A parsimonious modelling approach for water management in dryland areas

SAKET PANDE[1], HUBERT H. G. SAVENIJE[2], LUIS. A. BASTIDAS[3] & ASHVIN K. GOSAIN[4]

1 *Center for World Food Studies (SOW-VU), De Boelelaan 1105, VU University, 1081 HV Amsterdam, The Netherlands*
s.pande@sow.vu.nl

2 *Hydrology Division, Department of Civil Engineering and Geosciences, TU Delft, The Netherlands*

3 *Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan, Utah, USA*

4 *Department of Civil Engineering, Indian Institute of Technology, Delhi, India*

Abstract This paper deals with development of parsimonious models for dryland areas. The modelling approach is to capture the dominant processes of dryland areas in a data-limited environment. Two processes, evaporation and subsurface flows, are identified as dominant and are modelled at monthly time steps for a study area in western India. The area is represented by interconnected linear (in storage–discharge relationship) reservoirs, and each reservoir is parameterized to represent the two fluxes. The parameters are estimated based on GRACE (terrestrial storage change) and MERRA2D (evaporation flux) data simultaneously. Finally, parsimony in parameters of the overall model (of interconnected linear reservoirs) is achieved by regionalizing recession parameters in terms of soil characteristics. This study elicits an approach that is urgently needed in data and water scarce regions.

Key words parsimonious modelling; dryland modelling; water management in drylands; water management in India

INTRODUCTION

Development, data scarcity and need for parsimonious modelling

Water management is intricately linked to food security and to livelihoods of the agrarian poor in dryland areas of many developing countries. Most dryland agricultural production also tends to be rainfed (Bantilan *et al.*, 2006). Many governments have opted for reforms in water management with inclusive and sustainable economic growth in mind, and especially in watershed management with an emphasis on decentralized basin level water resource management (Dinar *et al.*, 2007). However, the success of such water policy reforms critically depends on reliable hydrological models. The robustness of such models is paramount, which becomes an exceptional challenge when faced with scarce relevant data for model selection. This is one wicked water management-related model selection problem.

Reliable hydrological models approximate the underlying set of processes as best as "data-possible". Models are conceptualizations (Savenije, 2009) and need to be evaluated with observations before selecting the best available conceptualization. A model that conceptualizes the processes better (low "approximation" error) is possibly more complex (Cucker & Smale, 2001). However, models selected from a set of complex models (in terms of number of parameters) are generally uncertain in prediction (high "estimation" error). Thus, there is an estimation/ approximation error trade-off in model selection.

This estimation/approximation error trade-off calls for a balance between complexity and process representation. A balance can be struck, when dominant processes are conceptualized via models that have a minimal set of parameters. One solution is to model dominant processes (at the scale of application) (Savenije, 2009) with concepts that have minimal sets of parameters, thus breaking the trade-off between "approximation" and "estimation" error.

We propose a parsimonious modelling approach as a solution to the wicked problem of model selection for water management in data scarce dryland areas. We conceptualize two dominant processes: evaporation and subsurface flows. Its parameterization is based on soil properties of the study area (Refsgaard & Storm, 1996; Vogel, 2006) and is robust. Additional simplicity is infused by modelling at monthly time steps, which is generally the scale for water policy.

STUDY AREA AND DATA

Study area

The study area comprises the arid/semi-arid states of Gujarat and Rajasthan in India, with an area of 538 346 km^2, minimum (maximum) average monthly precipitation and temperature of approx. 2 mm/month (202 mm/month) and 17°C (33°C), respectively. The relief is relatively flat except for the Aravali Hills extending from the northeast to southwest. The soil is sandy (northwest) to loamy-clay (southeast). The two states have impressive GDP growth (Planning Commission GOI, 2010) and are well-positioned for innovations in water management techniques.

