Knowledge Extraction from Rainfall-Runoff Artificial Neural Network Models

Shubham Mittal
Department of Civil Engineering
Indian Institute of Technology Kanpur
Kanpur-208016, Uttar Pradesh, India
Email: mshubh@iitk.ac.in

Abstract—**In the past decade, the revolutionary Artificial Neural Networks (ANNs) has won many accolades for improving the benchmark predictive accuracies in a variety of data-oriented problems in different domains. Rainfall-Runoff (RR) modelling process of a catchment (or watershed) is one such problem in Hydrology. Despite achieving better prediction accuracy compared to traditional modeling techniques, ANNs suffer from its lack of reliability in real-world applications. Usually, the reliability index of any predictive model is directly proportional to its comprehensibility score. ANNs' fundamental problem lies in their trained but random weights values which are difficult to physically comprehend or interpret. This study is aimed to prove the hypotheses if mathematically trained ANN models can explain the physical conceptual concepts inherent in the rainfall-runoff process i.e., whether there is a possibility of hidden neuron specialization or not. This has been achieved by employing qualitative and quantitative knowledge extraction techniques on the best ANN models developed for the data derived from three catchments, Kentucky River, Alsea River, and Bird Creek, at two different time scales. The results generated from this study support the preliminary hypotheses and also open a well-directed approach to knowledge extraction from trained ANN models.**

.

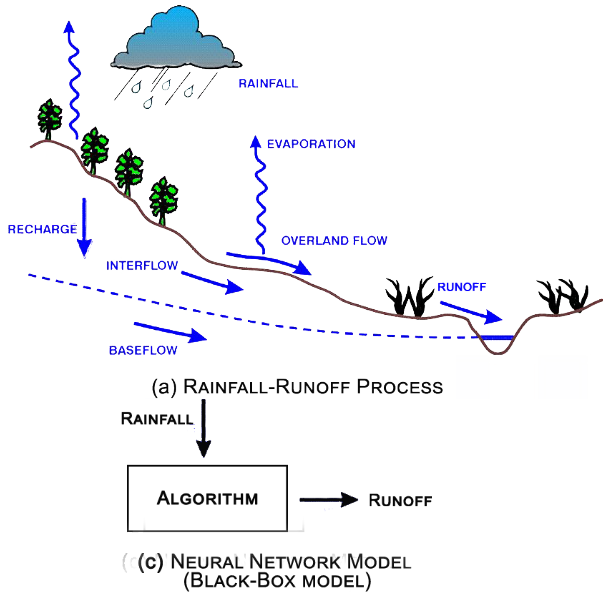
Keywords: *Rainfall-Runoff Modelling; Artificial Neural Networks (ANNs); Knowledge Extraction; Graphical Technique; Correlation Analysis*

I. Introduction

Rainfall-Runoff (RR) is a hydrologic process which can be arguably considered as one of the most researched topics in the field of hydrological processes modeling. Till now, different methodologies have been applied in an attempt to constantly model this highly complex and implicit but yet deterministic chaotic system [1]. Many researchers and engineers have tried both physical (conceptual) and mathematical modeling on RR process with both having advantages and disadvantages of their own. Modeling of this non-linear, dynamic (time-variant), and continuous process continually becomes a topic of research in the scientific community as soon as new modeling methodologies immerse. The objective of designing better hydraulic systems like culverts, dams, et al. heavily depends on a very accurate estimation of runoff in the surrounding catchments. Two of the most popularly applied modeling techniques to estimate runoff are physical modeling and mathematical modeling. In physical modeling, an output is directly estimated from analogs or theoretical simulations using the conceptual theory of the RR processes while mathematical modeling, with the help of various mathematical concepts, manipulates the interrelation (linear or non-linear) of dependent and independent variables to approximate the process' underlying complex function. Past studies show that each modeling process has been continuously scrutinized and analyzed thoroughly with a hope of an increase in their reliability and interpretability for the sensitive, costly and real-time risk-averse hydraulic systems.

Rainfall-runoff modeling, a non-linear process, estimates runoff or streamflow for any stream, river or catchment with the use of its inherent sub-processes as shown in the figure 1a. However, the relationship between rainfall and runoff is the most important. This is mainly due to the fact that rainfall data is the most influential factor in flow forecasting. And, good forecasting methods can only be developed with the most appropriate relationship between different factors of the underlying physical system which, in this case, are rainfall and runoff. Moreover, rainfall data has been abundantly (and more accurately) collected by governments, organizations, academic institutions, et al. in the past few decades or so, this availability can be efficiently as well as effectively used to fill the in-between missing values of already collected runoff records, or in extending them. Hence, for modeling purposes, the relationship between the two is extremely useful.

However, the physical process involved in RR modelling comprises an arguable level of uncertainty (because of its chaotic nature i.e., interdependency on a large number of variables including rainfall makes RR modelling more complex) due to which existing modelling techniques and results are often questioned on their reliability and hence, needed to be improved for gaining our utmost trust. This uncertainty can be attributed to the continuous human developments and the changes they bring to the hydrologic RR process. To mitigate this uncertainty as well as improve the accuracy of modeling results, researchers always play around with new theories and modeling research. One such theory is Neural Networks, a concept designed to imitate the functioning of the human brain. The cardinal concept was originally developed by *Warren McCulloch* and *Walter Pitts* during 1943 [2]. They created a simple computation model for processing large streams of data and extracting the useful insights hidden in them. The soft computing methodology, Neural networks, combine different theories from computer science, mathematics, and cognitive science and is one of the most popular predictive modeling technique of this past decade with applications in fields of both social and natural sciences. The framework of Neural network RR modeling is illustrated in the figure 1b. The input parameters (rainfall) are feed to a black-box model or algorithm which estimates the regression function between the output (runoff) and previously fed inputs (rainfall). The main advantage of Neural networks over other conventional modeling methods is its ability to easily as well as competently map the non-linear processes without even inheriting the fundamental concepts of the system being modeled. Hence, it is only reasonable to explore the application of this revolutionary technique on the rainfall-runoff modeling.



(a) Rainfall-Runoff Process

(c) Neural Network Model (Black-Box model)

Fig. 1: Rainfall-Runoff Process and its ANN Model

As of now, despite ANN's considerable benefits in RR modeling, the use of ANN models in the hydrologic application is still very much limited to the laboratories. The above-mentioned advantage of the ANNs is surprisingly one of its major drawbacks too. Being a data-driven technique, ANN doesn't concern with the ingrained details of the modeled process due to which the inherent knowledge an ANN model applies to generate a particular output is fundamentally non-interpretable. This bottleneck is still persistent in the real-world deployment of ANN RR Models and is a subject of current hydrological research.

Identifying or revealing the physics embedded in an ANN model is popularly called as *Knowledge Extraction* from trained ANNs. A plethora of research has already been conducted in this area and its initial success is enough of the reasons for more rigorous studies. Different methods, i.e, Sensitivity Analysis and Correlation Analysis, are already available to analyze a trained ANN model for interpreting the physics embedded inside them (However, there are also other methods outside the present author's limited knowledge and should be explored as per requirements). In Sensitivity Analysis of ANN models, there is an attempt to find the potential changes in the output relative to the changes in the inputs or interlinked weights of the network. But, being a multi-solution method, ANNs have different values of weights which can possibly result in nearly same outputs and that makes Sensitivity Analysis slightly non-effective as a Knowledge Extraction technique. The other method, Correlation Analysis, along with a complementary Graphical Method, is used to achieve this paper's primary objective of examining a possibility of inheritance of the modeled physical processes' conceptual components in a trained ANN model for the rainfall-runoff process. Details of the methods are illustrated in section IV.

