A Bayesian Neural Network for an Accurate Representation and Transformation of Runoff Dynamics: A Case Study of the Brazos River Basin in Texas

Hamidreza Ghasemi Damavandi[1], Dimitrios Stampoulis[1,*], John Sabo[1], Reepal Shah[1], Li Huang[1], Yuhang Wei[1], Yushiou Tsai[1], Jaishri Srinivasan[1], Tushar Sinha[2], Dragan Boscovic[3], and Glen Low[4]

[1]*Future H2O, Knowledge Enterprise Development, Arizona State University, Tempe, AZ, US*

[2]*Texas A&M Kingsville, TX, US*

[3]*Center for Assured and Scalable Data Engineering, Arizona State University, Tempe, AZ, US*

[4]*Earth Genome, Mill Valley, CA, US*

Abstract: Conventional physically based models have long yielded promising results, as they have been the main tool to depict the underpinnings of the physics governing the hydrological events. These models, however, suffer from certain issues such as the intense calibration time or the uncertainty in the estimation of hydrological variables. The development of the sophisticated data-driven techniques, and machine learning models in particular, combined with rapid increases in computational abilities (graphics processing units, computer clusters. etc.), has enabled hydrologists to utilize the data driven models in tandem with the well-established hydrological models to simulate miscellaneous environmental processes nimbly, and therefore circumvent the aforementioned conundrums associated with the physically based models. To this end, the present study aims at exploring a sophisticated neural network called variational Bayesian neural network, to improve the accuracy of physically based predictions such as runoff. Our neural network was able to accurately forecast the runoff rates with the mean Pearson correlation coefficient of $86.27\% \pm 0.0599$ within a randomly selected subset of cells in the Brazos River Basin. As these cells are selected randomly across the basin, we exclude the possibility of biasing our neural network by any specific cell. Moreover, this work for the very first time, to the best of our knowledge, suggests a similarity-based solution to transfer the learning model developed in a basin to be deployed across a different basin. In other words, there would be no need to develop a learning model for each basin from scratch. We, instead, utilize the models learnt from the previously studied basins. We cross-validated our proposed transfer learning solution via leave-one-out strategy within the grid cells of the Brazos River basin achieving a mean Pearson correlation coefficient of $85.83\% \pm 0.0592$.

Keywords: Variational Bayesian neural network, VIC Model, Similarity, Transfer-Learning, Pearson correlation coefficient.

1. INTRODUCTION

Accurate simulation of runoff rates is of crucial importance for reservoir operators, since it could serve as the fundamental indicator for the early flood warnings as well as the key point to devise a suitable water resource management scheme for dry seasons [1]. Decision-making relative to water resource management, however, necessitates the emulation of miscellaneous hydrological interventions to examine the diverse human-land-water dynamics. Conventional physically based models have long been successfully used for such purposes. These models, nonetheless, suffer from certain limitations associated with calibration processes including the physical distortion caused by an incorrect parameter tuning, and particularly, the tedious computational time [2-4]. These issues call for a robust and nimble solution to unravel the underlying intricate inter-relationships of hydrological parameters solely through probing the data. Apart from the physically based models, a variety of statistical methods has been explored to simulate the dynamics of hydrological events (Geetha *et al*., 2016 [5]; Graham *et al*., 2017 [6]; Stern *et al*., 1984 [7]; Chandler *et al*., 2002 [8]). These methods, however, often fail to meet the practical needs, as a prior input-output relationship assumption such as the order of non-linearity between the variables is required. However, environmental systems are characterized by non-linearity and heterogeneity, and therefore, a prior inter-relationship assumption might not be feasible. Specifically, one of the statistical methods that have also been implemented to improve the simulations of water dynamics is Hydrologic Data Assimilation. This approach however is also characterized by certain limitations. Combining the strengths of model estimates and remote sensing or in-situ observations, and often mitigating against their weaknesses, is achieved

*Address correspondence to this author at the Future H2O, Knowledge Enterprise Development, Arizona State University, Tempe, AZ, US;
E-mail: dstampou@asu.edu

E-ISSN: 2311-8741/20

through Data assimilation (DA). Through DA, the aforementioned sources of information are merged in the aim of increasing the spatiotemporal resolution, as well as the accuracy of the investigated variable. Hydrologic DA is a common practice, which results in the reduction of the ambiguity in model predictions, as well as the improvement over observations. A plethora of studies [9-15] show that hydrologic DA leads to statistically significant improvements in the accuracy of the models. Nevertheless, these results vary greatly among different assimilation techniques, hydrologic models, and geographical regions. As such, more studies are needed to provide further insight into novel approaches that can improve parameter calibration in hydrological modeling.

Data-driven models, and machine learning models in particular, have recently obtained immense applicability in hydrologic modeling as they tackle the conventional shortcomings in physically based models [16-18]. Several studies have investigated the potential of the application of artificial intelligence to emulate the physical-based hydrological events. To name a few, Kratzert *et al*., 2018 [19] proposed a novel data-driven approach, using the Long Short-Term Memory (LSTM) network for rainfall-runoff modeling using 241 catchments of the freely available CAMELS dataset. Hu *et al*., 2018 [1] deployed ANN and LSTM network models for simulating the rainfall-runoff process based on flood events from 1971 to 2013 in Fen River basin in china. Alizadeh *et al*., 2018 [20] compared the performance of a couple of learning models *i.e*. feedforward neural networks (FFNNs), time delay neural networks (TDNNs), radial basis neural networks (RBFNNs), recurrent neural network (RNN), a grasshopper optimization algorithm (GOA)-based support vector machine (SVM) and K-nearest neighbors (KNN) model for monthly flow prediction. Kenabatho *et al*., 2015 [21] applied artificial neural networks (ANNs), and Multiplicative Autoregressive Integrated Moving Average (MARIMA) to investigate the association between rainfall and large-scale rainfall predictors in Botswana. Fang *et al*., 2017 [22] examined the deep learning neural network to predict the soil moisture for Soil Moisture Active Passive (SMAP) satellite mission of NASA. Tokar *et al*., 1999 [23] employed an Artificial Neural Network (ANN) methodology to forecast daily runoff as a function of daily precipitation, temperature, and snowmelt for the Little Patuxent River watershed in Maryland.

