On the assessment of water resources model predictions

Raúl Fernando Vázquez Zambrano

Grupo de Ciencias de la Tierra y del Ambiente, Dirección de Investigación de la Universidad de Cuenca, Av. 12 de Abril S/N, Cuenca, Ecuador

Autor para correspondencia: raulfvazquezz@yahoo.co.uk

ABSTRACT

The analysis of the most commonly used measures of hydrological/hydraulic model performance was herein carried out by means of their statistical examination and illustrative modelling applications. In doing so, the model performance indexes were classified in two groups, according to the type of error (or residual) that those indexes are measuring: (i) statistics measuring the average systematic error (model bias); and (ii) statistics measuring the average combined systematic and random discrepancies among simulations and observations. The reader can in this way easily select a set of unrelated statistics to report on model performance. The manuscript addresses as well the main pitfalls of some of the most popular statistics used in scientific literature and suggests some approaches to overcome such potential pitfalls when addressing model performance.

Keywords: Hydrologic/hydraulic modelling, model performance, statistic, calibration, evaluation, validation, uncertainty, Monte Carlo simulations.

RESUMEN

En el contexto de modelación hidrológica/hidráulica, este artículo describe el análisis de los estadísticos de calidad de modelación empleados con mayor frecuencia en publicaciones científicas. El análisis se basó en el examen de los estadísticos y su aplicación en el contexto de modelación hidrológica. Así, estos estadísticos se clasificaron en dos grandes grupos de acuerdo al tipo de error que los mismos son capaces de percibir: (i) estadísticos que miden el error sistemático medio; y (ii) estadísticos que miden la combinación de los errores sistemático y aleatorio. De esta forma, el lector está en capacidad de seleccionar un grupo de estadísticos que midan información diferente de la población de errores (o residuos de modelación). El artículo se ocupa además de las debilidades principales de los estadísticos más populares, citados en la literatura científica, y sugiere algunas aproximaciones que podrían emplearse para mitigar dichas debilidades al momento de evaluar la calidad de la modelación numérica.

Palabras clave: Modelación hidrológica/hidráulica, calidad de modelación, estadístico, calibración, evaluación, validación, incertidumbre, simulaciones Monte Carlo.

1. INTRODUCTION

Numerical models, when properly applied, have the potential of being useful tools that can be used in decision making applications to select an optimal course of action (Vázquez, 2003). In this context, the modelling of the hydrological system is achieved through a numerical model that is, in its intrinsic nature, a simplified representation of reality. Under the pressure of being able to predict the impact of human activity on the environment there is a clear shift in the codes used in hydrology as well as in other fields, and this from lumped empirical approaches to fully distributed mechanistic codes (Refsgaard, 1997; Vázquez, 2003).

In recent years, considerable progress has been made in the conceptual approach and mathematical description of water flow and transport processes (Abbaspour *et al.*, 1997; Feyen *et al.*, 2000). Several mathematical codes of diverse complexity have been developed. For instance, in the past three decades, several physically based spatially distributed codes were developed such as the code by Freeze and Harlan (1969), MIKE SHE (Refsgaard and Storm, 1995), TOPMODEL (Beven, 1997) and THALES (Grayson *et al.*, 1992).

A model is a good representation of reality only if it can be used to predict, within a calibrated and evaluated range, an observable (and therefore measurable) phenomenon with acceptable accuracy and precision (Loague and Green, 1991). However, with regard to calibration and evaluation aspects, no unique terminology and/or approach are presently used by the hydrological community. The use of hydrological models for a broad range of applications has increased rapidly during the last few decades. The ease with which they are applied often leads to a less critical evaluation of the basin being modelled as well as the applied model evaluation tests (Legates and McCabe, 1999).

Besides some statistical analyses of timeseries of discharge and precipitation, few numerical modelling studies have been carried out in the Andean region (Braud *et al.*, 1999; Perrin *et al.*, 2001; Buytaert *et al.*, 2004; Buytaert *et al.*, 2006; Fleischbein *et al.*, 2006). They have concentrated on the analysis of the effects of diverse land use and soil scenarios on the hydrological production of the modelled study sites, without properly considering modelling issues such as the congruency of evaluation tests and model structure, as well as, modelling (i.e. model structure-data-parameter) uncertainty issues. The proper evaluation of model performance requires multi-criteria tests that combine visual (i.e. graphical) qualitative techniques as well as statistical quantitative analyses (Beven, 1993; Refsgaard, 1997; Vázquez, 2003; Vázquez *et al.*, 2009). These are required particularly when using physically based distributed models (Feyen *et al.*, 2000) as, owing to their intrinsic complex distributed nature, they are multi-input and multi-output.

