

Machine Learning: An Efficient Alternative to the Variable Infiltration Capacity Model for an Accurate Simulation of Runoff Rates

Hamidreza Ghasemi Damavandi, Dimitrios Stampoulis, Reepal Shah, Yuhang Wei, Dragan Boscovic, and John Sabo

Abstract—The present study aims to investigate the performance of the artificial intelligence to emulate the conventional physically-based hydrological models. Although these conventional models could accurately depict the underlying physical processes, but they require a lengthy preprocessing phase as well as a tedious calibration time. Therefore, a need to examine the potential efficient alternative for these models is highly felt. This need becomes imperative once we adopt fine temporal and spatial resolutions for our hydrological modeling, leading to a massive number of to-be-analyzed cells. To this end, we propose a learning framework towards an accurate prediction of runoff rates using meteorological variables, and hence, mimicking the Variable Infiltration Capacity (VIC) by a nimble systematized predictive model. We also present a novel strategy to optimally select the most informative subset of data to train our predictive model, out of the pool of accessible data. This strategy would then considerably enhance the performance of our prediction in terms of computation time. We reported our result as the Pearson correlation coefficient between the predicted and actual runoff rates. Our predictive model was able to forecast the runoff rates with the mean correlation coefficient of 0.9007 for the cells within the study basin.

Index Terms—Active learning, artificial intelligence, random forests, variable infiltration capacity.

I. INTRODUCTION

Conventional physically based models such as Variable Infiltration Capacity (VIC) (Liang *et al.*, 1994) have long been applied to simulate the hydrological behavior including but not limited to the runoff flow of the basins. Although these well-established hydrological models have been the main tool to depict the physical processes, they suffer from some deficiencies such as the intense computational time and the potential un-certainties in the estimation of hydrological variables. The artificial intelligence (AI) methods have offered prominent solutions towards tackling these shortcomings, and hence, provided efficient and robust alternative for the conventional physical-based hydrological models without a need for the costly calibration time. The

data deluge from the meteorological observations, measured at sporadically located gauges across the world, calls for an efficient and universal interpreting scheme to disentangle the intricate relationship of hydrological components, and this leads to a daunting signal-processing problem as well as an interesting machine learning challenge.

Several studies have revealed the potential of the application of artificial intelligence to emulate the physical-based hydrological events. Fang *et al.* 2017 [1] examined the deep learning neural network to predict the soil moisture for Soil Moisture Active Passive (SMAP) satellite mission of NASA. Sumi *et al.* 2012 [2] studied the potential of the artificial neural network, multivariate adaptive regression splines, the k-nearest neighbor, and radial basis support vector regression for rainfall prediction of Fukuoka city in Japan. Kim *et al.* [3] proposed Hybrid models such as the combination of wavelet transform and artificial neural network (WANN) with the Conchos River Basin as the case study. Karran *et al.* 2014 [4] compared the use of four different models, i.e. ANN, SVR, waveletANN, and wavelet-SVR in a Mediterranean, Oceanic, and Hemiboreal watershed. Khan *et al.* 2006 [5] studied the prediction of lake water levels using Support Vector Regression (SVR) and Multilayer Linear Perception where the intra-model comparison showed that SVR can be competitive to MLP. Among the state-of-the-art learning models, ensemble learning methods [6], and random forests as a widely used technique, have offered robust and less variant solutions for miscellaneous prediction purposes. This class of learning algorithms use multitude of learning models to deliver a robust predictive model via reducing the probability of over-fitting. For a detailed explanation of over-fitting issue, interested readers are referred to [7].

Motivated by the afore-mentioned offerings of AI, and machine learning strategies in particular, as a powerful tool to emulate the miscellaneous interventions of the physical-based hydrological models, we investigate the performance of ensemble learning model to forecast the runoff rates using meteorological components. Due to the inherent seasonality characteristics of the hydrological signals, we also harness the active learning techniques to capture the most informative portion of the data for the training and prune off the irrelevant or repetitive information, not contributing towards the accurate prediction of the runoff rates. Hydrological modeling is rife with miscellaneous statistical techniques and various surrogate models. However, to the best of our knowledge, this is for the first time in which

Manuscript received May 13, 2019; revised July 21, 2019.

