UNIVERSITY OF SZEGED
Faculty of Science
Doctorate School of Earth Sciences
Department of Physical Geography and Geoinformatics

LARGE SCALE GEOSTATISTICAL MODELLING OF THE SHALLOW GROUNDWATER TIME SERIES ON THE SOUTHERN GREAT HUNGARIAN PLAIN

TWO APPROACHES FOR SPATIOTEMPORAL STOCHASTIC SIMULATION OF A NON-COMPLETE MONITORING DATASET

Thesis of Dissertation

ZSOLT ZOLTÁN FEHÉR

Supervisors:

Dr. János Rakonczai
university teacher

Dr. János Geiger
associate professor

Szeged, 2019

1. Introduction and objectives

The conditions of freshwater resources on Earth (pollution, a rapid diminishing of per capita reserves) constitute a major environmental problem of our age. Shallow groundwater resources play a decisive role for vegetation (and thus the agricultural sector), the change to which has been impacted significantly by climate change in recent decades. Therefore, the precise analysis of the volume changes to these resources can be considered an important economic interest and is also included in the requirement formulated by the Water Framework Directive of the EU. Although the English *"groundwater"* phrase tends to be interpreted with confined and unconfined waters collectively, it is usually assessed independently in Hungary with regard to its significance in practice.

Despite the fact that research into shallow groundwater changes (the analysis of fluctuation characteristics, the mapping of water resource changes, etc.) goes back to a century in Hungary, analyses taking a geostatistical approach are quite left out of consideration. Due to the relatively sparse observation network in space and the quite heterogeneous relief in many territories, maps prepared based on conventional interpolation procedures are affected by significant estimation error, but this has received little attention so far, as they signalled major trends of change anyway. Moreover, a scientific interpretation of the problem in sufficient depth had been impossible for a long time in lack of a proper IT and geostatistical background. However, in view of the favourable change to such background conditions, the preparation of analyses based on substantiated professional grounds and with regard to several practical aspects have become essential by now.

The information on the actual state of groundwater may only be derived at specific temporal moments and spatial locations. Therefore we need to infer its depth at unsampled coordinates and long data deficit periods in order to determine water resource changes and the areal pattern (Mucsi et al 2013). The estimation of current groundwater conditions can be interpreted as a submodel of some chosen environmental models. Because of the sparse availability of information and the limits of interpretation of relations between the individual environmental factors, environmental models are affected by factors of uncertainty of various sources (Caers 2011). Such factors are present in every case, regardless of the physical model chosen (Fig. 1).

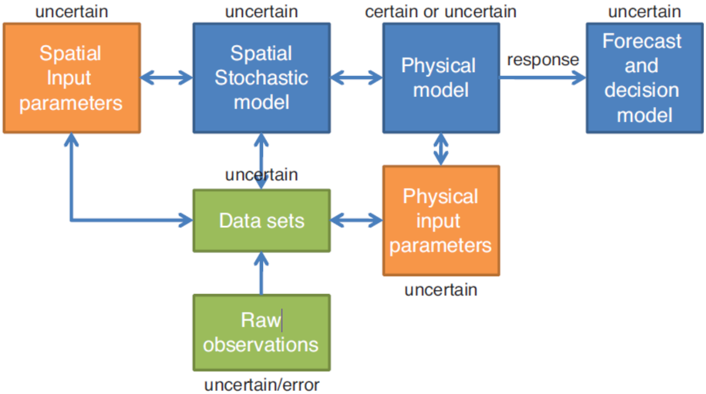

Fig. 1. Relationship between factors influencing the reliability of earth science models (Caers 2011)
(Arrows show the directions of relationships)

Shallow groundwater is often applied as the bottom boundary condition. However, a more realistic estimation of groundwater can give rise to a more successful application of data-driven environmental and hydrological models. The count of groundwater time series of a low areal density can be increased with stochastic, artificial hydrographs, which in turn makes the calibration of Artificial Intelligence based procedures possible with datasets of various spatial structures. However, for the application of alternative spatial structures to makes sense, the nature of spatial relationships (spatial continuity, heterogeneity, texture) must be preserved between the momentary values of the estimated artificial time series.