The primary objective of this paper to explore the inherent physical significance of trained ANN RR models can only be achieved after the development of the models. So, the objective is systematically divided into two parts which are Model Development (section III) and Knowledge Extraction (section IV). ANN models are prepared with a necessary exploitation of their cardinal architectural freedom. The process of development is briefly described in section III while the examination of these developed models for the physical significance is evaluated in section IV. The current paper also presents a comparative analytical study of different spatial locations' ANN RR models conducted to explore a generalization capability or the reserved nature of both modeling and knowledge extraction aspects. The three catchments used are *Kentucky River*, *Alsea River*, and *Bird-Creek River* the details of which are presented in section III.

II. Literature Review

A. *Rainfall-Runoff Artificial Neural Network Models*

Rainfall-Runoff modeling in hydrology has been a topic of interest among its researchers since the late 19th century (*Todini*, 1988). One of the main reasons to model hydrology's rainfall-runoff process is the limitation of hydrological measurement techniques in calculating the desired hydrological parameters of different space and time [3]. Since RR modeling can be easily and purely carried out on an analytical framework i.e., estimation or extrapolation of output parameters like runoff from its input parameters like rainfall for a catchment without any reference to the involved internal processes. And, Artificial Neural Networks (ANNs) satisfies all the above-mentioned criteria and have been constantly considered a robust tool for modeling the non-linear rainfall-runoff hydrological process [4].

The earliest study for developing a rainfall-runoff neural network model was performed in the early 1990s. The application of ANNs in RR modeling started with a preliminary study by *Halff et al.* (1993) who used a three layer feed-forward ANN for the prediction of hydrographs. Since then, *Karunanithi et al.* (1994), *Hsu et al.* (1995), *Smith & Eli* (1995), *Minns & Hall* (1996), *Sajikumar et al.* (1999), *Ehrman et al.* (2000), *Birikundavyi et al.* (2002), *Jain & Indurthy* (2003), *Jain & Kumar* (2007), *Narain & Jain* (2010), and others have developed and studied ANN RR models using the collected data from real catchments [5]. All these successful studies clearly demonstrate that ANN is a powerful tool to forecast runoff in catchments which is, otherwise, quite difficult for restrained conceptual models largely because of rainfall-runoff process' non-linear and non-deterministic nature.

In an ANN model, pre-defined and relevant input parameters have to be manually selected by the researchers or modelers. In that context, *Sudheer et al.* (2002) utilized statistical properties like auto-, cross-, and partial auto-correlation to determine an optimum number of input variables to feed in an ANN model which was also used for the development of this study's ANN RR models [6].

Currently, various neural networks' variants are being used to constantly improve on still-limitations of the earlier ones. The most popular and successful ANN which has been used to model RR process is *Multi-Layer feed-forward BackPropagation Artificial Neural Networks* (*ML-BPANN*) explanation of which is given in section IIIA. *Agarwal and Singh* (2003) developed ML-BPANN models to simulate RR process for two sub-basins of Narmada River, India. The BPANN models were developed using gradient descent optimization algorithm and generalized through cross-validation [7]. Other variants of ANNs suitable for modeling the temporal data of rainfall-runoff are *Recurrent Neural Networks* (RNNs), *Long Short-Term Memory* (LSTMs), et al. These methods are relatively new in rainfall-runoff modeling and are currently a part of ongoing research. This study, however, concerns only with ML-BPANN with an objective of determining best possible models for three different catchments. Different model parameters and functionalities of ML-BPANN i.e., architecture (number of hidden layers' neurons), features, etc. are exploited to achieve this objective (more on this in section III). Most of the above-mentioned studies founded that single hidden layer ANNs are very effective in RR modeling. Therefore, our study limits only to the single layer BPANN models.

B. Knowledge Extraction from Trained ANN Models

In spite of a considerable improvement in ANNs' accuracy as well as generalization capability in runoff-estimation, there is a stigma associated with their use in real-world applications. They are still underrated mostly because of their inexplicable nature of learning (*Arbatli and Akin*, 1977) [8]. One approach to shed some light on this limitation is to extract symbolic rules from trained ANNs which can best describe how it predicts a certain output. Following on the decoding of ANNs' "black-box" nature [9], *Taha and Ghosh* (1999) presented three rule-extraction techniques [10]. Possibly deriving the concept from Expert-Systems, the first technique attempts to extract a set of binary rules from any kind of ANN. The other two are specific to feed-forward networks, with a single hidden-layer of sigmoidal activation functions. The second technique, called partial rule extraction method, extracts partial rules explaining the network's adjusted but important embedded knowledge, while the third technique, full rule-extraction method, provides a more comprehensive and universal approach. Contrary to that, this study adopts qualitative (Graphical Method) and quantitative (Correlation Analysis) methods which are more suitable for the function approximation, modeling, or forecasting application of ANNs like rainfall-runoff modeling.

Decision Trees (DTs) algorithm, one of the most widely proclaimed decision analysis tool used for "white-box" models, was used to extract decision-based rules from a trained ANN which were further employed to the interpolated data generated from training sample by *Schmitz et al.* (1997) [11]. *Setiono et al.* (2002) described another knowledge-extraction approach of using regression in the extraction of decision-based rules by dividing the input space into sub-regions and the data corresponding to those sub-regions was approximated by a linear function involving the relevant input parameters of the RR process [12].

Castro et al. (2002) interpreted trained RR ANN model in terms of fuzzy rules. The paper also studied the extraction of knowledge from two or more hidden layer ANNs [13]. *Saad and Wunsch* (2007) developed a pedagogical knowledge extraction algorithm, HYPINV, which reverses the conventional network flow with the calculation of inputs from the output (i.e., rainfall from runoff). HYPINV extracts rules in the form of hyperplanes [14].

Studies specific to knowledge extraction in hydrological processes includes *Wilby et al.* (2003) who demonstrated that architectural features like hidden neuron outputs can be interpreted by conducting correlation analysis of those features with state variables and internal fluxes (including soil moisture, evaporation, base flow, surface flow and perlocation) of a conceptual RR model [15]. It suggested that different hidden neurons of hidden layers correlate with distinct and dominant components of RR process like the base flow and surface flow. Following on the same thought-process, *Jain et al.* (2004) successfully investigated ANN river flow model for a possible relationship between its hidden layer neurons' responses and its input variables as well as deterministic components of the RR process obtained from a conceptual RR model [16].

Another novel study in order to visualize and interpret the black-box ANN RR models, *Sudheer and Jain* (2004), hypothesized a technique of mathematically mapping flow duration curve with the trained ANN model's approximated function followed by its significant implementation through a case study [17]. *Sudheer* (2005) analyzed a river flow model of Narmada Basin, India through a perturbation analysis technique to objectively establish a hidden relationship between dependent output variable and input vectors' elements [18]. The results showed that each component of input vector at different antecedent time steps influences the hydrograph's shape in different ways.

From all of the above studies, there is a clear message that more work on knowledge extraction from trained ANNs in hydrology needs to be done in order to make the ANN RR models employable and simultaneously bring improvement in hydraulic systems. In that regard, the present study investigates trained ANN RR models, developed for three different catchments, with graphical and correlation analysis of partial network outputs and hidden neuron responses with conceptual components (surface and base flows) of the RR process.

III. Model Development

A. Study Area and Statistics

The data used to develop the ANN models was derived from three different watersheds of the USA at two different time scales. The three watersheds are *Kentucky River Basin, Kentucky; Alsea River Basin, Oregon*; and *Bird Creek River Basin, Oklahoma*. The basic statistics of the data (*minimum, maximum, mean, and standard deviation*) are shown in Table I.

For Kentucky River Basin, the data includes average daily streamflow (m^3/s) at Lock and Dam 10 (LD10) near Winchester, Kentucky. With an approximate drainage area of 10,240 Km^2, daily total rainfall (mm) data of five gauges at or near Manchester, Hayden, Jackson, Heidelberg, and Lexington Airport was taken. The data time length considered was 26 years which was further divided into two parts: training set consisting daily rainfall and runoff values from 1960 to 1972 and testing dataset from 1977 to 1989.

