

APEC기후센터

극단값 모형을 기반한 국내 강우 현상 모의

김무섭 기후분석팀

APEC기후센터
연구보고서

극단값 모형을 기반한 국내 강우 현상 모의

Korean Precipitation Simulation based on Extreme Value Model

김무섭 기후분석팀

연구보고서 2015-08

발간사

강수량 모형은 기상모의와 예측에 있어 핵심적인 부분이고, 여러 응용분야에서 주요한 역할을 하고 있다. 가뭄 및 홍수 등과 같은 자연재해를 대처하는데 있어 강수량 예측은 기본이며, 농업의 작물모형 운영, 수문의 이수과 치수분야에서 강수모의는 필수 불가결한 요소이다. 우리나라의 경우에는 여름철에 강수가 집중되는 경향이 있으며, 특히, 이 시기에 집중호우가 자주 나타난다. 그래서 여름철 호우를 잘 설명하는 강수모형구축은 기상과 기후, 그리고, 농업, 수문과 같은 응용분야에서 매우 흥미로운 연구주제이다.

본 연구는 이러한 강수모형의 필요성에 발맞추어 이를 구축하는 것을 목적으로 한다. 강수현상에 대한 연구가 긴 역사를 지닌 만큼, 많은 학술적 그리고 실용적 연구가 있다. 구체적으로, 치수분야에서는 극단값 이론을 기반으로 하여 여름철 극한강우의 위험성을 나타내는 여러 지표들이 개발되어 왔고, 농업과 이수분야에서는 강수와 함께 기상의 여러 요소를 모의하는 weather generator가 개발되어 각 분야의 목적에 맞게 쓰이고 있다. 본 연구에서는 극한강우의 위험성 측정에 쓰이는 극단값 모형을 weather generator의 강수모의알고리즘에 융합시켜, 새로운 강수모의알고리즘을 제시한다. 이를 통해 응용분야의 요구에 부합되는 강수자료를 생산할 수 있다.

이 연구에서 개발된 강수모의알고리즘은 다지점의 강수를 재현할 수 있다. 그리고, 앞으로 여타 기상변수의 모의알고리즘과 합쳐져 새로운 다지점 weather generator 개발에 활용될 것이다. 다지점 weather generator는 기후자료의 상세화(downscaling)의 도구로 주목을 받는다. 기후자료 상세화란 시공간적으로 큰 규모의 기후자료로부터 작은 시공간규모의 기상자료를 생산하는 과정으로 기후자료의 농업과 수문분야에 적용에 있어 필수적인 부분이다. 농업분야에서는 weather generator는 시간적 상세화에 활용되는데, 다지점 weather generator는 시공간적으로 이를 가능하게 한다는 점에서 매우 값진 연구일 것이라 확신한다.

앞서 언급하였듯이, 본 연구는 향후 다지점 weather generator 개발과 APCC가 역점을 두고 있는 상세화의 초석이 될 것이다. 이러한 연구방향의 가치를 인정하고 지원을 아끼지 않은 APCC의 모든 일원에게 감사의 말을 전한다.

2015. 3월
APEC 기후센터 소장
정진승

ABSTRACT

In this study, we analyze heavy precipitation of some regions in the Korea peninsula to simulate rainfalls in rainy season. We use hourly precipitation data and generate series of rainfall events that seem to be statistically independent. By following the procedure of extreme value analysis, a multivariate extreme value model is set up for extremely large rainfall amounts of several sites covering a region. Moreover, based on this model, we propose a rainfall simulation algorithm which will be a foundation for the study of multisite weather generator.

We have worked for the upper basins of Nakdong and Han rivers and established multivariate extreme value models. Moreover, in order to completely simulate rainfalls in these basins, we have also analyzed durations of heavy rainfall events and interval times between adjacent rainfall events. As results, Gamma Generalized Linear Models are set up for durations, while random sampling turns out to be appropriate for simulating intervals. Using the combination of these resulting models, we carry out simulations and compare the synthetic data with the observations in terms of cumulated amount in the whole rainy season, maxima of daily rainfalls and 3-day cumulated rainfalls, the number of wet days.

목차

극단값 모형을 기반한 국내 강우 현상 모의

발간사	i
ABSTRACT	iii
1. 서론	1
2. 연구 배경 및 방법	2
2.1 연구 배경	2
2.2 극단값 모형과 모의	4
2.2.1 극단값 모형	5
2.2.2 다변량 극단값 모형을 통한 다지점 강수량 모의	12
2.3 극단값 분석	14
2.3.1 Generalized Pareto model의 적합 및 진단	15
2.3.2 다변량 극단값 모형의 적합	21
2.3.3 다변량 극단값 모형의 진단	23
3. 연구내용	27
3.1 낙동강 상류 유역 호우 강수량 분석	27
3.1.1 강우사상자료(rainfall event data)의 생성	27
3.1.2 지점별 호우 강수량 분석	33
3.1.3 다지점 호우 강수량 분석	37
3.1.4 강수지속시간 및 사상간 시간간격 분석	47
3.2 한강 상류 유역 호우 강수량 분석	55
3.2.1 지점별 호우 강수량 분석	55
3.2.2 다지점 호우 강수량 분석	55
3.2.3 강수지속시간 및 사상간 시간간격 분석	60
3.3 호우기 강우 모의	63
3.3.1 강의 모의 방법	63
3.3.2 모의 유사성 평가	67
4. 결론 및 토론	139
REFERENCES	142
부록	144

1. 서론

우리나라 여름철 강우는 다른 계절에 비해 강수량이 많고 발생빈도가 높아서, 홍수, 산사태, 도시침수와 같은 재해로 이어질 가능성이 크다. 그래서 여름철에 자주 발생하는 호우는 여러 분야의 연구대상이다. 대표적으로 수문분야에서는 호우의 발생빈도와 강도 등을 분석하는 여러 연구가 있으며, 현재도 전지구적 기후변화의 영향을 받는 호우를 예측하고 대비하기 위한 노력이 이어지고 있다(김병식 외(2008), 윤선권 외(2014)).

강우를 포함한 다양한 기상현상을 설명하고 예측하기 위한 여러 연구가 있고, 그 결과가 농업, 수문과 같이 기상과 밀접한 관계가 있는 여러 응용 분야에 적용되고 있다. 기후학에서는 대기순환과 이로부터 비롯되는 기상을 설명하기 위해 기후역학모형(climate dynamical model)을 개발해 왔고, 응용분야에서는 이 모형에서 생산된 기상예측자료를 해당 분야의 연구에 적용해 오고 있다. 그런데, 기후역학모형에서 생산된 기상예측자료의 시공간적 세밀도가 응용 분야에서 요구되는 정도보다 낮아서, 실제 적용에 있어서는 이 차이에 대한 보완이 필요하다. 널리 쓰이는 보완의 방법 중 하나가 weather generator이다. Weather generator는 통상적으로 월단위로 생산되는 기상예측자료를 이용하여 응용분야에서 필요로 하는 하루단위의 자료를 생성하는 역할을 한다.

본 연구에서는 우리나라의 호우기인 6월부터 9월 동안 낙동강과 한강 상류유역의 강우를 모의(simulation)하는 것을 목표로 한다. 특히, 이 시기에 자주 발생하는 호우사상을 면밀히 분석하여 모의에 적극적으로 반영하고자 한다. 이들 지역은 산지지형으로 호우에 태생적으로 취약하면서 동시에 수자원 관리 측면에서 요충지역이다. 본 연구에서 이들 지역의 호우를 설명하는 통계모형(statistical model)을 세우고 이를 사용하여 호우기 강우를 모의할 것이다. 또한, 구축된 통계모형은 향후 연구에서 지역의 기상을 하루단위로 재현하는 weather generator의 초석으로 쓰이고자 한다. 이 weather generator의 개발은 이들 상류유역의 수자원 관리 연구에 큰 보탬이 될 것이다.

연구보고서의 구성은 다음과 같다. 2절에서는 본 연구의 배경과 호우사상의 강수량을 분석하기 위한 통계적 방법인 극단값 분석법을 소개한다. 세부적으로 연구의 배경이

되는 기존 강우모의알고리즘의 한계점을 논하고, 극단값 모형, 모의방법, 분석과정을 설명한다. 3절에서는 강수량자료의 분석과정과 결과를 제시한다. 극단값 분석법을 낙동강과 한강 상류유역에서 얻어진 강수량자료에 적용하고, 그 결과를 제시한다. 또한, 호우사상의 지속시간과 강우사상 간의 시간간격을 분석하여, 최종적인 강우통계모형을 구축한다. 그리고 구축된 모형을 이용한 새로운 강우모의 방법을 제안하고, 이 방법을 통해 얻어진 모의자료와 관측자료를 비교하여 재현성을 평가한다. 4절에서는 본 연구의 내용을 정리하고, 내년도 연구주제인 weather generator 개발과의 연계성을 간단히 설명하는 것으로 본 보고서를 마무리한다.

2. 연구 배경 및 방법

2.1 연구 배경

본 연구의 배경을 설명하기 위해 기존의 주요한 강우모의방법을 간단히 소개하고 한계점을 논한다. 여기서 소개할 모의방법은 weather generator로서 널리 쓰이고 있는 WGEN(Richardson(1981))와 LARS-WG(Racsko et al.(1991)), 그리고 k -NN(k -nearest neighbor) 알고리즘 기반의 semi-parametric weather generator(Apipattnanavis et al.(2007))의 강우알고리즘이다.

먼저, WGEN의 강우알고리즘을 설명한다. 강우알고리즘은 하루단위로 강우를 모의하며, 크게 습윤일 모의와 일강수량 모의로 나누어진다. 습윤일 모의는 날의 상태를 습윤과 건조로 이분한 2-state Markov chain 모형을 바탕으로 수행되며, 모형의 전이확률(transition probability)은 계절성을 반영하기 위해 통상적으로 월단위로 추정된다. 강수량 모의는 Gamma 분포를 쓰며, 습윤일 모의에서와 같이 월단위로 분포의 모수가 추정된다. 하지만 2-state Markov chain은 연속된 습윤일 또는 건조일을 잘 모의하지 못한다고 알려져 있다(Young(1994)). 또한 Gamma 분포를 통한 모의는 적합결여(lac of fit)로 인해 호우와 같이 일강수량이 많은 경우에 정확도가 떨어질 수 있다.

LARS-WG의 강우알고리즘은 WGEN의 단점을 많이 보완하였다. 우선, 습윤일 모의는 습윤기간과 건조기간을 번갈아 가며 생성하는 방식으로 진행된다. 여기서는 간단히 이를 ‘alternating method’라고 부르기로 한다. 기간의 길이는 준경험분포함수(semi-empirical distribution function)를 이용하여 모의한다. 그리고 강수량도 Gamma 분포를 대신해 역시 준경험분포함수를 적합하여 모의된다. 그래서 연속된 습윤일을 잘 모의할 수 있으며, Gamma 분포 적합으로 인해 나타날 수 있는 적합결여를 보완할 수 있다. 하지만 여전히 일강수량이 많은 강우에 대한 모의 정확도는 떨어진다. 예를 들어, 준경험분포함수를 쓰게 되면, 적합에 쓰인 일강수량 자료의 최댓값보다 더 큰 강수량을 모의할 수가 없다. 그래서 미래에 나타날 수 있는 극한강우에 대한 모의가 이루어지지 않아 홍수위험을 과소추정하게 된다.

WGEN과 LARS-WG의 강우알고리즘은 단일지점에서만 수행된다. 반면에 semi-parametric weather generator는 다지점 강우모의에 적용이 가능하다는 장점이 있다. 강우알고리즘은 역시 습윤일 모의와 일강수량 모의로 나누어진다. 습윤일 모의에서 WGEN에서와 같이 Markov chain 모형을 사용하지만 날의 상태를 dry, wet, extreme wet으로 세분하여 연속된 습윤일 모의를 향상시킨다. 모의 기간 동안의 습윤일 모의가 완료되면, 각 습윤일에 대해서 일강수량을 생성한다. 일강수량 생성은 모의기간 시작시점부터 순차적으로 진행된다. 한 습윤일의 강수량 y_t 을 모의하기 위해, 바로 전일의 상태 s_{t-1} 와 일강수량 y_{t-1} 그리고 해당일의 상태 s_t 등의 정보를 활용하여, 해당 습윤일의 상황과 유사한 여러 경우를 과거 자료에서 선별한다. 알고리즘에서 통상적으로 그 경우의 수를 k 로 나타내며, 선별된 경우를 ‘neighbor’라고 부른다. 선별된 각 neighbor는 이어진 두 날의 일강수량과 상태를 담고 있다. 유사성을 기준으로 확률을 부여하여 neighbor 중 하나를 선정한다. 그 선정된 neighbor를 $(s_{t-1}^*, y_{t-1}^*, s_t^*, y_t^*)$ 으로 나타낸다면, 이튿날 일강수량 y_t^* 을 해당 습윤일의 강수량으로 부여한다. 즉, $y_t = y_t^*$ 으로 둔다. 이 일강수량 생성과정은 s_{t-1} , y_{t-1} , s_t 가 주어졌을 때 y_t 의 조건부확률분포(conditional distribution)를 근사하는 것으로 k -NN 알고리즘의 한 예이다(Altman(1992), Coomans and Massart(1982)). 알고리즘의 세부사항은 Apipattanavis et al.(2007)을 참고하기를 바

란다. k -NN 방법은 적용에 있어 유연하여 확장가능성이 크다는 장점이 있다. 하지만 이 방법도 또한 LARS-WG에서처럼 자료값을 그대로 활용하여 강수량을 생성하기 때문에, 호우를 모의하는데 한계가 있다.

표 1에서 기존의 모의방법을 요약해서 비교해 보았다. 본 연구에서는 호우를 잘 설명하면서 지역의 강우를 모의하는 것이 목표이므로, 다지점 강우알고리즘이 필요하다. 그런데 WGEN과 LARS-WG 방법은 다지점 알고리즘이 아니며, semi-parametric weather generator는 다지점 알고리즘이기는 하지만 호우 모의에 한계가 있다. 그래서 본 연구에서 다지점 강우 모의가 가능하면서 호우모의에도 적합한 새로운 모의방법을 고려하고자 한다. 수문에서는 일찍이 호우의 위험도측정을 위해 강수량분석에 극단값 모형(extreme value model)을 도입하였다. 본 연구에서도 이 모형을 강우모의에 도입하여 기존 방법의 단점을 보완하고자 한다.

Table 1. 강우모의방법의 비교

	WGEN	LARS-WG	semi-parametric
습윤일 모의	2-state Markov chain	wet-dry period alternating method	3-state Markov chain
강수량 모의	Gamma 분포	준경험분포 (semi-empirical distribution)	k -NN 알고리즘
한계점	단일지점모형, Gamma 분포의 적합결여 가능성	단일지점모형, 호우모의의 한계	호우모의의 한계

2.2 극단값 모형과 모의

본 연구의 목표인 지역의 호우를 잘 설명하는 강우모의방법을 구축하기 위해서는, 기본적으로 호우사상의 다지점 강수량을 생성하는 확률모형이 필요하다. 이를 위해서 다변량 극단값 모형(multivariate extreme value model)을 고려한다. 본 소절에서는 간

단하게 극단값 모형을 소개하고, 이어 호우사상의 다지점 강수량 생성을 위한 다변량 극단값 모의방법을 소개한다.

2.2.1 극단값 모형

극단값 모형이란 극단값의 확률적 성질을 기술하는 확률 모형(stochastic model)이다. 크게 일변량 모형(univariate model)과 다변량 모형(multivariate model)으로 나누어 볼 수 있으며, 일변량 모형은 다시 그 표현과 분석적용방식에 따라 generalized extreme value distribution와 generalized Pareto model으로 나눌 수 있다. 전자는 연최대일별강수량(annual maximum of daily precipitation)과 같은 block maxima 분석에 활용되는 반면, 후자는 기준값(threshold) 이상의 변량을 분석하는 threshold model분석에 쓰인다. 사실, 후자는 전자의 경우처럼 block maxima를 이용해서 생기는 자료의 손실을 줄여보고자 하는 노력에서 나온 모형으로, 보다 더 효율적인(efficient) 분석을 할 수 있다(Coles et al.(2001)). Generalized extreme value distribution을 이용한 국내강수량연구로는 성용규, 손중권(2013), 신지예 외 (2013)를, generalized Pareto model의 연구로 엄명진 외 (2008)를 참고하기를 바란다. 본 연구에서는 generalized Pareto model을 지점 강수량 분석에 활용하고자 한다.

Table 2. 극단값 모형의 분류

분류	세부분류	비고
일변량 모형	generalized extreme value distribution	block maxima 분석에 활용
	generalized Pareto model	threshold model 분석에 활용
다변량 모형	multivariate extreme value distribution	주변확률분포는 모두 표준 Fréchet 분포로 표준화, spectral 밀도함수에 따라 분포형태가 결정

2.2.1.1 Generalized Pareto model

Generalized Pareto model을 설명하기 앞서, generalized Pareto 분포를 간단히 소개하고자 한다. 이름에서 알 수 있듯이 generalized Pareto 분포는 기존의 Pareto 분포를 확장하여 얻어진 확률분포이다. 분포함수의 형태는

$$GP(z; \sigma, \gamma) = \begin{cases} 0, & z < 0; \\ 1 - \left(1 + \gamma \frac{z}{\sigma}\right)^{-1/\gamma}, & z \geq 0, 1 + \gamma \frac{z}{\sigma} > 0; \\ 1, & z \geq 0, 1 + \gamma \frac{z}{\sigma} \leq 0, \end{cases} \quad (1)$$

으로, 두 개의 모수(parameter) $\sigma > 0$ 와 $\gamma \in \mathbb{R}$ 를 지니고 있다. 단, 식(1)의 우변에서 $\gamma = 0$ 일 경우는 $\gamma \rightarrow 0$ 일 때 극한인 $1 - \exp[-z/\sigma]$ 으로 간주한다:

$$GP(z; \sigma, 0) = \begin{cases} 0, & z < 0; \\ 1 - \exp\left(-\frac{z}{\sigma}\right), & z \geq 0, \end{cases} \quad (2)$$

각 모수는 분포함수에서 역할에 따라, σ 는 척도모수(scale parameter), γ 는 형태모수(shape parameter)로 불리고 있다. 형태모수인 γ 가 0보다 큰 경우가 일반적으로 말하는 Pareto 분포이다. 그림 1-2는 σ 와 γ 의 값에 따른 generalized Pareto 분포의 모습을 그림 3-4는 분위수(quantile)를 나타낸다. σ 가 커질수록 분포는 양의 방향으로 퍼지고, γ 가 커질수록 분포의 꼬리가 두꺼워진다. 강수량의 해석에서는, 두 모수의 값이 커지면 커질수록 강한 비가 내릴 확률이 높다는 것을 의미한다.

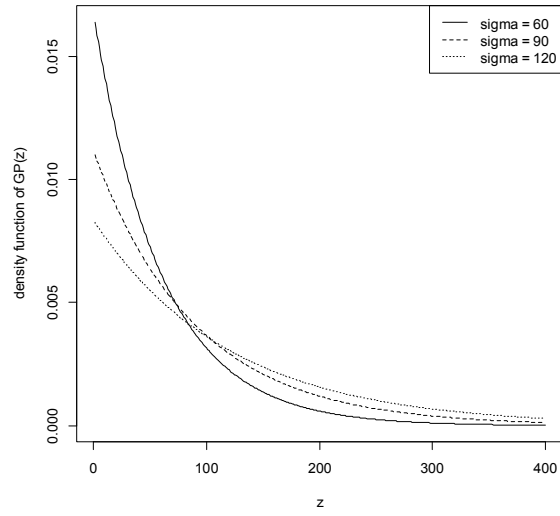


Figure 1. σ 에 따른 generalized Pareto 분포

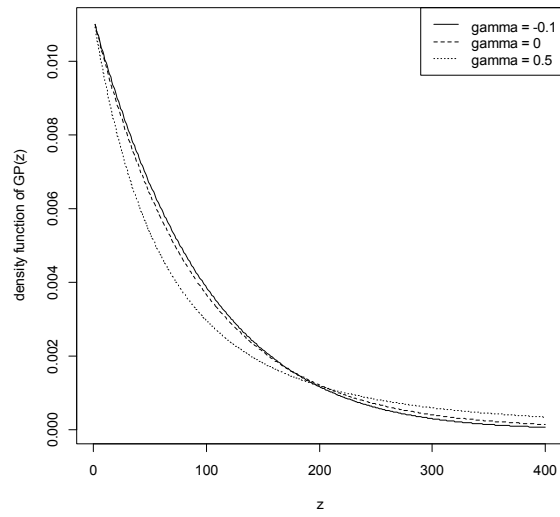


Figure 2. γ 에 따른 generalized Pareto 분포

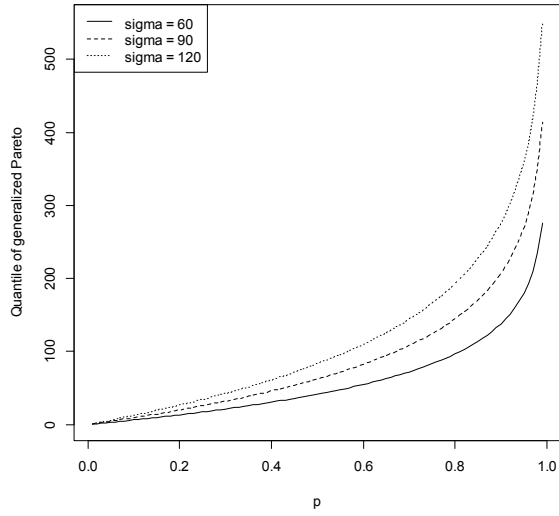


Figure 3. σ 에 따른 generalized Pareto 분포의 분위수: σ 가 클수록 더 빠르게 증가한다.

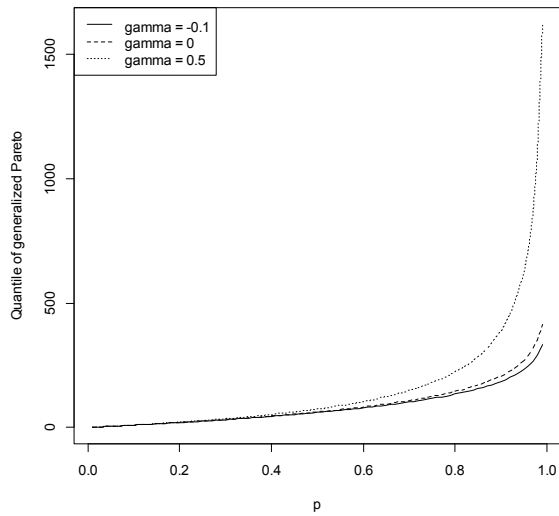


Figure 4. γ 에 따른 generalized Pareto 분포의 분위수: γ 가 클수록 p 가 1에 가까울 때, 증가가 가파르다.

Generalized Pareto model이란 일종의 threshold model로서, 주어진 기준값을 넘는 변량의 초과량이 근사적으로 generalized Pareto 분포를 따른다는 수리적 사실에 기인한다(Coles et al.(2001) Theorem 4.1). 구체적으로 변량 Y 가 확률분포 F 를 따르는 확률변수(random variable) - 모든 $y \in \mathbb{R}$ 에 대해서 $F(y) = \Pr(Y \leq y)$ 일 때, 충분히 큰 $b \in \mathbb{R}$ 대해서

$$\frac{1 - F(z + b)}{1 - F(b)} \approx 1 - \text{GP}(z; \sigma, \gamma), \quad z > 0, \quad (3)$$

을 만족하는 $\sigma > 0$ 와 $\gamma \in \mathbb{R}$ 이 존재하는 것이 알려져 있다. 여기서, γ 는 기준값 b 에 상관 없이 일정한 반면에 σ 는 달라질 수 있으며, 근사식은 b 가 커질수록 정확해 진다. 식(3)에서 좌변은 $Y > b$ 일 경우에, 초과량 $Y - b$ 가 z 보다 클 조건부 확률을 뜻한다. 즉, generalized Pareto model은 초과량의 조건부 확률분포에 대한 근사식을 제시하고 있다.

2.2.1.2 다변량 극단값 분포

이제 다변량 극단값 분포(multivariate extreme value distribution)를 소개하고자 한다. d -변량 극단값 분포 G 의 형태는 다음과 같다(d 는 2 이상의 자연수):

$$G(x_1, \dots, x_d) = \exp \left[- \int_{\mathbf{w} \in \mathbb{S}^{d-1}} \max \left(\frac{w_1}{x_1}, \dots, \frac{w_d}{x_d} \right) h(\mathbf{w}) dw_1 \cdots dw_{d-1} \right],$$

여기서,

$$\mathbb{S}^{d-1} = \{ \mathbf{w} = (w_1, \dots, w_d) : \min\{w_1, \dots, w_d\} \geq 0, w_1 + \dots + w_d = 1 \}$$

이고, h 는 \mathbb{S}^{d-1} 상의 양함수(positive function)로서, $i = 1, \dots, d$ 에 대해서

$$\int_{\mathbf{w} \in \mathbb{S}^{d-1}} w_i h(\mathbf{w}) dw_1 \cdots dw_{d-1} = 1 \quad (4)$$

를 만족한다.¹⁾ h 를 G 의 ‘spectral 밀도함수’라고 부르며, 사실상 G 와 동일시 할 수 있다. 또한, 식(4)에 의해 G 의 주변확률분포가 표준 Fréchet 분포이다. 즉, (X_1, \dots, X_d) 가 분포 G 를 따른다면, $i = 1, \dots, d$ 에 대해서,

$$\Pr(X_i \leq x) = \begin{cases} \exp\left(-\frac{1}{x}\right), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

이다. 그래서 문헌에 따라 G 를 ‘표준 다변량 극단값 분포’ (standard multivariate extreme value distribution)라고 부르기도 한다.

다변량 분포 G 에서 spectral 밀도함수 h 는 변량들 사이의 관계를 설명하는 역할을 하기 때문에, h 추정에는 다변량 극단값 분석에서 핵심적인 부분이다. 그런데, h 의 형태와 관련된 조건은 식(4)이 외에는 없기 때문에, 그 자체로는 매우 복잡하다. 그래서 실제 분석에서는 h 의 형태에 대한 적절한 regularization이 필요하다. 여기서는 대표적인 모형으로 logistic 모형을 소개한다. Logistic 모형은 널리 쓰이는 모형 중에서 가장 단순하고, 분포함수와 spectral 밀도함수의 형태가 명시적으로 알려져 있다:

$$G(x_1, \dots, x_d) = \exp\left[-\left(\frac{1}{x_1^{1/\alpha}} + \dots + \frac{1}{x_d^{1/\alpha}}\right)^\alpha\right], \quad 0 < \alpha < 1,$$

이고,

$$h(\mathbf{w}) = \left\{ \prod_{j=1}^{d-1} (j/\alpha - 1) \right\} \left(\prod_{j=1}^d w_j \right)^{-(1/\alpha+1)} \left(\sum_{j=1}^d w_j^{-1/\alpha} \right)^{\alpha-d} \quad (5)$$

이다. 단일 모수 모형이기 때문에 추정계산이 쉽다. 하지만 변량의 차수 d 가 증가할수록, 유연성(flexibility)이 떨어진다. 그림 5-7은 spectral 밀도함수의 그래프를 나타낸다. α 가 0에 가까울수록 변량사이의 연관성이 강하고, 반대로 1에 가까우면 독립(independence)에 가깝다.

1) 다변량 극단값 분포의 일반적인 형태는 Lebesgue integration으로 표현된다.

변량들 사이의 연관성을 나타내는 일반적인 도구로 copula가 있다. Copula란 균일분포(uniform distribution)를 따르도록 변환된 변량의 결합확률분포(joint probability distribution)로서 정의된다. 구체적으로, (X_1, \dots, X_d) 가 분포 G 를 따르고, G_i 가 X_i 의 주변확률분포라고 했을 때 ($i = 1, \dots, d$), (X_1, \dots, X_d) 의 copula C 는

$$C(z_1, \dots, z_d) = \Pr(G_1(X_1) \leq z_1, \dots, G_d(X_d) \leq z_d)$$

으로 정의된다. Copula C 로부터 극단영역에서 (X_1, \dots, X_d) 의 연관성을 나타내는 측도(measure)가 여러 연구에서 도입되었다(Coles et al.(1999), Frahm(2006)). 간단한 측도로서, $d = 2$ 인 경우,

$$\chi = \lim_{z \rightarrow 1} \frac{\Pr(G_1(X_1) > z, G_d(X_2) > z)}{\Pr(G_1(X_1) > z \text{ or } G_2(X_2) > z)} = \lim_{z \rightarrow 1} \frac{1 - z + C(z, z)}{1 - C(z, z)} \quad (6)$$

를 tail dependence coefficient라고 부르며, 이는 어느 한 변량이 극단적으로 클 때 두 변량이 동시에 극단적으로 클 조건부 확률로 해석된다. 특히, G 가 logistic 모형인 경우에는

$$\chi = 2 - 2^\alpha \quad (7)$$

가 성립한다(Coles et al.(2001) 식(8.29)). χ 은 항상 $0 \leq \chi \leq 1$ 를 만족하고, 1에 가까울수록 연관성이 강하는 것을 의미한다. 특히, 0일 경우는 두 변량이 ‘극단영역에서 점근적으로 독립’ (asymptotical extremal independence)이라고 부르며, 이는 강수량의 해석에서 X_1, X_2 가 두 지점의 강수량을 의미할 때, 극단호우가 두 지점에서 동시에 일어날 가능성이 없다는 것을 의미한다. $d > 2$ 인 경우에 대해서는 Frahm(2006)를 참고하기를 바란다.

2.2.2 다변량 극단값 모형을 통한 다지점 강수량 모의

d 개 지점의 강수량(Y_1, \dots, Y_d)의 다변량 극단값²⁾은 변량들의 주변확률분포와 표준 다변량 극단값 분포의 결합으로 재현된다. 우선, 변환

$$Z_j = \frac{1}{1 - F_j(Y_j)}, \quad j = 1, \dots, d \quad (8)$$

을 고려한다. 여기서, F_j 는 Y_j 의 분포함수이다. 각 F_j 가 연속이면,

$$\Pr(Z_j \leq z) = \begin{cases} 1 - \frac{1}{z}, & z > 1; \\ 0, & z \leq 1, \end{cases}$$

즉, 각 Z_j 가 모두 표준 Pareto 분포를 따른다. 그러면, 다변량 극단값 분포의 multivariate regular variation 성질에 의해, 표준 다변량 극단값 분포 G 가 존재해서,

$$\Pr(Z_1 > z_1 \text{ or } \dots \text{ or } Z_d > z_d) \sim 1 - G(z_1, \dots, z_d) \quad (9)$$

가 성립한다. 여기서, $\min(z_1, \dots, z_d) \rightarrow \infty$ 일 때, 근사식은 더 정확해 진다. (9)의 더 자세한 결과는 Resnick(1987)의 Proposition 5.17를 참고하기를 바란다.

한편, 식(8)의 변환은 실제 분석에서 F_j 를 모르기 때문에 활용이 불가능하다. 일반적인 분석에서는 경험분포함수(empirical distribution function)로써 F_j 를 근사하는데, 변량 Y_j 에 대한 자료 $y_{1,j}, \dots, y_{n,j}$ 가 주어진 경우, 경험분포함수는

$$\hat{F}_j(y) = \frac{1}{n} \sum_{i=1}^n I(y_{i,j} \leq y)$$

으로 정의된다.³⁾ 하지만 경험분포함수는 y 가 커질수록 정확도가 떨어지기 때문에, 극단

2) 다변량 극단값 또는 vector의 극단값은 원점으로부터 멀리 떨어진 값을 의미한다.

3) $I(\cdot)$ 는 indicator 로서 \cdot 의 내용이 참이면 1, 거짓이면 0이다.

값 분석에서는 이에 대한 보완이 필요하다. 이 보완책을 generalized Pareto model로 부터 얻을 수 있다. 충분히 큰 기준값 b_j 에 대해서, $p_j = 1 - \widehat{F}_j(b_j)$ 라고 두고,

$$\widetilde{F}_j(y) = \widetilde{F}_j(y; \sigma_j, \gamma_j) = \begin{cases} 1 - p_j \{1 - \text{GP}(y - b_j; \sigma_j, \gamma_j)\}, & y > b_j; \\ \widehat{F}_j(y), & y \leq b_j, \end{cases} \quad (10)$$

라고 정의한다. 단, $\sigma_j > 0$ 와 $\gamma_j \in \mathbb{R}$ 는 F_j 에 대한 generalized Pareto 근사에 의한 값이다. 그래서 우리는 Z_j 에 대한 근사값으로

$$\widetilde{Z}_j = \frac{1}{1 - \widetilde{F}_j(Y_j)} = \begin{cases} \frac{1}{p_j \{1 - \text{GP}(Y_j - b_j; \sigma_j, \gamma_j)\}}, & Y_j > b_j; \\ \{1 - \widehat{F}_j(Y_j)\}^{-1}, & Y_j \leq b_j, \end{cases}$$

를 얻는다. 요약하면 다음과 같다.

- 다지점 강수량의 극단확률분포⁴⁾는 주변확률분포 F_1, \dots, F_d 그리고 표준 다변량 극단값 분포의 spectral 밀도함수 h 에 의해 근사적으로 결정된다.
- F_j 는 generalized Pareto 근사로 개선된 경험분포함수(10)으로 근사된다.

이 이론을 바탕으로 다지점 강수량의 극단값 모의는 다음과 같은 단계로 수행된다:

- i. spectral 밀도함수 h 에 의한 표준 다변량 극단값 분포 G 로부터 Z_1^*, \dots, Z_d^* 는 생성한다. 단, 식(9)의 관점에서

$$p_1 Z_1^* > 1 \quad \text{or} \quad \dots \quad \text{or} \quad p_d Z_d^* > 1 \quad (11)$$

가 만족하도록 한다. 즉, 식(11)이 만족될 때까지 Z_1^*, \dots, Z_d^* 를 반복해서 생성하여, 최종의 Z_1^*, \dots, Z_d^* 을 택한다.

4) 극단확률분포는 원점에서 멀리 떨어진 영역에서의 확률분포를 의미한다.

ii. Generalized Pareto 근사로 얻어진 $\sigma_j > 0$, $\gamma_j \in \mathbb{R}$ 에 대해, 변환

$$Y_j^* = \begin{cases} \sigma_j \text{BC}(p_j Z_j^*; \gamma_j) + b_j, & Z_j^* > 1/p_j; \\ \widehat{F}_j^{\leftarrow}(1 - 1/Z_j^*), & Z_j^* \leq 1/p_j, \end{cases}$$

를 통해서 Y_1^*, \dots, Y_d^* 를 얻는다. 여기서, $\widehat{F}_j^{\leftarrow}$ 는 \widehat{F}_j 의 분위수함수(quantile function)으로서

$$\widehat{F}_j^{\leftarrow}(p) = \min\{y : \widehat{F}_j(y) \geq p\}, \quad 0 < p \leq 1,$$

으로 정의되고, 함수 $\text{BC}(\cdot; \gamma)$ 는 Box-Cox 변환으로

$$\text{BC}(y; \gamma) = \begin{cases} \frac{y^\gamma - 1}{\gamma}, & \gamma \neq 0; \\ \log(y), & \gamma = 0, \end{cases} \quad y > 0,$$

이다. 그러면

$$Y_1^* > b_1 \quad \text{or} \quad \dots \quad \text{or} \quad Y_d^* > b_d$$

이 충족되는 다지점 강수량을 얻는다.

2.3 극단값 분석

호우사상의 강수량을 생성하는 극단값 모형을 얻기 위해서는 이 모형이 적용될 영역을 지정하는 기준값 b_1, \dots, b_d 을 설정해야 하고, 모형식에서 등장하는 모수 $\sigma_1, \dots, \sigma_d, \gamma_1, \dots, \gamma_d, \alpha$ 를 자료를 이용해 추정해야 한다. 본 소절에서는 이를 위한 극단값 분석법(extreme value analysis)을 소개한다.

2.3.1 Generalized Pareto model의 적합 및 진단

다변량 분석에서 각 변량에 대한 일변량 분석을 통해 필요한 정보를 얻는 경우가 많다. 이는 다지점 강수량 분석에서도 예외가 아닐 것이다. 그래서 본 절에서는 지점별 분석에 쓰일 generalized Pareto model 적합과 진단에 대해 설명하고자 한다.

앞서 말하였듯이, generalized Pareto model은 threshold model의 일종으로, 이 모형을 이용한 분석에서는 우선 적절한 기준값을 선정해야 한다. 선정 방식으로 여러 가지가 있으나 기본적인 그림도구 중 하나인 mean residual life plot을 소개한다(Coles et al.(2001) 4.3.1절). 이 그림은 탐색적(exploratory) 성격을 띠고 있는데, 그림을 그리는 방식은 다음과 같다: 자료 y_1, \dots, y_n 가 주어졌을 때,

$$\left\{ \left(b, \frac{1}{k(b)} \sum_{i=1}^n (y_i - b) I(y_i > b) \right) : \min(y_1, \dots, y_n) < b < \max(y_1, \dots, y_n) \right\},$$

으로 정의되는 점들의 그래프이다. 단, $k(b)$ 는 y_1, \dots, y_n 중 b 보다 큰 자료값의 개수이다. mean residual life plot에서 그래프가 직선의 형태로 나타나는 b 의 영역을 찾는다(Coles et al.(2001) 4.3.1절). 그림 5은 안동지점 강수량의 mean residual life plot인데, 그래프가 기준값 b 가 50 mm 이상에서 그래프가 직선에 가깝게 나타난다. 그래서 50 mm 이상에서 기준값을 선정하는 것이 적절하다고 판단된다.

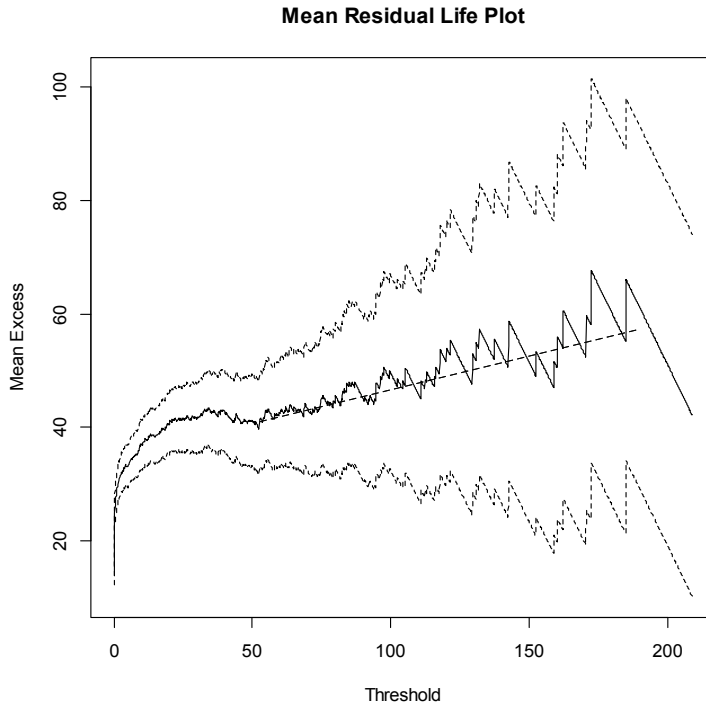


Figure 5. mean residual life plot: mean excess가 선형적으로 증가하는 영역에서 기준값을 정한다.

기준값을 선택하였다면, 다음으로 모수 $\sigma > 0$ 와 $\gamma \in \mathbb{R}$ 을 최대우도추정법(maximum likelihood estimation)으로 추정한다. log-likelihood 함수는

$$\ell(s, u) = \sum_i \log \frac{\partial}{\partial z} \text{GP}(z; s, u) \Big|_{z=y_i-b}$$

으로 정의되며, 여기서 \sum_i 에서 i 는 $y_i > b$ 가 만족하는 범위에서 움직인다. 추정치(estimate)는 함수 $\ell(s, u)$ 를 최대로 하는 (s, u) 의 값으로 정의가 되며, $(\hat{\sigma}, \hat{\gamma})$ 로 표기한다. 또한, 추정의 불확실성(uncertainty)을 나타내는 추정량(estimator)의 표준오차(standard error)와

공분산행렬은 행렬

$$\begin{bmatrix} -\frac{\partial^2}{\partial s \partial s} \ell(\hat{\sigma}, \hat{\gamma}) & -\frac{\partial^2}{\partial s \partial u} \ell(\hat{\sigma}, \hat{\gamma}) \\ -\frac{\partial^2}{\partial u \partial s} \ell(\hat{\sigma}, \hat{\gamma}) & -\frac{\partial^2}{\partial u \partial u} \ell(\hat{\sigma}, \hat{\gamma}) \end{bmatrix}$$

의 역행렬에서 얻어진다.

국내 강수량 분석에서는 종종 강수량의 분포가 Gumbel 분포인 것으로 가정된다. 이 가정은 $\gamma=0$ 인 경우에 해당한다. 강수량 분포가 Gumbel로 확신하는 경우에는 $\ell(s, u)$ 대신에 $\ell(s, 0)$ 를 최대화시켜 σ 의 추정치를 구한다. Gumbel 분포인지 확인이 필요하다면, 우도비 검정(likelihood ratio test)를 활용한다. $\ell(s, 0)$ 를 최대화시켜 얻은 σ 의 추정치를 $\hat{\sigma}^*$ 라고 했을 때,

$$2\{\ell(\hat{\sigma}, \hat{\gamma}) - \ell(\hat{\sigma}^*, 0)\} \leq \chi_{0.05}^2(1)$$

이면 유의수준 5%에서 $\gamma=0$ 이라고 판단한다. 여기서, $\chi_{0.05}^2(1)$ 는 자유도가 1인 χ^2 분포의 상위 5% 분위수이다.

사실, 추정치 $(\hat{\sigma}, \hat{\gamma})$ 는 기준값 b 에 따라 달라진다. Mean residual life plot과 더불어, 이러한 성질을 이용한 기준값 선택 방식으로 threshold choice plot이 있다. 추정치가 기준값 b 에 따라 달라진다는 점을 강조하기 위해 $\hat{\sigma}(b) = \hat{\sigma}$ 그리고 $\hat{\gamma}(b) = \hat{\gamma}$ 라고 표현한다면, threshold choice plot은

$$\{(b, \hat{\sigma}(b) - \hat{\gamma}(b)b) : a < b < c\}, \quad \{(b, \hat{\gamma}(b)) : a < b < c\}$$

으로 정의된다. 여기서, a 와 c 는 추정치가 잘 구해지는 범위에서 선택된다. $\hat{\sigma}(b) - \hat{\gamma}(b)b$ 를 편의상 ‘수정된 척도모수’(modified scale parameter)라고 부르는데, 이를 사용하는 것은 generalized Pareto 근사에서 $\sigma - \gamma \cdot b$ 가 일정하다는 사실에 의거한다. Threshold choice plot에서 그래프가 안정적으로 나타나는 기준값 b 의 영역을 찾는다(Coles et

al.(2001)4.3.4절). 그림 6는 안동지점의 threshold choice plot이다. Mean residual life plot에서와 비슷하게 대략 50mm 와 100mm 사이에서 그래프가 안정적으로 나타나기 때문에, 이 구간에서 기준값을 선정하는 것이 적절하다고 판단된다.

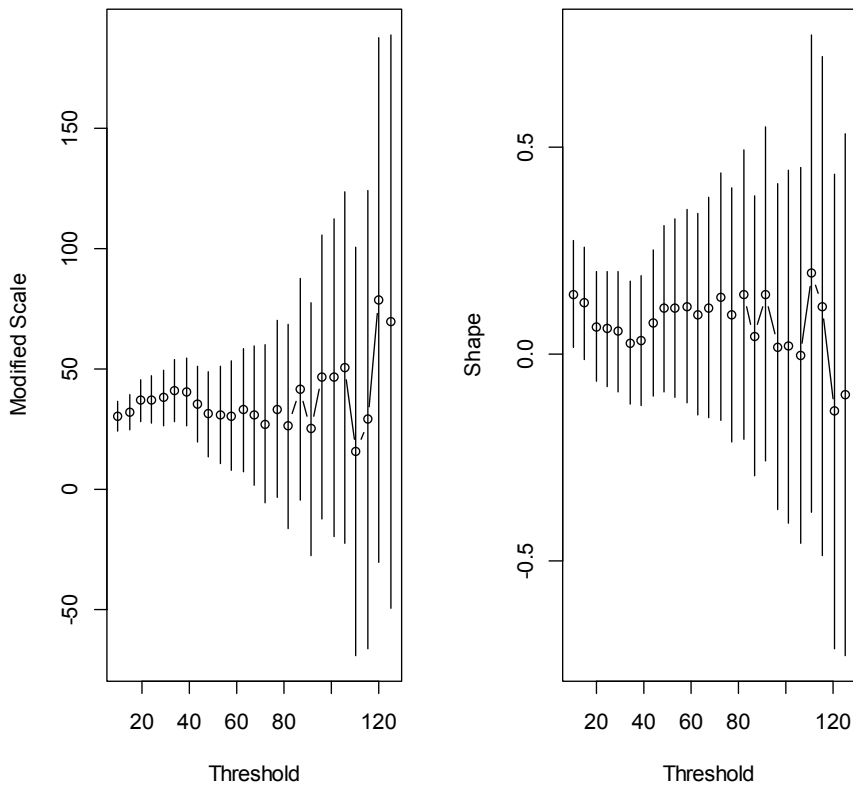


Figure 6. threshold choice plot: 추정치가 안정적으로 나타나는 영역에서 기준값을 정한다.

기준값 선정과 모형적합이 완료되면, 마지막으로 적합모형이 과연 자료에 적절한지를 살펴보아야 한다. 일반적으로 통계분석에서 이러한 단계를 모형진단(model diagnostic)이라고 부른다. Generalized Pareto 모형분석에서는 통상적으로 probability plot, 분위수그림(quantile plot), return level plot을 이용하거나, 추정된 밀도함수와 자료의 히스토그램의 비교를 통해 모형진단을 수행한다. 여기서는 분위수 그림과 return level

plot을 간단히 소개한다. 우선, 분위수 그림은 초과량의 표본분위수(sample quantile)와 적합모형의 분위수를 비교하는 그림으로,

$$\left\{ \left(\hat{\sigma} \cdot \text{BC} \left(\frac{k+1}{i}; \hat{\gamma} \right) + b, y_{(i)} \right) : i = 1, \dots, k \right\}$$

으로 정의된다. 단, $y_{(i)}$ 는 y_1, \dots, y_n 중에서 i 번째로 큰 값이고, k 는 y_1, \dots, y_n 에서 기준값 b 보다 큰 값의 개수를 나타낸다. 분위수 그림에서 그래프가 원점을 지나고 기울기가 1인 직선에 가까울수록 적합도가 높다는 것을 의미한다. 그림 7은 안동지점의 분위수 그림으로 generalized Pareto 분포가 적절한 것으로 판단된다.

