European Water 49: 77-88, 2015.

Key factors for improving large-scale hydrological model performance

J. C. M. Andersson*, I. G. Pechlivanidis, D. Gustafsson, C. Donnelly and B. Arheimer
Swedish Meteorological and Hydrological Institute (SMHI), 601 76, Norrköping, Sweden
* *e-mail: jafet.andersson@smhi.se*

Abstract: This study focuses on identifying key factors that influence large-scale hydrological model performance. It draws on experiences from modelling in the Arctic, West Africa, and Europe with the HYPE model (HYdrological Prediction of the Environment). We use multiple evaluation criteria to analyse the influence of catchment delineation, climate input data, model parameterisation, and water management. For each factor, we compare the model performance of reference models and refined models using time-series of observed river discharge ranging from ten to a thousand stations (depending on application). The results show that all investigated factors influence model performance to varying extents. Accurately representing catchment size is critical. European stations with small deviation between modelled and published catchment size performed better than stations with large deviations, even after aligning the stations to the modelled topography (*NSE*: +10%, *aRE*: −12%). Refining the climate input data substantially increased model performance in a number of cases. In the Niger River basin, the simulation of both daily discharge dynamics and cumulative volumes improved significantly (*NSE*: +40%, *aRE*: −26% on average). In Europe, a refined precipitation data set resulted in a similar performance enhancement (*NSE*: +2%, *aRE*: −8%) as a refined temperature data set (*NSE*: +2%, *aRE*: −4%), on average. However, the temperature refinement was more consistent spatially. Linking lake parameters to spatially varying hydrological characteristics improved model performance across the Arctic domain (*NSE*: +11%, *aRE*: −8%). Refining infiltration capacities in the Niger basin improved both flow dynamics (*NSE*: +60%) and cumulative volumes (*aRE*: −40%) through modified flow paths and enhanced evaporation. Irrigation water management in the Arctic only affected model performance locally. Model performance was generally better in large and wet catchments with high runoff coefficients compared with relatively small, dry catchments with low runoff. These factors are also likely to affect model performance in other areas of the world.

Key words: Hydrological modelling, HYPE, model performance, large-scale modelling

1. INTRODUCTION

Hydrological modelling is important for a wide range of applications, including operational forecasting, water resources planning and management, flood risk assessment, infrastructure design; and for coupled systems modelling, such as water quality, hydro-ecology and climate change (Pechlivanidis et al., 2011; Singh and Frevert, 2006). Considerable effort has been given on improving process understanding at local scale using basin-scale hydrological models. However, recent needs target towards a better understanding at regional, national and global scales. Large-scale hydrological modelling has the potential to encompass many river basins, cross regional and international boundaries and represent a number of different geophysical and climatic zones (Alcamo et al., 2003; Gupta et al., 2014; Samaniego et al., 2011). This type of modelling takes advantage of regional and global databases as driving inputs, whereas increased computer processing speeds allow model setup and use at high spatiotemporal resolution for multiple river basins (Strömqvist et al., 2012; Widén-Nilsson et al., 2007).

The performance of large-scale hydrological models is subject to several sources of uncertainty, which may be caused by imperfect process understanding, model parameterisation or input data (Gudmundsson et al., 2012). For instance, large river basins are often strongly influenced by human activities (e.g. irrigation, reservoirs, and groundwater use) for which information is rarely available (Döll et al., 2009). In addition, global data sets can be inconsistent, erroneous, or only available at a coarse resolution (Hunger and Döll, 2008). Moreover, physical properties (e.g. vegetation and soil type) exhibit high spatial variability, generally not resolved in large-scale data sets, which introduces high uncertainty on model parameter estimates and significant differences in simulated system behaviour (Teuling et al., 2009). Model uncertainty increases due to these issues, particularly for ungauged basins (Hrachowitz et al., 2013; Sivapalan, 2003) and when predicting consequences of hydrological and societal change (Montanari et al., 2013).

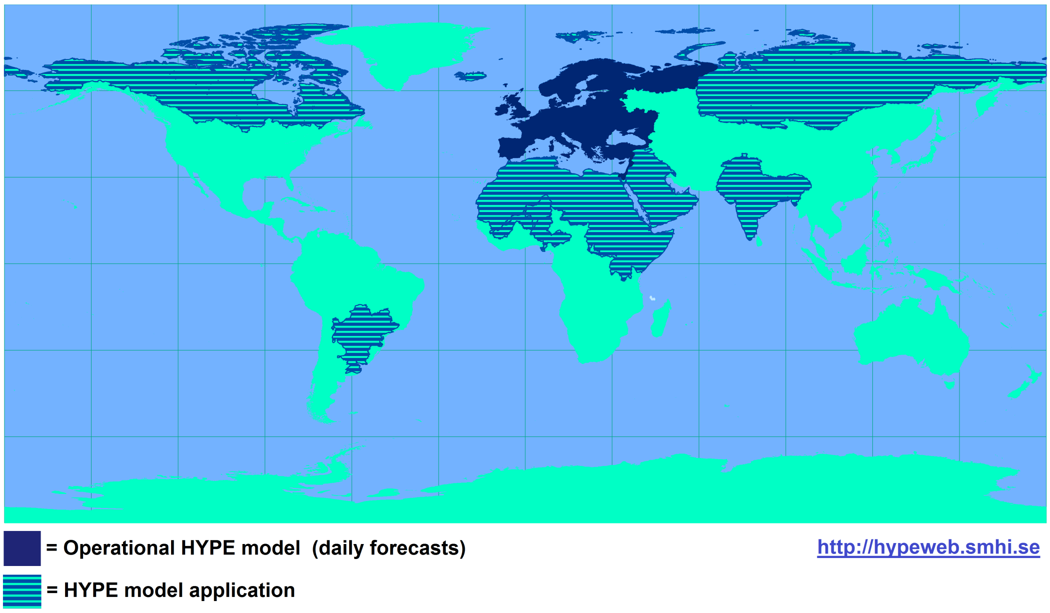

Figure 1. Map of HYPE model applications at SMHI.

