Article

A Novel Stacked Long Short-Term Memory Approach of Deep Learning for Streamflow Simulation

Majid Mirzaei ^{1,*}, Haoxuan Yu ² , Adnan Dehghani ¹, Hadi Galavi ³ , Vahid Shokri ¹, Sahar Mohsenzadeh Karimi ^{1,4} and Mehdi Sookhak ⁵

¹ Department of Civil Engineering, Faculty of Engineering, University of Malaya (UM), Kuala Lumpur 50603, Malaysia; adnan.dehghani@gmail.com (A.D.); vahidshokri93@yahoo.com (V.S.); sma229@sfu.ca (S.M.K.)

² School of Resources and Safety Engineering, Central South University, Changsha 410083, China; yuhaoxuan@csu.edu.cn

³ Department of Water Science and Engineering, University of Zabol, Zabol 98617, Iran; hadigalavi@uoz.ac.ir

⁴ Department of Geography, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

⁵ Department of Computer Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA; m.sookhak@ieee.org

* Correspondence: majid@um.edu.my

Citation: Mirzaei, M.; Yu, H.; Dehghani, A.; Galavi, H.; Shokri, V.; Mohsenzadeh Karimi, S.; Sookhak, M. A Novel Stacked Long Short-Term Memory Approach of Deep Learning for Streamflow Simulation. *Sustainability* **2021**, *13*, 13384. <https://doi.org/10.3390/su132313384>

Academic Editors: Chow Ming Fai, Ahmed Hussein Kamel Ahmed Elshafie and Al Mahfoodh Ali Najah Ahmed

Received: 3 October 2021

Accepted: 29 November 2021

Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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/>).

Abstract: Rainfall-Runoff simulation is the backbone of all hydrological and climate change studies. This study proposes a novel stochastic model for daily rainfall-runoff simulation called Stacked Long Short-Term Memory (SLSTM) relying on machine learning technology. The SLSTM model utilizes only the rainfall-runoff data in its modelling approach and the hydrology system is deemed a blackbox. Conversely, the distributed and physically-based hydrological models, e.g., SWAT (Soil and Water Assessment Tool) preserve the physical aspect of hydrological variables and their inter-relations while taking a wide range of data. The two model types provide specific applications that interest modelers, who can apply them according to their project specification and objectives. However, sparse distribution of point-data may hinder physical models' performance, which may not be the case in data-driven models. This study proposes a specific SLSTM model and investigates the SLSTM and SWAT models' data dependency in terms of their spatial distribution. The study was conducted in the two distinct river basins of Samarahan and Trusan, Malaysia, with over 20 years of hydro-climate data. The Trusan basin's rain gauges are scattered downstream of the basin outlet and Samarahan's are located around the basin, with one station within each basin's limits. The SWAT was developed and calibrated following its general modelling approach, however, the SLSTM performance was also tested using data preprocessing with principal component analysis (PCA). Results showed that the SWAT performance for daily streamflow simulation at Samarahan has been superior to that of Trusan. Both the SLSTM and PCA-SLSTM models, however, showed better performance at Trusan with PCA-SLSTM outperforming the SLSTM. This demonstrates that the SWAT model is greatly affected by the spatial distribution of its input data, while data-driven models, irrespective of the spatial distribution of their entry data, can perform well if the data adequacy condition is met. However, considering the structural difference between the two models, each has its specific application in a water resources context. The study of catchments' response to changes in the hydrology cycle requires a physically-based model like SWAT with proper spatial and temporal distribution of its entry data. However, the study of a specific phenomenon without considering the underlying processes can be done using data-driven models like SLSTM, where improper spatial distribution of data cannot be a restricting factor.

Keywords: machine learning; hydrological modelling; stochastic models; SLSTM; SWAT

1. Introduction

River flow estimation is an important factor in designing water infrastructures built for applications such as flood control, urban water supply [1,2], and irrigation network

design [3]. River flow is influenced by spatial and temporal variations of parameters such as temperature, precipitation, land use and land cover (LULC), and man-made structures on the river network within a catchment. Hydrological models fall into three main categories of physically-based, conceptual, and data-driven models. The physically-based models represent the physical description of the hydrological processes governing the basin's responses in the hydrology cycle. Conceptual models are based on the empirically observed relationships between different hydrological parameters. The data-driven models are built on the system state variables, such as input and output, and do not require in-depth knowledge of hydrological processes [4,5].