Hamidreza Ghasemi Damavandi, Dimitrios Stampoulis, Reepal Shah, Yuhang Wei, and John Sabo are with Future H2O, the Office of Knowledge Enterprise Development, Arizona State University, Tempe, AZ, USA (e-mail: {hghasemi, dstampou, Reepal.Shah, John.L.Sabo}@asu.edu, hhuweiyuhang@163.com).

Dragan Boscovic is with the Center of Assured and Scalable Data Engineering (CASCADE), Arizona State University, Tempe, AZ, USA (e-mail: dboscovi@asu.edu).

active learning is coupled by the state-of-the-art machine learning technique to predict the runoff rates while the most informative training data points are algorithmically extracted.

The rest of the paper is organized as follows. Section II briefly introduces the region of interest i.e. Brazos Basin in Texas. The dataset used in this work is introduced in Section III. Section IV presents the proposed methodology in which Section IV.a touches upon our calibrated VIC setup and Section IV.b to Section IV.e expatiate the predictive model and the strategy to selectively obtain the optimal training data (active learning). The results are presented in Section V. Section VI concludes the paper.

II. REGION OF INTEREST

Brazos 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 km² flowing 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 (Fig. 1). It also sustains various sunfish species such as blueill, longear and green, to name a few [8]. The afore-mentioned characteristics turn the basin into an appealing case study for further scientific explorations.

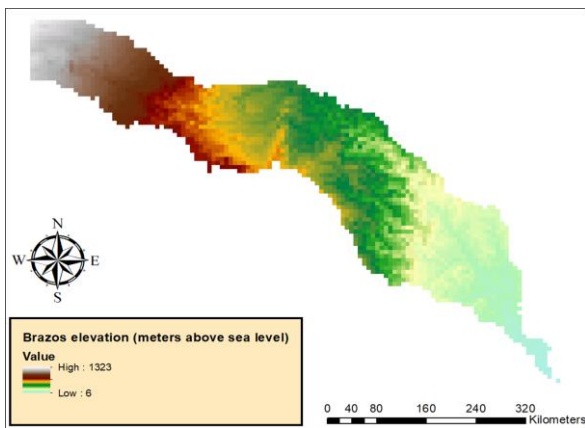


Fig. 1. The digital elevation model of Brazos Basin in Texas.

III. DATASET

The dataset utilized in this work is extracted from the publicly available Livneh database [9]. 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), 235.4688E-293.0312E (longitude). In this study, we focus on the last five years i.e. 2007-2011 and used the daily time series of the meteorological parameters, and not their long-term temporal average. We take into account the cells inside the Brazos basin and within the box of 28.9198° to 34.7323° (latitude) and -103.8352° to -95.2727° (longitude). We attempt to develop a learning predictive model for each individual cell to forecast the runoff rate for that particular location based upon its meteorological variables i.e. maximum and minimum temperature, wind speed and precipitation rate.

IV. METHODOLOGY

A. VIC Setup

Note that the actual runoff rates cannot be measured at the basin scale. 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. The VIC calibration was performed by employing a technique aiming at matching surface and subsurface runoff between a previously calibrated VIC version (4.0.3) used in Maurer et al. (2002 [10]) 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.

B. Predictive Model

Fig. 2 shows our proposed learning framework to forecast the runoff rates directly from the input meteorological data (precipitation rate, maximum and minimum temperature and the wind speed). Note that, we perform our prediction on a cell-by-cell basis. Starting from the first cell on the top left, we capture the temporal data across the five years for each of the input variables as well as the output runoff data. We would then use 90% of this data to train a model for that specific cell and evaluate the performance of the model against the remaining 10% testing data. A key distinction of our technique is the novel strategy to selectively obtain the optimal training data. Towards this goal, the active and machine learning blocks remain in a dynamic handshake with each other to jointly deliver the best predictive model (the best learning hyper-parameters) as well as the optimal training subset. Once the predictive model is learnt for a particular cell with a high prediction accuracy, we would be able to use this model to forecast the runoff rates for the future time stamps, without a need to run VIC model. We then proceed to the next adjacent cell to learn a new model for the current cell. We adopt one of the robust ensemble methods i.e. Random Forests [11] for our prediction purposes to remove the potential of overfitting issue, and hence, make our predictive model sufficiently generic to be applicable to various basins (of similar hydroclimatic conditions) across the world.

