Exploring Potentially Hazardous Areas for Water Quality Using Dynamic Factor Analysis

József Kovács[1], László Márkus[2], József Szalai[3], Márton Barcza[6], György Bernáth[1], Ilona Kovácsné Székely[4] and Gábor Halupka[5]
[1]*Eötvös Loránd University, Department of Physical and Applied Geology,*
[2]*Eötvös Loránd University, Department of Probability Theory and Statistics,*
[3]*VITUKI Environmental and Water Menagement Research Institute Non-Profit Ltd.,*
[4]*Budapest Business School, Institute of Methodology,*
[5]*Repét Ltd., Environmental Consulting Office,*
[6]*University of Szeged, Department of Mineralogy, Geochemistry and Petrology,*
Hungary

1. Introduction

Just like rivers, lakes and seas, subsurface drinking water resources are not safe from human impact and pollution. For sustainable development it is imperative to specify the areas where and the ways how these resources can be best protected. The aim of this study is to present a mathematical method for analysing spatially and temporally dependent data that is capable of assisting groundwater protection, by specifying potentially dangerous areas. In order to demonstrate the power of the proposed method the results obtained in three entirely different geological and hydrogeological environments are presented as case studies.

Natural and anthropogenic activities influence the fluctuation of groundwater level. Clearly, the main influencing factor is the precipitation recharging the subsurface water. Artificial influencing factors are the communal water withdrawals. If canals and rivers exist in the study area, they also control the natural random fluctuations of the groundwater table, they either feed or drain the shallow-groundwater depending on their water level. Considering the above-mentioned facts a mathematical model is formed on the basis of probability theory. Groundwater level data provided by monitoring wells are considered as time dependent random quantity, so hydrographs of each well are regarded as realizations of stochastic processes. Processes belonging to individual wells are not separate phenomena, but the occurrence of the same global natural phenomenon under different local conditions. That is why it is natural to treat these processes as the components of one multidimensional course, and as a matter of fact these components are probabilistically interdependent. It has to be emphasized that this interdependence is related to a spatial structure, and then the data is interpreted as the realization of a one-dimensional but *space-time* dependent stochastic process. We do not wish, however, to analyse the dynamics of this spatiotemporal process to a full extent. In our approach we suppose that the observed processes at different locations are governed by the same essential impacts, such as the mentioned recharge from precipitation, rivers and water withdrawal. It is the intensity of these impacts that depends on locality, creating the main aspect of the spatial dependence. So the goal is to identify these impacts, and to determine their spatially dependent intensities. In order to decompose the hydrograph time series into linear combination of the influencing effects – called factors - dynamic factor analysis has been applied. Dynamic factor analysis (DFA) produces factor time series and their loadings as an output; then the series must be identified with those processes of Nature that influence or drive the analysed phenomenon. The obtained loadings represent factor-intensities that provide essential information on the geological environment, improving the chance of correct decisions when environmental issues are on the agenda.

It has to be mentioned at this point that the whole factor analysis is scale invariant, and therefore, not absolute but only relative intensities can be inferred. These, however, are quite sufficient for our purposes, since we only wish to compare locations or subareas as less or more vulnerable to contamination. DFA readily provides the value of factor loadings belonging to a fixed factor and a fixed hydrograph that is to a fixed monitoring location. Since in our model the intensity is the sole source of dependence on location, the spatial aspect of interdependence can be revealed by the further analysis of the loadings. Consequently, if we wish to investigate a location where *no observation* is available, it can be done by spatially predicting the factor loadings and combining them with the factors; ultimately the hydrograph can be predicted. However, from hydrogeological point of view factor loadings and corresponding intensities are far more informative than the predicted hydrograph itself. At certain locations where no data are available, the loadings provides information about the expected intensity of infiltration, water-migration upward or downward, leakage effect or depression. Of course the uncertainty of the prediction has to be accounted for.

In the given applications the factors correspond to the infiltration, the water withdrawal and the river influence when relevant, hence the intensities appear to be connected to aquifer vulnerability. Intensive water withdrawal increases the danger of contamination of an aquifer, since overpumping may establish contact with a distant, already polluted storage. On the other hand high intensity infiltration increases vulnerability by helping local entry of any surface contaminant into the aquifer. Intense river impact may increase the danger of infiltrating contaminants carried by the river. Hence, the loadings, the measures of intensity can be regarded as *important quantitative markers of vulnerability* of the aquifer.

2. Dynamic factor analysis

Often the statistician encounters measurements of a very complex but well-observable time-dependent random phenomenon that is induced by only a few basic but unobservable (latent) effects of relatively simple dynamic structure. The behaviour of the measured phenomenon can be understood much better when these common driving forces are identified. In the conventional setup when independent observations are at hand factor analysis provides the variables representing the latent effects. As Anderson (1963) alerted, this technique can be misleading when applied to multidimensional time series with delayed interdependence among its components. Lagged interdependence invalidates the results of conventional factor analysis that has been elaborated for independent observations. Correct identification of the governing effects became possible with the invention of *Dynamic Factor Analysis (DFA),* that is capable to take into account the dynamic structure of both observations and factor time series.

The term dynamic factor analysis goes back to the pioneering work of Geweke (1977). Shortly after different concepts and procedures had become known as dynamic factor analysis, all of them being generalisations of this or that properties of the conventional factor model. For example Picci & Pinzoni (1986/1, 1986/2), Van Schuppen & Van Putten (1985) generalise the known property that observations are conditionally independent given the factors. Others, like Deistler & Scherrer (1991) decompose in every frequency the spectral density matrix function as sum of a diagonal and a singular matrix. So, by constructing factor spectral densities, they produce realisations of the factor time series on the base of the observed realisations. Gourieroux et al. (1995), Gourieroux & Monfort (1997), consider factor representations for Markov-processes. These as well as other early contributions to the literature on dynamic factor models like Sargent & Sims (1977), Engle & Watson (1981), Watson & Engle (1983), Connor & Korajczyk (1993) consider time series mostly with limited panel dimensions. Factor models for spatio-temporal processes became also a focus of interest, see e.g. Mardia et al. (1998) and Dryden et al. (2005).

The increasing availability of high-dimensional data sets has intensified the quest for computationally efficient estimation methods, leading to a renewed interest in dynamic factor analysis. DFA consolidated to state space representations of structural time series models. The new wave of literature was headed by Forni et al., (2000), Stock & Watson (2002) and Bai (2003). These methods are typically applied to high dimensional panels of time series. Exact maximum likelihood methods such as proposed in Watson & Engle (1983) have traditionally been dismissed as too computationally intensive. However Jungbacker & Koopman (2008) present new results that allow the application of exact maximum likelihood methods to large panels. Examples of recent papers employing likelihood-based methods for the analysis of dynamic factor models are Doz & Reichlin (2006), and Reis & Watson (2007). Alhtough the majority of publications for DFA is concerned with the problems of economics, environmental applications are also widespread (Márkus et al., 1999; Kovács et al., 2004; Ritter et al., 2007), a few of them assesses water quality (Kovács et al., 2002; Kovács et al., 2004; Muñoz-Carpena et al., 2005; Ritter et al., 2006).

The model, we briefly describe in the present paper, does not differ much of the ones based on state space representation, but the algorithmic solution lays on different concept. The idea, as we use, originates in Bánkövi et al. (1979), though it also relates to Box & Tiao (1977) on canonical transformations of time series vectors. While prescribing an autoregressive structure to the factor time series our factor analysis model minimises a cost function, which is a linear combination of the conditional variance of the prediction error and the state estimation error. The problem of finding the optimal factors leads to a minimisation problem on Stiefel manifolds much in the focus of recent investigations in operational research (see e.g. Rapcsák, 2002). The theoretical solutions are very complex and difficult to give, the explicit solutions hardly go as far as 10 dimensions, falling far short to our case. So instead, this optimisation problem can be solved by an iterative method, which relies on the maximization algorithm of sums of heterogeneous quadratic forms developed in Bolla et al. (1998).

Let's consider the usual static factor model equation

$$Y = \boldsymbol{A} \cdot F + \varepsilon \tag{1}$$

expressing that the observations Y are described by linear combinations of several latent factors F plus a random uncorrelated noise ε. Usually the number of observed variables *N* is significantly higher than that of the factors, *M*. The crucial difference when dynamic factor models are considered is that both observations and factors are empirical time series instead of independent samples/measurements of the variables, as is in ordinary models. To complete the model the dynamic structure of the factors has to be specified. The linear transformation expressed through the **A** matrix, however, should not depend on time. Supposing the observed N-dimensional time series

$$Y(t) = \left(Y_1(t), \ldots, Y_N(t)\right)', \quad 0 \le \mathrm{t} \le \mathrm{T}.$$

to be weakly stationary apart from a possible linear trend, and emphasising time dependence rewrite (1) as

$$Y(t) = \mathbf{A} \cdot F(t) + \varepsilon(t) \tag{2}$$

with the **N** x **M** matrix **A**, the factor time series vector

$$F(t) = \left(F_1(t), \ldots, F_M(t)\right)', \quad 0 \le \mathrm{t} \le \mathrm{T},$$

of M *uncorrelated, stationary* time series, and the N-dimensional Gaussian white noise

$$\varepsilon(t) = \left(\varepsilon_1(t), \ldots, \varepsilon_N(t)\right)', \quad 0 \le \mathrm{t} \le \mathrm{T}.$$

We are aimed at finding optimal, in a certain sense, *estimations* of the factors:

$$\hat{F}(t) = \left(\hat{F}_1(t), \ldots, \hat{F}_M(t)\right)'$$

The estimation of our model should focus on the following three natural requirements:

i. The estimation of the factors should be a *time-independent* homogeneous linear transformation of the observations.

$$\hat{F}(t) = \mathbf{B} \cdot Y(t) \tag{3}$$

ii. The factor time series components $F_j(t)$ should be linearly well predictable from their past. It is certainly fulfilled supposing them to be *autoregressive* processes of order L_j with a constant included in the autoregression:

$$F_j(t) = c_{j,0} + \sum_{k=1}^{L_j} c_{j,k} \cdot F_j(t-k) + \delta_j(t) \tag{4}$$

where the components of the Gaussian white noise $\delta(t) = \left(\delta_1(t), \ldots, \delta_M(t)\right)'$ are independent from each other and from $\varepsilon(\mathrm{t})$.

iii. Again, time-independent linear transformation of the factors should provide a "good" estimation - called factor-estimator - of $\mathrm{Y}(t)$, as expressed by equation

$$\hat{Y}(t) = \mathbf{D} \cdot \hat{F}(t) \tag{5}$$

where the **D** matrix is in fact the estimation of **A**.

The choice of autoregression in (ii) justified not only by its simple dynamic structure, but also by the fact that hydrographs of monitoring wells can reliably be modelled by autoregressive processes. In the process of model estimation it is supposed that the structure prescribed for the *unobservable factors* is inherited to its *estimations*. Were the components of $(\hat{F}_1(t),...,\hat{F}_M(t))$ observable, their *best forecast* $\widetilde{F}_j(t)$ could be obtained as

$$\widetilde{F}_j(t) = c_{j,0} + \sum_{k=1}^{L_j} c_{j,k} \cdot \hat{F}_j(t-k) \tag{6}$$

In order to relate $\widetilde{F}_j(t)$ to $F_j(t)$ we call $\widetilde{F}_j(t)$ the *empirical best forecast* of $F_j(t)$. In other words it is just the plug in of the predicted factors into the best forecast of the autoregression. As the true values are not known, the coefficients $c_{j,k}$ have to be estimated. The optimality of this forecast, guaranteed for a truly autoregressive process with known coefficients only, cannot in general be preserved for the plug in. Keeping this in mind we will use (6) for the forecast of the estimator $\hat{F}(t)$ given its past. Since the observations and thus the predictions of the factors can be computed for all $\mathrm{t}, 0 \le \mathrm{t} \le \mathrm{T}$, it is possible to compare the forecast with the estimator itself, and by centring, get an unbiased estimation $\hat{\delta}(t)$ of the noise $\delta_j(t)$ in (4) as

$$\hat{\delta}_j(t) = \widetilde{F}_j(t) - \hat{F}_j(t) - \left[\overline{\widetilde{F}_j - \hat{F}_j}\right]$$

(For any $X(t)$ $\overline{X}$ denotes the average

$$\frac{1}{T+1}\sum_{t=0}^{T} x(t))$$

The squared sum $\varepsilon^{(d)}$ of $\hat{\delta}_j(t)$ is called the estimated *dynamic error*:

$$\varepsilon^{(d)} = \sum_{j=1}^{M}\sum_{t=L_j}^{T} \hat{\delta}_j(t)^2$$

Similarly, $\hat{Y}(t)$, the factor-estimator of observations in (5), opens the way to estimate $\varepsilon(t)$, the noise in (2) by taking the centered difference as

$$\hat{\varepsilon}_i(t) = Y_i(t) - \hat{Y}_i(t) - \left[\overline{Y_i - \hat{Y}_i}\right],$$

the squared sum $\varepsilon^{(s)}$ of which is called the predicted *static error:*

$$\varepsilon^{(s)} = \sum_{i=1}^{N}\sum_{t=0}^{T} \hat{\varepsilon}_i(t)^2$$

If the importance of this or that observation is to be emphasized, or the precision of the forecast of this or that factor is of major concern then *weights* can be introduced in the definition of both dynamic and static errors, to achieve this end. To fulfil the requirement given in (iii), the estimation of the model is regarded to be "good" if the sum of the estimated static and dynamic errors is minimal. This means the minimization of the following functional:

$$\Psi(T) = \varepsilon^{(s)} + \varepsilon^{(d)} = \sum_{i=1}^{N}\sum_{t=0}^{T}\hat{\varepsilon}_i(t)^2 + \sum_{j=1}^{M}\sum_{t=0}^{T}\hat{\delta}_j(t)^2 \tag{7}$$

on the constraints,

$$\mathrm{var}(\hat{\mathrm{F}}) = I_M \tag{8}$$

stemming from the uncorrelatedness of the factors (I_M denotes the M×M unit matrix).

