1 The emulator

We treat the output of the simulator y as an uncertain function $f()$ of the simulator inputs x, so that $y = f(x)$. We wish to produce a predictive distribution for y at any model input, conditional on the points already run, or the design (Y, X). Throughout the study, we use a kriging function, similar to a Gaussian process regression emulator, as coded in the package DiceKriging (Roustant et al., 2012) in the statistical programming environment R (R Core Team, 2016), for prediction of climate simulator output at untried inputs. The kriging model or Gaussian Process regression is specified hierarchically with a separate mean and covariance function. For prediction purposes, *a priori* assume that the trend is a simple linear function of the inputs, and adjust with a Gaussian process.

$$f(x) = h(x)^T\beta + Z(x)$$

Where $h(x)^T\beta$ is the mean function, and the residual process Z is a zero mean stationary Gaussian process. The covariance kernel c of Z

$$Cov(Z, Z') = \sigma^2 c(x, x')$$

can be specified in a number of different ways: we use the default diceKriging option of a Matern $v = 5/2$ function so that

$$c(x, x') = (1 + \frac{\sqrt{5}|x - x'|}{\theta} + \frac{5|x - x'|^2}{3\theta^2})exp(-\frac{\sqrt{5}|x - x'|}{\theta})$$

where θ describes the *characteristic length scales* - a measure of how quickly information about the function is lost moving away from a design point, in any dimension. This and other hyperparameters are estimated via maximum likelihood estimation from the design (Y, X), meaning that the approach is not fully Bayesian (such an approach would find posterior distributions for the hyperparameters rather than point estimates). We use Universal Kriging, with no 'nugget' term, meaning that the uncertainty on model outputs shrinks to zero at the design points.

Full details of the Universal kriging process used can be found in (Roustant et al., 2012), section 2.1, details of the kernel can be found in section 2.3, and examples of the trend and hyperparameter estimation in section 3 the same publication.

2 Forest regions

Forest fraction data is taken by calculating the mean broadleaf forest fraction in the areas shown in figure 1. Mean temperature and precipitation from the model are calculated for the corresponding regions and time period. The regions are: Amazon 15°S - 15°N, 270°E - 315°E; Central Africa; 15°S - 10°N, 7.5°E - 30°E; SE Asia 12°S - 10°N, 90°E - 150°E.

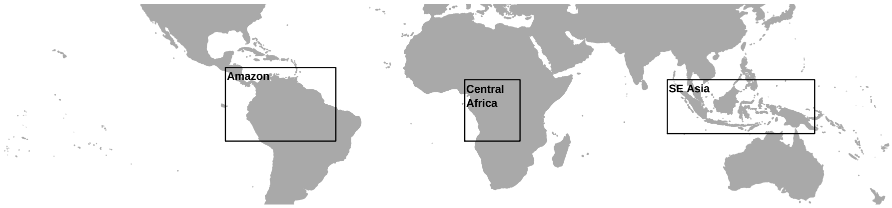

Figure 1. A map of the forest regions used in the study.

References

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2016.

Roustant, O., Ginsbourger, D., and Deville, Y.: DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization, Journal of Statistical Software, 51, 1–55, http://dx.doi.org/10.18637/jss.v051.i01, 2012.