연구보고서 2015-04

통계적 상세화를 위한 Weather Generator 개발

Development of Weather Generator for Statistical Downscaling

김무섭

APCC
APEC CLIMATE CENTER

연구보고서 2015-04

통계적 상세화를 위한 Weather Generator 개발

Development of Weather Generator for Statistical Downscaling

김무섭

APCC
APEC CLIMATE CENTER

발간사

GCM(General Circulation Model)은 기후 예측 및 변화연구의 근간이 되는 자료를 생산하는 중요한 모형이다. 하지만 기후자료의 시공간적 해상도가 타 응용분야의 요구에 미치지 못하여, 고해상도로 높이기 위한 기후자료 상세화(downscaling) 연구가 다양한 형태로 수행되어 왔다. 상세화 연구는 크게 RCM(Regional Circulation Model)을 기반으로 한 역학적 상세화(dynamical downscaling)와 경험적 통계모형을 기반으로 한 통계적 상세화(statistical downscaling)로 구분된다. 통계적 상세화는 역학적 방법에 비해 적은 계산자원으로 신속히 기후자료를 상세화할 수 있어, 응용분야에서 많이 활용되고 있다.

통계적 상세화는 여러 기법이 있는데, 본 연구에서는 계절예측자료의 시공간적 상세화를 위한 weather generator 기법을 고려한다. APCC에서는 매월 MME(Multi-Model Ensemble) 기법으로 계절예측자료를 생산하고 있는데, 자료 정확도와 신뢰성은 인정받고 있으나 시공간적으로 해상도가 낮다는 단점을 지니고 있다. 우리는 MME 계절예측자료의 시공간적 해상도를 높이는 위해 weather generator 기법에 주목한다. 하지만 기존의 weather generator는 통계모형의 기반인 관측자료의 평균적 기후의 재생을 목표로 하기 때문에 계절예측자료 상세화 적용에는 한계가 있다.

본 연구는 계절예측자료 상세화에 적합한 새로운 weather generator의 개발을 목표로 한다. 기존의 weather generator는 재추출 기법(resampling technique)을 기반으로 통계모형이 구축되어 이를 통해 손쉽게 기후변수의 분포를 재현할

수 있으나 기후의 경년변동성(inter-annual variability)을 표현하는데 한계가 있다. 본 연구에서는 재추출 기법에 대한 의존성을 줄이면서 경년변동성을 표현할 수 있는 모수(parameter)를 도입한 새로운 weather generator를 제안한다. 본 보고서에서 이 weather generator를 낙동강 유역에 적용하여 기본적인 성능을 검증하였다.

본 연구에서 제안한 weather generator는 평균적 기후를 재현할 뿐만 아니라 계절예측자료 상세화에 활용가능하다. 향후 이 weather generator를 활용하여 APCC MME 계절예측 상세화 자료를 생산할 계획이며 이를 통해 APCC 예측 자료의 활용도를 높일 수 있을 것이라 기대한다.

기후 자료의 상세화 연구는 우리 APCC 발전에 큰 기여할 것이라 기대한다. 마지막으로 이 연구의 수행을 위해 애쓰신 김무섭 박사에게 감사의 말씀을 전한다.

2016년 1월

APEC 기후센터

소장 **정 진 승**

ABSTRACT

GCM (Global Circulation Model) is a basic and fundamental tool for predicting future climate conditions. Unfortunately, its output is so coarse in space and time that it cannot be directly applied to application fields. Currently, there is a large amount of research focused on closing the gap between resolutions of GCM output and application data. This process, called downscaling, produces climate data of high resolution from coarse GCM outputs. Downscaling schemes have mainly been studied in two approaches - dynamical and statistical downscaling. Dynamical downscaling is based on RCM (Regional Circulation Model), which is a dynamical circulation model with higher spatial resolution and more detailed regional topography than GCM. Thus, it can produce climate data of fine resolution adapted for a target basin. On the other hand, statistical downscaling is based on statistical models describing basin weather and its relation with climate predictors. Since it can rapidly downscale GCM output without expensive computational costs, it has been employed in a variety of application fields. Recently, a hybrid approach has been studied to further increase the spatial resolution of dynamically downscaled data and correct the distributional bias using observation data.

In this report, we concentrate on statistical downscaling. There are several approaches for statistical downscaling. One of the most prominent approaches is a regression model which links basin weather variables with climate predictors so that climate prediction can be reflected on the basin weather. Since climate prediction is provided in coarse spatial resolution, it is viewed as spatial disaggregation. Another approach employs a weather generator, which generates

weather scenarios day to day by using a statistical model for basin weather. As such, it can be viewed as a tool for temporal disaggregation.

This report focuses on weather generators. Many weather generators have already been developed. Since weather generators are based on statistical models for basin weather, statistical modeling is the most prominent factor in developing a new weather generator. One of the most widespread models was proposed by Richardson, His model deals with precipitation, daily maximum/minimum temperatures, and solar radiation; it is a multivariate but site-specific model. The modeling approach is fully parametric: the precipitation model assumes that wet/dry-spells follow a first-order 2-state Markov model and rainfall amount is distributed according to a gamma distribution. In the temperature model, the daily mean trends are estimated by fitting Fourier series and the anomalies are modeled by the Vector AR (AutoRegressive) process that can describe their correlation structure. The precipitation model was modified by Racsko et al. As the parametric model may have model bias, the discrepancy between real world model and assumed model, thus non-parametric approach, is often employed. Racsko et al. adopted a wet/dry-spell alternation scheme instead of the Markov model. Moreover, rainfall amount is modeled by a semi-empirical distribution. Now in many publications, Richardson and Racsko models are called WGEN and LARS-WG, respectively. The site-specific model has been extended to a multi-site one for covering an entire basin rather than a specific site. Since a multi-site model is definitely more complicated than a site-specific one, non-parametric approach is preferred to parametric due to its flexibility. Apipattanavis et al. proposed a multi-site multivariate weather generator that assumes a parametric Markov model for wet/dry-spells but adopts the k-NN method for simulating conditional distribution of weather variables, which is a non-parametric resampling method.

In this study, we aim to develop a new weather generator for statistical downscaling

of climate prediction. For the purpose, the weather generator is required to have the following abilities:

- covering multiple sites;
- being linkable with climate predictors.

For spatial disaggregation of climate prediction, the weather generator has to be capable of simulating multi-site weather. Furthermore, for producing climate data of high resolution, spatial interpolation is applied to weather scenarios retaining spatial correlation. The second ability above is related to the inter-annual variation of climate. Typically, weather generator models are calibrated for generating climatological weather that ignores the inter-annual variation of climate. Thus, for such variations to be reflected, the weather generator has to be linkable with climate predictors. Motivated by these demands, we have developed a new multi-site multivariate weather generator for statistical downscaling. The weather generator deals with precipitation and daily maximum/minimum temperatures.

Now, we introduce the underlying statistical models of the proposed weather generator. First, the precipitation model is described in detail. As in the LARS-WG precipitation model, wet/dry-spells are simulated by a wet/dry-spells alternation model where each spell is generated independently. Especially, the lengths of dry-spells are generated by using a corresponding empirical distribution. As mentioned before, this method allows us to circumvent the model bias problem arising in Markov modeling. Generating wet-spell is more complicated than dry-spell since multi-site rainfall generation is involved. In particular, at-site distribution of rainfall amount is discontinuous at zero amount and multi-site distribution is far from multivariate normal. Thus, we take into account copula models that provide a general framework for modeling multivariate distributions. Actually, two approaches are considered for copula modeling: semi-empirical and Gaussian. The semi-empirical approach can be viewed as a multi-site version of the semi-empirical

distribution method of LARS-WG. As it is flexible, model fitting is rather simple. However, as the number of sites under consideration increases, it suffers from sparsity. On the other hand, Gaussian model is parametric, and thus sparsity problem vanishes but model bias may be involved.

In rainfall amount modeling, it is important to deal with extreme precipitation because it has enormous impacts on application fields. For appropriate extreme rainfall generation, extreme value models are employed; generalized Pareto distribution is fitted to extreme rainfall data at every site and Spatial Gaussian Extremes Modeling is adopted for describing inter-site coherence of extreme precipitation.

Next, the temperature model is described. The daily mean trends are estimated by local polynomial regression method instead of with Fourier series. The local polynomial regression seems to estimate mean trends more accurately than Fourier series. It is also notable that the daily mean is affected by precipitation. WGEN temperature model estimates daily mean trends of wet and dry days, separately, thus the magnitude of the effect is determined by whether the day is wet or dry. However, actual daily mean is observed to respond to rainfall intensity too: as the intensity gets stronger, daily mean falls down much more. This observation motivates the use of a regression model for mean down-shift against the intensity of precipitation that differentiates the proposed model from WGEN model. Temperature anomalies are obtained from standardization as in WGEN model, but they exhibit autocorrelation in space as well as time. To capture the correlations, we also consider the Vector AR model. Note that in multi-site model, the Vector AR model gets more complex (the number of parameters estimated is large) as the number of sites increases, thus we need to consider dimension reduction. It can be accomplished by EOF analysis, which provides a suitable set of principal component series. The Vector AR model is fitted to the vector series of the principal

components rather than anomalies directly. This enable us to capture the correlation structure efficiently.

We evaluated the performance of the proposed weather generator by carrying out a case study focusing on the Nakdong basin in the Korean peninsula. ASOS (Automatic Synoptic Observation System) daily data is readily available and 14 observation stations are located nearby. A 26 year observation period was used (January 1, 1988 - December 31, 2013). We fit the models described above to the ASOS data and generated 1,000 sets of simulation data, whose period is from January 1 to December 31. The performance was measured by the similarity between statistical characteristics of observation and simulation data. In the evaluation, we took the following characteristics into consideration: daily distribution; spatial correlation, temporal correlation of temperatures; correlation between maximum and minimum daily temperatures; monthly accumulation and maximum of precipitation; wet-day frequency; and monthly averages of temperatures. As the result, most of characteristics were reproduced well by the proposed weather generator, except monthly averages of temperatures where the synthetic averages revealed a narrower spread than observations. In calibrating the weather generator model, inter-annual variation of climate was ignored but significantly detected in real climate. We guess that excluding this variation causes the flaw in the monthly averages of temperatures. In future studies, we will investigate climate predictors and link them to weather generators for piping inter-annual variation of climate to basin weather.

목 차

1. 서론 ---- 1

2. Weather Generator의 모형과 알고리즘 ---- 8

2.1 강수모형과 알고리즘 ---- 9

2.1.1 은닉상태모형 ---- 9

2.1.2 Copula 모형 ---- 12

2.1.3 극단값 모형 ---- 14

2.1.4 습윤기와 강수 시나리오의 생성 알고리즘 ---- 16

2.2 기온모형과 알고리즘 ---- 20

2.2.1 일평균 모형 ---- 20

2.2.2 Anomaly 모형 ---- 23

2.2.3 기온 시나리오 생성 알고리즘 ---- 27

3. 사례 연구 - 낙동강 유역 ---- 29

3.1 모형추정과 Tuning ---- 30

3.2 모의 성능 검증 ---- 37

3.2.1 일별분포 검증 ---- 37

3.2.2 공간적 자기 상관 ---------- 56
3.2.3 기온의 시간적 자기상관 ---------- 69
3.2.4 최고/최저 기온 간 상관 ---------- 76
3.2.5 월단위의 강수특성 모의성능 검증 ---------- 79
3.2.6 기온 월평균 모의성능 검증 ---------- 90

4. 결론 및 향후 연구 계획 ---------- 95
4.1 결 론 ---------- 95
4.2 향후 연구 계획 ---------- 97

참고 문헌 ---------- 98

부 록 - 모형 추정법 ---------- 101

1. 서론

GCM(General Circulation Model)은 전지구적 기후예측 및 변화 연구의 가장 근원적인 도구이다. 이 모형은 대기역학을 기반으로 구축된 결정론적 모형으로 예측수행에 있어 많은 계산 자원이 소모되지만 가장 신뢰할 수 있는 미래 기후 정보를 생산한다. 현재, 여러 대학과 연구소에서 GCM 모형을 개발·운영 중이며, 우리 APCC(APEC Climate Center)에서도 CCSM3 모형을 계절 예측에 사용해 오고 있다. 아울러, 세계 유수의 대학 및 연구소와 협력으로 매월 예측 정보를 수집해 MME(Multi-model Ensemble, 다중모형앙상블) 계절예측자료를 생산하고 있고, 2015년에는 계절내 예측력(intra-seasonal prediction) 향상을 위해 SCoPS (Seamless Coupled Prediction System) 모형을 도입해 현업화를 앞두고 있다.

하지만, 대부분의 GCM에서 생산되는 기후자료들은 수문, 농업 등의 응용분야에 바로 활용되기에는 시공간적으로 낮은 해상도를 지니고 있다, (cf. Eum et al. (2010)). 그림 1은 GCM의 격자규모와 응용연구의 공간규모 간 차이를 보여 주고 있다.[1] 그림에서 응용연구의 대상 지역이 한 GCM 격자에 거의 포함되어 있다. 또한, 현재 APCC에서 생산하는 MME 자료는 2.5° × 2.5° 격자단위에서 월단위로 생산되는 반면, 농업과 수문분야에서는 수 km 단위 공간규모에서 일단위의 기후자료를 필요로 한다. 일찍이 응용분야에서는 이러한 필요를 충족시키기 위해서, GCM 자료로부터 시공간적으로 고해상도의 기후자료를 생산하는 연구가 다양한 형태로 수행되어 왔다. 일반적으로, 저해상도의 기후예측자료로부터 고해상도의 자료를 생산하는 과정을 '상세화'(downscaling)라고 부르는데, 크게 RCM(Regional Circulation Model, 지역순환모형)을 기반으로 한 역학적 상세화(dynamical downscaling)와 경험적 통계모형을 기반으로 한 통계적 상세화(statistical downscaling)로 구분된다. RCM은 GCM처럼 결정론적 역학모형으로 예측수행에 있어 상대적으로 많은 계산 자원이 요구되지만 대상지역의 지리적 특성(topography)을 반영한 기후예측정보를 생산할 수 있어 상세화의 좋은 도구이다, (cf. Jones et al. (1995)). 다른 한편, 통계적 상세화는 역학적 방법에 비해 접근이 용이하고 많은 계산 자원을 요구하지 않으면서 신속한 자료양산이 가능하기 때문에, 응용분야에서 많이 활용되고 있다. 최근에는 역학적으로 상세화

1) 그림 출처: http://consulclima.co.uk/climate-modelling/downscaling-gcm-outputs/

된 자료를 다시 통계적 상세화 기법으로 보정하는 hybrid 기법도 주목받고 있다. (cf. Guyennon et al.(2013)).

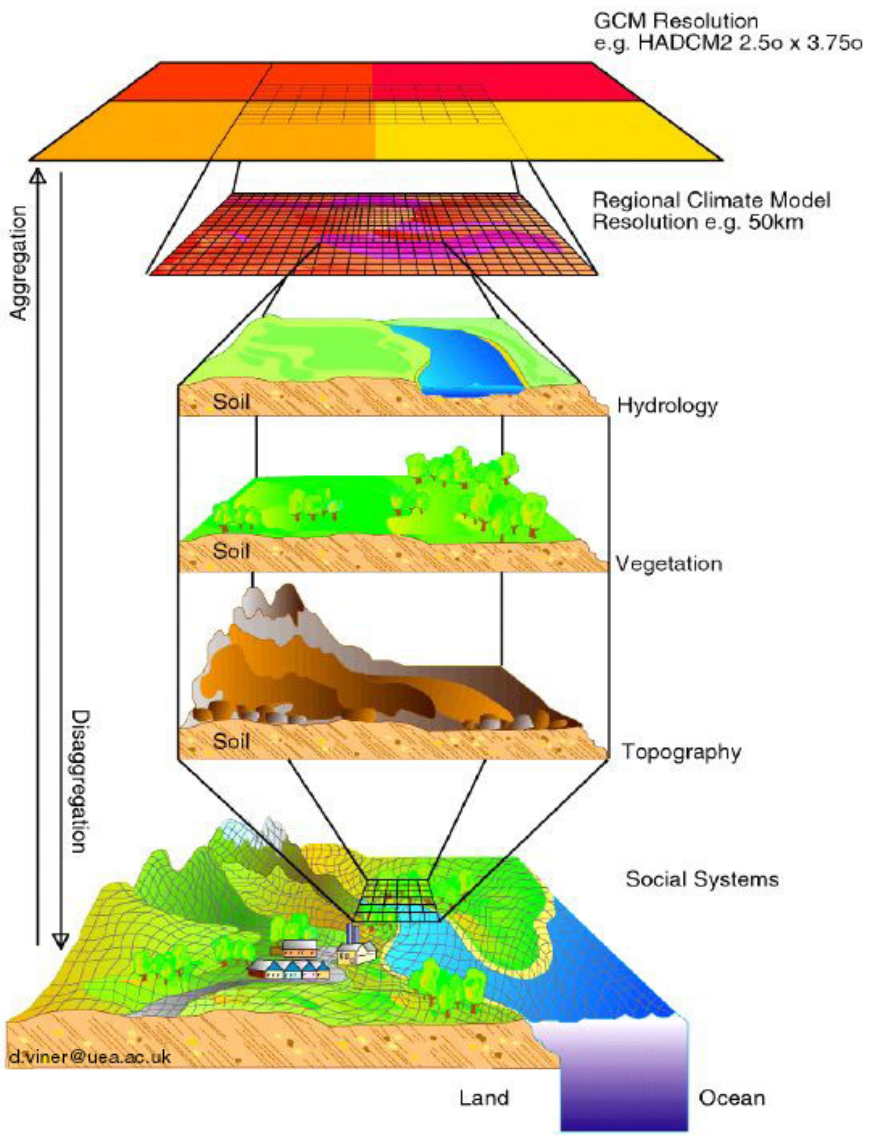


그림 1 GCM의 grid 규모(HadCM2, 2.5°×3.75°), RCM grid 규모 (50km×50km), 응용연구의 공간규모의 비교. 응용분야의 대상지역이 거의 CCM 한 격자내로 들어오고 있다.
(그림 출처: http://consulclima.co.uk/)

통계적 상세화 기법은 다양한 관점에서 연구되어 왔는데, 가장 대표적으로 회귀모형을 기반으로 한 상세화 기법을 들 수 있다, (cf. Semenov and Barrow (1997), Wilby et al. (1998)). 이 기법을 통해 대상 지점의 강수와 기온 같은 기상변수와 SST나 SLP와 같은 큰 공간규모의 기후인자 간 통계적 관계를 도출하여 기후의 경년변동(inter-annual variation)을 지점 기상에 반영할 수 있다. 즉, 대규모 기후장(climate field)과 소규모 유역의 기상을 연결짓는 공간적 상세화(spatial downscaling)의 일종이다. 공간적 상세화의 다른 방법으로 weather typing을 들 수 있다, (cf. Mehrotra and Sharma (2006b), Vrac and Naveu (2007)). 회귀모형이 유역 기상변수와 기후장 간 관계를 표현한다면, weather typing은 유역 기상 pattern을 기후장과 직접적으로 연결짓는다. 한편, 시간적 상세화(temporal downscaling)의 대표적 방법으로 weather generator를 들 수 있다. Weather generator란 기상관측자료로 구축된 통계모형을 이용하여 일별 기상시나리오를 모의생성하는 모형으로서, 농업, 수문, 환경 분야에서 널리 활용되고 있다, (cf. Semenov et. al. (1998)). 주어진 월단위 계절예측에 대응하는 일단위 기상시나리오는 무한정 다양한데, weather generator를 이용해 가능한 기상시나리오를 필요한 만큼 생성할 수 있다. 표 1에서 앞서 설명한 상세화 기법을 요약하였다.

표 1 상세화 기법의 분류와 특징

상세화 분류	기법/도구	특징
역학적 상세화	RCM	역학모형기반의 시공간적 상세화
통계적 상세화	회귀모형	기후인자와 기상변수의 통계적 관계를 회귀모형, 이를 기반으로 한 공간적 상세화
	weather typing	기후인자와 유역 기상 pattern 간 통계적 관계를 이용한 공간적 상세화
	weather generator	통계모형을 기반으로 한 시간적 상세화, 다지점 weather generator로 시공간적 상세화도 가능

Weather generator는 기상변수에 대한 통계모형을 기반으로 기상을 모의한다. 즉, weather generator의 연구는 내재된 통계모형의 개발 및 개선에 따른 것이다. 초창기 weather generator 연구로서, Richardson (1981)는 지점 기상관측자료를 이용해 강수, 기온, 일사량에 대해 통계모형을 구축하고 이를 기반한 기상시나리오 생

성을 연구하였다. 그는 2-state Markov 모형을 이용하여 습윤/건조일을 모의하고 감마분포로 습윤일의 강수량을 생성하였다. 또한, 기온의 순환추세를 습윤일과 건조일에 따라 Fourier 급수로 추정하고, anomaly를 벡터자기회귀모형(vector autoregressive model)[2]으로 모의하였다. 이후, Racsko et al. (1991)은 Richardson의 강수모형을 비모수적(nonparametric) 접근법으로 수정하였다. 2-state Markov 모형을 대신해 습윤/건조기 교대 모형(wet/dry-spell alternation model)을 사용하여 습윤일을 모의하였고, 강수량의 분포를 감마분포 대신해 준경험적 분포함수(semi-empirical distribution)로 추정하였다. 이를 통해 Richardson 모형이 지닐 수 있는 model bias를 줄일 수 있다.[3] 이후 Richardson과 Racsko의 weather generator는 작물생장모형, 병해충 모형의 입력기상자료 생산에 널리 쓰이고 있으며, 각각 WGEN, LARS-WG로 불리고 있다. 이처럼, 초창기 weather generator 연구는 농업분야에서 많이 이루어져 왔는데, 분야특성상 지점기상시나리오생성이 연구의 주된 내용이였다.

Weather generator의 활용이 다양한 분야로 확대되면서, 다지점(multi-site) 기상자료생성이 주목받기 시작하였다. 다지점 weather generator의 개발에서는 여러 지점을 동시에 다루기 때문에 내재된 통계모형이 복잡해짐은 불가피하다. 그래서 대다수의 다지점 weather generator 연구에서는 단일지점에서와 달리 비모수적 접근방식(non-parametric approach)으로 개발이 이루어졌다. Lall and Sharma (1996)은 시계열 재추출(re-sampling)을 위한 k-NN(k-nearest neighbor) 알고리즘이 제안하였고, Rajagopalan and Lall (1999)은 이를 이용해 단일 지점에서 강수를 비롯한 여러 기상변수의 모의를 연구하였다. Buishand and Brandsma (2001) 같은 알고리즘 기반으로 다지점 강수 및 기온 모의방법을 제안하고 라인강 유역에 적용하였다. Apipattanavis et al. (2007)은 역시 k-NN 알고리즘 기반한 다지점 다변량 weather generation 방법을 제안하였는데, Richardson의 모형에서와 유사하게 습윤/건조일을 parametric 모형인 3-state Markov 모형으로 모의하지만[4] (cf. Young (1994)), k-NN 알고리즘을 이용하여 기상변수의 조건부 확률분포를 비모수적 방식으로 모의

2] 다변량자기회귀모형이라고 부르기도 한다.

3] 사실, model bias를 줄이는데 대한 반대급부로 추정해야 할 모수가 증가하지만, 지점(site-specific) 모형에서는 그러한 단점이 두드러지는 않은 것으로 보인다.

4] 습윤일을 다시 non-extreme과 extreme 경우로 세분하였다.

하였다. 이 외의 주요한 관련연구로 Yates et al. (2003), Gangopadhyay et al. (2005), Mehrotra and Sharma (2006a), Eum et al. (2010) 등이 있는데, 언급된 모든 연구는 비모수적 방식을 채택하고 있다. 비모수적 방식을 통해 복잡한 parametric 모형을 피하면서 기상의 통계적 특성을 손쉽게 재현할 수 있지만 기후예측정보 상세화 기법 적용에 있어 제한적이다.

본 연구에서는 기후예측정보의 시공간적 상세화를 위해서 새로운 weather generator 개발을 목표로 한다. 이를 위해서 개발될 weather generator는

- 다지점의 기상을 모의할 수 있고,
- 기후예측정보 상세화를 위해 기후인자와 연결될 수 있어야 할 것이다.

다지점 weather generator는 수문분야에서 필요한 유역의 기상을 모의할 수 있을 뿐만 아니라, 고해상도 격자자료를 생산하기 위한 spatial interpolation의 기반자료를 생산할 수 있다. 한편, 통상적으로 weather generator는 관측기간의 climatology를 기준으로 기상시나리오를 생성하기 때문에, 기후의 경년변동을 유역의 기상에 반영하기 위해서 기후인자와 결부시킬 수 있는 특별한 장치가 필요하다. Bardossy (1997)는 GCM에서 얻어진 대기 순환 패턴과 유역 강수 및 기온 통계모형을 연결짓는 방법론을 제안하였다. Bardossy의 연결방식은 weather generator의 통계모형이 parametric 부분이 있을 것을 요구하기 때문에, 비모수적 weather generator에 적용이 쉽지 않다. 본 연구에서는 Bardossy 방식이 적용되도록 기후인자와 연결될 수 있는 다지점 weather generator를 개발하였다.

본 연구의 weather generator는 기상과 응용분야 모두에서 가장 기본이 되는 변수인 강수와 기온을 다룬다. 강수시나리오 생성 알고리즘으로 LARS-WG의 습윤/건조기 교대 방식을 고려하는데, 습윤기의 다지점 강수량 생성 모형은 강수량분포의 비연속성(discontinuity)과 비정규성(non-normality)으로 인해 매우 도전적인 주제이다. 우리는 weather generation의 유연성을 위해 재추출 방식(re-sampling method)을 고려하되 (cf. Lall et al. (1996)), 은닉상태모형(hidden state model)과 copula 모형으로 다양한 강수시나라오를 생성할 수 있도록 한다. 특히, copula 모형은 다지점 강수 생성의 핵심적인 역할을 할 것인데, 본 연구에서는 준경험적 copula 모형과 Gaussian copula 모형을 고려한다. 준경험적 copula 모형은 LARS-WG의

준경험적 분포처럼 유연성을 지니고 있는 반면에 Gaussian 모형은 parametric 모형으로 다양성을 추구하는 장점이 있다. 또한, 기상, 수문에서 중요한 의미는 지니는 극한강수(extreme precipitation)의 재현을 위해 극단값 모형(extreme value model)을 고려한다. 지점별 극한강수량 생성을 위해 추정의 효율성(estimation efficiency)이 높은 일반화 Pareto 모형(generalized Pareto model)을 사용하였고, 극한강수의 공간상관성을 재현을 위해 공간 극단값 모형(spatial extremes model)을 고려하였다. 사실, Gaussian copula 모형과 공간 극단값 모형의 도입은 통계적 추정문제를 야기하는데, 본 보고서의 마지막 부록에서 이 모형추정문제를 자세히 다룬다.

기온에 대해서는 비선형적인 연주기 순환추세를 추정하기 위해서 비모수적 방법인 local polynomial regression 방법을 쓴다. 이를 통해 WGEN의 Fourier 급수 적합에 비해 더 유연하고 정확한 추세를 얻을 수 있다. 또한, 기온에 대한 강수의 영향력을 회귀모형으로 표현함으로써, 영향력의 연속적인 정량화를 시도하였다. WGEN이나 LARS-WG에서는 단순히 건조일과 습윤일을 구분하여 기온 일평균을 추정하였는데, 우리의 기온모형에서는 강수량과 공간패턴에 따라 일평균을 연속적으로 조정함으로써 기존의 것과 차별화시킨다. 또한, 다지점 기온 anomaly에 대해 벡터자기회귀모형을 고려하되, 주성분분석을 통해 모형적합의 안정성을 기한다. 표 2에서 우리 weather generator의 특징을 기존의 것들과 비교하였다.

표 2 weather generator의 비교

Weather generator	강수모형	기온모형
WGEN, Richardson (1981), 지점(site-specific) 모형	• 습윤/건조일 모의: 2-state Markov 모형 • 강수량 분포 추정: 감마분포	• 추세추정: Fourier 급수, 습윤/건조일로 구분 • anomaly 모형: 벡터자기회귀모형
LARS-WG, Racsko et al (1991), 지점 모형	• 습윤/건조일 모의: 습윤/건조기 교대 방식 • 강수량 분포 추정: 준경험적 분포함수	• 추세추정: Fourier 급수, 습윤/건조일로 구분 • anomaly 모형: 벡터자기회귀모형
Apipattanavis의 weather generator, Apipattanavis et al. (2007), 다지점(multi-site) 모형	• 습윤/건조일 모의: 3-state Markov 모형. • 강수량, 기온의 조건부 분포를 k-NN 방식으로 모의,	

Weather generator	강수모형	기온모형
우리의 weather generator, 다지점 모형	• 습윤/건조일 모의: 습윤/건조기 교대 방식 • 다지점 강수량 분포: Copula 모형 • 극한강수모형: 일반화 Pareto 모형, 공간 극단값 모형.	• 추세추정: local polynomial regression, 강수영향의 회귀모형 • anomaly 모형: 주성분-벡터자기회귀모형

이후 본 보고서의 구성은 다음과 같다: 2.절에서는 2015년도 연구결과로서 강수/기온 모형과 알고리즘을 자세히 설명하고, 3.절에서는 제안한 방법을 낙동강 유역에 적용한 사례와 성능검증결과를 제시한다. 4.절에서는 연구의 결론을 짓고 향후 연구를 간단히 언급하면서 마무리한다. 마지막 부록에서는 본문에서 미루어 두었던 모형 추정문제를 다룬다.

2. Weather Generator의 모형과 알고리즘

본 연구에서 우리는 통계적 상세화를 위한 weather generator 개발을 목표로 한다. Weather generator란 대상 유역의 기상을 모의생성하는 프로그램으로, 개발을 위해서는 기상변수의 값을 생성하기 위한 모의알고리즘과 통계모형이 필요하다. 본 연구에서는 기상에서 가장 주요한 변수인 강수와 기온을 다룬다. 본 절에서 낙동강을 시범유역으로 선정하여 구축한 강수와 기온의 통계모형과 모의알고리즘을 제시한다.

설명의 간결성을 위해 여러 기호를 도입한다. 별도의 명시가 없는 한, 지점을 나타내는 기호로 s를, 일(日)단위의 시점을 나타내는 기호로 t를 쓴다. 사실상, 지점은 2차원 평면상의 위치로 표시될 것이기 때문에, $s \in \mathbb{R}^2$이다. 그리고 고려 중인 지점의 수는 $d \in \mathbb{N}$로, 각 지점은 $s_1, \cdots, s_d$로 나타낸다. $\mathrm{E}(\cdot)$와 $\mathrm{I}(\cdot)$는 각각 expectation과 indicator 함수를 뜻한다. 그 밖에 주요하게 쓰이는 기호를 정리하면 표 3와 같다.

표 3 기호 요약

기호	의미	관련세부정의
$\mathbf{R}(t) = (\mathrm{R}(s_1,t), \cdots, \mathrm{R}(s_d,t))$ $\mathbf{R} = (\mathrm{R}(s_1), \cdots, \mathrm{R}(s_d))$	지점 $s_1, \cdots, s_d$에서 t일의 다지점 강수량. 시점의 명시가 없는 경우 t 생략. 문맥에 따라 관측자료를 지시할 수도 있다.	
$\mathrm{T}(s,t)$	지점 s에서 t일의 일최고 또는 최저기온. 문맥에 따라 관측자료를 지시할 수도 있다.	
$\boldsymbol{\Psi}_1, \cdots, \boldsymbol{\Psi}_d$, $\boldsymbol{\Psi}_1^{\max}, \cdots, \boldsymbol{\Psi}_d^{\max}$, $\boldsymbol{\Psi}_1^{\min}, \cdots, \boldsymbol{\Psi}_d^{\min}$	공분산 행렬의 단위직교고유벡터. 문맥에 따라 공분산 행렬이 주어진다.	(1), (25), (30)
$\langle \cdot, \cdot \rangle$	벡터의 내적(inner product).	
$F_j(\cdot \mid \mathrm{w})$	지점 s_j의 강수량의 습윤일 조건부 누적분포함수.	(5)
$F_j^{\leftarrow}(\cdot \mid \mathrm{w})$, $\widetilde{F}_j^{\leftarrow}(\cdot \mid \mathrm{w})$	$F_j(\cdot \mid \mathrm{w})$의 분위수 함수와 근사함수	(8), (20)
N, Φ	정규분포, 표준정규분포의 분포함수	
det	행렬식	
G, f, GP, BC	공간극단값분포, profile 함수, 일반화 Pareto 분포, Box-Cox 변환	(12), (17), (21)
C, $(\mathrm{U}_1, \cdots, \mathrm{U}_d)$	copula와 확률벡터	(6), (7)
$\mathrm{b}_{j,l}$, $j = 1, \cdots, d$, $l = 0, \cdots, k+1$	준경험적 copula 모형의 경계값	(9)
$(\mathrm{Z}_1, \cdots, \mathrm{Z}_d)$, $(\mathrm{Y}_1, \cdots, \mathrm{Y}_d)$	$(\mathrm{U}_1, \cdots, \mathrm{U}_d)$의 Z-변환, Frechet-변환	(10), (14)
$\epsilon_{\max}(s,t)$, $\epsilon_{\min}(s,t)$	지점 s에서 t일의 일최고/최저기온의 anomaly.	(28), (29)

2.1 강수모형과 알고리즘

강수모의는 LARS-WG의 강수모의방식인 습윤/건조기 교대 알고리즘(wet/dry-spell alternation algorithm)을 따른다, (cf. Semenov et al. (1998)). 이 알고리즘에서 각 습윤기와 건조기는 번갈아 가며 독립적으로 생성된다. 이와 대비되는 방식으로 Markov chain 모형이 있다. 이 모형에서는 통상적으로 강수를 두 상태 습윤일(wet-day)와 건조일(dry-day)로 이루어진 chain으로 보고, 두 상태사이의 전이확률로써 chain을 모의생성한다, (cf. Richardson (1981)). Markov 모형과 비교할 때, 습윤/건조기 교대 알고리즘은 model mis-specification[5]에 대해 강건(robust)하다.

각 습윤기/건조기 생성에 있어, 우리는 기본적으로 재추출(re-sampling) 방식을 활용한다. 건조기의 경우, 건조일의 지속기간을 일단위로 생성해 간단하게 모의가 된다. 지속기간생성은 경험분포함수(empirical distribution function)를 이용할 수 있겠다. 반면에, 습윤기 생성은 유역의 강수량을 생성해야 하기 때문에 건조기 생성보다 훨씬 복잡하다. 다음 소절들에서 다지점 강수모형을 위한 통계모형을 소개하고자 한다.

2.1.1 은닉상태모형

우리는 습윤일의 다지점 강수량을 특징짓는 개념으로 '평활된(smoothed) 강수량공간분포'를 주목한다. 이는 강수량의 다지점 상세화와 관련성이 깊은 개념이다. 기후역학모형에서 강수량은 보통 유역 또는 그 보다 큰 격자규모에서 평균으로 주어지는데, 다지점 상세화에서 이를 다시 공간적으로 나누어야 한다. (이를 spatial dis-aggregation이라고 일컫는다.) 이 강수량의 공간평균이 '평활된 강수량공간분포'의 일종이다. 다지점 상세화 기법으로, 유역 내 지점사이의 강수량 통계모형으로부터 다수의 다지점 강수량 표본을 생성하여 이 중 기후모형의 유역강수량과 같은 공간평균을 지니는 표본을 선정하는 방식을 고려할 수 있다, (cf. Hwang and Graham (2013)). 우리의 강수량 생성 방식도 기본적으로 이와 유사한 기법을 고려한다.

---

5) model mis-specification: 일반적으로 모형을 real world model와 다르게 가정한 경우는 의미한다. 여기서는 실제 강수가 Markov 모형을 따르지 않는 경우를 뜻한다. 참고로, 모형 간 차이를 model bias라고 부른다.

유역의 평균으로 강수량공간분포를 평활하는 대신에, 우리는 공간분포의 특징을 더 잘 나타낼 수 있는 방식인 spectral 분해(spetral decomposition)를 고려한다. 다지점 강수량 $\mathbf{R}=(\mathrm{R}(s_1),\cdots,\mathrm{R}(s_d))$의 습윤일의 조건부 공분산 행렬은 대각화(diagonalization)를 통해

$$[\boldsymbol{\Psi}_1 \cdots \boldsymbol{\Psi}_d]\begin{bmatrix}\lambda_1 & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \lambda_d\end{bmatrix}[\boldsymbol{\Psi}_1 \cdots \boldsymbol{\Psi}_d]' \tag{1}$$

으로 표현된다. 여기서, $\lambda_1 \geq \lambda_2 \geq \cdots, \lambda_d \geq 0$는 공분산 행렬의 고유치(eigen value)이고, $\boldsymbol{\Psi}_1,\cdots,\boldsymbol{\Psi}_d$는 그에 대응하는 단위직교고유벡터(orthonormal eigen vector), 공간적 해석에서 단위직교의 공간모드(spatial mode)이다. 그러면

$$\mathbf{R}=\langle\mathbf{R},\boldsymbol{\Psi}_1\rangle\boldsymbol{\Psi}_1+\cdots+\langle\mathbf{R},\boldsymbol{\Psi}_d\rangle\boldsymbol{\Psi}_d \tag{2}$$

으로 표현되는데, 여기서 처음 주요한 모드의 선형결합

$$\langle\mathbf{R},\boldsymbol{\Psi}_1\rangle\boldsymbol{\Psi}_1+\cdots+\langle\mathbf{R},\boldsymbol{\Psi}_p\rangle\boldsymbol{\Psi}_p, \quad p<d \tag{3}$$

을 '평활된 강수량공간분포'로 볼 수 있다. 우리는 이를 간단히 $\mathbf{R}$의 은닉상태(hidden state)라고 부르고, 벡터 $(\langle\mathbf{R},\boldsymbol{\Psi}_1\rangle,\cdots,\langle\mathbf{R},\boldsymbol{\Psi}_p\rangle)$로 표현한다. 유역공간평균은 $p=1$ 그리고 $\boldsymbol{\Psi}_1=(1/d,\cdots,1/d)'$으로 정한 경우에 대응한다.

그림 2는 6월의 낙동강 유역의 공간모드를 남에서 북으로의 형태로 나타내고 있다. 첫 번째와 두 번째 모드가 전체 변동의 87%를 차지하며, 전자는 유역 전체의 고른 분포를 후자는 구미지점을 기준으로 남북으로 대비되는 분포를 대표한다. 이는 한반도를 거치는 강수역대(precipitation band)가 대체로 낙동강 유역 전체를 고루 덮지만, 간혹 남부나 북부 중심으로 덮을 때도 있음을 시사한다. (이는 유역이 남북으로 길게 자리잡고 있기 때문일 것이라 생각된다.) 그림 3은 실제 관측기간의 다지점 강수량 (2000년 6월 27일)과 식 (3)의 $p=2$로 두고 얻어진 평활된 공간분포를 나타내고 있다. 이 날 강수역대는 낙동강 유역 전체를 덮고는 있으나 그 중심이 북쪽으로 치우쳐 있었던 것으로 판단된다.

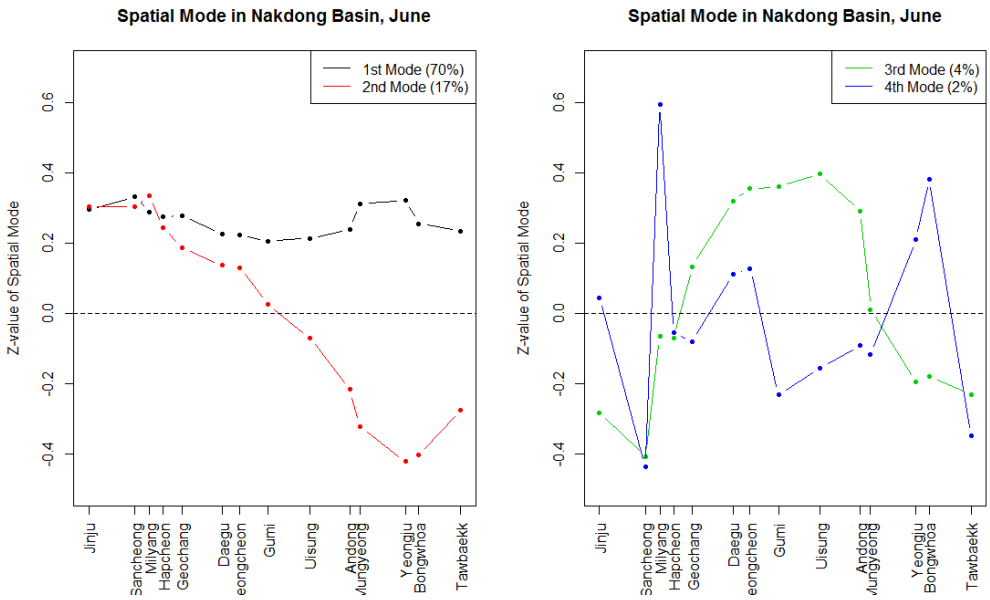


그림 2 6월 낙동강 유역의 주요한 공간모드: 첫 번째 모드는 유역전체 고른 분포를 두 번째는 구미를 기준으로 남북방향(진주(Jinju)가 최남, 태백(Tawbaekk)이 최북)으로 대비되는 분포를 나타내고 있다. 나머지 모드는 해석이 분명하지 않다.

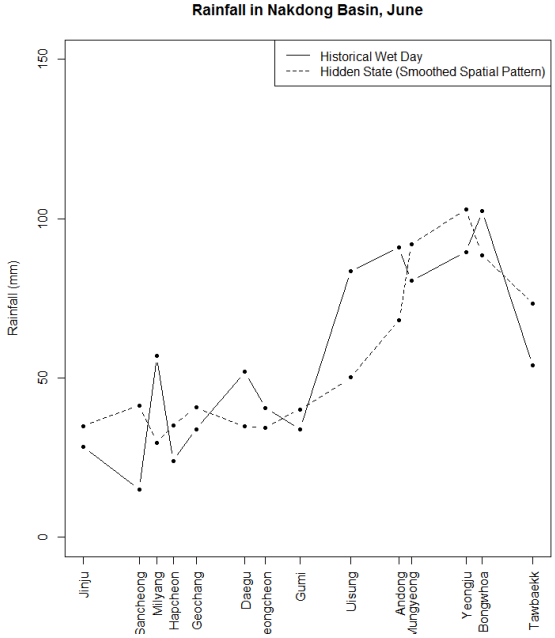


그림 3 2000년 6월 27일의 다지점 강수량 (실선)과 평활된 강수량공간분포 (점선). 강수역대가 유역 북부를 중심으로 위치한 것으로 판단됨.

이 후에 강수량 생성 알고리즘에서 다지점 강수량 사이의 근접성(nearness)을 고려할 것이다. 이 근접성의 측도로 은닉상태사이의 유클리드 거리를 쓴다: 두 다지점 강수량 $\mathbf{R}_1$과 $\mathbf{R}_2$에 대해서,

$$\rho(\mathbf{R}_1,\mathbf{R}_2)=\left\{\sum_{j=1}^{p}\left(\langle\mathbf{R}_1,\boldsymbol{\Psi}_j\rangle-\langle\mathbf{R}_2,\boldsymbol{\Psi}_j\rangle\right)^2\right\}^{1/2} \tag{4}$$

를 쓴다. 이는 평활된 강수량공간분포 사이의 평균제곱거리와 일치한다.

2.1.2 Copula 모형

앞서 강수량공간분포를 평활하기 위해 $\mathbf{R}$의 spectral 분해 (2)를 고려하였지만, $\langle\mathbf{R},\boldsymbol{\Psi}_1\rangle,\cdots,\langle\mathbf{R},\boldsymbol{\Psi}_d\rangle$은 서로 독립이 아니기 때문에, 다지점 강수량의 생성모형으로 쓸 수 없다. 우리는 여기서 다지점 강수량 생성을 위한 copula 모형을 소개한다.

Copula란 모든 주변확률분포가 0과 1사이의 균일분포(uniform distribution)인 다변량 분포의 총칭이다. Sklar (1959)에 따르면, 모든 다변량 분포는 그것의 주변확률분포와 copula의 합성으로 표현할 수 있다. 다지점 강수량 $\mathbf{R}=(\mathrm{R}(s_1),\cdots,\mathrm{R}(s_d))$에 대해 이를 적용하면,

$$F_j(x|\mathrm{w})=\Pr(\mathrm{R}(s_j)\le x\,|\,\text{wet day}),\qquad x\in\mathbb{R},\quad j=1,\cdots,d \tag{5}$$

이 각 지점의 유역의 습윤일(wet day)에 대한 조건부확률분포를 나타낼 때, 어떤 copula C에 대해

$$\Pr(\mathrm{R}(s_1)\le x_1,\cdots,\mathrm{R}(s_d)\le x_d|\,\text{wet day})=\mathrm{C}(F_1(x_1|\mathrm{w}),\cdots,F_d(x_d|\mathrm{w})),\quad x_1,\cdots,x_d\in\mathbb{R}, \tag{6}$$

가 성립한다. 이는 다지점 강수량 분포에 대한 모형에서 지점별 확률분포와 결합확률분포의 모형을 구분할 수 있음을 의미한다. 더 나아가, 모의생성에서는,

$$(\mathrm{U}_1,\cdots,\mathrm{U}_d)\sim\mathrm{C} \tag{7}$$

라고 할 때, $\mathrm{R}^*(s_j)=F_j^{\leftarrow}(\mathrm{U}_j|\mathrm{w})$, $j=1,\cdots,d$ 라고 둔다면, $(\mathrm{R}^*(s_1),\cdots,\mathrm{R}^*(s_d))$의 확률분포는 $\mathbf{R}$의 습윤일에 대한 조건부확률분포와 같다. 여기서,

$$F_j^{\leftarrow}(u|\mathrm{w}) := \min\{x : F_j(x|\mathrm{w}) \geq u\}, \quad 0 < u \leq 1, \tag{8}$$

으로 분위수 함수(quantile function)이라고 부른다.