Many distributed and semi-distributed models that have been developed over the years have gained popularity for their performance and capabilities. A few examples of such models are the MIKE SHE [6,7], TOPMODEL [8] and Soil and Water Assessment Tool (SWAT) [9,10]. The over-parametrization in such models entails uncertainties that can be quantified through uncertainty analysis techniques to facilitate their application in real world cases [11,12]. The SWAT model, among many hydrological models, has proven to be an efficient tool for many applications including, but not limited to, water quality simulation, pollutant loading estimates, conservation practices efficacy, source load allocation determinations, and streamflow simulation [13,14]. It is a physically-based, semi-distributed hydrological model capable of operating at different time scales [15]. The advancements made in physical and semi-distributed modeling features of the SWAT to simulate the hydrology, sediment and nutrient load transfer in a catchment has made it a comprehensive basin scale model suitable for catchment scale modeling with various degrees of data availability [16].

Due to the adverse impacts of data unavailability and insufficiency on water resources management practices, accurate estimation of the streamflow, especially at sites with missing or limited hydrological data, is an imperative for integrated water resources management. The physical and conceptual models are highly data-dependent and require proper temporal and spatial distribution of the data for proper performance. The data-driven models, on the other hand, do not require extensive knowledge of the case study and are well suited for sparsely gauged sites with limited data availability [17,18]. Application of data-driven models is relatively straightforward, and with proper pre-processing of the available data, accurate modelling of the hydrological variables is possible. Therefore, application of data-driven machine learning (ML) methods for streamflow forecasting such as neural networks (NNs), support vector machines (SVMs), fuzzy inference systems, and wavelet transform (WT) have received a great deal of attention in recent decades [19–21]. However, it should be highlighted that the two types of model, physically-based and data-driven, are different application-wise, and cannot be used interchangeably for the same project objectives and setting. Data-driven ones are only suited for estimation of output of a hydrologic system without studying its physical aspects, whereas the distributed ones need the underlying processes to be well defined in the model structure in order to study their interactions.

Among the different machine learning techniques, the ANN as a self-adaptive yet self-learning function approximator has exhibited remarkable aptitude for modeling non-linear hydrologic datasets [22]. Despite widespread application of ANNs, over-fitting, convergence to local minima and inability to capture long term dependency in time series data are their known drawbacks [23,24]. These drawbacks impede achievement of satisfactory model performance, especially when dealing with hydrological time series. To capture the long-term and short-term dependency in data, a new generation of recurrent neural networks (RNNs) with the ability to learn order dependence in time series data was developed. In this new network, information flows within different layers of network and between neurons in each layer, so it can analyse the information in different time steps. To overcome vanishing and exploding gradient problems of RNN training, combined with the aforementioned advances, the long short-term memory (LSTM) network was introduced by Hochreiter and Schmidhuber [25]. The LSTM models, unlike standard neural

networks, can capture the periodic and/or chaotic behaviour of time series and learn their long-term relationships with higher accuracy [26]. Kratzert [27] effectively characterised the rainfall-runoff behaviour of a large number of complex catchments on a daily basis using the LSTM model. Ni [28] built three hybrid models based on the traditional LSTM network for monthly rainfall and streamflow forecasts; their findings showed that the LSTM model is a highly capable model for time series forecasting. The LSTM model has been proven effective in learning long-term relationships between sequential data sets and performs well in flood forecasting [29,30].

With the above background, this paper applies a novel machine learning-based approach called Stacked-LSTM (SLSTM) and an improved version of it utilising principal component analysis (PCA) for data pre-processing (PCA-SLSTM) for daily streamflow simulation. The Stacked-LSTM model with the ability to extract complex data patterns from time-series through multiple layers of LSTM layers coupled with ANN layers is believed to outperform the traditional LSTM model in streamflow forecasting. The semi-distributed SWAT hydrological model is also developed and calibrated for streamflow simulation in the case studies. The two models' dependency on spatial distribution of rain and temperature data stations are also investigated. The case studies are the Samarahan and the Trusan river basins in Malaysia, with Trusan having improper data station distribution in and around the basin.

2. Methodology

2.1. Study Area

The two river basins selected for the present study are the Samarahan and Trusan river basins in East Malaysia. The Samarahan river basin lies within latitudes of $1^{\circ}18' \text{ E}$ and $110^{\circ}24' \text{ E}$ and longitudes of $5^{\circ}12' \text{ N}$ and $116^{\circ}54' \text{ N}$; the Trusan river basin's geographic location is within latitudes of $1^{\circ}18' \text{ E}$ and $110^{\circ}24' \text{ E}$ and longitudes of $5^{\circ}12' \text{ N}$ and $116^{\circ}54' \text{ N}$. The drainage areas of Samarahan and Trusan are 54 and 6322 km^2 , respectively. Various types of data have been collected in this study, including metrological, land use, soil, and digital elevation model (DEM), used in each model as needed. Figure 1 demonstrates basins and the location of each station corresponding to their details given in Table 1. The Trusan, with a larger area, only has two synoptic stations within its boundary and the rest are located downstream of the catchment outlet. The Samarahan, with a smaller drainage area, has synoptic stations distributed around the basin.

