8-1-2014

Comparative Study On Water Resources Assessment Between Kenya And England

Hesbon Otieno

Dawei Han

Recommended Citation

Otieno, Hesbon and Han, Dawei, "Comparative Study On Water Resources Assessment Between Kenya And England" (2014). *International Conference on Hydroinformatics.* Paper 3.
http://academicworks.cuny.edu/cc_conf_hic/3

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

COMPARATIVE PERFORMANCE ASSESSMENT OF HYDROLOGICAL MODELS

HESBON OTIENO, DAWEI HAN, ROSS WOODS

Water and Environmental Management Centre, Department of Civil Engineering, University of Bristol, UK

ABSTRACT

Water resources management decisions are made based on information from predictive models that are capable of simulating the behavior of hydrological systems. More of these models are in use today and it is becoming increasingly difficult to choose which model to use for particular space and time scales as well as climate. In addition, as a result of climate change there is a an increase in the degree of randomness in hydrological systems leading to reduced predictability of these systems and thus different models are prone to perform differently under varying conditions. Meta-analysis was conducted involving seven commonly applied models in hydrological assessment to try and establish patterns that these models exhibit under varying situations. This was achieved by looking at the homogeneity of the studies at the various space and time scales. In addition to the meta-analysis, a second stage of analysis looking at the variation in performance of the models with catchment characteristics such as climate, mean altitude and catchment size was assessed. Results from the review study showed varied performance with respect to the catchment characteristic and are important in aiding decision making regarding hydrological model selection.

INTRODUCTION

Hydrological models are useful tools in transforming the meteorological forces into hydrological response of a catchment [1]. Applying a hydrological model is usually in most cases the only economically possible way to obtain quantitative figures about the impact of environmental changes on the hydrological cycle. These computer-based catchment models save time and money in facilitating the simulation of outcomes that are crucial in effectively managing water resources. There is a growing pool of hydrological models and users are increasingly finding it difficult to decide on which model to pick for application, despite the fact that there is a common agreement that the models should in general solve the governing differential equations [2]. This form of intellectual diversity is healthy and is needed for the field of hydrological modelling to progress. The challenge is therefore to identify the most economic and efficient approach that fits the objectives of a particular study in an individual catchment. Numerous studies have been done using hydrological models; however, despite the large number of studies involving the models, there is a lack of systematic and comparative analysis on their performance. The objective of this review was thus to look at the homogeneity of the performance of various hydrological models across different geographical regions and catchment scales in order to help the hydrological community choose the suitable models in their applications.

Candidate models

Seven widely used models in water resources assessment were considered in the review which include Soil and Water Assessment Tool (SWAT), Hydrological Simulation Program-Fortran (HSPF), Hydrologiska Byrans Vattenbalansavdelning (HBV), Variable Infiltration Capacity (VIC), Systeme Hydrologique Europeen (MIKE-SHE), Precipitation- Runoff Modelling System (PRMS) and Hydrologic Engineering Center-Hydrologic Modelling System (HEC-HMS). SWAT incorporates features of several Agricultural Research Service (ARS) models and is a direct beneficiary of Simulator for Water Resources in Rural Basin (SWRRB), Chemical, Runoff and Erosion from Agricultural Management Systems (CREAMS), Groundwater Loading Effects on Agricultural Management Systems (GLEAMS) and Erosion-Productivity Impact Calculator (EPIC). It was developed in the early 1990's and has undergone continued improvement of capabilities as detailed by [3]. HSPF evolved out of the Stanford Watershed model, the US Environmental Protection Agency (USEPA) Agricultural Runoff Management (ARM) and Non-Point Source (NPS) models in the late 1970's. HBV was originally developed for use in the Scandinavian climate by the Swedish Meteorological and Hydrological Institute (SMHI) has been a standard tool for operational hydrology in the Nordic region, but has been adopted in other regions too. VIC model is hosted by the University of Washington and was developed by Xu Liang [4] and later improved by [5] and [6]. Originally developed as SHE, MIKE-SHE was improved by a consortium of three organisations: the Institute of Hydrology UK, Danish Hydrological Institute (DHI) and a French consulting firm SOGREAH; it became operational in 1982. Currently, DHI hosts the model and continues to enhance it as well as providing support services. PRMS was developed by the US Geological Survey (USGS) and is modular in design, whereas HEC-HMS was developed by the US Army Corps of Engineers (USACE). Important features of these seven models can be found in numerous peer reviewed publications.

