Large scale stochastic spatio-temporal modelling with PCRaster Python

Derek Karssenberg[1], Niels Drost[2], Oliver Schmitz[1], Kor de Jong[1], Marc Bierkens[1], [1]Faculty of Geosciences, Utrecht University, [2]Netherlands eScience Center, the Netherlands, info@pcraster.eu

Stochastic spatio-temporal modelling

Model

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{i}_t, \mathbf{p}_t) \quad \text{for all time steps } t = 1, 2, ..., T$$

state variables — *transition function* — *inputs* — *parameters*

Solution scheme

for each n in Monte Carlo samples:

for each t in time steps:

$$\mathbf{z}_t^{(n)} = f(\mathbf{z}_{t-1}^{(n)}, \mathbf{i}_t^{(n)}, \mathbf{p}_t^{(n)})$$

Building blocks

```
discharge = kinematic(flowDir,precipitation,..)
```

result map — *spatial function* — *input maps*

Building blocks to construct the transition function are functions on spatial data types (raster maps). Functions were developed in C++ and are available as Python functions (Python extension).

Solution framework (Python)

```
from PCRaster import *
from PCRaster.Framework import *

class SnowModel(DynamicModel, MonteCarloModel):
  def __init__(self):
    ...
  def premcloop(self):
    dem = self.readmap('dem')
    self.ldd = lddcreate(dem, ...)
    ...
  def initial(self):
    self.snow = scalar(0)
    ...
  def dynamic(self):
    runoff = accuflux(self.ldd, rain)
    self.report(runoff, 'q')
    ...
  def postmcloop(self):
    mcpercentiles('q',percentiles,..)
```

sets constant variables and parameters (premcloop)

is run at t = 0 for each Monte Carlo sample (initial)

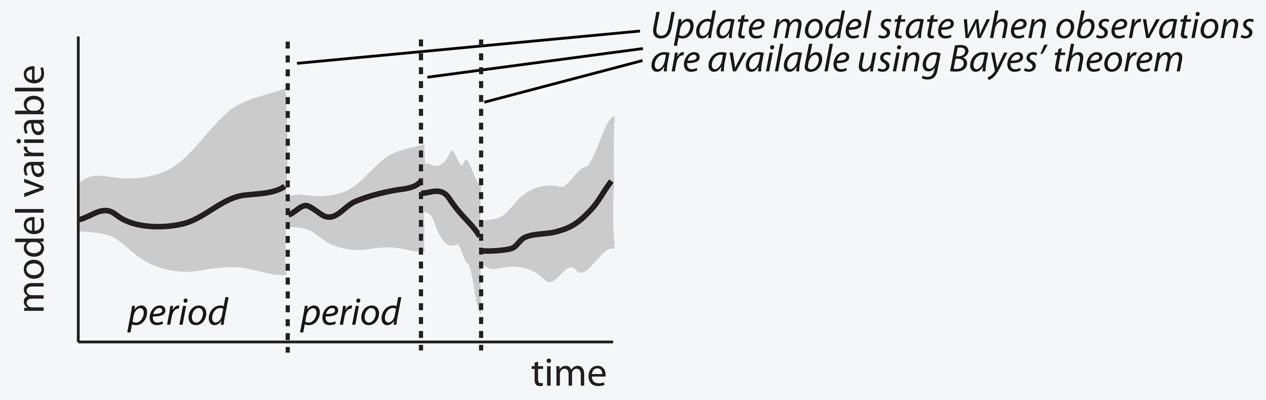
is run for each Monte Carlo sample and for each time step (dynamic)

is run at end calculating sampling statistics over Monte Carlo samples (postmcloop)

PCRaster

- Is targeted at the development of spatio-temporal models
- Fast model development and execution
- Scripting environments: PCRcalc and Python
- Rich set of model building blocks for manipulating raster maps
- Framework for stochastic spatio-temporal model building
- Framework for data assimilation
- Tool for visualisation of spatio-temporal stochastic data
- Runs on Linux, Microsoft Windows and Apple OS X
- Can be downloaded for free and is soon open source

Data assimilation



Solution scheme

for each period in periods:

for each n in Monte Carlo Samples:

for each t in period:

$$\mathbf{z}_t^{(n)} = f(\mathbf{z}_{t-1}^{(n)}, \mathbf{i}_t^{(n)}, \mathbf{p}_t^{(n)})$$

evaluate Bayes' theorem

Solution framework (Python)

```
def suspend(self):
  self.report(self.snow, 's')
  ...
def updateWeight(self):
  sum = exp(maptotal(((obs - mod)**2)/
       (2.0 * (observedStd ** 2))))
  weight = exp(sum)
  return weight
  ...
def resume(self):
  self.read('s')
  ...
```

