Performance comparison of a complex physics-based land surface model and a conceptual, lumped-parameter hydrological model at the basin-scale

THIAN YEW GAN[1], YEUGENIY GUSEV[2], STEPHEN J. BURGES[3] & OLGA NASONOVA[2]

1 *Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta T6G 2G7, Canada*
tgan@ualberta.ca

2 *Institute of Water Problem, Russian Academy of Sciences, Russia*

3 *Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195-2700, USA*

Abstract The Soil–Water–Atmosphere-Plants (SWAP) model of Gusev & Nasonova (2003) is a one-dimensional, land-surface model describing heat and water exchanges between the land surface and the atmosphere in a physics-based, analytical manner. The Sacramento model (SAC-SMA) is a deterministic, lumped-parameter, conceptual, rainfall–runoff model that does not explicitly consider spatial variability of terrain features, land-use, soil heterogeneities, and energy fluxes. SWAP is compared to SAC-SMA and both are tested using data from 12 MOPEX river basins located in the middle to southern eastern USA. Two model simulation comparisons are made: one using broad classifications of soil and vegetation to derive parameters (base case), and the second using calibrated parameters derived from measured data from the 12 MOPEX sites during the 20-year period (1960–1979). Only six of the SWAP parameters were automatically calibrated using a stochastic or Monte-Carlo technique. SAC-SMA was calibrated using manual and automatic means (Duan *et al.*, 1992). In terms of the goodness-of-fit between simulated and observed hydrographs, for the base case SAC-SMA was slightly better than SWAP. Both yielded results were unsuitable for any scientific inference. For the calibrated case SAC-SMA out-performed SWAP when comparing simulated hydrographs, but the simulations are still unsatisfactory. Analysis of the different behaviour of the models is presented.

Key words *a priori* parameter estimation; hydrological model; land surface model; MOPEX basins; parameter calibration

INTRODUCTION

As remotely sensed and digital elevation model (DEM) data have become more readily available, hydrological models used for simulating basin hydrological processes have been made more complex. More complex models may not necessarily produce better or more representative basin-scale simulations: simpler conceptual, lumped-parameter models can out-perform complex models, particularly when there is insufficient high quality data to support the complex models. The objective of this paper is to compare the performance of a relatively complex, physics-based model called SWAP (Gusev & Nasonova, 2003) with that of a simpler, conceptual, lumped-parameter model, the so-called Sacramento Soil Moisture Accounting model (SAC-SMA) (Burnash *et al.*, 1973). We tested model performance over a variety of hydroclimatic conditions using data from the 12 river basins of MOPEX (Model Parameter Experiment) (Duan *et al.*, 2005) in the southeastern United States (Fig. 1). The MOPEX basins represent a number of soil and vegetation types range from desert to very wet climatic conditions. Two model simulation comparisons are made: one using broad classifications of soil and vegetation to derive parameters (base case), and the second using calibrated parameters derived from measured data from the 12 MOPEX sites.

MOPEX RIVER BASINS

A full description of the 12 MOPEX basins is given in (Duan *et al.*, 2005); Fig. 1 and Table 1 provide summary information. All the basins are located below latitude 40ºN, and from Texas to the east. Consequently, there is minimal to no snow or frozen ground influence. Different hydrological and climatic conditions are well represented in these basins, ranging from desert conditions (basins 11 and 12) to very wet conditions (basins 1, 4, 5, 7 and 10) (Table 2). The catchment areas of these MOPEX basins range from about 1000 to 4400 km^2. (Table 1), the annual average precipitation ranges from 764 mm to 1564 mm, while the annual Runoff to Precipitation (*QP*) and Precipitation to Potential Evapotranspiration (*PPE*) ratios range from 0.09 to 0.59, and 0.45 to 2.34, respectively (Table 2).

MODEL DESCRIPTION

Land surface model SWAP

SWAP is a physically based land-surface model (LSM) describing heat and water exchange between the land surface and the atmosphere throughout a year at different

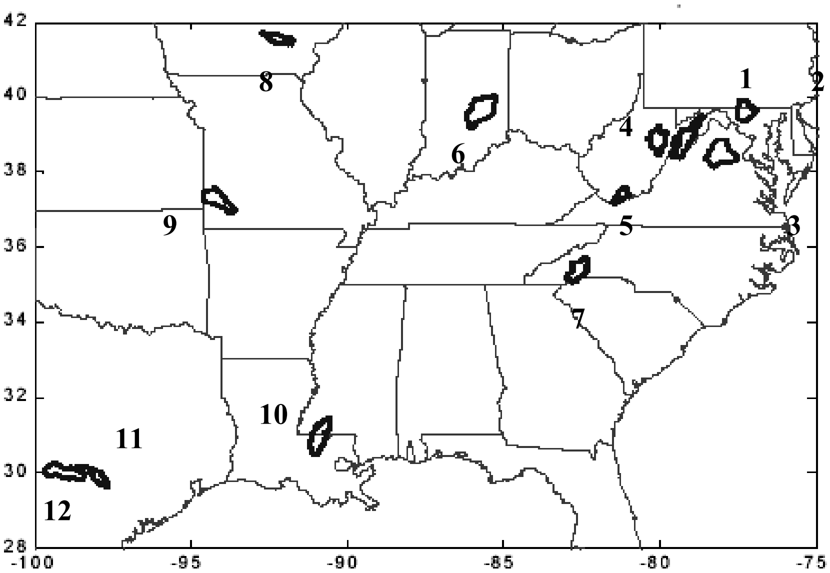

Fig. 1 Locations of 12 MOPEX river basins.

