SCALABLE GENERALIZATION OF HYDRAULIC CONDUCTIVITY IN QUATERNARY STRATA FOR USE IN A REGIONAL GROUNDWATER MODEL

Jānis JĀTNIEKS[1], Konrāds POPOVS[1], Ilze KLINTS[2], Andrejs TIMUHINS[3], Andis KALVĀNS[1], Aija DĒLIŅA[1], Tomas SAKS[1]

[1]Faculty of Geography and Earth Sciences, [2]Faculty of Physics and Mathematics, [3]Laboratory for Mathematical Modelling of Environmental and Technological Processes. University of Latvia, Riga, Latvia. Contact: janis.jatnieks@lu.lv, www.puma.lu.lv

This work was supported by the European Social Fund project "Establishment of interdisciplinary scientist group and modelling system for ground-water research", 2009/0212/1DP/1.1.1.2.0/09/APIA/VIAA/060

Workflow overview

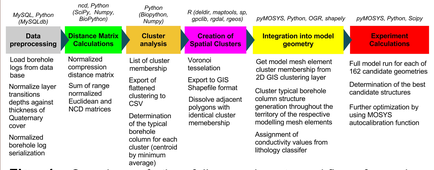

Fig. 1. Overview of the full experiment workflow for using Normalized Compression Distance and hierarchical clustering as an approch for scalable generalization of Quaternary sediment lithological structure in the MOSYS regional groundwater modelling system for the Baltic artesian basin territory.

Introduction

The cover of Quaternary sediments, especially in formerly glaciated territories usually is the most complex part of the sedimentary sequences. In regional hydro-geological models it is often assumed as a single layer with uniform or calibrated properties (Valner, 2003). However, the properties and structure of Quaternary sediments control the groundwater recharge: it can either direct the groundwater flow horizontally towards discharge in topographic lows or vertically, recharging groundwater in the bedrock. Building a representative structure of Quaternary strata in a regional hydrogeological model is important as this affects all the underlying strata.

Scope and objective

This work aims to present calibration results and detail our experience while integrating a scalable generalization of hydraulic conductivity for Quaternary strata into the regional groundwater modelling system for the Baltic artesian basin – MOSYS V1.

Methods

In this study the main unit of generalization is the spatial cluster and the main criteria for finding adequate representative borehole columns is the Normalized Copression Distance metric – a nonparametric datamining technique calculated by CompLearn ncd utility (Cilibrasi, 2007).

Spatial clusters were obtained from flattened complete-link hierarchical clustering dendrogram representations, supplying cluster membership identifiers for each borehole. Using Voronoi tesselation and GIS dissolve spatial analysis functionality on this data spatial clusters were obtained in GIS format. The procedure is described in detail in "Lithological Uncertainty Expressed by Normalized Compression Distance" by Jatnieks et al. 2012. Full overview of the experiment workflow with software components and key stages is shown in **Figure 1.**

We used two experiment matrices for hierarchical clustering. One where spatial clusters are obtained from Normalized Compression Distance (further identified as **NCD** matrix) matrix alone and a second where matrix is combined with range normalized Euclidean distance matrix of borehole locations in BalticTM93 projection (further identified as **NCD+E** experiment matrix).

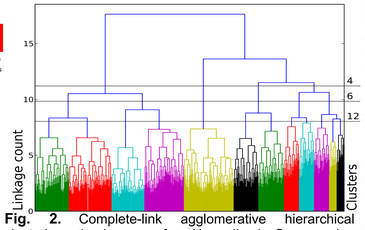

Fig. 2. Complete-link agglomerative hierarchical clustering dendrogram for Normalized Compression Distance matrix of serialized borehole log lithological structures.

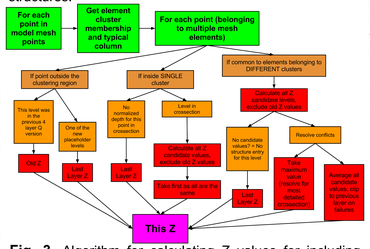

Fig. 3. Algorithm for calculating Z values for including spatial clustering results into the MOSYS finite-element mesh geometry and resolving structure conflicts, arising from different lithological structures, geological columns selected from the most typical borehole log in this cluster.

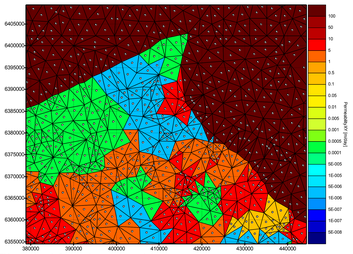

Fig. 4. After the algorithm described in Fig. 3 creates the layer structure into the model geometry, the respective conductivity values are assigned to mesh elements, created by the new placeholder layers. The cluster membership for each of mesh elements is determined by intersecting the element centroids, shown here in white with spatial clusters in GIS Shapefile layer. Log scale for conductivity values on Y axis.

This degree of similarity and the spatial heterogeneity of the cluster polygons can be varied by different flattening of the hierarchical cluster model into variable number of clusters. Such an approach **provides a scalable generalization** solution. It can be scaled from broad to more detailed generalization, depending on model and structural characteristics of the modelling territory, with the help of the clustering dendrogram (**Figure 2**).

