme is proposed with the following assumptions:

- A state $x_{i}, i=1$ to $n$, can participate in more than one cluster.
- For each state, we quantify its association with a cluster. Instead of a binary $0-1$ association, we propose a fuzzy association between 0 and 1 . Such a scheme can incorporate both decoupled and overlapping clusters.
- The number of clusters is determined from the initial condition uncertainty and the associated dynamics for every run.
The association of the state $\mathbf{x}$ in a particular cluster $j$ is defined by the variable $\mathbf{z}_{j} \in \mathbb{R}^{n}$, $\mathbf{z}_{j}=\left\{z_{1 j}, z_{2 j}, \ldots, z_{n j}\right\}$, where $z_{i j}, i=1$, to $n$ denotes the association of the state $x_{i}$ in cluster $j$. The variable $\mathbf{z}$ has the following properties:
- $\forall i=1$ to $n$ and $j=1$ to $m, 0 \leq z_{i j} \leq 1$,
- $\sum_{j=1}^{m} z_{i j}=1$.

Under this decomposition scheme, the state random variable $\mathbf{x}$ is represented in each cluster $j$ by the independently distributed component $\mathbf{x}^{j}$. A state $x_{i}$ has either 0,1 (full) or partial association with a cluster. Thus, $x_{i}$ is represented by a linear combination of $x_{i}^{j}$ 's, and its association $z_{i j}, j=1$, to $m$ is given as

$$
\begin{equation*}
x_{i}=\sum_{j=1}^{m} z_{i j} x_{i}^{j} . \tag{3.13}
\end{equation*}
$$

In the vector notation the state vector is given as

$$
\begin{equation*}
\mathbf{x}=\sum_{j=1}^{m} \mathbf{z}_{j} \odot \mathbf{x}^{j} \tag{3.14}
\end{equation*}
$$

where $\odot$ represents the element-to-element or Hadamard product and $\mathbf{x}^{j}, \mathbf{z}_{j} \in \mathbb{R}^{n}$ are $n$ dimensional random vectors. As mentioned earlier, for any two clusters $j_{1}, j_{2}=1$, to $m$, $j_{1} \neq j_{2}$, we assume $\mathbf{x}^{j_{1}} \Perp \mathbf{x}^{j_{2}}$.

An SCS is therefore defined as the subset $\mathbf{x}^{p}$ of $\mathbf{x}^{j}$ that is participating in cluster $j$,

$$
\begin{equation*}
\mathbf{x}^{p_{j}}=\left\{x_{i} \in \mathbf{x} \mid z_{i j}>0\right\} . \tag{3.15}
\end{equation*}
$$

A complimentary of the SCS is the nonparticipating variables of a cluster $j$ defined as

$$
\begin{equation*}
\mathbf{x}^{n p_{j}}=\left\{x_{i} \in \mathbf{x} \mid z_{i j}=0\right\} . \tag{3.16}
\end{equation*}
$$

The values of $Z$ matrix $\left(Z=\left\{z_{i j}\right\}\right)$ can be tabulated as shown in Table 3.1.
Table 3.1
The association values $z_{i j}$ tabulated in the matrix form.

| States | Clusters | 1 | 2 | $\ldots$ |
| :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |
| 1 | $z_{11}$ | $z_{12}$ | $\ldots$ | $z_{1 m}$ |
| 2 | $z_{21}$ | $z_{22}$ | $\ldots$ | $z_{2 m}$ |
| $\vdots$ | $\ldots$ | $\ldots$ | $\vdots$ | $\ldots$ |
| $n$ | $z_{n 1}$ | $z_{n 2}$ | $\ldots$ | $z_{n m}$ |

The detailed theory and application of SCS for enabling UQ are discussed in subsequent subsections.
3.4. Identification of SCS: Linearization and clustering. The combination of linearization and clustering technique facilitates the decomposition of the state space vector $\mathbf{x}$ of the overall dynamical system into SCSs. In our earlier work [35], scalable dynamical systems have been studied extensively, whereby it has been shown that appropriate linearization converts a dynamical system into an undirected graph adjacency matrix. This adjacency matrix takes the form of a block diagonal matrix. Suitable graph clustering techniques can then be used to identify the discernible blocks. These blocks correspond to the WCSs [33, 34]. The best possible technique of linearization that gives a better approximation of a nonlinear system has also been discussed in our earlier work [35]. The method for identifying SCSs uses a similar approach but a different clustering algorithm (described next). Given a nonlinear velocity function $f\left(\mathbf{x}_{t}\right)$ defined at a particular instance $\mathbf{x}_{t}$, an approximate linear form is given as [35, 41],

$$
\begin{equation*}
f\left(\mathbf{x}_{t}\right)=b_{s l}+A_{s l}\left(\mathbf{x}_{t}-\mu_{t}\right), \tag{3.17}
\end{equation*}
$$

where

$$
\begin{equation*}
\min _{A_{s l}, b_{s l}} J=\int_{\Omega_{t}}\left\|f\left(z_{t}\right)-A_{s l}\left(z_{t}-\mu_{t}\right)-b_{s l}\right\|_{2} p_{\mathbf{x}_{t}}(z) d z \tag{3.18}
\end{equation*}
$$

The estimates of this model is given by the stationary solutions to the optimization problem, given as

$$
\begin{align*}
\frac{\partial J}{\partial A_{s l}}=0 & \Rightarrow A_{s l} E\left[\left(\mathbf{x}_{t}-\mu_{t}\right)^{T}\left(\mathbf{x}_{t}-\mu_{t}\right)\right]-E\left[f\left(\mathbf{x}_{t}-\mu_{t}\right)^{T}\right]=0 \\
& \Rightarrow A_{s l}=E\left[f\left(\mathbf{x}_{t}-\mu_{t}\right)^{T}\right] P_{\mathbf{x}_{t} \mathbf{x}_{t}}^{-1},  \tag{3.19}\\
\frac{\partial J}{\partial b_{s l}}=0 & \Rightarrow b_{s l}-E[f]=0 \\
& \Rightarrow b_{s l}=E[f] .
\end{align*}
$$

The estimated linear system matrix $A_{s l}$ is used as the graph adjacency matrix. Given $A_{s l}$, a suitable graph clustering technique can be used to decouple the state space into the SCSs. Due to its ease of application and accuracy in estimating the graph clusters, a modification of the Louvain Modularity maximization method [10, 11] is used for detecting the cluster structure. Given the linear system matrix, the normalized graph adjacency $W \in \mathbb{R}^{n \times n}$ is formulated as

$$
\begin{equation*}
W_{i, j}=0.5 *\left(\left|A_{s l}\right|+\left|A_{s l}^{T}\right|\right) \tag{3.20}
\end{equation*}
$$

The cluster structure $Z$ is estimated by solving the following maximization problem:

$$
\begin{align*}
& \max _{z_{i c}} Q_{o v}^{C}=\frac{1}{2 n} \sum_{c \in C} i, j \in c\left[W_{i j}-\frac{1}{2 n}\right] z_{i, c} z_{j, c} \\
& \text { s.t. } \\
& 0 \leq z_{i c} \leq 1 \forall c \in C, \sum_{c=1}^{|C|} z_{i c}=1  \tag{3.21}\\
& z_{i c}=\frac{\sum_{k \in c W_{i k}}}{\sum_{c^{\prime} \in C_{i}} \sum_{k \in c^{\prime}} W_{i k}}
\end{align*}
$$

and $Z=\left[z_{i j}\right] \in \mathbb{R}^{n \times m}, m \ll n$ contains the fuzzy association of a node of a graph in a particular community. The variables of $Z$ that are below a certain threshold $\epsilon$ are converted to 0 , and the rows of the modified matrix are normalized to satisfy the constraint $\sum_{c=1}^{|C|} z_{i c}=1$. This association matrix $Z$ is used to model the propagation function for the SCSs under the principle of overlapping decomposition.
3.5. Propagation of a random sample $x \in \Omega$. The computational scheme for solving a scalable first-order ODE defined in section 3.1 uses the cluster matrix $Z=\left[z_{i j}\right] \in \mathbb{R}^{n \times m}$ obtained from solving (3.21). Each random sample $\mathbf{x} \in \mathbb{R}^{n}$ is carefully partitioned into $m$ overlapping clusters. Subsequently, the overall dynamical system is also broken down into $m$ independent equations. The state variable $\mathbf{x}$ is a union of overlapping variables $\mathbf{x}^{n o}$ and nonoverlapping variables $\mathbf{x}^{o_{i}}$ :

$$
\begin{gather*}
\mathbf{x}^{n o}=\left\{x_{i} \in \mathbf{x} \mid z_{i j}=1 \forall j=1 \text { to } m\right\}, \\
\mathbf{x}^{o}=\left\{x_{i} \in \mathbf{x} \mid 0<z_{i j}<1 \forall j=1 \text { to } m\right\} \\
\mathbf{x}^{n o}=\bigcup_{j} \mathbf{x}^{n o_{j}} \quad \mathbf{x}^{n o_{j_{1}}} \cap \mathbf{x}^{n o_{j_{2}}}=\emptyset \tag{3.22}
\end{gather*}
$$

