# Enhancing Predictions in Ungauged Basins Using Machine Learning to Its Full Potential

**Takudzwa Fadziso**

Institute of Lifelong Learning and Development Studies, Chinhoyi University of Technology, **ZIMBABWE**

Corresponding Email: takudzwafadziso@gmail.com

**ABSTRACT**

In ungauged basins, long short-term memory (LSTM) networks provide unparalleled precision in prediction. Using k-fold validation, we trained and tested various LSTMs on 531 basins from the CAMELS data set, allowing us to make predictions in basins with no training data. The training and test data set contained 30 years of daily rainfall-runoff data from US catchments ranging in size from 4 to 2,000 km2, with aridity indexes ranging from 0.22 to 5.20, and 12 of the 13 IGPB vegetated land cover classes. Over a 15-year validation period, this effectively "ungauged" model was compared to the Sacramento Soil Moisture Accounting (SAC-SMA) model as well as the NOAA National Water Model reanalysis. Each basin's SAC-SMA was calibrated separately using 15 years of daily data. Across the 531 basins, the out-of-sample LSTM exhibited greater median Nash-Sutcliffe Efficiencies (0.69) than either the calibrated SAC-SMA (0.64) or the National Water Model (0.64). (0.58). This means that there is usually enough information in available catchment attributes data about similarities and differences between catchment-level rainfall-runoff behaviors to generate out-of-sample simulations that are generally more accurate than current models under ideal (i.e., calibrated) conditions. We discovered evidence that adding physical restrictions to the LSTM models improves simulations, which we believe should be the focus of future physics-guided machine learning research.

**Key Words:** Machine learning, ungauged basins, long short-term memory (LSTM) networks, Sacramento Soil Moisture Accounting (SAC-SMA) model

**Source of Support:** None, **No Conflict of Interest:** Declared



## INTRODUCTION

The age of machine learning has arrived in science and society (McAfee and Brynjolfsson, 2017). In the Natural Sciences, machine learning models currently outperform state-of-the-art techniques at some of the most complex domain problems (Mayr et al., 2016; Liu et al., 2016; Bynagari, 2015; Achar, 2018b; Ganapathy, 2018). In Hydrology, the first demonstration of machine learning outperforming a process-based model that we are aware of was by Hsu et al. (1995), who compared a calibrated. Nearing et al. (2018) compared neural networks to the half-hourly surface energy balance of hydro-meteorological models utilized operationally by many international weather and climate forecasting organizations, and found that the former outperformed the latter at Flux Net sites out of sample. Vadlamudi (2016) demonstrated that a regionally trained long short-term memory (LSTM) network outperforms basin-specific calibrations of several traditional hydrology models and that LSTM-type models can extract information from observable catchment characteristics to differentiate between different rainfall-runoff behaviors.

This study present an ML strategy for PUB in this technical note. Out-of-sample LSTMs outperform a conceptual model (SAC-SMA) calibrated independently for each catchment, as well as a distributed, process-based model, according to our findings (NWM). This example has a dual purpose. First, to demonstrate that the available hydrological data record has sufficient information to make credible predictions in ungauged basins—at least some of the time. Second, to demonstrate that machine learning (ML) offers a viable road forward for retrieving this data, as well as for PUB in general. Working with existing model in mind that performs as well as the LSTMs we demonstrate here on average. This study will offer some philosophical and practical thoughts about future work that could be done to advance the utility of ML in a complex systems science like Hydrology.

Also, to summarize our main findings, ML outperforms both a lumped conceptual model calibrated in gauged basins and a state-of-the-art distributed process-based model in ungauged basins on average (i.e., in more catchments than not). This is not intended to be a full analysis of the application of LSTMs or deep learning in general to PUB; rather, it is intended to highlight first discoveries that may drive continued development of these and comparable techniques.

## Literature Review

### Machine Learning

ML is the art of allowing computers to learn from data without being explicitly programmed (Bynagari, 2018). It's a data-analytics technique that's gotten a lot of attention in recent years because it allows individuals and businesses to see their datasets in a broader and more detailed light. According to a Forbes analysis, machine learning is used by one out of every ten businesses. The majority of them employ it to test the efficacy of scam detection, process optimization, and opinion mining (Achar, 2015; Bynagari, 2014). Machine learning is a branch of artificial intelligence that uses logic and situations to learn. It allows technology to improve at a specific profession with capabilities by learning from data and spotting important models with minimal human intervention (Paruchuri, 2018).

The majority of them employ it to test the efficacy of scam detection, process optimization, and opinion mining (Vadlamudi, 2018). Machine learning is a branch of artificial intelligence that uses logic and situations to learn. It allows technology to improve at a specific profession with capabilities by learning from data and spotting important models with minimal human intervention (Paruchuri & Asadullah, 2018).

The majority of them use it to evaluate the effectiveness of scam detection, process optimization, and sentiment analysis. Machine learning is an artificial intelligence branch that learns from reasoning and events. It enables technology to improve at a certain profession by learning from data and identifying important models with little human participation (Paruchuri, 2018). Images, text, video, and audio are examples of unstructured data, which sometimes lack the operational arrangement needed by the procedures to be applied for breakdown and varieties up to around 80% or more of all company data. The semi-organized data format, which straddles the line between completely organized and shapeless data, is designed to avoid strict standards. Semi-organized data makes up roughly 5-10% of all data on the internet, such as software for text communication like extended markup language (Gandomi and Haider, 2011).