Overall, the overarching goal of this work is to employ a neural network within a Bayesian learning framework (Section 5.1) to accurately forecast the runoff rates in each grid cell of a river basin and transfer these learning models to grids of other basins. Here, however, due to the lack of data availability, we validated our proposed transfer-learning model within the grid cells of the same basin (Brazos) using leave-one-out strategy [24]. This method, however, can be deployed to transfer the learning models across the cells in two different basins. As indicated by the physics-based and the empirical modelings, the temporal variability in runoff can be dominantly explained by the variability of precipitation rates (Behrangi *et al*., 2018 [25]; Tayfur *et al*., 2006 [26]; Dawson *et al*., 1998 [27]). Hence, we merely consider the rainfall rates as the fundamental driving factor for prediction. Additionally, in this work, we are interested in the time-stamps in which a considerable amount of precipitation has occurred, or equivalently the runoff is above a certain threshold (γ_r). The rest of the paper is organized as follows. Section 2 introduces the region of study. The VIC setup and the dataset used in this work are discussed in Section 3 and 4, respectively. Section 5 expatiates on the proposed methodology. Section 5.3 describes the strategy to convey the learning models developed with one basin to be deployed across different basins. The results are presented in Section 6. Section 7 concludes the paper.

2. STUDY AREA

Brazos River Basin is the 11-th longest river in the United States and the second biggest basin by area within Texas with the size of 45,000 mile^2 covering the total area of 74051 mile^2. The Brazos Basin has a combined storage of capacity of 2.5 million acre-feet. It flows 840 mile from the confluence of Salt and Double Mountain forks in Stonewall County to the Gulf of Mexico, having the largest average annual flow volume among the rivers in Texas. The afore-mentioned characteristics make this basin a perfect case study for scientific scrutiny. The Digital Elevation Map (DEM) for this river basin is shown in Figure **1**.

3. VIC SETUP

Measuring runoff rates at the basin scale remains an intriguing challenge, and often not feasible in big basins as is the case in this work. Therefore, we treated the simulated runoff rates out of our well-calibrated VIC model as the actual runoff rates and compared our predictions against the VIC-modeled output. Here, we used the VIC.4.2.c version available from https://github.com/UW-Hydro/VIC.git website. VIC calibration was performed by employing a technique

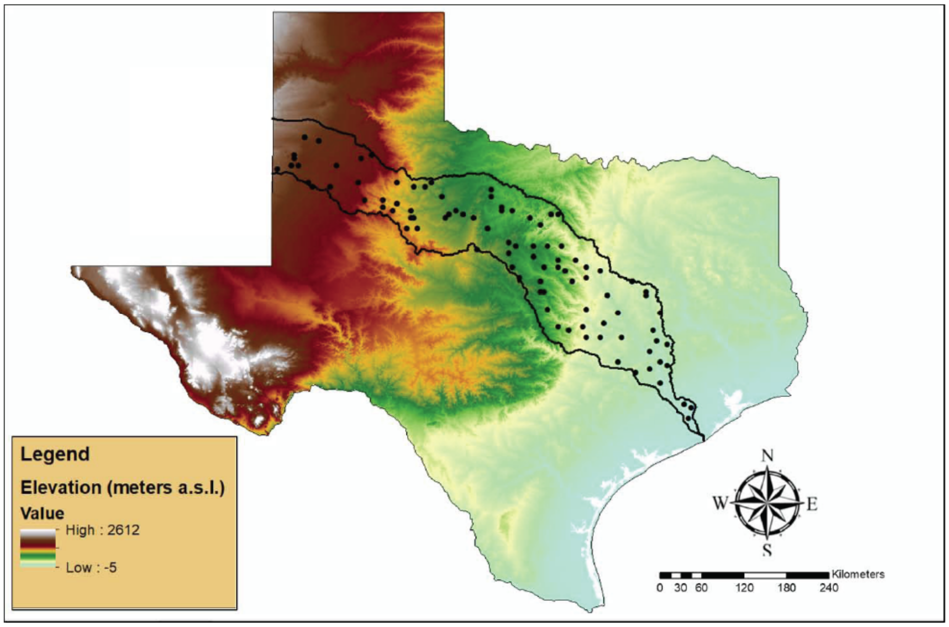

Figure 1: The Shuttle Radar Topography Mission (SRTM) Digital Elevation Map (DEM) resampled from its original resolution of 1 km to a 0.0625-degree grid for the entire state of Texas. The study area (*i.e*. Brazos River Basin) is shown with a solid black line. Black dots represent the 100 randomly selected cells used to validate the proposed method.

aiming at matching surface and subsurface runoff between a previously calibrated VIC version (4.0.3) used in Maurer *et al*. (2002 [28]) and the version used in this analysis (4.2.c). Specifically, three VIC soil parameters (the variable infiltration curve parameter, the maximum velocity of baseflow parameter, and the depth of the bottom soil layer) were optimized via the implementation of 200 Monte Carlo iterations, matching the runoff ratio between the two aforementioned versions of VIC. To avoid training neural network with zero values, we herein, take into account the days in which runoff is at least $\gamma_r = 0.1$ mm/day. For the rest of the paper, we will refer to the VIC generated runoff rates as "actual" to benchmark against the predicted runoff.

4. DATASET

The dataset utilized in this work is extracted from the publicly available Livneh database [29]. This database contains the US Continent (CONUS) near-surface gridded meteorological and derived hydrological data with daily temporal resolution spanning from 1915-2011, at the spatial resolution of 0.0625 degree and with a spatial coverage of 21.21875°-52.90625° (latitude), 235.4688°E-293.0312°E (longitude). In this study, we focused on the last five years *i.e*. 2007-2011 and used the daily time series of the meteorological parameters. We take into account a randomly selected subset of cells (a hundred cells) inside the Brazos basin and within the box of 28.9198° to 34.7323° (N) and 95.2727° to 103.8352° (W).

5. METHODOLOGY

In this section, we describe our proposed strategy to develop a learning model for each of a hundred of randomly selected cells within the Brazos River basin. We further, introduce our strategy to transfer the learning model trained for a subset of cells to accommodate for a new cell, utilizing the similarity between their hydrological parameters such as rainfall, temperature, wind speed, the three layers of soil moisture and the evaporation patterns (Section 5.3). The reason for choosing a subset of randomly selected cells, as opposed to using them all, was to ensure that the cells are scattered across the whole basin, ensuring accounting for all possible hydroclimatic conditions, and thus, for a given cell, its similar cells would not necessary be chosen as the nearest cells in Euclidean distance.