A systematic review to summarise findings of the seven selected hydrological models was conducted. Relevant studies published were accessed in various depositories including journals and dissertations/thesis and where possible downloaded for analysis. Special emphasis was on studies in which the model was used for hydrological assessment. Key information extracted from the studies included the model used, where the model was applied, authors of the study, size of the catchment, temporal step, climatic region of the catchment and the performance statistics (Nash and Sutcliffe Efficiency NSE) at the temporal time step. Table 1 shows the summary of the number of studies reviewed for each model.

Table 1: Summary statistics of candidate models (total 332 cases studies)

	n	Area (km^2)			NSE		
		Min.	Mean	Max.	Min.	Mean	Max.
SWAT	107	0.395	90179.1	4470500	0.39	0.79	0.99
HSPF	35	0.78	23061.8	160000	0.4	0.77	0.97
HBV	37	4.81	34116.8	403933	0.3	0.67	0.98
VIC	64	438	93659.3	1100000	0.4	0.75	0.98
MIKE-SHE	30	1.6	15762.2	350000	0.54	0.75	0.96
PRMS	35	1.4	353.3	2154	0.25	0.72	0.93
HEC-HMS	24	75	17116.7	177230	0.48	0.78	0.98

n is the number of case studies
NSE is Nash and Sutcliffe Efficiency

Performance assessment

Model performance assessment was done in two stages: 1) Meta-analysis to establish the homogeneity of the studies and 2) Analyzing the model performance with respect to climate, altitude and size of catchment. Meta-analysis employs statistical methods to combine and contrast the results of individual studies in order to reveal patterns, sources of disagreements or interesting relationships among the studies. Its fundamental index, the effect size provides information about treatment effect or relationships between two variables. Each study has an effect size used to evaluate its consistency in relation to the other studies. In this study the effect size was based on the explained variance captured by the Nash and Sutcliffe Efficiency (NSE) realized from the individual studies. The Nash and Sutcliffe Efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance [7]. It indicates how well the simulated data reproduces the observed data and is computed as shown in equation 1.

$$NSE = 1 - \left[\frac{\sum(Q_{obs} - Q_{Sim})^{2}}{\sum(Q_{obs} - \bar{Q})^{2}}\right] \qquad 1$$

where Q_{obs} is the observed value, Q_{sim} is the simulated value and $\bar{Q}$ is the mean of the observed values. NSE ranges between -∞ and 1, with values less than or equal to zero indicating that the mean observed value is a better predictor than the model simulated values, and thus indicating a unacceptable model performance. On the other hand positive NSE values close to 1 indicate acceptable levels of model performance. In the current study NSE values for the case studies were extracted for meta-analysis by looking through individual case studies. It is important to note that various studies have used different performance statistics in addition to NSE (and some studies have confused between R-squared and NSE), however for the sake of consistency the formula used to compute the model efficiency had to conform to Equation 1 for the model efficiency statistics to be considered. The NSE is preferred for hydrological model evaluation because it is the best objective function that reflects the overall modelling fit in addition to being sensitive to high extreme values due to the squared differences [8]. Where possible, for each model, the effect size for daily, monthly and annual time steps were used in the meta-analysis to investigate the similarity amongst the various studies. Assessment of the homogeneity of the studies was done visually on the Forest plot and statistically using I-squared statistic. In the Forest plot, a straight vertical line is drawn through the studies. If it passes through each confidence interval simultaneously, it is an indicator of homogeneity whereas if the line does not pass through each confidence interval simultaneously, then variations exist. Confidence intervals describe the uncertainty in the estimates and highlights the range of values within which the true effect lies. Narrower confidence intervals signal precision whereas wider confidence interval is indicative of greater uncertainty. The I^2 statistic serves as a kind of signal to noise ratio [9] and is given as shown in Equation 2.

$$I^{2} = \frac{Q - df}{Q} * 100 \qquad 2$$

where Q is a statistic that estimates the amount of variation due to sampling error at $p<0.05$, and df is the degree of freedoms (n-1). The threshold levels of the I^2 statistic are summarized in Table 2.