The data utilized for Alsea River Basin near Tidewater, Oregon, the USA with a drainage area of 885 km^2 comprised of daily streamflow (m^3/s) at station number 14306500 which is located on the right bank, 1.4 km downstream from Grass Creek, 0.4 km upstream from Scott Creek, and 6 km southeast of Tidewater. For this basin, runoff was the only input variable. The data was divided into 50% for both training and testing.

The catchment of Bird Creek basin, Oklahoma, USA is located near the northern state border which is also alongside Kansas and has a drainage area of 2344 km^2 with its outlet near Sperry, about 10 km north of Tulsa city. The data sampled at 6-hour intervals contains flow (m^3/s) and rainfall (mm) with the training set covering a period from 11 November 1972 to 17 April 1974 while the testing set comprising the 18 April 1974 to 30 December 1974 data.

TABLE I Statistical Properties of Three Catchments' Data

Statistics	**Flow (m^3/s)**		**Rainfall (mm)**	
	Training	**Testing**	**Training**	**Testing**
	Kentucky			
Minimum	1.932	1.274	0.000	0.000
Maximum	2449.217	2432.425	85.751	101.092
Mean	102.734	100.932	3.217	3.249
Standard Deviation	171.025	168.371	6.696	6.734
	Alsea			
Minimum	1.560	1.980	-	-
Maximum	373.240	434.000	-	-
Mean	30.500	37.490	-	-
Standard Deviation	43.370	55.580	-	-
	Bird Creek			
Minimum	0.340	0.280	0.000	0.000
Maximum	1290.480	1505.560	43.500	74.500
Mean	36.440	48.510	0.810	1.120
Standard Deviation	95.140	133.560	3.340	4.740

B. ANN Model Development

1) Artificial Neural Network (ANN): [19]An *artificial neural network* (ANN) is a modelling technique developed to mimic a ”brain-like” system of interconnected processing units (called *neurons*), first proposed by *McCulloch* and *Pitts* in 1943 [3], that learn patterns from available historical data (containing input and output variables) and predicts output for new input observations. They have vast applications for different purposes in different fields, but are hugely appraised & popular for forecasting, and classification problems. In figure 2, an illustration of a simple three-layer feed-forward ANN is given. As can be seen, an input layer with multiple neurons that depict each of different input variables feeds information (input variables data) to the middle layer and middle layer processes the information to further direct it to an output layer which finally produces the desired outputs in the output layer. Now, input or any other layer does not directly feed raw data to the next layer neurons. Each layer neuron is connected to the next layer's neurons with their connections representing a number which signifies their weights (w_{11}, et al.) and this weight is needed to be multiplied by the incoming input response before proceeding to any of the next layer neurons. Initially, the weights are random numbers but later learned over epochs during training. Most of the research on ANNs justifies that the initialization of weights should be done in such a way that the summation of all weights should be exactly 1. This heuristic makes ANN unbiased towards every input during training.

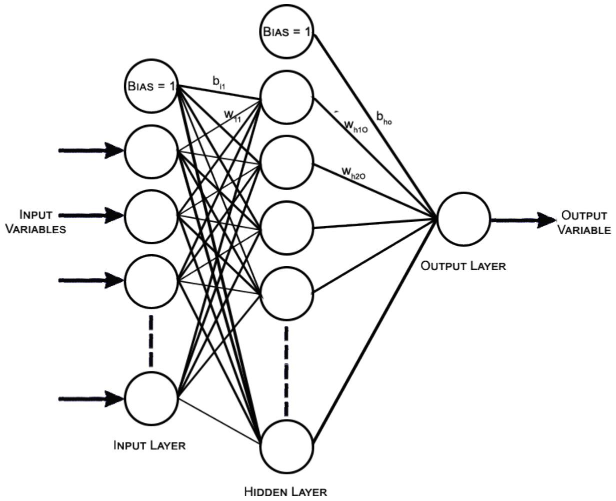


Fig. 2: Structure of a 3-layer feed-forward ANN Model

Middle & output layers, similar to the input layer, can also have one or more neurons depending upon the process. For RR modeling, only output variable is runoff. Similar to the input layer, each middle layer neuron is also characterized by a weight (w_{h1O}, w_{h2O}, et al.) mapped to output layer neurons. These weights are also initialized similar to input layer neurons and are finally learned as training epochs end. Another model parameter biases (b_{i1}, b_{i2},. . .,b_{hO} which are similar to weights and also learned during training) are multiplied by 1 and added

TABLE II Identified Input Variables Using Auto- & Cross-Correlation Analysis

Input Variables	**Basins**		
	Kentucky(daily)	**Alsea(daily)**	**Bird Creek(hourly)**
Precipitation, P (mm)	P(t), P(t-1), & P(t-2)	-	P(t-36), P(t-42), & P(t-48)
Flow, Q (m^3/s)	Q(t-1), Q(t-2), & Q(t-3)	Q(t-1), Q(t-2), Q(t-3), Q(t-4), & Q(t-5)	Q(t-6), Q(t-12), Q(t-18), & Q(t-24)
Total	6	5	7

into weighted inputs' summation. This summation is then fed to their next respective layer neurons before activation function is triggered. One bias neuron is provided in both input and middle layers (but not output layer) and taken as 1. Unlike input layer neurons, other layer neurons consist of activation functions (to be discussed later) which process the summation of all weighted inputs from input neurons and bound them. Usually, the activation function is deployed in middle layers of a network and used to bind the incoming value in order to make the processing of ANN fast and efficient. The outgoing activated values from middle layer neurons are again weighted, summed, forwarded to output layer neurons which may or may not apply activation function to finally produce the output.

A very popular ANN framework, Multi-Layer feed forward BackPropagation Artificial Neural Network (ML-BPANN) with one or more layers between input and output layer is very widely used for supervised learning of non-linear processes. ML-BPANN uses the very famous backpropagation algorithm for training or learning [20]. For RR modeling, it has been found out that one hidden layer ANNs are sufficient for good results as well as quick deployment & interpretability [21]. It can be easily seen that more hidden layers would result in more processing time and model interpretability loss. To understand the drawbacks of more hidden layers or complexity, one can refer to the theory of trade-off between variance and bias of statistical models.

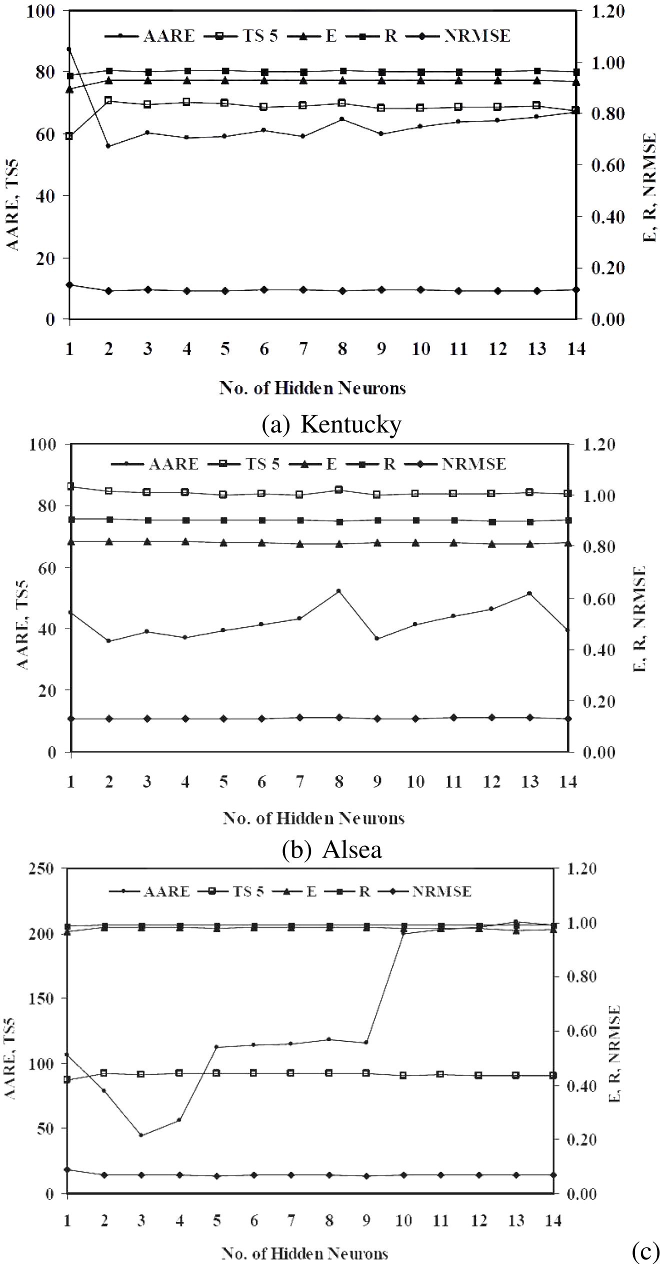
For this study, ANN models having one hidden layer and an output layer with multiple and one neuron respectively were developed. Non-linear activation function, Sigmoid [$y = 1/(1+e^{-x})$] was equipped in middle layer neurons with a linear activation function in output layer neuron. The linear activation function is a linear regression line [$y = x$] which gives output same as the input. From this, we can infer that output (rainfall) is a linear combination of inputs from hidden layer neurons with weights as parameters which will be useful for knowledge extraction part of the study.

