Runoff Prediction in Ungauged Watersheds Using Remote Sensor Datasets

Jong Pil Kim[*1], Won Kim[2], Il-Won Jung[3], Gwangseob Kim[4]

[1,2]Department of Hydro Science and Engineering, Korea Institute of Civil Engineering and Building Technology, Goyans-si, Gyenggi-do, Republic of Korea

[3]Climate Research Department, APEC Climate Center, Haeundae-gu, Busan, Republic of Korea

[4]School of Architecture and Civil Engineering, Kyungpook National University, Buk-gu, Daegu, Republic of Korea

[*1]jpkim@kict.re.kr; [2]wonkim@kict.re.kr; [3]bobilwon@apcc21.org; [4]kimgs@knu.ac.kr

***Abstract*-Runoff prediction in ungauged watersheds is regarded as one of the major issues in contemporary hydrology. This study examined the utility of remote sensor datasets for parameter regionalization. A distributed hydrologic model, the Coupled Routing and Excess Storage (CREST) model, was employed to simulate runoff conditions in the mountainous watersheds over South Korea. In gauged watersheds, the relationships between the optimized parameter set of the hydrologic model and the physiographic properties of the gauged watersheds were investigated using multiple linear regressions. The regression parameters for the ungauged watersheds were then validated and assessed. Results demonstrated that the hydrologic model and the proposed regression equations could acceptably simulate the discharge in both gauged and ungauged watersheds. However, they provided somewhat biased discharge for all the ungauged watersheds. In further studies, these biases should be reduced by investigating other watersheds and finding physiographic properties highly related to the model parameters.**

Keyword- Runoff Prediction; Ungauged Watersheds; TRMM; Multi-satellite; Regionalization

I. INTRODUCTION

Knowledge of hydrologic responses from a watershed is very important to the management of water resources and to cope with water-related disasters, including floods and droughts. General runoff predictions can be achieved by hydrological modelling, which includes preparation of input datasets, implementation of hydrologic simulations, and model calibration and validation. For a particular watershed, a hydrological model simulates runoff discharge and estimates an optimal parameter set by comparing the output with physical measurements. This parameter set can be applied to predict another runoff event in the same watershed.

However, hydrometeorological records including discharge, temperature, humidity, sunshine, and wind are insufficient for hydrological modelling in many watersheds in which precipitation is relatively abundant; this interferes with accurate and reliable prediction of runoff discharge. Moreover, recent hydrologic models contain complex nonlinear equations to simulate more practical watershed responses [1]. Thus, many parameters in such models must be estimated by calibration due to the deficits of measurements.

Model parameters can be transferred by using different regionalization methods, which transfer a parameter set from either the nearest watersheds or a watershed with the most similar physiographic properties to the locations of interest ([2, 3]), to classify watersheds by climatology ([4, 5]), or to use the mean values of parameters from neighbourhood watersheds. Most widely-used regionalization methods use linear regression equations to evaluate the relationship between model parameters and the physiographic attributes of watersheds such as drainage area, elevation, slope, land cover, and soil type ([6, 7, 8]).

The purpose of this study is to examine the utility of remote sensor datasets for runoff prediction in ungauged watersheds. To simulate runoff discharge in the watersheds over South Korea a distributed hydrological model, the Coupled Routing and Excess Storage (CREST) model [9], was employed. Hydrological simulation was implemented in gauged watersheds, and an optimal parameter set was estimated by model calibration. In addition, linear relationships between the model parameters and watersheds attributes were explored. New parameters from regression equations were generated for the ungauged watersheds. Runoff discharge with the estimated parameters was predicted and assessed by comparing observations. Fig. 1 shows the flowchart of this study.

II. STUDY AREA AND DATA

A. Study Watersheds

Eight watersheds over South Korea were selected for hydrological modelling. All eight watersheds are located in the upstream regions of multi-purpose dams and are no directly affected by any artificial structures such as dams, bridge, buildings, asphalt, etc. The drainage areas of the study watersheds vary from 103.1 km^2 to 6651.4 km^2. Six of the eight watersheds were regarded as gauged for parameter regionalization, and two watersheds were regarded as ungauged for validation (Fig. 2).

