Groundwater Estimation from Major Physical Hydrology Components Using Artificial Neural Networks and Deep Learning

Hassan Afzaal [1], Aitazaz A. Farooque [1,*], Farhat Abbas [1,*] , Bishnu Acharya [1] and Travis Esau [2]

[1] Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, PE C1A4P3, Canada; hafzaal2@upei.ca (H.A.); bacharya@upei.ca (B.A.)

[2] Engineering Department, Dalhousie University, Agriculture Campus, Truro, NS B2N5E3, Canada; tesau@dal.ca

* Correspondence: afarooque@upei.ca (A.A.F.); fabbas@upei.ca (F.A.); Tel.: +1-902-566-6084 (A.A.F.)

Abstract: Precise estimation of physical hydrology components including groundwater levels (GWLs) is a challenging task, especially in relatively non-contiguous watersheds. This study estimates GWLs with deep learning and artificial neural networks (ANNs), namely a multilayer perceptron (MLP), long short term memory (LSTM), and a convolutional neural network (CNN) with four different input variable combinations for two watersheds (Baltic River and Long Creek) in Prince Edward Island, Canada. Variables including stream level, stream flow, precipitation, relative humidity, mean temperature, evapotranspiration, heat degree days, dew point temperature, and evapotranspiration for the 2011–2017 period were used as input variables. Using a hit and trial approach and various hyperparameters, all ANNs were trained from scratched (2011–2015) and validated (2016–2017). The stream level was the major contributor to GWL fluctuation for the Baltic River and Long Creek watersheds (R^2 = 50.8 and 49.1%, respectively). The MLP performed better in validation for Baltic River and Long Creek watersheds (RMSE = 0.471 and 1.15, respectively). Increased number of variables from 1 to 4 improved the RMSE for the Baltic River watershed by 11% and for the Long Creek watershed by 1.6%. The deep learning techniques introduced in this study to estimate GWL fluctuations are convenient and accurate as compared to collection of periodic dips based on the groundwater monitoring wells for groundwater inventory control and management.

Keywords: deep learning; evapotranspiration; stream flow; stream level; watershed hydrology

1. Introduction

Groundwater is the major source of industrial and potable water supplies in Prince Edward Island, Canada [1]. Over the past few years, there has been increased demand in the agriculture sector for supplemental irrigation, which poses several challenges for water and resource managers. Because of the relatively small and non-contiguous watersheds in Prince Edward Island, pumping groundwater has also raised concerns for groundwater sustainability due to the island's uneven topography [2]. An inventory of groundwater is necessary for efficient water resource management, especially in relation to growing groundwater demands for agricultural use. It is neither feasible nor economical to install and manage monitoring groundwater wells in a place like Prince Edward Island, which consists of 260 watersheds for efficient water management. The inventory control of the groundwater resource can ensure the sustainability of water resources in the areas where groundwater pumping is common for supplemental irrigation or for domestic use.

Groundwater level (GWL) modeling provides useful information to water resource managers, engineers, and policy makers to make appropriate decisions. The modeling of GWLs is a complicated

procedure that requires thorough knowledge of physical hydrological parameters, big data, hydrological models, model inputs, and the geometry of watersheds [3]. Aspects of hydrogeology—i.e., geological factors affecting the distribution and movement of groundwater underneath the soil surface—need to be properly understood when modeling GWLs and manipulating the modeling results. Watershed scale fluctuations in GWLs occur over a period of several decades, and the resulting cumulative effects on streamflow depletion may not be fully realized for years [4]. Resultantly, depending upon the distance of the pumping station from the stream and the geologic characteristics of the aquifer, the groundwater system may take decades to recover from streamflow depletion caused by intermittent pumping. Components of the surface—and the sub-surface physical hydrology of a watershed—i.e., streamflow and groundwater flow, respectively—are interconnected, making the stream–aquifer interaction one of the key processes governing the groundwater flow pattern in a watershed [1]. Groundwater fluctuations affect streamflow and vice versa, as the pumping wells capture groundwater that would otherwise discharge to connected streams, rivers, and other surface-water bodies [4]. Francis [5] reported that, in typical watersheds of Prince Edward Island, the base flow represents almost 80% of the streamflow in the late summer and fall months. Stream length in these island watersheds ranges from less than 1 km to 20 km. Stream widths vary from less than 1 m at the head to 30 m at the estuary, with all the components of soil including sand, silt, and clay, contributing to the formation of the streambeds of the island watersheds, making them weakly permeable through reduced hydraulic conductivity [1].

Several numerical and conceptual methods have been reported in the literature for GWL estimation. For example, Mohammadi [6] tested artificial neural network (ANN) models and MODFLOW to simulate the monthly GWLs of Karstic aquifers in Iran. The results indicated that the ANN models require less input data and time to run as compared to conventional models, such as MODFLOW. Several experiments in GWL modeling have shown that ANNs could be the better alternative over conceptual models [7]. Mohanty et al. [8] found that ANNs are better predictors of GWLs than MODFLOW for short term predictions. Karandish and Šimůnek [9] compared artificial intelligence (AI) methods with physical modeling. Results indicated that the AI methods performed well in water-stressed conditions as compared to HYDRUS-2D. Therefore, AI methods provide promising tools for GWL predictions [10]. Several AI methods have been used by many researchers because of their simplicity and acceptable performance in different parts of the world [7].

Deep learning has been used to solve real-world problems relating to multi-magnitude data [11], determining the bearing capacity of concrete-steel columns [12], soil liquefaction [13], and solving genetic algorithms [14]. Deep learning and ANNs handle the non-linear behavior of time series better than regular regression [7]. Several hydrological parameters such as precipitation, streamflow, and GWLs exhibit time dependence and can be treated as time series functions. Time series functions usually exhibit non-linear relations, which are difficult to handle with simple static models. A multilayer perceptron (MLP), being the simplest kind of ANN, can approximate the functions related to dynamic hydrological problems that are difficult to model with numerical static methods. Similarly, an MLP can also predict values from a correlated input variable, which may be mapped with an output variable. For example, Sahoo and Jha [15] compared multiple regression and ANNs with a Levenberg Marquardt (LM) algorithm to predict GWLs in Kochi Prefecture of Shikoku Island, Japan. Graphical findings and analysis suggested the superiority of ANNs over multiple regression. Kouziokas et al. [16] used an MLP neural network with four different algorithms to forecast daily GWLs in Pennsylvania, USA. Temperature, humidity, and precipitation were used as input variables to predict GWLs. The four algorithms used in the study were Resilient Backpropagation (RB), LM, the Scaled Conjugate Gradient, and the BFGS Quasi-Newton. The LM algorithm performed better than the other algorithms in the prediction of GWLs. Juan et al. [17] developed two MLPs with an LM algorithm and two sets of input data, i.e., temperature–precipitation–GWL and temperature–precipitation in Qinghai-Tibet plateau, China. The MLP models with 3-input variables performed better than the 2-input variables.

Because of the relatively simple structure of the MLP, it is not possible for these models to store the previous information in the time series, unlike a recurrent neural network (RNN). An RNN contains a

memory block for storing previous steps of information in time series problems. Coulibaly et al. [18] compared three types of ANN models using GWL, precipitation, and temperature time series as input variables to simulate monthly GWLs in the Gondo aquifer, Burkina Faso. The results were also compared with radial basis function networks, generalized radial basis functions, and probabilistic neural networks. The results suggested that the RNN is the most efficient model compared to static structure ANNs. Furthermore, generalized radial basis functions are poorly performing models in monthly GWL simulations when compared to others. Müller et al. [19] compared an MLP with an RNN and more advanced ANNs to predict GWLs in Butte County, California, USA. They used three different methods to optimize model hyperparameters including two surrogate model-based algorithms and one random sampling method. They used stream flow, precipitation, and ambient temperature as input variables and estimated the GWLs after training these models. The results suggested that the MLP performed better than the ANNs, including long short term memory (LSTM) and the convolutional neural network (CNN). Babu et al. [20] compared different algorithms to evaluate the RNN in GWL forecasting in Karnataka, India. They compared the LM, gradient descent with momentum, and the adaptive learning rate back propagation algorithm in predicting GWLs with the RNN. Their results suggested that the LM algorithm performed better with the RNN than all other algorithms in GWL modeling. Although RNNs are enabled to store the previous information in their memory blocks, in large time series sequences the vanishing gradient problem hampers the learning of these models. The vanishing gradient problem in the RNN occurs when gradient updates become very small and add no significant difference to overall learning. To overcome the vanishing gradient problems of RNNs, more advanced models such as LSTM were introduced, and these are capable of storing important information with long sequences. Recently, Zhang et al. [10] conducted a similar study in which they compared LSTM with an MLP to predict GWLs in the Hetao irrigation district, China. Monthly water diversion, evaporation, precipitation, temperature, and time were used as input data to predict water table depth. Their results suggested that the LSTM model performed better than the MLP and could contribute to a strong learning ability on time series data. However, they did not include a detailed comparison of several input combinations for developing less data-intense models. Similarly, the more advanced time series models such as the CNN were not included [10].