In this context, several modelling studies have been carried out in the region (Braud *et al.*, 1999; Perrin *et al.*, 2001; Buytaert *et al.*, 2004; Buytaert *et al.*, 2006; Fleischbein *et al.*, 2006) by using model performance quantitative statistics that have been chosen by their "popularity" in scientific publications rather than by choosing them on the basis of a thorough analysis of the type of model error that they depict. Choosing an appropriate set of (unrelated) model performance statistics is rather important as the general objective of those statistics is then to extract the essential information from the usually large set of model predictions, reducing them to a small set of performance statistics. Thus, the current manuscript describes some of the most commonly used model performance statistics and addresses the type of model error that they are capable of perceiving. In doing so, the manuscript depicts, through some modelling results, the advantages and disadvantages of some "popular" indexes that are commonly used in model applications.

2. MATERIALS

2.1. The study site

Some of the concepts included in this manuscript are illustrated using results related to the hydrologic modeling of the Gete catchment (586 km^2), located to the east of Brussels, in the sandy loam region of Belgium. It comprises the sub-catchments of the Grote Gete (326 km^2) and the Kleine Gete (260 km^2) up to their confluence at the community of Budingen (Figure 1). The elevation of the area varies from approximately 27 m in the northern part to approximately 174 m in the southern parts. Land use is mainly agricultural, both pasture and cultivated fields, with some local forested areas. Nine soil units can be distinguished according to the legend of the Belgian soil map, e.g. loamy soils (Aba, Ada and Adc), sand-loamy soils (Lca, Lda and Ldc), clay soils (Eep and Uep) and soils with stony mixtures (Gbb). The dominant soil type in the catchment is the Aba soil unit. The geology of the catchment comprises nine geological units, some of which occur only in isolated parts of the catchment (Vázquez and Feyen, 2007).
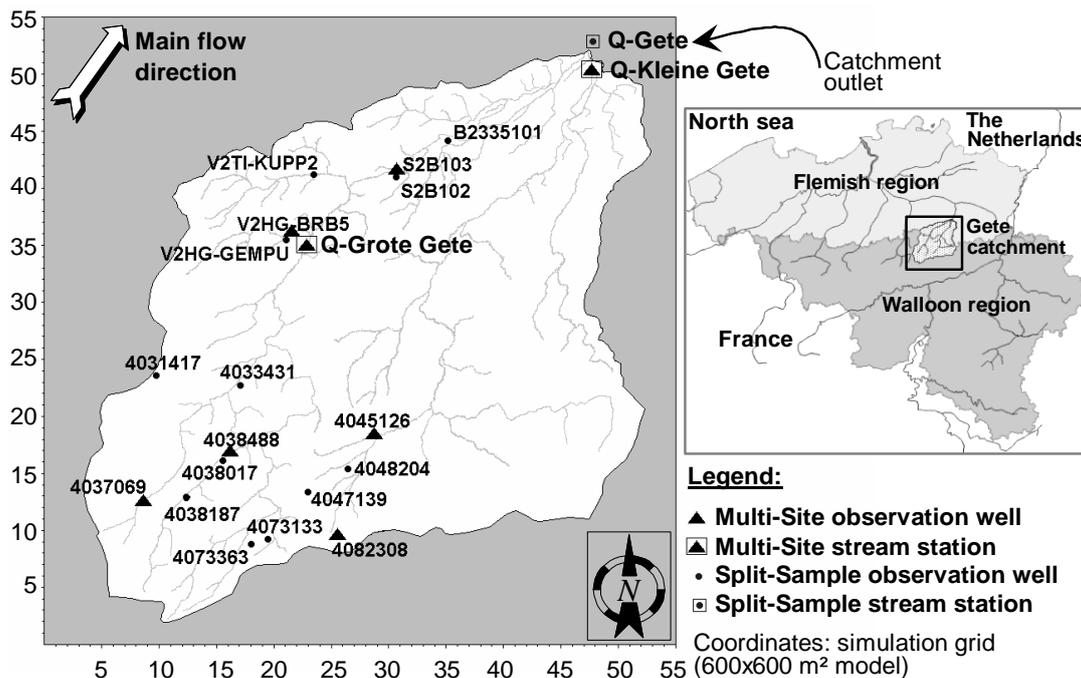
Figure 1. Location of the Gete catchment and distribution of the hydro-meteorological stations (after Vázquez and Feyen, 2007).
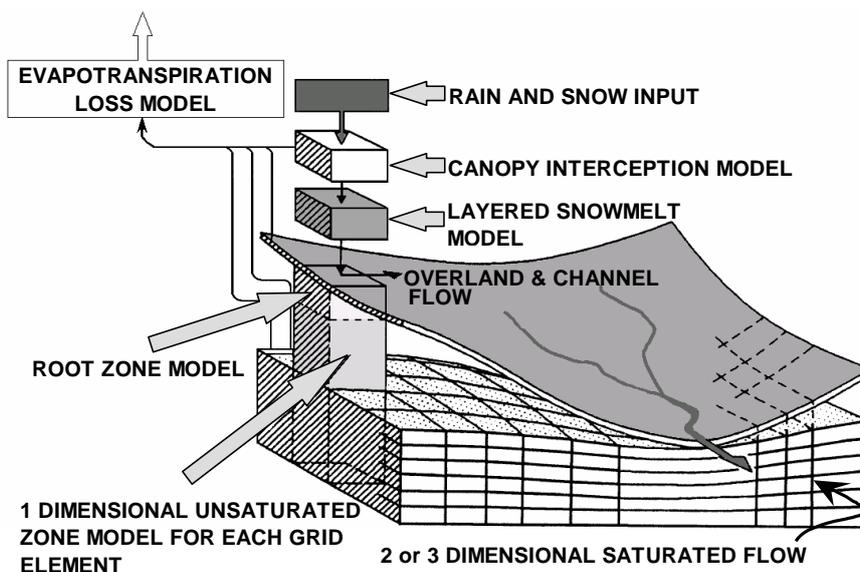

