An Uncertainty Multidisciplinary Propagation Analysis Method for Correlated High-Dimensional Data in Aircraft Design

Siyi Du [1], Chunna Li [1], Yang Liu [1]

[1] Shaanxi Aerospace Flight Vehicle Design Key Laboratory, School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

Abstract

The discipline decoupling of distributed Uncertainty Multidisciplinary Design Optimization method in aircraft reliability-based design result in the ignore of the correlation between the state variables, particularly high-dimensional variables (data), leading to inaccurate outcomes. To address this issue, we propose an uncertainty multidisciplinary propagation analysis method that can consider correlated high-dimensional data. Our approach involves utilizing the Nataf transformation method to convert uncorrelated high-dimensional variables into correlated variables. Additionally, we introduce a warm-up strategy to efficiently compute the marginal Cumulative Distribution Function of high-dimensional variables by the Maximum Entropy Model. This enables us to efficiently and accurately quantify the uncertainties of the high dimensional state variables. The results of the analytical problem and the solid rocket problem indicate that our method can enhance uncertainty quantification accuracy by accounting for the correlations between state variables. This improvement is particularly significant in the context of reliability-based design.

Keywords: multidisciplinary uncertainty propagation, correlated high-dimensional data, nataf transformation, maximum entropy model, warm-up strategy

Nomenclature

$\boldsymbol{X}$	=	Inputs of the training set of samples in the design space
$\hat{\boldsymbol{X}}$	=	Inputs of the testing set of samples in the design space
m	=	Size of the samples in the design space
$\boldsymbol{X}_{\Sigma}$	=	Inputs of all the samples in both of the design and uncertainty spaces
n	=	Size of the samples in the uncertainty space
$\boldsymbol{Y}_{\mathrm{C}}$	=	Matrix of correlated high-dimensional output data of the training set
$\boldsymbol{Y}_{\mathrm{C},i}$	=	The i^{th} high-dimensional correlated output data in $\boldsymbol{Y}_{\mathrm{C}}$, $i=1,\ldots,r$, where r is the size of high-dimensional outputs
$\hat{\boldsymbol{Y}}_{\mathrm{U}}$	=	The predicted high-dimensional uncorrelated output data
$\hat{\boldsymbol{Y}}_{\mathrm{U}}^{\mathrm{S}}$	=	The predicted high-dimensional uncorrelated output data in the standard normal space
$\hat{\boldsymbol{Y}}_{\mathrm{C}}$	=	The predicted high-dimensional correlated output data
$\boldsymbol{\rho}$	=	Pearson correlation coefficient matrix of output data
$\hat{\boldsymbol{\rho}}$	=	Pearson correlation coefficient matrix of the predicted output data

$\boldsymbol{\Phi}$	=	POD orthogonal basis matrix
$\boldsymbol{a}$	=	POD mode coefficient matrix
$\boldsymbol{\mu}$	=	The central moments of each POD mode coefficient vector
$g_{\mu_i}(\cdot)$	=	The surrogate model of the i^{th} central moments correlation $\boldsymbol{\mu}$
$g_{\rho}(\cdot)$	=	The surrogate model of the correlation coefficient matrix $\boldsymbol{\rho}$
$\boldsymbol{P}(\cdot)$	=	Probability density function
$\boldsymbol{F}(\cdot)$	=	Cumulative Distribution Function

1. Introduction

Complex systems such as aircraft systems are constituted of multiple disciplines, including aerodynamics, structure, trajectory, and so on. This complexity introduces a significant amount of uncertainty during design, manufacturing and application. To address these uncertainties, Uncertainty Multidisciplinary Design Optimization (UMDO) method is utilized to enhance the reliability and robustness of aircraft systems.

The typical UMDO framework involves an outer layer optimization nested within an inner layer uncertainty analysis. A common method for uncertainty analysis often utilizes the Monte Carlo method, which requires numerous high-fidelity numerical simulations. To improve efficiency, surrogate models are frequently utilized. Several uncertainty analysis methods based on surrogate models have been proposed, including the polynomial chaos expansions [1], Kriging [2], radial basis function [3], and neural networks [4]. These methods are only suitable for describing the uncertainty with mean and standard deviations. While, after uncertainty propagation across a discipline, the distributions of the discipline outputs are often not standard normal distributions [5-6] Shi proposed an accurate uncertainty surrogate modeling method for high-dimensional data with non-normal distributions employing the maximum entropy method (MaxEnt) [7].

Among various UMDO methods, the distributed UMDO method is particularly suitable for tackling complex multidisciplinary system problems. This method simplifies the propagation of uncertainties across multiple disciplines, allowing each discipline to conduct its own uncertainty analysis independently. Zaman proposes a single loop optimization design method that avoids coupled uncertainty propagation analysis [8]. Brevault proposes a new decoupled uncertainty multidisciplinary design optimization formulation ensuring the equivalence between the coupling and decoupling uncertain multi-disciplinary design optimization formulas [9]. More research in field of distributed UMDO method can be founded in the work of Qin [10] and An [11].

However, the above distributed UMDO methods do not consider the correlations between the state variables in uncertainty analysis, which will lead to accuracy loss in uncertainty propagation, limiting the performance of the multidisciplinary system. The Rosenblatt transform [12], Nataf transform [13] and Copula function [14] are often used to deal with one-dimensional correlated variables. Ghosh proposed a methodology for uncertainty propagation analysis within a distributed multidisciplinary architecture, which takes into account the correlation of one-dimensional state variables based on the Copula function [15].