Data

The Global 30 arc-second Elevation Data Set (GTOPO30) is used for the study area (available from the US Geological Survey). Gravity Recovery and Climate Experiment (GRACE) data from August 2002–August 2008 (yearly mean monthly) over land at 1° resolution is used for monthly changes in water storage (Chambers, 2006). The Modern Era Retrospective-analysis for Research and Applications (MERRA), a NASA reanalysis using the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5), is used to obtain monthly averages of surface evaporation fluxes for the year 2000 (Bosilovich, 2008). The data set used is at ½ × ⅔ degree (lat. × long.) resolution. The CRU TS2.1 data set (Mitchell & Jones, 2005) is used for yearly mean monthly total rainfall estimates, and mean temperature grid data from 1940 to 2000 at ½ degree resolution.

FAO's digital soil texture map (FAO, 2003) at 5° resolution is used. The European Union Joint Research Center's 30 arc seconds land cover map for the year 2000 over Southeast Asia (Beuchle *et al*., 2003) is also used. Finally, the agricultural census of India for the year 2000/01 (agcensus.nic.in/cendata/databasehome.aspx) provides data on the area of crops grown.

The study area is delineated using the DEM and the D8 algorithm of the ILWIS hydro-processing toolbox (ITC, 2009) to obtain a map of interconnected sub-basins. All spatial data sets are then re-sampled on a 8 × 8 km raster using a nearest neighbour method. Finally, basin-level average values of various variables are considered for later analysis.

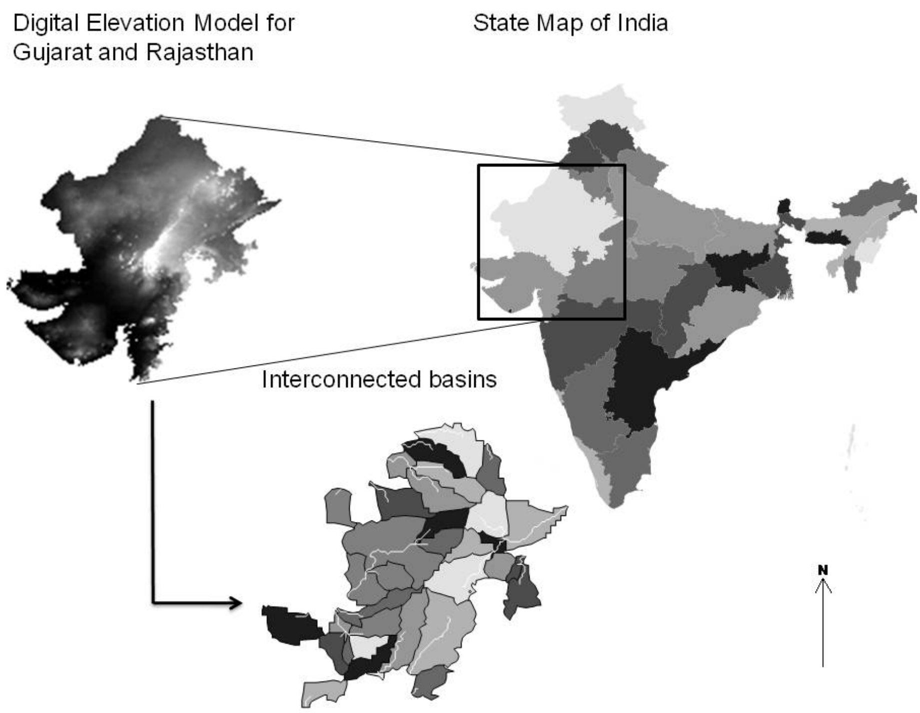

Fig. 1 Study area.

METHODOLOGY

Model conceptualization

Two dominant processes of dryland areas are modelled at monthly time steps and at the sub-basin scale (Liebe *et al*., 2009): evaporation and subsurface flows. Each sub-basin is conceptualized by a linear reservoir model, with precipitation and flows from upstream as influx, and flows downstream and evaporation as outflux. Flow is not decomposed into sub- and surface flows and is referred to as subsurface flow. However, intra-annual variation in water storage is modelled by enforcing storage at the end of the 12th month to be set as the storage at the beginning of the first month of the following year.

