Article

Assessment of Short Term Rainfall and Stream Flows in South Australia

Mohammad Kamruzzaman [1],*, Md Sumon Shahriar [2] and Simon Beecham [1]

[1] Centre for Water Management and Reuse (CWMR), School of Natural and Built Environments, University of South Australia, Mawson Lakes, SA 5095, Australia;
E-Mail: simon.beecham@unisa.edu.au
[2] CSIRO Computational Informatics (CCI), Hobart, TAS 7001, Australia;
E-Mail: mdsumon.shahriar@csiro.au

* Author to whom correspondence should be addressed;
E-Mail: Mohammad.kamruzzaman@unisa.edu.au; Tel.: +61-8-8302-3108; Fax: +61-8-8302-3386.

Abstract: The aim of this study is to assess the relationship between rainfall and stream flow at Broughton River in Mooroola, Torrance River in Mount Pleasant, and Wakefield River near Rhyine, in South Australia, from 1990 to 2010. Initially, we present a short term relationship between rainfall and stream flow, in terms of correlations, lagged correlations, and estimated variability between wavelet coefficients at each level. A deterministic regression based response model is used to detect linear, quadratic and polynomial trends, while allowing for seasonality effects. Antecedent rainfall data were considered to predict stream flow. The best fitting model was selected based on maximum adjusted R^2 values (R_{adj}^2), minimum sigma square (σ^2), and a minimum Akaike Information Criterion (AIC).

The best performance in the response model is lag rainfall, which indicates at least one day and up to 7 days (past) difference in rainfall, including offset cross products of lag rainfall. With the inclusion of antecedent stream flow as an input with one day time lag, the result shows a significant improvement of the R_{adj}^2 values from 0.18, 0.26 and 0.14 to 0.35, 0.42 and 0.21 at Broughton River, Torrance River and Wakefield River, respectively. A benchmark comparison was made with an Artificial Neural Network analysis. The optimization strategy involved adopting a minimum mean absolute error (MAE).

1. Introduction

A review of rainfall-runoff modeling has been given by [1]. Rainfall and stream flow models can be applied to a diverse range of purposes including daily control of reservoirs, projecting future stream flows and flood management. Rainfall and stream-flow models can be classified as physically based, conceptual and empirical. Physically-based models include the Système Hydrologique Européan with sediment and solute transport [2] and Gridded Surface Subsurface Hydrologic Analysis [3] both of which require extensive spatial and temporal data and typically are used for small catchments. An example of a conceptual based model is the Modèle du Génie Rural à 4 paramètres Journalier (GR4J), which has been developed for understanding catchment hydrological behavior [4]. Other examples of conceptual rainfall-runoff models are the Sacramento Soil Moisture Accounting Model [5] and the SIMulation and HYDrologic model (SIMHYD) [6], which can be applied either as a lumped or gridded application. SIMHYD estimates daily stream flows from daily rainfall and areal potential evapotranspiration data. The class of empirical models includes time series models [7–13]. An advantage of an empirical model is that it can be fitted to situations where the hydrological data are restricted to rainfall and stream flow time series. A further advantage is that in a parametric test, a distribution can be fitted for assessing the hydrological behavior for any time period in any region. In addition, they can represent either linear or non-linear relationships. Time series models perform as well as physically-based alternatives [14]. Combined a conceptual model with an artificial neural network (ANN) for forecasting inflow into the Daecheong Dam in Korea [15]. Compared the wavelet decompositions of rainfall and runoff at four sites in the Tianshan Mountains [16]. They aimed to distinguish between errors in timing and errors in magnitude of hydrograph peaks. They used a cross-wavelet technique to quantify timing errors and hence provided an empirical adjustment to model predictions of stream flow.

In this study, we have proposed a novel method for assessing short-term rainfall and stream flow models. The travel time between rainfall and stream flow gauges using cross-correlation functions [10,17]. They reported that the travel time was less than one day for the Onkaparinga catchment in South Australia. In this paper, we presume that there is a higher order relationship between rainfall gauge and stream flow data. It is, therefore, important in this study to construct the correlation structure. Linear regression models are commonly used for time series analysis [18], particularly for assessing evidence of trends, higher order changes and variability, including allowing for seasonality. We developed deseasonalized and detrended time series rainfall and stream flow models from deterministic regression models including linear, quadratic and cubic terms. These models take account of both lag rainfall and the influence of stream flow. The results of this study will be useful for water managers and policy makers involved in sustainable water resource management and climate change adaptation for the catchments used in this study. The approach is capable of modeling the non-linear relationships between inputs and outputs using ANNs [19]. The first advantage of ANN is that it only requires a

small number of parameters and learns through a number of training iterations involving adjusting the parameters (weights) of the network [20]. A second advantage is that it is useful in situations where it is complex to build a physical or conceptual model, such as hydrological modeling of rainfall- stream flow processes [21–25]. ANN models were useful to find the relationships between rainfall and river flow data in a river basin in India [26]. We present a statistical approach that uses the deterministic features of a regression model to build many neural networks with a combination of different lagged input patterns. A wavelet based regression model for stream flow using the discrete wavelet transform (DWT) of the entire time series [27]. They also provided a comparison of their model performance with ANN. A chaotic stream flow model using an ensemble wavelet network [28]. Used wavelet analyses of rainfall and runoff and wavelet rainfall–runoff cross-analyses to investigate the temporal variability of the rainfall-runoff relationship [17,29]. They found that wavelet transforms provide a physical explanation of the temporal structure of the catchment response.

2. Data Collection and Preparation

The analysis is based on data from three rainfall and stream flow stations in South Australia, as presented in Figure 1. The Broughton River (BR) station is at Mooroola, which is located approximately 40 km north of Port Broughton and 20 km south west of Port Pirie. Torrance River (TR) station is located at Mount Pleasant, and its rivers and tributaries are highly variable in flow and together drain an area of 508 km^2. Wakefield River (WR) is an ephemeral river near Rhynie, with a catchment area of approximately 1913 km^2.

