Chapter 13
Evaluation of Mathematical Models with Utility Index: A Case Study from Hydrology

Dr. Renji Remesan and Dr. Dawei Han

Abstract Conventional error-based statistical parameters like the Nash–Sutcliffe efficiency index are popular among hydrologists to check the accuracy of hydrological models and to compare the relative performance of alternative models in a particular modelling scenario. A major drawback of those traditional indices is that they are based on only one modelling attribute, i.e. the modelling error. This study has identified an overall model utility index as an effective error-sensitivity-uncertainty procedure which could serve as a useful quality indicator of data-based modelling. This study has also made an attempt to answer the question—should the increasing complexity of the existing model add any benefit to the model users? The study evaluates the utility of some popular and widely used data-based models in hydrological modelling such as local linear regression, artificial neural networks (ANNs), Adaptive neuro fuzzy inference system (ANFIS) and support vector machines (SVMs) along with relatively complex wavelet hybrid forms of ANN, ANFIS and SVM in the context of daily rainfall–runoff modelling. The study has used traditional error-based statistical indices to confirm capabilities of model utility index values in identifying better model for rainfall–runoff modelling. The implication of this study is that a modeller may use utility values to select the best model instead of using both calibration and validation processes in the case of data scarcity. The study comprehensively analysed the modelling capabilities of SVM and its waveform in the context of rainfall–runoff modelling.

Keywords Support vector machines • Wavelets • Local linear regression • Artificial neural network • Fuzzy inference system • Hydrology

---

R. Remesan (✉)
Cranfield Water Science Institute, Cranfield University, Vincent Building, College Road, Cranfield, Bedfordshire MK43 0AL, UK
e-mail: r.remesan@cranfield.ac.uk

D. Han
Department of Civil Engineering, University of Bristol, Bristol, UK

T. Islam et al. (eds.), *Computational Intelligence Techniques in Earth and Environmental Sciences*, DOI 10.1007/978-94-017-8642-3_13,

13.1 Introduction

Data-based modelling techniques have been popular in the field of hydrology for several decades. Even in the acclaimed success of different data models in hydrology, there are still many questions that need to be answered. The relevant questions in data-based modelling in hydrology are how useful is a model for predicting a particular component within the hydrological cycle? and does a complex model work better than simple ones? Visual judgment and statistical measures are two common approaches employed to establish the integrity of any data-based mathematical models. The usefulness of any model depends ultimately on its contribution to the success of decision making, not on its ability to generate unassailably correct numerical values (Pepelnjak 2009). It is often difficult in hydrology to decide which model should be used for a particular purpose, and the decision is often made on the basis of familiarity rather than the appropriateness and effectiveness of the model. Legates and McCabe (1999) have reviewed many major statistical measures that are used by the hydrologists to validate models, which includes the Nash–Sutcliffe (NS) efficiency index, the root mean square error (RMSE), the coefficient of correlation, the coefficient of determination, the mean absolute error and many more, out of which the NS-efficiency index is one of the most commonly used indicators for model comparison and performance evaluation in hydrology. A study by Jain and Sudheer (2008) has demonstrated the weakness of the NS-efficiency index in model comparison. Comparing different models just in terms of their better accuracy in simulating the numerical values is often misleading as there are many other aspects that need to be accounted for before declaring that one model with entirely different mathematical concepts is better than the other. It is a known fact that the best model is not necessarily the most complex, or the one which overtly reflects the most sophisticated understanding of the system (Barnes 1995). There is a hypothesis that more complex models simulate the processes in a better way but with high variability in sensitivity and relatively less bias (Snowling and Kramer 2001). On the other hand, a study by Oreskes et al. (1994) argues that there is no strong evidence that simple models are more likely to produce more accurate results than complex models. Snowling and Kramer (2001) have connected the usefulness of the model to model's uncertainty which was assessed through different modelling attributes like model error, model sensitivity and model complexity.

In hydrology and water resources research, there are two major bases of uncertainty attitudes; one is based on stochasticity as a necessary factor and the other is based on deterministic nature of the system. The definition of uncertainty is much more uncertain about the modelled numerical values; it relates to much deeper processes and pertains to the governing mechanisms of the model. Distinguishable uncertainties in hydrology are data uncertainties (mainly associated with measurements), sample uncertainties (e.g. number of data for calibration) and model uncertainty (Plate and Duckstein 1987). Klir (1989) made an attempt to consider uncertainty in terms of the complexity of the model. He found both categories have a conflictive nature, i.e. if complexity decreases, the uncertainty grows. In the last 20 years,

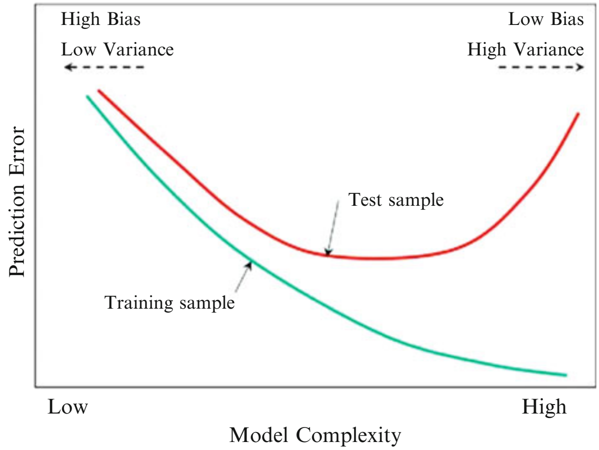

Fig. 13.1 Hypothesis showing effect of complexity during training and testing (Hastie et al. 2001)

the study of complexity in modelling systems has emerged as a recognised field in statistics. Though, the initial attempts to formalise the concept of complexity go back even further to Shannon's inception of information theory (Shannon 1948). The complexity of a model is closely related with the uncertainty of the system which can be defined in terms of model properties like model sensitivity and modelling error. The general hypothesis of model complexity and its influence during training and testing phases is shown in Fig. 13.1. The general hypothesis states that more complex models can simulate reality better than simpler models (i.e. less prediction error), and with a greater variance and low bias during training phase. Less complex models provide a relatively approximate simulation (i.e. more prediction error), but with less variance and high bias. But the case is a bit different in testing phase; highly complex models won't give best test results as the graph is parabolic with a minimum somewhere in the middle.

Figure 13.2 displays the hypothesis which shows the variation of different model parameters particularly with bias–variance interaction during the test phase.

Models of different complexity may show different modelling properties like sensitivity, flexibility, error and data requirements based upon their structure. Figure 13.3 illustrates the hypothetical relationship between model sensitivity, modelling error, model flexibility, training data requirement and model complexity.

