2022-11-29

Reservoir Computing Solutions for Streamflow Modeling and Prediction in Real World Scenarios

Paden Allsup

Honors Thesis

RESERVOIR COMPUTING SOLUTIONS FOR STREAMFLOW MODELING AND PREDICTION IN REAL WORLD SCENARIOS

by

Paden D. Allsup

Submitted to Brigham Young University in partial fulfillment

of graduation requirements for University Honors

Computer Science Department

Brigham Young University

December 2022

Advisor: Christophe Giraud-Carrier

Honors Coordinator: Seth Holladay

ABSTRACT

RESERVOIR COMPUTING SOLUTIONS FOR STREAMFLOW MONITORING AND PREDICTION IN REAL WORLD SCENARIOS

Paden D. Allsup

Computer Science Department

Bachelors of Science

With severe drought continuing in the western United States and the effects of climate change becoming more apparent across the world, it is becoming increasingly important to be able to predict the impact of extreme weather events like storms, droughts, and fires on streamflow dynamics [3]. This includes flow regime as well as biogeochemical behavior of river systems and their watersheds [5]. This project explores the use of Echo State Networks (ESN), a subset of Reservoir Computing, on modeling and predicting streamflow variability with a focus on biogeochemical patterns. In this project ESNs are tested and compared in the hope of creating more robust streamflow chemistry predictors that are applicable in broader scenarios than what are commonly needed for Machine Learning applications to Hydrological problems.

Reservoir Computing models are proven to be an effective model for multivariate time series problems like streamflow prediction, (problems with more than one time-dependent variable, where each variable depends on both its past readings, as well as

its relation to other variables) [14]. ESNs have been in use since the late 1990's, but remain less well-known than more modern Deep Learning models [15]. ESNs are an efficient Machine Learning model, and their inherent non-linearity makes them very dynamic and able to adapt to training quickly. This makes ESNs a good potential fit for large-scale environmental signal-processing and remote sensing problems [9]. We also compare ESNs with a modern Long Short-term Memory (LSTM) model, which is frequently used for streamflow problems, and provide a template for when one model should be picked over the other.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the generous help of Dr. Christophe Giraud-Carrier, Dr. Ben Abbott, and Brian Brown, who have all been instrumental in helping the author consolidate such an interdisciplinary project into a legible and coherent format. This also would not have been possible without the help of Katherine Allsup, an endless source of encouragement, enthusiasm, and support.

TABLE OF CONTENTS

Chapter

LIST OF FIGURES

Figure

CHAPTER 1

Introduction and Methodology

1.1 Introduction

Watershed health is incredibly important to surrounding ecosystems, with healthy watersheds providing benefits like carbon storage/transfer, erosion control, and soil/nutrient redistribution. The health of a watershed serves as a key metric for the health of the greater ecosystem. Accurate prediction of streamflow variance is increasingly important as the impact of climate change on our local environments grows. Streamflow dynamics, more specifically hydrochemical behavior or streams and rivers, are a complex web of interdependent variables which makes them difficult to accurately model. Two of the most important variables affecting biodiversity and resistance to extreme weather events are water temperature and dissolved oxygen levels [23]. Water temperature and dissolved oxygen are directly connected because of 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 model, but is key for understanding and predicting long-term changes to streamflow [24]. The relationship between these two variables is shown in figure 1.1.