Figure 1. Location of Broughton River (BR), Torrance River (TR), and Wakefield River (WR).
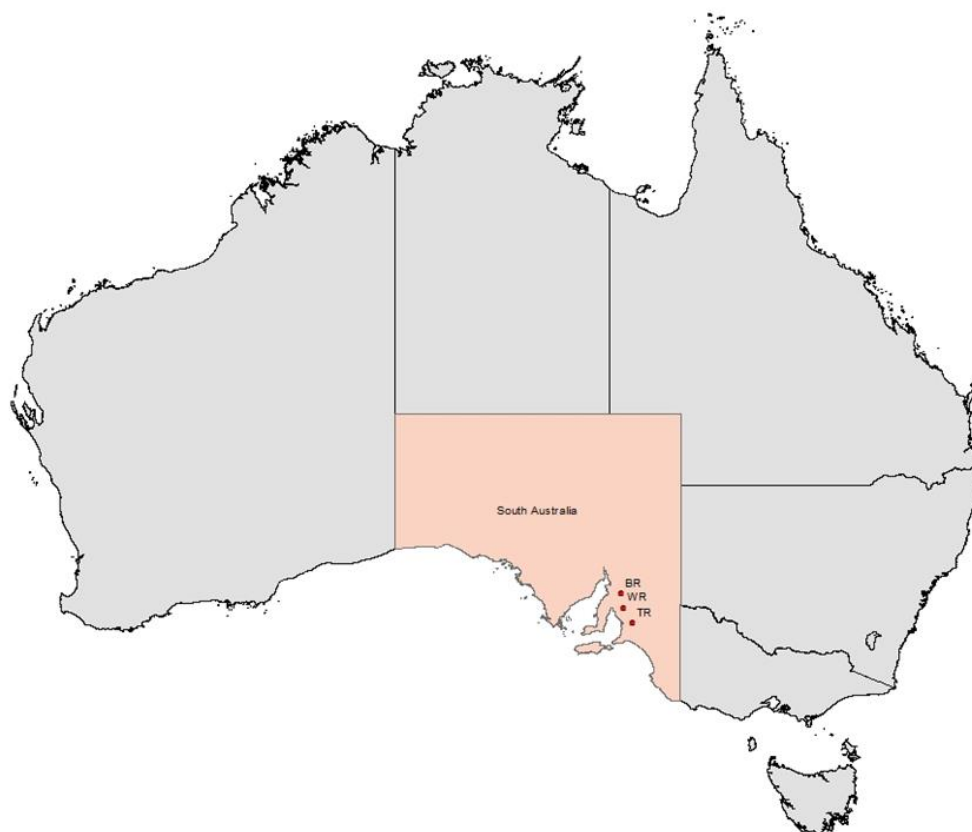
The elevation of the river may indicate the hydrological feature, presented in Table 1, Column 4. These stations were selected because they had long records of rainfall and stream flow and the highest quality control in terms of Australian Bureau of Meteorology, [30] and the Department for Environment, Water and Natural Resources [31] quality designations for rainfall and stream flow records. Information on these stations and data quality are presented in Table 1.

Table 1. Weather stations information, data quality and observations.

Stations name	ID	Location		Elevation	Variables	Data period		% of
		Latitude	Longitude			Start	End	Missing
Broughton River at Mooroola	A5070503	−33.53	138.51	196 m	Rainfall	Jun. 1989	Dec. 2011	0.1
					Stream flow	Jun. 1972	Dec. 2011	0.7
Torrance at Mount Pleasant	A5040512	−34.78	139.02	414.7 m	Rainfall	Jun. 1989	Dec. 2011	0.6
					Stream flow	May 1973	Dec. 2011	0.1
Wakefield river near Rhyine	A5060500	−34.13	138.63	202 m	Rainfall	Sep. 1985	Dec. 2011	0.9
					Stream flow	Jun. 1971	Dec. 2011	0.2

In this paper, there was less than 1% missing data and these were replaced by the mean of the series of rainfall and stream flow, to give an unbroken time series for analysis. Methods for replacing periods of missing values are discussed [18,32]. In this paper, we propose a dyadic signal time period (*i.e.*, 2^n where n is an integer and $n \geq 0$, for assessing the relationship between daily rainfall and stream flow during the period 1990–2012. We observe the discrete sequence of time series $\{y_t\}$ where $\{y_t\}$ is an integer ranging in length. We extract multi-level information of observed rainfall and stream flow series in three catchments in South Australia using the Haar wavelet decomposition. We split $\{y_t\}$ into 10 sub-time series of length power two *i.e.*, 2^n, where n is the level of the time series, starting from 0. We also investigate the correlation between rainfall and stream flow patterns for each sub-series from levels 0 to 8.

3. Statistical Analysis

3.1. Assessing the Relationship between Rainfall and Stream Flow

The open source software R [33] was used for the analyses in this paper. We calculate 10 subseries of rainfall and stream flow from 1990 to 2012 using the "wavethresh" R routine packages [34,35] for assessing the relationship between rainfall and stream flow. The length of time taken into account in 10 subseries for rainfall and stream flow is a period of 512 days.

The relationship between rainfall and stream flow within 10 subseries is presented in Figure 2. The maximum correlation coefficients are 0.08, 0.23 and 0.31 at Broughton River, Torrance River and Wakefield River, respectively. These values are between −1 and +1 in all cases, indicating the degree of linear dependence between rainfall and stream flow. For assessing short term spatial variability, a correlation coefficient of the sub-series of rainfall and stream flow less than 0.4 indicates a significant difference from 0 at each station. For example, in sub-series 2, the correlation coefficient was 0.04, 0.15 and 0.28 which indicates the independence of rainfall and stream flow at Broughton River, Torrance River and Wakefield River, respectively.

Figure 2. Correlation pattern subseries of rainfall and stream flow time series.
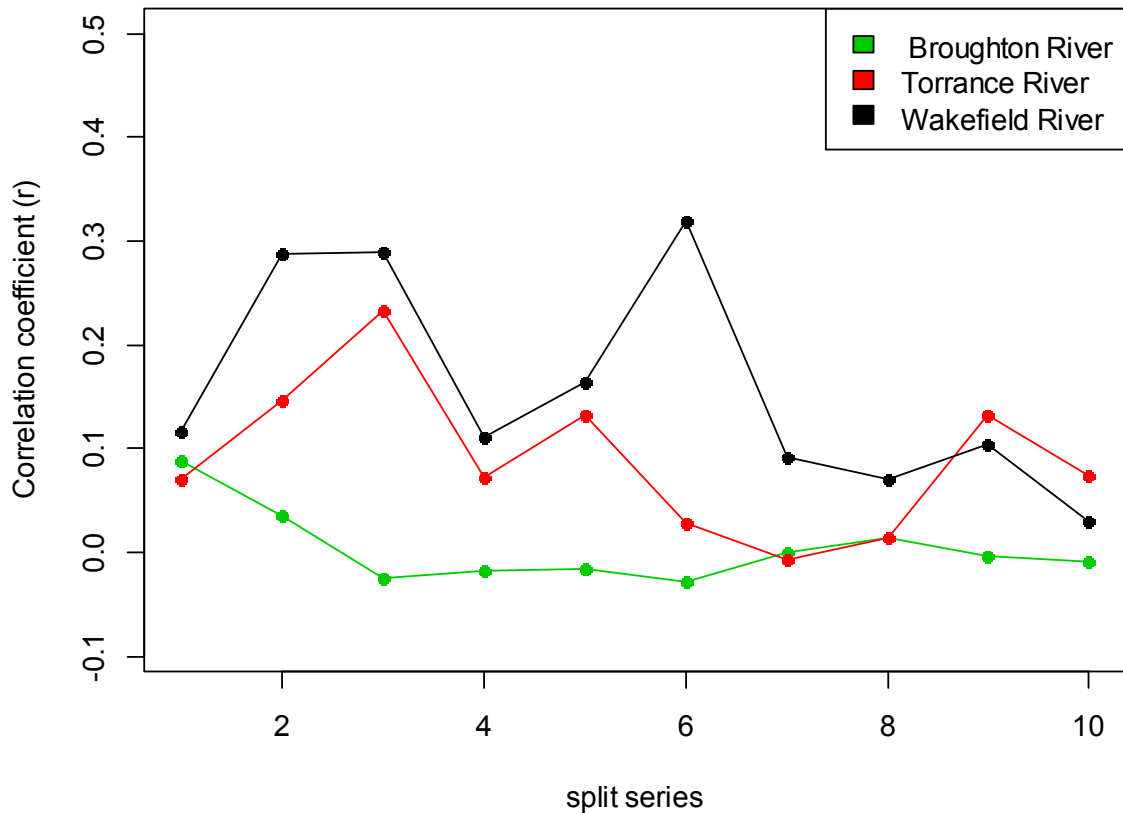
In order to understand stream flow availability under the climatic conditions in South Australia, we investigated the characteristics of rainfall and stream flow patterns, as categorized by climatic phenomena. A statistical measure of the dispersion of rainfall and stream flow patterns around the mean is defined as follows:

$$CV = \frac{S_x}{\mu_x} \tag{1}$$

where *CV* is defined as the coefficient of variation and is represented by the ratio of the standard deviation (S_x) to the mean (μ_x). Table 2 shows the degree of variation in rainfall and stream flow patterns.