One limitation in current research is the limited attention to the correlation between high-dimensional variables in uncertainty propagation analysis in the multidisciplinary system, such as the aerodynamic distributed loads, distributed stress of structural components, or thrust curves of solid engines. Therefore, it is crucial to develop a multidisciplinary uncertainty propagation analysis method considering the correlation between high dimensional data for the fine design of aircraft.

We propose an uncertainty multidisciplinary propagation analysis method for correlated high-dimensional data, which is validated and assessed through an analytical problem and a solid launch vehicle problem. The paper is organized as follows. Section 2 introduces the proposed uncertainty

multidisciplinary propagation analysis method in detail. Then the method is validated and assessed in Section 3. In section 4 we draw the conclusions.

2. Methodology

2.1 The procedure of the uncertainty multidisciplinary propagation analysis method for correlated high-dimensional data

The procedure of the uncertainty multidisciplinary propagation analysis method for correlated high-dimensional data is depicted in Figure 1, and detailed steps are introduced in the following.

Step 1: High-dimensional sample generation.

Determine the design space of the discipline design variables. Then the sampling plan with m samples are generated using the Random Latin Hypercube Sampling (RLHS), which is expressed as (1).

$$\boldsymbol{X} = \left[\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots \boldsymbol{X}_m\right] \tag{1}$$

where $\boldsymbol{X}_i$ is the i^{th} design. To describe the uncertainty distribution of each design point $\boldsymbol{X}_i$, an uncertainty space is specified around each design point for uncertainty sampling. It is defined to generate n samples within each uncertainty space, thus inputs of all the samples in both of the design and uncertainty spaces are expressed as (2).

$$\boldsymbol{X}_\Sigma = \left[\boldsymbol{X}_{11}, \cdots, \boldsymbol{X}_{1n}, \boldsymbol{X}_{21}, \cdots, \boldsymbol{X}_{2n}, \cdots, \boldsymbol{X}_{m\times n}\right] \tag{2}$$

Under the inputs of $\boldsymbol{X}_\Sigma$, the correlated high-dimensional output variables $\boldsymbol{Y}_{\text{C}} = \left[\boldsymbol{Y}_{\text{C},1}, \boldsymbol{Y}_{\text{C},2}, \cdots, \boldsymbol{Y}_{\text{C},r}\right]$ can be obtained by discipline analysis, where $\boldsymbol{Y}_{\text{C},i} = \left[\boldsymbol{Y}_{\text{C},i,1}, \boldsymbol{Y}_{\text{C},i,2}, \cdots, \boldsymbol{Y}_{\text{C},i,d}\right] \in \mathbb{R}^{mn\times d}$.

Step 2: Then the Pearson correlation coefficient matrix $\boldsymbol{\rho}$ of high-dimensional output variables $\boldsymbol{Y}_{\text{C}}$ is calculated.

Step 3: Build the surrogate model for central moments and correlations coefficient matrixes

The Proper Orthogonal Decomposition (POD) method is used to project $\boldsymbol{Y}_{\text{C},i}$ onto a low-dimensional space. The POD modes are truncated based on the energy proportion calculation, then the POD orthogonal basis matrix Φ and mode coefficient matrix $\boldsymbol{a}$ are obtained. Thereafter, the central moments $\boldsymbol{\mu}$ of each mode coefficient vector can be computed to describe its' uncertainty distribution. The size of the central moments is problem dependent. Further, the Back-Propagation Neural Network (BPNN) method is used to construct the surrogate model $\boldsymbol{g} = \left[g_{\mu_1}, \cdots, g_{\mu_k}, g_\rho\right]$, which maps the m-dimensional design variable $\boldsymbol{X}$ to the central moments correlation $\boldsymbol{\mu}$ and the correlation coefficient matrix $\boldsymbol{\rho}$.

$$\boldsymbol{\mu}_i = g_{\mu_i}\left(\boldsymbol{X}\right) \tag{3}$$

$$\boldsymbol{\rho} = g_\rho\left(\boldsymbol{X}\right) \tag{4}$$

Step 4: Compute high-dimensional output data by POD reconstruction

Using the built surrogate model $\boldsymbol{g}$, the predicted correlation coefficient matrix $\hat{\boldsymbol{\rho}}$ and central moments $\hat{\boldsymbol{\mu}}$ of the POD mode coefficients with a new $\hat{\boldsymbol{X}}$ can be obtained.

Then, the Probability Density Function (PDF) $\boldsymbol{P}\left(a\right)$ can be computed using the MaxEnt method. Moreover, we can generate samples of the mode coefficients $\hat{\boldsymbol{a}}$ using $\boldsymbol{P}\left(a\right)$, and obtain the high-dimensional output data $\hat{\boldsymbol{Y}}_{\text{U}}$ by POD reconstruction.

$$\hat{\boldsymbol{Y}}_{\text{U}} = \hat{\boldsymbol{a}}^T \Phi \tag{5}$$

Currently, each output data of $\hat{\boldsymbol{Y}}_{\mathrm{U}}$ is independent.

