

Prediction of Typhoon-Induced Flood Flows at Ungauged Catchments Using Simple Regression and Generalized Estimating Equation Approaches

Hyosang Lee ¹, Neil McIntyre ², Joungyoun Kim ³, Sunggu Kim ¹ and Hojin Lee ^{1,*}

¹ School of Civil Engineering, Chungbuk National University, Cheongju 28644, Korea; hyosanglee@chungbuk.ac.kr (H.L.); sunggukim@chungbuk.ac.kr (S.K.)

² Centre for Water in the Minerals Industry, University of Queensland, Brisbane 4072, Australia; n.mcintyre@uq.edu.au

³ Department of Information & Statistics, Chungbuk National University, Cheongju 28644, Korea; joungyoun@chungbuk.ac.kr

* Correspondence: Hojinlee@chungbuk.ac.kr; Tel.: +82-43-261-2379

Received: 29 March 2018; Accepted: 14 May 2018; Published: 16 May 2018

Abstract: Typhoons are the main type of natural disaster in Korea, and accurately predicting typhoon-induced flood flows at gauged and ungauged locations remains an important challenge. Flood flows caused by six typhoons since 2002 (typhoons Rusa, Maemi, Nari, Dienmu, Kompasu and Bolaven) are modeled at the outlets of 24 Geum River catchments using the Probability Distributed Moisture model. The Monte Carlo Analysis Toolbox is applied with the Nash Sutcliffe Efficiency as the criterion for model parameter estimation. Linear regression relationships between the parameters of the Probability Distributed Moisture model and catchment characteristics are developed for the purpose of generalizing the parameter estimates to ungauged locations. These generalized parameter estimates are tested in terms of ability to predict the flood hydrographs over the 24 catchments using a leave-one-out validation approach. We then test the hypothesis that a more complex generalization approach, the Generalized Estimating Equation, which includes properties of the typhoons as well as catchment characteristics as predictors of PDM model parameters, will provide more accurate predictions. The results show that the predictions of Generalized Estimating Equation are comparable to those of the simpler, conventional regression. The simpler approach is therefore recommended for practical applications; however, further refinements of the Generalized Estimating Equation approach may be explored.

Keywords: prediction of streamflow; rainfall runoff model; typhoon

1. Introduction

Typhoons are the main natural disaster in Korea. In recent years, Korea has experienced an increased frequency and strength of typhoons and an increase in the resulting damage [1,2]. The maximum annual flood peak historically is often caused by heavy rainfall in the wet season (“Changma”) as well as typhoons; however, since 2002, typhoons are perceived to have become the main source of flood risk, while the Changma has become less influential in Korea. Estimation of typhoon-induced flood hydrographs and associated flood peaks is therefore considered to be increasingly important for engineering applications, such as design of civil engineering structures, stabilization of river banks and flood warning and management.

Gauged flood hydrographs are not available for most of the medium to small-sized catchments in Korea. The need to predict flows at locations where observed flows do not exist for model calibration and validation, i.e., the ungauged catchment problem, is common [3]. The common

approach to this problem is selection of a preferred rainfall-runoff model followed by spatial generalization (regionalization) of its parameters. The regionalization methods depend on nature of the selected rainfall-runoff model and the catchments. Blöschl et al. [3] grouped the methods into three main categories: (a) a-priori estimation of model parameters from catchment characteristics (e.g., [4]); (b) transfer of calibrated model parameters from gauged catchments (e.g., [5]); (c) constraining model parameters by regionalized runoff characteristics and proxy data (e.g., [6]).

The literature on the topic of flood hydrograph regionalization is very well developed. Salinas et al. [7] summarize a comparative assessment of 20 flood prediction studies involving 3023 catchments, concluding that flood flow predictions in ungauged catchments are more accurate in humid than in arid catchments. However, in this and the broader literature, flood studies in Asian monsoon areas are not well represented compared to other humid areas. Although a number of studies have analyzed typhoon rainfall [8–11], there are particularly few studies focusing specifically on flood hydrology associated with East Asian typhoons combined with mountainous environments. One significant study was by [12], who analyzed how flood events in Korea are controlled by the mean, standard deviation, maximum rainfall intensity of the hourly rainfall time series. The typhoons' maximum rainfall intensity was found to be an important factor affecting the flood responses.

Classic statistical regionalization studies are based on the premise that there are predictable differences in rainfall-runoff model parameters between catchments, and that there are no changes in parameters between flood events. This means that the estimated model parameters are averaged over flood events to one representative parameter set per catchment [13,14]. This may underestimate uncertainty especially when large differences in rainfall properties between events, such as in typhoons, create response differences that cannot be captured by a simple rainfall-runoff model. This problem can be addressed empirically, by extending the classical regionalization to include inter-event variations in model parameters. One formal approach to do this is the Generalized Estimating Equation (GEE) method. GEE was proposed by [15] as an extension of the Generalized Linear Model (GLM) to longitudinal data analysis using quasi-likelihood estimation. GEE is widely used in biomedical studies to analyze the repeated measurements from a subject or correlated observation from a cluster [16,17]; however, the GEE method has not previously been applied in the rainfall-runoff regionalization context.

The aim of this study is to develop a method for prediction of typhoon-induced flood hydrographs for ungauged catchments, using a case study of six typhoon flood events in 24 Geum River catchments, Korea. The relationship between typhoon properties and flood hydrograph properties is explored empirically, as well as the ability to regionalize a simple rainfall runoff model across the catchments by using conventional regression and the GEE method.

2. Case Study Data and Preliminary Data Analysis

2.1. Geum River Region

The Geum River is one of the major rivers in Korea, draining the mid-western region of the Korean peninsula (Figure 1). Korea has a temperate climate with four distinct seasons, and the average annual temperature and rainfall in this region are 11.5 °C and 1285 mm, respectively [18]. The Geum River region experiences heavy rainfall in the summer from June through September including typhoons and tends to be cold and drier in the winter, although heavy snowfalls are common [19]. The region has a complex climate, influenced by both oceanic and continental climates.

The 24 gauged catchments used in the case study are shown in Figure 1. These catchments are relatively natural with the main land uses being agriculture and natural woodland. The catchments shown in white in Figure 1, although they include some water level stations, are not included in the analysis because their flood flows are heavily influenced by flood control effects of Daechung Multipurpose Dam and Geum River estuary barrier.