The CNNs are the more advanced ANNs with usually a high number of hidden layers as compared to the RNN and the MLP. CNNs have gained popularity in terms of image data; however, they can perform equally well on time series data if modeled accordingly. Because of the more advanced functions of CNNs, such as convolution and max-pooling, these models have proved to be very promising in time series prediction. However, there is very limited literature available on GWL modeling using CNNs. Lähivaara et al. [21] used CNNs to estimate the GWLs and groundwater storage using seismic data. They applied the Galerkin method to model wave propagation followed by deep CNNs for parameter estimations and found that CNNs can extract additional information from seismic data about the aquifer.

The review literature suggests that the MLP and LSTM are promising methods that have been successfully used in many time series problems related to GWL modeling. In addition to the MLP and LSTM, the CNN, a relatively new method, was used in this study to investigate GWL estimation in the Long Creek and Baltic River watersheds of Prince Edward Island. The reason behind conducting this research was that it is currently not feasible for groundwater managers to manage the large number of watersheds in Prince Edward Island due to the small number of monitoring wells. The challenges of this study involve the unavailability of guidelines for selecting appropriate input variables for the optimum prediction of GWLs. Therefore, most of the input variables have been selected based on data availability. The specific objective of this study was to select appropriate ANNs and the best combination of input variables for the accurate possible estimation of GWLs.

2. Materials and Methods

2.1. Site Selection

Prince Edward Island is the smallest but most populous Canadian Atlantic province, having a pastoral landscape consisting of several rolling hills, woods, reddish white sand beaches, ocean coves, and red soil. With the uneven topography of Prince Edward Island, its streams are mostly non-contagious. For this study, the Baltic River and Long Creek watersheds situated in the center of Prince Edward Island were selected as experimental watersheds because of the relatively large number of irrigation wells in the area that might have contributing fluctuations in GWLs (Figure 1).
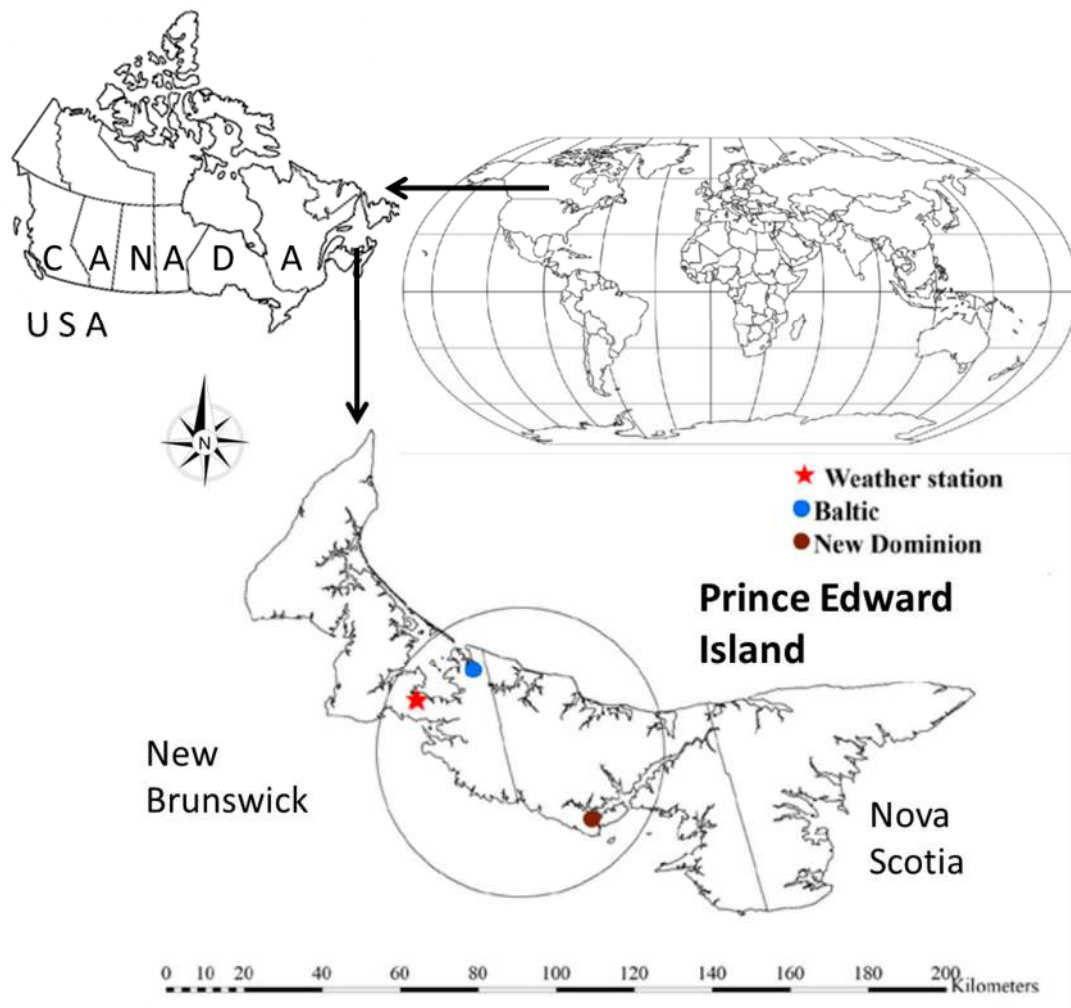

Figure 1. Locations of the experimental watersheds and the weather station in Prince Edward Island, Canada. The area of the relatively large number of irrigation wells has been encircled.

2.2. Data Collection

Two monitoring wells, the Baltic (46.51000° N, 63.648056° W) and New Dominion (46.170278° N, 63.250000° W) wells, were selected for collection of actual GWL data of the Baltic River and Long Creek watersheds, respectively. The Baltic groundwater monitoring well was installed in the Baltic River watershed at an elevation of 25 m above mean sea level. The New Dominion groundwater monitoring well was installed in the Long Creek watershed at an elevation of 19.93 m above mean sea level. Daily groundwater level data of these two monitoring wells for a seven-year period (2011–2017) was obtained from the Department of Communities, Land and Environment, Prince Edward Island. Weather data for this period was collected from a local weather station in Summerside

(46.3934° N, 63.7902° W). The data included mean temperature, dewpoint temperature, heat degree days, precipitation, relative humidity, reference evapotranspiration (RET), stream level, and stream flow. These data were used as input variables for the modeling of GWLs. The RET was calculated using the Penman Monteith equation below:

$$RET\left(\frac{mm}{day}\right) = \frac{\Delta(R_n - G) + \rho_a c_p\left(\frac{es-ea}{ra}\right)}{\Delta + \gamma\left(1 + \frac{r_s}{r_a}\right)} \tag{1}$$

where Δ is the slope of the vapor saturation pressure, R_n is the net radiation, G is the soil heat flux, ρ_a is the mean air density at constant air pressure, C_p is the specific heat of the air, $es - ea$ is the vapor pressure deficit, Υ is the psychrometric constant, Υ_s is the surface resistance, and Υ_a is the aerodynamic resistance m.

2.3. Regression Subset Analysis for Input Variable Selection

Regression subset analysis was conducted to choose the appropriate variables, several combinations of which were used to select input variables for the deep learning models. The basic subset regression analysis was conducted in Minitab software (Version 18). Best Subsets Regression compares the different regression models that contain subsets of the specified predictors. The best-fitting models are suggested by Minitab containing one predictor, two predictors, and so on. Based on the highest regression, four input variable combinations were tested, namely 1-input variable, 2-input variable, 3-input variable, and 4-input variable combinations.