Table 1 Twelve MOPEX river basins.

No	Station#	Longitude, Latitude (°W, °N)	Area km²	Description of river basin
1	1608500	–78.65, 39.45	3810	South Branch Potomac River Nr Springfield, WV
2	1643000	–77.38, 39.39	2120	Monocacy R At Jug Bridge Nr Frederick, MD
3	1668000	–77.52, 38.32	4130	Rappahannock River Near Fredericksburg, VA
4	3054500	–80.04, 39.15	2370	Tygart Valley River At Phillipi, WV
5	3179000	–81.01, 37.54	1020	Bluestone River Near Pipestem, WV
6	3364000	–85.93, 39.2	4420	East Fork White River At Columbus, IND
7	3451500	–82.58, 35.61	2450	French Broad River At Asheville, N. C.
8	5455500	–91.72, 41.47	1480	English River At Kalona, IA
9	7186000	–94.57, 37.25	3020	Spring River Near Waco, MO
10	7378500	–90.99, 30.46	3320	Amite River Near Denham Springs, LA
11	8167500	–98.38, 29.86	3410	Guadalupe River Nr Spring Branch, TX
12	8172000	–97.65, 29.67	2170	San Marcos River At Luling, TX

Table 2 Annual water budget and 8 major USDA soil texture classifications of the 12 MOPEX river basins.

	Annual water budget				8 Major USDA soil texture[b] (%)							
No.	*PPE*	*QP*	*AEPE*	*P* [a] mm	SL	SIL	L	SICL	CL	SIC	C	Br
1	1.64	0.47	0.86	1040	19	13	23	3	0	0	0	36
2	1.15	0.34	0.76	1040	9	64	10	6	0	0	4	0
3	1.2	0.36	0.77	1030	5	35	4	9	36	0	5	4
4	1.76	0.5	0.87	1170	0	25	49	9	0	0	0	17
5	1.5	0.45	0.83	1020	0	16	32	34	0	0	6	1
6	1.21	0.36	0.77	1020	0	31	16	20	31	0	0	0
7	2.34	0.59	0.96	1380	11	3	48	1	17	0	15	5
8	0.89	0.25	0.67	890	0	7	0	92	1	0	0	0
9	0.96	0.28	0.69	1080	0	29	10	29	14	14	0	0
10	1.46	0.44	0.83	1560	1	58	4	21	8	0	0	0
11	0.45	0.09	0.41	770	0	0	13	0	14	3	24	17
12	0.56	0.13	0.49	830	1	0	17	4	23	2	43	9

[b] SL, sandy loam; SIL, silt loam; L, loam; SICL, silty clay loam; CL, clay loam; SIC, silty clay; C, clay; BR, bedrock.

PPE, Precipitation/Potential ET; *QP*, Annual runoff/Precipitation; AEPE, actual ET/PET

[a] *P*, average annual precipitation in mm.

scales (from local to global). It was developed to use atmospheric forcing data from the lowest atmospheric layer of GCMs (Global Circulation Models) or from any reference height (Gusev & Nasonova, 2000, 2003). The main distinctive feature of SWAP is a combination of its physically based treatment of the main processes and rationality of modelling technique used. The latter is provided by application of analytical methods (contrary to the usual practice of application of numerical ones) to solve the systems of equations and by a relatively small number of model parameters (compared to other LSMs). This allows one to avoid many problems associated with solving numerical equations (such as instability, great consumption of computer resources and calculation time) and parameters estimation.

The direct use of analytical methods has led to a non-traditional model structure for SWAP. Thus, in SWAP, a calendar year is divided into two seasons: warm and cold. For each season, a separate submodel was developed. These two submodels were linked into one general model, named SWAP. SWAP operates at different time steps (from 30 minutes to 1 day), depending on available forcing data and includes the following processes: interception of rainfall/snowfall by the canopy; evapotranspiration (including transpiration by plants, soil/snow evaporation, canopy evaporation); formation of snowpack on the ground and on the trees' crowns (including snow accumulation, snow evaporation, snowmelt, water yield of snow cover, refreezing of meltwater); formation of surface runoff and drainage; water infiltration into soil; water exchange between soil layers; interaction between soil water and groundwater; formation of the energy balance at the land surface; soil freezing/thawing.

Input data for the model includes atmospheric forcings (incoming shortwave and longwave radiation, air temperature and humidity, surface air pressure, wind speed, precipitation) and land surface parameters. Model outputs include all the energy and water balance components, all evapotranspiration components, different surface and subsurface state variables, and cold season characteristics.

SWAP was one of the LSMs used in different international experiments, including PILPS, RhoneAGG, SnowMIP, GSWP-2.

