New Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009.

An evaluation of multi-basin hydrological modelling for predictions in ungauged basins

CHANTAL DONNELLY, JOEL DAHNE, GÖRAN LINDSTRÖM, JÖRGEN ROSBERG, JOHAN STRÖMQVIST, CHARLOTTA PERS, WEI YANG & BERIT ARHEIMER

Swedish Meteorological and Hydrological Institute, SE-601 76 Norrköping, Sweden

chantal.donnelly@smhi.se

Abstract Availability of data is the limiting condition for reliable hydrological predictions in many regions of the world. One possible solution is to increase the scale of hydrological models so as to encompass data sparse regions within larger regions where data are available. A growing number of hydrological model studies use the spatial distribution of soils and vegetation to predict spatially-varying catchment behaviour. This introduces the potential to setup and use hydrological models over larger scales and simultaneously over several river basins. Such models can be set up easily, over very large regions, using freely available global data sets. However, the questions remain: (a) can a multi-basin hydrological model be calibrated using a uniform parameter set? and (b) can the uniformly calibrated, multi-basin hydrological model be used to make reasonable predictions in ungauged basins? Two different multi-basin regions were set up and calibrated using uniform land-use and soil-type dependent parameters in the HYPE hydrological model. Results indicate that a reasonable calibration can be achieved for a large multi-basin model set using readily available global input databases and using a homogenous parameter set.

Key words distributed hydrological modelling; predictions in ungauged basins

INTRODUCTION

The ability to calibrate and model temporal and spatial streamflow variations in several major river basins simultaneously is an attractive prospect. For example, comparable estimates of the ecological status of water bodies across the European continent are requested by the European Water Framework Directive (EWFD) (Council of the European Communities, 2000), but estimates on this scale are determined by different assessment systems in different countries. Harmonised water quantity and quality models applied with high resolution on a multi-basin scale could provide a reference model across which these national models could be compared. In combination with local models, multi-basin models could contribute to a model ensemble for model uncertainty estimates. It is also hypothesised that reasonable predictions of flows in ungauged basins may be obtained by increasing the scale of hydrological models so as to encompass data sparse basins within larger regions where data are available. Increases in computational efficiency and access to global-scale databases of climatic driving data, watersheds, discharge, land use and soils (e.g. Feyen *et al.*, 2008) begin to make the use of such multi-basin models feasible. Multi-basin models may also help address the issue of global water cycle modelling in global circulation models (Ducharne *et al.*, 2003), transboundary water issues (e.g. Council of the European Communities, 2000), and large-scale climate impact studies (Andréasson *et al.*, 2004).

Large-scale hydrological models, distributed and homogenous both in input and in calibration, have been widely discussed in the literature (e.g. Beldring *et al.*, 2003; Feyen *et al.*, 2008). Large-scale distributed input/distributed calibration models such as that described by Andreasson *et al.* (2004), are used, e.g. operationally for flood forecasting in Sweden and nutrient transport mapping (Arheimer & Brandt, 1998). The disadvantage of such models is the time taken to conduct many local calibrations, and the choice of model parameters in ungauged basins.

To date, there have been few published studies of high resolution multi-basin models. Global-scale models are used to derive global water balances within global circulation models; however, these models generally have a low resolution and cannot reproduce measured daily flow variations in gauged basins (Miller *et al.*, 1994; Ducharne *et al.*, 2003). Beldring *et al.* (2003) constructed a multi-basin model for the whole of Norway using monthly simulations of a 1-km^2 resolution distributed HBV model. Results indicated large water balance errors; however, model validation on 43 independent catchments indicated some ability to predict monthly flow variation in ungauged basins.

Multi-basin models generally cover large heterogeneous areas, so a land-use/soil-type dependent approach is preferable if parameters are to be transferable to ungauged basins. Dunn & Lilly (2001) investigated linking parameter values to soil type in the DIY model using a hydrological soil type classification (HOST). Success of the calibrated parameters in independent basin validation depended on the processes represented by the parameter, indicating that some processes were better represented in the model and in-data formulation than others. Similar results were obtained by Marachal & Holman (2005) using a new conceptual daily rainfall–runoff model, CRASH, to resolve these differences on the catchment scale. These studies were limited to individual basins and sub-basins, so there remains scope to investigate land-use and soil dependent parameters over several basins. How does calibration quality vary with the scale of the calibration and the number of runoff stations used? Lumped and distributed parameter approaches were compared by Feyen *et al.* (2008), who considered the calibration quality both downstream and at internal streamflow stations for three different calibration approaches over a region of 10 000 km^2, a tributary of the Danube River. The results showed a small improvement in streamflow estimates when upstream streamflow records were included in the calibration, and a larger improvement when semi-distributed parameters were used.

