Regionalization of hydrological model parameters using data depth

András Bárdossy and Shailesh Kumar Singh

ABSTRACT

The parameters of hydrological models with no or short discharge records can only be estimated using regional information. We can assume that catchments with similar characteristics show a similar hydrological behaviour. A regionalization of hydrological model parameters on the basis of catchment characteristics is therefore plausible. However, due to the non-uniqueness of the rainfall/runoff model parameters (equifinality), a procedure of a regional parameter estimation by model calibration and a subsequent fit of a regional function is not appropriate. In this paper, a different procedure based on the depth function and convex combinations of model parameters is introduced. Catchment characteristics to be used for regionalization can be identified by the same procedure. Regionalization is then performed using different approaches: multiple linear regression using the deepest parameter sets and convex combinations. The assessment of the quality of the regionalized models is also discussed. An example of 28 British catchments illustrates the methodology.

Key words | halfspace depth function, parameter estimation, regionalization

András Bárdossy
Shailesh Kumar Singh (corresponding author)
Institute of Hydraulic Engineering,
University of Stuttgart,
Stuttgart D-70569,
Germany
E-mail: *sk.singh@niwa.co.nz*

Shailesh Kumar Singh
National Institute of Water and Atmospheric Research Ltd.,
Christchurch,
New Zealand

INTRODUCTION

Hydrological modelling has become a widely accepted theoretical tool for water resources engineering and management. Rainfall–runoff models are used both for short and medium time management (e.g. flood forecasting) and long-time design purposes. However, the application of hydrological models is limited due to several reasons.

One important limitation is data availability. Discharges are only measured at a few selected river cross-sections, leading to a small number of catchments for which the runoff calculated from the models might be verified. Further, the high spatial and temporal variability of the meteorological input (such as precipitation, temperature or wind) cannot fully be captured by the usually small number of meteorological stations. Radar measurement of precipitation can provide a more detailed space-time information on precipitation but, unfortunately, the reliability of the data is currently still low.

Other influencing factors, such as soil properties, also vary considerably in space and even to some extent in time (e.g. macropores in soils). These problems, among others, make models based on physical principles unsuitable for many practical applications. Models, which to some extent use analogous concepts, can partly smooth out the effects of variability and therefore can often be successfully used for practical purposes. The limitation of these models is that some of their parameters are not directly related to physically measurable quantities, and have to be estimated from observations using calibration techniques. For this purpose, observed discharge series are needed.

In order to use partly conceptual models for catchments without discharge observations, the model parameters have to be regionalized since they cannot be calibrated. In hydrology, regionalization is applied widely and successfully. One such example is the assessment of possible extremes floods (Burn *et al*. 2007) and low flows (Ouarda *et al*. 2008). Catchment properties are used as a basis in these procedures. The regionalization of hydrological model parameters is only possible if they can be related to catchment characteristics.

The link between model parameters and catchment properties has to be identified and quantified. This task is further

doi: 10.2166/nh.2011.031

complicated by the fact that model parameters are not measured quantities. These have to be identified, for example, by using an inverse procedure that maximizes some selected quality measures, such as the frequently used Nash–Sutcliffe coefficients (Nash & Sutcliffe 1970); these are based on the reproduction of the observed discharge series.

Hundecha & Bárdossy (2004) studied a regionalization method based on *a priori* defined transfer function. They found that a reasonable relationship between the model parameters and the catchment properties can be established by calibrating the parameters of the transfer function instead of the model parameters. In a similar study by Götzinger & Bárdossy (2007), the relationship between the model and the catchment parameters was established by imposing different conditions on the regionalization function. Furthermore, Merz & Blöschl (2004) examined eight regionalization methodologies in 308 catchments in Austria. They found that the best regionalization methods are those that make use of the average parameters of immediate upstream and downstream neighbours, followed by regionalization and by kriging.

In a similar study, Parajka *et al*. (2005) found that the kriging and similarity-based approaches performed best. Buytaert & Beven (2009) suggest a probabilistic shift of parameters of similar catchments as regionalization procedure. An overview of regionalization methodologies and different approaches can be found in Wagener *et al*. (2004).

The purpose of this paper is to develop a methodology for regionalization based on the geometry of the catchment property vectors. We concentrate on catchments which do not differ too much from the observed catchments. Instead of selecting the risky method of extrapolating results, we decided to restrict regionalization to catchments whose properties are inside the domain (formally in the convex hull) of the properties corresponding to the observed catchments. The suggested procedure makes it possible for a whole set of parameters to be derived for these catchments. However, this method does not allow for extrapolation. Nonetheless, we show that the corresponding methodology for the selection of the catchment properties can also be used for other regionalization approaches.

A case study with a number of small- to medium-sized British catchments modelled using HYMOD (hydrologic model) illustrates the methodology. The effect of relaxing the convexity assumption is discussed; the study shows that both the traditional linear regression approach and the suggested weighting fail similarly in extrapolation cases.

METHODOLOGY

The required steps to regionalize the hydrological model parameters are:

1. identification of the model parameters for the donor catchments (gauged catchments);
2. identification of the catchment properties to be used for the transfer of model parameters; and
3. assessment of the procedure to transfer model parameters to the ungauged catchments.

The first step is usually carried out by model calibration using a numerical optimization procedure. For the regionalization of model parameters, one of the major problems is that there is a large number of parameter vectors which perform nearly equally well. It is difficult to decide which of these should be taken for regionalization. Dotty plots, showing model performances as a function of individual parameters, show that a wide range of parameter values can lead to good model performance.

After performing step 1, classical multivariate regression takes care of steps 2 and 3 using a stepwise procedure as demonstrated in (Samaniego & Bárdossy 2006). The most significant variable is selected first and the corresponding linear regression is calculated. Then, stepwise, new variables enter the system until no significant improvement can be achieved. This procedure can be used to regionalize certain discharge characteristics such as annual discharge or extremes.

More sophisticated procedures combine all the three steps mentioned above. For example, in Hundecha & Bárdossy (2004) and Götzinger & Bárdossy (2007) the transfer function parameters are estimated simultaneously with the model parameters. A procedure for the stepwise estimation of the parameters is suggested in Lamb & Kay (2004), where the regionalization of a parameter is followed by a recalibration of the other parameters.

Another important step (although seldomly considered) is deciding whether the regionalization for the ungauged catchment can be performed on the basis of the gauged (donor) catchments or not. For ungauged catchments whose properties are very different to those of the donor catchments, a regionalization might be unreasonable. In order to make regionalization reasonable, one might allow the procedure to work for catchments whose properties are comparable in range to the donor catchments.

It is generally questionable to regionalize model parameters for catchments whose properties are outside the range of the properties of the donor catchments. However, this restriction might allow unreasonable model transfers. For example a set of small forested catchments and another set of large agricultural catchments cannot necessarily be sufficient to estimate parameters of small agricultural or large forested catchments. It is therefore reasonable to restrict the regionalization to ungauged catchments whose relevant properties for the regionalization are in between the properties of the donor catchments in a geometrical sense; i.e., they are in the convex hull of the property vectors of donor catchments. Figure 1 schematically explains the concept. The triangle represent the convex hull of properties of the observed catchments. The point marked with $*$ (star) represents a catchment with properties inside the convex hull, for which a regionalization is 'safe'. The point marked with a circle corresponds to a catchment for which extrapolation is required. This point is outside the convex hull.