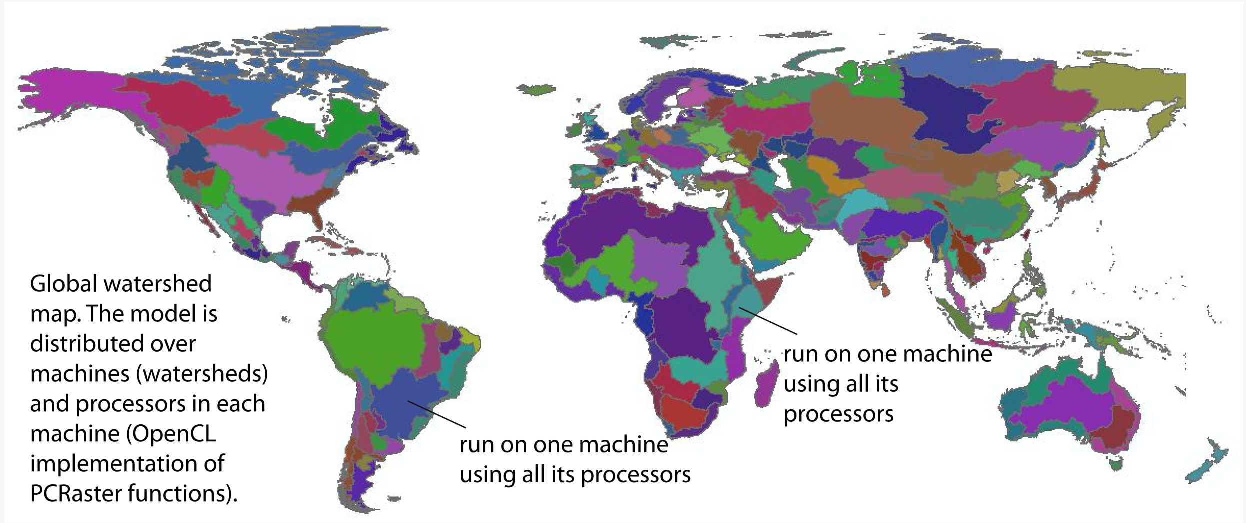
store model state at end of period (suspend)

calculate weight of Monte Carlo sample required for solution of Bayes' equation and return to framework (updateWeight)

read model state at start of next period (resume)

PCRaster on supercomputers (parallel execution)

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.



Global model runs are distributed over multiple compute nodes (using eScience Technology Platform eSTeP), where each node models one watershed. Each watershed is modelled by using all processors in the node (GPUs and CPUs), which is enabled by an OpenCL implementation of PCRaster functions. This will allow us to scale up to hundreds of machines, with thousands of compute cores.

PCRaster at EGU (selection)

Poster R187, EGU2013-11126 (Tuesday), Alberti et al. A web-application for visualizing uncertainty in numerical ensemble models.

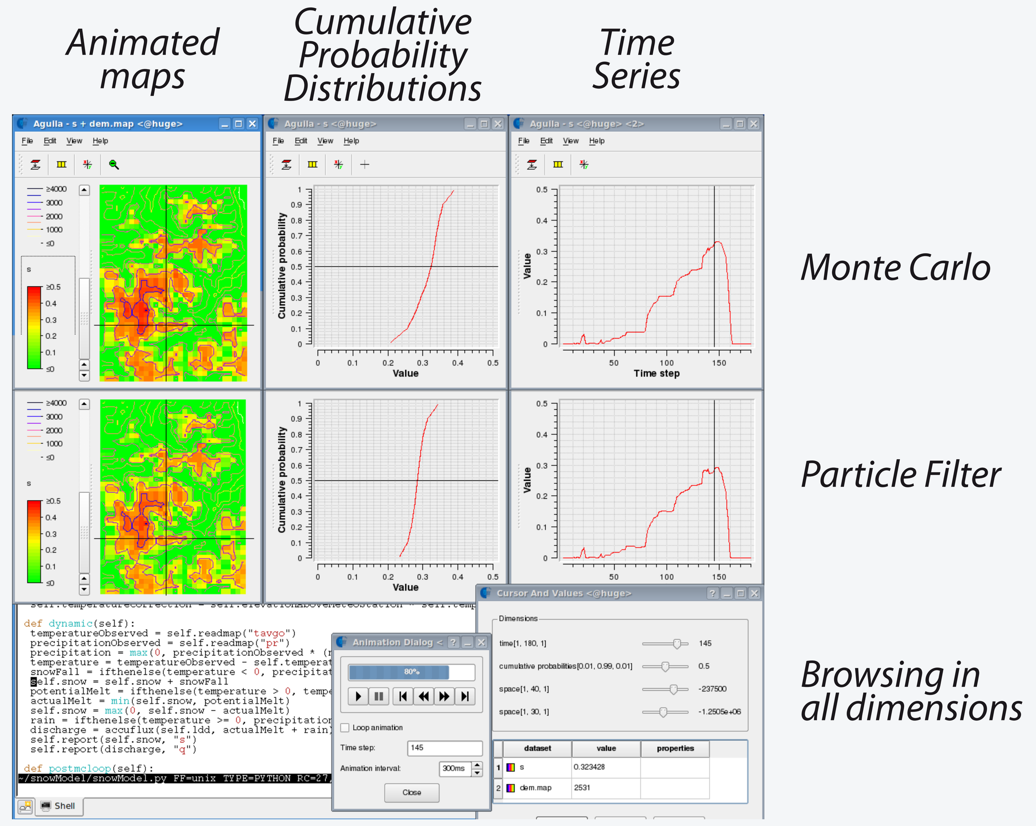
Poster R215, EGU2013-3337 (Wednesday), Sutanudjaja et al., eWaterCycle: Developing a hyper resolution global hydrological model.

