Highlights

Reservoir Computing for Modeling and Predicting Stream Chemistry

Paden Allsup,Benjamin W. Abbott,Brian Brown,Christophe Giraud-Carrier

- Streamflow and hydrochemistry time series are difficult to model because of variability.
- Echo State Networks are easier to configure and apply than many deep learning models.
- ESNs effectively reproduce chaotic hydrochemical time series compared to LSTMs.

Reservoir Computing for Modeling and Predicting Stream Chemistry

Paden Allsup[a], Benjamin W. Abbott[b], Brian Brown[c] and Christophe Giraud-Carrier[a]

[a]*Brigham Young University - Department of Computer Science*
[b]*Brigham Young University - Department of Plant and Wildlife Sciences*

ARTICLE INFO

Keywords:
Echo State Networks
Long Short-Term Memory Models
Hydrological Modeling
Machine Learning
Chaotic Time Series
Water Quality

ABSTRACT

This paper explores the use of Echo State Networks (ESN), a subset of Reservoir Computing, in modeling and predicting streamflow variability with a focus on biogeochemical patterns. Multiple ESNs were tested alongside a comparable Long Short-term Memory Model (LSTM), another Deep Learning model commonly used in time-series modeling, in the hope of finding a more robust streamflow chemistry predictor. Testing revealed that for our specific modeling of water temperature and dissolved oxygen (DO) levels, ESNs outperforms LSTMs in both model fit and time necessary for training and testing. Our conclusions are that for hydrological tasks where data forms a chaotic time series, ESNs provide a useful and efficient alternative to LSTMs, being quicker to train, providing better results, and being easier to apply to the given task.

CRediT authorship contribution statement

Paden Allsup: Conducted model training, testing and analysis, and writing of the manuscript. **Benjamin W. Abbott:** Advised on application of models to the data, as well as revised manuscript. **Brian Brown:** Advised in gathering data, as well as revised manuscript . **Christophe Giraud-Carrier:** Advised on model selection, application and comparison, as well as revised manuscript.

1. Introduction

The chemistry and flow of water through stream networks impacts human health, economy, and ecological functioning at global scales Díaz et al. (2019); Frei et al. (2021); Basu et al. (2022); Hannah et al. (2022). Hydrochemistry is in turn controlled by complex interactions in the contributing watershed and stream network, including vegetation, direct human disturbance, climate, groundwater dynamics, and water infrastructure Dupas et al. (2019); Godsey et al. (2019); Barbarossa et al. (2020); Goeking and Tarboton (2022); Brown et al. (2023). Accurate prediction of variance in stream flow and chemistry is increasingly important as the global human footprint and disruption of climate put more humans and habitat at risk from flooding, pollution, and ecological collapse Abbott et al. (2023); Hagen et al. (2023); Rockström et al. (2023); Willcock et al. (2023).

The fractal interactions of the factors controlling stream behavior make hydrochemical time series difficult to describe and predict (Blöschl et al., 2019; Kolbe et al., 2019; Brown et al., 2023). Machine Learning (ML) methods have recently been applied to hydrochemical problems with great success and have been shown to be more accurate than traditional physical-based models in some cases Jimeno-Saezm et al. (2022). Consequently, there has been a surge

ORCID(s): 0009-0006-2952-1565 (P. Allsup); 0000-0001-5861-3481 (B.W. Abbott); 0000-0002-0861-8564 (B. Brown); 0000-0002-4129-3239 (C. Giraud-Carrier)

in the use of ML tools for hydrochemical applications, particularly Artificial Neural Networks (ANNs) and the more recent Long Short-term Memory (LSTM) approaches Asadollah et al. (2020). LSTMs are a type of Recurrent Neural Network (RNN) that successfully avoid the vanishing- and exploding-gradient problems common in traditional RNNs, making them highly resistant to bifurcations, which historically have made RNNs difficult to train Doya (1992). LSTMs also integrate internal interactions across time scales, which lets them successfully model spatiotemporal datasets with time-dependent dynamics, such as when previous events influence current and future behavior. This makes them well suited to problems involving hydrochemical prediction Shen et al. (2021).

While LSTMs have successfully predicted streamflow forecasts Hunt et al. (2022a) and more recently water quality Liu et al. (2019); Wang et al. (2017), they are complex and costly to train, which makes it hard to apply them in areas where computational resources are limited. Echo State Networks (ESN), a subset of Reservoir Computing, are significantly simpler than LSTMs, yet they retain some of the beneficial attributes and remain robust to the chaotic variation inherent in water quality time series. ESNs are commonly used as an alternative to RNNs because of their accuracy and ease of use. ESNs and LSTMs differ in model architecture, training methods, and simplicity in modeling and forecasting applications. LSTMs consist of a series of interconnected cells that are made up of "gates" that handle signal propagation, enabling them to both forget unnecessary long-term information and retain important short-term information. ESNs, on the other hand, are composed of a single set of sparsely connected "nodes", called a reservoir, that propagates a signal through to a single output layer which decodes data and whose outcome is a final prediction. The single output layer, called a readout, is the only trainable piece of the network, saving both time and space when compared with other model architectures.

ESNs are notably simpler than more modern Deep Learning models, but are still commonly used for their efficiency and accuracy in spatio-temporal problems. When used for temporal problems, ESNs and LSTMs accept data in the form of a time series, where each data point represents a value, or set of values, at a specific point in time. Datasets are compilations of readings of the same set of features across a timescale and have a an ordering through time. Both ESNs and LSTMs make use of feedback connections which take into account previous timesteps' information while considering future timesteps' outcomes. While LSTMs possess non-linearity in each cell that helps to capture chaotic signal behavior, they often need large networks to handle increasingly complex signals. ESNs possess inherent non-linearity, which comes from the connectivity between reservoir nodes, that allows them to successfully handle largely chaotic time series, and are much easier to train on long-term natural signals Jaeger and Haas (2004).

Where resources and time are not limitations, LSTMs have been shown to provide accurate predictions at the cost of time and complexity Zhou et al. (2018). In cases where resources like memory and compute power are limited, or quick training and prediction are needed, ESNs can serve as a useful alternative to LSTMs. ESNs that have been correctly initialized also possess (and get their name from) the Echo State Property (ESP), which is very similar to the fading memory possessed by LSTMs. In order for ESNs to effectively handle chaotic signals, they must have this property Jaeger (2002). Building an accurate model correctly strikes a balance between the non-linearity of the signal propagation and the memory capability of the model Antonelo et al. (2017). When initialized correctly, Echo State Networks can be an efficient method for handling long-term, multivariate, temporal data. Part of this contribution is to serve as a user-friendly introduction to ESNs in the context of water quality time series.