Although there are various heuristics available to decide on a number of neurons in hidden layers to find the best model (means architecture), a trail and error method was used. ANN models with the number of hidden layer neurons varying from 1 to 14 were compared on performance using different evaluation metrics (to be described later). Since sigmoid functions have deeper descent (larger gradient values) at minimum input values, therefore, the input data was normalized between 0.1 and 0.9 for faster convergence to the desired optimum solution. For training, learning-constant of 0.005 and momentum-factor of 0.075 were used. More on model development and the result is provided in the later sections.

2) ANN Model Development: Development of ML-BPANN models for any process (like rainfall-runoff) requires stepwise procedure among which division of data into training and testing, identification of input and output variables, normalization of the data, selection of the network architecture (number of layers), determining appropriate number of hidden layer neurons, training of the network, and finally, the validation of the ANN model are more prevalent. With the division of data being done, other steps are illustrated below sequentially. Only output variable (runoff), Q(t), at time-step t was used in the development of ANN models. The input variables for the ANN models were determined based on an extensive cross-correlation and auto-correlation analysis of all the three catchments' data. The threshold autocorrelation values for Kentucky, Alsea, and Bird Creek of 0.5, 0.5, and 0.6; and cross-correlation values (with rainfall) for Kentucky and Bird Creek of 0.27 and 0.46, respectively, were used for the determination of input variables. Input variables for the three catchments were found out as mentioned in Table II.

With the identification of the relevant input and output variables; consideration of only single layer ANN models; division of the catchments' data; and scaling of data from 0.1 to 0.9, initial four steps of ANN model development are completed. Now, for the determination of an optimum number for the hidden layer neurons, a trail and error method was employed with a performance comparison of each catchment's trained ANN models having *I-N-1* architecture where I is the total number of input variables (I = 6, 5, and 7 for Kentucky, Alsea and Bird Creek, respectively), N is the number of hidden layer neurons varying from 1 to 14 and 1 is the number of output variable. As previously mentioned, back propagation algorithm was used for model training with learning constant as 0.005 and momentum factor as 0.075.

The five models' performance evaluation metrics used for this study are Sum Square Error (*SSE*), Average Absolute Relative Error (*AARE*), Coefficient of Correlation (*R*), Nash-Sutcliff Efficiency (*E*), Normalized Root Mean Square Error (*NRMSE*), and Threshold Statistics (*TS*). These are commonly used error statistics and interested readers would find their detailed description in studies by *Jain et al.* (2001). During training, the acceptable level of SSE was fixed at 0.0001 and the maximum iterations were limited to 50,000. Optimum architectures of catchments' ANN model were found by error visualization plots and error statistics. For all the three catchments, the plots between the error metrics and the number of hidden layer neurons are illustrated in figure 3. Both the



(a) Kentucky

(b) Alsea

(c) Bird Creek

Fig. 3: Visualization Plots of Error Metrics vs Number of Hidden-Layer Neurons for the three catchments

training and testing error statistics for the best ANN models are also presented, in Table III.

For Kentucky River Basin, 6-2-1 architecture, with the least SSE value of 0.00022 and 0.00027 during training and testing respectively, was the unanimous winner and hence, selected. During both training (fig.3a) and testing results, it was evidently prevalent that the chosen 6-2-1 ANN model had the preferable maximum values of TS5 (70.64 %, 72.64 %), E (0.930, 0.914) and R (0.965, 0.954) and minimum values of AARE (55.84, 57.67) and NRMSE (0.110, 0.124) amongst all of the possible 6-N-1 models.

In case of Alsea River basin, most of the metrics except AARE are not visually comparable as per fig.3b. Observing the objective least value of AARE during training, models having a number of hidden neurons 2, 9 and 14 with AARE values 35.94, 36.72 and 39.32 are deemed to be the best candidates. With further examination during testing, it was found that 5-2-1 didn't have the least AARE value with 34.88 compared to other architectures (least was 29.77 for 5-9-1). But, considering the parsimony principle and simplicity law, 5-2-1 was considered for further proceedings [22]. During training (fig.3c), R, E, TS5, and NRMSE values are not helpful in clearly selecting the optimum architecture of ANN models for Bird Creek catchment. Although based on the lowest AARE (44.42) value of model 7-3-1, it seemed a clear choice. Moreover, this structure choice was further verified with a second-best AARE value of 76.28 as compared to not much lower 75.34 of 7-4-1 during testing. So, in the end, 7-3-1 ANN model was selected for the third and final basin.

TABLE III Statistical Results of Three Catchments' Best ANN Models

Models	**Error Metrics**						
	NMRSE	**E**	**R**	**AARE**	**TS5**	**TS10**	**TS25**
	During Training						
6-2-1	0.110	0.930	0.965	55.84	70.64	86.91	98.36
5-2-1	0.130	0.821	0.906	35.94	85.47	93.00	98.21
7-3-1	0.068	0.981	0.992	44.42	91.59	96.83	99.47
	During Testing						
6-2-1	0.124	0.914	0.954	57.67	72.64	86.71	97.70
5-2-1	0.198	0.836	0.915	34.88	71.29	87.29	96.29
7-3-1	0.076	0.990	0.995	76.28	91.88	96.76	99.71

IV. Knowledge Extraction

Although ANNs have a well-proven capability of solving different as well as highly complex problems in the field of engineering, finance, medicine and many others. But, they still suffer from its cardinal "Black-box" nature. Hence, it is necessary to pursue research related to extraction of hidden knowledge embedded in the trained ANNs. Knowledge Extraction for function approximating and forecasting ANNs can be simply understood as an attempt to expose the internal mechanism which ANN models uses to reach a certain output for the given input i.e., an explanation of the approximated function generated by the ANNs. There are many approaches available to explore the physics of a trained ANN model and the examination of hidden neuron outputs' different forms is one of the widely employed one. In this study, different forms of the hidden neuron outputs from three ANN RR models, previously developed for the three catchments, are analyzed with an objective of extracting important and useful knowledge by using model's and preliminary-process' information.