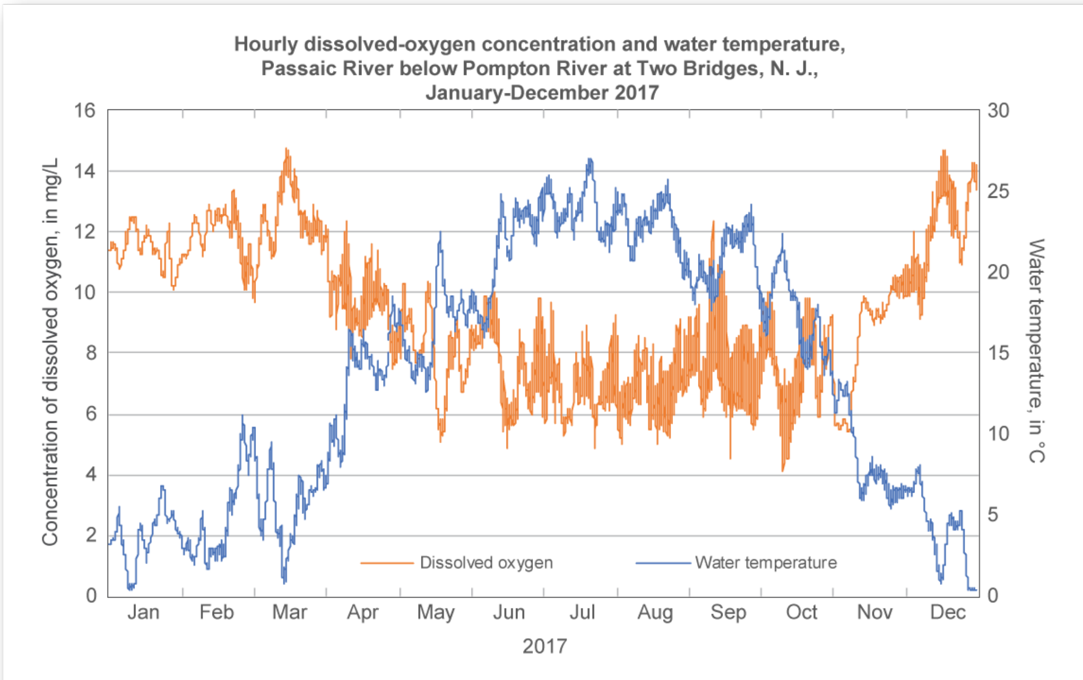

Figure 1.1: Relationship between Dissolved Oxygen and Water Temperature.

Machine Learning (ML) models have recently been applied to hydrological problems, and have been shown to be more accurate than traditional physical-based models [18]. As such, there has been a recent surge in the use of ML tools for hydrochemical and streamflow applications [22]. There are a few methods that have gained popularity for providing accurate and dynamical models for both monitoring and prediction, namely traditional Artificial Neural Networks (ANNs) and the more recent Long Short-term Memory (LSTM) model. LSTMs are a type of Recurrent Neural Network (RNN) that successfully avoids the vanishing- and exploding-gradient problems common in traditional RNNs, making them highly resistant to bifurcations, which historically have made RNNs difficult to train [13]. LSTMs also possess an internal context with helps them successfully model spatio-temporal datasets and time-dependent problems, where context relating to previous events is needed in order to understand future behavior. This makes them well suited to problems surrounding streamflow prediction [4]. Recently LSTMs have been used to predict water quality [19, 25].

While LSTMs have been shown to successfully predict streamflow forecasts [21], they are complex and costly to train, which makes it hard to apply the them in areas where computational resources are limited. Echo State Networks (ESN), a subset of Reservoir Computing, are significantly simpler, yet remain robust, and similar to LSTMs avoid the vanishing- and exploding-gradient problems found in more traditional RNNs. 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. Figure 1.2 shows a generic LSTM cell. ESNs, on the other hand, are composed of a single set of sparsely con-

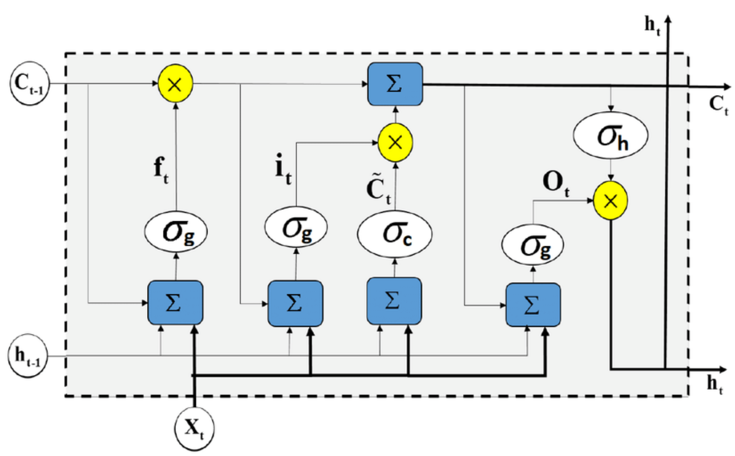

Figure 1.2: The structure of a vanilla LSTM model cell, where c_t is the cell input activation vector, h_t is the hidden state or output vector similar to the hidden state in a traditional RNN, and x_t represents the input vector to the unit itself, while f_t, i_t, O_t represent the forget gate, input gate, and output gate respectively.

nected "nodes", called a reservoir, that propagates a signal through to a single output layer which decodes data and whose outcome is a final prediction, as shown in Figure 1.3. The single output layer, called a readout, is the only trainable piece of the network, saving both time and space when compared with other models.