2.4. The Multilayer Perceptron for Groundwater Level Modeling

The multilayer perception is the simplest type of ANNs, which are biologically inspired computation models comprised of several layers, namely input, hidden, and output layers. All layers are connected each other with neurons—a basic processing unit of ANNs. Input layers take all the input variables such as temperature, precipitation, etc. to predict output variables such as GWLs. Hidden and output layers handle the weights and biases from input layers through activation functions. The most common activation functions are the rectified linear unit (relu), sigmoid, and tanh. In modeling, sigmoid activation function is the most common function [22]. However, for this study, the rectified linear unit was selected for its better performance than the others. Most of the reviewed papers used trial and error methods to model the ANN layers and neurons [7]. For this study, two hidden layers and 100 neurons performed better. The MLP requires training data to adjust weight and bias for optimal prediction. Several learning algorithms have been used by modelers, such as back propagation, the LM, Bayesian regularization, adaptive learning rate back propagation, and gradient descent with momentum. Krishna et al. [23] compared several training algorithms in the groundwater modeling of an urban coastal aquifer in Andhra Pradesh state, India. They found that the LM algorithm was among the best learning algorithms compared to others. It is the most common algorithm in groundwater modeling [9,17,18]. This algorithm works better for determining the local minima of error functions, resulting in increased prediction accuracy. For this study, the LM algorithm was selected to determine loss function. The MLP was used in this study for various input variable combinations (Figure 2).
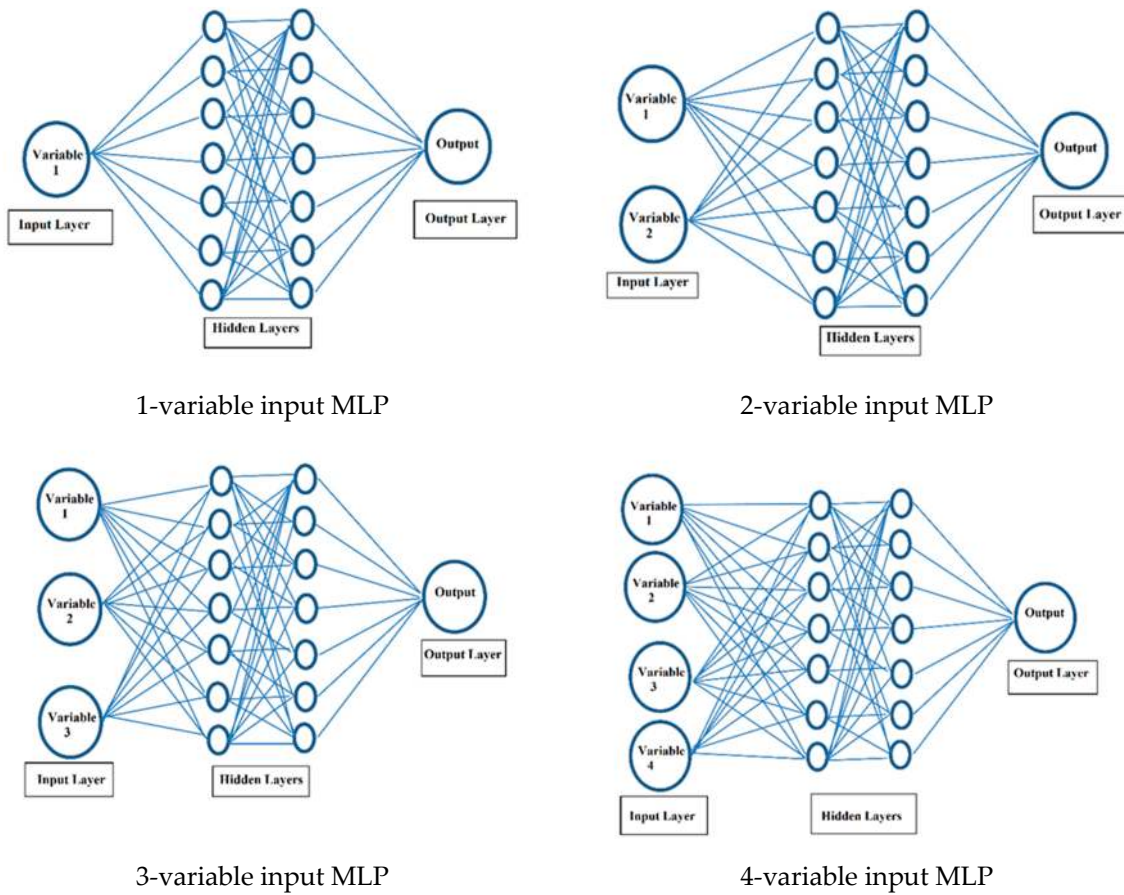
1-variable input MLP

2-variable input MLP

3-variable input MLP

4-variable input MLP

Figure 2. The multilayer perceptron (MLP) model for various input variable combinations.

2.5. Long Short-Term Memory Neural Networks

The LSTM neural network is a special kind of RNN, which is a sequence-based model that can store and relate the previous information in a sequence, enabling it to predict time series problems. However, RNNs cannot store longer sequences because of the gradient vanishing problem in early layers. Gradient vanishing in an early layer is sometimes referred to as short-term memory neural networks. All RNNs form a chain-like structure as the information flows through them (Figure 3a). The RNNs store information in each stage based on time/sequence steps in the form of a hidden state, i.e., h_t for each input X_t. The tanh function in the memory block of an RNN scales the input data between −1 and 1.
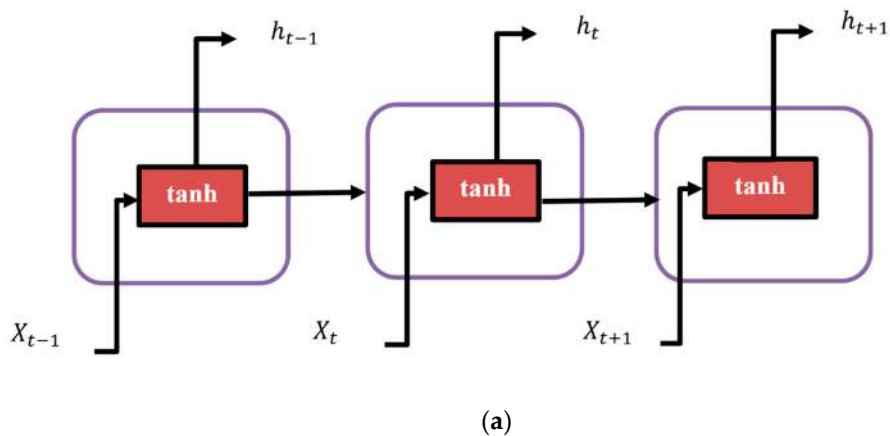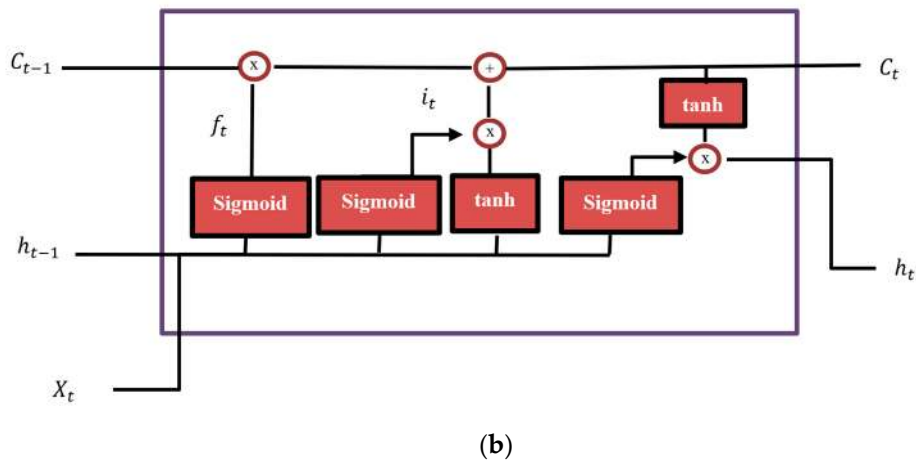
(**a**)

Figure 3. *Cont.*

(b)

Figure 3. The memory block of (**a**) recurrent neural networks (RNNs) and (**b**) long short term memory (LSTM) neural networks.

The LSTM addresses short-term memory problems by adding more states in the memory blocks of the RNN. In LSTM, a forget state (f_t) and a cell state (C_t) are added to retain the temporal and sequal dependence of previous blocks. The f_t in LSTM has the ability to discard or keep the information based on the sigmoid function output values. The values closer to 1 are kept, and those closer to 0 are forgotten. After passing f_t, the input variable X_t and the previous cell hidden state h_{t-1} are passed through the sigmoid and tanh functions. The dot product of the tanh and sigmoid functions are then added with f_t to compute C_t. The more detailed overview of LSTM information flow memory block is described in Figure 3b.

2.6. Convolutional Nural Networks

CNNs are mostly used for machine vision applcations in two-dimensional (2D) images. However, a one-dimensional (1D) CNN has several applications in time series classification and natural language processing. Use of a 1D CNN in GWL modeling is very limited in the literature. Convolution layers convolute the feature on 1D matrices. Because of fewer dimensions, the convolution filters have less processing work, making them computationally faster than 2D CNN. There has been very limited research conducted on GWL modeling using CNNs. However, based on the hit and trial method, five layers were used in the CNN in this study. The first layer was the convolulational layer followed by a maxpooling layer to extract features. To connect the maxpooling layer with the fully connected layers, a third, flatten layer was added. The last two layers were fully connected or dense layers to obtain the output vectors (Figure 4).
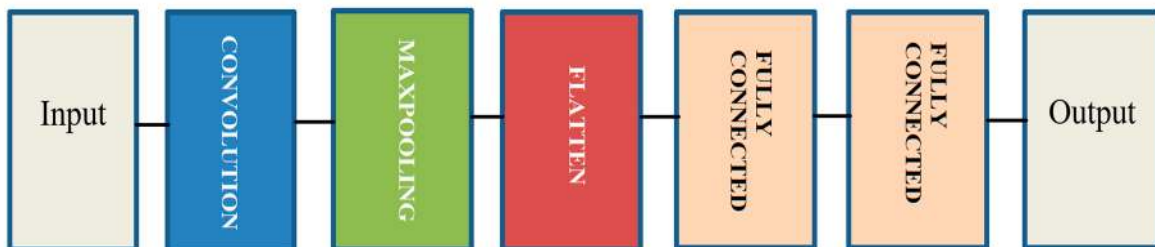

Figure 4. Architecture of the 1D CNN.