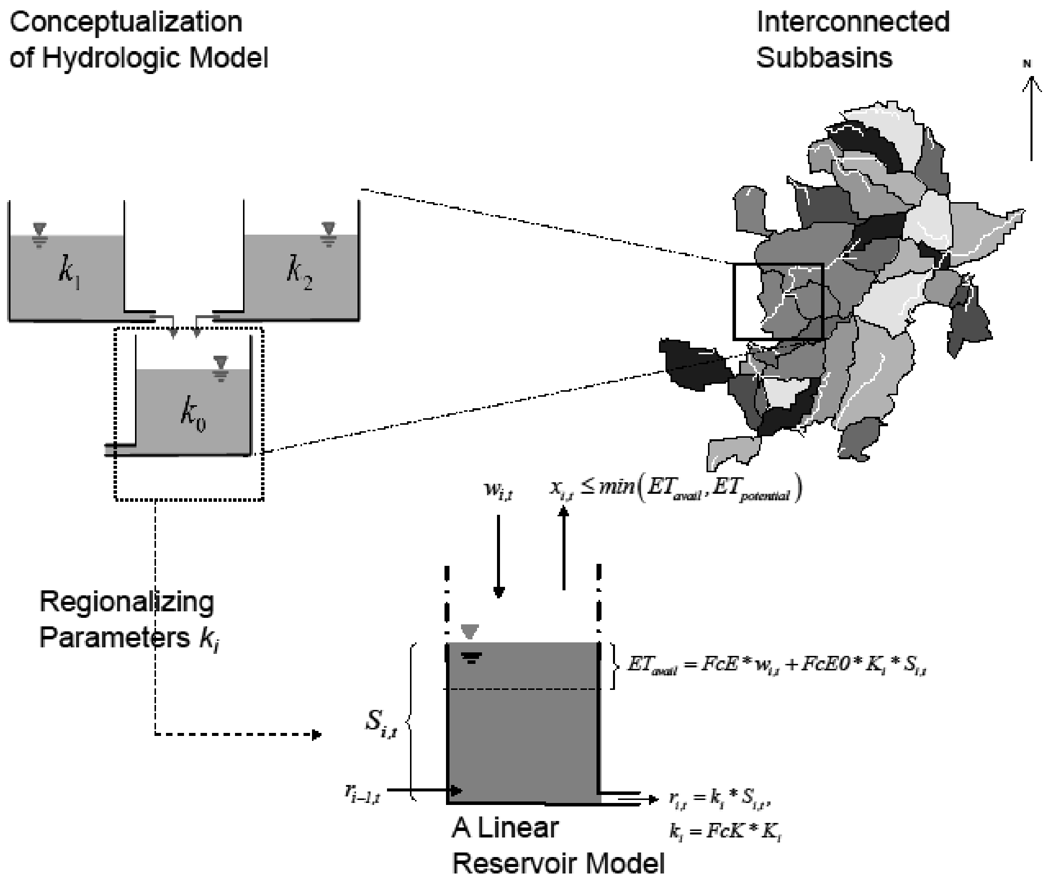

Fig. 2 A spatially explicit parsimonious model concept.

Actual evaporation for each store *i* (see Fig. 2), at monthly steps, is conceptualized as the sum of a fraction, *FcE*, of precipitation and a fraction, $FcE0 \times K_i$, of storage. Thus, evaporation is limited by water stress while the remaining precipitation fraction simultaneously contributes to infiltration (Reineker *et al*., 2007). The evaporation fraction of soil moisture conceptualizes that some of the available water storage also contributes to soil evaporation. The fraction is assumed to be proportional to hydraulic conductivity K_i, thus conceptualizing soil evaporation similarly to subsurface flow.

