



PROCEEDINGS BOOK

OF THE

20th Congress of Hungarian Geomathematicians and
9th Congress of Croatian & Hungarian Geomathematicians
“Geomathematics in multidisciplinary science -
The new frontier?”

2017



IMPRESSUM

Publisher: Pécs Regional Committee of the Hungarian Academy of Sciences

Editors: István Gábor Hatvani, Péter Tanos, Marko Cvetković, Ferenc Fedor

Circulation: 80 copies

ISBN 978-963-8221-65-0

Subject Collection: Geomathematics

Organizers



Hungarian
Geological Society
(HGS)



Geomathematical
& Informatics
Section of the HGC



MTA CSFK
2ka Palaeoclimate
Research Group



Cluster of Applied
Earth Sciences



Geomathematical
Section of the
Croatian
Geological Society



University of
Zagreb, Faculty of
Mining, Geology
and Petroleum
Engineering



Geochem Ltd.



Szent István
University, Faculty
of Mechanical
Engineering

The South Transdanubian Regional Branch
of the Hungarian Geological Society

Hungarian Academy of Sciences,
Geomathematical Subcommittee of
Committee on Geology and the Regional
Committee in Pécs

Sponsors



MOL Plc.



Mecsekérc Plc.



HUNGARIAN
ACADEMY
OF SCIENCES

Hungarian
Academy of
Sciences
(NKSZ 2017-28)



Chamber of
Commerce and
Industry of Pécs-
Baranya



Üzletfejlesztés karnyújtásnyira

Enterprise Europe
Network



European
Commission

SCIENTIFIC PROGRAM

Wednesday (10.05) – Geostatistics today – pre-conference short course by János Geiger

Thursday (11.05) –

09:00- Registration

11:00-11:30 Opening ceremony with speeches by:

Ferenc Fedor - President of the Geomathematical & Informatics Section of the HGC

Tamás Síkfői - President of the Chamber of Commerce and Industry of Pécs-Baranya

Zoltán Unger - Associate President of the Hungarian Geological Society

11:30-12:00 Opening lecture by János Geiger - Statistical Process Control in The Evaluation of Geostatistical Simulations

13:00-14:30 Climate modelling past and future – chair: István G. Hatvani

- Keynote speaker: **Gabriella Szépszó - Climate adaptation in Hungary: from the climate model outputs to the end-users**
- **Dániel Topál** Detecting breakpoints in annual $\delta^{18}\text{O}$ ice core records from North Greenland
- **Csaba Ilyés** Examination of 110 year long Rainfall Data using Spectral and Wavelet Analysis
- **Tímea Kalmár** Regional climate modelling with special focus on the precipitation-related fine scale processes
- **Péter Szabó** Sources of uncertainties in climate model results

15:00-17:30 (with break) Big data in geoscience – chair: Zoltán Nagy

- Keynote speaker: **Balázs Székely - Big data in geoscience – are we big enough to use and to understand that properly?**
- **Zoltan Nagy** The Repository Information Model
- **Éva Farics** Determination of sedimentological processes of a coarse-grained deposits in Buda Hills applying combined cluster and discriminant analysis
- **László Sőrés** How to Document Scientific Work? Preserving Memories for the Semantic Web
- **László G. Somos** Statistic evaluation of digitized geophysical well logs data. (Digitized logs, Statistical evaluation, Multi-correlation)
- **Gyula Mező** Use of Sequential Gaussian Simulation for modelling groundwater pathlines and travel times near an underground radioactive waste repository
- **Tamás Miklovicz** Application of 3D geological modelling and grade tonnage calculations on Recsk ore complex
- **Ágnes Krupa** “Data-Mine” software: Complex earth scientific documentation of excavations with a uniform, real 3D background

Friday (12.05) -

08:30-10:30 (with break) Mathematical aspects of reservoir geology – chair: János Geiger

- Keynote speaker: **Marko Cvetković - Geomathematical aspects of reservoir and exploration geology**
- **István Nemes** From Russia with love – On production since 1947
- **Szabolcs Borka** Handling a mature clastic HC reservoir without seismic – trends, facies proportions and depositional zones
- **János Blahó** Permeability problematics in modelling of highly heterogeneous reservoirs
- **Ivona Emanović** Variogram analysis of well derived lithofacies data in Eastern part of Drava Depression
- **Zsolt Kovács** Evaluation of the trends of hydrocarbon migration processes based on oil density - reservoir depths relationship in Hungary

11:00-12:15 Geomathematics in water quality protection – chair: Balázs Trásy

- Keynote speaker: **József Kovács - Determining Water-Bodies Using Multivariate Data Analysis Methods**
- **Zsuzsanna Szabó** A reactive transport model of CO₂ and brine inflow to a fresh water aquifer
- **István G. Hatvani** Estimation of chlorophyll-a in rivers, on the example of the River Danube and Tisza
- **Péter Tanos** Combined Cluster and Discriminant Analysis on the River Raab's Austrian and Hungarian sections

13:30-15:00 Data analysis in engineering geology – chair: László Kovács

- Keynote speaker: **Ákos Török - Geomathematical aspects of engineering geology**
- **Balázs Czinder** Statistical analyses of strength parameters and aggregate properties of andesites from Hungary
- **Nikoletta R.-Boissinot** Evaluation of curve characteristics of shear strength along discontinuities using state-of-the-art breakpoint detection methods
- **Alina Vattai** Effect of JRC value on shear strength along discontinuities of Hungarian sandstones
- **Dániel Borbély** Permeability of rock mass around a radioactive waste repository tunnel

15:30-17:15 System theory, future of geology & geoinformatics – chair: Péter Tanos

- Keynote speaker: **Sándor Molnár- The qualitative assessment of structured systems**
- **Viktor Feurer** The 3rd/4th industrial revolution - Challenges for researchers –
- **Ferenc Fedor** How to automate a petrophysical laboratory? - challenges and solutions

- **Neven Trenc** Application of semi-automated GIS procedure for river terrace delineation on high resolution LiDAR data in Sava river valley west of Zagreb, Croatia – First results
- **Ana Brcković** Application of artificial neural networks for lithofacies determination in absence of sufficient well data
- **Nikolina Mijic** Quarry Surveying and Analysis Using Drone-mounted LiDAR & AutoCAD Civil 3D

Saturday (13.05) –

08:45-10:15 Poster session – chair: Marko Cvetković

- **Janina Horváth** Application of cluster vs. directional classification method to identify lithofacies and sedimentary elements
- **Tímea Kocsis** Comparison of parametric and non-parametric time-series analysis methods on a long-term meteorological dataset
- **Dániel Erdélyi** Variogram analysis of precipitation $\delta^{18}O$ over the Iberian Peninsula on a monthly-, seasonal- and annual scale
- **Artúr Kőhler** Trends in contaminant concentration time series
- **Shwan O. Hussein** The geographical evolution of urban greenness in the city of Erbil based on Landsat imagery
- **Norbert Magyar** Combined Cluster and Discriminant Analysis, an efficient tool in taxonomical classification
- **Zsuzsanna Szabó** A numerical model of Na-montmorillonite validated by batch experiments
- **Gábor Somodi** Relationship between geotechnical parameters and discrete fracture network simulation results in Bataapáti National Radioactive Waste Repository
- **Sándor Gulyás** Data acquisition, pre-processing of 3D image data of artificially distorted skulls, archaeological artefacts, fossils for 3D geometric morphometric analysis using CT and laser scanning: a comparison

10:30-12:30 Geomathematics in applied geosciences – chair: Ferenc Fedor

- Keynote speaker: **Zoltán Unger - Fault Statistics and Fractal Geometry**
- **Dorottya Kovács** Image processing for fractal geometry-based Discrete Fracture Network modelling input data
- **Tamás Jaskó** Quantitative Determination of Standard Sphericity and Roundness
- **Noémi Jakab** Determining the sufficient number of stochastic realisations to represent spatial uncertainty
- **Mihály Apró** Spatial uncertainty quantification using distance-Kernel method
- **László Z.-Sebess** Geothermal potential estimation with Monte Carlo method
- **Mátyás Sanocki** Uncertainty of subsurface uncertainty analysis' – can we mitigate it?

12:30-13:00 Closing ceremony – István G. Hatvani – Ferenc Fedor – Marko Cvetković

Table of contents

Statistical process control in the evaluation of geostatistical simulations.....	10
Part I - Climate modelling past and future.....	18
Climate adaptation in Hungary: from the climate model outputs to the end-users	19
Detecting breakpoints in annual $\delta^{18}\text{O}$ ice core records from North Greenland*	20
Examination of 110 year long Rainfall Data using Spectral and Wavelet Analysis	28
Regional climate modelling with special focus on the precipitation-related fine scale processes.....	36
Sources of uncertainties in climate model results	44
Part II - Big data in geoscience	45
Big data in geoscience – are we big enough to use and to understand that properly?	46
The Repository Information Model (A Dream of the Future).....	47
Determination of sedimentological processes of a coarse-grained deposits in Buda Hills applying combined cluster and discriminant analysis.....	48
How to Document Scientific Work? Preserving Memories for the Semantic Web	49
Statistic evaluation of digitized geophysical well logs data. (Digitized logs, Statistical evaluation, Multi-correlation).....	50
Use of Sequential Gaussian Simulation for modelling groundwater pathlines and travel times near an underground radioactive waste repository.....	51
Application of 3D geological modelling and grade tonnage calculations on Reesk ore complex	59
“Data-Mine” software: Complex earth scientific documentation of excavations with a uniform, real 3D background	60
Part III - Mathematical aspects of reservoir geology	61
Geomathematical aspects of reservoir and exploration geology	62
From Russia with love – On production since 1947.....	63
Handling a mature clastic HC reservoir without seismic – trends, facies proportions and depositional zones.....	71
Permeability problematics in modelling of highly heterogeneous reservoirs.....	79
Variogram analysis of well derived lithofacies data in Eastern part of Drava Depression ..	80
Evaluation of the trends of hydrocarbon migration processes based on oil density - reservoir depths relationship in Hungary.....	88
Part IV - Geostatistics in water quality protection	96
Determining water-bodies using multivariate data analysis methods	97
A reactive transport model of CO ₂ and brine inflow to a fresh water aquifer	98
Estimation of chlorophyll-a in rivers, with the example of the Rivers Danube and Tisza* ..	106
Combined Cluster and Discriminant Analysis on the River Raab’s Austrian and Hungarian sections*	114

Part V - Data analysis in engineering geology	122
Geomathematical aspects of engineering geology	123
Statistical analyses of strength parameters and aggregate properties of andesites from Hungary*	124
Evaluation of curve characteristics of shear strength along discontinuities using state-of-the-art breakpoint detection methods*	132
The effect of JRC value on shear strength along discontinuities in Hungarian sandstones*	140
Permeability of rock mass around a radioactive waste repository tunnel.....	147
Part VI - System theory, future of geology & geoinformatics	155
The qualitative assessment of structured systems.....	156
The 3rd/4th industrial revolution - Challenges for researchers.....	160
How to automate a petrophysical laboratory? - challenges and solutions.....	167
Application of semi-automated GIS procedure for river terrace delineation on high resolution LiDAR data in Sava river valley NW of Zagreb, Croatia – First results	175
Application of artificial neural networks for lithofacies determination in absence of sufficient well data	183
Quarry Surveying and Analysis Using Drone-mounted LiDAR & AutoCAD Civil 3D.....	191
Part VII – Diverse faces of Geomathematics - Poster Session	199
Application of cluster vs. directional classification method to identify lithofacies and sedimentary elements.....	200
Comparison of parametric and non-parametric time-series analysis methods on a long-term meteorological dataset	208
Variogram analysis of precipitation $\delta^{18}\text{O}$ across the Iberian Peninsula on monthly-, seasonal- and annual scales*	216
Trends in contaminant concentration time series	224
The geographical evolution of urban greenness in the city of Erbil based on Landsat imagery*	232
Combined Cluster and Discriminant Analysis, an efficient tool in taxonomical classification	240
A numerical model of Na-montmorillonite validated by batch experiments	241
Relationship between geotechnical parameters and discrete fracture network simulation results in Bataapáti National Radioactive Waste Repository	242
Data acquisition, pre-processing of 3D image data of artificially distorted skulls, archeological artefacts, fossils for 3D geometric morphometric analysis using CT and laser scanning: a comparison	243
Part VIII - Geomathematics in applied geosciences	244
Fault Statistics and Fractal Geometry*	245
Image processing for fractal geometry-based Discrete Fracture Network model input data	246
Quantitative Determination of Standard Sphericity and Roundness	254

Determining the sufficient number of stochastic realisations to represent spatial uncertainty	259
Spatial uncertainty quantification using distance-Kernel method.....	260
Geothermal potential estimation with Monte Carlo method	268
Uncertainty of subsurface uncertainty analysis – can we mitigate it?	275

Note: In general, the content of proceedings book has not been proof read by a native English speaker, and that is why solely the authors are responsible for the quality of language usage. However, certain chapters marked with ‘*’ have gone through English proof reading by Mr. Paul Thatcher; pt.64@live.co.uk .

Statistical process control in the evaluation of geostatistical simulations

János Geiger

University of Szeged, Institute of Geography and Geology, Department of Geology and Paleontology,
matska@geo.u-szeged.hu

This paper deals with a question: how many stochastic realizations of sequential Gaussian and indicator simulations should be generated to get fairly stable description of the studied spatial process. The grids of E-type estimations and conditional variances were calculated from pooled sets of 100 realizations (the cardinality of the subsets increases by one in the consecutive steps). At each pooling step, a grid average was derived from the corresponding E-type grid, and the variance (calculated for all the simulated values of the pooling set) was decomposed into a within and a between group variances. The former one used as a measurement of numerical-uncertainty at grid points, while the between group variance was regarded as a tool to characterize the geological heterogeneity between grid nodes. By plotting these three values (grid average, within and between group variances) against the number of pooling steps, three time-series could be defined.

The ergodic fluctuations of the stochastic realizations may result in some “outliers” in these series. From a particular lag, beyond which there is not any “outlier”, the series can be regarded as being fully controlled by a background statistical process. The number of pooled realizations belonging to this step/lag can be regarded as the sufficient number of realization to be generated. In this paper ARIMA processes were used to describe the statistical process control, since these models strongly rely on the autocorrelation (serial correlation). The paper also study how the sufficient number of realizations depends on grid resolutions.

The method is illustrated on a CT slice of a sandstone core sample.

Key words: *statistical process control, arima chart, sequential gaussian simulation, sequential indicator simulation*

1. INTRODUCTION

From in the early 90's considerable amount of efforts has been taken to several issues of uncertainty. Nevertheless, the issue of the number of realizations

required to explore the space of uncertainty and the convergence of the sequence of growing numbers of realizations have received much less attention. Goovaerts (1999) proposed an idea of creating subsets of increasing cardinality through the random sampling of an initial set of 100 response values. For each subset, descriptive statistics were used to study the impact of the number of realizations. He found that the extent of the space of uncertainty increased with the number of realizations, but at different rates depending on the response variable and simulation algorithm.

Geiger et al. (2010) generated 100 subsets with increasing sizes from 100 realizations. For each set, they calculated grid averages of the E-type estimations along with the between and within groups variances. They studied the limits of the sequences of these statistical properties within the frame of the stochastic convergence.

Pyrzcz & Deutsch (2014) used two parameters in their analysis: (1) the deviation from the reported quantile and (2) the fraction of times, the true value falls within \pm of the reported statistic. With these parameters they could calculate the uncertainty of their results due to a small number of realizations.

The common problems of these approaches are that the limited number of available realizations and the ergodic fluctuations make difficult any generalization.

Recently Sancho et al. (2016) has demonstrated a new approach, the statistical process control (SPC), in geostatistical analysis by which the "assignable" ("special") sources of variation can be differentiated from "common" sources in a time series. Following the view of Sancho et al (2016), this paper is addressed to the application of ARIMA-based statistical process control in the definition of the critical number of stochastic realizations.

2. METHODS

When performing any stochastic simulation, several equally probable realizations (stochastic images) are generated which reflect the specific targeted multivariate distribution. As in a former work (Geiger et al., 2010), 100 sets were created from 100 realizations with increasing cardinality. The first set

contained the first, the second comprised the first two, while the 100th contained all the 100 realizations. At each pooling step, the grid average was calculated from the corresponding E-type grid. The average conditional variance (calculated for all the simulated values of the pooling set) was decomposed into a within and a between group variances. The within group variance served as a measurement of numerical-uncertainty at grid points, while the between group variance was regarded as a tool to characterize the geological heterogeneity between grid nodes (Geiger et al., 2010). Plotting these three values (grid average, within and between group variances) against the number of pooling steps, three finite point-sequences could be defined, which could be processed as time series (**Figure 1**). In the followings, these time series are analyzed by using ARIMA-type statistical process control.

2.1 Statistical Process Control (SPC)

Statistical process control (SPC) is a method of quality control of time series using statistical methods. SPC is applied to monitor and control a process. It can be applied to any process where the "conforming product" (product meeting specifications) output can be measured. The objective is to analyse data and detect anomalous values of the variable analysed.

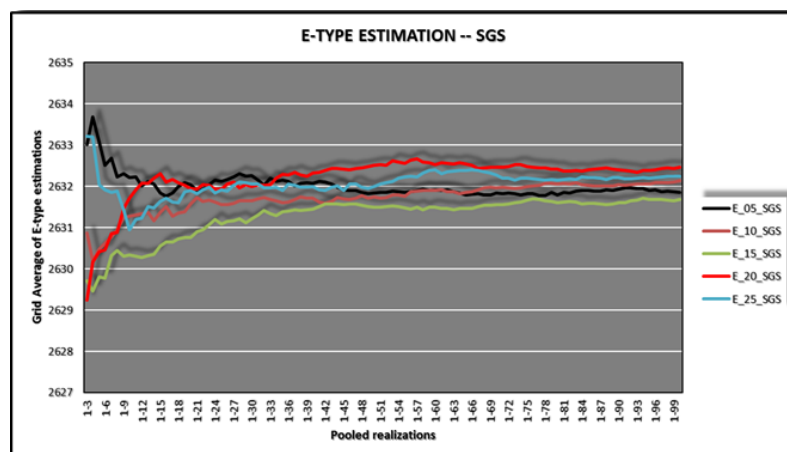


Figure 1: Time series representations of grid averages of pooled realizations

According to a certain tolerance margin and an objective value, control limits are defined. If the measurements are within the upper and lower control limits, there is not a non-random pattern in the distribution, and the process is under

statistical control. However, if there are points (measurements) outside the limits, there are external causes influencing the process.

2.2 ARIMA model and ARIMA chart

In this paper ARIMA chart was used in SPC. ARIMA model is a generalization of an autoregressive moving average (ARMA) model. In the most general form, ARIMA(p, d, q) models consist of three characteristic terms: (1) a set of autoregressive terms (p), (2) a set of moving average terms or non-seasonal differences (d) and a set of lagged forecast errors in the prediction equation (q). The model takes the following form:

$$Y_t = \mu + \Phi_1 \cdot Y_{t-1} + \dots + \Phi_p \cdot Y_{t-p} + e_t - \theta_1 \cdot e_{t-1} - \dots - \theta_q \cdot e_{t-q} \quad (1)$$

where μ is a constant, Φ_k is the autoregressive coefficient at lag k , θ_k is the moving average coefficient at lag k , and e_{t-k} is the forecast error that was made at period $(t - k)$.

The ARIMA Chart is a control chart for a single numeric variable where the data have been collected either individually or in subgroups. This model describes the serial correlation between observations close together in time. Out-of-control signals are based on the deviations of the process from this dynamic time series model (**Figure 2**). In this chart, the data are drawn around a centerline located at μ with control limits at

$$\mu \pm k \cdot \sigma^2 \quad (2)$$

where k is the multiple (in this paper $k = 1$). The mean and standard deviation depend on the specification of the ARIMA model.

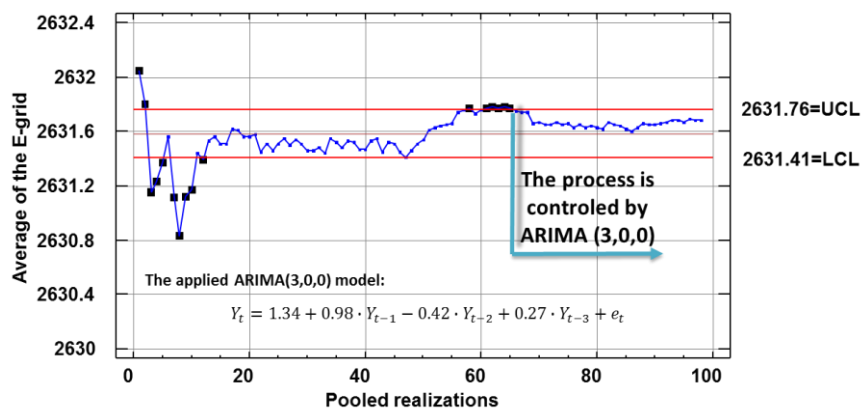


Figure 2: ARIMA(3,0,0) chart. The sufficient number of realization is 65. Legend:

UCL=upper control limit, LCL=lower control limit

In **Figure 2**, the black dots indicate those members of the series, where severe deviations can be detected from the used ARIMA(3,0,0) model. When the process does not trigger any deviation from the ‘detection rules’ of the control chart, it is said to be ‘stable’. In this paper the lag, from which this ‘stability’ is evident, is regarded as the sufficient number of realization to be generated. The significance of the selected ARIMA model was checked by partly statistical tests, and partly the residual autocorrelations of the ARIMA model.

3. THE WORK-FLOW

The suggested work-flow (**Figure 3**) is demonstrated by using a CT-slice oriented perpendicular to the vertical core axis of a sandstone core sample.

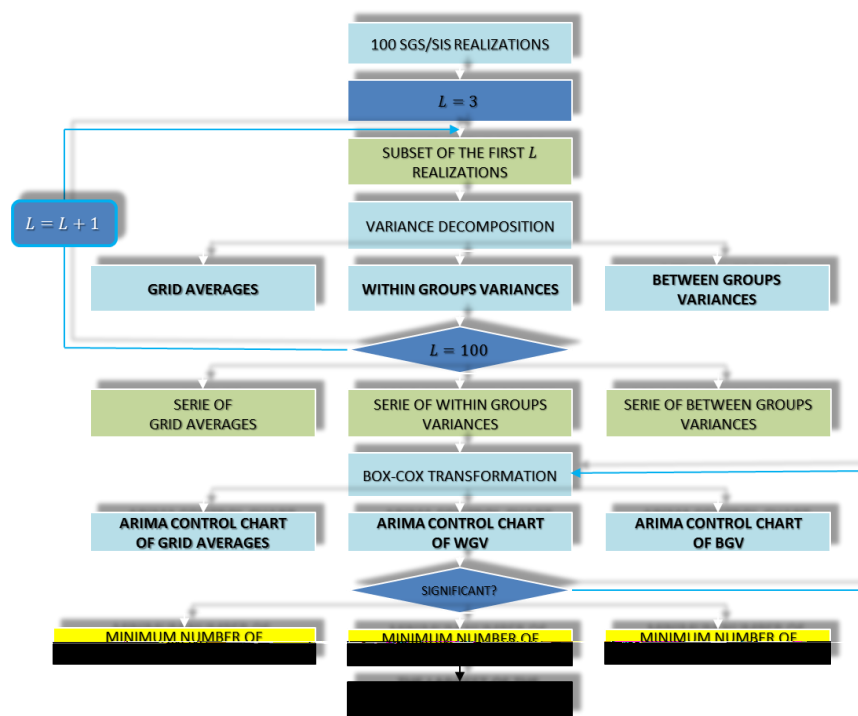


Figure 3: The work-flow

The image and the corresponding data set consisted of 16000 Hounsfield Unit values measured on a 125 x 128 regular grid. From this (exhausting) data set 100 data locations and the corresponding measurements were randomly chosen as the input data. With this sample data set a sequential Gaussian and a sequential indicator simulations were generated with 100 realizations in both approaches.

The reason of this choice was that Gaussian simulations are quite weeks to honor any lace-like geometry, while they are good in honoring small-scale heterogeneity. In contrast, indicator simulations are strong in honoring and revealing lace-like geometry, but they are quite week in describing small-scale heterogeneity. These simulations were calculated on five grid resolutions obtained by the methods suggested by Hengl (2006). The resolutions were the followings: (1) 0.5x0.5; (2) 1x1; (3) 1.5x1.5; (4) 2.0x.20; (5) 2.5x2.5.

The next step was the generation of pooled sets of realizations for the five grid resolutions (**Figure 3**). For each sets and grid resolutions, the grid averages of E-type estimations and the within group and between group variances were calculated. In this way, five time-series could be defined for each grid resolutions and simulation algorithms (an example is shown in **Figure 1**).

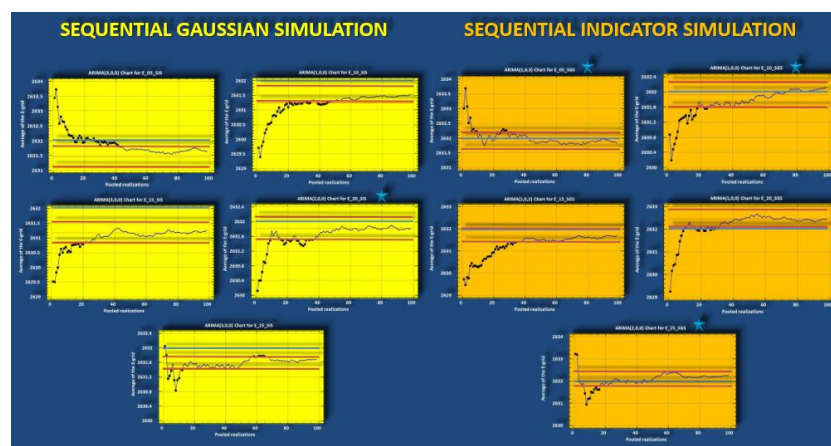


Figure 4: ARIMA Charts for the series of grid averages with different grid resolutions and simulation algorithms (For both simulations the grid resolutions from the top left to the right are (0.5x0.5), (1x1), (1.5x1.5), (2x2) and (2.5x2.5))

In the ARIMA modeling approaches, Box-Cox transformations ensured the requirement of normality. **Figure 4** gives an example of these models for the grid averages.

The final step was the identifications of those lags from which the series became to be fully controlled by the corresponding ARIMA models.

4. RESULTS

In **Figure 4a**, the number of realizations (lag) can be seen by simulation types and grid resolutions for all the three statistical properties. The highlighted two

columns contain the minimum of the row-wise numbers by simulation types. These results show that (1) SGS with (2.5x2.5) of grid resolution needs the smallest number of realizations. It is 22. The largest number of realizations, 89, belongs to the SGS with (0.5x0.5) of grid resolution. The **part B of Figure 4** shows the rank-number of realizations calculated row-wise for SIS and SGS. E.g. for the 0.5x0.5 grid resolution and sequential indicator simulation, the computation variance (WGV) needs the smallest number of realizations (rank 1 in **part B** which is 29 according to **part A of Figure 4**), then comes the geological variability (BGV, with rank 2 in **part B**, which is 35 in **part A of Figure 4**) and finally the grid average needs the largest number of realizations (E-type, with rank 3 in **part B**, which is 42 in **part A of Figure 4**) to serve stable description about the process.

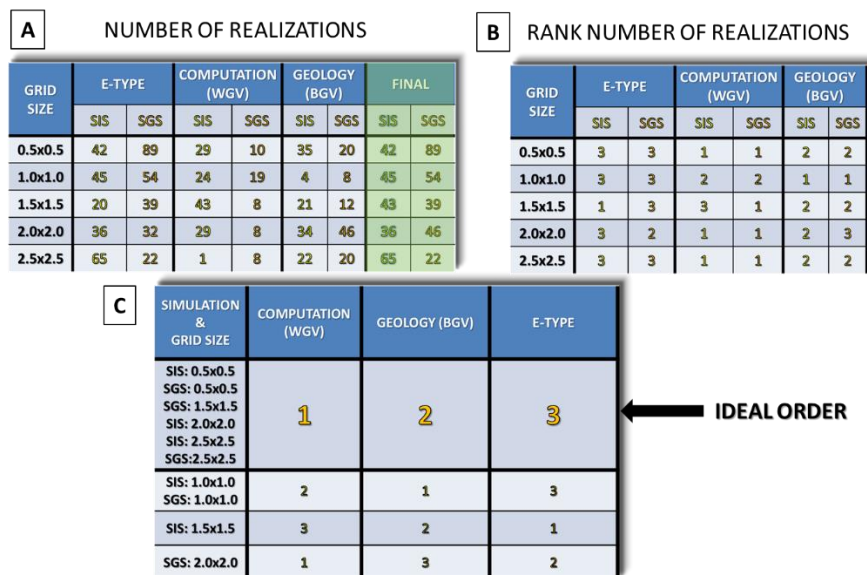


Figure 4: Summarization of the most important results

Ideally, the stabilization sequence should take the following order: firstly, the computation variability should be stabilized which should alleviate the geological variability (BGV) and this information should stabilize the grid average ('IDEAL ORDER', **part C of Figure 4**). This expectation comes true (1) for both the indicator and Gaussian simulations of (0.5x0.5) and (2.5x2.5) grid resolutions; (2) for (1.5x1.5) resolution with SGS; (3) for (2.0x2.0) resolution with SIS. That is why they can be suggested to use equally well in the description of lateral

variability of the CT slice. All the other studied portfolios are far from this ideal order (**part C of Figure 4**).

5. DISCUSSION

SPC can be used to any simulation properties. The method suggested does not need any extrapolation, it works well with the available number of realizations. To apply this approach, there is no need to know the analytical form of the normal distribution in the limit. The critical number of realizations (from which the SPC is complete) takes strongly the ergodic fluctuations into consideration. However, the method suggested needs enough number of realization to get stable information about the shape of the ,time series generated by the pooled realizations. The critical number of realizations strongly depends on the specification of the ARIMA model. It is also worth note that the critical number of realization does not give any information about the change of the spatial structure (e.g. connectivity) with the increasing number of realizations.

REFERENCES

- GEIGER, J, MALVIC, T., HORVÁTH, J. & NOVAK-ZELENIKA K. (2010): Statistical Characters of Realizations Derived from Sequential Indicator Simulation-In: Conference Book of 14th Annual Conference of the International Association for Mathematical Geosciences-IAMG 2010, 1-17.
- HENGL, T. (2006): Finding the right pixel size. *Computer & Geosciences*, 32, 1283-1298.
- GOOVAERTS, P. (1999): Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. *Stochastic Environmental Research and Risk Assessment*, 13, 161–182.
- PYRCZ, J.P. & DEUTSCH, C.V. (2014): *Geostatistical Reservoir Modeling*, Oxford University Press, Oxford, 449 p.
- SANCHO, J.; IGLESIAS, C., PIÑEIRO, J.; MARTÍNEZ, J.; PASTOR, J.; ARAÚJO, M.; TABOADA, J. (2016): Study of Water Quality in a Spanish River Based on Statistical Process Control and Functional Data Analysis. *Mathematical Geosciences*, 48, 163-186.

Part I - Climate modelling past and future

Climate adaptation in Hungary: from the climate model outputs to the end-users

Gabriella Szépszó¹

¹Hungarian Meteorological Service, szepszo.g@met.hu

High-quality, detailed and quantitative information on climate change is crucial to have targeted and sustainable adaptation strategies. Climate models provide objective tool to describe climate change with a given anthropogenic forcing. At the Hungarian Meteorological Service (OMSZ) two regional climate models (RCMs) are applied to estimate the future climate characteristics over the Carpathian Basin. RCM simulations are carried out on 10-50 km horizontal resolution using different (SRES, RCP) anthropogenic scenarios for the 21st century. Climate projections contain uncertainties originating from modelling the physical processes and the human activity. Therefore, instead of analysing results of a single model experiment, ensemble approach is used: available sets of regional and global climate model simulations (e.g., CMIP5, EURO-CORDEX) are examined to quantify the uncertainties and to estimate the role of different sources varying with the projection lead time, the target area and the given variable. RCM outputs serve as input data to investigate the climate change effects on different area. In Hungary, a National Adaptation Geo-information System (NAGiS) is established to support strategic planning and decision making related to the adaptation. NAGiS ensures the framework for the data provision, the impact and vulnerability assessments in different sectors, e.g., in hydrology, agriculture, forestry, biodiversity, tourism, critical infrastructures. Besides supplying RCM data for the impact studies, OMSZ is providing trainings and support for the users and end-users to properly utilize climate information.

The presentation is going to give an overview about the developments on adaptation to climate change impacts in last years in Hungary, with focus on the activity of the Hungarian Meteorological Service.

Key words: *regional climate models, projection uncertainties, impacts, adaptation*

Detecting breakpoints in annual $\delta^{18}\text{O}$ ice core records from North Greenland

Dániel Topál^{a,b}, István Gábor Hatvani^{b*}, Zoltán Kern^b

^aEötvös Loránd University, Department of Meteorology, H-1117 Budapest, Pázmány Péter stny. 1/A

^bInstitute for Geological and Geochemical Research, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, H-1112 Budapest, Budaörsi út 45.,

*hatvaniig@gmail.com

Interest in gaining a better understanding of past global changes is increasing, so additional and complementary information about past climates is needed. A generally employed approach is the use of proxy archives to extend the time span covered by instrumental temperature records. Water ice/firn stable isotopes have long been one of the most important and reliable proxies for temperature reconstructions. In the present study centennial $\delta^{18}\text{O}$ records from 9 ice cores of the *North Greenland Traverse* (NGT) ice coring campaign were examined to find common breakpoints (trend, shift, change in variance) in their time series. Two breakpoint detection methods with different algorithms were used to achieve this research aim. By studying the detected breakpoints in $\delta^{18}\text{O}$ data originating from different ice cores in the NGT, conclusions can be drawn concerning spatial impact of past climate changes in the least studied area of Greenland. In line with the dating uncertainty of the studied ice cores, a 10-year tolerance was allowed within the detected BPs are practically coincident. Results show a promising degree of coincidence between breakpoints detected by the applied methods, despite their differing computational backgrounds. This calls for involving further ice cores in the research from the area.

Key words: *breakpoint detection, cross entropy method, cpm method, greenland, paleoclimate, time series analysis*

1. INTRODUCTION

The Arctic has been experiencing major warming over the past few decades at a rate exceeding other parts of the Northern Hemisphere (Cohen et al., 2014). To understand the nature of such changes in the context of natural climate variability, it is essential to analyse long-term climate time series representative for the area (Steffensen et al., 2008). Since the available instrumental temperature records for the Arctic cover only a relatively short time-period (Cappelen et al., 2011), using proxy data – e.g. firn/ice $\delta^{18}\text{O}$ records which approximate ambient temperatures (Dansgaard, 1964) – is inevitable to extend the time span covered.

Motivated by a recent study on the *North Greenland Traverse* (NGT) ice coring campaign (Weissbach et al., 2016), the present research aims to find common breakpoints (BP, so-called ‘break point horizons’) in 9 NGT ice core $\delta^{18}\text{O}$ records spanning the last half millennium using two breakpoint detection methods.

2. MATERIALS AND METHODS

2.1. Ice core data

The ice/firn $\delta^{18}\text{O}$ (‰) records of the NGT ice cores were chosen for the following reasons: (i) North Greenland is the least studied area of the island, (ii) the NGT ice cores are proved to be more sensitive in recording late Holocene climate variations than cores originating from other parts of Greenland (Fischer et al., 1998) and (iii) the dating of the cores has recently been recalibrated (Weissbach et al., 2016) taking them all together, thereby ensuring a better comparability and easier handling of the

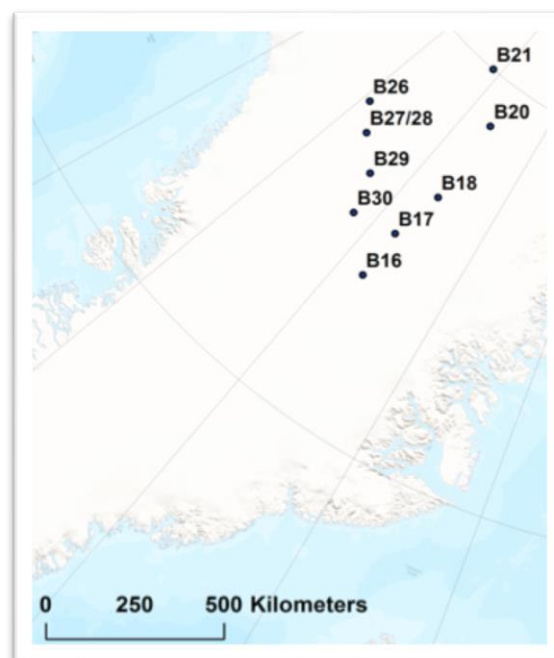


Figure 1: Map of Greenland with the NGT (B16-30) ice cores

ice/firn $\delta^{18}\text{O}$ records. Time-period 1471-1988 AD. was chosen in all the assessed records due to comparability reasons (**Figure 1**).

2.2. Breakpoint detection methods

Two different kinds of BP detection techniques were applied individually to 9 annualized ice/firn $\delta^{18}\text{O}$ records to find presumed BP horizons: a modified cross-entropy (CE) method (Priyadarshana & Sofronov, 2012) and a ‘Change Point Model’ (CPM) framework introduced by Hawkins et al. (2003) and further developed by Ross (2015). While the former is an iterative optimization procedure, using a modified Bayesian Information Criterion (Zhang & Siegmund, 2007) for detecting both the number and location of change points, the latter (CPM) consists of various parametric and non-parametric statistical tests. The methods are implemented in R (R Core Team, 2008) as `breakpoint` and `cpm` packages.

In the case of the CE method, due to continuous data the `CE.Normal.MeanVar` function was used with the distribution set to truncated normal (`distype=2`), the maximum number of breakpoints (`Nmax`) set to five, sample size (`M`) left as default (200) and `rho=0.25`, setting the number of elite samples to 129.

In the case of the CPM method, since almost all the ice/firn $\delta^{18}\text{O}$ time series were of Gaussian distribution the Student-t (detects changes in the mean), the Bartlett (detects changes in the variance), and the Generalized Likelihood Ratio (detects changes in both mean and variance) CPM packages were used (for further details see Topál et al., 2015; 2016). As some of the time series were not normally distributed, the non-parametric Mann-Whitney test of the CPM package was also applied. As for the parametrization, the Average Run Length (`ARL0`) was set to 200, which corresponds to $\alpha=0.95$, the startup sequence (`s`) was set to 33% of the total dataset length and the `processStream` function was used.

2.3. Coinciding BP horizons

A 10-year tolerance interval was chosen to detect coinciding BPs as common horizons based on both the dating uncertainty of the NGT cores (1-5 years; Weissbach et al., 2016) and the thorough tests on the uncertainty/efficiency of the applied methodologies (Topál et al., 2016).

3. RESULTS AND DISCUSSION

The applied methods were successful in finding BP horizons in the ice/firn $\delta^{18}\text{O}$ records of the 9 ice cores. Moreover, these frequently coincided with historical events proven to be reflected in climate records (**Figure 2**). Due to space limits, only the presence of a certain BP is discussed, rather than the distribution of the BPs between the CE and CPM methods. Note, that only those BPs are discussed out of the numerous BPs detected which represent the three chosen BP horizons.

3.1. 1795-1825 cooling event

Having run all the aforementioned BP detection methods one of the most characteristic BP horizons was 1795–1805 which is present in 7 of the 9 ice cores (B16: 1796, 1810; B17: 1801; B18: 1795/97, 1802; B20: 1797; B21: 1805; B23: 1794, 1802/3; B29: 1796/98, 1805; **Figure 2**). Also, noteworthy that the 1805 BP was found by all the different methods in core B21. Keeping in mind the introduced uncertainty interval, it is supposed that all BPs found within a 10-yr period can be considered to reflect the same environmental phenomenon. Thus, it is suspected that all the BPs detected between 1795–1805 indicate the beginning of a longer cooling period in climate (Fischer et al., 1998). The idea that the 1815 eruption of Tambora could not have caused such a long cooling period (Fisher et al., 1998) – starting from the end of the 18th century (1795–1825) – concurs with the presented findings that the change had already begun decades before. Thus, the eruption of the Tambora volcano in 1815 only further amplified the previously initiated cooling in climate. However, further research – including periodic component analyses also breakpoint analyses of stacked ice

cores (to improve the signal-to-noise ratio) – is needed (e.g. Weissbach et al., 2016).

3.2. Little Ice Age (LIA)?

An additional BP horizon was found (1621-1633) indicated by common BPs in 4 ice cores (B16: 1622, B17: 1629, B18: 1633, B21: 1621). These might show evidence of a start of a cooling period in the early 17th century (**Figure 2**) coinciding with the beginning of deepened cooling within the LIA (Fisher et al., 1998). Nevertheless, further research is needed for a clearer view. Following Fischer et al. (1998) who found evidence for the LIA in a stack of three NGT cores, BP analysis was conducted on the NGT stack data from Weissbach et al. (2016) consisting of 12 NGT ice cores. Thus, the CE method found two BPs in 1631 and 1700 which might also be related to the same deepened cooling within the LIA.

3.3. Detected volcanic horizons

Volcanic marker horizons play a major role in ice core dating, since e.g. from non-sea salt SO_4^{2-} concentration profiles, distinct volcanic horizons can be identified which can be used as match points in synchronizing ice cores (Sigl et al., 2015).

Several BPs were found in the 9 cores which may be connected to one-time short cooling events throughout the late Holocene (**Figure 2**). BPs were detected: (i) between 1816-24 in four ice cores simultaneously (B16, B27/28, B29, B30), (ii) in B20 (BP=1916/19) and (iii) between 1692-96 in 3 ice cores (B20: 1692, B21: 1696, B30: 1693).

Interestingly, the volcanic horizons used to synchronize ice cores in Weissbach et al. (2016) fall in line with some of the BP horizons discussed above: (i) BPs between 1816-24 may be related to the 1815 Tambora eruption; (ii) the 1916/19 might be linked to the 1912 Katmai eruption and (iii) the BP horizon of 1692-96 may coincide with the 1694 Hekla eruption.

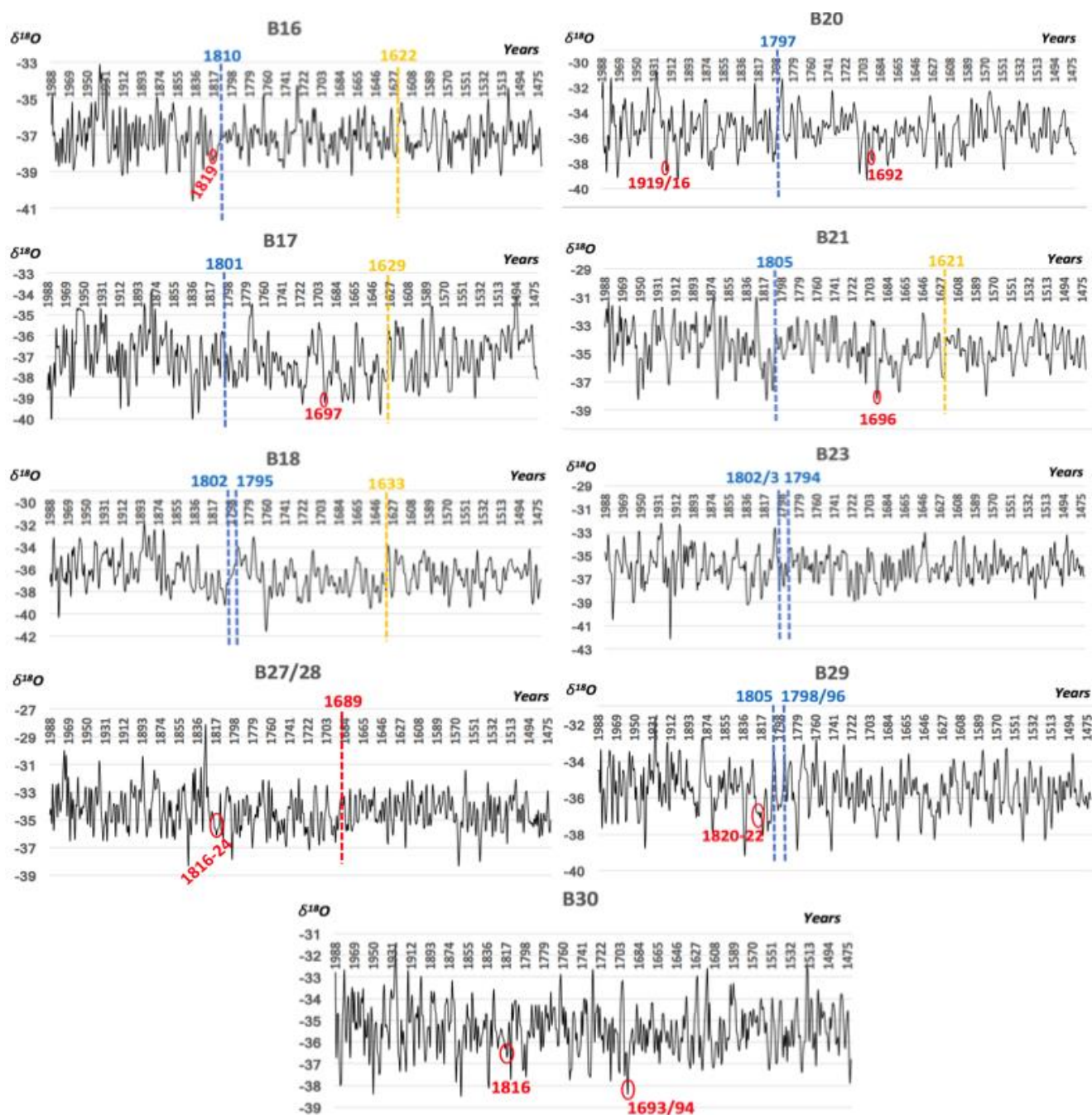


Figure 2: Detected BP horizons in $\delta^{18}\text{O}$ (‰) records of cores B16-B30 (1471-1988): 1795-1810 horizon (dashed blue lines); 1621-33 horizon (dashed yellow lines). Red circles indicate BPs suspected to reflect volcanic horizons documented in Weissbach et al. (2016)

4. CONCLUSIONS

In summary, variations of annual $\delta^{18}\text{O}$ values of 9 ice cores from North Greenland were studied using breakpoint analyses. Some very distinct breakpoint horizons were detected which either matched with previously known climate events or proposed further studies. Our analysis showed that studying mutual changes in annual $\delta^{18}\text{O}$ fluctuations originating from spatially distributed ice cores may be an efficient way of detecting late Holocene climate variations in the region.

ACKNOWLEDGEMENTS

The authors thank for the support of the MTA “Lendület” programme (LP2012-27/2012) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. This is contribution No. 46 of 2ka Palæoclimate Research Group.

REFERENCES

- CAPPELEN, J., LAURSEN, E.V., JØRGENSEN, P.V. & KERN-HANSEN, C. (2011): DMI monthly climate data collection 1768-2010, Denmark, the Faroe Islands and Greenland, *Technical Report 05-05*.
- COHEN, J. et al., (2014): Recent Arctic amplification and extreme mid-latitude weather, *Nature Geoscience*, 7, 637-637.
- DANSGAARD, W. (1964): Stable isotope in precipitation, *Tellus*, 16, 436-468.
- FISCHER, H. et al. (1998): Little Ice Age clearly recorded in northern Greenland ice cores, *Geophys. Res. Lett.*, 25, 1749-1752.
- HAWKINS, D.M., QIU P. & KANG C.W. (2003): The Changepoint Model for Statistical Process Control, *J. Quality Technology*, 35(4), 355-366.
- PRIYADARSHANA, W.J.R.M. & SOFRONOV G. (2012): A Modified Cross-Entropy Method for Detecting Multiple Change-Points in DNA Count Data, In *Proc. of the IEEE Conference on Evolutionary Computation (CEC)*, 1020-1027.
- R CORE TEAM, R (2008): A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

- ROSS, G.J. (2015): Parametric and Nonparametric Sequential Change Detection in R: The cpm Package, *J. Stat. Softw.*
- SIGL, M. et al. (2015): Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, 523, 543-54.
- STEFFENSEN, J.P. et al. (2008): High-resolution Greenland ice core shows abrupt climate change happens in few years, *Science*, 321, 680-684.
- TOPÁL, D., HATVANI, I.G., MATYASOVSKY, I. & KERN, Z. (2015): Break-point detection algorithms tested on artificial time series. - In: HORVÁTH, J., CVETKOVIĆ, M., HATVANI, I.G. (eds.): 7th Croatian - Hungarian and 18th Hungarian Geomathematical Congress: "The Geomathematical Models: The Mirrors of Geological Reality or Science Fictions?". Conference venue: Hungarian Geological Society, Mórahalom, 147-154.
- TOPÁL, D., MATYASOVSKY, I., KERN, Z. & HATVANI I.G. (2016): Detecting breakpoints in artificially modified and real-life time series using three state-of-the-art methods, *Open Geosciences*, 8, 78-98.
- WEISSBACH, S., WEGNER, A., OPEL, T., OERTER, H., VINTHER, B.M. & KIPFSTUHL, S. (2016): Spatial and temporal oxygen isotope variability in northern Greenland – implications for a new climate record over the past millennium, *Clim. Past.*, 12, 171-188.
- ZHANG, N.R. & SIEGMUND, D.O. (2007): A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data, *Biometrics*, 63, 22-32.

Examination of 110 year long Rainfall Data using Spectral and Wavelet Analysis

Csaba Ilyés¹, Endre Turai², Péter Szűcs¹

¹University of Miskolc, Faculty of Earth Science and Engineering, Institute of Environmental Management, MTA-ME Research Group of Geoengineering H-3515 Miskolc, Egyetemváros, Hungary, hgilyes@uni-miskolc.hu

²University of Miskolc, Faculty of Earth Science and Engineering, Institute of Geophysics and Geoinformatics, H-3515 Miskolc, Egyetemváros, Hungary

The aim of our research was, to better understand the precipitation part of the hydrological cycle, because through infiltration, this is the most important source of recharging the groundwater. To better understand the periodicity of these rainfalls we used monthly and annual rainfall data also. We examined 110 year long precipitation time series' from four different cities, across the Carpathian-basin, obtained from the Hungarian Meteorological Service. With Discrete Fourier-transformation (DFT) and Wavelet time series analysis, we defined local and country-wide cycles in the datasets.

Using DFT we calculated the time-period distributions (spectra) of monthly and annual rainfall data. Spectra shown 21 cycles in Budapest, 16 in Debrecen, 17 in Pécs and 20 in Szombathely from the annual rainfall data. The most dominant cycle was the 5 year long, with the most dominant relative amplitude in all four stations, and there were 12 other present in all four datasets. From the monthly datasets several other periodic components were calculated locally and nationwide also.

Using Wavelet-analysis the time dependence of the cycles was determined in the 110 year long dataset in case of two cities, Debrecen and Pécs.

The paper will present the mathematical algorithm of the spectral and wavelet analysis besides the numerical results.

Key words: *precipitation, spectral analysis, wavelet, cycles*

1. INTRODUCTION

On Earth approximately 400 000 km³ volume of water is being transported annually in the water cycle, which is affected by the changing climate and the meteorological extremities present in recent years. In Hungary most of the drinking water is produced from groundwater aquifers, therefore the effects of a changing climate and even the slightest changes in the water cycle can have a strong effect to these aquifers. These changes in the behaviour of precipitation have an impact on the groundwater resources through the recharge, so we chose to investigate the precipitation. Several mathematical methods, such as the Lomb-Scargle periodogram (Nason et. al. 1999), the Wavelet Time Series Analysis (Kovács et. al. 2010), and the analytic version of the Discrete Fourier-transformation can be used to define periodicity. These were used before in several studies to examine precipitation in California (Sangdan 2004), at the Sanjiang Plain (Liu et al 2009) and in Gannan County, China (Zheng et al 2014), the Bükk and Mátra Mountains of Hungary (Kovács et. al. 2014) and in Central-America (Hastenrath 1964).

2. THEORETICAL BACKGROUND

The long term hydrometeorological datasets are considered to be time series, containing several of periodic components therefore we chose to examine it with spectral analysis, based on the Fourier-Transformation (Meskó 1984). Working with harmonic functions in the analytic Fourier Transformation, a complex Fourier-spectrum $F(f)$ is obtained, which can be divided into a real and an imaginary part or can also be defined in an exponential form, by introduction two other real spectra.

The $A(f)$ spectrum is called the amplitude, while the $\Phi(f)$ spectrum is called the phase spectrum. The amplitude spectrum gives the weight in the formation of the signal of the harmonic component falling into a frequency band unit around any frequency. The phase spectrum shows, what part of the period length the maximum of this harmonic component shifts in relation to the maximum of base function $\cos(2\pi ft)$ (Ilyés et al, 2017).

Basically these meteorological processes are stochastic, but in this paper we searched for deterministic components in our datasets with the criteria, that the time series has no significant trend and the sampling rate is equidistant.

3. SPECTRAL ANALYSIS

To determine the cyclic parameters we used the data from the OMSZ (Hungarian Meteorological Service) online database (HMS 2015), which contains 110 years of meteorological parameters for five cities.

In this paper we present the results of Debrecen and Pécs, the two cities, which presented very different results, according to our measurements. After defining the cyclic parameters, the amplitude, the phase angle and the frequency, the cycles with the relative amplitude spectrum over 50 % were defined as major (dominant) cycles and cycles with relative amplitude spectrum between 20 % (in some cases 10 %) and 50 % were defined as additional (minor) cycles.

3.1 Annual precipitation

In this case, the registration period is 1901 – 2010, thus the length of the registration period, $t_{\text{reg}}=110$ years, the sampling rate is 1 year, and the numbers of samples are 110 from each city. The Nyquist frequency is 2 year (Meskó 1984). The Nyquist-frequency shows the minimal length of period of time that can be calculated correctly with this examination method.

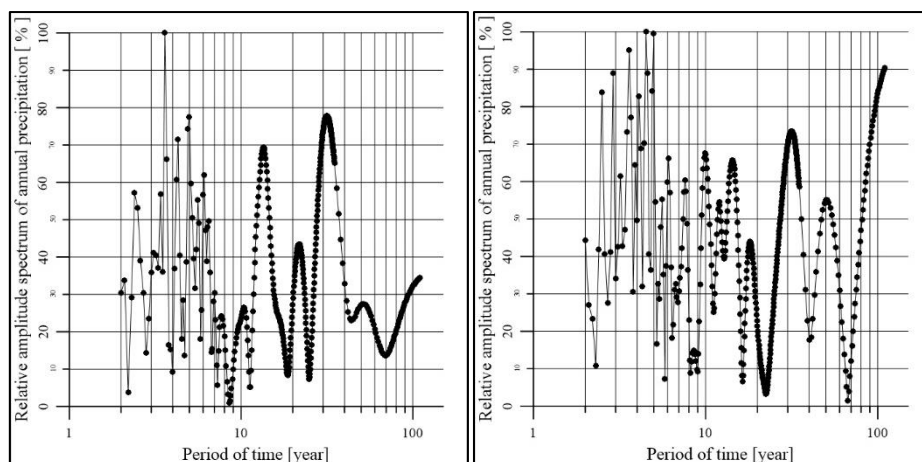


Figure 1: Relative amplitude spectra of Debrecen (l) and Pécs (r) according to annual precipitation data

In Debrecen 16 cycles were determined, 10 major cycles, and 6 additional cycles. In this city the 3.6 year long cycle was the most dominant, with 100 % relative amplitude, the other cycles amplitude spectra were between 27 – 77 % (**Figure 1**).

In Pécs 17 cycles were calculated, 16 of them are major, and only 1 is additional cycle, which means, the periodicity is much more dominant in the precipitation time series in Pécs, than in Debrecen. We can describe the precipitation in Pécs with few, but dominant cycles. The most dominant 4.5 year long cycle followed by the 5 year long, with almost the same relative amplitude value, and the third is the 3.6 year long with more than 90% relative amplitude.

3.2 Monthly precipitation

The registration period is January 1901 – December 2010, the length of the registration period, $t_{\text{reg}}=1320$ months, the sampling rate is 1 month, and the numbers of samples are 1320 from each city, the Nyquist frequency is 2 months.

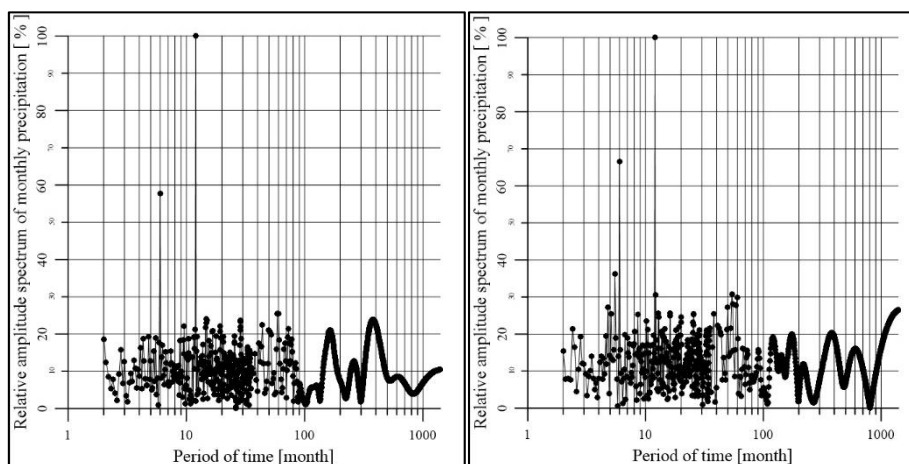


Figure 2: Relative amplitude spectra of Debrecen (l) and Pécs (r) according to monthly precipitation data

In Debrecen 43 cycles were detected, in Pécs 65 (**Figure 2**). The similarities of this two stations are the following: The 1 year long cycles are the most dominant with 100 % relative amplitude - normalized to maximum amplitude value of the dataset - , and the second is the 0.5 year long with 57.64 % relative amplitude

in Debrecen, and 66.53 % in Pécs. In Debrecen the other relatively dominant cycles are the 59, 14.7, 378 month long ones, most of the minor cycles have the relative amplitude range under the 20 % value, but were considered as important ones in the precipitation dataset. In Pécs, the dominant minor cycles are the 5.5, 54, 12.2, 60 month long ones.

3.3 Forecasting

With the $A(f)$ amplitude spectra and the $\phi(f)$ phase spectra, the original measured data can be recalculated, and with the major and minor cycles, and their period of time, amplitude and phase spectra values, the deterministic precipitation time series can be calculated (Eq. 1):

$$y(t)^{det} = \bar{Y} + \frac{2}{T_{reg}} \sum_{i=1}^{18} A_i \cos \left[\frac{2\pi}{T_i} (t - 1901) + \phi(T_i) \right] \quad (\text{Eq. 1})$$

If we use the t parameter, with a value of $t > 2010$, we can forecast to the future. In this paper we present the results of the 110 year long time series, calculated from the periodic components of the station of Debrecen from 1901 to 2010, and its correlation coefficient with the original measured data, with the forecast up to the year of 2030.

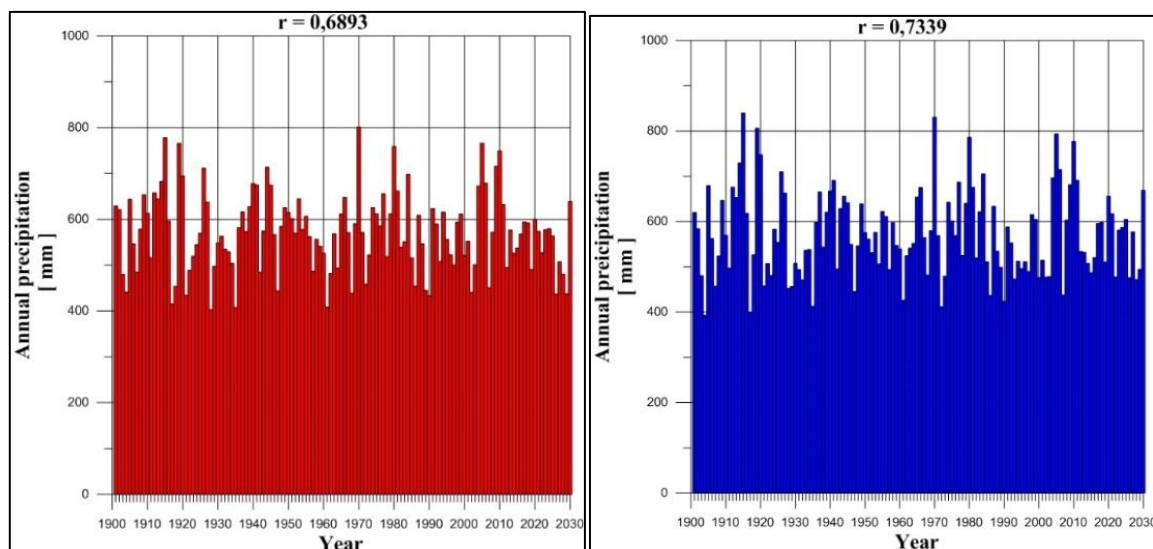


Figure 3: Forecasting up to 2030 from the periodic components of the Debrecen dataset.

The annual precipitation of Debrecen values range between 400 and 600 mm, with no year, when the 600 mm maximum would be exceeded (**Figure 3**). According to our calculations, large amount of precipitation within one year is not expected. The correlation coefficient between the original measured data and the calculated time series are 0.689 and 0.734 respectively.

4. WAVELET ANALYSIS

Wavelet time series analysis is a well-known method to investigate the time-dependence of a cycle within a time series. The periodic components from the previous examinations were used for this analysis.

The wave packet used for the calculation was a 1 year long period of time sine wave. The filtering has been derived by a cross-correlation function normalized to maximum value of the function. A discrete normalized cross-correlation matrix (R_{xy}) can be calculated from the cross-correlation functions of the wavelets (x) and the precipitation time series (y).

All of the 18 cycles from Debrecen were used for the wavelet analysis, and the results showed, that the cycles with period of time with a small value are less dominant in the 1910's, 30's and in the 70's.

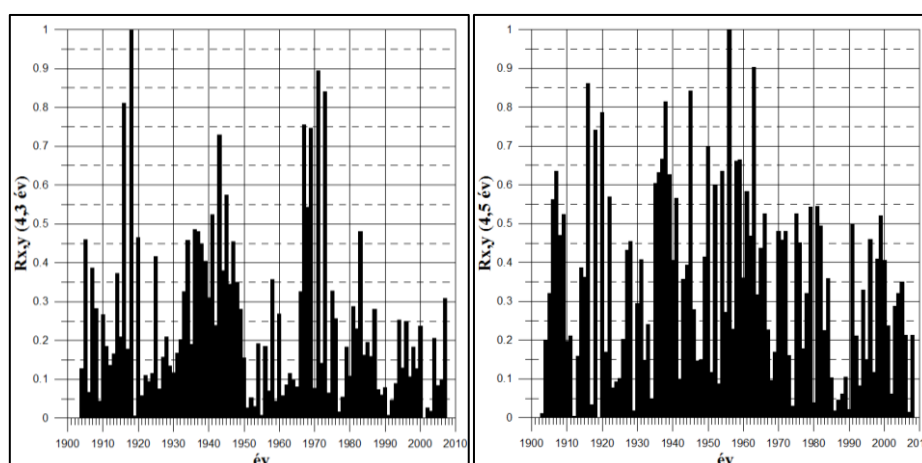


Figure 4: Wavelet cross-correlation of the 4.3 year long cycle from Debrecen (*l*), and the 4.5 year long from Pécs (*r*).

The wavelet of the 4.5 year long cycle shows 5 local maximum value (with more than 0.8 correlation coefficient), in the year 1956, 1963, 1916, 1945 and 1938.

The wavelet of the 4.3 cycle from Debrecen shows minimal similarities, with 4 maximum values, in the years: 1918, 1971, 1973 and 1916. The 1910's and the 1960-1970's decades have the more dominant values of the cycles in both cases, with much less dominance can be noticed after 2000 (**Figure 4**).

5. CONCLUSIONS

In recent years several studies explored the changes of the hydrological cycle. In this paper we examined the precipitation, to better understand the variability of it, with several mathematical methods based on DFT Fourier-transformation. According to our research, there are several different cycles in the rainfall datasets all over Hungary. We calculated 13 cycles which were present in all of the time series' and many others were calculated locally. We tested the forecasting method with the Debrecen precipitation records and it gave a strong correlation coefficient between the determined and the original measured data. Also the time-dependence of the cyclic parameters were determined with Wavelet analysis in case of two cities, Debrecen and Pécs. For further researches, the aim could be to better understand the nature of these periods, what are the main causes of it, and also to find a connection with the groundwater levels of the same time period.

ACKNOWLEDGEMENTS

The research was carried out in the framework of the GINOP-2.3.2-15-2016-00010 'Development of enhanced engineering methods with the aim at utilization of subterranean energy resources' project of the Research Institute of Applied Earth Sciences of the University of Miskolc in the framework of the Széchenyi 2020 Plan, funded by the European Union, co-financed by the European Structural and Investment Funds.

REFERENCES

HASTENRATH, S. L. (1967): Fourier Analysis of Central American Rainfall, Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B, vol. 16. No. 1. 81-94.

HMS – Hungarian Meteorological Service Online Database, Available from: http://met.hu/eghajlat/magyarorszag_eghajlata/eghajlati_adatsorok/ [1. June 2015.]

ILYÉS, Cs., TURAI, E., SZŰCS, P., & ZSUGA, J. (2017): Examination of the cyclic properties of 110-year-long precipitation time series. *Acta Montanistica Slovaca* 22, 1-11

KOVÁCS, J., KISZELY-PERES, B., SZALAI, J. & KOVÁCSNÉ SZÉKELY, I. (2010): Periodicity in Shallow Groundwater Level Fluctuation Time Series on the Trans-Tisza Region, Hungary. *Acta geographica ac geologica et meteorologica Debrecina* vol. 4-5, 65-70

KOVÁCS, F. & TURAI, E. (2014): Cyclic Variation in the Precipitation conditions of the Mátra-Bükkalja Region and the Development of a Prognosis Method, *ARPN Journal of Science and Technology*, vol. 4. No.8. 526-540.

LIU, D., FU, Q., MA, Y. & SUN, A. (2009): Annual Precipitation Series Wavelet Analysis of Well-irrigation Area in Sanjiang Plain, *IFIP International Federation for Information Processing, Volume 293, Computer and Computing Technologies in Agriculture II*, 1, 563–572.

MESKÓ, A. (1984): Digital Filtering Applications in Geophysical Exploration for Oil, Akadémia Press, Budapest

NASON, G. P. & von SACHS, R. (1999): Wavelets in time-series analysis, *Philosophical Transactions of the Royal Society A-mathematical Physical and Engineering Sciences*, vol. 357, 2511-2526

SANGDAN Kim (2004): Wavelet Analysis of Precipitation Variability in Northern California, U.S.A., *KSCE Journal of Civil Engineering*, vol 8, No. 4. 471-477.

ZHENG, W., SHI, S. & GONG, Z. (2014): Evolution of Growing Season Precipitation Series in the West Region of Heilongjiang Province Based on Wavelet Analysis, *Computer and Computing Technologies in Agriculture VII*, Volume 419 of the series *IFIP Advances in Information and Communication Technology* 25-31.

Regional climate modelling with special focus on the precipitation-related fine scale processes

Tímea Kalmár^{1*}, Ildikó Pieczka¹, Rita Pongrácz^{1,2}

¹ Eötvös Loránd University, Department of Meteorology, kalmar.d.timea@gmail.com

² Eötvös Loránd University, Faculty of Science, Excellence Center

To understand the climate, climate processes and its changes, climate models can be used. The higher resolution allows regional models to better account for topographic details, while also improving the ability to simulate surface variables such as air temperature and precipitation. In order to quantify the impact of the use of different dynamical cores and parameterization schemes on regional climate model outputs, hindcast experiments have been completed using the Regional Climate Model version 4.5 (RegCM4.5) for a 10-year-long period (1981–1990) for the Carpathian region and its surroundings at 10 km horizontal resolution. This resolution was chosen to allow both hydrostatic and non-hydrostatic approaches in the simulations. Our simulation matrix consists of hydrostatic and non-hydrostatic runs together with the different treatments of moisture, namely, (i) Subgrid Explicit Moisture Scheme (SUBEX) is used to handle non-convective clouds and precipitation resolved by the model, (ii) the new microphysics scheme allows a proper treatment of mixed-phase clouds and a physically more realistic representation of cloud microphysics and precipitation. For validation purpose, RegCM outputs are compared to the homogenized, gridded CarpatClim data. The results show that the outputs of the convection permitting simulation (using non-hydrostatic approach) overestimate the precipitation in the mountainous areas, which is greater than in the simulation using the hydrostatic approach.

Key words: *regional climate model, RegCM4.5, convection permitting model, Carpathian region*

1. INTRODUCTION

The climate models use physical equations to describe the influences of different factors on the Earth system and the observed climate. The higher resolution of models result in a better representation of land surface heterogeneity and fine-scale forcing, which are important for simulating local and regional climate systems accurately. Climate models are based on the hydro-thermodynamic equations (forming a partial differential equation system) and employ numerical approximation together with parameterization schemes. The hydrostatic approach is often used to simplify the equations since it ignores the vertical components of acceleration and the Coriolis force in the equations of motion. This approach is acceptable on global scale and also, on regional scale with some restrictions. The non-hydrostatic approach is necessary when the grid spacing is 10 km or less.

The climate system includes a variety of physical processes, such as cloud forming, developing processes and radiative processes, which interact with each other on many temporal and spatial scales. Many of these processes are not resolved adequately by the model grid and must therefore be parameterized. Therefore, we carry out a sensitivity analysis of the two cloud schemes (the SUBEX and the new microphysics scheme). The Department of Meteorology at the Eötvös Loránd University has gained an experience of using RegCM (see e.g., Torma et al., 2008; 2011; Pieczka et al., 2016) for about a decade. The current study focuses on the newest model version of RegCM (RegCM4.5), which was adapted with two dynamical cores and a new microphysics scheme for the Carpathian Region with 10 km horizontal resolution.

2. MODEL DESCRIPTION

The model used in this work was developed by the ICTP¹ (Elguindi et al., 2014). RegCM4.5 is based on the dynamics of NCAR² mesoscale model version 5 (MM5; Grell et al., 1994). One of the main improvements in this version is that the

¹ Abdul Salam International Centre for Theoretical Physics

² National Center for Atmospheric Research, USA

model can use a non-hydrostatic dynamical core, which allows for the small horizontal resolutions of the order of a few kilometres. The hydrostatic and non-hydrostatic model dynamic equations and numerical discretisation are described in detail by Grell et al. (1994).

2.1. Subgrid Explicit Moisture Scheme (SUBEX)

In the earlier RegCM versions, the resolved-scale cloud physics are treated by the SUBEX (Pal et al., 2000), which calculates fractional cloud cover as a function of grid point average relative humidity and includes only one prognostic equation for cloud water. Rain is calculated diagnostically and it forms when the in-cloud liquid water exceeds a temperature-dependent threshold.

2.2. The new cloud microphysics scheme

The new parameterization is based on a multi-phase one-moment cloud microphysics scheme built upon the implicit numerical framework recently developed and implemented in the ECMWF³ operational forecasting model (Forbes et al., 2011). The parameterization solves five prognostic equations for water vapour, cloud liquid water, rain, cloud ice, and snow mixing ratios. Compared to the pre-existing SUBEX scheme, it allows a proper treatment of mixed-phase clouds and a more realistic physical representation of cloud microphysics and precipitation (Nogherotto et al., 2016).

3. DESCRIPTION OF MODEL EXPERIMENTS AND VALIDATION DATA

Simulations were carried out for the period 1981–1990 with ICBCs⁴ from the 0.75° horizontal resolution ERA5-Interim data (Berrisford et al., 2011). Main differences between the model experiments are summarized in **Table 1**.

³ European Centre for Medium-Range Weather Forecasts

⁴ Initial and Lateral Boundary Condition

⁵ European Centre for Medium-Range Weather Forecasts Reanalysis

Table 1: *Main properties of simulations*

Experiment	Dynamical core	Microphysics
H_SUBEX	hydrostatic	SUBEX
H_NMIC	hydrostatic	new microphysics
NH_SUBEX	non-hydrostatic	SUBEX
NH_NMIC	non-hydrostatic	new microphysics

For the purpose of validation, we used the corresponding time period from the CarpatClim, which is a high resolution homogeneous gridded database for the Carpathian region with 0.1° horizontal resolution, covering the 1961–2010 period, containing all the major surface meteorological variables (Szalai et al., 2013; Spinoni et al., 2015). Daily temperature and precipitation datasets were separated from the downscaling portal, of which monthly, seasonal, and annual means were calculated for the validation domain and compared to the simulated values.

4. RESULTS

4.1. Temperature

Seasonal temperature bias fields for 1981–1990 are shown in **Figure 1**. The spatially averaged mean temperature bias for summer for the entire CarpatClim domain is around 3 °C in all the four experiments. In the mountainous areas, the signs of the annual and seasonal mean errors change within short distances, especially when the error is close to zero. This can be connected to the fact that the observation network is not so dense in these areas as our grid resolution is. The main differences of the runs are in winter and spring: the H_NMIC reproduces the average temperature better in Hungary (the bias is between – 0.5 °C and 0.5 °C), but it underestimates in the Carpathians (the bias is greater than in Hungary, between –1 °C and –2 °C). In autumn and winter the Lake Balaton appears with a positive temperature bias in all the simulations. The differences between the NH_SUBEX simulation and CarpatClim are between –1 °C and +5 °C. The spatial distributions of the bias fields show that the minimum values occur over the mountainous areas of the domain.

