Geostatistical analysis for Uncertainty Quantification in the SMART-SED model: a downscaling approach based on Digital Soil Mapping data

Federico Gatti

joint work with

Alessandra Menafoglio, Niccolò Togni, Luca Bonaventura, Davide Brambilla, Monica Papini, Laura Longoni

MOX-Department of Mathematics, Politecnico di Milano

Environmental-Civil Engineering Department, Politecnico di Milano

EGU 2020

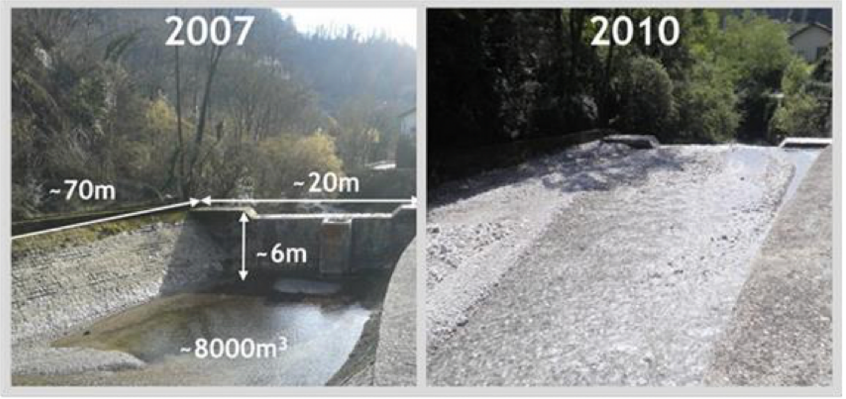
06 May, 2020 – Vienna, Austria

Funding Agencies: **Fondazione Cariplo**, SMART-SED Project grant number N. 2017-0722

Motivations: SMART-SED project

Objective: SMART-SED, i.e. Sustainable Management of sediment transpoRT in reposnSE to climate change conDitions. Numerical tool able to help smart-cities to assess hydro-geological risk. Need to model hydro-geological quantities e.g. infiltration

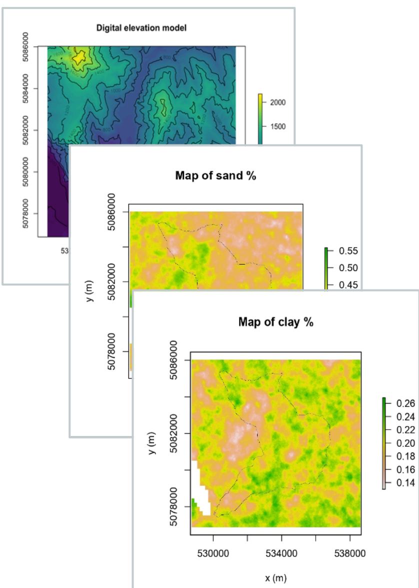
Case study: Caldone basin, Lecco northern Italy.



Modeling framework

Input parameters

Digital Terrain Model (DTM), with soil compositions



Model equations

De Saint Venant equations coupled with gravitational layer, sediment transport and snow equations

$$\partial_t H + \nabla \cdot (H\, u) = (1-\mu)\, p - f,$$

$$\partial_t u + g\nabla\eta + u \cdot \nabla u + \gamma(u)\, u = 0.$$

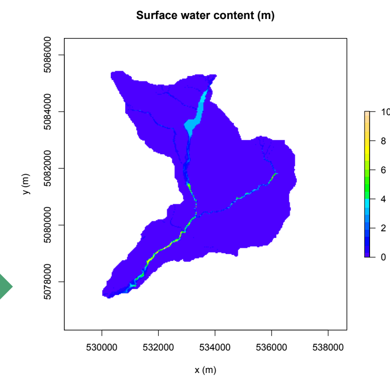
$$\partial_t h_g + \nabla \cdot f_g = f - ev + s\,,$$

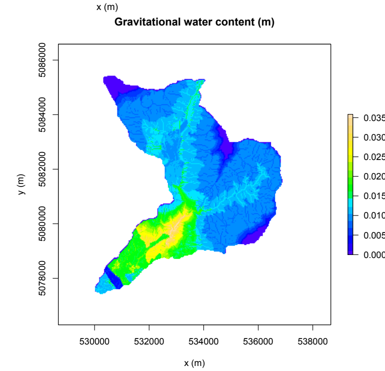
$$\partial_t h_{sd} + \nabla \cdot f_{sd} = w\,,$$

