Water Supply

© 2022 The Authors

Water Supply Vol 22 No 1, 589 doi: 10.2166/ws.2021.245

A new innovative method for model efficiency performance

Zekâi Şen [a,b], Eyüp Şişman[a,b,*] and Burak Kızılöz[c]

[a] School of Engineering and Natural Sciences, Civil Engineering Department, Istanbul Medipol University, Kavacık, Istanbul 34181, Turkey
[b] Istanbul Medipol University, Climate Change Researches Application and Research Center, (IKLIMER) Kavacık, Istanbul 34181, Turkey
[c] Kocaeli Metropolitan Municipality Water and Sewerage Administration, İzmit, Kocaeli, Turkey
*Corresponding author. E-mail: esisman@medipol.edu.tr

ZŞ, 0000-0003-3696-9967

ABSTRACT

In every aspect of scientific research, model predictions need calibration and validation as their representativity of the record measurement. In the literature, there are a myriad of formulations, empirical expressions, algorithms and software for model efficiency assessment. In general, model predictions are curve fitting procedures with a set of assumptions that are not cared for sensitively in many studies, but only a single value comparison between the measurements and predictions is taken into consideration, and then the researcher makes the decision as for the model efficiency. Among the classical statistical efficiency formulations, the most widely used ones are bias (BI), mean square error (MSE), correlation coefficient (CC) and Nash-Sutcliffe efficiency (NSE) procedures, all of which are embedded within the visual inspection and numerical analysis (VINAM) square graph as measurements versus predictions scatter diagram. The VINAM provides a set of verbal interpretations and then numerical improvements embracing all the previous statistical efficiency formulations. The fundamental criterion in the VINAM is 1:1 (45°) main diagonal along which all visual, science philosophical, logical, rational and mathematical procedures boil down for model validation. The application of the VINAM approach is presented for artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) model predictions.

Key words: efficiency, error, models, square graph, validation, visual

HIGHLIGHTS

- Objective assessment of model efficiency is presented by means of a new approach named visual inspection and numerical analysis method (VINAM).
- The visual inspection and validation are possible by means of square template.
- The VINAM provides a set of verbal interpretations and subsequent significantly numerical improvements.
- The proposed VINAM methodology improves all suggested model efficiency metrics.

LIST OF ABBREVIATIONS

The following symbols are used in this paper:

Acronym	Definition
ANFIS	Adaptive network-based fuzzy inference system
ASCE	American Society of Civil Engineering
ANN	Artificial neural network
BI	Bias
R^2	Coefficient of determination
CC	Correlation coefficient
DWST	Domestic water storage tank
FR	Failure Ratio
d	Index of agreement
KGE	Kling-Gupta efficiency
MSE	Mean square error
MPD	Mean pipe diameter

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

MIMO	Multi-input multi-output
NSE	Nash-Sutcliffe efficiency
NL	Network length
NRW	Non-revenue water
NSCF	Number of service connection failure
NJ	Number of junctions
NNF	Number of network failure
PBI	Percent Bias
PME	Persistence model efficiency
RSR	RMSE-observation standard deviation ratio
RMSE	Root mean square error
VINAM	Visual inspection and numerical analysis
WDQ	Water demand quantity
WM	Water meter

INTRODUCTION

Models are the reflection tools of the reality for simulation, prediction, automation and optimum management studies at the service to men, and they are required to produce outputs as close as possible to the measurements in an efficient manner. Whatever are the model types (analytical, probabilistic, statistical, stochastic or numerical) in practical studies, there are two sequences for comparison as the measurement series and corresponding model prediction series.

In general, the model predictions are related to measurements through a curve fitting methodology based on the least square analysis of some type. There are also other versions including empirical relationships, stochastic and more complex numerical solution algorithms. All these techniques have a visual basis, which can be appreciated by means of shapes in the forms of mathematical functions, flow charts, geometry, algorithms, and block diagrams. Any idea based on a geometrical shape provides visual inspections, examinations and inference deductions, perhaps at early stages verbally, but such statements can be converted to mathematical expressions after understanding the science philosophical, logical and rational fundamentals. Human philosophical thinking and logical rational trimming of blurted ideas lead to a set of logical rule bases, which are precedencies of mathematical equations and expressions by a set of convenient symbols.

In scientific researches, one is interested in relating measurements to model prediction outputs, which may be in the forms of single-input single-output (SISO) or various versions as multi-input multi-output (MIMO) models. In the literature, there is a set of standard coefficients that provides agreement between measurements and predictions through a set of single parameter values. In most of the cases, authors report that their model is suitably based on the comparison of statistical parameters composed of measurement and model output data series by consideration of one or few of the well-established agreement or association metrics among which the most commonly used, accepted, and recommended ones are bias (BI), Percent Bias (PBI), coefficient of determination (R^2), mean square error (MSE) or root mean square error (RMSE), correlation coefficient (CC), Nash-Sutcliffe efficiency (NSE) and index of agreement (d) (Pearson 1895; Nash & Sutcliffe 1970; Willmott 1981; Santhi *et al.* 2001; Gupta *et al.* 2002; Moriasi *et al.* 2007; Van Liew *et al.* 2007; Özger & Kabataş 2015; Tian *et al.* 2015; Zhang *et al.* 2016; Dariane & Azimi 2018).

In the literature, there are also other versions such as the modified index of agreement (d1) (Legates & McCabe 1999), prediction efficiency (P_e) as explicated by Santhi *et al.* (2001), persistence model efficiency (PME) (Gupta *et al.* 2002), RMSE-observation standard deviation ratio (RSR) as given by Moriasi *et al.* (2007), and Kling-Gupta efficiency (KGE), measured by Gupta *et al.* (2009).