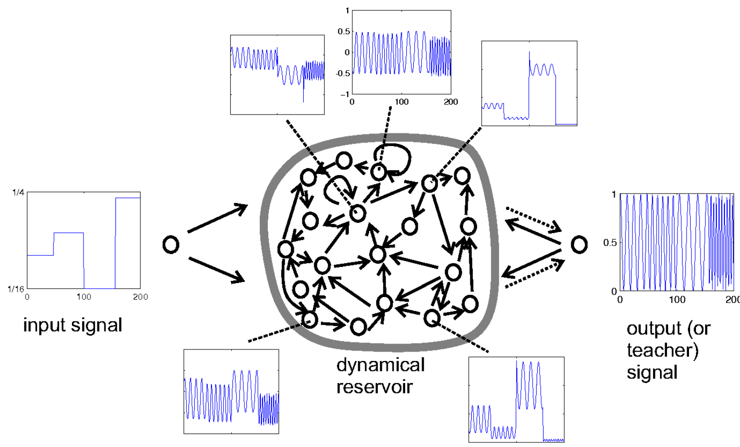

Figure 1.3: Generic structure for an Echo State Network, containing a dynamical reservoir and a single layer readout (in this example the layer consists of a single node)

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 time sequences, where each data reading represents values at a single point in time. Datasets are compilations of readings of the same set of features across a timescale, and are meant to be read in sequential order. 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 an 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.

Where resources and time are not limitations, LSTMs have been shown to provide accurate predictions at the cost of time and complexity [12]. In cases where resources like memory and compute power are limited, or quick training and prediction are needed, ESNs can serve as an alternative to an LSTM. 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 [7]. Building an accurate model correctly strikes a balance between the non-linearity of the signal propagation and the memory capability of the model [1]. When initialized correctly, Echo State Networks can be an efficient method for handling long-term, multivariate, temporal data.

1.2 Methodology

When predicting on time-series data, especially chaotic natural signals like streamflow, it helps to isolate the chosen features, and train the model on each feature of interest. This can help to highlight connections or relationships between 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 hydrochemical patterns. Once the model has been sufficiently trained with longterm 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 [10]). This project demonstrates the power of long-term future signal generation, and tests multiple networks on both water temperature and dissolved oxygen level prediction.

1.2.1 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 to determine, and is often found through trial and error. Optimal reservoir size is highly task-dependent, and dramatically impacts the ability to generate an accurate signal. Node connectivity, a hyperparameter governing the random connections between nodes in the reservoir, influences the signal generated by the reservoir, which may be either too chaotic, or not chaotic enough, which prevents 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 too large muddies the signal, which under-fits the time series and 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.

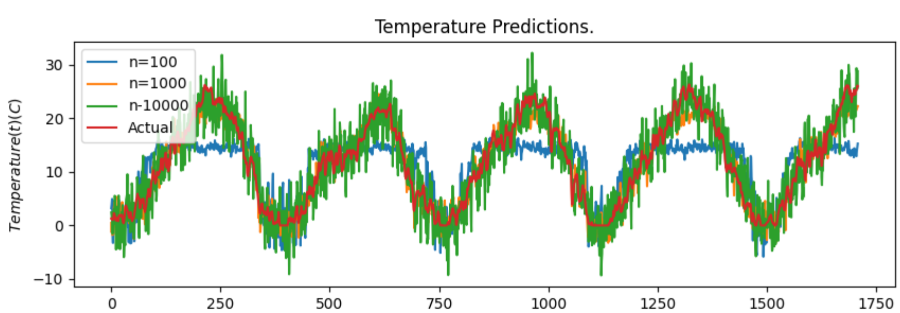

Figure 1.4: Various reservoir sizes and their effects on signal generation

The effects of various network sizes are shown in Figure 1.4. Here, three signals based on the same time series were generated by reservoirs with connectivity rates of 1%, and sizes $n = 100$, $n = 1000$, and $n = 10000$, representing too small, too large, and a close to optimal network size. 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 $n = 1000$ nodes provided the best signal generation for both temperature and dissolved oxygen 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 between 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.