The real statistical difficulty lays in the estimation of the model parameters, that is the matrices **B**, **C**, **D**. Remark, that **C** is the matrix of $c_{j,k}$-s from (4) endowed with zeros when necessary.

The usual ML methodology results in very complicated computations, and even though one can determine the density function, but it seems rather hopeless to find its place of global maximum. In state space models the EM algorithm provides a way to tackle the estimation problem. Alternatively, Markov Chain Monte Carlo (MCMC) estimation may also prove to be viable.

Our approach originates in Bánkövi et al. (1979), where instead of finding a direct optimal solution to (7) and (8) an iterative approximation by a criss-cross algorithm is suggested. It can be developed further by using the optimisation procedure of heterogeneous quadratic forms as described in Bolla et al. (1998), where the analysis of the optimisation can also be found. For a detailed description see Ziermann and Michaletzky (1995) or Márkus et al. (1999).

Introducing a new appropriate orthonormal system $\{\mathbf{e}_j\}_{j=1,\ldots,M}$ as in Márkus et al.(1999), the functional, rewritten as

$$\Psi = \sum_{j=1}^{M} \mathbf{e}_j^T \mathbf{Q}_j \mathbf{e}_j$$

with the $\mathbf{Q}_j$ matrices computable from the observations and the **C** matrix, has exactly the same structure - that is the sum of heterogeneous quadratic forms - as the one treated in Bolla et al. (1998). By applying Lagrange's multipliers it is easy to obtain a necessary condition for the existence of stationary point, namely that equation

$$[\mathbf{Q}_1\mathbf{e}_1,\ldots,\mathbf{Q}_M\mathbf{e}_M] = [\mathbf{e}_1,\ldots,\mathbf{e}_M]\mathbf{S}$$

must hold with **S** being a symmetric M×M matrix. Specifically the vectors $\mathbf{Q}_j\mathbf{e}_j$ are included in the subspace spanned by $\mathbf{e}_1,\ldots,\mathbf{e}_M$.

For the place of global maximum **S**≥0 holds. Introducing the notations $[\mathbf{e}_1,\ldots,\mathbf{e}_M] = \mathbf{E}$, $[\mathbf{Q}_1\mathbf{e}_1,\ldots,\mathbf{Q}_M\mathbf{e}_M] = \mathbf{Q}(\mathbf{E})$ and writing the condition formally, we have

$$\mathbf{Q}(\mathbf{E})=\mathbf{E}\mathbf{S}, \qquad \mathbf{E}^T\mathbf{E}=\mathbf{I}_M, \mathbf{S}\geq 0,$$

which is nothing, but the polar decomposition of the matrix **Q**(**E**). A one to one correspondence of the polar, and the singular value decompositions is established in Bolla et al. (1998) and a criterion for **S**≥0 i.e. **S** being positive semidefinit is also given there. So, the actual computations rely on the singular value decomposition instead of the polar one. It seems that this is the computationally most demanding step of the algorithm, and sometimes it does not converge fast enough.

Summarising, the algorithm is as follows. For a given $\mathbf{E}_1 = \mathbf{E} = [\mathbf{e}_1,...,\mathbf{e}_M]$ vector system the next one $\mathbf{E}_2 = \mathbf{H} = [\mathbf{h}_1,...,\mathbf{h}_M]$ is defined from the polar decomposition of **Q**(**E**) by

$$\mathbf{Q}(\mathbf{E})=\mathbf{H}\mathbf{S}, \qquad \mathbf{H}^T\mathbf{H}=\mathbf{I}_M, \qquad \mathbf{S}\geq 0,$$

It can be shown - as is in Bolla et al. (1998) - that the algorithm increases the value of the Ψ functional, and the $\mathbf{E}_1$, $\mathbf{E}_2$,... matrices are getting closer and closer to each other, and so the accumulation points of the algorithm are their stationary points as well. The set of accumulation points is a connected one. If any point obtained from the algorithm falls "near" to the place of global maximum then the algorithm will be *convergent*, and it will converge to this *place of global maximum*. However the algorithm is not necessarily globally convergent, because in general there may exist fixpoints of it different from the place of global maximum. Every stationary point of the functional is fixpoint of it at the same time.

After the detailed presentation of the mathematical method we turn to the case studies of hydrogeological applications. Although all of the presented applications study groundwater levels, the data are registered in three different parts of Hungary (see Fig. 1), and therefore the geological and hydrogeological environment is entirely different.

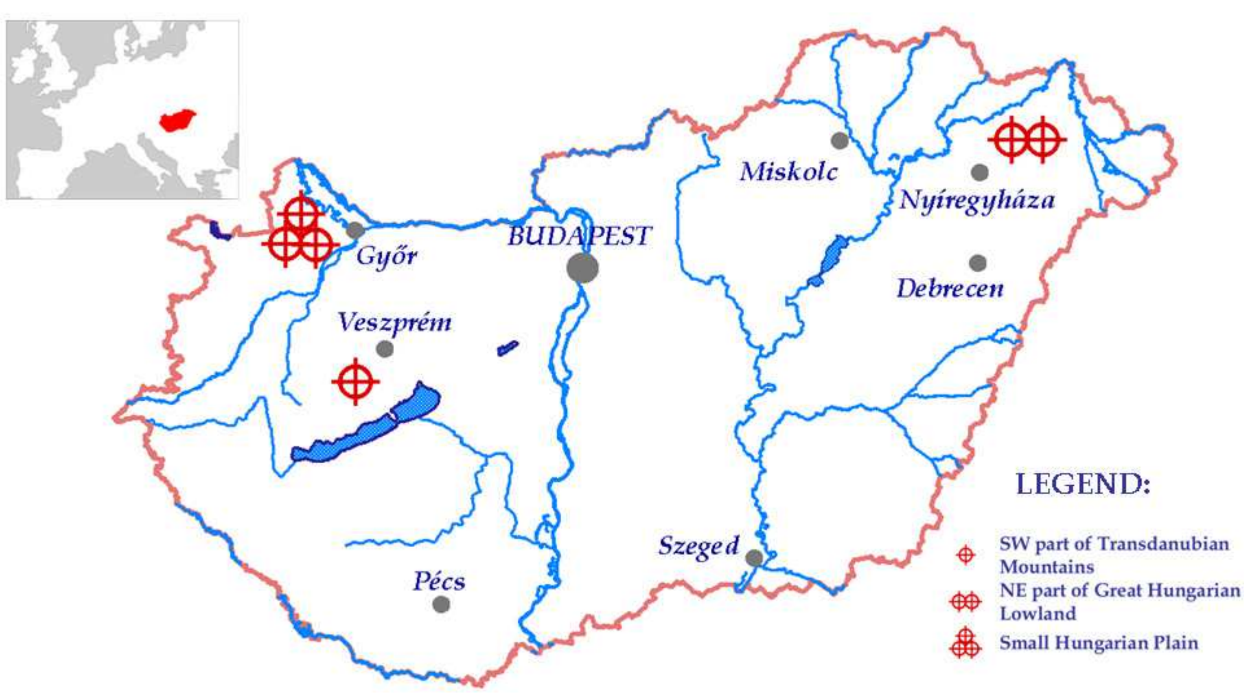


Fig. 1. The three different areas of Hungary, subject to case studies

3. Application of dynamic factor analysis in karstic environment

In Western Hungary the mining industry withdrew a large amount of water from the ground in the period of 1950-1990, when intense bauxite and coal mining characterised the Transdanubian region. In the south-western side of the Transdanubian Mountains (Bakony, Keszthelyi Mountains and Balaton-Highland) especially large quantity of water was pumped out from the main Triassic karst aquifer, the water conducting geological layer consisting mainly of Triassic dolomite and limestone. Both Bakony and Keszthelyi Mountains are built up mainly from this Triassic dolomite and limestone. In the Balaton-Highland a more various geological structure can be found, the Triassic aquifer consists mainly of middle and lower Triassic formations divided by aquitard layers. The Bakony and the Balaton-Highland are separated by a deep rift (Lesence-valley) from the Keszthelyi Mountains. This rift is filled with impermeable formations, and those hold back or slow down considerable part of the water flow. A reverse fault separates the Balaton Highland from the Bakony, and this creates an impermeable obstacle for flows between them (Fig. 2).

At its peak the withdrawal continuously exceeded the infiltrated water supply, and as a consequence water capacity of some springs declined to the half or one third and some others dried out completely. As the deterioration of the environment became more and more apparent, a monitoring system has been set up to observe karstwater levels. The system was extended over time and observations became more frequent. For our analysis it provided the *hydrographs*, the water levels registered in karstwater monitoring wells in the south-western side of the Transdanubian Mountains (Bakony, Keszthelyi Mountains and Balaton-Highland).

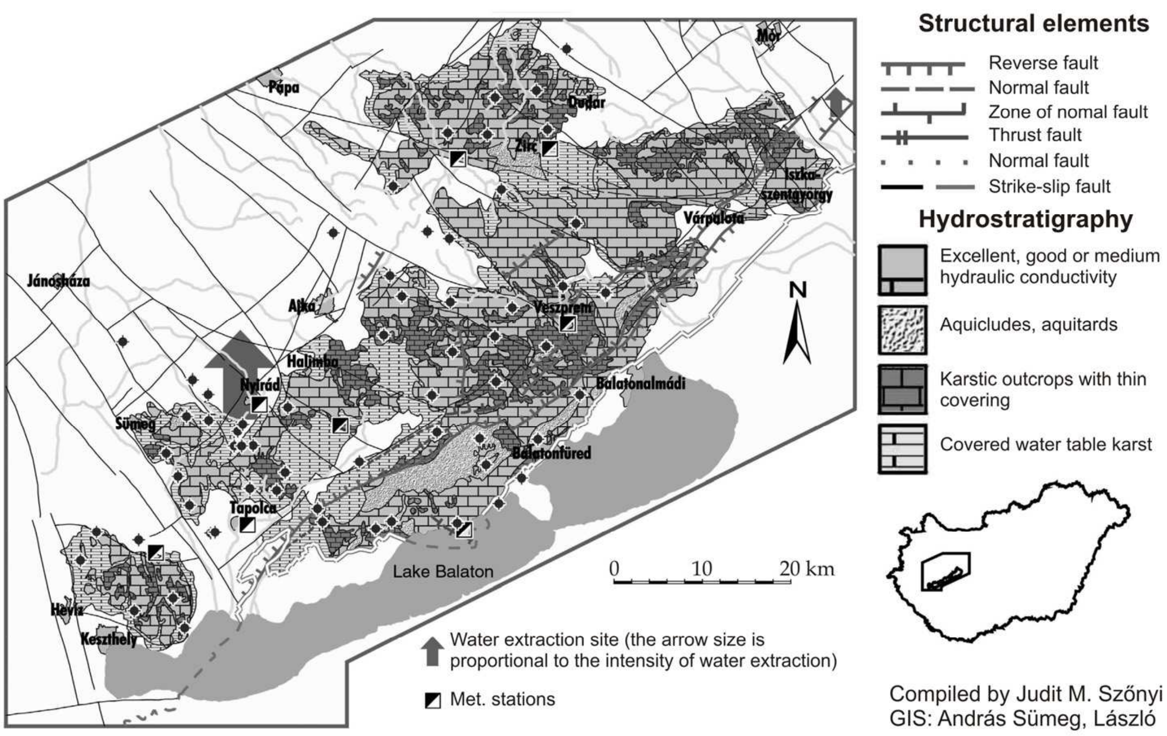


Fig. 2. Structural elements, hydrostratigraphy, water extraction and karstwater monitoring wells in the Transdanubian Mountanis (compiled by Mádlné, Szőnyi J. , GIS: L. Füle, 1998)

A hydrograph of a (monitoring) well reflects, on the one hand, all characteristics of the route of water from the surface to the observation point, and on the other hand, any kind of change of geology along this route. So, information on geological conditions can be obtained *after* water has reached the aquifer, moreover, the basis of this information-providing is just the water that completed its migration. This allows for the characterisation of the aquifer "bypassing" the direct study of uncertain geological parameters, since every influencing effect of the geological setting is already left behind by the water: the *final result* of roaming of water is reflected in the fluctuating level in the monitoring well. Consequently, ours is an "upside down" approach: the influence of geological parameters can indirectly be inferred by studying the hydrographs.