2. Methodology

When predicting time-series data, especially chaotic natural signals like streamflow, it helps to isolate the chosen features, and train the model separately on each feature of interest. This can help to highlight connections or relationships among tested features, and help the model to accurately predict some of the more chaotic components of streamflow. One common use for Echo State Networks is in future signal generation, which can be extremely valuable for modeling flow regime and longterm hydrochemical patterns. Once the model has been sufficiently trained with long-term data, the model can successfully highlight trends taking place over a long period of time (e.g., the growth of maximum temperature in recent years Pörtner and Roberts (2022)). This project demonstrates the use of ESNs in future signal generation, examines the impact or random reservoir initialization, and directly compares ESNs to a similar LSTM model.

2.1. Water Temperature and Dissolved Oxygen

Water temperature and DO are two of the most important hydrochemical variables affecting stream habitat and impact on human society **?**. Water temperature and DO are directly connected due to the temperature dependence of oxygen solubility and oxygen production by primary producers. These variables have both daily and seasonal variation. The relationship between seasonal and daily variation is difficult to accurately model, but is key to understanding and predicting long-term changes to streamflow Cao et al. (2021).

2.2. Reservoir Size and Connectivity

Similar to LSTM models, the main factors affecting performance of an Echo State Network are the overall network size and the regression regularization factor, (which helps to avoid over-fitting the data). ESN optimization is notoriously difficult, and is often found through trial and error. Optimal reservoir size is highly task-dependent; a reservoir too big or too small dramatically impacts model success in generating an accurate signal. Node connectivity, a hyper-parameter governing the random connections between nodes in the reservoir, also influences the signal generated by the reservoir, which may be either too chaotic, or not chaotic enough, which in either case leads to accurate predictions. ESNs will commonly be initialized with very sparse connectivity rates with the hope that less connectivity between nodes will increase the variation in reservoir response signals, which is good for overall training. Typically a connectivity rate of 1% is used, meaning each node is connected with approximately 1% of the other nodes in the reservoir. With the connectivity rate remaining constant, a network that is too large creates an insensitive signal, which cannot accurately predict minute daily variation. A network too small generates a signal that is too sensitive and becomes even more chaotic than the time series, which also gives inaccurate predictions.

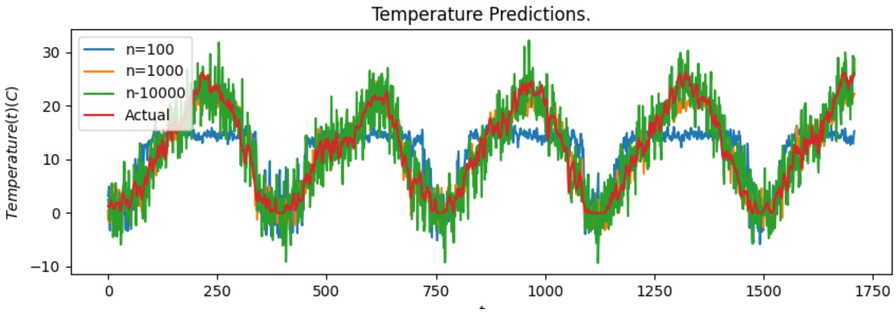

Figure 1: Various reservoir sizes and their effects on signal generation. Reservoir size is perhaps the most important component in developing an accurate model. Too small a reservoir, and the model cannot accurately generate large- or small-scale signal variation, while too large a reservoir places too much emphasis on daily variation, leading to highly chaotic and inaccurate signals.

The effects of various network sizes are shown in Figure 1. Here, three signals based on the same time series were generated by reservoirs with connectivity rates of 1%, and sizes $n = 100$, $n = 1,000$, and $n = 10,000$, representing too small, too large, and a close-to-optimal network sizes. As the size of the reservoir gets smaller, the signal generated cannot differentiate between large and small scale variations. This results in an inability to generate a signal with accurate seasonal variance. When the reservoir size becomes too big, it predicts too much small scale variation, and loses sensitivity. A plot of the actual recorded daily temperature is included for comparison. A round of testing various network sizes showed that in handling our particular datasets, a reservoir size of $n = 1,000$ nodes provided the best signal generation for both temperature and DO level prediction and modeling.

The reservoir state is updated at every timestep, and is governed by the equation

$$x(t+1) = f(Wx(t) + W^{in}u(t+1) + W^{fb}y(t))$$

where $x(t)$ is the reservoir state at timestep t, W is the randomly initialized $N * N$ weight matrix of weights among reservoir nodes, W^{in} represents the randomly initialized $N * K$ matrix of weights between input and reservoir nodes, W^{fb} is the feedback weight matrix of shape $N * L$ from output to reservoir nodes, and $u(t)$ and $y(t)$ represent the input signal of size K and output signal of size L, respectively. The extended state is given by

$$z(t) = [x(t); u(t)]$$

which is passed through an activation function (in our case a sigmoid function) g by multiplying a matrix of output weights W^{out} of shape $L * (K + N)$ and the extended state, $z(t)$:

$$y(t) = g(W^{out} * z(t))$$

. The output signal is then decoded by linear regression and a prediction is made.

2.3. Ridge Regularization

After testing to find the optimal network size, another round of testing various regression and regularization parameters helped to generalize the model for long-term future predictions. ESNs are able to make use of multiple kinds of on- and off-line regression models. For this project, we used a single readout layer which computes a simple Tikhonov linear ridge regression. This regression updates the output weight matrix W^{out} by using the form

$$W^{out} = (R + \lambda I)^{-1} * P$$

, with R being the correlation matrix of the extended reservoir state and P being the cross-correlation matrix of states vs. target outputs. λ, our regularization parameter, is a non-negative smoothing factor multiplied to I, the identity matrix. By experimenting with various values of the regularization parameter, we were able to find good generalization for both temperature and DO. This regularization helps control the signal propagation through the reservoir, and avoid over-fitting on either the daily or seasonal variation. Figure 2 shows the impact of various regularization parameters. While the difference between regularization values is not as noticeable in the generated signals as is the reservoir size, it is still important for maximizing the goodness-of-fit of the network to the chosen task. With a smaller-than-optimal regularization parameter, the generated signal becomes wild and predicts unrealistic daily variance. Larger-than-optimal parameters capture the general trends better, but ultimately produce a signal that is less sensitive to short-term variation. After multiple tests, a ridge regularization value of $1e - 7$ was chosen. This value helped to balance sensitivity between both large-scale seasonal trends and minute daily change.

