MODEL, IDENTIFICATION & ANALYSIS OF COMPLEX STOCHASTIC SYSTEMS: APPLICATIONS IN STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS AND MULTISCALE MECHANICS

by

Sonjoy Das

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(CIVIL AND ENVIRONMENTAL ENGINEERING)

August 2008

Epigraph

Knowledge is inherent in man; no knowledge comes from outside;
it is all inside· · · We say Newton discovered gravitation.
Was it sitting anywhere in a corner waiting for him?
It was in his mind; the time came and he found it out.
All knowledge that the world has ever received comes from the
mind; the infinite library of the universe is in your own mind.
The external world is simply the suggestion, the occasion, which
sets you to study your own mind.
∼ Swami Vivekananda (January 12, 1863 – July 4, 1902)

Dedication

To my জন্মভূমি (motherland), ভারতবর্ষ (Bharatbarsha aka India)[1]

[1]bengali typesetting by using Dr. Lakshmi K. Raut's package freely available online; see `http://www2.hawaii.edu/~lakshmi/Software/bengali-omega/index.html`

Acknowledgments

Reflecting on the help imparted on me during the course of this work, I am overwhelmed by the debt of gratitude.

At the very outset, I would like to express my sincere gratitude and appreciation to my research advisor, Professor Roger Ghanem, for his insightful and interesting suggestions, for having confidence in me, for innumerable critical comments not only on the work but also on the writing and presentation style, for silently and painstakingly arranging for the financial support, and for his patience, encouragement and never-ending push for the betterment of the work throughout the course of this work. Not only he introduced me to a splendor of fascinating technical subjects spanning from the wave propagation to abstract functional analysis to random matrix theory (RMT), he discretely taught me many other related non-technical subtleties. I earnestly thank him for many other things that he did to create a smooth and easy life for me, which perhaps I did not even realize or need to bother about that much.

I am thankful for the financial support received from the Office of Naval Research, the Air Force Office of Scientific Research and the National Science Foundation during the course of this work.

Several people extended their direct helps concerning the subject matter of this work. In particular, I would like to acknowledge most gratefully the cordial support that I received from Professor Eduardo Dueñez and Professor James C. Spall during the class hours and beyond the class hours. They, respectively, taught me the RMT, and simulation and Monte Carlo methods. Both the subjects helped me to a great extent to complete this dissertation work. In addition, I thank Professor L. Carter Wellford, the then chairman of the department when I transferred from the Johns Hopkins University to the University of Southern California (USC), for his support to my endeavor for the successful completion of this work.

It is too numerous to enlist all my past and current colleagues, friends and the prominent researchers from several parts of the worlds, who have often contributed in one way or the another to this work over

a cup of coffee, e-mails, simple meeting/chatting across the table or standing on the street. This list particularly and certainly includes Dr. Debraj Ghosh and Dr. Alireza Doostan for discussions on the work presented in chapter 2, Professor Plamen Koev for sharing his excellent and latest code to compute the confluent hypergeometric function of matrix argument extensively used in section 5.4, Dr. Steven Finette for sharing SWARM95 experimental data used in section 3.3, Professor Jack Chessa for sharing a few of his finite element codes that turned out to be handy for section 5.4, Arash Noshadravan for extensively helping me to apply the method proposed in chapter 5 on some experimental data (not a part of this dissertation work but it provides a certain confidence that the proposed method can be reliably employed in practical problems), Dr. Maarten Arnst for nice discussions on a few parts of section 5.3.4 as well as critical comments on several parts of this dissertation work, and Professor Boris Rozovsky (he was in my dissertation proposal guidance committee) for his critical comments on the work presented in chapter 2. Thanks are also due to Dr. Mohsen Heidary-Fyrozjaee for his sharing with me the latest LaTeX 2_ε class and template files for USC dissertation.

I acknowledge the members of my dissertation committee, Professor Sami F. Masri, Professor Erik A. Johnson, Professor Jean-Pierre Bardet and Professor Paul Newton, for their comments and suggestions.

While I cannot put my finger precisely on any specific part, I believe my view towards the uncertainty analysis, which is the general topic of this dissertation work, has surely an element of influence from Professor C. S. Manohar. He is the one who first taught me this subject during my tenure of MSc(Engg.) program back in Indian Institute of Science.

Finally, I would like to reserve my special thanks for my parents and sister for their implicit faith and unconditional love, without which this work would definitely have not been completed. I often feel guilty of being selfish to pursue my goal of higher study away from home at the cost of many unsaid sacrifices made by my parents and specially by my (younger) sister.

University of Southern California, Los Angeles — Sonjoy Das

April, 2008.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation focusses on characterization, identification and analysis of stochastic systems. A stochastic system refers to a physical phenomenon with inherent uncertainty in it, and is typically characterized by a governing conservation law or partial differential equation (PDE) with some of its parameters interpreted as random processes, or/and a model-free random matrix operator. In this work, three data-driven approaches are first introduced to characterize and construct consistent probability models of non-stationary and non-Gaussian random processes or fields within the polynomial chaos (PC) formalism. The resulting PC representations would be useful to probabilistically characterize the system input-output relationship for a variety of applications. Second, a novel hybrid physics- and data-based approach is proposed to characterize a complex stochastic systems by using random matrix theory. An application of this approach to multiscale mechanics problems is also presented. In this context, a new homogenization scheme, referred here as *nonparametric* homogenization, is introduced. Also discussed in this work is a simple, computationally efficient and experiment-friendly coupling scheme based on frequency response function. This coupling scheme would be useful for analysis of a complex stochastic system consisting of several subsystems characterized by, e.g., stochastic PDEs or/and model-free random matrix operators.

While chapter 1 sets up the stage for the work presented in this dissertation, further highlight of each chapter is included at the outset of the respective chapter.

Chapter 1

Introduction

Quantifying confidence in model-based predictions is an essential step towards model validation, requiring an analysis of uncertainty in the representation of physical phenomena, in data acquisition and representation, and in the numerical resolution of the resulting, possibly stochastic, governing equations. Casting this validation problem in a probabilistic context requires the probabilistic characterization of system parameters from experimental evidence, and the propagation of the associated uncertainty to the predictions of the corresponding mathematical model.

Two venues have generally been pursued for the probabilistic representation of system parameters, associated with *parametric* and *nonparametric* models. Parametric models typically refer to the governing conservation law or partial differential equation, interpreting some of the associated parameters as intrinsically random [MI99, GS91, KTH92] thereby modeling them as random processes or/and fields. The statistical characterization of these models is a well developed topic with a significant set of tools to draw from. Typical statistics derived from these representations include marginal and multi-point statistics (usually two and three-point statistics) [CBS00, Chapter 3], including correlation functions and spectral density functions. A physical phenomenon modeled as a stochastic system with a lower level of uncertainties or/and with a relatively fewer number of random system parameters is well suited for analysis within the parametric formalism.

Nonparametric models, on the other hand, refer to the predictive model as a random operator usually resulting in random matrix perturbations to some nominal deterministic matrix equation [Soi00, Soi01a]. While the initial development of nonparametric models has evolved around particular physical models in which specific probability distributions of matrix-valued random variables have been analytically derived [Meh04, TV04], their recent application to broader problems in science and technology has required novel adaptation of statistical estimation and identification methods [Soi99, Soi00, Soi01a, Soi05a, Soi05b]. Since it refers to the class of models in which the available information needs to be expressed only through a set of system matrices/tensors (for example, mass matrix, stiffness matrix, damping matrix or

elasticity tensor), a system having a higher level of uncertainties or/and a system with a large number of random *local* system parameters (for example, fluid permeability, Young's modulus, shear modulus, bulk modulus, Poisson's ratio etc.) is more amenable to the nonparametric approach. It does not require any information at the local system parameter level as needed in a parametric formulation. At the current stage, most of these methods are, however, still limited to assimilating first order statistics of experimental observations along with a certain (scalar-valued) second-order statistic. This promising technique has recently been applied in a number of practical applications [CLPP$^+$07, ACB08].

The work presented in this dissertation considers both the models in characterizing the uncertainty of stochastic systems. Works in chapter 2 and chapter 3 are carried out within the parametric framework. A simple coupling technique to combine nonparametric system and parametric system is described in chapter 4. Finally, the work presented in chapter 5 considers the nonparametric model in more detail and propels the existing nonparametric techniques a little bit forward. While motivation behind the work presented in each chapter is reflected in the corresponding chapter, a glimpse of the overall works is in order before proceeding further.

1.1 Outline

The topics of chapter 2 and chapter 3, within the parametric framework, primarily deal with the characterization of a non-stationary and non-Gaussian random process or field by using a set of measurement data. While the conventional probabilistic characteristics, e.g., probability density functions, correlation function etc., are informative in a descriptive context, they cannot be efficiently propagated through predictive physics-based models. This is largely due to the difficulty associated with synthesizing realizations of non-Gaussian and non-stationary random vectors and stochastic processes from the knowledge of their statistical moments.

In recent years, the polynomial chaos (PC) expansion [GS91] has been used to a great advantage in representing tensor-valued stochastic processes and characterizing solutions to the associated stochastic governing differential equations. Within the purview of the PC framework, the probability model of the random process refers to a spectral decomposition constructed with respect to (w.r.t.) a set of basis functions in a suitable linear space. The basis functions constitute a set of orthogonal functions w.r.t. a

properly chosen probability measure [GS91, XK02, SG04a]. The coordinates (often referred as PC coefficients in the literature) w.r.t. the basis functions are the representative statistics. The set of PC coefficients play the similar role within the PC framework as played by the parameters of a characterizing multivariate joint probability density function (mjpdf) (for example, the mean vector and the covariance matrix of a multidimensional Gaussian distribution) within the conventional probability framework. Representing the random process by using PC expansion has some added advantages over the conventional probability framework. It facilitates in performing rigorous convergence analysis of the error in representing the system parameters (modeled as a random process) and its effect on the model-based predictions by using the machinery already available in the field of functional analysis. Furthermore, the PC representation presents a mechanism for *easy* simulation of the random process thus making it a viable alternative even within the conventional ensemble-based probability framework.

The PC formalisms thus provides a theoretically sound backbone facilitating efficient construction of the probability model of the non-stationary and non-Gaussian second-order random process possibly representing some model parameters of a stochastic system [Gha99, XK02, DNP$^+$04, LMNGK04, SG04a]. It has been proven to be a useful tool for systematic propagation of the statistical properties of these stochastic system parameters to the response of the model in a diverse field of applications [GS91, Gha99, GRH99, PG00, XLSK02, RNGK03, DNM$^+$03, SG04a, GGRH05, GSD07, LMNP$^+$07, WSB07]. The works presented in chapter 2 and chapter 3, therefore, focus on constructing the PC representation of a random process from data.

Chapter 4 and chapter 5 investigate the issues of nonparametric models. A coupling technique, that can couple several systems each of which, in its uncoupled state, is most suitable to either parametric or nonparametric modeling, is presented in chapter 4. A new probabilistic formulation within the nonparametric framework is proposed in chapter 5 to characterize a positive-definite random system matrix that is bounded from below and above in positive-definite sense.

1.2 Notation and Terminology

Throughout this work, bold face character would be used to indicate that the quantity under consideration is either random or multidimensional. The realizations of a multidimensional random quantity are, however, denoted by the respective normal characters for distinction purpose. Though every attempt would be

made to follow this convention, there might be violations at a few places but only if there exist no room for ambiguity.

Since a part of the current work considers the problem of constructing the probability model of a non-stationary and non-Gaussian random process, a clarification of terminology for the present work is set forth now. When the indexing set of the stochastic process is multidimensional, reference is often made to a random field, and a stationary random process is then referred to as a homogeneous random field. In this work, and to emphasize the identical underlying mathematical structure, the term 'stochastic process' or 'random process' would be ubiquitously used and the equivalent concepts of homogeneity and stationarity would be implied by default.

In the context of works presented in chapters 2–3, the term 'probability model' refers to 'PC representation'.

The term 'random variate' or 'random variable' would be succinctly used throughout this work to indicate scalar-, vector-, matrix- or tensor-valued random variable, which would be clear from the context.

Chapter 2

Asymptotic Distribution for PC Representation from Data

A procedure is presented in this chapter for characterizing the asymptotic sampling distribution of estimators of the PC coefficients of a second-order non-stationary and non-Gaussian random process by using a collection of observations. The random process represents a physical quantity of interest, and the observations made over a finite denumerable subset of the indexing set of the random process are considered to form a set of realizations of a random vector, $\mathcal{Y}$, representing a finite-dimensional projection of the random process. The Karhunen-Loève (KL) decomposition and a scaling transformation are employed to produce a reduced order model, $\mathcal{Z}$, of $\mathcal{Y}$. The PC expansion of $\mathcal{Z}$ is next determined by having recourse to the maximum-entropy (MaxEnt) principle, Metropolis-Hastings (M-H) Markov chain Monte Carlo (MCMC) algorithm and the Rosenblatt transformation. The resulting PC expansion has random coefficients; where the random characteristics of the PC coefficients can be attributed to the limited data available from the experiment. The estimators of the PC coefficients of $\mathcal{Y}$ obtained from that of $\mathcal{Z}$ are found to be maximum likelihood estimators (MLE) as well as consistent and asymptotically efficient. Computation of the covariance matrix of the associated asymptotic normal distribution of estimators of the PC coefficients of $\mathcal{Y}$ requires knowledge of the Fisher information matrix (FIM). The FIM is evaluated here by using a numerical integration scheme as well as a sampling technique. The resulting confidence interval on the PC coefficient estimators essentially reflects the effect of incomplete information (due to data limitation) on PC representation of the stochastic process. This asymptotic distribution is significant as its characteristics can be propagated through predictive model for which the stochastic process in question describes uncertainty on some input parameters.

2.1 Motivation and Problem Description

Many applications in science and engineering involve modeling spatio-temporal phenomena. Within the confines of the probabilistic framework, the Gaussian stochastic process has been the most commonly used form for modeling such physical phenomena. In addition to the constraint provided by the form of the probability measure when using such a process, additional simplifying assumptions such as stationarity, separability and symmetry are usually made in constructing it for mathematical convenience and computational expediency. The construction of Gaussian processes from finite data continues to be an active field of research with issues such as multidimensionality, non-symmetry, and non-stationarity providing the motivation for much of the innovation [GGG05]. The development of non-Gaussian models, on the other hand, has been much slower; certainly to a slower extent than the Gaussian models, chiefly due to the scarcity of consistent mathematical theories for describing infinite-dimensional probability measures.

In addition to the mathematical challenges introduced by the quest for non-Gaussian stochastic models, a very important difficulty is presented by the scarcity of data on which these models are to be based. Since Gaussian processes are characterized only by their mean and covariance functions, they require a manageable amount of information and thus often provide a rational modeling alternative. This has limited the scope of non-Gaussian models to transformations of Gaussian vectors and processes, or to models that are completely characterized by their lower order statistics. These challenges notwithstanding, it remains a recognized fact that many processes representing physical phenomena rarely satisfy the assumptions and constraints associated with a Gaussian process. (See section 3.1 for a more exhaustive discussion on the currently existing procedures to characterize non-Gaussian random processes).

As highlighted in chapter 1, a significant benefit of the PC formalism lies in its ability to characterize non-Gaussian, non-stationary and multidimensional second-order stochastic processes and the potential for its efficient implementation into predictive models. Therefore, the work in this chapter focuses on the construction and characterization of PC representation of a non-stationary and non-Gaussian random process only from data.

It is assumed in the present work that the stochastic process under consideration is a second-order random process. This assumption guarantees the existence of its PC representation. From the vantage point of the fact that the most physically measurable random processes are second-order type, this assumption

is not a severe restriction. The PC expansion of a second-order (scalar-, vector-, matrix or tensor-valued) stochastic process is a spectral decomposition in terms of a set of orthogonal basis functions constructed w.r.t. a suitable and known probability measure of user's choice. Typical PC decompositions have been developed w.r.t. basis functions representing Hermite polynomials in Gaussian variables [GS91], polynomials that are orthogonal w.r.t. a variety of measures [XK02] and multi-wavelet basis [LMNGK04]. Convergence results for PC representations are well-established for functionals of Gaussian processes [CM47] and for functions of finite-dimensional random vectors with arbitrary measure [XK02, SG04a].

As indicated earlier, the particular statistics characterizing the PC representation consist of PC coefficients or coordinates of the process w.r.t. the chosen set of orthogonal basis functions. The algebraic character of the PC coefficients (scalars, vectors, functions or vector-, matrix- or tensor-valued functions) is inherited from the stochastic quantity they represent. Using these linear decompositions of stochastic vectors and processes, the mapping of probabilistic measure from stochastic system parameters to system state follows from a mapping between the PC coefficients of the system and state processes. This latter mapping is a deterministic transformation, obtained from the original stochastic governing equations through algebraic manipulations and projections in suitable linear spaces [GS91, Gha99, GRH99].

Consider a physical phenomenon defined over $\mathbf{D} \subset \mathbb{R}^d$ with $\mathbb{R}^d$ representing the Euclidean d-space, and $\mathbf{D}$ typically referring to a spatio-temporal domain. Assume that this physical process is modeled as a stochastic process, $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$, on $\mathbf{D} \times \boldsymbol{\Theta}$ with probability space $(\boldsymbol{\Theta}, \mathcal{F}_{\boldsymbol{\Theta}}, \mu)$. Consider a sequence of possible observations of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ at N locations over $\mathbf{D}$ with coordinates $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$. Denote the random variables associated with the random process, $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$, at these locations by $\mathbf{y}_q \equiv \mathbf{y}(\mathbf{x}_q, \boldsymbol{\theta})$, $q = 1, 2, \cdots, N$, and let $\boldsymbol{\mathcal{Y}} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]^T$ where T represents the transpose operator. It should be clear that $\boldsymbol{\mathcal{Y}}$ represents a finite-dimensional representation of the original (infinite-dimensional) stochastic process. Denote the multivariate joint probability distribution function (mjPDF; contrast it with mjpdf abbreviated earlier for multivariate joint probability density — not distribution — function) of $\boldsymbol{\mathcal{Y}}$ by $P_{\mathbf{y}_1, \cdots, \mathbf{y}_N}$. The probability measure of the underlying random process is then completely characterized by the family of mjPDF: $\{P_{\mathbf{y}_1, \cdots, \mathbf{y}_N}\}, \forall N \geq 1$. Since N is always finite in an experimental or numerical context, characterizing the underlying stochastic process has to be performed, in some approximate sense, through $\boldsymbol{\mathcal{Y}}$. The value of N, required to achieve a certain fidelity in the finite-dimensional representation, depends on the characteristics of the stochastic fluctuations of the original stochastic process over its spatio-temporal domain (think of, e.g., correlation length).

In many practical situations, each component of $\boldsymbol{\mathcal{Y}}$ can often be sufficiently well characterized by a finite-dimensional PC representation. In these cases, a first level approximation is thus introduced while selecting the dimension, $n_d < \infty$, of the PC representation. For a component of $\boldsymbol{\mathcal{Y}}$ associated with a specified $\mathbf{x}$, consider the leading P terms in this n_d-dimensional (D) PC representation of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ and, let $\mathbf{h}_\mathbf{x}$ be the P-D vector consisting of these PC coefficients. In most physical applications, it cannot be verified in general if such $\mathbf{h}_\mathbf{x}$ exists or not, and even if it exists, it cannot be specified exactly. It is assumed in this work that such an unknown $\mathbf{h}_\mathbf{x}$ exists. Further, let $\widehat{\mathbf{h}}_\mathbf{x}$ denote the P-D vector representing an estimator of $\mathbf{h}_\mathbf{x}$ based on available information. The elements of $\widehat{\mathbf{h}}_\mathbf{x}$ are computed based on a finite set of noisy measurements that are typically observed on $\boldsymbol{\mathcal{Y}}$. Let, furthermore, $\widetilde{\mathbf{h}}_\mathbf{x}$ be the P-D vector consisting of the PC coefficients of the appropriate random variable component of $\boldsymbol{\mathcal{Y}}$ satisfying $\lim_{N\to\infty} \widetilde{\mathbf{h}}_\mathbf{x} = \mathbf{h}_\mathbf{x}$.

While the error $\|\widetilde{\mathbf{h}}_\mathbf{x} - \mathbf{h}_\mathbf{x}\|$ ($\|\cdot\|$ represents a suitable norm, say, the Euclidean vector norm in $\mathbb{R}^P$) can be reduced by increasing N, the error $\|\widehat{\mathbf{h}}_\mathbf{x} - \widetilde{\mathbf{h}}_\mathbf{x}\|$, conditioned on $\widetilde{\mathbf{h}}_\mathbf{x}$, can be monitored and reduced by increasing the statistical significance of the sample from which the PC coefficients of $\boldsymbol{\mathcal{Y}}$ are estimated. The total error, $\|\widehat{\mathbf{h}}_\mathbf{x} - \mathbf{h}_\mathbf{x}\|$, is bounded by,

$$\|\widehat{\mathbf{h}}_\mathbf{x} - \mathbf{h}_\mathbf{x}\| \;\leq\; \|\widehat{\mathbf{h}}_\mathbf{x} - \widetilde{\mathbf{h}}_\mathbf{x}\| + \|\widetilde{\mathbf{h}}_\mathbf{x} - \mathbf{h}_\mathbf{x}\| \quad \text{a.s.}, \tag{2.1}$$

in which a.s. indicates that the above inequality is valid in almost sure (a.s.) sense w.r.t. the probability measure, μ. It is assumed here that the second error term, $\|\widetilde{\mathbf{h}}_\mathbf{x} - \mathbf{h}_\mathbf{x}\|$, is known either deterministically (for example, in the sense that the effect of finite N would be negligible if N is large enough so that $\boldsymbol{\mathcal{Y}}$ encompasses all the statistical characteristics of interests of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ with sufficient accuracy) or statistically. The work presented in this chapter, on the other hand, focuses on the first error term $\|\widehat{\mathbf{h}}_\mathbf{x} - \widetilde{\mathbf{h}}_\mathbf{x}\|$, conditioned on $\widetilde{\mathbf{h}}_\mathbf{x}$, that can be sharpened through data acquisition.

Recent work in this direction has relied on the maximum likelihood principle to estimate the PC coefficients based on an approximate mjpdf of the dominant KL random variable components of the stochastic process [DGS06, DSG07] that simplifies the form of the likelihood function for computational expediency (this computational scheme bears some resemblance to the composite likelihood method [Lin88]). Additional related work has assumed the KL components to be statistically independent and estimated their probability density functions using either Bayesian inference [GD06] or a histogram constructed from observations of the KL variables [BLT03]. It should be noted that a number of previous efforts

[BLT03], while constructing the probability density function (pdf) estimates of the KL variables, did not provide a method for constructing or using associated sampling distributions which could otherwise have been used as indicators to the sensitivity of the probabilistic model to additional observations.

As already explained, the work here focuses primarily on the error, due to the *inexact representation* of the stochastic process because of *data limitations*, for a general class of problems. A framework and its numerical implementation for the statistical analysis of this error, that would be useful to determine its impact on model-based prediction, is presented. Use of the PC representation of the stochastic process expressed with sufficient accuracy in terms of the statistically dependent dominant KL random variables makes the procedure very efficient in propagating the error to the model-based predictions. In particular, an asymptotic distribution, conditioned on $\widetilde{\mathbf{h}}_{\mathbf{x}}$, is identified for $\widehat{\mathbf{h}}_{\mathbf{x}} - \widetilde{\mathbf{h}}_{\mathbf{x}}$, and a computational scheme for its evaluation is presented. Given the Gaussian form of this distribution (see section 2.2.4), the propagation of this error through the system prediction can be readily formulated, thus enabling the assessment of the sensitivity of model-based predictions to refinement in the statistics of the model parameters.

Though the primary goal of the current chapter is to present a framework for analyzing the significance of data error in the background of PC formalisms, several tools, from the fields of MaxEnt principle and FIM, are also needed for successful completion of the work here. The MaxEnt principle is employed to estimate the mjpdf of a random vector, representing a reduced order model of $\boldsymbol{\mathcal{Y}}$, consisting of M dominant and statistically dependent KL random variables. It must be remarked here that though the MaxEnt principle is known for several decades, it is primarily and successfully used for the estimation of pdfs of scalar random variables and for a limited class of multivariate problems. Moreover, in the existing statistical literature, it is hard to find any appealing, reliable and computationally efficient density estimation technique *for* a set of statistically dependent random variables from a set of *finite* samples. A brief introduction of the principle of maximum entropy, its appealing features and a computational scheme for density estimation in the context of the present work are provided in section 2.3.1. The FIM, on the other hand, is required to compute the covariance matrix of the asymptotic normal distribution. This matrix is an indicator of the amount of information contained in the observed data about quantities of interest typically representing some model parameters. Prominent areas of applications of the FIM include, to name a few, confidence interval computation of model parameters [CLR96, HD97], determination of inputs to nonlinear models in experimental design [Spa03, Section 17.4] and determination of noninformative prior distribution (Jeffreys' prior) for Bayesian analysis [Jef46]. The FIM, in the present context, would

be useful to compute the confidence interval of the error term, $\|\widehat{\mathbf{h}}_{\mathbf{x}} - \widetilde{\mathbf{h}}_{\mathbf{x}}\|$, conditioned on $\widetilde{\mathbf{h}}_{\mathbf{x}}$. A brief discussion on this matrix in light of the present work and the required estimation technique is presented in section 2.3.4.

The chapter begins with the development of a reduced order model for $\boldsymbol{\mathcal{Y}}$ by using its KL decomposition. The resulting M-D (with $M < N$) random vector associated with the dominant subspace will be referred as the KL vector which is subsequently transformed to another M-D random vector supported on an M-dimensional hypercube, $[0\ 1]^M$. This new random vector will be referred as the normalized KL (nKL) vector. An estimation of the mjpdf of the nKL vector is then obtained by using the MaxEnt technique. Following that, a Markov chain is constructed and used to estimate the PC representation of the nKL vector from which estimators of the PC coefficients of $\boldsymbol{\mathcal{Y}}$ are determined. The asymptotic probability density function (apdf) of estimators of the PC coefficients of $\boldsymbol{\mathcal{Y}}$ is then identified in order to statistically characterize the first error term in (2.1). The procedure is demonstrated by an example and the final section contains the conclusion inferred from the work presented in this chapter.

The proposed use of the Rosenblatt transformation in constructing the PC representation of a *random vector* in section 2.2.3 and identification of the asymptotic distribution in section 2.2.4 are the original contributions of the present chapter to the literature of computational statistics. The computational scheme, as described in section 2.3.3, is also a noteworthy addition to the set of computational statistics tool for mjpdf estimation by matching a target set of higher order *joint* statistics of a random vector.

2.2 Representation and Characterization of the Random Process from Measurements

The KL expansion [Loe78, Chapter XI], [Jol02] is first employed to optimally reduce the number of random variables needed to characterize $\boldsymbol{\mathcal{Y}}$ yielding, in the process, a set of uncorrelated random variables. Then, the PC coefficients of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ are determined via estimating the PC coefficients of the reduced order model.

2.2.1 Karhunen-Loève Decomposition: Reduced Order Representation of the Random Process

Suppose that n observations of $\boldsymbol{\mathcal{Y}}$, denoted by $\mathcal{Y}_1, \cdots, \mathcal{Y}_n$, have been collected. An unbiased estimate of the mean vector of $\boldsymbol{\mathcal{Y}}$ is given by $\overline{\mathcal{Y}} = (1/n)\sum_{k=1}^{n} \mathcal{Y}_k$, and an estimate of the $N \times N$ covariance matrix by $C_{yy} = (1/(n-1))Y_o Y_o^T$ in which $Y_o = [\mathcal{Y}_{1o}, \cdots, \mathcal{Y}_{no}]$ represents an $N \times n$ matrix and $\mathcal{Y}_{ko} \equiv \mathcal{Y}_k - \overline{\mathcal{Y}}$, $k = 1, \cdots, n$. Let the i-th, $i = 1, \cdots, N$, largest eigenvalue of C_{yy} be denoted by ς_i and the associated eigenvector by V_i. Following the KL expansion procedure, let us now collect the dominant KL random variable components, $\{\mathbf{z}'_1, \cdots, \mathbf{z}'_M\}$, $M < N$, in an M-D random vector, $\boldsymbol{\mathcal{Z}}' = [\mathbf{z}'_1, \cdots, \mathbf{z}'_M]^T$. The M random variables, $\mathbf{z}'_i$, $i = 1, \cdots, M$, are zero-mean and uncorrelated (but not necessarily statistically independent), and have unbiased estimates of variances given by ς_i's. The value of M is chosen such that $\mathrm{tr}(C_{yy}) = \sum_{i=1}^{N} \mathrm{var}(\mathbf{y}_i) \approx \sum_{i=1}^{M} \varsigma_i = \sum_{i=1}^{M} \mathrm{var}(\mathbf{z}'_i)$ with var and tr, respectively, representing variance and trace operators. Here, $\boldsymbol{\mathcal{Z}}'$ is related to $\boldsymbol{\mathcal{Y}}$ by,

$$\boldsymbol{\mathcal{Z}}' = V^T(\boldsymbol{\mathcal{Y}} - \overline{\mathcal{Y}}), \tag{2.2}$$

in which $V = [V_1, \cdots, V_M]$ is the $N \times M$ matrix of eigenvectors, $V_1, \cdots, V_M$. The random vector, $\boldsymbol{\mathcal{Z}}'$, will be referred now on as the KL vector.

The set of experimental samples of $\boldsymbol{\mathcal{Z}}'$ can be immediately obtained by replacing $\boldsymbol{\mathcal{Y}}$ with $\mathcal{Y}_1, \cdots, \mathcal{Y}_n$ in (2.2) resulting in $\mathcal{Z}'_1, \cdots, \mathcal{Z}'_n$. To enhance the regularity of the ensuing numerical problem and improve the efficiency of the associated computation, the following scaling is applied to the data on $\boldsymbol{\mathcal{Z}}'$ obtaining a set of realizations of a new random vector,

$$\mathcal{Z}_k = \left[(\mathcal{Z}'_k - \underline{a}) \circ \left(\frac{1}{\underline{b} - \underline{a}} \right) \right], \quad k = 1, \cdots, n. \tag{2.3}$$

Here, the symbol $\circ$ represents element-wise product operator or the Hadamard product operator, $\underline{a} = [a_1, \cdots, a_M]^T$ and $\underline{b} = [b_1, \cdots, b_M]^T$ with $a_i = \min(z_i'^{(1)}, \cdots, z_i'^{(n)})$ and $b_i = \max(z_i'^{(1)}, \cdots, z_i'^{(n)})$, in which $z_i'^{(k)}$ is the i-th component of the k-th sample, $\mathcal{Z}'_k = [z_1'^{(k)}, \cdots, z_M'^{(k)}]$, and finally, $1/(\underline{b} - \underline{a})$ needs to be interpreted as $M \times 1$ column vector with its i-th, $i = 1, \cdots, M$, element being given by reciprocal of the i-th element of $(\underline{b} - \underline{a})$. Denote the resulting M-D random vector associated with the samples, $\{\mathcal{Z}_k\}_{k=1}^{n}$, by $\boldsymbol{\mathcal{Z}} = [\mathbf{z}_1, \cdots, \mathbf{z}_M]^T$ supported on M-dimensional unit hypercube, $\boldsymbol{\Xi} \equiv [0\ 1]^M \subset \mathbb{R}^M$. The

random vector, $\boldsymbol{\mathcal{Z}}$, having uncorrelated and non-zero mean components, will be referred as the normalized KL (nKL) vector. The following relation between $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{Y}}$ then holds,

$$\boldsymbol{\mathcal{Y}} \approx \boldsymbol{\mathcal{Y}}^{(M)} = [V\,(\underline{b} + \underline{a} \circ \boldsymbol{\mathcal{Z}})] + \overline{\mathcal{Y}}. \tag{2.4}$$

The approximation sign, '$\approx$', in (2.4) is indicated due to the fact that $\boldsymbol{\mathcal{Y}}$ is projected into the space spanned only by the largest M dominant eigenvectors of C_{yy} to obtain the reduced order representation, $\boldsymbol{\mathcal{Z}}$. Next, a sampling-based technique for computing an estimate of the vector, $\mathbf{h}_{\mathbf{x}}$, of the PC coefficients of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ is described via estimating the PC coefficients of $\boldsymbol{\mathcal{Z}}$. A description of the PC formalism is, however, first reviewed before estimating the PC coefficients of $\boldsymbol{\mathcal{Z}}$ from $\{\mathcal{Z}_k\}_{k=1}^{n}$.

2.2.2 Polynomial Chaos Formalism

The current state-of-the-art PC approach is the evolution of the repertoire of Cameron and Martin [CM47], where a second-order non-linear function(al), defined on the space, $\mathcal{C}$, of all real-valued continuous functions on a compact support, is approximated by a spectral decomposition constructed w.r.t. a set of multidimensional orthogonal Hermite polynomials. The set of Hermite polynomials is constructed w.r.t. a set of statistically independent Gaussian random variables. They particularly investigated the issue of convergence as the dimension (representing the number of Gaussian random variables) tends to infinity. It is shown [CM47] that the resulting spectral representation converges to the non-linear function(al) being approximated in mean-square sense as the dimension and order of the multidimensional Hermite polynomials tend to infinity. The mean-square error (MSE) is measured w.r.t. the Wiener measure [Wie38], on $\mathcal{C}$. The Wiener measure is used to represent the integral of a Brownian motion associated with (infinite-dimensional) Gaussian white noise process.

The aforesaid work involving infinite-dimensional Gaussian measures has been adapted to the finite-dimensional Gaussian and non-Gaussian measures by employing several novel schemes. Accordingly, the PC representation of second-order random process and random vector have been developed in terms of orthogonal polynomials constructed w.r.t. a set of statistically *independent* Gaussian random variables [GS91, Gha99, GRH99, DNP^{+}04] as well as non-Gaussian random variables [XK02]. The doubly orthogonal polynomials, assuming that the KL random variable components of the stochastic process are

statistically *independent* [BTZ05], and the wavelet basis functions, constructed w.r.t. statistically *independent* non-Gaussian random variables [LMNGK04] (also see [PB06] for a related application), have recently been implemented as basis functions in the construction of PC representation. The theoretical development of employing orthogonal polynomials, that are constructed w.r.t. a set of statistically *dependent* second-order random variables, has also been accomplished [SG04a].

Denote the number of random variables to be included in the PC representation (i.e., dimension of the PC representation) by n_d. While increasing the dimension of the PC expansion provides added freedom in the representation, it significantly increases the computational cost. A balance must be thus reached among flexibility of the representation, available computational resources and target accuracy. Let $\underline{\boldsymbol{\xi}} \equiv (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{n_d})$ be a $\mathbb{R}^{n_d}$-valued random vector defined on $(\boldsymbol{\Theta}, \mathcal{F}_{\boldsymbol{\Theta}}, \mu)$ with its induced mjPDF being denoted by $P_{\underline{\boldsymbol{\xi}}}$. The probability measure, $P_{\underline{\boldsymbol{\xi}}}$, is chosen such that it is best amenable [XK02, SG04a, LMNGK04] to the PC representation of $\mathbf{z}_k \equiv \mathbf{z}_k(\underline{\boldsymbol{\xi}})$ thus adapting its choice to the known probabilistic characteristics of $\mathbf{z}_k(\underline{\boldsymbol{\xi}})$. It is also assumed that $P_{\underline{\boldsymbol{\xi}}}$ admits a joint pdf, $p_{\underline{\boldsymbol{\xi}}}$, verifying $dP_{\underline{\boldsymbol{\xi}}}(\underline{\xi}) = p_{\underline{\boldsymbol{\xi}}}(\underline{\xi})\, d\underline{\xi}$, with $d\underline{\xi}$ being given by $d\underline{\xi} = \prod_{i=1}^{n_d} d\xi_i$ in which $d\xi_i$ is the Lebesgue measure on $\mathbb{R}$. Based on chosen $P_{\underline{\boldsymbol{\xi}}}$, the PC representation of each component of $\boldsymbol{\mathcal{Z}}$ can be expressed as,

$$\mathbf{z}_k \equiv \mathbf{z}_k(\underline{\boldsymbol{\xi}}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^{n_d}} z_{\boldsymbol{\alpha},k}\, \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}}), \quad k = 1, \cdots, M, \tag{2.5}$$

if $\mathbf{z}_k(\underline{\boldsymbol{\xi}})$ is second-order random variable, i.e., $E[\,|\,\mathbf{z}_k(\underline{\boldsymbol{\xi}})\,|^2] < \infty$ with $|\,x\,|$ representing the absolute value of x and $E[\,\cdot\,]$ representing the expectation operator w.r.t. the chosen probability measure, $P_{\underline{\boldsymbol{\xi}}}$ (this second-order condition is satisfied here since the underlying stochastic process is assumed to be second-order). Here, $\mathbb{N} = \{0, 1, 2, \cdots\}$, $\{z_{\boldsymbol{\alpha},k}, \boldsymbol{\alpha} \equiv (\alpha_1, \cdots, \alpha_{n_d}) \in \mathbb{N}^{n_d}\}$ is the set of PC coefficients representing the coordinates w.r.t. the set of basis functions, $\{\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$, given by [SG04a],

$$\begin{aligned} \boldsymbol{\Upsilon}_{\mathbf{0}}(\underline{\xi}) &= 1, && \text{if } \boldsymbol{\alpha} = \mathbf{0} \in \mathbb{N}^{n_d}, \\ \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\xi}) &= \left(\frac{\prod_{i=1}^{n_d} p_{\boldsymbol{\xi}_i}(\xi_i)}{p_{\underline{\boldsymbol{\xi}}}(\underline{\xi})} \right)^{1/2} \prod_{i=1}^{n_d} \boldsymbol{\Psi}_{\alpha_i}(\xi_i), && \text{if } \boldsymbol{\alpha} \neq \mathbf{0}. \end{aligned} \tag{2.6}$$

Here, $p_{\boldsymbol{\xi}_i}$ is the marginal pdf (marpdf) of $\boldsymbol{\xi}_i$ induced by $p_{\underline{\boldsymbol{\xi}}}$ and $\boldsymbol{\Psi}_{\alpha_i}(\xi_i)$ are polynomials of order α_i in ξ_i. These polynomials are orthogonal to each other in the sense that $E[\boldsymbol{\Psi}_j(\boldsymbol{\xi}_i)\boldsymbol{\Psi}_l(\boldsymbol{\xi}_i)] = 0$ for $j \neq l$, in which $E[\,\cdot\,]$ is the expectation operator w.r.t. the probability measure, $P_{\boldsymbol{\xi}_i}$, that admits $dP_{\boldsymbol{\xi}_i}(\xi_i) = p_{\boldsymbol{\xi}_i}(\xi_i)\, d\xi_i$.

As already indicated, this also implies [SG04a] the orthogonality of the set, $\{\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}}), \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$, w.r.t. $P_{\underline{\boldsymbol{\xi}}}$. In the case of statistically independent random variables, $(2.6)_2$ simplifies to,

$$\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}}) = \prod_{i=1}^{n_d} \boldsymbol{\Psi}_{\alpha_i}(\boldsymbol{\xi}_i). \tag{2.7}$$

The equality, '=', in (2.5) should be interpreted in the mean-square sense such that $E[\{\mathbf{y}_k(\underline{\boldsymbol{\xi}}) - \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq n_o} y_{\boldsymbol{\alpha},k}\ \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\}^2] \longrightarrow 0$ as $n_o \longrightarrow \infty$, where the expectation operator is w.r.t. $P_{\underline{\boldsymbol{\xi}}}$ [SG04a], $|\,\boldsymbol{\alpha}\,| = \sum_{i=1}^{n_d} \alpha_i$, and n_o is the maximum order (*i.e.*, order of the PC representation) of all the basic orthogonal polynomials, $\{\boldsymbol{\Psi}_{\alpha_i},\ \alpha_i \in \mathbb{N},\ i \in (1, \cdots, n_d)\}$, included in (2.5). Given n_d and chosen n_o, the number of basis functions retained in the infinite series of (2.5) is given by (including the 0-th order basis function), $(P+1) = (n_o + n_d)!/(n_o!\ n_d!)$ that clearly tends to infinity as $n_0 \longrightarrow \infty$. This implies that the accuracy (in the sense of mean-square error (MSE) reduction) of the PC representation can be improved by *only* increasing the order, n_0. However, for computational purpose, this infinite series is truncated after a finite number of terms that is typically determined by the available computational budget and target accuracy (usually in terms of MSE).

The flexibility and the accuracy of the PC representation also depend on the choice of $P_{\underline{\boldsymbol{\xi}}}$ and consequently, on the resulting set of orthogonal basis functions used in (2.5). The proper selection of the probability measure, $P_{\underline{\boldsymbol{\xi}}}$, may be dictated by the physical or experimental or modeling features involved in treating the physical process of interest as stochastic process. This stochastic process is, therefore, viewed as a (possibly nonlinear) transformation of $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{n_d}$ representing those features. However, different choice of suitable $P_{\underline{\boldsymbol{\xi}}}$ is also theoretically plausible, and might be preferred for a relatively less computational expense required to achieve an equivalent or more statistically significant representation (in some appropriate sense, for example, in the sense of minimum MSE). Once the choice for $P_{\underline{\boldsymbol{\xi}}}$ is made and the mapping, $\underline{\boldsymbol{\xi}} \longmapsto \mathbf{z}_k(\underline{\boldsymbol{\xi}})$, is identified (either explicitly or implicitly), the PC coefficients can be computed by using the orthogonality property of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}$'s,

$$z_{\boldsymbol{\alpha},k} = \frac{E\left[\mathbf{z}_k(\underline{\boldsymbol{\xi}})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right]}{E\left[\boldsymbol{\Upsilon}^2_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right]}, \quad \boldsymbol{\alpha} \in \mathbb{N}^{n_d} \quad \text{and} \quad k = 1, \cdots, M. \tag{2.8}$$

The denominator in (2.8) can be evaluated by using (2.6) or (2.7) as appropriate. When $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{n_d}$ are statistically independent, then $E\left[\boldsymbol{\Upsilon}^2_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right]$ reduces to, the denominator in (2.8) reduces to,

$$E\left[\boldsymbol{\Upsilon}^2_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right] = \prod_{i=1}^{n_d} E\left[\boldsymbol{\Psi}^2_{\alpha_i}(\boldsymbol{\xi}_i)\right], \tag{2.9}$$

in which $E\left[\boldsymbol{\Psi}^2_{\alpha_i}(\boldsymbol{\xi}_i)\right]$ can often be extracted from the existing literature [Leb72, Chapter 4], [GS91, XK02, SG04a] for many commonly employed measures, $P_{\boldsymbol{\xi}_i}$'s.

The numerator in (2.8), on the other hand, is given by,

$$E\left[\mathbf{z}_k(\underline{\boldsymbol{\xi}})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right] = \int_{S_{\underline{\boldsymbol{\xi}}}} \mathbf{z}_k(\underline{\xi})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\xi})\, p_{\underline{\boldsymbol{\xi}}}(\underline{\xi})\, d\underline{\xi}, \tag{2.10}$$

in which $S_{\underline{\boldsymbol{\xi}}} \subseteq \mathbb{R}^{n_d}$ is the support of $\underline{\boldsymbol{\xi}}$. In the discussion until now, the existence of the mapping, $\underline{\boldsymbol{\xi}} \longmapsto \mathbf{z}_k(\underline{\boldsymbol{\xi}})$, is implicitly implied. Recent developments in PC representations have predominantly treated problems where such a mapping is defined, either implicitly or explicitly. In the current work, however, this mapping is unknown since the information, that is assumed to be available, is *only* the measurement data on $\boldsymbol{\mathcal{Y}}$ (i.e., in turn, on $\boldsymbol{\mathcal{Z}}$). A sampling based-technique for evaluating the numerator in (2.8) is next described.

2.2.3 Polynomial Chaos Representation from Data

In general the random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{n_d}$, could be statistically dependent. The case of statistically independent components is, however, of particular interest because of the additional computational efficiency involved in evaluation of the integral in (2.10). After all the probability measure, $P_{\underline{\boldsymbol{\xi}}}$, is a *suitable* choice of the analyst! (It would be *more* clear later in chapter 3). For statistically independent $\{\boldsymbol{\xi}_i\}_{i=1}^{n_d}$, the integral in (2.10) reduces to,

$$\begin{aligned} &E\left[\mathbf{z}_k(\underline{\boldsymbol{\xi}})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right] \\ &= \int_{S_{\boldsymbol{\xi}_1}} \cdots \int_{S_{\boldsymbol{\xi}_{n_d}}} \mathbf{z}_k(\underline{\xi}) \left(\prod_{i=1}^{n_d} \boldsymbol{\Psi}_{\alpha_i}(\xi_i)\right) p_{\boldsymbol{\xi}_1}(\xi_1) \cdots p_{\boldsymbol{\xi}_{n_d}}(\xi_{n_d})\, d\xi_1 \cdots d\xi_{n_d}, \end{aligned} \tag{2.11}$$

in which $S_{\boldsymbol{\xi}_i} \subseteq \mathbb{R}$ is support of $\boldsymbol{\xi}_i$.

To establish the required mapping, $\underline{\boldsymbol{\xi}} \longmapsto \mathbf{z}_k(\underline{\boldsymbol{\xi}})$, an inverse approach is adopted now, which would facilitate in carrying out the integral in (2.11). The Rosenblatt transformation [Ros52] is used to relate the n_d-variate PDF, $P_{\underline{\boldsymbol{\xi}}}$, associated with (2.11), and an absolutely continuous M-variate PDF, $P_{\boldsymbol{\mathcal{Z}}}$, of $\boldsymbol{\mathcal{Z}}$. This step imposes the condition that $n_d = M$. The mapping defined by the Rosenblatt transformation (as described next) is continuous. A requirement for using the Rosenblatt transformation is absolute continuity of $P_{\boldsymbol{\mathcal{Z}}}$.

Note that $P_{\boldsymbol{\mathcal{Z}}}$ represents an estimate of the PDF of $\boldsymbol{\mathcal{Z}}$ obtained by using a suitable density estimation technique. It is assumed for the time being that an estimate of the PDF of $\boldsymbol{\mathcal{Z}}$ is available[1]. Suppose that $P_{\boldsymbol{\mathcal{Z}}}$ is characterized by p parameters, $\lambda_1, \cdots, \lambda_p$, represented as a $p \times 1$ column vector, $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_p]^T$. For example, the free elements characterizing a multivariate normal distribution function i.e., elements of the mean vector, $\boldsymbol{\mu}$, and elements on and above the diagonal of the covariance matrix, $\boldsymbol{\Sigma}$, might constitute the column vector $\boldsymbol{\lambda}$ or, for a second example, the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ could as well depend on $\boldsymbol{\lambda}$ by some known deterministic (functionally implicit or explicit) relationships, $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\lambda})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\lambda})$. This parameter vector, $\boldsymbol{\lambda}$, needs to be estimated by using a suitable density estimation technique and depends on the measurement data, and essentially characterizes the random process, $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$, through its reduced order representation, $\boldsymbol{\mathcal{Z}}$.

Consider the Rosenblatt transformation, $T : \boldsymbol{\mathcal{Z}} \longmapsto \underline{\boldsymbol{\xi}}$, defined by,

$$\left.\begin{array}{rcl} P_{\boldsymbol{\xi}_1}(\boldsymbol{\xi}_1) & \stackrel{d}{=} & P_1(\mathbf{z}_1), \\ P_{\boldsymbol{\xi}_2}(\boldsymbol{\xi}_2) & \stackrel{d}{=} & P_{2|1}(\mathbf{z}_2), \\ \vdots & & \\ P_{\boldsymbol{\xi}_M}(\boldsymbol{\xi}_M) & \stackrel{d}{=} & P_{M|1:(M-1)}(\mathbf{z}_M), \end{array}\right\} \Rightarrow \begin{array}{rcl} \boldsymbol{\xi}_1 & \stackrel{d}{=} & P_{\boldsymbol{\xi}_1}^{-1}\left(P_1(\mathbf{z}_1)\right), \\ \boldsymbol{\xi}_2 & \stackrel{d}{=} & P_{\boldsymbol{\xi}_2}^{-1}\left(P_{2|1}(\mathbf{z}_2)\right), \\ \vdots & & \\ \boldsymbol{\xi}_M & \stackrel{d}{=} & P_{\boldsymbol{\xi}_M}^{-1}\left(P_{M|1:(M-1)}(\mathbf{z}_M)\right), \end{array} \tag{2.12}$$

in which $P_{i|1:(i-1)}$, $i = 1, \cdots, M$, is the PDF of $\mathbf{z}_i$ conditioned on $\mathbf{z}_1 = z_1, \mathbf{z}_2 = z_2, \cdots, \mathbf{z}_{i-1} = z_{i-1}$ induced by $P_{\boldsymbol{\mathcal{Z}}}$. The equalities, "$\stackrel{d}{=}$", above should be interpreted in the sense of distribution implying that the PDFs of the random variables in the left-hand-side (lhs) and the right-hand-side (rhs) of each equality are identical [HLD04, Theorem 2.1]. For instance, consider $P_{i|1:(i-1)}(\mathbf{z}_i)$ and $P_{\boldsymbol{\xi}_i}(\boldsymbol{\xi}_i)$ that are two random variables (functions of $\mathbf{z}_i$ and $\boldsymbol{\xi}_i$, respectively), and the PDFs of both the random variables are uniform distributions supported over $[0,\ 1]$ [HLD04, Theorem 2.1]. It can be readily shown that the

[1] In the current work, the mjpdf of $\boldsymbol{\mathcal{Z}}$ is estimated based on available information (in the present context, in the form of a set of sample joint moments computed from measurements) and the normalization constraint on pdf by relying on the MaxEnt density estimation technique (see section 2.3.1 for details).

random variables, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_M$, as defined by (2.12) are statistically independent [Ros52]. Based on this transformation, (2.11) can be written as follows by a change of variable from $\underline{\xi}$ to $\mathcal{Z}$,

$$\begin{aligned} & E_{\boldsymbol{\lambda}}\left[\mathbf{z}_k \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{Z}})\right] \\ = & \int_{\boldsymbol{\Xi}} z_k \, \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\mathcal{Z}) \, p_1(z_1) \, p_{2|1}(z_2) \cdots p_{M|1:(M-1)}(z_M) \, dz_1 \cdots dz_M. \end{aligned} \tag{2.13}$$

Here, $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\mathcal{Z}) \equiv \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(P_{\boldsymbol{\xi}_1}^{-1} P_1(z_1), P_{\boldsymbol{\xi}_2}^{-1} P_{2|1}(z_2), \cdots, P_{\boldsymbol{\xi}_M}^{-1} P_{M|1:(M-1)}(z_M))$ is defined by (2.7) with $n_d = M$, the subscript on the expectation operator in (2.13) underscores the parametrization of the underlying PDF by $\boldsymbol{\lambda}$ and $p_{i|1:(i-1)}$, $i = 1, \cdots, M$, is the conditional pdf of $\mathbf{z}_i$ satisfying $dP_{i|1:(i-1)}(z_i) = p_{i|1:(i-1)}(z_i) dz_i$. Therefore, $z_{\boldsymbol{\alpha},k}$ in (2.8) clearly depends on $\boldsymbol{\lambda}$ and can be rewritten to emphasize this dependence in the form,

$$z_{\boldsymbol{\alpha},k}(\boldsymbol{\lambda}) = \frac{E_{\boldsymbol{\lambda}}\left[\mathbf{z}_k \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{Z}})\right]}{E\left[\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}^2(\underline{\boldsymbol{\xi}})\right]}, \quad \boldsymbol{\alpha} \in \mathbb{N}^{n_d} \quad \text{and} \quad k = 1, \cdots, M. \tag{2.14}$$

As indicated earlier, the denominator in (2.14) does not depend on $\boldsymbol{\lambda}$.

Based on the discussion above, for any given $k \in \{1, \cdots, M\}$, when a simulation technique is employed, $z_{\boldsymbol{\alpha},k}(\boldsymbol{\lambda})$ can be approximated by $\widehat{z}_{\boldsymbol{\alpha},k}(\boldsymbol{\lambda})$ that is given by,

$$\widehat{z}_{\boldsymbol{\alpha},k}(\boldsymbol{\lambda}) = \frac{\dfrac{1}{K} \displaystyle\sum_{r=1}^{K} z_k^{(r)} \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\mathcal{Z}^{(r)})}{E\left[\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}^2(\underline{\boldsymbol{\xi}})\right]}, \quad \boldsymbol{\alpha} \in \mathbb{N}^{n_d} \quad \text{and} \quad k = 1, \cdots, M. \tag{2.15}$$

Here, K is a large number indicating the number of independent samples of $\boldsymbol{\mathcal{Z}}$ (and also of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{Z}})$), $z_k^{(r)}$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\mathcal{Z}^{(r)})$ are the r-th sample of the respective variable and they must be simulated from the *same* seed. First, K realizations of $\boldsymbol{\mathcal{Z}}$ are sampled independently from $P_{\boldsymbol{\mathcal{Z}}}$. Application of the Rosenblatt transformation on the r-th realization results in r-th realization of $\underline{\boldsymbol{\xi}}$ that is then substituted in the expressions of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\mathcal{Z}^{(r)}) \stackrel{d}{=} \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\xi}^{(r)})$ (see below (2.13)) to obtain the corresponding r-th realization of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{Z}})$. This procedure ensures that the simulation of $\mathbf{z}_k$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{Z}})$ are associated with the same seed.

As already mentioned, use of the Rosenblatt transformation fixes the dimension, n_d, of $\underline{\boldsymbol{\xi}}$, to the value M, i.e., $n_d = M$. This condition ($n_d = M$), however, can be relaxed at the expense of increased computational cost by using the maximum likelihood formalism [DGS06, DSG07]. However, this maximum

likelihood approach at its current state is also relatively difficult to solve and does not guarantee a unique solution for the PC coefficients of $\boldsymbol{\mathcal{Z}}$ like many constrained nonlinear optimization problems[2].

It should also be noted that changing the ordering of the components, $\mathbf{z}_1, \cdots, \mathbf{z}_M$, of $\boldsymbol{\mathcal{Z}}$ yields a different transformation, T, defined by (2.12). As there is a total of $M!$ ways in which $\mathbf{z}_1, \cdots, \mathbf{z}_M$ could be ordered, there are $M!$ sets of estimates of the PC coefficients, $\{\widehat{z}_{\boldsymbol{\alpha},k}(\boldsymbol{\lambda}), \boldsymbol{\alpha} \in \mathbb{N}^{n_d}, \boldsymbol{\alpha} \neq \mathbf{0}\}$ (since $\boldsymbol{\Upsilon}_{\mathbf{0}} = 1$, $\mathbf{0} \in \mathbb{N}^{n_d}$, $\widehat{z}_{\mathbf{0},k}(\boldsymbol{\lambda})$ is not affected). Rosenblatt [Ros52] remarked "this situation can arise in any case where there is a multitude of tests in the same context". Unless the problem under study dictates the choice of a particular order, the order, which is associated with the most conservative decision, may be the most appropriate. In the present work, no attempt is made to determine which ordering yields the most critical design. Thus, the lexigraphic ordering i.e., $\{\mathbf{z}_1, \cdots, \mathbf{z}_M\}$ is considered. Nevertheless, any complete set out of those $M!$ sets of conditional PDFs uniquely characterizes $P_{\boldsymbol{\mathcal{Z}}}$.

The estimates of the PC coefficients of $\boldsymbol{\mathcal{Y}}$ are next obtained based on $\{z_{\boldsymbol{\alpha},k},\ \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$. Substituting the PC representation, $\mathbf{z}_k \overset{d}{=} \mathbf{z}_k(\underline{\boldsymbol{\xi}}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^{n_d}} z_{\boldsymbol{\alpha},k}\, \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})$, on the right-hand-side of (2.4) and noting that each component, $\mathbf{y}_q$, $q = 1, \cdots, N$, of $\boldsymbol{\mathcal{Y}}$ on the left-hand-side of (2.4), also has a PC representation, $\mathbf{y}_q \overset{d}{=} \mathbf{y}_q(\underline{\boldsymbol{\xi}}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^{n_d}} y_{\boldsymbol{\alpha}}(\mathbf{x}_q)\, \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})$ with $\{y_{\boldsymbol{\alpha}}(\mathbf{x}_q),\ \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$ being the set of the PC coefficients, the relationship between $\{y_{\boldsymbol{\alpha}}(\mathbf{x}_q),\ \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$ and $\{z_{\boldsymbol{\alpha},k},\ \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$ is obtained, by using the fact that $\boldsymbol{\Upsilon}_{\mathbf{0}} = 1$, $\mathbf{0} \in \mathbb{N}^{n_d}$, and equating the coefficient of each orthogonal polynomial $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in \mathbb{N}^{n_d}$, on both the sides, as $y_{\boldsymbol{\alpha}}(\mathbf{x}_q) \approx [(\overline{y}_q + \sum_{k=1}^{M} v_{qk} b_k)\boldsymbol{\delta}_{\boldsymbol{\alpha}\mathbf{0}} + \sum_{k=1}^{M} v_{qk} a_k z_{\boldsymbol{\alpha},k}]$ in which $\boldsymbol{\delta}_{\boldsymbol{\alpha}\mathbf{0}}$ is the kronecker delta, $\boldsymbol{\delta}_{\mathbf{rs}} = 0$ for $\mathbf{r} \neq \mathbf{s}$ and $\boldsymbol{\delta}_{\mathbf{rs}} = 1$ for $\mathbf{r} = \mathbf{s}$, $\mathbf{r}, \mathbf{s} \in \mathbb{N}^{n_d}$, and finally, $\overline{y}_q$ and v_{qk} are, respectively, the q-th element of $\overline{\mathcal{Y}}$ and the k-th eigenvector, V_k, of C_{yy}. If $\widehat{\boldsymbol{\lambda}}_n$, based on n noisy measurements of $\boldsymbol{\mathcal{Y}}$, denotes an estimator of $\boldsymbol{\lambda}$ that characterizes $P_{\boldsymbol{\mathcal{Z}}}$ and $z_{\boldsymbol{\alpha},k}$ is replaced by its estimators, $\widehat{z}_{\boldsymbol{\alpha},k}(\widehat{\boldsymbol{\lambda}}_n)$ (see (2.15)), then the required estimators of the PC coefficients of $\mathbf{y}_q$ can be written as, $\boldsymbol{\alpha} \in \mathbb{N}^{n_d}$ and $q = 1, \cdots, N$,

$$\widehat{y}_{\boldsymbol{\alpha}}(\mathbf{x}_q, \widehat{\boldsymbol{\lambda}}_n) = \Big(\overline{y}_q + \sum_{k=1}^{M} v_{qk} b_k\Big)\boldsymbol{\delta}_{\boldsymbol{\alpha}\mathbf{0}} + \sum_{k=1}^{M} v_{qk} a_k \widehat{z}_{\boldsymbol{\alpha},k}(\widehat{\boldsymbol{\lambda}}_n), \tag{2.16}$$

in which '$\approx$' is substituted by '=' by assuming that the error in considering M dominant eigenvectors in constructing the reduced order representation of $\boldsymbol{\mathcal{Y}}$ is negligible.