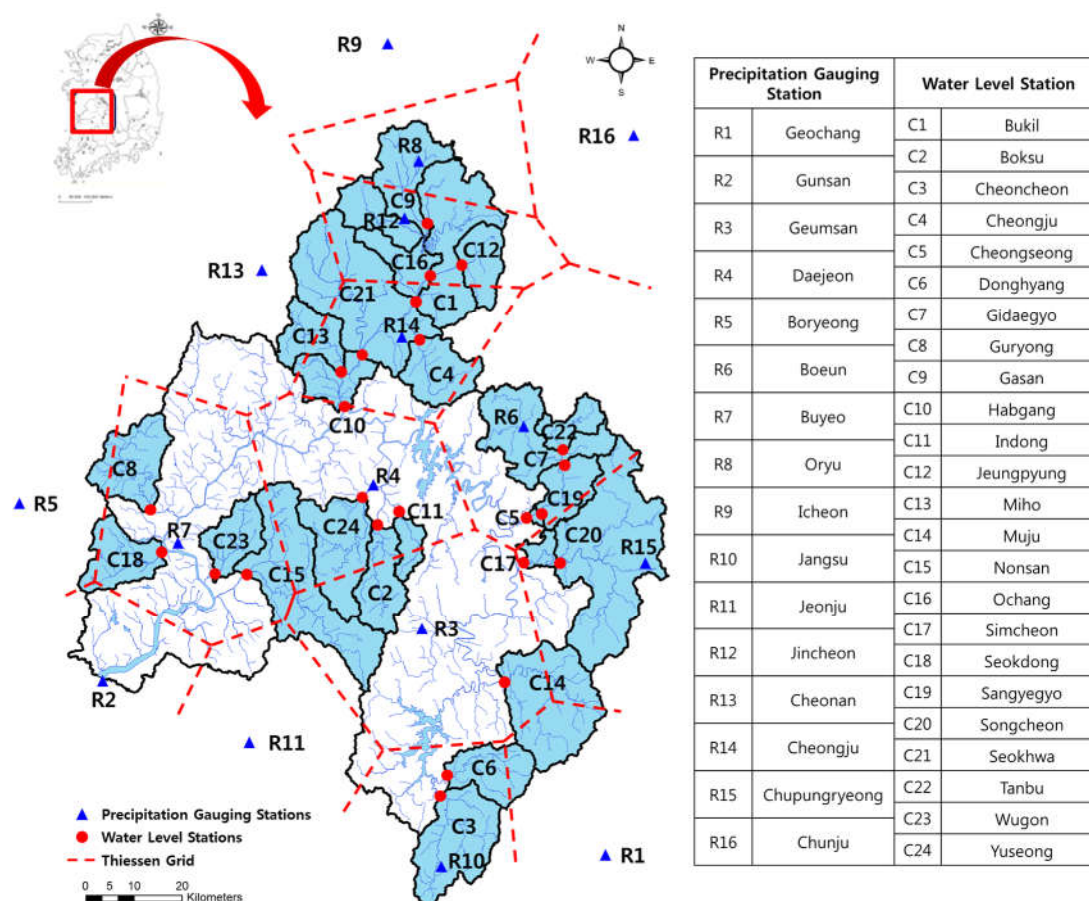


Figure 1. Study catchments; Geum River region, Korea.

Catchment Characteristics (CCs) are the basic descriptors that support understanding of differences between catchments. In this study, eight CCs were selected from a previous study of hydrological similarity [14]. The values of the selected CCs are given in Table 1.

- Area is size of the catchment, which affects flood volumes;
- ALTBAR is the mean altitude of catchment above sea level;
- DPS is the mean of the catchment slope, which affects surface runoff response times;
- DD is the Drainage Density, a measure of the total length of all the rivers in a catchment area divided by the total area of the catchment, which affects how efficiently a catchment is drained;
- FF is the Form Factor, the ratio of the catchment area to the squared value of the total catchment drainage length. It varies from zero (in a highly elongated area) to 1 (in a perfectly circular area), affecting runoff response times. DD and FF are calculated based on equations provided by National Water Resources Management Information System (WAMIS) in Korea;
- Curve Number (CN) is an empirical parameter for predicting direct runoff, developed by the US Soil Conservation System, and affects the volume of runoff during a storm [20];
- FARL is Flood Attenuation Factor by Reservoir and Lakes, estimated based on the reservoir data and the catchment terrain database in WAMIS, affecting flood attenuation [21];
- SAAR is Standard Annual Average Rainfall in the period 1981 to 2010, which represents effects of long-term catchment wetness.

Table 1. Catchment characteristics of the 24 catchments.

Catchments		Area (km ²)	ALTBAR (m)	FF (°)	DD (°)	SAAR (mm)	FARL (°)	CN (°)	DPS (°)	PG (°)
Bukil	C1	907.1	151	0.3	2.1	1222	0.92	67.4	9.7	4
Boksu	C2	162.5	229	0.1	4	1274	1	70.8	15.2	2
Cheoncheon	C3	291.2	555	0.3	3.3	1093	0.96	65.4	14.6	1
Cheongju	C4	166.3	159	0.2	2.4	1192	0.98	71.7	9.7	2
Cheongseong	C5	491	270	0.1	3.1	1198	0.95	65.6	12.9	2
Donghyang	C6	165.4	647	0.2	1.7	1231	0.96	60	16.2	3
Gideagyo	C7	353.1	284	0.3	3.1	1181	0.97	65.4	12.9	1
Guryong	C8	208.2	173	0.1	3	1258	0.99	72.3	12.3	2
Gasangyo	C9	183.9	141	0.3	1.7	1199	0.95	68.8	6.3	2
Habgang	C10	1853	136	0.2	2.4	1225	0.95	69.1	8.4	6
Indong	C11	58.7	217	0.1	3.4	1244	1	64.6	14.7	2
Juengpyung	C12	124.1	150	0.4	2.4	1206	0.96	69.3	8.6	2
Mihogyo	C13	1596	137	0.3	2.4	1192	0.98	63	9.7	2
Muju	C14	390.2	615	0.1	1.5	1154	0.99	59	18.9	4
Nonsan	C15	467.7	148	0.2	2.3	1318	0.91	62.5	11.4	4
Ochang	C16	564	154	0.2	1.9	1221	0.91	66	9.8	2
Simcheon	C17	664.3	372	0.2	2.2	968	0.98	61	14.7	2
Seokdong	C18	162.4	78	0.2	2.7	1271	0.96	66.2	8.1	3
Sangyegyo	C19	464.9	272	0.2	3.1	1195	0.96	65.6	12.9	2
Songcheon	C20	623.4	385	0.2	2.2	969	0.97	60.6	14.8	2
Seokhwa	C21	1594.1	143	0.3	2.3	1223	0.94	68.6	9.7	5
Tanbugyo	C22	81.7	363	0.2	2.5	1090	0.93	54.2	17	1
Woogon	C23	131.8	45	0.3	2.8	1235	0.98	68.9	5	1
Yuseong	C24	251.1	189	0.1	3.5	1315	0.99	62.8	12.1	2

PG: Number of precipitation gauges used in the Thiessen polygon method for calculating catchment average rainfall.

Hydrological data (rainfall, flow and climatic data) were obtained from the Water Resources Management Information System [22] and the Hydrological Investigation Report of Korea [19]. The rainfall data were provided by the Korea Meteorological Agency (KMA). As shown in Figure 1, 16 rain gauges operated by the KMA are used to produce catchment average rainfall on an hourly time-step with the Thiessen polygon method. The potential evapotranspiration data were estimated by using the metrological data of the nearest weather station to study catchment with the FAO Penman-Monteith method, developed by the Food and Agriculture Organization of the United Nations [21,23].

2.2. Typhoons in the Geum River Region

Typhoons in Korea have increased in strength after 2002 [24]. Table 2 lists the periods over which these six typhoons are analyzed in this paper.

Table 2. Selected six typhoons.

Typhoon	Period
Rusa	23 August 2002~6 September 2002
Maemi	6 September 2003~19 September 2003
Nari	13 September 2007~22 September 2007
Dianmu	8 August 2010~17 August 2010
Kompasu	29 August 2010~8 September 2010
Bolaven	20 August 2012~3 September 2012

Observed data from the six typhoons were prepared for the 24 Geum River catchments. Only three catchments have sufficient data for every typhoon—C1 (Bukil), C4 (Cheongju) and C11 (Indong)—however, all catchments have suitable data for two or more flood events, giving a total of 113 observed flood events and at least 8 events for each typhoon. Most of the events are from June to September, and most have single-peak flood hydrographs. The data quality is examined by looking at the runoff ratio (the total runoff divided by the total rainfall during an event, it should be less than one), the timing of peak rainfall and peak runoff (they should not be counter-intuitive), missing data and outliers.

The characteristics of each typhoon flood event are summarized as: total amounts of rainfall and runoff; maximum rainfall intensity; peak flow; lag time between rainfall and runoff peaks; rainfall duration; and antecedent rainfall over the 5 days preceding the event. Figure 2 shows box plots of the first 5 of these characteristics, showing the maxima and minima, and 25% and 75% quantiles. As shown in Figure 2, the flood characteristics vary widely within each typhoon (i.e., between catchments).