The Swedish Meteorological and Hydrological Institute (SMHI) has developed a set of large-scale hydrological models that use spatially distributed meteorological variables, topography, soil, and vegetation as inputs to predict spatially varying basin behaviour (Arheimer et al., 2012). The overall practical aim is to provide discharge, and often water quality, statistics and time-series to water authorities for: (i) flood forecasts, (ii) characterisation of water body status (e.g. water balance and dynamics), (iii) compliance assessment of environmental goals and directives (e.g. for the EU Water Framework Directive, using an integrated approach as suggested by Tsakiris and Alexakis (2012)), (iv) planning of remedial measures, and (v) projecting changes in water fluxes with climate change. Applications of large-scale hydrological modelling at SMHI are widely spread around the world, including Sweden, the Baltic Sea, Europe, Arctic, the Niger River, India, Middle East and North Africa, and the La Plata basin (Figure 1).

In this article, we identify and quantify key factors that improve performance of large-scale hydrological models drawing from our experience in Europe, Arctic and the Niger River. Among the several factors controlling model performance, we focus on: (1) catchment delineation and linking of contributory areas, (2) meteorological input data (i.e. precipitation and temperature), (3) model parameterisation, and (4) water management (i.e. irrigation).

2. MODELS AND METHODOLOGY

We simulate hydrological processes with the HYPE model (Lindström et al., 2010); a semi-distributed, process-oriented, rainfall-runoff model (Figure 2). We quantify the improvement in the model performance by implementing a number of model refinements. The analysis is based on comparing reference and refined models.

2.1 Reference models

The HYPE model was set up for all river basins in Europe (Donnelly et al., 2015), all river basins draining into the Arctic Ocean, and for the Niger River basin using available large-scale data sets on topography, land use, soil, precipitation, temperature, lakes, reservoirs, crop types, irrigation, evaporation, snow and discharge (Table 1).

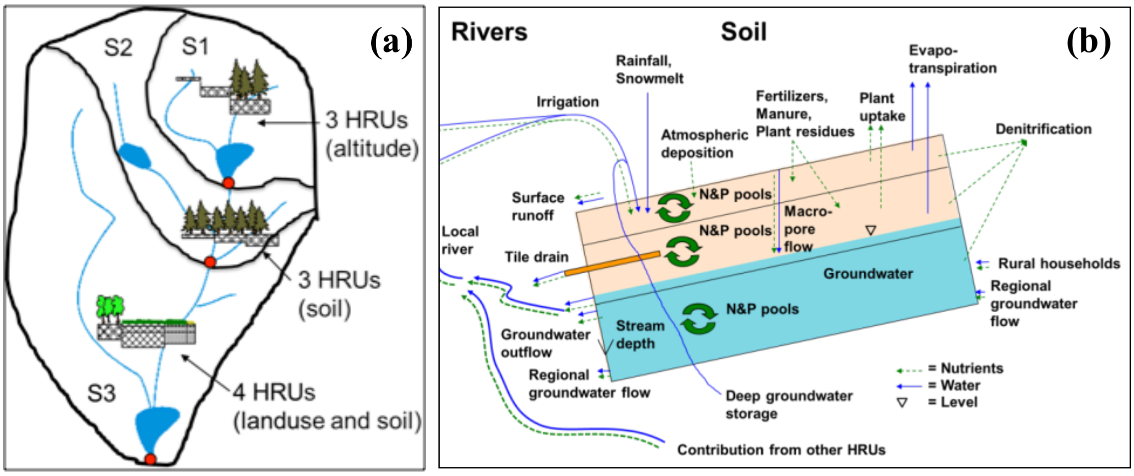

Figure 2. Conceptual overview of the HYPE model. (a) The model domain is divided into several nested spatial units: each domain contains one or more river basins, each river basin contains a set of sub-basins, each sub-basin contains a number of hydrological response units (HRUs, unique combinations of soil and landuse). (b) In each HRU, a number of key hydrological processes are simulated. Runoff from the soil and land surface is aggregated to sub-basin scale and subsequently routed through the river network and through lakes and reservoirs. See Lindström et al. (2010) for the full model description.

Table 1. Data sources and characteristics of the reference model setups.

	Europe	Arctic	Niger River
Acronym in this paper	E-HYPE	ARCTIC	NIGER
Total area (km^2)	9.6 million	23.3 million	2.1 million
No. of sub-basins	35000	30770	803
No. of discharge stations	1007	1349	71
Topography	HydroSHEDS[a], Hydro 1K[a]	Hydro 1K[a]	HydroSHEDS[a]
Land Use	CORINE[a], GLC2000[a]	GLC2000[a]	GLC2000[a]
Soil	ESD[a], DSMW[a]	HWSD[b]	HWSD[b], WISE[c]
Discharge observations	GRDC[a], EWA[a], BHDC[a]	GRDC[a], USGS[d], R-ArcticNET[e]	GRDC[a], ABN[f]
Precipitation	ERA-INTERIM[a]	WFDEI[g]	WFD[h]
Temperature	ERA-INTERIM[a]	WFDEI[g]	WFD[h]
Lakes and reservoirs	GLWD[a]	GLWD[a]	GLWD[a], GranD[i]
Irrigation	EIM[a], GMIA[a]	-	-
Crop types	MIRCA2000[j], CAPRI[a]	-	-

a:(see Donnelly et al., 2015), b:(FAO et al., 2009), c: (Batjes, 2012), d:(USGS, 2011), e:(R-ArcticNET, 2011), f:(ABN, 2008), g:(Weedon et al., 2012), h:(Weedon et al., 2011), i:(Lehner et al., 2011), j: (Portmann et al., 2010)

2.2 Model refinements

A number of model refinements were implemented to test their impact on model performance (Table 2). The *first* refinement concerns accurately representing the catchment size of the discharge stations, which we here illustrate in Europe with the “E-HYPE-Adiff” experiment. The locations of all the discharge stations were adjusted to match the underlying topography and drainage accumulation data based on published and computed upstream areas, respectively. We used an automated method that moved the stations to the location with the smallest difference in upstream area, within a pre-defined search radius. We then calculated the residual geometric relative error between estimated and published catchment area (sensu Donnelly et al., 2012) and compared the model performance at stations where the method reduced the residual error to within acceptable limits (<±10%) against stations where the residual error remained large (>±10%, Figure 3).