[2]On the other hand, use of the principle of maximum entropy in determining $P_{\boldsymbol{\mathcal{Z}}}$, as in the current work, ensures a unique solution for $\boldsymbol{\lambda}$ in a certain sense (see section 2.3.1 for details), and therefore, for the PC coefficients of $\boldsymbol{\mathcal{Z}}$.

Let the index of $(P+1)$ retained PC coefficients be changed from $\boldsymbol{\alpha}$, $|\boldsymbol{\alpha}| \leq n_o$, to $i \in \{0, 1, \cdots, P\}$ particularly for the notational convenience in the following discussion. Denote the $(P+1)$-D vector consisting of the PC coefficients, $y_0(\mathbf{x}_q, \boldsymbol{\lambda}), \cdots, y_P(\mathbf{x}_q, \boldsymbol{\lambda})$, by $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$, $q = 1, \cdots, N$. Note that $\widehat{\mathbf{h}}_{\mathbf{x}}$, at $\mathbf{x} = \mathbf{x}_q$, in (2.1) was essentially referred to indicate $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_n$ with $\widehat{\mathbf{h}}_{\mathbf{x}}$, $\widetilde{\mathbf{h}}_{\mathbf{x}}$ and $\mathbf{h}_{\mathbf{x}}$ each now containing $(P+1)$ elements. As mentioned earlier for $\mathbf{z}_k(\underline{\boldsymbol{\xi}})$, the associated PC decomposition of $\mathbf{y}_q$ approximates $\mathbf{y}_q$ in mean-squared convergence sense implying that $E[\{\mathbf{y}_q(\underline{\boldsymbol{\xi}}) - \sum_{i=0}^{[((n_o+n_d)!/(n_o!\, n_d!))-1]} y_i(\mathbf{x}_q)\, \boldsymbol{\Upsilon}_i(\underline{\boldsymbol{\xi}})\}^2] \longrightarrow 0$ as $n_o \longrightarrow \infty$ [SG04a]. It should, however, be bear in mind that the true mapping, $\underline{\boldsymbol{\xi}} \longmapsto \mathbf{y}_q(\underline{\boldsymbol{\xi}})$, that is unknown in reality, is defined here by (2.4) and (2.12) implying that $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)$ essentially refers to $\widetilde{\mathbf{h}}_{\mathbf{x}}$ at $\mathbf{x} = \mathbf{x}_q$, where $\boldsymbol{\lambda}^*$ is the true (in the absence of data error) value of $\boldsymbol{\lambda}$. The Rosenblatt transformation defined by (2.12) essentially guarantees the fact that the observed empirical mjpdf and therefore, the observed sample statistics match well with the same obtained from the constructed PC decomposition from which the digital realizations can be easily and efficiently simulated. Finally, note that $\widehat{\boldsymbol{\lambda}}_n$ is the MLE of $\boldsymbol{\lambda}$ since, in the present work, MaxEnt density estimation (MEDE) technique is employed to obtain $P_{\boldsymbol{\mathcal{Z}}}$ (see section 2.3.2 for details). This implies that $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is also the MLE of $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ [CB02, p. 320-321]. In the next section, an asymptotic distribution of $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is obtained.

2.2.4 Asymptotic Probability Distribution Function of $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$

The FIM has proven to be useful in determining the apdf of a deterministic mapping of a random parameter [CLR96, HD97]. If K in (2.15) is large enough and the effect of change in data is assumed to be manifested only through a change, $\Delta\boldsymbol{\lambda}$, in $\boldsymbol{\lambda}$, then $\boldsymbol{\lambda} \longmapsto \mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ is a deterministic function of $\boldsymbol{\lambda}$. Here, $\Delta\boldsymbol{\lambda}$ is a $p \times 1$ column vector of elements $\Delta\lambda_j$ in which $\Delta\lambda_j$ is a change in λ_j, $j = 1, \cdots, p$. The FIM would then be useful in constructing an apdf that can be used to obtain a confidence interval on $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$.

The second condition (manifestation of change in data only via $\Delta\boldsymbol{\lambda}$) implies that the sensitivities of v_{qk}, a_k and b_k, $q = 1, \cdots, N$, $k = 1, \cdots, M$, w.r.t. $\boldsymbol{\lambda}$ are very small. These assumptions would not have been required if $\boldsymbol{\lambda}$ were estimated directly from the observations of $\boldsymbol{\mathcal{Y}}$ without applying the mappings defined by (2.2) and (2.3). While this route would have directly (i.e., not via (2.16)) yielded the estimators of the PC coefficients of $\boldsymbol{\mathcal{Y}}$, (2.2) and (2.3) are useful, respectively, for reduction of the dimension of the problem (consequently, reduction of the computational cost) and enhancement of the efficiency of the numerical algorithm employed for MEDE technique.

Therefore, $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ becomes a deterministic function, by (2.16), of $\widehat{z}_{i,k}(\boldsymbol{\lambda})$, $i = 0, \cdots, P$ and $k = 1, \cdots, M$. Since many simulation techniques usually guarantee $O(1/K)$ rate of convergence of $\mathrm{var}\,[\widehat{z}_{i,k}(\boldsymbol{\lambda})]$ to some small number, $\epsilon > 0$, the enforcement of the first condition (large K) now is likely to ensure that the effect of the error of finite K on $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ can be neglected. Consequently, $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ is treated here as a *deterministic function of* $\boldsymbol{\lambda}$. The variability on $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ would then primarily be governed by the error in the estimator, $\widehat{\boldsymbol{\lambda}}_n$.

In addition to $\widehat{\boldsymbol{\lambda}}_n$ being the MLE of $\boldsymbol{\lambda}$, use of MEDE technique also has the following two consequences in the present work: (1) the density estimate belongs to exponential family [CB02, Section 3.4]; and (2) by (2.13) and by the first consequence as just mentioned in (1), $\mathbf{h}_{\mathbf{x}_q}(\cdot)$ is differentiable w.r.t. $\boldsymbol{\lambda}$. Let $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ be represented by a $(P+1) \times 1$ column vector, $[y_0(\mathbf{x}_q, \boldsymbol{\lambda}), \cdots, y_P(\mathbf{x}_q, \boldsymbol{\lambda})]^T$. Also assume that the $p \times (P+1)$ gradient matrix, $\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$, of $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$ is not a zero matrix. Then, by (1), (2) and the MLE property of $\widehat{\boldsymbol{\lambda}}_n$, it can be shown that [CB02, Theorem 10.1.12, p. 338-339], [Spa03, p. 359-360],

$$\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n) \overset{\text{approx.}}{\sim} N(\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}), \mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})^T \, \mathbf{F}_n(\boldsymbol{\lambda})^{-1} \, \mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})), \;\; q = 1, \cdots, N, \tag{2.17}$$

implying that $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is a consistent and asymptotically efficient estimator of $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda})$. Here, $N(\cdot)$ represents a $(P+1)$-D Gaussian distribution and $\mathbf{F}_n(\boldsymbol{\lambda})$ is the FIM. Equation (2.17) is true for $\boldsymbol{\lambda}$ close to (unknown) $\boldsymbol{\lambda}^*$ when n, representing the number of measurements of $\boldsymbol{\mathcal{Y}}$, is reasonably large. In practice, $\boldsymbol{\lambda}$ is often set to $\widehat{\boldsymbol{\lambda}}_n$ to evaluate the mean vector and covariance matrix of the asymptotic distribution in (2.17). Clearly, the prediction $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ has an uncertainty given by this approximate normal distribution. This uncertainty provides some sense of how much $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is likely to differ from $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*) \equiv \widetilde{\mathbf{h}}_{\mathbf{x}_q}$. This approximate distribution is useful in propagating the error, $\mathbf{h}_{\mathbf{x}}(\widehat{\boldsymbol{\lambda}}_n) - \widetilde{\mathbf{h}}_{\mathbf{x}}$, to the model-based predictions when $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ represents some stochastic parameter in the model.

Next, a discussion of the techniques for estimating the mjpdf of $\boldsymbol{\mathcal{Z}}$ parameterized by $\boldsymbol{\lambda}$, $\mathbf{F}_n(\boldsymbol{\lambda})$ and $\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$ are provided.

2.3 Estimations of the mjpdf of the nKL Vector, the Fisher Information Matrix and the Gradient Matrix

Since the mjpdf of $\boldsymbol{\mathcal{Z}}$, that is parameterized by $\boldsymbol{\lambda}$, is estimated by using MEDE technique, a brief discussion of this technique, its relationship to MLE and the specific estimation technique employed in the current work are included in section 2.3.1. The estimation techniques of the FIM, $\mathbf{F}_n(\boldsymbol{\lambda})$, and the gradient matrix, $\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$, are provided, respectively, in sections 2.3.4 and 2.3.5.

2.3.1 Multivariate Joint Probability Density Function of the nKL Vector

Given a finite data set, a density estimation technique consists of evaluating a pdf that is consistent, in some sense, with the data set. In general, this is an ill-posed problem because the solution is non-unique since many (possibly infinite) probability density functions can generate this specific data set with positive probability. The problem becomes more challenging in a multidimensional setting given the large amount of data required to estimate the density. If a priori information is available about the characteristics and functional form of the density, parametric estimation techniques making use of this information, can significantly reduce the amount of data required for density estimation. However, a priori or additional information may not always be available in many cases, and nonparametric density estimation techniques [Ize91] become useful in such situations. Kernel density estimation (KDE) techniques are some of the best-developed techniques in the literature and have been well-adapted to the multivariate case [Sco92, Chapter 6]. However, it suffers from a few drawbacks, for example, it usually shows spurious lobes and bumps in the estimates of the density functions (see Figure 2.5 showing a few bumps and lobes near the tail of the marpdf, based on measurement data, estimated by using KDE technique) and is computationally demanding for multivariate problems.

In the current work, an estimate, $p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) \equiv p_{\boldsymbol{\mathcal{Z}}}(z_1, z_2, \cdots, z_M)$, of the mjpdf of $\boldsymbol{\mathcal{Z}}$ is obtained by relying on the MEDE technique [SKR00, Wu03] that is based on the MaxEnt principle [Sha48, Jay57a, Jay57b, KK92]. Here, 'entropy' can be treated as a quantitative measure of uncertainty. The MaxEnt principle essentially states that in the absence of a priori knowledge about the probability model of the random quantity under consideration, a PDF should be selected that is most consistent with the available information contained in the given data set and closest to the uniform distribution (since uniform distribution has maximum entropy or uncertainty on a bounded support in absence of a priori knowledge) in

a space of probability distribution functions equipped with a suitable metric (not necessarily Euclidean metric). This is achieved by maximizing the entropy or uncertainty, $H(p_{\boldsymbol{\mathcal{Z}}})$, of $p_{\boldsymbol{\mathcal{Z}}}$ given by [KK92, p. 68],

$$H(p_{\boldsymbol{\mathcal{Z}}}) = -\int_{\boldsymbol{\Xi}} p_{\boldsymbol{\mathcal{Z}}}(z_1, \cdots, z_M) \ln \left[p_{\boldsymbol{\mathcal{Z}}}(z_1, \cdots, z_M)\right] d\mathcal{Z}, \tag{2.18}$$

subject to the available information and the normalization constraint on the pdf. In the current work, the available information is considered to be a set of sample joint-moment constraints based on the available finite data set of measurements. Since $H(\cdot)$ is a concave function of $p_{\boldsymbol{\mathcal{Z}}}$ and the moment constraints are linear in $p_{\boldsymbol{\mathcal{Z}}}$, MEDE technique guarantees the existence of a $p^*_{\boldsymbol{\mathcal{Z}}}$, satisfying the moment constraints, for which $H(\cdot)$ attains its global maximum.

The joint moments of $\boldsymbol{\mathcal{Z}}$ are defined by,

$$\beta_j = E\left[\mathbf{z}_1^{m_{1j}}\, \mathbf{z}_2^{m_{2j}}\, \cdots \mathbf{z}_M^{m_{Mj}}\right] = \int_{\boldsymbol{\Xi}} \left(\prod_{i=1}^{M} z_i^{m_{ij}}\right) p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, d\mathcal{Z}, \qquad j = 0, \cdots, p,$$

in which m_{ij}'s characterize the joint-moments. These joint-moments can be estimated from the given set of measurement as follows,

$$\widehat{\beta}_j = \frac{1}{n} \sum_{k=1}^{n} \left[\prod_{i=1}^{M} \left(z_i^{(k)}\right)^{m_{ij}}\right], \; j = 0, \cdots, p, \tag{2.19}$$

in which $z_i^{(k)}$ is the i-th component of the k-th sample, $\mathcal{Z}_k = [z_1^{(k)}, \cdots, z_M^{(k)}]^T$, of $\boldsymbol{\mathcal{Z}}$. Here, $j = 0$ refers to the normalization of pdf implying that $m_{i0} = 0, \; \forall \; i = 1, \cdots, M$, and $\beta_0 = \widehat{\beta}_0 = 1$.

The *primal problem* associated with the MaxEnt constrained optimization problem is, therefore, given by,

$$\begin{aligned} &\text{minimize} \;\; [-H(p_{\boldsymbol{\mathcal{Z}}})] \\ &\text{subject to } \beta_j = \widehat{\beta}_j, \;\; j = 0, \cdots, p. \end{aligned}$$

The Lagrangian function associated with the primal problem is defined by,

$$\mathcal{L}(p_{\boldsymbol{\mathcal{Z}}}, \boldsymbol{\lambda}) = -H(p_{\boldsymbol{\mathcal{Z}}}) + (\lambda_0 - 1)\left[\int_{\boldsymbol{\Xi}} p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, d\mathcal{Z} - 1\right] + \sum_{j=1}^{p} \lambda_j \left[\int_{\boldsymbol{\Xi}} \left(\prod_{i=1}^{M} z_i^{m_{ij}}\right) p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, d\mathcal{Z} - \widehat{\beta}_j\right],$$

in which $(\lambda_0 - 1)$ and λ_j's are Lagrange multipliers and $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_p]^T$. It is shown below that λ_0 depends on $\boldsymbol{\lambda}$, and therefore, $\mathcal{L}(\cdot)$ is only shown as a function of $p_{\boldsymbol{\mathcal{Z}}}$ and $\boldsymbol{\lambda}$. By using the theory of calculus of variations, the critical (stationary) point, representing the primal optimal solution, $(p^*_{\boldsymbol{\mathcal{Z}}}, \widehat{\lambda}_0, \widehat{\boldsymbol{\lambda}}_n)$, in which $\widehat{\boldsymbol{\lambda}}_n \equiv [\widehat{\lambda}_1, \cdots, \widehat{\lambda}_p]^T$, can be analytically determined to find that $p^*_{\boldsymbol{\mathcal{Z}}}$ belongs to the following exponential parametric family,

$$p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \boldsymbol{\lambda}) = \exp\Big[-\sum_{j=0}^{p} \lambda_j \Big(\prod_{i=1}^{M} z_i^{m_{ij}} \Big) \Big] \mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z}), \quad \boldsymbol{\lambda} \in \mathbb{R}^p, \tag{2.20}$$

in which $\mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z})$ is the indicator function implying $\mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z}) = 1$ if $\mathcal{Z} \in \boldsymbol{\Xi}$ and $\mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z}) = 0$ if $\mathcal{Z} \notin \boldsymbol{\Xi}$. The MaxEnt parameters, $\widehat{\lambda}_j$, $j = 0, \cdots, p$, are determined by solving the following non-linear equations representing the imposed constrains,

$$\int_{\boldsymbol{\Xi}} \Big(\prod_{i=1}^{M} z_i^{m_{ij}} \Big) p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, d\mathcal{Z} = \widehat{\beta}_j, \;\; j = 0, \cdots, p, \tag{2.21}$$

with $p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) = p^*_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})$. As the parametric family in (2.20) represents a pdf, it satisfies the normalizing constraint, $\int_{\mathbb{R}^M} p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \boldsymbol{\lambda})\, d\mathcal{Z} = 1$, implying,

$$\lambda_0 = \ln\Big[\int_{\boldsymbol{\Xi}} \exp\Big[-\sum_{j=1}^{p} \lambda_j \Big(\prod_{i=1}^{M} z_i^{m_{ij}} \Big) \Big]\, d\mathcal{Z} \Big] \equiv \xi(\boldsymbol{\lambda}). \tag{2.22}$$

Therefore, the form of the parametric family can be compactly written as,

$$p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \boldsymbol{\lambda}) = \exp\Big[-\boldsymbol{\lambda}^T T(\mathcal{Z}) - \xi(\boldsymbol{\lambda}) \Big] \mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z}), \tag{2.23}$$

in which $T(\mathcal{Z}) \equiv T(z_1, \cdots, z_M) = [t_1(\mathcal{Z}), \cdots, t_p(\mathcal{Z})]^T$ where $t_j(\mathcal{Z}) \equiv t_j(z_1, \cdots, z_M)$ is defined by,

$$t_j(\mathcal{Z}) = \prod_{i=1}^{M} z_i^{m_{ij}}, \quad j = 1, \cdots, p. \tag{2.24}$$

Note that the nonnegativity property of the pdf is already satisfied by the exponential family [CB02, Section 3.4] in (2.23). The next section describes the relationship between the MaxEnt probability model and the maximum likelihood probability model, which can also be found in earlier literature in the context of other applications [BTC79, BTTC88, BPP96, FRT97].

2.3.2 Relationship between MaxEnt and Maximum Likelihood Probability Models

Consider the following *dual function* associated with the primal problem defined earlier,

$$\Psi(\boldsymbol{\lambda}) = \min_{p_{\boldsymbol{\mathcal{Z}}} \in \mathcal{P}} \mathcal{L}(p_{\boldsymbol{\mathcal{Z}}}, \boldsymbol{\lambda}),\ \boldsymbol{\lambda} \in \mathbb{R}^p,$$

in which $\mathcal{P} = \{p_{\boldsymbol{\mathcal{Z}}} : \int_{\boldsymbol{\Xi}} p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, d\mathcal{Z} = 1\}$. By using (2.20) and (2.22), $\Psi(\boldsymbol{\lambda})$ can be explicitly calculated as,

$$\Psi(\boldsymbol{\lambda}) = -\xi(\boldsymbol{\lambda}) - \sum_{j=1}^{p} \lambda_j \widehat{\beta}_j, \tag{2.25}$$

and the corresponding *dual problem* can be formulated as,

$$\begin{aligned} &\text{maximize } \Psi(\boldsymbol{\lambda}) \\ &\text{subject to } \boldsymbol{\lambda} \in \mathbb{R}^p. \end{aligned} \tag{2.26}$$

This is an unconstrained optimization problem and the dual optimal solution is given by $\widetilde{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda} \in \mathbb{R}^p} \Psi(\boldsymbol{\lambda})$. Then, by duality theorem [Ber99, Section 5.1], under suitable conditions, the following is obtained,

$$\widehat{\boldsymbol{\lambda}}_n = \widetilde{\boldsymbol{\lambda}} \Rightarrow p^*_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) = p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \widetilde{\boldsymbol{\lambda}}).$$

Now, consider a set of statistically independent and identically distributed (i.i.d.) random data vector $\{\boldsymbol{\mathcal{Z}}_1, \boldsymbol{\mathcal{Z}}_2, \cdots, \boldsymbol{\mathcal{Z}}_n\}$ with each $\boldsymbol{\mathcal{Z}}_i \sim p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \boldsymbol{\lambda})$, $i = 1, \cdots, n$. Then, the empirical PDF of $\boldsymbol{\mathcal{Z}}$, $\widetilde{p}_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})$, based on this data set can be defined by,

$$\widetilde{p}_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) \equiv \frac{1}{n} \times (\text{number of times that } \mathcal{Z} \text{ appears in the i.i.d. data set})\,.$$

Consequently, $\widehat{\beta}_j$ in (2.19) can be alternatively represented as,

$$\widehat{\beta}_j = \sum_{\mathcal{Z} \in \boldsymbol{\Xi}} \widetilde{p}_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})\, t_j(\mathcal{Z}), \quad j = 1, \cdots, p. \tag{2.27}$$

Next, stacking the vectors of this random data in $\mathbf{Z}_n$, i.e., $\mathbf{Z}_n = [\boldsymbol{\mathcal{Z}}_1^T, \cdots, \boldsymbol{\mathcal{Z}}_n^T]^T$, the mjpdf of $\mathbf{Z}_n$ is given by $p_{\mathbf{Z}_n}(\mathbf{Z}_n \,|\, \boldsymbol{\lambda}) = \prod_{i=1}^n p_{\boldsymbol{\mathcal{Z}}_i}(\mathcal{Z}_i, \boldsymbol{\lambda})$ in which $p_{\boldsymbol{\mathcal{Z}}_i}(\mathcal{Z}_i, \boldsymbol{\lambda}) \equiv p_{\boldsymbol{\mathcal{Z}}}(z_1^{(i)}, \cdots, z_M^{(i)}, \boldsymbol{\lambda})$. The likelihood function of $\boldsymbol{\lambda}$ is then defined by $\ell(\boldsymbol{\lambda} \mid \mathbf{Z}_n) = p_{\mathbf{Z}_n}(\mathbf{Z}_n \mid \boldsymbol{\lambda})$. Finally, by using (2.25) and (2.27), the associated log-likelihood function can be shown to be given by,

$$\ln \ell(\boldsymbol{\lambda} \,|\, \mathbf{Z}_n) = n\Psi(\boldsymbol{\lambda}).$$

This facts implies that maximizing the log-likelihood function is equivalent to maximizing the dual function as defined by the dual problem in (2.26). With this interpretation, it can be stated that the MaxEnt mjpdf of $\boldsymbol{\mathcal{Z}}$, $p^*_{\boldsymbol{\mathcal{Z}}}(z_1, \cdots, z_M)$, is the mjpdf in the parametric family, $\{p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}, \boldsymbol{\lambda}), \boldsymbol{\lambda} \in \mathbb{R}^p\}$, that maximizes the log-likelihood function of $\boldsymbol{\lambda}$. This appealing fact reinforces the reasoning as to why the MaxEnt principle can be preferred in estimating the mjpdf of $\boldsymbol{\mathcal{Z}}$.

It is shown later in section 2.3.4 (see (2.33)) that the (r, s)-th element of the Hessian matrix of $\ln \ell(\boldsymbol{\lambda} \mid \mathbf{Z}_n)$ is given by $-n \,\mathrm{cov}\,[t_r(\boldsymbol{\mathcal{Z}})t_s(\boldsymbol{\mathcal{Z}})]$, $r, s = 1, \cdots, p$. If the moment constraints in (2.21) are imposed such that $\{1, t_1(\mathcal{Z}), \cdots, t_p(\mathcal{Z})\}$ is a linearly independent set, then the covariance matrix, $\mathrm{cov}\,[t_r(\boldsymbol{\mathcal{Z}})t_s(\boldsymbol{\mathcal{Z}})]$, is always positive definite implying that the Hessian matrix is always negative definite. Thus, $\ln \ell(\cdot \,|\, \mathbf{Z}_n)$ is a strictly concave function of $\boldsymbol{\lambda}$ guaranteeing the existence of $\widehat{\boldsymbol{\lambda}}_n$ for which $\ln \ell(\cdot \,|\, \mathbf{Z}_n)$ attains the global maximum value conforming to the fact that $H(\cdot)$ has a global maxima at $p^*_{\boldsymbol{\mathcal{Z}}}$.

2.3.3 MEDE Technique and Some Remarks on the Form of $p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})$

Based on (2.23), the estimate of the mjpdf of $\boldsymbol{\mathcal{Z}}$ is obtained as,

$$p^*_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) \equiv p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) = \exp\Big[-\boldsymbol{\lambda}^T T(\mathcal{Z}) - \xi(\boldsymbol{\lambda})\Big]\, \mathbb{I}_{\boldsymbol{\Xi}}(\mathcal{Z}), \tag{2.28}$$

in which λ_j, the elements of $\boldsymbol{\lambda}$, are obtained by solving the following set of nonlinear equations,

$$\int_{\boldsymbol{\Xi}} t_j(\mathcal{Z}) \exp\Big[-\boldsymbol{\lambda}^T T(\mathcal{Z}) - \xi(\boldsymbol{\lambda})\Big]\, d\mathcal{Z} = \widehat{\beta}_j, \;\; j = 1, \cdots, p. \tag{2.29}$$

Here, $\widehat{\beta}_j$ are the sample joint-moments computed by (2.19), and $\xi(\boldsymbol{\lambda})$ is given by,

$$\xi(\boldsymbol{\lambda}) = \ln\left(\int_{\boldsymbol{\Xi}} \exp\Big[-\boldsymbol{\lambda}^T T(\mathcal{Z})\Big]\, d\mathcal{Z}\right). \tag{2.30}$$

The set of equations in (2.29) forms a set of p nonlinear equations in p unknowns, λ_j's. The MaxEnt probability model in (2.28) is then obtained by solving this set of nonlinear equations that involves computation of M-dimensional integrations as shown in (2.29) and (2.30).

For a scalar random variable case, a numerical technique has been developed that uses a sequential updating procedure [Wu03] in conjunction with the Newton-Raphson algorithm. The sequential updating method imposes the sample moment constraints one at a time from the lower order moments to the higher order moments, and updates the pdf sequentially.

The implementation of the Newton-Raphson algorithm requires an initial guess for $\boldsymbol{\lambda}$ to which both the convergence and the rate of convergence of the algorithm are highly sensitive. Furthermore, in the multivariate case, a difficulty arises due to the fact that several moments are associated with a given order. Consequently, the additional information in a set of moments having the same given order ('additional' in the sense that the information in addition to that contained in the set of moments having order lower than the given order) are distributed among those moments in a disjoint fashion. The primary difficulty in such cases, therefore, becomes the issue of choosing a reasonable initial guess for $\boldsymbol{\lambda}$. The sequential updating method in conjunction with Newton-Raphson method is likely to fail in such a situation. Therefore, a method, that does not depend *much* on the choice of the initial guess for $\boldsymbol{\lambda}$, would be more useful. In the present work, the Levenberg-Marquardt method, a nonlinear least squares technique, as used in earlier literature [SKR00] in the context of a scalar-valued random variable case is used for this purpose. This method may suffer from a slower convergence rate in the event that a good initial guess is available for the Newton-Raphson method. The true potential of the Levenberg-Marquardt algorithm is, however, realized in the absence of such an initial guess.

To implement the Levenberg-Marquardt technique, the residuals of (2.29) are written as follows by using (2.30),

$$R_j = 1 - \frac{\displaystyle\int_{\boldsymbol{\Xi}} t_j(\mathcal{Z}) \exp\left[-\boldsymbol{\lambda}^T T(\mathcal{Z})\right] \, d\mathcal{Z}}{\widehat{\beta}_j \displaystyle\int_{\boldsymbol{\Xi}} \exp\left[-\boldsymbol{\lambda}^T T(\mathcal{Z})\right] \, d\mathcal{Z}}, \quad j = 1, \cdots, p. \tag{2.31}$$

The unknown parameter vector, $\boldsymbol{\lambda}$, can be evaluated by using the Levenberg-Marquardt method with or without the sequential updating method by minimizing the sum of squares of the residuals in (2.31). The interesting feature of the sequential updating method, however, is that it generates a sequence of $\boldsymbol{\lambda}$ (i.e., the mjpdf of $\boldsymbol{\mathcal{Z}}$) associated with the sequential activation of the joint-moment constraints. In the

present work, a hybrid MATLAB-FORTRAN program is written to perform this task. The main program written in MATLAB calls a FORTRAN numerical integration subroutine to speed up the process. The MATLAB command, `lsqnonlin`, is used to perform the nonlinear least-square technique with the Levenberg-Marquardt method option '`On`' to evaluate $\boldsymbol{\lambda}$. The vector, $\boldsymbol{\lambda}$, would be treated further as the model parameter that encompasses all the information contained in the measurements of $\boldsymbol{\mathcal{Y}}$. Denote the estimated model parameters collectively by $\widehat{\boldsymbol{\lambda}}_n = [\widehat{\lambda}_1, \cdots, \widehat{\lambda}_p]^T$.

It should be noted here that $\widehat{\boldsymbol{\lambda}}_n$ is a random column vector with randomness primarily being induced by the measurements of $\boldsymbol{\mathcal{Z}}$ (i.e., of $\boldsymbol{\mathcal{Y}}$). There also exist other factors, for example, measurement error, numerical error induced by the numerical method employed for solving the set of nonlinear equations etc., affecting the estimation of $\boldsymbol{\lambda}$. However, the effects of these factors are assumed to be within acceptable tolerance. It is clear that the true value, $\boldsymbol{\lambda}^*$, of $\boldsymbol{\lambda}$ is not known. Once the parameter $\boldsymbol{\lambda}$ is specified, an estimate of the mjpdf of $\boldsymbol{\mathcal{Z}}$ is known precisely by (2.28). Denote the associated joint PDF by MaxEPD($\mathcal{Z}, \boldsymbol{\lambda}$) in which MaxEPD stands for MAXimum-Entropy joint Probability Distribution. On p. 16, the estimate of the PDF of $\boldsymbol{\mathcal{Z}}$, in the context of the current work, refers to MaxEPD($\mathcal{Z}, \boldsymbol{\lambda}$) implying that $P_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{\mathcal{Z}}) \equiv \mathrm{MaxEPD}(\mathcal{Z}, \boldsymbol{\lambda})$.

2.3.4 Computation of the Fisher Information Matrix, $\mathbf{F}_n(\boldsymbol{\lambda})$

An estimate of the FIM as required in (2.17) is provided in this section. Consider a sequence of i.i.d. random data vectors, $\{\boldsymbol{\mathcal{Z}}_1, \boldsymbol{\mathcal{Z}}_2, \cdots, \boldsymbol{\mathcal{Z}}_n\}$, with each $\boldsymbol{\mathcal{Z}}_v \sim \mathrm{MaxEPD}(\mathcal{Z}, \boldsymbol{\lambda})$, $v = 1, \cdots, n$. The mjpdf of $\mathbf{Z}_n = [\boldsymbol{\mathcal{Z}}_1^T, \cdots, \boldsymbol{\mathcal{Z}}_n^T]^T$ is given by,

$$p_{\mathbf{Z}_n}(\mathbf{Z}_n \,|\, \boldsymbol{\lambda}) = \prod_{v=1}^{n} p_{\boldsymbol{\mathcal{Z}}_v}(\mathcal{Z}_v).$$

Consider the $p \times p$ FIM, $\mathbf{F}_n(\boldsymbol{\lambda})$, given by [Spa03, Section 13.3.2],

$$\mathbf{F}_n(\boldsymbol{\lambda}) \equiv E\left[\frac{\partial \ln\, \ell(\boldsymbol{\lambda}\,|\,\mathbf{Z}_n)}{\partial \boldsymbol{\lambda}} \cdot \frac{\partial \ln\, \ell(\boldsymbol{\lambda}\,|\,\mathbf{Z}_n)}{\partial \boldsymbol{\lambda}^T} \,\middle|\, \boldsymbol{\lambda} \right] = -E\left[\frac{\partial^2 \ln\, \ell(\boldsymbol{\lambda}\,|\,\mathbf{Z}_n)}{\partial \boldsymbol{\lambda}\, \partial \boldsymbol{\lambda}^T} \,\middle|\, \boldsymbol{\lambda} \right]. \tag{2.32}$$

Here, in a general case, the equality, '=', is followed [Spa03, p. 352-353] by assuming that $\ln \ell(\cdot \,|\, \mathbf{Z}_n)$ is twice differentiable w.r.t. $\boldsymbol{\lambda}$ and the regularity conditions [CB02, Section 10.6.2] hold for ℓ. Since $p_{\boldsymbol{\mathcal{Z}}}$ belongs to an exponential family [CB02, Section 3.4], the equality, '=', in (2.32) holds true in the current

context (see also [CB02, Section 2.4 and Lemma 7.3.11]). The log-likelihood function, in the present work, can be explicitly computed and is shown below,

$$\ln \ell(\boldsymbol{\lambda} \,|\, \mathbf{Z}_n) = \ln \left[p_{\mathbf{Z}_n}(\mathbf{Z}_n \,|\, \boldsymbol{\lambda}) \right] = -n \left[\xi(\boldsymbol{\lambda}) + \sum_{j=1}^{p} \lambda_j \left\{ \frac{1}{n} \sum_{v=1}^{n} \left(\prod_{i=1}^{M} \left(z_i^{(v)} \right)^{m_{ij}} \right) \right\} \right].$$

It is also straightforward to compute the second derivative of $\ln \ell(\cdot \,|\, \mathbf{Z}_n)$ w.r.t. the elements, λ_r and λ_s, $r, s = 1, \cdots, p$, of $\boldsymbol{\lambda}$, and the second derivative can be shown to be given by,

$$\frac{\partial^2 \ln \ell(\boldsymbol{\lambda} \,|\, \mathbf{Z}_n)}{\partial \lambda_r \, \partial \lambda_s} = -n \, \mathrm{cov} \left[t_r(\boldsymbol{\mathcal{Z}}) t_s(\boldsymbol{\mathcal{Z}}) \right], \tag{2.33}$$

in which $\boldsymbol{\mathcal{Z}} \sim \text{MaxEPD}(\mathcal{Z}, \boldsymbol{\lambda})$. The specific value of interest for $\boldsymbol{\lambda}$ here is $\widehat{\boldsymbol{\lambda}}_n$. It should also be noted that, by the definition of $t_j(\mathcal{Z})$ in (2.24), a few of these covariance terms should already be known by the right-hand-side of the imposed moment constraints as defined earlier by (2.21) or (2.29). This is particularly expected to happen for low values of r and s and consequently, in those cases, it is not required to evaluate the M-dimensional integration over $\boldsymbol{\Xi}$ required for the computation of $\mathrm{cov} \left[t_r(\boldsymbol{\mathcal{Z}}) t_s(\boldsymbol{\mathcal{Z}}) \right]$. Some of the elements of the upper diagonal block of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ would, therefore, be known and the rest of the elements unknown implying that $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ can be divided into known and unknown parts.

Since the (r, s)-th element of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ is given by $n \, \mathrm{cov} \left[t_r(\boldsymbol{\mathcal{Z}}) t_s(\boldsymbol{\mathcal{Z}}) \right]$ by (2.32)-(2.33), the FIM can be computed by estimating these covariance terms based on $p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})$ in (2.28) with $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_n$. The associated multidimensional integration can be carried out by employing a numerical integration technique or a simulation technique. The covariance terms associated with the elements of the known part need not be computed again since these elements are already known.

2.3.5 Computation of the Gradient Matrix, $\mathrm{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$

In this section, estimate of the gradient matrix as required in (2.17) is considered. Denote the i-th column of $\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$ by $\partial \widehat{y}_{i-1}(\mathbf{x}_q, \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}$, $i = 1, \cdots, (P+1)$, that is a $p \times 1$ column vector of elements $\partial \widehat{y}_{i-1}(\mathbf{x}_q, \boldsymbol{\lambda}) / \partial \lambda_j$, $j = 1, \cdots, p$. Clearly, this column vector can be determined by using (2.16) as follows,

$$\frac{\partial \widehat{y}_i(\mathbf{x}_q, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \sum_{k=1}^{M} v_{qk} \, a_k \, \widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda}), \quad i = 0, \cdots, P, \quad \text{and} \quad q = 1, \cdots, N. \tag{2.34}$$

Here, the gradient vector, $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda}) \equiv \partial\widehat{z}_{i,k}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$, is a $p \times 1$ column vector of elements, $\widehat{g}_{i,k}(\boldsymbol{\lambda}) \equiv \partial\widehat{z}_{i,k}(\boldsymbol{\lambda})/\partial\lambda_j$, $j = 1, \cdots, p$. By (2.14), the gradient vector, $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda})$, essentially is an estimator of the vector given by,

$$\mathbf{g}_{i,k}(\boldsymbol{\lambda}) \equiv \frac{\partial z_{i,k}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = \frac{1}{E\left[\boldsymbol{\Upsilon}_i^2(\underline{\boldsymbol{\xi}})\right]} \frac{\partial E_{\boldsymbol{\lambda}}\left[\mathbf{z}_k \boldsymbol{\Upsilon}_i(\boldsymbol{\mathcal{Z}})\right]}{\partial\boldsymbol{\lambda}}, \qquad \begin{matrix} i = 0, \cdots, P, \\ k = 1, \cdots, M. \end{matrix} \tag{2.35}$$

This can be calculated analytically by differentiating the resulting expression of (2.13) w.r.t. $\boldsymbol{\lambda}$ and substituting the result in (2.35). The integration and differentiation can be performed numerically. Another way to obtain an approximation of $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda})$ is to employ the classical finite-difference (FD) technique [Spa03, Section 6.3]. The two-sided FD approximation of $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda})$ for use with (2.34) is given by,

$$\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda}) \approx \widehat{\widehat{\mathbf{g}}}_{i,k}(\boldsymbol{\lambda}) = \begin{bmatrix} \dfrac{\widehat{z}_{i,k}(\boldsymbol{\lambda} + c\mathbf{1}_1) - \widehat{z}_{i,k}(\boldsymbol{\lambda} - c\mathbf{1}_1)}{2c} \\ \vdots \\ \dfrac{\widehat{z}_{i,k}(\boldsymbol{\lambda} + c\mathbf{1}_p) - \widehat{z}_{i,k}(\boldsymbol{\lambda} - c\mathbf{1}_p)}{2c} \end{bmatrix}, \tag{2.36}$$

in which $\widehat{\widehat{\mathbf{g}}}_{i,k}(\boldsymbol{\lambda})$ is the two-sided FD approximation of $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda})$, $\mathbf{1}_j$ denotes a $p \times 1$ column vector with 1 at the j-th place and 0 elsewhere and $c > 0$ is a small scalar.

The classical FD approximation technique requires $2p$ evaluations of $\widehat{z}_{i,k}(\cdot)$. Since the number of evaluations of $\widehat{z}_{i,k}(\cdot)$ grows with p for FD technique, simultaneous perturbation (SP) gradient approximation technique, introduced in the field of stochastic optimization [Spa92] (see [Spa03, Section 7.2] for a relatively simpler version of [Spa92]), might be useful for large p. This technique calls for averaging the gradient approximation over a multiple number of iterations and the number of estimation is only two per iteration regardless of the dimension of p. However, the FD approximation generally provides a superior approximation of $\widehat{\mathbf{g}}_{i,k}(\boldsymbol{\lambda})$ than its SP counterpart but the computational savings of SP technique might be a significant benefit for large p.

2.4 Numerical Illustration and Discussions

Consider a second-order random process, $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$, representing some random system parameter, evolved over a rectangular spatial domain, $\mathbf{D}$, of size 1.0×0.8 in appropriate length scale.

2.4.1 Measurement of the Stochastic Process

In the current work, experimental measurements of $\boldsymbol{\mathcal{Y}}$, the finite-dimensional representation of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$, are not available. Therefore, the realizations of $\boldsymbol{\mathcal{Y}}$ are digitally simulated and these simulated realizations are assumed to be proxy for the experimental measurements.

The measurements of the stochastic parameter are assumed to be available at $N = 100$ locations over $\mathbf{D}$ as shown in Figure 2.1. Instead of choosing the coordinates of the measurement locations at random,

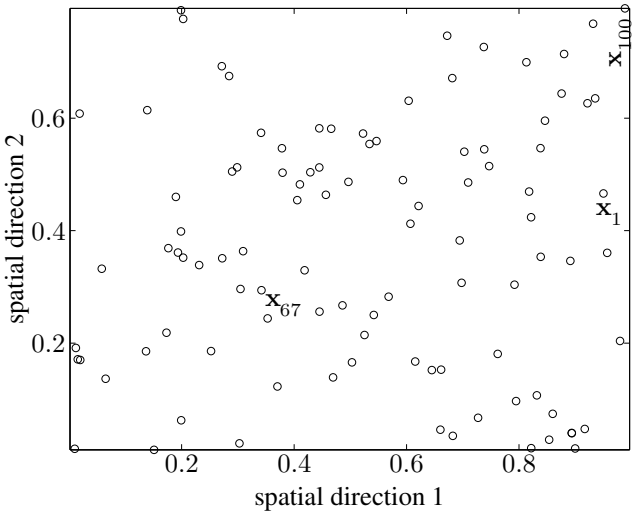


Figure 2.1: Measurement locations of $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ over spatial domain $\mathbf{D}$.

the following scheme is considered. Given the initial seed, the 100 horizontal coordinates are generated from $U(0, 1.0)$, in which $U(a, b)$ is the PDF of a uniform random variable supported on (a, b), by using the MATLAB's random number generator. Subsequently, the 100 vertical coordinates are generated from $U(0, 0.8)$. Given the initial seed, the coordinates of these locations represent a set of deterministic coordinates spread over the spatial domain $\mathbf{D}$. Each element of $\boldsymbol{\mathcal{Y}}$ is a random variable representing $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta})$ at a specific location shown in Figure 2.1 and $\dim(\boldsymbol{\mathcal{Y}}) = N = 100$.

The statistical dependency of the components of $\boldsymbol{\mathcal{Y}}$ is imposed here by assigning the Spearman's rank correlation (SRC) coefficient. The SRC function of the underlying random process is assumed to be isotropic and of the following form [Bar98],

$$R(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[- \sum_{k=1}^{\dim(\mathbf{D})} \gamma_k \mid x_{k,i} - x_{k,j} \mid^{\kappa} \right] \tag{2.37}$$

in which $R(\mathbf{x}_i, \mathbf{x}_j)$ is the SRC between two random variables associated with the locations having coordinates $\mathbf{x}_i$ and $\mathbf{x}_j$, γ_k, $k = 1, \cdots, \dim(\mathbf{D})$, is the inverse of the correlation length along the spatial direction k, $\mid x_{k,i} - x_{k,j} \mid$ is the absolute value of $(x_{k,i} - x_{k,j})$ in which $x_{k,i}$ and $x_{k,j}$ are, respectively, the k-th coordinates of $\mathbf{x}_i$ and $\mathbf{x}_j$ and κ is a constant. In the present case, $\dim(\mathbf{D}) = 2$ and it is assumed that $\gamma_k = 0.5$, $k = 1, 2$, and $\kappa = 2$. The marginal PDF of $\mathbf{y}_q$, $q = 1, \cdots, N$, is assumed to be lognormal with its pdf being given by,

$$f_{\mathbf{y}_q}(y) = \begin{cases} \dfrac{1}{y} g_{\mu_q, \sigma_q}(\ln y) & \text{if } y > 0, \\ 0 & \text{if } y \leq 0, \end{cases}$$

in which $g_{\mu_q,\sigma_q}(x) = [1/(\sqrt{2\pi}\sigma_q)] \exp[(x - \mu_q)/(2\sigma_q^2)]$ is the Gaussian density and μ_q and σ_q^2 are, respectively, mean and variance of the associated Gaussian random variable and given by,

$$\begin{aligned} \mu_q &= \ln \mu_{\mathbf{y}_q} - \frac{1}{2} \ln \left(\frac{\sigma_{\mathbf{y}_q}^2}{\mu_{\mathbf{y}_q}^2} + 1 \right), \\ \sigma_q^2 &= \ln \left(\frac{\sigma_{\mathbf{y}_q}^2}{\mu_{\mathbf{y}_q}^2} + 1 \right), \end{aligned} \tag{2.38}$$

in which $\mu_{\mathbf{y}_q}$ and $\sigma_{\mathbf{y}_q}^2$ are, respectively, mean and variance of $\mathbf{y}_q$.

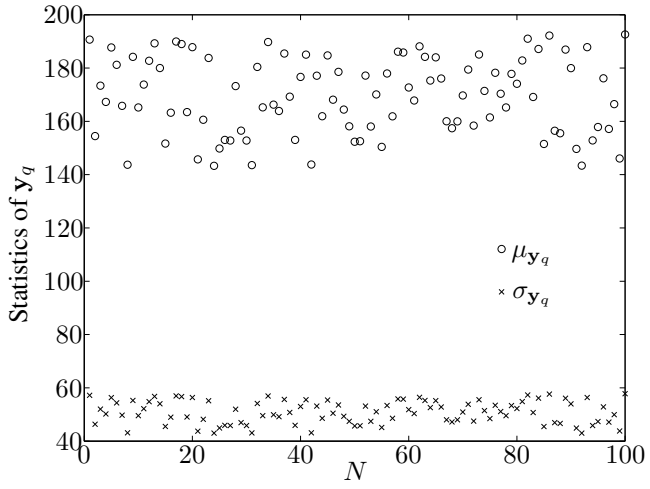


Figure 2.2: Statistics of $\mathbf{y}_q$, $q = 1, \cdots, N$.

To ensure the non-stationary characteristic of the random process, different values of mean are assigned to each $\mathbf{y}_q$. Like the coordinates of the 100 locations are generated given the initial seed, the 100

mean values of $\mathbf{y}_q$ are also generated from $U(0.85D, 1.15D)$ in which D is assumed to be 168. Coefficient of variation is assumed to be 0.3 for all $\mathbf{y}_q$. The values of $\mu_{\mathbf{y}_q}$ and $\sigma_{\mathbf{y}_q}$ thus selected are depicted in Figure 2.2. Denote the $N \times 1$ column vector consisting of $\{\mu_{\mathbf{y}_q}\}_{q=1}^{N}$ by $\boldsymbol{\mu}_y = [\mu_{\mathbf{y}_1}, \cdots, \mu_{\mathbf{y}_N}]^T$ and similarly the column vector of $\{\sigma_{\mathbf{y}_q}\}_{q=1}^{N}$ by $\boldsymbol{\sigma}_y = [\sigma_{\mathbf{y}_1}, \cdots, \sigma_{\mathbf{y}_N}]^T$.

The $N \times N$ SRC matrix, R_{yy}, of $\boldsymbol{\mathcal{Y}}$ is computed by using (2.37). The (i, j)-th element, $\rho_s(\mathbf{y}_i, \mathbf{y}_j)$, of R_{yy} is calculated by substituting $\mathbf{x}_i$ and $\mathbf{x}_j$ coordinates in (2.37). The SRC matrix is same [KC06, Section 3.2.2] for both the non-Gaussian vector, $\boldsymbol{\mathcal{Y}}$, and its underlying correlated Gaussian vector whose elements follow the standard normal distribution, $N(0, 1)$. The SRC matrix, R_{yy}, is then transformed to the Pearson's correlation coefficient (PCC) matrix (the 'usual' correlation coefficient matrix) of the underlying correlated Gaussian vector by using the following relation proposed by Pearson in 1904 [KC06, p. 51 and p. 75-77],

$$\rho_{ij} = 2 \sin\left(\frac{\pi}{6}\rho_s(\mathbf{y}_i, \mathbf{y}_j)\right), \quad i, j = 1, \cdots, N, \tag{2.39}$$

in which ρ_{ij} is the (i, j)-th element of the PCC matrix. This PCC matrix, $[\rho_{ij}]$, however, does not match with the PCC matrix of $\boldsymbol{\mathcal{Y}}$. Based on $[\rho_{ij}]$, the underlying correlated Gaussian vector is simulated by using its KL expansion. In this example, the number of terms retained in the KL expansion is 3 because 99% of the variance of the underlying correlated Gaussian vector is contributed by the 3 dominant KL random variables associated with the largest 3 eigenvalues of $[\rho_{ij}]$. Subsequently, the realizations of the underlying correlated Gaussian vector thus generated by using its KL expansion are shifted and scaled to enforce the required mean and variance vector whose elements are given by (2.38). The realizations of this Gaussian vector are finally transformed into the realizations of $\boldsymbol{\mathcal{Y}}$ by using the target marginal PDF of $\mathbf{y}_q$ by having recourse to the inverse-transform method, $\boldsymbol{\mathcal{Y}} = \exp[\boldsymbol{\mathcal{X}}]$, in which $\boldsymbol{\mathcal{X}}$ is the $N \times 1$ correlated Gaussian vector whose mean vector, variance vector and correlation matrix are, respectively, given by $[\mu_1, \cdots, \mu_N]^T$, $[\sigma_1^2, \cdots, \sigma_N^2]^T$ and $[\rho_{ij}]$.

In this numerical illustration, an additive Gaussian noise is also applied to the realizations of pure $\boldsymbol{\mathcal{Y}}$, say $\boldsymbol{\mathcal{Y}}_{(\text{pure})}$, to obtain noisy (as usually the case in practice) realizations of $\boldsymbol{\mathcal{Y}}$ as $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Y}}_{(\text{pure})} + N(\mathbf{0}, \text{diag}(0.04\boldsymbol{\sigma}_y))$, in which $N(\cdot)$ represents an N-D Gaussian vector, $\mathbf{0}$ is a $N \times 1$ column vector of zeros representing the mean vector and $\text{diag}(0.04\boldsymbol{\sigma}_y)$ is a diagonal matrix representing the covariance matrix whose i-th diagonal entry is $0.04\sigma_{\mathbf{y}_i}$, $i = 1, \cdots, N$. These noisy realizations of $\boldsymbol{\mathcal{Y}}$ would be treated further as the experimental measurements.

Relative difference in percentage (%) for		
Mean vector	Standard deviation vector	Spearman's rank correlation matrix
0.0272	0.9554	0.6545

Table 2.1: Comparison of sample statistics of noisy measurements of $\boldsymbol{\mathcal{Y}}$: Relative difference of each statistic is computed as $100\left(\|\mathbf{S}^{(\text{meas})}-\mathbf{S}\|_F\right)/\|\mathbf{S}\|_F$ in which $\mathbf{S}^{(\text{meas})}$ represents the sample statistic, $\mathbf{S}$ represents the appropriate population parameter ($\boldsymbol{\mu}_y$ or $\boldsymbol{\sigma}_y$ or R_{yy}) and $\|\cdot\|_F$ is Frobenius (matrix) norm defined by $\|\mathbf{S}\|_F=(\sum_{ij}\,|\,s_{ij}\,|^2)^{1/2}$, in which s_{ij} is the (i,j)-th element of $\mathbf{S}$.

A total of $n=1500$ noisy realizations of $\boldsymbol{\mathcal{Y}}$ are simulated. To ensure that the statistical characteristics of the simulated data are within acceptable tolerance, the observed values of the sample statistics based on these noisy data are compared to the given (exact) population parameters in Table 2.1 showing excellent match between the observed values of the statistics obtained from the digitally simulated measurements of noisy $\boldsymbol{\mathcal{Y}}$ and the given respective population parameters of $\boldsymbol{\mathcal{Y}}$.

It is noted that had actual measurements of $\boldsymbol{\mathcal{Y}}$ been available, digital simulation of a set of measurements of $\boldsymbol{\mathcal{Y}}$, as described above, would not have been necessary.

The rest of the numerical example is presented by following a linear pattern of the data processing procedure that an user would perhaps be required to follow to employ the strategy as proposed here.

2.4.2 Construction and MaxEnt Density Estimation of nKL Vector

Given the $n=1500$ noisy measurements of $\boldsymbol{\mathcal{Y}}$, the sample covariance matrix, C_{yy}, of $\boldsymbol{\mathcal{Y}}$ is evaluated first. Here, the KL vector, $\boldsymbol{\mathcal{Z}}'$, is determined such that $\sum_{i=1}^{M}\varsigma_i=0.99\sum_{i=1}^{N}\text{var}(\mathbf{y}_i)$. This choice of accuracy level dictates that 3 dominant KL random variables should be considered to construct the reduced order representation, $\boldsymbol{\mathcal{Z}}'$, of $\boldsymbol{\mathcal{Y}}$ implying that $\dim(\boldsymbol{\mathcal{Z}}')=M=3$.

Applying (2.2) on the noisy measurements of $\boldsymbol{\mathcal{Y}}$, the realizations of $\boldsymbol{\mathcal{Z}}'$ are obtained. To ensure that the information contained in the measurements of $\boldsymbol{\mathcal{Y}}$ are not lost as the dimension is reduced from $N=100$ to $M=3$, the realizations of $\boldsymbol{\mathcal{Y}}$ are reconstructed again from the realizations of $\boldsymbol{\mathcal{Z}}'$ by using the inverse transformation of (2.2), namely, $Y^{(\text{recons})}\approx VZ'+\overline{\mathcal{Y}}$, in which $Y^{(\text{recons})}$ is $N\times n$ matrix containing the horizontal stack of the reconstructed realizations of $\boldsymbol{\mathcal{Y}}$. The statistics evaluated from $Y^{(\text{recons})}$ are compared to the known population parameters in Table 2.2 showing that the sufficient information is propagated to the realizations of $\boldsymbol{\mathcal{Z}}'$.

Relative difference in percentage (%) for		
Standard deviation vector	Spearman's rank correlation matrix	Covariance matrix
1.0389	0.6345	0.0386

Table 2.2: Comparison of sample statistics of realizations contained in matrix $Y^{(\text{recons})}$: Relative difference of each statistic is computed as $100\left(\|\mathbf{S}^{(\text{recons})}-\mathbf{S}\|_F\right)/\|\mathbf{S}\|_F$ in which $\mathbf{S}^{(\text{recons})}$ represents the sample statistic and $\mathbf{S}$ represents the appropriate population parameter ($\boldsymbol{\sigma}_y$ or R_{yy} or C_{yy}).

Next, the realizations of $\boldsymbol{\mathcal{Z}}$ are obtained from the realizations of $\boldsymbol{\mathcal{Z}}'$ by using (2.3). Using these realizations, the sample joint moments, $\widehat{\beta}_j$, $j = 1,\cdots,p$, of $\boldsymbol{\mathcal{Z}}$, that are characterized by m_{ij}'s, are estimated by employing (2.19) to use in the MaxEnt constraints. Consider a set, $\mathbf{m} = \{\{m_{11},\cdots,m_{M1}\},\cdots,\{m_{1p},\cdots,m_{Mp}\}\}$. Here, $\mathbf{m}$ is assumed to be $\{\{1,0,0\},\{0,1,0\},\{0,0,1\},\{2,0,0\},\{0,2,0\},\{0,0,2\},\{1,1,0\},\{1,0,1\},\{0,1,1\},\{3,0,0\},\{0,3,0\},\{0,0,3\},\{2,1,0\},\{2,0,1\},\{1,2,0\},\{0,2,1\},\{1,0,2\},\{0,1,2\},\{1,1,1\},\{4,0,0\},\{0,4,0\},\{0,0,4\}\}$. The number of constraints defined by $\mathbf{m}$ is same as p implying $p = 22$. Denote the $p\times 1$ column vector consisting of elements, $\widehat{\beta}_j$, $j = 1,\cdots,p$, by $\widehat{\boldsymbol{\beta}}_n$. The MEDE technique as described in section 2.3.3 yields estimates of $\boldsymbol{\lambda}$. Two such estimates — one obtained without sequential updating method and the other with sequential updating method — are computed. A few representative elements of these estimates are reported in Table 2.3. While evaluating the former estimate, a $p\times 1$ column vector of zeros is chosen as an initial guess to start the Levenberg-Marquardt algorithm. Let $\boldsymbol{\beta}(\widehat{\boldsymbol{\lambda}}_n)$ be the vector of joint-moments having elements that are computed by setting $\boldsymbol{\lambda}$ to $\widehat{\boldsymbol{\lambda}}_n$ in (2.28) and using the resulting $p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z})$ in (2.19). Also reported in the table are $\xi(\widehat{\boldsymbol{\lambda}}_n)$, the relative difference of $\boldsymbol{\beta}(\widehat{\boldsymbol{\lambda}}_n)$ w.r.t. $\widehat{\boldsymbol{\beta}}_n$ and the value of entropy computed by using (2.18). Since the two estimates shown in Table 2.3 are drastically different from each other, it can be inferred that $\ln\ell(\cdot\mid\mathbf{Z}_n)$ is very flat in the neighborhood of $\widehat{\boldsymbol{\lambda}}_n\in\mathbb{R}^p$. The distance between the two associated pdf measured w.r.t. the symmetric cross-entropy measure [KK92, Section. 5.1.3] is found to be 0.0158. However, the estimates are almost equivalent in terms of the entropy, $H(p_{\boldsymbol{\mathcal{Z}}})$, and the relative difference of $\boldsymbol{\beta}(\widehat{\boldsymbol{\lambda}}_n)$ w.r.t. $\widehat{\boldsymbol{\beta}}_n$. The estimate, that is obtained without using the sequential updating method, is considered next for its relatively better numerical resolution.

Quantity	Without sequential updating method	With sequential updating method
$\widehat{\lambda}_1$	-74.9164	-85.4308
$\vdots$	$\vdots$	$\vdots$
$\widehat{\lambda}_4$	107.1809	107.7935
$\vdots$	$\vdots$	$\vdots$
$\widehat{\lambda}_{22}$	-41.7542	0
$\xi(\widehat{\boldsymbol{\lambda}}_n)$	29.1208	41.0928
$100\frac{\|\boldsymbol{\beta}(\widehat{\boldsymbol{\lambda}}_n)-\widehat{\boldsymbol{\beta}}_n\|}{\|\widehat{\boldsymbol{\beta}}_n\|}$	0.0054	0.0057
$H(p_{\boldsymbol{\mathcal{Z}}})$	-1.9721	-1.9725

Table 2.3: $\widehat{\boldsymbol{\lambda}}_n$, $\xi(\widehat{\boldsymbol{\lambda}}_n)$, relative difference of joint moment vector and $H(p_{\boldsymbol{\mathcal{Z}}})$.