Evaporation demand from land cover is calculated based on FAO guidelines (FAO, 1998). Reference transpiration is calculated using the Hargreaves equation (Hargreaves *et al*., 1985), corrected by crop coefficients, to obtain evaporation demand. Land cover-specific evaporation demand acts as an upper bound on actual evaporation, wherein agricultural cover type (for the year 2000) is further decomposed into constituent crop area shares based on the Indian agricultural census of 2000/01.

Subsurface flow is conceptualized as a linear function of water storage. The storage-flow parameter, k_i, is a linear function of hydraulic conductivity K_i via parameter *FcK*. Similar to transmissivity (Franchini *et al*., 1996), k_i is location specific (store specific) with the multiple *FcK* being a function of store specific depth, anisotropy ratio, and a scale factor (similar to the e-folding length) (Fan *et al*., 2007).

Parameter estimation

Parsimony is introduced at the conceptual level. Spatial heterogeneity of the distributed lumped hydrological model is introduced via estimated hydraulic conductivity (Refsgaard & Storm, 1996). The geometric mean (by area) of hydraulic conductivity classified by soil texture (Rawls *et al*., 1982) data (FAO, 2003) corresponding to each store is used as its approximate saturated hydraulic conductivity at ground surface, and is assumed to decay with depth (Franchini *et al.*, 1996; Fan *et al.*, 2007).

The model parameters, {*FcE0*, *FcE*, *FcK*}, are calibrated by minimizing a joint objective function – the equally weighted sum of two mean square errors (MSEs). One MSE corresponds to the deviance between simulated changes in water storage and observed GRACE data, and the second MSE corresponds to the deviance between simulated evaporation and MERRA2D values. Besides model constraints, a "soft" constraint on maximum water storage in each store is introduced such that water storage at any time never exceeds annual precipitation. Land cover-specific upper bounds on evaporation are also introduced. These are akin to the use of the "soft" information of Fenicia *et al*. (2008) in model selection and controls for model complexity, thereby lending robustness to parameter estimation (Vapnik, 2002).