Figure 2. Schematic representation of the water movement module of the MIKE SHE structure (after Vázquez, 2003; Vázquez and Feyen, 2008).

2.2. *The modelling system*

The hydrological modelling system MIKE SHE (Refsgaard and Storm, 1995) was used in the scope of the current manuscript. MIKE SHE is a spatially distributed, physically based system for the integral simulation of the hydrological processes of the land phase of the hydrological cycle. The code is based on the SHE (Système Hydrologique Européen) modelling concept (Abbott *et al.*, 1986). The reader is referred to Vázquez and Feyen (2008) for details on the main capabilities and characteristics of its respective structure.

3. METHODS

3.1. *Overview of main model performance statistics and analysis of their main statistical properties*

The model performance statistics that are commonly used in specialized literature can be classified according to different aspects, such as the mathematical form of the statistic as well as the type of modeling error that they are capable to map (Vázquez, 2003; Vázquez *et al.*, 2008). The approach that is followed herein is a combination of both aspects. This section of the manuscript not only presents an overview of the most common utilized model performance statistics but also reports an analysis of their main statistical properties. In doing so, their mathematical formulation has been examined thoroughly so as to derive potential relationships among them. It is important realizing that only a set of unrelated statistics should be used for model calibration and/or evaluation and other modeling aspects such as model development (Vázquez *et al.*, 2002).

Performance statistic	Optimum value	Feasible interval	Quoted in literature as	Units of measure
Residual BIAS $$BIAS = \mu_r = \frac{\sum_{i=1}^{n}(O_i - P_i)}{n}$$	0	$-\infty < BIAS < +\infty$	Residual Mean (RM) Mean Error (ME)	Same as variable
Maximum Absolute Error $$ABSMAX = MAX\left|O_i - P_i\right|_{i=1}^{i=n}$$	0	$0 \leq ME < +\infty$	Maximum Error (ME)	Same as variable
Mean Absolute Error $$MAE = \frac{\sum_{i=1}^{n}\left|O_i - P_i\right|}{n}$$	0	$0 \leq MAE < +\infty$	ABSERR	Same as variable

Figure 3. Statistics measuring the mean residual (after Vázquez, 2003; Vázquez *et al.*, 2008).

Analysis of residual errors can be utilized to evaluate model performance. A residual is defined as the difference between an observed and a computed dependable model variable. A first group of model performance statistics is measuring the residual bias, it is, and they measure the average systematic error among the simulated and the observed variables (Vázquez *et al.*, 2008). Some of these statistics are presented in Figure 3 depicting information on their mathematical formulation, optimal value, variation interval, units of measure and the way in which they are quoted in different scientific publications. Observe that there is not a single quotation terminology that is being used in scientific literature. Thus, it is rather important to carry out a through examination of the kind of statistic that has been used in a given (prior) research activity.