Table 2: I-squared statistic threshold table (Adapted from Cochrane Handbook)

I^2 Threshold (%)	Level of heterogeneity
0-40	Insignificant
30-60	Moderate
50-90	Substantial
75-100	Considerable

The I^2 statistic is preferred as a measure of heterogeneity to Q statistic because it removes the dependency on the number of studies, and expresses the result as a ratio. The second stage of the performance assessment considered NSE statistic from daily time steps scenarios and looked at its variations with respect to latitude, which was used as a proxy to climate [10], mean altitude and size of the catchment. The latitude (^{0}N) was grouped into >55, 40-55, 20-40, 0-20,

<0, altitude (m. asl) grouped into <300, 300-600, 600-900, 900-1500, >1500, whereas the area (km^2) was grouped into <250, 250-500, 500-1000, 100-10000, >10000.

RESULTS AND DISCUSSIONS

Confidence intervals and Heterogeneity

In all models, daily time step showed a considerable level of heterogeneity for both calibration and validation, while for the monthly time step, the levels of heterogeneity varied from insignificant (MIKE-SHE calibration) to considerable heterogeneity. As for the annual time step there were only studies using SWAT, VIC and PRMS and showed insignificant levels of heterogeneity (Table 3). In general, the confidence interval was seen to be increasing with the time step, suggesting that the precision (not accuracy) of the hydrological modelling studies is high at shorter time scale. However, when looking at individual studies the general trend was that the effect size (i.e., NSE which represents the model accuracy) was seen to improve with increasing time step.

With regard to space and time scales, SWAT, HSPF, HBV, MIKE-SHE, PRMS and HEC-HMS models showed a higher variation in the effect sizes at smaller catchment size and the variation reduced with the increase in catchment size. This situation was observed at all the three time steps for calibration and validation. The effect sizes are seen to approach the average effect size as the catchment size increases. VIC model also showed higher variation of the effect sizes at smaller catchments and variation reduced with the larger catchments, however in bigger catchments, the effect size was found to be generally lower than the average effect sizes for small catchments.

Table 3: I-squared statistics for the models

Model, Time step	Calibration	Validation
SWAT, Daily	99.4 (64)	99.9 (62)
SWAT, Monthly	89.2 (70)	98.0 (74)
SWAT, Annual	0.0 (12)	0.0 (9)
HSPF, Daily	99.2 (31)	98.7 (24)
HSPF, Monthly	77.6 (13)	68.1 (16)
HSPF, Annual	-	-
HBV, Daily	98.6 (32)	99.7 (17)
HBV, Monthly	89.7 (20)	97.3 (13)
HBV, Annual	-	-
VIC, Daily	98.7 (11)	99.3 (10)
VIC, Monthly	95.6 (48)	96.6 (5)
VIC, Annual	-	0.0 (6)
MIKE-SHE, Daily	99.2 (24)	99.1 (21)
MIKE-SHE, Monthly	0.0 (3)	85.1 (2)
MIKE-SHE, Annual	-	-
PRMS, Daily	99.7 (21)	99.4 (19)
PRMS, Monthly	46.9 (3)	74.5 (3)
PRMS, Annual	0.0 (15)	-
HEC-HMS, Daily	99.7 (23)	99.2 (24)
HEC-HMS, Monthly	93.0 (2)	90.1 (2)
HEC-HMS, Annual	-	-

Values in brackets indicate the number of studies

Performance based on climate

Figure 1 shows the performance of the models with respect to the latitude of the catchment. SWAT and HSPF showed decreasing performance as once moved away from the north towards the equator (latitude 0), whereas SHE, PRMS, HEC-HMS showed an increasing performance as one moved away from the north towards the equator. HBV studies were only in the region of latitude greater than 40 and also showed an increase in performance as one moved away from the north. VIC on the other hand did not reveal any significant pattern oscillating at a median NSE value of about 0.75. As one approaches the equator from the north the climate changes from temperate (cold) to tropical and the spatial homogeneity deteriorates making the hydrological processes less linear [11]. This therefore implies that SHE, PRMS and HEC-HMS appear to be good in capturing this spatial heterogeneity than SWAT and HSPF.

