Simulating runoff from MOPEX experimental river basins using the SWAP land surface model and different parameter estimation techniques

YEUGENIY M. GUSEV & OLGA N. NASONOVA

Institute of Water Problems, Russian Academy of Sciences, Gubkina St. 3, 119991 Moscow, Russia

gusev@aqua.laser.ru

Abstract Different techniques for the *a priori* estimation and calibration of parameters in the SWAP land surface model for the 12 MOPEX experimental river basins were investigated. Two approaches for *a priori* parameter estimation were applied: (1) derivation of parameters from spatial distribution of different soil and vegetation classes within a basin; (2) application of a global parameter data set provided within the framework of the Second Global Wetness Project (GSWP-2). Two parameter calibration techniques were also used: (1) manual calibration of hydraulic conductivity at saturation that is the most crucial for runoff generation; (2) automatic optimization of several soil parameters. The results of river flow simulations with all the derived sets of parameters were compared with observations to reveal optimal values of soil parameters. Since the optimal values differed from *a priori* estimations using the USDA soil classification, an attempt was made to connect the optimal parameter values with the USDA soil texture classes.

Key words *a priori* parameter estimation; land surface modeling; MOPEX; parameter calibration; river runoff

INTRODUCTION

Model simulation of river flow requires data on soil and vegetation parameters for a river basin. In physically-based models, some parameters have a physical meaning and can be measured directly, some can be estimated from the other measured characteristics, and some represent empirical coefficients obtained for a particular case and cannot be applied for an object with substantially different natural properties. In addition, most parameters are characterized by spatial variability while measurements have a point character. Besides that, these measurements, as a rule, are not carried out on a regular basis and do not cover all the watersheds of the Earth. As a result, there are many locations in the world where data on soil and vegetation parameters are not available. This raises the problem of *a priori* parameter estimation.

A priori estimation of the model parameters is usually based on relationships between the parameters and soil, vegetation, topographic and climatic characteristics of the object under study (Schaake *et al*., 2001). Some examples of such estimations for conceptual and physically-based hydrological models can be found in Koren *et al*. (2000) and Kuchment & Gelfan (2005), respectively. In the land surface models (LSMs), *a priori* parameters are usually derived from the relationships between model parameters and soil texture and vegetation classes. The shortcomings of existing *a* *priori* parameter estimation procedures are shown in (Duan *et al*., 2006). The Model Parameter Estimation Experiment (MOPEX) was initiated in 1996 to improve this situation by developing enhanced *a priori* parameter estimation techniques for hydrological and land surface models (Schaake *et al*., 2001). According to the MOPEX strategy the improved *a priori* parameters may be estimated using the results of model calibration, in particular, by application of new relationships between the calibrated parameters and basin characteristics.

In the present paper, following the MOPEX strategy, investigation of two *a priori* model parameter estimation techniques was undertaken using the SWAP land surface model (Soil Water–Atmosphere–Plants) (Gusev & Nasonova, 1998, 2002, 2003) for river runoff simulations. Another goal of this study was investigation of how different calibration techniques can improve simulation of river flow based on *a priori* estimated parameters. An attempt was made to use the results of calibration for construction of a new set of *a priori* parameters, which can be treated as effective parameters for river basins.

HYDROLOGICAL BASINS AND DATA

The hydrological objects under study represent 12 river basins (with an area of the order of 10^3 km^2), which were selected within the framework of the Second MOPEX Workshop (Duan *et al*., 2006). The basins are located within the southeastern part of the United States and characterized by a wide range of hydrological and climatic conditions, ranging from arid to a very wet regime. The main characteristics of the basins are given in Table 1, a full description can be found in Duan *et al*. (2006).

The data set for each of the 12 basins includes: (1) 1-hourly near-surface meteorological data (downward shortwave and longwave radiation, air temperature and humidity, atmospheric precipitation, air pressure and wind speed) for a 39-year period (1960–1998), and (2) basin land surface characteristics (basin boundary, elevation, spatial distribution of different soil and vegetation classes within a basin, soil texture,

Table 1 Characteristics of the 12 MOPEX river basins.

