Check for updates

Assessment of rainfall aggregation and disaggregation using data-driven models and wavelet decomposition

Sungwon Kim, Ozgur Kisi, Youngmin Seo, Vijay P. Singh and Chang-Joon Lee

ABSTRACT

The objective of this study is to develop hybrid models by combining data-driven models, including support vector machines (SVM) and generalized regression neural networks (GRNN), and wavelet decomposition for aggregation and disaggregation of rainfall. The wavelet-based support vector machines (WSVM) and wavelet-based generalized regression neural networks (WGRNN) models are obtained using mother wavelets, including db8, db10, sym8, sym10, coif6, and coif12. The developed models are evaluated in the Bocheong-stream catchment, an International Hydrological Program representative catchment, Republic of Korea. WSVM and WGRNN models with mother wavelet db10 yield the best performance as compared with other mother wavelets for estimating areal and disaggregated rainfalls, respectively. Among 12 rainfall stations, SVM, GRNN, WSVM (db10 and sym10), and WGRNN (db10 and sym10) models provide the best accuracies for estimating the disaggregated rainfalls at Samga (No. 7), and the worst accuracies for estimating the disaggregated rainfalls at Yiweon (No. 11) stations, respectively. Results obtained from this study indicate that the combination of data-driven models and wavelet decomposition can be a useful tool for estimating areal and disaggregated rainfalls satisfactorily, and can yield better efficiency than data-driven models.

Key words | generalized regression, kriging method, rainfall aggregation and disaggregation, support vector machines, wavelet decomposition

Sungwon Kim (corresponding author)
Chang-Joon Lee
Department of Railroad and Civil Engineering,
Dongyang University,
Yeongju 36040,
Republic of Korea
E-mail: *swkim1968@dyu.ac.kr*

Ozgur Kisi
Department of Civil Engineering, Architecture and Engineering Faculty,
Canik Basari University,
Samsun,
Turkey

Youngmin Seo
Department of Constructional Disaster Prevention Engineering,
Kyungpook National University,
Sangju 37224,
Republic of Korea

Vijay P. Singh
Department of Biological and Agricultural Engineering & Zachry Department of Civil Engineering,
Texas A & M University,
College Station,
TX 77843-2117,
USA

INTRODUCTION

Rainfall modeling is a complex task. The use of conventional approaches in modeling rainfall time series is far from trivial, since hydrometeorologic processes are complex and involve various factors, such as landscape and climatic factors, which are still not well understood (Wu *et al.* 2010).

Areal rainfall is the average rainfall over a region and is estimated by one of the popular methods, such as arithmetic mean, Thiessen polygon, isohyetal, spline, kriging, and copula among others (Chow *et al.* 1988; Goovaerts 2000; AghaKouchak *et al.* 2010). The arithmetic mean method is the simplest one for determining areal rainfall. The Thiessen polygon method assumes a linear variation in rainfall between two neighboring stations, and polygons are constructed which are essentially areal weights. This method is considered more accurate than the arithmetic mean method. The isohyetal method involves construction of isohyets using observed depths at rainfall stations and assumes a linear variation between two adjacent isohyets (Chow *et al.* 1988; Singh 1992). The spline method is an interpolation method that divides interpolation intervals into small subintervals, and each of these subintervals is interpolated by using the third-degree polynomial (Apaydin *et al.* 2004; Tait *et al.* 2006). The kriging method is an optimal interpolator, based on regression against observed rainfall values of surrounding rainfall points, weighted according to spatial covariance values (Goovaerts 2000; Ly

doi: 10.2166/nh.2016.314

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

et al. 2011). The copula method can be employed to describe the dependencies among n random variables on an n dimensional unit cube (uniform). Description of the spatial dependence structure independent of the marginal distribution is one of the most attractive features of copulas (Genest et al. 2007; Zhang & Singh 2007). In this study, rainfall aggregation means the estimation of areal rainfall using the conventional approaches such as arithmetic mean, Thiessen polygon, isohyetal, spline, kriging, and copula methods.

Rainfall disaggregation can be both temporal and spatial. Temporal rainfall disaggregation entails disaggregating hourly, daily or longer duration rainfall into shorter time rainfall, and many techniques for temporal rainfall disaggregation have been proposed (Hershenhorn & Woolhiser 1987; Ormsbee 1989; Koutsoyiannis & Xanthopoulos 1990; Glasbey et al. 1995; Connolly et al. 1998; Olsson 1998; Olsson & Berndtsson 1998; Durrans et al. 1999; Sivakumar et al. 2001; Socolofsky et al. 2001; Gyasi-Agyei 2005; Zhang et al. 2008; Knoesen & Smithers 2009). However, relatively limited research has been reported on spatial rainfall disaggregation (Perica & Foufoula-Georgiou 1996; Venugopal et al. 1999; Sharma et al. 2007) as compared with temporal rainfall disaggregation.

Data-driven models, including artificial neural networks (ANNs), neuro-fuzzy, and genetic programming, are computational methods that have been primarily used for pattern recognition, classification, and prediction (Haykin 2009). During the past decades, various data-driven models have been developed and applied for temporal rainfall disaggregation (Burian et al. 2000, 2001; Burian & Durrans 2002). Burian et al. (2000) evaluated ANNs for disaggregation of hourly rainfall into subhourly time increments. Results have shown that ANNs are comparable to other disaggregation methods, and improve the prediction of maximum incremental rainfall intensity. Burian et al. (2001) investigated the training performance of various ANN models' characteristics including data standardization, the geographic location of training data, quantity of training data, the number of training iterations, and the number of hidden neurons in ANNs. Burian & Durrans (2002) examined how the errors in the disaggregated rainfall hyetograph translate to errors in the prediction of the runoff hydrograph. However, research on the development and application using data-driven models for spatial rainfall disaggregation (Kim & Singh 2015) has been limited compared with temporal rainfall disaggregation. Kim & Singh (2015) developed ANN models, including multilayer perceptron (MLP) and Kohonen self-organizing feature map (KSOFM), for spatial disaggregation of areal rainfall in the Wi-stream catchment, an International Hydrological Program (IHP) representative catchment, Republic of Korea. Results showed that MLP and KSOFM models could disaggregate areal rainfall into individual point rainfall with spatial concepts successfully.

In recent years, wavelet decomposition and data-driven models have been combined and successfully implemented in hydrological applications including rainfall, streamflow, water stage, evapotranspiration, groundwater, reservoir inflow, and sediment load, etc. (Wang & Ding 2003; Cannas et al. 2006; Wang et al. 2009; Adamowski & Sun 2010; Kisi 2010; Rajaee 2010; Tiwari & Chatterjee 2010; Kisi & Cimen 2011; Adamowski & Chan 2011; Rajaee et al. 2011; Adamowski & Prasher 2012; Nejad & Nourani 2012; Okkan 2012; Wei et al. 2012; Okkan & Serbes 2013; Seo et al. 2015). The wavelet decomposition is a specific data-preprocessing method which can analyze a signal in both time and frequency so that it can overcome the drawbacks of the conventional Fourier transform method. The wavelet decomposition permits an effective decomposition of time series so that the decomposed data increase the performance of hydrological models by capturing the useful information at different decomposition levels (Nourani et al. 2009, 2011).

Adamowski & Sun (2010) suggested the method based on combining discrete wavelet transforms and ANNs for streamflow forecasting in non-perennial rivers. They found that the WA-ANN models provided more accurate streamflow forecasting than the ANN models. Tiwari & Chatterjee (2010) developed a hybrid wavelet-bootstrap-ANN (WBANN) model to investigate the potential of wavelet and bootstrapping techniques for developing an accurate and reliable ANN model for hourly flood forecasting. They found that the WBANN model improved the reliability of flood forecasting with greater confidence. Adamowski & Prasher (2012) compared support vector regression (SVR) and wavelet networks (WN) for daily streamflow forecasting in a mountainous watershed. They found that the best WN model performed slightly better than the best SVR model.

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Okkan & Serbes (2013) developed different models combining discrete wavelet transform (DWT) and different data-driven models, including multiple linear regression (MLR), feed forward neural networks (FFNN), and least square-support vector machines (LS-SVM) for reservoir inflow modeling. They found that the DWT-FFNN model performed better than the other models in terms of mean square error (MSE) and coefficient of determination (R^2). Nourani *et al.* (2014) recently reviewed the definition and advantages of wavelet-based models, as well as the history and potential future of their application in hydrology to predict important processes of the hydrologic cycle.

Although there have been investigations using the combination of data-driven models and wavelet decomposition, their applications for aggregation and disaggregation of rainfall has been limited. Mathematical formulas between areal and individual rainfalls on the catchment cannot be derived or developed using the conventional methods, including simple regression analysis. Therefore, the strong nonlinear behavior in nature, such as aggregation and disaggregation of rainfall, can be overcome by using the combination of data-driven models and wavelet decomposition successfully.

The objective of this study, therefore, is to develop and apply two different hybrid models, wavelet-based support vector machines (WSVM) and wavelet-based generalized regression neural networks (WGRNN), for aggregation and disaggregation of rainfall and evaluate them in the Bocheong-stream catchment, an IHP representative catchment, Republic of Korea. The paper is organized as follows: the second part describes the methodology including wavelet decomposition, support vector machines (SVM), generalized regression neural networks (GRNN), and WSVM and WGRNN, respectively. The third part describes the study area and data, and the fourth part presents the results and discussion. Conclusions are presented in the last part of the paper.

METHODOLOGY

Wavelet decomposition

Wavelet analysis is a multiresolution analysis in time and frequency domains. The wavelet transform decomposes a time series signal into different resolutions by controlling scaling and shifting. It provides good localization properties in both time and frequency domains (Nejad & Nourani 2012). It also has an advantage in that it has flexibility in choosing the mother wavelet, which is the transform function, according to the characteristics of the time series. The continuous wavelet transform (CWT) of a signal $x(t)$ is defined as (Mallat 1989; Nourani *et al.* 2009):

$$CWT_x^{\Psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \Psi^* \left(\frac{t-\tau}{s} \right) dt \quad (1)$$

where s = the scale parameter, τ = the translation parameter, * = the complex conjugate, and $\Psi(t)$ = the mother wavelet. CWT necessitates a large amount of computation time and resources, while DWT requires less computation time and is simpler to implement than CWT. DWT involves choosing scales and positions, which are called dyadic scales and positions, based on powers of two. This is achieved by modifying the wavelet representation as (Mallat 1989; Nourani *et al.* 2009):

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{\left|s_0^j\right|}} \Psi \left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right) \quad (2)$$

where j and k =the integers that control the wavelet dilation and translation, respectively. $s_0 > 1$ is a fixed dilation step, and τ_0 = the location parameter. The most common and simplest choice for parameters are $s_0 = 2$ and $\tau_0 = 1$ (Nourani *et al.* 2009). Using the wavelet discretization, the time scale can be sampled at discrete levels.