1.2.2 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 for the regularization parameter, we were able to find good generalization for both temperature and dissolved oxygen. This regularization helps control the signal propagation through the reservoir, and avoid over-fitting on either the daily or seasonal variation. Figure 1.5 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 muddier and less sensitive signal. 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.

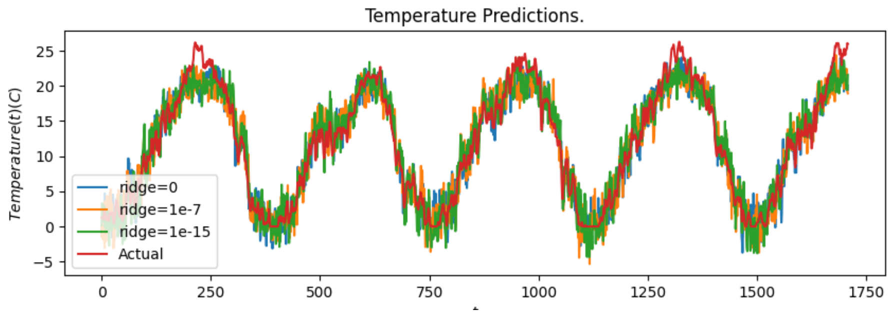

Figure 1.5: Ridge regularization

1.2.3 Data

Another important consideration relates to the 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. However, dissolved oxygen recordings were sporadic in most sites before the year 2018, which makes it difficult to find enough long-term data to both train and test on. We found in our initial testing that models trained on the limited amounts of dissolved oxygen 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 larger time-scale than was available. While this method does not produce the most accurate real-world predictions of long-term dissolved oxygen behavior, it is useful for both modeling purposes and general tracking of watershed reaction to extreme events like floods and wildfires. 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 in our results, as the signal to process becomes more complex, the amount of data needed to successfully train an ESN grows at a significant rate, which can make applying the model difficult in scenarios where total amount of data is a limitation. If, on the other hand, 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 difficult in situations where compute power is not an issue, but system storage is a limitation.

1.2.4 Training and Testing

Multiple drainage locations were chosen for training and testing based on similar elevation, discharge, hydrochemical behavior, and general topography. Data for this project came from various USGS sites on the Colorado and Green rivers near the Colorado-Utah border. The site numbers used were: USGS09095500, USGS09261000, and USGS09163500. Both temperature and dissolved oxygen data came from all 3 of these sites but individual models were trained on each site individually in order to test the model's fit for specific sites. Temperature results from all three models showed little variation in performance, . The dissolved oxygen dataset used on 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 dissolved oxygen. Dissolved oxygen data from the other sites was 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 dissolved oxygen metrics had maximum, minimum, and mean values recorded daily by the USGS. We found that each produced similar results after training, so the mean was chosen to report on.

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. In order to capture the effects of random reservoir initialization, 10 models identical in size and regularization, for both temperature and dissolved oxygen were initialized then trained and tested on the same datasets. 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 asked 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 outputted 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.

CHAPTER 2

Results, Discussion and Analysis

2.1 Results

2.1.1 Metrics

In testing model accuracy, multiple metrics were used to track model fit: 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 accuracy of a model. 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.

2.1.2 Temperature

With temperature data being plentiful, our model performed very well. Each of the ten models had high R^2 and NSE values, low values of RMSE, and successfully generated realistic water temperature series on both the seasonal and daily scale.