DOI: 10.5963/JWRHE0403007

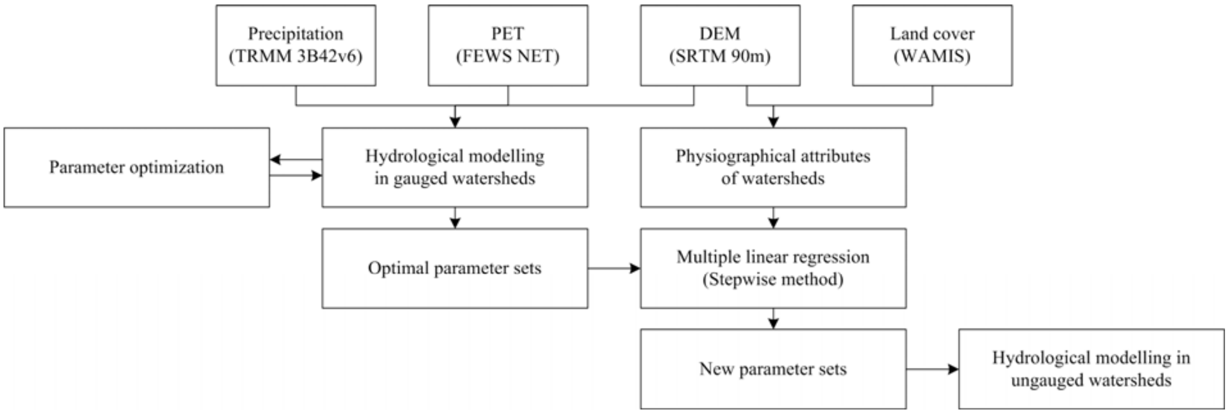

Fig. 1 Flowchart of this study

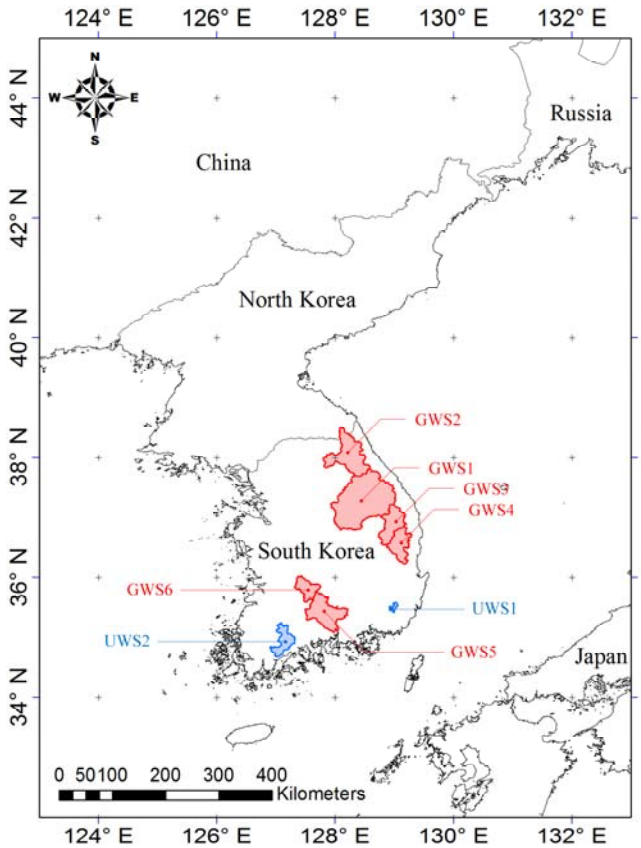

Fig. 2 Location map of the study area regarding gauged (red) and ungauged (blue) watersheds

B. Remotely-sensed Precipitation

The Goddard Space Flight Centre (GSFC) of the National Aeronautics and Space Administration (NASA) provides gridded quasi-global precipitation products as a result of the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) Algorithm [10]. The TMPA precipitation data is available in both near-real time and post-real time versions. The near-real time product, 3B42RT, is generated approximately 6-9 hours after observation, and the post-real time product, 3B42v6, is computed approximately 15 days after the end of each month.