Hydrological model-Sacramento Model (SAC-SMA)

The SAC-SMA is a deterministic, conceptually-based rainfall–runoff model with spatially lumped parameters (Burnash *et al*., 1973). The model does not explicitly consider spatial variability of the land surface, soil heterogeneities and energy fluxes. The Sacramento model is used world wide to estimate streamflow for river basins. It is one of the most extensively studied conceptual rainfall–runoff models, e.g. Gan & Burges (1990a,b). The SAC-SMA model is generally applied by the US National Weather Service to river basins ranging from 300 km^2 to 5000 km^2. It can run at daily, 6-hourly, hourly or a smaller time step. Input to the model includes precipitation, temperature, pan evaporation data and lumped parameters of the basin physical characteristics. Full details for how we used this model for the 12 MOPEX sites are given in Gan & Burges (2005)

RESEARCH METHODOLOGY

Atmospheric forcing data to drive the models were provided by the Second MOPEX Workshop organizers and described in details in Duan *et al*. (2005). Additionally, some common basin characteristics (e.g. spatial distributions of different soil and vegetation classes within a basin), which can be used to derive the required model parameters, were provided. Two sets of parameters were obtained for each model: *a priori* estimated (base case) and calibrated (CAB) parameters. Concurrent 20-year long calibration data sets (January, 1960–December, 1979) were used to calibrate the models for each of the 12 MOPEX river basins (Table 1). Calibration was affected by attempting to match the daily simulated and recorded streamflow time series. The calibrations for each basin were tested (or "validated") using 19 years of data independent of the calibration experience (January 1980–December 1998). We prefer to use the term "tested"; strict requirements for model "validation" are rarely met.

The base case parameters for SAC-SMA were derived using a set of physically based relationships between SAC-SMA parameters and United States Department of Agriculture (USDA) soil properties developed by Koren *et al*. (2000). SAC-SMA was calibrated using a combination of the global optimization algorithm, the shuffle complex evolution method, SCE-UA, of Duan *et al*. (1992) and manual effort. Duan *et al*. showed that SCE-UA has overcome calibration problems in five areas that most, if not all local calibration algorithms suffer from: (i) regions of attraction—where more than one main convergence region exists; (ii) minor local optima—where there are small "pits" in each region; (iii) roughness—when the response surface is rough with discontinuous derivatives; (iv) sensitivity–poor and varying sensitivity of the response surface in the region of optimum, and nonlinear parameter interaction; and (v) shape—resulting from a non-convex response surface with long curved ridges. After calibrating SCA-SMA using SCE-UA, the resulting model parameters were fine tuned manually to obtain final calibrated model parameters (referred to as CAB) for each of the 12 MOPEX river basins. SAC-SMA was calibrated using 16 parameters: 11 "land parameter" parameters, plus three unitgraph ordinates (channel features), a precipitation scaling factor (that helps to account for bias—see e.g. Burges (2003)) and the potential evapotranspiration (PET) adjustment factor.

The base case data for SWAP were derived using the University of Maryland vegetation classification, the USDA soil texture classification, and some information from the International Satellite Land-Surface Climatology Project 1 (ISLSCP1) database. The calibrated parameters were obtained by means of automatic procedure for optimization (by minimization of root-mean-square-deviation between simulated and measured daily total runoff from a basin) based on a stochastic or Monte-Carlo technique, which as far as we know, also does not suffer from the five shortcomings identified by Duan *et al.* (1992); more extensive research is needed to clarify this issue. Only the six SWAP soil parameters that influence runoff (saturated hydraulic conductivity k_0, porosity W_{sat}, field capacity W_{fc}, wilting point W_{wp}, the depth of soil column h_{soil}, root zone depth h_{root}) were calibrated.

DISCUSSION OF RESULTS

The performance of SAC-SMA, when applied to the 12 MOPEX river basins using both base case and calibrated parameters, was compared with that of SWAP.

SAC-SMA model performance

Model parameters from optimization and manual calibration (denoted CAB) are shown in Table 3. The SAC-SMA model represents basin hydrologic response based on storage and release of water from five conceptual storages. The sum of these five storages should be consistent with physical soil water holding capability (Gan & Burges, 1990b). We show these values as "ΣStorage" in Table 3. The CAB based values for ΣStorage range from 328 mm to 710 mm. The latter value requires a deep equivalent soil column. The closest matches in this ratio were 0.92 for basins 1 and 6. We expect the conceptual storages obtained from calibration to reflect the soil storages of each basin, which is likely the case since the range of "ΣStorage" between basins varies widely.

Table 3 Comparisons of calibrated (CAB) SAC-SMA parameters and their respective performance, in terms of coefficient of efficiency (E_f) and Bias obtained at the calibration (E_f CAB) (1960–1979) and validation (E_f VAL) stages (1980–1998), and base case (no calibration), for the twelve MOPEX river basins.

