

Article

Application of Neural Networks for Hydrologic Process Understanding at a Midwestern Watershed

Annushka Aliev ^{1,2} , Sinan Rasiya Koya ², Incheol Kim ², Jongwan Eun ², Elbert Traylor ³ and Tirthankar Roy ^{2,*} 

¹ Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA

² Department of Civil and Environmental Engineering, University of Nebraska, Lincoln, NE 68588, USA

³ Nebraska Department of Environment and Energy, Lincoln, NE 68521, USA

* Correspondence: roy@unl.edu

Abstract: The Shell Creek Watershed (SCW) is a rural watershed in Nebraska with a history of chronic flooding. Beginning in 2005, a variety of conservation practices have been employed in the watershed. Those practices have since been credited with attenuating flood severity and improving water quality in SCW. This study investigated the impacts of 13 different controlling factors on flooding at SCW by using an artificial neural network (ANN)-based rainfall-runoff model. Additionally, flood frequency analysis and drought severity analysis were conducted. Special emphasis was placed on understanding how flood trends change in light of conservation practices to determine whether any relation exists between the conservation practices and flood peak attenuation, as the strategic conservation plan implemented in the watershed provides a unique opportunity to examine the potential impacts of conservation practices on the watershed. The ANN model developed in this study showed satisfactory discharge–prediction performance, with a Kling–Gupta Efficiency (KGE) value of 0.57. It was found that no individual controlling variable used in this study was a significantly better predictor of flooding in SCW, and therefore all 13 variables were used as inputs, which resulted in the satisfactory ANN model discharge–prediction performance. Furthermore, it was observed that after conservation planning was implemented in SCW, the magnitude of anomalous peak flows increased, while the magnitude of annual peak flows decreased. However, more comprehensive assessment is necessary to identify the relative impacts of conservation practices on flooding in the basin.

Keywords: rainfall-runoff modeling; watershed hydrology; artificial neural network; conservation practices; flood frequency analysis; drought analysis



Citation: Aliev, A.; Koya, S.R.; Kim, I.; Eun, J.; Traylor, E.; Roy, T.

Application of Neural Networks for Hydrologic Process Understanding at a Midwestern Watershed. *Hydrology*

2023, 10, 27. <https://doi.org/10.3390/hydrology10020027>

Received: 10 November 2022

Revised: 7 January 2023

Accepted: 8 January 2023

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Flooding is a significant and recurring issue in Nebraska. In the past 100 years alone, five historic floods have been recorded [1]. The most recent of those historic floods was the March 2019 flood, during which several new records were set for river crests, snowfall prior to the event, and precipitable water values [2]. The flood caused extensive damage to infrastructure and personal property, including the destruction of bridges and roads, the breaching of 50 levees, and more. Five lives were lost, and thousands more were forced to evacuate their homes and businesses [1]. By August of 2019, flood damage cost estimates had reached \$3 billion [2].

In addition to the widespread infrastructure damage inflicted, Nebraska's agriculture industry also suffered heavy repercussions from the March 2019 flood. Cows and calves died or were stranded, crop fields were flooded or left covered in chunks of river ice, and grain stores were contaminated, resulting in approximately \$400 million in cattle losses and \$440 million in crop losses [3]. Destruction of roads due to the floods posed further challenges for farmers beyond immediate losses, as many farms suddenly became inaccessible or required detours of extra tens of miles to reach in order to tend to livestock. According

to an estimate made by the President of the Nebraska Farm Bureau, the additional transportation costs of these detours, as well as related costs of fuel and feed, were costing the Nebraska cattle industry roughly \$1 million a day at the time [4]. The flooding of March 2019 was a devastating example of the serious damage to crops and livestock that floods can inflict on farmers and rural areas.

The March 2019 flood occurred as a result of the culmination of several meteorological and hydrological events. The warm and wet start to winter resulted in unfrozen ground that was heavily saturated with moisture, which then froze following a shift to colder temperatures in late January; early winter conditions were approximately 1–2 °C warmer than average, while late January conditions ranged from 3–5 °C colder than average and resulted in frost depths of 60–90 cm. Record-breaking snowfall followed, developing a significant snowpack with a snow-water equivalent of 3–10 cm, and rivers froze, creating the potential for ice jams. When warmer temperatures returned coincident with a cyclone that brought 25–50 mm and 40–75 mm of rain across northeastern and central Nebraska, respectively, the excessive runoff generated from the combination of snowmelt, precipitation, and rain-on-snow events could not percolate into the frozen, saturated ground, thus overwhelming rivers and creating significant flooding [2].

The mechanism by which the March 2019 flood was produced, along with its repercussions, highlights the importance of understanding the interactions of antecedent hydrological and meteorological factors in creating flooding. Increased understanding of conditions that contribute to flooding can help with forecasting and decision-making [5]. As seen during the March 2019 flood, monitoring and modeling of several hydrological and meteorological inputs allowed for forecasts and warnings to be put out in advance of the flood event, which likely saved many lives and personal property [2]. Further improvements to flood models and additions to the understanding of flood development can aid in yielding more accurate and timely predictions for flooding.

This study aimed to contribute to the improvement of flood forecasting and hydrologic process understanding in a specific rural watershed in Nebraska, Shell Creek Watershed (SCW). Furthermore, preliminary insights were drawn from the results of this study regarding the effects of conservation practices on attenuating flooding in the watershed. SCW provides a unique example of a watershed that has had a comprehensive and strategic conservation plan aimed at addressing flooding and water quality issues, and this plan has been in place for a long enough span of time that observations can reasonably be drawn regarding the relationship between conservation practices and flood attenuation. This study opens the door to such discussion, laying groundwork and preliminary observations for future researchers to build upon when analyzing the impact of SCW's conservation plan on flooding.

SCW is a major tributary of the Platte River located in east-central Nebraska. The watershed is largely agrarian, with 93% of its area designated as agricultural land, and has a history of chronic flooding [6]. In the past 80 years, SCW experienced 17 river crests above flood stage, with an additional seven above action stage [7]. As previously exemplified, flooding can cause devastating damage and setbacks to Nebraskan farms through crop and cattle losses, making the frequent flooding of SCW an important matter of concern. In the early 1990s, some farmers began implementing conservation practices on their land in efforts to reduce flood risk on a small scale. Later, in 1999, the Shell Creek Watershed Improvement Group (SCWIG) organized to address flooding in SCW, then shifted their focus towards water quality impairments in the watershed and drove the development and implementation of the 2005 Shell Creek Watershed Management Plan. The plan was fully implemented in 2015; however, certain water quality impairments and other issues remained unresolved, warranting the development of the 2016 Shell Creek Watershed Environmental Enhancement Plan to improve soil health, reduce runoff, improve water quality, and address stream conditions impacted by watershed degradation. The 2016 plan is to be implemented in several phases extending to 2032 and beyond, and efforts will require a more detailed understanding of the hydrology of SCW [6]. The existing restoration

efforts and conservation practices implemented over the course of these management plans have been credited anecdotally with somewhat reducing flooding, an observation that warrants exploration and discussion and, as yet, lacks validation. This paper's research on flood modeling of SCW will aid in addressing the demand for further understanding of SCW hydrological processes, investigate potential changes in streamflow patterns, and support the improvement of forecasting for the watershed's frequent floods. Furthermore, analysis of the impact conservation practices have had on flood attenuation in the watershed will aid managers in making decisions regarding future conservation implementations.

To model flooding in SCW, a feedforward artificial neural network (ANN) was developed, trained, and tested in MATLAB. ANNs are useful for their ability to determine relationships between given inputs and recorded outputs of a process without explicit physical process information [8–10] and have been successfully employed in several hydrological applications [11–15].

This study designed and trained a single-hidden-layer feedforward ANN to model the rainfall-runoff process in SCW using varying combinations of potential flooding factors. Building a rainfall-runoff model allows for a deeper understanding of the hydrological processes, specifically flooding, in SCW, and aid in the accomplishment of the four main objectives of this study: (1) characterizing the efficacy of the selected controlling factors in SCW flood prediction, (2) building a rainfall-runoff model for the watershed using ANN, (3) assessing drought intensity and flood frequency in the watershed, and (4) drawing preliminary insights on the effects of conservation practices on flooding in the watershed.

2. Materials and Methods

2.1. Study Area

Shell Creek Watershed is a rural Nebraskan watershed that serves as a major tributary of the Lower Platte River, which eventually feeds into the Missouri River. Approximately 193 km long and spanning 123,391 hectares of land, the watershed runs through Antelope County, Boone County, Madison County, Platte County, and Colfax County. The cropland map of SCW is shown in Figure 1. Land use in SCW is predominantly agricultural, with 93% of the watershed area dedicated to farmland. The main agricultural crops of the watershed by land-use area are corn (48%) and soy (28%), and other agricultural products include swine, cattle, and alfalfa. Developed land only takes up about 4% of SCW. Land-use immediately adjacent to the channel (305 m to either side) is 73% crop cultivation and 11% grassland and grazing, with the rest populated by forest, wetlands, or development. In the upper half of SCW, cropland is often farmed up to the edge of the channel. In the lower half, the edges of the channel are typically lined with a narrow forest buffer [6].

The meteorological patterns of SCW are seasonal, and thus average temperature, precipitation, and streamflow in the region vary by season. Summers are warm, with an average temperature of 23.9 °C. Precipitation falls as rain in the form of showers and storms during summer, reaching a cumulative 50.8 cm from April to September out of the total annual 66 cm. The average summer surface water flow is 2.04 m³/s. During winter, temperatures average −4.4 °C, and precipitation is mainly in the form of snow. Average snowfall is about 63.5 cm, and average surface water flow is 0.71 m³/s. The average mid-afternoon relative humidity in SCW is 60% [6].

Like most of eastern Nebraska, SCW is characterized by rolling hills of easily erodible soils [6]. The majority of soils found in SCW are classified in the B Hydrologic Soil Group: moderately deep to deep soils that are between 10% and 20% clay and 50% to 90% sand in composition. These soils have a moderate infiltration rate when thoroughly wet [16]. SCW also contains a sizeable amount of loess [17], silt blankets of which eastern Nebraska has some of the thickest in the Midwest. Loess is highly erodible when wet and thus contributes significantly to the channel stability problems observed in SCW. There are three main types of bedrock in SCW, which divide the watershed into thirds; the western third is composed of mudstone and sandstone, the middle third of limestone, and the lower third of shale. Bedrock rarely impacts stream processes in SCW.