As stated by McCuen & Snyder (1975) and Willmott (1981) almost all the models have elusive predictions, which cannot be covered by the model efficiency measures and even, in general, by significance tests. Freedman *et al.* (1978) have mentioned that the statistical significance tests are concepts that must be viewed with skepticism. Along the same line, Willmott (1981) stated that it may be appropriate to test an agreement measure and report the value of a test statistic at a significance level, but the distinction between the significant and insignificant levels is completely unjustified. For instance, if the significance level is adapted as 0.05 then what are the differences, say, among 0.049, 0.048, 0.047 and 0.051, 0.052 and 0.053? Additionally, such significance levels depend on the number of data for depiction of the most suitable theoretical probability distribution function (PDF).

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

Even though ASCE (1993) accentuate the need to explicitly define model evaluation criteria, no widely accepted guidance has been established, but a few performance ratings and specific statistics have been used (Saleh *et al.* 2000; Santhi *et al.* 2001; Bracmort *et al.* 2006; Van Liew *et al.* 2007).

For a more objective assessment of model efficiency, calibration and validation, measurement association and comparison visual inspections must be preliminary conditions for better insights, interpretations and model modification possibilities. The basic statistical parameters such as arithmetic averages, standard deviations, and regression relationship between the measurement (independent variable) and model prediction (dependent variable) data through the scatter diagram are very important ingredients even for visual inspection to identify systematic and random components. Unfortunately, most often the model efficiency measure is obtained by available software, which does not provide any informative detailed visual inspection and assessment.

Even though there is extensive literature on model calibration and validation, it is difficult to compare the modeling results (Moriasi *et al.* 2012). Numerous models of calibration and validation have been the subject of discussion by scientists and experts (ASCE 1993; Van Der Keur *et al.* 2001; Li *et al.* 2009; Mutiti & Levy 2010; Palosuo *et al.* 2011; Moriasi *et al.* 2012; Zhang *et al.* 2012; Harmel *et al.* 2013; Ritter & Muñoz-Carpena 2013; Pfannerstill *et al.* 2014; Larabi *et al.* 2018; Rujner *et al.* 2018; Swathi *et al.* 2019).

The main purpose of this paper is to present visual inspection and numerical analysis (VINAM) methodology for effective model efficiency and ideal validation, and if necessary, modification or calibration of the model predictions to comply with the measurements. The visual inspection and validation are possible by means of a square template. It includes all basic information clearly in an objective manner first for verbal, science, philosophical, logical and rational inferences, which are then translatable to mathematical symbolic equations. The proposed VINAM methodology improves all suggested model efficiency metrics that are available in the literature.

METHODOLOGY

Statistical efficiency formulations

In the literature, all model efficiency standard indicators are dependent on three basic statistical parameters, among which are the arithmetic averages of the measurements, $\bar{M}$, and model predictions, $\bar{P}$; standard deviations, S_M and S_P, and the cross-correlation, $C_{M,P}$, between measurement and model prediction sequences. Additionally, the regression line between the measurements and predictions has two parameters as intercept, I, or regression line central point ($\bar{M}$ and $\bar{P}$) coordinates and the slope, S.

The simple and necessary, but not enough mathematical efficiency measure is the bias, BI, which measures the distance between the measurements and model predictions as:

$$BI = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i) = \bar{P} - \bar{M} \tag{1}$$

The ideal value for model efficiency is $BI = 0$; although this condition is necessary, but not enough. The second measure is the mean square error (MSE),

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2 \tag{2}$$

The ideal value is $MSE = 0$, but this condition is not valid in any hydro-meteorological model efficiency because there are always natural random errors. This is the main reason why the best model MSE should have the minimum level among all other alternatives. Equation (2) includes implicitly the standard deviations and the cross-correlation between the measurements and predictions.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

The Nash-Sutcliffe efficiency (NSE) measure includes the MSE with the standard deviation of the measurement data ratio as follows:

$$\mathrm{NSE} = \frac{\sum_{i=1}^{n}(M_i - \bar{M})^2 - \sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - \bar{M})^2} = 1 - \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - \bar{M})^2} \tag{3}$$

The second term on the r.h.s is greater than 1, hence NSE has negative values. The ideal value of NSE is 1, but this is never verified in practical applications, and therefore, the closer the value is to 1, the better is the model efficiency.

As for the cross-correlation, CC, between the measurement and prediction can be calculated as:

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(P_i - \bar{P})}{S_M S_P} \tag{4}$$

On the other hand, the straight-line regression intercept, I, and slope, S, values can be calculated according to the following expressions:

$$\bar{P} = S\bar{M} + I \tag{5}$$

and

$$S = \frac{n\sum_{i=1}^{n} P_i M_{i-} \sum_{i=1}^{n} P_i \sum_{i=1}^{n} M_i}{n\sum_{i=1}^{n} M_i^2 - \left(\sum_{i=1}^{n} M_i\right)^2} \tag{6}$$

respectively.

Apart from the above model efficiency measurements, there are others, which have been suggested for their rectification. One of the first versions is due to Willmott (1981), who gave agreement index d as:

$$d = 1 - \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(|P_i'| + |M_i'|)^2} \tag{7}$$

where $P_i' = P_i - \bar{P}$ and $M_i' = M_i - \bar{M}$ are the deviations from the respective arithmetic averages. The expression in the dominator is referred to as the potential error (PE). The significance of d is that it measures the degree to which model predictions are error free, and its values vary between 0 and 1; where 1 represents the perfect agreement between the measurements and predictions, which is never possible in practical applications, and therefore, the researchers take the closest value to $d = 1$ as the model efficiency acceptance, but there is no criterion that indicates objectively the limit value between acceptance and rejection, and hence, there appears subjectivity as in other efficiency measures. Equation (7) can be rewritten in terms of the MSE as follows.

$$d = 1 - \frac{n\mathrm{MSE}}{\mathrm{PE}} \tag{8}$$

Equations (1)–(8) include all the necessary numerical quantities that are useful in the construction of the VINAM template as will be explained and applied in the following sections.