$$\partial_t h_{sn} = \mu\, p - s,$$

Simulation

State variable maps and temporal sequences

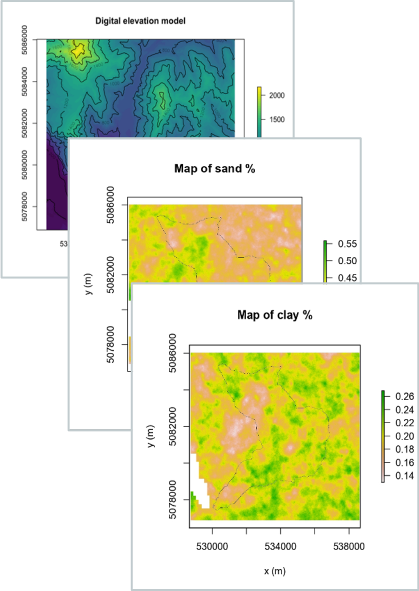




Modeling framework

Input parameters

Digital Terrain Model (DTM), with soil compositions



Model equations

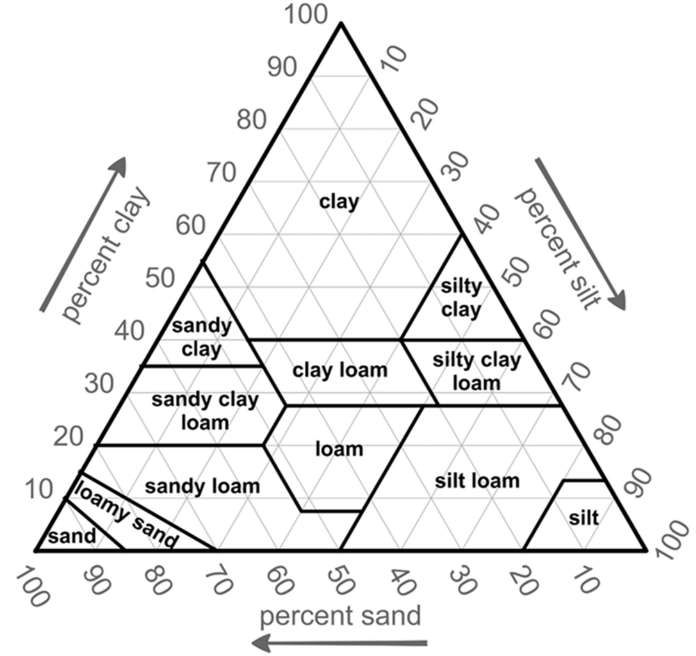
Simulation

Most of input variables, e.g. maximum soil moisture retention, hydraulic conductivity depends on terrain compositions

Soil texture: Particle Size Fractions (psf)

- In USDA classification soil texture is determined by relative percentage of "fine" length-scale L, i.e. less than $2\ mm$:
  - Clay: $L < 2\ \mu m$
  - Silt: $2\ \mu m < L < 50\ \mu m$
  - Sand: $50\ \mu m < L < 2\ mm$

- 12 classification in the soil texture triangle:



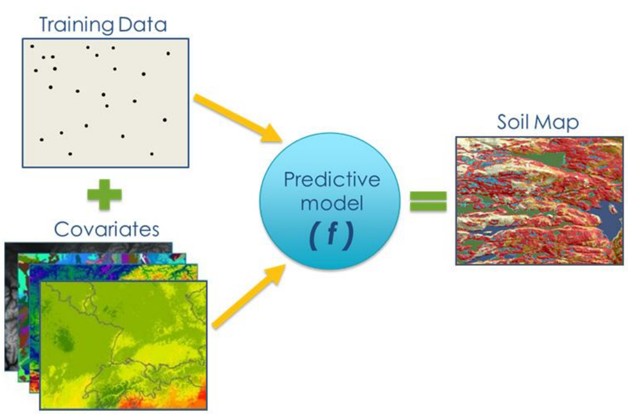
- Psf data are compositional, i.e.:
  - Clay + Silt + Sand$=1$ → constant sum
  - Clay, Silt, Sand ≥ 0 → semi-positivity

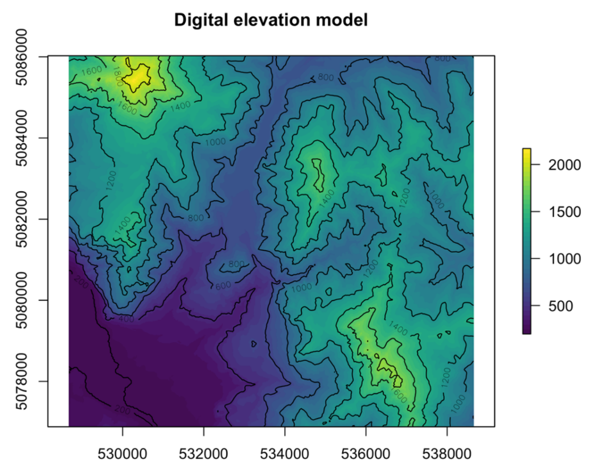
SoilGrids: digital soil maps

Need to have a psf map at a resolution given by DTM:

- Absense of field measurements
- Presence of an online open access repository SoilGrids:
  - Raster psf at $250\ m$ resolution
  - Result of training data, remote-sensing and machine learning algorithms

[Hengl, T., Mendes de Jesus, J., Heuvelink, G., Ruiperez Gonzalez, M., and Kilibarda, M. e. a. (2017). Soilgrids250m: Global gridded soil information based on machine learning. PLOS ONE, 12(2):1–40]

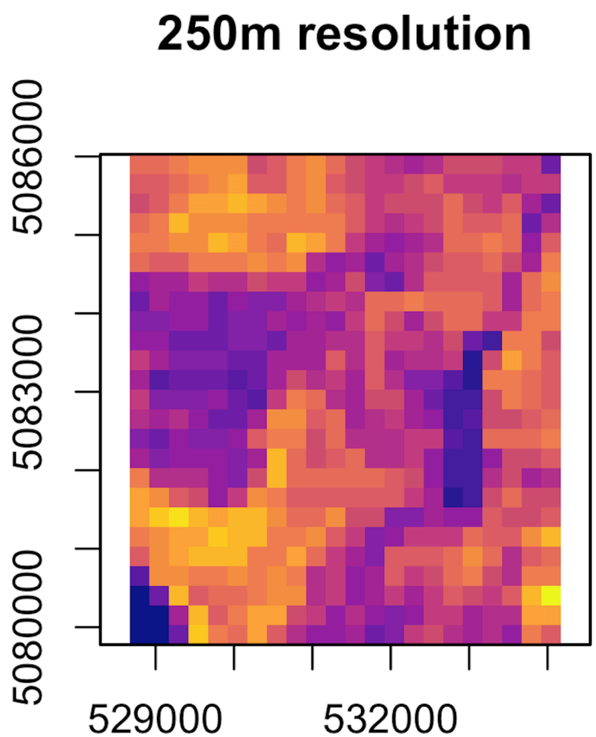


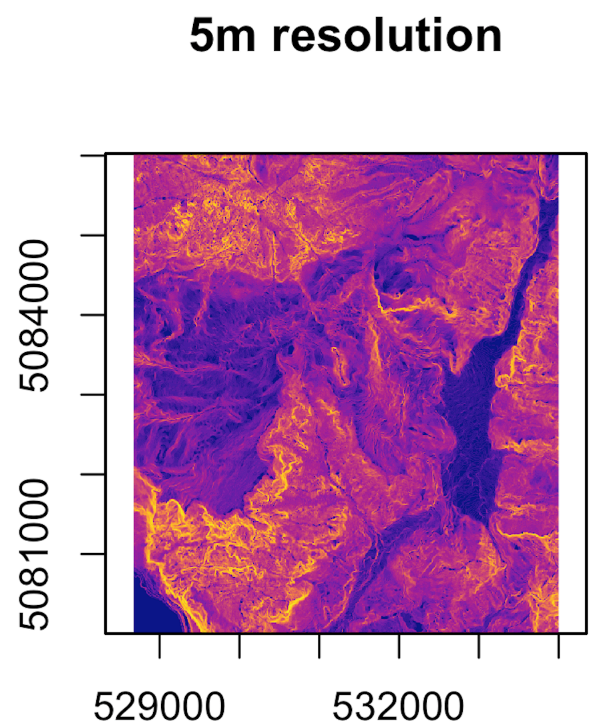


Downscaling: change of support

Our aim is to downscale SoilGrids raster data to DTM resolution, from the literature we have **Area-To-Point Regression Kriging** but it works only in an **Euclidean framework**,

→ need an extension to account for **constraints** on the variable to be downscaled.