2.7. Hyperparameter Tuning of ANNs

The tuned parameters were selected for different ANNs based on the highest accuracy using the hit and trial method. All ANNs were trained from scratch using hyperparameters including activation

functions, number of neurons, number of layers, neurons, optimization, and learning rate. For the MLP, these respective hyperparameters were Relu, 2, 100, Adam, and 10^{-3}. For LSTM and the CNN, the activation functions were Tanh and Relu, and the neurons were 50 and 64, respectively. The CNN had 5 layers. The rest of the hyperparameters for LSTM and the CNN were the same as those of the MLP.

2.8. Model Evaluation Criteria

The model performance was evaluated by loss functions. The root means square error (RMSE) is well known model evaluation criteria used in various studies to evaluate the model performance [24,25]. It squares the difference between predicted and actual value and may range between 0 and infinity. The RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(yi - \hat{y}i)^2}{N}} \tag{2}$$

where y_i is the actual value at the *i*th time, $\hat{y}_i$ is the estimated value at the *i*th time, and *i* ranges from 1 to *N*.

$$R^2 = \sqrt{\frac{\sum_{i=1}^{N}(yi - \overline{y})^2 - \sum_{i=1}^{N}(yi - \hat{y}i)^2}{\sum_{1=1}^{N}(yi - \overline{y})^2}} \tag{3}$$

where y_i, $\hat{y}_i$, and *i* are the same as above, and $\overline{y}$ is the mean value of y_i. The estimated GWLs were plotted against actual GWLs and against a 1:1 line to evaluate the over- or under-estimation of GWLs overlapping and its scattering/clustering. To overcome the noisy effects of the large data points used in this study, a data normalization technique was performed. The max-min normalization performed well with this set of data. However, after the model training, the data were back-transformed to show trends in the figures.

3. Results and Discussion

3.1. Descriptive Statistics of Input Variables

The descriptive statistics of several input variables for GWL modeling are displayed in Table 1. The mean temperature ranged between −22.5 and 26.5 °C in Summerside, Prince Edward Island, in the 2011–2017 period. The difference between mean temperature and median temperature is −0.291 °C, which depicts a slightly skewed distribution of mean temperature. The dewpoint temperature ranged between −27.692 and 21.775 °C. The dewpoint temperature distribution showed slightly more skewness than the mean temperature. Because of seasonality, both the mean temperature and the dewpoint temperature show a bimodal distribution. The relative humidity of selected region ranges between 46.7 to 98.25% for year 2011–2017. Because of the high variation, the relative humidity showed a high standard deviation for such a large dataset. The heating degree days ranged between 0 and 40.5 °C from 2011 to 2017. Because of a high frequency of 0 °C values of heat degree days in winter, the distribution has a high peak and skews towards the right. The RET ranges between 0.0839 and 7.2631 mm/day. Because of the high number of cold days in Prince Edward Island, the distribution of RET is right-skewed with a skewness of 0.74. However, because of a lower variability in the RET range, it has a lower standard deviation. The island received a total precipitation of 753 to 1070.3 mm/year for the years from 2011 to 2017. The daily precipitation received by the island is highly variable, i.e., 0–103.8 mm/day. Because of storms and a lack of rainfall on several days of any given year, the precipitation distribution is highly right-skewed with a skewness coefficient of 5.72. Daily water level fluctuations in the Baltic groundwater monitoring well located in the Baltic River watershed range between 13.1 and 18.6 m in the years 2011–2017. Because of lower fluctuations in GWLs, the standard deviation is lower, i.e., 0.93 m. The distribution of the Baltic River GWLs showed right skewness, with a skewness coefficient of 0.58. Daily water level fluctuations in the New Dominion groundwater monitoring well, located in the Long Creek watershed, range between 9.16 and 18.12 m in the years

between 2011 and 2017. Because of the high fluctuations in the Long Creek GWLs compared with the Baltic River GWLs, the standard deviation is relatively high, i.e., 1.58 m. The Long Creek distribution showed a right skewness, with a skewness coefficient of 0.58. The associated stream discharge for the Long Creek and Baltic River monitoring wells is 0.46–42.4 m^3/s and 0.548–51 m^3/s, respectively. The distribution of the stream discharge for both monitoring wells showed a right skewness, with a skewness coefficient greater than 7. A high skewness coefficient reflects the high occurrence of peak flow levels probably during storms. The associated stream levels for the New Dominion and Baltic monitoring wells are 1.055–2.835 m and 0.5–2.14 m, respectively. Similar trends can be seen with the stream levels associated with both monitoring wells, i.e., a high skewness coefficient.

Table 1. Descriptive statistics of input and output variables for groundwater level modeling.

Variable	Maximum	Minimum	Mean ± SD	Skewness
Mean temperature (°C)	26.5	−22.5	6.41 ± 10.4	−0.26
Dew point temperature (°C)	21.8	−28.0	2.66 ± 10.2	−0.33
Relative humidity (%)	98.3	46.7	77.7 ± 10.2	−0.37
Heat degree days (°C)	40.5	0.00	12.0 ± 9.90	0.40
Reference evapotranspiration (mm/day)	7.26	0.08	2.05 ± 1.50	0.74
Precipitation (mm)	103.8	0.00	2.41 ± 6.14	5.72
The Baltic River watershed daily groundwater levels (m)	18.6	13.1	14.6 ± 0.93	0.58
Stream flow for the Baltic River watershed (m^3/s)	51.0	0.548	2.65 ± 3.21	7.57
Stream level for the Baltic River watershed (m)	2.14	0.50	0.68 ± 0.15	2.99
The Long Creek watershed daily groundwater levels (m)	18.1	9.16	12.8 ± 1.58	0.06
Stream flow for the Long Creek watershed (m^3/s)	42.4	0.46	2.00 ± 2.21	7.04
Stream level for the Long Creek watershed (m)	2.84	1.06	1.25 ± 0.17	2.51

3.2. Regression Subset Analyis for Variable Selection

For the 1-input variable combination, stream level was found to be the highest contributor among the eight selected different variables for the Baltic and New Dominion monitoring wells. The correlation coefficient for the Baltic River GWLs and the associated stream level was found to be 50.8. Similarly, the correlation coefficient between the Long Creek GWLs and the associated stream level was found to be 49.1. For the 2-input variables combinations, stream flow and stream levels were selected for the Baltic River watershed with a coefficient correlation of 63%. For the Long Creek watershed, stream level and dew point temperature proved to be the best input variables for the 2-input variable combinations. For the 3-input variable combinations, stream level, stream flow, and evapotranspiration were selected for the Baltic River watershed, with a slightly high correlation coefficient of 65.8 compared with that of the 2-input variable combinations. Stream level mean temperature and evapotranspiration variables were selected for the Long Creek watershed, with a low coefficient correlation of 57.2 compared with that of the Baltic River watershed 3-input variable combinations (Table 2). Four-input variables did not add much toward defining GWL variability, i.e., only 0.4 and 2.2% increases in correlation coefficients for the Baltic River and Long Creek watersheds, respectively. Interestingly, the precipitation variable did not contribute towards GWL fluctuations for both watersheds. A comparative study of GWL modeling using AI methods [7] reported that precipitation was used 48 times in GWL modeling. Statistical analysis, specifically in regression analysis, should be conducted to select appropriate input variables for groundwater modeling using AI methods [7]. Several studies have used past GWLs for time series forecasting [17,18,23] however, in this study, prediction models were used to predict GWLs from other input variables.

Table 2. Correlation analysis of input variable selection plotted/regressed versus actual GWLs.

| Watershed	Number of Variables	Variables	R^2
Baltic River	1	Stream level	0.51
		Stream flow	0.28
	2	Stream level and stream flow	0.63
		Stream level and precipitation	0.51
	3	Stream level, stream flow and evapotranspiration	0.66
		Stream level, stream flow and Mean temperature	0.64
	4	Stream level, stream flow, heat degree days and evapotranspiration	0.66
		Stream level, stream flow, mean temperature and evapotranspiration	0.66
Long Creek	1	Stream level	0.49
		Stream flow	0.36
	2	Stream level and Dew point temperature	0.55
		Stream level and dew heat degree days	0.55
	3	Stream level, mean temperature and evapotranspiration	0.57
		Stream level, stream flow and heat degree days	0.57
	4	Stream level, dew point temperature, relative humidity, evapotranspiration	0.59
		Stream level, relative humidity, mean temperature and evapotranspiration	0.58

3.3. The 1-Input Variable Model

For the 1-input variable deep learning model, the last epoch training and validation losses of the MLP for the Baltic River watershed were recorded to be 0.0839 and 0.0818, respectively. A similar training loss of the MLP for the Long Creek watershed was recorded, i.e., 0.0856. However, the validation loss of the MLP for the Long Creek watershed was 25% higher than the Baltic River watershed (Table 3). The training and validation losses in the LSTM model for the Baltic River watershed were not substantially different from the MLP and were recorded to be, respectively, 0.0832 and 0.074. A trend of a higher validation loss in the LSTM model, similar to that of the MLP for the Long Creek watershed, was observed, i.e., the validation loss was 23% higher than the training loss. Both the MLP and LSTM models performed poorly in the validation stage for the Long Creek watershed. However, for both watersheds, the training and validation losses of the 1D CNN were lower than those of the LSTM and MLP models. For the Baltic River watershed, the training and validation loss of the CNN was 0.01. The training and validation losses of the CNN for the Long Creek watershed were recorded to be 0.0125 and 0.0174, respectively. Figure 5 shows the training and validation phases of the deep learning models for the Baltic River and Long Creek watersheds. For both watersheds, the deep learning models were unable to predict the lower peaks very well. Variation in estimations of peaks were found to be higher in the training phase compared with that in the validation phases of all ANNs.