Despite these results, a uniform calibration approach where parameters are linked to soil type and land use is preferable to a semi-distributed calibration in a multi-basin model. A uniform calibration approach is defined here as a single set of land-use and soil-type linked parameters, simultaneously optimised to all the available flow data in the model domain, regardless of whether or not an individual river within the domain is gauged. A semi-distributed calibration approach uses different parameter sets in different gauged sub-basins. Simply put, calibration time will be less for a single calibration of a uniform parameter set. Secondly, the uniform approach is more transparent, which is important where models are used in decision making (Arheimer *et al*., 2007). Finally, semi-distribution of parameters relies on gauged data downstream of the area of interest. Uniform calibration relies on a sufficient amount of gauged data surrounding the area of interest.

As it is for large and regional scale hydrological modelling, the quality of a multi-basin model will be a combination of: (1) calibration quality: how well the model can be calibrated to observations within the modelled area; (2) validation: how well the model compares to observations within the modelled area not used in calibration, i.e. a proxy-basin test (Klemes, 1986); and (3) uncertainty: the uncertainty of the model output being a combination of uncertainties in model inputs, process representations and calibrated parameters (Beck, 1987). By selecting a hydrological model in which variations in land use, soil type, lakes and river routing are explicitly accounted for in calibration, a uniform set of parameters determined by calibrating the model once over the entire modelled region should yield a reasonable model validation (Sivapalan *et al.,* 2003). This, in turn, suggests the ability to make reasonable predictions of flow in ungauged basins within the modelled multi-basin region.

This paper describes the development of multi-basin hydrological models over two different regions and addresses the first two model quality pre-requisites by stating the following research questions: (a) Can a multi-basin hydrological model be calibrated using a uniform parameter set? and (b) Can the uniformly calibrated, multi-basin hydrological model be used to make reasonable predictions in ungauged basins? Model sensitivity is addressed by Strömqvist *et al.* (2009). The first multi-basin model is the S-HYPE application, encompassing the land area of Sweden, and the second multi-basin model is the E-HYPE application, encompassing the European continent, from the Ural Mountains to the British Isles. The term “multi-basin” is used here to suggest a model set up and calibrated on a national or even continental scale, incorporating several entire river basins and their outlets to the sea. Modelling was performed using the HYPE hydrological model (Lindström *et al*., 2009), a new, daily time-stepping hydrological model developed at the Swedish Meteorological and Hydrological Institute (SMHI).

DATA AND METHODS

Two different multi-basin regions, encompassing wide variations in geomorphology, soil types, land uses and topography, were modelled using the HYPE model. The HYPE model (Lindström *et al.*, 2009) is a semi-distributed conceptual model for HYdrological Predictions for the Environment. Modelled river basins are divided into sub-basins. The sub-basin resolution is the resolution of the input forcing data and the geographic input data (elevation, slope and lake percent). Within each sub-basin the proportion of each soil and land-use combination, called class, is also specified. Digital Elevation Model data in addition to soil and land cover maps are used for these inputs and the model uses meteorological forcing in the form of daily precipitation and air temperature.

The processes of snow melt, evapotranspiration, surface runoff and infiltration, percolation and macropore flow through the soil, tile drainage, and groundwater outflow to the stream from soil layers with water content above field capacity are simulated for each class. Figure 1 shows the processes modelled and detailed water pathways. The model includes several soil layers and detailed pathways for water in the soil because it was developed with a principal focus on nutrient transport modelling (Arheimer *et al.*, 2008) and is therefore somewhat overparameterised for water discharge modelling. Here only the streamflow module is described. Outflow from the soil layers is then routed within and between sub-basins using a river routing routine which simulates delay in rivers and storage in lakes. Lakes may be either regulated or have a natural rating curve determining their outflow. The model has several parameters for calibration which are either soil type or land-use dependent (e.g. field capacity) or global (e.g. routing parameters). When calibrated on a large scale, the parameters are assumed to be transferable to ungauged neighbouring catchments.