Using the dissimilarity matrix of the NCD metric, a borehole, most similar to all the others from the lithological structure point of view, can be identified from the subset of all cluster member boreholes. The log structure of this borehole then is applied throughout the territory of the corresponding spatial cluster. The spatial locations of the same logical clusters can be disjoint - in different parts of the modelling territory and have the same borehole log column, representing the lithological structure throughout spatial clusters with the same logical cluster identifier.

162 model geometries were prepared, incorporating different candidate representations of Quaternary strata made from spatial clusters with the most typical borehole column, selected using NCD metric. The borehole log column lithological structure is composed using aggregate lithology types as shown in **Table 1** with their conductivity values.

To **integrate these results into the geometry** of regional groundwater model MOSYS V1, the algorithm, shown in **Figure 3,** calculates mesh point heights in meters above sea level for every mesh point in every newly created Quaternary layer.

After the layer surface calculations, the mesh element cluster memebership is determined by intersecting element centroids (white dots in **Figure 4)** with the spatial clusters.

The performance of resulting geometries is detailed in **Figures 8-11.** For both matrix variants (NCD and NCD+E) 82 model geometries, based on different flattenings of cluster dendrogram (**Figure 2**) were prepared and run in MOSYS V1 modelling environment. From the inital model run results in **Figures 8 and 10**, 3 better perfoming, one median and also the worst performing geometry was run through MOSYS autocalibration routine (Klints et al. 2012).

Class	Lithology	Horizontal conductivity, m/day	Vertical conductivity, m/day
1	sandstone	5	1
2	sandstone-low conductivity	1	0.5
3	sandstone-high conductivity	50	20
4	limestone	30	10
5	limestone-low conductivity	1	0.1
6	dolomite-high conductivity	100	30
7	dolomite	30	10
8	domerite (clayey dolomite)	0.1	0.01
9	clay	0.00001	0.00001
10	loam	0.001	0.001
11	sandy loam	0.001	0.001
12	silt	0.01	0.01
13	sand	5	5
14	gravel	20	20
15	gypsum	10	10
16	peat	10	10
17	gyttja and other biogenic sediments	0.0003	0.0003
18	soil	1	1
19	other	1	1
20	concrete	0	0

Table 1. Initial conductivity values from lithology classifier, used for generalization of borehole log structures for this study.

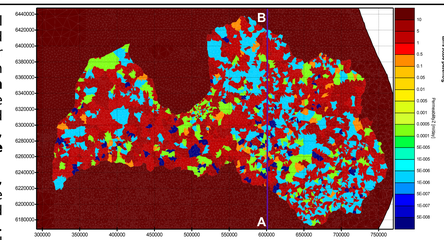

Fig. 5. Spatial distribution of vertical permeability values in the MOSYS V1 modelling system mesh in the clustering experiment territory in Latvia. Screenshot fragment of MOSYS visualization platform HiFiGeo, log scale, EPSG:25884 projection.

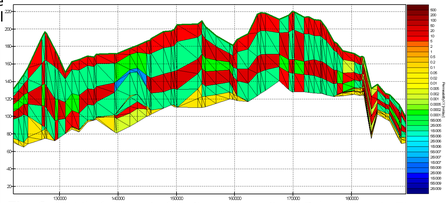

Fig. 6. Middle fragment of vertical crossection AB (overview in Figure 5) across the Vidzeme region, showing horizontal permeability values on the right Y axis in Quaternary represented by 8 modelling system layers (only Q layers shown, for simplicity). Left Y axis - Z values in meters above sea level. Distance from crossection A point on X axis.

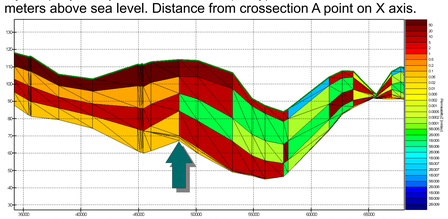

Fig. 7. Previous implementation of Quaternary strata in the MOSYS V1 modelling system consisted of 4 layers - two sets of aquifer-aquitard transitions. The arrow indicates the border between the previous Quaternary implementation (left) and the new Quaternary implementation, using hierarhical clustering of Normalized Compression Distanece, on the right.

This provided insight in their sensitivity to calibration through further optimization of hydraulic conductivity values in Quaternary strata (**Figure 12**). The model geometry, in which the Quaternary clustering experiment geometries were integrated, had already been calibrated until convergence before these experiements.

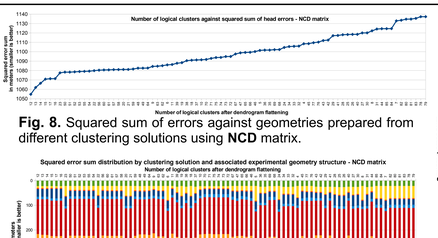

Fig. 8. Squared sum of errors against geometries prepared from different clustering solutions using **NCD** matrix.