A fast DWT algorithm, developed by Mallat (1989), is based on four filters, including decomposition low-pass and high-pass, reconstruction low-pass and high-pass filters. For practical implementation of Mallat's algorithm, low-pass and high-pass filters are used instead of father and mother wavelets, which are also called scaling and wavelet functions, respectively. The low-pass filter, associated with the scaling function, allows the analysis of low frequency components, while the high-pass filter, associated with the wavelet function, allows the analysis of high frequency components. These filters, used in Mallat's algorithm, are determined according to the selection of mother wavelets

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

(González-Audícana *et al.* 2005). Multiresolution analysis by Mallat's algorithm is a procedure to obtain 'approximations' and 'details' for a given time series signal. An approximation holds the general trend of the original signal, while a detail depicts high-frequency components of it. A multilevel decomposition process (Figure 1) can be achieved, where the original signal is broken down into lower resolution components (Catalão *et al.* 2011). Detailed information for Mallat's algorithm can be found in Nason (2010).

SVM

SVM models have found wide applications in several areas, including pattern recognition, regression, multimedia, bioinformatics, and artificial intelligence. An SVM model is a new kind of classifier that is motivated by two concepts. First, transformation of data into a high-dimensional space can transform complex problems into simpler problems that can use linear discriminant functions. Second, the SVM model is motivated by the concept of training, and uses only those inputs that are near the decision surface (Principe *et al.* 2000; Tripathi *et al.* 2006; Vapnik 2010). The solution of traditional neural network models may tend to fall into a local optimal solution, whereas a global optimum solution is guaranteed for the SVM model (Haykin 2009). The current study uses an ε-support vector regression (ε-SVR) model. It has been successfully applied for modeling hydrological processes (Tripathi *et al.* 2006; Kim *et al.* 2012; 2013a, 2013b). During the ε-SVR model training performance, the purpose is to find a nonlinear function

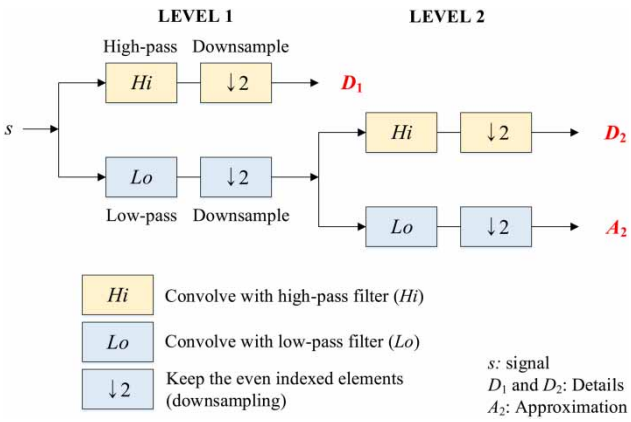

Figure 1 | Mallat's algorithm for two-level decomposition of a signal.

that minimizes a regularized risk function. This is achieved for the least value of the desired error criterion (e.g., root mean square error (RMSE)) for various constant parameters C_C, and ε and various kernel functions with various constant σ values. Detailed information on the SVM model can be found in Vapnik (2010), Principe *et al.* (2000), Tripathi *et al.* (2006), and Kim *et al.* (2012, 2013a, 2013b).

GRNN

GRNN is a neural network model based on the nonlinear regression theory. The GRNN model, as a universal approximation for smooth functions, is capable of solving any smooth function approximation problem. The process of GRNN modeling can solve the problem of local minimum (Specht 1991; Sudheer *et al.* 2003). GRNN is composed of four layers: that is, the input layer, the hidden layer, the summation layer, and the output layer. The input layer, the hidden layer, and the summation layer neurons are completely connected, whereas the output layer neuron is connected only with some of the summation layer neurons. The summation layer is composed of two types of neurons, including several summation neurons and one division neuron. Each output layer neuron is connected to the summation neuron and division neuron of the summation layer, and the connection weights are not composed between the summation layer and the output layer (Specht 1991; Wasserman 1993; Tsoukalas & Uhrig 1997).

GRNN training performance is very different from the training performance used in the MLP. The training performance between the input and hidden layers is composed of unsupervised training performance like the radial basis function (RBF). Thus, it requires a special clustering algorithm such as the K-means or orthogonal least squares (OLS) algorithms, and the radius of cluster should be set before the training performance starts. Also, the training performance between the hidden and the summation layers is composed of the supervised training performance based on a minimizing process of the mean square error for the output value from the hidden layer. Therefore, the parameters that need to be optimized during the training performance are centers, widths/spreads, and connection weights. The RBF is widely used for the transfer function of the hidden layer (Wasserman 1993; Tsoukalas & Uhrig

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

1997; Kim & Kim 2008b). A GRNN model has been successfully developed and investigated for hydrological modeling (Kisi 2006; Kim & Kim 2008b; Kim *et al.* 2012; 2013b). Detailed information on the GRNN model can be found in Tsoukalas & Uhrig (1997), Kim & Kim (2008b), and Kim *et al.* (2012, 2013b).

WSVM and WGRNN

WSVM is a combination of wavelet decomposition and SVM, whereas WGRNN is a combination of wavelet decomposition and GRNN. The wavelet decomposition is employed to decompose an input time series into approximation and detail components. The decomposed time series are used as inputs to SVM and GRNN for WSVM and WGRNN models, respectively. The application of WSVM and WGRNN models in hydrology and water resources can be found from the literature (Kisi 2011; Kisi & Cimen 2011, 2012).

WSVM and WGRNN consist of a two-step algorithm. The first step corresponds to a multilevel wavelet decomposition. The input data of SVM and GRNN are decomposed using the wavelet transform. In this study, DWT using Mallat's algorithm was used for decomposing the time series signals. The multiresolution analysis by Mallat's algorithm generates approximations and details for a given time series signal. An approximation holds the general trend of the original signal, whereas a detail depicts high-frequency components of it. Therefore, the original signal is broken down into lower resolution components. For example, two-level DWT decomposes a signal $x(t)$ into D_1, D_2, and A_2, where D_1 and D_2 are details and A_2 is an approximation. D_1, D_2, and A_2 are used as input to SVM and GRNN. The second step corresponds to training and testing phases using SVM and GRNN, respectively. Figure 2 shows the flowchart for rainfall aggregation using WSVM and WGRNN.

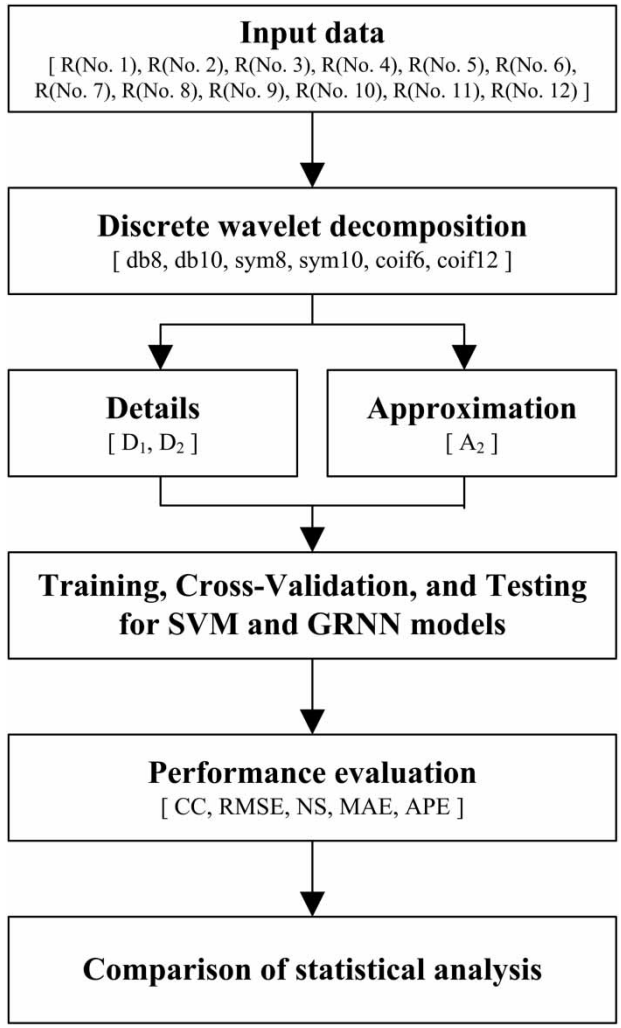

Figure 2 | Flowchart for rainfall aggregation using WSVM and WGRNN.

STUDY AREA AND DATA

Data were obtained from the Bocheong-stream catchment. The catchment, shown in Figure 3, is located at 36° 16′ to 36° 33′ latitude and at 127° 40′ to 127° 57′ longitude. It has an area of 475.68 km^2, a channel length of approximately 49.0 km, a channel slope of approximately 0.582%, a shape factor of approximately 0.166, and a river density of approximately 0.111. The catchment is short from east to west and long from south to north. There are 5 river stage stations, 5 groundwater stations, 12 rainfall stations, and 12 evaporation stations in the catchment (Ministry of Construction & Transportation 1982–2007). The hydrological data, such as rainfall, river stage, discharge, and groundwater table, had been recorded from 1982 to 2007.

