Selection of Catchment Descriptors for the Physical Similarity Approach. Part II: Application

Martin HEŘMANOVSKÝ and Pavel PECH

Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

Abstract

Heřmanovský M., Pech P. (2013): **Selection of catchment descriptors for the physical similarity approach. Part II: Application**. Soil & Water Res., **8**: 186–194.

This paper demonstrates an application of the previously published method for selection of optimal catchment descriptors, according to which similar catchments can be identified for the purpose of estimation of the Sacramento – Soil Moisture Accounting (SAC-SMA) model parameters for a set of tested catchments, based on the physical similarity approach. For the purpose of the analysis, the following data from the Model Parameter Estimation Experiment (MOPEX) project were taken: a priori model parameter sets used as reference values for comparison with the newly estimated parameters, and catchment descriptors of four categories (climatic descriptors, soil properties, land cover and catchment morphology). The inverse clustering method, with Andrews' curves for a homogeneity check, was used for the catchment grouping process. The optimal catchment descriptors were selected on the basis of two criteria, one comparing different subsets of catchment descriptors of the same size (*MIN*), the other one evaluating the improvement after addition of another catchment descriptor (*MAX*). The results suggest that the proposed method and the two criteria used may lead to the selection of a subset of conditionally optimal catchment descriptors from a broader set of them. As expected, the quality of the resulting subset of optimal catchment descriptors is mainly dependent on the number and type of the descriptors in the broader set. In the presented case study, six to seven catchment descriptors (two climatic, two soil and at least two land-cover descriptors) were identified as optimal for regionalisation of the SAC-SMA model parameters for a set of MOPEX catchments.

Keywords: a priori SAC-SMA model parameters; catchment characteristic; MOPEX catchments; regionalisation

The application of conceptual hydrological models in ungauged catchments is limited, due to the lack of rainfall-runoff time series for model calibration (Sivapalan *et al*. 2003). This problem instigated development of different methods for bypassing model calibration. One way is to identify appropriate mathematical relationships between the model parameters and some catchment descriptors (CDs) and then to apply these relationships in an ungauged catchment (Magette *et al*. 1976; Jakeman *et al*. 1992; Post *et al*. 1998). Another option is to search for the nearest gauged catchment (in the geographical sense or in terms of similarity of CDs) and then to transfer a model parameter set from the most similar gauged catchments to an ungauged one (Parajka *et al*. 2005; Oudin *et al*. 2008). In the following text, only the last of the options mentioned, i.e. the search for nearest similar gauged catchments defined in terms of a subset of CDs, will be addressed.

Why do we search for gauged catchments most similar to an ungauged one in terms of CDs? The answer draws on the assumption that the uniqueness of each catchment can be captured in unique combination of its descriptors (Wagener *et al*. 2004).

Since there is no established theory how to select an optimal subset of CDs, a trial-and-error approach is usually adopted. In this approach a wide range of available CDs is tested and an optimal subset of them is chosen from their all possible combinations (Parajka *et al.* 2005; Oudin *et al.* 2008). Certain disadvantages of the trial-and-error method are associated with the fact that a large number of CD subsets must be tested, while the importance of particular descriptors is not investigated (only the effect of the entire subset of CDs is tested). Heřmanovský and Pech (2013) proposed a method that is also partially based on the trial-and-error approach, but reduces the total number of the CD subsets tested, using a consecutive selection of CDs with the strongest influence. This method can also provide information about the effect of a particular CD on the estimates of a particular model parameter, as well as on the estimate of the whole model parameter set.

In this paper, the method proposed by Heřmanovský and Pech (2013) is applied to a strongly heterogeneous set of catchments of the Model Parameter Estimation Experiment (MOPEX) project and the associated set of CDs. A conditionally optimal set of CDs for these catchments is identified and the influence of the CDs selection on the accuracy of Sacramento – Soil Moisture Accounting (SAC-SMA) model parameter estimates is demonstrated. The first part of the paper describes the MOPEX data set, while the results of the analysis are presented in the second part. These results are discussed in the third part. The goal of this case study is to demonstrate the procedure on the data sets that are to some extent artificial. Further studies are needed in which a hydrological model with really calibrated parameters would be used.

MATERIAL AND METHODS

Datasets. An analysis was performed using the data from the MOPEX project (Duan *et al.* 2006). These data contain a few categories of catchment descriptors (CDs) and a priori parameter sets of the modified (11-parameter) version of the SAC-SMA model (Burnash 1995) for 438 catchments in the USA. The a priori parameters were identified by means of the methods proposed by Koren *et al.* (2003), who developed a set of physically based relationships between the SAC-SMA parameters and the United States Department of Agriculture (USDA) soil properties (Miller & White 1999). Due to these relationships, it is known how the a priori model parameters depend on soil properties. For our purposes, the a priori parameter sets were considered as "true" optimal model parameters for each catchment and, as such, they were used for comparison with the newly estimated parameter sets. Table 1 shows a list of the SAC-SMA model parameters with their units.