Square graph for visual inspection and numerical analysis method (VINAM)

For visual inspection of the model predictions associated with the measurements, one can regard the measurements as independent variables, predictions as dependent variables and plot them on a coordinate system. In the case of an ideal match

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

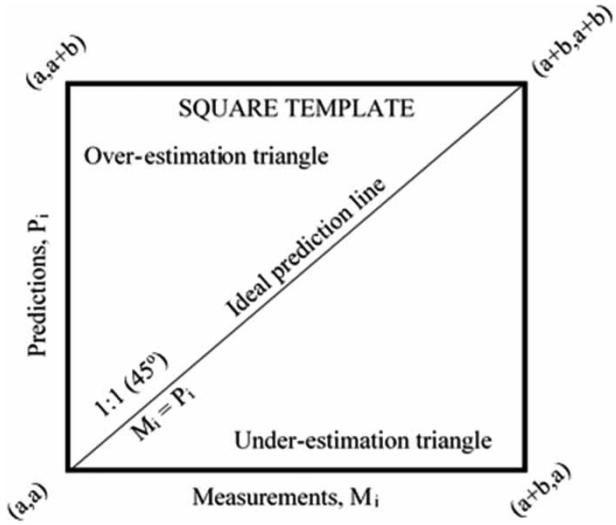


Figure 1 | Square template for VINAM.

between the two series, one expects that they fall on the 1:1 (45°) straight line, which appears as the diagonal straight-line on the square template graph as in Figure 1. This straight-line divides the square area into two half triangles with the upper (lower) one representing complete model over-estimation (under-estimation) domain provided that all scatter points fall completely in either of these triangles. It is also possible that the scatter points may have positions partially in each triangle, in which case the model has some points as over-estimation and others as under-estimation. The mathematical expression of the ideal model efficiency case as 1:1 line is:

$$P_i = M_i \tag{9}$$

In Figure 1, a and b coefficients correspond to the minimum and maximum values among the measurements and predictions.

The most significant feature of a square template is in its ability to reflect almost all the previously defined efficiency criteria properties in a single graph. For instance, in Figure 2(a), the scatter of points is shown in the upper triangular area (over-estimation), but there is no linear trend between the measurements and prediction scatter points, which are randomly distributed. This provides the message that the model is not capable of representing the measurements at all. It is necessary to try and model the measurements with another suitable model, which must yield at least some consistency among the scatter points.

For instance, in Figure 2(b), the scatter points have a linear tendency, which is the first indication that the model for predictions is suitable, because the scatters are around a regression line. The following features are the most important information pieces in this figure:

(1) The centroid point ($\bar{M}$ and $\bar{P}$) on the regression line is at a distance, D, from the ideal prediction line,
(2) The same straight-line has a slope, S, with the horizontal axis the value of which can be calculated from Equation (6),
(3) The model regression straight-line in the figure has an intercept, I, on the vertical axis. It also crosses from the centroid ($\bar{M}$ and $\bar{P}$) point.
(4) The straight-line passes through the centroid point which implies Equation (5),
(5) After the regression straight-line expression determination, one can calculate the vertical deviations of each scatter point from the ideal prediction line, which constitute the error sequence, ε_i.

Tian *et al.* (2015) suggested representing the straight line without any visual explanation as follows as (their Equation (10) with notations in this article):

$$P_i = SM_i + D + \varepsilon_i \tag{10}$$

which is exactly the reflection of the regression line in Figure 2(b). According to them, S is the scale error, and D is the constant or displacement error, but in this paper, they are referred to as the rotational error and shift error, respectively. It is