Downscaling: change of support

From literature we have **Area-To-Point Regression Kriging (ATPRK)**, given a spatial scalar field $Z(\mathbf{x}), \mathbf{x} \in D$, its fine scale prediction in a given k-th pixel, reads:

$$\widehat{Z}_k = \sum_l \beta^l \, u_k^l + \sum_{\overline{K}} \lambda^{\overline{K}} \, e_{\overline{K}}$$

It is a linear combination of two methods:

Linear Regression: first summation. Where $\beta^l \in \mathbb{R}$ are the regression coefficients and $u_k^l \in \mathbb{R}$ are some known fine scale covariates, e.g. DTM or some related functions.

$$\mathbb{E}[Z_k] = \mathbb{E}[\widehat{Z}_k] = \sum_l \beta^l \, u_k^l$$

Area-To-Point Kriging (ATPK): second summation. The fine residuals are linear combination of coarse residuals calculated after linear regression in the neighbourhood of the fine scale pixel k to be predicted. We assume second order stationarity on the signal.

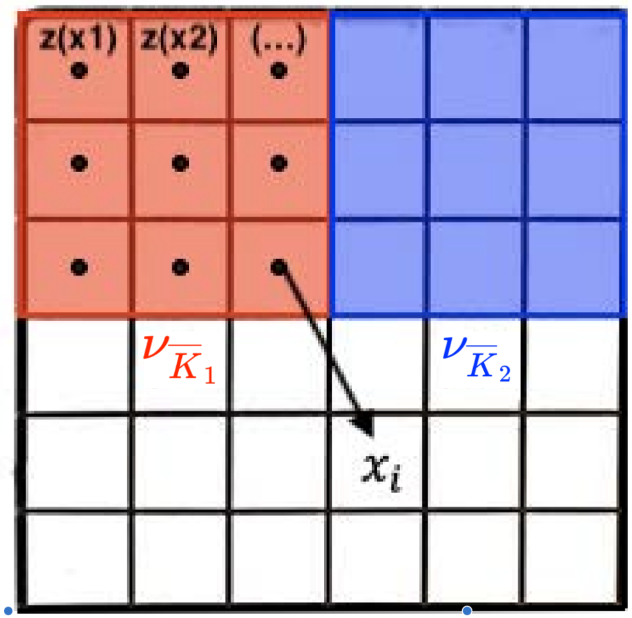
$$\lambda^{\overline{K}} \in \mathbb{R} \qquad e_k = \sum_{\overline{K}} \lambda^{\overline{K}} \, e_{\overline{K}} \quad \text{s.t.} \quad \sum_{\overline{K}} \lambda^{\overline{K}} = 1$$

Downscaling: change of support

Via minimization of the variance of the fine predicted residuals it is possible to determine the ATPK weights via solving the following linear system:

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \{\lambda^{\overline{K}}\}_{\overline{K}} \\ \mu \end{bmatrix} = \begin{bmatrix} \{\sigma^{\overline{K}}\}_{\overline{K}} \\ 1 \end{bmatrix}$$

[Kyriakidis, P. (2004). A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis, 36.]



[Goovaerts, P. (2008). Kriging and semivariogram deconvolution in presence of irregular geographical units. Mathematical geology, 40:101–128.]

- Block-Block covariance:

$$\Sigma_{\overline{K}_1,\overline{K}_2} = \frac{1}{P^2}\sum_{i=1}^{P}\sum_{j=1}^{P} C(\|\mathbf{x}_j - \mathbf{x}_i\|),$$
$$\mathbf{x}_i \in \nu_{\overline{K}_1}, \mathbf{x}_j \in \nu_{\overline{K}_2}.$$

- Point-Block covariance:

$$\sigma^{\overline{K}} = \frac{1}{P}\sum_{i=1}^{P} C(\|\mathbf{x}_k - \mathbf{x}_i\|), \quad \mathbf{x}_i \in \nu_{\overline{K}}.$$

- P: upscaling factor, i.e. ratio between coarse and fine pixel measure support

- $C(\cdot)$ covariance of the fine scale residuals unknown, cannot be directly estimated from coarse scale data

→ need for variogram deconvolution

Downscaling of compositional data

Psf data $\{Z_i\}_{i=1}^{n}$ are **compositional,** they live in a subset of $\mathbb{R}^n$, the simplex $\mathbb{S}^n$:

$$\mathbb{S}^n = \{(Z_1, \ldots, Z_n : Z_i \geq 0, \sum_{i=1}^{n} Z_i = 1\}$$

Aitchison (1986) first defined a set of operations in the simplex that define a Hilbert space::