A large set of CDs was available for the catchment grouping process. Four CD categories were used, namely, the climatic, soil, land-cover and catchment morphology descriptors. Table 2 shows an overview of them and explains the symbols used. Some CDs (*SS*, *LS*, *CS*, *Fo*, *Ur*, *Gr* and *Cr*) were aggregated in order to reduce their number and to increase the variability of their values.

The category of climatic CDs included two basic descriptors (P_{aa} and PE_{aa}) and two derived ones (P_{aa}/PE_{aa} and E_{aa}/PE_{aa}). The P_{aa} values were com-

Table 1. Symbols, descriptions and units of parameters of the modified Sacramento – Soil Moisture Accounting (SAC-SMA) model

Parameter symbol	Parameter description
UZTWM	upper-zone tension water capacity (maximum) (mm)
UZFWM	upper-zone free water capacity (maximum) (mm)
LZTWM	lower-zone tension water capacity (maximum) (mm)
LZFPM	lower-zone primary free water capacity (maximum) (mm)
LZFSM	lower-zone supplemental free water capacity (maximum) (mm)
UZK	fractional daily upper-zone free water withdrawal rate (–)
LZPK	fractional daily primary withdrawal rate (–)
LZSK	fractional daily supplemental withdrawal rate (–)
ZPERC	maximum percolation rate coefficient (–)
REXP	percolation equation exponent (–)
PFREE	fraction of percolated water going directly to lower-zone free water storage (%)

Table 2. Catchment descriptors (CD) tested within particular categories

Categories	Catchment descriptors tested
Climatic descriptors	mean annual precipitation, P_{aa} (mm)
	mean annual potential evaporation, PE_{aa} (mm)
	ratio of mean annual precipitation to mean annual potential evaporation, P_{aa}/PE_{aa} (–)
	ratio of mean annual evaporation to mean annual potential evaporation, E_{aa}/PE_{aa} (–)
Soil descriptors	porosity, *Po* (–)
	wilting point, *Wp* (–)
	saturated hydraulic conductivity, *SHC* (m/s)
	fraction of sandy soils, *SS* (–)
	fraction of loam soils, *LS* (–)
	fraction of clay soils, *CS* (–)
Land-cover descriptors	fraction of forest cover, *Fo* (–)
	fraction of cropland, *Cr* (–)
	fraction of grassland, *Gr* (–)
	fraction of urban areas, *Ur* (–)
	greenness factor, G_f (–)
Morphological descriptors	catchment area, *Ac* (km^2)

puted using PRISM (Parameter-elevation Regressions on Independent Slopes Model) (Daly *et al.* 1994) for the period 1961–1990. The PE_{aa} and E_{aa} values were computed using the NOAA Freewater Evaporation Atlas (Farnsworth *et al.* 1982).

The soil hydraulic descriptors (*Po*, *Wp* and *SHS*) were derived for each catchment from 1-km STATSGO soil texture data provided by the Penn State Earth System Science Center (Miller *et al.* 1999), using empirical relationships (Clapp & Hornberger 1978; Cosby *et al.* 1984). The soil cover parameters *SS*, *LS* and *CS* were obtained by aggregation of 12 USDA soil texture types that occurred in the MOPEX catchments – with *SS* comprising two soil texture types, *LS* four types and *CS* six types.

The IGBP (International Geosphere-Biosphere Programme) classification (Loveland *et al.* 2000) was a basis for determination of land-cover descriptors. The *Fo* descriptor was obtained as an aggregate of six land-cover descriptors, while *Cr* comprised two and *Gr* three land-cover descriptors. The descriptor of greenness factor outside the vegetation season (in February) was also included in the third category to represent the evergreen to deciduous forest fraction. The category of morphological CDs included only the catchment area (*Ac*).

In total, we must admit that this set of CDs may not be full enough, because it does not contain enough morphological and geological descriptors. Hence, the subsets discussed below as optimal CDs can be only considered as such with respect to the set of CDs available, i.e. as only conditionally optimal.

RESULTS

Analysis of input data. This part of the paper provides an analysis of the inputs. Pairwise correlation coefficients (*CC*) were computed, relating the model parameters to one another, CDs with one another and the model parameters with CDs. A strong correlation was found between some model parameters (e.g. the pairs UZK–LZSK with *CC* = 0.991 and REXP–PFREE with *CC* = 0.975).

A strong correlation was also found between the basic climatic CDs and the derived ones and between some soil descriptors (e.g. the pairs P_{aa}/PE_{aa}–E_{aa}/PE_{aa} with *CC* = 0.881 and *SS*–*SHC* with *CC* = 0.868). Furthermore, a strong correlation was found between the soil CDs and some model parameters (e.g. the pairs *Wp*–UZK with *CC* = –0.938 and *Wp*–LZSK with *CC* = –0.933).