C. Random Forests

Random Forest [11] as one of the ensemble learning methods, reveals a stable learning model in which the final regressed value is achieved by the combination of the outcomes from diverse regression trees. In other words, given K potential predictors, we begin constructing multitude of decision trees by randomly choosing $N(< K)$ of the predictors and build a decision tree [12] and output the regressed value. The final output would then be some combination of these regressed values (e.g. average value). Since, the final outcome is reckoned on multiple of trees and not just biased by one tree (set of predictors), this

method is considered as a robust learning model, decreasing the variance of model and reducing the chance of over-fitting issue.

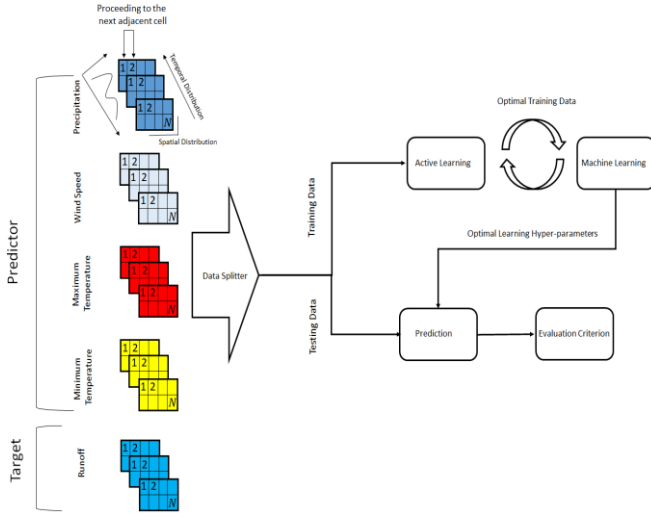


Fig. 2. The block-diagram illustrating our proposed learning framework to forecast the runoff data based upon the meteorological data (precipitation, maximum and minimum temperature and the wind speed). We start from the cell #1 to train a predictive model to forecast the runoff for this particular cell. We would then proceed to next adjacent cell to perform the same task for the current cell throughout the whole N cells of the region of interest.

D. Active Learning

In machine learning literature, active learning is a special learning strategy which attempts to provide an intelligent guidance towards exploring the most informative data points, and hence, considerably reduces the size of the required training set. Such a strategy seems crucial once collecting the training data is cumbersome or time consuming. We would then need to devise a smart strategy to extract the most informative data point for the training and set aside the un-necessary data points. This would also greatly expedite our training phase where an optimal model is learnt in a quiet time-efficient manner. A variety of strategies have been offered in the literature for this purpose such as the expected model change, pool-based sampling, uncertainty sampling, query by committee (used in this work) etc. Although the active learning method has been widely used in the classification mode [13]-[15] little has been done to harness the active learning strategies in the regression mode. In the current study, we examine a version of committee-based query active learning algorithm in the regression mode.

E. Optimal Training Data Points

Let's denote the input meteorological variables (predictors) and the output runoff data (target) as X and y , respectively. We split both the predictors and target into two disjoint training and testing sets. We randomly select 90% of the data for the training and the remaining 10% for testing:

$$\begin{aligned} X &= X^{train} \cup X^{test} \\ y &= y^{train} \cup y^{test} \\ X^{train} \cap X^{test} &= \emptyset \\ y^{train} \cap y^{test} &= \emptyset \end{aligned} \quad (1)$$

where \cup and \cap represent the union and intersection symbols, respectively, and \emptyset represents the empty set in set theory [16]. We further split the training data into two sets: the first set is the subset of training data which would definitely be used for training (denoted by X^{active}) and the second subset is the pool of remaining to-be-explored available data (denoted by X^{pool}). As mentioned earlier, the hydrological signals have an inherent seasonality or periodicity characteristics, so if we allocate the whole set of training data to learn a model, we would potentially use a massive amount of repetitive and redundant information. Using redundant information for training the machine learning model not only slows down the training phase, but would also make the model more erroneous in the testing session. We instead should selectively choose the training data points in order for the model to face miscellaneous scenarios and could perform well in the testing session. Hence, the goal is to optimally extract the most informative data point from this pool set and add it to X^{active} . We adopt the ten SVM committee-based active learning method (C_1, C_2, \dots, C_{10}) where each of the committee members is a Support Vector Machine (SVM) regression model to estimate the input value. Interested readers are referred to [17] and [15] for a comprehensive explanation of SVM in regression mode (SVR) and the pool-based active learning method.