2.4.3 Simulation of the nKL vector and Estimation of the Fisher Information Matrix

Having estimated $\widehat{\boldsymbol{\lambda}}_n$ by using MEDE technique, the next step is to estimate $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$. Here, $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ is computed by using a numerical integration technique as well as estimated by using a sampling technique. For the sampling-based estimate, the independent samples of $\boldsymbol{\mathcal{Z}}$ are generated by using M-H MCMC algorithm [Spa03, Section 16.2]. In this algorithm, given the k-th state, $\mathcal{Z}_k$, of $\boldsymbol{\mathcal{Z}}$, the candidate point, $\boldsymbol{\mathcal{W}}$, is generated according to a given proposal PDF. The following proposal PDF is considered for $\mathbf{q}(\cdot \,|\, \mathcal{Z}_k)$ for the present example,

$$\mathbf{q}(\boldsymbol{\mathcal{W}} \,|\, \mathcal{Z}_k) \sim U_M\left(\boldsymbol{\mathcal{W}} - \delta \mathbf{1}_M, \boldsymbol{\mathcal{W}} + \delta \mathbf{1}_M\right), \tag{2.40}$$

in which $U_M(\mathbf{a}, \mathbf{b})$ is a M-fold uniform PDF in which $\mathbf{a}$ and $\mathbf{b}$ are M-D vectors whose elements, respectively, represent the lower and upper bounds of the respective one-dimensional uniform random variable, δ is a positive constant and $\mathbf{1}_M$ is an M-D vector of 1s. Here, δ is assumed to be 0.3.

It must be noted that because $\boldsymbol{\mathcal{Z}}$ has a finite support $\boldsymbol{\Xi} = [0\ 1]^M$, the support, $\mathrm{supp}\,(\boldsymbol{\mathcal{W}} \,|\, \mathcal{Z}_k)$, of $\boldsymbol{\mathcal{W}} \,|\, \mathcal{Z}_k$ (and therefore, also the height of the proposal pdf in (2.40) to enforce the volume of the PDF to be unity) must be continuously changing during the runs of MCMC to guarantee the generation of $\boldsymbol{\mathcal{W}}$ within the $\mathrm{supp}\,(\boldsymbol{\mathcal{W}} \,|\, \mathcal{Z}_k) \subset \boldsymbol{\Xi}$. The conditional PDF, $\mathbf{q}(\cdot \,|\, \mathcal{W})$, of $\boldsymbol{\mathcal{Z}}_k$ is also required for M-H MCMC algorithm. Here, $\boldsymbol{\mathcal{Z}}_k$ is the possible k-th state given the value, $\mathcal{W}$, of $\boldsymbol{\mathcal{W}}$. The expression of $\mathbf{q}(\boldsymbol{\mathcal{Z}}_k \,|\, \mathcal{W})$ is analogous to the expression in (2.40). The support and height for $\mathbf{q}(\boldsymbol{\mathcal{Z}}_k \,|\, \mathcal{W})$ also need to be changed

during the runs of MCMC to guarantee that $\mathrm{supp}\,(\boldsymbol{\mathcal{Z}}_k\,|\,\mathcal{W}) \subset \boldsymbol{\Xi}$. It essentially implies that Eq. (16.3) on p. 441 of ref. [Spa03] cannot be reduced to the simplified version as shown on p. 442 of ref. [Spa03].

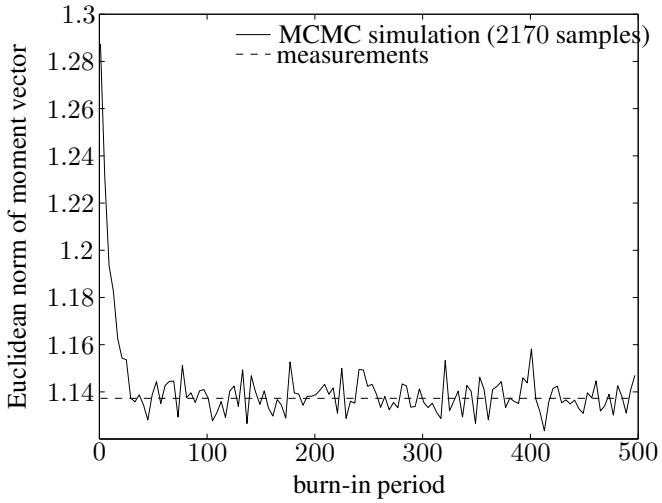


Figure 2.3: Euclidean norm, $\|\boldsymbol{\beta}^{(\mathrm{MCMC})}\|$, of $\boldsymbol{\beta}^{(\mathrm{MCMC})}$, representing the vector of sample joint-moments estimated by using 2170 independent MCMC samples and shown as solid line, is compared to $\|\widehat{\boldsymbol{\beta}}_n\|$ shown as dashed line.

In M-H MCMC, a burn-in-period of 300 is considered for the present example. The 301^{st} sample resulting from one run of MCMC yields one sample. A total of 2170 such independent runs of MCMC yields 2170 independent samples of $\boldsymbol{\mathcal{Z}}$. All the 22 sample joint-moments of $\boldsymbol{\mathcal{Z}}$, estimated based on these 2170 independent samples, are found to converge to the stationary values around the respective components of $\widehat{\boldsymbol{\beta}}_n$ justifying that the burn-in-period of 300 is sufficient enough (see Figure 2.3). A total of 950000 independent runs of MCMC is carried out to generate $N_s = 950000$ independent data vectors as a proxy for $\boldsymbol{\mathcal{Z}}$. The mean, standard deviation, maximum and minimum values of the acceptance rate over 950000 MCMC runs are found to be, respectively, 30.9217%, 4.0106%, 51.1628% and 1.6611%.

As indicated earlier, because of the use of MEDE technique, the 22×22 FIM here has an interesting feature in the sense that it can be divided into known and unknown parts as shown in Figure 2.4. The unknown elements of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ is estimated by using the independent MCMC samples as simulated above. The unknown elements of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ is also evaluated by a direct numerical integration scheme. Denote the later FIM by $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)^{(\mathrm{anal})}$ and the sampling-based estimate by $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)^{(\mathrm{MCMC})}$. The relative difference of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)^{(\mathrm{MCMC})}$ w.r.t. $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)^{(\mathrm{anal})}$ measured in terms of Frobenius norm is found to be 1.0941%.

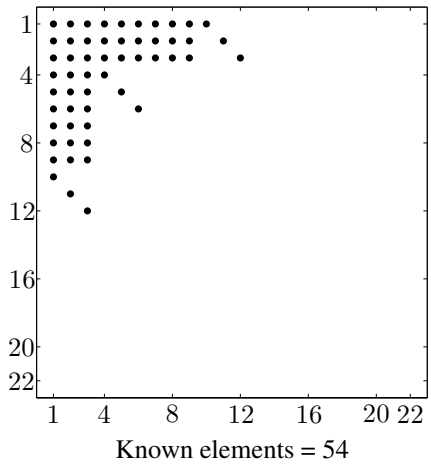


Figure 2.4: Fisher information matrix with known elements as marked; void part consists of unknown elements.

2.4.4 Estimation of PC coefficients of $\mathcal{Z}$ and $\mathcal{Y}$

To compute the estimators, $\{\widehat{z}_{i,k}(\widehat{\boldsymbol{\lambda}}_n)\}_{i=0}^{P}$, defined by (2.15), a set of statistically independent standard normal random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{n_d}$, is used. The basic polynomials, $\boldsymbol{\Psi}_j(\boldsymbol{\xi}_i)$, then turn out to be Hermite polynomials given by,

$$\begin{aligned} \boldsymbol{\Psi}_0(\boldsymbol{\xi}_i) &= 1, \quad \boldsymbol{\Psi}_1(\boldsymbol{\xi}_i) = \boldsymbol{\xi}_i, \\ \boldsymbol{\Psi}_j(\boldsymbol{\xi}_i) &= \boldsymbol{\xi}_i \boldsymbol{\Psi}_{j-1}(\boldsymbol{\xi}_i) - (j-1)\boldsymbol{\Psi}_{j-2}(\boldsymbol{\xi}_i), \quad \text{if } j \geq 2, \end{aligned} \tag{2.41}$$

and the variance of $\boldsymbol{\Psi}_{\alpha_i}(\boldsymbol{\xi}_i)$ in (2.9) is given by [SG04a],

$$E\left[\boldsymbol{\Psi}_{\alpha_i}^2(\boldsymbol{\xi}_i)\right] = \alpha_i!, \quad i = 1, \cdots, n_d.$$

The order, n_o, of the PC representation is considered to be 2 and the dimension, n_d, as already argued, is fixed to the value of $M = 3$. For computations described in this subsection, only the first $K = 2170$ samples of 950000 MCMC samples are considered. The mean, standard deviation, maximum and minimum values of the acceptance rate over these 2170 MCMC runs are found to be, respectively, 31.1659%, 4.0184%, 46.1794% and 15.9468%. The sample joint-moment vector, $\boldsymbol{\beta}^{(\mathrm{MCMC})}$, of $\mathcal{Z}$ based on these 2170 MCMC samples is compared to $\widehat{\boldsymbol{\beta}}_n$. The relative difference of $\boldsymbol{\beta}^{(\mathrm{MCMC})}$ w.r.t. $\widehat{\boldsymbol{\beta}}_n$ measured in

terms of Frobenius norm is found to be 0.9098% implying that $K = 2170$ samples are sufficient enough for this part of the example.

The number of terms to be included in a second order and third dimensional PC representation is $(P+1) = (2+3)!/(2!3!) = 10$. A total of 2170 realizations of $\underline{\boldsymbol{\xi}}$ is obtained by employing Rosenblatt transformation on 2170 independent MCMC realizations of $\boldsymbol{\mathcal{Z}}$. The realizations of $\underline{\boldsymbol{\xi}}$ is subsequently substituted in the expressions of $\boldsymbol{\Psi}_j(\boldsymbol{\xi}_i)$ in (2.41) to obtain the realizations of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})$ which is given by (2.7). These realizations of $\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})$ along with the respective realizations of $\boldsymbol{\mathcal{Z}}$ are used in (2.15) to compute $\widehat{z}_{i,k}(\widehat{\boldsymbol{\lambda}}_n)$, $i = 0 \cdots, P$ and $k = 1, \cdots, M$. Subsequently, a new set of 2170 realizations of $\boldsymbol{\Upsilon}_i(\underline{\boldsymbol{\xi}})$, that is statistically independent of the earlier realizations, is generated and substituted in the PC representation of $\mathbf{z}_k$, $\mathbf{z}_k \stackrel{d}{=} \sum_{i=0}^{P} \widehat{z}_{i,k}(\widehat{\boldsymbol{\lambda}}_n)\, \boldsymbol{\Upsilon}_i(\underline{\boldsymbol{\xi}})$, to generate a set of 2170 PC realizations of $\boldsymbol{\mathcal{Z}}$. The marpdf of each $\mathbf{z}_k$ is estimated by employing KDE technique based on these 2170 PC realizations. A plot is shown in Figure 2.5 for a typical value, $k = 3$. In this plot, also superimposed are the marpdf of $\mathbf{z}_k$ estimated by employing KDE technique based on earlier 2170 MCMC samples and 1500 measurement data of $\boldsymbol{\mathcal{Z}}$ along with the plot of analytical marpdf of $\mathbf{z}_k$ evaluated from MaxEPD($\mathcal{Z}, \widehat{\boldsymbol{\lambda}}_n$). Though only one plot is reported here, excellent matches are also found for $\mathbf{z}_k$, $k = 1, 2$. The relative difference of the sample joint-moment vector, $\boldsymbol{\beta}^{(\mathrm{PC})}$ (estimated by using 2170 PC realizations), w.r.t. $\widehat{\boldsymbol{\beta}}_n$ measured in terms of Frobenius norm is found to be 1.0695% showing that the joint statistical characteristics of $\boldsymbol{\mathcal{Z}}$ are also reproducible with sufficient accuracy within the framework of PC representation.

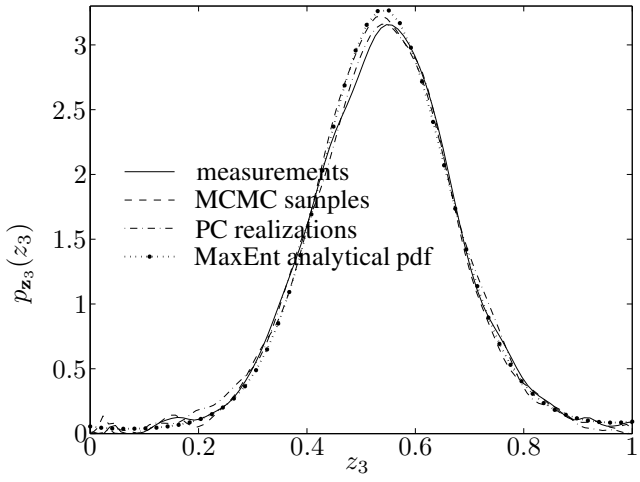


Figure 2.5: Marginal probability density function of $\mathbf{z}_3$, $p_{\mathbf{z}_3}(z_3)$.

Scheme	Relative difference in percentage (%) for		
	Standard deviation vector	Spearman's rank correlation matrix	Covariance matrix
MCMC	1.9632	0.8309	2.1962
PC	1.4831	0.8319	2.3407

Table 2.4: Comparison of sample statistics of MCMC and PC realizations of $\boldsymbol{\mathcal{Y}}$: Relative difference of each statistic is computed as $100\left(\|\mathbf{S}^{(\mathrm{schm})}-\mathbf{S}\|_F\right)/\|\mathbf{S}\|_F$ in which $\mathbf{S}^{(\mathrm{schm})}$ represents the sample statistic based on realizations obtained by using scheme - MCMC or PC.

Next, the MCMC samples and PC realizations of $\boldsymbol{\mathcal{Z}}$ are used, respectively, to generate the MCMC realizations and PC realizations of $\boldsymbol{\mathcal{Y}}$ by using (2.4). The statistics of $\boldsymbol{\mathcal{Y}}$ are compared to the given (exact) statistics in Table 2.4. Clearly, the sample statistics match well with the given statistics. The first element of $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ represents the mean of $\mathbf{y}_q$, $q=1,\cdots,N$. If these elements are collectively shown as a column vector, $\underline{\mathcal{Y}}=[\widehat{y}_0(\mathbf{x}_1,\widehat{\boldsymbol{\lambda}}_n),\cdots,\widehat{y}_0(\mathbf{x}_N,\widehat{\boldsymbol{\lambda}}_n)]^T$, then the relative difference of $\underline{\mathcal{Y}}$ w.r.t. $\boldsymbol{\mu}_y$ measured in terms of Frobenius norm is found to be 0.6393% showing the effect of finite K on $\widehat{z}_{0,k}$, $k=1,\cdots,M$.

2.4.5 Determination of Asymptotic Probability Distribution Function of $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$

The final task is to determine the apdf of $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ in (2.17). The gradient matrix, $\mathbf{h}'_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$, is approximated based on (2.36) with $c=0.0001$. The estimated approximate apdf, $(\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)-\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)) \overset{\text{approx.}}{\sim} N(\mathbf{0},\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})^T\,\mathbf{F}_n(\boldsymbol{\lambda})^{-1}\,\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda}))$, with the covariance matrix being evaluated at $\widehat{\boldsymbol{\lambda}}_n$ and $\mathbf{0}$ being a $(P+1)\times 1$ column vector of zeros, could be used to determine an uncertainty bound for $\|\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)-\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)\|$. However, it must be noted that for a general nonlinear problem, as considered in the current work, there is no known finite sample ($n<\infty$) distribution for $\widehat{\boldsymbol{\lambda}}_n$, and therefore, for its deterministic mapping $\mathbf{h}_{\mathbf{x}_q}(\cdot)$. The above apdf is only valid as the number of measurements, n, becomes reasonably large.

Exclusively for the last part of this numerical example, another set of noisy realizations of $\boldsymbol{\mathcal{Y}}$ is simulated with $n=100000$ and the sample joint-moments of $\boldsymbol{\mathcal{Z}}$ are computed. It is found that the relative difference of $\widehat{\boldsymbol{\beta}}_n$ at $n=1500$ w.r.t. $\widehat{\boldsymbol{\beta}}_n$ at $n=100000$, measured in terms of Frobenius norm, is 37.9973%. Clearly, $n=1500$ is not a reliably large value for using it in the computation of approximate apdf of $(\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)-\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*))$. However, it does not mean that MaxEPD($\mathcal{Z},\widehat{\boldsymbol{\lambda}}_n$) of $\boldsymbol{\mathcal{Z}}$ based on finite number ($n=1500$) of measurement data is not correct. Given $p=22$ numbers of sample joint-moments estimated by using $n=1500$ measurement data, the estimated MaxEPD($\mathcal{Z},\widehat{\boldsymbol{\lambda}}_n$) of $\boldsymbol{\mathcal{Z}}$ is least committed

to the information not given to us. As more data arrive over time, the estimate, $\widehat{\boldsymbol{\lambda}}_n$, changes. One of the goals of the present work is to determine the confidence level of $\|\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n) - \mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)\|$ as n becomes perceivably large.

For the purpose of illustration, however, suppose that the estimated $\widehat{\boldsymbol{\beta}}_n$ at $n = 1500$ would not significantly change from $\widehat{\boldsymbol{\beta}}_n$ at $n = 1 \times 10^{10}$ implying that $\widehat{\boldsymbol{\lambda}}_n$ at $n = 1 \times 10^{10}$ is almost equal to $\widehat{\boldsymbol{\lambda}}_n$ at $n = 1500$. The FIM, $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$, estimated earlier for $n = 1500$ can be scaled up by a factor of $(1 \times 10^{10}/1500)$ to approximately compute $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ at $n = 1 \times 10^{10}$. Hence, for $n = 1 \times 10^{10}$, using the new covariance matrix, $\mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})^T\, \mathbf{F}_n(\boldsymbol{\lambda})^{-1}\, \mathbf{h}'_{\mathbf{x}_q}(\boldsymbol{\lambda})$, evaluated at $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_{1500}$ with $\mathbf{F}_n(\boldsymbol{\lambda})$ being the scaled-up $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)^{(\mathrm{anal})}$ evaluated at $n = 1500$ as just mentioned, the 95% percentile confidence level of $\|\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n) - \mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)\|$ is obtained as 3.4887 for a typical value, $q = 1$. This confidence level is computed by simulation using 1×10^6 realizations. It must be again emphasized that the condition $\widehat{\boldsymbol{\lambda}}_{1500} \approx \widehat{\boldsymbol{\lambda}}_{1\times 10^{10}}$ (which also implies that $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_{1500}) \approx \mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_{1\times 10^{10}}) \approx \mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*) \approx \widetilde{\mathbf{h}}_{\mathbf{x}_q}$) is assumed to be valid simply for the sake of illustration and only for the last part of this numerical example when the confidence level of the error term, $\|\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n) - \mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*)\|$, is determined using its approximate apdf.

2.5 Conclusions

The work presented in this chapter investigates the effects of data uncertainty on the confidence interval of estimators of the PC coefficients of a random vector, $\boldsymbol{\mathcal{Y}}$, that is a finite-dimensional representation of a non-Gaussian, non-stationary and second-order stochastic process. The KL decomposition and a scaling transformation are employed, on a set of data measured on the random vector, to perform stochastic model reduction. The MaxEnt mjpdf of the resulting reduced random vector (normalized KL vector, $\boldsymbol{\mathcal{Z}}$) is subsequently estimated. Given the sample joint-moments estimated from the observations of $\boldsymbol{\mathcal{Z}}$, the estimated mjpdf is unique and most unbiased (any deviation from this probability density function implies a bias towards some unavailable information). The estimator, $\widehat{\boldsymbol{\lambda}}_n$, of $\boldsymbol{\lambda}$, that characterizes the MaxEnt mjpdf, is computed by employing a nonlinear least-squares technique (Levenberg-Marquardt optimization algorithm). By using the estimated mjpdf and the Rosenblatt transformation, the vector, $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$, consisting of the estimators of the PC coefficients of $\boldsymbol{\mathcal{Y}}$, is evaluated in order to obtain the PC representation of $\boldsymbol{\mathcal{Y}}$. This PC representation approximates $\boldsymbol{\mathcal{Y}}$ by projecting it on a finite-dimensional space spanned by a set of orthogonal basis functions providing access to all the tools available in the

area of functional analysis. This is useful for many purpose, for instance, convergence analysis of the PC representation of $\mathcal{Y}$.

It is reported, in the context of the numerical example presented here, that both the estimated MaxEnt mjpdf and the PC representation of the random vector could represent the probabilistic and statistical characteristics of the measured data with excellent accuracy even for a finite number of measurements which is typically the case in most practical problems. The estimator, $\widehat{\boldsymbol{\lambda}}_n$, is also the MLE of $\boldsymbol{\lambda}$ implying that $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is also the MLE of $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*) \approx \widetilde{\mathbf{h}}_{\mathbf{x}_q}$ representing the vector whose elements are the 'true' PC coefficients of the finite-dimensional representation of the stochastic process.

It should be noted that like mean, variance and other higher order joint-moments of the random process, the PC coefficients are population parameters characterizing the random process. Clearly, the probabilistic and statistical characteristics of estimators of these population parameters depend on inherent randomness embodied in the available measurement data. It is reported here that $\mathbf{h}_{\mathbf{x}_q}(\widehat{\boldsymbol{\lambda}}_n)$ is also consistent and asymptotically efficient estimator of $\mathbf{h}_{\mathbf{x}_q}(\boldsymbol{\lambda}^*) \approx \widetilde{\mathbf{h}}_{\mathbf{x}_q}$. The associated asymptotic normal distribution, estimated based on a large number of measurement data, is useful to determine a confidence interval as to how much the estimates of the PC coefficients are likely to differ from the 'true' PC coefficients that are typically unknown.

The computation of the asymptotic normal distribution of the PC coefficients requires estimation of the FIM, $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$. In the context of the current work, the FIM is found to have an interesting structure where some of the elements of $\mathbf{F}_n(\widehat{\boldsymbol{\lambda}}_n)$ are already known and the other elements are unknown. Because of the use of MEDE technique resulting in a pdf from exponential family, the unknown elements can be efficiently estimated in the current work without affecting the known elements. However, this is not possible in other cases where all the special advantages afforded by the MEDE technique are not readily available. A recent work [Das07] addresses this general case focusing on how the prior information, available in terms of the known elements, can be exploited to compute the better estimates for the unknown elements.

Chapter 3

Polynomial Chaos Representation of Random Field from Experimental Measurements

Two numerical techniques are proposed to construct polynomial chaos (PC) representation of an arbitrary second-order random vector. In the first approach, PC representation is constructed by matching a target joint probability density function (pdf) based on the concept of conditional probability and the Rosenblatt transformation. In the second approach, the PC representation is obtained by having recourse to the Rosenblatt transformation and matching simultaneously a set of all target marginal pdfs and a target Spearman's rank correlation coefficient (SRCC) matrix. Both the techniques are applied to model a spatio-temporal, non-stationary and non-Gaussian random temperature field, that is assumed to be a second-order random field, by using a set of oceanographic data obtained from a shallow-water acoustics transmission experiment [ABC$^+$97]. The set of measurement data, observed over a finite denumerable subset of the indexing set of the random process, is treated as a set of observed samples of a second-order random vector that can be treated as a finite-dimensional approximation of the original random field. A complete set of properly ordered conditional pdfs, that uniquely characterizes the target joint pdf, in the first approach and a set of all the target marginal pdfs and the target SRCC matrix in the second approach are estimated by using available experimental data. Digital realizations sampled from the constructed PC representations based on both the schemes capture the observed and target statistical characteristics of the experimental data with sufficient accuracy. The relative advantages and disadvantages of these techniques are also highlighted.

3.1 Motivation and Problem Description

Unlike matching a finite set of joint higher order statistics in chapter 2, a target mjpdf or a set of target marginal probability density functions (marpdfs) along with a target correlation coefficient (corrcoef) function are captured here by the PC representation. The use of the maximum-entropy (MaxEnt) principle for the estimation of the target pdfs is avoided in this chapter. While this is beneficial from a computational perspective if the construction of PC representation is the only goal, the identification of the asymptotic probability density function (apdf) of estimators of the PC coefficients cannot be done here because of the absence of the nice theory related to the maximum likelihood estimator (MLE) within a convex optimization setup that was present in chapter 2 (see section 2.2.4 and section 2.3.2 for further details). Therefore, the work in this chapter focuses more closely on the construction of the probability model (i.e., PC representation) of a non-stationary and non-Gaussian random process by using experimental measurements, and the associated simulation technique based on the constructed model. A brief literature survey in the context of simulation and characterization of non-Gaussian and non-stationary random processes is presented below to show the complete justification of carrying out the work presented here.

The most popular approach in digitally generating the realizations of non-Gaussian process is through specifying a set of target non-Gaussian marpdfs and a target corrcoef function or spectral density function (sdf) [CN97, Gri98, DM01]. The set of target marpdfs and the target corrcoef function or sdf can be determined by fitting a conformable set of statistics estimated from the available set of data. In synthesizing the realizations through this course, it is assumed that an "underlying" Gaussian process exists and a search technique is subsequently employed to find an "equivalent" and feasible (positive-definite) corrcoef function of the Gaussian process. The realizations of the Gaussian process synthesized based on the equivalent corrcoef function are then transformed to the realizations of the requisite non-Gaussian random process. The later transformation is based on the mapping introduced earlier by Nataf in 1962 [HM00, Section 4.3]. At this stage, an additional computational budget needs to be allocated to find the inverse functions of the target marginal probability distribution functions (marPDFs) if they are not readily available (imagine a distribution function with multi-modal characteristics). This may often be the case in many practical applications when the target marpdfs are estimated from the available set of data by employing nonparametric density estimation techniques [Ize91, KPU04, DGS08], [Sco92, Chapter 6] as employed in chapter 2. In addition to this computational overhead, counterexamples exist in the literature

showing that a non-Gaussian random vector with a specific target corrcoef matrix can still exist in spite of *non-existence* of an underlying Gaussian random vector [GH02].

It should also be noted that there exist several other different notions of correlation in the field of statistical literatures [EMS01]. Besides the usual corrcoef, the alternative measures of statistical dependency, that researchers had recently recourse to for characterizing non-Gaussian random processes, include SRCC or Spearman's rho and Kendall's tau. The usual corrcoef, on the other hand, is known as linear or Pearson's correlation coefficient (PCC) in the honor of Karl Pearson who first highlighted its usefulness as a measure of statistical dependency. A recent simulation study [HLD04, Section 12.5.2] investigates the feasibility (positive-definiteness) of the PCC matrix of an underlying Gaussian random vector when the statistical dependency of the non-Gaussian random vector is characterized by corrcoef matrices based on SRCC and Kendall's tau. It is found in their study that an underlying Gaussian vector is more likely to exist, particularly, in a high dimensional setting, when the statistical dependency among the random variable components of the non-Gaussian vector is characterized by a SRCC matrix. This feature of SRCC has a significant practical advantage from a simulation point of view. The realizations of the Gaussian vector, that are easy to sample digitally, can then be transformed to the realizations of the non-Gaussian vector by using the Nataf transformation. Therefore, only the SRCC will be considered in the ensuing discussion.

The topic on simulation of the non-Gaussian random process by specifying a set of target non-Gaussian marpdfs and a target SRCC function has already been considered in the literature. In this case, if the underlying Gaussian process *exists*, then *no* special search technique is required to determine the feasible PCC function [CR99, GH03, PQH04], [HLD04, Section 12.5.2] facilitating computational savings to a certain extent. However, efficient simulation still requires easy computation of the inverse functions of the target marPDFs because of the use of Nataf transformation. Clearly, simulation techniques, based on a target PCC/SRCC function or sdf and an underlying Gaussian process, are not computationally efficient, particularly, in the case when the target set of marpdfs are estimated by employing nonparametric techniques.

Another work [MB93, MB97], that does not assume the existence of an underlying Gaussian process, presents an optimization technique based on Kullback-Leibler minimum cross-entropy principle for bivariate distribution. It results in a Taylor expansion based pdf. Though, the generalization of this method

is theoretically feasible in higher dimensional distribution, the actual development becomes prohibitively complicated because of the high dimensional Taylor expansion.

Recently, two new methods based on undirected graph (referred as tree and vine) have been introduced in the literature [KC06, Chapter 4]. The technique based on a tree constructed for an N-dimensional (D) random vector allows specification of only $(N-1)$ elements of the SRCC matrix out of $N(N-1)/2$ off-diagonal elements. The method based on a vine relaxes this limitation, and therefore, can be theoretically used to realize every SRCC matrix. However, the use of the later method requires knowledge of a copula [Joe97, Nel06] that specifies the structure of the statistical dependence among the constituent random variable components. Only a limited class of copula is investigated until now to integrate them into the formulation based on a vine. It is unlikely that any arbitrary target SRCC matrix can be realized via this method at its current state. Nevertheless, this method has a promising future and needs further research attention.

The characterization of non-Gaussian random process continues to be an evolving research field drawing the motivation from the practically appealing issues of estimating the underlying family of mjpdfs *from* finite data [Ize91, Sco92, GH02, KPU04]. By making use of such techniques, the problem of non-existence of an underlying Gaussian random process or the complicated Taylor expansion based pdf can be overcome at the cost of additional computational expenses. However, advanced simulation techniques, for example, algorithms based on Markov chain Monte Carlo (MCMC) need to be invoked to sample from the resulting family of mjpdfs thus requiring further computational budget (as a side note, MCMC simulation technique is also required to sample from the Taylor expansion based pdf). This difficulty could be a major bottleneck particularly in the context of propagating the statistical characteristics of stochastic system parameters to the model-based predictions if the stochastic system parameters need to be modeled as non-Gaussian random processes. A number of studies [vdG98, PPS02, SG02a, SG02b] have been carried out to circumvent this particular difficulty by representing the non-stationary and non-Gaussian random processes through PC expansion [GS91]. The underlying concept of these studies is similar to the one introduced earlier by Lancaster [Lan57], which again assumes the existence of an underlying Gaussian process.

The work in this chapter presents two different computational techniques to estimate the probability model of a finite-dimensional approximation, $\boldsymbol{\mathcal{Y}}$, of the underlying non-stationary and non-Gaussian

spatio-temporal stochastic process whose inherent randomness is assumed to be completely characterized by the experimental measurements taken simultaneously over space and time. The first approach constructs the PC representation based on a target mjpdf, and the other approach is based on a set of all the target marpdfs and a target SRCC matrix. The target mjpdf, marpdfs and SRCC matrix, respectively, correspond to the observed joint histogram density, observed marginal histogram densities and sample SRCC matrix estimated by using the available measurements. No assumption about the existence of an underlying Gaussian vector is made for any of the approaches presented here, nonetheless the second approach can exploit the advantage of existence of such a vector (if any).

The two approaches are presented in section 3.2.1 and section 3.2.2. Since considerable use of the properties of SRCC is made in the second approach, the definition and relevant features of SRCC are highlighted before presenting the second approach. As an illustration of the two proposed techniques, a set of oceanographic data obtained from a shallow-water acoustics transmission experiment [ABC$^+$97] is used to model the spatio-temporal random temperature field and the results are discussed in section 3.3. Finally, the conclusions inferred from the work is presented in section 3.4.

3.2 Construction of PC Representation from Data

From the review of the PC formalism in section 2.2.2, the PC representation of each component of $\boldsymbol{\mathcal{Y}}$ can be expressed as,

$$\mathbf{y}_k \equiv \mathbf{y}_k(\underline{\boldsymbol{\xi}}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^{n_d}} y_{\boldsymbol{\alpha},k}\, \boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}}), \quad k = 1, \cdots, N, \tag{3.1}$$

since the underlying stochastic process, and therefore, $\mathbf{y}_k(\underline{\boldsymbol{\xi}})$, is assumed to be second-order satisfying $E[|\,\mathbf{y}_k(\underline{\boldsymbol{\xi}})\,|^2] < \infty$. Here, the set of orthogonal basis functions, $\{\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathbb{N}^{n_d}\}$, is given by (2.6) or (2.7) as appropriate and the set of PC coefficients is computed from,

$$y_{\boldsymbol{\alpha},k} = \frac{E\left[\mathbf{y}_k(\underline{\boldsymbol{\xi}})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right]}{E\left[\boldsymbol{\Upsilon}^2_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right]},\ \boldsymbol{\alpha} \in \mathbb{N}^{n_d},\ k = 1, \cdots, N. \tag{3.2}$$

The PC representation thus determined, albeit, with due care devoted to the concerns on choice of the appropriate probability measure, $P_{\underline{\boldsymbol{\xi}}}$, and the "most suitable and significant" representation, can capture the essential statistical characteristics of the random quantity of interest. The "most suitable and significant" PC representation is to be inferred in some appropriate sense, for example, based on a convergence

analysis by using the theory of functional analysis and statistical tests available in the theory of statistical inference.

The denominator in (3.2) can be determined by using (2.6) or (2.7) (see (2.8) and (2.9 for further references and details) with the corresponding discussions), and the numerator needs to be computed by evaluating the following integral,

$$E\left[\mathbf{y}_k(\underline{\boldsymbol{\xi}})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\boldsymbol{\xi}})\right] = \int_{S_{\underline{\boldsymbol{\xi}}}} \mathbf{y}_k(\underline{\xi})\boldsymbol{\Upsilon}_{\boldsymbol{\alpha}}(\underline{\xi})\, p_{\underline{\boldsymbol{\xi}}}(\underline{\xi})\, d\underline{\xi}, \tag{3.3}$$

that requires the knowledge of the mapping, $\underline{\boldsymbol{\xi}} \longmapsto \mathbf{y}_k(\underline{\boldsymbol{\xi}})$. This mapping is again not available in the present work. Two schemes are presented next defining this mapping, and consequently, enabling the computation of the integral in (3.3). Thus, the PC coefficients in (3.2) are determined yielding the required PC representation in (3.1).

The preliminary idea of the first approach is similar in some sense to the ones presented earlier in chapter 2 (see also [KPU04, DGS08]). This approach is based on the Rosenblatt transformation that makes use of a complete set of properly ordered conditional PDFs, and can be considered as a supplement to the work presented in chapter 2. The set of conditional PDFs uniquely defines the target mjPDF. The second approach, on the other hand, is strongly founded on the properties of SRCC and the Rosenblatt transformation (applied individually on each marPDF of the involved scalar-variate random variable components). It borrows ideas from the literature of computer simulation of a non-Gaussian random vector when the non-Gaussian vector is characterized by a set of marpdfs and a SRCC matrix.

3.2.1 Approach 1: Based on Conditional PDFs

The unknown mapping, $\underline{\boldsymbol{\xi}} \longmapsto \boldsymbol{\mathcal{Y}}$, in this case is defined by using the Rosenblatt transformation. While any suitable density estimation technique could be applied to compute the target mjPDF, $P_{\boldsymbol{\mathcal{Y}}}$, of $\boldsymbol{\mathcal{Y}}$ by using the available measurement data, the target mjPDF, in the present work, is simply obtained from the normalized $(N+1)$-D histogram of the available N-variate data of $\boldsymbol{\mathcal{Y}}$. The normalized histogram can be used to determine the corresponding target mjpdf, $p_{\boldsymbol{\mathcal{Y}}}$. The histogram is first estimated over a discrete array of finite number of grid points spread over the support, $S_{\boldsymbol{\mathcal{Y}}} \subset \mathbb{R}^N$, of $\boldsymbol{\mathcal{Y}}$. This discrete array of grid points typically represents the center points of the histogram bins. An N-D linear interpolation scheme is subsequently employed to determine the value of the histogram of $\boldsymbol{\mathcal{Y}}$ at any other arbitrary point, $\mathcal{Y} \in S_{\boldsymbol{\mathcal{Y}}}$,

thus resulting in the target mjpdf, $p_{\boldsymbol{\mathcal{Y}}}$, and therefore, the target mjPDF, $P_{\boldsymbol{\mathcal{Y}}}$, over the entire $S_{\boldsymbol{\mathcal{Y}}}$. Use of the normalized histogram to approximate $p_{\boldsymbol{\mathcal{Y}}}$ is acceptable. The density estimation techniques currently existing in the literature are founded on this primitive notion of normalized histogram. It should also be noted that the final objective of the present work is not estimation of the mjpdf of $\boldsymbol{\mathcal{Y}}$ but construction of the PC representation of $\boldsymbol{\mathcal{Y}}$. The resulting $P_{\boldsymbol{\mathcal{Y}}}$ is an absolutely continuous function on $S_{\boldsymbol{\mathcal{Y}}}$ because of the use of linear interpolation scheme. A requirement for using the Rosenblatt transformation is absolute continuity of $P_{\boldsymbol{\mathcal{Y}}}$.

Let us illustrate the approach now by using a 2-D random vector, say, $\boldsymbol{\mathcal{Y}} = [\mathbf{y}_1, \mathbf{y}_2]^T$. The formulation can be readily extended to the random vector with more than two random variable components. Consider the 2-D data set as shown in Figure 3.1. The corresponding histogram is shown in Figure 3.2. The target

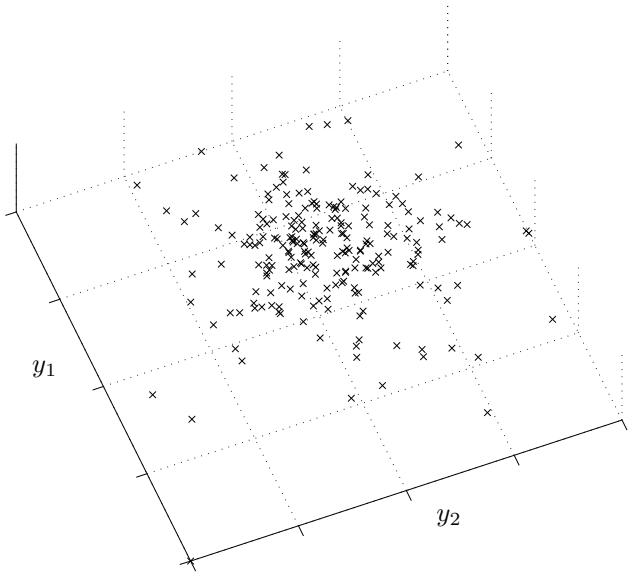


Figure 3.1: 2-D Illustration: data points.

mjpdf, $p_{\boldsymbol{\mathcal{Y}}}$, based on 2-D linear interpolation of the histogram is shown in Figure 3.3. The motive here is to pictorially describe the formulation; therefore, the specific values of the associated data or values of the resulting function and variables are not relevant.

Now, let $p_{1|2}$ be the conditional pdf of $\mathbf{y}_1$, given $\mathbf{y}_2 = y_2$, induced by $p_{\mathbf{y}_1\mathbf{y}_2}$ as shown in Figure 3.4 for different values of $y_2 \in s_{\mathbf{y}_2}$, in which $s_{\mathbf{y}_2} = [l_2,\, m_2] \subset \mathbb{R}$ is the support of $\mathbf{y}_2$. The slices representing

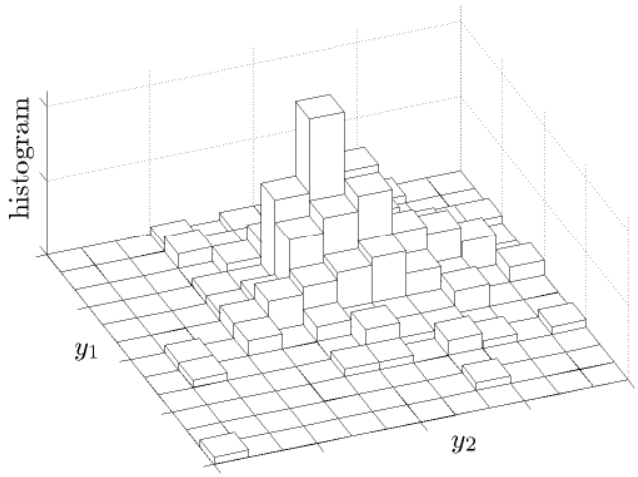


Figure 3.2: 2-D Illustration: histogram.

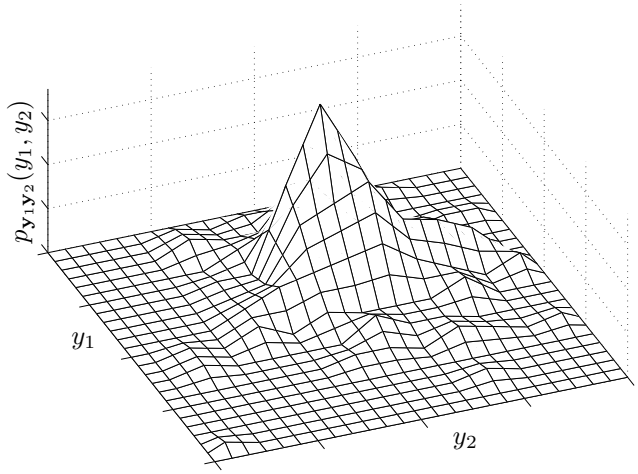


Figure 3.3: 2-D Illustration: the target mjpdf, $p_{\boldsymbol{\mathcal{Y}}} \equiv p_{\mathbf{y}_1\mathbf{y}_2}$, of $\boldsymbol{\mathcal{Y}} = [\mathbf{y}_1, \mathbf{y}_2]^T$.

$p_{1|2}$ as shown in this figure are obtained from the corresponding slices of Figure 3.3 by simply making the area under each slice unity because area under a pdf is always unity,

$$p_{1|2}(y_1\,|\,y_2) = \frac{p_{\mathbf{y}_1\mathbf{y}_2}(y_1, y_2)}{\int_{s_{\mathbf{y}_1}} p_{\mathbf{y}_1\mathbf{y}_2}(y_1, y_2)dy_1} = \frac{p_{\mathbf{y}_1\mathbf{y}_2}(y_1, y_2)}{p_{\mathbf{y}_2}(y_2)},$$

in which $s_{\mathbf{y}_1} = [l_1,\, m_1] \subset \mathbb{R}$ is the support of $\mathbf{y}_1$ and $p_{\mathbf{y}_2}$ is the marpdf of $\mathbf{y}_2$.

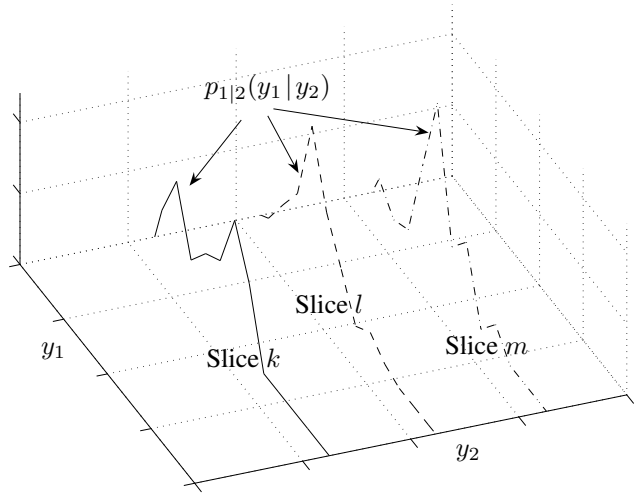


Figure 3.4: 2-D Illustration: three slices representing the conditional pdf of $\mathbf{y}_1$, given $\mathbf{y}_2 = y_2$, for three different y_2's.

Let the associated conditional PDF be denoted by $P_{1|2}$ given by,

$$P_{1|2}(y_1\,|\,y_2) = \frac{\int_{l_1}^{y_1} p_{\mathbf{y}_1\mathbf{y}_2}(y, y_2)dy}{p_{\mathbf{y}_2}(y_2)},$$

as depicted in Figure 3.5. Consider $P_{1|2}(\mathbf{y}_1|y_2)$ and $P_{\boldsymbol{\xi}_1}(\boldsymbol{\xi}_1)$ as two random variables (functions of $\mathbf{y}_1$ and

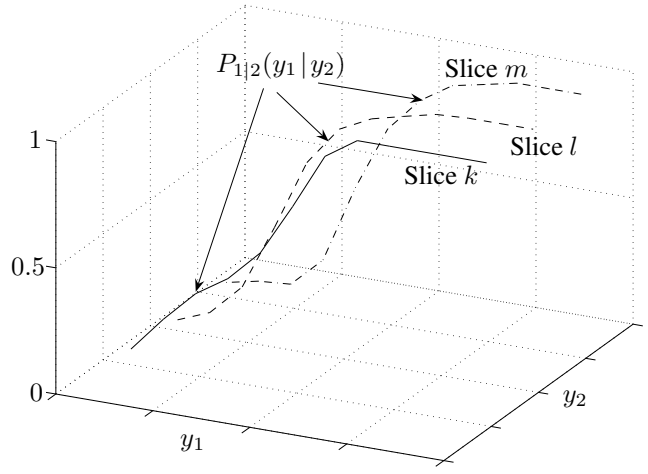


Figure 3.5: 2-D Illustration: three slices representing the conditional PDFs of $\mathbf{y}_1$, given $\mathbf{y}_2 = y_2$, for three different y_2's.

$\boldsymbol{\xi}_1$, respectively). The PDF of both the random variables are uniform distribution supported over $[0,\, 1]$

[HLD04, Theorem 2.1]. Then, the mapping, $T : \underline{\boldsymbol{\xi}} \longrightarrow \boldsymbol{\mathcal{Y}}$, can be defined by employing the Rosenblatt transformation [Ros52] as shown below,

$$\begin{aligned} P_{1|2}(\mathbf{y}_1 \,|\, y_2) &\stackrel{d}{=} P_{\boldsymbol{\xi}_1}(\boldsymbol{\xi}_1) && (3.4)\\ \Rightarrow \mathbf{y}_1 &\stackrel{d}{=} (P_{1|2}^{-1} P_{\boldsymbol{\xi}_1})(\boldsymbol{\xi}_1 \,|\, y_2) && (3.5)\\ &= \lim_{K\to\infty} \sum_{j=0}^{K} a_j(y_2)\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_1). && (3.6) \end{aligned}$$

Equation (3.5) ensures that the conditional PDF of $\mathbf{y}_1$, given $\mathbf{y}_2 = y_2$, is precisely given by $P_{1|2}$ as required [HLD04, Theorem 2.1].

It should be noted here that $f_{1|2} \equiv P_{1|2}^{-1} P_{\boldsymbol{\xi}_1}$ is piecewise smooth [Tol62, p. 18] on the support, $s_{\boldsymbol{\xi}_1} \subseteq \mathbb{R}$, of $\boldsymbol{\xi}_1$ by construction (because of use of linear interpolation), and second-order, $\int_{s_{\boldsymbol{\xi}_1}} f_{1|2}^2(\xi_1 \,|\, y_2) p_{\boldsymbol{\xi}_1}(\xi_1) d\xi_1 < \infty$, by choice of $\boldsymbol{\xi}_1$ and second-order assumption on $\mathbf{y}_1$. This results in the PC representation of $f_{1|2}$ as shown by the rhs of (3.6). It should be noted that while $\mathbf{y}_1$ and the rhs of (3.6) is equal in distribution,

$$\mathbf{y}_1 \stackrel{d}{=} f_{1|2}(\boldsymbol{\xi}_1 \,|\, y_2) = \lim_{K\to\infty} \sum_{j=0}^{K} a_j(y_2)\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_1), \tag{3.7}$$

the equality, "=", above or in (3.6) follows from $f_{1|2}(\boldsymbol{\xi}_1 \,|\, y_2)$ (not from $\mathbf{y}_1$) and is valid at every continuity point of $f_{1|2}$ [Leb72, Chapter 4] implying that this equality can also be interpreted in almost sure (a.s.) sense w.r.t. $P_{\boldsymbol{\xi}_1}$.

The deterministic (since, given y_2) PC coefficient, $\{a_j(y_2), j \in \mathbb{N}\}$, in (3.7) is given by,

$$a_j(y_2) = \frac{E\left[f_{1|2}(\boldsymbol{\xi}_1|y_2)\boldsymbol{\Psi}_j(\boldsymbol{\xi}_1)\right]}{E\left[\boldsymbol{\Psi}_j^2(\boldsymbol{\xi}_1)\right]}, \quad j \in \mathbb{N}. \tag{3.8}$$

The determination of $a_j(y_2)$ requires computation of the following integral,

$$\begin{aligned} & E\left[f_{1|2}(\boldsymbol{\xi}_1|y_2)\boldsymbol{\Psi}_j(\boldsymbol{\xi}_1)\right] \\ = & \int_{s_{\boldsymbol{\xi}_1}} (P_{1|2}^{-1} P_{\boldsymbol{\xi}_1})(\boldsymbol{\xi}_1|y_2)\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_1)\, p_{\boldsymbol{\xi}_1}(\xi_1)\, d\xi_1. \end{aligned}$$

The evaluation of this integral involves computation of the inverse of $P_{1|2}$. Since, in the current context, $P_{1|2}$ is based on observed histogram-based conditional PDF, no suitable analytical inverse function exists

for such nonparametric PDF. Therefore, inverse of this function needs to evaluated numerically while evaluating the above integral. This might be computationally expensive or/and numerically instable. A computationally efficient scheme based on a surrogate function (instead of using $P_{1|2}^{-1} P_{\boldsymbol{\xi}_1}$) is described in the Appendix.

For several different values of $y_2 \in s_{\mathbf{y}_2}$, the PC coefficients, $\{a_j(y_2)\}_{j\in\mathbb{N}}$, need to be computed. Let the support, $s_{\mathbf{y}_2} = [l_2, m_2] \subset \mathbb{R}$, be divided equally into $n_2 \in \mathbb{N}$ intervals. Then, coordinates of the points defining these intervals are given by $y_2^{(k)} = l_2 + k[(m_2 - l_2)/n_2]$, $k = 0, \cdots, n_2$. For each slice defined by $P_{1|2}(\mathbf{y}_1 \mid y_2^{(k)})$, compute the PC coefficients, $\{a_j(y_2^{(k)})\}_{j\in\mathbb{N}}$, by using (3.8). A few typical profiles of the mapping, $\mathbb{N} \ni j \longmapsto a_j(y_2) \in \mathbb{R}$, for given y_2 are depicted in Figure 3.6.

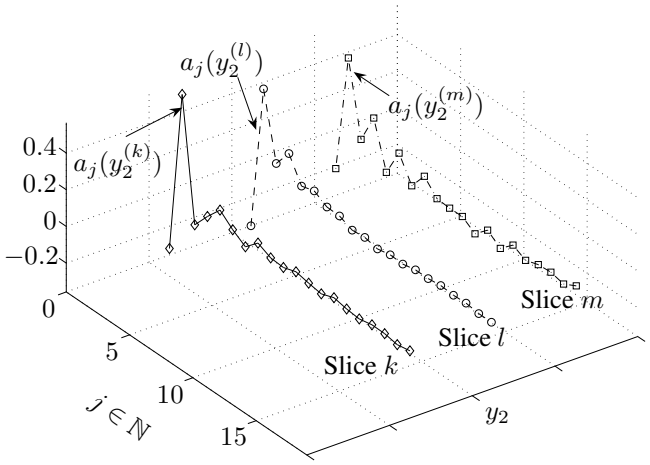


Figure 3.6: 2-D Illustration: $j \longmapsto a_j(y_2)$ for given y_2.

For any given $j \in \mathbb{N}$, the set of pairs, $\{y_2^{(k)}, a_j(y_2^k)\}_{k=0}^{n_2}$, as just determined is next used to construct the mapping, $s_{\mathbf{y}_2} \ni y_2 \longmapsto a_j(y_2) \in \mathbb{R}$, by simply employing a linear interpolation scheme (note that this is an 1-D version of a similar problem for estimating pdf from the histogram defined only over a discrete array of points as already encountered). A few profiles of this mapping are sketched in Figure 3.7.

Since $n_2 \in \mathbb{N}$ is a *finite* (but large) number, the mapping, $y_2 \longmapsto a_j(y_2)$, for any given $j \in \mathbb{N}$, defined via linear interpolation with $\{y_2^{(k)}, a_j(y_2^k)\}_{k=0}^{n_2}$, is piecewise smooth. By the second-order condition on $f_{1|2}$, it also implies that $\mid a_j(y_2) \mid < \infty$ for any given $j \in \mathbb{N}$. It is, therefore, straightforward to select a suitable weight, say, defined by $s_{\mathbf{y}_2} \ni y_2 \longmapsto w_2(y_2) \in (0, \infty)$, such that $\int_{s_{\mathbf{y}_2}} a_j^2(y_2) w_2(y_2) dy_2 < \infty$. Then, a set of basis functions, $\{\psi_k\}_{k\in\mathbb{N}}$, orthogonal w.r.t. the weight $w_2(\cdot)$,

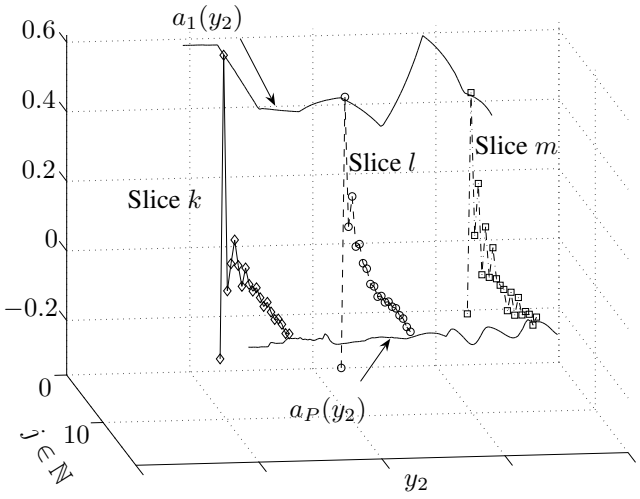


Figure 3.7: 2-D Illustration: $y_2 \longmapsto a_j(y_2)$ for given $j \in \mathbb{N}$.

$\int_{s_{\mathbf{y}_2}} \psi_m(y_2)\psi_n(y_2)w_2(y_2)dy_2 = 0$, $m \neq n$, can be employed to expand the function, $y_2 \longmapsto a_j(y_2)$, in the following series [Leb72, Chapter 4],

$$a_j(y_2) = \lim_{K\to\infty} \sum_{k=0}^{K} b_{jk}\, \psi_k(y_2). \tag{3.9}$$

This series expansion is valid at every continuity point of a_j with b_{jk} computed from,

$$b_{jk} = \frac{\int_{s_{\mathbf{y}_2}} a_j(y_2)\psi_k(y_2)w_2(y_2)dy_2}{\int_{s_{\mathbf{y}_2}} \psi_k^2(y_2)w_2(y_2)dy_2}. \tag{3.10}$$

The denominator are readily available in the literature for many commonly used orthogonal polynomials [Leb72, Chapter 4], [GS91, XK02, SG04a]. The numerator can be evaluated by using any standard numerical integration scheme.

Use of (3.9) in (3.7) results in,

$$\mathbf{y}_1 \stackrel{d}{=} f_{1|2}(\boldsymbol{\xi}_1\,|\,y_2) = \lim_{\substack{K_1\to\infty \\ K_2\to\infty}} \sum_{j=0}^{K_1}\sum_{k=0}^{K_2} b_{jk}\, \psi_k(y_2)\boldsymbol{\Psi}_j(\boldsymbol{\xi}_1). \tag{3.11}$$

Now, the marPDF, P_2, of $\mathbf{y}_2$ can be similarly (consider 1-D cases of the series of Figures 3.1-3.6) employed to obtain the following PC expansion for $\mathbf{y}_2$,

$$\mathbf{y}_2 \stackrel{d}{=} f_2(\boldsymbol{\xi}_2) = \lim_{K\to\infty} \sum_{j=0}^{K} c_j \, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_2), \tag{3.12}$$

in which $f_2 \equiv P_2^{-1} P_{\boldsymbol{\xi}_2}$ and c_j is given by,

$$c_j = \frac{E\left[f_2(\boldsymbol{\xi}_2)\boldsymbol{\Psi}_j(\boldsymbol{\xi}_2)\right]}{E\left[\boldsymbol{\Psi}_j^2(\boldsymbol{\xi}_2)\right]}, \quad j \in \mathbb{N}, \tag{3.13}$$

and can be efficiently computed by using the simple scheme described in the Appendix.

The PC expansions, (3.11) and (3.12), constructed from the available measurement data, together completely characterize the random vector, $\boldsymbol{\mathcal{Y}} = [\mathbf{y}_1, \mathbf{y}_2]^T$. In a computational set-up, the series in (3.11) and (3.12) are truncated after suitable large number of terms.

Sampling of $\boldsymbol{\mathcal{Y}}$ is straightforward. The random variables, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, are statistically independent. First, use (3.12) to generate a sample, y_2, of $\mathbf{y}_2$ and then use the realized value, y_2, in (3.11) to get y_1. Repeat the process until the desired number of samples of $\boldsymbol{\mathcal{Y}} = [\mathbf{y}_1, \mathbf{y}_2]^T$ is generated.

Extension of the above 2-D formulation to the N-variate $\boldsymbol{\mathcal{Y}}$ is now summarized below,

$$\begin{aligned}
\mathbf{y}_1 &\stackrel{d}{=} P_{1|2:N}^{-1} P_{\boldsymbol{\xi}_1}(\boldsymbol{\xi}_1 \,|\, y_2, \cdots, y_N) \\
&= \sum_{i_1=0}^{K_1^{(1)}} \cdots \sum_{i_N=0}^{K_N^{(1)}} b_{i_1 i_2 \cdots i_N}^{(1)} \, \psi_{i_N}(y_N) \cdots \psi_{i_2}(y_2) \boldsymbol{\Psi}_{i_1}(\boldsymbol{\xi}_1) \\
\mathbf{y}_2 &\stackrel{d}{=} P_{2|3:N}^{-1} P_{\boldsymbol{\xi}_2}(\boldsymbol{\xi}_2 \,|\, y_3, \cdots, y_N) \\
&= \sum_{i_2=0}^{K_2^{(2)}} \cdots \sum_{i_N=0}^{K_N^{(2)}} b_{i_2 \cdots i_N}^{(2)} \, \psi_{i_N}(y_N) \cdots \psi_{i_3}(y_3) \boldsymbol{\Psi}_{i_2}(\boldsymbol{\xi}_2) \\
&\vdots \\
\mathbf{y}_N &\stackrel{d}{=} P_N^{-1} P_{\boldsymbol{\xi}_N}(\boldsymbol{\xi}_N) = \sum_{i_N=0}^{K_N^{(N)}} b_{i_N}^{(N)} \, \boldsymbol{\Psi}_{i_N}(\boldsymbol{\xi}_N).
\end{aligned}$$

Here, $P_{i|(i+1):N}$ is conditional PDF of $\mathbf{y}_i$, given $\mathbf{y}_{i+1} = y_{i+1}, \cdots, \mathbf{y}_N = y_N$, induced by $P_{\boldsymbol{\mathcal{Y}}}$ and $b_{j_i j_{i+1} \cdots j_N}^{(i)}$ represents $(N-(i-1))$-dimensional matrix of PC coefficients of size $K_i^{(i)} \times \cdots \times K_N^{(i)}$ with $K_i^{(i)}, \cdots, K_N^{(i)}$ being the suitable large integers retained in the corresponding series expansion. The

random variables, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_N$, are statistically independent. Each digital sample of $\boldsymbol{\mathcal{Y}}$ is generated starting with sampling $\mathbf{y}_N$ and successively proceeding towards sampling $\mathbf{y}_{N-1}, \mathbf{y}_{N-2}, \cdots$, and in last sampling $\mathbf{y}_1$.