Determination of the optimal threshold value r_t^2. In this part of the analysis, the optimal threshold value of the coefficient of determination (r_t^2) for comparison of Andrews' curves is found. The theory based on Andrews' curves was elaborated by Heřmanovský and Pech (2013). The coefficient of determination r_t^2 quantifies the distance between CDs of two catchments. It is used to define the region, i.e. a group of catchments (regarded as gauged) similar to a particular catchment (regarded as ungauged) for which the parameters are to be estimated from those of the similar but gauged catchments. Within each such region, each catchment is in turn considered as ungauged and its parameters are estimated from those of the other catchments in the region, regarded as gauged. The higher the coefficient of determination, the smaller the size of the region of similar catchments. For this study, the following seven CDs were used: P_{aa}, PE_{aa}, CS, LS, G_f, Fo and Cr. They were selected on the basis of previous studies, e.g. Nathan and McMahon (1990), Xu (1999), Parajka *et al.* (2005) and Oudin *et al.* (2008), which indicate that these CDs are significant for runoff formation. The r_t^2 values tested were taken from the interval <0.00; 0.95> with a step 0.05.

The results obtained are shown in Figures 1–3. Figures 1 and 2 show the median ($\tilde{D}(\theta_i)$) and few other typical quantiles of the absolute values of relative deviations (in %) of the estimated values of a parameter θ_i for the ungauged catchments from the optimal (a priori) values of the same parameter, related to the coefficient of determination r_t^2 which defines the size of the region of similar catchments. This relation is demonstrated for the models parameters ZPERC, UZK and

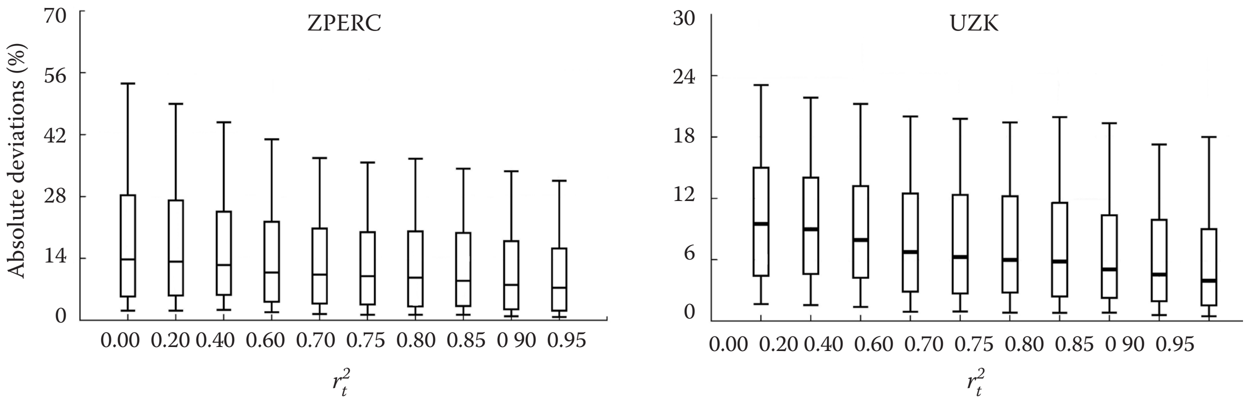

Figure 1. The quantiles (10th, 25th, 50th = $\tilde{D}(\theta_i)$, 75th and 90th percentiles) of the absolute values of relative deviations between the estimated and optimal model parameters for the parameters ZPERC (maximum percolation rate coefficient) and UZK (fractional daily upper-zone free water withdrawal rate), depending on the threshold value of the coefficient of determination r_t^2, for 438 tested catchments and seven CDs

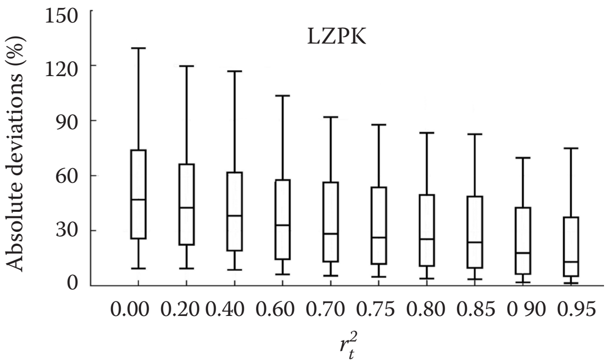

Figure 2. The quantiles (10th, 25th, 50th = $\tilde{D}(\theta_i)$, 75th and 90th percentiles) of the absolute values of relative deviations between the estimated and optimum values of the model parameter LZPK (fractional daily primary withdrawal rate), depending on the threshold value of the coefficient of determination r_t^2, for 438 tested catchments and seven CDs