Basin	Basin ID	Area (km^2)	Elevation (m)	Latitude	Longitude	Dominant soil type	Dominant vegetation type	Clay in soil (%)	Sand in soil (%)
1	01608500	3810	171	39.45	–78.65	Loam	Dec. broadleaf	43.5	30.6
2	01643000	2116	71	39.39	–77.38	Silt loam	Dec. broadleaf	17.1	23.9
3	01668000	4134	17	38.32	–77.52	Clay loam	Mixed forest	29.4	26.4
4	03054000	2372	390	39.15	–80.04	Loam	Dec. broadleaf	32.0	26.0
5	03179000	1020	465	37.54	–81.01	Si cl loam/loam	Dec. broadleaf	24.6	20.5
6	03364000	4421	184	39.20	–85.93	Si loam/cl loam	Croplands	25.0	24.6
7	03451500	2448	594	35.61	–82.58	Loam	Mixed forest	30.9	35.9
8	05455500	1484	193	41.47	–91.72	Clay loam	Cropland	32.6	10.8
9	07186000	3015	254	37.25	–94.57	Si loam/cl loam	Dec. broadleaf	27.4	17.5
10	07378500	3315	0	30.46	–90.99	Silt loam	Ever. needleleaf	20.5	21.0
11	08167500	3406	289	29.86	–98.38	Clay	Grassland	40.0	14.8
12	08172000	2170	98	29.67	–97.65	Clay	Grassland	49.1	23.5

monthly surface albedo, roughness length and greenness fraction). All these data were provided by the Second MOPEX Workshop organizers and described in detail in Duan *et al*. (2006).

Evidently, the information provided on soil and vegetation characteristics is rather general. Application of the SWAP model requires soil hydrophysical parameters including field capacity W_{fc}, wilting point W_{wp}, soil porosity W_{sat}, and relations of soil hydraulic conductivity K and soil matric potential ϕ with soil moisture W, i.e. $K(W)$ and $\phi(W)$. In SWAP, Clapp & Hornberger (1978) parameterizations for $K(W)$ and $\phi(W)$ are used. These parameterizations require information on soil matric potential at saturation ϕ_0, the B parameter, W_{sat}, and soil hydraulic conductivity at saturation k_0. Besides that, snow-free albedo of bare soil α_{soil} and soil depth h_{soil} are also required. As to vegetation parameters, along with provided ones, the SWAP model needs monthly values of the leaf area index LAI, zero plane displacement height d_0, snow-free albedo of vegetation α_{veg}, root depth h_{root}. Since these parameters were not available for the selected basins, we had to estimate them ourselves by means of the techniques described in the next section.

A PRIORI PARAMETER ESTIMATION AND CALIBRATION TECHNIQUES

Two approaches for *a priori* parameter estimation were applied. First, soil and vegetation parameters for each of the 12 basins were derived by us using available data on spatial coverage of the USDA soil texture classes and of the University of Maryland vegetation types. As to soil parameters, experimentally determined soil textural and bulk density data (corresponding to each soil class) were used as predictors for the calculation of hydraulic and soil water retention parameters using the Rosetta code that implements pedotransfer functions based on an artificial neural network predictions (Schaap *et al*., 1998). Thus, the values of W_{sat} and k_0 and Van Genuchten's (1980) shape parameters for the $K(W)$ and $\phi(W)$ functions were determined. Then, the values of ϕ_0 and B were obtained on the basis of interpolation of the estimated $\phi(W)$ by the Clapp and Hornberger equation within the range –33 to –1500 kPa. The values of W_{wp} and W_{fc} were determined as soil moisture at $\phi = -1500$ kPa and at $\phi = -33$ kPa, respectively. In such a manner, hydrophysical parameters were estimated for all soil classes with the exception of "bedrock" that is absent in the USDA classification. For bedrock, we roughly estimated the hydrophysical parameters on the basis of our experience. For each basin, soil parameters were calculated as weighted average values with accounting for the spatial coverage of each soil class within a basin. For k_0, logarithms of k_0 were averaged. As to vegetation parameters, we assigned their typical values for each vegetation class and then calculated their weighted average values for each basin. The constructed *a priori* parameters set hereafter will be referred to as the MOPEX *a priori* data set ("APR_MOP").