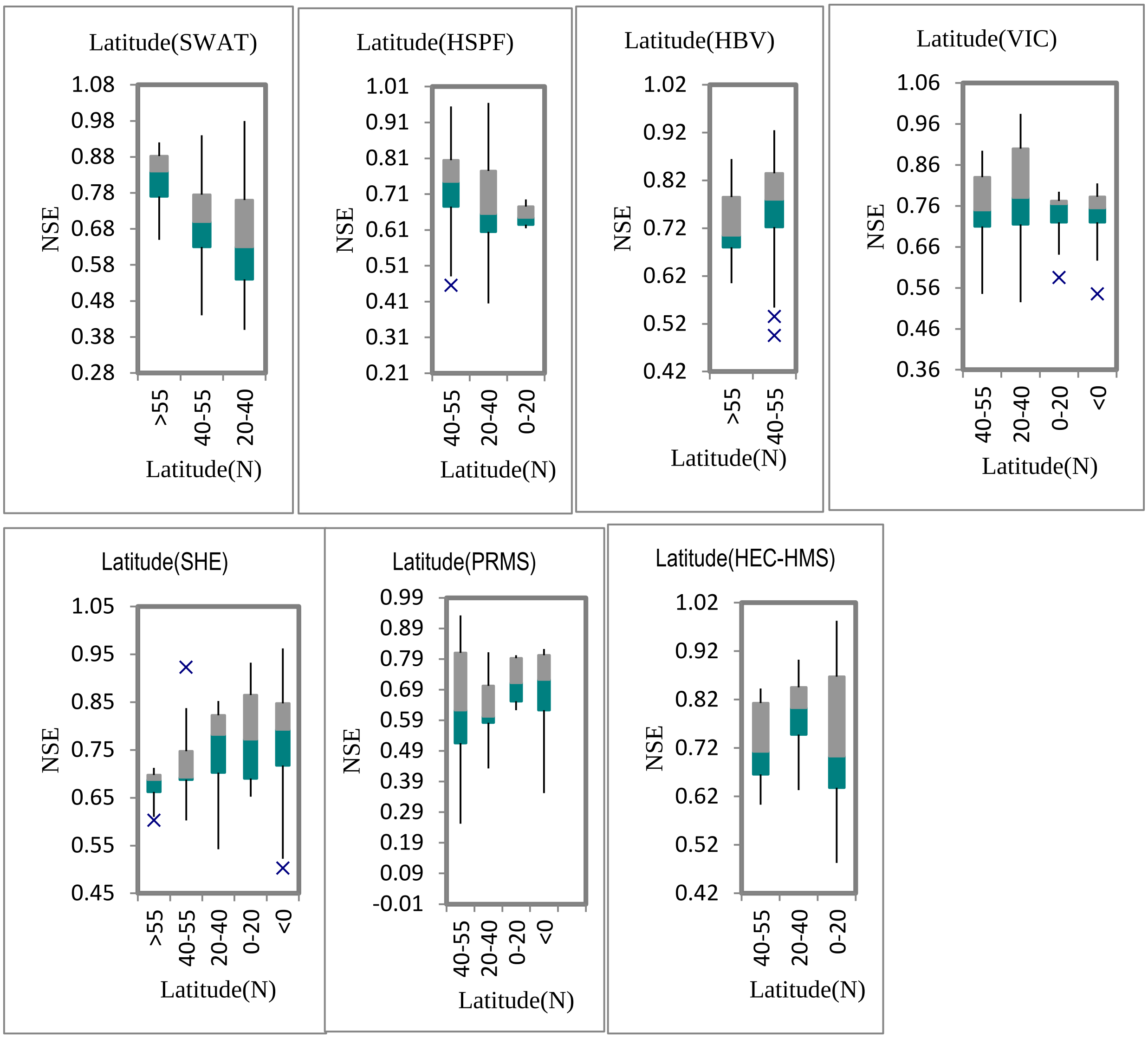

Figure 1: Variation of model performance with latitude

The above results also corroborate the findings by [12] in which they investigated the effect of performance of regionalization methods with climate and found out that the predictive performance decreases with increasing aridity.