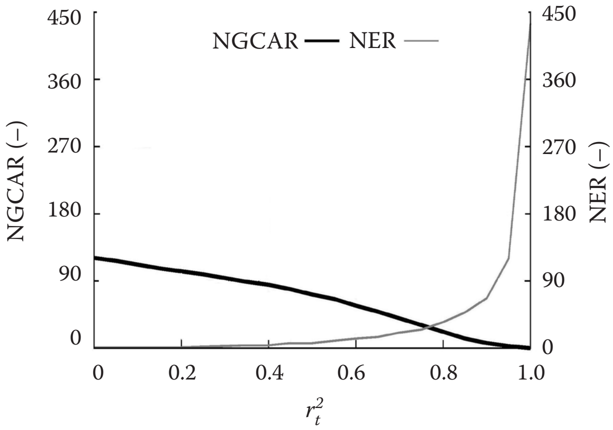

Figure 3. The median of the number of gauged catchments assigned to regions (NGCAR) and the number of empty regions (NER), depending on the threshold value of the coefficient of determination r_t^2, for 438 tested catchments and seven CDs

LZPK, representing three groups of parameters that behave differently in terms of the decrease of their $\widetilde{D}(\theta_i)$ and other quantiles with the increase of r_t^2. Note that the points of the horizontal axes of Figures 1 and 2 are not equidistant.

The first group comprises the parameters UZTWM and ZPERC, for which a similar decrease pattern of $\widetilde{D}(\theta_i)$ and other quantiles along with the increase in r_t^2 can be noted – an insignificant decrease in the interval r_t^2 ∈ <0.00; 0.60>, followed by a significant decrease in the interval r_t^2 ∈ <0.65;0.95> (for ZPERC behaviour, see Figure 1). The second group contains the parameters UZFWM, LZTWM, LZFPM, LZFSM, UZK and LZSK. Significant decrease in $\widetilde{D}(\theta_i)$ and other quantiles of these parameters is detected over the whole tested interval <0.00; 0.95> of r_t^2 (for the behaviour of UZK, see Figure 1). The third group contains the parameters LZPK, REXP and PFREE, for which a significant decrease of $\widetilde{D}(\theta_i)$ and other quantiles with increasing r_t^2 occurs in the intervals r_t^2 ∈ <0.00; 0.60> and r_t^2 ∈ <0.85; 0.95>, while their decrease is insignificant in the intermediate interval r_t^2 ∈ <0.65; 0.80>. For the behaviour of LZPK see Figure 2.

As already noted, the number of gauged catchments assigned to particular regions (NGCAR) decreases with the increase in the coefficient of determination r_t^2 value. This decrease is presented in Figure 3 in terms of the median of NGCAR and the number of empty regions (NER) depending on r_t^2. In Figure 3 we can see a relatively slow and regular decrease in the median of NGCAR value over the interval r_t^2 ∈ <0.00; 0.55>, followed by a more significant decrease in the interval r_t^2 ∈ <0.60; 0.85> and again a slower decrease in the interval r_t^2 ∈ <0.90; 0.95>. The number of the regions that do not contain any gauged catchment similar to the ungauged one (NER) is marginal in the interval r_t^2 ∈ <0.00; 0.75>. For example, there are only 25 empty regions from the total of 438 regions at r_t^2 = 0.75. However, a very significant increase in the number of empty regions was found in the interval r_t^2 ∈ <0.80; 0.95>. For example, 120 regions out of 438 are empty at r_t^2 = 0.95.

The results obtained suggest that the optimal r_t^2 values fall within the interval r_t^2 ∈ <0.80; 0.90>, where the minimum $\widetilde{D}(\theta_i)$ values, i.e., the minimum deviations of the estimated parameters from their optimal (a priori) values, were achieved. At the same time, the number of empty regions in this interval is still relatively low. The median of the size of the region of similar catchments (NGCAR) in this interval ranges from 3 to 13 gauged catchments.

Searching for optimal catchment descriptors. In this part of the analysis, a total of 438 MOPEX catchments were tested. The threshold value of coefficient of determination was chosen as r_t^2 = 0.90. Heřmanovský and Pech (2013) proposed two criteria (*MIN* and *MAX*), both associated with the median of deviation of parameters, $\widetilde{D}(\theta_i)$, for the identification of optimal CDs. The weights in the equations defining *MIN* and *MAX* were chosen in this particular study as follows: (1) for $\widetilde{D}(\theta_i)$ in the interval (0.00; 10.00] % the weight w_i is 0.01; (2) for $\widetilde{D}(\theta_i)$ in the interval (10.00; 20.00] % the weight w_i is 0.50; (3) for $\widetilde{D}(\theta_i)$ in the interval (20.00; +∞) % the weight w_i is 1.00.

The reasons for this choice relate to the equations for computation of *MIN* and *MAX*. The *MIN* value is calculated on the basis of the lowest $\widetilde{D}(\theta_i)$ value among all tested subsets of CDs of the same size. Without the weights, the more accurately estimated parameters (the deviations of which from the optimal values are low) would contribute more significantly to the overall *MIN* value than the less accurately estimated parameters (the deviations of which from the optimal values are high). This could bias the selection of the optimal subset of CDs. Similar reasoning led us to the adoption of the same weights for the calculations of *MAX*. Optimal subsets of CDs were searched for according to the algorithm described by Heřmanovský and Pech (2013).