	MOPEX River Basin #											
	1	2	3	4	5	6	7	8	9	10	11	12
Parameter	SCE–UA Optimized SAC–SMA Parameters with Manual Refinements											
UZTWM (mm)	5.4	39.5	3.4	6.3	20.2	49.	7.6	24.8	15.7	71.8	4.1	3.3
UZFWM (mm)	18.1	30	49.4	21.7	19.3	25.7	48.5	22.5	37.6	58.7	40.4	36.6
UZK	0.18	0.06	0.30	0.51	0.40	0.38	0.17	0.53	0.07	0.3	0.06	0.60
ZPERC	128.7	58.3	59.3	13.1	112.9	98.2	24.3	93.9	129.1	143.5	6.0	69.5
REXP	3.1	3.32	1.84	2.90	3.15	2.69	3.07	3.0	1.81	3.15	0.30	2.76
LZTWM (mm)	149.7	105	155.9	13.4	222.2	206.1	68.5	195.5	230.2	260.1	26.6	147.7
LZFSM (mm)	6.3	5.8	21.3	38.8	9.1	26.1	32.4	29.4	9.75	10.53	6.25	40.5
LZFPM (mm)	148.6	148.2	107.0	357.7	258.2	137.0	450.	80.0	30.	308.9	265.	147.3
LZSK	0.1	0.09	0.17	0.21	0.016	0.197	0.15	0.15	0.19	0.064	0.10	0.06
LZPK	0.01	0.015	0.007	0.013	0.010	0.008	0.01	0.018	0.004	0.005	0.01	0.011
PFREE	0.12	0.0	0.42	0.43	0.03	0.32	0.45	0.264	0.22	0.245	0.17	0.38
ΣStorage (mm)*	328.1	328.5	337	437.9	529	443.9	607	352.2	323.3	710.0	342.4	375.4
Ratio**	0.92	0.74	0.80	1.47	1.45	0.92	1.38	0.85	0.83	1.38	1.24	1.30
	Calibration (Jan. 1960–Dec. 1979) and Validation (Jan. 1980–Dec. 1998) Results											
E_f CAB	55.9	79.5	72.7	35.8	68.0	75.7	85.2	53.4	76.9	77.5	51.2	56.5
E_f VAL	60.1	63.5	66.4	39.0	60.8	79.6	84.6	61.7	81.6	80.6	51.5	78.3
Bias CAB	–0.9	0.6	–4.2	–13.8	–2.2	2.9	1.6	–1.8	–9.7	–5.9	0.8	3.3
Bias VAL	1.0	–0.2	–7.6	–13.6	2.3	8.9	2.1	20.8	–11.3	–7.8	–4.9	1.5
	A priori CAB (Jan. 1960–Dec. 1979) and VAL (Jan. 1980–Dec. 1998) Results											
E_f Base case C	48.9	69.8	23.3	50.3	55.4	75.1	40.2	46.9	75.0	72.2	10.2	16.2
E_f Base case V	45.1	55.1	19.4	53.6	45.3	79.4	29.9	55.0	80.5	81.0	49.6	66.4
Bias Base Case C	–5.0	–7.5	–71.6	–31.3	–0.4	1.0	–5.1	–8.7	–21.3	–5.7	–64.1	–69.5
Bias Base Case V	–2.3	–7.8	–72.0	–29.9	4.2	7.2	–6.3	14.4	–19.3	–7.7	–57.5	–63.3
Correlation CAB–BaseCase[a]	0.13	0.68	–0.16	–0.3	–0.36	–0.18	0.18	0.12	–0.07	0.05	–0.21	0.37

a = Correlation between each set of SAC-SMA CAB and Base Case parameters normalized by their respective median values; * ΣStorage = Sum of conceptual storages UZTWM, UZFWM, LZTWM, LZFPM, and LZFSM; **Ratio = CAB ΣStorage /BaseCase ΣStorage.

Table 3 shows for the calibration stage (1960–1979) two goodness-of-fit statistics between simulated and observed daily average flow rates for each river basin: the coefficient of efficiency (E_f) (also called the Nash and Sutcliffe coefficient of efficiency) and bias (Bias). None of these modelled hydrographs could be judged "satisfactory" given the stringent demonstrated requirements of Burges (2003); high quality calibrations should have values of E_f in excess of 95%. The median E_f for the 12 CAB cases is over 70% (Fig. 2). The CAB based simulations are generally satisfactory except for the two driest river basins, Guadalupe River and San Marcus River located in Texas (MOPEX Basins 11 and 12), and for Rappahannock River of VA (Basin 3) where E_f drops to less than 25%.