Table 1. Rainfall and streamflow gauge stations used in the study.

	Station	Type	Name	Longitude	Latitude
Trusan	Ulu Kuamut	Rainfall	RF1	117.44	5.08
	Tongod	Rainfall	RF2	116.97	5.27
	Kuamut Met	Rainfall	RF3	117.49	5.22
	Balat	Rainfall	RF4	117.6	5.31
	Tangkulap	Rainfall	RF5	117.28	5.3
	Bilit	Rainfall	RF6	118.19	5.49
	Sukau	Rainfall	RF7	118.28	5.53
	Milian	StreamFlow	SF1	117.32	5.3
Samarahan	Gayu	Rainfall	RF1	110.34	1.22
	Plaman Nyabet	Rainfall	RF2	110.44	1.21
	Dragon School	Rainfall	RF3	110.42	1.28
	Semongok	Rainfall	RF4	110.32	1.39
	Samarahan Estate	Rainfall	RF5	110.55	1.39
	Paya Paloh	Rainfall	RF6	110.49	1.34
	Baru	Rainfall	RF7	110.5	1.44

Table 1. Cont.

Station	Type	Name	Longitude	Latitude
Ketup	Rainfall	RF8	110.53	1.49
Semera	Rainfall	RF9	110.67	1.55
Asa Jaya	Rainfall	RF10	110.61	1.55
Similang	Rainfall	RF11	110.5	1.61
Batu Gong	StreamFlow	SF1	110.44	1.35

Figure 1. (a) Samarahan and (b) Trusan river basins.

2.2. Distributed Modelling—The SWAT Hydrological Model

The Soil and Water Assessment Tool (SWAT) model, used here for hydrological modelling of the two catchments, is a well-known semi-distributed basin scale model [31]. The SWAT as a comprehensive model simulates hydrology, chemicals, sediments, crop growth, agricultural management, and other phenomena at basin scale. There are three key modelling components in SWAT development: sub-basin delineation, reservoir routing, and channel routing. Regions with similar land use, slope, and soil types within a subbasin create similar hydrological response units (HRUs) building the basis of SWAT model hydrological processing structure. A tributary waterway and a main waterway may be present in the subbasin. Users have the option of editing inputs at the watershed, subbasin, or HRU level, as well as adding point sources. The variable storage approach and the Muskingum technique are the two options available in SWAT for channel routing. The DEM, meteorological inputs (precipitation, wind speed, temperature, humidity, and solar radiation,) besides the soil and land use maps are all required to run the SWAT model. Streamflow observations in the case of this study would be used for calibration and validation of the model. The SWAT model would be run in ArcSWAT, an ArcGIS extension, and the modelling workflow is shown in Figure 2.

Figure 2. Structure of the physically-based model in this study.

As in Figure 2, to calibrate the SWAT model, the Sequential Uncertainty Fitting version 2 (SUFI-2) has been used. It has a Bayesian framework with a stochastic algorithmic approach that can define the probability function of distributed models' parameters. SUFI-2 is most frequently used for uncertainty analysis, calibration of models, and validation of the SWAT model. The SUFI-2 method, by inverse modelling, calibrates and defines the distributed models' parameter uncertainty. The p -factor and r -factor indices are used to describe the calibration/uncertainty analysis outcome. The p -factor is counted based on the number of observations located within the 95 Percent Prediction Uncertainty (95PPU) bracket. The p -factor varies between 0 and 100 percent with 100% representing the best fit to the observed data.

R-factor is defined as the average thickness of the 95PPU divided by the standard deviation of the observed data, and it ranges from 0 to positive infinity. The model simulation exactly corresponds to observation data if the r -factor approaches zero. The SWAT model's performance is evaluated by consulting the coefficient of determination (R^2) and Nash-Sutcliffe Efficiency (NSE) measures for the calibration and validation processes.

2.3. Data-Driven Models—The Long Short-Term Memory Model (LSTM)

The LSTM architecture is built upon a Recurrent Neural Network (RNN) basis by Hochreiter and Schmidhuber [25] in order to take advantages of RNN capabilities, such as the ability to model sequential data, and solve some of its drawbacks, such as the exploding

and vanishing gradient issue. The Building blocks of LSTM architecture are its memory cells (a LSTM memory cell is illustrated in Figure 3) which despite simple neurons in neural networks, is able to learn and forget irrelevant parts of previous states, store relevant new information in the cell, update cell states selectively, and at the end control the information passed on. The physical intuition of LSTM architecture for runoff modelling allows it to retain the sequential nature of flow dependency to different rainfall stations by a series of LSTM units and time lags. It records the relation of different rainfall stations with runoff, and enables it to forget in each memory cell structure. Different LSTM-based models have been implemented recently for runoff modelling and forecasting [27,30]. The LSTM memory cell transition equations are written below.