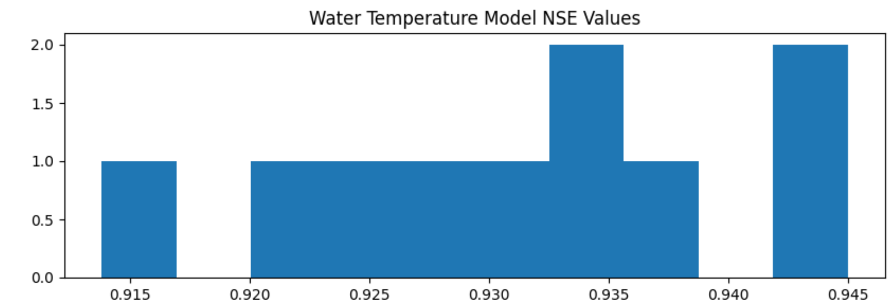

Figure 2.1: Water Temperature Model NSE values

The NSE distribution is shown in Figure 2.1. NSE values for each of the ten temperature models performed well, despite random reservoir generation. With an average NSE value of .933, our model provides a very good fit for predicting water temperature of our chosen section of the Colorado river. As shown in Figure 2.2, it seems that the most difficult part for the model to reproduce is the change in daily variance after significant weather events. During the second summer season in our

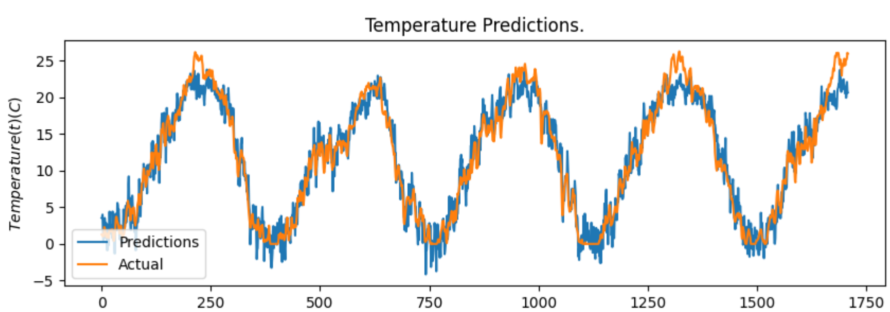

Figure 2.2: Temperature model prediction vs. observed temperature values

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 is not capped, the model would likely match the found temperature closer than in this dataset.

2.1.3 Dissolved Oxygen

With less data available for dissolved oxygen, our model was understandably less accurate. After adding random permutations of previous data to our training set, our model performed significantly better. Even after augmenting the data 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. Dissolved oxygen models were not as accurate as temperature models, but still had consistently good results. With the availability of more data, model accuracy would improve.

The NSE values in Figure 2.3 have an average of .664 and show that even

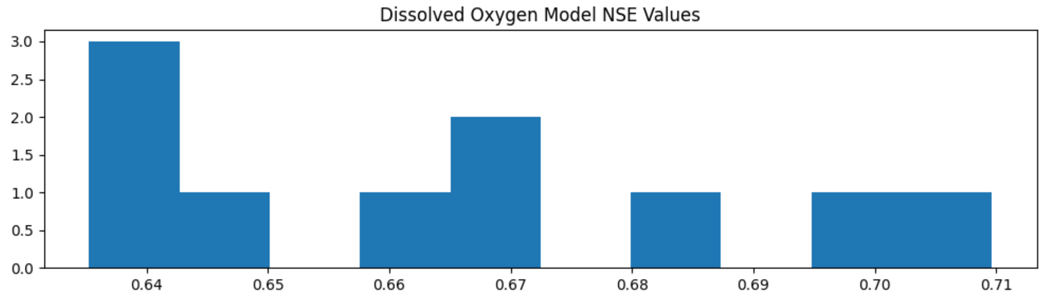

Figure 2.3: Dissolved Oxygen Model NSE values

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 2.4, the troubles our oxygen model had were in determining at what times there was significant daily variation, and when daily dissolved oxygen levels were more stable.