The results of this part of the analysis are shown in Table 3 and Figures 4 and 5. The optimal subset of CDs consists of two climatic CDs (PE_{aa}, E_{aa}/PE_{aa}), two soil characteristics (*Po*, *LS*) and three land-cover descriptors (G_f, *Ur*, *Gr*). The search for an optimal subset of CDs was terminated after finding

Table 3. Changes in the *MIN* and *MAX* values depending on the increase in the number of catchment descriptors (CDs) in the optimal subset

	Number of CDs						
	2	3	4	5	6	7	8
MIN	18.400	28.022	11.667	2.308	0.253	0.049	0.131
MAX	–	29.224	9.619	1.408	–0.362	1.729	–1.090

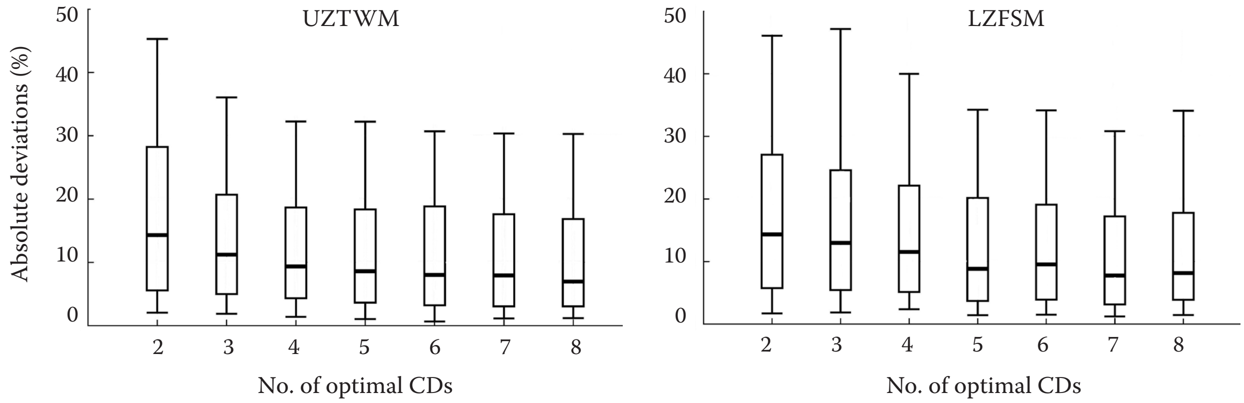

Figure 4. The quantiles (10^{th}, 25^{th}, $50^{th} = \tilde{D}(\theta_i)$, 75^{th} and 90^{th} percentiles) of the absolute values of relative deviations between the estimated and optimal model parameters for the parameters UZTWM (upper-zone tension water capacity) and LZFSM (lower-zone supplemental free water capacity), depending on the number of catchment descriptors (CDs) in the optimal subset, for 438 tested catchments with $r_t^2 = 0.90$

the seventh significant catchment descriptor. The eighth significant catchment descriptor (*Cr*) was subsequently identified in order to assess whether the search for an optimal subset of CDs was not terminated prematurely (the results are included in Figures 4 and 5 and are mentioned in text below).

Figures 4 and 5 show the medians ($\tilde{D}(\theta_i)$) and other selected quantiles of the absolute values of relative deviations (in %) of the estimated model parameters from their optimal (a priori) values, depending on the increasing size of the subset of optimal CDs. The model parameters can be divided into two groups according to similar patterns of decrease in $\tilde{D}(\theta_i)$ with the increase in the size of the subset of optimal CDs.

The first group contains eight model parameters (UZTWM, UZFWM, LZTWM, LZFSM, LZFPM, ZPERC, REXP and PFREE). A significant decrease in the $\tilde{D}(\theta_i)$ value is characteristic up to the addition of the fourth (in the case of UZTWM – Figure 4, and ZPERC – not shown), or the fifth (in the case of LZFSM – Figure 4, PFREE – Figure 5), UZFWM, LZTWM, LZFPM and REXP – not shown) CD to the optimal subset. The addition of the fifth (or sixth) to the eighth CD to the optimal CD subset only resulted in a marginal decrease in $\tilde{D}(\theta_i)$.

The second group comprises the model parameters UZK, LZPK and LZSK. A very significant decrease was observed (for UZK, see Figure 5) up to the addition of the third CD to the optimal subset. Negligible changes in $\tilde{D}(\theta_i)$ occurred after addition of the fourth to the eighth CD to the optimal subset.