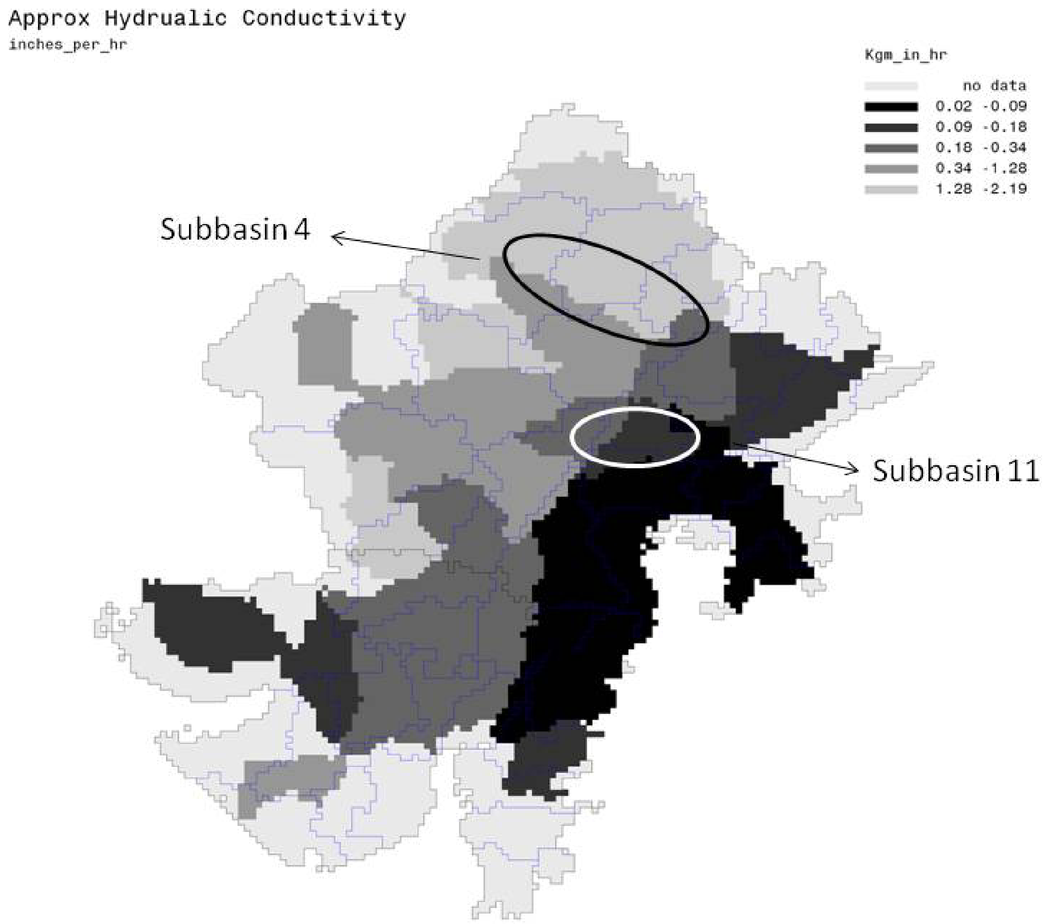

Fig. 3 Estimated hydraulic conductivity based on soil texture data of FAO (2003). Locations of sub-basins 4 and 11 are also shown. The map was made using software of van den Boom & Pande (2007).

We use complementary sources of information to identify and parameterize the two fluxes (Fenicia *et al*., 2008). The two data sources, GRACE and MERRA2D, are based on different satellite products. One measures changes in terrestrial water storage; the other is based on re-analysed data (Bosilovich, 2008). MERRA2D data controls evaporation parameters and the residual information in GRACE (including the effect of evaporation on storage changes) identifies subsurface flow parameters.

RESULTS

Using the described methodology and estimated hydraulic conductivity (see Fig. 3); Fig. 4 shows the performance of the estimated parameters for two sub-basins (from two different basins).

Sub-basin 4 belongs to the quicker response area, while sub-basin 11 belongs to the slower part. The left panel of Fig. 4 (Fig. 4(a) and (c)) compares model simulation of changes in monthly water storage with GRACE observations, while the right panel (Fig. 4(b) and (d)) compares simulated evaporation with MERRA2D data used for the two sub-basins. The GRACE and MERRA2D data shown were used to estimate the parameters.

The model is parsimonious in parameters and the data used for parameter estimation contains sufficient information to discriminate between the two out-fluxes. The additional "soft" constraints further control model complexity. The complexity control imbues confidence to the parameter estimation (may be limited by inadequate model structure) despite the limited observation data used.

The performance of estimated evaporation (Fig. 4(b) and (d)) appears to be adequate. One can observe that the model structure is inadequate to mimic the receding and the rising limbs of observed water storage change (Fig. 4(a) and (c)). This is due to suboptimal complexity of the chosen model structure and requires further investigation for diagnosis and correction.