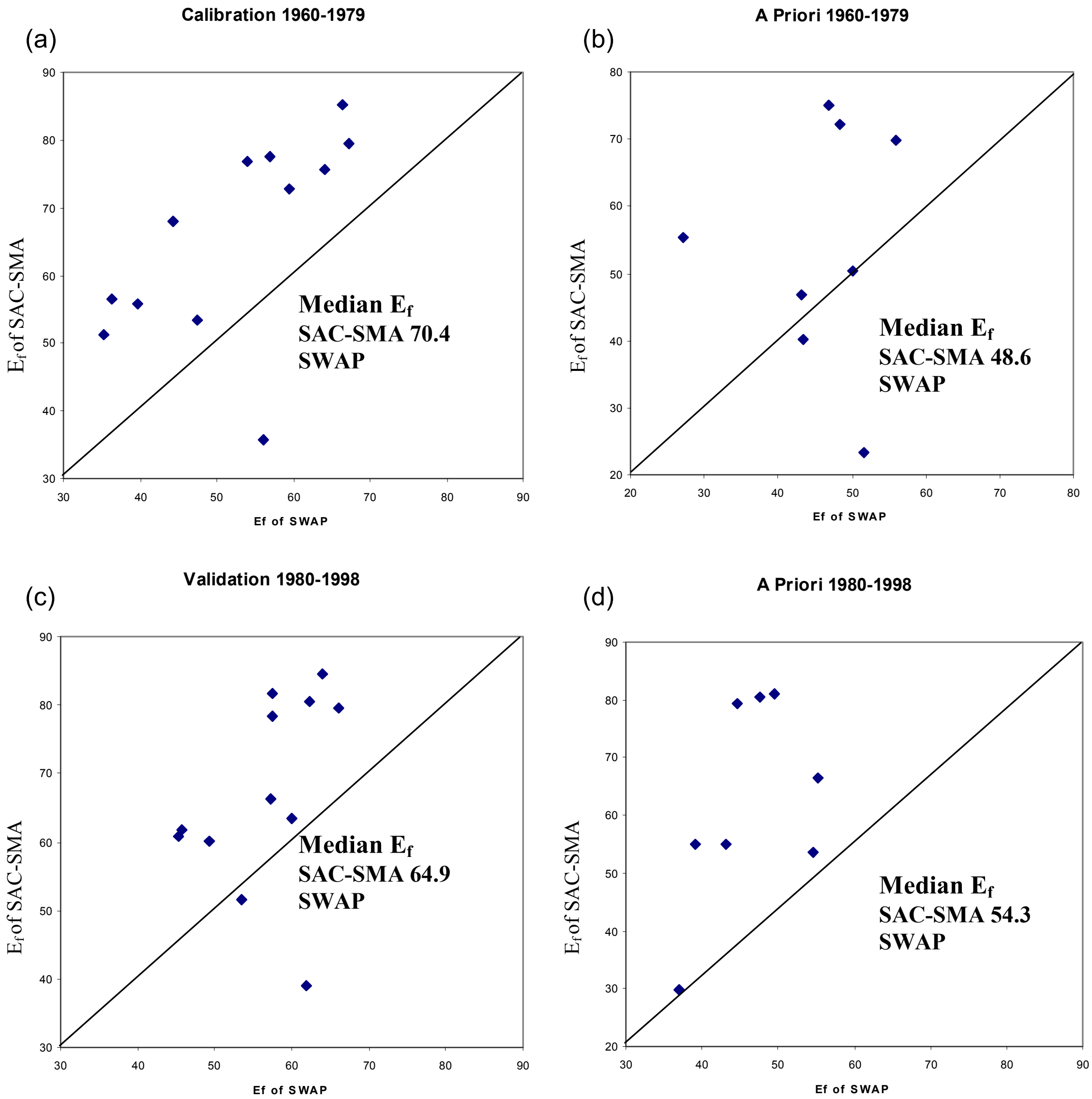

Fig. 2 Scatterplots of the coefficient of efficiency (E_f) of daily simulated streamflow for the 12 MOPEX river basins using SAC-SMA and SWAP for: (a) calibration stage (1960–1979), (b) base case stage (1960–1979), (c) validation stage (1980–1998), and (d) base case stage (1980–1998).

