Large-scale probabilistic optimization using non-stationary geostatistics for uncertainty assessment of groundwater flow and solute transport

Rogiers B1, Laloy E, Gedeon M2, Dassargues A2, Huyssmans M3, Batelaan O4,5, Mallants D1

1 Institute for Basic and Applied Sciences, SCK•CEN, P.O. Box 2400, BE-2420 Mol, Belgium. 2 Department of Environmental Sciences, Ghent University, Kruispoort 3, B-9000 Ghent, Belgium. 3 Environmental Research Institute, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. 4 School of the Environment, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia. 5 ONDRAF/NIRAS, Gate 4, Glen Osmond SA 5064, Australia.

* Corresponding author: brigers@sckcen.be

Introduction

Uncertainty quantification is very much needed to support decision making related to e.g. environmental impact assessment for waste disposal sites. A probabilistic result provides a much stronger basis for decision making compared to a single deterministic outcome. Accurate posterior exploration of high-dimensional and CPU-intensive models, which are often used for environmental impact assessment, is however a challenging task. To quantify the uncertainty associated with groundwater flow and solute transport in the framework of a near surface radioactive waste disposal in Mol/Dessel, Belgium (Fig 1), we investigate combining the adaptive Metropolis (AM; Haario et al. 2001) MCMC algorithm, and iterative spatial resampling (ISR; Mariethoz et al. 2009) for large-scale probabilistic optimization of a steady-state groundwater flow model (Rogiers et al. 2014).

Methods

Groundwater flow model

• Conditioned on borehole and direct push data (Fig 1), accounting for non-stationary heterogeneity in hydraulic conductivity (K) using distance-weighted geostatistics (Machuca-Mory & Deutsch 2012; Figs 2 & 3)
• Global parameters (i.e., spatially uniform)
  - HK_HUP_1D: K, multiplier for lower part model
  - R, BULL_AREA: bulk area recharge (% of meadow recharge)
  - R_TOTA: total recharge multiplier
• Spatially distributed parameters
  - XK and VANI (vertical anisotropy)
  - 41183 cells (each 5th grid cell, interpolated in between)
• Principal components are used for simulation, with hydraulic data as primary variable, and CPT and grain size based K predictions as secondary variable
• Compared to reference model parameterization with homogeneous hydrogeological units

Probabilistic optimization

• AM for updating global model parameters, ISR for spatially distributed K
• We use an exact variant of rejection sampling known as interrupted Markov chain, where we only accept better performing candidates, and interrupt the chain with probability \( \alpha = e^{-p(v_{\text{likelihood}})} \), where \( v \) is the log-likelihood

Random walk particle tracking

• Preliminary results from a basic implementation based on LaBolle et al. (2000), assuming constant porosity and isotropic dispersion
• Relative dispersivity fields estimated from outcrop investigations (Rogiers et al. 2013) and CPT-based K variance (Rogiers 2013)
• 1500-m scale dispersivity in the reference model, 250-m scale for the updated model simulations, using dispersivity-scale dependency from literature (Rogiers 2013)

Results

• 50 random realizations were optimized with the AM-ISR interrupted Markov chain approach (Fig 4)
• Global model parameter posterior distributions remain poorly characterized with only 50 samples (Fig 5)
• Model performance increased considerably, mainly due to better simulation of vertical head differences (Fig 6)
• Groundwater table elevation is mainly affected in the southern part; the variance is lowest near the river and drain network, and high observation-density areas (Fig 7)
• General solute plume evolution is similar, but differences exist in terms of arrival time and length of the plume in the lower aquifer (Fig 8)
• Correspondence between ensemble mean log10K fields and reference model depends on position in stratigraphy, as well as the variance, which additionally reflects the presence of primary and secondary data (Fig 9)

Conclusions

• Combination of AM and ISR proved to be effective for optimization of the non-stationary random fields
• More interrupted Markov chains should be run in order to get more robust posterior parameter distributions
• The random walk particle tracking code should be further optimized to reduce computation times

Acknowledgements

Several datasets on which the models used in this work rely, are provided by ONDRAF/NIRAS. Findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of ONDRAF/NIRAS.

References


Fig 1: Study area and site investigation points.

Fig 2: Non-stationarity in the variogram for the \( K_m \) and VANI principal components.

Fig 3: Non-stationarity in the mean and primary-secondary data correlation along the stratigraphical succession.

Fig 4: Evolution of the log-likelihood in function of the number of iterations, for 50 interrupted Markov chains.

Fig 5: Mean and 95% confidence interval for the model ensemble head observations and vertical head differences (inset graph). The reference model values are shown in red triangles.


Fig 7: A) Ensemble mean groundwater table elevation (masl). B) Difference with reference model. C) Variance.

Fig 8: Solute plume evolution based on a convolution of the random walk particle tracking results, with a unit mass per 100 days input flux at 12 50x50 m source cells.

Fig 9: K fields for the reference model and the ensemble mean and variance.

SCK•CEN | Boerentang 200 | BE-2400 Mol | www.sckcen.be | info@sckcen.be | 20140424 | 1 BRogeris