Table 3. Training and validation losses, and the root mean square error (RMSE) of the artificial neural networks (MLP, LSTM, and CNN) used in this study for the Baltic River and Long Creek watersheds using 1-, 2-, 3-, and 4-input variables.

Watersheds	No. of Variables	Method	Training Loss	Validation Loss	Training RMSE	Validation RMSE	R^2
Baltic River		MLP	0.084	0.082	0.567	0.558	0.65
		LSTM	0.083	0.074	0.557	0.530	0.66
		CNN	0.010	0.010	0.556	0.550	0.66
	1						
Long Creek		MLP	0.086	0.110	0.957	1.169	0.63
		LSTM	0.085	0.110	0.973	1.173	0.62
		CNN	0.013	0.017	0.948	1.180	0.63

Table 3. *Cont.*

Watersheds	No. of Variables	Method	Training Loss	Validation Loss	Training RMSE	Validation RMSE	R^2
Baltic River	2	MLP	0.081	0.072	0.560	0.512	0.66
		LSTM	0.083	0.076	0.561	0.531	0.66
		CNN	0.011	0.010	0.549	0.549	0.66
Long Creek		MLP	0.081	0.107	0.929	1.180	0.64
		LSTM	0.083	0.108	0.940	1.210	0.64
		CNN	0.011	0.018	0.898	1.215	0.65
Baltic River	3	MLP	0.079	0.065	0.545	0.474	0.69
		LSTM	0.079	0.067	0.539	0.483	0.69
		CNN	0.010	0.009	0.503	0.529	0.71
Long Creek		MLP	0.082	0.107	0.945	1.160	0.64
		LSTM	0.083	0.110	0.943	1.203	0.63
		CNN	0.011	0.018	0.888	1.209	0.66
Baltic River	4	MLP	0.079	0.064	0.543	0.471	0.69
		LSTM	0.077	0.066	0.534	0.480	0.69
		CNN	0.010	0.009	0.505	0.532	0.71
Long Creek		MLP	0.078	0.103	0.912	1.150	0.66
		LSTM	0.080	0.106	0.916	1.200	0.67
		CNN	0.009	0.017	0.813	1.170	0.70
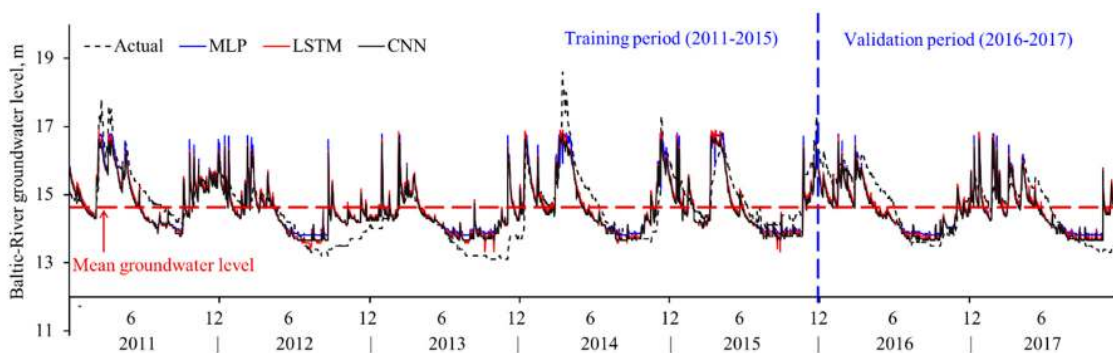
(a)
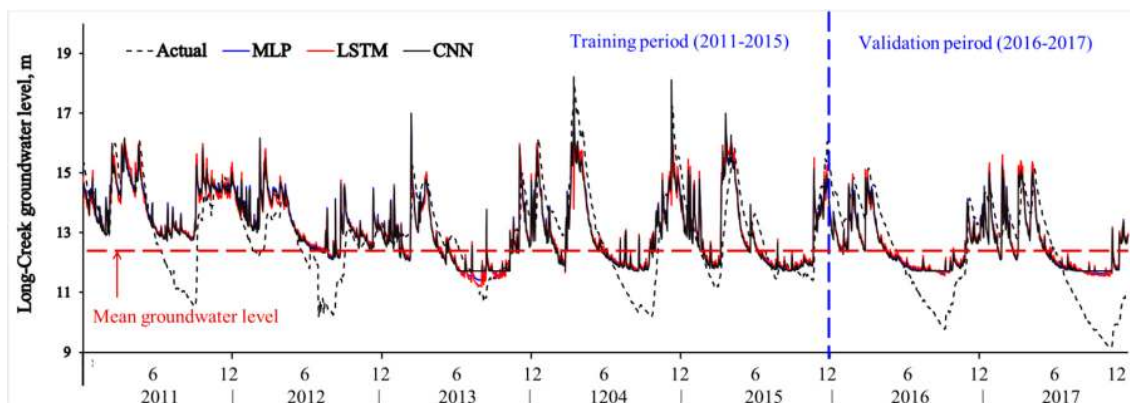
(b)

Figure 5. Training and validatdation phase of the 1-input variable models for (**a**) Baltic River watershed and (**b**) Long Creek watershed.

3.4. The 2-Input Variable Models

The 2-input variable MLP model for the Baltic River watershed showed training and validation losses of 0.0814 and 0.0723, respectively. These losses were slightly lower than the losses of the 1-input variable MLP models. Similarly, the 2-input variable MLP models training and validation losses for the Long Creek watershed were 0.0816 and 0.1068, respectively. The 2-input variable LSTM training losses for the Baltic River watershed improved by 0.8%, and the validation loss decreased by 0.2% in comparison with the 1-input variable models. In comparison with the training losses of the 1-input variable CNN, those of the 2-input variable CNN slightly increased from 0.0100 to 0.0113 and from 0.0100 to 0.0174, with no change in validation loss, for the Baltic River watershed. For the Long Creek watershed, the training loss was reduced from 0.0125 to 0.0112, and the validation loss increased from 0.0174 to 0.0184, in comparison with the 1-input variable CNN. The training and validation phases showed slight improvements in GWL estimation. However, no major difference was observed in GWL predictions in the 1-input and 2-inputs variable combinations, i.e., the lower peaks of the estimated GWLs did not match the lower peaks of the actual GWLs.

3.5. The 3-Input Variable Models

The 3-input variable MLP model for the Baltic River watershed showed slight improvements in training loss (0.079) and validation loss (0.0648) compared with the 2-input variable MLP model. However, the training loss of the 3-input variable MLP for the Long Creek watershed was lower, i.e., 0.0816, and the validation loss (0.1067) was slightly improved, compared with the 2-input variable MLP model. The 3-input variable LSTM model for the Baltic River watershed showed a higher validation loss (0.0673) as well as a training loss (0.079) that was similar to that of the 3-input variable MLP model. The 3-input variable LSTM model for the Long Creek watershed had a higher training loss (0.0832) and a higher validation loss (0.1095) compared with the 2-input variable LSTM model for the same watershed. The 3-input CNN model for the Baltic River watershed showed the lowest training loss (0.0101) and validation loss (0.0093) compared with all 3-input variable models. Similarly, the 3-input variable CNN model for the Long Creek watershed showed the lowest training loss (0.011) and validation loss (0.0182) compared with all 3-input variable models. There was close agreement between the actual and estimated GWLs for the Baltic River watershed. However, there was weak agreement between the lower peaks of the actual and estimated GWLs of the Long Creek watershed.