Step 5: Add correlated features to $\hat{\boldsymbol{Y}}_{\mathrm{U}}$

Calculate the central moments $\hat{\boldsymbol{\mu}}_j\left(j=1,\ldots d\right)$ for each uncorrelated output of $\hat{\boldsymbol{Y}}_{\mathrm{U}}$. Based on $\hat{\boldsymbol{\mu}}_j$, the MaxEnt method is used to compute the PDF $\boldsymbol{P}_j\left(\boldsymbol{Y}\right)$ of $\hat{\boldsymbol{Y}}_{\mathrm{U}}$, then the Cumulative Distribution Function (CDF) $\boldsymbol{F}_j\left(\boldsymbol{Y}\right)$ can be obtained by integrating $\boldsymbol{P}_j\left(\boldsymbol{Y}\right)$. To improve computation efficiency, the warm-up strategy is introduced. $\hat{\boldsymbol{\mu}}_j$ is first sorted in ascending order based on the deviation $\Delta=\sum_{i=1}^{k}\left(\mu_{j,i}-\overline{\mu}_i\right)$, where $\mu_{j,i}$ is the i^{th} central moment of $\hat{\boldsymbol{\mu}}_j$ and $\overline{\mu}_i$ is the mean of the i^{th} central moment. Therefore, in the optimization process in MaxEnt method, the result obtained based on $\hat{\boldsymbol{\mu}}_j$ can be used as the initial value of the latter process based on $\hat{\boldsymbol{\mu}}_{j+1}$.

Based on $\boldsymbol{F}_j\left(\boldsymbol{Y}\right)$, convert $\hat{\boldsymbol{Y}}_{\mathrm{U}}$ from the original space to the standard normal space, denoted as $\hat{\boldsymbol{Y}}_{\mathrm{U}}^{\mathrm{S}}$.

$$\hat{\boldsymbol{Y}}_{\mathrm{U}}^{\mathrm{S}}=\phi^{-1}\left(F\left(\hat{\boldsymbol{Y}}_{\mathrm{U}}\right)\right) \tag{6}$$

By employing inverse Nataf transformation method based on $\hat{\boldsymbol{\rho}}$ and $\boldsymbol{F}_j\left(\boldsymbol{Y}\right)$, we can further convert $\hat{\boldsymbol{Y}}_{\mathrm{U}}^{\mathrm{S}}$ to $\hat{\boldsymbol{Y}}_{\mathrm{C}}$.

$$\hat{\boldsymbol{Y}}_{\mathrm{C}}=Nataf^{-1}\left(\hat{\boldsymbol{Y}}_{\mathrm{U}}^{\mathrm{S}}\right) \tag{7}$$

Step 6: Pass $\hat{\boldsymbol{Y}}_{\mathrm{C}}$ as input data to the next discipline model.

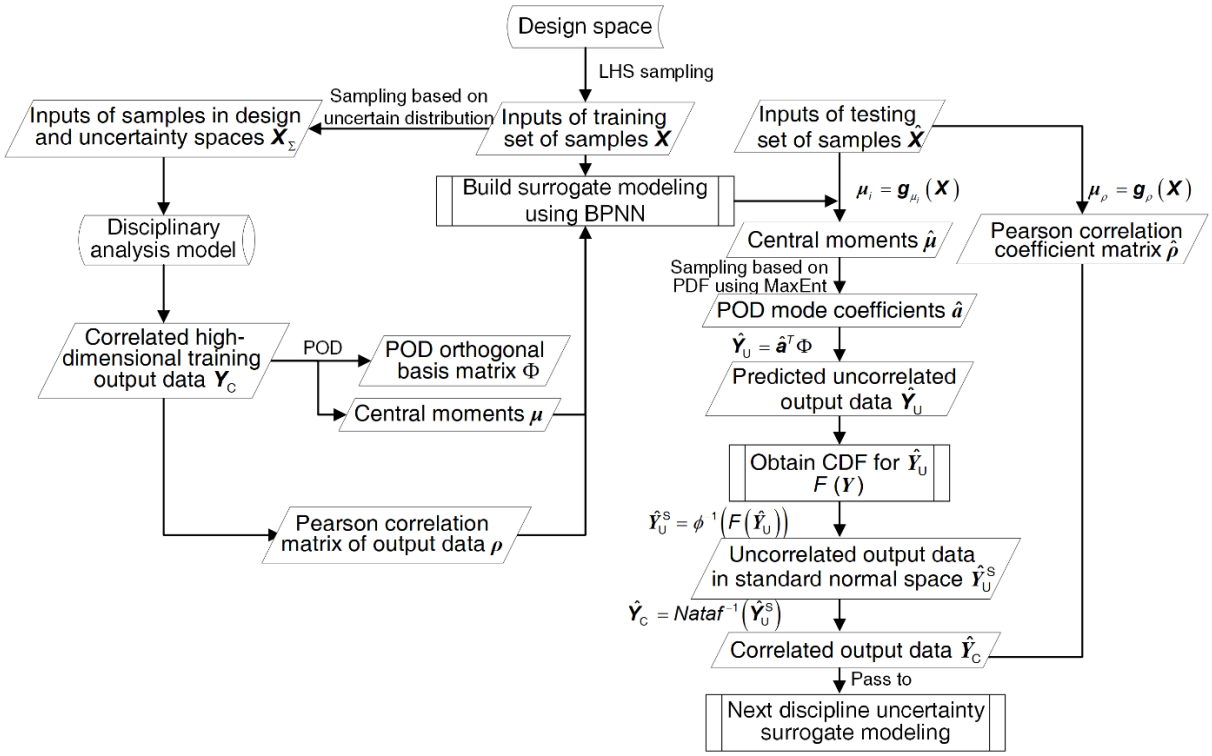

Figure 1 – Flowchart of the proposed uncertainty multidisciplinary propagation analysis method

2.2 POD model

POD can represent high-dimensional data by linear superposition of POD modes. As a reduced order

model, the POD can reduce high-dimensional data to low-dimensional data, which will enhance the efficiency of building surrogate models in the reduced order space.