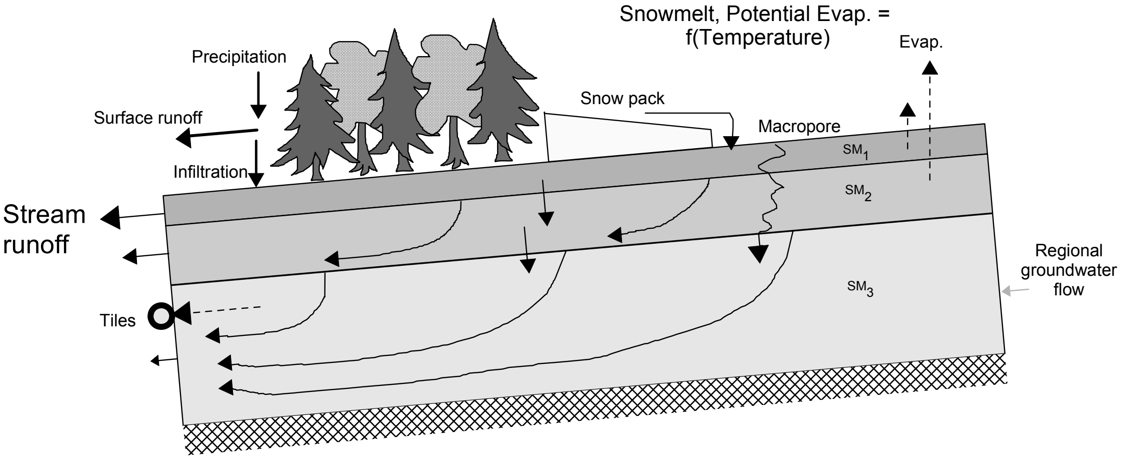

Fig. 1 Schematic HYPE model structure for a land class with three soil layers.

The region modelled first with S-HYPE, encompasses the land area of Sweden (including river basins that extend into Norway). The second region model, E-HYPE, encompasses most of the European continent, from the Ural Mountains to the British Isles and south to the Danube River basin. This application is split into six different multi-basin calibration regions. Table 1 summarises the databases used and resolution details of the two applications.

The S-HYPE application, developed to support Sweden's reporting to the EWFD, uses very high resolution sub-basins (median = 18 km^2) and flow routing between sub-basins delineated by hand by the Swedish Water Archive at the Swedish Meteorological and Hydrological Institute (SMHI). The E-HYPE application uses lower resolution sub-basins (median = 1000 km^2) in accordance with available input data resolution. Sub-basin delineation and direction of flow

Table 1 Model application set-up and input data.

	S-HYPE	E-HYPE
Area	476 000 km^2	10 200 000 km^2
Median sub-basin resolution	18 km^2	1000 km^2
No. of sub-basins	17 313	10 200
Topography/routing	Swedish Water Archive	Hydro1K (USGS, 2000)
Forcing data	PTHBV, 1961–2003 (Johansson, 2002) Resolution = 4 km	ERAMESAN 1980–2004 (Jansson *et al.*, 2007) Resolution = 11 km
Land cover	CORINE (CLC 2000)	ECOCLIMAP (Champeaux *et al.*, 2005)
Soil types	SGU	ECOCLIMAP (Champeaux *et al.*, 2005)
Runoff data	SMHI WISKI discharge database	GRDC (GRDC 2008)
No. of calibration regions	1	8
No. of calibration stations	30	1223
No. of validation stations	260	na
Station cover (km^2/station)	1641 km^2/station	82 800 km^2/station

between sub-basins for E-HYPE was made using the HYDRO1k Elevation Derivative Database from EROS (USGS 2000; Strömqvist *et al.*, 2009). For both applications, routing between sub-basins was checked by comparing the modelled area flowing to a gauging station with the station's reported runoff area. This revealed a number of problems with the HYDRO1K database used in E-HYPE and gauged basins with incorrect routing were excluded from model calibration and validation. Furthermore, this shortcoming was noted for further improvements to the E-HYPE application.

In both applications, re-analysed, gridded, daily precipitation and temperature data were used. For S-HYPE, these data was extracted from the SMHI applied climate database for hydrological modelling, PTHBV. The database contains values for the period 1961–2008 that have been interpolated on a 4 × 4 km grid over the whole of Sweden. The interpolation method is a geo-statistical method developed by Johansson (2002). The E-HYPE forcing data are obtained from the gridded ERAMESAN database. This database has a spatial resolution of 0.1° (approx. 11 km) and covers most of Europe for the period 1980–2004. ERA40 data (1° spatial resolution) are used in areas of the modelled domain not covered by the ERAMESAN data set.