The TMPA method combines precipitation estimates from four passive microwave sensors, including the TRMM Microwave Imager (TMI), the Special Sensor Microwave/Imager (SSM/I), the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E), and the Advanced Microwave Sounding Unit-B (AMSU-B). Infrared estimates from geosynchronous earth orbit (GEO) satellites are merged to fill any data gaps [11].

Unlike the near-real time product, 3B42v6 makes use of additional sources including the TRMM Combined Instruments (TCI) [12] from the TMI and the TRMM Precipitation Radar (PR), the Global Precipitation Climatology Project (GPCP) monthly rain gauge analysis [13] developed by the Global Precipitation Climatology Centre (GPCC), and the Climate Assessment and Monitoring System (CAMS) monthly rain gauge analysis [14, 15] developed by the Climate Prediction Centre

DOI: 10.5963/JWRHE0403007

(CPC). The multi-satellite merged microwave-infrared precipitation estimates are calibrated with TCI and then GPCP, and CAMS datasets are utilized to correct the bias for each calendar month in order to produce the final 3B42v6 product with 0.25° × 0.25° latitude-longitude at three-hour intervals.

C. Potential Evapotranspiration

A daily potential evapotranspiration (PET) dataset from the Famine Early Warning System Network (FEWS NET) was applied to the hydrologic model as meteorological input forcing. The PET data of FEWS NET is calculated from climate parameter data that was extracted from the Global Data Assimilation System (GDAS) analysis fields. As input to the PET computation, the GDAS fields include air temperature, atmospheric pressure, wind speed, relative humidity, and solar radiation. The daily PET is also calculated by the Penman-Monteith method on the basis of 1° × 1° grids on a global scale. Further details on this PET dataset can be found at the following website: http://earlywarning.usgs.gov/fews/.

D. Digital Elevation

The Shuttle Radar Topography Mission (SRTM) obtains land elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of the global land surface. SRTM is a joint research project between the National Geospatial Intelligence Agency (NGA) and NASA. The SRTM DEM dataset is available as three arc second (approximately 90 m resolution at the equator) over 80% of the globe, and is provided in 5° × 5° tiles for easy downloading and use. This study estimated the topographic parameters of the watersheds from the SRTM 90 m DEM data.

III. HYDROLOGICAL MODELLING

The Coupled Routing and Excess Storage (CREST) distributed hydrologic model [9] was employed to simulate runoff discharge in the selected mountainous watersheds. The model allows the computation of water balance components including infiltration, evaporation, runoff generation, and flow routing processes. The main components of the model are summarized as follows: 1) data flow module based on cell-to-cell routing; 2) three different layers within the soil profile that affect the maximum storage available in the soil layers; 3) coupling between the runoff generation and routing components via feedback mechanisms [16].

In this study, hydrological simulations were implemented by the CREST model from Jan-01-2004 to Dec-31-2009 at daily time-steps, and the optimal parameter sets were obtained by using the Adaptive Random Search (ARS) module provided in the model. As objective functions for calibration and validation, the Nash-Sutcliffe Coefficient of Efficiency (NSCE) and percent bias (PBIAS) were employed and computed as follows:

$$NSCE = 1 - \frac{\sum (Q_s - Q_o)^2}{\sum (Q_o - \overline{Q}_o)^2} \tag{1}$$

$$PBIAS = \frac{\sum Q_s - \sum Q_o}{\sum Q_o} \times 100(\%) \tag{2}$$

Where Q_s = simulated discharge; and Q_o = observed discharge.

IV. PARAMETER REGIONALIZATION

A. Physiographical Properties of the Watersheds

Watershed properties can be estimated from digital elevation and land cover maps. Drainage area is the entire geographical area drained by a river and its tributaries that can be identified by tracing a line along the ridge between two adjacent watersheds. The mean elevation and mean slope are the mean values of DEM and slopes at grid points within a watershed. The longest path is the distance from the outlet of a watershed to the furthest point. River length is the distance from the outlet to the start of the main stream. Shape factor is the ratio of river length to the diameter of a circle encompassing the drainage area of a watershed. The elongation ratio is defined as the ratio of the diameter of a circle with the same area as that of the watershed to the maximum watershed length. The watershed properties of the six gauged watersheds (GWS1-6) and two ungauged watersheds (UWS1-2) are presented in Table 2.