The $\mathbf{x}^{n o_{j}}$ represents the nonoverlapping variable participating in cluster $j$. Thus, the nonoverlapping variables participate in only one of the $m$ clusters, while the overlapping variables can participate in more than one cluster. This participation helps in defining the propagation equation for each cluster. Each overlapping variable $x_{i}^{o}$ is assumed to be a weighted sum of independently distributed auxiliary variables $x_{i}^{o_{j}}$,s, such that

$$
\begin{equation*}
x_{i}^{o}=\sum_{j} x_{i}^{o_{j}} z_{i j} . \tag{3.23}
\end{equation*}
$$

Equation (3.23) is in adherence with (3.14). In a cluster $j$, the overlapping variables and auxiliary variables form the propagating differential equation. These equations can be written differently based on whether the given system is a linear or nonlinear system.
3.5.1. Linear system. A linear system for the state variable $\mathbf{x} \in \mathbb{R}^{n}$ is represented by a constant square matrix of size $n \times n$ as given in (3.24),

$$
\begin{equation*}
\dot{\mathbf{x}}=A \mathbf{x}+b . \tag{3.24}
\end{equation*}
$$

The above equation is decomposed into $m$ independent equations. The overlapping and nonoverlapping variables $\mathbf{x}^{0_{j}}$ and $\mathbf{x}^{n 0_{j}}$ are identified for each cluster from the matrix $Z$. Each cluster $j$ is represented by the following linear equation:

$$
\left[\begin{array}{c}
\dot{\mathbf{x}}^{o_{j}}  \tag{3.25}\\
\dot{\mathbf{x}}^{0_{j}}
\end{array}\right]=A_{j}\left[\begin{array}{c}
\mathbf{x}^{o_{j}} \\
\mathbf{x}^{n o_{j}}
\end{array}\right] .
$$

The matrix $A_{j}$ contains the coefficients of the variables $\mathbf{x}^{o_{j}}$ and $\mathbf{x}^{n o_{j}}$ from (3.24).
3.5.2. Nonlinear system. A general decomposition scheme for a nonlinear system as represented in (3.26) is difficult to achieve:

$$
\begin{equation*}
\dot{\mathrm{x}}=f(\mathbf{x}) \tag{3.26}
\end{equation*}
$$

An easier approach is to formulate the approximate linear system as given in (3.17) and follow the decomposition scheme as given in section 3.5.1. However, such a linear model might not capture all the behaviors of a complex nonlinear system.

To establish a generalized scheme for decomposition of $\mathbf{x}$, the function $f(\mathbf{x})$ is rewritten in each cluster $j$ as

$$
\begin{equation*}
f(\mathbf{x})=f_{j}\left(\mathbf{x}^{o_{j}}, \mathbf{x}^{n o_{j}}\right)+g_{j}\left(\mathbf{x}^{o_{j}}, \mathbf{x}^{n o_{j}}, \mathbf{x}^{n p_{j}}\right)+h_{j}\left(\mathbf{x}^{n p_{j}}\right) \tag{3.27}
\end{equation*}
$$

The propagation equation for cluster $j$ is then written as

$$
\left[\begin{array}{c}
\dot{\mathbf{x}}^{o_{j}}  \tag{3.28}\\
\dot{\mathbf{x}}^{n o_{j}}
\end{array}\right]=f_{j}\left(\mathbf{x}^{o_{j}}, \mathbf{x}^{n o_{j}}\right)+\left.\nabla_{\mathbf{x}^{n p_{j}}} g_{j}\left(\mathbf{x}^{o_{j}}, \mathbf{x}^{n o_{j}}, \mathbf{x}^{n p_{j}}\right)\right|_{\mathbf{x}^{n p_{j}}}=\mathbf{x}^{n p_{j}}(0)
$$

After the simulation of the $m$ independent equations in parallel for either of the linear or nonlinear system, the overall state variable $\mathbf{x}$ is computed from (3.14).
3.6. Hadamard product and multivariate statistics. Once the statistical properties of the SCSs are determined from the propagation and moment equations as given in section 3.5, the first- and second-order moments for a state variable $x_{i}$ under the formulation of (3.13) is given as

$$
\begin{equation*}
E\left[x_{i}\right]=E\left[\sum_{j=1}^{m} z_{i j} x_{i} \mid j\right]=\sum_{j=1}^{m} z_{i j} E\left[x_{i} \mid j\right] \tag{3.29}
\end{equation*}
$$

and

$$
\begin{equation*}
E\left[x_{i}^{2}\right]=E\left[\left(\sum_{j=1}^{m} z_{i j} x_{i} \mid j\right)^{2}\right]=\sum_{j=1}^{m} z_{i j}^{2} E\left[x_{i}^{2} \mid j\right]+\sum_{j=1}^{m-1} \sum_{k=j+1}^{m} z_{i j} z_{i k} E\left[x_{i} \mid j\right] E\left[x_{i} \mid k\right] . \tag{3.30}
\end{equation*}
$$

Thus the covariance formulation is given as

$$
\begin{equation*}
\operatorname{var}\left(x_{i}\right)=\sum_{j=1}^{m} z_{i j}^{2} \operatorname{var}\left(x_{i}\right) . \tag{3.31}
\end{equation*}
$$

Similarly, a general expression for any $\alpha$ th order moment can be written using multinomial expansion as

$$
\begin{align*}
E\left[x_{i}^{\alpha}\right]=E\left[\left(\sum_{j=1}^{m} z_{i j} x_{i} \mid j\right)^{\alpha}\right] & =E\left[\left.\sum_{k_{1}+k_{2}+\cdots+k_{m}=\alpha}\binom{\alpha}{k_{1}, k_{2}, \ldots, k_{m}} \prod_{1 \leq j \leq m} x_{i}^{k_{j}} \right\rvert\, j z_{i j}^{k_{j}}\right.  \tag{3.32}\\
& =\sum_{k_{1}+k_{2}+\cdots+k_{m}=\alpha}\binom{\alpha}{k_{1}, k_{2}, \ldots, k_{m}} \prod_{1 \leq j \leq m} z_{i j}^{k_{j}} \prod_{1 \leq j \leq m} E\left[x_{i}^{\left.k_{j} \mid j\right]} .\right.
\end{align*}
$$

For the state vector $\mathbf{x}_{t}$, the moment equations for the first two order under the formulation of (3.14) are given as

$$
\begin{equation*}
E\left[\mathbf{x}_{t}\right]=E\left[\sum_{j=1}^{m} \mathbf{z}_{j} \odot \mathbf{x}_{t} \mid j\right]=\sum_{j=1}^{m} \mathbf{z}_{j} \odot E\left[\mathbf{x}_{t} \mid j\right] \tag{3.33}
\end{equation*}
$$

and

$$
\begin{align*}
E\left[\mathbf{x}_{t} \mathbf{x}_{t}^{T}\right] & =\sum_{j=1}^{m} E\left[\left(\mathbf{z}_{j} \odot \mathbf{x}_{t} \mid j\right)\left(\mathbf{z}_{j} \odot \mathbf{x}_{t} \mid j\right)^{T}\right]+\sum_{j=1}^{m-1} \sum_{k=j+1}^{m} E\left[\left(\mathbf{z}_{j} \odot \mathbf{x}_{t} \mid j\right)\left(\mathbf{z}_{k} \odot \mathbf{x}_{t} \mid k\right)^{T}\right]  \tag{3.34}\\
& =\sum_{j=1}^{m}\left(\mathbf{z}_{j} \mathbf{z}_{j}^{T}\right) \odot E\left[\mathbf{x}_{t}\left|j \mathbf{x}_{t}\right| j^{T}\right]+\sum_{j=1}^{m-1} \sum_{k=j+1}^{m}\left(\mathbf{z}_{j} \mathbf{z}_{k}^{T}\right) \odot E\left[\mathbf{x}_{t} \mid j\right] E\left[\mathbf{x}_{t} \mid k\right]^{T} .
\end{align*}
$$