To estimate areal rainfall using the Thiessen polygon, spline and kriging methods in the Bocheong-stream catchment, the hourly rainfall data from 12 rainfall stations, including Myogeum (No. 1), Cheongsan (No. 2), Neungweol (No. 3), Jungnyul (No. 4), Kwangi (No. 5), Pyeongon (No. 6),

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

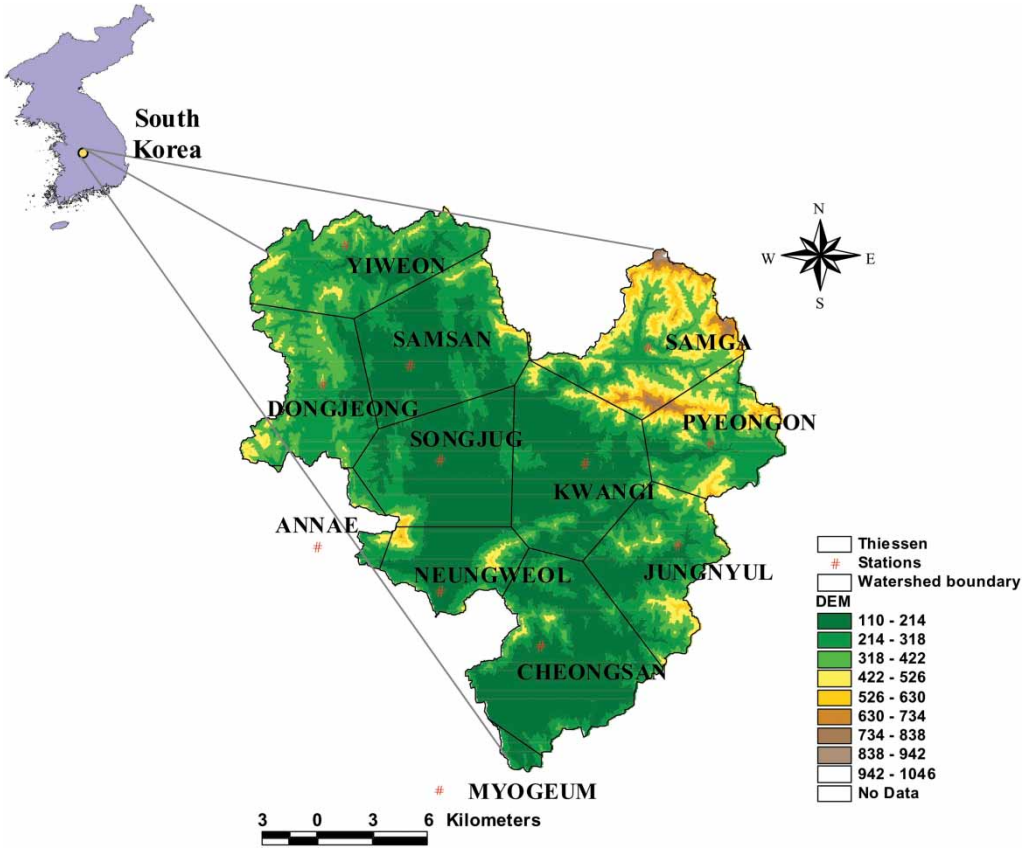

Figure 3 | Schematic diagram of the Bocheong-stream catchment.

Samga (No. 7), Songjug (No. 8), Samsan (No. 9), Dongjeong (No. 10), Yiweon (No. 11), and Annae (No. 12) stations were used. Only Myogeum (No. 1) and Annae (No. 12) stations are located outside the Bocheong-stream catchment. Since all stations are spread almost uniformly, the areal rainfall using the Thiessen polygon, spline and kriging methods can capture the natural phenomena of individual rainfall patterns in the catchment. In order for data-driven models to be able to make generalizations about rainfall, sufficient rainfall data should be available (Kim & Kim 2008a). Rainfall events must be recorded over 24 hours, including non-rainfall hours. Twelve rainfall events (events 1–12), including six floods and six typhoon events, were chosen from the mid-1980s to the mid-1990s to meet this condition. Since the kriging method includes considerable variables to estimate the areal rainfall compared with the Thiessen polygon and spline methods, the areal rainfall estimated using the kriging method was considered as observed areal rainfall.

For the data-driven model, data were split into training, cross-validation, and testing data. The training data were used for optimizing the connection weights and bias of the data-driven model, the cross-validation data were used to select the model variant that provides the best level of generalization, and the testing data were used to evaluate the chosen model against unseen data (Dawson & Wilby 2001; Izadifar & Elshorbagy 2010). The cross-validation method provides a rigorous test of a data-driven model's skill (Dawson & Wilby 2001) and is generally used to overcome the overfitting problem inherent in the data-driven models (Haykin 2009). This technique has often been applied at the end of training performance in the literature (Smith 1993; Haykin 2009) and is also employed for data-driven model selection (Stone 1974).

The training data consist of the rainfall events resulting in river floods, and the cross-validation and testing data consist of rainfall events when typhoons pass and affect the Republic of Korea. In all of these applications, 47% of data (events 1, 4,

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

7, 9, 11, and 12, $N = 459$ hours) were applied for training, 25% of data (events 5, 8, and 10, $N = 245$ hours) for cross-validation, and the remaining 28% of data (events 2, 3, and 6, $N = 280$ hours) for testing. Since floods and typhoons have occurred frequently during the summer season, the hourly rainfall data are sufficient to explain the rainfall patterns for floods and typhoons. However, it can be found that the data length does not significantly affect the performance of data-driven models. Tokar & Johnson (1999) indicated that the data length has less effect than the data quality on the performance of a neural network model. Sivakumar *et al.* (2002) indicated that it is imperative to select good training data from the available data series. They indicated that the best way to achieve a good training performance is to include most of the extreme events, such as very high and very low values, in the training data.

Table 1 summarizes statistical indices of training, cross-validation, and testing data. In Table 1, X_{mean}, X_{max}, X_{min}, S_x, C_v, C_{sx}, and SE denote the mean, maximum, minimum, standard deviation, coefficient of variation, skewness coefficient and standard error values for training, cross-validation, and testing data, respectively. Songjug (No. 8) and Dongjeong (No. 10) stations show high variation (see C_v values in Table 1) in training and cross-validation data. Pyeongon (No. 6), Songjug (No. 8), Dongjeong (No. 10), and Yiweon (No. 11) stations show high skewed distributions (see C_{sx} values in Table 1) in training and cross-validation data.

Table 2 summarizes statistical indices of areal rainfall data using the Thiessen polygon, spline and kriging methods. It is seen from Table 2 that the areal rainfall using three methods shows similar values for training, cross-validation, and testing data. The estimated rainfall values were compared with observed ones using five performance evaluation criteria: the correlation coefficient (CC), RMSE, Nash–Sutcliffe coefficient (NS) (Nash & Sutcliffe 1970; ASCE 1993), mean absolute error (MAE), and average performance error (APE). Although CC is one of the most widely used criteria for calibration and evaluation of hydrological models with observed data, it alone cannot discriminate which model is better than others. The standardization inherent in CC as well as its sensitivity to outliers yields high CC values, even when the model

Table 1 | Statistical indices of training, cross-validation, testing data

Statistical indices	Data	Rainfall stations											
		No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11	No. 12
X_{mean}	Training ($N = 459$)	1.25	1.04	1.13	1.56	1.38	1.40	1.20	0.91	1.36	0.60	0.97	1.25
	Cross-validation ($N = 245$)	1.17	1.19	1.11	1.80	1.28	2.65	1.26	0.69	0.96	0.30	1.22	0.69
	Testing ($N = 280$)	2.10	1.61	1.19	1.66	1.61	1.52	2.18	2.42	1.88	1.51	1.90	2.31
X_{max}	Training ($N = 459$)	26.00	30.00	26.50	49.00	32.00	53.00	34.50	47.00	37.00	23.00	58.00	33.00
	Cross-validation (N = 245)	18.00	24.50	24.00	38.00	17.50	111.00	35.50	44.50	15.00	22.00	42.00	19.00
	Testing ($N = 280$)	65.00	34.00	21.00	30.00	31.50	25.00	37.00	34.00	24.00	17.00	25.00	69.00
X_{min}	Training ($N = 459$)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Cross-validation ($N = 245$)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Testing ($N = 280$)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S_x	Training ($N = 459$)	3.85	3.38	3.81	4.71	4.54	4.42	3.97	4.21	4.31	2.72	3.82	4.34
	Cross-validation ($N = 245$)	3.30	3.49	3.55	5.30	3.27	9.12	4.29	3.45	2.68	1.91	3.89	2.84
	Testing ($N = 280$)	6.08	3.78	2.63	3.80	3.52	3.31	4.85	5.19	4.00	3.03	3.92	5.97
C_v	Training ($N = 459$)	3.07	3.26	3.37	3.02	3.29	3.15	3.30	4.64	3.17	4.55	3.95	3.48
	Cross-validation ($N = 245$)	2.81	2.93	3.19	2.95	2.56	3.44	3.40	4.98	2.80	6.44	3.20	4.10
	Testing ($N = 280$)	2.89	2.35	2.21	2.29	2.19	2.18	2.23	2.15	2.13	2.00	2.07	2.58
C_{sx}	Training ($N = 459$)	3.99	4.26	4.50	4.96	4.46	5.62	4.29	6.97	4.81	5.20	9.00	4.46
	Cross-validation ($N = 245$)	3.42	4.01	3.99	3.90	2.85	7.61	4.56	9.67	3.23	8.81	6.02	4.57
	Testing ($N = 280$)	6.32	4.10	3.20	4.08	3.89	3.74	3.69	3.11	3.09	3.21	2.88	6.36
SE	Training ($N = 459$)	0.18	0.16	0.18	0.22	0.21	0.21	0.19	0.20	0.20	0.13	0.18	0.20
	Cross-validation ($N = 245$)	0.21	0.22	0.23	0.34	0.21	0.58	0.27	0.22	0.17	0.12	0.25	0.18
	Testing ($N = 280$)	0.36	0.23	0.16	0.23	0.21	0.20	0.29	0.31	0.24	0.18	0.23	0.36

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Table 2 | Statistical indices of areal rainfall data

Data	Methods	Statistical indices						
		X_{mean}	X_{max}	X_{min}	S_x	C_v	C_{sx}	SE
Training	Thiessen polygon	1.16	20.01	0.00	2.77	2.38	3.22	0.13
	Spline	1.16	20.72	0.00	2.83	2.44	3.29	0.13
	Kriging	1.17	18.71	0.00	2.76	2.37	3.16	0.13
Cross-validation	Thiessen polygon	1.21	17.24	0.00	2.93	2.42	3.31	0.19
	Spline	1.16	16.63	0.00	2.77	2.38	3.19	0.18
	Kriging	1.19	17.61	0.00	2.94	2.46	3.33	0.19
Testing	Thiessen polygon	1.79	18.60	0.00	2.85	1.59	2.41	0.17
	Spline	1.80	18.39	0.00	2.87	1.60	2.39	0.17
	Kriging	1.82	19.66	0.00	2.95	1.62	2.63	0.18

performance is not perfect. Legates & McCabe (1999) suggested that various evaluation criteria (e.g., RMSE, MAE, NS, and APE) must be used to evaluate model performance. Table 3 shows mathematical expressions of performance evaluation criteria used in this study.