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

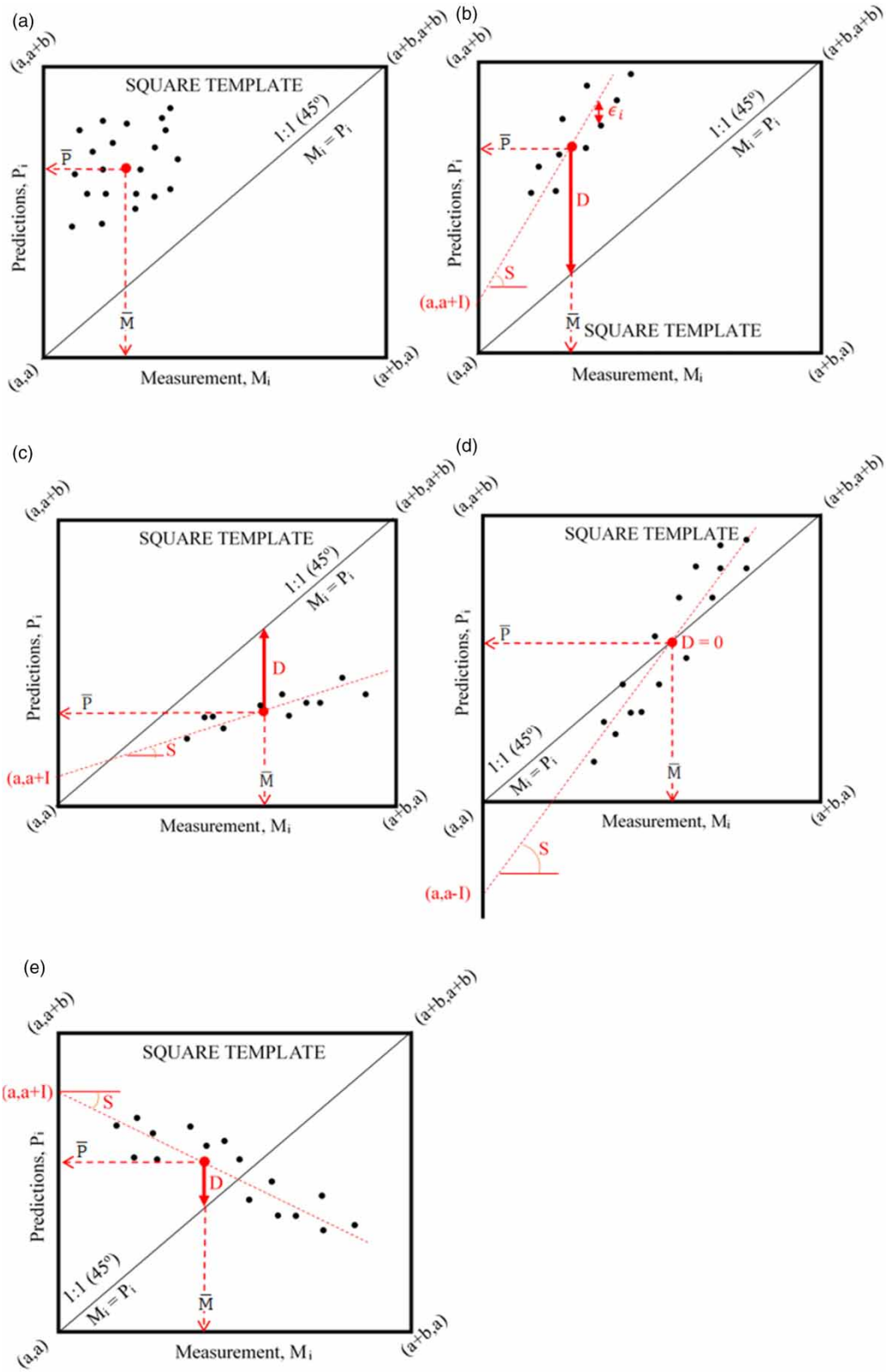


Figure 2 | Measurement and prediction scatter points in the various VINAM templates.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

obvious from Figure 2(b) that each of these are systematic deviations from the ideal prediction line and therefore, each is a systematic deviation, but their summation is total systematic error. They have significant duties, as will be explained in Section 4. No need to say, Figure 2(c) is the under-estimation alternative of Figure 2(b), and the same quantities are also available in this figure. Equation (10) is also valid for this case, but with opposite shift error.

Figure 2(d) and 2(e) is for the partial model over-estimation (under-estimation) case, where the central point of the regression line centroid coincides with the ideal prediction line ($D = 0$), and away from the ideal line, respectively. In the former case, there is no shift error, and for the other case everything is self-explanatory under the light of the above explanations.

According to the suggested template and algorithm, the measurement data is accepted as constant, and it is tried to systematically approach the predictions to these measurement data or to define the recalibration operation between the obtained model results and the measurement to obtain the best and optimum efficiency model. When the measurement data accepted as an independent variable is shown on the horizontal axis, and the prediction data that we can accept as a dependent variable on the vertical axis, the rotation and translation can be achieved by performing mathematical operations sequentially. In this way, a model prediction or calibration that is closer to actual measurement values is made by reducing systematic errors. In this case, a distortion occurs due to the change in the vertical distances with the ideal line as a result of these operations. This result is unavoidable to make better predictions. Thanks to the new approach in this study, the optimization of total vertical changes has been achieved by taking all available data into account. The positive contribution of the suggested method is seen by controlling the obtained results through the six different performance indicators.

Model modification

After all explanations in the previous section, an important question is, 'is it possible to improve the model performance, and how to increase its efficiency?'. The best and optimum efficiency is possible after shift and rotation operations on the VINAM template regression line. The following steps are necessary for arriving at the best solution:

(1) Shifting operation of the central regression point vertically such that it sits on the ideal prediction line (1:1). Only vertical shifts are possible for keeping the measurements as they are,
(2) After the shifting, the regression line is rotated according to the rotation angle as (1-S), so that the regression line coincides with the ideal prediction line (1:1),
(3) These two operations are preferable if there is no other choice to get the VINAM regression line to coincide with the ideal prediction line.

In shifting operation, there is no problem, because the whole scatter points are moved by the amount of D downwards or upwards. The shifting operation mathematical expression is:

$$P_i' = P_i \pm D \tag{11}$$

As for the rotation operation, the horizontal locations of each scatter point must remain the same, so as not to disturb measurement values. Such a rotation can be achieved by means of the following expression where P_i'' is for final data:

$$P_i'' = P_i - I - SM_i \tag{12}$$

The method suggested in this study aims to improve the model performances by reducing the differences between the systematic errors in the model prediction results and the predictive measurements. When the accuracy and reliability levels are analyzed, it is seen that there are systematic and random errors between the predictions and measurements. These errors vary depending on certain factors such as the experience of the modeler, the data quality, and the selected methodology. The error evaluation can be made according to the ideal line given in the Square template described in Figure 1, which is frequently preferred in studies, in addition to various performance indicators. When the first results of model studies are evaluated, it is seen that different alternatives may arise (Figure 2). The vertical differences of the prediction results are composed of the consistent difference between the mean values of the measurement and prediction, the angle between the 1:1 ideal line is expected to be between the model and the prediction, and the regression line is obtained according to the least-squares between the model and the prediction, and finally random differences. These three components showing the quality, accuracy, and reliability of the model vary depending on the established model, the used data, and the selected method, and so

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

on. The model performances depending on the first two components can be improved significantly through the suggested method. Therefore, the two important steps including the shift and rotation operations were described in this study. The model design needs to be revised to improve the model performance by developing the third component prediction.

RESULTS AND DISCUSSION

In the Appendix-A, the necessary software is given for the application of all VINAM steps. The applications of the VINAM procedure are presented for two well-known models, which are the artificial neural network (ANN) and adaptive network based fuzzy inference system (ANFIS). These applications are based on the water losses measurement in potable water distribution systems, for which water loss predictions are among the most important issues of water stress control (Şişman & Kizilöz 2020a). The most important component in the evaluation of a water distribution system with regards to water losses is the non-revenue water (Kanakoudis & Muhammetoglu 2014; Boztaş *et al.* 2019; Şişman & Kizilöz 2020b; Kizilöz & Şişman 2021). Jang & Choi (2017) built a model to calculate the NRW ratio of Incheon, Republic of Korea, by means of ANN methodology. When the best model was examined, R^2 was obtained as 0.397. It is seen that the models can be improved when the measurement values and model projections scatter plots appear along a regression line, as already explained in Figure 2.