The aim of this chapter is to highlight the need to have a statistical comparative index in data-based modelling which considers modelling attributes like model error, model complexity and model sensitivity. The study has made use of modified form of the overall model utility index proposed by Snowling and Kramer (2001) to identify "the best and right" model in data-based hydrological modelling, which was accomplished through a major case study, using the daily information of rainfall and runoff data from the Brue catchment in the United Kingdom. The utility-based results are compared with that of the traditional statistical indices. Another objective of this study is to ascertain if the usefulness of a model changes if one performs the wavelet-based input data splitting.

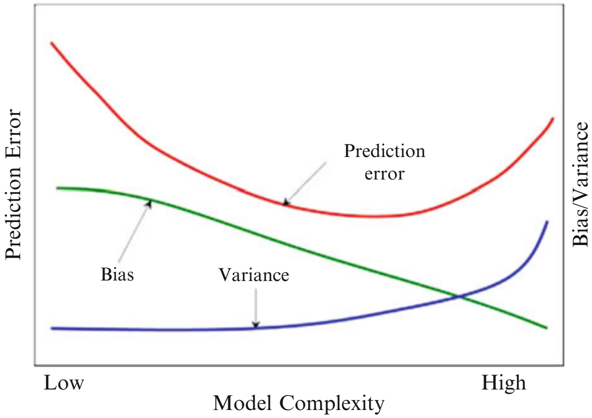

Fig. 13.2 Hypothesis showing effect of complexity on bias–variance interaction (Hastie et al. 2001)

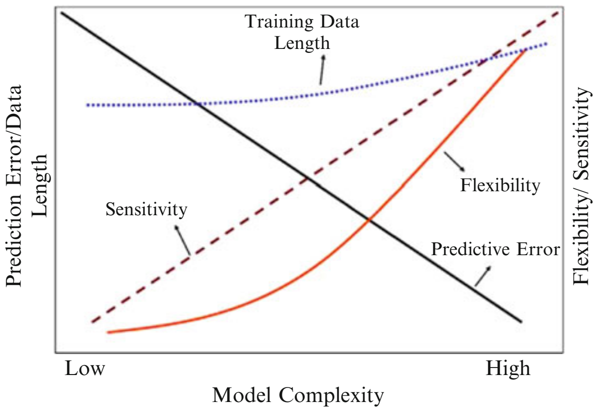

Fig. 13.3 Hypothetical relation of model complexity with sensitivity, flexibility, data requirements and predictive error

13.2 Study Area and Data Used

This study has used daily rainfall and runoff data from the Brue catchment of the United Kingdom. The River Brue catchment is located in Somerset, South West of England. It is considered as one of the best representative catchments to express hydrological responses in England, due to its data quality for a reasonably long time. This catchment has been extensively used in many good quality studies on weather radar, quantitative precipitation and flood forecasting and rainfall–runoff modelling. The location is famous among researchers because of its well-facilitated dense rain gauge network as well as the coverage by three weather radars. The River Brue catchment was the site of the Natural Environment Research Council

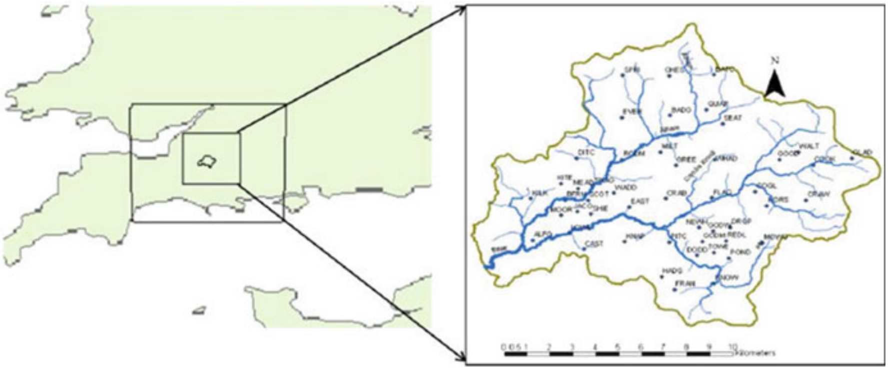

Fig. 13.4 The location map of the study area, the Brue River catchment

(NERC)-funded HYREX project (Hydrological Radar Experiment) from 1993 to 2000. The catchment was chosen for the HYREX project, as its size and relief were seen as representative of many catchments in the United Kingdom to demonstrate the hydrological dynamics and flood forecasting procedures. The catchment has a drainage area of 135 km^2 and an elevation range between 35 and 190 m above sea level. The catchment is located at 51.075°N and 2.58°W (Fig. 13.4). The river gauging point at the catchment is located at Lovington. An automatic weather station (AWS) and an automatic soil water station (ASWS) are located in the catchment which recorded the global solar radiation, net radiation and other weather parameters such as wind speed, wet and dry bulb temperatures and atmospheric pressure in hourly interval. Six years of daily rainfall–runoff data from the Brue catchment, spanning from 1993 to 2000, was used in this study. For the rainfall–runoff modelling, the study has used effective inputs like three-step antecedent runoff values ($Q(t-1)$, $Q(t-2)$, $Q(t-3)$), one-step antecedent rainfall ($P(t-1)$) and current rainfall information ($P(t)$) for hybrid modelling as observed in the previous studies (Remesan et al. 2009). The optimum training data length for this daily rainfall–runoff data set was identified as 1,056 data points (Remesan et al. 2009) which was used as training data set throughout the study and the rest is used for validation.

13.3 Models

The study has used several data-based models such as local linear regression (LLR) model, artificial neural networks (ANNs), Adaptive neuro fuzzy inference system (ANFIS), support vector machines (SVMs) and hybrid wavelet forms of ANN, ANFIS and SVMs in order to cover a wide range of models used in hydrology.

13.3.1 LLR Model

The LLR model is a widely accepted nonparametric regression method due to its better prediction capabilities in low dimensional forecasting and modelling problems. The attraction of LLR technique is its consistent performance even with a small amount of sample data. In the mean time, LLR can produce very accurate predictions in regions of high data density in the input space. The LLR procedure requires only three data points to obtain an initial prediction and then uses all newly updated data as they become available to make further predictions. The only problem with LLR is to decide the size of $p_{\max}$, the number of near neighbours to be included for the local linear modelling.

Given a neighbourhood of $p_{\max}$ points, we must solve a linear matrix equation

$$\mathbf{Xm} = \mathbf{y} \tag{13.1}$$

where $\mathbf{X}$ is a $p_{\max} \times d$ matrix of the $p_{\max}$ input points in d-dimensions, $\mathbf{x}_i (1 \leq i \leq p_{\max})$ are the nearest neighbour points, $\mathbf{y}$ is a column vector of length $p_{\max}$ of the corresponding outputs and $\mathbf{m}$ is a column vector of parameters that must be determined to provide the optimal mapping from $\mathbf{X}$ to $\mathbf{y}$, such that

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1d} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{x_{p_{\max}}1} & x_{x_{p_{\max}}2} & x_{x_{p_{\max}}3} & \cdots & x_{x_{p_{\max}}d} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_d \end{pmatrix} = \begin{pmatrix} y_1 \\ y_1 \\ \vdots \\ y_{p_{\max}} \end{pmatrix} \tag{13.2}$$

The rank $\boldsymbol{r}$ of the matrix x is the number of linearly independent rows, which will affect the existence or uniqueness of the solution for m.