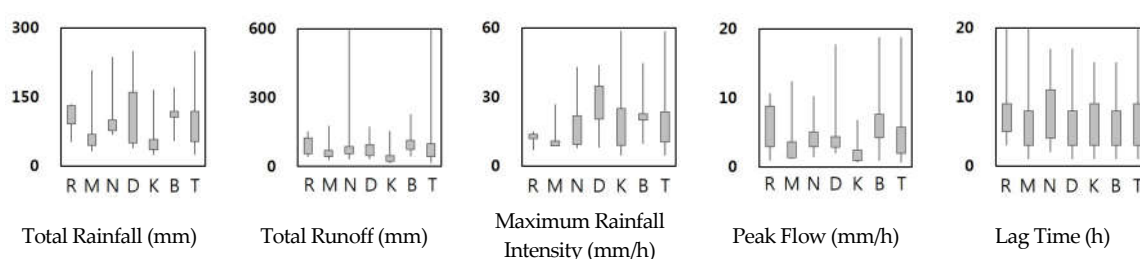


Figure 2. Box plot of the flood characteristics of the 113 flood events (Boxes are 25% and 75% quantiles, R = Rusa, M = Maemi, N = Nari, D = Dianmu, K = Kompas, B = Bolaven, T = Total).

Table 3 shows selected data to further illustrate the nature of the hydrological responses. Table 3 shows that the rainfall duration ranges from 2 to 17 h, and total amount of rainfall ranges from 41 mm to 129 mm. Most of the events have maximum rainfall intensity greater than 10 mm/h. The flood events of Rusa, Mamie and Nari have runoff ratios (total flow/total rainfall) greater than 0.85. However, the runoff ratio for Kompas is only 0.38, presumably related to the lower total rainfall of 40 mm in that typhoon. The wind rather than rainfall was the dominant source of damage in that typhoon event.

Table 3. Characteristics of the events at C4 (Cheongju).

Typhoon	Rusa	Maemi	Nari	Dianmu	Kompasu	Bolaven
TR (Total Rainfall, (mm))	129	72	92	90	41	107
TD (Total Discharge, (mm))	114	70	79	46	16	66
MRI (Maximum Rainfall Intensity, (mm/h))	14	10	10	42	28	21
PF (Peak Flood, (mm/h))	6	3	5	4	3	5
RD (Rainfall Duration, (h))	17	9	11	7	2	15
LT (Lag Time, (h))	4	3	6	4	1	3
PE (Potential Evapotranspiration, (mm))	21	12	5	3	3	15
RR (Runoff Ratio to rainfall, ())	0.88	0.98	0.85	0.52	0.38	0.62
P5 (mm/5 days)	31	43	91	153	44	10

Figure 3 gives some further insight into the variations between flood events and to what degree these are explained by the total rainfall, maximum intensity and P5 (the total amount of rainfall in previous 5 days). This exhibits the expected trends that peak flow and total flow tend to increase with increased rainfall and peak intensity. However, runoff ratio and lag time have no significant linear relationship with the rainfall properties; and none of the flow indices are related to P5. The high runoff ratios and the apparent lack of sensitivity to P5 may be due to the generally wet antecedent conditions (e.g., average P5 for C4 = 62 mm). The wide variation of responses for any given set of rainfall and P5 values is likely to be associated with the variation in responses between catchments.

The runoff ratios greater than 1 in Figure 3 may be due to rainfall or flow estimation errors and/or due to presence of baseflow. Understanding and modeling the variations in responses is the focus of the rest of this paper.

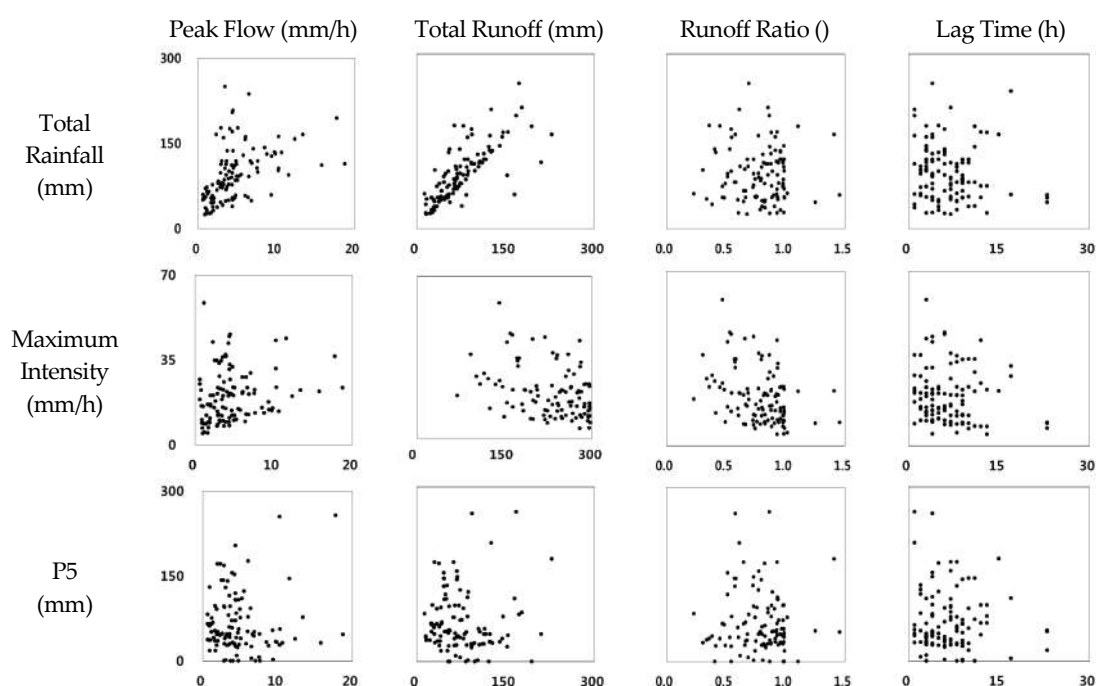


Figure 3. Scatterplots of the flood characteristics of the 113 events.

3. Method

The rainfall and flow data (both hourly) from six typhoon periods are prepared and model parameters (MPs) are calibrated for each. The MPs are regionalized by regressing the calibrated values against CCs. The conventional regression equations are developed using MP values that are averaged across the events, aiming solely to model the differences in response between catchments. The GEE regression model is developed using the MPs for each typhoon event, aiming to include the differences in responses between typhoons as well as between catchments. Descriptions of the PDM rainfall runoff model, the calibration and validation methods, and the regression and GEE methods are given below.

3.1. PDM Rainfall Runoff Model

The PDM (Probability Distributed Moisture) model is a lumped conceptual rainfall runoff model, which has been used as the main rainfall-runoff model in UK flood estimation procedures [25] and has been proposed as suitable for Korean flood estimation [26,27]. The version of the PDM developed by [25], illustrated schematically in Figure 4, is used here. The details of the PDM model structure are described in [28] and only a summary is provided here.

The PDM is composed of a soil moisture accounting model for estimating the effective rainfall and a routing model for converting effective rainfall into runoff at a defined point in the river system. The soil moisture accounting model represents the spatial distribution of soil moisture storage as a Pareto distribution function as specified in Equation (1).