Table 2. Characteristics of each reference model and each refined model, respectively.

Acronym	Reference model	Refined model
E-HYPE-Adiff	Stations with large difference in catchment area (>±10%)	Stations with small difference in catchment area (<±10%)
E-HYPE-Pcp	ERA-INTERIM precipitation	Scaled ERA-INTERIM precipitation
E-HYPE-Temp	ERA-INTERIM temperature	Adjusted ERA-INTERIM temperature
NIGER-Clim	WFD climate input	WFDEI climate input
NIGER-Flowpath	High surface runoff parameterization	Low surface runoff parameterization
ARCTIC-Lake*	Generic rating curve for all lakes	Specific rating curve for each lake
ARCTIC-Irr†	No irrigation	Irrigation abstraction included, based on GMIA and MIRCA2000

**: for all river mouths draining to the Arctic Ocean †: for the Ob River*

The *second* refinement focuses on meteorological input data. In Europe, the GPCC data set (Schneider et al., 2011) was used to bias adjust the daily ERA-INTERIM precipitation by applying a monthly scaling factor to each day in the month (E-HYPE-Pcp, see Donnelly et al., 2013). In addition, we adjusted the mean daily ERA-INTERIM air temperatures based on the elevation differences between grid cells and sub-basins using a scaling factor of 0.6°C/100m (E-HYPE-Temp). The model performance with and without each ERA-INTERIM adjustment was then compared. In West Africa, we replaced the precipitation and temperature from WFD (based on ERA-40) with that of WFDEI (based on ERA-INTERIM), and compared the model performance with each data set (NIGER-Clim).

The *third* category of refinements centres on model parameterisation. In the Niger basin, we examined the effect of adjusting the *mactrinf* model parameter controlling soil infiltration capacity and surface runoff. We compared a parameterisation of low infiltration capacity and high surface runoff (*mactrinf*=10) with a parameterisation of high infiltration capacity and low surface runoff (*mactrinf*=75). This parameter change was applied equally on all soils across the model domain (NIGER-Flowpath). Another approach is to distribute parameter values based on varying physical properties. We illustrate this in the Arctic domain, where all major lakes in the reference model had identical rating curves versus the refined model where we calculated specific rating curves for each lake using data on individual catchment and lake area (ARCTIC-Lake).

Finally, the *fourth* refinement exemplifies the impact of accounting for water management on model performance. In the Ob River basin (3 million km^2), we compared model performance with and without irrigation abstractions (ARCTIC-Irr). Irrigation practices were dynamically simulated based on data specifying the areal extent of irrigation, crop type, cropping season, temporal crop water demand profiles, sources of irrigation water, and irrigation efficiency. The required information was extracted from the GMIA and MIRCA2000 data sets.

2.3 Evaluation criteria

We used two performance measures to compare the reference and refined models: (1) the Nash-Sutcliffe Efficiency (*NSE*) and (2) the absolute value of the cumulative relative error (*aRE*, [%]):

$$NSE = 1 - \frac{\sum_{i=1}^{m}(c-r)^2}{\sum_{i=1}^{m}(r-\bar{r})^2} \qquad aRE = \left| \left(\frac{\sum_{i=1}^{m}(c-r)}{\sum_{i=1}^{m} r} \right) \times 100 \right| \tag{1}$$

where c is simulated value, r is observed value, $\bar{r}$ is the arithmetic mean of the observations, m is number of observations in the time-series. The *NSE* ranges between 1 (perfect fit) and $-\infty$, whereas *aRE* ranges between 0 (perfect fit) and ∞. Hence, a model improvement is signified by an increase in *NSE* or a reduction in *aRE*. The criteria were calculated on daily resolution for all comparisons except for ARCTIC, where a monthly resolution was used instead due to data availability. The model performance was judged against a set of discharge stations reporting observed river flow. Only stations with positive *NSE* (>0) in either the reference model or the refined model were included in each comparison in order to focus on locations where the model performance was acceptable. We used *NSE* to evaluate the model's capacity to capture daily dynamics and *aRE* to evaluate the long-term water volume, respectively.

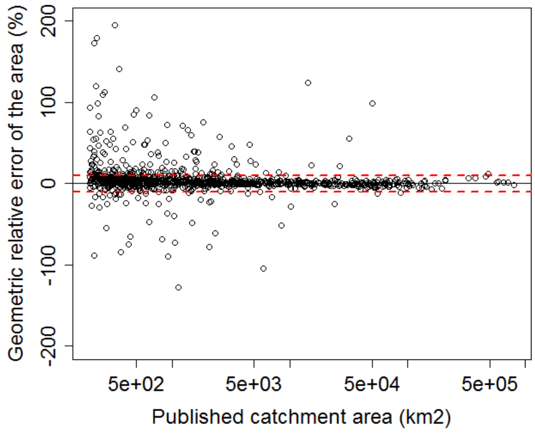

Figure 3. Geometric relative error vs. published catchment area for the E-HYPE-Adiff evaluation stations. The red dashed line indicates the selection of stations with acceptable residual errors (<±10%).

3. RESULTS AND DISCUSSION

Figure 4 summarises the performance of the reference and refined models, respectively, for all discharge stations analyzed in each model refinement. Furthermore, in Figure 5 we present example hydrographs for selected stations to illustrate the impact of various refinements on flow dynamics.

3.1 Impact of catchment delineation and linking of contributory areas

In E-HYPE-Adiff, we found that the automated method to align the topographic and observed discharge data sets improved model performance (Figure 4). The median *NSE* increased by 9.7%, whereas the median *aRE* decreased by 12%. The *aRE* distribution also narrowed, indicating an improvement across the model domain. Hence, accurately representing the catchment size of the discharge stations has a large impact on model performance. This involves correctly delineating the sub-basin boundaries, and correctly describing sub-basin connectivity. The topographic information is fundamental in this pursuit. However, data resolution and coordinate system differences often result in substantial catchment area differences, which require careful attention in model setup. If this type of alignment would be ignored in the model setup, the model errors would become enormous for HYPE and all other hydrological models relying on the principle that runoff is generated from precipitation falling on a given catchment area.

