Geophysical Research Abstracts
Vol. 15, EGU2013-10337, 2013
EGU General Assembly 2013

Large scale stochastic spatio-temporal modelling with PCRaster

Derek Karssenberg (1), Niels Drost (2), Oliver Schmitz (1), Kor de Jong (1), and Marc F.P. Bierkens (1)

(1) Faculty of Geosciences, Utrecht University, Department of Physical Geography, Utrecht, Netherlands (d.karssenberg@uu.nl), (2) Netherlands eScience Center, Science Park 140 (Matrix I) 1098 XG Amsterdam, The Netherlands

PCRaster is a software framework for building spatio-temporal models of land surface processes (http://www.pcraster.eu). Building blocks of models are spatial operations on raster maps, including a large suite of operations for water and sediment routing. These operations are available to model builders as Python functions. The software comes with Python framework classes providing control flow for spatio-temporal modelling, Monte Carlo simulation, and data assimilation (Ensemble Kalman Filter and Particle Filter). Models are built by combining the spatial operations in these framework classes. This approach enables modellers without specialist programming experience to construct large, rather complicated models, as many technical details of modelling (e.g., data storage, solving spatial operations, data assimilation algorithms) are taken care of by the PCRaster toolbox. Exploratory modelling is supported by routines for prompt, interactive visualisation of stochastic spatio-temporal data generated by the models.

The high computational requirements for stochastic spatio-temporal modelling, and an increasing demand to run models over large areas at high resolution, e.g. in global hydrological modelling, require an optimal use of available, heterogeneous computing resources by the modelling framework. Current work in the context of the eWaterCycle project is on a parallel implementation of the modelling engine, capable of running on a high-performance computing infrastructure such as clusters and supercomputers. Model runs will be distributed over multiple compute nodes and multiple processors (GPUs and CPUs). Parallelization will be done by parallel execution of Monte Carlo realizations and sub regions of the modelling domain. In our approach we use multiple levels of parallelism, improving scalability considerably. On the node level we will use OpenCL, the industry standard for low-level high performance computing kernels. To combine multiple nodes we will use software from the eScience Technology Platform (eSTeP), developed at the Netherlands eScience Center. This will allow us to scale up to hundreds of machines, with thousands of compute cores. A key requirement is not to change the user experience of the software. PCRaster operations and the use of the Python framework classes should work in a similar manner on machines ranging from a laptop to a supercomputer. This enables a seamless transfer of models from small machines, where model development is done, to large machines used for large-scale model runs.

Domain specialists from a large range of disciplines, including hydrology, ecology, sedimentology, and land use change studies, currently use the PCRaster Python software within research projects. Applications include global scale hydrological modelling and error propagation in large-scale land use change models. The software runs on MS Windows, Linux operating systems, and OS X.