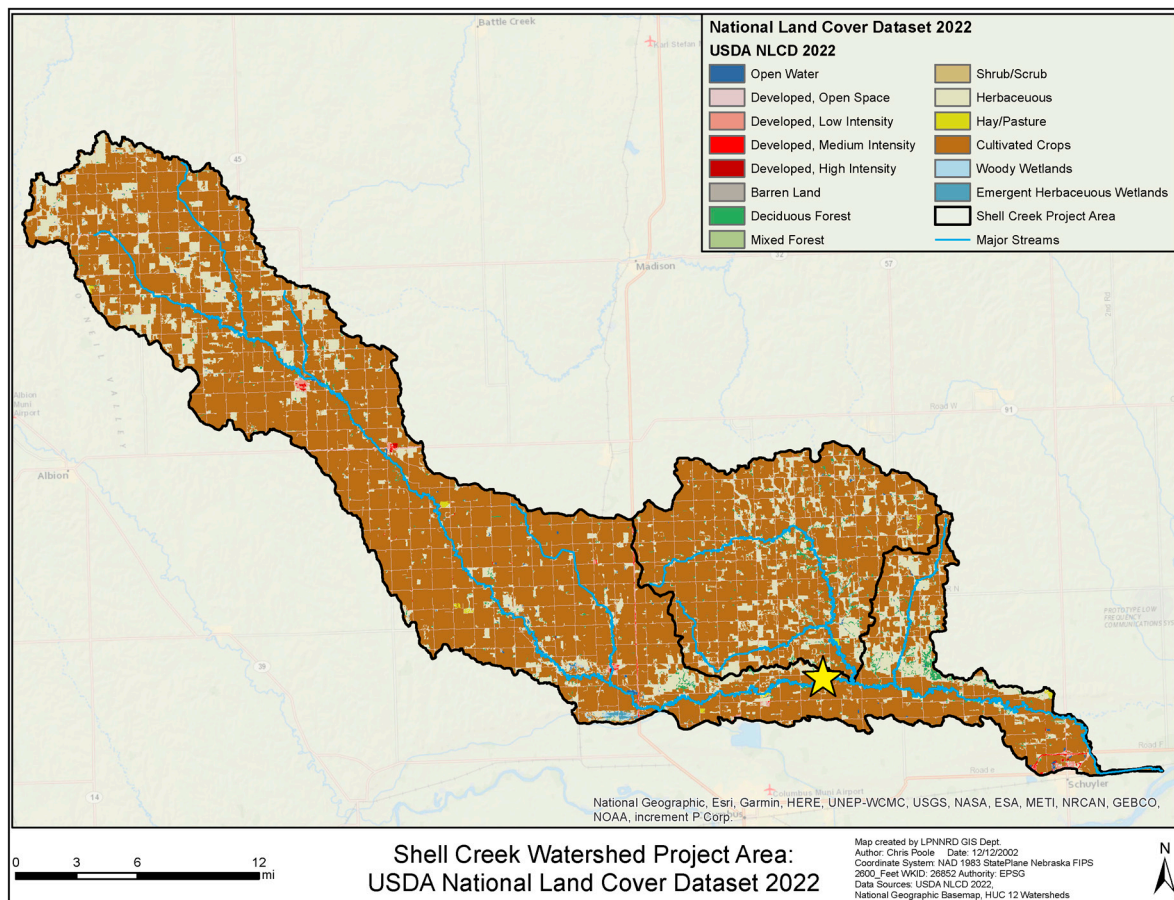


Figure 1. Shell Creek Watershed with USDA 2022 land cover data. The location of the streamflow gauge used for this study is indicated by a yellow star.

The designated beneficial uses of Shell Creek are recreation, aquatic life, agricultural water supply, and aesthetics. However, a history of agricultural use and anti-conservationist attitudes and practices left Shell creek impaired in the areas of recreation and aquatic life by Atrazine, selenium, and *Escherichia coli* (*E. coli*) [18]. Furthermore, historic anthropogenic hydrological and environmental modification in the way of clearing large tracts of land for cultivation has resulted in an increased rate and volume of storm flow in SCW, increasing the risk and frequency of flooding and accelerating erosion. Since 1999, SCWIG has been working to address these issues systematically on a watershed scale. The result of their efforts was the development of the 2005 Shell Creek Watershed Management Plan, which focused on both the quantity and quality of runoff to resolve flooding and water quality issues in Shell Creek. The 2016 Shell Creek Watershed Environmental Enhancement Plan was subsequently developed to address persisting water quality impairments not resolved by the 2005 plan. Through these plans and the work of SCWIG and SCW landowners, more than 340 conservation practices have been implemented on the land. These practices include no-till farming, cover crops, and filter and buffer strips and have resulted in the successful delisting of Shell Creek for aquatic life impairment due to Atrazine in 2018 [19]. Outside of improvements to water quality, these conservation practices have additionally been credited with alleviating flooding somewhat in SCW [20].

2.2. Data

The majority of hydrological and meteorological data for this study were obtained from Phase 2 of the North American Land Data Assimilation System [21,22]. The available dataset includes precipitation totals (PCP), above-ground convective available potential energy (CAPE), the fraction of total precipitation that is convective (CNFRAC), longwave

radiation flux downwards (DLRWF), shortwave radiation flux downwards (DSWRF), potential evaporation (PEVAP), surface pressure (S.P.), specific humidity (S.H.), temperature, zonal wind speed (UGRD), and meridional wind speed (VGRD), and starts from January 1979. Hourly data for these variables were masked to the study region, then aggregated to daily averages. Daily snow water equivalent (SWE) data were pulled from the National Snow and Ice Data Center [23], and daily groundwater level (GWL) data were obtained from the National Water Information System (NWIS). A large number of initial variables was selected intentionally, as it allows for the narrowing down of variables to determine what factors are relevant for flooding in SCW—one of the main objectives of this study.

Discharge data were obtained from the U.S. Geological Survey (USGS) from a stream-flow gauge near Columbus, NE. The Columbus gauge is the only gauge for Shell Creek that provides daily discharge data to date, and as discharge is the target variable in our rainfall-runoff model, the study region and time period for this research are restricted by the location and available data of this gauge. The Columbus gauge is located upstream from the point where SCW flows into the Lower Platte River and just slightly upstream from where Loseke Creek feeds into Shell Creek, so this study will focus on the portion of the watershed above the gauge and exclude the downstream area beyond Columbus. Approved daily discharge data from the Columbus gauge only extend as far back as 1990, and thus the total time interval used for variable selection and model training in this work is from 1990 to 2020.

While daily discharge data from the Columbus gauge are only available from 1990, annual peak flow data from the same gauge are documented and available all the way back to 1947, with the exception of 1976 and 1977. These annual peak flow data from the USGS are used for flood frequency analysis of the watershed. Drought analysis was conducted using NLDAS precipitation data from 1982 to 2020.

2.3. Model: Artificial Neural Network

Artificial neural networks are machine learning models loosely based on the structure of biological networks in the brain. These networks are composed of artificial neurons organized into distinct layers—the input layer, hidden layers, and output layer—each of which contains at least one neuron. Data inputs are run from the input layer through the hidden layers to the output layer, where a prediction is output. Neurons in the hidden layers and output layer take the outputs of other neurons as inputs, then compute nonlinear transformations of those inputs to generate their own outputs. In recurrent neural networks, a neuron may take the outputs of other neurons from the previous layer and from within the same layer. In feedforward neural networks (FFNNs), neurons only receive outputs as inputs from the previous layer.

In an ANN, a connection between two neurons has a weight associated with it that represents the connection strength. Changing these weights of an ANN changes the final output of the model, and thus it is necessary for such weights to be adjusted to optimize the performance of the model. This is done through training. Two widely used categories of ANN training are supervised and unsupervised training (note that there are other types of training as well, such as semi-supervised or self-supervised training, which are not discussed here). Supervised training is a method in which an established pair of inputs and outputs is compared against the model's outputs for the same established data inputs, and then feedback is given in order to minimize the deviation of the model outputs from the expected outputs. Unsupervised learning is a method in which unlabeled training data (data without a target variable) is given to an algorithm, from which the algorithm then identifies patterns and categories on its own. This study employed supervised learning to train the ANN model.

A common method used for adjusting network weights in supervised training is error backpropagation. Backpropagation compares the network output for a set of inputs with the observed target, then evaluates the error with a loss function. This error is then propagated backward to adjust connection weights and improve the accuracy of the model.

This study used a FFNN with a single hidden layer to model rainfall-runoff processes in SCW. A single-layer model was chosen for this research as preliminary testing of a single dataset on both a single-layer and double-layer model yielded nearly identical results, making the added complexity of the double-layer ANN redundant.

ANN codes were written in MATLAB, and all datasets were normalized using z-score normalization prior to input to the ANN according to the following expression:

$$z = \frac{x - x_{\text{mean}}}{\sigma}$$

where x is the data point, x_{mean} is the mean value of the data, and σ is the standard deviation of the data. Once data had been normalized, they were used in the training of an ANN whose number of neurons in the hidden layer was varied from 1 to 50. The model was trained on 70% of the dataset, then validated on the remaining 30% of the data. Mean squared error (MSE) from validation for the output of each model was compared, and a number of neurons between 10 and 50 were then selected for the prediction ANN model by determining which ANN had the lowest MSE after training and validation. These bounding values of 10 and 50 neurons were chosen in order to ensure the selection of a model that was neither oversimplified nor overly complex.

Following training, the selected ANN model was made to predict discharge, and the performance of the model was measured using the Kling–Gupta efficiency (KGE) [24]. KGE is a statistic that compares the bias, variability, and timing of a model's output to that of the observed data, and is calculated as follows:

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

where r is the linear correlation coefficient, α is a measure of relative variability in the simulated and observed values, and β represents bias. KGE values range from negative infinity to 1, with 1 indicating that a model's outputs perfectly match with the observed target data. The closer a KGE value is to 1, the better the model is performing.