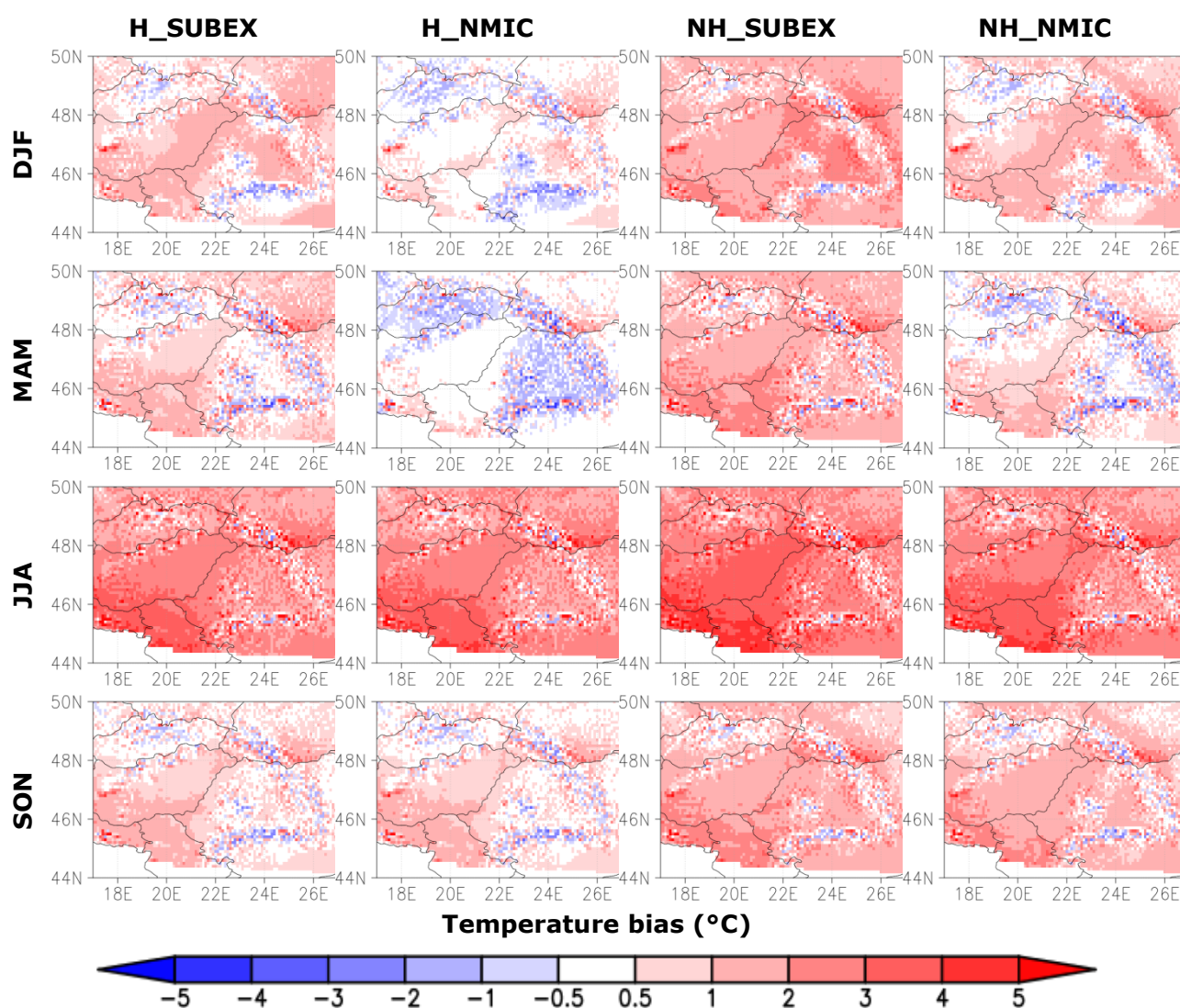


Figure 1: The seasonal average temperature bias of 10-km horizontal resolution RegCM4.5 simulations, 1981–1990. (Validation data: CarpatClim)

4.2. Precipitation

Seasonal precipitation bias fields for 1981–1990 are shown in **Figure 2**. Precipitation evidently depends on elevation, which is the most pronounced in the warm months of the year, especially in summer: the largest biases compared to the CarpatClim are found over the northern-northeastern part of the Carpathian Mountains. In all seasons the precipitation is overestimated over the Carpathian Mountains by as much as 50%. Compared to the CarpatClim data the H_SUBEX has the lowest precipitation mean bias (in summer and autumn only 5%) over the Great Hungarian Plain, except in spring when the largest bias

reaches 30%. Moreover, seasonal mean precipitation bias values for Hungary substantially decreased with the NH_NMIC simulation.

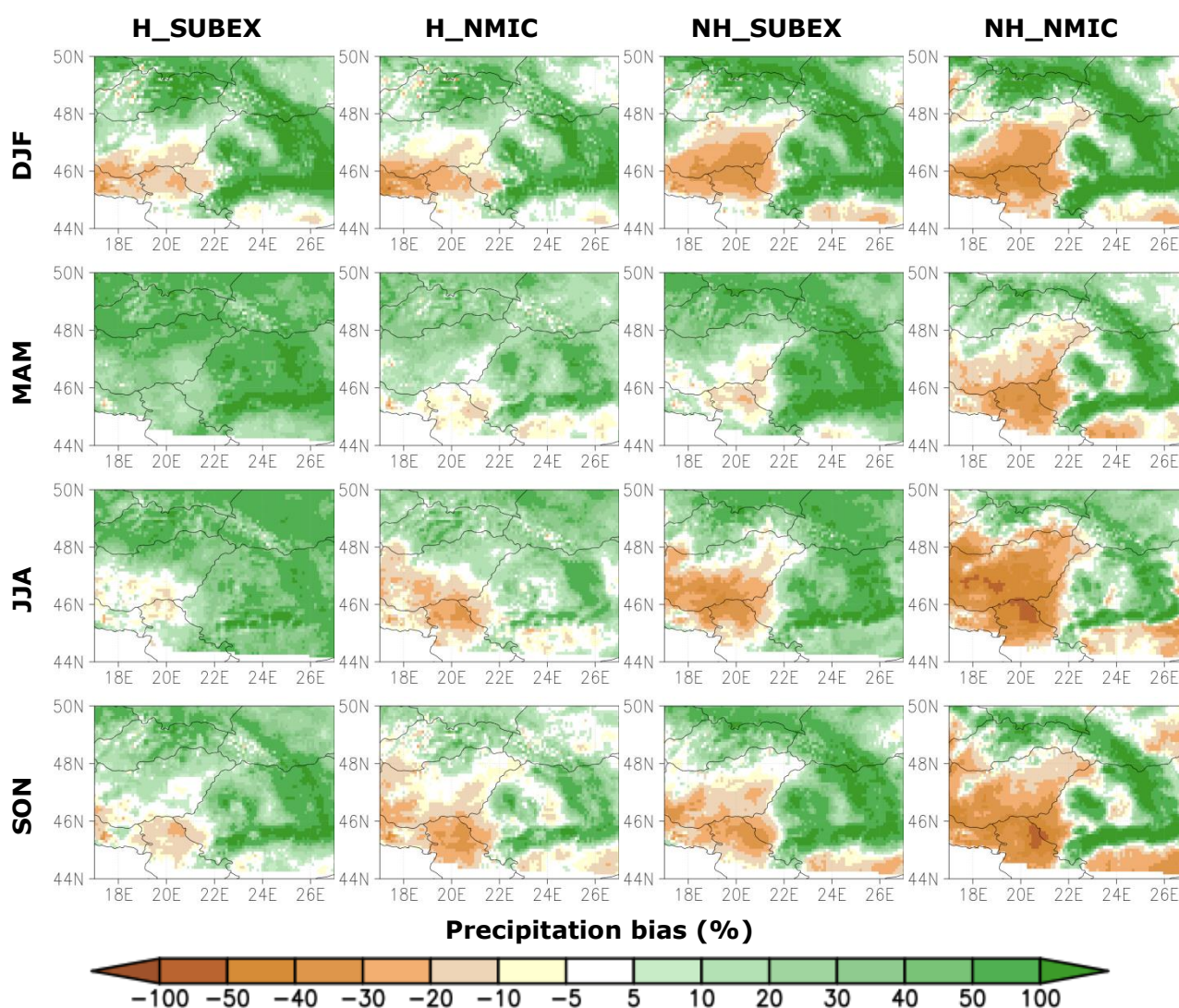


Figure 2: The seasonal average precipitation bias of 10-km horizontal resolution RegCM simulations, 1981–1990. (Validation data: CarpatClim)

5. CONCLUSIONS

We analysed the high-resolution (10-km grid spacing) simulation experiments of the RegCM4.5 for the period 1981–1990 over the Carpathian basin. Our simulation matrix consists of hydrostatic and non-hydrostatic runs together with the different treatments of moisture (the SUBEX and the new microphysics). The largest differences between simulated and observed temperatures occur in

summer. Comparing the simulations, the H_NMIC seems to be the most promising over Hungary; however, it underestimates temperature in the Carpathian Mountains. The largest positive precipitation biases are found over the Carpathian Mountains in all seasons. Negative biases appear over the lower elevated regions.

ACKNOWLEDGEMENTS

Research leading to this study has been supported by the following sources: the EEA Grant HU04 Adaptation to Climate Change Programme (EEA-C13-10), the AGRÁRKLIIMA2 project (VKSZ_12-1-2013-0034), the Széchenyi 2020 programme, the European Regional Development Fund and the Hungarian Government (GINOP-2.3.2-15-2016-00028).

REFERENCES

- BERRISFORD, P., DEE, D., POLI, P., BRUGGE, R., FIELDING, K., FUENTES, M., KALLBERG, P., KOBAYASHI, S., UPPALA, S. & SIMMONS, A. (2011): The ERA-Interim archive Version 2.0, ERA Report Series 1. ECMWF, UK, 27 p.
- ELGUINDI, N., BI, X., GIORGI, F., NAGARAJAN, B., PAL, J., SOLMON, F., RAUSCHER, S., ZAKY, A., O'BRIEN, T., NOGHEROTTO, R. & GIULIANI, G. (2014): Regional climatic model RegCM Reference Manual version 4.5. ITCP, Trieste, 37 p.
- FORBES, R. M., TOMPKINS, A. M. & UNTCH, A. (2011): A new prognostic bulk microphysics scheme for the IFS, Tech. rep., European Centre for Medium-Range Weather Forecasts, 30 p.
- GRELL, G. A., DUDHIA, J. & Stauffer, D. R. (1994): Description of the fifth generation Penn State/NCAR Mesoscale Model (MM5), Technical Report, NCAR, Boulder, Colorado, 128 p.
- NOGHEROTTO, R., TOMPKINS, A. M., GIULIANI, G., COPPOLA, E. & GIORGI, F. (2016): Numerical framework and performance of the new multiple-phase cloud microphysics scheme in RegCM4.5: precipitation, cloud microphysics, and cloud radiative effects. *Geoscientific Model Development*, 9, 7, 2533-2547.

- PAL, J.S., SMALL, E. & ELTHAIR, E. (2000): Simulation of regional-scale water and energy budgets — Representation of subgrid cloud and precipitation processes within RegCM. *Journal of Geophysical Research*, 105, 29, 567-594.
- PIECZKA, I., PONGRÁCZ, R., ANDRÉ, K., KELEMEN, F.D. & BARTHOLY, J. (2016): Sensitivity Analysis of Different Parameterization Schemes Using RegCM4.3 for the Carpathian Region. *Theoretical and Applied Climatology*, 1-14.
- SPINONI, J., SZALAI, S., SZENTIMREY, T., LAKATOS, M., BIHARI, Z., NAGY, A., NÉMETH, Á., KOVÁCS, T., MIHIC, D., DACIC, M. & PETROVIC, P. (2015): Climate of the Carpathian region in the period 1961–2010: climatologies and trends of 10 variables. *International Journal of Climatology* 35, 1322–1341.
- SZALAI, S., AUER, I., HIEBL, J., MILKOVICH, J., RADIM, T., STEPANEK, P., ZAHRADNICEK, P., BIHARI, Z., LAKATOS, M., SZENTIMREY, T., LIMANOWKA, D., KILAR, P., CHEVAL, S., DEAK, Gy., MIHIC, D., ANTOLOVIC, I., MIHAJLOVIC, V., NEJEDLIK, P., STASTNY, P., MIKULOVA, K., NABYVANETS, I., SKYRYK, O., KRAKOVSKAYA, S., VOGT, J., ANTOFIE, T. & SPINONI, J. (2013): Climate of the Greater Carpathian Region. Final Technical Report. www.carpatclim-eu.org
- TORMA, C., BARTHOLY, J., PONGRÁCZ, R., BARCZA, Z., COPPOLA, E. & GIORGI, F. (2008): Adaptation and validation of the RegCM3 climate model for the Carpathian Basin. *Időjárás*, 112, 3–4, 233–247.
- TORMA, C., COPPOLA, E., GIORGI, F., BARTOLY, J. & PONGRACZ, R. (2011): Validation of a high resolution version of the regional climate model RegCM3 over the Carpathian basin. *Journal of Hydrometeorology*, 12, 84–100.

Sources of uncertainties in climate model results

Péter Szabó¹

¹Hungarian Meteorological Service, szabo.p@met.hu

Climate model results inevitably contain uncertainties, which must be quantified in any sound assessments along with the results. These uncertainties are originating from the natural climate variability as an unavoidable feature of the climate system; the approximate description of physical processes in climate models; and the emission scenarios applied for describing possible future paths of the anthropogenic activity. We quantified these uncertainties separately for Northern and Southern Europe and the Carpathian Basin, as well. We analysed mean temperature and precipitation projections following the methodology of Hawkins and Sutton (2009) with some modifications.

Simulations from the newest coordinated global (CMIP5) dataset carried out with 15 global climate models (GCMs) and two (RCP 4.5 and 8.5) anthropogenic scenarios until 2100 were assessed. We seek for the answer whether the CMIP5 results are leading to different conclusions over Central Europe than the previous outcomes obtained from CMIP3 dataset. Furthermore, 24 regional climate model (RCM) experiments were investigated from the EURO-CORDEX database conducted with 12 RCMs and the same RCP scenarios on 0.44 and 0.11-degree horizontal resolution. We examined the added value of fine-resolution RCMs regarding model uncertainty against their coarser resolution counterparts and the role of internal variability in regional projections.

The current investigation is concentrating on the following specific issues: 1) fraction of total uncertainty on different lead times over the Carpathian Basin compared to projections for Northern and Southern Europe and its winter-summer variation; 2) time horizons when climate change signals exceed their total uncertainty (investigating signal-to-noise ratios) and when future changes are larger than natural variability (calculating time of emergence).

Key words: *climate models results, internal variability, model uncertainty, scenario uncertainty, signal-to-noise ratio*

Part II - Big data in geoscience

Big data in geoscience – are we big enough to use and to understand that properly?

Balázs Székely^{1,2,3}

¹Department of Geophysics and Space Science, Eötvös Loránd University, Budapest, Hungary,

balazs.szekely@ttk.elte.hu

²Interdisziplinäres Ökologisches Zentrum, TU Bergakademie Freiberg, Freiberg, Germany

³Department of Geodesy and Geoinformation, Technische Universität Wien, Vienna, Austria

It is a commonplace to state that the amount of available geoscientific data exploding unprecedentedly. As the observation platforms and sensors are being developed with rapid pace and the sampling rate multiplies almost yearly, beside the amount, the quality of these data is also improving: precision, data density are increasing, too. Besides terrestrial sources, data of other celestial bodies became available in resolutions never dreamt of.

Owing to this evolution the 2.5D/3D mapping of the various surfaces (canopy surface, topography), including their hyperspectral characteristics is getting more accurate and includes more details, better spectral resolution. More and more 3D data are being measured, the 2.5D approach is vanishing out, the microtopographic features are becoming detectable even in outcrops or in large coverages.

This long-awaited positive development has a dark side: the amount of data. Geoscience is also arrived in the BIG DATA era: It is not easy to handle this data mass, even if the hardware also is also developing. The fourth GIS component, the trained human resource is often lacking. New geoscientific models may include now big data, but the level of processing or the complexity of the derivatives did not increase considerably in the last years in geoscience.

This is mostly due to the yet unsatisfying absorption capability of the user community. A bimodality can be observed among geoscientists: big data enthusiasts and colleagues doing research in "conventional way". It seems to be difficult to abridge this gap. Partly this is an age-issue, but not completely. A toy too complex or too big for "the children"? Geoscientists should be sensibilized for the available data, processing technology and data integration opportunities. If this happens, the new geoscientist generation may revolutionize how we look at the topography, the rock surfaces, the rock volumes, their properties and their shaping processes.

Key words: *big data, machine learning, human resource, data integration*

The Repository Information Model (A Dream of the Future)

Zoltan Nagy

¹ Geophysicist, geologist, nagyzoltan@t-online.hu

The Repository Information Modelling (RIM) is simply a mean by which everyone can understand a radioactive waste repository through the use of a digital model. Modelling a radioactive waste repository in digital form enables those who interact with the repository to optimize their actions, resulting in a greater whole life value for the repository.

RIM is a way of working, it is information modelling and information management in a team environment, all team members should be working to the same standards as one another. If the establishment and operation of a radioactive waste repository is a fully collaborative 3D RIM (with all project and repository related information, documentation and data being digital) - RIM creates value from the combined efforts of people, process and technology.

The aim of the presentation is to begin a radioactive waste related industry journey to develop and agree a RIM metadata information framework, so that RIM objects can provide consistent levels of information and allow for measurable output-benefits to be realized from a completed RIM model.

Key words: *radioactive waste repository, information modelling, information management*

Determination of sedimentological processes of a coarse-grained deposits in Buda Hills applying combined cluster and discriminant analysis

Éva Farics^{1*}, Dávid Farics², József Kovács¹, János Haas³

¹Eötvös Loránd University, Department of Physical and Applied Geology; Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary, * eva.gyorfy@gmail.com

² Budapest University of Technology and Economics, Department of Electronics Technology; Egry József u. 18, H-1111 Budapest, Hungary

³ MTA-ELTE Geological, Geophysical and Space Science Research Group, Eötvös Loránd University; Pázmány Péter sétány 1/C, H- 1117 Budapest, Hungary

The Triassic formations are directly overlain by an Upper Eocene succession in Buda Hills, Hungary, usually starts with conglomerate beds. These contain predominantly dolomite clasts, but in many places the volcanic, mostly andesite clasts are also abundant. The main aim of the study is to demonstrate the applicability of a new grouping method Combined Cluster and Discriminant Analysis (CCDA) in facilitating the interpretation of sedimentological processes of this coarse-grained clastic deposits. First commonly used methods for characterization of the studied sections were applied: the size and the composition of the clasts; the roundness (Wadell's and Szádeczky Kardoss' methods) and the sphericity (Riley's method) of dolomite and andesite gravels were determined. For the obtained multivariate data, CCDA was applied to determine homogenous groups of sampling locations based on the quantitative composition of the conglomerate as well as the shape parameters (roundness and sphericity). CCDA combines linear discriminant analysis and cluster analysis and its aim is to find not only similar, but even homogenous groups of sampling locations based on the multivariate samples. The result is a spatial pattern, which shows the relationship of the sampling sites (homogeneity or difference and in particular the rate of the latter). According to our concept, those sampling sites which belong to the same homogenous groups were likely formed under similar geological circumstances and by similar geological processes. Various sedimentological environments has been observed within the Buda Hills area based on the results: alluvial fan, intermittent stream and marine.

Key words: CCDA, homogenous groups, roundness, sedimentological model, Buda Hills

How to Document Scientific Work? Preserving Memories for the Semantic Web

László Sőrés¹

¹MFGI, sores.laszlo@mfgi.hu

Metadata are most often used to support searching for results. Results are usually well documented in reports and publications, but details of the creating process and the background work are often lost, or difficult to find. Such information are often kept only in the memory of the authors. Unfortunately, in a few years' time those memories fade away. The Observations and Measurements metadata standard provide sufficient tools to preserve this kind of knowledge in a standard way.

A conceptual model based on generic patterns has been accepted by the Open Geospatial Consortium as a core standard for documenting observations and measurements in 2005. Since then it has become one of the foundation classes of INSPIRE, the Spatial Information Infrastructure for Europe.

Observations and measurements can be used in a lot broader sense that it is trivial at first sight. Any kind of act when trying to estimate values of properties can be considered an Observation and described by using the standard. No matter if it is a simple field observation by a human, data acquisition by some instrument, a mathematical calculation to create a simulation or a spatial model, the same pattern applies to all. Geoscience deals with endless types of features, geological, geophysical, geochemical methods and data types. All this can be hosted in one relatively simple database. However, the complexity of the “problem behind” requires intelligent dictionaries that are available as hierarchical semantic content on the web. The common concept and common dictionaries make it possible to connect logically and physically separates resources and to reach a higher level of data integration.

After a short introduction to the Observations and Measurements standard several applications from different disciplines are going to be presented.

Key words: *observation and measurement, sampling, inspire, semantic web*

Statistic evaluation of digitized geophysical well logs data. (Digitized logs, Statistical evaluation, Multi-correlation)

Laszlo G. Somos

More than 300 coal exploration holes were drilled during the last 60 years in the Eastern-Mecsek Coal Basin. The correlation of different coal seams presented the main challenge in the interpretation of the geological structure of the basin. The existence of the digitized geophysical well logs gave us a slight outlook to find concrete statistical parameters about different rocks. Archived geophysical curves were digitized by the author. The conversion of the obtained digits with rocks formed the main problem in our work. The main task consists on obtaining generalized approach of the morphologic structure of Mecsek coal basin. The main aspects of the geological structure of the region, the comparison of the statistic results (main, extreme data, standard deviation) of the coal bed rocks, sediments, and volcanic intrusions, getting functional relations between log types and the list of the calculated results and the probability level of the possible correlations were all considered. As part of our methodology we have selected geophysical log types (“carottage curves”) for statistical evaluations: Self potential, resistivity curves (potential and/or gradient), gamma radiations, gamma – gamma (density) curves.

Results showed that (i) there are different methods to specify the existence of coal layers in drilled core material, (ii) the numbering and identification of the different coal seams gave us dubious results. (Non characteristic averages, high standard deviation values), (iii) excellent and characteristic results were obtained for the identification of drilled geological formations and (iv) denoting the probability of the existence of graphite traces in contact metamorphic zones.

According to the results and the numbered characteristics of the different rocks (formations) we have received potentially generalized methods for better exploitation of digitized geophysical well logs data.

Key words: *coal mining, geophysics, Mecsek Basin*

Use of Sequential Gaussian Simulation for modelling groundwater pathlines and travel times near an underground radioactive waste repository

Gyula Mező¹

¹ Golder Associates (Magyarország) Zrt., GMezo@golder.com

Underground repositories intended for long term storage of radioactive waste are comprised of engineered and geological barriers to isolate radioactive materials from the biosphere. The geological barrier relies on the ability of the host rock to limit the underground transport of radionuclides. The performance assessment a potential or operating repository must provide estimates of advective transport velocity based on permeability and porosity of the host rock. At the Bataapáti site we have numerous data on rock permeability estimated from single well hydraulic tests performed during both the site characterisation and the repository construction phases. The hydraulic test were performed in individual sections of boreholes isolated from the other sections by single or double packers. The typical length of each test section is about 10 m. The total number of permeability data is more than 1500, however, for the present study we selected 827 data based on reliability and geographical position. The spatial distribution of the permeability data is rather uneven with respect to the area of the site scale groundwater model, which was constructed using the FEFLOW modelling software to forecast the post closure behaviour of the groundwater flow field. In order to populate the groundwater model with permeability data we performed 50 realisations of Sequential Gaussian Simulation of log K, using the open source geostatistical code SGeMS. The resulting grids consist of 96 000 cells, each cell having spatial extent of 20×20×10 m. The data of each individual realisation were used as an input of the site scale groundwater model in order to calculate hydraulic head distribution, velocity field and travel times of water particles from the repository to the biosphere.

Key words: *sequential gaussian simulation, groundwater flow, particle tracking, finite element modelling*

1. INTRODUCTION

The safe management and disposal of low and intermediate level radioactive waste produced by nuclear power plants relies on the concept of multiple acting barriers. The Engineered Barrier System (EBS) is constructed in a carefully selected host rock (Geological Barrier) which can isolate the waste from the biosphere and guarantees significant retardation and dilution of dissolved radionuclides along their pathways from the repository to potential receptors. The National Radioactive Waste Repository (NRWR) was put into operation in 2008 near the village of Bábaapáti, in the Southern Transdanubia region of Hungary (**Figure 3**). The geological environment is composed mainly of igneous sequences of the Mórágý Granite Formation (Benedek & Molnár, 2013).

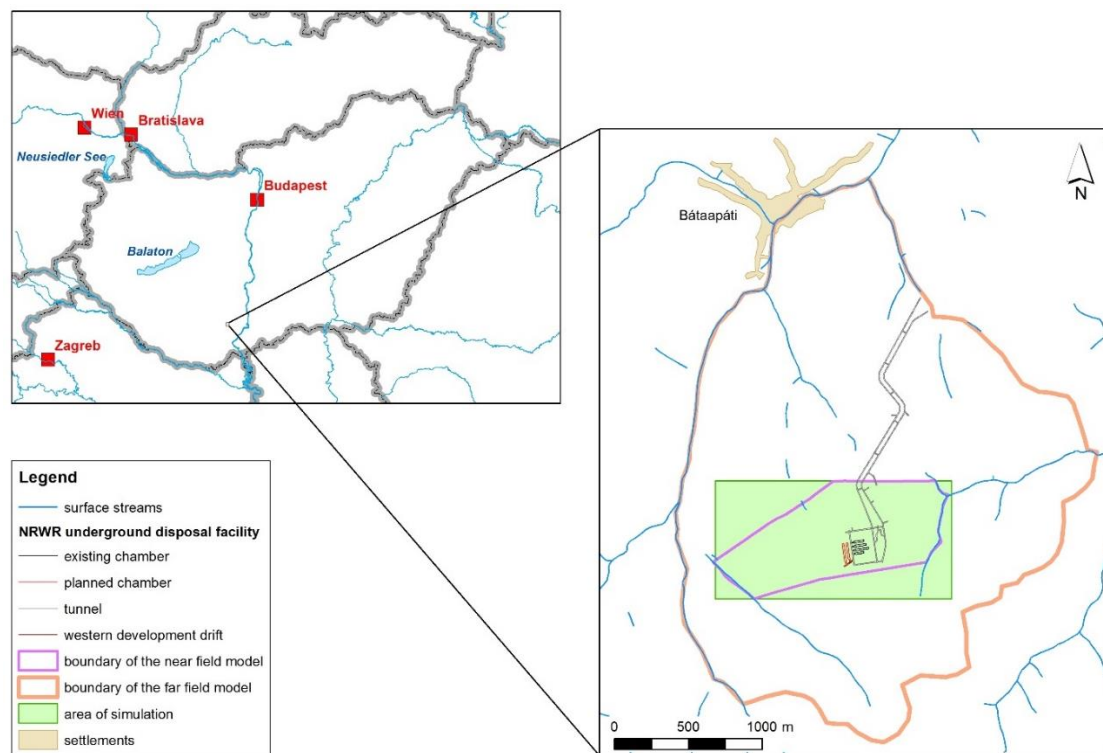


Figure 3: Location of the Bábaapáti site

For groundwater flow modelling through fractured rock formations two different modelling approaches are typically employed (e.g. Benedek et al., 2010): (1) the equivalent continuum (EC) and, (2) discrete fracture network (DFN). The application of the DFN approach to the Bábaapáti site is described in details

elsewhere (e.g. Benedek & Dankó, 2009; Benedek et al., 2010; Benedek & Molnár, 2013), here we only focus on the modelling of the near field groundwater conditions based on the EC approach.

The EC modelling approach was used to calculate travel times and pathline lengths from the repository to the boundary of the geosphere and the biosphere on near field scale and also on far field scale (**Figure 3**).

2. HYDROGEOLOGICAL CONDITIONS

The studied rock formation is subdivided into individual compartments between which the hydraulic connection is very limited (Benedek et al., 2009).

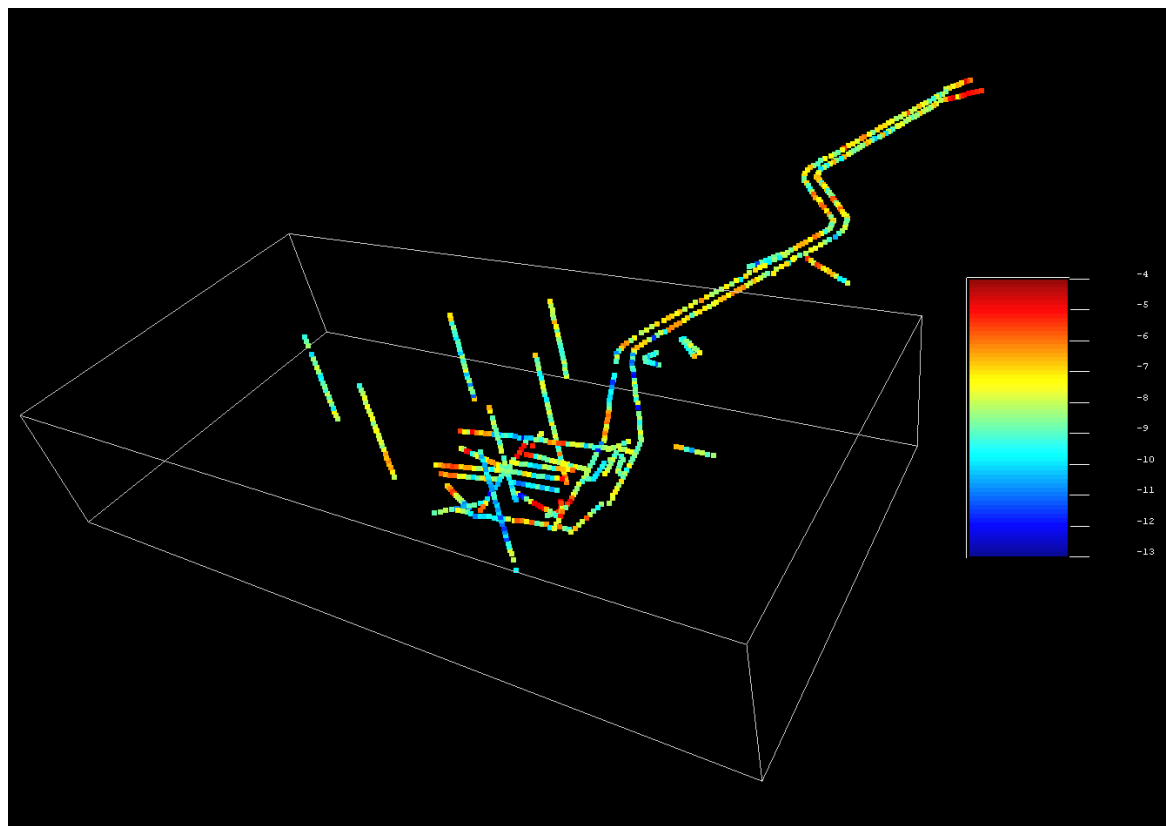


Figure 4: The spatial distribution of log K data and the box region of the Gaussian simulation. The simulation grid consists of $80 \times 40 \times 30 = 96000$ cells in X , Y and Z direction, respectively, each cell having spatial extent of $20 \text{ m} \times 20 \text{ m} \times 10 \text{ m}$.

The boundaries of the compartments are aligned with sheared fault zones with intense mineral alteration and formation of clay minerals (Maros et al., 2010).

Within each compartment the spatial distribution of the hydraulic head is almost uniform, composing a patchwork-like distribution of the heads at larger scale. The relatively uniform head distribution in individual compartments are attributed to the presence of large, well-connected fracture zones having high transmissivity (Benedek & Molnár, 2013). The hydraulic conductivity of the sound granitic rock is known from the interpretation of hydraulic tests conducted in individual boreholes. In each borehole hydraulic tests have been carried out in approximately 10 m long sections isolated by inflatable packers. Since 1996, the beginning of the site investigation more than 1500 values of hydraulic conductivity has been collected. For the present study we selected 827 data having good quality and relevant from geographical point of view (**Figure 4**).

3.GEOSTATISTICAL DATA MODELLING

The hydraulic conductivity (K) spans over several orders of magnitude. The minimum is around 4×10^{-13} m/s and the maximum is 3×10^{-5} m/s (**Figure 5**).

The spatial distribution of log-conductivity data is uneven with respect to the area of the near field flow model as the data are mainly concentrated near the underground tunnels and chambers (**Figure 4**). To feed the near field flow model with hydraulic conductivities we selected the sequential Gaussian simulation method in order to preserve the high variance of the observed values.

For the simulation we used the open-source Stanford Geostatistical Modeling Software (SGeMS), for detailed description of the software see Remy et al. (2011).

Variogram modelling (**Figure 6**) was also done using SGeMS. We found that the combination of a 0.5 valued nugget effect and a spherical variogram having partial sill of 1.5 is suitable. The spherical variogram has the range of 200 m in the vertical direction, 120 m in the 45° direction and 60 m in the 135° direction. The variogram model presented was used for conditioning the simulated values to observed data.

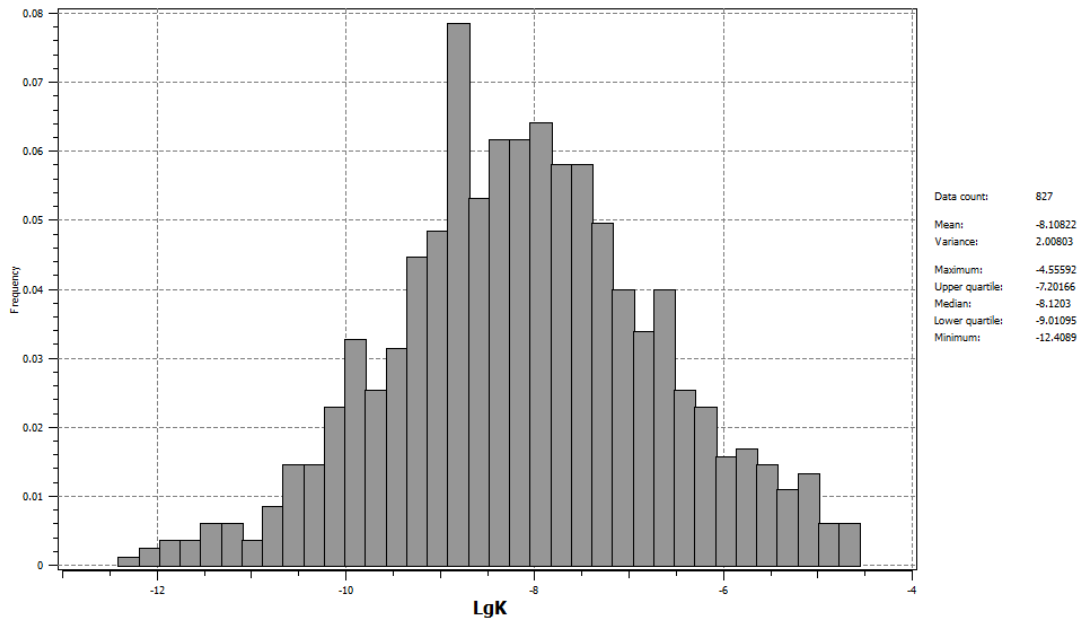


Figure 5: The histogram of the log K data

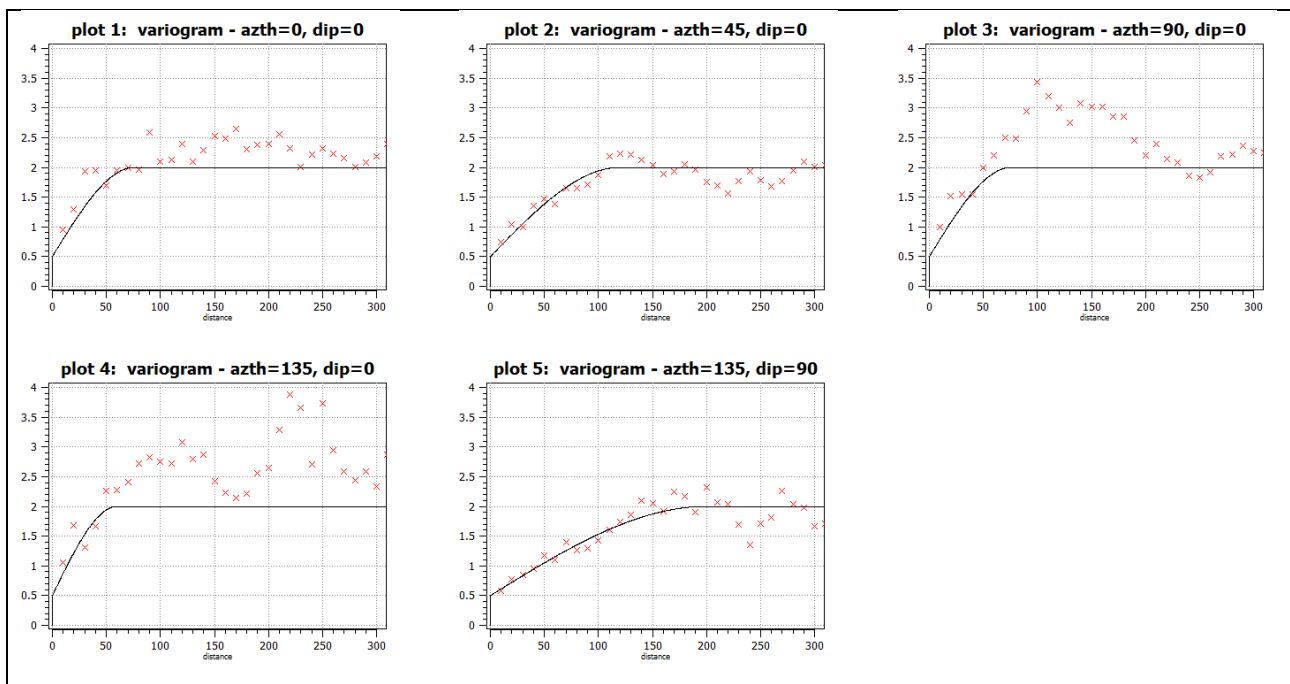


Figure 6: Directional experimental and model variograms of the log K data

We computed fifty realisations. As one may expect, the log K values of the individual realisations show greater spatial variability (**Figure 7**) than that of the mean of log K (**Figure 8**).

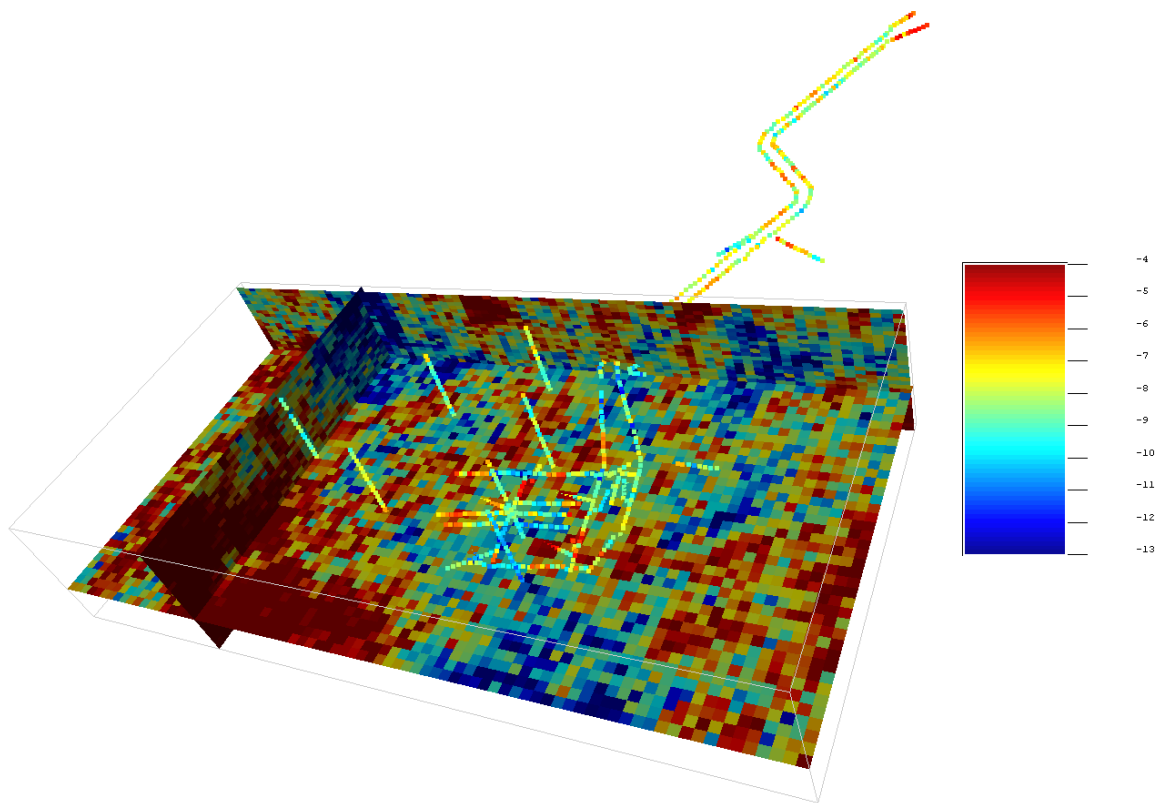


Figure 7: Spatial distribution of the simulated log K values in realisation #20

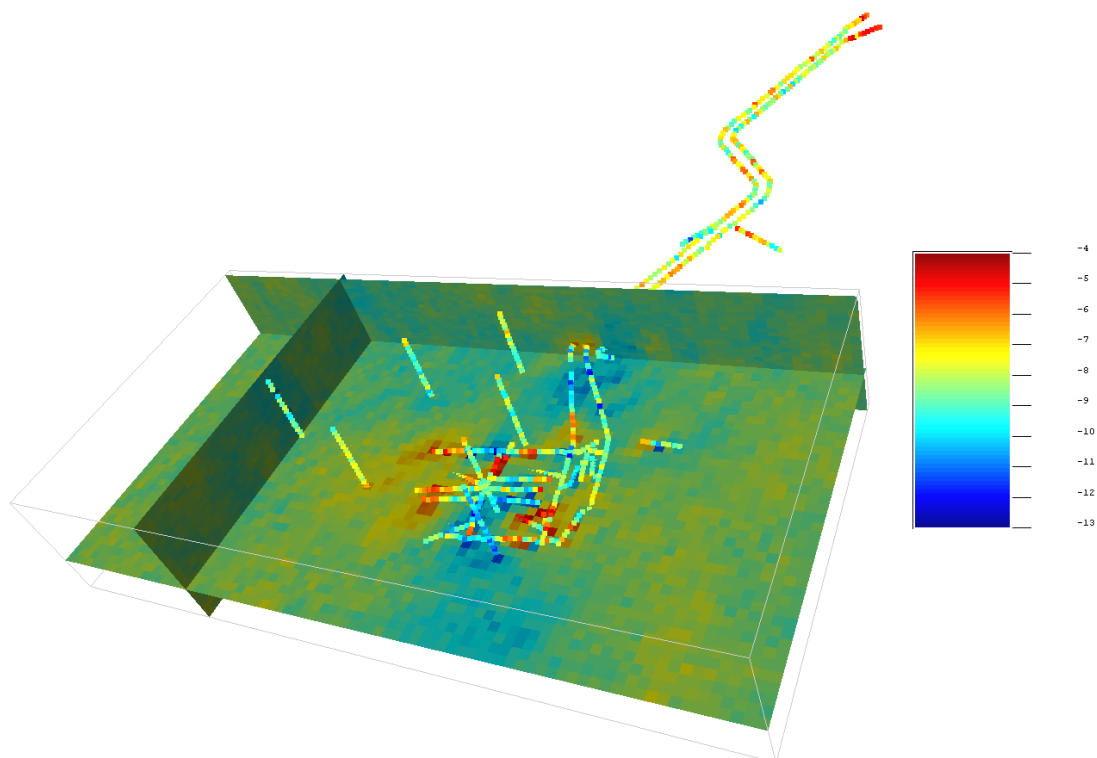


Figure 8: Spatial distribution of the mean of the simulated log K values

In order to preserve the spatial variability of the original data we decided to use the hydraulic conductivities of each individual realisation and calculate the flow field separately, instead of using the mean or other aggregate values.

4.MODELLING THE SUBSURFACE FLOW FIELD

All log K values calculated for each cell were transferred to the near field flow model constructed using FEFLOW®. This way 50 individual near field flow calculations were performed. The boundary conditions as well as the recharge/discharge conditions were derived from the deterministic far field flow model which also was created using the FEFLOW® code. The structure and results of the far field model were published by Benedek et al (2010).

The spatial distribution of the hydraulic head, the Darcy-fluxes and the advective velocity field were calculated for each of the 50 realisations. We also computed the pathlines of water particles starting from the disposal chambers to the model boundary as well as the travel time of each particle (**Figure 8**).

One can only maintain the spatial variability of the observed hydraulic conductivities when using individual realisations of the log-conductivity field for flow modelling. The high variability of the simulated log K values near the disposal chambers results in high variability of the mean flow direction as well. There is significant uncertainty related to our ability to forecast the path and fate of radionuclides potentially released from the repository. In order to reduce this uncertainty integrated, multidisciplinary assessment of data available from different sources (e.g. geology, geophysics, geomechanics, hydrogeology) is advisable.

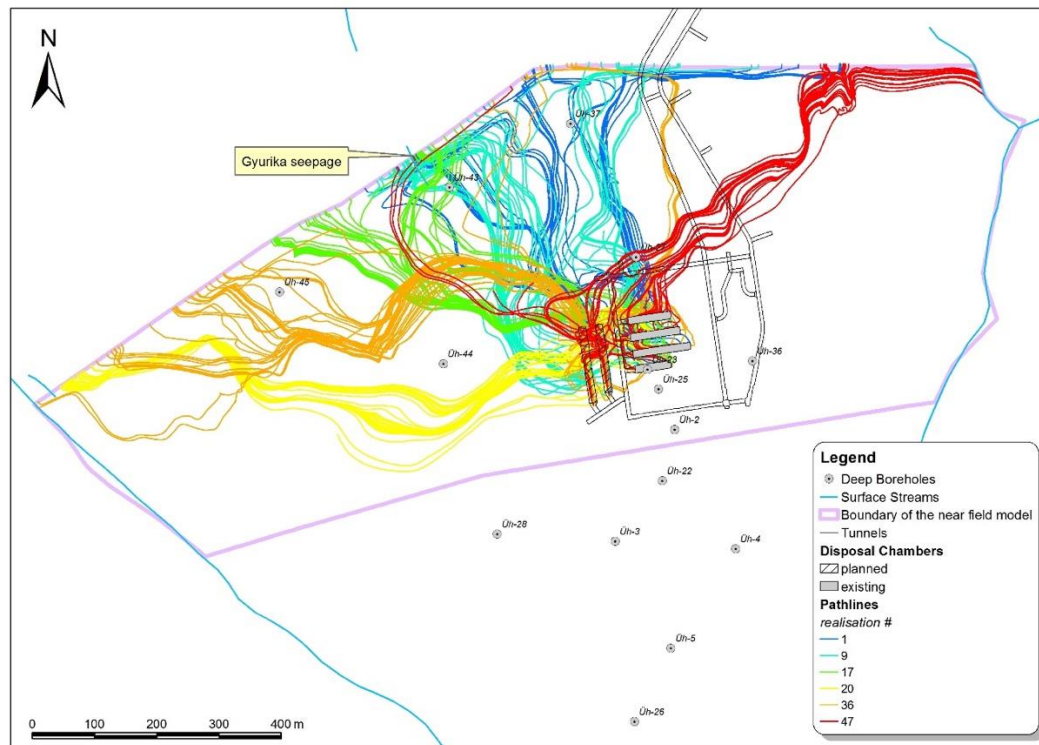


Figure 9: *Computed path lines of water particles starting from the disposal chambers in 6 selected realisations*

REFERENCES

- BENEDEK, K., BÓTHI, Z., MEZŐ, GY. & MOLNÁR, P. (2009): Compartmented flow at the Bátaapáti site in Hungary. *Hydrogeology Journal*, 17, 5, 1219–1232.
- BENEDEK, K. & DANKÓ, GY. (2009): Stochastic hydrogeological modelling of fractured rocks: a generic case study in the Mórággy Granite Formation (South Hungary). *Geologica Carpathica*, 60, 4, 271–281.
- BENEDEK, K., MEZŐ, GY., SZABÓ, ZS., MOLNÁR, P., BÓTHI, Z., SIDLÓ, T. & DANKÓ, GY. (2010): The Hydrogeological Modelling of the Bátaapáti Site. *Annual Report of the Geological Institute of Hungary 2009*, 149–165.
- BENEDEK, K. & MOLNÁR, P. (2013): Combining structural and hydrogeological data: Conceptualization of a fracture system. *Engineering Geology*, 163, 1–10.
- MAROS, GY., KOROKNAI, B., PALOTÁS, K., MUSITZ, B., FÜRI, J., BORSODY, J., KOVÁCS-PÁLFY, P., KÓNYA, P., VICZIÁN, I., BALOGH, K. & PÉCSKAY, Z. (2010): Brittle fault zones in the Mórággy Granite (South Transdanubia): new structural and K-Ar data. *Annual Report of the Geological Institute of Hungary 2009*, pp. 91–112.
- REMY, N., BOUCHER, A. & WU, J. (2011): *Applied Geostatistics with SGeMS: A User's Guide*. Cambridge University Press, Cambridge, 286 p.

Application of 3D geological modelling and grade tonnage calculations on Recsk ore complex

Tamás Miklovicz^{1,2},

¹University of Miskolc, tamas.miklovicz@lapalmacentre.eu

²La Palma Research Centre SL, Garafia, La Palma, Spain

Recsk polymetallic mineralization has a long exploration history including deep/underground drilling/sampling program with laboratory investigation, shaft sinking, geophysical measurements, number of academic/industry papers/reports, etc. These actions resulted a mass amount of data and knowledge on the deposit, however commercial exploitation has not started yet.

Within the scope of a master thesis, the author collected and synthesized many of those results into 3D geological models using Paradigm GOCAD software. Creating 3D visualization of the accumulated knowledge is a new approach and aims to trigger new discussions on the long investigated Recsk deposit.

The work started with collecting of all available information and importing them to a GOCAD project after data cleaning and sorting. The three major study lines were the 3D visualization of the geological structures until 1.2 km; grade tonnage calculations of the deposit; extension the intrusive body until the depth of 2.5 km. The key input for the geological model are the available geological cross sections, geophysical (local/regional gravity, resistivity, magnetic) maps, interpreted seismic sections, surface geology map and Shuttle Radar Tomography Mission data. The key input for the grade tonnage calculations are the geochemical measurements of the 125 Recsk deep drillings. The predictive modelling of the intrusive body is based on the geophysical maps.

As a result, the Eocene and Triassic layers, fault system, intrusive body and its depth extension has been visualized and grade tonnage models for Cu, Au, Zn metals has been calculated. All the results have been exported to commonly used file formats, such .dwg/AutoCAD files, so further work on this dataset can resume anytime. One of the suggested research directions is to investigate the Combined Heat Power and Metal (CHPM) extraction technology potential of the deposit.

Key words: *3D modelling, Recsk, geological model, grade tonnage model*

“Data-Mine” software: Complex earth scientific documentation of excavations with a uniform, real 3D background

László Kovács¹, Ágnes Krupa¹, Roland Schön², Imre Gaburi³

¹Kőmérő(RockStudy) Ltd., Pécs, Hungary, komero@komero.hu

²Smartcode Ltd. Pécs, Hungary,

³Mecsekérc Co., Pécs, Hungary

Detailed knowledge of geological-tectonical, hydrogeological and geotechnical features of host rock is essentially required for the design and safety/performance assessment of a radioactive waste repository. National Radioactive Waste Repository at Bábaapáti is constructed in a highly tectonised granitic formation. Since shotcrete layers are regularly used for supporting, the chance for subsequent data acquisition is limited. Accordingly the in-situ documentation became an indispensable element of each advance of tunnelling. During the earlier research and construction phases works have been performed by two separated expert teams based on different technologies. Geometrical basis of geological-tectonical and hydrogeological documentation provided by a photorobot produces a strongly deformed, not real 3D data-field. Geotechnical characterisation of advances has been performed with stereoscopic JointMetriX3D/ShapeMetriX3D system (JMX/SMX, produced by Austrian 3GSM GmbH), which able to generate real, georeferred 3D images (combining visual and spatial information) of rock surfaces but does not support acceptably other fields of surveying. Moreover, the 3GSM's system cannot create directly usable database for GIS applications. For solving the limitations a new complex system, the “Data-Mine” has been developed. “Data-Mine” is a HTML5/Javascript-based, optimized data-processing and analyzing software, which applies MySQL. The multi-user, user-friendly software works under browsers, so it is accessible from anywhere. The safety of data is guaranteed by passwords, different competency levels and systematic backup. JMX/SMX data files are converted and used as input. The experts can record all the general and field-specific data (together with their GIS-attributes) on the surface of the software. This development created the conditions for the complex earth scientific documentations and for accomplishing a combined, multi-purpose geoinformatical database. The construction of the last two disposal chambers in Bábaapáti was served successfully by the new system. The new software can be suited for any other projects and geological environments.

Key words: *radwaste repository, granitic host rock, complex documentation, geoinformatics*

Part III - Mathematical aspects of reservoir geology

Geomathematical aspects of reservoir and exploration geology

Marko Cvetković¹

¹University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, marko.cvetkovic@rgn.hr

A better understanding of the key elements in the subsurface has been an important factor in determining the successfulness exploration and exploitation of hydrocarbons. Employment of geomathematics in this kind of tasks has grown over time relating to the advancements in tools used for direct and indirect exploration of the subsurface (geophysics) along with computer science. Role of geomathematics today is twofold regarding reservoir modelling and exploration geology. Firstly, the numerous data now acquired with modern geophysical methods can be properly handled only by using geomathematics. Old data from mature fields can now be reinterpreted by using geomathematics either to find new reserves or to increase production. The same principle can be translated into the concept of exploration geology when mature basins are revaluated the use of geomathematics is invaluable as acquiring new data in the initial phase of exploration is not possible (pre-tender evaluation) or not cost effective. A general outline of the employment of deterministic and stochastic methods has been presented in both type of subsurface explorations (local and regional). Positive and negative aspects of the use of geomathematics has been addressed along with case studies from the Pannonian Basin.

Key words: *geomathematics, reservoir geology, hydrocarbons, exploration*

From Russia with love – On production since 1947

István Nemes¹

¹ MOL Group, isnemes@mol.hu

According to the global oil market changes one of the MOL Group's reactions was to take more ownership of its assets by reviewing them one-by-one in-house to cross-check historical values and possible third-party numbers. In line with the strategy one of the greatest challenges was to set up a comprehensive in-house database on one of MOL's largest assets in Russia, the Baitugan field. Historically the modelling, dynamic simulation and forecasting, field development optimization was planned by local, authenticated institutes and service companies (in correspondance to the Russian Federation's law) and MOL provided supervision and decision-making. The changes provided the opportunity and challenge to start an in-house subsurface project aiming to gain a well-established, up-to-date knowledge on the asset and become a more integrated party, as MOL HQ in the daily life of the field.

The interdisciplinary team faced/is facing several unexpected challenges, problems and pitfalls while building the database, quality-checking and interpreting the data with a starting vintage of 1947. To overcome these problems new approaches, “unorthodox” methods were needed while maintain the overall goal to arrive to a reliable output set of volumes and forecasts on time. The presentation would like to highlight some of the focal points during the workflow from the start until the finalization of reference case static model from a practical point of view, rather introducing the big frame than getting lost in the details, but highlighting the most interesting parts.

Key words: *mature field, database, reservoir geology, reservoir modelling*

1. INTRODUCTION

The Baitugan field geographically is located in South-Russia, approximately 1200 km far from Moscow, 200 km from Samara. Geologically, it belongs to the Volga-Ural basin.

The field was discovered and put on production in 1947, 70 years ago. Due to its relatively smaller size compared to neighbouring and other Russian hydrocarbon deposits, appraisal and development activities showed a modest

progress until MOL Plc's farm-in in 2007, gaining 100% equity in the Baituganskoye production block. In 2014, as a result of strategical partnership with Turkish Petroleum (TP), 49% share was sold to this company.

In 2014-2015, the crude prices underwent a significant drop, having a significant negative impact on the global oil industry (**Figure 1**). It was compelling for oil companies' upstream sector to adjust their strategies to the challenging market environment.

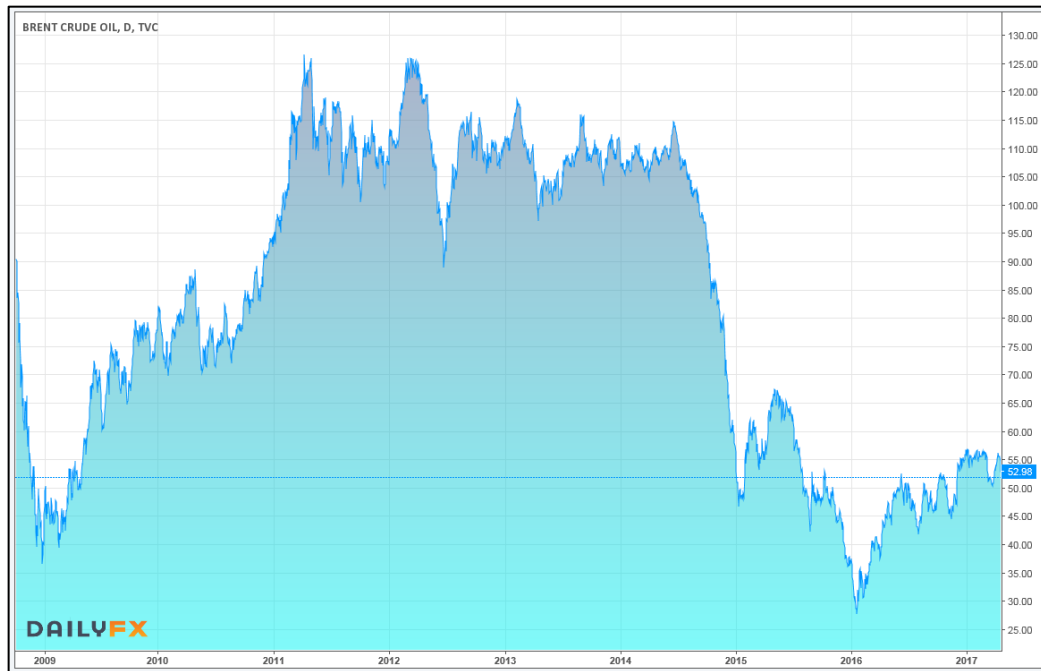


Figure 1: Brent oil prices since 2009 (Source: <https://www.dailyfx.com/crude-oil/>, 22.04.2017)

At MOL Group, the integrated step of the new strategy was a detailed revision of the existing portfolio to minimize risks and maximize value creation.

This company strategy vivified the urgent need of a thorough re-evaluation of Baitugan field's hydrocarbon and further development potential. Since historically a different approach was normally followed, namely supervision of subsurface evaluations and planning, conducted by local institutes (i.e. service companies), the task was huge and highly challenging.

In order to be able to re-estimate the field potential, a standardized, quality-checked and comprehensive subsurface database had to be built and the dataset had to be interpreted with the goal of achieving a well-established understanding of the field.

The key disciplines involved in this multidisciplinary task were geophysics, petrophysics, fluid and core laboratories, sedimentology, geology, reservoir geology and reservoir engineering (including well-testing also).

The optimal solution to meet the goals and deliverables of the work plan was to build a geological model, and based on that, a dynamic model. By appropriate history-matching, which can serve as an effective tool for development planning and estimation of remaining potential.

First, a low-complexity, one-realisation, deterministic geological model was built, and simultaneously preparations were made for a second, detailed and multi-realisation modelling aiming to reflect the actual behaviour of the field and incorporating data and understanding not available at the time of the first model. This paper aims to frame the work conducted during the first modelling job.

2. APPLIED METHODS

The Baitugan field has four productive formations, three carbonate and one clastic intervals. All four were part of the 3D geological model, **Figure 2** shows a schematic section of those, also indicating their depth in meters subsea.

- A4 – Bashkirian Formation (carbonate)
- C1s – Serpukhovian Formation (carbonate)
- Bb – Bobrikovian Formation (clastic)
- V1 – Tournasian Formation (carbonate)

A4 and C1s formations are separated from the Bb and V1 formations by approximately 300 meters of non-permeable layers. But the hydrodynamic communication of the stacked pairs is highly probable even though there is a shaly layer partly separating them (Flowbarrier, Malinov Shale) (**Figure 2**).

The method used was a repeatable, iterative modelling workflow outlined in Roxar's RMS 2013.1 software from data loading to 3D volumetric calculations. The workflow followed a common suit of procedures starting from fault and horizon modelling, structural modelling, building a 3D grid, populating the 3D grid with facies and petrophysical properties and calculating hydrocarbon in-place volumes.

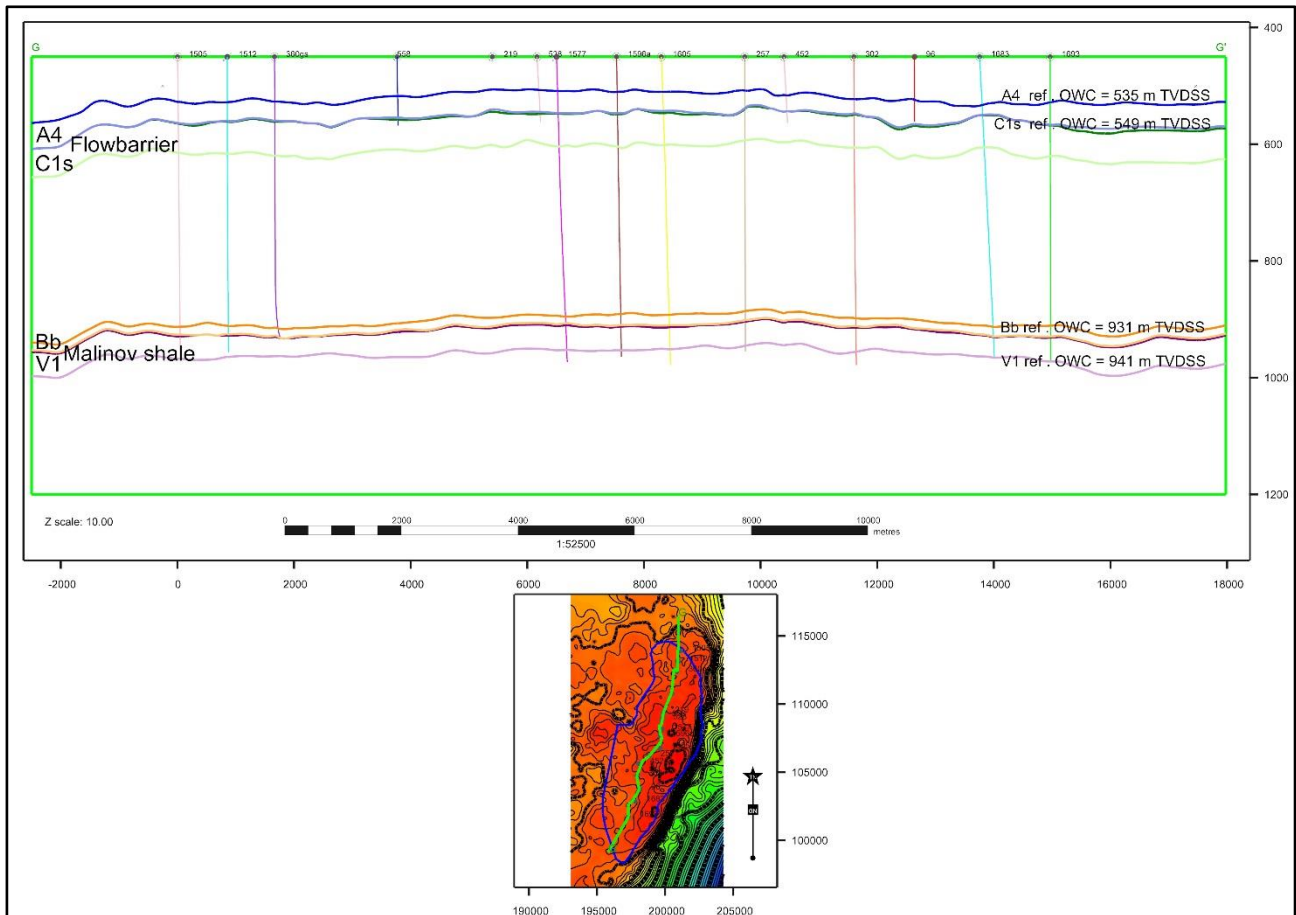


Figure 2: A N-S section showing the productive formations and their relative position

The number of wells penetrating at least one of four formations are shown in **Table 1**. Although it has to be noted, that the drilling of wells and related data acquisition span a period of over 70 years resulting in a highly varying quality level.

Table 1: Wellstock on 01.11.2016

Status of well	Number of wells
Producer	377
Injector	56
Disposal	2
Abandoned	52
Total	487

2.1. Structural modelling

The main input parameters for the structural model were the interpreted seismic horizons (stratigraphic tops and bottoms) and faultsticks. Well picks were filtered and used for adjusting the seismic interpretation to the well data, which were handled as hard data. The structural modelling was conducted in depth domain

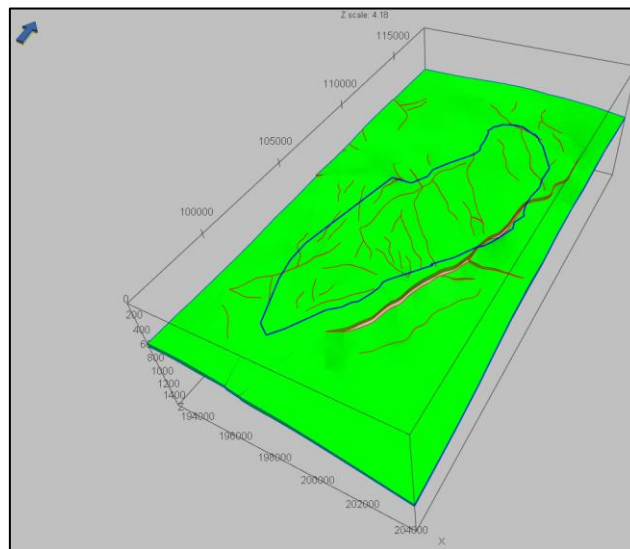
since no time seismic was available; the interpretation was also completed in a depth seismic cube.

Well pick filtering was necessary due to the reliability of input data: in some cases the well trajectory, the well log's depth or the interpretation was questionable.

The mapping increment was 20x20 meters due to the high density of wells (average well spacing is approximately 500 meters).

The oil-water contacts were identified based on formation testing data, but in several cases contradictions were revealed due to differing reasons like measurement quality, cement bond quality, log quality and/or influence of injection or production in the vicinity.

The structural modelling followed three main successive steps: fault modelling, and a nested two step horizon modelling resulting in a structural model as the main input for 3D gridding (**10. Figure 3**).



10. Figure: *The structural model combining seismic and well data*

According to the identified heterogeneity of individual formations, the vertical resolution of the 3D grid was set to 1 meter in the carbonate formations, while 0.4 meter in the clastic one. The horizontal resolution was set to 50 meters in the geological grid, and later, it was upscaled for flow modelling. Gridding was set up taking into consideration the unconformity at the top of Tournasian and

Serpukhovian formations, while stair-stepped fault handling was applied for dynamic modelling.

2.2. Property modelling

In order to synchronize the resolution of the 3D grid and the well data, the necessary well logs were needed to be blocked (upscaled) to the vertical steps of the 3D grid.

The well data used during property modelling included rocktype, interconnected porosity, absolute permeability and water saturation. Latter three were biased to rocktype.

The simplified rocktype model included only two discrete values 0 and 1, namely non-reservoir and reservoir. Cut-off values (interconnected porosity, shale content, water saturation) identified during petrophysical interpretation set the limits between the two types as per formation.

In the first step of property modelling, the 3D grid was populated with rocktypes using vertical proportion curves, variograms in indicator facies simulator (**Figure 4**).

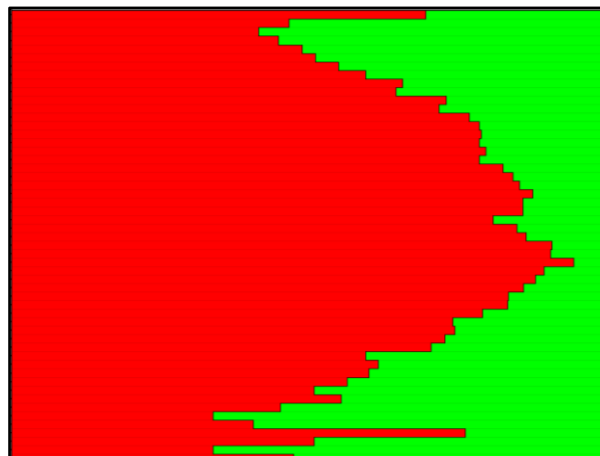


Figure 4: Vertical proportion curve of the Bobrikovian Formation, showing two subunits of the formation (green – reservoir, red – non-reservoir)

Interconnected porosity was modelled conditioned to the two rocktypes, while handling well information as hard data. In order to drive the spatial distribution, variograms were applied that were calculated based on upscaled, transformed data (**Figure 5**).

Permeability was directly derived based on the interconnected porosity using regression equations identified from core data. The petrophysical model followed a single-porosity approach due to the unavailability of any measurements targeting the quantification of (highly) possible fracture system.

Water saturation model was based on well logs, but taking into consideration the original deviation of saturation profile above and below the OWC.

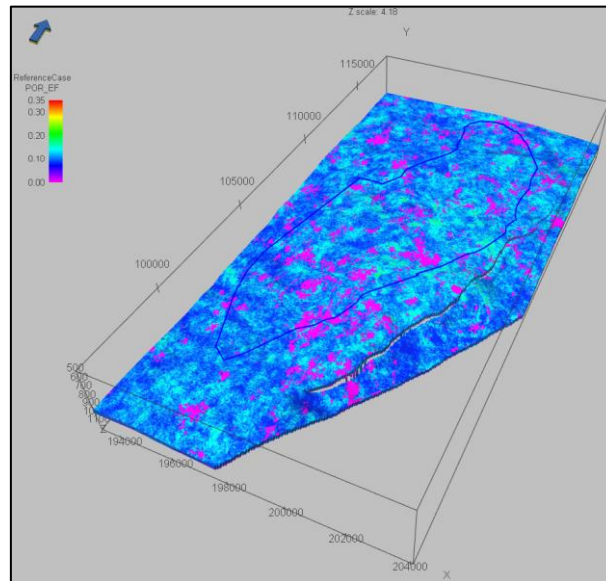


Figure 5: *Interconnected porosity model of the Tournasian Formation*

3. RESULTS AND DISCUSSION

The results and further use of the model and the associated database are varied. The main result of the geological model was that the in-place volumes identified earlier by third party companies could be verified (**Figure 6**). The in-place volumes showed an increase compared to the last evaluation, additionally, satellite structures were identified. The workflow applied during the simplified modelling, revealed several data gaps, contradictions that can be now addressed in data acquisition or re-interpretation campaigns.

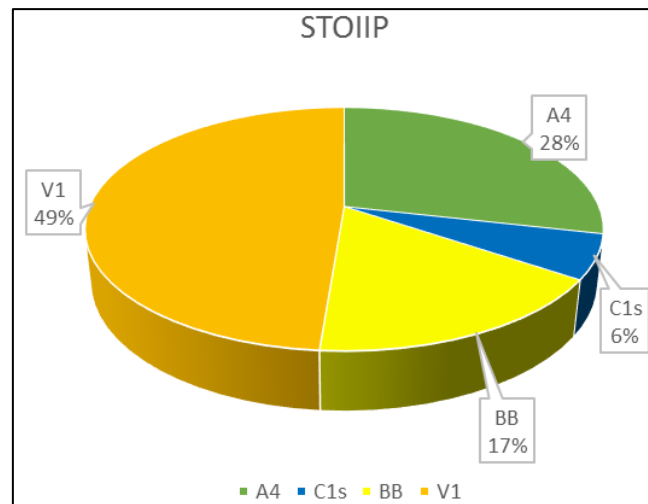


Figure 6: Stock tank oil initially-in-place distribution among the four formations

The model and the underlying database have been used the accomplishment for optimizing drilling target coordinates (in-fill well placement), well work-overs, well treatments (e.g. hydraulic fracturing), as well as a data acquisition program were launched to fill-in data gaps (seismic, petrophysical, pressure, well-test). It helped to identify possible new drilling targets and highlighted a possibility of unconventional appraisal target in the area.

The history-matching was also finished that showed valuable hints on the most uncertain elements of the understanding of the subsurface, as well as spatial uncertainties.

4. CONCLUSIONS

The main conclusion is that this work proved that even a mature field having been on production for 70 years, can have upside potential(s) and re-development opportunities that can yield additional value.