The common interpolation approaches, including the widely praised kriging estimations, are incapable of reproducing the texture between estimation nodes (Journel 1992), which actually projects the application of stochastic simulations. The sequential Gaussian simulation applied in my thesis reproduces the spatial or the temporal heterogeneity, as well as the statistical characteristics of groundwater observations, completely (Kyriakidis 1999, 2001, Ekstrom et al 2007).

However, the application of 3D variography based on two-point data statistics in a traditional way is hindered. The reason for this is that the classical approach is unable to reproduce spatial and temporal relations contemporaneously (Füst 2007, Gneitling et al 2007, Ma 2008, De Iaco 2010). Thanks to profit-oriented requirements aimed at the support of mining operations, determining the rate of uncertainty inherent in the models has become an essential means of earth sciences in recent decades. The purpose of stochastic models is to test hypotheses or to simulate alternative cases of specific geological processes in order to minimize the risks of making wrong decisions.

Only a few examples can be found in Hungarian literature for the practical application of research and procedures focusing on joint spatiotemporal stochastic geostatistics (Füst 2007, 2011, 2012, Füst and Geiger 2010, Geiger 2015). At the same time, from a geographical point of view, there is an increased demand e.g. for a survey of the new environmental challenges posed by climate change. On the one hand, the need to predict the future based on past environmental processes arises (principle of Hutton's actualism), while, on the other hand, it would be necessary to be known present conditions as precisely as possible, which is a prerequisite for quick decision-making involving the least amount of risk. Two different algorithms need to be applied to satisfy these two different data demands, which the classical geostatistical approach is unable to provide. My thesis discusses the joint spatiotemporal estimation of groundwater level. With consideration to the practical needs encountered, my core objectives are as follows:

- I wish to perform *methodological research* enabling a more precise estimation of groundwater based on the probability theory. To this end, I will analyse the accuracy of widely used and more complex geostatistical estimation methods applied in lack of certain data. On the one hand, I will create based on the experience gathered a fast, stochastic estimation algorithm that can be run real-time and produce groundwater estimation grids for the immediate decision-making process. On the other hand, I would like to develop an algorithm capable of providing even more accurate groundwater estimations exploiting time-series characteristics when the data gap period exceeds the temporal range.

- Based on the algorithms developed, I will estimate water resources in two regions. The Danube-Tisza Interfluve is severely affected by water discharge in the recent 30 years. The Southern Transtisza region is a valuable fertile land where the inland excess water causes problem occasionally. These findings will constitute the *earth science findings* of my thesis.
- Furthermore, I also wish to optimize the present groundwater level monitoring network in order to improve groundwater estimations. The current network was developed according to the professional requirements of the 1950s, however, novel geostatistical approaches demand a totally different design for spatial observation.

2. Study area and methodology

2.1. Study area

My research was mainly focused on the Danube-Tisza Interfluve and the operational area of the Lower Tisza District Water Directorate (Fig. 2). I chose the first one for its being the most distinctive landscape of the Hungarian groundwater discharge and the second one for its being composed of two units with different hydrological characteristics. The latter was essential for the spatial extension of the applicability of the models. In other forms of cooperation, I also conducted estimations of water resources in other regions of Hungary (Nyírség, Northern Great Hungarian Plain), which made it possible for us to compare the sensitivity of different kinds of geographical landscapes to climate change.