Table 2. Rainfall and stream flow variability at Broughton River, Torrance River and Wakefield River in South Australia (SA) from 1990 to 2011.

Statistics	Broughton River		Torrance River		Wakefield River	
	Rainfall	Stream flow	Rainfall	Stream flow	Rainfall	Stream flow
Mean	1.653	9.817	1.530	5.396	1.282	25.333
Estimated standard deviation	0.385	4.075	0.296	4.201	0.223	21.300
Coefficient of variation (CV)	23.31%	41.51%	19.36%	77.85%	17.40%	84.07%

In Table 2, the CV for stream flow patterns indicates higher variability than for the rainfall series.

Figure 3 shows the variability of the wavelet coefficients from levels 0 to 8. The evidence of association between the rainfall and stream flow coefficient is strongly correlated at the 5% significance level in Table 1.

Figure 3. Standard deviations of wavelet coefficients of rainfall and stream flow from level 0 to 8. (**a**) Rainfall; (**b**) Stream flow.

(**a**)   (**b**)
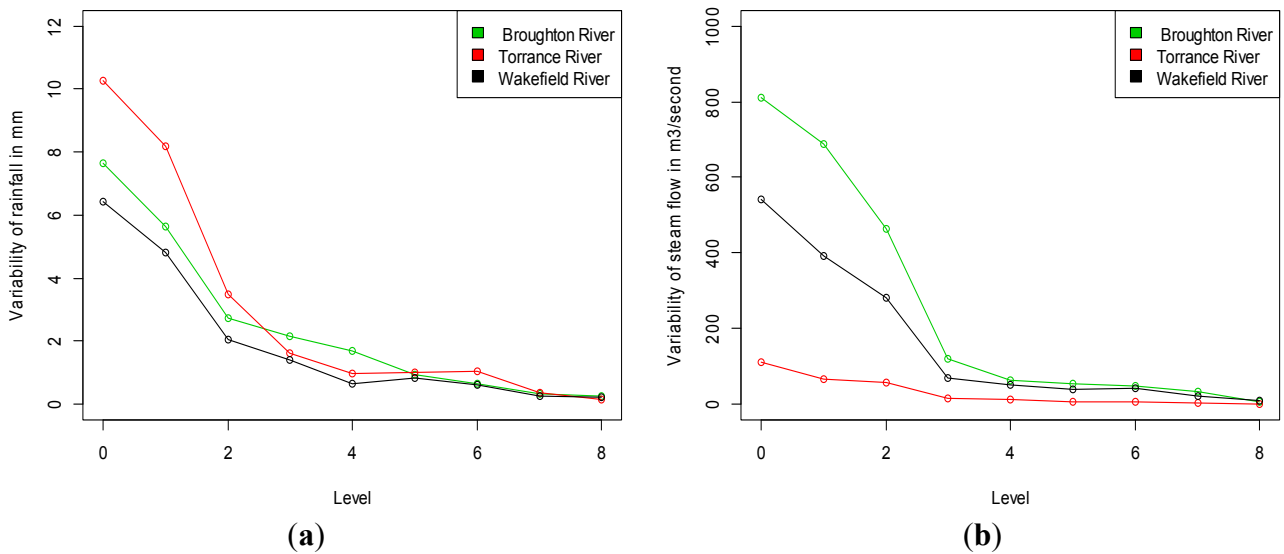
3.2. Correlation Structures between Rainfall and Stream Flow

In the previous sections, we calculated wavelet coefficients for each subset of the rainfall and stream flow series. In order to filter each of those series, we applied Haar wavelets.

The constructed correlation pattern for each rainfall and stream flow sub-series for levels 0 to 8 is given by:

$$r_k = \sum_{k=0}^{8} \frac{\sum_{j=1}^{10}(X_{wd_jk} - \overline{X}_{wd,k})(Y_{wd_jk} - \overline{Y}_{wd,k})}{S_{xwd}S_{ywd}} \tag{2}$$

and

$$\overline{X}_{wd} = \sum_{j=1}^{10} X_{wd_j}/10, \ \overline{Y}_{wd} = \sum_{j=1}^{10} Y_{wd_j}/10, \ (j = 1, 2, 3,\ldots10, \text{ and } k = 0, 1, 2,\ldots8) \tag{3}$$

where r_k is the constructed correlation with level n from 0 to 8 and X_{wd_j}, Y_{wd_j} is the *j*th sub-series of the rainfall and stream flow wavelet decomposed with the Haar procedure. The results are presented in Table 3.

The evidence of significant correlation ($r \geq 0.50$) between rainfall and stream flow wavelet coefficient series with at least a 5% significance level is shown in Table 3. Furthermore, to avoid co-linearity problems, the squared rainfall and stream flow wavelet coefficient series are also included. We found that a correlation structure ($r = 0.56$) such as stream flow is determined by rainfall on at least 4 days with 5% level at the Broughton River Basin and Torrance River Basin, as shown in Table 3. The adjusted squared stream flow and rainfall has a little evidence of correlations (*i.e.*, at 5% level) up to 64 days at Torrance River at, also a marginal correlation ($r = 0.51$) up to 128 days within squared adjusted rainfall and adjusted stream flow at Wakefield River. The rainfall and stream flow relationship was used to develop a response model for predicting stream flow.

Table 3. Constructed correlation pattern for different levels between (a) adjusted rainfall and adjusted stream flow; (b) squared adjusted rainfall and adjusted stream flow; (c) adjusted rainfall and squared adjusted stream flow; (d) squared adjusted rainfall and squared adjusted stream flow.