5.1. Variational Bayesian Neural Network

In this section, we discuss and utilize the recently introduced variational Bayesian neural network, *i.e*. Bayes by Backprop [30], towards building a neural network for runoff forecasting from rainfall rates. In this section, we briefly discuss the idea and the method proposed in [30]. The interested readers are encouraged to read [30] for a comprehensive explanation of the method. The classic Multi-layer neural network aims to optimize the weights over the neurons across the network towards an optimal mapping function from input features to the target. This approach, however, performs well with the presence of large amount of input data, and reveals uncertainty in regions with small amount of data. This issue promotes the application of Bayesian learning to neural networks, introducing probability distributions (Gaussian distribution in this work) over the weights of the network. As suggested by [30], instead of having a fixed value, the weights in a neural network should be extracted from a probability distribution function (PDF). In other words, the learning model is now trained using a multitude of networks in which the weights are drawn from a probability distribution function. This would then make the model reliable against the miscellaneous perturbations of the weights.

Mathematically speaking, given a dataset= $\{x_i, y_i\}^N{}_{i=1}$, we construct the probability [31] function conditioned upon the network weights as follows:

$$p(D \mid w) = \prod_i p(y_i \mid x_i, w) \tag{1}$$

with the probability density function and the network weights represented as $p(\cdot)$ and w, respectively. The optimal weights can now be achieved using the maximum likelihood [32] of the network weights:

$$w^* = \underset{w}{\arg\max}\ p(w \mid D) \tag{2}$$

Using the Bayesian theorem [33], $p(w/D)\,\alpha\, p(D/w)\,p(w)$ and hence, posterior distribution is a function of the distribution of the network weights. Therefore, the a posterior estimate [34] of $p(w/D)$ with a regularization term [35] of $\log\ p(w)$ - to avoid the overfitting [36] - is given by:

$$w^* = \underset{w}{\arg\max}\ \log p(w \mid D) \equiv \arg\max\ \log p(D \mid w)p(w) + \log p(w) \tag{3}$$

Equation 3 computes the point estimate of the weights within the network. Additionally, $p(w/D)$ is not tractable in a neural network, and hence, the attempt is to find the parameter θ of a distribution on the weights $q(w/\theta)$ (commonly referred to as the variational posterior) that minimizes the Kullback-Leibler divergence (KL divergence [37]) with the true posterior:

$$\begin{aligned}\theta^* &= KL[q(w \mid D) \parallel p(w \mid D)] \\ &= \underset{\theta}{\arg\min} \int q(w \mid \theta) \log \frac{q(w \mid \theta)}{p(w)p(D \mid w)} dw \\ &= \underset{\theta}{\arg\min}\ KL[q(w \mid \theta) \parallel p(w)] - E_{q(w|\theta)}[\log p(D \mid w)] \\ &= \underset{\theta}{\arg\min}\ F(D \mid \theta)\end{aligned} \tag{4}$$

Therefore, the to-be-minimized loss function is:

$$F(D,\theta) = KL[q(w \mid \theta) \parallel p(w)] - E_{q(w|\theta)}[\log p(D \mid w)] \tag{5}$$

The loss function is now estimated by the Monto Carlo sample from the variational posterior $p(D/w)$:

$$F(D \mid \theta) \approx \sum_{i=1}^{N} \log q(w^i \mid \theta) - \log p(w^i) - \log p(D \mid w^i) \tag{6}$$

where w^i is the $i-th$ Monte Carlo (MC) sample from the variational posterior. We used 500 MC samples in our simulations. The first term in Equation 6 is the variational posterior, which we consider a Gaussian distribution with the parameters of μ and σ^2 and hence the log-posterior would be:

$$\log q(w \mid \theta) = \sum_i \log N(w_i \mid \mu, \sigma) \tag{7}$$

For the second term in Equation 6, we consider weighted mixture of Gaussians for the prior of network weights:

$$p(w) = \prod_{i=1}^{\kappa} \pi N(w_i \mid 0, \sigma_1) + (1-\pi) N(w_i \mid 0, \sigma_2) \tag{8}$$

where k indicates the number of neurons or weights in the neural network and $\pi \in [0,1]$. In addition, we assume the weights are identically and independent distributed, and hence, the joint probability of them is equal to the multiplication of their distributions. The log-prior of $p(w)$ will be:

$$\log p(w) = \sum_i \log\left(\pi N(w_i \mid 0, \sigma_1)\ +\ (1-\pi) N(w_i \mid 0, \sigma_2)\right) \tag{9}$$

In this work, we use, $\pi = 0.2, \sigma_1 = 1$ and $\sigma_2 = 0.01$ for our simulations. Finally, a softmax layer is used to represent $p(D/w)$ (the third term in Equation 6). It is noteworthy to mention that the optimization should be solved for $\theta = (\mu, \sigma)$ of the variational posterior *i.e.* $q(w/\theta)$ to minimize the loss function $F(D,\theta)$. We, however, should ensure σ is always non-negative as it represents the variance of a Gaussian distribution. To this end, the σ is expressed as a function of a parameter ρ using a softplus function defined as:

$$softplus(\sigma) = \log[1 + e^{\rho}] > 0 \quad \forall \rho \tag{10}$$

Ultimately, we have used a shallow neural network with three hidden layers. For each layer we used ReLu activation function defined in Equation 11. Adam optimizer [38] with the learning rate of 0.01 is used for network backpropagation optimization.

$$\mathrm{Re}\,lu(x) = \max(0, x) \tag{11}$$

5.2. Anomaly Detection

As discussed in Section 3, we validate our proposed method via simulated runoff rates using our well-calibrated VIC setup. However, these runoff rates could have small simulation error propagating across the neural network, leading to a poor performance of our forecasting model. We, herein, employ an anomaly detection technique (explained in [39]) based upon the underlying distribution of VIC generated runoff rates to detect and remove these outliers. We assume the joint distribution of runoff and rainfall obey a Gaussian distribution. We then define an outlier as a data point, which has a low value in this joint distribution, and hence could be abnormal. We first construct the mean and standard deviation of all features *i.e*. rainfall and runoff rates using Equations 12 and 13, respectively. The joint Gaussian distribution is then given by Equation 14. This technique can be summarized in three steps:

1. For a feature x_i, we fit a Gaussian distribution with the mean and standard deviation of μ_i and σ_i, respectively.

$$\mu_i = \frac{1}{m}\sum_{j=1}^{m} x_i^j \tag{12}$$

$$\sigma_i^2 = \frac{1}{m}\sum_{j=1}^{m} (x_i^j - \mu_i)^2 \tag{13}$$

2. The joint probability distribution of features is calculated as:

$$p(x) = \prod_{i=1}^{N} p(x_i; \mu_i, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{14}$$