Performance based on altitude

Figure 2 shows the performance of the models with respect to the average altitude of the catchment. SHE and PRMS showed an increasing performance with a corresponding increase in the mean altitude, whereas HEC-HMS, SWAT and HBV showed a decreasing trend in NSE with respect to altitude. On the other hand VIC and HSPF did not reveal a significant trend

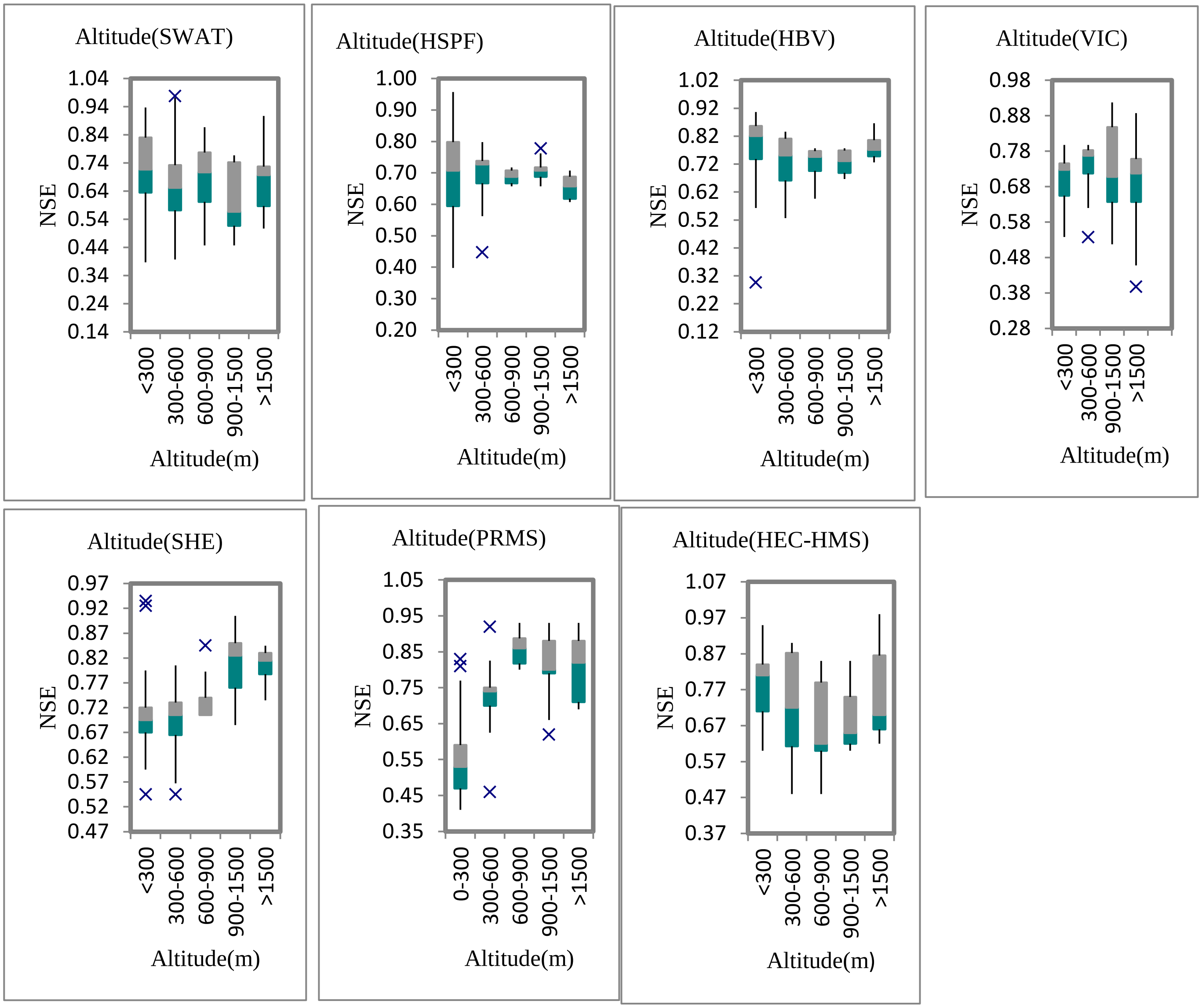

Figure 2: Variation of model performance with mean height above sea level

Performance based on the catchment size

Figure 3 shows the performance of the models with respect to the size of the catchment and reveals a pattern of increasing performance with a corresponding increase in the size of the catchment for all the models except SHE which shows an improved performance as the area increases from 250-10000km^2. However, catchment sizes in excess of 10000km^2 exhibited an increase in variability in performance for SWAT, HBV, and HEC-HMS. With the increase in the size of catchment an interaction of space-time processes leads to the averaging out of some hydrological variability which appears to lead to higher NSE values. Larger catchments are also associated with a high number of gauging stations which could also be a reason for their good performance. VIC and HEC-HMS were the best performers in smaller catchments with a median performance of around 0.7 and 0.8 respectively. SHE also performs best at areas below

250km^2. While PRMS is the worst performer in smaller catchments with a median performance of around 0.61 for areas less than 250km^2. However, PRMS performs best for areas greater than 500km^2 (no studies involving PRMS reviewed were in catchment size greater than 10,000km^2) with a median NSE of 0.8. For areas exceeding 10000km^2, VIC and HBV were the best performer with a median NSE of nearly 0.8.