2.4. Data

Another important consideration relates to the availability of chosen data. Water temperature has consistently been reported daily by the United States Geological Survey (USGS) in many sites dating back to the 1950's or earlier. However, DO and other nutrient recordings are sporadic in most sites before the year 2018, which makes it difficult to find enough long-term data for both training and testing. We found in our initial testing that models trained on the limited amounts of DO datasets were unreliable and inaccurate. In order to circumvent this problem, we added multiple random permutations of the same set of years to our dataset in order to simulate seasonal changes across a

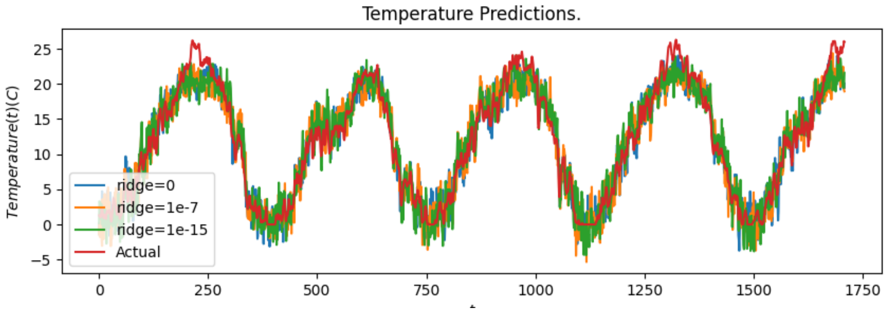

Figure 2: The effects of various ridge regularization parameters. While the effect is small, in combination with reservoir size, ridge regularization can either tame or enhance a generated signal to more accurately reflect small-scale signal generation.

larger time-scale than was available. Our results here are useful as a proof of concept and as a tool for hypothesizing about watershed reactions to extreme events. When extreme events happen, the model can be used to test possible reactions a watershed might have when not accustomed to dramatic events. As will be highlighted below, as the signal to process becomes more complex, the amount of data needed to successfully train an ESN grows at a significant rate. This can significantly affect model performance in scenarios where total amount of data is a limitation. If, on the other hand, the amount of data is not a limitation but the signal is extremely chaotic, and increased amounts of data only add more chaos, an ESN will not be able to successfully predict the signal without an extremely large reservoir. This makes the use of ESNs challenging in situations where compute power is not an issue, but system storage is a limitation.

2.5. Training and Testing

Multiple sites were chosen for training and testing based on similar elevation, discharge, hydrochemical behavior, and general topography. Data for this project came from various USGS gauge stations on the Colorado and Green rivers near the Colorado-Utah border. The site numbers used were: USGS09095500, USGS09261000, and USGS09163500. Both temperature and DO data came from all 3 of these sites but individual models were trained on each site in order to test the model's fit for specific sites. Temperature results from all three models showed little variation in performance. The DO dataset used to create the results provided below came from data gathered at site USGS09095500 on the Colorado river near Cameo, Colorado. This site contained the longest period of recording of daily DO. DO data from the other sites were tested, however we found that the recording periods were too short for our model to accurately reproduce the signal. Temperature data came from site USGS09163500 on the Colorado River near the Colorado-Utah border. Both temperature and DO had maximum, minimum, and mean values recorded daily by the USGS. We found that each produced similar results after training, but here we report results for the modeled mean values.

To build our models, we used a python library called reservoirpy, which makes building and optimizing Echo State Networks straightforward, and has many built-in tools to help fine-tune models for performance Trouvain et al.

(2020). In order to capture the effects of random reservoir initialization, 10 models each for temperature and DO were initialized then trained and tested on the same datasets. The models were identical in size and regularization parameters. A train/test split of approximately 70/30 was chosen (the first 70% of the recorded data was used to train the models and the remaining 30% was used for testing). After training, each model was used to predict the signal pattern for the test portion of the data. For each prediction, the model was given the previous day's value for temperature or oxygen, and asked to predict what the next day's value would be. Each model produced a new time series which was compared to the withheld portion of the data. Model accuracy was recorded and stored in a list for comparison to other models. These results were also plotted for visual comparison to the actual time series.

Comparable LSTM models were trained and tested on the same data splits and generation periods for both temperature and DO. For our LSTM models, we used the python library scalecast, which provides a wrapper over the commonly used TensorFlow Keras LSTM layer, and streamlines LSTMs for use with time-series problems Keith (2024). Scalecast automatically optimizes model performance based on chosen parameters for the given time series. Our LSTM models were initialized on the same training period with the temperature model having a time-lag of 100 steps (each prediction takes into account the previous 100 days' data) and the DO model having a timelag of 50 steps. Both models used the same train/test split as our ESN models, a standard Adam optimizer, and an early-stopping criterion monitoring validation loss for efficient training. Testing with various lengths of time-lag showed that finding the optimal lag input is a difficult problem. In order to compare the simplest usable form of LSTM, a time-lag of 100 days was chosen for temperature, while 50 days was chosen for DO, in order to strike a balance between length of time needed to train and quality of results.

3. Results

3.1. Metrics

Model accuracy was tested using several metrics: Root Mean Square Error (RMSE), R-squared (R^2), and Nash-Sutcliffe Efficiency (NSE). RMSE is a commonly used regression metric to test standard deviation of model predictions from true values, with values closer to 0 representing a more accurate model. A weakness of RMSE is that the return value can be highly relative (a value between 0 and infinity can be returned), which makes it difficult to judge real-world model performance. R^2 provides a solution to this problem, returning a value between 0 and 1, where a value of 1 represents a perfect correlation between predictions and true values, and values closer to 0 represent a lack of or no correlation between predicted and observed values. NSE is very similar to R^2, however, it is primarily used to judge model simulation fit and is commonly used to measure hydrological model accuracy. Together these metrics give a broad view of model performance and give insight into real-world application.

3.2. Temperature

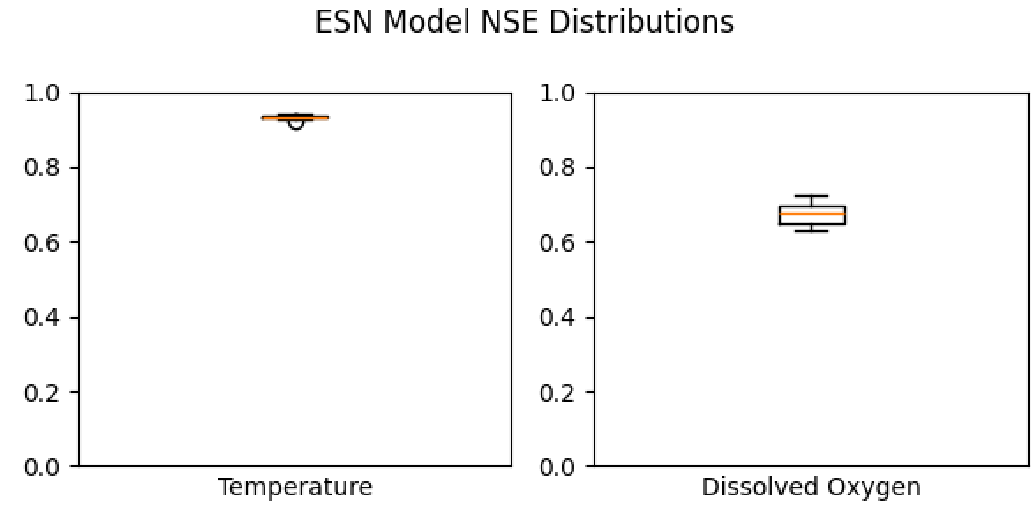

Figure 3: Echo-State Network (ESN) model Nash-Sutcliffe Efficiency (NSE) distribution. Small distributions in model performance suggest that random reservoir initialization has little effect on model performance, showing that the ESN model architecture is a good fit for chaotic, univariate time-series modeling.