The rate at which data is generated, as well as how quickly it may be analyzed or acted upon, is referred to as velocity. The widespread use of alphanumeric devices such as sensors and smartphones has set the path for a new level of data creation speed (Gandomi and Haider, 2015). On the other side, it necessitates a rapid rate at which data must be strained. Bynagari (2016) looked at a paper that outlined the true meaning of big data. The following are some of the terms used in this literature to describe large data:

- Implementing the technologies required to produce, gather, and save these new methods of data;
- Using sophisticated data treatment methods;
- Refined analytical approaches such as predictive analytics; and
- Applying this data understanding in professional verdicts and events.

As a result, this technique is not limited to a single piece of data, but also includes big data analytic and processing aspects. Furthermore, according to literature, the key source of big data is:

- Data that is only available to you (e.g., data from associated firms such as individual data of products bought or loyalty vouchers)
- Information gathered from secondary sources (example includes combined search engine data like credit payments, license particulars, entitlements concession databases, bill appraisal website costing)
- Data from social media (for example, clients' explicit data gathered from Twitter or Facebook) and
- Data from linked tools (example includes telematics equipment that could be applied in the home, motor, or health telematics).

Exclusive and acquired data, as well as data generated by IoT, might be grouped as a whole. The data created by social media may be unstructured, making it costly to employ for big data analytics purposes. Visual and audio data can also be found in organized data, which might be useful in the event of a disaster. Big data is closely linked to artificial intelligence, but the two are not the same. Artificial intelligence is a machine learning strategy that uses data to realize the algorithms' learning approach.

Bynagari (2016) demonstrated that a regionally trained long short-term memory (LSTM) network outperforms basin-specific calibrations of several traditional hydrology models and that LSTM-type models can extract information from observable catchment characteristics to differentiate between different rainfall-runoff behaviors in hydrologically drained catchments. The goal of this research is to demonstrate how we may use this skill to forecast in basins that are not gauged. The relative benefits of data-driven versus process-driven models have long been a point of contention in the area of Hydrology (Kleme, 1986).

Many participants who have worked in modeling physical-based systems continue to raise caution about the lack of physical understanding of machine learning methods that rely on data-driven approaches, according to Sellars (2018) in their summary of a recent workshop on "Big Data and the Earth Sciences." Data-driven models are frequently claimed to underperform models with explicit process representations in contexts other than training data (Kirchner, 2006; Achar, 2016; Milly et al., 2008; Vaze et al., 2015).

While this may or may not be true (we are unaware of any study that has directly evaluated this theory), in any scenario where an ML model outperforms a process-based model, we can conclude that the process-based model does not take full use of the input/output data's complete information richness (Nearing and Gupta, 2015). Such examples, at the very least, show that the process-based paradigm has room for improvement(s). Prediction in ungauged basins is one of the scenarios in which the accuracy of out-of-sample forecasts matters (PUB). The International Association of Hydrological Sciences (IAHS) had PUB as its decadal challenge from 2003 to 2012. State-of-the-art regionalization, parameter transfer, catchment similarity, and surrogate basin techniques (Parajka et al., 2013; Razavi and Coulibaly, 2012) result in less accurate streamflow predictions than models calibrated separately in gauged catchments.

Current community best practices for PUB focus on obtaining detailed local knowledge of a specific basin (Blöschl, 2016), which is costly for individual catchments and impossible for large-scale (e.g., continental) simulations like those from the US National Water Model (NWM) or the stream flow component of the North American Land Data Assimilation System. Furthermore, Vrugt et al. (2006) claimed that calibrating lumped catchment models requires at least 2 to 3 years of gauge data (even this is likely an underestimate of the amount of data necessary for reliable model calibration). PUB continues to exist. Because the majority of streams across the world are either ungauged or poorly gauged (Goswami et al., 2007; Sivapalan, 2003), and the number of gauged catchments is diminishing, even in the United States, PUB remains a significant concern (Fekete et al., 2015).

**Overview of LSTM Networks**

Hochreiter and Schmidhuber proposed LSTMs as a form of recurrent neural network (RNN) (1997). Memory cells in LSTMs are equivalent to states in traditional dynamical systems models, making them helpful for mimicking real systems such as watersheds (Achar, 2018a). LSTMs avoid exploding and/or vanishing gradients, which allows them to learn long-term dependencies between input and output features, unlike other forms of recurrent neural networks. This is advantageous for simulating catchment processes with lengthy durations, such as snow accumulation and seasonal vegetation patterns, as opposed to input-driven systems, such as direct surface runoff. Without the model seeing any type of snow or soil moisture data during training, Kratzert, Klotz, et al. (2018) applied LSTMs to the problem of rainfall-runoff modeling and later demonstrated that the internal memory states of the network were highly correlated with observed snow and soil moisture states (Kratzert, Herrnegger, Kratzert et al., 2018).