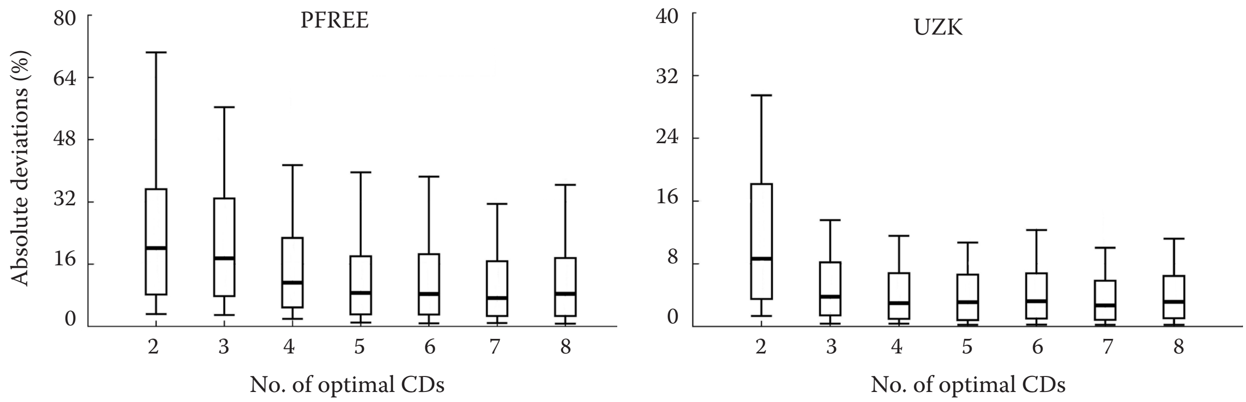

Figure 5. The quantiles (10^{th}, 25^{th}, $50^{th} = \tilde{D}(\theta_i)$, 75^{th} and 90^{th} percentiles) of the absolute values of relative deviations between the estimated and optimum model parameters for the parameters PFREE (fraction of percolated water going directly to lower-zone free water storage) and UZK (fractional daily upper-zone free water withdrawal rate), depending on the number of catchment descriptors (CDs) in the optimal subset, for 438 tested catchments with $r_t^2 = 0.90$

Table 3 presents the *MIN* and *MAX* values (summed over all model parameters) depending on the increasing size of the subset of optimal CDs. We can note an increase in the *MIN* value after the addition of the third CD to the optimal subset and a significant decrease in the *MIN* value up to the addition of the fifth CD. The addition of the sixth to the eighth CD caused only marginal changes in the *MIN* value.

A significant decrease in *MAX* was observed up to the addition of the fifth CD to the optimal subset. The addition of the sixth to the eighth CD led either to a negative *MAX* value (for the sixth and the eighth CD added) or to a relatively large positive *MAX* value (for the seventh CD added).

The results obtained show that, for the MOPEX data analysed, six to seven CDs from at least three different CD categories (climatic descriptors PE_{aa} and E_{aa}/PE_{aa}, soil descriptors *LS* and *Po* and at least two of the land-cover descriptors G_f, *Ur* and *Gr*) can define a sufficiently homogeneous region, in which accurate enough estimates of model parameters can be achieved. It is obvious from results presented in Figures 4 and 5 that the searching process was terminated correctly (i.e. after finding the seventh optimal CD), because any further improvements of the model parameter estimates after finding the eighth significant CD were negligible. Moreover, it is seen that even the improvements after finding the seventh significant CD were slight. Therefore, the searching process can be safely terminated after finding the sixth significant CD. For comparison, Oudin *et al.* (2008) reported that three to five catchment descriptors was the optimal number for a catchment grouping process.

DISCUSSION

Input data. The results presented above indicate that there are strong correlations among the SAC-SMA model parameters. These strong correlations arise, in particular, from the way in which the a priori model parameters are estimated. It is due to these correlations that some model parameters reveal similar patterns of decrease in $\tilde{D}(\theta_i)$ with the increasing r_t^2 or with the increasing the size of the optimal subset of CDs.

Strong correlations between the derived climatic descriptors (P_{aa}/PE_{aa} and E_{aa}/PE_{aa}) and the basic ones (P_{aa} and PE_{aa}) are explainable by the method of determination of the derived descriptors. Wagener *et al.* (2004) argue that the derived climatic CDs should be removed from an optimal CD set because they contain the same information as the basic climatic descriptors. Our optimal CD set (see the Results chapter) contains one basic climatic descriptor PE_{aa} and one derived climatic descriptor E_{aa}/PE_{aa}, between which a strong correlation was found. The reason for maintaining a derived CD is that the information carried by the derived CD is more explicit.

When we analyse the soil hydraulic descriptors – the porosity (*Po*), the wilting point (*Wp*) and the saturated hydraulic conductivity (*SHC*), we arrive at a conclusion that the strong positive correlation between *Po* and *Wp* probably relates to the content of clay fraction in the soil, on which both these descriptors depend. The values of *Po* increase, as well as the values of *Wp*, with the increase in the content of clay particles (< 2 μm) up to about 60% (Hillel 1998; Kutílek & Nielsen 1994). This effect is presumably related to a higher amount of fine capillary and subcapillary pores that have a high capacity to strongly retain soil water. The same argument can be used in relation to the strong negative correlation between the *SHC* and *Po* descriptors. On the other hand, the strong correlations among the soil hydraulic descriptors *Po*, *Wp* and *SHC* are probably also related to the way in which they were derived, namely, by using empirical relationships based on the soil texture (Clapp & Hornberger 1978; Cosby *et al.* 1984).