The second set of *a priori* parameters was constructed from the global parameter data set provided within the framework of the Second Global Wetness Project (GSWP-2) (Zhao & Dirmeyer, 2003). Hereinafter, this *a priori* parameter set will be referred to as GSWP-2 *a priori* data set ("APR_GSW")

To improve runoff simulations based on *a priori* parameters, some parameters related to soil, which are more important from the viewpoint of runoff generation in the SWAP model, were calibrated by means of minimization of the root-mean-square-deviation (RMSD) between simulated and measured daily total runoff from a basin. Two parameter calibration techniques were applied: (1) manual calibration of hydraulic conductivity at saturation (referred to as "CAL1") which is the most crucial for runoff generation, and (2) automatic procedure for optimization of six parameters (W_{fc}, W_{wp}, W_{sat}, k_0, h_{soil}, h_{root}) (referred to as "CAL6"). The latter is based on a stochastic or Monte-Carlo technique. The first 20 years (1960–1979) were used for calibration in both cases. Each calibration procedure was applied for each of the two *a priori* parameters sets. As a result, we obtained four sets of calibrated parameters, related to the MOPEX (referred to as "CAL1_MOP" and "CAL6_MOP") and GSWP-2 (referred to as "CAL1_GSW" and "CAL6_GSW") data sets.

RESULTS

Simulations of river flow from the 12 MOPEX basins for a 39-year period (1960–1998) were performed using the land surface model SWAP with two sets of *a priori* estimated and four sets of calibrated parameters. The results of the simulations were compared with each other and with measured daily streamflow. The results are given in Figs 1–3.

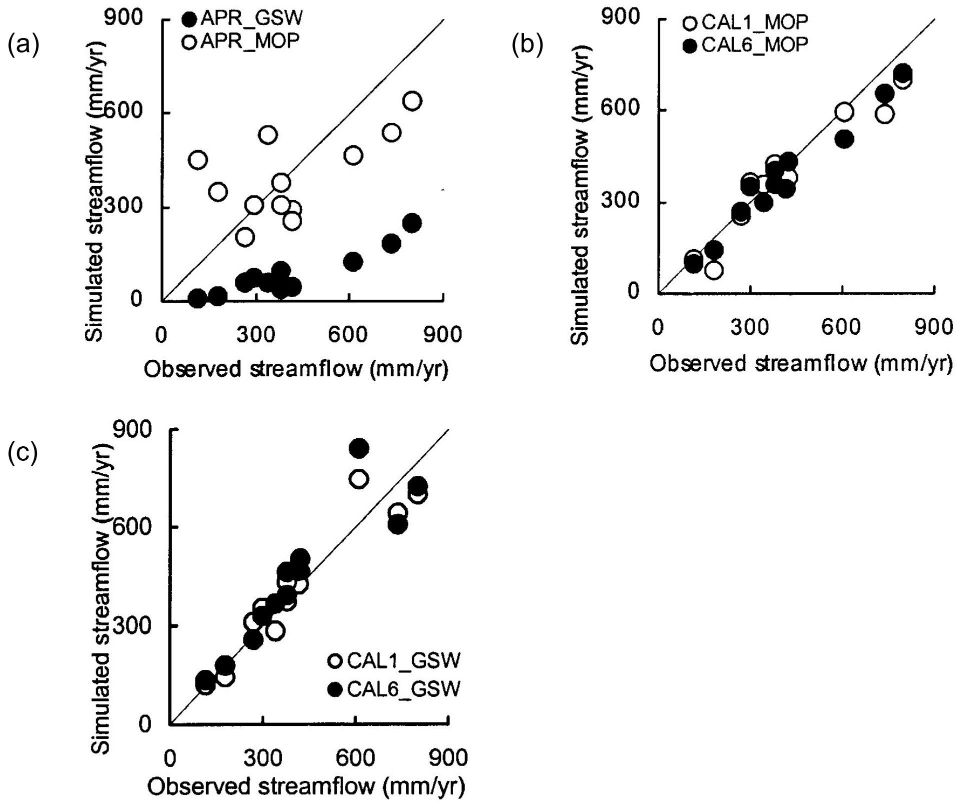

Fig. 1 Comparison of simulated and observed annual streamflow averaged over the 1960–1998 period for 12 basins when two sets of *a priori* estimated parameters (a), and one/six calibrated parameters from MOPEX (b) and GSWP-2 (c) parameter sets are used.