- Closure: $\mathcal{C}(\mathbf{x}) = \left(\frac{x_1}{\sum_{i=1}^{n} x_i}, \ldots, \frac{x_n}{\sum_{i=1}^{n} x_i} \right) \quad \forall \mathbf{x} \in \mathbb{R}_+^n$
- Sum: $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left(x_1 y_1, \ldots, x_n y_n\right) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^n$
- Product with a scalar: $\alpha \odot \mathbf{x} = \mathcal{C}\left(x_1^{\alpha}, \ldots, x_n^{\alpha}\right) \quad \forall \alpha \in \mathbb{R}, \quad \forall \mathbf{x} \in \mathbb{S}^n$
- Inner product: $\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} ln \frac{x_i}{x_j} \cdot ln \frac{y_i}{y_j} \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^n$

Exists a linear isometry called **isometric log-ratio transformation - ILR** that maps $\mathbb{S}^n$ in $\mathbb{R}^{n-1}$

[Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. Mathematical Geology, 35(3):279–300]

Downscaling of compositional data

In the Aitchison geometry $\mathbb{S}^n$ the ATPRK predictor reads:

$$\widehat{\mathbf{Z}}_k = \bigoplus_l \{\beta_n^l\}_n \odot u_k^l \oplus \bigoplus_K \lambda_K \odot \mathbf{e}_K$$

With respect to the classical Euclidean ATPRK predictor here the products and summations are in the simplex, in this way the predictor gives results in the simplex so in the feasibility region.

In particular, it is possible to use the standard ATPRK implementation thanks to the following **proposition**:

Given a compositional field $\mathbf{Z}(\mathbf{x}) \in \mathbb{S}^n$ *and a field* $\mathbf{Y}(\mathbf{x}) \in \mathbb{R}^{n-1}$ *s.t.* $\mathbf{Y} = ILR(\mathbf{Z})$ *.*
The ATPRK predictor for $\mathbf{Y}_k$ *is equal to the ILR-transformed predictor for* $\mathbf{Z}_k$ *in the Aitchison simplex:*

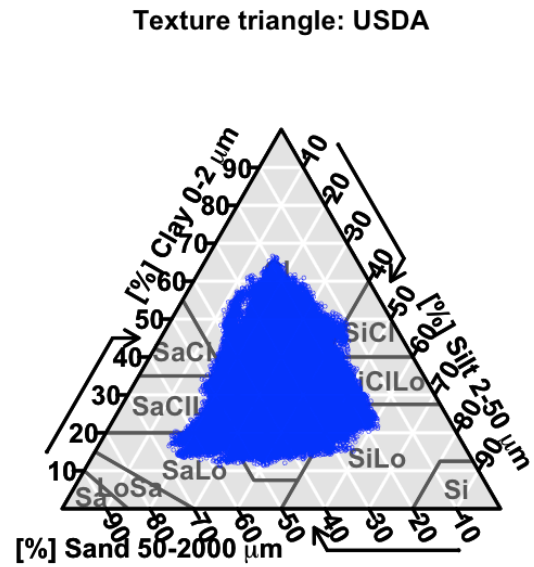
$$\widehat{\mathbf{Y}}_k = ILR(\widehat{\mathbf{Z}}_k)$$

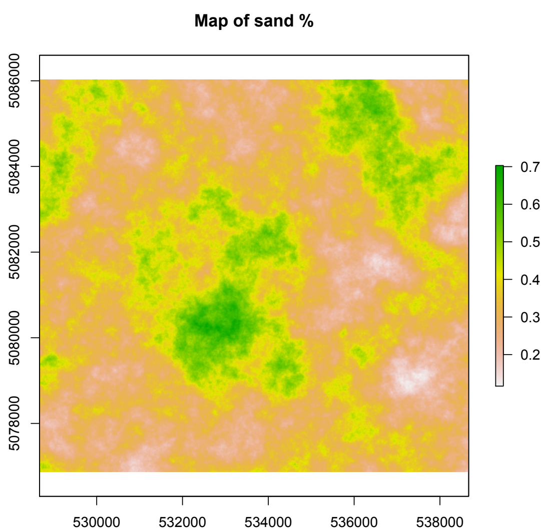
Downscaling in the Aitchison simplex imply the following **centre-preserving** property:

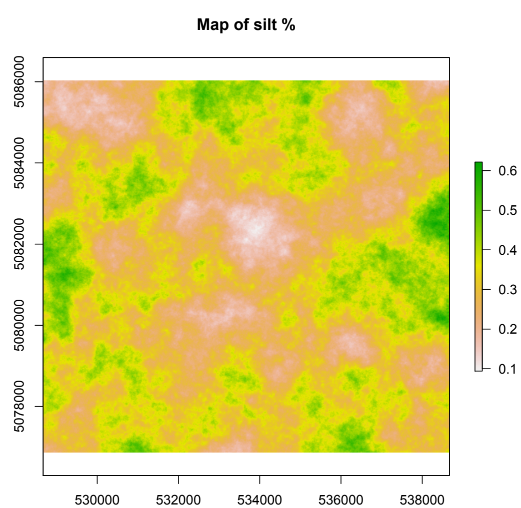
$$\mathbf{Z}_K = \frac{1}{P} \odot \bigoplus_{\mathbf{x}_k \in \nu_K} \widehat{\mathbf{Z}}_k$$

Synthetic data

Let us consider a set of 100 synthetic psf at a resolution of 20 m on a geographical domain of dimensions of the Caldone basin.
Each realization of this set is obtained via unconditional Gaussian simulation via using a spherical variogram model.