우리는 copula C에 대해 다음 두 가지 모형을 고려한다:

ⅰ. **준경험적(semi-empirical) Copula 모형**: 우리는 $(0,1]^d$을 서로 겹치지 않는 d차원 직사각구역

$$\mathrm{B}(l_1,\cdots,l_d) = \left(\mathrm{b}_{1,l_1-1},\mathrm{b}_{1,l_1}\right]\times\cdots\times\left(\mathrm{b}_{d,l_d-1},\mathrm{b}_{d,l_d}\right], \quad l_1,\cdots,l_d \in \{1,\cdots,k+1\}$$

으로 나눈다. 단, 고정된 $k\in\mathbb{N}$에 대해,

$$0 = \mathrm{b}_{j,0} < \mathrm{b}_{j,1} < \cdots < \mathrm{b}_{j,k+1} = 1, \quad j = 1,\cdots,d, \tag{9}$$

이다. 준경험적 모형에서는 각 직사각구역 내에서는 $\mathrm{U}_1,\cdots,\mathrm{U}_d$ 들이 조건부 독립이라고 가정한다.

ⅱ. **Gaussian Copula 모형**: 이 모형은 parametric 모형으로,

$$\mathrm{Z}_j := \Phi^{-1}(\mathrm{U}_j) \tag{10}$$

라고 할 때, $(\mathrm{Z}_1,\cdots,\mathrm{Z}_d)$가 다변량 정규분포를 따른다고 가정한다, 즉,

$$(\mathrm{Z}_1,\cdots,\mathrm{Z}_d) \sim \mathrm{N}(\mathbf{0},\Sigma). \tag{11}$$

여기서, Φ는 표준정규분포의 분포함수를 뜻한다.

준경험적 모형은 LARS-WG의 강수량 생성모형인 준경험적확률분포(semi-empirical distribution)방식의 다지점 확장으로 간주할 수 있다, (cf. Racsko et al. (1991)). 이 모형은 비모수적 모형으로 유연하고 비교적으로 손쉽게 다룰 수 있다. 준경험적 copula 모형 적합은 $(\mathrm{U}_1,\cdots,\mathrm{U}_d)$가 각 직사각구역에서 발생할 확률을 추정하는 것으로 수행된다. 그런데, d가 커질수록 직사각구역의 수는 기하급수적으로 늘어나서, 상당수의 직사각구역 확률추정값이 0이 된다. 이는 모의생성에서 해당직사각구역에서는 $(\mathrm{U}_1,\cdots,\mathrm{U}_d)$가 나타나지 않는다는 것을 의미하며, 기상의 다양성을 추구하는데 걸림돌이 된다. Gaussian copula 모형을 통해 준경험적 모형의 결점으로부터 벗어날

수 있다.

지점의 일강수량 분포 $F_j(x|\mathrm{w})$는 $x=0$에서 불연속이기 때문에, Gaussian copula 모형의 $(\mathrm{Z}_1,\cdots,\mathrm{Z}_d)$가 불완전하게 관측된다. 그래서 (11)의 Σ 추정에 이러한 관측상황을 반영해야 하는데, 자세한 추정방법은 부록에서 소개한다.

2.1.3 극단값 모형

앞서, 다지점 강수의 공간적 상관을 설명하기 위해서 copula 모형을 제시하였다. 하지만, 이들 모형이 극단영역에서의 공간확률분포 특성을 설명하기에는 적절하지 않을 수 있다. 이에 대비하기 위해, 공간극단값 분포(spatial extremes distribution)를 고려한다. d개 지점 공간 극단값분포 G의 형태는 다음과 같다:

$$\mathrm{G}(y_1,\cdots,y_d)=\exp\left[-\int_{w\in\mathbb{R}^2}\max\left(\frac{\mathrm{f}(w;s_1)}{y_1},\cdots,\frac{\mathrm{f}(w;s_d)}{y_d}\right)dw\right], \tag{12}$$

여기서, f는 양함수(positive function)로서, 모든 지점 s에 대해서

$$\int_{w\in\mathbb{R}^2}\mathrm{f}(w;s)dw=1 \tag{13}$$

를 만족한다. f를 G의 'profile 함수'라고 부르며, G와 동일시 할 수 있다. $(\mathrm{U}_1,\cdots,\mathrm{U}_d)$의 극단영역에서의 확률분포를 설명하기 위해서,

$$\mathrm{Y}_j=-\left\{\log \mathrm{U}_j\right\}^{-1} \tag{14}$$

이라고 두자. 그러면 $(\mathrm{Y}_1,\cdots,\mathrm{Y}_d)$가 max-stability 조건을 만족할 경우, 어떤 profile 함수 f가 존재하여, $\min(y_1,\cdots,y_d)\to\infty$ 일 때,

$$\Pr(\mathrm{Y}_1>y_1 \text{ or } \cdots \text{ or } \mathrm{Y}_d>y_d)\sim 1-\mathrm{G}(y_1,\cdots,y_d). \tag{15}$$

이는 $(\mathrm{Y}_1,\cdots,\mathrm{Y}_d)$의 극단영역에서 확률분포가 G로 근사됨을 의미한다. 공간극단값분포에 대한 자세한 내용은 Resnick (1987), Smith (1900), Coles et al. (2001), Davison and Gholamrezaee (2012) 를 참고하기를 바란다.

G 에서 profile 함수 f는 극한강수량의 공간적 관계를 설명하기 때문에, f 추정은 공간극단값 분석에서 가장 주요한 부분이다. 그런데, f의 추정을 위해서는 그 형태에 대한 적절한 모형이 필요하다. 여기서는 대표적인 모형으로 Gaussian 모형을 소개한다, (cf. Smith (1990)). 이 모형에서 f가 다변량 정규분포의 밀도함수이라고 가정한다:

$$\mathrm{f}(w;s)=\frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left\{-\frac{1}{2}(w-s)'\Sigma(w-s)\right\}. \tag{16}$$

Smith는 이 모형 하에서 연최대값을 이용한 Σ의 최소제곱추정을 고려하였다. 우리는 Smith의 추정법 대신, 강수량 peak 자료를 이용한 추정법을 고려한다. 자세한 추정법은 부록에서 소개한다.

극단영역에서 다른 문제로 관측분위수의 낮은 정확성을 들 수 있다. 분포의 중심부에서와 달리 극단영역에서는 관측자료의 수가 적기 때문에 관측분위수의 정확도가 떨어진다. 우리는 이에 대한 대비책으로 일반화 Pareto 모형을 고려한다. 이 모형을 설명하기에 앞서, 일반화 Pareto 분포(Generalized Pareto distribution)를 간단히 소개하고자 한다. 이름에서 알 수 있듯이 일반화 Pareto 분포는 기존의 Pareto 분포를 확장하여 얻어진 확률분포이다. 분포함수의 형태는

$$\mathrm{GP}(z;\sigma,\gamma):=\begin{cases}0, & z<0;\\ 1-\left(1+\gamma\dfrac{z}{\sigma}\right)^{-1/\gamma}, & z\geq 0,\ 1+\gamma\dfrac{z}{\sigma}>0;\\ 1, & z\geq 0,\ 1+\gamma\dfrac{z}{\sigma}\leq 0,\end{cases} \tag{17}$$

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

$$\mathrm{GP}(z;\sigma,0):=\begin{cases}0, & z<0;\\ 1-\exp\left(-\dfrac{z}{\sigma}\right), & z\geq 0,\end{cases} \tag{18}$$

각 모수는 분포함수에서 역할에 따라, σ는 척도모수(scale parameter), γ는 형태모수(shape parameter)로 불리고 있다. 형태모수인 γ가 0보다 큰 경우가 Pareto 분포이다. 강수량의 해석에서는, 두 모수의 값이 커지면 커질수록 강한 비가 내릴 확률이 높다는 것을 의미한다.

일반화 Pareto 모형은 주어진 threshold의 값을 넘는 변량의 초과량이 근사적으로 일반화 Pareto 분포를 따른다는 확률 이론에 기인한다.(Coles et al. (2001) Theorem 4.1 참조). 충분히 큰 threshold x_0 대해서

$$\frac{1-F_j(x|\mathrm{w})}{1-F_j(x_0|\mathrm{w})} \approx 1-\mathrm{GP}(x-x_0;\sigma_j,\gamma_j), \qquad x > x_0, \ j=1,\cdots,d \tag{19}$$

을 만족하는 $\sigma_j > 0$와 $\gamma_j \in \mathbb{R}$이 존재하는 것이 알려져 있고, 위의 근사식은 x_0가 커질수록 정확해 진다. 식 (19)로부터, $F_j^{\leftarrow}(\cdot|\mathrm{w})$의 근사함수 $\widetilde{F}_j^{\leftarrow}(\cdot|\mathrm{w})$을 구할 수 있다:

$$\widetilde{F}_j^{\leftarrow}(u|\mathrm{w}) = \begin{cases} \sigma_j \,\mathrm{BC}((1-\mathrm{b_{j,k}})/(1-\mathrm{u});\gamma_\mathrm{j}) + \mathrm{F_j^{\leftarrow}}(\mathrm{b_{j,k}}|\mathrm{w}), & u > \mathrm{b_{j,k}}; \\ \mathrm{F_j^{\leftarrow}}(\mathrm{u}|\mathrm{w}), & u \le \mathrm{b_{j,k}}, \end{cases} \tag{20}$$

여기서, 함수 $\mathrm{BC}(\cdot;\gamma)$는 Box-Cox 변환

$$\mathrm{BC}(y;\gamma) = \begin{cases} \dfrac{y^\gamma - 1}{\gamma}, & \gamma \neq 0; \\ \log(y), & \gamma = 0, \end{cases} \qquad y > 0, \tag{21}$$

을 나타낸다. σ_j와 γ_j은 최대우도법으로 추정한다. 추정의 자세한 방법은 Coles et al. (2001)의 4.3.2.절에 소개되어 있다. 앞서 언급하였듯이, $F_j^{\leftarrow}(\cdot|\mathrm{w})$는 실제에서 관측분위수로 대체될 것인데 극단영역에서는 정확도가 떨어진다. $\widetilde{F}_j^{\leftarrow}(\cdot|\mathrm{w})$의 적합은 관측분위수의 그러한 결점을 보완해 준다.

2.1.4 습윤기와 강수 시나리오의 생성 알고리즘

이제, 습윤기 모의 생성 알고리즘을 제안하고자 한다. 습윤기 생성은 과거 관측기간에서 습윤기를 무작위로 추출하는 것으로 시작한다. 우리는 추출된 습윤기의 강수량공간분포 평활형태를 지닌 다지점 강수량을 생성하고자 한다. 습윤일의 강수량 생성알고리즘은 copula 모형에 따라 구분된다. 우선, 준경험적 copula 모형의 알고리즘을 제시한다. 알고리즘 설명의 편의를 위해,

$$\mathrm{L}(s_j) := \sum_{l=1}^{k+1} l \cdot \mathrm{I}\big(\mathrm{b}_{j,l-1} < F_j(\mathrm{R}(s_j)|\mathrm{w}) \le \mathrm{b}_{j,l}\big), \quad j = 1,\cdots,d, \tag{22}$$

그리고, $\mathbf{L} := (\mathrm{L}(s_1),\cdots,\mathrm{L}(s_d))$으로 정의한다. 여기서, 등장하는 $\mathrm{b}_{j,l}$은 준경험적 copula 모형에서 도입된 경계값이다. (식 (9)를 참조.) 본 보고서에서는 $\mathrm{L}(s)$를 지점 s의 '수준', 그리고 $\mathbf{L}$를 '수준벡터'라고 칭한다. 그리고, 어느 지점에서 수준이 $k+1$인 경우를 '극단'이라고 부른다. 준경험적 copula 모형에서 수준벡터 $\mathbf{L}$을 기반으로 다지점 강수량을 생성한다. 구체적인 알고리즘 내용은 다음과 같다.

준경험적 Copula 모형하에서 습윤기 생성 알고리즘

1. 과거 관측기간에서 습윤기 $\mathbf{R}(t_0+1),\cdots,\mathbf{R}(t_0+m)$를 무작위로 추출한다.
2. k−NN 방법을 이용해, $i=1,\cdots,m$에 대해서, $\mathbf{R}(t_0+i)$의 nearest neighbor $\mathbf{R}(t_i)$를 선택한다. 여기서, 근접성측도로 (4)를 쓴다.
3. $i=1,\cdots,m$에 대해, $\mathbf{R}(t_i)$의 수준벡터 $\mathbf{L}(t_i)$를 계산한다. (식 (22) 참조.)
4. $i=1,\cdots,m$에 대해, $\mathbf{L}^* = \mathbf{L}(t_i)$라고 두고 하위 알고리즘을 수행한다.

하위알고리즘: 준경험적 Copula 모형하에서 다지점 강수량 생성

수준 $\mathbf{L}^* = (l_1,\cdots,l_d)$으로 주어졌다고 하자. 만약 극단이 아니라면,

1. $j=1,\cdots,d$에 대해서, $\left[\mathrm{b}_{j,l_j-1},\mathrm{b}_{j,l_j}\right]$의 균일 분포에서 U_j^*를 생성한다. 단, $\mathrm{U}_1^*,\cdots,\mathrm{U}_d^*$은 독립적으로 생성된다;
2. $j=1,\cdots,d$에 대해서 $\mathrm{R}^*(s_j) = F_j^{\leftarrow}(\mathrm{U}_j^*|\mathrm{w})$라고 둠으로써, 다지점 강수량 $(\mathrm{R}^*(s_1),\cdots,\mathrm{R}^*(s_d))$을 생성한다.

반대로 극단이라면, 다음 단계를 수행한다:

1. 공간극단값 분포 (12)의 $[0,\infty)^d \setminus \left\{\times_{j=1}^{d}[0,-1/\log \mathrm{b}_{j,k})\right\}$상에서 조건부분포로부터 $\mathrm{Y}_1^*,\cdots,\mathrm{Y}_d^*$를 생성한다;
2. $j=1,\cdots,d$에 대해서 $\mathrm{R}^*(s_j) = \widetilde{F}_j^{\leftarrow}\left[\exp\left(-1/\mathrm{Y}_j^*\right)|\mathrm{w}\right]$라고 둠으로써, 다지점 강수량 $(\mathrm{R}^*(s_1),\cdots,\mathrm{R}^*(s_d))$을 생성한다;
3. 각 $j=1,\cdots,d$에 대해서, $\mathrm{b}_{j,l_j-1} < F_j(\mathrm{R}^*(s_j)|\mathrm{w}) \le \mathrm{b}_{j,l_j}$가 성립하다면, 알고리즘을 끝낸다. 그렇지 않다면, 1로 돌아간다.

다음으로 Gaussian Copula 모형에서 습윤기 생성 알고리즘을 제안한다. 준경험적 copula 모형과 다르게, 추출된 습윤일의 다지점 강수량을 기반으로 새롭게 강수량을 생성한다.

Gaussian Copula 모형하에서 습윤기 생성 알고리즘

1. 과거 관측기간에서 습윤기 $\mathbf{R}(t_0+1),\cdots,\mathbf{R}(t_0+m)$를 무작위로 추출한다.
2. $i=1,\cdots,m$에 대해, $\mathbf{R}^\star=\mathbf{R}(t_0+i)$라고 두고 하위알고리즘을 수행한다.

하위알고리즘: Gaussian Copula 모형하에서 다지점 강수량 생성

반복횟수 N을 정한다. 다지점 강수량 $\mathbf{R}^\star$가 주어지고, 그것의 수준벡터를 $\mathbf{L}^\star$라고 하자. 만약 $\mathbf{L}^\star$가 극단이 아니라면, $i=1,\cdots,N$에 대해서 다음 두 단계를 반복한다:

1. $\mathrm{N}(\mathbf{0},\Sigma)$로부터 $(\mathrm{Z}^*_{i,1},\cdots,\mathrm{Z}^*_{i,d})$를 생성한다. 단, Σ는 식 (11)에서 주어진다;
2. $\mathrm{R}^*_i(s_j)=F_j^{\leftarrow}(\Phi(\mathrm{Z}^*_{i,j})|\mathrm{w})$라고 둔다.

이 두 단계를 통해 pool $\left\{(\mathrm{R}^*_i(s_1),\cdots,\mathrm{R}^*_i(s_d)):i=1,\cdots,N\right\}$을 구성한다.

3. 이 pool에서 $\mathbf{R}^\star$의 nearest neighbor 를 선정한다. 단, 근접성의 측도는 (4)를 쓴다.

반대로 $\mathbf{L}^\star$가 극단이라면, 다음 두 단계를 통해 pool을 구성한다.

1. 공간극단값 분포 (12)의 $[0,\infty)^d\setminus\left\{\times_{j=1}^d[0,-1/\log \mathrm{b}_{j,k})\right\}$상에서 조건부분포로부터 $\mathrm{Y}^*_{i,1},\cdots,\mathrm{Y}^*_{i,d}$를 생성한다;
2. $\mathrm{R}^*_i(s_j)=\widetilde{F}_j^{\leftarrow}(\exp(-1/\mathrm{Y}^*_{i,j})|\mathrm{w})$라고 둔다.

앞서처럼 pool $\left\{(\mathrm{R}^*_i(s_1),\cdots,\mathrm{R}^*_i(s_d)):i=1,\cdots,N\right\}$을 구성한 다음 단계를 따른다:

3. k-NN 방식으로 pool에서 $\mathbf{R}^\star=(\mathrm{R}^\star(s_1),\cdots,\mathrm{R}^\star(s_d))$의 nearest neighbor를 선정한다. 단, 근접성의 측도로는

$$\sum_{j=1}^{d}\left\{F_j(\mathrm{R}^\star(s_j)|\mathrm{w})-F_j(\mathrm{R}^*_i(s_j)|\mathrm{w})\right\}^2,\qquad i=1,\cdots,N$$

를 쓴다.

마지막으로 습윤기 생성 알고리즘을 기반으로 한 강수 시나리오 생성 알고리즘을 제시한다. 그림 4는 강수 시나리오 생성 알고리즘의 개념도이다.

강수 시나리오 생성 알고리즘

모의 기간이 $t_0+1,\cdots,t_0+n_{\text{sim}}$ 으로 주어졌다고 하자. 변수 t^* 의 값을 t_0으로 두고, '상태'를 '건조'와 '습윤' 중 하나를 무작위로 정한다. $\mathbf{R}(t_0+1),\cdots,\mathbf{R}(t_0+n_{\text{sim}})$의 값이 모두 생성될 때까지, 다음 단계를 반복한다.

1. $t^* \geq t_0+n_{\text{sim}}$ 이라면, 알고리즘을 멈춘다.
2. 상태가 '건조'라면,
   가. 관측기간에서 건조기를 무작위로 선정한다.
   나. 건조기의 지속일이 m이라면, $t=t^*+1,\cdots,t^*+m$에 대해, $\mathbf{R}(t)=(0,\cdots,0)$ 으로 둔다.
   다. t^* 의 값을 t^*+m으로 바꾸고, 상태를 '습윤'으로 전환하여 단계 1로 간다.
3. 상태가 '습윤'라면,
   가. 습윤기 생성 알고리즘으로 습윤기 $\mathbf{R}(t^*+1),\cdots,\mathbf{R}(t^*+m)$를 생성한다.
   나. t^* 의 값을 t^*+m으로 바꾸고, 상태를 '건조'으로 전환하여 단계 1로 간다.

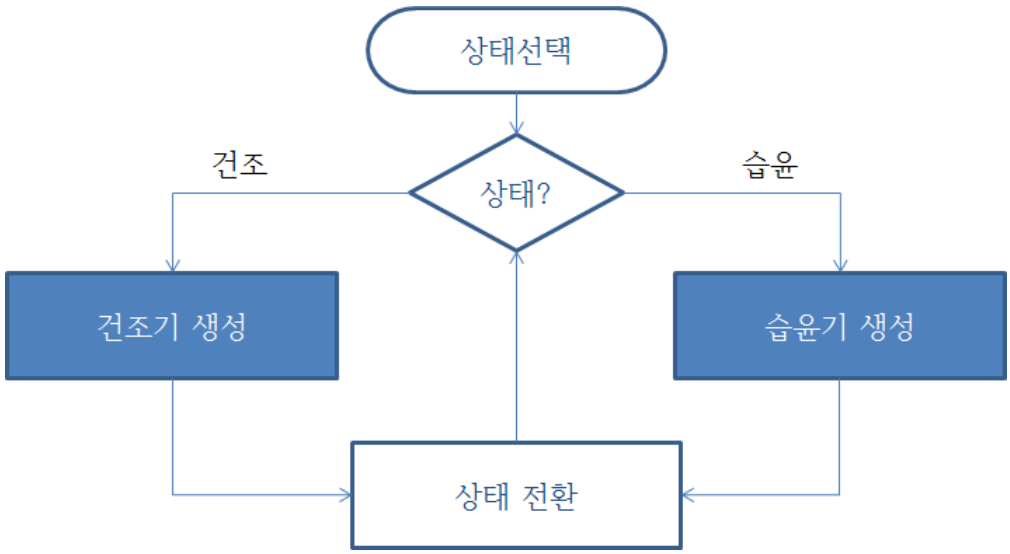


그림 4 강수 시나리오 생성 알고리즘 순서도. 순서도에서 '상태'는 건조상태 또는 습윤상태를 칭하며 '상태전환'은 건조상태에서 습윤상태로 또는 반대방향의 변화를 의미한다.

2.2 기온모형과 알고리즘

2.2.1 일평균 모형

기온은 뚜렷한 연주기의 순환추세를 지니고 있다. 그래서, 기온 모형은 이 연주기 순환추세를 추정하고, 이 추세를 제거한 anomaly를 구하는데서 시작한다. 순환추세를 추정하는 방법으로 local polynomial regression을 고려한다 (cf. Cleveland et al. (1992)): $\mathrm{T}(s,t)$가 지점 s의 t일의 최고 또는 최저기온을 나타낸다고 할 때, $\mathrm{E}\{\mathrm{T}(s,t)\}$는 t의 함수로 대해 매년 반복되는 비선형적인 추세를 드러내기 때문에, 비모수적 방식인 local polynomial regression이 적합하다. 모형의 설명변수로서 Julian day를 고려한다: t일의 Julian day $\tau(t)$는 속한 해에서 해 시작 후 경과일로 정의된다.[6] 또한, 추세가 매해 반복됨을 감안하여, t_1일과 t_2일 간 시차를 아래와 같이 순환형태(circular form)로 재정의한다[7]:

$$\mathrm{d}(t_1,t_2)=\begin{cases}\tau(t_1)-\tau(t_2), & |\tau(t_1)-\tau(t_2)|\le 365-|\tau(t_1)-\tau(t_2)|;\\ \mathrm{sign}\{\tau(t_2)-\tau(t_1)\}(365-|\tau(t_1)-\tau(t_2)|), & |\tau(t_1)-\tau(t_2)|>365-|\tau(t_1)-\tau(t_2)|.\end{cases}$$

그러면, 주어진 t_0일의 $\mathrm{E}\{\mathrm{T}(s,t_0)\}$는, 적절한 kernel 함수 Ker와 $h>0$에 대해 목적함수

$$\sum_{t=1}^{n}\{\mathrm{T}(s,t)-m\}^2\mathrm{Ker}\left(\frac{\mathrm{d}(t,t_0)}{h}\right) \tag{23}$$

를 최소로 하는 m의 값으로 추정된다. (단, 식 (23)에서 $\mathrm{T}(s,1),\cdots,\mathrm{T}(s,n)$은 관측기온값을 뜻한다.) 그림 5는 안동지점에서 local polynomial regression 방식으로 추정된 일최고기온의 일평균 순환추세선이고 회색 점은 실제 관측기온을 나타낸다. 그림에서 확인할 수 있듯이, 신뢰성이 높은 추세선을 얻을 수 있다.

6] 예를 들어, t일이 1월 1일인 경우에 $\tau(t)=1$, 2월 1일인 경우에 $\tau(t)=32$, 그리고 12월 31일 경우에 $\tau(t)=365$이다. 편의상, 2월 29일은 3월 1일로 간주한다.

7] 이 정의에 따르면 t_1일이 1월 1일이고 t_2일이 2월 1일 경우, 연도와 무관하게 차 $\mathrm{d}(t_1,t_2)$는 -31일이고, 2월 1일과 12월 31일의 차는 32일이다.

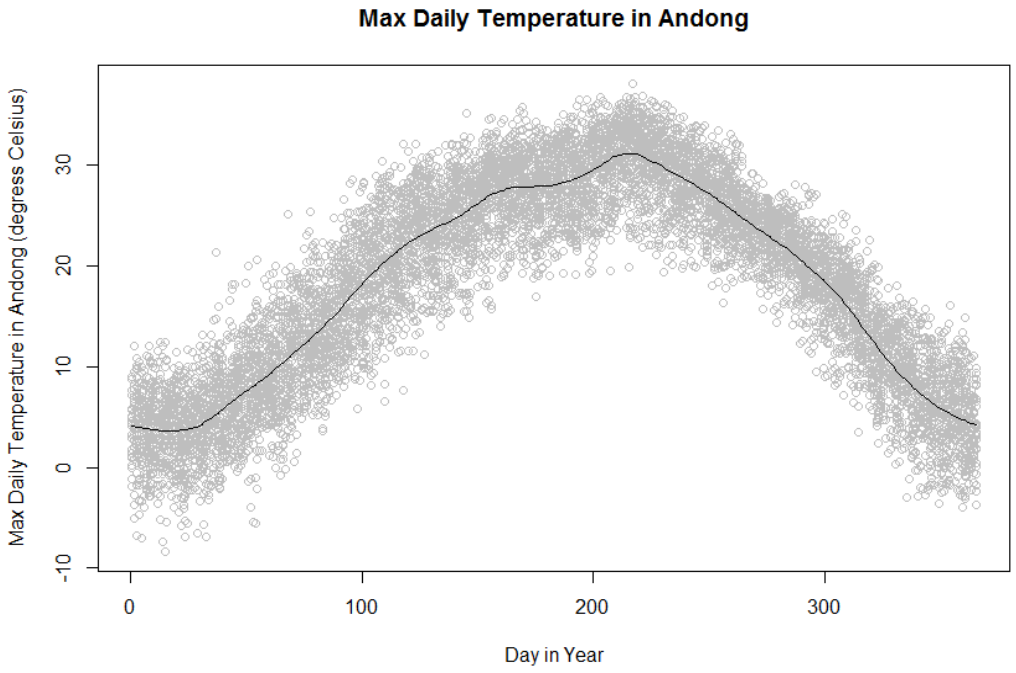


그림 5 Local polynomial regression으로 추정된 안동지점 일최고기온의 일평균 순환추세 (실선). 회색 점은 관측 일최고기온을 표시한다.

기온은 또한 강수에 영향을 받는다. 우리는 이 영향을 강수량에 따른 일평균의 증감으로 정량화한다. 구체적인 방식은 다음과 같다:

$$\mathbf{R}^{\gamma}(t) := \left(\{\mathrm{R}(s_1,t)\}^{\gamma},\cdots,\{\mathrm{R}(s_d,t)\}^{\gamma}\right), \quad \gamma > 0 \tag{24}$$

의 공분산행렬의 대각화를 통해 직교모드 $\boldsymbol{\Psi}_1,\cdots,\boldsymbol{\Psi}_d$를 얻는다. 우리는 기온편차에 대해 아래와 같은 회귀모형을 가정한다:

$$\mathrm{T}(s,t) - \mathrm{E}\{\mathrm{T}(s,t)\} = \sum_{i=1}^{p} \beta_i \langle \mathbf{R}^{\gamma}(t) - \mathrm{E}\{\mathbf{R}^{\gamma}(t)\}, \boldsymbol{\Psi}_i \rangle + (\mathrm{error}), \quad \gamma > 0. \tag{25}$$

여기서,

$$\Delta\{\mathbf{R}(t)\} = \sum_{i=1}^{p} \beta_i \langle \mathbf{R}^{\gamma}(t) - \mathrm{E}\{\mathbf{R}^{\gamma}(t)\}, \boldsymbol{\Psi}_i \rangle \tag{26}$$

라고 둔다면,

$$\mathrm{T}(s,t) = \mathrm{E}\{\mathrm{T}(s,t)\} + \Delta\{\mathbf{R}(t)\} + (\mathrm{error}) \tag{27}$$

로 표현할 수 있다. $\mathrm{E}[\Delta\{\mathbf{R}(t)\}] = 0$이기 때문에, $\mathrm{E}\{\mathrm{T}(s,t)\}$을 climatological mean으로 해석하는 것은 여전히 유효하다. LARS-WG의 기온모형과 대비해서, 모형 (27)은 강수영향력을 연속적으로 정량화시켰다는 면에서 의미가 있다. 그림 6은 실제, 강수의 기온에 대한 영향이 연속적임을 보여 주고 있다. 그림 7은 안동지점에서 회귀모형 (25)로 일최고기온의 추정된 일평균을 보여주고 있다. 흑색점선은 2013년 실제 관측기온, 청색 두꺼운 실선은 local polynomial regression 모형 (23)으로 추정된 순환추세, 적색 두꺼운 실선은 회귀모형 (25)로 추정된 일평균을 나타낸다.

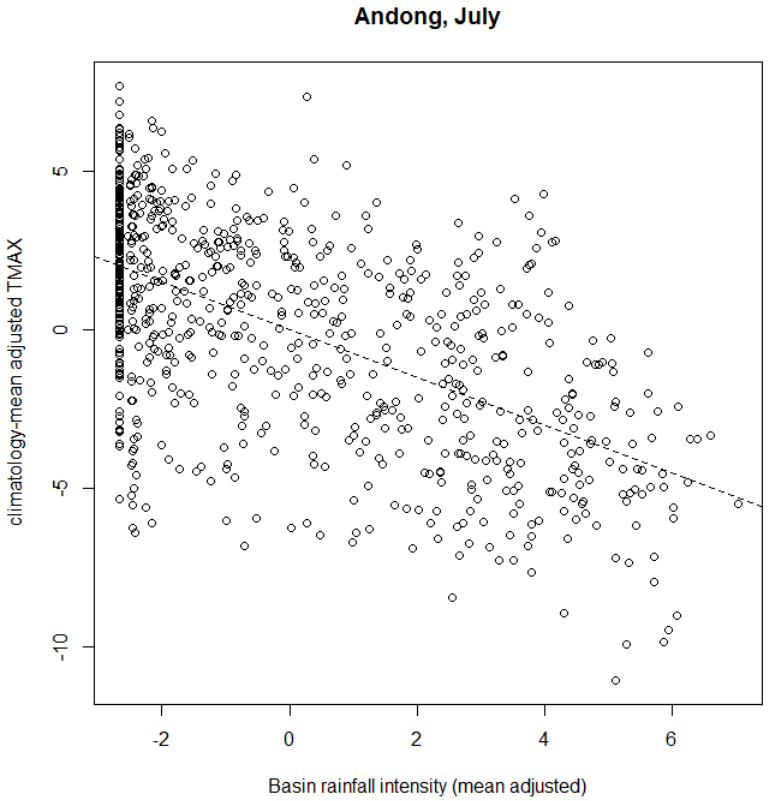


그림 6 안동 7월의 유역 강수 intensity $\langle \mathbf{R}^{\gamma}(\mathrm{t}) - \mathrm{E}\{\mathbf{R}^{\gamma}(\mathrm{t})\}, \boldsymbol{\Psi}_1 \rangle$에 대한 최고기온의 climatological-mean 편차 $\mathrm{T}(s,t) - \mathrm{E}\{\mathrm{T}(s,t)\}$의 관계. 편차 감소가 연속적이다.

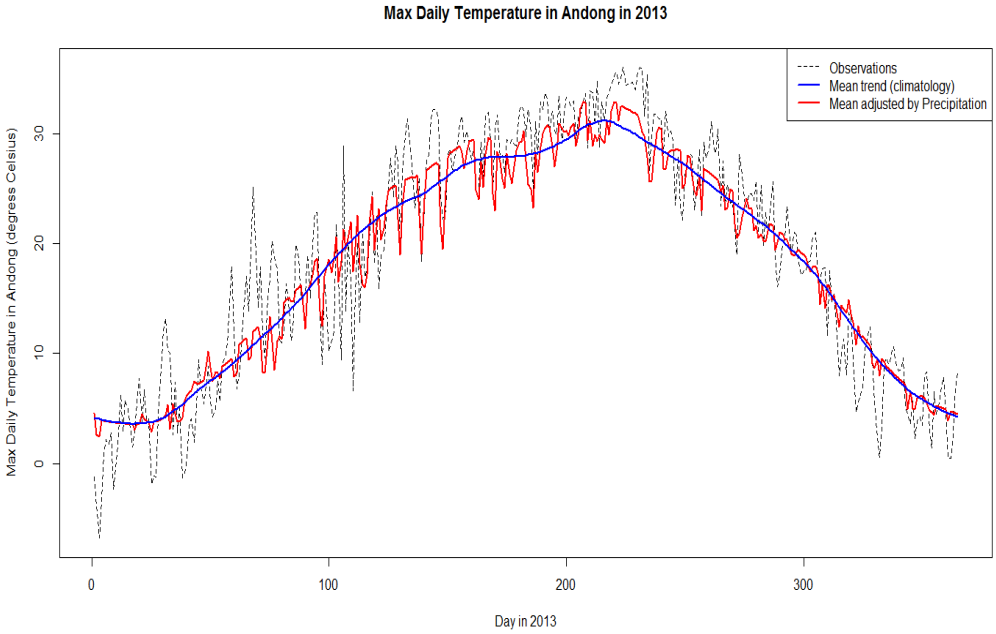


그림 7 안동지점 일최고기온 순환추세와 강수량에 의해 조정된 일평균. 흑색 점선은 2013년도 관측기온, 두꺼운 청색 실선은 순환추세, 두꺼운 적색 실선은 회귀모형으로 추정된 일평균이다.

식 (25)의 회귀모형은 최소제곱추정법으로 지점별로 그리고 계절성을 감안하여 월별로 구분되어 적합된다. 실제 낙동강 유역의 경우에 강수의 기온에 대한 영향력은 계절마다 상이하다. 그림 7에서 나타나듯 여름에는 강수에 따른 기온변화가 큰 반면에 겨울에는 그 영향력이 미미하다. 모형식에서 도입된 $\gamma > 0$는 회귀모형의 선형성을 확보하기 위한 tuning parameter이다.

2.2.2 Anomaly 모형

기온 $\mathrm{T}(s,t)$의 표준편차를 $\sigma(s,t)$라고 하자. 우리는

$$\epsilon(s,t) = [\mathrm{T}(s,t) - \mathrm{E}\{\mathrm{T}(s,t)\} - \Delta\{\mathbf{R}(t)\}]/\sigma(s,t) \tag{28}$$

를 지점 s에 t시점의 anomaly로 정의한다. 이러한 anomaly를 일최고/최저 기온에 각각에 대해 얻을 수 있는데, 이를 구분하기 위해 $\epsilon_{\max}(s,t)$과 $\epsilon_{\min}(s,t)$라고 표기한다.

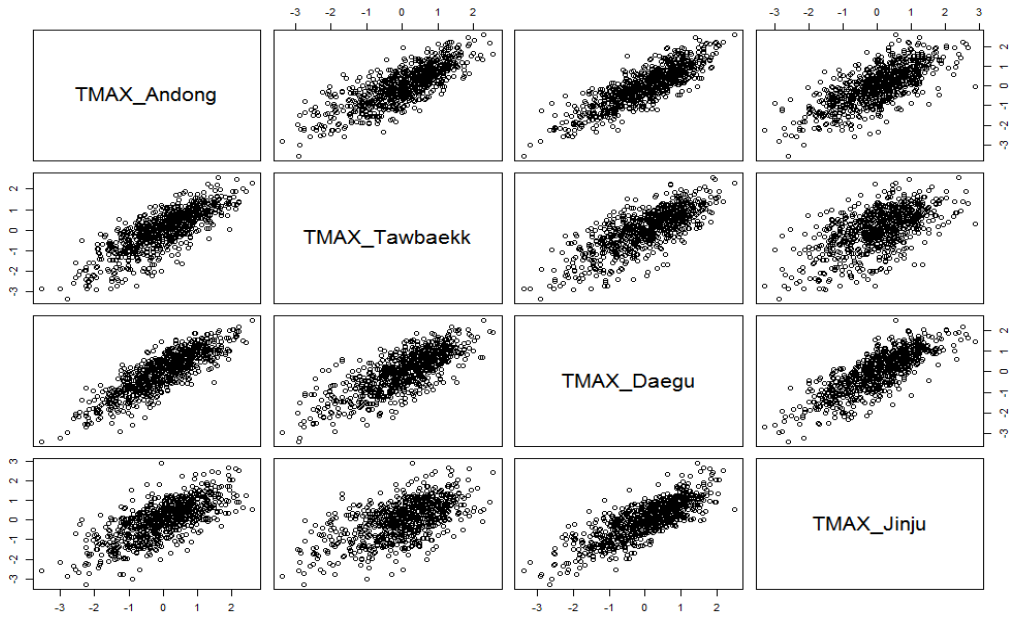


그림 8 6월 안동, 태백, 대구, 진주지점의 일최고기온 anomaly의 공간상관관계를 나타내는 산점도. 뚜렷한 양의 공간상관이 드러난다.

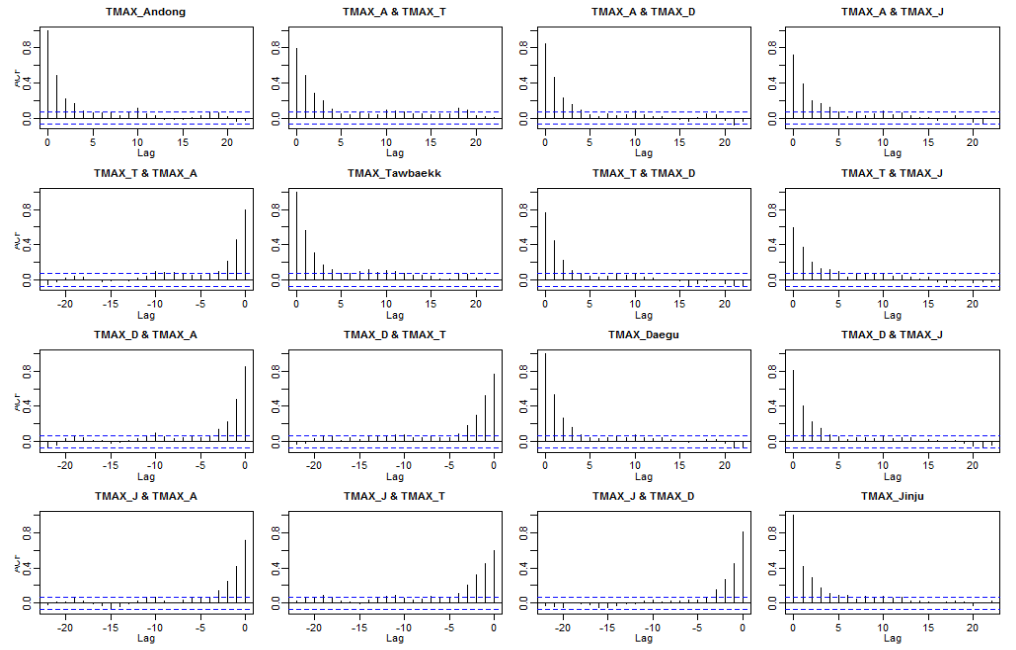


그림 9 6월 안동, 태백, 대구, 진주지점의 일최고기온 anomaly의 (시간적) 자기상관함수그림 (correlogram). 뚜렷한 양의 자기상관이 드러난다.

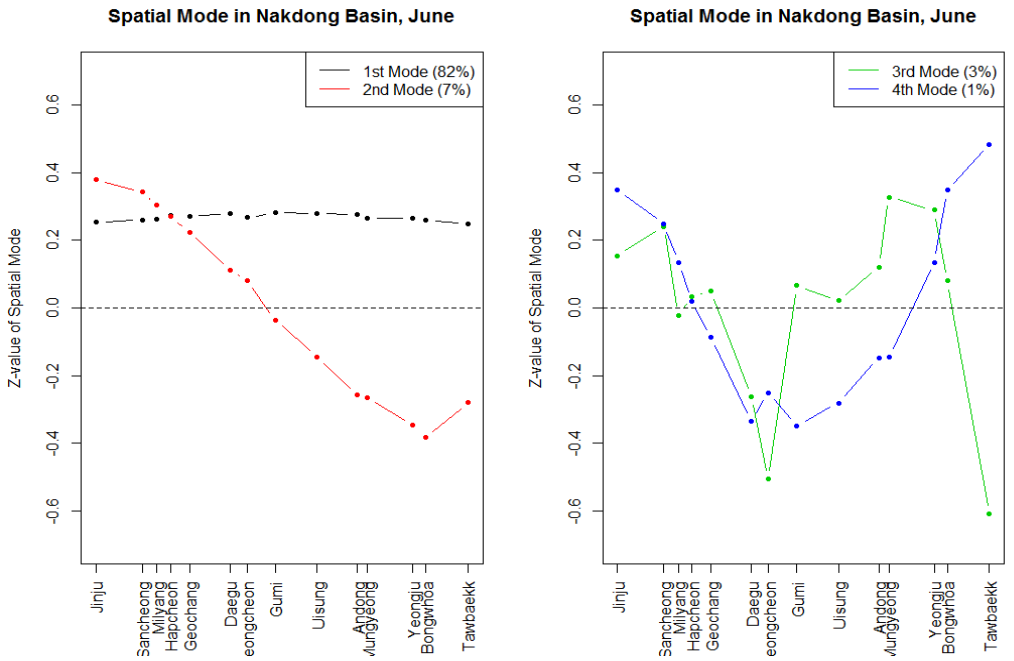


그림 10 6월 낙동강 유역 일최고기온 anomaly의 공간모드

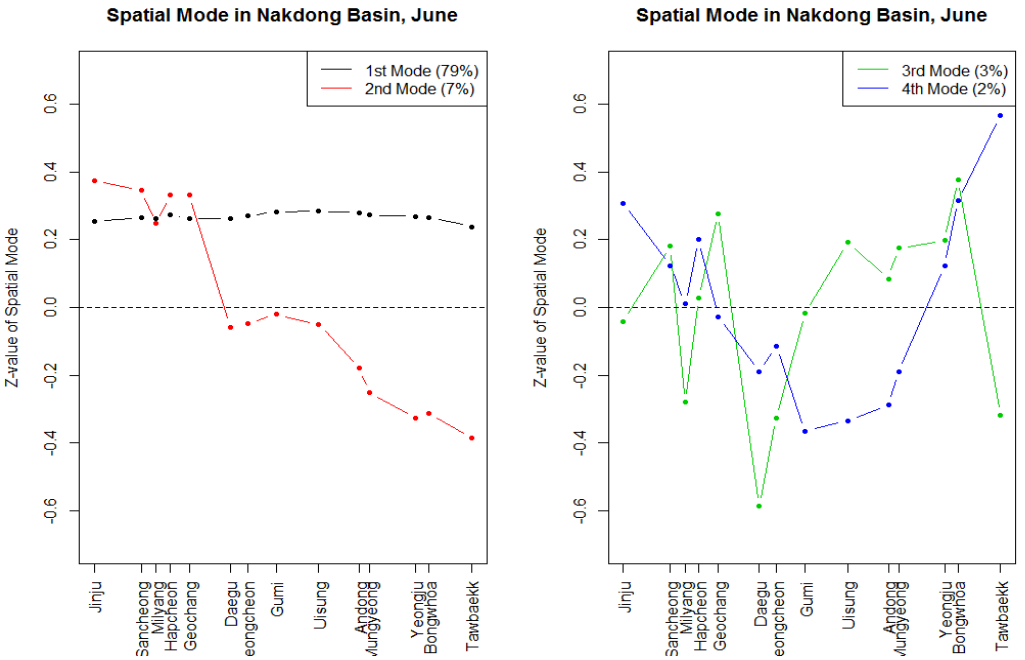


그림 11 6월 낙동강 유역 일최저기온 anomaly의 공간모드

그림 8-9에서처럼, anomaly는 시공간적으로 복잡한 상관관계를 지니는데, 이를 주성분의 벡터자기회귀모형으로 설명할 수 있다.

$$\boldsymbol{\epsilon}_{\max}(t)=(\epsilon_{\max}(s_1,t),\cdots,\epsilon_{\max}(s_d,t)),\ \ \boldsymbol{\epsilon}_{\min}(t)=(\epsilon_{\min}(s_1,t),\cdots,\epsilon_{\min}(s_d,t)) \quad (29)$$

라고 두고, 각각의 공분산행렬을 대각화하여 얻는 직교모드를 각각 $\boldsymbol{\Psi}_1^{\max},\cdots,\boldsymbol{\Psi}_d^{\max}$와 $\boldsymbol{\Psi}_1^{\min},\cdots,\boldsymbol{\Psi}_d^{\min}$라고 표시하자. 그림 10-11은 주요한 직교모드의 형태를 나타낸다. 이 때 주요 q개의 score의 벡터를

$$\begin{aligned}\boldsymbol{X}(t)&=(\boldsymbol{X}_{\max}(t)',\boldsymbol{X}_{\min}(t)')',\\ \boldsymbol{X}_{\max}(t)&=\left(\langle\boldsymbol{\epsilon}_{\max}(t),\boldsymbol{\Psi}_1^{\max}\rangle,\cdots,\langle\boldsymbol{\epsilon}_{\max}(t),\boldsymbol{\Psi}_q^{\max}\rangle\right)',\\ \boldsymbol{X}_{\min}(t)&=\left(\langle\boldsymbol{\epsilon}_{\min}(t),\boldsymbol{\Psi}_1^{\min}\rangle,\cdots,\langle\boldsymbol{\epsilon}_{\min}(t),\boldsymbol{\Psi}_q^{\min}\rangle\right)',\end{aligned} \quad (30)$$

라고 둔다면, 이 벡터시계열에 대해 p차-벡터자기회귀모형을 가정한다, 즉,

$$\boldsymbol{X}(t)=\sum_{i=1}^{p}\boldsymbol{A}_i\,\boldsymbol{X}(t-i)+\mathbf{e}(t),\qquad t\in\mathbb{Z}, \quad (31)$$

단, $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_p$는 자기회귀계수행렬이고 $\{\mathbf{e}(t):t\in\mathbb{Z}\}$는 독립오차항이다. 그러면

$$\begin{aligned}\boldsymbol{\epsilon}_{\max}(t)&=\left[\boldsymbol{\Psi}_1^{\max}\cdots\boldsymbol{\Psi}_q^{\max}\right]\boldsymbol{X}_{\max}(t)+\sum_{i=q+1}^{d}\langle\boldsymbol{\epsilon}_{\max}(t),\boldsymbol{\Psi}_i^{\max}\rangle\boldsymbol{\Psi}_i^{\max}\\ \boldsymbol{\epsilon}_{\min}(t)&=\left[\boldsymbol{\Psi}_1^{\min}\cdots\boldsymbol{\Psi}_q^{\min}\right]\boldsymbol{X}_{\min}(t)+\sum_{i=q+1}^{d}\langle\boldsymbol{\epsilon}_{\min}(t),\boldsymbol{\Psi}_i^{\min}\rangle\boldsymbol{\Psi}_i^{\min}\end{aligned} \quad (32)$$

으로 표현되며, $\left\{\langle\boldsymbol{\epsilon}_{\max}(t),\boldsymbol{\Psi}_i^{\max}\rangle:t\in\mathbb{Z}\right\}$와 $\left\{\langle\boldsymbol{\epsilon}_{\min}(t),\boldsymbol{\Psi}_i^{\min}\rangle:t\in\mathbb{Z}\right\}$, $i=q+1,\cdots,d$, 은 상호 독립인 백색잡음이라고 가정한다.