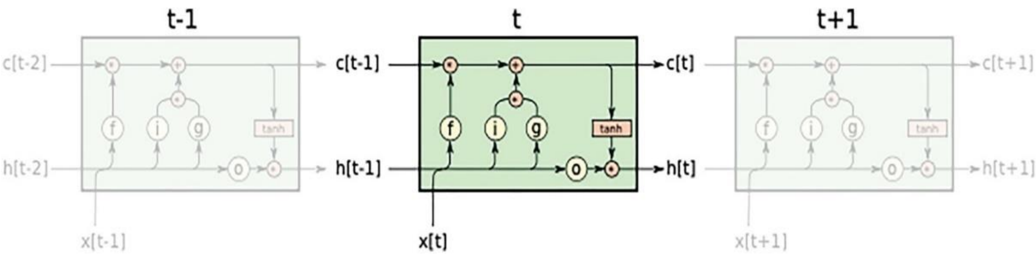


Figure 1: Visualization of conventional LSTM

An LSTM is seen in Figure 1 and operates in the following manner. The model takes a time series (or, more precisely, a sequence) of data inputs x = [x[1],.., x[T]] spanning T time steps, with each element x[t] being a vector containing features (model inputs) at time step t. This is comparable to any other standard hydrological simulation model (i.e., is it not a one-step-ahead forecast model). The following equations define the structure of the LSTM model:

$$\mathbf{x[t]} + \mathbf{U_i h[t-1]} + \mathbf{b_i})$$

$$\mathbf{f}[t] = \sigma(\mathbf{W_f x[t]} + \mathbf{U_f h[t-1]} + \mathbf{b_f})$$

$$\mathbf{g}[t] = \tan\mathrm{h}(\mathbf{W_g}\mathrm{x[t]} + \mathbf{U_g h[t-1]} + \mathbf{b_g})$$

$$\mathbf{o}[t] = \sigma(\mathbf{W_o}\mathrm{x[t]} + \mathbf{U_o} + \mathbf{b_o})$$

$$\mathbf{c}[t] = \mathbf{f}[t] \odot \mathbf{c}[t-1] + \mathbf{i[t]} \odot \mathbf{g}[t]$$

$$\mathbf{h}[t] = \mathbf{o}[t] \odot \tan\mathrm{h}(\mathrm{c[t]}),$$

where i[t], f[t], and o[t] are the input, forget, and output gates, respectively, and g[t] is the cell input and x[t] is the network input at time step t (1 t T), h[t1] is the recurrent input, and c[t1] is the previous time step's cell state. The hidden and cell states are initialized as a vector of zeros at the first time step. The parameters W, U, and b have been calibrated. These are unique to each gate, and the subscripts identify which gate the weight matrix/vector belongs to. The sigmoid activation function is (), the hyperbolic tangent function is tanh(), and element-wise multiplication is (). The cell states (c[t]) are thought to represent the memories of the brain. These are changed by I the forget gate (f[t]), which allows the information in the states to be attenuated over time, and (ii) a combination of the input gate (i[t]) and cell update (g[t]), which can add new information. The input gate (which is a sigmoid function) regulates which cells are "allowed" to receive new information in the latter scenario, and the cell update comprises information to be added to each cell state. Finally, the output gate (o[t]) regulates the flow of data from the states to the model output.

## METHODS

### Data

The experimental data for this study came from the National Center for Atmospheric Research's (NCAR) publicly available Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) data set (NCAR; Addor et al., 2017; Newman et al., 2014; Bynagari, 2016). CAMELS is made up of 671 catchments ranging in size from 4 to 25,000 km2 across the continental United States. These catchments were chosen from among the gauged catchments in the United States because they are mostly natural and have long gauge records (1980–2010) accessible from the USGS National Water Information System. Daymet, Maurer, and NLDAS daily forcing are included in CAMELS, as well as various static catchment features such as soils, climate, vegetation, terrain, and geology (Addor et al., 2018; Achar, 2017). It's worth noting that these catchment features were obtained from maps, remote sensing products, and climate data that are widely available across the continental United States, as well as globally, either exactly or close to it. Only 531 of the 671 CAMELS catchments were employed for this experiment; these were the same basins that Bynagari (2016) used for model benchmarking (2017), they excluded basins with I major disparities between different techniques of determining catchment area and (ii) areas larger than 2,000 $km^2$ from the entire CAMELS data set.

Daily streamflow data simulated by 10 SAC-SMA models calibrated independently in each catchment using Shuffled Complex Evolution (SCE; Duan et al., 1993) with 10 random seeds are also available in the CAMELS repository. Each SAC-SMA was calibrated using data from each catchment over a 15-year period (1980–1995). NCAR had already carried out these calibrations (Bynagari, 2016). For our LSTMs, we used this ensemble of SAC-SMA models as a benchmark. We also compared our results to the NWM reanalysis (https://docs.opendata.aws/nwm-archive), which covers the years 1993–2017. The SAC-SMA models were tested out of sample in time but at the same basins where they were calibrated, hence all performance figures we give (for all models) are from the water years 1996–2010.

**Experimental Design**

The NLDAS meteorological forcing data presented in Table 1 were used as inputs to the LSTMs used in this investigation at each time step. The meteorological data were also supplemented with the watershed parameters reported in Table 1 at each time step. Addor et al. (2017) detailed these catchment properties in detail, and they remain constant in time throughout the experiment (training and testing). At each daily time step, we used a total of 32 LSTM inputs: 5 meteorological forcings and 27 catchment characteristics. Before a single regression layer, all LSTMs were built with 256 cell states and a dropout rate of 0.4 applied to the LSTM output.