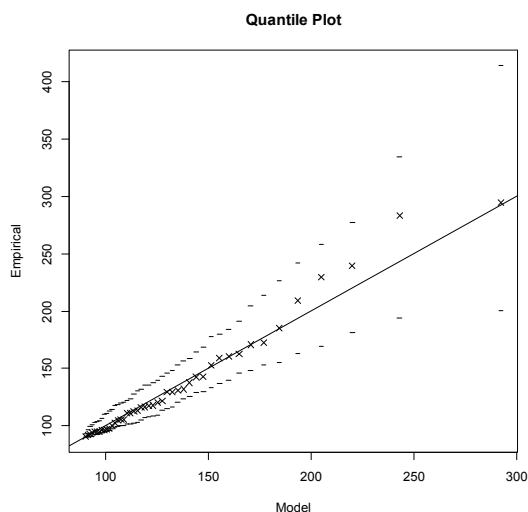


Figure 7. 분위수 그림: 적합모형의 분위수와 표본분위수가 일치할수록 모형이 적합하다.

분위수 그림과 유사한 return level plot도 진단도구로 많이 활용된다. Return level은 주어진 return period에 대응되는 분위수로서, return period m 에 대한 적합모형의 return level은

$$b + \hat{\sigma} \cdot \text{BC} \left(\frac{mk}{n}; \hat{\gamma} \right)$$

으로 주어진다. Return level plot은 return period m 에 따라 적합모형의 return level 과 경험치(empirical estimate)인 $y_{([n/m])}$ 을 나타낸 그림으로 통상적으로 두 그래프

$$\left\{ \left(m, b + \hat{\sigma} \cdot \text{BC} \left(\frac{mk}{n}; \hat{\gamma} \right) \right) : m = 1, 2, \dots \right\}, \quad \left\{ (m, y_{([n/m])}) : m = 1, \dots, n \right\}$$

을 겹쳐 나타낸다. 두 그래프가 일치할수록 적합도가 높다. 그림 8은 안동지점의 return level plot이다. 적합모형과 표본의 return level이 적절히 일치해 보이므로, 모형은 적격한 것으로 판단할 수 있다.

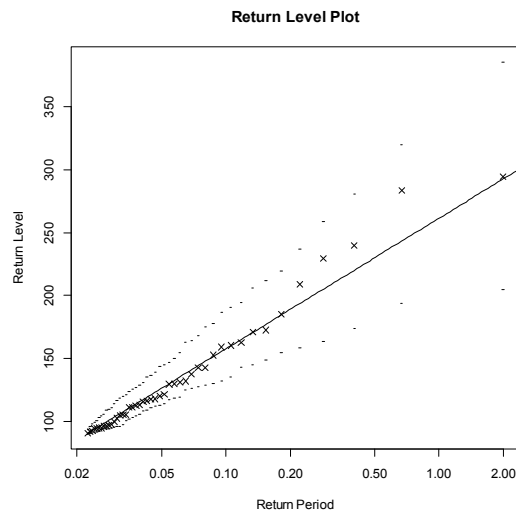


Figure 8. 적합모형의 return level(실선)과 경험치('x'마크). 적합모형의 return level과 경험치가 일치할수록 모형이 적합하다.

2.3.2 다변량 극단값 모형의 적합

미지의 모수 $\sigma_1, \dots, \sigma_d, \gamma_1, \dots, \gamma_d, \alpha$ 의 동시추정(joint estimation)을 위해 최대우도추정법을 고려한다. 지점별 분석을 통해, $\sigma_1, \dots, \sigma_d, \gamma_1, \dots, \gamma_d$ 을 추정할 수 있으나, 동시추정이 더 효율적인 것으로 알려져 있다(Coles and Tawn(1991)).

먼저, 추정에서 쓰일 목적함수(objective function)인 log-likelihood 함수를 정의한다. log-likelihood 함수의 계산법은 다음과 같다:

$$\mathbf{y}_1, \dots, \mathbf{y}_n, \quad \mathbf{y}_i = (y_{i,1}, \dots, y_{i,d})$$

가 n 회의 강우사상 관측에서 얻어진 강수량 자료를 나타낸다고 하자. 여기서, $y_{i,j}$ 는 i 번째 강우사상에서 지점 j 의 강수량을 나타낸다. 주어진 극단값 모형 적용의 기준값 b_1, \dots, b_d 에 대해서

$$k_j = \sum_{i=1}^n I(y_{i,j} > b_j), \quad p_j = \frac{k_j}{n} \quad (12)$$

으로 정의하면, k_j 는 해당지점에서 기준값 b_j 이상의 강수량을 기록한 횟수, p_j 는 상대빈도이다. 우리는 식(10)에서 generalized Pareto 근사법으로 경험분포함수를 개선하였는데, 이 관점에서 다음과 같은 변환을 정의한다:

$$\tilde{F}_j(y) = \tilde{F}_j(y; s_j, u_j) = \begin{cases} 1 - p_j \{1 - \text{GP}(y - b_j; s_j, u_j)\}, & y > b_j; \\ \hat{F}_j(y), & y \leq b_j. \end{cases}$$

이 변환을 통해 얻어진 $y_{i,j}$ 의 변환값을

$$\tilde{z}_{i,j} = \tilde{z}_{i,j}(s_j, u_j) = \frac{1}{1 - \tilde{F}_j(y_{i,j})}, \quad i = 1, \dots, n, \quad j = 1, \dots, d$$

으로 둔다. 다음으로 $(\tilde{z}_{i,1}, \dots, \tilde{z}_{i,d})$ 의 극좌표(polar coordinate)

$$\begin{cases} r_i = r_i(s_1, \dots, s_d, u_1, \dots, u_d) = \tilde{z}_{i,1} + \dots + \tilde{z}_{i,d} \\ \mathbf{w}_i = \mathbf{w}_i(s_1, \dots, s_d, u_1, \dots, u_d) = (\tilde{z}_{i,1}/r_i, \dots, \tilde{z}_{i,d}/r_i), \end{cases} \quad i = 1, \dots, n,$$

를 구한다. 최종적으로 log-likelihood 함수를 다음과 같이 정의한다: spectral 밀도함수 h 에 대한 parametric model의 밀도함수 $h(\cdot; v)$ 와 표준 다변량 극단값 분포함수 $G(\cdot; v)$ 에 대해서,

$$\begin{aligned} & \ell(s_1, \dots, s_d, u_1, \dots, u_d, v) \\ &= \sum_i \left[\log h(\mathbf{w}_i; v) - (d+1) \log r_i + \sum_j \left\{ \log \frac{\partial}{\partial y} \tilde{F}_j(y_{i,j}) - 2 \log(1 - \tilde{F}_j(y_{i,j})) \right\} \right] \quad (13) \\ & \quad + \log G(k_1^{-1}, \dots, k_d^{-1}; v). \end{aligned}$$

여기서, \sum_i 에서 i 는

$$y_{i,1} > b_1 \text{ or } \dots \text{ or } y_{i,d} > b_d \quad (14)$$

이 성립하는 범위에서, 대괄호 []의 내부 \sum_j 에서 j 는 $y_{i,j} > b_j$ 을 만족하는 범위에서 움직이고, $s_1, \dots, s_d, u_1, \dots, u_d$ 는 모든 $j = 1, \dots, d$ 에 대해,

$$\frac{u_j}{s_j} > \frac{1}{\max\{y_{i,j} - b_j : i = 1, \dots, n\}}$$

을 만족하는 범위에서 움직인다.

미지의 모수값 $\sigma_1, \dots, \sigma_d, \gamma_1, \dots, \gamma_d, \alpha$ 에 대한 추정치는 목적함수 ℓ 를 최대로 하는 $s_1, \dots, s_d, u_1, \dots, u_d, v$ 의 값으로 정해지고, $\hat{\sigma}_1, \dots, \hat{\sigma}_d, \hat{\gamma}_1, \dots, \hat{\gamma}_d, \hat{\alpha}$ 으로 표기한다. 최대우도추정의 이론에 따라서, 추정의 불확실성을 나타내는 표준오차와 공분산행렬은 ℓ 의 negative Hessian 행렬

$$\begin{bmatrix} -\frac{\partial^2}{\partial s_1 \partial s_1} \ell(\hat{\sigma}_1, \dots, \hat{\alpha}) \cdots -\frac{\partial^2}{\partial s_1 \partial v} \ell(\hat{\sigma}_1, \dots, \hat{\alpha}) \\ \vdots \quad \quad \quad \ddots \quad \quad \quad \vdots \\ -\frac{\partial^2}{\partial v \partial s_1} \ell(\hat{\sigma}_1, \dots, \hat{\alpha}) \cdots -\frac{\partial^2}{\partial v \partial v} \ell(\hat{\sigma}_1, \dots, \hat{\alpha}) \end{bmatrix}$$

의 역행렬에서 얻어진다.

많은 연구에서 국내 강수량은 Gumbel 분포를 따른다고 결론을 내리고 있다. 이러한 사전정보를 활용하는 방식으로, 해당지점 j 에 대해 $u_j = 0$ 으로 고정해놓고 $\ell(s_1, \dots, v)$ 를 최대화시켜 추정치를 얻는 것을 고려할 수 있다. 이로써, 형태모수가 실제로 0일 때 효율적인(efficient) 모수추정을 할 수 있다. 실제로 Gumbel 분포인지를 확인하기 위해 우도비 검정법을 고려한다. 해당지점에 대해 $u_j = 0$ 으로 고정해놓고 $\ell(s_1, \dots, v)$ 를 최대화시켜 얻어진 추정치를 $\hat{\sigma}_1^*, \dots, \hat{\sigma}_d^*, \hat{\gamma}_1^*, \dots, \hat{\gamma}_d^*, \hat{\alpha}^*$ 으로 나타내고, 해당지점이 수가 q 라고 했을 때,

$$2\{\ell(\hat{\sigma}_1, \dots, \hat{\alpha}) - \ell(\hat{\sigma}_1^*, \dots, \hat{\alpha}^*)\} \leq \chi_{0.05}^2(q)$$

이면 유의수준 5%에서 해당지점에 대해 $\gamma_j = 0$ 이라고 판단한다. 여기서, $\chi_{0.05}^2(q)$ 는 자유도가 q 인 χ^2 분포의 상위 5% 분위수이다.

2.3.3 다변량 극단값 모형의 진단

다변량 극단값 모형의 적합결과에 대한 간단한 진단법(diagnostic procedure)을 설명한다. 진단방법의 설명을 위해서, 모수의 추정치를 $\hat{\sigma}_1, \dots, \hat{\sigma}_d, \hat{\gamma}_1, \dots, \hat{\gamma}_d, \hat{\alpha}$ 으로 나타내고,

$$\tilde{F}_j(y) = \begin{cases} 1 - p_j \{1 - \text{GP}(y - b_j; \hat{\sigma}_j, \hat{\gamma}_j)\}, & y > b_j \\ \hat{F}_j(y), & y \leq b_j \end{cases} \quad \tilde{z}_{i,j} = \frac{1}{1 - \tilde{F}_j(y_{i,j})}, \quad i = 1, \dots, n, \quad j = 1, \dots, d$$

으로 두어

$$\begin{cases} r_i = \widetilde{z}_{i,1} + \cdots + \widetilde{z}_{i,d} \\ \mathbf{w}_i = (\widetilde{z}_{i,1}/r_i, \cdots, \widetilde{z}_{i,d}/r_i), \end{cases} \quad i = 1, \cdots, n,$$

를 정의한다. 단, p_j 의 정의는 (12)에서 주어진다. 일반적으로, 모형진단은 분석의 기본 가정을 확인하는 것으로 수행된다. 우리는 다변량 극단값 이론을 의거하여, 다음의 가정을 확인한다:

(i) r_i 와 \mathbf{w}_i 가 독립인가?, 단, 첨자 i 는

$$y_{i,1} > b_1 \quad \text{or} \quad \cdots \quad \text{or} \quad y_{i,d} > b_d \quad (15)$$

가 만족하는 범위에서 움직인다.

(ii) $r_1/d, \cdots, r_n/d$ 이 표준 Pareto 분포를 따르는가?

(iii) \mathbf{w}_i 가 확률밀도함수 $d^{-1}h(\cdot, \hat{\alpha})$ 를 따르는가? 단, 첨자 i 는 식(15)가 만족하는 범위에서 움직인다.

가정 (i)의 확인은 각 $j=1, \cdots, d$ 에 대해서,

$$\{(w_{i,j}, \log(r_i))\} \quad (16)$$

의 산점도(scatter plot)로 확인한다. 단, 첨자 i 는 식 (15)가 만족하는 범위에서 움직인다. 가정 (i)가 충족된다면, 산점도에서 점들이 $w_{i,j}$ 의 값에 관계없이 r_i 가 일정하게 퍼져 있을 것이다. 그림 9는 산점도의 예인데, $\log(r_i)$ 의 분포가 $w_{i,5}$ 에 따라 다르게 나타나지 않는다. 일반적으로 $d > 2$ 이상인 경우에는 산점도 (16)이 r_i 와 \mathbf{w}_i 사이의 독립성을 완전하게 설명하지는 못한다. 그래서 더 완전한 독립성 확인을 위해서는 면밀한 조사가 필요할 것이다.

가정 (ii)은 표준 Pareto 분포에 대한 $r_1/d, \cdots, r_n/d$ 의 분위수그림을 그려서 확인된다. $r_1/d, \cdots, r_n/d$ 의 분포가 표준 Pareto 분포에 가까울수록 가정 (ii)이 만족되고 있음을 의미한다. 그림 10은 표준 Pareto 분위수 그림의 예인데, 점들이 기울기가 1인 직선에 위

치해 있는 것으로 보아, 가정 (ii)이 충분히 만족되고 있다고 할 수 있다.

가정 (iii)에 대한 일반적인 진단방식은 \mathbf{w}_i 의 분포형태와 추정된 spectral 밀도함수 $d^{-1}h(\cdot, \hat{\alpha})$ 를 비교하는 것이다. 여기서 d^{-1} 은

$$\int_{\mathbf{w} \in \mathbb{S}^{d-1}} h(\mathbf{w}; \alpha) dw_1 \cdots dw_{d-1} = d$$

에 기인한 normalizing constant이다. 그래서 $d^{-1}h(\cdot, \hat{\alpha})$ 는 확률밀도함수이다. 그런데, 차수 d 가 커지면 두 분포의 형태를 직접적으로 비교하기가 어려워져서, 차원축소(reduction of dimension)가 불가피하다고 생각된다. 여기서는 spectral 밀도함수에 대해 logistic 모형을 선택하였을 경우를 중심으로, 차원축소를 통한 진단방식을 설명하고자 한다. Logistic 모형의 다음과 같은 특징에 주목한다: \mathbf{w}_i 가 식 (5)의 형태로 주어진 \mathbb{S}^{d-1} 상의 logistic 분포에 따라 근사적으로 퍼져 있다고 하자. 그러면 $1 \leq j_1 < \cdots < j_m \leq d$, $m \leq d$ 에 대해서,

$$w_{i,j_1}^* = \frac{w_{i,j_1}}{w_{i,j_1} + \cdots + w_{i,j_m}}, \cdots, w_{i,j_m}^* = \frac{w_{i,j_m}}{w_{i,j_1} + \cdots + w_{i,j_m}}$$

으로 정의했을 때,

$$(w_{i,j_1}^*, \cdots, w_{i,j_m}^*), \quad \text{단, } y_{i,j_1} > b_{j_1} \text{ or } \cdots \text{ or } y_{i,j_m} > b_{j_m},$$

은 \mathbb{S}^{m-1} 상의 logistic 분포를 근사적으로 따르고, 분포의 모수 α 도 일치한다. 이 사실을 이용해서 차원축소를 통한 모형진단을 실시할 수 있다. $m=2$ 인 경우 w_{i,j_1} 의 히스토그램과 \mathbb{S}^1 상의 logistic 밀도함수의 그래프를 비교해서 분포의 적합도(goodness of fit)를 확인할 수 있다. 그림 11은 w_{i,j_1} 의 히스토그램과 추정된 spectral 밀도함수의 비교한 그림의 예이다. 두 그림이 일치도가 높을수록 적합도가 높은 것이다.

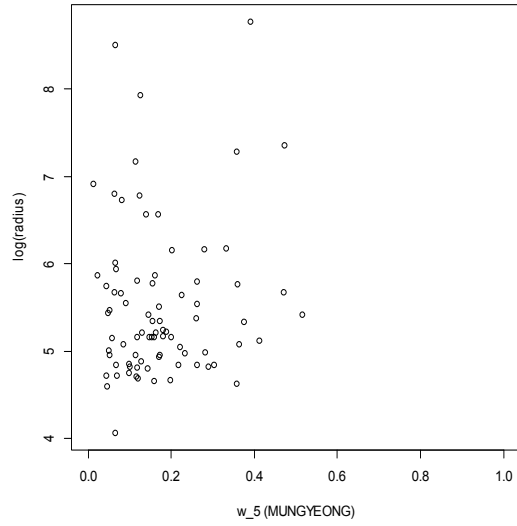


Figure 9. $\log(r_i)$ 대 $w_{i,5}$ 의 산점도: $\log(r_i)$ 의 분포형태가 $w_{i,5}$ 에 무관하게 일정할수록, 모형이 적합하다.

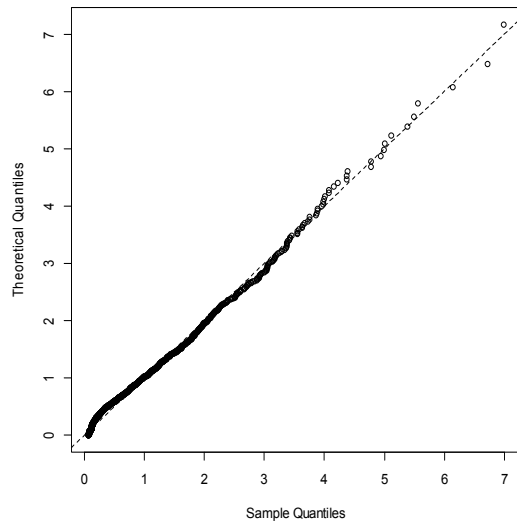


Figure 10. 표준 Pareto 분위수 그림: 점들이 점선에 가까이 분포되어 있을수록, 모형이 적합하다.

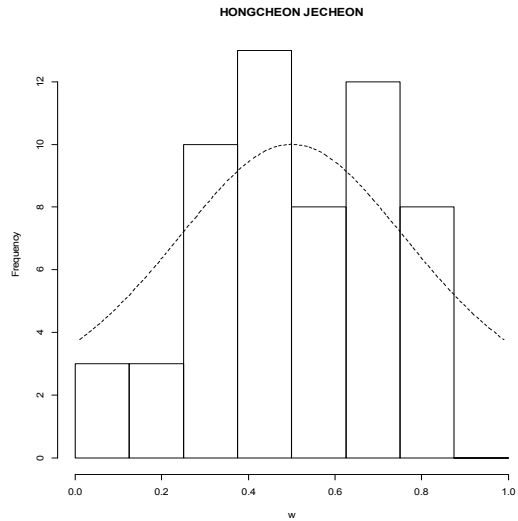


Figure 11. spectral 밀도함수의 적합도 그림: 히스토그램과 밀도함수의 그래프가 일치할수록 모형이 적합하다.

3. 연구내용

3.1 낙동강 상류 유역 호우 강수량 분석

3.1.1 강우사상자료(rainfall event data)의 생성

국내지역 호우기(6월에서 9월까지)동안 강우현상에 대한 통계모형을 세우기 위해, 국내 63개 관측지점에서 1973년부터 2012년까지 기록된 시간별 강수량 자료(hourly precipitation data)를 사용한다.

낙동강 상류 유역의 강우를 모의하기 위해서 안동, 봉화, 태백, 영주, 문경, 의성지점의 강우현상을 분석한다. 모의를 위해서 우리는 이들 지점의 시간별강수량자료를 이용하여 적절히 강우사상(rainfall event)을 정의한다. 본 연구에서의 강우사상은

(i) 적어도 한 지점에서 비가 오는 또는 그에 준하는 연속된 시간대의 균집으로,

(ii) 군집은 서로 확률적으로 독립(stochastically independent)이다.

본 연구에서 사상 간 독립성을 충분히 확보하기 위해 6시간의 간격을 둔다.

강우사상을 얻는 과정을 설명하고자 한다. 각 지점에서 시간대에 강수량이 0.1mm 이상이면, 습윤시간대(wet hour), 그렇지 않으면 건조시간대(dry hour)로 정의한다. 그리고, 적어도 한 지점이 습윤시간대이면 유역이 습윤시간대인 것으로, 반대이면 건조시간대인 것으로 정의한다. 우리는 유역의 습윤시간대가 연속되거나 6시간 이내이면 하나의 강우사상이 지속되는 것으로 간주한다. 표 3에서 2개의 강우사상 또는 습윤기간이 등장한다. 첫 번째 군집은 연속적으로 이어진 습윤시간대로 이루어져 있고, 두 번째는 중간에 건조시간대가 나타나지만, 1시간의 비교적 짧은 시간대이므로 한 사상의 연장으로 본다.

Table 3. 시간별 자료(W: 습윤시간대, D: 건조시간대)

시	안동	봉화	태백	영주	문경	의성	유역	기간
1	D	D	D	D	D	D	D	건조기간
2	W	D	D	W	D	D	W	
3	W	D	D	W	W	D	W	
4	W	D	W	W	W	D	W	
5	W	D	W	W	W	W	W	
6	W	W	W	W	W	W	W	
7	D	W	W	W	W	W	W	
8	D	D	D	D	D	W	W	
9	D	D	D	D	D	D	D	건조기간
10	D	D	D	D	D	D	D	
11	D	D	D	D	D	D	D	
12	D	D	D	D	D	D	D	
13	D	D	D	D	D	D	D	
14	D	D	D	D	D	D	D	
15	D	D	D	D	D	D	D	
16	W	W	W	D	D	W	W	습윤기간 (강우사상)
17	W	W	W	D	W	W	W	
18	D	D	D	D	D	D	D	
19	W	W	W	W	W	W	W	
20	W	W	W	W	W	W	W	
21	W	D	D	D	W	W	W	
22	D	D	D	D	D	D	D	건조기간
23	D	D	D	D	D	D	D	

시간별자료에 대해 강우사상을 생성하면, 표 3에서처럼 습윤기간과 건조기간이 번갈아가며 등장하는 시간흐름이 등장하는데, 습윤기간 동안의 지점별 누적강수량, 습윤기간의 길이(사상의 지속시간), 건조기간의 길이(사상 간 시간간격)을 계산하여 새로운 자료를 구축한다. 이를 '강우사상자료' (rainfall event data)라고 부른다. 표 4는 낙동강 유역에 대한 강우사상자료를 나타낸다. 이 자료는 호우 사상의 강수량을 분석할 때에 뿐만 아니라, 강우 현상 전반을 모의할 때 주요하게 쓰일 것이다.

Table 4. 낙동강 유역의 강우사상자료

No.	월	지속시간 (hr)	시간간격 (hr)	강수량(mm)					
				안동	봉화	태백	영주	문경	의성
1	6	19	100	14	0.5	0.7	0.3	18.9	9.3
2	6	1	94	0	0	0.4	0	0	0
3	6	1	44	0	0.6	0	0	0	0
4	6	2	27	0	0.3	0	0.1	0	0
5	6	6	13	1.4	5.1	0.3	0.3	0.2	4.6
6	6	8	58	0	0	4.9	0	0	0
7	6	7	30	3.7	26	27	6.2	0.9	17.8
8	6	35	61	2.2	0.9	23.5	0.2	0.5	1.7
9	6	5	32	0	0	1.5	0	0	0
10	6	31	61	12.2	4.7	3.3	6.8	20.9	6.2

우리는 생성된 강우사상이 앞서 제시한 그것의 정의에 따라 독립적인지 확인해 볼 필요가 있다. 이를 위해 강우사상자료에서 다지점 강수량과 지속시간의 자기상관함수와 교차상관함수(cross-correlation function)를 그려보았다(그림 12-13). 자기상관과 1 이상 시차(lag)에서 교차상관이 드러나지 않는다. 이러한 사실에 근거하여, 강수량과 지속시간이 복잡한 비선형적 자기상관이 있을 것이라 우려되지는 않기 때문에, 생성된 강우사상들이 독립이라고 가정한다.

강우사상의 독립성은 모의에서 중요한 역할을 한다. 이는 모의에서 매 습윤기간마다 앞선 강우사상과 독립적으로 새로운 사상을 생성할 수 있다는 것을 의미한다. 만약, 앞선 강우사상에 영향을 받는다면, Markov 모형과 같은 복잡한 확률모형을 고려해야 한다. 하지만 사상 간 최소 시간간격을 6시간으로 정함으로써 충분한 독립성이 확보된 것으로 판단되고, 이 덕분에 복잡한 모형을 피할 수 있다.

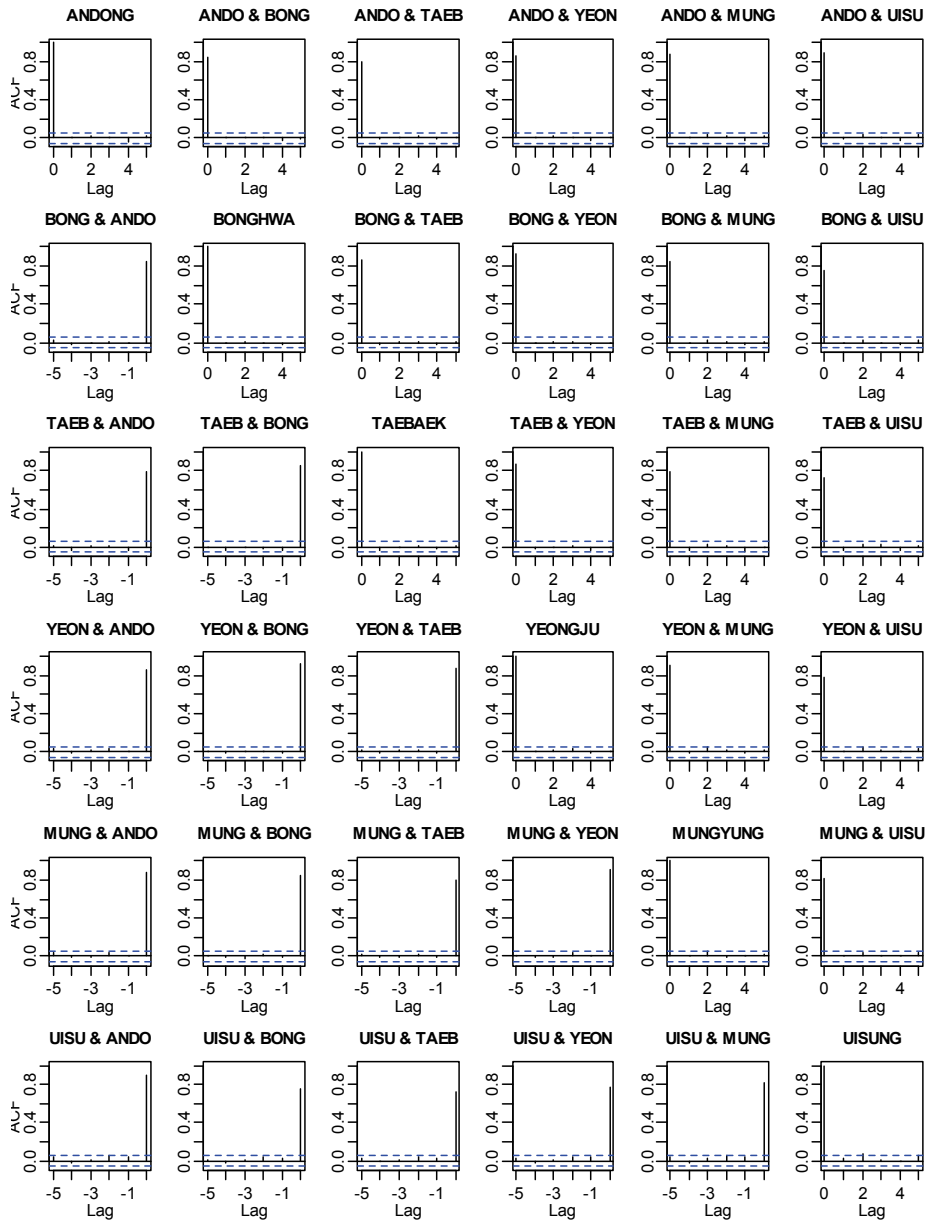


Figure 12. 안동, 봉화, 태백, 영주, 문경, 의성지점 강수량의 자기상관함수와 교차상관함수: 다지점 강수량 시계열은 1이상의 시차에서 유의한 자기상관과 교차상관을 드러내지 않는다.

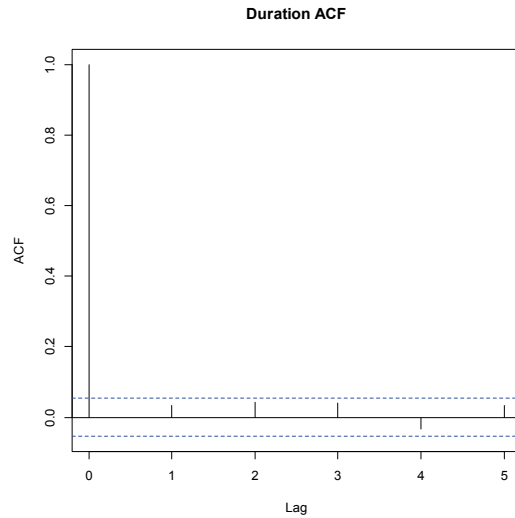


Figure 13. 지속시간의 자기상관함수: 지속시간 시계열은 유의한 자기상관을 드러내지 않는다.

3.1.2 지점별 호우 강수량 분석

낙동강 상류유역의 호우사상의 강수량을 분석하기 위해서는 기준값 b_1, \dots, b_d 을 정해야 한다. 이 값을 정하는데 여러 가지 방법이 있으나, 지점별 호우 강수량 분석을 통해 시작하는 것이 보편적이다(Coles and Tawn(1991)). 하지만 이렇게 정해진 기준값은 다변량 극단값 분석에서는 적절하지 않을 수 있기 때문에, 차후 적합된 다변량 극단값 모형의 진단에 의해 조정될 수도 있다.

지점별로 호우 강수량 분석을 실시한다. 우선, mean residual life plot과 threshold choice plot을 이용하여 적절한 기준값의 범위를 조사한다. 그림 14은 안동지점의 mean residual life plot과 threshold choice plot을 나타낸다. 그림을 통해, 안동지점에 대한 기준값을 50 mm로 선정해 본격적인 분석을 시작한다. 50 mm 기준의 초과 강수량에 generalized Pareto 분포를 최대우도추정법으로 적합시킨다. 그리고 적합을 통

해 얻어진 모형을 진단하여 자료에 적절한지를 판단한다. 만약 적절하지 않다면 기준값을 조정해서, 다시 모형을 적합시켜 본다. 이 과정을 반복해서 적절한 기준값을 찾아간다.

기준값을 50 mm로 선정하여 나온 적합모형을 진단한 결과, 적합모형이 적절한 것으로 판단된다(그림 15). 그런데, 표 5의 적합 결과를 보면, 형태모수 γ 에 대한 가설 ' $\gamma \neq 0$ '가 유의하지 않음을 알 수 있다. 그래서 $\gamma=0$ 이라고 두고 모형적합을 다시 실시하였다. $\gamma=0$ 은 강수량 분석에서 자주 등장하는 Gumbel 모형에 대응한다. 표 6과 그림 16는 각각 Gumbel 모형에 대한 추정결과와 진단그림을 나타낸다. 진단 결과, 분위수그림과 return level plot에서 점들이 $\gamma \neq 0$ 인 generalized Pareto model 적합 결과와 비교하여 봤을 때, 변동이 심해졌다. 하지만 기준선(reference line)의 내부에 머물고 있기에, 우리는 Gumbel 모형을 최종 모형으로 선정한다.

다른 지점에 대해서도 안동지점과 같은 방식으로 호우 강수량을 분석하였다. 표 7에서 적합의 최종결과를 나타낸다. 보고서 지면 활용상, 각 지점의 mean residual life plot과 threshold choice plot, 모형진단의 그림은 부록에서 제공한다.

Table 5. 안동지역의 generalized Pareto model 적합 결과(기준값: 50mm)

모수	추정치	표준오차
σ	36.79	5.058
γ	0.1064	0.1402

Table 6. 안동지역의 Gumbel model 적합 결과(기준값: 50mm)

모수	추정치	표준오차
σ	41.14	3.71
γ	0	-

Table 7. 각 지점의 모형분석 최종결과(괄호 속은 추정의 표준오차)

지점	기준값(mm)	σ	γ
안동	50	41.14(3.71)	0(-)
봉화	80	45.24(8.92)	0.279(0.159)
태백	70	49.07(8.92)	0.357(0.152)
영주	80	48.50(7.98)	0.178(0.126)
문경	70	42.98(6.39)	0.112(0.112)
의성	70	48.25(5.81)	0(-)

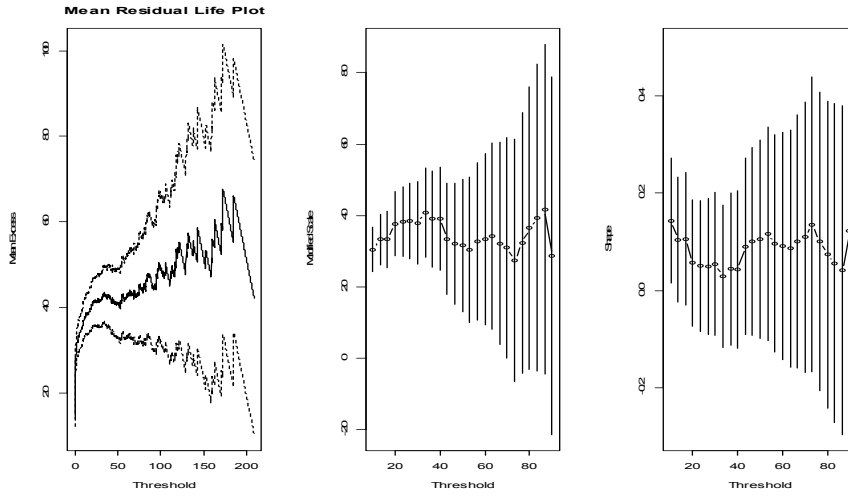


Figure 14. 안동지점의 mean residual life plot과 threshold choice plot

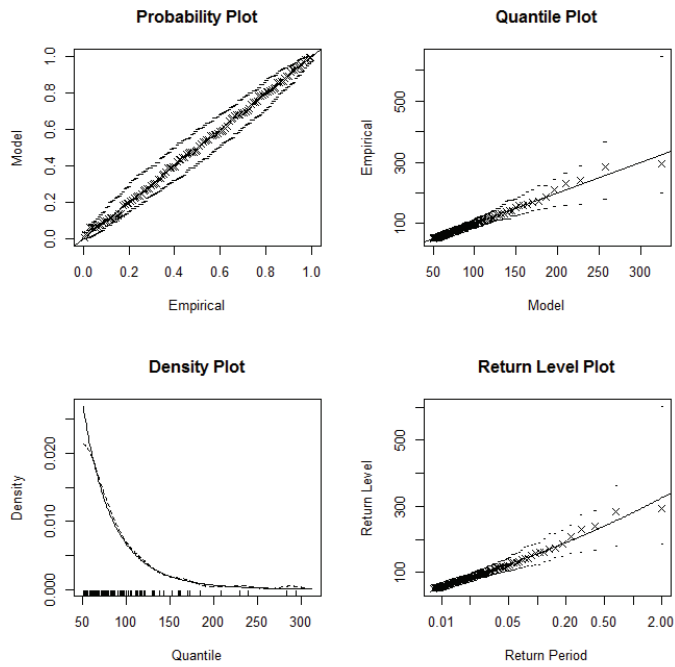


Figure 15. 안동지점의 generalized Pareto 모형에 대한 진단

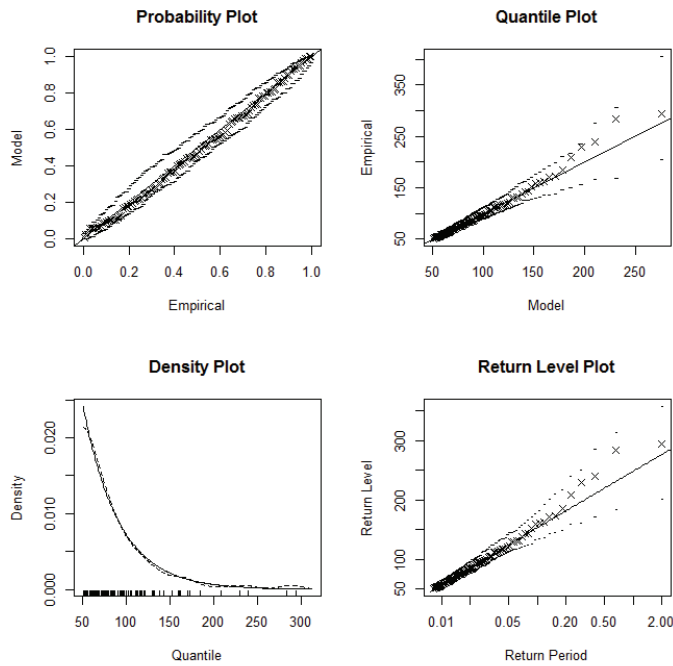


Figure 16. 안동지점의 Gumbel 모형에 대한 진단

3.1.3 다지점 호우 강수량 분석

이제, 다지점 분석을 수행하고자 한다. 앞서, 안동, 봉화, 태백, 영주, 문경, 의성지점의 기준값을 각각 50, 80, 70, 80, 70, 70 mm로 정하였다. 표기의 편의를 위해 표 8에서처럼 안동: 1, 봉화: 2, 태백: 3, 영주: 4, 문경: 5, 의성: 6으로 나타내기로 한다. 다변량 극단값 분포의 spectral 밀도함수에 대한 모형으로 logistic을 선택하고, 최대우도추정법으로 모수를 추정한다. 표 9는 추정결과를 나타낸다.

Table 8. 지점의 번호부여

지점	안동	봉화	태백	영주	문경	의성
번호	1	2	3	4	5	6

Table 9. 추정치와 표준오차(기준값: 50, 80, 70, 80, 70, 70mm)

모수	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	
추정치	40.56	55.23	56.58	60.82	52.80	56.35	
표준오차	4.535	9.031	9.313	8.260	6.329	8.666	
모수	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	α
추정치	0.3016	0.5225	0.6146	0.3241	0.1935	0.1691	0.4250
표준오차	0.1070	0.1766	0.1628	0.1313	0.1117	0.1512	0.0112

이제, 적합모형을 진단한다. 그림 17-18은 진단 그림을 나타낸다. 진단 결과, 모든 $j=1, \dots, 6$ 에 대해 $\log(r_i)$ 과 $w_{i,j}$ 는 독립적으로 보이나, $r_1/d, \dots, r_n/d$ 의 분포가 표준 Pareto 분포와 다른 것으로 판단된다. 특히, 적합모형에서 호우 강수량이 과대추정 (over-estimation) 되는 것으로 보인다. 그래서 기준값을 조정하여 다시 분석을 해야 한다. 여러 번의 기준값 조정 시도를 통해 우리는 안동, 봉화, 태백, 영주, 문경, 의성에 대해 각각 95, 106, 111, 119, 110, 96 mm로 기준값을 선정하였다.

Table 10. 추정치과 표준오차(기준값: 95, 106, 111, 119, 110, 96mm)

모수	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	
추정치	51.07	55.08	99.48	55.52	54.87	45.15	
표준오차	11.62	13.59	23.44	14.18	11.20	10.48	
모수	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	α
추정치	0.1338	0.4394	0.1771	0.3830	0.1259	0.1957	0.5279
표준오차	0.2020	0.2286	0.2128	0.2325	0.1803	0.2009	0.0168

표 10은 조정된 기준값에 대한 적합결과를 나타낸다. 이 적합모형에 대해서 모형 진단을 실시하였다. 그림 19-20은 진단그림을 나타낸다. 진단결과, 그림 18에서 보였던 호우 강수량의 과대추정경향은 더 이상 나타나지 않았다. 그런데, 표 10의 추정결과를 보면 봉화를 제외한 모든 지점에서 $\gamma=0$ 일 것이라 생각된다. 우도비 검정을 통해 이를 확인해 본 결과, 적절한 추측인 것으로 판단되어(표 11), $\gamma_1 = \gamma_3 = \dots = \gamma_6 = 0$ 인 하위모형을 다시 적합시켰다. 표 12은 추정결과를 나타낸다. 그림 20-21은 추정된 모형에 대한 진단그림을 나타낸다. 진단 결과, 적합모형은 logistic 모형의 단순성을 감안했을 때 적정 수준에서 낙동강 상류유역의 호우 강수량을 설명한다고 보여 진다.

Table 11. 우도비 검정(대립가설: 포화모형)

귀무가설	χ^2 -검정통계량	자유도	유의확률
$\gamma_1 = \gamma_3 = \dots = \gamma_6 = 0$	6.423	5	26.7%

Table 12. 추정치와 표준오차(기준값: 95, 106, 111, 119, 110, 96 mm, $\gamma_1 = \gamma_3 = \dots = \gamma_6 = 0$)

모수	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	
추정치	58.90	53.69	121.40	83.05	61.99	55.97	
표준오차	6.681	13.12	13.53	9.595	6.855	6.508	
모수	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	α
추정치	0	0.4664	0	0	0	0	0.5256
표준오차	-	0.2286	-	-	-	-	0.0168

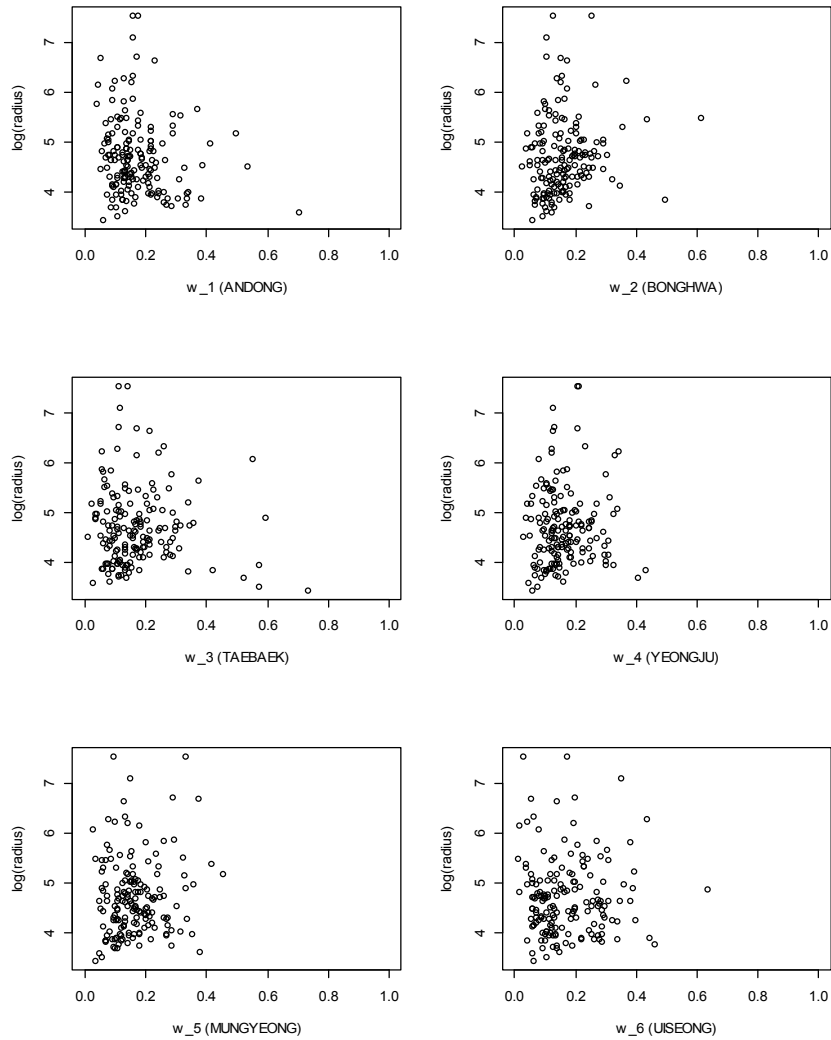


Figure 17. 진단그림: $\log(r_i)$ 대 $w_{i,j}$ (기준값: 50, 80, 70, 80, 70, 70mm).
 $\log(r_i)$ 의 분포에 특이한 경향성은 없는 것으로 판단된다.

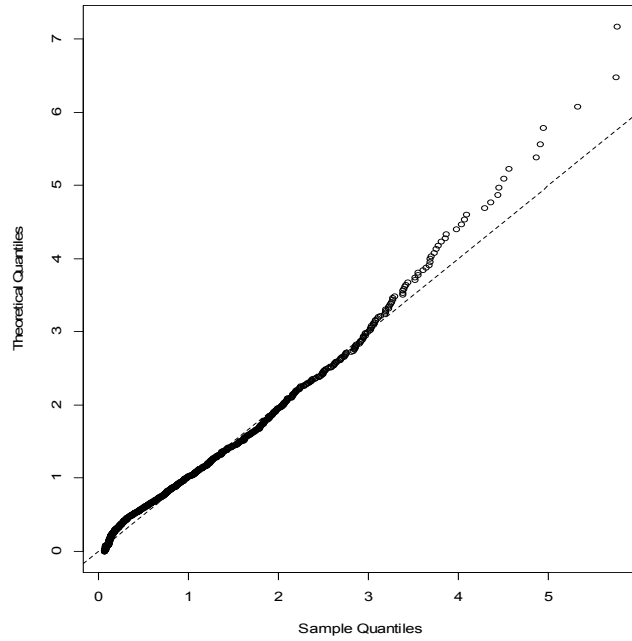


Figure 18. 진단그림: Pareto 분위수 그림(기준값: 50, 80, 70, 80, 70, 70mm).
 r_1, \dots, r_n 의 분포가 Pareto 분포와 어긋나는 것으로 보인다.