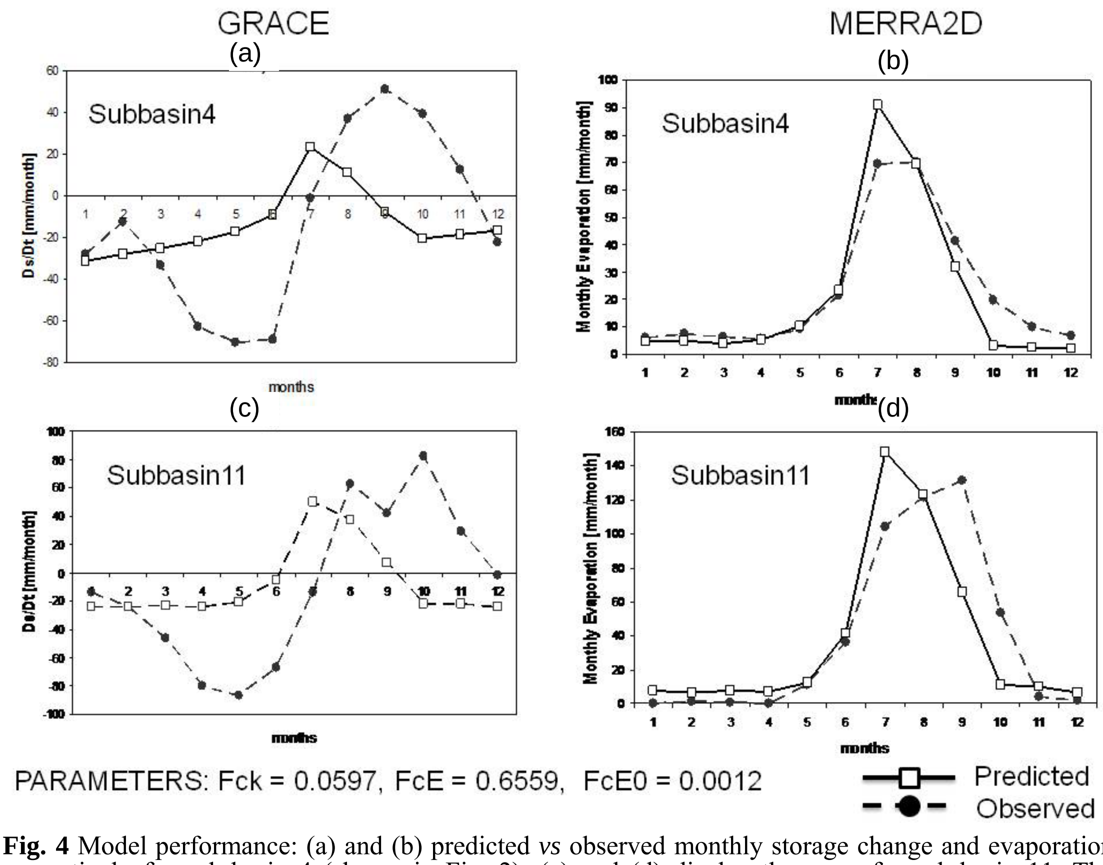

Fig. 4 Model performance: (a) and (b) predicted *vs* observed monthly storage change and evaporation respectively for sub-basin 4 (shown in Fig. 2); (c) and (d) display the same for sub-basin 11. The parameters of the model are also shown at the bottom.

KEY DISCUSSION POINTS

We have introduced a parsimonious modelling approach for water management in dryland areas that relies on fewer data and aims to provide reliable predictions for water policy design and management. The key to reliable predictions in this approach is to represent only dominant processes and to regionalize parameter estimation. Controlling model complexity while estimating parameters for a given model structure ensures reliability.

The utility of such an approach is significant, from the development perspective, in dryland regions that lack infrastructure, investment, and good quality data. The approach can be easily and quickly adopted with minimal costs to run models at a monthly time scale for water management objectives such as scenarios of water availability, testing water allocation mechanisms for equity and efficiency, interventions in times of water scarcity, etc.

REFERENCES

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. (1998) Crop evapotranspiration – Guidelines for computing crop water requirements, *FAO Irrigation and Drainage Paper 56*.

Bantilan M. C. S., A. Babu, A., Anupama, G. V., Deepthi, H. & Padmaja, R. (2006) Dryland Agriculture: Dynamics, Challenges and Priorities. *Research Bulletin no. 20. Patancheru 502324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics*. ISBN 92-9066-496-7.

Bosilovich, M. (2008) NASA's Modern Era Retrospective-analysis for Research and Applications: Integrating Earth Observations. *Earthzine*.