If the matrix $\mathbf{X}$ is square and non-singular then the unique solution to Eq. 13.1 is $\mathbf{m} = \mathbf{X}^{-1}\mathbf{y}$. If $\mathbf{X}$ is not square or singular, we modify Eq. 13.1 and attempt to find a vector $\mathbf{m}$ which minimises

$$|\mathbf{Xm} - \mathbf{y}|^2 \tag{13.3}$$

13.3.2 ANN and ANFIS Models

The theory of ANNs started in the early 1940s when McCulloch and Pitts developed the first computational representation of a neuron (McCulloch and Pitts 1943). The ANNs are nonlinear formations which work based on the function of human neural system. ANNs have become focus of much attention in last few decades in hydrology due to their immense capabilities in the implementation of nonlinear static and dynamic systems. The most commonly used learning algorithm in

ANNs is the back-propagation algorithm. Algorithms like conjugate gradient, quasi-Newton and Levenberg–Marquardt (LM) are considered as some of the faster algorithms, all of which make use of standard numerical optimisation techniques. The Levenberg–Marquardt (LM) learning algorithm was used in this study. There are several types of ANNs like multilayer perceptron, radial basis functions and Kohonen networks. ANN structure defines its structure including number of hidden layers, number of hidden nodes, number of input and output nodes and activation function. For hidden layer the sigmoid activation function and for output layer linear activation function were used in this study. Three-layer feed-forward neural network (one input layer, one hidden layer and one output layer) is the most commonly used topology in hydrology. This topology has proved its ability in modelling many real-world functional problems. The selection of hidden neurons is the tricky part in ANN modelling as it relates to the complexity of the system being modelled and there are several ways of doing it, such as the geometric average between input and output vector dimensions (Maren et al. 1990), the same as the number of inputs used for the modelling (Mechaqrane and Zouak 2004), twice the input layer dimension plus one (Hecht-Nielsen 1990), etc. In this study, the Hecht-Nielsen (1990) approach has been adopted according to our past experimental experience with it.

Adaptive neuro-fuzzy inference system (ANFIS) model is a well-known artificial intelligence technique that has been used in modelling hydrological processes. The ability of neural network to learn fuzzy structure from the input–output data sets in an interactive manner has encouraged many researchers to combine the ANN and the fuzzy logic effectively to organise network structure itself and to adapt parameters of a fuzzy system. The ANFIS model used in this study is based on the Sugeno fuzzy model, which is based on a systematic approach to generate fuzzy rules and membership function parameters for fuzzy sets from a given hydrological time series data set (Sugeno and Kang 1988; Jang 1993). The learning algorithm for ANFIS is a hybrid algorithm, which is a combination between the gradient descent method and the least squares method for identifying nonlinear input parameters and the linear output parameters, respectively. The ANFIS modelling was performed using the "subtractive fuzzy clustering" function due to its good performance with a small number of rules.

For a first-order Sugeno fuzzy model, a typical rule set with two fuzzy If/Then rules can be expressed as

$$\text{Rule 1 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ Then } f_1 = p_1 x + q_1 y + r_1 \tag{13.4}$$

$$\text{Rule 2 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ Then } f_2 = p_2 x + q_2 y + r_2 \tag{13.5}$$

where x and y are the crisp inputs to the node i, A_i and B_i are the linguistic labels (low, medium, high, etc.) characterised by convenient membership functions and p_i, q_i and r_i are the consequence parameters ($i = 1$ or 2). In the ANFIS, nodes in the same layer have similar functions as described below.

(a) *Layer 1* (*input nodes*): Nodes of this layer generate membership grades of the crisp inputs which belong to each of the convenient fuzzy sets using the membership functions. The generated bell-shaped membership function given below was used:

$$\mu_{A_i}(x) = \frac{1}{1 + ((x - c_i)/a_i)^{2b_i i}} \tag{13.6}$$

where μ_{A_i} is the appropriate membership function for A_i fuzzy set, and $\{a_i, b_i, c_i\}$ is the membership function's parameter set (premise parameters) that changes the shape of membership function from 1 to 0.

(b) *Layer 2* (*rule nodes*): In this layer, the rule operator (AND/OR) is applied to get one output that represents the results of the antecedent for a fuzzy rule. The outputs of the second layer, called as firing strengths O_i^2, are the products of the incoming signals obtained from the layer 1, named as w below:

(c) *Layer 3* (*average nodes*): In this layer, the nodes calculate the ratio of the ith rule's firing strength to the sum of all rules' firing strengths

$$O_i^3 = \overline{w}_i = \frac{w_i}{\sum_i w_i}, \quad i = 1, 2 \tag{13.8}$$

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1, 2 \tag{13.7}$$

(d) *Layer 4* (*consequent nodes*): In this layer, the contribution of ith rule towards the total output or the model output and/or the function is calculated as follows:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i), \quad i = 1, 2 \tag{13.9}$$

where $\overline{w}_i$ is the output of Layer 3 and $\{p_i, q_i, r_i\}$ are the coefficients of a linear combination in Sugeno inference system. These parameters of this layer are referred to as consequent parameters.

(e) *Layer 5* (*output nodes*): This layer is called the output nodes. This layer's single fixed node computes the final output as the summation of all incoming signals.

$$O_i^5 = f(x, y) = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{13.10}$$

13.3.3 Support Vector Machines

Just like ANNs, SVMs can be represented as two-layer networks (where the weights are nonlinear in the first layer and linear in the second layer).

Mathematically, a basic function for the statistical learning process is:

$$y = f(x) = \sum_{i=1}^{M} \alpha_i \varphi_i(x) = w\varphi(x) \tag{13.11}$$

where the output is a linearly weighted sum of M. The nonlinear transformation is carried out by $\varphi(x)$.

The decision function of SVM is represented as:

$$y = f(x) = \left\{ \sum_{i=1}^{N} \alpha_i K(x_i, x) \right\} - b \tag{13.12}$$

where K is the kernel function, α_i and b are parameters, N is the number of training data, x_i are the vectors used in training process and x is the independent vector. The parameters α_i and b are derived by maximising their objective function.