3.2 Impact of climatic information

The primary dynamic input data for the HYPE model is daily precipitation and temperature. Consequently, one would expect the model performance to be sensitive to inaccuracies in these data sets. E-HYPE-Pcp, E-HYPE-Temp, and NIGER-Clim all verify this hypothesis (Figure 4). The GPCC precipitation correction of E-HYPE-Pcp decreased *aRE* by 8% on average, and increased *NSE* by 2%. The worst performing stations in the reference model improved the most, as shown by the reduction in the 75th percentile of the *aRE* distribution. Figure 5a illustrates the greater capacity of the refined model to capture the observed monthly flow dynamics at Le Mas d'Agenais station. However, the impact of the E-HYPE-Pcp refinement was very variable: *aRE* increased more than 100% for some stations, and the inter-quartile range of the change in *aRE* was large (94 percentage points). A possible reason could be the method's uniform scaling approach applying a monthly scaling factor across all daily precipitation intensities, which may have exaggerated precipitation extremes. Alternatively, the accuracy of GPCC and ERA-INTERIM varies substantially depending on location (e.g. being worse in areas with low station density). This suggests that a more refined approach is needed to achieve spatiotemporally consistent results (e.g. based on the methods introduced by Andersson *et al.*, 2012 and Yang *et al.*, 2010).

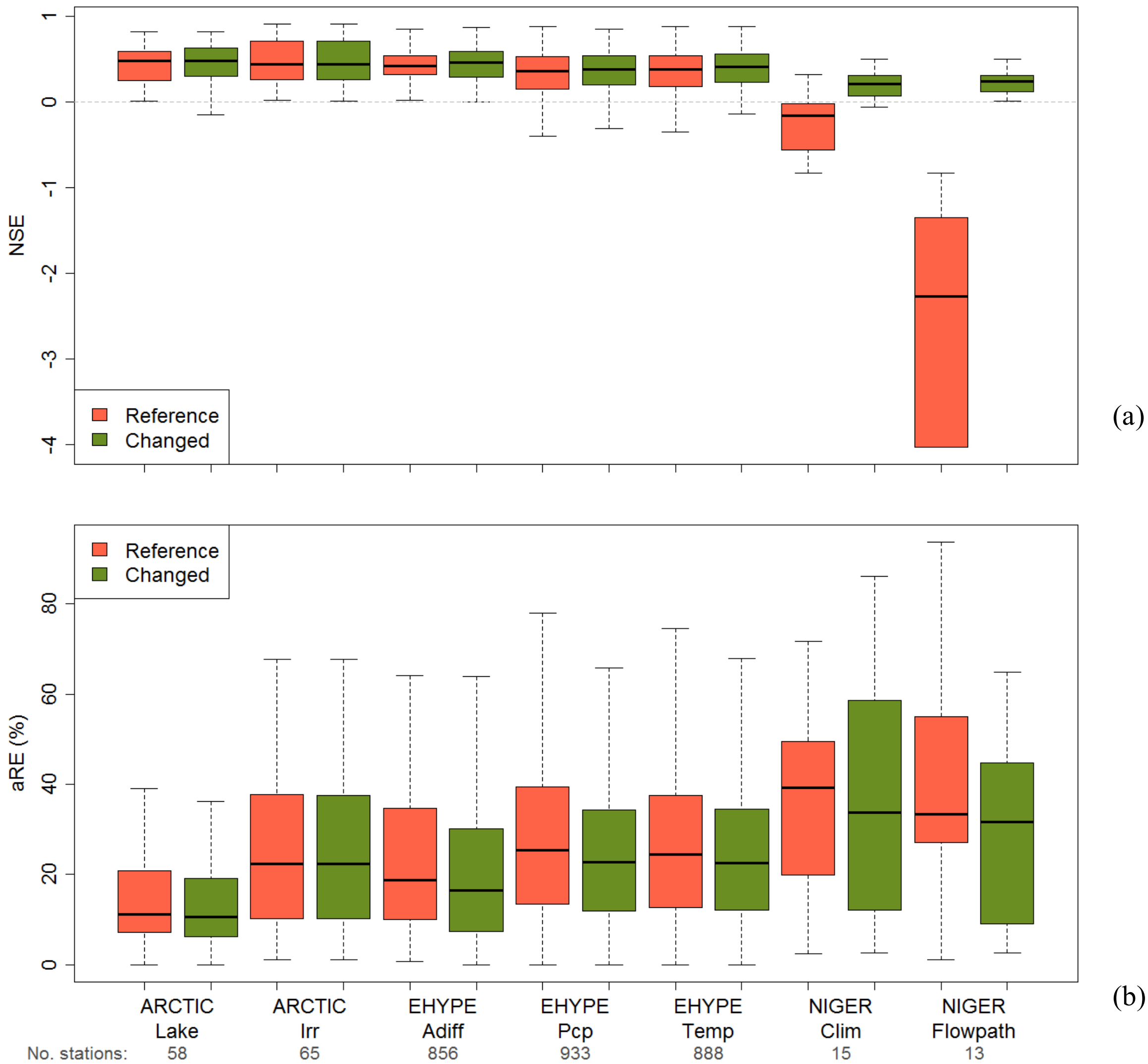

Figure 4. Distribution of (a) NSE and (b) aRE values for the reference and the refined models respectively (Table 2). Each boxplot summarizes the distribution for all discharge stations with positive NSE in either the reference model or the refined model. The number of stations used in each analysis is given in grey at the bottom of the figure.

E-HYPE-Temp led to more consistent improvements of model performance (inter-quartile range of the change in *aRE*: 22 percentage points). The influence of the simple elevation correction of temperature was surprisingly large (median improvement in *NSE*: 2% and *aRE*: 4%, i.e. similar to E-HYPE-Pcp). Temperature controls evaporation, snowfall and snowmelt in HYPE. Hence, E-HYPE-Temp shows the importance of accurately representing these hydrological processes in different parts of Europe. The importance of accurate climate information has also recently been assessed and confirmed for five large-scale models across Europe (E-HYPE, LISFLOOD, LPJmL, VIC and WBM; Greuell et al., 2015).