벡터자기회귀모형 (31)에서 자기회귀계수행렬 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_p$는 최소제곱법으로 추정되며, 관측기간의 오차항 $\{\mathbf{e}(t):t=p+1,\cdots,n\}$에 대한 추정량으로 잔차

$$\hat{\mathbf{e}}(t)=\boldsymbol{X}(t)-\sum_{i=1}^{p}\widehat{\boldsymbol{A}}_i\,\boldsymbol{X}(t-i),\qquad t=p+1,\cdots,n,$$

를 쓴다. 여기서, $\widehat{\boldsymbol{A}}_1,\cdots,\widehat{\boldsymbol{A}}_p$은 최소제곱법으로 추정된 계수행렬이다. 최소제곱법의 자세한 내용은 Helmut (1993)에서 소개되어 있다.

2.2.3 기온 시나리오 생성 알고리즘

기온은 강수량의 영향을 받으므로, 모의기간 $t_0+1,\cdots,t_0+n_{\text{sim}}$의 강수시나리오

$$\{\mathbf{R}(t): t=t_0+1,\cdots,t_0+n_{\text{sim}}\}$$

를 먼저 생성한다. 단, $n_{\text{sim}}\in\mathbb{N}$은 모의기간 길이를 뜻한다. 이어서, 다음의 알고리즘에 따라 기온시나리오를 생성한다.

기온 시나리오 생성 알고리즘

1. 일평균추세생성: 각 지점 s마다 모의기간 $t_0+1,\cdots,t_0+n_{\text{sim}}$ 동안 일평균추세를 식
$$\mu(s,t)=\mathrm{E}\{\mathrm{T}(s,t)\}+\Delta\{\mathbf{R}(t)\},\quad t=t_0+1,\cdots,t_0+n_{\text{sim}}$$
에 따라 생성한다. 단, $\mathrm{E}\{\mathrm{T}(s,t)\}$는 식 (23)에 따라 추정되고, $\Delta\{\mathbf{R}(t)\}$의 정의는 식 (26)에서 주어진다.
2. Anomaly 생성: 모의기간 $t_0+1,\cdots,t_0+n_{\text{sim}}$ 동안 모형 (31)에 따라
$$\boldsymbol{X}(t),\quad t=t_0+1,\cdots,t_0+n_{\text{sim}}$$
을 생성한다. 또한, $\langle \boldsymbol{\epsilon}_{\max}(t),\boldsymbol{\Psi}_i^{\max}\rangle$와 $\langle \boldsymbol{\epsilon}_{\min}(t),\boldsymbol{\Psi}_i^{\min}\rangle$, $i=q+1,\cdots,d$ 를 각각 경험분포함수로부터 $t=t_0+1,\cdots,t_0+n_{\text{sim}}$ 동안 독립적으로 생성한다. 다음, 식 (32)에 따라 $\boldsymbol{\epsilon}_{\max}(t)$와 $\boldsymbol{\epsilon}_{\min}(t)$, $t=t_0+1,\cdots,t_0+n_{\text{sim}}$을 얻는다.
3. 기온생성: 각 지점 s와 $t=t_0+1,\cdots,t_0+n_{\text{sim}}$에 대해서,
$$\mathrm{T}(s,t)=\mu(s,t)+\sigma(s,t)\cdot\epsilon(s,t)$$
에 따라 일최고/최저기온값을 얻는다.

그림 12는 weather generator의 기상 시나리오 생성 알고리즘의 순서도를 나타낸다. 순서도에서 강수생성의 순서도는 그림 4에서 주어졌다.

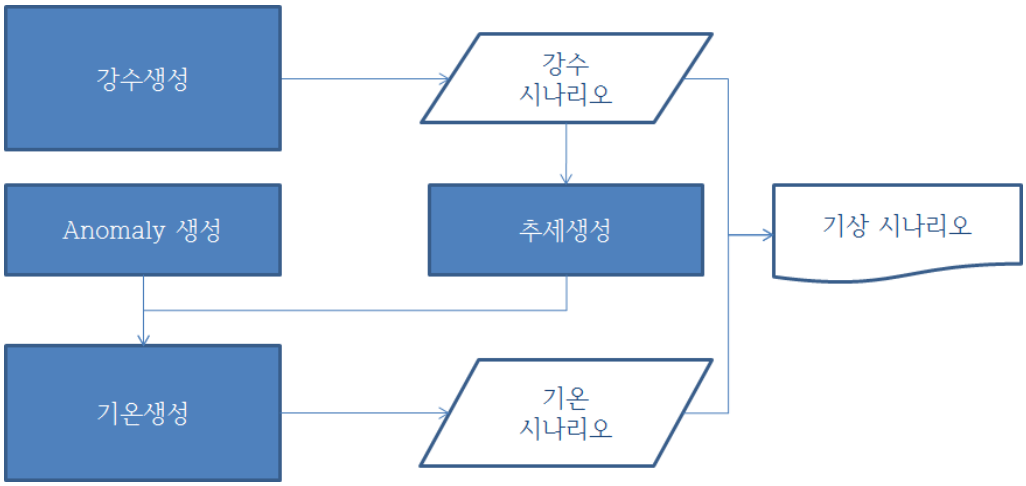
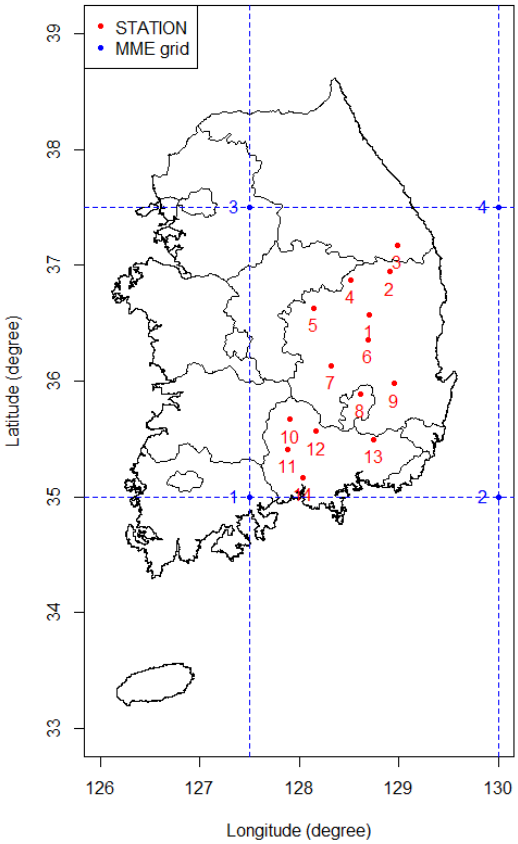


그림 12 weather generator 기상 시나리오 생성 알고리즘 순서도

3. 사례 연구 - 낙동강 유역

본 절에서는 앞서 제안한 통계모형과 알고리즘을 낙동강 유역에 적용한 결과를 제시하고자 한다. 통계모형구축을 위한 관측자료로 ASOS(자동종관관측시스템)자료를 썼는데, 낙동강 유역에는 총 14개 관측소가 위치해 있다, 즉, $d = 14$이다. 관측기간은 1988년 1월 1일부터 2013년 12월 31일까지이다. 그림 13은 관측소의 위치와 이름을 나타내고 있다.



관측소 번호와 이름

No.	이름	No.	이름
1	안동 (Andong)	8	대구 (Daegu)
2	봉화 (Bongwhoa)	9	영천 (Yeongcheon)
3	태백 (Tawbaekk)	10	거창 (Geochang)
4	영주 (Yeongju)	11	산청 (Sancheong)
5	문경 (Mongyeong)	12	합천 (Hapcheon)
6	의성 (Uisung)	13	밀양 (Milyang)
7	구미 (Gumi)	14	진주 (Jinju)

APCC MME 격자

No.	위치
1	127.5° E, 35.0° N
2	130.0° E, 35.0° N
3	127.5° E, 37.5° N
4	130.0° E, 37.5° N

그림 13 낙동강 유역, ASOS 관측소(적색), APCC MME 격자(청색)

3.1 모형추정과 Tuning

앞서 2.절에서 강수와 기온에 대한 통계모형을 소개하면서 모형추정법을 간단히 설명하였다. 그런데, 기온 순환추세 추정을 위한 local polynomial regression 적합(23)을 제외한 나머지 추정은 모두 관측자료의 정상성(stationarity)를 전제하기 때문에, 이를 보장할 수 있도록 관측자료를 월 단위로 나누어 수행되었다. (단, 경년변동(inter-annual variation)이 통계적으로 유의하지 않다고 전제한다.) 표 4에서 추정방식을 정리하였다. 추정 이외에도 모형적합과 알고리즘에서 세부적인 선택사항이 있다. 낙동강 유역에 대해 통계모형과 알고리즘을 tuning한 결과를 표 5와 같다.

추정결과 모두를 제시하기는 어려우므로 몇 가지를 선정하였다. 그림 14은 극한강수의 intensity 추정결과를 각각 지점 고도에 따라 나타내었다. 여기서 극한강수의 기준은 상위 2% 분위강수량으로, 극한강수 intensity는 습윤일에 극한강수가 발생했을 때 기대강수량(mm/day)으로 정의하였으며, 지점 s_j의 극한강수 intensity는 $\sigma_j + F_j^{\leftarrow}(0.98|\mathrm{w})$으로 계산하였다.[8] (기호의 정의는 식 (8)과 (19)를 참조.) 그림에서 점(•)은 intensity의 추정치, I빔은 신뢰수준 95%의 신뢰구간을 나타낸다. 그림은 극한강수의 세기가 지점고도에 따른 뚜렷한 경향이 없음을 보여 준다. 그림 15-18은 local polynomial regression 방법으로 추정된 최고/최저기온의 일평균 순환추세이다. 모든 지점에서 적절히 순환추세를 얻은 것으로 판단된다. 그 외 Gaussian copula 모형과 Gaussian 공간 극단값 모형 추정결과는 부록에서 추정방법과 함께 제시한다.

8] $\gamma_j = 0$, 즉, Gumbel 모형 하에서 계산하였다.

표 4 낙동강 유역 통계모형 추정 방식

Parameter	추정방식
(1), (24), (29)의 공분산 행렬	월별로 구분하여 표본 공분산 행렬로 추정
식 (5)의 지점 강수량 분포함수 $F_j(\cdot\|\mathrm{w})$와 식 (8)의 분위수 함수 $F_j^{\leftarrow}(\cdot\|\mathrm{w})$, $j=1,\cdots,d$	월별로 구분하여 경험분포함수와 관측분위수로 추정. 단, 극단영역에서 식 (20)의 $\widetilde{F}_j^{\leftarrow}(\cdot\|\mathrm{w})$를 사용한다.
Gaussian copula 모형 (11)의 Σ	월별로 구분하여 부록에서 소개된 방법으로 추정
공간극단값 모형 (12)의 profile 함수 f의 추정	호우기(6-9월) 각 월에 대해서 부록에서 소개된 방법으로 추정
일반화 Pareto 모형 (19)의 σ_j와 γ_j, $j=1,\cdots,d$	호우기(6-9월) 각 월에 대해서 $\gamma_j=0$두고, 최대우도추정법 (Coles et al. (2001)의 4.3.2. 절)으로 추정
2.2.1.절의 local polynomial regression	식 (23)의 방법으로 월 구분없이 추정
회귀모형 (25)	지점별, 월별로 구분하여 최소제곱법으로 추정
벡터자기회귀모형 (31)	월별로 구분하여 최소제곱법으로 추정

표 5 낙동강 유역 통계모형과 알고리즘 tuning 결과 또는 방식

Parameter	tuning 결과 또는 방식
준경험적 copula 모형의 $\mathrm{b}_{j,1},\cdots,\mathrm{b}_{j,k}$	$k=6$, $\mathrm{b}_{j,1}=0.5,\ \mathrm{b}_{j,2}=0.6,\cdots,\ \mathrm{b}_{j,5}=0.9,\ \mathrm{b}_{j,6}=0.98$
식 (3)의 p (은닉상태의 차수)	$p=2$
k-NN 방식의 k (neighbor의 갯수)	$k=10$
습윤기 생성 알고리즘의 N (pool 크기)	$N=100$
식 (23)의 kernel 함수 Ker	Ker: Gaussian
식 (30)의 q (선택된 주성분의 수)	$q=2$
식 (31)의 p (벡터자기회귀모형의 차수)	AIC를 최소로 하는 값으로 선정
식 (25)의 γ (회귀모형 선형성 위한 tuning parameter)	$\{0.1, 0.2,\cdots, 1\}$중에서 회귀모형 R^2의 최대로 하는 값으로 선정

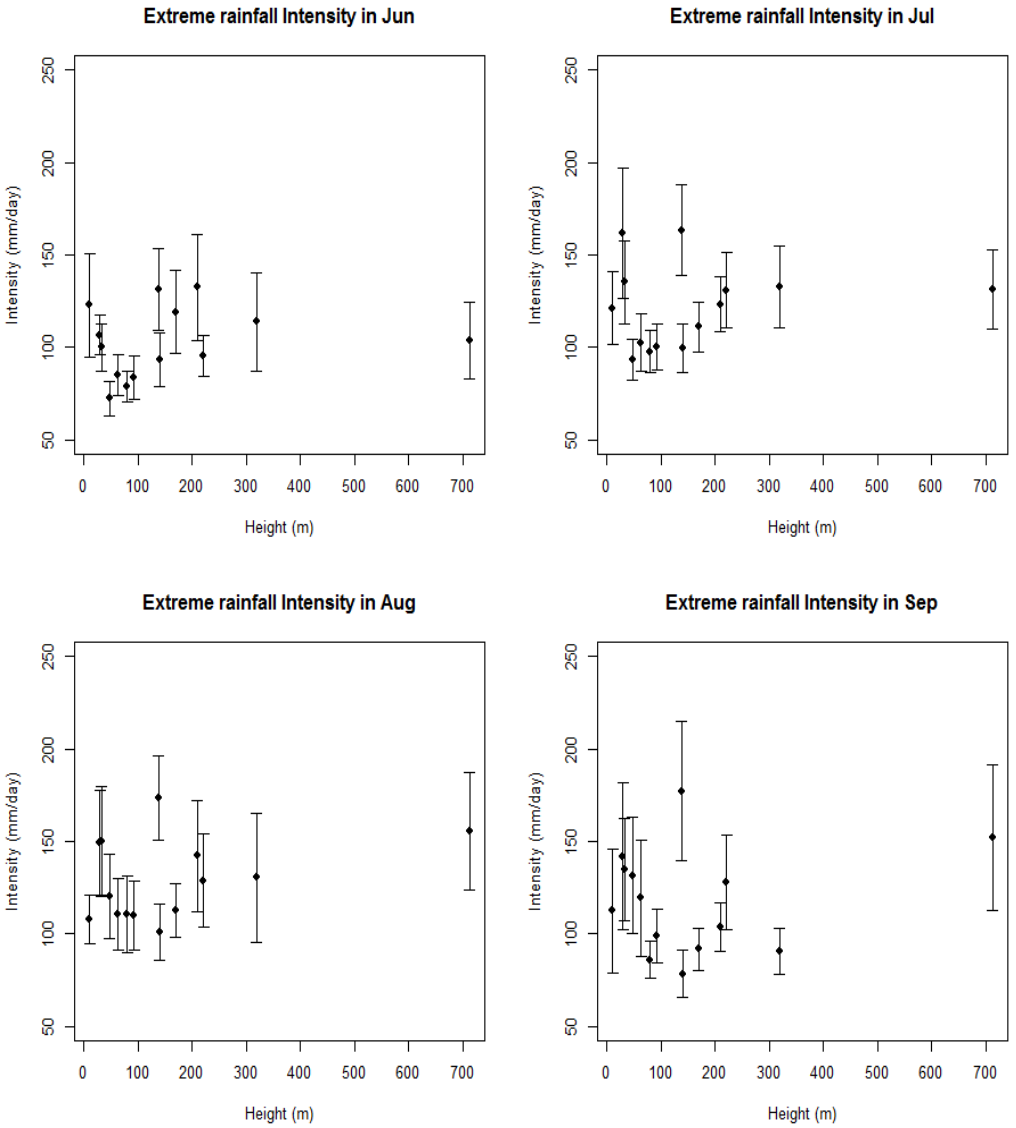


그림 14 낙동강 유역 호우기 (6월-9월) 극한강수의 intensity 추정결과. 극한강수의 기준량은 상위 2% 분위강수량으로 선정하였다. 점(•)은 추정치, I빔은 신뢰수준 95% 신뢰구간을 나타낸다. 지점 고도에 따른 intensity의 뚜렷한 경향은 나타나지 않는다.

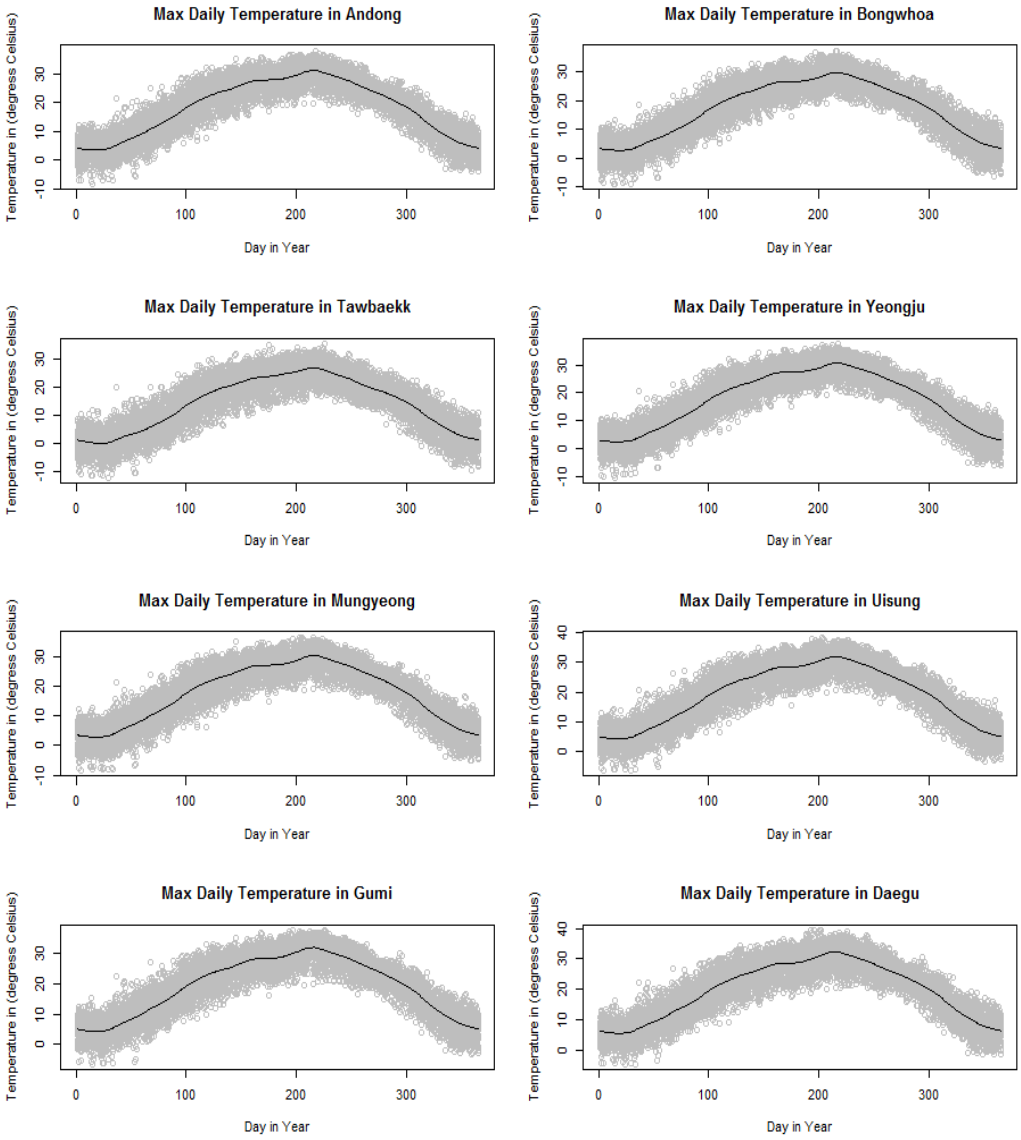


그림 15 각 지점의 최고기온 일평균 local polynomial regression 통한 추정 결과

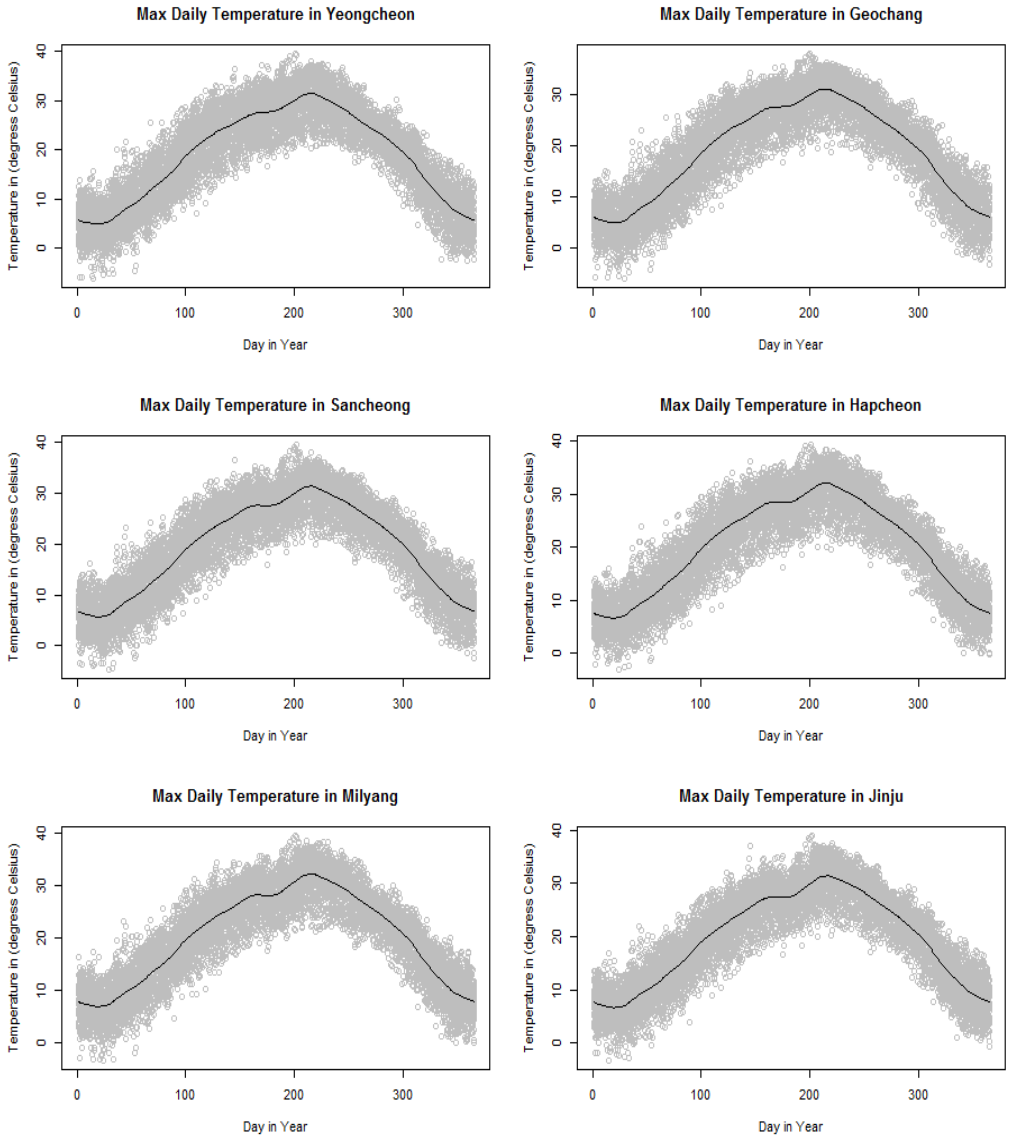


그림 16 각 지점의 최고기온 일평균 local polynomial regression 통한 추정 결과

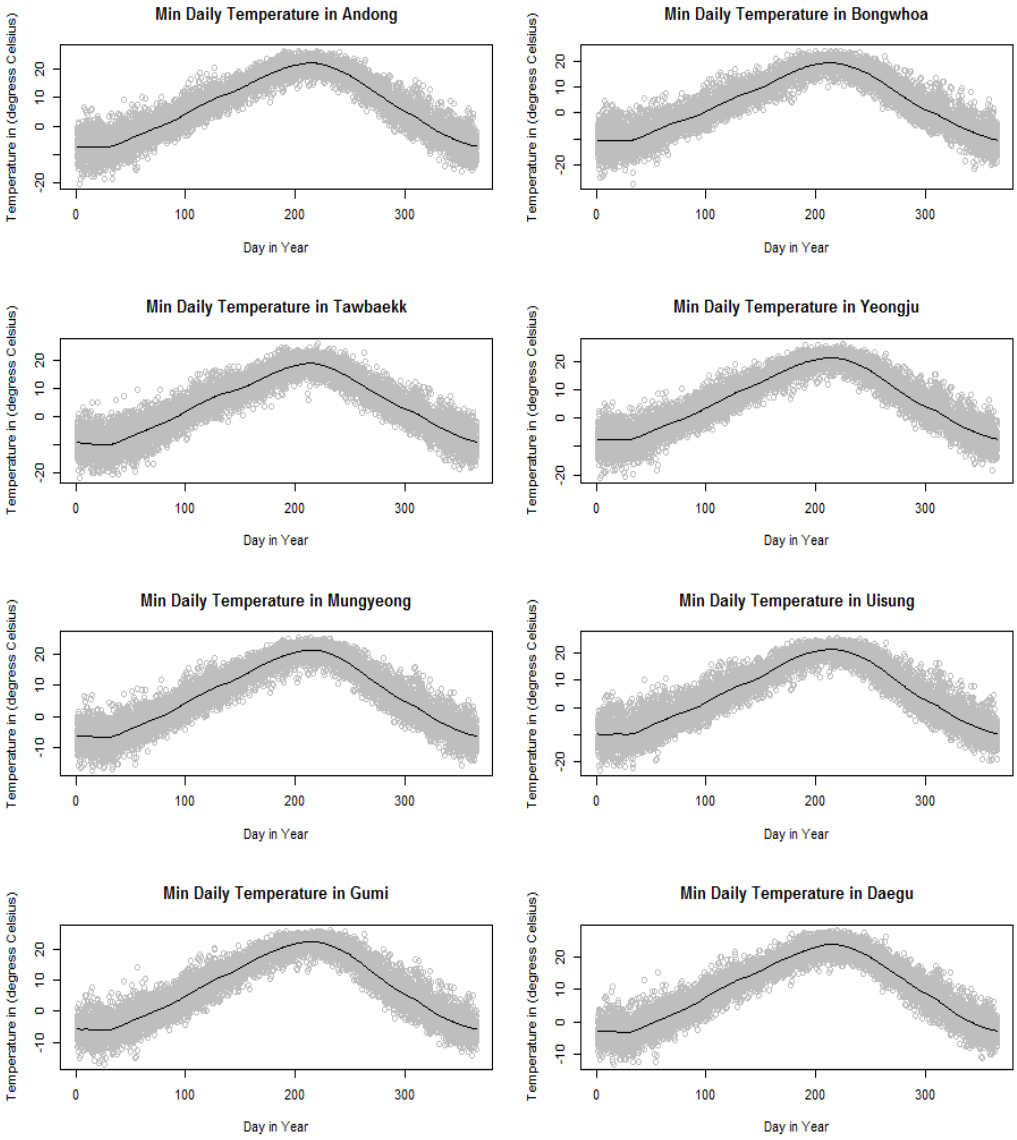


그림 17 각 지점의 최저기온 일평균 local polynomial regression 통한 추정 결과

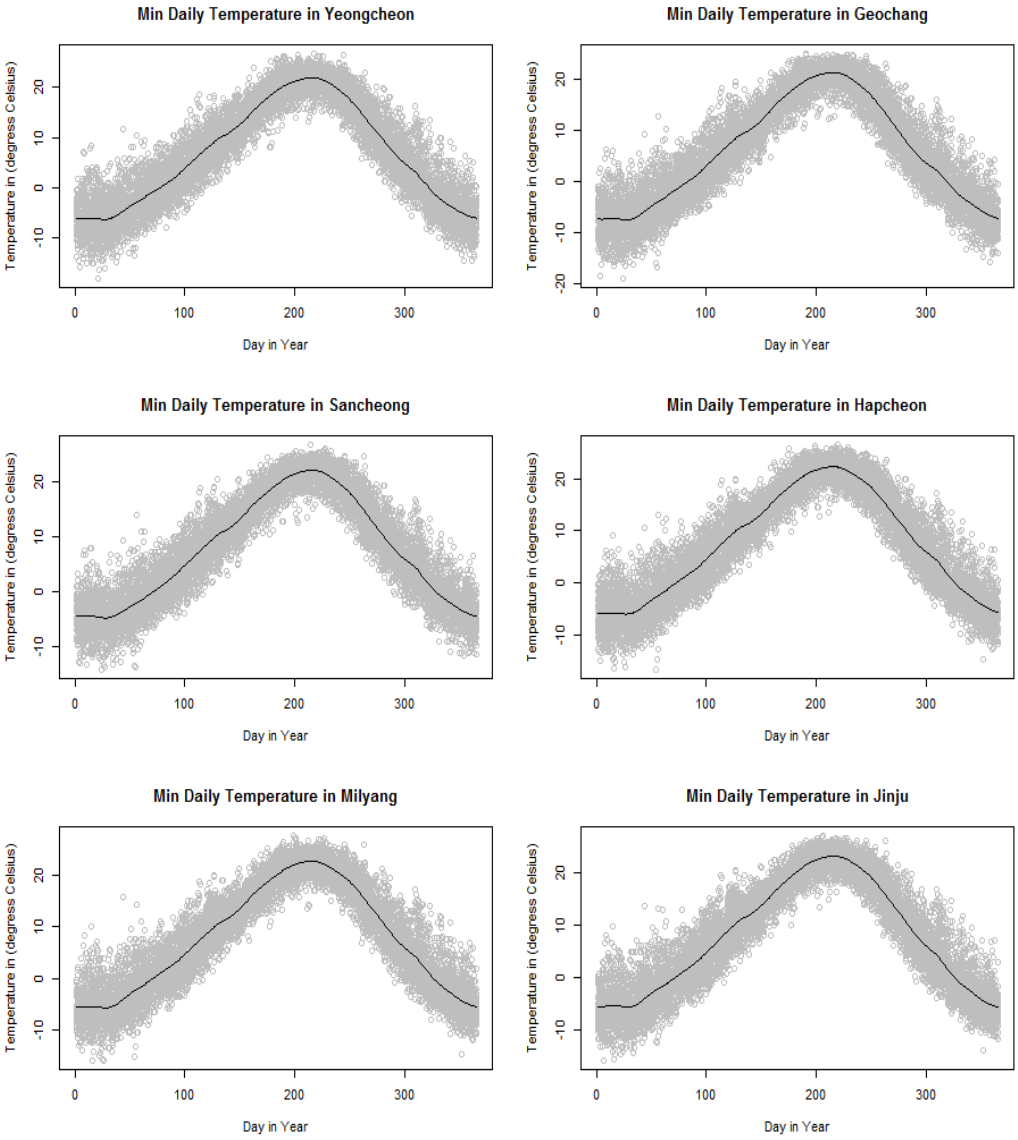


그림 18 각 지점의 최저기온 일평균 local polynomial regression 통한 추정 결과

3.2 모의 성능 검증

Weather generator의 모의 성능은 관측자료 통계적 특성에 대한 재현정확도로 평가된다. 본 보고서에서 제안한 모의알고리즘의 성능을 평가하기 위해서, 1월 1일부터 12월 31일을 모의기간으로 한 모의자료를 총 1000개 생성하였다. 여기서 살펴볼 통계적 특성은 다음과 같다:

ⅰ. 강수와 기온의 일별분포;

ii. 강수와 기온의 공간적 자기상관;

iii. 기온의 시간적 자기상관;

iv. 최고/최저 기온 간 상관;

ⅴ. 강수의 월 단위 통계적 특성: 월누적강수량, 최대일강수량, 습윤일수;

vi. 기온의 월평균.

3.2.1 일별분포 검증

강수량, 최고/최저기온의 일별분포에 대한 모의성능을 모든 지점에서 검증한다. 일별분포는 각 모의자료와 관측자료에 대해 각 월별로 경험분포함수로써 계산하고, 분포 간 차이가 통계적으로 유의한지를 확인하기 위해 Kolmogorov–Smirnov 검정을 실시하였다. 또한, 그림 도구로 관측자료의 히스토그램, 모의자료 확률밀도함수의 비모수적(nonparametric) 추정 결과, 그리고 분위수그림(quantile–quantile plot)을 쓴다.

우선, 강수량의 일별분포를 살펴본다. 앞서 설명하였듯이, 강수생성알고리즘은 준경험적 copula 모형과 Gaussian copula 모형의 알고리즘으로 나누어진다. 먼저 준경험적 copula 모형 기반 알고리즘의 검증결과를 제시한다. 그림 19은 유의성 검정의 결과를 나타내는데, 모든 지점과 월에서 분포간 차이가 유의하지 않다. 면밀한 검토를 위해 분위수 그림으로 검정결과는 재확인하였다. 그림 20–23은 각 계절과 지점을 대표하여 1월, 4월, 7월, 10월의 안동, 태백, 대구, 진주의 일별강수량 분포를 비교한

것이다. 그 결과, 모든 경우에서 관측자료와 모의자료가 서로 유사한 분포 형태를 보이고 있다.

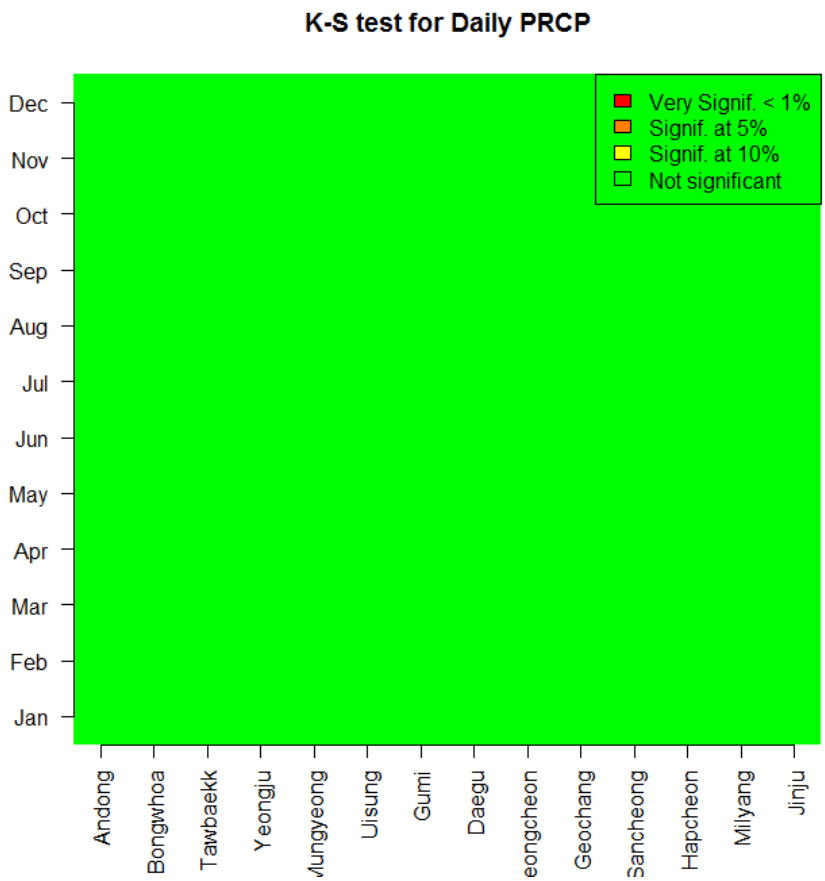


그림 19 준경험적 copula 모형 기반 알고리즘의 강수량 일별분포 차이 유의성 검정 결과. 유의수준 10%, 5%, 1%에서 유의한 차이로 보이면, 각각 노란색, 주황색, 빨강색으로 표시되고, 유의하지 않으면 녹색으로 표시됨. 모든 지점과 월에서 유의한 차이가 나타나지 않는다.

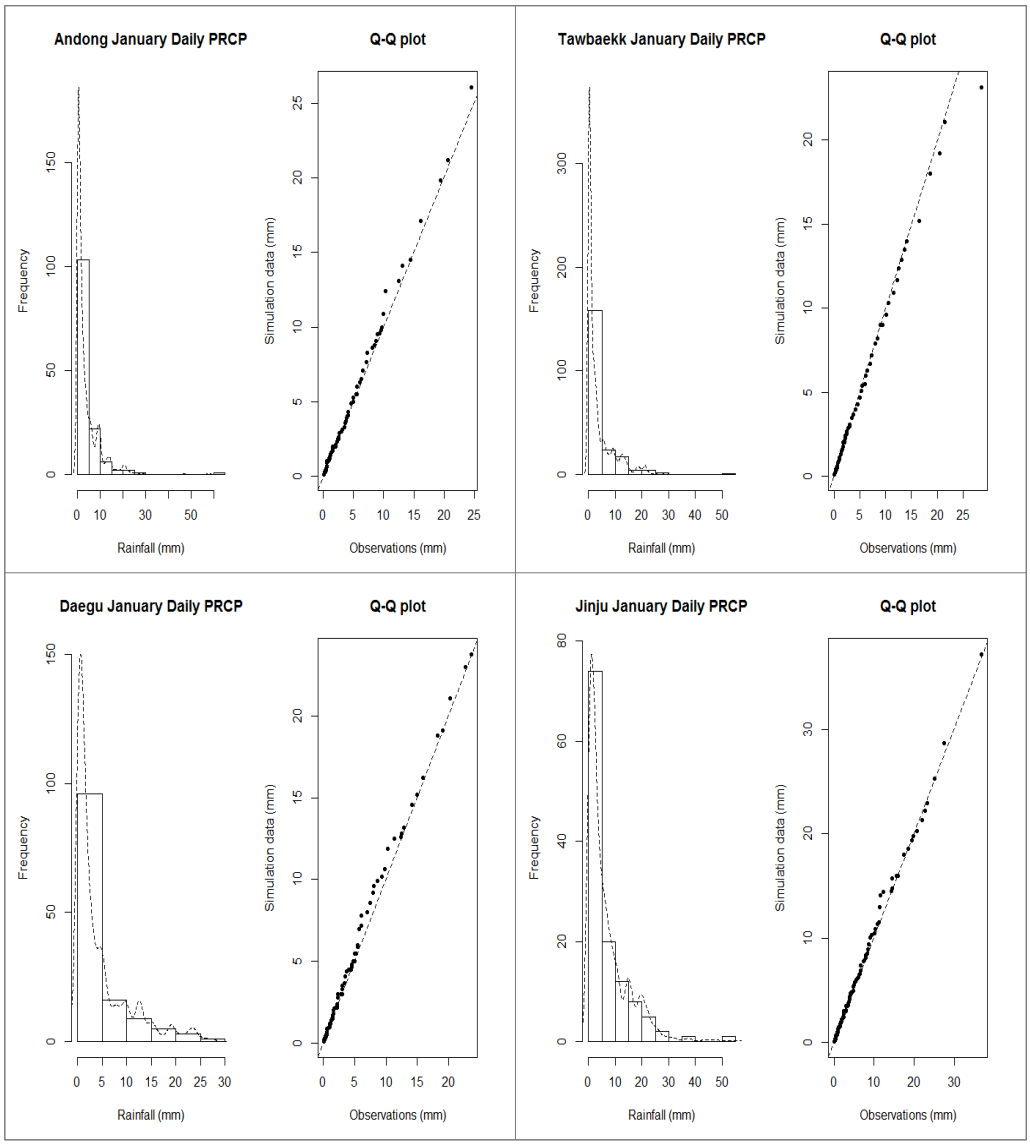


그림 20 준경험적 copula 모형 기반 알고리즘의 모의자료에 대한 1월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 각 지점에서 왼편 히스토그램은 관측자료분포를, 왼편 점선그래프는 모의자료의 확률밀도함수를 나타낸다. 오른편은 분위수 그림이다. 모든 경우에서 분포가 매우 유사하다.

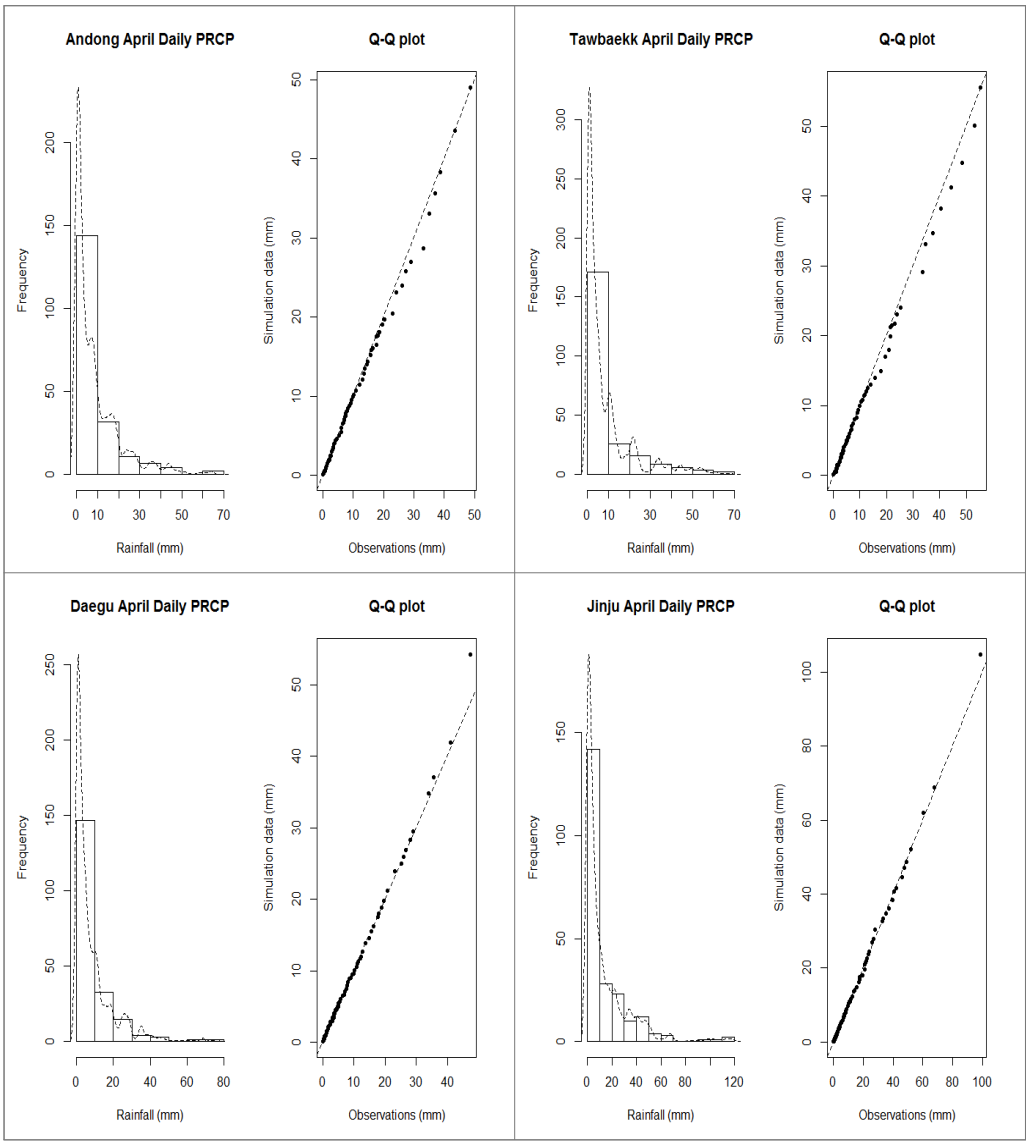


그림 21 준경험적 copula 모형 기반 알고리즘의 모의자료에 대한 4월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 모든 경우에서 분포가 매우 유사하다.