The role of the kernel function simplifies the learning process by changing the representation of the data in the input space to a linear representation in a higher dimensional space called a feature space. A suitable choice of kernels allows the data to become separable in the feature space despite being non-separable in the original input space. Four standard kernels are usually used in classification problems and also used in regression cases: linear, polynomial, radial basis and sigmoid:

Linear: $u' \times v$
Polynomial: $(\gamma \times u'v + \text{coef})^{\text{degree}}$
Radial basis: $e^{-\gamma \left| u-v \right|^2}$
Sigmoid: $\tanh(\gamma \times u'v + \text{coef})$

The SVM software used in this study was LIBSVM developed by Chih-Chung Chang and Chih-Jen, and supported by the National Science Council of Taiwan. The basic algorithm is a simplification of both SMO by Platt and SVMLight by Joachims. The source code is written in C++. The choice of this software was made on its ease of use and dependability. It has been tried and tested in several research institutions worldwide including the Computer and Information Sciences Department, University of Florida, USA, and the Institute for Computer Science, University of Freiburg, Germany. The LIBSVM is capable of C-SVM classification, one-class classification, ν-SV classification, ν-SV regression and ε-SV regression. The model first trains the SVM with a list of input vectors describing the training data. It outputs a model file, which contains a list of support vectors and hence the description of the hypothesis for the particular regression problem (Bray and Han 2004).

13.3.4 Wavelet Hybrid Models

This study has used three different types of wavelet hybrid models namely neuro-wavelet (NW) models, wavelet-adaptive-network-based fuzzy inference system

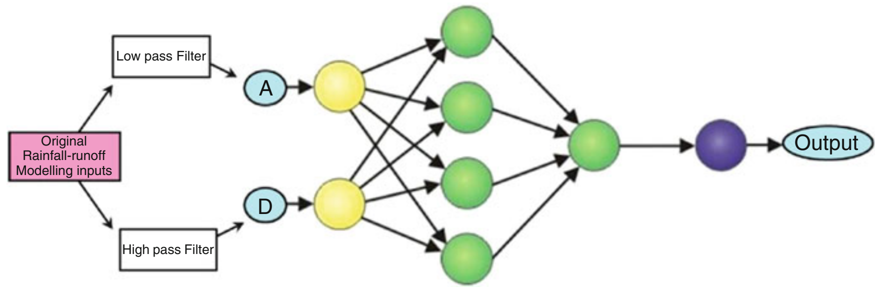

Fig. 13.5 The proposed hybrid scheme for NW model

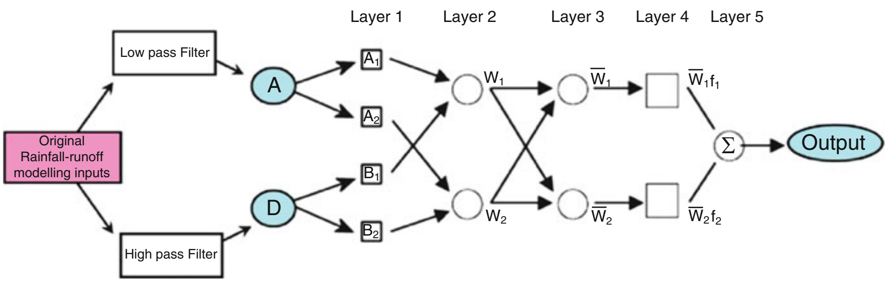

Fig. 13.6 The proposed hybrid scheme for W-ANFIS model

(W-ANFIS) model and wavelet-support vector machine (W-SVM) model in conjunction with ANNs, ANFIS and SVMs, respectively. A multilayer feed-forward network type of ANN and discrete wavelet transfer (DWT) model were combined together to obtain a neuro-wavelet (NW) model. The DWT model is functioned through two set of filters, viz. high-pass and low-pass filters, which decompose the signal into two set of series namely detailed coefficients (*D*) and approximation (*A*) sub-time series, respectively. Please refer to Remesan et al. (2009) for further details of wavelet model construction. In the proposed NW model, these decomposed sub-series obtained from DWT on the original data directly are used as inputs to the ANN model. This study has used another conjunction model: wavelet-neuro-fuzzy is applied in subsequent sections for daily rainfall–runoff modelling. The W-ANFIS model utilises the time–frequency representation ability of DWT to display the data in the time domain in conjunction with a conventional ANFIS model. The input antecedent information data considered are decomposed into wavelet sub-series by discrete wavelet transform and then the neuro-fuzzy model is constructed with appropriate wavelet sub-series as input, and desired time step of the target time series as output. In the case of W-SVM, the DWT model is combined with the SVM model with the best kernel function and best regressor and several trial and error evaluations. The detailed model structure and proposed specifications are given in Figs. 13.5, 13.6 and 13.7 which correspond to NW, W-ANFIS and W-SVM models.

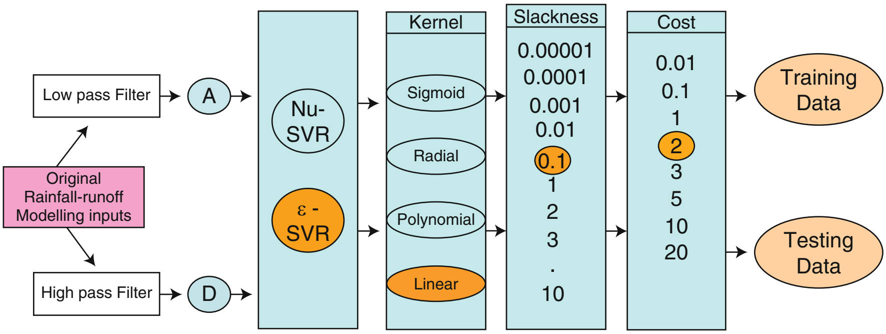

Fig. 13.7 The proposed hybrid scheme for W-SVM model

13.3.5 *Index of Model Utility* (U)

This study has adopted an index of model utility to make a decision about which is the "best and right" model for a hydrological modelling exercise. The adopted approach is a modified version of Snowling and Kramer (2001) for the suitability in data-based modelling. Statistically the proposed "index of model utility" of a model can be defined as scaled distance from origin on a graph of sensitivity vs. modelling error of different models to the point corresponding to that model in the graph. Mathematically it can be written as

$$U_i = 1 - \sqrt{\frac{K_s S_i^2 + K_e E_i^2}{(K_s + K_e)}} \tag{13.13}$$

where

U_i is the utility index for model i
S_i, sensitivity value for model i (relative to the maximum sensitivity). In this study the value is obtained from the mean value of slope of all sensitivity curves obtained from all inputs
E_i, error value for model i (relative to the maximum error; this study has adopted RMSE as the indicator of model error)
K_s and K_e are weighting constants for sensitivity and error, respectively

U value varies between 0 and 1 and if the value of U is larger the model has higher utility. The values of S and E for each model should be normalised to satisfy the equation; that's the reason for dividing all values by the maximum sensitivity and error value. The values of K_s and K_e depend on how the model values error and sensitivity. If error and sensitivity are valued equally, then K_s and K_e should both be set to 1. In this study both values were set to 1, and the model utility indexes (U) were calculated for this case study using different models detailed above.