Finally, let us conclude this section by emphasizing that $P_{i|(i+1):N}$ should not be computed by integrating $p_{\boldsymbol{\mathcal{Y}}}$ since it would then involve a substantial computational effort to perform several multidimensional integrations while approximating the corresponding function, $P^{-1}_{i|(i+1):N} P_{\boldsymbol{\xi}_i}$ (see Appendix). Instead, $P_{i|(i+1):N}$ should be computed from estimate of mjpdf of $(\mathbf{y}_i, \cdots, \mathbf{y}_N)$ determined by considering only the measurement data associated with $\mathbf{y}_i, \cdots, \mathbf{y}_N$, and completely ignoring the data associated with $\mathbf{y}_1, \cdots \mathbf{y}_{i-1}$. This would always involve 1-D integration in computation of $P_{i|(i+1):N}$,

$$P_{i|(i+1):N}(y_i \,|\, y_{i+1}, \cdots, y_N) = \frac{\int_{l_i}^{y_i} p_{\mathbf{y}_i, \cdots, \mathbf{y}_N}(y_i, \cdots, y_N)\, dy_i}{p_{\mathbf{y}_{i+1}, \cdots, \mathbf{y}_N}(y_{i+1}, \cdots, y_N)}.$$

Here, the integration is carried over the domain, $[l_i,\, y_i] \subseteq s_{\mathbf{y}_i} = [l_i,\, m_i] \subset \mathbb{R}$, where $s_{\mathbf{y}_i}$ is the support of $\mathbf{y}_i$. This scheme would be relatively computationally inexpensive even after the additional computational overhead required to estimate the set of pdfs, $p_{\mathbf{y}_2, \cdots, \mathbf{y}_N},\ p_{\mathbf{y}_3, \cdots, \mathbf{y}_N}, \cdots, p_{\mathbf{y}_N}$, (from the corresponding data) that need to determined only once at the outset.

3.2.2 Approach 2: Based on Marginal PDFs and SRCC

In this approach, the unknown relationship between $\underline{\boldsymbol{\xi}}$ and $\boldsymbol{\mathcal{Y}}$ is defined again by having recourse to the Rosenblatt transformation establishing a set of N mappings, each of which is similar to (3.12), between the corresponding k-th components, $\mathbf{y}_k$ and $\boldsymbol{\xi}_k$, $k = 1, \cdots, N$. It should be noted that the Rosenblatt transformation, when applied on marPDF of a scalar-variate random variable, is similar to the Nataf transformation. Only marPDF of $\mathbf{y}_k$ is used in this approach. Unlike $\boldsymbol{\xi}_k$'s in Approach 1, the random variables, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_N$, here are, however, statistically dependent enforcing the required statistical dependencies among $\mathbf{y}_k$'s. The statistical dependency is characterized via SRCC. In the following, the definition and the relevant properties of SRCC are briefly reviewed first before describing Approach 2.

Spearman's Rank Correlation Coefficient

The rank correlation coefficient or Spearman's rho is named after Charles Edward Spearman who first introduced it [Spe04]. The rank correlation coefficient between random variables, $\mathbf{y}_i$ and $\mathbf{y}_j$, is simply

the PCC applied to the rank of the observed samples of $\mathbf{y}_i$ and $\mathbf{y}_j$ rather than to their observed or measured values. When there is no tie in the observed values of the data, a simple formula exists for the calculation of SRCC [Man01, p. 655]. Further theoretical treatment and calculation procedural of SRCC including the case of tied data values can be found in the literature (see e.g., [Leh75, p. 297-303], [PTVF96, p. 634-637]). Statistical toolbox of MATLAB provides function, `corr`, that can be used to calculate SRCC.

Definition 3.2.1 *The* **Spearman's rank correlation coefficient** *between two random variables,* $\mathbf{y}_i$ *and* $\mathbf{y}_j$*, with marginal probability distribution functions, respectively, being given by* $P_{\mathbf{y}_i}$ *and* $P_{\mathbf{y}_j}$*, is defined as,*

$$\rho_s(\mathbf{y}_i, \mathbf{y}_j) = \rho(P_{\mathbf{y}_i}(\mathbf{y}_i), P_{\mathbf{y}_j}(\mathbf{y}_j)) = 12\,\mathrm{cov}(P_{\mathbf{y}_i}(\mathbf{y}_i), P_{\mathbf{y}_j}(\mathbf{y}_j)). \tag{3.14}$$

Here, ρ *is the Pearson's correlation coefficient (or the usual product-moment correlation coefficient),* cov *is the covariance and the multiplying factor,* 12*, emanates from variance of* $P_{\mathbf{y}_k}(\mathbf{y}_k)$*,* $k = i, j$*, since* $P_{\mathbf{y}_k}(\mathbf{y}_k) \sim U(0,\,1)$ *with* $U(0,\,1)$ *being uniform distribution on* $[0,\,1]$ *(see, e.g.,[HLD04, Theorem* 2.1*]).*

It must be noted from the above definition that SRCC and PCC coincide if PDFs of $\mathbf{y}_i$ and $\mathbf{y}_j$ are $U(0,\,1)$. In general, they are, however, different.

A collection of a few salient properties of ρ_s is enlisted below [EMS01, Section 4.3], [KC06, Section 3.2.2]. It

- always exists and is symmetric;
- is independent of marpdf of $\mathbf{y}_i$ and $\mathbf{y}_j$;
- is invariant under strictly monotone transformation of $\mathbf{y}_i$ and $\mathbf{y}_j$;
- can take any values from the closed interval, $[-1,\,1]$;
- is zero if $\mathbf{y}_i$ and $\mathbf{y}_j$ are statistically independent, the converse is not true.

The most important property to be used in the present work is invariance under monotone transformation property of SRCC.

Now that the relevant information on SRCC is set forth, Approach 2 is described below by introducing the mapping, $\boldsymbol{\xi}_k \longmapsto \mathbf{y}_k$, $k = 1, \cdots, N$, through the use of the Rosenblatt transformation [Ros52] applied on each $\boldsymbol{\xi}_k$ separately,

$$\mathbf{y}_k \overset{d}{=} q_k(\boldsymbol{\xi}_k) = \lim_{K_k \to \infty} \sum_{j=0}^{K_k} c_{jk}\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_k),\ q_k(\boldsymbol{\xi}_k) \equiv P_{\mathbf{y}_k}^{-1} P_{\boldsymbol{\xi}_k}. \tag{3.15}$$

This PC representation is similar to (3.12). The marPDF, $P_{\mathbf{y}_k}$, is estimated from the normalized and linearly interpolated 1-D histogram of the measurement data on each random variable component, $\mathbf{y}_k$, separately. This can be readily performed as already discussed in section 3.2.1. The PC representation of q_k in (3.15) is, therefore, valid at every continuity of q_k implying that the equality, '=', can also be interpreted in a.s. sense w.r.t. $P_{\boldsymbol{\xi}_k}$. The PC coefficient, c_{jk}, is given by,

$$c_{jk} = \frac{E\left[q_k(\boldsymbol{\xi}_k)\boldsymbol{\Psi}_j(\boldsymbol{\xi}_k)\right]}{E\left[\boldsymbol{\Psi}_j^2(\boldsymbol{\xi}_k)\right]}, \quad j \in \mathbb{N}. \tag{3.16}$$

A simple and computationally efficient scheme based on 1-D interpolated surrogate function, approximating $P_{\mathbf{y}_k}^{-1} P_{\boldsymbol{\xi}_k}$, is described in Appendix to determine $\{c_{jk}\}_{j \in \mathbf{N}}$, $k = 1, \cdots, N$.

The series in (3.15) is truncated after a large number of terms, K_k. Since SRCC is preserved under monotone transformation, the SRCC matrices of $\underline{\boldsymbol{\xi}} = [\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_N]^T$ and $\boldsymbol{\mathcal{Y}}$ are identical. The target $N \times N$ SRCC matrix, $[\rho_s]$, is simply estimated from the available measurement data on $\boldsymbol{\mathcal{Y}}$. If (i, j)-th, $i, j = 1, \cdots, N$, element of $[\rho_s]$ is denoted by $(\rho_s)_{ij}$, then $(\rho_s)_{ij} = \rho_s(\mathbf{y}_i, \mathbf{y}_j)$. The samples of $\underline{\boldsymbol{\xi}}$, with SRCC matrix, $[\rho_s]$, are first generated. Subsequently, samples of each $\boldsymbol{\xi}_k$ are substituted in the corresponding PC expansion of $\mathbf{y}_k$ to obtain the realizations of $\mathbf{y}_k$. The resulting samples of $\boldsymbol{\mathcal{Y}}$ are consistent with the target set, $\{p_{\mathbf{y}_k}\}_{k=1}^N$, of marpdfs and the target SRCC matrix, $[\rho_s]$.

The commonly used PC random variables, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_N$, that are often chosen to construct PC representation, are standard Gaussian random variables, uniform random variables on $[-1,\ 1]$, beta type I random variables on $[-1,\ 1]$ or gamma random variables. The generation of samples of such statistically independent random variables, as required in Approach 1, is straightforward. The samples of statistically dependent random variables, particularly when the statistical dependency is characterized by a specified SRCC matrix, $[\rho_s]$, as required in Approach 2, can also be readily generated by using the existing simulation techniques. However, these simulation schemes are scattered across the spectrum of literatures:

MC simulation to management science. Therefore, for the sake of completeness of the present work, two useful and easily implementable techniques are summarized in the next two subsections. These two techniques are directly related to concept of copula [Joe97, EMS01, Nel06, KC06] knowledge of which, though useful, is not required here.

Normal Copula Technique

This technique assumes existence of an underlying correlated N-D standard Gaussian random vector, $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]^T$, in which each component, $\mathbf{x}_i$, is a standard Gaussian random variable. If $\mathbf{X}$ exists, i.e., if a feasible (positive-definite) covariance matrix is found, then it is the fastest method among all the currently existing methods. In such situation, the correlation (also, covariance) matrix, $[\rho]$, of $\mathbf{X}$ is determined as follows.

It was shown by Pearson in 1904 that [KC06, p. 51 and p. 75-77],

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = 2 \sin\left(\frac{\pi}{6}\rho_s(\mathbf{u}_i, \mathbf{u}_j)\right), \tag{3.17}$$

in which $\mathbf{u}_i \sim U(0,\, 1)$ and $\mathbf{u}_j \sim U(0,\, 1)$ are uniform random variables. If PDF of the standard Gaussian random variable is denoted by $\Phi(\cdot)$, then $\Phi(\mathbf{x}_i) \sim U(0,\, 1), \forall i$ [HLD04, Theorem 2.1]. Let us then select $\mathbf{u}_i$'s in (3.17) as $\mathbf{u}_i \equiv \Phi(\mathbf{x}_i)$. Consider now the following mapping based on the Rosenblatt transformation,

$$\mathbf{u}_i \equiv \Phi(\mathbf{x}_i) \stackrel{d}{=} P_{\mathbf{y}_i}(\mathbf{y}_i), \quad i = 1, \cdots, N, \tag{3.18}$$

since $P_{\mathbf{y}_i}(\mathbf{y}_i) \sim U(0,\, 1)$ [HLD04, Theorem 2.1]. By the invariance under monotone transformation property of the SRCC, we have $\rho_s(\mathbf{y}_i, \mathbf{y}_j) = \rho_s(P_{\mathbf{y}_i}(\mathbf{y}_i), P_{\mathbf{y}_j}(\mathbf{y}_j))$. Then, by (3.14) and "$\stackrel{d}{=}$" in (3.18), the SRCC matrix of $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N]^T$ is given by $[\rho_s]$ with its (i, j)-th element being given by $\rho_s(\mathbf{y}_i, \mathbf{y}_j)$ estimated based on the measurement data on $\boldsymbol{\mathcal{Y}}$. The correlation (or covariance) matrix, $[\rho]$, of $\mathbf{X}$ then follows from (3.17), with the (i, j)-th, $i, j = 1, \cdots, N$, element, ρ_{ij}, of $[\rho]$ being given by $\rho(\mathbf{x}_i, \mathbf{x}_j)$. Simulation of the standard Gaussian random vector, $\mathbf{X}$, with covariance matrix, $[\rho]$, is then straightforward. In the literature, PDF of $\mathbf{U}$ is usually referred as normal copula.

Since $P_{\boldsymbol{\xi}_i}(\boldsymbol{\xi}_i) \sim U(0,\,1)$ [HLD04, Theorem 2.1], use of the following transformation (again based on the Rosenblatt transformation),

$$\left.\begin{array}{rcl} P_{\boldsymbol{\xi}_i}(\boldsymbol{\xi}_i) & \stackrel{d}{=} & \Phi(\mathbf{x}_i) \equiv \mathbf{u}_i \\ \Rightarrow \boldsymbol{\xi}_i & \stackrel{d}{=} & P_{\boldsymbol{\xi}_i}^{-1}\Phi(\mathbf{x}_i), \end{array}\right\}, \quad i = 1, \cdots, N, \tag{3.19}$$

yields the samples of $\underline{\boldsymbol{\xi}} = [\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_N]^T$. The SRCC matrix of $\underline{\boldsymbol{\xi}}$ again turns out to be $[\rho_s]$ by the invariance under monotone transformation property of the SRCC. The closed form expression of, or the efficient algorithm to compute the inverse function, $P_{\boldsymbol{\xi}_i}^{-1}$, associated with the commonly used PC random variables can be readily extracted from the standard textbook on MC simulation (see e.g., [Fis96, Section 3.2], [HLD04, Section 2.1]). The MATLAB statistical toolbox provides many such useful functions. Clearly, the simulation of $\underline{\boldsymbol{\xi}}$ with SRCC matrix, $[\rho_s]$, essentially reduces to the simulation of an N-D standard Gaussian random vector with covariance matrix $[\rho]$ (if it exists).

Let us consider the last remark about the existence of feasible covariance matrix of $\mathbf{X}$ more carefully. Denote the set of symmetric $N \times N$ positive definite real matrices by $\mathbb{M}_N^+(\mathbb{R})$ and $\mathbb{S}_N(\mathbb{R}) = \{A : A \in \mathbb{M}_N^+(\mathbb{R}),\ A_{ii} = 1\}$, in which A_{ij} is (i,j)-th element of A. Then, for any $[\rho_s] \in \mathbb{S}_M(\mathbb{R})$, there always exists a random vector with uniform marPDFs and SRCC matrix, $[\rho_s]$ [KC06, Theorem 4.4, p. 100, 124-125]. It does not, however, necessarily mean that its uniform random variable components can be given by $\Phi(\mathbf{x}_i)$'s. Counterexamples exist in the literature (see e.g., [GH02, GH03], [HLD04, Section 12.5.2], [KC06, Section 4.2]) showing that application of the mapping defined by (3.17) on each element, $(\rho_s)_{ij}$, of $[\rho_s]$ may produce a matrix, $[\rho^{(1)}]$, that is not a positive-definite matrix, thus rendering the normal copula technique of no use. This problem becomes increasingly severe as the dimension, N, of the random vector increases. The following alternative technique might be then useful.

Augmented Normal Copula Techniques

Application of these techniques ensures that the samples of $\mathbf{U}$ follow the uniform marPDFs but the SRCC or PCC matrix (identical by definition for uniform distribution) is approximate in the sense that the target correlation matrix is modified to a 'new' correlation matrix that is close, in some sense, to the originally estimated correlation matrix. Let us denote the original matrix by $[\rho_s^{(1)}]$ and the modified positive-definite

correlation matrix by $[\rho_s]$. With this new target correlation matrix, $[\rho_s]$, the use of the normal copula technique, as described in the previous subsection, becomes feasible.

One such technique [vdG98, Section 5] suggests to adapt $[\rho_s^{(1)}]$ and $[\rho^{(1)}]$, to new positive-definite correlation matrices, $[\rho_s]$ and $[\rho]$, by using a simple iterative scheme based on the spectral decomposition of real Hermitian matrices. While this scheme might work in practice, it is likely to be little unwieldy, particularly in high dimension, requiring too many iterations often resulting in relatively large error between the old and modified matrices.

Another technique [GH03] is a constrained minimization problem in the space of $\mathbf{X}$ and relatively more robust. Two metrics, in particular, $L_1 = \sum_{i<j} \mid \rho_{ij} - \rho_{ij}^{(1)} \mid$ norm and $L_\infty = \max_{i<j} \mid \rho_{ij} - \rho_{ij}^{(1)} \mid$ norm, are minimized subject to $[\rho] \in \mathbb{S}_N(\mathbb{R})$ [GH03]. It is, however, not guaranteed that the resulting 'new' correlation matrix, $[\rho_s]$, of $\mathbf{U}$ (by applying the inverse transformation of (3.17) on $[\rho]$) would be positive-definite and close to the originally specified target correlation matrix, $[\rho_s^{(1)}]$, of $\mathbf{U}$. In such situation, an iterative scheme like the one proposed earlier in the literature [vdG98, Section 5] might be adopted.

In the present work, the following constrained minimization problem, similar to the works presented in the literature [GH03], is recommended,

$$\begin{aligned} &\text{minimize} && \|[\rho] - [\rho^{(1)}]\|_F \\ &\text{subject to} && [\rho] \in \mathbb{S}_N(\mathbb{R}), \end{aligned} \tag{3.20}$$

or/and other meaningful constraints (see e.g., [GH03, Section 5]). The Frobenius norm is preferred (over L_1 and L_∞ norms) since it shows relatively much smaller error (even in high dimension). The above optimization problem can be efficiently solved by employing the semi-definite program (SDP) [VB96], [Dat05, Chapter 4]. Many efficient freely available softwares[1] exist to solve such SDP. In the present work, a public domain MATLAB toolbox, YALMIP, developed by Löfberg ([Lof04]), is used.

The techniques as discussed above should be applied only if the new correlation matrix, $[\rho_s]$, of $\mathbf{U}$ is positive-definite and is close, in the appropriate sense, to the originally specified target correlation matrix, $[\rho_s^{(1)}]$. Otherwise, alternative techniques [MB93, MB97, GH02, KC06] at the expense of significantly additional computational time and resource might be interrogated. In many practical applications, the two

[1] `http://www-user.tu-chemnitz.de/~helmberg/semidef.html`

recommended techniques — normal copula technique and augmented normal copula technique — are, nevertheless, likely to be satisfactory.

3.3 Practical Illustration and Discussion

The proposed techniques are employed here to construct the PC representation of a spatio-temporal random temperature field by using a set of oceanographic data obtained from a shallow-water acoustics transmission experiment. This experiment would be referred now onwards as SWARM95 (Shallow Water Acoustics in Random Medium) experiment. It was conducted during the month of July-August in 1995 in the Mid-Atlantic Bight continental shelf region off the coast of New Jersey [ABC$^+$97].

The primary objective of the SWARM95 experiment is to investigate the effects of random variations of the oceanographic parameters, for example, temperature and salinity fields, on the statistical properties of the acoustic field. The acoustic field is perturbed significantly by a small change in water column sound speed distribution. The sound speed variation depends on the internal wave field and the oceanographic parameters through an integral equation. This internal wave field is also governed by partial differential equations with random coefficients depending on the oceanographic parameters. Further details and precise objective of the experiment are documented and discussed in other research papers [ABC$^+$97, FOT$^+$00]. In the current chapter, modeling of the spatio-temporal random temperature field from the oceanographic measurements of SWARM95 experiment would only be considered. The PC representation of the spatio-temporal random field modeling the oceanographic parameters would be useful in propagating the uncertainty in a rational manner to the prediction of the acoustic field and in estimating the confidence interval of the associated statistical parameters by employing the techniques available elsewhere [GRH99, PG04, DG04, GD06, DGS08] (also see chapter 2).

There are 3 vertical strings through 72 m depth of water column each with 11 temperature sensors measuring the temperature histories. These temperature senors are located at depth $h \in D = \{16,\ 21,\ 26,\ 31,\ 36,\ 38.5,\ 41,\ 46,\ 48.5,\ 51,\ 56\}$ m. The temperature data are sampled every minute and there is a total of 17281 samples from each sensor. The three strings would hereafter be referred as tav309, tav307 and tav598 as per nomenclature rule decided earlier for a different analysis (not a part of the current work) conducted on this set of temperature data. A few typical time histories obtained from tav309 are shown in Figure 3.8. However, it is imperative to separate the background internal wave field from the

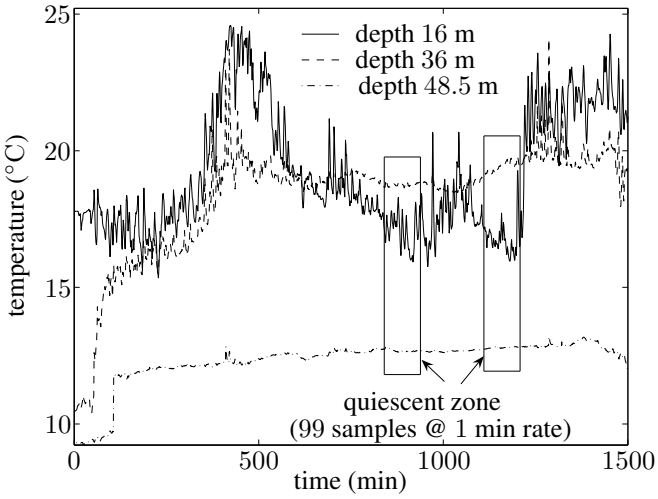


Figure 3.8: A few experimentally measured time histories (shown only for a segment of the total experimental time span).

solitary wave contribution while computing some intermediate oceanographic parameters, for example, buoyancy frequency, that is required to compute the sound speed fluctuation [FOT$^+$00]. Therefore, only the "quiescent" part of the measurement data excluding the solitary waves must be used while computing such intermediate parameters. The most active solitary wave region is in the upper half of the water column.

3.3.1 Selecting the Regions of Low Internal Solitary Wave Activity

There is some subjectivity in choosing the quiescent part of the temperature data because it is next to impossible to completely separate the background internal wave field from the solitary wave contribution. The mathematical decomposition of the sound speed distribution into deterministic, time-dependent field and a random fluctuation about this deterministic field, as discussed in previous work [FOT$^+$00], is an idealization. In a real ocean experiment, the situation is much more complicated. In order to estimate the oceanographic parameters, e.g., buoyancy frequency, it is important to try to stay away from regions containing the obvious large fluctuations that often start with a jump discontinuity. These regions are usually associated with the main components of the solitary wave train. Therefore, the highly variable regions, containing the strong solitary wave activity, are not used in the following analysis.

By visual inspection, the regions in the boxes, for example, as shown in Figure 3.8, are examples of "low" internal solitary wave activity and suitable for reliable estimation of the buoyancy frequency, and

consequently, selected for further analysis. A total of such 8 time-segments each with 99 temperature measurements at any $h \in D$ are selected from the whole span of the experimentally measured time history. Out of 17281 samples available from each sensor, only 8×99 samples are deemed to be useful in constructing the PC representation of the spatio-temporal random temperature field. The resulting PC representation would be useful for other analysis involving (stochastic) oceanographic parameters that depend on the random temperature field.

More detailed features of a typical quiescent segment showing the time histories collected from a few sensors (at different depths) attached to one of the 3 strings (tav309) are shown in Figure 3.9. Each

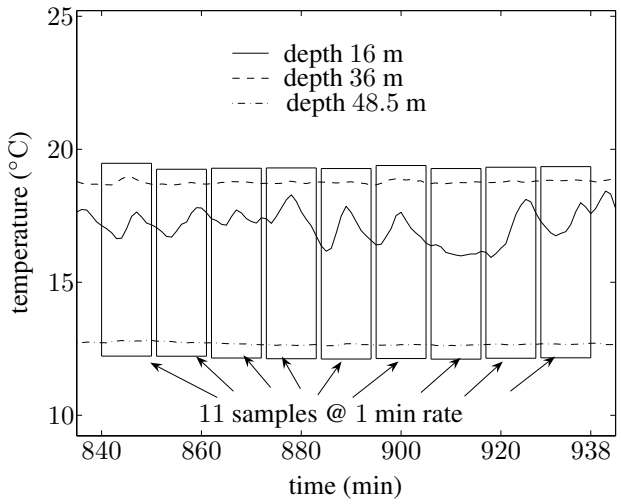


Figure 3.9: A typical quiescent zone divided into 9 smaller segments with 11 samples (shown for a few sensors).

quiescent segment with 99 samples is further divided into 9 smaller segments with each containing 11 samples as shown in this figure.

At any given time instant, all the 11 sensors located at $h \in D$ are measuring the temperature (at 1 min sampling rate) simultaneously. Consider a spatio-temporal domain defined by one *smaller* segment associated with the quiescent zone and the 72 m depth of water column along which SWARM95 experiment was conducted. Let us assume that the random temperature field is statistically independent and identically distributed (i.i.d.) both across the smaller segments with 11 samples as shown in Figure 3.9 within a given quiescent zone as well as across the different quiescent zones as shown in Figure 3.8. Without any further loss of generality, time can be, therefore, conveniently reset to $t = 0$ at the beginning of each of these smaller segments as illustrated in Figure 3.10. Denote the spatio-temporal domain thus described by

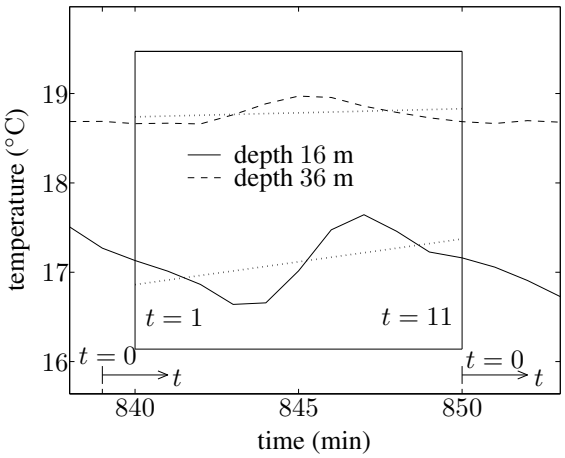


Figure 3.10: A typical subset of $(\mathcal{T} \times \mathcal{D})$ with two time histories collected from tav309; dotted lines indicate linear fit to the experimental data.

$(\mathcal{T} \times \mathcal{D})$ in which $\mathcal{T} = (0,\, 11)$ min and $\mathcal{D} = (0,\, 72)$ m. Denote the random temperature field evolving over $(\mathcal{T} \times \mathcal{D})$ by $(\mathcal{T} \times \mathcal{D}) \ni (t, h) \longmapsto \mathbf{\Gamma}(t, h) \in \mathbb{R}$.

3.3.2 Detrending the Data

The average trends of the oscillatory time histories are obtained by fitting the data linearly within each smaller segment as shown as dotted lines in Figure 3.10. Within a given segment, suppose that the experimentally measured data, for any given $h \in D$, is represented by $\Gamma^{(\mathrm{meas})}(t, h)$ and the linear trend of the measurement by $\overline{\Gamma}(t, h)$. Then, define a normalized spatio-temporal random temperature field, $\mathbf{\Gamma}^{(\mathrm{n})}(t, h)$, as,

$$\mathbf{\Gamma}^{(\mathrm{n})}(t, h) = \frac{\mathbf{\Gamma}(t, h) - \overline{\Gamma}(t, h)}{\overline{\Gamma}(t, h)}. \tag{3.21}$$

The experimental samples of $\mathbf{\Gamma}^{(\mathrm{n})}(t, h)$ can be readily deduced by substituting $\mathbf{\Gamma}(t, h)$ with $\Gamma^{(\mathrm{meas})}(t, h)$ in (3.21). A few such typical experimental samples of $\mathbf{\Gamma}^{(\mathrm{n})}(t, h)$ are shown in Figure 3.11.

In the following, $\mathbf{\Gamma}^{(\mathrm{n})}(t, h)$ is modeled by employing the approaches as proposed in the present work based on the resulting experimental samples. Once the PC representation of $\mathbf{\Gamma}^{(\mathrm{n})}(t, h)$ is available, PC representation of the original random temperature field, $\mathbf{\Gamma}(t, h)$, immediately follows from $\mathbf{\Gamma}(t, h) = \overline{\Gamma}(t, h)\mathbf{\Gamma}^{(\mathrm{n})}(t, h) + \overline{\Gamma}(t, h)$. The linear fit, $\overline{\Gamma}(t, h)$, has already been deduced by using the experimental samples of $\mathbf{\Gamma}(t, h)$. The separation of this average trend from $\mathbf{\Gamma}(t, h)$ essentially adds a certain flexibility to the scheme of modeling $\mathbf{\Gamma}(t, h)$ as adopted in this numerical illustration. This, in particular,

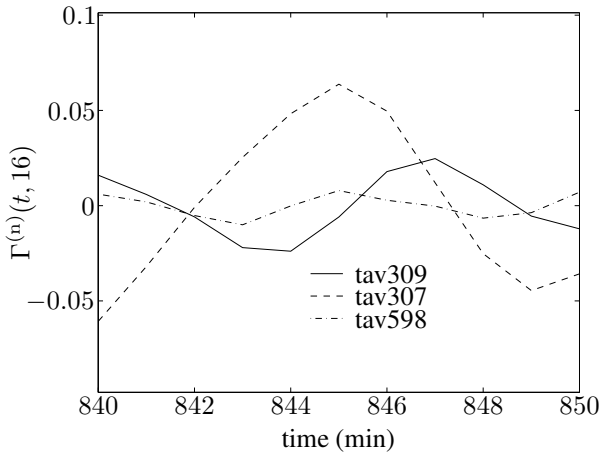


Figure 3.11: A few typical profiles of experimental samples of $\mathbf{\Gamma}^{(\mathrm{n})}(t,h)$; $(t,h) \longmapsto \Gamma^{(\mathrm{n})}(t,h)$ at $h =$ 16 m depth.

facilitates in inferring the PC coefficients of $\mathbf{\Gamma}(t,h)$, $(t,h) \notin (\mathcal{T} \times \mathcal{D})$ (assuming that the corresponding $\overline{\Gamma}(t,h)$ can be reliably estimated from the experiment or is available from other sources/experiments). The normalization by $\overline{\Gamma}(t,h)$ as shown in (3.21) also facilitates in achieving certain numerical stability to the ensuing analysis since values of the experimental measurements collected from sensors at different depths show significant variations (see Figure 3.12). This variation should be compared with the variation

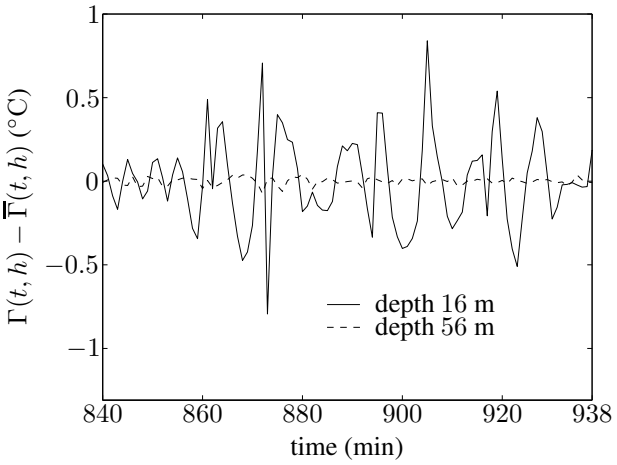


Figure 3.12: Experimental variation of temperature measurements after removing the linear trends and before normalization (shown for two time histories and over a quiescent zone).

after the normalization as shown in Figure 3.13.

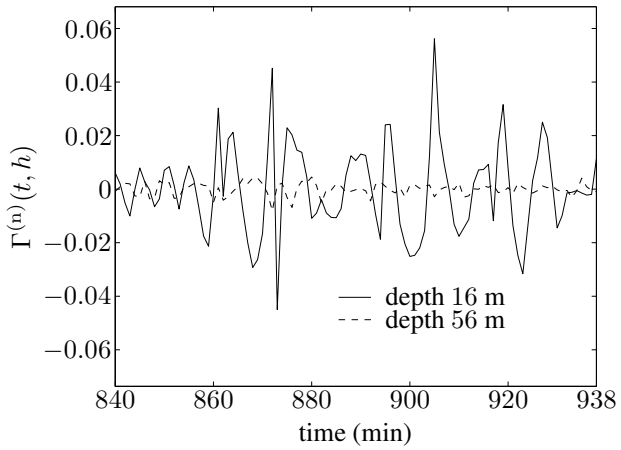


Figure 3.13: Variation of the normalized temperature measurements (shown for two time histories and over a quiescent zone).

3.3.3 Stochastic Modeling of $\mathbf{\Gamma}^{(\mathrm{n})}(t,h)$

For any given $(t,h) \in (\mathcal{T} \times \mathcal{D})$, $\mathbf{\Gamma}^{(\mathrm{n})}(t,h)$ represent a random variable. Clearly, the experimental measurements essentially represent the samples of a finite set of these random variables. Recall that D represents the set of coordinates of the sensors attached to the strings along 72 m water column. Let us now also denote the set of time instants (per convention of Figure 3.10) of collection of experimental samples by $T = \{1,\, 2, \cdots, 10,\, 11\}$ (since sampling rate is 1 min). Note the difference between the continuous space, $(\mathcal{T} \times \mathcal{D})$, over which $\mathbf{\Gamma}^{(\mathrm{n})}(t,h)$ is evolving and the discrete space, $(T \times D)$, consisting of only a finite set of points at which the experimental samples are available.

Let us denote the set of 11×11 random variables, $\{\mathbf{\Gamma}^{(\mathrm{n})}(t,h)\}_{(t,h)\in(T\times D)}$, collectively by $\boldsymbol{\mathcal{Y}}$, i.e., $N = \dim(\boldsymbol{\mathcal{Y}}) = 121$. Since each quiescent zone is divided into 9 smaller segments (see Figure 3.9) and 8 quiescent zones are selected (see section 3.3.1) across the whole span of the experimental time histories, there are $8 \times 9 = 72$ statistically independent samples of $\boldsymbol{\mathcal{Y}}$ from each string. In the present work, a space and time separability condition of statistical dependency of the original random temperature field is assumed for the time and spatial extent spanning the sea surface. However, no such space-time separability is assumed for the time and depth, i.e., for $(\mathcal{T} \times \mathcal{D})$. The random variable components of $\boldsymbol{\mathcal{Y}}$ are, therefore, statistically dependent. From three vertical strings, tav309, tav307 and tav598 (about 10 km away from each other), a total of $n = 3 \times 72 = 216$ samples of $\boldsymbol{\mathcal{Y}}$ are available.

The task is now to construct PC representations of $\boldsymbol{\mathcal{Y}}$ by using the approaches as proposed in the present work with 210 samples of $\boldsymbol{\mathcal{Y}}$. The PC representations would be consistent with the information

extracted from these 210 experimental samples. Further details and results are discussed in the next subsections.

3.3.4 Modeling of $\boldsymbol{\mathcal{Y}}$ via Approach 1

The Karhunen-Loève (KL) decomposition is first employed to construct a reduced order model of the non-Gaussian random vector, $\boldsymbol{\mathcal{Y}}$. Though the resulting non-Gaussian KL random variable components are uncorrelated, they are, in general, statistically dependent. Approach 1 is subsequently used to characterize this reduced order model of $\boldsymbol{\mathcal{Y}}$.

KL Decomposition of $\boldsymbol{\mathcal{Y}}$

Let n experimental samples of $\boldsymbol{\mathcal{Y}}$ be denoted by $\mathcal{Y}_1, \cdots, \mathcal{Y}_n$. An estimate of $N \times N$ covariance matrix of $\boldsymbol{\mathcal{Y}}$ is computed by using the samples as $C_{yy} = (1/(n-1))Y_o Y_o^T$. Here, $Y_o = [\mathcal{Y}_{1o}, \cdots, \mathcal{Y}_{no}]$ represents an $N \times n$ matrix and $\mathcal{Y}_{ko} \equiv \mathcal{Y}_k - \overline{\mathcal{Y}}$, $k = 1, \cdots, n$, with $\overline{\mathcal{Y}}$ being unbiased estimate of the mean vector of $\boldsymbol{\mathcal{Y}}$, i.e., $\overline{\mathcal{Y}} = (1/n)\sum_{k=1}^{n} \mathcal{Y}_k$. Following the KL expansion procedure [Loe78, Chapter XI], [Jol02], let us collect the dominant KL random variable components, $\{\mathbf{z}'_1, \cdots, \mathbf{z}'_M\}$, $M < N$, in an M-D random vector, $\boldsymbol{\mathcal{Z}}' = [\mathbf{z}'_1, \cdots, \mathbf{z}'_M]^T$. Here, $\boldsymbol{\mathcal{Z}}'$ is related to $\boldsymbol{\mathcal{Y}}$ by (2.2).

The set of experimental samples of $\boldsymbol{\mathcal{Z}}'$ can be immediately obtained by replacing $\boldsymbol{\mathcal{Y}}$ with $\mathcal{Y}_1, \cdots, \mathcal{Y}_n$ in (2.2) resulting in $\mathcal{Z}'_1, \cdots, \mathcal{Z}'_n$. To enhance the regularity of the ensuing numerical problem and to improve the efficiency of the associated computation, this data set is further scaled to obtain another data set as shown below,

$$\mathcal{Z}_k = 2\left[(\mathcal{Z}'_k - \underline{a})^{\circ}\left(\frac{1}{\underline{b} - \underline{a}}\right)\right] - \mathbf{1}_M, \quad k = 1, \cdots, n. \tag{3.22}$$

Here, $\mathbf{1}_M$ is an M-D column vector of 1's, $\underline{a} = [\alpha_1, \cdots, \alpha_M]^T$ and $\underline{b} = [\beta_1, \cdots, \beta_M]^T$ with $\alpha_i = \min(z_i'^{(1)}, \cdots, z_i'^{(n)})$ and $\beta_i = \max(z_i'^{(1)}, \cdots, z_i'^{(n)})$, in which $z_i'^{(k)}$ is the i-th component of the k-th sample, $\mathcal{Z}'_k = [z_1'^{(k)}, \cdots, z_M'^{(k)}]$. Denote this M-D normalized KL random vector associated with the samples, $\{\mathcal{Z}_k\}_{k=1}^n$, by $\boldsymbol{\mathcal{Z}} = [\mathbf{z}_1, \cdots, \mathbf{z}_M]^T$. The scaling in (3.22) is particularly chosen so that $\boldsymbol{\mathcal{Z}}$ is supported on $[-1, 1]^M$, which would be in concordance, in some sense, of the support, $[-1, 1]$, of the

uniform distributions used as measures of the PC random variables in section 3.3.4. The following relation between $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{Y}}$ then holds,

$$\boldsymbol{\mathcal{Y}} \approx \boldsymbol{\mathcal{Y}}^{(M)} = V\left[\underline{a} + \left\{(\underline{b} - \underline{a}) \circ \frac{1}{2}(\boldsymbol{\mathcal{Z}} + \mathbf{1}_M)\right\}\right] + \overline{\mathcal{Y}}. \tag{3.23}$$

The approximation sign, '$\approx$', in (3.23) emphasizes that $\boldsymbol{\mathcal{Y}}$ is projected into the space spanned only by the largest M dominant eigenvectors of C_{yy} to obtain the reduced order representation, $\boldsymbol{\mathcal{Z}}$.

Based on $n = 216$ samples, the sample covariance matrix, C_{yy}, is first determined. Here, M is chosen such that $\sum_{i=1}^{M} \varsigma_i = 0.999 \sum_{i=1}^{N} \mathrm{var}(\mathbf{y}_i)$ dictating that $M = 78$ dominant KL random variables should be considered (recall that $N = \dim(\boldsymbol{\mathcal{Y}}) = 121$). Use of the dominant M eigenvectors, along with the samples of $\boldsymbol{\mathcal{Y}}$, in (2.2) yields the set of samples of $\boldsymbol{\mathcal{Z}}'$ which, in turn, yields the samples of $\boldsymbol{\mathcal{Z}}$ through (3.22). At this stage, a crosscheck is performed to ensure that enough information is propagated from $\boldsymbol{\mathcal{Y}}$ to $\boldsymbol{\mathcal{Z}}$ as the dimension is reduced from $N = 121$ to $M = 78$. The samples of $\boldsymbol{\mathcal{Y}}$ are reconstructed back from the samples, $\{\mathcal{Z}_k\}_{k=1}^n$, by using (3.23), i.e., $\mathcal{Y}_k^{(\mathrm{recons})} = V\left[\underline{a} + \left\{(\underline{b} - \underline{a}) \circ \frac{1}{2}(\mathcal{Z}_k + \mathbf{1}_M)\right\}\right] + \overline{\mathcal{Y}}$. The MSE of the relevant statistics computed from $\{\mathcal{Y}_k^{(\mathrm{recons})}\}_{k=1}^n$ are compared with the corresponding statistics computed from the original experimental samples, $\{\mathcal{Y}_k\}_{k=1}^n$. The results are shown in Table 3.1.

Relative MSE in percentage (%) for		
Mean vector	Covariance matrix	SRCC matrix
0	0	0.1220

Table 3.1: Comparison of statistics based on $\{\mathcal{Y}_k^{(\mathrm{recons})}\}_{k=1}^n$ and $\{\mathcal{Y}_k\}_{k=1}^n$: Relative MSE as shown below is computed as $\mathrm{relMSE}(\mathbf{S}^{(\mathrm{recons})}, \mathbf{S}) = 100\left(\|\mathbf{S}^{(\mathrm{recons})} - \mathbf{S}\|_F^2\right) / \|\mathbf{S}\|_F^2$, in which $\mathbf{S}$ represents sample statistic of the experimental samples, $\{\mathcal{Y}_k\}_{k=1}^n$, i.e., $\mathbf{S}$ represents either $\overline{\mathcal{Y}}$ or C_{yy} or $[\rho_s]$ as appropriate, and $\mathbf{S}^{(\mathrm{recons})}$ represents the corresponding sample statistic of the reconstructed samples, $\{\mathcal{Y}_k^{(\mathrm{recons})}\}_{k=1}^n$.

Approach 1 as proposed in section 3.2.1 is now employed to construct the PC representation of $\boldsymbol{\mathcal{Z}}$ based on $n = 216$ experimental samples of $\boldsymbol{\mathcal{Z}}$.

Construction of PC Representation of $\boldsymbol{\mathcal{Z}}$ via Approach 1

In order to gain computational advantage, it is assumed here that 78 random variable components of $\boldsymbol{\mathcal{Z}}$ are pairwise statistically independent; particularly, the mjpdf of $\boldsymbol{\mathcal{Z}}$ has the following form,

$$p_{\boldsymbol{\mathcal{Z}}}(\mathcal{Z}) = p_{\mathbf{z}_1\mathbf{z}_2}(z_1, z_2)p_{\mathbf{z}_3\mathbf{z}_4}(z_3, z_4)\cdots p_{\mathbf{z}_{77}\mathbf{z}_{78}}(z_{77}, z_{78}). \tag{3.24}$$

In the present work, this form is found to be capable of accurately capturing the practically relevant and important information as demonstrated later at the end of this section while discussing the results. Note that the random variable components, $\mathbf{z}_1, \cdots, \mathbf{z}_{78}$, are ordered in the descending order of values of the associated eigenvalues, $\varsigma_1 = \mathrm{var}(\mathbf{z}'_1) > \cdots > \varsigma_{78} = \mathrm{var}(\mathbf{z}'_{78})$, obtained in section 3.3.4 (this, however, does not imply that $\{\mathrm{var}(\mathbf{z}_i)\}_{i=1}^{78}$ is also similarly ordered).

Let us generically indicate any of the pairs in (3.24) by $(\mathbf{z}_l, \mathbf{z}_u)$, $l \in \mathcal{L} \equiv \{1, 3, \cdots, 77\}$ and $u \in \mathcal{U} \equiv \{2, 4, \cdots, 78\}$. For any given $l \in \mathcal{L}$ and $u \in \mathcal{U}$, the target bivariate pdf, $p_{\mathbf{z}_l\mathbf{z}_u}$, is determined by using a normalized histogram of the corresponding experimental samples appropriately collected from $\{\mathcal{Z}_k\}_{k=1}^n$. Each bivariate histogram is estimated with 12×12 bins on equally spaced grids on the support, $s_{\mathbf{z}_l\mathbf{z}_u} \equiv [-1,\, 1]^2$, of $(\mathbf{z}_l, \mathbf{z}_u)$. By using the set of pdfs, $\{p_{\mathbf{z}_l\mathbf{z}_u}\}_{l\in\mathcal{L}, u\in\mathcal{U}}$, PC representations (similar to (3.11)-(3.12)) of all the pairs are constructed. The set of PC random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{78}$, is assumed here to be a set of statistically independent uniform random variables, all of which are supported on $[-1,\, 1]$. For such $\boldsymbol{\xi}_i$'s, the orthogonal polynomials are Legendre polynomials given by,

$$\left.\begin{aligned} \boldsymbol{\Psi}_0(\boldsymbol{\xi}_i) &= 1, \quad \boldsymbol{\Psi}_1(\boldsymbol{\xi}_i) = \boldsymbol{\xi}_i, \\ \boldsymbol{\Psi}_j(\boldsymbol{\xi}_i) &= \tfrac{1}{j}(2j-1)\boldsymbol{\xi}_i\boldsymbol{\Psi}_{j-1}(\boldsymbol{\xi}_i) - \tfrac{1}{j}(j-1)\boldsymbol{\Psi}_{j-2}(\boldsymbol{\xi}_i), \\ &\qquad\qquad\qquad\qquad \text{if } j \geq 2, \end{aligned}\right\} \tag{3.25}$$

and the variance of $\boldsymbol{\Psi}_j(\boldsymbol{\xi}_i)$ is given by,

$$E\left[\boldsymbol{\Psi}_j^2(\boldsymbol{\xi}_i)\right] = \frac{1}{2j+1}. \tag{3.26}$$

While computing the PC coefficients (see (3.8) and (3.13)), the proxy function, $\tilde{q}$, for $f_{l|u} \equiv P_{l|u}^{-1}P_{\boldsymbol{\xi}_l}$ or $f_u \equiv P_u^{-1}P_{\boldsymbol{\xi}_u}$, as appropriate, is based on dividing the support, $s_{\mathbf{z}_l} \equiv [-1,\, 1]$, of $\mathbf{z}_l$ (when approximating $f_{l|u}$) or $s_{\mathbf{z}_u} \equiv [-1,\, 1]$ of $\mathbf{z}_u$ (when approximating f_u), into 199 equal intervals (see Appendix). In

determining the series expansion (similar to (3.9)) of $z_u \longmapsto a_j(z_u)$, the basis functions, ψ_k, are also selected as Legendre polynomials that are orthogonal w.r.t. the weight $w(z) = 1/2$ on $[-1,\, 1]$ implying that the denominator of (3.10) is given by $1/(2k+1)$. The set of PC coefficients of $f_{l|u}(\boldsymbol{\xi}_l \mid z_u)$ is computed for 200 slices, that are equally spaced along the support, $s_{\mathbf{z}_u}$, resulting in $\{z_u^{(k)}, a_j(z_u^k)\}_{k=0}^{199}$ (see Figures 3.5-3.7). The function, $z_u \longmapsto a_j(z_u)$, based on this set is first formed by employing linear interpolation scheme and later used to compute the PC coefficients, b_{jk}. The resulting PC representation of $(\mathbf{z}_l, \mathbf{z}_u)$ given by expressions similar to (3.11)-(3.12) is truncated at $K_1 = K_2 = K = 19, \forall l \in \mathcal{L}$ and $\forall u \in \mathcal{U}$.

Now that PC representations for all the pairs of random variables, $\{(\mathbf{z}_l, \mathbf{z}_u)\}_{l\in\mathcal{L},u\in\mathcal{U}}$, are obtained, a set of $n_{\mathrm{PC}} = 50000$ samples of statistically independent uniform random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{78}$, is generated to test the quality of the constructed PC representations. Use of these samples in PC representations yields a set, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, of 50000 samples of $\boldsymbol{\mathcal{Z}}$. Let estimate of the bivariate pdf of each pair, $(\mathbf{z}_l, \mathbf{z}_u)$, based on 50000 PC samples be denoted by $p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}(z_l, z_u)$. This bivariate pdf is simply determined by employing linear interpolation scheme on a normalized histogram of PC samples with 25×25 bins constructed on equally spaced grids on the support, $s_{\mathbf{z}_l\mathbf{z}_u}$. Introduce the following relative MSE for pdf,

$$\begin{aligned} & \mathrm{relMSE}_p(p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}, p_{\mathbf{z}_l\mathbf{z}_u}) \\ = \;& 100\frac{\displaystyle\int_{[-1,\,1]^2} \left\{p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}(z_l, z_u) - p_{\mathbf{z}_l\mathbf{z}_u}(z_l, z_u)\right\}^2 dz_l\, dz_u}{\displaystyle\int_{[-1,\,1]^2} p^2_{\mathbf{z}_l\mathbf{z}_u}(z_l, z_u)\, dz_l\, dz_u}. \end{aligned}$$

It is found that $\max_{l\in\mathcal{L},u\in\mathcal{U}}[\mathrm{relMSE}_p(p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}, p_{\mathbf{z}_l\mathbf{z}_u})] = 2.4136\%$ and $\min_{l\in\mathcal{L},u\in\mathcal{U}}[\mathrm{relMSE}_p(p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}, p_{\mathbf{z}_l\mathbf{z}_u})] = 0.1217\%$. Bivariate pdf based on 216 experimental samples and 50000 PC realizations are plotted in Figure 3.14 corresponding to $\max_{l\in\mathcal{L},u\in\mathcal{U}}[\mathrm{relMSE}_p(p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}, p_{\mathbf{z}_l\mathbf{z}_u})] = 2.4136\%$. The associated contour plots are shown in Figure 3.15. In Table 3.2, a few practically significant statistics of experimental samples, $\{\mathcal{Z}_k\}_{k=1}^{n}$, and PC samples, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, are compared.

While random variable components, $\{\mathbf{z}_i\}_{i=1}^{78}$, of the normalized KL vector, $\boldsymbol{\mathcal{Z}}$, are uncorrelated by construction resulting in zero off-diagonal elements of the covariance matrix of $\boldsymbol{\mathcal{Z}}$, the SRCC matrix of $\boldsymbol{\mathcal{Z}}$ would be fully populated (since $\{\mathbf{z}_i\}_{i=1}^{78}$ are statistically dependent in the present work). The covariance matrix and SRCC matrix estimated from the experimental samples, $\{\mathcal{Z}_k\}_{k=1}^{n}$, that contain the information

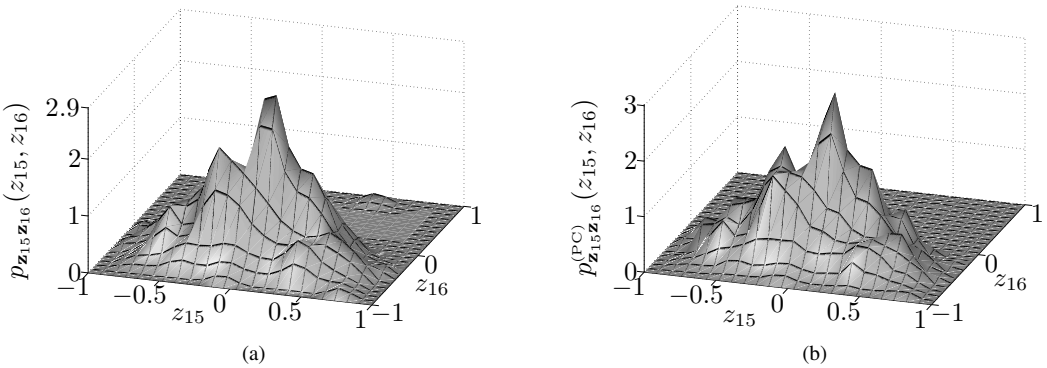


Figure 3.14: Bivariate pdf of $(\mathbf{z}_l, \mathbf{z}_u)$ corresponding to $\max_{l\in\mathcal{L},u\in\mathcal{U}}[\mathrm{relMSE}_p(p^{(\mathrm{PC})}_{\mathbf{z}_l\mathbf{z}_u}, p_{\mathbf{z}_l\mathbf{z}_u})] = 2.4136\%$ based on: (a) 216 experimental samples and (b) 50000 PC samples.

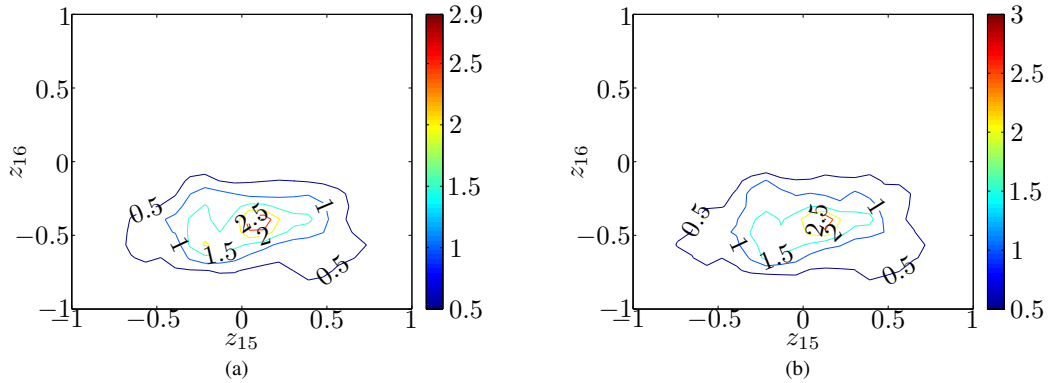


Figure 3.15: Contour plots associated with the bivariate pdfs shown in Figure 3.14: (a) based 216 experimental samples and (b) based on 50000 PC samples.

Relative MSE in percentage (%) for		
Mean vector	Covariance matrix	SRCC matrix
0.1103	0.1581	6.5934

Table 3.2: Comparison of statistics based on $\{\mathcal{Z}_k\}_{k=1}^{n}$ and $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$: Relative MSE as shown below is computed as $\mathrm{relMSE}(\mathbf{S}^{(\mathrm{PC})}, \mathbf{S}) = 100\left(\|\mathbf{S}^{(\mathrm{PC})} - \mathbf{S}\|_F^2\right)/\|\mathbf{S}\|_F^2$, in which $\mathbf{S}$ represents the appropriate sample statistic of experimental samples, $\{\mathcal{Z}_k\}_{k=1}^{n}$, and $\mathbf{S}^{(\mathrm{PC})}$ represents the corresponding sample statistic of PC realizations, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$.

about this statistical dependency indeed display the respective characteristics. The effect of assumption of pairwise statistical independence in (3.24), therefore, can be assessed, in some sense, by deviation of the SRCC matrix estimated from PC samples, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, that cannot capture the effect of statistical

dependency among the pairs, $\{\mathbf{z}_l, \mathbf{z}_u\}_{l\in\mathcal{L},u\in\mathcal{U}}$, from the SRCC matrix estimated from experimental samples, $\{\mathcal{Z}_k\}_{k=1}^n$. The value of relative MSE for SRCC matrix as shown in the third column of Table 3.2 implies that assumption of pairwise statistical independence might be practically acceptable.

Finally, the set, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, are used to obtain the set, $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, of PC samples of $\boldsymbol{\mathcal{Y}}$ by having recourse to (3.23). The statistics of the resulting samples, $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, are compared to that of the experimental samples, $\{\mathcal{Y}_k\}_{k=1}^n$, and the results are shown in Table 3.3.

Relative MSE in percentage (%) for		
Mean vector	Covariance matrix	SRCC matrix
2.8525	0.0124	6.3700

Table 3.3: Comparison of statistics based on $\{\mathcal{Y}_k\}_{k=1}^n$ and $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$: Relative MSE as shown below is computed as $\mathrm{relMSE}(\mathbf{S}^{(\mathrm{PC})},\mathbf{S}) = 100\left(\|\mathbf{S}^{(\mathrm{PC})}-\mathbf{S}\|_F^2\right)/\|\mathbf{S}\|_F^2$, in which $\mathbf{S}$ represents the sample statistic of experimental samples, $\{\mathcal{Y}_k\}_{k=1}^n$, i.e., either $\overline{\mathcal{Y}}$ or C_{yy} or $[\rho_s]$ as appropriate and $\mathbf{S}^{(\mathrm{PC})}$ represents the corresponding sample statistic of PC realizations, $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$.

In the next section, modeling of $\boldsymbol{\mathcal{Y}}$ via Approach 2 is considered.