The individual hydrographs can be regarded as temporal measurements or samples from the above mentioned aggregated phenomenon, the fluctuating karstwater level. Only 64 monitoring sites provided sufficiently regular readings and even in the period of 1970-1990. The scattered areal distribution of the selected 64 monitoring wells can be seen on Fig. 2. The sporadic timing of observations forced us to study hydrogeological time series of yearly averages, derived from the non time-equidistant raw data. Observations after 1990, though available, are not taken into consideration in this study, since the social and economic changes taking place at that time in Hungary triggered sharp decline in mining and mounting concern towards environmental damages. This led in particular to the closure of the mines at the main water withdrawal site. Since recovery processes of nature are usually different of human interference, new changes were induced in the karstwater reservoir system. Therefore, it seems to be reasonable to exclude this time period from the present investigations.

We applied dynamic factor analysis (as described above) to the data in order to decompose the observed empirical time series (the hydrographs) into linear combinations of several underlying factors. Since there was no reason to emphasise the behaviour of any particular monitoring well, or factor, we used equal unit weights in the static and dynamic error terms.

On the different possibilities to determine the number of factors as well as the order of autoregressions we refer the reader to Márkus & Kovács (2003). In our case three factors proved to be sufficient to describe the behaviour of hydrographs by satisfactory precision, and among the concurrent models the AR(1) - AR(1) - AR(1) proved to be the easiest and most natural to be interpreted. The three effects, that is the corresponding three dynamic factors explain about 92% of the overall variance. It has to be noted though that generalization of this customary characterization of ordinary factor analysis for the dynamic case is not quite straightforward. We were able to identify the first two factors only, and conjecture that the third is describing a small territorial variance in infiltration.

The immediate aim of the application of dynamic factor analysis was to *identify the main underlying effects* shaping the hydrographs, and give spatial prediction of the intensity of these effects. The interpretation of the obtained factors reiterate the well known fact that infiltration and water withdrawal (where relevant) are the two main effects shaping the hydrographs. The more valuable gain from the analysis is that the loadings of the dynamic factors serve as measures of the intensity of the effects at a given location and thus endowing a certain area with a comprehensive hydrogeological characteristic.

Remark here, that throughout the considered period monitoring wells near the main mines registered about 100 meters drop in karstwater level. Even in wells further away from the main site decline could exceed 20 meters. The decline was not uniform, its pace changed with time. In the area of study there was one main pumping centre at Nyírád, near several closely located bauxite mines, and some other, much smaller water withdrawal sites near coal and bauxite mines, furthermore, water has also been withdrawn for community use (Fig. 2). The withdrawn amount of water at these latter sites combined, was of considerably smaller magnitude than that of the main site at Nyírád. For this reason it seems to be sufficient to use the data of the dominating Nyírád withdrawal in the analysis, noting that local withdrawals can still play an important role regionally, but the use of these data does not improve significantly the identification of the factor.

However, the water withdrawal data cannot be used without transformations for comparisons with the decline in water levels. On the one hand, water withdrawal is partly balanced by the decrease in water capacity of springs, and that has to be compensated for in comparisons. On the other hand the amount, dynamically supplied to the aquifer, has to be subtracted, because only the remaining amount lowers the static water reserve of the aquifer. The detailes of the actual computations can be found in Márkus et. al., 1999, Kovács, 2007.

The result is surprisingly spectacular. In Fig. 3 we compare the graph of the differences of the first factor (blue line) with the sum of the registered water withdrawal and the spring capacities (red and green lines). In order to present these data on the same graph one has to bring them to the same scale; the simplest way - as it is done in Fig 3 - is to standardise them. The red and green lines differ in the amount of withdrawn water, the red includes the Nyírád withdrawal only, whereas the green takes into account all the registered withdrawals in the area.

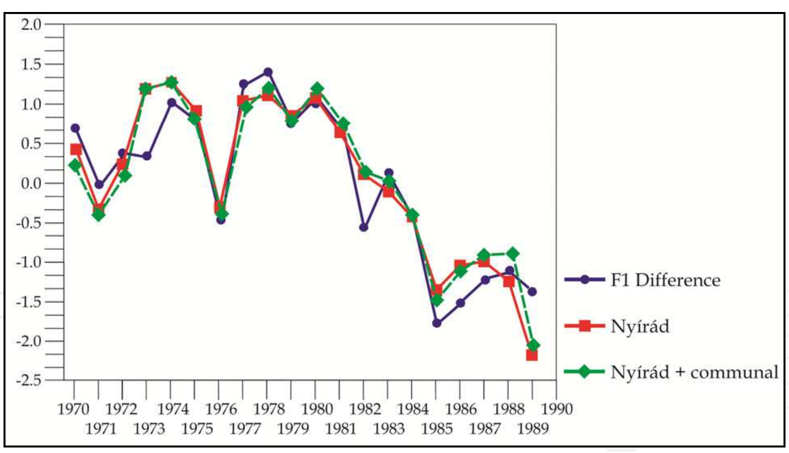


Fig. 3. Comparison of the differenced 1st factor (blue) with the rescaled water withdrawal at Nyírád (red) and Nyírád plus communal (green) withdrawals

Second dynamic factor represent the overall tendencies in infiltration throughout the whole investigated area. In order to prove it, infiltration has to be computed by conventional methods.

Infiltration is a very complex process, because at a given location the amount of infiltrating water depends on geomorphology, water conductivity of superficial rocks, air temperature, amount, duration and physical state of precipitation, and also vegetation and many other factors. Therefore, the effect of infiltration varies throughout the mountains. Nevertheless, the area is not so large that the main tendencies in the yearly amount of infiltrated water would differ too heavily among locations. This circumstance allows us to identify only one factor (the second) with infiltration and obtain comparable weights (intensities) of the infiltration effect in every monitoring well.

Different empirical methods exist for the computation of the infiltration. When used in the analysis, infiltrations are computed by methods developed or refined on the basis of observations in Hungary, rather then elsewhere, (cf. e.g. Primault (1963), Dracos (1980), Burman and Pochop (1994), Jeannin & Grasso (1995) and references therein), because karstic infiltration strongly depends on local characteristics. For comparisons we used four methods, called by the names of their creators as Böcker, Kessler, Maucha and Morton.

The method of Kessler (1954) is based on several decades long observations in the Mecsek Mountains. Böcker's method (1974) was evaluated at several different locations. Maucha (1990) computes infiltration on the basis of observations in the northern part of Hungary, where the karst knowingly has different characteristics (e.g. water conductivity) than the one we study. Using Morton's evapotranspiration model (Morton, 1983) Csepregi (1995) developed a way to calculate infiltration on monthly basis. As it was pointed out in Márkus et al. (1999), the different methods produce pretty different results - especially so in temporal dynamics - correlating only 0.5-0.7 among themselves. All the methods mentioned, compute infiltrations on the basis of precipitation data, and do not depend on location, or local geological structure. On the contrary to this, the factor representing infiltration was computed from the hydrograph - that is from the groundwater level data.

To identify the second factor with the infiltration, it has to be compared to the computed conventional ones. We got highly significant correlations (around 0.8) among the factor and the conventionally computed infiltrations, for details see Márkus et al., 1999; Kovács, 2007. The corresponding graph (Fig. 4) emphasizes the fit further.

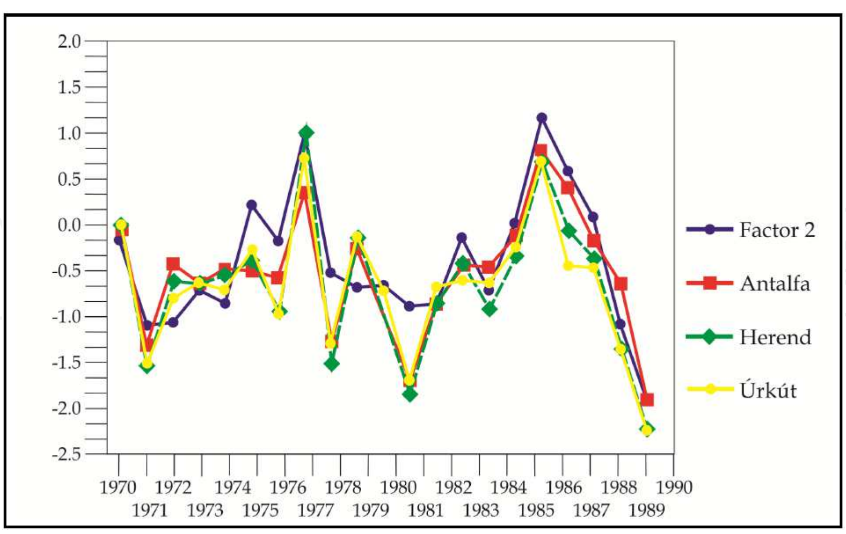


Fig. 4. Graphs of factor 2 and standardised Kessler's infiltration computed from the data of three different meteorological stations

The dynamic factor loadings reflect the proportion of the underlying effects in influencing the fluctuation of the hydrographs. When these loadings indicate a dominant role among the underlying effects either for withdrawal or infiltration then important conclusions can be drawn for the aquifer vulnerability or sensitivity for contamination.

We would like to present a relationship between the above-mentioned factors and vulnerability, in two steps. First we describe a connection between the factors and geological parameters influencing hydrographs. Afterwards, factors are considered as influencing effects in vulnerability of an aquifer by means of a different scale of groundwater flow regimes.

To reach this goal it is essential to define vulnerability. The following few approaches can help to see essence of the concept despite the lack of a widely accepted definition.

* "*Vulnerability is an intrinsic property of a groundwater system. It depends both on the geological and hydrogeological characteristics of an area, and on the sensitivity of groundwater system to human and natural impacts.*" *(COST Action 65 Final Report, 1995)*
* "*Vulnerability as a system-approached term means the ability of an environmental system and it surroundings to compensate human activities.*" *(Mádl-Szőnyi & Füle, 1998)*
* "*Intrinsic vulnerability is the term used to define the vulnerability of groundwater to contaminants generated by human activities. It takes account of the inherent geological, hydrological and hydrogeological characteristics of an area, but is independent of the nature of the contaminants.*" *(Daly et al., 2001)*

Briefly: *sensitivity* expresses a chance of contamination of an aquifer, whereas *vulnerability* is like a summary of conditions (human and/or other factors), that can activate the risk of any geological settings into real danger. Nevertheless it is important to take into consideration that sensitivity refers to an aquifer as rock-type, while vulnerability concerns a complete groundwater providing system, as Alföldi (1994) pointed out. Our opinion is that the presented results below refer to vulnerability, since water withdrawal is a kind of human activity and infiltration reflects the existing climatical, and other meteorological conditions. So neither of them is part of intrinsic attributes of an aquifer.

Mathematical study of the time series have disclosed two factors which account for more than 90% of fluctuations of well-hydrographs. One of them is *water withdrawal*, the other one is *infiltration*. How to connect these factors to geological parameters and what kind of parameters are these?

1. Water withdrawal. It is a kind of anthropogenic activity, non-geological parameter. Nevertheless it provides indirectly information about hydraulic conductivity, karstic network, and material of the aquifer.
2. Infiltration. It is itself an important characteristic of an aquifer, on the other hand it has relationships with precipitation, features of cover material (soil, as well), and it is in inverse ratio to runoff.

If the referred parameters are considered, it can be seen that all of them are key parameters in one or more vulnerability-assessment methods. To show this table 1. summarizes 6 very remarkable methods, in point of view of the used parameters.

It can be claimed that a close connection exists among the identified factors and the geological key parameters of vulnerability-assessment methods. Consequently factors (and so the mathematical method, as well) should have an indirect relationship to vulnerability, itself. The next point assesses the question, how.