Strong correlations among the soil descriptors (*SS*, *CS*, *Po*, *Wp* a *SHC*) and some model parameters (e.g. the fractional daily primary and supplementary withdrawal rates LZPK and LZSK, the parameters of percolation equation ZPERC and REXP and the PFREE parameter) arise due to the relationships used for estimation of the a priori model parameters, in which some of the soil characteristics appear. It is for this reasons that the largest improvement of accuracy of model parameter estimates was obtained after adding the soil descriptors to the subset of optimal CDs.

Threshold coefficient of determination r_t^2 and the number of optimal CDs. It is interesting that the median deviations $\tilde{D}(\theta_i)$ for the model parameters LZTWM, UZK, LZSK and REXP are very small even at low r_t^2 values (e.g. $\tilde{D}(\theta_i)$ < 10% for r_t^2 = 0), while it was a priori expected that the $\tilde{D}(\theta_i)$ values at low r_t^2 should be very high for all model parameters, because the regions of similar catchments formed at low r_t^2 values should be considerably heterogeneous, and a decrease in $\widetilde{D}(\theta_i)$ should be associated with the increase in r_t^2, because in this way the homogeneity of the regions is supposed to increase. However, a change in $\widetilde{D}(\theta_i)$ for the above-mentioned four parameters was not significant over the whole interval of r_t^2, although the concurrent changes in the number of the gauged catchments within the region were substantial.

This paradox could be attributed to several factors. The first factor is the existence of a strong correlation between the model parameters LZTWM, UZK, LZSK and REXP on the one hand and some CDs (namely, the soil descriptors) on the other hand. The regions formed with these descriptors at low r_t^2 may show a significant similarity among the catchments in terms of the model parameters mentioned. Consequently, we can expect relatively accurate estimates of these model parameters even at low r_t^2 values. On the other hand, the regions formed with these descriptors at low r_t^2 values may be considerably heterogeneous in terms of other model parameters.

The second factor is the small overall variability of the model parameters LZTWM, UZK, LZSK and REXP (this problem, related to a priori SAC-SMA parameters, was discussed by Gan and Burges (2006)). The coefficients of variation (CV) for the above-mentioned model parameters were low in comparison with the other model parameters. The lowest CV = 0.16 was found for REXP, while the highest CV = 0.97 was computed for LZPK). This indicates a lower variability of LZTWM, UZK, LZSK and REXP in comparison with the other model parameters. Therefore, a region formed at low r_t^2 may consist of catchments that are little different according to these parameters, although they can be very different in terms of the CDs used and belong to a single region only because of the low r_t^2 threshold.

Another explanation is related to the fact that the first two climatic CDs (PE_{aa} and E_{aa}/PE_{aa}) identified as optimal are important overall climate characteristics that affect the catchment descriptors of the other categories (soil characteristics and land cover). Therefore, the catchment similarity defined on the basis of these climatic descriptors may lead to regions in which the catchments are also similar according to other catchment descriptors.

Some remarks can be made on the variation of *MIN* and *MAX* depending on the increase in the number CDs in the optimal subset (Table 3). The decrease in *MIN* and *MAX* with the increased number of CDs is primarily caused by the gradual improvement of the model parameter estimates. This decrease is most pronounced in the initial phase of the search (for the first two to five CDs) and less perceivable in its final phase (for the sixth to the eighth CD).

The negative *MAX* value obtained after addition of the sixth optimal CD (Table 3) indicates that the estimate of at least one model parameter did not improve. In this case, the total *MAX* value was significantly influenced by the partial *MAX* value (a single term of the sum) for LZPK. The median deviation $\widetilde{D}(\theta_i)$ of this model parameter increased after addition of the urban area fraction (*Ur*).

CONCLUSIONS

The algorithm presented by Heřmanovský and Pech (2013) was applied to a heterogeneous set of catchments previously used in the MOPEX project. The results obtained show that the algorithm may lead to a reasonable subset of catchment descriptors (CDs), suitable for regionalisation based on the physical similarity approach. The identified conditionally optimal subset of CDs was used in the catchment grouping process. Within each region of similar catchments, created on this basis, the parameters of the SAC-SMA model were estimated as weighted averages of model parameters of gauged catchments in the region.

The result of selection of optimal CDs mainly depends on the number and type of CDs that are available in the basic set. It is possible that a repeated application of the same algorithm with the same catchment set and the same model parameters but with a different set of basic set of CDs will lead to a different set of optimal CDs. It is also possible that the subset of CDs identified in this study as optimal is only optimal for the estimation of parameters of the SAC-SMA model and for the catchment set used. Notwithstanding this reservation, the proposed scheme can be applied to other catchment sets and other hydrological models of interest.