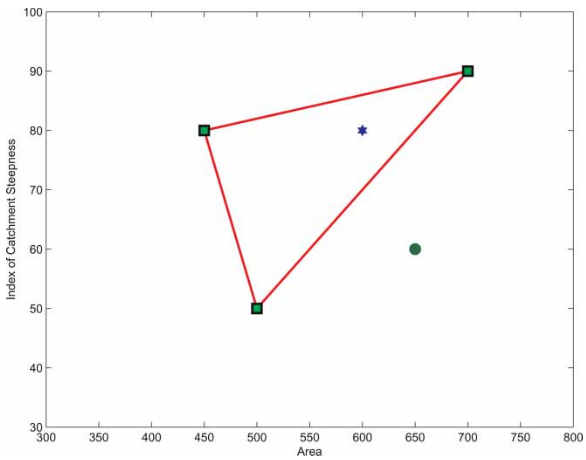

Figure 1 | Given three catchments with two properties (area and steepness), the three observed catchments represented by squares allow parameter interpolation for the catchment with properties represented by the star and require extrapolation for the catchment represented by the circle.

Formally, a vector $\mathbf{v} = (v_1, \ldots, v_k)$ is in the convex hull of the vectors $\mathbf{u^i} = (u_1^i, \ldots, u_k^i)$ for $i = 1, \ldots, m$ if and only if

$$v_j = \sum_{i=1}^{m} \lambda_i u_j^i$$

for all $j = 1, \ldots, k$ with $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$. For more details, please refer to the appendix.

If this is the case, then step 2 of the procedure might become unnecessary for linear regionalization. To see this, let us assume that $(b_1^*, \ldots, b_J^*)$ denotes the parameters of the hydrological model for catchment *. Assuming a linear function can be used to regionalize the model parameters, this would mean that any selected model parameter b_j^k for catchment k can be calculated as a linear function of the catchment characteristics c_i^k:

$$b_j^k = a_{0,j} + \sum_{i=1}^{I} a_{i,j} c_i^k. \tag{1}$$

If we restrict regionalization to catchments whose properties are not outside the ranges of the catchments with observations, or even restricting them to their convex hull, then all selected properties of a catchment k in the convex hull can be written as a linear combination of the properties of catchments. We then have:

$$c_i^k = \sum_{l=1}^{L} \lambda_l c_i^l \quad i = 1, \ldots, I. \tag{2}$$

Further, $\lambda_l \geq 0$ and $\sum \lambda_l = 1$. (Note that the weights are the same for each property.)

Combining Equations (1) and (2) yields:

$$b_j^k = a_{0,j} + \sum_{i=1}^{I} a_{i,j} \sum_{l=1}^{L} \lambda_l c_i^l. \tag{3}$$

Exchanging the summations gives:

$$b_j^k = a_{0,j} + \sum_{l=1}^{L} \lambda_l \left(\sum a_{i,j} c_i^l \right). \tag{4}$$

Assuming that $\sum_{l=1}^{L} \lambda_l = 1$, we have:

$$b_j^k = \sum \lambda_l b_j^l. \tag{5}$$

The above equation states that, in the case of a linear regionalization function, the model parameters can be calculated directly using the weights λ_l without the explicit calculation of the coefficients $a_{i,j}$. A weighted combination of the model parameters was suggested in Kay *et al.* (2006). However, their weights are based on catchment similarity indices, while the weights in this case can be directly identified from the catchment properties.

Note that for a given catchment, weights λ_l are either non-existent (for catchments with property vectors outside the convex hull) or are generally non-unique (for catchments inside the convex hull). This leads to a set of parameter vectors even in the case of a single parameter vector being considered for each catchment. The estimator obtained under the mentioned constraints will be refered to as the *convex estimator* in the rest of this paper.

A perfect regionalization is of course not possible, so we can assume a random error ε_j^k where

$$b_j^k = a_{0,j} + \sum_{i=1}^{I} a_{i,j} c_i^k + \varepsilon_j^k. \tag{6}$$

This modifies Equation (5) to

$$b_j^k = \sum_{l=1}^{L} \lambda_l b_j^l + \sum_{l=1}^{L} \lambda_l \varepsilon_j^l. \tag{7}$$

Thus, assuming that the errors are independent (which is the assumption for multiple linear regression),

$$\mathrm{Var}\left[b_j^k - \sum_{l=1}^{L} \lambda_l b_j^l \right] = \sum_{l=1}^{L} \lambda_l^2 \left(\varepsilon_j^l\right)^2. \tag{8}$$

Equation (8) shows that the estimation error variance depends on the weights λ_l.

Combining $\lambda_l \geq 0$ with $\sum_{l=1}^{L} \lambda_l = 1$ implies that

$$\sum_{l=1}^{L} \lambda_l^2 \leq 1;$$

the estimator defined in Equation (5) is therefore at least as good as the linear estimation defined in Equation (1).

However, for the estimator (5) note that we only need the geometrical properties of the catchment characteristics (relevant for regionalization) and, consequently, no fit of the regression parameters $a_{i,j}$ is required. This means that a regionalization on the basis of a few catchments becomes possible (in contrast to the multiple linear regression approach where a small number of samples makes the parameter estimation very uncertain). The eventual non-uniqueness of the weights λ leads to a set of model parameters, even if only one parameter vector is used for each catchment with observations. On the other hand, the convex estimator can only be used for catchments which are in the convex hull of the catchments with observations (in the sense of relevant catchment properties).

A weighted sum estimator could also be used outside the convex hull (which is extrapolation case) without considering the $\lambda_l > 0$ condition; however, in this case, the squared sum of the λ_ls has to be restricted. Furthermore, due to the extrapolation, the estimator may become unreliable. This problem is demonstrated and discussed in the application section.

Another problem – that the model parameters cannot be identified as unique parameters for the catchments with observations (equifinality) – remains and leads to the problem of deciding on which model parameters b to combine in Equation (5). This problem is discussed in the following sections. Parameter interactions provide a further obstacle for the regionalization of model parameters using multiple linear regression. From this viewpoint, the convex estimator is advantageous as it does not require an explicit estimation of function parameters.

Choice of relevant catchment properties

The selection of appropriate catchment properties is a central problem for regionalization. Usually a great number of candidates are used and the most important are selected sequentially. For the above-proposed regionalization, the method is restricted to the convex hull of the donor catchment property vectors. As more properties are selected, the number of catchments which remain in convex hull becomes smaller. It is therefore essential to keep the number of relevant catchment properties as small as possible. For this purpose we develop a procedure to investigate whether, for a choice of relevant catchment properties, the regionalization could be performed using the convex combination (Equation (1)) for all catchments with observations. The procedure considers each observed catchment individually and checks whether the regionalization for this catchment based on the others would work or not.