While capturing the seasonal trends relatively well, a longer training period would likely provide better results in predicting the levels of daily variation, and though this model is perfectly usable in a modeling application, we would hesitate to use its predictions to make decisions regarding real-world watershed health. As the period of recording grows larger, the results will become more applicable in real-world scenarios. Though these results are not particularly groundbreaking they provide a key insight into the potential of ESNs in dissolved oxygen prediction, 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. Dissolved Oxygen levels are affected by more than just temperature, relying on groundwater discharge, the atmosphere itself, and light levels which affect the

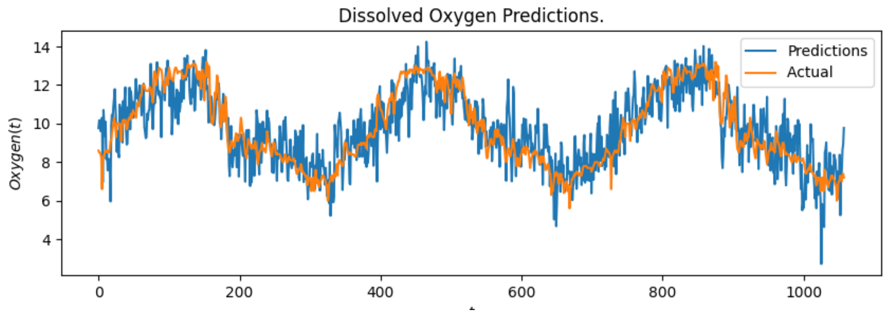

Figure 2.4: Dissolved oxygen model predicted vs. observed oxygen levels

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 recorded amount. Because the chaos 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.

2.1.4 Comparison with a similar LSTM

We also measured the performance of a comparable LSTM on the same data splits and generation periods. For our LSTM model, we used a python library called scalecast, which provides a wrapper over the commonly used TensorFlow Keras LSTM layer, which streamlines LSTMs for use with time series problems, and automatically optimizes model performance based on chosen parameters. Our LSTM model was initialized on the same training period with a time-lag of 50 steps (each prediction takes into account the previous 50 days' data), the same train/test split as our ESN, 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 period to use as a lag input takes an incredibly long time. In order to compare the simplest usable form of LSTM, an arbitrary value of 50 days was chosen in order to strike a balance between length of time needed to train and quality of results.

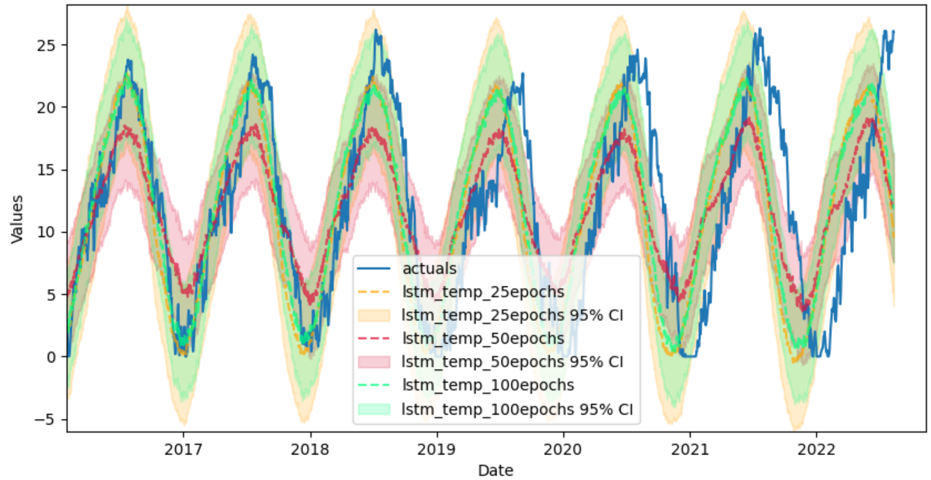

Figure 2.5: LSTM Temperature Predictions.

Training the LSTM on our water temperature dataset delivered reasonably good results, though the model performed slightly worse than our ESN as shown in Figure 2.5. 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 is not a good fit for the time-series. Though a more complex model would perform markedly better, that eliminates the benefit of having a simple model to be used where resources are limited. Our results show that the simplest LSTM model does not provide as good a fit for this dataset even though it had multiple optimizers and took into account a longer prior period in order to make predictions than did our ESN. These results were not unexpected, but the difference in accuracy was surprising considering our ESN had almost no optimization, and was purely predicting based on the previous day's output, compared to the much longer 50-day period taken into account by the LSTM.