The ANN was trained and run numerous times, and on data for three different time periods: the full period, 1990–2020; the pre-planning period, 1990–2004; and the post-planning period, 2005–2020. In this paper, the pre-planning period is sometimes noted as the non-conservation period, and the post-planning period is sometimes noted as the conservation period. Discharge was the target variable for all trials. The first trial for each period included all predictive variables, and its KGE was recorded as a reference KGE for future trials. Leave-one-out analysis was then conducted where the model for each period was run multiple times, each with a different variable left out, and the resulting KGE of each trial was compared to the reference KGE for that time period in order to determine the respective influence of each variable on the model's output accuracy and performance. The least influential variables were eliminated, and the process was repeated until the most effective model had been obtained.

2.4. Flood Frequency Analysis

Flood frequency analysis is a method employed in hydrology to estimate the exceedance probabilities corresponding to specific streamflow values for a given river. Annual peak flow or peak-over-threshold data existing over a sizeable number of consecutive years (typically more than 30) were collected and used to fit probability distribution functions from which the exceedance probabilities can be calculated. In this study, flood frequency analysis was conducted using the U.S. Army Corps of Engineers Hydrologic Engineering Center's (HEC) Statistical Software Package (HEC-SSP 2.2). Flood frequency analysis for SCW was also broken into three different time periods to allow for the examination and identification of flood frequency differences before and after the implementation of conservation practices. These periods are the full period, 1947–2020, the pre-planning period, 1947–2004, and the post-planning period, 2005–2020. However, only results from

the full period and pre-planning period were analyzed in this paper, as the post-planning period extends only 15 years back, so it would not be a reliable source to draw conclusions from regarding flood frequencies. Conclusions about flood frequency changes during the post-planning period were instead drawn from the comparison of the full period and the pre-planning period. It is to be noted that there always has been some level of conservation in the watershed, but the key to success in the Shell Creek watershed was the development and implementation of the comprehensive and strategic plan to address the flooding and water quality issues as opposed to random acts of conservation.

2.5. Drought Analysis

Droughts are often measured using drought indexes, which analyze data for selected drought indicators over various time intervals in order to output a drought index value. This drought index value is a single number interpreted on a range from abnormally wet to abnormally dry. In order to conduct a drought analysis for SCW, this study employed the Standard Precipitation Index (SPI) [25], which compares actual precipitation accumulation over a region for a certain time period to the probability of precipitation according to historical records for that same time period. SPI values indicate the number of standard deviations from mean moisture conditions, with positive SPI values representing wet conditions and negative SPI values representing dry conditions. SPI values ranging in magnitude from 0 to 0.99 indicate mild conditions, from 1 to 1.49 indicate moderate conditions, from 1.5 to 1.99 indicate severe conditions, and 2 or above indicate extreme conditions [25]. In this study, SPI was run for 1-month, 3-month, 6-month, and 12-month time intervals. Inferences about soil moisture levels in an area can be drawn from the 3-month SPI, while SPI values corresponding to longer timescales (e.g., 6- to 12-month) relate information about wet and dry periods.

2.6. Analysis Framework

Figure 2 shows the analysis framework followed in this study. In order to determine variables relevant to flooding, both model-free and model-based elimination were employed. Once data were masked to study area and aggregated to daily data, pair plots and cross-correlation plots were created that plotted each variable individually against the target variable discharge, as well as all variables against each other. These plots were generated for the three time intervals previously mentioned: the full time period, 1990–2020; the pre-planning period, 1990–2004; and the post-planning period, 2005–2020. This step allowed for the examination of relationships between each variable and the target variable and the identification of multicollinearity between independent variables. The variables were then run through the ANN, which first underwent training and validation with the data, then generated predictions for the target variable discharge. These predictions were evaluated using KGE, and using those evaluations, the variables most relevant to flooding were determined. This identification of factors and conditions conducive to flooding allowed for an improved understanding of SCW hydrological processes.

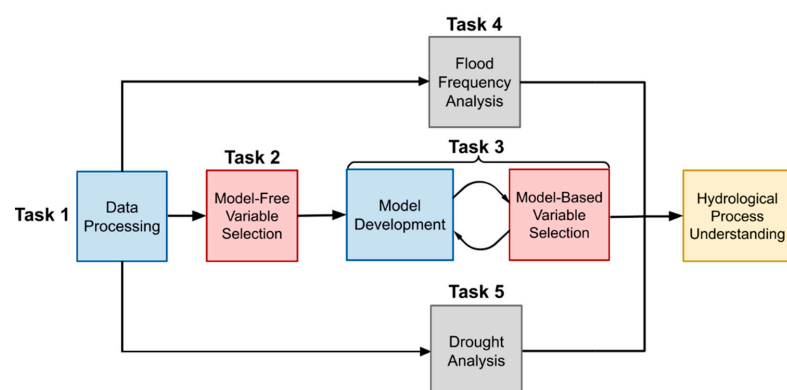


Figure 2. Flowchart depicting the analysis framework for this study.

Additional pathways towards improved hydrologic understanding taken on in this research are flood frequency analysis and drought analysis. Using USGS annual peak flow data, flood frequency analysis was carried out to identify the exceedance probabilities of streamflow in SCW. Using long-term precipitation data, drought analysis was conducted with the SPI in order to better understand patterns of dryness and wetness in SCW.

3. Results

3.1. Model-Free Variable Selection

Figure 3 shows a pair plot of the cross-correlation between the selected predictive variables for the full time period of study, 1990–2020. From the figure, it can be seen that temperature shows a relatively high correlation with specific humidity, potential evaporation, and longwave radiation flux downwards. Specific humidity additionally showed a good correlation with above-ground convective available potential energy and longwave radiation flux downwards. Shortwave radiation flux downwards displayed a fairly good correlation with potential evaporation. These correlation patterns between variables held true not only for the full time period but for the pre-planning and post-planning periods as well (see Appendix A for cross-correlation pair plots for each period). Similar patterns of occurrence between variables can indicate similar potential efficiency in flood prediction, making some of these variables candidates for removal in order to reduce redundancies before inputting data into the ANN. However, none of the predictive variables showed a good correlation with the target variable, discharge, as evidenced in Figure 4. The correlation coefficient (r) is a measure of the relationship between two variables and exists on a scale from -1 to 1 ; 1 represents perfect positive correlation, -1 represents perfect negative correlation, and 0 indicates no correlation. All correlation coefficients found for the SCW data between the predictive variables and target variable were within ± 0.2 of zero, indicating a very poor correlation. For both the full period and the post-planning period, precipitation was the most closely correlated variable to discharge, with r values of only 0.16 and 0.18 , respectively. The most highly correlated variable in the pre-planning period was the fraction of total precipitation that is convective, with an r value of 0.14 . Due to the low across-the-board low correlations, it was decided that all variables would be carried over to the ANN phase of variable selection instead of eliminating variables with low target correlations or selected variables with close cross-correlations.

3.2. Model-Based Variable Selection

The ANN was initially trained, validated, and run for all three time periods with all variables included as inputs in order to determine reference KGE values for model performance. The resulting reference KGEs for the total time period, pre-planning period, and post-planning period were 0.3624 , 0.1740 , and 0.5666 , respectively. The components of the KGE score—correlation (r), bias ratio (β), and variability ratio (α)—for the total time period were 0.5103 , 0.9037 , and 0.6033 , respectively. For the pre-planning period, r was 0.4397 , β was 0.8876 , and α was 0.4036 . For the post-planning period, r , β , and α were 0.6891 , 0.7573 , and 0.8204 , respectively.

Following the determination of the reference KGEs, the leave-one-out process was then conducted with the model for each time period, and the percent change in model performance from the reference KGE was recorded for each trial. Figure 5 displays these percentage changes. In both the full period and the post-planning period, there were no instances of variable removal improving model performance. However, the pre-planning period showed four instances of model performance improving with the removal of certain variables. For the full period, the three greatest performance drops resulted from the removal of snow water equivalent (-104.64%), shortwave radiation flux downwards (-102.98%), and specific humidity (-97.05%). The post-planning period saw its three greatest model performance drops from the removal of shortwave radiation flux downwards (-102.54%), precipitation (-97.49%), and specific humidity (-91.86%). The pre-planning period saw model performance improvement with the removal of temperature, poten-

tial evaporation, convective available potential energy, and meridional wind speed, the highest of which was potential evaporation (34.14%). The largest performance drop in the pre-planning period was due to the removal of groundwater level (−96.72%).

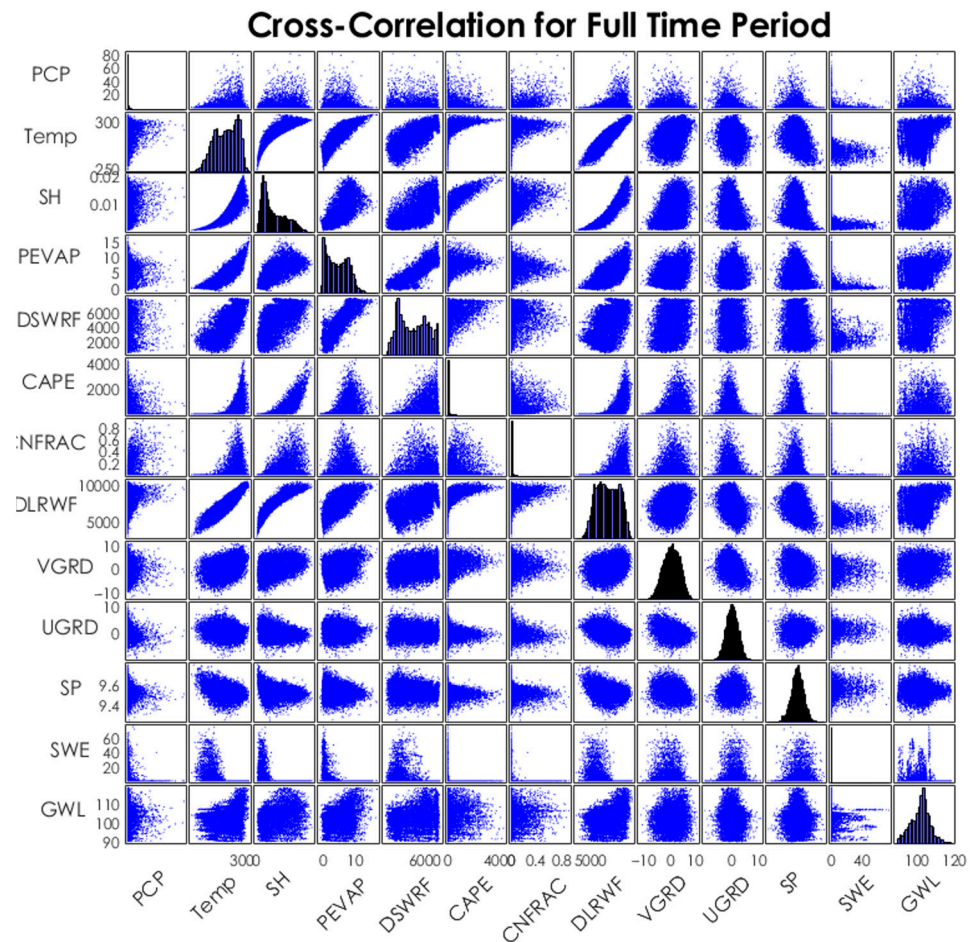


Figure 3. Cross-correlation pair plot for all 13 predictive variables over the full time period.