The strategy to obtain the most informative training data points is given by the following four-step algorithm:

While $|X^{pool}| \neq \emptyset$ DO:

Step 1: Train a model using a small portion of training data or X^{al} , which means we look for the best transfer function i.e. R to map the input meteorological predictors to the runoff rates with the minimum mapping error. Random Forest algorithm would provide us with such a transfer function.

$$f(X^{active}) \xrightarrow{R} y^{active} \quad (2)$$

Step 2: Find the most informative data point from the pool and add it to the training where the most informative data point is the one "having the highest disagreement among the committee member or having the highest uncertainty among the committee members":

$$\begin{aligned} ix^* &= \arg \max_{1 \leq i \leq |X^{pool}|} d(C_1(X^{pool}(i)), C_2(X^{pool}(i)), \dots, C_N(X^{pool}(i))) \\ &1 \leq i \leq |X^{pool}|, \quad N=10 \end{aligned} \quad (3)$$

which ix^* is the index of the optimal data point to be added to training and the function $d(\cdot)$ is a measure of disagreement between the committee members for that particular data point. Here, such function is the standard deviation. In other words, the data point with the highest standard deviation of prediction of committee members will be chosen to be added to X^{active} .

Step 3: Add the informative data point to the training set and exclude it from the pool:

$$\begin{aligned} X^{active} &= \{X^{active}\} + \{X^{pool}(ix^*)\} \\ X^{pool} &= \{X^{pool}\} - \{X^{pool}(ix^*)\} \end{aligned} \quad (4)$$

Step 4: Repeat the same procedure to obtain the next informative data point (go to *Step 1*)

As the algorithm indicates, at each iteration one of the data point will be extracted from the pool and will be added to the training set. We repeat this procedure to extract the desired percentage of data to train the model. Without losing any generality, we may, for example, just need the “optimal 60% of training data” to train the model. We repeat the procedure until we have collected the desired 60% of data and would then terminate the algorithm.

V. RESULT

For each of the cells within the basin, we trained a model according to Fig. 1 and evaluated the performance of the prediction via the Pearson correlation coefficient between the actual and the predicted runoff rates while active learning extracted the optimal 60% of the training data points.

A. Pearson Correlation Coefficient

Let the observed runoff rate at particular latitude (lat), longitude (lon) and day (t) be $r(lat, lon, t)$. The predicted runoff rate also denoted as $\tilde{r}(lat, lon, t)$. The Pearson correlation coefficient between $r(\cdot)$ and $\tilde{r}(\cdot)$ is defined as:

$$\begin{aligned} \rho_{rr} &= \frac{cov(r(\cdot), \tilde{r}(\cdot))}{\sigma_r \sigma_{\tilde{r}}} \\ |\rho_{rr}| &\leq 1 \end{aligned} \quad (5)$$

where the σ_r and $\sigma_{\tilde{r}}$ represent the standard deviation of actual and predicted runoff rates, respectively and the $cov(\cdot)$ indicates the covariance between the variables in its argument. The absolute value of ρ_{rr} is a value between 0 and 1, indicating the minimum and maximum correlation between r and \tilde{r} , respectively.

B. Active Learning Gain

Fig. 3 illustrates the Pearson correlation coefficient between the actual and predicted run off rates in terms of the percentage of data needed for the training for both active and passive learning strategies. The passive learning strategy here is simply a random sampling method in which at each iteration we randomly add one data point from the pool (X^{pool}) to the training set (X^{active}). In this figure, the x-axis is the percentage of data used for training out of the pool of

available training data, or equivalently the ratio of the number of data points in X^{active} over number of data points of X^{train} . Y-axis is Pearson correlation coefficient between the actual and predicted run off rates using the corresponding percentage of training data. The active learning gain is readily observed within the curve where it requires far less training data points to result in the same performance as the passive learning strategy.

