Hybrid modeling of the soil water regime with physically based and data driven approach

M. Cisty, J. Skalova, M. Pasztorova and J. Bezak

***Abstract*—**Global warming impacts the water cycle not only by changing regional precipitation and temperatures and their temporal variability, but consequently also by affecting soil moisture dynamics, which is a crucial condition of crop production. Adaptation strategies, such as improved water management or development of the more efficient irrigation systems, will be important tools in limiting the adverse effects of expected climate changes. The characteristics of a water regime for such decision making can be obtained by the mathematical modeling of the soil water regime. This modeling depends on the knowledge of the input data which are necessary for the numerical simulations. The aim of this paper is to contribute to solving this data acquisition task by using a data-driven methodology, namely support vector machines (SVMs). It is used for acquiring of the important inputs in such a modeling – namely ground water levels. The results of the paper confirm that in the soil moisture modeling, influence of the limited data availability which naturally exists in the context of the climate change impact studies, has an acceptable influence on the final precision and could be substituted by described data modelling approach. This paper emphasizes the usefulness of the proposed symbiosis of the data-driven and physically-based types of modeling.

***Keywords*—**climate change, data generation, soil water modelling, support vector machines.

I. Introduction

AGRICULTURE is considered to be one of the most exposed sectors with respect to the effects of climate change. The higher temperatures expected can reduce crop yields, and changes in the distribution of precipitation complicate water management and may increase the likelihood of crop failure. Global warming impacts the water cycle not only by changing regional precipitation and temperature levels and their temporal variability, but also by affecting soil moisture dynamics, which are a crucial condition of crop production. Adaptation strategies, such as improved water management or the development of more efficient irrigation systems, will be important in limiting the adverse effects of changes to climates. The characteristics of a water regime for such decision making can be obtained by the mathematical modeling of the soil water regime. Various models could be used for such a study. The GLOBAL [1], MOVOREP [2] or HYDRUS-ET [3] mathematical models are among some of the most frequently used tools. However, the most common and most widely used model in Slovakia is the GLOBAL mathematical model, which is also used for the simulation of the soil water regime in this paper.

M. Cisty is with the Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia (phone: +421259274628, e-mail: milan.cisty@stuba.sk)

J. Skalova is with the Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia (e-mail: jana.skalova@stuba.sk)

M. Pasztorova is with the Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia (e-mail: maria.pasztorova@stuba.sk)

J. Bezak is with the Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovakia (e-mail: juraj.bezak@stuba.sk)

The above-mentioned types of models are based on knowledge of the governing physical laws, their mathematical representation, and the resulting equation's system algorithmization. In addition to understanding the physical processes, this modeling crucially depends on the availability of the input data needed for the numerical simulations. This study is concerned with the data acquisition for soil moisture modeling in the context of climate change impact studies. Some of the data necessary for modeling water transport in soil, e.g., temperatures or precipitation, could be derived from climate change scenarios for this purpose. A stochastic weather generator is one possible tool for climate change downscaling, which produces a time series of weather data for a location based on the statistical characteristics of the observed weather at that location and the selected climate change scenario. The generator produces consistent series of meteorological variables, such as precipitation, temperature, humidity, wind, sunshine, etc.

Various studies have been performed to examine the influence of other phenomena, namely, water table fluctuations in soil moisture modeling, [4] – [10]. Fluctuations in water table depths have important impacts on hydrological, agricultural, and environmental issues. The soil water modeler must therefore also deal with the task of how to obtain a time series with water table values which is consistent with other data, e.g., the mentioned meteorological data. A number of numerical models, which are usually governed by the Boussinesq equation, have been developed for this task [11] – [18]. However, such physically-based models require an explicit understanding of the complicated physical processes and relationships and a great amount of the meteorological, hydrological, and geological data of the study area as inputs [19-20]. Spatially and temporally variable aquifer parameters and boundary and initial conditions should be known if one wants to perform numeric simulations. While such modeling constitutes a powerful tool, it also presents the formidable challenge of overcoming parameter uncertainty, which, to date, has not been especially satisfactorily resolved for groundwater problems, with the consequence of producing model prediction errors or even disabling the possibility of using this modeling approach [21].

The purpose of this paper is to contribute to solving the task of obtaining data about groundwater levels in the context of climate change impact studies of a soil water regime by using a different approach, i.e., a data-driven methodology, namely support vector machines (SVMs). As was indicated in the previous paragraph, it seems unsuitable to apply the above-mentioned physically-based modeling of a water table in the context of climate change studies, mainly because of the unavailability of the necessary data. On the other hand, data-driven modeling could be preferable because of its lower data and knowledge demands. In some hydrology studies SVMs have already been successfully applied [23] but usually in the context of prediction tasks in which the present conditions are known, and the prediction is realized in not very distant time steps. In such a context, the predicted variable in a previous time step could be used as input, which is not possible for the extremely distant time horizons examined in climate change impact studies. For this reason the authors of this paper are dealing with predictions only on the basis of variables which are readily available from weather generators. This limitation could lower the degree of precision to some extent in comparison to when the known values of a predicted variable from previous time steps are used as predictors. The authors of this paper are investigating the task of estimating and evaluating this impact of the input data limitation in the generation of water table levels on the final modeling of the soil water moisture, which is accomplished by means of physically-based modeling (by the GLOBAL model).