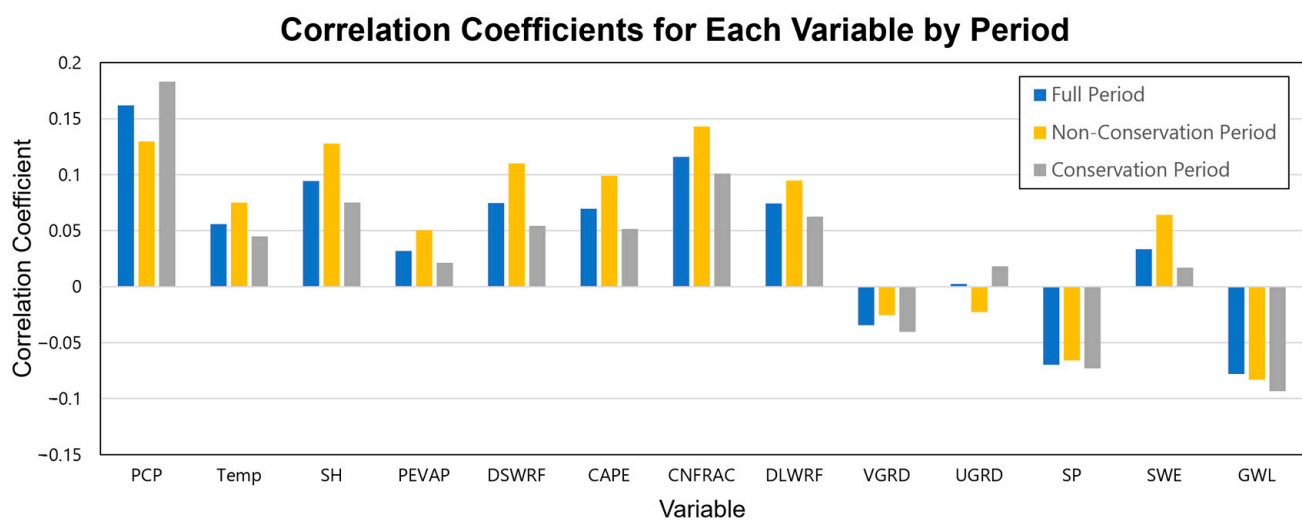


Figure 4. Correlation coefficients between each variable and the target variable by time period.

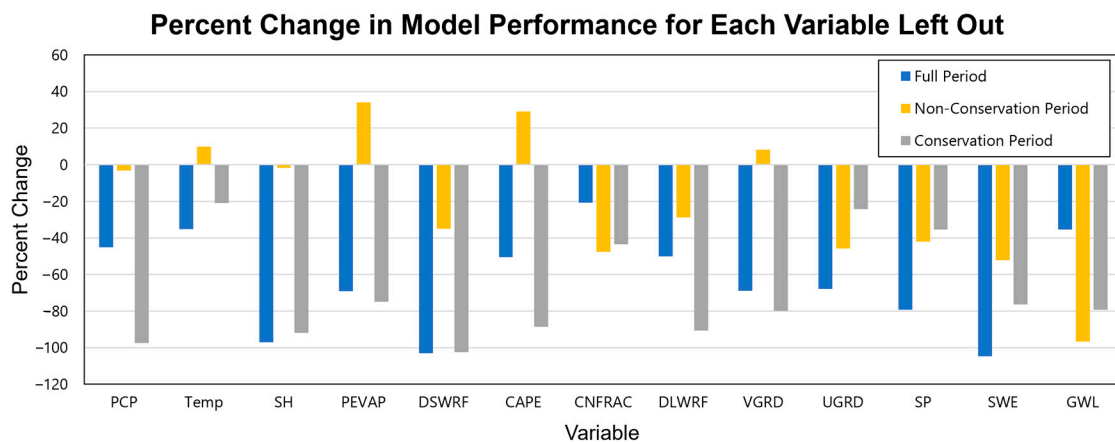


Figure 5. Percent change in model performance by time period due to each variable left out.

3.3. Flood Frequency Analysis

Computed streamflow and their corresponding return periods for both the full period and pre-planning period are shown in Figure 6. For each return period, computed curve flow is consistently higher for the full period flood frequency analysis than the pre-planning period flood frequency analysis, with the exception of the 2-year return period.

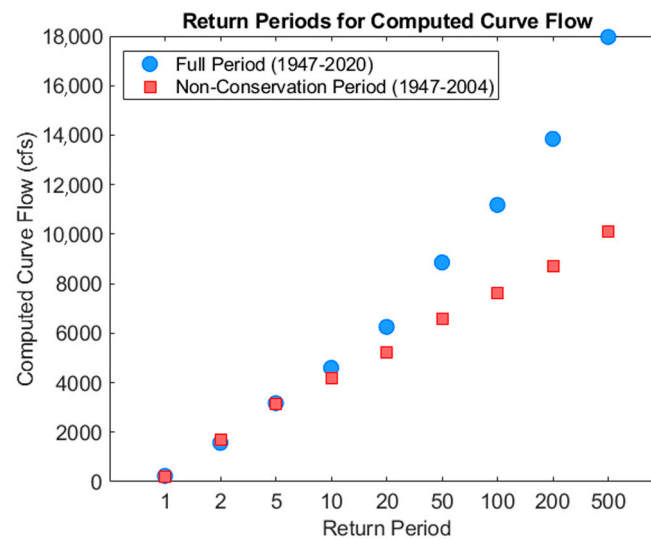


Figure 6. Return periods for the computed curve flows of the full period and the non-conservation period.

Flood frequency analysis for the post-planning period was not included in this analysis as the post-planning period only extends 15 years, which is too short of a time span to produce reliable flood frequency results. To see flood frequency analysis results for the post-planning period and two halves of the pre-planning period, as well as tabulated results for the flood frequency analysis, see Appendix B.

3.4. Drought Analysis

Figure 7 shows the 1-month, 3-month, 6-month, and 12-month SPI values for SCW plotted over time from 1982–2020 (See Appendix C for each SPI timescale plotted individually). From the 12-month SPI, moderate wet periods can be seen during 1984, 1987, 1994, 1995, 2010, 2011, 2016, and 2019. Severe wet periods were observed in 1983, 1993, 1999, 2007, and 2008 (which reached extreme levels mid-year). Moderate dry periods can be seen during 1991, 1997, 2002, and 2014, severe dry periods occurred during 1988, 2000, and 2020, and extremely dry periods took place in 1989, 1990, and from the end of 2012 to the beginning

of 2013. From the 3-month SPI, one can speculate some patterns of dryer soil moisture conditions in winter, and wetter conditions in spring, summer, and fall. However, summer soil moisture occasionally drops, so spring and fall appear to show more consistently moist conditions. Wet and dry periods indicated by the drought analysis were used for deeper understanding and discussion of individual variable prediction efficacy in Section 4.1.

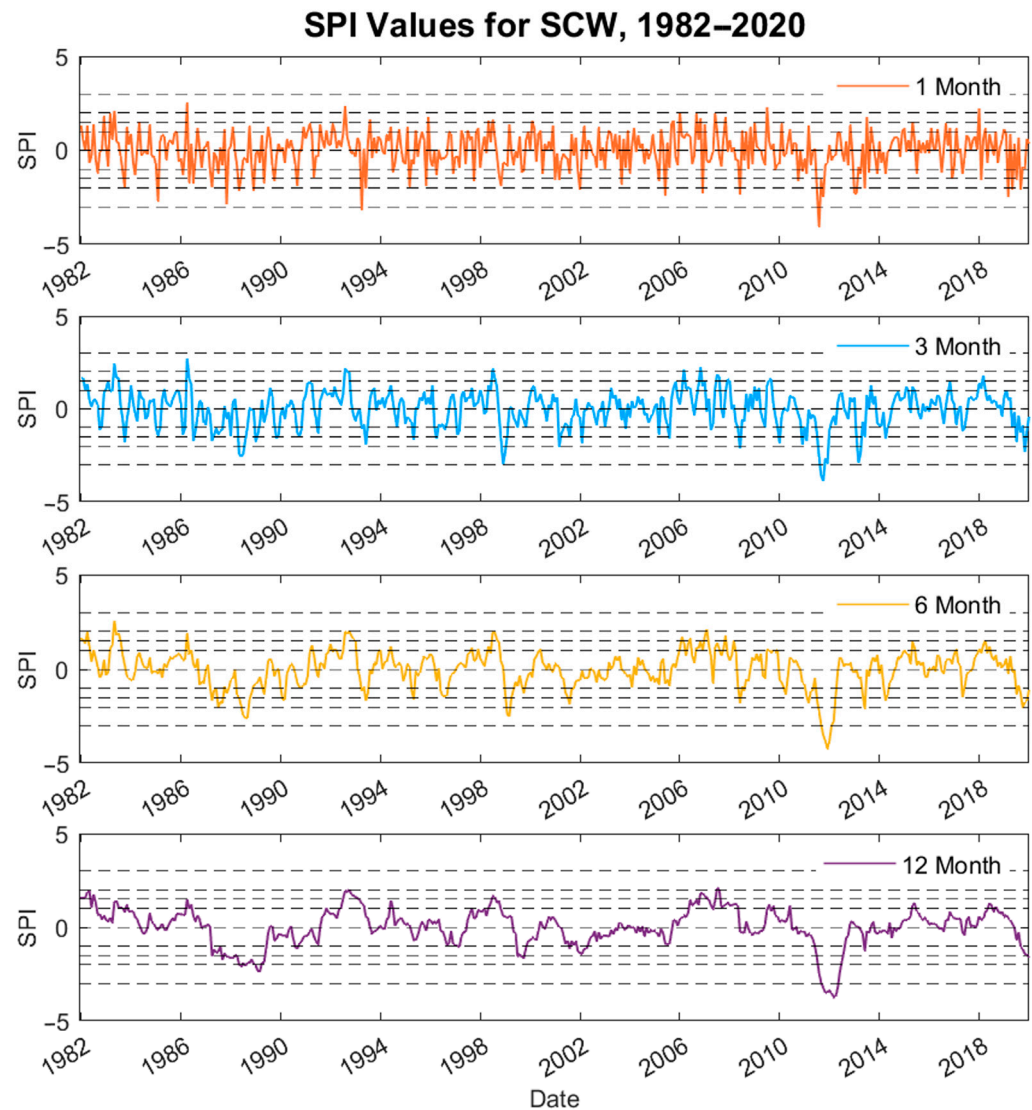


Figure 7. 1-month, 3-month, 6-month, and 12-month SPI values for SCW plotted over time from 1982–2020.