The high-dimensional data $\boldsymbol{Y}_i(n\times d)$ is expressed as follows, where n is the number of $\boldsymbol{Y}_i$, and d denotes the dimension.

$$\boldsymbol{Y}_i=\begin{bmatrix}\boldsymbol{Y}_{i,11} & \boldsymbol{Y}_{i,21} & \cdots \boldsymbol{Y}_{i,d1}\\ \boldsymbol{Y}_{i,12} & \boldsymbol{Y}_{i,22} & \cdots \boldsymbol{Y}_{i,d2}\\ \vdots & \vdots & \vdots \ \vdots\\ \boldsymbol{Y}_{i,1n} & \boldsymbol{Y}_{i,2n} & \cdots \boldsymbol{Y}_{i,dn}\end{bmatrix} \tag{8}$$

The mean vector $\overline{\boldsymbol{y}}_i$ is computed using $\boldsymbol{y}_{i,j}\left(j=1,\ldots,d\right)$. Then subtract the mean $\overline{\boldsymbol{y}}_i$ from $\boldsymbol{Y}_i$ to obtain the normalized data $\overline{\boldsymbol{Y}}_i$.

$$\overline{\boldsymbol{Y}}_i=\left[\boldsymbol{y}_{i,1}-\bar{\boldsymbol{y}}_i \quad \boldsymbol{y}_{i,2}-\bar{\boldsymbol{y}} \quad \cdots \quad \boldsymbol{y}_{i,d}-\bar{\boldsymbol{y}}\right] \tag{9}$$

Then the correlation matrix for the normalized data is calculated by (10).

$$\boldsymbol{C}=\overline{\boldsymbol{Y}}_i^{\boldsymbol{T}}\overline{\boldsymbol{Y}}_i/n \tag{10}$$

Thus the eigenvalues and eigenvectors of the correlation matrix can be obtained by solving (11).

$$\boldsymbol{C\xi=\lambda\xi} \tag{11}$$

The POD mode coefficients $\boldsymbol{a}$ corresponding to (12).

$$\boldsymbol{a}_i=\boldsymbol{\xi}^T\overline{\boldsymbol{Y}}_i \tag{12}$$

To minimize the number of POD modes, a mode truncation method can be applied using the mode energy ratio. This ratio, denoted by $I\left(h\right)$, is defined as the proportion of the eigenvalue corresponding to a mode relative to the sum of all eigenvalues, represented by (13).

$$I\left(h\right)=\sum_{i}^{k}\lambda_i \ / \ \text{sum(all } \lambda_i) \tag{13}$$

Typically, $I\left(h\right)$ should exceed 99% to ensure accuracy [16].

2.3 BPNN algorithm

As a type of neural network, BPNN demonstrates strong nonlinear fitting capabilities for multi-input and multi-output data of lower dimensions [17]. The input layer components of the typical architecture of a single hidden layer BPNN are denoted as $x_i\left(i=1,2,\cdots n\right)$, the hidden layer components are denoted as $z_j\left(j=1,2,\cdots b\right)$, and output layer components are denoted as $y_k\left(k=1,2,\cdots s\right)$. By adjusting parameters such as the number of hidden layers, neurons in each layer, interlayer activation function type, learning rate, and other hyperparameters [18], the model can enhance its fitting and generalization abilities, enabling robust prediction of output values for new samples. The specific process of BPNN is elaborated in reference [19].

2.4 Nataf transformation

The Nataf transformation is an essential mathematical method that facilitates the mapping of a group of correlated random variables onto a set of uncorrelated standard normal variables. This transformation hinges on two key components: the marginal Cumulative Distribution Function (CDF) for each individual variable and the matrix of Pearson correlation coefficients that describe the relationships between the variables.

The Pearson correlation coefficient, denoted as $\boldsymbol{\rho}$, quantifies the linear correlation between two variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, with its value bounded between -1 to 1. When $\boldsymbol{\rho}$ is greater than zero, it

signifies a positive linear relationship between the variables; when $\boldsymbol{\rho}$ is less than zero, it indicates a negative linear relationship. The closer the absolute value of $\boldsymbol{\rho}$ is to 1, the stronger the correlation.

$$\boldsymbol{\rho} = \frac{\mathrm{cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)}{\sigma_{X_1}\sigma_{X_2}} \tag{14}$$

Suppose random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots \boldsymbol{X}_n$ with linear correlation, and denote the correlation coefficient matrix as $\boldsymbol{\rho} = \left[\rho_{ij}\right]_{n\times n}$. The correlated matrix $\boldsymbol{X} = \left[\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots \boldsymbol{X}_n\right]^T$ can be transformed into the correlated standard normal vector $\boldsymbol{Z}$ using the following formula.

$$\boldsymbol{Z}_i = \phi^{-1}\left(F_i\left(\boldsymbol{X}_i\right)\right) \tag{15}$$

where $\phi^{-1}(\cdot)$ denotes the inverse CDF of a standard normal variable, and $F_i(\cdot)$ is the CDF corresponding to $\boldsymbol{X}_i$.