3. Given a new data point x^*, we claim x^* as an outlier if $p(x^*) \leq \varepsilon$

In this work, we have two features ($N = 2$) corresponding to rainfall (x_1) and runoff (x_2), respectively. We, herein, detect and remove the outliers for each cell individually and m represents the number of runoff data points in each cell. Obviously, m would change from a cell to another, as the number of data points varies across the cells after applying the runoff threshold (γ_r). Figure **2** shows the rainfall-runoff scatter plot for two cells in which the detected outliers are marked in red. It is worth mentioning that the optimal value for the $\in$ could be achieved by cross-validation using a ground-truth data. In this work, the optimal $\in$ for each cell is selected from the set $S_{\in} = \{0.1, 0.01, 0.001\}$ leading to the maximum

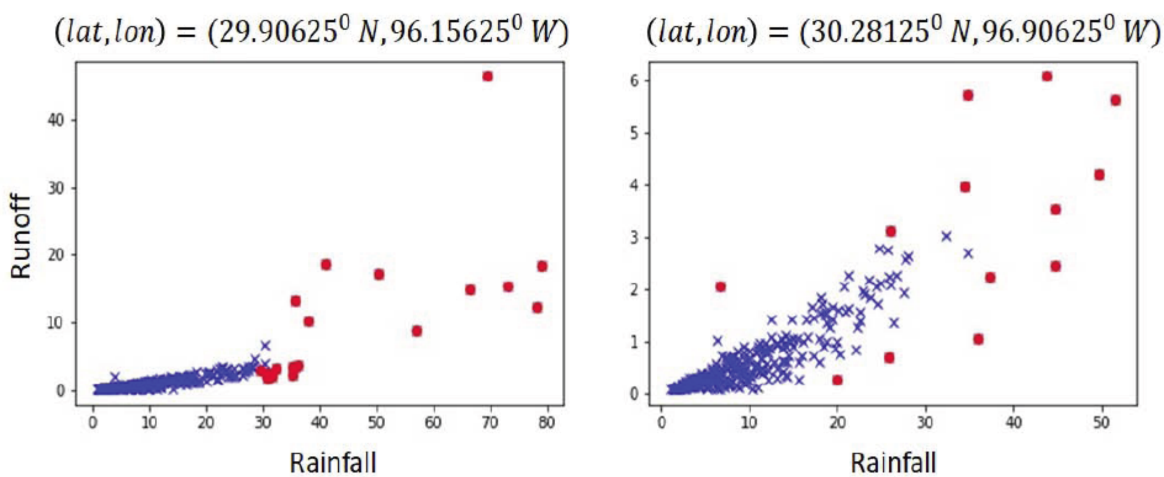

Figure 2: Rainfall-runoff scatter plot for two cells located at (29.90625° N, 96.15625° W) and (30.28125° N, 96.90625° W), respectively. The detected outliners using the algorithm in Section 5.2 are depicted with red markers. The optimal value for both of these cells are $\in^* = 0.001$.

correlation coefficient between the actual and predicted runoff:

$$\varepsilon^{*}(lat,lon) = \underset{\varepsilon \in S_{\varepsilon}}{\arg\max}\ \rho(r,\hat{r},lat,lon,\varepsilon) \qquad (15)$$

Figure **3** represents the histogram of the frequency of optimal $\in$′ s across the whole cells. As shown here, most of the cells have the maximum prediction correlation coefficient with $\in^{*} = 0.001$.

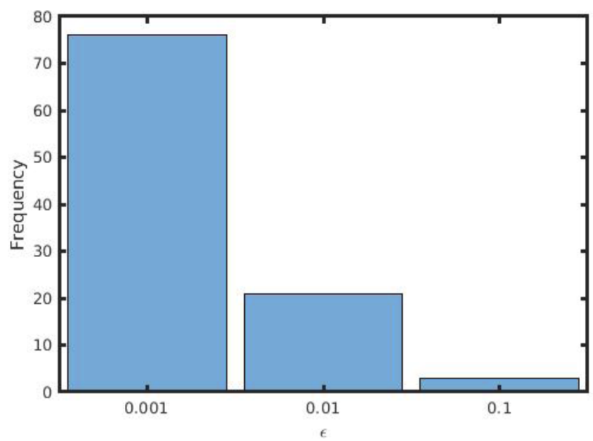

Figure 3: The histogram plot of the frequency of optimal $\in$. As shown here the optimal $\in$ for most of the cells is 0.001.

5.3. Transfer Learning

Hydrologists have observed the broad scope of variant machine and deep learning models to emulate miscellaneous hydrologic-hydrodynamic conditions. Although the benefits of these models have been extensively examined, the reliability implications regarding the deployment of one model across various hydrologic conditions, for example across different basins, have not been scrutinized. These models perform well under the base assumption that the training and testing data are governed by an identical feature space and distribution (Dai *et al*., 2009 [40]). Therefore, they frightfully fail to generalize to the testing examples, which are different from the ones encountered during training, as they have inherited the bias of the training data. Now, a model, or equivalently the input-output inter-relation, trained for one specific river basin with exclusive hydrologic characteristics, will not necessarily serve for another river basin. In this work, we focus our investigations to devise an engine, which is adaptable to cater for miscellaneous river basins with variant hydrologic characteristics. This strategy would in turn expedite the modeling procedure, and thus, further hone the learning skills of the engine to be applicable across basins with various hydrologic conditions. To this end, we employ transfer-learning literature (Pan and Yang, 2010 [41]) to leverage the learning parameters out of a pre-trained model extracted in one domain and modularize them for the newly investigated basin. In the current study, we do not perform traditional fine-tuning in the transfer learning literature. Instead, we exploit the learning models of the cells to forecast the runoff rates for the similar cells. This module would then exclude the need to build the individual learning model, from scratch, for each basin. Due to data availability limitations, we have validated our proposed transfer learning solution across the cells within a basin using leave-one-out cross-validation. This method, however, is generic and can be deployed across various cells.