The NRW ratio estimates are modeled through ANN and ANFIS for Kocaeli district, Turkey, and the implementation of the suggested VINAM method is carried out on similar model outputs for further improvements.

A total of eight models (four ANN and four ANFIS) with nine input measurements are developed through the modeling procedures. Water demand quantity, domestic water storage tank, number of network failure, number of service connection failure and failure ratio, network length, water meter, number of junctions, and mean pipe diameter are the model input parameters. All models are validated by the VINAM approach, and the model efficiency evaluations are carried out through statistical values according to BI, MSE, CC, R^2, d and NSE.

For ANN model performance, 55% of the available data is arranged as training, 35% as validation and 10% as testing. These models are developed with one hidden layer including four neurons and feed forward back propagation training procedure with support of the Levenberg-Marquardt back propagation algorithm (Coulibaly *et al.* 2000; Kermani *et al.* 2005; Kızılöz *et al.* 2015; Rahman *et al.* 2019; Şişman & Kizilöz 2020a, 2020b).

As for the ANFIS model implementation, 66% of the obtained data are taken as training and the remaining 34% for validation (testing) purposes. For this model, various membership functions (MFs) are considered as triangular (Trimf), Gaussian bell-shaped (gbellmf) and trapezium (trammf) with 'low', 'medium', and 'high' linguistic terms. The statistical properties of input components and model outputs are given in Table 1 for ANN and ANFIS models.

The resultant VINAM graphs are presented in Figure 3 for ANN model versions, with the model's efficiency classical and VINAM improvements in Table 2.

In this study, the NRW prediction rate of the selected model was calculated over nine different parameters through the ANN and ANFIS methodologies. The performance indicator results of the NRW rate predictions, which are made through three different combinations of input parameters given in Table 2, are available and it seems that the model results are not at the desired level. On the other hand, when the NRW rate predictions are analyzed through the Square template described in Figure 1, it is seen that the combinations determined by certain systematic errors can make good predictions. So, a

Table 1 | Input-output parameters

	Model parameters		Range	Unit
Input	Water demand quantity	WDQ	315.445–2.844.526	m^3
	Domestic water storage tank (Şişman & Kizilöz 2020a)	DWST	4000–85.901	m^3
	Number of network failure	NNF	34–628	Number
	Number of service connection failure	NSCF	23–541	Number
	Failure Ratio	FR	0.01–3.43	–
	Network length (Şişman & Kizilöz 2020a)	NL	306–1600	km
	Water meter (Şişman & Kizilöz 2020a)	WM	15.124–160.135	Number
	Number of junctions (Şişman & Kizilöz 2020a)	NJ	9.616–52.565	Number
	Mean pipe diameter (Şişman & Kizilöz 2020a)	MPD	108–159	mm
Output	Non-revenue water ratio	NRW	0.13–0.54	–