This problem can be treated from a geometrical viewpoint. In Bárdossy & Singh (2008), it is shown that parameters which perform well (e.g. parameters with which the model has an Nash–Sutcliffe coefficient (NS) above a given threshold, or the root mean squared error below a threshold) of a model (denoted as B^* for a given catchment *) can be effectively embedded in a convex set in the J-(number of model parameters) dimensional space of model parameters. In fact, for several models, the convex set can be selected such that all parameter vectors of this set are good (meaning that B^* is itself convex). The relevant catchment properties have been well-selected if there is a linear function of these properties which provides parameters which perform well for each catchment. This means that this linear function intersects with each B^*.

In order to select the catchment characteristics, the sets B^* of the observed catchments are investigated. The goal is to find out whether, for a selected set of catchment properties, there is a linear function (Equation (1)) such that this function intersects with all sets of good performing model parameters. If a set of properties is selected, then a catchment with observations for which the regionalization on the basis of the other catchments can be performed using Equation (5) serves as control. For this, we can try all possible parameters of the donor catchments for the regionalization. If none of the combinations work well for the selected catchment, then the regionalization using these properties will fail (and other properties have to be selected).

The formal description of this procedure is as follows. Assume the regionalization procedure should be checked for a catchment with a vector of characteristics $(c_1^*, \ldots, c_I^*)$. If this vector is in the convex hull of the parameter vectors corresponding to the other catchments, then

$$C = \{(c_1^l, \ldots, c_I^l); \quad l = 1, \ldots, L\},$$

which means $c_i^* = \sum_{l=1}^{L} \lambda_l c_i^l$ with $\sum_{l=1}^{L} \lambda_l = 1$ and $\lambda_l \geq 0$. If the catchment properties are such that a regionalization using them is reasonable, Equation (7) should lead to good model parameters for the catchment indicated by $*$.

For each catchment * whose selected catchment properties are in the convex hull of the other catchments properties and which have discharge observations, we can check whether the selected catchment properties would allow a regionalization for it. This also means that there are good parameters for * in the convex hull of the parameter sets B^l. If, on other hand,

$$B^* \cap \mathrm{conv}(B^1, \ldots, B^L) = \emptyset, \tag{9}$$

then this property is not fulfilled. Consequently, any regionalization using the selected catchment properties cannot provide good parameters. The advantage of the formulation Equation (9) is that it provides a purely geometrical condition which can be checked without performing the regionalization explicitly. This can be used to select an appropriate set of explaination variables for the regionalization. The corresponding selection of appropriate variables (SAV) algorithm is as follows.

1. Take all catchments with the set of possible catchment characteristics.
2. Identify the sets of good parameters B^l for each catchment $l = 1, \ldots, L$ using the ROPE (robust parameter estimation) algorithm described in Bárdossy & Singh (2008).
3. Select a catchment property c_{i_1}; set $m = 1$ and define a set of selected indices $S_m = \{i_1\}$.
4. Select a catchment $l = 1$.
5. Select an $m + 1$ catchments with observations which do not include l ($\{l_1, \ldots, l_{m+1}\} \subset \{1, \ldots, L\} - \{l\}$); check if the vector of selected catchment properties $(c_{i_1}^l, \ldots, c_{i_m}^l)$ is in the convex hull of the vectors corresponding to the above chosen $m + 1$ catchments $((c_{i_1}^{l_1}, \ldots, c_{i_m}^{l_1}), \ldots, (c_{i_1}^{l_{m+1}}, \ldots, c_{i_m}^{l_{m+1}}))$. If yes, then regionalization can be checked for this combination; the convex combination should perform for the selected catchment l as target and the $m + 1$ catchments as donors.
6. If the above condition holds, check if

$$B^l \cap \mathrm{conv}(B^{l_1}, \ldots, B^{l_{m+1}}) \neq \emptyset$$

holds. This means all convex combinations of all good parameters for the donor catchments are considered. If none of them gives good parameters for the target, then the regionalization cannot be performed.

7. If the above condition does not hold, then the set of selected catchment properties is not sufficient to perform the regionalization. In this case, an additional property $c_{i_{m+1}}$ has to be selected. The new $S_{m+1} = S_m \cup \{i_{m+1}\}$ is defined and $m = m + 1$. The algorithm continues with step 4.
8. Repeat steps 5–7 until all $m + 1$ element subsets of the set $\{1, \ldots, L\} - \{l\}$ have been visited.
9. Repeat steps 5–8 until all catchments l are considered.
10. The resulting set S_m consists of the indices of catchment properties which are good candidates for linear regionalization.

The above algorithm might have two singular outcomes as described below.

- No set of the catchment properties fulfills the above conditions. In this case, we have to select additional properties and might weaken the conditions defining the good sets B^l.
- The catchment properties lead to a case where condition 5 of the above algorithm is never fulfilled. In this event, the selected properties make all catchments singular and the regionalization using Equation (7) cannot be performed.

Note that the increase in the selected catchment properties in step 7 leads to a decrease in the number of intersections to be checked in step 6. The selection of the catchment property in steps 3 and 7 should be based on hydrological understanding; the purpose of the algorithm is to decide whether a selection is reasonable or not.

Once a set of catchment properties is selected, a method for the selection of the appropriate model parameter vectors has to be chosen.

The calculation of the intersection in step 6 is carried out by the Monte Carlo simulation that generates elements in B^l and checks if they belong to the convex hull of the union of the sets corresponding to the other catchments.

Figures 2 and 3 schematically explain the algorithm for the case of two model parameters and one and two catchment properties, respectively. In the first case, a single

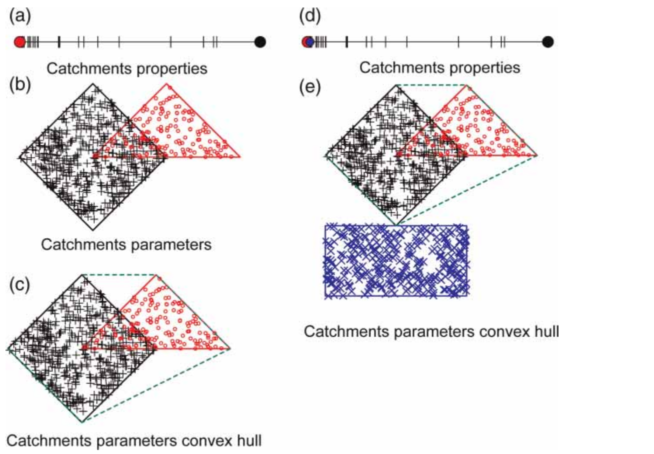

Figure 2 | Step 7 of the catchment property selection algorithm; one property is selected. (a) The black and the red circles represent the low depth points; (b) the corresponding 2 model parameters obtained as convex sets via the ROPE algorithm; (c) convex hull of the two parameter sets. The appropriateness of the selected property is now checked for each observed catchment; (d) selected property represented with a blue circle. (e) The convex set of parameters corresponding to the blue circle obtained using the ROPE algorithm. As the convex hull of the model parameters corresponding to the catchments represented with the red and black circles does not intersect with the set corresponding to the blue, the regionalization using the selected single property is not possible. A full colour version of this figure can be found online at http://www.iwaponline.com/nh/042/5/default.htm