SVMs - a tool used in this study for generating water table levels consistent with other variables - were first developed to solve a classification problem and then were extended to regression problems [22], which is also accomplished in the present study. In a support vector regression algorithm, the input data are nonlinearly mapped into a higher dimensional feature space, in which the training data may exhibit linearity, so that a linear regression problem is solved in this feature space. Similarly as with the usual linear regression, the basic goal is to find a function that approximates the training points well by minimizing the prediction errors. The essential difference in SVM applications is that all the deviations smaller than a user-specified value are discarded. Also, when minimizing an error, the risk of over-fitting is reduced by simultaneously trying to maximize the flatness of the regression function. SVMs are gaining popularity due to these attractive features, which equip SVMs with a greater ability to generalize - the main goal in data-driven modeling. For these reasons SVMs have been applied to the prediction of various water resource variables. Asefa [23] and Khalil [24] used SVMs to capture the spatial distribution features of a groundwater´s surface and quality, respectively. Based on past measurements of soil moisture and meteorological data, Gill [25] predicted soil moisture using SVMs. More recently, Gill [26] compared the performance of an artificial neural network (ANN) and SVMs for predicting groundwater levels under conditions of incomplete data that were assumed to be randomly missing. Most researchers have showed that performance of SVMs is superior to various other methods, e.g., the well-known ANN methodology.

One of the important emphases of this paper is to highlight the usefulness of the proposed symbiosis of the data-driven and physically-based types of modeling. The rest of the paper is organized as follows: in Section 2 the methodology used is described. Firstly, the physically-based modeling and the GLOBAL model's basic principles, and then a brief description of the SVMs, which is used for input generation to the GLOBAL model, are given. In Section 3 a case study and description of the data are presented. Section 4 presents and explains the computational experiments and the results obtained. Finally, some conclusions about the possibility of input generation by SVMs for the purpose of soil moisture modeling by a physically-based model are drawn in Section 5.

II. METHODOLOGY

A. Physically-based modelling of a soil water regime

A simulation of a soil water regime was carried out using the GLOBAL mathematical model, which was developed at the Institute of Hydrology of the Slovak Academy of Sciences in Bratislava [1]. It is a mathematical simulation model of the movement of water in soil, which allows for the calculation of the distribution of the soil water potential and the soil moisture over time, based on the flow in the soil matrix. The model is based on a numerical solution of nonlinear partial differential equations of the movement of water in an unsaturated soil zone in the form [27]:

$$\frac{\partial h_w}{\partial t} = \frac{1}{C(h_w)} \frac{\partial}{\partial z}\left[k(h_w)\frac{\partial h_w}{\partial z} + 1\right] - \frac{S(z,t)}{C(h_w)} \tag{1}$$

where:

h_w is soil water potential [cm],
z - vertical coordination [cm],
$k(h_w)$ - unsaturated soil hydraulic conductivity [cm.s^{-1}],
$S(z,t)$ - root extraction rate [cm.s^{-1}],
$C(h_w) = \partial\theta/\partial h_w$ - specific water capacity [cm^{-1}],
θ - volumetric soil water content [cm^3.cm^{-3}],
t - time [s].

The model calculates the characteristics in a daily step and provides an original method of calculating evapotranspiration and its components (transpiration, evapotranspiration). To enhance the accuracy of the modelling, a function characterizing the abstraction of water by plant roots is provided. The water retention curves are approximated by the Van Genuchten [28] method, and the $k=f(h_w)$ dependences are calculated according to Mualem's theory [29]. The GLOBAL model also includes the water retention curve's hysteresis.

B. Generating inputs for modeling a soil water regime by SVMs

A support vector machine (SVM) is a supervised learning method that produces input-output mapping functions from a set of labeled training data [30]. Support vector machines are gaining popularity due to their many attractive features and promising empirical performance.

In data-driven modeling a model is usually trained using a data set (x,y), where x is the predictor and y the modeled variable, by fine-tuning the model's parameters so as to minimize an error function, which is a corresponding measure between the predicted and actual values of y. This type of objective function is based on the so-called empirical risk minimization principle, and a problem with its use is that it does not guarantee a small generalization error (i.e., an error on data other than that which was used in the training). To resolve this problem Vapnik [22] employs in SVM the concept of so-called structural risk minimization (SRM), described e.g., in [31]. The SRM principle defines a trade-off between the quality of the approximation of the given training data and the complexity of the approximating function. Less complex models tend to have a better generalization ability, which has been reported many times, e.g., while comparing SVM results with neural networks.

Another basic idea behind SVMs is to project the input data by means of kernel functions into a higher dimensional space called the feature space, where a linear regression can be performed, although a nonlinear problem is initially to be solved (real world problems are usually nonlinear). The results of the regression are then mapped back to the original nonlinear input space.

The next important concept in SVM methodology is to fully ignore small errors, while evaluating the precision of the modeling (by introducing the variable ε, which defines what the "small" error is). This makes the regression task dependent on a smaller number of inputs than were given in the original task, which makes the methodology much more computationally treatable. These crucial vectors of the inputs are called the support vectors.

