Weighted space-filling designs for dependent variables with application to deterministic computer codes

M. Feddag*, D. Woods* and V. Rapley**
*University of Southampton, UK
**Dstl, UK

The simulation of physical and engineering systems via complex mathematical models has become a common method of gaining knowledge about mechanisms where physical experimentation would be time consuming, costly or hazardous. In order to explore and understand these often computationally expensive computer models, experiments are often performed where the treatments are combinations of values of the input variables and the responses are deterministic outputs from the computer code. The designs used are often chosen to cover, or space-fill, the design region of possible input values. Many of these space-filling designs, such as computer-generated coverage designs based on Euclidean distance, do not take into account of dependencies between variables which often exist in applications. Further, these dependencies may result in substantial areas of the design region which are of little relevance to the application, for example, where it is known that no response can occur. Standard space-filling techniques can therefore result in highly inefficient designs, in terms of the numbers of points in interesting regions of the design space.

This work is motivated by dispersion models, which typically have the following features:

(i) the input variables are usually of two types (meteorological and source), and can be quantitative or qualitative
(ii) there is substantial prior information about the distribution of the input variables from, for example, empirical observations (meteorological) or expert prior knowledge (source)
(iii) these distributions are not usually independent, either within type (for example, wind direction and speed is defined via a wind rose) or between type (wind

direction and source location)
(iv) the distributions define a joint probability density (or weight function) on the design region, which is likely to have substantial areas of low weight.

Although dispersion models are generally quick to evaluate, when used routinely in, for example, sensor placement algorithms or as part of other optimisation functions, there is a need to reduce the number of code evaluations through carefully designed experiments.

Methods of incorporating known dependencies between variables into design selection are investigated. An adaptation of computer-generated coverage designs is considered, which accounts for dependencies between variables through redefining “distance” between two points, $x1$ and $x2$, to include a weight function, $w(x1 , x2)$. The weight function is based upon a specified joint distribution for the k input variables, chosen to reflect application specific dependencies. This method can include quantitative and qualitative (ordered and unordered categorical) variables, and different types of prior information. We compare this approach with the use of Latin Hypercube designs (for quantitative variables), where the design points are chosen so that the values of each input variable will approximate the marginal distribution of that variable.

The different approaches are demonstrated through interrogation of a computer model for atmospheric dispersion. Both the weighted space-filling and Latin Hypercube designs are critically compared with computationally expensive Monte Carlo techniques.