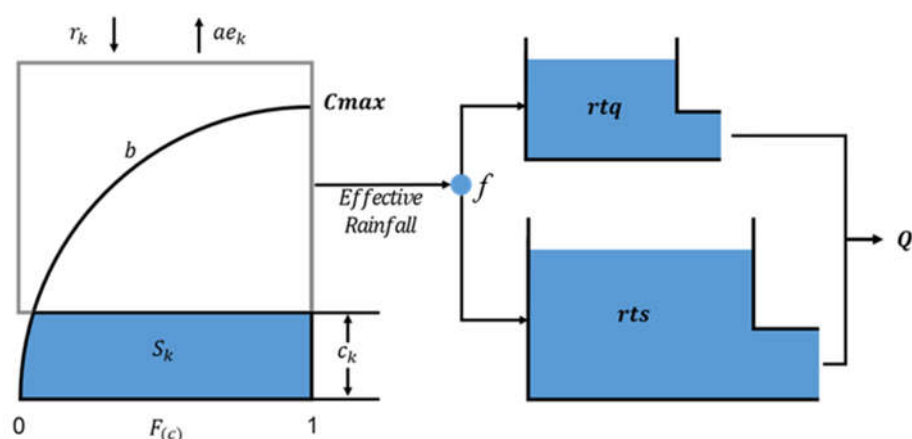


Figure 4. Schematic description of the PDM model [28].

$$F(c) = 1 - \left(1 - \frac{c}{C_{max}}\right)^b, 0 \leq c \leq C_{max} \quad (1)$$

r_k is the rainfall, ae_k is the actual evapotranspiration, S_k is the soil moisture storage, Q is the runoff, c_k is the soil moisture storage capacity, C_{max} is the maximum soil moisture storage capacity of the catchment, and b is a parameter that defines the strength of spatial variation of c . c is distributed uniformly between values of 0 and C_{max} if b is equal to 1; and tends towards a single value, C_{max} , as b tends towards 0.

The routing model consists of two parallel reservoirs that represent quick and slow responses of catchment runoff. The parameters of the PDM model are summarized in Table 4, along with the ranges of values assumed feasible prior to model calibration [29].

Table 4. Parameters of the PDM model.

MPs	Description	Range (Units)
C_{max}	Maximum storage capacity of the catchment	1–500 (mm)
b	Degree of spatial variability of storage capacity in the catchment	0–2 ()
rtq	Residence time of the quick flow reservoir	0–15 (h)
rts	Residence time of the slow flow reservoir	50–500 (h)
f	Fraction of effective rainfall that enters the quick flow reservoir and 1- f of effective rainfall that enters the slow flow reservoir	0–1 ()

3.2. Calibration

For calibration of the MPs in Table 4, a Uniform Random Search was implemented using the Monte-Carlo Analysis Toolbox (MCAT) [28]. 10,000 MP sets were randomly sampled from the prior ranges in Table 4, and objective functions were calculated for each sample. This approach to calibration achieves calibration performances that are sufficient approximations to the optimal performances. The initial soil moisture condition of each case is adopted from the results of preliminary, daily continuous-time modelling [26]. The first 10% of the calibration period was used as a warm-up period during which the errors were neglected in calculation of the objective function in order to reduce sensitivity to initial conditions.

The calibration objective function is the Nash Sutcliffe Efficiency (NSE), which measures all-round model fit, although, is often observed to have a bias towards fitting high flows [25,30].

$$NSE = 1.0 - \frac{\sum_{i=1}^n (\mathbf{O}_i - \mathbf{S}_i(\theta))^2}{\sum_{i=1}^n (\mathbf{O}_i - \bar{\mathbf{O}})^2} \quad (2)$$

where, $\mathbf{O}_i$ is observed flow at time i , $\bar{\mathbf{O}}$ is the mean of observed flows, $\mathbf{S}_i(\theta)$ is simulated flow at time i with parameter set θ , 1 of NSE indicates a perfect match of the simulated and observed flow.

3.3. Statistical Regression

In this study, we use the following notations. y_{ij} is defined as a calibrated MP value of the j -th flood event from the i -th catchment. $x_i = (x_{i1}, \dots, x_{ip})$ is the set of CCs for the i -th catchment and $z_{ij} = (z_{ij1}, \dots, z_{ijq})$ is the set of rainfall characteristics for the j -th flood event at the i -th catchment. For each i -th catchment, let $y_i = \sum_{j=1}^{n_i} y_{ij} / n_i$ and $z_i = (z_{i1}, \dots, z_{iq}) = \left(\sum_{j=1}^{n_i} \frac{z_{ij1}}{n_i}, \dots, \sum_{j=1}^{n_i} \frac{z_{ijq}}{n_i} \right)$ where n_i denotes the number of flood events in the catchment.

3.3.1. Conventional Regression

In the conventional regression analysis, for each catchment, the PDM MPs are calibrated for each event and the values averaged over all events [9]. A linear least-squares regression equation is developed for each of the five averaged model MPs as follows:

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i \quad (3)$$

$E(\cdot)$ is an expectation and e_i represents all random errors in the estimation approach, including randomness between catchments and non-convergence of the optimization of the MPs during calibration. The significant CCs are identified using forward selection with a significance level (p value) of 0.05. That is, we added the CC with the smallest p -value from a wald test and stop adding CCs when no variable has a p -value less than 0.05.

3.3.2. Generalized Estimating Equations (GEEs)

Multiple flood events in a catchment can be considered as repeated measurements from a subject, which corresponds to a catchment in this study [20,31,32]. Therefore, using event-aggregated MPs in (3) leads to loss of information about variability of MPs between events and is likely to underestimate the standard error of the MP estimate for any event. One possible solution to this is to consider a subject-specific random effect in the model fitting using a generalized linear mixed model; however, this would involve specifying the error distribution of y_{ij} . Instead, a GEE approach is adopted that avoids the need to fully account for the error distributions while still potentially achieving robust results [16]. For our flood event data, the following GEE model can be considered:

$$E(y_{ij}) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_{p+1} z_{ij1} + \dots + \beta_{p+q} z_{ijq} \quad (4)$$

Note that, while a non-linear transformation of y is generally present in a GEE, here the identity link function is used. Equation (4) is separately estimated for each MP. Then, $E(y_{ij})$ corresponds to the mean value of an MP for the j -th flood event from the i -th catchment.

The estimation of the parameters β requires the within-subject covariance, that is “Exchangeable” structure assuming that flood events over time have the same correlation, to be specified [33] and See Appendix A. The estimation is based on iteratively reweighted least squares with quasi-likelihood, incorporating the forward selection method and $p < 0.05$ as the significance criterion. All the statistical analysis was performed using Statistical Analysis Systems (SAS) version 9.4.

3.4. Validation

Leave-one-out cross validation is applied. One catchment is kept as a target catchment for validation, and the regression models are developed with the remaining 23 catchments. The simulated hydrographs are visualized and the NSE is evaluated to represent the applicability of the approach for ungauged catchment prediction.

4. Results

4.1. Calibration of the PDM Model

A separate PDM MP set was calibrated for all 113 events with sufficient observed data, in order to examine differences between them. The results for the 6 flood events at the C4 (Cheongju) catchment are presented here as an example. Figure 5 shows the scatter plots of MP value versus NSE obtained from the Uniform Random Search. Optimal values of rtq and f are relatively identifiable, whereas optimal values for rts are relatively non-identifiable, with c and b having intermediate identifiability. Figure 5 shows that, even for the more identifiable MPs, the optimum values are different across typhoons. A similar result is observed for the other gauges.