The equations in Figure 3 are traditionally applied to each model parameter set to be calibrated on the basis of multi-objective optimization; P_i is the i-th predicted output as a function of the model parameter set; O_i is the i-th measured historical data; and n is the total number of available measurements in the considered time interval.

The most common representative index from this group is the residual BIAS. This statistic is a rough estimator of the total inaccuracy of the model for predicting observed data. For a good calibrated model, residuals certainly balance each other in the simulation period. However, inspection of Figure 3 reveals that very significant errors having similar order of magnitude but opposite signs may cancel out each other such that at the end the residual mean is close to zero whilst the model performance remains poor (Vázquez, 2003). In this sense, the mean absolute error (MAE) is a better measure of the average error in the modeled timeseries since huge discrepancies do not cancel out each other.

Performance statistic	Optimum value	Feasible interval	Quoted in literature as	Units of measure
Mean Squared Error $$MSE = \frac{\sum_{i=1}^{n}(O_i - P)^2}{n}$$	0	$0 \leq MSE < +\infty$		Units of the variable to the square
Root Mean Squared Error $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P)^2}{n}}$$	0	$0 \leq RMSE < +\infty$	Daily Root Mean Squared Error (DRMS)	Same as variable Dimensionless if normalized with $\overline{O}$
Standard Deviation of residuals $$SDE = \sqrt{\frac{\sum_{i=1}^{n}\left[(O_i - P_i) - \mu_r\right]^2}{n}}$$	0	$0 \leq SDE < +\infty$		Same as variable Dimensionless if normalized with L_{range}
Standard Error $$SE = \sqrt{\frac{\sum_{i=1}^{n}(w_i)(O_i - P_i)^2}{n - m}}$$	0	$0 \leq MAE < +\infty$		Same as variable

Figure 4. Statistics measuring the average combined systematic and random modeling errors (after Vazquez, 2003; Vazquez *et al.*, 2008).

A second group of model performance statistics, that are measuring the average combined systematic and random discrepancies among simulations and observations, is shown in Figure 4, where μ_r is the BIAS (see Figure 3), w_e is a weighting factor associated to the observed data and m is the number of calibration parameters. Perhaps the most representative and common statistic of this group is the mean squared error (MSE). The root mean squared error (RMSE) is a measure of the overall spread of residuals with respect to its mean value. It is usually expressed with the same units as the variable under analysis. It is the maximum likelihood estimator under the assumption that measurement errors, approximated by $(O_{il} - P_i)$, are normally distributed with mean ($\mu_r = 0$) and constant variance (Bindley *et al.*, 1991).

Observe that the range of variation of the statistics listed in Figure 4 have no upper bound. Other statistics that belong to this group and that allow a more easy judgment of the model performance in the relative interval of variation [0, 1] are the Pearson's product-moment correlation coefficient (CC), also known as cross-correlation between two timeseries (Salas, 1993) and the Nash and Sutcliffe (1970) modeling efficiency (EF_2). CC is depicted in Eq. (1), the analysis of which reveals that $-1 < CC < 1$. The square of CC, termed as R^2, is commonly reported in literature. Its interval of variation is then $0 < R^2 < 1$. Thus, R^2 is a Pearson type correlation index.

$$CC = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2}} \qquad (1)$$

The mathematical formulation of EF_2 is shown in Eq. (2). Its range of variation is $-\infty < EF_2 \leq 1$, with a meaningful variation in the period $0 < EF_2 < 1$. Equation (2) shows that EF_2 is directly proportional

to MSE or RMSE. Hence, $EF_2 = 1$ when MSE = 0 that is for perfect conditions. $0 < EF_2 < 1$ when MSE $< \sigma^2_O$, that is for acceptable modeling conditions, with σ^2_O the variance of the observed data. If $EF_2 = 0$ when MSE $= \sigma^2_O$, the O (baseline or benchmark) is as good predictor as the model. Care should be exerted when referring to either R^2 or EF_2, since in several scientific publications the same quotation is used to refer to either of them (Vázquez, 2003).

$$EF_2 = 1.0 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} = 1.0 - \frac{MSE}{\sigma^2_O} = 1.0 - \frac{(RMSE)^2}{\sigma^2_O} \tag{2}$$

3.2. *Examining the congruence and usefulness of common model performance indexes*

Besides their statistical analysis (depicted in the previous section), some of the most common model performance statistics have been used in the context of some hydrologic modeling, so as to examine their congruence and usefulness. In this context, the modeling analysis considers some synthetic and real world modeling (for the Gete catchment, Belgium) and applied the performance indexes to evaluate on the goodness of fit of the model predictions with regard to the observations.