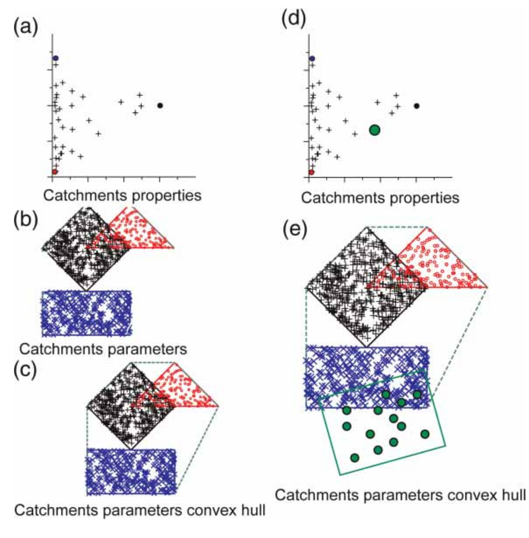

Figure 3 | Step 7 of the catchment property selection algorithm with two properties selected. (a) The black, blue and red circles represent the low depth points; (b) corresponding 2 model parameters obtained as convex sets via the ROPE algorithm; (c) convex hull of the three parameter sets. The appropriateness of the selected property is now checked for each observed catchment; (d) selected properties represented with a green circle. (e) Convex set of parameters corresponding to the green circle obtained using the ROPE algorithm. As the convex hull of the model parameters corresponding to the catchments represented with the red, blue and black circles does intersect with the set corresponding to the green, the regionalization using the selected two properties might be possible if the same procedure repeated for the other catchments also leads to non-empty intersections. A full colour version of this figure can be found online at http://www.iwaponline.com/nh/042/5/default.htm

catchment property is taken for all the catchments available. The two extreme catchments are identified from the catchment property under consideration and shown by circles (red and black) in Figure 2(a). Figure 2(b) illustrates the good parameters B in the model parameter space corresponding to these two extreme catchments. Using the good model parameters (B^l) of these two extreme catchments, we can construct their convex hull as shown in Figure 2(c).

In order to check the feasibility of the regionalization we need to take a catchment whose catchment property lies within the properties corresponding to the two extreme catchments; this is then marked by a circle (blue point) and shown in Figure 2(d). If this (blue) catchment can be regionalized using the selected property, then the parameter space of the set of good parameters corresponding to the (blue) catchment (in the blue rectangle) should have an intersection with the convex hull of the parameters of the catchment corresponding to the two extreme catchments (black and red points). This can be verified in Figure 2(e) where, in our case, the intersection is empty. In this case no linear function of the selected catchment property using the black and red catchments would lead to good performing model parameters for the blue catchment. Thus at least one more catchment property has to be taken into account.

This is explained in Figure 3, where the catchment property space is two-dimensional (two properties). The three catchments corresponding to the corners of the triangle in the property space possess good parameters in the convex sets (denoted by the same colours). For an inside catchment (green point), the corresponding good set intersects with the convex hull corresponding to the three corners. The condition of step 7 is therefore fulfilled. To see if this condition is fulfilled throughout, it should be checked for all catchments to decide whether two properties are enough to perform regionalization. If not, an additional catchment property has to be considered.

Note that the above algorithm can also be used to identify the properties for a regionalization with a non-linear function, which is monotonic in each of its variables. The reason for this is that monotonic functions are mapping the interior of a set to the interior of the image set. The above algorithm selects the properties independently from the target catchment. For a given target catchment, we first have to check if the selected properties are in the convex hull of the observed catchments or not. If yes, the convex estimator can be used; otherwise a regionalization based on the observed catchments cannot be performed for the target catchment.

Equation (5) could also be used for extrapolation if negative weights $\lambda < 0$ are allowed. Unfortunately in this case, all possible parameter vectors could be obtained in the form of Equation (5). Thus, for extrapolation with Equation (5), another criterion is needed; for example a *best* estimator minimizing the estimation variance can be used. This means that $\sum_{l=1}^{L} \lambda_l^2$ is minimized under the conditions imposed by the catchment properties and $\sum_{l=1}^{L} \lambda_l = 1$.

This estimator can be used both for inside and boundary catchments. The λ_l weights are determined by solving the minimization problem:

$$\sum_{l=1}^{L} \lambda_l^2 \rightarrow \min$$

under the constraints:

$$c_i^* = \sum_{l=1}^{L} \lambda_l c_i^l, \quad i = 1, \ldots, I \tag{10}$$

and

$$\sum_{l=1}^{L} \lambda_l = 1.$$

Due to the possibility of negative weights, this estimator becomes unstable if catchment properties show high correlations.

How to perform regionalization

It would be ideal if all good model parameters of the catchments could be combined with good parameters of the catchment of interest using Equation (5). Unfortunately, this is unlikely to ever be the case. In Bárdossy & Singh (2008), we showed that for a catchment the deepest parameter vectors are robust and transferable to other time periods. In light of this, a reasonable choice of model parameters for Equation (5) is to take the deepest points of the good sets B^l. The definition of data depth can be found in the appendix of this paper. Another possibility is to perform an explicit multiple linear regression according to Equation (1) using the deepest parameters for the observed catchments. Alternatively, an implicit multiple regression, as described in Hundecha & Bárdossy (2004), can be used where model parameters are restricted to the sets B^l.

CASE STUDY

The concept of this paper will be illustrated with examples from British catchments. The hydrological model chosen for this study is HYMOD. A short overview of the study area and model concept description is provided in this section.

Study area

This study was carried out on different catchments located in the United Kingdom (UK). Eleven years of data (1980–1990) from 28 small- to medium-sized watersheds (50–1,100 km^2) were available for this research (Figure 4). The watersheds are located throughout the UK. Discharge and precipitation data were obtained from the Top-Down Working Group website (http://tdwg.catchment.org/datasets.html). The temperature data were provided by the British Atmospheric Data Centre (http://badc.nerc.ac.uk/home/index.html). The physical characteristics came from the National River Flow Archive (http://www.nwl.ac.uk/ih/nrfa) and the data CD of the Flood Estimation Handbook. An overview of catchments can be found in Yadav *et al.* (2007).