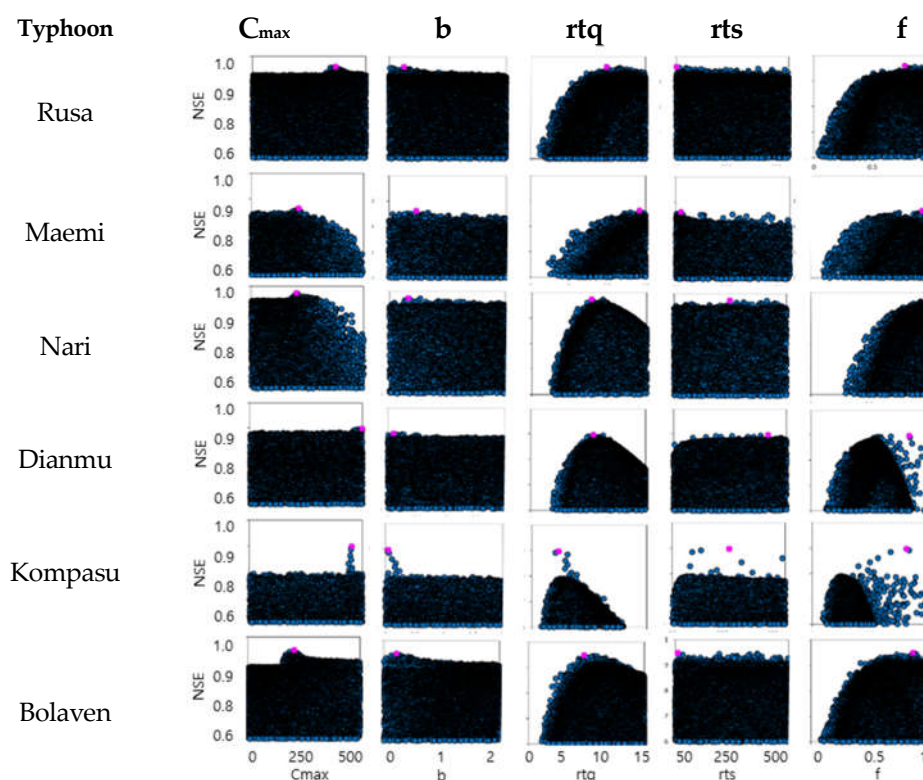


Figure 5. Scatter plots of calibration NSE against MP for each Typhoon flood events at C4.

Table 5 shows the MP and NSE values for C4, and Figure 6 shows the corresponding modeled and observed hydrographs. Table 6 shows the NSE values representing calibration performance of the model applied to the 113 events.

Table 5. Calibrated model parameters for the six typhoon floods at C4.

Typhoon	C_{max} (mm)	b ($\circ$)	rtq (h)	rts (h)	f ($\circ$)	NSE
Rusa	367.6	0.25	9.9	51.0	0.77	0.96
Maemi	211.4	0.47	14.2	69.9	0.96	0.86
Nari	201.0	0.32	7.8	273.1	0.99	0.97
Dianmu	497.8	0.10	8.2	427.3	0.84	0.90
Kompasu	453.0	0.05	3.5	276.0	0.81	0.90
Bolaven	206.5	0.22	7.2	59.3	0.87	0.95
Averaged	322.9	0.23	8.5	192.8	0.87	0.80

In rainfall runoff modelling, the estimated MPs are generally assumed to be the same over all flood events; however, there are wide variations in optimum model parameters in Table 5 and Figure 5. This could be due to a data quality problem, a limitation of model structures or different hydrological responses to flood events. We suppose that these variations in MP values may be partly due to the changes in hydrological response between events, i.e., the MP value must change to allow the PDM model to capture that changed response. Thus, the flood characteristics of each typhoon event are included in the GEE.

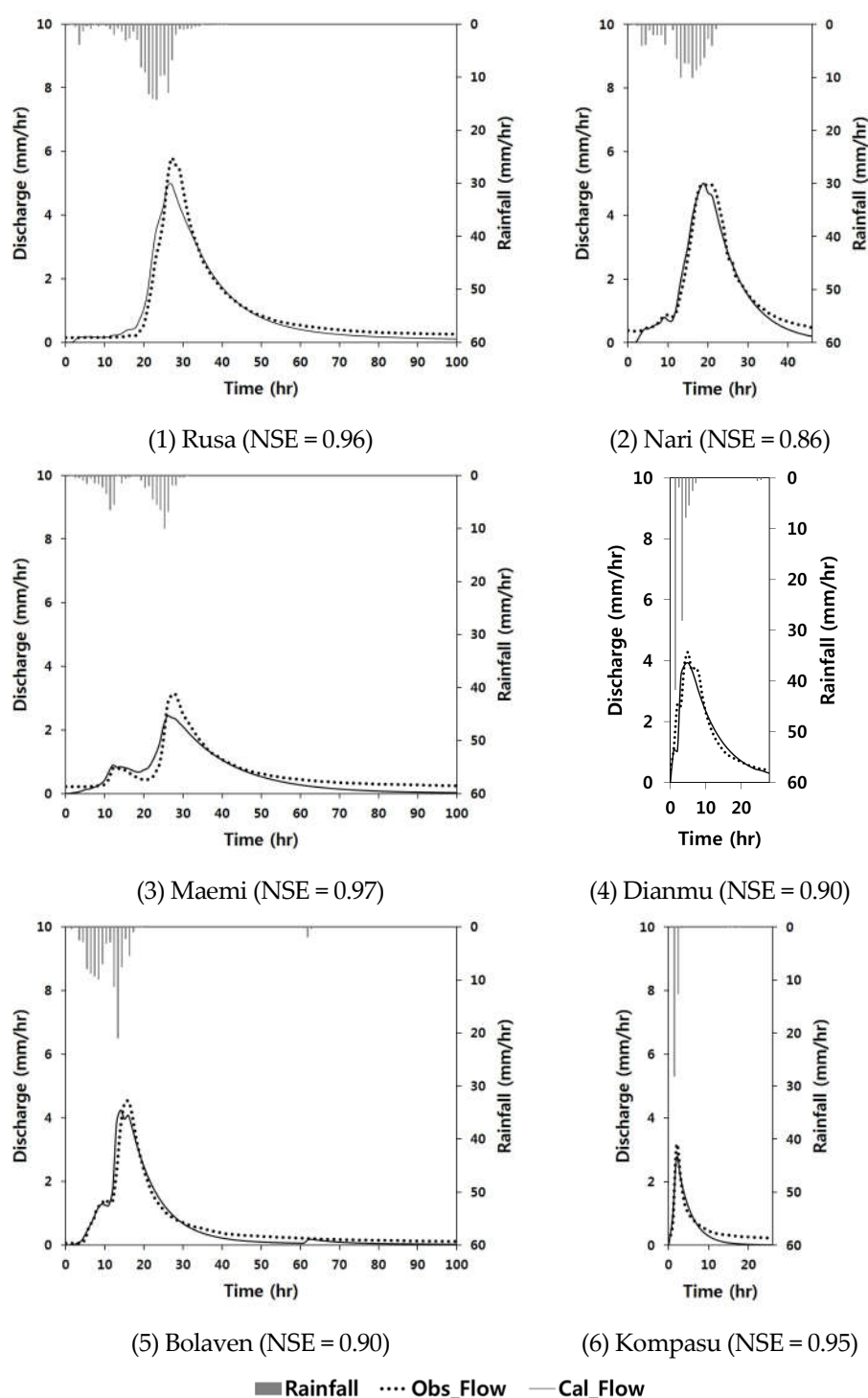


Figure 6. Observed and calibrated hydrographs for the six events at C4.