In an ε-SVM regression (as opposed to classification or clustering), the goal is to find a function f(x) that at most has an ε deviation from the actually obtained targets y for the training data:

$$y = w.\Phi(x) + b \tag{2}$$

Where y is the model output, x is the input mapped into a feature space, and w and b are the parameter vectors of the searched regression function. The goal of a regression algorithm is to fit a flat function to the data points. A smaller w means a smoother and less complex approximating function, which means that for the sake of a good generalizing solution, one seeks a small w. Thus, the regression problem can be written as a quadratic optimization problem:

$$\text{minimize} \quad \frac{1}{2}\|w\| + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{3}$$

$$\text{subject to:} \quad y_i - \left(w.\Phi(x) + b\right) \le \varepsilon + \xi_i$$

$$\left(w.\Phi(x) + b\right) - y_i \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0$$

where ξi, ξi* are slack variables that specify the upper and lower training errors, subject to an error tolerance ε (soft margin), and C is a positive constant that determines the degree of the penalization loss for the model's output errors. In equation system (2), the objective function simultaneously minimizes both the empirical risk and the model's complexity; the trade-off between these two goals is controlled by parameter C. In most cases the optimization problem (2) can be solved more easily in its dual formulation. Hence, a standard dualization method utilizing Lagrange multipliers is usually applied for the final formulation of the support vector regression problem. Various approaches to the quadratic optimization problem for solving this system could be used. The details are described in, e.g., [32].

III. Area of study and data description

The proposed methodology was applied to the "Poiplie" region in Slovakia in the Ipeľ River basin, which is in the southern part of Slovakia (Fig. 1). This area has a typical flat relief of a lowland river floodplain, with an average slope of the terrain of around 1°. The prevailing soils in the investigated area are fluvisols and regosols, and the soil types are heavy and medium heavy soils, namely clay-loam, clay soils, loam and loam-sandy soils [30].

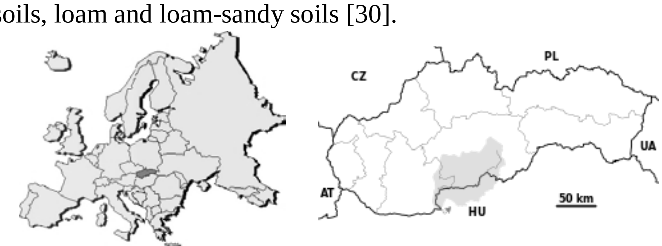

Fig. 1 Location of Poiplie in Slovakia

From the point of view of climatic conditions, the Poiplie area has a warm and dry climate with mild winters. The average annual air temperature is around 9.6 °C and approximately 16.5 °C during the growing season. The trend in the air temperatures is shown in Fig. 2, where a slight increase in air temperatures during the period 1977-2010 is clearly seen, which could be considered as a possible expression of climate change.

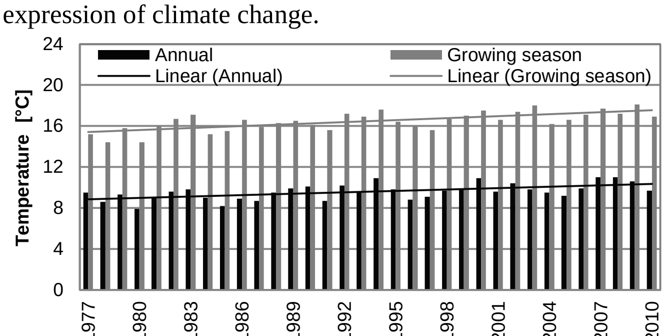

Fig. 2 Average annual and growing season temperatures in the area investigated during the period 1977-2010

The area investigated is relatively dry; the amount of annual precipitation is around 600 mm. The maximum precipitation usually occurs in the summer months, especially in June, which is due to the influence of the tidal marine polar air from the Atlantic to the European hinterland. Another peak is usually in November, which can be explained by the influence of Mediterranean Adriatic disorders. Fig. 3 shows the increasing trend of the annual precipitation totals and precipitation totals during the growing season for the years 1977-2010.

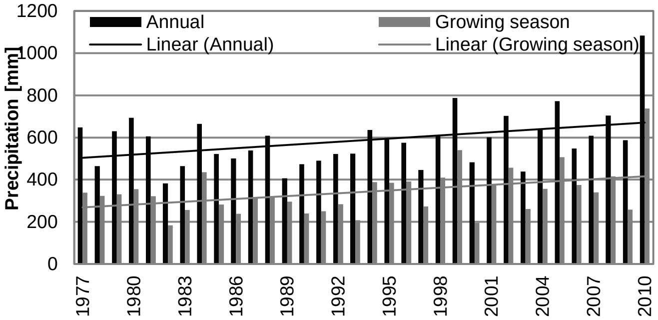

Fig. 3 Annual precipitation totals and precipitation totals during the growing season for the period 1977-2010

The Ipeľ River springs in central Slovakia in the Slovenské Rudohorie mountains. It flows south to the Hungarian border, and then southwest, west and again south along the border until it flows into the Danube near Szob, with an average annual flow of 21 $m^3.s^{-1}$. The annual water level on the Ipeľ River near the location where the modeling using GLOBAL was accomplished for the period 1989-2010 is shown in Fig. 4.