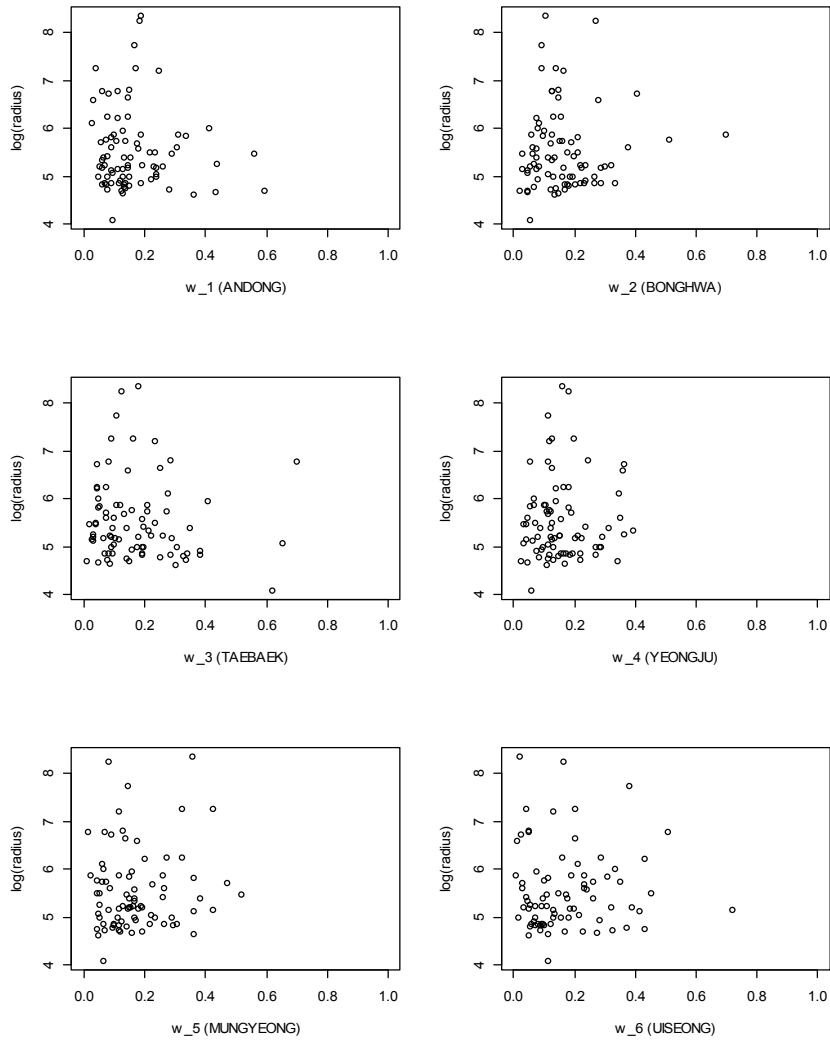


Figure 19. 진단그림: $\log(r_i)$ 대 $w_{i,j}$ (기준값: 95, 106, 111, 119, 110, 96mm).
 $\log(r_i)$ 의 분포에 특이한 경향성은 없는 것으로 판단된다.

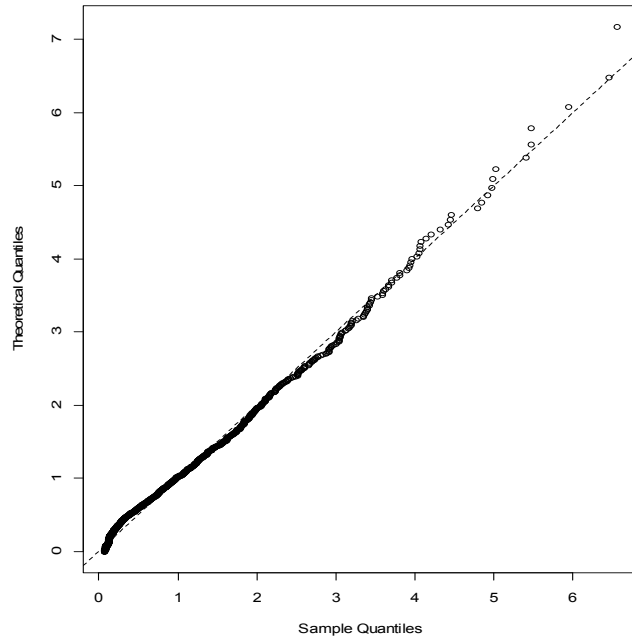


Figure 20. 진단그림: Pareto 분위수 그림 (기준값: 95, 106, 111, 119, 110, 96 mm).
 r_1, \dots, r_n 가 Pareto 분포를 따르는 것으로 보인다.

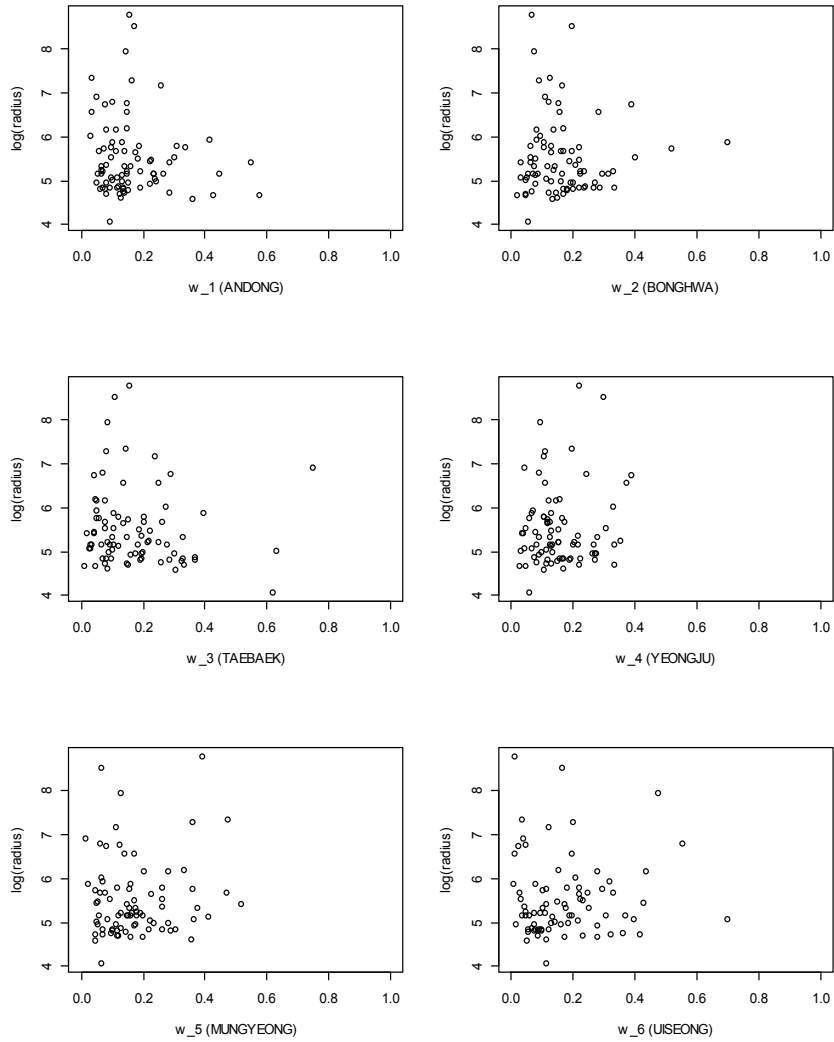


Figure 21. 진단그림: $\log(r_i)$ 대 $w_{i,j}$ (기준값: 95, 106, 111, 119, 110, 96 mm, $\gamma_1 = \gamma_3 = \dots = \gamma_6 = 0$). $\log(r_i)$ 의 분포에 특이한 경향성은 없는 것으로 판단된다.

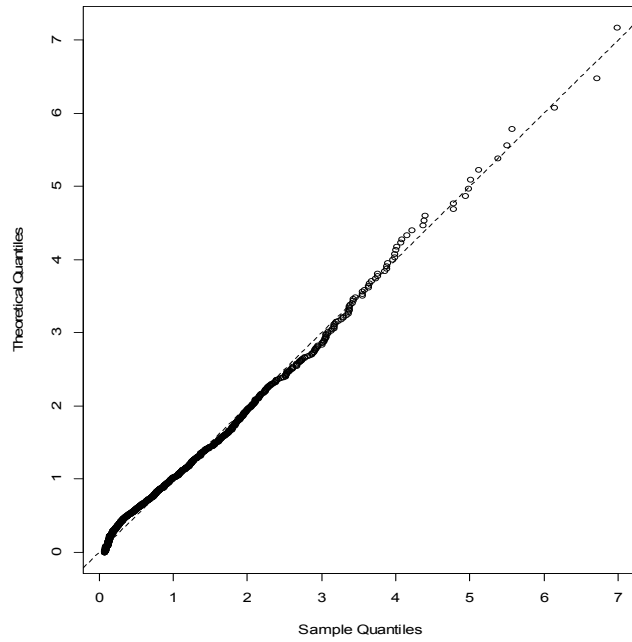


Figure 22. 진단그림: Pareto 분위수 그림 (기준값: 95, 106, 111, 119, 110, 96 mm, $\gamma_1 = \gamma_3 = \dots = \gamma_6 = 0$). $\log(r_i)$ 의 분포에 특이한 경향성은 없는 것으로 판단된다.

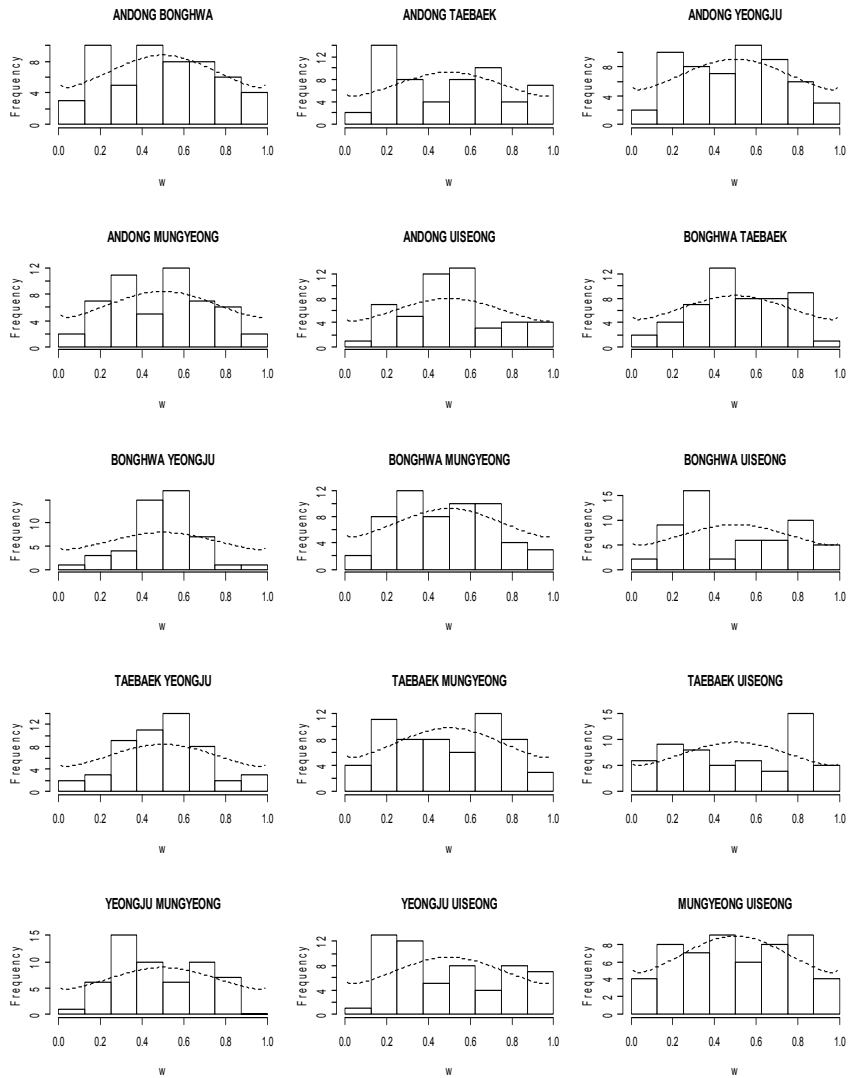


Figure 23. spectral 밀도함수 적합도 그림. Logistic 모형이 모든 지점 간에
연관성을 완전하게는 표현하지는 못하는 것으로 보인다.

3.1.4 강수지속시간 및 사상간 시간간격 분석

강우모의를 위해 강수량 모의와 더불어 강수지속시간과 습윤일 과정(wet day process)에 대한 통계모형이 필요하다. 이를 위해 본 소절에서는 호우사상의 지속시간과 강우사상간 시간간격을 분석하고자 한다.

우선, 호우사상의 지속시간을 분석한다. 그림 24는 지점별로 호우사상의 강수량과 지속시간의 산점도를 나타낸다. 그림의 관찰을 통해 강수량과 지속시간 사이에 양의 상관관계가 있는 것으로 보이고, 이를 근거로 다중회귀모형(multiple regression model)

$$\tau_i = \beta_0 + \beta_1 y_{i,1} + \dots + \beta_6 y_{i,6} + \epsilon_i$$

을 적합시켜 본다. 여기서, τ_i 은 지속시간, $y_{i,1}, \dots, y_{i,6}$ 은 지점별 강수량, ϵ_i 는 오차항이며, 첨자 i 는 적어도 한 지점에서 강수량이 기준값을 넘는 강우사상, 즉,

$$y_{i,1} > 95 \text{ or } y_{i,2} > 106 \text{ or } y_{i,3} > 111 \text{ or } y_{i,4} > 119 \text{ or } y_{i,5} > 110 \text{ or } y_{i,6} > 96$$

이 만족하는 범위에서 움직인다. 최소제곱추정법(least square estimation)으로 β_0, \dots, β_6 와 오차항의 분산을 추정하였다. 표 13은 적합결과를 나타낸다. 이어서, 적합 결과를 회귀 분석의 통상적인 방법으로 진단해 보았다. 그림 25은 진단그림을 나타낸다. 진단결과, 오차항의 정규성(normality) 가정이 약간 어긋나는 것으로 판단된다.

지속시간에 대한 다른 모형으로 일반화 선형모형(generalized linear model)을 고려하고자 한다. 지속시간의 분포로는 Gamma를 택하고, 연결함수(link function)은 Gamma 분포족의 정준연결함수(canonical link function)인 $g(\mu) = \mu^{-1}$ 으로 둔다. 즉, τ_i 는 형태 모수가 $\nu > 0$ 이고 척도모수가 μ_i/ν 인 Gamma 분포를 따르며,

$$\frac{1}{\mu_i} = \beta_0 + \beta_1 y_{i,1} + \dots + \beta_6 y_{i,6}$$

으로 가정한다. 일반화 선형 모형의 모수인 ν 와 β_0, \dots, β_6 를 IRLS(iterative reweighted

least square) 방법으로 추정하였다. 표 14는 일반화 선형모형의 적합 결과이다. 이 적합 모형의 대한 진단은 그림 26에서 주어진다. 진단 결과, 적합모형은 적절한 것으로 판단되어, 우리는 최종모형으로 선정한다.

Table 13. 회귀모형의 적합 결과

모수	β_0	β_1	β_2	β_3	β_4	β_5	β_6	오차항의 표준편차
추정치	23.16	0.0112	-0.0343	0.1065	0.0543	0.0263	0.1224	20.91
표준 오차	5.531	0.077	0.060	0.032	0.736	0.0605	0.0574	

Table 14. 일반화 선형 모형의 적합 결과

모수	추정치	표준오차
β_0	2.594×10^{-2}	1.798×10^{-3}
β_1	-1.619×10^{-6}	2.556×10^{-5}
β_2	1.482×10^{-5}	1.896×10^{-5}
β_3	-2.304×10^{-5}	9.720×10^{-6}
β_4	-9.013×10^{-6}	2.646×10^{-5}
β_5	-1.024×10^{-5}	2.070×10^{-5}
β_6	-3.119×10^{-5}	1.431×10^{-5}
ν	6.265	

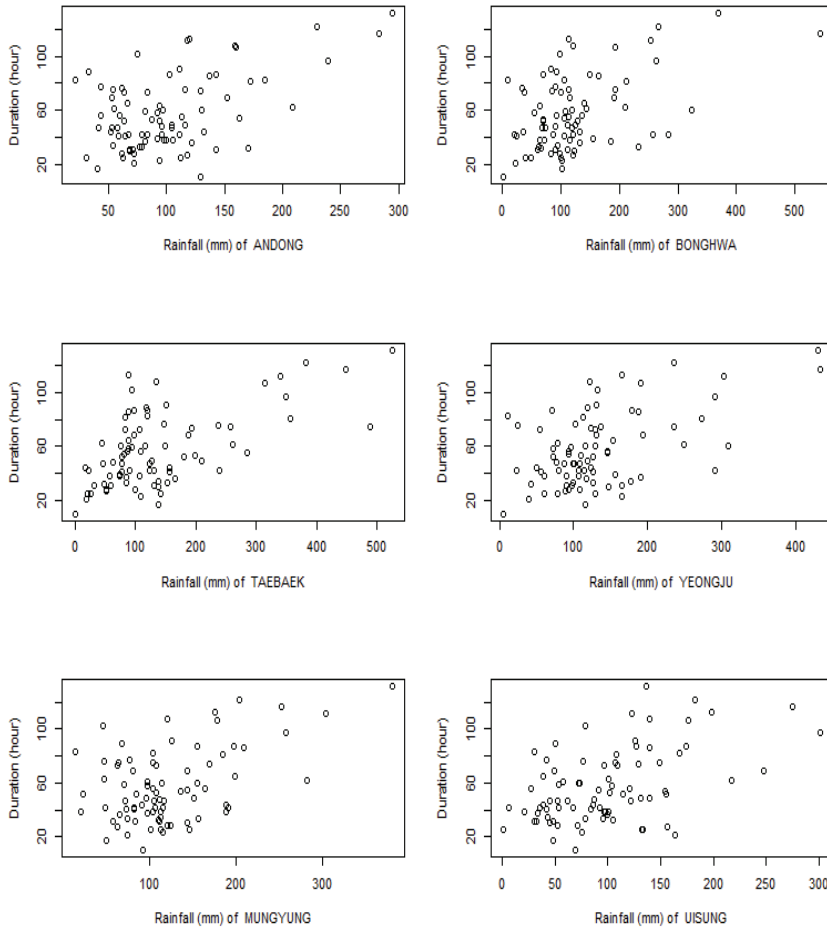


Figure 24. 호우사상의 지점별 강수량과 지속시간의 산점도. 양의 상관관계가 있는 것으로 판단된다.

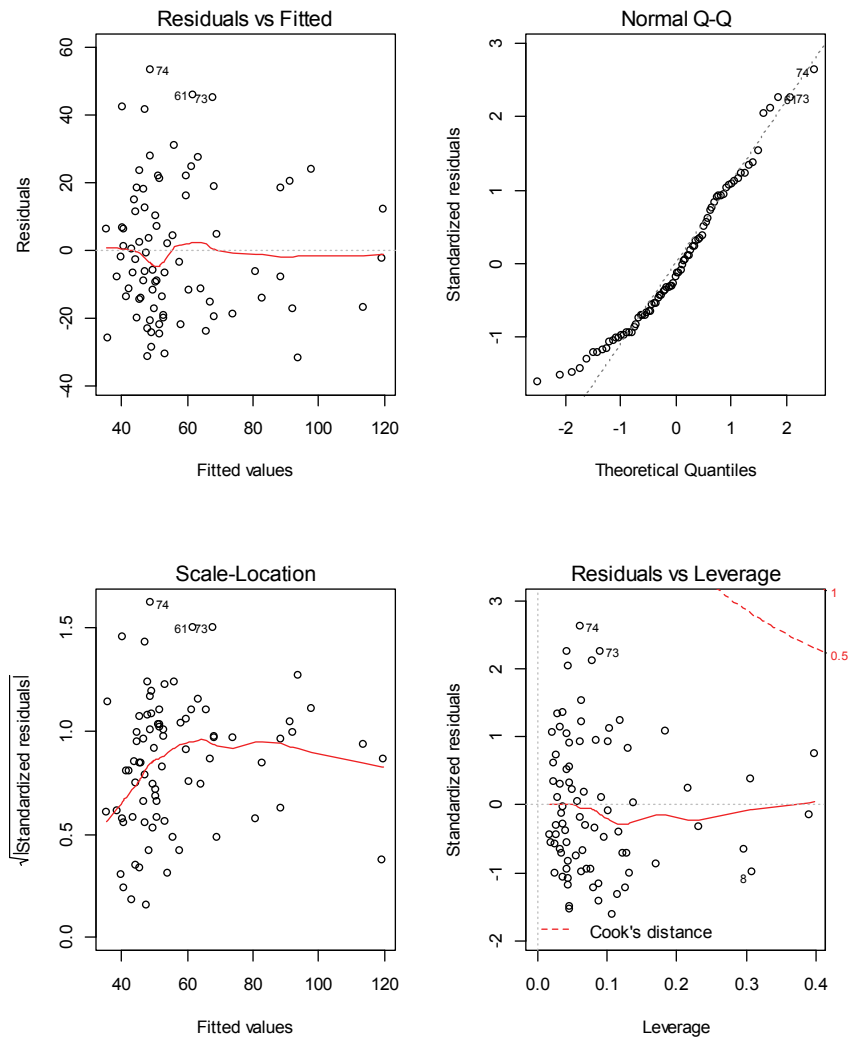


Figure 25. 회귀모형 진단. 오차항의 정규성 가정이 어긋나는 것으로 보인다.

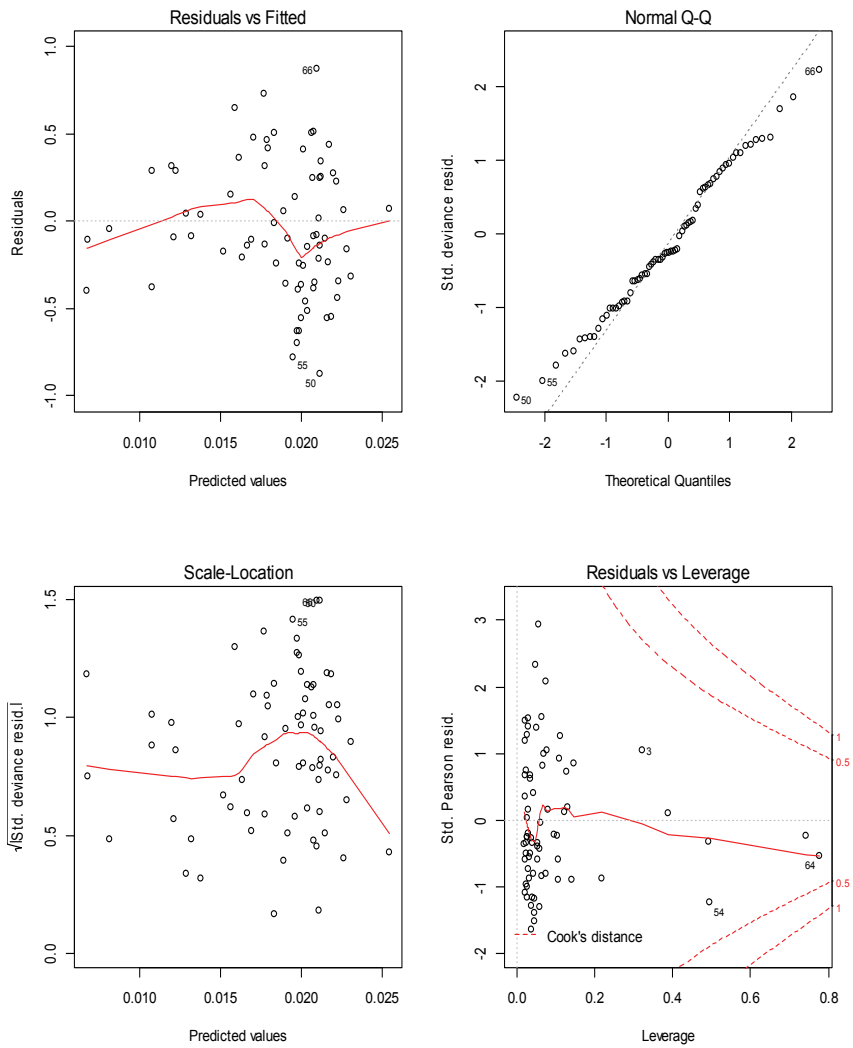


Figure 26. 일반화 선형 모형의 진단그림

이제, 사상 간 시간간격을 분석한다. 강우사상자료를 생성에서 사상 간의 시간간격을 계산할 때, 원자료가 시간별자료이므로 시간간격도 시(時)단위로 계산된다. 또한, 월별 강우현상의 차이를 감안하여 기존의 연구에처럼 시간간격을 월 단위로 나누어 분석한다 (Semenov et al.(1998)). 우선, 강우사상 간 시간간격의 독립성을 확인해 보기 위해 자기상관함수를 그려보았다. 그림 27은 자기상관함수를 나타내는데, 자기상관계수가 모두 유의하지 않게 나온다. 즉, 시간간격은 자기상관이 없는 것으로 보여서, 시간간격은 독립적이라고 가정한다. 시간간격의 확률분포를 추정하기 위해 Gamma 분포를 적합시켜 보았다. 표 15는 적합결과를 나타낸다. 적합결과에 대한 진단으로 분위수 그림을 그려 보았다. 그림 28은 분위수 그림을 나타내는데, 6-8월에서는 Gamma 분포가 적합하지 않다. 그래서 시간간격에 대한 모의에서는 임의추출(random sampling) 기법을 쓰는 것이 적절해 보인다.

장시간의 시간간격을 모의하기 위해서 극단값 모형을 고려해 볼 수 있다. 그래서 시간간격에 대한 극단값 분석을 해보고자 한다. 표 16은 최종 선택된 시간간격의 기준값과 generalized Pareto 분포의 적합결과를 나타낸다. 모수추정에서 형태모수는 모든 월에서 $\gamma=0$ 으로 선택되었다. 기준값 선택과 진단을 위한 그림은 부록에서 제공한다.

Table 15. Gamma 분포 적합 결과(괄호 속은 표준오차)

월	형태모수	척도모수
6월	1.203(0.090)	43.13(3.98)
7월	1.434(0.093)	20.86(1.61)
8월	1.174(0.077)	29.71(2.42)
9월	1.011(0.077)	47.47(4.66)

Table 16. generalized Pareto 분포 적합 결과(괄호 속은 표준오차)

월	기준값(시간)	형태모수	척도모수
6월	80	0(-)	61.64(8.467)
7월	70	0(-)	47.10(8.747)
8월	30	0(-)	47.03(4.241)
9월	80	0(-)	47.02(6.121)

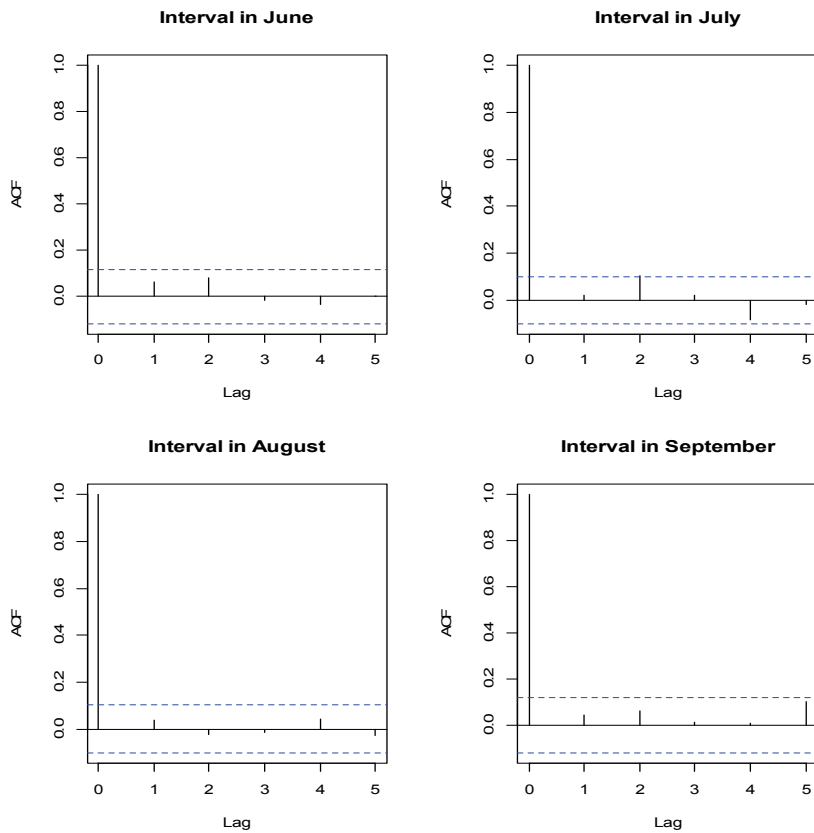


Figure 27. 강우사상 간 시간간격의 자기상관함수

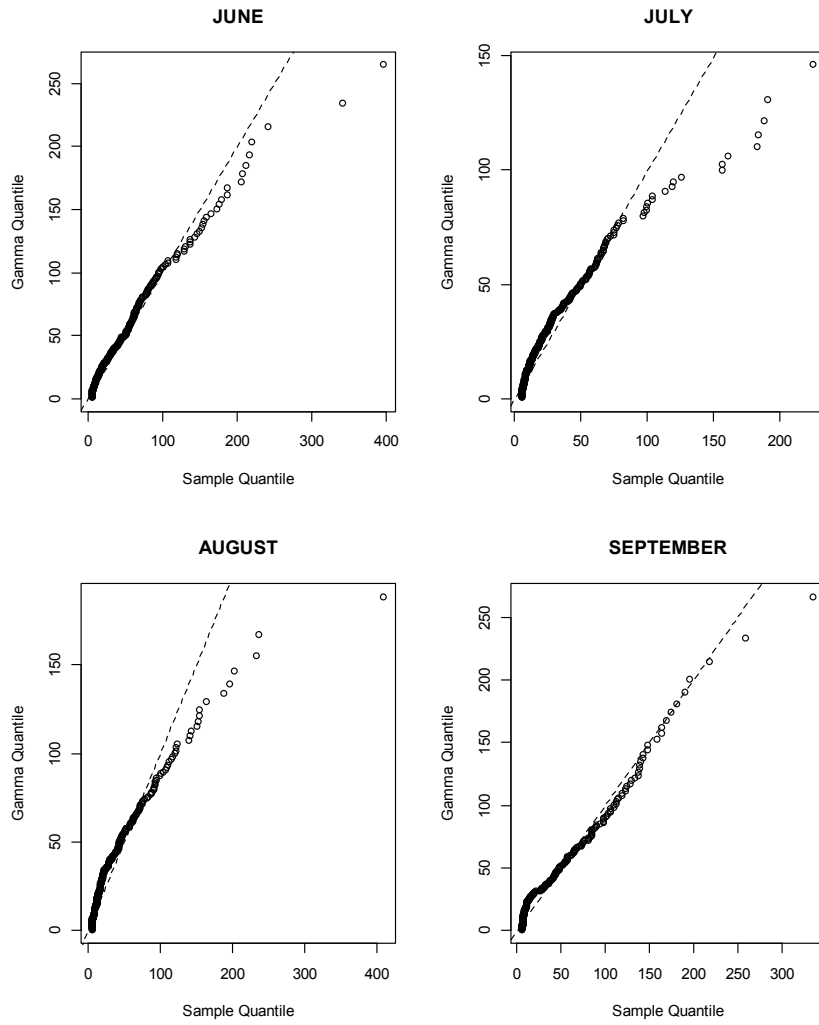


Figure 28. 시간간격의 Gamma 분위수 그림

3.2 한강 상류 유역 호우 강수량 분석

3.2.1 지점별 호우 강수량 분석

한강 상류 유역에 위치한 철원, 춘천, 원주, 영월, 홍천, 제천지점의 강우현상을 분석하고자 한다. 우선, 낙동강 상류유역에서와 동일한 방법으로 강우사상자료를 생성한다. 생성된 강우사상자료에 대해서 다지점 강수량과 지속시간의 자기상관을 조사해 본 결과, 독립으로 가정해도 무방한 것으로 판단된다.

생성된 강우사상자료를 이용해 지점별로 호우 강수량을 분석한다. 분석방법은 낙동강 유역과 동일하다. 표 17은 한강유역 지점의 generalized Pareto model 적합결과인데, 철원은 제외한 다른 모든 지점에서 Gumbel 모형이 선택되었다. 기준값 선택과 모형진단의 그림은 부록에서 제공된다.

Table 17. 각 지점의 모형분석 최종결과(괄호 속은 추정의 표준오차)

지점	기준값(mm)	σ	γ
철원	50	52.79(7.77)	0.297(0.118)
춘천	100	113.2(16.51)	0(-)
원주	80	81.38(9.94)	0(-)
영월	80	63.59(8.35)	0(-)
홍천	80	85.23(9.77)	0(-)
제천	80	71.68(8.50)	0(-)

3.2.2 다지점 호우 강수량 분석

이제 한강 상류유역의 호우 강수량에 대한 다지점 분석을 실시한다. 분석방법은 역시 낙동강유역에서와 동일하다. 지점별 분석을 통해 각 지점의 기준값을 선택하였지만, 이를 써서 얻은 적합모형은 적절하지 않은 것으로 판단되었다. 그래서 기준값을 조정하였다. 각 지점의 번호와 더불어 기준값 선택결과는 표 18과 같다.

Table 18. 한강유역 지점의 기준값 및 번호부여

지점	철원	춘천	원주	영월	홍천	제천
번호	1	2	3	4	5	6
기준값(mm)	141	176	165	124	180	143

지점별 분석을 통해서, 철원을 제외한 다른 모든 지점에서 호우 강수량에 대해 Gumbel 모형이 선택되었다. 이 선택결과를 우도비 검정을 통해 재확인하였는데, 검정결과는 표 19와 같다. 우도비 검정결과 유의수준 5%에서 가설 $\gamma_2 = \dots = \gamma_6 = 0$ 이 기각되지 않아, 다지점 분석에서도 이 선택을 유지하였다. 그래서 지점별 강수량 모형은 지점별 분석결과를 따르고, spectral 밀도함수에 대한 모형으로 logistic을 선택하여, 적합을 실시하였다. 표 20은 그 적합결과를 나타낸다. 그림 29-31은 적합모형의 진단그림을 나타낸다. 진단결과, 적합모형은 적정 수준에서 한강 상류유역의 호우 강수량을 설명한다고 보여 진다.

Table 19. 우도비 검정(대립가설: 포화모형)

귀무가설	χ^2 -검정통계량	자유도	유의확률
$\gamma_2 = \dots = \gamma_6 = 0$	9.337	5	9.6%

Table 20. 추정치와 표준오차(기준값: 141, 176, 165, 124, 180, 143 mm)

모수	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	
추정치	62.66	114.52	94.45	96.50	98.12	107.70	
표준오차	20.01	16.96	13.87	14.85	14.59	15.96	
모수	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	α
추정치	0.796	0	0	0	0	0	0.502
표준오차	0.375	-	-	-	-	-	0.020

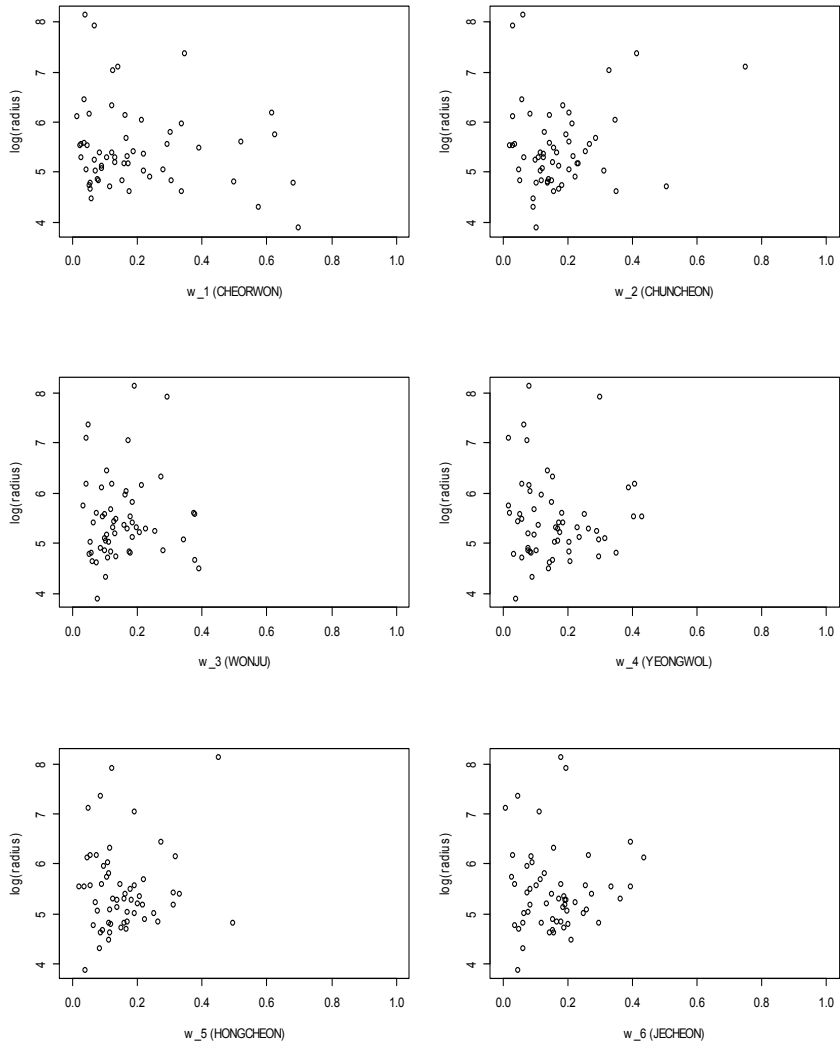


Figure 29. 진단그림: $\log(r_i)$ 대 $w_{i,j}$. $\log(r_i)$ 의 분포에 특이한 경향성은 없는 것으로 보인다.

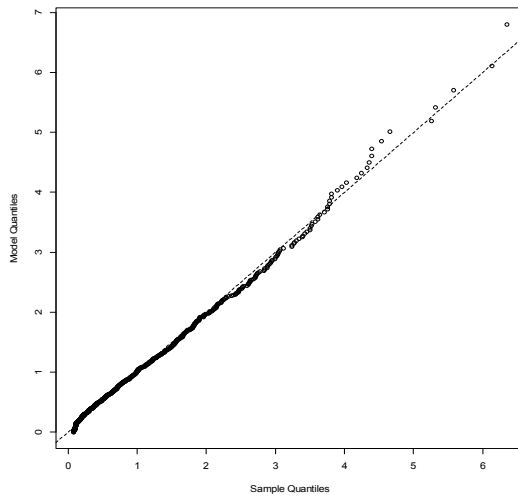


Figure 30. 진단그림: Pareto 분위수 그림. r_1, \dots, r_n 은 Pareto 분포를 따르는 것으로 보인다.

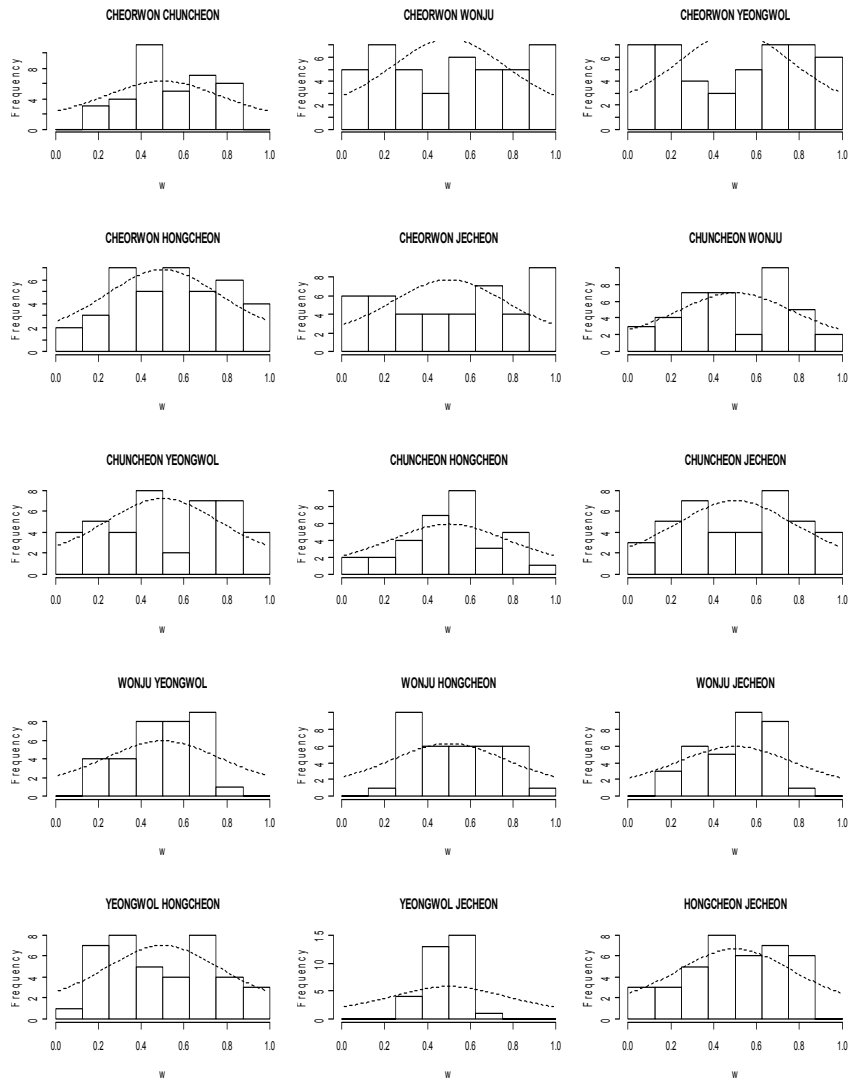


Figure 31. spectral 밀도함수의 적합도 그림. Logistic 모형이 모든 지점 간에 연관성을 완전하게는 표현하는데 한계가 있어 보인다.

3.2.3 강수지속시간 및 사상간 시간간격 분석

이제 한강 상류유역에서 호우사상의 지속시간과 사상간 시간간격을 분석하고자 한다. 우선 호우사상의 지속시간을 분석하는데, 그림 32에서 나타나는 것처럼 강수량과 지속시간이 양의 상관관계가 있으므로, 낙동강 상류유역의 경우에서와 동일한 일반화 선형모형을 고려한다. 표 21은 적합결과를 나타낸다. 그림 34은 적합모형의 진단그림이다. 진단결과, 적합모형은 호우사상의 지속시간을 설명하는데 큰 무리는 없는 것으로 보여진다.

다음으로 사상간 시간간격을 분석한다. 우선, 자기상관함수를 통해 시간간격의 독립성을 확인해 보았다. 그림 34은 자기상관함수를 나타내는데, 유의하지 않거나 무시할 수 있는 크기의 자기상관을 드러내므로, 우리는 시간간격이 독립적이라고 가정한다. 또한 시간간격에 Gamma 분포를 적합시켜 보았다. 하지만, 낙동강 상류유역의 경우에서처럼 적합도가 좋지 않았다. 그래서 시간간격 모의에 임의추출기법을 쓰는 것이 합리적인 것으로 판단된다. 또한 장시간의 시간간격 모의를 위해 극단값 분석을 고려한다. 표 22는 월별로 generalized Pareto 분포의 적합결과이다. 특이하게 6월에서 형태모수가 음으로 나왔다. 기준값 선택과 진단을 위한 그림은 부록에서 주어진다.

Table 21. 일반화 선형 모형의 적합 결과

모수	추정치	표준오차
β_0	2.230×10^{-2}	1.988×10^{-3}
β_1	-1.857×10^{-6}	8.754×10^{-6}
β_2	-8.200×10^{-6}	1.381×10^{-5}
β_3	2.058×10^{-6}	2.070×10^{-5}
β_4	-3.428×10^{-5}	3.216×10^{-5}
β_5	-2.166×10^{-5}	1.760×10^{-5}
β_6	2.274×10^{-5}	3.206×10^{-5}
ν	0.209	

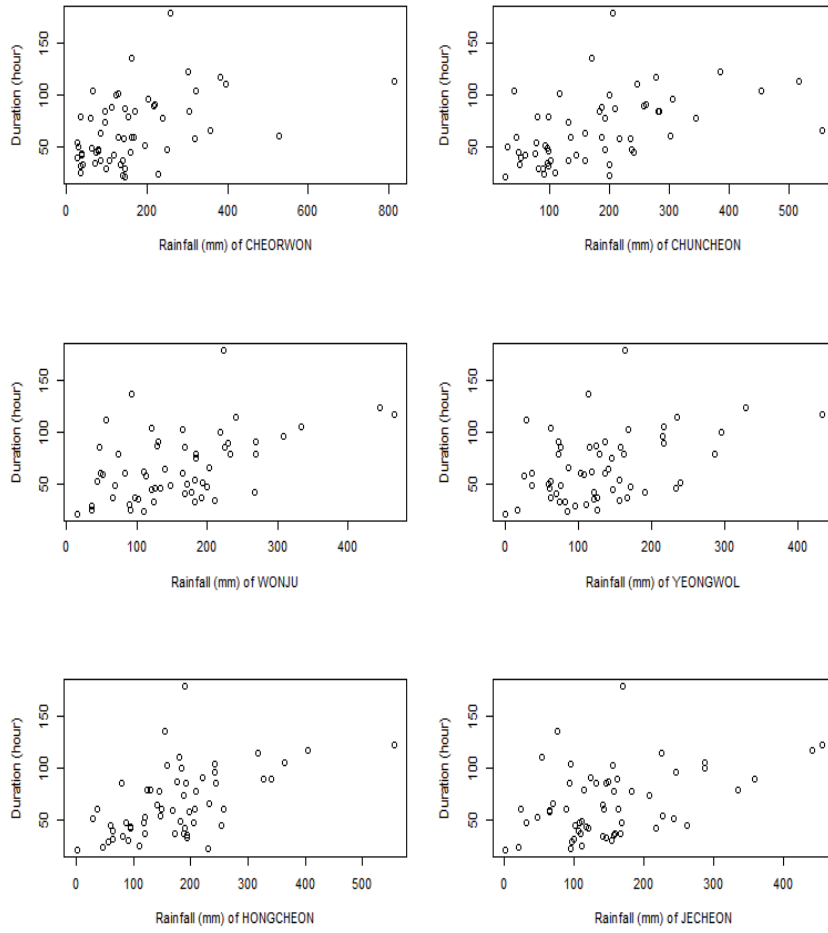


Figure 32. 호우사상의 지점별 강수량과 지속시간의 산점도.