This study is heavily dependent on an initial logical assumption which states that hidden layer neurons of a trained ANN model might represent the modeled physical system's different sub-processes. For example, the two hidden layer neurons of

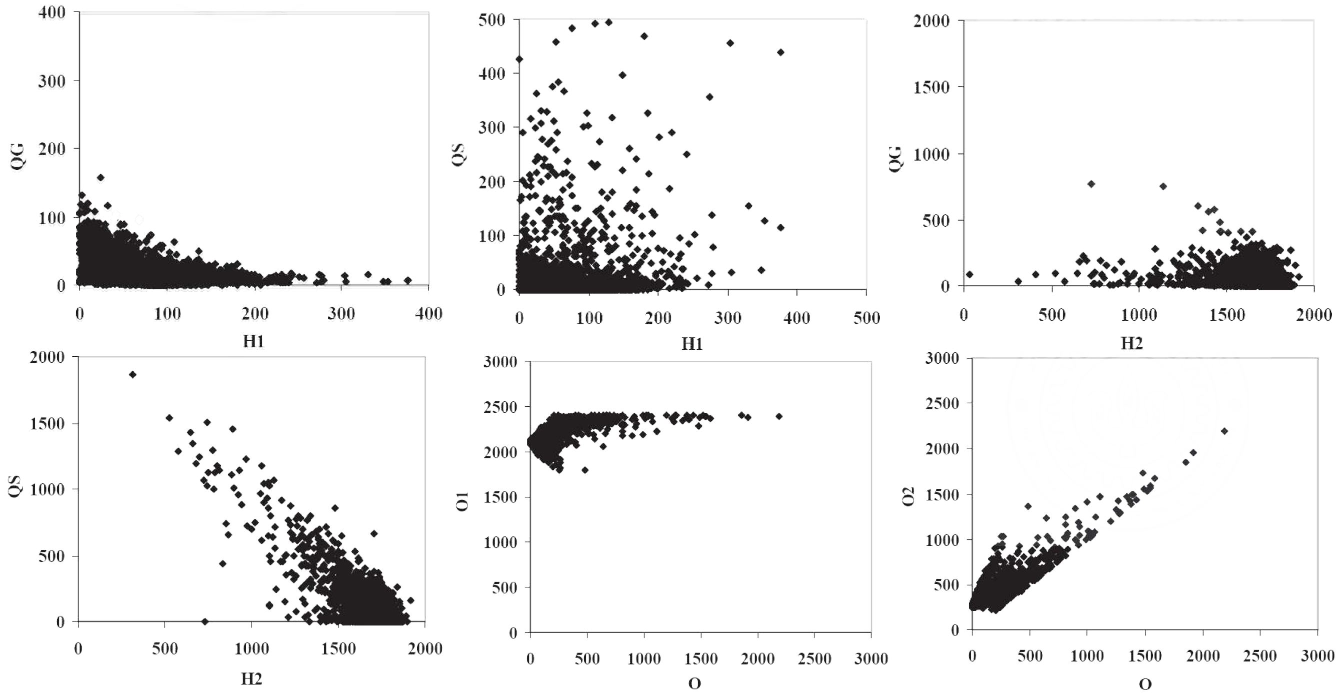


Fig. 4: Scatter Plots of Hidden Neuron Outputs vs Base, QG; and Surface, QS flows for Kentucky River Basin

6-2-1 ANN model of Kentucky River basin may be separately modeling the internal sub-components of the rainfall-runoff process. Now, these sub-components can be either infiltration, base flow, sub-surface flow, evaporation, soil moisture or others. However, the purpose of this study is limited to only two major conceptual components of the rainfall-runoff process which are surface flow *QS(t)*, and sub-surface (or base) flow *QG(t)*. A linear addition of these components forms the observed streamflow QO(t) i.e., $QO(t) = A*QS(t)+B*QG(t)$.

Base flow was calculated using the base flow recession concept, $QG(t) = KG(t) * QG(t-1)$ where *KG(t)* is the recession coefficient of the base flow at time *t*. The mathematical expression for KG(t) is $KG(t) = QG(t-1)/QG(t-2)$. And, then from the above observed-streamflow equation, surface flow QS(t) values were found.

The different forms of hidden neuron outputs considered in this study are hidden neuron outputs (*Hi*) and partial network outputs (*Oi*). Hidden neuron outputs, Hi, are immediate outputs from *i*th hidden layer neuron while partial network outputs, Oi, are calculated by turning off the connections between the output neuron and all the other hidden neurons except ith. Here, output neuron is provided a linear activation function which implies that partial network outputs Oi are nothing but the multiplication of hidden neuron outputs Hi (and Hb_h) and its respective connecting weight Wio (and Wb_ho) with the output neuron. Mathematical formulations of Hi and Oi are as follows:

$$Hi = \frac{1}{1+e^{-net_i}}$$

$$Oi = \frac{1}{1+e^{-[HiWio+Hb_hWb_ho]}}$$

where, $net_i = \sum_{j=1}^{m} I_j * W_{ij}$,
I_j = jth input layer neuron,
W_{ij} = weight connecting jth input neuron to the ith hidden layer neuron,
m = total number of input layer neurons,
Hb_h = hidden layer bias output,
Wb_ho = weight connecting hidden layer bias neuron to the output neuron,
Wio = weight connecting ith hidden neuron to the output neuron.

The knowledge extraction methods, Graphical Techniques and Correlation Analysis, used in this study considered either hidden neuron outputs (Hi) or partial network outputs (Oi) or both. The detailed investigation of the selected three ANN models for exploring their physical significance through the two methods is described below.

A. *Graphical Technique*

In the graphical techniques, insights about physical conceptual components' (sub-processes) significance which might be possibly captured by the above best ANN models chosen for the three catchments were found. For all the three models, scatter plots were generated between Hidden neuron outputs, Hi and Surface flow, QS(t); Hidden neuron outputs, Hi and Sub-surface flow, QG(t); and, Partial network outputs, Qi and Network output, O(t). These plots are provided and discussed in the following sub-sections for each of the catchments.

1) Kentucky River Basin: Figure 4 shows the scatter plots for the 6-2-1 ANN model of Kentucky river basin. Before inferencing the graphs, let's establish an obvious but important

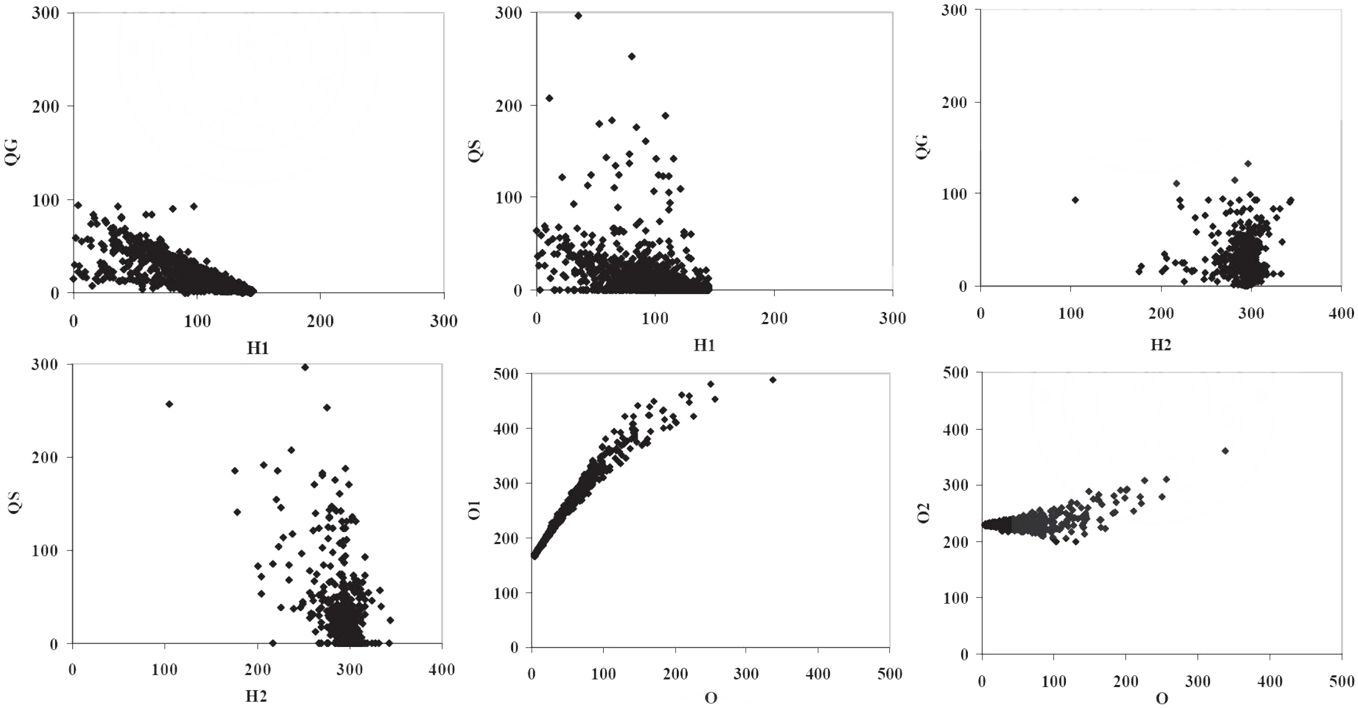


Fig. 5: Scatter Plots of Hidden Neuron Outputs vs Base, QG; and Surface, QS flows for Alsea River Basin

characteristic of stream flow. At the outlet of the catchment, stream flow is generated majorly from two components that are base flow and surface flow (QG and QS respectively). Now, surface flow comprises a major portion of the end flow of the stream while base flow is smaller in magnitude which is almost self-understood because of its dense traveling medium.