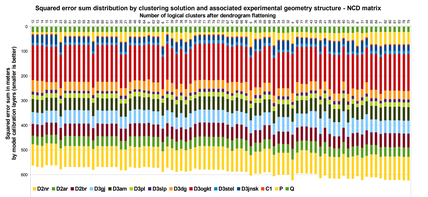

Fig. 9. Squared error distribution by model calibration layers for each of the geometries derived from different logical clustering solutions using **NCD** matrix.

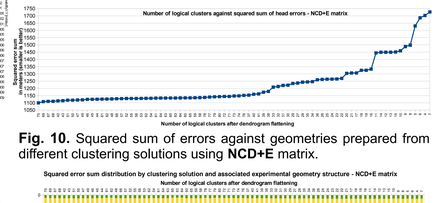

Fig. 10. Squared sum of errors against geometries prepared from different clustering solutions using **NCD+E** matrix.

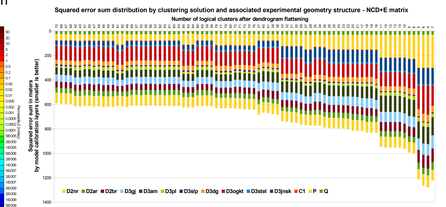

Fig. 11. Squared error distribution by model calibration layers for each of the geometries derived from different logical clustering solutions using **NCD+E** matrix.

The response of the selected new model geometries to calibration indicates their influence on the underlying model strata, not just improved performance in the part of the model representing the Quaternary deposits.

Geometries NCD 12c, NCD 13c were calibrated further, including the remaining bedrock layers as optimization parameters and yielding further decrease of squared error sum, used as overall fitness function for model performance against observed pressure heads.

Examples of crossections from NCD 12c structure are shown in **Figures 6 and 7.**

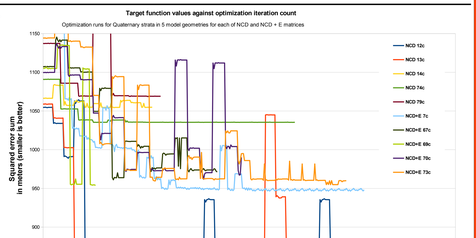

Fig. 12. Selected experimental model geometries and their response to calibration of model Quaternary strata.

Results

The experiment **geometries based on NCD matrix perform better** overall.

The various Quaternary representations tested have a measurable effect on model results - the worst geometries, such as NCD+E 7, 3 and others with low logical cluster count, perform more than twice as bad compared to the best NCD clustering based geometries.

NCD clustering based geometries with the better initial calculation results (in **Figures 8 and 10**), such as the 12 and 13 cluster variants, also tend to respond better to optimization with MOSYS auto-calibration routine (**Figure 12**).

Some the geometries with lower results from intial model run, show very minimal improvement with calibration (**Figure 12**).

For our experiment territory, consisting of formerly glaciated territory of Latvia, the **best performing generalizations have 12 and 13 clusters.**

This could have been successfully deduced from the clustering dendrogram in **Figure 2.**

The best performing model geometry before applying the NCD metric based clustering solution, had squared error sum of 814 meters. The best NCD clustering geometry provides **9.7% improvement** to 735. This may seem modest in nominal terms. It isn't, because the experimental Quaternary strata were inserted in a previously calibrated model geometry with initial conductivity values as estimates. Model autocalibration works by multiplying layer material properties with coefficients determined by the the Scipy L-BFGS-B routine (Zhu et al. 1997). Since this works on a layer basis, the proportional differences in conductivity values between elements in each layer remain constant. Ideally, the model geometry area built using clustering and the rest of the territory would be calibrated seperately. This could yield further improvements.

References

1. Bennett, C. H., Gacs P., Li M., Vitanyi P., Zurek W. 1998, Information Distance, IEEE Transactions on Information Theory, 44(4), 1407-1423., IEEE.
2. Cilibrasi, R. 2007., Statistical Inference Through Data Compression, ILLC Dissertation Series DS-2007-01, Institute for Logic, Language and Computation, Universiteit van Amsterdam.
3. Klints I., Virbulis J., Timuhins A., Sennikovs J. and Bethers U., Calibration of the hydrogeological model of the Baltic Artesian Basin, EGU Geophysical research abstracts EGU2012-10003 [this section poster board **A255**].
4. Takšdi, E. 1999. Datu bāzes "Urbumi" dokumentācija [Documentation of the database "Boreholes"]. Valsts ģeoloģijas dienests, Rīga. [In Latvian].
5. Valner, L. 2003. Hydrogeological model of Estonia and its applications. Proc. Estonian Acad. Sci. Geol., 52, 3, 179-192.
6. Virbulis, J., Bethers, U., Saks, T., Sennikovs, J. & Timuhins, A. Hydrogeological model of the Baltic Artesian Basin. Hydrogeology Journal. [Under revision].
7. C. Zhu, R. H. Byrd and J. Nocedal. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization (1997), ACM Transactions on Mathematical Software, 23, 4, pp. 550 - 560.