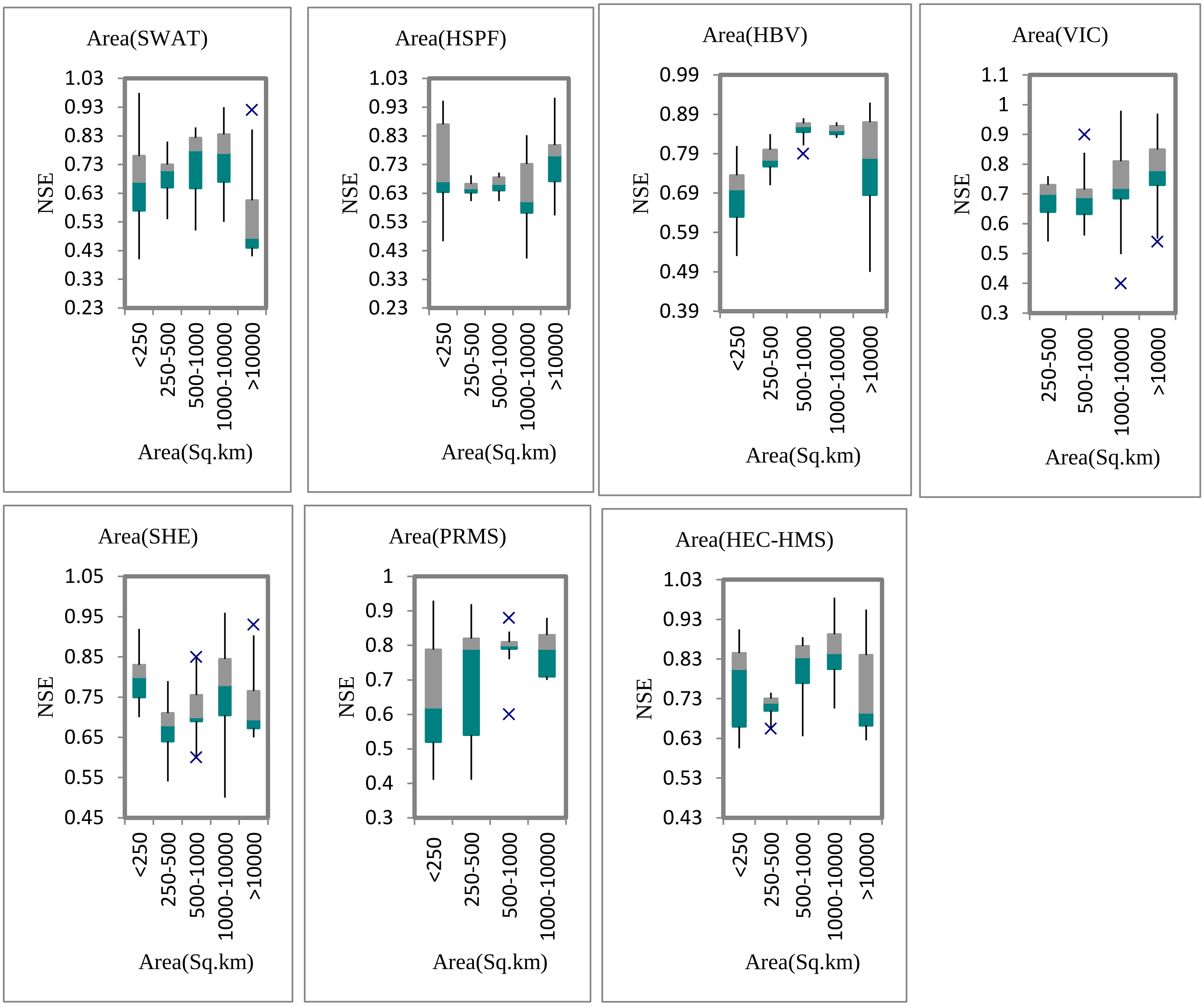

Figure 3: Variation of model performance with size of catchment

CONCLUSION

Understanding the behavior of hydrological models with respect to various aspects of the catchment in which they are to be applied is an important concern in hydrological modelling studies. Hydrological model users are constantly faced with a difficult task of choosing from an increasing number of models to use. This review study has attempted to explore the variations in performance of seven selected models with respect to general catchment characteristic such as climate, mean altitude and catchment size. The results realized indicate varying levels of performance indicating varying level of strengths in performance with respect to a particular parameter, and could assist in the selection of the best model for specific conditions even though certain factors such as models data requirements and the available data to the user are also fundamental in decision making. Further reviews' involving more models and more case studies is recommended to ascertain whether the findings in the current study will be vindicated.

REFERENCES

[1] Bronstert, A. (2004). Rainfall-runoff modelling for assessing impacts of climate and landuse change. Hydrological Processes, 18(3), 567-570.

[2] Fleming, S. W. (2009). An informal survey of watershed model users: Preferences, applications, and rationales. Streamline Watershed Manage. Bull, 13(1), 23-35.