With temperature data being plentiful, our ESN models performed very well. With an average R^2 and NSE value of .97 each model successfully generated realistic water temperature time-series on both the seasonal and daily scale, showing that random reservoir generation had little impact on model fit. The NSE distribution is shown in Figure 3. As shown in Figure 4, it seems that the most difficult part for the model to reproduce is the change in daily variance after significant weather events, and at the peaks of the winter and summer seasons. During the second summer season

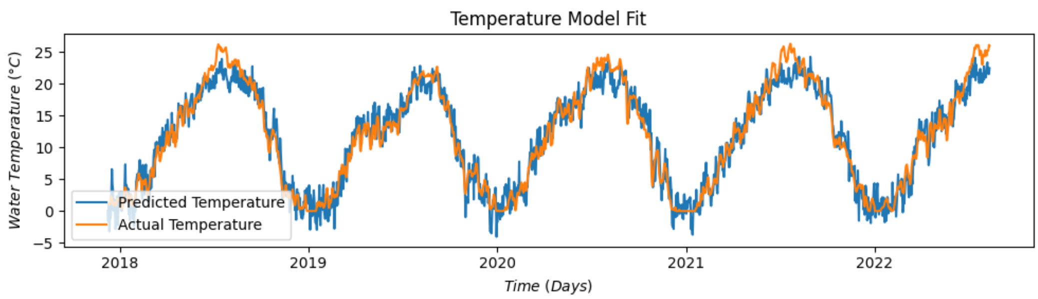

Figure 4: ESN temperature model predicted vs actual time-series. The signal accurately predicts daily and seasonal variation, however struggles to match peak variation at the coldest and warmest seasons. This generated signal matches more closely to the actual signal than the LSTM temperature model does, showing ESNs to be a better fit for this problem than LSTMs for this problem.

in our test years there was a relatively stable period before a large spike just before the peak of the season, where the temperature remained relatively stable for a period of approximately 60 days. During that period, the recorded daily variance of the water temperature was minimal, whereas our model predicted more temperature variance. Other places where the model struggled seem to be during the autumn season where there were dramatic drops in daily temperature. In the winter, days where the actual recorded temperature reaches 0°C represent days when likely the water surrounding the sensors at the USGS gauge site was frozen and therefore a minimum bound was recorded before the water froze, or where water was visibly frozen and so a temperature of 0°C was manually recorded. Our model incorrectly predicted values below freezing for water temperature, although it could be argued that artificially capping the temperature data at 0°C is more problematic given that ice can have temperatures well below zero, and water can still flow beneath ground when the surface is frozen. In other cases where temperature recording accounts for the temperature range of ice, the model would likely match the recorded temperature closer than in this dataset.

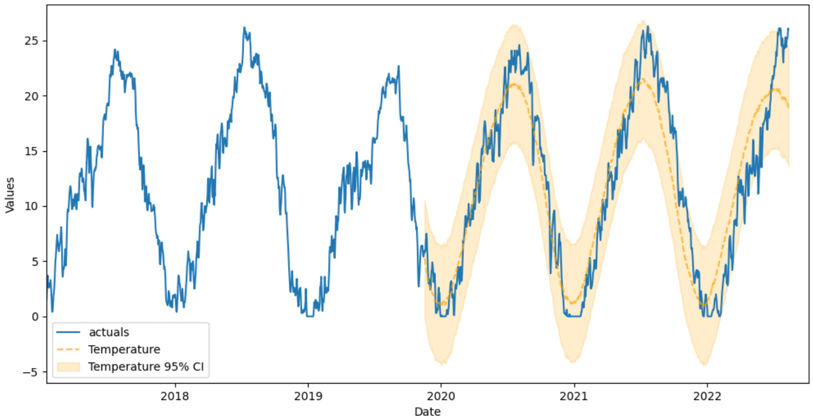

Figure 5: LSTM temperature model predicted vs actual time-series. Despite ample data, the chaotic nature of the training signal prevented the LSTM from accurately generating a signal matching either seasonal or daily variation of the actual signal.

Training an LSTM on our water temperature dataset delivered comparable results to our ESNs, with an NSE value of .965, as shown in Figure 5. The LSTM was able to recreate seasonal variation, but not daily variation. Training also took significantly longer than our ESN, although this was expected because ESNs require little training compared to a more complex LSTM. Though the 95% confidence interval contains almost all the correct test values, the actual predicted signal does not accurately model the short term variation of the time-series. Though a larger and more complex model would likely perform better, that eliminates the benefit of having a simple model to be used where resources or time are limited. Our results show that although a simple and optimized LSTM model provides almost as good a fit for this dataset as out ESN models do, the time needed to achieve the same results is a significant disadvantage towards using an LSTM for time series modeling. These results were not unexpected, but the difference in training time and performance was surprising considering our ESN had almost no optimization, and was predicting based on only the previous day's output, compared to the much longer 100-day period taken into account by the LSTM. Directly

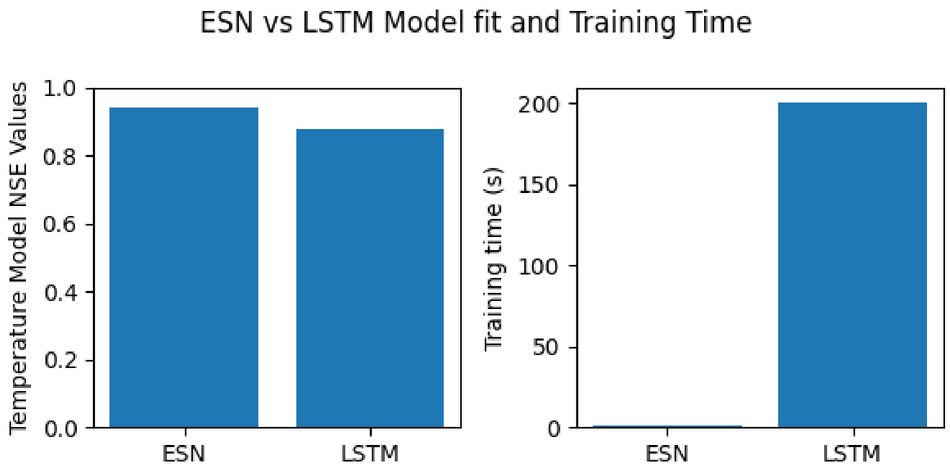

Figure 6: Performance comparison between one of the 10 ESN temperature models and an LSTM trained on the temperature dataset. The LSTM model not only had a worse model fit, but trained almost 100 times slower than the average time needed to train an ESN model. This perfectly highlights the main key advantage ESNs have over LSTMs.