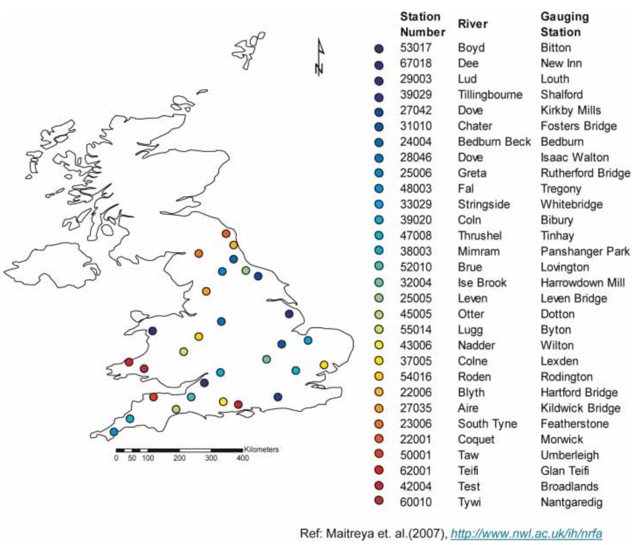

Figure 4 | Study area: 28 selected catchments in UK. A full colour version of this figure can be found online at http://www.iwaponline.com/nh/042/5/default.htm

Hydrological model

HYMOD is a simple conceptual hydrological model. This model has two main components: rainfall excess (two parameters) and two series of linear reservoirs (three parameters, three identical quick and single for the slow response) in parallel as routing components. The model is based on the characteristics of the runoff production process at a point in a catchment. A probability distribution which describes the spatial variation in the catchments is derived from an algebraic expression given by Moore (1985). This model makes the assumption that the soil structure, texture and water storage capacity varies across the catchment. The distribution function of different storage capacity is therefore described as

$$F(C) = 1 - (C/C_{\max})^{\beta} \quad 0 \leq C \leq C_{\max}. \tag{11}$$

The model structure is given in Figure 5. The five parameters of this model are: the maximum storage capacity in the catchment ($C_{\max}$); the degree of spatial variability of the soil moisture capacity within the catchment (β); the factor distributing the flow between the two series of reservoirs (α); and the residence times of the linear reservoirs (R_q) and (R_s). Additional information about the HYMOD

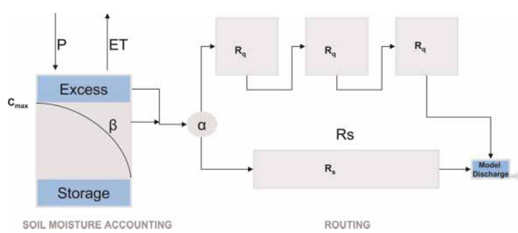

Figure 5 | Schematic representation of HYMOD model.

model can be found in Moore (1985), Boyle *et al.* (2001) and Wagener *et al.* (2001).

APPLICATION AND RESULTS

Choice of catchment properties

Nine possible catchment properties were selected to be considered for regionalization and are listed in Table 1. In order to check the quality of the regionalization, the good set of parameters was identified for each catchment using the ROPE algorithm. As for the model performance measure NS_p, a combination of the Nash–Sutcliffe coefficient for the whole time period (N_a) and the Nash–Sutcliffe coefficient calculated for each year t (N_y) $y = 1, \ldots, NY$ (NY being the number of years) was selected:

$$NS_p = N_a + \min(N_1, \ldots, N_{NY}). \quad (12)$$

This performance measure restricts the possible parameter set by taking only those which are nearly equally good for all years.

The selection of appropriate variables (SAV) algorithm described in 'Choice of relevant catchment properties' was applied to the selected catchments with the list of properties defined in Table 2. The intersection of the sets was calculated using the Monte Carlo simulation. This means that the points in the possible parameter domain were simulated uniformly (after rescaling). Then each simulated point was checked to see if the point belonged to the set corresponding to the catchment. In the case when a single catchment property was selected, there were always cases with empty intersections. A stepwise increase of the catchment property led to non-empty intersections, if three catchment properties were selected. In this case, the selected catchment properties were: AREA (catchment area), BFIHOST (base flow index) and DPSBAR (index of catchment steepness).

Table 1 | The catchments properties to be considered for regionalization (Yadav *et al.* 2007)

Characteristic	Unit	Description
AREA	km^2	Watershed drainage area
BFIHOST	–	Base-flow index derived using HOST classification
DPSBAR	$m\ km^{-1}$	Index of watershed steepness
APSBAR	–	Index representing the dominant aspect of watershed slopes
APSVAR	–	Index representing the invariability of aspect of watershed slopes
RMED-1H	mm	Median annual maximum 1-hour precipitation
MED PERM	–	Percentage soil within watershed with medium/mixed permeability
LOW PERM	–	Percentage soil within watershed with low permeability

Regionalization

The transfer of parameters to unobserved catchments was tested using the classical linear regionalization approach (Equation (1)), using multiple linear regression and with the convex combinations (Equation (5)). In the first step, regionalization was restricted to catchments whose selected properties were in the convex combination of those of the other catchments; two groups were therefore obtained.

The first group contained the so-called *boundary* catchments whose properties cannot be obtained as a convex combinations of the others. The second group contained the *inside* catchments whose properties are in the convex hull of the properties of the boundary catchments. Table 3 lists the *boundary* and the *inside* catchments.

Regionalization was performed and checked for each inside catchment. As an illustration of the methodology, Table 4 shows a set of possible boundary catchments and the corresponding weights for the target inside catchment 20. The linear parameter estimation was carried out by using the deepest parameters for the catchments in the boundary set. For comparison, a set of randomly selected good parameters were also used.

The convex combinations were applied both for the deepest parameters and also for randomly selected good ones. As it is often inside catchments, the catchment properties can be expressed as a large number of convex combinations of the boundary catchments. Some of the special cases were selected.

Table 2 | Numerical values of the considered catchment properties

Cat. No.	AREA	BFIHOST	DPSBAR	APSBAR	APSVAR	RMED-1H	MOD PERM	LOW PERM
1	569.8	0.49	112.10	109.8	0.19	9.1	16.4	26.2
2	269.4	0.46	32.83	93.2	0.25	0.4	0.0	0.0
3	321.9	0.33	125.70	12.3	0.20	10.7	0.0	0.0
4	74.9	0.46	109.67	94.7	0.34	10.0	0.0	0.0
5	196.3	0.43	75.96	300.6	0.24	10.1	0.0	87.4
6	86.1	0.21	67.67	55.6	0.31	10.7	0.0	0.0
7	282.3	0.37	101.84	169.8	0.13	10.5	28.3	1.1
8	59.2	0.60	147.69	173.7	0.24	10.1	0.0	46.7
9	83.0	0.78	144.58	185.4	0.09	10.2	21.0	0.0
10	55.2	0.90	61.13	69.3	0.24	10.6	0.0	0.0
11	68.9	0.51	62.65	108.6	0.17	12.2	0.0	71.1
12	194.0	0.55	40.79	97.0	0.16	12.6	0.0	44.5
13	98.8	0.86	13.50	203.0	0.18	11.1	0.0	7.7
14	238.2	0.53	30.98	121.3	0.15	11.0	0.0	73.3
15	106.7	0.94	78.39	147.7	0.18	10.6	0.0	12.4
16	1,040.0	0.90	50.10	176.0	0.15	10.6	0.1	3.1
17	220.6	0.81	80.13	125.0	0.15	10.9	25.5	13.2
18	202.5	0.54	85.00	195.0	0.11	11.4	31.7	40.6
19	112.7	0.39	91.19	227.2	0.17	12.0	0.0	100.0
20	87.0	0.69	80.74	242.4	0.15	11.8	0.0	76.9
21	826.2	0.42	106.94	219.9	0.07	11.9	2.3	95.4
22	135.2	0.47	72.51	218.3	0.17	10.7	12.8	44.3
23	47.9	0.46	64.00	270.1	0.24	10.8	4.3	84.1
24	259.0	0.61	22.76	119.3	0.11	9.4	0.0	73.2
25	203.3	0.67	161.66	122.8	0.19	9.8	0.0	83.7
26	1,090.4	0.48	157.20	214.0	0.09	11.3	0.1	99.2
27	893.6	0.53	112.35	285.7	0.10	11.0	0.0	100.0
28	53.9	0.27	152.19	89.3	0.17	11.2	0.0	100.0