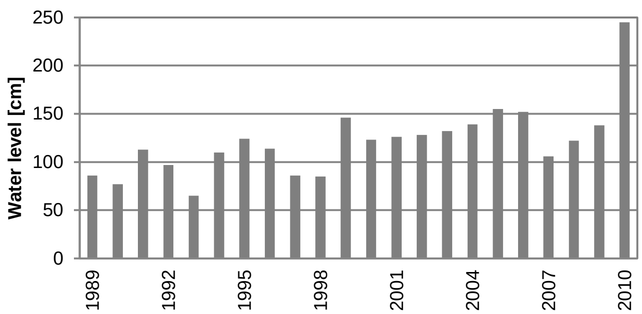

Fig. 4 Course of the average annual water levels on the Ipeľ River at the SHI Vyškovce above the Ipeľ station for the period 1989-2010

The groundwater resources in the area are supplemented by atmospheric precipitation, and fluctuations in the level are caused by fluctuations in the water level of the Ipeľ River. The trend in the groundwater levels at the Šahy station is shown in Fig. 5. An evaluation of the fluctuations in the groundwater level is shown in Fig. 6.

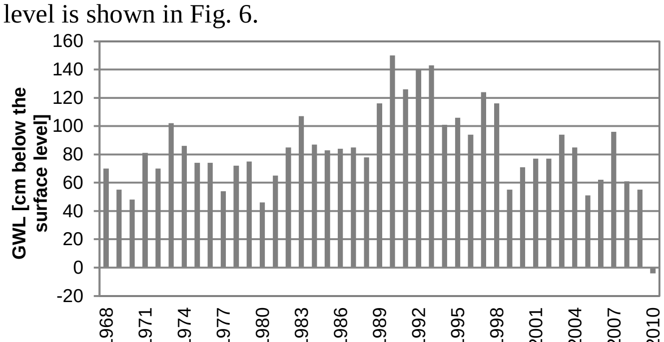

Fig. 5 Course of the average annual groundwater levels at the Šahy station for the period 1968-2010

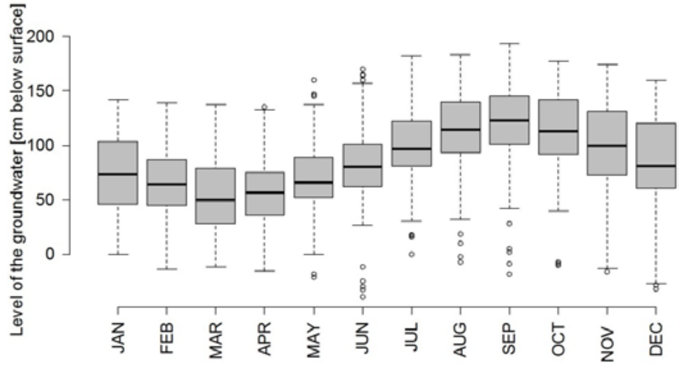

Fig. 6 Statistical evaluation of the groundwater level data and its fluctuations during a year

The input data used in the simulation of the soil water regime by the GLOBAL mathematical model is as follows: The upper boundary condition between the soil surface and the atmosphere was defined by meteorological data: the precipitation totals Z [mm], the average temperatures T [˚C], the totals of the duration of sunshine S [hours], the average atmospheric water vapor pressure p [hPa], and the average wind speed v_v [$m.s^{-1}$] for each day of the modeled period. The bottom boundary condition was the position of the groundwater level [cm] below the surface level. Because of the aim of this study, the measured groundwater levels were used in the first case, and the groundwater levels calculated with the SVM methodology were used in the second case. The soil in the model was characterized by its water retention curve (WRC) and saturated hydraulic conductivity. For this purpose soil samples from the Poiplie area were collected and evaluated in a laboratory. The points of the drainage branch of the WRC were calculated using pedotransfer functions. These points were further approximated by function according to Van Genuchten [28]. These functions or, rather, its parameters α and n, are other inputs of the GLOBAL model. The soil profile was divided into 3 layers. The soil parameters which were finally used as the inputs into the GLOBAL model are shown in Table 1. For the purposes of this analysis, 0-50 and 0-100 cm soil layers below the surface level were used.

Table 1 Physical and hydro-physical characteristics of the soil used in the simulation (θ_s - water content at saturation, θ_r - residual soil water content, K - saturated hydraulic conductivity, ρ_d - reduced bulk density, α, n - Van Genuchtens parameters of WRC)

Sampling depth [cm]	θ_s [$cm^3.cm^{-3}$]	θ_r [$cm^3.cm^{-3}$]	K [$cm.d^{-1}$]	ρ_d [$g.cm^{-3}$]	α [cm^{-1}]	n [-]
0-20	0.5340	0.084	17.8	1.4275	0.1328	1.0777
20-80	0.5543	0.090	18	1.2991	0.0419	1.0791
80-150	0.5403	0.1059	19.8	1.1702	0.1231	1.0690

The input parameters characterizing the vegetation cover include the leaf area index (LAI), the evaporating surface roughness, the albedo of the evaporating surface, and the root zone depth. The vegetation cover of soft mead with a predominance of poplars, which prevail in the investigated area, was considered within the simulation.