Days	Broughton River				Torrance River				Wakefield River			
	a	b	c	d	a	b	c	d	a	b	c	d
1	0.71 **	0.89 ***	0.70 **	0.86 ***	0.76 **	−0.184	−0.513	0.447	0.76 **	−0.53 *	−0.59 *	0.86 ***
2	0.65 *	0.264	0.469	0.449	0.72 **	0.57	0.158	0.265	0.384	0.413	0.201	0.038
4	0.56 *	−0.189	0.257	0.146	0.63 *	0.061	−0.27	0.125	0.324	−0.28	−0.341	0.115
8	0.087	0.369	−0.296	0.103	0.032	0.55 *	0.173	−0.51 *	−0.369	0.483	0.191	−0.418
16	0.233	0.08	−0.009	−0.326	0.275	−0.15	0.009	−0.311	0.306	−0.652	0.081	−0.393
32	0.094	0.055	−0.059	−0.248	0.68 *	−0.84 **	−0.81 ***	0.97 ***	−0.116	0.002	−0.382	−0.121
64	0.411	0.036	0.292	0.238	0.488	0.67 *	0.456	0.71 **	−0.091	0.166	−0.005	−0.301
128	0.423	−0.409	0.604 *	0.68 *	0.299	−0.162	0.186	−0.223	0.279	−0.575	0.128	−0.51 *
512	0.456	0.354	−0.292	0.343	−0.218	−0.405	0.007	−0.405	0.094	0.117	0.098	−0.163

Notes: * Coefficients are statistically significant at 5%; ** Coefficients are statistically significant at 1%; *** Coefficients are statistically significant at 0.1%

3.3. Rainfall-Stream Flow Response Modeling

The constructed correlations described in the previous section may be partly due to common seasonal variations and trends, so a first step is to estimate these deterministic features with regression models for entire period from 1990 to 2010. The residuals from these regressions are reformed to the deseasonalized and detrended (dsdt) time series. For all three stations, a cubic trend gave a statistically improved fit over a linear or quadratic trend over the study period. The seasonal variation was reasonably modelled by a sinusoidal curve. Therefore, the regression models are of the form:

$$T_i = \beta_0 + \beta_1 \times time + \beta_2 \times time^2 + \beta_3 \times time^3 + \beta_4 \times C + \beta_5 \times S + \varepsilon_t \tag{4}$$

where, T_i represents either rainfall or stream flow; time is the mean adjusted time, that is $(t - \bar{t})$ where t is the number of days from the start of the record and $\bar{t}$ is the mean of t, $time^2$ and $time^3$, which allows for possible quadratic and cubic trends; C is $\cos(2\pi t/365.25)$ and S is $\sin(2\pi t/365.25)$ and together these allow for seasonal variation of period one cycle per year; β_j are the unknown coefficients to be estimated; and ε_t are random variations with mean 0 and constant standard derivation.

For the estimated coefficients, only a few values are significantly different from 0 even at the 5% significance level, as shown in Table 4. There is evidence of significantly different trends in rainfall at Wakefield River, which may have corresponded to increased stream flows if rainfall is increased. We have predicted the stream flow (Y_t) on day t from rainfall (X_t) with corresponding lags k. This is referred to as a Response Model (RM). The regression is defined as:

$$Y_t = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \ldots \ldots \ldots \ldots \beta_{128} \times X_{128} + \varepsilon_t \tag{5}$$

We assess stream flow in response to rainfall at lags 0 to 128. The best fitted model is selected based on the adjusted coefficient of determination; (R^2_{adj}); minimum sigma squared (σ^2) and the Akaike Criterion Information (AIC); The AIC is defined as:

$$AIC = 2 \times \text{number of parameters} - 2 \text{Log}(L) \tag{6}$$

where L is the maximized value of the likelihood function for the estimated model. Comparisons of the AIC for different model is as shown in Table 5. The R^2_{adj} value significantly reduces and the estimated stream flow influence is close to zero after the exogenous rainfall at lag 7. Therefore, we reduced the exogenous rainfall at lags from 128 to 7 in the response model; referred to as RM0 in Table 5. This strategy is sub-optimal inasmuch as rejected terms might meet the retention criterion if added back individually. However; any small improvement in R^2_{adj} would be balanced by increased complexity in the model; which is undesirable if interaction and squared terms are added. The regression model is defined as RM:

$$Y_t = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_4 \times X_4 + \beta_5 \times X_5 + \beta_6 \times X_6 + \beta_7 \times X_7 \tag{7}$$

In the second model, we add deterministic features to the regression model including linear, quadratic and cubic terms of *t*, allowing for seasonality effects. This model is defined as RM_D:

$$Y_t = \beta_0 + L + \sum_{l=1}^{7} \beta_{5+l} \times X_l \tag{8}$$

where $L = \beta_1 \times time + \beta_2 \times time^2 + \beta_3 \times time^3 + \beta_4 \times C + \beta_5 \times S_t$.

The third model is defined as RMD_AR[1] and is an autoregressive model of order 1 (AR[1]) with RM_D. It can be written in the form:

$$Y_t = \beta_0 + L + \sum_{l=1}^{7} \beta_{5+l} \times X_l + \beta_{13} \times Y_{t-1} \tag{9}$$

The fourth model is defined as RMD_AR[2], and is an autoregressive model of order 2 (AR[2]) with RMD_AR[1]. It can be written in the form:

$$Y_t = \beta_0 + L + \sum_{l=1}^{7} \beta_{5+l} \times X_l + \beta_{13} \times Y_{t-1} + \beta_{14} \times Y_{t-2} \tag{10}$$

Table 4. Estimated coefficients of rainfall and stream flow variability from 1990 to 2012.

Station	Statistical Summary	Intercept (β_0)	Linear Term t	Quadratic Term t	Cubic Term t
Broughton River	Estimated rainfall	1.58	−0.000042	−0.000000001	−0.000000000003
	Variability of rainfall	0.106	0.00008	0.000000017	0.000000000008
	Estimated stream flow	52.08	−0.01244 *	0.0000031 *	−0.000000000258
	Variability of stream flow	6.18	0.004661	0.0000009	0.000000000485
Torrance River	Estimated rainfall	1.424	−0.00007	0.000000019	0.000000000004
	Variability of rainfall	0.077	0.00006	0.000000012	0.000000000006
	Estimated stream flow	3.47	−0.00174 *	0.0000003 *	0.000000000149 *
	Variability of stream flow	0.607	0.0004573	0.000000092	0.000000000048
Wakefield River	Estimated rainfall	1.226	−0.000123 *	0.0000000037	0.00000000001 *
	Variability of rainfall	0.067	0.000051	0.00000001	0.000000000005
	Estimated stream flow	15.95	−0.01144 *	0.0000007	0.0000000008 *
	Variability of stream flow	3.576	0.002694	0.0000005	0.000000000280

Note: * statistical significance at 5%.

Table 5. Fitted regression model for Broughton River, Torrance River and Wakefield River.

Model	Broughton River				Torrance River				Wakefield River			
	R^2_{adj}	Std. Error	AIC	RMSE *	R^2_{adj}	Std. Error	AIC	RMSE *	R^2_{adj}	Std. Error	AIC	RMSE *
RMO	0.16	333.6	1104.9	3.5107	0.24	31.19	742.77	5.074	0.13	195.8	1023.53	1.1777
RM_D	0.18	331.3	1103.9	3.1507	0.26	31.02	741.9	5.012	0.14	195.3	1023.1	1.1777
RMD_AR[1]	0.35	292.9	1085.1	0.0353	0.42	27.35	722.7	0.052	0.21	187.4	1016.9	0.11777
RMD_AR[2]	0.36	291.7	1084.5	0.0313	0.43	27.35	722.5	0.0452	0.22	187.4	1016.8	0.10777
RMD_tau	0.39	285.8	1081.4	0.0035	0.42	27.32	722.1	0.0411	0.23	187.3	1016.1	0.10178

Note: Asterisk (*) units are in m^3s^{-1}.