Table 3 | Mathematical expressions of performance evaluation criteria

Evaluation criteria	Equation
CC	$\frac{\frac{1}{n}\sum_{i=1}^{n}[y_i(x)-u_y][\hat{y}_i(x)-\hat{u}_y]}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i(x)-u_y]^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}[\hat{y}_i(x)-\hat{u}_y]^2}}$
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i(x)-\hat{y}_i(x)]^2}$
NS	$1-\frac{\sum_{i=1}^{n}[y_i(x)-\hat{y}_i(x)]^2}{\sum_{i=1}^{n}[y_i(x)-u_y]^2}$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i(x)-\hat{y}_i(x)\rvert$
APE	$\frac{\sum_{i=1}^{n}\lvert y_i(x)-\hat{y}_i(x)\rvert}{\sum_{i=1}^{n}y_i(x)}\times 100$

$y_i(x)$ = the observed hourly rainfall (mm); $\hat{y}_i(x)$ = the estimated hourly rainfall (mm); u_y = the mean of observed hourly rainfall (mm); $\hat{u}_y$ = the mean of estimated hourly rainfall (mm); and n = the total number of hourly rainfall values considered.

RESULTS AND DISCUSSION

Rainfall aggregation using data-driven models

The development of an optimal model is a major problem in data-driven modeling (Kisi 2007; Kim & Kim 2008b). Since the number of input–output nodes is problem dependent, there is no precise way of choosing the optimal number of hidden nodes. The model structure, therefore, is generally determined using a trial and error method (Coulibaly *et al.* 2001; Makarynskyy *et al.* 2005). Conventional data-driven models adopt one hidden layer for model construction, since it is well known that one hidden layer is enough to represent the nonlinear complex relationship (Kumar *et al.* 2002; Makarynskyy *et al.* 2005). The number of hidden nodes of data-driven models for rainfall aggregation was determined using a trial and error approach. Figure 4(a) shows the developed structure of SVM (12-12-1) comprising input (12 nodes), hidden (12 nodes), and output (1 node) layers for estimating areal rainfall in this study. Figure 4(b) shows the developed structure of GRNN (12-12-2-1) comprising input (12 nodes), hidden (12 nodes), summation (1 summation and 1 division node), and output (1 node) layers for estimating areal rainfall in this study.

The input data were decomposed by DWT to develop and apply WSVM and WGRNN models. The optimal decomposition level must be selected in advance to determine the performance of the model in the wavelet domain. Several researchers have used an empirical equation to determine the decomposition level (Nourani *et al.* 2009; Tiwari & Chatterjee 2010; Adamowski & Chan 2011; Nejad & Nourani 2012). In this study, the

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

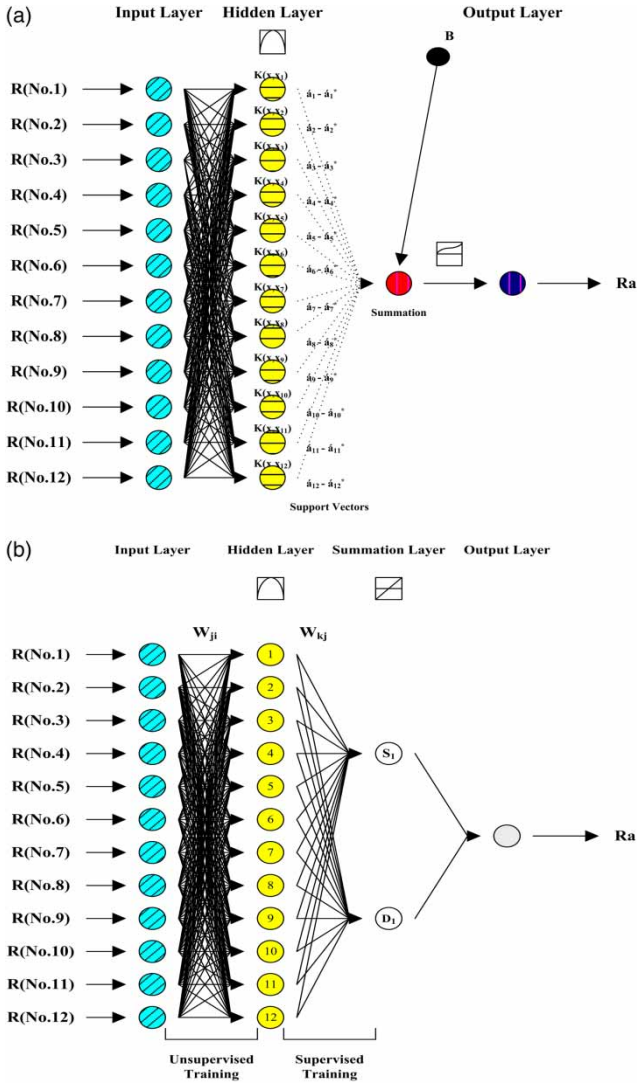

Figure 4 | Developed structure for estimating areal rainfall. (a) SVM (12-12-1), (b) GRNN (12-12-2-1).

decomposition level was determined using the following empirical equation (Nourani *et al.* 2009):

$$L = \text{int}[\log(N)] \tag{3}$$

where L = the decomposition level, N = the number of time series data, and int[·] = the integer-part function. In this study, two decomposition levels were obtained. Thus, input times series were decomposed using different mother wavelets, and details mode (D_1, D_2), and approximation mode (A_2) for individual input data were obtained for the training, cross-validation, and testing periods.

This study also aims at examining the effects of different mother wavelets on the efficiency of developed models. For this purpose, the performance of applied models was investigated for different mother wavelets, including Daubechies-8 (db8), Daubechies-10 (db10), Symmlet-8 (sym8), Symmlet-10 (sym10), Coiflet-6 (coif6), and Coiflet-12 (coif12). For discrete wavelet analysis, Daubechies wavelets have been commonly used as mother wavelets, and Symmlets and Coiflets wavelets have also been applied in hydrologic wavelet-based studies (Alikhani 2009; Adamowski & Sun 2010; Tiwari & Chatterjee 2010; Nejad & Nourani 2012; Evrendilek 2014; Santos *et al.* 2014). Daubechies, Symmlet, and Coiflet wavelets provide compact support (Vonesch *et al.* 2007; Mathworks 2014), indicating that the wavelets have non-zero basis functions over a finite interval, as well as full scaling and translational orthonormality properties (Popivanov & Miller 2002; de Artigas *et al.* 2006). These features are important for localizing events in the time-dependent signals (Popivanov & Miller 2002). Based on these features, Daubechies, Symmlet, and Coiflet wavelets were selected as mother wavelets in this study. Figure 5 shows an example of the original time series and sub-time series (D1, D2, and A2) decomposed using db10 wavelet for the training period.

Selection of effective wavelet components is important for model performance. Previous studies selected effective wavelet components using the CC between wavelet components and observed values (Alikhani 2009; Tiwari & Chatterjee 2010; Kisi & Cimen 2011). The effective wavelet components have also been selected using other methods, including Mallow's C_p (Okkan 2012; Okkan & Serbes 2013), CC, mutual information, Shannon entropy (Khanghah *et al.* 2012), and self-organizing map (Nourani *et al.* 2012). Several researchers also used all decomposed components as effective wavelet components (Adamowski & Sun 2010; Adamowski & Chan 2011; Kisi 2011; Adamowski & Prasher 2012).

To construct new input time series from the wavelet components, several methods have been used, including summing the effective components (Partal & Cigizoglu 2008; Alikhani 2009; Kisi 2010; Kisi & Cimen 2011), summing the components for different levels (Adamowski & Chan 2011; Adamowski & Prasher 2012), using all components for different levels without summing the components (Adamowski & Sun 2010; Kisi 2011), and using only effective components without summing them (Okkan 2012). Based