Figure 3. LSTM memory cell architecture.

The LSTM memory cell takes in various information through three specific gates; in the first gate, known as the forget gate, it is decided which element of C_{t-1} must be forgotten by multiplying it with f_t , ranging from 0 to 1. In additions, W_f and b_f are parameters within the forget gate that can be trained, and x_t is the current input to the cell while h_{t-1} is the previous cell's state in the layer:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

The next gate is called the input gate, and here it is decided which value should be updated:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

where i_t is an output variable with value ranging from 0 to 1. W_i and b_i are parameters within the input gate that can be trained. Then, a potential vector of the cell state is computed by the current input (x_t) and the last hidden state h_{t-1} in the following equation:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (3)$$

$\tilde{C}_t$ is a vector with values ranging from 0 to 1, and W_C and b_C are parameters within the input gate that can be trained during the training process.

Then, C_t is calculated by the following equation:

$$C_t = f_t * C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

At the end, by applying a sigmoid activation function, the output is calculated in the output gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

where o_t is a vector with values ranging from 0 to 1. W_o and b_o are trainable parameters. The new hidden state h_t is calculated by element-wise multiplication of the output by the hyperbolic tangent of C_t as follows.

$$h_t = o_t \times \tanh(C_t). \quad (6)$$

2.4. SLSTM Setup

The proposed deep neural network architecture consists of two LSTM layers combined with two fully connected dense layers for detecting the sequential nature of the problem, and a dropout layer with a dropout rate of 0.4 on the first fully connected layer for preventing overfitting; in addition the L2 normalization method was utilised to reduce the likelihood of the model overfitting by keeping the values of the weights and biases small. Since streamflow has an autocorrelated nature, input variables are turned to sequences to capture the impact of autocorrelation. The neural network structures used for both basins are identical, the only difference being the number of hidden neurons in each layer. The first two LSTM layers and the dense middle layer in the Samarah basin had 200, 200 and 150 hidden neurons respectively, while in the Trusan basin all three layers each had 100 hidden neurons; the last dense layer for both basins was the same, with one neuron representing runoff output. In both basins, Principal Component Analysis transform was used to find normally independent variables, and simple normalized datasets were used to study the impact of multicollinearity between variables in runoff modelling too. In the Samarah basin the study period was from 2003 to 2008 with 80% of the data being used for training and 20% for validation. In the Trusan basin, the available datasets of 2003–2007 were also divided with the same fractions for training and validation.

In Figure 4, the proposed sequential machine learning, coupled with a fully connected dense layer, is depicted. The algorithm takes sequential data with n features and p days lag as input sequence window.

Figure 4. Proposed SLSTM Architecture.

The SWAT and SLSTM models developed in this study were compared using coefficient of determination (R^2), and Nash-Sutcliffe efficiency coefficient (NSE) measures.

3. Results and Discussion

3.1. The SWAT Performance Evaluation

The calibration and validation of the SWAT model were performed for both river basins. Calibration using the SUFI-2 method in the SWAT-CUP computing package is performed based on the parameters that have a high impact on the model output. These parameters are selected through performing a model sensitivity analysis. The selection of parameters is an essential step because some parameters in SWAT hinder the possibility of manual calibration. By reviewing the relevant literature, 20 parameters that are generally the most sensitive in the SWAT model are given in Table 2. These are from the different model components, namely surface retention, groundwater flow, and soil characteristics. Through the sequencing and fitting process, a Bayesian framework, the SUFI-2, was used to calibrate the SWAT model and compute SWAT output uncertainties, as indicated in the methodology. The SUFI-2 uses the Latin Hypercube sampling (LHS) method, which reduces the amount of sample points as compared to the Monte Carlo sampling methodology. The uncertainty in the calibration procedure was assessed in this study using three iterations with 1000 model runs each iteration. The Nash-Sutcliffe Efficiency Coefficient was chosen as the objective function; the parameter uncertainty and the simulation results were acceptable if the NSE value was greater than 0.4. Lower band, upper band, and adjusted values (calibrated values) are listed in Table 2 for the two catchments.

Table 2. Calibrated parameters of SWAT model with range and calibrated values.