3.3.5 Modeling of $\boldsymbol{\mathcal{Y}}$ via Approach 2

The experimental samples, $\{\mathcal{Y}_k\}_{k=1}^n$, of $\boldsymbol{\mathcal{Y}}$ as obtained in section 3.3.3 are used here again to deduce PC representation of $\boldsymbol{\mathcal{Y}}$ by employing Approach 2. In this case, application of KL decomposition in order to obtain a reduced order representation of $\boldsymbol{\mathcal{Y}}$ is not plausible since statistical dependency here would be characterized by SRCC not by PCC. However, for the sake of improved efficiency and regularity of the following numerical task, the samples of $\boldsymbol{\mathcal{Y}}$ is scaled to obtain a set of samples of another N-D random vector $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N]^T$ supported on $[-1,\ 1]^N$ by employing a transformation similar to (3.22). In this case, $\boldsymbol{\mathcal{Y}}$ is related to $\boldsymbol{\mathcal{Z}}$ by,

$$\boldsymbol{\mathcal{Y}} = \underline{a} + \left[(\underline{b}-\underline{a})\circ\frac{1}{2}(\boldsymbol{\mathcal{Z}}+\mathbf{1}_N)\right], \tag{3.27}$$

and the experimental samples, $\{\mathcal{Z}_k\}_{k=1}^n$, of $\boldsymbol{\mathcal{Z}}$ follow from,

$$\mathcal{Z}_k = 2\left[(\mathcal{Y}_k-\underline{a})\circ\left(\frac{1}{\underline{b}-\underline{a}}\right)\right]-\mathbf{1}_N,\quad k=1,\cdots,n. \tag{3.28}$$

In (3.27) and (3.28), unlike Approach 1, $\underline{a}$ and $\underline{b}$ are now given by $\underline{a} = [\alpha_1, \cdots, \alpha_N]^T$ and $\underline{b} = [\beta_1, \cdots, \beta_N]^T$ with $\alpha_i = \min(y_i^{(1)}, \cdots, y_i^{(n)})$ and $\beta_i = \max(y_i^{(1)}, \cdots, y_i^{(n)})$, in which $y_i^{(k)}$ is the i-th component, $i = 1, \cdots, N$, of the k-th sample, $\mathcal{Y}_k = [y_1^{(k)}, \cdots, y_N^{(k)}]$.

The normalized marginal histogram of each random variable component, $\mathbf{z}_i$, $i \in \mathcal{I} = \{1, 2, \cdots, 121\}$ (recall $N = 121$), is constructed based on corresponding $n = 216$ experimental samples appropriately collected from $\{\mathcal{Z}_k\}_{k=1}^n$. Marginal histogram is based on 12 equal-sized bins on the support, $s_{\mathbf{z}_i} \equiv [-1, 1]$, of $\mathbf{z}_i$. Similar to previous approach, subsequent use of 1-D linear interpolation scheme on this normalized histogram results in an estimate of the target marPDF of $\mathbf{z}_i$ to be denoted by $P_{\mathbf{z}_i}$. Based on this $P_{\mathbf{z}_i}$, PC representation of each $\mathbf{z}_i$ (see (3.15)) is determined. In constructing these PC representations, orthogonal polynomials are again chosen as Legendre polynomials, given by (3.25), in terms of a set of uniform random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{121}$, each of which is supported on $[-1, 1]$. In computing the corresponding PC coefficients, the approximate function, $\tilde{q}_i$, to be used in lieu of $q_i \equiv P_{\mathbf{z}_i}^{-1} P_{\boldsymbol{\xi}_i}$ in (3.16) is based on dividing $s_{\mathbf{z}_i}$ into 199 equal intervals (see Appendix). The resulting PC representation, $\mathbf{z}_i \overset{d}{=} \lim_{K_i \to \infty} \sum_{j=0}^{K_i} c_{ji}\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_i)$, is truncated at $K_i = 14$, $\forall i \in \mathcal{I}$.

Now, in order to digitally generate realizations of $\boldsymbol{\mathcal{Z}}$ (and $\boldsymbol{\mathcal{Y}}$), a set of $n_{\rm PC} = 50000$ samples of random vector, $\underline{\boldsymbol{\xi}} = [\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{121}]^T$, is simulated first as follows. Unlike Approach 1, the random variables, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{121}$, here are statistically dependent. The statistical dependency of $\underline{\boldsymbol{\xi}}$ is characterized by the SRCC matrix estimated based on experimental samples, $\{\mathcal{Z}_k\}_{k=1}^{216}$, of $\boldsymbol{\mathcal{Z}}$. Application of the mapping defined by (3.17) on the resulting sample SRCC matrix of $\boldsymbol{\mathcal{Z}}$, however, yields a non-positive-definite matrix, $[\rho^{(1)}]$, rendering the normal copula technique inapplicable. Samples of the associated Gaussian random vector, $\mathbf{X}$, consisting of correlated standard normal random variables, $\mathbf{x}_1, \cdots, \mathbf{x}_{121}$, therefore, need to be generated by using the augmented normal copula technique as highlighted in section 3.2.2. The constrained optimization problem defined by (3.20) is solved to determine the feasible positive-definite covariance (or correlation) matrix, $[\rho]$, of $\mathbf{X}$. It is found that $\mathrm{relMSE}([\rho], [\rho^{(1)}]) = 0.0006\%$ and $\mathrm{relMSE}([\rho_s], [\rho_s^{(1)}]) = 0.0006\%$, in which $[\rho_s^{(1)}]$ is the sample (positive-definite) SRCC matrix estimated based on $\{\mathcal{Z}_k\}_{k=1}^{216}$ and $[\rho_s]$ is (again) a positive-definite matrix resulting from the application of inverse mapping of (3.17) on $[\rho]$, i.e., (i,j)-th, $i, j = 1, \cdots, 121$, element of $[\rho_s]$ is obtained as $(\rho_s)_{ij} = (6/\pi) \arcsin(\rho_{ij}/2)$. Then, 50000 samples of $\underline{\boldsymbol{\xi}}$ consisting of statistically dependent uniform random variables, $\{\boldsymbol{\xi}_i\}_{i=1}^{121}$, supported on $[-1, 1]^{121}$, with its SRCC or PCC matrix being given by $[\rho_s]$, can be readily generated by using the augmented normal copula technique. Use of these samples in

the constructed PC representations for $\{\mathbf{z}_i\}_{i\in\mathcal{I}}$ yields a set, $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, of 50000 samples of $\boldsymbol{\mathcal{Z}}$, and subsequently, the set, $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$, of samples of $\boldsymbol{\mathcal{Y}}$ follows from (3.27).

Let the estimate of marpdf of $\mathbf{z}_i$, $i \in \mathcal{I}$, be denoted by $p_{\mathbf{z}_i}^{(\mathrm{PC})}$ that is again determined as an approximation of the corresponding marginal normalized linearly interpolated histogram. The histogram is based on 25 equal-sized bins on the corresponding support, $s_{\mathbf{z}_i}$. A comparison between two marpdfs based on 50000 PC realizations and 216 experimental samples is shown in Figure 3.16 for $\mathbf{z}_i$ corresponding to $\max_{i\in\mathcal{I}}[\mathrm{relMSE}_p(p_{\mathbf{z}_i}^{(\mathrm{PC})}, p_{\mathbf{z}_i})] = 2.1833\%$, in which $\mathrm{relMSE}_p(p_{\mathbf{z}_i}^{(\mathrm{PC})}, p_{\mathbf{z}_i})$ is now defined by,

$$\mathrm{relMSE}_p(p_{\mathbf{z}_i}^{(\mathrm{PC})}, p_{\mathbf{z}_i}) = 100\frac{\displaystyle\int_{s_{\mathbf{z}_i}} \left\{p_{\mathbf{z}_i}^{(\mathrm{PC})}(z_i) - p_{\mathbf{z}_i}(z_i)\right\}^2 dz_i}{\displaystyle\int_{s_{\mathbf{z}_i}} p_{\mathbf{z}_i}^2(z_i)\, dz_i},$$

with $p_{\mathbf{z}_i}$ being the marpdf based on 216 experimental samples of $\mathbf{z}_i$. Let us also report the minimum value of relMSE_p, $\min_{i\in\mathcal{I}}[\mathrm{relMSE}_p(p_{\mathbf{z}_i}^{(\mathrm{PC})}, p_{\mathbf{z}_i})] = 0.0729\%$.

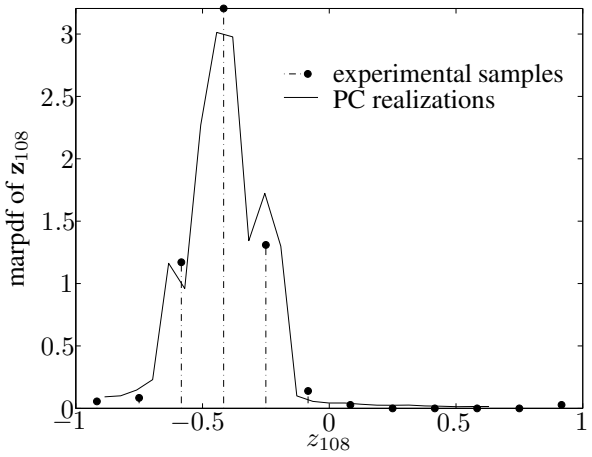


Figure 3.16: Marginal pdf of $\mathbf{z}_i$ corresponding to $\max_{i\in\mathcal{I}}[\mathrm{relMSE}_p(p_{\mathbf{z}_i}^{(\mathrm{PC})}, p_{\mathbf{z}_i})] = 2.1833\%$.

Finally, summaries of practically significant statistics based on PC realizations are compared with that based on experimental samples for $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{Y}}$, respectively, in Table 3.4 and Table 3.5. It must remarked here that while covariance matrix is not used as a measure of statistical dependency in Approach 2, the corresponding results are still shown in the second columns of these tables for the sake of probing (if any) by the inquisitive readers.

Relative MSE in percentage (%) for		
Mean vector	Covariance matrix	SRCC matrix
0.0339	5.4139	0.0040

Table 3.4: Comparison of statistics based on $\{\mathcal{Z}_k\}_{k=1}^{n}$ and $\{\mathcal{Z}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$ (see caption of Table 3.2 for further explanation).

Relative MSE in percentage (%) for		
Mean vector	Covariance matrix	SRCC matrix
2.5569	1.3123	0.0040

Table 3.5: Comparison of statistics based on $\{\mathcal{Y}_k\}_{k=1}^{n}$ and $\{\mathcal{Y}_k^{(\mathrm{PC})}\}_{k=1}^{n_{\mathrm{PC}}}$ (see caption of Table 3.3 for further explanation).

3.3.6 Reconstructing the Original Random Temperature Field

Construct the PC representation of $\boldsymbol{\mathcal{Z}}$ either by using Approach 1 or Approach 2 as appropriate. The PC coefficients of the random variable components of $\boldsymbol{\mathcal{Z}}$ and those of $\boldsymbol{\mathcal{Y}}$ are related by linear mappings as can be readily verified by using (3.23) and (3.27) (see chapter 2 for further details). Since the set of 11×11 random variables, $\{\mathbf{\Gamma}^{(\mathrm{n})}(t,h)\}_{(t,h)\in(T\times D)}$, constitute $\boldsymbol{\mathcal{Y}}$, PC coefficients of $\mathbf{\Gamma}(t,h)$, $(t,h) \in (T \times D)$, immediately follow from the PC coefficients of $\boldsymbol{\mathcal{Y}}$ by using the relation, $\mathbf{\Gamma}(t,h) = \overline{\Gamma}(t,h)\mathbf{\Gamma}^{(\mathrm{n})}(t,h) + \overline{\Gamma}(t,h)$. Inference of PC coefficients of the original random process, when $(t,h) \notin (T \times D)$, from those of $\mathbf{\Gamma}(t,h)$, $(t,h) \in (T \times D)$, is essentially a task of interpolation or/and approximation technique as shown in numerous other occasions in the present work. Digital generation of realizations of the original random process similarly needs no further explanation.

3.4 Conclusion

Two approaches for constructing PC representation of a non-Gaussian and second-order random vector, $\boldsymbol{\mathcal{Y}}$, is presented by only using the experimental measurements. The random vector, $\boldsymbol{\mathcal{Y}}$, can be viewed as a finite-dimensional representation of a non-stationary, non-Gaussian and second-order random field evolving over space or/and time. The experimental data is measured on a finite countable subset of the space-time indexing set of the random field. In many practical applications, e.g., prediction of acoustic field involving oceanographic parameters as indicated in the previous section, use of spatio-temporal

random field would be a more appropriate model for characterizing the inherent uncertainty in system parameters of a stochastic system. The PC representation of the random field representing such random system parameters has been proven to be an efficient tool in systematically propagating the uncertainty to the model-based predictions of response of the stochastic system.

Approach 1 attempts to capture the complete information of a target mjpdf of $\boldsymbol{\mathcal{Y}}$. This approach uses the knowledge of a complete set of properly ordered target conditional PDFs estimated from the experimental measurements and the concept of Rosenblatt transformation. The set of target conditional PDFs, that uniquely defines the target mjpdf of $\boldsymbol{\mathcal{Y}}$, are approximations, based on linear interpolation, of the corresponding set of normalized histograms of the appropriate set of experimental samples. Approach 2, on the other hand, satisfies the target marPDFs and the target SRCC matrix of $\boldsymbol{\mathcal{Y}}$. The set of target marPDFs and the target SRCC matrix are similarly estimated by using the experimental samples. The second approach is also founded on the Rosenblatt transformation. In both the approaches, appropriate functions based on the Rosenblatt transformation are first defined in terms of the selected PC variables, $\boldsymbol{\xi}_k$'s. The defined functions are equal to $\boldsymbol{\mathcal{Y}}$ in the sense of distribution. Subsequently, construction of PC expansion of these functions results in appropriate PC representations that can be readily employed within the PC framework to propagate the associated uncertainty. It should, however, be realized that the existences and (if they exist) the true forms of these functions are never known in reality. Nevertheless, the proposed approaches guarantee [HLD04, Theorem 2.1] that such functions can always be constructed. For efficient and fast computation of the PC coefficients, the Rosenblatt transformation based functions are further substituted by the appropriate interpolated functions.

One important distinction between two proposed approaches is that while PC random variables, $\boldsymbol{\xi}_k$'s, are statistically independent in Approach 1, the corresponding set of PC random variables are statistically dependent in Approach 2. Approach 1 is computationally expensive relative to Approach 2. Additional model reduction technique, e.g., use of KL decomposition as discussed in the context of numerical illustration in section 3.3.4, is recommended to reduce the computational cost. Further probabilistic assumptions as made in section 3.3.4, while modeling the spatio-temporal random temperature field, would also alleviate the computational burden at the expense of accuracy. But it is certainly recommended if the achieved accuracy is practically acceptable. Accuracy level of Approach 2 is expected to be higher than the accuracy level of Approach 1 if additional probabilistic assumptions and model reduction scheme, as just indicated, are incorporated into Approach 1. Approach 2 would also be computationally cheaper

manyfold for low and moderate dimension of $\mathcal{Y}$ (quantification of the qualifiers, 'low' and 'moderate', however, directly depends on the available computational resources).

Chapter 4

Hybrid Representations of Coupled Nonparametric and Parametric Models

Parametric modeling of stochastic systems has proven useful for systems with well-defined and well structured sources of uncertainty. The suitability of such models is usually indicated by small levels of uncertainty associated with their parameters. The parametric model may not efficiently employed to deal with problems whose level of uncertainty is high, involving spatially distributed sources of uncertainty. The class of so-called nonparametric stochastic models has recently been introduced in mechanics to address this specific issue and found to be useful. This chapter presents a coupling technique, adapted to the receptance frequency response function (FRF) matrix, for combining these two approaches. This will be useful for the analysis of complex dynamical systems having spatially non-homogeneous uncertainty that is otherwise difficult to analyze. The existing nonparametric approach was, till date, applied to the positive definite/semi-definite system matrices, for example, to the matrices of mass, damping and stiffness. In the current work, the nonparametric approach is also employed to the complex symmetric receptance FRF matrix, now acting as the system matrix, by having recourse to the Takagi's factorization.

4.1 Introduction and Motivation

Two types of uncertainty are of particular interest in connection to the dynamical systems: *modeling uncertainty* and *data uncertainty*. Modeling uncertainty can be further decomposed into *mechanical uncertainty* and *probabilistic uncertainty*. While mechanical uncertainty results from several simplified assumptions in developing a mechanical/mathematical model, referred further here as predictive model, of the physical phenomena, the latter one stems from the introduction of probabilistic assumptions associated with the statistical/probabilistic characteristics of the random system/model parameters (geometry, boundary conditions, parameters of the constitutive equation etc.) of the predictive model. Examples

in which mechanical uncertainty is present include simplified mechanical models of a complex junction, one or two dimensional beam and plate models instead of their three-dimensional (3D) elasticity models. Even the 3D theory of elasticity encompasses many assumptions introduced for mathematical convenience that might not satisfy the 'true' behavior of a complex system. The assumption of statistical independence between two random system parameters and assigning a particular probability distribution law to a random system parameter are the instances of injecting probabilistic uncertainty. Data uncertainty, on the other hand, is characterized by the uncertainty associated with the data collected from experimental measurements for estimation of the statistical/probabilistic features of the model parameters. The *data limitation* (because of finite sample size) and *experimental uncertainty* (caused due to, for example, imperfect set-up and condition of the experiment, human error and environmental condition) introduce data uncertainty.

In addition, the models of certain parts of a complex system are more accurate than those in the other parts; for instance, the mechanical model of a part constituted of a slender beam structure is generally better than a simplified mechanical model of a complex joint. This indicates that the uncertainties resulting from the mechanical model are not homogeneous throughout the system. Uncertainties resulting from the probabilistic uncertainty are also not spatially homogeneous because, e.g., some of the parameters of a certain part of the complex system might be truly statistically independent whereas the assumption of statistical independence for other parameters may be a mere mathematical convenience. Uncertainties resulting from the data uncertainty are again not spatially homogeneous because the data may not be uniformly available throughout the complex system. Clearly, in general, uncertainties in a complex system are expected to be spatially non-homogeneous.

The analysis of a complex dynamical system with such non-homogeneous uncertainties is typically quite involved both to set-up as well as to numerically resolve. It may not even be possible to analyze the built-up structure because of the presence of the spatially non-homogeneous uncertainties. In order to model all the uncertainties in such a complex dynamical systems and to solve the global stochastic equations, it might be useful and convenient to decompose the system into several smaller subsystems such that uncertainty in each subsystem is spatially homogeneous. Each of these subsystems can be analyzed separately using the method most suitable for it, and finally can be assembled to obtain the response of the built-up system. A subsystem having a lower level of modeling and data uncertainty

or/and having a relatively fewer number of random system parameters can be analyzed by using the parametric approach requiring knowledge of the *local* system parameters (for example, Young's modulus, shear modulus, bulk modulus, Poisson's ratio etc.). On the other hand, a subsystem having a higher level of uncertainties due to modeling uncertainty and data uncertainty or/and having a large number of random system parameters can be analyzed by using the recently proposed approach, called nonparametric approach [Soi00, Soi01a, Soi05a, Soi05b, CLPP$^+$07], that does not require the knowledge of the local system parameters.

The objective of the work presented in this chapter is to propose a hybrid approach that permits the coupling between subsystems having been analyzed by using these two different approaches, in order to determine quantities of interest of the built-up structure. The approach is based on the point-wise enforcement of dynamic equilibrium condition for each realization of the stochastic system.

4.2 Nonparametric Model

In this section, an overview of Soize's pioneering work [Soi00, Soi01a], that proposed the nonparametric approach to model uncertainties in dynamical systems in the low-frequency regime, is provided. This model differs from the parametric modeling of uncertainties in that it does not require information of the local parameters of the system being analyzed. The nonparametric framework allows one to construct a probability density function (pdf) of the random system matrices directly based on partial knowledge of the systems. It is directed towards constructing the probability models of the random system matrices without consideration of the parametrization of the dynamical model of the system. The work considers the problem of constructing the associated probability space, namely, $\big(\mathbb{M}_n^S(\mathbb{R}), \mathcal{F}, P_{\mathbf{A}}\big)$ in which $\mathbb{M}_n^S(\mathbb{R})$ is the set of all real $n \times n$ symmetric matrices, $\mathcal{F}$ is the σ-algebra of subsets of $\mathbb{M}_n^S(\mathbb{R})$ and $P_{\mathbf{A}}$ is the probability measure on $\mathcal{F}$ such that the support of the random system matrix variate is the set of symmetric positive definite real matrices denoted by $\mathbb{M}_n^+(\mathbb{R}) \subset \mathbb{M}_n^S(\mathbb{R})$. It implies that $\text{supp}(\mathbf{A}) = \{A : p_{\mathbf{A}}(A) > 0\} = \mathbb{M}_n^+(\mathbb{R})$, where $\mathbf{A}$ is the random matrix variate, $p_{\mathbf{A}}$ is its pdf, $p_{\mathbf{A}} : \mathbb{M}_n^S(\mathbb{R}) \longmapsto \mathbb{R}^+ = [0, \infty)$, which can be related to the probability measure, $P_{\mathbf{A}}$, as follows,

$$dP_{\mathbf{A}}(A) = p_{\mathbf{A}}(A)\, \tilde{d}A.$$

Here $\tilde{d}A$ can be interpreted as a volume element in $\mathbb{M}_n^S(\mathbb{R})$ and is defined to be the wedge product or the exterior product [Fla89] of the independent elements of the differential form of the matrix dA and the (i,j)th element of dA is simply defined [For08] as dA_{ij} in which A_{ij} is the (i,j)th element of A. There are $n(n+1)/2$ such independent elements. In the context of random matrix theory [Meh04, For08], for symmetric matrix, $\tilde{d}A$ is then given by $\bigwedge_{1\leq i\leq j\leq n} dA_{ij}$, where $\bigwedge$ indicates the wedge product. In the present context, the wedge product reverts to be the natural volume element, $\prod_{1\leq i\leq j\leq n} dA_{ij}$, in the Euclidean space $\mathbb{R}^{n(n+1)/2}$ that is topologically equivalent to $\mathbb{M}_n^S(\mathbb{R})$. Soize, however, considers an Euclidean structure on $\mathbb{M}_n^S(\mathbb{R})$ to define the following volume element,

$$\tilde{d}A = 2^{n(n-1)/4} \prod_{1\leq i\leq j\leq n} dA_{ij} \ ,$$

that differs by a multiplication factor, $2^{n(n-1)/4}$, from the natural volume element. It should be noted that it is, however, feasible to reformulate the theory behind the nonparametric approach by using the natural volume element, $\prod_{1\leq i\leq j\leq n} dA_{ij}$, with minor changes (please see section 5.3 for further details).

In constructing the probability measure, the principle of maximum entropy [KK92], as initially introduced by Jaynes [Jay57a, Jay57b] for discrete random variables, has been used [Soi00]. The MaxEnt principle yields a constrained optimization problem with the objective being to maximize Shannon's measure of entropy [Sha48] constrained by the given statistics (mean, variance etc.) of the random variate representing the available information. In the present case, entropy can be interpreted as a measure of relative uncertainty [KK92] associated with the probability distribution of the random matrix variate. This uncertainty is not about which realization of the random matrix variate will be observed but it represents the uncertainty of the probability distribution of the random matrix variate. The basic idea of the MaxEnt principle is to choose the probability distribution with maximum uncertainty out of all the probability distributions that are consistent with the given set of constraints. Any other probability distribution would be associated with unwarranted assumptions about the systems which are not available [Jay57a, Jay57b].

It should be noted that the uniform distribution is often considered to represent a state of maximum uncertainty. Interestingly, the probability distribution resulting from the application of Jaynes's maximum-entropy (MaxEnt) principle on the continuous pdf [KK92, Section 2.5.2] is the same as the probability distribution resulting from the use of Kullback-Leibler's [KL51, Kul59] principle of minimum directed divergence (minimum cross-entropy) provided the prior probability distribution is uniform

[SJ80, SJ83, Jay68]. Out of all the probability distributions satisfying the given constraints, the principle of minimum directed divergence chooses the one that is closest to the uniform distribution. Geometrically, it corresponds to the probability distribution, in a space consisting of probability distributions (a point in this space represents a probability distribution), that has minimum *directed* distance, computed by using the Kullback-Leibler's measure [KL51, Kul59], with respect to the point in this space representing the uniform distribution.

Let us denote the mass matrix of the mean dynamical system by $\underline{M} \in \mathbb{M}_n^+(\mathbb{R})$ and its Cholesky decomposition [Har97, Theorem $14.5.11$] by,

$$\underline{M} = L_M^T L_M, \tag{4.1}$$

in which L_M is an upper triangular matrix in the set, $\mathbb{M}_n(\mathbb{R})$, of all real matrices of size $n \times n$, where n is the total degree of freedoms (dof) of the mean finite element model (FEM). The mean stiffness matrix $\underline{K}$ is positive definite for a fixed structure and positive semi-definite for free structure and consequently has the following Cholesky decomposition [Har97, Theorem $14.5.16$],

$$\underline{K} = S_K^T S_K,$$

in which S_K is an upper triangular matrix in $\mathbb{M}_n(\mathbb{R})$ with a fixed structure and is an almost upper triangular matrix in $\mathbb{M}_{m,n}(\mathbb{R})$ for free structure with $(n-m) \leq 6$ being the number of rigid body modes of the system; where $\mathbb{M}_{m,n}(\mathbb{R})$ is the set of all real matrices of size $m \times n$. The random system mass matrix $\mathbf{M}$ and random system stiffness matrix $\mathbf{K}$ are then written as [Soi99],

$$\mathbf{M} = L_M^T \mathbf{G}_M L_M, \quad \text{and} \quad \mathbf{K} = S_K^T \mathbf{G}_K S_K,$$

in which $\mathbf{G}_M$ and $\mathbf{G}_K$ are second order (see below (4.4)) random matrix variate, respectively, in $\mathbb{M}_n^+(\mathbb{R})$ and $\mathbb{M}_m^+(\mathbb{R})$ with $E\{\mathbf{G}_M\} = I_n$ and $E\{\mathbf{G}_K\} = I_m$, where E is the mathematical expectation operator and I_p is an identity matrix of size $p \times p$, $p = n, m$. A similar decomposition exists for random system damping matrix.

Next step is to generate the ensembles of the random matrices $\mathbf{G}_M$ and $\mathbf{G}_K$ in order to simulate the realizations of $\mathbf{M}$ and $\mathbf{K}$. Let us denote the random matrix to be generated ($\mathbf{G}_M$, $\mathbf{G}_K$ or the corresponding

matrix for damping) by $\mathbf{A} \in \mathbf{M}_n^+(\mathbb{R})$ generically. The pdf, $p_{\mathbf{A}}$, is then determined by using the principle of maximum entropy. The nonparametric approach as proposed by Soize only assumes the information of the ensemble means of the system matrices (mass, stiffness and damping matrices) to be known *a priori*. These ensemble means can be taken as the system matrices obtained by discretizing a nominal continuous system in view of analyzing it using the finite element method (FEM). Subsequent uses of

1. **(normalization constraint)** the axioms of probability, specifically that the total probability must be unity,

2. **(ensemble mean constraint)** the given ensemble mean matrix, $\underline{A}$, that is given by the matrix corresponding to the mean system, and

3. **(existence of moments of response)** the existence of the moments of the response random variables, that is expressed in terms of the existence of the moments of the random system matrices,

as constraints in the MaxEnt principle yields the pdf, $p_{\mathbf{A}}$. Here, the last constraint also implies that [Soi01a, p. 1985],

$$E\left\{\|\mathbf{A}^{-1}\|_F^{\gamma}\right\} < \infty, \quad \text{for mass, damping and stiffness matrices,}$$

where $\gamma \geq 1$ is a positive integer and $\|\cdot\|_F$ is the Frobenius norm defined by $\|A\|_F = \langle A, A^T \rangle^{1/2} \equiv \left\{\mathrm{tr}\left(A\,A^T\right)\right\}^{1/2} = \left(\sum_{ij} |\,a_{ij}\,|^2\right)^{1/2}$, in which a_{ij} is the (i,j)-th element of A. The existence of these moments is required in order to guarantee the existence of moments of the response that is obtained by solving a dynamical system, for example, $\mathbf{AX} = \mathbf{F} \Rightarrow \mathbf{X} = \mathbf{A}^{-1}\mathbf{F}$, in which $\mathbf{X}$ and $\mathbf{F}$, respectively, represent the response of and the external disturbance on the stochastic system represented by the random matrix operator $\mathbf{A}$. Subsequently, the pdf, $p_{\mathbf{A}}$, can be determined [Soi00, Soi01a] by maximizing the entropy of $p_{\mathbf{A}}$ subject to the above constraints.

The pdf, $p_{\mathbf{A}}$, thus obtained is found to be characterized by three parameters, $\underline{A}$, λ and n. Here, $(1-\lambda)$ is one of the Lagrange multipliers, with $\lambda > 0$, associated with the last constraint. When $\underline{A}$ is identity matrix (as is the case for $\mathbf{G}_M$ and $\mathbf{G}_K$), λ is given by [Soi01a],

$$\lambda = \frac{(1-\delta_A^2)}{2\delta_A^2} n + \frac{1+\delta_A^2}{2\delta_A^2}, \tag{4.2}$$

where $\delta_A > 0$ is the dispersion parameter defined by,

$$\delta_A = \left\{ \frac{E\left\{ \|\mathbf{A} - \underline{A}\|_F^2 \right\}}{\|\underline{A}\|_F^2} \right\}^{1/2}. \tag{4.3}$$

From the convergence study and existence of the second order moment of the inverse random matrix (guaranteing the existence of moments of the response random variables), δ_A must satisfy the following relation [Soi01a],

$$0 < \delta_A < \sqrt{\frac{n_0 + 1}{n_0 + 5}}, \quad \forall\, n \geq n_0, \tag{4.4}$$

in which n_0 is an integer. This condition then guarantees the existence of the mean and the second order moment of $\mathbf{A}^{-1}$ (as required for the random matrices associated with mass, damping and stiffness). The upper bound of δ_A in (4.4) is a monotonically and strictly increasing function of n_0 with $\min \sqrt{\frac{n_0+1}{n_0+5}} = 0.5774$ for $n_0 = 1$ and $\sup \sqrt{\frac{n_0+1}{n_0+5}} = 1$. In the context of practical problems, $n \geq 1$, and δ_A typically satisfies the relation expressed by (4.4) implying $\lambda \gg 1$ by (4.2).

The PDF associated with MaxEnt pdf is the Wishart distribution or matrix-variate gamma distribution, $W\left(\frac{1}{n-1+2\lambda}\underline{A}, (n-1+2\lambda)\right)$ (see, e.g., [Mur82, Section 3.2], [GN00, Chapter 3] [And03, Section 7.2]). Procedures for generations of realizations of random matrix $\mathbf{A}$ are well documented in the literature [Mur82, Theorem 3.2.5], [GN00, Theorem 3.3.1 or Theorem 3.3.11] and used extensively by Soize [Soi00, Soi01a] to describe the simulation technique for sampling from the Wishart distribution. A quick summary, as to how Monte Carlo simulation (MCS) technique is employed to generate the realizations of $\mathbf{A}$, is provided below. (See pp. 116–117 for further theoretical and algorithmic details).

4.2.1 Monte Carlo Simulation of A

For many applications, n is sufficiently large and in such cases, there exists a simple form of the random matrix, $\mathbf{A}$, given by,

$$\mathbf{A} = \frac{1}{m_A} \sum_{j=1}^{m_A} \left(L_A^T \mathbf{U}_j\right) \left(L_A^T \mathbf{U}_j\right)^T, \tag{4.5}$$

Here, L_A is defined by the Cholesky decomposition of $\underline{A}$, i.e., $\underline{A} = L_A^T L_A$, $\mathbf{U}_j$'s are independent and identically distributed (i.i.d) $\mathbb{R}^n$-valued normal random vector, i.e., $N(\mathbf{0}, I_n)$, and finally, $m_A = (n+1)/\delta_A^2$. This form is more amenable to the practical calculation for the purpose of MCS of the random matrix, $\mathbf{A}$.

The form defined by (4.5) is useful when λ is an integer implying that m_A is also an integer. For high n, m_A can be rounded off to the nearest integer without causing significant limitation in the nonparametric model. However, this introduces probabilistic uncertainty in the model. To avoid introducing this probabilistic uncertainty (though very small), the exact simulation technique can be used if m_A is not integer. When m_A is not an integer, simulation of $\mathbf{A}$ involves simulation of Gamma random variables and Gaussian random variables (see Algorithm 5.3.3 in pp. 116).

It should be noted here that the above nonparametric formulation is developed by following a similar procedure for constructing a probability space for Gaussian orthogonal ensemble (GOE) [Meh04]. However, unlike the statistically independent elements of matrix belonging to GOE, the elements $\mathbf{A}_{ij}$ of the random matrix $\mathbf{A} \in \mathbb{M}_n^+(\mathbb{R})$ are not statistically independent. A comparative study by using the ensemble of the matrix computed in the framework of the nonparametric approach and the matrix of *GOE* has been conducted to show the superiority of the nonparametric approach over the *GOE* approach in the context of structural dynamics problems [Soi03]. This work also compares the nonparametric approach and the parametric approach to validate the nonparametric technique in the low-frequency range. The nonparametric approach has been applied both to problems of frequency domain [Soi00, Soi03] and time domain [Soi01a, Soi01b].

This section is concluded by noting that given the partial information separately for each of the system matrices with *no* information of the statistical dependency among these system matrices, the MaxEnt principle also implies that these system matrices are statistically independent of each other, i.e., $p_{\mathbf{M},\mathbf{C},\mathbf{D},\mathbf{K}}(M, C, D,\ K) = p_{\mathbf{M}}(M)\, p_{\mathbf{C}}(C)\, p_{\mathbf{D}}(D)\, p_{\mathbf{K}}(K)$ [Soi00]. Here $\mathbf{C}$ and $\mathbf{D}$, respectively, represent the matrix of viscous damping and structural damping and C and D are their realizations. The resulting fact of this independence is discernible because no knowledge of the statistical dependency of the random system matrix variate, $\mathbf{M}, \mathbf{C}, \mathbf{D}, \mathbf{K}$, has been used in the development. Only knowledge of the ensemble means of these matrices are separately incorporated in the formulation while solving a dynamical system. In section 4.3, this implied condition of statistical independency on the system matrices is removed by considering the complex FRF matrix of the system at the expense of additional computational burden. Use of FRF matrix, instead of mass, stiffness and damping matrices, automatically takes care of the issue of statistical dependency among the mass, stiffness and damping matrices of the system.

4.3 Nonparametric Model for Complex FRF Matrix

The receptance FRF matrix, $H(\omega)$, is expressed by,

$$H(\omega) = \left(K - \omega^2 M + \iota(\omega C + D)\right)^{-1},$$

where K, M, C and D are, respectively, the system matrices of stiffness, mass, viscous damping and structural damping, ω is the forcing frequency and $\iota = \sqrt{-1}$. Here, $H(\omega)$ is a complex symmetric matrix and consequently, has the following Takagi's factorization [HJ85, Corollary. 4.4.4],

$$H(\omega) = U\Sigma U^T, \tag{4.6}$$

where $U \in \mathbb{U}(n)$ is a unitary matrix with $\mathbb{U}(n)$ being the unitary group of unitary matrices of size $n \times n$. The set of orthonormal eigenvectors of $H(\omega)H(\omega)^*$ constitutes the columns of U and the positive square roots of the corresponding eigenvalues of $H(\omega)H(\omega)^*$ are the corresponding diagonal entries of Σ. Here * represents the element-wise conjugate operator. The Takagi's factorization for a complex symmetric matrix is a special case of singular value decomposition (SVD) for symmetric matrix. The SVD exists for any matrix $A \in \mathbb{M}_{m,n}(\mathbb{C})$ (where $\mathbb{M}_{m,n}(\mathbb{C})$ is the set of all complex matrices of size $m \times n$) such that $A = U\Sigma W^\dagger$ with $U \in \mathbb{U}(m)$ and $W \in \mathbb{U}(n)$ being unitary matrices and the diagonal entries of the diagonal matrix, Σ, being the non-negative square roots of the eigenvalues of $AA^\dagger$ (where † represents the conjugate-transpose operator). In Takagi's factorization for complex symmetric matrix, it turns out that $U = W^*$. Now, at any fixed ω, define $A = H(\omega)H(\omega)^*$. As $H(\omega)$ is symmetric, $H(\omega)^* = H(\omega)^\dagger$. Consequently, $A = A^\dagger$ is Hermitian and also positive definite because $x^\dagger A\, x = x^\dagger H(\omega)H(\omega)^*\, x = \left(H(\omega)^\dagger x\right)^\dagger \left(H(\omega)^\dagger x\right) = \|H(\omega)^\dagger x\|_2^2 > 0 \;\forall$ non-zero $x \in \mathbb{C}^n$. Here $\|\cdot\|_2$ is Euclidean norm on $\mathbb{C}^n$. Therefore, all the eigenvalues of $H(\omega)H(\omega)^*$ are positive. If we denote the group of all the diagonal matrices of size $n \times n$ with positive diagonal entries by $\mathbb{D}_n^+$, then $\Sigma \in \mathbb{D}_n^+ \subset \mathbb{M}_n^+(\mathbb{R})$. Note that (4.6) can also be written as $H(\omega) = V^T V$ where $V = (U\Sigma^{1/2})^T$ with $\Sigma^{1/2} = \text{diag}(+\sqrt{\sigma_1}, \cdots, +\sqrt{\sigma_n})$ in which σ_j is the j-th diagonal element of Σ. Hence, the receptance FRF matrix of the mean dynamical system $\underline{H}(\omega)$ has the following decomposition,

$$\underline{H}(\omega) = V_{H(\omega)}^T V_{H(\omega)}$$

which can be compared to (4.1). Now, the random receptance FRF matrix, $\mathbf{H}(\omega)$, can be written as,

$$\mathbf{H}(\omega) = V_{H(\omega)}^T \mathbf{G}_{H(\omega)} V_{H(\omega)} \tag{4.7}$$

in which $\mathbf{G}_{H(\omega)}$ is the random matrix variate (with all of its moments being finite) in $\mathbb{M}_n^+(\mathbb{R})$ and $E\left\{\mathbf{G}_{H(\omega)}\right\} = I_n$. The last constraint condition on "existence of moments of response" as mentioned earlier also guarantees the existence of the moments of $\mathbf{A}$ implying [Soi01a, p. 1985],

$$E\left\{\|\mathbf{A}\|_F^\gamma\right\} < \infty, \quad \text{for receptance FRF matrix,}$$

as required for the random matrix associated with the receptance FRF matrix in order to enforce the condition for the existence of random response quantity.

Then, a probability model for $\mathbf{G}_{H(\omega)}$ can be developed in exactly the same way as described earlier for $\mathbf{G}_M$ and $\mathbf{G}_K$ in section 4.2. Simulation of $\mathbf{G}_{H(\omega)}$, and therefore, of $\mathbf{H}(\omega)$, follows from the constructed probability model of $\mathbf{G}_{H(\omega)}$ by virtue of (4.7).

The additional computation burden of this FRF-based nonparametric formulation due to different probability model of $\mathbf{H}(\omega)$ at different values of ω in the frequency band of interest must be noted as a computational drawback. However, the question of statistical independent mass, damping and stiffness matrices does not arise in this case since the FRF matrix, that depends on these mass, damping and stiffness matrices, is alternately being characterized. This FRF-based nonparametric formulation is also more suitable from practical consideration because experimental measurement is directly available on the FRF matrix not on the mass, damping or stiffness matrices. In this case, there exists no modeling uncertainty since the 'true' physical process is directly measured. However, there do exists experimental/measurement uncertainty that is now characterized within the nonparametric formalism along with the inherent or irreducible uncertainty in the 'true' process.

4.4 Coupling Nonparametric Model and Parametric Model

The coupling technique used in the current work to combine the nonparametric model and the parametric model is based on the receptance FRF matrices of the uncoupled subsystems. The standard FRF coupling technique for two subsystems (say, denoted by a and b) is expressed by [JBF88],

$$\begin{bmatrix} {}_{aa}H & {}_{ac}H & {}_{ab}H \\ {}_{ca}H & {}_{cc}H & {}_{cb}H \\ {}_{ba}H & {}_{bc}H & {}_{bb}H \end{bmatrix} = \begin{bmatrix} H^{(a)}_{rr} & H^{(a)}_{rc} & \mathbf{0} \\ H^{(a)}_{cr} & H^{(a)}_{cc} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & H^{(b)}_{rr} \end{bmatrix}$$

$$- \begin{bmatrix} H^{(a)}_{rc} \\ H^{(a)}_{cc} \\ -H^{(b)}_{rc} \end{bmatrix} \begin{bmatrix} H^{(a)}_{cc} + H^{(b)}_{cc} \end{bmatrix}^{-1} \begin{bmatrix} H^{(a)}_{rc} \\ H^{(a)}_{cc} \\ -H^{(b)}_{rc} \end{bmatrix}^{T} . \tag{4.8}$$

The left-hand-side represents the whole FRF matrix of the built-up system which consists of the two subsystems, a and b. This FRF matrix is partitioned such that each partition can be expressed by the FRF matrices of the uncoupled subsystems a and b as shown in the right-hand-side (rhs). The FRF matrices of the uncoupled subsystems are denoted by $H^{(j)}_{..}$, $j = a, b$. The subscript, c, of these FRF matrices denotes the coupling dof involved in the common physical connection of the subsystems, a and b, and the subscript, r, represents the remaining internal dof of the corresponding subsystems, a and b. The built-up FRF matrix must be symmetric and so, we have the relations, ${}_{ca}H = {}_{ac}H^T$, ${}_{ba}H = {}_{ab}H^T$ and ${}_{bc}H = {}_{cb}H^T$, that can be realized upon expressing each partition of the built-up FRF matrix in terms of the FRF matrices of the uncoupled subsystems, $H^{(a)}_{..}$'s and $H^{(b)}_{..}$'s. This basic approach, however, is not directly adapted to joints with more than two components, and enhancements have been proposed in the literature [Urg91, RB95, SK97, Liu00, LL04]. One such method has already been used by the author in the context of a different class of vibration problem [DM03] and would be used in the current work. A description of this coupling technique follows next.

Consider the mean built-up system consisting of several subsystems. Also consider one constituent subsystem. Denote this subsystem by j and the complex FRF matrix of the subsystem by $\underline{H}^{(j)}(\omega)$ when it is isolated from the other adjoining subsystems of the built-up structure. The procedure of construction of the matrix, $\underline{H}^{(j)}(\omega)$, is not relevant in the coupling technique described here. The receptance FRF

matrix can be constructed by using any available technique, for example, modal analysis, direct inversion of the dynamic stiffness matrix (DSM) of the corresponding uncoupled finite element model, or even by experimentally identifying this FRF matrix. It should be noted here that the subsystem, j, could be either classically or non-classically damped and viscously or hysteretically damped. We need to adopt an appropriate method to analyze the subsystem in order to compute $\underline{H}^{(j)}(\omega)$. Given $\underline{H}^{(j)}(\omega)$, the response of the subsystem in frequency domain can be obtained from the following equation,

$$\underline{X}^{(j)}(\omega) = \underline{H}^{(j)}(\omega)\underline{F}^{(j)}(\omega), \tag{4.9}$$

where $\underline{X}^{(j)}(\omega) = \left[\underline{X}_1^{(j)}\ \underline{X}_2^{(j)}\ \cdots\ \underline{X}_{n_j}^{(j)}\right]^T \in \mathbb{C}^{n_j}$ is the response of subsystem j, where n_j is the total number of dof (displacements or/and rotations) of the subsystem, j. Similarly, $\underline{F}^{(j)}(\omega) = \left[\underline{F}_1^{(j)}\ \underline{F}_2^{(j)}\ \cdots\ \underline{F}_{n_j}^{(j)}\right]^T \in \mathbb{C}^{n_j}$ is the vector of forcing components (including the coupling forces and the externally applied forces) at the n_j dof, when the subsystem, j, is isolated from the other adjoining subsystems. Then, $\underline{H}^{(j)}(\omega)$ is an $n_j \times n_j$ complex symmetric matrix. The (s,t)-th element of this matrix, $\underline{H}_{s,t}^{(j)}(\omega)$, represents the response (displacement/rotation) of the subsystem j at the frequency ω at the s-th dof due to an unit force (load/moment) acting at t-th dof.

Let all the excitation points on subsystem j be denoted by $(\hat{I}, \cdots, \hat{Z})$ and all the coupling points by $(I, \cdots, Z)$. Consider one of these coupling points denoted by k, $k \in (I, \cdots, Z)$. Also suppose that there are N_k number of subsystems denoted by $(j, l, \cdots, s)$ connected at this coupling point, k. In addition, consider that at this coupling point, there are total p_k number of dof ($p_k \leq 6$) that must maintain the continuity of the corresponding responses among the subsystems, $(j, l, \cdots, s)$, meeting at the coupling point. For subsystem j, let us denote these dof that would generate either a coupling force or a coupling moment at the coupling point, k, when the subsystem j would be isolated from the other adjoining subsystems, by $m_j(o_k)$, $o_k = 1, 2, \cdots, p_k$. The response at the dof, $m_j(o_k)$, $o_k = 1, 2, \cdots, p_k$, then can be expressed by,

$$\underline{X}_{m_j(o_k)}^{(j)}(\omega) = \sum_{r=1}^{n_j} \underline{H}_{m_j(o_k),r}^{(j)}(\omega)\, \underline{F}_r^{(j)}(\omega) \tag{4.10}$$

It should be noted here that the force components, $\underline{F}_r^{(j)}(\omega)$ contain both the known externally applied forces and the unknown coupling forces resulting from the isolation of the subsystem j from the other adjoining subsystems, $(l, \cdots, s)$. Hence it is useful to decompose the right hand side of (4.10) into two

parts — one containing the contributions from the known components of the externally applied forces and the other containing the contributions from the unknown coupling forces. This is done next.

Let S_j be the set containing the dof associated with the unknown coupling force components for subsystem j, $S_j = \{m_j(o_k), o_k = 1, 2, \cdots, p_k, k \in (I, \cdots, Z)\}$. If we denote the total number of dof (at an excitation point, q, $q \in (\hat{I}, \cdots, \hat{Z})$), that are associated with the non-zero externally applied known force components, by p_q ($p_q \leq 6$) and the corresponding dof by $m_j(o_q)$, $o_q = 1, 2, \cdots, p_q$, we have $\hat{S}_j = \left\{m_j(o_q), o_q = 1, 2, \cdots, p_q, q \in (\hat{I}, \cdots, \hat{Z})\right\}$. As there are a total N_k number of subsystems, $(j, l, \cdots, s)$, connected at the coupling point k, one can write a total N_k number of expressions of response for each dof at this point. Now during the process of assembling the subsystems, we need to merge the appropriate dof of their coupling points. For convenience, we sort the elements of the sets of the dof of the subsystems, $(j, l, \cdots, s)$, connected at the coupling point k, $\{m_j(o_k), o_k = 1, 2, \cdots, p_k\}, \{m_l(o_k), o_k = 1, 2, \cdots, p_k\}, \cdots, \{m_s(o_k), o_k = 1, 2, \cdots, p_k\}$, such that the o_k-th element of each of the sets, $m_i(o_k)$, $i = j, l, \cdots, s$, $\forall\, o_k = 1, 2, \cdots, p_k$, refers to the same dof of the assembled structure. Hence, from the condition of compatibility of the response at the merged dof of the assembled system, following $(N_k - 1)$ number of equations can be formed (after performing some rearrangement-type operations) for each o_k, $o_k = 1, 2, \cdots, p_k$,

$$
\begin{aligned}
& \sum_{r \in S_j} \underline{H}^{(j)}_{m_j(o_k), r}(\omega)\, \underline{F}^{(j)}_{r}(\omega) - \sum_{r \in S_l} \underline{H}^{(l)}_{m_l(o_k), r}(\omega)\, \underline{F}^{(l)}_{r}(\omega) \\
= & -\sum_{\hat{r} \in \hat{S}_j} \underline{H}^{(j)}_{m_j(o_k), \hat{r}}(\omega)\, \underline{F}^{(j)}_{\hat{r}}(\omega) + \sum_{\hat{r} \in \hat{S}_l} \underline{H}^{(l)}_{m_l(o_k), \hat{r}}(\omega)\, \underline{F}^{(l)}_{\hat{r}}(\omega) \\
& \qquad\qquad\qquad\qquad\qquad\qquad \cdots\cdots \text{ 1-st equation} \\
\vdots & \\
& \sum_{r \in S_j} \underline{H}^{(j)}_{m_j(o_k), r}(\omega)\, \underline{F}^{(j)}_{r}(\omega) - \sum_{r \in S_s} \underline{H}^{(s)}_{m_s(o_k), r}(\omega)\, \underline{F}^{(s)}_{r}(\omega) \\
= & -\sum_{\hat{r} \in \hat{S}_j} \underline{H}^{(j)}_{m_j(o_k), \hat{r}}(\omega)\, \underline{F}^{(j)}_{\hat{r}}(\omega) + \sum_{\hat{r} \in \hat{S}_s} \underline{H}^{(s)}_{m_s(o_k), \hat{r}}(\omega)\, \underline{F}^{(s)}_{\hat{r}}(\omega) \\
& \qquad\qquad\qquad\qquad\qquad\qquad \cdots\cdots (N_k - 1)\text{-th equation.}
\end{aligned}
\tag{4.11}
$$

Here, subsystem, l, contains $(P, \cdots, T)$ coupling points (with $k \in (P, \cdots, T)$) and $(\hat{P}, \cdots, \hat{T})$ exciting points, and subsystem, s, contains $(S, \cdots, U)$ coupling points (with $k \in (S, \cdots, U)$) and $(\hat{S}, \cdots, \hat{U})$ exciting points so that $S_l = \{m_l(o_k), o_k = 1, 2, \cdots, p_k, \quad k \in (P, \cdots, T)\}$, $\hat{S}_l =$

$\{m_l(o_q), o_q = 1, 2, \cdots, p_q,\ \ q \in (\hat{P}, \cdots, \hat{T})\Big\}$, $S_s = \{m_s(o_k), o_k\ = 1, 2, \cdots, p_k, k \in (S, \cdots, U)\}$ and $\hat{S}_s = \{m_s(o_q), o_q = 1, 2, \cdots,\ \ p_q, q \in (\hat{S}, \cdots, \hat{U})\Big\}$.

It should be noted that rhs of this set of equations are completely known since it contains the known external loads acting on the subsystems. The unknown is the coupling force vector denoted by the concatenated column vector, $\left[\underline{F}_{S_j}^{(j)}\ \underline{F}_{S_l}^{(l)}\ \cdots\ \underline{F}_{S_s}^{(s)}\right]^T$, of the column vectors, $(\underline{F}_{S_t}^{(t)})^T, t \in (j, l, \cdots, s)$, that consist of the force components $\underline{F}_r^{(t)}$, $r \in S_t, t \in (j, l, \cdots, s)$. Set of equations, similar to (4.11), are further developed, in a similar manner, for all the coupling points of the built-up structure.

The next step is to consider the force equilibrium conditions of coupling forces (loads and moments) of different subsystems at a common dof. For coupling point, k, as the dof, $m_j(o_k)$, $m_l(o_k)$,$\cdots$ and $m_s(o_k)$, $\forall\, o_k\ = 1, 2, \cdots, p_k$, refer to the same dof of the assembled structure (because we choose to sorted them in this fashion), we have the following force equilibrium condition at the coupling point k,

$$\underline{F}_{m_j(o_k)}^{(j)}(\omega) + \underline{F}_{m_l(o_k)}^{(l)}(\omega) + \cdots + \underline{F}_{m_s(o_k)}^{(s)}(\omega) = 0, \qquad o_k = 1, 2, \cdots, p_k. \tag{4.12}$$

In this manner, it is possible to form the force equilibrium conditions of the coupling forces at all the coupling points of the assembled structure.

The sets of equations representing the deflection compatibility conditions (see (4.11)) at all the coupling points lead to a total of $\sum_{k=1}^{N_c}[N_k - 1]D_k$ equations. Here, N_c is the total number of coupling points in the assembled system, $N_k, N_k \geq 2$, is the number of subsystems coupled at the coupling point $k, k = 1, 2, \cdots, N_c$ and $D_k, D_k \leq 6$, is the total number of dof that would generate either a coupling load or a coupling moment at the coupling point, k, when one of the subsystems connected at the coupling point would be isolated from the other adjoining subsystems. The sets of equations representing the force equilibrium conditions (see (4.12)) at all the coupling points of the assembled system result in a total of $\sum_{k=1}^{N_c} D_k$ equations. Consequently, there is a total of $\sum_{k=1}^{N_c} N_k D_k$ equations representing the deflection compatibility and the force equilibrium conditions. On the other hand, there is a total of $\sum_{k=1}^{N_c} N_k D_k$ unknown coupling force components denoted by the unknown concatenated coupling force vector $\left[\underline{F}_{S_1}^{(1)}\ \underline{F}_{S_2}^{(2)}\ \cdots \underline{F}_{S_j}^{(j)}\ \underline{F}_{S_l}^{(l)}\ \cdots\ \underline{F}_{S_s}^{(s)}\ \cdots \underline{F}_{S_{N_s}}^{(N_s)}\right]^T$ with N_s being the total number of subsystems into which the built-up structure is decomposed. This unknown coupling force vector can be readily obtained by solving the above equations representing the deflection compatibility and the force equilibrium conditions. Having calculated all the coupling forces at all the coupling points, the response of

any subsystem, j, can be readily computed by using (4.9), $\underline{X}^{(j)}(\omega) = \underline{H}^{(j)}(\omega)\underline{F}^{(j)}(\omega)$. Though the procedure is described for mean built-up system, it remains precisely the same if applied to any other realization of the ensemble of the built-up systems. Therefore, the formulation can be applied to each realization of the ensemble of the systems to compute the ensemble of the responses which can be further processed to evaluate the statistics of the response quantities of interests. This procedure results in dynamic equilibrium of the stochastic system being satisfied sample-wise.

This coupling technique allows one to treat a complex dynamical structure as being formed of several simple subsystems, each of which could be analyzed individually and independently of others without having recourse to the global mode shapes of the built-up structure. Analysis of each constituent subsystem is performed by using a method that is most adapted to it. Assemblage of all such analyses at subsystem level results in equations for the built-up structure. This coupling technique requires, as inputs, the subsystems' FRFs over a given frequency range and produces the output of the built-up system. Basic output of this technique is the displacement field from which other response quantities, i.e., velocity, acceleration, stress and strain fields could be readily obtained.

The formulation is exemplified by considering a structure that consists of a set of three free-free Euler-Bernoulli beam (parametric) subsystems that are discretely coupled by a set of six axially vibrating rod (nonparametric) subsystems.

4.5 Illustration and Discussion on Results

Consider the built-up structure shown in Figure 4.1. Subsystems $2, 3, 4, 6, 7$ and 8 are analyzed by using nonparametric approach and subsystems $1, 5$ and 9 are analyzed by using parametric approach to compute the respective realizations of FRFs of the uncoupled subsystems. These computed FRFs are used to determine the realizations of the response of the built-up structure. The parametric subsystems are modeled as Euler-Bernoulli beams and the mean subsystems of all the connecting nonparametric subsystems are modeled as axially vibrating rods. The parametric subsystems are analyzed by using the classical dynamic analysis for continuous systems and subsequently, lowest 10 modes (including the 2 rigid-body modes) of each subsystem have been retained to compute the realizations of the FRFs by using the modal superposition method. On the other hand, the mass and stiffness matrices of the mean nonparametric subsystems are computed by using commercially available finite element analysis (FEA) software,

namely, ABAQUS. These mean mass and stiffness matrices are then used to generate the realizations of the mass and stiffness matrices by following the nonparametric approach as described earlier. Damping of these subsystems, however, are treated parametrically. In this sense, the nonparametric subsystems have themselves been characterized by both the parametric (with respect to damping) and nonparametric components (with respect to mass and stiffness matrices) (A problem of nonparametric-parametric nature has earlier been reported in literature [DSC04]). Three modes are retained (including the 1 rigid-body mode) in the FRF computation of nonparametric subsystems. An uniform distribution $U(5 \times 10^{-4}, 1.5 \times 10^{-3})$ with mean value of 0.001 is assumed for modal damping over all the modes for all subsystems.

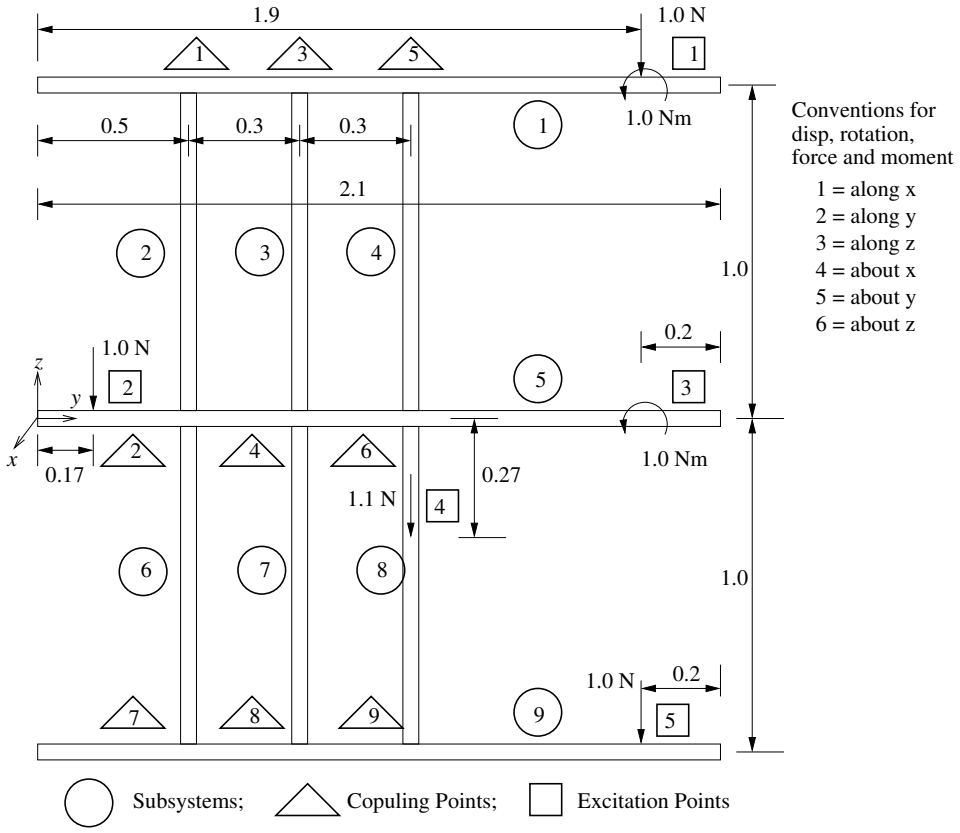


Figure 4.1: Mean built-up structure; $E = 2.0 \times 10^{11}$ N/m^2, $\rho = 7850$ kg/m^3, Circular section with radius $r = 0.025$ m, Modal critical damping $\xi = 0.001$ over all modes; all dimensions are in m.