Parameter / Method	1	2	3	4	5	6
Infiltration, recharge and/or its type	*	*	*	*	*	*
Precipitation, climate						*
Epikarst (thickness, quality)			*			
Deepness of groundwater level (thickness of unsaturated zone)	*	*				
Run-off				*		*
Hydraulic conductivity	*	*		*	*	
Karstic network			*			*
Material of unsaturated zone (thickness, quality – [no epikarst specifically])	*	*		*	*	*
Topography	*	*				
Protective cover material, soil (thickness, quality)	*	*	*	*		*
Material of the aquifer (thickness, quality)	*	*				

Table 1. Parameters of vulnerability assessment methods. The numbers refer to the following vulnerability-assessment methods: 1.) *DRASTIC* (Aller et al., 1985), 2.) *SINTACS* (Civita & Regibus, 1995), 3.) *EPIK* (Doerfliger, 1996), 4.) *PI* (Goldscheider et al., 2000), *Irish method* (Daly & Drew, 1998), 6.) *European Approach* (Daly et al., 2001)

One of the above mentioned hydrograph-determining factors is *water withdrawal*. Its effect is indirect. If an aquifer is under over-pressure, it means that it is protected against the entry of contaminant. The effect of extensive withdrawal decreases the resistance, so entry of pollutant may occur by either vertical (local) or horizontal (regional) recharge – as Mádl-Szőnyi pointed out (1997). Consequently, vulnerability of the aquifer increases in both local and regional scales. However, pressure-fluctuation spreads throughout an area faster than water or contaminant, decreasing of overpressure by water withdrawal in large, closed or quasi-closed aquifers does not mean immediately increasing vulnerability (Alföldi, 1994). However our opinion is that it is essential to take into consideration role of *time*. Dynamic factor analysis used data of 21 years, a period that seems long enough to assume that the process of decreasing pressure definitely has negative effect on vulnerability. When a distant aquifer is contaminated and the available time is long enough for spreading of the contaminant a direct connection can be established between extensive water withdrawal and increasing regional vulnerability. On the other hand increasing water withdrawal and local vulnerability can only be connected indirectly, because in a period when there is no significant precipitation to induce remarkable infiltration the medium carrying any soluble contaminant is missing. So pollution will not immediately follow the decrease or disappearance of overpressure protection of the aquifer.

The other determining factor is *infiltration*. In areas of high intensity infiltration a contaminant (general, soluble type is assumed) can enter in great quantities into the aquifer, directly causing high local vulnerability. At the same time infiltration can generate an increase in regional vulnerability when the investigated area has a regional recharge function, too (ignoring here dilution, degradation of the contaminant). If the investigated area is situated in a non-regional recharge area, fluctuations in hydrographs refer to local vulnerability.

Summarising, it can be stated that the factor of *water withdrawal* has direct connection with regional, but indirect connection with local vulnerability, while factor of *infiltration* directly refers to local and indirectly to regional vulnerability, depending on which hydraulic regime does the pilot area belong to. Observing, that in withdrawal-dominated wells infiltration-generated fluctuations are covered up by the effect of water-withdrawal, *regional vulnerability can be described by the factor of water-withdrawal*. When infiltration-dominated wells are in non-regional recharge areas, *local vulnerability can be characterized by the factor of infiltration.* Both statements concern the whole aquifer as the subject of vulnerability assessment.

As it could be read above, we have revealed a connection between vulnerability and factors by dynamic factor analysis, approaching both from basic geological parameters used in vulnerability assessments and from groundwater flow regimes in different scales. So a new, exact description of vulnerability of an aquifer can arise from this mathematical method.

4. A case study in an alluvial environment

The next case study concerns shallow-groundwater in alluvial environment, where the influence of surface waters is negligible. The study presents the results of dynamic factor analysis on the shallow-groundwater level time series of monitoring wells located in the Nyírség and Hajdúság area, which is a somewhat higher elevated plain in the northeastern part of the Great Hungarian Plain. The studied area lays between and around two major cities of Hungary, Debrecen and Nyíregyháza (with a population of 200,000 and 120,000 people), and it is surrounded by other plains of lower altitude. The hydrology of the studied area is quite complex. Most of the higher elevated Nyírség is a recharge area, while in the surrounding lower elevated areas features of discharge indicated for example by the saline lakes and soil.

The Nyírség and its vicinity was part of the Pannonian Sea that separated from the Central-Paratethys about 5.4 million years ago, and was slowly filled by clastic sediments. Later, due to the Pliocene erosion and tectonic events during Pliocene/Pleistocene considerable parts of the area lost sedimentation. As a result of vertical plate movements some basin areas, including the studied one had subsided 150-700 meters while the surrounding areas ascended forming the Carpathian Mountains. During Pleistocene, those basins were filled by coarse sediments (gravel, coarse sand) from northeast and northwest (Urbancsek, 1965).

The layers of alluvial sediments wedge out suddenly, they have lenticular appearance, and toward south both frequency and thickness of their occurrence increases.

During middle Pleistocene, tectonical activity lessened, so the amount of sediment and its grain size were decreasing, clay intercalations became more frequent. During late Pleistocene, tectonical activity intensified again, and average grain size of the alluvium increased. During Holocene, the typical sediment of the study area was sand.

The Pliocene sediment sequence contains thermal water with high salt and gas contents, so it is not suitable for drinking water. The lower Pleistocene sequence is the most favourable drinking water aquifer from both qualitative and quantitative point of view; it contains water suitable for communal use.

Middle Pleistocene sediments are poorer aquifers, though the drinking water supply of the largest city of the area Debrecen is supplied partially from the sand intercalations of this sequence. The hydraulic conductivity of the upper Pleistocene sediments is lower than that of lower Pleistocene. Shallow aquifers should also be mentioned though those are not thick enough for serving considerable demands (Fig. 5, Virág et al., 2005).

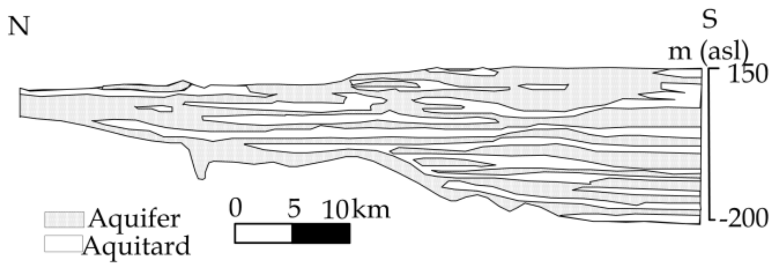


Fig. 5. Geological cross-section of Pleistocene sequence in the studied area (based onVirág et al., 2005)

Strata of the studied area suddenly wedge out. These have lenticular appearance. As a hydrogeological consequence there can be significant communication between the aquifers that is very complex in both levels and directions (Marton & Szanyi, 1997a, 1997b). Regional and local flow systems of the area were explored in earlier studies. Water age measurements data proved that the Nyírség region is a recharge area and the recharge intensity may reach 8 to 16 mm/year. On the other hand the margins of the area are the discharge zones of the flow system (Fig. 6 & 7) (Deák, 1979; Marton, 1981; Deák et al., 1987; Marton & Szanyi, 2000; Székely, 2003, 2006).

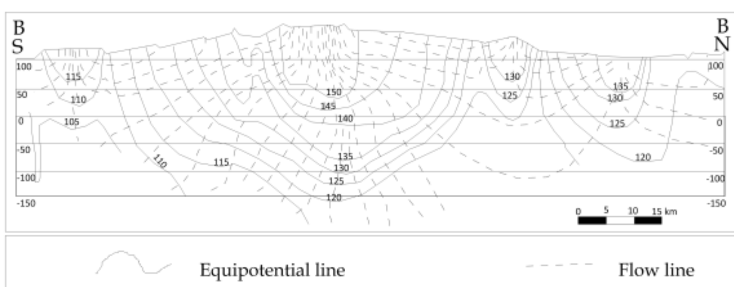


Fig. 6. Hydrological cross section (north to south) of Nyírség (Marton, 1981)

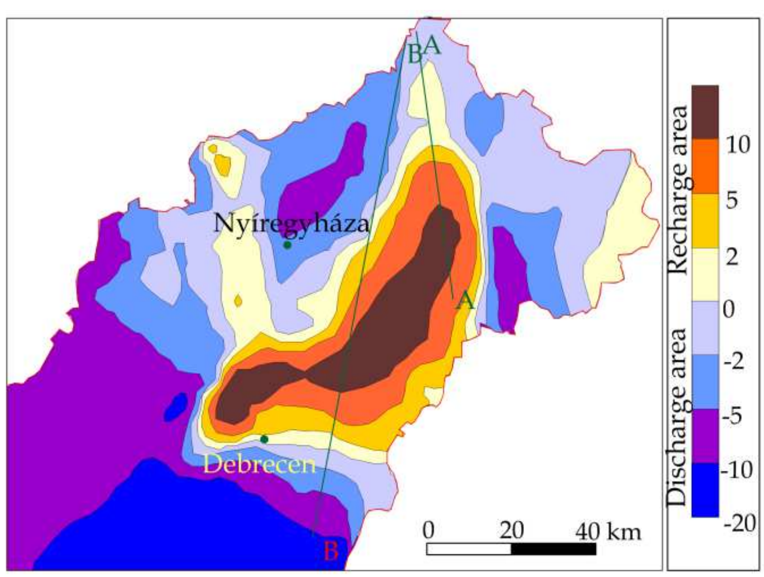


Fig. 7. Calculated piezometric head difference (m) between the layers of upper and lower Pleistocene, and the location of cross sections based on Székely, 2003

In the vicinity of larger cities, significant depressions are formed due to the withdrawal of subsurface water. Piezometric levels of the lower Pleistocene main aquifer declined 30 meters at Debrecen and 7 meters at Nyíregyháza. In the surroundings of Debrecen the compaction due to water withdrawal resulted in a 62 cm descent of the surface (Szanyi, 2004). According to other studies, descent of the shallow groundwater level can reach 1 meter at certain locations, e.g. at Debrecen. At other, smaller settlements (with a population of 20,000-30,000 people) the depressions are shallower, but still of substantial extent. A future equilibrium state expected to commence at a 4% decrease in the groundwater supply. The decreasing water table causes decreasing evapotranspiration and increasing infiltration from surface water flow (Table 2) (Székely, 2003, 2006).

Replenishment	Percentage
From small surface flows	30.75
From rivers	5.87
Decrease of Evapotranspiration	59.57
Decline of stored supply	3.81

Table 2. Replenishment forms related to the 209 m^3/day of production plus the 146,000 m^3/day marginal outflow of 2020 (after Székely, 2003)

Based on hydrogeological models the piezometric water levels predicted to rise due to decline of water withdrawal in the early 1990's. However, due to the increasing number of individual, private and often illegal well drilling and the connected water withdrawal, shallow groundwater level may decline even further (Székely, 2003, 2006).

Hydrographs of 215 monitoring wells were at our disposal for the analysis. The period of observation varied in time. Observation of some well started in 1933 and lasts up to today. The measurement frequencies were variable as well, changing from monthly, to daily. To correct for this situation and make the results comparable with other explanatory data we had to consider only the annual mean water level as our input data.

In the period of 1986-2000 122 hydrographs proved to be sufficiently homogeneous in order to apply dynamic factor analysis to them. Spatial distribution of the wells was uneven, the north part of the study area had less wells.

To be able to identify the factor time series it is necessary to have the data of the water balance. These are the time series of water withdrawal, precipitation and temperature. Though water withdrawal data is not registered with sufficient accuracy for the whole area, fortunately the communal withdrawal for the largest city at Debrecen Water Works has proper representative power (Fig. 8). So, we used it for the comparison. The necessary precipitation data were provided by 20 gauge stations and 4 other stations served us with daily temperature data.

So, we were in the position to apply dynamic factor analysis and analyse its results. The estimated factor time series were correlated to the mentioned elements of water balance, such as precipitation, evapotranspiration and withdrawal.

The first factor time series identifies the most influential background effect. The estimation of reference evapotranspiration (ET) was computed by the Hargreaves method, and then the actual evapotranspiration was calculated. The estimation is based on the daily maximum and minimum temperature and the solar radiation values (Allen et al., 1998). This method is preferred by a number of authors (Diodato & Geccarelli, 2006; Magliulo, 2003) and is widely regarded as a reliable one. The correlation coefficient between time series of the first factor and evapotranspiration is 0.73, which indicates that the most important effect is related to the evapotranspiration. To obtain a clearer relationship it is worth considering the subprocesses of the evapotranspiration as well.