Finally, we develop a model for a benchmark comparison of stream flow on day t based on the entire previous period of stream flow and their influence (τ) adding with model RM_D. This model is defined as RMD_tau. Tau (τ) is 0 if there is no stream flow influence from the previous day's rainfall. We have demonstrated an example of count stream flow influence in Table 6.

Table 6. An example of count tau and stream flow influence rainfall over time.

Stream flow	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13
	8	9	0	0	0	2	9	22	3	5	8	8	6
Rainfall	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
	3	2	5	3.2	3	2.8	2.6	2.4	2.2	2	1.8	1.6	1.4

In the Table 6, when the day t = 6, Y6 = 2, then we count tau = 3 (number of 0), and Y6-3-1 = 9, can be applied in the referred model RMD_tau.

The model RMD_tau can be written in the form:

$$Y_t = \beta_0 + L + \sum_{l=1}^{7} \beta_{5+l} \times X_l + \beta_{13} \times \tau + \beta_{14} \times Y_{t-\tau-1} \tag{11}$$

The fitted model for predicted stream flow in response to exogenous rainfall, deterministic features of the regression model, and previous stream flow influence, is presented in Table 5. The best fitting model selection was based on minimum AIC and minimum root mean square Error (RMSE). The RMSE is defined as:

$$RMSE = \sqrt{E(\hat{Y}_t - Y_t)^2} \tag{12}$$

where, $\hat{Y}_t$ is defined as the estimated stream flow and Y_t is the observed stream flow, respectively.

The response model RM0 has 128 predictor variables namely the rainfall lags at 0 to 128. Therefore, there are 129 parameters to estimate including the intercept. The estimated rainfall effects belong to 0 up to 7 days lag, therefore we reduced the rainfall lags from 128 to 7 days and the optimized R^2_{adj} values for this model are 0.16, 0.24 and 0.13 for Broughton River, Torrance River and Wakefield River, respectively, as presented in Table 5. We also offset the cross product term of lags to further reduce the complexity of this model. The second model included linear quadratic and cubic terms, and this model is denoted as RM_D. The number of parameters to be estimated is therefore 8 + 3 = 11 and the R^2_{adj} increased to 0.18, 0.26 and 0.14 for Broughton River, Torrance River and Wakefield River, respectively, which is a practical and statistically significant improvement. We then

added a first order autoregressive term, referred to as a RMD_AR[1] model, and a second order autoregressive term referred to as a RMD_AR[2] model. We also made a benchmark comparison by using the entire stream flow record and this model is denoted RMD_tau, as presented in Table 5.

In Table 5, there is evidence of improvement of R^2_{adj} values, RMSE in m^{3}s^{-1} from RM to RM_D. Adding autoregressive order 1 (AR[1]) with RM_D results in substantially improved R^2_{adj} values (from 0.18, 0.26, and 0.14 to 0.35, 0.42 and 0.21 for Broughton River, Torrance River and Wakefield River, respectively. Furthermore, when adding autoregressive order 1 (AR[1]) with RM_D, there is evidence of improvement but this may be offset by the increasing number of parameters that affect the complexity of the model. In addition, the RMD_tau model represents a small improvement for two of the three river basins. The best fitted models are RMD_tau for Broughton River, RMD_AR[2] for Torrance River and RMD_tau for Wakefield River, were selected based on the minimum Akaike Information Criterion (AIC) and minimum root mean square error (RMSE) in m^{3}s^{-1}. The residuals from the best fitted models were transformed to normalized form by factor multiplication. A factor was calculated, which allows for the fact that the mean of a non-linear function of a random variable is not equal to that function of the mean. The transform series follow an identically normalized form with mean (μ) of zero, standard deviation (σ^2) of 1 and a random disturbance term (ε_t) which is uncorrelated. The transformed series were used to predict the stream flow on day *t* based on the predicted stream flow influence over the short term, as shown in Figure 4.

Figure 4. Predicted stream flow based on dsdt rainfall for (**a**) Broughton River; (**b**) Torrance River; and (**c**) Wakefield River from 1990 to 2010.

(**a**)   (**b**)   (**c**)
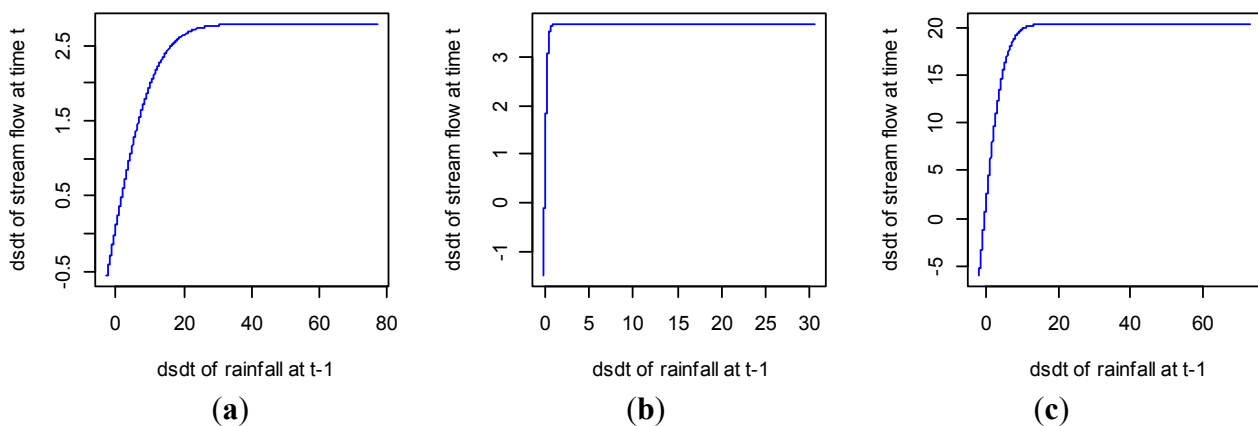
In Figure 4, we demonstrate the versatility of stream flow prediction. It can be seen that this is a non-linear relationship when expressed in terms of the physical interpretation of stream flow based on rainfall.

3.4. Modeling Stream Flow Using an Artificial Neural Network

Artificial neural network (ANN) techniques are motivated by the principles of biological nervous systems [36]. Although there are different types of ANN, the multilayer feed forward network is the most commonly used technique. For example, a common approaches of training using back-propagation in a multi-layer feed forward network [23]. The network consists of input, hidden

and output layers. Each layer is fully connected with the proceeding layer with weights in each connection, as shown in Figure 5.