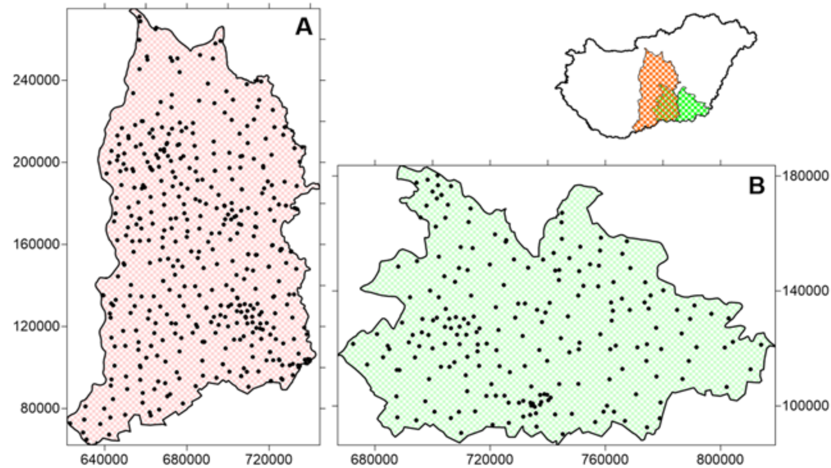

Fig. 2. The Danube–Tisza Interfluve (A) and the Lower Tisza District Water Directorate (B) study area with the considered groundwater observation network

2.2 The data used

Data were derived from the most detailed official Hungarian hydrological database (MAHAB) for 1950–2017. Previously, I had obtained three types of data from the database set up by the Environmental Protection and Water Management Research Institute (VITUKI) from each local water directorate: (1) well coordinates, (2) geometric data relevant to the wells, including ground elevation (above sea level), pipe edge (above sea level) and well depth and (3) water levels measured in the individual wells. I developed a uniform databased from the non-organized data obtained, thus enabling absolute and relative ground water level calculations and the interactive spatial selection of wells based on their coordinates. I also performed error filtering on the data organized based on previous experience.

Occasionally, some unusual leaps could be observed in some of the groundwater time series. In most of the cases, unusual values were caused by data digitalization error, data conversion error during various database export-import operations or the confusion of data series in the course of the organisation of unstructured raw data. Problematic data series were analysed based on water management considerations with the contribution of an external expert. Thereafter, only data series deemed to be reliable were used for the different types of evaluations. 533, 207 and 1196 time series were involved in the analysis from the regions of the Danube-Tisza Interfluve, the Lower Tisza District Water Directorate and the Great Hungarian Plain, respectively, which means that more than 2 million measurements were considered in total.

2.3. Geostatistical background

Geostatistics examines natural phenomena varying in space and/or time (such as shallow groundwater) based on a particular, location-dependant random variable, the so-called regionalized variable, by means of statistical statistic and deterministic tools (Matheron 1967, Deutsch – Journel 1998). Locally, the regionalized variable acts like a random variable and presumes some kind of a spatial and temporal regularity between the individual points of observation (Cressie 1985, Pannatier 1996).

It is possible to make estimations at unsampled locations based on the spatial structure considered to be known (such as the semivariogram model), and geostatistics supplies various kriging interpolators to this end (Goovaerts 2000, Szatmári – Pásztor 2015).

In the course of cokriging, the estimation can be improved significantly by the consideration of auxiliary data (Kohán 2014), however, the construction of the variogram model becomes much more complicated (Isaaks-Srivastava 1992, Deutsch-Journel 1998, Fehér-Rakonczai 2012). The advantage of collocated cokriging is the elimination of the overweighting effect of the auxiliary data, if it is much more densely sampled than the primary variable (Xu et al 1992, Almeida 1993, Almeida-Journel 1994). The application of the two Markov models of collocated cokriging simplifies the construction of the semiovariogram model considerably (Journel 1999, Shmaryan–Journel 1999, Xianlin–Journel 1999).

Groundwater level is influenced by large-scale effects (e.g. precipitation, elevation) in the first place. However, the impact of local factors (permeability, land use, hydrography) give rise to significant small-scale, apparently random heterogeneities (Marton 2009, Mucsi et al 2013, Rakonczai – Fehér 2015).

Small-scale heterogeneities can be reproduced by means of stochastic simulations (Carr–Myers 1985, Journel 1989, 1993, Deutsch–Journel 1998, Boisvert–Deutsch 2011). The procedures are meant to generate a substantial amount of alternative, equiprobable estimations at unsampled locations. Sequential Gaussian simulation has been used for groundwater estimations since the early 2000s in Hungary (Mucsi et al. 2013, Fehér– Rakonczai 2012).