Thus, the covariance matrix is given as

$$
\begin{equation*}
\operatorname{cov}\left(\mathbf{x}_{t}\right)=\sum_{j=1}^{m}\left(\mathbf{z}_{j} \mathbf{z}_{j}^{T}\right) \odot \operatorname{cov}\left(\mathbf{x}_{t} \mid j\right) \tag{3.35}
\end{equation*}
$$

Furthermore, the moment generating function (mgf) for $x_{i}$ can be written as

$$
\begin{align*}
M_{x_{i}}(\tau) & =E\left[e^{\tau x_{i}}\right]=E\left[e^{\tau \sum_{j=1}^{m} z_{i j} x_{i} \mid j}\right]=\prod_{j=1}^{m} E\left[e^{\tau\left(z_{i j} x_{i} \mid j\right)}\right]  \tag{3.36}\\
& =\prod_{j=1}^{m} M_{x_{i} \mid j}\left(z_{i j} \tau\right) .
\end{align*}
$$

Hence, the mgf of $\mathbf{x}_{t}$ is given as

$$
\begin{align*}
M_{\mathbf{x}_{t}}(\boldsymbol{\tau}) & =E\left[e^{\boldsymbol{\tau}^{T} \mathbf{x}_{t}}\right]=E\left[e^{\boldsymbol{\tau}^{T} \sum_{j=1}^{m} \mathbf{z}_{j} \odot \mathbf{x}_{t} \mid j}\right]=\prod_{j=1}^{m} E\left[e^{\left(\mathbf{z}_{j} \odot \boldsymbol{\tau}\right)^{T} \mathbf{x}_{t} \mid j}\right]  \tag{3.37}\\
& =\prod_{j=1}^{m} M_{\mathbf{x}_{t} \mid j}\left(\mathbf{z}_{j} \odot \boldsymbol{\tau}\right) .
\end{align*}
$$

In this current work, (3.33) and (3.35) have been used to estimate the statistical properties of the state variable $\mathbf{x}$ obtained from the solution to (3.6).
3.7. Integration with nonintrusive spectral projection methods. The linearization and clustering paradigm can be easily integrated with any of the Non-Intrusive Spectral Projection (NISP) methods of UQ. The representation of the state variable $\mathbf{x} \mid j$ by a multidimensional truncated Polynomial Chaos Expansion (PCE) is given as

$$
\begin{equation*}
\mathbf{x} \mid j=\sum_{k=0}^{P} c_{k} \Phi_{k}\left(\xi_{1}, \ldots, \xi_{n}\right) \tag{3.38}
\end{equation*}
$$

where, $P+1=\binom{n+p}{p}$ are the number of PC points and $p$ is the maximum order of the class of orthogonal polynomial $\Phi$. Once the coefficients $c_{k}$ 's are determined, quadrature-based methods such as full-tensor or sparse-grid collocation methods are applied to generate samples from the orthogonal polynomials $\Phi_{k}$ 's.

The numerical integral computation by quadrature rule involves a generation of quadrature points and associated weights $\left[\mathcal{X}, W_{i}\right]$ 's for each orthogonal polynomial $\Phi_{k}$.
3.8. Recomputation of cluster structure by solving the inverse UQ problem. In this section, the different subroutines are grouped together to describe the overall framework. The linearization and clustering method described in section 3.4 is coupled with the standard UQ method of Unscented Kalman Filter (UKF) [23] to analyze a high-dimensional system defined in section 3.1. UKF has been the preferred method of UQ owing its speed and accuracy in
estimating moments up to the second order of a Gaussian random vector by generating fewer sample points. UKF proposes the following deterministic formulation for generating Sigma Points from a Gaussian random variable $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma), \mathbf{x} \in \mathbb{R}^{n}$ :

$$
\begin{array}{lr}
\mathcal{X}_{0}=\mu, & W_{0}=\kappa /(n+\kappa), \\
\mathcal{X}_{i}=\mu+(\sqrt{(n+\kappa) \Sigma})_{i}, & W_{i}=1 / 2(n+\kappa),  \tag{3.39}\\
\mathcal{X}_{i+n}=\mu-(\sqrt{(n+\kappa) \Sigma})_{i}, & W_{i+n}=1 / 2(n+\kappa),
\end{array}
$$

$(\sqrt{(n+\kappa) \Sigma})_{i}$ is the $i$ th row or column of the matrix square root of $(n+\kappa) \Sigma$. These points accurately estimate the moment of $\mathbf{x}$ up to the second order. These moment equations are given as

$$
\begin{aligned}
\mu & =\sum_{\substack{i=1 \\
2 n+1}} W_{i} \mathcal{X}_{i}, \\
\Sigma & =\sum_{i=1}^{2 n+1} W_{i}\left(\mathcal{X}_{i}-\mu\right)\left(\mathcal{X}_{i}-\mu\right)^{T} .
\end{aligned}
$$

Due to the coupling between the state variables, it is assumed that the cluster structure may change with time. Subsequently, there is a requirement of changing the cluster structure after a finite time evolution of the SCSs. This time is determined by the availability of measurement data. Measurements or observations are obtained from the actual running of the system at different time intervals. The measurement data is used to periodically update the statistical properties of the state variables using the UKF. The updated mean and covariance is used as the uncertainty information required to cluster the state space for the next finite time interval until the next measurement becomes available. All the above mentioned methods are combined to analyze a given large system. The different steps involved are as given below:

1. Given the problem

$$
\begin{align*}
& \dot{\mathbf{x}}_{t}=f\left(\mathbf{x}_{t}\right), \quad \mathbf{x}_{t_{0}}=\mathbf{x}_{0},  \tag{3.40}\\
& \mathbf{z}_{t}=\mathbf{x}_{t}+\nu,
\end{align*}
$$

where $\mathbf{z}_{t}$ is the measurement at a given time, and $\nu \sim \mathcal{N}\left(0, R_{\mathbf{z}}\right)$ is the measurement error.
2. At a time $t_{0}$, the nonlinear system is linearized as

$$
\begin{equation*}
f\left(\mathbf{x}_{t_{k}}\right)=A_{s l} \mathbf{x}_{t_{k}}+b_{s l} \tag{3.41}
\end{equation*}
$$

and clustered as per methods described in section 3.4. At the end of this step, the number of subgroups or clusters in the state variables is identified, along with the association matrix $Z$ as given in Table 3.1.
3. The state-space is decomposed into SCS from the values of the association matrix $Z$ using the method of mapping described in section 3.5.
4. Sigma points are generated using (3.39) from the reduced order SCSs. The submanifold for each SCS is propagated as described in section 3.5 up to a certain time $t_{k}$. The mean and covariance for each SCS are estimated using (3.40).
5. Once the UQ of all the SCSs is complete, the mean and covariance of $\mathbf{x}_{t_{k}}$ are computed using (3.33) and (3.35).
6. Steps 4 and 5 are continued up to a time $t_{k}+h$ until a measurement $\mathbf{z}_{t_{k}+h}$ is available. Once a measurement update is available, the mean and covariance values are updated using the following steps of the UKF:
(a) The innovation covariance is then computed as

$$
\begin{equation*}
P_{\nu \nu}=R_{\mathbf{z}}+P_{k+h \mid k+h-1} \tag{3.42}
\end{equation*}
$$

(b) The cross correlation matrix is computed as $P_{k+h \mid k+h-1}$.
(c) The Kalman gain is computed as

$$
\begin{equation*}
K=P_{k+h \mid k+h-1} P_{\nu \nu}^{-1} . \tag{3.43}
\end{equation*}
$$