B. Multiple Linear Regression

The relationships between the model parameters and watershed properties including drainage area, elevation, slope, longest path, river length, elongation ratio, impervious area, forest area, paddy field, and crop land were computed by multiple linear regressions in the six gauged watersheds. New parameter sets were then generated by the regression equations and physiographic properties of the six gauged watersheds and two ungauged watersheds.

DOI: 10.5963/JWRHE0403007

In gauged watersheds, the relationship between the *i*th model parameter (ψ_i) as a dependent variable, and the watershed properties (Φ_j) as independent variables, are expressed as:

$$\psi_i = \alpha_i + \beta_{i,1}\phi_1 + \cdots + \beta_{i,j}\phi_j + \varepsilon_i \tag{3}$$

Where α_i = intercept; $\beta_{i,j}$ = coefficient; and ε = error term.

An important process in building linear regression models is to determine which variables affect the dependent variable. There are several model-selection methods including forward selection, backward elimination, stepwise method, etc.; we utilized the stepwise method to select a linear regression model.

TABLE 1 CREST MODEL PARAMETERS REQUIRING OPTIMIZATION

Parameter	**Description**	**Range**
coeM	Slope flow speed multiplier	1.0-150.0
expM	Slope flow speed exponent	0.2-0.9
River	Multiplier used to convert slope flow speed to channel flow speed	0.1-3.0
Under	Multiplier used to convert slope flow speed to interflow speed	0.01-1.0
LeakO	Overland reservoir discharge multiplier	0.1-1.0
LeakI	Interflow reservoir discharge multiplier	0.01-0.5
Th	Flow accumulation needed for a cell to be marked as a channel cell	10.0-100.0
GM	Change in DEM used to calculate the slope when the DEM for the downstream cell is higher than the upstream cell, or when the downstream cell is a nodata/outside region cell	0.5-2.0
pWm	Maximum soil water capacity of three soil layers	10.0-500.0
pB	Exponent of the variable infiltration curve	0.05-1.5
pIM	Impervious area ratio	0.0-100.0
pKE	Multiplier to convert between input PET and local actual ET	0.1-3.0
pFc	Soil saturated hydraulic conductivity	1.0-10.0
iWU	Initial value of soil water	1.0-100.0
iSO	Initial value of overland reservoir	1.0-10.0
iSU	Initial value of interflow reservoir	1.0-50.0
AreaC	Multiplier that modifies the area of grid cells	0.5-1.5

TABLE 2 PHYSIOGRAPHICAL ATTRIBUTES OF THE WATERSHEDS

Watersheds	**Drainage area (km^2)**	**Mean elevation (m)**	**Mean slope (%)**	**Longest path (km)**	**River length (km)**	**Shape factor**	**Elongation ratio**	**Impervious area (%)**	**Forest area (%)**	**Paddy field (%)**	**Crop land (%)**
GWS1	6651.4	615.6	32.4	300.0	232.6	2.528	0.307	1.7	83.1	10.4	1.6
GWS2	2783.5	643.2	33.3	164.1	142.1	2.387	0.363	2.3	90.8	2.7	2.8
GWS3	1590.3	562.5	29.1	164.1	109.2	2.428	0.274	1.3	85.7	5.9	6.1
GWS4	1363.6	400.1	26.5	99.1	121.5	2.916	0.421	2.3	82.0	10.1	5.5
GWS5	2288.2	430.8	27.1	111.8	86.7	1.605	0.483	0.9	75.2	16.8	7.0
GWS6	103.1	549.8	38.7	28.6	20.2	1.762	0.401	0.0	92.0	4.0	4.0
UWS1	931.5	514.8	27.4	64.6	60.6	1.761	0.533	2.1	80.2	9.2	5.9
UWS2	1024.9	274.4	23.4	96.6	77.5	2.146	0.374	4.5	72.2	12.3	8.7

V. RESULTS

A. Regression Analysis

Multiple linear regressions were utilized to develop equations for parameter regionalization. For eight of seventeen parameters in the model, the equations of parameters and watershed properties were estimated through the multiple linear regressions as given in Table 3. Because we could not find the regression equations for the rest of the parameters, they were replaced by the mean value of the parameters from the gauged watersheds. Although the equations did not appeared in this study, it is natural that the physiographical attributes of watersheds directly or indirectly affect hydrologic response. Therefore, further study should continue to investigate the relationships by applying additional watershed properties to other regions.