comparing our 10 ESN models with the results from our LSTM temperature model was very interesting. As shown in Figure 6, each of our ESN models was not only a better fit for the time series, but our ESN models trained on average almost 100 times faster than the corresponding LSTM.

3.3. Dissolved Oxygen

Even after augmenting the DO dataset by adding multiple random permutations of the total set in order to simulate extra years' data, there was still significantly less total data than was available for water temperature. With less data available, our models were understandably less accurate than the temperature models. DO ESN models were not as accurate as temperature models, with an average NSE and R^2 value of .71 compared to .92 for the temperature models, however they still performed reasonably well. With the availability of more data, model accuracy would likely improve. As seen in Figure 3, The NSE distribution for dissolved oxygen models show that even with a significantly shorter training period our model is still a reasonably good fit for the watershed, though not good enough to rely on for real-world predictions. In the predicted signal seen in Figure 7, the troubles our DO models had were in discerning at what times there was significant daily variation and when daily DO levels were more stable.

LSTM models were trained on the original DO dataset (with no added random permutations), and the augmented DO data in order to better understand how limited data would affect a more complex model architecture. Similar to ESN performance on the original dataset, the LSTM could not accurately recreate the signal, and struggled to generate correct seasonal or daily generation. Interestingly, even for the modified dataset, LSTM DO predictions were significantly worse than our ESNs, with an NSE and R^2 of 0 and 0.17 respectively. This highlights the same problem experienced by our ESN models above. When presented with a limited amount of data, a basic LSTM model cannot

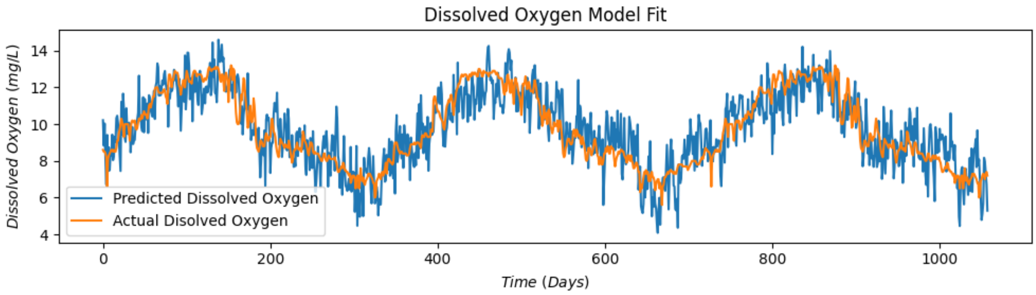

Figure 7: ESN dissolved oxygen model predicted vs. actual time series. The extra sensitivity to small-scale (daily) variation show that the length of the training dataset is significant in determining the accuracy of the generated signal.

accurately replicate highly chaotic signals. The lack of long-term data in such a chaotic series would likely inhibit any model's accuracy, though some might perform better than others. Similar to the temperature results above, the 95% confidence interval contains most of the values, however the actual predicted values rarely overlapped the time-series. With more data the results would likely have resembled the temperature spread from the temperature LSTM model. DO results from the model are shown in Figure 8.

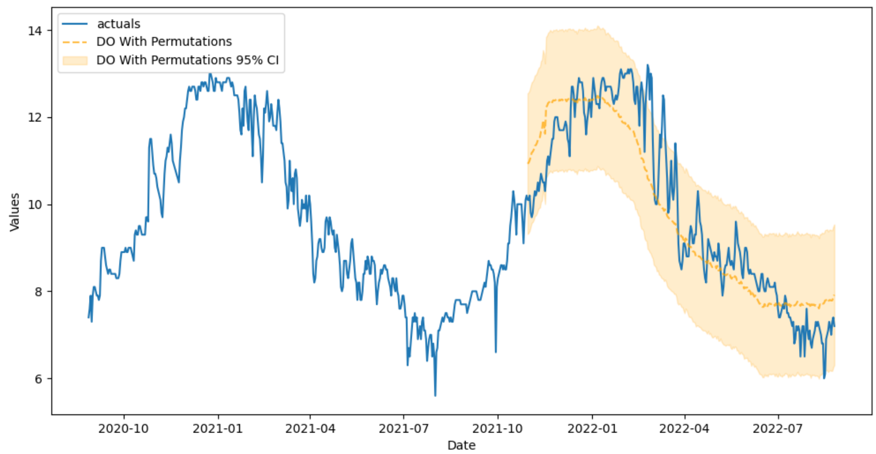

Figure 8: LSTM dissolved oxygen model predicted vs actual time series. Model performance shows adequate seasonal variation in the signal, however, it fails to correctly generate small scale (daily) variation. Similar to the dissolved oxygen ESN above, this suggests that as training sets become smaller, difficulty in generating accurate small-scale signals increases. This makes LSTM models an unoptimal choice for modeling time-series where the amount of training data is a concern.