13.4 Results and Discussions

In this section different popular data-based models in hydrology were established and compared for the case study: daily rainfall–runoff modelling. The first subsection evaluates the utility of these models in different case studies, in terms of model attributes like model error, model sensitivity and model complexity. In the second subsection the models were compared and contrasted with finding of the overall model utility index, in terms of traditional statistical parameters like RMSE, Nash–Sutcliffe efficiency (E), mean bias error (MBE), slope and correlation coefficient (CORR).

In this case study, several models were constructed and tested for predicting daily runoff values (using models ranging from relatively simple LLR model to relatively complex W-SVM). The nonparametric procedure based on LLR models does not require training in the same way as that of neural network models. The optimal number of nearest neighbours for LLR (principally dependent on the noise level) was determined by trial and error method and 16 nearest neighbours were implemented.

The adaptive Adaptive neuro fuzzy inference system (ANFIS) model was used for daily rainfall–runoff modelling, in which a set of parameters in ANFIS were identified through a hybrid learning rule combining the back-propagation gradient descent and a least squares method. The ANFIS model in this modelling used a set of fuzzy IF–THEN rules with appropriate membership functions. The subtractive fuzzy clustering was used to establish the rule-based relationship between input data series and output data variable. The subtractive clustering was used to automatically identify the natural clusters in the input–output data pool. In this ANFIS model, there were 32 parameters to determine in the layer 2 because of five input variables. The three rules generated 3^6 nodes in the subsequent layer. The study set the number of membership functions for each input of ANFIS as three with Gaussian (or bell-shaped) and linear membership functions at the inputs and outputs, respectively.

For SVM modelling, C++-based LIBSVM with ν-SV and ε-SV regressions was used as explained in methodology section. Normalisation of input vectors and proper identification of different parameters are very important in SVM modelling. Initial analysis results in rainfall–runoff modelling at the Brue catchment on both ν-SVM and ε-SVM using different kernel functions are shown in Fig. 13.8. The SVM analysis on the Brue catchment daily data has used different kernel functions like linear, polynomial, radial and sigmoid and compared in terms of mean-squared errors. Out of these eight analysis results, the results from two SVM stood out quite considerably from the remaining six. These were ε-SVM and ν-SVM with linear kernel function (Fig. 13.8). The performance of E-SVM with linear kernel was better than that of ν-SVM with linear kernel. However, it was unclear whether this performance was due to the regression algorithm implemented or whether optimising the parameters within each algorithm would enhance the performance of one SVM over the other.

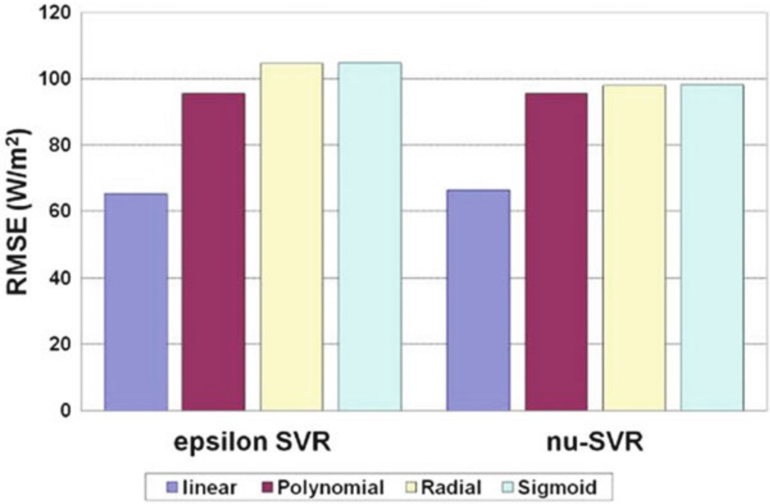

Fig. 13.8 Variation of performance in daily rainfall–runoff modelling at the Brue catchment when applying different support vector machines on different kernel functions

This analysis on daily data from the Brue catchment was performed after fixing the parameters to the default values (degree in kernel function is set as 3, coef0 in kernel function is set as zero, cache memory size is set as 40 Mb, tolerance of termination criterion is set as a default value of 0.001). The SVM hypothesis suggested that the performance of SVM depended on the slack parameter (ε) and the cost factor (C). The study has performed the modelling analysis varying the E values between $\varepsilon = 1$ to $\varepsilon = 0.00001$ and the cost parameters $C = 0.1$–1,000. The analysis results have shown that the least error increases rapidly for E greater than 0.1. So the study set the value of E to 0.1 for reliable results and less computation time.

The cost factor of error (C) assigns a penalty for the number of vectors falling between the two hyperplanes in the hypothesis. It suggests if the data is of good quality the distance between the two hyperplanes is narrowed down. If the data is noisy it is preferable to have a smaller value of C which will not penalise the vectors (Bray and Han 2004). So it was important to find the optimum cost value for SVM modelling. The cost value was chosen to be 2 because it produced the least error at that value, with the minimum running time, which was identified after several trial and error analyses.

The study has used a neuro-wavelet (NW) model for modelling; for this purpose, a multilayer feed-forward network type of ANN and DWT model were combined together to obtain an NW model. The DWT model is functioned through two set of filters, viz. high-pass and low-pass filters, which decompose the signal into two set of series namely detailed coefficients (D) and approximation (A) sub-time series, respectively. The present value of runoff has been estimated using the three

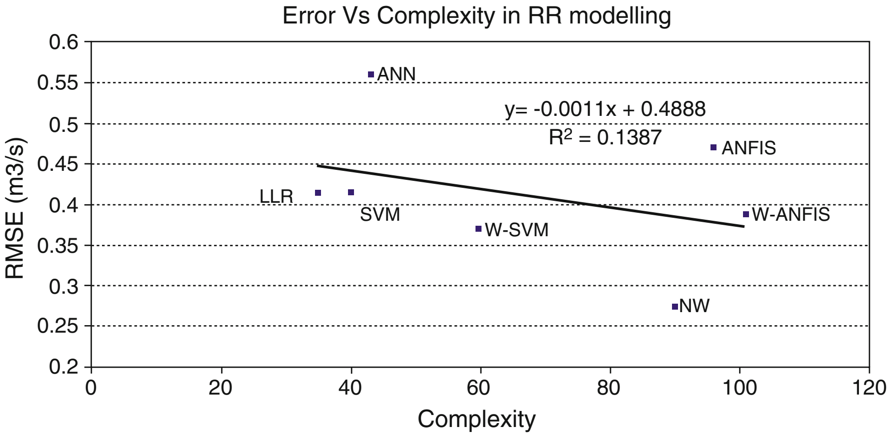

Fig. 13.9 Complexity vs. training error—case study: rainfall–runoff modelling

resolution levels of antecedent runoff and rainfall information (2-day mode (D^j_q1, D^i_p1), 4-day mode (D^j_q2, D^i_p2), 8-day mode (D^j_q3, D^i_p3) and approximate mode A^j_q3, A^i_p3 where q denotes runoff, p denotes rainfall and i and j denote number of antecedent data sets of rainfall and runoff, respectively). The above-mentioned decomposed sub-series were used as inputs in the case of wavelet hybrid forms of ANFIS and SVM (viz. W-ANFIS and W-SVM).