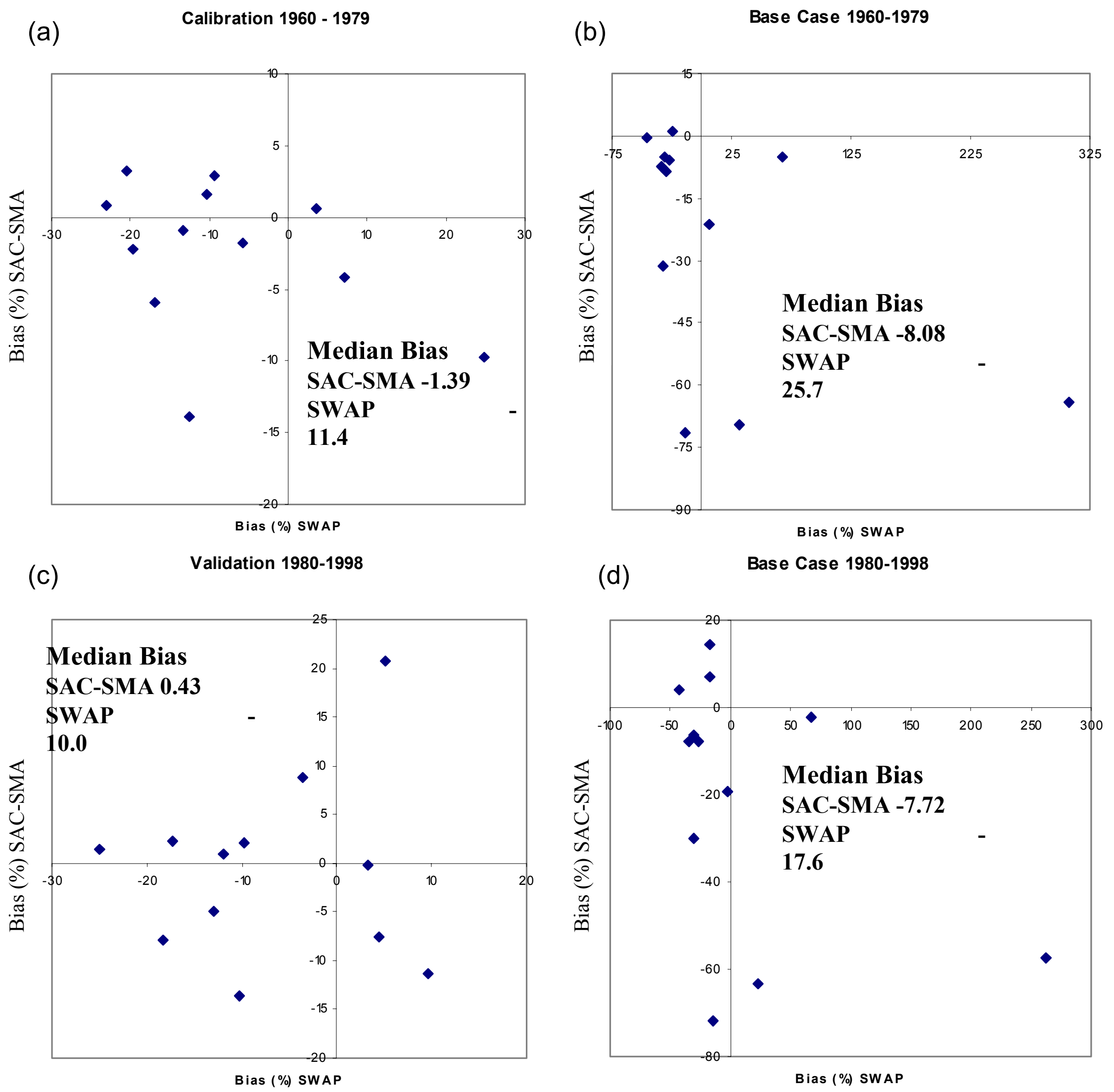

Fig. 3 Scatterplots of the bias (BIAS) of the 12 MOPEX river basins using SAC-SMA and SWAP for: (a) calibration stage (1960–1979), (b) base case stage (1960–1979), (c) validation stage (1980–1998), and (d) base case stage (1980–1998).

At the testing (validation) stage (1980–1998), the relative modelled hydrograph performances based on calibrated parameters (denoted as VAL) are similar, with the former mostly better than the latter except for Basin 4. We suspect that data inadequacies or errors in Basin 4 data make a meaningful calibration (E_f = 35.8%) an almost impossible task. The median E_f for VAL drops to 64.5%. It is to be expected that, moving from the calibration to the validation stages, the performance of calibrated parameters (CAB) to generally decrease because the validation (VAL) data used are independent of the calibration experience (see Gan & Biftu (1996), and Gan & Burges (1990a,b)).

Among the 12 sets of CAB parameters, the standard deviations of the normalized moisture storage parameters (UZTWM, LZFSM and LZFPM) are 1.88, 0.83 and 0.84, respectively, while the corresponding standard deviations of the three normalized extraction parameters (UZK, LZSK and LZPK) are 0.84, 0.64, and 0.50, respectively. Physical reasoning supports our finding for CAB parameters that the standard deviation of UZK is higher than LZSK and LZPK. UZK controls release from the upper zone storage (related to surface runoff) while LZPK controls storage release from the lower zones (related to sub-surface runoff). Wetter basins that are more dominated by surface runoff have a wider range of upper zone storages than dry basins that are dominated by sub-surface runoff. Given that the 12 MOPEX basins represent a wide range of hydrologic conditions, from wet (Basin 7), to medium, to very dry (Basins 11 and 12), we expect their dominant hydrologic processes to differ widely, which should be more reflected in surface runoff parameters like UZK than in sub-surface parameters like LZPK or LZSK, or similarly in moisture storages UZTWM than in LZTWM (Table 3).

Performance of SWAP Model

The results of SWAP model simulations are given in Table 4. In the case of *a priori* (base case) parameters, the efficiency of simulated runoff is negative for basins 11, 12 and 1 (for 1960–1979): the *a priori* estimated parameters for these basins are inadequate. This could be expected because *a priori* parameters values for SWAP were derived using the USDA soil texture classification where class “bedrock” (Br) is absent; Br soil parameters could only be estimated very roughly. Bedrock covers 9, 17, and 36 % of the area of basins 12, 11 and 1, respectively. Consequently, the error of *a priori* estimation of model parameters for these basins may be large. Basins 11 and 12 are the driest of the 12 MOPEX basins. The influence of this factor may be evident when comparing the results for basins 4 (E_f = 50 and 55% for 1960–1979 and 1980–1998 years, respectively) and 11 (E_f <0). The bedrock coverage is the same (17%) in basins 11 and 14; basin 11 is the driest and basin 4 is the second wettest.

Table 4 Comparisons of calibrated SWAP parameters and their respective performance, in terms of coefficient of efficiency (E_f) and Bias obtained at the calibration (E_fCAB) (1960–1979) and validation (E_fVAL) stages (1980-1998), and Base Case (no calibration), for the twelve MOPEX river basins.

