# Towards Improved Global River Discharge Prediction in Ungauged Basins Using Machine Learning and Satellite Observations

**Aggrey Muhebwa**
STIMA Lab, Department of Electrical and Computer Engineering.
University of Massachusetts, Amherst
MA 01003 U.S.A.
amuhebwa@umass.edu

**Sungwook Wi**
Department of Civil and Environmental Engineering.
University Of Massachusetts, Amherst
MA 01003 U.S.A.
sungwookwi@ecs.umass.edu

**Colin J. Gleason**
Department of Civil and Environmental Engineering.
University Of Massachusetts, Amherst
MA 01003 U.S.A.
cjgleason@umass.edu

**Jay Taneja**
STIMA Lab, Department of Electrical and Computer Engineering.
University of Massachusetts, Amherst
MA 01003 U.S.A.
jtaneja@umass.edu

## Abstract

The recent increase in frequency and severity of natural disasters is a clear indication of an immediate need to address the cascading impacts of climate change. However, climate change cannot be measured directly. In a weather cycle, river discharge is the end result of any hydrologic process, and thus directly measures the effect of two major parameters used to measure impacts of climate change; Temperature and Precipitation. Unlike current methods that are able to infer climate change patterns over a long period of time, river discharge is an effective proxy for measuring effects of climate change within a short period of time. Unfortunately, current statistical and physics-based models neither take full advantage of hydrometeorological information encoded in over 100 years of historical hydrologic data nor are they applicable on a global scale. In this work, we train Long Short Term Memory (LSTM) Recurrent Neural Network models on satellite observations and daily discharge from gauged basins to predict discharge in ungauged basins. Our models show Kling-Gupta and Nash-Sutcliffe Efficiency scores of 85% and 81% respectively in ungauged basins with limited to no existing data, while the latest state-of-the-art process-based hydrology models show performance between 0% and 50% in similar circumstances. Applying techniques like ours will allow accurate predictions in river basins across the world, the majority of which do not have *in situ* measurements.

## 1 Introduction

Anthropogenic climate change and explosive population growth are straining already scarce water resources and the resulting impact is borne in many crucial sectors; Agriculture, renewable energy, and manufacturing among others [9, 19, 20]. Therefore, there is a need for near real-time and accurate systems to measure the direct impact of climate change on water resources. River discharge is the end result of all hydrologic processes within a river basin, and as such, can be used as a proxy for measuring increased surface melting and runoff, temporary injection of melt water to the bed of

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

grounded glaciers, and hydrofracturing, i.e., melt water-induced ice shelf collapse [2], all of which are key indicators of increase in global temperature. However, there is limited measurement of river discharge on a global scale, which has hampered the ability to measure the true depth, scale, and pace of climate change.

Traditionally, river discharge has been measured *in situ* using water gauges strategically placed along the river. However, this approach does not scale well to the global level. In a weather cycle, hydrometeorology variables combine to produce the flow of water in rivers (discharge) and as such, can be used to estimate the amount of river discharge in a hydrologic cycle. Fortunately, these variables are recorded globally using numerous satellite constellations that rotate the earth at regular intervals. Machine Learning approaches are able to encode domain knowledge and leverage the spatial-temporal relationship between hydrometeorology variables (satellite data) and *in situ* discharge data. This opens up the opportunity for more accurate river discharge predictions on a global scale, especially for the majority of the global rivers, which have no *in situ* data.

In this work, we demonstrate improved performance of machine learning methods that leverage both spatial and temporal information existing in hydrometeorologic data to improve daily discharge prediction. We demonstrate that using a Long Short Term Memory (LSTM) Recurrent Neural Network, we are able to achieve Kling-Gupta Efficiency (KGE) and Nash-Sutcliffe Efficiency (NSE) [1] scores of 85% and 81% respectively on held-out discharge data drawn from a different distribution, outperforming the latest state-of-the-art process-based hydrology models in ungauged basins with limited to non-existing data.

These experiments and results demonstrate the impact of integrating spatial and temporal information in improving prediction of daily river discharge using modern machine learning algorithms in a physical sciences field that relies heavily on both conventional time-series and process-based models for analysis

## 2 Related Work

**Discharge measurement**: *In situ* measurements are the standard approach for measuring daily river discharge where water gauges are strategically placed at gauge stations along a river network. In places where gauge stations do not exist, process-based models, for example the Manning Equation (Eq. 1) for daily discharge is used if the geomorphological characteristics of the river are known.

$$Q_t = \frac{1}{n} A_{it}^{5/3} W_{it}^{-2/3} S_{it}^{1/2} \tag{1}$$

Where Q is discharge $(m^3 S^{-1})$, A is the cross-sectional area $(m^2)$, W is width $(m)$, S is slope(unitless), the index $i$ specifies the cross section and $t$ specifies the day. However, process-based models tend to degrade when trained on non-independent and identically distributed data(i.i.d), i.e., data drawn from varying geographical regions. This means that it is difficult to transfer hydrological information learnt about one river basin to another river basin, making it difficult to predict discharge for basins with little to no data.