Groundwater displays some kind of a structural relationship in time, too. The temporal characteristics of hydrographs can be clearly detected with dynamic factor, wavelet spectrum and periodogram analysis (Kovács J. et al 2004a, b, 2011a, b). The relationship between groundwater time series and precipitation can also be well detected (Rétháti 1977a, b, Kovács F.–Turai 2004, Kovács F. 2014a, b).

However, the construction of the model of a joint spatiotemporal structure is problematic in terms of geostatistics (Kyriakidis – Journel 1999). The reason for this is that we need to determine some regularity – instead of using some 3D semivariogram (Füst 2007) – which is capable of defining the system of locally varying relationships between space and time only (Kyriakidis 1998, 1999, Kyriakidis – Journel 1999, 2001a, b, Gneitling et al. 2007, Ma 2008, Geiger 2015).

To sum it up, my task was to perform the joint spatiotemporal stochastic simulation of groundwater in reliance on the temporal pattern of precipitation and the spatial pattern of relief.

2.3.2. Real-time recursive sequential simulation of water table elevation above sea level based on the digital elevation model

Data are practically never available to us contemporaneously for each and every well. The procedure I developed determines groundwater level at a given time instant based on a spatial reference pattern, without the lack of data influencing the estimation of water resource changes.

I compared three patterns of estimation in my analyses: the classical co-simulation (Shmaryan-Journel 1999), the chain of simulations (Kyriakidis 1999, Geiger 2015) and the recursive chain of simulations with the backfeeding of the digital elevation model

(DEM) (Fehér–Rakonczai 2019) (Figure 3). The interpolator of the sequential Gaussian cosimulation applied is simple collocated cokriging, for which I used the Markov 2 type variogram structure.

I used the time series of 213 groundwater wells in the Danube-Tisza Interfluve as a whole during the development. The performance of the models run for the period of 1976–2003 were analysed in light of the maps obtained for 2003, the maps of changes to groundwater level between 1976 and 2003 and the time series of water resources.

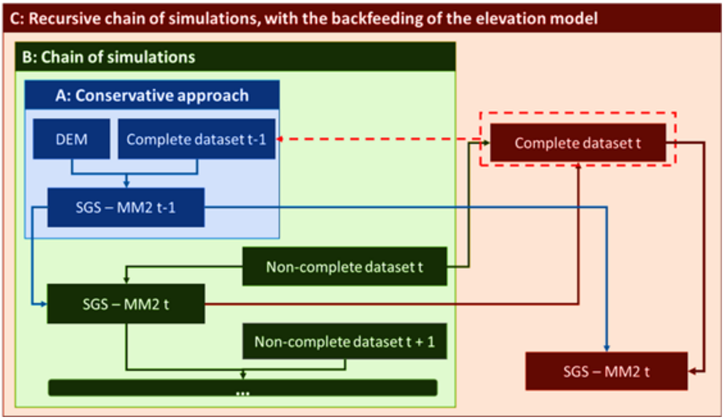

Fig. 3. Comparison of the three cosimulation schemes applied

The procedure developed under the title *"recursive chain of simulations with the backfeeding of the DEM"* reproduces both the spatial pattern of changes to water resources and the relevant initial statistical distributions properly.

In the case of the *"chain of simulations"*, the spatial pattern is becoming more and more obscure, therefore it is unsuitable for estimating resources. In the case of *"classical co-simulation"*, there is no temporal relationship between subsequent coverages, but the problem with it is the overemphasizing of relief contours in the mapping of changes to water resources.