IV. RESULTS

While dealing with an evaluation of the impact of climate change on a soil water regime, scientists are logically confronted with a lack of input data. The same is true in the case of soil water regime modeling and the required data about the future positions of the groundwater levels. The authors of the paper recommend a methodology to obtain these data using support vector machines (SVMs). The groundwater levels are calculated on the basis of the air temperature, precipitation amount and supposed similarity of the future water regime (water level fluctuations) in the Ipel' River up to the present. The groundwater levels calculated using the SVM methodology are used as the lower boundary condition for the simulation of the soil water regime by the GLOBAL model; later in this chapter those computations are compared with the computations for the measured groundwater levels during the testing period.

The aim of this work is to evaluate the consequences of applying the SVM methodology to computing groundwater levels which should be inputs in the subsequent modeling of a soil water regime. For this purpose the authors evaluated the final precision of the modeling in the known (present) conditions, so a computation of the water regime in distant horizons, e.g., 2050 or 2100, as is usual in climate change impact studies, is not provided here. The aim of this study is to verify this possibility. For this reason the soil water storage was calculated: 1) using the measured groundwater levels, and was compared with 2) the modeling, which uses the values of the groundwater levels calculated by the SVM methodology. Through their analyses of the GLOBAL modeling results, the authors will demonstrate the sensitivity of the GLOBAL model to the precision of the groundwater level specification, e.g., its impact on the final precision of the soil water storage modeling. As has already been mentioned in the data description part, an important characteristic of applying an SVM to groundwater level predictions in climate change studies, which is also incorporated in this verification, is the data limitation. Therefore, only data which could be obtained from weather generators could be used.

The years 2000-2010 were chosen as the testing period. The other years were used for training the SVM (or GLOBAL model, respectively). The basic meteorological variables serve as the input (the temperatures and precipitation from the nearby Dudince measurement station). These variables are available from weather generators (or by other means), when the expected climate change study will be resolved. Because the groundwater levels for the purposes of this study were available in weekly time steps, the meteorological variables (precipitation and temperatures) were also summed respectively averaged to also have these values in the weekly time step. In addition to the values of these variables from the previous time step, their previous values from eight weeks before the predicted groundwater level were also used. This relatively long history was included in the inputs because the groundwater levels from the previous time step could not be used in the context of climate change studies as was hereinbefore explained. This meteorological history should compensate for this lack of such an important and usual predictant in time series modeling. One additional variable was constructed from the average monthly values of the water levels in the nearby gauging station on the Ipel' River. When a certain row of data belongs to a week in some month, the average value of the particular month's water level in the Ipel' is added, as this additional variable should represent the typical Ipel' River's influence on the groundwater in that time in the area. The averages are computed from the data in the training period; the same values are used in the testing period. It is presumed that there is no significant change in the water level regime in the Ipel' between the training and testing periods and also that it would not change in the period of years which would be investigated for climate change impacts. Of course, this is a simplification, and this feature of the proposed modeling could be improved in future studies. On the basis of the considerations just described, the final dataset has 18 features and 1,760 rows. It is not a very huge data set and is suitable for the proposed SVM modeling (from a practical point of view, perhaps the most serious problem with SVMs is the high algorithmic complexity and extensive memory requirements of the required quadratic programming in large-scale tasks).

The estimation of the practical steps of the SVM regression are as follows: 1) selecting a suitable kernel and the appropriate kernel's parameter; 2) specifying the ε (tube parameter in Equation (3)); and 3) specifying the capacity C (Equation 3).

The radial basis function was chosen as the kernel function on a trial and error basis:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma^* \left\|x_i - x_j\right\| 2\right), \quad \gamma \rangle 0 \tag{4}$$

The parameter γ of this kernel function, the tube size ε for the ε-insensitive loss function, and the parameter C should be found, which the crucial step is when SVMs are applied to a particular task [33]. A genetic algorithms (GA) heuristic search combined with a cross-validation methodology in its fitness function was used for finding the mentioned parameters. In this approach a set of SVM parameters generated by the genetic algorithms is sent to the parameter-evaluating algorithm. Basically, a k-fold cross validation was used. The data set was divided into k subsets, and the training-testing-evaluation was repeated k times. Each time, one of the k subsets is used as the test set, and the other k-1 subsets are put together to form a training set. Then the average error across all the k trials is computed, which is the SVM parameter's combination (called the "chromosome" in GA terminology) "fitness" in the context of the genetic algorithms.

In contrast to the usual approach, the cross-validation used in this study does not randomly split the data, but partitions defined in advance were used. This was motivated by the cyclic (sinusoidal) character of the groundwater level fluctuations (Fig. 6). The authors have divided the training set in that manner, so that every partition contains the data for one calendar year. This ensures that this typical seasonal fluctuation is included in each cross validation test set. Better results were obtained by this strategy in comparison with the standard approach in which the folds are formed randomly (verified by trial and error).