13.4.1 Comparison of Data Models Using Utility Index

The study has used the overall model utility index to evaluate and compare different data-based models from relatively simple LLR model to the hybrid W-SVM model. This index gives a measure of the "utility" of the model in a particular modelling scenario, with respect to modelling uncertainty (assuming that model uncertainty connects to its sensitivity, error and complexity). Thus we have assessed model attributes like model complexity (the study has used the training time as the indicator of complexity), model sensitivity (output response to changes in training input) and model error (closeness of simulation to measurement) of all seven data-based models used for daily rainfall–runoff modelling.

Figure 13.9 shows the variation of error (RMSE) with the model complexity for this case study i.e. daily rainfall–runoff modelling. The RMSE decreases with increasing complexity as this study has hypothesised. However, relatively complex ANN and ANFIS models have shown more erroneous prediction than the relatively simpler LLR model. The better prediction in terms of error was exhibited by the NW model, followed by W-SVM, W-ANFIS and SVM models. Even though, the hypothetical relation is a straight line, we observed a decreasing linear relation with R^2 value of 0.138.

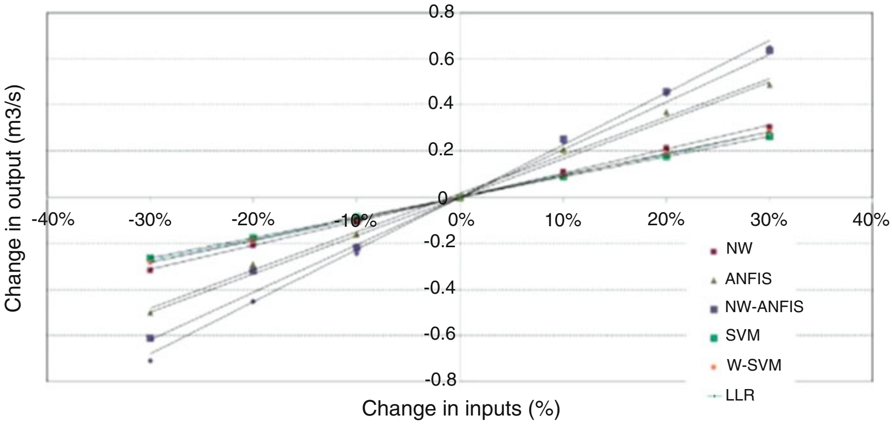

Fig. 13.10 Sensitivity curves for different data-based models

The sensitivity of the model with change in inputs used for training was assessed varying the inputs in the range of certain percentages. To find the training sensitivity of the model to the inputs, the study has changed all the inputs in the range of −30 to +30 % and checked the change in outputs produced in each scenario. These values were averaged to plot sensitivity diagram of each model as shown in Fig. 13.10. The slopes of these sensitivity diagrams were estimated and these slope values were considered as the measure of sensitivity.

Figure 13.11 shows the results of the variation of the sensitivity of different models with the corresponding complexity values. The sensitivity showed an increasing linear trend with increasing complexity with an R^2 value of 0.58. Even though the SVM model was a bit complex in structure the sensitivity value was observed close to that of LLR model. The highest value of sensitivity was observed with NW model, followed by W-ANFIS and ANFIS. The complex W-SVM showed relatively low sensitivity compared to other hybrid wavelet models like NW and W-ANFIS.

Now a modeller can make a decision in terms of uncertainty (expressed in terms of error and sensitivity) and complexity (expressed in terms of modelling time). The overall model utility statistic requires the error and the sensitivity to express in relative ratio to the maximum value. Table 13.1 shows different data models and the corresponding overall utility indices.

The value of the overall model utility index (U) varies between 0 and 1, where the larger the value of U, the greater the model utility considering aspects like uncertainty and complexity. The resultant figure shows the utility values corresponding to different models for the case study: rainfall–runoff modelling is shown in Fig. 13.12. Due to the relatively high sensitivity of NW/W-ANFIS models and relatively high error of ANN models, these three models stand out as the three lowest utility models in the rainfall–runoff modelling study. It means that even

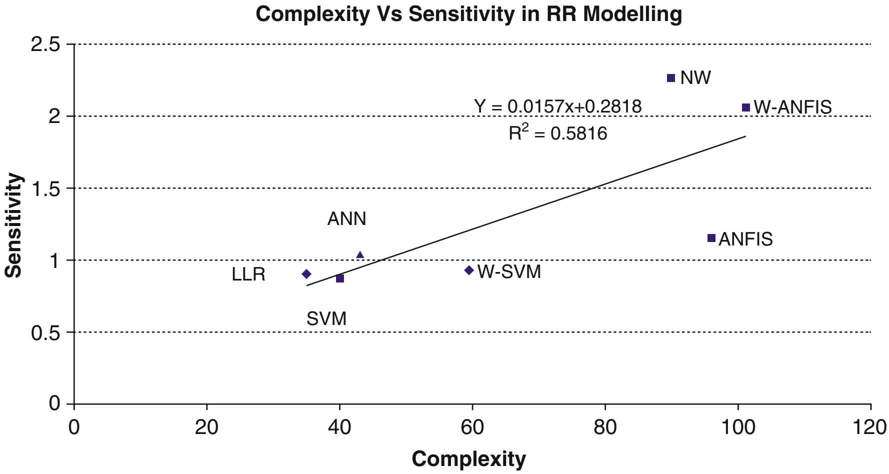

Fig. 13.11 Sensitivity vs. complexity—case study: rainfall–runoff modelling

Table 13.1 Different models and their attributes which decide overall model utility in rainfall–runoff modelling

Model	Complexity (function of modelling time)	Sensitivity (function of slope of sensitivity curves)	RMSE (m^3/s)	RMSE (%)	Sensitivity (%)	Utility (U)
ANN	43	1.04	0.558	1	0.458	0.222
NW	90	2.27	0.274	0.491	1	0.212
ANFIS	96	1.15	0.47	0.842	0.506	0.304
W-ANFIS	101	2.06	0.39	0.698	0.907	0.190
SVM	40	0.88	0.415	0.743	0.387	0.406
W-SVM	60	0.95	0.37	0.663	0.418	0.445
LLR	35	0.9	0.414	0.741	0.396	0.405

though the models are better at predicting numerical values, when considering other attributes, which decide consistency in modelling, complex models may stand out of "better and useful" model groups. Among all seven models, the W-SVM model has the best model utility followed by the models like SVM and LLR. It was interesting to note that the complex model SVM and relatively simple model LLR have very close utility values.