It is also a strong proof that without a well-established, strictly quality-checked database, no useful work and prediction can be made while saving a dollar today can cause a loss of ten dollars tomorrow.

ACKNOWLEDGEMENTS

I wish to express my gratitude to all former and actual colleagues at MOL HQ, Mol-Russ, and Baitex, who participated and contributed to this project in the last years up to now.

Handling a mature clastic HC reservoir without seismic – trends, facies proportions and depositional zones

Szabolcs Borka

University of Szeged, Institute of Geography and Geology, Department of Geology and Paleontology

borka.szabolcs@gmail.com

One can use net-to-gross, lithological or depositional facies to build a static reservoir model. Considering the latter, it is hard to reveal and put them back to our 3D model in the absence of seismic data because of their existing trends and geometries. Thus, we should rely on the well data to define and visualize depositional facies. This article provides a useful method to handle the apparent lateral “blindness”.

The area of the case study is located in the southern part of the Great Hungarian Plain (Algyő field, Hungary). The formation is a submarine fan system. The reservoir is well explored by 132 wells.

Well-logs (GR, RES, SP, Sand- and shale-content, porosity etc.) were used to determine a simplified channel-lobe-background (depositional facies) system. Core samples from 7 wells were also available. In each well the facies proportions were calculated and encoded (0-1) corresponding to the dominant facies in the wells. After applying a simple local B-spline interpolation algorithm on these codes, the facies trends were unfolded as a property map.

The truncated Gaussian simulation was a suitable method to make this information spatially robust regarding facies trends (channel/proximal part or lobe/distal part) and to obtain facies zones. These facies zones were incorporated as trends into a sequential indicator simulation algorithm, thereby establishing a consequent 3D reservoir model. In two parts of the reservoir, channel-dominated zones can be recognized with the surrounding terminal lobes.

Key words: *depositional facies, facies proportion, trend, truncated gaussian simulation, sequential indicator simulation, submarine fan*

1. INTRODUCTION

A static clastic-reservoir facies model can be built in three different ways using net-to-gross, lithological or depositional facies units. Due to its good lateral information richness, seismic data may aid all the three methods. In the case of

the latter it is perhaps the most important soft data because it can show geometries associated with depositional facies. Therefore, in the absence of seismic, we may feel lost regarding any kind of trend, especially if there are huge amount of well data waiting to be processed. A pure interpolated average porosity map can often help to obtain trends, but there are other ways to reveal them. For example, visualization of facies proportions in each well can show a clear picture about the distributions of facies which can be used as trends or guides for facies modelling. The aim of this study is to provide a workflow for depositional facies modelling based on facies proportions.

The case study is located in the southern part of the Great Hungarian Plain (Algyő HC field, Hungary). It is a submarine fan system penetrated by 132 wells with all wells possessing the main well-logs (SP, GR, RES, DEN, CN, CAL, Sand- and Shale-content, porosity, permeability). Core samples from 7 wells were also available to obtain the depositional facies.

The whole workflow was executed in Roxar's IRAP RMS platform (academic license).

2. METHODS

2.1. Defining the depositional facies in the wells

The very first step was to identify the sedimentary facies in a simplified channel (fining-upward sections)-lobe (coarsening-upward sections)-background (fine-grained sections) system in the 132 wells by using mainly the GR, SP, V_{sand} , V_{shale} , porosity and permeability logs. Considering the core samples, the channels (FU sequences) consist of thick (max: 10 m, mean: 4.5 metres (from defined channel facies in the wells) massive, structureless sandstones and they are always followed by thin parallel-, cross-bedded and/or lenticular bedded fine-grained parts. In the case of submarine fans, this kind of sequence refers to a channel-levee system. On the other hand, lobe sections (CU sequences) can be described by also massive sandstones, but they are followed by the sudden appearance of separating shales (mudstones, argillaceous marlstone) with an average thickness of 4 metres (max: 14 m). The background facies is massive or parallel-bedded mudstones or argillaceous marlstones.

Based on these information, the conceptual geological model is a mud-rich submarine fan system with meandering channel-levee systems (mid-fan) and terminal lobes or sand sheets (lower-fan) according to Richards and Bowman's model (1998).

It was quite obvious that in the northern part of the reservoir, both channel-levee systems and terminal lobes were present in a general sequence: the part of lobes are followed by the channel-levee systems. Southwards the latter gradually decreases in thickness, in the southernmost part of the reservoir it is completely missing (pinches out), while the lobes are present in the whole area of the reservoir. This suggests a prograding system towards south (**Figure 1**).

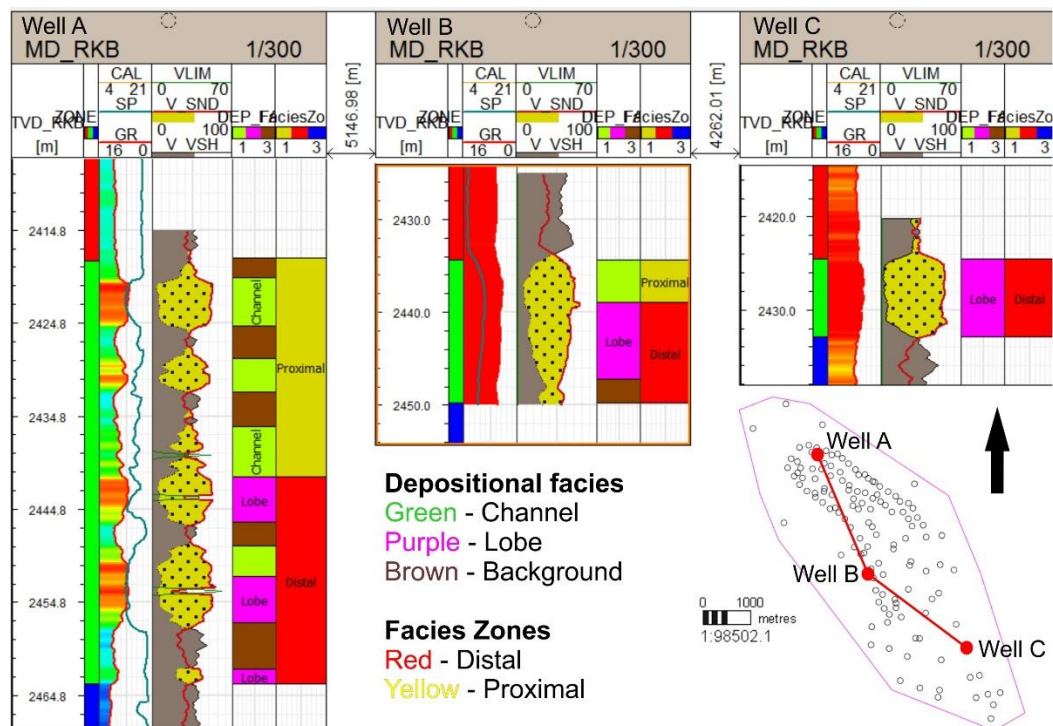


Figure 1: A prograding system seems to appear in the reservoir through 3 wells.

Note that southwards the channel-dominated parts (proximal parts) are absent, only lobe-dominated parts (distal parts) are present

2.2. Establishing facies proportion maps as main guides

After defining the depositional facies in the wells, it was possible to calculate the facies proportions regarding each facies (channel, lobe and background). For instance, in the case of 'channel fraction dominance', if the channel facies

dominated relatively over the two others in a well, then that well was given a parameter '1', else a parameter '0' (the classical 'IF-THAN-ELSE-ENDIF' logical expression). The same method was followed in the matters of the lobe and background too. By applying a simple local B-spline interpolation method of these facies codes, the estimated facies trend property maps were unfolded. Using the well data and these property maps as guides, the so-called 'proximal' (channel-levee dominated) and 'distal' (lobe and sheets dominated) variables were defined in every well (**Figure 1**)

2.3. Creating depositional zones for possible 3D trends and the final facies model

In Roxar's Irap RMS it is possible to use truncated Gaussian simulation ('Facies Belts' module, TGS) to model depositional zones or belts, which interfinger with each other. The results of this module are often used as additional 3D trends for the sequential indicator simulation (SIS).

The defined proximal and distal facies were suitable input data for the TGS to reveal 3D trends for SIS. Two main parameters were needed: the depositional direction and the so-called stacking angle (this describes the average angle of climb of the facies belts within the grid model). The depositional direction could be predicted from the thickness map of the proximal zone. Furthermore, it was possible to calculate dip angles in degrees (from node to node, corresponding to the thickness values) based on the thickness map of the proximal zone. Theoretically, if a zone gradually becomes thinner strictly in one direction (azimuth, in case of the reservoir, north to south), the mean of these angles represents approximately the main stacking angle of the zone, because both the mean angle and the dip angle of the zone as well as the main stacking angle are alternate interior angles. The theoretical background can be seen on **Figure 2**. Using these input parameters, the TGS method was executed (with altogether 50 realizations). The resulting P_{50} (evaluated from all the realizations) 3D model regarding the proximal zone was incorporated into the SIS as a 3D trend for the channel facies.

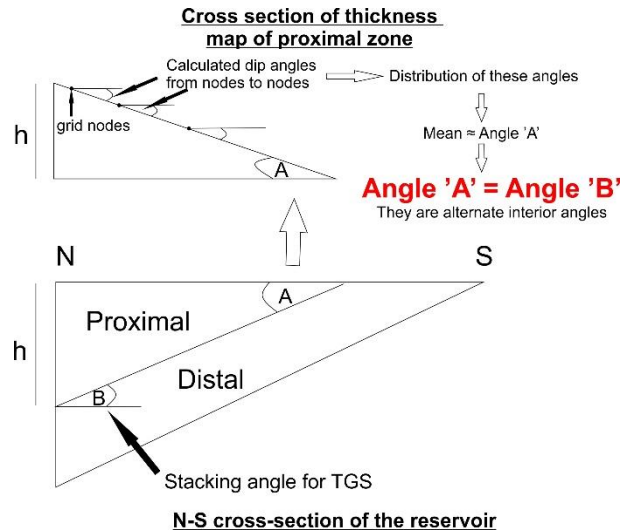


Figure 2: The theoretical background for calculating the main stacking angle in the case of the proximal zone which gradually becomes thinner strictly in one direction.

It is important, that the result of the distal zone was not used as a trend for the lobe facies, as the result pertains to both the terminal lobe and the background facies on the edge of the reservoir, which lacks hard well data (**Figure 1**). Instead, the modelling of the lobe and background facies relied on their global facies proportions coming from the upscaled well data and their semivariograms, respectively. The geological interpretation of the used semivariograms was necessary in the case of every facies (Gringarten and Deutsch, 2001). **Table 1** shows the trend parameters of SIS for the depositional facies model. Using the SIS 100 realizations were generated, and the final facies model was the P₅₀ case computed from the distribution of all realizations.

Table 1: The used trend parameters of depositional facies

Depositional facies	Trend	Semivariogram
Channel	3D parameter trend coming from TGS 'proximal zone'	Yes
Lobe	Global proportion coming from upscaled well data	Yes
Background	Global proportion coming from upscaled well data	Yes

3. RESULTS

After calculating the facies fractions in each well, it was possible to visualize them as pie charts (**Figure 3**). As mentioned before, for overviewing a mature

field it may be better to parametrize regarding each facies in every well. The property maps of the channel and lobe dominance can be seen on **Figure 3**.

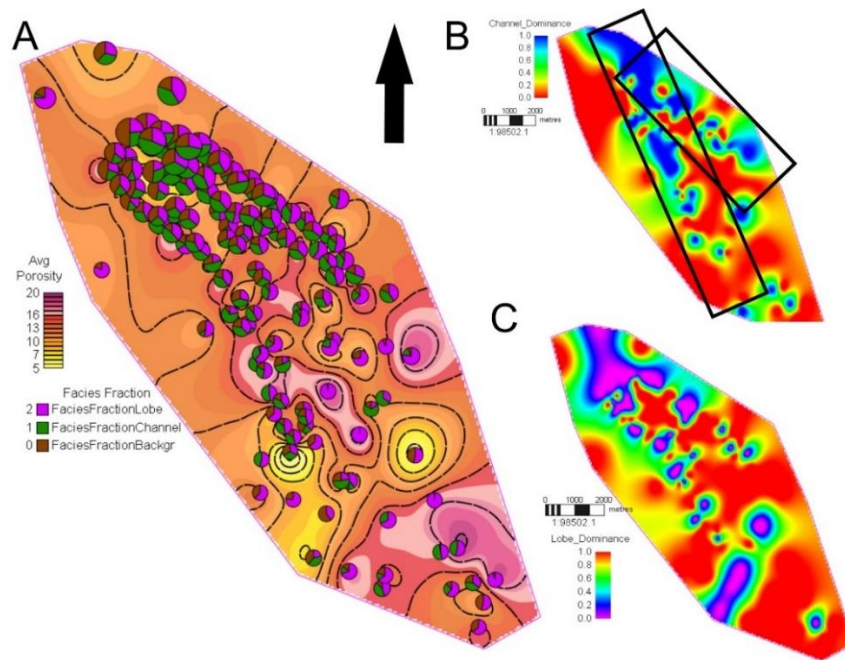


Figure 3: A: average porosity map with the facies pie charts, B: property map of channel dominance, C: property map of lobe dominance

In the case of the channel facies, two sharp zones are recognizable from north to south, and from north to south east.

Using these information, the proximal and distal parts were defined in all wells. The thickness maps of the distal and proximal zones as well as the calculated dip angles regarding a filtered proximal (excluded the parts with zero thickness → sedimentary pinching out) zone can be seen on **Figure 4**.

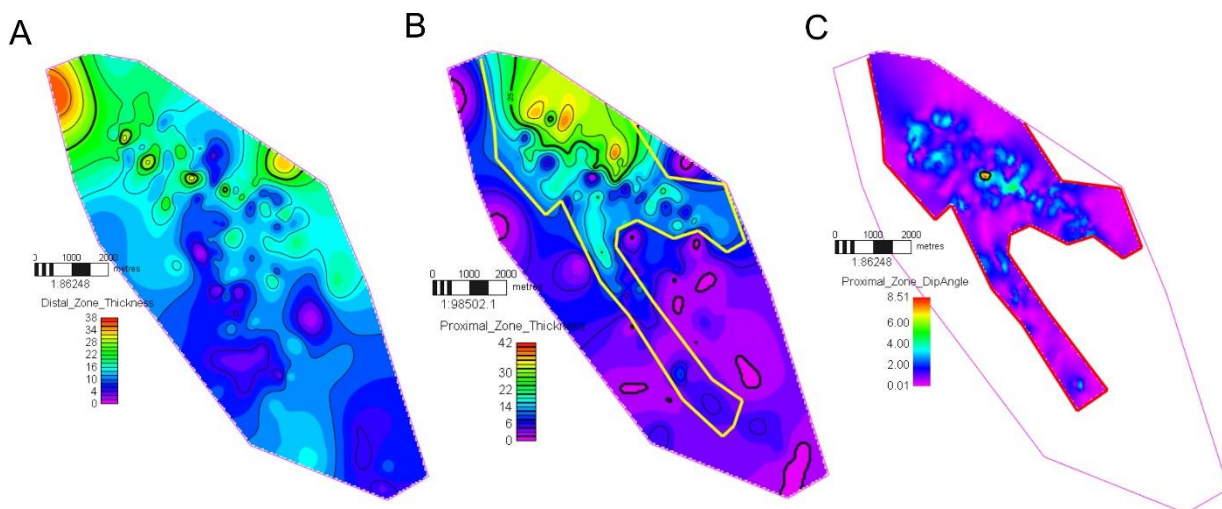


Figure 4: A: thickness map of distal zone, B: thickness map of proximal zone with the filtering polygon, C: dip angles of the filtered proximal zone

The distribution parameters of the calculated dip angles of the filtered proximal zones (main stacking angle) are: mean=0.98, min=0.01, max=8.51, std. dev.=0.79. The depositional direction is about 160 azimuth degree (std. dev.=10). A filtering polygon was also used to denote the facies border between the proximal and distal depositional zones in TGS. As mentioned before, only the P₅₀ of the proximal zone was provided as a 3D trend for the SIS.

Table 2 shows the semivariograms' parameters and interpretations corresponding to each facies in the SIS.

Table 2: The semivariogram parameters and their interpretations (SV: subvariogram, in the case of a nested semivariogram structure; *range: parallel to azimuth/normal to azimuth)

	Channel SV1	Channel SV2	Lobe SV1	Lobe SV2	Background
Type	Spherical	Exponential	Nugget	Exponential	Exponential
Fraction of sill	0.7	0.3	0.25	0.75	1
Azimuth	15/195	150/330	-	130/310	100
Range*	700/500	3500/1600	-	3900/2800	2500/1000
Interpretation	Rapid changes due to the meandering process itself (probably represents a wavelength)	the main channel direction (smoother, persistent change)	-	terminal part of the depositional processes but covers a greater area	terminal part of the depositional processes but covers a greater area

A layer (layer 4) of the final P₅₀ model of the reservoir using SIS (evaluated from the 100 realizations) can be seen on **Figure 5**, compared to the same layer of result of TGS.

It is apparent that the method performed better at revealing the channel compared to the reproduction of lobe features, perhaps due to the lack of a 3D trend of the lobe in the SIS.

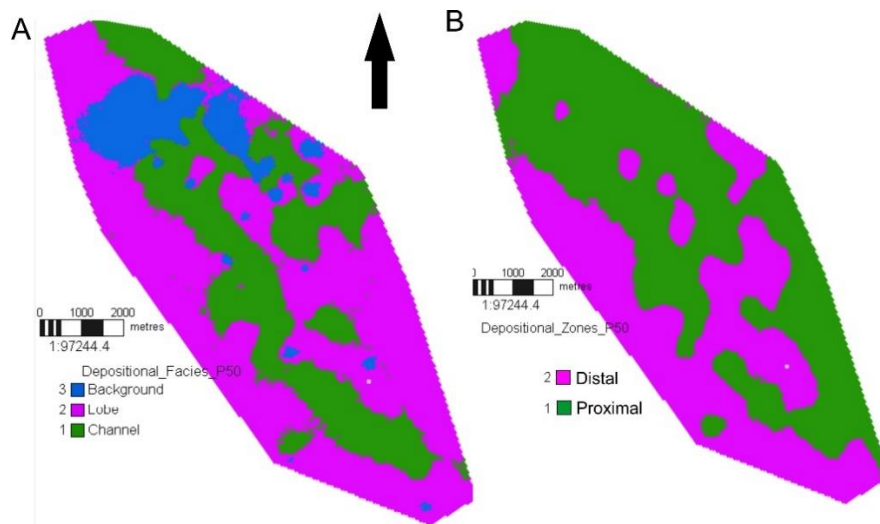


Figure 5: A: P_{50} result of SIS, layer 4; B: P_{50} result of TGS, layer 4; Note that in the case of 'A', two well visible main channels are recognizable

4. CONCLUSIONS

Based on the results, it can be stated that in contrast with the common assumption that pixel-based (e.g. SIS, TGS) methods are not able to reproduce geometrical information of (depositional) facies (Pyrcz and Deutsch, 2014), the use of facies fraction maps and depositional zones can aid these methods to honour geometrical features. Moreover they can solve the lateral 'blindness' caused by the initial confusion related to dense well data and by the lack of seismic data.

ACKNOWLEDGEMENTS

The well logs and core samples were provided by MOL Plc. The academic license of Roxar Irap RMS was provided by Emerson.

REFERENCES

- GRINGARTEN, E. & DEUTSCH, C.V. (2001): Teacher's aide – Variogram interpretation and modelling, *Mathematical Geology*, Vol. 33, No. 4, 507-534.
- PYRCZ, M.J. & DEUTSCH, C.V. (2014): *Geostatistical reservoir modeling – 2nd edition*. Oxford University Print, University of Oxford, 448 p.
- RICHARDS, M. & BOWMAN M., (1998): Submarine fans and related depositional systems II: variability in reservoir architecture and wireline log character, *Marine and Petroleum Geology*, 15, 821-839.

Permeability problematics in modelling of highly heterogeneous reservoirs

János Blahó¹, István Nemes ²,

¹MOL Plc, JBlaho@mol.hu

²MOL Plc

In the modelling process the main challenge is the spatial modelling of permeability. The higher the heterogeneity in hydrocarbon reservoirs the bigger the challenge for the geomodeller. One of the problems is the scale effect. How to give a valid permeability for a grid cell with 50-200 m lateral dimension based on centimeter scaled core samples and on the 10-20 cm vertical length of log-interpreted “average” values. How to validate the permeability of cells? What can be expected, when upscaling for dynamic modelling? What are the roots of calculation, predicted failures?

Overwhelming number of questions without reasonable answers. We can deliver explanations but our tools are poor to reach appropriate solutions. In case of other reservoir parameters analyzing uncertainties is an acceptable result for engineers. However the permeability is not an additive parameter and in heterogeneous reservoirs the statistical variance is high namely the correlation with other parameters is so low that geostatistical methods seem to fail correct modelling, mainly dynamic modelling. Validation of permeability is very important based on well tests result and production experience but in heterogeneous clastic reservoirs and fractured carbonate or metamorphic reservoir it is also a mission impossible.

Some case studies will be described about the challenges of correct permeability interpretation and the spatial modeling of permeability, as well as a number of methods and practices to overcome this ever-lasting phenomena.

Key words: *permeability, modelling, heterogeneous, scale effect, upscaling, validation*

Variogram analysis of well derived lithofacies data in Eastern part of Drava Depression

Ivona Emanović¹, Marko Cvetković²

¹University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, 10 000
Zagreb, Croatia, ivona.emanovic@gmail.com

One of the initial steps for the building of a proper subsurface model for the basin analysis, after solving structural relations, is the determination of lithofacies distribution. This can be either performed as a general mixed lithology based on well log data from a large interval or with a properly-layered model. In the latter, a variogram analysis has to be first employed to determine the spatial relation of lithofacies distribution for each mapped interval. Thus, variogram analysis was performed on five lithostratigraphic intervals in the eastern part of Drava Depression. The analysis was performed on four models of different geometry (cell size and layer number) with the same geographic and stratigraphic extent.

Results of variogram analysis mainly do not show significant differences. In most cases, analysis shows that the major direction of variograms is similar as the “Dinaric” orientation in relation to the paleotransport direction.

Some variograms show different orientation regarding stratigraphic level which could be interpreted to the regional shift in the depositional environments and sediment transport. Four lithofacies models were built upon the results of the variogram analysis which describe the spatial relation of each lithofacies.

Key words: *Croatia, Drava Depression, facies modelling, Pannonian Basin, variogram analysis*

1. INTRODUCTION

Variogram presents one of the basic geostatistical tools. It is usually graphically represented with parameters: lag, sill, nugget effect, distance and range. The Variograms are used for assessing the selected variables in the space and their correlation for subsequent mapping. In this case, variogram analysis was used for facies modelling prior to basin modelling.

2. AREA OF EXPLORATION AND GEOLOGICAL SETTINGS

The entire Drava Depression covers the surface area of about 12 000 km², where approximately 9 100 belong to the Croatia (**Figure 1**). The total thickness can reach more than 7 000 m in the central part of that depression (Malvić & Cvetković, 2013). Infill of the depression can be subdivided into five lithostratigraphic units in rang of formations. These are: Vukovar, Valpovo, Vinkovci, Vera and Vuka Formation.

The oldest Vukovar Formation of approximate Middle Miocene age (Hernitz, 1980; Malvić & Cvetković, 2013), overlies directly on the pre-Neogene basement rocks. It contains of breccia, conglomerate, limestone, sandstone and marl deposits which can be locally organic rich (Zečević et al., 2010).

Valpovo formation consists of limestone as well as clay-enriched limestone. However, it can locally contain large amount of coarse-grained sediments. The deposition was only local according to Hernitz (1980). Upper boundary is defined by E—log border Rs5 which separates the Valpovo Formation from the younger one, Vinkovci Formation. The border between the aforementioned Vukovar Formation and Valpovo Formation is defined by E-log border Rs7.

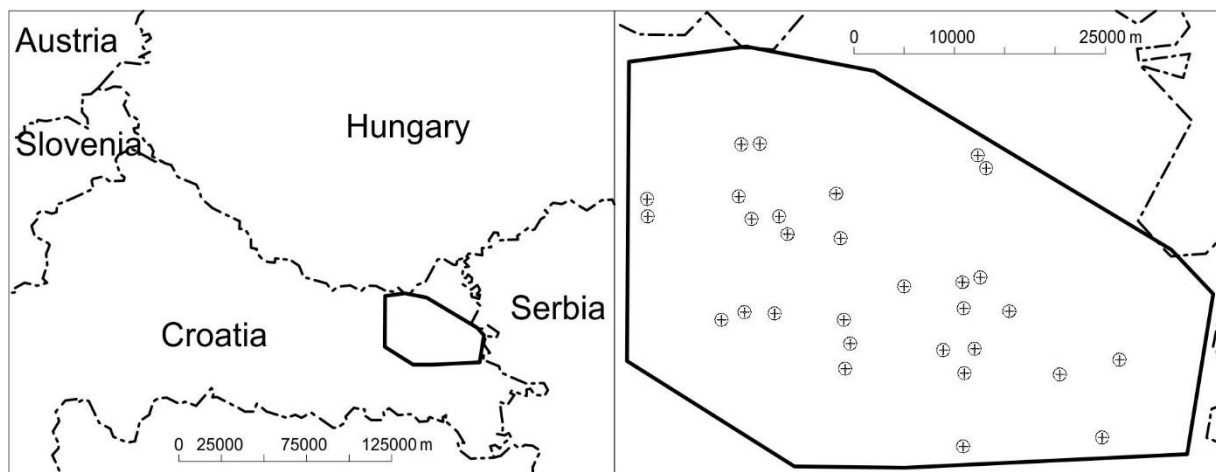


Figure 1: Area within Drava depression used for variogram analysis for further facies modelling

Valpovo Formation overlays Vinkovci Formation, which is made of dominantly sandstone and marl intercalations with sporadic occurrences of source rocks.

Vera Formation has been noted as approximately of Late Pontian age (Hernitz, 1980), and its constituents are marl, sandstones, and coal deposits. The formation is confined with E-log border B in its base and A on top.

Vuka Formation is the youngest formation and it is made of gravel, sand, sandy clay, clay and coal.

3. MATERIALS AND METHODS

Before the Variogram analysis, the structural model had to be prepared along with the defined lithologies within the wells. Surfaces and faults for the model were derived from published structural maps (Hernitz, 1980) which were subsequently digitized and modelled with a 100m cell grid (Stopar, 2015). Lithology was determined based on well log analysis and rock cuttings description.

3.1. Model properties

As the end goal of the research was basin analysis of the selected area (**Figure 1**), several models of different grid spacing and layer numbers were constructed from the original model. The grid spacing was coarser than the initial model (e.g. 200 or 500 m respectively) as basin analysis computation requires a lot of time when number of cells is high in the model. Layering was also observed in two cases – high and low case. Model parameters are shown in **Table 1**. Lithology data was defined in 14 different classes from clastic to carbonate facies.

Table 1: Model properties

Model name	Cell size	Number of cells	Number of sublayers				
			Vuka Fm.	Vera Fm.	Vinkovci Fm.	Valpovo Fm.	Vukovar Fm.
Drava A	200 m	8.3×10^6	15	15	15	7	15
Drava B	200 m	4.2×10^6	30	30	30	15	30
Drava C	500 m	1.3×10^6	15	15	15	7	15
Drava D	500 m	0.7×10^6	30	30	30	15	30

3.2. Variography

A semi-variogram (Nikraves, et al., 2003; Molnar et al., 2010; Hatvani et al., 2017) (**Eq. 1**) is described by three parameters:

- *Range*- defines the lag distance at which the semi-variogram reaches the maximum value;
- *Sill*- the maximum value that the semi-variogram attains at the ‘range’;
- *Nugget Effect*- In theory, semi-variogram should have a value of 0 for original distance (lag distance 0) (**Figure 2**, Nikraves et al., 2003), the amount of nugget defines the initial spatial correlation of the variable.

Matheron algorithm (Matheron, 1965):

$$2\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(u_n) - z(u_n + h)]^2 \quad \text{Eq. 1}$$

$N(h)$ - number of data pairs compared at a distance ‘h’

$z(u_n)$ - values at location ‘h’

$z(u_n+h)$ - values at location ‘ u_n+h ’

The process of choosing the right fit parameters of a semi-variogram that captures the spatial variability of data is subjective as fitting of the optimal one has to be attained through spatial analysis. Once the semi-variogram is fitted on the empirical one, a model describing the functional relationship between the semi-variogram value and the lag distance is selected. The semi-variogram parameters, namely, ‘range’, ‘sill’, and ‘nugget’ are used to define the model (Nikraves et al., 2003).

Almost all the semi-variogram models can be mathematically approximated with many theoretical models, the most frequently used are: Spherical model, Exponential model, Gaussian model (**Figure 3**), Linear model and Logarithmic model.

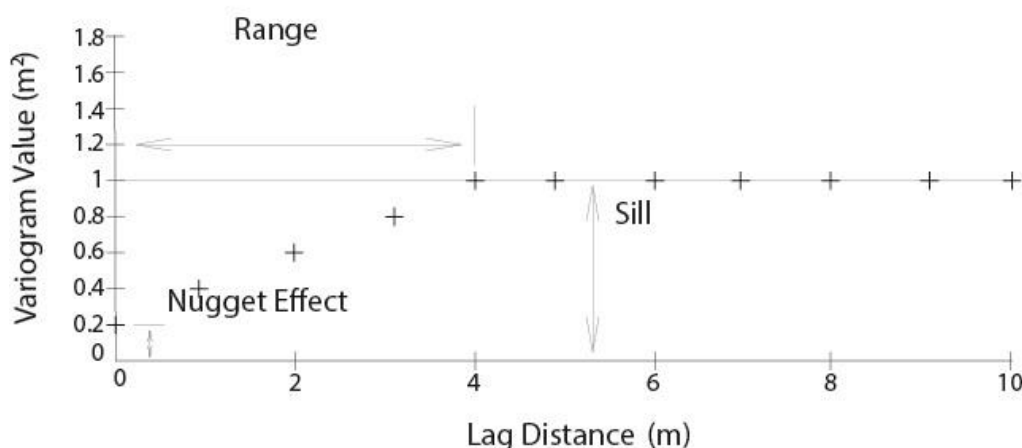


Figure 2: Parameters of variogram (Nikraves et al., 2003)

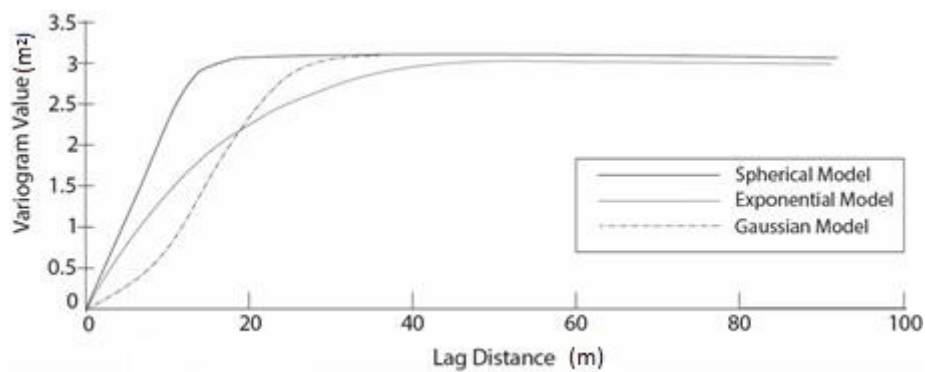


Figure 3: Three popular semi-variogram models (from Nikraves et al., 2003)

4. RESULTS

The variogram analysis produced 20 anisotropic variogram models, for the area of exploration. The analysis was made on the four models with different cell dimension and layer number (**Table 1**). Spherical variogram was used for fitting the variance values of the lithology through the each layer of the model data distribution.

The best attained value of the range has been in high layer models (**Table 2**), except for the variogram made in interval “Base Upper Pannonian–Base Pannonian” in 200x200 model. The range in this model shows the range value of 28,000 m. When the range is reached, the data correlation is minimal (**Figures 4 & 5**).

Table 2: Range of the large axis of the variograms in different models

MODEL 200X200		MODEL 500X500	
RANGE		RANGE	
High layer	Low layer	High layer	Low layer
7372	7163	11496	10832
10440	9698	9727	9109
5366.8	3010.7	5357	5606
4259.9	28000	15536	13093
8596.9	5900	6958	5037

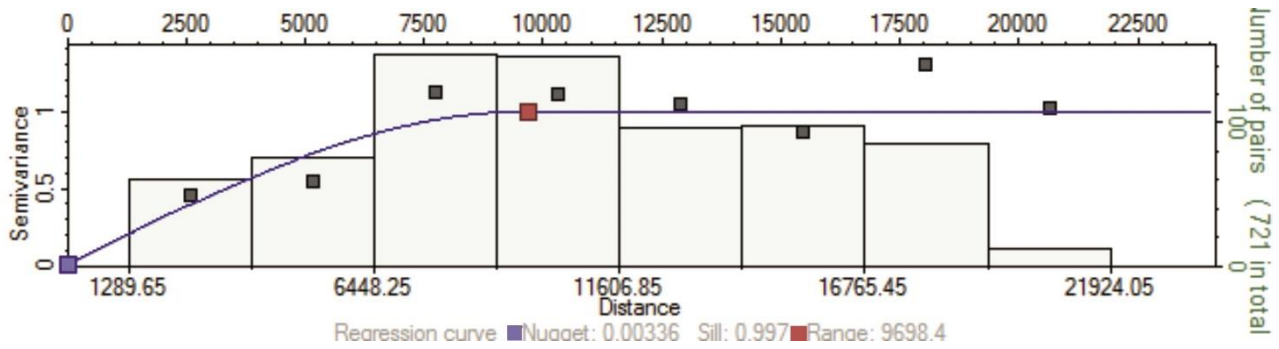


Figure 4: The variogram model of interval of Vera Formation (major direction)

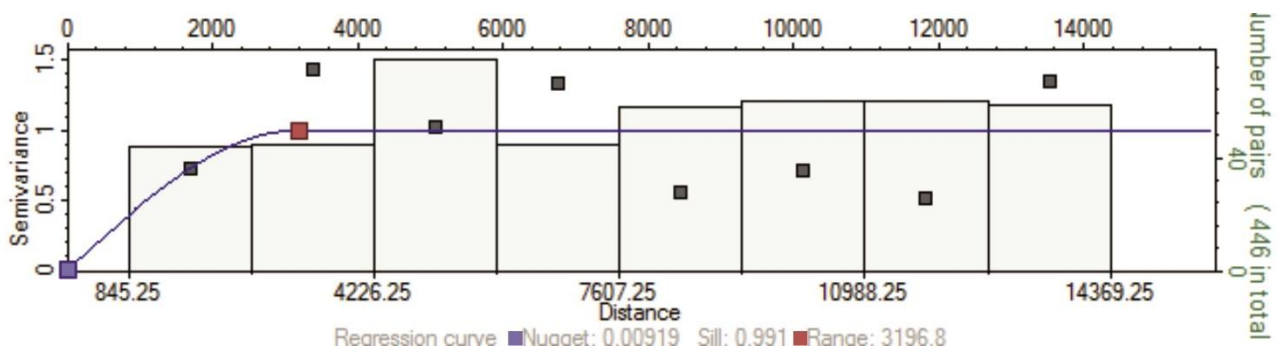


Figure 5: The variogram model of interval of Vera Formation (minor direction)

Optimal variogram was selected on the base of maximum range and minimal nugget value to attain the highest lateral correlation of facies.

Eleven of twenty variogram models show similar large axis orientation to the “Dinaric” orientation (**Figure 6**). Other models show different orientation, which is probably a result of the different transportation regime - local rather than basin wide as in Pliocene Vuka Formation when sedimentary environments were

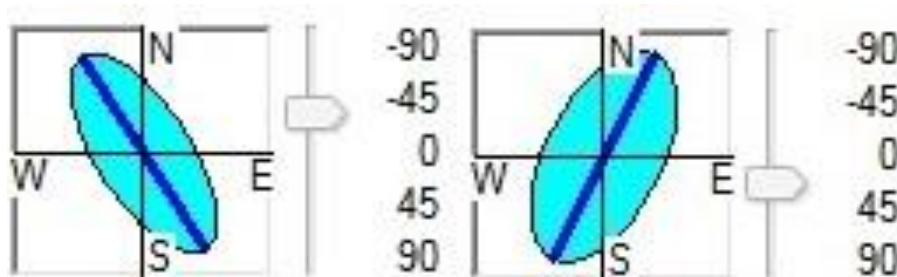


Figure 6: The orientations of variogram models from interval of Valpovo (left) and Vukovar Formation (right)

5. CONCLUSIONS

Variogram analysis was performed for five different lithostratigraphic intervals in the eastern part of Drava Depression. The differences in the orientation of the variograms were probably the result of variable paleotransport direction during the Neogene and tectonic regimes. Based on the results of the analysis four facies models were built (**Figure 7**) which will subsequently be used in basin analysis.

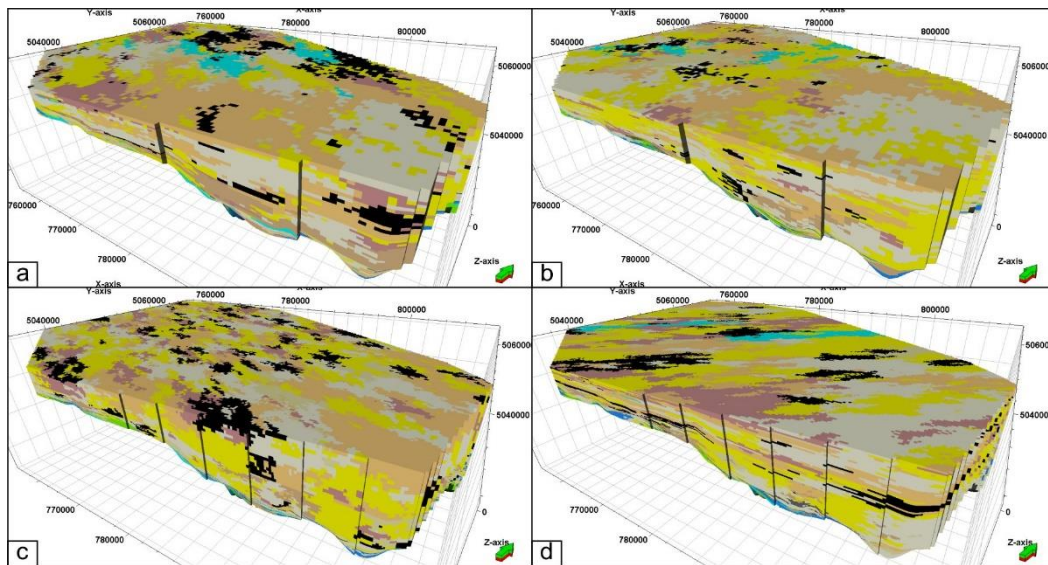


Figure 7: Result of facies modelling in four different models with Gaussian simulations (a-Drava D; b- Drava C; c – Drava B; d – Drava A)

ACKNOWLEDGEMENT

Authors would like to thank the University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering for financial support and the Schlumberger company for donating the Petrel licenses without which these analysis in their extent could not be possible. Authors would also like to than the Ministry of Environment and Energy, Croatian Hydrocarbon Agency and Vermillion Zagreb Exploration d.o.o. for the well data.

REFERENCES

HATVANI, I., LEUENBERGER, M., KOHÁN, B., & KERN, Z. (2017): Geostatistical analysis and isoscape of ice core derived water stable isotope records in an

Antarctic macro region. Polar Science, in press.

<http://doi.org/10.1016/j.polar.2017.04.001>

HERNITZ, Z. (1980): Dubinski strukturno-tektonski odnosi u području istočne Slavonije [*Subsurface structural-tectonic relations in the Eastern Slavonia*]. – Unpub. PhD Thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb, 230 p.

MALVIĆ, T. & CVETKOVIĆ, M. (2013): Lithostratigraphic units in the Drava Depression (Croatian and Hungarian parts)- a correlation. *Nafta*, 64, 1, 34-48.

MATHEON, G., 1965. Les Variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson et Cie, Paris, 307 p

MOLNÁR, S., FUST, A., SZIDAROVSKY, F. & MOLNÁR, M. (2010): Models in Environmental Informatics II. Szent István University, Department of Informatics, Gödöllő, Hungary.

NIKRAVESH, M., AMINZADEH, F. & ZADEH, L.A. (2003): Soft computing and intelligent data analysis in oil exploration, Elsevier, 724 p.

STOPAR, A. (2015): Potpovršinsko modeliranje područja istočnog dijela Dravske depresije [*Subsurface modelling of eastern part of Drava depression*]. – Unpub. Master's Thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb, 36 p.

ZEČEVIĆ, M., VELIĆ, J., SREMAC, J., TROSKOT-ČORBIĆ, T. & GARAŠIĆ, V. (2010): Significance of the Badenian petroleum source rocks from the Krndija Mt. (Pannonian Basin, Croatia). *Geologia Croatica: journal of the Croatian Geological Survey and the Croatian Geological Society*. 63, 2, 225-239.

Evaluation of the trends of hydrocarbon migration processes based on oil density - reservoir depths relationship in Hungary

Zsolt Kovács¹, László Zilahi-Sebess¹

¹Geological and Geophysical Institute of Hungary, kovacs.zsolt@mfgi.hu

Depth of reservoirs of Hungarian oil fields and related oil density data were collected from the database of the Hungarian Mineral Resource Inventory. Our purpose was to point out correlation between the oil density and reservoir depth in some Hungarian hydrocarbon productive region. Oil density related to reservoir depth in a particular area generally refers to the migration mechanism. These mechanisms can be divided into two basic types: the overflow mechanism movement along leading surfaces or semi-permeable slow migration through the seal rock. The latter mechanism is present in the Great Plain formations, while the overflow mechanism is specific for some local reservoir groups in Transdanubia.

For four Transdanubian trends, the deeper the average depth migration is taking place, the stronger the tendency of the inversely proportional change of oil density with depth. If Nagylengyel, Sávolgy, Budafa trends were examined together, we found that generally the lower average density oils occur shallower. This suggests that the primary segregation process was a kind of separation mechanism through a molecular filter in regional scale. The decreasing density with increasing depth occurred only within some local group of reservoirs, and suggest the tertiary migration occurred in a hydraulically connected system.

In the case of Great Plain reservoirs, it seems clear that lighter density oil can be found on top within the majority of trends; however, there is a big difference in the measure of depth dependency. In shallower depth, stronger correlation with depth can be observed. This is presumably related to the inversely proportional depth change to the change of the depth of compaction. Two-narrower steep trend relates to the traditional separation mechanism through a semi-permeable marls taking place in shallower depths.

Key words: *hydrocarbon migration, oil density, Zala Basin, Hungarian Great Plain*

1. INTRODUCTION

Secondary migration is the process by which hydrocarbons are transported from the matured source rock to the places of accumulations (traps). This type of migration is happened through more permeable and porous carrier beds and reservoir rocks, as opposed to primary migration through dense, less permeable and porous source rocks. Variations in crude oil composition are to a certain extent inherited from different source rocks. The primary driving force for migration is the vertical buoyancy force due to the lower density of oil and gas compared to that of formation water, which can be modified by subsurface water potential gradient. Secondary migration terminates in hydrocarbon reservoirs, but tectonic events such as folding, faulting or uplifting may cause redistribution of filled oil or gas pools (tertiary migration) (**Figure 1**, Tissot & Welte, 1984)

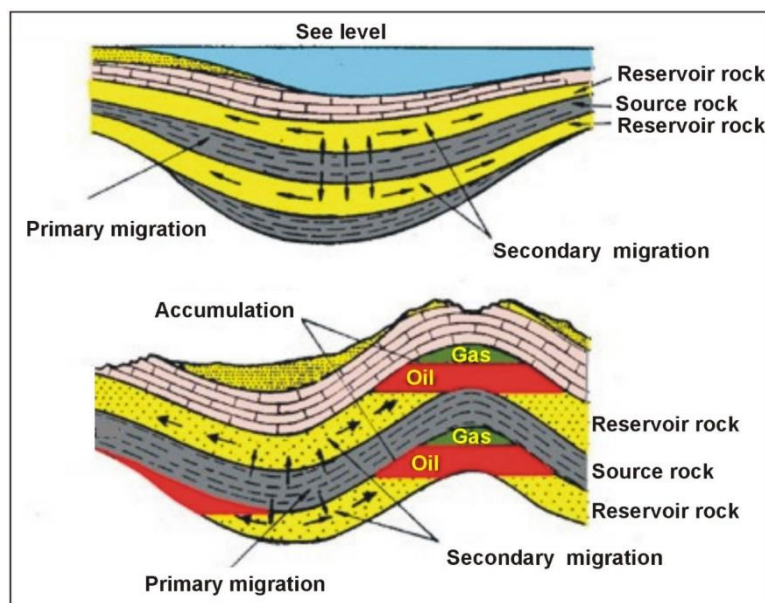


Figure 1: Primary and secondary migration pathways (Tissot & Welte, 1984)

Considering hydrocarbon exploration and production in Hungary, four geographical regions can be distinguished with some smaller units: (1) the Great Hungarian Plain (including Kiskunság, Szeged Basin, Battonya-High, Nagyunság, Hajdúság, Nyírség and Jászság; for unconventional aspects the Makó Trough, Békés Basin and Derecske Trough), (2) the Zala and the Dráva

basin area (Zala Basin, Somogy, Dráva Basin), (3) the Hungarian Paleogene Basin and (4) the Danube Basin (Little Plain) as seen in the **Figure 2**.

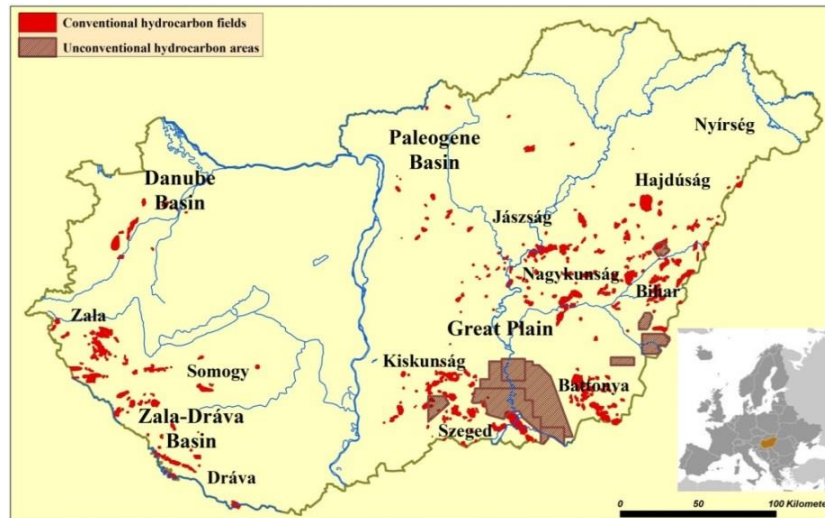


Figure 2: Hydrocarbon fields and the main hydrocarbon exploration and production regions of Hungary

Our purpose was to point out correlation between the oil density and reservoir depth in some Hungarian hydrocarbon productive region. Oil density related to reservoir depth in a particular area generally refers to the migration mechanism. The aim of work was a better understanding of direction of different migration paths, hydrocarbon accumulation possibilities, and hydrostatic-hydrodynamic conditions of the subbasins. We supposed that the expected result would help us in the estimation of potential resources of different regions in Hungary in connection with the designation of new hydrocarbon exploration areas could be marked to concession activity.

2. DATA AND METHODS

Depth of oil-water contact data of oil field reservoirs and related oil density data were collected from the database of the Hungarian Mineral Resource Inventory. The inventory database consist data for 769 reservoirs of 142 oil fields. The initially in place resources of reservoirs are 332 million tons, in which almost 100 million tons were produced in the last 80 years (Kovács et al., 2017). We have collected data for 259 reservoirs in 31 oil fields, which seemed to be suitable for our investigation. The main aspect of the selection was to provide as much data

as possible within a given field. Finally, Zala region had information on 62 reservoirs in seven fields, Szeged-Kiskunság region had 116 reservoirs in nine fields, Nagykunság region had 32 reservoirs on 10 fields, and Bihar region had 49 reservoirs in five fields. Relation of oil density vs. reservoir depths plots can be observed on the **Figures 4-7**.

3. RESULTS

3.1. Zala Basin

The prevailing oil density vs. depth relation in the Zala region - as can be seen on **Figure 3** - is that in general lighter oils are in shallower position, but locally, field by field, the shallower the depth the higher the oil density. The deeper the average depth migration is taking place, the stronger the tendency of the inversely proportional change of oil density with depth. This suggests that the primary segregation process was a kind of separation mechanism according to carbon chain length through the semi-permeable sediments. The increasing density with decreasing depth trends occur only separately in some cases of fields, and suggest tertiary migration mechanisms within an earlier formed reservoir system which hydraulically connected to each other.

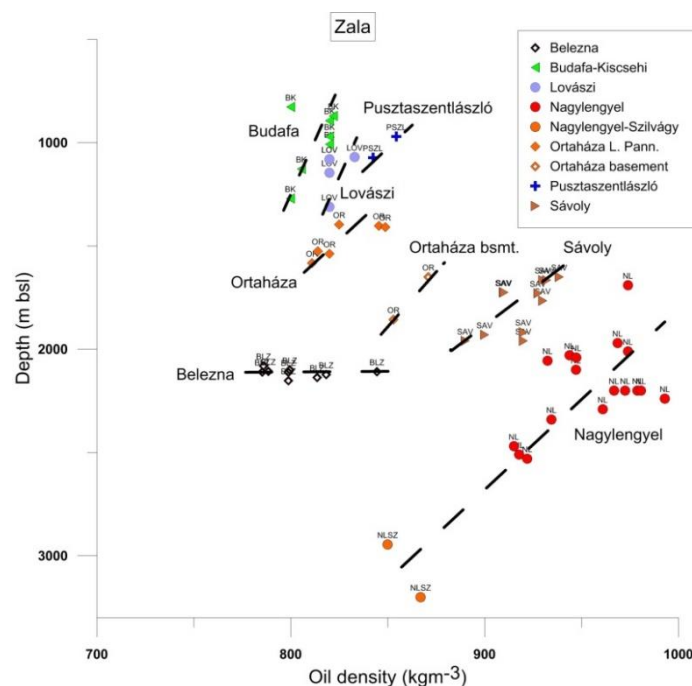


Figure 3: Oil density vs. depth relations in the Zala Basin, Transdanubia

In the case of Zala region this difference may also follows from the fact that the covering seem rocks of reservoirs are more compacted in the Transdanubian region than in the Great Plain because that were buried previously in higher depth.

3.2. Szeged-Kiskunság

In the case of Algyő, Szank and Üllés the specific trend is that the shallower the depth the lower the oil density (**Figure 4**). This may be caused by the separation of oil transported along the fine-grained sediments of carrier bed. Steep trends of Dorozsma, Ruzsa, and Kiskunhalas can be caused by the oil flow along tectonic surfaces. In the opposite trend of Öttömös post trapping alteration (degassing, water washing, biodegradation) of crude oil can be supposed because the relatively near surface position. In some case of Algyő Upper Pannonian reservoirs the dissolved gas contain of oils is much higher than in other oils in this field.

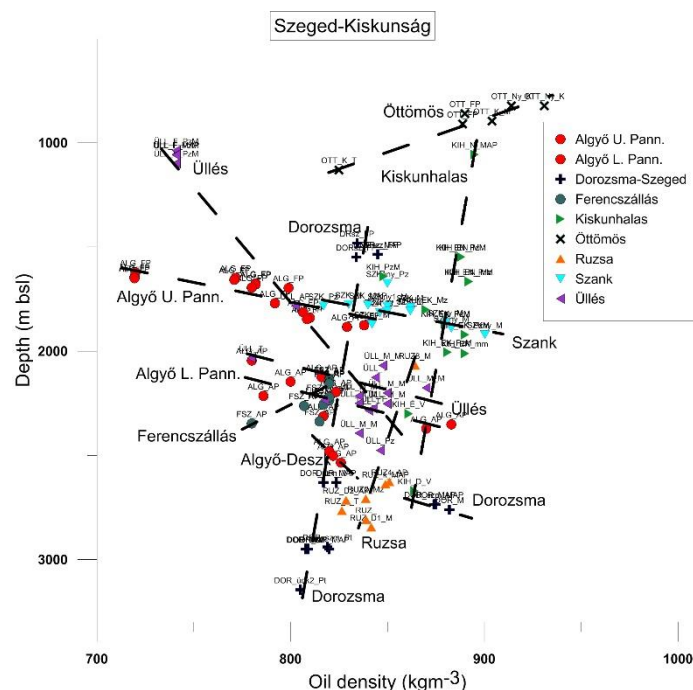


Figure 4: Oil density vs. depth relations in the Szeged-Kiskunság region, Hungarian Great Plain

3.3. Nagykunság

In the deeper parts of the Nagykunság region subbasin the main trend is that, the more dense oils are the higher position (**Figure 5**). This can be caused by degassing, the overflow mechanism of the oil flow, the migration along fault surfaces and possibly biodegradation at the zone above 2000 meters.

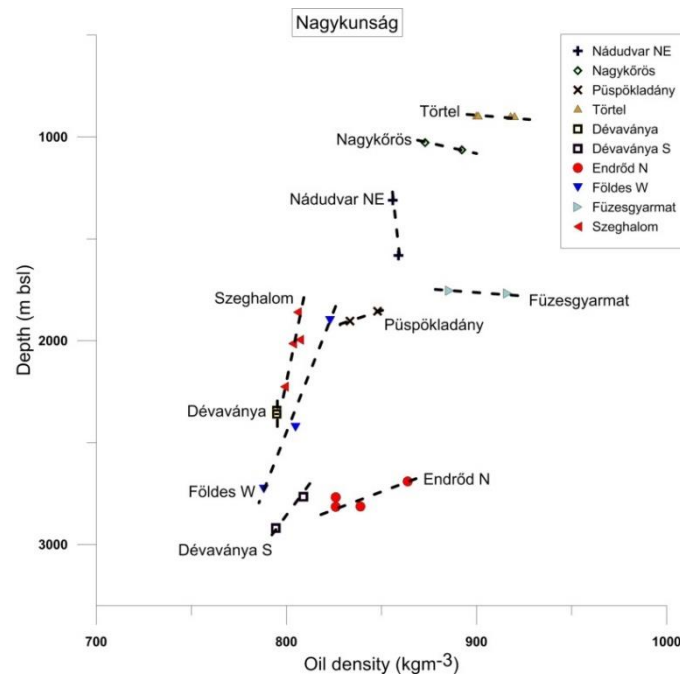


Figure 5: Oil density vs. depth relations in the Nagykunság region, Hungarian Great Plain

In the case of reservoirs in moderate depth an opposite, slightly sloping trend can be seen. It is explicable with low compaction, horizontally interconnected coarser grained carrier bed sandstones, and better porosity-permeability properties.

3.4. Bihar

It seems to be a clear trend in the case of Bihar region of Great Plain, that the higher the average depth the higher the density of oils (**Figure 7**). This may be caused by the braks of HC chains, the separation while migration along the carrier bed in fine grain sandstones against the capillary forces; by the growing of dissolved gas contents, or maybe early oil window generation of oil. In the case of the oil migration from near basement reservoirs of Komádi to shallower

depth, the role of flow along faults can be significant. In Kismarja field case the alteration of trapped oil may be supposed which may be caused by biodegradation or water washing.

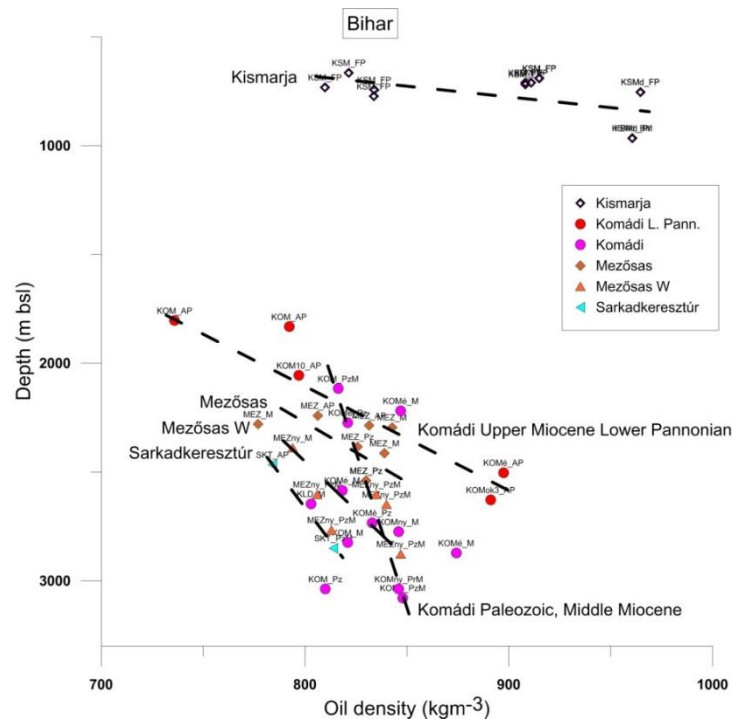


Figure 6: Oil density vs. depth relations in Bihar region, South-East Great Plain

In the depth deeper than 2000 m, the mostly argillitic sediments and randomly deposited turbidite sandstones may not be interconnected to each other, while in the case of Kismarja the relatively younger Late-Miocene Pliocene Upper Pannonian channel fill sandstones are horizontally interconnected. At this shallower depth the rate of compaction is low, and the horizontal groundwater flow effect may be more significant compared to the deeper situated Middle Miocene – Lower Pannonian argillitic sandstones.

4.DISCUSSION & CONCLUSION

The scope of our work was the investigation of trends between the depth of reservoirs and the density of crude oils in four characteristic subbasins, in those hydrocarbon fields when enough sufficient data were available for the study. The Zala Basin is situated in southwest of Hungary, while the other three subbasins are in the area of the Hungarian Great Plain, south and southeast of Hungary.

Zala Basin trends show a different migration process regionally and locally, tertiary migration by overflow mechanism can be supposed in the latter case. In the case of the Szeged-Kiskunság region, locally and regionally, the migration along carrier beds across semipermeable sediments is present, but the role of faults is very significant. In the Nagykunság region the migration processes are similar to those in Zala, but the presence of faults seems more important. In the depth below 2000 m the Bihar region oil vs. depth trends are similar to those of Szeged-Kiskunság region. In the shallower zone the hydrodynamic effects can be recognizable.

Results may be important for the evaluation of hydrocarbon potential estimation studies in the Geological and Geophysical Institute of Hungary. The knowledge of the hydrocarbon migration process, the trend evaluation results help us in the understanding the situation of different places of HC accumulations along the migration routes and the recognition of the different hydrostatic, hydrodynamic, tectonic characteristics of the subbasins. Similar published comparative study has not known for us for the Hungarian subbasins, so this work can be considered as a pioneer attempt. We plan to follow this study, the continuation of the work will be completed with the hydrocarbon generation type and maturity data of possible source rocks.

REFERENCES

- KOVÁCS, ZS., BABINSZKI, E., CSERKÉSZ-NAGY, Á., JÁMBOR, Á., KERCSMÁR, ZS., MAIGUT, V., SELMECZI, I., THAMÓNÉ BOZSÓ, E. & TÓTHNÉ MAKK, Á. (2017): Hydrocarbon potential of Hungary. Hydrocarbon subbasins, hydrocarbon resources. -Manuscript, Report for the Hungarian Office for Mining and Geology. Hungarian Geological and Geophysical Institute, Budapest, 351 p.
- TISSOT, B.P. & WELTE, D.H. (1984): Petroleum formation and occurrence. - Springer-Verlag, Berlin, 699 p.

Part IV - Geomathematics in water quality protection

Determining water-bodies using multivariate data analysis methods

József Kovács

Department of Physical and Applied Geology, Eötvös Loránd University, kevesolt@geology.elte.hu

The European Water Framework Directive (WFD) obliged EU Member states to achieve good qualitative and quantitative status for all water bodies by 2015. According to the WFD's definition, a “water body” as a basic unit has to be ecologically “near homogeneous”. Hence, single surface waters are considered to be one “water body”. This approach can be questioned in the case of Lake Balaton. We wanted to design a method that is capable of refining the "one lake, one water body" approach. We wanted to find a method with which we can delimit sub-areas characterized by differing water quality. Another aim of the research was to determine the minimum number of sampling locations needed to represent the sub-areas.

To determine the sub-areas of Lake Balaton cluster analysis, discriminant analysis were used. The clustering is performed in three steps. Step 1: 85 points in time were examined individually using hierarchical cluster analysis. As a result, 85 dendrograms were produced. Step 2: After each group at every point in time received a code, these were clustered in time (1985-2004). As a result, time intervals with similar behavior were obtained from which homogeneous sub-areas were determined. There were two time intervals with different behaviors, namely 1985-1997 and 1998-2004. Between 1985 and 1997 two patterns were present at the same time so we separated this interval into two parts. Step 3: the points in time from the three time intervals (1985-1997a, 1985-1997b, and 1998-2004) were extracted and clustered. The cluster analysis was conducted on the sampling locations from the points in time belonging to the same time intervals. This led to the final results. Because of the coding, a “new dataset” was obtained and this permitted the use of cluster analysis. For the period of 1985-2004 we identified five distinct sub-areas in Lake Balaton's water body. In order to be able to follow the changes between sub-areas of the water body, we suggested a minimum of five sampling locations, each one in a different sub-area. Verification of the cluster results by using discriminant analysis showed that the determined sub-areas in the two time intervals were 91% and 84% correctly classified, respectively. For the time interval 1998-2004 77.6% of classification was correct.

Key words: *multivariate statistics, sampling site grouping, combined cluster and discriminant analysis*

A reactive transport model of CO₂ and brine inflow to a fresh water aquifer

Zsuzsanna Szabó¹, Nóra Edit Gál², Éva Kun², Teodóra Szőcs², György Falus¹

¹Department of Geochemistry and Laboratories, ²Department of Hydrogeology,
Geological and Geophysical Institute of Hungary, Stefánia út 14., Budapest 1143, Hungary,
szabo.zsuzsanna@mfgi.hu, zsszabo86@gmail.com

CO₂ geological storage (CCS) is a transitional technology to reduce greenhouse gas emissions and to mitigate climate change. During the study of potentially suitable saline water formations in Hungary, it became necessary to extend the evaluation to the storage complex, as a whole. In worst-case leakage scenarios of the technology, CO₂ or brine may contaminate shallower drinking water aquifers. If this first happens, the dissolution of CO₂ significantly changes the pH and induces mineral dissolution and precipitation in the aquifer and therefore, changes in solution composition and rock porosity. Mobilization of heavy metals may also be a concern. Brine replacement of fresh water due to the pressure increase in the geological system induces mineral reactions as well. This work integrates an advanced geochemical toolset to predict and understand the effects of the above mentioned contamination scenarios. The improvement of a modelling methodology for the subject of the study has happened through several levels, such as equilibrium batch, kinetic batch and finally, 1D kinetic reactive transport simulations. These have been implemented in the widely applied PHREEQC code. Since the visualization of the modelling output faced inconveniences, the production of figures and animations from the results has been automatized by R programming. Both scenarios (CO₂ or brine leakage) indicate the increase of ion-concentrations in the fresh water, which might exceed drinking water limit values. Transport models also provide a possibility to identify the most suitable chemical parameter for leakage monitoring. For example, in the CO₂ models the potassium concentration increases significantly and as the fluid flows, it runs ahead of other parameters. In the rock, feldspar, montmorillonite, dolomite and illite dissolves, calcite, chlorite, kaolinite and silica precipitates, and dawsonite traps a part of CO₂. The work presents a generalizable methodology with a case-study at the CCS field using open-source PHREEQC and R.

Key words: *CO₂ geological storage, geochemical interactions, PHREEQC, R programming*

1. INTRODUCTION

Carbon Capture and Storage (CCS) is a transitional technology for the reduction of greenhouse gas emission and the mitigation of climate change. Following the implementation of the 2009/31/EC Directive in the Hungarian legislation, the Geological and Geophysical Institute of Hungary (MFGI) is required to evaluate the potentially suitable Hungarian geological structures for CCS, the geological sequestration of CO₂. An initial assessment period closed in 2014 which included determining the geological and technical viability, estimating storage volume and evaluating the basic environmental aspects of selected structures. A second assessment period started in 2015 including the study of the geochemical reactivity of reservoir and cap rocks, as well as evaluating disaster scenarios of CO₂ escape and brine displacement to protected shallower drinking, i.e. fresh water aquifers of the neighbouring area. To predict potential CO₂ or brine induced effects, reactive transport modelling of rock – water (– CO₂ gas) interaction is carried out for the first time for the Pannonian Basin. The creation of the transport modelling approach had to follow the batch modelling and inspection of several scenarios. Another aim of the work is to provide recommendations for the most suitable chemical parameters for leakage monitoring.

2. STUDIED AREA

In total, 13 potential CO₂ geological storage reservoirs, i.e. high porosity saline water structures were selected for the study. These are primarily located in the South-East part of Hungary, at the Great Hungarian Plain and are all overlaid by thick Quaternary and Upper Pannonian (Miocene) multi-layered aquifers. Above the brine, protected drinking water is found at shallower depths both in the Quaternary and in the upper part of the Upper Pannonian (Miocene A) layers. The worst-case CO₂ or brine leakage modelling approach is presented through the example of the well-studied Mezőtúr area which is characterized by regional and intermediate flow systems with a dominant upward flow, comprising a regional discharge area.

3. MODELLING METHODOLOGY

All of the geochemical modelling work is carried out in PHREEQC ver. 3 (Parkhurst and Appelo, 2013) and its results are visualized by several codes written in R programming language (R Core Team, 2017)

3.1. Model types

So far, three levels of geochemical modelling methodology have been applied providing complementary information to each other and the possibility to build up the most representative reactive transport concepts of worst-case leakage scenarios. These levels are as follows:

- Thermodynamic batch models: CO₂ added to drinking water in 1 mmol steps up to 10 mol, and brine displacement of fresh water in 0-100%;
- Kinetic batch models: several years of reactions with CO₂ saturated fresh water, and 10% saline+90% fresh water in the aquifer;
- 1D kinetic transport models: inflow and reactive transport of CO₂ saturated fresh water, and 10% saline+90% fresh water in the aquifer.

3.2. Model input data

3.2.1. Thermodynamic

For all simulations the PHREEQC.DAT thermodynamic database was used. Any missing information was taken from the LLNL.DAT database (Wolery, 1979).

3.2.2. Kinetic

Kinetic dissolution and precipitation equations of mineral phases have been defined similarly to an earlier work of the authors published in Szabó et al. (2016), which modelling methodology is primarily based on the USGS report of Palandri and Kharaka (2004). Specific surface areas of minerals originate from the RES³T database (HZDR, online).

3.2.3. Fluid and matrix properties

The solution composition of saline and fresh waters and their temperature information are accessible from MFGI databases. The mineral composition applied in the models has been determined based on XRD measurements of core samples and estimated 15% porosity for the formation. In transport models 2 m/day flow rate was used based on previous regional 3D hydrodynamic modelling.

4. RESULTS AND DISCUSSION

4.1. Understanding reactions in a static concept – batch models

4.1.1. Thermodynamic batch models

Thermodynamic models provide information on equilibrium state of different rock – water – gas mixtures in infinite time. Mineral compositions predicted in brine and CO₂ contamination models of fresh water aquifer of Mezőtúr area are presented in **Figure 1**. These models benefit from the possibility of running several scenarios of interest at a time and support selection of most representative ones for further modelling tasks. Brine displacement continuously changes the equilibrium state (increasing illite, dolomite dissolution and kaolinite, calcite precipitation), whereas different amounts of CO₂ inflow changes reactions in three major steps characterized by the full dissolution of albite and calcite.

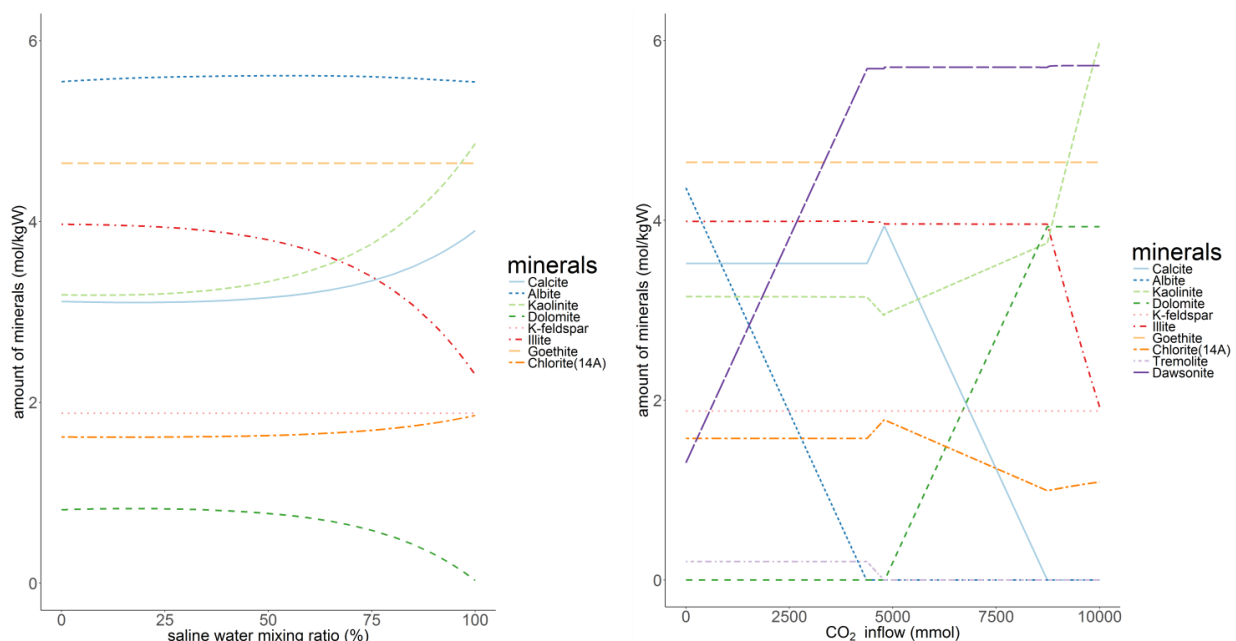


Figure 1: Equilibrium mineral compositions predicted by thermodynamic batch models of brine and CO₂ inflow, respectively.

4.1.2. Kinetic batch models

Geochemical systems are usually far from equilibrium therefore, kinetic modelling provides the necessary information for evaluating induced reactions. For this type of models specific concepts were selected to present in **Figure 2**,

which were 10% brine + 90% fresh water and CO₂ saturated fresh water reaction in the aquifer simulated for 125 years. In the batch modelling approach, no significant difference was observed between the pure fresh water model and that of the brine mixture, unlike the CO₂ model. It becomes clear how the change of the flow rate in transport simulations corresponds to different parts of the presented curves (**Figure 2**) and how it influences the outcome of the models presented in the next section.

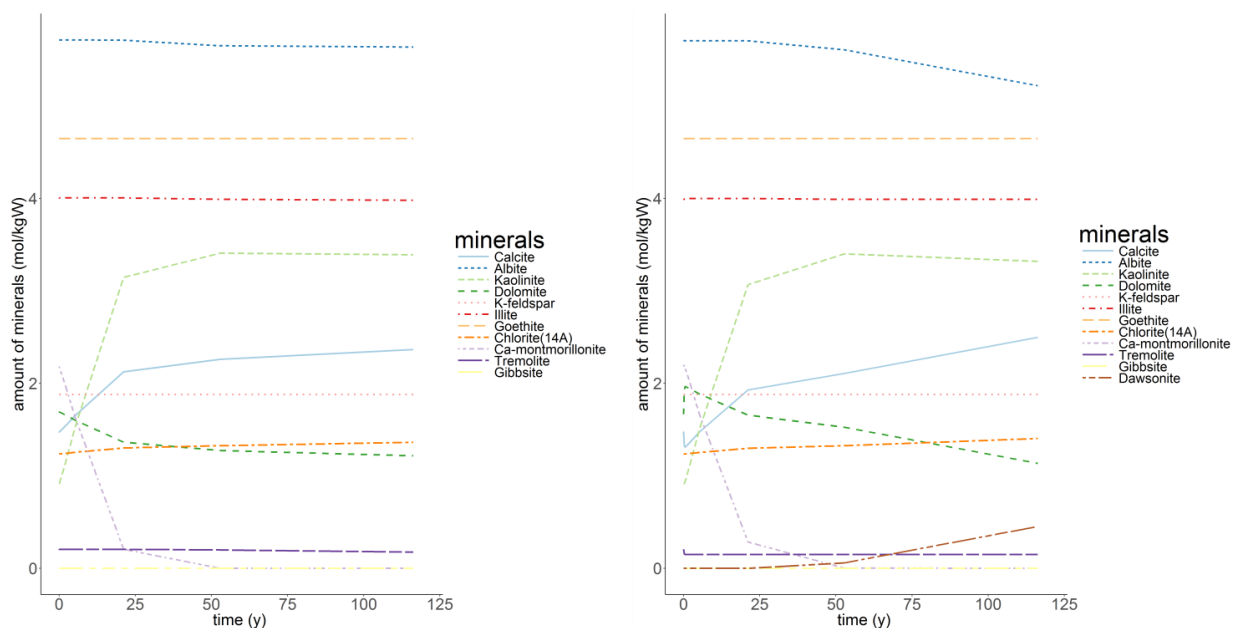


Figure 2: Mineral compositions vs. time in kinetic batch models of 10% brine + 90% fresh water and CO₂ saturated fresh water, respectively.