4. Discussion

4.1. Individual Variable Prediction Efficacy

Individual variable prediction efficacy was primarily determined through analysis of the correlation coefficients for each variable found in Section 3.2 of Results. The low across-the-board correlation between predictive variables and the target variable, discharge, suggests that no one factor (of those included in this study) is a good indicator of flooding. However, if one were to select for the highest individual variable prediction efficacy, precipitation is the variable most tied to discharge in SCW. Precipitation saw the highest correlation with discharge for the full period and post-planning period and was the second most correlated variable for the pre-planning period after the fraction of precipitation that is convective, another precipitation-related statistic.

The precipitation-to-discharge relationship is also supported by the SPI results found in the drought analysis. For example, from the 12-month SPI, one can see multiple instances of marked wet periods coinciding with past floods. Major examples include 1982, 1993, and 2008: 1982 had peak SPI values above 1.5 (well within the wet range) and 2 large floods; 1993 had major flooding and showed a sharp increase in SPI (and therefore precipitation) mid-year that reached into the moderately wet range ($SPI > 2$); 2008 had an SPI value that reached the moderately wet range, and had an annual peak flow that is the highest of those on record. However, there were also some cases where flooding coincided with dry or average moisture periods, so the SPI results were fairly consistent with the low precipitation–discharge relationship observed in the model-free variable selection.

One variable not considered in this study was soil moisture, however results of the drought analysis suggest that this variable be taken into account for flood prediction in future studies. The majority of flooding in SCW takes place during spring and summer [7], and according to results from the 3-month SPI, spring and summer in SCW typically have higher soil moisture levels than winter. This suggests a potential correlation between discharge and soil moisture, though that relationship should be validated with data and cannot be confirmed from the results of this research.

4.2. Model Prediction

The across-the-board low variable-target correlations found in this study could have the implication that a combination of factors is the more reliable route for discharge prediction. Indeed, from the ANN leave-one-out analysis results detailed in Section 3.2, it can be seen that model prediction performance is better when all 13 predictive variables are included as inputs (with the exception of the pre-planning period), as shown in Figure 6. Since the full-period and post-planning period models showed performance drops solely with the removal of variables, their final and most successful versions included all input variables. For the full period, the best model included 32 hidden neurons and had a KGE value of 0.36. The best model for the post-planning period had 45 neurons in the hidden layer and a KGE of 0.57. For the pre-planning period, performance increased slightly with the individual removal of certain variables; however, the removal of combinations of multiple variables again resulted in decreased performance. Thus, the pre-planning period model with the highest performance was simply the one that excluded potential evaporation—the variable that caused the greatest performance increase when left out—and included all other input variables. This pre-planning model had 11 hidden neurons and a KGE of 0.23.

The KGE values for these models indicate functional prediction performance of varying degrees of acceptability, as all final model KGEs were positive. Of these final proposed models, the post-planning period model showed the best relative performance as its KGE value of 0.57 was closest to 1. The low variable-target correlations of the input variables used in the models justify acceptance of this 0.57 KGE value for the post-planning period as satisfactory performance. To increase the robustness of the model, more data would be needed for both before and after the initiation of conservation practices in the watershed.

4.3. Changes in Flood Intensity

Comparison between the results of the flood frequency analysis of the full period and the pre-planning period (from Section 3.3) allows for insight into patterns of flood frequency and magnitude during the post-planning period. From the examination of higher-probability flows (those with lower return periods) for both periods, it can be seen that the full period has lower corresponding streamflow values than the pre-planning period, indicating that more common flows have decreased in magnitude during the post-planning period. However, for lower-probability flows (those with higher return periods), differences between corresponding flow values for the full period and the pre-planning period become increasingly pronounced, with the full period having higher flow values for return periods greater than 2 years. This indicates a marked increase in the

magnitude of extreme flows during the post-planning period. These anomalous, high-magnitude flows could potentially be attributed in part to climate intensification, as was the case with the 2019 flood. In the Midwest, climate change has been linked to increased annual precipitation and higher intensity rainstorms [26], which in turn could contribute to increased flooding. In fact, increased precipitation in the region is projected to drive 10–30% increases in 100-year peak daily streamflows by the 2080s [27]. Historical data further support the possibility that climate change may be linked to increased peak flow magnitudes, as analysis of 100 years of US watershed data has brought to light increases in flooding in the northern eastern prairies and parts of the Midwest, particularly over recent decades [28].

The findings of the flood frequency analyses comparison are supported by Figure 8; Figure 8a shows annual peak flows and their 15-year moving average for SCW. Figure 8b shows the same SCW peak flow data, but with outliers above 4000 cfs, or 113.27 m³/s, reduced to 4000 cfs in order to more closely examine the typical behavior of streamflow by minimizing added variability from very high extremes. From Figure 8a, instances of abnormally high magnitude flows during the post-planning period can be observed, consistent with the notably increasing difference between computed flows for the full period and pre-planning period in the flood frequency analysis. Conversely, the 15-day average for the modified dataset in Figure 8b shows an overall downward trend in annual peak flow during the post-planning period, consistent with the implications of the lower streamflow values corresponding to lower return periods in the full period flood frequency analysis. In the early post-planning period, this downward trend is very mild, but it becomes more pronounced starting roughly 10 years after the advent of the post-planning period in 2005. During the 10 years immediately prior to the post-planning period, peak flows appeared to have stabilized at just above 2000 cfs, or 56.63 m³/s, before which could be observed an increasing trend in peak flows. The period of stabilization beginning in the 1990s could be explained by the preliminary conservation practices that began in the early 1990s, prior to the 2005 Shell Creek Watershed Management Plan. This trend of flood peak stabilization and subsequent decrease, with the exception of some high outliers later in the record, suggests that the conservation practices implemented beginning in the early 1990s and adopted more widely in 2005 through the SCW Management Plan may have had an impact on attenuating flood peaks, while climate change may be playing a role in intensifying flood peaks. Further research is necessary to disentangle the impacts of conservation practices and climate change on the watershed.

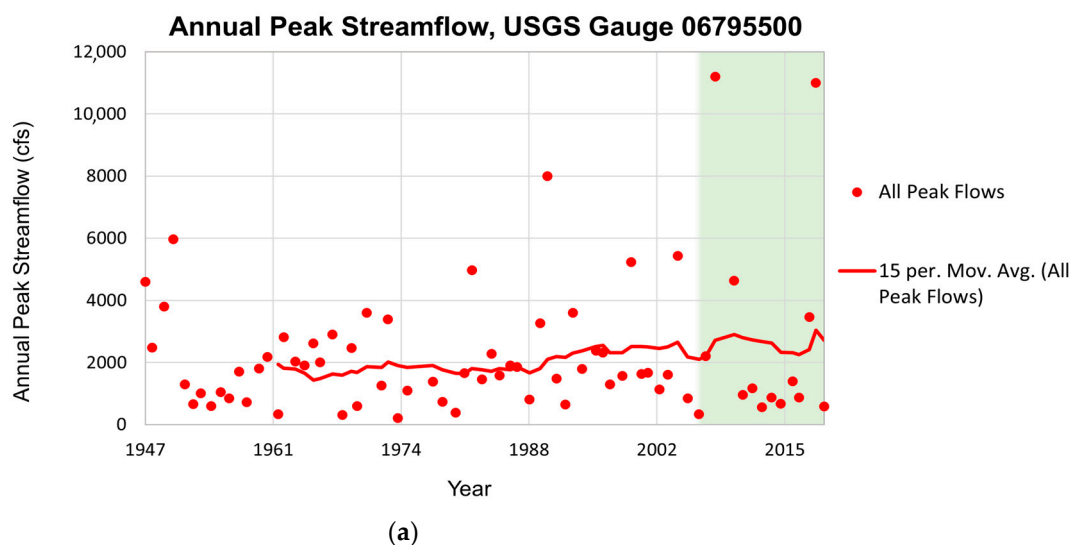


Figure 8. Cont.