Let $\boldsymbol{\rho}_0 = \left[\rho_{0ij}\right]_{n\times n}$ be the correlation coefficient matrix of $\boldsymbol{Z}$. Then the transition from the correlation coefficient ρ_{ij} in original space to the correlation coefficient ρ_{0ij} in standard normal space, can be achieved using the following integral function, as elaborated in the previous studies [20].

$$\rho_{ij} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{F_i^{-1}\left(f\left(\boldsymbol{Z}_i\right)\right) - \mu_i}{\sigma_i}\right]\left[\frac{F_j^{-1}\left(f\left(\boldsymbol{Z}_j\right)\right) - \mu_j}{\sigma_j}\right] f\left(\boldsymbol{z}_i, \boldsymbol{z}_j, \rho_{0ij}\right) dz_i dz_j \tag{16}$$

where $f\left(\boldsymbol{z}_i, \boldsymbol{z}_j, \rho_{0ij}\right)$ is the joint probability density function (PDF) of variables $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$, with the correlation coefficient ρ_{0ij}. Given that $\boldsymbol{\rho}_0$ is typically a symmetric matrix, it can be decomposed using the Cholesky Decomposition method, as shown in (17).

$$\boldsymbol{\rho}_0 = \boldsymbol{L}_0\boldsymbol{L}_0^T \tag{17}$$

where $\boldsymbol{L}_0$ denotes the lower triangular matrix obtained by the Choleskey decomposition. Then the correlated standard normal vector $\boldsymbol{Z}$ can be linearly transformed into an independent standard normal vector $\boldsymbol{Y}$ utilizing the matrix $\boldsymbol{L}_0$ by (18).

$$\boldsymbol{Y} = \boldsymbol{L}_0^{-1}\boldsymbol{Z} \tag{18}$$

The Nataf transformation, which encompasses these steps, can be succinctly denoted as (19).

$$\boldsymbol{Y} = \boldsymbol{L}_0^{-1}\left[f^{-1}\left(F\left(\boldsymbol{X}\right)\right)\right] \tag{19}$$

This comprehensive process can be inverted, thereby enabling the sampling to produce correlated random vectors with specified distributions.

$$\boldsymbol{X} = F^{-1}\left[f\left(\boldsymbol{L}_0\boldsymbol{Y}\right)\right] = Nataf^{-1}\left(\boldsymbol{Y}\right) \tag{20}$$

2.5 Maximum Entropy Model

Entropy serves as a measure of the information content inherent in a random variable X. The principle that less probable events carry more information makes entropy an effective metric for quantifying uncertainty. A larger entropy value signifies a higher degree of uncertainty, which in turn minimizes the need for subjective assumptions. The self-information refers to the amount of information inherent in a specific outcome of X=*x*, representing the uncertainty that exists prior to the occurrence of event X.

$$I(x) = -\ln P(x) \tag{21}$$

Information entropy is calculated by aggregating the uncertainties present across the entire spectrum of a probability distribution.

$$H(X)=\mathrm{E}_{x\sim P}[I(x)]=-\mathrm{E}_{x\sim P}[\ln P(x)] \tag{22}$$

The principle of the maximum entropy provides a criterion for evaluating probability models.

It asserts that the optimal model is the one that maximizes the entropy while satisfying all the constraints. Thus, we can identify the model with the highest entropy by solving optimization problem in (23).

$$\begin{aligned} \max \quad & H(x)=-\int_{\Omega} p(x)\ln p(x)\mathrm{d}x \\ \text{s.t.} \quad & \int_{\Omega} p(x)\mathrm{d}x=1 \\ & \int_{\Omega}(x-\mu)^{i} p(x)\mathrm{d}x=\mu_i, 1\le i\le n \end{aligned} \tag{23}$$

where $p(x)$ is the PDF of the random variable x; μ_i denotes the i^{th} central moment of x, with n denoting the order of the moments. The information entropy is maximized to obtain a more precise PDF.

The initial value for the optimization using is quite important, as it can significantly affect optimization outcomes and the efficiency of the convergence process. To enhance solution efficiency of MaxEnt in calculating multiple CDFs, a warm-up strategy is proposed.

The center moments $\boldsymbol{\mu}$ of multiple groups of data to be computed are sorted in ascending order based on the following formula.

$$\Delta=\sum_{i=1}^{k}\left(\mu_i-\overline{\mu}_i\right) \tag{24}$$

where k is the size of the central moments , μ_i is the i^{th} central moment and $\overline{\mu}_i$ is the mean of the i^{th} central moment. Then in the optimization process in MaxEnt method, the result obtained based on the former set of central moments can be used as the initial value of the optimization process of the latter set of central moments.

3. Numerical Examples

3.1 Analytical Problem

The analytical problem consists of three systems coupling with each other, as is shown in Figure 2. There are four uncertain source variables for the whole system, of which $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are the inputs for subsystem 1, and $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_4$ are the inputs for subsystem 2. Both of the output of subsystem 1 and 2 are state variables, respectively expressed by $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, which are passed to subsystem 3 as inputs. The final output of the whole system is $\boldsymbol{g}$.