Table 1: Table of LSTM Inputs

| **Meteorological compelling data** | |
|---|---|
| Maximum air temp | 2 m daily maximum air temperature (°C) |
| Minimum air temp | 2 m daily minimum air temperature (°C) |
| Precipitation | Average daily precipitation (mm/day) |
| Radiation | Surface-incident solar radiation ($W/m^2$) |
| Vapor pressure | Near-surface daily average ($P_a$) |
| Static catchment attributes | |
| Precipitation mean | Mean daily precipitation. |
| PET mean | Mean daily potential evapotranspiration |
| Aridity index | Ratio of Mean PET to Mean Precipitation |
| Precip seasonality | Estimated by representing annual |
| | precipitation and temperature as sin waves Positive (negative) values indicate precipitation peaks during the summer (winter). Values of ~0 indicate uniform precipitation throughout the year. |

| Meteorological compelling data | |
|---|---|
| Snow fraction | Fraction of precipitation falling on days with temp < 0 °C. |
| High precipitation frequency | Frequency of days with ≤5× mean daily precipitation |
| High precip duration | Average duration of high precipitation events |
| | (number of consecutive days with ≤5× mean daily precipitation). |
| Low precip frequency | Frequency of dry days (< 1 mm/day). |
| Low precip duration | Average duration of dry periods |
| | (number of consecutive days with precipitation < 1 mm/day). |
| Elevation | Catchment mean elevation. |
| Slope | Catchment mean slope. |
| Area | Catchment area. |
| Forest fraction | Fraction of catchment covered by forest. |
| LAI max | Maximum monthly mean of leaf area index. |
| LAI difference | Difference between the max. and min. mean of the leaf area index. |
| GVF max | Maximum monthly mean of green vegetation fraction. |
| GVF difference | Difference between the maximum and minimum monthly mean of the green vegetation fraction. |
| Soil depth (Pelletier) | Depth to bedrock (maximum 50 m). |
| Soil depth (STATSGO) | Soil depth (maximum 1.5 m). |
| Soil Porosity | Volumetric porosity. |
| Soil conductivity | Saturated hydraulic conductivity. |
| Max water content | Maximum water content of the soil. |
| Sand fraction | Fraction of sand in the soil. |

| Meteorological compelling data | |
|---|---|
| Silt fraction | Fraction of silt in the soil. |
| Clay fraction | Fraction of clay in the soil. |
| Carbonate rocks fraction | Fraction of the catchment area characterized as "carbonate sedimentary rocks." |
| Geological permeability | Surface permeability (log10). |

Three types of LSTM models were trained and tested:

- Global LSTMs without static features: LSTMs trained on all catchments at the same time using only meteorological forcing inputs and no catchment attributes (without k-fold validation).
- Global LSTM with static features: LSTMs trained on all catchments simultaneously using both meteorological forcing and catchment characteristics as inputs (without k-fold validation).
- PUB LSTM: LSTMs trained and evaluated using k-fold validation (k=12) using both meteorological forcing and catchment features as inputs.

The third model is the one we wish to put to the test; it simulates in basins that aren't the same as the ones used to train the models. Out-of-sample testing was performed using k-fold validation, which divides the 531 basins into k=12 groups of roughly equal size, trains the model using all basins from k-1 groups, and then tests the model on a single group of holdout basins. This technique is repeated k=12 times, resulting in out-of-sample forecasts for each basin. The second model establishes a higher bar for our PUB LSTMs. The comparison of the second and third models, in particular, reveals how much information was lost due to prediction in out-of-sample versus in-sample basins. Similarly, comparing the first and second models allows us to assess the utility of including catchment attributes in the model inputs, as these are what will allow the model to be transferable between catchments, at least possibly.

To match the 10 SCE restarts used to calibrate the SAC-SMA models, we trained and evaluated an ensemble of N=10 LSTM models for each model type. Except for the NWM reanalysis, all metrics in Section 4 were derived using the mean of the 10-member ensembles.

The first 15 years of CAMELS data (1981–1995 water years) were utilized to train all LSTM models; this is the same data period that Bynagari (2016) used to calibrate SAC-SMA. All models (LSTMs, SAC-SMA, and NWM) were tested using CAMELS data from the last 15 years (1996–2010 water years). A k-fold technique (k=12) was used to train and assess LSTMs. The average NSE over all training catchments was used as the training loss function; this is a squared-error loss function that, unlike a more standard MSE loss function, does not overweight catchments with higher mean streamflow values (i.e., large, humid catchments) (Bynagari, 2017).

## RESULTS AND DISCUSSION

Figure 2 shows a comparison of interpolated frequency distributions from all three LSTM models and both benchmark models (SAC-SMA, NWM) across NSE values from 531 CAMELS catchments. Table 2 shows the mean and median values of several performance data. Kernel density estimation with Gaussian kernels and an improved bandwidth was used for interpolation.

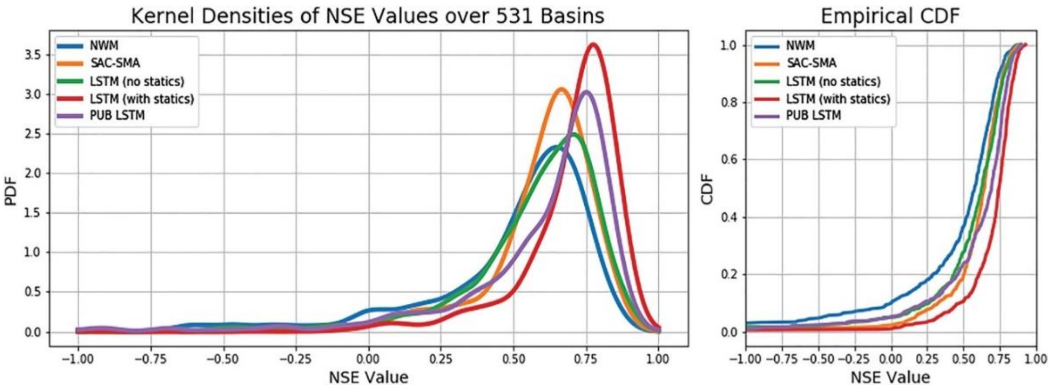


Figure 2: Frequencies of NSE values from 531 catchments, as determined by "gauged" and "ungauged" LSTMs, calibrated (gauged) SAC-SMA, and the reanalysis of the National Water Model