The modeling analyses herein include traditional model calibration through tests such as the Split-Sample (Refsgaard, 1997) as well as non trivial calibration tests based on Monte Carlo simulations (Vázquez *et al.*, 2009). In this context, it should be noticed that model performance statistics can be used in the scope of any statistical analysis to characterize the quality of the model predictions.

4. RESULTS AND DISCUSSION

Legates and McCabe (1999) quoted that correlation-based measures such as the R^2 are insensitive to the additive and proportional differences between the observed and predicted timeseries. For instance large values of R^2 can be obtained even when the simulated timeseries departs significantly from the observed one. It is illustrated in Figure 5 by considering a synthetic timeseries that represents the model prediction. This synthetic timeseries was prepared on the basis of a given timeseries of observations by considering a linear relationship $P_i = (A) O_i + B$, linking the predicted (i.e. synthetic) to the observed timeseries by means of a proportional factor $A \neq 0$ influencing the variability of the predicted series and an additive term $-\infty < B < +\infty$ influencing the magnitude of the predicted series. The following conditions are depicted in Figure 5: (a) better prediction of peak flows; (b) the timeseries are completely parallel to each other; and (c) better prediction of base flows.

The scatter plots indicate that the timeseries are markedly correlated to each other (through the linear relationship between both timeseries), but they do not match to each other; that is the correlation line has a different inclination than the 45° inclination of the one-to-one line or equal value line that is the centre of gravity of the scatter plot (Vázquez, 2003). In all the illustrated cases, the value of R^2 equals 1.0, reflecting the perfect correlation of both timeseries, despite the strong discrepancies that are however highlighted by the EF_2 index. Particularly, for the case depicted by Figure 5b, corresponding to parallel timeseries, the EF_2 value is significantly low, whilst the R^2 value is still 1. Thus, the EF_2 index represents an improvement over the R^2 in that EF_2 is sensitive to additive and proportional links between observed and simulated timeseries and less sensitive than R^2 and other correlation-based statistics to extreme values (Legates and McCabe, 1999).

$$d_2 = 1.0 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (|P_i - \overline{O}| + |O_i - \overline{O}|)^2} \tag{3}$$
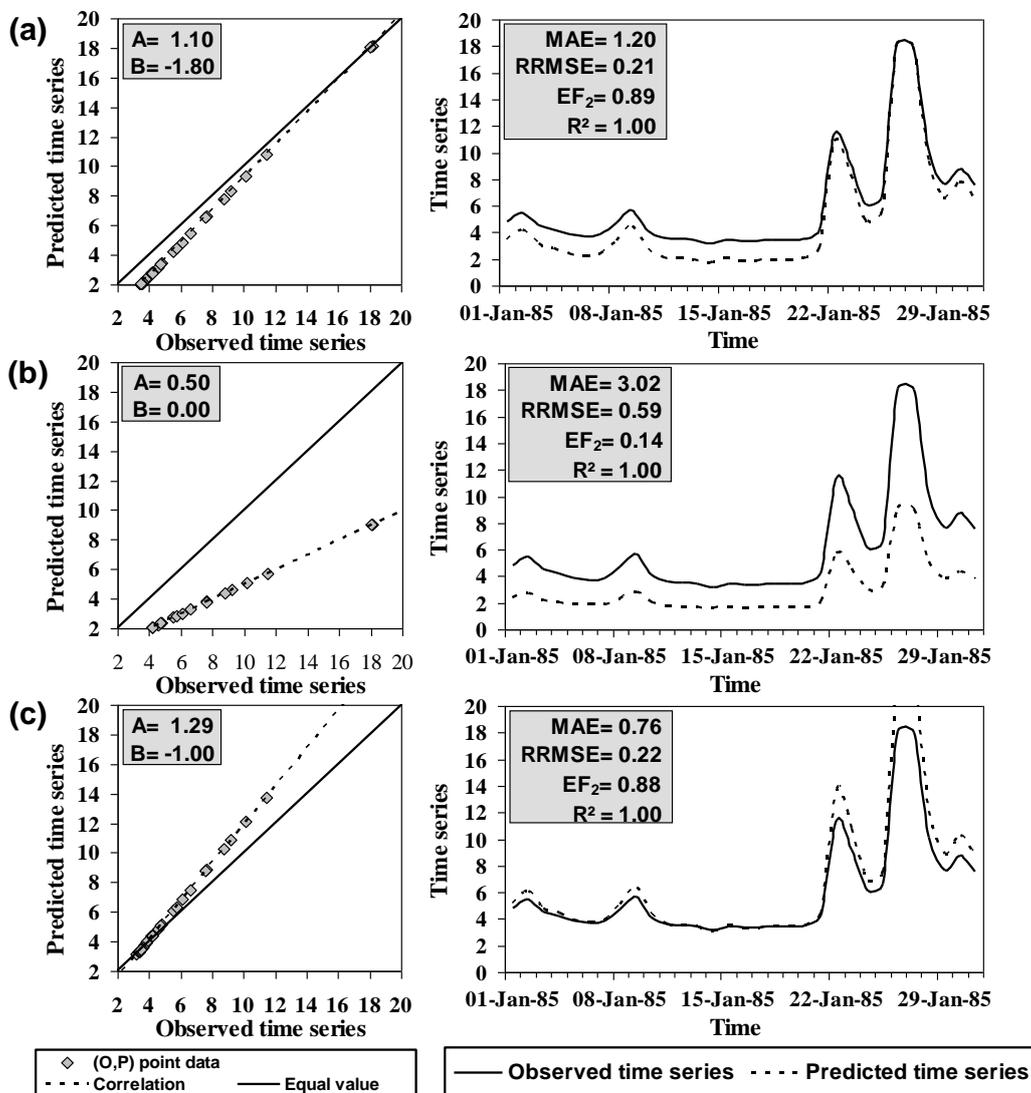