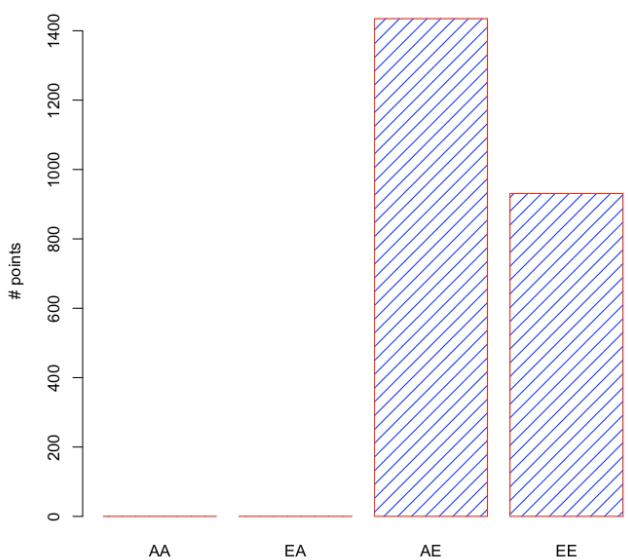


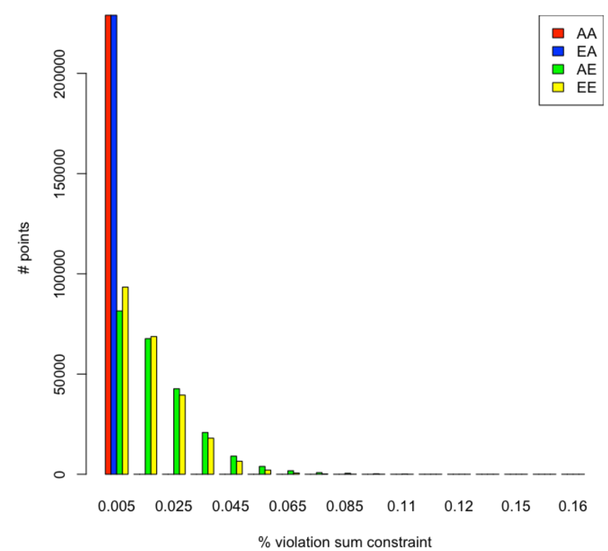
Synthetic data

For each of these realizations we perform upscaling and downscaling, both in Euclidean and Aitchison geometry, on a set of upscaling factors P. In this way, we obtain a set of 100 reconstruction of the initial 20 m psf for each method.

Below we show, on the left the mean violation of the positivity constraints, i.e. Clay, Silt, Sand ≥ 0 while on the right the mean percentage violation of the constant sum constraint, i.e. Clay + Silt + Sand = 1 .

AA: upscaling Aitchison, downscaling Aitchison
EA: upscaling Euclidean, downscaling Aitchison
AE: upscaling Aitchison, downscaling Euclidean
EE: upscaling Euclidean, downscaling Euclidean





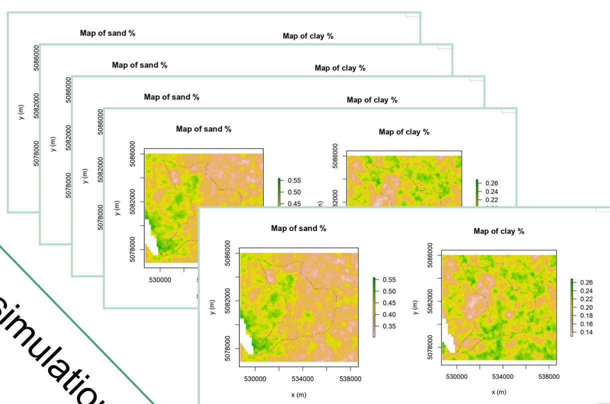
Case study results

A consequence of the **proposition** is that, in order to analyze how uncertainty propagates to the output of a numerical tool like SMART-SED tool, one can perform *block sequential simulations* directly on the ILR transformed datas, this is done in order to account for the absense of the SoilGrids psf confidence interval.