In the Niger River basin, the NIGER-Clim refinement led to significant improvements in model performance, particularly for *NSE* (median improvement: 40%), but also for *aRE* (median improvement: 26%). Using the WFD data to drive the model, 11 out of 15 stations had negative *NSE*; whereas with WFDEI data, 13 out of 15 stations had positive *NSE*. The *NSE* distribution also became substantially narrower, indicating a more consistent model performance (Figure 4a). Although water volumes generally improved, *aRE* remained quite large for some stations. This suggests that other factors are also critical to address in order to obtain better performance across the model domain.

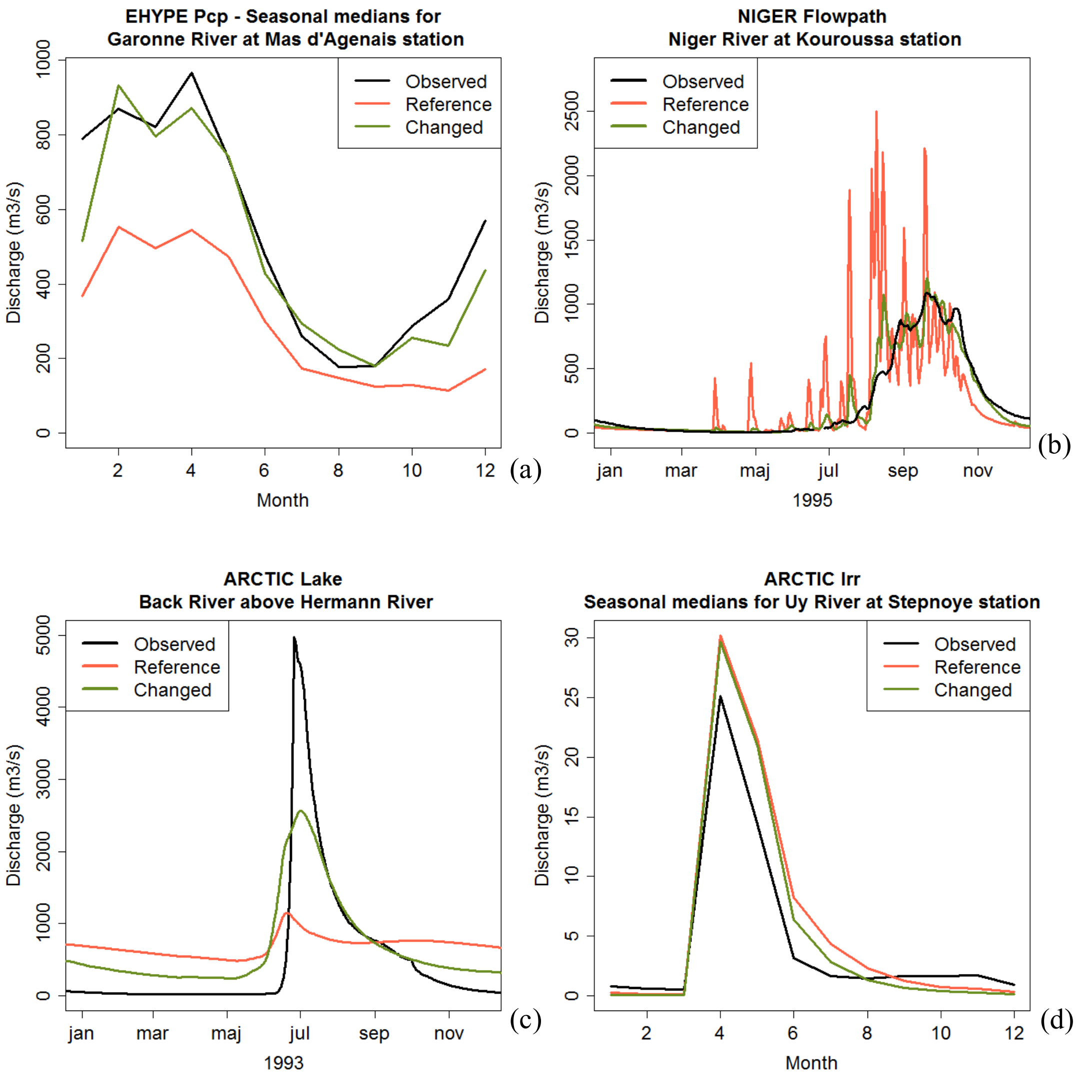

Figure 5. Example hydrographs to illustrate specific impacts of selected model refinements. (a) Median monthly discharge at Le Mas d'Agenais station illustrating E-HYPE-Pcp. (b) Daily discharge for year 1995 at Kouroussa station illustrating NIGER-Flowpath. (c) Daily discharge for year 1993 at Herman River station illustrating ARCTIC-Lake. (d) Median monthly discharge at Stepnoye station illustrating ARCTIC-Irr.

3.3 Impact of model parameterisation

The flow of information through the model is conducted by equations and parameters defining process dynamics. HYPE model parameters regulate, for example, hydrological flow path: whether water flows slowly to the streams through sub-surface flow, or more rapidly as surface runoff. The

NIGER-Flowpath experiment shows that the flow path parameterisation has a large impact on model performance (median improvement in *NSE*: 60% and *aRE*: 40%). The large percentage change is partly due to the poor reference model performance, but moving from only negative *NSE* to only positive *NSE* and substantially narrowing the *NSE* distribution still indicates a significant performance improvement (Figure 4a). Figure 5b exemplifies the importance of the flow path parameterisation for simulating discharge in the Niger River basin. The large improvement in *aRE* is at first glance surprising since no changes were made that directly affect the water inputs and outputs (precipitation and evaporation). However, the improved *aRE* is the result of a second-order effect where the increase in infiltration capacity subjected more water to the strong evaporative demand in the region, which increased evaporation by, on average, 9% and reduced discharge volumes accordingly. This effect would probably be more subdued if HYPE had simulated evaporation from surface runoff.

Another way to improve model performance is to vary parameter values according to the varying physical conditions of the catchments instead of applying constant parameter values across the model domain. In large-scale modelling, such spatial parameter distribution needs to be simple enough to be applicable in ungauged basins and for thousands of sub-basins where individual fine-tuning is impracticable. In the ARCTIC-Lake experiment, we applied a simple method based on widely available data, which improved model performance (mean *NSE* improvement: 11%, mean *aRE* improvement: 8%, and narrower *NSE* and *aRE* distributions in Figure 4). This model refinement was particularly helpful in areas with a high lake density (Figure 5c).