Figure 5. Illustration of the insensitivity of the R²-index to measure linear departures of predicted timeseries from historical time data (after Vázquez, 2003; Vázquez *et al.*, 2002). The units of measure of both timeseries are irrelevant for the current illustrative scope and as such are not indicated in the figure.

The oversensitivity of the R^2 index to extreme (peak) values is given by the amplification effect produced by the squared differences in Eq. (1). This amplification effect should be also associated to EF_2 as suggested by Eq. (2), although EF_2 seems to be less affected by this oversensitivity (than R^2). To overcome this oversensitivity of EF_2 to peak values a similar index (d_2) could be used (Legates and McCabe, 1999).

The feasible interval of the d_2 index is only between 0 and 1 as it can be only positive. The denominator in Eq. (3) is the Potential Error (PE). It represents the largest value that $(O_i - P_i)^2$ can adopt for each pair of observed and simulated points, as every expression within the absolute value operators is a measure of distance from the observed mean). Despite the different formulation of the d_2 index, the amplification effect produced by the squared differences is still present in Eq. (3).

It should be noticed that the interpretation associated to the above indexes is different from each other and is indeed much different from the interpretation associated to the R^2 index. The value of the R^2 index can be directly associated to the percentage of the variance in the observed data explained by the model. However, it is not a trivial matter either to associate a variance percentage to the EF^2 and d^2 values or to adopt threshold values for them to assess model simulations as acceptable. Generally, researchers select arbitrarily threshold values for judging model results as acceptable. Thus, important shortcomings of the EF_2 and the d_2 indexes are their sensitivity to extreme values, and their associated difficulty for deciding what values should be used as thresholds to identify appropriate model performances.

An alternative approach to reduce the oversensitivity of the EF_2 index to the peak values is by considering suitable transformations of the simulated variable. In this context, Figure 6 illustrates the variation of results when considering variable transformations to calculate the EF_2 index. The figure shows the results from a Monte Carlo simulation based analysis, for one of the hydrogeological parameters included in the modeling of the Gete catchment. Figure 6a corresponds to the traditional evaluation of the EF_2 index; Figure 6b depicts the EF_2 distribution when a logarithmic transformation of the predicted variable has been considered (EF_2^{ln}); whilst Figure 6c shows the distribution of a statistic calculated as a multiplicative combination of the prior statistics (EF_2^{Comb}; Vázquez and Feyen, 2008).