The performance of the explicit multiple linear regression (Equation (1)) using the deepest and randomly selected parameter vectors for the estimation of the regression coefficients, and the corresponding results using the convex estimator (Equation (5)) are shown in Table 5. As we can see, the performances of the two estimators are similar. The deepest parameter vectors lead to the best estimations for both cases. The performance of the models using the regionalized parameters is comparable to the performance which was obtained using calibration. Note that we cannot expect a better performance for the target catchment than the performance of the model parameters on the catchments which were used for regionalization. Figure 6 shows the observed and the simulated hydrographs for catchment 17 using the convex estimator (using four boundary catchments only) and multiple linear regression with the deepest point.

For the regionalization, which is not restricted to the convex hull (allowing negative weights according to Equation (10)), a cross-validation was performed. The corresponding results are listed in Table 6. We can see that for extrapolation, both multiple linear regression and the

Table 3 | List of the boundary and inside catchments

Boundary	Inside
2	1
3	4
6	5
8	7
9	12
10	17
11	18
13	19
14	20
15	22
16	
21	
23	
24	
25	
26	
27	
28	

Table 4 | A set of possible boundary catchments for catchment 20 and the corresponding weights

Inside catchment	Boundary catchment	Weights	AREA	BFIHOST	DPSBAR
	2	0.128	269.4	0.46	32.83
	6	0.110	86.1	0.21	67.67
	8	0.260	59.2	0.60	147.69
	10	0.502	55.2	0.90	61.13
20			87.0	0.69	80.74

weighted sum have failed for catchment 16 (which is a boundary catchment). Its properties differ strongly from those of the other catchments, meaning that the parameter estimation for this catchment could be risky. The corresponding sum of the squared weights is the largest for this catchment, indicating the highest uncertainty of the estimation. We could assume that the bad performance is due to hydrologically different behavior of this catchment; this is not the case, however, as will be shown in the following.

To test the effectiveness of the method with fewer boundary catchments, all $J+1$ element subsets of the catchment property vectors corresponding to the boundary set, which contained the property vector of the target catchment in their convex hull, were identified. For our case, this means that for each inside catchment as a target, all sets of the four boundary catchments whose properties contained those of the target catchment in their convex hull were obtained.

Table 7 shows the number of all possible four catchment combinations of boundary catchments for describing the inside catchments. For any $J+1$ selected catchments, the weights λ_l are unique (due to the constraint that $\sum \lambda_l = 1$). The estimation was carried out using these weights and the performance of the resulting parameter set was calculated.

Figure 7 shows the performance of the estimator for selected target catchments using the deepest and most randomly selected parameters. The x axis shows the minimal performance corresponding to the selected boundary catchments. On the y axis, we can read the performance of the model for the target catchment using the convex estimator (Equation (5)) for the selected combination. We cannot expect a better performance for the target catchments than for the catchments used for the regionalization; the points above the main diagonal are obtained more or less by chance. Note also that a large number of combinations can be formed. All regionalizations which used the deepest parameter vectors performed well. On the other hand, the quality from the randomly selected (i.e. less deep) parameters was poorer, showing larger scattering. This may consequently lead to poorer performances.

In order to check whether catchment 16 behaves irregularly compared to the others, as might assumed from the cross-validation results for multiple linear regression estimation, all combinations for which this catchment were used for regionalization were selected. If catchment 16 was an irregularly behaving catchment, then the combinations where catchment 16 is used should not perform well. However, this is not the case. Figure 8 shows that all combinations which include catchment 16 perform as well as the other combinations. The impossibility of the regionalization for catchment 16 is mainly caused by the inability to

Table 5 | The performance (NS_p) of the convex estimation (Equation (5)) and the explicit multiple linear regression using the deepest and randomly selected parameter vectors for inside catchments

		Weighted combination					Multiple regression				
			Random					Random			
	Opt	Deep	Mean	Max	Min	Std	Deep	Mean	Max	Min	Std
1	1.16	1.13	1.12	1.13	1.12	0.0026	1.14	1.14	1.14	1.12	0.0044
4	0.94	0.85	0.85	0.86	0.82	0.0071	0.87	0.87	0.88	0.85	0.0065
5	1.20	1.11	1.11	1.12	1.09	0.0059	1.12	1.12	1.13	1.11	0.0034
7	1.67	1.65	1.65	1.65	1.64	0.0030	1.64	1.63	1.64	1.62	0.0042
12	1.33	1.20	1.22	1.25	1.17	0.0168	1.25	1.25	1.26	1.24	0.0053
17	1.63	1.37	1.36	1.42	1.31	0.0242	1.43	1.42	1.50	1.34	0.0341
18	1.41	1.00	1.02	1.06	0.98	0.0181	1.02	1.02	1.05	0.98	0.0137
19	1.69	1.63	1.63	1.64	1.63	0.0025	1.60	1.60	1.61	1.59	0.0033
20	1.70	1.45	1.45	1.49	1.39	0.0176	1.52	1.51	1.53	1.47	0.0135
22	1.47	1.26	1.26	1.28	1.23	0.0117	1.26	1.26	1.28	1.24	0.0064

estimate the overall slope of the linear function from the other catchments.

DISCUSSION AND CONCLUSIONS

In this paper the possibilities to regionalize conceptual model parameters using explicit and implicit methods were discussed. It was shown that for a catchment whose property vector is in the convex hull of the property vector of observed catchments, a linear regionalization can be replaced by a convex combination for which an explicit estimation of the coefficients of the regionalization function was not required.

The idea of convex combinations could be used to check whether a set of properties should be used for regionalization. The selection of appropriate variables (SAV) algorithm was developed for this purpose.

After selection of the catchment properties, regionalization can be carried out using either the classical linear regression approach or using the convex combinations.

The developed methodology was applied to a set of 28 UK catchments using the HYMOD model. For the case study, three catchment properties were identified as necessary for the regionalization of the five model parameters.

Data depth is a useful tool to identify unique parameter vectors for regionalization. If parameter vectors with a low data depth (which are near the boundary of the good parameter set) are used for regionalization, the performance of the target catchment varies strongly. The deepest parameter sets lead to good regionalizations both for the linear regression and the convex combination methods.

