An S4 Object structure for emulation - the approximation of complex functions

Rachel Oxlade[1], **Peter Craig**[1]

1. Durham University, UK
*Contact author: r.h.oxlade@durham.ac.uk

Keywords: Bayesian, computer model, emulation, nested models, S4 objects

We aim to facilitate the *R* user who has data from a complex function which is costly to evaluate directly (for example a deterministic computer model, see for example Craig et al. (1997), Kennedy and O'Hagan (2001)). We approximate the function by a gaussian process, and thereby produce a posterior distribution for the function's behaviour throughout its domain. Where the function's value is already known, that is at the input points belonging to the data mentioned above, the posterior will predict the correct value with certainty. Elsewhere, the distribution will give an approximation which represents our uncertainty. Typically the domain of the function may cover twenty or more dimensions, and there may be several thousand data points. An object oriented structure is presented for dealing with this problem, enabling the user to interact with the posterior distribution of the function across the domain. This is similar in its goals to the package **BACCO**, but with a different approach.

A particular scenario of interest to us is where we are given a second model (or function), whose inputs extend those of the first. It may contain an additional process, and it is useful for us to be able to compare the two models, to assess the value of including this new process, and to see how the two functions differ across the input domain. We deal with this problem by creating a hierarchical structure, which we again use S4 objects to capture. Joint emulation of models with differing input spaces has, to our knowledge, not yet been addressed, and so this approach is a novel one. The object oriented structure will be presented, along with some methods for dealing with such objects.

In constructing these emulators there are many choices to make, for example whether or not to include a regression surface to capture large scale variation, and if so, how it is to be formed. Incorporated into the *R* code described above are functions which automate such choices given the user's stipulations. This means that, given some initial data showing the function's behaviour and some directions as to how *R* is to proceed at certain points, the functions' behaviour at any set of previously unexplored points can be predicted, and a measure of uncertainty given. The S4 object structure allows the user to interact with the objects, and in particular provides methods to adjust them by making alterations to the original choices made.

References

Craig, P. S., M. Goldstein, A. H. Seheult, and J. A. Smith (1997). Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments. *Case Studies in Bayesian Statistics 3*, 37–93.

Kennedy, M. C. and A. O'Hagan (2001). Bayesian Calibration of computer models. *Journal Of The Royal Statistical Society Series B 63*, 425–464.