Land cover types were reduced to a number of land-use classes thought to have the largest effects on evapotranspiration and rainfall–runoff processes. Similarly, soils were categorised into a small number of different soil types based on soil texture, as this was thought to have the largest effect on soil water transport parameters (Marshall & Holmes, 1979). The soil and land-use types making up the classes modelled are listed in Table 2 for each model. This categorisation differs between model applications because of the different input data sources used. In particular, the E-HYPE application demonstrates the ability to construct, run and validate a multi-basin scale, high resolution regional hydrologic model using freely available global data (Strömqvist *et al.*, 2009).

Table 2 Soil and land-use types in the S-HYPE and E-HYPE categories.

	S-HYPE	E-HYPE
Soil types	Clay, sand, moraine, peat, thin soil layers, above tree line	Clay, loam, thin soil layers
Land-use types	Forest, agriculture, urban, lake, wetlands, glacier, above tree line, other vegetated land, other non-vegetated land	Forest, agriculture, urban, lake, bare rock

Calibration methodology

The goal of calibration was to obtain a uniform set of calibration parameters optimised to match observations at a large number of spatially distributed runoff stations within the model domain. A number of simple, yet effective criteria were used. For each station, a relative volume error (VE), a Nash-Sutcliffe Efficiency (NSE) and an All Flow Efficiency (AFE) were calculated (equations (1)–(3)). Because the NSE is based on the square of the anomalies, it biases optimisation to match flow peaks. The AFE is based on the absolute of the anomaly and therefore weights low and high flows equally.

$$\mathrm{VE} = \frac{\sum_{t=1}^{n} (q_t^{\mathrm{obs}} - q_t^{\mathrm{obs}})}{\sum_{t=1}^{n} q_t^{\mathrm{obs}}} \tag{1}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n} (q_t^{\mathrm{obs}} - q_t^{\mathrm{sim}})^2}{\sum_{t=1}^{n} (q_t^{\mathrm{obs}} - q^{\mathrm{mean}})^2} \tag{2}$$

$$\mathrm{AFE} = 1 - \frac{\sum_{t=1}^{n} \left| q_t^{\mathrm{obs}} - q_t^{\mathrm{sim}} \right|}{\sum_{t=1}^{n} \left| q_t^{\mathrm{obs}} - q^{\mathrm{mean}} \right|} \tag{3}$$

where q_t^{obs} and q_t^{sim} are the observed and simulated runoff, respectively, at time t, and n is the number of time-steps. Values of NSE and AFE may extend from minus infinity to one, values greater than zero suggest a model giving better results than the average of the recorded values, and a value of 1 indicates a perfect model.

During calibration, it was attempted to optimise the mean, and median of the VE for all calibration stations towards zero in order to ensure a realistic water balance in the model. A tolerance of ±1% was set for the mean VE. The main focus during calibration; however, was to minimise the standard deviation of the VE, ensuring as many sub-basins as possible had a small VE, and to maximise the mean and median of the NSE and AFE for all stations towards 1. This was done using mainly manual calibration; however, Monte-Carlo simulations were also done to investigate parameter sensitivity.

For the S-HYPE application, 36 gauging stations were used for model calibration, while a separate 115 were used for model validation. Calibration stations were chosen to represent gauged areas with dominant areas of each class, e.g. forest on moraine soils, open land on clay soils, lakes etc. Land-use and soil type specific parameters were therefore determined by optimising a group of stations containing dominant areas of the relative class. The model was then run using fixed calibration parameters to validate the results against the extra 115 validation stations. Finally, the model was run using all discharge stations. A number of these stations were regulated, so simple regulation routines were added to the model based on observed regulation patterns at the station.

This paper presents only a preliminary calibration of the E-HYPE application. The modelled region was divided into 6 multi-basin calibration regions (or separate multi-basin models), roughly based on European physiography, but constrained to major river basin boundaries. Calibration was not attempted in southern Europe due to lack of observed discharge data. The European calibration regions are listed in Table 4. Manual calibration was used to optimise the mean and median NSE and AFE while maintaining a VE within ±10%. It should be noted that this preliminary calibration is currently under improvement using both manual and automatic calibration methods, and hence represents the state of the model application at the time of writing. The two applications represent multi-basin modelling conducted at different resolutions and scales, and using different calibration methodologies, allowing the research questions to be explored using two different model applications.