(d) The measurement update is carried out as

$$
\begin{align*}
& \mu_{k+h \mid k+h}=\mu_{k+h \mid k+h-1}+K\left(\mathbf{z}_{t_{k}+h}-\mu_{k+h \mid k+h-1}\right), \\
& P_{k+h \mid k+h}=P_{k+h \mid k+h-1}-K P_{\nu \nu} K^{T} . \tag{3.44}
\end{align*}
$$

(e) The above defined recalibration of mean and covariance helps in updating the cluster structure of $\mathbf{x}$. Using the updated statistical properties and the velocity function $f(\cdot)$, the association matrix $Z$ is recomputed using the methods described in section 3.4. The state variables are further mapped into clusters using the method of mapping (section 3.5). Steps 2-5 are repeated, and step 6 is used when a measurement is available. These steps are repeated in iterative fashion until the whole analysis is complete.
4. Numerical example: High dimensional diffusively coupled Van der Pol oscillators. The state space equations of weakly coupled Van der Pol Oscillators is given as [49]

$$
\begin{align*}
& \dot{x}_{i}=y_{i}+\epsilon\left(x_{i-1}-2 x_{i}+x_{i+1}\right), \quad i=1,2, \ldots, N .  \tag{4.1}\\
& \dot{y}_{i}=\mu\left(1-x_{i}^{2}\right) y_{i}-x_{i},
\end{align*}
$$

The system in (4.1) is linearized and the state space is clustered following the methods described in sections 3.4 and 3.5. With $e=0$, the equilibrium point for any set of parameters for an individual oscillator is $[00]^{T}$. With slight perturbation, the system repels from the equilibrium point and orients itself around two steady convections given. To study the effectiveness of our proposed methodology, 42 test cases of the problem in (4.1) are considered by taking seven values of $\epsilon=\{0.1,1,5,10,20,50,100\}$ and six values of $N=\{5,10,25,50,100,250\}$. The value of the $\epsilon$ is kept very low to introduce a weak coupling between the oscillators. Strong coupling induces change in the behavior of the individual oscillator. The changing cluster structure needs to be identified with time to avoid accumulation of error. Uncertainties in the
state variables $\left\{\dot{x}_{i}, \dot{y}_{i}\right\}$ are introduced by random sampling from $\mathcal{N}(0,1)$. Table A. 1 shows the average mean of error in propagation obtained for our proposed UQ method for some of the test cases.
4.1. Computational advantage in integration with quadrature-based methods. The computational advantage of the SCS-based decomposition scheme can be leveraged if the underlying UQ methodology used has an exponential or high-order polynomial complexity. Quadrature (or Cubature) based methods [2, 47] involve a deterministic scheme to generate the pair of sigma points or collocation points $\left[\mathcal{X}, W_{i}\right]$ similar to (3.39) depending on the type of probability distribution function [1]. The high-dimensional cubature methods such as Gaussian cubature [2], sparse-grid cubature [19], and MC [13] requires the generation of collocation points $\mathcal{N}$ to exactly compute moments of multivariate polynomial functions up to order $d$. For UKF formulation given in (3.39), $N=2 n+1$ for $d=3$. Table 4.1 down the number of collocation points $N$ required to solve the high-dimensional UQ problem for the system given in (4.1) for Gaussian cubature and sparse-grid collocation method for $d=3$. Also, the number of collocation points required for SCS-based decomposition is also listed.

Table 4.1
Collocation points required for Gaussian Cubature, Sparse-grid collocation along with SCS-based decomposition.

| Size of the problem | Gaussian Quadrature (GQ) | Sparse-grid Collocation (SG) | Number of clusters | GQ with SCS | SG with SCS |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 10 | 32768 | 286 | 5 | 20 | 50 |
| 20 | $1.049 \mathrm{E}+06$ | 1771 | 10 | 40 | 100 |
| 50 | $1.126 \mathrm{E}+15$ | 23426 | 25 | 250 |  |
| 100 | $1.268 \mathrm{E}+30$ | 176851 | 50 | 200 |  |

5. Numerical example: Shallow water equation. The tidal water flow in a long narrow channel with varying bathymetric depth (Figure 5.1) is modeled by the Shallow Water Equations (SWE) given as [50]

$$
\begin{align*}
\frac{\partial h}{\partial t}+D \frac{\partial u}{\partial x}+\frac{\partial D}{\partial x} u & =0 \\
\frac{\partial u}{\partial t}+g \frac{\partial h}{\partial x}+c_{f} u & =0 \\
h(x=0, t) & =h_{b}(t),  \tag{5.1}\\
h(x, t=0) & =0 \\
u(x, t=0) & =0 \\
h(x=L, t) & =0,
\end{align*}
$$

where $h$ denotes the water surface and $u$ is the flow velocity along the channel. The model is used to study storm surges and has been researched in details to study the effect of such surges in long narrow channels [50]. The water wave is under the influence of thesurge $h(0, t)=h_{b}(t)$.


Figure 5.1. Schematic of the Tidal Water Flow.

The slope is assumed as $\frac{\partial D}{\partial x}=\theta(x)$ to further simplify the system. To solve the problem, (5.1) is discretized as per the Leendertse and Stelling scheme [29, 45, 50, 52] as follows:

$$
\begin{align*}
\frac{h_{i}^{k+1}-h_{i}^{k}}{\triangle t}+\frac{1}{2} D_{i} \frac{u_{i+\frac{1}{2}}^{k}-u_{i-\frac{1}{2}}^{k}}{\triangle x}+\frac{1}{2} D_{i} \frac{u_{i+\frac{1}{2}}^{k+1}-u_{i-\frac{1}{2}}^{k+1}}{\triangle x}+\frac{1}{2} \theta_{i}\left(u_{i+\frac{1}{2}}^{k}+u_{i+\frac{1}{2}}^{k+1}\right) & =0, \\
\frac{u_{i+\frac{1}{2}}^{k+1}-u_{i+\frac{1}{2}}^{k}}{\triangle t}+\frac{1}{2} g \frac{h_{i+1}^{k}-h_{i}^{k}}{\triangle x}+\frac{1}{2} g \frac{h_{i+1}^{k+1}-h_{i}^{k+1}}{\triangle x}+\frac{1}{2} c_{f} u_{i+\frac{1}{2}}^{k}+\frac{1}{2} c_{f} u_{i+\frac{1}{2}}^{k+1} & =0,  \tag{5.2}\\
h_{0}^{k}-h_{b}(k \triangle t) & =0, \\
h_{i}^{0} & =0, \\
u_{i}^{0} & =0, \\
u_{N+\frac{1}{2}}^{k} & =0 .
\end{align*}
$$

Equation (5.2) is recast as a state space equation given by

$$
\begin{align*}
& \bar{D} \mathbf{x}_{k+1}=\bar{A} \mathbf{x}_{k}+\bar{B} \mathbf{u}_{k}, \\
& \mathbf{h}_{x+1}=\bar{C} \mathbf{x}_{k+1}, \tag{5.3}
\end{align*}
$$

where

$$
\begin{align*}
& \mathbf{x}_{k}=\left[\begin{array}{c}
h_{0}^{k} \\
u_{\frac{1}{2}}^{k} \\
h_{1}^{k} \\
u_{1+\frac{1}{2}}^{k} \\
\vdots \\
h_{N}^{k}
\end{array}\right], \quad \mathbf{h}_{k}=\left[\begin{array}{c}
h_{0}^{k} \\
h_{1}^{k} \\
\vdots \\
h_{N}^{k}
\end{array}\right], \\
& \begin{aligned}
\bar{D} & =\left[\begin{array}{ccccccc}
1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
-\frac{g}{2 \Delta x} & \frac{1}{\Delta t}+\frac{1}{2} c_{f} & \frac{g}{2 \Delta x} & 0 & 0 & \cdots & 0 \\
0 & -\frac{D_{i}}{2 \Delta x} & \frac{1}{\Delta t} & \frac{D_{i}}{2 \Delta x}+\frac{1}{2} \theta_{i} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\
0 & 0 & 0 & 0 & -\frac{D_{i}}{2 \Delta x} & \frac{1}{\Delta t} & \frac{D_{i}}{2 \Delta x}+\frac{1}{2} \theta_{i} \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right],
\end{aligned} \\
& \bar{A}=\left[\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
\frac{g}{2 \Delta x} & \frac{1}{\Delta t}-\frac{1}{2} c_{f} & -\frac{g}{2 \Delta x} & 0 & 0 & \ldots & 0 & \\
0 & \frac{D}{2 \Delta x} & \frac{1}{\Delta t} & -\frac{D}{2 \Delta x}-\frac{1}{2} \theta_{i} & 0 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\
0 & 0 & 0 & 0 & \frac{D}{2 \Delta x} & \frac{1}{\Delta t} & -\frac{D}{2 \Delta x}-\frac{1}{2} \theta_{i} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 &
\end{array}\right],  \tag{5.4}\\
& \bar{B}=\left[\begin{array}{c}
1 \\
0 \\
0 \\
\vdots \\
0 \\
0
\end{array}\right], \\
& \bar{C}=\left[\begin{array}{cccccccc}
1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0
\end{array}\right] .
\end{align*}
$$