Figure 5. A schematic ANN including input, hidden and output layers.
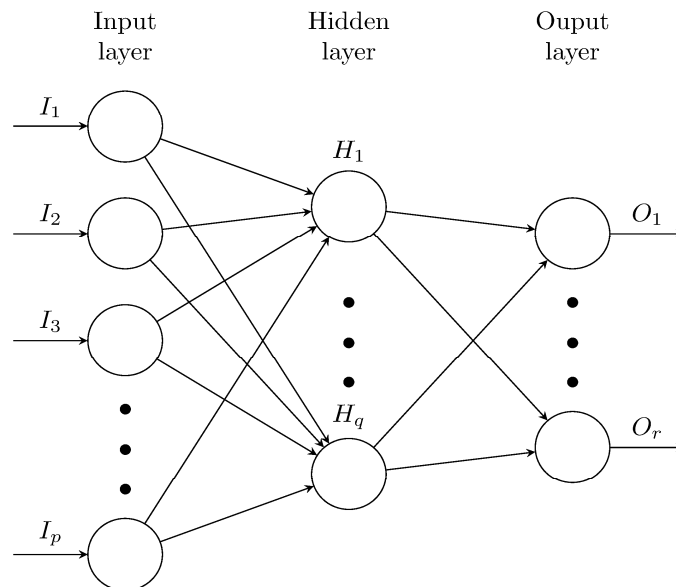
In Figure 5, the number of nodes in the input layer is p, the number of nodes in the hidden layer is q and the number of nodes in the output layer is r. The initial assigned random weights are updated during the training process by comparing the predicted output and the known output for errors. Errors are then back-propagated to adjust the weights. The dsdt of daily rainfall and stream flow data from the regression model developed in the previous section are considered for developing a prediction model for each of the three river basins for the years 1990 to 2010. A certain methods proposed such as input selection, model architecture selection, model calibration (training) and validation (testing) [37]. In addition, we emphasize the fact that ANN set-up has to be carefully achieved and described to get the reliable results. This study described the steps in building the prediction models for stream flow. We consider the prediction function as: $S_{t+1} = f(S_t, S_{t-1}, S_{t-2}, \ldots, S_{t-m}, R_t, R_{t-1}, R_{t-2}, \ldots, R_{t-n})$ where S represents stream flow, R represents rainfall, t is the current day, $m = \{3,\ldots,8\}$, $n = \{3,\ldots,8\}$ and f represents the ANN as a regression function. We investigate necessary lagged inputs of rainfall and river flow for modeling the river flows at three locations in South Australia. We apply an artificial neural network (ANN) technique for modeling river flow. ANN models are developed with all combinations of rainfall and river flow input ranges. In addition, a standard range of nodes in the hidden layer are also considered. Among all models based on inputs and hidden nodes, the best model is selected based on mean absolute error criteria. This entire process is applied to all three locations. ANN models capture the non-linear relationships of rainfall and river flow patterns in modeling river flows from large time series data. For example, if we consider 3 days lag of stream flow and 5 days lag of rainfall, then the total number of input nodes in the ANN structure will be 8 and we consider the number of nodes in the hidden layers ranging from 1 to 10. To achieve the best model using ANN for each location, all inputs not only apply in combination, but we also consider setting a range of parameters, such as different number of nodes in the hidden layer, for each combination of inputs.

In predicting stream flow one day ahead as output, we consider stream flow and rainfall with combinations of consecutive lags where the minimum lag is 3 days and the maximum lag is 8 days. Thus, for each location, the total number of models to be trained becomes 36. As the data set is large, one year of data is considered initially for testing. For training ANN models at each location, we consider stream flow and rainfall data for the period 1990 to 2009. The remaining data for the year 2010 is used for testing the best model found in the training phase.

For the Multilayer Perceptron (MLP) function, the ANN stream flow prediction model was built using the RWeka package in R Language [38]. One of the important parameters to specify is the number of nodes in the hidden layer, which may vary for time series modeling in different locations. Using trial and error, the number of nodes in the hidden layer is considered from 1 to 10. This range is widely used in hydrological time series modeling [21]. We consider the learning rate (the amount the weights are updated) to be 0.3, momentum is 0.2 and the number of epochs to train is 500.

Application of back propagation in ANN with a sigmoidal function was used to set the normalized data in the MLP function. Furthermore, the mean absolute error (MAE) in m^3s^{-1} was minimized through an iteration process that varied the number of nodes in the hidden layer.

The best lag combination at each location is presented in Figure 6.

Figure 6. MAE for training data (1990–2009) using ANN with best lag combinations at each location, units in m^3s^{-1}.
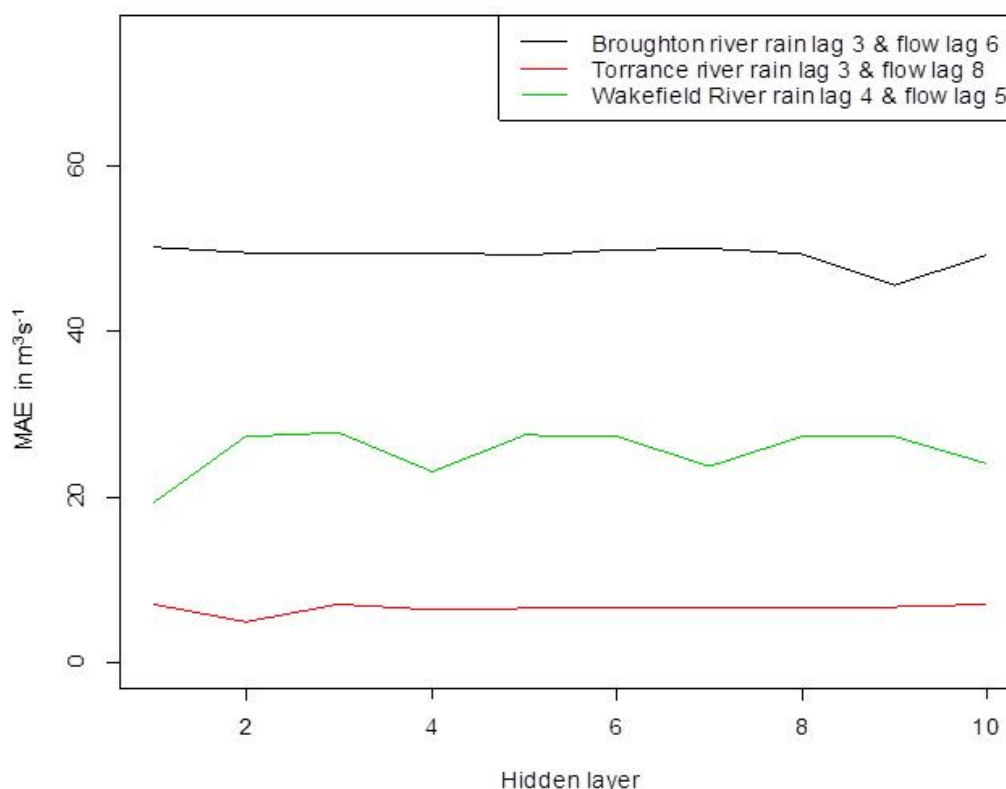
We find that both input lags and nodes in the hidden layer are different for each location. The best model based on correlation coefficient (R^2_{adj}) and the lowest root mean square error (RMSE) and mean absolute error (MAE) for each location is presented in Table 7. For Broughton River, 3 days rainfall and 6 days stream flow as lagged inputs with 9 nodes in the hidden layers produces the lowest MSE.

At Torrance River, 3 days rainfall and 8 days stream flow as lagged inputs with 2 nodes in the hidden layers produces the lowest MSE. For Wakefield River, 4 days rainfall and 5 days stream flow as lagged inputs with only one node in the hidden layer produces the lowest MSE. This indicates the variability in the ANN models for different locations.

When the best model is identified based on the training data for each location, we use this model on testing data prediction. This study show the prediction results for the testing data for each location. Figure 7 shows the predicted and observed stream flows using testing data for the locations Broughton River, Torrance River and Wakefield River, respectively.