4.2. Kinetic reactive transport models

4.2.1. Reactions in fresh water aquifer – a reference model

Since, even without any brine or CO₂ leakage, there are reactions taking place in the fresh water aquifer, it is important to provide a reference model of this continuous fluid flow in porous media. In **Figure 3**, the most important observation which can be made is the dissolution of Ca-montmorillonite and the precipitation of kaolinite along the whole path.

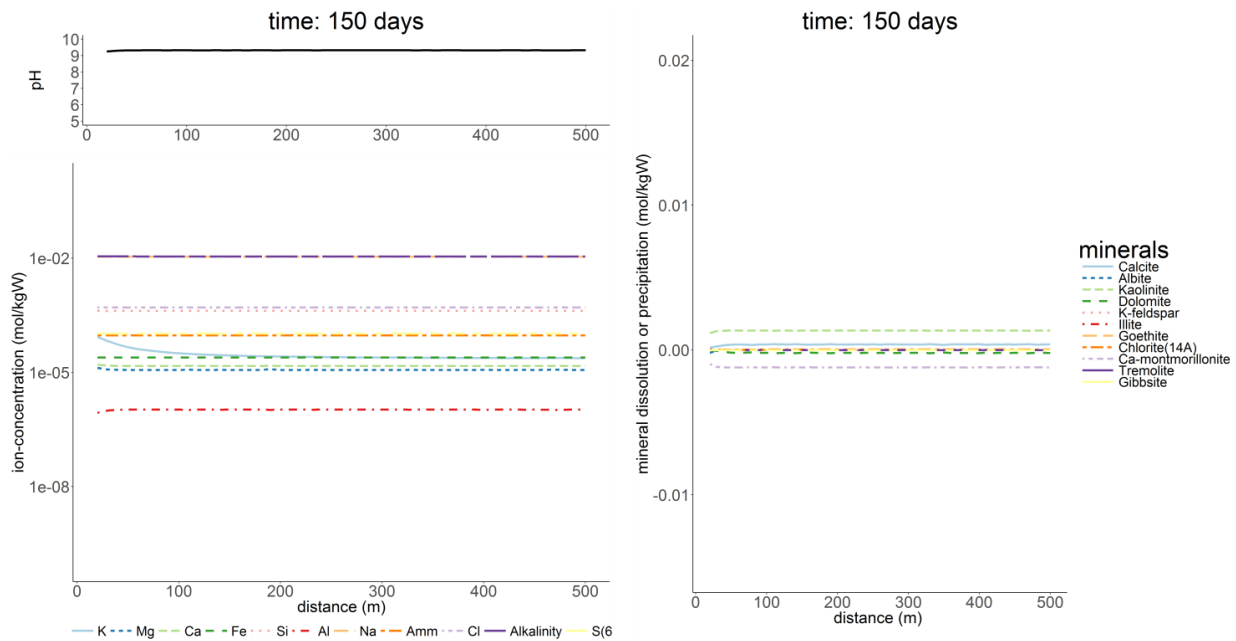


Figure 3: Solution compositions and mineral reactions of fresh water flow

4.2.2. Effect of brine replacement of fresh water

Brine-fresh water mixture inflow into the studied aquifer changes solution compositions corresponding to the chemistry of the brine. The front is at around 500 m for all parameters on the 150th day of the inflow (**Figure 4**), therefore the most suitable monitoring parameter could be selected solely based on the brine composition, i.e. conductivity, a low cost parameter.

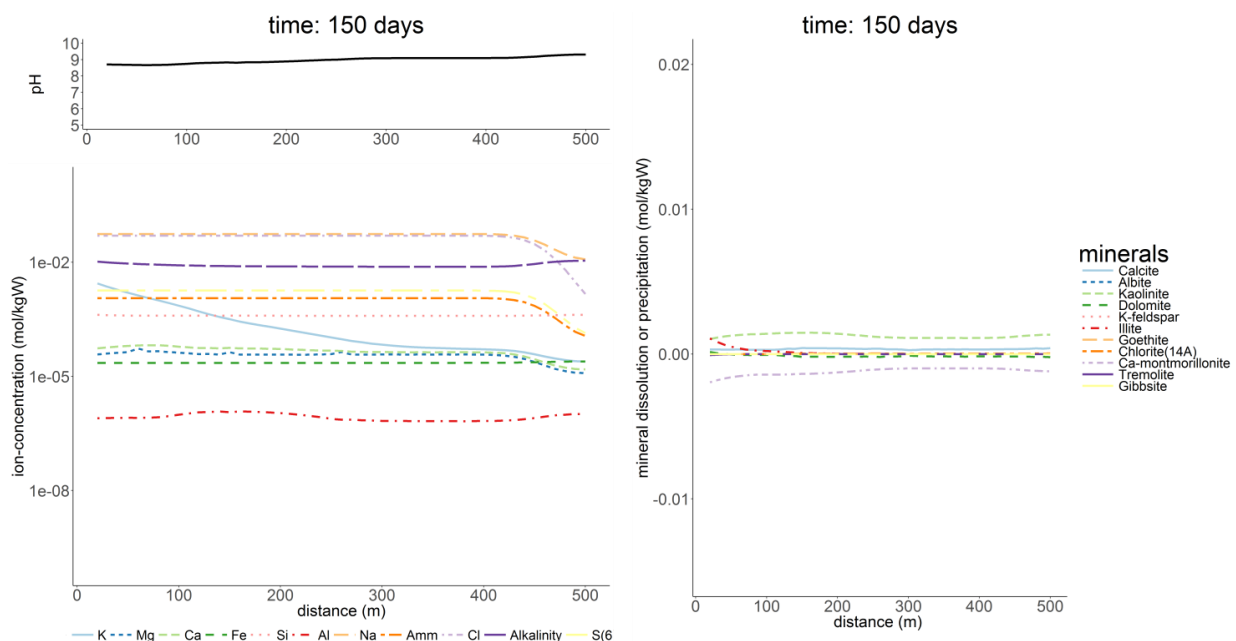


Figure 4: Predicted consequences of 10% brine + 90% fresh water inflow

4.2.3. Reactions in fresh water aquifer induced by CO₂ inflow

CO₂ dissolution in the fresh water significantly reduces its pH (**Figure 5**, top left). This induces much more significant and different mineral dissolution and precipitation reactions in the rock than in the above cases (**Figure 5**, right). This makes it understandable why the reaction front reaches shorter distance on the 150th day than for the transport of brine mixture, 250 m vs. 500 m (**Figure 5** vs. **Figure 4**). Calcite, tremolite, illite and albite dissolve, while dolomite, Ca-montmorillonite and kaolinite precipitate. At the reaction front illite precipitation is predicted. These processes change the solution composition (**Figure 5**, bottom left). K, Mg, Ca, Si concentrations increase among which K runs ahead of the front and suggests that it could be a useful parameter for monitoring beside conductivity.

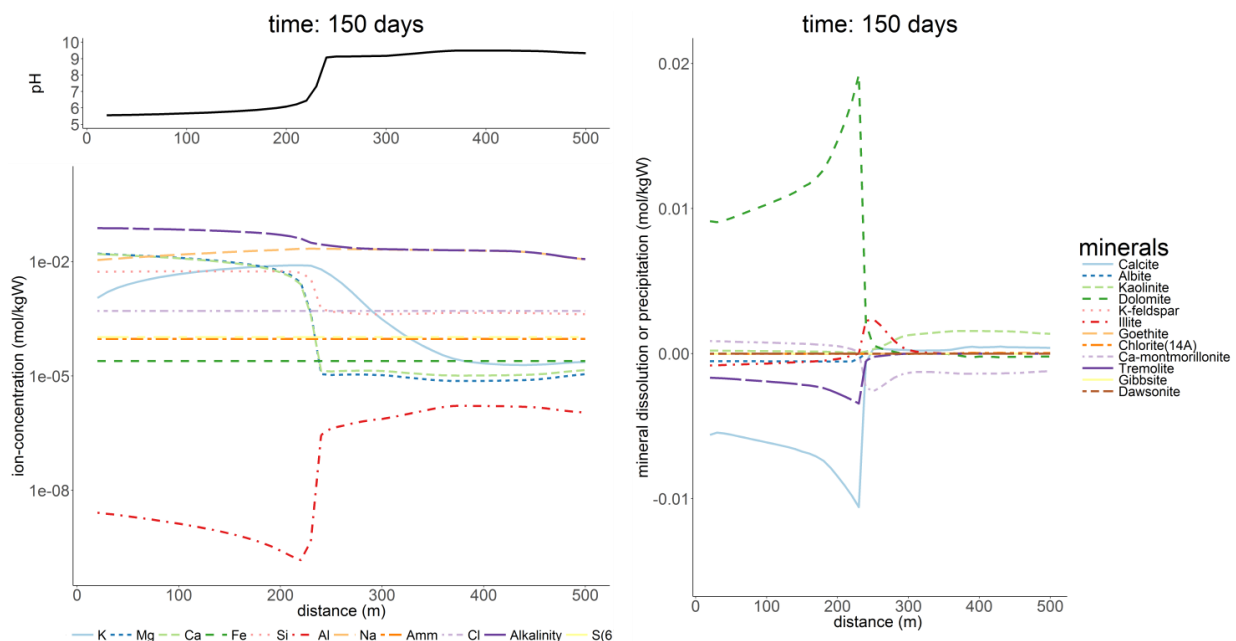


Figure 5: Predicted consequences of CO₂ saturated fresh water inflow

5. CONCLUSIONS

Three main types of PHREEQC models were generated to evaluate unlikely, but potential saline water or CO₂ contamination of drinking water aquifers first time in the Pannonian Basin. Their results are complementary to each other and suggest that certain ion-concentrations may exceed drinking water limit or indicator values in both of the studied worst-case leakage scenarios of CCS. 1D

reactive transport models identify the most suitable chemical parameter for monitoring purposes, i.e. conductivity and K-concentration. Reactivity between rock and fluid slows down the reaction front in the CO₂ inflow case. Further improvement is planned by incorporating cation exchange capacity into models, and their extension to 2-3 dimensions. The applied methodology is also available for other study cases outside of CCS field.

ACKNOWLEDGEMENTS

The research was supported by the OTKA program of the National Research, Development and Innovation Office of Hungary (K 115927 for Gy. Falus).

REFERENCES

HZDR, HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF, RES³T - Rossendorf Expert System for Surface and Sorption Thermodynamics.

<http://www.hzdr.de/db/RES3T.queryData>

PALANDRI, J.L. & KHARAKA, Y.K. (2004): A compilation of rate parameters of water–mineral interaction kinetics for application to geochemical modelling. USGS Water-Resources Investigations Report 04/1068.

PARKHURST, D.L. & APPELO, C.A.J. (2013): Description of input and examples for PHREEQC version 3—A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. USGS Techniques and Methods, book 6, chap. A43, 497.

R CORE TEAM (2017): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

SZABÓ, ZS., HELLEVANG, H., KIRÁLY, CS., SENDULA, E., KÓNYA, P., FALUS, GY., TÖRÖK, SZ. & SZABÓ, CS. (2016): Experimental-modelling geochemical study of potential CCS caprocks in brine and CO₂-saturated brine. International Journal of Greenhouse Gas Control, 44, 262–275.

WOLERY, T.J. (1979): Calculation of chemical equilibrium between aqueous solution and minerals—The EQ3/6 software package: Livermore, Calif., Lawrence Livermore National Laboratory Report UCRL-52658, 41 p.

Estimation of chlorophyll-a in rivers, with the example of the Rivers Danube and Tisza

István Gábor Hatvani^{1*}, József Kovács², Péter Tanos^{3,4}, Gábor Várbíró^{5,6}, Angéla Anda³

¹ Institute for Geological and Geochemical Research, Research Center for Astronomy and Earth Sciences, MTA, H-1112 Budapest, Budaörsi út 45, Hungary; hatvaniig@gmail.com

² Eötvös Loránd University, Department of Physical and Applied Geology, H-1117 Budapest, Pázmány Péter stny. 1/C, Hungary; kevesolt@gmail.com

³ University of Pannonia Georgikon Faculty, Department of Meteorology and Water Management, H-8360 Keszthely, Festetics u. 7, Hungary; tanospeter@gmail.com,
anda-a@georgikon.hu

⁴ Szent István University, Department of Mathematics and Informatics, H-2100 Gödöllő, Páter Károly utca 1, Hungary; molnar.sandor@gek.szie.hu

⁵ MTA Centre for Ecological Research, Danube Research Institute Department of Tisza River Research, H-4026 Debrecen, Bem tér 18/C, Hungary varbiro.gabor@okologia.mta.hu

⁶ MTA Centre for Ecological Research, GINOP Sustainable Ecosystems Group, H-8237 Tihany, Klebansberg Kuno u. 3, Hungary;

Phytoplankton biomass is a key driver of water quality of rivers and lakes. Thus, in cases where its measurements are lacking, its estimation is highly important. The aim of the present study was to estimate the chlorophyll-a concentration in the two largest rivers in Hungary for the period 1993-2005 using the results of wavelet spectrum- and regression analyses conducted on 15 water quality variables measured in both rivers at numerous sampling sites. Results show that the presence of annual periodicity in the water quality parameters varied over time and space; in general, however, an increase was observed downstream. Based on the spatial distribution of the annual periodic behavior of the water quality parameters, significant and acceptable models were derived for both rivers. The annual periodic behavior of ions and streamflow (River Danube) or nutrients (River Tisza) accounted for >50% of the variance of chlorophyll-a.

Key words: annual periodicity, Morlet wavelet spectrum analysis, multiple regression analysis, phytoplankton biomass, River Tisza, River Danube

1. INTRODUCTION

Eutrophication is one of the focal points of modern day water quality protection. In general, the main indicator of eutrophication, phytoplankton biomass, unquestionably shows an increase downstream, as a river may become more easily dominated by the planktonic element (Várbíró et al., 2007). Unfortunately,

in Eastern Europe, long term data are generally only available from the 1990s for the primary proxy for phytoplankton biomass: chlorophyll-a (Borics et al., 2007).

Phytoplankton play a vital role in fluvial ecosystems, especially in cases of changing climatic and environmental conditions (Villegas & de Giner, 1973) and as such are important indicators of water quality, due to their short life cycle (Wu et al., 2014). Therefore, by modeling phytoplankton biomass, real life phenomena governed by water quality may be approximated. Successful models have been derived previously for rivers (e.g. Wu et al., 2014) and lakes (e.g. Hussain et al., 2016) in harmony with the general laws of environmental monitoring (Molnár et al., 2014). However, none of these models has the periodic behavior of different water quality parameters incorporated into it as possible driving factors.

Annual periodicity is a natural behavior of riverine systems in the moderate climate zone, and has been shown (Tanos et al., 2015) to play a major role in driving the periodic behavior of water quality parameters and shaping natural phytoplankton dynamics. This is particularly important in the context of imminent climate change (Molnár & Molnár, 2012).

It has also been documented that the complex nature- and the superimposed presence of (i) anthropogenic, as well as (ii) other natural processes can disturb the natural periodic behavior of different ecosystems (e.g. Kovács et al., 2010; Fehér et al., 2016). With changes in the periodic behavior of water quality variables comes a change in the growth of phytoplankton. As primary producers, planktonic algae in aquatic environments are driven by the available nutrients and factors such as turbidity, water temperature and transparency, and the periodic characteristic of these factors, while the algae themselves have a determining role in shaping the composition of aquatic ecosystems through their production of organic carbon and oxygen, as well as providing a source of food for herbivorous grazers (Wehr & Descy, 1998), thereby enhancing or diminishing the periodic characteristic of a river section.

Although, models have been derived to describe the relationship between macronutrients - mainly total phosphorus and total nitrogen - and chlorophyll-

a, these models mainly focused on lakes (e.g. Phillips et al., 2008; Borics et al., 2016). Thus, if the periodic behavior of the general water quality parameters (streamflow, ions, nutrients etc.) can be proven to play a significant role in driving the variance of chlorophyll-a content and quantify that effect, it could serve as a direct link in the creation of a new way of estimating phytoplankton chlorophyll-a presence.

Therefore, the aims of the study are to determine the change in annual periodic behavior of the water quality parameters of the riverine system of the Rivers Tisza and Danube and with the information obtained derive a model for the estimation of chlorophyll-a values from it.

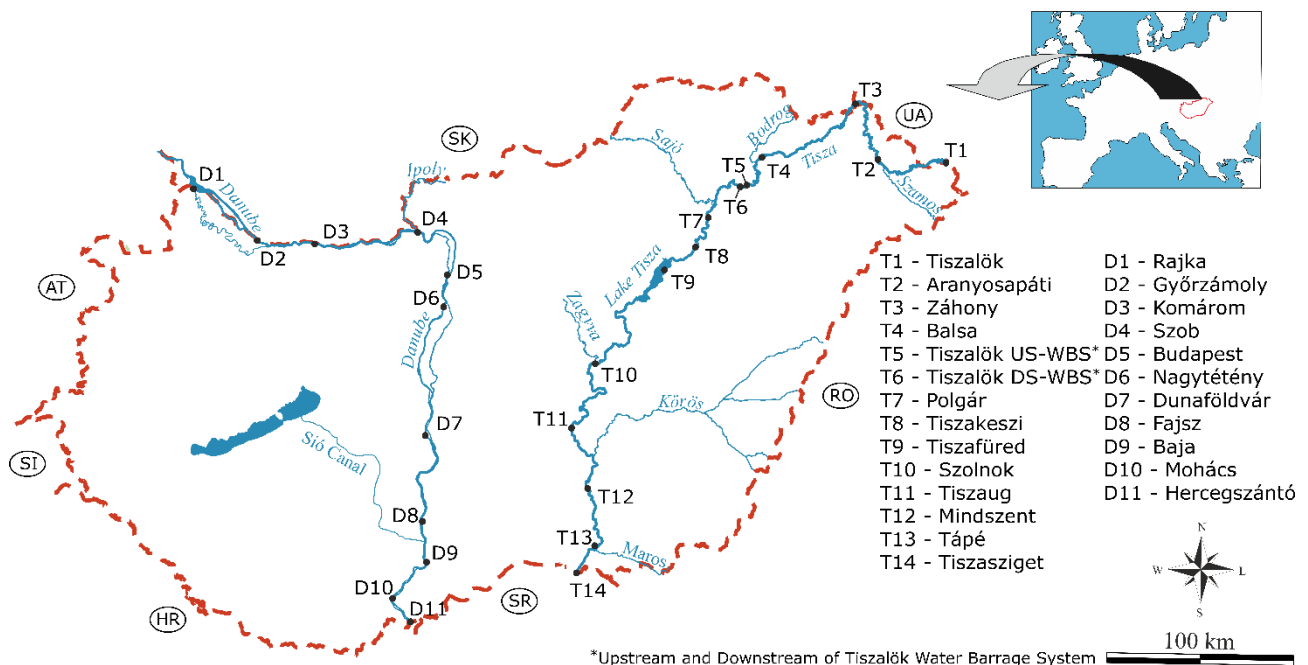


Figure 1: Location of the sampling sites of the Rivers- Danube and Tisza

2. MATERIALS AND METHODS

The monthly time series – after statistical resampling - of 14 physical, chemical and biological water quality parameters were assessed for 1993-2005 from 14 monitoring stations from the River Tisza and 11 on the Danube.

In this study first Morlet wavelet spectrum analysis (Goupillaud et al., 1984) was conducted on the time series of each variable at each sampling location to determine the percentage of time in which the annual period was present, if the

full time period was considered as 100%. Periodicity indices were then derived from these for easier interpretation (**Table 1**).

Table 1: Definitions of the periodicity indices (taken from Kovács et al., 2017)

Name	Abbreviation	Definition
PI of each variable	PI _v	The ratio of time where the annual period is present to the full assessed time period in percentages for a particular variable.
PI of a group of variables	PI _{gv}	Average PI _v of a particular variable group at a certain sampling site.
PI of a particular sampling site	PI _{sl}	The ratio of the sum of the length of time where the annual period is present to the sum of the time periods assessed at a particular sampling site considering all variable together.

These periodicity indices were subsequently employed as the independent variables in the multiple regression models calculated with the aim of estimating chlorophyll-a.

3. RESULTS AND DISCUSSION

In the River Tisza, the annual periodicity of the water quality parameters indicated an average 36% increase in its value downstream, while in the case of the River Danube this was 16% (Tisza PI_{sls} range from 22 to 58%, Danube PI_{sls} range from 71 to 87%). The smallest PI_{sl} was observed at the first sampling site, while the largest was at the penultimate and last sites in the Rivers Tisza and Danube, respectively. The increase was not strictly monotonic, due to the disturbing effect of the Tisza Lök water barrage system (Kovács et al., 2017). In the River Danube the PI percentages were higher, but their increase was smaller downstream. The two most noticeable drops in the PI_{sl} were observed at D4 after the mouth of the Ipoly Stream and at D9 after the junction of the Sió Canal and the River Danube; these are most characteristic in the model which incorporates all the variables (**Figure 2**: blue line).

Since both rivers can be considered as linear systems (Kovács et al., 2015), their PIs were plotted against the distance between the first and the last sampling

locations, creating linear regression models, e.g. in the case of the River Danube (**Figure 2**). These particular increasing patterns along with the regression models formed the backbone of the chlorophyll-a estimation.

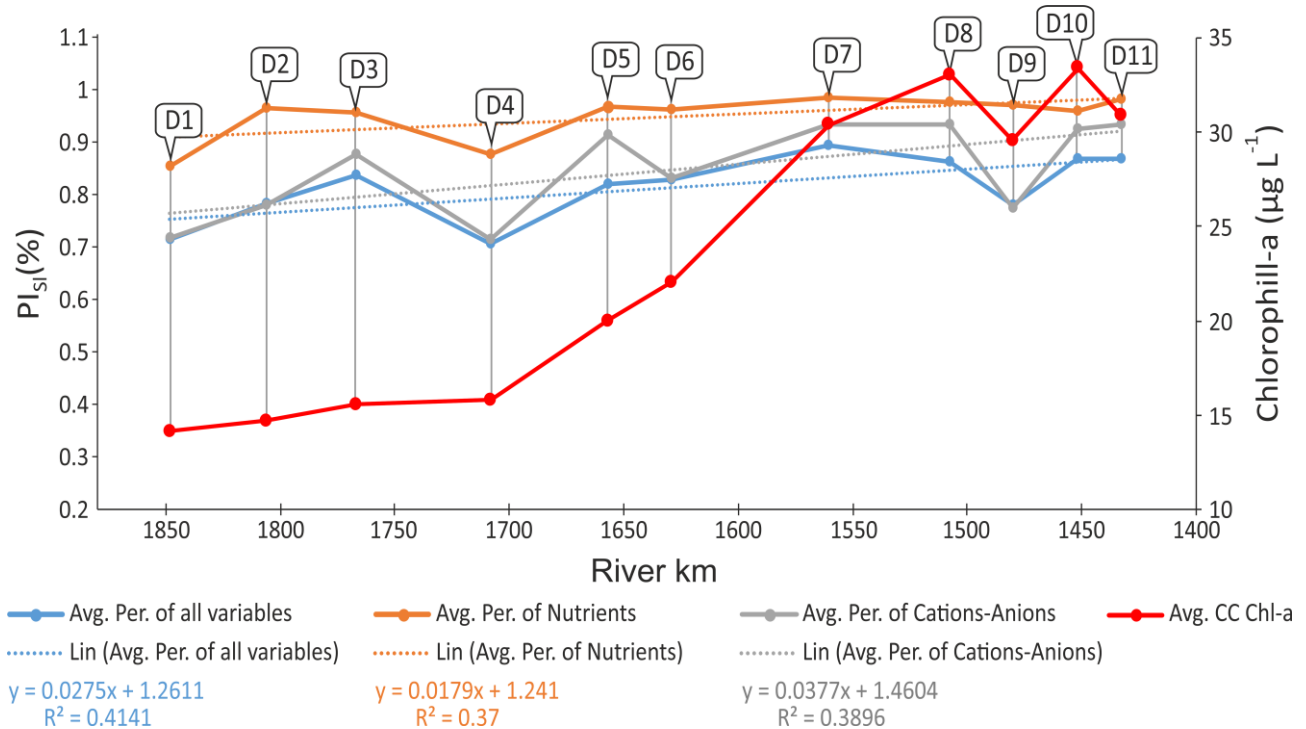


Figure 2: Linear regression models of PIs (Table 1) with different combinations of water quality variables incorporated into them vs. river km for the River Danube

Seven multiple regression models incorporating different combinations of water quality variables were derived and evaluated taking into account adjusted R^2 (R^2) root mean square error, model p-value, and variance inflation factor (VIF) to estimate the chlorophyll-a content of the two rivers using the PIs of the water quality parameters at the different sites (Table 3).

The estimated and measured chlorophyll-a values correlated at $r > 0.6$ in the Tisza and $r > 0.56$ in the River Danube, while all the models were proven to be significant and a VIF of $< \max. 2.63$ indicated that there is no alarming multicollinearity in either model of the two rivers. The average root mean squared error (RMSE) was $1.05 \mu\text{g L}^{-1}$, the avg. $R^2 = 0.439$ in the River Tisza and avg. RMSE = $2.01 \mu\text{g L}^{-1}$ and avg. $R^2 = 0.497$ in the Danube.

In summary, the periodic behavior of ions and streamflow in the River Danube, and nutrients in the River Tisza accounted for >50% of the variance of chlorophyll-a (**Table 2**).

Table 2: Parameters of the linear regression models used to estimate chlorophyll-a as the dependent variable (after Kovács et al., 2017)

Independent variable(s)	River Tisza			River Danube		
	R ²	p-value	RMSE	R ²	p-value	RMSE
PI-streamflow	0.36	0.01	1.16	0.46	0.01	2.22
PI-nutrients	0.46	0.01	1.06	0.31	0.04	2.25
PI-ions	0.41	0.01	1.11	0.40	0.02	2.09
PI-streamflow, nutrients	0.45	0.02	1.03	0.58	0.01	1.92
PI-streamflow, ions	0.43	0.02	1.04	0.71	0.00	1.74
PI-nutrients, ions	0.50	0.01	0.98	0.35	0.07	2.11
PI-streamflow, nutrients, ions	0.46	0.03	0.97	0.67	0.01	1.74

As both rivers slow down, their water becomes more transparent as sediment carried in suspension settles out and their water quality changes to that of a lower section type-of-river downstream. At the same time, their periodic behavior increases and so does chlorophyll-a content, due to the previous phenomena (Abonyi et al., 2012). The disturbances, however, in the monotonic increase of the annual periodic behavior of the combination of the water quality parameters were always caused by an explainable anomaly. From this perspective, particular attention should be paid to the side-arms, canals and man-made obstacles. It should be noted that both anthropogenic- and natural disturbances which reduce periodicity decrease primary production as well (Kovács et al., 2017), directly affecting the whole riverine ecosystem through the food web (Ou and Winemiller, 2016).

4. CONCLUSIONS

Rivers, one of the most endangered ecosystems, produce a wide range of ecosystem services. Their monitoring is a focal point of action strategies aiming to conserve/improve environmental conditions. With the present study we hope to fill a niche in water quality monitoring by facilitating the estimation of chlorophyll-a through the combination of periodicity- and multiple regression

analyses. By applying the previously presented methodology to two large rivers over a wide time interval (1993-2005) it was shown that the periodicity of water quality variables has a quantifiable and significant effect on riverine ecosystems, specifically phytoplankton biomass. To assess this relationship further, data from different rivers, or river sections (e.g. lower section, river delta) and additional parameters (meteorological, physical, etc.) should be incorporated into this basic model.

ACKNOWLEDGMENTS

Thanks to the MTA “Lendület” program (LP2012-27/2012), GINOP-2.3.2-15-2016-00019 project and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. This is contribution No. 47 of 2ka Palæoclimate Research Group.

REFERENCES

- HUSSIAN, A.F., HOWAYDA, H.A.E., MEF, T. & VARBIRO G. (2016): Phytoplankton Structure biochemical Stoichiometry and elemental composition in Lake Nasser Egypt. *Int. J. Of Applied Environmental Sciences Research* 11, 1, 211-228
- ABONYI, A., LEITÃO, M., LANÇON, A.M. & PADISÁK, J. (2012): Phytoplankton functional groups as indicators of human impacts along the River Loire (France). *Hydrobiologia*, 698,1, 233-249
- BORICS, G., TÓTHMÉRÉSZ B., VÁRBÍRÓ G., GRIGORSZKY I., CZÉBELY A. & GÖRGÉNYI J. (2016): Functional phytoplankton distribution in hypertrophic systems across water body size. *HYDROBIOLOGIA* 764, 1, 81-90
- BORICS, G., VÁRBÍRÓ, G., GRIGORSZKY, I., KRASZNAI, E., SZABÓ, S. & KISS, K.T. (2007): A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers. *Large Rivers* 17,3-4, 466-486
- FEHÉR, K., KOVÁCS, J., MÁRKUS, L., BORBÁS, E., TANOS, P. & HATVANI, I.G. (2016): Analysis of drip water in an urban karst cave beneath the Hungarian capital (Budapest). *Acta Carsologica*, 45,3, 213-231
- GOUPILLAUD, P., GROSSMAN, A. & MORLET, J. (1984): Cycle-Octave and Related Transforms in Seismic Signal Analysis. *Geoexploration*, 23,85-102
- KOVÁCS, J., TANOS, P., VÁRBÍRÓ, G., ANDA, A., MOLNÁR, S. & HATVANI, I.G. (2017): The role of annual periodic behavior of water quality parameters in

- primary production: chlorophyll-a estimation Ecological Indicators (in press)
DOI: 10.1016/j.ecolind.2017.03.002
- KOVÁCS, J., KOVÁCS, S., HATVANI, I.G., MAGYAR, N., TANOS, P., KORPONAI, J. & BLASCHKE A.P. (2016): Spatial optimization of monitoring networks on the examples of a river, a lake-wetland system and a sub-surface water system. *Water Resources Management* 29, 5275
- KOVÁCS, J., HATVANI, I.G., KORPONAI, J. & SZÉKELY KOVÁCS, I. (2010): Morlet wavelet and autocorrelation analysis of long-term data series of the Kis-Balaton water protection system (KBWPS). *Ecological Engineering*, 36, 10, 1469-1477
- MOLNÁR, S. & MOLNÁR, M. (2012): Comprehensive assessment of climate change policies and measures in Hungary: concerns and tasks in an underestimated challenge, *Időjárás / Quarterly Journal of the Hungarian Meteorological Service* 116, 297-321.
- MOLNÁR, S., MOLNÁR, M. & FÜST, A. (2014): Környezeti folyamatok modellezése, *Szent István Egyetem, Gödöllő*.
- OU, C. & WINEMILLER, K.O. (2016): Seasonal hydrology shifts production sources supporting fishes in rivers of the Lower Mekong Basin. *Can. J. Fish. Aquat. Sci.* 73, 1342–1362
- PHILLIPS, G., PIETILÄINEN, O., CARVALHO, L., SOLIMINI, A., LYCHE SOLHEIM, A. & CARDOSO, A. (2008): Chlorophyll–nutrient relationships of different lake types using a large European dataset. *Aquatic Ecology* 42, 2, 213–226
- TANOS, P., KOVÁCS, J., KOVÁCS, S., ANDA, A. & HATVANI, I.G. (2015): Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environmental Monitoring and Assessment* 187, 9, 575
- VÁRBÍRÓ, G., ACS, É., BORICS, G., ÉRCES, K., FEHÉR, G., GRIGORSZKY, I., JAPOPORT, T., KOCSI, G., KRASZNAI, E., NAGY, K., NAGY-LÁSZLÓ, ZS., PILINSZKY, ZS. & KISS, K.T. (2007): Use of Self-Organizing Maps (SOM) for characterization of riverine phytoplankton associations in Hungary. *Archiv für Hydrobiologie*, 17, 3-4, 383-394
- VILLEGAS, I. & DE GINER, G. (1973): Phytoplankton as a biological indicator of water quality. *Water Res.* 7, 479-487
- WEHR, J.D. & DESCY, J-P. (1998): Use of phytoplankton in large river management. *Journal of Phycology*, 34, 5, 741–749
- WU, N., HUANG, J., SCHMALZ, B. & FOHRER, N. (2014). Modeling daily chlorophyll a dynamics in a German lowland river using artificial neural networks and multiple linear regression approaches. *Limnology*, 15, 1, 47–56

Combined Cluster and Discriminant Analysis on the River Raab's Austrian and Hungarian sections

Péter Tanos^{1,2*}, Angéla Anda², Gábor Várbiro^{3,4}, Solt Kovács⁵, Alfred Paul Blaschke⁶, Molnár Sándor¹

¹Szent István University, Department of Mathematics and Informatics, H-2100 Gödöllő, Páter Károly utca 1, Hungary; tanospeter@gmail.com, molnar.sandor@gek.szie.hu

²University of Pannonia Georgikon Faculty, Department of Meteorology and Water Management, H-8360 Keszthely, Festetics u. 7, Hungary; anda-a@georgikon.hu

³MTA Centre for Ecological Research, Danube Research Institute Department of Tisza River Research, H-4026 Debrecen, Bem tér 18/C, Hungary varbiro.gabor@okologia.mta.hu

⁴MTA Centre for Ecological Research, GINOP Sustainable Ecosystems Group, H-8237 Tihany, Klebelsberg Kuno u. 3, Hungary;

⁵Vienna University of Technology, Institute of Hydraulic Engineering and Water Resources Management, A-1040 Wien, Karlsplatz 13.

The primary requirements of any water management system are that it be efficient and representative. It is therefore a challenge to operate water quality monitoring networks so that they will be both scientifically and economically viable, especially in the case of river sections ranging over hundreds of kilometres. Therefore, the spatio-temporal optimization of such systems is vital task. The aim of the present study was to optimize the monitoring system of a river, the Raab, which flows through both Austria and Hungary. Combined Cluster and Discriminant Analysis was applied to data for 14 water quality parameters from 10 sampling sites equally distributed between the two countries for the period 1994-2004. Results indicate that the dissimilarities/similarities between the neighbouring sampling sites do not appear to depend on their proximity to each other, although the watershed of the section is quite compact. On the contrary, it would appear that the tributaries of the Raab should be considered significant in separating the sampling sites, and that the anthropogenic activity along the river influences the water quality and consequently the grouping of the sites as well to a great degree.

Key words: *Combined Cluster and Discriminant Analysis, River Raab, water quality monitoring system*

1. INTRODUCTION

Due to the ever increasing amount of data, the discipline of geomathematics is growing in importance in the field of environmental science and, therefore, water

quality assessment, too (Hatvani & Horváth, 2016). There are numerous studies which have applied stochastic and/or deterministic methods successfully to group water quality sampling sites, whether these be surface or subsurface (e.g. Naddeo et al., 2013; Lee et al., 2014; Tanos et al., 2015), lakes (Hatvani et al., 2014; Magyar et al., 2013), subsurface water bodies (Kovács et al., 2012; Hatvani et al., 2014; Kovács et al., 2015), or large scale wetland mitigation projects as the Kis-Balaton (Hatvani et al., 2009; 2011; Hatvani, 2014).

The paper presents a new approach, combining the aforementioned geomathematical methods (Combined cluster and discriminant analysis (CCDA) Kovács et al., 2014), to suggest a solution to the problem of finding homogeneous groups of water quality monitoring sites. This procedure was carried out on the water quality monitoring network of the Raab using CCDA to examine the spatial monitoring networks of the river for redundancy.

This is a vital question, because the aims of the EC Water Framework Directive 2000/60/EC cannot be met without proper water quality monitoring. Moreover, in order to obtain representative results, the monitoring networks in question have to be representative of the phenomena they monitor. Unfortunately, sustaining and managing such systems is costly and time consuming, but their optimal functioning is vital from a scientific, environmental and economic perspective (Kovács et al., 2015; Tanos et al., 2015).

2. MATERIALS AND METHODS

The river chosen as the subject of our study is the River Raab, a tributary of the Danube. It has a watershed of 10,270 km² within the Danube basin. The full length of the river is 283 km, 72 km in Austria, 211 km in Hungary (**Figure 1**). The River Raab is affected by external pressures mainly of anthropogenic origin. Numerous industrial facilities (such as leather tanneries, iron works or food factories) and municipal sewage treatment plants can be found along the river. The river's 14 sampling sites were evaluated (1994-2004) using water quality parameters: pH, oxygen demand [%], oxygen content, BOD-5, Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄⁻, NH₄-N, NO₂-N, NO₃-N [mg l⁻¹], PO₄-P [µg l⁻¹] to find their homogeneous groups.

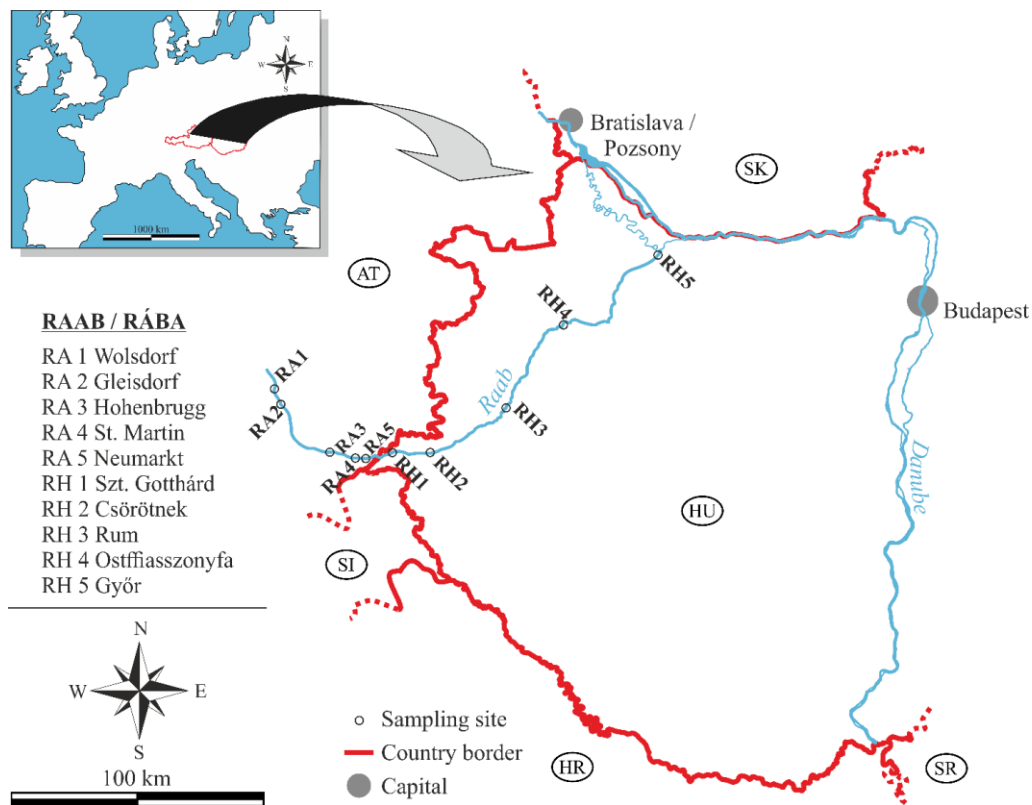


Figure 1: Location of River Raab and its studied water quality monitoring sites

The method used to find homogeneous, and not just similar, groups was Combined cluster and discriminant analysis (CCDA) implemented in R as package `ccda`, and was first introduced by Kovács et al. (2014). During the grouping process a decision has to be made whether further division of certain groups is necessary, or not. Cluster analysis is a frequently used grouping procedure; however, if multiple groupings are possible a decision has to be made which classification to choose. While there are different methods determining an optimal classification with similar group members (Dunn, 1973), CCDA goes a step further and aims to find homogeneous groups. It consists of three main steps: i) a basic grouping procedure to determine possible groupings; ii) a core cycle where the correctness of the groupings from step i) and the correctness of random classifications is determined using linear discriminant analysis; and (iii) a final evaluation step, where a decision about the further iterative investigation of sub-groups is taken. “Hence, the main idea of CCDA is that once the ratio of correctly classified cases for a grouping is higher than at least 95% of the ratios

for the random classifications (i.e. the difference $d = \text{ratio} - q_{95}$ is positive), then at the level of $\alpha = 0.05$ the given classification is not homogeneous” (Kovács et al., 2014).

3. RESULTS AND DISCUSSION

In the River Raab, no homogeneous group of sampling sites was found; even the smallest differences were not enough to reach the level of homogeneity (**Figure 2**). The highest difference was observed between sampling sites RA1 and RA2 (35%), while the smallest between sites RH3 and RH4 (6%). The average difference between the sites was 24%. There was no continuous change to be observed between the pairwise difference of the sampling locations downstream, unlike in the case of the River Danube, where differences frequently exceeded 20-30%, indicating a continuous decrease downstream (Chapman et al., 2016).

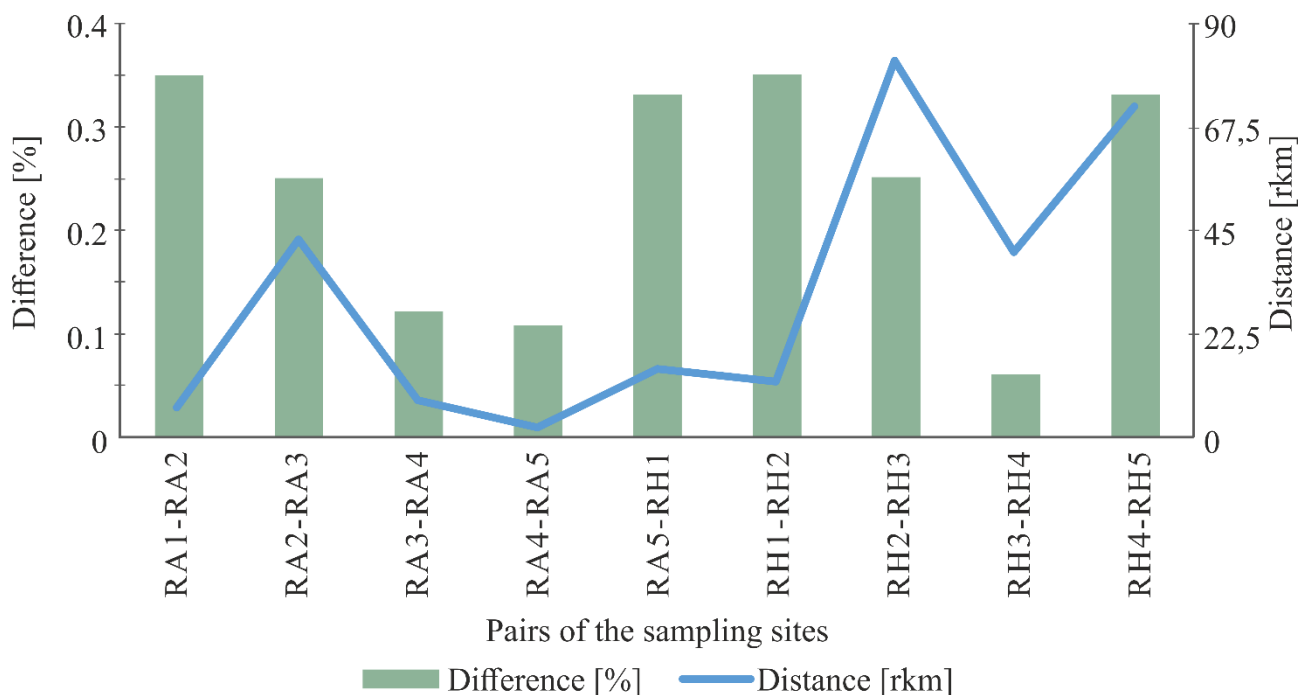


Figure 2: Pairwise comparison of the Danube’s sampling sites (based on Chapman et al., 2014).

The differences between the examined sampling site pairs cannot be simply explained by the distance between them. In many cases, even sites quite close

to each other (e.g. RA5-RH1 or RH1-RH2 in **Figure 2**) show large differences. The following phenomena are suspected to play a major role in determining the difference of the sampling site on the River Raab. In first place, the tributaries, as in numerous other cases (Sharp, 1971; Sanders & Adrian, 1978; Sanders, 1980) one of the most important separating factors are those which we consider as point sources. Even in the case of sites close to each other a tributary can clearly cause their separation and cause a change water composition; e.g. sites RH1 and RH2 separated by the Lafnitz (**Figure 3**). Thus, it is important to note that samples should be taken below the confluence of a tributary where the two different water masses have fully mixed (Chapman et al., 2016) and the location of full mixing should be explored with profiles across and along the river in order to select the sampling site. As an equally important factor, the heterogeneity of the sites of the River Raab can be caused by anthropogenic activity, for example the heavy industry on the banks of the river (e.g. leather tanneries, iron works and food factories). The outlets of the factories are suspected of causing the lack of homogeneity of the sampling sites on the River Raab in both Austria and Hungary (**Figure 3**). Comparing the above results with ones from the largest rivers of the Danube Basin, striking differences can be observed. Both on the Danube and on its largest tributary, the River Tisza, differences between neighbouring sampling sites, clearly smaller than those on the River Raab were seen. This implies the higher vulnerability of the Raab compared to the larger ones, although similar anthropogenic activities take place on the Danube and on the Tisza as well (Tanos et al., 2015; Chapman et al., 2016).

4. CONCLUSIONS

With the application of CCDA to the River Raab, the river's water quality monitoring network was reassessed. It was found that all sites hold unique information, do not form homogeneous groups, and therefore should be retained. Nevertheless, it is suspected that not merely the natural differences in water quality are responsible for the observed pattern, but the tributaries and the intense industrial activity along the river.


Flow direction	Sampling sites	Point sources	Tributaries
	RA1 Wollsdorf		
		Lederfabrik Schmidt Wollsdorf	
	RA2 Gleisdorf		
		STP Gleisdorf	
		STP St. Margarethen	
		STP Fladnitz Raabtal	
		Lederfabrik Boxmark Feldbach	
		STP Feldbach/Raabtal	
		AT&S Fehring	
		ARA Fehring	
	RA3 Hohenbrugg		
	RA4 St. Martin		
		Lederfabrik Boxmark Jennersdorf	
	RA5 Neumarkt		
			border AT
		LURATEX	HU
	RH1 Szentgotthárd		
			Lafnitz
	RH2 Csörötnek		
			Pinka
		STP Körmend	
	RH3 Rum		
		STP Sárvár	
	RH4 Ostffyasszonyfa		
		STP Szany	
		STP Rábacsécsény	
			Marcal
	RH5 Győr		

Figure 3: Homogeneous groups of sampling site in the Austrian Hungarian section of the Raab with the suspected reasons behind their separation (taken from Chapman et al., 2016)

REFERENCES

CHAPMAN, D. V., BRADLEY, C., GETTEL, G. M., HATVANI, I. G., HEIN, T., KOVÁCS, J., LISKA, I., OLIVER, D. M., TANOS, P., TRÁSY, B. & VÁRBÍRÓ, G. (2016): Developments in water quality monitoring and management in large river catchments using the Danube River as an example. *Environmental Science & Policy* 64, 141-154

DUNN, J.C. (1973): A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32-57.

EUROPEAN COMMISSION. DIRECTIVE 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Off J Eur Communities 2000;L 327:1.

HATVANI, I. G. & HORVÁTH, J. (2016): A Special Issue: Geomathematics in practice: Case studies from earth- and environmental sciences: Proceedings of the Croatian-Hungarian Geomathematical Congress, Hungary 2015. Open Geosciences, 8, 1, 1-4.

HATVANI, I. G. (2014): Application of state-of-the-art geomathematical methods in water protection: - on the example of the data series of the Kis-Balaton Water Protection System - Ph.D. thesis, Eötvös Loránd University, Budapest, 110 p.

HATVANI, I. G., CLEMENT, A., KOVÁCS, J., SZÉKELY KOVÁCS, I. & KORPONAI, J. (2014): Assessing water-quality data: the relationship between the water quality amelioration of Lake Balaton and the construction of its mitigation wetland. Journal of Great Lakes Research, 40, 1, 115-125.

HATVANI, I. G., KOVÁCS, J., SZÉKELY KOVÁCS, I., JAKUSCH, P. & KORPONAI, J. (2011): Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. Ecological Engineering, 37, 4, 629-635.

HATVANI, I. G., MAGYAR, N., ZESSNER, M., KOVÁCS, J. & BLASCHKE, A. P. (2014): The Water Framework Directive: Can more information be extracted from groundwater data? A case study of Seewinkel, Burgenland, eastern Austria. Hydrogeology Journal, 22, 4, 779-794.

HATVANI, I., KOVÁCS, J., KORPONAI, J. & KOVÁCSNÉ SZÉKELY, I. (2009): A Kis-Balaton Vízvédelmi Rendszer (KBVR) hosszú távú fizikai, kémiai és biológiai paramétereinek elemzése. Hidrológiai Közlöny, 89, 6, 15-18.

KOVÁCS, J., KOVÁCS, S., HATVANI, I. G., MAGYAR, N., TANOS, P., KORPONAI, J. & BLASCHKE, A. P. (2015): Spatial Optimization of Monitoring Networks on the Examples of a River, a Lake-Wetland System and a Sub-Surface Water

- KOVÁCS, J., KOVÁCS, S., MAGYAR, N., TANOS, P., HATVANI, I. G. & ANDA, A. (2014): Classification into homogeneous groups using combined cluster and discriminant analysis. *Environmental Modelling & Software* 57, 52-59.
- KOVÁCS, J., NAGY, M., CZAUNER, B., SZÉKELY KOVÁCS, I., KÉRINÉ BORSODI, A. & HATVANI I. G. (2012): Delimiting sub-areas in water bodies using multivariate data analysis on the example of Lake Balaton (W Hungary). *Journal of Environmental Management*, 110, 151-158.
- LEE, C., PAIK, K., YOO, D. G. & KIM, J. H. (2014): Efficient method for optimal placing of water quality monitoring stations for an ungauged basin. *Journal of Environmental Management*, 132, 24-31.
- MAGYAR, N., HATVANI, I. G., KOVÁCSNÉ SZÉKELY, I., HERZIG, A., DINKA, M. & KOVÁCS, J. (2013): Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See. *Ecological Engineering*, 55, 82-92.
- NADDEO, V., SCANNAPIECO, D., ZARRA, T. & BELGIORNO, V. (2013). River water quality assessment: Implementation of non-parametric tests for sampling frequency optimization. *Land Use Policy*, 30, 197-205.
- System. *Water Resources Management*, 29, 14, 5275-5294.
- TANOS, P., KOVÁCS, J., KOVÁCS, S., ANDA, A. & HATVANI, I. G. (2015): Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environmental Monitoring and Assessment*, 187, 9, 575.

Part V - Data analysis in engineering geology

Geomathematical aspects of engineering geology

Ákos Török¹

¹Department of Engineering Geology and Geotechnics, Budapest University of Technology and Economics, H-1111 Budapest, Hungary, torokakos@mail.bme.hu

Major engineering geological projects require the handling of large data sets; such as core description, geographic information, on site test results and laboratory analyses. The present paper brings some examples of how these data sets can be managed and dealt with. In the first example the multivariate analysis of geotechnical parameters are given. The studied parameters were obtained from laboratory analyses of core drillings. Water content, dry bulk density, water saturated density, void ratio, index of plasticity, coefficient of skewness, angle of friction, cohesion, compressive strength, modulus of elasticity and Poisson-ratio were considered for multivariate statistical analyses. From these parameters, five engineering geological index properties were selected as coherent ones: dry bulk density, void ratio, angle of friction, cohesion and compressive strength. All together more than 1200 data points were evaluated using SPSS. One of the main aims of the study was to link lithologies to parameters. Results indicate that discriminant analysis did not allow the differentiation of the various lithological categories, since a significant overlap was found in the physical parameters of the five identified lithological units. However, a very strong correlation between cohesion and compressive strength was found.

The second example deals with rock mechanical laboratory analyses of travertine. The apparent density of both dry and water saturated samples, the effective porosity, the uniaxial unconfirmed compressive strength (UCS) and the ultrasonic pulse velocity were determined. A data set of 600 test specimens representing 40 travertine blocks was analysed. Linear regressions have been established between the air dry- and water saturated densities, and air dry- and water saturated ultrasonic pulse velocities. Porosity does not necessarily reflect the strength of travertine but a fairly good exponential correlation was found between density and UCS; ultrasonic wave velocity and UCS for both dry and saturated conditions.

The above listed examples demonstrate the applicability of geomathematics in engineering geology. The support of NKFI Fund (K 116532) is appreciated.

Key words: *multivariate analyses, bulk density, cohesion, uniaxial compressive strength*

Statistical analyses of strength parameters and aggregate properties of andesites from Hungary

Balázs Czinder¹, Ákos Török¹

¹Department of Engineering Geology and Geotechnics, Budapest University of Technology and Economics, H-1111 Budapest, Hungary, czinder.balazs@epito.bme.hu, torokakos@mail.bme.hu

The present paper deals with the strength parameters and aggregate properties of andesites obtained from the quarries of Recsk (Mátra Mountains, Hungary) and of Nógrádkövesd (Cserhát Mountains, Hungary). The measured rock mechanical parameters included bulk density, uniaxial compressive strength, modulus of elasticity and tensile strength. The specimens were tested by being subjected to air-dry, water saturated and freeze-thaw conditions.

The aggregate properties were measured using micro-Deval abrasion tests. Though the relevant standards recommend 12,000 rotations to determine the loss of aggregate samples in the course of the micro-Deval tests, measurements were made on samples subjected to larger numbers of rotations. The number of the rotations was increased in 11 steps, reaching the maximum number of 120,000 rotations in the end.

The aim of the study was to describe statistically the properties of the andesites tested under laboratory conditions. Within the framework of the statistical analyses the extreme overhanging elements of the database were selected and then excluded from further analyses. The distribution of the parameters was tested for normal distribution. Correlation and regression analyses were performed to determine the dependence between the rock mechanical parameters. A linear correlation was found between the uniaxial compressive strength and the modulus of elasticity of the andesite from Nógrádkövesd.

An additional part of the research focused on the change of the resistance to abrasion by using the micro-Deval test results of aggregate samples subjected to different number of rotations. The main aim of this research was to find the most suitable mathematical model to describe the behaviour of the studied andesite aggregates in the course of continuous abrasive impact. The interpreted test results fit well to quadratic models.

Key words: *andesite, rock mechanical parameters, aggregate properties, statistical analyses*

1. INTRODUCTION

Igneous rocks are frequently used as aggregate in Hungary. Volcanic rocks such as basalt and andesite are the most common types found in Hungary. The favourable properties of andesite (high uniaxial compressive strength, resistance to abrasion) make it suitable for use in road and railway construction in the form of aggregate or armourstone and as blocks in hydraulic engineering.

Several test methods exist to determine the aggregate properties of rock materials (the polished stone value, Nordic, aggregate crashing and aggregate impact value tests) but the most common, such as Los Angeles (LA) and micro-Deval (MDE) tests, usually focus on the determination of durability. The LA and MDE test results are taken as the basis of the different classification systems.

Long-term durability is also important for the description of the behaviour of the various rocks undergoing continuous abrasive impact. Erichsen (2015) tested the long-term durability of jasper and greenstone by LA tests; the samples were subjected to different numbers of rotations.

Relationships could be found between rock mechanical parameters and aggregate properties. According to Rigopoulos et al. (2013) the MDE coefficient correlates with uniaxial compressive strength in the case of ophiolite complexes from Greece. A dependence has been found between the polished stone value and the Nordic abrasion value of various igneous rocks (Krutilová & Přikryl, 2017). Török (2015) revealed the relationship between the LA and MDE values of Hungarian andesite and basalt.

The aim of the research is to describe the properties of the studied andesites and determine the degree of dependence between the different parameters. The authors also propose to describe the behaviour of the studied rocks in the course of continuous abrasive impact using suitable mathematical models.

2. MATERIALS

Andesites from two quarries in the northern part of Hungary were used to prepare specimens for the laboratory tests. The andesite from Nógrádkövesd (Cserhát Mountains) was formed during the Miocene volcanism and the andesite from Recsk (Mátra Mountains) was formed in the Eocene. Both rock materials

show quite homogenous fabrics. Whereas the andesite from Nógrádkövesd is coarser porphyritic, the andesite from Recsk is fine porphyritic.

3. METHODS

3.1. Determination of rock mechanical parameters

The rock mechanical parameters were determined under laboratory conditions. The cylindrical specimens were drilled from rock blocks. The mechanical parameters (bulk density, propagation speed of ultrasonic pulse wave, uniaxial compressive strength, modulus of elasticity and indirect tensile strength) were measured for specimens subjected to air-dry, water saturated and freeze-thaw conditions. The laboratory tests were carried out according to the relevant standards and suggested methods (**Table 1**).

Table 1: *The laboratory test methods, the relevant standards and suggestions.*

Laboratory test methods	Relevant standard
Bulk density	EN 1936:2000
Water absorption	EN 13755:2008
Propagation speed of the ultrasonic wave	EN 14579:2005
Determination of frost resistance	EN 12371:2010
Uniaxial compressive strength	ISRM 1979
Modulus of elasticity	ISRM 1979
Tensile strength (Brazilian)	ISRM 1978

The laboratory test results of the different rock block specimens were managed separately during the statistical analyses.

3.2. Determining aggregate properties

The aggregate quality was described using micro-Deval tests (relevant standard: EN 1097-1:2012). The aggregate samples were made from two fractions representing two ranges of grain-size: 150g 10.0/11.2 mm and 350g 11.2/14.0 mm were used for the tests. The samples were tested in steel drums. The abrasive impact is provided by 5000g steel balls and 2500ml water in the drums. The result of the test is the micro-Deval coefficient (MDE), which is the ratio of

the abraded mass and the original mass of the aggregate sample. The relevant standard recommends 12,000 rotations to determine the MDE coefficient. The long-term durability was analysed using further rotations: the number of rotations was increased to 120,000 in 11 steps. The micro-Deval tests were made in pairs, thus two test results represent one experimental setting, i.e. number of rotations.

3.3. Methods of the statistical analyses

The laboratory test results provided the input parameters for the statistical analyses. Calculations were made using IBM SPSS Statistics and Microsoft Excel software. Following the exploratory statistical approach suggested by Kovács et al. (2012), one- and two-variable analyses were also conducted. First descriptive statistics were obtained (mean, minimum, maximum, range, standard deviation, coefficient of variation). The one-variable analyses included the selection of the extreme overhanging elements of the database. Boxplot diagrams were used for this. The elements were excluded from the further analyses if the difference between them and the quartiles was more than three times that of the interquartile range (Sajtos & Mitev, 2007). The distributions of the rock mechanical parameters were compared graphically with the normal distribution. Correlation and regression analyses were conducted within the framework of the two-variable statistical analyses. A relationship between rock mechanical parameters was qualified as strong relationship if $R > 0.7$ (Kovács et al., 2012; Sajtos and Mitev, 2007). Regression analyses were used to find suitable mathematical models which describe the behaviour of the aggregates during the long-term durability tests.

4. RESULTS AND DISCUSSION

4.1. Results of the statistical analyses of the rock mechanical parameters

According to the one-variable analyses the measured parameters of the andesite from Recsk fall within a narrow range, just a few data points were excluded from the further research. The number of the extreme overhanging elements of the

rock mechanical parameters of the andesite from Nógrádkövesd is higher. The measured parameters of both andesites follow normal distribution.

The mean values of the strength parameters didn't change significantly when the specimens were tested in different conditions (**Table 2**). In the course of the two-variable analyses the laboratory test results in different conditions were not separable.

Table 2: The mean values of the strength parameters of the studied andesites subjected to a-d: air-dry, w-s: water saturated, f-t: freeze-thaw conditions.

Rock mechanical parameters	Conditions	Recsk	Nógrádkövesd
Uniaxial compressive strength [MPa]	a-d	270.8	85.4
	w-s	265.7	76.0
	f-t	244.8	78.0
Modulus of elasticity [GPa]	a-d	48.7	22.5
	w-s	48.2	28.8
	f-t	47.0	23.8
Tensile strength [MPa]	a-d	9.8	7.9
	w-s	6.4	6.4
	f-t	8.3	5.8

The relationships between the determined parameters were analysed employing regression analyses. A linear correlation was found between the bulk densities in dry and water saturated condition (Nógrádkövesd: $R^2 = 0.95$, Recsk: $R^2 = 0.97$) and between the uniaxial compressive strength and the modulus of elasticity of the andesite from Nógrádkövesd ($R^2 = 0.55$). There is no correlation between the uniaxial compressive strength and the MDE coefficient of the studied andesites.

4.2. Results of the long-term micro-Deval tests

The long-term abrasive impact caused mass reduction of the aggregate and the grains became subspherical. The long-term durability test results fit well to quadratic curves, and the coefficient of correlation reached 0.96 in both cases (**Figure 1**). The IBM SPSS Statistics and Microsoft Excel programs facilitate the fitting of cubic functions without a constant member. If the functions contain

constant members the Pearson coefficients are 0.998 (Nógrádkövesd) and 0.985 (Recsk).

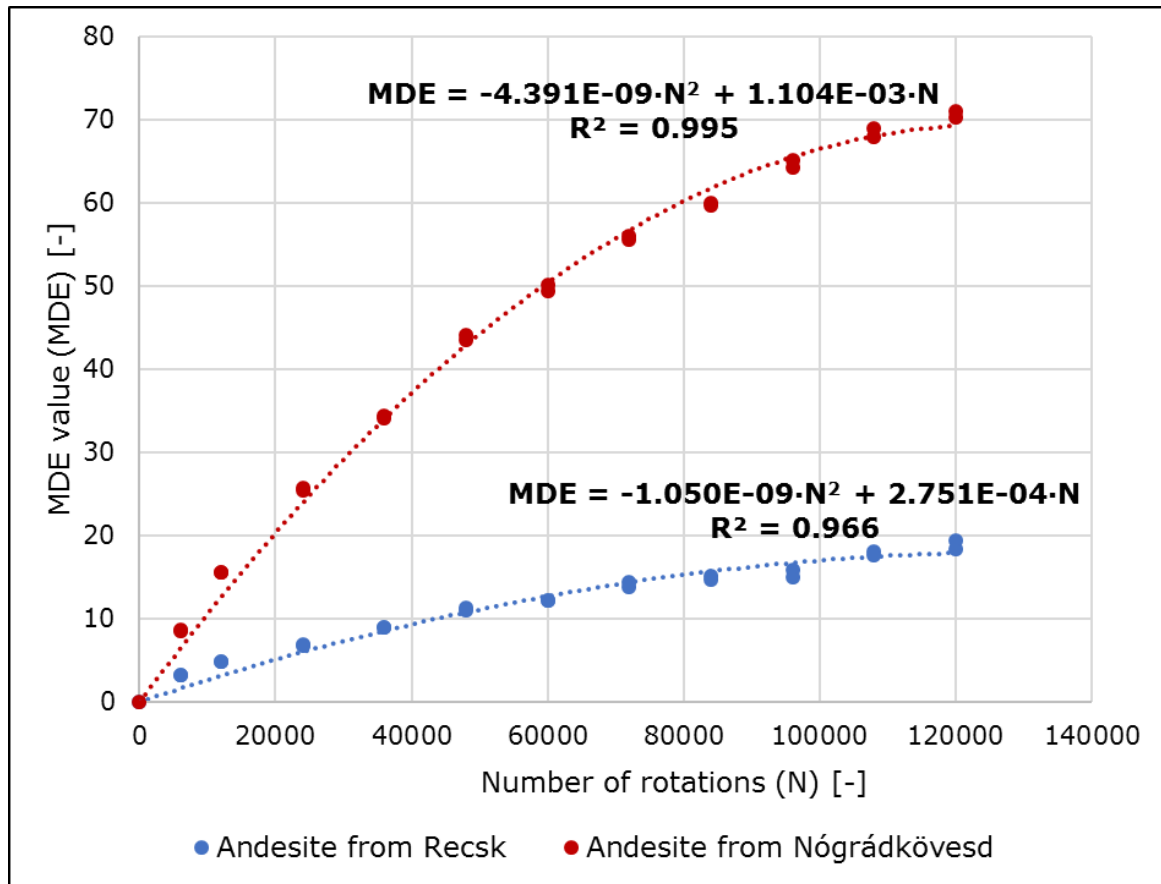


Figure 1: The MDE values according to the number of rotations.

The long-term micro-Deval test results (**Table 3**) show significant difference between the long-term behaviour of the two studied andesites. The abraded mass of the andesite from Nógrádkövesd reaches 70% of the original mass at 120,000 rotations.

Table 3: The long-term micro-Deval test results of the andesites from Recsk and Nógrádkövesd.

Number of rotations	MDE value	
	Recsk	Nógrádkövesd
12,000	4.9	15.6
60,000	12.2	49.8
120,000	18.9	70.6

5. CONCLUSIONS

The andesite from Recsk possessed higher uniaxial compressive strength (270 MPa) than the andesite from Nógrádkövesd (85 MPa). In the framework of the statistical analyses of the laboratory test results a linear correlation was found between the bulk densities in air-dry and water saturated conditions of the studied andesites. In addition a linear correlation was found between the uniaxial compressive strength and the modulus of elasticity of the andesite from Nógrádkövesd.

According to the results of the long-term micro-Deval tests quadratic models are suitable to describe the behaviour of the studied andesite aggregates in the course of continuous abrasive impact. The long-term durability properties of the andesite from Recsk is preferable to the properties of the andesite from Nógrádkövesd.

ACKNOWLEDGEMENTS

The authors wish to acknowledge their participations in the laboratory tests for Gyula Emszt and Bálint Pálinkás. The authors are thankful to László Ézsiás and the COLAS Észekő Ltd. for providing andesite blocks for the laboratory tests.

REFERENCES

- ERICHSEN, E. (2015): Plotting aggregate degradation results from the Los Angeles test on a triangular diagram: proposal of a new quality ranking for aggregates. *Bulletin of Engineering Geology and the Environment*, 74, 667-671.
- KOVÁCS, J., TANOS, P., KORPONAI, J., KOVÁCSNÉ SZÉKELY, I., GONDÁR, K., GONDÁR-SŐREGI, K. & HATVANI, IG (2012): Analysis of water quality data for scientists. –In: VOUDOURIS, K., VOUTSA, D. (ed.): *Water Quality Monitoring and Assessment*, InTech Open Access Publisher, Rijeka, 614 p.
- KRUTILOVÁ, K. & PŘIKRYL, R. (2017): Relationship between polished stone value (PSV) and Nordic abrasion value (AN) of volcanic rocks. *Bulletin of Engineering Geology and the Environment*, 76, 85-99.

RIGOPOULOS, I., TSIKOURAS, B., POMONIS, P. & HATZIPANAGIOTOU, K. (2013): Determination of the interrelations between the engineering parameters of construction aggregates from ophiolite complexes of Greece using factor analysis. *Construction and Building Materials*, 49, 747-757.

SAJTOS, L. & MITEV, A. (2007): SPSS Kutatási és Adatelemzési Kézikönyv. [*SPSS Manual for Research and Induction – in Hungarian*] Alinea Kiadó, Budapest, 404 p.

TÖRÖK, Á. (2015): Los Angeles and micro-Deval values of volcanic rocks and their use as aggregates, examples from Hungary. –In: LOLLINO, G., MANCONI, A., GUZZETTI, F., CULSHAW, M., BOBROWSKY, P. & LUINO, F. (eds.): *Engineering Geology for Society and Territory - Volume 5: Urban Geology, Sustainable Planning and Landscape Exploitation*, Springer International Publishing, 1400 p.

Evaluation of curve characteristics of shear strength along discontinuities using state-of-the-art breakpoint detection methods

Nikoletta Rozgonyi-Boissinot¹, Ildikó Buocz¹, István Gábor Hatvani^{2*}, Ákos Török¹

¹ Department of Engineering Geology and Geotechnics, Budapest University of Technology and Economics, H-1111 Budapest, Műegyetem rkp. 3., rozgonyi.nikoletta@epito.bme.hu,
torokakos@mail.bme.hu

² Institute for Geological and Geochemical Research, Research Center for Astronomy and Earth Sciences, Hungarian Academy of Sciences, H-1112 Budapest, Budaörsi út 45., hatvaniig@gmail.com

The detection of the main characteristics of shear strength along discontinuities is not straightforward. Such curves consist of a quasi linear section, depending on the rock assessed, followed by a “kick” representing the peak shear strength, and a residual part mostly parallel to the abscissa. In general, the end of the rising section and the peak shear strength are determined visually by experts. The aim of the present study is to facilitate the detection of these crucial characteristics. Therefore, breakpoint detection methods (Cross-Entropy, Change Point Model) were applied to laboratory test curves describing the shearing along discontinuities of 9 Mont Terri claystone samples. This is one of the target rocks for radioactive waste disposal in Switzerland. Results indicated that the end of the rising section and the kick can be effectively approximated. This can be considered as one of the first steps towards the automation of the evaluation of shear strength curves along discontinuities.

Key words: *automation, break point detection, claystone, shear strength along discontinuities*

1. INTRODUCTION

The shear strength of rock discontinuities can be determined by two types of test: those conducted under constant normal loading conditions, and those under constant normal stiffness loading conditions (Mularha et al., 2014). In the case of the former, it is the shear resistance of the discontinuity at the given normal load which is tested, rather than the joint strength of the discontinuity. This test is used to understand rock slope stability and the stability of

excavations. In the case of the latter, the ultimate shear strength of a joint is determined, because the dilation of the discontinuity is constrained while sliding. This test is used in the understanding of excavation problems.

The problem of the automation of rock mechanical data evaluation is to this day unsolved. The different rock types and the varying mechanical behavior of rock beds, even those of the same rock type, still make the automatic evaluation of strength curves difficult. Under normal circumstances, these curves consist of varying nonlinear and linear sections with one or more breakpoints. The present study therefore chose as its aim the suggestion and presentation of a new approach to facilitate the automation of determining the linear section(s) and the possible breakpoint(s) on the outputs of direct shear test results along discontinuities.

2. MATERIALS AND METHODS

2.1 Geological setting and materials

The studied samples belong to the Opalinus Claystone Formation from Mont Terri (MT), Jura Mountains of the Swiss Alps (**Figure 1**). The Aalenian (Middle Jurassic) Opalinus Claystone forms a part of the Folded Jura system and is covered by Middle Jurassic limestone and Late Jurassic clay and limestone.

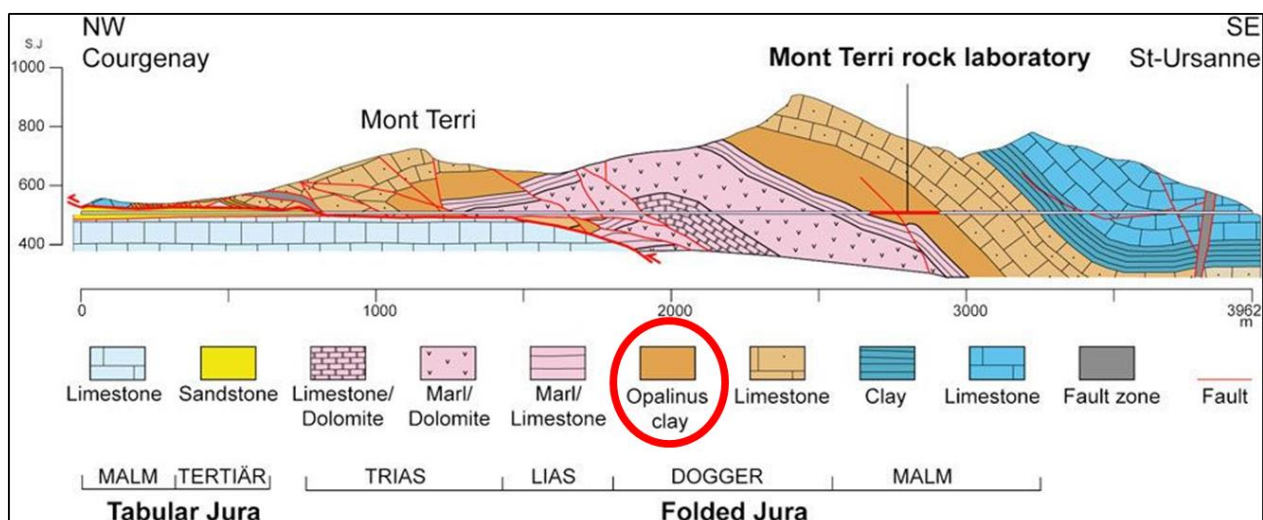


Figure 1: The strike-slip tectonic in analysed hydrocarbon zone (taken from Nussbaum & Bossart, 2008)

The folding is related to the Alpine orogeny (Upper Miocene-Lower Pleistocene). The Opalinus Claystone is an over-consolidated clay, which has an apparent thickness of 160 m and an actual thickness of 90 m nowadays (Nussbaum & Bossart, 2008).

2.2 Test method and data

In this research 9 shear strength tests of discontinuities under 1 MPa constant normal load on Claystones from Mont Terri were studied. The testing method was based on the International Society for Rock Mechanics (ISRM) guidelines (Mularha et al., 2014). The shear displacement was controlled by a stepper motor set to 0.8 mm min^{-1} and the vertical and horizontal displacements were detected. By evaluating the data measured, the relations between shear stress and shear displacement were determined (Buocz et al., 2015). With the help of 3D surface detection results, a regression plane was calculated for each sample surface. Using 3D surface roughness detection methods, smooth and moderately rough types of surface roughness of the claystone discontinuities were differentiated (Buocz et al., 2014; 2016). Thus, the tested specimens belong to one lithology, but two groups, so far as discontinuity surfaces are concerned. The ‘smooth’ ones are characterized by surface irregularities smaller than 0.25 mm from the regression plane, while the ‘moderately rough’ ones have surface irregularities between 0.25 and 1 mm measured from the regression plane.