Parameter	Definition	Lower Band	Upper Band	Adjusted Value (Trusan)	Adjusted Value (Samarahan)
R_RCHRG_DP.gw	Deep aquifer percolation	−0.2	0.1	−0.14555	0.0529
R_SOL_BD(..).sol	Moist bulk density	0	0.4	0.2518	0.0202
V_TLAPS.sub	Temperature lapse rate	−8	−5	−7.1495	−7.7434
V_PLAPS.sub	Precipitation lapse rate	100	300	105.5	110
R_SLSUBBSN.hru	Slope sub-basin	−0.1	0.3	−0.0639	0.234
R_SHALLST.gw	Initial depth of water in the shallow aquifer	−0.3	0.1	−0.2469	0.046
R_GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur	−0.2	0.4	0.3	0.15
V_CH_N2.rte	Manning's "n" value for the main channel	0.1	0.4	0.21145	0.3216
R_CN2.mgt	Curve number for moisture condition II	−0.15	0.15	−0.13395	−0.0549
R_SOL_AWC(..).sol	Available Water Capacity is calculated as the difference between field capacity at the wilting point	0	0.6	0.1515	0.5542
R_SOL_K(..).sol	Saturated hydraulic conductivity	−0.1	0.3	0.0927	0.2356
R_OV_N.hru	Manning's N	5	9.5	6.9	9.1
V_GW_DELAY.gw	Groundwater delay	200	500	218.925003	491.684204
R_GW_REVAP.gw	Groundwater "revap" coefficient	−0.2	0	−0.1895	−0.1962
V_REVAPMN.gw	Threshold depth of water for "revap" to occur	600	1200	1039.300049	804.144592
R_ALPHA_BNK.rte	Baseflow alpha factor for bank storage	0	0.2	0.1895	0.0878
R_ALPHA_BF.gw	Base flow alpha factor	0	0.6	0.105875	0.5973
R_CH_K2.rte	Effective hydraulic conductivity in main channel alluvium	−0.1	0.2	0.02495	−0.0565
V_EPCO.hru	Plant uptake compensation factor	0.8	1	0.8455	1.0705
V_ESCO.hru	Soil evaporation compensation factor	0.1	0.3	0.2485	0.2834

The hydrograph of the 95PPU-simulated runoff is compared with the observed runoff using SUFI-2 in Figure 5. The green band is the 95 percent prediction interval for the best estimates' parameter set, and it may cover the bulk of peak flow and dry periods. After three cycles, the best estimating parameter choices yield NSE = 0.75.

Figure 5. The optimally simulated runoff with 95PPU to the observed runoff compared runoff using the SUFI-2 approach.

Referring to Figure 5 and Table 3, it can be concluded that the SWAT model in the Samarahan Basin has shown better results compared with the Trusan; however, the performance values are not as high. The reason behind this is associated with the dearth of geographical data. Therefore, this impact on the SWAT model as a semi-distributed model is significant because these models are sensitive to the availability of geographical data. Considering that the Trusan is a much bigger area than the Samarahan, with a lower distribution of rainfall stations, the SWAT model has demonstrated low performance compared with the Samarahan.

Table 3. Models' performance at the Samarahan and Trusan river basins.

	Samarahan						Trusan					
	R^2		RMSE (cms)		NSE		R^2		RMSE		NSE	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
PCA-SLSTM	0.95	0.76	1.47	2.09	0.89	0.76	0.91	0.84	73	68	0.8	0.82
SLSTM	0.9	0.74	2.07	2.26	0.78	0.72	0.88	0.78	88	77	0.86	0.77
SWAT	0.5	0.67	2.8	2.53	0.49	0.66	0.65	0.42	157.3	132.5	0.45	0.34

3.2. The SLSTM Performance Evaluation

3.2.1. Data Preprocessing

Two main strategies for data preparation were used in this work. The first approach is normalizing data to distribute it around 0, and since standard deviation was equal to 1 in most cases, this approach made the training easier. In order to address the multicollinearity problem, and to preserve maximum statistical variability, the Principal Component Analysis (PCA) transform approach was chosen. PCA is arguably the most popular unsupervised machine learning method that can be used for dimension reduction (O'Farrell et al., 2005). As illustrated in the correlation heat map for both basins in Figure 6b,d, PCA is used for

extracting new independent features based on original features. New extracted features are linearly independent and sorted by correlation with the target variable i.e., flow. By using independent and some first high-correlated features extracted by PCA, machine learning models train faster due to the scaled and relatively low dimensional dataset and score higher accuracy due to lower noise input to the model. The new variables are called NRFX (X is 1 to 11) in PCA analysis.

Samarahan

Trusan

Figure 6. (a) normalized data correlation heat map for Samarahan, (b) PCA transformed data correlation heat map for Samarahan, (c) normalized data correlation heat map for Trusan, (d) PCA transformed data correlation heat map for Trusan.