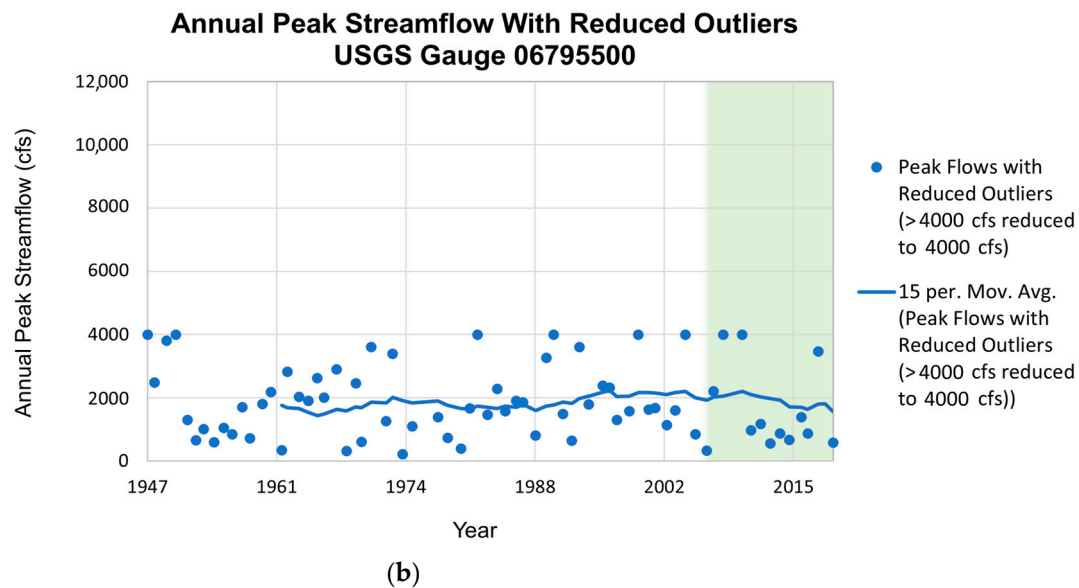


Figure 8. Unmodified and modified peak flow data for SCW, with the green section of each graph indicating the years of the post-planning period. (a) Annual peak flow data for SCW are plotted in red points. The 15-year moving average for the data is indicated by a red line. (b) Annual peak flow data for SCW with outliers above 4000 cfs reduced to 4000 cfs is plotted above in blue points. The blue line indicates the 15-year moving average for the modified peak streamflow data.

5. Conclusions

In this study, Shell Creek Watershed was examined in order to identify the respective influence of selected controlling factors on the hydrology of the catchment, improve understanding of hydrological processes there, and develop an effective rainfall-runoff model for the watershed using ANN. An additional intention of the work was to identify any possible relationship between changes in flood severity and the implementation of conservation practices. Variable selection was broken into three time periods determined by the start of widespread conservation practice implementation in SCW: the full time period (1990–2020), the pre-planning period (1990–2004), and the post-planning period (2005–2020). Model-free and model-based variable selection were used to identify variable influence on flooding, following which ANN rainfall-runoff models were developed using the optimal combination of input variables for each period. Additionally, drought and flood frequency analyses were conducted to improve hydrological understanding of the watershed. Their results contributed to analyses of the efficacy of certain variables in discharge prediction and the identification of changes in peak flow trends.

From the results and analysis of this work, several conclusions can be drawn, as well as some suggestions for future study. In terms of individual variable predictability, our results suggest that of the variables considered, precipitation is most tied to flooding in SCW. However, the correlation between precipitation and discharge for SCW is still quite low, so dependence solely on this factor for discharge prediction is discouraged. Instead, a combination of all variables is encouraged for this purpose. Out of all of the models developed in this study, the model for the post-planning period more closely reflects the current conditions of the watershed and has overall good prediction performance, making it the strongest candidate for discharge prediction in SCW. This model included all variables used in this study as inputs; however, future work may benefit from the investigation and inclusions of soil moisture, as the 3-month SPI results implied cases of high soil moisture coinciding with flooding. With regard to the relationship between flood-change and conservation implementation, flooding has seen an overall decrease in intensity during the post-planning period (with the exception of a few outliers); however, more intensive investigation is warranted to determine a cause for this trend. Furthermore, the

incorporation of climate change effects on SCW in such future studies may aid in shedding light on flood trend shifts and the appearance of anomalous flows. Understanding the potentially countering effects of climate change and conservation practices on flooding is crucial to advancing our fundamental understanding of the hydrological processes in the basin. This will require a more in-depth investigation with high-resolution remote sensing datasets and advanced hydrologic modeling.

Author Contributions: Conceptualization, A.A. and T.R.; methodology, A.A. and T.R.; software, A.A. and T.R.; formal analysis, A.A.; investigation, A.A.; resources, A.A. and T.R.; data curation, S.R.K. and I.K.; writing—original draft preparation, A.A.; writing—review and editing, A.A., S.R.K., I.K., J.E., E.T. and T.R.; visualization, A.A.; supervision, T.R.; funding acquisition, A.A. and T.R. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for Annushka Aliev was provided in part by the National Science Foundation (Award EEC-1950597; Title: REU Site: Sustainability of Horizontal Civil Networks in Rural Areas). Tirthankar Roy acknowledges funding from the Nebraska Department of Environment and Energy (EPA Grant 9007403-29, Title: Impacts of conservation practices on the water quality and quantity in the Shell Creek Watershed, Nebraska).

Data Availability Statement: Publicly available datasets were analyzed in this study. The NLDAS-2 data presented in this study is openly available in NLDAS Primary Forcing Data L4 Hourly 0.125×0.125 degree V002 at 10.5067/6J5LHHOHZHN4. The NWIS data can be found here: <https://maps.waterdata.usgs.gov/mapper>, accessed on 1 November 2022. The USGS data can be found here: https://waterdata.usgs.gov/ne/nwis/uv/?site_no=06795500. The NSIDC can be found here: <https://nsidc.org/data/NSIDC-0719/versions/1>, accessed on 1 November 2022.

Acknowledgments: Special thanks to Mark Seier, Gene Wissenburg, and Matt Bailey for their help and enthusiasm in connecting us with background information about SCW and touring us around the watershed. We thank Chris Poole from the Lower Platte North Natural Resources District for his help in creating Figure 1.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix includes cross-correlation plots with all input variables included for all three time periods of study.

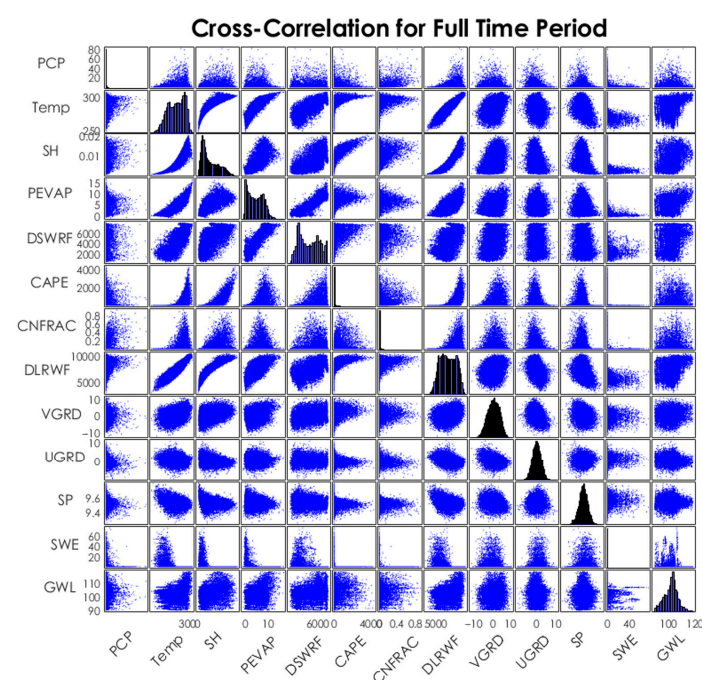


Figure A1. Cross-correlation pair plot for all 13 predictive variables over the full time period (1990–2020).

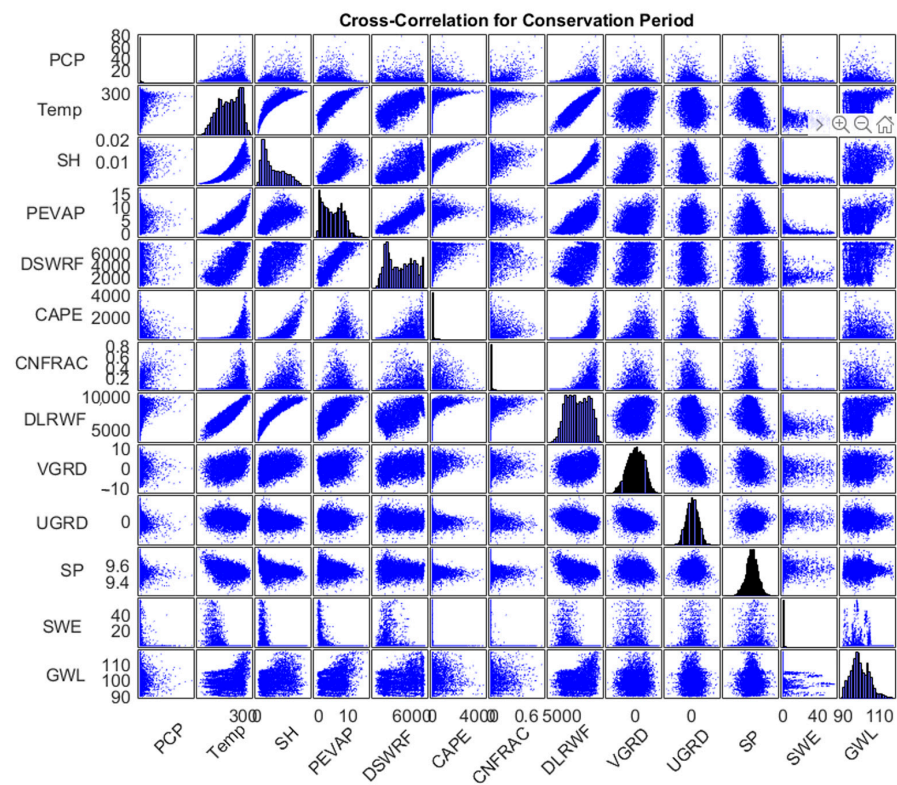


Figure A2. Cross-correlation pair plot for all 13 predictive variables over the conservation period (2005–2020).