2.3 Preprocessing and break point detection

Besides the original data, the first differences of the shear stress and shear displacement curves were determined for each point, so if the gradient of the linear slope is quasi-constant, a change point will be found at the end of the linear section and/or around the peak shear strength.

2.3.1 Multiple Breakpoint Detection

The Multiple Breakpoint Detection method is based on a modified cross-entropy (CE) approach using a modified Bayesian Information Criterion to estimate both the number and the corresponding locations of the breakpoints. It uses an

iterative optimization procedure, to detect changes in the mean and/or in the variance (for details see Priyadarshana and Sofronov, 2015). Due to continuous data the `CE.Normal.MeanVar` function was used with `distyp` set to 2, the maximum number of breakpoints (N_{\max}) was set to five, sample size (M) was left on default (200) and $\rho=0.25$. All other options were left as defaults. It has been observed (Topál et al., 2016) that after each run the CE method produces a slightly different result due to its stochastic characteristics, thus in every case it was run 100 times and out of the distribution of results, the 3 most abundantly occurring breakpoints were taken as the final result.

2.3.2 Change point model framework

The Change Point Model (CPM) framework was developed to detect change points in variance and/or the mean by evaluating the distribution of the data before and after a given point in the data stream using a two sample hypothesis test (Shirvani, 2015). In the present study, due to the non-normal characteristics of the shearing the obtained data were non-parametric. Hence, the Mann-Whitney test was applied within the CPM framework (CPM-MW), tuned to detect changes in the mean (Ross, 2011). The Average Run Length (ARL_0) was set to 200, corresponding to $\alpha=0.95$, the startup sequence (s) was $n/1.5$, where n denotes the length of the data stream.

Both methods (CE and CMP) are capable of handling a stream of observations and multiple breakpoints, and it should be noted that based on experience (Topál et al., 2015; 2016) the locations of the breakpoints indicated by the tests give the locations of the last datum occurring in the sequence before the particular change takes place. In the case of the CE method (Priyadarshana & Sofronov, 2015) the `CE.Normal.MeanVar` function of the `breakpoint` package, while in the case of the CPM (Ross, 2015) the `processStream` function of the `cpm` package was used both implemented to the R statistical environment (R Core Team, 2014).

3. RESULTS AND DISCUSSION

The CPM-MW method on the original data usually approximated the end of the linear rising section (**Figures 2 & 3**), but found a few additional unexplainable breakpoints as well. Where the amount of data describing the residual section of the strength curve was sufficiently dense, the CPM-MW method placed a breakpoint at the end of the linear rising section of the original data in the case of both surfaces (**Figures 2 & 3**). It is basically insignificant from an engineering geological aspect whether the breakpoint is placed at the beginning or the end of the residual section, because the important result is that the BP was found in the residual section (quasi constant shear strength on the ordinate) rather than the BP's location on the abscissa.

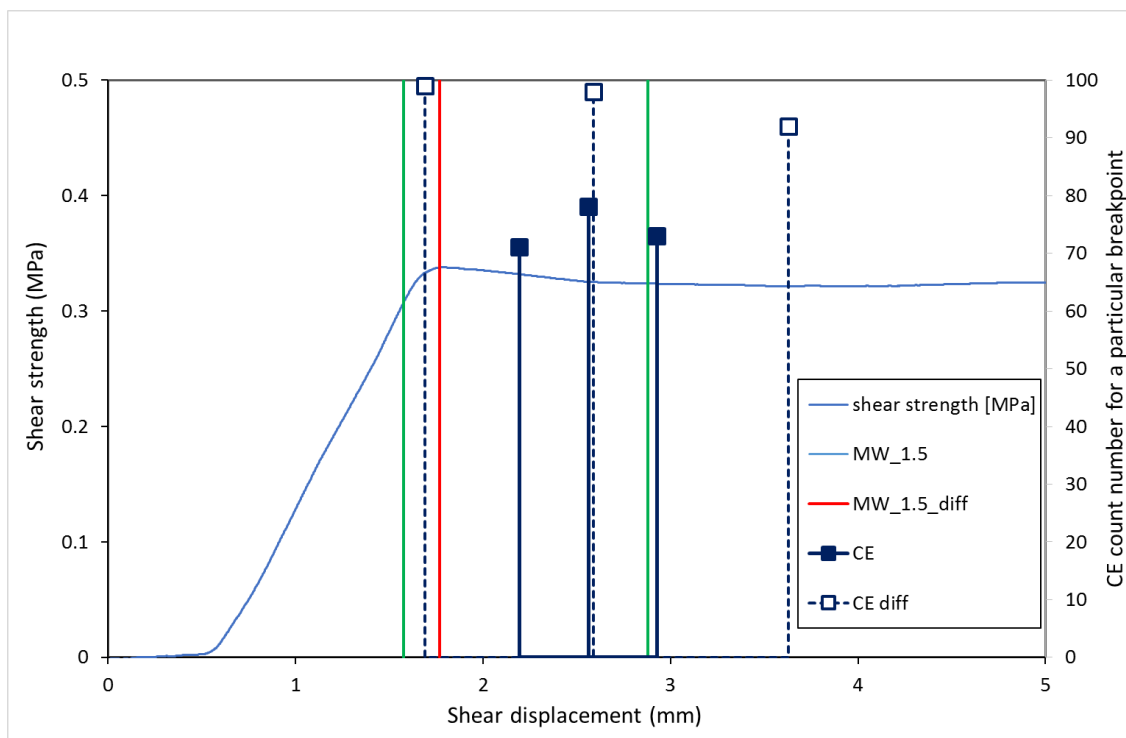


Figure 2: Breakpoints detected on the S8-1 smooth surface MT claystones. For further description, the reader is referred to the figure legend

On the first differences, however, it accurately found one breakpoint, indicating the peak shear strength for the smooth (**Figure 2**) and moderately rough (**Figure 3a**) types of surface roughness, except for in the case of one of the moderately rough samples (**Figure 3b**).

The CE method was less successful in outlining the linear section and finding the peak. Its results on the original data were meaningless, while on the first differences, although it did point out the peak, the overall picture obtained was confusingly unclear. Thus, while it can provide help in evaluating these curves, it is definitely not the final solution of the automation problem.

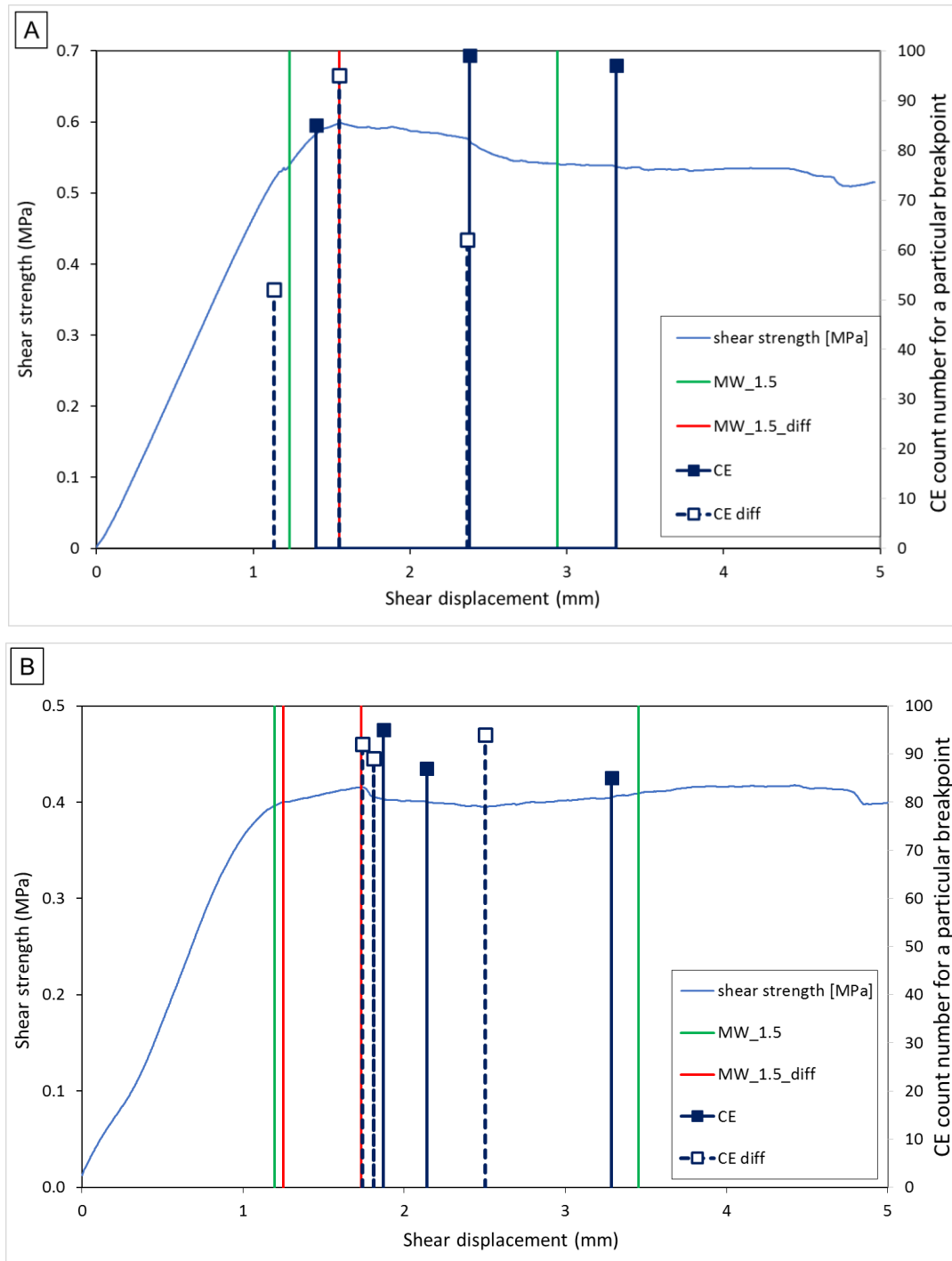


Figure 3: Breakpoints detected on the S7-3 A) and S9-2 B) moderately rough surface MT claystones. For further description, the reader is referred to the figure legend.

4. CONCLUSIONS AND OUTLOOK

The study presents a new direction in facilitating the otherwise empirical practice of evaluating shear strength curves along discontinuities. On the one hand, the Cross-Entropy method appeared to be unsuccessful in aiding the procedure, thus further tests and research are needed. On the other hand, however, the CPM-MW method approximated the end of the linear sections, and found the peak shear strength precisely on the first differential of the shear stress and shear displacement curves. Thus, we think this result can be considered as one of the first steps towards automating the procedure by further investigations of evaluating these curves.

ACKNOWLEDGMENTS

Thanks for the support of the MTA “Lendület” program (LP2012-27/2012) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

- BUOCZ I., ROZGONYI-BOISSINOT, N., . & TÖRÖK, Á. (2017): Influence of Discontinuity Inclination on the Shear Strength of Mont Terri Opalinus Claystones. Period. Polytech. Civil Eng., paper 10017. DOI: 10.3311/PPci.10017
- BUOCZ, I., ROZGONYI-BOISSINOT, N., & TÖRÖK Á., (2015): Shear strength behavior of Mont Terri Opalinus Claystone in fault zone. Pollack Periodica, 10/1, 31–38
- BUOCZ, I., TÖRÖK, Á., ZHAO, J. & ROZGONYI-BOISSINOT N. (2014): Direct shear strength test on Opalinus Clay, a possible host rock for radioactive waste, –In: LOLLINO G., GIORDAN D., THURO K., CARRANZA-TORRES C., WU F., MARINOS P., DELGADO C. (eds.): Engineering Geology for Society and Territory 6, 901-904
- MURALHA, J., GRASSELLI, G., TATONE, B., BLÜMEL, M., CHRYSSANTHAKIS, P., & YUJING J. (2014):: ISRM Suggested Method for Laboratory Determination of the Shear Strength of Rock Joints: Revised Version. -In: ULUSAY R. (ed.): The

ISRM Suggested Methods for Rock Characterization, Testing and Monitoring: 2007–2014. 131-142.

NUSSBAUM, C., BOSSART, P. (2008): Geology. -In: BOSSART P. and THURY M. (eds.): Mont Terri Rock Laboratory Project, Programme 1996 to 2007 and Results, Rep. Swiss Geol. Surv., 3,; 29-38

PRIYADARSHANA, W.J.R.M. & SOFRONOV G. (2015): Multiple Breakpoints Detection in array CGH Data via the Cross-Entropy Method, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12, 2, 487-498

R CORE TEAM (2014): R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

<http://www.R-project.org>

Ross, G.J. (2015): Parametric and Nonparametric Sequential Change Detection in R: The cpm Package, Journal of Statistical Software

SHIRVANI, A. (2015): Change point analysis of mean annual air temperature in Iran, Atmospheric Research, 160. 91-98

TOPÁL, D., HATVANI, I.G., MATYASOVSKY, I. & KERN, Z. (2015): Break-point detection algorithms tested on artificial time series. -In: HORVÁTH, J., CVETKOVIĆ, M., HATVANI, I.G. (eds.): 7th Croatian - Hungarian and 18th Hungarian Geomathematical Congress: "The Geomathematical Models: The Mirrors of Geological Reality or Science Fictions?". Conference venue: Hungarian Geological Society, Mórahalom, 147-154.

TOPÁL, D., MATYASOVSKY, I., KERN, Z. & HATVANI, I.G. (2016): Detecting breakpoints in artificially modified and real-life time series using three state-of-the-art methods, Open Geosci., 8, 78–98

The effect of JRC value on shear strength along discontinuities in Hungarian sandstones

Alina Vattai¹, Nikoletta Rozgonyi-Boissinot¹

¹ Department of Engineering Geology and Geotechnics, Budapest University of Technology and Economics, H-1111 Budapest, Műegyetem rkp. 3., rozgonyi.nikoletta@epito.bme.hu,
vattai.alina@epito.bme.hu

Shear strength along discontinuities has been investigated for decades. It is one of the most common parameters employed in the design of models of slopes, quarries, tunnels or indeed rocks surrounding of almost every kind of engineering structure. Many factors have an effect on the value of shear strength along discontinuities, including joint roughness coefficient (JRC), joint compressive strength (JCS), friction angle, asperity angle, filling, dilatation, and so on. There are several methods for the detection their values. In this study the adaptability of failure criteria developed by Barton and Choubey is investigated in the case of low-strength sandstones typical of those found in Hungary. The detection of the correct value of the JRC is quite subjective and difficult in many cases. Changes in Barton and Choubey's failure curves, based on values of shear strength along discontinuity measured in laboratory conditions, were investigated as a function of the JRC values obtaining between probable minimum and maximum values detected by visual comparison, employing ten standard profiles given by Barton and Choubey.

Key words: *shear strength along discontinuities, failure criterion, joint roughness coefficient, Hungarian sandstone*

1. INTRODUCTION

An essential part of the design process in underground engineering structures such as dam foundations, storage reservoirs or tunnels is knowing how the surrounding rock behaves. One of the main parameters used for modelling such rock engineering problems is shear strength along discontinuities. Estimating this parameter is also required as part of the stability assessment of rock slopes and quarry walls. In rock engineering practice, the value of shear strength along discontinuities can be predicted in a theoretical or empirical way. Several models have been developed by researchers. The most common empirical models are based on the Mohr-Coulomb theory:

(1)

$$\tau = \sigma_n * \operatorname{tg}(\varphi)$$

where τ, σ_n, φ are shear strength, normal stress, and friction angle, respectively. These models in question are: Patton's (1966), which is developed with angle of saw tooth of the surface of discontinuity; that of Ladanyi and Archambault (1970), in which the value of shear strength of discontinuity is estimated in an energetic way; Jaeger's (1971), where shear stress comes from both friction and failure; and the model of Barton and Choubey (1977):

(2)

$$\tau = \sigma_n * \operatorname{tg}[\varphi_r + JRC * \log_{10}\left(\frac{JCS}{\sigma_n}\right)]$$

where $\tau, \sigma_n, \varphi_r, JRC, JCS$ are shear strength, normal stress, residual friction angle, joint roughness coefficient, and joint compressive strength, respectively. An accurate value for the JRC is quite difficult to determine, and several methods have been developed for estimating it from the profile geometry. The fractal dimension of roughness profiles of a joint surface has been suggested for its estimate by many researchers but one of the most common methods is given by Barton and Choubey, and is based on visual comparison with standard profiles. This method is simple to use but quite subjective.

2. MATERIALS AND METHODS

2.1. Sandstone

In this study core samples from bore-holes situated in the area of Sopron, a town in the west of Hungary, were analysed. The samples were characterised as compact, fine-grained and medium-grained, gravelly sandstone (calcarenite) coming from sandstone layers at depth of 30 metres from the surface.

2.2. Methods

In this study the empirical failure criterion developed by Barton and Choubey is used. According to this model shear strength along discontinuities defined as a function of normal stress, joint roughness, joint compressive strength and friction angle. The most common method for the shear testing of discontinuities is the direct shear test, carried out in the field as in-situ shear testing or in laboratory conditions. In both cases the shear test is carried out with constant normal load or constant normal stiffness. In this paper, the results of direct shear tests carried out in laboratory conditions at constant normal loads were analysed. The laboratory test samples ranged from approximately 10 x 10 centimetres, while normal stress ranged between 0.08 and 0.23 MPa. Concrete was used for fixing both the lower part and the upper part of the specimen inside the shear test equipment. During the test process, the lower part of the shear box is fixed in all directions and the upper part can move in the direction of shearing. During the process shear displacements, normal stresses and shear stresses are detected continuously. A graph of shear stresses and normal stresses versus shear displacements is plotted.

The effect of joint roughness is built into the Barton and Choubey's formula (**Eq.2**) as the joint roughness coefficient (JRC), which is quite subjective and difficult to estimate accurately. Many methods have been developed for estimating the value of JRC. One of the most common ways of estimating is measuring the profiles of joint surface with a special profilometer. This method is based on a visual comparison with ten standard profiles (each profile represents three values, e.g. 18-20) given by Barton and Choubey. It is simple

to use, but it is quite difficult to estimate the correct average JRC value of the joint surface because visual comparison is subjective and three values are associated with each profile. In this study four profiles of each specimen were measured and integer JRC values (between the probable minimum and maximum, based on visual comparison) were applied using the empirical method instead of applying the average JRC value of surface of each specimen.

The value of joint compressive strength (JCS) in the Barton and Choubey's formula (**Eq.2**) is normally detected using a Schmidt hammer but in case of the unweathered joint surface JCS is equal to the uniaxial compressive strength (UCS) of the rock mass (Barton 1976). Whereas all of the sandstone joints were unweathered, values of UCS measured by uniaxial compressive test under laboratory conditions were applied.

The value of residual friction angle is established from the incline of the graph of shear stresses versus normal stresses detected under direct shear testing in the laboratory. Differences between the characteristics of failure curves based on Barton and Choubey's model, and the curve derived from experimental data pairs of shear stresses and normal stresses demonstrate how the different JRC values gained by subjective detection change the range of the interval between the curve based on the empirical method and that based on experimental results. The applied values of parameters in model of Barton and Coubey are shown in **Table 1**.

Table 1: The applied values of parameters in model of Barton and Coubey

	Joint Roughness Coefficient	Joint Compressive Strength	Residual friction angle
	JRC [-]	JCS [Mpa]	ϕ_r [°]
Fine-grained sandstone	7-9	9.61	41
Medium-grained sandstone	16-18	4.96	48

3. RESULTS AND DISCUSSION

There are several methods for estimating the value of JRC: i) the comparison of the measured profiles of the specimens with standard profiles (Barton and Choubey 1977), ii) estimating the value of JRC as a function of amplitude of joint roughness surface (Barton 1982), iii) the use of fractal dimensions (M. Sanei et

al.), and iv) 3D scanning (Buocz et al., 2017). In this study the comparison method developed by Barton and Choubey was used.

The experimental shear stress versus normal stress, and the three empirical failure curves calculated by **Eq.(2)** of fine-grained sandstone are presented in **Figure 1**. The results of the test on medium-grained sandstone are different (**Figure 2**). The curves derived from experimental data pairs were calculated using a linear regression. In **Figure 1**, the experimental data points are in the same interval as the three failure curves based on three different low values of JRC (7;8;9, respectively). The three failure curves are parallel to each other, the shear stress differences in between close-by curves are less than 0.03 MPa, under 0.6 MPa normal stress.

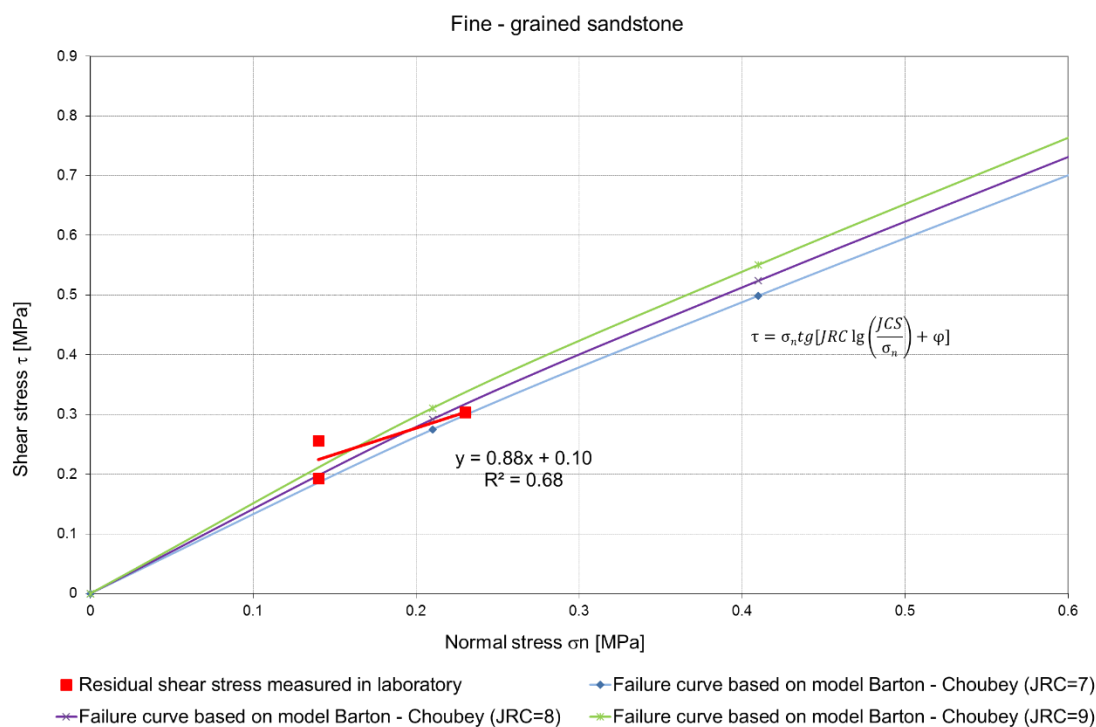


Figure 1: *Experimental shear stress versus experimental normal stress and three empirical failure curves calculated by Eq.(2) of a fine-grained sandstone sample*

The three failure curves with JRC values of 16;17; and 18, respectively, are parallel to each other with the shear stress difference approximately 0.05 MPa,

under 0.6 MPa normal stress (**Figure 2**), but consist of significantly higher values of shear stress than the detected data of direct shear test.

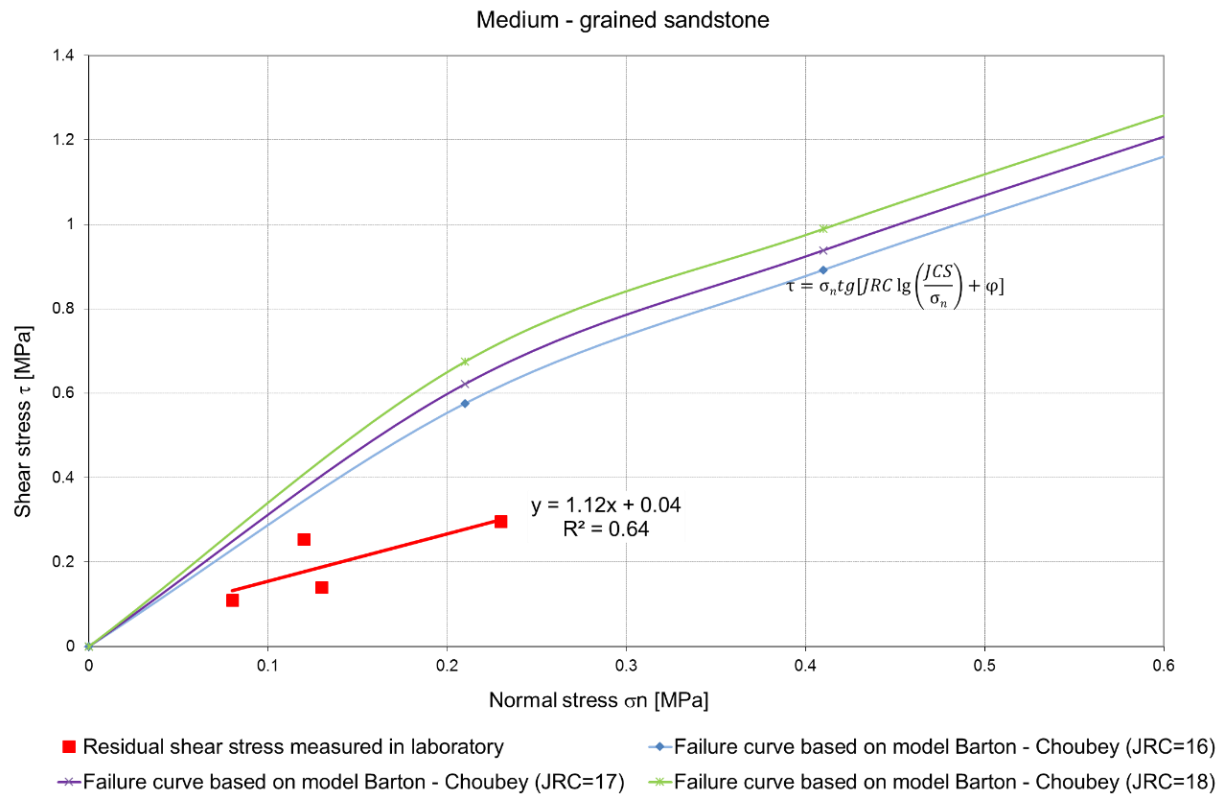


Figure 2: *Experimental shear stress versus experimental normal stress and three empirical failure curves calculated by Eq.(2) of a medium-grained sandstone sample*

4. CONCLUSIONS

The results of this study show that failure curves based Barton and Choubey's model (1977) using lower values of JRC measured on the surface of a fine-grained sandstone specimen shows a good fit to the curve consisting of experimental data pairs of shear stress and normal stress. To the contrary at higher values of JRC measured on the surface of a medium-grained sandstone specimen it is less consistent, although differences between own failure curves of both types of sandstone were not significant. Calculating values of JRC by using different methods can provide different results. Furthermore, changing in joint compressive strength and in friction angle have to be taken into account

when failure curves are analysed. For estimating the correct value of JRC and the other parameters and their effect on shear strength along discontinuities additional direct shear tests are required.

REFERENCES

- BARTON, N. (1976): The Shear Strength of Rock and Rock Joints, International Journal of Rock Mechanics and Mining Science & Geomechanics Abstracts 13(9):255-279
- BARTON, N. & BANDIS, S.C. (1982): Effects of block size on the shear behaviour of jointed rock, 23rd U.S. sym. on rock mechanics, Berkeley, 739-760
- BARTON, N., CHOUBEY, V. (1977): The shear strength of rock joints in theory and practice, Rock Mechanics 10, 1-54
- BUOCZ, I., ROZGONYI-BOISSINOT, N. & TÖRÖK, Á. (2017): Influence of Discontinuity Inclination on the Shear Strength of Mont Terry Opalinus Claystones, Periodica Polytechnica Civil Engineering, Online First (2017) paper 10017. DOI:10.3311/PPci.10017
- JAEGER, JC. (1971): Friction of rocks and stability of rock slopes, Géotechnique 21, 148-158
- LADANYI, B. & ARCHAMBAULT, G. (1970): Simulation of shear behaviour of a jointed rock mass, Proc.11 Symp. Rock Mech., AIME, New York, 105-125
- PATTON FD. (1966): Multiple modes of shear failure in rock and related materials, Proc.1. ISRM Cong. Lisbon 1: 509-513
- SANEI, M., FARAMARZI, L., FAHIMIFAR, A., GOLI, S., MEHINRAD, A. & RAHMATI, A., (2015): Shear strength of discontinuities in sedimentary rock masses based on direct shear tests. International Journal of Rock Mechanics and Mining Sciences 75, 119-131.
- VÁSÁRHELYI, B. (2016): Az alkalmazott kőzetmechanika alapjai [Basics of rock mechanics – in Hungarian]. Department of Engineering Geology and Geotechnics, Faculty of Civil Engineering, Budapest University of Technology and Economics, Budapest, 299 p.

Permeability of rock mass around a radioactive waste repository tunnel

Daniel Borbély¹, Tamás Megyeri², Péter Görög¹

¹ Budapest University of Technology and Economics, Department Of Engineering Geology And Geotechnics, borbely.daniel@epito.bme.hu

²Mott MacDonald Magyarország Kft

Upon completion, the National Radioactive Waste Repository in Bábaapáti will provide safe storage for low and medium-level radioactive waste. The emplacement chambers were excavated in fractured, blocky, granitic rock mass in app. 240 m below surface. One of the tasks related to the repository development is on the feasibility demonstration of the permanent repository closure including long term rock mass associated issues.

The required life time exceeds the usual life time of an engineering structure. Long term behaviour of the repository is needs to be extrapolated from observation in shorter time period or from analogous natural caverns. Geo-mathematical methods or numerical methods are the most promising techniques to carry out the extrapolation. It is commonly understood that there are significant uncertainties in long term predictions. Uncertainties can be mitigated by utilizing independent methods to assess long term behaviour and by improve prediction capability of the calculation model in short term.

Aim of the paper is to:

1. Create a numerical model effectively capture wide range of observed behaviour of the rock mass, including stress state, and stress dependent permeability
2. Identify possible cause of long term creep and show that the long term creep can be captured by the selected calculation method.

The moderately fractured rock mass is modelled using distinct element modelling code UDEC released by Itasca. The joints in the rock mass are explicitly modelled; the blocky nature of the rock mass is captured. The model is verified with actual field observations and monitoring results. Based on the predicted stress state of the rock mass, potential cause of long term creep is identified. By fulfilling the two aims explained above, it is concluded that the model can be used to extrapolate in time, serve as one possible estimation on the long term behaviour of the repository.

Key words: *radioactive waste repository, tunnel, monitoring, permeability*

1. INTRODUCTION

For rock engineers need to extrapolate small scale, short term lab test results to the actual extent and time scale of the project. In rock engineering practice 3 main methods are used: Empirical, analytical and numerical methods.

In cases where sufficient amount of previous experience is available, empirical methods can be used. Most widely used methods are Terzaghi rock load theory Terzaghi (1946), Rock Mass Rating (RMR) introduced by Bieniawski (1976), Barton et al (1974) Tunnelling Quality Index (Q) and Geological Strength Index (GSI) introduced by (Hoek, 1994).

In cases where joint properties, intact rock properties and scale effect are known, analytical tools can be used. Analytical relations can be derived for the elastic moduli of the fractured rock mass (Lekhnitskii, 1977 and Oda, 1982).

Recent years, numerical modelling, synthetic rock mass approach (Ivars et al., 2011) is a viable option to determine large scale or long term rock mass properties (Vallejos, et al. 2013). It takes under consideration stochastic discrete fracture networks, and complex intact rock properties such as brittle failure. Extrapolation of coupled hydro-mechanical properties was investigated by Min et al. (2014).

The required life time of the engineering barriers is way above the usual life time of an engineering structure. Long term behaviour of the repository is needs to be extrapolated from observation in shorter time period (decades). In the lack of previous experience, geo-mathematical methods and numerical models are the most promising techniques to carry out the extrapolation.

Aim of the paper is to set up a reasonable estimation on the construction events, compare model behaviour to observed behaviour of the actual rock mass and identify the cause/source of the time dependent displacements measured on site (Kovács et. al., 2013).

2. MATERIALS AND METHODS

2.1. Typical rock support

The present investigation was based on the measurements in National Radioactive Waste Repository in Bábaapáti, the I-K1 emplacement chamber. Rock class 2b (moderately fractured rock mass) was considered. The rock

support system consists of: 120mm steel fibre reinforced sprayed concrete lining with systematic 4m long rock bolting in a 1.5m x 1.5m raster.

2.2. Rock mechanical characterisation of surrounding ground

2.2.1. Intact rock properties

The results of systematic rock sampling are presented in Kovács, et. al. (2012) and summarized in **Table 2**. According to Damjanac and Fairhurst (2010) the intact rock creep does not occur below crack initiation limit. Considering minor cracks in the blocks, properties are calculated using the Hoek-Brown failure criteria assuming that the GSI value is equal to 85.

Table 2: Intact rock properties

Property name		Value	Unit
Uniaxial compressive strength	UCS	129	MPa
Crack initiation Limit	σ_{ci}	34.3	MPa
Hoek-Brown parameter	m_i	15	-
Disturbance factor	D	0	-
Young modulus	E	67000	MPa
Poisson ratio		0.232	-
Unit weight	γ	0.027	MPa/m ³

2.2.2. Joint properties

The behaviour of jointed rock mass (or an associated distinct element model) highly depends on the joint characteristics. The shear strength of the joints had been measured by laboratory tests (Buocz, et al., 2010) and had been verified by plain strain distinct element models (Horváth, et al., 2012).

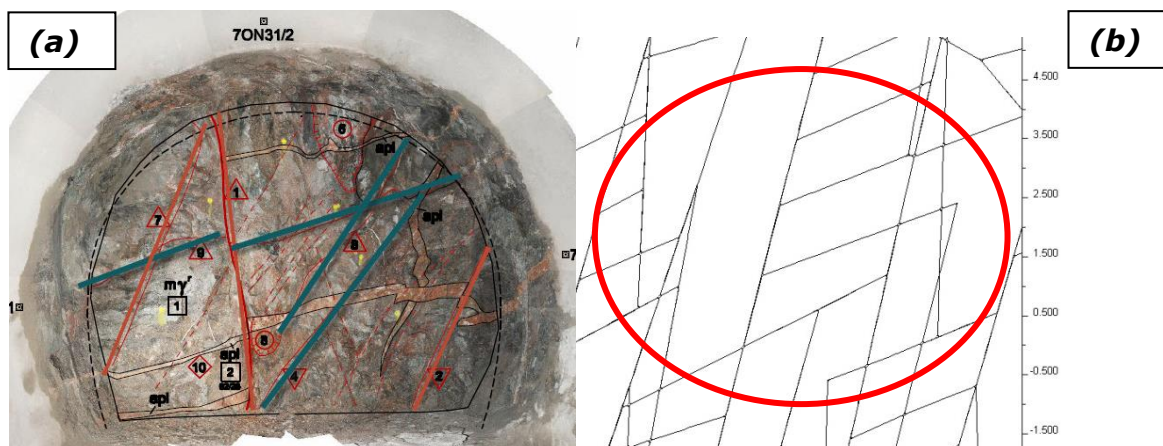


Figure 11: (a) Actual and (b) modelled joint pattern, red ellipse represents app. same area as the actual tunnel

2.2.3. Joint pattern

During excavation a systematic face mapping programme was performed, and the Q, GSI and RMR values were all determined. The joint patterns selected to characterize the rock mass is shown on **Figure 11**).

2.3. In situ tests

2.3.1. AE monitoring

The inelastic deformation in rock masses is usually coupled with acoustic emission events. To monitor the inelastic deformations, acoustic emission monitoring system was installed around the chambers (Szűcs & Bakai, 2012). The AE events are shown in **Figure 12**.

2.3.2. Hydraulic conductivity

The rock mass hydraulic conductivity was determined with field test. The field tests were verified with numerical modelling (Mező, 2010). According to the field measurements, the hydraulic conductivity of the rock mass is between $1.0 \cdot 10^{-12}$ m/s and $1.0 \cdot 10^{-6}$ m/s to with an average of $2 \cdot 10^{-9}$.

2.4. Numerical model

Two-dimensional distinct element code UDEC 5.0 is applied in numerical modelling. In order to appropriately represent the stress path and block movement induced by the chamber excavation the following main modelling stages were considered:

1. Set up the in-situ condition. The horizontal stress ratio is 1.4 at the tunnel axis, and 1.2 along the tunnel axis (Kovács, et al., 2012).
2. The tunnel is excavated in one go. The three-dimensional redistribution of the stresses were considered with a relaxation factor of 30%.
3. Rock support installation: the ageing of the sprayed concrete lining was considered in the model according to the paper of Chang and Stille (1993).

3. RESULTS AND DISCUSSION:

3.1. Displacements

As it is presented in Kovács, et. al. (2012), rock mass stiffness was determined using rock mass classification and was validated with convergence monitoring.

The displacements in the UDEC model were compared to a continuum finite element model generated by Phase2. The displacements are in good agreement: 2.0-3.6mm and 2.6-3.4mm in the UDEC model and in the Phase2 model respectively.

3.2. Acoustic emission monitoring

According to Damjanac and Fairhurst (2010) acoustic emission (AE) does not initiate in the intact rock if it is loaded under the crack initiation limit. The difference between the major and minor principal stress is below 23.5MPa in 100% of the zones and 15MPa in 99.75% of the zones. Crack initiation of the intact rock occurs around 34MPa, so the stresses are significantly below the crack initiation limit of the intact rock. According to these results, it can be concluded that the AE events are caused by the slips of the joints. The assumed AE activity in the model (app. 9-10m, see **Figure 2-a**) is in good agreement with the field observations (app. 10-12m, see **Figure 2-b**).

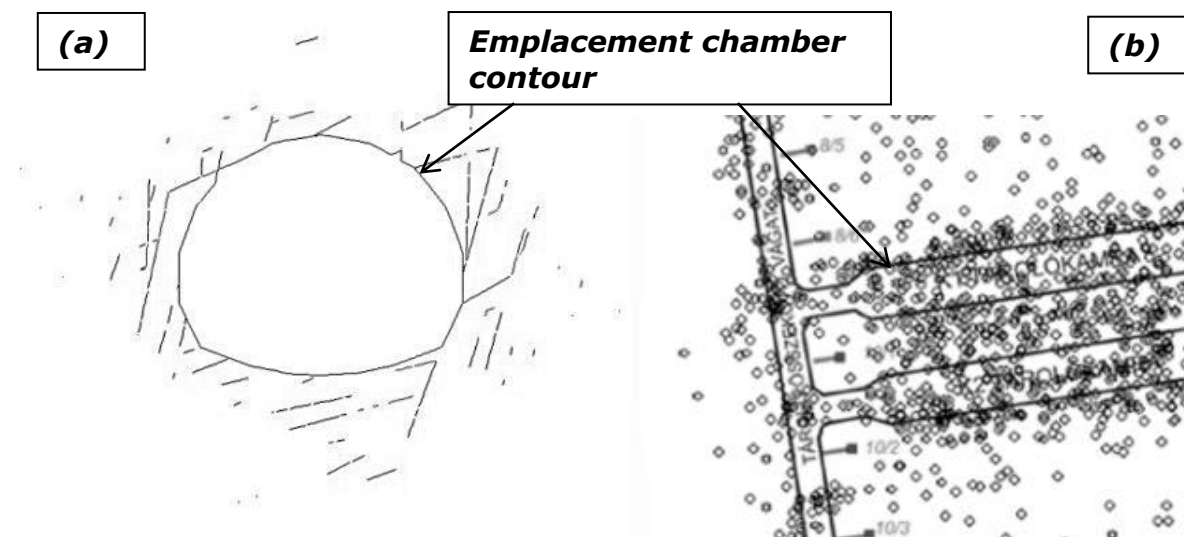


Figure 12: *(a) locations of AE events in the model, cross section of the emplacement chamber; (b) locations of AE events measured on site, plan view of the emplacement chamber*

3.3. Hydraulic conductivity

The equivalent hydraulic conductivity of the rock mass (considered in the UDEC model) was calculated using the crack tensor (Oda, 1985). The hydraulic

conductivity of an individual joint was determined according to Barton et. al. (1985). The hydraulic conductivity of the rock mass considered in the UDEC model is $1.76 \cdot 10^{-8}$ m/s. The average hydraulic conductivity of the actual rock mass is $2 \cdot 10^{-9}$ m/s (Mező, 2010). There is one order of magnitude difference in the modelled and measured hydraulic conductivity.

The hydraulic conductivity of each joint has been calculated according to Oda (1985). The aperture change during the excavation was directly calculated by a tailor made solution built in the distinct element code. The hydraulic conductivity of an individual joint is plotted on **Figure 13**. As shown, the hydraulic conductivity is increasing in the vicinity of the tunnel. At the stage of rock support system installation, the radius of the disturbed zone is within the extents of 2 tunnel radius measured from the chamber axis (**Figure 13**).

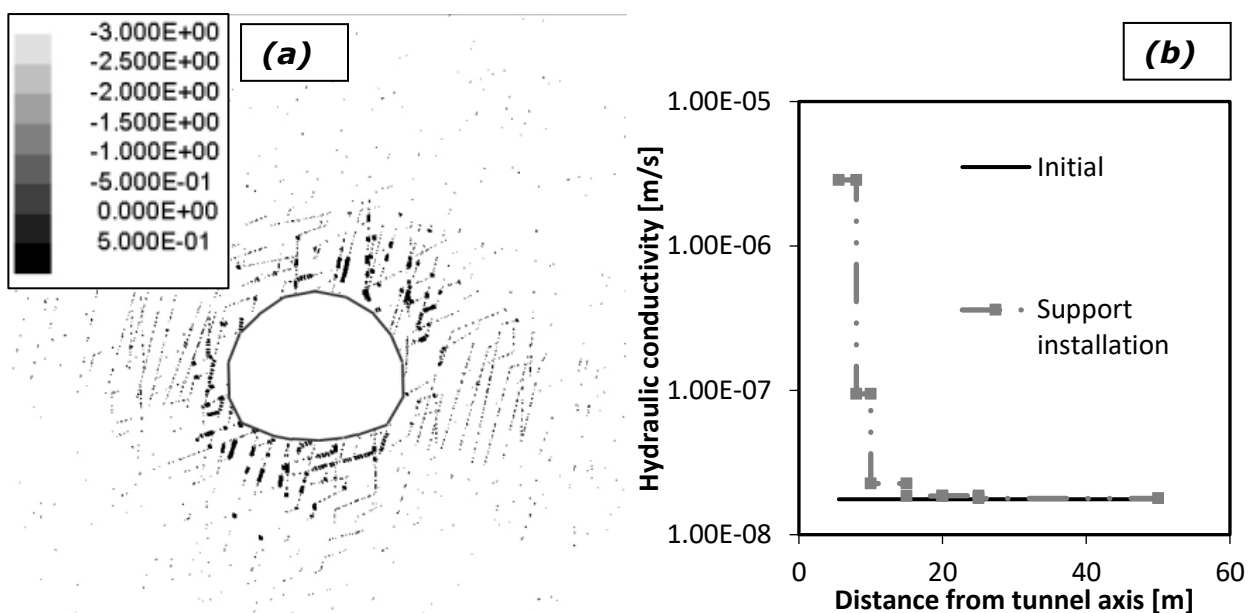


Figure 13: (a) Decimal logarithm of joint hydraulic conductivity [m/s] of an individual joint; (b) equivalent conductivity of the rock mass

It has been noted that the model overestimate the change in hydraulic conductivity induced by tunnelling. However, the predicted tendencies are believed to be a faithful representation of the phenomenon, and a promising method to compare backfilling options.

4. CONCLUSIONS

Results are significant, because models are a starting point to the extrapolation for further long term investigations. Cause/source of the time dependent displacements are identified. Based on the research presented here, joint creep is the primary source of long term displacements. Next steps for the current research are to carry out joint creep tests and implement creep model to a numerical model. Modelling results can then be compared with the experience gained in the first few years (Kovács, et al., 2013).

The inherent uncertainty of long term predictions can be reduced by using independent methods, with various assumptions. The presented research aimed to be one of the several valid methods.

REFERENCES

- BARTON, N., BANDIS, S. & BAKHTAR, K. (1985): Strength, deformation and conductivity coupling of rock joints. *International Journal of Rock Mechanics and Mining Sciences*, 22(3), 121-140.
- BIENIAWSKI, Z.T. (1976) Rock mass classification in rock engineering. In *Exploration for rock engineering, proc. of the symp.*, (ed. Z.T. Bieniawski) 1, 97-106. Cape Town: Balkema
- BUOCZ, I., ROZGONYI-BOISSINOT, N., GÖRÖG, P. & TÖRÖK, Á. (2010): Laboratory determination of direct shear strength of granitoid rocks; examples from the host rock of the nuclear waste storage facility of Bábaapáti (Hungary). *Central European Geology*, 405-417.
- CHANG, Y. & STILLE, H. (1993): Influence of early age properties of shotcrete on tunnel construction sequences. In *Shotcrete for Underground Support VI* pp110-117. American Society of Civil Engineers.
- DAMJANAC, B. & FAIRHURST, C. (2010): Evidence for a long-term strength threshold in crystalline rock. *Rock Mechanics and Rock Engineering*, 43(5), 513-531.
- HOEK, E. (1994) Strength of rock and rock masses. *ISRM News Journal*, 2(2). pp. 4-16.

- HORVÁTH, Z., MEGYERI, T., VÁRÓ, Á., & GÖRÖG, P. (2012) Discrete element modelling of the Mórággy Granite Formation in Southern Hungary. Horváth T (ed.) 1st Eastern European tunneling Conference, pp22. Veszprém.
- IVARS, D. M., PIERCE, M. E. & DARCEL, C., REYES-MONTES, J., POTYONDY, D. O., YOUNG, R. P., & CUNDALL, P. A. (2011): The synthetic rock mass approach for jointed rock mass modelling. International Journal of Rock Mechanics and Mining Sciences, 48(2), 219-244.
- Kovács, L., Deák, F., Somodi, G., Mészáros, E., Máté, K., Jakab, A., Vásárhelyi, B., GEIGER, J., DANKÓ, GY., KORPAI, F, MEZŐ, GY., DARVAS, K., VÁN, P., T., FÜLÖP & ASSZONYI, CS. (2012): Geotechnical investigation report. Pécs: RHK-K-032/12.
- KOVÁCS, L., MÉSZÁROS, E., DEÁK, F. & SOMODI, G. (2013): Time-dependent deformations in the underground openings of NRWR at Bataapáti and their possible causes. Budapest: Mérnökgeológia Közvetmechanika 2013.
- LEKHNITSKII, S. G. (1977): Theory of elasticity of an anisotropic body. Mir Publishers, Moscow.
- MEZŐ, G. (2010): Hydrogeologic modeling, FeFlow model (RHK-K-158/09. kötet). Paks: RHK.
- MIN, K., RUTQVIST, J., TSANG, C. & JING, L. (2004): Stress-dependent permeability of fractured rock masses: a numerical study. International Journal of Rock Mechanics and Mining Sciences, 41. kötet, pp. 1191-1210.
- ODA, M. (1985): Permeability tensor for discontinuous rock masses. Geotechnique , 35.4, 483-495.
- ODA, M. (1982): Fabric tensor for discontinuous geological materials. Soils and Foundations, Japanese Society of Soil Mechanics and Foundation Engineering, 1982;22(4):96-108.
- SZÚCS, I. & BAKAI, J. (2012): Acoustic Emission monitoring system in NRHT. „Geoscientific, mining and technological results of the construction of NRHT, first two chamber. (2012. JUN 13. Pécs) pp. 69-73.
- TERZAGHI, K. (1946): Rock defects and loads on tunnel supports. In Rock tunneling with steel supports, (eds R. V. Proctor and T. L. White) 1, 17-99. Youngstown, OH. Commercial Shearing and Stamping Company.
- VALLEJOS, J. A. (2013): Application of the Synthetic Rock Mass approach to characterize rock mass behavior at the El Teniente Mine, Chile. FLAC/DEM Symposium, Minneapolis. 2013

Part VI - System theory, future of geology & geoinformatics

The qualitative assessment of structured systems

Sándor Molnár

Szent István University, Department of Informatics, Molnar.Sandor@gek.szie.hu

In the following brief paper the main aspects of the qualitative assessment of structured systems will be discussed with a special focus on the development of verticum type of systems throughout the past cca. 30 years.

Key words: *linear systems, verticum type systems, Wei-Norman reorganisation*

1. INTRODUCTION TO HIERARCHICAL SYSTEMS

Hierarchical systems are one of the major foci of research in control and system theory from the middle of the 80's (Molnár & Szidarovszky, 1986; Molnár, 1988a; Molnár, 1988b; Molnár, 2000).

Economical, biological, engineering and mining systems are structured hierarchically by their fundamental characteristics due to specialisation, technological conditions, geographical attributes and technical limitations. Social considerations, economic policy or other reasons can divide economic systems further. The description of such systems can be very complicated and demanding. In this sense we can consider the restructuring of dynamic complex systems into simpler structure systems a very significant and important task from the mathematical and system theoretical aspect.

2. DISCUSSION ON THE DEVELOPMENT OF VERTICUM TY-OF SYSTEMS

We consider a system to be of simpler structure if it is of a verticum type see e.g. Molnár & Szidarovszky (1986). Such systems have a wide range of application and relevance in e.g. engineering sciences, like the modelling of an upstream or downstream oil industry, or in business or economic sciences where any extraction-production chain can be easily organised as a verticum-type system (Molnár, 1988b; Molnár & Szidarovszky, 1994; Molnár et al., 2003).

Recently the widespread applications of verticum type systems can be observed in the analysis and assessment of biological and environmental control systems (Molnár et al., 2007; Molnár et al., 2008; Molnár et al., 2012; Molnár et al., 2013; Molnár, 2013).

The principal challenge can be considered as follows. Starting from a parametric variable control system using algebraic and differential algebraic means the Wei-Norman reorganisation of the generalised verticum system should be achieved. It is straightforward to assume that the derivatives of the input variables can occur both in the state space control equations and the output equations. This provides the so called canonical Fliess-form which is a more general form than the canonical Kalman-form (Fliess, 1991; Sira-Ramírez & Agrawal, 2004). We can also mention linear systems with equations which have linear state variables and input and derivatives. An interesting result will be shown for the above which generalises the Kalman-type rank condition and the constant parametric linear systems given in a canonical Fliess-form.

The system examined is as follows:

$$\begin{aligned} dx &= Ax + \sum_{i=0}^I B_i d^i u \\ y &= Cx + \sum_{i=0}^I D_i d^i u. \end{aligned}$$

The sufficient and necessary condition of reachability is:

$$\text{rank} \left[\sum_{i=0}^I A^i B_i, A \left(\sum_{i=0}^I A^i B_i \right), \dots, A^{n-1} \left(\sum_{i=0}^I A^i B_i \right) \right] = n$$

This is the generalisation of the Kalman-type rank condition.

An especially important parametric variable control system is a specially structured system with a verticum-type hierarchy. This will be described with

a differential algebraic formalism. The special case with a state dependent parametrisation will be discussed using partial algebraic methodology.

REFERENCES

- FLIESS, M (1991): Controllability Revisited in Mathematical System Theory: The influence of R.E. Kalman, A.C. Antoulas (Ed.) Springer-Verlag , Berlin.
- SIRA-RAMÍREZ, H. & AGRAWAL, S. K. (2004): Differentially Flat Systems (Automation and Control Engineering), Marcel Dekker Inc., New York, Basel, p. 450, 2004. ISBN10: 0824754700, ISBN 13: 9780824754709
- MOLNÁR, S., LÓPEZ, I. & GÁMEZ, M. (2007): Observability and observers in a food web, Applied Mathematics Letters, Vol. 20, Issue 8, August 2007, pp. 951-957.
- MOLNÁR, S., GÁMEZ, M. & LÓPEZ, I. (2012): Observation of nonlinear verticum-type systems applied to ecological monitoring, International Journal of Biomathematics, Vol. 5, No. 6, pp. 1250051-1-1250051-15, 2012, DOI: 10.1142/S1793524512500519,
- MOLNÁR, S., GÁMEZ, M., LÓPEZ, I. & T. CABELLO (2013): Equilibrium control of nonlinear verticum-type systems, applied to integrated pest control, BioSystems, Vol. 113, pp. 72-80. DOI: 10.1016/j.biosystems.2013.05.005
- MOLNÁR, S., GAMEZ, M. & LOPEZ, I. (2008): Monitoring Environmental Change in an Ecosystem, Biosystems, Vol. 93. No. 3, pp. 211-217. ISSN 0303-2647
- MOLNÁR, S., SZIDAROVSKY, F. & OKUGUCHI, K. (1988): On a General Scheme in the Theory of Conflicts, Math. Anal. and System Theory, Vol. 5., pp. 31-37, (Marx Károly Közgazdaságtudományi Egyetem).
- MOLNÁR, S. & SZIDAROVSKY, F (1994): A dinamikus termelői-fogyasztói modell irányíthatóságáról, SZIGMA, Vol. 26. No. 1-2, pp. 49-54.
- MOLNÁR, S. & SZIDAROVSKY F. (1986): A Stochastic Multiobjective Dynamic Programming. Method with Application to Energy Modelling, Book Series: Lecture Notes in Control and Information Sciences, Book: System Modelling and Optimization, (Eds.: Molnár, S; Szidarovszky, F,) Springer, Vol. 84, pp. 601-609, Berlin/Heidelberg, 1986. ISBN: 978-3-540-16854-6, ISSN: 0170-8643, DOI:10.1007/BFb0043885

MOLNÁR, S., SZIDAROVSKY, F. & MOLNÁR, M. (2003): Controllability of Time-varying Oligopolies, Proceedings of the 4th International Conference on Control and Automation, 2003, Montreal, Canada, WA05-5, IFAC-IEEE, pp. 570-573. ISBN: 0-7803-7777-X, Library of Congress: 10-12 June 2003.

MOLNÁR, S. (2001): Időtől függő vertikum-típusú lineáris rendszerekről, in MTA Közgyűlési előadások, 2000 május II. kötet, pp. 645-657, 2001, MTA. ISSN: 1585-1915

MOLNÁR, S. (1988a): Observability and Controllability of Decomposed Systems I., Math.Anal. and System Theory, Vol. 5., pp. 57-66, (Marx Károly Közgazdaságtudományi Egyetem), 1988a.

MOLNÁR, S. (1988b): Realization of Verticum-Type Systems, Math. Anal. and System Theory, Vol. 5., pp. 11-30, (Marx Károly Közgazdaságtudományi Egyetem).

MOLNÁR, S., LOPEZ, I., GAMEZ, M. & GARAY, J. (2016): A two-agent model applied to the biological control of the sugarcane borer (*Diatraea saccharalis*) by the egg parasitoid *Trichogramma galloi* and the larvae parasitoid *Cotesia flavipes*, BioSystems, 141, 45-54. doi:10.1016/j.biosystems.2016.02.002

The 3rd/4th industrial revolution - Challenges for researchers

Viktor Feurer¹, Ph.D. Ferenc Fedor ²

¹ Intercomp Ltd, feurer.viktor@intercomp.hu

² Geochem Ltd, fedor.ferenc@geochem-ltd.eu

We live in the era of the 3rd and 4th industrial revolution and technology is advancing with a speed never seen before. The communication between the members of the economy is moving towards automated business-to-business systems. In the “new” industry, the big data heaps are valuable resources waiting to be processed by advanced systems more-and-more resembling artificial intelligence. The service providers must connect to the huge information transfer systems connecting their customers because the information must be present at the required place and time with the required quality. Deadlines getting shorter and not participating in this information flow will cause loss of business. Researchers working in the field of geosciences must be prepared for the following challenges (sometimes threats): (i) automatization: the deep/machine learning and the early artificial intelligence-like systems can speed up the data evaluation, increase the quality of the result and some processes can be completely automatized; (ii) integration: the devices and computers in a laboratory must be integrated into one network (Internet of Things), enabling the semi-autonomous operation and communication; (iii) standards: the business-to-business communication requires standards, some are already developed but there is a lot of work ahead; (iv) non-standard solutions: there can be business entities having so much power that they are able to enforce non-standard solutions and (v) security: in the era of Internet of Things the information security is definitely a very important issue and the valuable scientific data is the primary target of industrial espionage. The proper response for these challenges is to build integrated systems providing: highly automated processes to reduce the possible errors made by humans; rich business-to-business interfaces for communication and secure environment for producing and delivering valuable results for the customers. The SmartLab system which is under development at the Geochem Ltd. is designed to provide an affordable solution to the previously enumerate challenges and requirements.

Key words: *automatization, business-to-business communication, internet of things, information security*

1. INDUSTRIAL REVOLUTIONS

The first industrial revolution happened in the 18th/19th century, converting the former rural and agricultural society into a modern urban and industrial society. The main actors were the iron and textile industries while the main achievement was the steam engine.

The second industrial revolution at the beginning of the 20th century brought the era of electric power and mass production. The main actors were the steel- and oil industry while the main achievements were the telephone and the internal combustion engine.

The 3rd or “digital” industrial revolution began after World War II and converted analogue electronic and mechanical devices into digital technology. The main actor is the software industry, affecting everyone and everything. The main achievements are the personal computer and the internet.

The 21st century brought the 4th industrial revolution, although, the 3rd is not yet finished. Based on the digital technologies there are many interesting breakthroughs and completely innovative technologies created every day, such as in robotics, nanotechnology, biotechnology, augmented reality, Internet of Things (IoT) (Greengard, 2015).

This 4th industrial revolution is fundamentally different from the previous ones which were mainly “only” advances in technology. Now the efficiency of business and organisations are dramatically improved and asset management is more important than any time before. The impact of this industrial revolution changes everything, not just how the industry works, it changes our everyday life with a speed never seen before. The Industry 4.0 (Schwab, 2016) is a product of this revolution.

2. INDUSTRY 4.0

The Industry 4.0 (Schwab, 2016) is an emerging new concept, trending of automation and automated data exchange. The main principles are:

- interoperability to connect every player;
- information transparency by using virtual copies of the physical world;

- technical assistance to humans provided by semi-autonomous machines and systems;
- decentralized decisions mostly made by automatic system.

2.1. Interoperability

Everything and everyone is connected. Today the Internet of Things (IoT) (Greengard, 2015) connects billions of devices, in 2020 there will be 50 billion devices connected (Statista, 2017).

The Internet of People is the same for the human beings: by having a digital footprint, some bigger than others, everybody is on-line.

2.2. Information transparency

Information transparency is provided by the ability of information systems to create a virtual copy of the physical world using aggregated raw sensor data. The devices connected to the IoT are mostly sensors producing many terabytes of raw sensor data every day. Using these virtual models, it is possible to produce higher-value context information.

2.2. Technical assistance

The new smart devices and systems help the humans by aggregating and visualizing information comprehensibly for making informed decisions and solving urgent problems on short notice. The short notice is a very important aspect, because the well-known industrial principle called “just-in-time” is also applied to the information. The second aspect of the technical assistance is the use of robots to conduct wide range of tasks that are unpleasant, too exhausting, or unsafe for their human co-workers.

2.3. Decentralised decisions

The new IT systems can make decisions on their own using the data collected by the sensors, and they are able to perform their tasks as autonomously as possible. Only in the case of exceptions, interferences, or conflicting goals are

tasks delegated to a higher level, which means fewer interruption, quicker production and less errors.

3. CHALLENGES FOR SCIENCE LABORATORIES

The new Industry 4.0 also affects the science laboratories and service providers, because – not surprisingly – the players from the industry are those, who pay the services.

The members of the economy are using Business-to-Business (B2B) networks for data interchange, which means the necessary information must be present and the required time, at the required place having the required quality. This means the science labs must be part of these B2B networks by providing access to the internal databases by interfaces which can be used by remote systems without direct human interactions.

This requires the usage of data interchange standards, for example ones like GeoSciML (2017), although there can be powerful players who can enforce using of non-standard data formats.

This integration works also in the other way around: the raw data for scientific evaluation and processing can and will be offered by remote systems when the source of the data is a far-away or dangerous location (like in case of space mining or remote oil fields in deserts or oceans).

3.1. Automation

The automation of laboratories is a very important aspect. First, more and more equipment can operate automatically, which means higher quality result (less errors) and continuous operation, and provides a good basis for integration. Using deep/machine learning tools the analysis of the data can be also automated resulting reduced processing type, increased output quality and even the possibility to process big heaps of data and turn them into valuable resources. These tools are a “must have” because not having them is an obvious disadvantage.

3.2. Integration

The integration of all the devices and systems in a laboratory is the basis of the continuous operation, so when one device or system finishes its task, the next one can start its own work on the results. From a technological point of view using workflow-based systems provide ideal solutions to build a highly-integrated laboratory where the everyday work is governed by a well-defined, “standard” procedure which helps to reduce the error-prone and boring manual data entry, enables parallel processing and continuous operations, so there will be less idle time for expensive devices and experts. In this way, the machines work while the creator sleeps.

3.3 Security

Security is a very delicate aspect. When everything is automated, integrated and the operation has been made as seamless and continuous as possible, the most valuable business asset is produced: information.

Unfortunately, the “networks everywhere” also means that “intruders everywhere”. The internet of things is full of devices with poor security implementations and settings, and most software products also have security flaws.

While the system must be made secure, it must be also made easy to use by the researchers and by the customers. Researchers want to work efficiently, without being interrupted by security systems and customers want to access the information they have paid for from remote locations.

4. GEOCHEM SMARTPORTAL SCIENTIFIC INFORMATION SYSTEM (SPSIS)

As a response to these challenges, a new system is under development at the Geochem Ltd, called smartPORTAL Scientific Information System (SPSIS) (smartPORTAL, 2017).

This new product is intended to be a cloud-based modular system for managing every aspect of the laboratory, including the daily workflow, the knowledge and all the research data.

The main emphasis on the workflow based operation to reduce the idle time, the unnecessary repeated tasks. This provides elevated level of automation and employee and customer self-service functions, while keeping all the data in secure environment.

4.1. Main modules

There are four main modules in the system:

- the scientific database stores all the research and context data in a uniform and searchable format,
- the knowledge base contains all the support data necessary for the employees, meaning publications, previously made reports etc.,
- the resource management module enables planning and accounting of the available and used resources, including all the equipment, support materials, human resources etc.,
- the financial module provides functions for managing the business issues, including support for project-based reporting.

4.2. Cloud-based

Because the system is cloud-based, it is accessible using a web interface, practically by a web browser only. This provides maximum flexibility, since no need for installing a client application, only a live internet connection is necessary. This way a secure platform for data storage and interchange provided.

The 3rd party systems can access the system using secure connectors, so they can access all the information granted to them but only those ones.

4.3. Workflow based

The workflows are designed to model the real task chain performed in the laboratory, so the users can work uninterrupted and at the same time it is possible to plan and manage resources.

The consecutive tasks are using the data produced by the previous tasks, so the error-prone manual data entry is minimized.

The flexible configuration of the workflows enable adaptation of new processes quite easily.

The repeating of the processes is made easy because every data produces by every task is stored in a database with its version and context information, so any workflow task can be repeated easily without repeating the previous tasks, and all data versions are available later for re-use.

4.4. Geochem SmartLab

This scientific information system is the fundamental service which the SmartLab (SmartLab, 2017) mobile laboratory system is built on, however this mobile laboratory concept is a topic of another presentation.

5. CONCLUSIONS

The Industry 4.0, big data, machine learning and all the innovative technologies change the ways of data collection, information processing, project management and execution. Adaptation to this changing world and using the newest IT technologies is necessary for everybody to remain in the business and being able to run successful projects.

5 REFERENCES

- Geochem SmartLab: <http://www.geochem-ltd.eu/hu/smlab> (accessed on 2017.04.10)
- GeoSciML: <http://www.geosci.ml.org/> (accessed on 2017.04.10)
- GREENGARD, S. (2015): The Internet of Things – The MIT Press, 2015 - ASIN: B00VB7I9VS
- SCHWAB, K. (2016): The Fourth Industrial Revolution - PENGUIN GROUP, 2016 - ISBN-13: 978-0241300756
- smartPORTAL: <http://www.intercomp.hu/hu/smartportal> (accessed on 2017.04.10)
- Statista: <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/> (accessed on 2017.04.10)

How to automate a petrophysical laboratory? - challenges and solutions

Ferenc Fedor¹, Viktor Feurer², Schön Roland¹

¹Geochem Ltd, fedor.ferenc@geochem-ltd.eu

²Intercomp Ltd, feurer.viktor@intercomp.hu

The economic strategic concept of industry known as Industry 4.0 has a strong effect on the service sector and in this way on the laboratory service sector as well, enforcing the adaptation of its principles. Geological laboratories are no exceptions from this trend. Geology has to prepare itself for the new challenges made up of the conquest of partly unknown, extreme fields and partly of the application of the “design as you go” principle. A possible solution to this is the Smartlab concept that enables in situ, widely or fully automated high-tech measurements according to protocols combined with an interpretation method via secure internet communication. At the same time, the application of the concept raises a number of problems that the authors intend to present in details on the example of the automation of a petrophysical laboratory.

Key words: *automation, SmartLab, petrophysics, mobilization*

1. INTRODUCTION

The Fourth Industrial Revolution taking place in our days, just like the former ones, is fundamentally altering the global economic and social environment. Its key words are horizontal and vertical integration, i.e. enhancing the effectiveness of the supply chain and manufacturing processes and their automation, decentralized and automated decision-making, mass customization, effective analysis of large data bases related to processes (big data), virtual reality. The emergence of the economic strategic concept of the so called Industry 4.0 (Schwab, 2016) is based on the demand of production automation (unreliability of human workforce in the process of production) and the informatics systems that enable it, and especially on the rapidly developing data storage- and analysing ability, as well as on the desire for unique consumer

products instead of mass products. The changes accompanying the concept have a fundamental effect on the service sector and on the process of studying/education (lexical knowledge and basic skills vs. project based learning). Geology as a profession has to face new challenges as well in the near future let us think, for instance, on space mining or on conquering the Mars. The Smartlab project would be an effective option to answer these future challenges on the market where traditional laboratory measurements characterized by excess capacity.

2. THE SMARTLAB CONCEPT

The basic idea of the Smartlab concept is that not the core sample should be carried to a well-equipped laboratory, but the fully equipped laboratory should be transported to the drilling site/field/mine performing the measurement on-the-spot (Fedor, 2016). This idea appears to be unusual and unreasonable at first, but it becomes clear and gains meaning on the basis of the explanation presented in this paper.

2.1. The justification of in-situ measurements

Considering the core sample's way from sampling (on-site or drilling) to laboratory measurements, there are several arguments in favour of in situ investigations (Rathi 2015, Fedor, 2016). These are as follows:

- The process from sampling to interpretation requires a lot of time (months). For instance in case of drilling the sample is carried to storage after quick documentation. In the storage facility the samples are documented in details and then, the sample or a part of it is transported to a laboratory for non-destructive measurements or in case of appropriate equipment they undergo non-destructive measurements in the storage facility (scanning, surface investigations, tests on half cores, computer tomography). This is followed by a review, where the places of sampling are identified in detail and the chosen samples are taken to laboratories for special measurements (mineralogy, geochemistry, petrophysics, rock mechanics, pore fluid tests etc.). After the measurements, the reports are delivered to customer within 3-6 months and the

integrated interpretation is done either by the customer or an assigned company.

- During transportation and storage the sample is exposed to different physical and chemical effects (destruction, fracturing, inversion, drying, wetting, geochemical changes etc.) that might lead to incorrect results or no interpretation.
- In case of certain raw materials strict rules apply to transportation and a lot of countries prohibit the shipment across their borders (e.g. uranium, hydrocarbon exploration, etc.).
- In the latter case the measurements should be carried out in the given country, which is not always possible in a high quality and the research data should also be stored in a secure way.
- The companies usually order the most necessary measurements only, and hence, sometimes has to re-sample the core or in the field later.

2.2. Problems of the mobile laboratory (Smartlab)

A laboratory that is mobile and at the same time suitable for sophisticated measurements provides appropriate solutions for these issues, since measurements are made on-the-spot, transportation of cores is eliminated (damage, permissions, laboratory measurement possibilities), the customer receives the results much faster (no shipping, fast analysis, etc.), and the measurements suit the requirements of mass customization (fewer limits in sampling and investigation).

At the same time, such a laboratory has to solve the following difficulties:

- Meet the quality control and market requirements (verification, validation, acceptance of measurements);
- Operation of the laboratory under extreme conditions;
- Installing high-tech equipment in a limited space;
- Automation, online interpretation, sporadically limited network traffic (satellite, internet), data security;
- Operation without qualified personnel.

2.3. The solutions given by Smartlab

The Smartlab concept (**Figure 1**) is able to provide complete sequences of measurements, almost anywhere under any circumstances with full or partial automation, working with local technicians according to protocols (quality control). The measured raw data are transferred via secured internet to a database where after automatic pre-processing they are interpreted by professionals who can be practically located anywhere (time saving and data security). The professionals' tasks are integrated interpretation, setting the measurement parameters and in given case the control of the measurement. The results are summarized in a report and uploaded to an online platform where its directly accessible after online payment. The laboratory is basically using renewable energy sources (solar collector, wind turbine) but can rely on diesel generator or grid connection if necessary. The consistence of power supply is provided by uninterruptable power supply (UPS) and accumulators. Water and nitrogen is retrieved from the environment using only a minimal amount of chemicals. In order to save space, different instruments can be combined or built together.

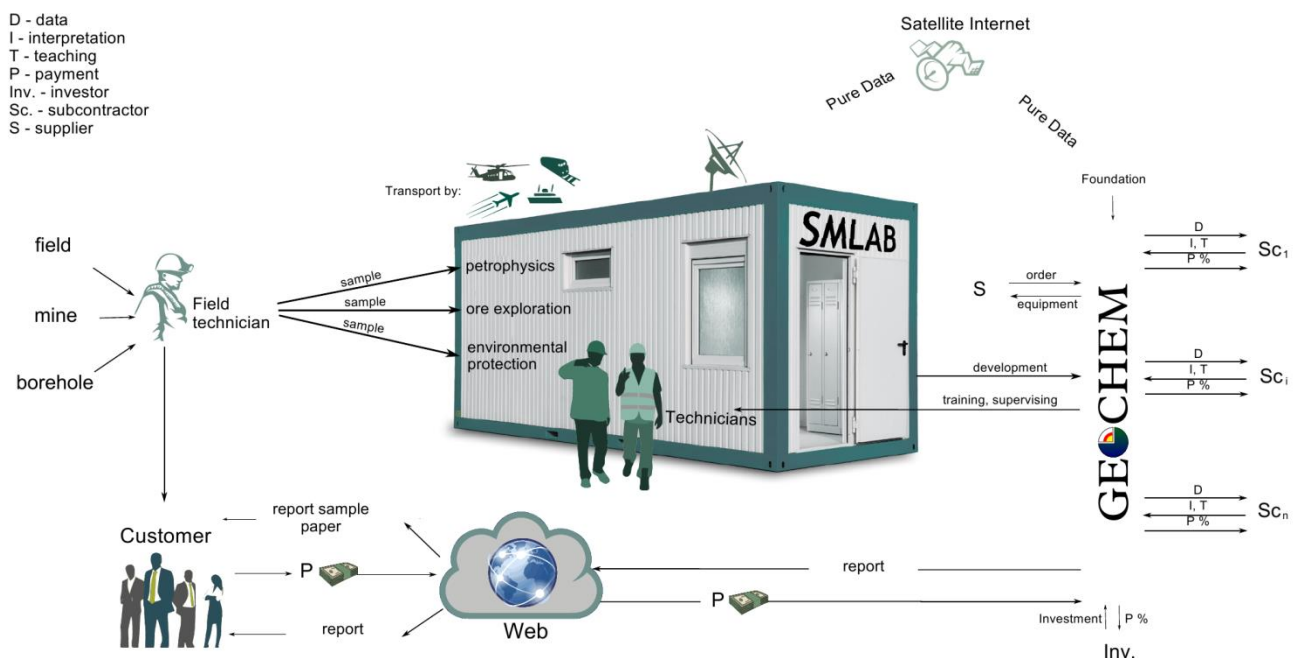


Figure 1: The Smartlab conception

3. APPLICATION OF THE SMARTLAB CONCEPT ON THE EXAMPLE OF A PETROPHYSICAL LABORATORY

The application of the Smartlab project will be practically demonstrated on the example of GEOCHEM Ltd.'s petrophysical laboratory. Although, the automation of the laboratory has only been started, it will be realized along the principles mentioned before (Nagy et al., 2013). The first task is to design a functioning, but not moveable prototype and by incorporating the experiences the development of a mobile version.

3.1. Steps of laboratory automation

In the first instance, some problems can be neglected in the case of the given laboratory, as electricity (UPS), gas and water supply and internet network are provided and few chemicals are used in the facility, but mobilization is not possible.

Each high-tech instrument is a separate unit with individual, own processing software displaying and reporting data only by their own software. At the same time, these programs can often be run only on the given attached computers. The task is to retrieve raw data and directly transmit them to a relational database. The procession of raw data should be done directly from the database, thus the necessary algorithms have to be programmed. This, in addition enables interpretation on a deeper level.