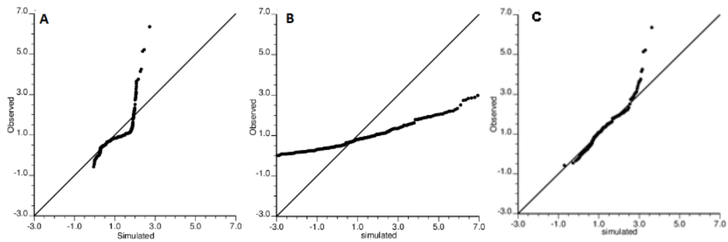

Fig. 4. Cross-validation of actual and simulated water change values in reliance on the DEM exclusively (A), as a chain of cosimulations (B) and in a recursive manner, with backfeeding DEM (C)

I used the procedure developed to estimate monthly groundwater resources in the Danube-Tisza Interfluve between 1950 and 2017. I drew conclusions as to the changes to water resources in space and time and the pace of the change to groundwater level based on the relative value of the individual grid node values obtained.

2.3.3. The approach based on spatially correlated time series

The majority of groundwater time series do not cover the entire period subject to the simulation. Recursive approaches only take the areal pattern of the neighbouring time instant, but, if data are missing for an extended period, the estimation regarding the part of space in question will be distorted. The time series pattern is interpreted as a trend and processed by means of mathematical functions by the approach based on spatially correlated time series (Dimitrakopoulos–Luo 1994, 1997, Kyriakidis 2001a, b, Ekstrom et al 2007).

I pointed out that trends generate three times less fitting errors, if detailed temporal auxiliary data are available (Fehér 2015b). In my thesis, I use as auxiliary data an artificial time series the values of which at the individual time instants were determined, following the standard-normal transformation of hydrographs, as the medians of the values for the time instants. (Satellite gravimetric measurements might be capable of generating similar time series in the future.)

I treated fitting parameters as spatially structured, regionalized variables, which makes the complexity of the trend function limited. Temporal parameters were orthogonalized by principal component analysis, which was followed by spatial estimation based on sequential

Gaussian simulation. The next step of the procedure was the stochastic simulation of the residual time series obtained as the error of the fitted trend models with consideration to locally varying temporal ranges. Finally, estimated groundwater level was derived as the sum of the back transformed trend and the residual components (Figure 5).

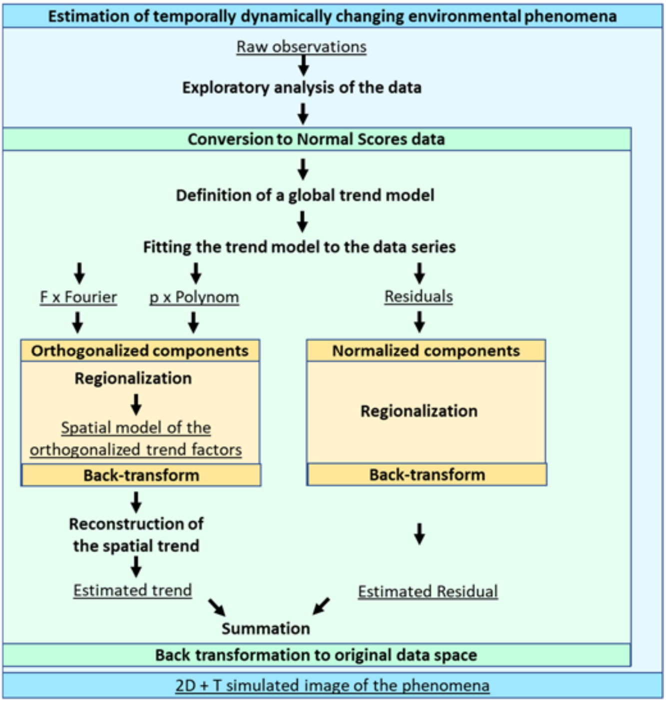

Fig. 5. Simulation algorithm of the spatially correlated time series

I applied the procedure developed to estimate the relative groundwater level of a daily resolution in the operational area of the Lower Tisza District Water Directorate and evaluated spatial and temporal changes to groundwater resources between 2005 and 20015 based on the results obtained.

The disadvantage to this procedure is its high computational cost, therefore I accelerated it by parallel processing of the simulation.