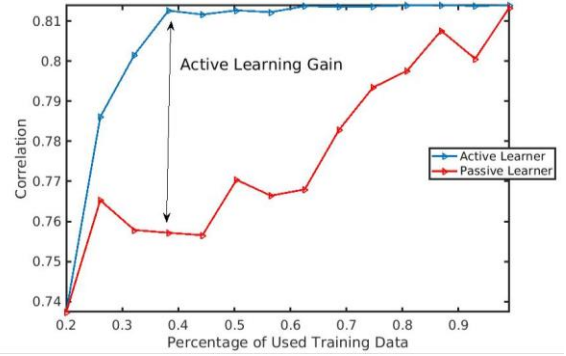


Fig. 3. Comparison between the active and passive learning in terms of the percentage of used training data points one of the cells within the basin. As shown, active learning needs far less training data to achieve the same correlation coefficient as passive learning.

C. Prediction Performance

Fig. 4(a) and 4(b) show the predicted versus actual runoff time series across 178 days (10% of the data for testing session) for two top-performing cells, the pair of latitude and longitude of (34.46875, -95.65625) and (33.40625, -96.96875), respectively. As shown in these figures, our learning framework could accurately predict both the low and the high peaks.

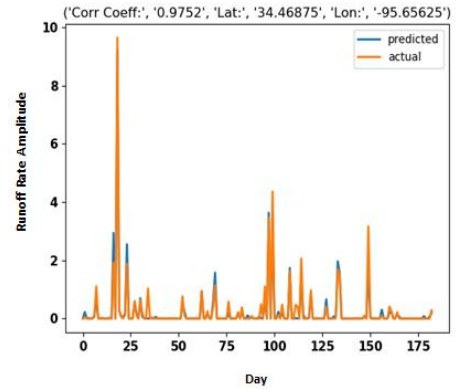


Fig. 4 (a). Pearson correlation coefficient equals 0.9752.

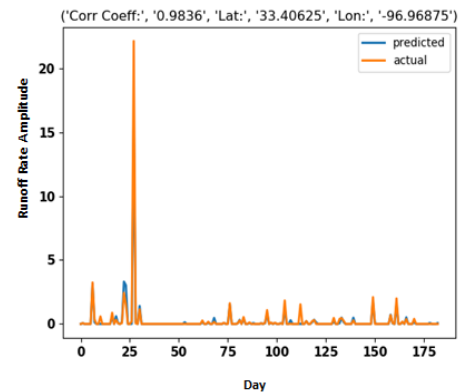


Fig. 4 (b). Pearson correlation coefficient equals 0.9836.

Fig. 5 illustrates the correlation coefficient for all cells within the basin. The average correlation coefficient with one standard deviation across the whole basin is 0.9007 ± 0.0547 . The histogram of these correlation coefficients is plotted in Figure 6. It is noteworthy that the correlation coefficients are concentrated around 93%, certifying the robustness of our proposed learning framework.

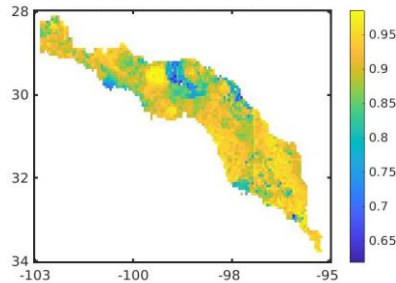


Fig. 5. The Pearson correlation coefficient of predicted and actual runoff data across the whole cells within the basin.

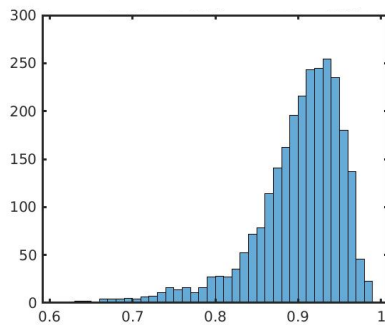


Fig. 6. The histogram of the Pearson correlation coefficient of predicted and actual runoff data across the whole cells within the basin. As shown, the histogram is concentrated around 0.93.