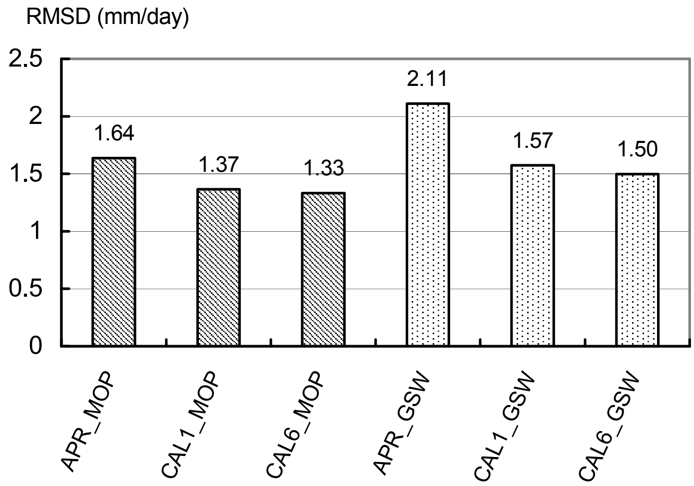

Fig. 2 The root-mean-square deviation between simulated (with six parameter sets) and observed daily streamflow, for 1960–1998, averaged over the 12 basins.

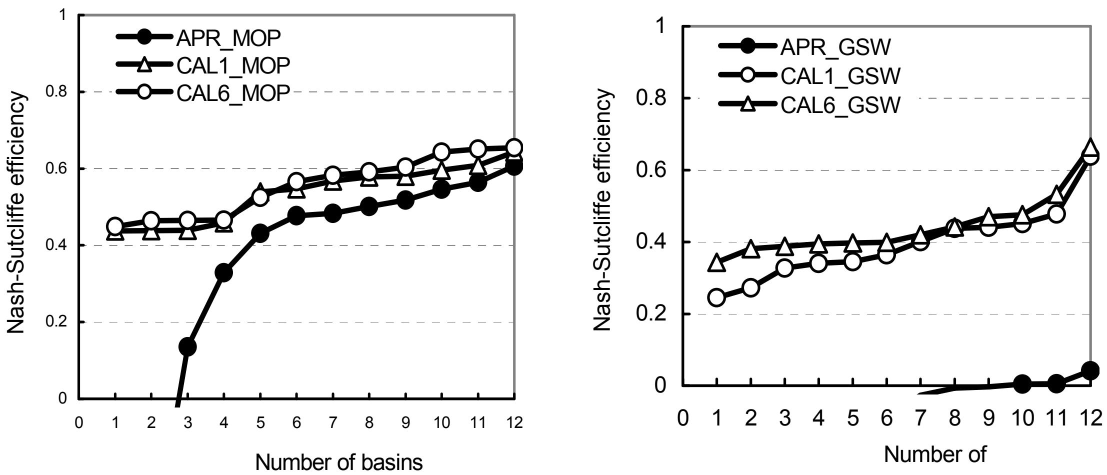

Fig. 3 Daily Nash-Sutcliffe efficiency of runoff simulation for the 12 MOPEX basins with *a priori* estimated and calibrated parameters sorted in increasing order.

Figure 1(a) shows the comparison between observed and simulated annual runoff (averaged over the 39-year period) with *a priori* parameters from the 12 basins. As seen, in the case of GSWP-2 *a priori* parameters, the simulated runoff is greatly underestimated (by 80%), however, the coefficient of correlation (r) is very high (0.91). For the MOPEX *a priori* parameters, runoff is underestimated by only 5%, while the scatter of points is much greater ($r = 0.58$).

Application of the calibrated parameters sets allowed us to substantially improve the results (Fig. 1(b),(c)). This is also confirmed by the values of RMSD between simulated and observed daily runoff (Fig. 2) and by the daily Nash-Sutcliffe efficiency of runoff simulations with different parameters sets (Fig. 3). As seen from these figures, the values of simulated runoff in the case of one calibrated parameter are close to those when six parameters were calibrated (both for MOPEX and GSWP-2 data sets) and the agreement between simulations and observations is better for the MOPEX parameters sets. The former means that soil hydraulic conductivity at saturation k_0 is the main soil parameter effecting runoff generation. The latter means that non-calibrated parameters derived from the MOPEX data are more adequate than those taken from the global GSWP-2 data set. Consequently, the main focus of the rest of this paper is on k_0 and on the MOPEX data set.

As seen from Fig. 4, *a priori* estimated values of k_0 differ from the optimal values obtained when six parameters from the MOPEX data set were calibrated. Again, application of the MOPEX *a priori* parameters shows better results compared to the GSWP-2 *a priori* parameters. So, it is reasonable to concentrate only on the MOPEX data set and to use the optimal values of k_0 to improve the procedure of estimation of *a priori* values of k_0 for ungauged basins. We attempted to do this in the following manner.