The performance of the model using regionalized parameters is comparable to the performance of the model

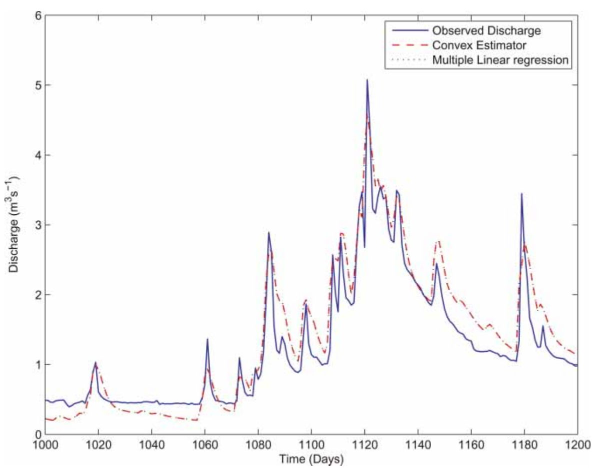

Figure 6 | Observed and simulated hydrographs for catchment 17 using the convex estimator (using four boundary catchments only) and multiple linear regression with the deepest point.

Table 6 | Cross-validated performance (NS_p) of the relaxed convex combination (negative weights allowed) and multiple linear regression using the deepest and randomly selected parameter vectors

		Weighted combination					Multiple regression				
			Random					Random			
	Opt	Deep	Mean	Max	Min	Std	Deep	Mean	Max	Min	Std
1	1.16	1.13	1.13	1.14	1.12	0.0031	1.14	1.14	1.15	1.12	0.0068
2	1.18	1.15	1.15	1.16	1.14	0.0042	1.13	1.13	1.15	1.12	0.0050
3	1.34	1.25	1.25	1.27	1.24	0.0069	1.26	1.25	1.27	1.24	0.0059
4	0.94	0.87	0.87	0.89	0.85	0.0089	0.88	0.87	0.89	0.85	0.0062
5	1.20	1.13	1.13	1.14	1.12	0.0038	1.12	1.13	1.14	1.12	0.0044
6	1.02	1.02	1.02	1.03	1.01	0.0046	0.99	1.09	1.04	0.98	0.0164
7	1.67	1.65	1.65	1.66	1.64	0.0028	1.64	1.64	1.65	1.63	0.0032
8	1.04	0.93	0.93	0.95	0.91	0.0071	0.95	0.95	0.97	0.91	0.0111
9	1.55	1.23	1.22	1.28	1.16	0.0230	1.31	1.23	1.39	1.06	0.0871
10	1.25	1.03	1.03	1.11	0.98	0.0217	1.03	1.03	1.11	0.999	0.0211
11	1.10	0.96	0.96	0.97	0.95	0.0048	0.98	0.97	0.99	0.94	0.0112
12	1.33	1.26	1.25	1.27	1.24	0.0063	1.24	1.24	1.25	1.23	0.0043
13	1.69	1.65	1.63	1.67	1.55	0.0247	1.64	1.65	1.68	1.55	0.0233
14	1.11	0.99	0.99	0.99	0.98	0.0023	0.98	0.98	0.98	0.97	0.0026
15	1.29	1.26	1.25	1.31	1.17	0.0280	1.18	1.17	1.26	1.04	0.0543
16	1.59	−4.22	−4.08	−1.30	−7.456	1.4227	−3.81	-5.15	−0.51	−13.80	4.1568
17	1.63	1.48	1.47	1.52	1.43	0.0171	1.44	1.44	1.51	1.36	0.0341
18	1.41	1.08	1.07	1.11	1.04	0.0127	1.07	1.07	1.19	1.04	0.0129
19	1.69	1.64	1.63	1.64	1.63	0.0036	1.62	1.68	1.64	1.61	0.0087
20	1.70	1.50	1.49	1.51	1.48	0.0060	1.52	1.50	1.53	1.47	0.0173
21	1.54	1.45	1.45	1.48	1.43	0.0091	1.42	1.43	1.46	1.40	0.0186
22	1.47	1.30	1.30	1.31	1.29	0.0042	1.27	1.28	1.30	1.26	0.0088
23	1.40	1.38	1.38	1.39	1.37	0.0030	1.36	1.37	1.38	1.36	0.0077
24	1.35	1.08	1.08	1.11	1.04	0.0158	1.15	1.14	1.16	1.07	0.0098
25	1.43	1.06	1.05	1.15	0.93	0.0410	1.09	1.06	1.13	0.96	0.0391
26	1.55	1.37	1.37	1.42	1.34	0.0173	1.29	1.30	1.37	1.23	0.0352
27	1.73	1.54	1.54	1.57	1.51	0.0111	1.62	1.60	1.64	1.52	0.0341
28	1.36	1.38	1.38	1.42	1.34	0.0181	1.31	1.34	1.38	1.29	0.0267

on the catchments used for regionalization. The restriction of the estimator (Equation (5)) to convex combinations (with non-negative weights) allows regionalization toward the *inside* of the observed set. This restriction is reasonable, since the transfer of parameters to catchments which differ strongly from those observed is always risky.

As the methodology does not require the estimation of transfer function parameters; it can also be used in the case of a small number of catchments. This was demonstrated for the selected British catchments. Depending on the number of boundary catchments, the convex combination became non-unique, but even using the minimum number of catchments, the estimation quality did not decrease.

Formally, the non-negativity condition can be relaxed but this leads to an increase of the squared sum of the weights and thus the estimation variance also increases. Further, this allows extrapolation which, as stated above, might be problematic.

Table 7 | Number of possible combinations for the choice of four boundary catchments which include the properties of the given inside catchments

Inside catchment no.	Total number of possible combinations (four catchments combinations)
1	359
4	183
5	249
7	151
12	211
17	256
18	470
19	168
20	256
22	323

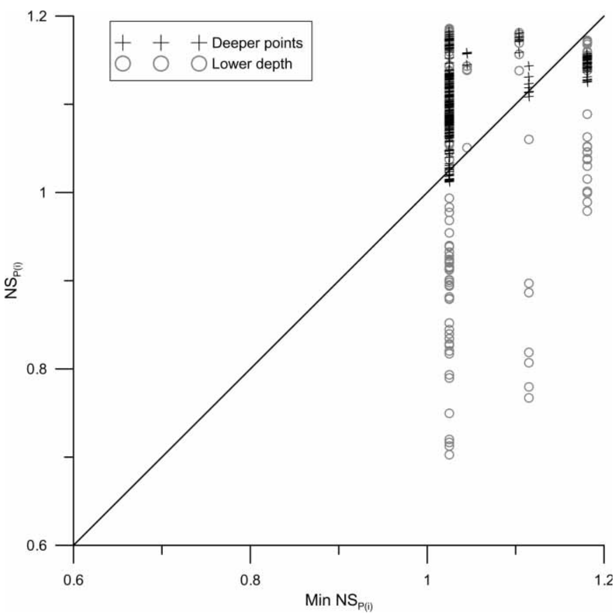

Figure 7 | The performance (NS_P) of the convex estimator for target catchment 5 (*y* axis) using all possible convex combinations using four catchments. The minimal performance of the model on the selected four catchments is represented by the *x* axis. The combinations obtained using deepest parameters are represented by the crosses and using lower depth parameter sets by the circles.