Table 6. Calibration Results (NSE) for the six typhoon flood events for all 24 catchments

Catchment	Rusa	Maemi	Nari	Dianmu	Kompasu	Bolaven	No. > 0.5 NSE	No. Events	%
C1	0.87	0.79	0.79	0.48	0.77	0.83	5	6	83
C2	0.92	-	-	0.53	0.76	0.72	3	4	75
C3	-	-	0.76	0.89	0.90	0.96	4	4	100
C4	0.96	0.86	0.97	0.90	0.90	0.95	6	6	100
C5	-	0.24	0.27	-0.43	-0.44	0.87	1	5	20
C6	-	-	0.75	0.69	0.8	0.89	4	4	100
C7	-	-	0.8	0.55	0.55	0.83	4	4	100
C8	0.73	0.42	0.74	0.35	0.28	-0.02	2	6	33
C9	-	-	-	0.73	0.74	0.86	3	3	100
C10	-	-	0.26	0.04	0.66	0.7	1	3	33
C11	0.85	0.79	0.96	0.4	0.62	0.67	5	6	83
C12	0.83	0.83	-	0.67	0.55	0.84	5	5	100
C13	-	-	-	0.12	-13.5	0.5	1	3	33
C14	-	-	0.43	0.68	0.34	0.3	1	4	25
C15	0.56	-0.64	0.28	-0.92	0.28	0.71	2	6	33
C16	0.76	0.25	0.64	0.68	-0.32	-	3	5	60
C17	-	-	0.49	0.77	0.24	0.79	2	4	50
C18	-	-	-1.5	-0.56	0.49	0.39	0	4	0
C19	-	-	0.57	-0.14	-0.60	0.71	2	4	50
C20	-	0.82	0.73	0.83	0.47	0.47	3	4	75
C21	0.79	0.72	0.43	0.67	-1.12	0.68	3	6	50
C22	-	-	0.9	-0.37	-0.14	0.03	1	4	25
C23	0.12	-0.32	0.34	-0.23	0.44	0.47	0	6	0
C24	0.72	-	0.82	0.52	0.73	0.87	5	5	100
Max	0.92	0.86	0.97	0.89	0.9	0.96	Missing values in observed data		
Min	0.12	-0.64	-1.5	-0.92	-13.5	-0.02			
No. > 0.5 NSE	10	6	12	13	12	17			
No. Cat.	11	11	20	24	24	23			
%	91	55	60	54	50	74			

There is a wide variation in results across flood events. For typhoon Kompasu period, the range of NSE values across 18 catchments is -13.5 to 0.9; and for typhoon Bolaven the range across 23 catchments is -0.02 to 0.96. 68 of the 113 events have NSE values greater than 0.5. C18 and C23 have consistently low calibration performances with none of the modelled events having NSE > 0.5. This may be because these catchments have comparatively large data errors or have controls on flow, such as out-of-bank flood storage, which the simple PDM model cannot represent. The calibration results for typhoon Dianmu and Kompasu are poor in 12 and 13 of the catchments. This result is due to the inability of the model to simulate the low runoff ratios observed in many of the catchments (e.g., Table 3). It is likely that in these cases the observed rainfall was overestimated. Only the MP sets giving calibration NSE values greater than 0.5 are used for the statistical regression.

4.2. Statistical Regression

Table 7 shows conventional regression equations based on the set of calibrated, event-averaged MPs (excluding C4, which will be used for validation of these equations). The R^2 of these equations varies between 0.1 (Cmax) to 0.79 (f).

The regression equations in Table 7 are applied to catchment C4 as an example. The estimated MPs are shown in Table 8, and Figure 7 shows the predicted hydrographs. The hydrographs at Kompasu and Bolaven are not considered acceptable in terms of volumes and peaks, indicating that MP estimates are not applicable to all flood events. The result at Dianmu is better than the calibrated model in terms of peak flow, at the expense of volume performance.

Table 7. The conventional regression equations for PDM model parameters when using C4 as the validation catchment.

Model Parameter	Regression Equation	R ²	p-Value
C _{max}	248.3 – 0.04 Area	0.10	0.16
b	2.35 – 0.0238 CN	0.10	0.17
rtq	32.05 + 0.003 Area – 0.01 ALTBAR – 0.3 CN	0.52	0.00
rts	–728.56 – 0.06 Area – 77.16 DD + 1211.31 FARL	0.55	0.00
f	–0.6 – 0.0002 Area – 0.12 DD + 1.84 FARL	0.79	0.00

Table 8. Estimated PDM model parameters for catchment C4 using conventional regression.

Model Parameter	C _{max}	b	rtq	rts	f
Value	242	0.64	9.86	264	0.88

Table 9 shows the developed GEEs. In the GEE model of parameter b, the total amount of rainfall (TR) in a typhoon event was selected as an explanatory variable. The b parameter represents the degree of spatial variability of storage capacity in the catchment, therefore a possible explanation for the model of b is as follows: As total rainfall increases, the effect of variability of the soil moisture storage capacity is expected to reduce and hence the value of b is expected to reduce. However, this effect should be implicit in the PDM model, and so the reason b is related to TR is more complex. For the other MPs, only CCs are selected. The R² values in Table 9 are low compared to those in Table 7. This is expected because of the much lower variance of the MPs when they are averaged over the typhoons so that only 23 data points are used, as was done to obtain the results in Table 7, compared to the 107 data points used to obtain the results in Table 9 (107 points rather than 113 because the 6 points from C4 are kept back for model validation). In particular, the non-identifiable MPs would increase variance of the errors more if they are not averaged over events. The results in Table 9 may also be influenced by the covariance of the MP values between flood events in each catchment, whereby if the matrix R Equation (6) could be more accurately specified it may improve the GEE result.

Table 9. The GEEs for PDM model parameters when using C4 as the validation catchment.

Model Parameter	GEE	R ²
C _{max}	232.33 – 0.03 Area	0.01
b	1.29 – 0.004 TR	0.11
rtq	22.75 + 0.003 Area – 0.007 ALTBAR – 0.15 CN	0.24
rts	678.12 – 0.09 Area – 50.46DD – 0.26 SAAR	0.13
f	1.70 – 0.0002 Area – 0.06 DD – 0.0006 SAAR	0.08

In Figure 7, the GEE results (shown in solid lines) are generally better performance in the peak flow than the conventional linear regression results (shown in dotted lines) for typhoons Rusa, Maemi, Bolaven and Kompasu. For typhoon Kompasu, the GEE matches closely the peak of the hydrograph, although it has over-predicted the recession flows. For typhoon Nari, the hydrographs simulated using GEE are slightly worse than those simulated using the conventional regression model. However, overall, the GEE gave slightly better performances compared to the conventional regression model.

The results of cross validation are shown in Figure 8. The selected CCs for each MP are the same CCs shown in Tables 7 and 9. The results for the conventional regression and GEEs are compared with the locally calibrated model results for each typhoon flood event. In Figure 8a,b, if the points are located on the 45-degree line, the performances using regionalization are the same as those of calibration; whereas if the points are located above the 45-degree line, the performances using regionalization are better than those using the calibrated model. The calibration results are better than both regionalization approaches. The results of the two regionalization methods are compared

in Figure 8c). The performances using the two methods are similar. This indicates that there is no practical benefit adopting the more complex GEE method, although there may be scope for refining its specification, in particular the link function and correlation structure. Furthermore, in this study we did not incorporate error models in order to estimate confidence intervals on predictions, and these may be more successful using the GEE approach.