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

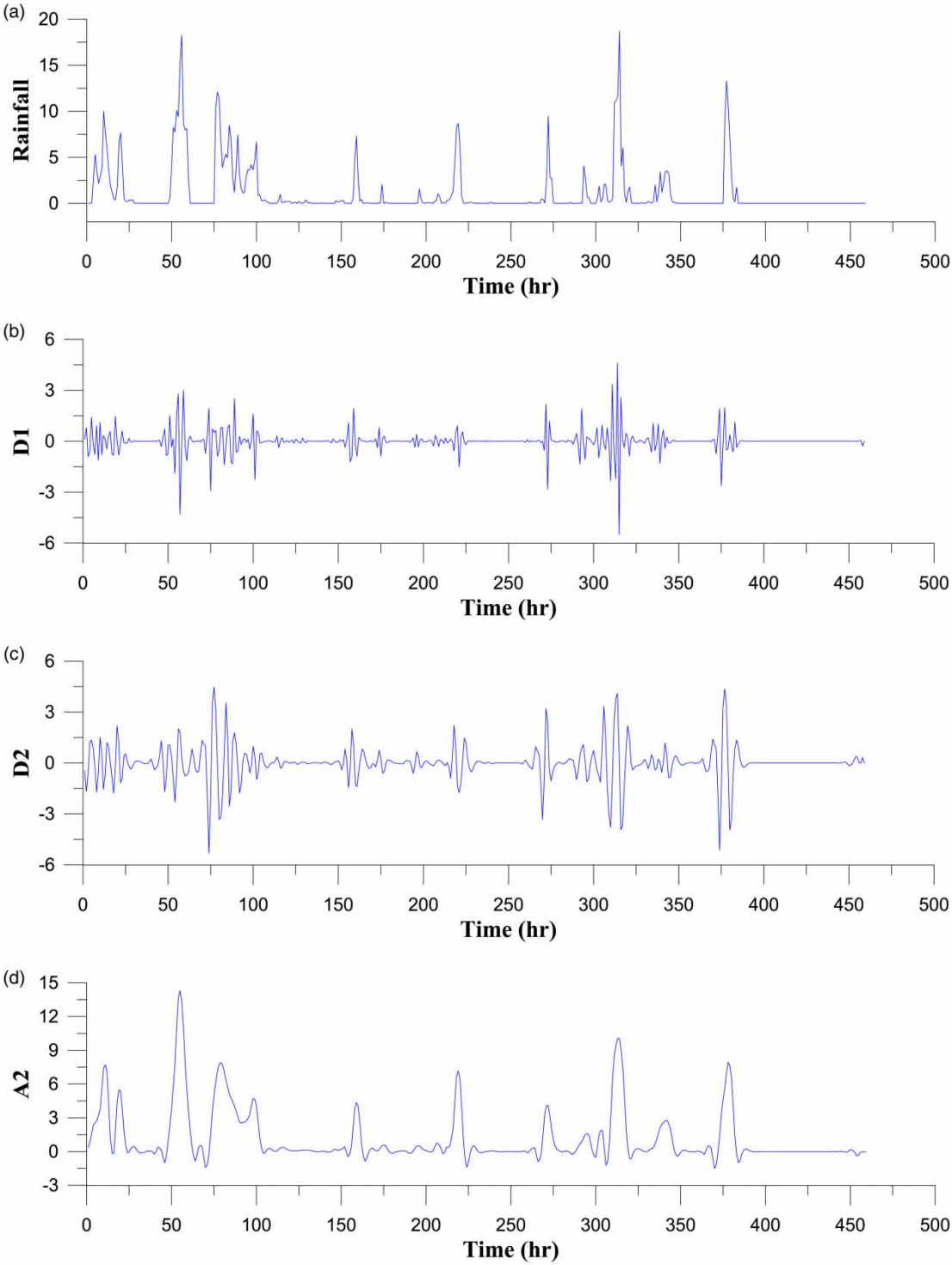

Figure 5 | Original and decomposed time series (D1, D2, and A2) using db10 wavelet for training period. (a) Original time series, (b) D1, (c) D2, (d) A2.

on the modeling strategies, one SVM, one GRNN, six WSVM, and six WGRNN models were developed for the rainfall aggregation.

Table 4 summarizes statistical results for rainfall aggregation models during the testing performance. It is clear from Table 4 that all the models generally perform well.

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Table 4 | Performance statistics for spatial aggregation of areal rainfall models during the testing performance

	Evaluation criteria				
Models	**CC**	**RMSE (mm)**	**NS**	**MAE (mm)**	**APE (%)**
SVM	0.950	0.958	0.895	0.516	31.347
WSVM_db8	0.967	0.770	0.932	0.486	29.493
WSVM_db10	0.972	0.711	0.942	0.437	25.546
WSVM_sym8	0.954	0.914	0.905	0.508	29.771
WSVM_sym10	0.962	0.881	0.913	0.486	28.843
WSVM_coif6	0.952	0.941	0.901	0.515	30.766
WSVM_coif12	0.955	0.924	0.903	0.510	30.889
GRNN	0.891	1.460	0.756	0.905	56.532
WRNNN_db8	0.935	1.119	0.858	0.525	34.540
WGRNN_db10	0.944	1.052	0.874	0.460	27.765
WGRNN_sym8	0.903	1.350	0.792	0.821	51.078
WGRNN_sym10	0.919	1.190	0.838	0.682	42.136
WGRNN_coif6	0.899	1.393	0.778	0.885	54.276
WGRNN_coif12	0.904	1.338	0.796	0.771	47.550

Comparison of SVM and WSVM models with different mother wavelets indicates that the results of WSVM models are better than those of the SVM model. Comparison of GRNN and WGRNN models with different mother wavelets also indicates that the results of WGRNN models are better than those of the GRNN model. Furthermore, it can be found from Table 4 that the results of SVM and WSVM are better than those of the GRNN and WGRNN with respect to different mother wavelets, respectively. Comparison of different mother wavelets reveals that db10 yields the best accuracy for the rainfall aggregation for the WSVM and WGRNN models. This indicates that wavelet decomposition using mother wavelet, db10, can improve the performance of SVM and GRNN models as compared with the other mother wavelets. These results are consistent with those reported by Seo *et al.* (2015).

Figure 6(a)–6(d) compare observed and estimated areal rainfall for testing data using the SVM and WSVM models with different mother wavelets, including db10, sym10, and coif12. It is clear from the fit-line equations that the SVM and WSVM models with different mother wavelets perform well with high correlations. It is clear from Figure 6 that the WSVM model with mother wavelet db10 shows the best accuracy. Figure 7(a)–7(d) compare observed and estimated areal rainfall for testing data using the GRNN and WGRNN models with different mother wavelets, including db10, sym10, and coif12. It is clear from the fit-line equations that the WGRNN model with mother wavelet db10 performs better than the other models. Comparison of SVM and GRNN models given in Figures 6 and 7 clearly reveals that the SVM and WSVM models are better than the GRNN and WGRNN models for estimating areal rainfall, respectively.

Disaggregation of areal rainfall using data-driven models

In this section, the WSVM and WGRNN models, which yielded the best performance for estimating areal rainfall, including SVM and GRNN, were used for disaggregating the areal rainfall. Therefore, the WSVM and WGRNN models with mother wavelets db10 and sym10 were used in this study. Only two performance evaluation criteria (CC and RMSE) were applied for disaggregating the areal rainfall. The number of hidden nodes of data-driven models for disaggregating the areal rainfall was also determined using a trial and error approach. Figure 8(a) shows the developed structure of SVM (1-12-12), comprising input (1 node), hidden (12 nodes), and output (12 nodes) layers for estimating the disaggregated rainfall in this study. Figure 8(b) shows the developed structure of GRNN (1-12-13-12), comprising input (1 node), hidden (12 nodes), summation (12 summation and 1 division nodes), and output (12 nodes) layers for estimating the disaggregated rainfall in this study.

Figure 9(a) and 9(b) show the influence of individual rainfall stations with respect to the performance evaluation criteria (CC and RMSE) of SVM and WSVM models during the test period. SVM and WSVM models were generally sensitive to individual rainfall stations, as seen from large fluctuations. The disaggregated rainfall for individual rainfall stations yielded very different performance based on evaluation criteria (CC and RMSE). For Samga (No. 7) station, the values of CC and RMSE are 0.786 and 2.22 mm for the SVM model, 0.827 and 2.04 mm for the WSVM model with mother wavelet db10, and 0.811 and 2.16 mm for the WSVM model with mother wavelet sym10, respectively. For Yiweon (No. 11) station, the values of CC and RMSE are 0.411 and 5.58 mm for the

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

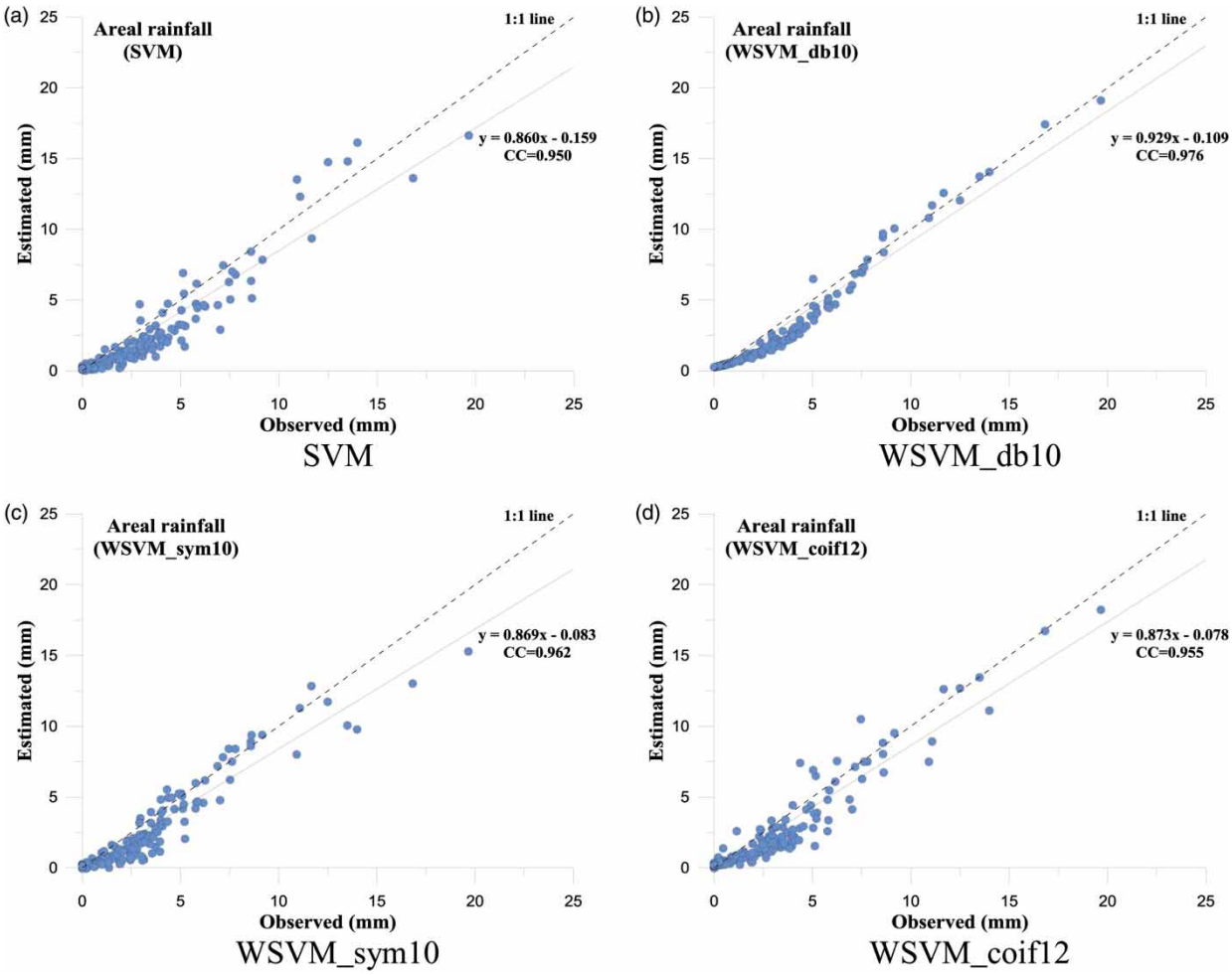

Figure 6 | Comparison of observed and estimated areal rainfalls using SVM and WSVM.