5.1. Results for a deterministic problem setup. Reviewing the results for a deterministic case, the constant slope $\theta$ in (5.3) is set to 0 . The height of the wave at the left end is provided as an input to the equation at each time instance. The water level rises with time and the wave moves along the stretch of the channel. It then dissipates with time once the maximum height is reached. The discretized linear model in (5.3) is clustered by the method explained in section 3.5. The adjacency matrix used in this case is $A_{s l}=\bar{D}^{-1} \bar{A}$. The parameters in (5.3) are assumed to be as follows:

- Length of the channel $L=60 \mathrm{~km}$.
- Constant Water Depth $D=10 \mathrm{~m}$.
- Friction constant $c_{f}=0.0002$ in $1 / \mathrm{sec}$.
- Total time $T=240 \mathrm{~min}$.

The discrete model in (5.2) is scaled using a spatial discretization of $N=80$, with $\triangle x=$ $L /(N+0.5)$ and a temporal discretization of $\Delta t=300 \mathrm{~s}$. Thus, the number of time steps is 80. Figure 5.2 shows the analysis of (5.3) using the full model and the clustered model given by the method of clustering. For the given test setup, the clustered model shows a visual accuracy as compared to the true simulation.


Figure 5.2. True versus Estimated solution using SCS-based decomposition for a deterministic test case.
In the subsequent section, (5.3) is solved for the same time limit of $T=400 \mathrm{mins}$ and the length of the channel $L=60 \mathrm{~km}$. The number of discrete spatial points $N$ has been varied from 150 to 1200, and temporal discrete points Tstep $=T / \Delta t$ has been ranged from 60 to 1200. To test the accuracy of the discrete model for solving the physical system, the stability and convergence of the discrete model is tested for the above resolutions of discretization.
5.2. Stability conditions. The stability conditions for a discretized model of (5.10) is defined by the Von Neumann theory [9]. This condition puts a restriction on the degree of discretization for (5.2). To derive the condition, the following nondimensional variables are
defined:

$$
\begin{equation*}
\tilde{x}=\frac{x}{L}, \quad \tilde{t}=\frac{t}{T}, \quad \tilde{h}=\frac{h}{E}, \quad \tilde{u}=\frac{u}{U} . \tag{5.5}
\end{equation*}
$$

Thereafter, the following relations are defined:

$$
\begin{equation*}
L^{2}=D g T^{2}, \quad E^{2}=\frac{D U^{2}}{g}, \quad T=\frac{\beta}{c_{f}}, \quad \frac{T U}{E}=\alpha \tag{5.6}
\end{equation*}
$$

Introducing (5.5) and (5.6), we see that (5.1) yields the set of the following dimensionless equations:

$$
\begin{align*}
\frac{\partial \tilde{h}}{\partial \tilde{t}}+\frac{\partial \tilde{u}}{\partial \tilde{x}}+\theta \alpha \tilde{u} & =0 \\
\frac{\partial \tilde{u}}{\partial \tilde{t}}+\frac{\partial \tilde{h}}{\partial \tilde{x}}+\beta \tilde{u} & =0 \tag{5.7}
\end{align*}
$$

The solution to (5.7) is expressed in terms of discrete Fourier modes as

$$
\begin{align*}
& \tilde{h}_{i}^{k}=E^{k} \exp \left(j k_{x} i \triangle x\right),  \tag{5.8}\\
& \tilde{u}_{i+\frac{1}{2}}^{k}=U^{k} \exp \left(j k_{x}\left(i+\frac{1}{2}\right) \triangle x\right), \quad j^{2}=-1
\end{align*}
$$

Putting the expression for $\tilde{h}_{i}^{k}$ and $\tilde{u}_{i+1}^{k}$ into the dimensionless form of (5.2), we get

$$
\begin{align*}
& E^{k+1}-E^{k}=-j \frac{\Delta t}{\Delta x}\left(U^{k}+U^{k+1}\right) \sin \left(\frac{k_{x} \Delta x}{2}\right)-\theta \alpha \frac{1}{2} \triangle t\left(U^{k}+U^{k+1}\right) \exp \left(j \frac{k_{x} \Delta x}{2}\right) \\
& U^{k+1}-U^{k}=-j \frac{\Delta t}{\Delta x}\left(E^{k}+E^{k+1}\right) \sin \left(\frac{k_{x} \Delta x}{2}\right)-\beta \frac{1}{2} \triangle t\left(U^{k}+U^{k+1}\right) \tag{5.9}
\end{align*}
$$

Equation (5.9) is expressed in the matrix form as

$$
\begin{equation*}
\binom{E^{k+1}}{U^{k+1}}=A\binom{E^{k}}{U^{k}} \tag{5.10}
\end{equation*}
$$

where

$$
\begin{align*}
& A=c\left[\begin{array}{cc}
-j \exp \left(j \frac{k_{x} \Delta x}{2}\right) \sin \left(\frac{k_{x} \Delta x}{2}\right) \alpha \Delta t^{2} \theta \Delta x+2 \sin ^{2}\left(\frac{k_{x} \Delta x}{2}\right) \Delta t^{2}-\beta \Delta t \Delta x^{2}-2 \Delta x^{2} & 2 \theta \alpha \Delta t \exp \left(j \frac{k_{x} \Delta x}{2}\right) \Delta x^{2}+4 j \Delta t \Delta x \sin \left(\frac{k_{x} \Delta x}{2}\right) \\
4 j \Delta t \Delta x \sin \left(\frac{k_{x} \Delta x}{2}\right) & -j \exp \left(j \frac{k_{x} \Delta x}{2}\right) \sin \left(\frac{k_{x} \Delta x}{2}\right) \alpha \Delta t^{2} \theta \Delta x+2 \sin ^{2}\left(\frac{k_{x} \Delta x}{2}\right) \Delta t^{2}+\beta \Delta t \Delta x^{2}-2 \Delta x^{2}
\end{array}\right],  \tag{5.11}\\
& c=\frac{1}{j \exp \left(j \frac{k_{x} \triangle x}{2}\right) \sin \left(\frac{k_{x} \triangle x}{2}\right) \alpha \triangle t^{2} \theta \triangle x-2 \sin ^{2}\left(\frac{k_{x} \triangle x}{2}\right) \triangle t^{2}-\beta \triangle t \triangle x^{2}-2 \triangle x^{2} .}
\end{align*}
$$

By Von Neumann theory of stability analysis [9], for a given resolution of spatio-temporal discretization or a combination of $\triangle x$ and $\Delta t$, the system is stable if the norm of the matrix $A$ is less than 1. The discrete model in (5.2) has been found to be stable for all the combinations for $N$ and $\Delta t$ that have been chosen in subsection 5.1.
5.3. Convergence analysis. Once the discrete model is determined to be stable for a varying resolutions of discretization, the convergence analysis is performed. The discrete model is supposed to converge to the exact solution with increase in resolution. Since there is no analytic solution available, a very high resolution discretization with $N=3200$ and $\Delta t=15 \mathrm{~s}$ is chosen. The true solution for the height of the water level is given by a $1601 \times 3201$ matrix $U_{\text {true }}$. This matrix records the height at each point along the $x$-direction for each time instance. For a given resolution of discretization $r=(\Delta t, \Delta x)$, the same matrix $U_{r}$ is computed and upscaled it to the resolution of the matrix $U_{\text {true }}$ as $\hat{U}_{r}$. The rate of convergence for $r$ is given as

$$
\begin{equation*}
e r r_{r}=\left\|\frac{U_{\text {true }}-\hat{U}_{r}}{U_{\text {true }}}\right\|_{2} . \tag{5.12}
\end{equation*}
$$