To visualize the correlation between variables, a heatmap is used. The heatmap in Figure 6 for both basins represents the intercorrelation of rainfall stations with one another and with the flowrate. Figure 6 demonstrates the correlation between all rainfall variables. In the PCA-transformed variables, it can be seen that newly generated variables are independent from one another and have only shown correlation with the flow, which helps ML models to fit more easily.

3.2.2. Tuning the Model

Selecting window size

Because of the sequential nature of the problem, window size is an important parameter and should be chosen precisely. By investigating flow autocorrelation in basins, window size is selected by maximum autocorrelation lag for each basin.

ACF plots for both basins are depicted in Figure 7. ACF is an autocorrelation function which gives values of autocorrelation of any series with lagged values. When the bar rises out of the shaded area, it can be interpreted that there is meaningful autocorrelation. Based on Figure 7 for the Samarahan basin, selected window size p is 9, and for the Trusan p is 16 days; since Trusan is bigger than Samarahan, this amount seems logical.

Figure 7. Autocorrelation for finding best lag for analysis.

Hyperparameters tuning

Neural networks contain some parameters which can be tuned to minimize a defined loss function in order for the model to operate in a local optimum place on a given dataset. In this study, the activation function for all layers except for the last dense layer was the Rectified Linear Unit (ReLU), and for the tuning purposes an Adaptive Moment Estimation (Adam) optimizer with Huber loss function was used. Huber loss function is a combination of the mean squared error function and the absolute value function. When the difference between the predicted and observed value is less than the specific value δ , the loss function will turn the squared error and in larger differences the loss function will act in an absolute way. In this fashion, both the sensitivity of squared loss and the robustness of the absolute loss function will be utilized.

$$L_{\delta} = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta \\ \delta(|y - f(x)| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases} \quad (7)$$

Figure 8 shows the loss function changes in the process of model fitting at both the training and testing steps for the two basins. The goal is to minimize the loss function at both the train and test datasets, but usually, after several epochs no further improvements happen, which indicates that the learning process has ended. The best epoch for stopping training is when both train and test loss have stopped learning (it is somehow flat or oscillates around a fixed value). In addition, Figures 9 and 10 show that the PCA-SLSTM has outperformed the SLSTM model at both basins with higher R^2 magnitudes, proving the positive impacts of the data preprocessing on the SLSTM model. Therefore, it can be recommended to use PCA combined with the LSTM model in general.

3.2.3. The Models Performance Comparison

The SWAT is a semi-distributed hydrological model that employs precipitation data from only one station closest to the centroid of each sub-basin. Therefore, the model becomes sensitive to the spatial distribution of the rainfall stations. It is then difficult to calibrate the model if the near centroid station is assigned for model simulation and has no significant correlation with other rainfall stations or streamflow data. On the other hand, data-driven models, since they use only rainfall stations, eliminate this shortcoming of distributed models and reduce the sensitivity of the model to the data of one station. If the data of one station is not accurate, these models are able to reduce its impact on the final model output.

The impacts of the improper spatial distribution of rainfall stations on physical models are significant. At the Trusan basin, since there is no proper distribution of rainfall stations upstream of the streamflow station, the SWAT model has weaker performance when compared with the Samarahan basin. On the other hand, the SLSTM model employs all the rainfall stations, reduces the negative impact of improper spatial distribution of data stations and has yielded higher R^2 and NSE, compared with that of SWAT (Table 3). Table 3 shows that the data preprocessing with PCA has improved the SLSTM model's performance as well.

As mentioned before, physical models such as SWAT, in addition to climate data of temperature and precipitation, need physical data such as soil and land use maps, and slope data to form a hydrological response unit (HRU). Inaccurate and inadequate data significantly impacts the model output. This impact is well demonstrated in the Trusan basin's results, where SWAT performance is weak due to improper spatial distribution of rainfall and temperature data stations within and around the basin. SWAT at the Samarahan basin, which is a smaller basin with rainfall stations close to its borders has shown acceptable results (Table 3).

Figure 8. Loss function changes in the modelling process.

Figure 9. Models' performance comparison in Samarahan basin for test and train datasets.

Figure 10. Models' performance comparison in Trusan basin for test and train datasets.

3.3. Conclusions

Effective simulation models are imperative for effective water management. However, limited hydro-meteorological and geospatial data availability impedes successful hydrological modelling in basins with sparse and/or ill-located data measurement stations. The SWAT hydrological model was used here to investigate the above statement. This study had also aimed to illustrate the usefulness of data-driven models in simulating rainfall-runoff processes in data-scarce basins. The data-driven models of SLSTM and PCA-SLSTM in the class of machine learning models were applied. The simulated daily streamflow data of two case studies were compared based on different performance measures.