SVM model, 0.532 and 4.83 mm for the WSVM model with mother wavelet db10, 0.478 and 5.35 mm for the WSVM model with mother wavelet sym10, respectively. Figure 9(a) and 9(b) clearly show that the disaggregated rainfalls at Samga (No. 7) and Yiweon (No. 11) stations yield the best and worst accuracies among the 12 rainfall stations for the SVM and WSVM models. Results show that the SVM and WSVM models are capable of disaggregating the areal rainfall into individual point rainfall. The reliability of disaggregating the areal rainfall on individual rainfall stations, however, shows much difference. From Figure 9(a) and 9(b), it can be judged that the WSVM model with mother wavelet db10 is an optimal model for disaggregating the areal rainfall.

Figure 10(a) and 10(b) show the influence of individual rainfall stations on the performance evaluation criteria (CC and RMSE) of the GRNN and WGRNN models during the test period. Here also, the GRNN and WGRNN models were generally sensitive to individual rainfall stations, as seen from large fluctuations. The disaggregated rainfall for individual rainfall stations yielded very different performance based on evaluation criteria (CC and RMSE). For Samga (No. 7) station, the values of CC and RMSE are 0.679 and 2.32 mm for the GRNN model, 0.718 and 2.14 mm for the WGRNN model with mother wavelet db10, and 0.691 and 2.29 mm for the WGRNN model with mother wavelet sym10, respectively. For Yiweon (No. 11) station, the values of CC and RMSE are 0.303

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

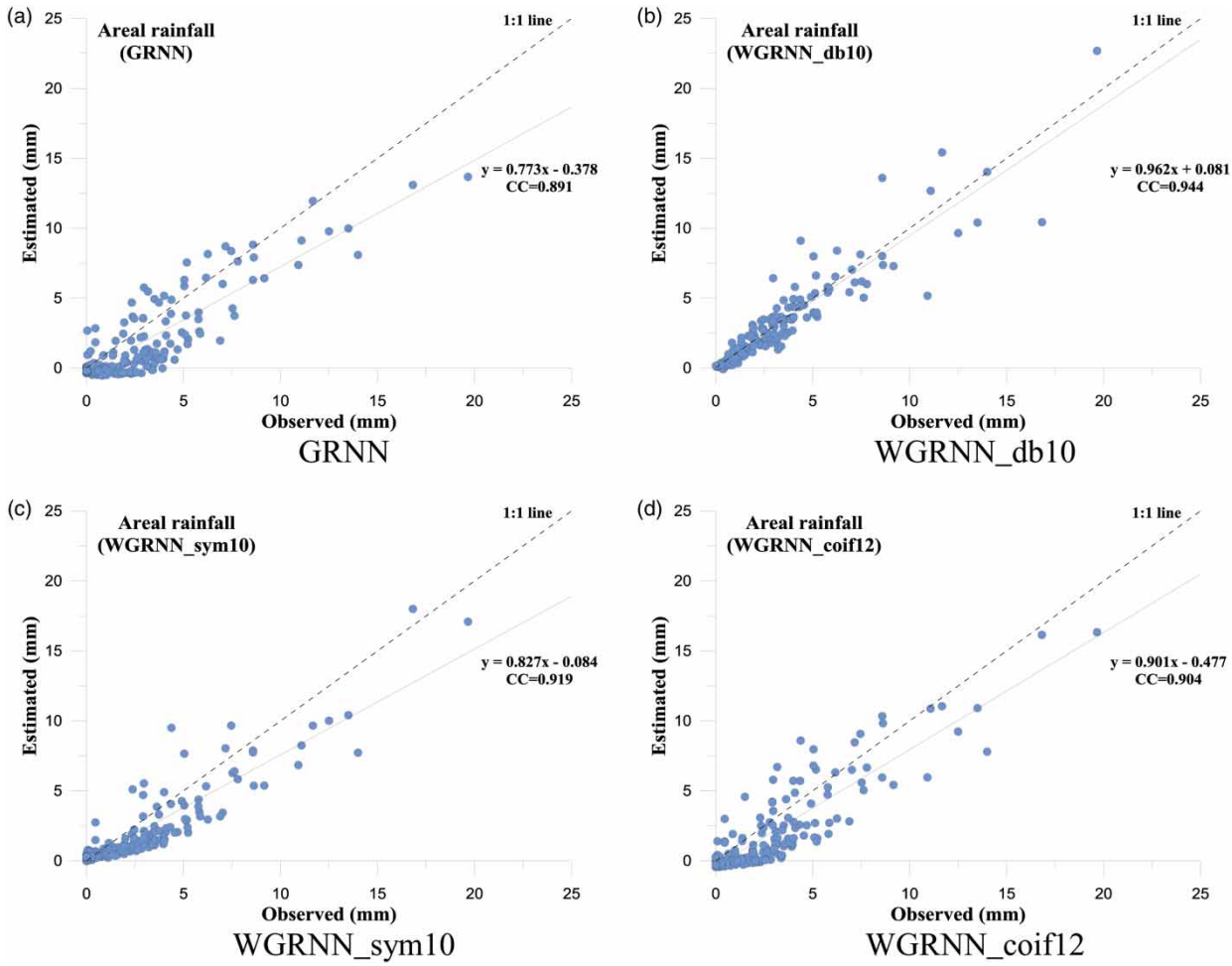

Figure 7 | Comparison of observed and estimated areal rainfalls using GRNN and WGRNN.

and 5.67 mm for the GRNN model, 0.396 and 4.97 mm for the WGRNN model with mother wavelet db10, 0.372 and 5.48 mm for the WGRNN model with mother wavelet sym10, respectively. Figure 10(a) and 10(b) show that the disaggregated rainfalls at Samga (No. 7) and Yiweon (No. 11) stations yield the best and worst accuracies among the 12 rainfall stations for the GRNN and WGRNN models. Results show that the GRNN and WGRNN models are capable of disaggregating the areal rainfall into individual point rainfall. The reliability of disaggregating the areal rainfall on individual rainfall stations, however, shows much difference. From Figure 10(a) and 10(b), it can be judged that the WGRNN model with mother wavelet db10 is an optimal model for disaggregating the areal rainfall.

The specific rainfall stations (e.g., No. 6, No. 8, No. 10, and No. 11), which have high skewed distributions in training and cross-validation data, did not generally show satisfactory results in the performance evaluation criteria (CC and RMSE (mm)) of SVM and GRNN models. However, wavelet composition techniques can overcome the weakness of SVM and GRNN models effectively. It can be found from this observation that data quality can affect the performances of data-driven models. This is in agreement with Tokar & Johnson (1999) and Sivakumar *et al.* (2002). Comparison of SVM and GRNN models indicates that the results of the SVM model are better than those of the GRNN model for disaggregating the areal rainfall. Also, comparison of WSVM and WGRNN models with mother wavelets

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

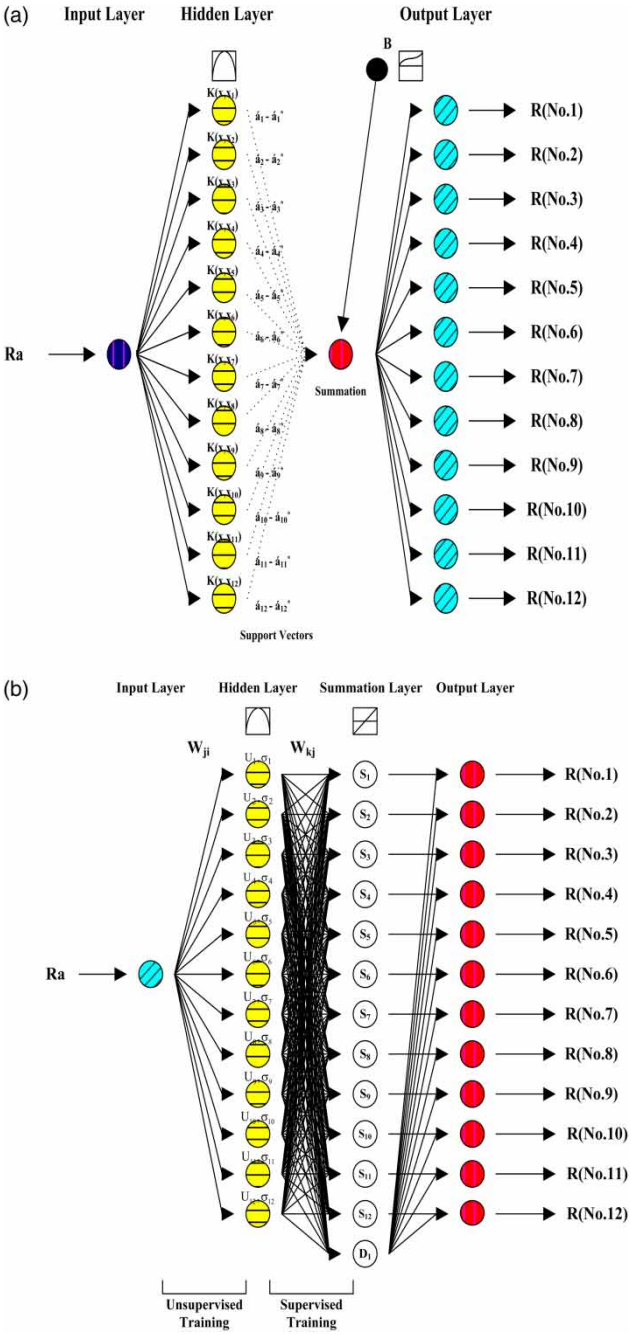

Figure 8 | Developed structure for estimating spatial disaggregated rainfall.

db10 and sym10 indicates that the results of the WSVM model with mother wavelets db10 and sym10 are better than those of the WGRNN model with mother wavelets db10 and sym10 for disaggregating the areal rainfall, respectively.