Figure 5.3 depicts the rate of convergence for different resolutions of discretization. The convergence rate decreases with increase in resolution of the discrete model. Following the result of convergence and stability, it can be concluded that the discrete model in (5.2) can approximate the solution to the continuous model in (5.1).
5.4. Cluster length versus resolution of discretization. In this section, the cluster structures for different test setups are determined using the methodology discussed in section 3.4 to give a better physical interpretation of the state-space clusters. The state space matrix $D^{-1} A$ as per given in (5.4) is clustered. It is to be noted that for a given discrete point $x_{k}$ along the spatial domain, there are two state variables $u_{i}$ and $h_{i+\frac{1}{2}}$. The clustering on the $D^{-1} A$ gives us the association vectors $\mathbf{z}_{j}$ 's of the state space $\mathbf{x}$ in $m$ clusters, as discussed in section 3.4. To have a meaningful insight into the physical meaning of the clusters, the length of the first cluster has been chosen as the subject of interest. The length of the first cluster is given as

$$
\mathcal{L}_{1}=\frac{\sum_{i=1}^{2(N+1)} \mathbb{I}_{1}\left(z_{1 i}\right)}{2(N+1)} L, \quad \mathbb{I}_{1}\left(z_{1 i}\right)= \begin{cases}1 & z_{1 i}=1  \tag{5.13}\\ 0 & \text { otherwise }\end{cases}
$$

The length $\mathcal{L}_{1}$ in kilometer represents the physical space along the channel length. Since both $D$ and $A$ are functions of $\triangle t$ and $\triangle x$, it can be assumed that $\mathcal{L}_{1}$ is also a function of both $\Delta t$ and $\triangle x$. Variation in the cluster lengths with different resolutions of discretizations is illustrated in Figures 5.4 and 5.5.

Figure 5.4 shows the variation of the cluster length by varying $N=80$ to 200 for a constant $\triangle=300 \mathrm{~s}$. The plots show that the cluster length is almost constant for a given $\Delta t$ and is unaffected by the change in $N$. The cluster lengths represent the physical space in which the waves can travel for a given time independent of another cluster. This length is bound to change with an increase in $\Delta t$ or decrease in Tstep.

Figure 5.5 shows a very small variation in the values of $\mathcal{L}_{1}$ by varying $\triangle=75$ to 300 s for a constant $N=80$. The plots show that the cluster length varies different values of $\triangle t$. Detailed variation of the length $\mathcal{L}_{1}$ with change in $r$ is shown in Figure 5.6. This analysis provides an insight into the physical meaning of the clusters. The clusters correspond to decomposed domain that can be analyzed in parallel. The variation of the cluster lengths are consistent


Figure 5.3. Rate of Convergence versus Discretization Resolution.
with the actual physical system. The subsequent UQ is carried out on the clustered model of (5.3) rather than the whole system.
5.5. Uncertainties associated with the model. The bathymetric height of the system is assumed to be random for the purpose of uncertainty analysis. The height and subsequently the water depth $D(x)$ is assumed to be a Gaussian random field in one dimension. The mean of $D(x)$ is assumed to have a bimodal profile shown in Figure 5.7 and the covariance function is assumed to be given as

$$
\begin{equation*}
\mathbf{C}(x, y)=\exp \left(-\frac{|x-y|}{a}\right) . \tag{5.14}
\end{equation*}
$$

$D(x)$ admits a spectral decomposition as

$$
\begin{equation*}
D(x)=\mu(x)+\sum_{i=1}^{\infty} \sqrt{\lambda_{i}} \psi_{i}(x) D_{i}(x) \tag{5.15}
\end{equation*}
$$

where $Y \sim \mathcal{N}(0,1)$ are i.i.d. Gaussian random variables and $\lambda_{i}, \psi(x)$ is the solution to the


Figure 5.4. Variation of Cluster length with varying spatial discretization and constant temporal discretization of $\triangle t=300 \mathrm{~s}$.
eigenvalue problem

$$
\begin{equation*}
\int_{0}^{L} \mathbf{C}\left(x_{1}, x_{2}\right) \psi_{i}\left(x_{2}\right) d x_{2}=\lambda_{i} \psi_{i}\left(x_{1}\right) . \tag{5.16}
\end{equation*}
$$

The expansion of (5.15) is truncated according to the decay of the eigenvalues $\left\{\lambda_{i}\right\} \mathrm{s}$. The random field $D(x)$ is expressed in terms of the first $M$ eigenvalues as

$$
\begin{equation*}
D(x)=\mu(x)+\sum_{i=1}^{M} \sqrt{\lambda_{i}} \psi_{i}(x) D_{i}(x) . \tag{5.17}
\end{equation*}
$$

The eigenvalue trend for $a=0.1 L$ is depicted in Figure 5.8. This trend helps us in truncating the expression for KL expansion. From the figure, $M$ can be approximated as 12 .


Figure 5.5. Variation of Cluster length with constant spatial discretization $N=80$ and varying temporal discretization.


Figure 5.6. Cluster Length versus Discretization Resolution.


Figure 5.7. Mean of the Bathymetric profile.


Figure 5.8. Trend of Eigenvalue for exponential covariance function with $a=0.1 L$.

Measurements for $\mathbf{h}_{k}$ are generated using the deterministic model explained in the section 5.1. The following section details the method of clustering and UQ applied to estimate the uncertainties in the state $\mathbf{h}_{k}$ and $\mathbf{u}_{k}$.
5.5.1. Clustering and uncertainty quantification. The state space of the model includes the discretized height of the water level $\mathbf{h}_{k}$ and the velocity $\mathbf{u}_{k}$. The uncertainty in $\mathbf{h}_{k}$ and $\mathbf{u}_{k}$ is attributed to $D(x)$ and the initial state uncertainty. To cluster the state space, the statistically linearized matrix is obtained from the state space (5.3) as follows:.

$$
\begin{equation*}
A_{s l}=\int_{0}^{L} \bar{D}^{-1}(D(x)) \bar{A}(D(x)) d P_{D(x)} . \tag{5.18}
\end{equation*}
$$

For simplicity, the linearized matrix is assumed as $A_{s l}=\bar{D}^{-1}(E[D(x)]) \bar{A}(E[D(x)])$. For $N=80$ and $\triangle t=300 \mathrm{~s}$, the cluster structure of $A_{s l}$ is shown in Figure 5.9. The clustering output results in five almost equal sized clusters. The clustering decomposes the one-dimensional domain into five overlapping domains $D_{j}, j=1,2, \ldots, 5$. Hence, the bathymetric profile is also decomposed into five random fields.


Figure 5.9. Cluster Structure for Shallow Water Model for uncertain bathymetry.
The effect of clustering the state space is reflected from the eigenvalue plot of the clustered random field. Figure 5.10 shows the eigenvalue plot for the bathymatry random field clustered into five overlapping fields. It is observed that each cluster requires a lesser number of eigenvalue for the truncated expansion for the random field. For each cluster $j$, the bathymetry
for the domain $D_{j}$ can be expressed as

$$
\begin{equation*}
D_{j}(x)=f(x)+\sum_{i=1}^{M_{j}} \sqrt{\lambda_{i}} \psi_{i}(x) D_{i}(x), \quad x \in D_{j} \tag{5.19}
\end{equation*}
$$

The value of $M_{j}$ is approximately 4 for each cluster. This reduces the total number of samples required for the random field generated by any standard quadrature method that is used to estimate high-order moments. The state variables corresponding to the water level $\mathbf{h}_{k}$ and the velocity $\mathbf{u}_{k}$ are also clustered. The number of states in each cluster can be estimated from Figure 5.9.


Figure 5.10. Eigenvalue plot for the bathymetry random field clustered into five overlapping fields.

The corresponding filtering problem is solved as described in section 3.8. The effect of filtering is displayed by plotting the observed and estimated height with $\pm 3 \sigma$ limit. Figure 5.11 shows the results of filtering out noise with the availability of measurement.