As mentioned above, the MOPEX *a priori* values of k_0 for each basin were estimated with the help of the soil texture classes provided (from the USDA classification) and the values of k_0, assigned for each class by the US Salinity Laboratory on the basis of investigation of different soil patterns of local character. At the same time, k_0 is known to be highly variable in space. The optimal values of k_0 obtained by us on the basis of model calibration may be treated as effective values for inhomogeneous areas and may be used for the calculation of new (effective) values of k_0 for the USDA soil classes. For this purpose, we derived nonlinear regression equations connecting the optimal k_0 for each basin with the content of clay and sand in a soil averaged over each basin (the average values of clay and sand are given in Table 1):

$$\mathrm{Ln}(k_0) = a_0 + a_1 \cdot \mathrm{CLAY} + a_2 \cdot \mathrm{SAND} + a_3 \cdot \mathrm{CLAY}^2 + a_4 \cdot \mathrm{SAND}^2 + a_5 \cdot \mathrm{CLAY} \cdot \mathrm{SAND} \qquad (1)$$

Using equation (1) it is easy to calculate new values of k_0 for the USDA soil textural classes because the average content of clay and sand for each class is known. However, the basins under study are very few in number and the average contents of clay and sand in their soils are within narrow ranges: 17% < clay < 43%, and 10% < sand < 35% (Table 1). For these ranges, the coefficients in equation (1) were obtained. Figure 5 shows the location of each basin (according to the averaged-over-a-basin

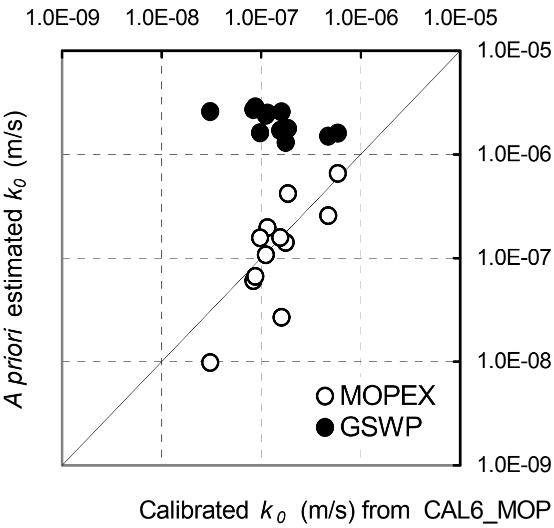

Fig. 4 Comparison of *a priori* estimated and calibrated hydraulic conductivity at saturation (in logarithmic scale).

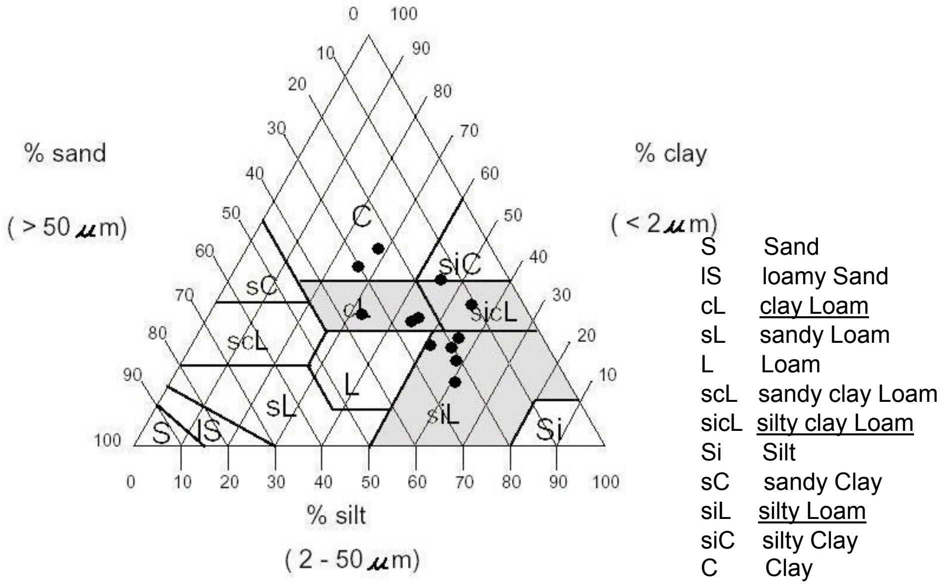

Fig. 5 Location of the 12 MOPEX basins (black circles) within the USDA-SCS soil textural triangle.

clay/sand content in a soil) within the USDA-SCS soil textural triangle. As shown, in the main, three soil classes correspond to the clay and sand ranges mentioned: clay loam, silty clay loam, and silty loam. This means that we can derive the new values of k_0 only for these three classes. These values were calculated by equation (1) using average contents of clay and sand for each soil class (Table 2). Increasing the number of basins will allow the calculation of the new values of k_0 for all soil classes and the construction of a new set of *a priori* parameters.