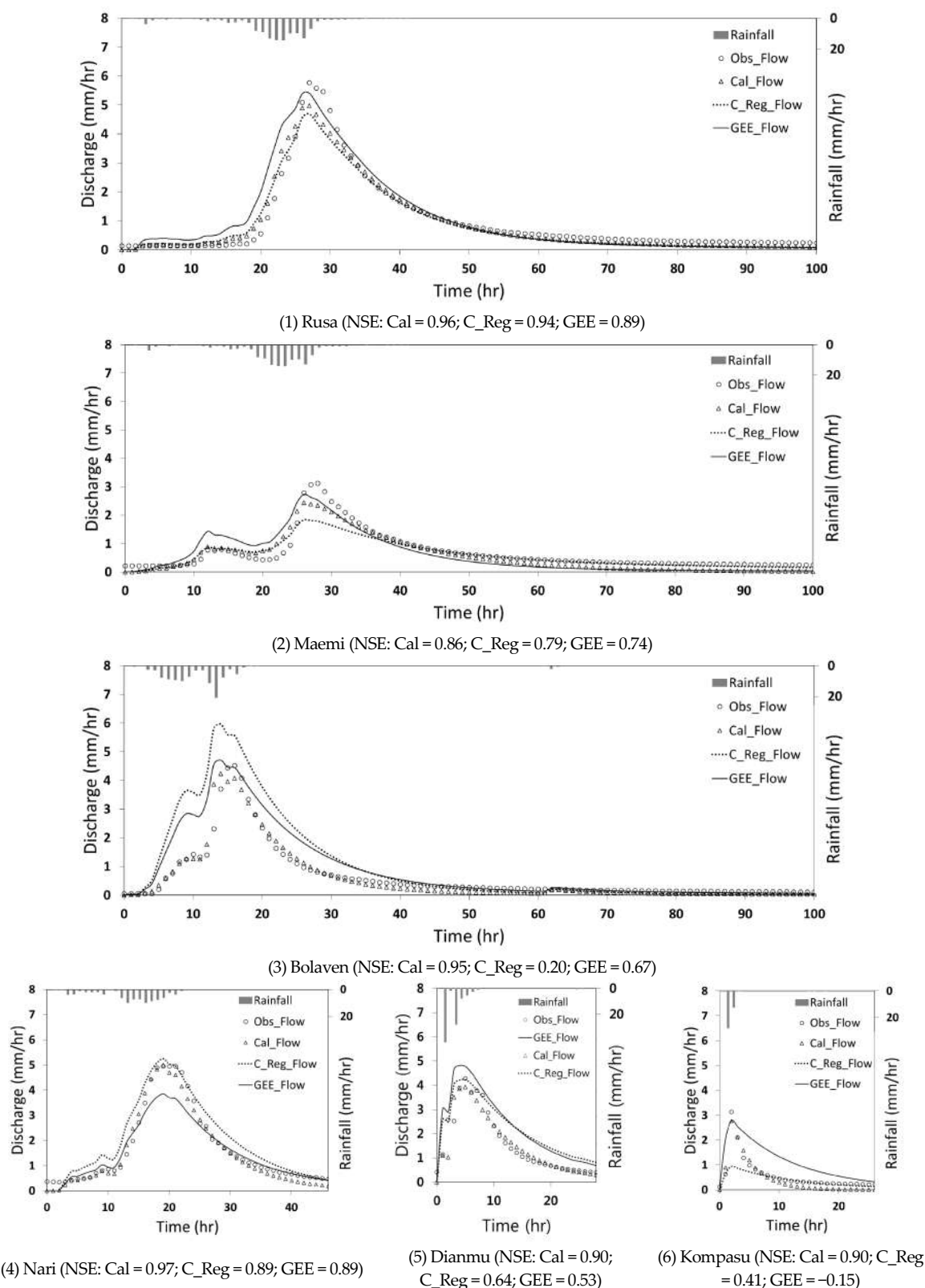


Figure 7. Modelled and observed hydrographs at catchment C4.

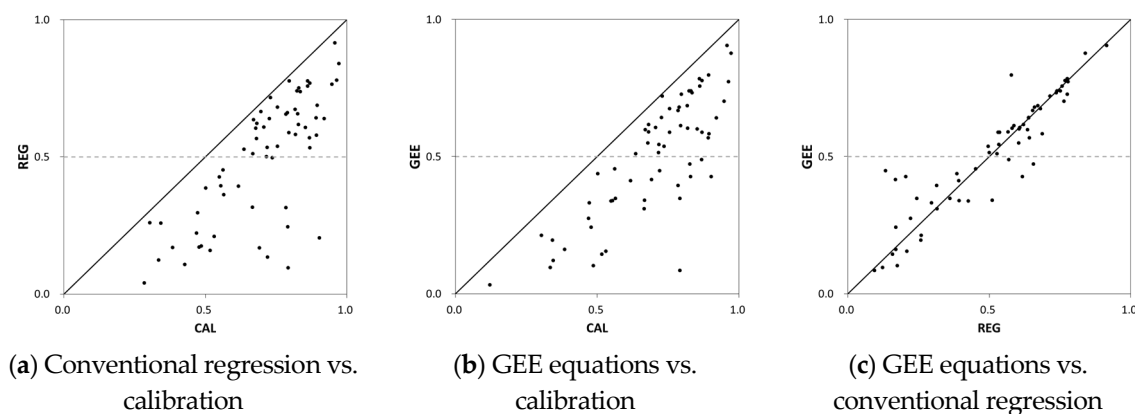


Figure 8. Comparison between the NSE values obtained using calibration, conventional regression and GEEs.

5. Conclusions

This study analyses six typhoon flood events (Rusa, Maemi, Nari, Dienmu, Kompas and Bolaven) at 24 Geum River catchments and applies the Probability Distributed Moisture (PDM) rainfall-runoff model to predict river flows. Two alternative statistical models, conventional regression and Generalized Estimating Equations (GEEs), are used to regionalize the PDM model parameters, so that typhoon-induced flows can be simulated at ungauged locations within the same region. The main outcomes are:

- The application of a calibrated PDM model to modelling flood event flows in Geum River catchments shows, overall, an acceptable model performance. This supports the use of the PDM model or comparable rainfall-runoff models for simulating extreme flood events in Korea.
- Using conventional regression equations to regionalize model parameters to ungauged catchments showed mixed success, for example when treating the C4 catchment as ungauged, there were good results for two flood events (typhoons Nari and Dianmu) and underestimated peaks for two events (typhoons Rusa and Kompas).
- The GEE model extends the conventional regression by including the inter-event variability in PDM model parameters as well as the inter-catchment variability. However only the model parameter b was found to be related to event properties; and validation results showed only slight improvement on the simpler regression approach. While for practical applications we would therefore recommend the simpler regression approach, refinements to the GEE approach may be explored, in particular its potential advantage for estimating an error model and confidence limits on predictions.

Future research should also increase the number of extreme flood events and study catchments to increase the statistical significance of the analysis and increase confidence in the comparison between the two approaches.

Author Contributions: Hy.L. designed the study; N.M. and Ho.L. conducted the majority of the data analysis and wrote the majority of the paper; S.K. provided the data from the models. J.K. contributed to the statistical analysis and interpretation.

Acknowledgments: This research was supported by a grant (11-TI-C06) from Advanced Water Management Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government and 2015 International Hydrological Program (IHP) funded by the Ministry of Land, Infrastructure and Transport. The statistical analysis was supported by the Statistical Analysis and Consulting Center in Chungbuk National University. Joungyoun Kim's work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2017R1C1B5015192).

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

Appendix A. GEE for Regionalization of Rainfall Runoff Model

GEE is a semi-parametric method which uses moment assumptions, instead of assuming a certain distribution as in a Generalized Linear Mixed Model. Besides, GEE is instead of attempting to model the within-subject covariance structure, to treat it as a nuisance and simply model the mean response. Sandwich estimator [32,34,35] enables to be robust against to the mis-specified within-subject covariance structure. The estimation is based on iteratively reweighted least squares (IRLS) with quasi-likelihood. For our flood event data, the following GEE model can be considered:

$$g(E(y_{ij})) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_{p+1} z_{ij1} + \dots + \beta_{p+q} z_{ijq} \quad (A1)$$

Note that $g(\cdot)$ represents a link function, which can be identity, logit and log functions depending on if response y_{ij} is normally distributed, binary or a Poisson process. In this paper, we assume our predicted variables of interest (i.e., PDM MPs) are continuous, hence, we consider the identity link function only and Equation (A1) can be rewritten as:

$$E(y_{ij}) = \mu_{ij} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_{p+1} z_{ij1} + \dots + \beta_{p+q} z_{ijq} \quad (A2)$$

Based on the CCs and flood characteristics, the GEE model is

$$E(y_{ij}) = \mu_{ij} = \beta_0 + \beta_1 Area_i + \beta_2 ALTBAR_i + \beta_3 FF_i + \beta_4 DD_i + \beta_5 SAAR_i + \beta_6 FARL_i + \beta_7 CN_i + \beta_8 DPS_i + \beta_9 TR_{ij} + \beta_{10} MRI_{ij} + \beta_{11} P5_{ij} \quad (A3)$$

Equation (A3) is separately estimated for each PDM MP. Then, μ_{ij} corresponds to the mean value of an MP for the j -th flood event from the i -th catchment.