(a) Time $=300 \mathrm{~s}$.

(b) Time $=9300 \mathrm{~s}$.

(c) Time $=21300 \mathrm{~s}$.

Figure 5.11. Results of the filtering problem. The noise is filtered out with availability of measurement for different time.
6. Conclusion. This paper formulates a clustering method to identify Strongly Coupled Subsystems (SCSs) in high-dimensional uncertain dynamical systems. A new method of clustering the state-space is developed based on overlapping community detection algorithms. The decomposition of state-space based on outlined clustering methods facilitates effective UQ of a high-dimensional system with fewer sample points with the help of any standard UQ method. The method of linearization and clustering shows to be very effective in identifying the common states. The propagation method gives low error in estimated statistical property. The SWE model is carefully chosen for analysis in which the input at one end is propagated throughout the model. Hence, a proper communication between the clusters is essential to have the continuous flow of information in between two neighboring clusters over time. This communication is established very well by the SCS-based decomposition method.

The decision to adopt this clustering method should be taken based on the problem and physical properties of the system. It is also shown that the method drastically reduces the computational cost when integrated with quadrature-based methods while maintaining the accuracy. This makes the method suitable to coupled oscillator problems with varying coupling strengths and scalable problem sizes.

The accuracy of the element-to-element or Hadamard product lies in the assumption of linear relationship of the estimated association values with the adjacency matrix and the statistical independence of the clusters. Thus, the method accurately estimates the deterministic as well as uncertain flow in the Shallow Water model because of the linearity in the original model of (5.2). For a nonlinear model, the association values are to be periodically updated with availability of measurement, or after a certain period of time.

Finally, the clustering method is proved to be effective for not only state-space decomposition but also for clustering a random field. Each cluster of a random field is shown to require a lesser number of eigenvalues for approximate KL expansion than the whole field. Thus, the clustering not only aids in reduced sample points for the state space, but also for the random field. The extension of this study to higher-order random fields (two dimensional and three-dimensional) is the scope for future work.

Appendix A. Results of numerical experiment. We begin with Table A.1.

Table A. 1
Root Mean Squared time-averaged mean for different test setups in Coupled Van der Pol Oscillators.

| $N$ | $\epsilon$ | $\bar{e}_{\mu}$ |
| :---: | :---: | :---: |
| 5 | 0.1 | 0.0019 |
|  | 1 | 0.00339 |
|  | 5 | 0.00323 |
| 10 | 0.1 | 0.00082 |
|  | 1 | 0.00473 |
|  | 5 | 0.01282 |
|  | 20 | 0.01937 |
|  | 100 | 0.02112 |
| 25 | 0.1 | 0.00046 |
|  | 10 | 0.01053 |
|  | 20 | 0.01096 |
|  | 50 | 0.01131 |
| 50 | 0.1 | 0.00028 |
|  | 1 | 0.00153 |
|  | 5 | 0.00394 |
|  | 10 | 0.00519 |
|  | 20 | 0.00539 |
|  | 50 | 0.00558 |
| 100 | 1 | 0.00072 |
|  | 5 | 0.00187 |
|  | 10 | 0.00241 |
|  | 50 | 0.0026 |
|  | 100 | 0.00264 |
| 250 | 0.1 | 0.00005 |
|  | 1 | 0.00001 |
|  | 5 | 0.00027 |
|  | 10 | 0.00094 |

## Appendix B. Some useful results.

B.1. Result 1. Let us consider $n$-dimensional Gaussian random vector $\mathbf{y} \in \mathbb{R}^{n}$ with $\mathbf{y} \sim$ $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and mgf as $M_{\mathbf{y}}(\boldsymbol{\tau})=\exp \left(\mu^{\prime} \boldsymbol{\tau}+\frac{1}{2} \boldsymbol{\tau}^{\prime} \Sigma \boldsymbol{\tau}\right)$. Let us consider an arbitrary $n$-dimensional vector $\alpha$. Under Hadamard product transformation $\mathbf{z}=\alpha \odot \mathbf{y}$, we get

$$
\begin{align*}
M_{\mathbf{z}}(\boldsymbol{\tau})=E\left[e^{\boldsymbol{\tau}^{T}(\alpha \odot \mathbf{y})}\right] & =M_{\mathbf{y}}(\alpha \odot \boldsymbol{\tau})=\exp \left(\mu^{\prime}(\alpha \odot \boldsymbol{\tau})+\frac{1}{2}(\alpha \odot \boldsymbol{\tau})^{\prime} \Sigma(\alpha \odot \boldsymbol{\tau})\right)  \tag{B.1}\\
& =\exp \left((\alpha \odot \mu)^{\prime} \boldsymbol{\tau}+\frac{1}{2} \boldsymbol{\tau}^{\prime}\left(\left(\alpha^{\prime} \alpha\right) \odot \Sigma\right) \boldsymbol{\tau}\right) .
\end{align*}
$$

Thus, $\mathbf{z}$ is also a Gaussian random vector with $\mathbf{z} \sim \mathcal{N}\left(\alpha \odot \mu,\left(\alpha^{\prime} \alpha\right) \odot \Sigma\right)$.
B.2. Result 2. Let us consider a random vector $\mathbf{y} \in \mathbb{R}^{n}$, whose pdf can be approximated by a Gaussian mixture

$$
\begin{equation*}
p(\mathbf{y})=\sum_{i} w_{i} \mathcal{N}\left(\mathbf{y} \mid \mu_{i}, \Sigma_{i}\right) . \tag{B.2}
\end{equation*}
$$

The mgf of $\mathbf{y}$ is given as

$$
\begin{align*}
M_{\mathbf{y}}(\boldsymbol{\tau})=E\left[e^{\boldsymbol{\tau}^{T} \mathbf{y}}\right] & =\int_{\mathbb{R}^{n}} e^{\boldsymbol{\tau}^{T} \mathbf{y}} \sum_{i} w_{i} \mathcal{N}\left(\mathbf{y} \mid \mu_{i}, \Sigma_{i}\right) d \mathbf{y} \\
& =\sum_{i} w_{i} \int_{\mathbb{R}^{n}} e^{\boldsymbol{\tau}^{T} \mathbf{y}} \mathcal{N}\left(\mathbf{y} \mid \mu_{i}, \Sigma_{i}\right) d \mathbf{y}  \tag{B.3}\\
& =\sum_{i} w_{i} M_{\mathbf{y}}\left(\boldsymbol{\tau} ; \mu_{i}, \Sigma_{i}\right) .
\end{align*}
$$