5.3.1. Transfer Learning Based Upon the Similarity

As discussed earlier, we develop a neural network model for each of the cells within the basin. Now, in order to forecast the runoff rates for a new cell, we employ the “similar” cells and their corresponding models, where the similarity between two cells are given by the similarity between their hydrological parameters such as precipitation pattern, the soil moisture layers, evaporation, temperature and wind speed. Each of these parameters constitutes a temporal pattern representing a particular property of their cell. In this work, we exploit multi-dimensional dynamic time warping to compute the similarity between theses parameters across the cells. We will touch upon this method in Section 5.3.2. Once the similar cells for a particular cell like C_i is detected, the weighted average of the predictions of the learning models of similar cells are utilized to forecast the runoff rate for C_i. In the general case, the similarity between two cells are given by the similarities among their hydrological variables as well as their topographical features such as elevation and land cover. We could model such a similarity as the convex combination of the hydrological and topographical similarities:

$$s(C_i \to C_j) = \lambda s^{H}(C_i \to C_j) + (1-\lambda)s^{T}(C_i \to C_j) \qquad 0 \le \lambda \le 1 \qquad (16)$$

In which S^{H} and S^{T} are the hydrological and topographical similarities, and the parameter λ dictates the importance weight for each of these two sets of features. In this study, we focus our investigations into the hydrological features, and hence λ = 1.

5.3.2. Dynamic Time Warping

We utilize the multi-dimensional dynamic time warping (DTW) to align and compute the similarity among the hydrological variables of cells as a measure of their similarities. This method calculates the distance between two temporal sequences even if they are of different lengths in which one time-series could be non-linearly warped along its axis to get aligned to the other. The traditional time-series similarity measurements such as Euclidean distance is extremely restrictive, as they require the time-series to be of equal lengths. These methods simply measure the point distance of two time-series at the same locations, ignoring the potential temporal drift of the sequences. Here, however, we are aware that the runoff rate at one particular day could be due to the rainfall rates at prior days, and hence, DTW is an appropriate similarity measurement. As DTW is computational intense, we employ the fast DTW implementation (FastDTW). Interested readers are referred to [42] for a comprehensive explanation of dynamic time warping. Here we consider seven hydrological variables including rainfall, three layers of soil moisture, evaporation, wind speed, and the difference between maximum and minimum temperature patterns for a particular cell. The DTW between each of these features are calculated and converted to similarity using Equation 17, and the final similarity between two cells are the average of the similarities among the features (Equation 18).

$$\theta(f_i^k \to f_j^k) = 1 - \frac{dtw(f_i^k, f_j^k)}{\sum_{i=1}^{\kappa} dtw(f_i^k, f_j^k)} \quad for \quad k = 1,2,\ldots,\kappa \tag{17}$$

$$s^H(C_i \to C_j) = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\theta(f_i^k, f_j^k) \tag{18}$$

where κ is the number of hydrological parameters used to define a particular cell (here, we have considered seven hydrological parameters). Figure **4** illustrates the similarity heat-map of the cells selected for this study. The tuple (i, j) in this figure gives the similarity of cell C_i to C_j and the diagonal elements indicate the similarity between a cell and itself, and thus, has a 100% similarity. With the definition of similarity given in Equation 18, we will have:

$$s^H(C_i \to C_j) = s^H(C_j \to C_i) \tag{19}$$

And hence, the similarity heat-map is symmetric along its main diagonal.

Now, let us assume, for a particular cell C_i, we have detected the top M similar cells *e.g.* $S_{Ci} = \{C_i^1, C_i^2, \ldots, C_i^M\}$, with the corresponding learning models of $S_{\phi_i} = \{\phi_i^1, \phi_i^2, \ldots, \phi_i^M\}$. Each ϕ_i^j in $S_{\phi i}$ is a variational Bayesian neural network mapping function to forecast the runoff rate from the its rainfall rate (p_{Ci}) for cell C_i, located at (latitude, longitude) = $(lt_i, \ln_i)$ and for the day d, using the cell C_j:

$$r_{C_i|C_j}(lt_i, \ln_i, d) \approx \phi_i^j\left(p_{C_i}(lt_i, \ln_i, d)\right) \tag{20}$$

$$j \in \{1,2,\ldots,M\}$$

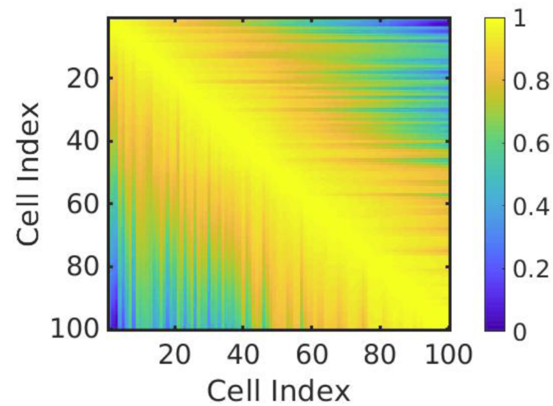

Figure 4: Similarity heatmap between the cells using Equations 17 and 18.

Then the forecasted runoff for C_i is the weighted average of runoff rates given by the learning model of each of these similar cells:

$$\hat{r}_{C_i|C_j} = \frac{\sum_{j=1}^{M} s^H(C_i \to C_j) \times r_{C_i|C_j}}{\sum_{j=1}^{M} s^H(C_i \to C_j)} = \frac{\sum_{j=1}^{M} s^H(C_i \to C_j) \times \phi_i^j(p_{C_i(\cdot)})}{\sum_{j=1}^{M} s^H(C_i \to C_j)} \tag{21}$$

With S^H defined in Equation 18. The more similar C_j is to C_i, the more its learning model would affect the forcasted runoff for C_i. The results shown in this work is presented with $M = 3$.

6. RESULT

6.1. Pearson Correlation Coefficient

Let the actual and predicted runoff at a specific latitude (lt) and longitude $(\ln)$ and for a particular day *(d)* be given by $r(lt, \ln, d)$ and $\hat{r}(lt, \ln, d)$, respectively. The Pearson correlation coefficient between $r(\cdot)$ and $\hat{r}(\cdot)$ is given by:

$$\rho(r, \hat{r}) = \frac{\mathrm{cov}[r(lt, \ln, d), \hat{r}(lt, \ln, d)]}{\sigma_r \sigma_{\hat{r}}} \tag{22}$$

$$|\rho(r, \hat{r})| \le 1$$

where the σ_r and $\sigma_{\hat{r}}$ represent the standard deviation of actual and predicted runoff rates, respectively and the $cov(\cdot)$ indicates the co-variance between the variables in its argument. The absolute value of $p(\cdot)$ is a value between 0 and 1, indicating the minimum and maximum correlation between r and $\hat{r}$, respectively. We use this metric to evaluate the performance of our learning models.