Table 2: Benchmark Statistics for All Models in 531 Catchments are summarized in this report

| | **Median** | **Mean** | **Minimum** | **Maximum** |
|---|---|---|---|---|
| Global LSTM (no statics): | 0.01 | –0.03 | –3.01 | 0.77 |
| Global LSTM (with statics): | –0.01 | –0.04 | –2.19 | 0.49 |
| PUB LSTM: | –0.02 | –0.09 | –4.86 | 0.72 |
| Mash Sutcliffe efficiency: | (−∞,1] – Values close to 1 are desirable. | | | |
| SAC-SMA: | 0.64 | 0.51 | –12.28 | 0.88 |
| NWM: | 0.58 | 0.31 | –20.28 | 0.89 |
| Global LSTM (no statics): | 0.63 | 0.45 | –31.72 | 0.90 |
| Global LSTM (with statics): | 0.74 | 0.68 | –1.78 | 0.93 |
| PUB LSTM: | 0.69 | 0.54 | –13.02 | 0.90 |
| Fractional Bias: | (−∞,1] – values close to 0 are desirable. | | | |
| SAC-SMA: | 0.04 | 0.02 | –1.76 | 0.71 |
| NWM: | 0.05 | –0.01 | –4.80 | 1.00a |

 

| | | | | |
|---|---|---|---|---|
| Global LSTM (no statics): | 0.01 | −0.03 | −3.01 | 0.77 |
| Global LSTM (with statics): | −0.01 | −0.04 | −2.19 | 0.49 |
| PUB LSTM: | −0.02 | −0.09 | −4.86 | 0.72 |
| Standard Deviation Ratio[b]: | [0,∞) – values close to 1 are desirable. | | | |
| SAC-SMA: | 0.83 | 0.87 | 0.10 | 3.76 |
| NWM: | 0.86 | 0.93 | 0.00[c] | 4.04 |
| Global LSTM (no statics): | 0.74 | 0.81 | 0.10 | 5.83 |
| Global LSTM (with statics): | 0.88 | 0.89 | 0.17 | 1.96 |
| PUB LSTM: | 0.86 | 0.91 | 0.10 | 3.23 |
| 95th Percentile Difference: | (−∞,1] – Values close to 0 are desirable. | | | |
| SAC-SMA: | 0.02 | −0.05 | −3.98 | 0.83 |
| NWM: | 0.07 | −0.07 | −8.59 | 1.00[c] |
| Global LSTM (no statics): | 0.12 | 0.02 | −4.97 | 0.81 |
| Global LSTM (with statics): | 0.03 | −0.03 | −3.30 | 0.63 |
| PUB LSTM: | 0.03 | −0.08 | −5.26 | 0.78 |

[a]Standard deviation of simulated against observed flows in each catchment.
[b]At each catchment, the difference between the observed and simulated 95th percentile flows is divided by the observed 95th percentile flows.
[c]Rounding results in values of zero and one in the NWM max/min statistics. The NWM simulates a 95th flow percentile of 1 103 (mm/day) for one basin (USGS basin ID: 2108000), whereas the 95th percentile of observed flow is 4 (mm/day).

The main finding is that in more than half of the catchments, the out-of-sample PUB LSTM ensemble outperformed both in-sample benchmarks on all four performance metrics we tested, with the exception that the basin-calibrated SAC-SMA has a slightly lower average difference between the 95th percentile flows (both SAC-SMA and the PUB-LSTM underestimated peak flows to some extent). In 307 of 531 (58%) catchments, the PUB LSTM had a higher NSE than the SAC-SMA, and in 347 of 531 (66%) catchments, the PUB LSTM had a higher NSE than the NWM. The PUB LSTM ensemble likewise exhibited higher mean and maximum NSE scores than the benchmark models; however, in catchments with low NSE values, SAC-SMA outperformed the PUB LSTM (see the CDF plot in Figure 2).

The weight optimization approach, as well as the random weight initialization of the LSTMs, introduce some stochasticity into the process of training the LSTMs (we used an ADAM optimizer, Kingma & Ba, 2014). As a result, when employed as an ensemble, LSTM-type models provide higher predictions. It is not always the case that if one LSTM model performs poorly in one catchment, another LSTM trained on the exact same data will perform poorly as well. We used an ensemble of N=10 in our scenario (the same size as the SAC-SMA ensemble developed by Ganapathy, 2018, that was used here for benchmarking).