Based on the plots of H1 with QS and QG, it is observed that H1 is not correlated with the surface flow and hence, might have a plausible relationship with the other conceptual component, base flow. From the other scatter plots involving H2, it can be evidently concluded that H2 is only related to surface flow if any. This claim is also backed by a strong negative correlation of it with the surface flow. Plots of network's output O with its partial outputs O1 and O2 points out that partial output O2 is strongly correlated with O but only at larger values and O1 with O at smaller values. Since larger values of surface flow influence the stream flow much more as compared to larger values of base or ground flow, one can inference that O2 signifies the presence of surface flow in its numeric values. Hence, it is finally concluded that H2 related to O2 is modeling the surface flow.

A most interesting observation which is striking enough for more valid confirmation is that H1 and H2 are demarcated with H1 bounded between 0 to 400 while H2 varying from 500 to 2000. So, H1 and H2 clearly represent low and high flows corresponding to the base and surface flows. With these findings, there is a clear suggestion of the hidden neuron specialization in the 6-2-1 ANN model of the Kentucky river basin. The first hidden neuron H1 is observationally modeling the base flow while the surface flow is most probably being modeled by H2. These observations are further explored by the 6-2-1 ANN model's correlation analysis which is to be described later under sub-section B.

2) Alsea River Basin: Similar to the Kentucky river basin's 6-2-1 ANN model, Alsea river basin's 5-2-1 ANN model is also explored to extract the significant and useful knowledge by observing same kind of scatter plots between different forms of hidden neuron outputs (H1, H2, O1, and O2), network output (O), and rainfall-runoff process' conceptual components (QS and QG) as shown in figure 5.

From scatter plots of O1 and O2 with O, it is prominent that O has a very strong positive relationship with O1 at smaller magnitudes while a reasonable relationship with O2 at higher magnitudes exactly similar to the previous case of Kentucky. This clearly indicates that O1 and O2 are idiosyncratically modeling respective base (QG) and surface (QS) flows. The observation is further boosted by the hidden neuron outputs (preceding the partial network outputs), H1 and H2, showing a relationship with QG and QS, respectively, with no hint of correlation with the other component. Also, there is a notable and insightful demarcation of hidden neuron outputs' range between 0 to 150 for H1 and 150 to 350 for H2 which further supports the concluded hypotheses. Hence, the inferences of hidden neuron specialization for 5-2-1 ANN model of Alsea are exactly similar to the previous Kentucky case with H1 clearly modeling the base flow while H2 the surface flow. These results are again explored and testified by model parameters' correlation analysis.

3) Bird Creek River Basin: Compared to first two catchments, ANN model (7-3-1) of Bird Creek river basin has

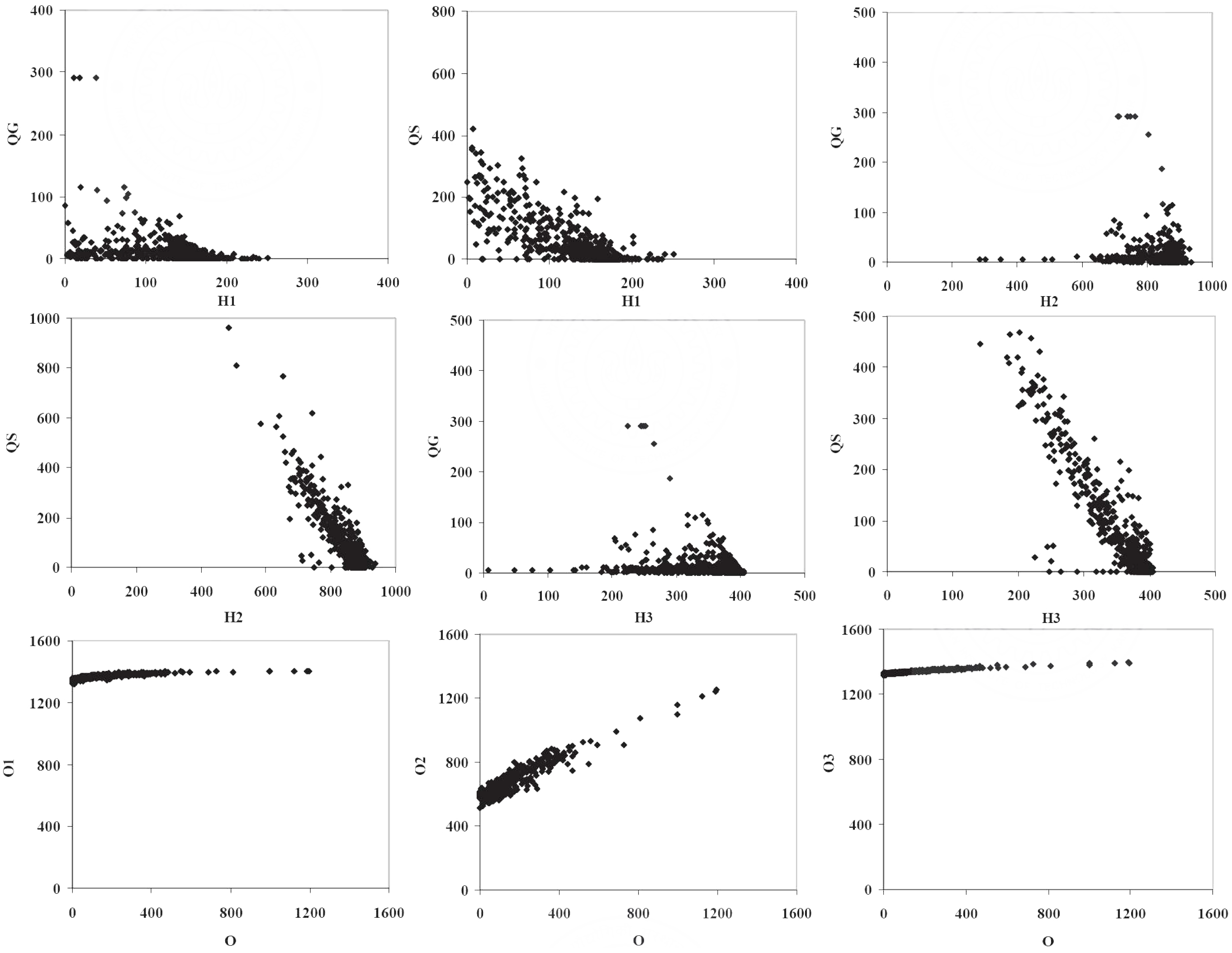


Fig. 6: Scatter Plots of Hidden Neuron Outputs vs Base Flow, QG; and Surface Flow, QS for Bird Creek River Basin

3 hidden neurons which poses a certain degree of difficulty in inferences as there were only two hidden neurons in the previous two cases. Moving forward, it can be seen from the plots that H1, H2, and H3 (Figure 6) are demarcated within 0 and 250, 300 and 1000, and 175 and 400 respectively which points towards a possibility of H1 modelling the base flow, H2 modelling the high surface flow, and H3 modelling the medium surface flow.