6.2. Prediction Performance

Figure **5** represents the Pearson correlation coefficient between the actual and predicted runoff rates across all hundred cells, achieving the average with one standard deviation of $86.27\% \pm 0.0599$ $(\rho \subset [0.6227; 1])$. As shown in this figure, except one cell, the rest achieved the correlation coefficient of above 0.7. In this figure the minimum, maximum and mean of these correlation coefficients are shown as, $\rho_{\min}, \rho_{\max}$ *and* ρ_{mean}, respectively.

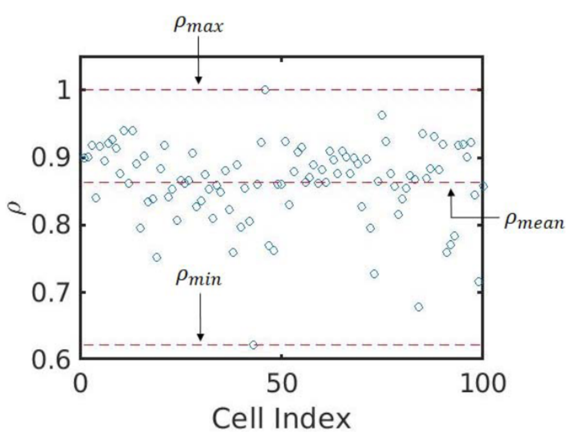

Figure 5: The Pearson correlation coefficient between the predicted and actual runoff rates for all cells. In this figure minimum, (ρ_{min}) maximum (ρ_{max}) and the average (ρ_{mean}) correlation coefficients are 0.6227, 1 and 0.8627, respectively.

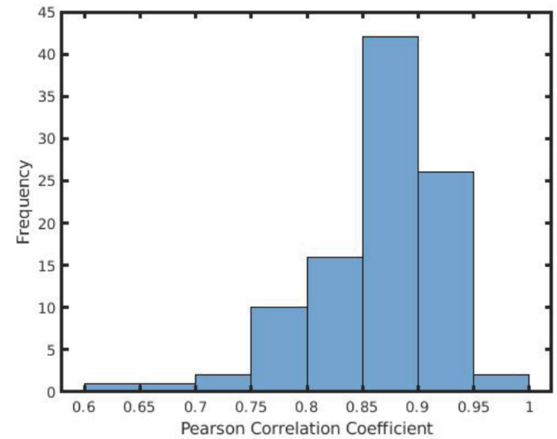

Figure 6: Histogram of Pearson correlation coefficient between the predicted and actual runoff rates.

The histogram of the correlation coefficients is shown in Figure **6**. As this figure illustrates, these coefficients are centered around 90% as an indicative of the robustness of our proposed model. Figure **7** illustrates the scatter plot of the predicted versus actual runoff for two of the top performing cells located at $(29.53125^{o}\,N, 95.71875^{o}\,W)$ and $(33.21875^{o}\,N, 99.15625^{o}\,W)$, respectively. Figure **8** illustrates the Pearson correlation coefficient between the predicted and actual runoff for each of the hundred cells using (i) trained model of the cell itself (ρ_{self}) and (ii) the transferred models of the top three most similar cells (ρ_{sim}) with the average mean absolute difference with one standard deviation of 0.0134 ± 0.012. As can easily be inferred from this figure, these two values are pretty close per each individual cell. This observation would in turn support our prior claim that the transferred models could be harnessed for forecasting, without an explicit need to learn an individual model per each cell. We, therefore, require to train our neural networks for a

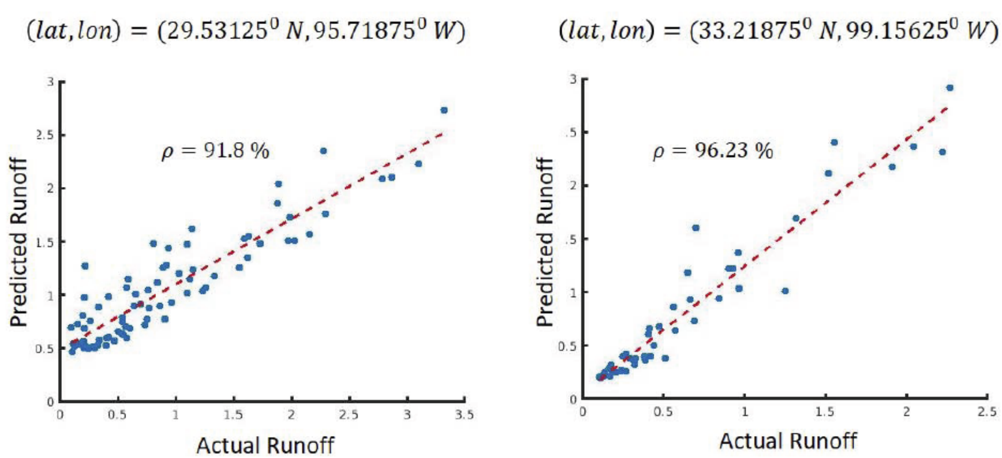

Figure 7: Scatter plot of the predicted versus actual runoff rates for the cell located at (29.53125° N, 95.71875° W) (left) and (33.21875° N, 99.15625° W) (right).

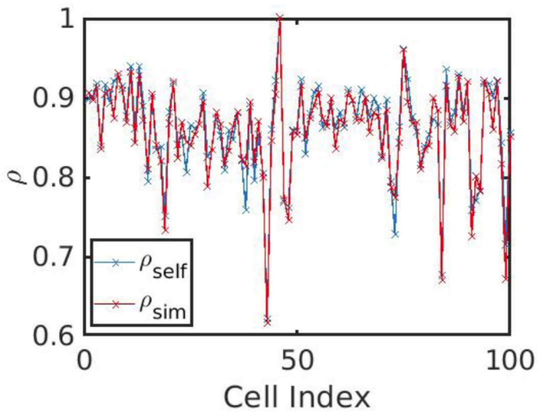

Figure 8: The Pearson correlation coefficient between the predicted and actual runoff rates for each cell, using their individual learning model and the transferred models of the similar cells.

subset of the cells within the basin, and transfer these networks to the similar cells, which indeed, reduces the computational time considerably. In an ideal case, we train the models across the hydrological cells of a big basin and transfer these models into a smaller basin having similar hydroclimatic conditions. In this work, the correlation coefficients of the transferred models have been achieved via solely considering the hydrological features. Integrating the topographical features could further improve the transferred models, representing a more realistic scenario. Moreover, once both set of the features (hydrological and topographical) are considered, the optimal weighting between them as well as an appropriate choice of similarity definition would be of paramount importance and could definitely further hone the skills of the transferred models.