The cross-sections of the parametric subsystems are assumed to be circular with radius having an uniform distribution $U(0.0248, 0.0253)$ m mean radius $r = 0.025$ m. The material is assumed to be isotropic and homogeneous with material density having an uniform distribution $U(7457.5, 8242.5)$ kg/m^3 with mean $\rho = 7850$ kg/m^3 and Young's modulus having an uniform distribution $U(1.9 \times 10^{11}, 2.1 \times$

10^{11}) N/m^2 with mean $E = 2.0 \times 10^{11}$ N/m^2. The length of all the parametric subsystems is treated as deterministic with a value of 2.1 m. However, the y-coordinates of the coupling points and the excitation points (see Figure 4.1) are assumed to be random variables having uniform distribution as follows: $cp_1 = U(0.495, 0.505)$ m on subsystem 1, $cp_3 = U(0.795, 0.805)$ m on subsystem 1, $cp_5 = U(1.095, 1.105)$ m on subsystem 1, $cp_2 = U(0.495, 0.505)$ m on subsystem 5, $cp_4 = U(0.795, 0.805)$ m on subsystem 5, $cp_6 = U(1.095, 1.105)$ m on subsystem 5, $cp_7 = U(0.495, 0.505)$ m on subsystem 9, $cp_8 = U(0.795, 0.805)$ m on subsystem 9, $cp_9 = U(1.095, 1.105)$ m on subsystem 9, $et_1 = U(1.895, 1.905)$ m on subsystem 1, $et_2 = U(0.165, 0.175)$ m on subsystem 5, $et_3 = U(1.895, 1.905)$ m on subsystem 5, $et_5 = U(1.895, 1.905)$ m on subsystem 9, where cp_k represents the y-coordinate of the coupling point k, $k = 1, \cdots, 9$ and et_k represents the y-coordinate of excitation point k, $k = 1, 2, 3, 5$ on parametric subsystems.

The mean subsystems of all the nonparametric subsystems are assumed to have circular cross-sections with radius $r = 0.025$ m, isotropic and homogeneous material with material density $\rho = 7850$ kg/m^3 and Young's modulus $E = 2.0 \times 10^{11}$ N/m^2. The lengths of all the mean nonparametric subsystems are assumed to be 1.0 m. The dispersion parameters of the associated system matrices are assumed to be, $\delta_{K,j} = 0.4$, $\delta_{M,j} = 0.4$, where $\delta_{M,j}$ and $\delta_{K,j}$ denote, respectively, the dispersion parameters for mass matrix and stiffness matrix of the nonparametric subsystem j, $j = 2, 3, 4, 6, 7, 8$. This is considered as **case 1**.

Total number of realizations used in the MCS technique to find the statistics of the response of the built-up structure is 575. This number of realizations of random mass and stiffness matrices and random modal damping parameter (each realization of modal damping parameter remains constant over all modes included in the modal superposition method) are generated to calculate a total of 575 realizations of the receptance FRF matrix at each $\omega = \{1, 2, \cdots, 300\}$ Hz in the frequency band of interest, $[1, 300]$ Hz, for each uncoupled nonparametric subsystem. Subsequently, these realizations of the FRF matrix, $\mathbf{H}^{(j)}(\omega), j = 2, 3, 4, 6, 7, 8$, are used to estimate the values of the associated dispersion parameters per (4.3) by using,

$$\delta_{H,j} = \left[\frac{1}{m} \left(\sum_{u=1}^{m} \| H^{(j)}(\omega; u) - \underline{H}^{(j)}(\omega) \|_F^2 \right) \Big/ \| \underline{H}^{(j)}(\omega) \|_F^2 \right]^{1/2}. \tag{4.13}$$

Here, $m = 575$ and $H^{(j)}(\omega; u), u = 1, \cdots, m$, is the u-th realization of $\mathbf{H}^{(j)}(\omega)$. The results are plotted in Figure 4.2. It can be seen that the dispersion parameter of random FRF matrix of each uncoupled nonparametric subsystem remains almost constant over the frequency range of interest, $[1, 300]$ Hz, and consequently, $\delta_{H,2} = 0.2513$, $\delta_{H,3} = 0.2542$, $\delta_{H,4} = 0.2462$, $\delta_{H,6} = 0.2406$, $\delta_{H,7} = 0.2510$ and $\delta_{H,8} = 0.2799$ have been considered in the second phase of analysis (**case 2**) when realizations of the random recptance FRF matrices of the nonparametric subsystems are generated instead of generating the realizations of the random mass and stiffness matrices and random modal damping parameters of the nonparametric subsystems. In this second case, the subsystems $2, 3, 4, 6, 7$ and 8 are precisely characterized

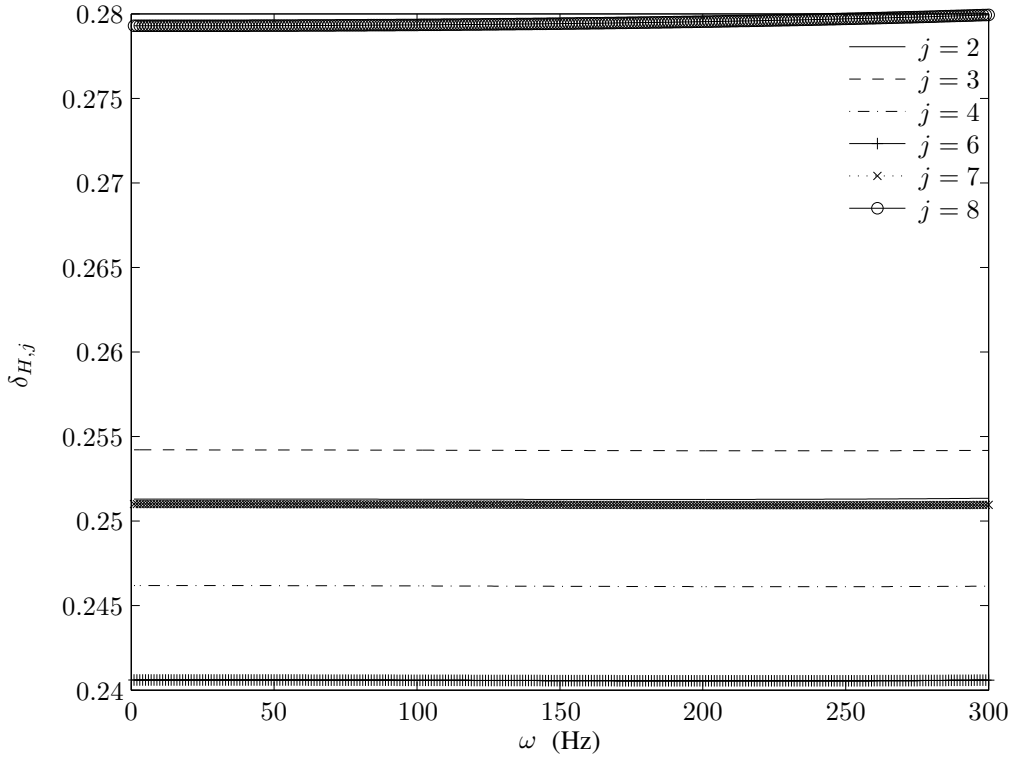


Figure 4.2: Dispersion parameters of receptance FRF matrices of nonparametric subsystems.

by the nonparametric model unlike the first case. The mean matrices of $\mathbf{H}^{(j)}(\omega), j = 2, 3, 4, 6, 7, 8$, in the second case are considered to be the same as the mean matrices of the FRF matrices (calculated based on the mean mass and stiffness matrices and mean modal damping parameter considered in case 1) in the first case. The characteristics of the random parameters of the parametric subsystems in the second

case are considered to be the same as the respective characteristics in the first case. Again, a total of 575 realizations of the built-up system are generated to use in the MCS technique in the second case.

Based on 575 realizations of the receptance FRF matrices of each uncoupled subsystems in the frequency band $[1, 300]$ Hz for each case, a total of 575 realizations of the responses of the built-up structure for each case is computed by using the coupling technique as described earlier.

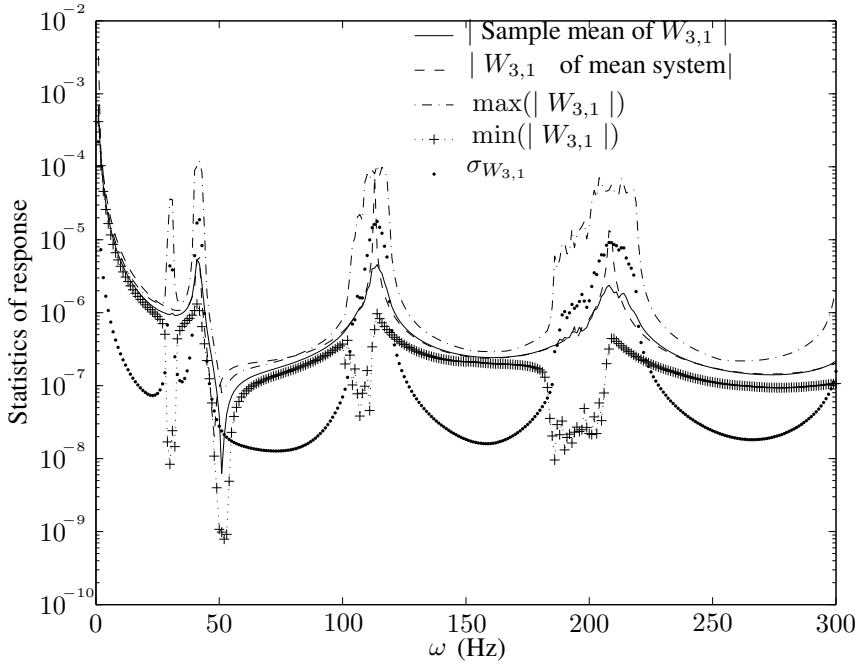


Figure 4.3: Statistical details of the deflection, $W_{3,1}$, of the built-up structure (**case 1**).

Use of these 575 realizations of the response of the built-system yields Figure 4.3 and Figure 4.4 showing the statistics of the response $|W_{3,1}|$ of the built-up system, respectively, for case 1 and case 2. In this figures, $|\cdot|$ represents the magnitude of the response and $W_{m,k}$ represents the displacement or rotation denoted by $m, m = 1 \cdots, 6$, (according to the convention described in Figure 4.1) at the coupling point $k, k = 1, \cdots, 9$. It can be seen that all the statistics (sample mean, sample standard deviation, sample maximum and sample minimum) of the displacement along z direction at coupling point 1 computed in both the cases match very closely. In this figures, the response of the mean built-up system is also superimposed. It is noted that the response (same for both the cases) of the mean system usually lies in the interval bounded by the sample maximum and sample minimum except in the frequency range $50 - 75$ Hz near anti-resonance for both the cases. The responses of the built-up system near a few resonance

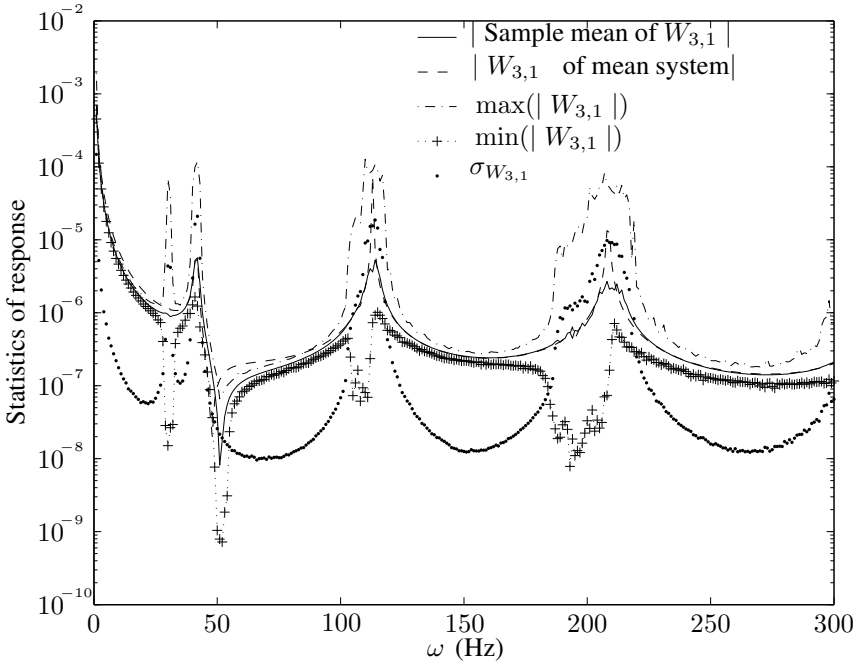


Figure 4.4: Statistical details of the deflection, $W_{3,1}$, of the built-up structure (**case 2**).

frequencies of the mean system, i.e., at 42 Hz, 113 Hz and 208 Hz, have been separately investigated to see if they follow some of the usual distributions, specially, normal or log-normal distribution. It was found (see Figures 4.5–4.8; not all plots are included here) that the response does not follow either distribution. Comparisons of the respective plots of the two cases, however, show that the patterns of the simulated response in both the cases are of the similar nature. This also validates the correctness of the nonparametric formulation of the complex symmetric receptance FRF matrix as presented here.

4.6 Conclusion

A FRF-based coupling technique to combine the parametric and nonparametric models of stochastic systems has been described. This technique enables the analysis of a complex dynamical systems having spatially non-homogeneous uncertainty. Complex dynamical system with spatially non-homogeneous uncertainty can be decomposed into several smaller components such that each component separately shows spatially homogeneous uncertainty over its domain. Consequently, each smaller subsystem is analyzed by using an approach most pertinent to it, and then the results at the subsystem level are assembled

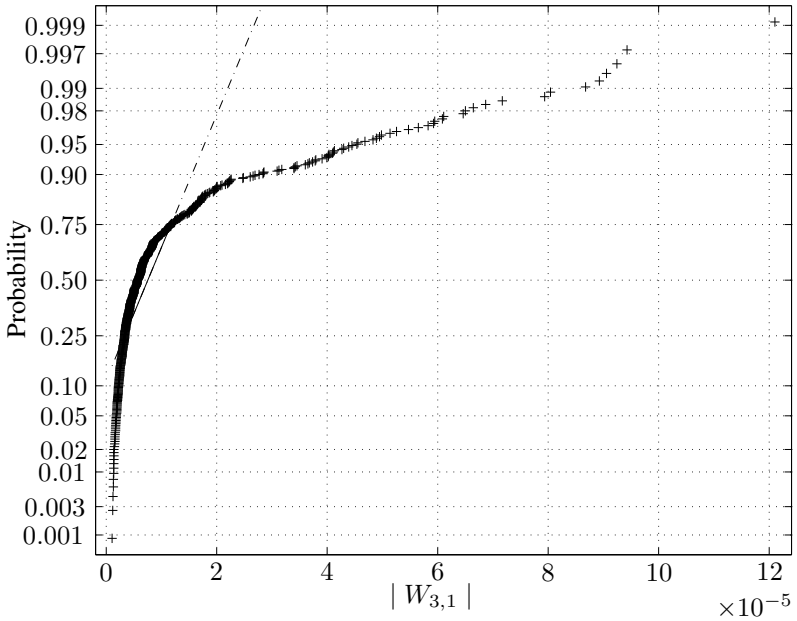


Figure 4.5: Normal probability plot of $|W_{3,1}|$ at $\omega = 42$ Hz (case 1).

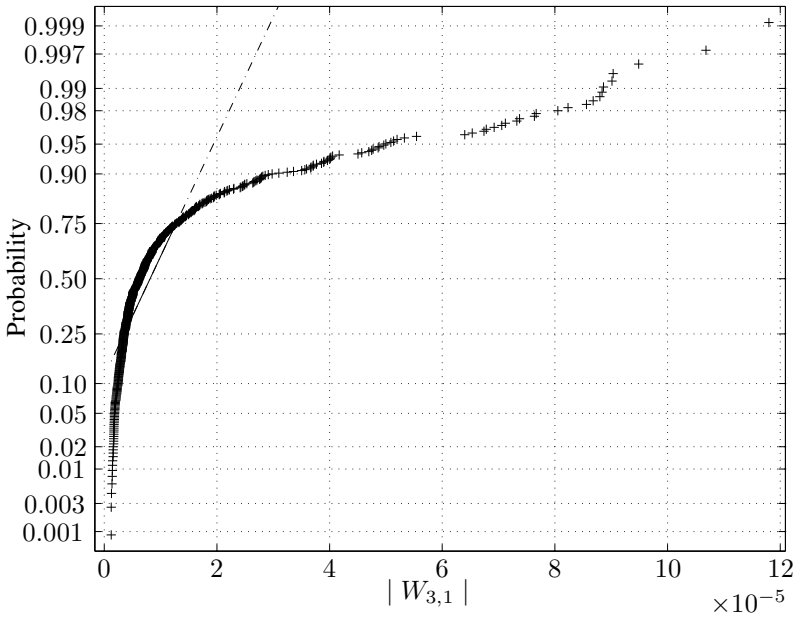


Figure 4.6: Normal probability plot of $|W_{3,1}|$ at $\omega = 42$ Hz (case 2).

by using the coupling technique described here to obtain the response quantities of interest at the built-up system level. The FRFs associated with the uncoupled subsystems as required in this coupling technique

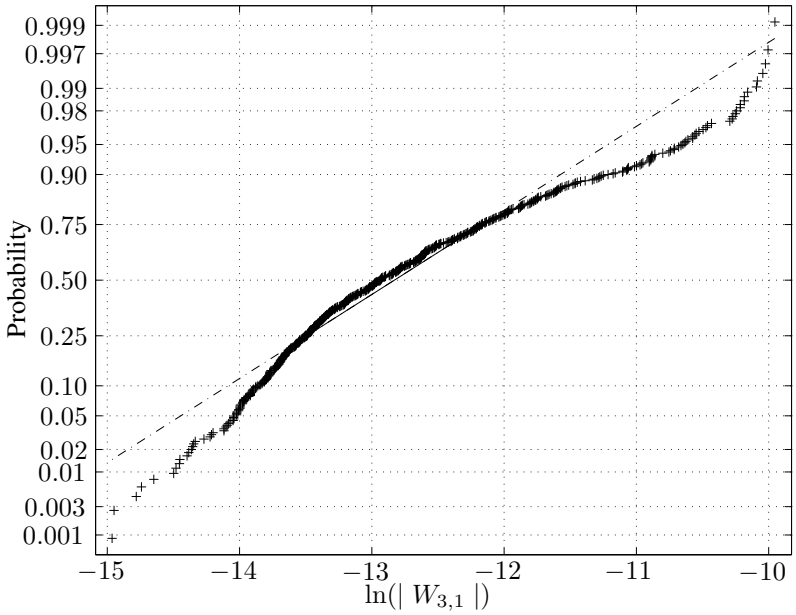


Figure 4.7: Normal probability plot of $\ln(|W_{3,1}|)$ at $\omega = 208$ Hz (case 1).

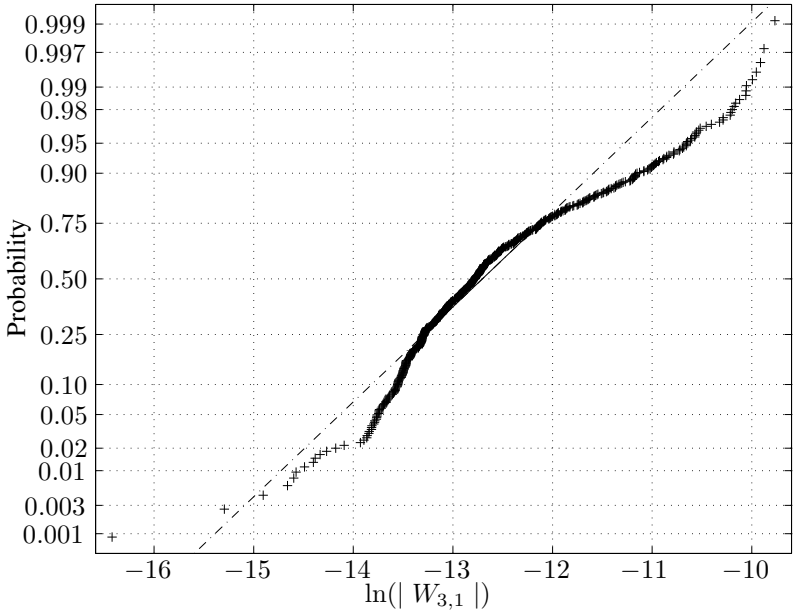


Figure 4.8: Normal probability plot of $\ln(|W_{3,1}|)$ at $\omega = 208$ Hz (case 2).

can be determined analytically or numerically (e.g., by using FEM) as well as experimentally (e.g, through laboratory test).

Not only the usual real positive definite/semi-definite mass, stiffness and damping matrices can be modeled within the framework of the nonparametric approach, it is also shown here that the nonparametric formulation can also be employed to model the uncertainty in the complex symmetric FRF matrix of the system. More generally, even if any system matrix does not show any symmetry (e.g., rotating systems having skew-symmetric damping matrix or/and skew-symmetric stiffness matrix), it can be effectively dealt with by using the nonparametric approach by having recourse to the SVD that exists for any matrix $A \in \mathbb{M}_{m,n}(\mathbb{C})$.

Chapter 5

A Bounded Random Matrix Approach

All fixed set patterns are incapable of adaptability or pliability.

The truth is outside of all fixed patterns.

$\sim$ Bruce Lee (November 27, 1940 – July 20, 1973)

A random matrix approach is proposed in the present chapter to model a stochastic mechanical system characterized by symmetric positive definite random matrix that is bounded by two deterministic matrices in positive definite sense from below and above. The existing random matrix approach in the field of computational mechanics is only adapted to the Wishart matrix supported over the entire space of the symmetric positive definite matrices, and therefore, unable to exploit the additional information available through the lower and upper bounds when appropriate. Such a bounded positive definite random matrix is naturally encountered in the homogenization of a heterogeneous material. A new concept, nonparametric homogenization, in this context, is introduced. It is also highly unlikely that the system matrices of an ensemble of nominally different mechanical structures could span the entire space of the symmetric positive definite matrices.

5.1 Motivation

The present work proposes a maximum-entropy (MaxEnt) [Jay57a, Jay57b, Kap89, KK92] based probabilistic formulation within a nonparametric framework by using the random matrix theory (RMT). The resulting probability model is useful to characterize a positive definite random matrix, $\mathbf{C}$, that is bounded in the following sense,

$$\mathbf{0} < C_l < \mathbf{C} < C_u \quad \text{a.s.}, \tag{5.1}$$

in which $\mathbf{0}$ is a zero matrix, C_l and C_u are two positive definite deterministic matrices, and the inequalities should be interpreted in the positive definite sense (for instance, $C_l < \mathbf{C}$ a.s. implies that $(\mathbf{C} - C_l)$ is

positive definite matrix a.s.). Here, a.s. (almost surely, i.e., with probability one) should be interpreted with respect to (w.r.t.) the joint probability measure of all the associated random variate(s) characterizing the uncertainties. These uncertainties are induced by several errors, for example, data error, modeling error, etc., typically involved in modeling a physical phenomenon of interest.

As reviewed in section 4.2, the existing nonparametric approach results in Wishart or matrix-variate gamma distribution (see, e.g., [Mur82, Section 3.2], [GN00, Chapter 3]). The Wishart distribution is supported over the entire interior of the positive semi-definite cone [Dat05, Section 2.9] rendering the existing nonparametric model inapplicable when the random system matrices are positive definite and bounded in the sense as defined by (5.1). Another important concern expressed by a recent article [Adh07] is about the *large* discrepancy between the inverse of the ensemble average or mean of a Wishart random matrix and the ensemble average of the inverse of the Wishart matrix. It was further proposed there to modify the parameters of the Wishart distribution in order to minimize such difference if it is practically unacceptable. Instead of such heuristic scheme, the present work advocates collecting or inferring more information from the underlying physics of the problem (for example, the bounds as shown in (5.1)) and encapsulating all the available information to the extent possible in constructing the probability model of $\mathbf{C}$. It should be noted that the bounds on $\mathbf{C}$ automatically imply similar bounds on $\mathbf{C}^{-1}$, i.e., $\mathbf{0} < C_u^{-1} < \mathbf{C}^{-1} < C_l^{-1}$ a.s. [HJ85, Corollary. 7.7.4].

The current chapter begins with describing in the next section a problem of significant practical importance where a positive definite and bounded random matrix is naturally encountered in determining the *effective* (or macroscopic or overall) material property of a heterogeneous material. The existing schemes for determining the effective material property and the relevant concept, in the context of the present work, are also reviewed in section 5.2 followed by the probabilistic formulation in section 5.3 for characterizing such a positive definite and bounded random matrix. The proposed probabilistic formulation is just not limited to characterizing the random effective material property but it is equally applicable to other similar problems if the associated random system matrices are positive definite and bounded. Therefore, if the problem of effective material property is not of the readers' interests, they may wish to skip to the self-contained section 5.3. It would not impede the flow of reading. Previous works primarily in the context of computational mechanics, that make use of the RMT, are briefly reviewed or referred to the appropriate scholarly literature as and when required. Sampling schemes are also highlighted in section 5.3 for simulation of such bounded positive definite random matrix. The proposed approach is

numerically illustrated in section 5.4. Finally, section 5.5 is reserved for the conclusion inferred from the work presented in this chapter.

The main contributions of the proposed work are the introduction of the new concept of the nonparametric homogenization of a heterogeneous material in the multiscale field and the mathematical formulation presented in section 5.3. This formulation would be useful for constructing the probability model of a positive definite and bounded random matrix such as random effective elasticity matrix.

5.2 Parametric Homogenization

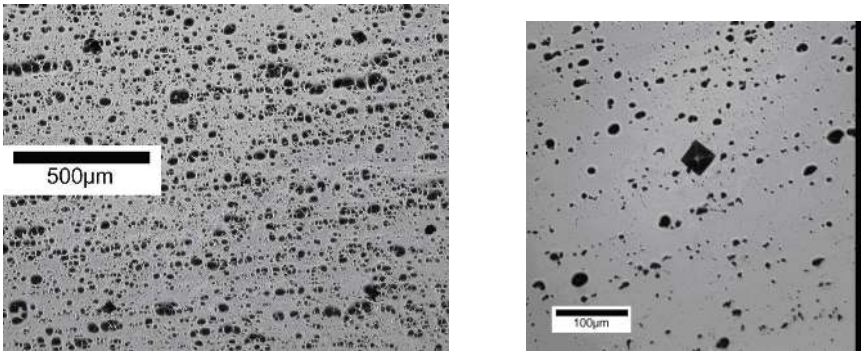
Physical phenomena associated with many problems of fundamental and practical importance exhibit a broad spectrum of rich complexity coupled with multiple space and time scale features. Prominent areas of multiscale applications include, to name a few, heterogeneous materials (concrete structures), composite materials (ships and aircrafts), flow and transport in porus media, marine structures subjected to underwater detonation, and living cells (biomolecular mechanics). Exhaustive review and highlight of many different facets of this multidisciplinary area are available in the prevailing scholarly articles and monographs (see, e.g., [Baz00, TPO00, LKP06, VG08, OS08]).

It is generally acknowledged that even with the state-of-the-art computer hardware and computing technology, studying multiscale problem ab initio atomistic level or quantum level is a formidable undertaking primarily due to the massive requirement of hard-disk space and memory [RKL$^+$02, KGL04]. This challenge brings forward a class of appealing hierarchical (coupled) downscaling approaches [LLY99, EE03, ZG04, ZKG05, CF06, FNS$^+$07, LKP06], [LKP06, Chapter 5–8]. Within the confine of such approaches, the entire spatio-temporal domain of a multiscale process is mostly represented by the well-established or conventional coarse-scale law (for example, the continuum theory) with the exception of only a very tiny subset of the domain that is enriched, if needed, by the fine-scale theory (for example, atomistic description near a crack tip). The information between the coarse-scale domain and the fine-scale domain is typically exchanged through a virtual 'boundary' defined for mathematical and computational tractability.

The focus of the current work, on the other hand, is another class of multiscale research efforts, namely, the hierarchical (uncoupled) upscaling approach. It is particularly useful if the overall response

of the coarse-scale model is of primary concern, albeit, with due care by incorporating the significant fine-scale mechanism in some approximate sense. Through this upscaling approach, the constitutive law of the conventional coarse-scale model is updated or constructed, or parameters of the coarse-scale constitutive law are gleaned by incorporating the effects of the fine-scale regime.

The present work is specifically concerned with the effective material property of a heterogeneous material. Zooming into the neighborhood of a macroscopic (continuum) material point at the scale of micron level ($1\,\text{micron} = 1\,\mu\text{m} = 1 \times 10^{-6}\,\text{m}$), that belongs to the mesoscopic domain, would show a wide variety of nonuniform and non-regular characteristics. This variation results from the fluctuations in the textural or morphological features of the microconstituents, for example, volume fraction, geometrical shape and size, spatial orientation, clustering etc. A typical sample of aluminium at the mesoscale regime is shown in Figure 5.1. It is a polycrystalline material whose mesoscopic texture (associated with an



(a) (b)

Figure 5.1: Heterogeneity of Al2024 at two different scales (mesoscopic regime).[1]

arbitrary material point in the macroscopic regime) typically shows a dominant matrix phase consisting of an assembly of single crystal grains connected by grain boundaries, several inclusions or precipitates (copper, magnesium, silica, manganese etc.) as a secondary phase and a large number of crystallographic defects. In the rest of the chapter, the term, *microstructural characteristics* or *microstructural features* or sometime simply *microstructure*, will be used to coherently indicate the features or characteristics induced by the constitutive laws, texture or morphology, interfacial nature (bonding/debonding) and interactions of the microconstituents.

[1] via private communication with Professor Pedro Peralta, Department of Mechanical & Aerospace Engineering, Arizona State University, Tempe, AZ 85287-6106.

Since the underlying microstructural features, and therefore, the macroscopic property of the heterogeneous material, randomly vary across samples, a probabilistic formalism clearly is more suitable to characterize the random macroscopic property with the microstructural features being modeled as random fields. Consider a volume of heterogeneous material of interest, over domain, $\mathbf{D} \subseteq \mathbb{R}^d$, $\mathbb{R}^d$ representing the Euclidean d-space, being subjected to a specified deterministic loading condition for which the heterogeneous material can be approximated as a linear elastic material. The specific focus of the current work is the positive definite fourth-order effective modulus or elasticity tensor.

Let $\mathbf{C}^{\text{eff}}$ be the matrix representation of the fourth-order effective elasticity tensor. The positive definite effective elasticity matrix, $\mathbf{C}^{\text{eff}}$, is determined by invoking the concept of representative volume element (RVE) — a classical notion that was first introduced by Hill [Hil63]. The classical RVE can be interpreted as a large enough "macroscopically uniform" or "statistically uniform" or "statistically representative" [Hil63, Hue90], [NNH99, Section 2] material volume cut from the heterogeneous material around *any* macroscopic point $\mathbf{x} \in \mathbf{D}$. This classical RVE essentially implies satisfaction of the following two postulations.

Assumption 5.2.1 ***(Spatial homogeneity and ergodicity)*** *The microstructural random fields of the heterogeneous material must be spatially homogeneous (stationary) and ergodic.*

Assumption 5.2.2 ***(Independence of boundary conditions)*** *The classical RVE-based* $\mathbf{C}^{\text{eff}}$ *must be independent of the boundary conditions applied on the boundary of the RVE.*

Consider two boundary conditions as shown below applied on a volume element, V, of the heterogeneous material with boundary ∂V.

Kinematic uniform boundary condition (KUBC): The prescribed displacement vector, $\mathbf{u}(\mathbf{x})$, is of the following form,

$$\mathbf{u}(\mathbf{x}) = \boldsymbol{\varepsilon}_o \, \mathbf{x}, \quad \forall \mathbf{x} \in \partial V,$$

where $\boldsymbol{\varepsilon}_o$ is a constant symmetric second-order strain tensor i.e., a constant symmetric strain matrix with the values of its components in the order of the magnitude for which $\mathbf{C}^{\text{eff}}$ needs to be determined.

Static uniform boundary condition (SUBC): The applied traction vector surface density, $\mathbf{t}(\mathbf{x})$, takes the following form,

$$\mathbf{t}(\mathbf{x}) = \boldsymbol{\sigma}_o \, \mathbf{n}(\mathbf{x}), \quad \forall \mathbf{x} \in \partial V, \tag{5.2}$$

where $\boldsymbol{\sigma}_o$ is a constant and consistent symmetric stress matrix and $\mathbf{n}(\mathbf{x})$ denotes the unit vector normal to ∂V at $\mathbf{x}$.

For V to be an RVE in the sense defined above, it is required that V be an infinite-sized volume element in a rigorous mathematical sense [OS02, BP04]. An infinite volume element, however, cannot be considered in an experimental or a computational setup. It is proved in literature [Hue90], [NNH99, Section 2] that, for a finite-sized V, $\mathbf{C} \equiv \mathbf{C}^{\text{eff}}$ satisfies (5.1) with $C_l \equiv C^{\text{app}}_{\boldsymbol{\sigma}}$ and $C_u \equiv C^{\text{app}}_{\boldsymbol{\varepsilon}}$. Here, $C^{\text{app}}_{\boldsymbol{\sigma}}$ and $C^{\text{app}}_{\boldsymbol{\varepsilon}}$ represent the positive definite *apparent* — a notion introduced by Huet [Hue90] — elasticity matrices. These matrices, $C^{\text{app}}_{\boldsymbol{\sigma}}$ and $C^{\text{app}}_{\boldsymbol{\varepsilon}}$, are essentially the ensemble average or mean of the positive definite elasticity matrices resulting from the numerical analysis of an ensemble of finite-sized V subjected to, respectively, KUBC and SUBC. It should be noted though that the final goal of obtaining such bounds, within a parametric framework, is to determine the appropriate bounds for the (random) system parameters (see, e.g., [Hil63, HS63, Hue90, HH94], [NNH99, Section 2, 9 and Appendix D], [Tor02, Chapter 14, 20-23]). The utility of the resulting bounds is, however, not rigorously articulated in the existing multiscale literature.

One remedial route in such situation is to consider, in lieu of Assumption 5.2.1, other statistical conditions that are weaker than Assumption 5.2.1. For instance, spatial homogeneity and ergodicity conditions only w.r.t. a selective statistical estimators (e.g., lineal path function, marked correlation function, or a mixture of a few such statistics) of a limited set of random microstructural features are often considered in the literature [Zv01, SGP04, SG04b, Zv07], [Tor02, Part I]. Certain periodicity structure in the displacement field is also often enforced for computational convenience [Zv01, SGP04, SG04b, Zv07]. The effective elasticity matrix based on such approaches is usually determined (typically in a deterministic sense) by considering a reasonably large V and applying either KUBC or SUBC or a series of *selective* strain-based boundary conditions. The extracted effective elasticity matrix, nevertheless, also satisfy (5.1) [Hue90, KFG$^+$03].

From the purview of probabilistic reasoning, consideration of a large V is implicitly geared towards reducing the uncertainty induced by *data error* caused by a *finite* set of experimental samples. The reason follows from the fact that the ensemble average of a function of spatially homogeneous and *ergodic*

microstructural random fields can be replaced with the corresponding volume average. The effect of increasing the size of V on the proximity between the two bounds has been extensively studied by Ostoja-Starzewski and his co-workers [OS01, OS02, DOS06, OSDKL07, OS08] (see also [KFG$^+$03]).

While a large enough material volume, V, certainly reduces the data error, it fails to efficiently capture many other errors inherently involved in characterizing $\mathbf{C}^{\text{eff}}$, namely, experimental error, modeling error and numerical error as explained by the following bounded inequality,

$$\|\mathbf{C}^{\text{eff}} - \mathbf{C}^{\text{eff}}_{(\text{true})}\| \leq \|\mathbf{C}^{\text{eff}} - \mathbf{C}^{\text{eff}}_{(\text{data})}\| + \|\mathbf{C}^{\text{eff}}_{(\text{data})} - \mathbf{C}^{\text{eff}}_{(\text{exp})}\| + \|\mathbf{C}^{\text{eff}}_{(\text{exp})} - \mathbf{C}^{\text{eff}}_{(\text{mod})}\| + \|\mathbf{C}^{\text{eff}}_{(\text{mod})} - \mathbf{C}^{\text{eff}}_{(\text{true})}\| \quad \text{a.s.} \quad (5.3)$$

Here, $\|\cdot\|$ is a suitable matrix norm, for example, Frobenius norm and $\mathbf{C}^{\text{eff}}_{(\text{true})}$ is the "true" (unknown) effective elasticity matrix of a heterogeneous material with spatially homogeneous and ergodic microstructural random fields. In reality, it cannot be verified in general if such a true $\mathbf{C}^{\text{eff}}_{(\text{true})}$ exists or not, and even if it exists, such a $\mathbf{C}^{\text{eff}}_{(\text{true})}$ would remain elusive. Nevertheless, the concept of such a true and unknown effective elasticity matrix facilitates in understanding the meaning of, e.g., modeling error, data error etc., more clearly. The other notations in (5.3) can be explained similarly and would be more clear as the associated error terms are explained below.

The last error term, $\|\mathbf{C}^{\text{eff}}_{(\text{mod})} - \mathbf{C}^{\text{eff}}_{(\text{true})}\|$, in (5.3) represents the modeling error. While modeling error can be reduced by considering several detailed modeling issues (for example, the stress concentrations at the vicinity of grain boundaries, orientations of the contiguous grains, defects and heterogeneities within an individual grain etc.) [DLRC05, SG04b, DWR07, AS07, RDD07], it cannot be effectively characterized within the conventional parametric framework. It should be noted that characterization and reduction are two different issues.

The third error term, $\|\mathbf{C}^{\text{eff}}_{(\text{exp})} - \mathbf{C}^{\text{eff}}_{(\text{mod})}\|$, represents the experimental error in characterizing $\mathbf{C}^{\text{eff}}_{(\text{mod})}$. To explain it further, consider the optical microscopy or the orientation imaging microscopy that is typically used to identify the microtexture of polycrystal at the micron level. A grain boundary is generated by comparing orientations between each pair of neighboring observation points in the micrograph scan with a specified tolerance angle. It is feasible that the two adjacent microscopic regions with the difference in their orientations smaller than the specified tolerance angle can be recognized as one single grain. It is reported by Ren and Zheng [RZ04] that the effective material property is influenced by the grain sizes, shapes and spatial distribution. Moreover, the currently available microscopy techniques can only be

used to create 2D micrographs. Construction of 3D microtexture based on experimentally identified 2D micrographs is another current research field on its own right [SFED$^+$04, BAS$^+$06]. This experimental error can only be reduced by employing a better experimental set-up or scheme, and it is not the focus of the present work. Therefore, it is assumed in the present work that $\mathbf{C}^{\text{eff}}_{(\text{exp})} \approx \mathbf{C}^{\text{eff}}_{(\text{mod})}$.

Now, consider the second error term, $\|\mathbf{C}^{\text{eff}}_{(\text{data})} - \mathbf{C}^{\text{eff}}_{(\text{exp})}\|$, that represents the data error due to *finite* set of experimental samples. This error can be reduced by collecting more data and enhancing the statistical qualities of the data.

Finally, consider the first error term in (5.3) caused by the numerical error. This error can only be reduced by invoking better numerical scheme (for example, improved finite element discretization scheme and better element type) [DLRC05]. In the present work, it is assumed that $\mathbf{C}^{\text{eff}} \approx \mathbf{C}^{\text{eff}}_{(\text{data})}$. Hence, based on the discussion thus far, the inequality in (5.3) can be simplified to,

$$\|\mathbf{C}^{\text{eff}} - \mathbf{C}^{\text{eff}}_{(\text{true})}\| \leq \|\mathbf{C}^{\text{eff}} - \mathbf{C}^{\text{eff}}_{(\text{mod})}\| + \|\mathbf{C}^{\text{eff}}_{(\text{mod})} - \mathbf{C}^{\text{eff}}_{(\text{true})}\| \quad \text{a.s.} \tag{5.4}$$

The present work deals with only these remaining two error terms as shown in (5.4). As already indicated, the second error term in (5.4) is due to the modeling error, and the first error term is due to the data error, $\|\mathbf{C}^{\text{eff}} - \mathbf{C}^{\text{eff}}_{(\text{mod})}\|$. The parametric formulation is efficient in characterizing the data error [DGS06, GD06, DGS08] as illustrated in chapters 2–3 but not the modeling error at least not to the extent the data error is characterized within a parametric formulation. The nonparametric formulation, on the other hand, is more efficient and effective in characterizing the modeling error (as well as the data error) [Soi05a, Soi05b, CLPP$^+$07]. The detailed modeling schemes as indicated earlier in the context of reduction of modeling error (see pp.107) can also be seamlessly integrated within the nonparametric formulation to reduce the modeling error.

The present work, therefore, presents a rigorous probabilistic framework based on nonparametric formulation to characterize $\mathbf{C}^{\text{eff}}$. The uncertainties associated with the resulting probability model of $\mathbf{C}^{\text{eff}}$ can be ascribed not only to the data error but also to the (possibly significant) modeling errors which are intrinsically present in characterizing $\mathbf{C}^{\text{eff}}$ based on a fragment of heterogeneous material volume element, V.

Starting with the definition of $\mathbf{C}^{\text{eff}}$ for a heterogeneous material, the concept of $\mathbf{C}^{\text{eff}}$ is reviewed and the notion of nonparametric $\mathbf{C}^{\text{eff}}$ is elaborated below setting up the stage for the probabilistic formulation in section 5.3.

5.2.1 The Concept of Effective Elasticity Matrix

If the stress and strain states resulting from a specified deterministic loading condition can be approximated by a linear relationship and are independent of the boundary condition, then the *local* effective elasticity matrix of a nonhomogeneous and nonergodic heterogeneous material is defined by,

$$E[\boldsymbol{\sigma}(\mathbf{x})] = \mathbf{C}^{\text{eff}}(\mathbf{x})\, E[\boldsymbol{\varepsilon}(\mathbf{x})]. \tag{5.5}$$

Here, $E[\cdot]$ is the expectation operator w.r.t the joint probability measure of all the microstructural random fields, $\mathbf{C}^{\text{eff}}(\mathbf{x})$ is the local effective elasticity matrix at the macroscopic material point, $\mathbf{x} \in \mathbf{D}$, and finally, $\boldsymbol{\sigma}(\mathbf{x})$ and $\boldsymbol{\varepsilon}(\mathbf{x})$ are, respectively, the vector-valued random field representations of the second-order tensor-valued random stress and strain fields that depend on the underlying microstructural random fields.

Computation of $E[\boldsymbol{\sigma}(\mathbf{x})]$ and $E[\boldsymbol{\varepsilon}(\mathbf{x})]$ requires *joint* probabilistic characterization of all the microstructural random fields. By the assumption of spatial homogeneity on the microstructural random fields w.r.t. the mean of $\boldsymbol{\varepsilon}(\mathbf{x})$, (5.5) simplifies to,

$$E[\boldsymbol{\sigma}] = \mathbf{C}^{\text{eff}}\, E[\boldsymbol{\varepsilon}], \tag{5.6}$$

showing the invariance w.r.t. the spatial translation, thus, resulting in the effective elasticity matrix valid for the entire domain, $\mathbf{D}$. In (5.6) the left-hand-side (lhs) follows from the fact that the assumption of spatial homogeneity w.r.t. the mean of $\boldsymbol{\varepsilon}(\mathbf{x})$ immediately implies the same w.r.t. the mean of $\boldsymbol{\sigma}(\mathbf{x})$ because the relationship in (5.5) involves constant coefficients that do not depend on $\boldsymbol{\varepsilon}(\mathbf{x})$.

Let us further assume that the microstructural random fields are ergodic w.r.t. the mean of $\boldsymbol{\varepsilon}(\mathbf{x})$ (or $\boldsymbol{\sigma}(\mathbf{x})$) so that the local spatial fluctuation over any one sample is identical to the statistical fluctuation over

a single neighborhood in an ensemble of samples. Then, the ensemble average in (5.6) can be replaced by the volume average over an infinite domain of heterogeneous material,

$$\begin{aligned} \lim_{V\to\infty} \frac{1}{V} \int_V \boldsymbol{\sigma}(\mathbf{x})\, d\mathbf{x} &= \mathbf{C}^{\text{eff}} \lim_{V\to\infty} \frac{1}{V} \int_V \boldsymbol{\varepsilon}(\mathbf{x})\, d\mathbf{x} \\ \langle \boldsymbol{\sigma} \rangle_V &= \mathbf{C}^{\text{eff}} \langle \boldsymbol{\varepsilon} \rangle_V, \end{aligned} \tag{5.7}$$

in which $\langle \cdot \rangle_V$ represents the volume average over V as $V \longrightarrow \infty$. Equation (5.7) recovers the most commonly used definition of $\mathbf{C}^{\text{eff}}$ based on classical RVE with V being a reasonably large material volume and manifesting the underlying microstructural characteristics of the heterogeneous material at *any* macroscopic point, $\mathbf{x} \in \mathbf{D}$. Then, carrying out the averaging operation in (5.7) over a rather finite volume, V, results in $\mathbf{C}^{\text{eff}}$ within a given degree of accuracy.

The size of the resulting RVE defined by $L_{\text{meso}} \approx V^{1/d}$ dictates the minimum size beyond which the continuum theory based on a fictitious homogeneous material, whose property is defined by $\mathbf{C}^{\text{eff}}$, is no longer valid. One of the significance of L_{meso} is that a finite element (FE) model with material property defined by $\mathbf{C}^{\text{eff}}$ and mesh size no smaller than L_{meso} can be used as a proxy for a detailed fine-scale FE model with the actual heterogeneous material property in the following sense. The mesh size of the later FE model must be sufficiently smaller than L_{meso} in order to accurately capture the actual heterogeneous material property. The response of the later FE model at a point $\mathbf{x} \in \mathbf{D}$ would be same, within a given accuracy, as the homogenized or averaged (averaged over V) response of the former FE model at the same point.

While considering a large V helps one to reduce the variability (typically due to data error) in $\mathbf{C}^{\text{eff}}$, there exists a different notion of $\mathbf{C}^{\text{eff}}$ that allows homogenization of the heterogeneous material at a remarkably small length scale by only insisting that the mean of $\mathbf{C}^{\text{eff}}$ be captured accurately while compromising on the variability in $\mathbf{C}^{\text{eff}}$ [DW96, Gus97]. The present work can be readily employed to extend this concept, developed within parametric framework, to the nonparametric framework as explained in section 5.4.1 while concurrently characterizing the data error and the modeling error.

The two bounds, $C_{\boldsymbol{\sigma}}^{\text{app}}$ and $C_{\boldsymbol{\varepsilon}}^{\text{app}}$, might be given by, respectively, Reuss and Voigt bounds. In fact, if the volume element, V, can be assumed to contain homogeneous and linear elastic constituent phases with perfectly bonded interfaces, then $C_{\boldsymbol{\sigma}}^{\text{app}}$ and $C_{\boldsymbol{\varepsilon}}^{\text{app}}$ refer to the Reuss and Voigt bounds, respectively [Hil63, Hue90]. Here, perfectly bonded interfaces essentially imply no defects, no slips and continuity of

displacement and traction across interfaces, but no assumptions about the characteristic of stress gradients at the interfaces are made. These bounds are *independent* of the detailed microtextural features and depend only on the volume fractions and elasticity matrices of the constituent phases given by [Hil63, NNH99, p. 209-213],

$$(C^{\text{app}}_{\boldsymbol{\sigma}})^{-1} = \sum_{i=1}^{n_p} v_i (C^{(i)})^{-1}, \quad \text{and} \quad C^{\text{app}}_{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n_p} v_i C^{(i)}, \tag{5.8}$$

in which v_i and $C^{(i)}$ are, respectively, volume fraction and elasticity matrix of the i-th phase with n_p being the number of phases identified in V and $\sum_{i=1}^{n_p} v_i = 1$. Clearly, substantial experimental, modeling and numerical details are not required in obtaining these two bounds. Only a minimal information about the microstructural features is required to obtain these two bounds. However, the "gap" between these bounds may be large if the constituent phases vary considerably from rigid phase to weak phase in the sense that some quantitative measure of the magnitude of the constituent elasticity matrices (for example, trace of the matrix) representing the strength of the constituent phases vary from 0 to ∞. In general, $C^{\text{app}}_{\boldsymbol{\sigma}}$ and $C^{\text{app}}_{\boldsymbol{\varepsilon}}$, respectively, underestimate and overestimate $\mathbf{C}^{\text{eff}}$ [Hil63]. In the probabilistic characterization of $\mathbf{C}^{\text{eff}}$ as developed in section 5.3, this particular issue has been tackled by imposing two constraints that enforce negligible probability mass around the boundaries of the support of the resulting pdf of the matrix-variate random variable, $\mathbf{C}^{\text{eff}}$. This will guarantee negligible realizations of $\mathbf{C}^{\text{eff}}$ near $C^{\text{app}}_{\boldsymbol{\sigma}}$ and $C^{\text{app}}_{\boldsymbol{\varepsilon}}$ from the resulting pdf.

The nonparametric notion of $\mathbf{C}^{\text{eff}}$ stems from the fact that the entire matrix, $\mathbf{C}^{\text{eff}}$, is characterized by the resulting pdf estimate. Individual characterization of several random system parameters, for example, Young's modulus, Poisson's ratio etc., is neither required nor the goal of such nonparametric formulation. It, therefore, implies that a typical sparsity structure, that is observed for a *parametric* $\mathbf{C}^{\text{eff}}$, is not preserved in the individual realization of nonparametric $\mathbf{C}^{\text{eff}}$ that are sampled from the resulting pdf estimate. Nevertheless, if the mean of $\mathbf{C}^{\text{eff}}$ is enforced while estimating the pdf and it shows some sparsity structure, then the resulting pdf, of course, yields the same mean matrix with the same sparsity structure even if the individual realization of $\mathbf{C}^{\text{eff}}$ does not display any sparsity structure. The bounds in (5.1) or (5.8) can be characterized either in a nonparametric or a parametric sense.

Contrary to the classical (first-order) homogenization scheme, a second-order homogenization scheme based on first order spatial-gradient of the macroscopic strain by using Taylor series expansion has recently been proposed [KGB02, KGB04]. In another recent work [MVL06], [LKP06, Chapter 9], the

microscopic strain-gradient information, instead of the macroscopic strain-gradient information, is used. In the present work, the effective elasticity matrix based on only the classical homogenization (within nonparametric formalism) is considered. Nevertheless, the probabilistic formulation as described in the next section would still be applicable for the positive definite matrices associated with the second-order homogenization provided the suitable lower and upper bounds associated with these matrices are available.

5.3 Probability Model for Positive Definite and Bounded Random Matrix

Let $\mathbb{M}_N^S(\mathbb{R})$ be the set of all real symmetric matrices of size $N \times N$ and $\mathbb{M}_N^+(\mathbb{R}) \subset \mathbb{M}_N^S(\mathbb{R})$ be the set of symmetric positive definite real matrices of size $N \times N$. Then, $\mathbf{C} \in \mathbb{M}_N^+(\mathbb{R})$ a.s. and $\mathbf{0} < C_l \in \mathbb{M}_N^+(\mathbb{R})$ and $C_u \in \mathbb{M}_N^+(\mathbb{R})$.

Let $\mathcal{C}$ be the set of all real symmetric positive definite matrices of size $N \times N$ bounded in the sense as defined by (5.1), i.e., $\mathcal{C} = \{C \in \mathbb{M}_N^+(\mathbb{R}) : C_l < C < C_u\}$. Denote the probability space, on which $\mathbf{C}$ is defined, by $(\mathcal{C}, \mathcal{F}, P)$ in which $\mathcal{F}$ represents the $\sigma-$algebra of subsets of $\mathcal{C}$ and P represents the probability measure on $\mathcal{F}$. Assume that P admits a pdf, $p_{\mathbf{C}} : \mathcal{C} \longrightarrow \mathbb{R}^+ =]0, \infty[$, that is supported on $\mathcal{C}$, $\text{supp}(\mathbf{C}) = \{C : p_{\mathbf{C}}(C) > 0\} = \mathcal{C}$. Therefore, $dP(C) = p_{\mathbf{C}}(C)\, dC$, in which dC is the volume element on $\mathbb{M}_N^S(\mathbb{R})$ given by,

$$dC = \prod_{1 \le i \le j \le N} dC_{ij},$$

with C_{ij} being the (i, j)-th element of C. In the following, $p_{\mathbf{C}}$ is estimated by having recourse to the MaxEnt principle [Jay57a, Jay57b, Kap89, KK92].

Given information, the MaxEnt principle allows one to estimate the pdf of a random variate that is least committal to the unavailable information and most consistent with the partial knowledge available about the quantity modeled as random variate. This is achieved by extending the unique concept of entropy, that was proposed by Shannon [Sha48] in the context of discrete random variable, to the continuous random variate, and maximizing it subjected to the available information [Kap89, KK92]. The entropy of a pdf can be treated as a measure of uncertainty associated with the pdf, i.e., it is a quantitative measure

of ignorance in the state of our knowledge about the quantity modeled as random variate. Therefore, maximizing the entropy (uncertainty) of a pdf defined by,

$$H(p_{\mathbf{C}}) = -\int_{\mathcal{C}} p_{\mathbf{C}}(C)\ \ln[p_{\mathbf{C}}(C)]\, dC, \tag{5.9}$$

subjected to meaningful constraints cast from the available information, yields the sought-after MaxEnt pdf estimate.

In section 5.3.1, the MaxEnt principle is first employed to estimate $p_{\mathbf{C}}$ by having recourse to the two constraints that facilitate negligible probability mass around C_l and C_u resulting in matrix-variate beta type I distribution. The additional information about the ensemble average or mean of $\mathbf{C}$ is used next to estimate an updated pdf in section 5.3.2. The later distribution is known as matrix-variate Kummer-Beta distribution. Simulation of $\mathbf{C}$ from both the distributions are highlighted. A comparison with the Wishart distribution that results from Soize's work is also noted.

5.3.1 Matrix Variate Beta Type I Distribution

The pdf, $p_{\mathbf{C}}$, is determined by solving the following MaxEnt problem,

minimize $\quad [-H(p_{\mathbf{C}})]$

subject to

$$\int_{\mathcal{C}} p_{\mathbf{C}}(C)\, dC \ = \ 1, \tag{5.10}$$

$$E\left\{\ln[\det(\mathbf{C} - C_l)]\right\} \ = \ \int_{\mathcal{C}} \ln[\det(C - C_l)]\, p_{\mathbf{C}}(C)\, dC \ = \ c_l, \tag{5.11}$$

$$E\left\{\ln[\det(C_u - \mathbf{C})]\right\} \ = \ \int_{\mathcal{C}} \ln[\det(C_u - C)]\, p_{\mathbf{C}}(C)\, dC \ = \ c_u, \tag{5.12}$$

where c_l and c_u either are assumed to be known and consistent or need to be estimated from the samples, $C^{(1)}, \cdots, C^{(n)}$, of $\mathbf{C}$ that are assumed to be available (see section 5.4.1-5.4.2 for a scheme that can be readily implemented by using a combination of experimental and computational techniques to obtain samples of nonparametric $\mathbf{C}$ by considering n specimens of the heterogeneous material). The first constraint expresses the normalization of $p_{\mathbf{C}}$. The second and third constraints are properly modified version of

Soize's constraints [Soi00, Soi01a]. These two constraints effectively guarantee that the inverse moments of $(\mathbf{C} - C_l)$ and $(C_u - \mathbf{C})$, respectively, exist a.s. [Soi00, Soi01a] provided $\mid c_l \mid < \infty$ and $\mid c_u \mid < \infty$ which can be assumed without any loss of generality for most of the practical systems. Existences of such inverse moments are feasible if $p_{\mathbf{C}}$ decreases sufficiently in the neighborhood of $C - C_l = \mathbf{0}$ and $C_u - C = \mathbf{0}$, respectively, ensuring negligible probability mass around C_l and C_u [Spa03, p. 184-185].