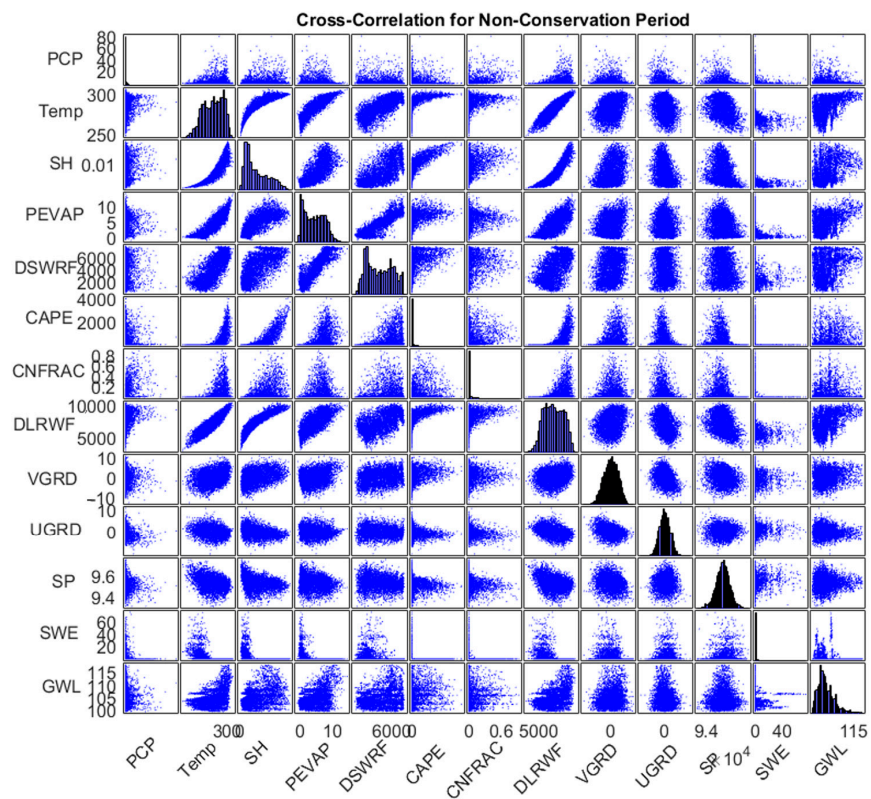


Figure A3. Cross-correlation pair plot for all 13 predictive variables over the non-conservation period (1990–2004).

Appendix B

This appendix includes flood frequency analysis results for the post-planning period and two halves of the pre-planning period, as well as tabulated results for flood frequency analysis.

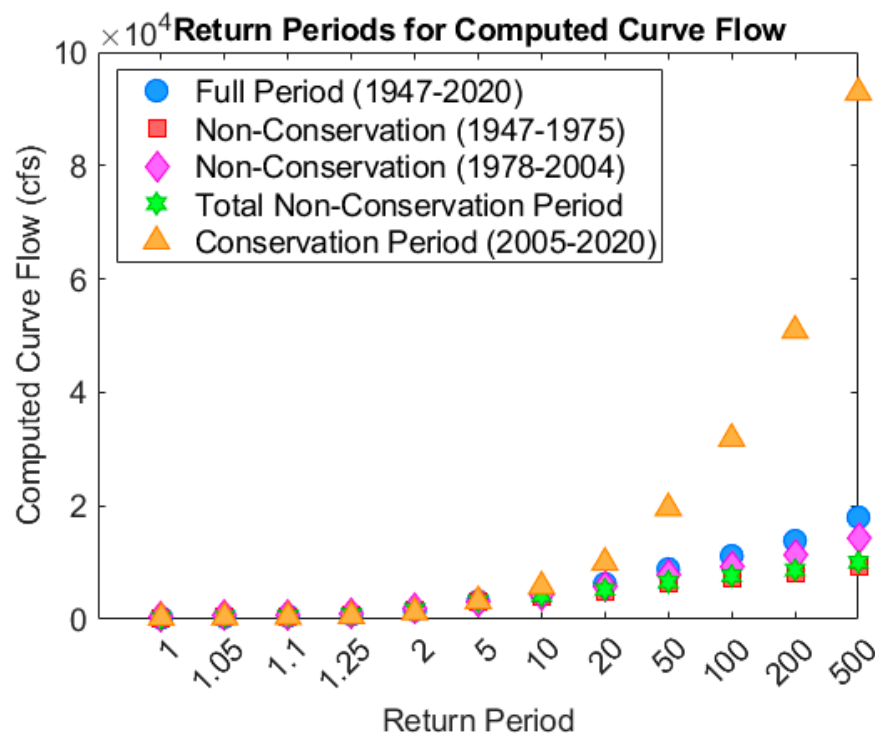


Figure A4. Flood frequency analysis results for 5 different time periods: The full period (1947–2020), the total non-conservation period (1947–1975), the first half of the non-conservation period (1947–1975), the second half of the non-conservation period (1978–2004), and the conservation period (2005–2020). Peak flow data was missing for the years 1976 and 1977, accounting for the break between the two non-conservation time periods.

Table A1. Return periods and their corresponding computed curve flows for the full time period, total pre-planning period, and the post-planning period.

Full Time Period		Total Pre-Planning Period		Post-Planning Period	
Return Period	Computed Curve Flow (cfs)	Return Period	Computed Curve Flow (cfs)	Return Period	Computed Curve Flow (cfs)
1	259.9	1	214.2	1	259.9
1.05	354	1.05	420.2	1.05	354
1.1	436.4	1.1	588.5	1.1	436.4
1.25	588.9	1.25	867.2	1.25	588.9
2	1222.1	2	1713.5	2	1222.1
5	3181.5	5	3136.1	5	3181.5
10	5783.7	10	4177.3	10	5783.7
20	9990	20	5216.1	20	9990
50	19,604.1	50	6594.8	50	19,604.1
100	31,838.9	100	7643.1	100	31,838.9
200	50,899.6	200	8695.1	200	50,899.6
500	92,883.2	500	10,088.9	500	92,883.2

Table A2. Return periods and their corresponding computed curve flows for the two halves of the pre-planning period.

Pre-Planning Period (1947–1975)		Pre-Planning Period (1978–2004)	
Return Period	Computed Curve Flow (cfs)	Return Period	Computed Curve Flow (cfs)
1	146.2	1	396.5
1.05	322.7	1.05	610.1
1.1	476	1.1	771.2
1.25	739.5	1.25	1028.5
2	1571.1	2	1807.7
5	2985	5	3232.6
10	4002.8	10	4410.9
20	4995.7	20	5722.3
50	6274	50	7700.4
100	7214.6	100	9407.8
200	8131.1	200	11,318.9
500	9303.8	500	14,192.1

Appendix C

This appendix contains the 1-month, 3-month, 6-month, 12-month, and combined SPI results expanded and plotted in individual figures.

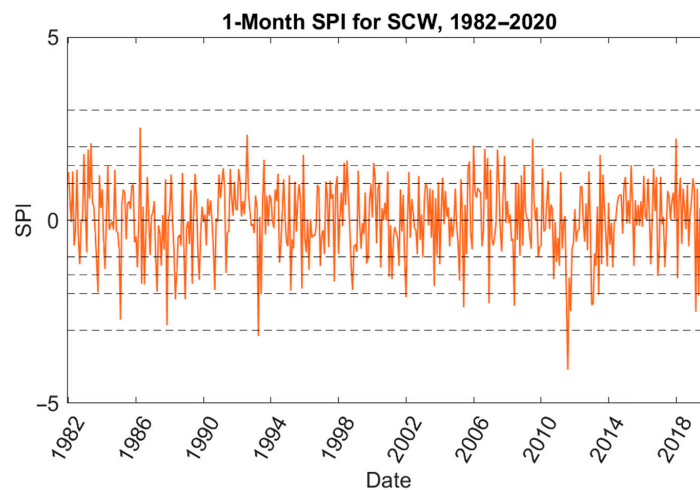


Figure A5. 1-Month SPI from 1982 to 2020.

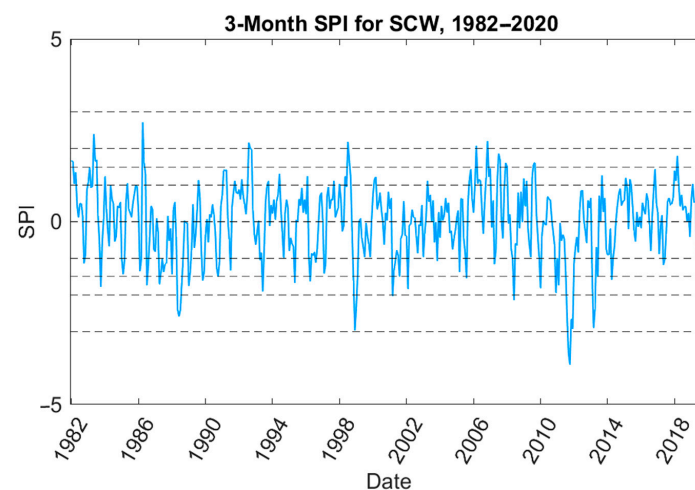


Figure A6. 3-Month SPI from 1982 to 2020.

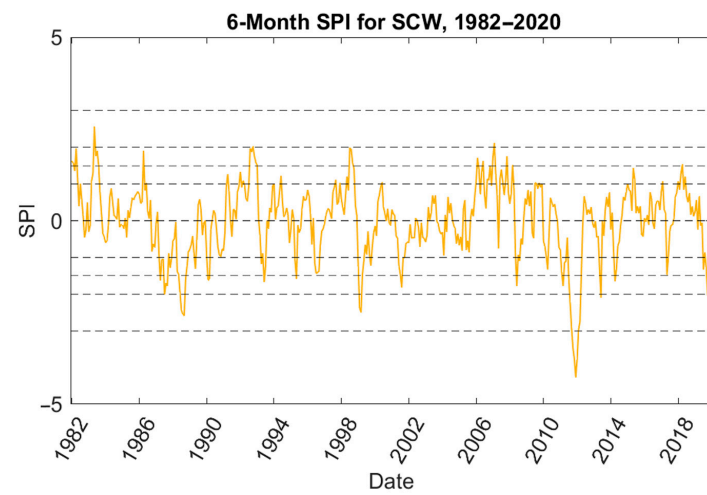


Figure A7. 6-Month SPI from 1982 to 2020.

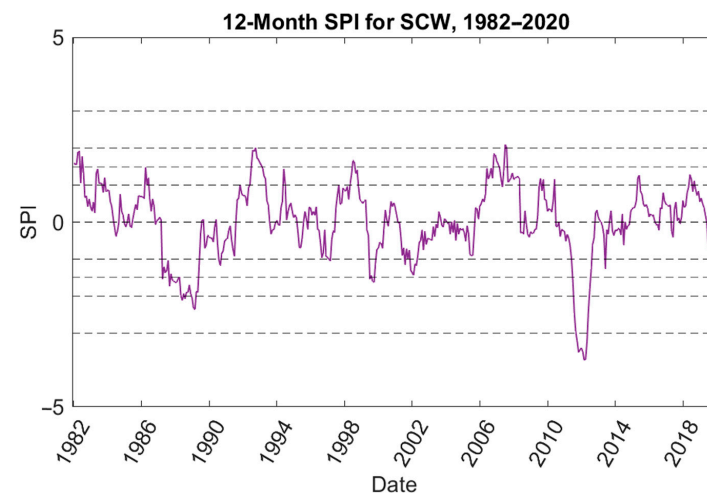


Figure A8. 12-Month SPI from 1982 to 2020.