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

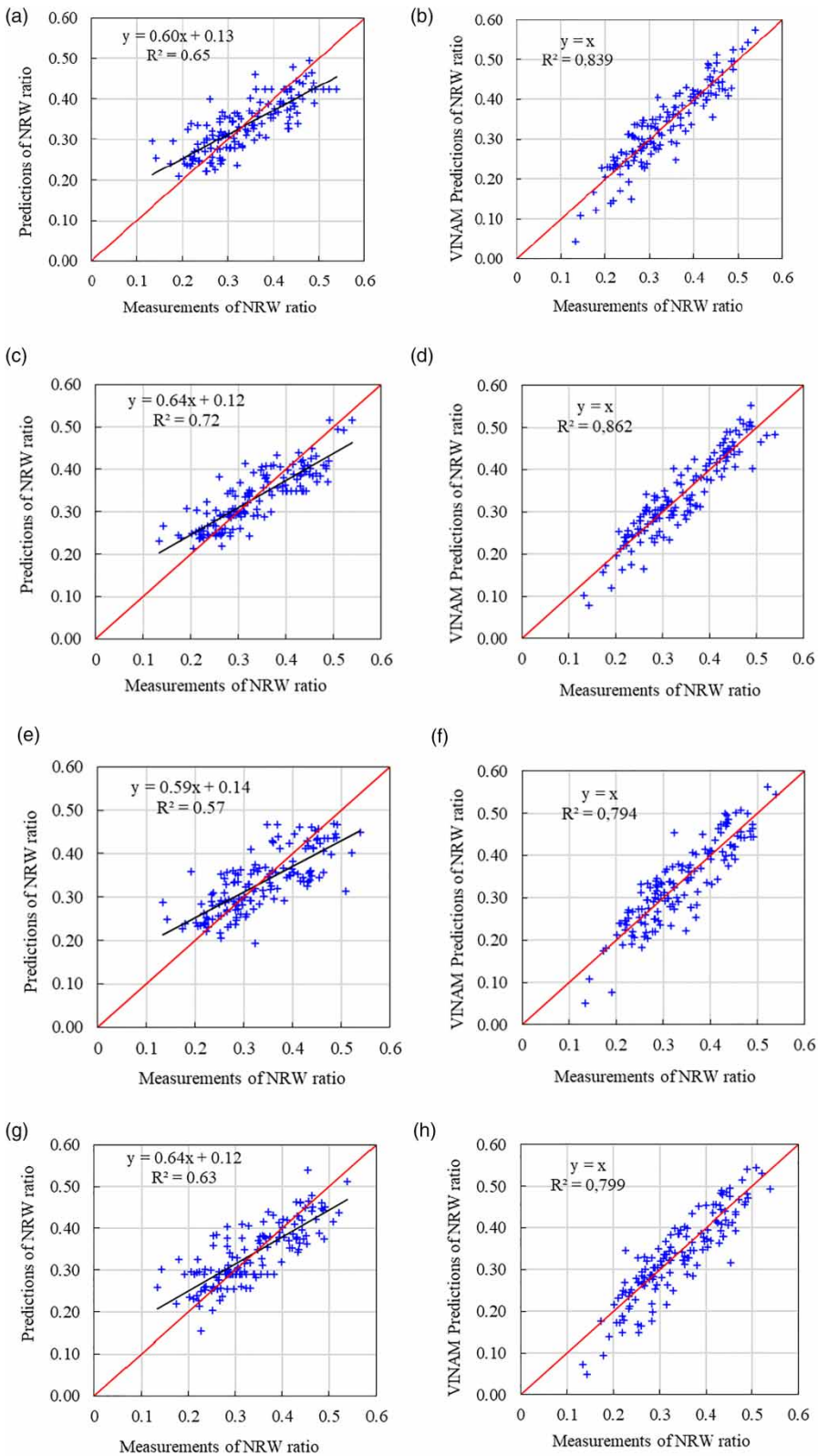


Figure 3 | ANN classical (a, c, e, g) and VINAM (b, d, f, h) approach.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

Table 2 | ANN and VINAM ANN models result

Model No	Input combinations	R^2	MSE	NSE	BI	CC	d
ANN 1	WDQ – MPD – NNF	0.65	0.0028	0.648	−0.0014	0.808	0.874
ANN 2	WDQ – NL – NNF	0.72	0.0022	0.715	−0.0024	0.851	0.901
ANN 3	WDQ – WM – FR	0.57	0.0034	0.572	−0.0022	0.757	0.851
ANN 4	WDQ – DWST – FR	0.63	0.0290	0.625	0.0018	0.791	0.877
VINAM_ANN 1	WDQ – MPD – NNF	0.84	0.0015	0.809	−0.0004	0.916	0.955
VINAM_ANN 2	WDQ – NL – NNF	0.86	0.0013	0.839	−0.0004	0.928	0.962
VINAM_ANN 3	WDQ – WM – FR	0.79	0.0020	0.740	−0.0004	0.891	0.940
VINAM_ANN 4	WDQ – DWST – FR	0.80	0.0020	0.745	−0.0049	0.894	0.941

considerable improvement has been achieved by calibrating the models (according to the ideal line) over the classical approaches through the suggested methodology. It is possible to predict the NRW rates with specific levels that can be accepted with only three parameters, and evaluate the network losses over three parameters such as WDQ, WM, FR.

The second application graph links to the measurements and ANFIS models VINAM diagram with Figure 4. The classical efficiency and VINAM improvements are available in Table 3.

The NRW rate, which is predicted through certain parameters in this study, actually varies depending on many variables and it is also affected by many uncertainties in the water distribution infrastructure. Leaks that cause physical losses in the system, number of failures, network age, network pressure, meter ages that cause apparent losses, meter measurement errors, illegal water use, and so on show the importance of these uncertainties. The effects of these uncertainties and their management are more important, especially for the administrations with high NRW rates. Developing a model in which all uncertainties that cause increases in NRW rates are evaluated together is possible with significant time and cost resources. Since the parameter changes in the water distribution system are partially related to the aforementioned issues, the water distribution system management can be carried out by these parameters by predicting the models through the suggested methodology with many fewer parameters. Evaluating the NRW rate performances through the parameter combinations provides much faster and economical solutions for systems. In a conclusion, an improvement in predictive power could help designing investment incentives more effectively and in a more targeted way, and therefore determining the effects of the parameters for which the best predictions are made on the NRW rate makes the revision of field practices on investment programs possible with the advantages of the effective models suggested in this study.

CONCLUSIONS

The existing model efficiency criteria have statistical mathematical expressions, which yield a single value about the association between the measurement and prediction sequences. Accordingly, the researcher may adopt with subjective acceptance one of the classical efficiency metrics, because the closer the efficiency measure value to the ideal value, the better the model representativeness of the measurement data. In these criteria assessments, single significant tests are inappropriate in many cases because they do not provide preliminary visual information. Rather than depending on such expressions without visual impressions, this paper presents an effective model efficiency evaluation methodology by means of the visual inspection and numerical analysis (VINAM) square template concept. It provides visualization of all the methodological details first by eye for science philosophical, logical and rational inferences, which lead to the fundamentals of the mathematical model efficiency expressions explicitly. The main ideal is to assess the scatter plot diagram between measurement and model predictions. In the case of random scatter, the model is not suitable at all. On the other hand, if the scatter points appear along an acceptable regression line on the VINAM square template, then by means of the shift and rotation procedures the scatter points can be formed around the 1:1 (45°) straight-line with improvements in the classically available statistical model efficiency results. The application of the VINAM procedure is checked with the artificial neural network (ANN) and adaptive network based fuzzy inference system (ANFIS) models. It is observed that the VINAM method improves all the cases with very significant percentages.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