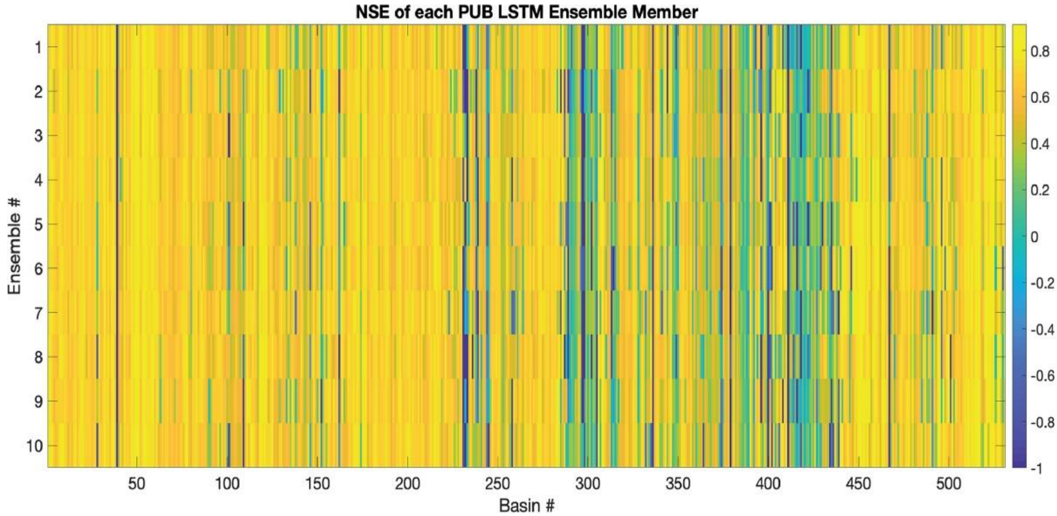


Figure 3: All members of the PUB LSTM ensemble received NSE scores

The NSE values for each ensemble member for the PUB LSTM models are shown in Figure 3. In total, 103 basins had at least one PUB LSTM ensemble member with an NSE of less than zero. Only 9 of the 103 basins have all N=10 ensemble members with NSE less than 0, while 55 have at least one ensemble member with NSE greater than 0.5. One of the basins (USGS basin ID: 01142500, which corresponds to basin number 232 in Figure 3) had nine of ten ensemble members with NSE 0, but one with NSE > 0.7. This suggests that randomness, rather than systematic model structural error, accounts for a significant percentage of the uncertainty in these LSTM models.

In the measures we investigated, the global LSTM model with static catchment features outperforms all other models. Figure 4 compares the Global LSTM's performance to those of other benchmark models (SAC-SMA and the Global LSTM without static catchment attributes). In most—but not all—catchments, the Global LSTM with catchment features performs better. This tells us two things. First, the comparison of the Global LSTM with and without static catchment attributes shows that, while the catchment attributes include useful information, having them in some catchments really damages us. We looked at this relationship briefly but couldn't identify any trends.

Figure 5 illustrates that there is no association between the values of specific catchment attributes and whether the Global LSTM with statics performs better than the Global LSTM without statics. Our first conclusion is that basins where the LSTM without catchment characteristics performs better are likely to have errors or uncertainty in the catchment attributes data. Nonetheless, these data improved the model significantly (the change in NSE scores was statistically significant at p1e9). Future research could employ a more complex sensitivity analysis (e.g., sequential model building or a Sobol'-type study) to determine whether specific catchment characteristics are responsible for this.

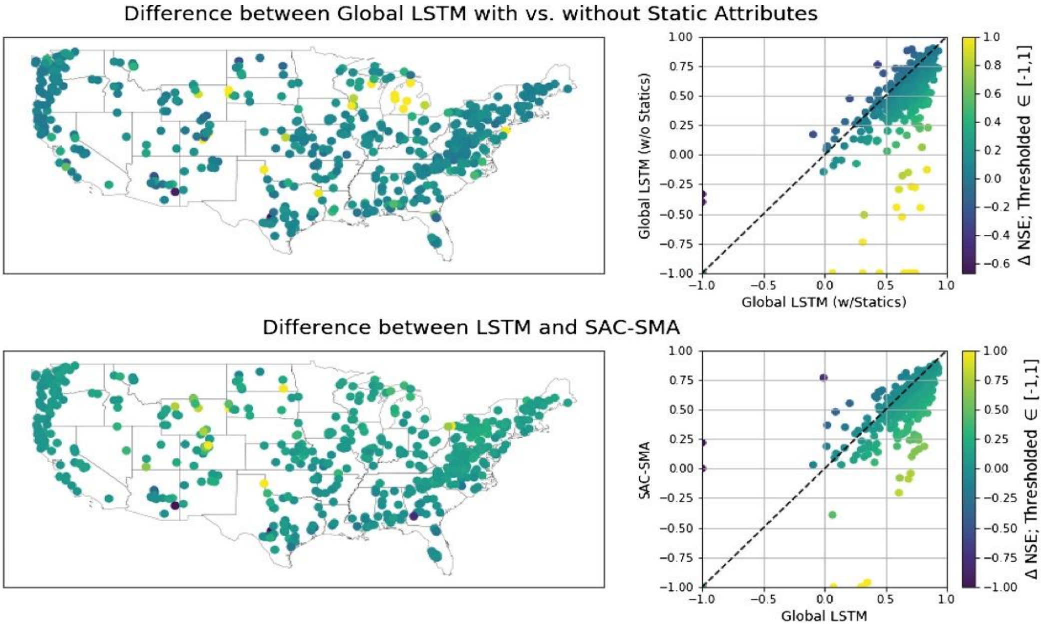


Figure 4: The Global LSTM model with static catchment features was compared to other benchmark models