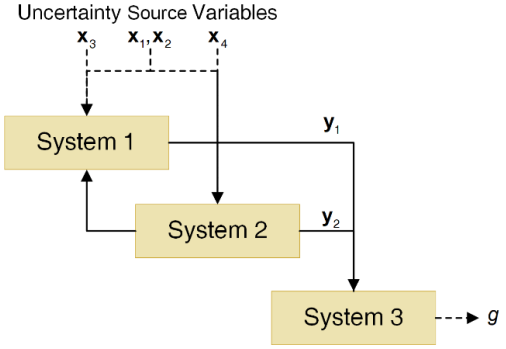

Figure 2 – The uncertainty propagation relation in the analytical problem

The numerical models for the three systems are given as follows.

System 1:

$$\boldsymbol{y}_1 = \boldsymbol{x}_1 a_1^2 + \boldsymbol{x}_2 b_1 + \boldsymbol{x}_3 \tag{25}$$

System 2:

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 a_2^2 + \boldsymbol{x}_1 b_2 + \boldsymbol{x}_4 \tag{26}$$

System 3:

$$g = \int_0^5 \boldsymbol{y}_1 + \boldsymbol{y}_2 dy \tag{27}$$

Where $\boldsymbol{a}$ and $\boldsymbol{b}$ are both system parameters, and $a \in [0,2]$, $b \in [2,4]$. $\boldsymbol{x}_i \left(i = 1,\ldots 4\right)$ is the input variable with the desian space and uncertainty space detailed in Table 1. $\boldsymbol{y}_i$ represents the i^{th} high-dimensional state variable. g is the output variable of subsystem 3.

Table 1 – Design and uncertainty space for analytical problem

Uncertainty variables	Design space		Uncertainty space		
	Upper bound	Lower bound	Distribution	Mean	Standard deviation
$\boldsymbol{x}_1$	1	2	Gaussian	-	0.1
$\boldsymbol{x}_2$	-1	0	Gaussian	-	0.1
$\boldsymbol{x}_3$	0	1	Gaussian	-	0.1
$\boldsymbol{x}_4$	0	1	Gaussian	-	0.1

Table 2 presents the energy proportion respectively of the first one, two and three modes of $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ with the correlation coefficient matrix $\boldsymbol{\rho}$. To ensure the accuracy of the POD reconstruction, the first 2, first 1 and first 3 orders are selected as the casting spaces for $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, and $\boldsymbol{\rho}$.

Table 2 – POD mode energy ratio for analytical problem

Number of eigenvalues / Energy ratio	1st order	2nd order	3rd order
Energy ratio of $\boldsymbol{y}_1$	94.56%	99.98%	100%
Energy ratio of $\boldsymbol{y}_2$	99.85%	100%	100%
Energy ratio of $\boldsymbol{\rho}$	80.18%	98.64%	99.86%

To compare the impact of ignoring the correlation between the hign-dimensional state variables, the original correlated data $\boldsymbol{y}_C = \left[\boldsymbol{y}_{C,1}, \boldsymbol{y}_{C,2}\right]$, the predicted correlated data $\hat{\boldsymbol{y}}_C$, and the predicted uncorrelated data $\hat{\boldsymbol{y}}_U$ are then passed to System 3. The output of System 3 for different input is compared in Figure 3 and Table 3.

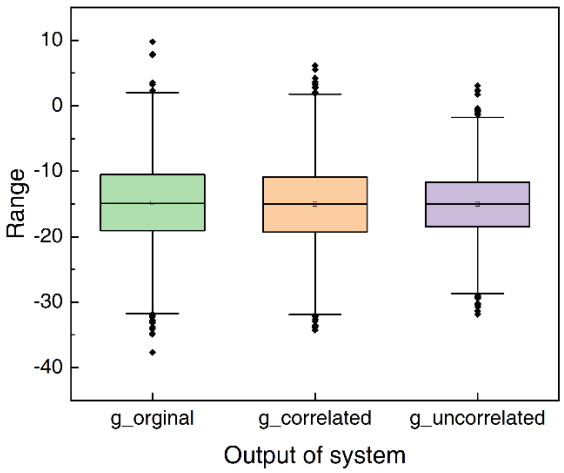

Figure 3 – The output of System 3 for different input in analytical problem

Table 3 – Comparison of the outputs of System 3 in analytical problem

Variables \ Relative error	g_{mean}	g_{std}
Relative error of $\boldsymbol{y}_C$ and $\hat{\boldsymbol{y}}_C$	1.67%	1.64%
Relative error of $\boldsymbol{y}_C$ and $\hat{\boldsymbol{y}}_U$	1.69%	19.75%

In Figure 3, g_original, g_correlated and g_uncorrelated represent the output of System 3 respectively obtained by passing $\boldsymbol{y}_C$, $\hat{\boldsymbol{y}}_C$ and $\hat{\boldsymbol{y}}_U$ to subsystem 3. In Table 3, g_{mean} and g_{std} denote the mean and standard deviation of the ouput g, respectively. It is observed that g_original and g_correlated exhibit similar means and standard deviations, whereas the standard deviation shows a difference of 19.75% between g_uncorrelated and g_original. The method's capacity to accurately propagate uncertainty across various disciplines is thereby validated, highlighting ignoring correlation between hign-dimensional state data can did result in great loss in uncertainty propagation analysis.