Manufactured instruments are usually not prepared for special tasks, making it necessary to review their operation and further develop them for the desired purpose. Some measurements (especially those made under reservoir conditions) can only be conducted on more than one instruments, although temperature and pressure range values and the used sample are the same in each measurement. It would be practical to include these measurements in a single instrument. In effect, this means applied research and experimental development work resulting in a new instrument prototype. The task is to further develop the device according to the purpose.

Redundancy is one of the biggest questions concerning the laboratory. Since the measurements are interconnected, the output data of individual measurements

serve partly as input data for the other ones. The control- and processing program of each device is different, so the data have to be copied and/or written into the given program and/or file, and this inevitably leads to mistakes when handling large amount of data. The task is to eliminate human errors resulting from data copying.

The measurements are partly performed on the high-tech instruments by technicians under supervision, but dominantly by professionals. Carrying out the measurements takes a lot of time instead of doing development and interpretation work. It is important to elaborate protocols that enhance the role of technicians during measurements, so the highly qualified personnel will be able to concentrate on interpretation and development. This also requires the interpretation process to be separated from the instrument, i.e. the professional should be able to retrieve information from a platform independent software, directly from the database when making the interpretation from the measurement results.

Data queries and their follow-up mean a serious challenge in managing the majority of databases. This problem means that on the level of metadata it is important to exactly know on which data the interpretation is based and the given query should be repeatable at any time. The task is to record the data of queries and attach them to the given interpretation.

Managing this, ensures that the technician can carry out the measurement according to protocol and the data are automatically processed and transmitted into a database, thus the output data of individual measurements are conveyed with references (not by manual copying) to the input data position of the other measurement. The interpretation can be done anywhere on a platform independent surface by a professional with access permission. Every data movement is recorded, thus data theft cannot be fully eliminated but is traceable. Regarding hardware the system is constituted by terminals where data saving can be blocked and a local server. The framework is provided by a project management system that supports the allocation of resources, economic calculations and the delegation of authorities.

3.2. Example of automation in He-pycnometry

He-pycnometry serves for the calculation of porosity and matrix density of rocks by measuring/calculating their matrix volume at given temperature and low pressure according to the Boyle-Marriott law with the use of He gas, if sample's mass and geometrical volume are known. Its input data are the geometry and volume (measured before measurement), the output data are rock density, specific density and porosity. The measurement produces a txt file that contains the information of the measurement. The information is retrieved directly from the txt file from which the porosity value is calculated. Porosity is the input information for example permeability measurement and density for Hg-porosimetry. Matrix volume are calculable from p and V pairs but this was not the aim in this case, because the algorithm integrated into the instrument makes the calculations in an automated and controlled way. A testing algorithm will be built in (in progress) for checking the acceptability of geometry, mass and porosity.

3.3 Example of automation in acoustic velocity measurements

The instrument for acoustic velocity measurement used at GEOCHEM Ltd. is able to detect the arrival time of wave packets induced by a given central frequency at given pressure, temperature and pore pressure. The device was reconstructed in order to improve the quality of measurement. The arrival of “P” and “S” waves is detected from two, independent algorithms. The measurement itself can be conducted by a technician based on the protocol. Direct data transmission into a database is in the programming phase. Data procession is done on a platform independent software.

3.3 The integration of instruments

The measurements for reservoir characterization (electric and acoustic measurements, permeability, porosity) on a single core sample are theoretically done under the same reservoir conditions. Here the question arises as to whether it is possible to measure all parameters in a single work phase. The

answer is yes. The practical realization would be time and resource consuming that are lacking at the present.

4. CONCLUSIONS

The realization of the Smartlab concept is not a trivial process, it incorporates the integrated knowledge of several scientific disciplines from device design to the understanding of given physical, chemical processes, including the programming of related calculations and algorithms as well as tasks (planning and realization) related to energetics, building engineering, logistics, informatics, automation and process control. At the same time, the implementation of the Smartlab project makes way for conquering hitherto unknown fields under extreme conditions and for a significantly faster exploration process. The improvement of the petrophysical laboratory in this direction is only a part of the complete task, while several problems are to be solved, but with appropriate planning and professional experience it is realizable, because technology and informatics are on an adequate level of development. In order to achieve this, it is time to break with the traditional approach and to apply devices and methods that are new and unknown in geology.

REFERENCES

- FEDOR, F. (2016): SmartLab – a new concept of global geological laboratory service, 35th IGC, Instrumental, experimental and laboratory – based developments in the geosciences, 31.08.2016., Cape Town, SA, Paper Number: 5239, <http://www.americangeosciences.org/igc/16621>
- NAGY, A., FEDOR, F., FEURER, V. (2013): Purpose and limits of automation in laboratory practice, 5th HR-HU and 16th HU geomathematical congress “Geomathematics as science”, 30 May-01 June, 2013, Mórahalom
- RATHI, C. (2015): Outsource Drill Core Storage To Increase Oil & Gas Operational and Financial Efficiency, Frost & Sullivan White Paper, www.frost.com
- SCHWAB, K. (2016): The Fourth Industrial Revolution - PENGUIN GROUP, 2016 - ISBN-13: 978-0241300756

Application of semi-automated GIS procedure for river terrace delineation on high resolution LiDAR data in Sava river valley NW of Zagreb, Croatia – First results

Neven Trenc¹, Bojan Matoš², Josipa Velić², Dario Perković²

¹Croatian Agency for Environment and Nature, Radnička cesta 80/7, Zagreb, Croatia,

(Neven.Trenc@dzzp.hr),

²Faculty of Mining, Geology and Petroleum Engineering (University of Zagreb), Pierottijeva 6, Zagreb, Croatia.

Three aggradational terraces of Sava River generally composed of poorly consolidated gravels and sands were formed during Holocene time. In spite of their economic significance, spatial relationships and extent of the two older Holocene Sava River terraces and the recent terrace near Zagreb are not undoubtedly established. This is due to relatively weak morphological expression, vegetation cover and anthropogenic changes in the terrain associated with flood protection structures, traffic infrastructure and agriculture. In the past two decades stratigraphic history of Quaternary sediments in the area has been reconstructed to a great detail through extensive analysis of borehole data. However, present knowledge about spatial relationship and extent of Sava River terraces is mostly based on the geological map published in 1979 and represents summary of the field information and geological exploratory work completed at that time. LiDAR data with high vertical and horizontal resolution have been proven in various studies as a good basis for the application of automatic algorithm based procedures for extraction of river terraces. In this work *TerEx* ArcGIS extension developed by Justin C. Stout and Patrick Belmont was applied. Several morphologic parameters were varied in the iterative process until most probable result coinciding with the known extent of terraces was obtained. After manual editing of the presumed terrace geometry, the heights of the terraces above the current river course were also extracted by the same tool. The results of the applied semi-automated river terrace extraction process indicated in some sections different spatial relations than previously established. To assess their validity, the results will be checked in the field and compared to the data obtained during preparatory works for gravel excavation and to a new 3D visualisation of water supply borehole data.

Key words: *Sava River terraces, Holocene, NW Croatia, TerEx (GIS)*

1. INTRODUCTION

The Quaternary evolution of Sava River and its sediments in the NW Croatia has been studied in detail by many authors in the past several decades (Šikić et al., 1972a; Velić and Saftić, 1991; Velić et al. 2002; Saftić et al. 2003 with references). In these studies detailed insight in the evolution of Quaternary deposits was obtained but the spatial extent and fluvial dynamics of Sava River system (i.e., Sava alluvial terraces and floodplains) were not discussed in detail. Geological map of the study area (sheet Zagreb, Šikić et al., 1972b) produced in late 1970-ies, shows three characteristic terrace risers that delineate two older Holocene and one recent Sava River terrace. Identified terrace risers are characterized by relatively poor landscape expression (<2m), and in many cases are either covered by dense vegetation or influenced by anthropogenic changes (agriculture, flood protection dikes and traffic infrastructure). Considering the fact that some segments of terrace risers and Sava River terraces may be misallocated by previous studies, in this work we employed high resolution LiDAR data (Light detection and ranging) to address possible true geometries of risers and terraces and their heights. Data collected in this study on Sava River fluvial system may be significant geomorphological and paleoenvironmental indicators and provide the framework of present-day geomorphic processes as well as paleoclimate and tectonic events that affected study area in Quaternary. In this study we present first findings on GIS-based LiDAR processing methodology that was used to delineate Sava River terraces and their comparison with existing data.

2. TECTONIC AND GEOLOGIC SETTINGS

The tectonic evolution of the study area, i.e., Sava Depression is linked with evolution of Pannonian Basin System (PBS) through Neogene and Quaternary, which was characterized by several distinct phases. Tectonic evolution onset with lithospheric extension trough Early to Late Miocene (c. 26-11.5 Ma) where PBS was extended by “back-arc-type” extension along dominantly NNW-striking listric normal and strike-slip faults. (Royden et al., 1983; Horváth et al. 2006;). Sava Depression was characterized by deposition of syn-rift alluvial and

freshwater lake clastics (Ottungian-Karpatian in age) that due to marine inundation in Badenian transgressed to basal breccia-conglomerates and coarse grained sandstones or lithothamnion and bioclastic limestones (**Figure 1**) (Lučić et al., 2001; Velić et al., 2002; Saftić et al., 2003). Beside tectonic inversion that commenced at the end of Middle Miocene (c. 13.0-11.6 Ma), in the Sava Depression as well as PBS in the Late Miocene to early Pliocene (c. 11.5-5.3 Ma) deepening and rapid thermal subsidence were dominant processes. This processes were located along inherited structures yielding deposition of delta-related clastic sequences in the shallow parts, whereas in the deepest parts, deep-water turbidites sediments (Pavelić, 2001; Tomljenović and Csontos, 2001; Saftić et al., 2003).

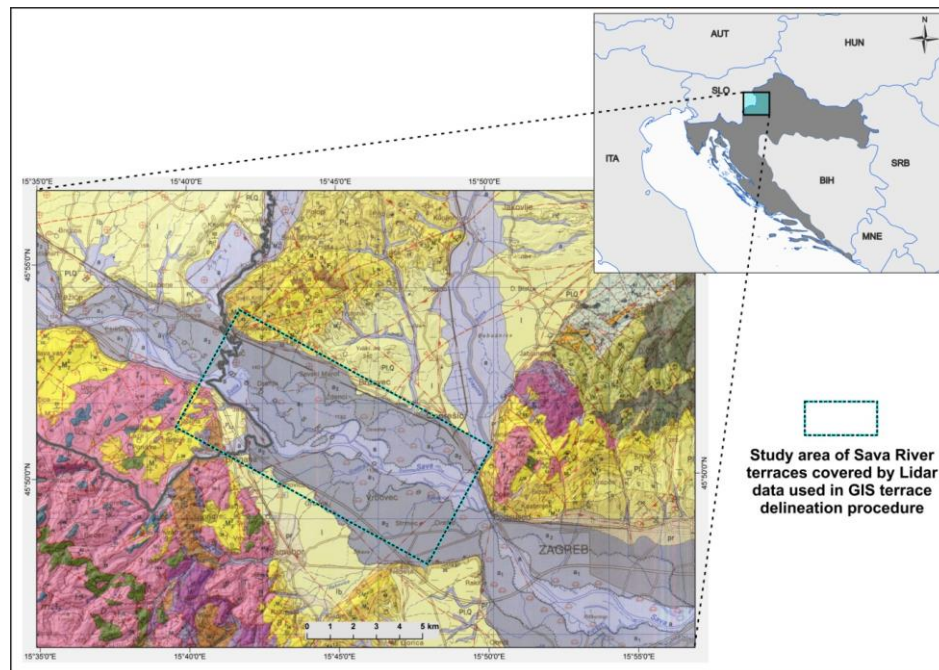


Figure 1: The geological map of study area (after Šikić et al., 1972b). Study area with an area of available LiDAR data is indicated with rectangle

With the NE-ward translation and CCW rotation of the Adria Microplate during Pliocene and Quaternary, PBS as well as Sava Depression was affected by regional compression, i.e., transpression motions, tectonic inversion and structural reactivation along the existing and newly formed faults (Prelogović et al., 1998; Fodor et al., 1999; Rukavina et al., 2016 with references). Beside uplifted inselbergs and neotectonic active faults (Prelogović et al., 1998;

Tomljenović and Csontos, 2001, Rukavina et al., 2016), Pliocene and Quaternary (**Figure 1**) was characterized by sedimentation within remnants of local-scale sub-depressions, filled with lacustrine deposits, gravel-rich beds and finally Holocene loess cover deposits (Šikić et al., 1972a; Velić and Saftić, 1991; Saftić et al., 2003 with references). In the area of Sava Depression, Quaternary deposition involved fluvial deposition of Sava River, which resulted in three Holocene aggradational alluvial terraces (**Figure 1**) of c. 30 to 115 m total sediment thickness (Šikić et al., 1972b).

3. MATERIALS AND METHODS

The limits of the study area were defined by the obtained LiDAR dataset that covered section of NW Croatia between the city of Zagreb and Slovenian border. The path of the image acquisition roughly followed NW-SE oriented course of the Sava River. The DEM created on the basis of the LiDAR data has a shape of irregular polygon extended in the flight direction with an area of about 69 km² (12km length, 5km wide). The created DEM had horizontal resolution of a centimetre scale, with a RMS error of 0.075 m, whereas its vertical resolution was decimetre scale, with an overall RMS error of 0.055 m. The morphometric analysis for this research was performed on the prepared “bare earth” digital elevation model in the GeoTIFF format with 1X1 m grid resolution. This digital elevation model covered almost the full width of the Sava River alluvial deposits except the oldest raiser that indicates contact between Quaternary and older Neogene sediments.

The river terrace's can be described as a staircase form that consists of flat areas, terraces or treads which are separated by steep risers (Hugget, 2007). The main research objective of this study was to extract the spatial extent of each Sava terrace and position of the associated terrace raiser. This was achieved by LiDAR-based semi-automatic GIS procedure. The GIS extension TerEx by Stout & Belmont (2013) was applied since i) it uses typical morphological characteristics to delineate potential terraces on the basis of DEM, and ii) it has been tested in the river alluvial valleys of various sizes and characteristics (Stout & Belemont, 2013).

The initial datasets necessary for terrace extraction are high resolution DEM, and the vector line representing the river path. The set of parameters necessary for procedure can be divided the two groups. Parameters in the first group are defined by the characteristics of the river and its valley (max. valley width, stream width at the mouth and reach length). Whereas second group of parameters (focal zone size, maximum terrace area, elevation change and smoothing tolerance) is changed through iterative process until the most plausible result is obtained by the operator based on his previous knowledge of the area. Focal zone size and max. elevation change parameters describe local relief over the selected unit area and directly relate to the morphological characteristics of the flat terrace surfaces, while the minimal terrace size defines minimal identified flat area that is considered a terrace. Using described parameters and DEM, application creates polygonal layer where each polygon represents a potential terrace surface. This layer is then manually edited i.e., polygons or terraces that are connected with small artifacts called “tails” are separated (Stout & Belmont, 2013). In the second TerEx step river path is divided in the multiple reaches, and terrace polygons elevations with respect to the closest river reach are calculated. In this research, terrace polygons were further aggregated based on their average heights above corresponding reach in three groups. All of them were considered as a potential single terrace. Afterwards, these extracted, DEM-based river terraces and terrace raisers were compared to the known extent of river terraces and risers shown on the geological map of Zagreb (Šikić et al., 1972b).

4. RESULTS WITH DISCUSSION

Due to similar elevation above river obtained from DEM, the results (**Figure 2**) reveals that in the NW part of the study area, a significantly wider area could be assigned to the older Sava terrace (marked by a2 in **Figure 2**). Similarly, in the easterly part of a Sava River valley, results obtained from DEM indicate wide areas of lowermost elevation as an areas of the younger Sava terrace (marked by a1 in **Figure 2**). The same area according to the geological map was either classified as the recent Sava River terrace or as a younger Sava terrace (marked

by a1 in **Figure 2**), however, with a significantly smaller assigned area. This extent difference may be explained by misinterpretations of numerous oxbow lakes and meanders that were considered as a features related to recent (abandoned) Sava River course. Instead, due to their relative higher elevation with respect to recent Sava River course, these features probably were constructed by Sava course migration that was concurrent with formation of younger Holocene Sava River terrace

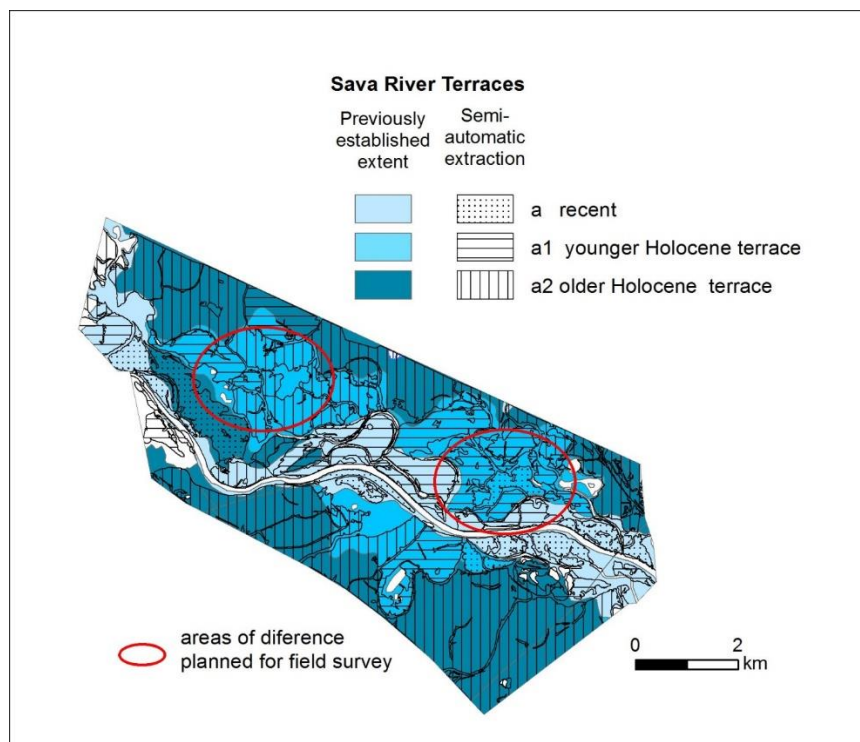


Figure 2: Comparison of the extent of Sava river terraces obtained by GIS semi-automatic procedure to extent shown on the geological map of study area (after Šikić et al., 1972b).

5. CONCLUSIONS

The results obtained in this study support geologically determined spatial extent of the terraces in most but not all areas. The similarity of the GIS derived Sava River terraces and geological map shows that this method represents a viable approach to extract spatial extent of river terraces and raisers. The semi-automatic procedure may be influenced by operator decisions or various artifacts caused i.e. by vegetation or infrastructure and must be validated with field

surveying. Accordingly, our future work will try to clarify whether our results represent true extent of terrace surfaces and positions of the risers.

ACKNOWLEDGEMENTS

This study was supported by the Faculty of Mining, Geology and Petroleum Engineering (University of Zagreb) and Croatian Agency for Environment and Nature (Republic of Croatia). The authors would like to thank to PROGRAM SAVA d.o.o. for providing LiDAR data for this research and to Justin C. Stout and Patric Belemont for instruction in usage of TerEx application.

REFERENCES

- FODOR, L., CSONTOS, L., BADA, G., GYÖRFI, I. & BENKOVICS, L. (1999): Tertiary tectonic evolution of the Pannonian basin system and neighbouring orogens: a new synthesis of paleostress data. *The Mediterranean Basins: Tertiary extension within the Alpine Orogen*. Geological Society of London, Special Publications 156, 295-334.
- HORVÁTH, F., BADA, G., SZAFIÁN, P., TARI, G., ÁDÁM, A. & CLOETINGH, S. (2006): Formation and deformation of the Pannonian basin: constraints from observational data. In: Gee, D.G., Stephenson, R.A. (Eds.), *European Lithosphere Dynamics, Memoirs, Geological Society, London, Vol. 32.*, 191–206.
- HUGGET, R. J. (2007): *Fundamentals of geomorphology*, Second Edition, Routledge, 447p.
- LUČIĆ, D., SAFTIĆ, B., KRIZMANIĆ, K., PRELOGOVIĆ, E., BRITVIĆ, V., MESIĆ, I. & PAVELIĆ, D. (2001): Tectonostratigraphic model for the North Croatian and North Bosnian sector of the Miocene Panonian Basin System. *Basin Research*, 12, 359-376.
- PRELOGOVIĆ, E., SAFTIĆ, B., KUK, V., VELIĆ, J., DRAGAŠ, M. & LUČIĆ, D. (1998): Tectonic activity in the Croatian part of the Pannonian basin. *Tectonophysics*, 297, 283-293.
- ROYDEN, L.H., HORVÁTH, F. & RUMPLER, J. (1983): Evolution of the Pannonian basin system, 1. tectonics. *Tectonics* 2, 63-90.

RUKAVINA, D., MATOŠ, B. TOMLJENović, B. & SAFTIĆ, B. (2016): Neotectonic active faults in the Eastern Part of Sava Depression : Implications to tectonic evolution based on 2D seismic data and 3D subsurface structural modelling. In: Cvetković, M., Novak Zelenika, K., Horvath, J., Hatvani, I. G. (eds.): 8th Croatian-Hungarian and 19th Hungarian geomathematical congress "Geomathematics - present and future of geological modelling", Trakošćan, Proceedings, 129-137.

SAFTIĆ, B., VELIĆ, J., SZTANÓ, O., JUHÁSZ, G. & IVKOVIĆ, Ž. (2003): Tertiary Subsurface Facies, Source Rocks and Hydrocarbon Reservoirs in the SW Part of the Pannonian Basin (Northern Croatia and South-Western Hungary). *Geologia Croatica*, 56/1, 101-122.

STOUT J. C & BELMONT P. (2014): TerEx Toolbox for semi-automated selection of fluvial terrace and floodplain features from lidar., *Earth Surface Processes and Landforms*, 39, 569–580.

ŠIKIĆ, K., BASCH, O. & ŠIMUNIĆ, A. (1972a): Basic geological map of Yugoslavia, M 1:100 000, sheet Zagreb. Geological Survey Zagreb, published by Federal Geological Institute Beograd.

ŠIKIĆ, K., BASCH, O. & ŠIMUNIĆ, A. (1972b): Explanatory notes for geological map of Zagreb (L 33-80; 1: 100 000) (in Croatian). Geological Survey Zagreb, published by Federal Geological Institute Beograd.

TOMLJENović, B. & CSONTOS, L. (2001): Neogene-Quaternary structures in the border zone between Alps, Dinarides and Pannonian Basin (Hrvatsko zagorje and Karlovac Basin, Croatia). *International Journal of Earth Sciences (Geologische Rundschau)*, 90, 560-578.

VELIĆ, J. & SAFTIĆ, B. (1991): Subsurface spreading and facies characteristics of Middle Pleistocene deposits between Zaprešić and Samobor. *Geološki vjesnik*, 44, 69-82, Zagreb.

VELIĆ, J., WEISSER, M., SAFTIĆ, B., VRBANAC, B. & IVKOVIĆ, Ž. (2002): Petroleum-geological characteristics and exploration level of the three Neogene depositional megacycles in the Croatian part of the Pannonian basin. *Nafta*, 53/6–7, 239–249.

Application of artificial neural networks for lithofacies determination in absence of sufficient well data

Ana Brcković¹, Monika Kovačević¹, Marko Cvetković¹

¹University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6,
10 000 Zagreb, Croatia, ana.brckovic@gmail.com

Lithofacies definition in the subsurface is an important factor in modelling, regardless of the scale being on the reservoir or basin level. In areas with low exploration level, modelling of lithofacies distribution presents a complicated task as very few inputs are available. For this purpose, a case study in Požega Valley was selected with only one well and several seismic profiles on an area covering roughly 850 km². A well to seismic correlation of lithofacies was performed with three different neural networks - multi-layer perceptron, radial-basis function and probabilistic neural network. Variables for the analysis were lithofacies determined from well logs, cores and cuttings, and twelve seismic attribute data. Lithofacies as the output variable was adequately presented in numerical values as a categorical variable and predictions on it were tested through a process of artificial neural network training in analytics software StatSoft Statistica. All of the neural networks were successful in the training process. However, the probabilistic neural network showed better results than other two. By extrapolating the lithofacies data onto the seismic profile, a greater input dataset for facies modelling was obtained. In this way, an application of neural networks in early phase of exploration has been confirmed for better definition of lithofacies distribution in the subsurface.

Key words: *artificial neural networks, Croatia, facies prediction, Pannonian Basin, Požega Valley*

1. INTRODUCTION

The area of Požega Valley in Požega-Slavonia County is weakly explored region. For such a wide area, of approximately 850 km², there are only few seismic profiles and one well dataset found. In order for a complete geological model of the Neogene Quaternary infill to be made, lithology data has to be established. For this purpose, artificial neural networks were used as they were previously proved successful in such a task (Bhatt, 2002; Cvetković et al., 2009).

2. MATERIALS AND METHODS

2.1. Area of exploration and geological settings

Požega Valley is located in the central part of northern Croatia (**Figure 1**). It is positioned between mountain ranges that are surrounding it from three sides: Papuk and Krndija on the north, Psunj on the west, as well as Požega and Dilj Mountain on the south. Further on the east there is Đakovo-Vinkovci plain. (Najdenovski, 1988).

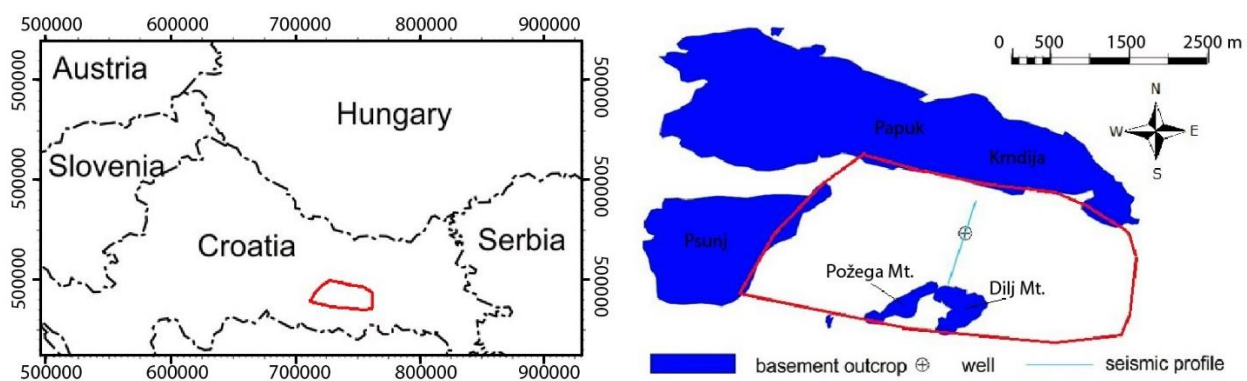


Figure 1: Outline of the Požega Valley (left) and enlarged area showing basement outcrop, with well Tek-1 and profile indicated (right)

Geotectonically, Požega Valley represents graben type subsidence and filling, while a horst type uplift is associated with surrounding highlands (mountains) in between Sava and Drava Depressions, that subside along vertical fault system and are being filled with sediments. The mentioned specific position allows for (represents) a great drainage basin, where sediments have been deposited by different mechanisms during Quarternary (alluvial sand, gravel, clay, loess) in alluvial and terrestrial environment (Ivanišević et al., 2015).

Geomorphologically, The Basin represents a spacious asymmetrical longitudinal plateau mainly bounded by a 300 m isohypse and covered by Quarternary deposits (Najdenovski, 1988).

Geological composition of The Valley is comprised of tectonically and stratigraphically different sets of rocks. There are roughly two rock complexes with different geological, lithological, petrographical and genetical

characteristics. First complex contains paleozoic and mesozoic rocks that are all together know as pre Neogene basement (**Figure 1**, basement outcrop). The other one incorporates Neogene and Quarternary sediments as sedimentary infill (Najdenovski, 1988).

The valley represents a closed area of 870 km², which in initial exploration did not prove potential it is underexplored regarding seismic and well network.

2.2. Methods and input data

Determination of lithofacies is crucial for modelling a certain area. In cases of low number of exploration data, problem of determination can be possibly resolved with a help of artificial neural networks (ANN). Input data that can be used for this purpose are ranging from well data and cores, to seismic attributes. What differentiates ANN from other available methods is the ability to lessen the uncertainty and errors in lithofacies determination through the incorporation of hard data (well logs and seismic attributes) and soft data (predictions) through the training process. The addition of lithology probabilities directly based on well observations will help to reduce uncertainty in reservoir prospecting and qualification (Hami-Eddine et al., 2014).

2.3. Artificial neural networks

Artificial neural networks are powerful geostatistical analysis method. It is based on the process of iteration which is called training process of a network. Every network is represented by the model of artificial neuron with a specific structure, with different inputs and, in our case, one output. Adequate weight is assigned to the values of the inputs (**Figure 2**). Certain amount of neurons are assembled into a network in the shape of layers. First layer is made of selected data and is called input layer. Signal travels from input layer to interlayers, where the processing of the data takes place. Because there is no interaction with its surroundings, interlayers are commonly called hidden layers (**Figure 2**). Different types of networks are distinguished based on the connection between the two mentioned layers. Every neuron is connected to the previous layer's

output and based on the given result or output, neuron can be activated. The activation is controlled by the activation function. Because the model works on feedback, that is accounted for assimilating weight coefficient and therefore reducing error, the network is known as backpropagation network (Cvetković & Velić, 2012).

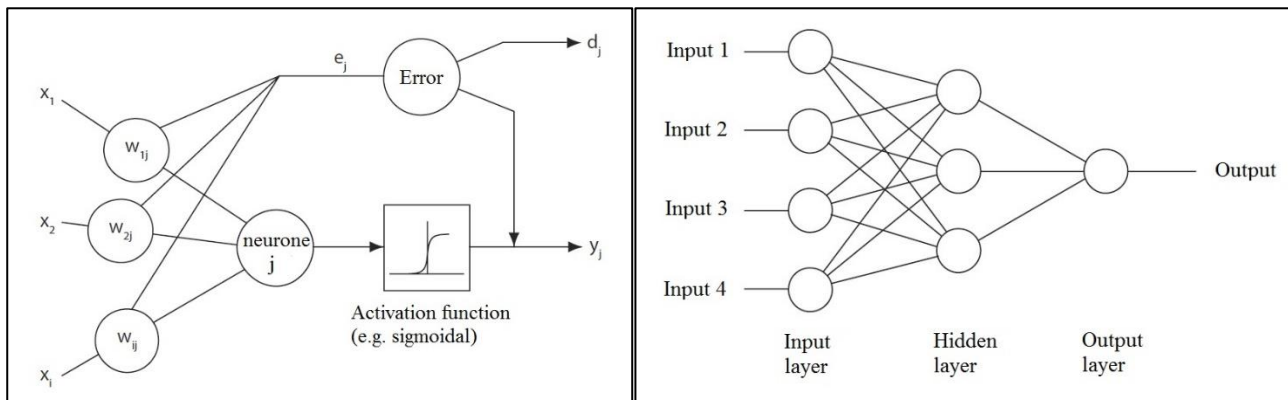


Figure 2: Principle of artificial neural network (left) and general structure (right)

In this paper we were using three different neural network architectures, in order to get the most likeable lithofacies predictions. Multi-layer perceptron network (MLP) are popular network architectures used in most the research applications in engineering, mathematical modelling, etc. The MLP network is based on a backpropagation algorithm that has one or more hidden layers. The MLP is more successfully applied in classification and prediction problems (Rumelhart et al., 1986).

The Radial-basis function network (RBF) is also a commonly used neural network but is more successfully and frequently applied in solving classification problems than in solving prediction problems (Cvetković et al., 2014). RBF has a hidden layer of radial units (neurons), each modelling a Gaussian response surface. They are also good at modelling nonlinear data and can be trained in one stage rather than using an iterative process as in MLP and also learn the given application quickly (Venkatesan & Anitha, 2006; Csábrági et al., 2017). They have similar structure as MLP with a distinction of having only one hidden layer.

Probabilistic Neural Networks (PNN) are useful for automatic pattern recognition, nonlinear mapping, and estimation of probability of class membership and likelihood ratios. For databases ranging from small to moderate, the training speed of PNN can be orders of magnitude faster than the well-known back-propagation (BP) paradigm, and yet the ability of the network to generalize to classify unknown patterns is approximately the same (Specht, 1988; 1990).

3. RESULTS

Neural network analysis has been conducted on the well Tek-1 data in Požega Valley, as well as the seismic attributes from seismic profile (**Figure 1**). For the sake of better output results, multiple networks have been used for processing data. Networks have been trained in StatSoft's analysis software STATISTICA. By using "Intelligent Problem Solver" and "Custom Network Designer" options, multi-layer perceptron, radial-basis function and probabilistic neural networks were tested alongside (**Table 1**). The analysis was made on thousand networks for each type with learning rate 0.01, which made for a relatively smooth training graph. Twenty of the best ones in the training process were retained. After comparing the predictions and errors, one of each type was selected for running on a new set of data. Training dataset was constructed of a set of 12 seismic attributes derived from the single seismic profile. Best results were obtained with all attributes used for the lithology training and prediction.

MLP provided us with acceptable lithostratigraphic predictions, but comparing to other network architectures, the errors were the highest (**Table 1**).

RBF network did not perform well when it comes to the ability to predict lithofacies. Even though the errors were small, the actual predictions were geologically unsound when prediction was performed on the whole profile.

Finally, even though the PNN network is expected to perform best in classification, it gave us adequate predictions and highest prediction/error ratio (**Table 1**). Rising the number of neurons in the hidden layer showed little to no deviation in error values, but at the same time the gained results were improved.

Table 1: Comparison of three different neural networks

Network architecture	Train error	Selection error
MLP 12-40-1	0.66	0.79
PNN 12-250-1	0.19	0.26
RBF 9-28-1	0.32	0.35

*Structure of neural networks is represented with input layers (first number), output layers (last number) and hidden layer (number in between)

The above mentioned PNN model (**Table 1**) was applied on traces from seismic profile and acquired results (predictions) of lithostratigraphy were imported in Schlumberger’s software platform Petrel. Modelled lithology was then overlayed on the seismic profile for better visual representation (**Figure 3**). Acquired lithology distribution shows dominantly marls with sandstone lenses within them, which can be correlated to the general geologic stings for the Upper Miocene.

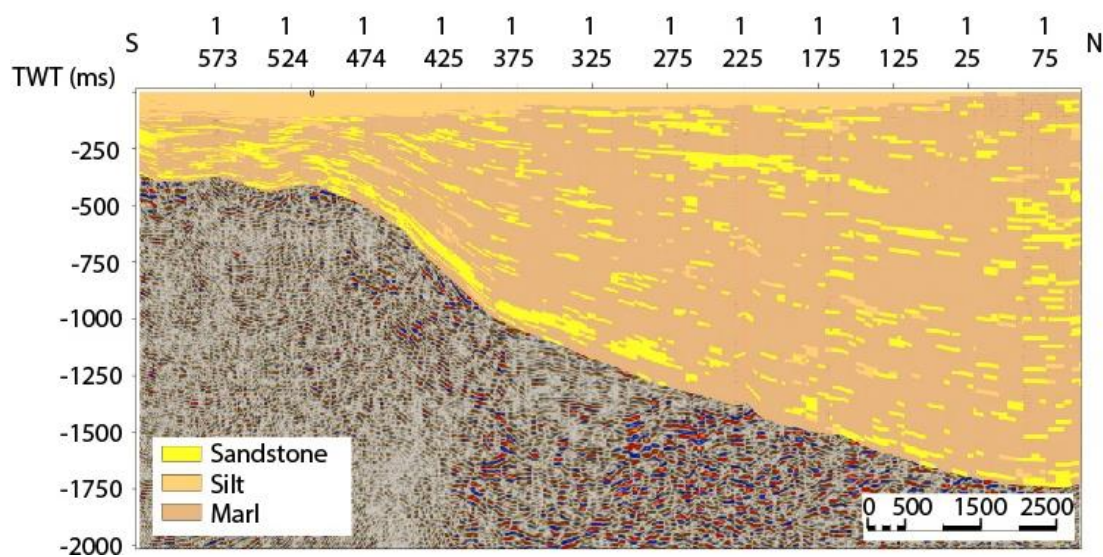


Figure 3: Lithofacies analysis overlaying seismic profile (vertical blue line represents position of Tek-1 well).

4. CONCLUSION

Usage of artificial neural networks has shown adequate results when it comes to predicting lithologic relations in the area of low exploration level, if no other data

is available (e.g. more wells). Networks are initially trained on obtained data, to calculate the most similar stratigraphy sequence to the given one. Trained network, with acceptable deviation from the original data, is then applied for predicting lithology on other seismic profiles which can then be used in further geologic research; e. g. Basin modelling.

ACKNOWLEDGEMENT

Authors would like to thank the University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering for financial support and the Schlumberger Company for donating the Petrel licenses without which these analysis in their extent could not be possible. Authors would also like to thank the Ministry of Environment and Energy, Croatian Hydrocarbon Agency and Vermillion Zagreb Exploration d.o.o. for the well data.

REFERENCES

- BHATT, A. (2002): Reservoir properties from-well logs using neural networks.– Unpubl. PhD Thesis, NTNU, Trondheim, 151 p.
- CSÁBRÁGI, A., MOLNÁR, S., TANOS, P. & KOVÁCS, J. (2017): Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecological Engineering* 100, 63-72.
- CVETKOVIĆ, M., VELIĆ, J. & MALVIĆ, T. (2009): Application of neural networks in petroleum reservoir lithology and saturation prediction. *Geologia Croatica*, 62, 2, 115-121.
- CVETKOVIĆ, M., VELIĆ, J. & VUKIČEVIĆ, F. (2014): Selection of the most successful neural network algorithm for the purpose of subsurface velocity modelling, example from Sava Depression, Croatia. 6th HR-HU and 17th HU geomathematical congress, 21-28.
- CVETKOVIĆ, M. & VELIĆ, J. (2012): Successfulness of inter well lithology prediction on Upper Miocene sediments with artificial neural networks. 4th HR-HU and 15th HU geomathematical congress, 13-20.

HAMI-EDDINE, K., KLEIN, P., RICHARD, L., DE RIBET, B. & GROUT, M. (2015): A new technique for lithology and fluid content prediction from prestack data: An application to a carbonate reservoir. *Interpretation*, 3,, SC19–SC32

IVANIŠEVIĆ, D., GALOVIĆ, L., HORVAT, M. & WACHA, L. (2015): Visualisation and interpretation of the mineralogical data from the boreholes in Quaternary deposits of the southern part of the Pannonian basin system, by using multivariate statistics combined with volume interpolation.— In: Horvat, M. & WACHA, L. (ed.): Abstracts Book, 5th Croatian Geological Congress with international participation, 117-118.

NAJDENOVSKI, J. (1988): Dubinski geološki odnosi i razvitak struktura u terciarnim sedimentima Požeške kotline [*Subsurface Geological Relations and Structures Development in Tertiary Sediments of Požega Valley – in Croatian*]. Unpubl. PhD Thesis, Faculty of Mining, Geology and Petroleum Engineering ,University of Zagreb, Zagreb, 146 p.

RUMELHART D.E., HINTON, G.E. & WILLIAMS, R.J. (1986): Learning internal representations by error propagation. -In: RUMELHART, D.E., McCLELLAND, J.L. (ed.) *Parallel Distributed Processing*, Vol.1. MIT Press, Cambridge, 381-362.

VENKATESAN, P. & ANITHA, S. (2006): Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Current Science*, Vol. 91, No. 9, 1195-1199.

SPECHT, D.F. (1988): Probabilistic Neural Networks for Classification Mapping or Associative Memory. *Proceedings IEEE International Conference on Neural Networks*, Vol. 1, 525-532.

SPECHT, D.F. (1990): Probabilistic Neural Networks, *Neural Networks*, 3, 109-118.

Quarry Surveying and Analysis Using Drone-mounted LiDAR & AutoCAD Civil 3D

Nikolina Mijic¹,

¹PhD Candidate, Institute for Geophysics and Geoinformatics, University of Miskolc,
nikolinamijic7@gmail.com

While quarry and plant managers recognize the necessity of carrying out physical inventories of material stockpiles for accounting purposes, many regard it as a chore that takes up time and which has the potential to slow production. Times are quickly changing - AutoCAD Civil 3D provides rich set of geodetic tools and add-ons to dramatically speed-up surveyed data post-processing, visualization and analysis. Manual surveys using total stations and traditional global positioning system (GPS) equipment are time consuming, requiring 10 to 20 shots on a typical 15,000 cubic-meter stockpile. Drone-based laser scanning speeds up the data collection stage of the workflow and, when compared to aerial photogrammetry, offers much faster turnaround of the physical quantities. AutoCAD Civil 3D allows to compute volumes and generate profile views within a matter of hours, so that stockpile quantities recorded are accurate on a set day, rather than reflecting a historic situation. Drone-based 3D laser scanning not only reduces the time spent on stockpile surveying while enhancing the safety of workers, but also offers a level of surface detail that is incomparable to that collected from total stations: 100,000+ 3D points can be collected in just a few minutes. AutoCAD Civil 3D enables creating TIN surface from points within RCS format point cloud scanned object created with Autodesk ReCap. Drone-mounted 3D laser scanner includes a GPS receiver and inertial measurement unit (IMU), so data can be geo-referenced to an exact location. Each operation can be referenced to the same co-ordinate system. Key benefit of this operation is better accuracy and traceability of these methods. Drone-mounted LiDAR represents a safe way to survey dangerous and hostile environments. Once a point cloud-based surface is created within AutoCAD Civil 3D, it can be used built-in tools to perform quick volumetric calculations, and easily create alignment profile cross-sections using only polylines drawn atop of a generated TIN.

Key words: *surface, point cloud, AutoCAD Civil 3D, LIDAR.*

1. INTRODUCTION

LIDAR is today one of the most modern technology that is used in the survey and development of topographic maps for different purposes. The technology is based on collection of three different sets of data. Position sensors are determined using Global Positioning System (GPS), using phase measurements in the relative kinematics, use of Inertial Measurement Unit (IMU). The last component is a laser scanner. The laser sends infrared light to the ground and is reflected to the sensor. The time between the broadcast signal reception to the knowledge of the position and orientation sensor, allows the three dimensional coordinates to calculate the Earth. LiDAR has a very simple principle of measurement. The scanner emits pulses with a high frequency and is reflected from the surface back to the instrument. Mirror inside the laser transmitter is moved by rotating perpendicular to the tack allowing measurement in a wider band. Time elapsed from the emission to return every impulse and inclination angle from the vertical axis the instrument used to determine the relative position of each measured point. The absolute position sensor is determined by GPS every second. Data laser scanning combined with modern scanners and orientation to obtain three-dimensional coordinates of the laser footprint on the surface of the field (Mijic et al., 2017). Drone-based laser scanning speeds up the data collection stage of the workflow. In this paper techniques of drone-mounted 3D laser scanning are presented. It will be used a orthophoto input which is later processed. Along with combination of the appropriate software a surface model based on point clouds and Delenay triangulation will be created.

2. METHODOLOGY OF THE DATA COLLECTION

For Surveying and Civil Engineering the most important applications are aerial scanning and terrestrial scanning. Terrestrial Scanning creates 3D models of complex objects: piping networks, roadways, archeological sites, buildings, bridges, etc. Aerial Scanning has many uses - measuring agricultural productivity, distinguishing faint archeological remains, measuring tree canopy heights, determining forest biomass values, advancing the science of

geomorphology, measuring volcano uplift and glacier decline, measuring snow pack, and providing data for topographic maps (Rankin, 2013). In **Figure 1** the process of the using LiDAR techniques has been shown.



Figure 1: *Terrestrial LiDAR technique measuring roadways*

Process of creating point clouds isn't easy and takes a lot of time. Only for processing of the data set few different softwares have been used.

3. LIDAR PLATFORMS

Airborne topographic LiDAR systems are the most common LiDAR systems used for generating digital elevation models for large areas. LiDAR was first developed as a fixed-position ground-based instrument for studies of atmospheric composition, structure, clouds, and aerosols and remains a powerful tool for climate observations around the world. Modern navigation and positioning systems enable the use of water-based and land-based mobile platforms to collect LiDAR data. Data collected from these platforms are highly accurate and are used extensively to map discrete areas, including railroads, roadways, airports, buildings, utility corridors, harbors, and shorelines. Two different techniques of collecting data, from a boat and from vehicle has been shown on **Figure 2**.

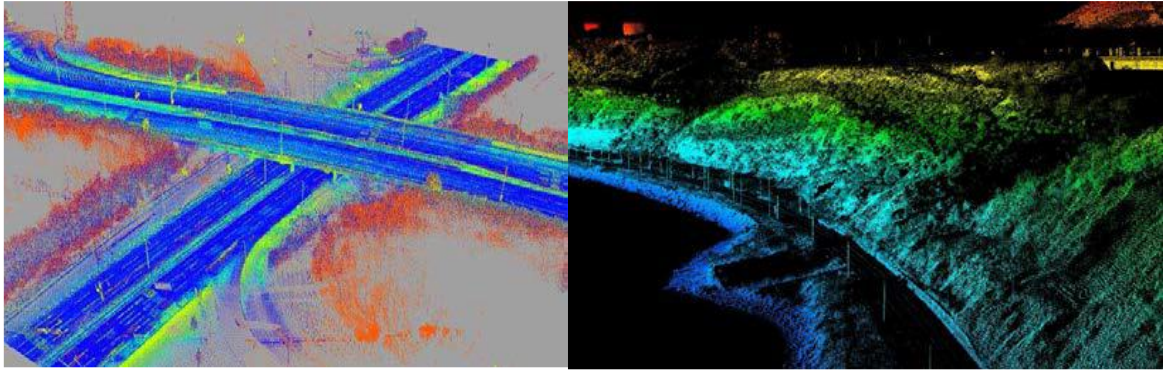


Figure 2: Mobile LiDAR collected from a vehicle (left) and a boat (right)

Airplanes and helicopters are the most common and cost-effective platforms for acquiring LiDAR data over broad, continuous areas. Airborne LiDAR data are obtained by mounting a system inside an aircraft and flying over targeted areas (Carter et al., 2012).

3.1. Techniques used for creating Digital Elevation Models

Basis for the terrain modeling is a set of scattered points in 3D space. For subsequent modeling various mathematical functions can be used. In any case, the representation of the terrain surface is realized by small surface elements. Terrain Modeling using the network quadrilaterals (GRID) is more suitable for organizing and storing data in the matrix form, and later to use various algorithms for data processing (Janic et al., 2015). DEM data are commonly in raster files (**Figure 3**) with formats that include GeoTiff (.tif), Esri Grid (.adf), floating point raster (.flt), or ERDAS Imagine (.img). In some cases the data are available in a TIN format (e.g., Esri TIN). In the raster cases, they are created using point files and can be interpolated using many different techniques. The techniques used to create DEMs range from simple (e.g., nearest neighbor) to complex (e.g., kriging) gridding routines can create slightly different surface types.

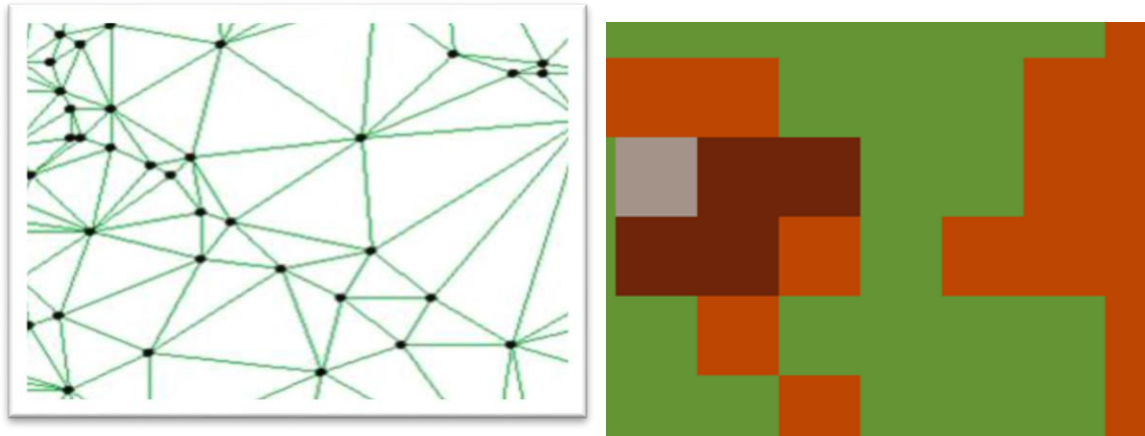


Figure 3: Surface represented like a TIN (left) and as a raster (GRID) (right)

The most common are the surfaces created by TIN or the inverse distance weighted (IDW) routines. The appropriate interpolation method depends on the data and the desired use of the DEM (Carter et al., 2012).

4. EXPERIMENTAL RESULTS – CASE STUDY OF THE QUARRY

Experimental results of this paper are based on creating of 3D models of the quarry. These 3D models are based on the point clouds which are part of airborne images created with new modern technology. First of all a drone was used to acquire all area of the stone pit which is located in Bosnia and Herzegovina. These images are processed in Pix4D software. Processing of the images is the first stage after drone shooting of the area. After processing this software gave us point clouds which can now be use for creating 3D model of the quarry. Besides of the software Pix4D, for the processing of the data Agisoft PhotoScan can also be used. During the data processing we must do the classification of the objects and vegetation and separate terrain from these objects. After processing airborne images and getting point clouds which are in .las or .laz files we must do the conversion if we want to create 3D models of the terrain. Conversion of the .las and .laz files were done in Autodesk ReCap convertor. Procedure for conversion is very easy, we just import .las or .laz files and create new project in Autodesk ReCap and export it like a .rcp files. These .rcp files can be used to import them in AutoCAD Civil 3D and analyse and create

a 3D model of the surface, in this case quarry surface. Imported .rcp file in AutoCAD Civil 3D is shown on the **Figure 4**.

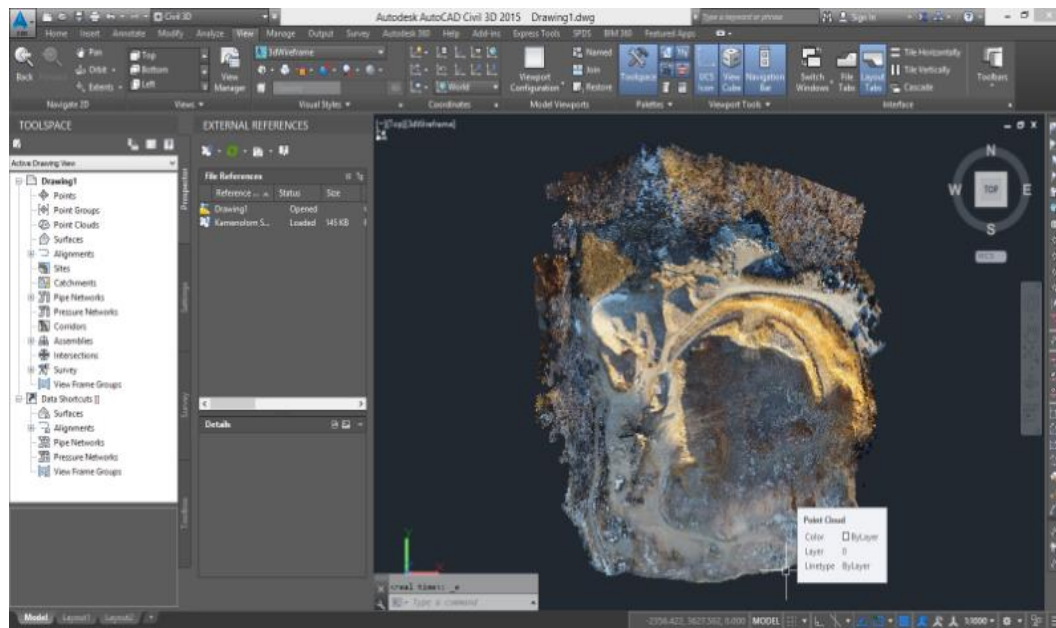


Figure 4: Quarry processed point cloud

One can use the Create Surface from Point Cloud command to create a surface from several point clouds, selecting only the areas that one wants to include and filtering out non-ground points so they are not included in the resulting surface. When using this command, user can select entire point clouds or areas of point clouds to include in the surface. One can select areas of point clouds by using window selections, by defining polygon areas, or by selecting existing closed polylines in the drawing. On the **Figure 5** is shown a surface model of quarry.

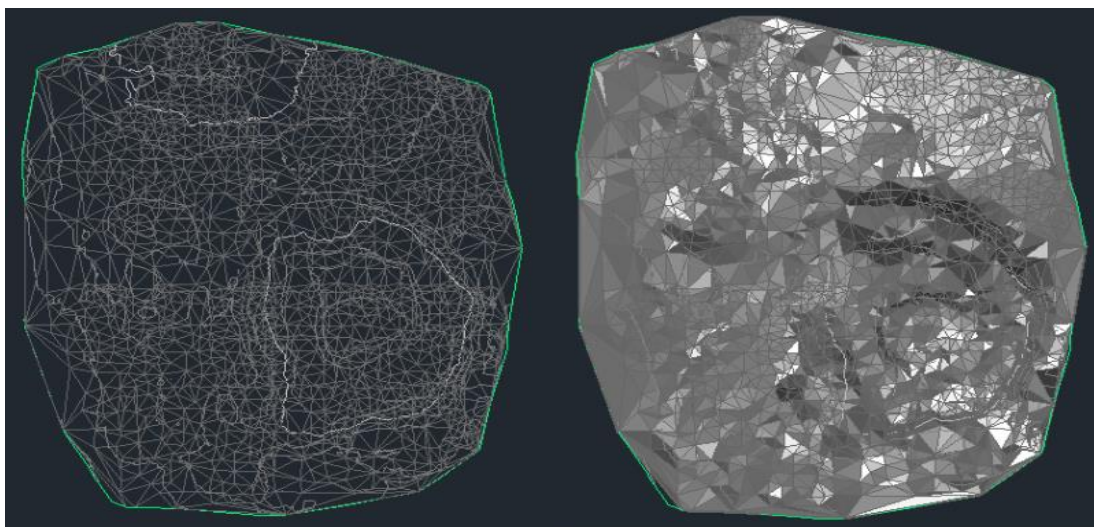


Figure 5: Surface model of the quarry, triangulation network (left) and 3D model of surface (right)

Because drone-mounted 3D laser scanner includes a GPS receiver and inertial measurement unit (IMU), data can be geo-referenced to an exact location. This ensures that each inventory operation can be referenced to the same coordinate system and pile limits each time the inventory is performed: a key benefit when accountants require accuracy and traceability of methods. On the **Figure 6** different ways of seeing an quarry is shown.

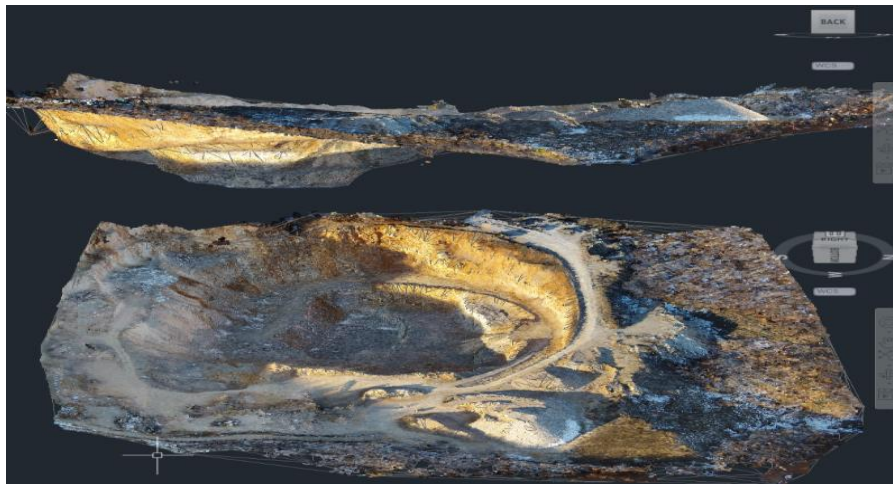


Figure 6: Point clouds of the quarry shown in different projections

Once a point cloud-based surface is created within AutoCAD Civil 3D, it can be used built-in tools to perform quick volumetric calculations (e.g. against previously measured structure), and easily create alignment profile cross sections using nothing but polylines drawn a top of a generated TIN like it is shown on **Figure 7**.



Figure 7: Creating of the alignment using 3D surface model of the quarry

5. CONCLUSIONS

For the creating 3D models of the surface we can use very different platforms. In this paper it was shown how to convert and process different files and after that use them for different purposes. It was used a specific case of the quarry which is located in Bosnia and Herzegovina. In this experimental research it was used different platforms and software tools which are based on processing LiDAR data. Creating a point clouds are not easy and process of import data sets which are recorded with drones. The main purpose of this work was to show different tools and platforms for processing LiDAR data. On the other hand it is shown how we can create from this data cross sections after we created 3D surface model, in this case model of the quarry.

REFERENCES

- CARTER, J., SCHMID, K., WATERS, K., BETZHOLD, L., HADLEY, B., MATAOSKY, R. & HALLERAN, J. (2012): Lidar 101: An Introduction to Lidar Technology, Data, and Applications. Charleston: NOAA Coastal Services Center, 76 p.
- JANIC, M., DJUKANOVIC, G., GRUJOVIC, D. & MIJIC, N. (2015): Eartwork volume calculation from digital terrain models. Journal od industrial design and engineering graphics, 10, pp. 27-30.
- MIJIC, N., SESTIC, M. & KOLJANCIC, M. (2017): CAD—GIS BIM Integration—Case Study of Banja Luka City Center. In: Springer International Publishing (ed.): Advanced Technologies, Systems, and Applications, pp. 267-281.
- RANKIN, F. A. (2013): LiDAR Applications in Surveying and Engineering. Raleigh, GIS Conference Raleigh, NC.

Part VII – Diverse faces of Geomathematics - Poster Session

Application of cluster vs. directional classification method to identify lithofacies and sedimentary elements

Janina Horváth

University of Szeged, Institute of Geography and Geology, Department of Geology and Paleontology;

th.janina@geo.u-szeged.hu

This study compares two multivariate methods to identify lithofacies and sedimentary elements. The applied techniques were tested on the log data obtained from a Miocene clastic depositional reservoir (Algyó field, Hungary). Conventionally direct observations of underground cores are applied to define lithofacies but this process is not always feasible (e.g. lack of cores, expensive method). The basic idea behind the comparison of the two applied grouping methods is that the low number of core samples can be difficult to extend as lithology information. This paper focuses on two statistical processes which use core information in two different ways. Both processes use neural network technology: in this work the results of an unsupervised neural network clustering are compared with those of a directional classification technique. The first one is based on interpreted well-log parameters (porosity, permeability, shale and sand content) while it separates more or less homogeneous subsets. This method uses the core information as supplementary data (for the proper number and labelling of clusters). The directional classification also uses petrophysical attributes besides the core samples in a direct way. Classification analysis is concerned with finding a class membership for a set of input variables. The class membership is determined by the categorical levels of the target variable based on the core information. In this process the small number of core samples can be handled by a random sampling with replacement during the classification. These two applied grouping processes have very similar results. However, the differences can reveal some corner-points of the identification of lithofacies and sedimentary elements.

Key words: *lithofacies, supervised and unsupervised grouping analysis*

1. INTRODUCTION

The basic idea behind the study is that the low number of core samples can be difficult to extend as lithology information. The reason for comparing the two different applications of grouping methods is that the clustering and the

directional classification are able to use core information in different ways. Originally, the goal of the study is to compare the clustering and the directional classification using core samples as well as several petrophysical well-logs in the facies analysis. A related study (Horváth et al., in press) was concerned with the optimization of clustering and the evaluation of the proper number of clusters which are able to describe the sedimentary and lithological facies. The results of this study are compared with another grouping process which is a kind of directional classification. Furthermore, this attempt raises other question besides the comparison. The directional classification may be also a potential validation method of the optimized clustering.

This paper demonstrates the method through a case study. In the study core samples were also available from four wells and one of them included continuous data from a thickness interval of about 35 meters. In this study, the descriptions of core samples were obtained from well-reports and published descriptions by Borka (2016).

Due to the low number of core samples, it is difficult to extend the lithology information to the whole area which contains 141 wells. Hence, the interpreted logs (porosity, permeability, sand content and shale content) are used to define the lithology types and facies in the case of clustering, and the core samples represent supplementary information. In the case of the directional classification these core data are applied directly and these define the target variables.

The study area is located in the Algyő sub-basin of the Pannonian-basin geographically belonging to the Great Hungarian Plain. According to the paper by (Grund & Geiger, 2011; Borka, 2016) this study area can be characterized as sequences of prodeltaic submarine fan which is a typical mixed sand-mud submarine fan complex.

2. METHODS

2.1. Pre-processing

The same process was applied to handle the petrophysical data for both methods. The grouping processes may be more efficient if a good structure exists for the transformed variable, which can approximate the symmetric

distribution (Templ et al., 2006). Due to the significant skewness of the distribution of the variables, especially in the shale content and permeability (Horváth et al., in press), the data was transformed by a Box-Cox transformation. A principal component analysis (PCA) was used to reduce redundancy and to create new components. The first component is based on permeability and porosity while the second component is based on sand content and shale content (HORVÁTH et al., in press). This component was used in the case of both grouping processes.

2.2. Grouping processes

2.2.1. UNN SOM clustering

Both applied grouping methods are based on neural network (NN) processes. In the clustering an unsupervised neural network (UNN) clustering the so-called Self Organised Map (SOM) is applied. This method uses the core information as supplementary data for the proper number and labelling of clusters. The input variables of the UNN-SOM are transformed petrophysical data. **Figure 1** illustrates the workflow of the clustering process.

To determine the stable separation DFA with LOO cross validation technique is a possible approach (Asante & Kreamer, 2015). A cluster structure was declared stable if DFA predicted at least 80% of the members in each cluster grouping. Nevertheless, the appropriate separation is also based on the number of clusters. However, as clustering is an unsupervised technique the researcher has only little or no information about the number of clusters. At the same time, the number of clusters is a required parameter. The optimisation of the right number of separated clusters is also a feedback point in the workflow (**Figure 1**). This problem can be solved by several of the suggested indexes depending on the sum of squares called WB indexes (e.g. Hartigan, 1975; Calinsky & Harabasz, 1974).

The core samples as supplemental information was able to support the right number of clusters in the initialization of UNN SOM. According to the core information, in the related study (Horváth et al., in press) the number of clusters was increased from the value of 3 to 8 one by one.

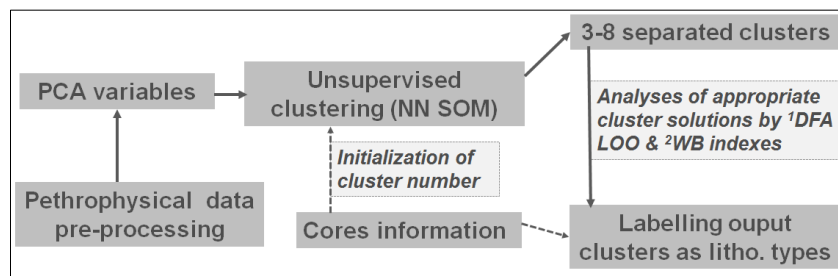


Figure 1: Workflow of the UNN SOM clustering with additional processes (¹Discriminant analysis, leave one out - cross validation; ²Sum of squares indexes - determining the right number of clusters)

After the NN SOM clustering, evaluation of the optimal clustering solution relied on more WB indexes. All of them determined the "best fitting clustering" to be the five clusters solution and 92% of cross-validated grouped cases were correctly classified.

The five groups were labelled based on the core samples as: (1) siltstones and marls, interbedded sandstones; (2) spatially dispersed, low permeability sandstones; (3) alternation of siltstones and sandstones; (4) silty sand; (5) massive sandstones.

2.2.2.MLP NN classification

In the directional classification a kind of the Multi Layer Perceptron model (MLP) was used. This grouping method also uses petrophysical attributes besides the core samples in a direct way. Classification analysis is concerned with finding a class membership for a set of input variables. The class membership is determined by the categorical levels of the target variable based on the core information.

Four categories were defined after Borka (2016): (1) siltstones-marlstones with interbedded sand; (2) sandstones with interbedded marlstones; (3) alternation of siltstones and fine sandstones; (4) alternation of fine sandstones and silty sandstones.

These were the target variables in the classification and the related petrophysical data were the test samples in MLP NN. **Figure 2** illustrates the workflow of classification process.

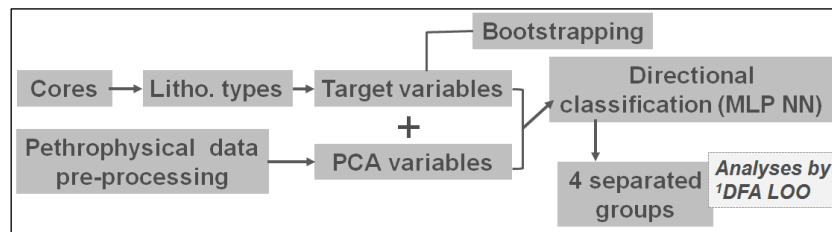


Figure 2: Workflow of the UNN SOM clustering with additional processes

In this process the small number of core samples can be handled by a random sampling (bootstrapping in **Figure 2** with replacement during the classification). After the classification, the DFA with LOO cross validation were also applied to measure the stability of the separated sub-sets. Members in each cluster groupings were validated by over 88% prediction.

3. COMPARING THE RESULTS OF THE DIFFERENT METHODS

It is very important that in the case of clustering the final number of separated sub-sets were five. In contrast, in the case of classification only four categorical levels were determined. Naturally, this raises the question, why only four lithofacies were determined as target variable in the classification instead of five types. This happened for the quite simple reason - there is a scaling difference between the description of core samples and the applied well-logs. Consequently, it would be hard to identify smaller scaling lithofacies as target variables. However, based on the general statistical characters (**Table 1-2**) of the separated groups, it is easy to identify the related sub-sets from the different methods. Also, the statistical parameters (**Table 1-2**) reveal that cluster 4-5 together (from NN SOM) are related with the category 4 in the case of NN MLP classification. This statement is supported by the distributions of petrophysical parameters in these sub-sets (**Figure 3**). In addition, the statement is also verified by the comparison of thickness maps of the defined lithofacies.

Table 1: Statistical characterisation of clusters by NN SOM method

	POR					PERM				
	1	2	3	4	5	1	2	3	4	5
N	182	252	556	503	328	182	252	556	503	328
Mean	12.86	14.41	16.39	18.39	20.25	2.18	7.84	12.04	32.2	87.2
Median	12.84	14.49	16.45	18.35	20.23	1.22	5.56	11.54	31.1	79
	VSHA					VSND				
	1	2	3	4	5	1	2	3	4	5
N	182	252	556	503	328	182	252	556	503	328
Mean	33.07	11.29	22.83	15.3	8.79	52.6	66.67	59.54	65.9	71.2
Median	32.13	11.38	22.5	15.45	9.04	53.87	68.92	60.47	65.8	70.3

Table 2: Statistical characterisation of classified groups by NN MLP method

	POR				PERM			
	1	2	3	4	1	2	3	4
N	342	383	345	751	342	383	345	751
Mean	14.5343	15.0556	16.716	19.2796	6.14914	10.2745	14.1492	57.2057
Median	15.0803	15.1411	16.8335	19.101	5.69825	5.966	14.2686	44.7525
	VSHA				VSND			
	1	2	3	4	1	2	3	4
N	342	383	345	751	342	383	345	751
Mean	28.0783	12.3347	23.6819	12.8359	56.628	66.0796	59.0639	67.8459
Median	26.8847	12.2138	22.451	12.9245	57.932	67.949	60.319	67.561

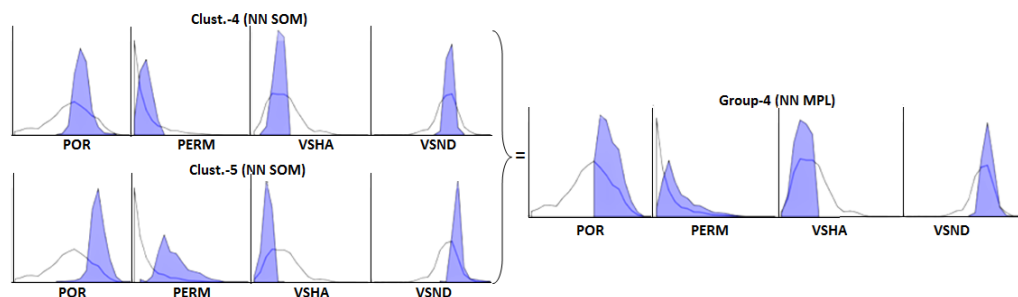


Figure 3: Distributions of petrophysical parameters in the Clust.-4, Clust.-5 (NN SOM), and Group-4 (NN MLP) sub-sets

This comparison shows that the clustering results in a more detailed division of *alternation of fine sandstones and silty sandstones* lithofacies (Group-4 NN MLP). In the case of clustering it is defined as (Clust.-4) silty sand; (Clust.-5) massive sandstones.

The statistical parameters (**Tables 1-2**) of separated sub-sets in both method cases show that there is a lithofacies (Clust.-2 and Group-2) which is

characterized by large sand content (~67%) but relatively low porosity (~14-15%) and permeability properties (8-10mD). These are defined as Clust.-2: *spatially dispersed, low permeability sandstones* and Group-2: *sandstones with interbedded marlstones*. These sub-sets which come from different methods are characterized by very similar features and what is more the spatial extension of these facies is almost the same while the thickness maps of these two lithofacies are also similar.

The sub-sets marked by Clust.-3 & Group-3 are also related based on statistical characters and spatial extensions. The significant differences may be revealed between the sub-sets marked by Clust.-1 & Group-1. In contrast to the clustering method, the directional method based on core information classifies much more data points into the group which is characterized by low por-perm and sandcontent features. However, the spatial extensions of these sub-sets are almost the same. The difference originates from the fact that the directional classification method separated sub-sets with smaller ranges in case of all petrophysical features (according to the distribution of petrophysical parameters).

4. CONCLUSION

It can be stated that there are no significant differences between the results of the various separation techniques. The comparison reveals only little difference which means e.g. (1) more detailed separation in the case of one lithofacies: *alternation of fine sandstones and silty sandstones* lithofacies (2) refinement of petrophysical features by smaller range in the case of *siltstones-marlstones with interbedded sand* lithofacies.

Accordingly, this comparison between the two separation processes shows that the directional classification is a kind of a validation process of the optimized clustering results. The directional classification can be a validation method, especially because it uses the lithology information based on core samples. Consequently, the separated facies can be defined as depositional facies by the depositional structures. Thus, based on the core descriptions by Borka (2016)

the separated sub-sets are identified as the following facies: (1) deposit of slump; (2) bottom current; (3) sandy debris flow; (4) sandy debris flow/turbidity current.

ACKNOWLEDGEMENT

This paper was supported by the New National Excellence Program of the Ministry of Human Capacities.



REFERENCES

- ASANTE, J. & KREAMER D. (2015): A New Approach to Identify Recharge Areas in the Lower Virgin River Basin and Surrounding Basins by Multivariate Statistics, *Mathematical Geosciences*, 47/7, 819-842.
- BORKA, SZ. (2016): Markov chains and entropy tests in genetic-based lithofacies analysis of deep-water clastic depositional systems, *Open Geosciences*, 8, 45–51.
- CALINSKI, T., & HARABASZ, J. (1974): A dendrite method for cluster analysis, *Communications in Statistics*, 3, 1, 1-27.
- GRUND, S.Z., & GEIGER, J. (2011): Sedimentologic modelling of the Ap-13 hydrocarbon reservoir, *Central European Geology*, 54, 4, 327-344.
- HARTIGAN, J. A. (1975): *Clustering Algorithms*, John Wiley and Sons, Inc., NY, USA, 351 p.
- HORVÁTH, J., BORKA, SZ. & GEIGER, J. (in press): Cluster defined sedimentary elements of deep-water clastic depositional systems and their 3D spatial visualization using parametrization: a case study from Pannonian-basin, *Geologia Croatica*
- TEMPL, M., FILZMOSER P. & REIMANN C. (2006): Cluster analysis applied to regional geochemical Data: problems and possibilities, Research report - CS-2006-5, Vienna University of Technology, 39 p.

Comparison of parametric and non-parametric time-series analysis methods on a long-term meteorological dataset

Tímea Kocsis¹, Angéla Anda²

¹ Budapest Business School College of Commerce, Catering and Tourism Department of Methodology,
H-1054 Budapest, Alkotmány Str. 9-11. Tel.: +36-1/374-6244, jakuschneKocsis.Timea@uni-bge.hu

²University of Pannonia Georgikon Faculty Department of Meteorology and Water Management, H-
8360 Keszthely, Festetics Str. 7.

In the present explorative study different time-series analysis methods, such as moving-average, deterministic methods (linear trend with seasonality) and non-parametric Mann-Kendall trend test were used on 144 years of precipitation data with the aim of to compare the results of these methods and to detect the signs of climate change. Dataset was provided by the University of Pannonia, and it contains monthly precipitation data of 144 years measurements at Keszthely Meteorological Station. This dataset is special because few stations in Hungary have continuous measurements of more than 140 years with detailed historical background. Results of the research can give insight to the signs of climate change in the past for the region of West-Balaton.