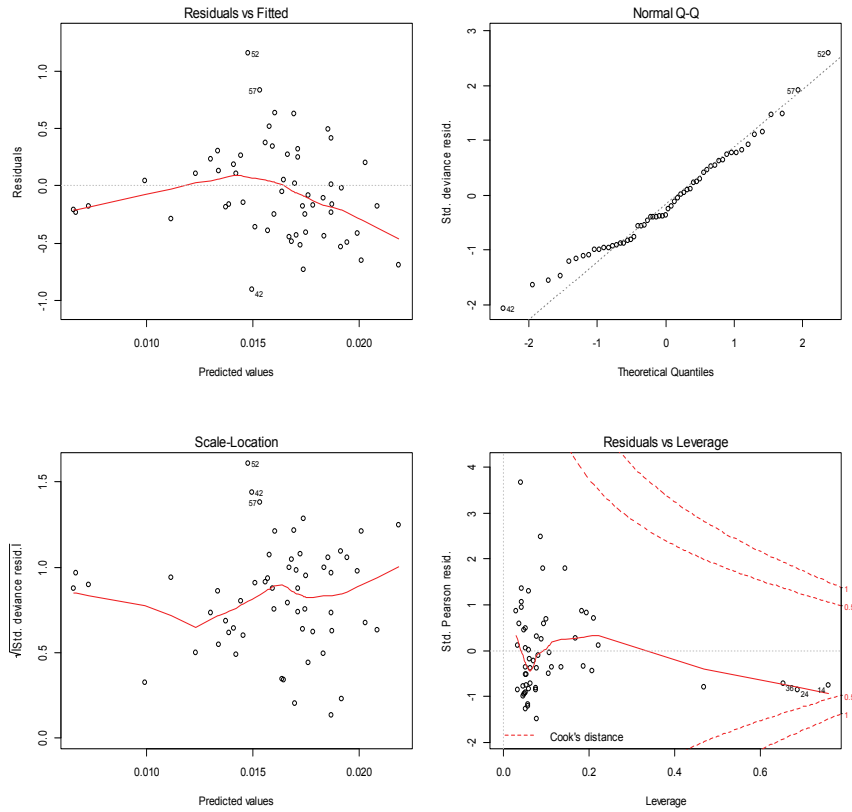


Figure 33. 일반화 선형모형의 진단그림

Table 22. generalized Pareto 분포 적합 결과(괄호 속은 표준오차)

월	기준값(시간)	형태모수	척도모수
6월	80	-0.3933(0.1345)	86.52(17.09)
7월	70	0(-)	36.54(7.459)
8월	90	0(-)	52.85(14.66)
9월	120	0(-)	40.86(8.712)

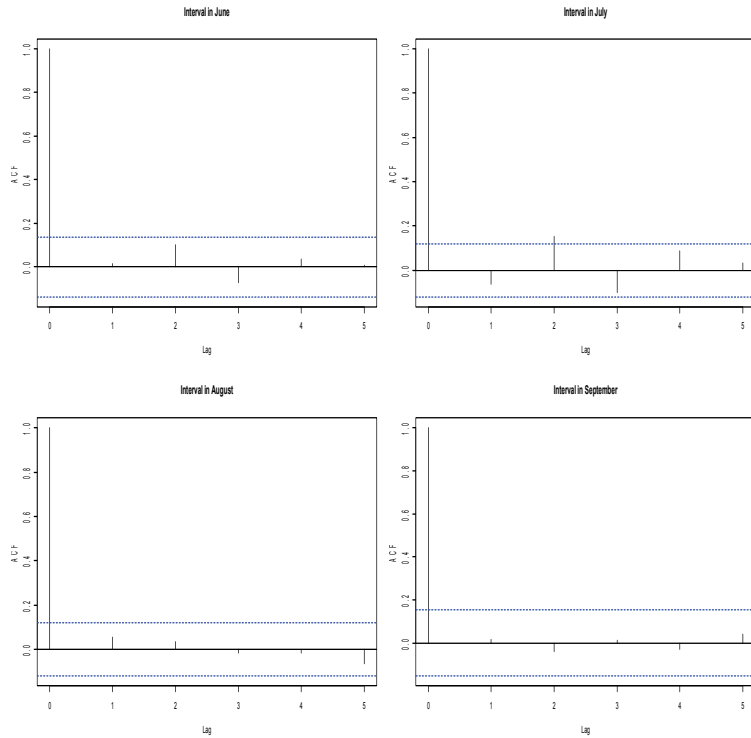


Figure 34. 강우사상 간 시간간격의 자기상관함수

3.3 호우기 강우 모의

3.3.1 강의 모의 방법

앞서 낙동강과 한강 상류유역에 대해 강우사상자료를 생성하였고, 이 자료를 이용해 호우사상의 강수량과 지속시간, 그리고 강우사상 간 시간간격을 분석하였다. 본 절에서는 이 분석결과를 바탕으로 우리나라 호우기인 6월에서 9월 동안에 해당지역의 강우를 모의해 보기로 한다. 낙동강과 한강 유역에 대한 통계모형을 정리하면 표 23과 같다.

Table 23. 낙동강과 한강 상류 유역의 모형 선택과 적합 결과

항목	모형선택	모형적합결과
호우사상의 강수량	Generalized Pareto 모형, Logistic 모형	표 12(낙동강 상류유역) 표 20(한강 상류유역)
호우사상의 지속시간	Gamma 일반화 선형모형	표 14(낙동강 상류유역) 표 21(한강 상류유역)
강우사상간 시간간격	경험분포(empirical model) generalized Pareto 모형	표 16(낙동강 상류유역) 표 22(한강 상류유역)

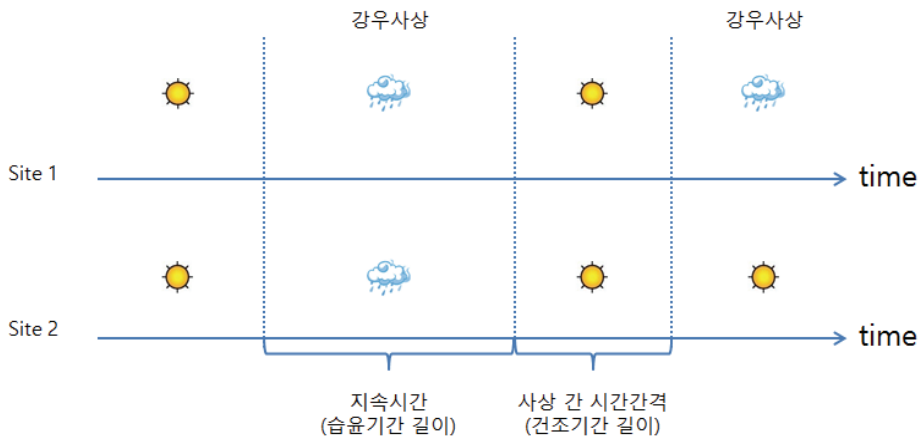


Figure 35. 강우모의 개념도

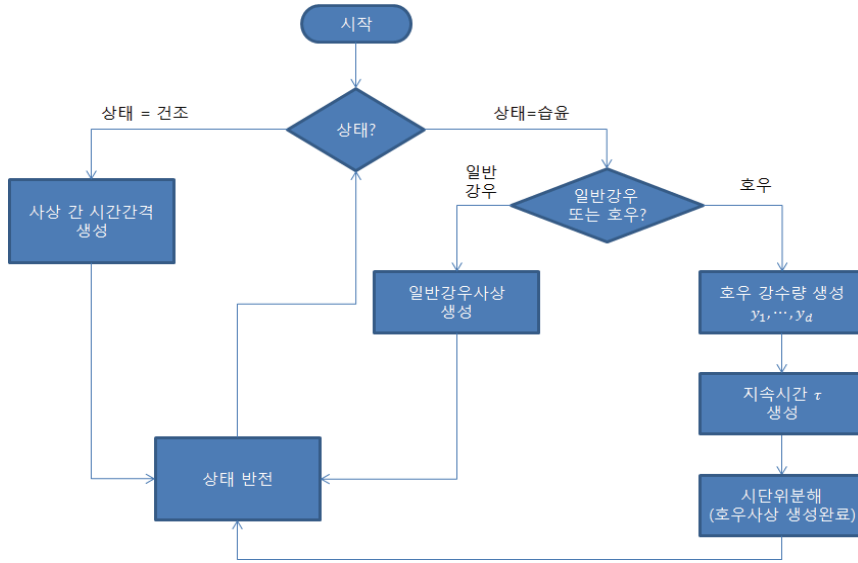


Figure 36. 강우모의의 순서도

이제, 표 23에서 주어진 분석결과를 이용한 모의방법을 제안한다. 모의 순서도는 그림 36와 같다. 시간별 자료를 쓰는 관계로 자료의 활용도를 높이기 위해 모의시간단위를 시(時, hour)로 한다. 습윤일 모의 방법으로 LARS-WG와 같이 건조기간과 습윤기간을 번갈아서 생성시키는 alternating method를 선택한다. 단, 여기서 습윤기간이란 모의 대상지점 중 적어도 하나의 지점에서 비가 내리는 기간으로, 건조기간은 그 반대의 경우로 정의한다(그래서, 그림 35에서처럼 습윤기간 동안에도 비가 내리지 않는 지점이 있을 수 있다). 그런데, LARS-WG에서는 모의기간동안 습윤일을 모두 생성해 놓고 각 습윤일마다 강수량을 독립적으로 모의하는 반면에 제안된 방법에서는 습윤기간과 건조기간이 번갈아가며 순차적으로 생성한다. 습윤기간 생성단계에서는 다지점 강우사상을 발생시켜 강우사상의 지속시간(습윤기간의 길이)과 강수량을 동시에 모의하며, 건조기간 생성단계에서는 사상 간 시간간격의 경험분포를 이용해 건조기간 길이를 생성한다.

강우사상은 다시 ‘일반강우’와 ‘호우’로 나눈다. 여기서 ‘일반강우’란 모든 지점에

서 기준값 이하의 강수량이 내리는 사상을 말하고, ‘호우’는 그 반대의 경우를 뜻한다. 우리는 강우사상자료를 활용해 월별로 강우사상이 ‘호우’일 확률을 계산한다. 즉, 해당 월에 발생한 모든 강우사상 중에서 ‘호우’로 분류되는 사상의 비율을 확률로 정한다. 이 확률에 따라 생성될 강우사상이 ‘호우’인지 ‘일반강우’인지를 결정한다.

일반강우는 강우사상자료에서 ‘일반강우’로 분류된 강우사상을 임의로 추출하는 방식으로 모의한다. 이는 강우사상의 경험분포를 쓰는 것과 동일하다. 단, 강우의 월별 특성을 감안하여, 월별로 경험분포를 구한다. 그런데 시(時)단위의 모의를 위해서는 강우사상의 시간별 강수량을 생성해야 하는데, 일반 강우의 모의에서는 추출된 강우사상의 시간별 강수량이 자료에서 가용하므로 이를 쓴다.

한편, ‘호우’는 모의과정이 복잡하다. 우선 다지점 강우량은 logistic 모형으로부터 생성된 난수를 Box-Cox 변환하여 얻어진다. 이 생성된 강우량을 y_1, \dots, y_d 라고 두면, 지속시간 τ 은 형태모수가 ν 이고 평균이 $(\beta_0 + \beta_1 y_1 + \dots + \beta_d y_d)^{-1}$ 인 Gamma 분포로 모의된다($\nu, \beta_0, \dots, \beta_d$ 는 앞서 일반화 선형모형 적합을 통해 추정되었다). 하지만 사상의 시간별 강수량에 대한 통계모형은 앞서 고려되지 않았다. Rodriguez-Iturbe et al. (1987)와 김병식 외 3인(2008) 등에서 한 지점에서 일강수량의 시단위분해(hourly disaggregation)에 대한 확률모형을 고려하였으나, 다지점에 관한 연구는 아직 없는 것으로 보인다. 여기서는 이에 대한 대안으로 k -NN 알고리즘을 이용하여 강우사상자료에서 강우량과 지속시간이 각각 y_1, \dots, y_d 과 τ 와 유사한 강우사상을 추출하고 이것의 시간별 강수량을 차용하는 방법으로 강우량 시단위분해를 하고자 한다. 표 24는 시단위분해의 예시로서, 총강우량이 50mm 지속시간이 4시간인 강우사상의 시단위분해를 위해, 총강우량 30mm이고 지속시간은 8시간인 강우사상이 추출된 경우이다. 예시에서처럼, 지속시간이 임의추출의 특성상 다를 경우가 대부분이다. 이 경우 추출된 사상의 지속시간 8시간을 4시간 기준으로 환산하여, 각 시간의 강우량 비율을 계산한다. 그리고 계산된 비율에 맞추어 총강수량 50mm를 각 시간대에 배분해 최종적으로 모의된 강우사상을 얻는다.

Table 24. 총강수량 50mm, 지속시간 4시간인 강우사상의 시단위분해의 예

	시간별 강수량								합 계
	5mm	5mm	6mm	2mm	1mm	1mm	3mm	7mm	
추출된 강우사상	5mm	5mm	6mm	2mm	1mm	1mm	3mm	7mm	30mm/ 8시간
비율	16.7%	16.7%	20%	6.7%	3.3%	3.3%	10%	23.3%	100%/ 8시간
4시간 기준	33.4%		26.7%		6.6%		33.3%		100%/ 4시간
최종 강우사상 모의결과	16.7mm		13.35mm		3.3mm		16.65mm		50mm/ 4시간

강우모의는 시단위로 시행되지만 기상생성의 단위가 하루이기 때문에, 나온 결과를 다시 하루단위로 환산한다. 표 25-28은 한강 상류유역에 대해 6월부터 9월 사이 강우 모의자료를 하루단위로 환산한 것이다. 모의 자료는 WGEN과 LARS-WG와 같은 기존 weather generator에서 달리 강우의 지점 간 연관성을 표현한다.

3.3.2 모의 유사성 평가

본 소절에서는, 제안된 강우모의알고리즘을 통해 얻어진 모의자료가 실제 강우현상을 잘 반영하는지를 평가해 보고자 한다. 우선, 표 25-28과 같은 모의자료를 10000개 생성하고, 호우기 강우 특징을 나타내는 대푯값-총누적강수량, 최대일강수량(maximum of daily rainfall), 강우일수를 계산하였다. 또한 모형구축에 사용된 관측자료에 대해서도 같은 값을 계산하여 유사성평가에 활용한다. 이들 대푯값을 선정한 이유는 모의알고리즘이 시단위로 수행되기 때문에 짧은 시간단위에서는 강우현상을 잘 모의할 것으로 기대되나, 호우기 전체 기간과 같은 긴 시간단위에서 강우의 특성을 잘 반영하는지 확인이 필요하기 때문이다.

3.3.2.1 지점별 모의 유사성 평가

먼저, 지점별로 모의 유사성을 평가한다. 유사성은 모의자료와 관측자료의 대푯값이 따르는 확률분포의 일치도로 평가할 수 있다. 각 지점별로 관측대푯값(관측자료의 대푯값)의 히스토그램과 모의대푯값(모의자료의 대푯값)의 확률밀도함수를 구해 비교하였고, t-검정과 Kolmogorov-Smirnov 검정(K-S 검정)을 통해 분포의 일치성을 검정하였다. 그림 38-67은 확률분포비교를 위한 그림을, 표 29-58은 각각 분포차이 유의성 검정의 유의확률을 나타낸다. 낙동강 상류유역의 경우, 분포형태그림에서 두 분포가 유사하다는 것을 확인할 수 있고, 검정도 또한 유의수준 5%에서 두 분포가 일치함을 시사한다. 그러므로 낙동강 상류유역의 호우기와 월별 누적강수량, 최대일강수량, 강우일수가 모두 잘 재현되는 것으로 판단된다. 한편, 한강 상류유역도 역시 누적강수량과 최대일강수량은 대체적으로 잘 재현되나, 9월 강우일수는 상대적으로 유사성이 낮은 것으로 보인다.

3.3.2.2 지점 간 강우연관 재현성 평가

이제, 지점 사이 강우연관성이 잘 모의되는지 확인해 보기로 한다. 우선, 낙동강 상류유역부터 살펴보도록 하겠다. 모의대푯값 지점 쌍의 결합확률분포(joint probability distribution)를 Gaussian kernel 함수를 이용해 추정하고 등고선그림으로 추정결과를 표현하였다. 그리고 등고선 그림 위에 관측대푯값을 점으로 나타내어, 비교를 원활하게 하였다(그림 68-79). 아울러 상관계수로 함께 계산하였다(표 59-61). 단, 누적강수량과 강우일수에 대해서는 일반적인 Pearson 상관계수를 사용하였으나, 최대일강수량의 경우는 모의자료가 이상점(outlier)을 많이 지니고 비정규성(non-normality)이 의심되기 때문에 보다 강건(robust)하다고 알려진 Spearman 상관계수를 썼다. 누적강수량과 강우일수의 경우, 등고선 그림에서 관측과 모의의 분포가 유사하게 나타난다. 또한, 상관계수의 일치성 검정 결과, 모든 지점 간에 대해 차이가 5%수준에서 유의하지 않았다. 그래서 우리는 호우기 누적강수량과 강우일수의 지점 간 상관관계가 잘 모의되는 것으로 판단한다. 한편, 최대일강수량의 경우는 이상점이 자주 등장하며, 상관관계도 역시 여러 지점 간에서 차이가 있는 것으로 나타났다.

한강 상류유역에 대해서도 동일한 방식으로 모의 유사성을 시험해 보았다. 그림 80-88은 모의와 관측대푯값들의 확률분포를, 표 62-64은 상관계수 계산 결과를 나타낸다. 단, 누적강수량과 최대일강수량은 모의자료에 이상점이 많아서 Spearman 상관계수를 활용하였다. 강우일수의 상관관계는 등고선그림과 상관계수를 볼 때 잘 모의가 되는 것으로 판단된다. 하지만 누적강수량과 최대일강수량의 경우는 여러 지점 간에서 상관정도가 일치하지 않는 것으로 나타났다.

3.3.2.3 극한강우의 재현성 - 극단값 모형의 도입 효과

본 연구에서는 호우의 재현성을 제고하기 위해 강우모의알고리즘에 극단값 모형을 도입하였다. 이 절에서는 시도된 도입의 효과를 검증하고자 한다. 일반적으로 극단값 모형 도입의 주목적은 관측자료의 시간범위를 넘어선 긴 return period에서 return level을 얻기 위함이다. 그래서, 우선 지점별로 호우의 극한강수량에 대한 return level을 구해보고자 한다. 구체적으로는 통상적으로 고려되는 최대일강수량의 return level과 더불어, 호우사상의 지속시간이 평균적으로 이틀이 넘는 것을 감안하여, ‘최대3일연속누적강수량’ (maximum of 3-day cumulated rainfalls)도 함께 고려한다. 최대3일연속누적강수량은 호우기의 모든 날에 대해 전후 하루씩 포함한 3일 동안 누적강수량을 계산한 다음에 그 중 최대를 구하는 과정으로 얻어진다. 도입의 효과를 확인하기 위한 비교대상으로 본 연구에서 제안된 알고리즘(그림 36)에서 ‘호우’ 경우를 제거하여 단순화된 알고리즘을 선택하였다(그림 37). 대신, 강우사상생성 단계에서는 강우사상자료에서 ‘호우’와 일반강우’의 구분없이 강우사상을 임의추출하는 방식을 택하고, 그 외의 세부방식은 동일하다. 편의상, 원래의 알고리즘을 ‘제안알고리즘’, 단순화된 알고리즘을 ‘대조알고리즘’이라고 부르도록 한다.

그림 92-95에서는 제안알고리즘, 대조알고리즘의 return level과 더불어, 관측자료의 최대일강수량 또는 최대3일연속누적강수량에 대해 GEV 분포(generalized extreme value distribution)를 적합시켜 얻은 return level과 그것의 95% 오차범위를 함께 나타내었다. 대조알고리즘의 return level은 어느 기간 이상의 return period에서는 더 이상 증가하지 않아 음의 편의(negative bias)가 있음을 알 수 있다. 한편, 최대일강수량

에 대해 제안된 알고리즘의 return level은 몇몇 지점에서 GEV 모형의 return level 기준으로 양의 편의(positive bias)가 있는 것으로 보이는 반면, 최대3일연속누적강수량에 대해서는 GEV 모형 return level에 근접하거나 오차범위 내에서 안정적으로 return level을 제시하고 있다. 봉화화 철원지점에서는 제안알고리즘의 return level이 발산하는데, 이는 이들 지점에 Pareto 모형($\gamma > 0$)이 선택되었기 때문인 것으로 보인다.

다음으로 극한강우의 지점 간 연관성이 잘 재현되는지 확인해 보고자 한다. 두 지점 간 연관성을 나타내는 측도로 식(6)에서 주어진 계수 χ 를 고려한다. 우리는 호우기 최대일강수량과 최대3일연속누적강수량의 χ 값을 계산한다. 관측값에 대해서는 이변량(bivariate) logistic 모형의 적합을 통해 얻어진 α 의 추정치로부터 식(7)을 이용해 χ 값을 얻는다(Coles et al.(2001), Section 8.2). 모의값에 대해서는 식(6)에 대응하는 경험치(empirical estimate)

$$\sum_{l=1}^N I(x_{l,i}^* > a_i, x_{l,j}^* > a_j) \left\{ \sum_{l=1}^N I(x_{l,i}^* > a_i \text{ or } x_{l,j}^* > a_j) \right\}^{-1}$$

을 구한다. 위 식에서, $N=10000$ 으로 모의의 반복횟수를 나타내고, $(x_{1,i}^*, x_{1,j}^*), \dots, (x_{N,i}^*, x_{N,j}^*)$ 는 N 번의 모의를 통해 얻어진 지점 i 와 j 의 최대일강수량 또는 최대3일연속누적강수량 쌍을 나타내며, a_i 와 a_j 는 충분히 큰 기준값으로 각각 $x_{1,i}^*, \dots, x_{N,i}^*$ 와 $x_{1,j}^*, \dots, x_{N,j}^*$ 의 상위 1% 분위수로 잡았다.

표 65-72은 계산된 χ 값을 나타낸다. 전반적으로 대조알고리즘은 연관성을 과소추정하는 성향이 있는 것으로 보여 진다. 특히, 유의하게 연관성이 있는 지점 간에 대해서 독립이거나 그에 가까운 모의결과를 제시하여 편의가 심하다. 이에 반해 제안알고리즘은 유의한 연관성이 있는 지점 간에서는 관측값의 χ 와 오차범위 내에서 안정적인 값을 제시한다.

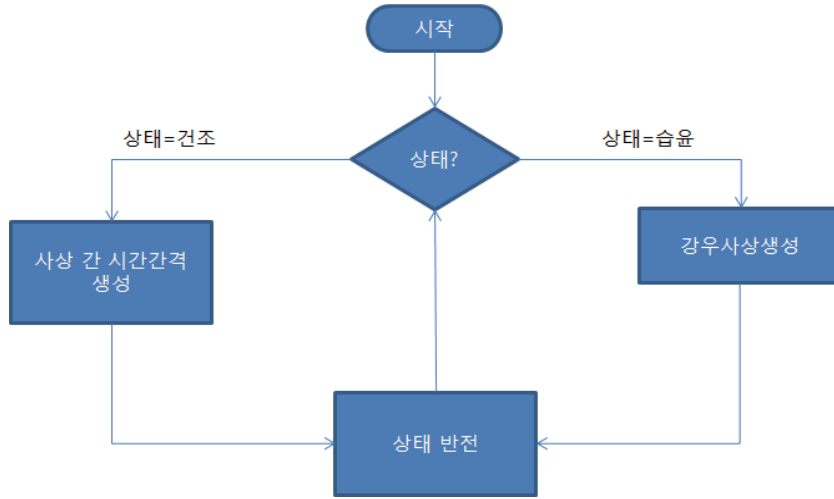


Figure 37. 대조알고리즘. 강우사상생성단계에서는 강우사상자료로부터 ‘호우’와 ‘일반강우’의 구분없이 강우사상을 임의추출하여 모의한다.

Table 25. 모의 자료(6월)

월	일	철원	춘천	원주	영월	홍천	제천
6	1	5.6	4.3	0	0	0.5	0
6	2	0	0.5	1	2	1	0.4
6	3	0	0	0	0	0	0
6	4	0	0	0	0	0	0
6	5	0	0	0	0	0	0
6	6	0	0	0	0	0	0
6	7	0	0	0	0	0	0
6	8	0	0	0	0	0	0
6	9	0	0	0	0	0	0
6	10	0	0	0	0	0	0
6	11	5.5	0.3	5.5	0	1	0.1
6	12	0	0.1	0	0	0.1	0
6	13	4.5	8	1.5	0	0.5	0
6	14	0	0	0	0	0	0
6	15	0.5	0	0	0	0.5	0
6	16	0	0	0	0	0	0
6	17	0	0	0	0	2	0
6	18	0.2	0.4	2.6	3.2	1.5	0.5
6	19	0.5	0.4	0.5	0	4	0
6	20	0.5	0.3	2.5	0	1	1
6	21	3.4	0.5	5.7	0	0	0
6	22	16.6	28.7	50.9	40.1	54.5	43.5
6	23	0	0	0	0	0	0
6	24	0	0	0	0	0	0
6	25	51.5	65.5	32.5	22	85	41
6	26	4.5	5.5	16.5	11.5	16	17.5
6	27	0	0	0	0	0	0
6	28	0	0	0	0	0	0
6	29	0.4	0	0.4	0	0	0
6	30	0	0	0	0	0	0

Table 26. 모의 자료(7월)

월	일	철원	춘천	원주	영월	홍천	제천
7	1	0	0	0	0	0	0
7	2	0	0	0	0	0	0
7	3	0	0	1.2	1	2.9	1.8
7	4	68.2	68.7	145.1	31.2	101.9	36.2
7	5	6.8	11.8	98.7	40.7	18.8	28.3
7	6	108.9	95.8	27.5	52.4	47.8	65.6
7	7	39.9	51.5	43.3	55.8	124.7	81.5
7	8	16.5	12.2	29.5	14	12	12.3
7	9	27.5	7.5	5.5	4.1	3.5	1
7	10	0	0.2	0	0	0	0
7	11	0	0.1	0	0	0	0
7	12	0	0	0	8	0	4
7	13	22	28.5	0.2	5	12	0.3
7	14	29	21.5	58.8	51.5	28	55.2
7	15	27	28	33	50.6	56.5	21
7	16	0	0	0	0	1	0.5
7	17	4	5	0	0	0	0
7	18	0	0.1	3.5	38	0.5	29
7	19	0.5	0.5	0	1	0	7
7	20	0.5	0.1	0.1	0	0	0
7	21	0	0	1	0	0	0
7	22	0	0	0	0	0	0
7	23	0.2	0.3	0	0.1	0.4	0.1
7	24	89.8	40.2	11.5	0.5	13.1	2.6
7	25	2.5	0.5	0.5	0	0	0.5
7	26	0	0	0	5.2	0	1.5
7	27	29.9	12.9	0	0	6.5	0
7	28	0	0	0	0	0	0
7	29	0	0	0	0	0	0
7	30	0	0	0	0	0	0
7	31	0.5	0.5	0	0	0	0

Table 27. 모의 자료(8월)

월	일	철원	춘천	원주	영월	홍천	제천
8	1	0	0	0	0	0	0
8	2	0	0	0	0	0	0
8	3	0	0	0	0	0	0
8	4	0	0	0	0	0	0
8	5	0	0	0	0	0	0
8	6	0	0	0.2	0.2	0	0
8	7	0	1	0.5	0.5	0	0.5
8	8	0	0	0	0	0	0
8	9	0	0	39	0	0	0
8	10	24	62.5	4.5	6.5	33.5	11
8	11	7.5	9.5	3	10.5	3	7
8	12	24	5	2.5	5.5	0	13.5
8	13	0	0	0.1	0	0	0
8	14	3	0	0	0	0	0
8	15	0	0	0	0	0	0
8	16	0	0	0	0	0	0
8	17	0	0	0	0	0	0
8	18	0	0	0	0	0	0
8	19	0.5	0	9	1.5	1	0.5
8	20	0	0.2	6.5	12.9	15	4.1
8	21	0	0	0	0	0	0
8	22	5	0.3	3.4	0	5.5	0
8	23	0	2	25.5	21	18.5	18
8	24	0	0	0	0	0	0
8	25	0	0	0.1	0	0	0
8	26	3	2.5	7.4	15.8	15.5	13
8	27	8	0.1	0	0	6.5	0
8	28	27	18.5	0	0	36	0
8	29	0	0	0	0	0	0
8	30	0	0	0	0	0	0
8	31	0	0	0	10	0	12.5

Table 28. 모의 자료(9월)

월	일	철원	춘천	원주	영월	홍천	제천
9	1	2.5	2.9	3.7	37.5	12.7	38.5
9	2	0	0	0.6	0	0	0.1
9	3	0	0	0	0	0	0
9	4	0	0	0	0	0	0
9	5	0	0	0	0	0	0
9	6	1.5	3	57.5	31	12.5	19.5
9	7	19	11.5	66	37.5	113.5	52
9	8	1.7	5.9	0	0	0.5	0
9	9	0	0.2	0.1	0	0	0
9	10	0	0	0	0	0	0
9	11	0	0	0.1	0	0	0
9	12	2	3.5	2.4	0.5	2	1
9	13	19.5	32	30	27.5	10	15
9	14	64	35	2	0	48.5	0
9	15	1.5	11.5	47.5	88	10.5	124.5
9	16	0	0	0	0	0	0
9	17	0	0.5	7.5	9	4	9.5
9	18	48	43.7	14.1	5.4	38	11.5
9	19	31	42.8	42.4	21.1	70.5	24
9	20	37.5	24	35	4	33.5	7.5
9	21	7	1	0	0	0.5	0
9	22	76	46.5	70.5	31	84	52.5
9	23	19	18	17	25	32.5	30
9	24	0	0	0	9.5	0	2.5
9	25	0	0.4	0.9	2.5	0	0
9	26	0	0	0	0	0	0
9	27	0.1	0	0	0	0	0
9	28	0	0	0	0	0	0
9	29	0	0	0	0	0	0
9	30	0	0	0	0	0	0

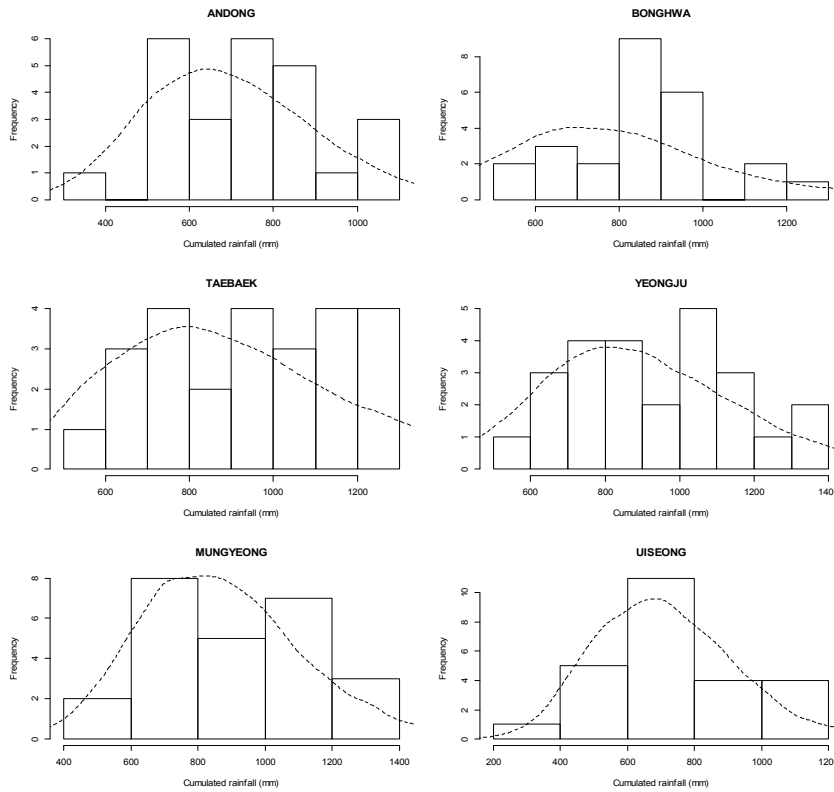


Figure 38. 낙동강 상류 6~9월 지점별 누적강수량: 관측(히스토그램)과 모의(파선)

Table 29. 낙동강 상류 6~9월 누적강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	42%	72%	63%	48%	46%	67%
K-S 검정 유의확률	57%	6%	56%	67%	79%	94%

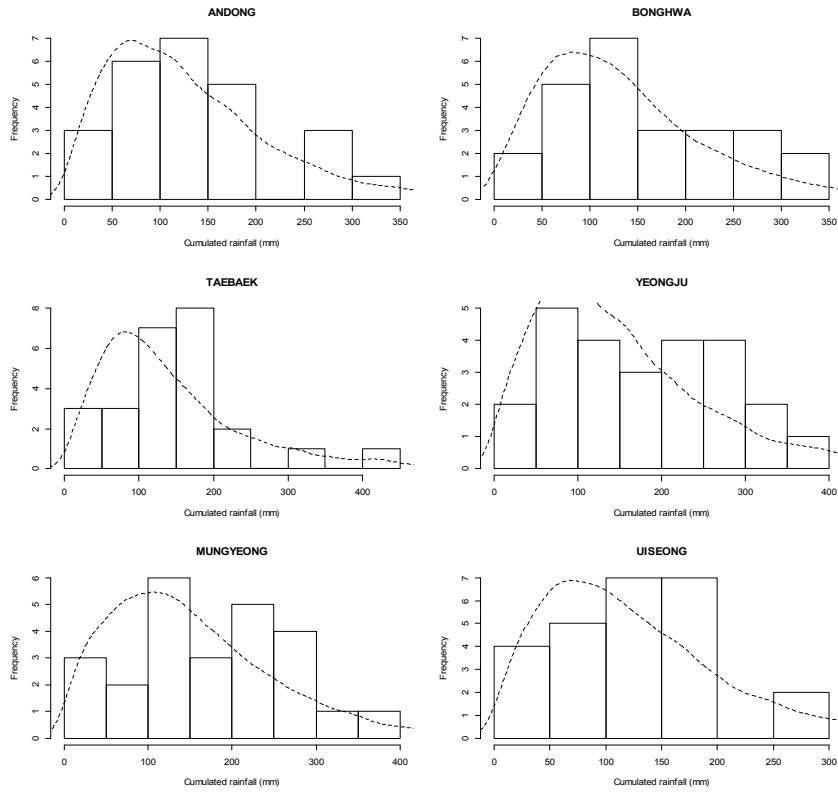


Figure 39. 낙동강 상류 6월 지점별 누적강수량: 관측(히스토그램)과 모의(파선)

Table 30. 낙동강 상류 6월 누적강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	58%	58%	87%	30%	32%	97%
K-S 검정 유의확률	44%	41%	35%	27%	40%	50%

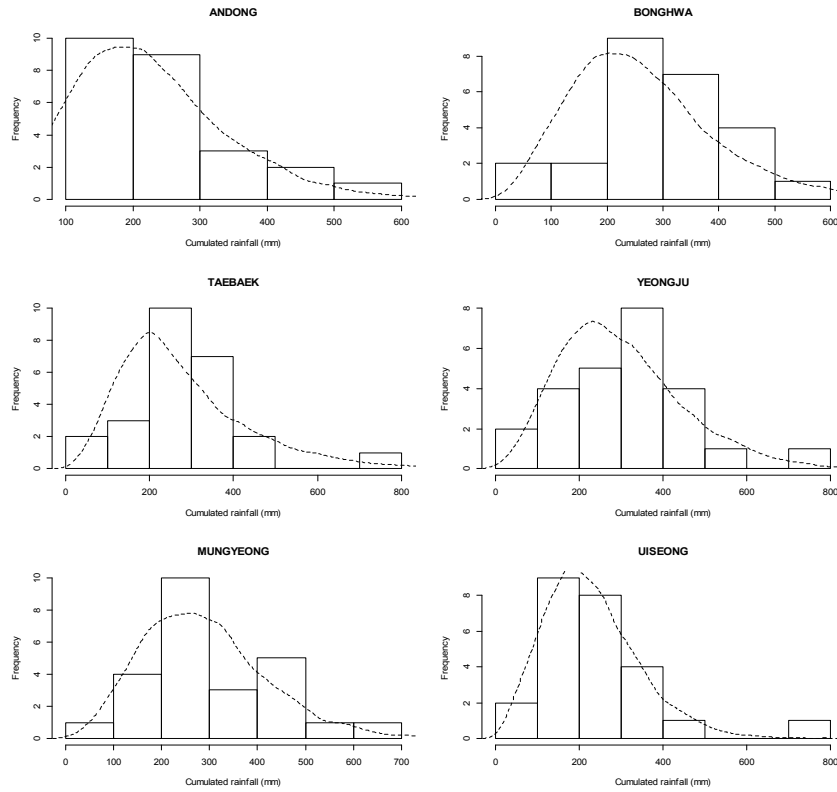


Figure 40. 낙동강 상류 7월 지점별 누적강수량: 관측(히스토그램)과 모의(파선)

Table 31. 낙동강 상류 7월 누적강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	50%	39%	88%	62%	41%	71%
K-S 검정 유의확률	56%	10%	22%	60%	70%	99%

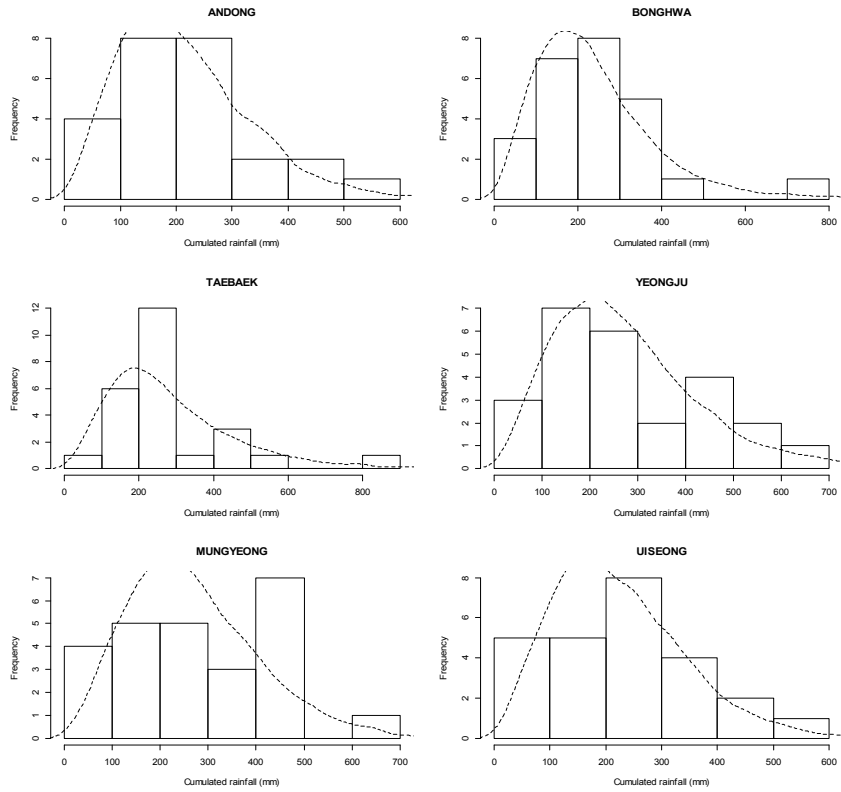


Figure 32. 낙동강 상류 8월 지점별 누적강수량: 관측(히스토그램)과 모의(파선)

Table 32. 낙동강 상류 8월 누적강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	66%	87%	91%	82%	81%	60%
K-S 검정 유의확률	86%	61%	70%	83%	53%	81%

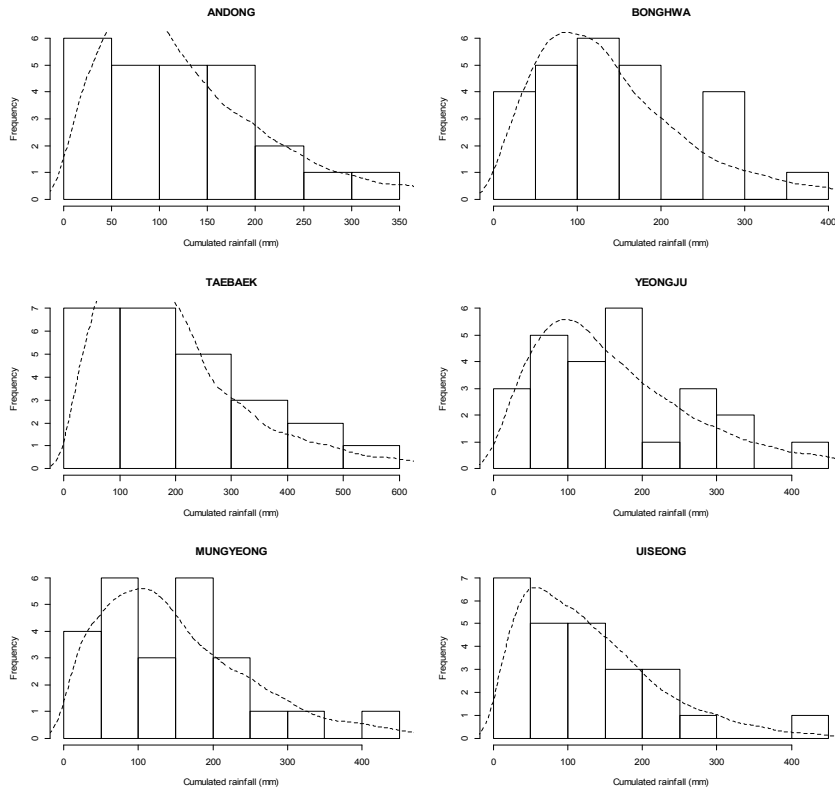


Figure 42. 낙동강 상류 9월 지점별 누적강수량: 관측(히스토그램)과 모의(파선)

Table 33. 낙동강 상류 9월 누적강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	78%	50%	49%	67%	63%	80%
K-S 검정 유의확률	95%	99%	76%	99%	93%	92%

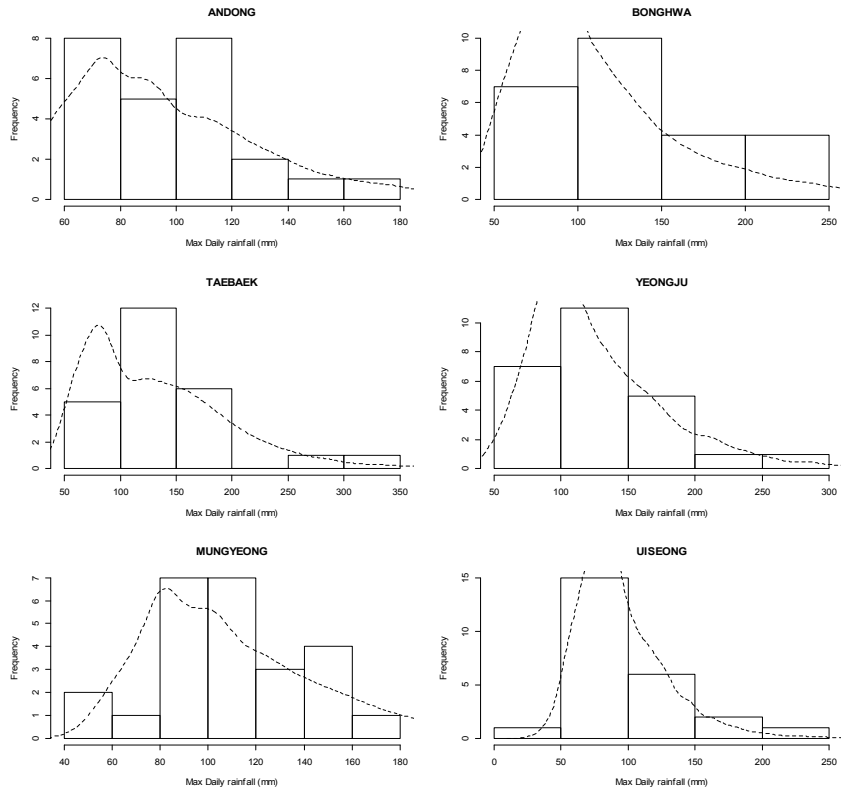


Figure 43. 낙동강 상류 6~9월 지점별 최대일강수량: 관측(히스토그램)과 모의(파선)

Table 34. 낙동강 상류 6~9월 최대일강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	78%	98%	44%	41%	64%	68%
K-S 검정 유의확률	28%	11%	7%	57%	57%	89%

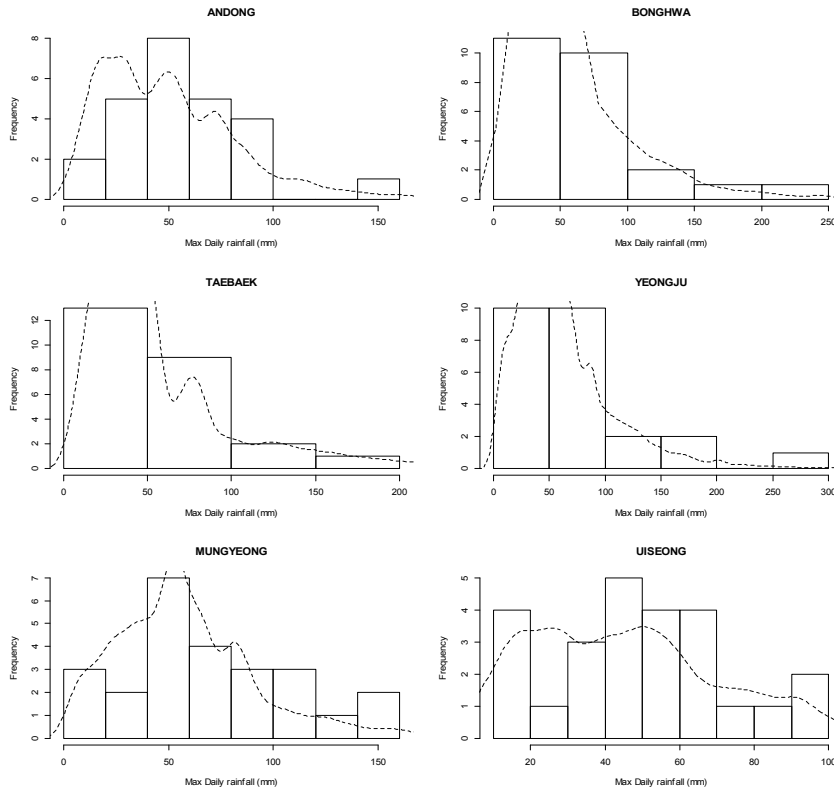


Figure 44. 낙동강 상류 6월 지점별 최대일강수량: 관측(히스토그램)과 모의(파선)

Table 35. 낙동강 상류 6월 최대일강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	39%	43%	84%	19%	14%	99%
K-S 검정 유의확률	27%	33%	95%	22%	13%	59%

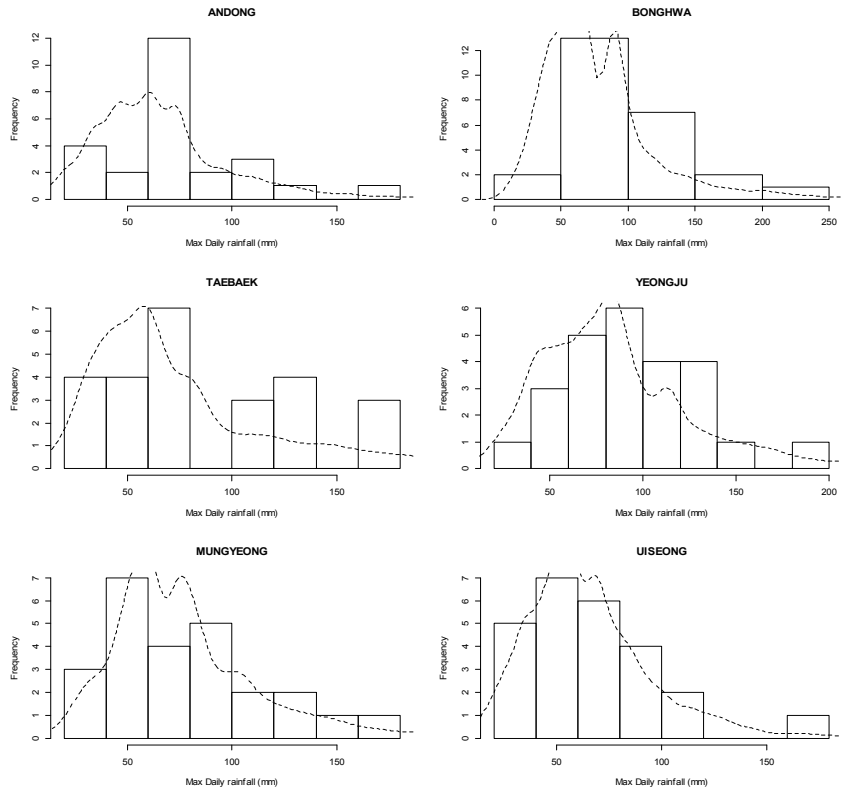


Figure 45. 낙동강 상류 7월 지점별 최대일강수량: 관측(히스토그램)과 모의(파선)