3.2 Solid Rocket Problem

The proposed method is applied in quantifying the uncertainty propagation between the engine and the trajectory disciplines of a solid rocket. The zero-dimensional internal ballistic numerical simulation method is employed as the engine discipline model. There are six uncertain input variables of the engine discipline: star side half angle, angle fraction, outer diameter of the charge column, length of the charge column, throat radius, and spread ratio. The uncertain output variables of the engine discipline are the time-varying thrust $\boldsymbol{P}$ and time-varying mass $\boldsymbol{M}$, which are also input variables for the trajectory discipline. The trajectory discipline focuses solely on the motion of the center mass in rocket, and the output is the velocity $\boldsymbol{V}$ and total energy $\boldsymbol{E}$ at the moment of first stage engine separation.

Three first modes are chosen for POD modeling according to the energy proportion analysis by achieving 99.9%. Thus, the high-dimensional state variables $\boldsymbol{P}$ and $\boldsymbol{M}$ can be represented respectively by three parameters. For the correlation coefficient matrix $\boldsymbol{\rho}$ between $\boldsymbol{P}$ and $\boldsymbol{M}$, first eight modes are chosen with an energy proportion of 99.2%.

To compare the impact of ignoring the correlation between the hign-dimensional state variables, the original correlated data $\boldsymbol{y}_C = \left[\boldsymbol{P}_C, \boldsymbol{M}_C\right]$, predicted correlated data $\hat{\boldsymbol{y}}_C$, and predicted uncorrelated data $\hat{\boldsymbol{y}}_U$ are then passed to the trajectory discipline. The entire uncertainty propagation analysis runs

the MaxEnt method 300 times within 2 minutes. The outputs of the trajectory discipline are compared in Table 4,Table 5 and Figure 4.

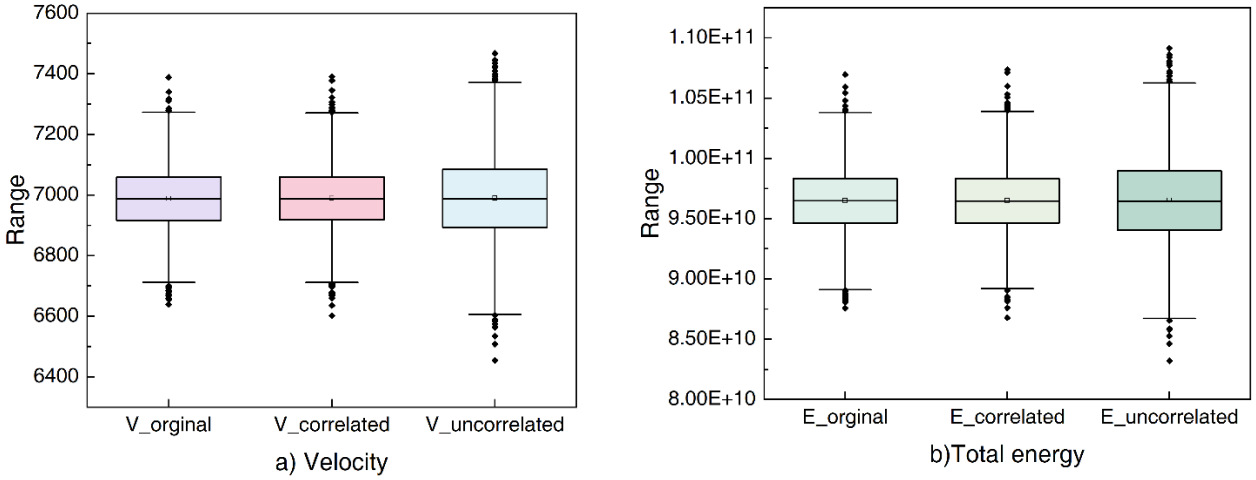

Figure 4 – The output of the whole rocket system for different inputs

Table 4 – Comparison of the outputs of the whole rocket system

Variables \ Relative error	V_{mean}	V_{std}	E_{mean}	E_{std}
Relative error of $\boldsymbol{y}_C$ and $\hat{\boldsymbol{y}}_C$	0.01%	2.29%	0.02%	3.18%
Relative error of $\boldsymbol{y}_C$ and $\hat{\boldsymbol{y}}_U$	0.02%	36.63%	0.05%	35.77%

Table 5 – Time cost computing the CDFs in the whole rocket system

	Time cost(s)
Set the initial coefficient to 0	809
Adopt the warm-up strategy	81

Table 4 and Figure 4 illustrate the uncertainties in $\boldsymbol{V}$ and $\boldsymbol{E}$. After considering the correlation between the high-dimensional state variables, the accuracy of the standard deviation in $\boldsymbol{V}$ is improved by 34.34%, and the accuracy of the standard deviation in $\boldsymbol{E}$ is improved by 32.59%. Thereby the method's capacity to accurately propagate uncertainty across various disciplines is validated.

Additionally, Table 5 compares the time cost for computing the CDFs in the whole rocket system with and without a warm-up strategy. The results indicate that the warm-up strategy significantly improves the calculation efficiency.

4. Conclusion

We proposed an innovative uncertainty propagation analysis method that can consider correlated high-dimensional state variables across various disciplines of a complex system. This method combines the Nataf transformation method with POD reduced order modeling, to effectively enhance the accuracy of uncertainty propagation across multidisciplinary systems. The main conclusions are drawn as follows.