DOI: 10.5963/JWRHE0403007

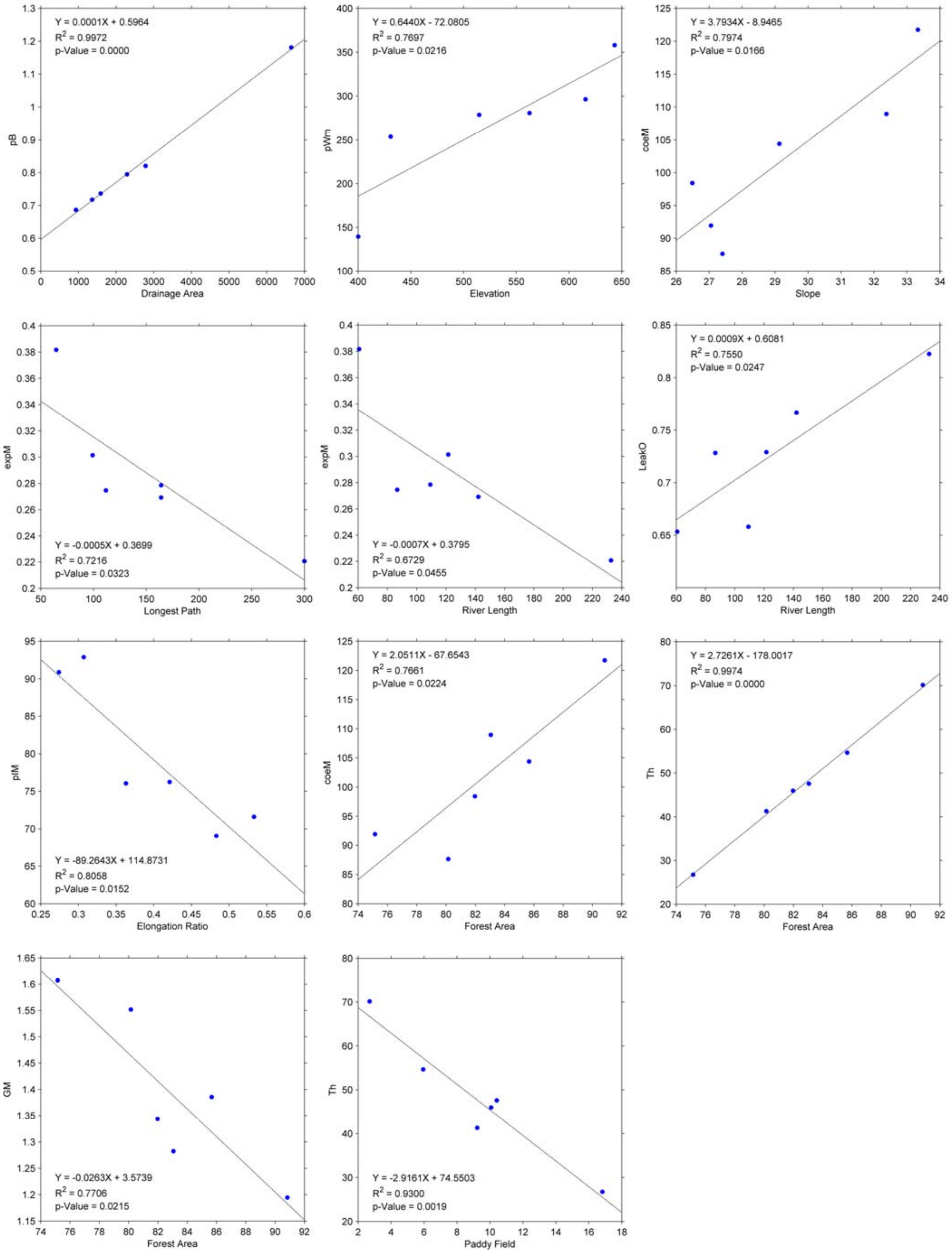

Fig. 3 Scatter plots and regression equations between the model parameters and physiographical attributes of the watersheds