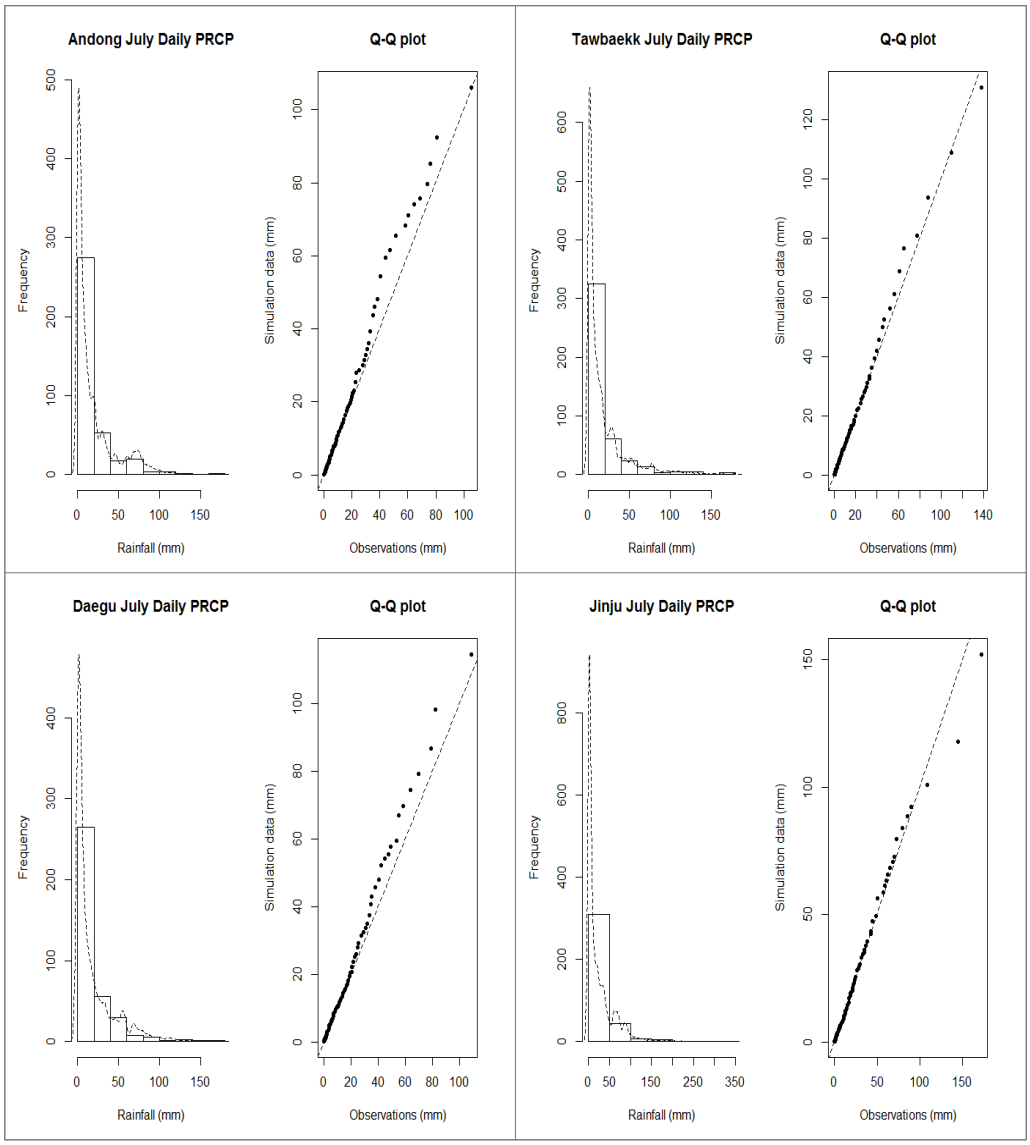


그림 22 준경험적 copula 모형 기반 알고리즘의 모의자료에 대한 7월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 모든 경우에서 분포가 유사하다.

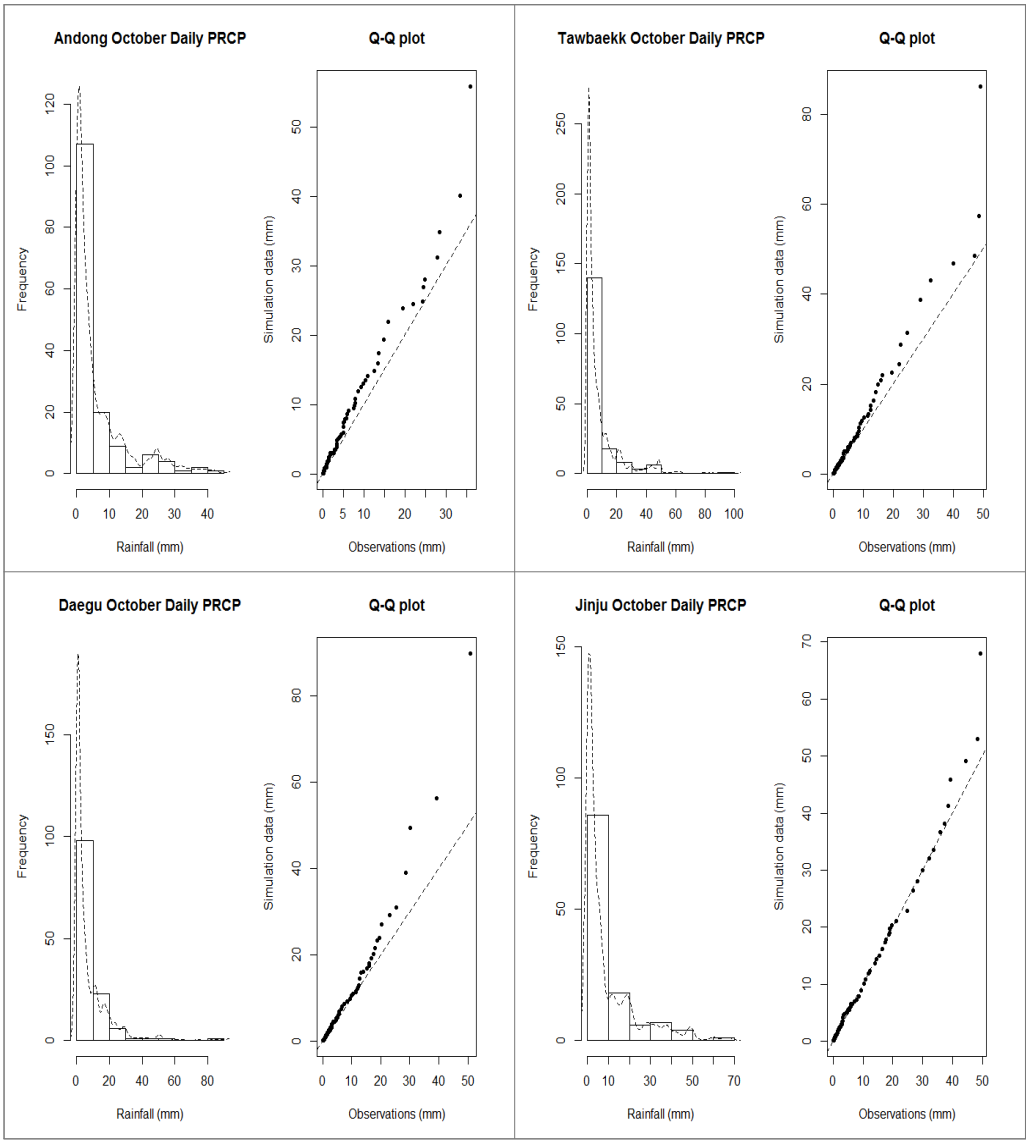


그림 23 준경험적 copula 모형 기반 알고리즘의 모의자료에 대한 10월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 모든 경우에서 두 분포가 대체로 유사하다.

다음으로 Gaussian copula 모형 기반 알고리즘의 모의자료를 검증한다. 그림 24는 유의성 검정 결과인데, 역시 모든 지점과 월에서 분포 간 차이가 유의하지 않다. 또한, 그림 25-28에서 나타나듯이, 히스토그램과 분위수그림에서도 두 분포가 유사함을 확인할 수 있다. 결론적으로, 두 copula 모형 모두에서 강수량일별분포는 강수생성알고리즘에 의해 잘 모의되고 있다.

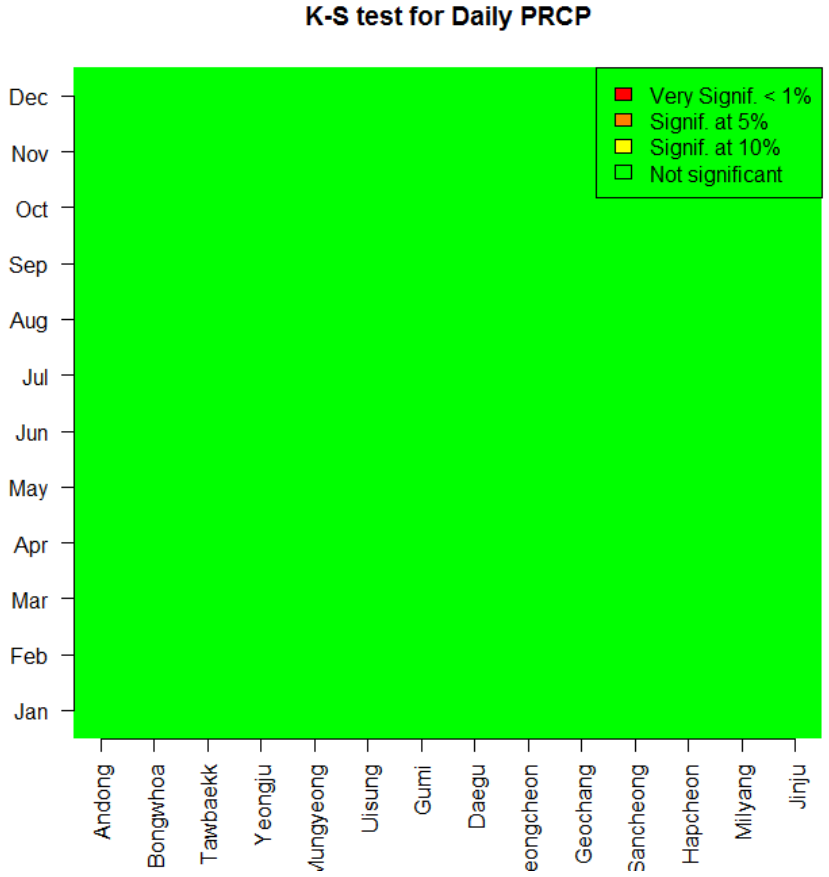


그림 24 Gaussian copula 모형 기반 알고리즘의 강수량 일별분포 차이 유의성 검정 결과. 유의수준 10%, 5%, 1%에서 유의한 차이로 보이면, 각각 노란색, 주황색, 빨강색으로 표시한다. 유의하지 않으면 녹색으로 표시된다. 모든 지점과 월에서 유의한 차이가 나타나지 않는다.

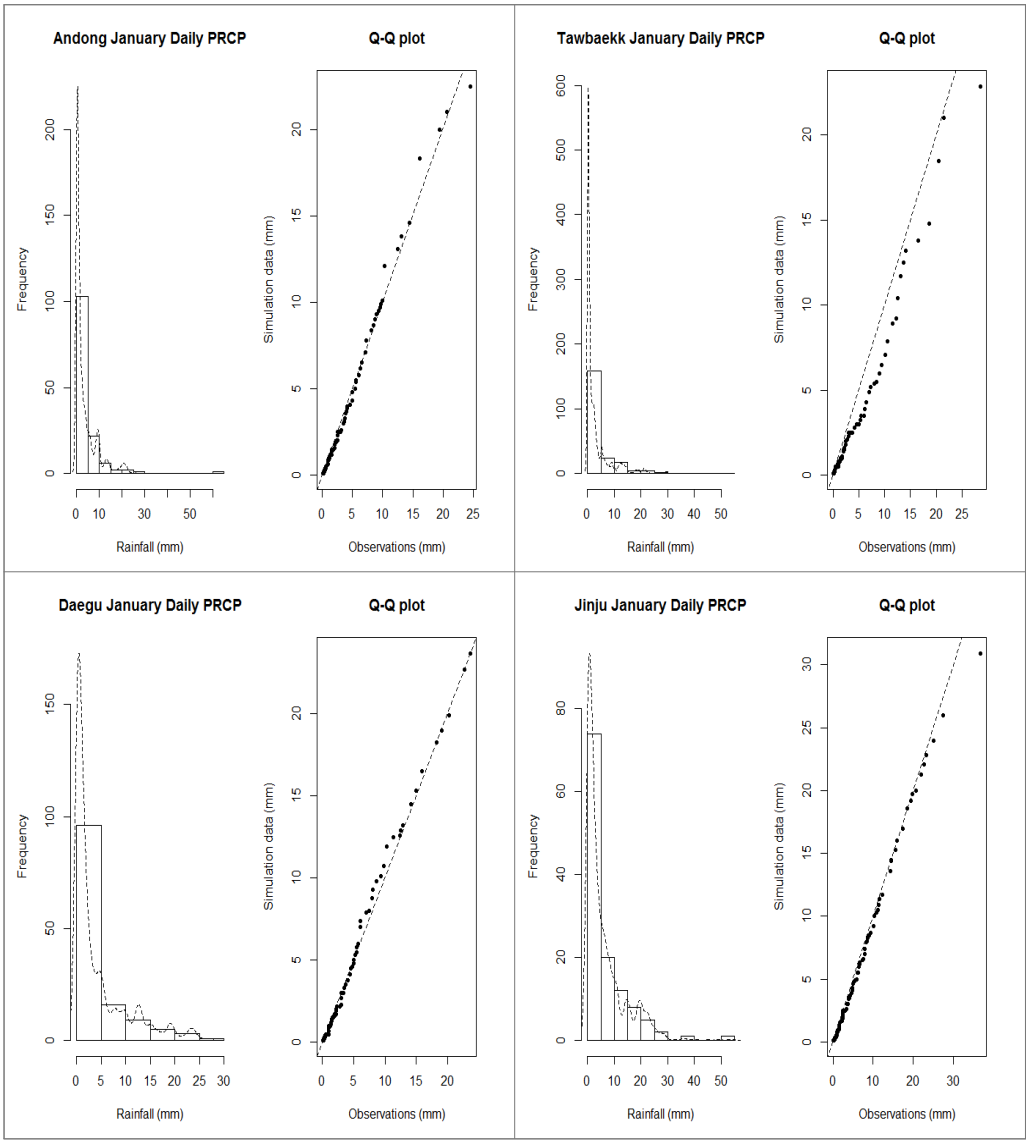


그림 25 Gaussian copula 모형 기반 알고리즘의 모의자료에 대한 1월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 각 지점에서 왼편 히스토그램은 관측자료분포를, 왼편 점선그래프는 모의자료의 확률밀도함수를 나타낸다. 오른편은 분위수 그림이다. 모든 경우에서 분포가 유사하다.

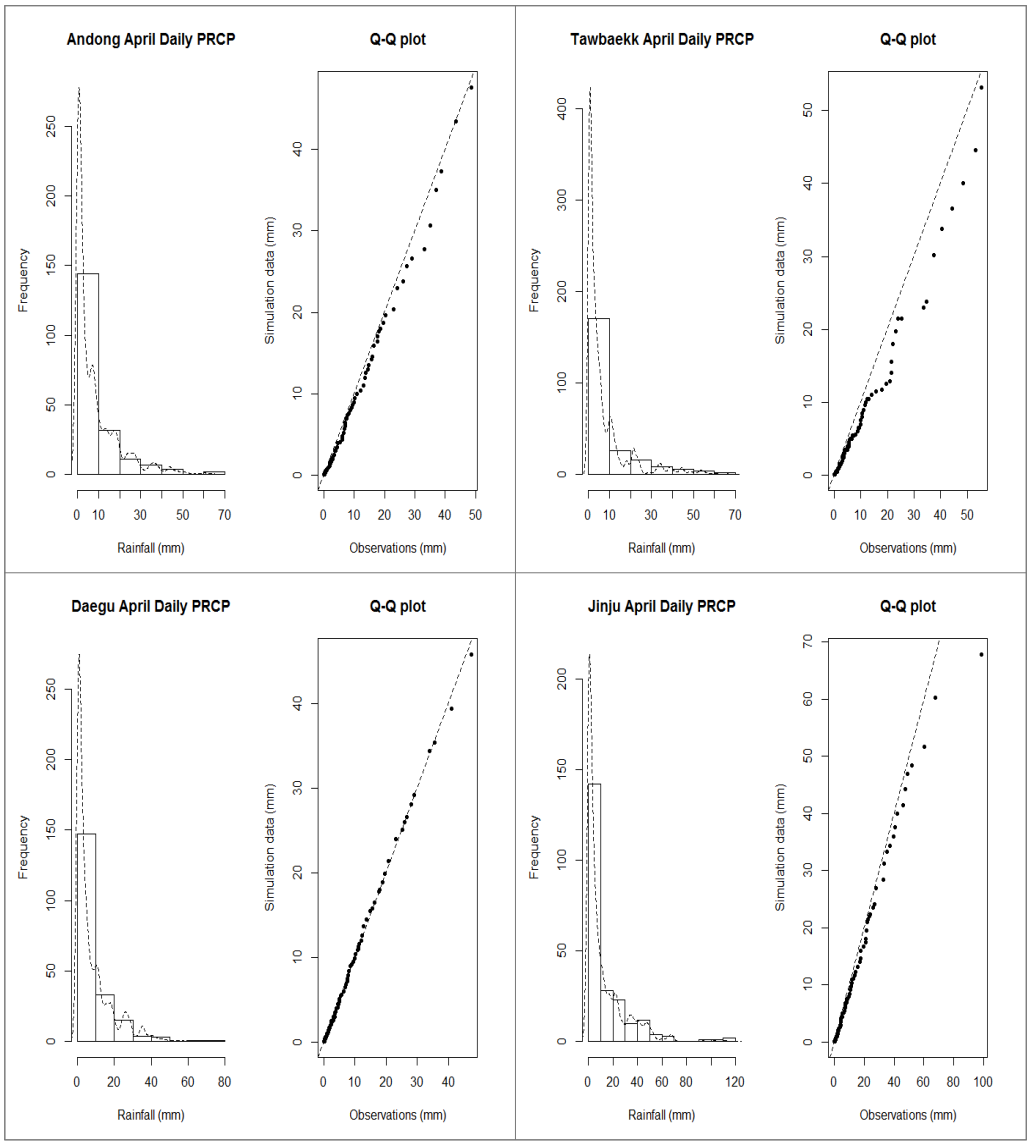


그림 26 Gaussian copula 모형 기반 알고리즘의 모의자료에 대한 4월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 모든 경우에서 분포가 유사하다.

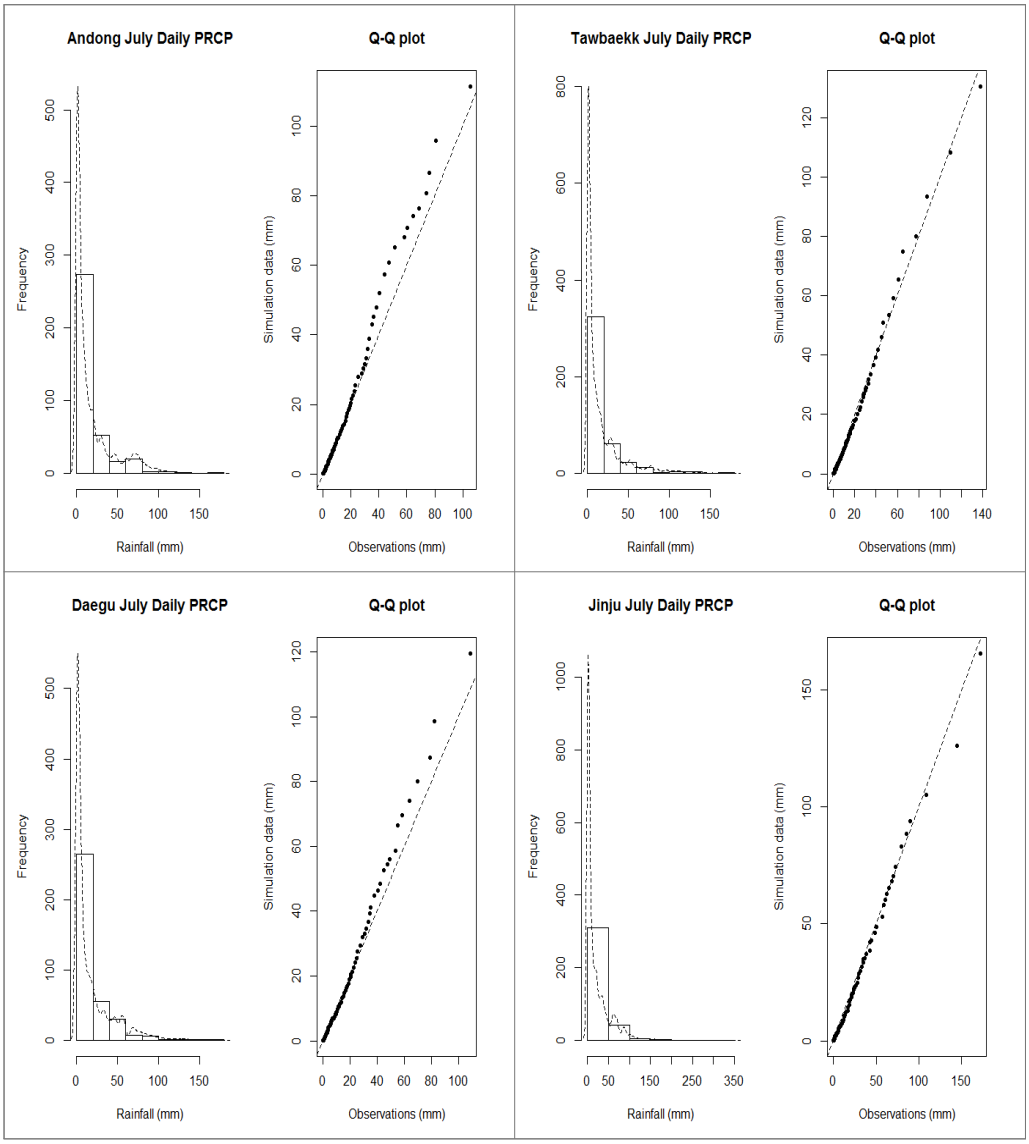


그림 27 Gaussian copula 모형 기반 알고리즘의 모의자료에 대한 7월 안동, 태백, 대구, 진주의 일별강수량 분포 비교. 모든 경우에서 분포가 유사하다.

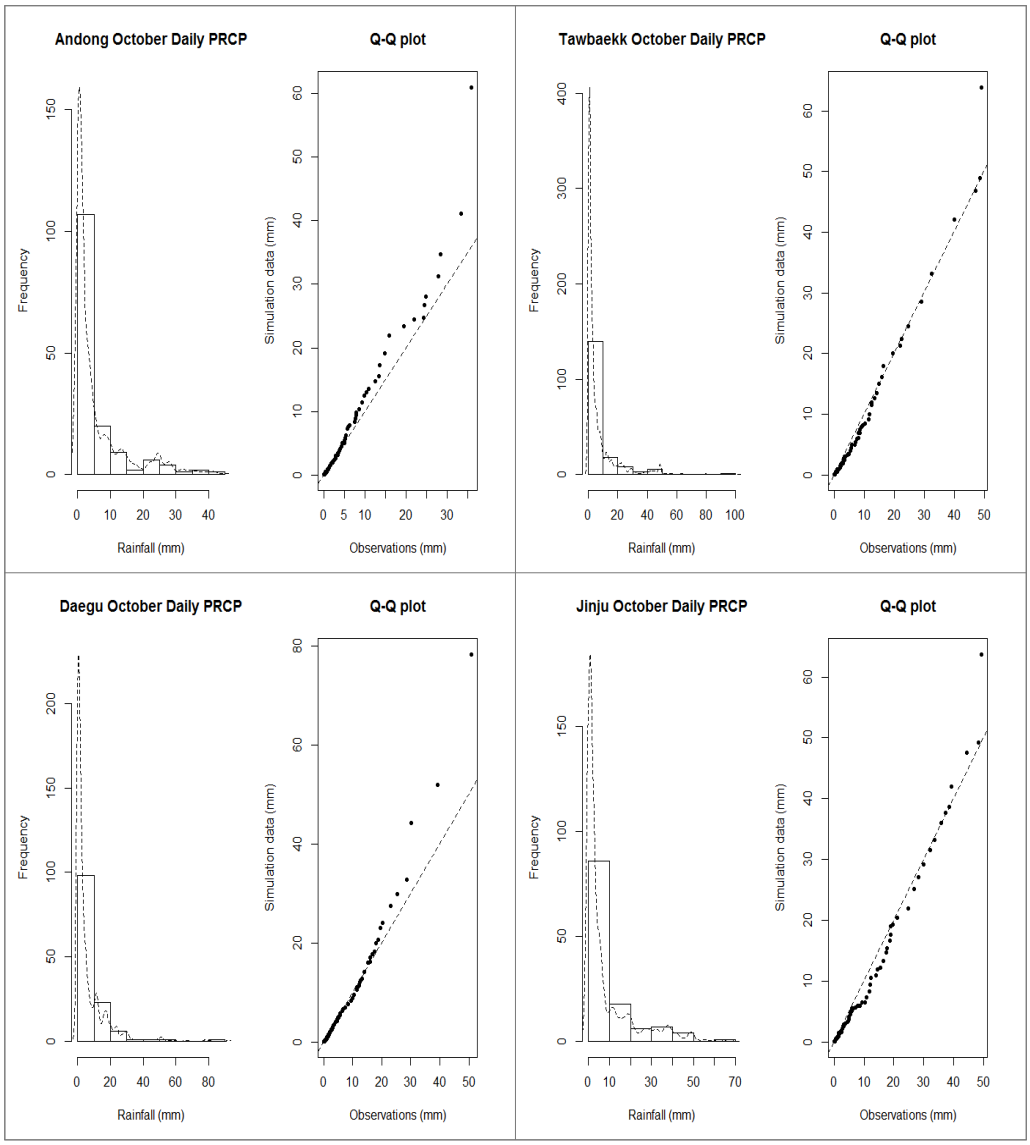


그림 28 Gaussian copula 모형 기반 알고리즘의 모의자료에 대한 10월 안동, 태백, 대구, 진주의 일별 강수량 분포 비교. 모든 경우에서 분포가 유사하다.

이제, 기온의 대한 검증으로 넘어 가겠다. 기온은 강수의 영향을 받으므로, 엄밀히는 기온생성알고리즘의 성능은 강수생성알고리즘의 선택에 영향을 받을 것이다. 하지만 그 차이가 크지 않을 것이라 예상되어, Gaussian copula 모형 기반의 강수생성알고리즘의 경우만 살펴보기로 한다. 검증방법은 강수량의 경우와 동일하다.

그림 29은 최고기온의 분포간 차이 유의성 검정의 결과이다. 최고기온은 강수의 경우와 달리 6월과 7월에 몇몇 지점에서 유의수준 5%에서 유의한 분포차이를 보인다. 그 차이정도를 확인하기 위해, 해당 월에서 유의한 차이를 보이는 지점을 포함한 4개 지점에서 분위수그림을 그려 보았다, (그림 30-31). 확인결과, 그 차이가 크지 않고, 대두 분포가 대체적으로 유사하다.

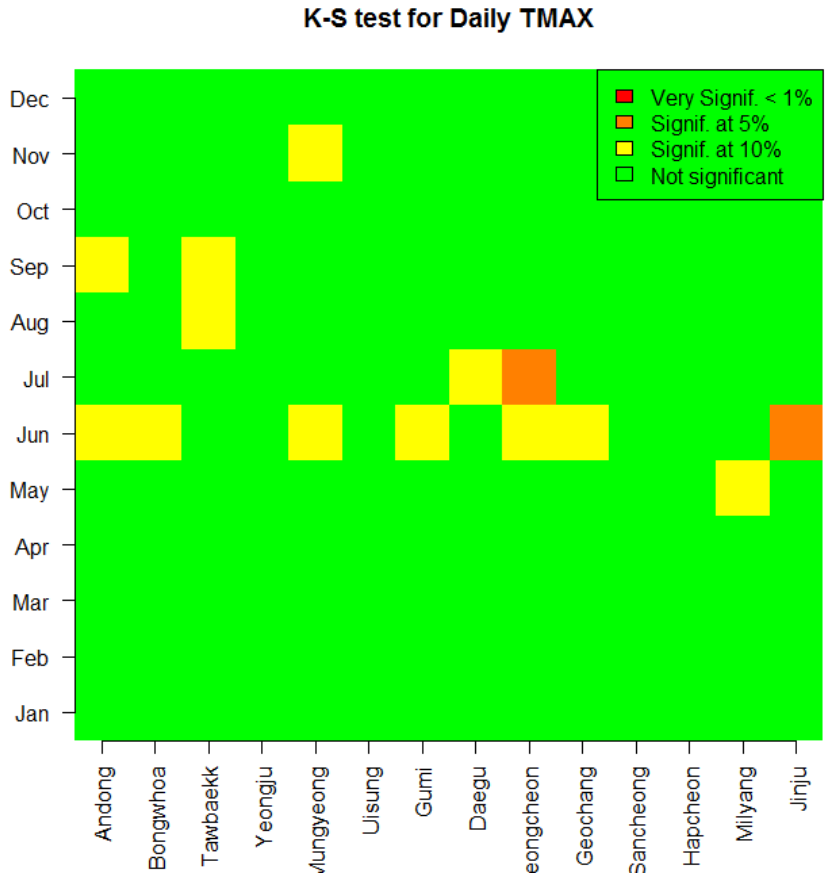


그림 29 모의자료의 최고기온 일별분포 차이 유의성 검정 결과. 유의수준 10%, 5%, 1%에서 유의한 차이로 보이면, 각각 노란색, 주황색, 빨강색으로 표시한다. 녹색은 유의하지 않음을 의미한다. 몇몇 지점과 월에서 유의수준 5%에서 유의한 차이가 나타난다.

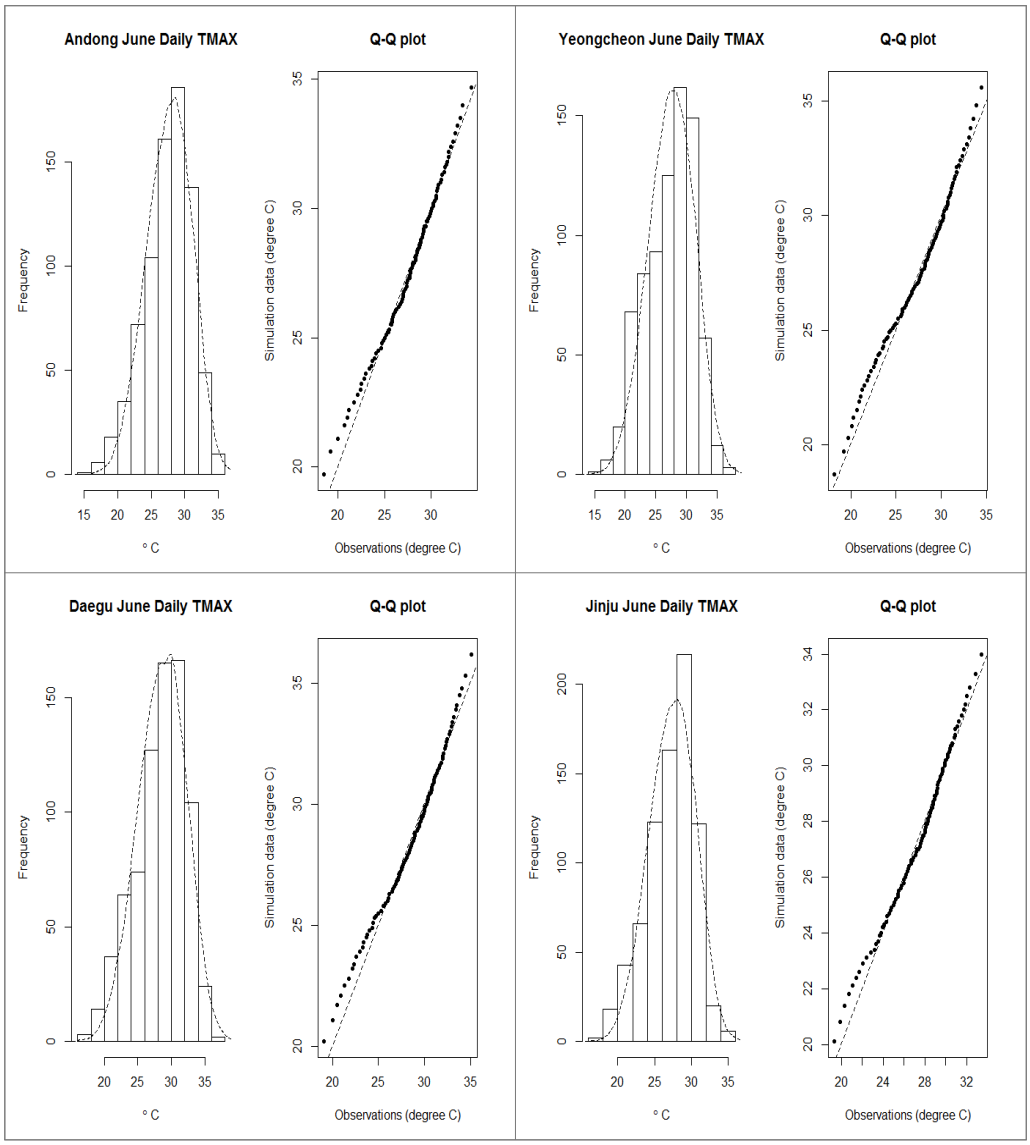


그림 30 6월 안동, 영천, 대구, 진주의 최고기온 일별 분포 비교. 각 지점에서 왼편 히스토그램은 관측자료분포를, 왼편 점선그래프는 모의자료의 확률밀도함수를 나타낸다. 오른편은 분위수 그림이다. 분포가 대체적으로 유사하다.

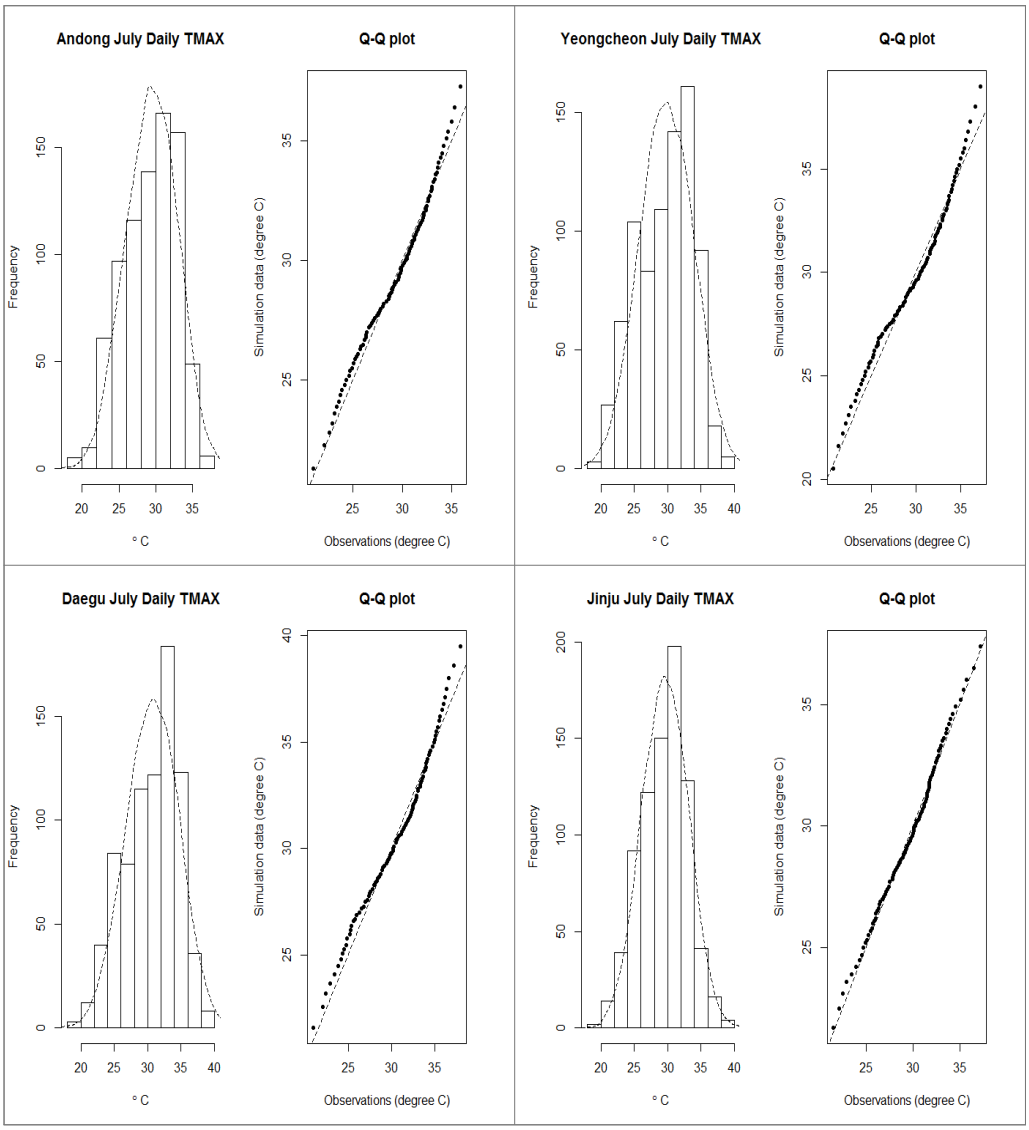


그림 31 7월 안동, 영천, 대구, 진주의 최고기온 일별 분포 비교. 분포가 대체적으로 유사하다.

최저기온에 대해서도 동일하게 검증을 실시한다. 그림 32은 분포 간 차이 유의성 검정결과인데, 7월 산청(Sancheong)지점에서만 유의수준 10%에서 차이가 유의하게 나타난다. 즉, 최고기온에 비해 모의성능이 나은 것으로 판단된다. 면밀한 검증을 위해 분위수그림을 그려 보았다. 그림 33-36에서 확인할 수 있듯이, 일별분포가 잘 모의되고 있다. 전체적으로, 강수와 기온의 일별분포가 만족할 만한 수준으로 잘 모의되고 있다.

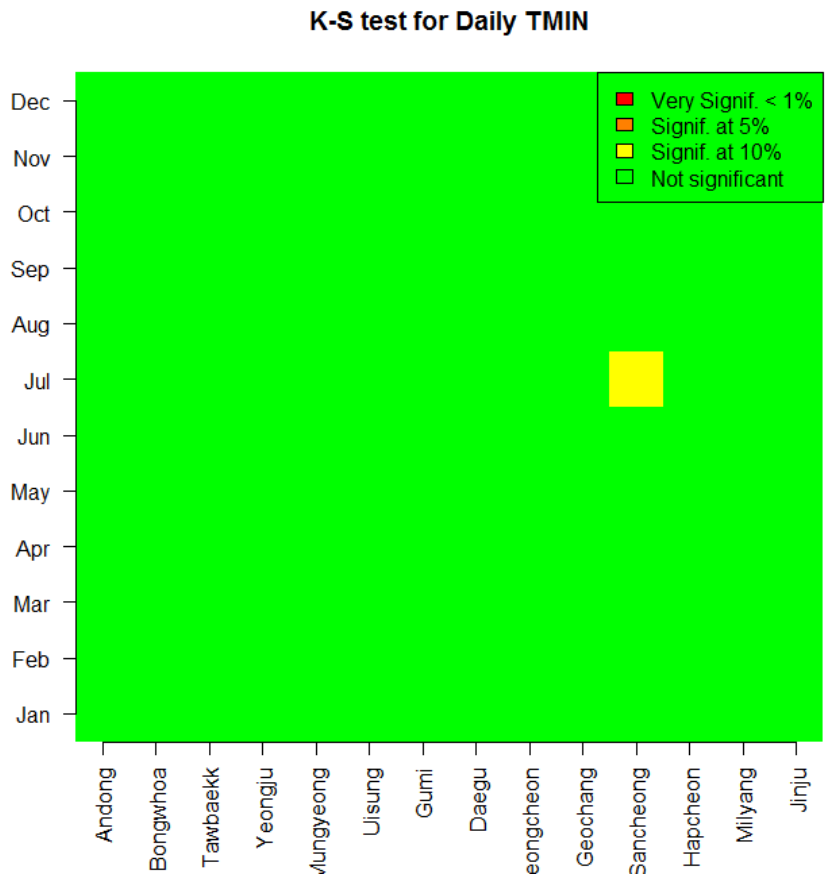


그림 32 모의자료의 최저기온 일별분포 차이 유의성 검정 결과. 유의수준 10%, 5%, 1%에서 유의한 차이로 보이면, 각각 노란색, 주황색, 빨강색으로 표시한다. 녹색은 유의하지 않음을 뜻한다. 7월 산청 지점에서 유의수준 10%에서 유의한 차이가 나타난다.

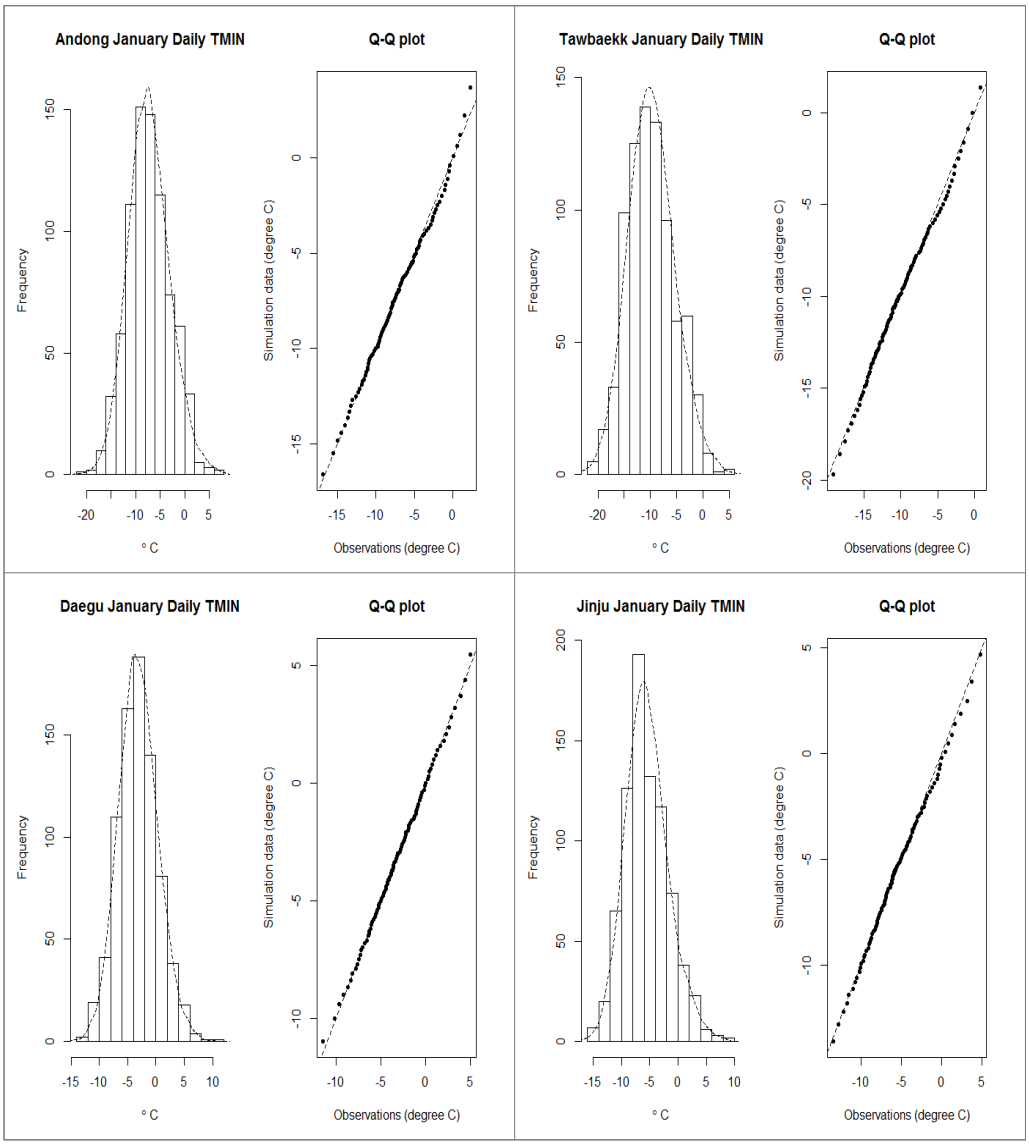


그림 33 1월 안동, 태백, 대구, 진주의 최저기온 일별 분포 비교. 각 지점에서 왼편 히스토그램은 관측자료분포를, 왼편 점선그래프는 모의자료의 확률밀도함수를 나타낸다. 오른편은 분위수 그림이다. 분포가 매우 유사하다.