Table 7. Best prediction model based on R^2_{adj}, lowest RMSE and MAE are in m^3s^{-1} on the training data.

| Location	Input Lags	Nodes in Hidden Layer in ANN(H)	R^2_{adj}	RMSE *	MAE *
Broughton River	3 days rain, 6 days stream flow	9	0.68	270.33	45.53
Torrance River	3 days rain, 8 days stream flow	2	0.71	24.54	4.89
Wakefield River	4 days rain, 5 days stream flow	1	0.45	179.42	19.28

Note: Asterisk (*) units are in m^3s^{-1}.

Figure 7. Observed and predicted stream flow for (**a**) Broughton River; (**b**) Torrance River; and (**c**) Wakefield River for the year 2010.

(**a**)　　　　　　　　　　　　(**b**)　　　　　　　　　　　　(**c**)
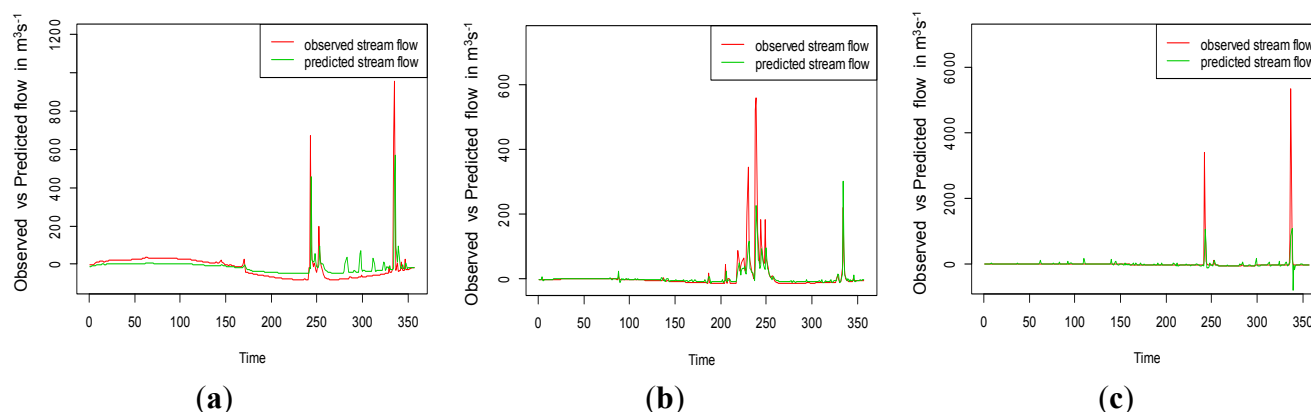
The MAE for training and testing data is shown in Figure 8 for all three locations. We observed that the MAE for the training and testing data at Broughton and Torrance Rivers do not vary significantly.

For Broughton, in training, the best ANN model structure includes 3 days lagged rainfall and 6 days lagged stream flow as inputs with 9 nodes in the hidden layer. This model has the lowest MAE, at 45.53 m^3s^{-1}. We further use this best model for testing and we find the MAE of 32.43 m^3s^{-1}. For Torrance, the ANN best model in training has 3 days lagged rainfall and 8 days lagged stream flow as inputs with 2 nodes in the hidden layer achieving the MAE of 4.89 m^3s^{-1}. For testing data, this model gives a MAE of 9.27 m^3s^{-1}. In case of Wakefield, the best ANN model has 4 days lagged rainfall and 5 days lagged stream flow as inputs with 1 node in the hidden layer achieving the MAE of 19.28 m^3s^{-1}. For the testing data, this model achieves an MAE of 42.88 m^3s^{-1}. The reason for the difference in MAE between the training and testing phases could be due to this river's ephemeral nature, and its substantial dependence on rainfall.

Figure 8. Comparison of MAE for training and testing data, units are in m^{3}s^{-1}.
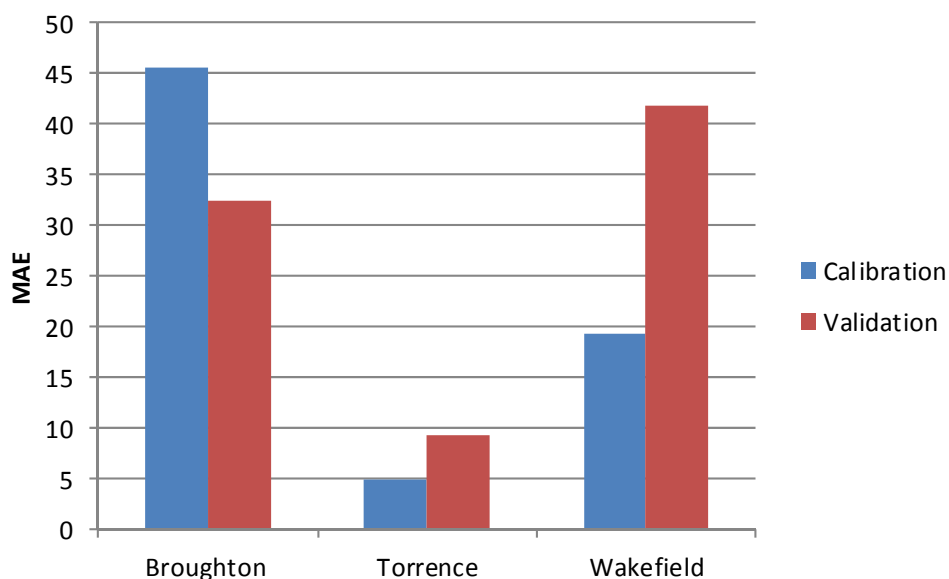
4. Conclusions

Initially, we split the whole series with a dyadic signal process for assessing the short term relationship between rainfall and stream flow including correlation using Haar wavelets. We have presented an innovative idea for the hydrological community for assessing stream flow for any catchment. In particular, the end user could assess the variability of changes and construct higher order correlations from 2 days up to as long as required. In addition, this study would be helpful for predicting stream flows using deterministic regression techniques, particularly where there is evidence of changes of statistical distribution characteristics, which is important for Water Sensitive Urban Design, as clearly demonstrated [39]. Using a deterministic regression based response model we found an increasing trend in stream flow when rainfall increased significantly. Predicted stream flow was more influenced by the previous few days' stream flows than when considering the entire previous period of stream flow. We also developed artificial neural network models for three locations. The results show that the influence of lagged rainfall and stream flow lies within a short temporal window. The results demonstrate that the ANN models perform better for Broughton and Torrance River in capturing the rainfall and stream flow relationships.