DOI: 10.5963/JWRHE0403007

TABLE 3 WATERSHED PROPERTIES AFFECTING THE MODEL PARAMETERS

Parameter	R^2	Adjusted R^2	Watershed properties
coeM	0.8959	0.8699	Mean slope, forest area
expM	0.7244	0.6555	Longest path, river length
LeakO	0.7783	0.7229	River length
Th	0.9979	0.9974	Forest area, paddy field
GM	0.7706	0.7132	Forest area
pWm	0.7697	0.7122	Mean elevation
pB	0.9974	0.9968	Drainage area
pIM	0.8058	0.7573	Elongation ratio

VI. RUNOFF SIMULATION

The evaluation of the hydrologic model was assessed using three common statistical indices: NSCE, PBIAS, and the Root Mean Squared Error (RMSE) - Observation Standard Deviation Ratio (RSR) [17]. The NSCE index is a normalized statistic that determines the relative magnitude of the noise compared to the variance of the measurements, and varies from negative infinity to one. PBIAS assesses the systematic bias of the simulated runoff discharge; positive and negative values indicate an underestimation and overestimation, respectively. RSR is the standardized RMSE using the standard deviation of the measurements; it varies from the optimal value of zero to a large positive value and is computed as follows:

$$RSR = \frac{\sqrt{\sum \left(Q_o - Q_s\right)^2}}{\sqrt{\sum \left(Q_o - \overline{Q}_s\right)^2}} \tag{4}$$

A previous study [18] recommended model evaluation techniques, including statistical and graphical skills, and proposed criteria in terms of the accuracy of simulated discharge compared to the observed data (Table 4). We focused on and discuss the results from this study in monthly time-steps using the performance ratings presented in Table 4.

TABLE 4 GENERAL PERFORMANCE RATINGS FOR STREAMFLOW SIMULATION IN A MONTHLY TIME-STEP

Rating	NSCE	PBIAS	RSR
Very good	$0.75 < \text{NSCE} \leq 1.00$	$\text{PBIAS} < \pm 10\%$	$0.00 \leq \text{RSR} \leq 0.50$
Good	$0.65 < \text{NSCE} \leq 0.75$	$\pm 10 \leq \text{PBIAS} < \pm 15\%$	$0.50 < \text{RSR} \leq 0.60$
Satisfactory	$0.50 < \text{NSCE} \leq 0.65$	$\pm 15 \leq \text{PBIAS} < \pm 25\%$	$0.60 < \text{RSR} \leq 0.70$
Unsatisfactory	$\text{NSCE} \leq 0.50$	$\text{PBIAS} \geq \pm 25\%$	$\text{RSR} > 0.70$

Fig. 4 shows the comparison between the simulated and observed discharge time series in the ungauged watersheds (UW1 and UW2). Overall seasonal and daily variations of the simulated data agree with that of the observations. However, it can be seen that biased discharge was generated in both watersheds.

Table 5 shows the statistical indices to evaluate the simulated discharge in the ungauged watersheds using the parameter sets estimated from regression equations. In UWS1, the simulated data obtained “satisfactory” ratings in terms of NSCE and RSR while it received an “unsatisfactory” rating in terms of PBIAS, based on the values presented in Table 4. Similarly, in UWS2 the simulation obtained “good” ratings in terms of NSCE and RSR, while receiving an “unsatisfactory” rating in terms of PBIAS.

Although the biased results were obtained by parameter regionalization, we observed that the results from the hydrologic simulation with parameter regionalization are better than the results using the medians of the model parameters.