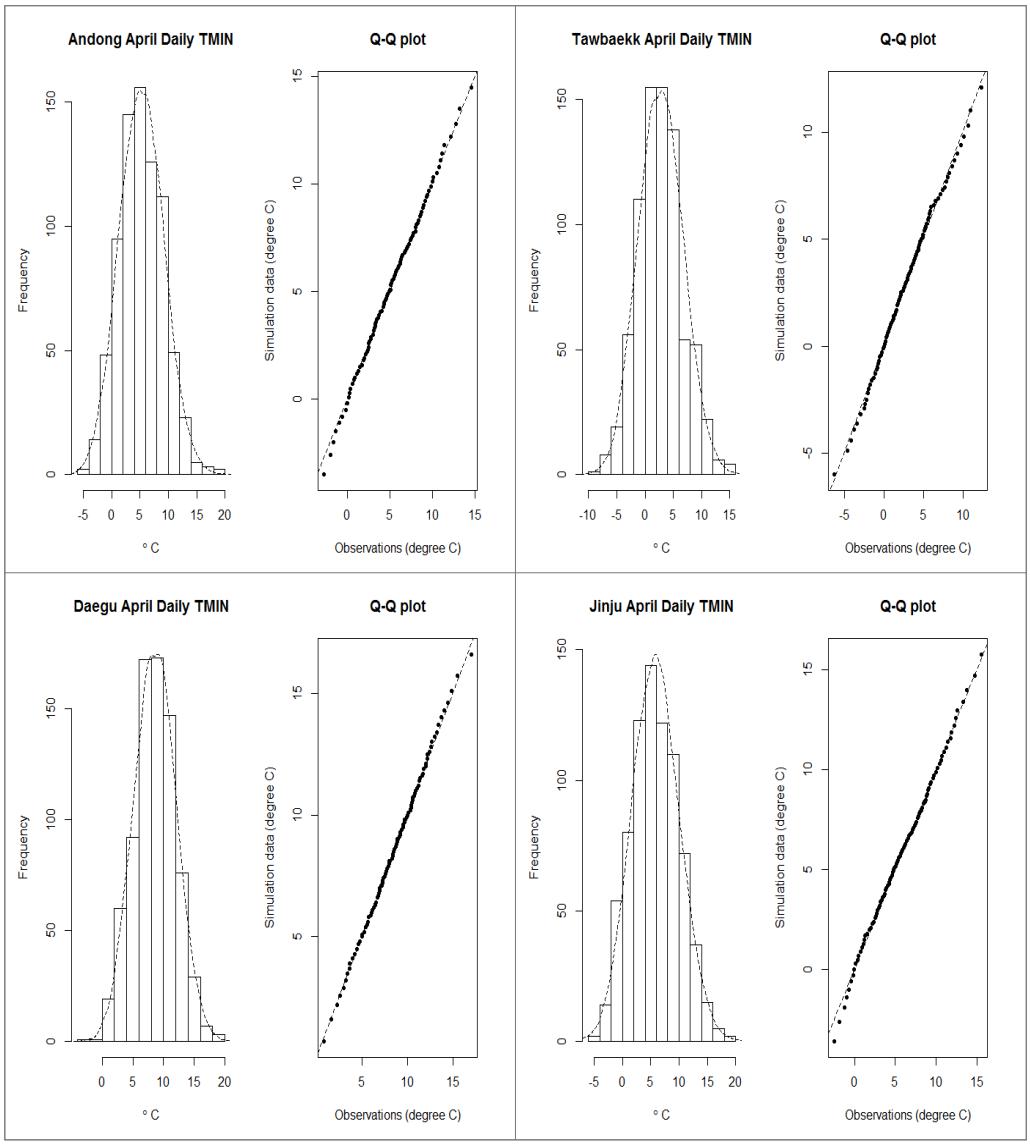


그림 34 4월 안동, 태백, 대구, 진주의 최저기온 일별 분포 비교. 분포가 매우 유사하다.

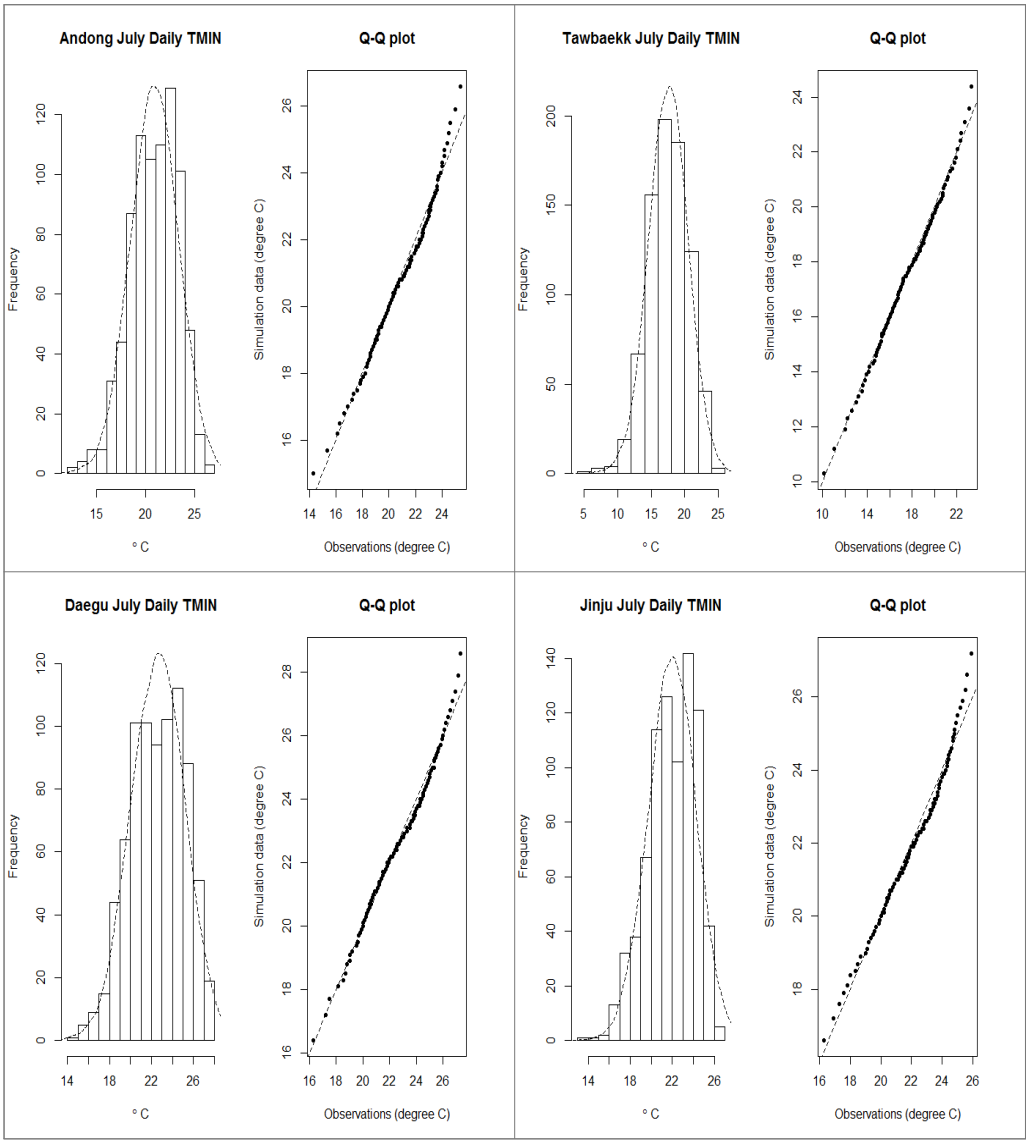


그림 35 7월 안동, 태백, 대구, 진주의 최저기온 일별 분포 비교. 분포가 매우 유사하다.

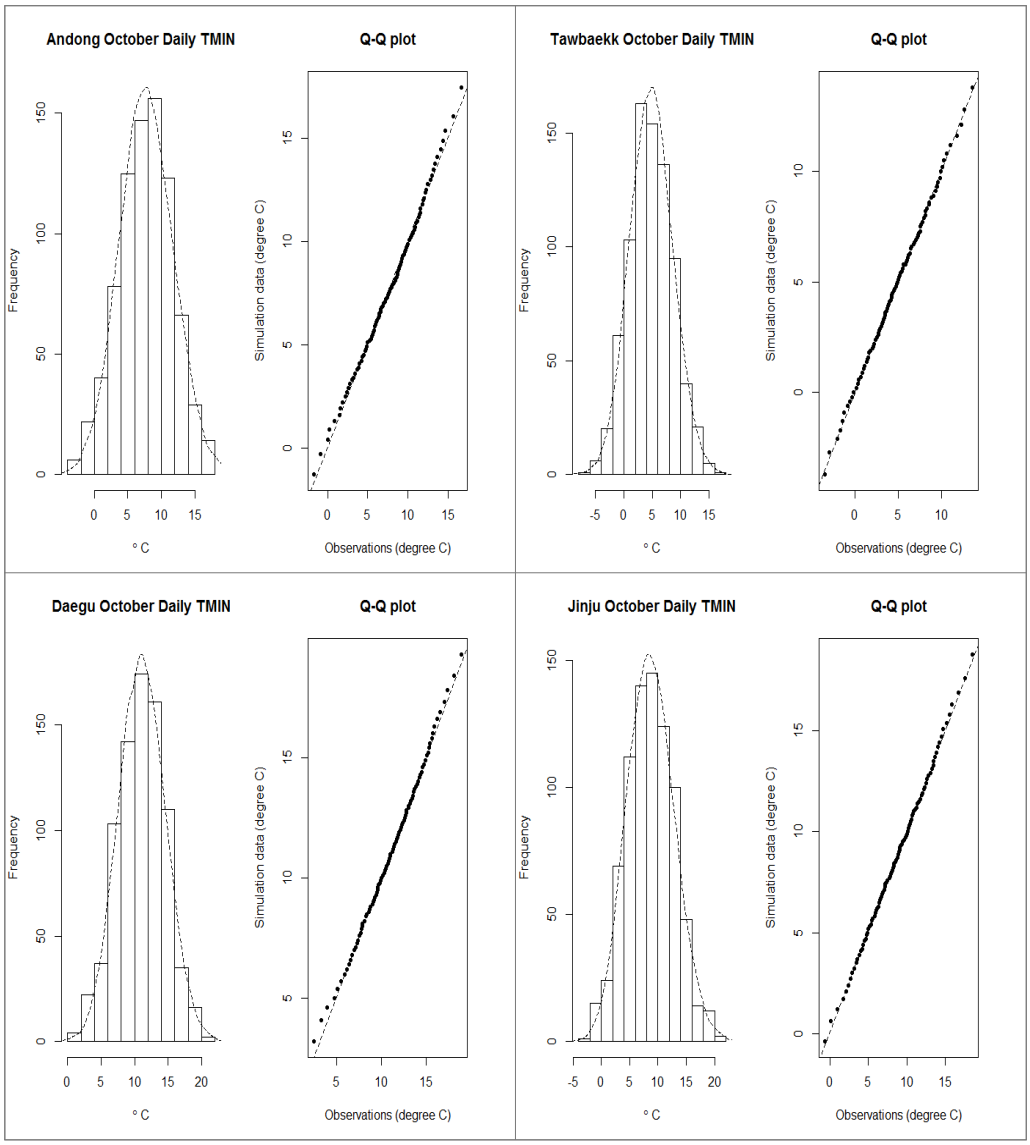


그림 36 10월 안동, 태백, 대구, 진주의 최저기온 일별 분포 비교. 분포가 매우 유사하다.

3.2.2 공간적 자기 상관

이제 강수와 기온의 공간적 자기상관을 검증해 보고자 한다. 이 통계적 특성은 다지점 weather generator의 개발 동기로서 여러 다지점 weather generator 연구에서 중요하게 고려되고 있다. 공간 상관의 지표로 지점 간의 Spearman 상관계수를 쓰고자 한다. 통상적으로 쓰이는 Pearson 상관계수보다 이상점(outlier)에 강건하다고 알려져 있어, 지점별 분포특성의 영향을 받지 않으면서 공간상관을 나타내는데 적절하다.

먼저, 강수의 공간적 자기상관을 살펴본다. 앞서 언급하였듯이, 강수생성알고리즘이 copula 모형에 따라 구분되기 때문에, 모형선택에 따라 모의성능을 평가할 것이다. 그림 37-39은 준경험적 copula 모형 기반 강수생성 알고리즘의 공간상관 모의정확도를 나타낸다. 그림에서 open circle (◦)은 지점 간의 상관계수를 나타내는데, x-좌표는 관측자료, y-좌표는 모의자료의 상관계수를 나타낸다. 즉, 두 좌표값이 유사할수록 (점선으로 표시된 기울기 1의 직선에 가까울수록) 모의정확도가 높음을 의미한다. 아울러, 관측자료의 표본상관계수의 불확실성을 감안하여, 강수모형의 표본상관계수의 95% 분포범위[9] (I-빔형태)를 함께 표시하였다. 즉, 그림에서 범위구간이 점선의 보조직선과 겹쳐 있으면, 상관계수의 차이가 유의수준 5%에서 유의하지 않다. 그림을 통해 준경험적 copula 모형 기반의 강수생성알고리즘이 대체적으로 공간상관을 잘 모의하고 있음을 알 수 있다. 한편, 그림 40-42은 Gaussian copula 모형 기반 알고리즘의 공간상관 정확도를 나타내는 그림인데, 모든 월에서 공간상관이 역시 잘 모의되고 있다. 그러므로, 두 알고리즘 모두가 공간 상관을 잘 모의하고 있다고 결론을 내린다.

앞서, 준경험적 copula 모형의 한계점을 2.1.2.절에서 지적한 바 있다. 한편, Gaussian copula 모형은 model bias를 지닐 수 있다. 그런데, 검증을 통해, model bias 가능성이 현저히 낮다는 점을 확인하였다. 이 경우에, 우리는 다양한 다지점 강수량을 생성할 수 있는 Gaussian copula 모형을 선호한다.

---

9) 95% 분포범위: 표본상관계수의 확률분포에서 하위 2.5% 분위수에서 상위 2.5% 분위수까지의 구간을 의미한다.

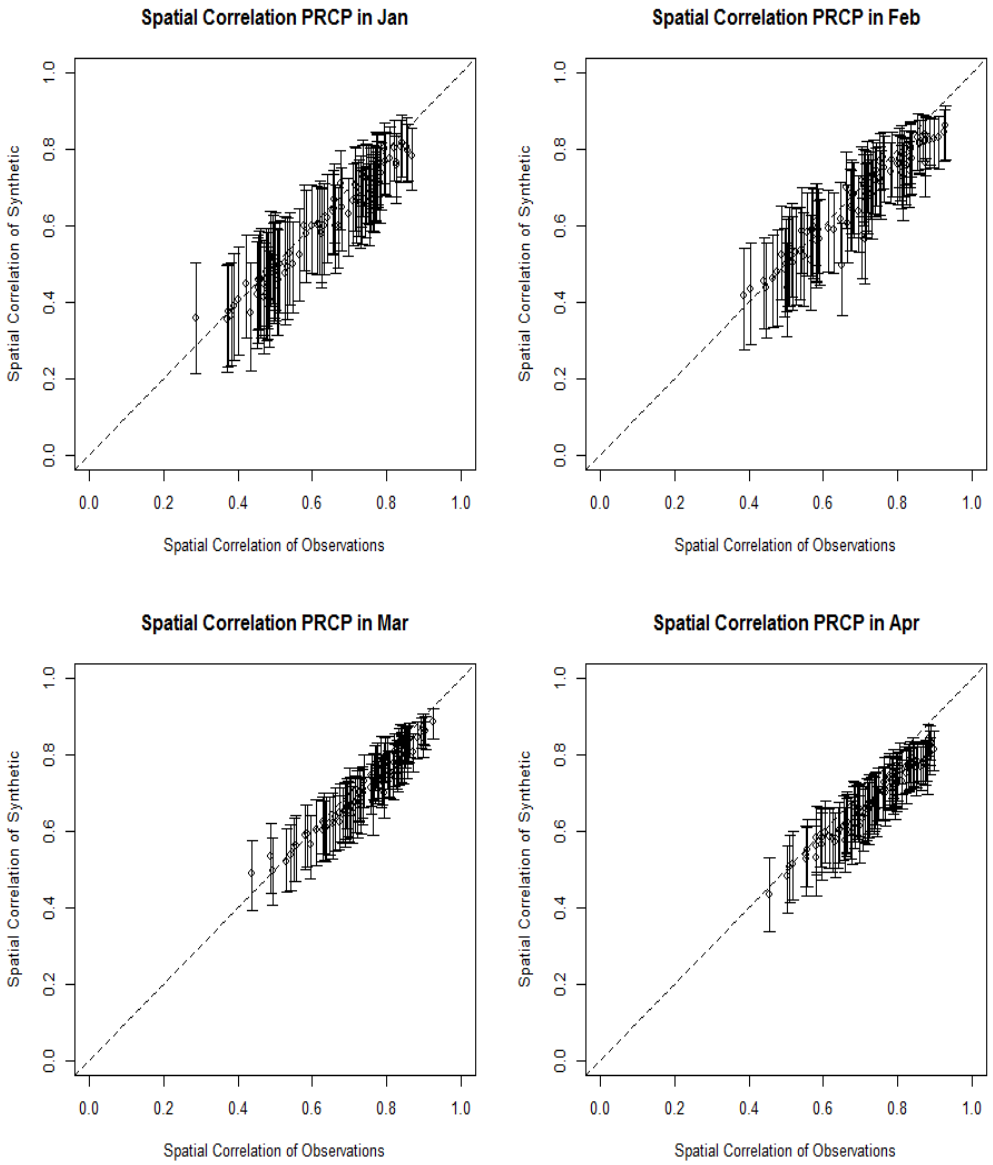


그림 37 준경험적 copula 모형에 기반한 강수생성알고리즘에 대한 1월-4월 강수의 공간적 자기상관 비교. 각 open circle ∘ 은 지점 간의 상관계수를 나타내는데, x-좌표는 관측자료, y-좌표는 모의자료의 상관계수를 뜻한다. 즉, 점선으로 표시된 직선에 가까울수록 모의정확도가 높다는 의미이다. 표본상관계수의 불확실성을 감안하여, 95% 분포범위(I-빔형태)를 함께 표시하였다. 거의 모든 지점 간 상관이 잘 모의되고 있다.

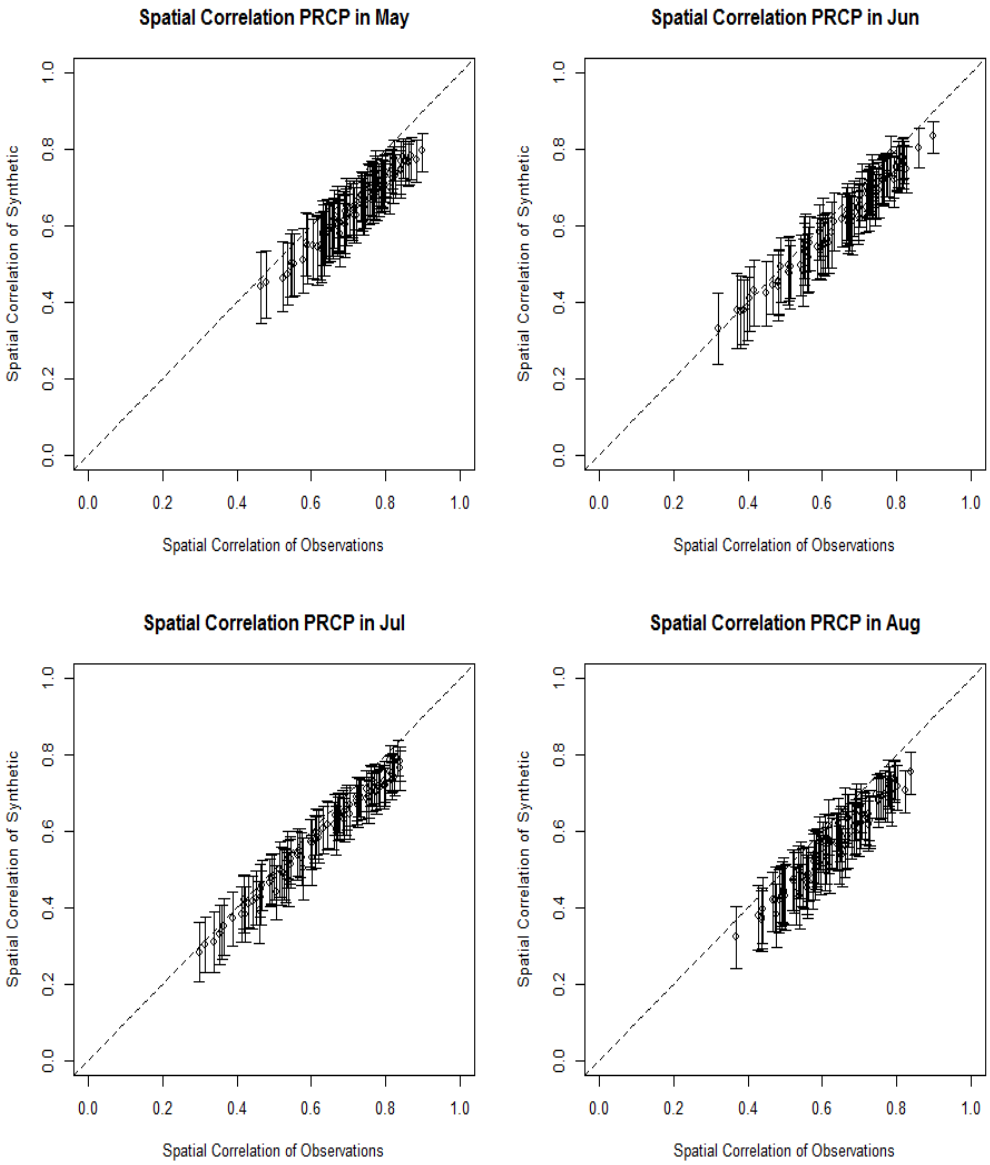


그림 38 준경험적 copula 모형에 기반한 강수생성알고리즘에 대한 5월-8월 강수의 공간적 자기상관 비교. 5월과 8월에서 다소 관측에 비해 약한 모의상관이 나타나지만, 대체적으로 공간상관이 잘 모의되고 있다.

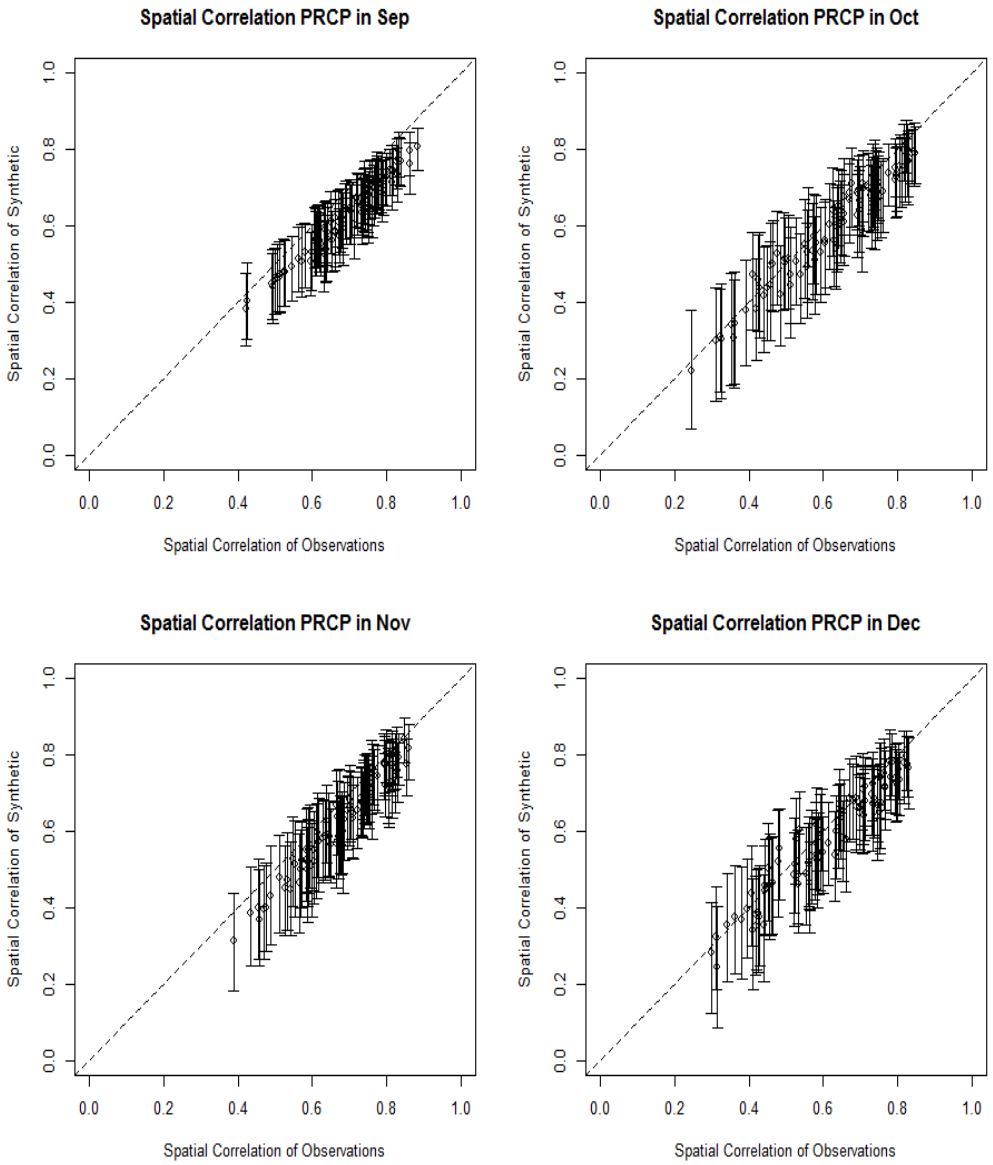


그림 39 준경험적 copula 모형에 기반한 강수생성알고리즘에 대한 9월-12월 강수의 공간적 자기상관 비교. 거의 모든 지점 간 상관이 잘 모의되고 있다.

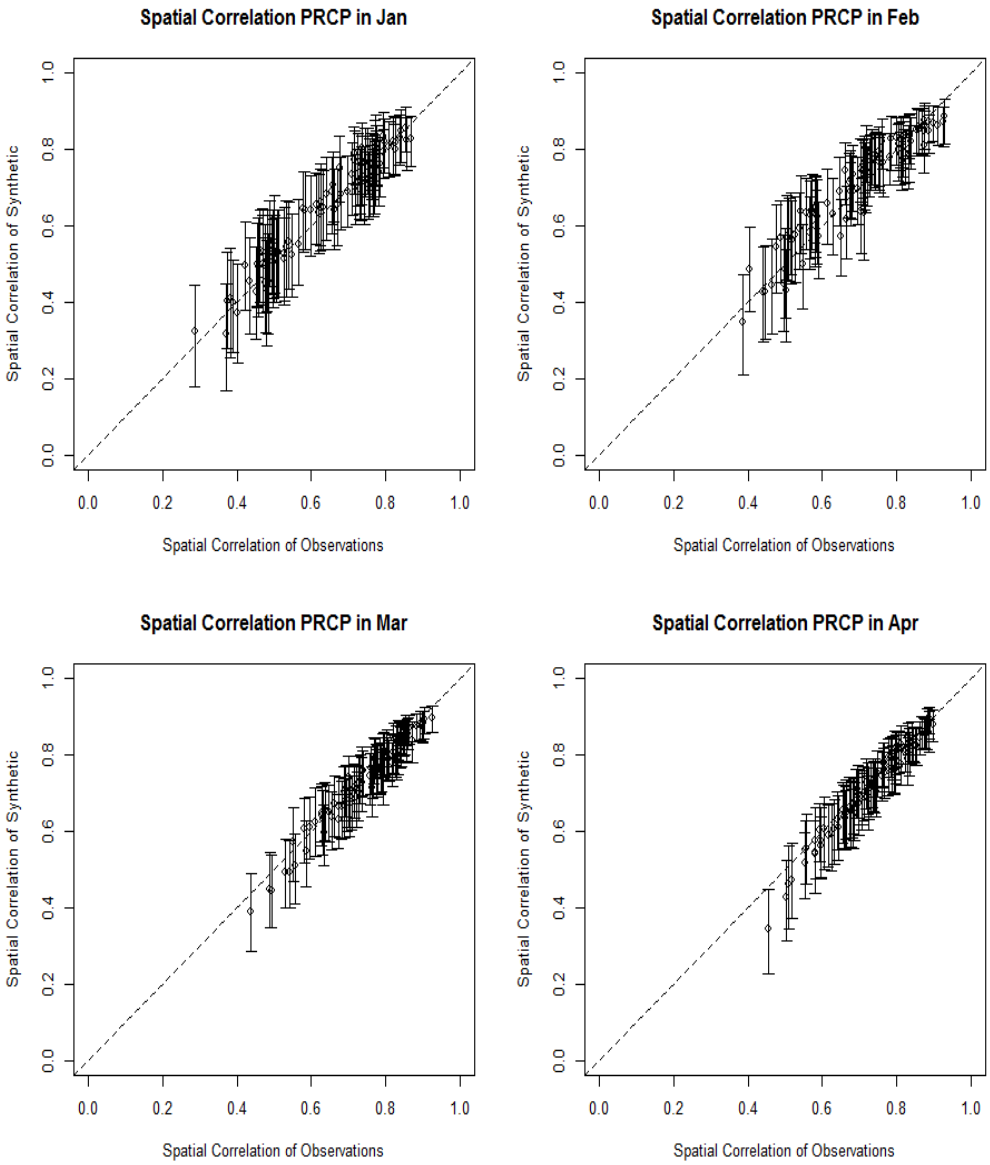


그림 40 Gaussian copula 모형에 기반한 강수생성알고리즘에 대한 1월-4월 강수의 공간적 자기상관 비교. 거의 모든 지점 간 상관이 잘 모의되고 있다.

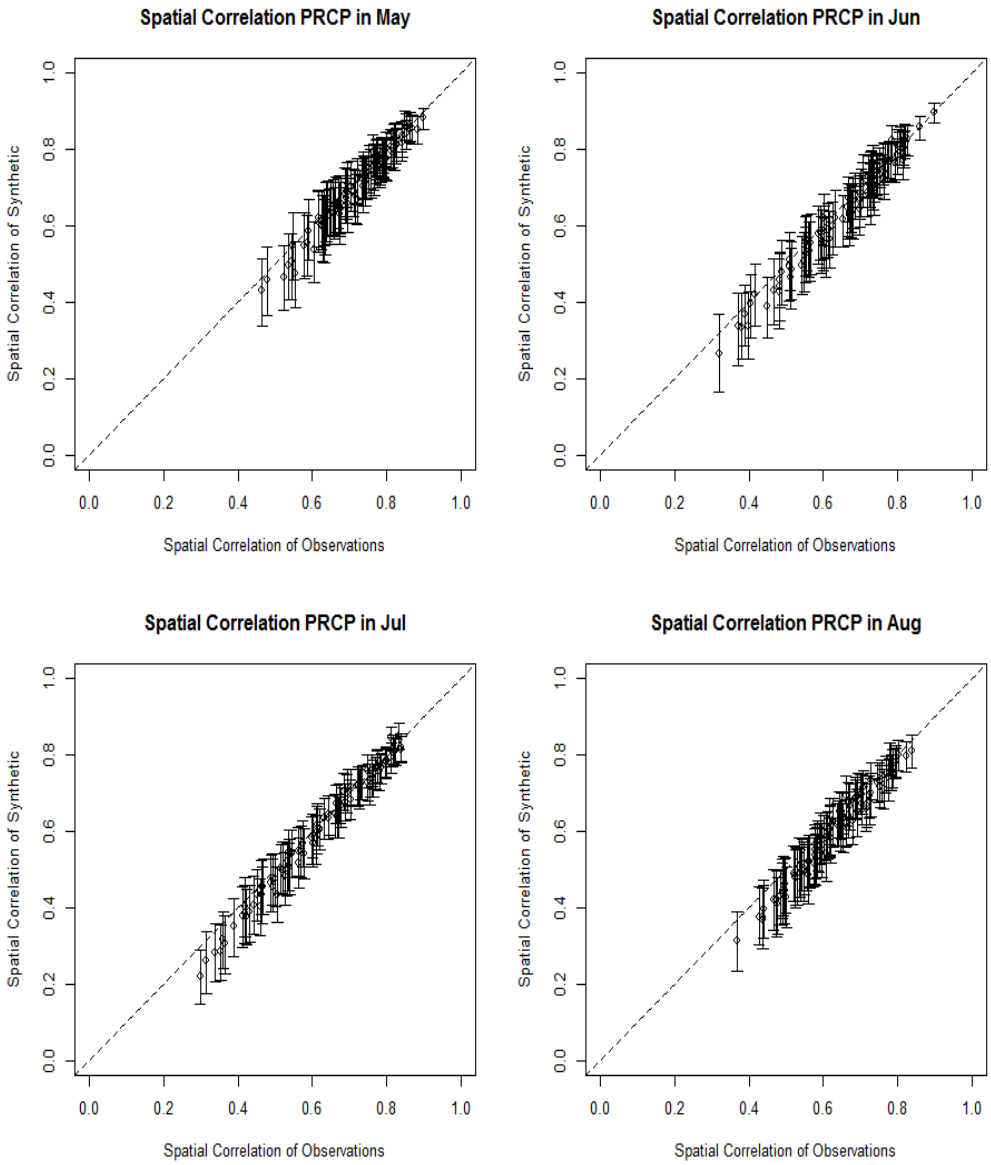


그림 41 Gaussian copula 모형에 기반한 강수생성알고리즘에 대한 5월-8월 강수의 공간적 자기상관 비교. 거의 모든 지점 간 상관이 잘 모의되고 있다.

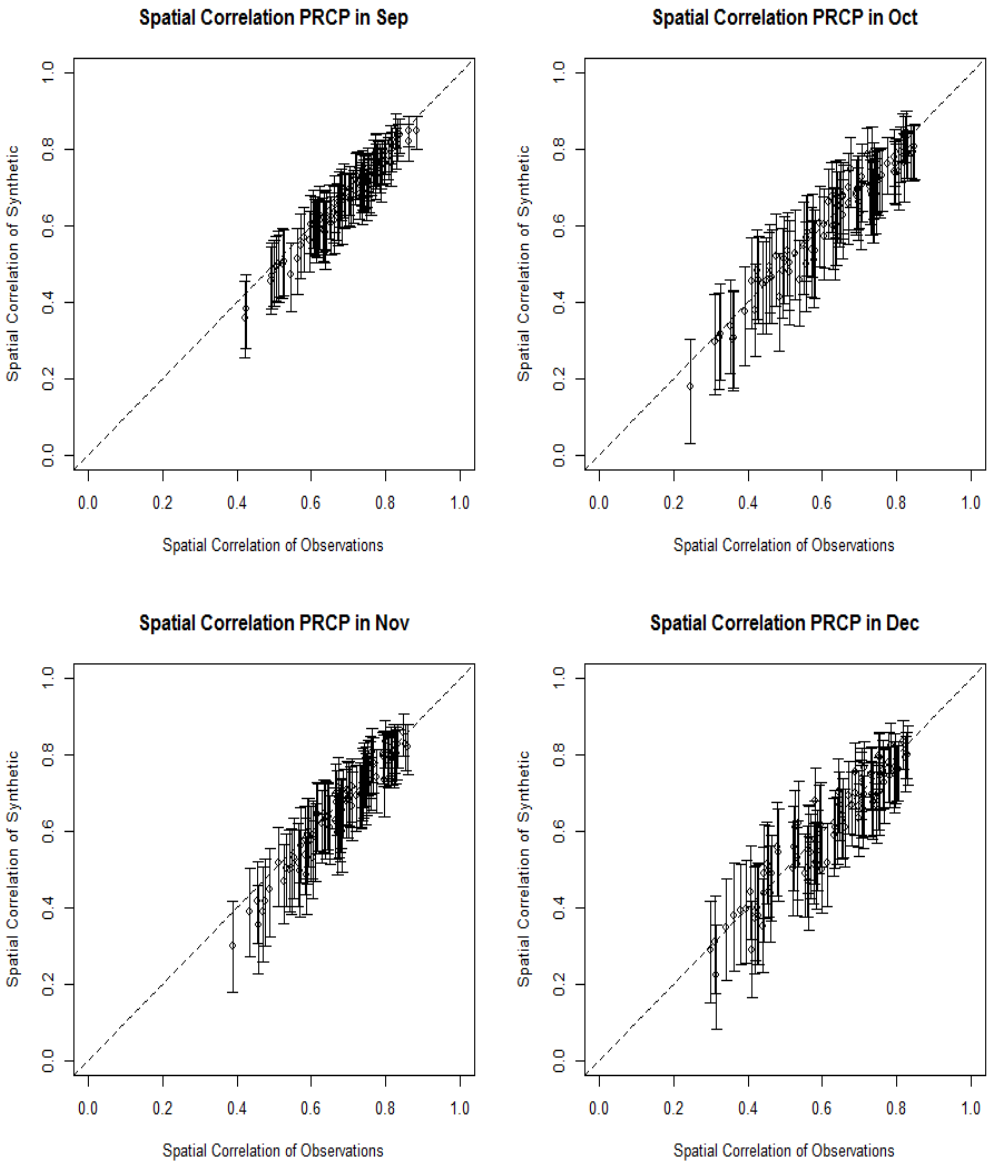


그림 42 Gaussian copula 모형에 기반한 강수생성알고리즘에 대한 9월-12월 강수의 공간적 자기상관 비교. 거의 모든 지점 간 상관이 잘 모의되고 있다.

이제, 기온에 대한 공간상관을 살펴본다. 기온의 일별분포를 검증할 때와 같이, 기온생성에서 Gaussian copula 모형 기반 강수생성 알고리즘을 쓴 경우를 고려한다. 그림 43–45은 최고기온의 공간적 자기상관 정확도를 나타낸 그림이다. 9월을 제외

한 모든 월에서 공간상관이 대체적으로 잘 모의되고 있다. 그림 46-48은 최저기온에 대한 공간상관 정확도를 나타낸다. 역시 모든 월에서 공간상관이 잘 모의되고 있다. 그러므로, 우리는 전반적으로 기온의 공간적 자기상관이 잘 모의되고 있다고 결론을 내린다.

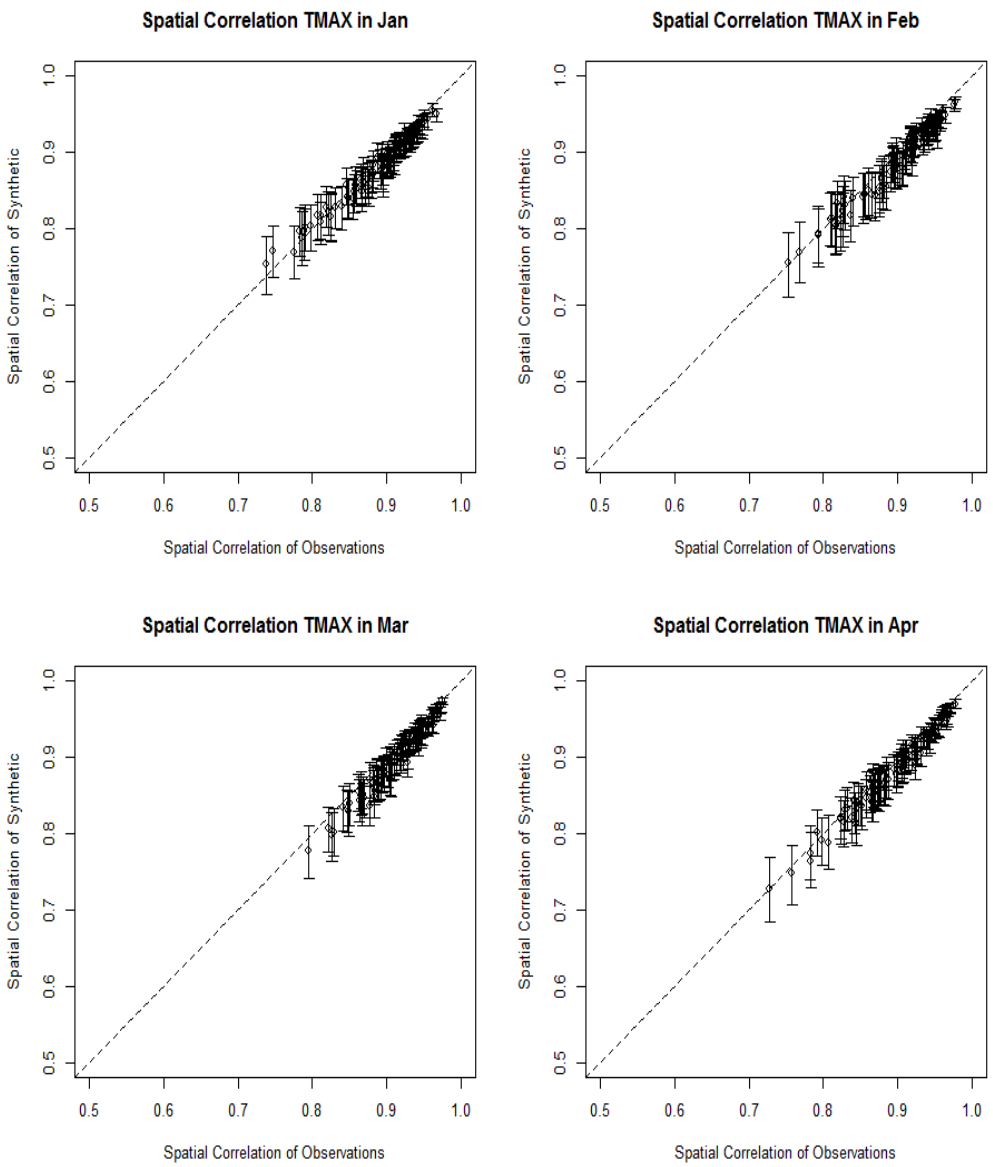


그림 43 1월-4월 최고기온의 공간적 자기상관 비교. 공간 상관이 잘 모의되고 있다.

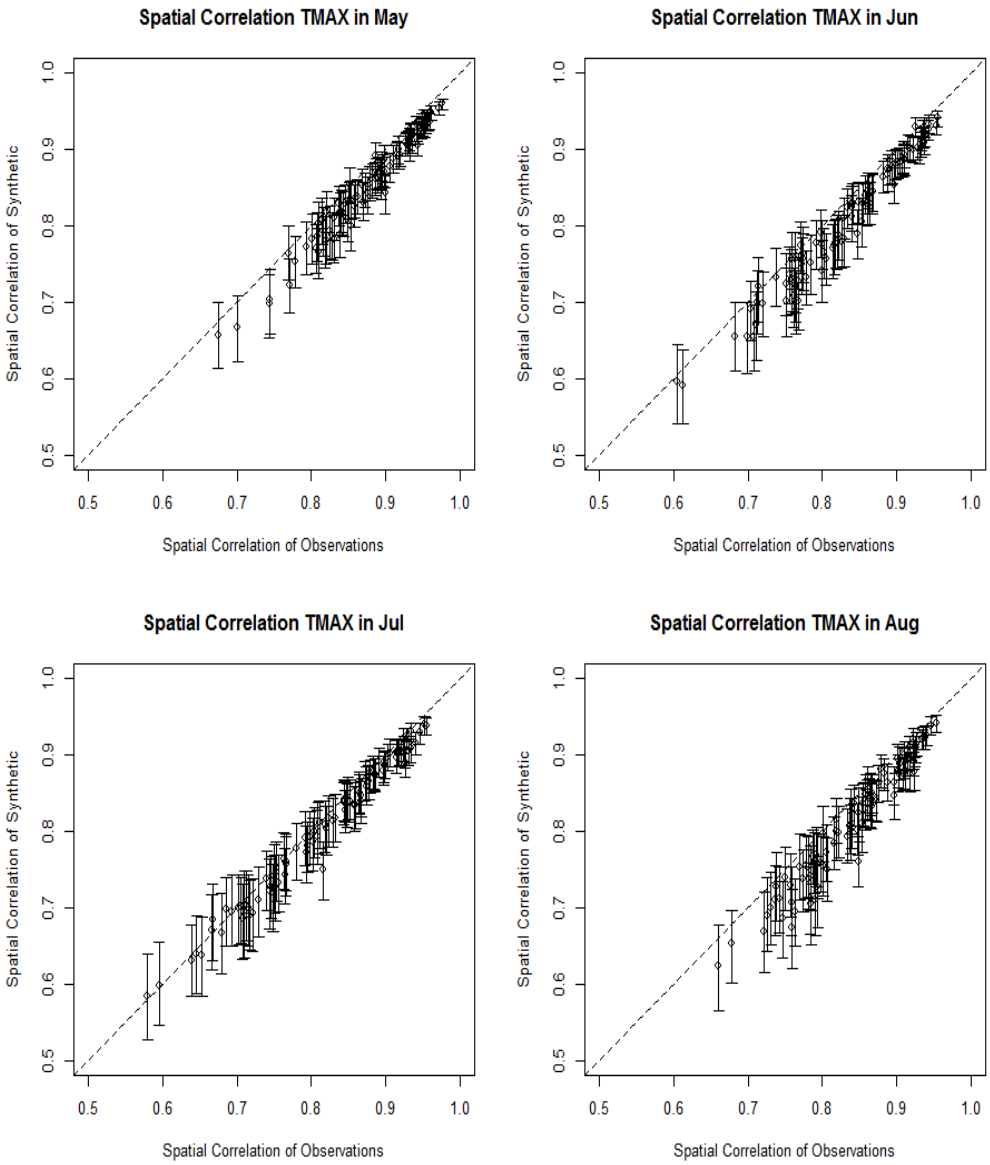


그림 44 5월-8월 최고기온의 공간적 자기상관 비교. 관측에 비해 약한 공간상관이 보이지만, 대체적으로 잘 모의되고 있다.

