UNCERTAINTY QUANTIFICATION WITH MISSING VALUES

David Aguilar

SmartUQ, USA

Contact@SmartUQ.com

KEYWORDS

Uncertainty quantification, polynomial chaos expansion, sparse grid, experimental design, data augmentation, Gaussian, Kriging, kernel, emulation, surrogate model, meta-model, sensitivity analysis, propagation of uncertainty.

ABSTRACT

Uncertainty quantification (UQ) has emerged as the science of quantitative characterization and reduction of uncertainties in simulation and testing. Stretching across applied mathematics, statistics, and engineering, UQ is a multidisciplinary field with broad applications. Popular UQ methods include generalized Polynomial Chaos Expansion (gPCE), Gaussian processes analysis, sensitivity analysis, and propagation of uncertainty. Many of these methods were developed using applied mathematics and do not require knowledge of a simulation's physics. Thus they may be used across disparate industries and are applicable to both individual component and system level simulations. Because of their ability to bring greater degrees of confidence to simulation based decisions, UQ methods are playing an ever larger role in design and engineering of high precision turbomachinery.

A short coming of many existing UQ methods is that they assume simulation results can be collected precisely on a set of carefully-chosen input configurations. These methods encounter problems when any of the input configurations fail to produce valid simulation results. For example, gPCE requires that results be collected on a sparse grid Design of Experiment (DOE), which is generated based on probability distributions of the input variables. A failure to run the simulation at any one input configuration can result in a large decrease in the accuracy of a gPCE. The decrease can necessitate the generation and simulation of an entirely new sparse grid input configuration. In practice, simulation data sets with missing values are common because simulations regularly yield invalid results due to physical restrictions or numerical instability. Such failures render existing UQ methods unreliable. Evaluation of new input configurations requires a great deal of time, risks further failures and wastes the original computational effort.

We propose a statistical framework to mitigate the cost of missing values. This framework decreases the additional simulation runs necessary to yield accurate UQ

results in the event that simulation failure makes gPCE methods unreliable. By careful data augmentation, the proposed framework is able to recycle existing simulation results when constructing a space filling DOE. The new simulation results from this space filling design can be used to fit robust statistical emulators. The processed of choosing new points and fitting emulators has minimal computational overhead even for large data sets. Once fitted, the emulators may be used to perform a variety of UQ tasks in place of gPCE methods. These emulators may also be used for additional analytics tasks such as model calibration, design space exploration, and optimization.

Several examples are used to demonstrate the proposed frame work and its utility across a number of problem types.