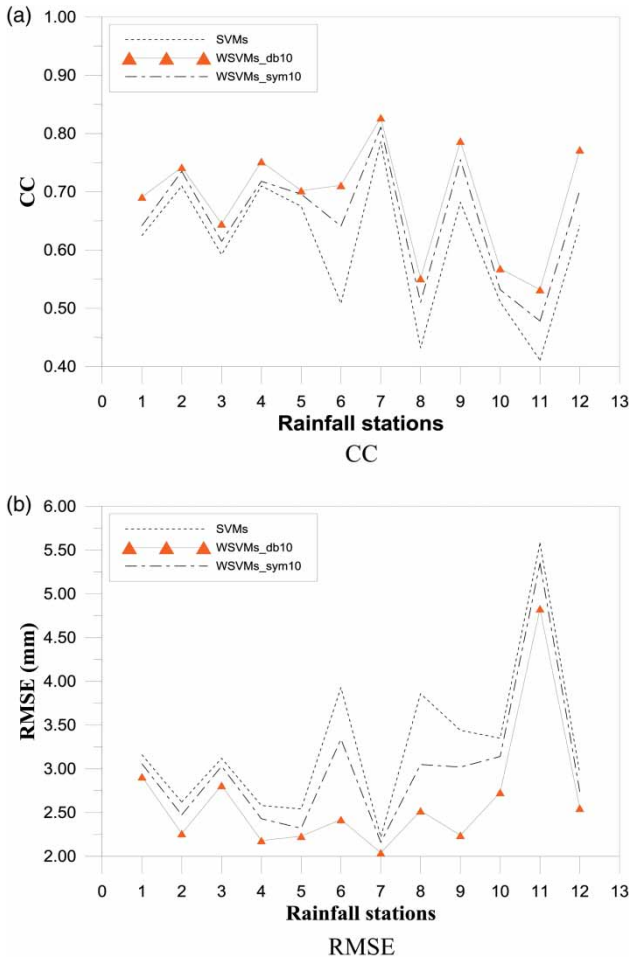

Figure 9 | Influence of individual rainfall stations for SVM and WSVM models (test period).

CONCLUSIONS

This study develops and evaluates the combination of wavelet decomposition and data-driven models for aggregation and disaggregation of rainfall in the Bocheong-stream catchment, an IHP representative catchment, Republic of Korea.

The SVM and GRNN models are used to estimate areal rainfall and individual point rainfall. Wavelet decomposition is employed and sub-components are used as input to SVM and GRNN to obtain WSVM and WGRNN models, respectively. Comparison of SVM and WSVM models with different mother wavelets indicates that the results of WSVM models with different mother wavelets are better than those of the SVM model. Comparison of GRNN and WGRNN models with different mother

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

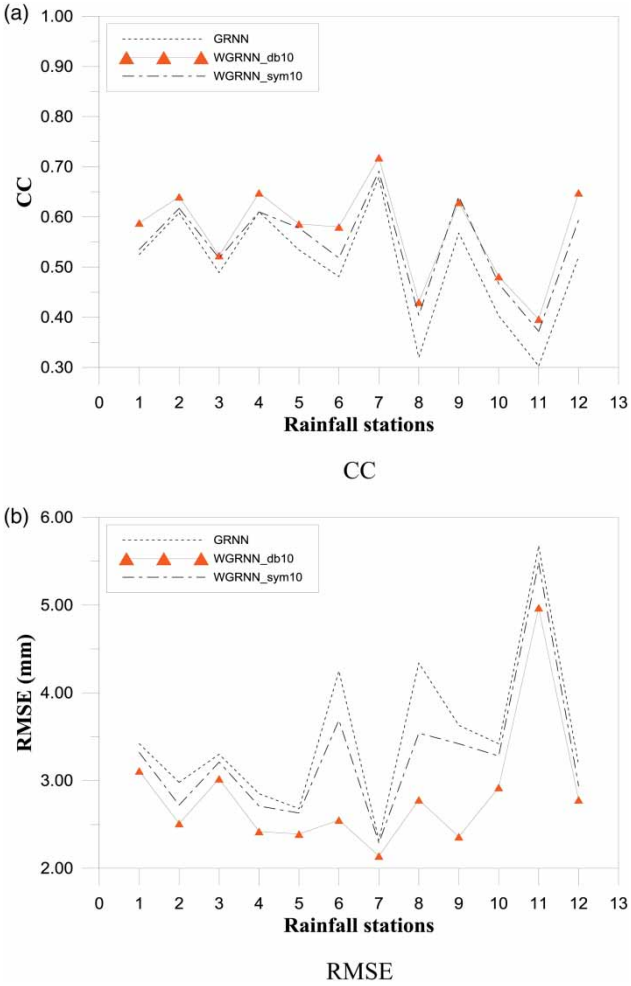

Figure 10 | Influence of individual rainfall stations for GRNN and WGRNN models (test period).

wavelets indicates that the results of WGRNN models with different mother wavelets are better than those of the GRNN model.

The WSVM models with mother wavelet db10 yield the best performance for rainfall aggregation among SVM and WSVM models with different mother wavelet models. The WGRNN models with mother wavelet db10 yield the best performance for rainfall aggregation among GRNN and WGRNN models with different mother wavelet models.

The SVM, GRNN, WSVM (db10 and sym10), and WGRNN (db10 and sym10) models are used for estimating the disaggregated rainfall. The SVM, GRNN, WSVM (db10 and sym10) and WGRNN (db10 and sym10) models are generally found to be sensitive to individual rainfall stations. The disaggregated rainfall at Samga (No. 7) yields the best results among the 12 rainfall stations for the SVM, GRNN, WSVM, and WGRNN models. The disaggregated rainfall at Yiweon (No. 11) station yields the worst results among the 12 rainfall stations for the SVM, GRNN, WSVM, and WGRNN models.

Comparison of the SVM and GRNN models indicates that the results of the SVM model are better than those of the GRNN model for disaggregating the areal rainfall. Comparison of the WSVM and WGRNN models with mother wavelets db10 and sym10 indicates that the results of the WSVM model with mother wavelets db10 and sym10 are better than those of the WGRNN model with mother wavelets db10 and sym10 for disaggregating the areal rainfall, respectively. The WSVM and WGRNN models with mother wavelet db10 are found to be optimal models for disaggregating the areal rainfall in this study.

REFERENCES

Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **407** (1–4), 28–40.

Adamowski, J. & Prasher, S. O. 2012 Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *J. Water Land Dev.* **17**, 89–97.

Adamowski, J. & Sun, K. 2010 Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* **390** (1–2), 85–91.

AghaKouchak, A., Bárdossy, A. & Habib, E. 2010 Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. *Adv. Water Resour.* **33** (6), 624–634.

Alikhani, A. 2009 Combination of neuro fuzzy and wavelet model usage in river engineering. *Int. J. Energy Environ.* **3** (3), 122–134.

Apaydin, H., Sonmez, F. K. & Yildirim, Y. E. 2004 Spatial interpolation techniques for climate data in the GAP region in Turkey. *Clim. Res.* **28** (1), 31–40.

ASCE Task Committee 1993 Criteria for evaluation of watershed models. *J. Irrig. Drain Eng.* **119** (3), 429–442.

Burian, S. J. & Durrans, S. R. 2002 Evaluation of an artificial neural network rainfall disaggregation model. *Water Sci. Technol.* **45** (2), 99–104.

Burian, S. J., Durrans, S. R., Tomić, S., Pimmel, R. L. & Wai, C. N. 2000 Rainfall disaggregation using artificial neural networks. *J. Hydrol. Eng.* **5** (3), 299–307.

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Burian, S. J., Durrans, S. R., Nix, S. J. & Pitt, R. E. 2001 Training artificial neural networks to perform rainfall disaggregation. *J. Hydrol. Eng.* **6** (1), 43–51.

Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Phys. Chem. Earth* **31** (18), 1164–1171.

Catalão, J. P. S., Pousinho, H. M. I. & Mendes, V. M. F. 2011 Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Trans. Power Syst.* **26** (1), 137–144.

Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill, New York, USA.

Connolly, R. D., Schirmer, J. & Dunn, P. K. 1998 A daily rainfall disaggregation model. *Agr. Forest. Meteorol.* **92** (2), 105–117.

Coulibaly, P., Anctil, F., Aravena, R. & Bobée, B. 2001 Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* **37** (4), 885–896.

Dawson, C. W. & Wilby, R. L. 2001 Hydrological modelling using artificial neural networks. *Prog. Phys. Geog.* **25** (1), 80–108.

de Artigas, M. Z., Elias, A. G. & de Campra, P. F. 2006 Discrete wavelet analysis to assess long-term trends in geomagnetic activity. *Phys. Chem. Earth* **31** (1–3), 77–80.

Durrans, S., Burian, S. J., Nix, S. J., Hajji, A., Pitt, R. E., Fan, C. Y. & Field, R. 1999 Polynomial-based disaggregation of hourly rainfall for continuous hydrologic simulation. *J. Am. Water Resour. Assoc.* **35** (5), 1213–1221.

Evrendilek, F. 2014 Assessing neural networks with wavelet denoising and regression models in predicting diel dynamics of eddy covariance-measured latent and sensible heat fluxes and evapotranspiration. *Neural Comput. Appl.* **24** (2), 327–337.

Genest, C., Favre, A., Béliveau, J. & Jacques, C. 2007 Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* **43** (9), W09401.1–W09401.12.

Glasbey, C. A., Cooper, G. & McGechan, M. B. 1995 Disaggregation of daily rainfall by conditional simulation from a point process model. *J. Hydrol.* **165** (1–4), 1–9.

González-Audícana, M., Otazu, X., Fors, O. & Seco, A. 2005 Comparison between Mallat's and the 'à trous' discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. *Int. J. Remote Sens.* **26** (3), 595–614.

Goovaerts, P. 2000 Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **228** (1–2), 113–129.

Gyasi-Agyei, Y. 2005 Stochastic disaggregation of daily rainfall into one-hour time scale. *J. Hydrol.* **309** (1), 178–190.

Haykin, S. 2009 *Neural Networks and Learning Machines*, 3rd edn. Prentice Hall, Upper Saddle River, NJ, USA.

Hershenhorn, J. & Woolhiser, D. A. 1987 Disaggregation of daily rainfall. *J. Hydrol.* **95** (3–4), 299–322.

Izadifar, Z. & Elshorbagy, A. 2010 Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models. *Hydrol. Process.* **24** (23), 3413–3425.

Khanghah, T. R., Nourani, V., Parhizkar, M. & Sharghi, E. 2012 Application of information content to extract wavelet-based feature of rainfall-runoff process. In: *Proceedings of the 12th WSEAS International Conference on Applied Computer Science*, WSEAS, Greece, pp. 148–153.

Kim, S. & Kim, H. S. 2008a Uncertainty reduction of the flood stage forecasting using neural networks model. *J. Am. Water Resour. Assoc.* **44** (1), 148–165.