In GEE, within-subject covariance has some structure and this is called “working covariance”. For estimation, we need to specify the correlation structure (denoted as R_i). For example, if there are three modeled flood events in the i -th catchment, we can consider R_i as the following structure:

$$R_i = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \quad (A4)$$

The matrix in (A4) is the most general case and is called “unspecified” or “unstructured”. In addition, there exist other structures such as (i) independent ($\rho_{ij} = 0$ for all i and j); (ii) Exchangeable ($\rho_{ij} = \rho$ for all i and j) and (iii) AR (1) ($\rho_{ij} = \rho^{|i-j|}$). In this paper, we use the “Exchangeable” structure assuming that flood events over time have the same correlation.

References

- Altman, J.; Doležal, J.; Cerný, T.; Song, J.S. Forest response to increasing typhoon activity on the Korean peninsula: Evidence from oak tree-rings. *Glob. Chang. Biol.* **2013**, *19*, 498–504, doi:10.1111/gcb.12067.
- National Disaster Safety Portal. Available online: <http://safekorea.go.kr> (accessed on 1 April 2018).
- Blöschl, G.; Sivapalan, M.; Wagener, T.; Viglione, A.; Savenije, H. *Runoff Prediction in Ungauged Basins—Synthesis across Processes, Places and Scales*; Cambridge University Press: Cambridge, UK, 2013.
- Vieux, B.E. Distributed Hydrologic Modeling Using GIS. In *Distributed Hydrologic Modeling Using GIS. Water Science and Technology Library*; Springer: Dordrecht, The Netherlands, 2001; Volume 38, pp. 1–17.
- Perrin, C.; Oudin, L.; Andreassian, V.; Rojas-Serna, C.; Michel, C.; Mathevet, T. Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. *Hydrol. Sci. J.* **2007**, *52*, 131–151.
- Boughton, W.; Chiew, F. Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model. *Environ. Model. Softw.* **2007**, *22*, 943–956.
- Salinas, J.L.; Laaha, G.; Rogger, M.; Parajka, J.; Viglione, A.; Sivapalan, M.; Blöschl, G. Comparative assessment of predictions in ungauged basins—Part 2: Flood and low flow studies. *Hydrol. Earth Syst. Sci.* **2013**, *15*, 2421–2435.
- Choi, C.; Han, D.; Kim, J.; Jung, J.; Kim, D.; Kim, H. Mega flood simulation assuming successive extreme rainfall events. *J. Wetlands Res.* **2016**, *18*, 76–83.
- Kim, K. A Study on Precipitation and Runoff Analysis of Gangneung Area of Korea by the Typhoon Lusa in 2002. Master’s Thesis, University of Seoul, Seoul, Korea, 2005.

10. Kim, T.; Maeng, S.; Choi, H.; Kim, D.; Jang, I. Analysis of runoff characteristic using rainfall-runoff model. *J. Saf. Crisis Manag.* **2014**, *10*, 119–134.
11. Oh, T.; Moon, Y.; Ahn, J. A Study on special quality of hourly precipitation of typhoon happened in Korea. *J. Korea Water Resour. Assoc.* **2007**, *40*, 709–722.
12. Oh, K. A Study on the Evaluation of the Extreme Rainfall Using Climatic Indices in Korea. Ph.D. Thesis, University Of Seoul, Seoul, Korea, 2009.
13. Joo, J.; Kjeldsen, T.; Kim, H.J.; Lee, H. A comparison of two event-based flood models (ReFH-rainfall runoff model and HEC-HMS) at two Korean catchments, Bukil and Jeungpyeong. *KSCE J. Civ. Eng.* **2014**, *18*, 330–343, doi:10.1007/s12205-013-0348-3.
14. Kjeldsen, T.R.; Kim, H.; Jang, C.H.; Lee, H. Evidence and implications of nonlinear flood response in a small mountainous watershed. *J. Hydrol. Eng.* **2016**, *21*, 04016024.
15. Liang, K.-Y.; Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, *73*, 13–22, doi:10.1093/biomet/73.1.13.
16. Potthoff, R.F.; Roy, S.N. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **1964**, *51*, 313–326.
17. Wang, M. Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments. *Adv. Stat.* **2014**, *2014*, 11p, doi:10.1155/2014/303728.
18. Korea Meteorological Administration. Available online: http://www.kma.go.kr/info_open/public_data/guidepage.jsp (accessed on 1 June 2016).
19. MLTM. *Statistical Yearbook of Ministry of Land, Transportation and Maritime Affairs 2012*; Ministry of Land, Transportation and Maritime Affairs: Sejong, Korea, 2012.
20. Donner, A. Linear regression analysis with repeated measurements. *J. Chronic Dis.* **1984**, *37*, 441–448.
21. Rim, C. Comparison of evapotranspiration estimation approaches considering grass reference crop. *J. Korean Water Resour. Assoc.* **2008**, *41*, 212–228.
22. Water Resources Management Information System. Available online: www.wamis.go.kr (accessed on 1 June 2016).
23. Chapter 4—Determination of ETo. Available online: www.fao.org/docrep/X0490E/x0490e08.htm (accessed on 1 June 2016).
24. NEMA. *Natural Disaster Year Book of 2010*; National Emergency Management Agency: Seoul, South Korea, 2010.
25. Moore, R.J. The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 483–499.
26. Chang, H.; Lee, H. Assessment of continuous simulations of rainfall runoff models based on soil moisture accounting models at Geum River catchments, Korea. *J. Korean Soc. Hazard Mitig.* **2015**, *15*, 231–242.
27. Choi, H.H.; Lee, H.S. Assessment of conceptual rainfall runoff models for regionalization at Miho catchment. *Korean Soc. Hazard Mitig.* **2012**, *12*, 193–203.
28. Korea Meteorological Administration. Available online: <http://www.imperial.ac.uk/environmental-and-water-resource-engineering/research/software/> (accessed on 1 June 2016).
29. Wagener, T.; Wheater, H.S. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *J. Hydrol.* **2006**, *320*, 132–154.
30. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290.
31. Ugrinowitsch, C.; Fellingham, G.; Ricard, M. Limitations of ordinary least squares models in analyzing repeated measures data. *Med. Sci. Sports Exerc.* **2004**, *36*, 2144–2148.
32. Ware, J. Linear Models for the Analysis of Longitudinal Studies. *J. Am. Stat. Assoc.* **1985**, *39*, 95–101.
33. White, H. A Heteroskedasticity—Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *J. Econom. Soc.* **1980**, *48*, 817–838.
34. Lee, H.; McIntyre, N.R.; Wheater, H.S.; Young, A.R. Selection of conceptual models for regionalisation of rainfall runoff relationships. *J. Hydrol.* **2005**, *312*, 125–147.
35. Hong, Y.; Adler, R.F. Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *Int. J. Remote Sens.* **2008**, *29*, 471–477.