While both model types were able to predict seasonal variation, a longer training period would likely provide better results in predicting the levels of daily variation for both model types. The clear difference in performance between the temperature and dissolved oxygen models highlights that no model can overcome a lack of sufficient training data. As the period of recording grows larger, the results will likely become more realistic and applicable in real-world scenarios. Interestingly, these results show that with smaller amounts of data, the time needed to train an LSTM was

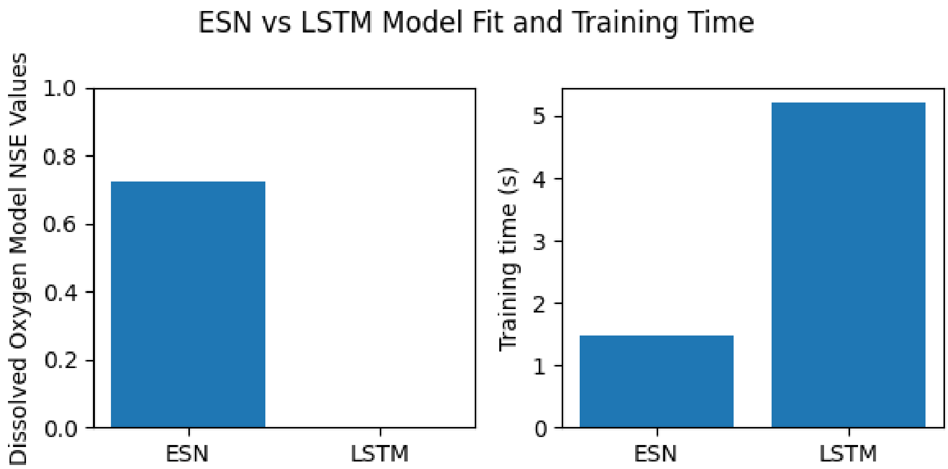

Figure 9: Performance comparison between one of the 10 dissolved oxygen models and an LSTM trained on the augmented dataset. Although the disparity between training time was much smaller, LSTM model fit was significantly worse. This shows that data quantity is a major concern when using LSTMs for time-series modeling.

much closer to the ESN training time, though model fit was significantly worse, shown in Figure 9. Although these results are not particularly groundbreaking they provide a key insight into the potential of ESNs when working with small or incomplete datasets, and highlight the importance of having access to sufficient data for training and testing.

Similar to temperature recordings, the more dramatic nature of the variance during spring and fall seasons made it hard for the model to differentiate between the more stable winter months, and the rest of the year where the recorded levels varied greatly. DO levels are affected by more than just temperature, relying on groundwater discharge, the atmosphere, and light levels which affect the amount of oxygen primary producers (plants) add to the water. During the summer, spring, and fall seasons these contributions from other sources could be responsible for the greater variance found in the recorded amount. Because variance of the signal differs from season to season, it is difficult to build a model that can accurately predict these trends without access to each contributing variable.

4. Discussion and Analysis

This comparison highlights the major advantage ESNs have over LSTMs: in order to generate accurate time series, LSTM models must be deep enough, and have a training set large enough, to handle the chaotic signal variance and balance between small- and large-scale signal behavior. This often means that a sufficiently trained model is too complex and costly to be realistic in a real world scenario. The simplicity of ESNs allows for almost any machine to build and run a model that provides accurate results. A sufficiently deep LSTM would almost certainly be more accurate than our relatively simple ESN architecture, however for the quality of results given, ESNs are a viable choice for quick predictions and signal generation, especially where immediate results are needed. ESNs provide very quick and efficient training and handle chaotic signals well with little optimization compared to more modern Deep Learning models.

4.1. Necessity of Consistent Data

As shown by our results, Echo State Networks can provide effective modeling and generation in long-term streamflow and hydrochemistry prediction problems, and can perform better than state-of-the-art model architectures like LSTMs. The efficiency of their initialization and training make them a good choice for hydrological modeling problems, and they can be extremely sensitive to changes in streamflow dynamics. The temperature models had markedly better results because of the length of the training sets, and by augmenting the available data for DO, we were also able to produce reasonably good results with that model as well. While the lack of sufficient training data inhibits successful real world modeling in this specific watershed, for any watershed where DO data from a longer period is available, our model could be used as a more accurate predictor than a traditional LSTM, with less time and effort needed for model initialization, training and tuning. Where there is sufficient long-term recording periods, a fully trained model could be used either as a control, tracking what a healthy watershed should look like, or as a model of watershed reaction to major events.

Other variables that were initially considered as key metrics were discharge, specific conductance, turbidity, and pH, however no sites were found with enough consistent daily recordings to enable successful training. Many of the recorded periods were far apart, with inconsistent period lengths. These metrics were significantly less autocorrelated than temperature or DO which, combined with the lack of consistent recording periods, made developing an accurate model unrealistic. More advanced Deep Learning architectures may have produced better results with the data available, and if training data were not an issue, ESNs trained on these parameters would likely generate similar results to the models developed in this experiment. For these variables, most sites with large periods of recorded data only contained seasonal recordings (e.g., daily recordings for the summer or winter season, or a few years of monitoring after a major event), which prevented our model from generating an accurate seasonal spread. This highlights the importance of finding consistent, long-term data in developing a model that holds real-world importance.

4.2. One Model, Many Applications

Some water quality metrics are dependant on others, which are more readily available in large quantities. In cases where some variables directly depend on one or more independent feature in the data, it is worth exploring the use of a model fully trained on the independent feature and passed through a relational function to predict dependent variables of interest. In our case, DO levels directly depend on water temperature. Further experiments could use our fully trained temperature model along with salinity values and percent oxygen saturation levels to predict a range for DO levels for modeling or planning purposes. ESNs can also be used with higher-dimensional data, or to generate a single prediction based on multiple previous state outputs. With streamflow chemistry being a dynamic web of interactions between variables, it is worth future effort exploring how a model trained on a specific target could be used to predict other variables contained in the training set by switching the target with the desired variable for prediction.

4.3. Analysis of Echo State Networks

Echo state networks have been shown to be effective in signal processing applications as described above, and we have shown them to be effective in hydrological applications as well. In problems with temporal datasets, ESNs shine as a simple and efficient model architecture that provides accurate temporal predictions and time-series generation. When early RNN algorithms were introduced, they suffered from many problems related to gradient descent (such as bifurcations). This made them hard to apply in real-world scenarios, and led many researchers to explore the use of ESNs as an alternative. Today, thanks to developments like autodifferentiation, RNNs are much more useful. Because of this, Echo State Networks' only advantage over modern RNN architectures is their highly adaptive and quicker training. RNNs today are very effective in solving highly complex signal processing problems like speech recognition Graves et al. (2013). For problems like this, ESNs would likely need unrealistic amounts of memory to create a model sensitive enough to compete with an RNN. It remains to be seen whether ESNs are subsumed or even made irrelevant by modern deep learning techniques in these types of applications. Regardless, in many signal processing problems, ESNs remain a simple, highly effective, and broadly applicable architecture. For their ease of use and accuracy alone, ESNs are an extremely viable ML architecture for time series modeling, especially where compute power, storage, or time are limited resources.

In regards to streamflow dynamics and hydrochemical modeling, Echo State Networks can be used to create realistic models of high-dimensional scenarios, as well as single variable applications like the one shown here. Streamflow dynamics is a challenging area of hydrology, with individual watershed catchments having dramatically different reactions to similar weather events. It is worth exploring the differences in ESN model reaction to extreme weather events when models have been trained on different watershed catchments of similar landscape and topography. In order for this to work, there must be well documented extreme event data on a scale large enough to compare models.