3.4 Impact of water management

Water management and regulation is another factor that can affect river flows substantially. This includes impoundments (e.g. dams), transfers (e.g. inter-basin canals) and abstractions (e.g. pumping from rivers, lakes and groundwater). In large-scale modelling, most rivers are affected by regulation in some form; however, the impact is highly dependent on location. Irrigation abstractions in the Ob River illustrate this in the Arctic domain (ARCTIC-Irr). On the river basin scale, irrigation did not significantly affect the model performance (average improvement in *NSE*: 0.5% and *aRE*: 1%). This is probably because irrigation only occupies 0.2% of the river basin area. However, locally, irrigation withdrawals from lakes and rivers improved model performance noticeably (reducing *aRE* by >10%; Figure 5d). In river basins with more extensive irrigation, the impact is more substantial (e.g. Donnelly et al., 2015; Yilmaz and Harmancioglu, 2012).

3.5 Hydro-climatic catchment characteristics and model performance

We finally investigated the relationship between model performance and long-term hydro-climatic catchment characteristics to better understand the model behaviour. We calculated the runoff coefficient (RC, the ratio of mean annual runoff to mean annual precipitation) the wetness index (WI, the ratio of mean annual precipitation to mean annual potential evaporation), the energy availability index (Energy, the inverse of WI), the moisture availability index (Moisture, the ratio of mean annual actual evaporation to mean annual precipitation), and the upstream drainage area (A) for each gauging station to assess their varying impact on model performance (Figure 6).

Overall, *aRE* performance seems to increase from dry to wet regions, with more impervious area, and with larger drainage area (Figure 6b). In contrast, relatively small, dry catchments with low runoff coefficients (RC < 0.4) in the Niger basin displayed the largest water balance errors. These stations also had low *NSE* performance, although *NSE* and *aRE* were not significantly correlated overall. The *aRE* performance is generally controlled by parameters that are sensitive to the volume rather than the shape of the hydrograph, which impacts the long-term water balance (e.g. the *cevp* parameter controlling potential evaporation). Moreover, the *aRE* tends to increase the further away the station is from the idealized Budyko curve (Gerrits et al., 2009), particularly for dry catchments

(Energy > 1, Figure 6d). This indicates that accurately partitioning between evaporation and runoff is critical for improving model performance, also beyond the stations assessed in the NIGER-Flowpath experiment.

No clear relationship between *NSE* and the catchment characteristics was observed. A wide range of *NSE* values were obtained for all model domains, in small and large basins, in wet and dry basins, for a range of runoff coefficients, and in areas with both energy and moisture limitations (Figure 6a&c). However, the highest concentration of relatively high *NSE* values were obtained in the middle of the spectrum (i.e. for stations with RC ≈ 0.6 and WI ≈ 2). Additional analysis against remotely sensed snow cover data (http://www.globsnow.info/) showed a discrepancy in the rate of snow-melt in the Arctic domain. This led to a mismatch in the simulated temporal dynamics of the system, and is a probable reason for the low *NSE* observed in some areas. Inadequate parameter identification could also be a reason for low *NSE*, since the shape of the simulated hydrograph is sensitive to several HYPE model parameters. In addition, inadequate representation of river regulations, affecting daily dynamics, could cause low *NSE* across the spectrum of catchment characteristics.