D. Predictor Importance

Fig. 7 illustrates the relative importance score for each of the predictors to forecast the runoff rates. We used Gini criterion [18] to calculate these scores. As this figure suggests, the precipitation plays a pivotal role to forecast the runoff rates. This finding, however, has been well-established using the physics-based and the empirical modeling indicating that the temporal variability in runoff can be dominantly explained by the variability of precipitation rates (Tayfur 2006 [19], Dawson 1998 [20]).

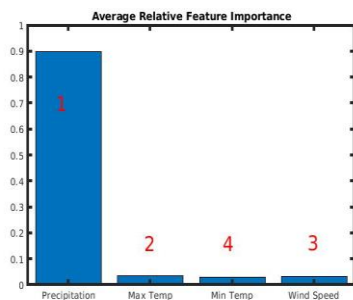


Fig. 7. Relative importance score of each meteorological predictors to forecast the runoff rates. As shown, precipitation rate with the importance score of ~90% is the most contributive parameter. Maximum temperature, wind speed and the minimum temperature rates are consequently the next contributive parameters. The reported scores are the average importance scores across the whole cells.

VI. CONCLUSION

Inspired and motivated by the recent advances in data-driven models across environmental [21], [22] and hydrological sciences [1]-[5] as a powerful tool to approximate the physical-based models, we investigate the potential of the artificial intelligence (AI) methods, and in particular random forest coupled by active learning strategy to predict the runoff rate using the optimal portion of meteorological variables over Brazos Basin in Texas. To the best of our knowledge, for the first time, this paper presents a novel strategy to predict the runoff rates within a learning framework in which the optimal training data are suggested by an advanced active learning technique. We report the prediction performance as the Pearson correlation coefficient between the actual and predicted runoff rates with average correlation coefficient of 0.9007 across all cells within the Brazos Basin. Our results also constantly suggest that the precipitation rate is the main driving parameter to forecast the runoff rate within all of the cells of the basin which is in agreement with prior studies [19],[20].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

HGD contributed significantly towards the data preparation as well as the algorithm development and its implementation; HGD, DS, RS and YW contributed significantly towards drafting the manuscript, verifying the findings of the work and the related interpretation of the results; DB and JS supervised the project and have been significantly involved in revising the results for the intellectual content as well as providing critical feedback; all authors had approved the final version.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation award (grant: # GR10458) and conducted at Future H2O, Office of Knowledge Enterprise Development (OKED) at Arizona State University.

REFERENCES

- [1] F. Kuai *et al.*, "Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network," *Geophysical Research Letters*, vol. 44, no. 21, 2017, pp. 11-030.
- [2] H. Hirose *et al.*, "A rainfall forecasting method using machine learning models and its application to the Fukuoka city case," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 4, 2012, pp. 841-854.
- [3] J. Aparicio *et al.*, "Frequency and spatial characteristics of droughts in the Conchos River Basin, Mexico," *Water International*, vol. 27, no. 3, 2002, pp. 420-430.
- [4] J. Adamowski *et al.*, "Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes," *Journal of Hydroinformatics*, vol. 16, no. 3, 2014, pp. 671-689.
- [5] Khan, M. Sajjad, and P. Coulibaly, "Application of support vector machine in lake water level prediction," *Journal of Hydrologic Engineering*, vol. 11, no. 3, 2006, pp. 199-205.
- [6] G. Thomas, "Ensemble learning," 2002.
- [7] D. Tom, "Overfitting and undercomputing in machine learning," *ACM Computing Surveys*, vol. 27, no. 3, 1995, pp. 326-327.
- [8] S. P. Bratton *et al.*, "Urbanization is a major influence on microplastic ingestion by sunfish in the Brazos River Basin, Central Texas, USA," *Environmental Pollution*, vol. 210, 2016, pp. 380-387.