From Hi (i=1, 2 or 3) vs QG (or QS) scatter plots, the envelope curve over H1 vs QG is similar to a base flow recession curve, H1 vs QS shows marginal correspondence, H2 and H3 are highly correlated with QS at relatively higher and smaller values which are indicating towards the similar conclusions as made above. Further examination of O1, O2, and O3 vs O scatter plots reveal that O1 and O3 have a small correlation but only at smaller values along with a relatively steeper relationship for O1. And, O2 is clearly in a strong harmony with O. These observations finally concludes that 7-3-1 ANN model of Bird Creek had a hidden neuron specialization with H1, H2 and H3 modeling base, high surface, and medium surface flows respectively. More on this is in the correlation analysis part.

B. *Correlation Analysis*

Until now, the graphical method has resulted in several visually observational important hypotheses which are still in a need for more verification. Here, Correlation analysis is used as a decisive technique to testify the previous graphical method's findings. Correlation analysis also assumes a similar hypothesis that trained ANN model's hidden neurons represent conceptual components of the modeled physical process. The forms of hidden neuron outputs (i.e., Hi and Oi which are ith hidden neuron output, and partial network output when only ith hidden neuron is remained turned on) used to develop scatter plots in the graphical method are also used here to determine correlation coefficients with their respective parent model's input variables and study-restrained conceptual components (QG and QS). In addition to that, correlation coefficients of Oi and O are also calculated. The results for each catchment are displayed in tables 4, 5 and 6 (for Kentucky, Alsea and Bird Creek, respectively) which are also supported by subsequent explanations or data inferencing.

1) Kentucky River Basin: In 6-2-1 ANN model of Kentucky river basin, there are 6 input variables (flow at t-1, t-2, & t-3 and rainfall at t, t-1, & t-2) and one output (flow at t, O(t)). From the correlation coefficients of these variables (Table IV) and two conceptual components (QS(t) and QG(t)) with hidden neuron outputs (H1, H2, O1, and O2), the present author observes that H1 and H2 are comparatively better correlated with runoff and rainfall input variables respectively. Because rainfall is the primary source of most of the surface flow, H2 might be modeling the surface flow which is responsible for producing the flow hydrograph's rising limb.

TABLE IV Correlation Statistics for Kentucky River Basin

Variables	Hidden Neuron Outputs		Partial Network Outputs	
	H1	H2	O1	O2
Q(t-1)	-0.8221	-0.6583	0.7902	0.6588
Q(t-2)	-0.8280	-0.3859	0.7966	0.3743
Q(t-3)	-0.7624	-0.2363	0.7397	0.2221
P(t)	0.1558	-0.5951	-0.1809	0.5725
P(t-1)	0.0015	-0.7405	-0.0272	0.7332
P(t-2)	-0.3101	-0.3868	0.2973	0.3888
QG(t)	-0.6071	-0.3349	0.6041	0.3128
QS(t)	-0.5494	-0.8232	0.5231	0.8412
O(t)	-	-	0.6733	0.8822

Moreover, the correlation strengths of H1 with QG and H2 with QS are high (-0.6071 and -0.8232) compared to the reversed ones (-0.5494 and -0.3349), and this further indicates that H1 and H2 might separately be modeling base and surface flows. The preliminary inference is also strengthened by partial network output O1's poor correlation values with rainfall input variable (ranging from -0.1809 to 0.2973) compared to that of runoff (from 0.7397 to 0.7966), and O2's strong correlation with the surface flow (0.8412) compared to O2 with the base flow (0.3128). Hence, the correlation analysis' inferential findings are clearly consistent with the previous results of the applied graphical method on the 6-2-1 ANN model of Kentucky catchment. Now, the conclusions can be drawn that hidden neuron outputs H1 and H2 are separately modeling the base and surface flows of the rainfall-runoff process.

2) Alsea River Basin: 7-2-1 ANN model of Alsea river basin has only runoff input variables at time steps t-1, t-2, t-3, t-4, and t-5. And, from the correlation coefficients of hidden neuron outputs as given in Table V, Hi and partial network outputs, Oi with these input variables, it is observed that both H2 and O2 are very strongly correlated with Q(t-1) and QS(t) having values -0.5020, -0.5392, 0.5176 and 0.5439 as compared to coefficients with other runoff inputs and base flow (-0.1392 and 0.1383). These values lead to a reasoning that H2 is possibly modeling the surface flow.

And, H1 and O1 are both heavily associated with all input variables and considerably more correlated with the base flow than surface flow. All of these observations can be very-well reasoned to establish the feasibility of H1 and H2 modeling the base and surface flows. The inferences, similar to the ones made for Kentucky river basin's ANN model, are exactly consistent with the graphical method's findings and hence, can be considered as the verification of probably existent hidden neuron specialization in rainfall-runoff ANN models of different catchments.

TABLE V Correlation Statistics for Alsea River Basin

Variables	Hidden Neuron Outputs		Partial Network Outputs	
	H1	H2	O1	O2
Q(t-1)	-0.9450	-0.5020	0.9696	0.5176
Q(t-2)	-0.9210	-0.1354	0.9158	0.1298
Q(t-3)	-0.8261	0.0785	0.8362	-0.0981
Q(t-4)	-0.7664	0.1634	0.7673	-0.1833
Q(t-5)	-0.7198	0.1598	0.7099	-0.1916
QG(t)	-0.7477	-0.1392	0.7385	0.1383
QS(t)	-0.5764	-0.5392	0.5845	0.5439
O(t)	-	-	0.9749	0.5529

3) Bird Creek River Basin: The optimum ANN model structure for Bird Creek river basin's hourly data (dissimilar to the daily data of other two catchments) consists of seven input variables (i.e., runoff at t-6, t-12, t-18, & t-24 and rainfall at t-36, t-42, & t-48). First glance on these variables' and conceptual components' (QS(t) and QG(t)) correlation coefficients with hidden neuron outputs (Hi and Oi), as mentioned in Table VI, helps us make a preliminary observation that H1 is more strongly correlated with discharge and less with rainfall inputs than H2 and H3. And, same is true for partial network outputs Oi. With this, a conclusion of H1 modeling the base flow would be too far-fetched as both H1 and O1 are better correlated with QS (-0.8465 and 0.8131) than QG (-0.4134 and 0.3976). However, H1's and O1's correlation with QG is more than that with QS relative to all the other hidden neuron outputs or partial network outputs. And, although the behavior of H2 and H3 with both types of input are almost similar, they are more correlated with the surface flow (-0.9433 and -0.9433) than base flow (-0.2822 and -0.3254). Further, O2 and O3 are showing more correlation with rainfall than 01 is. So, it can be finally concluded that H1 and both H2 and H3 are modeling base and surface flows respectively.

TABLE VI Correlation Statistics for Bird Creek River Basin

Variables	Hidden Neuron Outputs			Partial Network Outputs		
	H1	H2	H3	O1	O2	O3
Q(t-6)	-0.9220	-0.9043	-0.6583	0.8902	0.8991	0.9611
Q(t-12)	-0.9184	-0.7838	-0.3859	0.8867	0.7793	0.9202
Q(t-18)	-0.8831	-0.6414	-0.2363	0.8546	0.6381	0.8446
Q(t-24)	-0.8222	-0.5023	-0.5951	0.7990	0.5005	0.7495
P(t-36)	-0.3788	-0.5113	-0.7405	0.3744	0.5055	0.5004
P(t-42)	-0.2926	-0.5779	-0.3868	0.2729	0.5839	0.5520
P(t-48)	-0.3960	-0.4933	-0.3349	0.3805	0.5044	0.4698
QG(t)	-0.4134	-0.2832	-0.3254	0.3976	0.2668	0.3376
QS(t)	-0.8465	-0.9433	-0.9421	0.8131	0.9347	0.9191
O(t)	-	-	-	0.8557	0.9572	0.9538

In addition to the above, it can be seen that H2 is possibly more interrelated with rainfall inputs than H1 and H3 can be with an only exception at P(t-36) where H3 is the most correlated hidden neuron output. Also, H2's relationship with more closer inputs i.e., Q at t-6, t-12 and t-18 is very strong compared to that of H3's, and P(t-36) rainfall would definitely influence medium surface flows because of the already spent 36 hours. High surface flow at a particular time step mostly consists the past discharge and more recent rainfalls. So, H2 and H3 are respectively modeling high and medium surface flows.