The above optimization problem can be solved by using the Lagrange multiplier theory. The Lagrangian function associated with this optimization problem is given by,

$$\begin{aligned}\mathcal{L}(p_{\mathbf{C}}, \lambda_l, \lambda_u) &= -H(p_{\mathbf{C}}) + (\lambda_0 - 1)\Big[\int_{\mathcal{C}} p_{\mathbf{C}}(C)\, dC - 1\Big] + \lambda_l\Big[\int_{\mathcal{C}} \ln[\det(C - C_l)] \\ &\quad \times p_{\mathbf{C}}(C)\, dC - c_l\Big] + \lambda_u\Big[\int_{\mathcal{C}} \ln[\det(C_u - C)]\, p_{\mathbf{C}}(C)\, dC - c_u\Big],\end{aligned}$$

in which $(\lambda_0 - 1)$, λ_l and λ_u are Lagrange multipliers. It is shown below that λ_0 depends on λ_l and λ_u, and therefore, λ_0 is not shown in the argument of $\mathcal{L}(\cdot)$ above. By using the theory of calculus of variations, it can be inferred immediately that $p_{\mathbf{C}}$ assumes the form, $p_{\mathbf{C}}(C) = \exp(-\lambda_0)\det(C - C_l)^{a-\frac{1}{2}(N+1)}\ \det(C_u - C)^{b-\frac{1}{2}(N+1)}\, \mathbb{I}_{\mathcal{C}}(C)$, in which $a = (1/2)(N+1) - \lambda_l$ and $b = (1/2)(N+1) - \lambda_u$ can be treated as modified Lagrange multipliers, and $\mathbb{I}_{\mathcal{C}}(\cdot)$ is the indicator function implying that $\mathbb{I}_{\mathcal{C}}(C) = 1$, if $C \in \mathcal{C}$, and $\mathbb{I}_{\mathcal{C}}(C) = 0$, otherwise. Assume that $a > (1/2)(N-1)$ and $b > (1/2)(N-1)$. Then, from the already existing results in the literature of RMT [GN00, Eq. 5.2.4], it can be immediately concluded that $p_{\mathbf{C}}$ is a generalized matrix-variate beta type I density given by,

$$\begin{aligned}p_{\mathbf{C}}(C) &= \frac{\det(C - C_l)^{a-\frac{1}{2}(N+1)}\ \det(C_u - C)^{b-\frac{1}{2}(N+1)}}{\beta_N(a,b)\det(C_u - C_l)^{(a+b)-\frac{1}{2}(N+1)}}\, \mathbb{I}_{\mathcal{C}}(C), \\ &\qquad a > \frac{1}{2}(N-1),\ b > \frac{1}{2}(N-1),\end{aligned} \tag{5.13}$$

that must satisfy (5.10) implying that the normalization constant, $\exp(-\lambda_0)$, is given by, $\exp(-\lambda_0) = 1/[\beta_N(a,b)\det(C_u - C_l)^{(a+b)-\frac{1}{2}(N+1)}]$, showing the dependence of λ_0 on $a = a(\lambda_l)$ and $b = b(\lambda_u)$. Here, $\beta_N(\cdot)$ is the multivariate beta function given by [GN00, p. 20],

$$\beta_N(x,y) \equiv \int_{\mathcal{I}} \det(U)^{x-\frac{1}{2}(N+1)}\ \det(\mathbf{I} - U)^{y-\frac{1}{2}(N+1)}\, dU = \frac{\Gamma_N(x)\Gamma_N(y)}{\Gamma_N(x+y)}, \tag{5.14}$$

in which $\Re(x) > (1/2)(N-1)$, $\Re(y) > (1/2)(N-1)$, $\mathbf{I}$ is a $N \times N$ identity matrix, $\mathcal{I} = \{U \in \mathbb{M}_N^+(\mathbb{R}) : \mathbf{0} < U < \mathbf{I}\}$ and $\Gamma_N(\cdot)$ represents the multivariate gamma function given by [GN00, Theorem 1.4.1],

$$\Gamma_N(z) = \pi^{\frac{1}{4}N(N-1)} \prod_{i=1}^{N} \Gamma[z - \frac{1}{2}(i-1)], \quad \Re(z) > \frac{1}{2}(N-1), \tag{5.15}$$

with $\Gamma(\cdot)$ being gamma function defined by $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)dt$, $\Re(z) > 0$ [AS70, Chapter 6]. For integer z, the gamma function, however, reduces to, $\Gamma(z+1) = z!$. Let the PDF associated with pdf in (5.13) be denoted by $GB_N^I(a, b; C_u, C_l)$.

Computation of Parameters of $GB_N^I(a, b; C_u, C_l)$

The two parameters, a and b, are now required to be computed by using (5.11) and (5.12). However, solving these two integral equations for a and b is a formidable problem. An alternative efficient technique is proposed in this section to determine a and b by following the course of solution as already adopted in a scalar-variate problem [Kap89, p. 66-67]. By (5.10), equation (5.13) implies that,

$$\beta_N(a, b) = \int_{\mathcal{C}} \frac{\det(C - C_l)^{a - \frac{1}{2}(N+1)} \ \det(C_u - C)^{b - \frac{1}{2}(N+1)}}{\det(C_u - C_l)^{(a+b) - \frac{1}{2}(N+1)}} \, dC.$$

Differentiating this equation w.r.t. a and b, and substituting (5.13) in (5.11) and (5.12), and subsequently using all the resulting expressions, it can be shown that,

$$\begin{aligned} \frac{\partial \ln[\beta_N(a, b)]}{\partial a} + \ln[\det(C_u - C_l)] &= c_l, \\ \frac{\partial \ln[\beta_N(a, b)]}{\partial b} + \ln[\det(C_u - C_l)] &= c_u. \end{aligned}$$

Solving these two differential equations, instead of the two integral equations, (5.11) and (5.12), is extremely efficient and computationally very cheap since the differentiation of $\ln[\beta_N(a, b)]$ as shown above can be efficiently expressed in terms of the psi or digamma function [AS70, Chapter 6]. This psi function, $\psi(\cdot)$, is defined by the logarithmic derivative of the gamma function as $\psi(z) = d(\ln(\Gamma(z)))/dz$. The two differential equations can be readily cast into a nonlinear least squares problem to solve for a and b with lower bounds of $(1/2)(N-1) + \epsilon$ and $(1/2)(N-1) + \epsilon$, respectively, in which ϵ is a very small number, say, 1×10^{-7}. For MATLAB users, it might be useful to invoke the `lsqnonlin` function

with Levenberg-Marquardt method option 'On' since the Levenberg-Marquardt algorithm is relatively robust against poorer initial guess (but may be slow in convergence, see also section 2.3.3) [DGS08]. To compute the psi function, MATLAB function, `psi`, would be handy as well.

Simulation from $GB_N^I(a, b; C_u, C_l)$

Samples of $\mathbf{C}$ can be digitally generated from $GB_N^I(a, b; C_u, C_l)$ by making use of the theoretical propositions available in the field of RMT. For that, we need the definition of the matrix-variate gamma distribution and two lemmas as outlined next.

Definition 5.3.1 *A random positive definite matrix,* $\mathbf{S}$*, is said to follow the matrix-variate gamma distribution,* $G_N(\alpha, \Lambda_S)$ *parameterized by* α *and* Λ_S*, if its pdf is given by,*

$$p_{\mathbf{S}}(S) = \left\{2^{\alpha N}\ \Gamma_N(\alpha) \det\left(\frac{1}{2}\Lambda_S^{-1}\right)^{\alpha}\right\}^{-1} \det(S)^{\alpha - \frac{1}{2}(N+1)}\ \mathrm{etr}(-\Lambda_S S)\ \mathbb{I}_{\mathbb{M}_N^+(\mathbb{R})}(S), \tag{5.16}$$

in which $\alpha > (1/2)(N-1)$ *is a real number,* $\Lambda_S \in \mathbb{M}_N^+(\mathbb{R})$ *and* $\mathrm{etr}(\cdot)$ *is defined by* $\mathrm{etr}(A) = \exp\{\mathrm{tr}(A)\}$.

This pdf is often known (see, e.g., [Mur82, p. 87], [Mat97, p. 264], [Soi00, Soi01a]) as Wishart density [GN00, Chapter 3]. A 'pure' Wishart density is, however, defined for $m = 2\alpha \geq N$ an integer and $\Lambda_S = (1/2)\Sigma^{-1}$ (e.g., [SK79, Chapter 3], [Mur82, Section 3.2], [GN00, p. 89]). The slightly different parametrization as shown in (5.16) from the usual notation (as in, [SK79, p. 76], [Mur82, p. 85], [Mat97, p. 87], [GN00, p. 87], [And03, Section 7.2]) is adopted here to be consistent and for the sake comparison with the Kummer-Beta distribution as developed in section 5.3.2.

Now, let us introduce the following lemma, based on Barlett's decomposition [Bar33], that is the corner stone of sampling $\mathbf{S}$ from $G_N(\alpha, \Lambda_S)$. Generation of samples of $\mathbf{S} \sim G_N(\alpha, \Lambda_S)$ is an important intermediate step in sampling from the generalized matrix-variate beta type I distribution as shown after this lemma.

Lemma 5.3.2 *Let* $\mathbf{S} \sim G_N(\alpha, \frac{1}{2}\mathbf{I})$ *and* $\mathbf{S} = \mathbf{T}\mathbf{T}^T$*, where* $\mathbf{T}$ *is a lower triangular matrix with its* (i, j)*-th element being given by* $\mathbf{t}_{ij}$ *and* $\mathbf{t}_{ii} > 0$*. Then,* $\mathbf{t}_{ij}$*,* $1 \leq j \leq i \leq N$*, are statistically independent,*

$\mathbf{t}_{ij} \sim N(0,1)$, $1 \leq j < i \leq N$, *and* $\mathbf{t}_{ii}^2 \sim G(\alpha - \frac{1}{2}(i-1), \frac{1}{2})$, $i = 1, \cdots, N$. *Here,* $N(0,1)$ *represents the standard normal distribution and* $G(k,\gamma)$ *represents the gamma distribution whose pdf is given by,*

$$p(t) = \frac{\gamma^k}{\Gamma(k)}\, t^{k-1}\, \exp(-\gamma t)\, \mathbb{I}_{\mathbb{R}^+}(t), \quad k, \gamma > 0.$$

Proof The proof immediately follows from the literature (see, e.g., [SK79, Corollary 3.2.4], [Mur82, Theorem 3.2.14], [GN00, Theorem 3.3.4]) by considering $\alpha = (m/2)$ with the only exception that the chi-squared distribution with $(m - i + 1)$ degrees of freedom as indicated in the literature now needs to be interpreted as gamma distribution with real parameters, $k = (m - i + 1)$ and $\gamma = 1/2$. ∎

In the present work, $(m - i + 1) = (2\alpha - i + 1)$ is allowed to be a real number while the chi-squared distribution is a special case of gamma distribution with positive integer $(m - i + 1)$ and $\gamma = 1/2$ [Fis96, p. 193]. It is emphasized here that no other part of the already available proofs as indicated above need to be changed because of $m = 2\alpha$ being a real number. The decomposition, $\mathbf{S} = \mathbf{T}\mathbf{T}^T$, as shown in Lemma 5.3.2, is followed by the fact that every symmetric positive definite matrix always has a unique Cholesky decomposition [Har97, Theorem 14.5.11].

Since samples of $\mathbf{S} \sim G_N(\alpha, \frac{1}{2}\mathbf{I})$ is required for generating the samples from the generalized matrix-variate beta type I distribution, the procedure of sampling from $G_N(\alpha, \frac{1}{2}\mathbf{I})$ is sketched in Algorithm 5.3.3 next.

Algorithm 5.3.3 *Matrix Variate Gamma Distribution,* $G_N(\alpha, \frac{1}{2}\mathbf{I})$

Input: Dimension of matrix, N, and real parameter $\alpha > (1/2)(N-1)$.

Output: Samples of $\mathbf{S} \sim G_N(\alpha, \frac{1}{2}\mathbf{I})$.

Step 0: Generate statistically independent $\mathbf{t}_{ij}$, $1 \leq j \leq i \leq N$, as follows.

Step 1: $\mathbf{t}_{ij} \sim N(0,1)$, $1 \leq j < i \leq N$.

Step 2: $\mathbf{y}_i \sim G(\alpha - \frac{1}{2}(i-1), \frac{1}{2})$, $i = 1, \cdots, N$. Take $\mathbf{t}_{ii} = \sqrt{\mathbf{y}_i}$.

Step 3: Form $\mathbf{S} = \mathbf{T}\mathbf{T}^T$ to get a sample from $G_N(\alpha, \frac{1}{2}\mathbf{I})$.

Sampling scheme from a scalar-variate gamma distribution as required in **Step** 2 is available in many standard textbook on Monte Carlo (MC) simulation (see, e.g., [Fis96, Section 3.14]). The MATLAB function, `gamrnd`, might be useful here. In fact, the whole algorithm above can be substituted by the recently

introduced MATLAB function, `wishrnd`[2]. Generating a sample from $G_N(\alpha, \Lambda_A)$, when $\Lambda_A \neq (1/2)\mathbf{I}$, is straightforward. Since $\Lambda_A \in \mathbb{M}_N^+(\mathbb{R})$, then $\Lambda_A^{-1} \in \mathbb{M}_N^+(\mathbb{R})$ and Λ_A^{-1} has a Cholesky decomposition, $\Lambda_A^{-1} = L_A L_A^T$, with L_A being a lower triangular matrix. Then, $\mathbf{A} = (L_A/\sqrt{2})\mathbf{S}(L_A^T/\sqrt{2})$, in which $\mathbf{S} \sim G_N(\alpha, \frac{1}{2}\mathbf{I})$, follows $G_N(\alpha, \Lambda_A)$ ([Mur82, Theorem 3.2.5], [GN00, Theorem 3.3.1 or Theorem 3.3.11]). This feature has been exploited well by Soize [Soi00, Soi01a] to digitally simulate samples from the Wishart or matrix-variate gamma distribution.

Now, the final lemma required as a last piece for chalking out a sampling scheme for the generalized matrix-variate beta type I distribution is introduced.

Lemma 5.3.4 *Let* $\mathbf{S}_1 \sim G_N(a, \Lambda_S)$ *and* $\mathbf{S}_2 \sim G_N(b, \Lambda_S)$ *be statistically independent. Then,* $\mathbf{U} = (\mathbf{S}_1 + \mathbf{S}_2)^{-1/2}\mathbf{S}_1((\mathbf{S}_1 + \mathbf{S}_2)^{-1/2})^T \sim GB_N^I(a, b; \mathbf{I}, \mathbf{0})$*, where* $A^{1/2}(A^{1/2})^T = A$ *is the symmetric matrix square root factorization of* $A \in \mathbb{M}_N^+(\mathbb{R})$.

Proof See p. 149-151 of the monograph by Mathai [Mat97]. ∎

The symmetric matrix square root factorization is valid for every $A \in \mathbb{M}_N^+(\mathbb{R})$ [Har97, Theorem 21.9.1]. It may be noted here that the factorization, $(\mathbf{S}_1+\mathbf{S}_2)^{1/2}((\mathbf{S}_1+\mathbf{S}_2)^{1/2})^T = (\mathbf{S}_1+\mathbf{S}_2)$, may be safely replaced by the Cholesky factorization, $\mathbf{T}\mathbf{T}^T = (\mathbf{S}_1 + \mathbf{S}_2)$ [Mur82, Theorem 3.3.1]. In fact, any reasonable nonsingular factorization should work depending on $(\mathbf{S}_1 + \mathbf{S}_2)$ [SK79, Theorem 3.6.3], [GN00, p. 186]. It is common in the literature of RMT to denote the distribution of $\mathbf{U} \sim GB_N^I(a, b; \mathbf{I}, \mathbf{0})$, supported on $\mathcal{I}$, by $B_N^I(a, b)$ referred hereafter as the standard matrix-variate beta type I distribution. The distribution of $\mathbf{U}$ is free of Λ_S, thus often referring it as a density-free approach to the matrix-variate beta distribution [Mit70, Kha70]; compare also Example 1.15 and Example 2.11 of the monograph by Mathai [Mat97]. Finally, if $\mathbf{U} \sim B_N^I(a, b)$ and $\mathbf{C}$ is defined by,

$$\mathbf{C} = (C_u - C_l)^{\frac{1}{2}}\ \mathbf{U}\ (C_u - C_l)^{\frac{1}{2}} + C_l, \tag{5.17}$$

then $\mathbf{C} \sim GB_N^I(a, b; C_u, C_l)$ [GN00, Theorem 5.2.1].

Now, it is only a matter of putting things together as just described in this subsection to prescribe a sampling technique for $GB_N^I(a, b; C_u, C_l)$.

[2]The Algorithm 5.3.3, however, generates samples of relatively better statistical qualities. If $\mathbf{S} \sim G_3(\alpha, \frac{1}{2}\mathbf{I})$, then $E\{\mathbf{S}\} = 2\alpha\mathbf{I}$ by (5.42). Consider $\alpha = 1.0653$. A set of 100 samples of $\mathbf{S}$ based on Algorithm 5.3.3 results in a sample mean estimate, $\underline{S}$, with relative mean-squared error (see (5.51) for its definition) relative to $E\{\mathbf{S}\}$ as $\mathrm{relMSE}(\underline{S}, E[\mathbf{S}]) = 0.0163\%$ as opposed to $\mathrm{relMSE}(\underline{S}, E[\mathbf{S}]) = 1.3054\%$ based on 100 samples generated by using `wishrnd` function.

Algorithm 5.3.5 *Matrix Variate Generalized Beta Type I Distribution,* $GB_N^I(a, b; C_u, C_l)$

Input: Dimension of matrix, N, the real parameters $a, b > (1/2)(N-1)$ and the bounds $C_u, C_l \in \mathbb{M}_N^+(\mathbb{R})$.

Output: Samples of $\mathbf{C} \sim GB_N^I(a, b; C_u, C_l)$.

Step 1: Generate statistically independent $\mathbf{S}_1 \sim G_N(a, \frac{1}{2}\mathbf{I})$ and $\mathbf{S}_2 \sim G_N(b, \frac{1}{2}\mathbf{I})$ by employing Algorithm 5.3.3.

Step 2: Form $\mathbf{U} = (\mathbf{S}_1 + \mathbf{S}_2)^{-1/2}\mathbf{S}_1((\mathbf{S}_1 + \mathbf{S}_2)^{-1/2})^T \sim B_N^I(a, b)$.

Step 3: Get a sample $\mathbf{C} \sim GB_N^I(a, b; C_u, C_l)$ by employing (5.17) based on $\mathbf{U}$.

The MATLAB users may find the function, `sqrtm`, useful in executing **Step** 2-3 above.

Remark 5.3.6 *If* $N = 1$*, then the matrix-variate gamma distribution* $G_N(\alpha, \frac{1}{2}\mathbf{I})$ *reduces to [KK92, p.* 67*], [DSC04] the scalar-variate gamma distribution* $G(\alpha, \frac{1}{2})$*, and the matrix-variate beta type I distribution* $B_N^I(a, b)$ *reduces to [KK92, Section* 2.6.1*] the scalar-variate beta type I distribution* $B(a, b)$ *whose pdf is defined by,*

$$p(u) = \frac{1}{\beta(a, b)}\, u^{a-1}\,(1-u)^{b-1}\,\mathbb{I}_{(0,1)}(u), \quad a, b > 0,$$

in which $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$.

Remark 5.3.7 *In the context of the effective material property, we draw attention of the readers to an interesting previous work [OS01, p.* 124-131*], [OS08, Section* 8.1.2*] in which a scalar-variate beta distribution was proposed to be the most convenient distribution, within a parametric formulation, to characterize the trace of* $\mathbf{C}$ *(note that the pdf in the above literature is wrongly printed; the correct form of the pdf can be readily obtained from (5.13) by substituting* $N = 1$ *and the appropriate upper and lower bounds). The lower and upper bounds were grossly assumed to be given by, respectively, the most flexible (matrix phase) and the most stiff (inclusion) material properties of the constituents of the multiphase material. Similar results on the compliance or flexibility matrix were also reported.*

5.3.2 Matrix Variate Kummer-Beta Distribution

It is assumed here that the ensemble average, $\underline{C}$, of $\mathbf{C}$ is known or can be estimated from the available samples, $C^{(1)}, \cdots, C^{(n)}$, of $\mathbf{C}$. Therefore, the pdf, $p_{\mathbf{C}}$, can be determined by solving a similar MaxEnt problem, as formulated in section 5.3.1, along with the following additional constraint,

$$E[\mathbf{C}] = \int_{\mathcal{C}} C \; p_{\mathbf{C}}(C)\, dC = \underline{C} \in \mathbb{M}_N^+(\mathbb{R}). \tag{5.18}$$

Following the Lagrange multiplier method as employed in section 5.3.1 and assuming that the two modified Lagrange multipliers, a and b, associated with the constraints (5.11) and (5.12) are greater than $(1/2)(N-1)$, it can be immediately concluded that $p_{\mathbf{C}}$ is given by,

$$\begin{aligned} p_{\mathbf{C}}(C) \;\; = \;\; & C(a, b, \Lambda_C, C_u, C_l)\ \text{etr}(-\Lambda_C C)\ \det(C - C_l)^{a - \frac{1}{2}(N+1)} \\ & \times \det(C_u - C)^{b - \frac{1}{2}(N+1)}\ \mathbb{I}_{\mathcal{C}}(C), \\ & a > \frac{1}{2}(N-1),\ b > \frac{1}{2}(N-1),\ \Lambda_C \in \mathbb{M}_N^S(\mathbb{R}), \end{aligned} \tag{5.19}$$

in which $C(a, b, \Lambda_C, C_u, C_l)$ is the normalization constant whose explicit expression is given later in section 5.3.2 and $\Lambda_C \in \mathbb{M}_N^S(\mathbb{R})$ is the matrix-valued Lagrange multiplier associated with (5.18). This pdf has recently been introduced and studied by Nagar and Gupta [NG02]. The associated PDF is referred as the generalized matrix-variate Kummer-Beta distribution to be denoted hereafter by $GKB_N(a, b, \Lambda_C; C_u, C_l)$.

Computation of Parameters of $GKB_N(a, b, \Lambda_C; C_u, C_l)$

Direct computation of the parameters, a, b and Λ_C, of $GKB_N(a, b, \Lambda_C; C_u, C_l)$ is likely to cause a computer overflow problem in the ensuing optimization technique since elements of the elasticity matrix associated with a practical system often tend to have values of high order, say, 1×10^{10}. An alternative MaxEnt optimization problem formulated in terms of the standard matrix-variate Kummer-Beta distribution, that is supported on $\mathcal{I}$, is, therefore, recommended to circumvent this machine overflow problem.

If $\mathbf{C} \sim GKB_N(a, b, \Lambda_C; C_u, C_l)$ and $\mathbf{U}$ is defined by (5.17), i.e., $(C_u - C_l)^{-\frac{1}{2}}\ (\mathbf{C} - C_l)\ (C_u - C_l)^{-\frac{1}{2}}$, then $\mathbf{U}$ follows the standard matrix-variate Kummer-Beta distribution, to be denoted henceforth as $KB_N(a, b, \Lambda_U)$, with its pdf, $p_{\mathbf{U}}$, given by [NG02],

$$\begin{aligned} p_{\mathbf{U}}(U) &= K(a, b, \Lambda_U)\ \text{etr}(-\Lambda_U U)\ \det(U)^{a-\frac{1}{2}(N+1)} \\ &\qquad \times \det(\mathbf{I} - U)^{b-\frac{1}{2}(N+1)}\ \mathbb{I}_{\mathcal{I}}(U), \\ &\qquad a > \frac{1}{2}(N-1),\ b > \frac{1}{2}(N-1),\ \Lambda_U \in \mathbb{M}_N^S(\mathbb{R}). \end{aligned} \tag{5.20}$$

Here, Λ_U is related to Λ_C by $\Lambda_U = [(C_u - C_l)^{1/2}\Lambda_C(C_u - C_l)^{1/2}] \in \mathbb{M}_N^S(\mathbb{R})$ and $K(a, b, \Lambda_U)$ is the normalization constant given by,

$$\{K(a, b, \Lambda_U)\}^{-1} = \int_{\mathcal{I}} \text{etr}(-\Lambda_U U) \det(U)^{a-\frac{1}{2}(N+1)}\ \det(\mathbf{I} - U)^{b-\frac{1}{2}(N+1)}\, dU \tag{5.21}$$

$$\Rightarrow \{K(a, b, \Lambda_U)\}^{-1} = \beta_N(a, b)\ {}_1F_1(a; a + b; -\Lambda_U), \tag{5.22}$$

in which ${}_1F_1(\cdot)$ is the confluent hypergeometric function of matrix argument [Mur82, Chapter 7], [Mat97, Section 5.2], [GN00, Section 1.6] defined by,

$$\begin{aligned} {}_1F_1(\alpha; \gamma; X) &= \frac{1}{\beta_N(\alpha, \gamma - \alpha)} \int_{\mathcal{I}} \text{etr}(SX)\ \det(S)^{\alpha-\frac{1}{2}(N+1)} \\ &\qquad \times \det(\mathbf{I} - S)^{\gamma-\alpha-\frac{1}{2}(N+1)}\, dS, \end{aligned}$$

with $\Re(\alpha) > (1/2)(N-1)$, $\Re(\gamma) > (1/2)(N-1)$ and $\mathbf{X}$ being a $N \times N$ complex symmetric matrix. The computation of ${}_1F_1(a; a + b; -\Lambda_U)$ was a hopeless task until very recently even in the simplest cases [BW02] before the arrival of the excellent algorithm by Koev and Edelman [KE06]. We refer the readers to these literatures for a discussion and the numerical algorithm for computing ${}_1F_1(a; a + b; -\Lambda_U)$. Finally, it can also be shown [NG02] that the normalization constant, $C(a, b, \Lambda_C, C_u, C_l)$, of $GKB_N(a, b, \Lambda_C; C_u, C_l)$ is related to $K(a, b, \Lambda_U)$ through $C(a, b, \Lambda_C, C_u, C_l) = K(a, b, \Lambda_U)\ \text{etr}(\Lambda_C C_l)\ \det(C_u - C_l)^{-(a+b)+(N+1)/2}$.

In short, since $KB_N(a, b, \Lambda_U)$ and $GKB_N(a, b, \Lambda_C; C_u, C_l)$ are directly related as just discussed, the following MaxEnt optimization problem in terms of $\mathbf{U}$ would be solved instead of the MaxEnt problem as originally formulated in terms of $\mathbf{C}$.

$$\begin{aligned}
&\text{minimize} \qquad [-H(p_{\mathbf{U}})] \\
&\text{subject to} \\
&\int_{\mathcal{I}} p_{\mathbf{U}}(U)\, dU \;=\; 1, && (5.23)\\
&\int_{\mathcal{I}} \ln[\det(U)]\, p_{\mathbf{U}}(U)\, dU \;=\; u_l, && (5.24)\\
&\int_{\mathcal{I}} \ln[\det(\mathbf{I} - U)]\, p_{\mathbf{U}}(U)\, dU \;=\; u_u, && (5.25)\\
&\int_{\mathcal{I}} U\; p_{\mathbf{U}}(U)\, dU = \underline{U} \in \mathbb{M}_N^+(\mathbb{R}). && (5.26)
\end{aligned}$$

The integral domain, $\mathcal{C}$, as originally used in defining the entropy of a pdf in (5.9) also needs to be replaced with $\mathcal{I}$. By using (5.17), u_l, u_u and $\underline{U}$ can be readily extracted from the information already available for the ensemble of $\mathbf{C}$ as shown below,

$$\begin{aligned}
u_l &= c_l - \ln[\det(C_u - C_l)] && (5.27)\\
u_u &= c_u - \ln[\det(C_u - C_l)] && (5.28)\\
\underline{U} &= (C_u - C_l)^{-\frac{1}{2}}\; (\underline{C} - C_l)\; (C_u - C_l)^{-\frac{1}{2}}. && (5.29)
\end{aligned}$$

Solving the above MaxEnt optimization problem results in an estimate of the pdf of $\mathbf{U}$ as shown in (5.20).

The parameters, a, b and $\Lambda_{\underline{U}}$, can be determined by solving (5.24)-(5.26). Solving these integral equations, however, is a notoriously challenging problem even with the cutting-edge of computer hardware and computing techniques. An alternative scheme is, therefore, described next.

Before dealing with the mean matrix constraint defined by (5.26), let us first consider (5.24) and (5.25). Following a scheme, that is similar in essence as already suggested in section 5.3.1, is adopted

here again. Differentiating both sides of (5.21) w.r.t. a and b, and substituting (5.20) in (5.24) and(5.25), and subsequently using all the equations, it can be shown that,

$$\frac{\partial \ln[\beta_N(a,b)]}{\partial a} + \frac{\partial \ln[{}_1F_1(a;a+b;-\Lambda_{\underline{U}})]}{\partial a} = u_l, \tag{5.30}$$

$$\frac{\partial \ln[\beta_N(a,b)]}{\partial b} + \frac{\partial \ln[{}_1F_1(a;a+b;-\Lambda_{\underline{U}})]}{\partial b} = u_u. \tag{5.31}$$

In deriving the final forms as shown above, the identity in (5.22) is also used.

Now, let us tackle the mean constraint defined by (5.26). It requires the characteristic function, $\phi_{\mathbf{U}}(\Theta)$, of $\mathbf{U}$ defined by,

$$\begin{aligned}
\phi_{\mathbf{U}}(\Theta) &\equiv E[\mathrm{etr}(\iota\Theta U)] \\
&= K(a,b,\Lambda_U) \int_{\mathcal{I}} \mathrm{etr}(\iota\Theta U - \Lambda_U U)\ \det(U)^{a-\frac{1}{2}(N+1)}\ \det(\mathbf{I}-U)^{b-\frac{1}{2}(N+1)}\, dU \\
&= \frac{{}_1F_1(a;a+b;\iota\Theta-\Lambda_{\mathbf{U}})}{{}_1F_1(a;a+b;-\Lambda_{\mathbf{U}})}, \quad \iota=\sqrt{-1}, \quad \Theta \in \mathbb{M}_N^S(\mathbb{R}),
\end{aligned} \tag{5.32}$$

where the second equality follows by (5.20) and the last equality follows by (5.21) and (5.22). From the above characteristic function, it immediately follows that the (i,j)-th element of $E[\mathbf{U}]$ is given by,

$$E[\mathbf{U}_{ij}] = \frac{-\iota}{2-\delta_{ij}} \left[\frac{\partial \phi_{\mathbf{U}}(\Theta)}{\partial \Theta_{ij}}\right]_{\Theta=\mathbf{0}} \tag{5.33}$$

$$= \frac{-\iota}{2-\delta_{ij}} \frac{1}{{}_1F_1(a;a+b;-\Lambda_{\mathbf{U}})} \left[\frac{\partial\{{}_1F_1(a;a+b;\iota\Theta-\Lambda_{\mathbf{U}})\}}{\partial \Theta_{ij}}\right]_{\Theta=\mathbf{0}}, \tag{5.34}$$

in which δ_{ij} is Kronecker's delta defined by $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ otherwise.

Solving now (5.30), (5.31) and equating the right-hand-side (rhs) of (5.34) to $\underline{U}$, the parameters, a, b and $\Lambda_{\mathbf{U}}$, can be determined. This can be effectively performed by solving the following nonlinear constrained minimization problem,

$$\min_{\substack{a>\frac{1}{2}(N-1),\, b>\frac{1}{2}(N-1),\\ \Lambda_{\mathbf{U}}\in\mathbb{M}_N^S(\mathbb{R})}} \quad \epsilon_1^2 + \epsilon_2^2 + \|E[\mathbf{U}] - \underline{U}\|_F^2 \tag{5.35}$$

subject to

$$E[\mathbf{U}] \in \mathbb{M}_N^+(\mathbb{R}). \tag{5.36}$$

Here, ϵ_1 and ϵ_2 are, respectively, the residuals of (5.30) and (5.31) defined by,

$$\epsilon_1^2 = \left\{ u_l - \frac{\partial \ln[\beta_N(a,b)]}{\partial a} + \frac{\partial \ln[{}_1F_1(a;a+b;-\Lambda_{\underline{U}})]}{\partial a} \right\}^2, \tag{5.37}$$

$$\epsilon_2^2 = \left\{ u_u - \frac{\partial \ln[\beta_N(a,b)]}{\partial b} + \frac{\partial \ln[{}_1F_1(a;a+b;-\Lambda_{\underline{U}})]}{\partial b} \right\}^2. \tag{5.38}$$

In the above minimization problem, while the differentiation of $\ln[\beta_N(a,b)]$ can be determined readily as indicated in section 5.3.1, the logarithmic derivative of the hypergeometric function is not available in terms of any known mathematical functions, and therefore, a numerical technique, e.g., a two-sided classical finite-difference (FD) approximation [Spa03, Section 6.3], should be employed to compute this derivative. For instance, the FD approximation of the second term of the lhs of (5.30) is given below,

$$\begin{aligned} \frac{\partial \ln[{}_1F_1(a;a+b;-\Lambda_{\underline{U}})]}{\partial a} &= \frac{1}{{}_1F_1(a;a+b;-\Lambda_{\underline{U}})} \frac{\partial\ {}_1F_1(a;a+b;-\Lambda_{\underline{U}})}{\partial a} \\ &\approx \frac{1}{{}_1F_1(a;a+b;-\Lambda_{\underline{U}})} \left(\frac{{}_1F_1^{(+)} - {}_1F_1^{(-)}}{2\Delta} \right), \end{aligned}$$

in which ${}_1F_1^{(+)} = {}_1F_1[(a+\Delta);(a+\Delta)+b;-\Lambda_{\underline{U}}]$ and ${}_1F_1^{(-)} = {}_1F_1[(a-\Delta);(a-\Delta)+b;-\Lambda_{\underline{U}}]$ with Δ being a very small number, say, 1×10^{-6}. Similar expressions exist for the second term of the lhs of (5.31) and the square-bracketed quantity of (5.34). This specific step of computations of several derivatives, as would be required to feed in an optimization algorithm for a set of consistent values of a, b and Λ_{U}, can be readily performed in parallel by using the above FD approximation. Here, the computation of ${}_1F_1$ can be executed by using the algorithm of Koev and Edelman [KE06] available in public domain[3].

Since the commonly available optimization algorithms are typically formulated in terms of vector-valued parameters, the matrix-valued parameter, $\Lambda_{\mathrm{U}} \in \mathbb{M}_N^S(\mathbb{R})$, needs to be mapped to a suitable vector

[3] `http://www-math.mit.edu/~plamen/software/mhgref.html`; more recent and updated version of the code was kindly made available to the authors by Professor Plamen Koev.

before invoking such optimization algorithms. This can be achieved by introducing the vec-operator for a symmetric matrix, $X = [x_{ij}] \in \mathbb{M}_N^S(\mathbb{R})$, as defined below, $\text{vec} : \mathbb{M}_N^S(\mathbb{R}) \longrightarrow \mathbb{R}^{N(N+1)/2}$,

$$\text{vec}(X) = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{22} \\ \vdots \\ x_{1N} \\ \vdots \\ x_{NN} \end{pmatrix}. \tag{5.39}$$

However, before computing ${}_1F_1$, it is also necessary to map $\text{vec}(\Lambda_{\underline{U}})$ back to $\Lambda_{\underline{U}} \in \mathbb{M}_N^S(\mathbb{R})$ by employing an inverse vec^{-1}-operator, $\text{vec}^{-1} : \mathbb{R}^{N(N+1)/2} \longrightarrow \mathbb{M}_N^S(\mathbb{R})$. One convenient way, among many other possibilities, to tackle the positive definite nonlinear constraint defined by (5.36) is enforcement of the following inequality constraint,

$$\lambda_{\min}(a, b, \Lambda_{\underline{U}}) > 0. \tag{5.40}$$

Here, $\lambda_{\min}(a, b, \Lambda_{\underline{U}})$ is the minimum eigenvalue of $E[\mathbf{U}]$, in which the dependence on the current values of a, b and $\Lambda_{\underline{U}}$ is made explicit. Finally, the MATLAB users may like to use the function, `fmincon`, with Levenberg-Marquardt method option '`On`', to solve the above minimization problem defined by (5.35), (5.37), (5.38) and (5.40).

5.3.3 Simulation from $GKB_N(a, b, \Lambda_C; C_u, C_l)$

Once the parameters, a, b and Λ_U, are determined, the samples of $\mathbf{U} \sim KB_N(a, b, \Lambda_U)$ need to be generated. Samples of $\mathbf{C} \sim GKB_N(a, b, \Lambda_C; C_u, C_l)$ then can be readily obtained from the samples of $\mathbf{U}$ by employing (5.17).

It should be noted that $KB_N(a, b, \Lambda_U)$ is a joint pdf of the functionally independent elements, $\{\mathbf{u}_1, \mathbf{u}_{12}, \mathbf{u}_{22}, \cdots, \mathbf{u}_{1N}, \cdots, \mathbf{u}_{NN}\}$, of $\mathbf{U}$. In other words, $\text{vec}(\mathbf{U}) \sim KB_N(a, b, \Lambda_U)$. Several algorithms based on MCMC method, namely, M-H algorithm and Gibbs sampling, exist for generating samples from such multivariate PDF (see, e.g., [Spa03, Chapter 16]). However, application of such algorithms either requires a good proposal pdf (for M-H algorithm) or full conditional pdf of the components of $\text{vec}(\mathbf{U})$ (for

Gibbs sampling). Therefore, the present work recommends the use of slice sampling technique [Nea03] to sample $\text{vec}(\mathbf{U}) \sim KB_N(a, b, \Lambda_U)$. The slice sampling technique needs neither any proposal distribution nor the conditional distributions.

The key idea behind the slice sampling technique is to alternately sample from a vertical interval and a horizontal slice as sketched below:

1. Assume an initial guess, $\text{vec}(U^o)$, such that $U^o \in \mathcal{I}$.

2. Given $\text{vec}(U^k)$, obtain the $(k+1)$-th sample as follows:

   (a) Draw a scalar $Y \sim$ uniform on vertical interval $(0, p_{\mathbf{U}}(U^k))$. Define a horizontal "slice", $S = \{X \in \mathbb{R}^{N(N+1)/2} : Y < p_{\mathbf{U}}(\text{vec}^{-1}(X))\}$.

   (b) Draw the new sample, $\text{vec}(U^{k+1}) \sim$ uniform on S.

3. Increase $k \to (k+1)$ and repeat the above step until the desired number of samples of $\text{vec}(\mathbf{U}) \sim KB_N(a, b, \Lambda_U)$ is obtained. Map samples of $\text{vec}(\mathbf{U})$ to the samples of $\mathbf{U}$ and use the later samples to obtain the samples of $\mathbf{C} \sim GKB_N(a, b, \Lambda_C; C_u, C_l)$ by employing (5.17).

The step 2 is essentially the slice sampling technique and can be executed by using the softwares made available by Neal [Nea03] on-line[4]. Recently, MATLAB also introduced its function, `slicesample`, to implement this slice sampling algorithm.

5.3.4 A Note on Comparing Wishart Distribution and Standard Matrix Variate Kummer-Beta Distribution

The Wishart distribution or the matrix-variate gamma distribution, $G_N(\alpha, \Lambda_U)$, whose pdf is defined by (5.16), can also be shown to be an outcome of a MaxEnt optimization problem. In this MaxEnt formulation, the entropy would be defined by $(5.9)_{\mathbb{M}_N^+(\mathbb{R})}$ and the constraints by $(5.23)_{\mathbb{M}_N^+(\mathbb{R})}$, $(5.24)_{\mathbb{M}_N^+(\mathbb{R})}$ and $(5.26)_{\mathbb{M}_N^+(\mathbb{R})}$, in which the subscript, $\mathbb{M}_N^+(\mathbb{R})$, indicates that the support, $\mathcal{I}$, of $\mathbf{U}$ now needs to be replaced with $\mathbb{M}_N^+(\mathbb{R})$ keeping everything else the same.

[4] `http://www.cs.toronto.edu/~radford/fbm.software.html`

Consider the characteristic function of the Wishart distribution that can be immediately extracted from the existing literature of the RMT (see, e.g., [Mat97, p. 364]),

$$\phi_{\mathbf{U}}(\Theta) = \det(\mathbf{I} - \iota\Lambda_U^{-1}\Theta)^{-\alpha}, \quad \Theta \in \mathbb{M}_N^S(\mathbb{R}). \tag{5.41}$$

See, for alternate derivations, the books by Srivastava and Khatri [SK79, Section 3.3.6], Murihead [Mur82, Section 3.2.2], Gupta and Nagar [GN00, Theorem 3.3.7] and Anderson [And03, Section 7.3.1]. Based on this characteristic function and (5.33), the ensemble mean, $E[\mathbf{U}]$, can be obtained as,

$$E\{\mathbf{U}\} = \alpha\,\Lambda_U^{-1}. \tag{5.42}$$

In deriving this expression, the differentiation of determinant is required and it is readily available in the literature as shown below [Mat97, Theorem 1.3 and Lemma 1.3] in component form for any nonsingular matrix, $X = [x_{ij}]$,

$$\frac{\partial \det(X)}{\partial x_{ij}} = \det(X)[X^{-1}]_{ji}. \tag{5.43}$$

See the books by Murihead [Mur82, p. 90] and Gupta and Nagar [GN00, Theorem 3.3.15] for alternative derivations of (5.42).

Clearly, the role of the confluent hypergeometric function in the case of standard matrix-variate Kummer-Beta density in (5.20) is essentially being played by the determinant in the case of Wishart density in (5.16). In fact, it can "perhaps" be guessed simply by comparing the normalization constant, $K(a, b, \Lambda_{\underline{U}})$, of the standard matrix-variate Kummer-Beta density given by (5.22) and the normalization constant of the matrix-variate gamma density given by,

$$C(\alpha, \Lambda_U) = \left\{2^{\alpha N}\,\Gamma_N(\alpha)\det\left(\frac{1}{2}\Lambda_U^{-1}\right)^{\alpha}\right\}^{-1} = \frac{1}{\Gamma_N(\alpha)\det\left(\Lambda_U^{-1}\right)^{\alpha}}. \tag{5.44}$$

Since, the differentiation of determinant w.r.t. the element of its matrix argument can be readily obtained in closed form as shown in (5.43), the analytical derivation of the mean matrix of Wishart distribution as shown in (5.42) is straightforward. Such simple analytical result does not exist for the matrix-variate Kummer-Beta distribution until now.

A New Recommendation for Computing the Parameters of the Wishart Distribution

By the constraint, $(5.26)_{\mathbb{M}_N^+(\mathbb{R})}$, which implies $E\{\mathbf{U}\} = \underline{U}$, the use of (5.42) immediately results in the associated matrix-valued Lagrange multiplier, Λ_U,

$$\Lambda_U = \alpha\, \underline{U}^{-1}. \tag{5.45}$$

Here, the mean matrix, $\underline{U}$, is already known (either from a FE model or previous experience or from a set of samples of $\mathbf{U}$) but the other parameter (i.e., the modified Lagrange multiplier, α, associated with $(5.24)_{\mathbb{M}_N^+(\mathbb{R})}$) is still unknown and needs to be determined.

Determination of α typically requires a set of samples of $\mathbf{U}$ from which either α or some other scalar-valued parameter, that directly depends on α, is estimated. Particularly, a "dispersion parameter", that explicitly depends on α, is proposed by Soize [Soi00, Soi01a] as discussed in section 4.2 (see equation 4.3). The dispersion parameter, δ_U, is defined as $\delta_U = (E[\|\mathbf{U} - E[\mathbf{U}]\|_F^2]/\|E[\mathbf{U}]\|_F^2)^{1/2}$. The rhs of this expression can be estimated from the available set of samples resulting in $\widehat{\delta}_U$. On the other hand, knowing that $\mathbf{U} \sim G_N(\alpha, \Lambda_U)$, the dispersion parameter can also be explicitly expressed in terms of α and the known matrix, $\underline{U}$, (after using (5.45)) by using the already available analytical results on the fourth-order covariance tensor of $\mathbf{U}$ (see e.g. [Mur82, p. 90], [GN00, Theorem 3.3.15]) resulting in $\delta_U(\alpha, \underline{U})$. Then, solving $\delta_U(\alpha, \underline{U}) = \widehat{\delta}_U$ immediately yields α [Soi00, Soi01a, Soi06, Adh07]. The maximum likelihood approach or the minimum relative entropy approach, instead of using the dispersion parameter, is also applied to estimate α and Λ_U [Soi05b, ACB08].

While such schemes (based on covariance tensor or likelihood or relative entropy) are physically appealing from the end user's perspective, they also explicitly consider information that was not used in formulating the MaxEnt problem. If the dispersion parameter is known from previous experience or other reliable source, then the approach based on covariance tensor is perhaps more easy to apply. On the other hand, likelihood or relative entropy based approach *implicitly* assumes that the underlying 'true' PDF of $\mathbf{U}$ is exactly given by the family of PDF, $G_N(\alpha, \Lambda_U)$. The crucial premise behind using the MaxEnt principle is to estimate a pdf based on partial information so that the entropy (uncertainty) of the estimated pdf is maximized. Inverse techniques for estimation of α based on some other derived statistics are inconsistent with the original MaxEnt formulation and philosophy. We believe that it is more important to satisfy $(5.24)_{\mathbb{M}_N^+(\mathbb{R})}$ if a set of samples of $\mathbf{U}$, $\{U^{(1)}, \cdots, U^{(n)}\}$, can be used to estimate

$u_l = (1/n)\sum_{i=1}^{n} \log[\det(U^{(i)})]$ or a consistent, u_l, can be reliably obtained from other source. Then, α can be readily estimated by following the similar procedure as already described in section 5.3.1 and section 5.3.2. By (5.16) and (5.44), $(5.23)_{\mathbb{M}_N^+(\mathbb{R})}$ implies that,

$$\{C(\alpha, \Lambda_U)\}^{-1} = \int_{\mathbb{M}_N^+(\mathbb{R})} \det(U)^{\alpha - \frac{1}{2}(N+1)} \ \mathrm{etr}(-\Lambda_U U) \ dU.$$

Differentiating both sides w.r.t. α and substituting (5.16) in $(5.24)_{\mathbb{M}_N^+(\mathbb{R})}$, and subsequently using all the resulting expressions along with (5.45), it can be shown that,

$$\frac{\partial \log[\Gamma_N(\alpha)]}{\partial \alpha} - N[1 + \log(\alpha)] + \log[\det(\underline{U})] = u_l.$$

Here, $\underline{U}$ and u_l are known or given, and the logarithmic derivative of $\Gamma_N(\alpha)$ can be conveniently expressed in terms of psi or digamma function. A nonlinear least squares technique can be readily employed to solve the above equation for α with lower bound of $(1/2)(N-1) + \epsilon$, in which ϵ is a very small number, say, 1×10^{-7}.

5.4 Numerical Illustration

A two phase material with a dominant matrix phase and a secondary phase (inclusions) is considered here. The application of the proposed probability model and the related numerical strategies are illustrated in a step-by-step sequential fashion.

5.4.1 Computational Experiment

It is explained here, as indicated earlier in section 5.2.1, how the present work can be adapted to the non-parametric homogenization at a very small length scale. In an experimental set-up, the volume fractions of the different phases of a heterogeneous material can be identified from the micrograph obtained by scanning a heterogeneous test specimen. If a reasonably large V is chosen, then the volume fractions are not expected to vary across different samples of test specimens. On the other hand, the volume fractions would vary considerably if the size of V is very small. It is assumed here that the volume fraction, v_{i}, based on a fragment of material volume of size, V, varies across the heterogeneous test specimens. The volume fraction, v_{m}, of the corresponding matrix phase follows from $v_{\mathrm{i}} + v_{\mathrm{m}} = 1$.

It is assumed here that the observed minimum (over test specimens) value of $v_{\rm i}$ is 0.01 and the maximum value is 0.05. Of course, the actual values would depend on the selected physical size of V, and a large enough V would ensure that the minimum and maximum value would be approximately equal. Since V could be small, this difference is allowed. In the present work, C_l is determined based on the minimum value, 0.01, of the volume fraction of the inclusion (and, therefore, $v_{\rm m} = 0.99$), and C_u is determined based on the maximum value, 0.05, of $v_{\rm i}$ (with $v_{\rm m} = 0.95$). It is also assumed here that the matrix phase and the inclusion are individually homogeneous, linearly elastic and isotropic. This implies that V is linearly elastic but it is still, in general, heterogeneous and anisotropic. The lower and upper bounds of $\mathbf{C}^{\rm eff}$ are computed within a parametric set-up by further assuming that the plane stress linear elasticity theory is valid for both the matrix phase and the inclusion, and the Young's modulus and Poisson's ratio of the matrix phase are assumed to be, respectively, $E_{\rm m} = 73$ GPa and $\nu_{\rm m} = 0.33$, and that of the inclusion, respectively, $E_{\rm i} = 730$ GPa and $\nu_{\rm i} = 0.15$. It should be noted that this parametric approach is used only to obtain the two bounds of $\mathbf{C}^{\rm eff}$. The effective elasticity matrix, $\mathbf{C}^{\rm eff}$, would still be modeled by using the nonparametric approach. Based on the isotropic and homogeneous material properties of the constituent phases and the plane stress condition, the elasticity matrices of the constituent phases can be computed. Subsequently, using these elasticity matrices of the constituents phases, the bounds, C_l and C_u, are determined, respectively, by using $(5.8)_1$ (with $v_{\rm i} = 0.01$ and $v_{\rm m} = 0.99$) and $(5.8)_2$ (with $v_{\rm i} = 0.05$ and $v_{\rm m} = 0.95$). The matrices, C_l and C_u, are reported below.

$$C_l = 1.0 \times 10^{10} \begin{bmatrix} 8.27 & 2.73 & 0 \\ 2.73 & 8.27 & 0 \\ 0 & 0 & 2.77 \end{bmatrix}, \quad C_u = 1.0 \times 10^{10} \begin{bmatrix} 11.52 & 3.13 & 0 \\ 3.13 & 11.52 & 0 \\ 0 & 0 & 4.19 \end{bmatrix}. \tag{5.46}$$

Now that the bounds are available, either the knowledge about c_l in (5.11), c_u in (5.12) and the mean matrix, $\underline{C}$, in (5.18) are required or a set of samples are needed to estimate these statistics in order to characterize $\mathbf{C}^{\rm eff}$ by employing the nonparametric probability model as proposed in section 5.3.2. In the absence of a suitable experimental database, a set of $n = 100$ specimens of the heterogeneous material is digitally generated. Consider a laboratory test set-up where a test specimen is typically subjected to a specified tensile or compressive loading and the digital image processing technique is used to identify the associated strain field over the entire domain of the test specimen. The set of n digitally generated test specimens are tested through a computational experiment virtually simulating this laboratory test set-up.

Since such laboratory tests use specimens that are larger than the typical sizes of micrograph scans, the volume fractions of the different phases of the material based on such relatively larger sized test specimens are likely to show very small fluctuation. The volume fraction of the inclusion of such test specimen can be expected to be around the middle of the range, $(0.01,\ 0.05)$, of $v_{\rm i}$ as used earlier in determining C_l and C_u based on a relatively small sized V. It is, therefore, presumed here that $v_{\rm i}$ varies between 0.028 and 0.032 across the n test specimens. The set of n values of $v_{\rm i}$ for n test specimens is simply obtained from $U(0.028,\ 0.032)$ *given* an initial seed (resulting a deterministic set of n values of $v_{\rm i}$), in which $U(x, y)$ is the PDF of uniform random variable supported on $(x,\ y)$. Since the volume averaged stress and strain are required in determining the samples of $\mathbf{C}^{\rm eff}$ (see (5.7)), the size of such *computational* test specimens can be conveniently selected as long as they are consistent with the volume fractions of the different phases of the *laboratory* test specimens. Therefore, an unit area, that manifests the volume fractions in a statistically uniform manner, is selected as the size for all the n computational test specimens. Based on the set of n values of $v_{\rm i}$ as determined earlier, a set of n computational test specimens are digitally generated. Two such typical samples showing the two different phases are shown in Figure 5.2. Of course, these samples

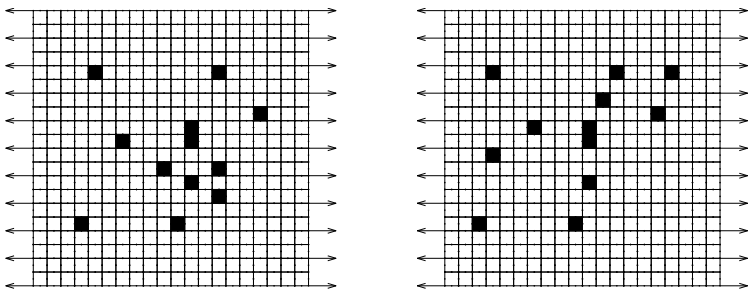


Figure 5.2: Typical test samples of unit area; the black phase represents inclusion and the spatial regions of the inclusions are randomly selected; FE analysis done with 9-node quadrilateral plane stress elements.

are grossly simplified versions of the heterogeneous material. All the simplifications as mentioned here nevertheless are general enough within the context of the present work, and are employed only to focus more closely on the primary contributions of the proposed method and also because of non-availability of experimental data[5].

[5]An application of the proposed method on aluminium specimen containing inclusions of different materials has recently been completed and presented in the recent SPIE conference 6926 on Modeling, Signal Processing, and Control for Smart Structures 2008.

The set of n strain fields, from which the volume averaged strain for n computational test specimens can be obtained, is determined by performing FE analysis on each test specimen subjected to an applied traction. In the FE models, the material properties of the matrix phase and inclusion as mentioned earlier are assigned appropriately to the corresponding phases. The 9-node quadrilateral plane stress elements are used. The FE mesh and the applied boundary traction, both of which are same for all the n computational test specimens, are also shown in Figure 5.2. The applied boundary traction is a particular SUBC (see (5.2)) and the vector-valued representation of the associated $\boldsymbol{\sigma}_o$ is given by $[\sigma_{11}^{(o)}\ \sigma_{22}^{(o)}\ \sigma_{12}^{(o)}]^T$ with $\sigma_{11}^{(o)} = 60$ KPa and $\sigma_{22}^{(o)} = \sigma_{12}^{(o)} = 0$. It must be realized that the set of n strain fields are generated within a parametric framework simply because of the lack of experimental strain fields. The computationally generated database of these consistent strain fields can, therefore, be treated as a proxy for the set of experimental strain fields typically needed to be identified by employing the digital image processing technique. Since v_{i} varies between 0.028 and 0.032 across the test specimens and the spatial regions of the inclusions are randomly selected (see Figure 5.2), the determined n strain fields would vary across the test specimens even though the applied boundary traction remains same for all the samples.

While the n strain fields as determined above yield the set, $\{\langle\boldsymbol{\varepsilon}^{(i)}\rangle_V\}_{i=1}^n$, of volume averaged strain (with V being unity representing the area of each computational test specimen), the applied boundary traction immediately results in the volume averaged stress since perfect interfaces between the inclusions and matrix phase are assumed to be valid here [NNH99, Section 2.1]. The vector-valued representation of the volume averaged stress for all the specimens, thus, is immediately given by [Hue90, NNH99, Section 2.3.1],

$$\langle\boldsymbol{\sigma}\rangle_V = [60 \quad 0 \quad 0]^T \,\mathrm{KPa}. \tag{5.47}$$

In a laboratory test set-up, the applied tensile loading can be changed to control the value of $\sigma_{11}^{(o)} = 60$ KPa to attain a desired order of stress and strain.

Based on the set, $\{\langle\boldsymbol{\varepsilon}^{(i)}\rangle_V\}_{i=1}^n$, and $\langle\boldsymbol{\sigma}\rangle_V$ in (5.47), the next section describes a computational scheme (within a nonparametric formalism) to determine a set, $\{C^{(1)}, \cdots, C^{(n)}\}$, of samples of $\mathbf{C}^{\mathrm{eff}}$. From this set, the sample estimates of c_l, c_u and $\underline{C}$ can be obtained as $c_l^{(\mathrm{samp})} = (1/n)\sum_{i=1}^n \ln[\det(C^{(i)} - C_l)]$, $c_u^{(\mathrm{samp})} = (1/n)\sum_{i=1}^n \ln[\det(C_u - C^{(i)})]$ and $\underline{C}^{(\mathrm{samp})} = (1/n)\sum_{i=1}^n C^{(i)}$, respectively.

5.4.2 Nonparametric Homogenization: Determination of Experimental Samples of $\mathbf{C}^{\text{eff}}$

Given the set $\{\langle\boldsymbol{\varepsilon}^{(i)}\rangle_V\}_{i=1}^n$ and $\langle\boldsymbol{\sigma}\rangle_V$ in (5.47), the i-th sample, $C^{(i)}$, of $\mathbf{C}^{\text{eff}}$ is given by (5.7), $\langle\boldsymbol{\sigma}\rangle_V = C^{(i)}\langle\boldsymbol{\varepsilon}^{(i)}\rangle_V$. The i-th sample, $C^{(i)}$, is obtained here by solving the following optimization problem,

$$\begin{aligned} &\text{minimize} \qquad 100\,\frac{\mid \langle\boldsymbol{\sigma}\rangle_V - C^{(i)}\langle\boldsymbol{\varepsilon}^{(i)}\rangle_V \mid_1}{\mid \langle\boldsymbol{\sigma}\rangle_V \mid_1} \\ &\text{subject to} \qquad C_l < C^{(i)} < C_u. \end{aligned}$$

Here, $|\cdot|_1$ is the l_1-norm defined by $\mid \mathbf{x} \mid_1 = \sum_{i=1}^d \mid x_i \mid$, $\mathbf{x} = (x_1, \cdots, x_d) \in \mathbb{R}^d$. Since the values of the components of $C^{(i)}$ could be of high order (say, 1×10^{10}), it might be useful to solve the following equivalent optimization problem instead to obtain the samples of $\mathbf{C}^{\text{eff}}$ in order to avoid machine overflow problem,

$$\text{minimize} \qquad 100\,\frac{\mid S^{(i)}\langle\boldsymbol{\sigma}\rangle_V - \langle\boldsymbol{\varepsilon}^{(i)}\rangle_V \mid_1}{\mid \langle\boldsymbol{\varepsilon}^{(i)}\rangle_V \mid_1} \tag{5.48}$$

$$\text{subject to} \qquad C_u^{-1} < S^{(i)} < C_l^{-1}, \tag{5.49}$$

and then computing $C^{(i)} = (S^{(i)})^{-1}$. This optimization problem can be conveniently solved by using the semi-definite programming (SDP) [VB96], [Dat05, Chapter 4]. A very efficient public domain MATLAB toolbox, YALMIP, developed by Löfberg [Lof04], is used in the present work to solve a set of n semi-definite optimization problems to obtain the samples, $\{C^{(1)}, \cdots, C^{(n)}\}$, of $\mathbf{C}^{\text{eff}}$. Based on the samples thus determined, $c_l^{(\text{samp})}$, $c_u^{(\text{samp})}$ and $\underline{C}^{(\text{samp})}$ are estimated as,

$$\begin{aligned} c_l^{(\text{samp})} &= 66.3893, \\ c_u^{(\text{samp})} &= 71.1065, \end{aligned} \quad \text{and } \underline{C}^{(\text{samp})} = 1.0 \times 10^{10} \begin{bmatrix} 8.5833 & 2.9949 & 0.0033 \\ 2.9949 & 9.1587 & 0.0007 \\ 0.0033 & 0.0007 & 3.0915 \end{bmatrix}. \tag{5.50}$$

It must be realized that the scheme presented in this section is strikingly different than the currently existing methods available in the multiscale literature for determining the samples of $\mathbf{C}^{\text{eff}}$. It does not require a sequence of traction or displacement boundary conditions in order to determine the individual components of $C^{(i)}$'s as often done in a parametric set-up. Each sample, $C^{(i)}$, can be obtained based on

only one operational or representative boundary condition — traction or displacement or combination of both. The matrix samples, $\{C^{(1)}, \cdots, C^{(n)}\}$, thus obtained by employing the scheme proposed in this section, are characterized by the nonparametric models even if the bounds, C_l and C_u, are obtained by using the parametric approach.