Acknowledgments

Author Contributions

This manuscript, draft design by Mohammad Kamruzzaman and Md Sumon Shahriar, complete revised by Simon Beecham.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Moradkhani, H.; Sorooshian, S. General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. *Water Sci. Technol. Libr.* **2008**, *63*, 1–23.
2. Birkinshaw, S.J. Introduction. Available online: http://research.ncl.ac.uk/shetran/Introduction.htm (accessed on 26 July 2013).
3. Downer, C.W.; Ogden, F.L. GSSHA: A model for simulating diverse stream flow generating processes. *J. Hydrol. Eng.* **2004**, *9*, 161–174.
4. Perrin, C.; Michel, C.; Andreassian, V. Improvement of a parsimonious model for stream flow simulation. *J. Hydrol.* **2003**, *279*, 275–289.
5. Burnash, R.J.C.; Ferreal, R.L.; McGuire, R.A. *A Generalized Stream Flow Simulation System: Conceptual Modeling for Digital Computers*; U.S. Department of Commerce, National Weather Service and Department of Water Resources: Sacramento, CA, USA, 1973.
6. Chiew, F.H.S.; Peel, M.C.; Western, A.W. Application and testing of the simple rainfall-runoff model SIMHYD. In *Mathematical Models of Watershed Hydrology*; Singh, V.P., Frevert, D., Meyer, S., Eds.; Water Resources Publications: Littleton, CO, USA, 2002.
7. Beven, K.J. *Rainfall-Runoff Modelling: The Primer*; Wiley: Chichester, UK, 2011; pp. 1–360.
8. Castellano-Méndez, M.; Gonzàlez-Manteigo, W.; Febrero-Bande, M.; Prada-Sànchz, J.M.; Lozano-Calderon, R. Modelling of the monthly and daily behaviour of the runoff of the Xallas rivers using Box-Jenkins and neural networks methods. *J. Hydrol.* **2004**, *296*, 38–58.
9. Hipel, K.W. (Ed.) *Stochastic and Statistical Methods in Hydrology and Environmental Engineering: Time Series Analysis in Hydrology and Environmental Engineering, Water Science and Technology*; Springer: New York, NY, USA, 2010.
10. Kamruzzaman, M.; Metcalfe, A.; Beecham, S. Wavelet based rainfall-stream flow models for the South-East Murray Darling Basin. *ASCE J. Hydrol. Eng.* **2014**, *19*, 1283–1293.
11. Kamruzzaman, M.; Beecham, S.; Metcalfe, A. Climatic influence on rainfall and runoff variability in the South-East region of the Murray Darling Basin. *Int. J. Climatol.* **2013**, *33*, 291–311.
12. Kamruzzaman, M.; Beecham, S.; Metcalfe, A. Evidence for Changes in Daily Rainfall Extremes in South Australia. In Proceedings of 9th International Workshop on Precipitation in Urban Areas, IWA/IAHR, St. Moritz, Switzerland, 6–9 December 2009.
13. Solomatine, D.; See, L.; Abrahart, R. Data-driven modelling: Concepts, approaches and experiences. In *Practical Hydroinformatics*; Abrahart, R., See, L., Solomatine, D., Eds.; Springer Berlin: Heidelberg, Germany, 2008; Volume 68, pp. 17–30.
14. McIntyre, N.; Al-Qurashi, A.; Wheater, H. Regression analysis of rainfall-runoff events from an arid catchment in Oman. *Hydrol. Sci. J.* **2007**, *52*, 1103–1118.
15. Kim, Y.; Jeong, D.; Ko, I. Combining rainfall-runoff model outputs for improving ensemble stream flow prediction. *J. Hydrol. Eng.* **2006**, *11*, 578–588.
16. Liu, H.L.; Bao, A.; Chen, X.; Wang, L.; Pan, X.L. Response analysis of rainfall-runoff processes using wavelet transform: A case study of the alpine meadow belt. *Hydrol. Processes* **2011**, *25*, 2179–2187.

17. Kamruzzaman, M.; Beecham, S.; Metcalfe, A. Wavelet Based Assessment of Relationship between Hydrological Time Series in South East Australia. In Proceedings of 2nd Conference on Practical Responses to Climate Change, Engineers Australia, Canberra, Australia, 1–3 May 2012.

18. Kamruzzaman, M.; Beecham, S.; Metcalfe, A. Non-stationarity in rainfall and temperature in the Murray Darling Basin. *Hydrol. Process* **2011**, *25*, 1659–1675.

19. Hsu, K.-L.; Gupta, H.V.; Sorooshian, S. Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res*. **1995**, *31*, 2517–2530.

20. Sajikumar, N.; Thandaveswara, B.S. A non-linear rainfall-runoff model using an artificial neural network. *J. Hydrol.* **1999**, *216*, 32–55.

21. Kisi, O.; Shiri, J.; Tombul, M. Modelling rainfall-runoff process using soft computing techniques. *Comput. Geosci.* **2013**, *51*, 108–117.

22. Kisi, O. River flow forecasting and estimation using different artificial neural network techniques. *Hydrol. Res.* **2008**, *39*, 27–40.

23. Kisi, O. Stream flow forecasting using different artificial neural network algorithms. *J. Hydrol. Eng.* **2007**, *12*, 532–539.

24. Rezaeian Zadeh, M.; Amin, S.; Khalili, D.; Singh, V.P. Daily outflow prediction by Multi Layer perceptron with logistic sigmoid and tangent sigmoid activation functions. *Water Resour. Manag.* **2010**, *24*, 2673–2688.

25. Rezaeianzadeh, M.; Stein, A.; Tabari, H.; Abghari, H.; Jalalkamali, N.; Hosseinipour, E.Z.; Singh, V.P. Assessment of a conceptual hydrological model and Artificial Neural Networks for daily outflows forecasting. *Int. J. Environ. Sci. Technol.* **2013**, *10*, 1181–1192.

26. Sudheer, K.P.; Gosain, A.K.; Ramasastri, K.S. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol. Processes* **2002**, *16*, 1325–1330.

27. Kisi, O. Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resour. Manag.* **2011**, *25*, 579–600.

28. Dhanya, C.T.; Kumar, D.N. Predictive uncertainty of chaotic daily stream flow using ensemble wavelet networks approach. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2010WR010173.

29. Labat, D.; Ababou, R.; Mangin, A. Rainfall-runoff relations for karstic springs. Part II: Continuous wavelet and discrete orthogonal multi resolution analyses. *J. Hydrol.* **2000**, *238*, 149–178.

30. Australian Bureau of Meteorology (BoM). Climate Data Online. Available online: http://www.bom.gov.au/climate/data/index.shtml (accessed on 12 January 2013).

31. Department for Environment, Water and Natural Resources (DEWNR) 2013. Available online: https://www.waterconnect.sa.gov.au/Systems (accessed on 12 March 2013).

32. Weerasinghe, S. A missing values imputation method for time series data: An efficient method to investigate the health effects of sulphur dioxide levels. *Environmetrics* **2010**, *21*, 162–172.

33. R Development Core Team. *Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011.

34. Nason, G.P. *Wavelet Methods in Statistics with R*; Springer: New York, NY, USA, 2008.

35. R routine in open access. Wavethresh package, 2013. Available online: http://cran.ms.unimelb.edu.au/web/packages/wavethresh/index.html (accessed on 10 November 2013).

36. Haykin, S. *Neural Networks—A Comprehensive Foundation*, 2nd ed; Prentice-Hall: Upper Saddle River, NJ, USA, 1998; pp. 26–32.

37. Maier, R.H.; Jain, A.; Graeme, C.D.; Sudheer, K.P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* **2010**, *25*, 891–909.

38. R routine in open access. RWeka package, 2014. Available online: http://cran.r-project.org/web/packages/RWeka/index.html (accessed on 28 April 2014).

39. Beecham, S.; Chowdhury, R. Effects of changing rainfall patterns on WSUD in Australia. *Proc. ICE Water Manag.* **2012**, *165*, 285–298.