RESULTS

The results of calibration and validation of the two applications are shown in Tables 3 and 4. The Swedish application was calibrated over one region and the European application in six different regions. This large-scale approach allows for rapid calibration for many river basins simultaneously. The Swedish application was then validated for 115 independent sub-basins. Finally, results are presented for all gauged sub-basins in the Swedish application, including regulated stations, where a simple regulation routine helps reproduce an averaged seasonal variation in regulation.

The standard deviation of the AFE and NSE results at all gauging stations in the region indicates the spread of the predictability over the region. The standard deviation of the VE indicates how well evapotranspiration is modelled over the region, with 0 % indicating an overall perfect water balance for a mean water balance of 0 %.

Table 3 Summary of calibration and validation results, S-HYPE application.

Set	No. stns	Mean			Median			Standard Deviation		
		AFE	NSE	VE (%)	AFE	NSE	VE (%)	AFE	NSE	VE (%)
Calibration	36	0.43	0.51	–0.35	0.45	0.55	–0.75	0.15	0.25	17.34
Validation	115	0.46	0.59	–1.23	0.47	0.63	–2.81	0.14	0.23	13.89
All(inc. Reg)	285	0.34	0.38	1.28	0.41	0.55	1.31	0.26	0.57	13.19

Table 4 Summary of calibration results, E-HYPE application.

	Calibration region		Mean:			Median:			Standard deviation:		
		No. stns	AFE	NSE	VE (%)	AFE	NSE	VE (%)	AFE	NSE	VE (%)
1	Nordic countries	19	0.37	0.53	1.17	0.40	0.62	–1.15	0.13	0.19	11.94
2	France & Italy	67	0.30	0.31	–1.07	0.35	0.48	–0.06	0.28	0.76	34.63
3	Central Europe	92	–0.12	–0.64	–11.61	–0.06	–0.12	–6.74	0.33	1.14	30.46
4	UK & Ireland	77	0.31	0.17	–0.22	0.38	0.46	–0.81	0.36	1.24	33.15
5	Northern Europe	13	0.13	0.26	–1.61	0.20	0.36	–5.92	0.32	0.46	18.62
6	East Europe & Russia	32	0.18	0.19	3.14	0.25	0.26	–3.65	0.36	0.47	56.36

DISCUSSION

For all calibration regions, excluding Central Europe, NSE and AFE exceeded 0, indicating a model efficiency greater than the mean of the observations. These values were highest for the S-HYPE application and the E-HYPE application region, which included Sweden and Finland. Peak flows (NSE) were better estimated than overall flows (AFE). The regions best modelled are relatively homogenous, dominated by lakes and forested land on moraine soils. Results were worst in Central Europe where it is suspected that a number of glaciers and lakes did not show up in the land cover data (observed runoff indicates the presence of lakes); however, the AFE indicates that the overall flows were represented better than the peak flows. Moreover, it is recognised that the precipitation data (from ERAMESAN) has less accuracy over mountainous areas (Jansson *et al.*, 2007). In general, however, the results suggest that a reasonable calibration can be achieved in a multi-basin model using a homogenous parameter set, and for the case of E-HYPE, using freely available input databases.

In order to compare a simultaneously calibrated multi-basin model with locally calibrated models, the S-HYPE results were compared with the Swedish operational flood forecasting