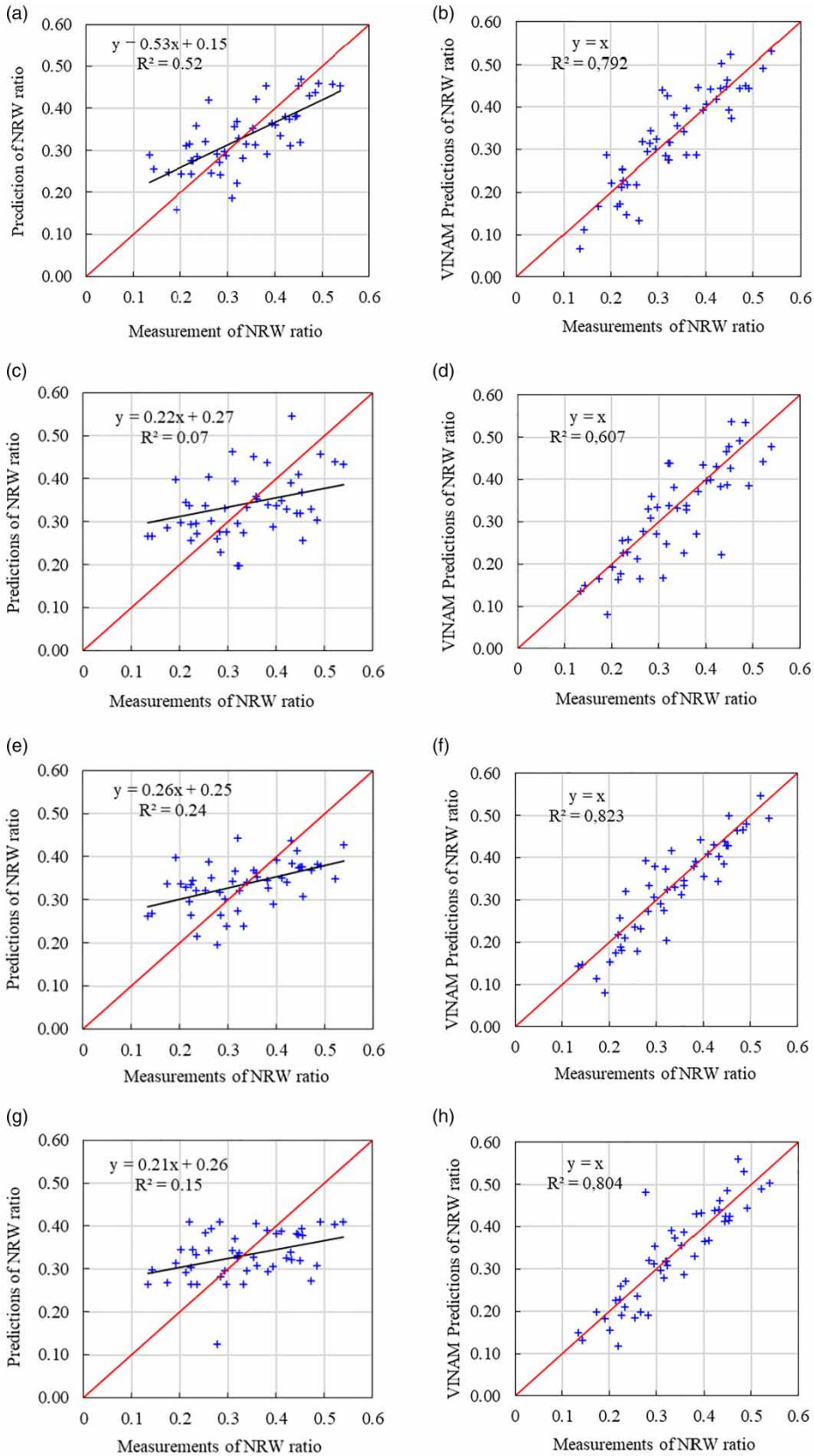


Figure 4 | ANFIS classical (a, c, e, g) and VINAM (b, d, f, h) approach.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

Table 3 | ANFIS models with three inputs

Model No	Input combinations	R^2	MSE	NSE	BI	CC	d
ANFIS 1	WDQ – MPD – NNF	0.52	0.0051	0.523	– 0.0014	0.724	0.825
ANFIS 2	WDQ – NL – NNF	0.07	0.0138	– 0.292	0.0086	0.258	0.536
ANFIS 3	WDQ – WM – FR	0.24	0.0081	0.238	0.0031	0.491	0.621
ANFIS 4	WDQ – DWST – FR	0.15	0.0093	0.123	– 0.0011	0.384	0.575
VINAM_ANFIS 1	WDQ – MPD – NNF	0.79	0.0028	0.738	– 0.0012	0.89	0.939
VINAM_ANFIS 2	WDQ – NL – NNF	0.607	0.0075	0.296	– 0.0224	0.779	0.856
VINAM_ANFIS 3	WDQ – WM – FR	0.823	0.0024	0.776	– 0.0095	0.907	0.948
VINAM_ANFIS 4	WDQ – DWST – FR	0.804	0.0026	0.756	– 0.002	0.897	0.944

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

ASCE 1993 Criteria for evaluation of watershed models (definition of criteria for evaluation of watershed models of the watershed management committee, irrigation and drainage division). *Journal of Irrigation and Drainage Engineering* **119** (3), 429–442.

Boztaş, F., Özdemir, Ö., Durmuşçelebi, F. M. & Firat, M. 2019 Analyzing the effect of the unreported leakages in service connections of water distribution networks on non-revenue water. *International Journal of Environmental Science and Technology* **16** (8), 4393–4406.

Bracmort, K. S., Arabi, M., Frankenberger, J. R., Engel, B. A. & Arnold, J. G. 2006 Modeling long-term water quality impact of structural BMPs. *Transactions of the ASABE* **49** (2), 367–374.

Coulibaly, P., Anctil, F. & Bobée, B. 2000 Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology* **230** (3–4), 244–257.

Dariane, A. B. & Azimi, S. 2018 Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection. *Journal of Hydroinformatics* **20** (2), 520–532.

Freedman, D., Purves, R. & Pisani, R. 1978 *Statistics*. W.W. Norton & Co, New York.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. 2002 Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrologic Engineering* **4** (2), 135–143.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91.

Harmel, R. D., Smith, P. K. & Migliaccio, K. W. 2013 Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. *Transactions of the ASABE* **53** (1), 55–63.