Table 36. 낙동강 상류 7월 최대일강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	12%	16%	52%	43%	78%	77%
K-S 검정 유의확률	10%	11%	95%	49%	82%	92%

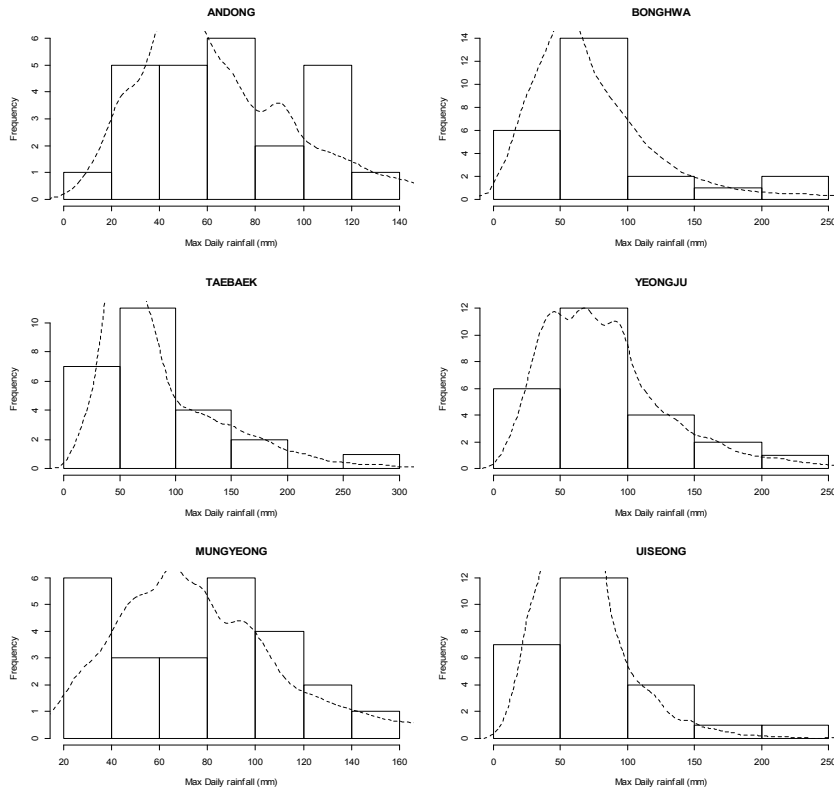


Figure 46. 낙동강 상류 8월 지점별 최대일강수량: 관측(히스토그램)과 모의(파선)

Table 37. 낙동강 상류 8월 최대일강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	86%	65%	87%	84%	70%	53%
K-S 검정 유의확률	81%	99%	99%	96%	69%	93%

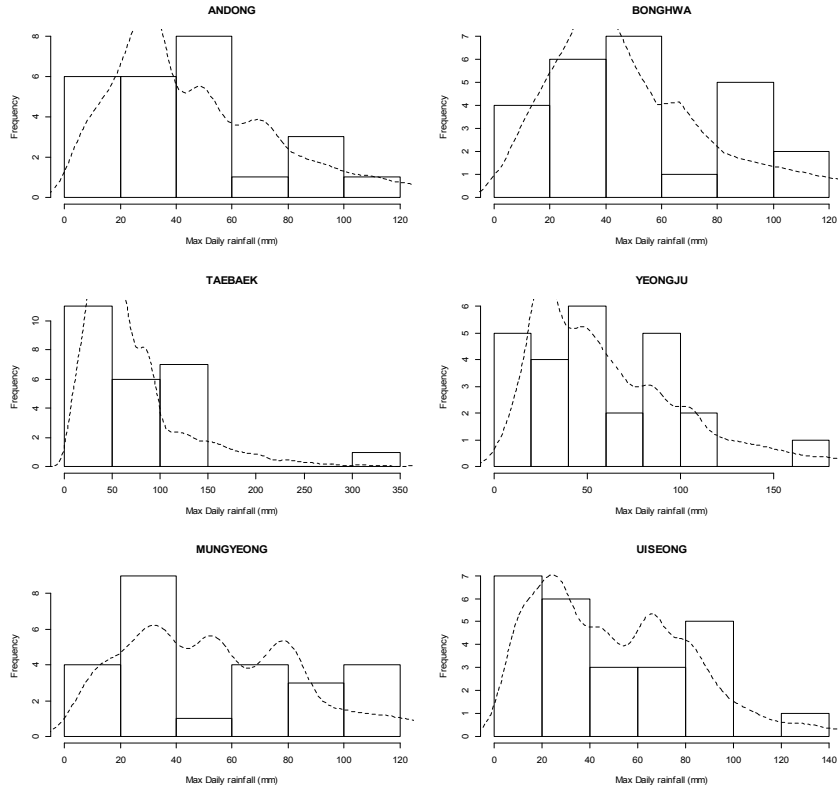


Figure 47. 낙동강 상류 9월 지점별 최대일강수량: 관측(히스토그램)과 모의(파선)

Table 38. 낙동강 상류 9월 최대일강수량 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	44%	21%	41%	45%	31%	29%
K-S 검정 유의확률	95%	93%	35%	91%	68%	55%

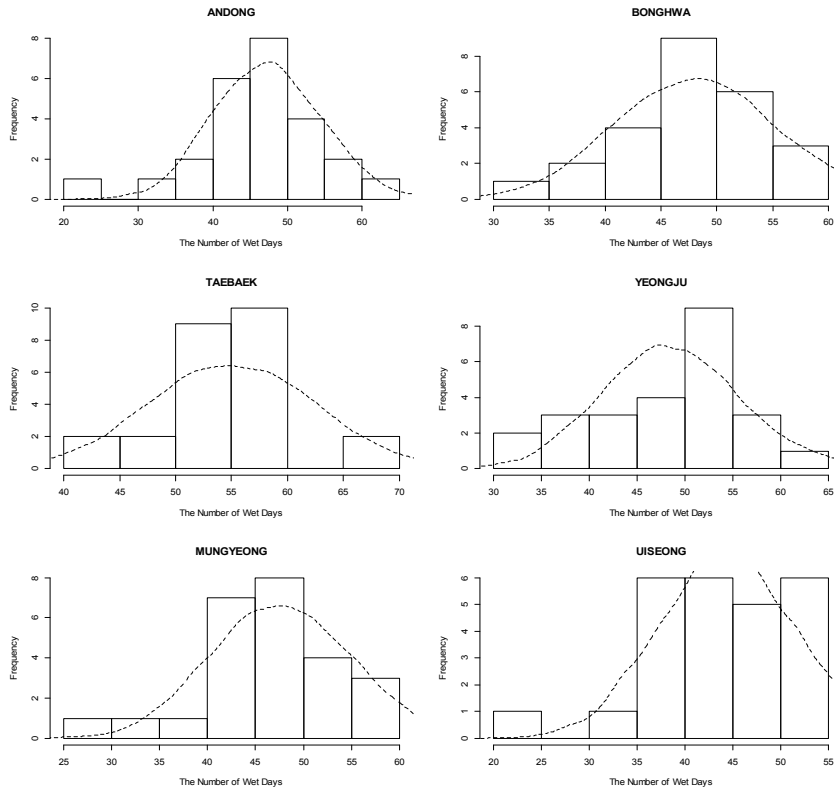


Figure 48. 낙동강 상류 6~9월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 39. 낙동강 상류 6~9월 강우일수 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	80%	96%	82%	91%	60%	53%
K-S 검정 유의확률	95%	99%	56%	36%	99%	81%

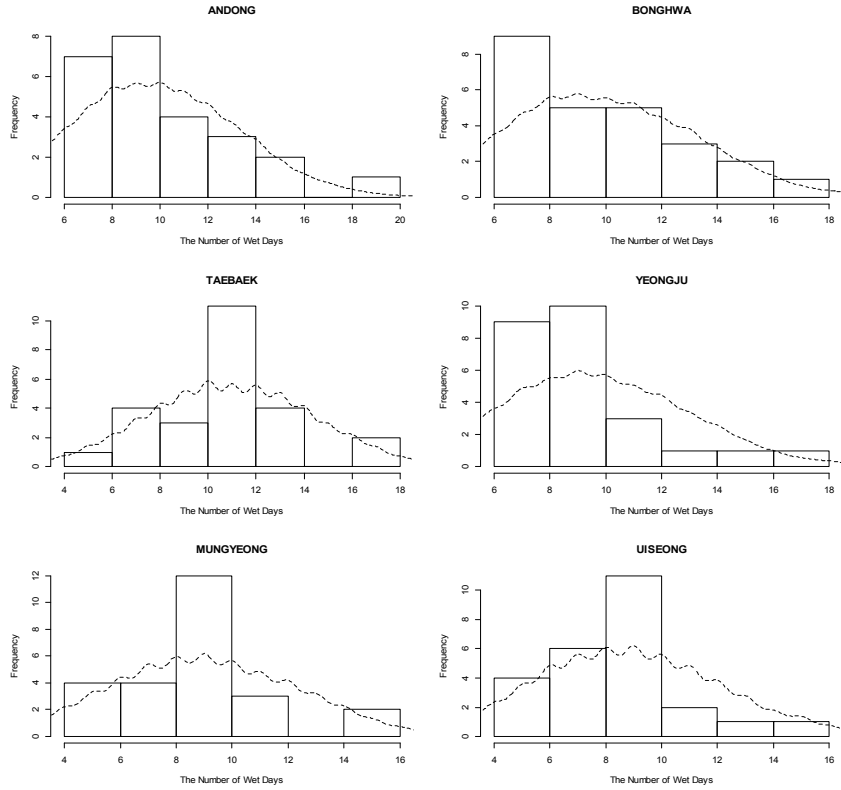


Figure 49. 낙동강 상류 6월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 40. 낙동강 상류 6월 강우일수 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	46%	51%	75%	93%	70%	93%
K-S 검정 유의확률	83%	53%	77%	70%	77%	65%

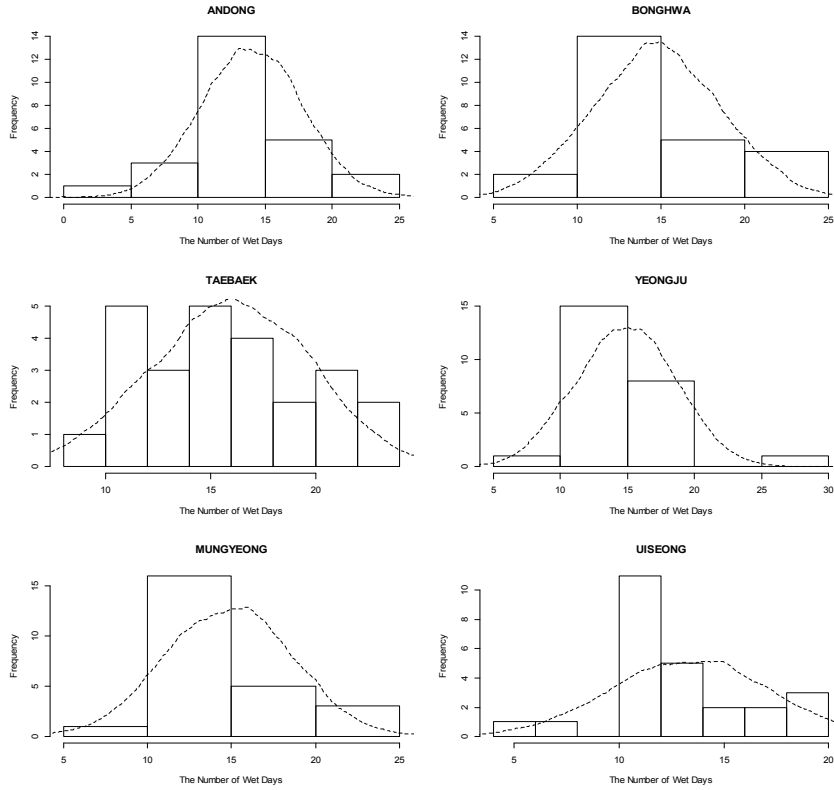


Figure 50. 낙동강 상류 7월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 41. 낙동강 상류 7월 강우일수 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	98%	82%	72%	71%	78%	91%
K-S 검정 유의확률	99%	98%	99%	99%	60%	48%

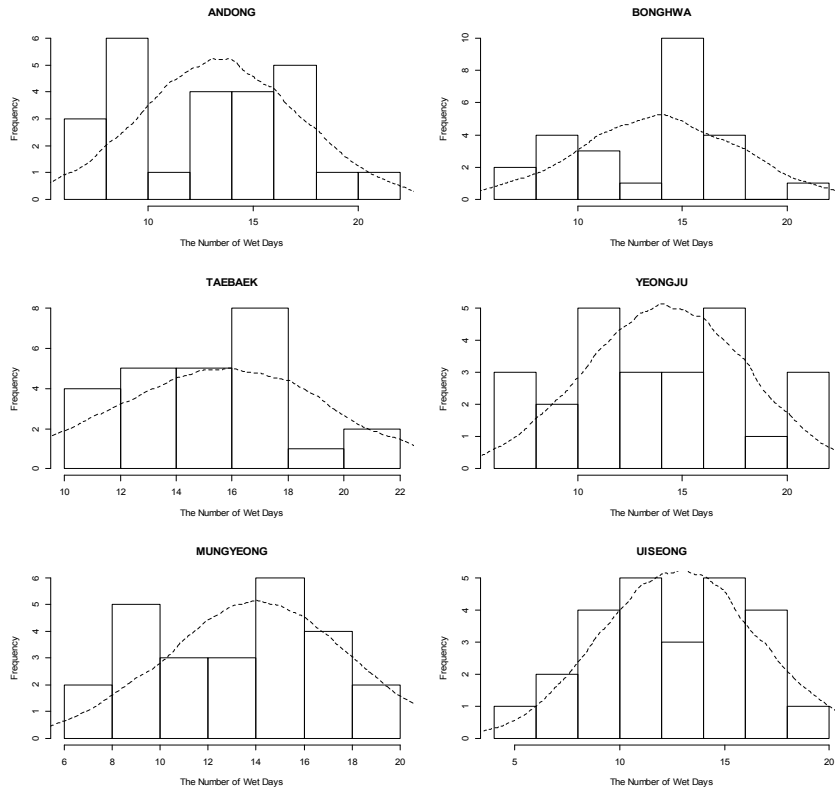


Figure 51. 낙동강 상류 8월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 42. 낙동강 상류 8월 강우일수 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	82%	96%	96%	96%	53%	90%
K-S 검정 유의확률	49%	51%	93%	89%	93%	98%

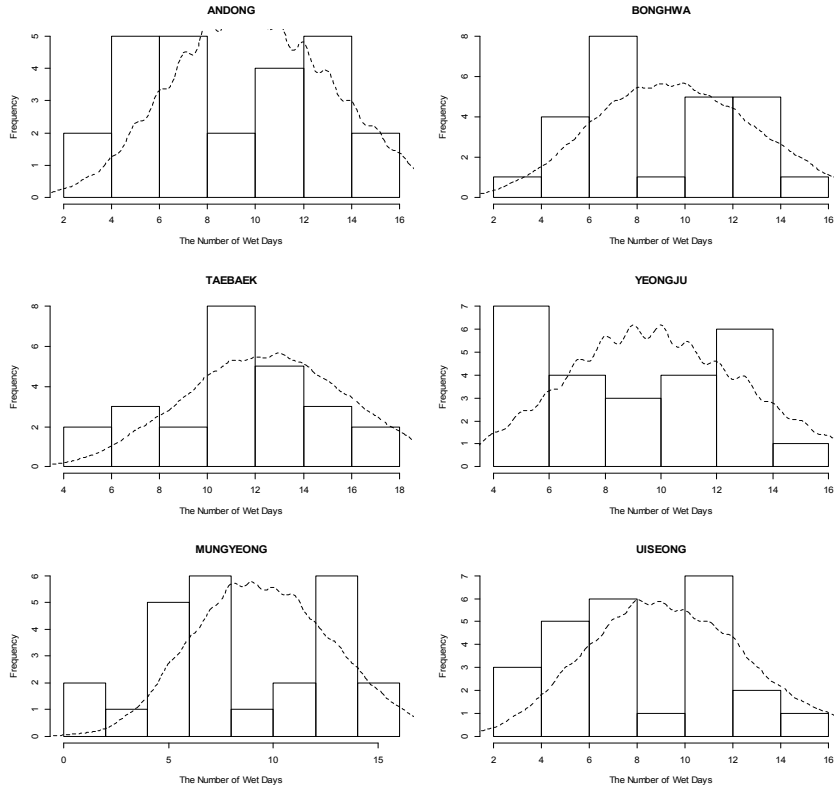


Figure 52. 낙동강 상류 9월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 43. 낙동강 상류 9월 강우일수 분포검정

	안동	봉화	태백	영주	문경	의성
t-검정 유의확률	33%	47%	23%	52%	39%	27%
K-S 검정 유의확률	50%	48%	92%	58%	23%	40%

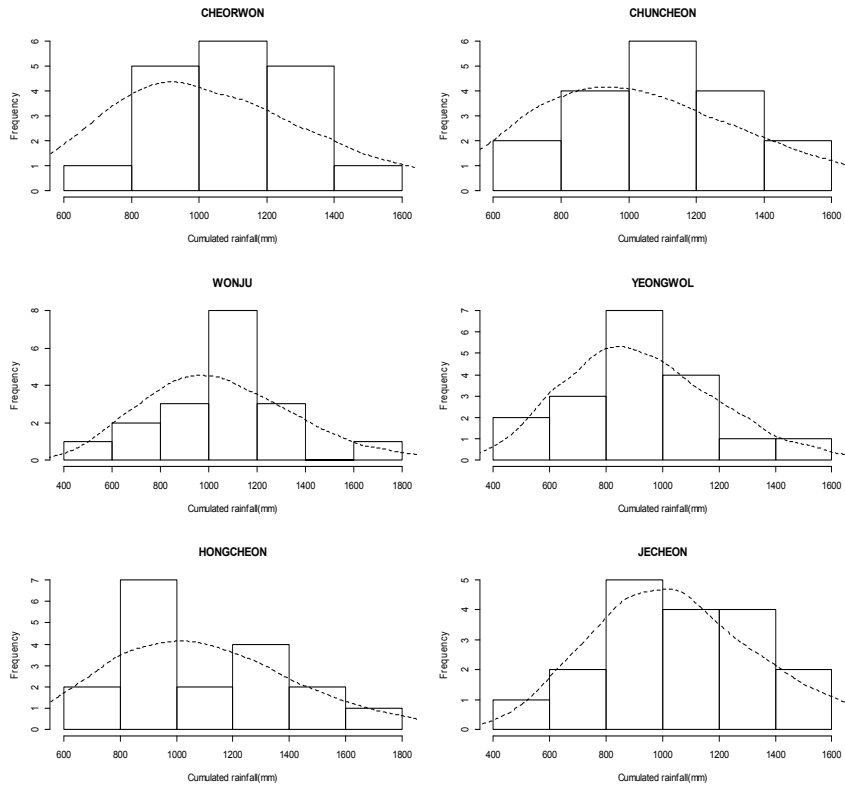


Figure 53. 한강 상류 6~9월 지점별 누적강우량: 관측(히스토그램)과 모의(파선)

Table 44. 한강 상류 6~9월 누적강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	84%	75%	98%	84%	93%	91%
K-S 검정 유의확률	41%	29%	32%	40%	68%	99%

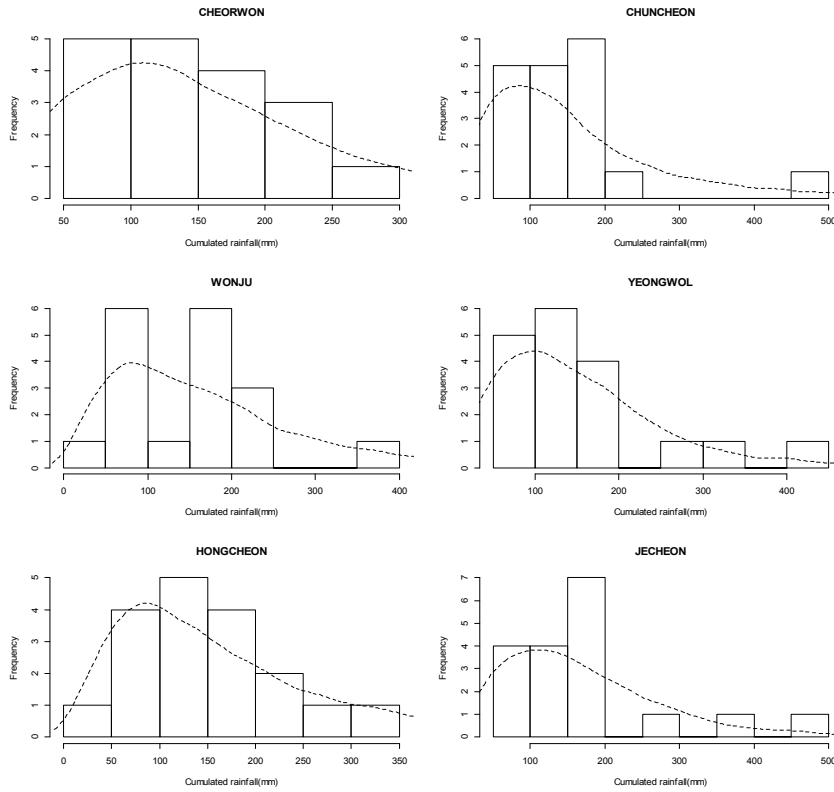


Figure 54. 한강 상류 6월 지점별 누적강우량: 관측(히스토그램)과 모의(파선)

Table 45. 한강 상류 6월 누적강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	34%	43%	43%	83%	22%	85%
K-S 검정 유의확률	79%	45%	78%	45%	77%	61%

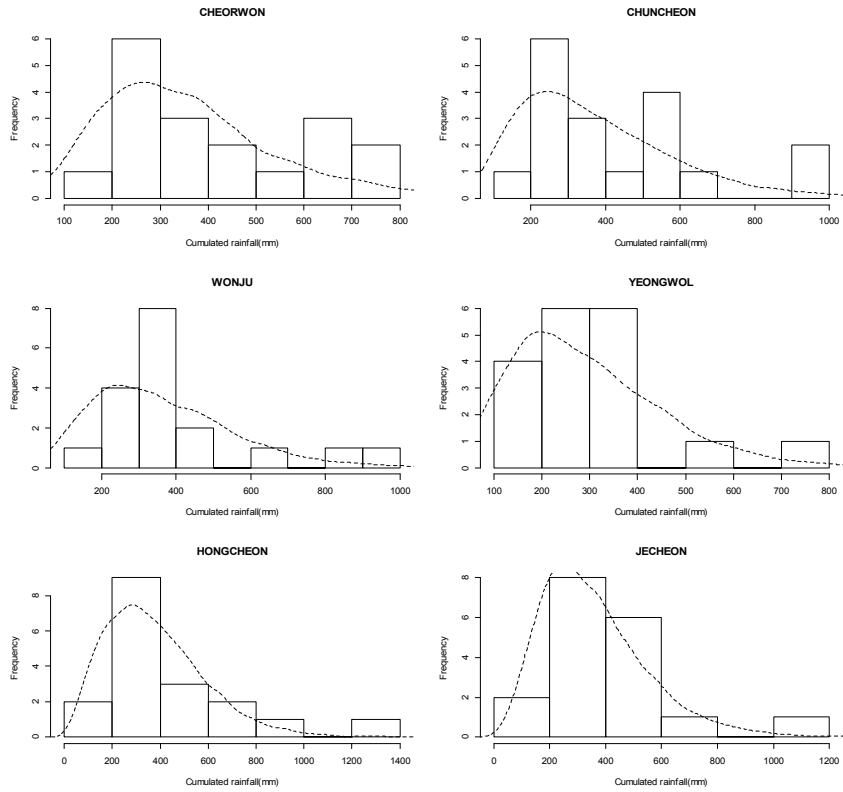


Figure 55. 한강 상류 7월 지점별 누적강우량: 관측(히스토그램)과 모의(파선)

Table 46. 한강 상류 7월 누적강우량 분포검정

	철원	춘천	원주	영일	홍천	제천
t-검정 유의확률	30%	33%	41%	75%	30%	61%
K-S 검정 유의확률	53%	54%	40%	54%	84%	75%

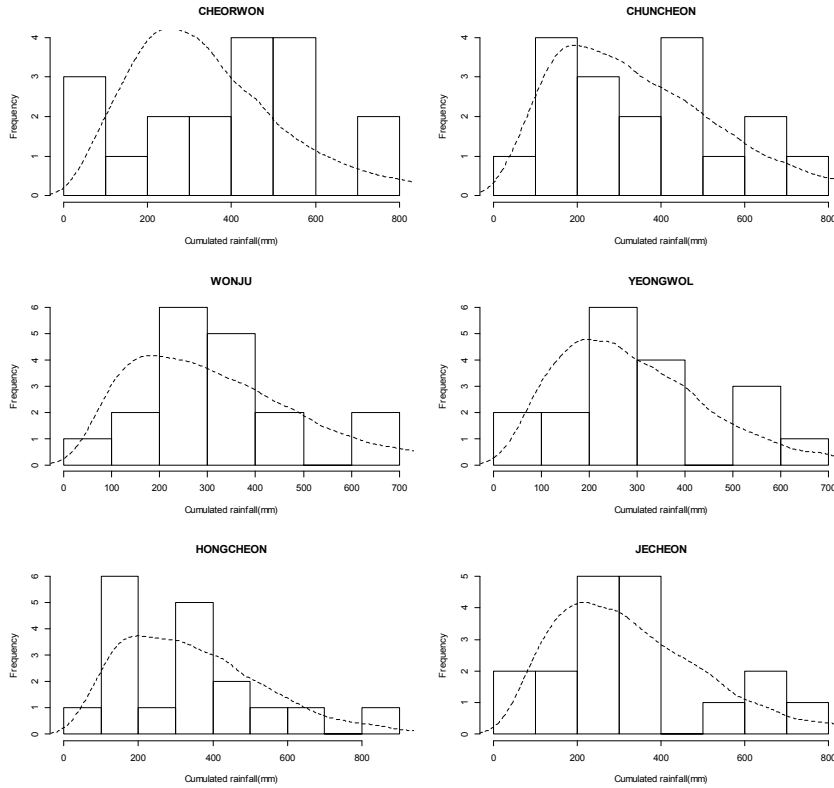


Figure 56. 한강 상류 8월 지점별 누적강우량: 관측(히스토그램)과 모의(파선)

Table 47. 한강 상류 8월 누적강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	53%	84%	95%	98%	87%	93%
K-S 검정 유의확률	17%	99%	71%	97%	74%	77%

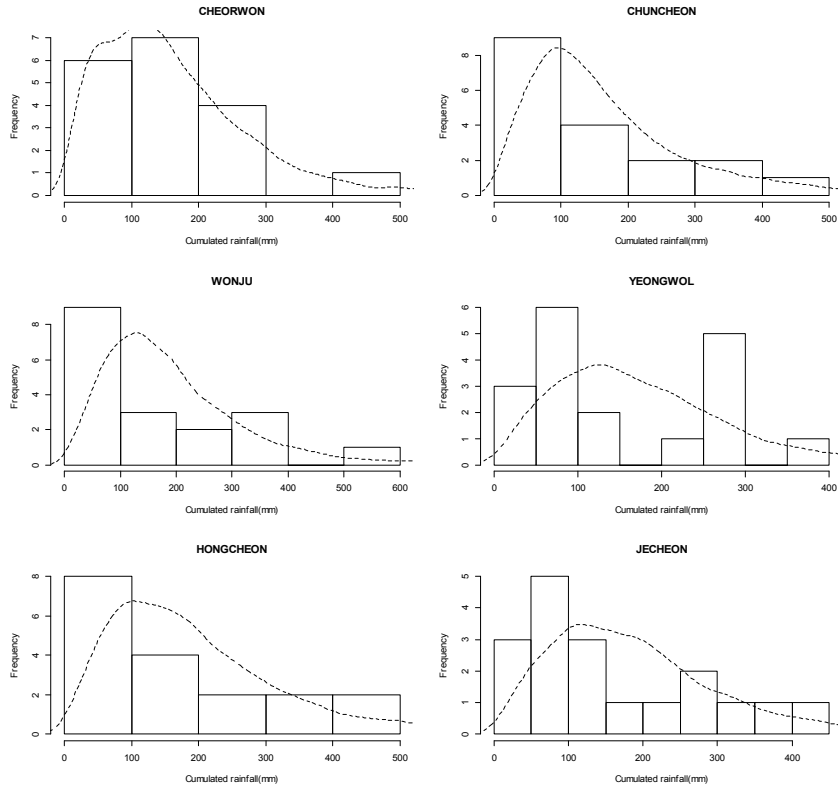


Figure 57. 한강 상류 9월 지점별 누적강우량: 관측(히스토그램)과 모의(파선)

Table 48. 한강 상류 9월 누적강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	37%	38%	54%	29%	41%	25%
K-S 검정 유의확률	25%	20%	12%	8%	27%	15%

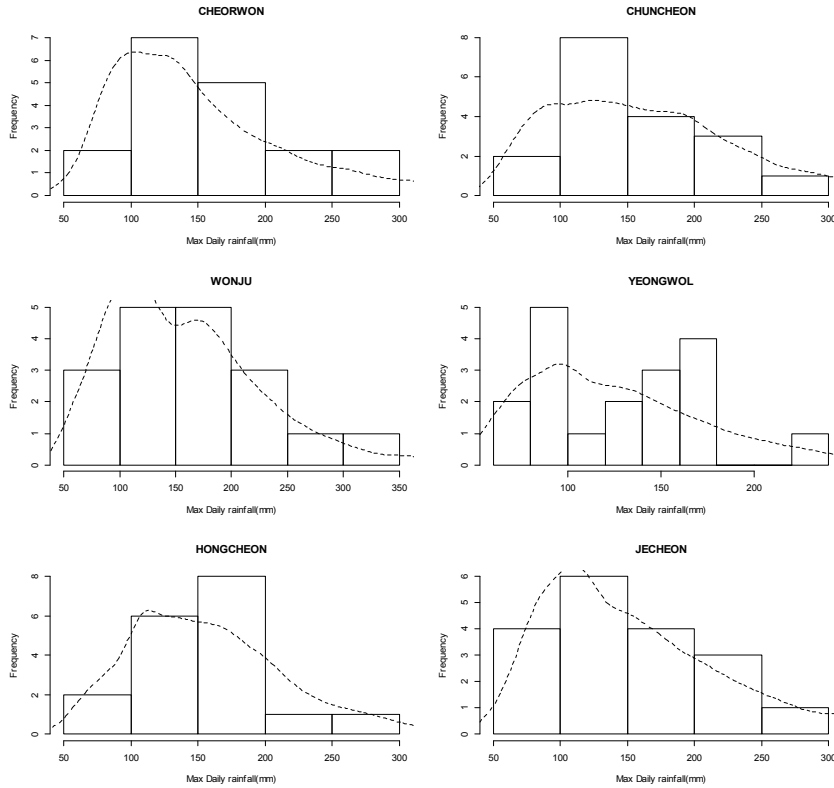


Figure 58. 한강 상류 6~9월 지점별 최대일강우량: 관측(히스토그램)과 모의(파선)

Table 49. 한강 상류 6~9월 최대일강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	25%	36%	81%	79%	35%	27%
K-S 검정 유의확률	75%	71%	82%	79%	56%	62%

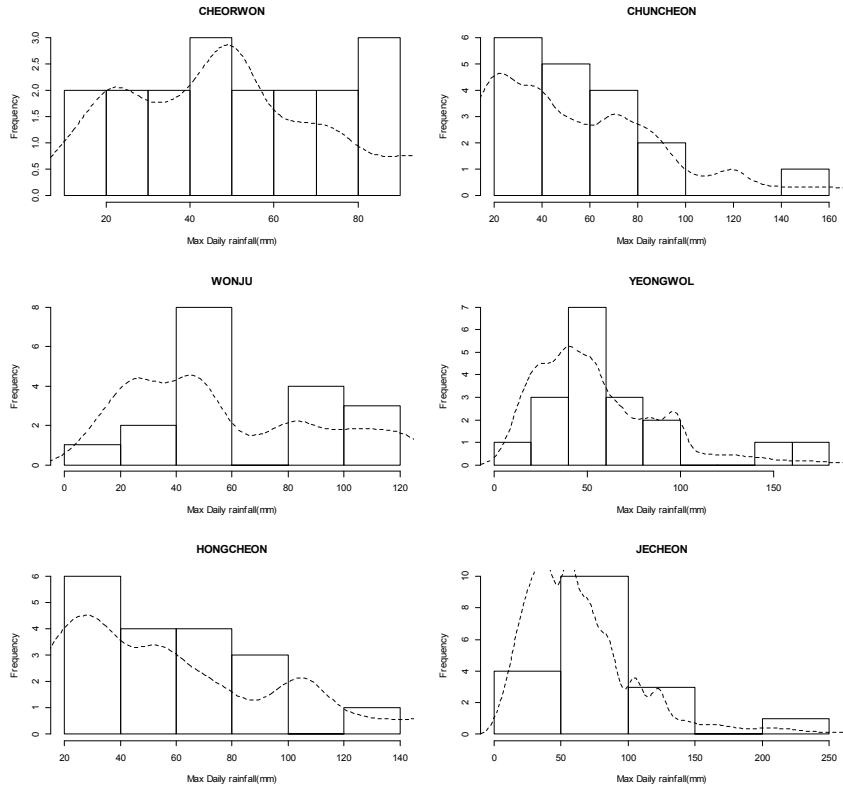


Figure 59. 한강 상류 6월 지점별 최대일강우량: 관측(히스토그램)과 모의(파선)

Table 50. 한강 상류 6월 최대일강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	8%	40%	46%	42%	14%	32%
K-S 검정 유의확률	41%	73%	53%	39%	50%	34%

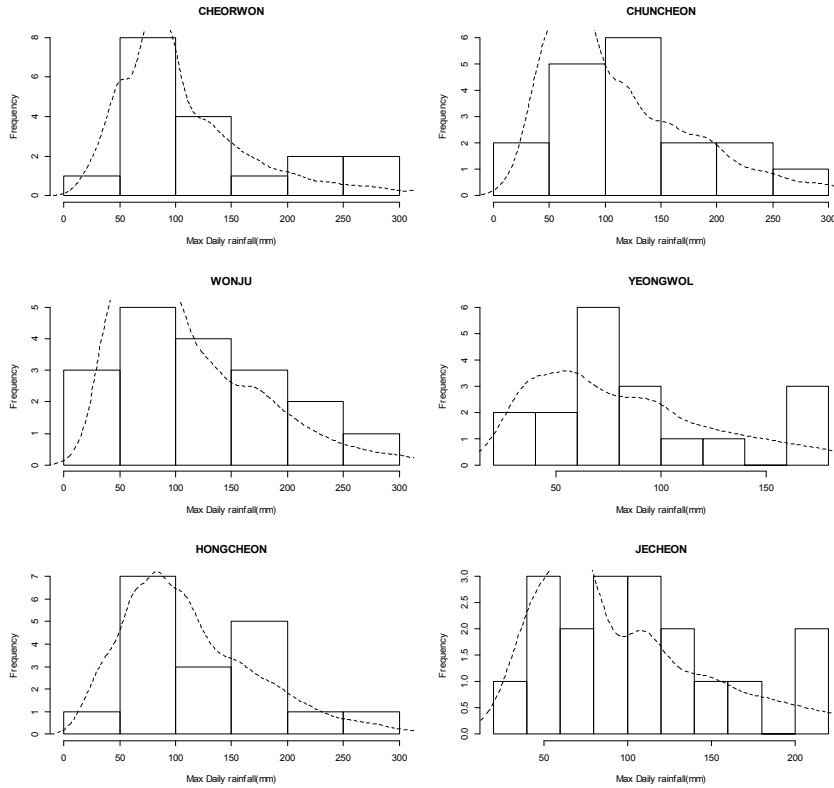


Figure 60. 한강 상류 7월 지점별 최대일강우량: 관측(히스토그램)과 모의(파선)

Table 51. 한강 상류 7월 최대일강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	40%	40%	44%	80%	56%	82%
K-S 검정 유의확률	73%	42%	49%	85%	74%	51%

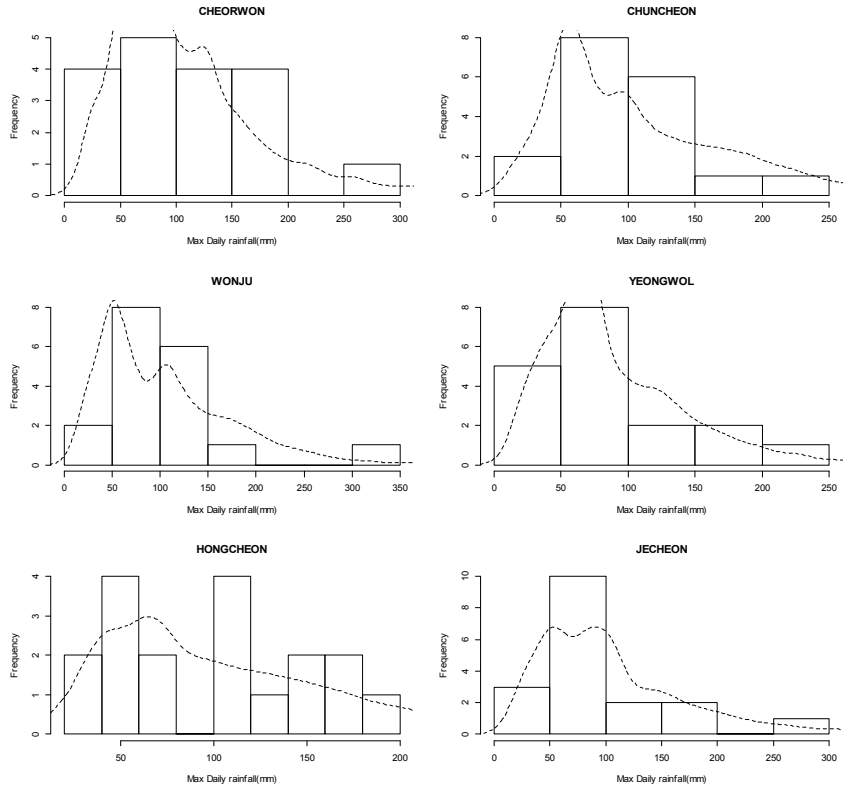


Figure 61. 한강 상류 8월 지점별 최대일강우량: 관측(히스토그램)과 모의(파선)

Table 52. 한강 상류 8월 최대일강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	93%	25%	93%	66%	56%	53%
K-S 검정 유의확률	94%	66%	74%	99%	93%	76%

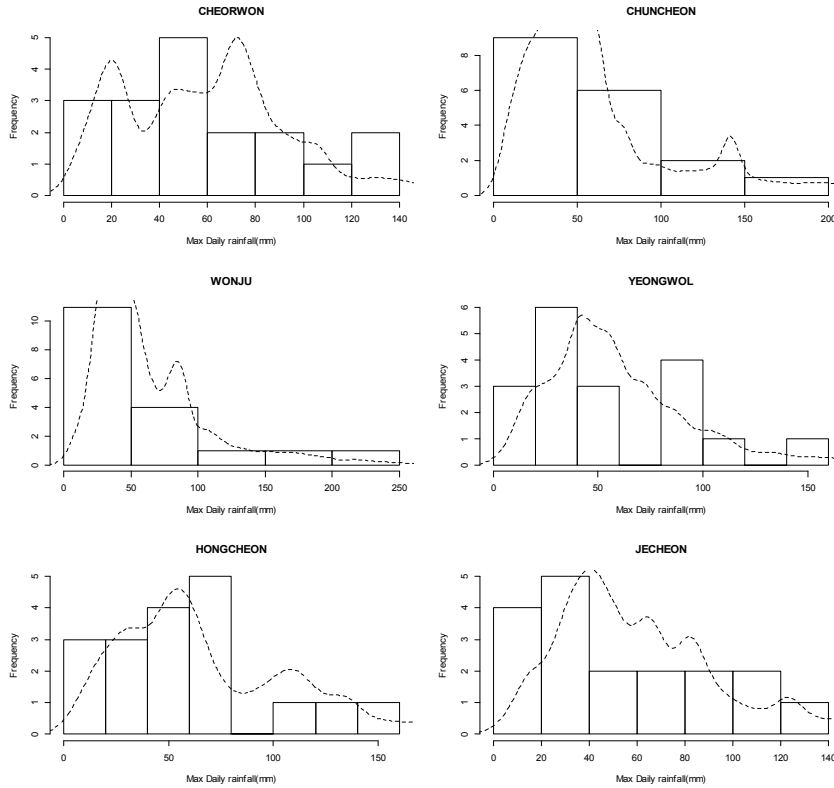


Figure 62. 한강 상류 9월 지점별 최대일강우량: 관측(히스토그램)과 모의(파선)

Table 53. 한강 상류 9월 최대일강우량 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	48%	43%	70%	51%	15%	11%
K-S 검정 유의확률	77%	95%	68%	36%	44%	14%

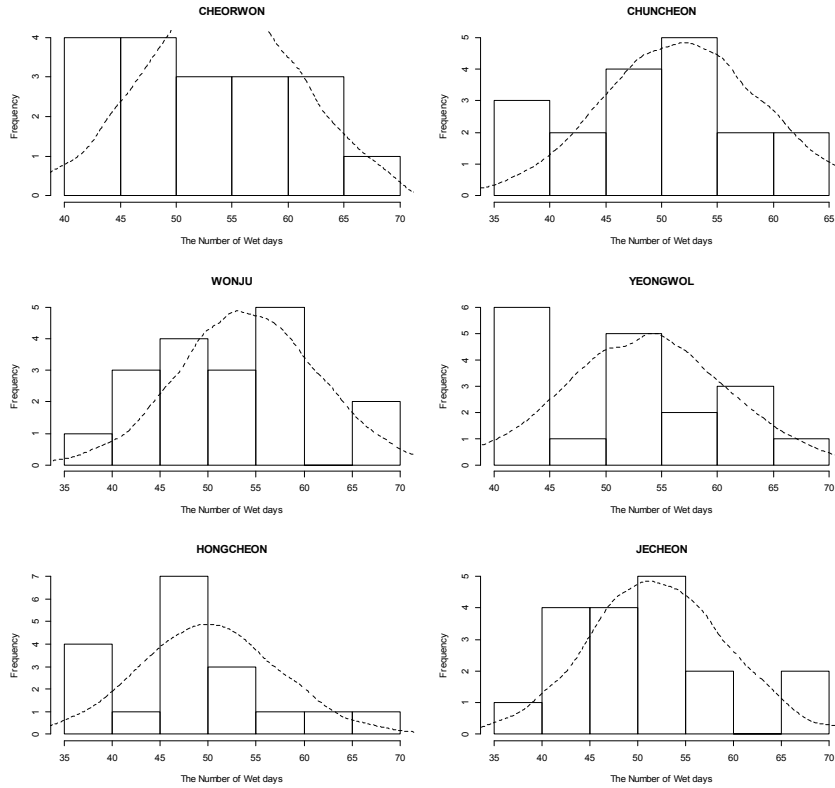


Figure 63. 한강 상류 6~9월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 54. 한강 상류 6~9월 강우일수 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	46%	31%	53%	42%	48%	58%
K-S 검정 유의확률	46%	78%	86%	48%	50%	90%

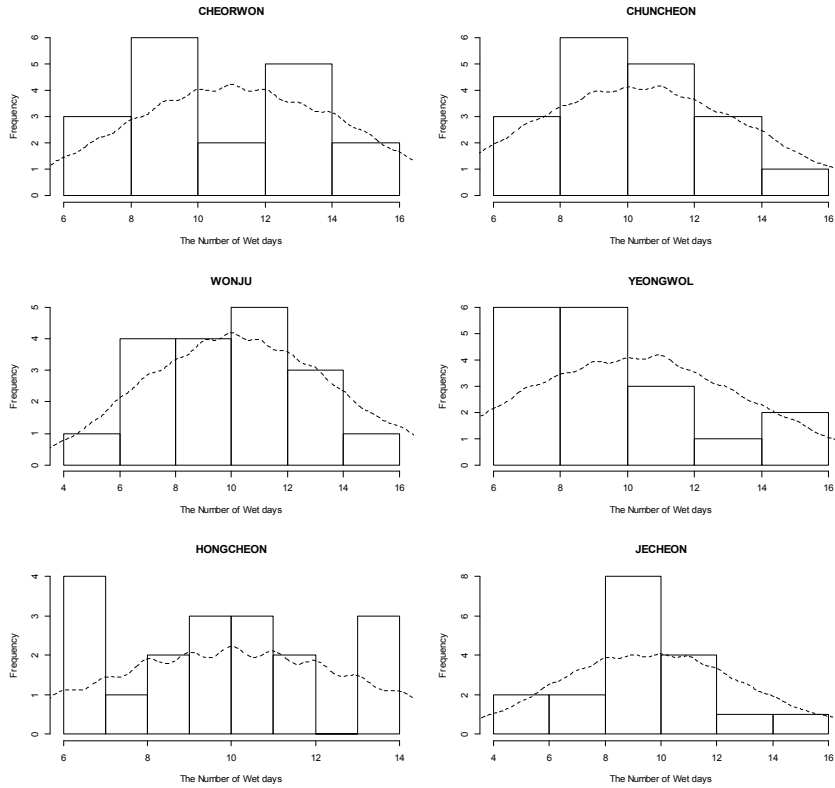


Figure 64. 한강 상류 6월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 55. 한강 상류 6월 강우일수 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	89%	84%	89%	53%	82%	93%
K-S 검정 유의확률	83%	90%	92%	57%	98%	86%