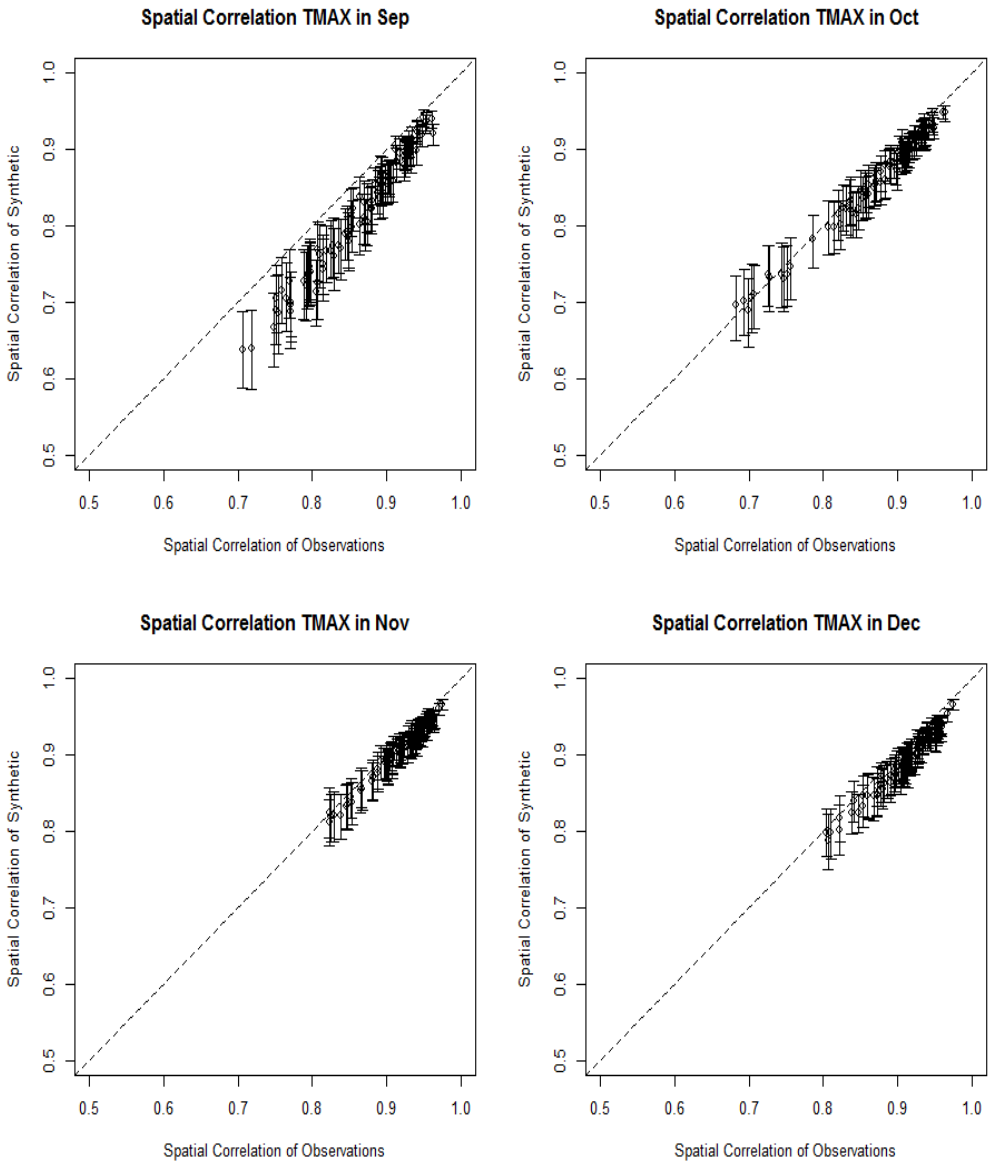


그림 45 9월-12월 최고기온의 공간적 자기상관 비교. 9월에서 유의한 공간상관 차이를 보인다. 나머지 월에서는 공간상관이 잘 모의되고 있다.

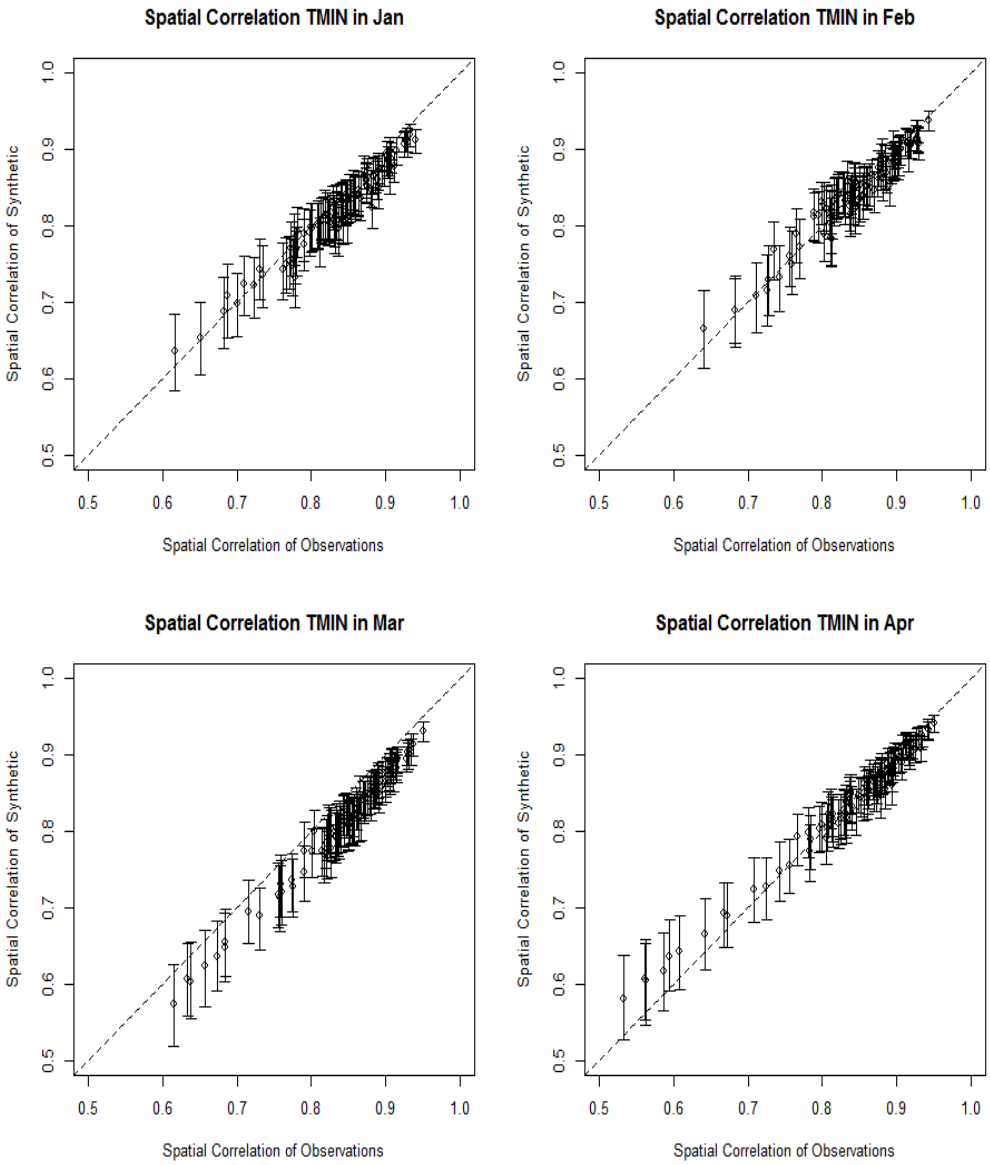


그림 46 1월-4월 최저기온의 공간적 자기상관 비교. 공간상관이 잘 모의되고 있다.

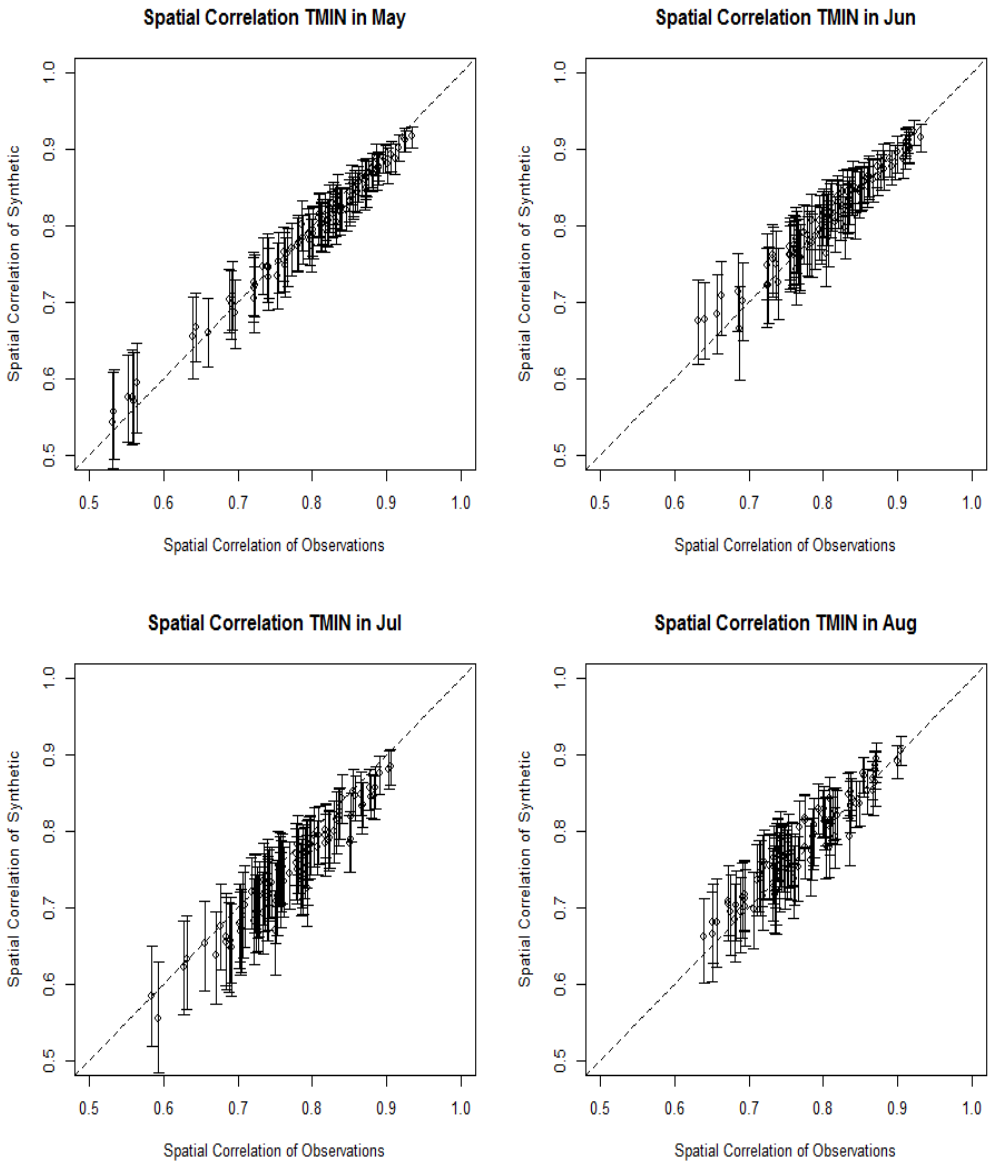


그림 47 5월-8월 최저기온의 공간적 자기상관 비교. 공간 상관이 잘 모의되고 있다.

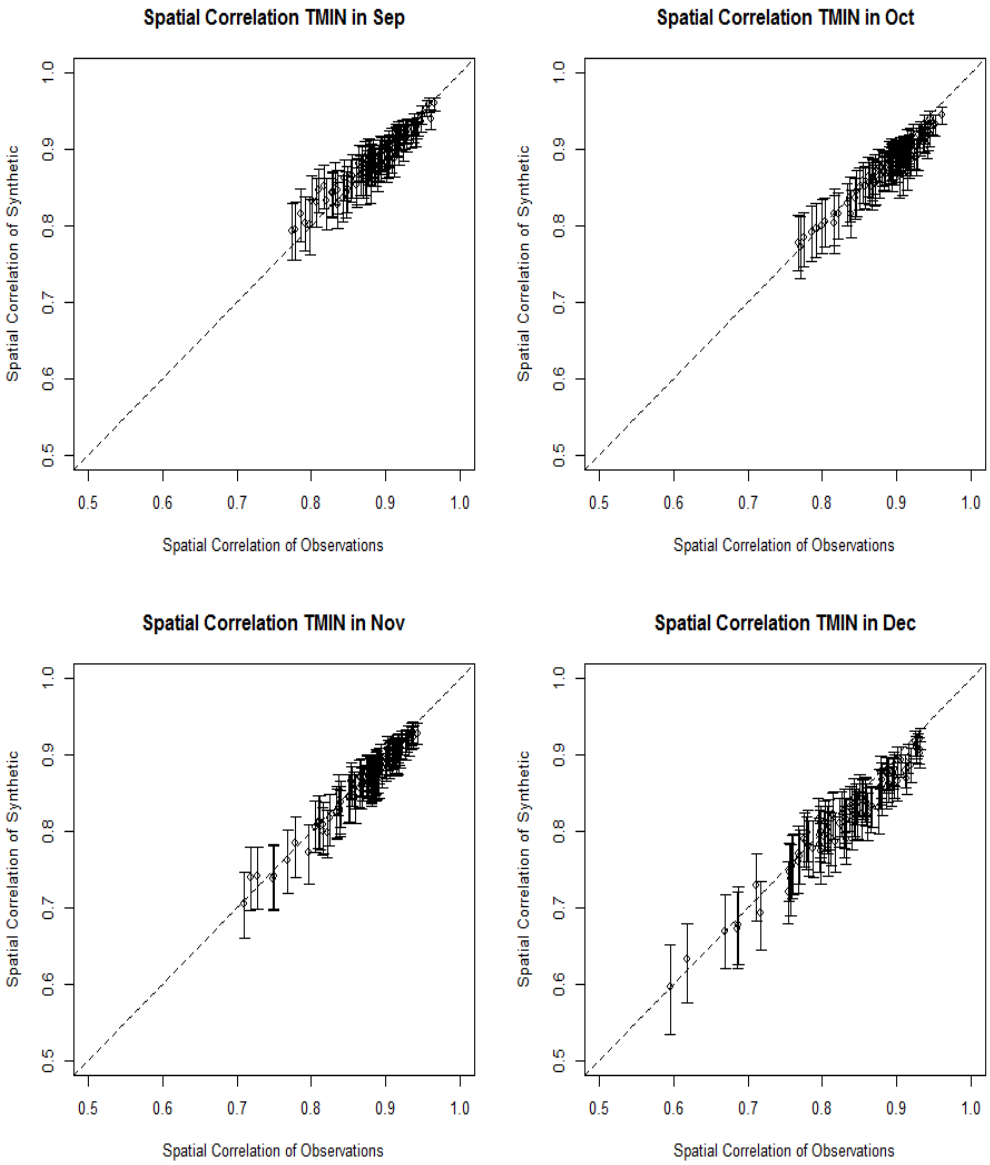


그림 48 9월-12월 최저기온의 공간적 자기상관 비교. 공간상관이 잘 모의되고 있다.

3.2.3 기온의 시간적 자기상관

기온의 시간적 자기상관의 모의성능을 평가한다. 기온은 강수와 다르게 뚜렷한 시간적 자기상관을 드러낸다. 그래서, 그러한 특징이 기온 생성 알고리즘을 통해 잘 모의되는지 확인할 필요가 있다. 시간적 자기상관 검증은 관측과 모의자료에 대해 1일차 자기상관계수를 계산하여 비교하는 것으로 실시한다. 그림 49-51는 각 월과 지점에 대해서 최고기온의 시간적 자기상관, 그림 52-54은 최저기온의 그것을 비교하고 있다. 그림에서 open circle (◦)은 기온모형의 자기상관계수이고, solid circle (•)은 관측자료의 것이다. 아울러, 공간적 상관검증에서와 같이, 관측자료의 표본자기상관계수의 불확실성을 감안하여 기온 모형의 표본자기상관계수의 95% 분포범위를 함께 제시한다. 만약 관측자료의 자기상관계수가 범위구간 내에 있다면, 유의수준 5%에서 차이가 유의하지 않다는 의미이다. 비교결과, 대체적으로 시간적 자기상관이 잘 모의되고 있다.

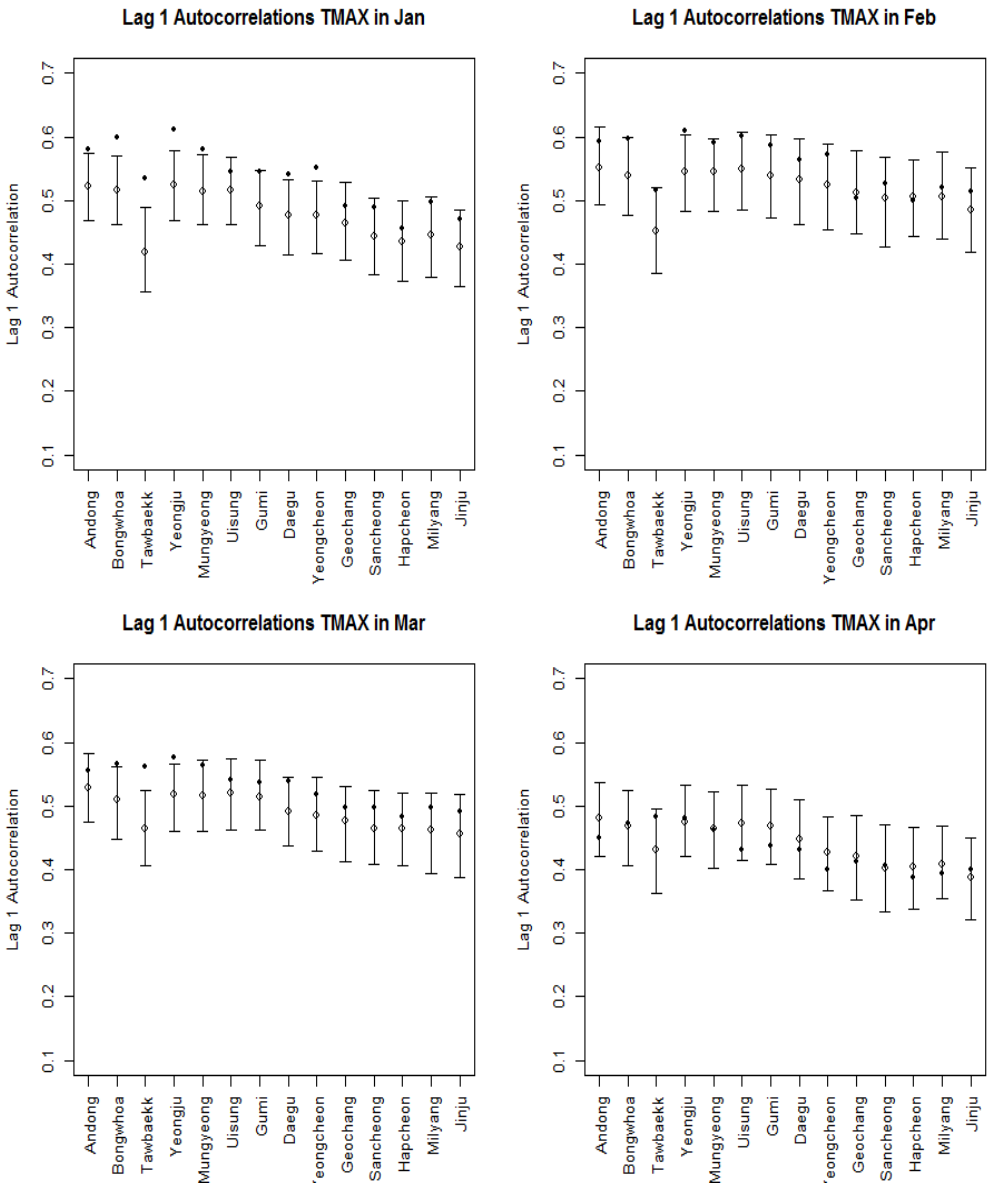


그림 49 1월-4월 각 지점에서 최고기온의 1일차 자기상관계수 비교. open circle (◦)은 기온모형의 자기상관계수, solid circle (•)은 관측자료의 것을 뜻한다. I빔형태의 구간은 기온모형의 표본상관계수의 95% 분포범위이다. 관측의 자기상관이 잘 모의되고 있다.

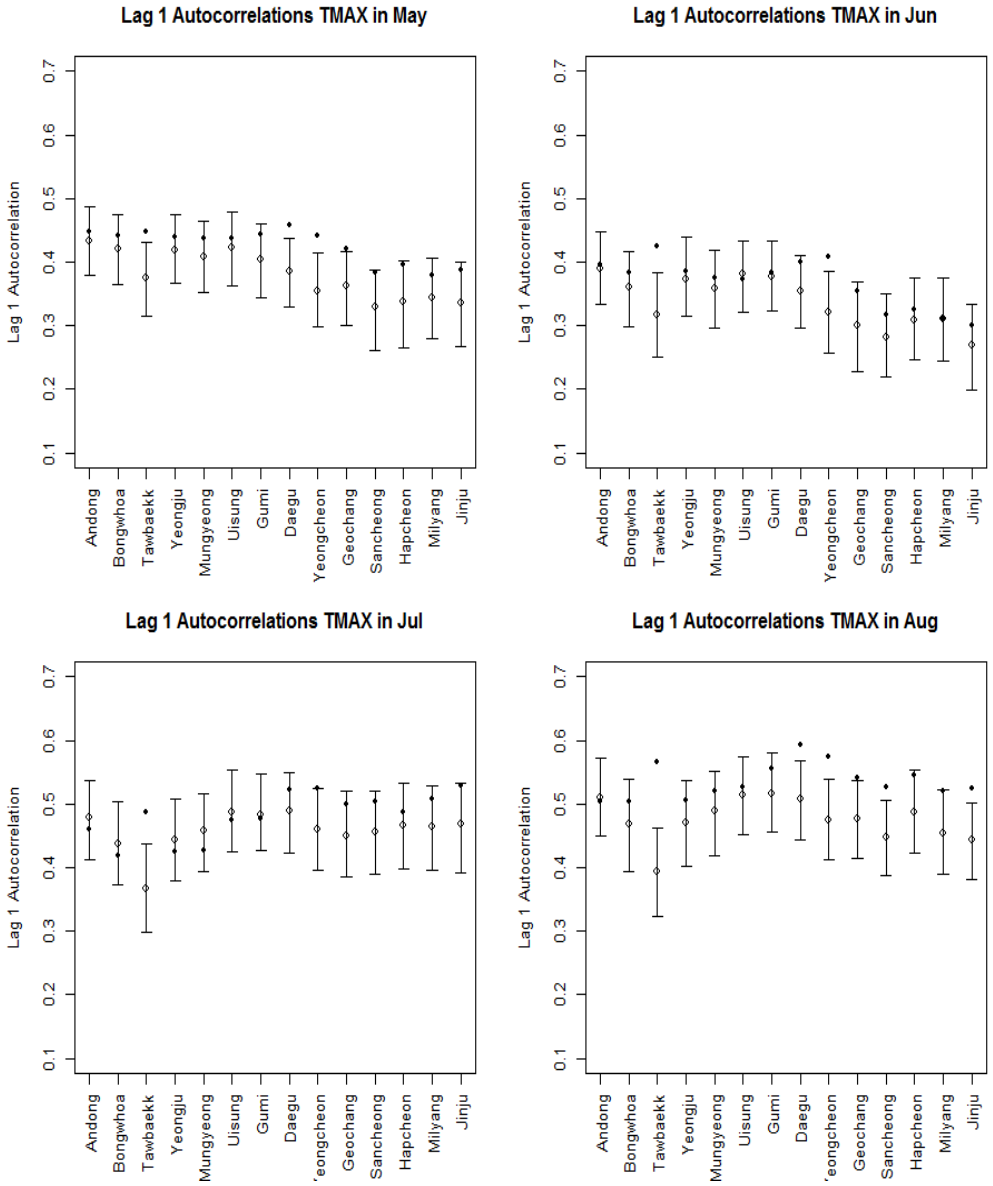


그림 50 5월-8월 각 지점에서 최고기온의 1일차 자기상관계수 비교. 관측의 자기상관이 대체적으로 잘 모의되고 있다.

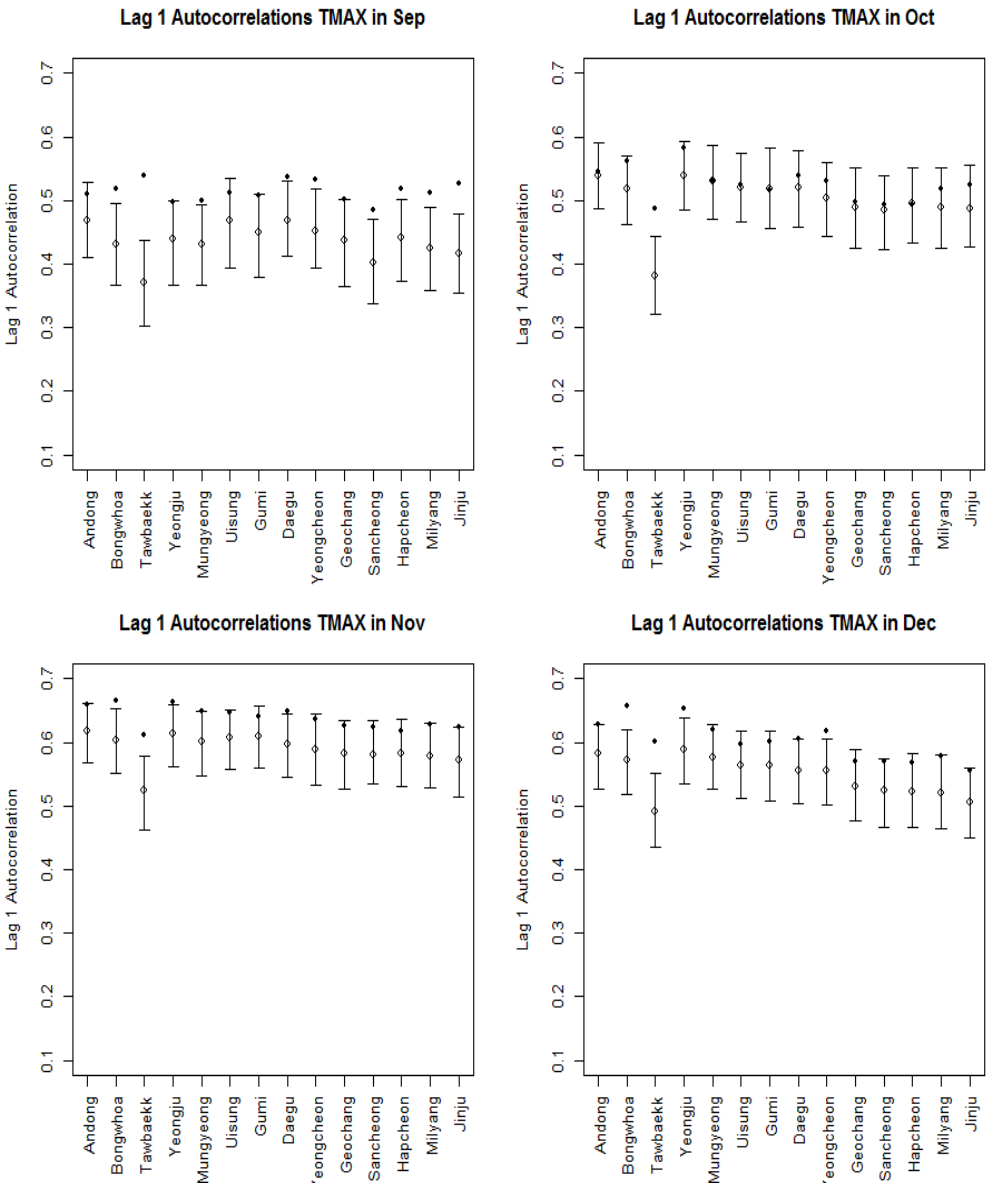


그림 51 9월-12월 각 지점에서 최고기온의 1일차 자기상관계수 비교. 관측의 자기상관이 대체적으로 잘 모의되고 있다.

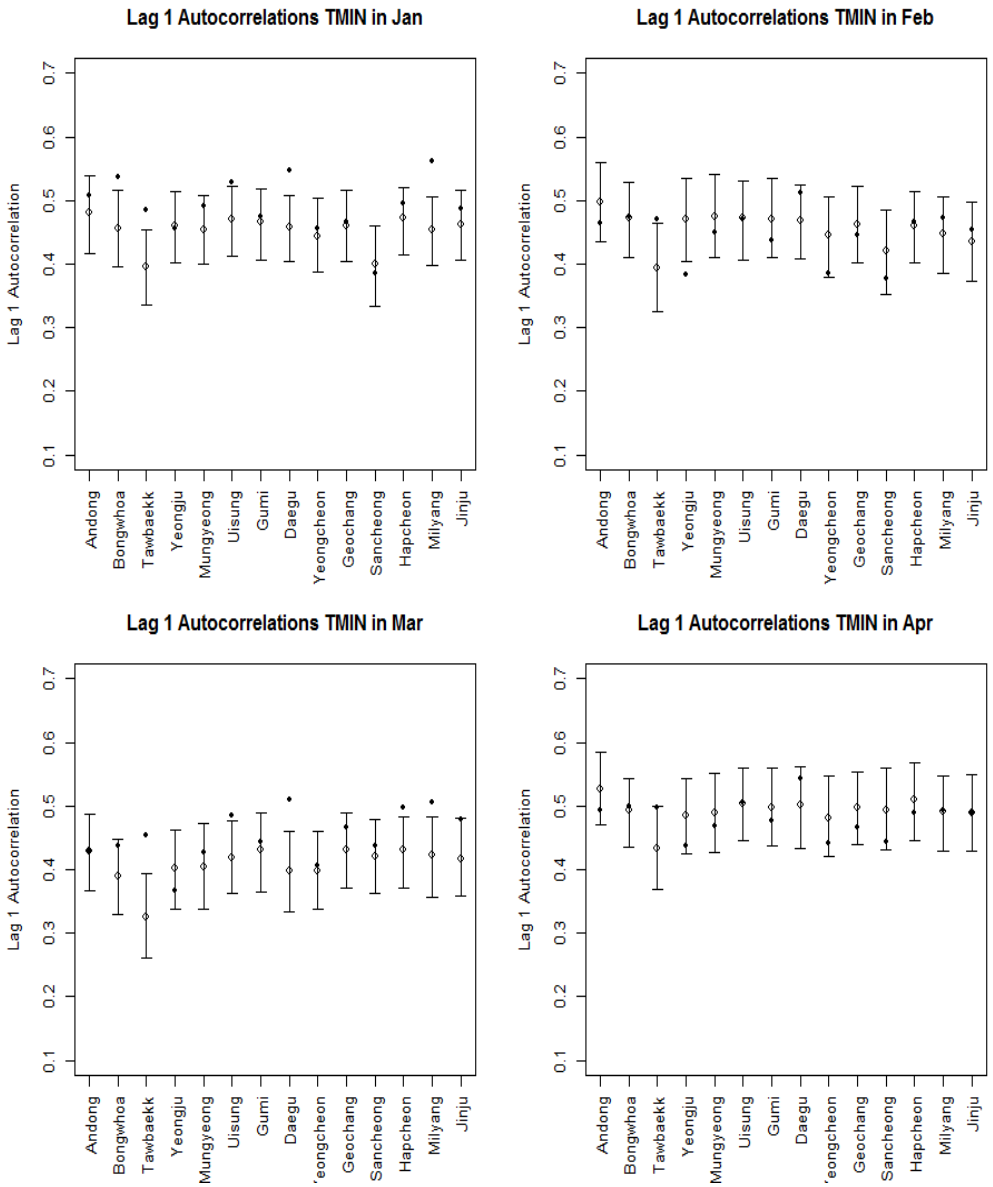


그림 52 1월-4월 각 지점에서 최저기온의 1일차 자기상관계수 비교. 관측의 자기상관이 잘 모의되고 있다.

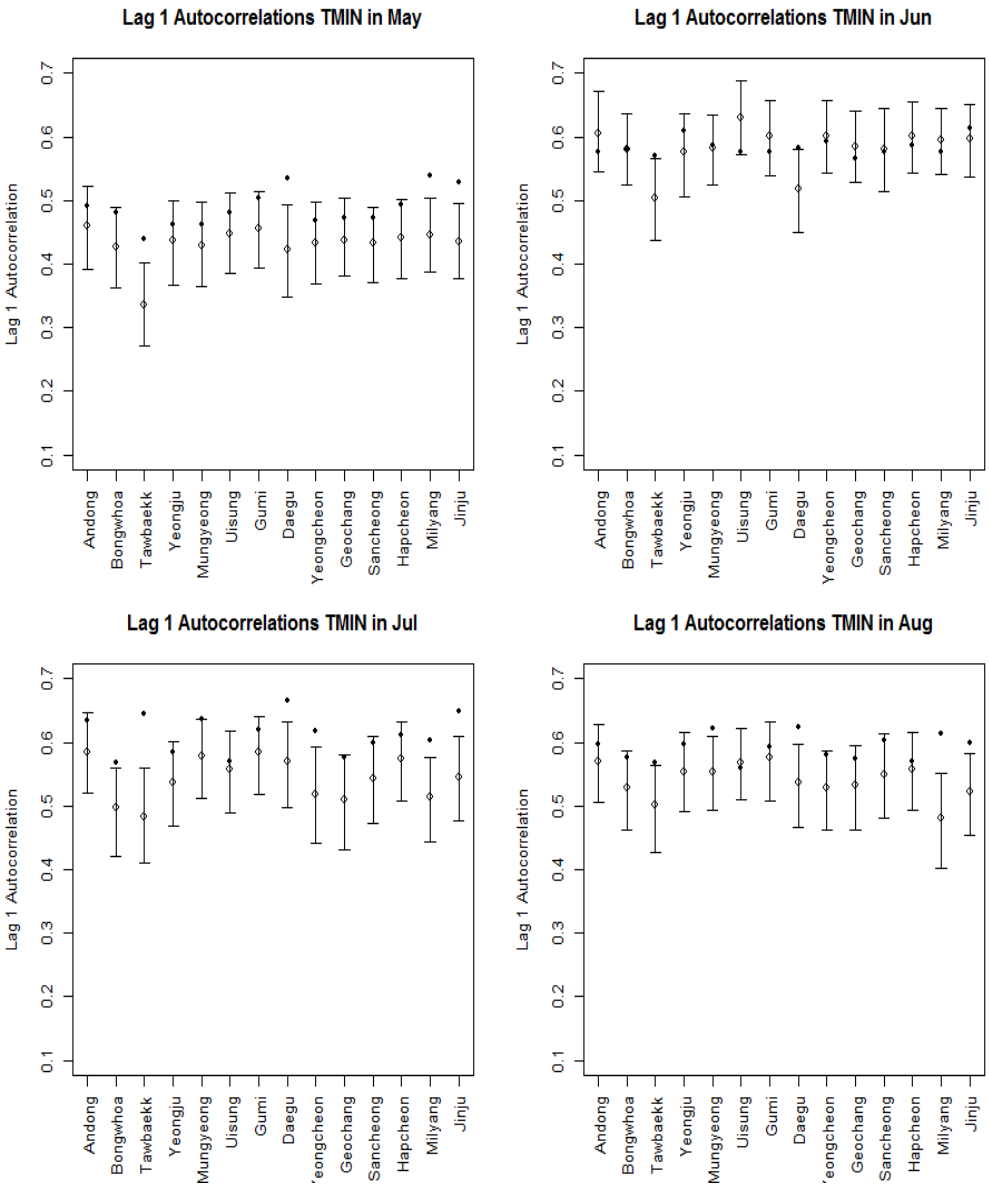


그림 53 5월-8월 각 지점에서 최저기온의 1일차 자기상관계수 비교. 관측의 자기상관이 대체적으로 잘 모의되고 있다.

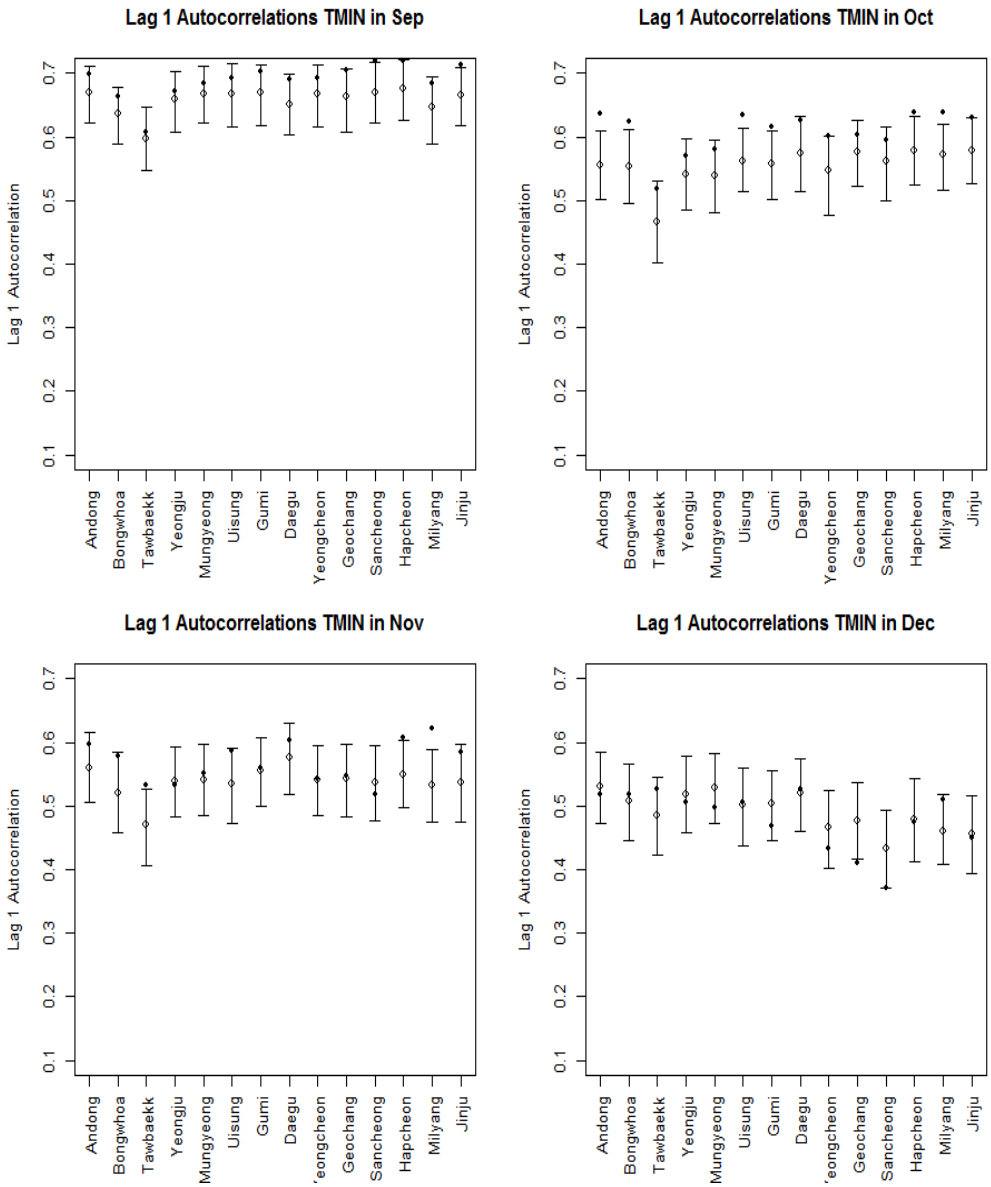


그림 54 9월-12월 각 지점에서 최저기온의 1일차 자기상관계수 비교. 관측의 자기상관이 잘 모의되고 있다.

3.2.4 최고/최저 기온 간 상관

이제, 최고/최저 기온 간 상관에 대한 모의성능을 검증한다. 하루 간 최고와 최저기온 사이에는 양의 상관관계가 있는데, 이 특성이 모의를 통해 재현되고 있는지 확인한다. 그림 55-57은 각 월에서 최고/최저 기온 간 상관관계를 나타낸다. 그림에서 기호의 의미는 시간적 자기상관그림에서 것과 대응한다. 관측자료와 모의자료 간 비교 결과, 전반적으로 최고/최저 기온 간 상관관계가 잘 모의되고 있는 것으로 판단된다.

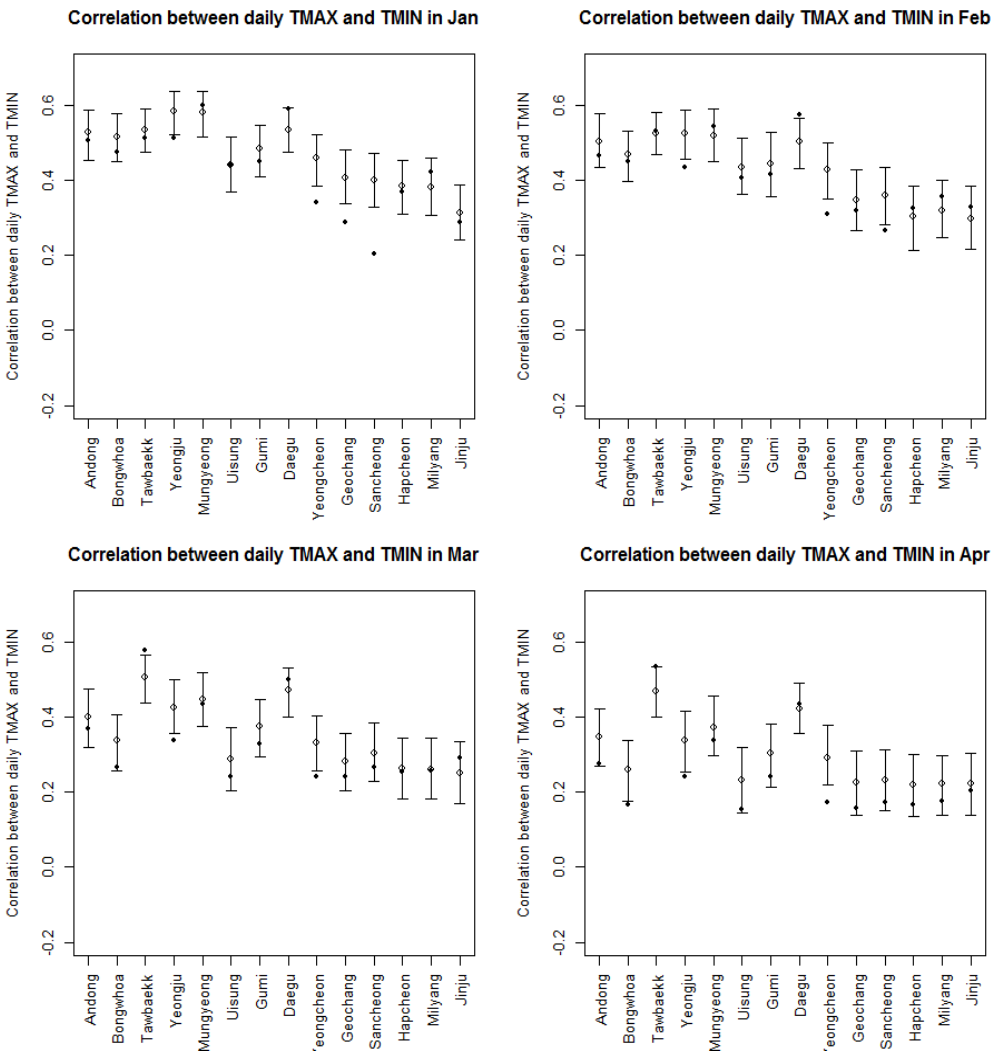


그림 55 1월-4월 각 지점에서 최고/최저기온 간 상관관계 비교. 몇몇 지점에서 관측과 모의자료 간 유의한 차이가 보이나, 대체적으로 잘 모의되고 있다.