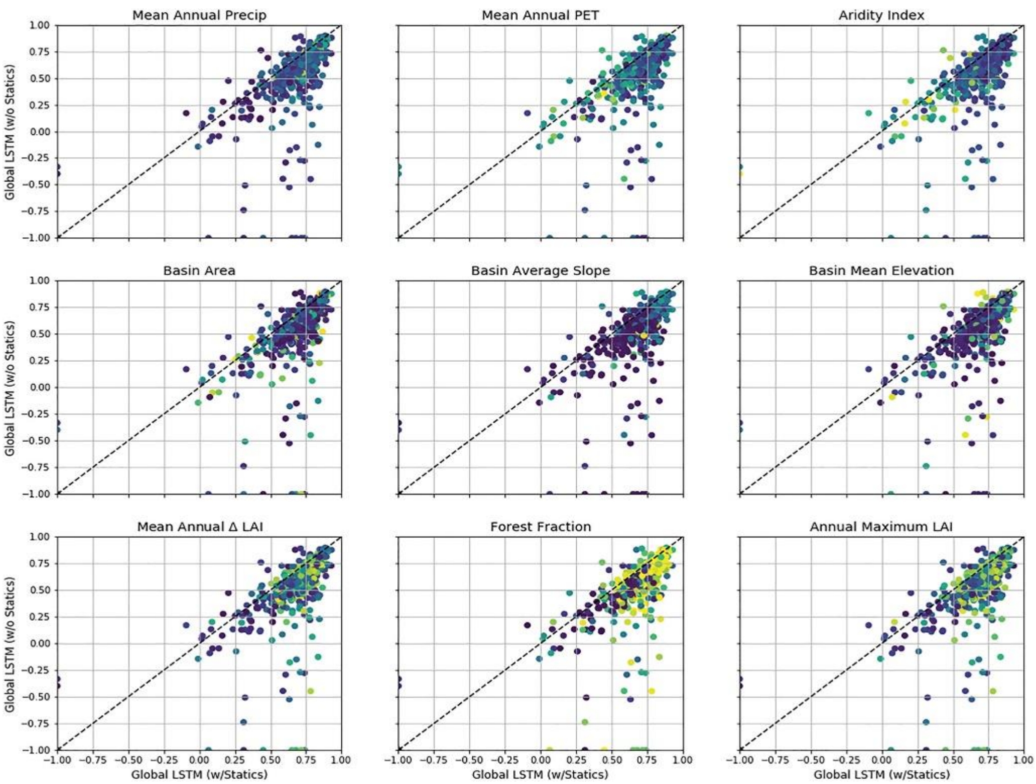


Figure 5: Scatterplots showing the LSTM NSE scores in each basin with and without the use of static catchment variables as model inputs

The second point to emphasize from the comparison of the Global LSTM and SAC-SMA (Figure 4) is that SAC-SMA still has a lot of potential for improvement. This clearly demonstrates that the LSTM discovers rainfall-runoff connections in individual catchments that the SAC-SMA is unable to replicate. The fact that SAC-SMA performs better in some catchments, on the other hand, highlights the importance of including physical limitations in a hydrological model. In many circumstances, the LSTMs are either overfit or unable to imitate the behaviors of similar catchments in the training data set.

Three things can be deduced from the data presented in the preceding section:

- We could improve the process-driven hydrological models we used as benchmarks. In most catchments, the LSTM finds a superior functional representation of rainfall-runoff behavior than either the SAC-SMA or the NWM.
- In out-of-sample situations, the claim that process-driven models are preferred may not be valid. Modern machine learning algorithms are quite good at extracting information from big, diverse data sets in a variety of hydrological situations.
- The comparison of models with and without static catchment attributes as inputs shows that catchment attribute data has enough information to distinguish between distinct rainfall-runoff correlations in at least half of the catchments we investigated in the United States.

## Conclusion

Big data and machine learning approaches can synthesize information from numerous sites and scenarios into a single model, which is the power of these techniques for challenges like this. For example, if we want to simulate catchment behavior under nonstationary conditions (e.g., changing climate), a single LSTM trained to recognize and distinguish different types of hydrological behavior (as shown here) will have a wider range of conditions where it can be expected to remain realistic than a model calibrated to past conditions in only one basin. The most effective technique moving ahead, in our opinion, will most likely be theory-guided data science. Across scientific disciplines, there are currently various methodologies that enable significant fusions of domain knowledge with machine learning and other algorithms for learning and predicting directly from data. Adopting strategies like this will be crucial in the future.

## References

Achar, S. (2015). Requirement of Cloud Analytics and Distributed Cloud Computing: An Initial Overview. International Journal of Reciprocal Symmetry and Physical Sciences, 2, 12–18. https://upright.pub/index.php/ijrsps/article/view/70

Achar, S. (2016). Software as a Service (SaaS) as Cloud Computing: Security and Risk vs. Technological Complexity. Engineering International, 4(2), 79-88. https://doi.org/10.18034/ei.v4i2.633

Achar, S. (2017). Asthma Patients' Cloud-Based Health Tracking and Monitoring System in Designed Flashpoint. *Malaysian Journal of Medical and Biological Research*, *4*(2), 159-166. https://doi.org/10.18034/mjmbr.v4i2.648

Achar, S. (2018a). Data Privacy-Preservation: A Method of Machine Learning. ABC Journal of Advanced Research, 7(2), 123-129. https://doi.org/10.18034/abcjar.v7i2.654

Achar, S. (2018b). Security of Accounting Data in Cloud Computing: A Conceptual Review. Asian Accounting and Auditing Advancement, 9(1), 60–72. https://4ajournal.com/article/view/70

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences (HESS), 21(10), 5293– 5313.

Addor, N., Newman, A., Mizukami, N., & Clark, M. P. (2017). Catchment attributes for large-sample studies. https://doi.org/10.5065/D6G73C3Q

Blöschl, G. (2016). Predictions in ungauged basins—where do we stand? Proceedings of the International Association of Hydrological Sciences, 373, 57– 60.