A separate LSTM model was also tested on the original dissolved oxygen dataset, (with no added random permutations), to see if a more complex Deep Learning model could handle the smaller amounts of data. Interestingly, dissolved oxygen predictions were significantly worse than our other model, and highlighted the same problem experienced by our ESN above. When presented with a limited amount of data, a basic LSTM model cannot 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 were very far off. With more data the results would likely have resembled the temperature spread from the temperature LSTM model. Dissolved Oxygen results from the model are shown in Figure 2.6.

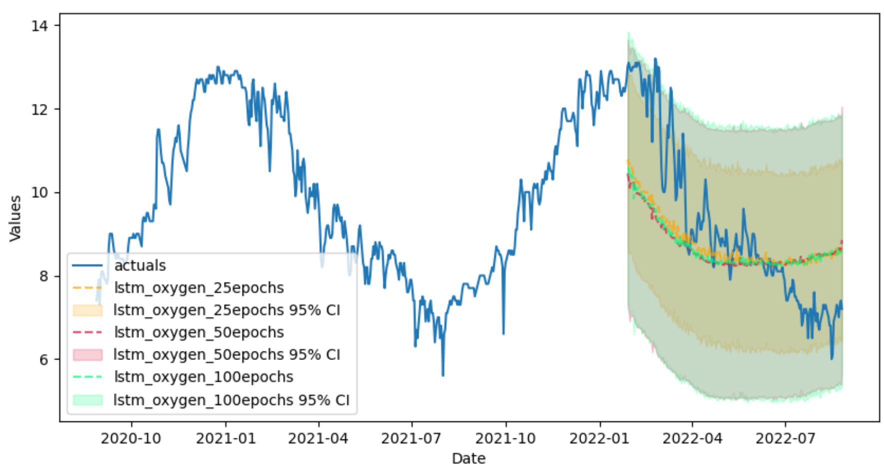

Figure 2.6: LSTM Dissolved Oxygen Predictions.

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 short-term and long-term 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 allow for almost any machine to build and run a model that provides accurate results. A sufficiently deep LSTM would 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. Where time and computational resources are not a limiting factor, an optimized LSTM would likely provide better results than the more simple ESN.

2.2 Discussion and Analysis

2.2.1 Necessity of Consistent Data

As shown by our results above, Echo State Networks can provide very good signal modeling and generation in long-term streamflow and hydrochemistry prediction problems. 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. This can be very helpful when studying the impact of extreme weather events on watersheds. The temperature models had markedly better results because of the length of the training sets, though by augmenting the available data our dissolved oxygen model was also able to produce reasonably good results. While the lack of data eliminates our dissolved oxygen model's use in this specific watershed, any watershed where dissolved oxygen data from a longer period is available, our model could be used as a predictor. Where there is sufficient data, 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. Much of the recorded periods were far apart, with no consistent periods, or long-term recordings. These variables were significantly more chaotic than temperature or dissolved oxygen which, combined with the lack of consistent data, made designing an accurate model very difficult. More advanced Deep Learning models may have produced better results with the data available, but ESNs would likely produce similar results to the variables focused on in this experiment if data were not an issue. 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). This highlights the importance of finding consistent, long-term data in developing a model that holds real-world importance.

2.2.2 One Model, Many Applications

As an alternative to having separate models handle individual variables, in cases where some variables directly depend on one or more independent variables, it is worth exploring the use of a model fully trained on the independent variable and passed through a relational function to predict dependent variables of interest. In our case, dissolved oxygen 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 dissolved oxygen 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 an dynamic web of interactions between variables, it is worth 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.

2.2.3 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 discussed above). 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 the quicker and highly adaptive training. RNNs today are very effective in solving highly complex signal processing problems like speech recognition [2]. In this kind of application, 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.

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.

2.2.4 Ensemble Learning for Hydrological Problems

Recent publications in Hydrology have used many Machine Learning models for various modeling and prediction purposes, and have shown that ML applications to hydrological problems have generally been successful, especially when used as part of an ensemble [17]. 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 [20]. 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.

CHAPTER 3

Conclusion

3.1 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 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 the growing amount of extreme weather events such as wildfires, droughts, and floods. 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 [11]. This project explores the use of ESNs in predicting hydrochemical behavior of streams and river systems, and in long-term modelling of these systems, and provides a template for when ESNs would provide a good fit for a chosen problem, and when other models should be considered. 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.