5.4.3 Matrix Variate Kummer-Beta Probability Model for $\mathbf{C}^{\text{eff}}$

Having obtained the bounds of $\mathbf{C}^{\text{eff}}$ given by (5.46) and the sample statistics given by (5.50), the parameters of the matrix-variate Kummer-Beta pdf is determined by following the scheme as described in section 5.3.2 by setting $c_l = c_l^{(\text{samp})}$, $c_u = c_u^{(\text{samp})}$ and $\underline{C} = \underline{C}^{(\text{samp})}$. In solving the constrained minimization problem defined by (5.35), (5.37), (5.38) and (5.40), a hybrid global and local optimization technique is employed to determine the triplet of parameters, (a, b, Λ_U). A set of several random points (4000 points here) are first generated in the joint domain of the parameters, (a, b, Λ_U), as a set of possible initial points, and then a subset of the best initial points (200 points) from these randomly generated 4000 points are carefully chosen. Subsequently, a local optimization algorithm is invoked at each of these best initial points. The local optimization algorithm successfully converges only for 62 initial points out of these chosen 200 initial points resulting in a set of 62 triplets of the optimized parameters. The minimum (over the set of 62 optimized triplets of parameters) value of the objective function defined by (5.35) is 0.0071. The associated relative mean-squared error (relMSE) of the analytical ensemble mean, $E[\mathbf{U}]$, relative to $\underline{U}^{(\text{samp})}$ defined by,

$$\text{relMSE}(E[\mathbf{U}], \underline{U}^{(\text{samp})}) = 100 \times \frac{\|\underline{U}^{(\text{samp})} - E[\mathbf{U}]\|_F^2}{\|\underline{U}^{(\text{samp})}\|_F^2}, \tag{5.51}$$

turns out to be 4.7492%, in which $E[\mathbf{U}]$ is calculated by (5.34) and the sample estimate, $\underline{U}^{(\text{samp})}$, is computed by the rhs of (5.29) with $\underline{C}^{(\text{samp})}$ being given by $(5.50)_3$ and C_l and C_u by (5.46). However, $\text{relMSE}(E[\mathbf{U}], \underline{U}^{(\text{samp})})$ attains a minimum value of 4.6993% for a different optimized triplet for which the objective function assumes a value of 0.0074. It clearly indicates the existence of multiple local solutions. The optimized triplet, (a, b, Λ_U), that yields the minimum $\text{relMSE}(E[\mathbf{U}], \underline{U}^{(\text{samp})})$ is, however, chosen as the parameters of the associated standard matrix-variate Kummer-Beta distribution. This optimized triplet of parameters is more meaningful in this example since capturing the mean of $\mathbf{C}^{\text{eff}}$

as accurately as possible is one of the primary criteria for homogenization at a small length scale. The optimized parameters thus selected are reported below,

$$a = 7.5926, \quad b = 36.3529, \quad \text{and} \quad \Lambda_U = \begin{bmatrix} 14.2379 & -17.5549 & 1.2482 \\ -17.5549 & -13.3125 & -3.8675 \\ 1.2482 & -3.8675 & -6.9579 \end{bmatrix}. \tag{5.52}$$

While the analytical value of $B(u_l, u_u) = [u_l,\ u_u]$ given by the lhs of (5.30) and (5.31), based on the parameters in (5.52), turns out to be $B^{\text{anal}}(u_l, u_u) = [-5.4119,\ -0.6782]$, the sample estimate of $B(u_l, u_u)$ based on the rhs of (5.27) and (5.28) is $B^{\text{samp}}(u_l, u_u) = [-5.3846,\ -0.6674]$ with $c_l^{(\text{samp})}$ and $c_u^{(\text{samp})}$ being given by $(5.50)_{1,2}$.

Based on (5.52), the $\text{relMSE}(E[\mathbf{C}], \underline{C}^{(\text{samp})})$, $\mathbf{C} \equiv \mathbf{C}^{\text{eff}}$, turns out to be 0.0286%, in which the sample estimate, $\underline{C}^{(\text{samp})}$, is given by $(5.50)_3$ and the analytical ensemble mean, $E[\mathbf{C}]$, is determined by using (5.17) as $E[\mathbf{C}] = (C_u - C_l)^{\frac{1}{2}}\ E[\mathbf{U}]\ (C_u - C_l)^{\frac{1}{2}} + C_l$. The analytical value of $B(c_l, c_u)$ is found to be $B^{\text{anal}}(c_l, c_u) = [66.3620,\ 71.0957]$ which is determined from (5.27) and (5.28) by using the already determined $B^{\text{anal}}(u_l, u_u)$. The value of $B^{\text{anal}}(c_l, c_u)$ thus determined should be compared with the corresponding sample estimate, $B^{\text{samp}}(c_l, c_u) = [66.3893,\ 71.1065]$, as given in $(5.50)_{1,2}$.

5.4.4 Sampling of $\mathbf{C}^{\text{eff}}$ Using the Slice Sampling Technique

Using the probability model as determined above, a set of 0.1 million samples of $\mathbf{U}$ is digitally generated by employing the slice sampling technique as indicated in section 5.3.3. A burn-in period of 500 is used here. The 501-st sample resulting from one run of the slice sampling algorithm yields one sample of $\mathbf{U}$. Each run of the slice sampling algorithms is initiated with a sample of the standard matrix-variate beta type I distribution generated on the fly by using Algorithm 5.3.5. A total of 0.1 million independent such runs is carried out thus generating $n_s = 0.1 \times 10^6$ statistically independent samples of $\mathbf{U}$. The samples of $\mathbf{C}^{\text{eff}}$ follow from the samples of $\mathbf{U}$ through the use of (5.17).

Let us denote the mean matrix based on n_s digital samples of $\mathbf{U}$ obtained via the slice sampling technique by $\underline{U}^{(n_s)}$, and similarly the mean matrix for $\mathbf{C} \equiv \mathbf{C}^{\text{eff}}$ by $\underline{C}^{(n_s)}$. Then, we have $\text{relMSE}(\underline{U}^{(n_s)}, \underline{U}^{(\text{samp})}) = 3.8824\%$, $\text{relMSE}(\underline{U}^{(n_s)}, E[\mathbf{U}]) = 3.4869\%$ for $\mathbf{U}$ and $\text{relMSE}(\underline{C}^{(n_s)}, \underline{C}^{(\text{samp})}) = 0.0249\%$, $\text{relMSE}(\underline{C}^{(n_s)}, E[\mathbf{C}]) = 0.0248\%$ for $\mathbf{C}$. The values of

$B(u_l, u_u)$ and $B(c_l, c_u)$ based on the respective n_s slice samples are estimated as $B^{(n_s)}(u_l, u_u) = [-5.4102,\ -0.6787]$ and $B^{(n_s)}(c_l, c_u) = [66.3637,\ 71.0952]$.

5.4.5 Analyzing a Cantilever Beam by Using Nonparametric $\mathbf{C}^{\text{eff}}$

Consider a 2D cantilever beam subjected to a downward load, $P = 1$ N, at $x = L = 16.0$ m and fixed at $x = 0$ m as shown in Figure 5.3. The total load, P, is distributed along y as $f(y) = -P\{(h/2)^2 - y^2\}/(2I)$ as shown at $x = L$ with a dashed line, where $h = 4$ m is the height of the cantilever beam and $I = (1/12)h^3$ is the second polar moment of inertia of the beam cross-section with unit width. The material property is characterized by the random effective elasticity matrix, $\mathbf{C}^{\text{eff}}$, the samples of which are obtained by the nonparametric homogenization scheme as proposed in section 5.4.2. The FE analysis and MC simulation are employed to characterize the associated random response by considering a set of $N_s = 45000$ samples of $\mathbf{C}^{\text{eff}}$. These N_s samples is simply selected from the previously generated n_s samples of $\mathbf{C}^{\text{eff}}$. Each sample of $\mathbf{C}^{\text{eff}}$ characterizes the material property over the entire spatial domain of the corresponding sample of the beam. The FE mesh as shown in Figure 5.3 consists of 9-node quadrilateral plane elements and remains same for all the beam samples. The material property of each beam sample is characterized by the corresponding sample of $\mathbf{C}^{\text{eff}}$.

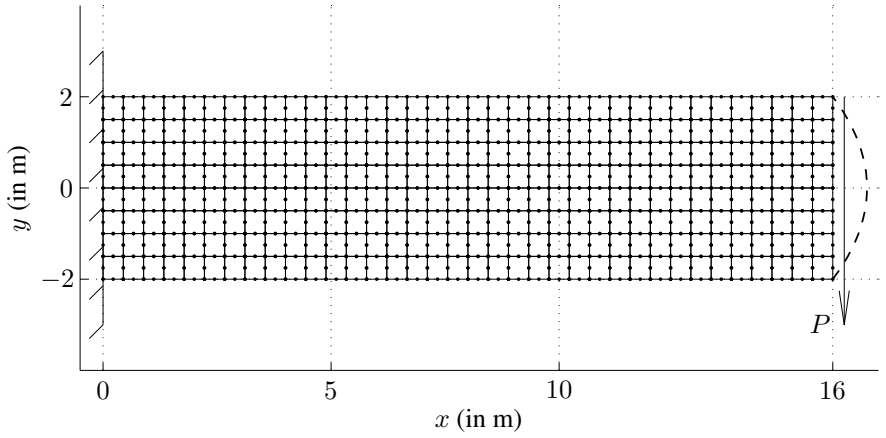


Figure 5.3: A 2D homogenized cantilever beam modeled with 9-node quadrilateral nonparametric plane elements; the total load P is distributed parabolically as shown with a dashed line at $x = L$.

Based on N_s samples of $\mathbf{C}^{\text{eff}}$, estimate of the pdf of $\text{tr}(\mathbf{C}^{\text{eff}})$ and of the volume averaged strain energy, $\varphi = (1/2)\langle \boldsymbol{\varepsilon}^T(\mathbf{x})\mathbf{C}^{\text{eff}}\boldsymbol{\varepsilon}(\mathbf{x})\rangle_V$, $V = L \times h \times 1\ \text{m}^3$, resulting from the FE analysis are shown in Figure 5.4. A few representative statistics of the random response of the beam are plotted in Figure 5.5. Profile of

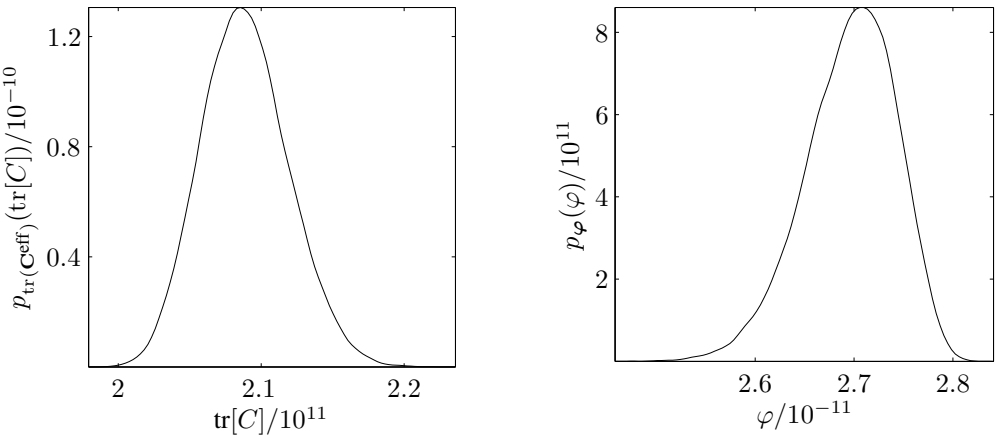


Figure 5.4: Estimate of the pdf of $\mathrm{tr}(\mathbf{C}^{\mathrm{eff}})$ and of the volume averaged strain energy, $\boldsymbol{\varphi}$, based on N_s samples.

the sample mean of y-displacement (based on N_s samples) of the beam along its center line, $y = 0$, is shown in the top half of the figure. In the bottom half, estimates of three typical pdf of the random y-displacement along $y = 0$ at three different locations are depicted based on N_s samples.

Remark 5.4.1 *The computational burden in modeling a system by using the matrix-variate beta type I distribution as proposed in this work and the Wishart or matrix-variate gamma distribution as proposed by Soize [Soi00, Soi01a] are comparable. Both the models are easy to implement in characterizing a suitable stochastic system. The associated computational overhead are substantially less relative to that involved in characterizing a stochastic system by employing the matrix-variate Kummer-Beta distribution as proposed as another nonparametric model in the present work. The numerical example, therefore, is purposefully selected to illustrate the step-by-step application procedure of the matrix-variate Kummer-Beta model and the required computational workloads are executed solely in a single processor machine to highlight the affordable computational expenses. Nevertheless, the several modules of the numerical tasks required in using the nonparametric Kummer-Beta model can be readily parallelized reducing the computational cost to a great extent. Specific instances of such parallelizable tasks are the FD approximation of the gradient of ${}_1F_1$ (section 5.3.2), generation of samples of $\mathbf{C}^{\mathit{eff}}$ by using the SDP (section 5.4.2), the hybrid global-local optimization scheme (section 5.4.3) and the use of slice sampling technique (section 5.3.3 or section 5.4.4) and of MC simulation (section 5.4.5).*

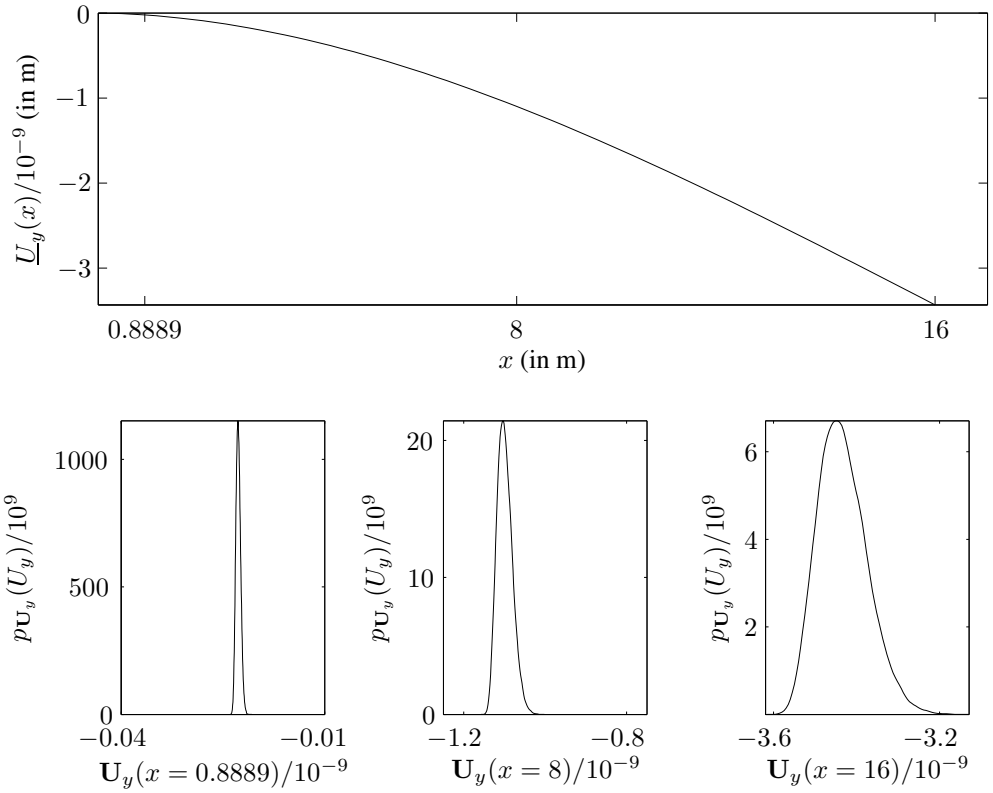


Figure 5.5: Statistics of the random response of the cantilever beam; $\mathbf{U}_y(x)$ represents the random displacement of the beam in y-direction along the center line of the beam, $y = 0$; $E[\mathbf{U}_y(x)]$ is the sample mean of $\mathbf{U}_y(x)$ estimated based on N_s samples.

5.5 Conclusions

A MaxEnt based random matrix formalism is proposed in the present work to construct the probability model of a stochastic mechanical structure whose system matrix, $\mathbf{C}$, is symmetric positive definite and *strictly* bounded from below and above in the positive definite sense by two deterministic symmetric positive definite matrices, C_l and C_u, as shown in (5.1). Two distinct distributions result from the MaxEnt formulation depending on the information made use of in constructing the probability model of $\mathbf{C}$. The first probability model is the matrix-variate beta type I distribution that results from imposing two particular constraints in the MaxEnt framework guaranteeing negligible probability mass near the two bounds thus taking care of the strict-bound conditions. The second probability model is the outcome of enforcing the additional constraint of the mean matrix of $\mathbf{C}$ yielding the matrix-variate Kummer-Beta distribution. The former probability model would be more useful if the "gap" between the two bounds, C_l and C_u,

is narrow so that imposing the mean matrix constraint is not essential. The later probability model, on the other hand, would be particularly suitable if the "gap" between the two bounds are relatively wide yielding higher level of scatter so that enforcing the mean matrix would be beneficial in capturing the known first-order statistic of $\mathbf{C}$.

While the proposed approach can be applied to characterize a broader spectrum of stochastic systems, the motivation behind choosing the particular numerical example on homogenization of a heterogeneous material is close in spirit with two recent works [SZ06, Soi08]. But it is also distinct in key respects as explained next. One of these works [SZ06] relies on the higher-order statistics of the morphological features of the experimental micrographs to characterize the random effective material property (e.g., Young's modulus etc.) within the parametric framework. The corresponding bounds of the effective material property are not used in constructing the associated probability model in that work but a *finite* set of simulated samples are used to cross-check if the samples lie within the bounds. The present work, besides being developed within a nonparametric framework, explicitly enforces the bounds in constructing the pdf of $\mathbf{C}^{\text{eff}}$. It also requires minimal information about the morphological features, or the experimentally identified complete morphological features are precisely and implicitly incorporated into the bounds, C_l and C_u. The information about the morphological features are also embedded into the statistics such as c_l, c_u and $\underline{C}$ while estimating them from the experimentally identified samples of $\mathbf{C}^{\text{eff}}$. Minimal information (i.e., volume fractions of the different phases of the heterogeneous material) is used if (5.8) is employed. The experimentally identified complete morphological features, on the other hand, are inherently taken into account while computing the bounds as well as c_l, c_u and $\underline{C}$ from the samples of $\mathbf{C}^{\text{eff}}$ by employing numerical analysis (e.g., FE technique) of the micrograph specimens. In such cases, the individual material properties are appropriately assigned to the different phases and the resulting specimens are subsequently subjected to SUBC, KUBC and computational tension test (or operational/representative boundary conditions including a combination of traction or displacement boundary conditions depending on the test-setup) [Hue90, HH94]. Therefore, no attempt is made in the present work to explicitly incorporate the higher-order statistics of the morphological features of the micrographs in characterizing the random effective material property. The other work [Soi08], published after the present work was completed, is carried out within the nonparametric formulation but proposes to use the Wishart or matrix-variate gamma distribution supported over $\mathbb{M}_N^+(\mathbb{R})$ thus violating the significant bound constraints as indicated by (5.1).

The nonparametric homogenization scheme as proposed in section 5.4.2 is a much less time-consuming technique both experimentally and computationally since it requires application of only one boundary condition to extract the full matrix (samples of $\mathbf{C}^{\text{eff}}$) simultaneously. The similar scheme, in fact, can also be efficiently used in determining the lower and upper bounds, $C_l, C_u \in \mathbb{M}_N^+(\mathbb{R})$, of $\mathbf{C}^{\text{eff}}$ when C_l and C_u are obtained by applying the SUBC and KUBC, instead of using the Reuss and Voigt type bounds as used in the present work, in order to achieve slightly tighter bounds.

The present work also critically investigates the existing schemes for estimating the parameter, α, of $G_N(\alpha, \Lambda_U)$. A new scheme is proposed to estimate α if a set of samples of $\mathbf{U} \sim G_N(\alpha, \Lambda_U)$ are available or a consistent value of u_l in $(5.24)_{\mathbb{M}_N^+(\mathbb{R})}$ can be specified, which would be fundamentally more congruous with the very basic motive of seeking a MaxEnt based pdf estimate.

Chapter 6

Current and Future Research Tasks

The greater danger for most of us is not that our aim is too high and we miss it,

but that it is too low and we reach it.

$\sim$ Michelangelo di Lodovico Buonarroti Simoni (March 6, 1475 – February 18, 1564)

Based on the study carried out in this dissertation, the followings are a few ongoing works and suggestions for future research plan:

1. (**Chapter 2**) While the asymptotic probability density function (apdf) of the estimators of the polynomial chaos (PC) coefficients is identified as multivariate normal probability density function (pdf), a careful reflection from consideration of the moment-based constraints would reveal that the support of the estimators of the PC coefficients should be a proper subset of $\mathbb{R}^{(P+1)}$. Thus, there is a further need to investigate this issue along with the effect of truncation of the infinite-series based PC expansion at a finite $P < \infty$.

2. (**Chapter 3**) Identification of the appropriate apdf of the estimators of the PC coefficients based on histogram could be another research topic since it would also be useful in determining the confidence interval as discussed in the context of the work presented in chapter 2.

3. (**Chapter 2–3**) The constraint, $n_d = M$, is required to be relaxed to identify an optimal functional (PC) dimension of the random variate under investigation. A maximum-entropy (MaxEnt) based procedure, with a goal to minimize the mean-square error (MSE) of the resulting PC expansion, has recently been formulated but the associated algorithm still requires to be validated on a benchmark problem.

4. (**Chapter 4**) A time-domain coupling technique for coupling the parametric and nonparametric subsystems has also recently been formulated. However, its potential applicability and computational efficiency are yet to be tested.

5. (**Chapter 5**) A more efficient and robust algorithms, or reformulation of the optimization problem, for computing the parameters of the matrix-variate Kummer-Beta distribution could be another interesting topic. Special attentions are also due for more user friendly sampling schemes.
   As a part of this task, one mathematical research topic could be to search for a closed-form expression for the differentiation of the confluent hypergeometric function of matrix argument with respect to the elements of the associated matrix.

6. (**Chapter 5**) The proposed approach needs to be extended or/and applied to a several other significant practical problems including, but not limited to, microstructures with cracks (a simple semi-definite programming (SDP) based algorithm has recently been formulated but yet to be tested on practical data), marine structures subjected to underwater detonation (this would probably also require coupling approach similar to the one presented in chapter 4), living cells, network systems (e.g., transportation network system) etc.

7. (**Chapter 4–Chapter 5**) Incorporation of additional higher order statistics (for example, covariance tensors) of the random matrix or/and mechanics-based constraints into the MaxEnt formulation would also be a practically appealing research topic.

8. (**Chapter 5**) Extension of the (discretized) random matrix based approach to the case with continuous stochastic operators would be another potential research task.

9. (**Chapter 4-Chapter 5**) Another interesting research topic would be to characterize a random variate by using orthogonal polynomials of matrix argument.

References

[ABC+97] J. R. Apel, B. Badiey, C. S. Chiu, S. Finette, J. Headrick, J. N. Kemp, J. F. Lynch, A. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut, K. von der Heydt, and S. Wolf. An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment. *IEEE Journal of Oceanic Engineering*, 22(3):465–500, 1997.

[ACB08] M. Arnst, D. Clouteau, and M. Bonnet. Inversion of probabilistic structural models using measured transfer functions. *Computer Methods in Applied Mechanics and Engineering*, 197(6-8):589–608, 2008.

[Adh07] Sondipon Adhikari. Matrix variate distributions for probabilistic structural dynamics. *AIAA Journal*, 45(7):1748–1762, July 2007.

[And03] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, New Jersy, 2003.

[AS70] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., New York, 1970.

[AS07] Y. Aoyagi and K. Shizawa. Multiscale crystal plasticity modeling based on geometrically necessary crystal defects and simulation on fine-graining for polycrystal. *International Journal of Plasticity*, 23(6):1022–1040, 2007.

[Bar33] M. S. Barlett. On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, 53:260–283, 1933.

[Bar98] R.R. Barton. Simulation metamodels. In D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan, editors, *Proceedings of the 1998 Winter Simulation Conference*, 1998.

[BAS+06] A. Brahme, M. H. Alvi, D. Saylor, J. Fridy, and A. D. Rollett. 3D reconstruction of microstructure in a commercial purity aluminum. *Scripta Materialia*, 55(1):75–80, July 2006.

[Baz00] Zdeněk P. Bažant. Size effect. *International Journal of Solids and Structures*, 37(1-2):69–80, 2000.

[Ber99] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts, USA, 1999.

[BLT03] I. Babuška, Kang-M. Liu, and R. Tempone. Solving stochastic partial differential equations based on the experimental data. *Mathematical Models and Methods in Applied Sciences*, 13(3):415–444, 2003.

[BP04] Alain Bourgeat and Andrey Piatnitski. Approximations of effective coefficients in stochastic homogenization. *Annales de l'Institut Henri Poincaré – Probabilités et Statisques*, 40(2):153–165, 2004.

[BPP96] A.L. Berger, S.A.D. Pietra, and V.J.D. Pietra. A maximum entropy approach to natural language processing. *Comput. Linguist.*, 22(1):39–71, 1996.

[BTC79] A. Ben-Tal and A. Charnes. A dual optimization framework for some problems of information theory and statistics. *Problems of Control and Information Theory*, 8:387–401, 1979.

[BTTC88] A. Ben-Tal, M. Teboulle, and A. Charnes. The role of duality in optimization problems involving entropy functionals with applications to information theory. *J. Optim. Theory. Appl.*, 58(2):209–223, August 1988.

[BTZ05] I. Babuška, R. Tempone, and G.E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1251–1294, 2005.

[BW02] R.W. Butler and A.T.A Wood. Laplace approximation for hypergeometric functions with matrix argument. *The Annals of Statistics*, 30(4):1155–1177, 2002.

[CB02] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury, USA, 2002.

[CBS00] G. Christakos, P. Bogaert, and M. Serre. *Temporal GIS: Advanced Functions for Field-Based Applications*. Springer, 2000.

[CF06] Wen Chen and Jacob Fish. A mathematical homogenization perspective of virial stress. *International Journal for Numerical Methods in Engineering*, 67:189–207, 2006.

[CLPP$^+$07] E. Capiez-Lernouta, M. Pellissetti, H. Pradlwarter, G. I. Schueller, and C. Soize. Data and model uncertainties in complex aerospace engineering systems. *Journal of Sound and Vibration*, 295(3-5):923–938, 2007.

[CLR96] G. Chryssolouris, M. Lee, and A. Ramsey. Confidence interval prediction for neural networks models. *IEEE Transactions on Neural Networks*, 7(1):229–232, January 1996.

[CM47] R.H. Cameron and W.T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. Math. (2)*, 48(2):385–392, April 1947.

[CN97] M.C. Cario and B.L. Nelson. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of Industrial Engineering and 35 Management Sciences, Northwestern University, Evanston, Illinois, 1997. `http://citeseer.ist.psu.edu/cario97modeling.html`.

[CR99] R.T. Clemen and T. Reilly. Correlations and copulas for decision and risk analysis. *Management Science*, 45(2):208–224, February 1999.

[Das07] Sonjoy Das. Efficient calculation of Fisher information matrix: Monte Carlo approach using prior information. Master's thesis, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland, USA, May 2007. `http://dspace.library.jhu.edu/handle/1774.2/32459`.

[Dat05] Jon Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing USA, 2005. Corrected version in 2007.

[DG04] Sonjoy Das and Roger Ghanem. Uncertainty analysis for surface ship subjected to underwater detonation. In *Proceedings of the IMAC-XXII Conference & Exposition on Structural Dynamics*, 26-29 January 2004.

[DGS06] C. Desceliers, R. Ghanem, and C. Soize. Maximum likelihood estimation of stochastic chaos representation from experimental data. *International Journal for Numerical Methods in Engineering*, 66(6):978–1001, 2006.

[DGS08] Sonjoy Das, Roger Ghanem, and James C. Spall. Asymptotic sampling distribution for polynomial chaos representation of data : A maximum-entropy and Fisher information approach. *SIAM Journal on Scientific Computing*, (accepted), 2008.

[DLRC05] O. Diard, S. Leclercq, G. Rousselier, and G. Cailletaud. Evaluation of finite element based analysis of 3D multicrystalline aggregates plasticity: Application to crystal plasticity model identification and the study of stress and strain fields near grain boundaries. *International Journal of Plasticity*, 21(4):691–722, 2005.

[DM01] G. Deodatis and R.C. Micaletti. Simulation of highly skewed non-Gaussian stochastic processes. *Journal of Engineering Mechanics - ASCE*, 127(12):1284–1295, December 2001.

[DM03] Sonjoy Das and C.S. Manohar. Prediction of vibration energy flow variability in random built-up structures. In *Proceedings of the 74th Shock and Vibration Symposium*, 2003.

[DNM+03] B.J. Debusschere, H.N. Najm, A. Matta, O.M. Knio, R.G. Ghanem, and O.P. Le Maître. Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation. *Physics Of Fluids*, 15(8):2238–2250, August 2003.

[DNP+04] B.J. Debusschere, H.N. Najm, P.P. Pébay, O.M. Knio, R. Ghanem, and O.P. Le Maître. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM J. Sci. Comput.*, 26(2):698–719, 2004.

[DOS06] X. Du and Martin Ostoja-Starzewski. On the scaling from statistical to representative volume element in the thermoelasticity of random materials. *Networks and heterogeneous media*, 1(2):259–274, June 2006.

[DSC04] C. Desceliers, C. Soize, and S. Cambier. Non-parametric-parametric model for random uncertainties in non-linear structural dynamics: Application to earthquake engineering. *Earthquake Engineering and Structural Dynamics*, 33(4):315–327, 2004.

[DSG07] C. Descelliers, C. Soize, and R. Ghanem. Identification of chaos representations of elastic properties of random media using experimental vibration tests. *Computational Mechanics*, 39(6):831–838, 2007.

[DW96] W. J. Drugan and J. R. Willis. A micromechanics-based nonlocal constitutive equation and estimates of representative volume element size for elastic composites. *Journal of the Mechanics and Physics of Solids*, 44:497–524, 1996.

[DWR07] F. P. E. Dunne, A. Walker, and D. Rugg. A systematic study of hcp crystal orientation and morphology effects in polycrystal deformation and fatigue. *Proceedings of the Royal Society of London. Series A*, 463(2082):1467–1489, March 2007.

[EE03] W. E and B. Engquist. The heterogeneous multiscale methods. *Communications in Mathematical Sciences*, 1(1):87–133, 2003.

[EMS01] P. Embrechts, A.J. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls. In M.A.H. Dempster, editor, *Risk Management : Value at Risk and Beyond*, pages 176–223. Cambridge University Press, 2001.

[Fis96] G.S. Fishman. *Monte Carlo: Concepts, Algorithms and Applications*. Springer, New York, 1996.

[Fla89] H. Flanders. *Differential Forms with Applications to the Physical Sciences*. Dover Publications Inc., New York, 1989.

[FNS+07] Jacob Fish, Mohan A. Nuggehally, Mark S. Shephard, Catalin R. Picu, Santiago Badia, Michael L. Parks, and Max Gunzburger. Concurrent AtC coupling based on a blend of the continuum stress and the atomistic force. *Computer Methods in Applied Mechanics and Engineering*, 196(45-48):4548–4560, 2007.

[For08] P. Forrester. *Log-gases and Random matrices*. Book in progress, 2008. `http://www.ms.unimelb.edu.au/~matpjf/matpjf.html`.

[FOT+00] S. Finette, M.H. Orr, A. Turgut, J.R. Apel, C.S. Badiey, B. Chiu, J. Headrick, J.N. Kemp, J.F. Lynch, A.E. Newhall, K.von der Heydt, B.H. Pasewark, S.N. Wolf, and D. Tielbuerger. Acoustic field variability induced by time evolving internal wave fields. *Journal of the Acoustical Society of America*, 108(3):957–972, September 2000.

[FRT97] Shu-Cheng Fang, J.R. Rajasekera, and H.S.J. Tsao. *Entropy Optimization, and Mathematical Programming*. Kluwer Academic Publishers, Boston, USA, 1997.

[GD06] R. Ghanem and A. Doostan. On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *J. Comput. Phys.*, 217(1):63–81, September 2006.

[GGG05] T. Gneiting, M.G. Genton, and P. Guttorp. Geostatistical space-time models, stationarity, separability and full symmetry. Technical Report 475, Department of Statistics, University of Washington, 2005. Also references therein. `http://www.stat.washington.edu/www/research/reports/2005/tr475.pdf`.

[GGRH05] D. Ghosh, R. Ghanem, and J. Red-Horse. Analysis of eigenvalues and modal interaction of stochastic systems. *AIAA Journal*, 43(10):2196–2201, October 2005.

[GH02] S. Ghosh and S.G. Henderson. Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research*, 50(5):820–834, September-October 2002.

[GH03] S. Ghosh and S.G. Henderson. Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation*, 13(3):276–294, July 2003. (for a short and conference version of the paper, see `http://www.informs-cs.org/wsc02papers/034.pdf`).

[Gha99] R. Ghanem. Ingredients for a general purpose stochastic finite elements implementation. *Computer Methods in Applied Mechanics and Engineering*, 168(1-4):19–34, 1999.

[GN00] A.K. Gupta and D.K. Nagar. *Matrix Variate Distribution*. Chapman & Hall/CRC, Boca Raton, 2000.

[GRH99] R. Ghanem and J. Red-Horse. Propagation of uncertainty in complex physical systems using a stochastic finite element approach. *Phys. D*, 133(1-4):137–144, 1999.

[Gri98] M. Grigoriu. Simulation of stationary non-Gaussian translation processes. *Journal of Engineering Mechanics - ASCE*, 124(2):121–126, February 1998.

[GS91] R. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, USA, 1991. Revised edition published by Dover Publications Inc. in 2003.

[GSD07] R. Ghanem, G. Saad, and A. Doostan. Efficient solution of stochastic systems: Application to the embankment dam problem. *Structural Safety*, 29(3):238–251, 2007.

[Gus97] Andre1 A. Gusev. Representative volume element size for elastic composites : A numerical study. *Journal of the Mechanics and Physics of Solids*, 45(9):1449–1459, 1997.

[Har97] D.A. Harville. *Matrix Algebra from a Statistician's Perspective*. Springer, 1997.

[HD97] J.T.G. Hwang and A.A. Ding. Prediction intervals for artificial neural networks. *Journal of the American Statistical Association*, 92(438):748–757, June 1997.

[HH94] S. Hazanov and C. Huet. Order relationship for boundary conditions effect in heterogonous bodies smaller than the representative volume. *Journal of the Mechanics and Physics of Solids*, 42(12):1995–2011, 1994.

[Hil63] R. Hill. Elastic properties of reinforced solids: Some theoretical principles. *Journal of the Mechanics and Physics of Solids*, 11(5):357–372, 1963.

[HJ85] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[HLD04] W. Hörmann, J. Leydold, and G. Derflinger. *Automatic Nonuniform Random Variate Generation*. Springer, 2004.

[HM00] A. Haldar and S. Mahadevan. *Reliability Assessment using Stochastic Finite Element Analysis*. John Wiley & Sons, 2000.

[HS63] Z. Hashin and S. Shtrikman. A variational approach to the theory of the elastic behavior of multiphase materials. *Journal of the Mechanics and Physics of Solids*, 11:127–140, 1963.

[Hue90] C. Huet. Application of variational concepts to size effects in elastic heterogenous bodies. *Journal of the Mechanics and Physics of Solids*, 38(6):813–841, 1990.

[Ize91] A.J. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, March 1991.

[Jay57a] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, May 1957.

[Jay57b] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 108(2):171–197, October 1957.

[Jay68] E.T. Jaynes. Prior probabilities. *IEEE Transactions on Systems and Cybernetics*, SSC-4(3):227–241, 1968.

[JBF88] B. Jetmundsen, R.L. Bielawa, and W.G. Flannelly. Generalized frequency domain substructure synthesis. *Journal of the American Helicopter Society*, 33(1):55–64, 1988.

[Jef46] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461, September 1946.

[Joe97] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 1997.

[Jol02] I.T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, October 2002.

[Kap89] J.N. Kapur. *Maximum Entropy Models in Science and Engineering*. Wiley Eastern Limited, India, 1989. Revised edition in 1993.

[KC06] D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons Ltd, 2006.

[KE06] P. Koev and A. Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. *Mathematics and Computation*, 75(254):833–846, 2006.

[KFG+03] T. Kanit, S. Forest, I. Galliet, V. Mounoury, and D. Jeulin. Determination of the size of the representative volume element for random composites: statistical and numerical approach. *International Journal of Solids and Structures*, 40(13-14):3647–3679, 2003.

[KGB02] V. Kouznetsova, M. G. D. Geers, and W. A. M. Brekelmans. Multi-scale constitutive modelling of heterogeneous materials with a gradient-enhanced computational homogenization scheme. *International Journal for Numerical Methods in Engineering*, 54(8):1235–1260, 2002.

[KGB04] V. Kouznetsova, M. G. D. Geers, and W. A. M. Brekelmans. Size of a representative volume element in a second-order computational homogenization framework. *International Journal for Multiscale Computational Engineering*, 2(4):575–598, 2004.

[KGL04] Kai Kadau, Timothy C. Germann, and Peter S. Lomdahl. Large-scale molecular-dynamics simulation of 19 billion particles. *International Journal of Modern Physics C*, 15(1):193–201, 2004. Also see `http://www.lanl.gov/orgs/t/publications/research_highlights_2005/docs/SF05_Kadau_WorldRecord.pdf`.

[Kha70] C. G. Khatri. A note on Mitra's paper "A density-free approach to the matrix variate beta distribution". *Sankhyā: Series A*, 32:311–318, 1970.

[KK92] J.N. Kapur and H.K. Kesavan. *Entropy Optimization Principles with Applications*. Academic Press Inc., Boston, USA, 1992.

[KL51] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[KPU04] Y. Kouskoulas, L.E. Pierce, and F.W. Ulaby. A computationally efficient multivariate maximum-entropy density estimation (MEDE) technique. *IEEE Transactions on Geoscience and Remote Sensing*, 42(2):457–468, February 2004.

[KTH92] M. Kleiber, D.H. Tran, and T.D. Hien. *The Stochastic Finite Element Method*. John Wiley & Sons, 1992.

[Kul59] S. Kullback. *Information Theory and Statistics*. John Wiley & Sons, New York, 1959. Republished by Dover Publications Inc. in 1997.

[Lan57] H.O. Lancaster. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44(1-2):289–292, June 1957.

[Leb72] N.N. Lebedev. *Special Functions and Their Applications*. Dover Publications, New York, 1972.

[Leh75] E.L. Lehmann. *Nonparametrics: Statistical methods based on ranks*. McGraw-Hill International Book Company, New York, 1975.

[Lin88] B.G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

[Liu00] W. Liu. *Structural Dynamic Analysis and Testing of Coupled Structures*. PhD thesis, Department of Mechanical Engineering, Imperial College of Science, Technology and Medicine, University of London, October 2000.

[LKP06] Wing Kam Liu, Eduard G. Karpov, and Harlod S. Park. *Nano Mechanics and Materials: Theory, Multiscale Methods and Applications*. John Wiley & Sons, Ltd., 2006.

[LL04] C.Q. Liu and X. Liu. A new method for analysis of complex structures based on FRF's of substructures. *Shock and Vibration*, 11(1):1–7, 2004.

[LLY99] Ju Li, Dongyi Liao, and Sidney Yip. Nearly exact solution for coupled continuum/MD fluid simulation. *Journal of Computer-Aided Materials Design*, 6:95–102, 1999.

[LMNGK04] O.P. Le Maître, H. Najm, R. Ghanem, and O. Knio. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J. Comput. Phys.*, 197(2):502–531, 2004.

[LMNP+07] O.P. Le Maître, H. Najm, P.P. Pébay, R.G. Ghanem, and O.M. Knio. Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.*, 29(2):864889, 2007.

[Loe78] M. Loève. *Probability Theory II*. Springer-Verlag, New York, 1978.

[Lof04] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the IEEE Conference on Computer Aided Control Systems Design (CACSD)*, Taipei, Taiwan, 2004. `http://control.ee.ethz.ch/~joloef/yalmip.php`.

[Man01] P.S. Mann. *Introductory Statistics*. John Wiley & and Sons, 4th edition, 2001.

[Mat97] A.M. Mathai. *Jacobians of Matrix Transformations and Functions of Matrix Argument*. World Scientific, 1997.

[MB93] A.M.H. Meeuwissen and T. Bedford. Probability distributions with given marginals and given correlation that have maximal entropy. Technical Report 93-81, Department of Technical Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands, 1993.

[MB97] A.M.H. Meeuwissen and T. Bedford. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57(1-4):143–174, 1997.

[Meh04] M.L. Mehta. *Random Matrices*. Academic Press, 3rd edition, 2004.

[MI99] C.S. Manohar and R.A. Ibrahim. Progress in structural dynamics with stochastic parameter variations: 1987-98. *Applied Mechanical Reviews*, 52(5):177–197, 1999. also references therein.

[Mit70] S. K. Mitra. A density-free approach to the matrix variate beta distribution. *Sankhayã: Series A*, 32:81–88, 1970.

[Mur82] R.J. Murihead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons Inc., 1982. Revised printing in 2005.

[MVL06] Cahal McVeigh, Franck Vernerey, and Brinson L. Cate Liu, Wing Kam. Multiresolution analysis for material design. *Computer Methods in Applied Mechanics and Engineering*, 195:5053–5076, 2006.

[Nea03] Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–741, June 2003. with discussions.

[Nel06] R.B. Nelsen. *An Introduction to Copulas*. Springer, 2006.

[NG02] D.K. Nagar and A.K. Gupta. Matrix-variate Kummer-Beta distribution. *Journal of the Australian Mathematical Society*, 73:11–25, 2002.

[NNH99] Sia Nemat-Nasser and Muneo Hori. *Micromechanics: Overall properties of heterogeneous materials*. Elsevier, Amsterdam, 2nd revised edition, 1999.

[OS01] Martin Ostoja-Starzewski. Mechanics of random materials: Stochastics, scale effects and computation. In Dominique Jeulin and Martin Ostoja-Starzewski, editors, *Mechanics of Random and Multiscale Microstructures*, pages 93–161. Springer-Verlag, 2001.

[OS02] Martin Ostoja-Starzewski. Microstructural randomness versus representative volume element in thermodynamics. *Journal of Applied Mechanics*, 69(1):25–35, January 2002.

[OS08] Martin Ostoja-Starzewski. *Microstructural Randomness and Scaling in Mechanics of Materials*. Chapman & Hall/CRC, 2008.

[OSDKL07] M. Ostoja-Starzewski, X. Du, Z.F. Khisaeva, and W. Li. Comparisons of the size of the representative volume element in elastic, plastic, thermoelastic, and permeable random microstructures. *International Journal for Multiscale Computational Engineering*, 5(2):73–82, 2007.

[PB06] C. L. Pettit and P. S. Beran. Spectral and multiresolution Wiener expansions of oscillatory stochastic processes. *Journal of Sound and Vibration*, 294:752–779, 2006.

[PG00] M.F. Pellissetti and R.G. Ghanem. Iterative solution of systems of linear equations arising in the context of stochastic finite elements. *Advances in Engineering Software*, 31:607–616, 2000.

[PG04] M. Pellissetti and R. Ghanem. A method for validation of the predictive computations using a stochastic approach. *Journal of Offshore Mechanics and Arctic Engineering*, 126:227–234, August 2004.

[Phi03] G.M. Phillips. *Interpolation and Approximation by Polynomials*. Springer, 2003.

[PPS02] B. Puig, F. Poirion, and C. Soize. Non-Gaussian simulation using Hermite polynomial expansion: convergences and algorithms. *Probabilistic Engineering Mechanics*, 17(3):253–264, 2002.

[PQH04] K.K. Phoon, S.T. Quek, and H. Huang. Simulation of non-Gaussian process using fractile correlation. *Probabilistic Engineering Mechanics*, 19:287–292, 2004.

[PTVF96] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. Cambridge University Press, 1996.

[RB95] Y. Ren and C.F. Beards. On substructure synthesis with FRF data. *Journal of Sound and Vibration*, 185(5):845–866, 1995.

[RDD07] D. Rugg, M. Dixon, and F. P. E. Dunne. Effective structural unit size in titanium alloys. *The Journal of Strain Analysis for Engineering Design*, 42(4):269–279, 2007.

[RKL+02] Cindy L. Rountree, Rajiv K. Kalia, Elefterios Lidorikis, Aiichiro Nakano, Laurent Van Brutzel, and Priya Vashishta. Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular dynamics simulations. *Annual Review of Materials Research*, 32:377–400, 2002.

[RNGK03] M.T. Reagan, H.N. Najm, R.G. Ghanem, and O.M. Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion And Flame*, 132(3):545–555, February 2003.

[Ros52] M. Rosenblatt. Remarks on multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, September 1952.

[RZ04] Z.-Y. Ren and Q.-S. Zheng. Effect of grain sizes, shapes, and distribution on minimum sizes of representative volume elements of cubic polycrysrals. *Mechanics of Materials*, 36:1217–1229, 2004.

[Sco92] D.W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley-Interscience, August 1992.

[SFED+04] David M. Saylor, Joseph Fridy, Bassem S. El-Dasher, Kee-Young Jung, and Anthony D. Rollett. Statistically representative three-dimensional microstructures based on orthogonal observation sections. *Metallurgical and Materials Transactions A — Physical Metallurgy and Materials Science*, 35A(7):1969–1979, July 2004.

[SG02a] S. Sakamoto and R. Ghanem. Polynomial chaos decomposition for the simulation of non-gaussian nonstationary stochastic processes. *Journal of Engineering Mechanics - ASCE*, 128(2):190–201, February 2002.

[SG02b] S. Sakamoto and R. Ghanem. Simulation of multi-dimensional non-Gaussian non-stationary random fields. *Probabilistic Engineering Mechanics*, 17(2):167–176, April 2002.

[SG04a] C. Soize and R. Ghanem. Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, 26(2):395–410, 2004.

[SG04b] Shriram Swaminathan and Somnath Ghosh. Statistically equivalent representative volume elements for unidirectional composite microstructures: Part ii — with interfacial debonding. *Journal of Composite Materials*, 40(7):605–621, 2004.

[SGP04] Shriram Swaminathan, Somnath Ghosh, and N. J. Pagano. Statistically equivalent representative volume elements for unidirectional composite microstructures: Part i - without damage. *Journal of Composite Materials*, 40(7):583–604, 2004.

[Sha48] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3,4):379–423, 623–656, July, October 1948.

[SJ80] J.E. Shore and R.W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26(1):26–37, January 1980.

[SJ83] J.E. Shore and R.W. Johnson. Comments on and correction to "axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy". *IEEE Transactions on Information Theory*, IT-29(6):942–943, 1983.

[SK79] M. S. Srivastava and C. G. Khatri. *An Introduction to Multivariate Statistics*. North Holland, New York, 1979.

[SK97] K. Shankar and A.J. Keane. Vibrational energy flow analysis using a substructure approach: The application of receptance theory to fea and sea. *Journal of Sound and Vibration*, 201(4):491–513, 1997.

[SKR00] M. Srikanth, H.K. Kesavan, and P.H. Roe. Probability density function estimation using the minmax measure. *IEEE Transactions on Systems, Man, and Cybernatics - Part C: Applications and Reviews*, 30(1):77–83, February 2000.

[Soi99] C. Soize. A nonparametric model of random uncertainties in linear structural dynamics. *Publications du LMA-CNRS, ISBN 2-909669-16-5*, 152:109–138, 1999.

[Soi00] C. Soize. A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, 15:277–294, 2000.

[Soi01a] C. Soize. Maximum entropy approach for modeling random uncertainties in transient elastodynamics. *Journal of the Acoustical Society of America*, 109(5):1979–1996, 2001. Pt. 1.

[Soi01b] C. Soize. Transient responses of dynamical systems with random uncertainties. *Probabilistic Engineering Mechanics*, 16:363–372, 2001.

[Soi03] C. Soize. Random matrix theory and non-parametric model of random uncertainties in vibration analysis. *Journal of Sound and Vibration*, 263:893–916, 2003.

[Soi05a] C. Soize. A comprehensive overview of non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics. *Journal of Sound and Vibration*, 288:623–652, 2005.

[Soi05b] C. Soize. Random matrix theory for modeling uncertainties in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 194:1333–1366, 2005.

[Soi06] C. Soize. Non-Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. *Computer Methods in Applied Mechanics and Engineering*, 195:26–64, 2006.

[Soi08] C. Soize. Tensor-valued random fields for meso-scale stochastic model of anisotropic elastic microstructure and probabilistic analysis of representative volume element size. *Probabilistic Engineering Mechanics*, (in press) doi:10.1016/j.probengmech.2007.12.019, 2008.

[Spa92] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Control*, 37(3):332–341, 1992.

[Spa03] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*. Wiley-Interscience, 2003.

[Spe04] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, January 1904.

[SZ06] Sethuraman Sankaran and Nicholas Zabaras. A maximum entropy approach for property prediction of random microstructures. *Acta Materialia*, 54:2265–2276, 2006.

[Tol62] G.P. Tolstov. *Fourier Series*. Prentice-Hall, 1962.

[Tor02] Salvatore Torquato. *Ranom Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, 2002.

[TPO00] E.B. Tadmor, R. Phillips, and M. Ortiz. Hierarchical modeling in the mechanics of materials. *International Journal of Solids and Structures*, 37(1-2):379–389, 2000.

[TV04] A.M. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications*. now Publishers Inc., 2004.

[Urg91] A.P.V. Urgueira. Using the s.v.d. for the selection of independent connection coordinates in the coupling of substructures. In *Proceedings of the 9th International Modal Analysis Conference*, pages 919–925, 1991.

[VB96] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[vdG98] P.A.G. van der Geest. An algorithm to generate samples of multi-variate distributions with correlated marginals. *Computational Statistics & Data Analysis*, 27(3):271–289, 1998.

[VG08] Ashkan Vaziri and Arvind Gopinath. Cell and biomolecular mechanics *in silico*. *Nature Materials*, 7(1):15–23, January 2008.

[Wie38] N. Wiener. The homogeneous chaos. *Amer. J. Math.*, 60(4):897–936, 1938.

[WSB07] J.A.S. Witteveen, S. Sarkar, and H. Bijl. Modeling physical uncertainties in dynamic stall induced fluidstructure interaction of turbine blades using arbitrary polynomial chaos. *Computers and Structures*, 85(11-14):866–878, June-July 2007.

[Wu03] X. Wu. Calculation of maximum entropy densities with application to income distribution. *J. Econometrics*, 115:347–354, 2003.

[XK02] D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

[XLSK02] D. Xiu, D. Lucor, C.-H. Su, and G.E. Karniadakis. Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering*, 124:51–59, March 2002.

[ZG04] Yu Zou and Roger Ghanem. A multiscale data assimilation with the ensemble Kalman filter. *Multiscale Modeling & Simulation*, 3(1):131–150, 2004.

[ZKG05] Yu Zou, Ioannis Kevrekidis, and Roger Ghanem. Equation-free dynamic renormalization: Self-similarity in multidimensional particle system dynamics. *Physical Review E*, 72:940–956, 2005.

[Zv01] J. Zeman and M. Šejnoha. Numerical evaluation of effective elastic properties of graphite fiber tow impregnated by polymer matrix. *Journal of the Mechanics and Physics of Solids*, 49:69–90, 2001.

[Zv07] J. Zeman and M. Šejnoha. From random microstructures to representative volume elements. *Modelling and Simulation in Materials Science and Engineering*, 15:S325–S335, 2007.

Appendices

Appendix A

Computation of PC Coefficients

This appendix describes an 1-dimensional (D) scheme based on interpolation technique. It would be useful for the efficient computation of

1. $\{a_j(y_2)\}_{j\in\mathbb{N}}$ in (3.7), or
2. $\{c_j\}_{j\in\mathbb{N}}$ in (3.12), or
3. $\{c_{jk}\}_{j\in\mathbb{N}}$ for any given $k \in \{1, \cdots, N\}$ in (3.15).

Since all these cases are similar, only the last case involving $\{c_{jk}\}_{j\in\mathbb{N}}$, $k \in \{1, \cdots, N\}$, would be demonstrated below. Any other case can be readily tackled by considering the appropriate PC coefficients and the PDFs.

Let the pdf and support of $\mathbf{y}_k$ be denoted, respectively, by $p_{\mathbf{y}_k}$ and $s_k = [l_k,\, m_k] \subset \mathbb{R}$. The computation of c_{jk} in (3.16) based on $q_k \equiv P^{-1}_{\mathbf{y}_k} P_{\boldsymbol{\xi}_k}$ needs solving an integral equation. For some given ξ_k, the integral equation, $P_{\boldsymbol{\xi}_k}(\xi_k) = \int_{l_k}^{y_k} p_{\mathbf{y}_k}(y)dy$, is to be solved for y_k. Solving this integral equation several times *within* the numerical integration algorithm, that is employed to compute c_{jk}, significantly increases the computational burden, and might also lead to certain numerical instability. To overcome these difficulties and to increase the computational expediency and efficiency, a surrogate function, $\tilde{q}_k$, determined based on 1-D interpolation scheme, is used for q_k in (3.16) to compute the PC coefficients, c_{jk}. The approximate function, $\tilde{q}_k$, needs to be determined only once $\forall j \in \mathbb{N}$.

Consider $u_k \equiv P_{\boldsymbol{\xi}_k}(\xi_k) \stackrel{d}{=} P_{\mathbf{y}_k}(y_k)$ (u_k here should not be confused with the components of $\mathbf{U}$ in sections 3.2.2–3.2.2). For a given $y_k \in s_{\mathbf{y}_k} = [l_k,\, m_k]$, finding u_k as $u_k = P_{\mathbf{y}_k}(y_k)$ is, in general, much cheaper than finding y_k as $y_k = P^{-1}_{\mathbf{y}_k}(u_k)$ for a given $u_k \in [0,\, 1]$.

For each $k \in \{1, \cdots, N\}$, let the support, s_k, be divided equally into $n_k \in \mathbb{N}$ intervals. Then, the coordinates of the points defining these intervals are given by $y_k^{(j)} = l_k + j[(m_k - l_k)/n_k]$, $j = 0, \cdots, n_k$. For each of these points, first compute $u_k^{(j)}$ as $u_k^{(j)} = P_{\mathbf{y}_k}(y_k^{(j)})$, and then compute $\xi_k^{(j)}$ as $\xi_k^{(j)} = P^{-1}_{\boldsymbol{\xi}_k}(u_k^{(j)})$. Since $P_{\boldsymbol{\xi}_k}$'s are suitably chosen standard measures associated with the commonly

used PC random variables, computation of $P^{-1}_{\boldsymbol{\xi}_k}$ via closed form expression or efficient algorithms is available in the statistical literature (see e.g., [Fis96, Section 3.2], [HLD04, Section 2.1]). As already indicated, the statistical toolbox of MATLAB provides functions to evaluate the inverse of PDF for many such standard PC random variables. Since $P_{\mathbf{y}_k}$ and $P_{\boldsymbol{\xi}_k}$ are monotonically increasing function, the set of values in $\{\xi_k^{(j)}\}_{j=0}^{n_k}$ would be in the increasing order, $\xi_k^{(0)} < \cdots < \xi_k^{(n_k)}$. The set, $\{\xi_k^{(j)}, y_k^{(j)}\}_{j=0}^{n_k}$, thus determined is now used to construct the approximate function, $s_{\boldsymbol{\xi}_k} \ni \xi_k \longmapsto \tilde{q}_k(\xi_k) \in s_{\mathbf{y}_k}$, by using standard interpolation technique (see e.g., [Phi03, Chapter 1], [PTVF96, Chapter 3]). The basic MATLAB package offers a function, `interp1`, use of which should be sufficient enough for determining $\tilde{q}_k$ for many practical purposes. The approximate function, $\tilde{q}_k$, is used as a proxy for q_k in (3.16) to compute the PC coefficients, c_{jk}'s.

The error in approximating $q_k(\boldsymbol{\xi}_k)$ by the resulting PC representation, $\tilde{q}_k^{(K_k)}(\boldsymbol{\xi}_k) = \sum_{j=0}^{K_k} c_{jk}\, \boldsymbol{\Psi}_j(\boldsymbol{\xi}_k)$, for some large $K_k \in \mathbb{N}$, is bounded above by the following relation,

$$\begin{aligned} |\, q_k(\boldsymbol{\xi}_k) - \tilde{q}_k^{(K_k)}(\boldsymbol{\xi}_k)\,| \;\; &\leq \;\; |\, q_k(\boldsymbol{\xi}_k) - \tilde{q}_k(\boldsymbol{\xi}_k)\,| \\ &\quad + \; |\, \tilde{q}_k(\boldsymbol{\xi}_k) - \tilde{q}_k^{(K_k)}(\boldsymbol{\xi}_k)\,| \;\; \text{a.s.} \end{aligned} \tag{A.1}$$

The second error term is bounded above by some $e_{K_k}(\boldsymbol{\xi}_k)$ satisfying $\lim_{K_k \to \infty} e_{K_k}(\boldsymbol{\xi}_k) = 0$ [Leb72, Chapter 4] a.s. When a linear interpolation scheme is employed, the interpolated function, $\tilde{q}_k$, is piecewise linear in ξ_k and the first error term is then bounded above by $O(h_k^2)$, in which $h_k = \max_{1 \leq i \leq n_k}(\xi_k^{(i)} - \xi_k^{(i-1)})$, [Phi03, Example 1.1.4] a.s. In establishing this error bound, $O(h_k^2)$, it is necessary for the second derivative of q_k to be piecewise bounded by some finite $\mathcal{K}$, i.e., $|\, \partial^2 q_k(\xi_k)/\partial \xi_k \,| \leq \mathcal{K}$ on $s_{\boldsymbol{\xi}_k}$ except possibly a finite number of points. As it is already mentioned that an assumption of piecewise smoothness is required to arrive at the PC representation in a.s. sense, the piecewise linear function, $\tilde{q}_k$, that is actually being represented by PC formalism, automatically satisfies the assumption of piecewise smoothness. Therefore, in order to satisfy (A.1), the assumption of piecewise smoothness of the original function, q_k, needs to be replaced by a relatively stronger assumption of the piecewise boundedness of the second derivative of the function, q_k.