3. Results

3.1. Methodological developments and geostatistical results

1. I found that groundwater behaves as a regionalized variable when studied at a scale of 500–1000 m. I pointed out that, in the case of absolute groundwater levels, the effect of the elevation screened out the impact of groundwater resource changes on spatial continuity and statistical distribution. I revealed a relationship between the hydrometeorological factors and the variogram parameters of relative groundwater levels, which had been considered to be free of trends.

2. I demonstrated that the uneven spatial structure of the observation network and the inappropriate handling of data gaps might give rise to an incorrect mathematical-statistical interpretation of the findings, which could practically lead to an erroneous estimation of the change in water resources over time and across space. Moreover, there is a difference in the weight of representation of the water level values of the individual areas in the statistical sample, which in turn has an impact on the parameterising of geostatistical applications. I showed based on the missing data sensitivity of various widely used and lesser-known interpolators that it was the SKlm approach based on regional regression which produced the least data-point errors, it was not worth relying on it with regard to the lower rate of representation of ridge areas in the Danube-Tisza Interfluve. Simple cokriging is capable of managing the spatial heterogeneity of groundwater based on the local covariance relations between primary and auxiliary variables. With the application of collocated cokriging, the construction of the variogram model can be simplified on the one hand and, the overrepresentation of auxiliary data in estimations, which is typically the case with other classical bivariate geostatistical approaches, can be avoided on the other hand. In light of the foregoing considerations and cross-validation errors, the cokriging model of Markov 2 type seems to be most suitable for the interpolation of the elevation of the groundwater table above sea level.

3. I pointed out that groundwater is characterized by extreme spatial continuity, with gravity and the geological environment being the constant factors of its flow. The local relationship between precipitation and groundwater level does not change significantly on

a large spatial and temporal scale. However, groundwater level is influenced by small scale, spatially heterogeneous natural and anthropogenic effects locally. Considering all these aspects, I decided to apply sequential Gaussian simulation. *The procedure chosen allowed the incorporation of small-scale heterogeneities in the model. Furthermore, simulation also provides equiprobable alternative estimations for making decisions under uncertainty.* The sets of grid node values from the individual alternative estimations honour both the histogram of the data set observed and the semivariogram model considered.

4. I established based on the change of the confidence interval of the statistical distribution derived from the grid note values of the sequential Gaussian simulation that the estimation of 125 realisations was sufficient for determining the confidence interval in both areas subject to my study at a significance level of 95%.

5. I found based on the comparison of the results of the three different simulation structures that absolute groundwater levels could be estimated most effectively by the application of a recursive cosimulation structure, with the backfeeding of the DEM. In this case, the spatial results generated by the simulations are accurate in respect of both the actual water level and the change to water resources and the base point and grid node groundwater change values represent the appropriate statistical distributions. I theoretically pointed out that, in case the cosimulations are used with non-complete dataset, then: in the case of relative groundwater levels, where the spatial continuity is temporally varying, it was the *Markov 1 model* and in the case of the joint simulation of absolute groundwater levels and elevation, it was the *Markov 2 model* which could be applied.

6. I developed a sequential Gaussian simulation algorithm for relative groundwater levels, considering the parameters describing the shape of the time series as regionalized valuables. While approximating hydrographs with functions, I showed that there was a special joint theoretical time series pattern the use of which allows us a significantly better fitting of the function for the time series compared to the application of mere mathematical functions. The theoretical time series pattern mentioned above shows slight similarities to the shape obtainable from the gravimetric satellite data

series supplied by mission GRACE of NASA. However, low spatial and temporal resolutions and the frequent missing of data make the practical application of gravimetry-based modelling is quite problematic for the time being.

7. I prepared the daily resolution database of relative groundwater levels on the Southern Hungarian Great Plain based on the methodology of the spatiotemporally correlated time series. This was followed by the spatial optimization of the observation network of wells in the area. I considered for the optimization synthetic time series derived for the selected spatial coordinates from the above-mentioned database. In light of my findings, 9 properly selected new wells could ensure an estimation uncertainty of 150 cm on 95% of the study area.