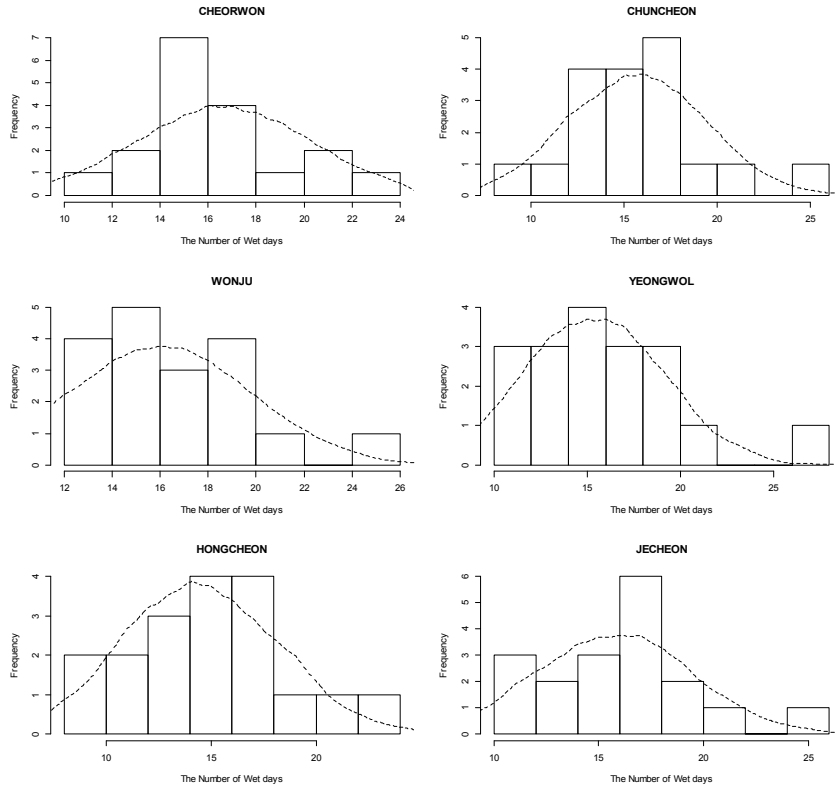


Figure 65. 한강 상류 7월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 56. 한강 상류 7월 강우일수 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	65%	54%	29%	24%	44%	32%
K-S 검정 유의확률	96%	67%	94%	99%	96%	87%

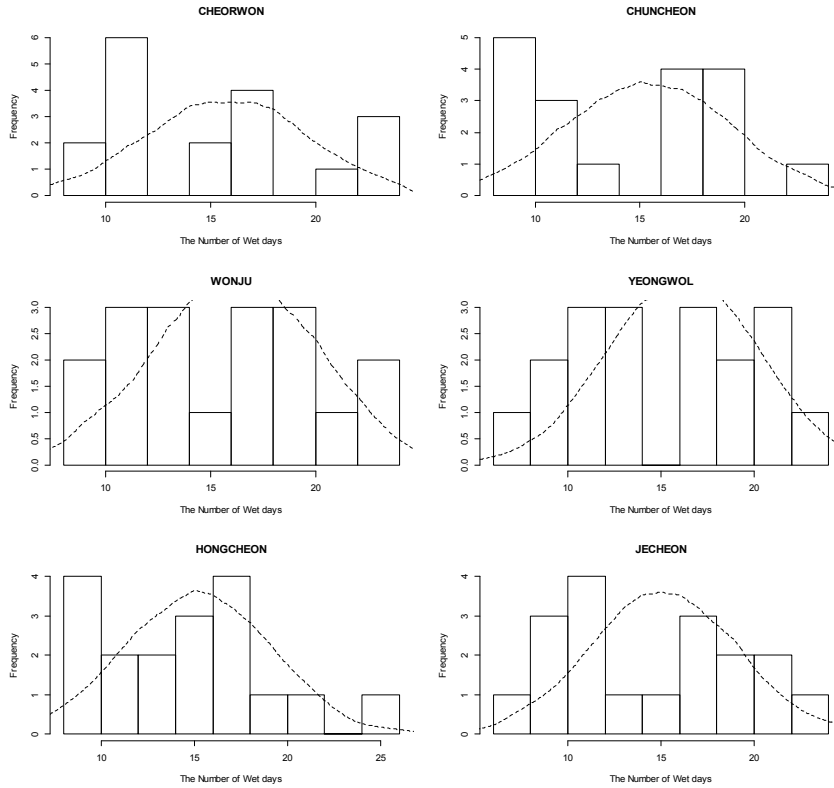


Figure 66. 한강 상류 8월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 57. 한강 상류 8월 강우일수 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	82%	48%	73%	57%	75%	71%
K-S 검정 유의확률	26%	33%	88%	76%	78%	16%

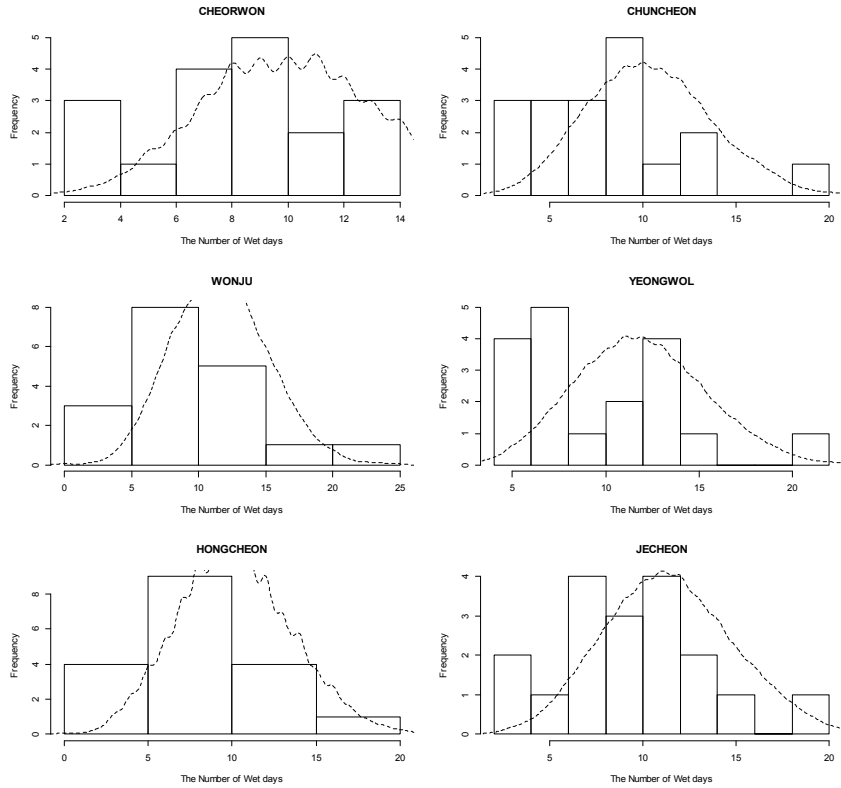


Figure 67. 한강 상류 9월 지점별 강우일수: 관측(히스토그램)과 모의(파선)

Table 58. 한강 상류 9월 강우일수 분포검정

	철원	춘천	원주	영월	홍천	제천
t-검정 유의확률	10%	7%	12%	11%	11%	12%
K-S 검정 유의확률	57%	23%	13%	4%	11%	49%

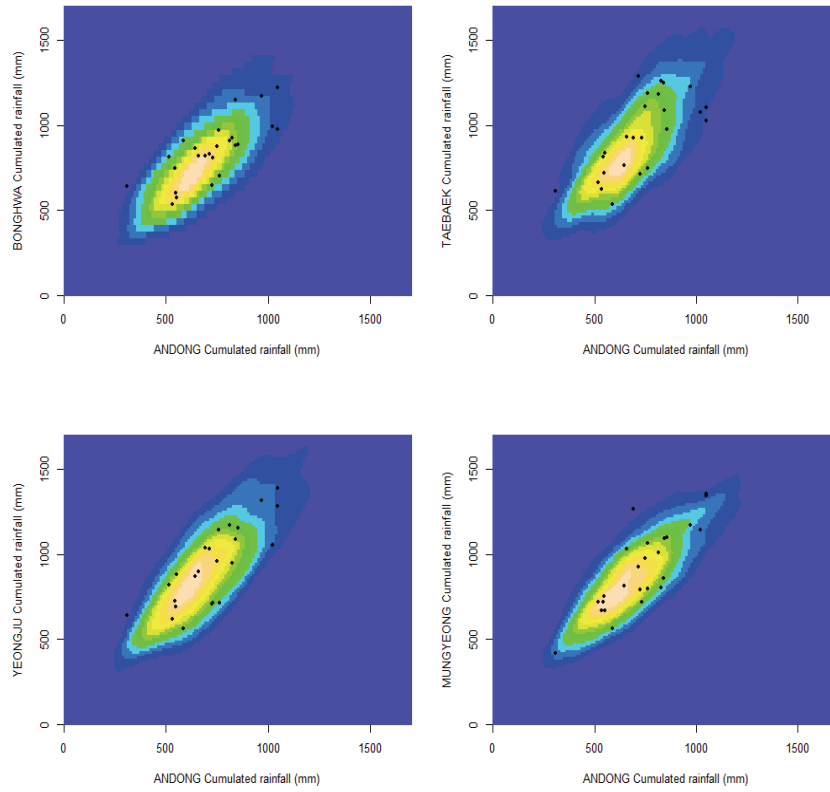


Figure 68. 낙동강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

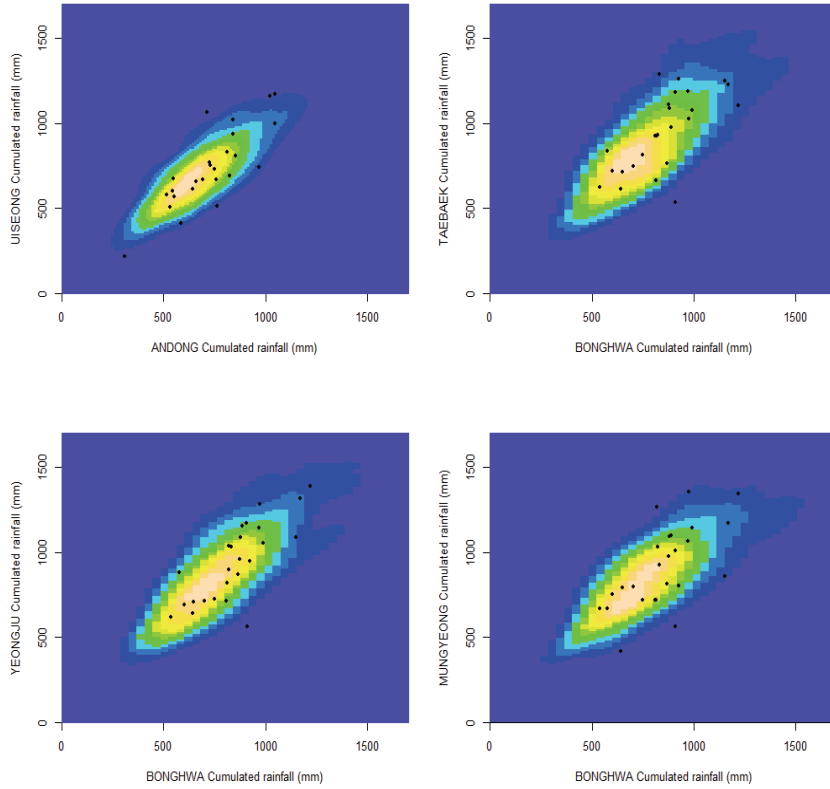


Figure 69. 낙동강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

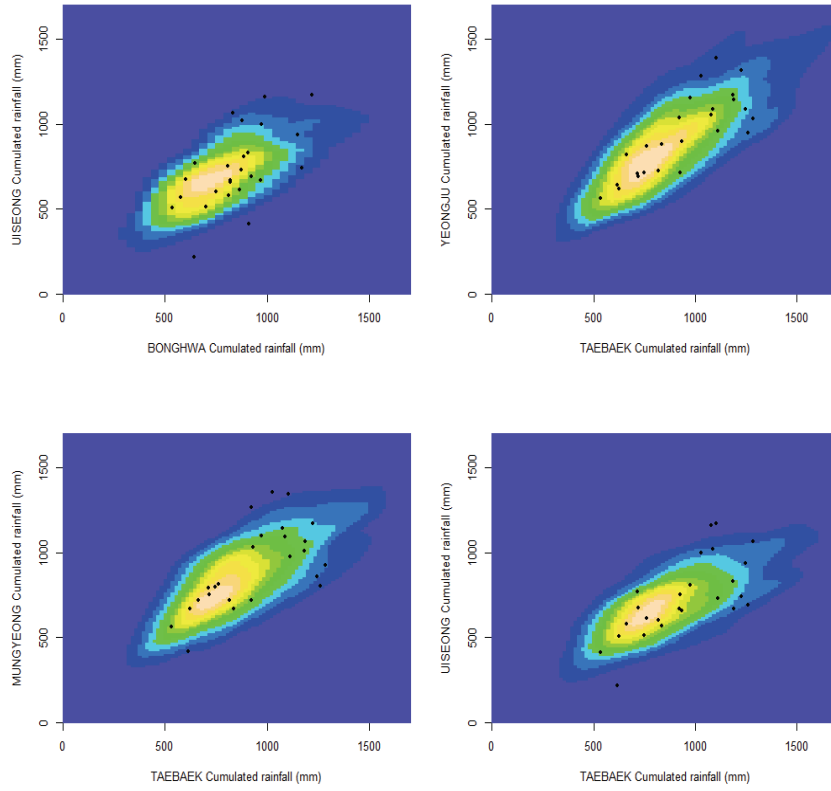


Figure 70. 낙동강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

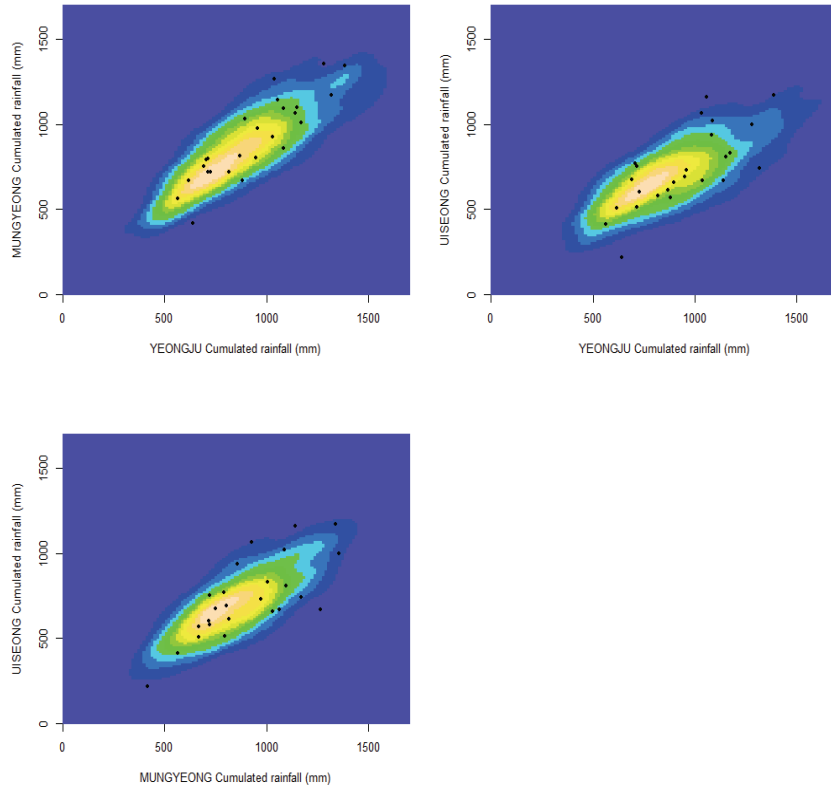


Figure 71. 낙동강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

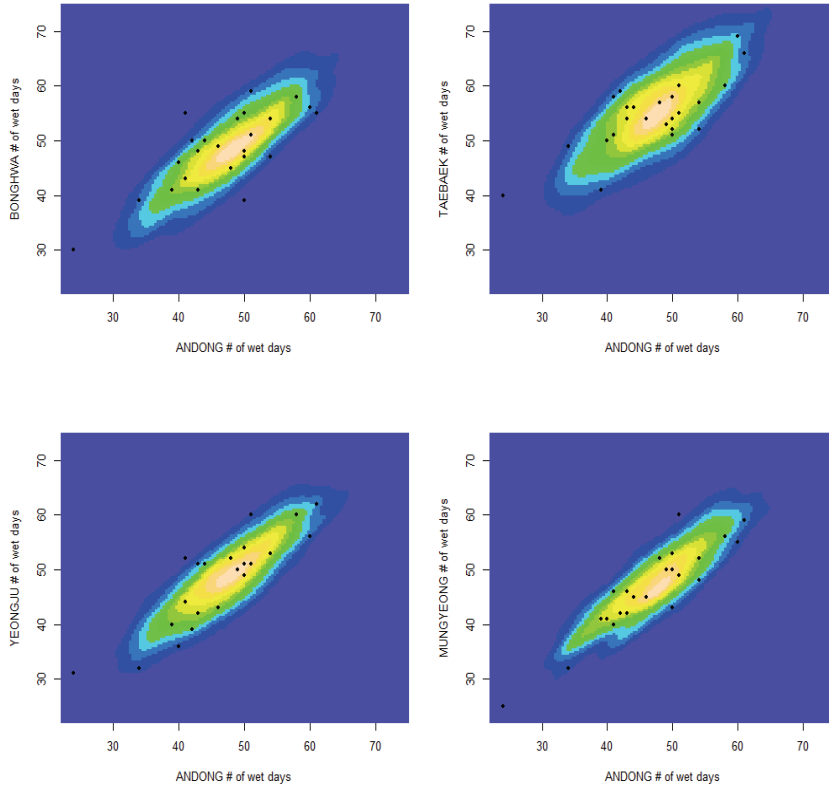


Figure 72. 낙동강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

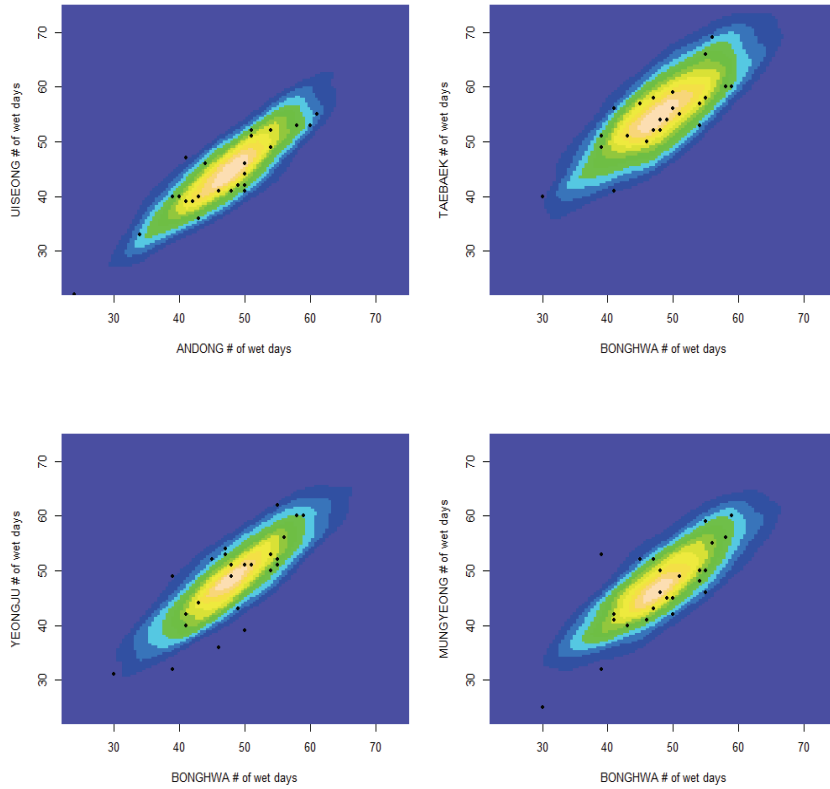


Figure 73. 낙동강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

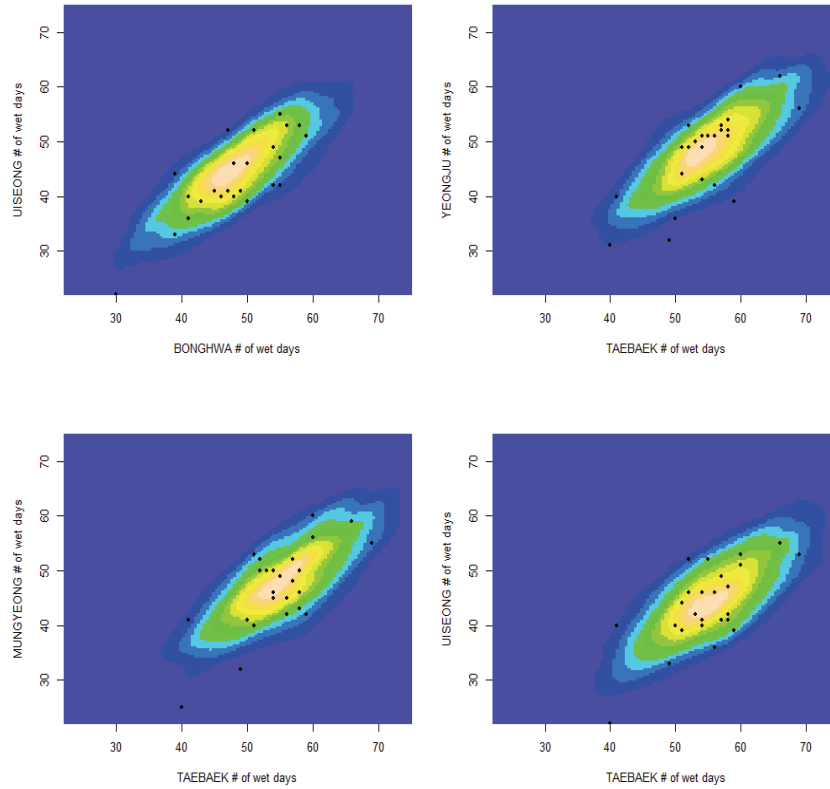


Figure 74. 낙동강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

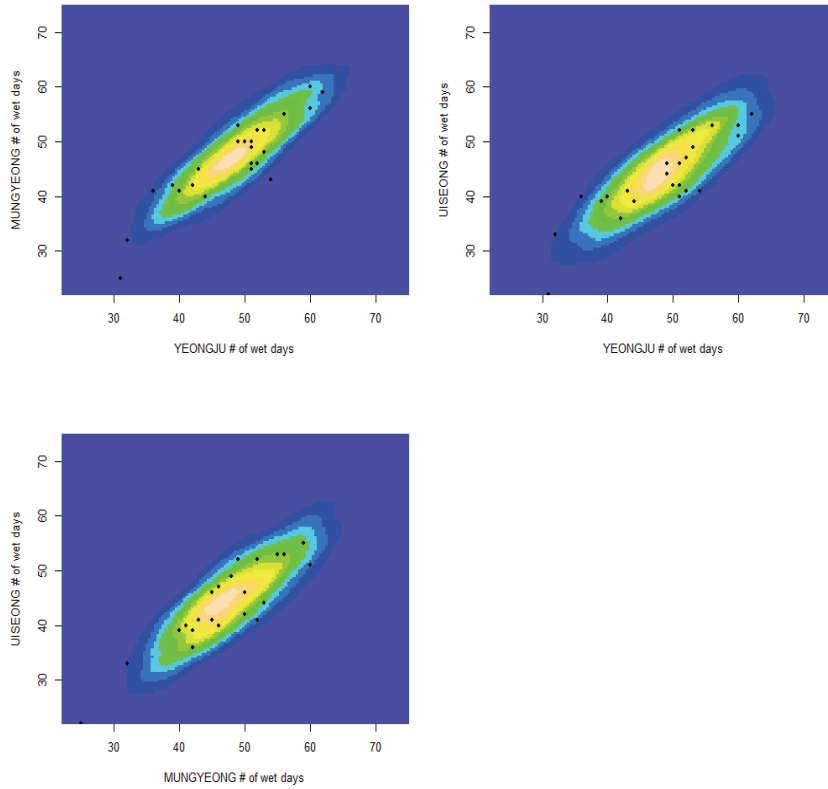


Figure 75. 낙동강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

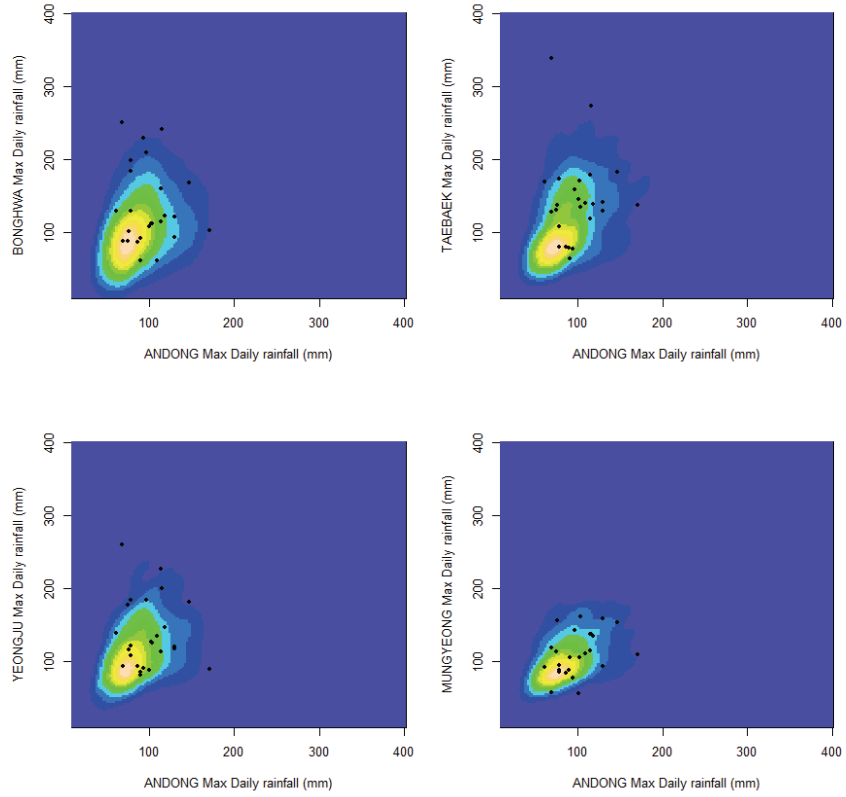


Figure 76. 낙동강 유역 6~9월 최대일강수량 연관성 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

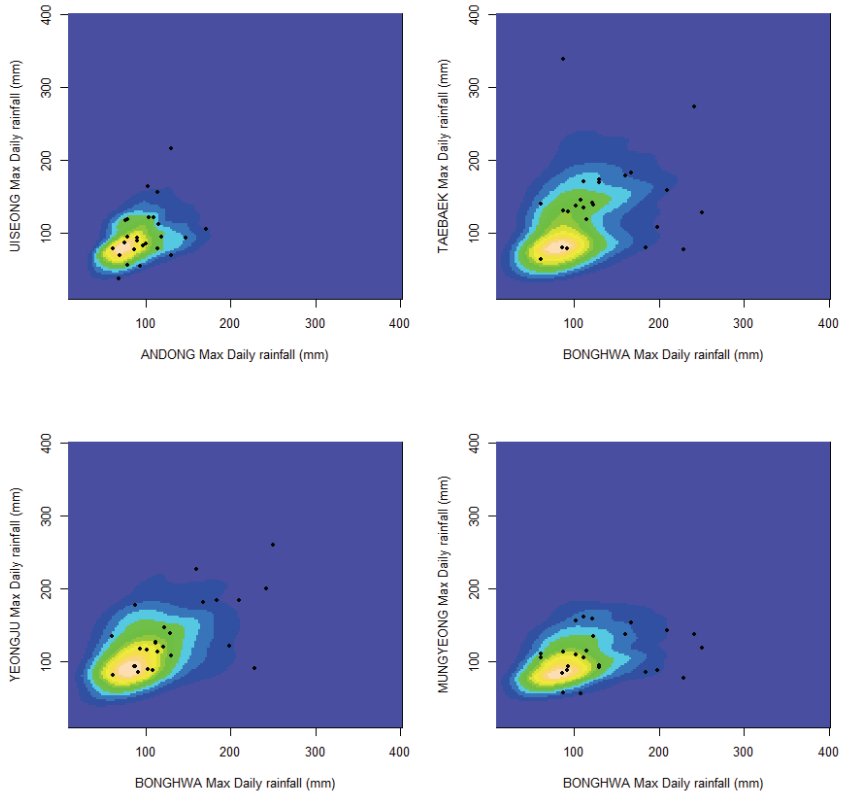


Figure 77. 낙동강 유역 6~9월 최대일강수량 연관성 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

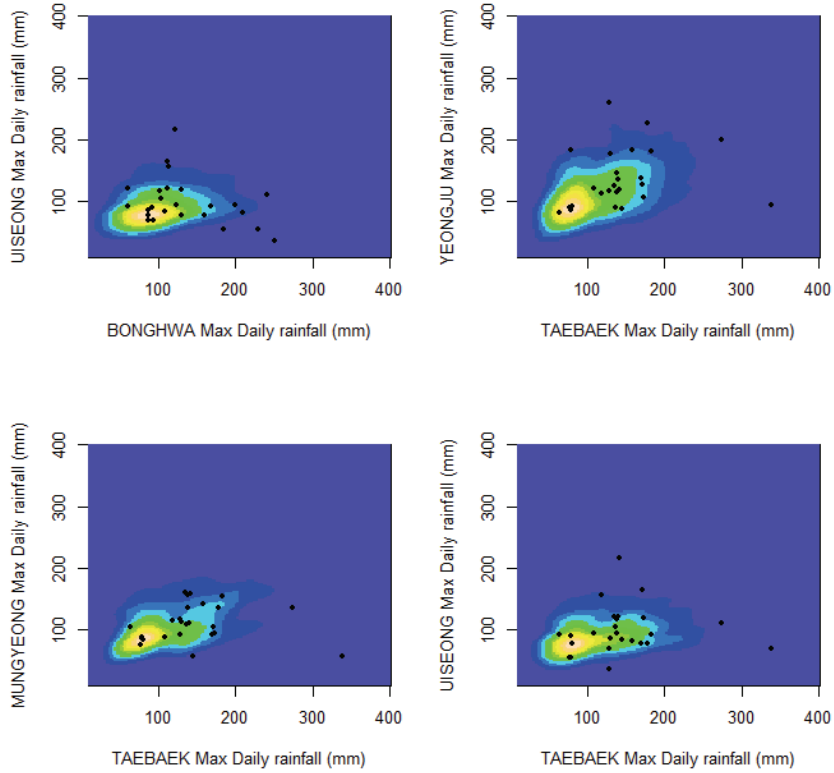


Figure 78. 낙동강 유역 6~9월 최대일강수량 연관성 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

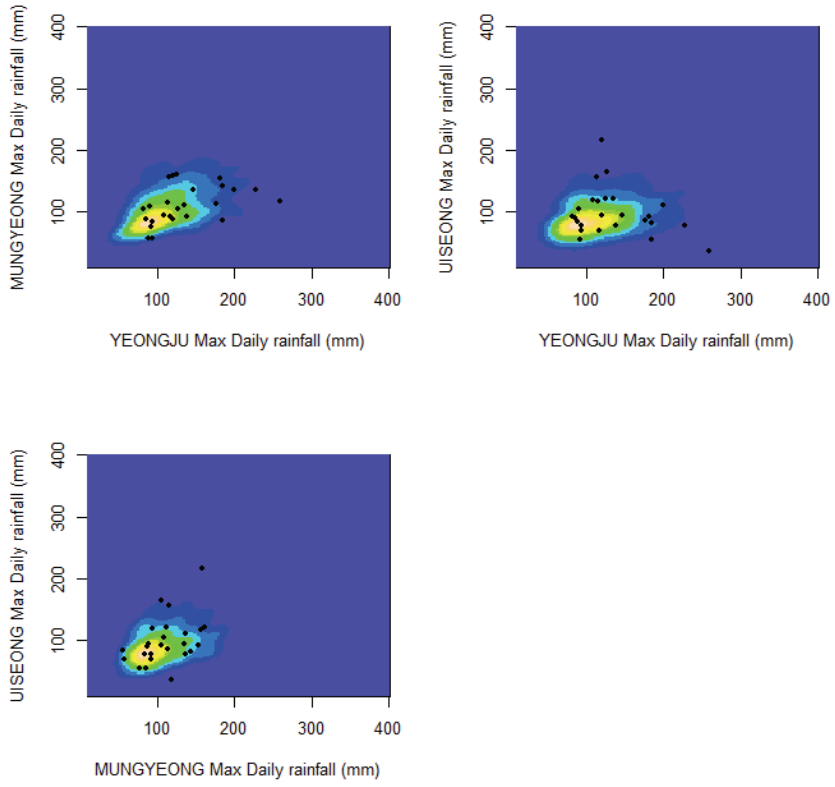


Figure 79. 낙동강 유역 6~9월 최대일강수량 연관성 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

Table 59. 낙동강 유역 6~9월 누적강수량 관측과 모의의 Pearson 상관계수: 모두 쌍에서 차이가 5% 수준에서 유의하지 않다.

관측 / 모의						
	안동	봉화	태백	영주	문경	의성
안동	1	0.76 / 0.62	0.70 / 0.77	0.81 / 0.81	0.83 / 0.81	0.82 / 0.82
봉화		1	0.68 / 0.66	0.78 / 0.64	0.65 / 0.61	0.60 / 0.60
태백			1	0.80 / 0.79	0.64 / 0.76	0.60 / 0.73
영주				1	0.88 / 0.83	0.73 / 0.77
문경					1	0.74 / 0.78
의성						1

Table 60. 낙동강 유역 6~9월 강우일수 관측과 모의의 Pearson 상관계수: 모두 쌍에서 차이가 5% 수준에서 유의하지 않다.

관측 / 모의						
	안동	봉화	태백	영주	문경	의성
안동	1	0.74 / 0.84	0.76 / 0.77	0.87 / 0.86	0.91 / 0.86	0.90 / 0.87
봉화		1	0.79 / 0.80	0.78 / 0.86	0.76 / 0.83	0.80 / 0.80
태백			1	0.74 / 0.79	0.70 / 0.77	0.69 / 0.75
영주				1	0.90 / 0.86	0.85 / 0.81
문경					1	0.88 / 0.83
의성						1

Table 61. 낙동강 유역 6~9월 최대일강수량 관측과 모의의 Spearman 상관계수. 음영영역에서 유의한 차이가 있을 것이라 짐작된다.

관측 / 모의						
	안동	봉화	태백	영주	문경	의성
안동	1	0.02 / 0.43	0.20 / 0.42	0.00 / 0.43	0.41 / 0.42	0.43 / 0.42
봉화		1	0.17 / 0.49	0.57 / 0.49	0.22 / 0.46	-0.18 / 0.40
태백			1	0.38 / 0.48	0.29 / 0.45	0.24 / 0.37
영주				1	0.50 / 0.46	-0.11 / 0.38
문경					1	0.51 / 0.38
의성						1