[P. Kyriakidis. A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36, 08 2004.]

In this way we can generate B psf maps and perform for each psf a numerical simulation. In order to analyze e.g. in a **Monte Carlo setting** how input uncertainty affects numerical model output.

Parameters' uncertainty



B simulations

Numerical model

$$\partial_t H + \nabla \cdot (H\, u) = (1-\mu)\, p - f,$$

$$\partial_t u + g \nabla \eta + u \cdot \nabla u + \gamma(u)\, u = 0.$$

$$\partial_t h_g + \nabla \cdot f_g = f - ev + s\,,$$

$$\partial_t h_{sd} + \nabla \cdot f_{sd} = w\,,$$

$$\partial_t h_{sn} = \mu\, p - s,$$

Output uncertainty

Modeling framework

Input parameters

Digital Terrain Model (DTM), with soil compositions

Model equations

De Saint Venant equations coupled with gravitational layer, sediment transport and snow equations

$$\partial_t H + \nabla \cdot (H\, u) = (1-\mu)\, p - f,$$

$$\partial_t u + g \nabla \eta + u \cdot \nabla u + \gamma(u)\, u = 0.$$

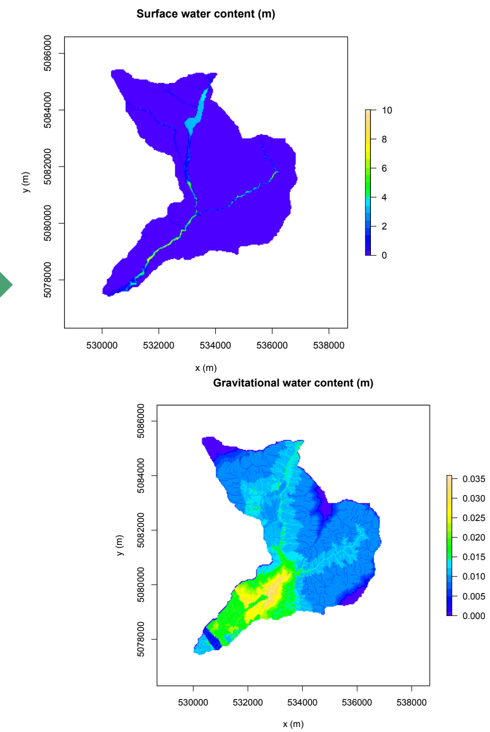
$$\partial_t h_g + \nabla \cdot f_g = f - ev + s\,,$$