**Machine Learning in Hydrology**: The success of machine learning has largely been due to its ability to extract complex spatial and temporal patterns existing in the training data, thus overcoming the drawbacks of conventional time-series models. Long-Short Term Memory (LSTM) Recurrent Neural networks [11] have demonstrated exceptional performance in predicting discharge in gauged basins [14, 13, 3] at both local- and continental-scale. Models trained on over 100 years' worth of historical data have demonstrated the ability to extract inherent patterns in large hydrological datasets whose dynamics are dependent on various direct and indirect interconnected phenomenon, thus opening up the possibility of solving a longstanding problem of regional modelling via transfer learning [17]. However, machine learning models are stochastic and non-deterministic in that they tend to encode correlation in the training data instead of causation. Furthermore, machine learning models require large training data in order to make better predictions, which do not exist for a majority of the basins in the world. Finally, unlike process-based models, ML models provide blackbox predictions, which are not easily explainable or interpretable. These make them less useful for modelling physics-driven processes in which the interactions between the underlying variables must be interpretable in order to enhance broader understanding.

[1] **KGE** and **NSE** are the common performance metrics for measuring accuracy of river discharge predictions in hydrology

# 3 Methods

## 3.1 Dataset and Problem Definition

Our ultimate goal is to predict the average amount of water flowing through a particular gauge station per day. Our data is from 1980 to 2010. To achieve this goal, we leverage *in situ* discharge values obtained from the Government of Canada [7], climate forcing variables from Google Earth Engine [6], simulated discharge from the Princeton discharge database [15], river reach widths obtained from Landsat images [4], and river classes originally defined by C. B. Brinkerhoff *et. al* [1].

Although 17 classes were initially defined in [1], we focus on the five largest classes as a proof-of-concept for our proposed approach. We make the following data selection decisions. First, although previous studies [5] have shown that width is a strong predictor of daily river discharge, Landsat4-8 have repeat cycles of 16 days, with some overhead days being too cloudy to pick out river width outlines. As such, we use other features to train an intermediate model to impute widths for the missing days. Secondly, we only consider gauge stations with more than two years of *in situ* discharge data and at least five upstream reaches. This is to ensure that there is sufficient data to quantify the impact of upstream hydrometeorological factors towards daily discharge at a given gauge station. Finally, we difference the non-static independent variables to remove temporal dependence and normalize all data to ensure that they are within the same range, thus maintaining general distribution and ratios in the training data.

## 3.2 Sequential Learning

The standard approach in machine learning is to train, validate, and test models on data drawn from the same distribution (i.i.d); applications of these techniques for river discharge predictions are common in the literature [3, 13, 14]. However, we focus on training models that can perform well on previously unseen data (i.e., ungauged river basins), which is needed for the majority of basins, where *in situ* data are unavailable. Section 4 reports results obtained via transfer learning. By modelling daily discharge prediction as a sequential problem, we can utilize the full power of LSTMs and the historic context of related physics of the hydrologic systems to improve predictions across time and space, both in gauged and ungauged basins. Our preliminary analysis led us to use a Bi-directional LSTM model with 4 layers because additional layers showed no substantial improvement in performance. Furthermore, we choose Swish [18] as the activation function after comparison with existing state-of-the-art activation functions. Finally, we train our Bi-directional LSTM model with L2 regularization to prevent over-fitting and present the results in Section4. In practice, we train *n* models where *n* corresponds to the number of classes selected.

## 3.3 Training and Evaluation Metrics

Both single model and ensemble models [13, 3] trained on basin-wide datasets have demonstrated remarkable results in predicting daily discharge. However, the Mackenzie River basin (where we perform our analyses) has extreme variations in the average discharge across its tributaries and as such, a single model performed relatively similar to the current state-of-the-art process-based-models[10].

As stated in 3.1, we train five models, one for each class of rivers considered. Whereas we designed multiple experiments with varying volumes of observations and meteorological variables to quantify the impact of data quantity and quality towards the model performance, we only report results for one experiment that combines dynamic and static features at a particular gauge station and one upstream reach.