The authors do not document the following finding in the paper, but they realized that when searching of the parameters by genetic algorithm is used, better results were obtained in comparison with a standard grid search in which the possible values of the parameters are de facto predefined. The reason for this is that the search space in this case is a typical example of so-called equifinality. This means that there is not a unique optimal parameter set for which Beven [34] in the field of hydrological modeling introduced the term “equifinality”; this means the existence of multiple parameter sets, which are all acceptable, albeit not equivalent. In this situation a genetic algorithm is capable of finding one of these regions containing the “optimal solution” and then scans it more thoroughly, and thus finds a better solution because the search is not limited by grid settings. On the other hand, some mechanisms for avoiding overtraining and over-fitting were implemented - the maximum number of generations of the GA was set to 20 and the population size to 15 chromosomes, which is not abundant. Moreover, the GA is terminated if three subsequent generations without any improvement occurs.

The testing results of modeling the groundwater levels by SVM are in Fig. 7. The correlation coefficient was 0.721 and the mean absolute error 26.2. As a consequence of the data limitation, this degree of precision is not very high, but the question is what impact such a degree of accuracy has on the final soil water storage modeling.

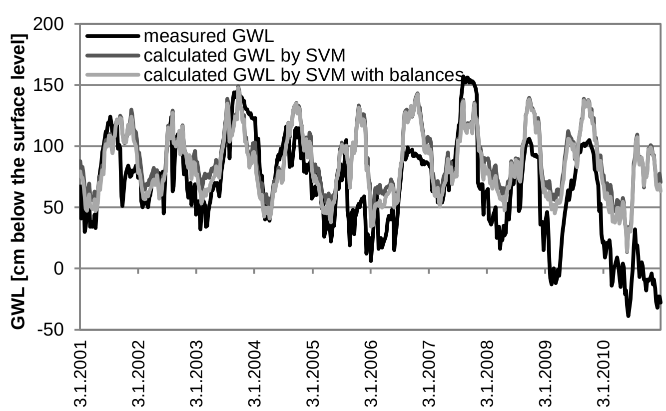

Fig.7 Comparison of the measured and computed groundwater levels in the period 2001-2010

In the next step modeling carried out by the GLOBAL model was accomplished and evaluated by a correlation analysis. In this second stage the authors compared the soil water storages 1) simulated when using as input the measured groundwater levels with the soil water storage and 2) simulated using the groundwater levels calculated by the SVM methodology. The results of this analysis are shown in Table 2. The correlation coefficients reached high levels, which mean that between the evaluated parameters is a high degree of commitment.

Table 2 Correlation of GLOBAL results produced by modelling with the measured GWL and computed by SVM

year	0-50 cm soil layer	0-100 cm soil layer
2001	0.87	0.71
2002	0.99	0.84
2003	0.97	0.98
2004	0.99	0.94
2005	0.89	0.79
2006	0.99	0.99
2007	0.99	0.99
2008	0.99	0.99
2009	0.81	0.92
2010	0.49	0.47
Average	**0.93**	**0.86**

As an illustration, the development of the soil water storage during the year 2007 is shown in Fig. 8 and 9, when the correlation coefficient had one of the highest values in each case of the evaluation (R=0.99). From the figures it can be seen that the soil water storage computations are very similar in the cases with the measured and computed groundwater levels. There are minimal differences in the 0-50 cm soil layer and in the 0-100 cm soil layer; there are some insignificant differences during the growing season.

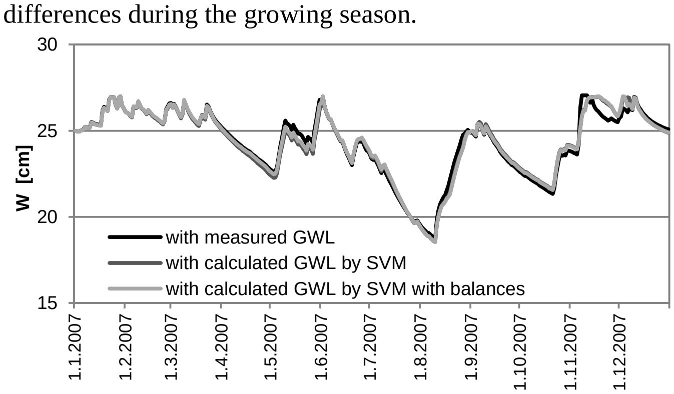

Fig. 8 Daily soil water storage in the 0-50 cm soil layer below the surface level for the year 2007

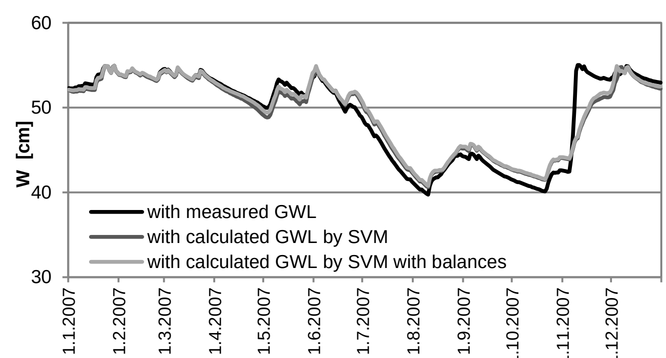

Fig. 9 Daily soil water storage in the 0-100 cm soil layer below the surface level for the year 2007