4.4. Ensemble Learning for Hydrological Problems

There is significant potential for future work exploring the use of ESNs in conjunction with models like LSTMs as part of an ensemble to solve water quality problems. Ensemble learning is an effective approach which has been shown to be successful in hydrological applications Zounemat-Kermani et al. (2021). Ensemble learning is a type of meta-learning where multiple models' predictions are combined on a task, and then results are given to a parent model which will learn through training which model is best for the given problem. Models can be chosen based on some threshold or accuracy level in order to maximize model performance on a difficult task, or based purely on predictions from the parent model. Because they are so efficient and easy to implement, Echo state networks can be used in collaboration with other models as part of an ensemble in order to maximize ensemble performance in difficult hydrological tasks.

Ensembles can also be used to increase ESN performance, by helping to stabilize the training and tuning process Wu et al. (2018). One downside to Echo State Networks we found was that our ESN models were relatively unstable, with good results being highly dependent on an optimal combination of hyper-parameters. Because finding the perfect set of parameters was a very difficult problem, this provides an opportunity for ensemble learning to improve robustness and help to stabilize model performance. Because of the natural simplicity of ESNs, many individual models of various layouts and levels of optimization, with different combinations of hyper-parameters, can be combined in an ensemble in order to maximize performance on specific problems. In conjunction with other well-known Machine Learning models for hydrological problems, ESNs can provide insight and help to validate insights and findings gained from other models.

5. Conclusions

5.1. Importance of Monitoring and Prediction Tools

As the effects of climate change become more visible around us, it becomes increasingly important to monitor vital resources in locations where those resources are strained. In the western United States, drought has significantly affected the lives of the approximately 80 million people who live there. In order to consciously and ethically manage resources and keep people safe, there is a great need for tools that can give accurate predictions of water resources. Streamflow chemistry is a key indicator of the quality of those resources, and their importance for biodiversity and overall ecosystem health make successful prediction and monitoring tools an essential part of our efforts to understand and mitigate the effects of climate and land-use change. There is growing interest in applying Machine Learning tools to predict and model streamflow, which has proven to be a very effective combination and helped to better manage limited water resources. Streamflow is made up of chaotic natural signals, which are difficult to model and predict in physical-based or statistical models. Echo State Networks are another application of Machine Learning used to create more robust streamflow predictors which are sensitive to these types of signals. ESNs handle chaotic signals well, and provide another opportunity for real-world modeling and prediction that is accessible to a wider range of scientists due to their ease of use and broad application.

ESNs have already been proposed as an alternative to traditional neural networks and RNNs in rainfall forecasting De Vos (2013), and while LSTMs have been shown to be effective under certain conditions, Hunt et al. (2022b), this project gives an introductory comparison between the two and gives an introduction to 1) predicting hydrochemical behavior of streams and river systems, 2) long-term modelling of these systems, and 3) provides a template for when

ESNs would provide a better fit than LSTMs and other model architectures for water quality time series problems. The success we have shown in applying ESNs to this problem warrants further exploration of their use in the broader field of Hydrology, and more specifically in the field of streamflow hydrochemistry.

5.2. Future Work

One of the most impressive features of ESNs is their dynamic reservoir memory, and how that memory is affected by model feedback. Many forms of online training make special use of these feedback connections, which can be beneficial as the signals become more complex. Future efforts comparing and contrasting the use of these forms of training, and their effects on model feedback in cases with extremely complex signals, have the potential to benefit humanity by creating more robust and accurate tools for water quality prediction. It is also worth exploring the use of ESNs in predicting reaction patterns of DO to other key variables like turbidity, percent oxygen saturation, and primary producer activity in a more high-dimensional space. This problem is of particular interest in areas where flow regimes are affected by discharge from joining river systems, dam construction and regulation, and unique biochemical processes Zhong et al. (2021). ESNs could provide key insights into this problem in areas where remote sensing and monitoring are essential to measuring watershed health.

Code and Data Availability

Language: python 3.11.0

Software required: reservoirpy, scalecast, keras, tensorflow, pandas, hydroeval, matplotlib

The source code, data, and manuscript are available for downloading at the link: https://doi.org/10.5281/zenodo.12584470

Contact: r.allsup123@gmail.com +1-801-427-7243

References

Abbott, B.W., Abrahamian, C., Newbold, N., Smith, P., Merritt, M., Sayedi, S.S., Bekker, J., Greenhalgh, M., Gilbert, S., King, M., Lopez, G., Zimmermann, N., Breyer, C., 2023. Accelerating the Renewable Energy Revolution to Get Back to the Holocene. Earth's Future 11, e2023EF003639. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023EF003639, doi:10.1029/2023EF003639.

Antonelo, E.A., Camponogara, E., Foss, B., 2017. Echo state networks for data-driven downhole pressure estimation in gas-lift oil wells. Neural Networks 85, 106–117.

Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M., 2020. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. Journal of Environmental Chemical Engineering 9.

Barbarossa, V., Schmitt, R.J.P., Huijbregts, M.A.J., Zarfl, C., King, H., Schipper, A.M., 2020. Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. Proceedings of the National Academy of Sciences 117, 3648–3655. URL: https://pnas.org/doi/full/10.1073/pnas.1912776117, doi:10.1073/pnas.1912776117.

Basu, N.B., Van Meter, K.J., Byrnes, D.K., Van Cappellen, P., Brouwer, R., Jacobsen, B.H., Jarsjö, J., Rudolph, D.L., Cunha, M.C., Nelson, N., Bhattacharya, R., Destouni, G., Olsen, S.B., 2022. Managing nitrogen legacies to accelerate water quality improvement. Nature Geoscience 15, 97–105. URL: https://www.nature.com/articles/s41561-021-00889-9, doi:10.1038/s41561-021-00889-9.

Brown, B.C., Fullerton, A.H., Kopp, D., Tromboni, F., Shogren, A.J., Webb, J.A., Ruffing, C., Heaton, M., Kuglerová, L., Allen, D.C., McGill, L., Zarnetske, J.P., Whiles, M.R., Jones, J.B., Abbott, B.W., 2023. The Music of Rivers: The Mathematics of Waves Reveals Global Structure and Drivers of Streamflow Regime. Water Resources Research 59, e2023WR034484. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023WR034484, doi:10.1029/2023WR034484.

Cao, W., Zhang, Z., Liu, Y., Bend, L., Weng, S., Xu, H., 2021. Seasonal differences in future climate and streamflow variation in a watershed of northern china. Journal of Hydrology: Regional Studies 38. URL: https://doi.org/10.1016/j.ejrh.2021.100959.

De Vos, N.J., 2013. Echo state networks as an alternative to traditional artificial neural networks in rainfall-runoff modelling. Hydrol. Earth Syst. Sci. 17, 253–267. URL: https://doi.org/10.5194/hess-17-253-2013.