Consider a class with $n$ stations, we can create all possible combinations of classes using Equation( Eq. 2) that vary the type and volume of data available to the model.

$$^{n}C_k = \frac{n!}{k!(n-k)!}; k = 1, 2, ..., n-1 \tag{2}$$

Then, we train a model on each of the selected sets and test on $(n-k)$ held-out stations. For large sets, we randomly select 20 sets at most. Our results consist of distributions across these sets to reduce bias towards a single set of high-performing gauge station datasets. Finally, we choose to report our results based on three major metrics used in hydrology to evaluate river discharge prediction performance; Nash-Sutcliffe Efficiency (NSE) [16], Kling-Gupta Efficiency (KGE) [8], and Relative Bias (RBIAS).

NSE is a normalized statistic that determines the relative magnitude of residual variance compared to the measured data variance. NSE ranges between $(-\infty, 1]$ with $NSE = 1$ being the optimal value. Values between 0.0 and 1.0 are generally viewed as acceptable while values $\leq 0.0$ indicate that the mean of observed values is a better predictor than the predicted value. KGE is based on decomposition of NSE into its constituent components (correlation, variability bias and mean bias). Like NSE, KGE ranges between $(-\infty, 1]$ with $KGE = 1$ being the desired value that indicates perfect agreement between observed and simulated values. Positive KGE values are an indicator of good model performance while negative values are considered undesirable.Finally, RBIAS quantifies the relative systematic bias in the predicted discharge values. A positive or negative value indicates corresponding bias in predicted values respectively while 0.0 shows no bias in the predicted values

Overall, a stable performance should always have KGE values higher than NSE, although it should be noted that NSE and KGE values cannot be directly compared Knoben et al. [12].

# 4 Results

In Table 1, we report statistics of ${}^{n}C_{k}$ combinations of predicted discharge across the five selected classes in ungauged basins (previously unseen data). We compare our results to to the existing state-of-the-art process based models [1] with average scores of NSE and KGE in the range of 0.0 to 0.5. Class one performs poorly as compared to other classes. This is mainly attributed to the smaller widths for rivers in this class as compared other classes. River width is a stronger predictor of discharge relative to other features [5]. Overall, models across the remaining classes are able to generalize well across ungauged basins, as indicated by high values of NSE and KGE, and values of RBIAS close to 0.0, indicating less deviation of models' predictions from the actual observations. These results strongly suggest that machine learning models are better at generalizing hydrological information across ungauged basins compared to the existing state-of-the-art process based models.

Table 1: Statistical distribution of discharge prediction results in ungauged basins. With the exception of class one, mean discharge across the remaining classes outperform state-of-the-art process-based model predictions, which report NSE and KGE values in the range of 0.0 to 0.5.

| | River class | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| KGE | Mean | 0.17 | 0.60 | 0.71 | 0.47 | 0.54 |
| | Median | 0.26 | 0.61 | 0.72 | 0.47 | 0.58 |
| | Max | 0.73 | 0.88 | 0.86 | 0.81 | 0.86 |
| | Min | -1.05 | 0.41 | 0.31 | -0.04 | 0.07 |
| NSE | Mean | -0.28 | 0.58 | 0.72 | 0.27 | 0.47 |
| | Median | 0.10 | 0.62 | 0.74 | 0.41 | 0.50 |
| | Max | 0.62 | 0.84 | 0.87 | 0.84 | 0.81 |
| | Min | -4.77 | 0.26 | 0.35 | -0.72 | -0.54 |
| RBIAS | Mean | 0.23 | -0.03 | -0.06 | 0.01 | 0.09 |
| | Median | 0.17 | -0.03 | -0.07 | -0.01 | 0.07 |
| | Max | 1.95 | 0.30 | 0.47 | 0.79 | 0.71 |
| | Min | -0.57 | -0.29 | -0.42 | 0.77 | -0.44 |

# 5 Conclusion and Future Work

In this paper, we have demonstrated improved performance of machine learning approaches over process-based models for predicting discharge in ungauged basins. However, categorizing basin-wide rivers into classes is a less efficient method because of varying hydrometeorology characteristics across basins. Future work will improve river classification by adopting stream orders or Pfafstetter units since these are more hydrologically-informed approaches for grouping rivers based on climatic regions, geomorphological, and tributary characteristics. Furthermore, we hope to statistically quantify the impact of additional training data, both qualitatively and quantitatively, towards model performance. Finally, this work sets the stage to enable examination of constraints of process-based modelling approaches for predicting river discharge and better characterizing how machine learning based models can be used to model physical processes, not only in hydrology, but also in other other physical sciences.