	MOPEX River Basin #											
	1	2	3	4	5	6	7	8	9	10	11	12
Parameter	Normalized by median Optimized SWAP Parameters											
Wfc (m^3 m^3)	1.11	0.82	0.91	0.90	0.99	0.97	1.06	1.03	1.15	1.01	1.04	0.97
Wwp(m^3 m^3)	0.51	0.77	0.76	1.14	0.80	1.20	1.12	0.88	1.25	1.49	1.02	0.98
Por (m^3 m^3)	1.01	1.10	1.13	1.02	1.13	0.95	0.81	1.13	0.93	0.99	0.89	0.97
Ko (m s^{-1})	1.45	0.735	0.90	0.14	0.68	0.78	0.86	1.61	1.79	1.10	5.82	7.52
hroot (m)	1.06	0.75	0.63	0.54	0.57	1.37	0.58	1.11	1.05	1.46	1.32	0.95
ho (m)	1.29	1.23	1.06	0.96	1.06	0.69	1.01	0.97	0.85	1.0	0.91	1.13
	Calibration (Jan. 1960–Dec. 1979) and Validation (Jan. 1980–Dec. 1998) Results											
E_fCAB	39.6	67.2	59.4	56.1	44.4	64.2	66.5	47.4	54	57	35.3	36.3
E_fVAL	49.9	60	57.2	61.8	45.4	66	64	45.8	57.4	62.3	53.4	57.5
Bias CAB	–13.4	3.5	7.1	–12.5	–19.7	–9.4	–10.3	–5.8	24.8	–16.8	–23.0	–20.5
Bias VAL	–11.9	3.2	4.5	–10.2	–17.3	–3.5	–9.8	5.1	9.7	–18.3	–12.9	–25.0
	A priori CAB (Jan. 1960–Dec. 1979) & VAL (Jan. 1980–Dec. 1998) Results											
E_f Base Case C	–36.4	56	51.6	50.1	27.1	46.9	43.3	43.1	46.8	48.4	–367	–75
E_f Base Case V	20.2	39.2	43.5	54.6	28.2	44.6	36.9	43.1	47.7	49.5	–227	55.3
Bias Base Case C	67.8	–34.0	–13.8	–32.9	–45.5	–24.4	–31.5	–30.3	6.1	–26.9	307	31.2
Bias Base Case V	67.7	–33.8	–14.7	–30.4	–41.9	–17.7	–30.4	–17.5	–1.9	–26.3	262	22.7

Application of the calibrated parameters resulted in better performance of SWAP compared to the base case. For the calibration period, E_f varies among the basins between 35 and 67%, Bias varies from –23 to 25%; for the validation period, E_f and bias range from 45–66%, and –25 to 10%, respectively. The median E_f for 12 basins increased by 10% for 1960–1979 and 14% for 1980-1998; the median Bias reduced by 14 and 8%, respectively. Basins 11 and 12 (with large bedrock coverage and driest conditions) are among the worse with respect to both Bias and E_f. For 8 basins the simulated runoff is underestimated. According to our experience underestimation of runoff modelled by SWAP usually results from underestimation of precipitation.

Comparing SAC-SMA with SWAP model performance

SAC-SMA and SWAP are comparable in their performance for the base case parameter simulations (without calibration), though SAC-SMA has a slight edge over SWAP. For the 12 MOPEX sites the median E_f for simulated daily streamflow was 49% (SCA-SMA) and 45% (SWAP) for the calibration period (1960–1979), and, respectively, 54% and 43% for the "test period" (1980–1998). The corresponding median bias values were –26% (1960–1979) and –18% (1980–1998) for SWAP, which are higher than that –8.1% and –7.7% for SAC-SMA. Why did SAC-SMA appear to perform better than SWAP? We offer several possibilities. A hydrological model developed for streamflow simulations should reproduce streamflow better than a land surface model which places emphasis on heat and water exchange processes occurring in a complex and multifactor soil-vegetation/snow cover-atmosphere system. A second explanation may be connected with better estimation of *a priori* parameters values for SAC-SMA. It should be noted, however, that both models produce simulations that are completely inadequate for any question of science and may have limited use for hydrological decision making. This is disappointing in that the model parameters are derived as if "ungauged catchments" were being modelled.

The second set of comparisons involved simulations from calibrated models of each of the 12 MOPEX catchments. The simple Sacramento model consistently out-performed the much more complex SWAP model when considering both the calibration period and the testing "validation" period (Tables 3 and 4, Figs 2 and 3).

The first possible reason that SAC-SMA outperformed SWAP in both calibration and validation stages could be because SWAP was only calibrated in terms of six soil parameters whereas SAC-SMA was calibrated with the 11 parameters shown in Table 3, plus the three unitgraph ordinates, one precipitation scaling factor (that helps account for bias—see e.g. Burges (2003)) and one potential ET adjustment factor, a total of 16 parameters. This was partly why after calibration (20 years of data), the median streamflow E_f of SAC-SMA was 70%, *vs* 55% for SWAP while for the test (validation) period (19 years), the gap between them is more modest, 65% of SAC-SMA over 58% for SWAP. For SWAP, calibration reduced the Bias relative to the base case parameters by about 8 to 14% (–26 to –11% for 1960–1979 and –18 to –10% for 1980–1998). However, calibration further reduced the Bias of SAC-SMA to –1.4% (1960–1979) and 0.4% (1980–1998), respectively, i.e. for SAC-SMA, the bias was reduced by 7–8%.