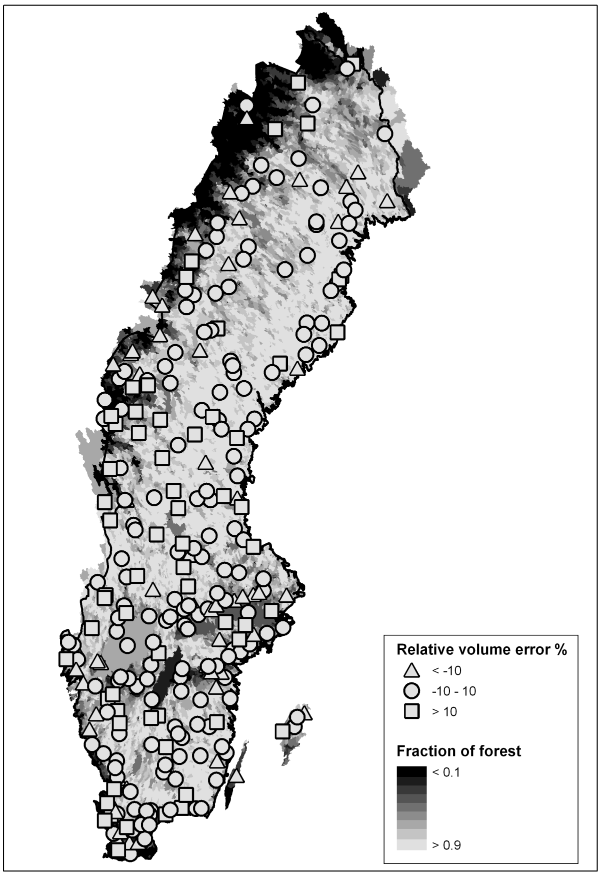

Fig. 2 Relative volume error (VE) and percentage of forest in sub-basin, S-HYPE application.

application of the HBV model. This model application is set up using the HBV model (Lindström *et al.*, 1997) calibrated in four districts. The median of the validation NSE values, from all districts, for the locally calibrated operation model was 0.71 and the standard deviation of the water balance was 11.4%. This is somewhat higher than the median validation NSE value of 0.63 and water balance standard deviation of 13.9% for S-HYPE; however, it may be concluded that the performance of the multi-basin model is near to comparable with the current operational model in terms of water balance and daily flow variation.

The spread of the water balance (i.e. the standard deviation) indicates either problems with forcing data (e.g. measurement of precipitation in mountains), processes not considered in evapotranspiration algorithms, or the removal or appearance of water from processes not considered in the rainfall–runoff model, e.g. water removed for irrigation or regional groundwater flow. An advantage with homogenous modelling over very large regions is that spatial visualisation of the volume error might help indicate processes missed. Figure 2 shows the spatial distribution of the water balance over Sweden. Note that there is a general trend indicating a modelled excess of water in the southwest and central eastern parts of the country. This corresponds somewhat with regions having low amounts of forest, and may therefore indicate that evapotranspiration should be more for the open land classes than for forested areas. There is also a tendency to underestimate volume in the northwest of Sweden. This is a mountainous region of Sweden, and may indicate underestimation of precipitation at high altitudes in the gridded input data set (i.e. PTHBV).

Water balance errors over Europe were, in general, within a standard deviation of 35%. Problems with water balance were particularly pronounced in mountainous regions with orographic lifting, a problem also acknowledged by Beldring *et al.* (2003) for a homogenously calibrated monthly rainfall–runoff model for all of Norway with VE ranging from 40 to 200%. For the European application, particularly towards the south, the larger resolution of the sub-basin forcing data input may have led to the larger errors in water balance.

A validation was performed for the S-HYPE application over 155 independent sub-basins. The validation of the S-HYPE application showed actually improved performance over the independent stations. This is good evidence that the calibrated parameters transferred well to other basins and demonstrates the usefulness of multi-basin modelling for predictions of streamflow in ungauged basins. Some validation results for E-HYPE are shown by Strömqvist *et al.* (2009).

The study presented here shows that it is possible to construct, calibrate and to some extent validate multi-basin models; however, there remains large scope to further develop input data and calibration methods for better validation results. For the E-HYPE application, there remains potential for calibration improvement, but also for input data improvement. Further research is needed to test higher resolution input data, in particular meteorological forcing and landcover data for lakes and glaciers. For multi-basin modelling in general, there is scope for testing automatic calibration using multi-objective criteria, improving process descriptions in arid regions in the HYPE model and for testing the ability to use auxiliary data such as recorded snow depths, lake levels, and remotely sensed soil moisture in calibration/verification.

CONCLUSIONS

The ability of a multi-basin hydrological model to predict flow in both gauged and ungauged basins was assessed. Two multi-basin hydrological models were set up for all of Sweden and most of Europe to calculate daily runoff at high resolution. It was shown that it is possible to obtain a good calibration using spatially homogenous parameter sets over very large regions, in the Swedish case, comparable to individually calibrated models. The Swedish application was also validated on a large number of independent sub-basins in order to evaluate the model's usefulness for predictions in ungauged basins. The mean and median of the performance criteria for the validation of the S-HYPE model were at least as good as those for the calibration of the model, indicating the model's ability to make reasonable predictions of streamflow in ungauged basins. The results indicate that although model performance may decrease somewhat as calibration scale increases, the predictive capacity of the model remains useful. Using a larger model domain and simultaneous calibration can thus be a method for hydrological predictions in ungauged basins in data sparse regions.