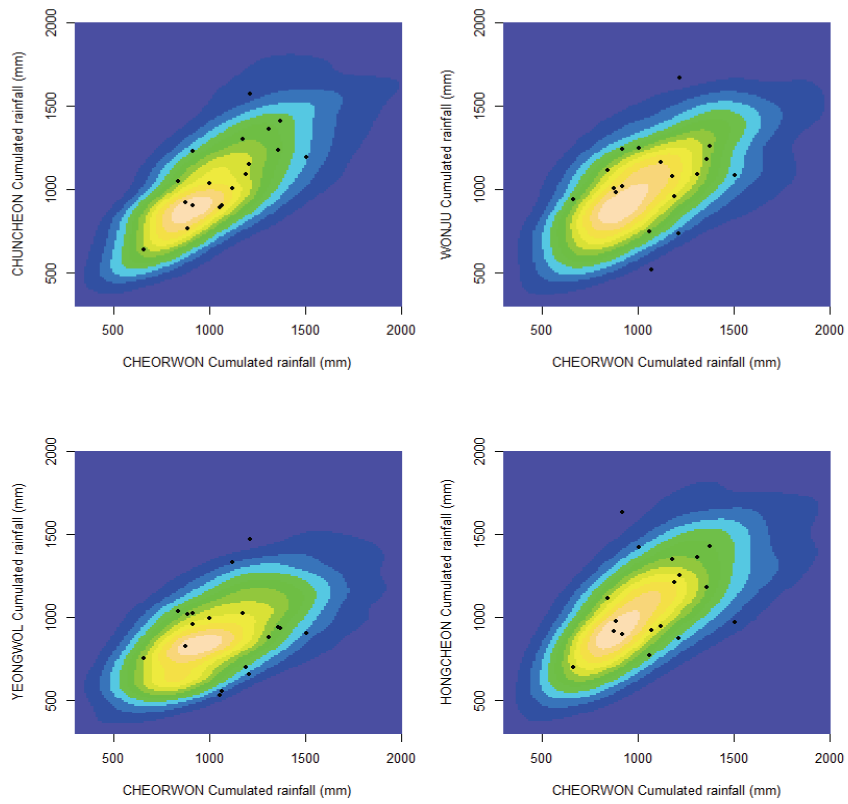


Figure 80. 한강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

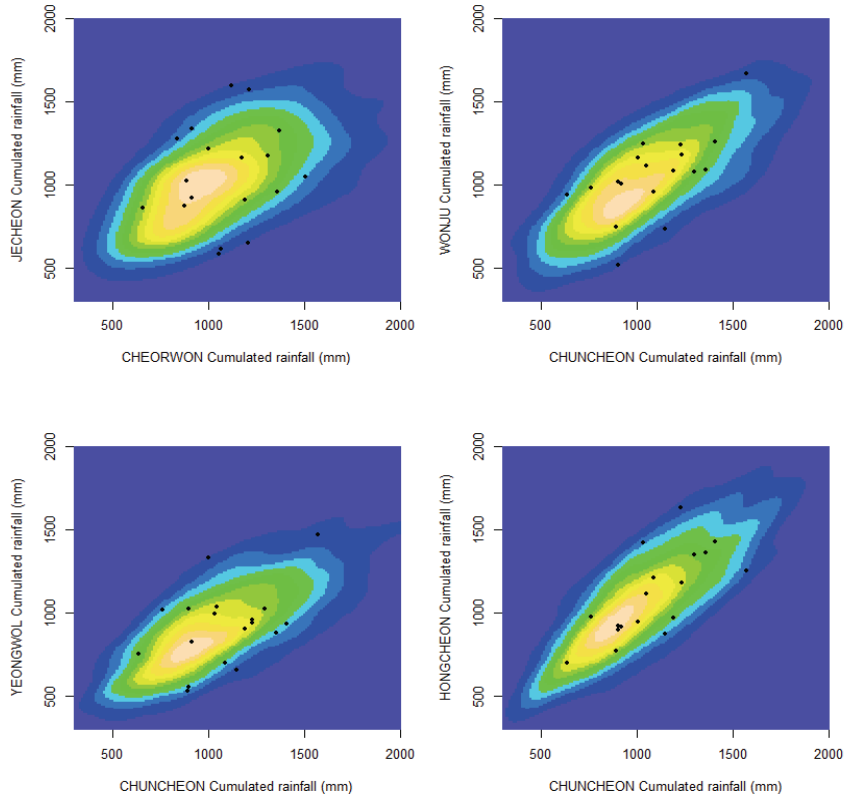


Figure 81. 한강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

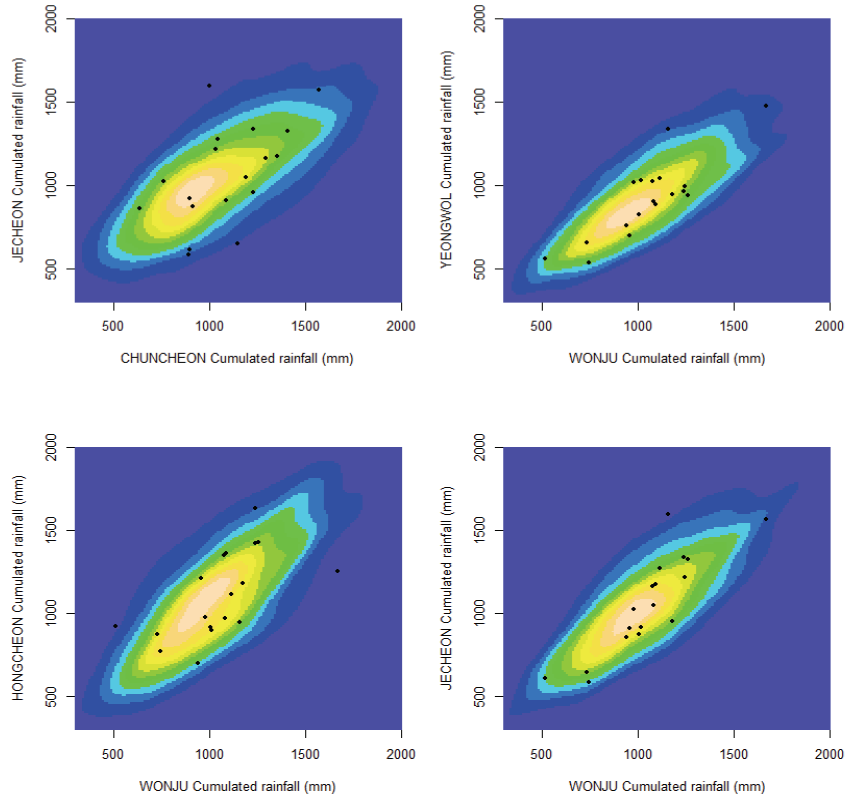


Figure 82. 한강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

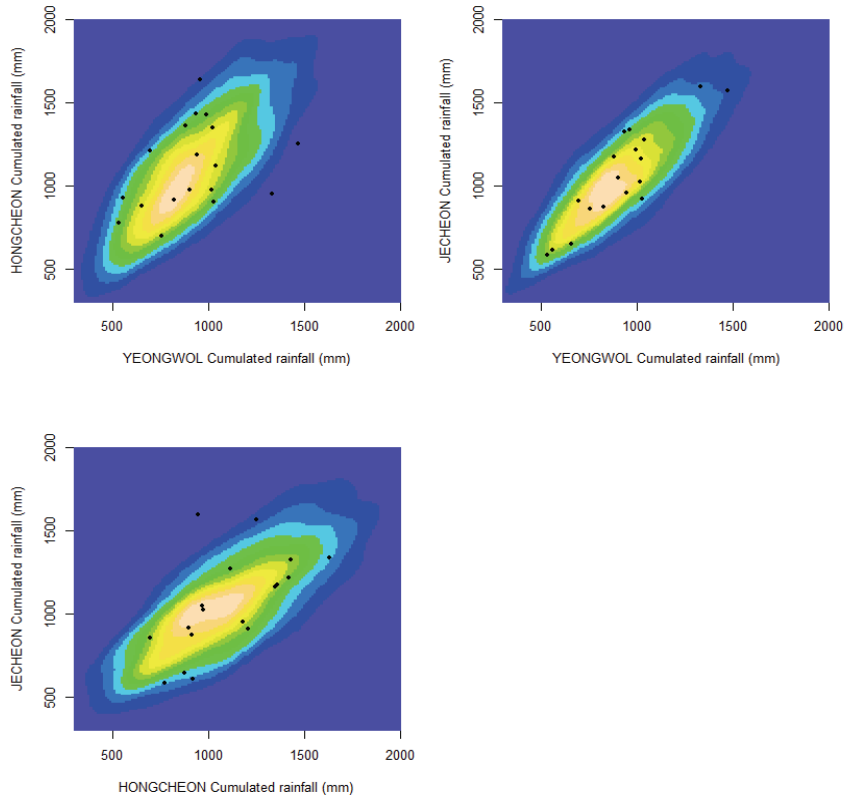


Figure 83. 한강 유역 6~9월 누적강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

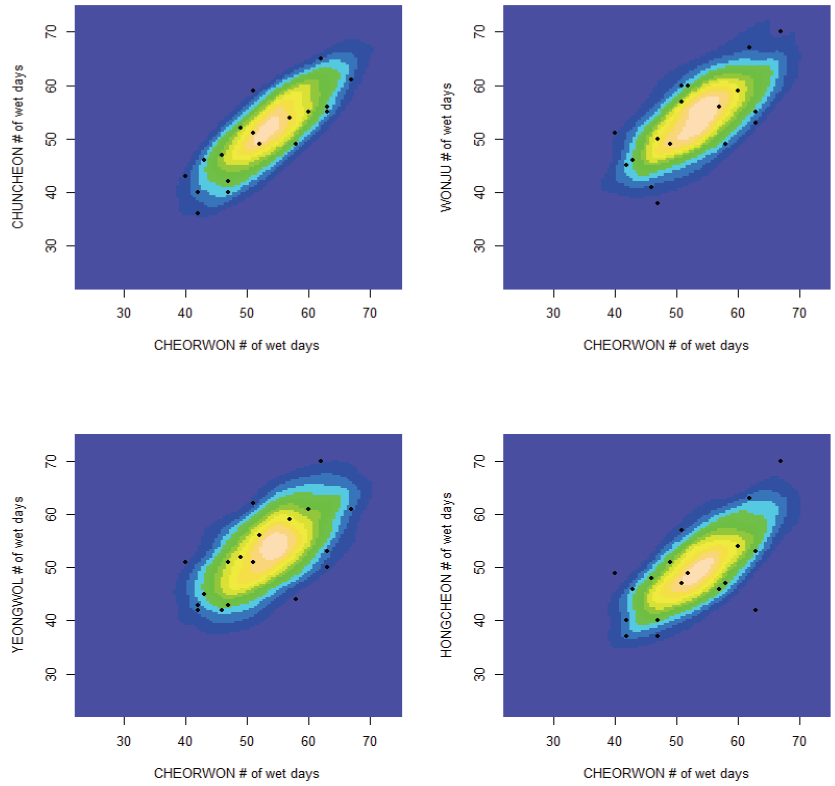


Figure 84. 한강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

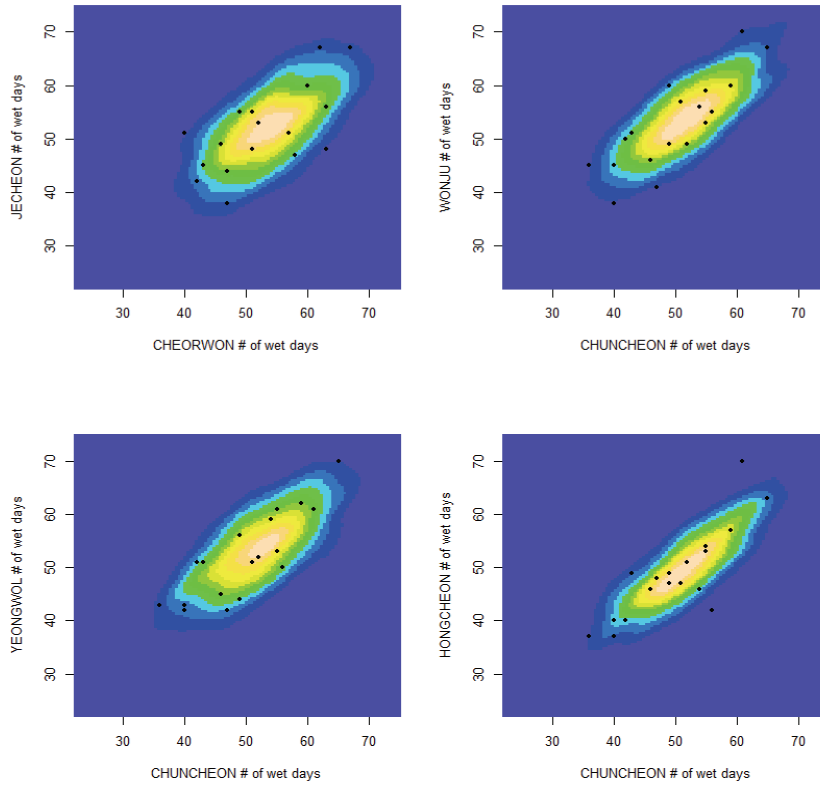


Figure 85. 한강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

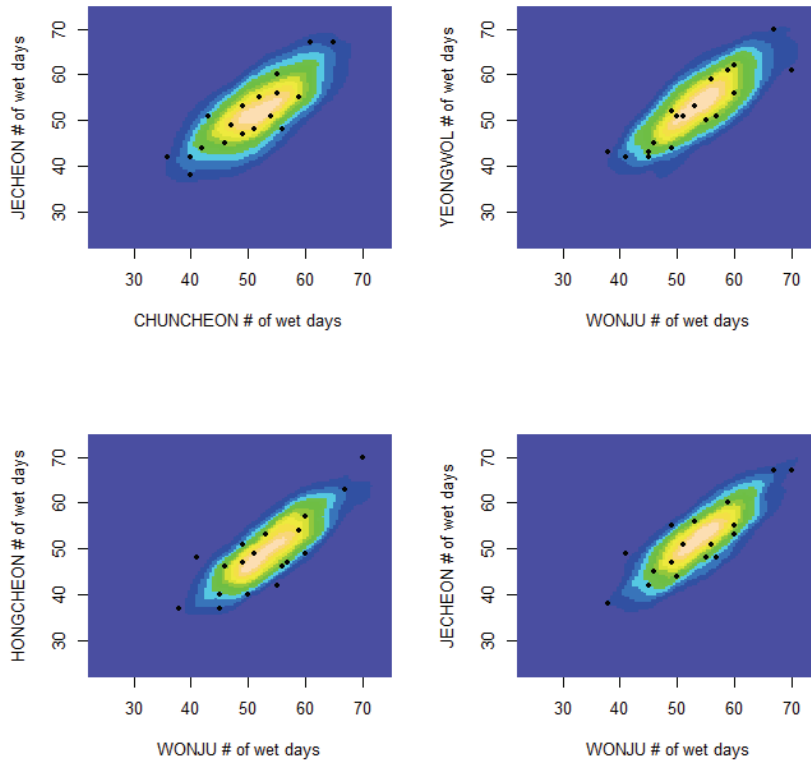


Figure 86. 한강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

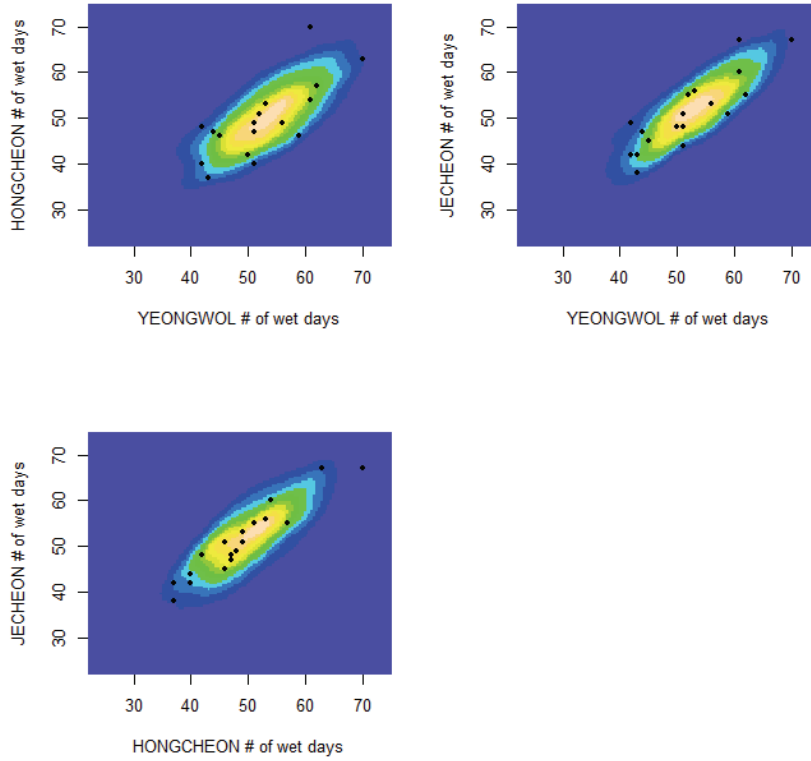


Figure 87. 한강 유역 6~9월 강우일수 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

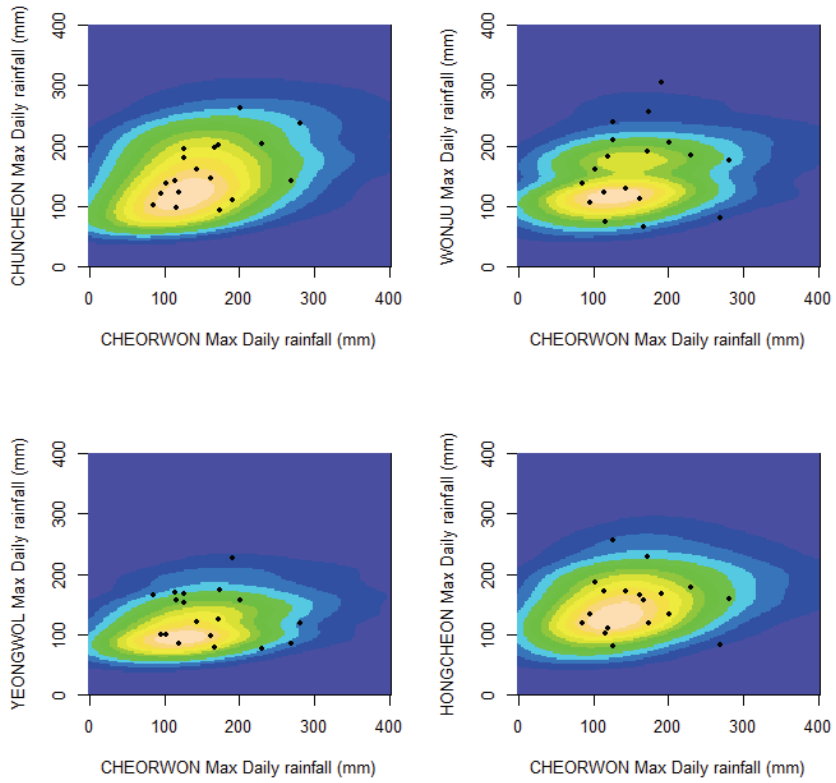


Figure 88. 한강 유역 6~9월 최대일강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

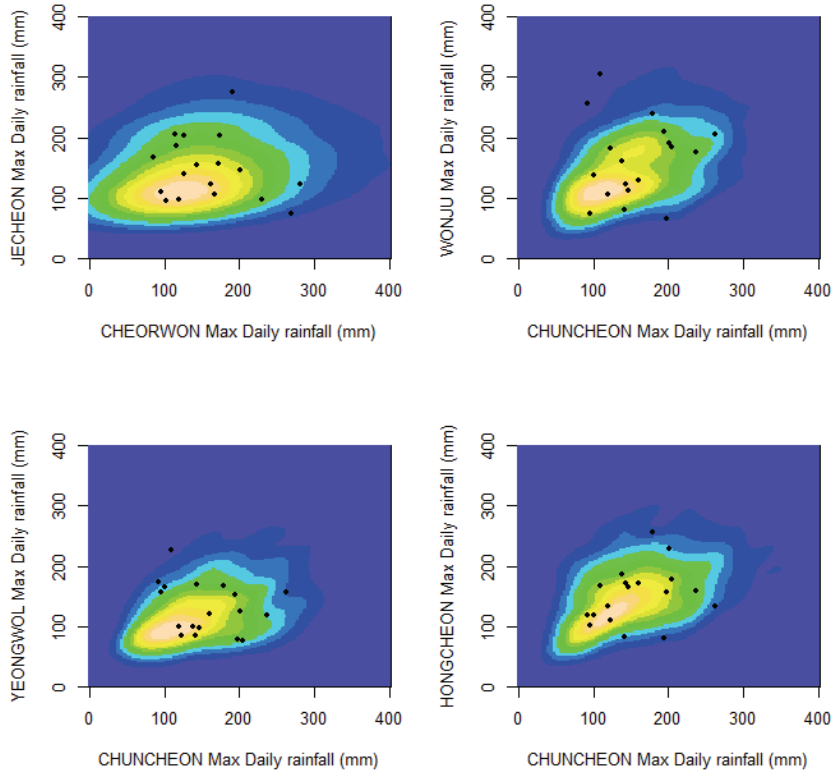


Figure 89. 한강 유역 6~9월 최대일강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

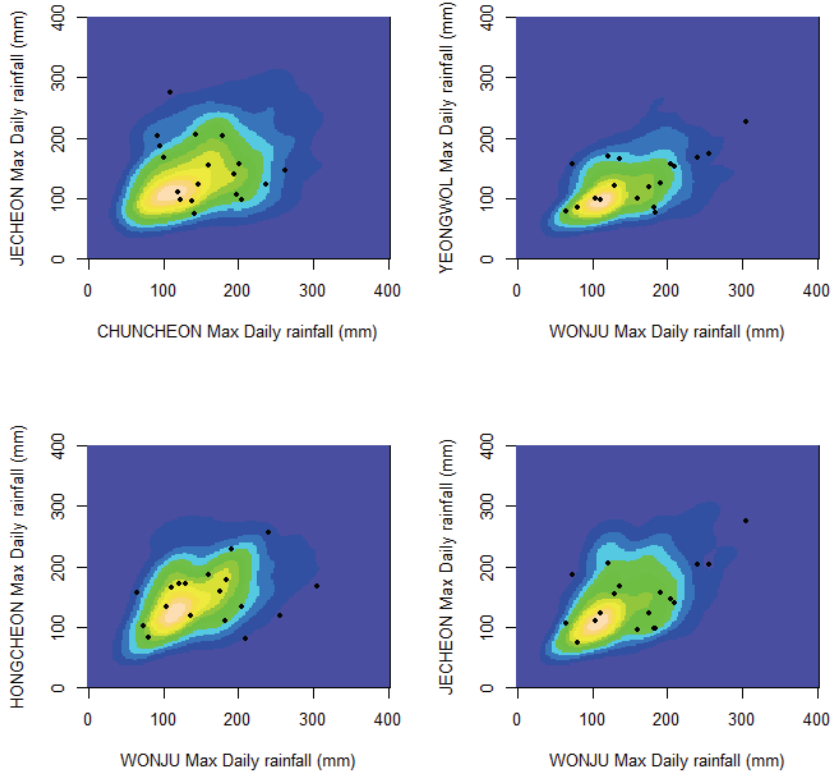


Figure 90. 한강 유역 6~9월 최대일강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림(topographical color)

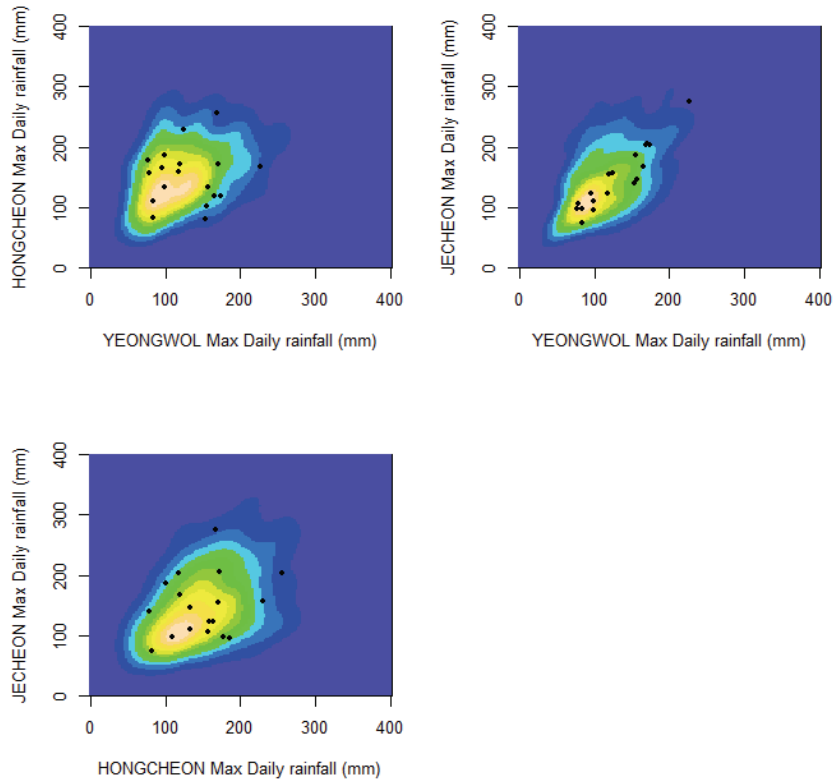


Figure 91. 한강 유역 6~9월 최대일강수량 상관관계 비교 - 관측자료(black dot)와 모의자료의 등고선 그림 (topographical color)

Table 62. 한강 유역 6~9월 누적강수량 관측과 모의의 Spearman 상관계수. 음영영역에서 유의한 차이가 있을 것으로 짐작된다.

관측 / 모의						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.74 / 0.77	0.31 / 0.71	-0.09 / 0.69	0.37 / 0.75	0.21 / 0.70
춘천		1	0.66 / 0.76	0.29 / 0.73	0.73 / 0.80	0.58 / 0.74
원주			1	0.69 / 0.80	0.75 / 0.80	0.89 / 0.80
영월				1	0.39 / 0.76	0.80 / 0.83
홍천					1	0.72 / 0.77
제천						1

Table 63. 한강 유역 6~9월 강우일수 관측과 모의의 Pearson 상관계수. 음영영역에서 5% 수준에서 차이가 유의하다.

관측 / 모의						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.83 / 0.84	0.71 / 0.74	0.62 / 0.68	0.62 / 0.80	0.69 / 0.70
춘천		1	0.83 / 0.79	0.83 / 0.75	0.83 / 0.87	0.87 / 0.76
원주			1	0.90 / 0.83	0.81 / 0.84	0.85 / 0.86
영월				1	0.77 / 0.78	0.86 / 0.87
홍천					1	0.95 / 0.81
제천						1

Table 64. 한강 유역 6~9월 최대일강수량 관측과 모의의 Spearman 상관계수. 음영영역에서 유의한 차이가 있는 것으로 짐작된다.

관측 / 모의						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.51 / 0.47	0.28 / 0.46	-0.14 / 0.48	0.05 / 0.43	-0.13 / 0.45
춘천		1	0.12 / 0.46	-0.33 / 0.43	0.38 / 0.46	-0.25 / 0.44
원주			1	0.52 / 0.49	0.22 / 0.46	0.41 / 0.52
영월				1	0.07 / 0.41	0.91 / 0.55
홍천					1	0.19 / 0.44
제천						1

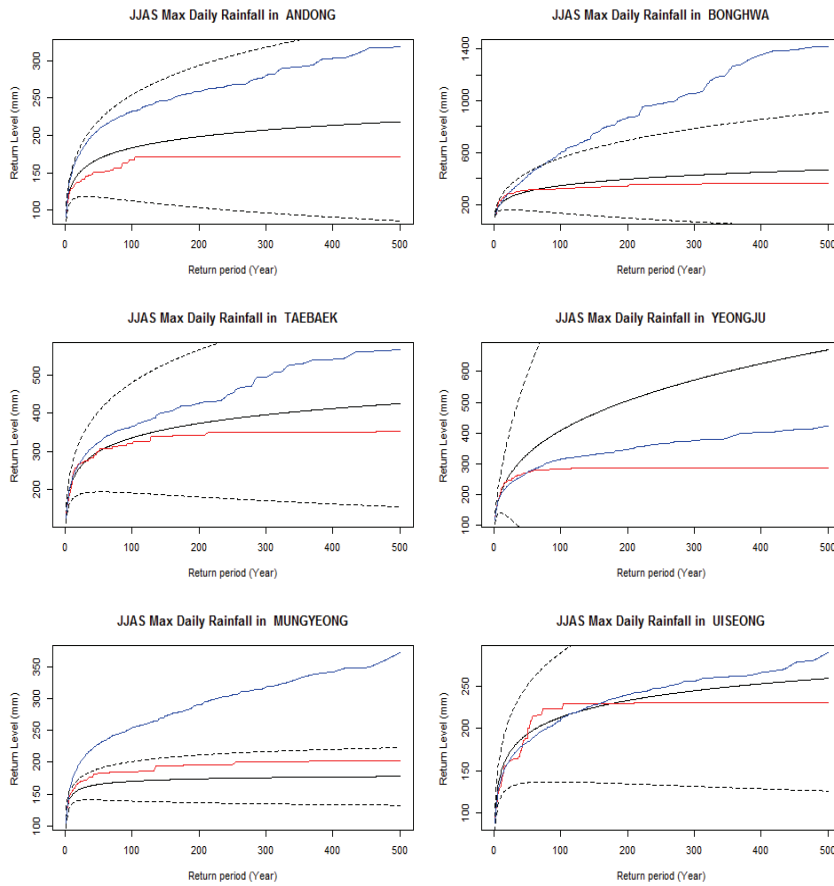


Figure 92. 낙동강 상류유역 호우기 최대일강수량에 대한 return level의 그래프.

- 제안된 강우알고리즘(파란실선), 대조알고리즘(빨간실선), GEV 모형(검은실선)의 return level. 파선은 GEV 모형에서 얻어진 return level의 95% 오차범위를 나타내는 기준선이다.
- 대조알고리즘은 관측자료의 유한성으로 인해 어느 return period 이상에서는 return level이 더 이상 증가하지 않는다. 한편, 봉화, 문경 지점에서 제안된 강우알고리즘은 GEV 모형 기준으로 최대일강수량의 return level을 과대추정하는 것으로 판단된다.

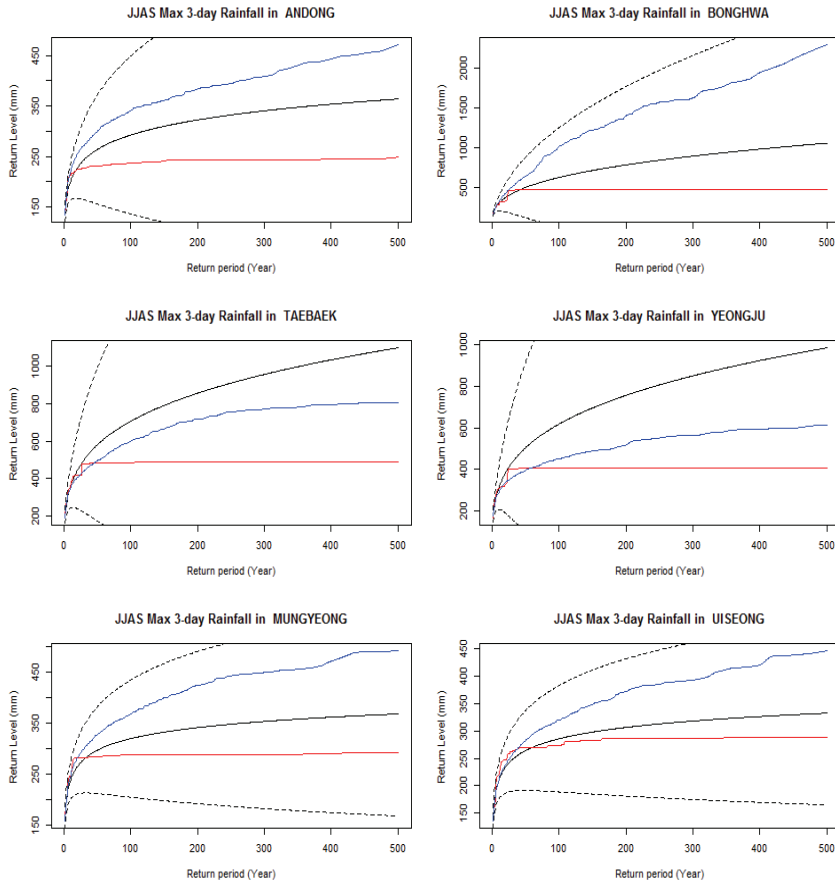


Figure 93. 낙동강 상류유역 호우기 최대일강수량에 대한 return level의 그래프.

- 제안된 강우알고리즘(파란실선), 대조알고리즘(빨간실선), GEV 모형(검은실선)의 return level. 파선은 GEV 모형에서 얻어진 return level의 95% 오차범위를 나타내는 기준선이다.
- 3일연속누적강수량을 고려한 이유는 호우사상의 지속시간이 평균적으로 2일이 넘기 때문이다.
- 대조알고리즘의 return level이 어느 return period 이상에서 더 이상 증가하지 않는다. 한편, 제안된 강우알고리즘은 GEV 모형 return level의 오차범위 내에서 값을 제시한다.

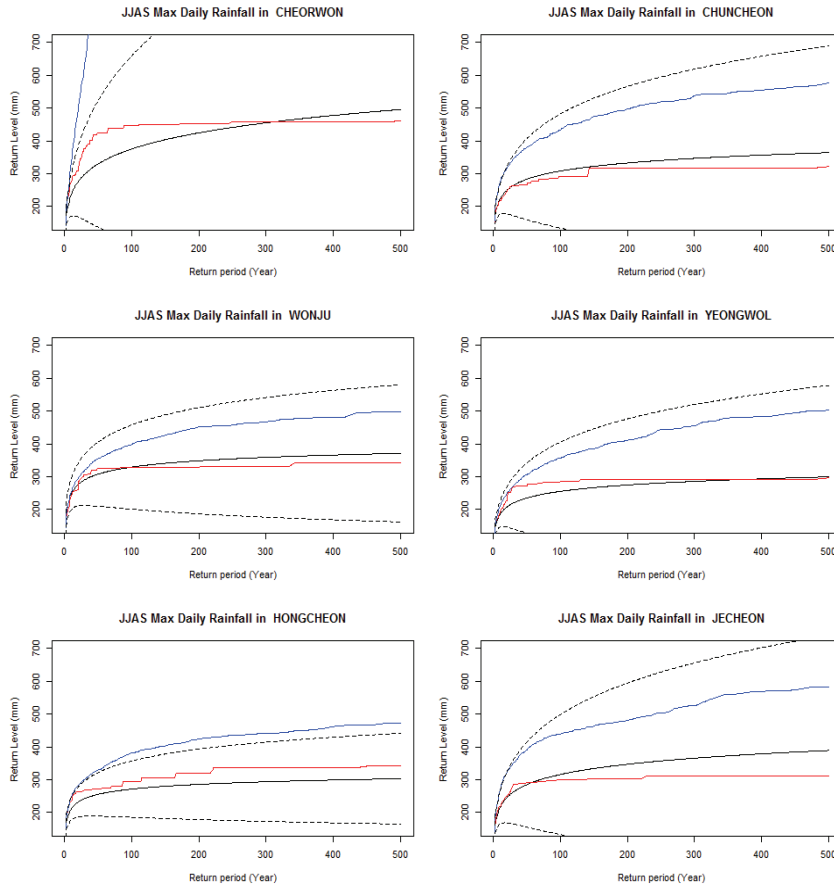


Figure 94. 한강 상류유역 호우기 최대일강수량에 대한 return level의 그래프.

- a. 제안된 강우알고리즘(파란실선), 대조알고리즘(빨간실선), GEV 모형(검은실선)의 return level. 파선은 GEV 모형에서 얻어진 return level의 95% 오차범위를 나타내는 기준선이다.
- b. 대조알고리즘은 관측자료의 유한성으로 인해 어느 return period 이상에서는 return level이 더 이상 증가하지 않는다. 한편, 제안된 강우알고리즘은 GEV 모형 기준으로 최대일강수량의 return level을 과대추정하는 것으로 판단된다.

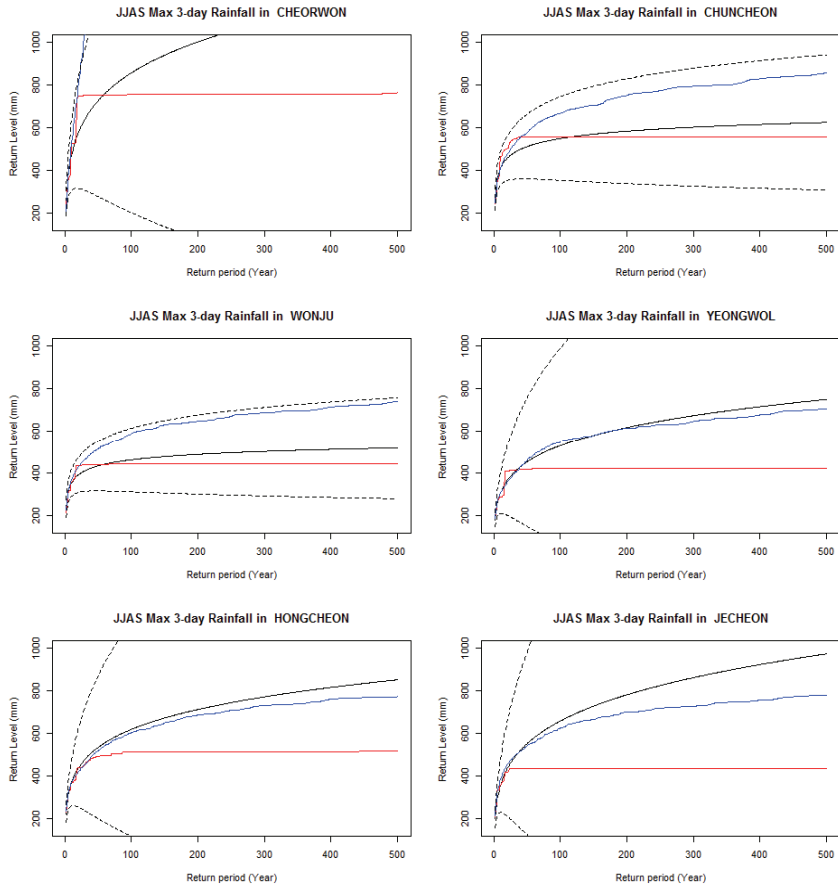


Figure 95. 한강 상류유역 호우기 최대3일연속누적강수량의 return level의 그래프.

- 제안된 강우알고리즘(파란실선), 대조알고리즘(빨간실선), GEV 모형(검은실선)의 return level. 파선은 GEV 모형에서 얻어진 return level의 95% 오차범위를 나타내는 기준선이다.
- 3일연속누적강수량을 고려한 이유는 호우사상의 지속시간이 평균적으로 2일이 넘기 때문이다.
- 대조알고리즘은 어느 return period 이상에서는 return level이 더 이상 증가하지 않는데, 그 정도가 최대일강수량 경우에 비해 심해졌다. 한편, 제안된 강우알고리즘은 GEV 모형의 return level에 근접하거나 오차범위 내의 값을 생산한다.
- 철원 지점은 여전히 제안된 알고리즘의 return level이 발산하고 있다. 이는 Pareto 모형을 쓴 데 따른 현상으로 보인다.

Table 65. 낙동강 유역 6~9월 최대일강수량 관측과 대조알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 대조알고리즘						
	안동	봉화	태백	영주	문경	의성
안동	1	0.00 / 0.01	0.03 / 0.01	0.00 / 0.0	0.33 / 0.01	0.30 / 0.00
봉화		1	0.18 / 0.01	0.52 / 0.2	0.10 / 0.58	0.00 / 0.01
태백			1	0.44 / 0.0	0.19 / 0.01	0.06 / 0.00
영주				1	0.44 / 0.19	0.00 / 0.00
문경					1	0.29 / 0.00
의성						1

Table 66. 낙동강 유역 6~9월 최대일강수량 관측과 제안알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 제안알고리즘						
	안동	봉화	태백	영주	문경	의성
안동	1	0.00 / 0.24	0.03 / 0.21	0.00 / 0.20	0.33 / 0.21	0.30 / 0.25
봉화		1	0.18 / 0.36	0.52 / 0.36	0.10 / 0.31	0.00 / 0.28
태백			1	0.44 / 0.32	0.19 / 0.28	0.06 / 0.22
영주				1	0.44 / 0.26	0.00 / 0.21
문경					1	0.29 / 0.18
의성						1

Table 67. 낙동강 유역 6~9월 최대3일 연속누적강수량의 관측과 대조알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 대조알고리즘						
	안동	봉화	태백	영주	문경	의성
안동	1	0.19 / 0.00	0.41 / 0.01	0.37 / 0.12	0.52 / 0.21	0.50 / 0.25
봉화		1	0.38 / 0.00	0.57 / 0.00	0.47 / 0.00	0.09 / 0.00
태백			1	0.48 / 0.01	0.39 / 0.00	0.20 / 0.00
영주				1	0.61 / 0.02	0.35 / 0.00
문경					1	0.44 / 0.00
의성						1

Table 68. 낙동강 유역 6~9월 최대3일 연속누적강수량의 관측과 제안알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 제안알고리즘						
	안동	봉화	태백	영주	문경	의성
안동	1	0.19 / 0.38	0.41 / 0.41	0.37 / 0.39	0.52 / 0.37	0.50 / 0.44
봉화		1	0.38 / 0.52	0.57 / 0.46	0.47 / 0.43	0.09 / 0.47
태백			1	0.48 / 0.49	0.39 / 0.44	0.20 / 0.46
영주				1	0.61 / 0.43	0.35 / 0.44
문경					1	0.44 / 0.45
의성						1

Table 69. 한강 유역 6~9월 최대일강수량의 관측과 대조알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 대조알고리즘						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.42 / 0.00	0.12 / 0.00	0.00 / 0.00	0.00 / 0.01	0.00 / 0.01
춘천		1	0.10 / 0.00	0.00 / 0.00	0.27 / 0.00	0.00 / 0.00
원주			1	0.51 / 0.16	0.20 / 0.01	0.45 / 0.45
영월				1	0.00 / 0.43	0.81 / 0.47
홍천					1	0.17 / 0.00
제천						1

Table 70. 한강 유역 6~9월 최대일강수량의 관측과 제안알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 제안알고리즘						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.42 / 0.30	0.12 / 0.30	0.00 / 0.36	0.00 / 0.26	0.00 / 0.23
춘천		1	0.10 / 0.27	0.00 / 0.29	0.27 / 0.34	0.00 / 0.19
원주			1	0.51 / 0.32	0.20 / 0.25	0.45 / 0.24
영월				1	0.00 / 0.28	0.81 / 0.30
홍천					1	0.17 / 0.17
제천						1

Table 71. 한강 유역 6~9월 최대3일 연속누적강수량의 관측과 대조알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 대조알고리즘						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.51 / 0.01	0.00 / 0.00	0.00 / 0.01	0.16 / 0.01	0.00 / 0.00
춘천		1	0.37 / 0.01	0.18 / 0.00	0.53 / 0.01	0.19 / 0.00
원주			1	0.69 / 0.15	0.53 / 0.00	0.64 / 0.18
영월				1	0.42 / 0.00	0.80 / 0.27
홍천					1	0.56 / 0.00
제천						1

Table 72. 한강 유역 6~9월 최대3일 연속누적강수량의 관측과 제안알고리즘의 χ 값. 음영영역에서 5%수준에서 유의한 차이가 있다.

관측 / 제안알고리즘						
	철원	춘천	원주	영월	홍천	제천
철원	1	0.51 / 0.39	0.00 / 0.38	0.00 / 0.32	0.16 / 0.41	0.00 / 0.33
춘천		1	0.37 / 0.46	0.18 / 0.45	0.53 / 0.47	0.19 / 0.44
원주			1	0.69 / 0.45	0.53 / 0.43	0.64 / 0.44
영월				1	0.42 / 0.43	0.80 / 0.43
홍천					1	0.56 / 0.43
제천						1

4. 결론 및 토론

본 연구에서는 낙동강과 한강 상류 유역의 강우현상-강우사상의 강수량과 지속시간, 사상 간 시간간격을 분석하여, 이 지역에서 호우기 강우를 모의하였다. 통계적 분석을 통해 호우사상의 강수량에 대해서 다변량 극단값 모형을, 호우사상의 지속시간에 대해서는 Gamma 일반화 선형모형을 구축하였다. 한편, 사상 간 시간간격에 대해서는 임의추출방식으로 이를 모의하는 것이 적절한 것으로 확인되었다. 구축된 통계모형을 바탕으로 새로운 강우모의알고리즘을 제안하였다. 이 알고리즘은 다지점의 강우현상을 설명하고 극한강우모의의 제고를 위해 극단값 모형이 도입되었다는 점에서 기존의 연구와 차별적이다.

제안된 모의알고리즘으로 모의자료를 생성하고, 관측자료와 비교하였다. 호우기 강우현상을 특징짓는 대푯값으로 호우기누적강수량, 최대일강수량, 강우일수, 그리고 호우사상의 강우량을 대변하는 최대3일연속누적강수량을 고려하였다. 모의자료와 관측자료에 대해 이들 대푯값을 계산하여 비교한 결과를 정리하면 다음과 같다.

- 지점별로 비교한 결과, 강우알고리즘은 두 유역 모두에서 호우기 강우현상을 만족할만한 수준으로 모의하였다.
- 지점 간 강우연관의 재현정도는 두 유역에서 차이를 보였다. 낙동강 상류유역의 경우에는 누적강수량과 강우일수의 지점연관성을 잘 모의하였지만, 최대일강수량에서는 상관정도가 완전히 일치하지는 않았다. 한편, 한강 상류 유역에서는 강우일수의 연관성은 잘 재현하였지만, 누적강수량과 최대일강수량의 상관정도는 일치하지 않았다.
- 제안된 강우알고리즘은 극한강우의 재현능력이 있다. 그러나 최대일강수량의 return level을 과대추정하는 성향이 있어 보인다. 반면에, 최대3일연속누적강수량에 대해서는 적절히 추정하는 것으로 판단된다. 또한, 극한강우의 지점 연관성을 재현은 하지만, 연관의 복잡성을 표현하는 데는 한계가 있어 보인다.

제안된 강수모의알고리즘은 기존의 것에 비해 여러 장점이 있으나, 지역의 강우를 완벽히 재현하는 데는 아직 부족함이 있다. 그래서 더 정교하고 활용도 높은 모의를 위해서는 통계모형 구축에 더 많은 노력을 쏟아야 할 것이다. 개선의 방향은 다음과 같다.

- 호우사상에 대한 다변량 극단값 모형으로 logistic 모형을 고려하였다. 그런데, 이 모형은 여러 지점의 연관성을 하나의 모수로 표현하기 때문에 강우현상을 완전하게 설명하는 데 한계가 있다. 그래서 더 복잡한 연관성을 표현할 수 있는 모형을 고려해야겠다.
- 활용도를 높이기 위해서는 보다 넓은 지역을 고려해야 한다. 지역이 넓을수록 지점 수가 늘어나며 모형이 복잡해 질 수밖에 없으나, 통계분석의 효율성을 위해서는 복잡도를 조절해야 한다. 향후 적정수준의 복잡도를 지니는 다변량 극단값 모형의 선택을 고민할 필요가 있겠다.

향후, 이 연구결과를 바탕으로 지역 대상의 weather generator를 개발할 계획이다. 서론에서 언급하였듯이, weather generator는 기후예측자료의 시공간적 세밀도를 높이는 상세화(downscaling)의 주요한 방편이다. Weather generator의 기본 골격은 그림 96과 같은데, 강수모의가 기상생성의 기반을 이루고 있음을 알 수 있다. 그래서 본 연구를 통해 얻어진 강우통계모형은 지역 대상의 weather generator 개발에서 초석이 될 것이다.

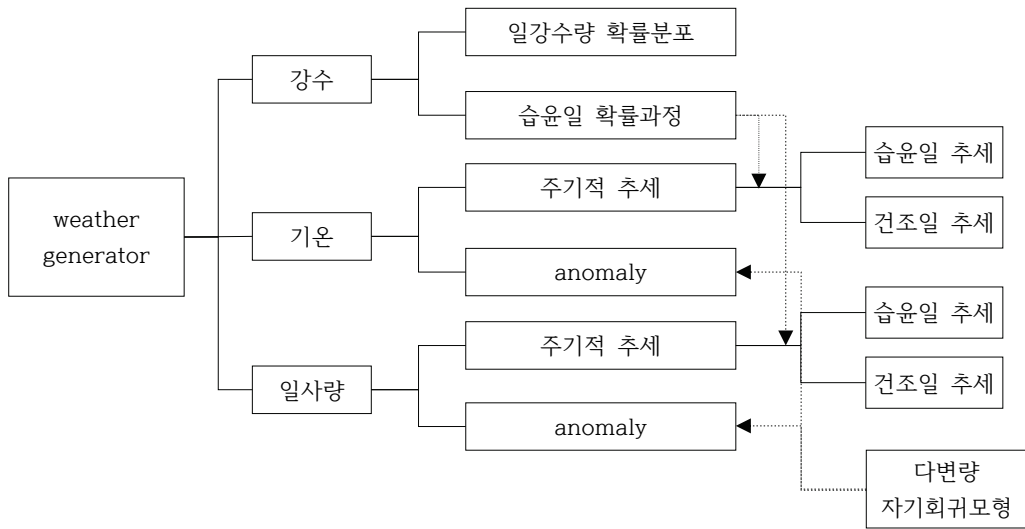


Figure 96. weather generator의 통계모형

REFERENCES

- 김병식, 김보경, 경민수, 김형수. 2008: 기후변화가 극한강우와 IDF 분석에 미치는 영향 평가. *한국수자원학회논문집, 한국수자원학회*, **41**(4), 379-394.
- 성용규, 손중권. 2013: 일반화 극단 분포를 이용한 강우량 예측. *한국데이터정보과학회지*, **24**(4), 857-865.
- 신지예, 박예준, 김태웅. 2013: 비정상성 GEV 모형을 이용한 행정구역별 미래 확률강우량 산정, *Journal of KOSHAM*. **13**(3) 147-156.
- 염명진, 조원철, & 허준행. 2008: GPD 모형 산정을 위한 부분시계열 자료의 임계값 산정방법 비교. *한국수자원학회논문집, 한국수자원학회*, **41**(5), 527-544.
- 윤선권, 문영일. 2014: 한국의 주요도시지점 기준강수량 초과 강수의 최근 증가경향 분석. *대한토목학회논문집*, **34**(1), 117-133.
- Altman, N. S. 1992: An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46**(3) 175-185.
- Apipattanavis, S., Podestá, G., Rajagopalan, B., and Katz, R. W. 2007: A semiparametric multivariate and multisite weather generator. *Water Resources Research*, **43**(11). W11401, doi:10.1029/2006WR005714.
- Coles, S., Bawa, J., Trenner, L., and Dorazio, P., 2001: *An introduction to statistical modeling of extreme values(Vol. 208)*. London: Springer.
- Coles, S., Heffernan, J., & Tawn, J. 1999: Dependence measures for extreme value analyses. *Extremes*, **2**(4), 339-365.
- Coles, S. G., and Tawn, J. A., 1991: Modelling extreme multivariate events. *J. Roy. Statist. Soc. Series B (Methodological)*, **53**(2). 377-392.
- Coomans, D. and Massart, D.L. 1982: Alternative k-nearest neighbour rules in supervised

- pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta* **136** 15-27
- Frahm, G. 2006: On the extremal dependence coefficient of multivariate distributions. *Statistics & probability letters*, **76**(14), 1470-1481.
- Racsko, P., Szeidl, L., and Semenov, M. 1991: A serial approach to local stochastic weather models. *Ecological modelling*, **57**(1), 27-41.
- Resnick, S. I., 1987: *Extreme values, regular variation, and point processes (Vol. 4)*. Springer Verlag, 320 pp.
- Richardson, C. W. 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, **17**(1), 182-190.
- Rodriguez-Iturbe, I., Cox, D. R., & Isham, V. 1987: Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, **410**, 269-288.
- Semenov, M. A., Brooks, R. J., Barrow, E. M., and Richardson, C. W., 1998: Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate research*, **10**(2), 95-107.
- Young, K. C. 1994: A multivariate chain model for simulating climatic parameters from daily data. *J. Appl. Meteorol.*, **33** 661-671.

부 록

본 부록에서는 보고서 본문에서 제시되지 않았던 그림을 제공한다. 우선, 낙동강 유역 각 지점(안동 제외)의 기준값 선정과 모형진단의 그림이다.

