The DiceKriging package: kriging-based metamodeling and optimization for computer experiments

Olivier Roustant[1], David Ginsbourger[2], Yves Deville[3]

1. Ecole des Mines de St-Etienne (France). Contact author: roustant@emse.fr
2. Université de Neuchâtel (Switzerland)
3. Statistical Consultant (France)

Keywords: Computer Experiments, Gaussian Processes, Spatial Statistics, Kriging, Global Optimization

The package DiceKriging has been developed for analyses involving computer intensive experiments as met in various industrial contexts (automotive, aeronautics, nuclear, ...) where numerical simulations are required. Kriging stands for the well-known model from spatial statistics, which has recently gained popularity in the computer experiments community, and is sometimes referred to as Gaussian Process Regression (Rasmussen, Williams, 2006). DICE (*Deep Inside Computer Experiments*) is the name of a consortium within the frame of which the works were conducted, joining the industrial partners Armines, Renault, EDF, IRSN, Onera and Total S.A (`www.dice-consortium.fr`). DICE members have shared a growing interest for R, due to its efficiency in including the most recent statistical methods, and its ease of use. The development was both guided by applications and based on published efficient algorithms. DiceKriging was tested on toy & industrial case studies in dimensions 2 to more than 30. It was found to be a valuable complement to other existing packages like *tgp* or *mlegp*. The package contents is the following:

1. Kriging models for deterministic simulators, stochastic simulators with unknown homogenous noise and stochastic simulators with controllable heteroscedastic noise:
   - Maximum Likelihood (ML) estimation of unknown parameters, with possible penalization.
   - Cross Validation (Leave-One-Out, k-fold CV),
   - Prediction: Simple, Ordinary, and Universal Krigings with many kinds of trends,
   - Covariance kernels: Anisotropic Gaussian, Power Exp., and Matérn with $\nu = 3/2$ or $\nu = 5/2$.
2. Kriging-based black-box optimization:
   - Efficient Global Optimization (EGO) algorithm,
   - Several parallelized versions of the EGO algorithm for synchronous distributed computing.

Among the innovations proposed in the package, much effort was devoted to the efficient use of optimization routines, both for the problem of ML estimation and within black-box optimization algorithms.

- To address the known difficulties in ML estimation (multimodality and possible numerical instability) the algorithm proposed by Park and Baek (2001) has been implemented. This allows the use of analytic gradient in the optimisation either with the classical BFGS (optim{stats}) or with the hybrid evolutionary genoud{rgenoud}. The available covariance structures can include a "nugget effect" for stochastic simulators and their list can be extended in the future.
- The EGO algorithm was proposed by (Jones, Schonlau, Welch, 1998). Additionally to EGO, a class of variants for synchronous parallel computing is proposed. In both cases, the maximization of the highly multimodal Expected Improvement is performed using the hybrid evolutionary strategy *genoud*.
- The package also includes Kriging with known parameters, which can be useful for Bayesian Kriging.

The package documentation includes some validation tests (as ML estimation of simulated processes), classical analytical test functions, and several case studies proposed by the members of the DICE Consortium.

References

Jones D.R., Schonlau M. and Welch W.J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global optimization*, 13, 455–492

Park J-S, Baek J. (2001). Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computer Geosciences*, 27, 1–7.

Rasmussen C.E., Williams C.K.I. (2006). *Gaussian Processes for Machine Learning*, the MIT Press. `www.GaussianProcess.org/gpml`.