7. CONCLUSION

Inspired and motivated by the recent advances in data-driven models across environmental [43, 44] and hydrological sciences [22, 45-50] as a powerful tool to approximate the physical-based models, we investigate the potential of the artificial intelligence (AI) methods, and in particular variational Bayesian neural network to predict the runoff rate using the rainfall pattern over Brazos River Basin in Texas. We reported the prediction performance as the Pearson correlation coefficient between the actual and predicted runoff rates with average correlation coefficient of $86.27\% \pm 0.0599$ cross a hundred randomly selected cells within the Brazos River Basin. Moreover, this work provides the first step towards suggesting an efficient similarity-based methodology for transferring a learning model developed for a specific hydrologic basin to another, characterized by different hydroclimatic conditions. In other words, we developed a basin-agnostic learning framework that has significant implications in improving and enhancing the representation of hydrologic conditions over different regimes. The extensibility of our work allows for implementation over catchments across the globe. Although, we cross validated our findings among the cells within Brazos, the proposed method is generic enough to be deployed across miscellaneous basins. Our leave-one-out cross-validation achieved the average Pearson correlation coefficient of $85.83\% \pm 0.0592$ Future work includes developing a similarity criterion to compare the time-series of the hydrological features. Such criterion not only captures the temporal drift of the time-series, as is the case in DTW, it will also add a parameter of importance for each data point of the time-series, and hence, propose a more realistic definition of similarity.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation award (award #1735579) and conducted at Future H2O, Office of Knowledge Enterprise Development (OKED) at Arizona State University.

REFERENCES

[1] Hu C, Wu Q, Li H, Jian S, Li N, Lou Z. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water. 2018 Oct 30; 10(11): 1543. https://doi.org/10.3390/w10111543

[2] Sorooshian S, Gupta VK. Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness. Water Resources Research. 1983 Feb; 19(1): 260-8. https://doi.org/10.1029/WR019i001p00260

[3] Beven K, Freer J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of hydrology. 2001 Aug 1; 249(1-4): 11-29. https://doi.org/10.1016/S0022-1694(01)00421-8

[4] Madsen H. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in water resources. 2003 Feb 1; 26(2): 205-16. https://doi.org/10.1016/S0309-1708(02)00092-1

[5] Geetha A, Nasira GM. Time-series modelling and forecasting: Modelling of rainfall prediction using ARIMA model. International Journal of Society Systems Science. 2016; 8(4): 361-72. https://doi.org/10.1504/IJSSS.2016.081411

[6] Graham A, Mishra EP. Time series analysis model to forecast rainfall for Allahabad region. Journal of Pharmacognosy and Phytochemistry. 2017; 6(5): 1418-21.

[7] Stern RD, Coe R. A model fitting analysis of daily rainfall data. Journal of the Royal Statistical Society: Series A (General). 1984 Jan; 147(1): 1-8. https://doi.org/10.2307/2981736

[8] Chandler RE, Wheater HS. Analysis of rainfall variability using generalized linear models: a case study from the west

of Ireland. Water Resources Research 2002 Oct; 38(10): 10-. https://doi.org/10.1029/2001WR000906

[9] Stampoulis D, Reager JT, David CH, Andreadis KM, Famiglietti JS, Farr TG, Trangsrud AR, Basilio RR, Sabo JL, Osterman GB, Lundgren PR. Model-data fusion of hydrologic simulations and GRACE terrestrial water storage observations to estimate changes in water table depth. Advances in Water Resources. 2019 Jun 1; 128: 13-27. https://doi.org/10.1016/j.advwatres.2019.04.004

[10] Girotto M, De Lannoy GJ, Reichle RH, Rodell M. Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. Water Resources Research. 2016 May; 52(5): 4164-83. https://doi.org/10.1002/2015WR018417

[11] Girotto M, De Lannoy GJ, Reichle RH, Rodell M, Draper C, Bhanja SN, Mukherjee A. Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. Geophysical research letters. 2017 May 16; 44(9): 4107-15. https://doi.org/10.1002/2017GL072994

[12] Su H, Yang ZL, Dickinson RE, Wilson CR, Niu GY. Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. Journal of Geophysical Research: Atmospheres. 2010 May 27; 115(D10). https://doi.org/10.1029/2009JD013035

[13] Houborg R, Rodell M, Li B, Reichle R, Zaitchik BF. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. Water Resources Research. 2012 Jul; 48(7). https://doi.org/10.1029/2011WR011291

[14] Forman BA, Reichle RH, Rodell M. Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. Water Resources Research. 2012 Jan; 48(1). https://doi.org/10.1029/2011WR011239

[15] Reager J, Thomas A, Sproles E, Rodell M, Beaudoing H, Li B, Famiglietti J. Assimilation of GRACE terrestrial water storage observations into a land surface model for the assessment of regional flood potential. Remote Sensing. 2015 Nov; 7(11): 14663-79. https://doi.org/10.3390/rs71114663

[16] Tripathi S, Srinivas VV, Nanjundiah RS. Downscaling of precipitation for climate change scenarios: a support vector machine approach. Journal of hydrology. 2006 Nov 15; 330(3-4): 621-40. https://doi.org/10.1016/j.jhydrol.2006.04.030

[17] Shrestha DL, Solomatine DP. Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. International Journal of River Basin Management. 2008 Jun 1; 6(2): 109-22. https://doi.org/10.1080/15715124.2008.9635341

[18] Selle B, Muttil N. Testing the structure of a hydrological model using Genetic Programming. Journal of Hydrology. 2011 Jan 24; 397(1-2): 1-9. https://doi.org/10.1016/j.jhydrol.2010.11.009

[19] Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M. Rainfall-runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences. 2018 Nov 22; 22(11): 6005-22. https://doi.org/10.5194/hess-22-6005-2018

[20] Alizadeh Z, Yazdi J, Kim J, Al-Shamiri A. Assessment of Machine Learning Techniques for Monthly Flow Prediction. Water. 2018 Nov 17; 10(11): 1676. https://doi.org/10.3390/w10111676

[21] Kenabatho PK, Parida BP, Moalafhi DB, Segosebe T. Analysis of rainfall and large-scale predictors using a stochastic model and artificial neural network for hydrological applications in southern Africa. Hydrological Sciences Journal. 2015 Nov 2; 60(11): 1943-55.