The SLSTM model has shown NSE and R^2 of 0.72 and 0.74 in the Samarahan river basin and NSE and R^2 of 0.77 and 0.78 in the Trusan river basin, and when combined with PCA its performance has been improved at both basins. The proposed SLSTM model, based on the findings of this study, proved to be a practical model for hydrological modelling of catchments with insufficient data and sparse distribution of the gauging network.

The physically-based SWAT model, considering its underlying processes for the modelling approach, is suitable for integrated modelling and management of catchments. However, it has shown to be highly sensitive to the spatial variability of precipitation data and has not been successful at simulating the weakly-gauged Trusan basin while it has shown moderate performance at Samarahan basin. The machine learning models of SLSTM and PCA-SLSTM are only applicable to studies where the catchment output/streamflow irrespective of underlying processes and interactions is to be studied. For thorough investigation of the catchment hydrology and integrated water resources management, the application of distributed models is deemed necessary to encompass all the phenomena affecting the catchment hydrology cycle. Therefore, selection of a proper model is application-dependent, highlighting that each model type has its own merits and demerits.

Author Contributions: Conceptualization, M.M. and H.G.; Methodology, V.S., S.M.K. and H.Y.; software A.D., V.S. and M.S.; validation, A.D. and M.M.; formal analysis, A.D., H.G. and H.Y.; writing original draft preparation, M.M. and H.G.; writing—review and editing, H.G. and S.M.K.; supervision, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Department of Civil Engineering, Faculty of Engineering, University of Malaya, Malaysia. (GPF044A-2019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Mirzaei, M.; Huang, Y.F.; El-Shafie, A.; Chimeh, T.; Lee, J.; Vaizadeh, N.; Adamowski, J.; Valizadeh, N. Uncertainty analysis for extreme flood events in a semi-arid region. *Nat. Hazards* **2015**, *78*, 1947–1960. [\[CrossRef\]](#)
2. Ravazzani, G.; Valle, F.D.; Gaudard, L.; Mendlik, T.; Gobiet, A.; Mancini, M. Assessing Climate Impacts on Hydropower Production: The Case of the Toce River Basin. *Climate* **2016**, *4*, 16. [\[CrossRef\]](#)
3. Ramireddygar, S.; Sophocleous, M.; Koelliker, J.; Perkins, S.; Govindaraju, R. Development and application of a comprehensive simulation model to evaluate impacts of watershed structures and irrigation water use on streamflow and groundwater: The case of Wet Walnut Creek Watershed, Kansas, USA. *J. Hydrol.* **2000**, *236*, 223–246. [\[CrossRef\]](#)
4. Galavi, H.; Mirzaei, M.; Shul, L.T.; Valizadeh, N.; Shui, L.T. Klang River-level forecasting using ARIMA and ANFIS models. *J. Am. Water Work. Assoc.* **2013**, *105*, E496–E506. [\[CrossRef\]](#)
5. Valizadeh, N.; El-Shafie, A.; Mirzaei, M.; Galavi, H.; Mukhlisin, M.; Jaafar, O. Accuracy Enhancement for Forecasting Water Levels of Reservoirs and River Streams Using a Multiple-Input-Pattern Fuzzification Approach. *Sci. World J.* **2014**, *2014*, 1–9. [\[CrossRef\]](#)
6. Rujner, H.; Leonhardt, G.; Marsalek, J.; Viklander, M. High-resolution modelling of the grass swale response to runoff inflows with Mike SHE. *J. Hydrol.* **2018**, *562*, 411–422. [\[CrossRef\]](#)
7. Sonnenborg, T.O.; Christiansen, J.R.; Pang, B.; Bruge, A.; Stisen, S.; Gundersen, P. Analyzing the hydrological impact of afforestation and tree species in two catchments with contrasting soil properties using the spatially distributed model MIKE SHE SWET. *Agric. For. Meteorol.* **2017**, *239*, 118–133. [\[CrossRef\]](#)