In this study, only six to seven catchment descriptors, belonging to three categories (climatic – PE_{aa}, E_{aa}/PE_{aa}, soil – *LS*, *Po* and land cover – G_f, *Ur* and *Gr*) were identified as optimal. These optimal catchment descriptors were similar to the descriptors used in other regionalisation studies, e.g. Nathan and McMahon (1990), Parajka *et al.* (2005) and Oudin *et al.* (2008). However, choosing the a priori SAC-SMA model parameters derived from soil properties as optimal led to overestimation of the role of soil CDs.

Acknowledgements. Supported by the Ministry of the Interior of the Czech Republic, Project No. BV II/2-VS. The authors would like to thank reviewers for their constructive suggestions, which helped them to improve the text.

References

Burnash R.J.C. (1995): The NWS river forecast system – catchment modelling. In: Singh V.P. (ed.): Computer Models of Watershed Hydrology. Water Resources Publication, Highlands Ranch, 69–118.

Clapp R.B., Hornberger G.M. (1978): Empirical equations for some soil hydraulic properties. Water Resources Research, **14**: 601–604.

Cosby B.J., Hornberger G.M., Clapp R.B., Ginn T.R. (1984): A statistical relationship of soil moisture characteristics to the physical properties of soils, Water Resources Research, **20**: 682–690.

Daly C., Neilson R.P., Phillips D.L. (1994): A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology, **33**: 140–158.

Duan Q., Schaake J., Andréassian V., Franks S., Goteti G., Gupta H.V., Gusev Y.M., Habets F., Hall A., Hay L., Hogue T., Huang M., Leavesley G., Liang X., Nasonova O.N., Noilhan J., Oudin L., Sorooshian S., Wagener T., Wood E.F. (2006): Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology, **320**: 3–17.

Farnsworth R.K., Thompson E.S., Peck E.L. (1982): Evaporation Atlas for the Contiguous 48 United States. NOAA Technical Report, NWS 33, Washington, DC.

Gan T.Y., Burges S.J. (2006): Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. Journal of Hydrology, **320**: 117–131.

Heřmanovský M., Pech P. (2013): Selection of catchment descriptors for the physical similarity approach. Part I: Theory. Soil & Water Research, **8**: 133–140.

Hillel D. (1998): Environmental Soil Physics. Elsevier, San Diego.

Jakeman A.J., Hornberger G.M., Littlewood I.G., Whitehead P.G., Harvey J.W., Bencala K.E. (1992): A systematic approach to modelling the dynamic linkage of climate, physical catchment descriptors and hydrological response components. Mathematics and Computers in Simulations, **33**: 359–366.

Koren V., Smith M., Duan Q. (2003): Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, Calibration of Watershed Models. Water Science and Application, **6**: 239–254.

Kutílek M., Nielsen D.R. (1994): Soil Hydrology. Catena-Verlag, Cremlingen-Destedt.

Loveland T.R., Reed B.C., Brown J.F., Ohlen D.O., Zhu J., Yang L., Merchant J.W. (2000): Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. International Journal of Remote Sensing, **21**: 1303–1330.

Magette W.L., Shanholtz V.O., Carr J.C. (1976): Estimating selected parameters for the Kentucky Watershed Model from watershed characteristics. Water Resources Research, **12**: 472–476.

Miller D.A., White R.A. (1999): A Conterminous United States multi layer soil characteristics data set for regional climate and hydrology modeling. Earth Interactions, 2 (available at http://EarthInteractions.org).

Nathan R.J., McMahon T.A. (1990): Identification of homogeneous regions for the purposes of regionalisation. Journal of Hydrology, **121**, 217–238.

Oudin L., Andréassian V., Perrin C., Michel C., Le Moine M. (2008): Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resources Research, **44**: W03413 doi:10.1029/2007WR006240.

Parajka J., Merz R., Blöschl G. (2005): A comparison of regionalisation methods for catchment model parameters. Hydrology and Earth System Sciences, **9**: 157–171.

Post D.A., Jones J.A., Grant G.E. (1998): An improved methodology for predicting the daily hydrologic response of ungauged catchments. Environmental Modelling & Software, **13**: 395–403.

Sivapalan M., Takeuschi K., Franks S., Gupta V.K., Karambiri H., Lakshmi V., Liang X., McDonnel J.J., Mendiondo E.M., O'Connel P.E., Oki T., Pomeroy J.W., Schertzer D., Uhlenbrook S., Zehe E. (2003): IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences, **48**: 857–880.

Wagener T., Wheater H.S., Gupta H.V. (2004): Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London.

Xu C.Y. (1999): Estimation of parameters of a conceptual water balance model for ungauged catchments. Water Resources Management, **13**: 353–368.

Received for publication April 25, 2012
Accepted after corrections October 14, 2013

Corresponding author:

Ing. Martin Heřmanovský, Česká zemědělská univerzita v Praze, Fakulta životního prostředí, Kamýcká 129, 165 21 Praha 6-Suchdol, Česká republika
e-mail: hermanovsky@fzp.czu.cz