[22] Fang K, Shen C, Kifer D, Yang X. Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. Geophysical Research Letters. 2017 Nov 16; 44(21): 11-030. https://doi.org/10.1002/2017GL075619

[23] Tokar AS, Johnson PA. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering. 1999 Jul; 4(3): 232-9. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(232)

[24] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing. 2017 Sep 1; 27(5): 1413-32. https://doi.org/10.1007/s11222-016-9696-4

[25] Behrangi A, Yin X, Rajagopal S, Stampoulis D, Ye H. On distinguishing snowfall from rainfall using near-surface atmospheric information: C omparative analysis, uncertainties and hydrologic importance. Quarterly Journal of the Royal Meteorological Society. 2018 Nov; 144: 89-102. https://doi.org/10.1002/qj.3240

[26] Tayfur G, Singh VP. ANN and fuzzy logic models for simulating event-based rainfall-runoff. Journal of hydraulic engineering. 2006 Dec; 132(12): 1321-30. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:12(1321)

[27] Dawson CW, Wilby R. An artificial neural network approach to rainfall-runoff modelling. Hydrological Sciences Journal. 1998 Feb 1; 43(1): 47-66. https://doi.org/10.1080/02626669809492102

[28] Maurer EP, Wood AW, Adam JC, Lettenmaier DP, Nijssen B. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. Journal of climate. 2002 Nov; 15(22): 3237-51. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

[29] Livneh B, Rosenberg EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. Journal of Climate. 2013 Dec; 26(23): 9384-92. https://doi.org/10.1175/JCLI-D-12-00508.1

[30] Blundell C, Cornebise J, Kavukcuoglu K, Wierstra D. Weight uncertainty in neural networks. arXiv preprint arXiv: 1505.05424. 2015 May 20.

[31] Durrett R. Probability: theory and examples. Cambridge university press; 2019 Apr 18. https://doi.org/10.1017/9781108591034

[32] Myung IJ. Tutorial on maximum likelihood estimation. Journal of mathematical Psychology. 2003 Feb 1; 47(1): 90-100. https://doi.org/10.1016/S0022-2496(02)00028-7

[33] Korb KB, Nicholson AE. Bayesian artificial intelligence. CRC press; 2010 Dec 16. https://doi.org/10.1201/b10391

[34] Gauvain JL, Lee CH. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE transactions on speech and audio processing. 1994 Apr; 2(2): 291-8. https://doi.org/10.1109/89.279278

[35] Scholkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press; 2001 Dec 1.

[36] Dietterich T. Overfitting and undercomputing in machine learning. ACM computing surveys. 1995 Sep 1; 27(3): 326-7. https://doi.org/10.1145/212094.212114

[37] Hershey JR, Olsen PA. Approximating the Kullback Leibler divergence between Gaussian mixture models. In2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07 2007 Apr 15 (Vol. 4, pp. IV-317). IEEE. https://doi.org/10.1109/ICASSP.2007.366913

[38] Zhang Z. Improved Adam Optimizer for Deep Neural Networks. In2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS) 2018 Jun 4 (pp. 1-2). IEEE. https://doi.org/10.1109/IWQoS.2018.8624183

[39] https: //www.coursera.org/lecture/machine-learning/anomaly-detection-using-the-multivariate-gaussian-distribution-DnNr9

[40] Dai W, Jin O, Xue GR, Yang Q, Yu Y. Eigentransfer: a unified framework for transfer learning. InProceedings of the 26th Annual International Conference on Machine Learning 2009 Jun 14 (pp. 193-200). ACM. https://doi.org/10.1145/1553374.1553399

[41] Pan SJ, Yang Q. A survey on transfer learning. IEEE Transaction on Knowledge Discovery and Data Engineering, 22 (10). https://doi.org/10.1109/TKDE.2009.191

[42] Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis. 2007 Jan 1; 11(5): 561-80. https://doi.org/10.3233/IDA-2007-11508

[43] Damavandi HG, Gupta AS, Nelson RK, Reddy CM. Interpreting comprehensive two-dimensional gas chromatography using peak topography maps with application to petroleum forensics. Chemistry Central Journal. 2016 Dec; 10(1): 75. https://doi.org/10.1186/s13065-016-0211-y

[44] Damavandi HG, Gupta AS, Reddy C, Nelson R. Oil-spill forensics using two-dimensional gas chromatography: Differentiating highly correlated petroleum sources using peak manifold clusters. In2015 49th Asilomar Conference on Signals, Systems and Computers 2015 Nov 8 (pp. 1589-1592). IEEE. https://doi.org/10.1109/ACSSC.2015.7421415

[45] Sumi SM, Zaman MF, Hirose H. A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. International Journal of Applied Mathematics and Computer Science. 2012 Dec 1; 22(4): 841-54. https://doi.org/10.2478/v10006-012-0062-1

[46] Kim TW, Valdés JB, Aparicio J. Frequency and spatial characteristics of droughts in the Conchos River Basin, Mexico. Water International. 2002 Sep 1; 27(3): 420-30. https://doi.org/10.1080/02508060208687021

[47] Karran DJ, Morin E, Adamowski J. Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. Journal of Hydroinformatics. 2013 Nov 6; 16(3): 671-89. https://doi.org/10.2166/hydro.2013.042

[48] Khan MS, Coulibaly P. Application of support vector machine in lake water level prediction. Journal of Hydrologic Engineering. 2006 May; 11(3): 199-205. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:3(199)

[49] Damavandi HG, Shah R, Stampoulis D, Wei Y, Boscovic D, Sabo J. Accurate prediction of streamflow using long short-term memory network: A case study in the Brazos river basin in Texas. International Journal of Environmental Science and Development. 2019 Jan 1; 10(10): 294-300. https://doi.org/10.18178/ijesd.2019.10.10.1190

[50] Damavandi HG, Stampoulis D, Shah R, Wei Y, Boscovic D, Sabo J. Machine learning: An efficient alternative to the variable infiltration capacity model for an accurate simulation of runoff rates. International Journal of Environmental Science and Development. 2019 Jan 1; 10(9): 288-93. https://doi.org/10.18178/ijesd.2019.10.9.1189

Received on 20-12-2019 Accepted on 17-1-2020 Published on 8-2-2020

DOI: https://doi.org/10.12974/2311-8741.2020.08.5