Table 2 Optimized values of hydraulic conductivity at saturation k_0 compared to initial values from the Rosetta code for the three USDA soil textural classes.

Soil class	Average content of clay (%)	Average content of sand (%)	k_0 (cm h^{-1}) (from the Rosetta code)	k_0 (cm h^{-1}) (calculated with equation (1))
Clay loam	34	32	0.038	0.006
Silty clay loam	34	10	0.047	0.079
Silty loam	14	18	0.053	0.127

CONCLUSIONS

1. When using the SWAP model on the 12 MOPEX basins, the application of *a priori* parameters estimated from the MOPEX database, results in better simulations of river flow than *a priori* parameters obtained from the GSWP-2 database.
2. Analysis of the soil parameter calibration results shows that: (a) the main soil parameter (in the model SWAP) controlling streamflow is the soil hydraulic

conductivity at saturation, and (b) an increase in the number of calibrated parameters only slightly improves the quality of streamflow simulation.

3. Following the MOPEX strategy, it seems possible to link the optimal soil hydrophysical parameters (from the viewpoint of streamflow simulation by SWAP) to more integral pedological characteristics (e.g. soil texture). For this purpose, it is necessary to increase the number of basins under study in order to cover all existing soil types.

Acknowledgements This work was supported by the Russian Foundation for Basic Research (Grant 05-05-64161). We acknowledge the MOPEX and GSWP-2 experiment organizers for providing us with the data to run the model.

REFERENCES

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

Duan, Q., Schaake, J., Andreassian, 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. *J. Hydrol.* **320**, 3–17.

Gusev, Ye. M. & Nasonova, O. N. (1998) The Land Surface Parameterization scheme, SWAP: description and partial validation. *Global Planet. Change*. **19**(1–4), 63–86.

Gusev, Ye. M. & Nasonova, O. N. (2002) The simulation of heat and water exchange at the land–atmosphere interface for the boreal grassland by the land-surface model SWAP. *Hydrol. Processes* **16**(10), 1893–1919.

Gusev, Ye. M. & Nasonova, O. N. (2003) Modelling heat and water exchange in the boreal spruce forest by the land-surface model SWAP. *J. Hydrol*. **280**, 162–191.

Koren, V. I., Smith, M., Wang, D. & Zhang, Z. (2000) Use of soil property data in the derivation of conceptual rainfall–runoff model parameters. In: *Preprints, 15th Conf. on Hydrol.* (Long Beach, California, January 2000), Paper 2.16. Am. Met. Soc.

Kuchment, L. S. & Gelfan, A. N. (2005) Toward parameter estimation of physically-based models of river flow formation under insufficient hydrological observations. *Meteorology & Hydrology* **12**, 77–87 (in Russian).

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models: 1. A discussion of principles. *J. Hydrol*. **10**(3), 282–290.

Schaake, J., Duan, Q., Koren V. & Hall A. (2001) Toward improved parameter estimation of land surface hydrology models through the Model Parameter Estimation Experiment (MOPEX). In: *Soil–Vegetation–Atmosphere Transfer Schemes and Large-Scale Hydrological Models* (ed. by A. J. Dolman, A. J. Hall, M. L. Kavvas, T. Oki & J. W. Pomeroy), 91–97. IAHS Publ. 270. IAHS Press, Wallingford, UK.

Schaap, M. G., Leij, F. J. & van Genuchten, M. Th. (1998) Neural network analysis for hierarchical prediction of soil water retention and saturated hydraulic conductivity. *Soil Sci. Soc. Am. J.* **62**, 847–855.

van Genuchten, M. Th. (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **44**, 892–898.

Zhao, M. & Dirmeyer, P. A. (2003) Production and analysis of GSWP-2 near-surface meteorology data sets. *COLA Technical Report 159*.