Similar to the other two catchments, the correlation analysis' findings for the Bird Creek exactly matches to that of graphical methods' and hence, strengthens the conclusion that rainfall-runoff ANN models have a hidden-neuron specialization.

V. Results & Conclusion

As far as the primary objective of this study is concerned which is to extract the hidden physics embedded in a rainfall-runoff ANN model, the current study strongly suggests that there is a definite hidden-neuron specialization during training of lesser hidden-neurons single-layer BPANN hydrological models. In case of all three river basins, the first neuron is found to be modeling the base flow, rainfall-runoff process' conceptual component with the help of conducted graphical and correlation techniques as Knowledge Extraction methods. And, the second neuron of Kentucky and Alsea river basins is modeling the other considered conceptual component, surface flow. Since Bird Creek had a three hidden neurons ANN model, a more discrete conclusion was made for the second and third neurons which are found to be modeling high and medium surface flows. It was also found that rainfall as an input plays an important role in helping ANN RR models' hidden neurons discretely capture the sub-processes of the overall physical system being modeled. The results indicate that the time scales of the modeled data has an important effect on the optimum architecture of the ANN models as well as the knowledge extracted from them.

These preliminary results are subject to studies of different catchments with varying temporal and spatial characteristics in proving ANN models' generalization capability during the knowledge extraction. And, such studies would be very exciting to conduct and also resourceful to carry forward the initial findings of this research study. In addition, there is a possible scope for a similar research study with more advanced ANN models or others like Recurrent Neural Networks (RNNs), Genetic Algorithms (GAs) [9]. It would be interesting to see the application of the currently employed knowledge extraction methods on those models. One limitation of this study was that it only explored two conceptual components among many others like evaporation, perlocation, infiltration, soil moisture. A further study including more conceptual components would definitely provide much deeper insights into the functioning of the "black-box" ANN rainfall-runoff models and help us reveal the hidden magic tricks being performed by them [23]. With similar positive revelations like this study's in the near future, we will only bring a new revolution in the development of many more (and hopefully better) useful hydrological applications.

Acknowledgment

The author would like to thank Dr. Bhaskar Dasgupta and Dr. Nisheeth Srivastava for extending this writing opportunity, and providing their enriching guidance during the course of this study.

In regard to this paper, the present author also acknowledges the immense contribution of the thesis ”Knowledge Extraction from Trained ANN Hydrologic Models” accessible at ”http://172.28.64.70:8080/jspui/handle/123456789/10374” by ”Kumar, Sumant” submitted to ”Department of Civil Engineering, Indian Institute of Technology Kanpur” in ”2008”.

References

[1] Sivakumar, B.; Berndtsson, R.; Olsson, J.; and Jinno, K. (2001), "*Evidence of chaos in the rainfall-runoff process*", Hydrological Sciences Journal, 46(1), 131-145.

[2] McCulloch, W.S.; and Pitts, W. (1943), "*A logical calculus of the ideas immanent in nervous activity*", Bulletin of Mathematical Biophysics, 5(4), 115-133.

[3] Beven, K.J. (2012), "*Rainfall-Runoff Modelling: The Primer*", 2nd ed., pp. 1-2. John Wiley & Sons, Ltd.

[4] Abrahart, R.J.; and See, L.M. (2007), "*Neural Network modeling of non-linear hydrological relationships*", Hydrology and Earth System Sciences, 11, 1563-1579.

[5] Narain, Seema; Jain, A. (2010), "*Modelling Hydrological Process using Conceptual, Neural System, and Hybrid Approaches*", Ph.D. thesis, IIT Kanpur.

[6] Sudheer, K.P.; Gosain, A.K.; and Ramasastri, K.S. (2002), "*A data-driven algorithm for constructing artificial neural network rainfall-runoff models*", Hydrological Processes, 16, 1325-1330.

[7] Agarwal, A.; and Singh, R.D. (2004), "*Runoff modeling through back propagation artificial neural network with variable rainfall-runoff data*", J. Water Resources Management, 18, 285-300.

[8] Arbatli, A.D.; and Akin, H.L. (1997), "*Rule Extraction from Trained Neural Networks using Genetic Algorithms*", Non-linear Analysis, Theory, Methods & Applications, 30(3), 1639-1648.

[9] Benitez, J.M.; Castro, J.L.; and Requena, I. (1997), "*Are Artificial Neural Networks Black Boxes?*", IEEE Transactions on Neural Networks, 8(5), 1156-1164.

[10] Taha, I.A.; and Ghosh, J. (1999), "*Symbolic Interpretation of Artificial Neural Networks*", IEEE Transactions on Data and Knowledge Engineering, 11(3), 448-463.

[11] Schmitz, G.P.J.; Aldrich, C.; and Gouws, F.S. (1997), "*ANN-DT: An algorithm for Extraction of Decision Trees from Artificial Neural Networks*", IEEE Transactions on Neural Networks, 10(6), 1392-1401.

[12] Setiono, R.; Leow, W.K.; and Zurada, J.M. (2002), "*Extraction of Rules from Artificial Neural Networks for Nonlinear Regression*", IEEE Transactions on Neural Networks, 13(3), 564-577.

[13] Castro, J.L.; Mantas, C.; and Benitez, J.M. (2002), "*Interpretation of Artificial Neural Networks by Means of Fuzzy Rules*", IEEE Transactions on Neural Networks, 13(1), 101-116.

[14] Saada, E.W.; and Wunsch, D.C. (2007), "*Neural Network Explanation using Inversion*", Science Direct, Neural Networks, 20, 78-93.

[15] Wilby, R.L.; Abrahart, R.J.; and Dawson, C.W. (2003), "*Detection of Conceptual Rainfall-Runoff Processes inside an Artificial Neural Network*", Hydrological Sciences-Journal-des Sciences Hydrologiques, 48(2), 163-181.

[16] Jain, A.; Sudheer, K.P.; and Srinivasulu, S. (2004), "*Identification of Physical Processes Inherent in Artificial Neural Network Rainfall-Runoff Models*", Hydrological Processes, 18, 571-581.

[17] Sudheer, K.P.; and Jain, A. (2004), "*Explaining the internal behavior of artificial neural network river flow models*", Hydrological Processes, 18, 833-844.

[18] Sudheer, K.P. (2005), "*Knowledge Extraction from Trained Neural Network River Flow Models*", Journal of Hydrological Engineering, 10(4), 264-269.

[19] Zurada, J.M. (1994), "*Introduction to Artificial Neural Networks*", Jaico Publishing House, Mumbai, India.

[20] Rumelhart, D.E.; Hinton, G.E.; and Williams, R.J. (1986), "*Learning representations by back-propagating errors*", Nature, 323, 533-536, doi:10.1038/323533a0.

[21] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). "*Artificial Neural Networks in Hydrology II: Hydrological Applications*", Journal of Hydrological Engineering, ASCE, 5(2), 124-137.

[22] Vos, N.J.; and Rientjes, T.H.M. (2005), "*Constraints of artificial neural networks for rainfall-runoff modeling: trade-offs in hydrological state representation and model evaluation*", Hydrology and Earth System Sciences Discussions, 2, 365-415.

[23] Tickle, A.B.; Andrews, R.; Golea, M.; and Diederich, J. (1998), "*The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded Within Trained Artificial Neural Networks*", IEEE Transactions on Neural Networks, 9(6), 1057-1068.