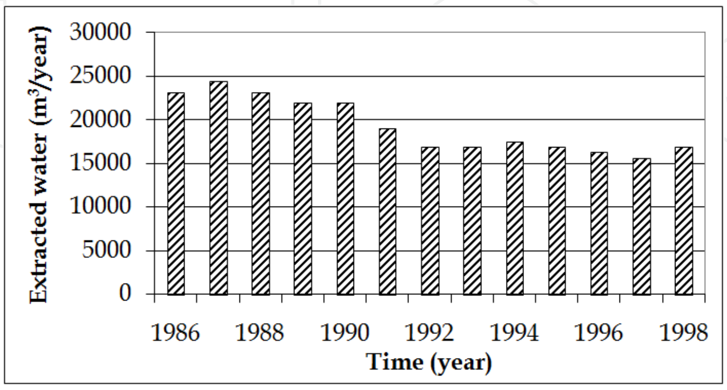


Fig. 8. Water consumption (in thousand m^3) from confined aquifers at Debrecen between 1951 and 1998 (Marton & Szanyi, 2000)

We employed the soil water characteristic curve to determine the evaporation of the shallow-groundwater (Kovács, 1981). The evaporation was computed for the 122 observation wells, and then the mean values were used to describe the study area. Fig. 9 displays the time series of evaporation and of factor one. The 0.96 correlation between the two time series indicates a very strong relation. Based on the results the first factor is identified as the evaporation of shallow-groundwater.

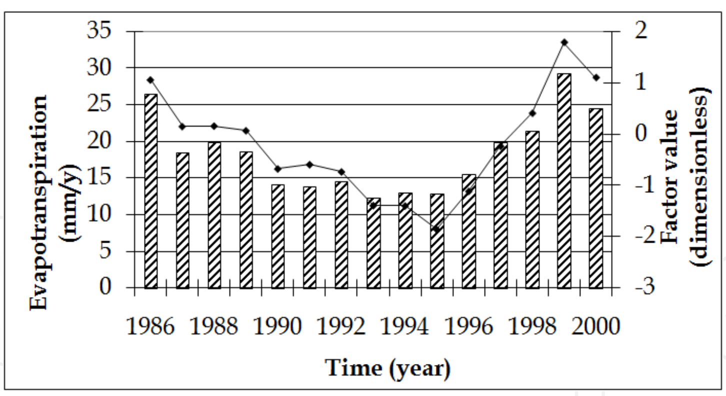


Fig. 9. The evaporation (columns) and the first factor (line) between 1986 and 2000

The other important element of the water-balance is the infiltrated part of precipitation. We compared this time series to the second factor. Because the studied area's climatic conditions, the majority of winter precipitation reaches the water table, while during vegetation it is absorbed by plants. The start and end of the vegetation period could shift by one or two months. On the long run, however, the sum of precipitation between December and March represents best the yearly infiltration. The water supply infiltrating to the shallow groundwater was calculated from the precipitation data of 20 gauge stations. It is more useful to consider the relation of each observation year to the long time average. For this purpose we created the time series of deviation from that average. The created deviation time series has 0.79 correlation with the second factor time series. Fig. 10 also underpins the strong relationship.

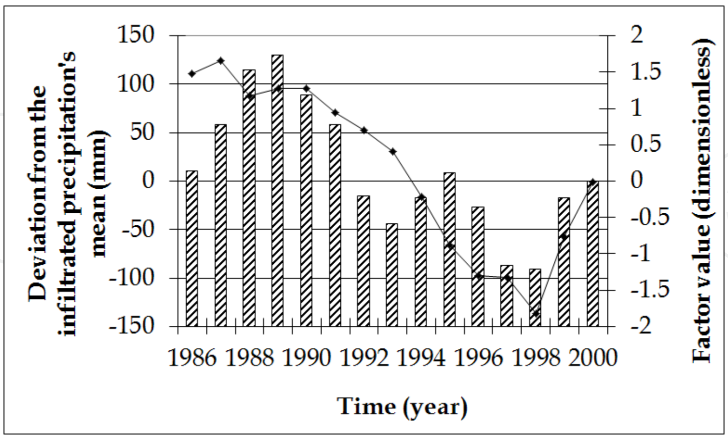


Fig. 10. Deviation from the mean of infiltrated precipitation (columns) and the second factor (line) between 1986 and 2000

Aside of infiltration, another downward moving effect exsits in water balance; it is the leakage effect, due to water withdrawal from deeper (100-150 m deep) layers. Several authors (Juhász, 1987; Marton & Szanyi, 2000) draw attention to its significance. For this reason we estimated the strength of the relationship between the shallow-groundwater level and the communal water withdrawal. Since leakage is a long-term process, the influence of the withdrawal will appear in water table with some delay; hence an appropriate time lag has to be taken into account. So, the relationship should be represented by cross-correlation. Table 3 shows the obtained cross-correlations. The dependence between the shallow-groundwater level and the withdrawal increases up to two-years lag, then it starts to decrease. We found that the second factor has strong relation to water withdrawal that reaches its maximum in a two-year lag at 0.89 correlation.

Time lag (year)	0	1	2	3	4	5
Groundwater level	0.63	0.68	0.69	0.58	0.39	0.35
2nd factor	0.83	0.88	0.89	0.85	0.76	0.60

Table 3. Cross correlation coefficients of water withdrawal with lagged shallow-groundwater level (first row) and the 2nd factor (second row)

As it was mentioned earlier, not only the factors themselves but also their loadings contain valuable information as they represent the intensity of a given effect in creating the actually observed time series, i.e. the hydrographs. The DFA-provided values of factor loadings belong to a fixed monitoring location. As we wish to extend the analysis to locations where no observation is available, the factor loadings should be spatially predicted by standard geostatistical methods, such as kriging etc. To this end the empirical semivariograms has to be computed first from the values of the factor loadings given at each observation points. In case of first factor the obtained semivariogram has two thresholds, and the corresponding ranges are 10 and 25 km-s. This allows us to conclude that the loadings possess a genuine spatial structure, that has a finer and a coarser scale, meaning that the effects appear at two different scale. Evaporation depends on the depth of water table, and is affected by two impacts. One of them is the local decline of the water level caused by water withdrawal from the deeper strata: this may correspond to the finer scale. The other impact may be related to topography, since the surface of shallow-groundwater follows the surface morphology. Lower values on the map of factor loadings appear in the vicinity of larger cities where the water level declined, indeed, by the extensive water withdrawal of the cities, thus decreasing evaporation (Fig. 11). Higher values mark the recharge areas and surroundings of rivers.

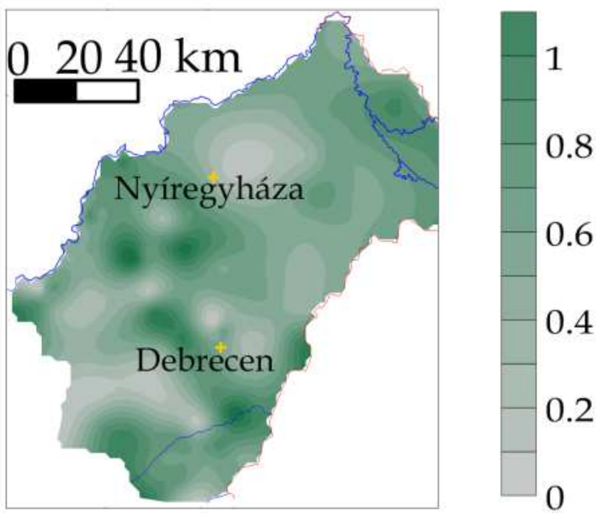


Fig. 11. Map of intensities (loadings) for the first factor

The loadings of the second factor (Fig. 12) were analysed in a completely similar way. In order to characterise the spatial structure of the loadings the empirical semivariogram was computed and fitted with a theoretical one. This time, too, we got a two-threshold semivariogram and the corresponding ranges were 12 and 35 kilometers. In our opinion the appearance of two ranges is the consequence of the existence of two processes that create the second factor.

The increase of the communication of water among strata is a local impact, hence creates a smaller range, while the infiltration from precipitation is influenced by the variability of the uppermost layers near to the surface, and that is reflected in the appearance of the larger range. The smaller range coincides with the one obtained for the first factor, and that is quite understandable in view of the generating process, the local decline of shallow-groundwater level.

The monitoring sites are also marked on the map of the loadings of the second factor, and notably their distribution is quite uneven. In subareas where observation locations are dense we could get an accurate information on the intensity of downward migration of water. On elevated ridges, that can be found in recharge areas, factor loadings are larger; here the effect of infiltrating precipitation is stronger hence more detectable in the hydrograph. The map displays higher loadings in the northeastern edge of the study area, however this does not reflect the reality; it is caused by an interpolation error that is the consequence of the lack of observation sites.

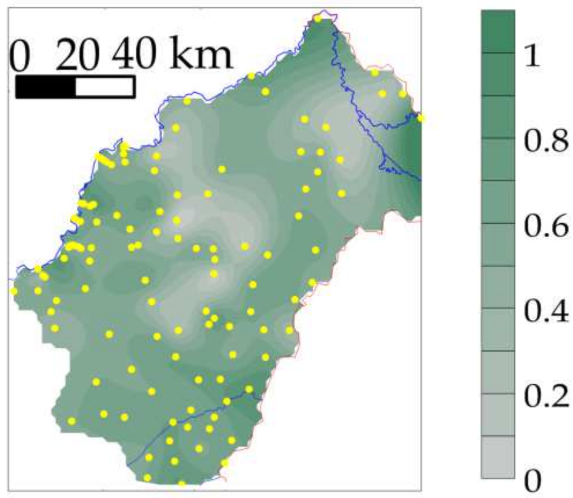


Fig. 12. Map of intensities (loadings) for the second factor (Yellow points indicate the monitoring wells.)

5. Study of a river influenced area

Our third area of case studies located on the southern part of the West Pannonian Basin, so called Kisalföld (Small Hungarian Lowland) at the NW state border of Hungary (Fig. 13).

The West Pannonian Basin is the second largest alluvial plain of the Carpathian-basin. It is located among the Transdanubian Central Range, the Alps and the Western Carpathians forming an irregular square depression (Bulla, 1962). Its basin was formed after periods of sedimentary deposition and depression surrounded by transversal and edge faults. The basin bottom at Northwest is formed from Variscian altered rock on top of which Miocene and Pannon-sea sediment was deposited. In the middle and Southeast regions the bottom is covered by thick Mesozoic limestone and dolomite beds on top of which Neogene marine sediment was deposited. After the Pannonian inland sea recessed the area was first occupied by the Pannonian Lake and in the Upper Pliocene fluviocarstic lake system (Bulla, 1962).

After the disappearance of the lake system in the beginning of the Pleistocene, the Danube and other rivers arriving from the Western-Carpathians passed through the area in a southern direction. In the Lower Pleistocene as a result of tectonics the area subsided, on which the Danube was able to form an extensive alluvial cone (Pécsi, 1959). At the Southwest part of the West Pannonian Basin the Rába and the smaller tributaries (arriving from the Alps) created the alluvial cone sequence. In the second part of the Pleistocene at the middle of the basin another subsidence was observable resulting in an alluvial cone with a diameter of approximately 200 km (Fig. 13) formed by the Danube, arriving from North-west – across the Gate of Devin (Porta Hungarica) - at that time. This subsidence (which was here and there still observable in the Holocene) defined the actual hydrographical picture of the Danube and its branches (Mosoni-Duna, Maly Dunaj) and the tributaries. The thickness of the Pleistocene and Holocene sediments together is 200-250 m (Don et al., 1993; Góczán, 1999; Neppel et al., 1999).

The top capping stratum of the alluvial cone was formed from the deposition of the fine grained, muddy river load in the flood plain originating from the Danube’s floods and the branches. The fine grained Holocene top capping is 0-5 m thick and shows diverse regional distribution. In the valleys of the Danube's tributaries it consists of Holocene alluvial sediment and in case of the Hanság region it consists of the lacustrine and wetland sediments. The NW part of the Small Hungarian Lowland (Kisalföld) is covered by coherent Pleistocene sand and pebbles that reaches the Rába in South-east direction in a narrow line. In the middle regions spots with alluvial and eolic sand make the surface more diverse.