Doya, K., 1992. Bifurcations in the learning of neural networks, in: IEEE International Symposium on Circuits and Systems, IEEE.

Dupas, R., Abbott, B.W., Minaudo, C., Fovet, O., 2019. Distribution of Landscape Units Within Catchments Influences Nutrient Export Dynamics. Frontiers in Environmental Science 7, 43. URL: https://www.frontiersin.org/article/10.3389/fenvs.2019.00043/full, doi:10.3389/fenvs.2019.00043.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I., Willis, K.J., Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science 366, eaax3100. URL: https://www.science.org/doi/10.1126/science.aax3100, doi:10.1126/science.aax3100.

Frei, R.J., Lawson, G.M., Norris, A.J., Cano, G., Vargas, M.C., Kujanpää, E., Hopkins, A., Brown, B., Sabo, R., Brahney, J., Abbott, B.W., 2021. Limited progress in nutrient pollution in the U.S. caused by spatially persistent nutrient sources. PLOS ONE 16, e0258952. URL: https://dx.plos.org/10.1371/journal.pone.0258952, doi:10.1371/journal.pone.0258952.

Godsey, S.E., Hartmann, J., Kirchner, J.W., 2019. Catchment chemostasis revisited: Water quality responds differently to variations in weather and climate. Hydrological Processes 33, 3056–3069. URL: https://onlinelibrary.wiley.com/doi/10.1002/hyp.13554, doi:10.1002/hyp.13554.

Goeking, S.A., Tarboton, D.G., 2022. Variable Streamflow Response to Forest Disturbance in the Western US: A Large-Sample Hydrology Approach. Water Resources Research 58, e2021WR031575. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021WR031575, doi:10.1029/2021WR031575.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech recognition with deep recurrent neural networks, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE.

Hagen, J.S., Hasibi, R., Leblois, E., Lawrence, D., Sorteberg, A., 2023. Reconstructing daily streamflow and floods from large-scale atmospheric variables with feed-forward and recurrent neural networks in high latitude climates. Hydrological Sciences Journal 68, 412–431. URL: https://www.tandfonline.com/doi/full/10.1080/02626667.2023.2165927, doi:10.1080/02626667.2023.2165927.

Hannah, D.M., Abbott, B.W., Khamis, K., Kelleher, C., Lynch, I., Krause, S., Ward, A.S., 2022. Illuminating the 'invisible water crisis' to address global water pollution challenges. Hydrological Processes 36, e14525. URL: https://onlinelibrary.wiley.com/doi/10.1002/hyp.14525, doi:10.1002/hyp.14525.

Hunt, K.M.R., Matthews, G.R., Pappenberger, F., Prudhomme, C., 2022a. Using a long short-term memory (lstm) neural network to boost river streamflow forecasts over the western united states. Hydrology and Earth System Sciences [preprint] 53.

Hunt, K.M.R., Matthews, G.R., Pappenberger, F., Prudhomme, C., 2022b. Using a long short-term memory (lstm) neural network to boost river streamflow forecasts over the western united states. Hydrological Earth Systems Science .

Jaeger, H., 2002. Short term memory in echo state networks. Technical Report. Fraunhofer Institute for Autonomous Intelligent Systems.

Jaeger, H., Haas, H., 2004. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science 304, 78–80.

Jimeno-Saezm, P., Martinez-Espana, R., Casali, J., Perez-Sanchez, J., Senet-Aparicio, J., 2022. A comparison of performance of swat and machine learning models for predicting sediment load in a forested basin, northern spain. CATENA 212.

Keith, M., 2024. scalecast. URL: https://scalecast.readthedocs.io/en/latest/.

Liu, P., Wang, J., Sangaiah, A.K., Xie, Y., Yin, X., 2019. Analysis and prediction of water quality using lstm deep neural networks in iot environment. Sustainability 11.

Pörtner, H.O., Roberts, D.C.e.a., 2022. Climate change 2022: Impacts, adaptation and vulnerability working group ii contribution to the sixth assessment report of the intergovernmental panel on climate change.

Rockström, J., Gupta, J., Qin, D., Lade, S.J., Abrams, J.F., Andersen, L.S., Armstrong McKay, D.I., Bai, X., Bala, G., Bunn, S.E., Ciobanu, D., DeClerck, F., Ebi, K., Gifford, L., Gordon, C., Hasan, S., Kanie, N., Lenton, T.M., Loriani, S., Liverman, D.M., Mohamed, A., Nakicenovic, N., Obura, D., Ospina, D., Prodani, K., Rammelt, C., Sakschewski, B., Scholtens, J., Stewart-Koster, B., Tharammal, T., Van Vuuren, D., Verburg, P.H., Winkelmann, R., Zimm, C., Bennett, E.M., Bringezu, S., Broadgate, W., Green, P.A., Huang, L., Jacobson, L., Ndehedehe, C., Pedde, S., Rocha, J., Scheffer, M., Schulte-Uebbing, L., De Vries, W., Xiao, C., Xu, C., Xu, X., Zafra-Calvo, N., Zhang, X., 2023. Safe and just Earth system boundaries. Nature 619, 102–111. URL: https://www.nature.com/articles/s41586-023-06083-8, doi:10.1038/s41586-023-06083-8.

Shen, C., Chen, X., Laloy, E., 2021. Editorial: Broadening the use of machine learning in hydrology. Water .

Trouvain, N., Pedrelli, L., Dinh, T.T., Hinaut, X., 2020. Reservoirpy: an efficient and user-friendly library to design echo state networks. International Conference on Artificial Neural Networks , 494–505.

Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L., 2017. Water quality prediction method based on lstm neural network, in: 12th International Conference on Intelligent Systems and Knowledge Engineering.

Willcock, S., Cooper, G.S., Addy, J., Dearing, J.A., 2023. Earlier collapse of Anthropocene ecosystems driven by multiple faster and noisier drivers. Nature Sustainability 6, 1331–1342. URL: https://www.nature.com/articles/s41893-023-01157-x, doi:10.1038/s41893-023-01157-x.

Wu, Q., Fokoue, E., Kudithipudi, D., 2018. An ensemble learning approach to the predictive stability of echo state networks. Journal of Informatics and Mathematical Sciences 10, 181–199.

Zhong, M., Liu, S., Li, K., Jiang, H., Jiang, T., Tang, G., 2021. Modeling spatial patterns of dissolved oxygen and the impact mechanisms in a cascade river. Frontiers in Environmental Science .

Zhou, J., Wang, Y., Xiao, F., Wang, Y., Sun, L., 2018. Water quality prediction method based on igra and lstm. Water 10.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. Journal of Hydrology .