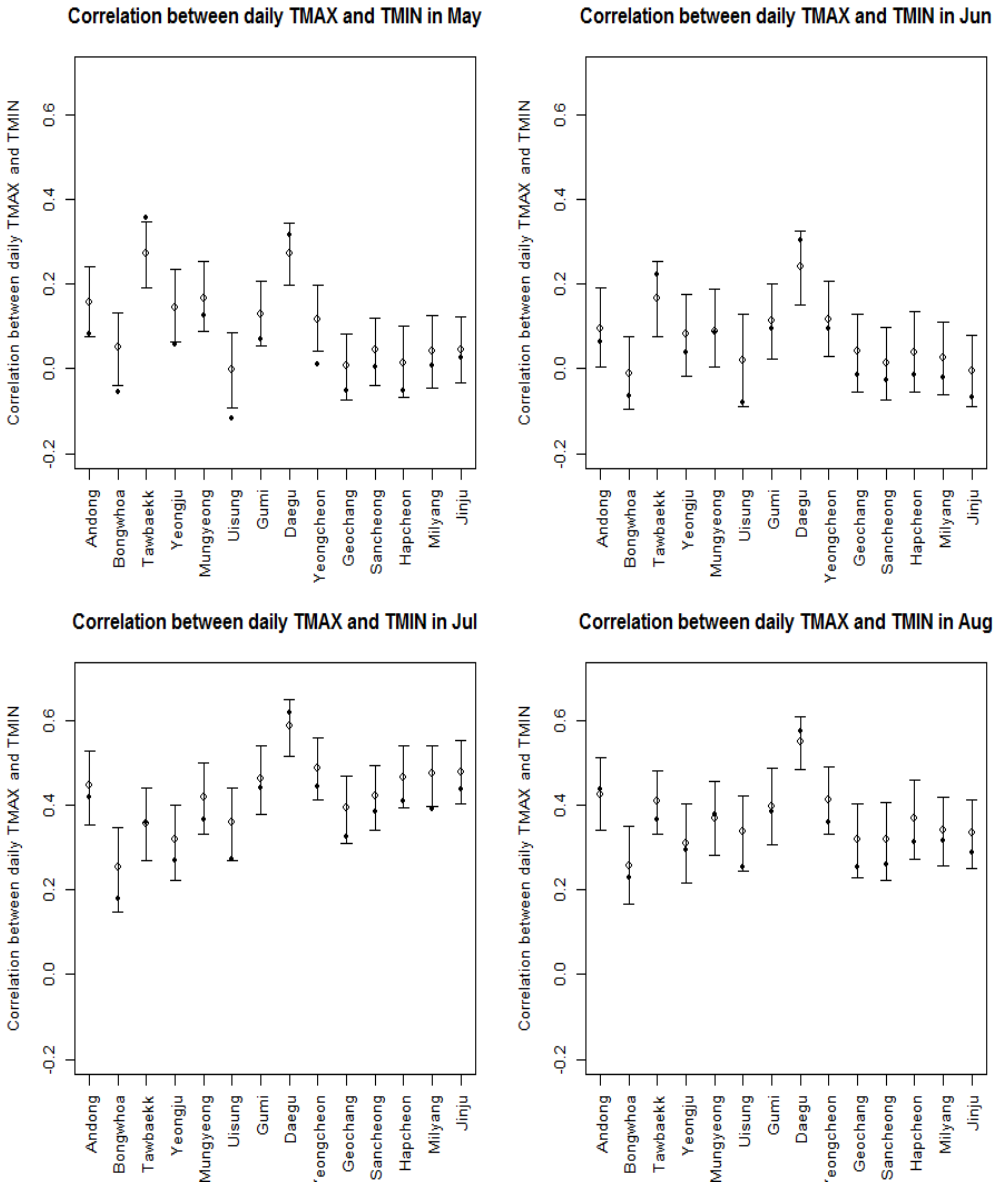


그림 56 5월-8월 각 지점에서 최고/최저기온 간 상관관계 비교. 대체적으로 잘 모의되고 있다.

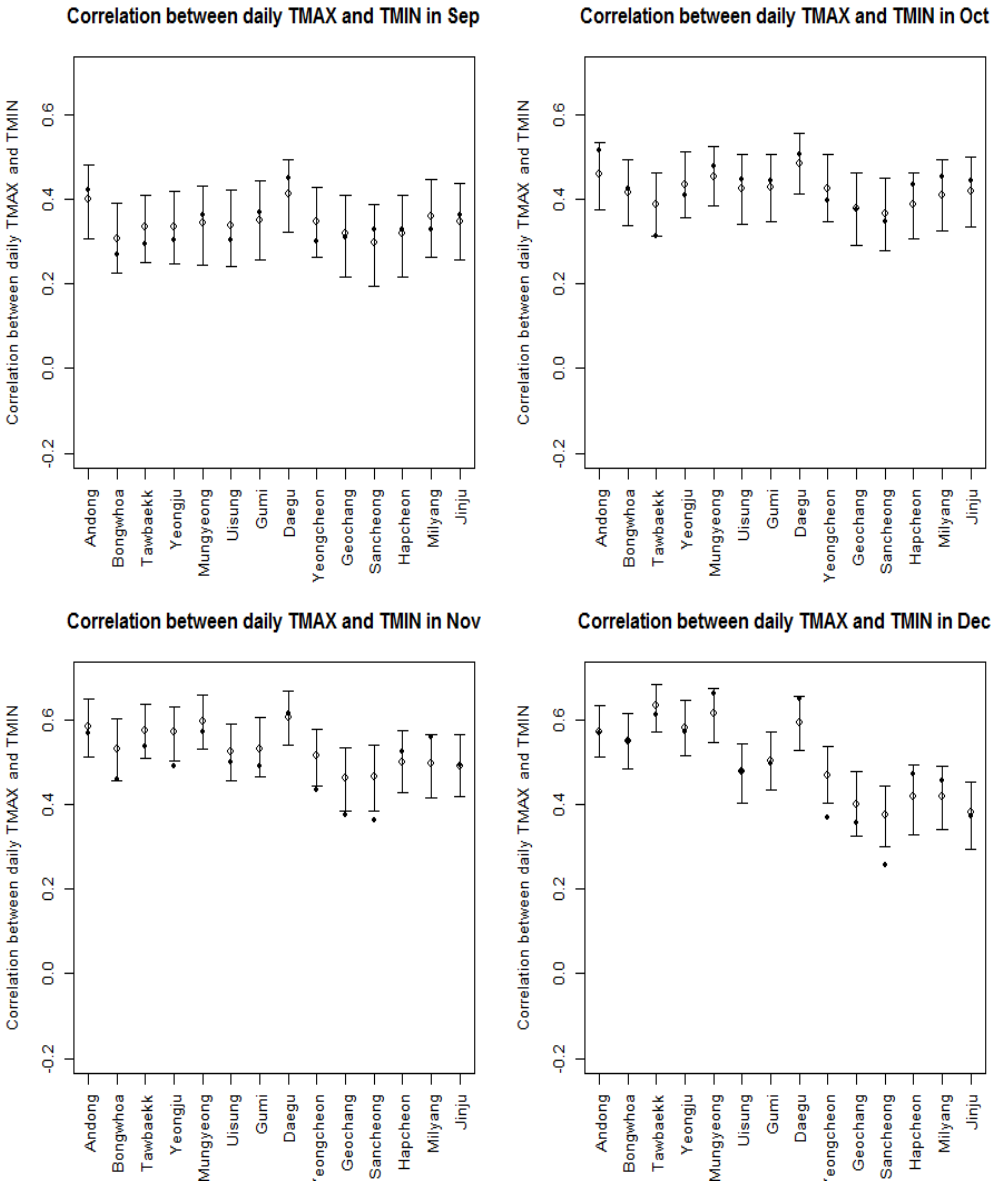


그림 57 9월-12월 각 지점에서 최고/최저기온 간 상관관계 비교. 대체적으로 잘 모의되고 있다.

3.2.5 월단위의 강수특성 모의성능 검증

지금까지 일단위에서 강수/기온 통계적 특성 재현정확도로 모의성능을 검증하였다면, 여기서부터는 월단위의 특성 재현성을 고려한다. 우선, 강수의 월누적 강수량, 최대일강수량, 습윤일수를 살펴보기로 한다. 지면 절약을 위해, Gaussian copula 모형 기반 알고리즘의 경우만 살핀다. 그림 58-64은 관측자료와 모의자료의 월누적 강수량 분포를 상자그림으로 비교한 것이다. 선택된 월은 1월, 4월, 6월, 7월, 8월, 9월, 10월인데, 1월, 4월, 10월은 각각 겨울, 봄, 가을을 대표해서, 6-9월은 우리나라의 호우기에 해당하기 때문에 선정되었다. 대체적으로 두 분포가 서로 유사하나, 모의자료에서 다수의 이상점(outlier)이 나타난다. 관측기간이 26년이고 1000개의 모의자료를 쓴 점을 감안하면, 다수의 이상점이 충분히 발생할 수 있다. 또한, 이상점의 등장은 제안한 강수생성알고리즘이 관측기간에서 나타나지 않았지만 미래에 발생할 수 있는 극단현상을 생성한다는 것으로 받아드릴 수 있다. 그림 65-71은 월 최대일강수량, 그림 72-78은 습윤일수의 분포를 비교하고 있는데, 대체적으로 분포가 서로 유사하다. 그래서, 우리는 강수의 월 특성이 잘 모의되고 있다고 결론을 내린다.

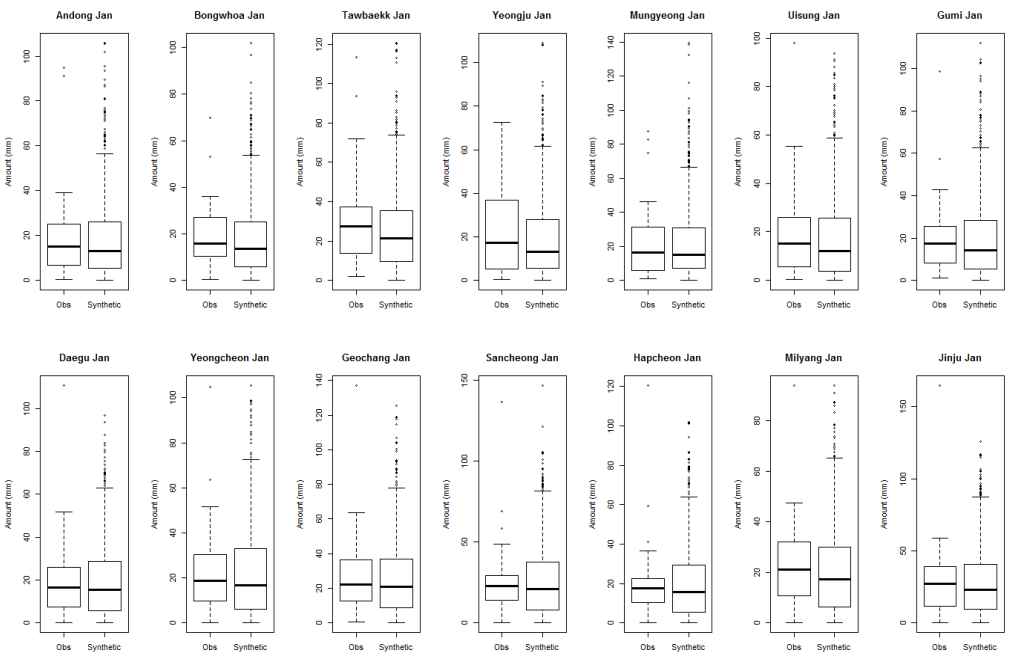


그림 58 1월 지점 별 월 누적 강수량의 분포 비교.

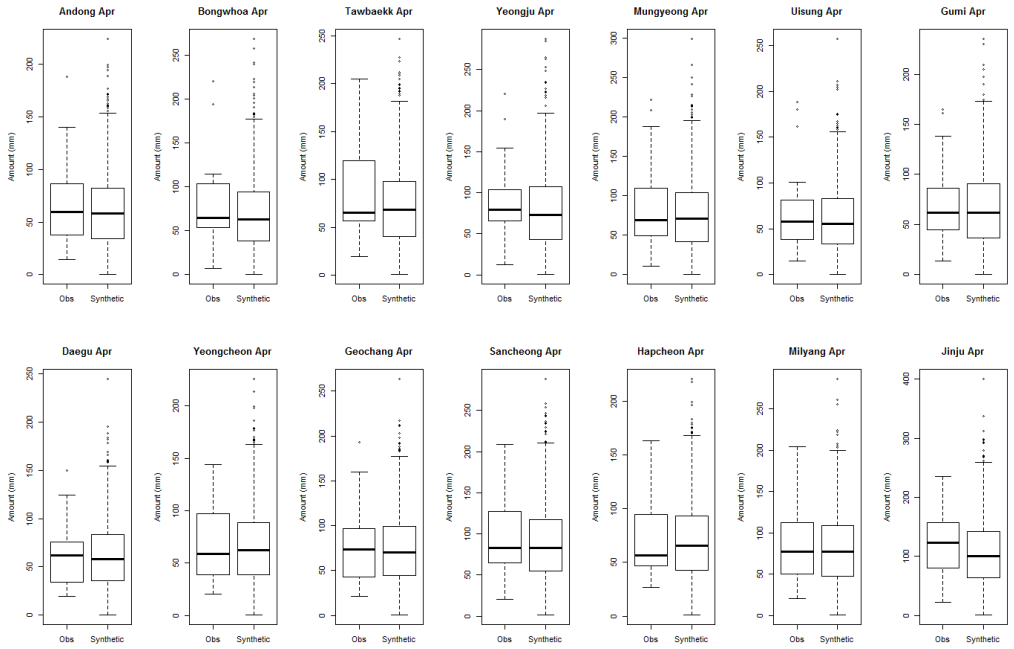


그림 59 4월 지점 별 월 누적 강수량의 분포 비교.

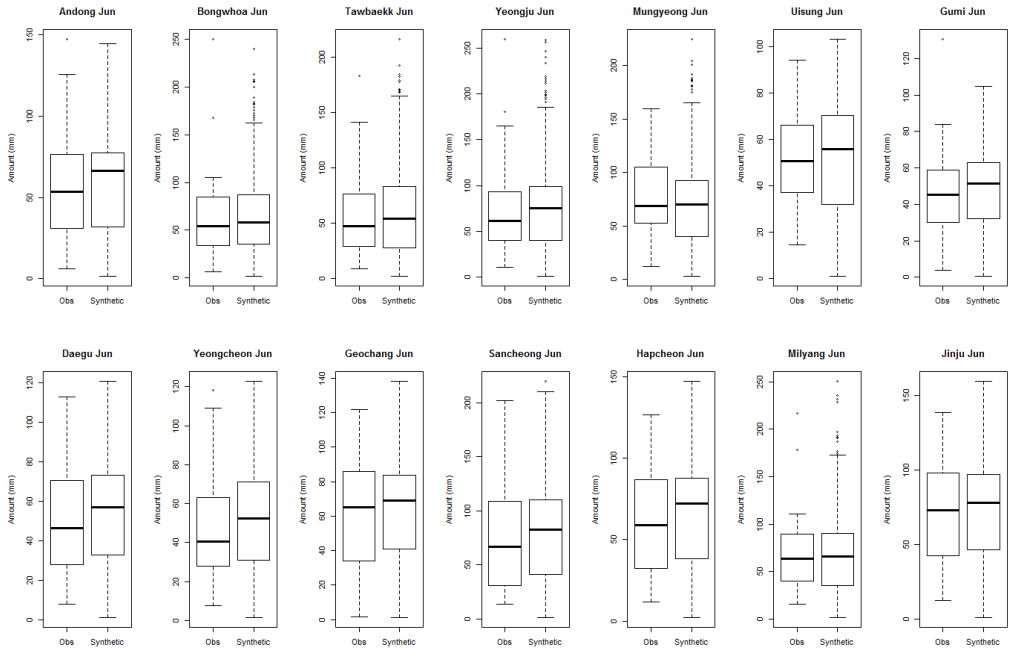


그림 60 6월 지점 별 월 누적 강수량의 분포 비교.

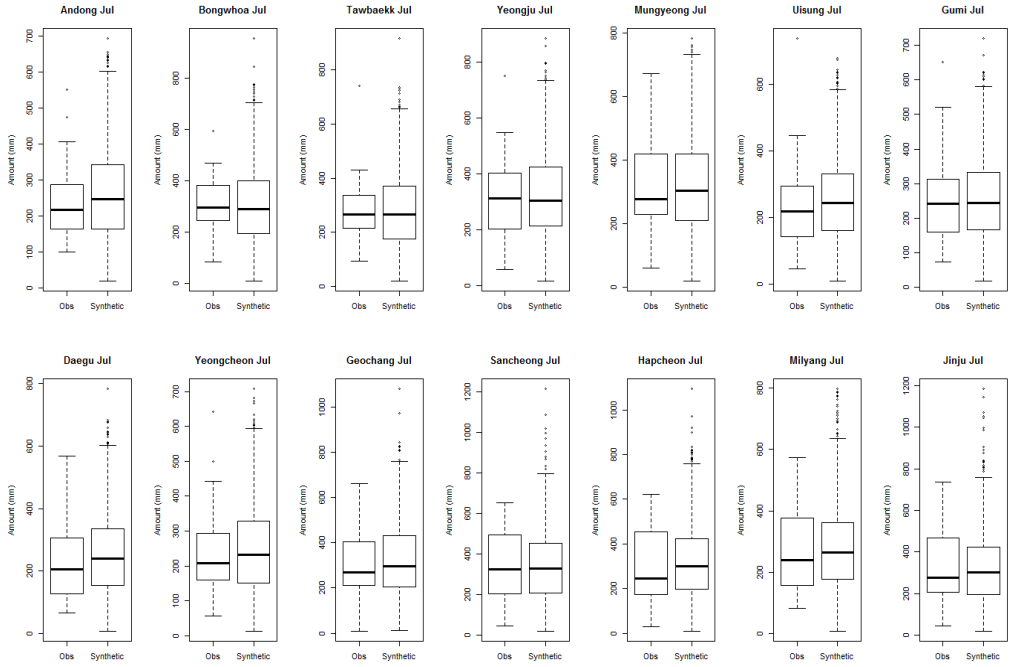


그림 61 7월 지점 별 월 누적 강수량의 분포 비교.

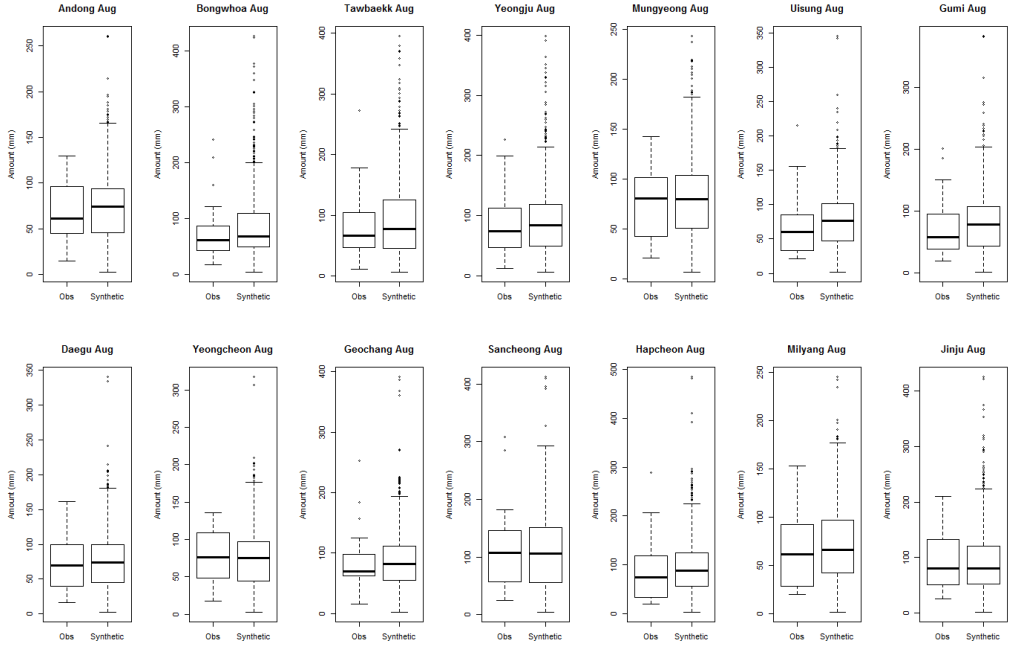


그림 62 8월 지점 별 월 누적 강수량의 분포 비교.

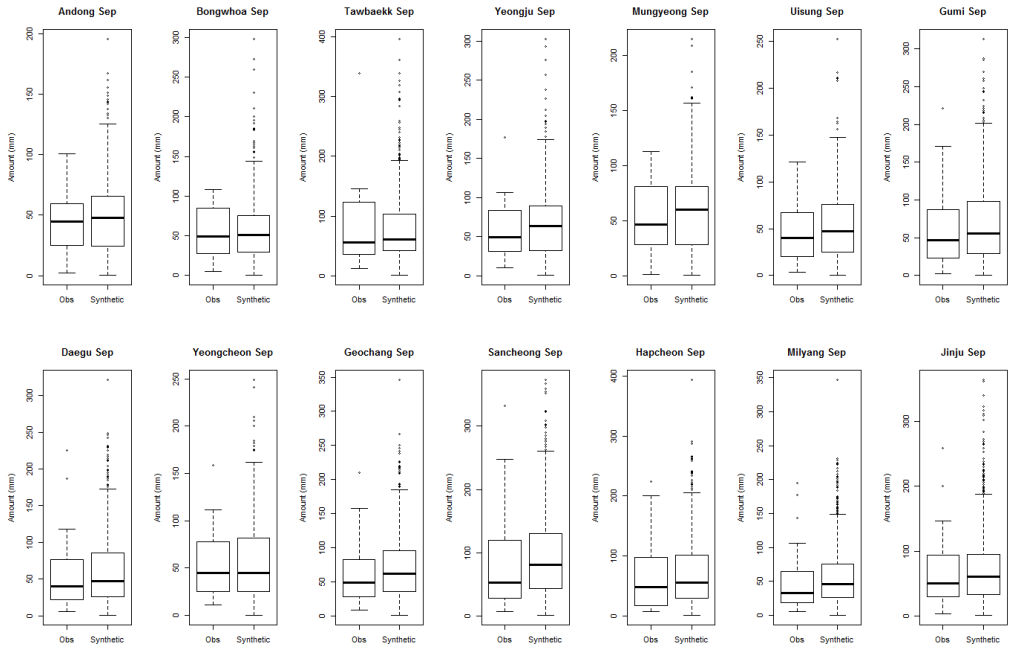


그림 63 9월 지점 별 월 누적 강수량의 분포 비교.

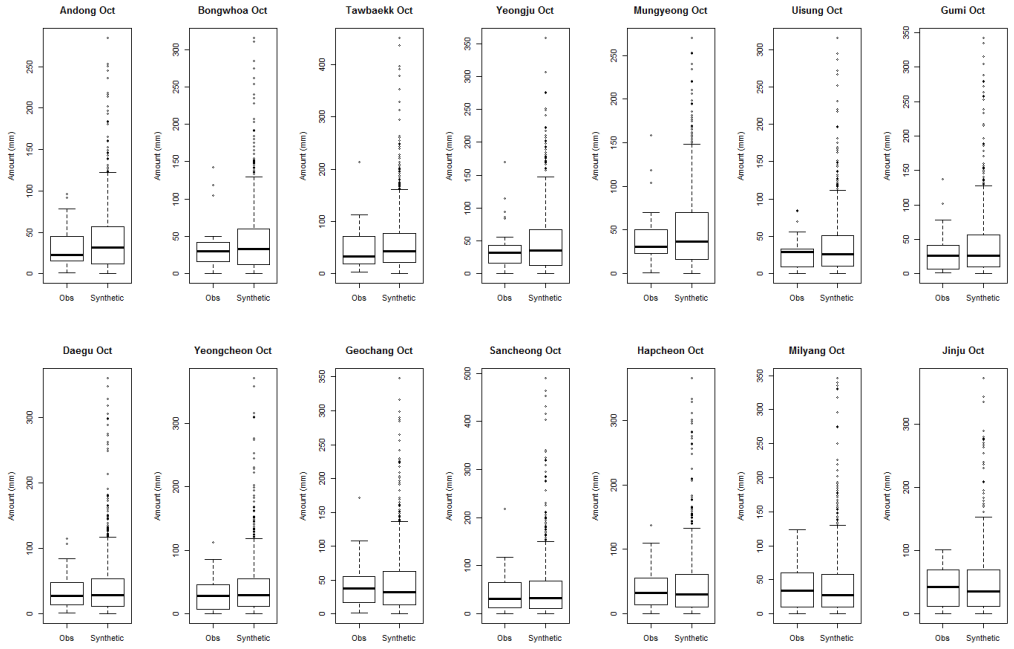


그림 64 10월 지점 별 월 누적 강수량의 분포 비교.

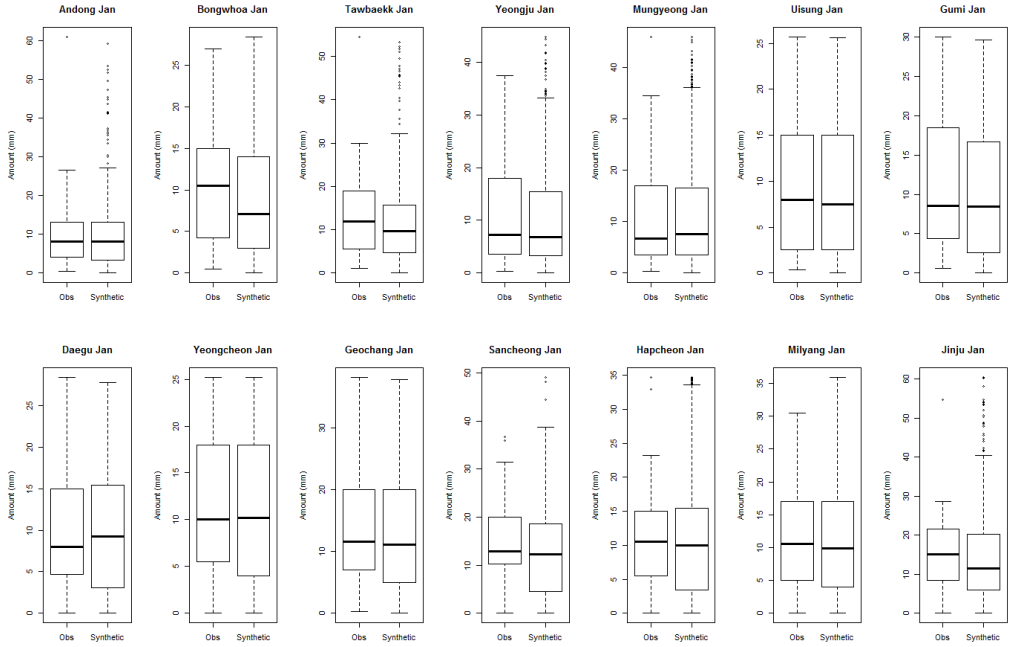


그림 65 1월 지점 별 월 최대일강수량의 분포 비교.

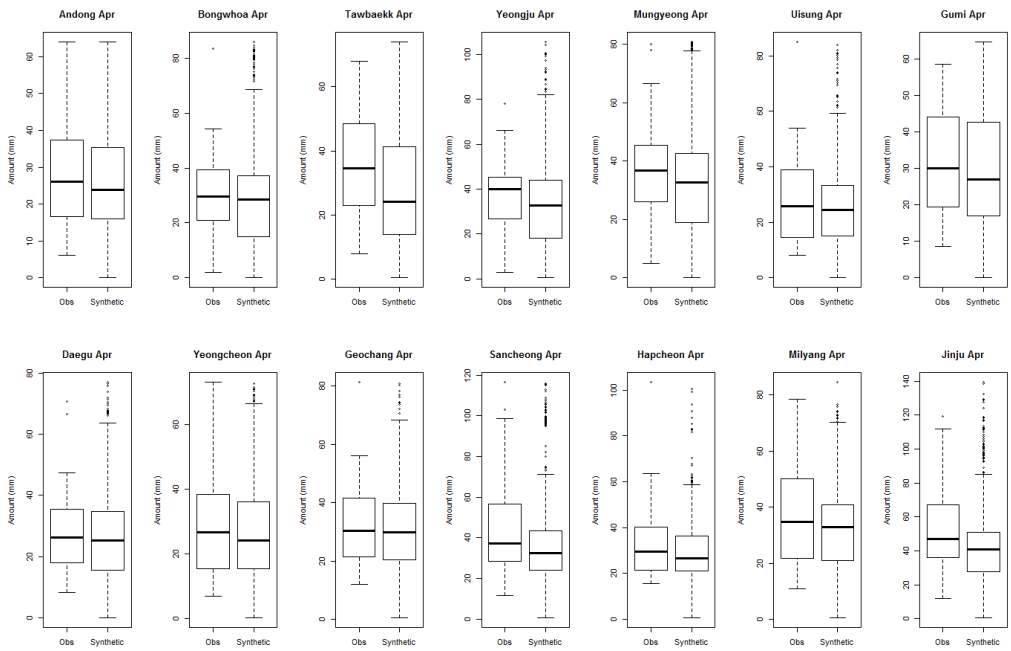


그림 66 4월 지점 별 월 최대일강수량의 분포 비교.

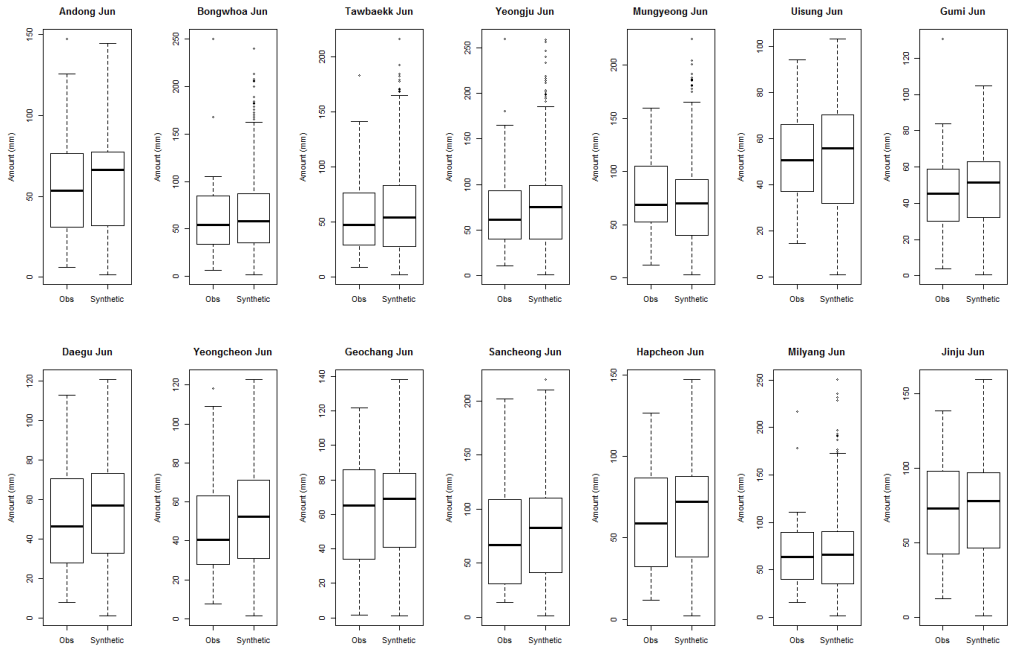


그림 67 6월 지점 별 월 최대일강수량의 분포 비교.

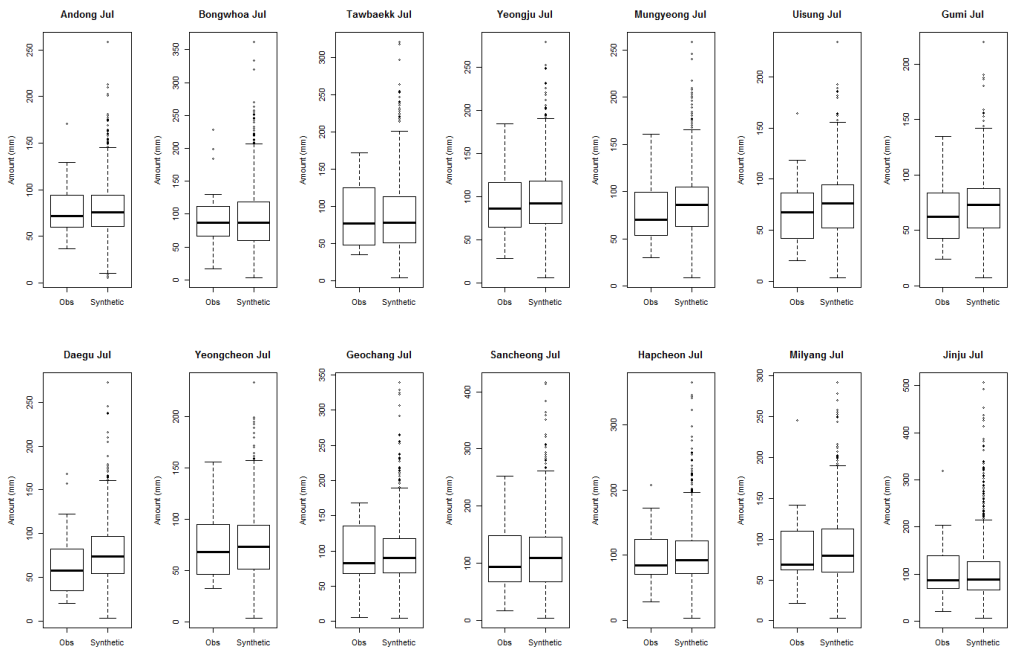


그림 68 7월 지점 별 월 최대일강수량의 분포 비교.

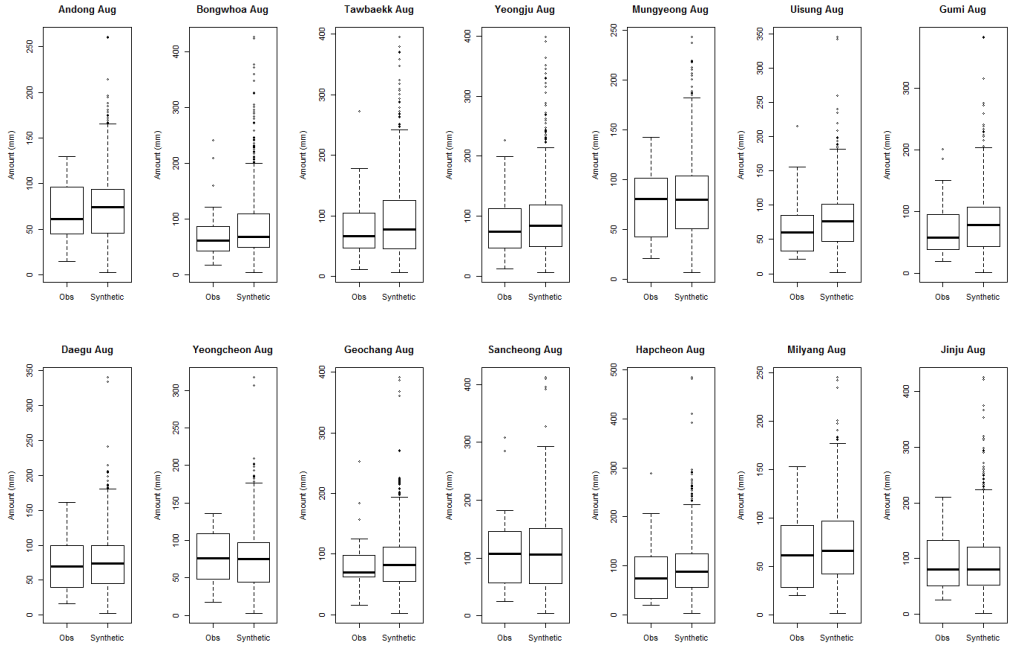


그림 69 8월 지점 별 월 최대일강수량의 분포 비교.

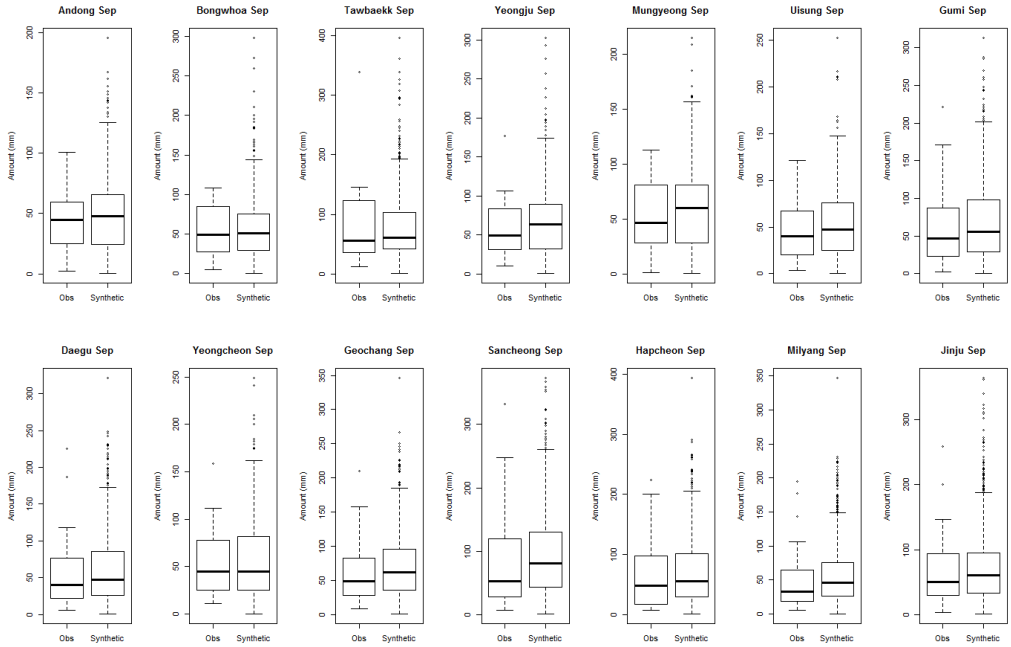


그림 70 9월 지점 별 월 최대일강수량의 분포 비교.

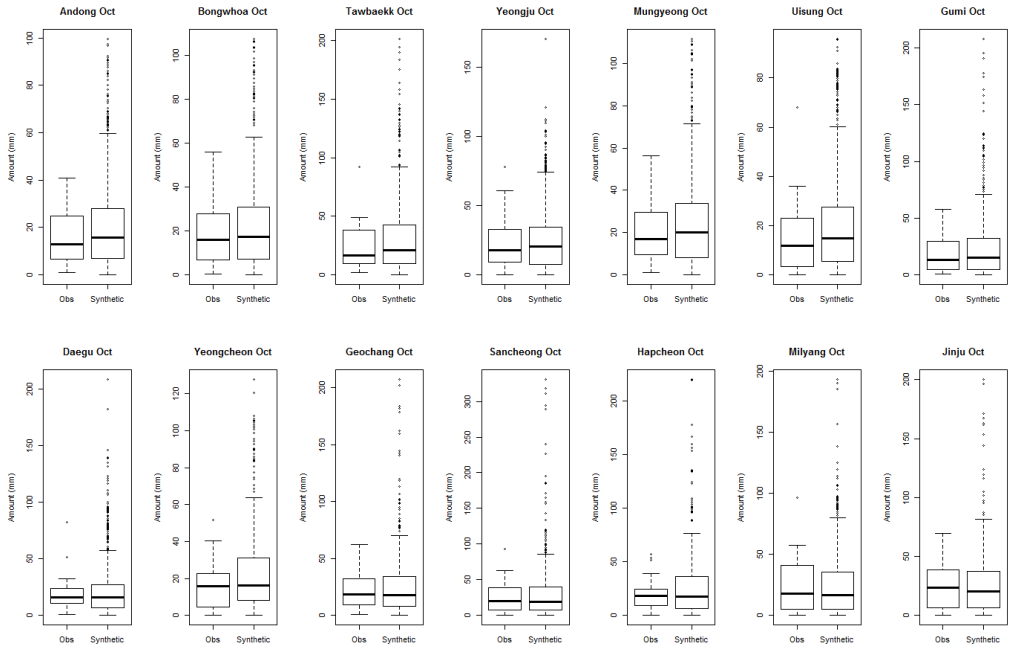


그림 71 10월 지점 별 월 최대일강수량의 분포 비교.

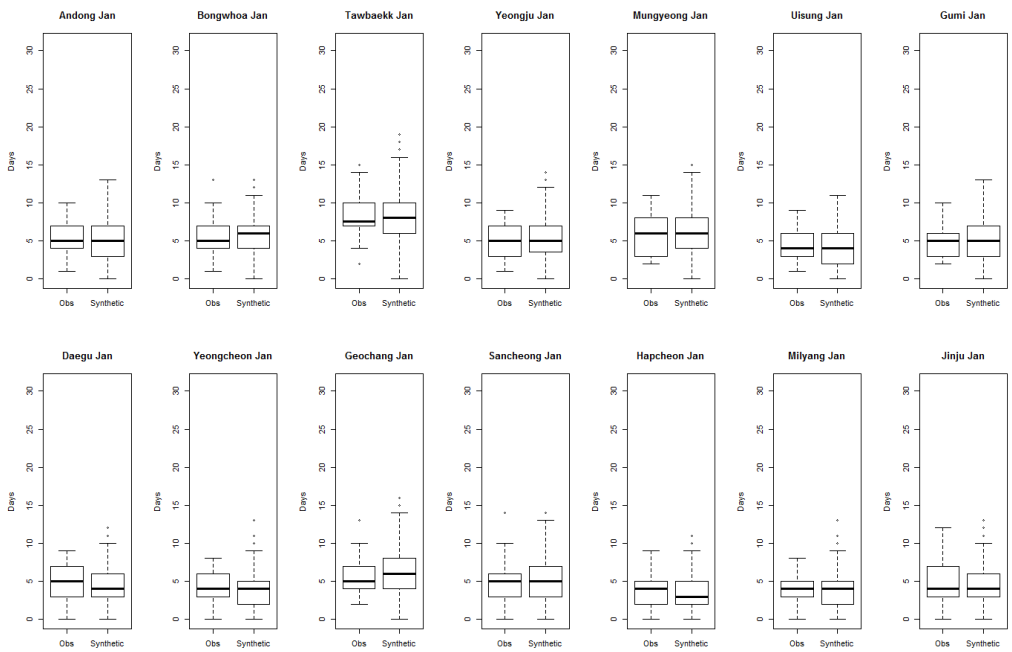


그림 72 1월 지점 별 월 습윤일수의 분포 비교.

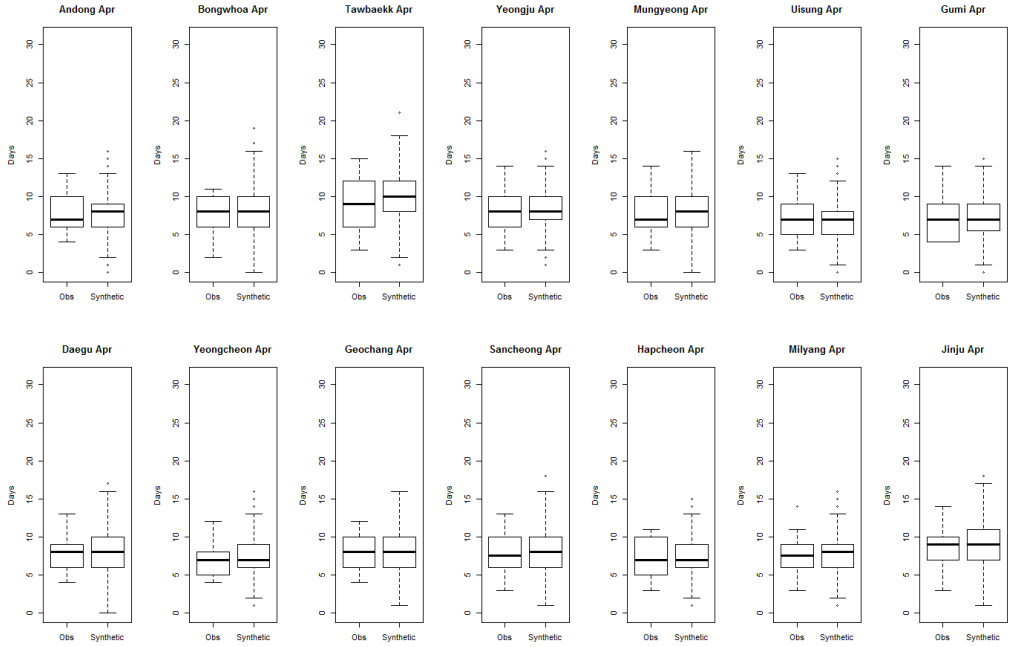


그림 73 4월 지점 별 월 습윤일수의 분포 비교.

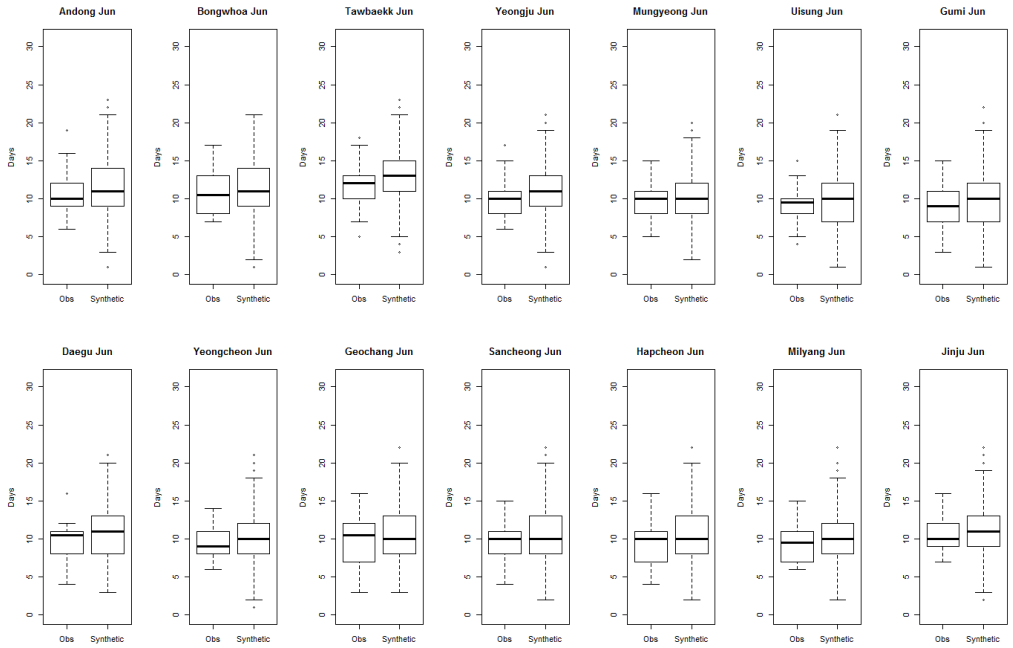


그림 74 6월 지점 별 월 습윤일수의 분포 비교.