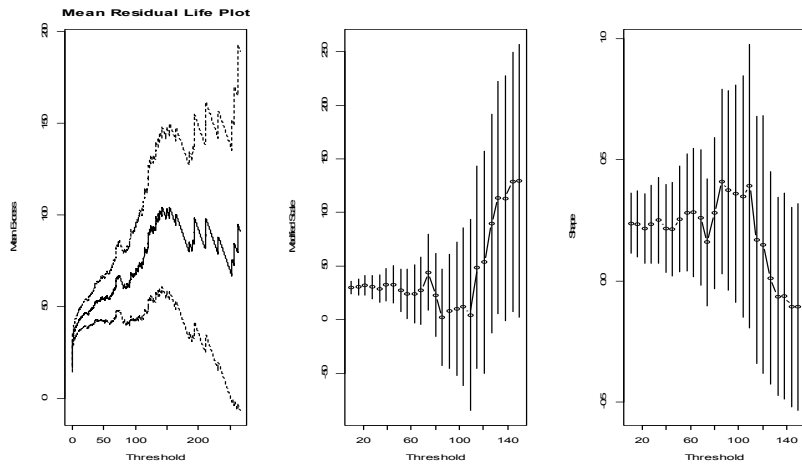


Figure 97. 봉화지점의 mean residual life plot과 threshold choice plot

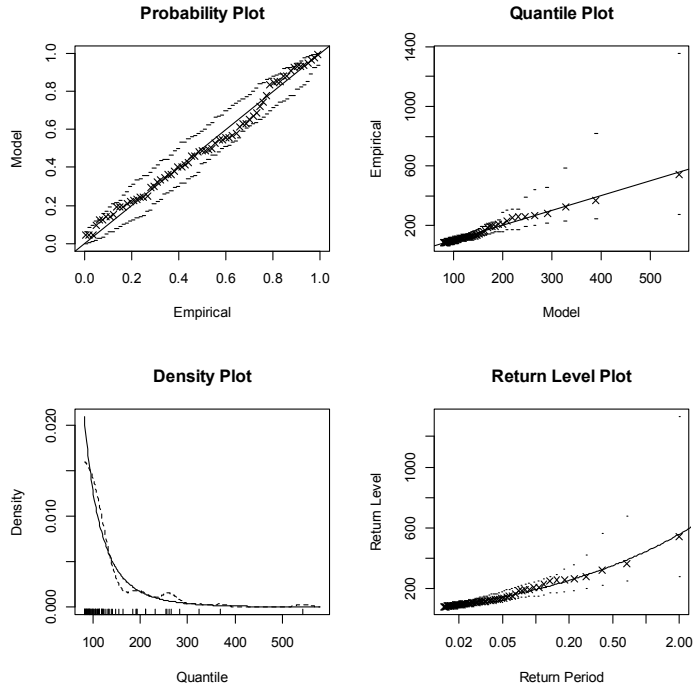


Figure 98. 봉화지점 generalized Pareto model의 모형진단

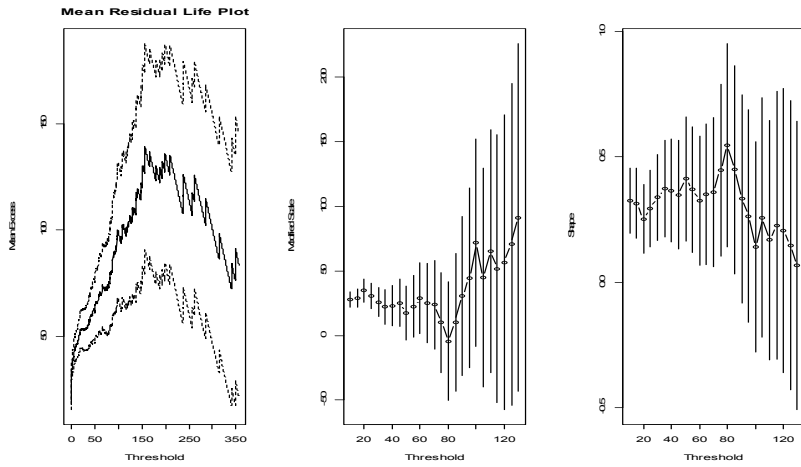


Figure 99. 태백지점의 mean residual life plot과 threshold choice plot

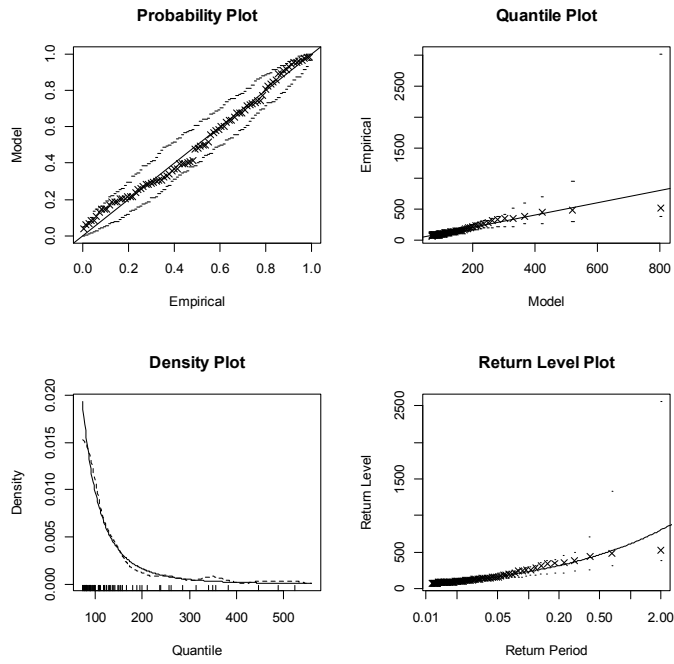


Figure 100. 태백지점 generalized Pareto model의 모형진단

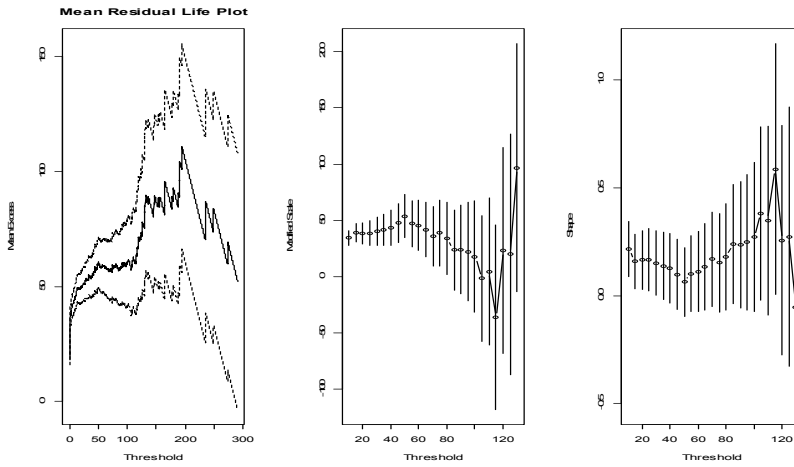


Figure 101. 영주지점의 mean residual life plot과 threshold choice plot

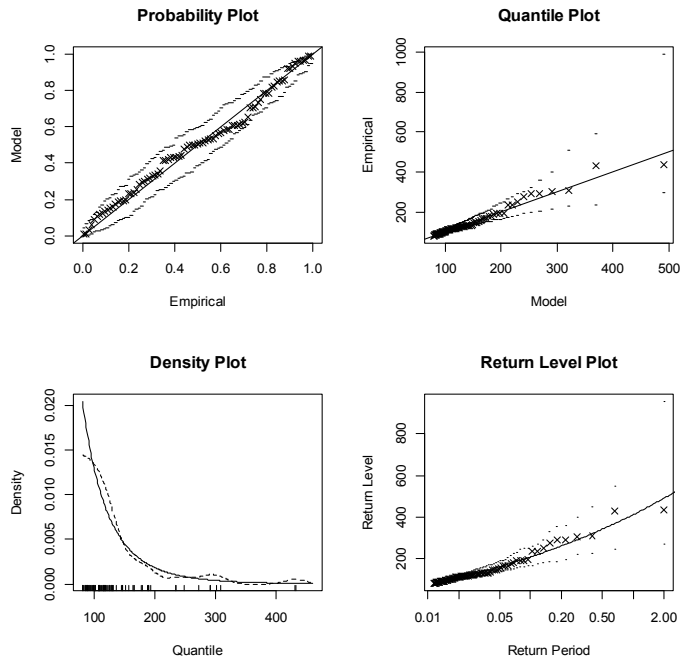


Figure 102. 영주지점 generalized Pareto model의 모형진단

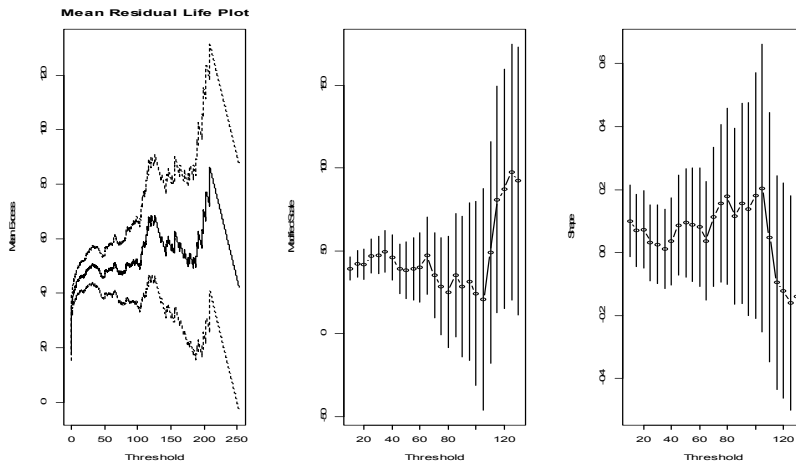


Figure 103. 문경지점의 mean residual life plot과 threshold choice plot

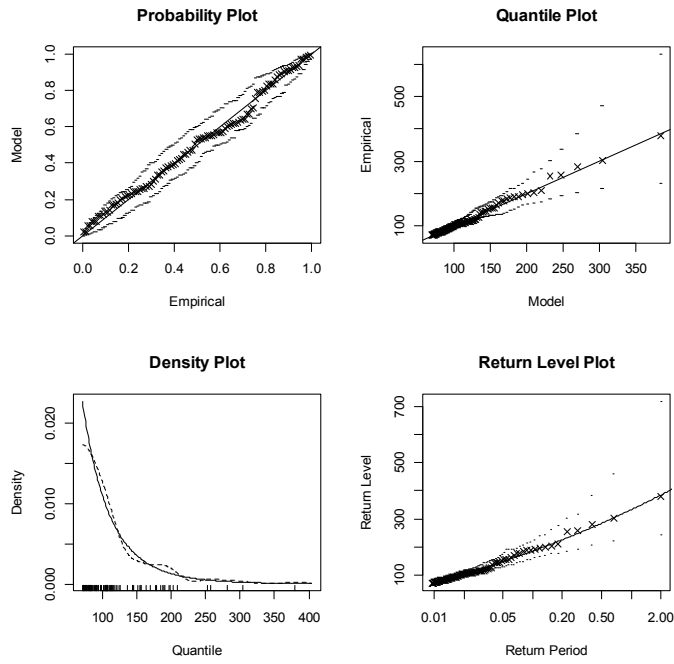


Figure 104. 문경지점 generalized Pareto model의 모형진단

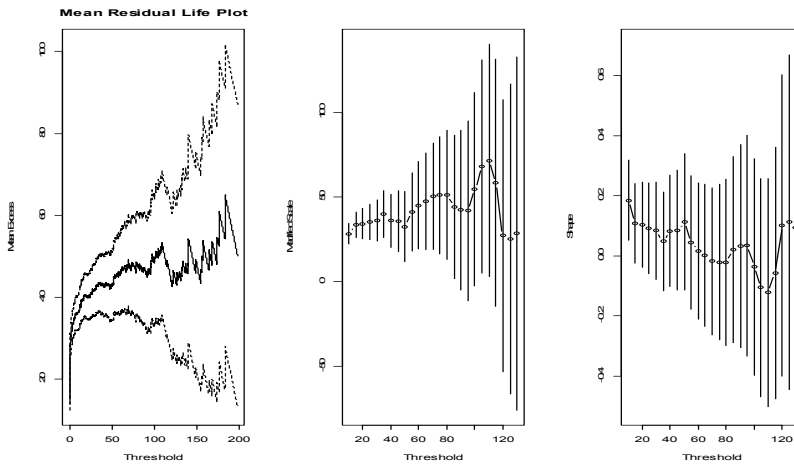


Figure 105. 의성지점의 mean residual life plot과 threshold choice plot

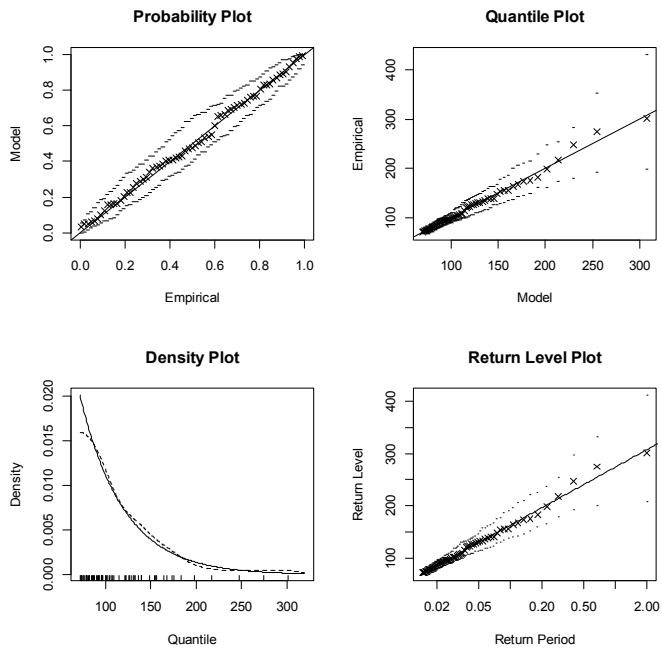


Figure 106. 의성지점 Gumbel model의 모형진단

낙동강 유역에서 사상 간 시간간격의 기준값 선정과 진단 그림이다.

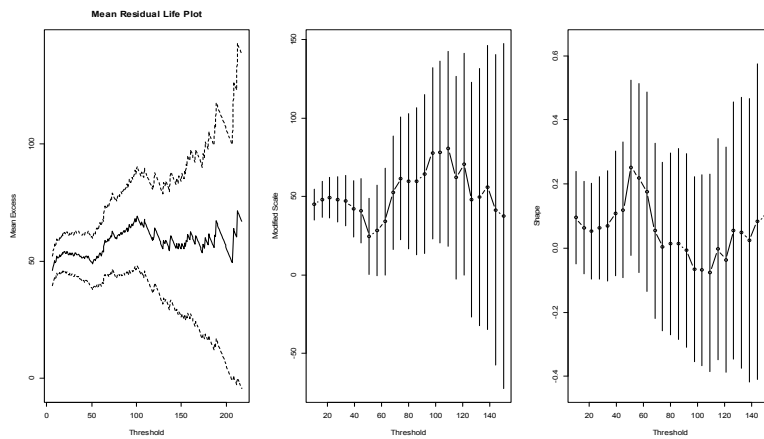


Figure 107. 6월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

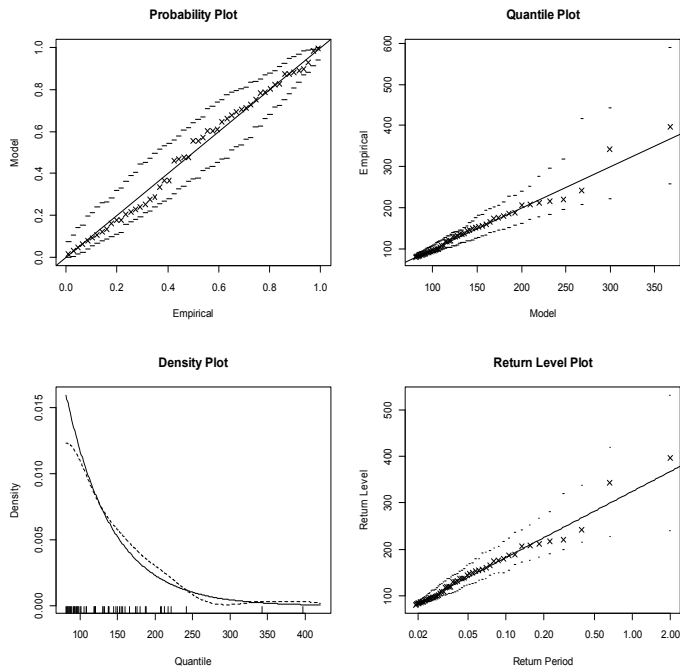


Figure 108. 6월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

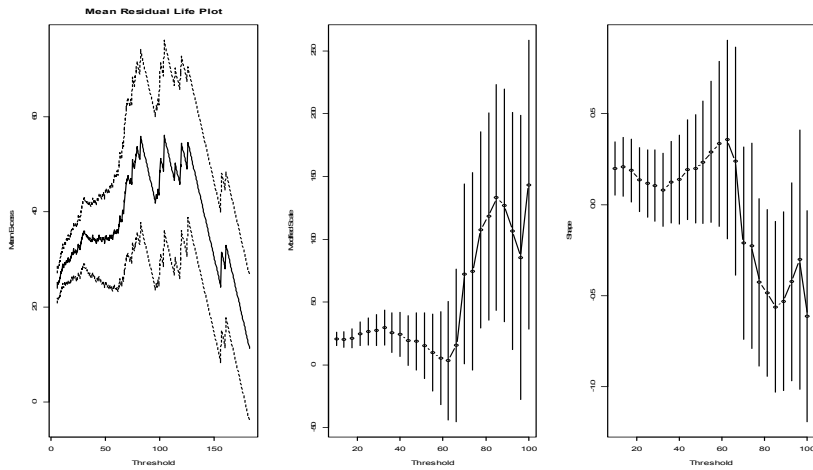


Figure 109. 7월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

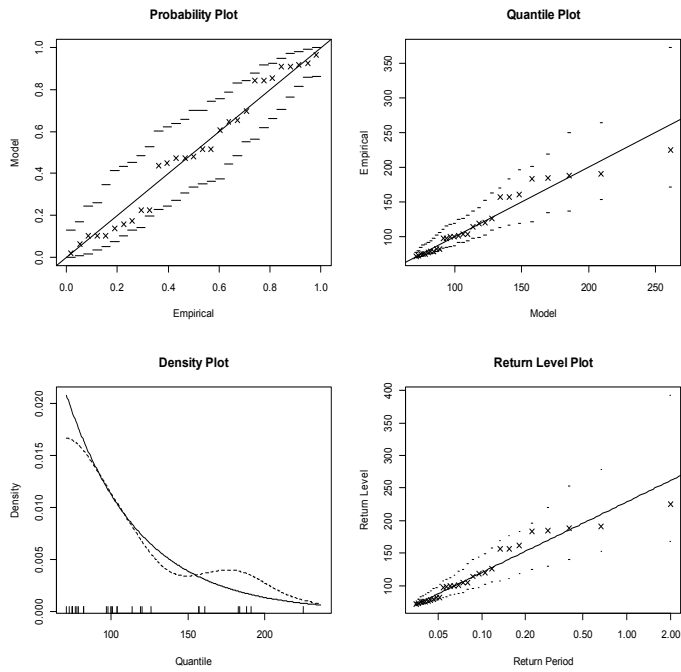


Figure 10. 7월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

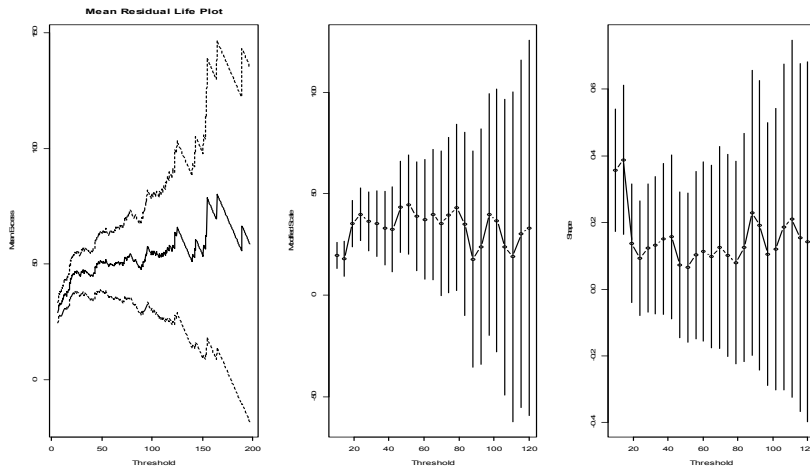


Figure 11. 8월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

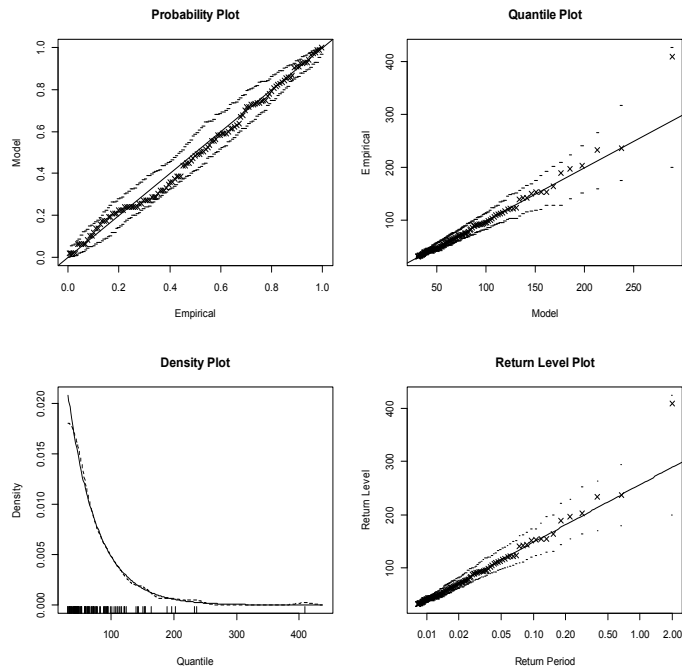


Figure 112. 8월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

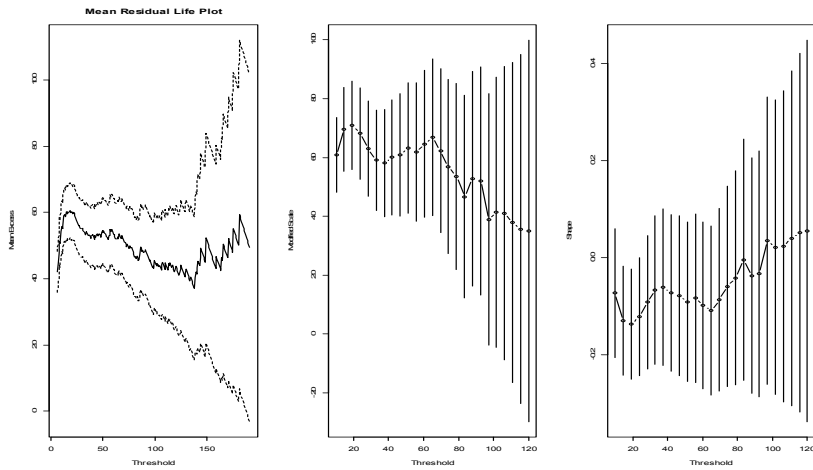


Figure 113. 9월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

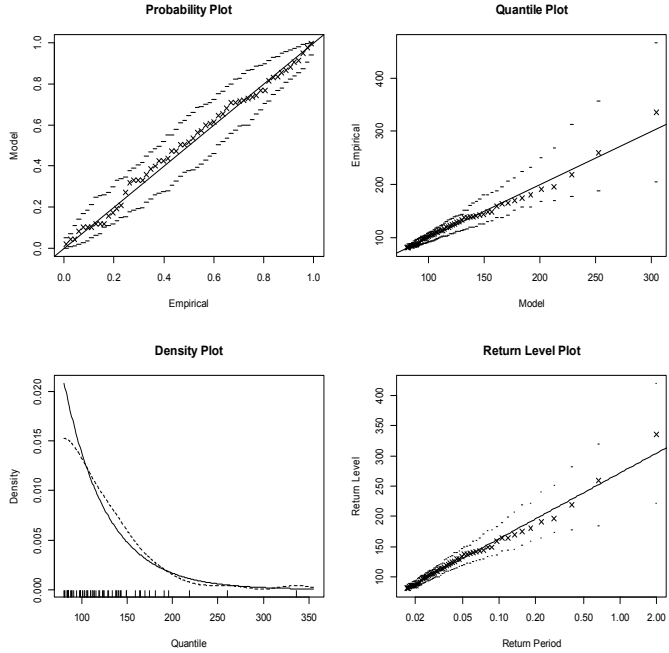


Figure 114. 9월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

한강 유역 각 지점의 기준값 선정과 모형진단의 그림이다.

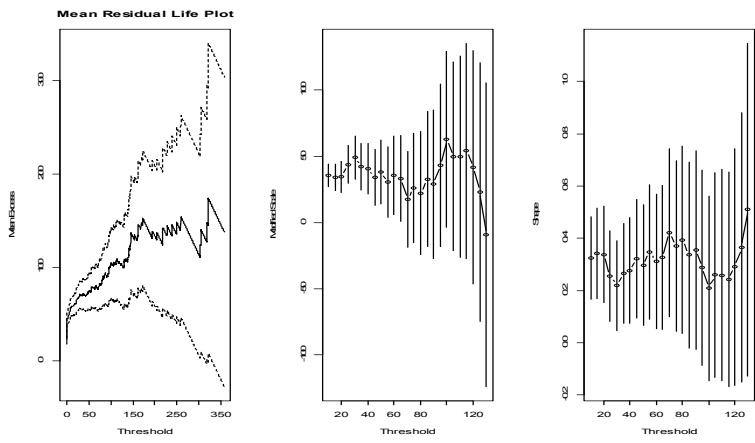


Figure 115. 철원지점의 mean residual life plot과 threshold choice plot

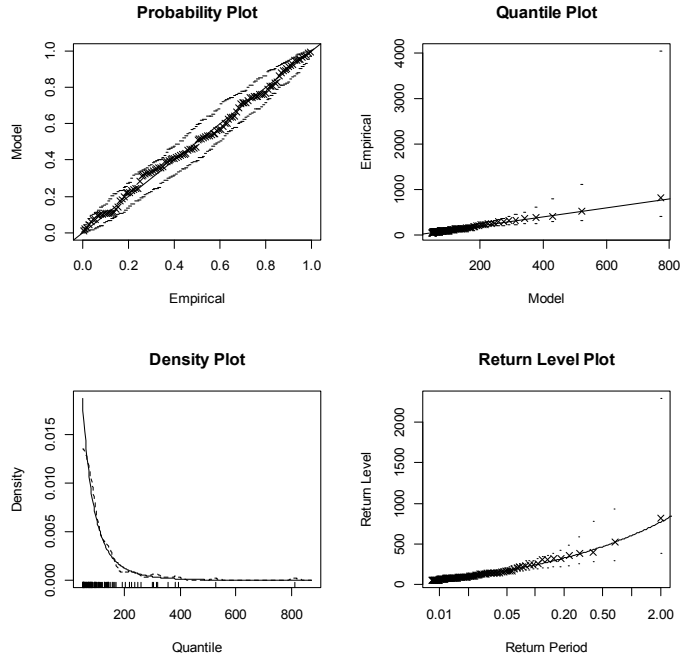


Figure 116. 철원지점 generalized Pareto model의 모형진단

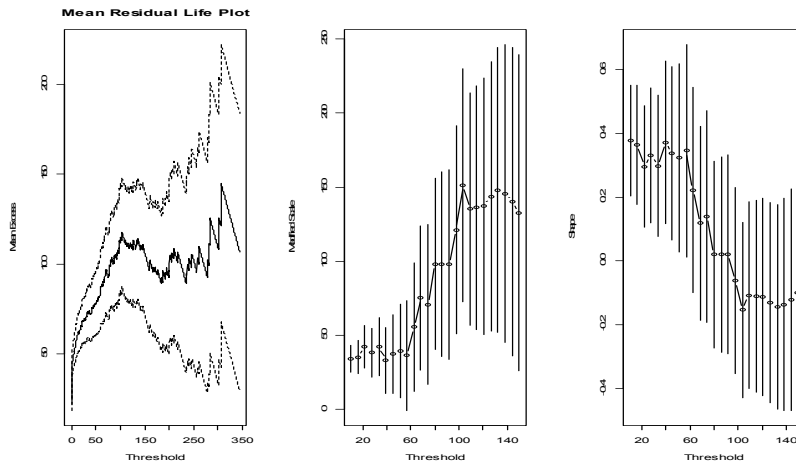


Figure 117. 춘천지점의 mean residual life plot과 threshold choice plot

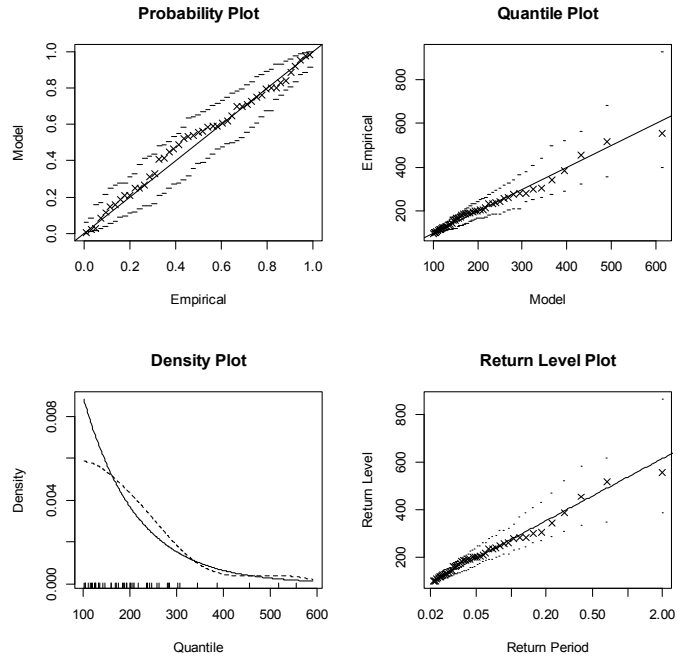


Figure 118. 춘천지점 Gumbel model의 모형진단

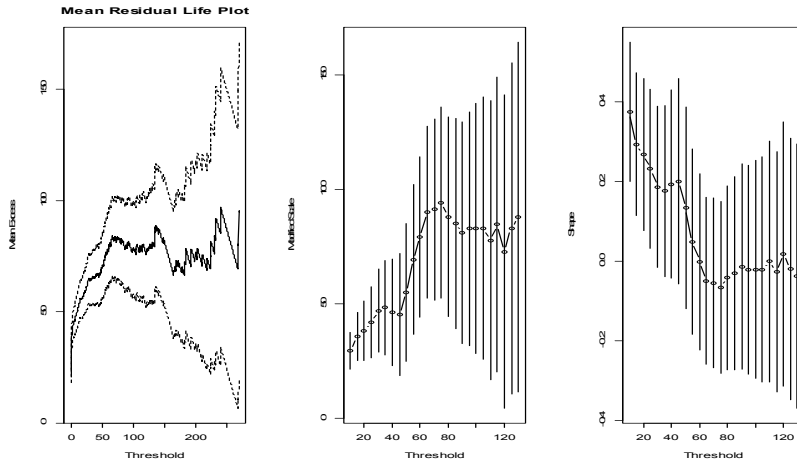


Figure 119. 원주지점의 mean residual life plot과 threshold choice plot

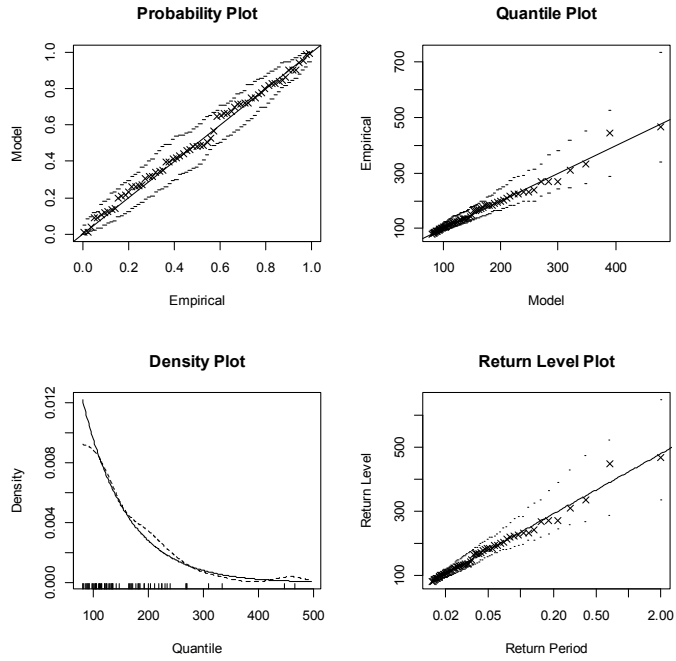


Figure 120. 원주지점 Gumbel model의 모형진단

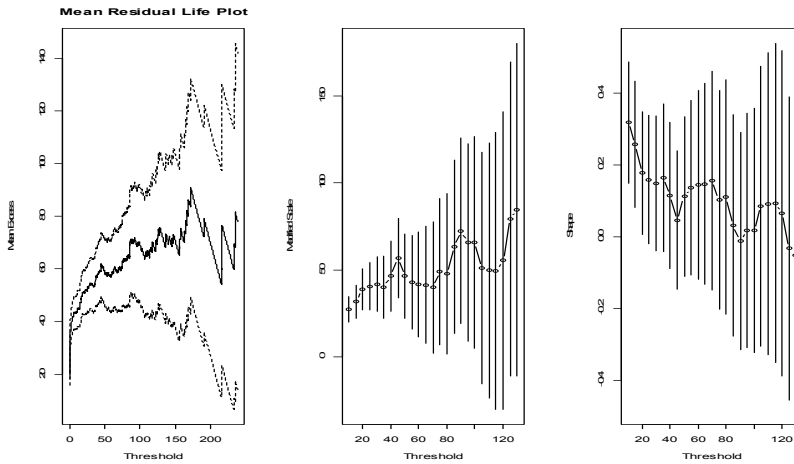


Figure 121. 영월지점의 mean residual life plot과 threshold choice plot

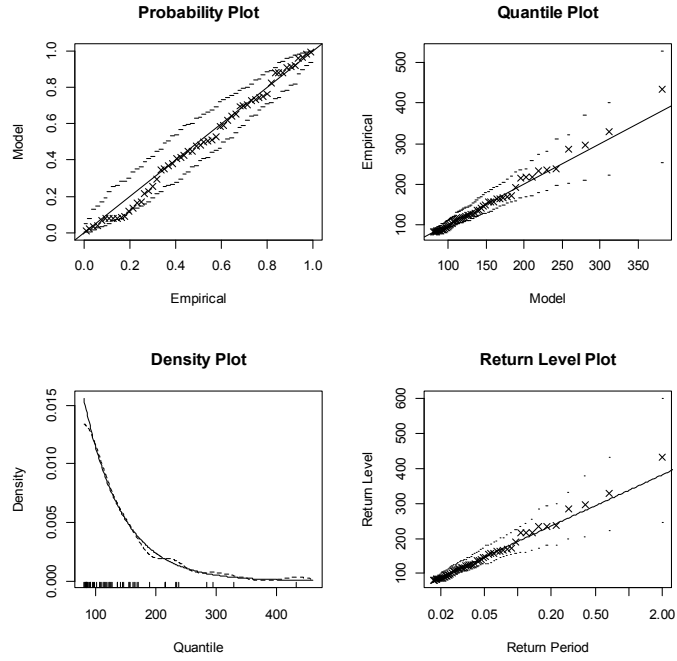


Figure 122. 영월지점 Gumbel model의 모형진단

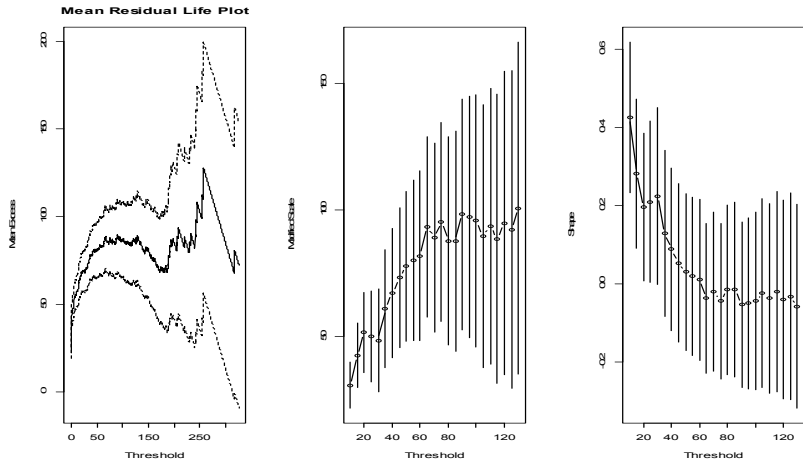


Figure 123. 흥천지점의 mean residual life plot과 threshold choice plot

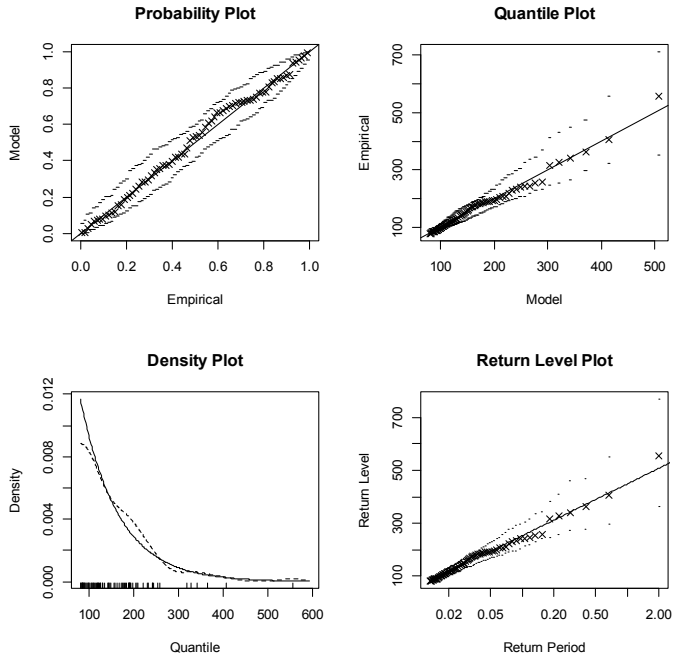


Figure 124. 홍천지점 Gumbel model의 모형진단

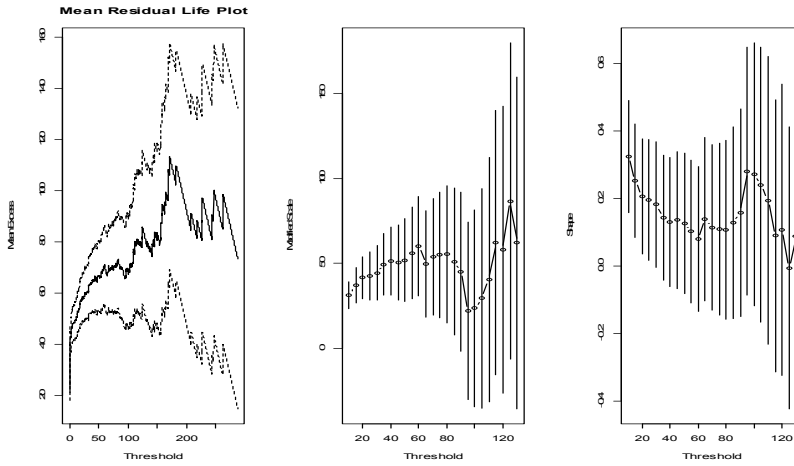


Figure 125. 제천지점의 mean residual life plot과 threshold choice plot

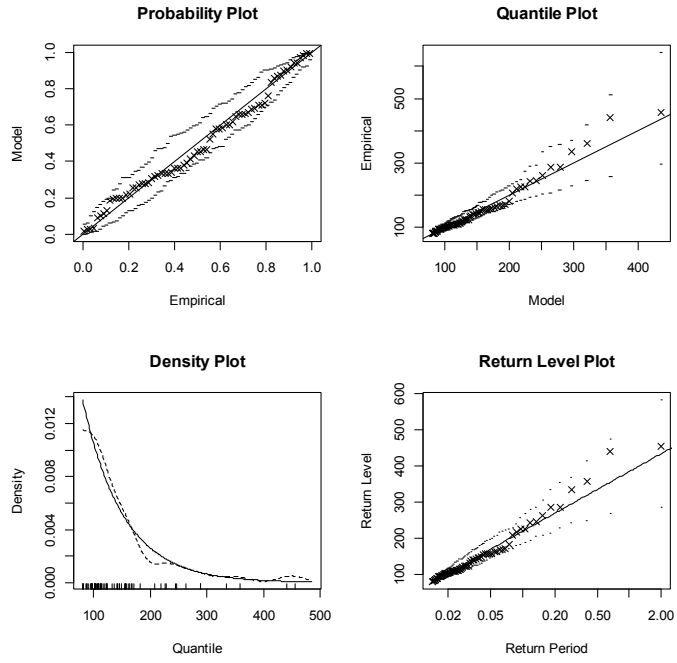


Figure 126. 제천지점 Gumbel model의 모형진단

한강 유역에서 사상 간 시간간격의 기준값 선정과 진단 그림이다.

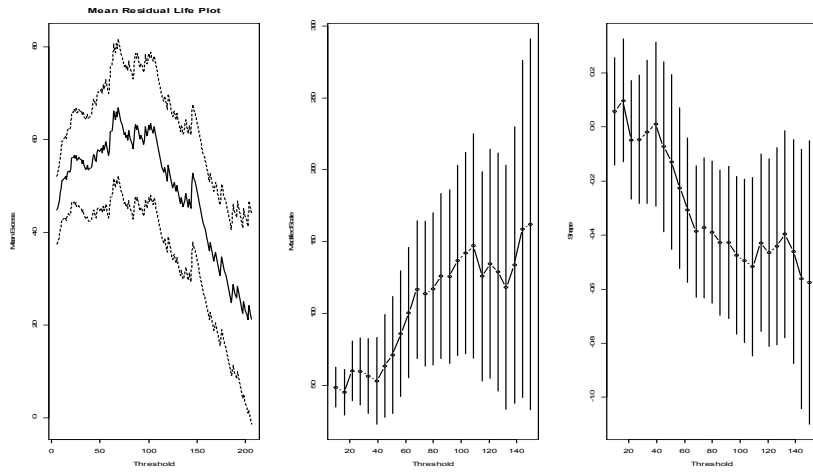


Figure 127. 6월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

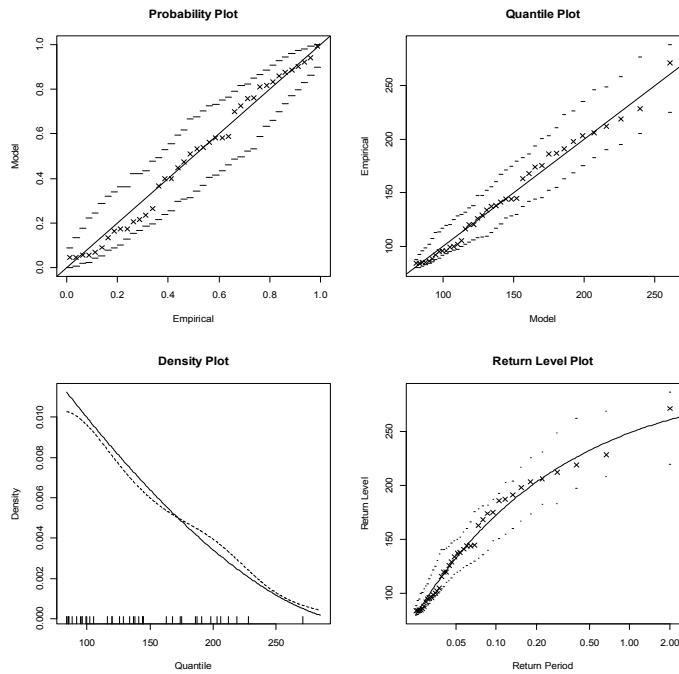


Figure 128. 6월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

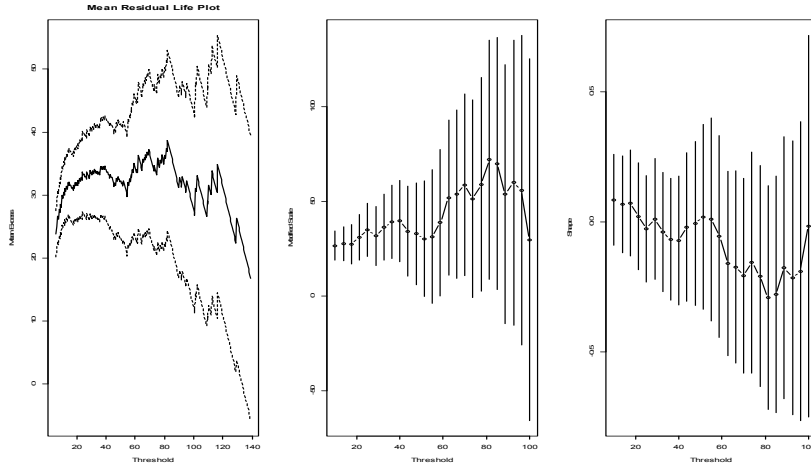


Figure 129. 7월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

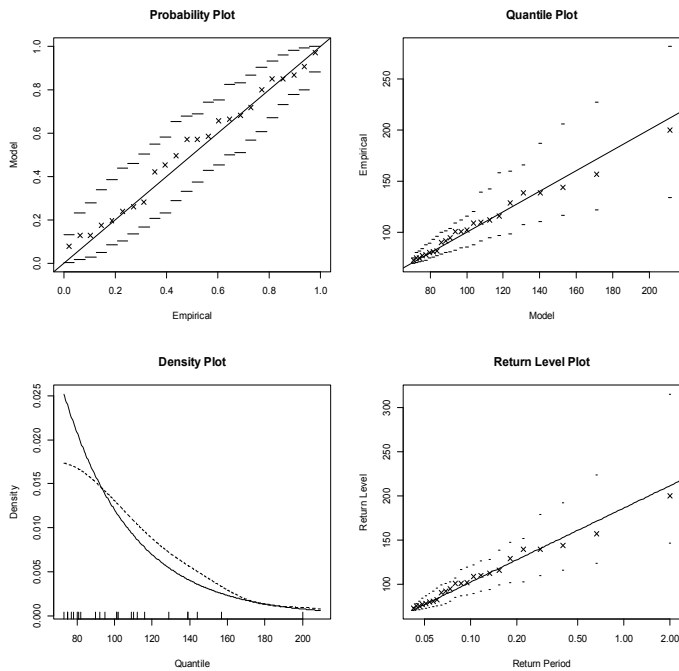


Figure 130. 7월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

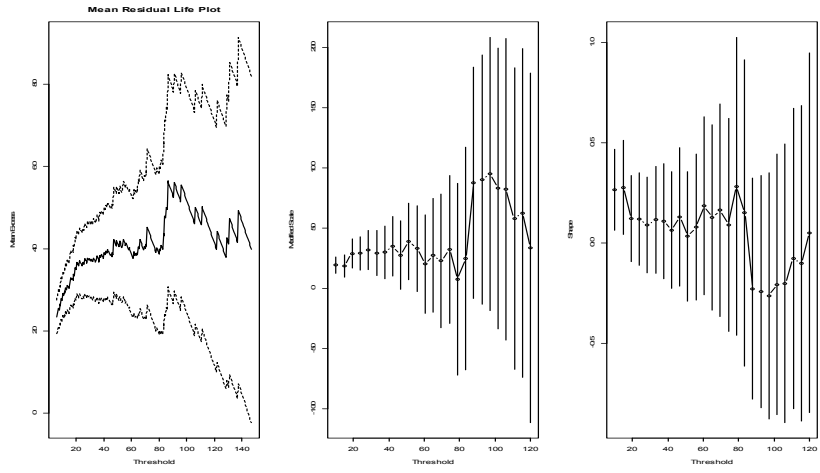


Figure 131. 8월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

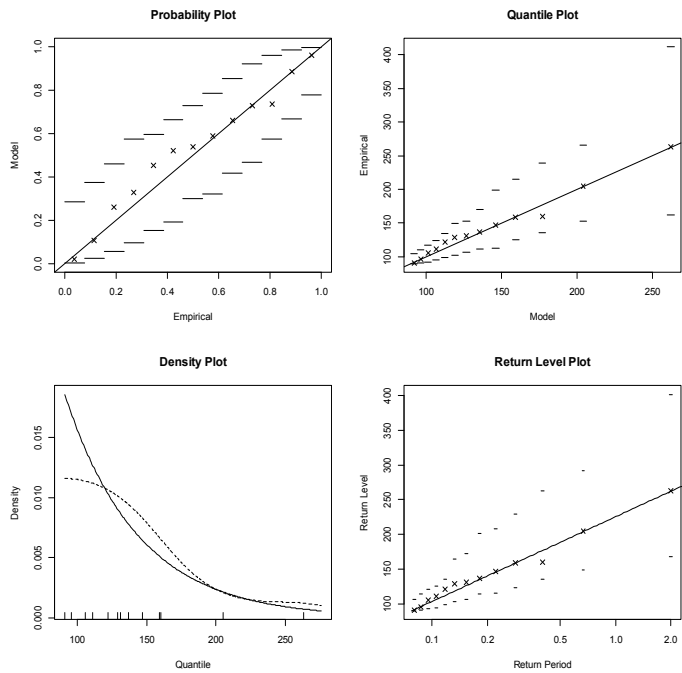


Figure 132. 8월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

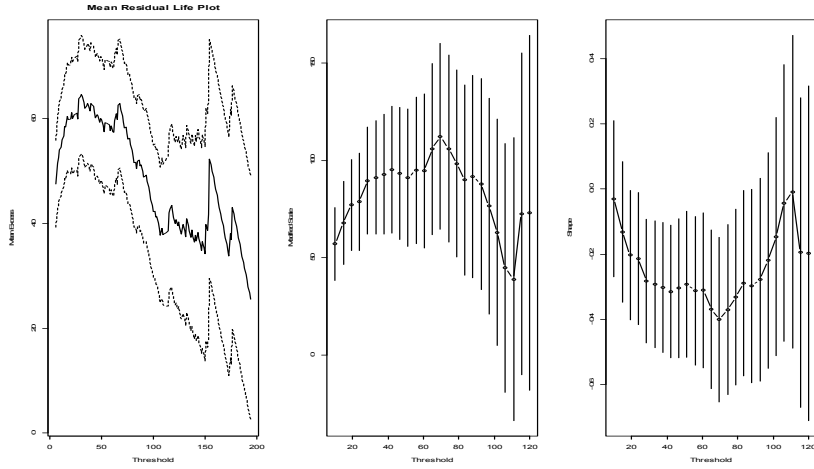


Figure 133. 9월 강우사상간 시간간격의 mean residual life plot과 threshold choice plot

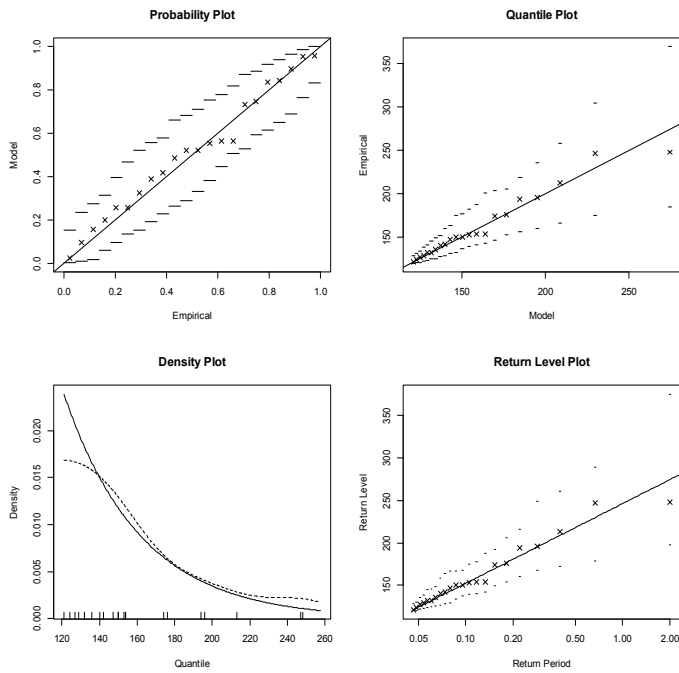


Figure 134. 9월 강우사상간 시간간격의 극단값 모형에 대한 진단그림

연구보고서 2015-08

극단값 모형을 기반한 국내 강우 현상 모의

Korean Precipitation Simulation based on Extreme Value Model

김무섭 기후분석팀



APEC Climate Center

612-020 부산광역시 해운대구 센텀7로 12

Tel: 051-745-3900 Fax: 051-745-3949

www.apcc21.org

 www.facebook.com/apcc21

 www.youtube.com/APECClimateCenter21

 www.twitter.com/apcc21

 www.plus.google.com/+APECClimateCenter21

 www.flickr.com/apcc21