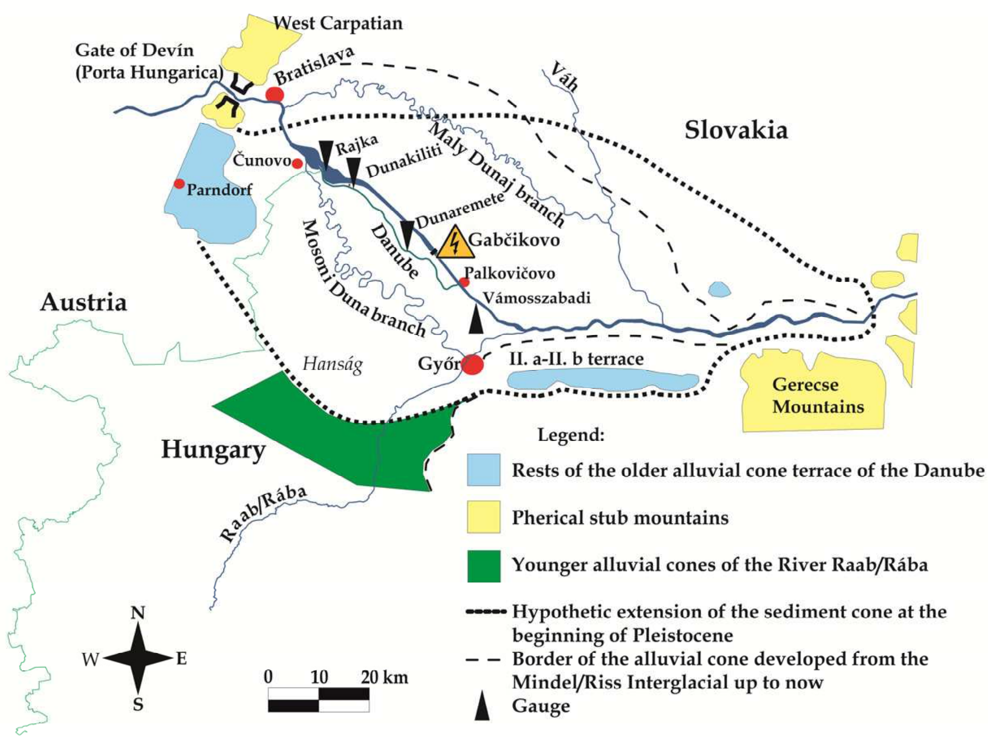


Fig. 13. Older and younger alluvial cones of the Danube on the West Pannonian Basin (based on Pécsi, 1959), with the names of the main geographical units occurring in the text

The West Pannonian Basin is not only a geographic but a hydrogeological unit as well because of the characteristics of the alluvial cones. From south the sandy-pebbly sediments of Rába valley connect to the alluvial cone of the Danube in Hungary (Erdélyi, 1979; Góczán, 1999).

Central Europe's largest potable water supply is located at the Kisalföld in its Pleistocene porous aquifer alluvial cone. The water coming from the Danube and its branches influenced the shallow groundwater level of the alluvial cone. Based on the data obtained from the shallow groundwater monitoring wells set up in 1950's and later, it was possible to determine the main direction of the groundwater migration. The direction of the flow was confirmed by isotope-hydrological measurements and its speed became measurable as well (Szalai, 2008). Further, it was also possible to determine those specific areas where the major part of the shallow groundwater supply is provided by the Danube (Erdélyi, 1979; Major, 1976). This hydrological situation has changed in the early 1990's.

In October 1992, on the Slovakian territory at Čunovo a dam was constructed at the 1851+750 river km, therefore natural fluctuation of the Danube has changed. Since then majority of the Danube flow has been diverted to the insulated power-plant channel that rejoins to the original riverbed only at the 1811 river km, at Palkovičovo. Therefore, the original 2000 m3*s-1 mean discharge dropped drastically in the natural river network. The Mosoni-Duna branch had only a flow rate of 10-20 $m^3 s^{-1}$, while the original Danube bed had only a flow rate of 250-350 $m^3 s^{-1}$. Consequently, water level in the riverbeds dropped several meters, and by 1993, some of the Danube's branch rivers had dried out. As the subsurface aquifer of the area is in a close connection with the surface water network, the level of shallow-groundwater also dropped significantly. Since 1995, to improve the situation, Slovakia has increased the annual mean water flow to the original Danube bed up to 400 $m^3 s^{-1}$. Furthermore, Hungary has built a submerged weir at Dunakiliti with a spillover at 123 m a.s.l., so water level increased in the original bed backward 10 km from the weir. The uplifted water level supports the artificial recharge of shallow groundwater from the rehabilitated riverbed and across the built up channel-system. The water supply to Mosoni-Duna branch has been increased to 40 $m^3 s^{-1}$ that has resulted in a slight growth of discharge.

In our approach the observed processes at different locations are supposed to be governed by the same essential impacts, such as recharge from precipitation and rivers and water withdrawal, while intensity of these impacts depends on locality, creating the main aspect of the spatial dependence.

Our goal was to identify these impacts, and to determine their spatially dependent intensities. In order to do this, dynamic factor analysis has been applied to the data, registered from the Danube's diversion (1992) to 2009. A total of 93 shallow groundwater monitoring wells have been sampled. Annual averages were formed from the data of each well. For the purpose of identifying factor-time series precipitation and Danube water level data were used as well. During the analysis three factor-time series have been estimated by DFA.

The standardised first factor-time series corresponded to the combined deviance of precipitation during the period of January-April. These months are known to be the periods that determine the major part of the supply for the shallow ground water. The two data series showed strong correlation (r=0.79) (Fig. 14).

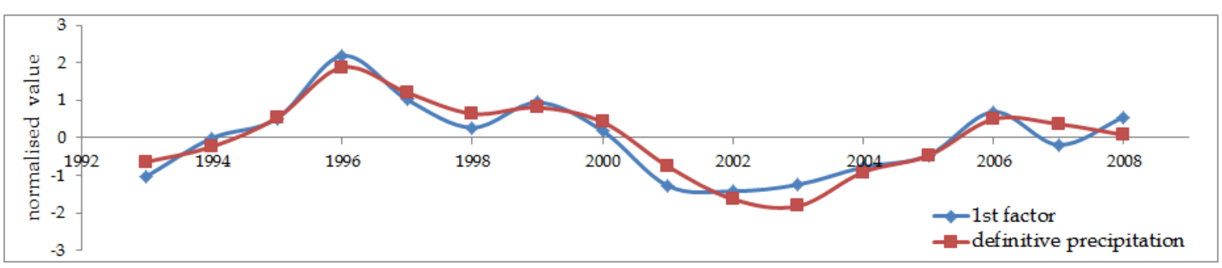


Fig. 14. Graphs of factor 1 and definitive precipitation

The isoline map prepared from the factor loadings contain information regarding the environmental impacts (Fig.15). Along the Danube the predicted factor loading values are small, because the dominant factors are the water levels of the river and the natural and manmade water supplementing channels. Higher values are observable North-west near to the state border. This particular area is covered by Pleistocene alluvial sand and pebbles, and their greater permeability increases infiltration.

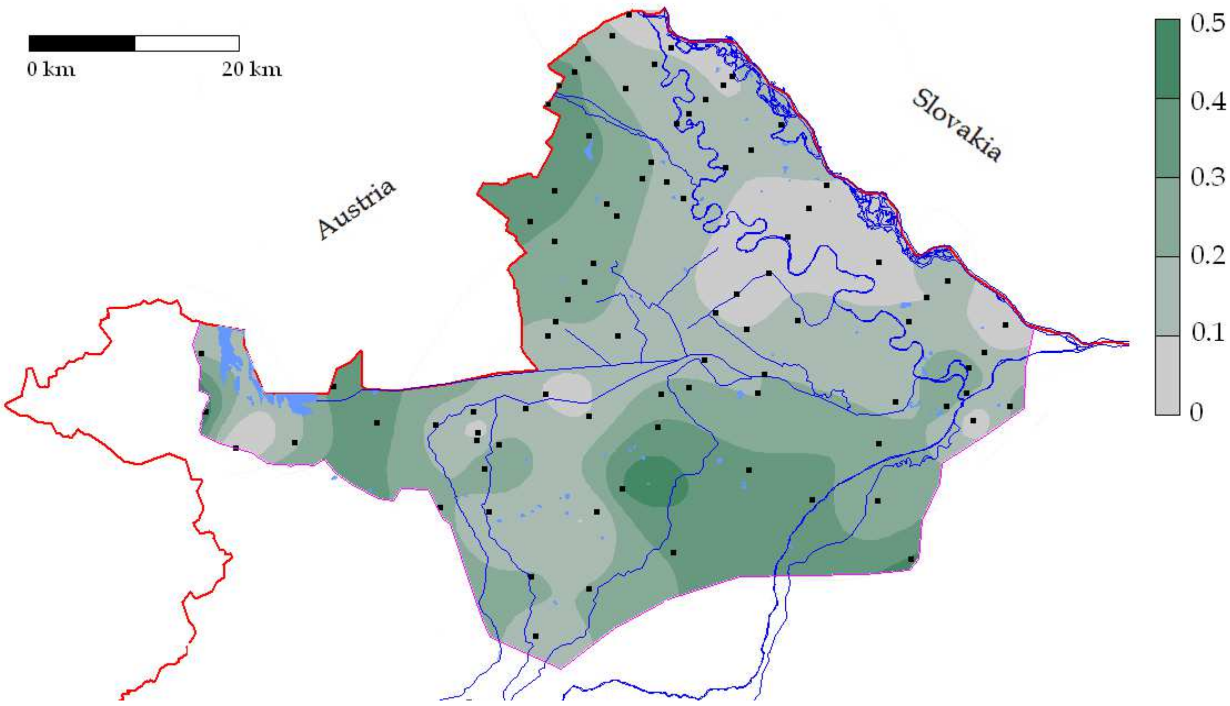


Fig. 15. Map of intensities (loadings) for the first factor

Second factor shows strong correlation (r=0.89) with the mean of registered water levels of Danube at two gauges, Dunaremete and Rajka (Fig. 16). Because of their location both water gauges are indirect indicators of river diversion and water level regulation. In the diagram the effect of the underwater weir installed in 1995 is remarkable. With the exception of floods this structure fixes the level of the upper water at a nearly constant value which is approximately 200-250 cm higher than what was measured immediately after the diversion. The increased water level with the recharging facilities make the gravitational recharge of underground water possible.

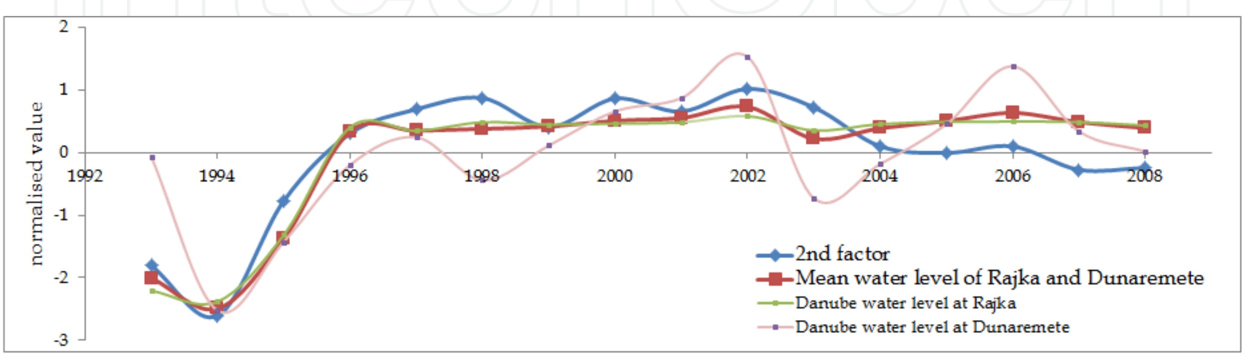


Fig. 16. Rescaled graphs of the factor 2 and water levels of Danube

Based on the map prepared from the loadings of the second factor (Fig. 17) the areas increasingly vulnerable to the diversion of Danube can be outlined. The highest values i.e. highest vulnerabilities were observed in the NW part of the area. In addition to manmade influence the reason may lay in the geological setting as well. The high factor loadings S and SE from this area are suspected to be indicating the effect of the gravitational recharge and the increased runoff of the Mosoni-Duna.