One more additional explanation why the site specific calibrated SAC-SMA model produced better simulated streamflow than the SWAP model may be related to the different calibration techniques used. SWAP parameter calibration is based on a stochastic or Monte-Carlo technique which may not be as effective as that of SCE-UA. More complete testing of the stochastic optimization scheme is needed before this issue can be resolved.

CONCLUSIONS

On the basis of results obtained from driving a complex, physics-based land surface model (SWAP) and a simple, conceptual lumped-parameter hydrological model (SAC-SMA) on 12 MOPEX river basins ranging from 1000 to 4400 km^2 in area, and representing a wide range of hydrological and climatic conditions (dry to wet), it seems that a land surface model may not necessarily reproduce streamflow better than a simple, hydrological model. This could be expected at the first stage of this study (when *a priori* parameters were used) because the model designed for streamflow simulation should perform better than the model that emphasizes heat and water exchange between the land surface and the atmosphere. Better performance (with respect to streamflow simulation) of SAC-SMA at the second stage (when calibrated parameters were used) may have resulted because SWAP was not as completely calibrated as SAC-SMA (6 versus 16 parameters). More extensive tests involving more models and basins are needed to confirm the finding of this study, that conceptual hydrologic models might model basin-scale hydrological processes more reliably than complex models.

Acknowledgements This research was partly funded by the Natural Science and Engineering Research Council (NSERC) of Canada. All the MOPEX data were provided jointly by the National Oceanic and Atmospheric Administration/National Weather Service, Silver Spring, and SHARA/University of Arizona.

REFERENCES

Burges, S. J. (2003) Process representation, measurements, data quality, and criteria for parameter estimation of watershed models. In: *Calibration of Watershed Models* (ed. by Q. Duan, H. Gupta, S. Sorooshian, A. N. Rousseau & R. Turcotte), 283–299. AGU Monograph, Water Science and Application 6.

Burnash, R. J. C., Ferral, R. L. & McGuire, R. A. (1973) A generalized streamflow simulation system-conceptual modeling for digital computers. US Department of Commerce, National Weather Service and State of California, Department of Water Resources, California, USA.

Clapp, R. B. & Hornberger, G. M. (1978) Empirical equations for some soil hydraulic properties. *Water Resour. Res.* **14**(4), 601–604.

Duan, Q., Sorooshian, S. & Gupta, V. K. (1992) Effective and efficient global optimization for conceptual rainfall runoff models. *Water Resour. Res.* **28**(4), 1015–1031.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., 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. (2005) Model parameter estimation experiment (MOPEX): overview and summary of the second and third workshop results, MOPEX Special Issue. *J. Hydrol.* **320**, 3–17, doi:10.1016/j.jhydrol.2005.07.031.

Gan T. Y. & Burges, S. J. (1990a) An assessment of a conceptual rainfall–runoff model's ability to represent the dynamics of small hypothetical catchments, 1: models, model properties, and experimental design. *Water Resour. Res.* **26**(7), 1595–1604.

Gan T. Y. & Burges, S. J. (1990b) An assessment of a conceptual rainfall–runoff model's ability to represent the dynamics of small hypothetical catchments, 2: hydrologic responses for normal, extreme rainfall. *Water Resour. Res.* **26**(7), 1605–1619,.

Gan, T. Y. & Biftu, G. F. (1996) Automatic calibration of conceptual rainfall–runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resour. Res.*, AGU, Dec. **32**(12), 3513–3524.

Gan, T. Y. & Burges, S. J. (2005) Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. *J. Hydrol.* **320**, 117–131, doi:10.1016/j.hydrol.2005.07.008.

Gusev, Ye. M. & Nasonova, O. N. (2000) An experience of modeling heat and water exchange at the land surface on a large river basin scale. *J. Hydrol.* **233**(1–4), 1–18.

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**(1–4), 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: *15th Conf. Hydrology* (Long Beach, California, USA), Am. Meteor. Soc., 10–14 January, 103–106.

Lohmann, D., Lettenmaier, D. P., Liang, X., Wood, E. F., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R. E., Duan, Q., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchell, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z. -L. & Zeng, Q. -C. (1998) The project for intercomparison of land-surface parameterization schemes phase-2(c) Red-Arkansas River basin experiment. 3. Spatial and temporal analysis of water fluxes. *Global Plan. Change* **19**(1–4), 161–179.

Luo, L., Robock, A., Vinnikov, K. Y., Schlosser, A. C., Slater, A. G., Boone, A., Braden, H., Cox, P., de Rosnay, P., Dickinson, R. E., Dai, Y. -J., Duan, Q., Etchevers, P., Henderson-Sellers, A., Gedney, N., Gusev, Ye. M., Habets, F., Kim, J., Kowalczyk, E., Mitchell, K., Nasonova, O. N., Noilhan, J., Pitman, A. J., Schaake, J., Shmakin, A. B., Smirnova, T. G., Wetzel, P., Xue, Y., Yang, Z. -L. & Zeng, Q.-C. (2003) Effects of frozen soil on soil temperature, spring infiltration, runoff: results from the PILPS 2(d) experiment at Valdai, Russia. *J. Hydrometeorol.* **4**(2), 334–352.