In the case study region, one catchment was identified whose properties are on the boundary of the property vector set and for which regionalization using other catchments failed. However, this catchment could be applied to

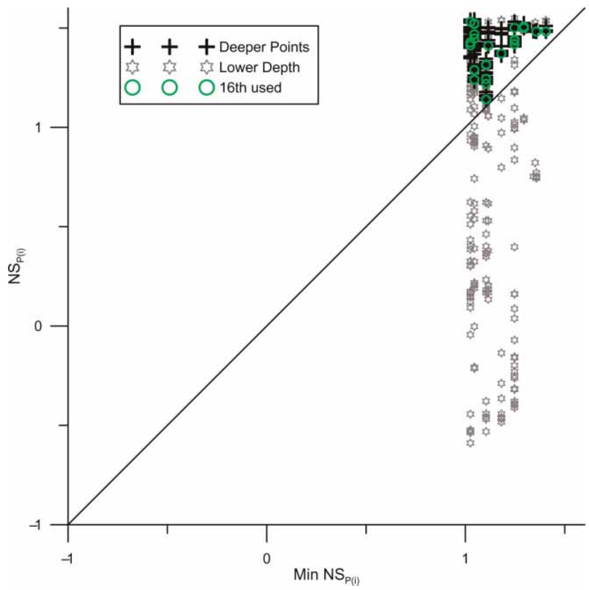

Figure 8 | The performance (NS_P) of the convex estimators for catchment number 20 using the deepest (crosses) and lower depth (stars). The estimators using the deepest parameter vectors including catchment 16 are highlighted with green circles. A full colour version of this figure can be found online at http://www.iwaponline.com/nh/042/5/default.htm

the regionalization of the inside catchments. This result showed that extrapolation might lead to problems even for catchments which behave reasonably.

ACKNOWLEDGEMENTS

This research was partly supported by the IPSWAT scholarship for the second author.

REFERENCES

Bárdossy, A. & Singh, S. K. 2008 Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences* **12** (6), 1273–1283.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z. & Smith, M. 2001 Toward improved streamflow forecasts: value of semidistributed modeling. *Water Resources Research* **37** (11), 2749–2759.

Burn, D., Ouarda, T. & Shu, C. 2007 Estimation of extreme flow quantiles and quantile uncertainty for ungauged catchments. *IAHS-AISH Publication* **313**, 417–424.

Buytaert, W. & Beven, K. 2009 Regionalization as a learning process. *Water Resources Research* **45** (11), W11419.

Götzinger, J. & Bárdossy, A. 2007 Comparison of four regionalisation methods for a distributed hydrological model. *Journal of Hydrology* **333**, 374–384.

Hundecha, Y. & Bárdossy, A. 2004 Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology* **292**, 281–295.

Kay, A. L., Jones, D. A., Crooks, S. M., Calver, A. & Reynard, N. S. 2006 A comparison of three approaches to spatial generalization of rainfall-runoff models. *Hydrological Processes* **20** (18), 3953–3973.

Lamb, R. & Kay, A. L. 2004 Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resources Research* **40** (7), W07501.

Liu, R. Y., Parelius, J. M. & Singh, K. 1999 Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Annals of Statistics* **27** (3), 783–858.

Merz, R. & Blöschl, G. 2004 Regionalisation of catchment model parameters. *Journal of Hydrology* **287**, 95–123.

Moore, R. J. 1985 The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal* **30** (2), 273–297.

Nash, J. & Sutcliffe, J. 1970 River flow forecasting through conceptual models. 1. a discussion of principles. *Journal of Hydrology* **10**, 282–290.

Ouarda, T., Charron, C. & St-Hilaire, A. 2008 Statistical models and the estimation of low flows. *Canadian Water Resources Journal* **33**, 195–206.

Parajka, J., Merz, R. & Blöschl, G. 2005 A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences* **9** (1–2), 157–171.

Rousseeuw, P. J. & Struyf, A. 1998 Computing location depth and regression depth in higher dimensions. *Statistics and Computing* **8**, 193–203.

Samaniego, L. & Bárdossy, A. 2006 Simulation of the impacts of land use/cover and climatic changes on the runoff characteritics at the mesoscale. *Ecological Modelling* **196**, 45–61.

Tukey, J. 1975 Mathematics and picturing data. In *Proceedings of the 1975 International 17 Congress of Mathematics* (R. D. James, ed.), Vol. 2, Canadian Mathematical Congress, Montreal, pp. 523–531.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. & Sorooshian, S. 2001 A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* **5** (1), 13–26.

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

Yadav, M., Wagener, T. & Gupta, H. 2007 Regionalization of constraints on expected watershed response behavior for improved prediction in ungauged basin. *Advance in Water Resources* **30** 1756–1774.

Zuo, Y. & Serfling, R. 2000 General notions of statistical depth function. *Annals of Statistics* **28** (2), 461–482.

First received 12 April 2010; accepted in revised form 11 November 2010. Available online June 2011

APPENDIX

Data depth function

Depth functions were first introduced by Tukey (1975) to identify the centre (a kind of generalized median) of a multivariate dataset. Several generalizations of the concept have been defined by Rousseeuw & Struyf (1998), Liu *et al*. (1999) and Zuo & Serfling (2000).

Definition

The halfspace depth of a point p with respect to the finite set X in the d dimensional space $\Re^d$ is defined as the minimum number of points of the set X lying on one side of a hyperplane through the point p. The minimum is calculated over all possible hyperplanes.

Formally, the halfspace depth of the point p with respect to set X is:

$$D_X(p) = \min_{n_h}(\min(|\{x \in X\langle n_h, x-p\rangle > 0\}|), (|\{x \in X\langle n_h, x-p\rangle < 0\}|)). \quad (13)$$

Here, $\langle x, y\rangle$ is the scalar product of the d dimensional vectors and n_h is an arbitrary unit vector in the d dimensional space representing the normal vector of a selected hyperplane.

If the point p is outside the convex hull of X then its depth is 0. Points on and near the boundary have low depth while points *deeply* inside have high depth.

One advantage of this depth function is that it is invariant to affine transformations of the space. This means that the different ranges of the parameters have no influence on their depth.

Convex hull

The convex hull of a set of points S is the smallest area polygon which encloses S. A formal example of convex hull is given in Figure 9.

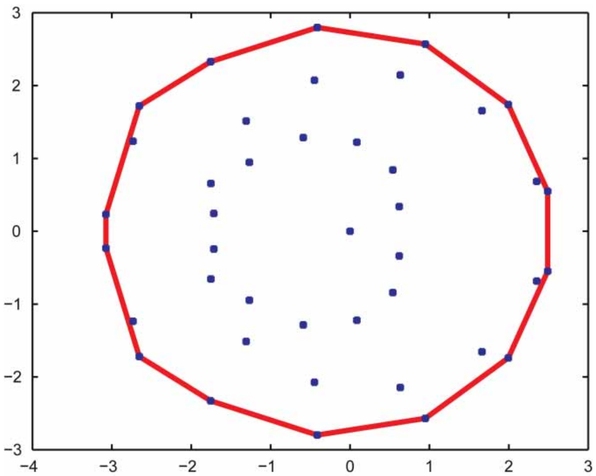

Figure 9 | Convex hull.