Kim, S. & Kim, H. S. 2008b Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *J. Hydrol.* **351** (3–4), 299–317.

Kim, S. & Singh, V. P. 2015 Spatial disaggregation of areal rainfall using two different artificial neural networks models. *Water* **7** (6), 2707–2727.

Kim, S., Shiri, J. & Kisi, O. 2012 Pan evaporation modeling using neural computing approach for different climatic zones. *Water Resour. Manage.* **26** (11), 3231–3249.

Kim, S., Seo, Y. & Singh, V. P. 2013a Assessment of pan evaporation modeling using bootstrap resampling and soft computing methods. *J. Comput. Civ. Eng.* **29** (5).

Kim, S., Shiri, J., Kisi, O. & Singh, V. P. 2013b Estimating daily pan evaporation using different data-driven methods and lag-time patterns. *Water Resour. Manage.* **27** (7), 2267–2286.

Kisi, O. 2006 Generalized regression neural networks for evapotranspiration modeling. *Hydrol. Sci. J.* **51** (6), 1092–1105.

Kisi, O. 2007 Evapotranspiration modeling from climatic data using a neural computing technique. *Hydrol. Process.* **21** (14), 1925–1934.

Kisi, O. 2010 Wavelet regression model for short-term streamflow forecasting. *J. Hydrol.* **389** (3–4), 344–353.

Kisi, O. 2011 A combined generalized regression neural network wavelet model for monthly streamflow prediction. *KSCE J. Civ. Eng.* **15** (8), 1469–1479.

Kisi, O. & Cimen, M. 2011 A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J. Hydrol.* **399** (1–2), 132–140.

Kisi, O. & Cimen, M. 2012 Precipitation forecasting by using wavelet-support vector machine conjunction model. *Eng. Appl. Artif. Intell.* **25** (4), 783–792.

Knoesen, D. & Smithers, J. 2009 The development and assessment of a daily rainfall disaggregation model for South Africa. *Hydrol. Sci. J.* **54** (2), 217–233.

Koutsoyiannis, D. & Xanthopoulos, T. 1990 A dynamic model for short-scale rainfall disaggregation. *Hydrol. Sci. J.* **35** (3), 303–322.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W. & Pruitt, W. O. 2002 Estimating evapotranspiration using artificial neural networks. *J. Irrig. Drain. Eng.* **128** (4), 224–233.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35** (1), 233–241.

Ly, S., Charles, C. & Degré, A. 2011 Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci.* **15**, 2259–2274.

Makarynskyy, O., Pires-Silva, A. A., Makarynska, D. & Ventura-Soares, C. 2005 Artificial neural networks in wave predictions at the west coast of Portugal. *Comput. Geosci.* **31** (4), 415–424.

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Mallat, S. G. 1989 A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **11** (7), 674–693.

Mathworks 2014 *Wavelet toolbox user's guide*. The Mathworks, Inc. http://www.mathworks.com/help/pdf_doc/wavelet/wavelet_ug.pdf (accessed 6 July 2014).

Ministry of Construction and Transportation 1982–2007 *Collection and fundamental analysis of hydrologic data of the representative basin*. International Hydrological Program (IHP), South Korea.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models, Part 1 – a discussion of principles. *J. Hydrol.* **10** (3), 282–290.

Nason, G. 2010 *Wavelet Methods in Statistics with R*. Springer, New York, USA.

Nejad, F. H. & Nourani, V. 2012 Elevation of wavelet denoising performance via an ANN-based streamflow forecasting model. *Int. J. Comput. Sci. Manage. Res.* **1** (4), 764–770.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2009 A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Eng. Appl. Artif. Intell.* **22** (3), 466–472.

Nourani, V., Kisi, O. & Komasi, M. 2011 Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *J. Hydrol.* **402** (1–2), 41–59.

Nourani, V., Parhizkar, M., Khanghah, T. R., Baghanam, A. H. & Sharghi, E. 2012 Wavelet-based feature extraction of rainfall-runoff process via self-organizing map. In: *Proceedings of the 12th WSEAS International Conference on Applied Computer Science*, WSEAS, Greece, 101–106.

Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. 2014 Applications of hybrid wavelet–Artificial Intelligence models in hydrology: a review. *J. Hydrol.* **514**, 358–377.

Okkan, U. 2012 Using wavelet transform to improve generalization capability of feed forward neural networks in monthly runoff prediction. *Sci. Res. Essays* **7** (17), 1690–1703.

Okkan, U. & Serbes, Z. A. 2013 The combined use of wavelet transform and black box models in reservoir inflow modeling. *J. Hydrol. Hydromech.* **61** (2), 112–119.

Olsson, J. 1998 Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrol. Earth Syst. Sci.* **2** (1), 19–30.

Olsson, J. & Berndtsson, R. 1998 Temporal rainfall disaggregation based on scaling properties. *Water Sci. Technol.* **37** (11), 73–79.

Ormsbee, L. E. 1989 Rainfall disaggregation model for continuous hydrologic modeling. *J. Hydraul. Eng.* **115** (4), 507–525.

Partal, T. & Cigizoglu, H. K. 2008 Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. *J. Hydrol.* **358** (3–4), 317–331.

Perica, S. & Foufoula-Georgiou, E. 1996 Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. *J. Geophys. Res.* **101** (D21), 26347–26361.

Popivanov, I. & Miller, R. J. 2002 Similarity search over time-series data using wavelets. In: *Proceedings of 18th International Conference on Data Engineering*, San Jose, CA, USA, pp. 212–221.

Principe, J. C., Euliano, N. R. & Lefebvre, W. C. 2000 *Neural and Adaptive Systems: Fundamentals through Simulation*. John Wiley & Sons, Inc., New York, USA.

Rajaee, T. 2010 Wavelet and neuro-fuzzy conjunction approach for suspended sediment prediction. *Clean-Soil, Air, Water* **38** (3), 275–286.

Rajaee, T., Nourani, V., Mohammad, Z. K. & Kisi, O. 2011 River suspended sediment load prediction: application of ANN and wavelet conjunction model. *J. Hydrol. Eng.* **16** (8), 613–627.

Santos, C. A. G., Freire, P. K. M. M., Silva, G. B. L. & Silva, R. M. 2014 Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias reservoir. In: *Proceedings of the International Association of Hydrological Sciences*, Bologna, Italy, pp. 100–105.

Seo, Y., Kim, S., Kisi, O. & Singh, V. J. 2015 Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **520**, 224–243.

Sharma, D., Das Gupta, A. & Babel, M. S. 2007 Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrol. Earth Syst. Sci.* **11** (4), 1373–1390.

Singh, V. P. 1992 *Elementary Hydrology*. Prentice Hall, Upper Saddle River, NJ, USA.

Sivakumar, B., Sorooshian, S., Gupta, H. V. & Gao, X. 2001 A chaotic approach to rainfall disaggregation. *Water Resour. Res.* **37** (1), 61–72.

Sivakumar, B., Jayawardena, A. W. & Fernando, T. M. K. G. 2002 River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *J. Hydrol.* **265** (1–4), 225–245.

Smith, M. 1993 *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold Company, New York, USA.

Socolofsky, S., Adams, E. E. & Entekhabi, D. 2001 Disaggregation of daily rainfall for continuous watershed modeling. *J. Hydrol. Eng.* **6** (4), 300–309.

Specht, D. F. 1991 A general regression neural network. *IEEE Trans. Neural Network* **2** (6), 568–576.

Stone, M. 1974 Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. B Met.* **36** (2), 111–147.

Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. 2003 Estimating actual evapotranspiration from limited climatic data using neural computing technique. *J. Irrig. Drain. Eng.* **129** (3), 214–218.

Tait, A., Henderson, R., Turner, R. & Zheng, X. 2006 Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *Int. J. Climatol.* **26** (14), 2097–2115.

Tiwari, M. K. & Chatterjee, C. 2010 Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J. Hydrol.* **394** (3–4), 458–470.

Tokar, A. S. & Johnson, P. A. 1999 Rainfall-runoff modeling using artificial neural networks. *J. Hydrol. Eng.* **4** (3), 232–239.

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest

Tripathi, S., Srinivas, V. V. & Nanjundish, R. S. 2006 Downscaling of precipitation for climate change scenarios: a support vector machine approach. *J. Hydrol.* **330** (3–4), 621–640.

Tsoukalas, L. H. & Uhrig, R. E. 1997 *Fuzzy and Neural Approaches in Engineering*. John Wiley & Sons, New York, USA.

Vapnik, V. N. 2010 The Nature of Statistical Learning Theory, 2nd edn. Springer, New York, USA.

Venugopal, V., Foufoula-Georgiou, E. & Sapozhnikov, V. 1999 A space-time downscaling model for rainfall. *J. Geophys. Res.* **104** (D16), 19705–19721.

Vonesch, C., Blu, T. & Unser, M. 2007 Generalized Daubechies wavelet families. *IEEE Trans. Signal Process.* **55** (9), 4415–4429.

Wang, W. & Ding, J. 2003 Wavelet network model and its application to the prediction of hydrology. *Nat. Sci.* **1** (1), 67–71.

Wang, W., Jin, J. & Li, Y. 2009 Prediction of inflow at Three Gorges Dam in Yangtze River with wavelet network model. *Water Res. Manage.* **23** (13), 2791–2803.

Wasserman, P. D. 1993 *Advanced Methods in Neural Computing*. Van Nostrand Reinhold, New York, USA.

Wei, S., Song, J. & Khan, N. I. 2012 Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. *Hydro. Process.* **26** (2), 281–296.

Wu, C. L., Chau, K. W. & Fan, C. 2010 Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J. Hydrol.* **389** (1–2), 146–167.

Zhang, L. & Singh, V. J. 2007 Bivariate rainfall frequency distributions using Archimedean copulas. *J. Hydrol.* **332** (1–2), 93–109.

Zhang, J., Murch, R., Ross, M., Ganguly, A. & Nachabe, M. 2008 Evaluation of statistical rainfall disaggregation methods using rain-gauge information for West-Central Florida. *J. Hydrol. Eng.* **13** (12), 1158–1169.

First received 22 October 2015; accepted in revised form 30 January 2016. Available online 1 March 2016

Downloaded from http://iwaponline.com/hr/article-pdf/48/1/99/367220/nh0480099.pdf
by guest