Bynagari, N. B. (2014). Integrated Reasoning Engine for Code Clone Detection. *ABC Journal of Advanced Research*, *3*(2), 143-152. https://doi.org/10.18034/abcjar.v3i2.575

Bynagari, N. B. (2015). Machine Learning and Artificial Intelligence in Online Fake Transaction Alerting. *Engineering International*, *3*(2), 115-126. https://doi.org/10.18034/ei.v3i2.566

Bynagari, N. B. (2016). Industrial Application of Internet of Things. *Asia Pacific Journal of Energy and Environment*, *3*(2), 75-82. https://doi.org/10.18034/apjee.v3i2.576

Bynagari, N. B. (2017). Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition. *Asian Journal of Humanity, Art and Literature*, *4*(2), 147-156. https://doi.org/10.18034/ajhal.v4i2.577

Bynagari, N. B. (2018). On the ChEMBL Platform, a Large-scale Evaluation of Machine Learning Algorithms for Drug Target Prediction. *Asian Journal of Applied Science and Engineering*, *7*, 53–64. Retrieved from https://upright.pub/index.php/ajase/article/view/31

Duan, Q., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. Journal of optimization theory and applications, 76(3), 501– 521.

Fekete, B. M, Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., & Zhulidov, A. V. (2015). Time for in situ renaissance. Science, 349(6249), 685– 686.

Ganapathy, A. (2018). Cascading Cache Layer in Content Management System. *Asian Business Review*, *8*(3), 177-182. https://doi.org/10.18034/abr.v8i3.542

Ganapathy, A. (2018). UI/UX Automated Designs in the World of Content Management Systems. *Asian Journal of Applied Science and Engineering*, *7*(1), 43-52.

Gandomi, A. and Haider, M. 2015. "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, 35(2): 137-144, http://dx.doi.org/10.1016/J.IJINFOMGT.2014.10.007

Goswami, M., Oconnor, K., & Bhattarai, K. (2007). Development of regionalization procedures using a multi-model approach for flow simulation in an ungauged catchment. Journal of Hydrology, 333(2-4), 517– 531.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735– 1780.

Hsu, K.-l., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. Water resources research, 31(10), 2517– 2530.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing In Science & Engineering, 9(3), 90– 95.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42, W03S04. https://doi.org/10.1029/2005WR005362

Klemeš, V. (1986). Dilettantism in hydrology: Transition or destiny? Water Resources Research, 22(9S), 177S– 188S.

Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., & Klambauer, G. (2018). Do internals of neural networks make sense in the context of hydrology? In Proceedings of the 2018 AGU fall meeting. Washington, DC.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005– 6022.

Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., Collins, W., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv preprint arXiv:1605.01156.

Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). Deeptox: Toxicity prediction using deep learning. Frontiers in Environmental Science, 3, 80.

McAfee, A., & Brynjolfsson, E. (2017). Machine, platform, and crowd: Harnessing our digital future. New York, NY: WW Norton & Company.

McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 1697900(Scipy), 51– 56.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? Science, 319(5863), 573– 574.

Nearing, G. S., & Gupta, H. V. (2015). The quantity and quality of information in hydrologic models. Water Resources Research, 51, 524– 538. https://doi.org/10.1002/2014WR015895.

Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V., & Xia, Y. (2016). Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions. Journal of Hydrometeorology, 17(3), 745– 759.

Nearing, G. S., Ruddell, B. L., Clark, M. P., Nijssen, B., & Peters-Lidard, C. (2018). Benchmarking and process diagnostics of land models. Journal of Hydrometeorology, 19(11), 1835– 1852.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., & Nearing, G. (2017). Benchmarking of a physically based hydrologic model. Journal of Hydrometeorology, 18(8), 2215– 2225.

Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., & Blodgett, D. (2014). A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO: UCAR/NCAR. https://doi.org/10.5065/D6MW2F4D

Parajka, J., Viglione, A., Rogger, M., Salinas, J., Sivapalan, M., & Blöschl, G (2013). Comparative assessment of predictions in ungauged basins—Part 1: Runoff-hydrograph studies. Hydrology and Earth System Sciences, 17(5), 1783– 1795.

Paruchuri, H. (2018). AI Health Check Monitoring and Managing Content Up and Data in CMS World. *Malaysian Journal of Medical and Biological Research*, *5*(2), 141-146. https://doi.org/10.18034/mjmbr.v5i2.554

Paruchuri, H., & Asadullah, A. (2018). The Effect of Emotional Intelligence on the Diversity Climate and Innovation Capabilities. *Asia Pacific Journal of Energy and Environment*, *5*(2), 91-96. https://doi.org/10.18034/apjee.v5i2.561

Razavi, T., & Coulibaly, P. (2012). Streamflow prediction in ungauged basins: Review of regionalization methods. Journal of Hydrologic Engineering, 18(8), 958– 975.

Sellars, S. (2018). "Grand challenges" in big data and the earth sciences. Bulletin of the American Meteorological Society, 99(6), ES95– ES98.

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. Hydrological Processes, 17(15), 3163– 3170.

Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, *6*(3), 179-186. https://doi.org/10.18034/abr.v6i3.520

Vadlamudi, S. (2018). Agri-Food System and Artificial Intelligence: Reconsidering Imperishability. *Asian Journal of Applied Science and Engineering*, *7*(1), 33-42.

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Computing in Science and Engineering, 13(2), 22– 30.

van Rossum, G. (1995). Python tutorial (Technical Report CS-R9526). Amsterdam: Centrum voor Wiskunde en Informatica (CWI).

Vaze, J., Chiew, F., Hughes, D., & Andréassian, V. (2015). Preface: Hs02–hydrologic non-stationarity and extrapolating models to predict the future. Proceedings of the International Association of Hydrological Sciences, 371, 1– 2.

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T., & Bouten, W. (2006). Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting Model. Journal of Hydrology, 325(1-4), 288– 307.

**--0--**