8. Metcalfe, P.; Beven, K.; Freer, J. Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps. *Environ. Model. Softw.* **2015**, *72*, 155–172. [\[CrossRef\]](#)
9. Amirabadizadeh, M.; Ghazali, A.H.; Huang, Y.F.; Wayayok, A. Assessment of impacts of future climate change on water resources of the Hulu Langat basin using the swat model. *Water Harvest. Res.* **2017**, *2*, 13–29.
10. Salimirad, H.; Dehvari, A.; Galavi, H.; Ebrahimian, M. Identification and Uncertainty Analysis of Sensitive Parameter of SWAT model in Kardeh Streamflow Simulation. *Iran Water Resour. Res.* **2020**, *16*, 212–221.
11. Mirzaei, M.; Galavi, H.; Faghih, M.; Huang, Y.F.; Lee, T.S.; El-Shafie, A. Model calibration and uncertainty analysis of runoff in the Zayanderood River basin using generalized likelihood uncertainty estimation (GLUE) method. *J. Water Supply Res. Technol.* **2013**, *62*, 309–320. [\[CrossRef\]](#)
12. Mirzaei, M.; Huang, Y.F.; Lee, T.S.; El-Shafie, A.; Ghazali, A.H. Quantifying uncertainties associated with depth duration frequency curves. *Nat. Hazards* **2014**, *71*, 1227–1239. [\[CrossRef\]](#)
13. Galavi, H.; Mirzaei, M. Analyzing Uncertainty Drivers of Climate Change Impact Studies in Tropical and Arid Climates. *Water Resour. Manag.* **2020**, *34*, 2097–2109. [\[CrossRef\]](#)
14. Gassman, P.W.; Reyes, M.R.; Green, C.H.; Arnold, J.G. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. *Am. Soc. Agric. Biol. Eng.* **2007**, *50*, 1211–1250. [\[CrossRef\]](#)
15. Srinivasan, R.; Ramanarayanan, T.S.; Arnold, J.G.; Bednarz, S.T. LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART II: MODEL APPLICATION. *JAWRA J. Am. Water Resour. Assoc.* **1998**, *34*, 91–101. [\[CrossRef\]](#)
16. Amini-Zad, A.; Galavi, H.; MohammadRezaPoor, O. Hydrological Modeling of Pishin Dam Watershed Using SWAT. Development and Applications of Soil and Water Assessment Tool (SWAT) in Water Resources Management. 2018; pp. 26–30. Available online: <https://civilica.com/doc/820016/> (accessed on 28 November 2021).
17. Galavi, H.; Lee, T.S. Neuro-fuzzy modelling and forecasting in water resources. *Sci. Res. Essays* **2012**, *7*, 2112–2121. [\[CrossRef\]](#)
18. Valizadeh, N.; Mirzaei, M.; Allawi, M.F.; Afan, H.A.; Mohd, N.S.; Hussain, A.; El-Shafie, A. Artificial intelligence and geo-statistical models for stream-flow forecasting in ungauged stations: State of the art. *Nat. Hazards* **2017**, *86*, 1377–1392. [\[CrossRef\]](#)
19. Mohsenzadeh Karimi, S.; Karimi, S.; Poorrajabali, M. Forecasting monthly streamflows using heuristic models. *ISH J. Hydraul. Eng.* **2021**, *27*, 73–78. [\[CrossRef\]](#)
20. Pakdaman, M.; Falamarzi, Y.; Babaeian, I.; Javanshiri, Z. Post-processing of the North American multi-model ensemble for monthly forecast of precipitation based on neural network models. *Theor. Appl. Clim.* **2020**, *141*, 405–417. [\[CrossRef\]](#)
21. Palizdan, N.; Falamarzi, Y.; Huang, Y.F.; Lee, T.S. Precipitation trend analysis using discrete wavelet transform at the Langat River Basin, Selangor, Malaysia. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 853–877. [\[CrossRef\]](#)
22. Nourani, V.; Komasi, M. A geomorphology-based ANFIS model for multi-station modeling of rainfall-runoff process. *J. Hydrol.* **2013**, *490*, 41–55. [\[CrossRef\]](#)
23. Shortridge, J.E.; Guikema, S.D.; Zaitchik, B.F. Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2611–2628. [\[CrossRef\]](#)
24. Sun, A.; Wang, D.; Xu, X. Monthly streamflow forecasting using Gaussian Process Regression. *J. Hydrol.* **2014**, *511*, 72–81. [\[CrossRef\]](#)
25. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
26. Mouatadid, S.; Adamowski, J.F.; Tiwari, M.K.; Quilty, J.M. Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. *Agric. Water Manag.* **2019**, *219*, 72–85. [\[CrossRef\]](#)
27. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [\[CrossRef\]](#)
28. Ni, L.; Wang, D.; Singh, V.P.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J. Streamflow and rainfall forecasting by two long short-term memory-based models. *J. Hydrol.* **2020**, *583*, 124296. [\[CrossRef\]](#)
29. Hu, Z.; Liu, W.; Bian, J.; Liu, X.; Liu, T.-Y. Listening to Chaotic Whispers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*; ACM Press: New York, NY, USA, 2018; pp. 261–269. [\[CrossRef\]](#)
30. Le, X.-H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. [\[CrossRef\]](#)
31. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: Model development. *JAWRA J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89. [\[CrossRef\]](#)