3.2. Geographical consequences of groundwater changes

8. I compiled the database of areal and temporal estimates for ground water levels in the Danube-Tisza Interfluve for the period of 1950–2017 by the application of the recursive stochastic approach developed based on monthly median absolute groundwater levels. The spatiotemporal tendencies of groundwater resources and the rate of groundwater discharge were evaluated based on E-type estimations. I also calculated the mean inter-annual fluctuation of the groundwater level, as well as its variability according to elevation zones.

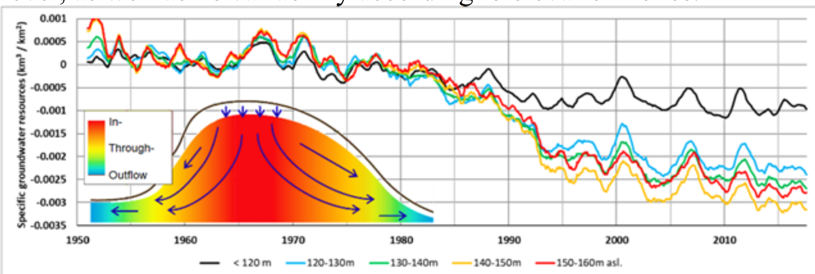

Fig. 6. The specific water resources over time by elevation zones on the Danube–Tisza Interfluve (1950-2017)

9. I produced based on daily resolution data a database containing the spatiotemporal estimation of relative groundwater resources for the operational area of the Lower Tisza District Water Directorate by the application of the spatially correlated time series approach. I made a probability-based estimate regarding the temporal change to water resources and its annual rate in the period of 2005–2015 by using the database. I also estimated the likelihood of a water level drop in the study area during the period under survey.

10.I revealed the role of underground runoff based on the estimation of groundwater resources by elevation zones. During drought years, groundwater resources discharge faster in higher areas. I also showed that in the areas where water could be replenished from nearby areas (e.g. the sediment cone of river Maros), the negative effects of climate change were less pronounced as long as there were periods of extremely high precipitation in between years of drought. Furthermore, I estimated the change to water resources in three additional areas of the Great Plain, which in turn made it possible for me to explore, by applying the evaluation method developed, the sensitivity of groundwater resources to climate change in areas with diverse geographical settings.

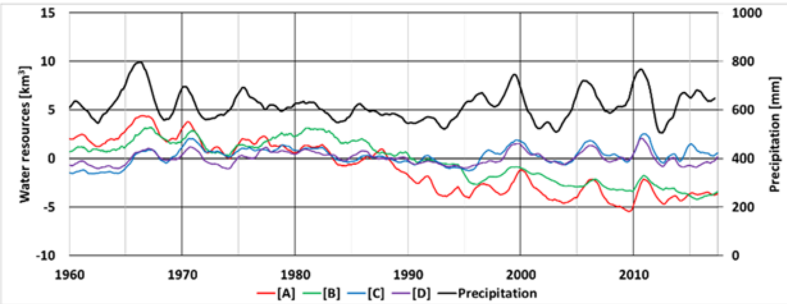

fig 7. Groundwater resources over time (1960 – 2017) on four areas of the Great Plain: Danube – Tisza Interfluve (A), Nyírség (B), Southern Transtisza (C) and the southern periphery of the North Hungarian Mountains (D)

Publications used in the dissertation

Fehér, Zs.–Rakonczai, J. 2012: Reliability enhancement of groundwater estimations. In: Malvic T. (ed.): *Conference book "Geomathematics as Geoscience": 4th Croatian–Hungarian and 15th Hungarian Geomathematical Congress*, Croatian Geological Society, Opatija. pp. 3–40.

Rakonczai J.–Deák J.–Ladányi Zs.–Fehér Zs. 2012: A klímaváltozás és tájváltozás kapcsolata alföldi mintaterületeken. In: Rakonczai, J.–Ladányi Zs.–Pál-Molnár E. (szerk.): Sokarcú klímaváltozás. GeoLitera, pp. 37-62.