13.4.2 Comparison of Data Models Using Statistical Indices

Though the study has presented the utility evaluation as an alternative method for model comparison, it is essential to have a look into how the traditional statistical

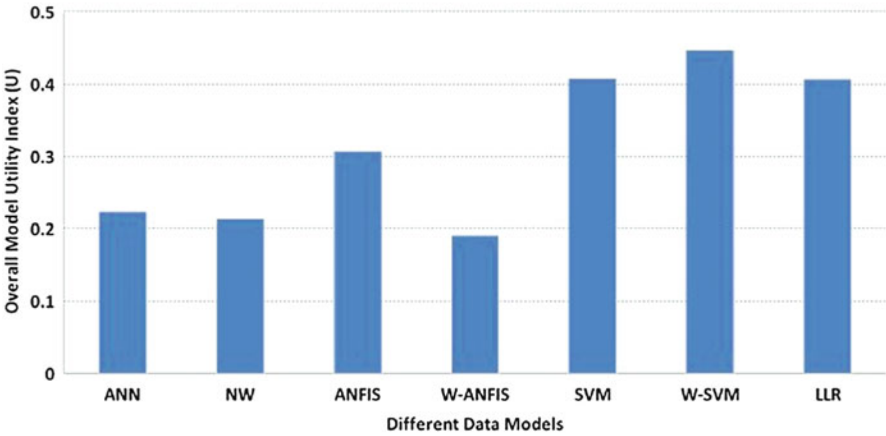

Fig. 13.12 Overall model utility—case study: rainfall–runoff modelling

indices behave in this modelling case study. The performance of the developed LLR technique was compared with other models using major global statistics as shown in Table 13.2. The estimated daily runoff values using the LLR model for 1,056 data points resulted in the overall RMSE value of 0.414 m^3/s which is 20.7 %; compared to observed daily runoff values and the MBE observed as -0.028 m^3/s. The Levenberg–Marquardt algorithm-based ANN underperformed to that of LLR in both the training and validation phases (RMSE value of 0.558 m^3/s (27.9 % of mean observed) and MBE value of -0.085 m^3/s during validation phase). It was seen that the LLR model's performance had a superior efficiency and performance compared with Levenberg–Marquardt ANN model with lower RMSE and higher CORR, for the training period and validation periods.

Table 13.2 implies that the performance of ANFIS model is remarkably better than that of LM-based ANN model in both validation and training data. The ANFIS model showed an efficiency of 88.9 % (increase of 5.45 % from ANN model) for the training data, and a validation efficiency of 77.3 % (an increase of 7.81 % compared to ANN). The correlation coefficients between the computed and observed are found to be 0.88 during training and 0.75 during validation. In order to assess the robustness of the models developed, evaluation criterion like MBE was used. From the MBE value one can deduce that both ANN and ANFIS showed underestimation for the training data and validation data. The underestimation is less for ANFIS during validation phase compared with ANN as a low value of MBE was observed. However, the performance of LLR was observed better than that of ANFIS model during training phase while ANFIS model outperformed LLR in the validation phase.

The analysis results in Table 13.2 have shown that the NW model is superior in predicting runoff values in comparison to all other models. The performance efficiency of the NW model is 4.13 % higher than that of the ANFIS model for validation and the corresponding value for the training data is 8.21 % higher.

Table 13.2 Comparison of some basic performance indices of different models employed in the study for daily rainfall–runoff modelling at the Brue catchment

	Training data (1,056 data points)					Validation data				
Models used	RMSE[a] (m^3/s and %)	CORR	Slope	MBE (m^3/s)	*E*	RMSE[a] (m^3/s and %)	CORR	Slope	MBE (m^3/s)	*E*
LLR	0.414 (20.7)	0.92	0.93	−0.028	0.923	0.922 (37.7)	0.70	0.80	−0.171	0.72
ANN-LM	0.558 (27.9)	0.83	0.90	−0.007	0.843	0.877 (36.4)	0.68	0.80	−0.085	0.717
ANFIS	0.470 (23.4)	0.88	0.93	−0.038	0.889	0.796 (33.2)	0.75	0.85	−0.039	0.773
SVM	0.415 (20.75)	0.89	0.91	−0.062	0.91	0.692 (28.3)	0.79	0.82	−0.012	0.831
W-SVM	0.370 (18.8)	0.90	0.90	−0.051	0.90	0.670 (27.2)	0.75	0.83	−0.112	0.770
W-ANFIS	0.39 (19.5)	0.90	0.89	−0.059	0.905	0.702 (28.6)	0.80	0.89	−0.103	0.802
NW	0.274 (13.6)	0.96	0.97	−0.0002 5	0.962	0.699 (28.6)	0.81	0.92	−0.0068	0.805

[a]Root mean square error is also shown in percentage of the mean value of observed runoff

Compared with the ANN model, the efficiency values of the NW model are 14.1 % and 12.27 % higher for the training and validation data, respectively. In terms of MBE, the performance of the NW model outperformed all other tested models in both the training and validation phases.

As shown in the above section, the study chose the ε-SVM with linear kernel for modelling applying the value of ε to 0.1 and values of C to 2; the modelling results are shown in Table 13.2. The SVM model made a better modelling with RMSE value of 0.415 m^3/s (20.75 %) and CORR of 0.89 during the training phase. The corresponding values during the validation phase were 0.692 m^3/s (28.3 %) and 0.79, respectively. The SVM model has shown better statistical performance compared to ANN, and ANFIS with an efficiency of 0.91 during training. The ε-SVM is applied with DWT to form a W-SVM model. Likewise, the ANFIS model was combined with DWT to form a hybrid W-ANFIS model. In the case of W-ANFIS, the DWT decomposed the input data sets into three wavelet decomposition levels (2–4–8) as mentioned in the previous sections and are used for rainfall–runoff modelling. The performance analysis of wavelet-based ε-SVM (W-SVM) is shown in Table 13.2 along with its comparison with the W-ANFIS model. The table implies that the performance analysis of both the W-ANFIS model and the W-SVM models was remarkably well in both validation and training data. The W-SVM model showed an efficiency of 90.0 % (increase of 6.76 % from ANN model) for the training data, and a validation efficiency of 77.0 % (an increase of 7.39 % compared to ANN). The correlation coefficient between the computed and observed are found to be 0.90 during training and 0.75 during validation. The RMSE for the LM-based ANN model is lower (0.558 m^3/s (27.9 %)) compared with the W-SVM model (0.37 m^3/s (18.8 %)) during training. From MBE value one can see the significant improvements while using hybrid wavelet forms of SVM models. The performance of W-ANFIS model in predicting runoff values is observed superior to the conventional LM-based ANN model and inferior to hybrid wavelet-based SVM model. The runoff prediction was underestimated by all models including W-ANFIS model for both the training and validation phases as indicated by the MBE values in Table 13.2. The RMSE value in the validation phase obtained by W-ANFIS model was 0.702 m^3/s (28.6 %), which was higher than that of the NW model and W-SVM model. The performance efficiency of W-ANFIS model in the rainfall–runoff modelling was 3.75 % lower than that of W-SVM model for validation and corresponding value for training data was 1.12 % lower. Compared with the ANN model, the efficiency values of the W-ANFIS model are 5.57 % and 11.85 % higher for training and validation data, respectively. However, in terms of MBE values, the performance of NW model outperformed all other tested models including W-ANFIS and W-SVM in both training and validation phases. Though many statistical parameters are used for evaluation of robustness of the model, the major index used for comparison of model performance is the Nash–Sutcliffe efficiency. Figure 13.13 shows comparison of overall model utility index and the Nash–Sutcliffe efficiency (average of training and validation) so in general the comparison using general error-based statistical indices and the N-S efficiency has shown that the rainfall–runoff modelling capabilities of data-based models