VII. CONCLUSIONS

This research utilized remote sensory datasets to predict runoff discharge in ungauged watersheds. The CREST distributed hydrological model was employed for hydrological simulation. First, we simulated the runoff discharge in the gauged watersheds, and an optimal parameter set was estimated by model calibration. Next, linear regression equations of the CREST model parameters and watershed attributes were investigated. New parameter sets derived from the regression equations were generated for the ungauged watersheds. Runoff discharge in ungauged watersheds was predicted by the hydrologic model with the parameters, and assessed by comparing observations.

Results demonstrated that the CREST hydrological model and the proposed regression equations could acceptably simulate the runoff discharge as generated by NSCE and RSR in both the gauged and the ungauged basins, but they provided somewhat biased streamflow in all the study basins. In further studies, these biases should be reduced by applying parameters to other

DOI: 10.5963/JWRHE0403007

watersheds and finding watershed properties highly related to model parameters.

TABLE 5 PERFORMANCE STATISTICAL INDICES IN A MONTHLY TIME-STEP

Watershed	NSCE	PBIAS	RSR
GWS1	0.67	29.83	0.58
GWS2	0.62	40.34	0.62
GWS3	0.87	-29.87	0.36
GWS4	0.71	-1.49	0.54
GWS5	0.74	-40.42	0.51
GWS6	0.45	31.15	0.74
UWS1	0.53 (-0.03)*	39.89 (82.98)	0.69 (1.02)
UWS2	0.68 (-15.34)	-75.44 (-481.25)	0.56 (4.04)

* The values in parentheses indicate the performance statistics for the simulation of the CREST model using the medians of the model parameters without parameter regionalization.

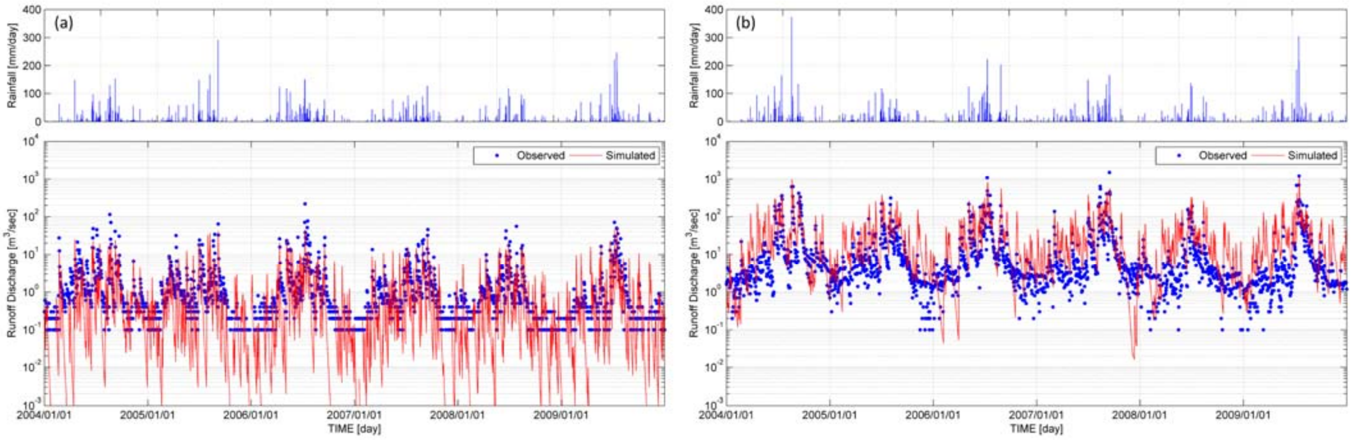

Fig. 4 Comparison between the simulated and observed runoff discharge in (a) UWS1 and (b) UWS2 for validation of the regression equations

ACKNOWLEDGMENT

This research was supported by Korea Institute of Civil Engineering and Building Technology (Project name: Development of a micro raingauge using electromagnetic wave).

REFERENCES

[1] M.G. Kang, J.H. Lee, and K.W. Park, “Parameter regionalization of a Tank model for simulating runoffs from ungauged watersheds,” *J. Korea Water Resour. As.*, vol. 46(5), pp. 519-530 (in Korean), 2013.