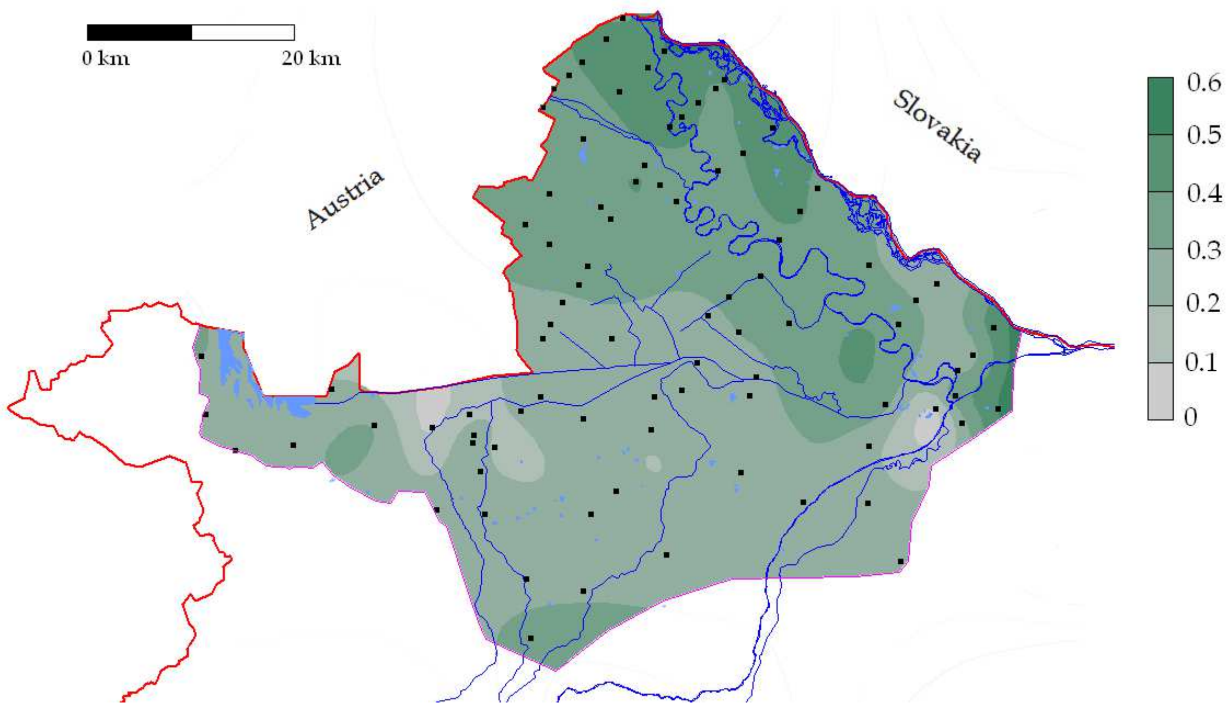


Fig. 17. Map of intensities (loadings) for the second factor

The third factor shows a relationship with Danube level values measured at Vámosszabadi water gauge (r=0.81) (Fig. 18). This gauge is located under the downstream channel of the hydroelectric power plant. By the installation of the plant hydraulic relations have radically changed. In addition to natural ones, considerable daily water level fluctuations appear according to the operation of the power plant. Pressure waves caused by water level fluctuations may be beneficial from the point of view of groundwater recharge, but their expansion is rather restrained by the fine-grained top capping stratum which is rather thick here (Fig. 19).

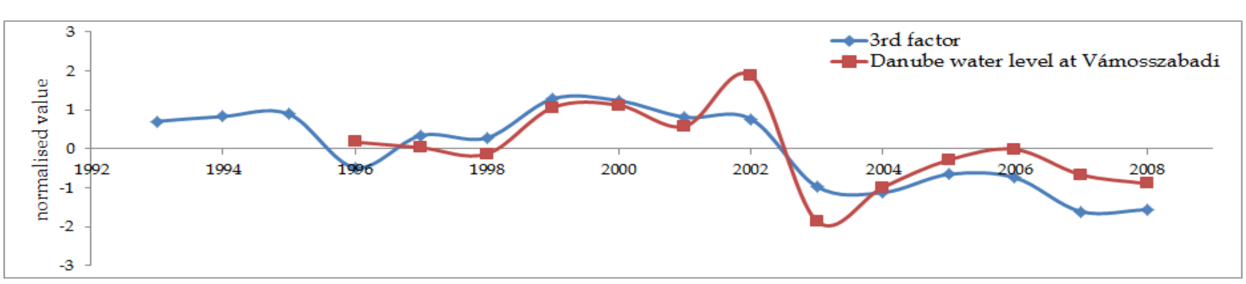


Fig. 18. Comparison of the third factor with the water level of the Danube at Vámosszabadi

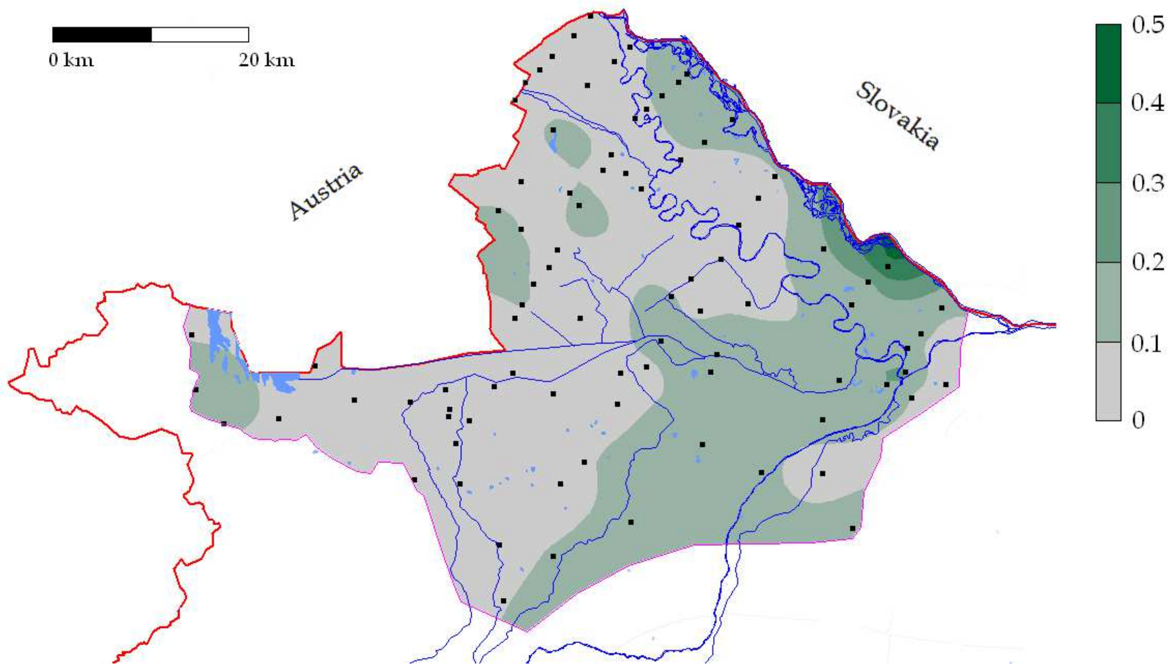


Fig. 19. Map of intensities (loadings) for the third factor

6. Summary

In the process of complex evaluation of water resources the quantitative and qualitative parameters cannot be considered separately. These parameters are influenced by several background effects that are variable in both time and space, hence it is dynamic factor analysis that makes their estimation mathematically possible. The identification of these effects and their intensity provide information on the area characteristics from either geological, hydrogeological or environmental point of view. The subject of the application of our proposed method is a well observable, measurable, physical parameter, the temporal fluctuation of water level at several monitoring sites of the area. Dynamic factor analysis decomposes the time series of water levels into the combination of factors that are then identified as the various occurrences of the precipitation, water withdrawal, and river effects. These effects primarily govern the water quantitatively but may also have an impact on the qualitative characteristics, therefore they might be informative and important markers on the water quality. That is why the factor loadings represent a quantitative measure of vulnerability of the studied areas and the corresponding maps constitute the main results of our works. As an extension of the present study, a joint analysis of the quantitative and qualitative characteristics can be considered in the sequel.

7. Acknowledgement

We the authors would like to thank all of our collegues from the Department of Phisycal and Applied Geology at the Eötvös Loránd University and for the help of Edit Borbás and Péter Tanos, and last but not least István Gábor Hatvani for reviewing our English version.

The European Union and the European Social Fund have provided financial support to the research under the grant agreement no. TÁMOP 4.2.1./B-09/KMR-2010-0003 and TÁMOP-4.2.1/B-091/1/KONV-2010-0005.

8. References

Alföldi, L. (1994). Ideas for the groundwater contamination's sensitivity questions (in Hungarian) *Hidrológiai Közlöny*, Vol.74, No.1, pp. 15-22, ISSN 0018-1823

R. G., Pereira, L. S.; Raes, D. & Smith, M., (1998). *Crop evapotranspiration: Guidelines for computing crop requirements.* Irrigation and Drainage Paper, No.56, FAO, Rome, Italy, pp. 300, ISBN 92-5-104219-5

Aller, L.; Bennet, T.; Lehr, J. H. & Petty, R. J. (1985). DRASTIC: A standardized system for evaluating groundwater pollution potential using hydrologic settings, U.S. EPA, Sobert S. Kerr Environmental Research Laboratory, Ada, OK, EPA/600/2-85/0108, p. 163, In: *A short overview on groundwater vulnerability assessment – basic statements for use in the framework of the COST 620 Action*, Gogu, R. & Dassargues, A. (1998).

Anderson, R. L. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, Vol.28, pp. 1-25, ISSN 0033-3123

Bai, J. (2003). Inference for factor models of large dimensions. *Econometrica 71*, pp. 135–172, ISSN 0012-9682

Bánkövi, Gy.; Veliczky, J. & Ziermann, M. (1979). Dynamic models for prediction of the development of national economies, In: *Models and Decision Making in National Economies*, Janssen, J. M. L.; Pau, L. F. & Straszak, A., (Eds.), pp. 257-267, ISBN 0-444-85280-8, Amsterdam, North-Holland

Böcker, T. (1974). *Determining infiltration in a karstic area using seasonal beoundary precipitaion method (in Hungarian)* VITUKI beszámoló, pp. 207-216, Budapest, Hungary

Bolla, M.; Michaletzky, Gy.; Tusnády, G. & Ziermann, M. (1998). Extremes of sum of heterogeneous quadratic forms. *Linear Algebra and Its Applications*, Vol.269, pp. 331-365, ISSN 0024-3795

Box, G. E. P. & Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika*, Vol.64, pp. 355-405, ISSN 0006-3444

Bulla, B. (1962). *Geography of Hungary* (in Hungarian), Tankönyvkiadó, Budapest, Hungary

Burman, R. & Pochop, L. O. (1994). Evaporation, Evapotranspiration and Climatic Data. *Developments in Atmospheric Science*, 22, Elsevier, p. 278, ISSN 0167-5117, Amsterdam, Holland

Civita, M. C. & De Regibus (1995). Determining aquifer vulnerability using different methods (in Italian: Sperimentazione de alcune metodologie per la valutazione della vulnerabilita degli aquiferi.) *Quaderni de Geologia Applicata*, 3, pp. 311-324, ISSN 1593-8433, Bologna, Italy

Connor, G. & Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance* 48 (4), 1263–91, ISSN 0022-1082

COST Action 65 (1995). Hydrogeological aspects of groundwater protection in karstic areas, Final Report, ISBN 92-827-4682-8, Luxembourg

Csepregi, A. (1995). Investigation of groundwater recharge on the Hungarian Plain by precipitation and areal evapotranspiration data (in Hungarian), Technical Report of the Hungarian Water Researches Institute

Daly, D. & Drew, D. (1998). Irish methodologies for karst aquifer protection, Paper submitted for conference "Seventh Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst", Pennsylvania, USA

Daly, D.; Dassargues, A.; Drew, D.; Dunne, S.; Goldscheider, N.; Neale, S.; Popescu, Ch. & Zwahlen, F. (2001). Main concepts of the "European Approach" for (karst) groundwater vulnerability assessment and mapping, Draft for Working Group 1 of the COST 620 Action "Vulnerability and risk mapping for the protection of carbonate (karst) aquifers", p. 12

Deák, J. (1979). Environmental isotopes and water chemical studies for groundwater research in Hungary. *Isotope Hydrology 1978*, Vol.1, International Atomic Agency, Vienna, pp. 221-249, ISBN 92-0-040079-5

Deák, J.; Stute, M.; Rudolph, J. & Sonntag, C. (1987). Determination of the flow regime of quaternery and pliocene layers in Great Hungarian Plain (Hungary) by D, 18O, 14C and noble gas measurements. *Isotope Techniques in Water Research Development,* International Atomic Agency, Vienna, ISBN 92-0-040087-6.