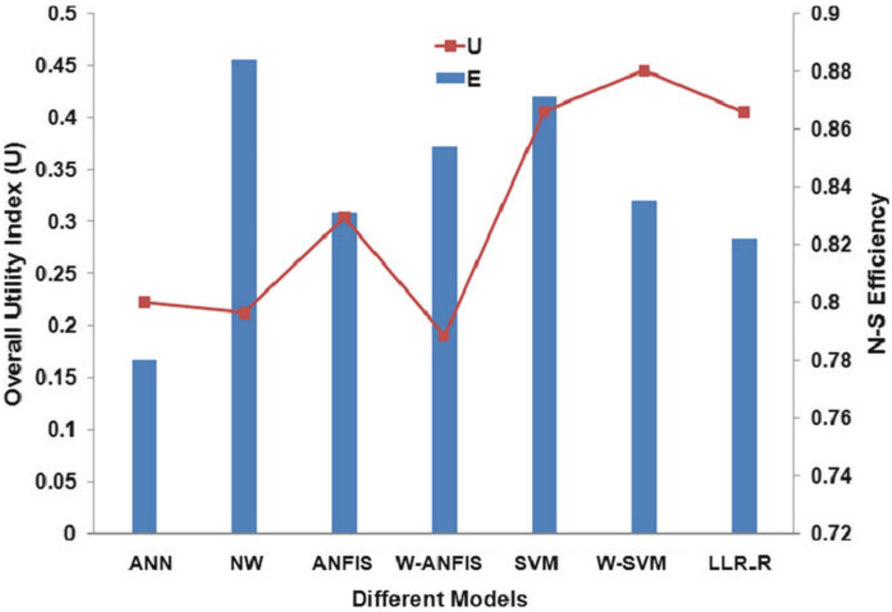

Fig. 13.13 Comparison of *S–E* efficiency and overall model utility index

are in the form of NW > SVM > W-ANFIS > W-SVM > ANFIS > LLR > ANN. However, the results are different when we consider other modelling attributes like model sensitivity and model complexity along with modelling errors. The utility index identified that the usefulness of models in this case study are in the form of W-SVM > SVM > LLR > ANFIS > ANN > NW > W-ANFIS. Both the approaches have acknowledged the better performance of SVM, giving second position in terms of efficiency and utility values. The N-S approach gave high ranking for both W-ANFIS and NW model as the approach couldn't account for the influence of higher sensitivity. The higher utility value of the wavelet-based SVM has shown the capabilities of SVM framework to handle large input space without any difficulty of sensitivity.

13.5 Conclusions

The study adopted a utility index to critically evaluate the acceptance of a model in terms of different modelling properties and contrasted the results with that of traditional statistical indices (particularly the N-S efficiency). This study has demonstrated that such an error-sensitivity-uncertainty procedure could help modellers make effective comparison of different data-based models and it can give an answer on how much the model benefits by increased complexity on data-based models. The study extensively analysed the capabilities of SVM in the context of rainfall–runoff modelling and demonstrated its ability to perform better even in larger wavelet-decomposed input space. The study has demonstrated the weakness of NW and W-ANFIS models. Even though these models had better numerical prediction results, the utility evaluation has shown their limitation in making a useful model for rainfall–runoff modelling due to their inclination towards sensitivity. The overall utility analysis based on the utility index has identified W-SVM as the best model, followed by SVM and LLR models.

The approach would be very useful in the data scarce situation where there is insufficient data for validation. The modeller could use this method for selection of the best possible model for the available data without diverting valuable data away from calibration of model. Though the study presented with useful information, there is room for improvement regarding the sensitivity assessment. The study has assessed local sensitivity of the model with respect to the variation of inputs. The term sensitivity is rather complicated; the local sensitivity slopes of these nonlinear models vary depending on the range of inputs. However the choice of the sensitivity and complexity values of these nonlinear models requires further research to develop the presented utility assessment to a robust method in data-based model comparisons.

References

Barnes CJ (1995) Commentary: the art of catchment modelling: what is a good model? Environ Int 21(5):747–751

Bray M, Han D (2004) Identification of support vector machines for runoff modelling. J Hydroinf 6(4):265–280

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

Hecht-Nielsen R (1990) Neurocomputing. Addison-Wesley, Reading, MA

Jain SK, Sudheer KP (2008) Fitting of hydrologic models: a close look at the Nash–Sutcliffe Index. J Hydrol Engrg 13(10):981–986. doi:10.1061/(ASCE)1084-0699(2008)13:10(981)

Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference systems. IEEE Trans Syst Man Cybern 23(3):665–685

Klir GJ (1989) Methodological principles of uncertainty in inductive modelling: a new perspective. In: Erickson GJ, Smith CR (eds) Maximum-entropy and Bayesian methods in science and engineering, vol 1. Kluwer, Dordrecht

Legates DR, McCabe GJ (1999) Evaluating the use of goodness of fit measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241

Maren AJ, Harston CT, Pap RM (1990) Handbook of neural computing applications. Academic, San Diego, CA

McCulloch WS, Pitts W (1943) A logical calculus of the ideas imminent in nervous activity. Bull Math Biophys 5:115–133

Mechaqrane A, Zouak M (2004) A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. Neural Comput Appl 13:32–37

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the Earth sciences. Science 263:641–646

Pepelnjak JC (2009) Validating the combined approach to online reach prediction. Atlas® Institute Digital Marketing Insights. http://www.atlassolutions.com /uploadedFiles/Atlas/Atlas_Institute/Published_Content/CombinedApproach.pdf

Plate EJ, Duckstein L (1987) Reliability in hydraulic design. In: Duckstein L, Plate EJ (eds) Engineering reliability and risk in water resources. NATO ASI Series, Series E: Applied Sciences, No. 124. Nijhoff, Dordrecht

Remesan R, Shamim MA, Han D, Mathew J (2009) Runoff prediction using an integrated hybrid modelling scheme. J Hydrol 372(1–4):48–60. doi:10.1016/j.jhydrol.2009.03.034

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Snowling SD, Kramer JR (2001) Evaluating modelling uncertainty for model selection. Ecol Model 138:17–30

Sugeno M, Kang GT (1988) Structure identification of fuzzy model. Fuzzy Set Syst 28:15–33