3.2 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. It is worth future efforts comparing and contrasting use of these forms of training and their effects on model feedback in cases with extremely complex signals. It is also worth exploring the use of ESNs in predicting reaction patterns of dissolved oxygen 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 [16]. ESNs could provide key insights into this problem in areas where remote sensing and monitoring are essential to measuring watershed health.

BIBLIOGRAPHY

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

[2] Graves A., Mohamed A., and Hinton G. Speech recognition with deep recurrent neural networks. In IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2013.

[3] Schneider C., Laizé C., Acreman M. C., and Flörke M. How will climate change modify river flow regimes in europe? Hydrology and Earth System Sciences, 17:325–339, 2013.

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

[5] Bunn S. E. and Arthington A. H. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management, 30:492–507, 2002.

[6] ElMoaqet H. A typical architecture of a long short-term memory (lstm) cell, 2020. Online; accessed September 14, 2022.

[7] Jaeger H. Short term memory in echo state networks. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems, 2002.

[8] Jaeger H. The basic schema of an echo state network, 2007. Online; accessed September 14, 2022.

[9] Kwon H. and Dai Q. Remote sensing for streamflow simulation. Remote Sensing, 2020.

[10] Pörtner H.-O. and Roberts D. C. et al. Climate change 2022: Impacts, adaptation and vulnerability working group ii contribution to the sixth assessment report of the intergovernmental panel on climate change. 2022.

[11] De Vos N. J. Echo state networks as an alternative to traditional artificial neural networks in rainfall-runoff modelling. Hydrol. Earth Syst. Sci., 17:253–267, 2013.

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

[13] Doya K. Bifurcations in the learning of neural networks. In IEEE International Symposium on Circuits and Systems. IEEE, 1992.

[14] Bianchi F. M., Scardapane S., Løkse S., and Jenssen R. Reservoir computing approaches for representation and classification of multivariate time series. CoRR, abs/1803.07870, 2018.

[15] Lukoševičius M. A practical guide to applying echo state networks. In Neural Networks: Tricks of the Trade: Second Edition. Springer Berlin, Heidelberg, 2012.

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

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

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

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

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

[21] Hunt K. M. R., Matthews G. R., Pappenberger F., and Prudhomme C. 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, 2022.

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

[23] Water Science School. Dissolved oxygen and water. 2018.

[24] Cao W., Zhang Z., Liu Y., Bend L., Weng S., and Xu H. Seasonal differences in future climate and streamflow variation in a watershed of northern china. Journal of Hydrology: Regional Studies, 38, 2021.

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

APPENDIX A

Additional Facts and Figures

About appendices: This appendix contains additional facts and figures regarding both streamflow monitoring and ESNs

(1) Reservoir size plays a vital part in determining model fit.

```
Reservoir Sizes: 100, 1000, 10000
RMSE: [4.7494515413933325, 2.3203492661214824, 3.292617636918171]
R^2: [0.6427957827915836, 0.9143668801020041, 0.8275548777113881]
NSE: [0.641178472729699, 0.9143557341663798, 0.8275456439187848]
```

Figure A.1: Reservoir Size effects on model accuracy

(2) Ridge regularization values have less impact than reservoir size, but do influence network sensitivity to long-term trends and minute patterns.

```
Ridge Values: 1e-3, 1e-7, 1e-9
Temperature RMSE's:  [2.0357654890487225, 2.105800785519578, 2.146845084073016]
Temperature R^2:  [0.9324607407806352, 0.9277133161693457, 0.9248975212395806]
Temperature NSE's:  [0.9324321190162821, 0.9277031508511248, 0.9248574000896701]
```

Figure A.2: Ridge Regularization Parameters

(3) The USGS Water Science School contains helpful information on key metrics for measuring watershed health. The home page can be found at: Water Science School Home Page

(4) This project made use of a python reservoir library called reservoirpy, that makes Echo State Network initialization, and training very easy, with high levels of customization. Users can experiment with various on- and off-line training methods, hyperparameter optimization tools, and simple graphical interfaces to make training and testing intuitive for users. Their home page and documentation can be found at: ReservoirPy