Poster R373, EGU 2013-10215 (Thursday), Wanders et al., The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models.

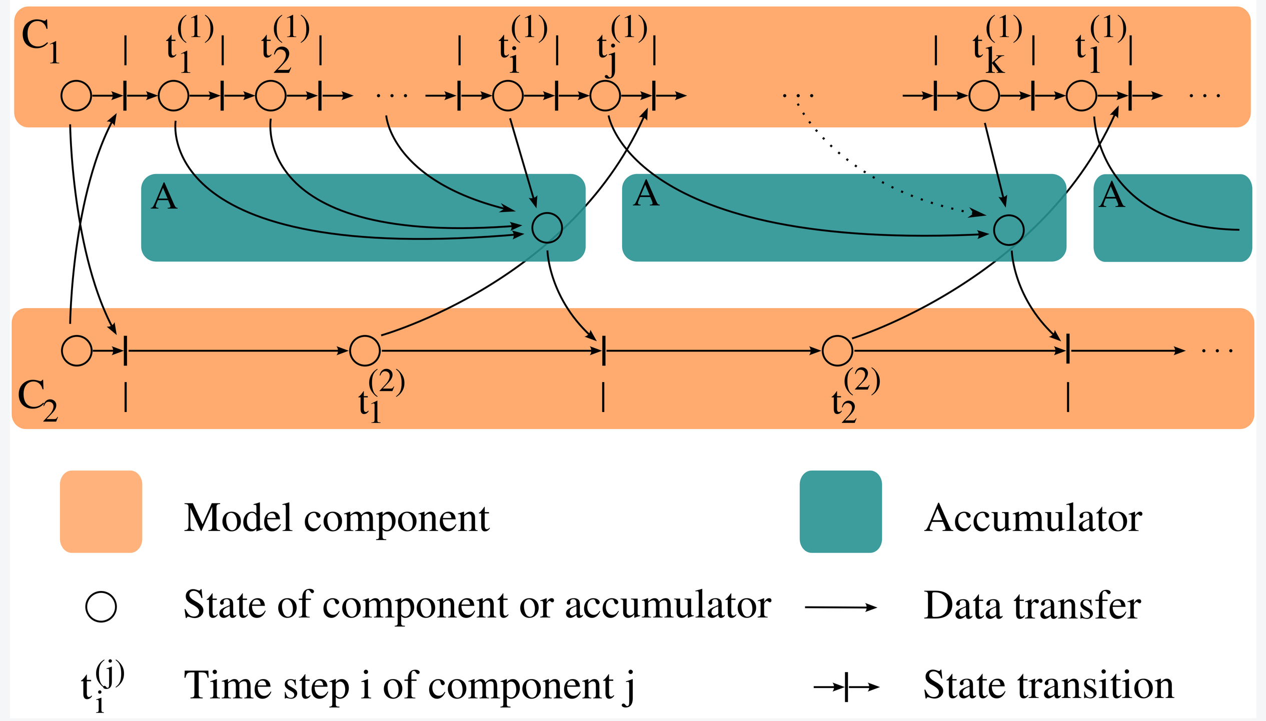
Poster R293, EGU2013-10355 (Friday), Straatsma et al., Water2Invest: Global facility for calculating investments needed to bridge the climate-induced water gap.

Oral Room B6, 14.15 h (Friday), Bernhard et al., Consequences of secondary succession on water availability in Mediterranean areas: a study case in northeastern Spain.

Visualisation of stochastic spatio-temporal data



Current work: integrated modelling



Temporal control flow between model components with shorter (C1) and longer (C2) time steps. Each model component requests output from the other component. C1 directly accepts the input of C2 , C2 expects aggregated values from C1, provided by the accumulator A.

Information and download at:
http://www.pcraster.eu

References

Karssenberg, D., Schmitz, O., Salamon, P., De Jong, K. and Bierkens, M.F.P., 2010, A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. Environmental Modelling & Software, 25, pp. 489-502.