This uncertainty propagation analysis method for correlated high-dimensional data effectively converts uncorrelated high-dimensional data into correlated data, preserving the inherent correlation among state variables. This aspect is often neglected in traditional distributed analysis methods. The method's effectiveness has been validated through applications to both an analytical problem and a solid rocket problem.

During the conversion process, multiple CDFs need to be computed. To increase efficiency, the warm-up strategy strategy is used when computing CDFs with MaxEnt. In the solid rocket problem, the time cost for computing the CDF in the whole rocket system is 81 seconds, which confirms its

practicability in practical application. Despite the large number of CDF required, the warm-up strategy significantly improves calculation efficiency, making the calculation consumption borned.

The data in numerical examples further emphasize the effectiveness of the proposed method. In analytical problems, this method improves the prediction accuracy for the standard deviation of output in System 3 by 18.11%. In solid rocket problem, it enhances the prediction accuracy for the standard deviation of the velocity and the total energy at the separation time of the first stage engine by 6.2% and 23%, respectively. These improvements underscore the importance of considering the correlation of high-dimensional variables in uncertainty propagation analysis.

While the proposed method effectively addresses issues related to the correlation of high-dimensional data, it predominantly focuses on linear correlations between variables. Future research should extend this approach to encompass the complex nonlinear correlations which are often present in aircraft design.

5. Contact Author Email Address

chunnali@nwpu.edu.cn

6. Copyright Statement

References

[1] Thomas J, Dowell E H. Non-Intrusive Polynomial Chaos Approach for Nonlinear Aeroelastic Uncertainty Quantification. *AIAA AVIATION 2022 Forum*, Chicago, 2022.

[2] Han Z. Kriging surrogate model and its application to design optimization: a review of recent progress(in Chinese). *Acta Astronautica et Astronautica Sinica*, Vol. 37, No. 11, pp 3197-3225, 2016.

[3] Jan H U, Uddin M and Abdeljawad T. Numerical study of high order nonlinear dispersive PDEs using different RBF approaches. *Applied Numerical Mathematics*, Vol. 182, No. pp 365-369, 2022.

[4] Qian E, Farcas I G and Willcox K. Reduced operator inference for nonlinear partial differential equations. *SIAM Journal on Scientific Computing*, Vol. 44, No. 4, pp 1934-1959, 2022.

[5] Xie B, Zhang Z and Jiang C. An uncertainty propagation analysis method considering multimodal random distribution(in Chinese). *Scientia Sinica Technologica*, Vol. 52, No. 8, pp 1259-1273, 2022.

[6] Wei X, Li H and Huang C. Application of uncertainty design optimization based on polynomial chaos expansions and maximum entropy method in ship design(in Chinese). *Chinese Journal of Ship Research*, Vol. 18, No. 3, pp 13-25, 2023.

[7] Shi M, Li C and Liu Y. Accurate surrogate modeling method for performance of solid rocket motor under uncertainty(in Chinese). *Journal of Propulsion Technology*, 2024.

[8] Zaman K, Mahadevan S. Robustness-Based Design Optimization of Multidisciplinary System Under Epistemic Uncertainty. *AIAA Journal*, Vol. 51, No. 5, 2013.

[9] Brevault, Loïc and Balesdent. Decoupled Multidisciplinary Design Optimization Formulation for Interdisciplinary Coupling Satisfaction Under Uncertainty. *AIAA Journal*, Vol. 54, No. 1, pp 186-205, 2016.

[10]Qin Z, Liu T and Jiang Z. Distributed Optimization of Nonlinear Uncertain Systems: An Adaptive Backstepping Design. *IFAC-PapersOnLine*, Vol. 53, No. 2, pp 5653-5658, 2020.

[11]An B, Huang B and Zou Y. Distributed optimization for uncertain nonlinear interconnected multi-agent systems. *Systems & Control Letters*, Vol. 168, 2022.

[12]Rosenblatt M. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, Vol. 23, No. 3, pp 470-472, 1952.

[13]Liu P, Kiureghian A D. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, Vol. 1, No. 2, pp 105-112, 1986.

[14]Nelsen R B. *An Introduction to Copulas*. 2st edition, Springer New York, 2006.

[15]Ghosh S, Mavris D N. Methodology for Probabilistic Analysis of Distributed Multidisciplinary Architecture. *Journal of Aircraft*, Vol. 58, No. 2, pp 1-13, 2020.

[16]Jia X, Gong C, Ji W and Li C. Flow sensing method for fluid-structure interaction systems via multilayer proper orthogonal decomposition. *Journal of Fluids and Structures*, Vol.124, p.104023, 2024.

[17]Li C, Jia X and Gong C. Airfoil inverse design method based on two-layer POD and BPNN. *Advances in Aeronautical Science and Engineering*, Vol. 12, No. 2, pp 30-37, 2021.

[18]Phil K. *Deep learning for beginners:with MATLAB examples*. Zou W, Wang Z and Wang Y, Translation. 1st edition, Beihang University Press, 2018.

[19]Kiureghian A D, Liu P L. Structural Reliability under Incomplete Probability Information. *Journal of Engineering Mechanics*, Vol. 112, No. 1, pp 85-104, 1986.

[20]Liu P L, Kiureghian A D. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, Vol. 1, No. 2, pp 105-112, 1986.