Rakonczai J.–Deák J.–Ladányi Zs.–Fehér Zs. 2012: Indicators of the climate change in the landscape:: Investigation on the soil - groundwater - vegetation connection system in the Great Hungarian Plain. In: Rakonczai J.–Ladányi Zs. (szerk.) Review of climate change research program at the University of Szeged (2010-2012): A klímaváltozás kutatás válogatott eredményei a Szegedi Tudományegyetemen (2010-2012). pp. 41-58

Rakonczai J.–Fehér Zs. 2014: A talajvízváltozások új szempontú értékelése az Alföldön. In: Kóródi T.–Sansumné Molnár J.–Siskáné Szilasi B.–Dobos E. (szerk.): *A VII. Magyar Földrajzi Konferencia kiadványa*, pp. 491–501.

Fehér Zs. 2015a: Talajvízkészletek változásának geostatisztikai alapú elemzése. A rendelkezésre álló információk természete és feldolgozása. *Hidrológiai Közlöny*. 85 (2): 15–31.

Fehér Zs. 2015b: A Spatiotemporal Stochastic Framework Of Groundwater Fluctuation Analysis On The South-Eastern Part Of The Great Hungarian Plain. *J. Environ. Geography* 8: (3–4): pp. 41–52.

Rakonczai J.–Fehér Zs. 2015: A klímaváltozás szerepe az Alföld talajvízkészleteinek időbeli változásaiban. *Hidrológiai Közlöny*. 95 (1): pp. 1–15.

Fehér, Zs.–Rakonczai, J. 2019: Analysing the sensitivity of Hungarian landscapes based on shallow groundwater fluctuation affected by climate change. *Hungarian Geographical Bulletin.* (in press).

Other publications

Barton G.–Tobak Z.–Fehér Zs.–van Leeuwen, B.–Kovács F.–Mészáros M.–Dolinaj, D.–Szatmári J. 2013: Téradat infrastruktúra építése és webes térképi publikációs felület készítése a belvíz monitoring tevékenység támogatásához. In: Lóki J. (szerk.): *Az elmélet és a gyakorlat találkozása a térinformatikában*. Debrecen, pp. 89–96.

Farsang A.–Barta K.–van Leeuwen, B.–Fehér Z. 2017: Az öntözés talajtani feltételei, lehetőségei és körülményei az ATIVIZIG működési területén. *Tanulmány*. K&K Mérnöki Iroda, Szeged. 30p.

Kiss A.–Sümeghy Z.–Fehér Zs. 2008: A Maros 18. századi áradásai és egy jellemző téli árvizének területi hatásai. In: Füleky Gy. (szerk.) A táj változásai a Kárpát-medencében. Az erdélyi táj változásai.: A Sapientia Magyar Tudományegyetemen, Marosvásárhelyen 2008. július 2-4. között tartott XII. tájtörténeti tudományos konferencia kiadványa. SZIE, Gödöllő, pp. 94 – 100.

Mezősi G.–Sipos Gy.–Rakonczai J.–Fehér Zs.–Szatmári J. 2017: Vízrajzi hálózatfejlesztés – Felszínközeli (talaj)vizek – monitoringjának optimalizálása. *Kutatási jelentés,* 87 p.

Rakonczai J.–Fehér Zs. 2014: A vízhiány hatása a talajvízszintre. In: Blanka V.–Ladányi Zs. (szerk.) Aszály és vízgazdálkodás a Dél-Alföldön és a Vajdaságban SZTE TTIK. pp. 110-117.

Szatmári J.–Barta K.–Csépe Z.–Fehér Zs.–Miodrag B.–Obradovic D. – Dudarin Z.–Radonic V. 2014: A talajvíz megfigyelésének lehetőségei. In: Blanka V.–Ladányi Zs. (szerk.): Aszály és vízgazdálkodás a Dél-Alföldön és a Vajdaságban. SZTE – TFGT, Szeged. pp. 289-299.