V. Conclusions

Predictions of soil moisture have drawn great interest from researchers, because soil moisture plays an important role in water resources planning and management. This parameter is important in research on climate, hydrology, agriculture, and forestry. This is also an important research issue in climate change impact studies. Different methods and models have been applied for predictions of soil water storage. They can be sorted into various categories; one possible classification distinguishes physically-based models and data-driven models. Physically-based models resolve exact physically justified governing equations. Data-driven models analyze and derive results only from observed inputs and outputs. This study has demonstrated the predictive value of a hybrid approach which involves both methods - SVM and GLOBAL - and underlines the usefulness of the proposed symbiosis of the data-driven and physically-based types of modeling.

In this paper the authors are evaluating the consequences of groundwater level (GWL) specifications on the subsequent modeling of a soil water regime, when these groundwater levels are obtained by support vector machines. Soil water storage was simulated by GLOBAL - a physically-based model - comparing cases, when the measured GWL and computed ones by the SVM methodology are taken as input. The results of this analysis are shown in Table 2. The accuracy of the final prediction of soil moisture between computations with measured versus computed GWL is very satisfactory; in the 0-100 cm soil layer there is a little less precision than in the 0-50 cm soil layer. This is caused due to the greater depth of the soil profile; consequently there are more complicated nonlinear relationships and probably a greater influence of the groundwater level from a neighbouring location. The GWLs are simulated by SVM because such data is rarely available in climate change impact studies, but this study successfully tested this approach.

Further studies are recommended, especially using more data for the prediction of the water flows of rivers in the area investigated or water levels in significant water bodies, which could have a significant influence on groundwater and could be taken as additional inputs. Nevertheless, the proposed methodology does show an acceptable degree of precision and could be an option in climate change impact studies. It is also desirable to evaluate the production of some other variables which are also not outputs of weather generators or to compare this approach with existing methodologies.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under Contract No. APVV-0139-10 and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, Grant No. 1/1044/11.

REFERENCES

[1] J. Majerčák, V. Novák, "GLOBAL, one-dimensional variable saturated flow model, including root water uptake, evapotranspiration structure, corn yield, interception of precipitations and winter regime calculation. Research report," Bratislava, 1994, 75 p.

[2] J. Benetin, et al., *Balance mathematical model for detailed analysis of the time variability of soil water regime components.* Water management journal, vol. 33, 1985,585-609.

[3] J. Šimůnek, et al., "The HYDRUS ET- Software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version1.1." U.S. Salinity Laboratory, USDA, ARS, Riverside, California and Institute of Hydrology SAS, Bratislava, Slovakia. IH SAS Bratislava, 1997.

[4] R. S. Kanwar, "Water table management and groundwater quality research at Iowa State University," Pap. Amer. Soc. Agric. Eng., 1990, 10 p.

[5] D. L. Thomas, P. G. Hunt, and J. W. Gilliam, *Water table management for water quality improvement*, Journal of Soil and Water Conservation 47(1), 1992, 65-70.

[6] R. L. Bengtson, R. S. Garzon, and J. L. Fouss, *A fluctuating water-table model for the management of a controlled-drainage subirrigation system*, Transaction of the ASAE 36 (2), 1993, 437-443.

[7] H. W. Belcher, and G. E. Merva, "Water table management at Michigan State University," Pap. Am. Soc. Agric. Eng., 1991, 19 p.

[8] S. Chuenchooklin, T. Ichikawa,S. Patamatamkul, CH. Kirdpitugsa, and V. Sriboonlue, "Surface water and phreatic aquifer interaction in floodplain with paddy field in Thailand, " in *Proc. of the 2006 IASME/WSEAS Int. Conf. on Water Resources, Hydraulics & Hydrology*, ISBN: 960-8457-44-0, Chalkida, Greece, 2006, pp.1-6

[9] G. Kim, J. Park, "Soil moisture estimation using classification and regression trees and neural networks, " in *Recent Researches in Hydrology, Geology and Continuum Mechanics*, ISBN: 978-960-474-275-2, WSEAS Press, 2011, pp. 57-60

[10] K. Kumarci, A. Ziaie and A. Kyioumarsi, "Land subsidence modeling due to ground water drainage using WTAQ software, " in *10th WSEAS Int. Conf. on AUTOMATIC CONTROL, MODELLING & SIMULATION (ACMOS'08)*, ISBN: 978-960-6766-63-3, Istanbul, Turkey, May 27-30, 2008.

[11] R. Lagace, R. W. Skaggs, and J. E. Parsons, *Predicting water table drawdown for two-dimensional drainage Boussinesq equations*, Advances in drainage: proceedings, Fourth National Drainage Symposium 13-14, 1982, 6-15.

[12] Singh, R. N., Rai, S. N. and Ramaua, R. N.: Water-table fluctuation in a sloping aquifer with transient recharge, Journal of Hydrology 126 (3-4), 315-326, (1991)

[13] S. N. Rai, and R. N. Singh, *Water-table fluctuations in an aquifer system owing to time-varying surface infiltration and canal recharge*, Journal of Hydrology 136(1-4), 1992, 381-387.