Parametric methods (linear trend, t-test for slope) for analysing time series are the simplest methods to get insight to the changes in a variable over time. These methods have a requirement for normal distribution of the population that can be a limit for application. Non-parametric methods are distribution-free methods and investigators can have a more sophisticated view to the variable tendencies in time series.

Key words: *Keszthely, linear trend, long-term precipitation data set, Mann-Kendall trend test, seasonality, moving-average, signs of climate change*

1. INTRODUCTION

Climate change is one of the problems that mankind should face in the 21st century. According to the last *IPCC report* (2013) human role in the process is without doubt and 95% is the probability that human influence has been dominant on the present changes of climate system. Hydrological cycle is one among the elements of the climate system that is expected to change and the signs of these modifications can already be detected. Changes are different in signs and volumes on the different parts of the Earth.

Changes can be detected not only in the changes of the annual sums, but within seasons as well. Between 1901 and 2009 the highest precipitation decline over Hungary (Central-Europe) occurred in the spring, nearly 20% (Lakatos & Bihari, 2011). There have been several extreme precipitation indices examined that suggest the increase of the regional intensity and frequency of extreme precipitation in the Carpathian-Basin in the second half of the last century, while the total precipitation decreased (Bartholy & Pongrácz, 2005; 2007; 2010). However the hydroclimate of the region is quite variable in space and time (Kern et al., 2016).

The aim of this study is to analyse the long-term data series of the meteorological measurements of precipitation amount at Keszthely (Western Hungary, N 46°44', E 17°14') between 1871 and 2014 from the point of view of climate change, and to compare different statistical methods on the results of time series analysis based on this dataset.

2. DATASET AND METHODS

Monthly amounts of precipitation were analysed from 1871 to 2014 (1728 data) measured at the beginning in the territory of the ancient Georgikon Academy of Agriculture at Keszthely, then at the meteorological station of the Hungarian Meteorological Service. The dataset was provided by the Department of Meteorology and Water Management of University of Pannonia Georgikon Faculty (Keszthely). This dataset is special because few stations in Hungary have continuous measurements of more than 140 years with detailed historical background (Kocsis & Anda, 2006). Different time-series analysis methods, such

as moving-average, deterministic methods (linear trend with average seasonality) and non-parametric Mann-Kendall trend test were used on 144 years of monthly precipitation data with the aim of to compare the results of these methods and to detect the signs of change in the hydrological cycle. Parametric methods (linear trend, t-test for slope) for analysing time series are the simplest methods to get insight to the changes in a variable over time. A non-parametric method, Mann-Kendall trend test /see detailed descriptions given by Gilbert (1987) and Hipel & McLeod (1994)/ is widespread for analysing meteorological (such as precipitation sums) and hydrological data and it has no requirements for normal distribution as linear trend (parametric “regression on time” method) has. Mann-Kendall trend test is based upon the work of Mann (1945) and Kendall (1975), and is closely related to Kendall’s rank correlation coefficient. Hirsch et al. (1982) and Hirsch & Slack (1984) developed the method and introduced seasonal Mann-Kendall test for data that are serially dependent. Sen’s slope estimator (Sen, 1968) was applied as non-parametric method for determine the slope of the tendencies. Kolmogorov-Smirnov test was applied to check the normal distribution. In all hypothesis tests significance level was fixed at $\alpha=5\%$. Addinsoft’s XLSTAT 2016 was applied for carrying out the results.

3. RESULTS

3.1. Results of “regression on time” and average seasonality

1728 monthly precipitation data were analysed and mean monthly precipitation at Keszthely is 56 mm with standard deviation of 37 mm. Linear tendency cannot be detected with two-tailed ($H_a: \beta_1 \neq 0$) t-test ($\hat{y}_t = 59 - 0.003 * t$) at $\alpha=5\%$ (p-value was 6.2%), but in case the sign of the changes is determined (one-tailed test of $H_a: \beta_1 < 0$), significant modification can be proved with p-value of 3.1% (**Figure 1.-2.**). In 100 years 0.3 mm of decrease can be detected in monthly precipitation sums.

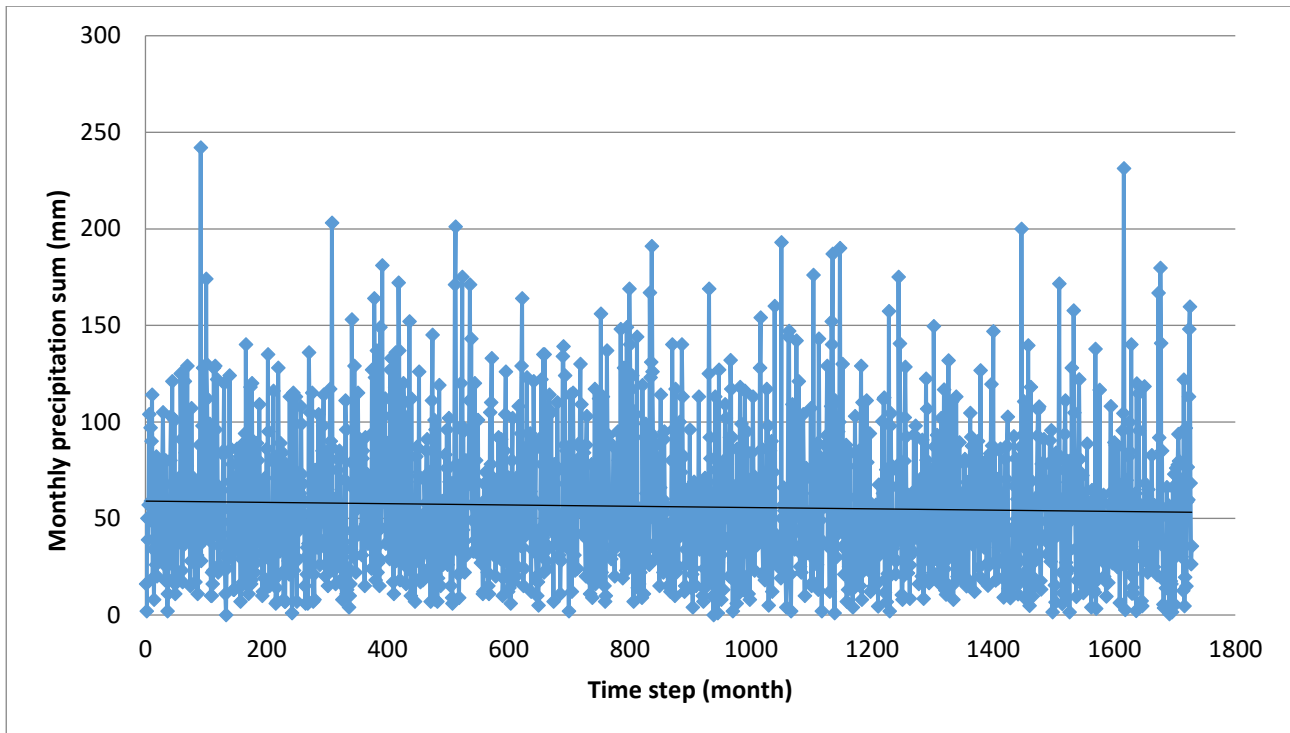


Figure 1: Time series of monthly precipitation sum at Keszthely between 1871 January and 2014 December

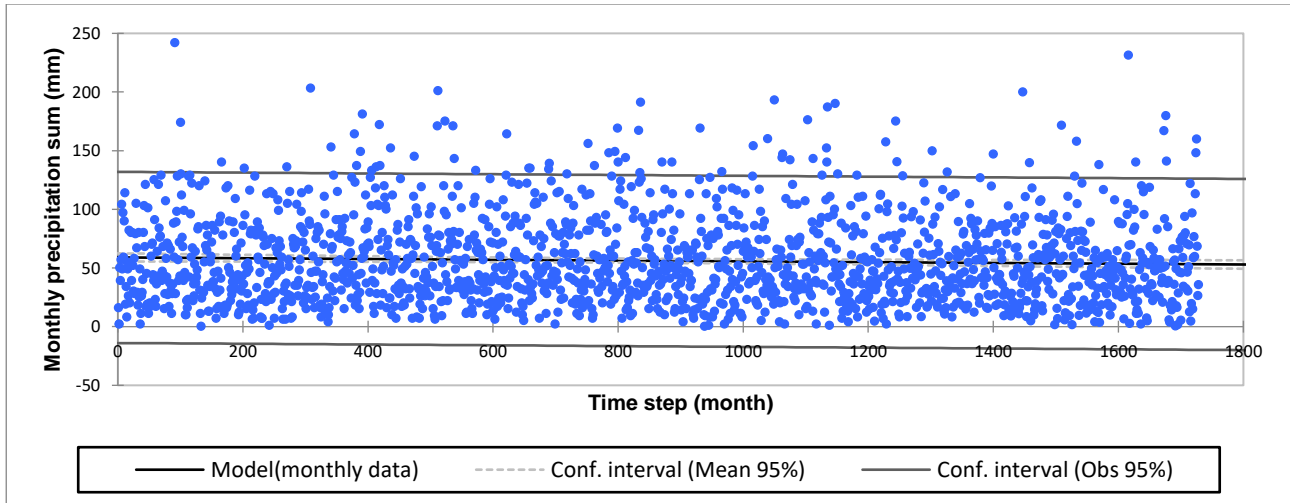


Figure 2: Regression of monthly precipitation sum by time step ($R^2=0.002$)

Moving-average with tags of 12 sums can be used as a smoothing method that can eliminate the effect of the seasonality in the data series. Tendency of the 12MA (moving average) is not so clear on **Figure 3**.

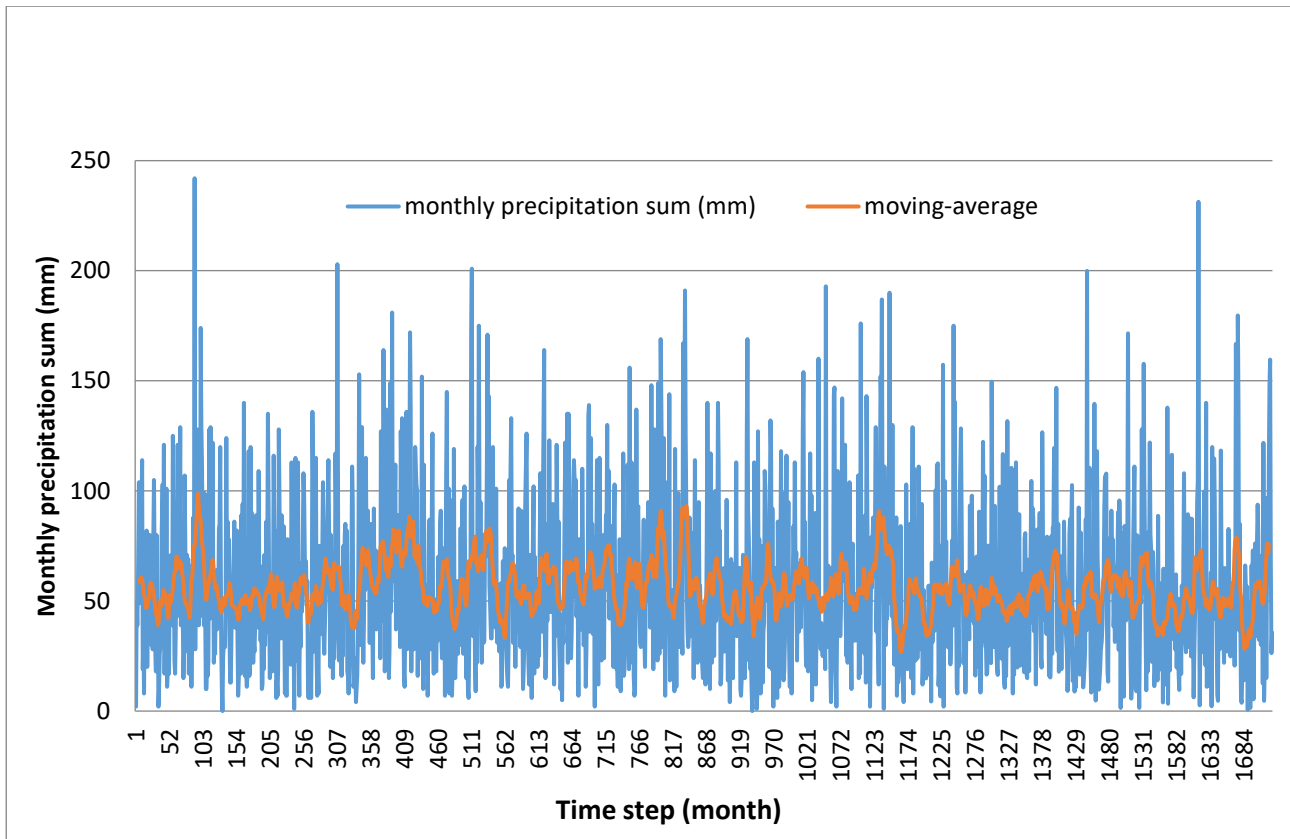


Figure 3: Monthly precipitation sum and moving average (12 MA) between 1871 and 2014

Deterministic trend analysis can be followed by the decomposition of the time series data to trend, average seasonality and random component. The mean tendency is modified by seasonal effect that can be described by corrected mean seasonal deviation. Corrected mean seasonal deviation gives the average volume how much the seasonality increase or decrease the value given by the main trend (**Table 1.**).

Table 1. *Corrected mean seasonal deviation in each season (1871-2014)*

Season	Corrected mean seasonal deviation from trend (mm)
January	-23.5
February	-22.8
March	-17.0
April	-2.8
May	13.3
June	19.9
July	17.7
August	16.9
September	4.9
October	2.3
November	1.9
December	-10.8

3.2. Result of non-parametric methods

A parametric method, as “regression on time” is commonly used method to determine the main tendency of the time series, but t-test for the slope has requirement for normal distribution. Kolmogorov-Smirnov test was used to check normality, and it can be concluded that this dataset does not follow normal distribution (p-value is lower than 0.01%). In this case non-parametric methods can give more appropriate results about the trend. Two-tailed Mann-Kendall trend test gave non-significant result with p-value of 6.48%, but in case the sign of the changes is determined (one-tailed test of $H_a: \tau < 0$), significant decreasing modification can be proven with p-value of 3.24%. Sen’s slope estimator gives a slope of -0.003 similarly to linear trend.

As the time series contains seasonal component, the values are not serially independent. Seasonal Mann-Kendall trend test was also applied, and two-tailed test showed that there is no significant trend in the time series (p-value was 7.73%). One-tailed test proved significant negative tendency at p-value of

3.86%. Sen's slope was -0.033 that means 3.3 mm decline in 100 years taking into account the effect of seasonality.

4. CONCLUSIONS

Parametric and non-parametric methods both proves significant negative trend in the time series of monthly precipitation sums at Keszthely. As the data does not follow normal distribution non-parametric Mann-Kendall trend test gives more adequate result. When analysing monthly precipitation sums, the effect of seasonality leads to the serial dependence of the data. This fact must be taken into account. Sen's slope showed 3.3 mm decreasing tendency of the monthly sums in 100 years taking into account also the effect of seasonality. As an outlook the time series assessed in the study should be taken into account in climate studies dealing with low frequency signals (Sen & Kern, 2016).

REFERENCES

- BARTHOLY, J. & PONGRÁCZ, R. (2005): Extremes of ground-based and satellite measurements in the vegetation period for the Carpathian Basin. *Physics and Chemistry of the Earth* 30: 81-89.
- BARTHOLY, J. & PONGRÁCZ, R. (2007): Regional analysis of extreme temperature and precipitation indices for the Carpathian Basin from 1946 to 2001. *Global and Planetary Change* 57: 83-95.
- BARTHOLY, J. & PONGRÁCZ, R. (2010): Analysis of precipitation conditions for the Carpathian Basin based on extreme indices in the 20th century and climate simulation for 2050 and 2100. *Physics and Chemistry of the Earth* 35: 43-51.
- GILBERT, R. O. (1987): *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Company, NY, USA: 208-224.
- HIPEL, K. W. & McLEOD, A. I. (1994): *Time series modelling of water resources and environmental systems*. Elsevier, Amsterdam, The Netherlands: 864-866, 924-925.
- HIRSCH, R. M. & SLACK, J. R. (1984): A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20 (6): 727-732.
- HIRSCH, R. M., SLACK, J. R. & SMITH, R. A. (1982): Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18 (1): 107-121.

IPCC AR5 (2013): Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change –In: Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

KENDALL, M. G. (1975): Rank correlation methods. Charles Griffin, London

KERN, Z., NÉMETH, A., HOROSZNÉ GULYÁS, M., POPA, I., LEVANIC, T. & HATVANI, I. G. (2016): Natural proxy records of annual temperature- and hydroclimate variability from the Carpathian-Balkan Region for the past millennium: review & recalibration. Quaternary International 415: 109-125.

KOCSIS, T. & ANDA, A. (2006): History of the meteorological observations at Keszthely. Published by University of Pannonia Georgikon Faculty, Keszthely ISBN 963 9639 07 9 (In Hungarian)

LAKATOS, M. & BIHARI, Z. (2011): Temperature- and precipitation tendencies observed in the recent past. - In: Climate Change 2011 (Eds: Bartholy, J., Bozó, L., Haszpra, L.): 159-169. (In Hungarian)

MANN, H. B. (1945): Nonparametric tests against trend. Econometrica 13: 245-259.

SEN, A. & KERN, Z. (2016): Wavelet analysis of low-frequency variability in oak tree-ring chronologies from east Central Europe. Open Geosciences 8: 478–483.

SEN, P. K. (1968): Estimates of the regression coefficient based on Kendall's tau. Journal of American Statistical Association 63 (324): 1379-1389.

XLSTAT 2016, Addinsoft (<https://www.xlstat.com/en/>)

Variogram analysis of precipitation $\delta^{18}\text{O}$ across the Iberian Peninsula on monthly-, seasonal- and annual scales

Dániel Erdélyi^{1,2}, Zoltán Kern², István Gábor Hatvani^{2*}

¹Eötvös Loránd University, Centre for Environmental Sciences, H-1117 Budapest, Pázmány Péter stny.

1/A, danderdelyi@gmail.com

²Institute for Geological and Geochemical Research, Research Centre for Astronomy and Earth

Sciences, Hungarian Academy of Sciences, H-1112 Budapest, Budaörsi út 45.,

zoltan.kern@gmail.com, hatvaniig@gmail.com

The aim of the study was to explore the spatial variance of precipitation $\delta^{18}\text{O}$ across the Iberian Peninsula. Monthly precipitation stable isotope data (October 2002-September 2003) from 32 GNIP stations in conjunction with 21 stations from a regional monitoring campaign (October 2004-June 2006) were acquired. Monthly semivariograms were calculated for 10.2002-09.2003 on monthly stable isotope data, and seasonal/annual ones on amount weighted means. A common critical weakness that may be a major cause of nugget-type semivariograms is underdeveloped sampling in space. More than half of the monthly semivariograms, showed a rising section and stabilization around the mean variance. Unfortunately, due to the lack of data from short distances, theoretical semivariograms were unsatisfactory. However, for the dataset as extended by the 21 stations an exemplary empirical semivariogram was obtained with two sills, corresponding to ~100 km and ~500km spatial ranges.

Key words: *geostatistics, isotope hydrometeorology, precipitation stable isotope, spatial variance*

1. INTRODUCTION

The isotopic composition of precipitation gives insights into the origin of water vapor and the conditions obtaining during its condensation and precipitation (Dansgaard, 1964). A major part of tropospheric water vapor originates from the evaporation of warm subtropical oceans. In the clouds, continuous equilibrium fractionation takes places between the liquid and vapor phases of water. During the process, condensate water becomes enriched in heavy

isotopes which leave the system as precipitation, and the residual depleted in heavy isotopes. This is known as the Rayleigh distillation process (Dansgaard, 1964).

The analysis of the stable oxygen isotope composition of precipitation ($\delta^{18}\text{O}_p$) can provide valuable information in such diverse fields of science as hydrogeology and ecology. Stable isotope ratios are incorporated into the tissues of animals by the metabolization of food originating from their environment. Thus, if the stable isotope ratios found in hair or feathers are compared with those of the animal's current environment, their movements can be tracked (Hobson, 1999). Another use is the determination of the origin and the flow-direction of sub-surface waters (Clark & Fritz, 1997; Fórizs 2003, Rodriguez-Arévalo et al., 2013). For such applications the development of a proper monitoring network is an unavoidable necessity. The inclusion of point data (from the stations) and external factors in a geostatistical model (isoscape; Bowen, 2010) provides a spatially continuous picture of the isotopic content of precipitation, and therefore contains information essential to the ability to track phenomena related to $\delta^{18}\text{O}_p$. A global isotope-hydrometeorological monitoring network was launched in the 1960s (IAEA, 2016), and a European sub-region of this network with a relatively high abundance of stations is found on the Iberian Peninsula.

The aim, therefore, was to explore the spatial variability of $\delta^{18}\text{O}_p$ on the Iberian Peninsula using variography in order (i) to determine the representativity of its precipitation monitoring network and (ii) provide a stepping stone towards the derivation of a regional isoscape.

2. MATERIALS AND METHODS

2.1. Used materials

Precipitation oxygen stable isotope ratios ($\delta^{18}\text{O}_p$) are expressed as the ratio between the heavy and light stable oxygen isotopes in the sample compared to the Vienna Standard Mean Ocean Water given in per mill:

$$\delta^{18}\text{O} = \frac{{}^{18}\text{O}/{}^{16}\text{O}_{\text{sample}} - {}^{18}\text{O}/{}^{16}\text{O}_{\text{VSMOW}}}{{}^{18}\text{O}/{}^{16}\text{O}_{\text{VSMOW}}} \cdot 1000[\text{‰}]$$

Monthly $\delta^{18}\text{O}_p$ records from 32 Iberian stations of the Global Network of Isotopes in Precipitation (GNIP) (October 2002-September 2003; IAEA, 2016) were acquired, together with those from 21 stations of a regional monitoring campaign (October 2004-June 2006) (Oliveira & Lima, 2010). Precipitation amount data from the GNIP network was found to be in some cases missing/erroneous. Therefore, gridded precipitation data from the GPCC database (Becker et al., 2013) were used to derive the amount weighted stable isotope averages. The available data was unequally distributed in time and space, the highest abundance being in January 2003, when 24 monitoring sites were operating simultaneously.

2.2. Methodology

2.2.1. Variography

The basic function of geostatistics, the variogram, was used to describe the spatial autocorrelation structure of the $\delta^{18}\text{O}_p$ in the Iberian Peninsula to obtain the weights necessary for kriging (Chilès & Delfiner, 2012). The most important properties of the semivariogram are as follows: the nugget holds information regarding the error of the sampling; the sill is the level at which the variogram stabilizes; and the range, which is the distance within which the samples have an influence on each other and outside of which they are quasi-independent (Chilès & Delfiner, 2012). In order to meet the intrinsic hypothesis of geostatistics, that the co-variances between the station pairs of $\delta^{18}\text{O}_p$ depend not on their actual locations, but on their relative distances (Hohn, 1999), trend has to be removed from the data (Füst and Geiger, 2010; Kohán and Szalai, 2014). To achieve the most uniform number of station pairs in the analysis, the semivariograms derived from the monthly GNIP data had 10 bins and a maximum lag distance of 540 km, while those from the extended dataset had 15 bins and a 570 km maximum lag distance.

2.2.2. Steps of the analysis

Since previous studies have pointed out seasonal differences in the spatial trend of $\delta^{18}\text{O}_p$ (Kern et al., 2014), variogram analysis was first conducted on 12

monthly datasets including exclusively GNIP stations. Then, from these two amount weighted seasonal datasets (winter: October-March; summer: April-September) and an annual one were derived. Those GNIP stations where more than 20% of the fallen precipitation did not have corresponding $\delta^{18}\text{O}_p$ were omitted in that particular period (i.e. month, season, year). Where this ratio was 10%-20% the variogram cloud was used to decide whether to keep the data of the particular station or discard them.

Based on our pre-analysis and literature (Oliveira & Lima, 2010) it became clear that the removal of the governing effect of geography (“trend”) is necessary. The best fit models were determined using multiple regression analysis with the independent parameters: POINT X, -Y and Elevation; and were removed from the raw $\delta^{18}\text{O}_p$ data. The 16 models thus obtained were all found to be significant ($p < 0.01$), variance inflation was negligible ($\text{VIF} < 1.157$), and avg. adjusted $R^2 = 0.63$. For the obtained residuals first semivariogram clouds were calculated to facilitate preprocessing/data filtering for outliers, then variography was conducted.

3. RESULTS AND DISCUSSION

3.1. Variography using GNIP data

Seven out of 16 semivariograms (5 monthly: January, March, May, June, July; and the two seasonal: summer, winter) were of nugget-type. As an example, the values of the semivariogram for June 2003 vary around the variance without a rising section (**Figure 1A**). A common critical weakness that may be a major cause for the occurrence of nugget-type of semivariograms is the underdeveloped sampling of $\delta^{18}\text{O}_p$ in space, which, in turn, provides data unsatisfactory for the determination of a sampling range. Unfortunately, due to the lack of data from short distances, a fair portion of the theoretical semivariograms was unsatisfactory. There is no month in which more than two stations were sampling simultaneously within a 100 km distance; moreover, within 50 km, there were none. So, for example, in July 2003 only 13 stations had usable data, which was a critically low number. Therefore, the expected

small variances over short distances were not observed, only the large variance over long distances.

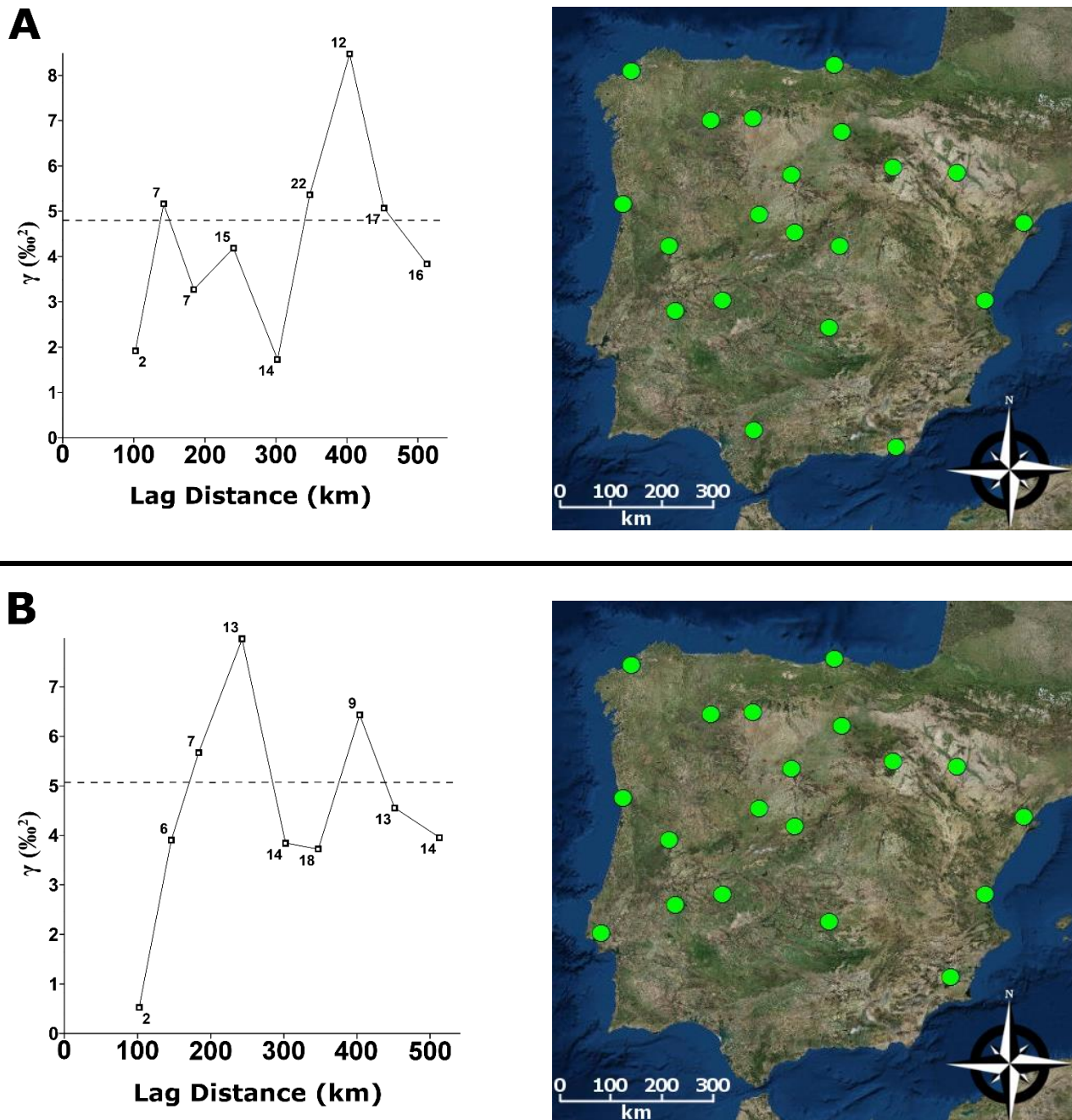


Figure 1: Empirical semivariograms of June- **A)** and August 2003 **B)** and the maps representing the corresponding sampling sites (green dots). The numbers next to certain values of the empirical semivariogram (black squares) represent the data pairs within a given distance (bin width was 54 km) behind the semivariogram. The dashed line represents the variance.

Unfortunately, even in those cases where the semivariogram did outline a rising section (**Figure 1B**), we had to restrain ourselves from fitting theoretical semivariograms, because the number of station pairs was low and the first bin was empty.

3.2. Variography using the extended database

A unique opportunity to extend the GNIP station $\delta^{18}\text{O}_p$ records presented itself, with an additional dataset (Oliveira & Lima, 2010) covering a two year period (October 2004-June 2006). The merged dataset had many more short distance data pairs than that of the GNIP stations (**Figure 2A**).

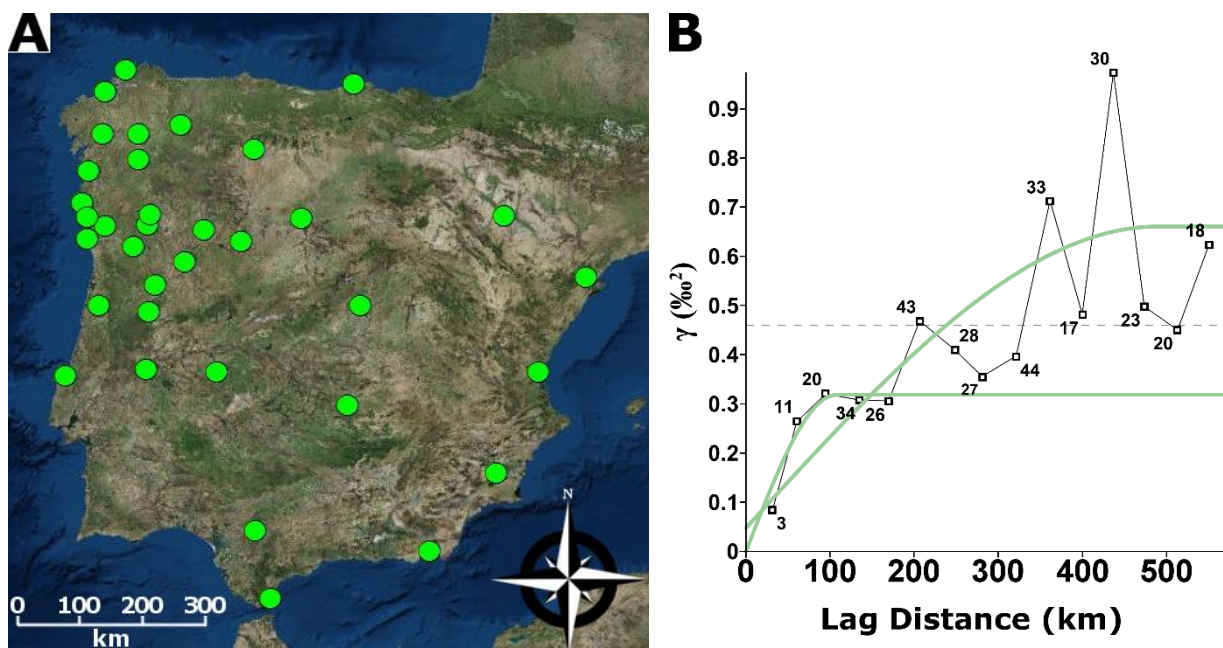


Figure 2: The map representing the sampling sites (green dots) the Empirical- (black line on left panel) and fitted theoretical semivariograms (green line on left panel) were derived from. The parameters of the semivariogram model fitted to the first sill $C0=0.001$; $C0+C=0.318$; range=107km; $r^2=0.979$; $RSS=0.001$; and to the second sill $C0=0.048$; $C0+C=0.613$; range=489km; $r^2=0.57$; $RSS=0.291$. Bin width was 38 km; the dashed line represents the variance.

The semivariogram obtained from the extended dataset had two sills, indicating two ranges: ~ 100 km and ~ 500 km (**Figure 2B**). While the former range concurs with that determined for $\delta^{18}\text{O}_p$ in the Alps (Kern et al., 2015), the latter

is difficult to interpret because it is comparable to the entire East-West extent of the peninsula.

4. CONCLUSIONS

The monthly GNIP $\delta^{18}\text{O}_p$ records (October 2002-September 2003) for the 32 stations distributed across the Iberian Peninsula seem to be unsatisfactory for the determination of a sampling range of the isotope hydrometeorological parameters with variography. However, extended by a 21 station dataset (October 2004-June 2006), a much more dense monitoring network was obtained, one which was now suitable for determining a semivariogram - with two sills at ranges of ~ 100 and ~ 500 km - describing complex isotope-hydrometeorological processes on the peninsula.

ACKNOWLEDGMENTS

Thanks for the support of the MTA “Lendület” program (LP2012-27/2012) and the National Research, Development and Innovation Office (SNN118205) and the Slovenian Research Agency (ARRS:N1-0054). This is contribution No. 45 of 2ka Palæoclimate Research Group.

REFERENCES

- BECKER, A., FINGER, P., MEYER-CHRISTOFFER, A., RUDOLF, B., SCHAMM, K., SCHNEIDER, U. & ZIESE, M. (2013): A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data* 5: 71-99.
- BOWEN, G.J. (2010): Isoscapes: spatial pattern in isotopic biogeochemistry. *Annual Review of Earth and Planetary Sciences*, 38: 161-187.
- CHILÈS, J.-P. & DELFINER, P. (2012): *Geostatistics*. Wiley, Canada.
- CLARK, I. D. & FRITZ, P. (1997): *Environmental isotopes in hydrogeology*, Lewis Publishers, New York, 328 p.
- DANSGAARD, W. (1964): Stable isotopes in precipitation. *Tellus*, 16, 436-468.

- FÓRIZS, I. (2003): Isotopes as natural tracers in the watercycle: examples from the Carpathian Basin. *Studia universitatis Babes-Bolyai, Physica XLVIII*, 69-77.
- FÜST, A. & GEIGER, J. (2010) Setting up monitoring networks using geostatistics I 306 p.
- HOBSON, K.A. (1999): Tracing origins and migration of wildlife using stable isotopes: A review. *Oecologia* 120, 314-326.
- HOHN, M.E. (1999): *Geostatistics and Petroleum Geology* 2ed. Springer Science+Business Media Dordrecht, The Netherlands 263p.
- IAEA. (2016): Global Network of Isotopes in Precipitation. The GNIP Database, <http://www.isohis.iaea.org>; last accessed on 12.12.2015
- KERN, Z., KOHÁN, B., CZINDER, A., CSONKA, D. & LEUENBERGER, M. (2015): Variogram analysis of precipitation $\delta^{18}\text{O}$ in the Great Alpine region - implications for isoscape modeling. *International Symposium on Isotope Hydrology, BOOK OF EXTENDED SYNOPSES*, pp.99-102.
- KERN, Z., KOHÁN, B. & LEUENBERGER, M. (2014): Precipitation isoscape of high reliefs: interpolation scheme designed and tested for monthly resolved precipitation oxygen isotope records of an Alpine domain, *Atmos. Chem. Phys.*, 14, pp. 1897–1907.
- KOHÁN, B. & SZALAI, J. (2014): Spatial analysis of groundwater level monitoring network in the Danube–Tisza Interfluve using semivariograms. *Hun. Geo. Bull.*, 63(4) 379–400.
- de OLIVEIRA, A.C.V. & da LIMA, A.S. (2010): Spatial variability in the stable isotopes of modern precipitation in the northwest of Iberia. *Isotopes in Environmental and Health Studies* 46, 13-26.
- RODRIGUEZ-ARÉVALO, J., DIAZ-TEIJEIRO, M.F. & CASTANO, S. (2013): Modelling and mapping oxygen-18 isotope composition of precipitation in Spain for hydrologic and climatic applications. In: *Isotopes in Hydrology, Marine Ecosystems and Climate Change Studies. Vol. I. Proceedings of an International Symposium (IAEA-CN-186/214)*, 171-177.

Trends in contaminant concentration time series

Köhler Artúr

Eötvös Loránd University of Sciences, kohler@ftr2000.hu

FTR 2000 Ltd

The groundwater at many industrial sites is often contaminated by various chemicals. The affected groundwater regime is a complex and dynamic system by itself, which interacts with the released chemicals increasing the complexity of the system. The actual dissolved concentration of the contaminants is dependent on far too many factors to be possible to account individually for each of them. Measurable concentration changes are sum results of these factors, therefore trend evaluation of concentration changes provide a tool that can support decision making process. However, trends in concentration time series are often difficult to evaluate. Different time series evaluation methods can lead to different results, and as new data is added to a time series the resulting trends can change using the same method. The two most common trend evaluation method groups are regressions using least squares and Mann-Kendall statistic. Both groups contain several slightly different procedures, like the most commonly applied linear regression and the seasonal Mann-Kendall. Linear regression and Mann-Kendall statistics applied on the same concentration time series data can result in different trends, increasing or decreasing trends in concentration time series may not be prevailing. It is not sufficient to evaluate trends solely at a certain point of the data collection, rather series of evaluation procedures are required, backwards in time if necessary, and as new data is added to the time series, the evaluation procedure should be repeated. The proposed evaluation of a time series with n data point ($n \geq 4$) should include $n-3$ evaluation procedures for 4, 5, 6 ... n data points. The results also form a time series. Finally it is proposed, that at some cases cumulative time average is also a suitable approach. Results indicate, that one-time trend evaluation may to provide inconclusive answers, trend persistence is a more important aspect of the time series.

Key words: *concentration, contamination, groundwater, time series*

1. INTRODUCTION

Chemical releases damage basic resources like groundwater, one of our main resources for drinking water. Once released, chemicals can reach the groundwater, they can move unnoticed in dissolved phase to water wells far from the point of release. While in transport they react with the constituents of the groundwater and the aquifer matrix, transforming by themselves and changing original water chemistry and sometimes the matrix also. These are summarized as “fate and transport” processes (Nyer & Gearhart., 1997).

Groundwater regime is a complex and dynamic system. The released chemicals superimpose on this system to form an anthropogenic anomaly called the contaminant plume. The resulting feature is similar to solid mineral resources, however the spatial scale is smaller, but it differs from most geologic phenomena because of its time-scale behaviour.

Fate and transport processes may be modelled by sophisticated hydrodynamic and transport models, however these models must necessarily be simplified. These models are validated by water chemistry data collected from various locations within and outside the plume. Groundwater chemistry data are the sum result of all ongoing fate and transport processes. Plume dynamics can be assessed by these data, and trend analysis is one of the means to understand the behaviour of a contaminant plume.

2. BACKGROUND

2.1. Site History

The studied site was used from the early 1960's to manufacture civil telecommunication devices (radios, etc.). In the mid-late 70's the facility was transformed to produce military-used electronic devices. Documented chemical usage indicated the use of tetrachloro-ethene (perchloro-ethene, tetrachloro-ethylene, PCE) between 1984 and 2001. Groundwater data indicated the use of Freon113 (1,1,2 trifluoro-trichloro-ethane) sometimes in the past. The groundwater contamination was detected in 1998.

The plume extent was assessed between 2002 and 2004. Monitored natural attenuation (MNA) was proposed, government administration conceived that

active remedial action required. Subsequent negotiations lead to the conclusion, that MNA is adequate, if there is no increasing trend in the concentrations. Trend in concentration time series is evaluated since.

2.2. Site Geology and Hydrogeology

The geological and hydrogeological factors provide the context of the plume. The subsurface of the site has a relatively simple geology and hydrogeology. The topography of the site is essentially plain, there is one shallow confined aquifer between 2 and 6 m below ground surface underlain by a minimum 30 m thick aquiclude. Groundwater flow is uniform, the direction is constant and the hydraulic potential surface has a low gradient.

3. DATA

3.1. Data description

Groundwater contamination data from the site is available from November 2000. Regular (quarterly) sampling started in December 2005. **Table 1** summarizes the sampling history:

Table 1: Sampling history

Time period	Number of wells				Total number of wells
	irregular	quarterly	yearly	not sampled	
2000.11.09-2001.05.08	3	0	0	0	3
2003.05.08-2004.01.21	0	10	0	0	10
2005.12.08-2007.12.07	0	13	1	1	15
2008.02.01-2009.12.10	0	15	1	1	17
2010.03.22-2012.09.18	0	19	1	1	21
2012.12.13-2013.09.25	0	23	1	1	25
2013.12.19-recent	0	24	1	1	26

All samples were analysed for halogenated aliphatic hydrocarbons. Between 2000 and April 2009 the components were specified by the 10/2000 Ministry decree following that by the 6/2009 Ministry decree (24 components). 1,1,1 Trichloro-ethane (1,1,1TCA) was added to the component list in May 2008. In March 2011 Hexachloro-butadiene was also added to the analytical list in order to comply with regarding regulations.

3.2. Data preparation

3.2.1. Theoretical background

Groundwater contamination datasets consist exclusively of a distinctive type of data: compositional data. The attention for distinctiveness of compositional data occurs as early as in Pearson 1897. Compositional datasets have properties that hinder or block straightforward application of even simple statistical procedures, like calculation of the average concentration. Typical datasets for instance usually contain missing values.

Missing values are classified into 4 categories (Boogaart et al., 2006) (1) rounded zeros (below detection limit) (2) structural (true) zeros, (3) data missing at Random (MAR) and (4) data Not Missing at Random (NMAR). Regarding literature mostly deals with MAR and NMAR type missing data. MAR and NMAR values differ from rounded zeros, because rounded zeros carry information about the missing value, namely an upper limit (the detection limit) and zero as a lower limit (negative concentration is impossible).

The percentage of missing values is mainly concern for MAR and NMAR data. Schafer recommended 5% cut off limit (Schafer, 1999), later Bennett allowed 10% (Bennett, 2001). Peng used 20% (Peng et al., 2006), Schlomer suggested a different approach, where the percentage of missing data is not definitive, rather the effect of the missing data should be taken into consideration (Schlomer et al., 2010).

Nondetect values common and frequent actors of compositional datasets. Their presence is especially self-explanatory when the dataset originates from a groundwater contamination problem. Wells around the plume and close to the perimeter of the plume yield almost exclusively nondetect data. As for time series in a case of diminishing or exhausting phase plume as time passes the number of nondetect data increases.

3.2.2. Methods for dealing with nondetects

The methods for dealing with nondetects can basically grouped into two categories, conventional and advanced methods. Conventional methods include category grouping, case deletion and simple imputation (Soley-Bori, 2013).

Advanced methods overcome the problem via highly sophisticated imputation methods (Soley-Bori, 2013).

3.2.3. Applied practice

For this study conventional methods were applied. Components that were detected in less than 5% of all the samples were excluded (17 components). One additional component was excluded on the basis, that it never had a value over regarding contamination limit. Five monitoring wells were excluded from trend analysis having 90% nondetects for all data. Time series with less than 5 data above nondetect were also excluded. For the remaining 99 time series, the nondetects were imputed with 65% value of detection limit (Martin-Fernandez et al., 2003).

4. METHODS

From purely theoretical point of view a minimum of three consecutive data required to facilitate trend evaluation (Hyndman & Kostenko, 2007). For this study trend evaluation was completed first place for the initial four data using linear regression (LR). For consistency reasons LR was used on as data accumulated, even when some of the time series exhibited exponential decrease. Mann-Kendall statistics (M-K) (Mann, 1945; Kendall, 1975) was drawn into in order to support LR results. Trend evaluation was repeated every time a new data augmented to the time series, for each time series with “n” ($n \geq 4$) data $n-3$ calculations were performed. Trend significance was evaluated at 90%, 92,5% and 95% levels. M-K statistics for $n \leq 10$ was not evaluated for 92,5 %.

4.1. Linear regression

Linear regression was done using MS Excel LIN.ILL function. Excel calculates (beside other statistical parameters) slope, degree of freedom, and F-statistic.

4.2. Mann-Kendall test

Mann-Kendall tests were performed similarly to linear regression calculations (i.e. for $n-(n-3)$, $n-(n-4)$... $n-3$ data). For $n \leq 10$ significance was calculated using the probability table from Gilbert For $n > 10$ Z-statistics is used to determine significance (Gilbert, 1978).

5. RESULTS

Concentration time series were plotted with $\pm 10\%$ laboratory uncertainty, and cumulative average concentration. For LR the $n-3$ set of results, and F-test critical values were plotted against time, a horizontal line added representing zero slope, started with the date of the fourth data. Mann-Kendall test calculations produced two plots for $n \leq 10$ starting with the date of the fourth data and $n > 10$, starting with the data of the 11th data. Examples of output plots are on **Figure 1**.

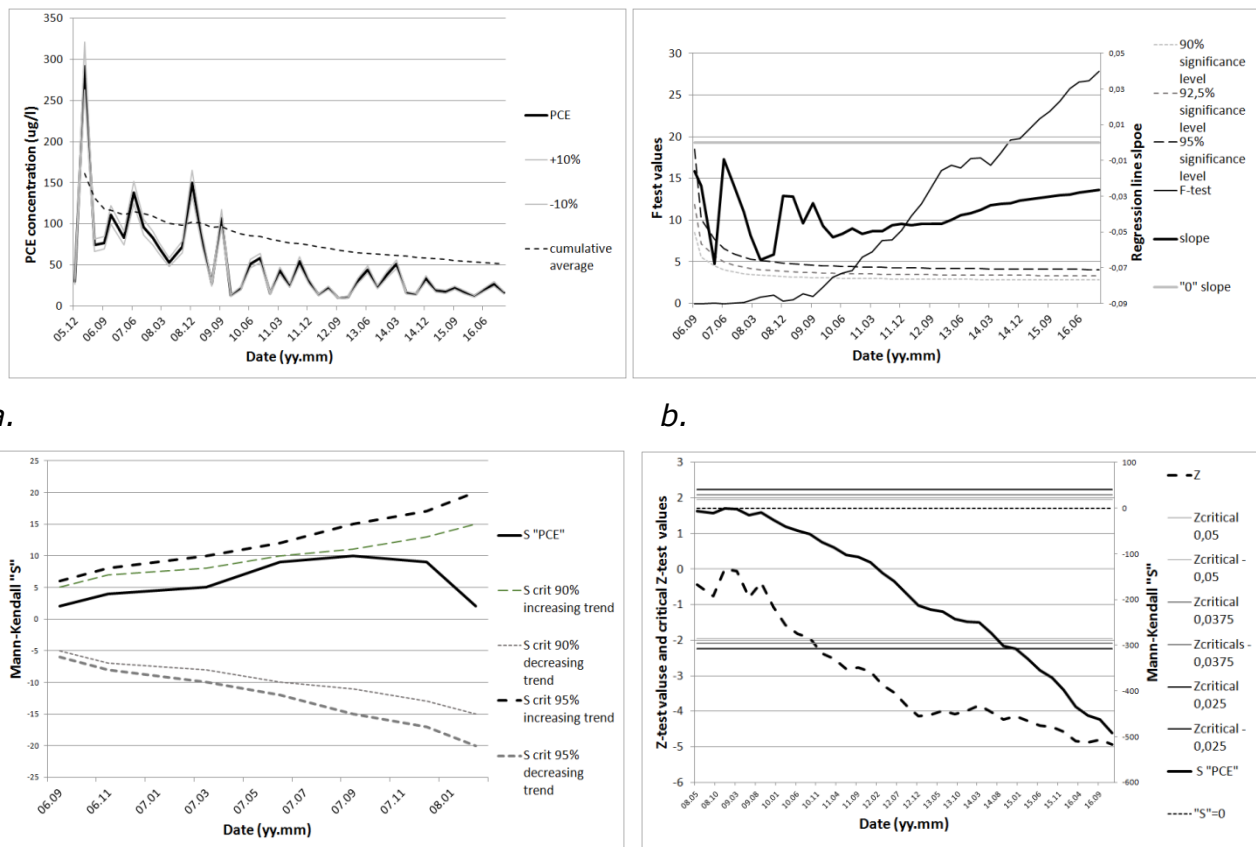


Figure 1: Graphical output of trend analyses. a.: concentration time series with $\pm 10\%$ laboratory uncertainty and cumulative average concentrations. b.: LR slope, F-test and critical F-levels with zero slope line added. c.: Mann-Kendall test and significance for $n \leq 10$. d.: Mann-Kendall test and Z-test for $n > 10$, Z test significance levels, and "0" S value line added.

The majority of the evaluations resulted in decreasing trends or no trends (less than 90% probability trends). Increasing trends were associated mainly with Freon 113. The persistence of significant trends varied, in some of the time

series a 95% significant decreasing trend turned later into a 95% increasing trend. 19 cases out of the 99 simultaneous LR and M-K calculations, the two methods produced inconsistent results, i.e. where LR indicated 95% significance trend, M-K indicated no trend or vice versa. LR proved to be not applicable when the first n data were uniform (imputed nondetect values). LIN.ILL function produces very small (10^{-35}) slopes for data at $n=10$ and $n=13$, associated with an extreme high F-test value for these cases. Cumulative time average seems to be insensitive and robust alternative, which at the same time refers to long-term exposition levels to calculate human health risks.

6. DISCUSSION

Contaminant concentration time series often evaluated for trend. The results obtained in this calculation is similar to the results published in affordable documents (HydroGeoLogic, Inc. 2004), in a sense that at the same contaminant plume monitoring wells with increasing and decreasing trends exist at the same time, together with wells that show no significant trend. Generally applied trend evaluation procedures usually use one selected method and do not consider trend persistence. They are usually restricted to evaluate trends at a certain time on all available or selected data. The calculations presented here indicate, that trends may not be prevalent, and that different methods could lead to different results applied on the same dataset.

7. CONCLUSION

The monitoring of groundwater contamination plumes may produce large datasets that are evaluated from various points of view. Trend detection is an often emerging task, and results can lead to decisions associated with financial consequences. Applied practice may not always ensure sound basis for these decisions. Trend persistence is the question that is not asked, and there is no readily available statistical procedure to use for trend persistence evaluation. It is proposed that evaluation of trend persistence is to be assessed, like calculating the probability of a value that would change our trend perception based on the actual dataset.

REFERENCES

- BENNETT, D. A. (2001). How can I deal with missing data in my study? Australian and New Zealand Journal of Public Health, 25, 464–469.
- VAN DEN BOOGAART, K.G., TOLOSANA-DELGADO, R. & BREN, M. (2006): Concepts for handling of zeros and missing values in compositional data. Int. Assoc. for Mathematical Geology XIth International Congress.
- GILBERT, R.O. (1987): Statistical Methods for Environmental Pollution Monitoring, 208-217., table A18.
- HYDROGEOLOGIC, INC (2004): OU-1 2004 Annual Groundwater Monitoring Report – Former Fort Ord, California.
- HYNDMAN, R.J. & KOSTENKO, A.V. (2007): Minimum sample size requirements for seasonal forecasting models. FORESIGHT Issue 6 Spring.
- KENDALL, M.G. (1975): Rank Correlation Methods, 4th edition. Charles Griffin, London, 202-231.
- MANN, H.B. (1945): Non-parametric tests against trend. Econometrica 13,163-171.
- MARTIN-FERNANDEZ, J. A., BARCELO-VIDAL, C. & PAWLOWSKY-GLAHN, V. (2003): Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. Math. Geol. Vol 35 No. 3.
- NYER, E.K. & GEARHART, M.J. (1997): Plumes Don't Move. Groundwater Monitoring and Remediation, Winter 1997, 52-55.
- PEARSON, K., (1897): Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs: London, Proceedings of the Royal Society, v. 60, p. 489-498.
- PENG, C.-Y. J., HARWELL, M., LIOU, S.-M. & EHMAN, L. H. (2006): Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), Real data analysis (pp. 31–78). Greenwich, CT: Information Age.
- SCHAFER, J. L. (1999). Multiple imputation: A primer. Statistical Methods in Medical Research, 8, 3–15.
- SOLEY-BORI, M. (2013): Dealing with missing data: Key assumptions and methods for applied analysis. PM931 Directed Study in Health Policy and Management Technical Report No. 4, Boston University.

The geographical evolution of urban greenness in the city of Erbil based on Landsat imagery

Shwan Hussein

University of Szeged, Physical Geography and Geoinformatics, H-6722, Egyetem u. 2-6, Hungary.

Email: shwan.huseen1@su.edu.krd

Worldwide, most cities have experienced major development in the last 20-25 years, in most cases resulting in a decrease in green areas, as shown by previous research. Therefore, this paper aims to assess the spatiotemporal variations of urban green areas in the 1990-2015 timeframe in the city of Erbil, one of Iraq's fastest developing cities, using remote sensing techniques. Given its vast spatiotemporal coverage, Landsat imagery was considered the best option for a remotely sensed data source.

Considering the medium spatial resolution of the Landsat data, we anticipated poor results regarding the traditional per-pixel classification, so it was decided to use sub-pixel classification, since recent studies have demonstrated this method's higher degree of accuracy in classification.

The study uses a combination of Linear Spectral Mixture Analysis and Fuzzy functions for the classification of Landsat imagery from 1990, 2000 and 2015 in order to extract four main classes of land use: agricultural land, vacant land, built-up land, and green vegetation. The same land-use classes were extracted using a traditional per-pixel classification algorithm, such as Maximum Likelihood Classification.

The accuracy of both methods was assessed by using a series of 50 trust points for each class, providing a confidence level of 95% and a confidence interval of +/-10% for both classification methods, resulting in an overall accuracy over 85%. Therefore, the results of the spatiotemporal evolution of urban greenness assessment in the Erbil urban area can be used both for spatial planning purposes and as an urban greenness assessment method in dry climate areas.

Key words: *urban greenness assessment, linear spectral mixture analysis, sub-pixel classification.*

1. INTRODUCTION

Cities worldwide have experienced a major development in the last 20-25 years. Since green areas within a city play a crucial role in the urban environment, knowing the spatial distribution of green areas within urban areas is crucial for decision-makers, and even more so for Erbil's local authorities, since, according to Erbil tourism (2014), green areas cover 12% of the city's built-up area, and the authorities would like to increase this to 15%, which is the minimum ratio provided for in International standards.

According to the United Nations (2014), global urban population has increased from 746 million in 1950 to 3.9 billion in 2014. Therefore, the urbanization process has experienced an acceleration the like of which has never been seen before, so the interest in the study of greenness within cities has undergone a directly proportional increase.

One of the first studies to use remote sensing on Landsat imagery to assess the urban greenness areas dates back to 1987 and it focused on defining the spatial patterns of urban vegetation and separate woody and herbaceous vegetation (Sadowski, Sturdevant, & Rowntree, 1987).

The 1990-2000 timeframe saw two other studies focusing on the use of remotely sensed data for urban greenness assessment, both dating from 1997. These studies used Landsat data to derive vegetation indices as one of the multiple variables used for the assessment of quality of life in urban areas.

Since 2000, the number of studies involving urban greenness has increased exponentially, reaching 124 in 2016. The majority of these studies tried to correlate the impact of urban greenness with different other urban variables, such as crime, health, life expectancy, or household income. However, the most relevant studies regarding the use of remotely sensed data are those of (Lu., Batistella., & Moran., 2002; Song, 2005; Chang & Ji, 2006; Mozaffar, Zoej, Sahebi, & Rezaei, 2008; Rhew, A, A, NL, & MD., 2011; Dawelbait & Morari, 2011; Gupta, Kumar, Pathan, & Sharma, 2012; Tang, Chen, & Schwartz, 2012; Wang, et al., 2013; Sakti & Tsuyuki, 2015). Most of the above-mentioned studies are focusing on sub-pixel classification techniques and end-member extraction.

2. MATERIALS AND METHODS

The study area for this research is the city of Erbil, the capital of the Iraqi Kurdistan Region it can be seen on (**Figure 1**), which has seen a big population growth in the last 11 years. According to Iraq City Population (2014), Erbil's population has increased from 494,194 in 1993 to 910,827, by 84%. As one of the fastest developing cities in Iraq, the city of Erbil is perfect for a spatiotemporal assessment of urban greenness, if we consider the fact that the city is located in a region characterized by a hot-summer Mediterranean climate, therefore, urban greenness plays an important role.



Figure 1: Study area, Erbil City, covered by one Landsat scene in the patch/row 169/36

For the spatiotemporal assessment of urban greenness, a set of three Landsat scenes was used. In order to avoid seasonally-derived errors, it was decided to use scenes from the same season, dating from 10th June 1990, 14th of August 2000 and 8th of August 2015. A series of corrections has been made, such as orthorectification and the conversion of the digital numbers to normalized at sensor reflectance, all done using the method described by Markham & Barker (1987).

The methodology involves a three steps framework: Step I, that implies the use of Linear Spectral Mixture Analysis (LSMA) and Maximum Likelihood Classification (MLC) for the classification of the three Landsat scenes; Step II, where the results from the LSMA are fuzzyfied in order to extract land use classes and then, both LSMA and MLC results undergo a classification accuracy assessment, which must be greater than 80% in order for the classification to be used in Step III, where the green vegetation change magnitude will be determined as shown in the flowchart (**Figure 2**).

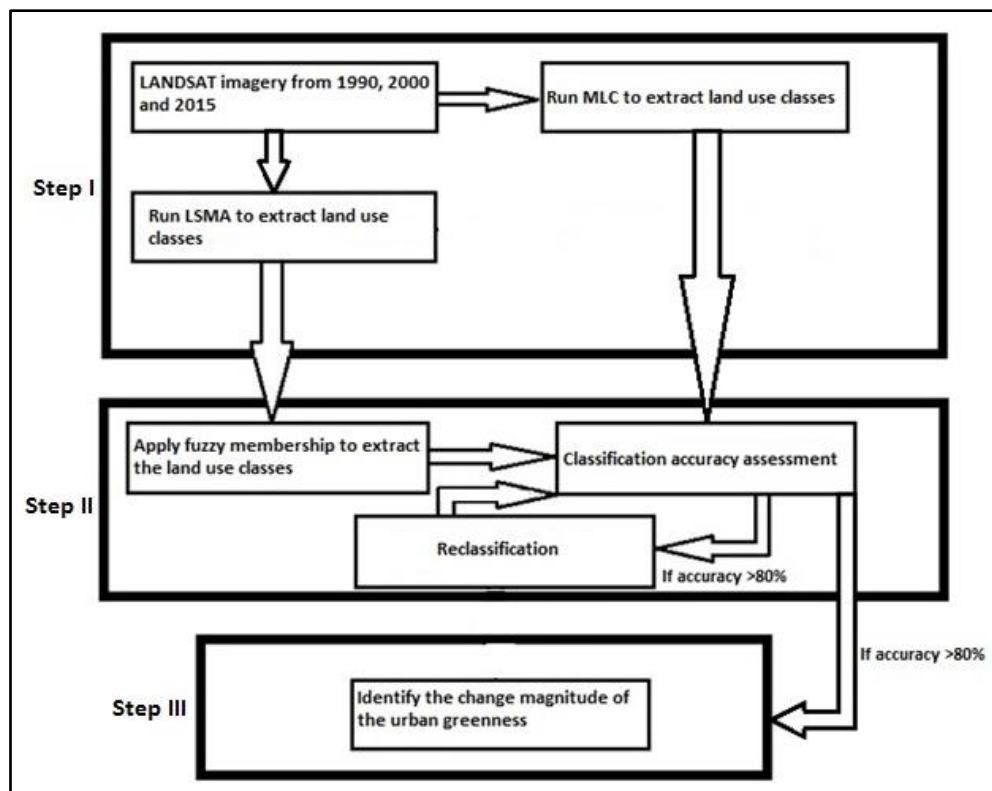


Figure 2: Work methodology flowchart

3. RESULTS AND DISCUSSIONS

The evolution of the vegetation cover within cities is crucial to urban planners and local authorities, since the presence of relatively large areas of green space can significantly improve the quality of life in urban areas (Ambrey & Fleming, 2013).

Both classification approaches used in our research produced good results, obtaining an overall accuracy >80%. The LSMA classification produced more accurate results, since all three classifications had an accuracy >92% (92.08%

in 1990, 95.12% in 2000 and 92.28% for the 2015 image). Meanwhile, the MLC approach was inconsistent, dropping from a 95.81% accuracy in 1990 to an 80.78% accuracy for the 2000 image.

Land use within the city of Erbil has undergone a series of major changes, especially between 2000 and 2015, when the city’s built-up land area increased by 229%, according to the LSMA classification and here it seems something is missing – what increased by 225% (**Figure 3**) according to the MLC classification. The expansion of the built-up area has led to the disappearance of agricultural land within the city limits. Unlike other cases, where green vegetation areas have decreased as a result of urbanization, in our case we can clearly see a direct link between the urbanization process and a 308% increase in green vegetation.

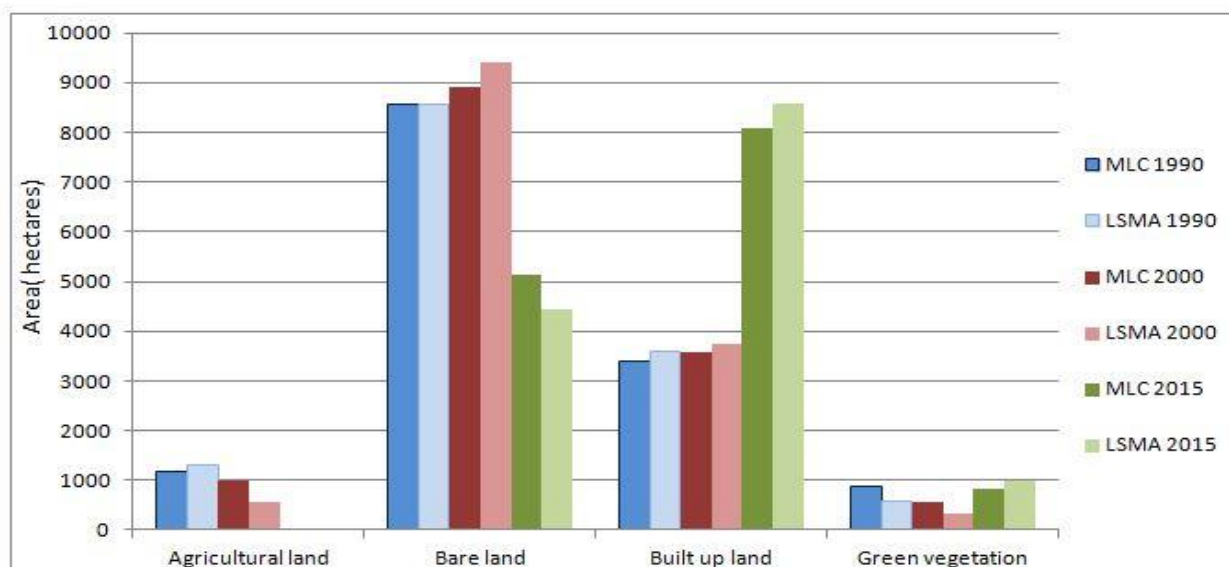


Figure 3: 1990-2015 land use evolution in the city of Erbil

The increase of area covered by green vegetation can be explained by the construction of green areas, such as the Sami Abdulrahman Park, built in 2006 and covering 200 hectares. Therefore, the construction of the Sami Abdulrahman Park alone accounts for about half of the increase in green vegetation area. The Sami Abdulrahman Park was accurately identified by both classification methods, representing the largest and most compact area covered by green vegetation, as it can be seen in **Figure 4 c** and **f**.

The spatiotemporal analysis of the green vegetation distribution reveals a continuous migration of the green vegetation clusters, from north-west in 1990 to south-east in 2000, then back to north-west again.

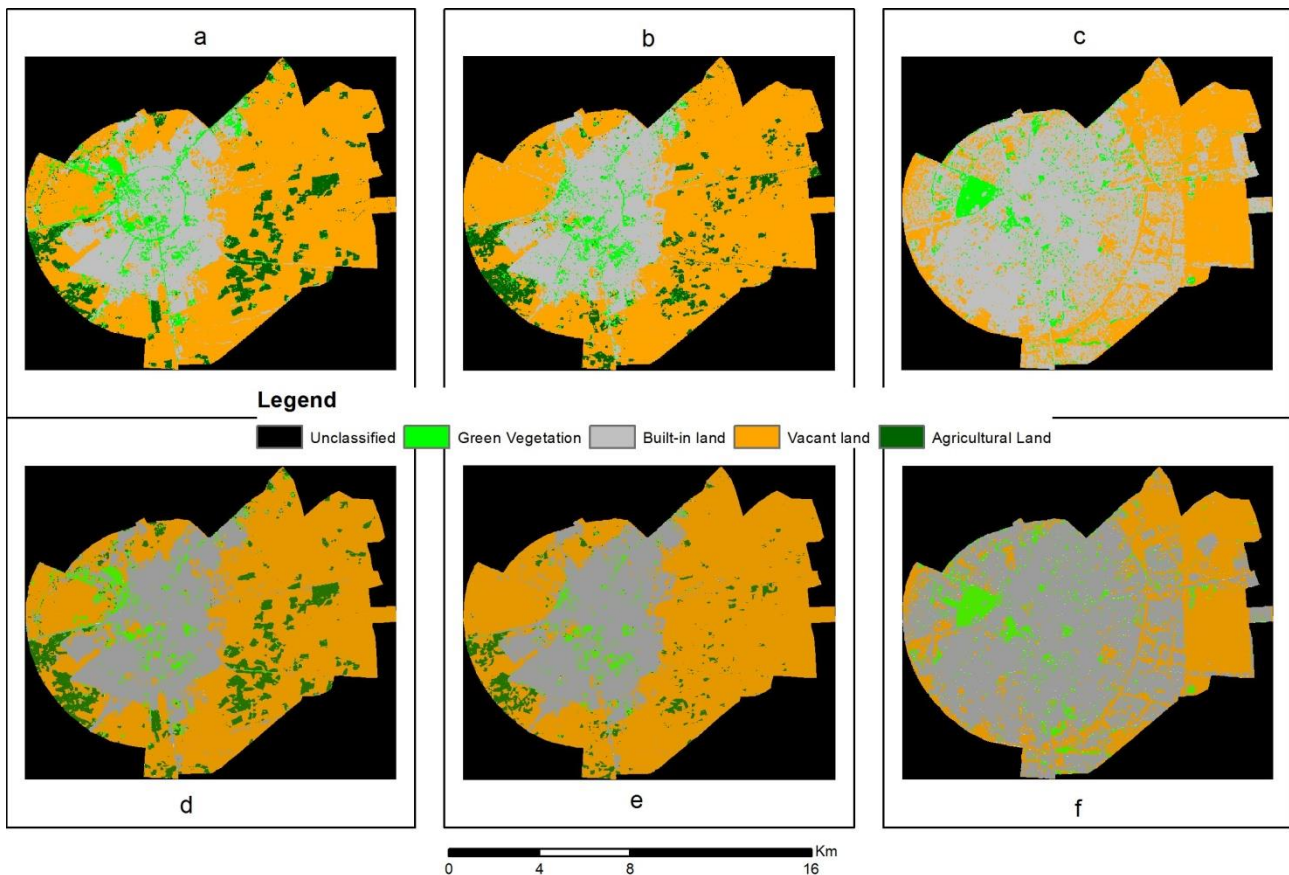


Figure 4: Landsat imagery classification results (a-MLC 1990, b-MLC 2000, c-MLC 2015, d-LSMA 1990, e-LSMA 2000, f-LSMA 2015)

4. CONCLUSIONS

This research provides critical information regarding the spatiotemporal evolution of the urban greenness in the Erbil city area over a 25 year timeframe (1990-2015), highlighting the increase in urban greenness by approx. 308%. The accuracy of our results proves the usefulness of the LSMA in extracting end-members from satellite imagery in order to provide an accurate quantitative analysis of land use change.

The research also proves the usefulness of medium spatial resolution satellite imagery, such as Landsat, to assess the geographical evolution of urban greenness. Therefore, similar studies can be carried out to assess the urban greenness evolution at the national or regional scale in order to help decision

makers identify cities that have experienced a loss in urban greenness as a result of urban development and don't reach the minimum ratio of urban greenness required by international standards.

REFERENCES

AMBREY, C. L. & FLEMING, C. M. (2013): Public Greenspace and Life Satisfaction in Urban Australia. *Urban Studies*, pp. 1290 - 1321.

CENTRAL STATISTICAL ORGANIZATION-IRAQ (2014): Population Statistics. [Online] Available at: <https://cosit.gov.iq> Accessed 28 12 2016].

CHANG, C.-I. & JI, B. (2006): Fisher's Linear Spectral Mixture Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, pp. 2292-2304.

DAWELBAIT, M. A. & MORARI, F. (2011): LANDSAT, Spectral Mixture Analysis and Change Vector Analysis to Monitor Land Cover Degradation in a Savanna Region in Sudan (1987-1999-2008). *International Journal of Water Resources and Arid Environments*, pp. 366-377.

ERBIL TOURISM (2014). Erbil Seeking Access To International Standards of Green Areas. [Online] Available at: <http://www.erbiltourism2014.com/en/erbil/news/113-erbil-seeking-access-to-international-standards-of-green-areas> Accessed 20.12.2016

GUPTA, K., KUMAR, P., PATHAN, S. K. & SHARMA, K. P. (2012): Urban Neighborhood Green Index – A measure of green spaces in urban areas. *Landscape and Urban Planning*, pp. 325-335.

LU., D., BATISTELLA., M. & MORAN., E. (2002): Linear Spectral Mixture Analysis Of TM Data For Land-use And Land Cover Classification In Rondonia, Brazil Amazon. Ottawa, Department of Natural Resources Canada.

MARKHAM, B. L. & BARKER, J. L. (1987): Thematic mapper bandpass solar exoatmospheric irradiances. *International Journal of Remote Sensing*, pp. 517-523.

MOZAFFAR, M. H., ZOEJ, M. V., SAHEBI, M. & REZAEI, Y. (2008): Vegetation Endmember Extraction in Hyperion Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 37,409-412.

RHEW, I DR. (2011): Validation of the Normalized Difference Vegetation Index as a Measure of Neighborhood Greenness. *Annals of Epidemiology, Annals of Epidemiology*.

SADOWSKI, F. G., STURDEVANT, J. A. & ROWNTREE, R. A. (1987): Testing the consistency for mapping urban vegetation with high-altitude aerial photographs and landsat MSS data. *Remote Sensing of Environment*, 21 129-141.

SAKTI, A. D. & TSUYUKI, S. (2015): Spectral Mixture Analysis (SMA) of Landsat Imagery for Land Cover Change Study of Highly Degraded Peatland in Indonesia. *Berlin, an.*, pp. 495-501.

SONG, C. (2005): Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?. *Remote Sensing of Environment*, 95, 248-263.

TANG, J., CHEN, F. & SCHWARTZ, S. S. (2012): Assessing spatiotemporal variations of greenness in the Baltimore-Washington corridor area. *Landscape and Urban Planning*, pp. 296-306.

WANG, H.-F. I DR. (2013): Variations of urban greenness across urban structural units in Beijing, China. *Urban Forestry & Urban Greening*, 12, 554-561.

Combined Cluster and Discriminant Analysis, an efficient tool in taxonomical classification

Norbert Magyar^{1,2}, Bálint Polonkai³, József Kovács², Ágnes Görög³, Emese Réka Bodor⁴

¹Budapest Business School, Magyar.Norbert@uni-bge.hu

²Eötvös Loránd University, Department of Physical and Applied Geology

³Eötvös Loránd University, Department of Paleontology

⁴MFGI, Geological and Geophysical Collections

The classification of records into homogeneous groups is an important aim on many scientific fields. The main goal of this study is to test the applicability of a new data analysis method (Combined Cluster and Discriminant Analysis - CCDA) to the field of palaeontology. CCDA combines to well-known grouping methods, the hierarchical cluster analysis and linear discriminant analysis. During the application of CCDA preconceived groupings obtained from hierarchical cluster analysis with random groupings are compared. If the preconceived grouping is better than the random ones, the samples does not form homogeneous groups. CCDA provide an objective index number of the homogeneity. The method has been implemented into the R statistical software as a package called “ccda”.

The applicability of the CCDA was tested on well preserved irregular sea urchin fossils of the family Scutellidae Hungarian Middle Miocene (Badenian). Despite the large number of specimens there are some open question regarding the taxonomical classification of Scutellidae.