All the patterns exhibited in the figure are similar. However, plot b indicates that considering the logarithmic transformation of the discharges (alone) reduced the number of behavioral simulations, since it accentuated better simulations of low flows in comparison to the analysis without transformation (plot a). Furthermore, plot c shows that the multiplicative combination of indexes (EF_2^{Comb}) gives lower model performance values than the individual consideration of either model performance measure EF_2 and EF_2^{ln} (or an additive combination of them), since it requires both model performance measures to be high for depicting a good model performance.
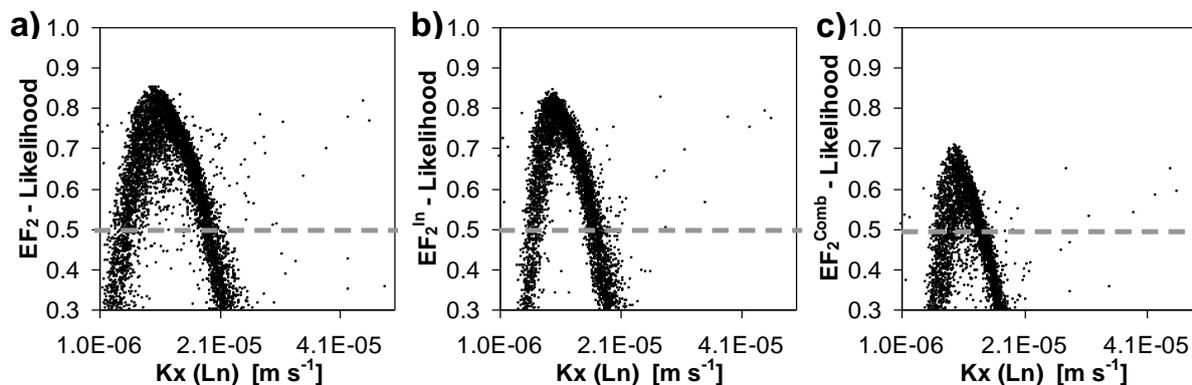

Figure 6. Scatter plots of likelihood measures versus values of a model parameter unit (after Vázquez *et al.*, 2009). The likelihood measure is the EF_2 index calculated (a) without any variable transformation; (b) considering a logarithmic transformation of the simulated variable; and (c) combining both prior statistics.

5. CONCLUSIONS

The analysis of the most commonly used measures of hydrological model performance was carried out by means of their statistical examination and illustrative modeling applications. In doing so, the model performance indexes were classified in two groups, according to the type of error (or residual) that those indexes are measuring.

The study revealed some relationships among statistics that, at a first sight (and indeed have been used in scientific literature assuming so), are measuring different properties from the population of

residuals (i.e. differences between the predicted and observed timeseries). In this context, the multi-criteria analysis of model performance should be based on the use of a set of statistics (or indexes) that are unrelated, so that different types of errors are measured, avoiding as such multiplicity. On the basis of the presented analysis, the reader can select a set of statistics depicting different information from the population of model residuals. It is recommended that, in general, the selected set of model performance statistics should be used in combination with other appropriate model performance tools, such as, visual inspection of auxiliary graphical information like scatter plots and cumulative hydrographs of the model outputs, or appropriate evaluation tests, differing from the traditional split-sample test for calibrating hydrologic models.

The numerical study revealed that model performance statistics are useful not only for model calibration and evaluation but also in the context of uncertainty analysis, that is, for defining model prediction limits. Further, model performance statistics can be very useful as well to validate modeling tools that are under development, either by comparing the model predictions to analytic solutions of a given problem or by comparison to the predictions of previously developed modeling tools.

REFERENCES

Abbaspour, K.C., M.T. van Genuchten, R. Schulin, E. Schläppi, 1997. A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters. *Water Resour. Res.*, 33(8), 1879-1892.

Abbott, M.B., J.C. Bathurst, P.E. Cunge, P.E., O'Connell, J. Rasmussen, 1986. An introduction to the European Hydrological System Système Hydrologique Européen, 'SHE', 1. History and philosophy of a physically based distributed modeling system. *J. Hydrol.*, 87, 45-59.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resourc.*, 16, 41-51.

Beven, K.J., 1997. TOPMODEL: A critique. *Hydrol. Process.*, 11(9), 1069-1085.

Binley, A.M., K.J. Beven, A. Calver, L.G. Watts, 1991. Changing Responses in Hydrology: Assessing the Uncertainty in Physically Based Model Predictions. *Water Resour. Res.*, 27(6), 1253-1261.

Braud, I., P. Fernandez, F. Bouraoui, 1999. Study of the rainfall-runoff process in the Andes region using a continuous distributed model. *J. Hydrol.*, 216(3-4), 155-171.