3.6. The 4-Input Variable Models

The 4-input variable MLP model for the Baltic River watershed showed no difference in training loss (0.079) and validation loss (0.0644) compared with the 3-input variable MLP model. Similarly, no major difference was observed in training and validation losses for the Long Creek watershed compared with the 4-input MLP model; they were found to be 0.0783 and 0.1031, respectively. For the 4-input variable LSTM model, the training and validation losses for the Baltic River watershed were recorded as 0.077 and 0.0663, respectively. The 4-input variable LSTM model for the Long Creek watershed showed a slight improvement in training (0.0798) and validation (0.1057) losses compared with the other LSTM models. The 4-input variable CNN models showed the lowest training and validation losses for both the Baltic River and Long Creek watersheds compared with all other 4-input variable models. The addition of a fourth variable reduced the difference between the lower peaks of the estimated and actual GWLs as observed in the case of the 1-input and 2-input combinations (Figure 6). A close agreement between the actual and the estimated GWLs for the Baltic River watershed was observed, as in the output of the 3-input variable model, and a comparatively weak agreement was observed between the lower peaks of the actual and the estimated GWLs of the Long Creek watershed.
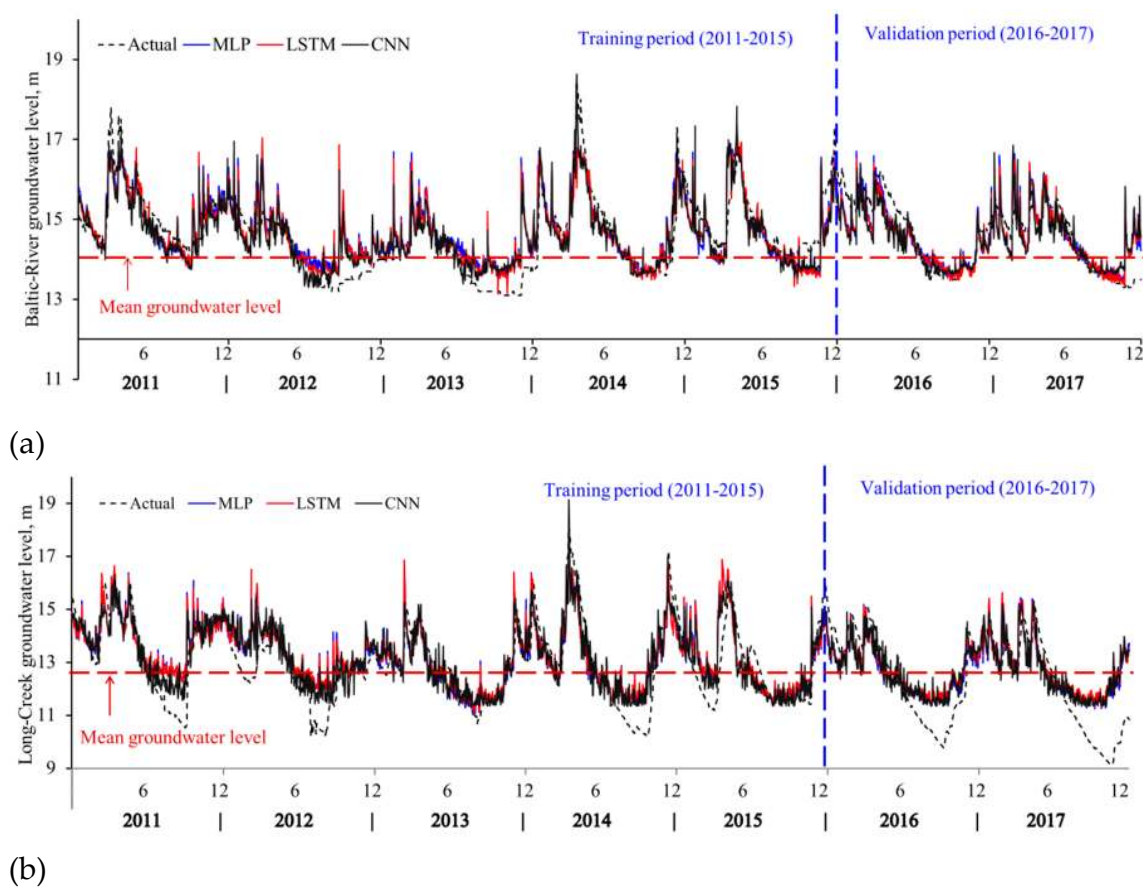
(a)

(b)

Figure 6. Training and validation phases of the 4-input variable models for (**a**) Baltic River watershed and (**b**) Long Creek watershed.

3.7. Model Evaluation

The RMSE of the 1-input variable model for the Baltic River watershed was lower than the Long Creek watershed for all ANN models (Table 3). For the 1-input variable Baltic River watershed LSTM, the RMSE (0.53 m) was lower at the validation stage than both the MLP and the CNN. However, in the training stage of the 1-input variable model for the Baltic River watershed, the CNN performed better than the other models, with an RMSE of 0.556 m. For the 1-input variable models of the Long Creek watershed, higher validation losses, as compared to training losses, were recorded. The 1-input variable model of the MLP performed well in the validation stage, with a recorded RMSE of 1.169 m for the Long Creek watershed. At the training stage for the Long Creek watershed, the 1-input variable CNN performed well, with a slightly lower RMSE, i.e., 0.948 m. Similar trends were recorded for the 2-input variable ANNs, with a slight improvement in the recorded RMSE for both training and validation stages. For the Baltic River watershed, the 2-input variable MLP model performed well at the validation stage, with a slightly lower RMSE, i.e., 0.56 m. However, it was not much different than the 1-input variable model, as the RMSE was improved only by 0.046 m. Similarly, the CNN improved the training stage error of the 2-input variable models by 0.007 m. The slightly reduced RMSE indicates the better performance of the 2-input variable MLP and CNN models in comparison with the 1-input variable models for the Baltic River watershed. For the Long Creek validation stage, the 2-input variable models performed poorly compared with the 1-input variable models. The 2-input variable MLP performed better than the LSTM and CNN models, with a recorded RMSE of 1.18, which was higher in comparison with the 1-input variable MLP model. The 2-input variable CNN model performed better at the training stage for the Long Creek watershed, with a slight improvement in RMSE (0.898) compared with the 1-input variable CNN model. The 3-input variable MLP model

for the Baltic River watershed performed well, reducing the RMSE by 0.038 m in comparison with the 2-input variable MLP model. The CNN performed well at the training stage for the Baltic River watershed, with an improved RMSE of 0.066 m. For the Long Creek watershed, the 3-input variable MLP model performed well, with an improved RMSE of 0.02 m, as did the CNN, with an RMSE of 0.01 m in the validation and training stages, respectively. The smaller differences in RMSE indicate the slightly improved performance of the 3-input variable models over the 2-input variable models for the same site. The 4-input variable MLP model for the Baltic River watershed performed well at the validation stage, with an RMSE of 0.471, compared with all the models used in this study. Similarly, the 4-input variable MLP model for the Long Creek watershed achieved the lowest RMSE (1.15) in the validation stage. The reduction of the RMSE, achieved by increasing the number of variables, suggests that the selected variables slightly increased the performance and overall efficiency of the ANNs. In general, no major effect on model performance was observed by increasing the number of variables. The major contributor in defining the GWL variation was the stream level. The reason for this could be the inability of the remaining three variables to account for the variability in GWLs. Evidence of this trend is presented in Table 2. The increase in the number of variables does not show any major effects on the coefficient of determination.

The potential reason for the low performance of ANNs on the Long Creek watershed could be the absence of a lurking input variable, such as pumping data. There is no data available on actual daily pumping usage by high-capacity well owners. Most of the lower peaks were recorded mostly in the summer season for the Long Creek watershed, as depicted in Figures 5 and 6; wells might be pumped by well owners in drought season for supplemental irrigation to replenish crop water requirements. However, pumping problems or lower peaks were not a major problem for the Baltic-River watershed, whereas high-capacity wells in that area was.

It is noticeable that the MLP, being the simplest type of ANN, performed well at the validation stages of both watersheds compared with the more advanced ANNs, including the LSTM and CNN models. In a similar study of GWL estimation using ANNs, Mueller et al. [19] reported that the MLP outperformed both the LSTM and CNN models. At the training stage, the performance of the CNN was greater than that of the other ANNs. This could be due to the additional parameter-learning functionalities in the CNN, such as the convolutional and max-pooling filters, which were not the part of the MLP and LSTM models.