- [9] L. Ben *et al.*, "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions," *Journal of Climate*, vol. 26, no. 23, 2013, pp. 9384-9392.
- [10] E. P. Maurer *et al.*, "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States," *J. Climate*, vol. 15, pp. 3237-3251, 2002.
- [11] L. Andy and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, no. 3, 2002, pp. 18-22.
- [12] Safavian, S. Rasoul, and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, 1991, pp. 660-674.
- [13] T. X. He *et al.*, "An active learning approach with uncertainty, representativeness, and diversity," *The Scientific World Journal*, 2014.
- [14] D. Fouzi *et al.*, "SVR active learning for product quality control," presented at 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), 2012.
- [15] K. Nigamy *et al.*, "Employing EM and pool-based active learning for text classification," in *Proc. International Conference on Machine Learning (ICML)*, 1998.
- [16] J. Thomas, *Set Theory*, Springer Science & Business Media, 2013.
- [17] B. Schölkopf *et al.*, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, 2004, pp. 199-222.
- [18] Ke.tu-darmstadt. [Online]. Available: <http://www.ke.tu-darmstadt.de/lehre/archiv/ws0809/mlbm/dt.pdf>
- [19] T. Gokmen and V. P. Singh, "ANN and fuzzy logic models for simulating event-based rainfall-runoff," *Journal of Hydraulic Engineering*, vol. 132, no. 12, 2006, pp. 1321-1330.
- [20] R. Wilby *et al.*, "An artificial neural network approach to rainfall-runoff modelling," *Hydrological Sciences Journal*, vol. 43, no. 1, 1998, pp. 47-66.
- [21] D. Ha. Ghasemi *et al.*, "Interpreting comprehensive two-dimensional gas chromatography using peak topography maps with application to petroleum forensics," *Chemistry Central Journal*, vol. 10, no. 1, 2016, p. 75.
- [22] D. H. Ghasemi *et al.*, "Signal variability and data compression considerations for petroleum forensics in two-dimensional gas chromatography," presented at OCEANS 2017, 2017.

Copyright © 2019 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Hamidreza Ghasemi Damavandi received his doctoral degree at the Department of Electrical and Computer Engineering, the University of Iowa, summer 2016. Prior to Ph.D, he received his B.S.c degree at the Department of Electrical and Computer Engineering, University of Tehran, summer 2013. His main research interest is to apply machine and deep learning strategies to mimic the environmental and hydrological events. He currently works as postdoctoral research associate at office of knowledge enterprise development, Future H2O, Arizona State University. He attempts to emulate the physical-based hydrological models via optimization theories as well as the deep machine learning tools. He held the same position at University of California- Los Angeles before joining ASU.



Dimitrios Stampoulis received his doctoral degree from the Civil & Environmental Engineering Department of the University of Connecticut. He is currently a research scientist working at the Arizona State University. His primary research interests are directed toward investigating the potential for synergies between remote sensing science and hydrology as well as eco-hydrology in the aim of understanding the linkages and interactions between the water dynamics and different ecosystem processes. He uses hydrologic modeling, remote sensing observations and data integration techniques to assess the impact of climatic variability on freshwater availability globally.



Reepal Shah is currently an assistant research scientist with Arizona State University. Prior to joining Arizona State University, he received his PhD from IIT Gandhinagar in 2017. His current research are surrounding following areas: hydrology, climate variability and climate change impact assessment, and crop modeling.



Yuhang Wei is currently a postdoctoral research associate with Arizona State University. He received his PhD and BEng from Hohai University in Water Conservancy and Hydropower Engineering. His research interests are water resources modeling and optimization, decision support tool.



Dragan Boscovic is the technical director of Center for Assured and Scalable Data Engineering (CASCADE) as well as the founder and director of Blockchain Research Lab at Arizona State University. His research interest is on the envisions, communicate and direct development of technologies, products and services in support of IoT smart spaces vision focusing on Smart City, Smart Office, eHealth and Mobility services and applications.



John Sabo is currently an full professor of School of Life Sciences, College of Liberal Arts and Sciences - the senior sustainability scientist at Julie Ann Wrigly Global Institute of Sustainability, as well as the director of Future H2O, Knowledge Enterprise Development at Arizona State University. He received his PhD at University of California – Berkely 2000. Dr. Sabo's research employs large-scale field experiments, stable isotopic tracers, and lab physiology to understand links between the water cycle and animal performance, abundance, and species diversity. Most of his work focuses on riparian and river ecology.