Buytaert, W., B. De Bièvre, G. Wyseure, J. Deckers, 2004. The use of the linear reservoir concept to quantify the impact of land use changes on the hydrology of catchments in the Ecuadorian Andes. *Hydrol. Earth Syst. Sc.*, 8, 108-114.

Buytaert, W., V. Iñegues, R. Célleri, B. De Bièvre, 2006. The impact of pine plantations on water yield: a case study from the Ecuadorian Andes. *3rd Int. Symp. on Int. Water Manage.*, *Bochum, Germany*.

Feyen, L., R.F. Vazquez, K., Christiaens, O. Sels, J. Feyen, 2000. Application of a distributed physically-based hydrological model to a medium size catchment. *Hydrol. Earth Sys. Sci.*, 4(1), 47-63.

Fleischbein, K., W. Wilcke, C. Valarezo, W. Zech, K. Knoblich, 2006. Water budgets of three small catchments under montane forest in Ecuador: experimental and modeling approach. *Hydrol. Process.*, 20, 2491-2507.

Freeze, R.A., R.L. Harlan, 1969. Blueprint for a physically based digitally simulated hydrologic response model. *J. Hydrol.*, 9, 237-258.

Grayson, R.B., I.D. Moore, T.A. McHanon, 1992. Physically-based hydrological modelling: a terrain-based model for investigative purposes. *Water Resour. Res.*, 28(10), 2639-2658.

Legates, D.R., G.J. McCabe, 1999. Evaluating the use of 'goodness-of-fit' measures in hydrological and hydroclimatic model validation. *Water Resour. Res.*, 35(1), 233-241.

Loague, K., R.E. Green, 1991. Statistical and graphical methods for evaluating solute transport models: overview and applications. *J. Contam. Hydrol.*, 7, 51-73.

Nash, J.E., J.V. Sutcliffe, 1970. River flow forecasting through conceptual models. Part I: A discussion of principles. *J. Hydrol.*, 10(3), 282-290.

Perrin, J.L., C. Bouvier, J.L. Janeau, G. Menez, F. Cruz, 2001. Rainfall/runoff processes in a small peri-urban catchment in the Andes mountains. The Rumihurcu Quebrada, Quito (Ecuador). *Hydrol. Process.*, 15(5), 843-854.

Refsgaard, J.C. 1997. Parameterisation, calibration and validation of distributed hydrological models. *J. Hydrol.*, 198, 69-97.

Refsgaard, J.C., B. Storm, 1995. MIKE SHE. In Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. *Water Resources Publications, Colorado*, USA, 809-846.

Refsgaard, J.C., B. Storm, 1996. Construction, calibration and validation of hydrological models. In: Abbott, M.B. and Refsgaard , J.C. (Eds.), Distributed Hydrological Modelling, *Kluwer Academic, The Netherlands*, 41-54.

Salas, J.D., 1993. Analysis and modelling of hydrologic timeseries. In: Maidment, D.R. (Ed.), Handbook of hydrology, *McGraw-Hill Inc.*, USA, 19.1-19.71.

Vázquez, R.F., 2003. Assessment of the performance of physically based distributed codes simulating medium size hydrological systems. *Tesis Ph.D.* ISBN 90-5682-416-3, *K.U.Leuven*, Bélgica, 335 págs.

Vázquez, R.F., J. Feyen, 2007. Assessment of the effects of DEM gridding on the predictions of basin runoff using MIKE SHE and a modelling resolution of 600 m. *J. Hydrol.*, 334, 73- 87.

Vázquez, R.F., J. Feyen, 2008. Application of distributed hydrologic models. En: Pilar García-Navarro, P., E. Playán (Eds.): Numerical modelling of Hydrodynamics for Water Resources. *Taylor & Francis*, Londres, Reino Unido, 153-174.

Vázquez, R.F., P. Willems, J. Feyen, 2008. Improving the predictions of a MIKE SHE catchment-scale application by using multi-criteria approach. *Hydrol. Process.*, 22(13), 2159-2179.

Vázquez, R.F., K. Beven, J. Feyen, 2009. GLUE based assessment on the overall predictions of a MIKE SHE application. *Water Resour. Manag.*, 23(7), 1325-1349.

Vázquez, R.F., L. Feyen, J. Feyen, J.C. Refsgaard, 2002. Effect of grid-size on effective parameters and model performance of the MIKE SHE code applied to a medium sized catchment. *Hydrol. Process.*, 16(2), 355-372.