Jang, D. & Choi, G. 2017 Estimation of non-revenue water ratio for sustainable management using artificial neural network and Z-score in Incheon, Republic of Korea. *Sustainability* **9** (11), 1933.

Kanakoudis, V. & Muhammetoglu, H. 2014 Urban water pipe networks management towards non-revenue water reduction: two case studies from Greece and Turkey. *Clean – Soil, Air, Water* **42** (7), 880–892.

Kermani, B. G., Schiffman, S. S. & Nagle, H. T. 2005 Performance of the Levenberg-Marquardt neural network training method in electronic nose applications. *Sensors and Actuators, B: Chemical* **110** (1), 13–22.

Kizilöz, B. & Şişman, E. 2021 Exceedance probabilities of non-revenue water and performance analysis. *International Journal of Environmental Science and Technology* 1–12. https://doi.org/10.1007/s13762-020-03018-y.

Kizilöz, B., Çevik, E. & Aydoğan, B. 2015 Estimation of scour around submarine pipelines with artificial neural network. *Applied Ocean Research* **51**, 241–251.

Larabi, S., St-Hilaire, A., Chebana, F. & Latraverse, M. 2018 Using functional data analysis to calibrate and evaluate hydrological model performance. *Journal of Hydrologic Engineering* **23** (7), 1–12.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35** (1), 233–241.

Li, Z., Liu, W., Zhang, X. & Zheng, F. 2009 Impacts of land use change and climate variability on hydrology in an agricultural catchment on the Loess Plateau of China. *Journal of Hydrology* **377** (1–2), 35–42.

McCuen, R. H. & Snyder, W. M. 1975 A proposed index for comparing hydrographs. *Water Resources Research* **11** (6), 1021–1024.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50** (3), 885–900.

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest

Moriasi, D. N., Wilson, B. N., Douglas-Mankin, K. R., Arnold, J. G. & Gowda, P. H. 2012 Hydrologic and water quality models: use, calibration, and validation. *Transactions of the ASABE* **55** (4), 1241–1247.

Mutiti, S. & Levy, J. 2010 Using temperature modeling to investigate the temporal variability of riverbed hydraulic conductivity during storm events. *Journal of Hydrology* **388** (3–4), 321–334.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I - a discussion of principles. *Journal of Hydrology* **10** (3), 282–290.

Özger, M. & Kabataş, M. B. 2015 Sediment load prediction by combined fuzzy logic-wavelet method. *Journal of Hydroinformatics* **17** (6), 930–942.

Palosuo, T., Kersebaum, K. C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J. E., Patil, R. H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, F., Mirschel, W., Şaylan, L., Šiška, B. & Rötter, R. 2011 Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *European Journal of Agronomy* **35** (3), 103–114.

Pearson, K. 1895 Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* **58** (347–352), 240–242.

Pfannerstill, M., Guse, B. & Fohrer, N. 2014 Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *Journal of Hydrology* **510**, 447–458.

Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A. & Shufeng, T. 2019 Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment* **3** (3), 585–601.

Ritter, A. & Muñoz-Carpena, R. 2013 Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology* **480**, 33–45.

Rujner, H., Leonhardt, G., Marsalek, J. & Viklander, M. 2018 High-resolution modelling of the grass swale response to runoff inflows with Mike SHE. *Journal of Hydrology* **562**, 411–422.

Saleh, A., Arnold, J. G., Gassman, P. W., Hauck, L. M., Rosenthal, W. D., Williams, J. R. & McFarland, A. M. S. 2000 Application of SWAT for the Upper North Bosque River Watershed. *Transactions of the American Society of Agricultural Engineers* **43** (5), 1077–1087.

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R. & Hauck, L. M. 2001 Validation of the SWAT model on a large river basin with point and nonpoint sources. *Journal of the American Water Resources Association* **37** (5), 1169–1188.

Şişman, E. & Kizilöz, B. 2020a Artificial neural network system analysis and Kriging methodology for estimation of non-revenue water ratio. *Water Science and Technology: Water Supply* **20** (5), 1871–1883.

Şişman, E. & Kizilöz, B. 2020b Trend-risk model for predicting non-revenue water: an application in Turkey. *Utilities Policy* **67**, 1–9.

Swathi, V., Srinivasa Raju, K., Varma, M. R. R. & Sai Veena, S. 2019 Automatic calibration of SWMM using NSGA-III and the effects of delineation scale on an urban catchment. *Journal of Hydroinformatics* **21** (5), 781–797.

Tian, Y., Nearing, G. S., Peters-Lidard, C. D., Harrison, K. W. & Tang, L. 2015 Performance metrics, error modeling, and uncertainty quantification. *Monthly Weather Review* **144** (2), 607–613.

Van Der Keur, P., Hansen, S., Schelde, K. & Thomsen, A. 2001 Modification of DAISY SVAT model for potential use of remotely sensed data. *Agricultural and Forest Meteorology* **106** (3), 215–231.

Van Liew, M. W., Veith, T. L., Bosch, D. D. & Arnold, J. G. 2007 Suitability of SWAT for the conservation effects assessment project: comparison on USDA agricultural research service watersheds. *Journal of Hydrologic Engineering* **12** (2), 173–189.

Willmott, C. J. 1981 On the validation of models. *Physical Geography* **2** (2), 184–194.

Zhang, X., Hörmann, G., Fohrer, N. & Gao, J. 2012 Parameter calibration and uncertainty estimation of a simple rainfall-runoff model in two case studies. *Journal of Hydroinformatics* **14** (4), 1061–1074.

Zhang, R., Moreira, M. & Corte-Real, J. 2016 Multi-objective calibration of the physically based, spatially distributed SHETRAN hydrological model. *Journal of Hydroinformatics* **18** (3), 428–445.

First received 27 September 2020; accepted in revised form 22 July 2021. Available online 4 August 2021

Downloaded from http://iwaponline.com/ws/article-pdf/22/1/589/992984/ws022010589.pdf
by guest