## REFERENCES

[1] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series 55, U.S. Government Printing Office, Washington, D.C., 1964.
[2] I. Arasaratnam and S. Haykin, Cubature Kalman filters, IEEE Trans. Automat. Control, 54 (2009), pp. 1254-1269.
[3] D. A. Barajas-Solano and D. M. Tartakovsky, Stochastic collocation methods for nonlinear parabolic equations with random coefficients, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 475-494, https: //doi.org/10.1137/130930108.
[4] J. Baumes, M. K. Goldberg, M. S. Krishnamoorthy, M. Magdon-Ismail, and N. Preston, Finding communities by clustering a graph into overlapping subgraphs, IADIS AC, 5 (2005), pp. 97104.
[5] P. Bratley and B. L. Fox, Algorithm 659: Implementing Sobol's quasirandom sequence generator, Trans. Math. Software, 14 (1988), pp. 88-100.
[6] R. E. Caflisch, Monte Carlo and quasi-Monte Carlo methods, Acta Numer. 7 (1998), pp. 1-49.
[7] X.-C. CAI, Additive Schwarz algorithms for parabolic convection-diffusion equations, Numer. Math., 60 (1991), pp. 41-61.
[8] R. Cazabet, F. Amblard, and C. Hanachi, Detection of overlapping communities in dynamical social networks, in Proceedings of the 2010 IEEE Second International Conference on Social Computing (SocialCom), IEEE, 2010, pp. 309-314.
[9] J. G. Charney, R. Fjörtoft, and J. V. Neumann, Numerical integration of the barotropic vorticity equation, Tellus, 2 (1950), pp. 237-254.
[10] D. Chen, M. Shang, Z. Lv, and Y. Fu, Detecting overlapping communities of weighted networks via a local algorithm, Phys. A, 389 (2010), pp. 4177-4187.
[11] M. Chen, K. Kuzmin, and B. K. Szymanski, Extension of modularity density for overlapping community structure, in Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2014, pp. 856-863.
[12] H. Сho, X. Yang, D. Venturi, and G. E. Karniadakis, Algorithms for propagating uncertainty across heterogeneous domains, SIAM J. Sci. Comput., 37 (2015), pp. A3030-A3054, https://doi.org/ 10.1137/140992060.
[13] R. Cools, Monomial cubature rules since stroud a compilation part 2, J. Comput. Appl. Math., 112 (1999), pp. 21-27.
[14] R. Cools, An encyclopaedia of cubature formulas, J. Complexity, 19 (2003), pp. 445-453.
[15] J. C. Dunn, Well-separated clusters and optimal fuzzy partitions, J. Cybernet., 4 (1974), pp. 95-104.
[16] I. Farkas, D. Ábel, G. Palla, and T. Vicsek, Weighted network modules, New J. Phys., 9 (2007), 180.
[17] G. S. Fishman, Monte Carlo: Concepts, Algorithms, and Applications, Springer Series in Operations Research, Springer, New York, 1996.
[18] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell., 6 (1984), pp. 721-741.
[19] T. Gerstner and M. Griebel, Numerical integration using sparse grids, Numer. Algorithms, 18 (1998), pp. 209-232.
[20] G. Grinstein, C. Jayaprakash, and Y. He, Statistical mechanics of probabilistic cellular automata, Phys. Rev. Lett., 55 (1985), pp. 2527-2530.
[21] F. Havemann, M. Heinz, A. Struck, and J. Gläser, Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels, J. Stat. Mech: Theory Exp., 01 (2011), P01023.
[22] R. L. Iman, Latin hypercube sampling, Encyclopedia of Quantitative Risk Analysis and Assessment, John Wiley \& Sons, Hoboken, NJ, 2008, pp. 1-5.
[23] S. J. Julier and J. K. Uhlmann, New extension of the Kalman filter to nonlinear systems, in Proceedings of the 11th International Sysmpsium on Aerospace / Defence Sensing, Simulation and Controls, Volume 3068, Orlando, FL, 1997, pp. 182-193.
[24] D. Kohler, J. Müller, and U. Wever, Cellular probabilistic automata-A novel method for uncertainty propagation, SIAM/ASA J. Uncertainty Quantification, 2 (2014), pp. 29-54, https://doi.org/ 10.1137/120897183
[25] J. M. Kumpula, M. Kivelä, K. Kaski, and J. Saramäki, Sequential algorithm for fast clique percolation, Phys. Rev. E, 78 (2008), 026109.
[26] A. Lancichinetti, S. Fortunato, and J. Kertész, Detecting the overlapping and hierarchical community structure in complex networks, New J. Phys., 11 (2009), 033015.
[27] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, Finding statistically significant communities in networks, PloS one, 6 (2011), e18961.
[28] P. Latouche, E. Birmelé, and C. Ambroise, Overlapping stochastic block models with application to the French political blogosphere, Ann. Appl. Stat., 5 (2011), pp. 309-336.
[29] J. J. Leendertse, Aspects of a Computational Model for Long-period Water-wave propagation, Ph.D. thesis, Rand Corporation, Santa Monica, CA, 1967.
[30] Q. Liao and K. Willcox, A domain decomposition approach for uncertainty analysis, SIAM J. Sci. Comput., 37 (2015), pp. A103-A133, https://doi.org/10.1137/140980508.
[31] M. Magdon-Ismail and J. Purnell, SSDE-cluster: Fast overlapping clustering of networks using sampled spectral distance embedding and GMMs, in Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), IEEE, 2011, pp. 756-759.
[32] R. Melchers, Importance sampling in structural systems, Structural Safety, 6 (1989), pp. 3-10.
[33] A. Mukherjee, R. Rai, P. Singla, T. Singh, and A. Patra, Laplacian graph based approach for uncertainty quantification of large scale dynamical systems, in American Control Conference (ACC),

2015, IEEE, 2015, pp. 3998-4003.
[34] A. Mukherjee, R. Rai, P. Singla, T. Singh, and A. Patra, Non-negative matrix factorization based uncertainty quantification method for complex networked systems, in ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2015, V02AT03A048.
[35] A. Mukherjee, R. Rai, P. Singla, T. Singh, and A. K. Patra, Comparison of linearization and graph clustering methods for uncertainty quantification of large scale dynamical systems, Int. J. Uncertain. Quantif., 7 (2017), pp. 23-56.
[36] A. Padrol-Sureda, G. Perarnau-Llobet, J. Pfeifle, and V. Muntés-Mulero, Overlapping community search for social networks, in Proceedings of the 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010), IEEE, 2010, pp. 992-995.
[37] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature, 435 (2005), pp. 814-818.
[38] A. Papanicolaou and K. Spiliopoulos, Dimension reduction in statistical estimation of partially observed multiscale processes, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 1220-1247, https: //doi.org/10.1137/16M1085930.
[39] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon, Overlapping community detection using Bayesian non-negative matrix factorization, Phys. Rev. E, 83 (2011), 066114.
[40] S. Rahman, A generalized anova dimensional decomposition for dependent probability measures, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 670-697, https://doi.org/10.1137/120904378.
[41] J. Roberts and P. Spanos, Stochastic averaging: An approximate method of solving random vibration problems, Internat. J. Non-Linear Mech., 21 (1986), pp. 111-134.
[42] A. Sarkar, N. Benabbou, and R. Ghanem, Domain decomposition of stochastic PDEs: Theoretical formulations, Internat. J. Numer. Methods Engrg., 77 (2009), pp. 689-701.
[43] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, Detect overlapping and hierarchical community structure in networks, Phys. A, 388 (2009), pp. 1706-1712.
[44] B. Smith, P. Bjorstad, and W. Gropp, Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations, Cambridge University Press, Cambridge, 2004.
[45] G. S. Stelling, On the Construction of Computational Methods for Shallow Water Flow Problems, Ph.D. thesis, TU Delft, Delft University of Technology, Delft, The Netherlands, 1983.
[46] A. H. Stroud, Approximate Calculation of Multiple Integrals, Prentice-Hall, Englewood Cliffs, NJ, 1971.
[47] A. H. Stroud and D. Secrest, Gaussian Quadrature Formulas, Prentice-Hall Series in Automatic Computation, Prentice-Hall, Englewood Cliffs, NJ, 1966.
[48] M.-B. Tran, A parallel four step domain decomposition scheme for coupled forward-backward stochastic differential equations, J. Math. Pures Appl. (9), 96 (2011), pp. 377-394.
[49] B. Van der Pol, A theory of the amplitude of free and forced triode vibrations, Radio Review, 1 (1920), pp. 701-710.
[50] M. Verlaan, Effcient Kalman Filtering Algorithms for Hydrodynamic Models, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1998.
[51] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput., 17 (2007), pp. 395-416.
[52] P. Wesseling, Principles of Computational Fluid Dynamics, Springer Ser. Comput. Math. 29, SpringerVerlag, Berlin, 2001.
[53] O. Widlund and M. Dryja, An Additive Variant of the Schwarz Alternating Method for the Case of Many Subregions, Technical Report 339, Ultracomputer Note 131, Department of Computer Science, Courant Institute, New York, 1987.
[54] S. Zhang, R.-S. WANG, AND X.-S. Zhang, Identification of overlapping community structure in complex networks using fuzzy c-means clustering, Phys. A, 374 (2007), pp. 483-490.


[^0]:    *Received by the editors July 16, 2018; accepted for publication (in revised form) April 14, 2020; published electronically July 8, 2020.
    https://doi.org/10.1137/18M1200567
    Funding: This work was supported by the National Science Foundation under grant NSF CMMI 1301235. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
    ${ }^{\dagger}$ Mechanical and Aerospace Engineering, University at Buffalo - SUNY, Buffalo, NY 14260 (arpanmuk@buffalo. edu, rahulrai@buffalo.edu, tsingh@buffalo.edu, abani@buffalo.edu).
    ${ }^{\ddagger}$ Aerospace Engineering, Pennsylvania State University, University Park, PA 16802 (psingla@psu.edu).