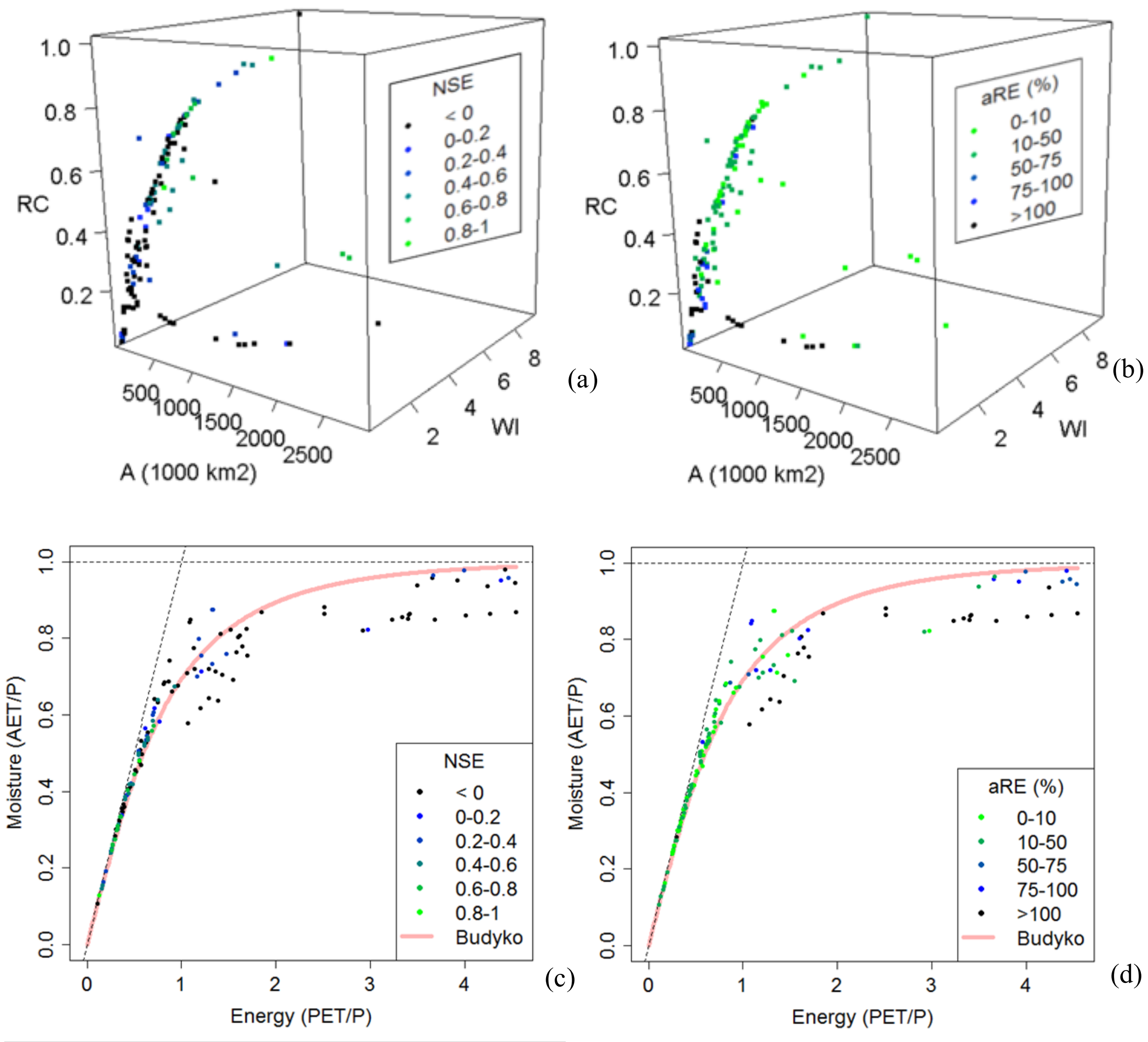

Figure 6. Scatter plots demonstrating model performance as a function of (a,b) the runoff coefficient (RC), the wetness index (WI), and the drainage area (A); and (c,d) the Energy availability index, the Moisture availability index, and the Budyko curve, evaluated with (a,c) NSE, and (b,d) aRE.

3.6 Limitations of this study and future work

It is important to note the various assumptions and limitations in these modelling experiments. The reliability of the input data used to drive large-scale multi-basin hydrological models, particularly those derived from global data sets, has been questionable (Kauffeldt et al., 2013). In most experiments, a comparison of the applied climatic and physiographic data against national data was conducted if available. For instance, our analysis showed some inconsistencies of soil types in Europe. Hence, our results are subject to the quality of the selected input data. However, this study also illustrates that significant gains can made by making simple improvements to input data. Future research should involve additional sensitivity analysis to make further improvements of the different input data sets and models relying on them. This is critical for improving risk assessments in integrated water resources management (Walczykiewicz et al., 2012), since models are typically used to quantify management alternatives.

Given that the experiments were tested simultaneously on many stations covering a large range of physiographic conditions in each model application, there is already some evidence of generality of the conclusions. Future research could assess this further by testing all experiments in all regions.

Performances from each of the three model setups were grouped and analysed; however the predictability and consistency of each model (here described by the model's performance) varies in every application. All models can adequately represent the long-term average fluxes and seasonal variation in their domain; however poor performances have been identified at various regions. It is expected that the models' performance can be increased from a better spatial resolution of input data and a finer spatial representation of the existing sub-basins (Lobligeois et al., 2014). Model consistency can also be improved by introducing new discharge observations both in space (representing different hydro-climatic systems) and time (increasing data length and temporal resolution), while additional variables (e.g. evapotranspiration, snow covered area) could provide information to enhance our process understanding (Döll et al., 2008). Finally, our study is limited by the assumed stationarity in the hydrological systems during the investigated periods. It is recognised that non-stationarity exists as a characteristic of the natural world due to various environmental changes (land use and man-made alterations) (Wagner et al., 2013). Future research should aim to model the system responses in a dynamic way addressing changes due to individual factors, i.e. climate, land use, man-made, and population as they evolve (Ehret et al., 2014).

4. CONCLUSIONS

Several factors affect HYPE model performance in large-scale applications. Key factors to refine in order to improve performance include the catchment delineation, meteorological input data, and model parameterisation. The impact of the simple temperature refinement was surprisingly large, reflecting its linkage to several hydrological processes in the model. Irrigation water management was important locally, but not for the entire Ob River basin. The impact would likely be more significant in other river basins dominated by irrigated agriculture. The model setups included in this study cover a substantial variation in climate and physiographic characteristics. Hence, our conclusions are likely applicable for improving large-scale hydrological models in other areas of the world exhibiting similar variability in catchments characteristics.

ACKNOWLEDGEMENTS

We are very grateful for the funding of this cross-cutting research by EU FP7 (through the ECLISE, SIRIUS and IMPACT2C projects), Formas (through the ADSIMNOR and HYDROIMPACTS2.0 projects), and Sida/VR (through the Niger-HYPE project). The study was performed at the SMHI Hydrological Research unit, where much work is done jointly. Hence, we would like to acknowledge contributions from Joel Dahné, Kristina Isberg and Jörgen Rosberg for assistance with background material for this study. We also thank Thorsten Wagener for valuable suggestions on an earlier version of this article.

REFERENCES

ABN, 2008. Niger-HYCOS Hydrological Information System [WWW Document]. URL http://nigerhycos.abn.ne/ (accessed 3.1.12).

Alcamo, J., Doll, P., Henrichs, T., Kaspar, F., Lehner, B., Rosch, T., Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability. Hydrol. Sci. J.-J. Sci. Hydrol. 48, 317–337.

Andersson, J.C.M., Zehnder, A.J.B., Wehrli, B., Yang, H., 2012. Improved SWAT Model Performance With Time-Dynamic Voronoi Tessellation of Climatic Input Data in Southern Africa. J. Am. Water Resour. Assoc. 1–14. doi:10.1111/j.1752-1688.2011.00627.x

Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., Strömqvist, J., 2012. Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale. Hydrol. Res. 43, 315–329. doi:10.2166/nh.2012.010

Batjes, N.H., 2012. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2) (No. 2012/01). ISRIC - World Soil Information, Wageningen, The Netherlands.

Döll, P., Berkhoff, K., Bormann, H., Fohrer, N., Gerten, D., Hagemann, S., Krol, M., 2008. Advances and visions in large-scale hydrological modelling: findings from the 11th Workshop on Large-Scale Hydrological Modelling. Adv Geosci 18, 51–61. doi:10.5194/adgeo-18-51-2008

Döll, P., Fiedler, K., Zhang, J., 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrol Earth Syst Sci 13, 2413–2432. doi:10.5194/hess-13-2413-2009

Donnelly, C., Andersson, J.C.M., Arheimer, B., 2015. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrol. Sci. J. doi:10.1080/02626667.2015.1027710

Donnelly, C., Arheimer, B., Capell, R., Dahné, J., Strömqvist, J., 2013. Regional overview of nutrient load in Europe – challenges when using a large-scale model approach, E-HYPE, in: Proceedings of the IAHS-IAPSO-IASPEI Assembly, IAHS Publ. Presented at the IAHS-IAPSO-IASPEI Assembly, IAHS Press, Gothenburg, Sweden.