Furthermore, polynomial relationships between the predicted and actual groundwater level suggest that increasing the number of variables from 1 to 4 increased the R^2 from 0.62 to 0.70 for the Long Creek watershed (Figure 7). A similar trend was observed for the Baltic River watershed as the R^2 was raised from 0.66 to 0.71 for the 1-input and 4-input variable models, respectively (Figure 8). These results suggest that the stream level and stream flow variables should be considered in defining the overall variability of GWLs. These variables contribute to the role of stream level and stream flow in a watershed, as reported in the literature [26]. Recent study conducted by Lee et al. [27] found the similar relationship between GWLs and stream level. Zhang et al. [10] compared feedforward neural networks with LSTM in GWL prediction in the Hetao irrigation district, China. Contrary to our findings, their results suggest that the LSTM model is a better predictor of GWL modeling than the MLP. The CNN performed better with a higher number of input variables compared with the LSTM and MLP models. For example, for both watersheds, the MLP performed slightly better than the CNN for the 1- and 2-input variable models. However, the accuracy of the CNN was higher than that of the MLP in the 3-and 4-input variable models for both watersheds (Figures 7 and 8). The higher accuracy of the CNN with a certain number of variables may indicate the advantage of the CNN in dealing with more complex relationships in comparison with the MLP and LSTM models. The additional functionalities of convolutions and max pooling helped the CNN to obtain higher accuracies with a higher number of inputs.
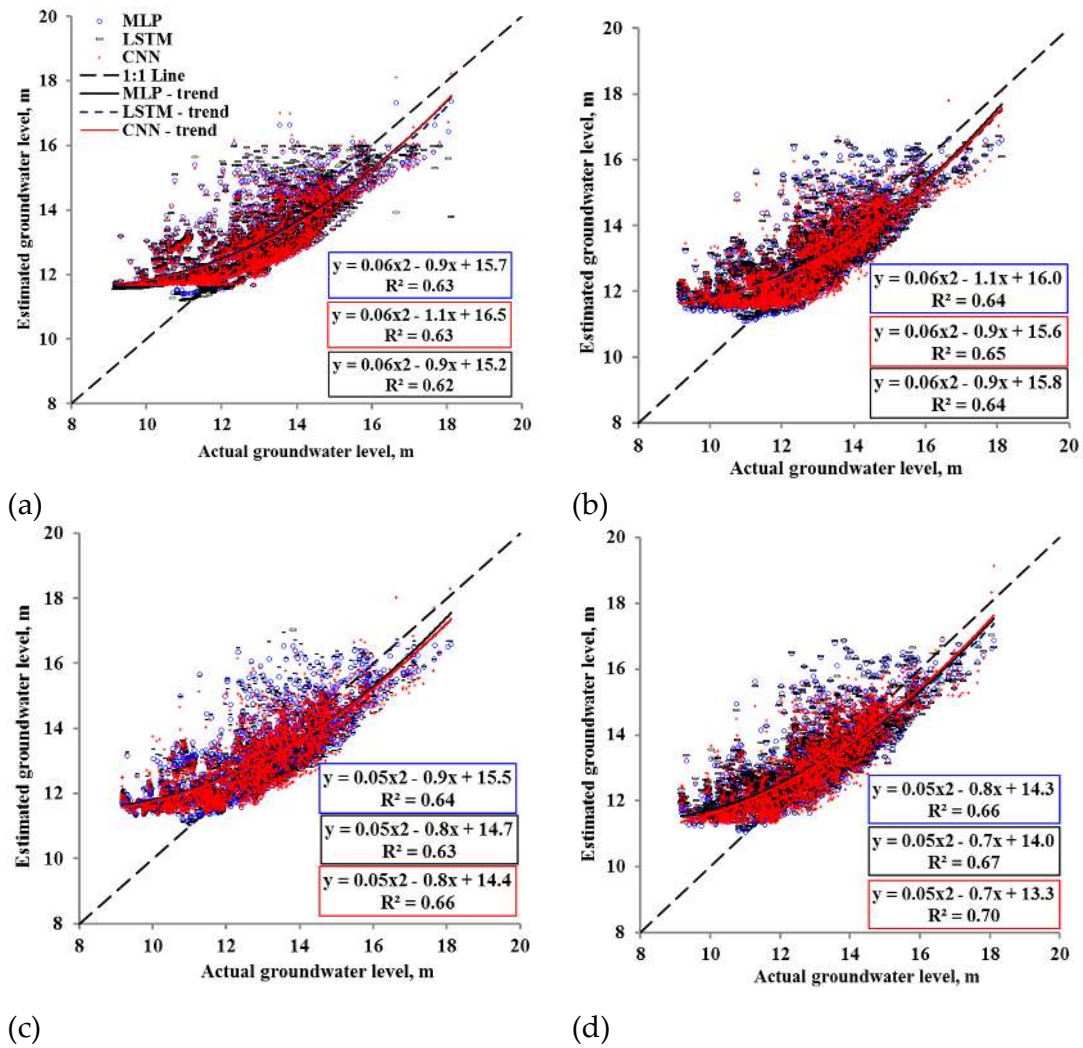
Figure 7. The Long Creek watershed estimated groundwater levels for (**a**) 1-input, (**b**) 2-input, (**c**) 3-input, and (**d**) 4-input variable models versus actual groundwater levels. Under-estimation and scattering is reflected in 1- and 2-variable models. Clustering and overlapping of the 1:1 line is seen for 3- and 4-variable models. Because of a large number of data points (7671; 2557 for each ANN), the symbols overlap and hide one another.
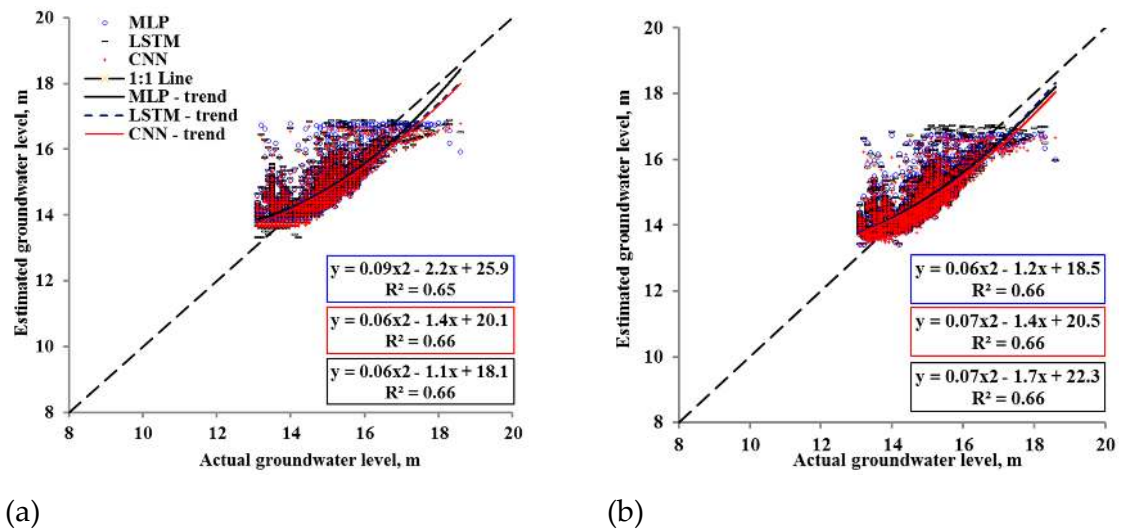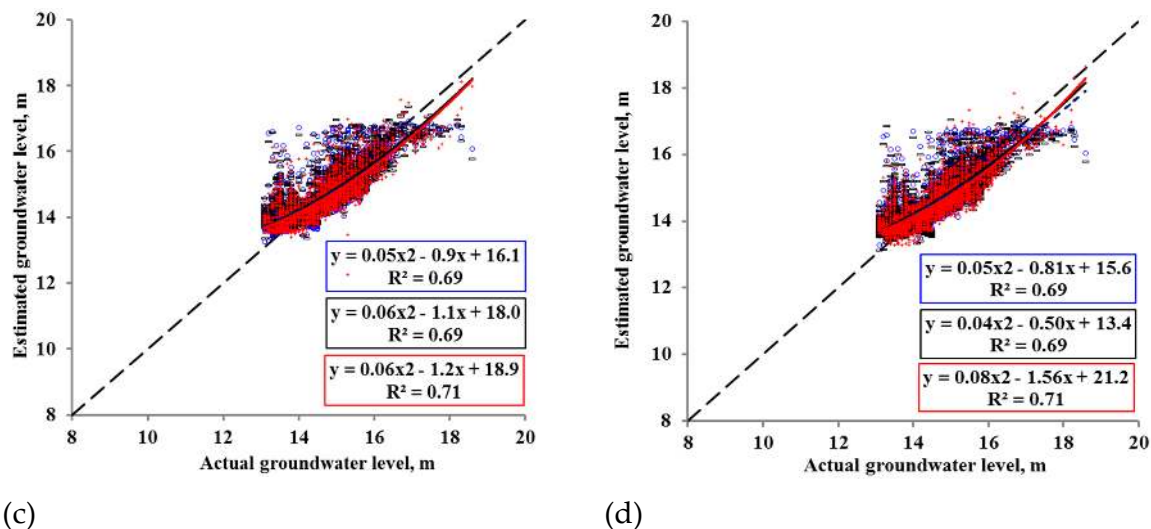

Figure 8. *Cont.*

(c)　　　　　　　　　　　　　　　　　　　　　　　　　　(d)

Figure 8. The Baltic River estimated groundwater levels for (**a**) 1-input, (**b**) 2-input, (**c**) 3-input, and (**d**) 4-input variable models versus actual groundwater levels, showing under-estimated and scattered data for the 1- and 2-variable models and a 1:1 line overlapping clustered data for the 3- and 4-variable models, respectively. Because of a large number of data points (7671; 2557 for each ANN), the symbols overlap and hide one another.