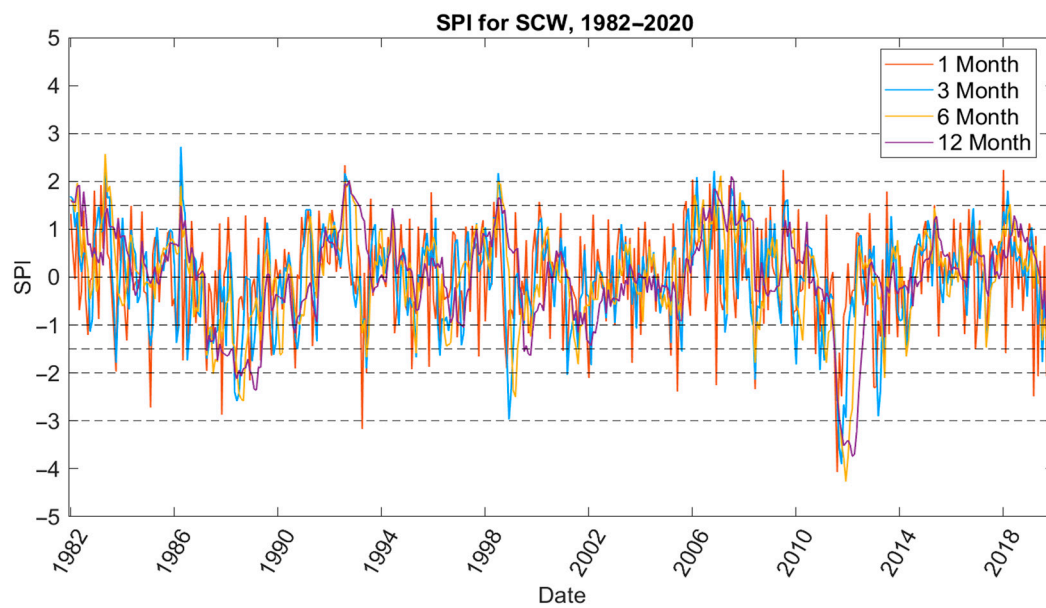


Figure A9. 1-, 3-, 6-, and 12-Month SPI from 1982 to 2020 plotted on the same axes.

Appendix D

This appendix includes some information about the current state of flood forecasting in the basin.

SCW's floodplains have been recorded and made publicly available on the Nebraska Department of Natural Resources (NeDNR) Floodplain Management Interactive Map, which can be found at the following link: <https://gis.ne.gov/portal/apps/webappviewer/index.html?id=7bc8738d3d8f4e87823cc604543b7ddf>, accessed on 1 November 2022.

Forecasting takes place primarily at Shell Creek near Columbus, NE, as that is the only gauge that provides daily streamflow data to date. River forecasts in this location rely on past precipitation and predicted precipitation amounts for the 48 h following the time of issuance for the forecast. Forecasts are only issued as needed during times of high water, and thus are not routinely available [7].

Information about data and measurements taken at the USGS 06795500 Shell Creek near Columbus, Nebraska stream site can be found at the following link: https://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=06795500 (accessed on 9 November 2022).

References

1. NWS. Flooding in Nebraska. National Weather Service. 13 October 2020. Available online: <https://www.weather.gov/safety/flood-states-ne> (accessed on 28 June 2021).
2. Flanagan, P.X.; Mahmood, R.; Umphlett, N.A.; Haacker, E.; Ray, C.; Sorensen, W.; Shulski, M.; Stiles, C.J.; Pearson, D.; Fajman, P. A Hydrometeorological Assessment of the Historic 2019 Flood of Nebraska, Iowa, and South Dakota. *Bull. Am. Meteorol. Soc.* **2020**, *101*, E817–E829. [\[CrossRef\]](#)
3. Salter, P. 'Just a Terrible Mess'—Ranchers, Farmers Left with Dead Animals, Flooded Fields, Work to Be Done; Journal Star: Lincoln, NE, USA, 2019.
4. Ducey, M. 'An Utter Disaster': Ag Losses from Nebraska Flooding Could Top \$1 Billion; Omaha World-Herald: Omaha, NE, USA, 2019.
5. Holman, K.D. Characterizing Antecedent Conditions Prior to Annual Maximum Flood Events in a High-Elevation Watershed Using Self-Organizing Maps. *J. Hydrometeorol.* **2018**, *19*, 1721–1730. [\[CrossRef\]](#)
6. NDEE. *Shell Creek Watershed Environmental Enhancement Plan*; Nebraska Department of Environment and Energy: Lincoln, NE, USA, 2016.
7. NWS. National Weather Service Advanced Hydrologic Prediction Service. Advanced Hydrologic Prediction Service. Available online: <https://water.weather.gov/ahps2/hydrograph.php?gage=clbn1&wfo=oax> (accessed on 22 June 2021).
8. Govindaraju, R.S. Artificial Neural Networks in Hydrology. II: Hydrologic Applications. *J. Hydrol. Eng.* **2000**, *5*, 124–137.
9. Dawson, C.; Wilby, R.L. Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr. Earth Environ.* **2001**, *25*, 80–108. [\[CrossRef\]](#)
10. Sahoo, S.; Jha, M.K. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: A comparative assessment. *Hydrogeol. J.* **2013**, *21*, 1865–1887. [\[CrossRef\]](#)
11. Hsu, K.-L.; Gupta, H.V.; Sorooshian, S. Artificial Neural Network Modeling of the Rainfall-Runoff Process. *Water Resour. Res.* **1995**, *31*, 2517–2530. [\[CrossRef\]](#)
12. French, M.N.; Krajewski, W.F.; Cuykendall, R.R. Rainfall forecasting in space and time using a neural network. *J. Hydrol.* **1992**, *137*, 1–31. [\[CrossRef\]](#)
13. Jain, A.; Roy, T. Evaporation modelling using neural networks for assessing the self-sustainability of a water body. *Lakes Reserv. Res. Manag.* **2017**, *22*, 123–133. [\[CrossRef\]](#)
14. Roy, T.; Schütze, N.; Grundmann, J.; Brettschneider, M.; Jain, A. Optimal groundwater management using state-space surrogate models: A case study for an arid coastal region. *J. Hydroinform.* **2016**, *18*, 666–686. [\[CrossRef\]](#)
15. Rajurkar, M.; Kothiyari, U.; Chaube, U. Modeling of the daily rainfall-runoff relationship with artificial neural network. *J. Hydrol.* **2004**, *285*, 96–113. [\[CrossRef\]](#)
16. Mockus, V.; Werner, J.; Woodward, D.E.; Nielson, R.; Dobos, R.; Hjelmfelt, A.; Hoeff, C.C. Chapter 7: Hydrologic Soil Groups. In *National Engineering Handbook Part 630 Hydrology* (pp. 7–2–7–2). *Essay*; United States Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA, 2009.
17. LPNNRD. 2009 Lower Platte North Natural Resources District Hydrogeologic Evaluation and Subarea Delineation Study. Olsson Associates: 2009. Available online: https://www.enwra.org/media/LPN_Subarea%20Delineation%20Report%20Final.pdf (accessed on 1 November 2022).
18. NDEE. *Total Maximum Daily Loads for Shell Creek: LP1-20700, Atrazine*; Water Quality Planning Unit, Water Quality Division, Nebraska Department of Environmental Quality: Lincoln, NE, USA, 2007.
19. Hill, E.; TePoel, D. Nebraska's Shell Creek Watershed Makes History for Successful Water Clean-Up Efforts. LPNNRD. Available online: https://lpnnrd.org/wp-content/uploads/2018/08/Shell_Creek_Press_Release.pdf (accessed on 1 November 2022).

20. Traylor, E.; McCullough, C. Shell Creek Watershed: A Tough Nut to Crack. NDEE. 2018. Available online: <http://neiwpc.org/wp-content/uploads/2018/11/McCullough-Shell-Creek-A-Tough-Nut-to-Crack-NPS-Workshop-Final-10-26.pdf> (accessed on 1 July 2021).
21. Xia, Y.; Al, E. NCEP/EMC (2009), NLDAS Primary Forcing Data L4 Hourly 0.125×0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Available online: <https://doi.org/10.5067/6j5lhohzhn4> (accessed on 1 November 2022).
22. Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmos.* **2012**, *117*, 25. [CrossRef]
23. Broxton, P.; Zeng, X.; Dawson, N. Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data Over the Conterminous US, Version 1. January 1982 to December 2020, 42°N, 99°W; 41°N, 96°W. Boulder, Colorado, USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. 2019. Available online: <https://doi.org/10.5067/0GGPB220EX6A> (accessed on 1 July 2021).
24. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
25. McKee, T.B.; Doesken, N.J.; Kleist, J. 1993: The relationship of drought frequency and duration of time scales. In Proceedings of the Eighth Conference on Applied Climatology, Anaheim, CA, USA, 17–23 January 1993; American Meteorological Society: Boston, MA, USA; pp. 179–186.
26. Harrington, S. Did Climate Change Cause Midwest Flooding? Yale Climate Connections. 2019. Available online: <https://yaleclimateconnections.org/2019/04/did-climate-change-cause-midwest-flooding/> (accessed on 1 November 2022).
27. Byun, K.; Chiu, C.-M.; Hamlet, A.F. Effects of 21st century climate change on seasonal flow regimes and hydrologic extremes over the Midwest and Great Lakes region of the US. *Sci. Total. Environ.* **2018**, *650*, 1261–1277. [CrossRef] [PubMed]
28. Peterson, T.C.; Heim, R.R., Jr.; Hirsch, R.; Kaiser, D.P.; Brooks, H.; Diffenbaugh, N.S.; Dole, R.M.; Giovannettone, J.P.; Guirguis, K.; Karl, T.R.; et al. Monitoring and Understanding Changes in Heat Waves, Cold Waves, Floods, and Droughts in the United States: State of Knowledge. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 821–834. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.