[2] A.L. Kay, D.A. Jones, S.M. Crooks, T.R. Kjeldsen, and C.F. Fung, “An investigation of site-similarity approaches to generalization of a rainfall-runoff model,” *Hydrol. Earth Syst. Sci.*, vol. 11, pp. 500-515, 2007.

[3] C. Perrin, L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel, and T. Mathevet, “Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models,” *Hydrolog. Sci. J.*, vol. 52, pp. 131-151, 2007.

[4] Z. Xie, F. Yuan, Q. Duan, J. Zheng, M. Liang, and F. Chen, “Regional parameter estimation of the VIC land surface model: Methodology and application to river basins in China,” *J. Hydromet.*, vol. 8(3), pp. 447-468, 2007.

[5] K.H. Son, M.H. Lee, and D.H. Bae, “Runoff analysis and assessment using land surface model on East Area,” *J. Korea Water Resour. As.*, vol. 45(2), pp. 165-178 (in Korean), 2012.

[6] B.H. Cho, “A study on regionalization of long-term runoff model parameters,” M.S. Thesis, Sejong Univ., Republic of Korea, p. 86, (in Korean), 2004.

[7] A.R. Young, “Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model,” *J. Hydrol.*, vol. 320, pp. 155-172, 2006.

[8] S.H. Lee and S.U. Kang, “A parameter regionalization study of a modified Tank model using characteristic factors of watersheds,” *KSCE J. Civ. Eng.*, vol. 27(4B), pp. 379-385 (in Korean), 2007.

[9] J. Wang, Y. Hong, L. Li, J.J. Gourley, S.I. Khan, K.K. Yilmaz, R.F. Adler, F.S. Policeli, S. Habib, D. Irwin, A.S. Limaye, T. Korme, and L. Okello, “The coupled routing and excess storage (CREST) distributed hydrological model,” *Hydrolog. Sci. J.*, vol. 56(1), pp. 84-98, 2011.

[10] G.J. Huffman, R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, and D.B. Wolff, “The TRMM

DOI: 10.5963/JWRHE0403007

Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *J. Hydrometeo.*, vol. 8, pp. 38-55, 2007.

[11] J.E. Janowiak, R.J. Joyce, and Y. Yarosh, "A real-time global half-hourly pixel-resolution infrared dataset and its applications," *Bull. Am. Meteorol. Soc.*, vol. 82, pp. 205-217, 2001.

[12] Z.S. Haddad, E.A. Smith, C.D. Kummerow, T. Iguchi, M.R. Farrar, S.L. Durden, M. Alves, and W.S. Olson, "The TRMM "day-1" radar/radiometer combined rain-profiling algorithm," *J. Meteorol. Soc. Jpn.*, vol. 75, pp. 799-809, 1997.

[13] B. Rudolf, "Management and analysis of precipitation data on a routing basis," *Proc. intern. Symp. on precipitation and evaporation*, vol. 1, pp. 69-76, 1993.

[14] P. Xie and P.A. Arkin, "Gauge-based monthly analysis of global land precipitation from 1971 to 1994," *J. Geophys. Res.*, vol. 101, pp. 19023-19034, 1996.

[15] F. Su, Y. Hong, and D.P. Lettenmaier, "Evaluation of TRMM multi-satellite precipitation analysis (TMPA) and its utility in hydrologic prediction in La Plata basin," *J. Hydrometeorol.*, vol. 9(4), pp. 622-640, 2008.

[16] S.I. Khan, Y. Hong, J. Wang, K.K. Yilmaz, J.J. Gourley, R.F. Adler, G.R. Brakenridge, F. Policelli, S. Habib, and D. Irwin, "Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins," *IEEE Trans. Geosci. Remote Sens.*, vol. 49(1), pp. 85-95, 2011.

[17] J. Singh, H.V. Knapp, G. Arnold, and M. Demissie, "Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT," *J. Am. Water Resour. As.*, vol. 41(2), pp. 361-375, 2005.

[18] D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Trans. of the ASABE*, vol. 50(3), pp. 885-900, 2007.

DOI: 10.5963/JWRHE0403007