[3] Arnold J.G., and Fohrer N. (2005). SWAT2000: current capabilities and research opportunities in applied watershed modelling. Hydrological Processes 19:563-572

[4] Liang X., Lettenmaier D.P., Wood E.F.,and Burges S.J., (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys Res. 99: 14415-14458

[5] Lohmann, D., Raschke, E., Nijssen, B., & Lettenmaier, D. (1998). Regional scale hydrology: II. Application of the VIC-2L model to the Weser River, Germany. Hydrological Sciences Journal, 43(1), 143-158.

[6] Liang, X., & Xie, Z. (2001). A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. Advances in Water Resources, 24(9), 1173-1193.

[7] Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3), 282-290.

[8] Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R. and Veith, T. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50(3), 885-900.

[9] Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. BMJ: British Medical Journal, 327(7414), 557.

[10] Peel M.C., Finlayson B.L., and T.A. McMahon (2007). Updated map of the Koppen-Geiger climate classification. Hydrology and Earth Systems Sciences (HESS), 11:1633-1644

[11] Parajka J., Viglione, A., Rogger M., Salinas J.L., Sivapalan M., and Bloschl G. (2013). Comparative assessment of predictions in ungauged basins- Part 1: runoff hydrograph studies, Hydrology and Earth Systems Sciences (HESS), 17:1783-1795

[12] Salinas J.L., Laaha G., Rogger M., Parajka J., Viglione, A., Sivapalan M., and Bloschl G. (2013). Comparative assessment of predictions in ungauged basins- Part 2: Flood and low flow studies, Hydrology and Earth Systems Sciences (HESS), 17:2637-2652