Three species (*Scutella hungarica*, *Scutella pygmea*, *Scutella vindobonensis*) were analysed based on morphological parameters. Based on the results it can be concluded, these species did not form homogeneous groups, there are significant differences between them. These results can provide useful informations to classify the species more precisely, and make the palaeoenvironmental and palaeobiogeographical reconstructions better.

Key words: *classification into homogeneous groups, Combined cluster and discriminant analysis, middle Miocene, Scutellidae.*

A numerical model of Na-montmorillonite validated by batch experiments

Zsuzsanna Szabó¹, Csaba Hegyfalvi², Ágnes Freiler¹, Beatrix Udvardi¹, Péter Kónya¹,
Edit Székely², György Falus¹

¹Department of Geochemistry and Laboratories, Geological and Geophysical Institute of Hungary,
Stefánia út 14., Budapest 1143, Hungary, szabo.zsuzsanna@mfgi.hu, zsszabo86@gmail.com

²Department of Chemical and Environmental Process Engineering, Budapest University of Technology
and Economics, Műegyetem rakpart 3., Budapest 1111, Hungary

One of the challenges of the present century is to limit the greenhouse gas emissions for the mitigation of climate change which is possible for example by a transitional technology, CCS (Carbon Capture and Storage). Clay minerals are considered to be responsible for the low permeability of CCS caprocks sealing off stored CO₂. For the better understanding of their reactivity, this work aims to create a kinetic geochemical model of Na-montmorillonite clay standard Swy-2 in the widely used PHREEQC code, supported by solution and mineral composition results from batch experiments. Several four-days batch experiments have been carried out at atmospheric conditions, and with scCO₂ at 100 bar and 80 °C. Solution samples have been taken during experiments and their compositions were measured by ICP-OES. The treated solid phase has been analysed by XRD and ATR-FTIR and compared to in-parallel measured references (dried Swy-2). Fitting the numerical modelling output on laboratory measurement results has been automatized by a code written in R programming language. Both experiments and models show fast reactions with an increased rate in the presence of scCO₂. A model sensitivity analysis has pointed out that the continuously changing solution composition results cannot be described by the change of the uncertain reactive surface area of mineral phases in the model and still several orders of magnitude different ion-concentrations are predicted. However, by considering the clay standard's cation exchange capacity divided proportionally among interlayer cations of Na-montmorillonite, the measured variation can be described on an order of magnitude level. It is furthermore indicated that not only the interlayer cations take part in this process but a minor proportion of other, structural ions as well, differently in the reference and scCO₂ environments. In this work, an improved model of clay behaviour in geological environments became available.

Key words: CO₂ geological storage, geochemical interactions, PHREEQC, R programming

Relationship between geotechnical parameters and discrete fracture network simulation results in Bábaapáti National Radioactive Waste Repository

Gábor Somodi¹, Krisztina Istovics², László Kovács¹, Tivadar M. Tóth²

¹Kőmérő Kft., somodigabor@komero.hu, kovacslaszlo@komero.hu

²University of Szeged, Department of Mineralogy, Geochemistry and Petrology

In rock engineering, the determination of geotechnical parameters of rock masses is a key issue for the prediction of expected mechanical behaviour of rock mass. The geotechnical parameters of a fractured rock mass are dominated by the spatial distribution of fracture sets and the mechanical behaviour of the discontinuities. The research area is situated in the host rock of Hungarian National Radioactive Waste Repository, which is a fractured granitic body. Previous field observations and model results suggested that the repository formation is hydraulically strongly compartmented, dividing the underground flow system into several blocks of limited hydraulic connection. Each block characterized by different fracture orientation also. It presumes that the geotechnical properties must have influenced by these patterns. Geotechnical documentation would provide data for hydraulic modelling and simulation result can be useful for prediction geotechnical behaviours but clear correlation between properties of the fracture system is needed.

In this paper we continue a previous research of finding possible connection between the simulated DFN network and data of geotechnical documentation. Two boreholes were investigated which explored the area of the subsequent repository chambers. Several DFN model have created from BHTV data. Repsim software was used for simulations. Results show one communicating fracture network consists of fractures of EN to NS strike. This coincide the strike of the main transmissive zones and perpendicular to the dominant fracture orientation determined by the geotechnical mapping of the repository area. Based on the spatial distribution of fractal dimension (D) and the exponent of the fracture length (E) we can conclude that both parameters define the hydraulic connections.

Our work tends to provide results for helping further design and construction works of the waste repository.

Key words: *fractured rock, radioactive waste repository, DFN model, rock mass classification.*

Data acquisition, pre-processing of 3D image data of artificially distorted skulls, archeological artefacts, fossils for 3D geometric morphometric analysis using CT and laser scanning: a comparison

Sándor Gulyás ¹, László Nyúl, ² Gábor Németh², Csilla Balogh ³, Lilla Kardos ⁴, Pál Sümegi ¹

¹ University of Szeged, Department of Geology and Paleontology, Szeged, Hungary.
gulyas.sandor@geo.u-szeged.hu.

²University of Szeged, Department of Image Processing and Computer Graphics, Szeged, Hungary

³University of Istanbul, Research Institute of Turkology, Istanbul, Turkey

⁴Affidea Diagnostics Ltd., Budapest, Hungary

Novel techniques for capturing 3D image data used in statistical shape analysis are readily available at a reasonable price. In cases where shape differences are especially important along entire surfaces, 3D geometric morphometrics performs much better than combined 2D GM implemented on sections from various views. 3D GM has strong mathematical and computational backgrounds. Nevertheless, little detailed information is available on the foundation of the entire process: data acquisition and pre-processing. Nowadays two methods prevail in image data acquisition in 3D: computer tomography and laser surface scanning. CTs provide highly detailed spatial information of a given object at the sub-centimetre scale. Visualization is relatively easy. In addition, slices can be used for sectional GM along well-identified landmarks. Data analysis can be rather costly and interpretation of Hounsfield values might be problematic on taphonomically altered material. 3D laser scanning has similar hardships to optimize: number of slices set for scanning, surface reflectivity, object's distance to laser source, merging, segmentation algorithms for accurate fit. Processing times must be optimized but by no means at the cost of resolution. This work presents initial results of technical settings developed for data acquisition using both methods, computational algorithms for merging and segmentation and the assignment of landmarks of received dataset for 3D GM on artificially distorted skulls, archaeological artefacts and fossils. Our work is supported by OTKA Grant 109-510.

Key words: *geometric morphometrics, 3D data acquisition processing, surface and volume data, development of technical settings, segmentation techniques, merging algorithms*

Part VIII - Geomathematics in applied geosciences

Fault Statistics and Fractal Geometry

Zoltán Unger¹

¹ELTE-TTK-SEK –Szombathely, Károlyi Gáspár tér 4., unger.zoltan@freemail.hu

One of the aims of research activity in this field is to analyse the validity and applicability of fault-statistics calculations in different geological fields.

In mining for metal ores and coal this procedure is familiar, with its utilisation in areas such as following surface geomorphologic evolution, the interpretation of tectonic maps, hydrocarbon exploration and production, and lineament analysis.

After the theoretical introduction, we will deal with an array of applications of this research methodology, the implementation of these methods in the hydrocarbon industry and the development of new ones. The intent is to show the methods developed by case studies dealing with the relative position of the fractured and non-fractured blocks.

Following this, we examine the practical implementation in geology of this new branch of research based on fractal geometry. Afterwards, there follows a review of the results being achieved nowadays in the field of fault statistics calculation and then a summing-up of the calculations used to process the results from tectonic elements in the research areas. After the prediction of the possible number of faults based on their length, there follows the approximation of the power function and a solution to its calculation by iteration.

Finally, it can be shown mathematically that the exponents of the power functions are equal to the corresponding fractal dimensions.

Key words: *fault statistic, fractal geometry, power law function, exponent, iteration*

Image processing for fractal geometry-based Discrete Fracture Network model input data

Dorottya Kovács¹, Gergely Dabi², Tivadar M. Tóth³ and Balázs Vásárhelyi⁴

¹Ph.D. Student and ⁴Associate Professor at Budapest University of Technology and Economics,
Department of Engineering Geology and Geotechnics, kovacs.dorottya@epito.bme.hu

²Assistant Lecturer and ³Head of Department, Associate Professor at University of Szeged, Department
of Mineralogy, Geochemistry and Petrology

Investigation of the spatial distributions of microfracture networks in fractured rock bodies is essential for the accurate understanding of mechanical behaviour and fluid flow within the rock mass in case of designing underground facilities. Rock mechanics suggests that most mechanical parameters are related to the inherent microcrack system of the rock. In order to investigate the fracture network in a supposedly Excavated and/or Environmental Damage Zone (EDZ) fractal geometry-based Discrete Fracture Network (DFN) models were composed, which necessitated several parameters of the microcrack system, such as fractal dimension of the fracture midpoints (D_c), length exponent (E) and relative dip data. Since these data are very tedious and time consuming to be collected, the extraction of the necessary parameters should be automatized. With the intent of making data acquisition time efficient and automatized a new method was established. In order to enhance the visibility of fractures, drilled core samples were impregnated with tinted epoxy resin during preparation. The appropriate input parameters for the modelling were extracted from binary images of the studied microcrack systems. The images of the studied sections were pre-processed in the ImageJ based Fiji environment in advance for further image processing in MATLAB. The determination of the fractal dimension parameters was carried out with the use of the box-counting method. For the automation of data retrieval from the images, a series of MATLAB algorithms have been developed. The method allows the retrieval of more than one thousand fracture length data per cm^2 of one sample in several minutes. This methodology tends to be a useful tool in studies of fracture network geometries, DFN models are presented in this work as an example.

Key words: *Image processing, Fracture network, Fractal geometry, MATLAB*

1. INTRODUCTION

Retrieval of data relevant for the characterisation of fracture geometry is a tedious and time consuming task, but it is essential in Discrete Fracture Network (DFN) modelling. DFN modelling is a widely used tool in rock mechanics (Ling, 2003), and hydrogeological practice (Long, 1996), and thus underground designing and reservoir modelling related researches may use it beneficially. In this study a fractal geometry-based DFN model was developed, which needed the following input data:

- Fractal dimension for fracture centres (D_c);
- Parameters of the fracture length distribution (E, F);
- Lower and upper extreme fracture length values;
- Aperture, which must be given as a function of length with two input parameters (Vermilye and Scholz 1995), because it is not treated as an independent variable;
- Strike and dip data can be given as an explicit set of data pairs instead of a distribution function.

For the automatization of data retrieval, a sequence of image processing algorithms was developed, which is introduced in this paper.

2. SAMPLES

Specimens used in this study were extracted from the floor of an investigation tunnel, which was constructed by using the careful Drill & Blast (D&B) technique. The specimens were prepared to 68 mm diameter and 50 mm length cylinders. The samples used in this study were very coarse grained leucocratic granites (LGR) and banded gneisses (BGN).

3. METHODS

3.1. Sample preparation and image creation

5 mm thick disks were cut from the cylindrical specimens, and their surfaces were polished after impregnation with UV fluorescent tinted epoxy. This made direct observation of the crack-geometries possible in refracted UV illumination (**Figure 1a**). Images of the discs were taken with an Olympus DP73 camera

mounted on an Olympus BX41 microscope with 1.25x magnification. The microscope was used in reflected-light mode with a 100 watt U-LH100HG mercury vapour lamp as light source and an Olympus U-MWBV2 filter cube installed (excitation range between 400-440 nm).

3.2. Image processing

3.2.1. Segmentation

First step of the image processing is the segmentation of the UV fluorescence images, i.e. the division of the pixels of the images into microcracks and host rock. For this, the original RGB images were processed by Trainable Weka Segmentation plug-in, which is available in the ImageJ based Fiji environment (Schindelin et al., 2012). The Trainable Weka Segmentation plug-in enables the user to train the software to distinguish and classify segments of the image by drawing lines and marking out areas, thereby collecting information on pixels' grey level and their spatial variation.

3.2.2. Morphological processing

The segmentation step results 8-bit grayscale images (**Figure 1b**). It is unavoidable to get misidentified image segments after the segmentation treatment. These objects are mainly fluorescent mineral grains with nearly isometric shape. Because of the obvious difference of the shape of fractures and that of the nonlinear objects these group of pixels can be easily filtered out in the following steps of the image processing in MATLAB (e.g. Arena et al., 2014). First, the 8-bit grayscale image (**Figure 1b**) is converted into a binary black-and-white image (**Figure 1c**). In the black-and-white image, white pixels represent fractures and voids, and black pixels represent the intact host rock. Next, noise and small objects (generally fluorescent mineral grains) are filtered from the data by selecting retaining the 1000 largest objects (i.e. largest groups of associated white pixels) on the image, which effectively removes isolated pixels or small pixel groups of pixels (**Figure 1d**). The morphological processing begins with dilatation (**Figure 1e**), which thickens the objects and removes

interior irregularities (i.e. holes inside pixel groups representing microfractures) and those along their boundaries (i.e. embayments).

Originally continuous cracks can appear as a sequence of individual parallel microfractures in line, which can reunite during dilation. However, originally distinct microcracks can also coagulate during dilation, which are separated in the next step, based on their relative angles with the reference direction of the image (see later). Dilatation is necessary, because internal irregularities in the image yield additional spurious microcracks during thinning in the final step of the process. The dilatation should be repeated as many times as needed. This should be carefully controlled by the user and should be kept close to the original image. After achieving the ideal connectivity of the objects (as defined by the user by checking the image between subsequent dilation steps), a filtering is carried out based on the objects' Euler Number (see MATLAB's Image Processing Toolbox for details). The Euler Number specifies the number of objects in a region minus the number of holes in those objects. This step extracts circular objects, which are generally inherited from planar objects, like UV fluorescent minerals. Removing objects with negative Euler Numbers, i.e. objects that have internal holes, braided crack systems can also be removed from the image, thus this step needs to be carried out carefully. Finally, the remaining objects should be thinned until they are one pixel wide. This results a skeletonized image of the fracture network (**Figure 1f**).

3.2.3. Data retrieval

To extract geometric parameters, the fractures need to be divided into single cracks. To do this, a MATLAB algorithm, which detects branchpoints from the skeletonized image was developed. The algorithm calculates the angle between the cracks emanating from the branchpoint and a reference line (the horizontal line of the studied image). It is assumed that the fractures with similar angles are part of the same crack, and the pixel that connects the third fracture to the branchpoint is removed.

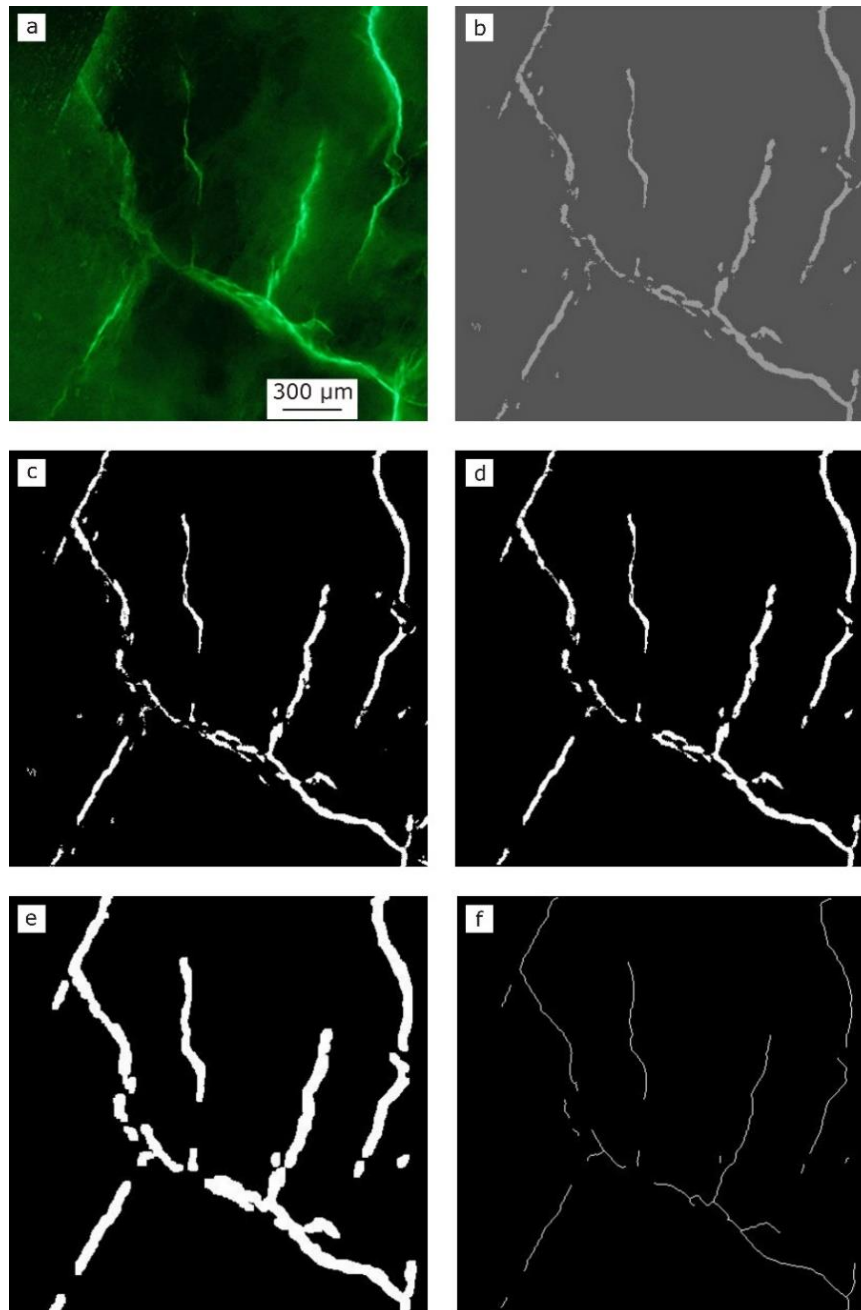


Figure 1: The performed image processing steps are detailed in the text. (a-crack geometry in refracted UV illumination; b- grey-scale image; c-binary image; d-image after filtration; e-crack geometry after dilatation; f- skeletonized image of the fracture network)

By defining the branchpoint pixels, the continuous fracture network is divided into single cracks, increasing the total number of cracks, and decreasing their length. As the final step of the image processing, another MATLAB algorithm pairs fractures' end point coordinates and calculates their lengths and centre coordinates.

Box counting measurements (Mandelbrot, 1983; Barton, 1995) were carried out with Benoit 1.3 software from the skeletonized images that were processed in MATLAB.

4. FRACTURE NETWORK MODELLING

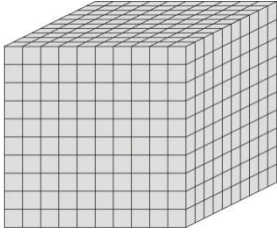
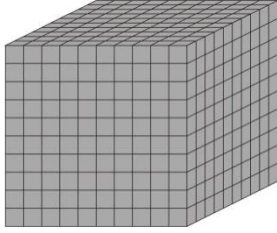
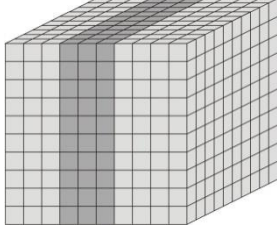
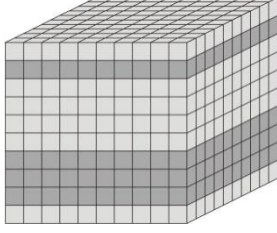
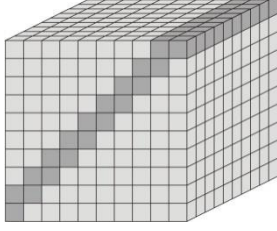
Based on the above geometric parameters fracture networks have been simulated in 3D using the fractal geometry-based DFN algorithm, RepSim code, which was developed by M. Tóth et al. (2004).

RepSim is stochastic in the sense that it obtains a certain number of equally likely realizations. The final output form of the fracture network corresponds to the initial expectations, that the parameters of this network are equal to the originally studied fracture network's parameter-distributions.

The RepSim fracture network models were built to represent 10 m³ volumes, the constituent cells of which are 1 m³ cubes. For the length distribution function, an interval of the fracture lengths of interest needs to be given by the user, which was between 50 mm - 1500 mm in this work. Each configuration was run five times, thus five equally likely realizations were generated based on the measured parameters.

In case if the whole 10 m³ rock volume was composed of gneiss (LGN) the simulations yielded 181 000 fractures in the model (**Table 1.**). According to these simulations, the highest number of communicating fracture clusters was 17, while the highest number of communicating fractures in one cluster was 109. Five simulations of the model with the whole rock volume built of granite (BGR) show that the highest number of communicating fracture clusters was 52, while the highest number of communicating fractures in one cluster was 583 out of 231 000 fractures (**Table 1.**). The fractures are oriented in two preferred orientations and there are relatively longer fractures comparing them to the ones in the gneiss. The models that consisted of both rock types vary between the values characteristic for the studied gneiss and granite, depending on the quantity of each component.

Table 1: Summary of the built models with their five realizations. The light grey color represents the leucocratic granite ($E=2.05$ and $F=16.64$) and the dark grey represents the banded gneiss ($E=2.23$ and $F=18.2$) in the modeled volume.

	Number of clusters, that are larger than 50 pieces	Number of fractures in the largest cluster
	9	84
	14	100
	17	101
	7	109
	10	74
	Number of clusters, that are larger than 100 pieces	Number of fractures in the largest cluster
	47	498
	52	583
	47	310
	52	256
	43	393
	Number of clusters, that are larger than 100 pieces	Number of fractures in the largest cluster
	16	208
	12	150
	13	220
	18	225
	11	203
	Number of clusters, that are larger than 100 pieces	Number of fractures in the largest cluster
	15	252
	15	179
	14	204
	20	193
	19	217
	Number of clusters, that are larger than 100 pieces	Number of fractures in the largest cluster
	11	215
	11	194
	11	201
	7	213
	10	237

5. CONCLUSIONS

The data acquisition method that has been developed for defining the geometrical parameters of fracture systems is a convenient tool for determination of the variables required for Discrete Fracture Network modelling. According to the simulated fracture network models, the largest percolation cluster sizes are between 0.06 % (gneiss) and 0.25 % (granite) of all fracture systems. According to the very low number of interconnected fractures, the hydrological communication between fractures is constantly negligible.

REFERENCES

- ARENA, A., DELLE PIANE, C. & SAROUT, J. (2014): A new computational approach to cracks quantification from 2D image analysis: Application to micro-cracks description in rocks. *Computers & Geosciences* 66, 106–120.
- BARTON, C.C. (1995): Fractal analysis of scaling and spatial clustering of fractures. -In: BARTON C.C., LA POINTE P.R. (eds.): *Fractals in the Earth sciences*. Plenum Press, New York, 168 p.
- JING, L. (2003): A review of techniques, advances and outstanding issues in numerical modelling for rock mechanics and rock engineering. *International Journal of Rock Mechanics & Mining Sciences* 40, 283–353.
- LONG, J.C.S (Ed.) (1996): *Rock fractures and fluid flow: contemporary understanding and applications*. National Academy Press, Washington D.C., 551 p.
- M. TÓTH, T., SZÚCS, É., SCHUBERT, F. & HOLLÓS, CS. (2004): Conceptual fracture network model of the crystalline basement of the Szeghalom Dome (Pannonian Basin, SE Hungary). *Acta Geologica Hungarica* 47, 1, 19–34.
- MANDELBROT B.B. (1983): *The fractal geometry of nature*. Freeman, New York, 468 p.
- MIN, K.B., JING, L. & STEPHANSSON, O. (2004): Determining the equivalent permeability tensor for fractured rock masses using a stochastic REV approach: Method and application to the field data from Sellafield. *Hydrogeology Journal* 12, 5, 497–510.
- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J.-Y., WHITE, D.J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. (2012): Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9, 676–682.
- VERMILYE, J. M. & SCHOLZ, C. H. (1995): Relation between vein length and aperture. *Journal of Structural Geology* 17, 423–434.
- YIELDING, G., WALSH, J.J. & WATTERSON, J. (1992): The prediction of small-scale faulting in reservoirs. *First Break* 10, 449–460.

Quantitative Determination of Standard Sphericity and Roundness

Tamás Jaskó¹

¹16 Melrose Place, Watford WD17 4LN, England, tamas.jasko@jasko.info

A new method is proposed to measure API standard sphericity and roundness by image analysis. The method is both fast and practical and the results are more precise than those obtained by chart comparison.

Key words: *Sphericity, Roundness, Image processing, Clastics, Sand, Pebbles*

1. INTRODUCTION

Sphericity and roundness are important shape characteristics for clastic grains, sand, gravel etc. In particular, the suitability of sand used for fracking is defined in terms of sphericity and roundness. The agreed method is the API standard (American Petroleum Institute, 2011) which reproduces the visual comparison chart of 1951 by Krumbein & Sloss (Krumbein & Sloss, 1963). Szádeczky, Wadell and others proposed various other index numbers for the shape of grains (Szádeczky-Kardoss, 1933; Wadell, 1932). Some of these are precise but so time consuming as to make them unsuitable for routine use. The quicker methods are subjective and less precise (Schneiderhöhn, 1954; Folk, 1955).

2. PROPOSED METHODOLOGY

The limitations of processing time and complexity that made the precise methods impractical in the past apply less today. Modern instrumentation and computer methods allow fast evaluation and more precise results. The proposed method starts with digital photographs of sand and gravel size particles, from which an outline (perimeter or intercept) of the grain is extracted by image processing. Further computation gives the centroid, perimeter, and area of the particle, and the axes: axis a (long axis) and axis b (short/middle axis) (**Figure 1**). From

these a number of derived characters e.g. projection sphericity (Riley, 1941), circularity and roughness were calculated and evaluated for usefulness.

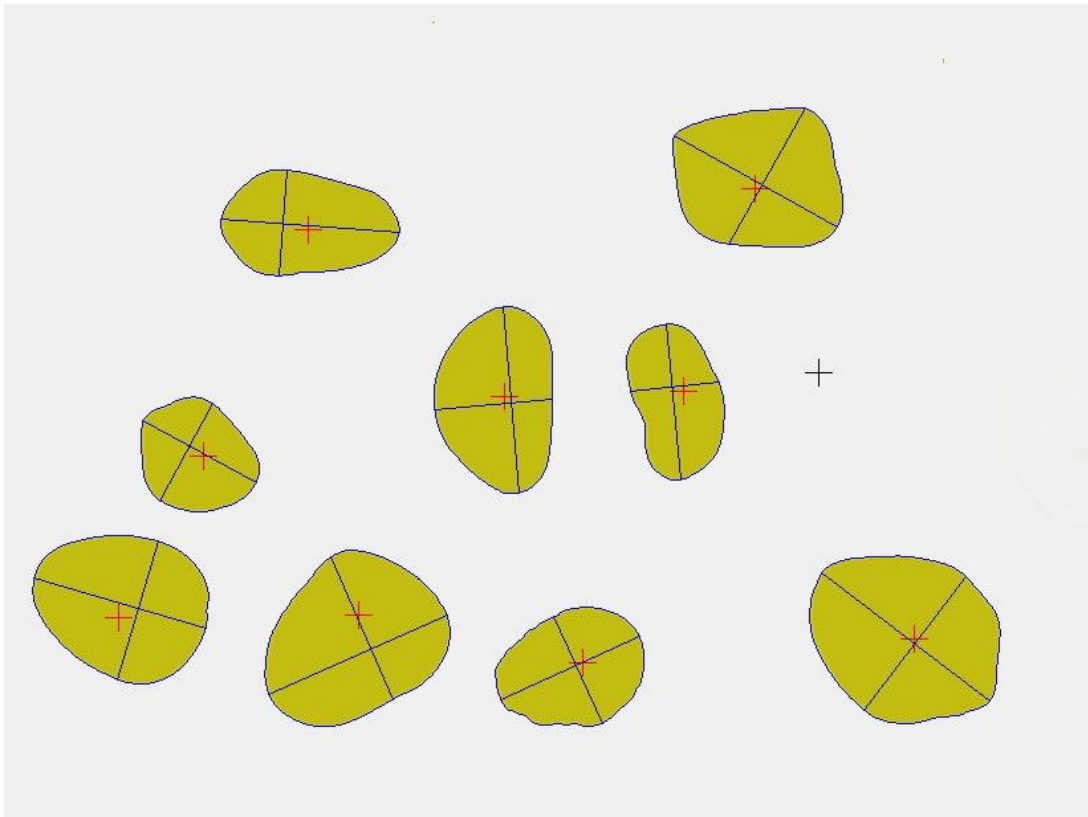


Figure 14: *Processed image of gravel showing extracted outlines and axes of pebbles*

3. SPHERICITY

Sphericity computed from the b/a aspect ratio of axes correlates well with the API standard, allowing sphericity computed as a simple function of b/a .

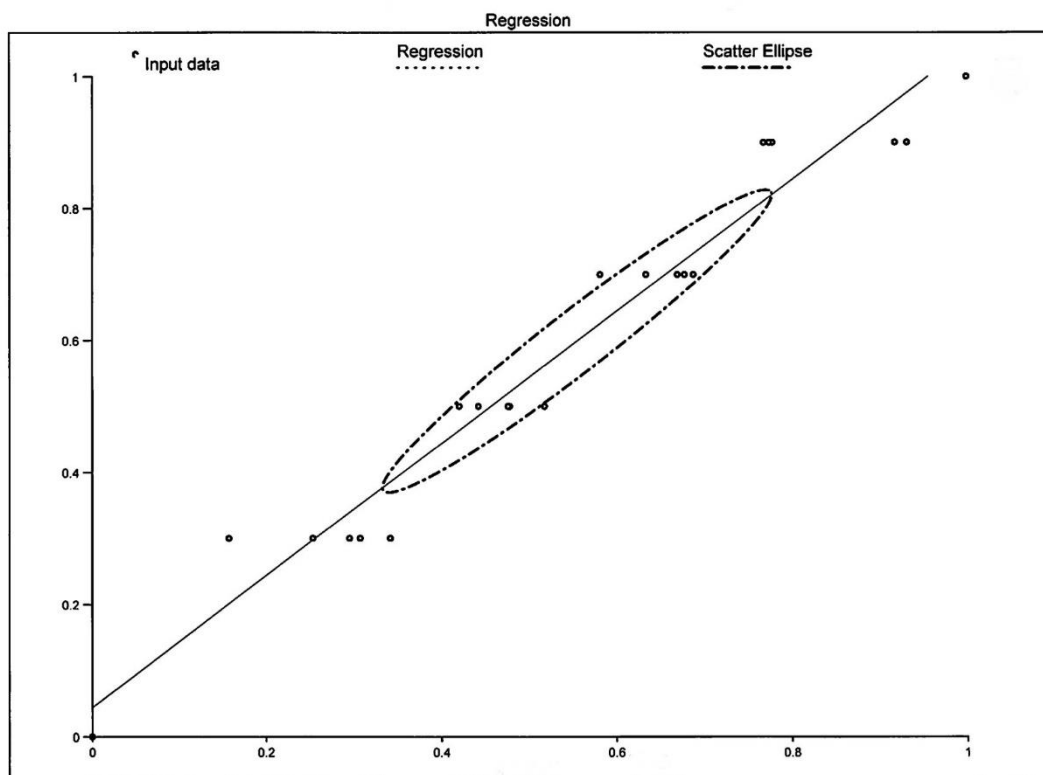


Figure 15: Regression of roundness with scatter ellipse.

X axis - computed roundness, Y axis - standard roundness of particles defining API standard

The computed sphericity matches the nominal sphericity of the standard visual charts within a few percent (Fig. 2). E.g. the chart shows a number of particles with a nominal sphericity 0.70. The range of computed sphericities for these is 0.63 - 0.73. Overall standard deviation of estimation is 0.06 through the sphericity range of the standard (0.30 to 0.90).

4. ROUNDNESS

Roundness is somewhat different as none of the simple measurements approximates the standard as defined by Krumbein & Sloss. This should not be surprising as the standard is not defined in terms of simple geometry, and, fractal considerations state that it can not be defined so. On the other hand, the usefulness and practical value of this heuristical standard cannot be denied. The

chart serving as the base of the standard differs significantly from the more detailed roundness chart of 1941 (Krumbein, 1941) despite the identical theoretical basis. A view of the charts suggests the use of Fourier analysis (Hung & Stroeve, 2006) and, indeed, this yields a measure reasonably agreeing with the chart. Overall standard deviation of estimation is 0.13. The use of this formula reduces both operator error and subjectivity, and speeds up processing.

5. CONCLUSIONS

Computerised image analysis is a preferred alternative to visual comparison in measuring sphericity and roundness, suggesting a modification of the current standard.

REFERENCES

- AMERICAN PETROLEUM INSTITUTE (2011): Fracking Sand Sphericity and Roundness. American Petroleum Institute, API RP-56, section 5, 1995/2011.
- BOWMAN, E.T., SOGA, K., & DRUMMOND, W. (2001): Particle shape characterization using Fourier descriptor analysis. *Geotechnique*, 51, 545-554.
- CLARK, N.N. (1987): A new scheme for particle shape characterization based on fractal harmonics and fractal dimensions. *Powder Technology*, 51, 243-249.
- DAVIS, J.C. (1973): *Statistics and Data Analysis in Geology*. Wiley, New York, 550 p.
- FOLK, R.L. (1955): Student operator error in determination of roundness, sphericity, and grain size. *J. Sedim. Petr.*, 25, 297-301.
- HUNG, J. & STROEVEN, P. (2006): The Shape of Aggregates. *Image Anal. Stereol.*, 25, 43-53
- JASKÓ, T. (2016): Quantitative Determination of Sphericity and Roundness of Pebbles and Sand Grains. Abstracts, 35th International Geological Congress, Cape Town
- KRUMBEIN, W.C. (1941): Measurement and geological significance of shape and roundness of sedimentary particles. *J. Sedimen. Petr.*, 11, 64-72.
- KRUMBEIN, W.C. & SLOSS, L.L. (1963): *Stratigraphy and Sedimentation*, Second Edition, W.H. Freeman & Co, San Francisco, 660 p.

- MELOY, T.P. (1977): Fast Fourier transforms applied to shape analysis of particle silhouettes to obtain morphological data. Powder Technol., 17, 27-35.
- RILEY, N.A. (1941): Projection sphericity. J. Sedim. Petr., 11(2), 94-97
- SCHNEIDERHÖHN, P. (1954): Bestimmung von Abrundung an Sandkörnern. Heidelb. Beitr. Miner. Petr., 4, 172-191
- SZÁDECZKY-KARDOSS, E. (1933): Die Bestimmung des Abrollungsgrades. Geol. Paleont. 34B, 389-401
- WADELL, H. (1932): Volume, shape, and roundness of rock particles. J. Geol., 40, 443-451.

Determining the sufficient number of stochastic realisations to represent spatial uncertainty

Noémi Jakab

University of Szeged, Department of Geology and Paleontology, Egyetem utca 2-6., 6722 Szeged,
Hungary, j.noemi@geo.u-szeged.hu

Finding the optimal number of stochastic realisations to represent spatial uncertainty is still a relevant question in geostatistics. The method used in this paper demonstrated this fact on the outputs of a sequential Gaussian simulation. The idea is to represent the stochastic images in an attribute space constructed using their static connectivity attributes. By choosing attributes with values occurring in closed intervals, it is possible to confine the realisations to a finite part of the attribute space. Generating new realisations can be considered as sampling this finite attribute space. In the long run, such a sampling results in a point structure where most of the points are located in areas of high point densities with low point densities surrounding them. High point densities represent typical realisations that show very similar connectivity characteristics, while low point densities correspond to atypical realisations. These deviate strongly from the bulk as they represent more extreme connectivity attributes caused by higher ergodic fluctuations.

In this sense, reaching the optimal number of realisations is the equivalent of reaching a state in the sampling process where high and low point densities are present at the same time, yet high point densities do not dominate the overall structure of the attribute space. A way to identify the optimum state is to regard the realisations in the attribute space as a spatial point process. In this framework, the optimum is reached when based on the inhomogeneous L-function complete spatial randomness is achieved, which is the equivalent of a point structure that is neither dispersed nor aggregated. This suggests that the sampling of the attribute space has reached a state where the information content of any additional realisations is most likely to be redundant, as reflected by the increasing aggregation of the point process.

Key words: *spatial uncertainty, geostatistical simulation, connectivity metrics, inhomogeneous Poisson process*

Spatial uncertainty quantification using distance-Kernel method

Mihály Apró

University of Szeged, Department of Geology and Paleontology, apromisi@gmail.com

Assessing the uncertainty in reservoir performance is a necessary step during the exploration phase. To examine the uncertainty in flow response, a large set of realizations has to be processed. There are several stochastic geostatistical algorithms, which can simulate multiple (100–1000) equiprobable realizations. Although these can show us the possible realities highlighting the spatial uncertainty, their handling is time- and CPU-consuming in the further processes, such as flow simulations. Consequently, just a small number of realizations can be post-processed in industrial practice. The purpose of this work is to find a method which can reduce the huge number of realizations in a way that the remaining ones retain the information of spatial uncertainty, regarding the flow behaviour shown by the whole set of realizations. To solve this problem ranking methods can be applied. The traditional ranking techniques, like probability selection, are highly dependent on the used static properties. In this paper, an alternative selection method is parameterized for measuring the pairwise dissimilarity between geostatistical models with a distance function based on hydrodynamic properties. Thus, in this case the distance function refers to the flow responses and allows visualizing the space of uncertainty through Multidimensional Scaling (MDS). The clustering of the given samples of the space of uncertainty could be done by K-means algorithm. However, the K-means clustering sometimes leads to wrong partitioning. In these cases Kernel transformation of the MDS data set is needed in order to get a feature space where the K-means algorithm can discover non-linear structures in the basic data set. After the back transformation, new clusters can be found, that, if the data contains non-linear structures, are different from the original K-means clustering, although the validation of the clusters needs further evaluation. The final step of the method is the selection of the Earth models closest to the cluster centres. This tool allows for the selection of a subset of representative realizations containing similar properties to the larger set.

Key words: *uncertainty, realization, clustering, Kernel, response*

1. INTRODUCTION

Stochastic spatial simulation is a widely-used method to quantify and display spatial uncertainty. Multiple, alternative, equiprobable realizations are used to represent the spatial uncertainty of the given, simulated variables. This set of realizations is a sampling of space of uncertainty of a spatial phenomenon. In most cases these realizations are not enough to assess uncertainty; further processing must be applied in the mirror of the preliminary issue. For instance, in reservoir engineering, the generated geological models are submitted for flow simulation. These can be petrophysical and/or lithological models. These are suitable to assess reservoir flow performance, to assess the effect of drilling new wells, and to locate their placements (Scheidt & Caers, 2007).

Handling of many realizations (100–1000) is time-, and CPU-consuming. Nowadays, traditional ranking techniques can be used to select representative models from a set, such as probability based, or quantile selection (P10, P50, P90), which is often based on static parameters. In general, a “transfer function” can be applied to post-process the set of Earth models to reduce the huge amount of data, focusing on the flow response of the reservoir, and retaining the information of spatial uncertainty. This method takes into account the hydrodynamic properties of the realizations and defines the dissimilarity between each model.

There are several similar methods, which can cluster the realizations based on flow properties, such as streamline simulations, or OOIP estimations, but these ones need dynamical assessments, such as fast-flow simulations, while the flow parameters of the presented technique were originated from static properties (Caers, 2011).

The aims of the research is to create a “transfer function” which can make the realizations pairwise comparable, and to quantify and visualize the spatial uncertainty regarding to the flow response of the reservoir.

2. METHODS

The input data set contained porosity logs from 22 wells. 100 realizations were simulated by Sequential Gaussian Simulation in SGeMS environment. These were generated in 3D, but averaging the values for every grid node along the Z axis was necessary (2D) to ensure their easier handling (**Figure 1A**). The further processes were done in Matlab (R2016a).

2.1. Definition of distance

A GIS tool, called TopoToolbox, was used to examine the 100 maps as topography maps and to derive hydrodynamic attributes from each one. There is a command package in the toolbox, which can calculate the drainage basins of each porosity/topography maps (**Figure 1B**) (SCHWANGHART & KUHN, 2010).

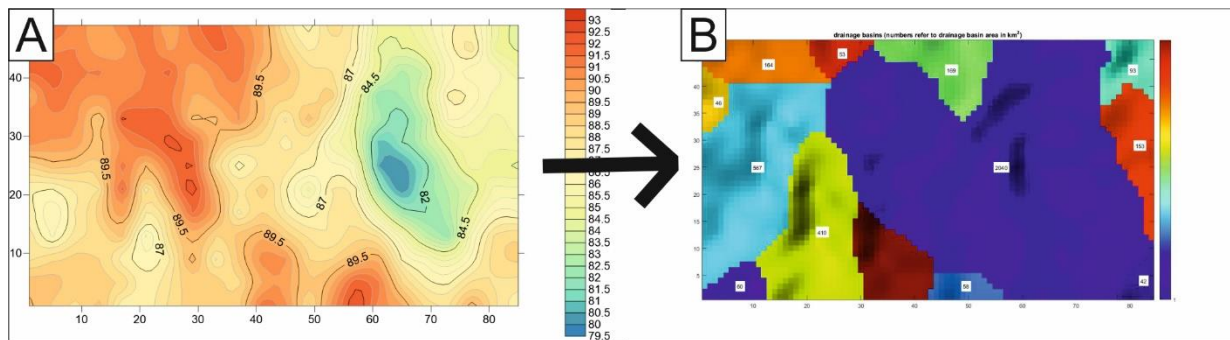


Figure 1: An average, inverse porosity map (A), and its drainage basins (B)

For the distance definition, the number of drainage basins, the area and the heterogeneity of the biggest basin (**Figure 2A**) were considered. The biggest basin always occurred at nearly the same area of the maps, representing the HC reservoir (**Figure 1B**), that is why it was chosen as a distinguishing property. The heterogeneity means the median and the standard deviation of the cells at different velocity of the biggest drainage basin (**Figure 2B**) (Schwanghart & Kuhn, 2010).

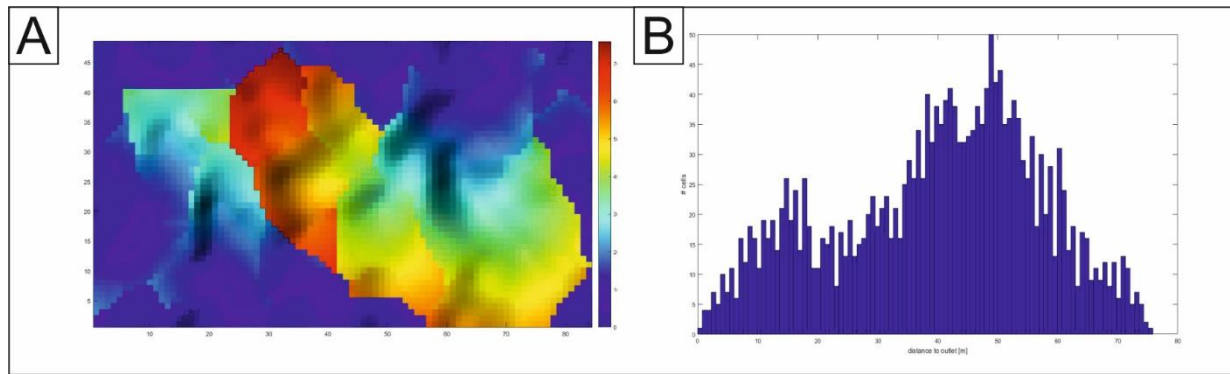


Figure 2: The velocity map of the biggest drainage basin (A), its histogram of number of cells, associated with different velocity (B)

In addition to the above described 2D properties, the volume of higher than 15% of porosity was measured as well in each 3D realization. Measuring the pairwise dissimilarities of the realizations, based on these properties, resulted in a dissimilarity distance matrix.

2.2. Multidimensional Scaling (MDS)

MDS is a technique applied to transform the dissimilarity matrix into a configuration of points (realizations) in n-dimensional Euclidean space. The distances between the points correspond as much as possible to the dissimilarities of the maps. There are two kind of algorithms designed for examining a single dissimilarity matrix: classical and nonmetric MDS (Scheidt & Caers, 2007).

In this case, classical MDS was employed because it assumes that the matrix displays metric properties, such as distance measured from the map, thus, the distances retain the intervals and ratios between the points as well as possible (Scheidt & Caers, 2007).

2.3. Clustering of the realizations

Cluster analysis can discover inner organizations of a dataset by searching for structures within the point cloud. Thus, the data is divided into a number of groups, containing similar number of objects. The number of clusters was

specified by density based hierarchical clustering, called OPTICS, and dendrogram (Ankerst et al., 1999).

After choosing the proper number of clusters, k , K-means algorithm was applied to assign points to each of k clusters, S_i , by minimizing the expected squared distance between points and the cluster centres, μ_i . Generally, the algorithm works properly, but it is sensitive to the distribution of points, because the initialization of randomly selected cluster centres is determined by the arrangement of points, at the beginning of clustering process. A data set with less linearity can result in an unfortunate initialization which may group the object in a wrong way (**Figure 3**). To avoid this situation radial basis function (Gaussian kernel) was applied to transform the points into feature space, and arrange them more linearly (CAERS, 2011). The process is shown in Figure 3. The meaningful K-means clustering is ensured in this space and inner structures may be discovered within the data set after a back-transformation.

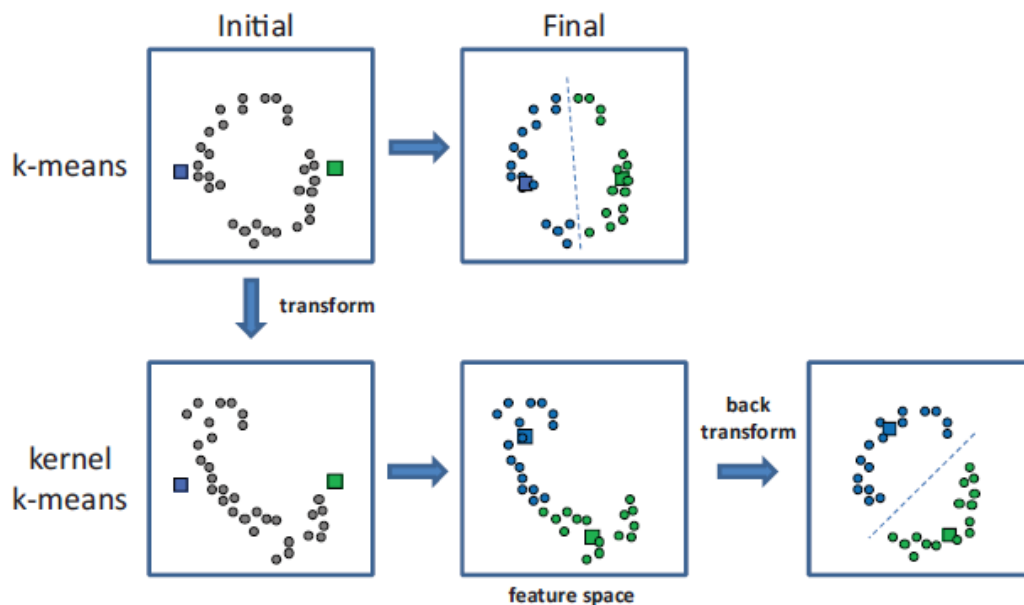


Figure 3: The difference between K-means and Kernel K-means clustering (Caers 2011)

2.4. Selection of the representative realizations and evaluation of the clusters

Finally, the selection of the representative Earth models being closest to the cluster centres is last phase of the method of uncertainty visualization (Caers,

2011). The evaluation of cluster membership is the key point to assess the quality of the clustering.

3. RESULTS AND DISCUSSION

The method was applied to a Lower-Pannonian turbidity sand body. The input data for the Sequential Gaussian Simulation were porosity logs from 22 boreholes, resulting in 100 equiprobable realizations, with application of an omnidirectional, exponential variogram model and 84X48X48 grid resolution. After the dimensionality reduction (3D->2D), applying the GIS tools of Matlab, the dissimilarity distance matrix was computed, and the MDS coordinates were plotted (**Figure 4**).

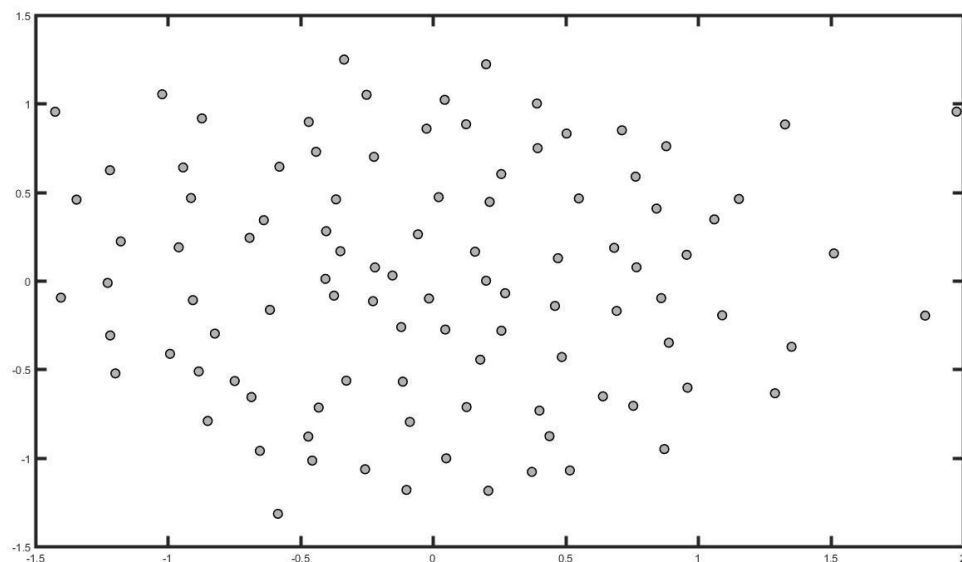


Figure 4: Result of Multidimensional Scaling (X axis: components of dimension 1, Y axis: components of dimension 2)

In the most cases, 3 realizations (P10, P50, and P90) are chosen for dynamical simulations in industrial circumstances. However, in this case the density based hierarchical clustering resulted in 4 clusters.

The data structure of MDS plot had less linearity, thus, the radial basis function (RBF) was applied with $\sigma=25$, which is signed the flexibility of the function (**Figure 5**).

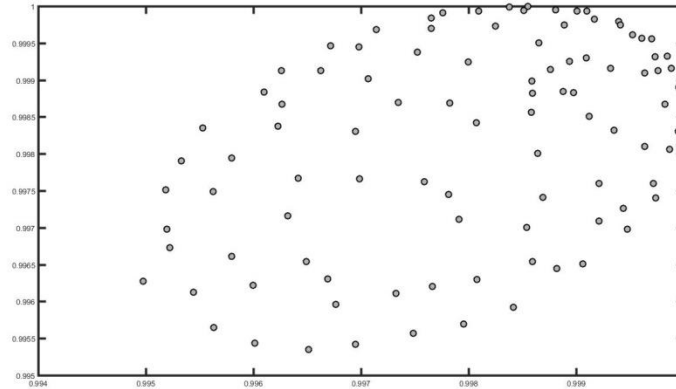


Figure 5: The data set after the RBF transformation, in feature space, (X axis: components of dimension 1, Y axis: components of dimension 2)

In this work the classic K-means algorithm was applied in feature space to determine the previously defined 4 subsets. As can be seen in the Figure 6, among the 4 clusters having been revealed the back-transformation one took a central position in coordinate system and the other three clusters surrounded the central one. Realizations, being closest to the corresponding cluster centres represent undoubtedly real groups. The distances between any two group-centres are proportional with the dissimilarity between these groups (**Figure 6**).

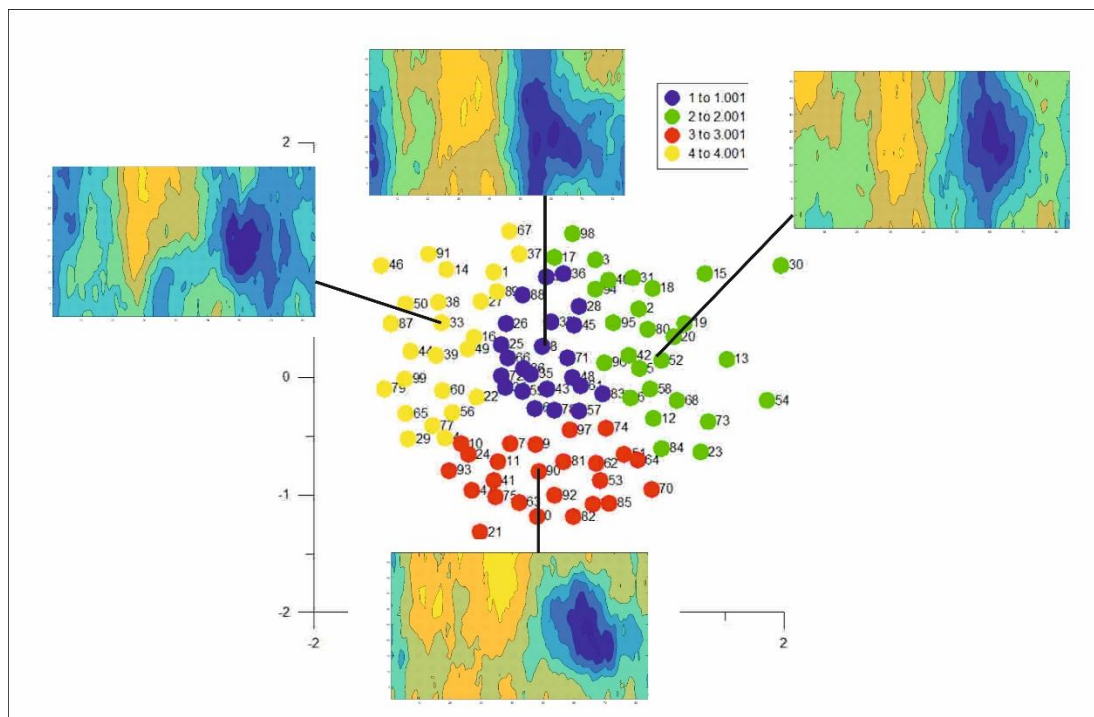


Figure 6: Result of the clustering with the selected realizations

4. CONCLUSIONS

The presented realizations-grouping method does not require dynamical simulations, or original oil in place (OOIP) estimations. It relies on flow-related and volume properties, and on the relation between the distance measure and flow behaviour of the realizations. The selected 4 models represent the 4 clusters and they also provide information about the spatial uncertainty of the reservoir, focusing on its performance. The application of this new process showed promising results, but it needs further improvements and other distance measurements should be tested in more complex cases.

REFERENCES

- ANKERST, M., BREUNIG, M. M., KRIEGEL, H-P & SANDER, J. (1999): OPTICS Ordering Points To Identify the Clustering Structure, ACM SIGMOD international conference on Management of data, Philadelphia, 49–60.
- CAERS, J. (2011): Modeling uncertainty in the Earth Sciences, Department of Geological Sciences, Stanford University, Stanford 156–186.
- SCHEIDT, C. & CAERS, J. (2007): A workflow for Spatial Uncertainty Quantification using Distances and Kernels, Stanford Center for Reservoir Forecasting Annual Meeting report 20, Stanford University
- SCHWANGHART, W., KUHN, N.J. (2010): TopoToolbox: a set of Matlab functions for topographic analysis, Environmental Modelling & Software, 25, 770–781.

Geothermal potential estimation with Monte Carlo method

László Zilahi-Sebess¹, Ágnes Gulyás¹

¹Geological and Geophysical Institute of Hungary (MFGI), zilahi.sebess.laszlo@mfgi.hu

There are two main ways of geothermal reserves estimates: the physical geological modelling and the prediction from production data sets. The speciality of geothermal potential and resource estimation is that the raw material (geothermal energy) can be found everywhere, and yet still keep supply, therefore should only be considered as reservoir where the thermal energy can be exploited. Without reinjection only the heat content of movable water can be exploitable. While total amount of heat-in-place is known with small error (only couple of percent relative standard deviation), the estimation of heat energy content of movable water is more uncertain therefore it is advisable to specify the degree of uncertainties. One of the best ways of uncertainty characterisation is Monte Carlo method. In Monte Carlo Simulation the elements of physical model (components of rock model, temperature, and geothermal gradient) are handled as probability variables. The expected value, standard deviation stems from measurement or from apriori knowledge. Sorting in ascending order, the results (using variables with noise originated from random number generator performed at least thousand times) provides a distribution. From this distribution the resource can be calculated according the probability levels – P90 is the most pessimistic, P50 is the most probable and P10 is the most optimistic estimation in UNFC-2009 system. The most important parameters for estimating exploitable heat content are effective porosity, permeability and temperature (handled as statistical variables). In the presentation the results of Monte Carlo Simulation is shown: the geothermal resources and geothermal potential of a certain area is calculated.

Key words: *Probability, Monte Carlo, Estimated geothermal resource, Movable water*

1. INTRODUCTION

Theoretically there is no limit of how many percent of the stored heat energy could be exploited from a reservoir, but in practical reason the heat content of the movable water means the exploitable part of the stored heat. This approach is justified by that the conductive heat transfer is negligible to the convective

heat transfer. In first approximation in the geothermal prognosis it is not important to deal with the natural heat supply, conductive geothermal heat flow is negligible. The amount of energy recovered with water production is primarily controlled by the temperature and the effective porosity. Therefore, the assessment of reserves (based on the volumetric porosity component model), and potential needs estimated permeability as well. An important problem with the geological mathematical model is that even the measured physical parameters without any stochastic error have uncertainties in geological meaning because these are representative only for small volumes and not for the whole reservoir. If there are not enough data to build in the model, input parameters of the model can only be depth trends with uncertainties. Monte Carlo Simulation can be a potential solution to handle this problem. In Monte Carlo Estimation the elements of the physical model are handled as probability variables (Grant 2015, Garg & Combs 2010). The main goal of the study is defining the most probabilistic exploitable heat from the effective pore space and the heat content of the rock matrix and bound water.

2. MATERIALS & METHODS

2.1. Monte Carlo Simulation

The Monte Carlo Estimation can be performed on different complexity levels. In the simplest case a data set of a previous calculation is used and combined with white noise. It is a more sophisticated approach if productivity is involved as well.

The main elements of the model are the total porosity and „shaliness” that may be used as stochastic variables. The effective porosity is calculated from the above two parameters. In prognostic model the porosity is derived from the porosity-depth trends (Mészáros & Zilahi-Sebess 2001). An even better solution is a more complex model with stochastic variables such as total porosity, shaliness and temperature. In prognostic estimation of resources and reserves the original stochastic variables are the coefficients of porosity depth trends, shaliness and temperature depth trends.

In the geological model, any of the above mentioned objects are required to calculate the following volumes of the earth's crust:

1. Free water-filled pore volume and permeability of Pannonian sediments;
2. Pore volume of alteration zone of the crystalline basement and basal conglomerate at the bottom of Pannonian Basin;
3. Pore volume of the carbonate rocks in the basement under Pannonian;
4. Pore volume of the tectonised rocks of neotectonic fault zones.

2.2. Elements of simplified geological-physical model

The amount of energy recovered with water production is primarily controlled by temperature, and effective porosity. The exploitable amount of heat vs. time unit depends on leakage factor. Leakage factor strongly depends on the shaliness that could be estimated from well logs, and petrophysical analysis of core samples. Even the hydrodynamic tests are not representative for the entire volume of reservoir thus they have uncertainty as well. Substantially all factors involved in the inventory and potential estimation may be considered as probability variables. In Monte Carlo Simulation the original deterministic factors are handled as statistical distributions with the expected value of the original deterministic factor with a given standard deviation (originated from previous measurements of the area or based on apriori information). Standard functions can be applied for the estimation of the physical parameters. The parameters of standard functions are considered as statistical variables.

Elements of geothermal potential estimation:

1. Local porosity model (porosity components);
2. Porosity vs. depth (compaction trends in basin);
3. Thermal parameters (specific heat conductivity and capacity);
4. Estimation of permeability using porosity;
5. Temperature as function of depth (geothermal gradient);
6. Shaliness (depends on the formation);
7. Carbonates in the basement (maps, and estimated thickness);
8. Thickness of altered zones and basic conglomerates;
9. Geothermal heat flow (measured and estimated from basin depth);

10. Estimation of technical parameters.

An important problem with the geological-mathematical model is that even the measured physical parameters with no error have uncertainties in geological meaning because they are representative only for small volumes not for whole reservoir (**Figure 1**). The regional primary heat flow (conductive heat flow) is a constant value that can be determined from a few well logs. It is important when the temperature-depth function is estimated from the porosity compaction trend. Ideally it can be determined from a well in nearly steady-state (unperturbed for a long time, being in thermal equilibrium with its environment), or by measuring time series of the continuous temperature of the same well.

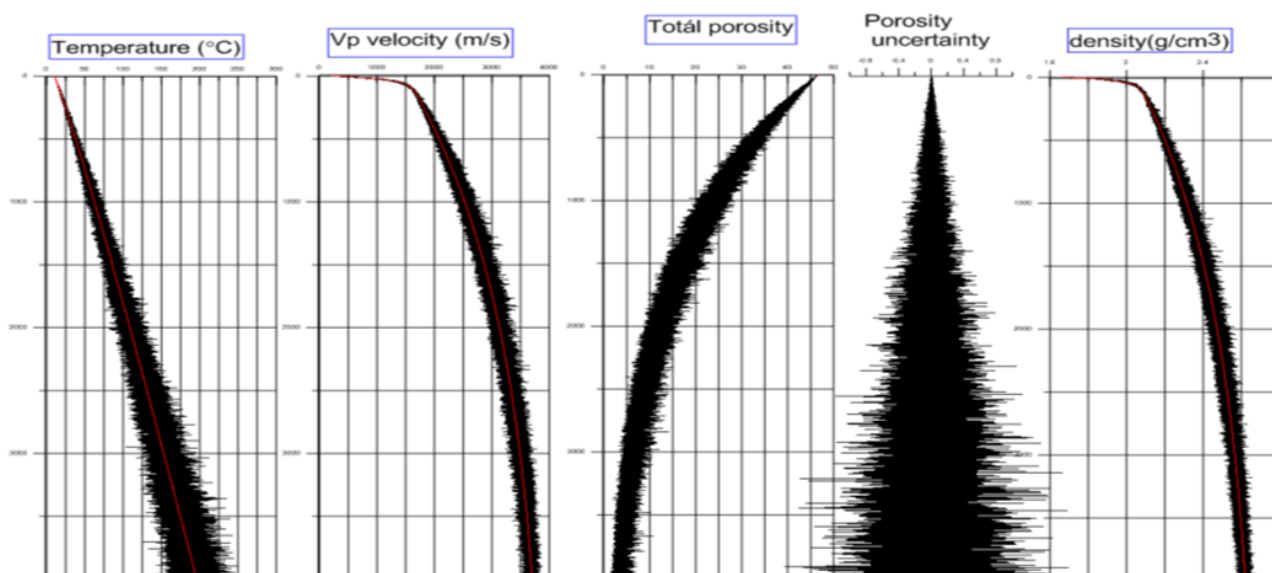


Figure 1: Physical parameters vs. depth trends with noise

In the case history shown in **Figure 2**, there are wells with known yields. The catchment area around the wells depends on the production, effective porosity and permeability. There are two opportunities for the estimation of the necessary volume of sustainable production. In the first case the true production is applied to estimate the reserve. In the second case the calculation is based on the possible maximum production, controlled by porosity, permeability and maximum possible applied depression (supposed that the production is under the possible level).

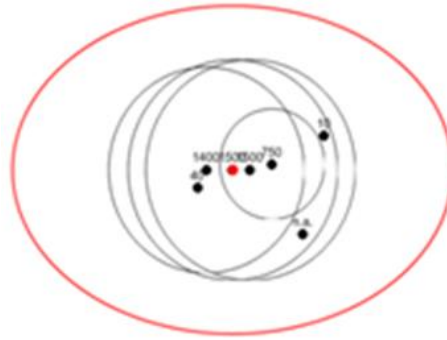


Figure 2: Group of wells (with black dots) with the calculated effective catchment area (with red ellipse)

3. RESULTS

The Monte Carlo Calculation helps us to estimate the limits of exploitable heat energy, and proves us information on the estimated catchment area.

The reserves of the individual wells were estimated with Monte Carlo method from the porosity, permeability and known thickness of the layers (**Figure 3**). The thickness was also handled as statistical variable.

The reserves were estimated for the whole country divided it into 14 areas and for 3 stratigraphical units. Within each unit heat content of the total porosity, heat content of effective porosity, heat content of the rock matrix and the exploitable heat were calculated on the most pessimistic (P90), the most probable (P50), and the most optimistic (P10) levels (**Figure 4**).

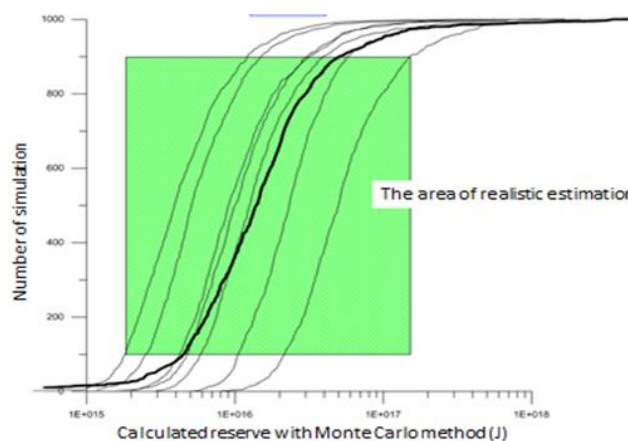


Figure 3: The individual distributions (thin black lines) and average distribution (thick black line) of the calculated geothermal reserve (in Joules)

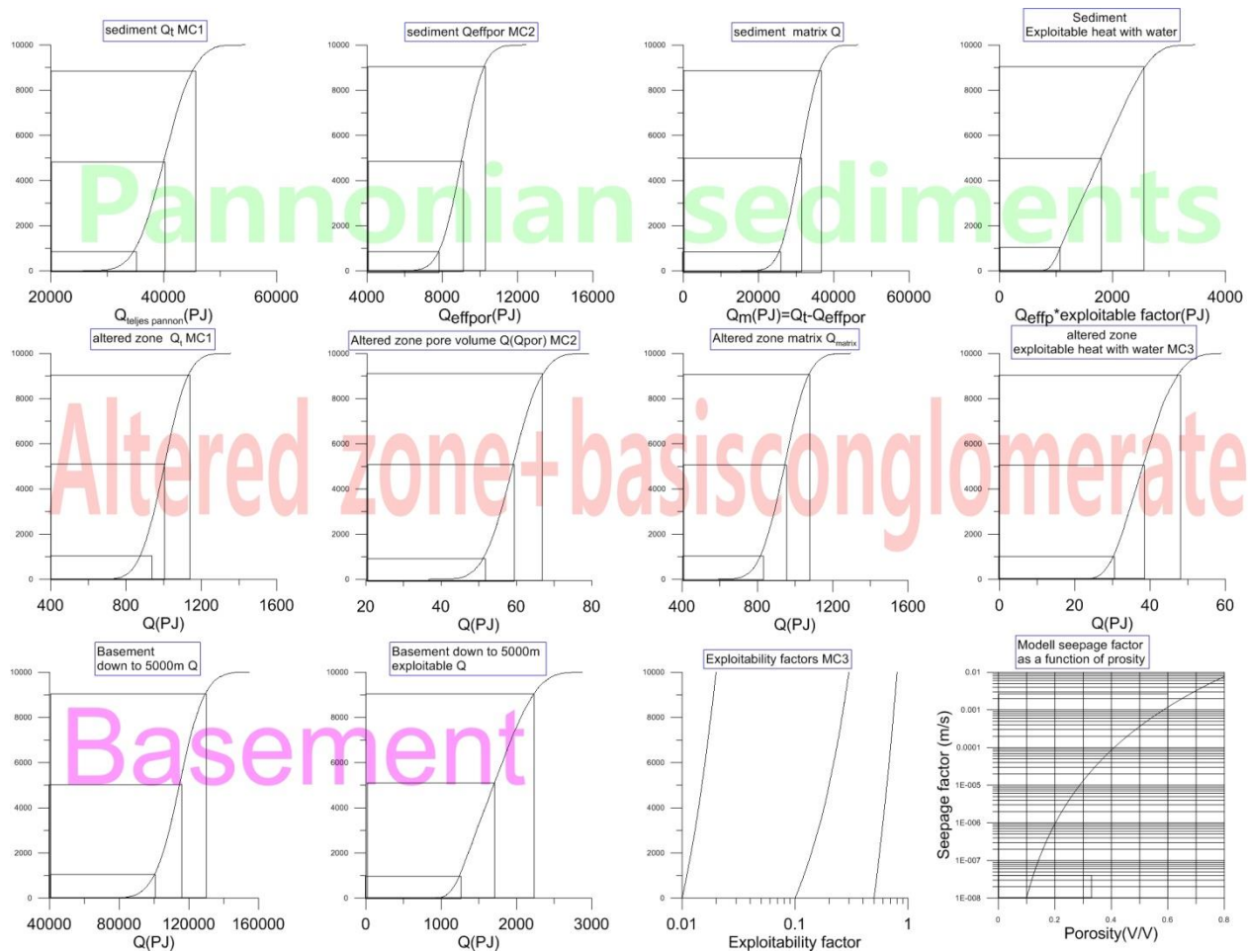


Figure 4: Distributions of estimated heat content in various lithological and stratigraphical units

The upper row represents the Pannonian sediments. The middle row represents the altered zone of basement and basic conglomerate. The third row represents the basement.

In each row the first simulated distribution is for heat content of total rock volume, the second one is for effective porosity, the third is for rock matrix and the fourth is for the exploitable heat.

4. CONCLUSIONS

Based on former deterministic resource calculation a new Monte Carlo Estimation was calculated for the whole country focusing on the exploitable heat energy from the sediments, and altered basement.

The total 'heat in place' is relatively less variable. The exploitable heat energy varies within rather wide range (difference between P10 and P90) even within relatively accurately known parameters with 10% relative standard deviation.

Another, more precise estimation is planned based on the newly calculated elements of the assessments where the standard deviation of the elements will be based on the measured values of each area of evaluation.

REFERENCES

GARG, S.K., & COMBS, J. (2010): Appropriate use of USGS volumetric “heat in place” method and Monte Carlo calculations. Proceedings of the 35th Workshop on geothermal reservoir engineering. Stanford University. 2010.

GRANT, M. A (2015): Resource Assessment, a Review, with Reference to the Australian Code. Proceedings World Geothermal Congress 2015. Melbourne, Australia, 19-25, April 2015

MÉSZÁROS F. & ZILAHÍ-SEBESS L. (2001): Compaction of the sediments with great thickness in the Pannonian Basin (Nagyvastagságú üledékek kompakciója a Pannon-medencében). 2001. Budapest. Geophysical Transactions 44, 1, 21-48

UNFC-2009: United Nations Framework Classification for Fossil Energy and Mineral Reserves and Resources 2009

http://www.unece.org/energy/se/unfc_2009.html

Uncertainty of subsurface uncertainty analysis – can we mitigate it?

Matyas Sanocki¹

¹Reservoir Geologist, MOL Plc., msanocki@mol.hu

Modelling subsurface environments have always been challenging due to the many factors that could not be measured directly. Albeit, oil and gas industry needs good enough description of the subsurface reservoirs which can serve as a basis for capital intensive business decisions. In the past, the life cycle of an oil&gas field followed the so-called funnel diagram of the subsurface uncertainties – as moving towards the project life cycle, uncertainty range was being narrowed down by time. Today, sophisticated uncertainty analysis workflows are run during each project step as a mandatory part of exploration-appraisal-development-abandonment project stages. As gathered data quantity and also quality skyrocketed in the last two decades, now we are dealing with more data and more knowledge about of oil & gas reservoirs than ever. Was the uncertainty range narrowed with the data revolution? The answer is no – today's funnel diagrams are reversed in many cases. With the introduction of more and more uncertainty parameters, our overall range of uncertainty seems to form an ever-increasing trend. But what do we call as uncertainty? The uncertainty we face with when performing each geomodelling step or the uncertainty we face with when deciding how to deal with those geomodelling steps? This presentation shows a case study of a Hungarian fractured & karstified (mainly) limestone oil reservoir where evolution of different uncertainty evaluation methods can clearly be seen. This project provides a good example where stochastic and thus easily quantifiable uncertainty ranges are significantly narrower than any of the conceptual uncertainty cases. The biggest challenge of uncertainty analysis is to add and quantify those conceptual uncertainty cases to the overall uncertainty analysis workflow.

Key words: *uncertainty analysis, subsurface uncertainty, geomodelling, reservoir geology*

Postscript

**The 20th Congress of Hungarian Geomathematicians and
9th Congress of Croatian & Hungarian Geomathematicians**

was organized by the

Hungarian Geological Society and

co-organized by

the Geomathematical and Informatics Section of the Hungarian Geological Society,
the South Transdanubian Regional Branch of the Hungarian Geological Society,
the Croatian Geological Society,
the Hungarian Academy of Sciences' Geomathematical Subcommittee of the Committee
on Geology,
the 2ka Palæoclimate Research Group of the Hungarian Academy of sciences,
the Pécs Regional Committee of the Hungarian Academy of Sciences,
the Cluster of Applied Earth Sciences,
the Enterprise Europe Network Chamber of Commerce and Industry of Pécs-Baranya.

Conference venue



Hotel Makár
Pécs Középmakár dűlő 4, 7635



Chamber of Commerce and Industry of
Pécs-Baranya