## 6 Broader Impact

The authors acknowledge that machine learning can be misused, but recognize no situation in which their work, both written and implemented, can be misused. The authors believe that leveraging machine learning in hydrology will greatly improve discharge prediction and help to further understand climate change and its impact on environmental health and economic development.

## References


[1] CB Brinkerhoff, CJ Gleason, D Feng, and P Lin. 2020. Constraining Remote River Discharge Estimation Using Reach-Scale Geomorphology. *Water Resources Research* 56, 11 (2020), e2020WR027949.

[2] Mariel Dirscherl, Andreas J Dietz, Christof Kneisel, and Claudia Kuenzer. 2020. Automated mapping of antarctic supraglacial lakes using a machine learning approach. *Remote Sensing* 12, 7 (2020), 1203.

[3] Dapeng Feng, Kuai Fang, and Chaopeng Shen. 2020. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research* 56, 9 (2020), e2019WR026793.

[4] Dongmei Feng, Colin J Gleason, Xiao Yang, and Tamlin M Pavelsky. 2019. Comparing discharge estimates made via the BAM algorithm in high-order Arctic rivers derived solely from optical CubeSat, Landsat, and Sentinel-2 data. *Water Resources Research* 55, 9 (2019), 7753–7771.

[5] Colin J Gleason and Laurence C Smith. 2014. Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences* 111, 13 (2014), 4788–4791.

[6] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* (2017). `https://doi.org/10.1016/j.rse.2017.06.031`

[7] govermentOfCanada. [n.d.]. Canadian Water Office. `https://wateroffice.ec.gc.ca/`.

[8] Hoshin V Gupta, Harald Kling, Koray K Yilmaz, and Guillermo F Martinez. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology* 377, 1-2 (2009), 80–91.

[9] Ingjerd Haddeland, Jens Heinke, Hester Biemans, Stephanie Eisner, Martina Flörke, Naota Hanasaki, Markus Konzmann, Fulco Ludwig, Yoshimitsu Masaki, Jacob Schewe, et al. 2014. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences* 111, 9 (2014), 3251–3256.

[10] MW Hagemann, CJ Gleason, and MT Durand. 2017. BAM: Bayesian AMHG-Manning inference of discharge using remotely sensed stream width, slope, and height. *Water Resources Research* 53, 11 (2017), 9692–9707.

[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[12] Wouter JM Knoben, Jim E Freer, and Ross A Woods. 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences* 23, 10 (2019), 4323–4331.

[13] Frederik Kratzert, Daniel Klotz, Mathew Herrnegger, Alden K Sampson, Sepp Hochreiter, and Grey S Nearing. 2019. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research* 55, 12 (2019), 11344–11354.

[14] Frederik Kratzert, Daniel Klotz, Guy Shalev, Günter Klambauer, Sepp Hochreiter, and Grey Nearing. 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences* 23, 12 (2019), 5089–5110.

[15] Peirong Lin, Ming Pan, Hylke E Beck, Yuan Yang, Dai Yamazaki, Renato Frasson, Cédric H David, Michael Durand, Tamlin M Pavelsky, George H Allen, et al. 2019. Global reconstruction of naturalized river flows at 2.94 million reaches. *Water resources research* 55, 8 (2019), 6499–6516.

[16] J Eamonn Nash and Jonh V Sutcliffe. 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology* 10, 3 (1970), 282–290.

[17] Sinno Jialin Pan and Qiang Yang. 2010. A Survey on Transfer Learning IEEE Transactions on Knowledge and Data Engineering. *22 (10): 1345* 1359 (2010).

[18] Prajit Ramachandran, Barret Zoph, and Quoc V Le. 2017. Searching for activation functions. *arXiv preprint arXiv:1710.05941* (2017).

[19] Nick Watts, W Neil Adger, Paolo Agnolucci, Jason Blackstock, Peter Byass, Wenjia Cai, Sarah Chaytor, Tim Colbourn, Mat Collins, Adam Cooper, et al. 2015. Health and climate change: policy responses to protect public health. *The lancet* 386, 10006 (2015), 1861–1914.

[20] Thomas Wilbanks, Vatsal Bhatt, Daniel Bilello, Stanley Bull, James Ekmann, William Horak, Y Joe Huang, Mark D Levine, Michael J Sale, David Schmalzer, et al. 2008. Effects of climate change on energy production and use in the United States. *US Department of Energy Publications* (2008), 12.