Deistler, M. & Scherrer, W. (1991). Identification of linear systems from noisy data, In: *Modeling, Estimation and Control of Systems with Uncertainty* (Sopron, 1990) Progr. Systems Control Theory, 10. Birkhauser, Boston, pp. 91-112, ISBN 0817635807

Diodato, N. & Geccarelli, M. (2006). Computational uncertainty analysis of groundwater recharge in catchment. *Ecological Informatics*. Vol.1, pp. 377-389, ISSN 1574-9541

Doerfliger, N. (1996). Advances in karst groundwater protectin strategy using artificial tracer tests analysis and multiattribute vulnerability mapping (EPIK method) – Thesis presented to the Faculty of Sciences of the Neuchatel University for the obtention of the title Docteur ès sciences, University of Neuchatel (Switzerland), Faculty of Sciences, p. 308

Don, Gy.; Kaiser, M.; Marsi, I.; Scharek, P.; Síkhegyi, F.; Szeiler, R.; Tullner, T. & Zsámbok, I. (1993). Geological Setting of the Szigetköz - p. 63 +14 annexs. MÁFI Library, Budapest, Hungary (manuscript)

Doz, C.D. Giannone & Reichlin, L. (2006). *A quasi maximum likelihood approach for large approximate dynamic factor models*. European Central Bank, p. 37, ISSN 1725-2806, Frankfurt am Main, Germany

Dracos, Th. (1980). Hydrology, introduction for engineers (in German), Springer-Verlag, p. 194, ISBN 978-3540632924

Dryden, I.; Markus, L.; Taylor, C.; Kovacs, J. (2005). Non-Stationary spatio-temporal analysis of karst water levels, *Journal of the Royal Statistical Society, Series C-Applied Statistics*, Vol. 54., No.3., pp. 673-690, ISSN 1467-9876

Engle, R. F. & Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association* 76, pp. 774–81, ISSN 0162-1459

Erdélyi, M. (1979). Hydrogeology and Hydrodinamics of the West Pannonian Basin (In Hungarian). *Hidrológiai Közlöny,* Vol. 59, No.7, pp. 290-301, ISSN 0018-1823

Forni, M.; Hallin, M.; Lippi, M. & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics* 82, pp. 540–54, ISSN 1530-9142

Geweke, J. F. (1977). The Dynamic factor analysis of economic time series models In: Latent Variables in Socio-economic Models, Aigner, D.J. & Goldberger, A.S., (Eds.), pp. 365-382, Amsterdam, North-Holland

Góczán, L. (1999). The Little Hungarian Plain (in Hungarian), In: *Pannon Encyclopaedia Magyarország földje* (Pannon Encyclopaedia The land that is Hungary – geology, geography and cartography of Hungary and its surroundings) Karátson, D., 302-305, Kertek 2000 Publishing House, ISBN 963-85792-3-4, Budapest, Hungary

Goldscheider, N.; Klute, M.; Sturm, S. & Hötzl, H. (2000). The PI method – a GIS-based approach to mapping groundwater vulnerability with special consideratin of karst aquifers, Z.angew. Geol., Vol.46, No3., pp. 157-166

Gourieroux, C. & Monfort, A. (1997). Time Series and Dynamic Models. *Cambridge University Press*, p. 665

Gourieroux, C.; Monfort, A. & Renault, E. (1995). Inference in Factor Models, In: *Advances in Econometrics and Quantitative Economics*, Maddala, G.; Phillips, P.C.B. & Srinivasan, T., (Eds.) pp. 311-353, ISB**N** 1-55786-382-2, Blackwell

Jeannin, P. Y. & Grasso, D. (1995). Estimation des infiltrations efficaces journali\`eres sur le bassin karstique de la Milandrine, (Ajoie, JU, Suisse). *Bulletin d'Hydrogéologie,* No.14., pp. 83-93, ISSN 1421-4946

Juhász, J. (1987). *Hydrogeology* (in Hungarian). Akadémiai Kiadó, pp. 466-548, 871, ISBN 963 05 4381 8, Budapest, Hungary

Jungbacker, B. & Koopman, S. J. (2008). Likelihood-based Analysis for Dynamic Factor Models, Tinbergen Institute Discussion Paper TI 2008-0007/4

Kessler, H. (1954). The percentage of infiltration and the permanently extractable water in karst-terrains (in Hungarian) *Vízügyi Közlemények (Journal of Hydraulic Engineering)*, 36, ISSN 0733-9429

Kovács, Gy. (1981). *Subterranean Hydrology.* Water Resources Publications, pp. 440-460. ISBN 978-0918334350 Colorado, Littleton, USA

Kovács, J.; Mákus, L.; Halupka, G. (2002) Measuring contamination–vulnerability of aquifers by dynamic factor analysis, 8th Annual Conference of the International Association for Mathematical Geology, *Terra Nostra*, Schriften der Alfred-Wegener-Stiftung, ISSN 0946-8978 Volume 2, pp.: 463 – 468 Berlin, Germany

Kovács, J. (2007). The application of modern geomathematical methods in practical hydrogeology (in hungarian, with english abstract), Ph. D. dissertation, p. 122 (manuscript)

Mádlné, Szőnyi, J. & Füle, L. (1998). Groundwater vulnerability assessment of the SW Trans-Danubian Central Range, Hungary. *Environmental Geology*, 35, pp. 9-18, ISSN 0943-0105

Mádlné, Szőnyi, J. (1997). Vunlerability analysis of aquifers on the example of the Transdanubian Central Range's karst aquifer (in Hungarian) *Földtani Közlöny*, Vol.127, No.1-2, pp. 19-83, ISSN 0015-542X

Magliulo, V.; d'Andria, R. & Rana, G. (2003). Use of the modified atmometer to estimate reference evapotranspiration in Mediterranean environments. *Agricultural Water Management,* Vol.63, No.1, (November 2003), pp. 1-14, ISSN 0378-3774

Major, P. (1976). Effects of the Dunakiliti hydropower station on the subsurface water body in Szigetköz, (in Hungarian) Vízgazdálkodási Tudományos Kutató Intézet (VITUKI report), VITUKI, Budapest

Mardia, K. V.; Goodall, C. R.; Redfern, E. J. &Alonso, F. J. (1998). The kriged Kalman filter. *Test.* Vol.**7,** pp. 217-285

Márkus, L. & Kovács J. (2003). On the Effects of Water Extraction of Mines in Karstic Areas. *Mathematische Geologie,* Vol.7, ISSN 1431-8016

Márkus, L.; Berke, O.; Kovács, J. & Urfer, W. (1999). Spatial Prediction of the Intensity of Latent Effects Governing Hydrogeological Phenomena. *Environmetrics,* Vol.10, pp. 633-654, ISSN 1180-4009

Marton, L. & Szanyi, J. (1997a). Geostatistical investigation of pleistocene sediments in Eastern-Hungary: 1. Mapping of the transmissivity (in Hungarian), *Journal of the Hungarian Hydrological Society*, Vol.77, No.5, pp. 233-240, ISSN 0018-1823

Marton, L. & Szanyi, J. (1997b). Geostatistical investigation of pleistocene sediments in Eastern-Hungary: 2. Mapping of the leakage (in Hungarian), *Journal of the Hungarian Hydrological Society*, Vol.77, No.5, pp. 241-248, ISSN 0018-1823

Marton, L. & Szanyi, J. (2000). Connection between the level of the shallow groundwater and the communal water extraction (in Hungarian), *Journal of the Hungarian Hydrological Society* Vol.80, No.1, pp. 3-13, ISSN 0018-1823

Marton, L. (1981). *Investigation of the groundwater in Nyírség using environmental isotopes (in Hungarian)*, C. Sc. Thesis, Debrecen, p. 133.

Maucha, L. (1990). Calculation of karstic infiltration (in Hungarian). *Hidrológiai Közlöny* (*Journal of the Hungarian Hydrological Society*), Vol.70, pp. 153-161, ISSN 0018-1823

Morton, F. I. (1983). Operational estimates of areal evapotranspiration and their significance to the science and practice. *Journal of Hydrolog*y, Vol.66, pp. 1-76, ISSN 0022-1694

Munoz-Carpena, R.; Ritter, A.; Li, Y.C. (2005). Dynamic factor analysis of groundwater quality trends in an agricultural area adjacent to Everglades National Park, *Journal of Contaminant Hydrology*, Vol. 80., No. 1-2, pp. 49-70, ISSN 0169-7722

Neppel, F.; Somogyi, S.; Domokos, M. and the experts of the cooperating Danube Countries (1999). Paleogeography of the Danube and its cathment. A hydrographic monograph. Follow-up volume V. The riverbed conditions of the Danube. Part 2, Regional cooperation of the Danube Countries in the Framework of the International Hydrological Programme of UNESCO, Published by VITUKI, ISBN 963 511 128 2

Pécsi, M. (1959). A magyarországi Duna-völgy kialakulása és felszínalaktana (Genesis and surface morphology of the Hungarian Danube Valley, (in Hungarian). *Földrajzi Monográfiák*, Vol.3, Akadémiai Kiadó, Budapest, Hungary.

Picci, G. & Pinzoni, S. (1986/1). A new class of dynamic models for stationary time series, In: *Time Series and Linear Systems*, Bittanti, S.,(Ed.) Lecture Notes in Control and Information Sciences, Vol.86, pp. 69-114, ISBN 3-540-16903-2

Picci, G. & Pinzoni, S. (1986/2). Factor analysis models for stationary stochastic processes, In: *Analysis and Optimization of Systems (Antibes, 1986)* Lecture Notes in Control and Information Sciences Vol.83, pp. 411-424

Primault, B. (1963). Du calcul de l'evapotranspiration. *Arch. Meteorologie, Geophysik, Biologie,* Vol.12, pp. 124-150, ISSN 00039861, Zürich

Rapcsák, T. (2002). On minimization on Stiefel Manifolds. *European Journal of Operational Research*, Vol.143, pp. 365-376, ISSN: 0377-2217

Reis, R. & Watson, M. W. (2007). Relative Goods' Prices, Pure Inflation, and the Phillips Correlation (November 2007). NBER Working Paper Series, Vol. w13615, pp. -, 2007. Available from http://ssrn.com/abstract=1032868

Ritter, A. & Munoz-Carpena, R. (2006). Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park, *Journal of Hydrology*, Vol. 317, No. 3-4, pp 340-354, ISSN 0022-1694

Ritter, A.; Munoz-Carpena, R.; Bosch, D. D. (2007). Agricultural land use and hydrology affect variability of shallow groundwater nitrate concentration in South Florida, *Hydrological Processes* Vol. 21, No. 18., pp. 2464-2473, ISSN 1099-1085

Sargent, T. J. & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. A. S. et al. (Ed.), New methods in business cycle research. Minneapolis: Federal Reserve Bank of Minneapolis

Stock, J. H. & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, pp. 1167–79, ISSN 0162-1459

Szalai, J. (2008). Recharging efforts of the subsurface water body in Szigetköz (in Hungarian) report, VITUKI, 2008

Szanyi, J. (2004). *Environmental effects of groundwater production from South-Nyírség's example (in Hungarian).* Ph.D. thesis, University of Szeged, Departement of Mineralogy, Geochemistry and Petrology, p. 81

Székely, F. (2003). *Developping of the hydrogeological model in Northeastern part of the Great Hungarian Plain (in Hungarian),* Report of the Environmental and Water Researh Institute, p. 61, Budapest

Székely, F. (2006). Results of hydrogeological modelling of porous sediments in the Northeastern part of the Great Hungarian Plain (in Hungarian). *Journal of the Hungarian Hydrological Society,* Vol.86, No.4, pp. 23-28, ISSN 0018-1823

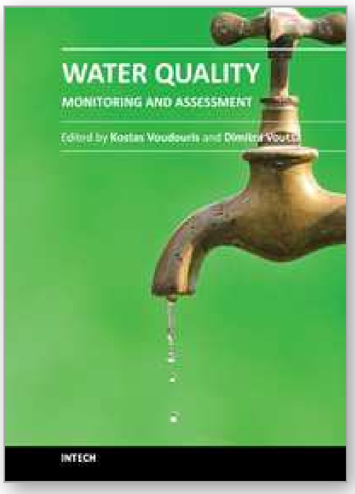
Urbancsek, J. (1965). Structural geology of the Quarter in the Great Hungarian Plain (in Hungarian). *Journal of the Hungarian Hydrological Society,* Vol.47, No.3, pp. 111-124. ISSN 0018-1823

Van Schuppen, J. & Van Putten, C. (1985). Invariance properties of the conditional independence relation. *The Annals of Probability,* Vol.15, No.3, pp. 934-945, ISSN 1050-5164

Virág, M.; Szűcs, P. & Lakatos, A. (2005). Local hydrogeological models in the Nyírség area (in Hungarian), *Proceeding of the XXIII. Conference of the Hungarian Hydrological Society,* Nyíregyháza, July 2005

Watson, M. W. & Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *Journal of Econometrics* 23, pp. 385–400, ISSN 0304-4076

Ziermann, M. & Michaletzky, Gy. (1995). Factoranalysis of time series, Szigma, XXVI No. (3-4), pp. 77-91



Water Quality Monitoring and Assessment
Edited by Dr. Voudouris

ISBN 978-953-51-0486-5
Hard cover, 602 pages
Publisher InTech
Published online 05, April, 2012
Published in print edition April, 2012

The book attempts to covers the main fields of water quality issues presenting case studies in various countries concerning the physicochemical characteristics of surface and groundwaters and possible pollution sources as well as methods and tools for the evaluation of water quality status. This book is divided into two sections: Statistical Analysis of Water Quality Data;Water Quality Monitoring Studies.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821