However, the over-estimation of GWLs can be visualized for the Long Creek watershed for approximately below 11 m, and under-estimation can be visualized beyond the 13 m level. For the Baltic River watershed, the over- and under-estimations were approximately below 14 and beyond 15 m, respectively. The possible reason for the poorly matching peaks of the extreme GWLs may be due to short-time variations in the GWLs because of the pumping during water-deficient months and the quick recharge of groundwater from the higher level of groundwater neighboring the monitoring well. This means that there is a need to further investigate how the extreme GWLs can be estimated, especially for the lower peaks. Mismatch between the predicted and actual levels of groundwater could be due to several hydrogeological reasons, including the fact that groundwater pumping can also draw streamflow into connected aquifers, where pumping rates are relatively large or where the locations of pumping stations are relatively close to a stream [4]. The ANN models used in this study were unable to predict the lower groundwater peaks. This drawback can be addressed by adding more variables, such as actual pumping withdrawal, to develop more accurate models. In general, ANNs can predict human interventions if modeled with proper inputs. However, in this study, the groundwater pumping data were not available. Emphasis on the collection and storing of the pumping data is therefore recommended so that provincial water management authorities can track this important variable for groundwater sustainability on the island.

The application of these models can be further extended to other watersheds to estimate GWLs, as there are more than 250 watersheds in Prince Edward Island. The installation and maintenance of a large number of monitoring wells is neither feasible nor economical. These modeling techniques provide economical and convenient methods of equitable water distribution for water resource managers and policymakers in large areas.

4. Conclusions

In this study, deep learning has been used to test different ANN models with several input combinations to predict GWLs in the Baltic River and Long Creek watersheds in Prince Edward Island. The results suggest that ANNs can be used to predict GWLs in Prince Edward Island. The stream level was the most highly correlated factor in defining GWL fluctuations for both watersheds. All ANN models performed well on the Baltic River watershed in comparison with the Long Creek watershed. The number of variables had no major effect on the RMSE of either watershed. The results suggest that

the GWLs of the Baltic River and Long Creek watersheds can be efficiently modeled with the stream level variable in the absence of GWL data. Similarly, for the Long Creek watershed, the GWLs can be modeled with stream levels with a slightly high RMSE as compared to the Baltic River watershed. The performance of the MLP at the validation stage was slightly greater than that of the MLP and LSTM models. However, at training stages, the CNN performed very well, with the lowest RMSE for both watersheds. The ANN models were unable to predict the lower peaks in the summer season, specifically for the Long Creek watershed, which could be due to lurking variables such as the pumping from high-capacity irrigation wells. Because of the non-availability of daily pumping data in the island, this critical input variable was not included in this study. However, it is recommended that provincial water resource management authorities monitor pumping data on a daily basis. It is also notable that the increased number of variables from 1 to 4 improved the RMSE for the Baltic River watershed by 11% only and for the Long Creek watershed by 1.6% only. It may be concluded that the GWLs in the Baltic River and Long Creek watersheds can be modeled using stream levels with RMSEs of only 0.53 and 1.169 m, respectively. The results also suggest that the CNN performed better with a higher number of input variables; however, the performance of the MLP was better with lower input variable models. Since the physical measurement of GWLs is challenging with respect to the maintenance of groundwater wells and the collection of periodic dips, the techniques introduced in this study are convenient and accurate. Stream flow gauges are easier to manage than physical measurement of GWLs. In areas where groundwater pumping is common for supplemental irrigation or for domestic use, the inventory control of groundwater resources could become more convenient with the technique used in this study.

References

1. Jiang, Y.; Somers, G.; Mutch, J. Application of Numerical Modeling To Groundwater Assessment and Management in Prince Edward Island. In Proceedings of the 57th Canadian Geotechnical Conference, Quebec, QC, Canada, 24–26 October 2004; Volume 57, pp. 2–9.
2. Department of Communities, Land and Environment. Watershed Management on PEI|Prince Edward Island. Available online: https://www.princeedwardisland.ca/en/information/communities-land-and-environment/watershed-management-pei (accessed on 22 November 2019).
3. Gao, Z.; Long, D.; Tang, G.; Zeng, C.; Huang, J.; Hong, Y. Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin. *J. Hydrol.* **2017**, *550*, 478–496. [CrossRef]
4. USGS. How does Groundwater Pumping Affect Streamflow? Available online: https://www.sciencedaily.com/releases/2012/11/121116124557.htm (accessed on 12 December 2019).
5. Francis, R.M. Hydrogeology of the winter river basin-Price Edward Island. Department of the Environment, PEI report; 1989. Available online: http://www.gov.pe.ca/photos/original/cle_WinterR.pdf (accessed on 22 November 2019).

6. Mohammadi, K. Groundwater Table Estimation Using MODFLOW and Artificial Neural Networks. In *Practical Hydroinformatics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 127–138.

7. Rajaee, T.; Ebrahimi, H.; Nourani, V. A review of the artificial intelligence methods in groundwater level modeling. *J. Hydrol.* **2019**, *572*, 336–351. [CrossRef]

8. Mohanty, S.; Jha, M.K.; Kumar, A.; Panda, D.K. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi–Surua Inter-basin of Odisha, India. *J. Hydrol.* **2013**, *495*, 38–51. [CrossRef]

9. Karandish, F.; Šimůnek, J. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *J. Hydrol.* **2016**, *543*, 892–909. [CrossRef]

10. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2018**, *561*, 918–929. [CrossRef]

11. Zhang, N.; Shen, S.L.; Zhou, A.; Xu, Y.S. Investigation on Performance of Neural Networks Using Quadratic Relative Error Cost Function. *IEEE Access* **2019**, *7*, 106642–106652. [CrossRef]

12. Sarir, P.; Shen, S.L.; Wang, Z.F.; Chen, J.; Horpibulsuk, S.; Pham, B.T. Optimum model for bearing capacity of concrete-steel columns with AI technology via incorporating the algorithms of IWO and ABC. *Eng. Comput.* **2019**, 1–11. [CrossRef]

13. Atangana Njock, P.G.; Shen, S.L.; Zhou, A.; Lyu, H.M. Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/t-SNE model. *Soil Dyn. Earthq. Eng.* **2020**, *130*, 105988. [CrossRef]

14. Elbaz, K.; Shen, S.-L.; Zhou, A.; Yuan, D.-J.; Xu, Y.-S. Optimization of EPB Shield Performance with Adaptive Neuro-Fuzzy Inference System and Genetic Algorithm. *Appl. Sci.* **2019**, *9*, 780. [CrossRef]

15. Sahoo, S.; Jha, M.K. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: A comparative assessment Sasmita. *Hydrogeol. J.* **2013**, *21*, 1865–1887. [CrossRef]

16. Kouziokas, G.N.; Chatzigeorgiou, A.; Perakis, K. Multilayer Feed Forward Models in Groundwater Level Forecasting Using Meteorological Data in Public Management. *Water Resour. Manag.* **2018**, *32*, 5041–5052. [CrossRef]

17. Juan, C.; Genxu, W.; Tianxu, M. Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. *J. Hydrol.* **2015**, *529*, 1211–1220.

18. Coulibaly, P.; Anctil, F.; Aravena, R.; Bobée, B. Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* **2001**, *37*, 885–896. [CrossRef]

19. Mueller, J.; Park, J.; Sahu, R.; Varadharajan, C.; Arora, B.; Faybishenko, B.; Agarwal, D. Surrogate Optimization of Deep Neural Networks for Groundwater Predictions. *arXiv* **2019**, preprint. arXiv:1908.10947.

20. Maheshwara Babu, B.; Srinivasa Reddy, G.; Satishkumar, U.; Kulkarni, P. Simulation of Groundwater Level Using Recurrent Neural Network (RNN) in Raichur District, Karnataka, India. *Int. J. Curr. Microbiol. Appl. Sci.* **2018**, *7*, 3358–3367.

21. Lähivaara, T.; Malehmir, A.; Pasanen, A.; Kärkkäinen, L.; Huttunen, J.M.J.; Hesthaven, J.S. Estimation of groundwater storage from seismic data using deep learning. *Geophys. Prospect.* **2019**, *67*, 2115–2126. [CrossRef]

22. Ravansalar, M.; Rajaee, T. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environ. Monit. Assess.* **2015**, *187*, 366. [CrossRef]

23. Krishna, B.; Satyaji Rao, Y.R.; Vijaya, T. Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. *Hydrol. Process.* **2008**, *22*, 1180–1188. [CrossRef]

24. Tsanis, I.K.; Coulibaly, P.; Daliakopoulos, I.N. Improving groundwater level forecasting with a feedforward neural network and linearly regressed projected precipitation. *J. Hydroinformatics* **2008**, *10*, 317–330. [CrossRef]

25. Gholami, V.; Chau, K.W.; Fadaee, F.; Torkaman, J.; Ghaffari, A. Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers. *J. Hydrol.* **2015**, *529*, 1060–1069. [CrossRef]

26. Safeeq, M.; Fares, A. Groundwater and Surface Water Interactions in Relation to Natural and Anthropogenic Environmental Changes. In *Emerging Issues in Groundwater Resources*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 289–326. [CrossRef]

27. Lee, S.; Lee, K.K.; Yoon, H. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. *Hydrogeol. J.* **2019**, *27*, 567–579. [CrossRef]