$$\partial_t h_{sd} + \nabla \cdot f_{sd} = w\,,$$

$$\partial_t h_{sn} = \mu\, p - s,$$

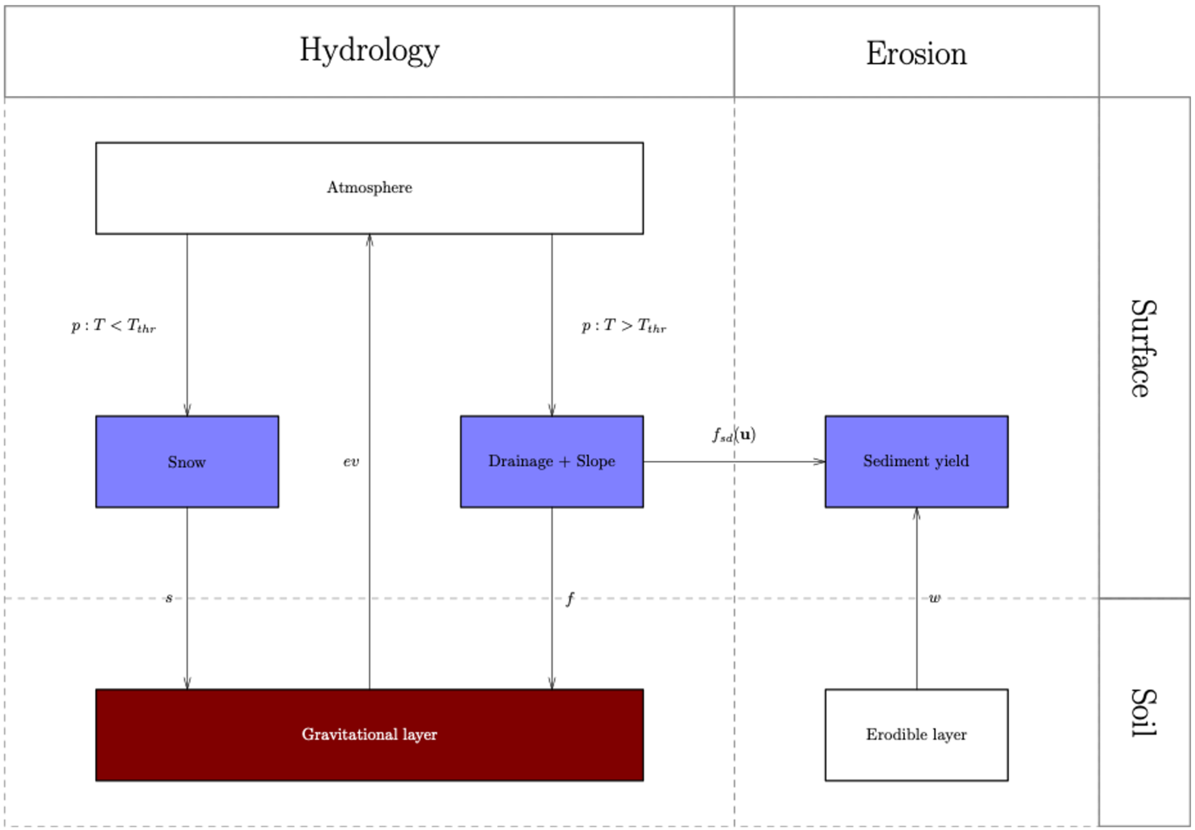
Simulation

State variable maps and temporal sequences



SMARTSED: simulation tool model structure overview

- Vertical arrows represent vertical fluxes, they link the surface processes with soil dynamics.
- Hydrology links to sediment yield transport via horizontal flux function of mass surface flux, horizontal arrow.
- Box colored in brown and blues are modeled via proper conservation law while atmosphere dynamics is neglected.

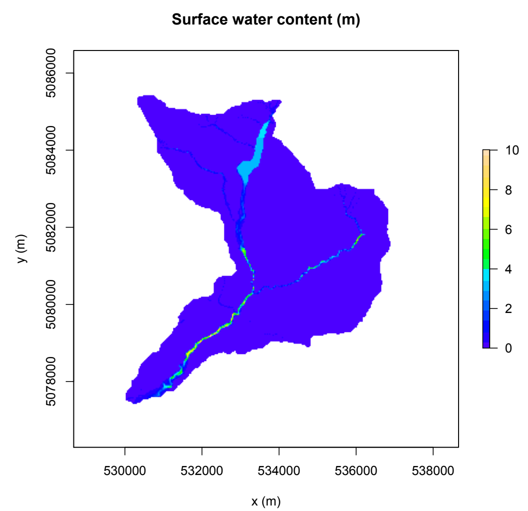


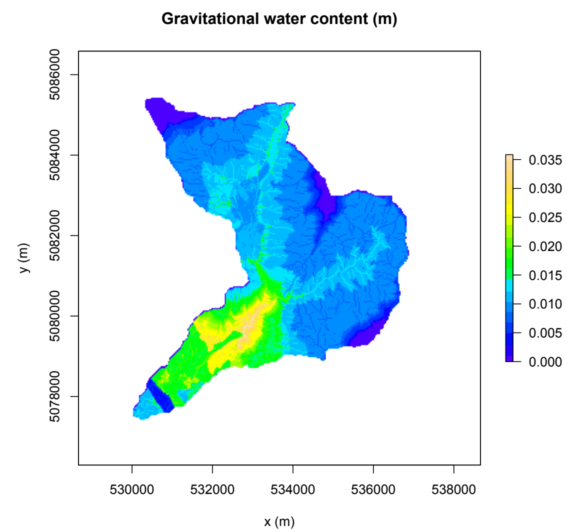
SMARTSED: simulation tool generalities

Soil texture information is needed to model properly infiltration (our model implements a SCS-CN method in a dynamic way).

Simulation tool main features:

- Ability to perform basin frequency response subject to multi-event rainfall
- Automatic determination of river bed, no distinction between drainage and slopes regions
- Output raster format in ESRI ASCII, data can be processed in applications like QGIS
- Basin scale sediment yield modeling





Conclusions and future developments

Conclusions:

We have shown the capability of the geostatistical model to deal with **downscaling of compositional data**, psf data for us, and shown one output of the numerical model for one day uniform and constant rain on the whole basin for one psf realization.

Future developments:

We are planning to perform a **global sensitivity analysis,** in a **Monte Carlo setting**, on the output of the numerical model subject to the variability of the input psf data for a given rainfall event.