REFERENCES

Andréasson, J., Bergström, S., Carlsson, B., Graham, L. P. & Lindström, G. (2004) Hydrological change—climate change impact ssimulations for Sweden. *Ambio* **XXXIII**(4-5), 228–234.

Arheimer, B. & Brandt, M. (1998) Modelling nitrogen transport and retention in the catchments of southern Sweden. *Ambio* **27**(6), 471–480.

Arheimer, B., Andersson, L., Alkan-Olsson, J. & Jonsson, A. (2007) Using catchment models for establishment of measure plans according to the WFD. *Water Sci. Technol.* **56**(1), 21–28.

Arheimer, B., Lindström, G., Pers, C., Rosberg, J. & Strömqvist, J. (2008) Development and test of a new Swedish water quality model for small-scale and large-scale applications. In: *XXV Nordic Hydrological Conference* (Reykjavik, 11–13 August 2008), 483–492. Nordic Hydrology Publications Report no. 50, Sweden.

Beck, M. B. (1987) Water quality modeling: a review of the analysis of uncertainty. *Water Resour. Res.* **23**(8), 1393–1442.

Beldring, S., Engeland, K, Roald, L.A., Saelthun, N. R. & Voksoe, A. (2003) Estimation of parameters in a distributed precipitation–runoff model for Norway. *Hydrol. Earth System Sci.* **7**(3), 304–331.

Champeaux, J. L., Masson, V. & Chauvin, F. (2005) ECOCLIMAP: a global database of land surface parameters at 1 km resolution. *Met. Appl.* **12**, 29–32.

Council of the European Communities (2000) Directive 2000/60/EG of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Off. J. Eur. Commun. 22.12.2000, L327/1.*

Ducharne, A., Golaz, C., Leblois, E., Laval, K., Polcher, J. Ledoux, E. & de Marsily, G. (2003) Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM. *J. Hydrol.* **280**, 207–228.

Dunn, S. M. & Lilly, A. (2001) Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment-scale hydrological model. *J. Hydrol.* **252**, 157–173.

Feyen, L., Kalas, M. & Vrugt, J. A. (2008) Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization. *Hydrol. Sci. J.* **53**(2), 293–308.

GRDC (2008) Global Runoff Data Centre, World Meteorological Organisation & The German Federal Institute of Hydrology. http://www.bafg.de/cln_007/nn_301072/GRDC/Home/homepage_node.html?_nnn=true [date accessed: 16 Feb 2009].

Jansson, A., Persson, C. & Strandberg, G. (2007) 2D meso-scale re-analysis of precipitation, temperature and wind over Europe – ERAMESAN Time period 1980–2004. *SMHI Reports: Meteorology and climatology no. 112.*

Johansson, B. (2002) Estimation of areal precipitation for hydrological modelling in Sweden. PhD Thesis, Earth Sciences Centre, Dept Phys. Geog., Göteborg University. Sweden.

Lindström, G., Pers, C. Rosberg, J., Strömqvist, J. & Arheimer, B. (2009) Development and test of the HYPE (Hydrological Predictions in the Environment) model – A water quality model for different spatial scales. *Hydrol. Res.* (submitted).

Marachel, D. & Holman, I. P. (2005) Development and application of a soil classification-based conceptual catchment-scale hydrological model. *J. Hydrol.* **312**, 277–293.

Marshall, T. J. & Holmes, J. W. (1979) *Soil Physics*. Cambridge University Press, Cambridge, UK.

Miller, J. R., Russell, G. L. & Caliri, G. (1993) Continental-scale river flow in climate models. *J. Climate* **7**, 914–928.

SGU (Sveriges Geologiska Undersökning) Swedish Geological Survey's soil maps in digital form. Sveriges Geologiska Undersökning, Sweden.

Strömqvist, J., Dahné, J., Donnelly, C., Lindström, G., Rosberg, J., Pers, C., Yang, W. & Arheimer, B. (2009) Using recently developed global data sets for hydrological predictions. In: *New Approaches to Hydrological Prediction in Data-sparse Regions* (ed by K. K. Yilmaz *et al.*) (Joint IAHS & IAH Convention, Hyderabad, India, 6–12 September 2009). IAHS Publ. 333, this volume. IAHS Press, Wallingford, UK.

USGS (US Geological Survey) (2000) Hydro1k Elevation Derivative Database. http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html [date accessed: 16 Feb 2009].