[14] I. S. Teloglou, T. S. Zissis, and D. N. Karamouzis, *Water table fluctuation in soils overlying semi-confined aquifers in response to time-varying replenishment*, Journal of Hydrology 202(1-4), 1997, 78-94.

[15] S. N. Rai, and A. Manglik, *Modeling of water table variation in response to time-varying recharge from multiple basins using the linearlized Boussinesq equation*, 220(3/4), 1999, 141-148.

[16] A. Manglik, and S. N. Rai, *Modeling of water table fluctuations in response to time-varying recharge and withdrawal*, Water Resources Management 14(5), 2000, 339-347.

[17] A. Upadhyaya, and H. S. Chauhan, *Water table fluctuations due to canal seepage and time varying recharge*, Journal of Hydrology 244(1/2), 2001, 1-8.

[18] Z. Song, L. Li, J. Kong, and H. Zhang, *A new analytical solution of tidal water table fluctuations in a coastal unconfined aquifer*, Journal of Hydrology 340(3-4), 2007, 256-260.

[19] C. C. Yang, S. O. Prasher, and R. Lacroix, *Applications of artificial neural networks land drainage engineering*, Trans. ASAE 39(2), 1996, 525-533.

[20] P. Coulibaly, F. Anctil, R. Aravena, and B. Bobee, *Artificial neural network modeling of water table depth fluctuations*, Water Resource Research 37(4), 2001, 885-896.

[21] F. Szidarovszky, F.E. Coppola Jr., J. Long, A.D. Hall, M.M Poulton, *A Hybrid Artificial Neural Network-Numerical Model for Ground Water Problems.* Ground Water 45, no. 5, 2007, 590-600.

[22] V. Vapnik, S. Golowich, and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, Submitted to Advances in Neural Information Processing Systems," 1996.

[23] T. Asefa, M. W. Kemblowski, G. Urroz, M. McKee, and A. Khalil, *Support vectors-based ground water head observation networks design.* Water Resour. Res. 40, 2004, W11509.

[24] A. Khalil, M. McKee, M. W. Kemblowski, and T. Asefa, *Sparse Bayesian learning machine for real-time management of reservoir releases.* Water Resour. Res. 41, 2005, W11401.

[25] M. K. Gill, T. Asefa, M.W. Kemblowski, M. McKee, *Soil moisture prediction using support vector machines.* J. Am. Water Resour. Assoc. 42, 2006, 1033–1046.

[26] M. K. Gill, T. Asefa, Y. Kaheil, M. McKee, *Effect of missing data on performance of learning algorithms for hydrologic predictions: implications to an imputation technique.* Water Resour. Res. 43, 2007, W07416, doi:10.1029/2006WR005298.

[27] L. A. Richards, *Capillary conduction of liquids through porous media*. Physics, vol. 1, no. 5, 1931, 318-333.

[28] M., Th. Van Genuchten, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils.* Soil Sci. Soc. Am. J. 44, 1980, 987 – 996.

[29] Y. Mualem, *A new model for predicting the hydraulic conductivity of unsaturated porous media.* Water Resour. Res., 12 (3), 1976, 513-522.

[30] L. Wang, *Support Vector Machines: Theory and Applications*. Springer-Verlag, Berlin Heidelberg, 431 p., (2010), ISBN 978-3-540-24388-5.

[31] V. Vapnik, "Statistical Learning Theory." John Willey & Sons, 1998.

[32] A.J. Smola, B. Scholkopf, "A Tutorial on Support Vector Regression." NeuroCOLT Technical Report TR Royal Holloway College, London, 1998.

[33] H. B. Alwan, K. R. Ku-Mahamud, "Incremental Continuous Ant Colony Optimization Technique for Support Vector Machine Model Selection Problem," *Mathematical Methods for Information Science and Economics*, ISBN: 978-1-61804-148-7, 2012, pp.165-170.

[34] K. J. Beven, *Prophecy, reality and uncertainty in distributed hydrological modelling*, Adv. Water Resource. 16, 1993, 41–51.

[35] J. Skalová, "Pedotransfer functions of Záhorská lowland soil and their application in modeling of soil water regime. Inaugural dissertation." Slovak Technical University in Bratislava, Slovakia, 2001.

[36] L. A. Richards, *Capillary conduction of liquids through porous media*. Physics, vol. 1, no. 5, 1931, 318-333.

[37] M. Gomboš, A. Tall, *Monitoring of water content in the mould layer of clay soil profile.* Acta Hydrologica Slovaca. vol. 2, no.1., 2001, 29-34.

[38] V. Štekauerová, J. Skalová, K. Nováková, *Assignment of hydrolimits for estimation of soil ability to supply plants by water.* Növénytermelés - Crop production. Supplement, vol. 59, 2010, 195-198.

[39] J. Šútor, J. Majerčák, V. Štekauerová, *Quantification of water storage in soil aeration zone in agricultural ecosystems using of data from simulation (in Slovak).* Acta Hydrologica Slovaca, vol. 1, no. 2, 2001, 72-77.