Chambers, D. P. (2006) Evaluation of new GRACE time-variable gravity data over the ocean. *Geophys. Res. Lett.* **33**(17), LI7603.

Cucker, F. & Smale, S. (2001) On the mathematical foundations of learning. *Bull. Am. Math. Soc.* **39**(1), 1–49.

Dinar, A., Kemper, K., Blomquist, W. & Kurukulasuriya, P (2007) Whitewater: decentralization of river basin water resource management. *J. Policy Modeling* **29**, 851–867.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R. & Robock, A. (2007) Incorporating water table dynamics in climate modeling: 1. water table observations and equilibrium water table simulations. *J. Geophys. Res.* **112**, D10125, doi:10.1029/2006JD008111.

FAO (2003) *The Digital Soil Map of the World*, version 2.3. Land and Water Development Division, FAO, Rome.

Fenicia, F., McDonnell, J. J. & Savenije, H. H. G. (2008) Learning from model improvement: on the contribution of complementary data to process understanding. *Water Resour. Res.* **44**, W06419, doi:10.1029/2007WR006386.

Franchini, M., Wendling, J., Obled, C. & Todini, E. (1996) Physical interpretation and sensitivity analysis of the TOPMODEL. *J. Hydrol.* **175**, 293–338.

Hargreaves L. G., Hargreaves, G. H. & Riley, J. P. (1985) Irrigation water requirements for Senegal River Basin. *J. Irrig. Drain. Engng* **111**(3), 265–275. doi:10.1061/(ASCE)0733-9437(1985)111:3(265)

ITC (2009) ILWIS 2.1 Application Guide. Downloaded from www2.itc.nl/ilwis/documentation/version_2/aguide.asp on 8 December 2009.

Liebe, J. R., van de Giesen, N., Andreini, M., Walter, M. T. & Steenhuis, T. S. (2009) Determining watershed response in data poor environments with remotely sensed small reservoirs as runoff gauges. *Water Resour. Res.* **45**, W07410, doi:10.1029/2008WR007369.

Mitchell, T. D. & Jones, P. D. (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**, 693–712, doi:10.1002/joc.1181.

Planning Commission, Government of India (2010) Growth Rate – GSDP % (current prices) as per CSO Data 1997–98 onwards as on 12.11.2009. Downloaded from: http://planningcommission.nic.in/data/datatable/0110/tab64.pdf.

Rawls, W. J., Brakensiek, D. L. & Saxton, K. E. (1982) Estimation of soil water properties. *Trans. ASAE* **25**(5), 1316–1320 & 1328.

Refsgaard, J. C. & Storm, B. (1996) Construction, calibration and validation of hydrological models. In: *Distributed Hydrological Modelling* (ed. by M. B. Abbott & J. C. Refsgaard), 41–54, Springer, New York, USA.

Rienecker, M. M. and co-authors (2008) The GEOS-5 Data Assimilation System – Documentation of Versions 5.0.1 and 5.1.0. *NASA GSFC Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2007-104606, vol. 27*.

Savenije, H. H. G. (2009) The art of hydrology. *Hydrol. Earth Syst. Sci.* **13**, 157–161.

Stibig H.-J., Upik, R. Beuchle, R., Hildanus, R. & Mubareka, S. (2003) The Land Cover Map for South East Asia in the Year 2000. GLC2000 database, European Commission Joint Research Centre, http://www-gem.jrc.it/glc2000.

van den Boom, B. & Pande S. (2007) A user manual for the SAS facility to plot maps. Working Paper 07-03. Amsterdam, SOW-VU.

Vapnik, V. (2002) *The Nature of Statistical Learning Theory*, 2nd edn. Springer, New York, USA.

Vogel, R. M (2006) Regional calibration of watershed models. In: *Watershed Models* (ed. by V. Singh & D. Frevert), 47–69. Taylor & Francis, UK.