Donnelly, C., Rosberg, J., Isberg, K., 2012. A validation of river routing networks for catchment modelling from small to large scales. Hydrol. Res. 44, 917–925. doi:10.2166/nh.2012.341

Ehret, U., Gupta, H.V., Sivapalan, M., Weijs, S.V., Schymanski, S.J., Blöschl, G., Gelfan, A.N., Harman, C., Kleidon, A., Bogaard, T.A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M.F.P., Di Baldassarre, G., Parajka, J., van Beek, L.P.H., van Griensven, A., Westhoff, M.C., Winsemius, H.C., 2014. Advancing catchment hydrology to deal with predictions under change. Hydrol Earth Syst Sci 18, 649–671. doi:10.5194/hess-18-649-2014

FAO, IIASA, ISRIC, ISS-CAS, JRC, 2009. Harmonized World Soil Database (version 1.1). FAO and IIASA, Rome, Italy.

Gerrits, A.M.J., Savenije, H.H.G., Veling, E.J.M., Pfister, L., 2009. Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model. Water Resour. Res. 45, W04403. doi:10.1029/2008WR007308

Greuell, W., Andersson, J.C.M., Donnelly, C., Feyen, L., Gerten, D., Ludwig, F., Pisacane, G., Roudier, P., Schaphoff, S., 2015. Evaluation of five hydrological models across Europe and their suitability for making projections under climate change. Hydrol. Earth Syst. Sci. Discuss. submitted.

Gudmundsson, L., Wagener, T., Tallaksen, L.M., Engeland, K., 2012. Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. Water Resour. Res. 48, W11504. doi:10.1029/2011WR010911

Gupta, H.V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., Andréassian, V., 2014. Large-sample hydrology: a need to balance depth with breadth. Hydrol Earth Syst Sci 18, 463–477. doi:10.5194/hess-18-463-2014

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58, 1198–1255. doi:10.1080/02626667.2013.803183

Hunger, M., Döll, P., 2008. Value of river discharge data for global-scale hydrological modeling. Hydrol Earth Syst Sci 12, 841–861. doi:10.5194/hess-12-841-2008

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., Westerberg, I.K., 2013. Disinformative data in large-scale hydrological modelling. Hydrol Earth Syst Sci 17, 2845–2857. doi:10.5194/hess-17-2845-2013

Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front. Ecol. Environ. 9, 494–502. doi:10.1890/100125

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrol. Res. 41, 295–319. doi:10.2166/nh.2010.007

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. Hydrol Earth Syst Sci 18, 575–594. doi:10.5194/hess-18-575-2014

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D.A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., Belyaev, V., 2013. "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. Hydrol. Sci. J. 58, 1256–1275. doi:10.1080/02626667.2013.809088

Pechlivanidis, I.G., Jackson, B.M., McIntyre, N.R., Wheather, H.S., 2011. Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Glob. NEST J. 13, 193–214.

Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Glob. Biogeochem. Cycles 24, GB1011. doi:10.1029/2008GB003435

R-ArcticNET, 2011. R-ArcticNET Pan-Arctic River Discharge Data [WWW Document]. URL http://www.r-arcticnet.sr.unh.edu/v4.0/AllData/index.html (accessed 12.7.11).

Samaniego, L., Kumar, R., Jackisch, C., 2011. Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data. Hydrol. Res. 42, 338. doi:10.2166/nh.2011.156

Schneider, U., Becker, A., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2011. Global Precipitation Analysis Products of the Global Precipitation Climatology Centre (GPCC) [WWW Document]. URL http://gpcc.dwd.de (accessed 9.1.12).

Singh, V.P., Frevert, D.K., 2006. Watershed Models. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA.

Sivapalan, M., 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrol. Process. 17, 3163–3170. doi:10.1002/hyp.5155

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. Hydrol. Sci. J. 57, 229–247. doi:10.1080/02626667.2011.637497

Teuling, A.J., Uijlenhoet, R., van den Hurk, B., Seneviratne, S.I., 2009. Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. J. Hydrometeorol. 10, 751–765. doi:10.1175/2008JHM1033.1

Tsakiris, G., Alexakis, D., 2012. Water quality models: an overview. Eur. Water 37, 33–46.

USGS, 2011. USGS National Water Information System [WWW Document]. URL http://waterdata.usgs.gov/nwis (accessed 11.2.11).

Wagner, P.D., Kumar, S., Schneider, K., 2013. An assessment of land use change impacts on the water resources of the Mula and Mutha Rivers catchment upstream of Pune, India. Hydrol Earth Syst Sci 17, 2233–2246. doi:10.5194/hess-17-2233-2013

Walczykiewicz, T., Biedroń, I., Kwiecień, M., 2012. How to quantify risk in integrated water resources management? Eur. Water 38, 33–42.

Weedon, G.P., Gomes, S., Balsamo, G., Best, M.J., Bellouin, N., Viterbo, P., 2012. WATCH forcing data based on ERA-INTERIM [WWW Document]. URL ftp://rfdata:forceDATA@ftp.iiasa.ac.at (accessed 12.15.12).

Weedon, G.P., Gomes, S., Viterbo, P., Shuttleworth, W.J., Blyth, E., Österle, H., Adam, J.C., Bellouin, N., Boucher, O., Best, M., 2011. Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. J. Hydrometeorol. 12, 823–848. doi:10.1175/2011JHM1369.1

Widén-Nilsson, E., Halldin, S., Xu, C., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. J. Hydrol. 340, 105–118. doi:10.1016/j.jhydrol.2007.04.002

Yang, W., Andréasson, J., Phil Graham, L., Olsson, J., Rosberg, J., Wetterhall, F., 2010. Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. Hydrol. Res. 41, 211. doi:10.2166/nh.2010.004

Yilmaz, B., Harmancioglu, N., 2012. Comparative assessment of water management indicators in the Gediz river basin with foreseen climate scenarios. Eur. Water 37, 27–32.