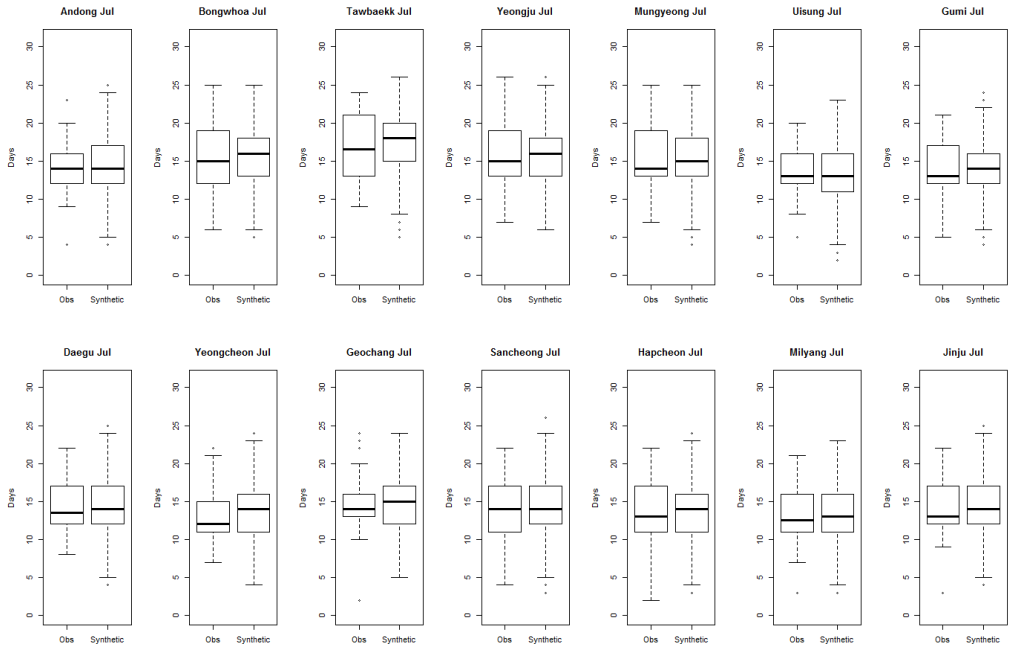


그림 75 7월 지점 별 월 습윤일수의 분포 비교.

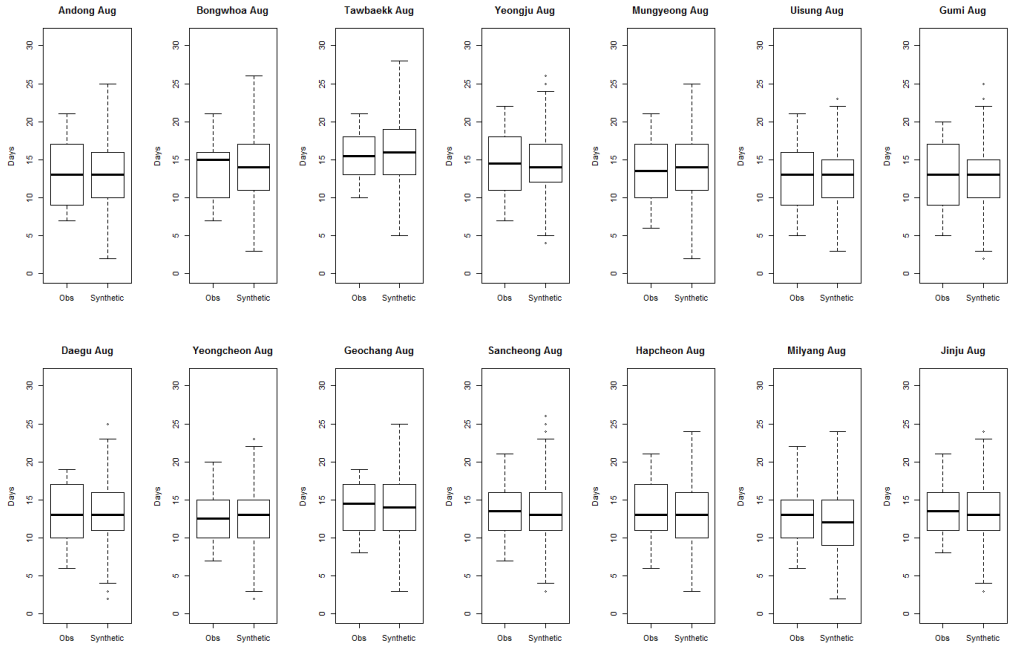


그림 76 8월 지점 별 월 습윤일수의 분포 비교.

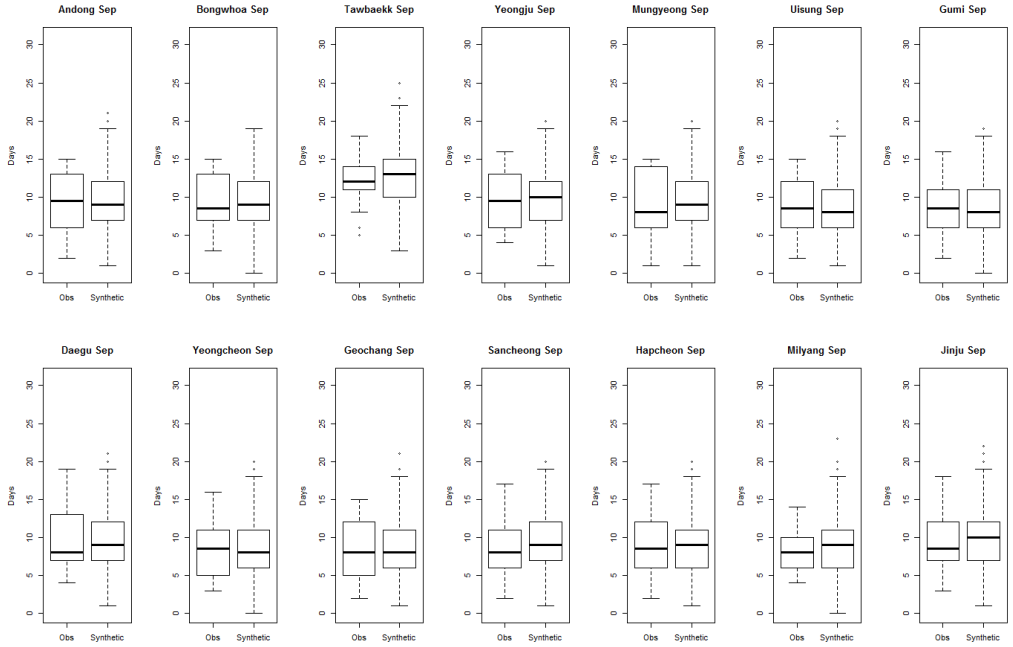


그림 77 9월 지점 별 월 습윤일수의 분포 비교.

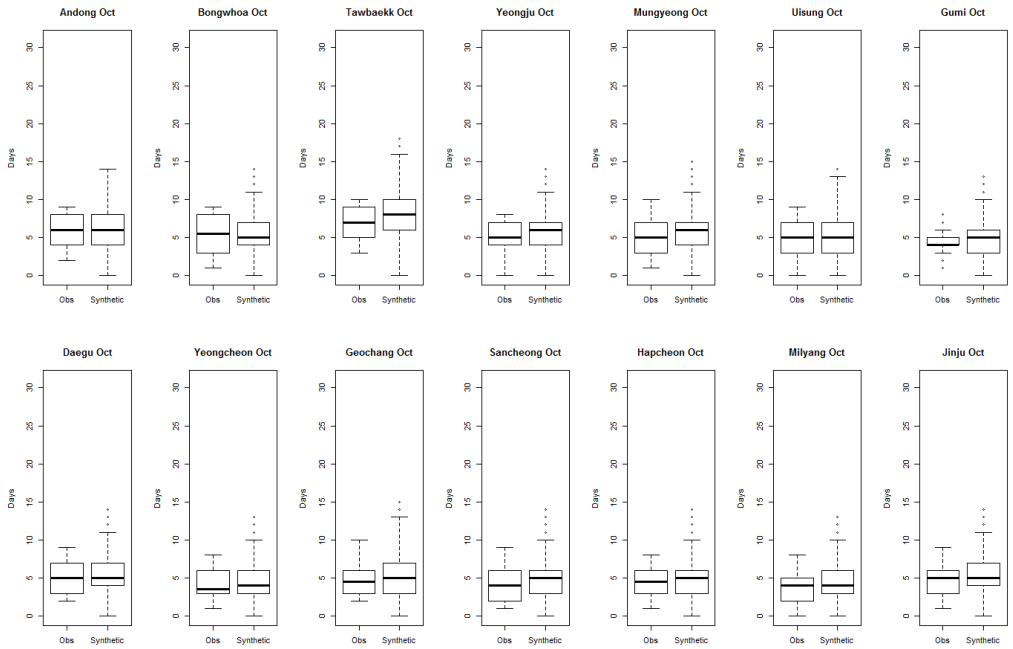


그림 78 10월 지점 별 월 습윤일수의 분포 비교.

3.2.6 기온 월평균 모의성능 검증

이제, 기온의 월단위 특성에 대한 모의성능을 검증한다. 월단위 특성으로 최고/최저기온의 월평균을 살핀다. 그림 79-82은 일최고기온 월평균, 그림 83-86은 일최저기온의 것을 비교한다. 최고기온의 경우에 1월, 4월, 10월에는 대체적으로 서로 유사한 분포형태를 보이나, 7월에 산포도 간 유의한 차이가 나타난다. 한편, 최저기온은 모든 월에서 두 분포가 대체로 유사하다.

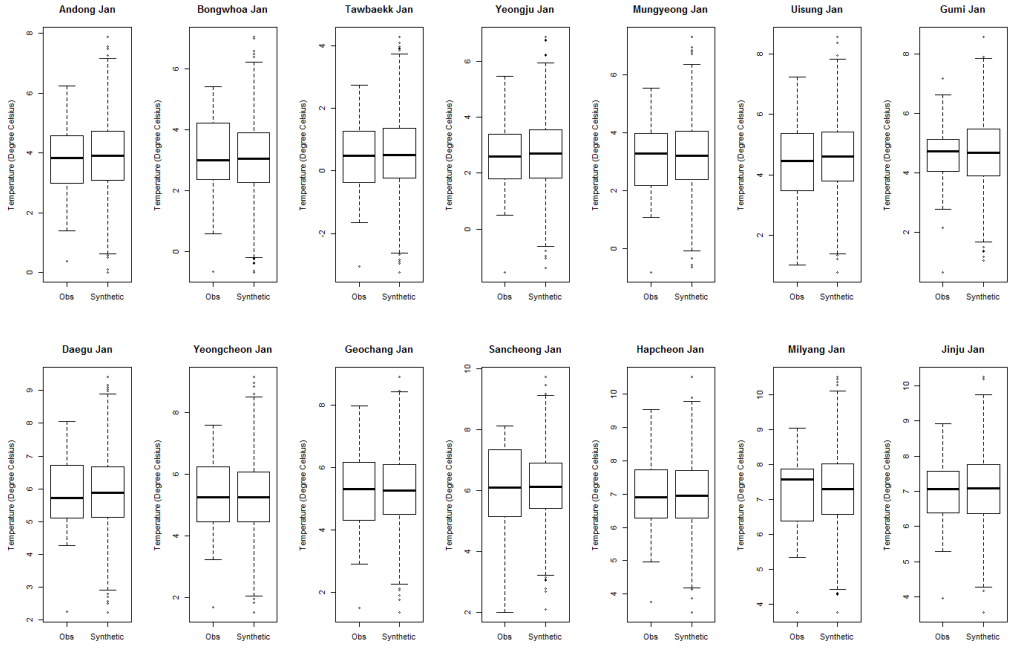


그림 79 1월 지점 별 일최고기온 월평균의 분포 비교.

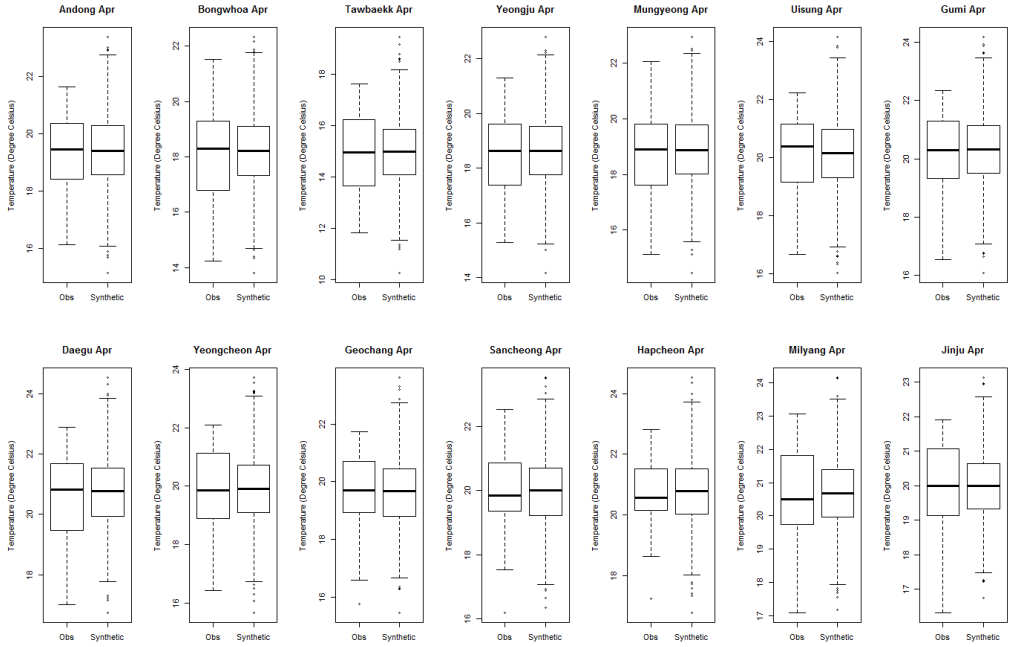


그림 80 4월 지점 별 일최고기온 월평균의 분포 비교.

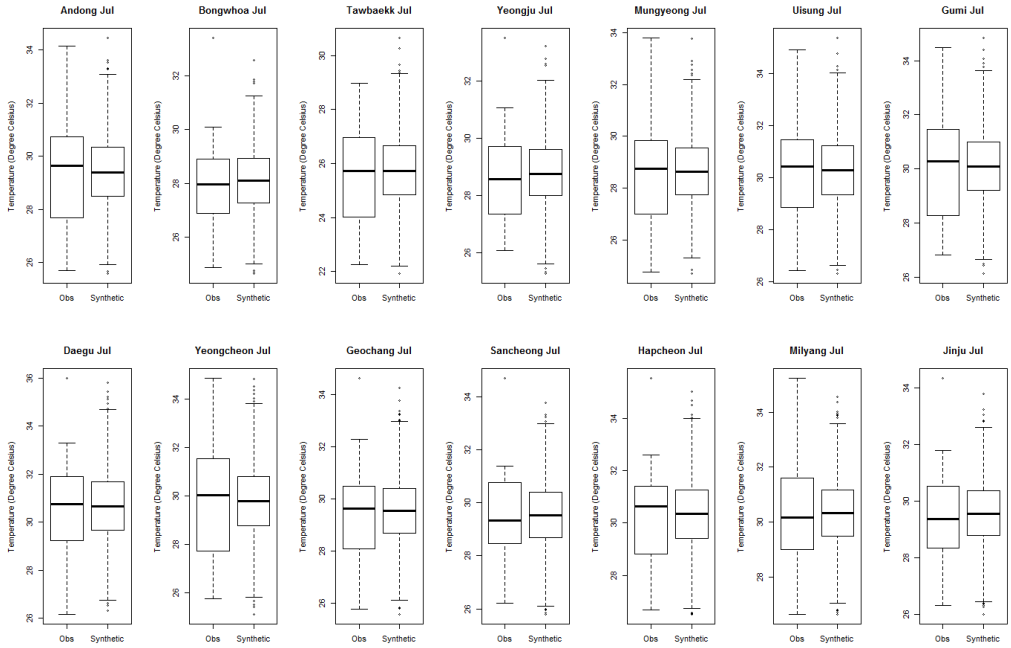


그림 81 7월 지점 별 일최고기온 월평균의 분포 비교.

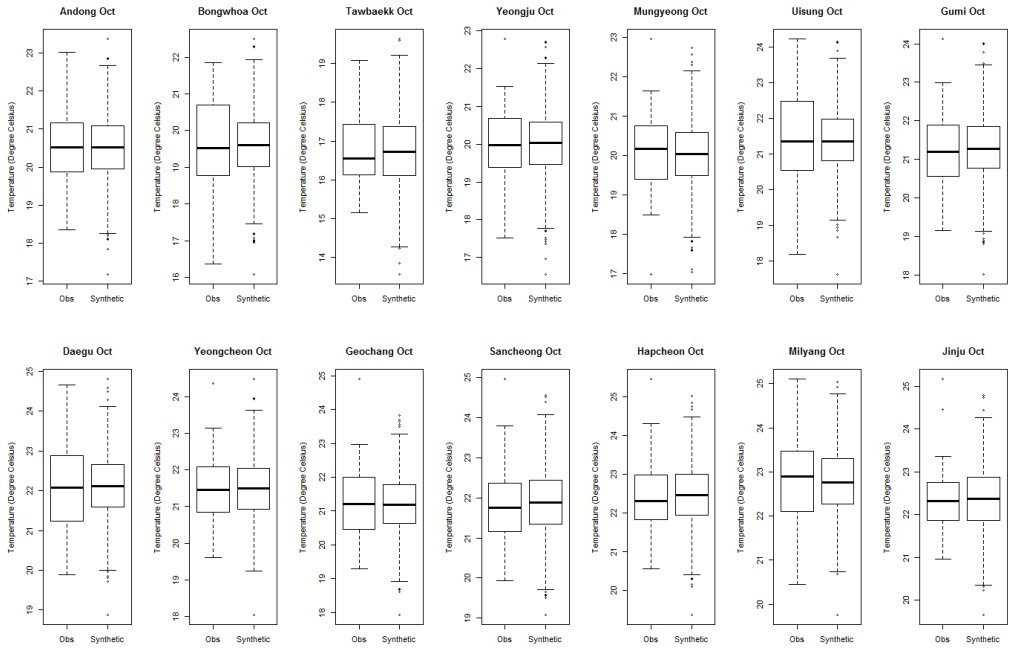


그림 82 10월 지점 별 일최고기온 월평균의 분포 비교.

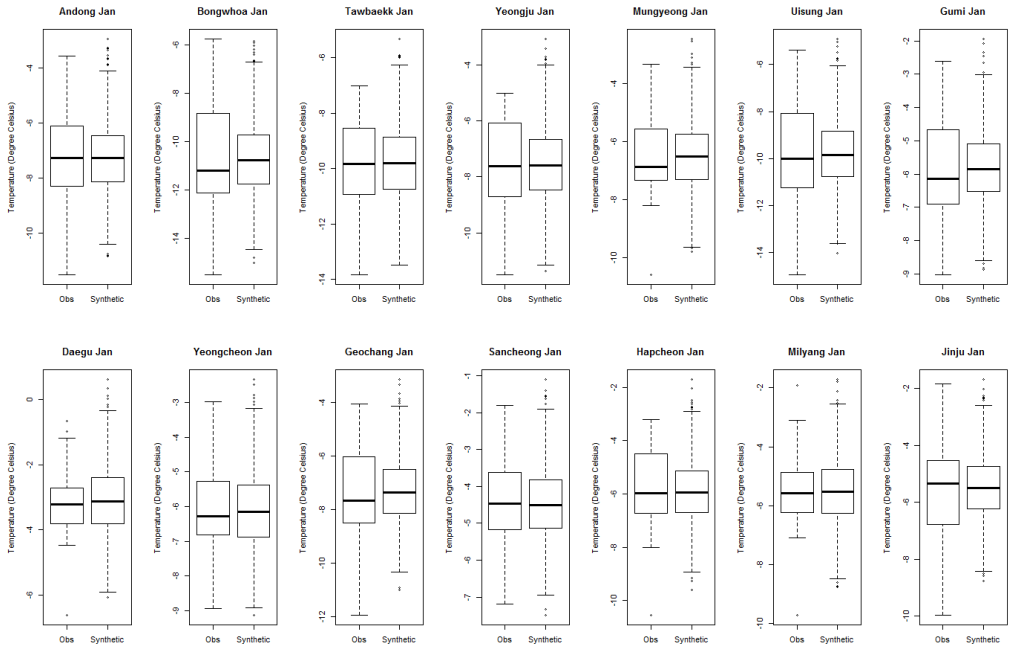


그림 83 1월 지점 별 일최저기온 월평균의 분포 비교.

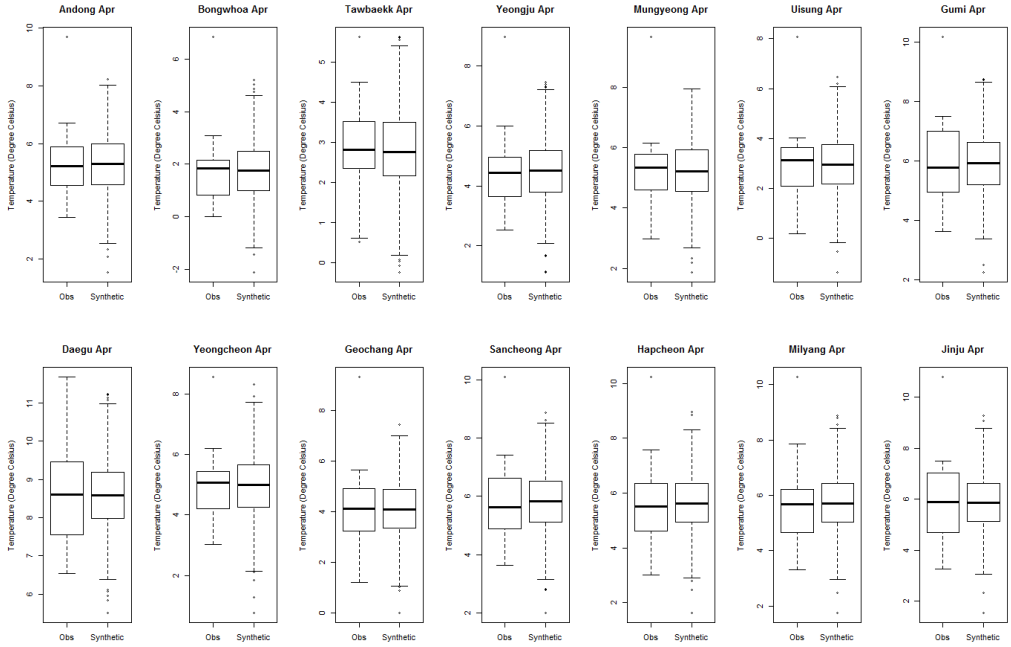


그림 84 4월 지점 별 일최저기온 월평균의 분포 비교.

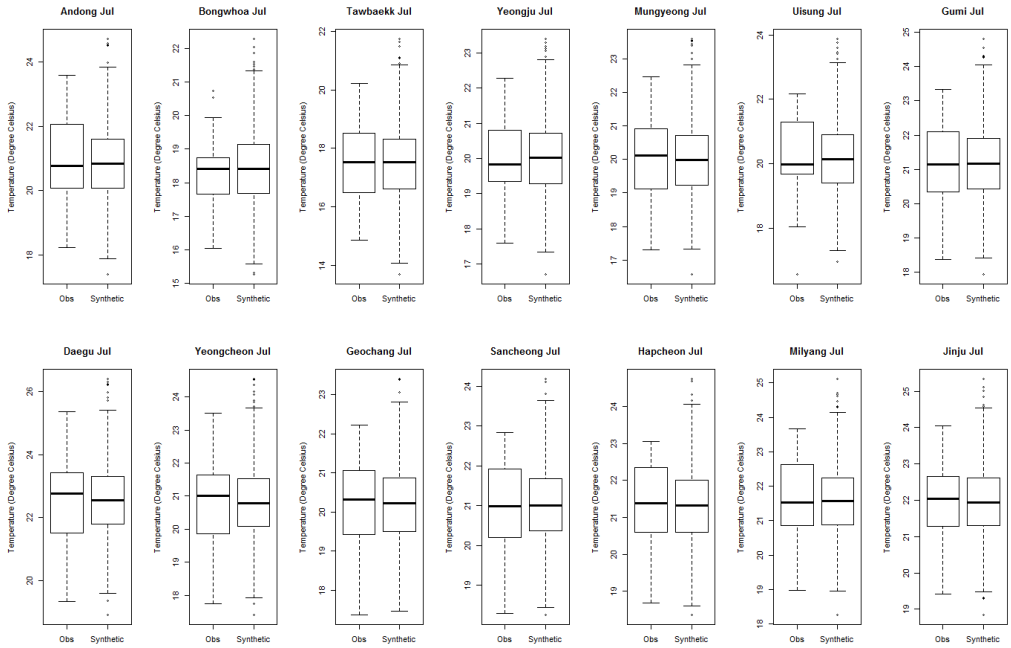


그림 85 7월 지점 별 일최저기온 월평균의 분포 비교.

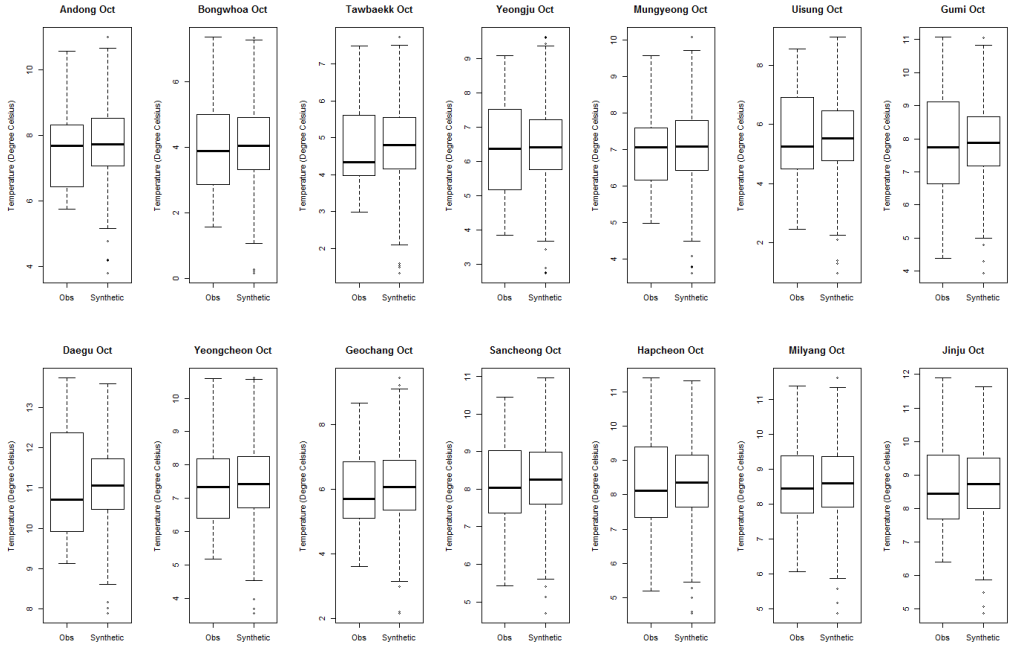


그림 86 10월 지점 별 일최저기온 월평균의 분포 비교.

4. 결론 및 향후 연구 계획

4.1 결 론

본 연구에서 통계적 상세화를 위한 weather generator를 개발하였으며, 그 특징을 다음과 같이 요약할 수 있다:

- **강수모의를 위한 습윤/건조기 교대 방식**: 이 방식은 LARS-WG의 습윤기 생성 방법으로 Markov 모형이 지닐 수 있는 model mis-specification을 피할 수 있다.
- **다지점 강수량 생성을 위한 Copula 모형**: 본 연구에서는, 다지점 강수량을 생성하기 위해 copula 모형을 고려하였다. 이 모형은 특히 강수량 모형에서처럼 다변량 정규분포를 가정하기 힘든 경우에서 공간 상관을 표현하는 좋은 수단이다. 본 연구에서는 copula에 대해 준경험적 모형과 Gaussian 모형을 고려하였다. 두 모형의 비교 결과, Gaussian 모형이 parametric 모형으로서 준경험적 모형과 유사한 성능을 보이므로, 우리는 최종적인 모형으로 정하였다.
- **극한 강수 생성을 위한 극단값 모형**: 극한강수의 재현성을 위해 극단값 모형을 고려한다. 지점모형으로 일반화 Pareto 모형을, 지점 간 모형으로 공간극단값모형을 고려하였다. 지점 강수량 모형으로 많이 쓰이는 감마분포 경우에 극한강수 모의에서 부정확성을 드러낼 수 있는데, 극단값 모형을 통해 그러한 단점을 보완할 수 있다. 또한, 극한강우의 공간적 상관성은 copula 모형으로 추정되는 상관정도와 다를 수 있어, 이에 대한 보완으로 공간극단값모형을 쓴다.
- **강수의 의한 기온 일평균에 대한 회귀모형**: WGEN과 LARS-WG에서는 습윤일과 건조일로 구분해 일평균을 추정함으로써, 기온에 대한 강수의 효과를 나타내었다. 하지만 강수의 효과는 강수량에 따라 연속적이며, 본 연구에서는 연속적 효과를 회귀모형을 통해 표현하였다.
- **기온 anomaly에 대한 주성분-벡터자기회귀모형**: WGEN과 LARS-WG에서는 anomaly의 상관성을 표현하기 위해 벡터자기회귀모형을 고려하였다. 다지점에서는 추가적으로 공간적 상관을 고려해야 하는데, 고려되는 지점의 수가 늘어날

수록 벡터자기회귀모형이 복잡해진다. 이러한 복잡성을 피하기 위해서 EOF 분석을 통해 anomaly의 주성분을 얻고 이에 대해 벡터자기회귀모형을 적합해 시공간적 자기상관을 모의하였다.

- **기후인자와의 연결성**: 기후인자와 weather generator의 연결을 위해 Bardossy (1997)의 방식을 고려하는데, 이 방식은 내재된 통계모형의 적절한 모수화(parametrization)를 요구한다. 내재된 통계모형에서 기후인자와 연결될 성분은 dry-spell 지속일 분포, 지점강수량 분포, 최고/최저기온의 월평균이다. Dry-spell 지속일 분포는 음이항분포로 modeling이 가능하며, 지점강수량 분포는 감마분포 적합을 통해 intensity(mm/day)에 대한 일반화 선형모형(generalized linear model)이 고려된다. 한편, 기온의 월평균와 기후인자의 연결을 위해서는 climatological mean과 강수효과가 제거된 일기온 잔차(residual)의 회귀모형을 고려할 수 있다.

제안된 weather generator의 모의성능을 평가한 결과는 아래와 같다.

- 강수와 기온의 일별분포와 공간상관, 기온의 시간적 자기상관, 최고/최저 기온간 상관은 대체적으로 만족할 만한 수준으로 모의되는 것으로 판단된다. 또한, 강수의 월누적, 월최대량과 습윤일수도 큰 차이없이 모의되는 것으로 보인다.
- 기온 월평균의 모의분포는 관측분포와 비교했을 때 climatological mean은 유사하지만 산포도가 낮게 나온다. 이 차이에 대한 심도 있는 연구가 진행되어야 할 것이다. 현재까지의 분석에 따르면, 기온의 경우 월평균이 통계적으로 유의한 경년변동을 보이고 있지만, weather generator의 통계모형은 이를 무시해서 발생한 차이로 판단된다. 향후 경년변동성이 반영된 기온 통계모형 개발하여 월평균 모의 성능 개선을 시도할 계획이다.

4.2 향후 연구 계획

통상적으로 weather generator는 관측기간의 climatology를 기반으로 기상시나리오를 생성한다. 그래서 기후의 경년변동이 반영된 기상시나리오 생성을 위해서는 기후인자와 weather generator를 연결짓는 장치가 필요하다. 실제, 기상관측자료를 분석하면 유역 기후의 경년변동이 드러내는데, 향후에 이를 설명할 수 있는 상세화 모형을 연구할 계획이다.

실제 응용분야 연구에서는 강수와 기온 외에도 일사량, 상대습도와 같은 변수도 중요하게 다루어진다. 그래서, weather generator의 활용성을 높이기 위해 다양한 기상변수를 모의하도록 확장하는 것이 필요하다. 또한 weather generator의 기상시나라오의 공간해상도는 관측소(station)의 공간밀도로 결정되는데, 더 높은 고해상도를 얻기 위해서는 공간내삽(spatial interpolation)을 실시해야 한다. 향후 weather generator의 기상시나리오의 공간내삽법을 개발하여 응용분야에서 필요로 하는 고해상도 격자자료를 생산하고자 한다.

참고 문헌

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.

Bardossy, A. 1997: Downscaling from GCM to local climate through stochastic linkages. *J. Environ. Manage*. **49** 7-17.

Buishand, T. A. and Brandsma, T. 2001: Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest neighbor resampling *Water Resourc*. Res. **37**(11) 2761-2776

Cleveland, W. S., Grosse, E., & Shyu, W. M. 1992: Local regression models. Statistical models in S, 309-376.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P., 2001: *An introduction to statistical modeling of extreme values (Vol. 208)*. London: Springer.

Davison, A. C., & Gholamrezaee, M. M. 2012: Geostatistics of extremes. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. **468** (2138), 581-608.

Eum, H., Simonovic, S. P., and Y. Kim 2010: Cimate change impact assessment using K-nearest neighbor weather generator: case study of the Nakdong river basin in Korea. *Journal of Hydrologic Engineer*. **15**(10) 772-785.

Gangopadhyay,S, Clark, M., and Rajagopalan 2005: Statistical downscaling using K-nearest neighbors *Water Resour*. Res. **41** W02024 1-24.

Guyennon, N., Romano, E., Portoghese, I., Salerno, F., Calmanti, S., Petrangeli, A. B., Tartari, G. and Copetti, D. 2013: Benefits from using combined dynamical-statistical downscaling approaches-lessons from a case study in the Mediterranean region. *Hydrol. Earth Syst. Sci*. **17**, 705-720.

Hwang, S., & Graham, W. D. 2013: Development and comparative evaluation of a stochastic analog method to downscale daily GCM precipitation. *Hydrology and Earth System Sciences*, **17**(11), 4481-4502.

Helmut, L. 1993: *Introduction to multiple time series analysis*. Springer-Verlag. 545 pp.

Jones, P. D., Murphy, J. M. and Noguer, M. 1995: Simulation of climate change over Europe using a nested regional-climate model. 1: Assessment of control climate, including sensitivity to location of lateral boundaries. *Q. J. R. Meteorol. Soc*. **121** 1413-1449.

Lall, U. and Sharma, A. 1996: A nearest neighbor bootstrap for time series resampling *Water Resour. Res.* **32**(3) 679-693.

Lall, U., Rajagopalan, B. and Torboton, D. G. 1996: A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resour.* Res. **32**(9) 2803-2823.

Mehrotra, R. and Sharma, A 2006a: Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbor approach *Adv. Water Resour.* **29** 987-999

Mehrotra, R. and Sharma, A. 2006b: A nonparametric stochastic downscaling framework for daily rainfall at multiple locations *J. Geophys. Res., [Atmos.]*, **111**, D15101

Racsko, P., Szeidl, L., and Semenov, M. 1991: A serial approach to local stochastic weather models. *Ecological modelling*, **57**(1), 27-41.

Rajagopalan, B. and Lall, U. 1999: A k-nearest neighbor simulator for daily precipitation and other variables. *Water Resour. Res.* **35**(10) 3089-3101

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.

Semenov, M. A., and Barrow, E. M. 1997: Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change*, **35** 397-414.

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.

Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris 8,* 229-231

Smith, R. L. 1990: Max-stable processes and spatial extremes. Unpublished manuscript, University of Surrey, Guildford GU2 5XH, England.

Vrac, M. and Naveau, P. 2007: Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resour. Res.* **43** W07402.

Wilby, R. L. , Hassan, H. and Hanaki, K. 1998: Statistical downscaling of hydrometeorological variables using general circulation model output *J. Hydro.l* **205** 1-19.

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Stzepek, K. 2003: A technique for generating regional climate scenario using a nearest-neighbor algorithm. *Water*

Resour. Res. **39**(7), SWC 7-1-SWC 7-14

Young, K. C. 1994: A multivariate chain model for simulating climatic parameters from daily data. *J. Appl. Meteorol.*, **33** 661-671.

부록 – 모형 추정법

본 절에서는 2.1.2.–2.1.2.절에서 언급되었던 모형추정법을 자세하게 소개한다. 이 절에서는

$$(\mathrm{R}(s_1,t_i),\ \cdots\ ,\mathrm{R}(s_d,t_i)),\quad i=1,\cdots,n,$$

가 습윤일 다지점 강수량의 관측자료를 뜻한다.

Gaussian copula 모형 추정

모형 (11) 하에서,

$$\mathrm{R}(s_j,t_i)=F_j^{\leftarrow}(\Phi(\mathrm{Z}_{i,j})|\mathrm{w}),\quad i=1,\cdots,n,\ j=1,\cdots,d$$

를 만족하는 독립확률벡터 $(\mathrm{Z}_{i,1},\cdots,\mathrm{Z}_{i,d})\sim \mathrm{N}(\mathbf{0},\Sigma)$, $i=1,\cdots,n$가 존재한다. 하지만, $F_j(x|\mathrm{w})$가 $x=0$에서 불연속이기 때문에, $(\mathrm{Z}_{i,1},\cdots,\mathrm{Z}_{i,d})$가 완전히 관측되지 않는다:

$$\hat{\mathrm{Z}}_{\mathrm{i,j}}:=\Phi^{-1}(\mathrm{F_j}(\mathrm{R}(s_j,t_i)|\mathrm{w})),\quad \mathrm{l}(j):=\Phi^{-1}(F_j(0|\mathrm{w}))>-\infty,$$
$$i=1,\cdots,n,\ j=1,\cdots,d$$

이라고 했을 때, 항상 $\hat{\mathrm{Z}}_{i,j}\geq \mathrm{l}(j)$이며,

- $\hat{\mathrm{Z}}_{i,j}>\mathrm{l}(j)$이면, $\mathrm{Z}_{i,j}=\hat{\mathrm{Z}}_{i,j}$;
- $\hat{\mathrm{Z}}_{i,j}=\mathrm{l}(j)$이면, $\mathrm{Z}_{i,j}\leq \mathrm{l}(j)$, 즉, 정확한 $\mathrm{Z}_{i,j}$의 값을 알 수 없다.

Σ를 추정하기 위해서, 우리는 위의 불완전한 관측상황이 반영된 최대 쌍별우도추정법(maximum pair-wise likelihood estimation)을 고려한다. $\sigma(j_1,j_2)$가 Σ의 (j_1,j_2)번째 원소를 나타낸다고 하자. 단, $j_1,\ j_2\in\{1,\cdots,d\}$이고 $j_1\neq j_2$이다. 목적함수

$$\ell(\sigma)=\ell_1(\sigma)+\ell_2(\sigma)+\ell_3(\sigma)+\ell_4(\sigma),$$

를 최대로 하는 σ값을 $\sigma(j_1,j_2)$의 추정값으로 한다. 여기서,

$$\ell_1(\sigma) = \sum_{i=1}^{n} \log\left\{ \frac{\partial^2}{\partial z_1 \partial z_2} \Phi(z_1, z_2; \sigma) \Big|_{z_1 = \hat{Z}_{i,j_1}, z_2 = \hat{Z}_{i,j_2}} \right\} \mathrm{I}\left(\hat{Z}_{i,j_1} > \mathrm{l}(j_1), \hat{Z}_{i,j_2} > \mathrm{l}(j_2)\right),$$

$$\ell_2(\sigma) = \sum_{i=1}^{n} \log\left\{ \frac{\partial}{\partial z_1} \Phi(z_1, \mathrm{l}(j_2); \sigma) \Big|_{z_1 = \hat{Z}_{i,j_1}} \right\} \mathrm{I}\left(\hat{Z}_{i,j_1} > \mathrm{l}(j_1), \hat{Z}_{i,j_2} = \mathrm{l}(j_2)\right),$$

$$\ell_3(\sigma) = \sum_{i=1}^{n} \log\left\{ \frac{\partial}{\partial z_2} \Phi(\mathrm{l}(j_1), z_2; \sigma) \Big|_{z_2 = \hat{Z}_{i,j_2}} \right\} \mathrm{I}\left(\hat{Z}_{i,j_1} = \mathrm{l}(j_1), \hat{Z}_{i,j_2} > \mathrm{l}(j_2)\right),$$

$$\ell_4(\sigma) = \sum_{i=1}^{n} \log\left\{ \Phi(\mathrm{l}(j_1), \mathrm{l}(j_2); \sigma) \right\} \mathrm{I}\left(\hat{Z}_{i,j_1} = \mathrm{l}(j_1), \hat{Z}_{i,j_2} = \mathrm{l}(j_2)\right),$$

$$\Phi(z_1, z_2; \sigma) = \frac{1}{2\pi\sqrt{1-\sigma^2}} \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \exp\left\{ -\frac{1}{2(1-\sigma^2)} (u_1^2 - 2\sigma u_1 u_2 + u_2^2) \right\} du_1 du_2,$$

이다. 그림 87-88은 낙동강 유역에 대해 각 월에서 $\sigma(j_1, j_2)$의 추정치를 지점 s_{j_1}와 s_{j_2} 간 거리에 따라 나타내었다. 우리의 직관과 일치하게 지점 간 거리가 멀수록 상관 정도가 작아진다.

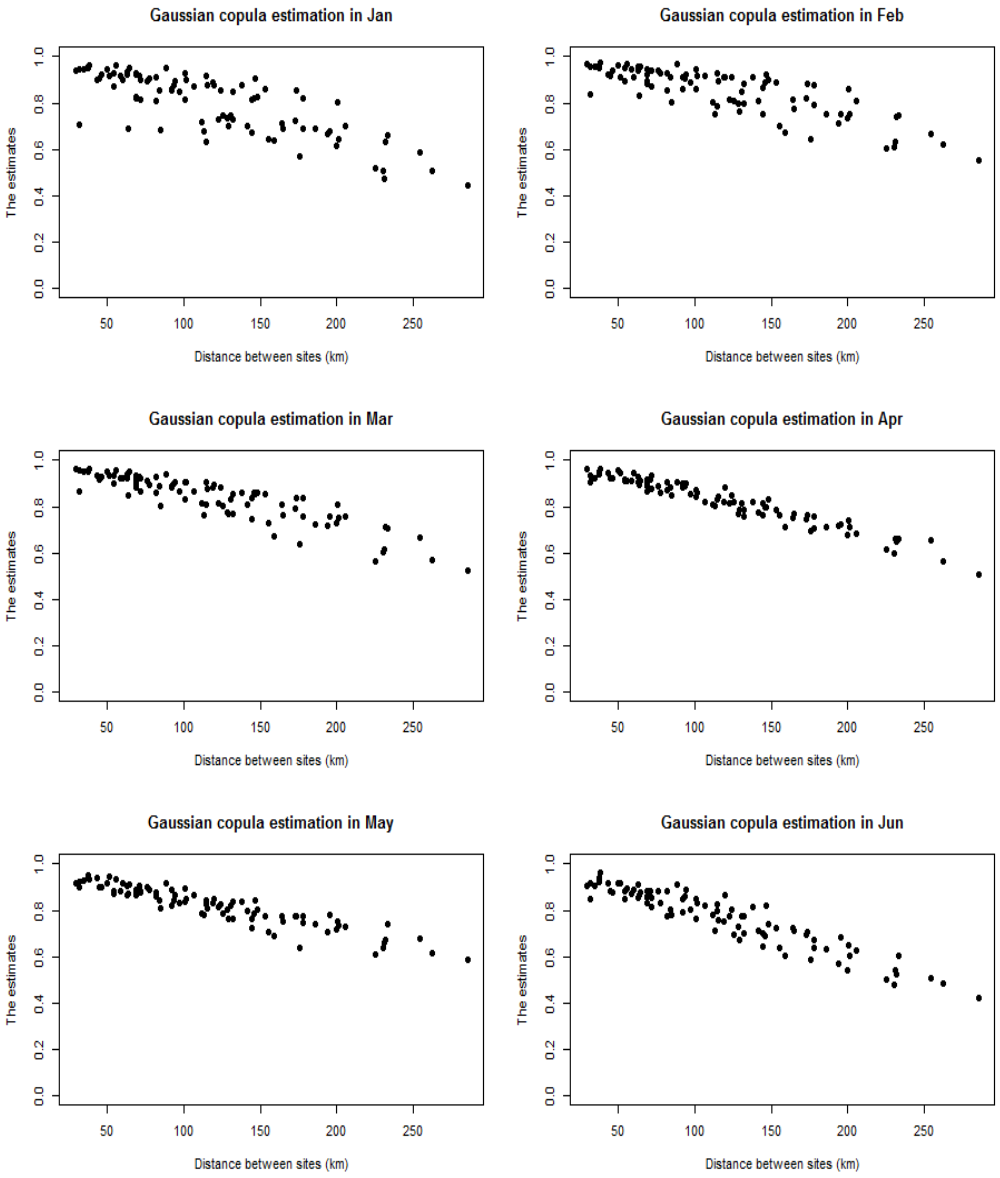


그림 87 1월-6월에 대한 Gaussian copula 추정 결과. 지점 s_{j_1}과 s_{j_2} 사이의 거리(km)에 따른 $\sigma(j_1,j_2)$의 추정치를 나타낸다. 대체적으로 지점간 거리가 멀수록 상관관계가 작아진다.

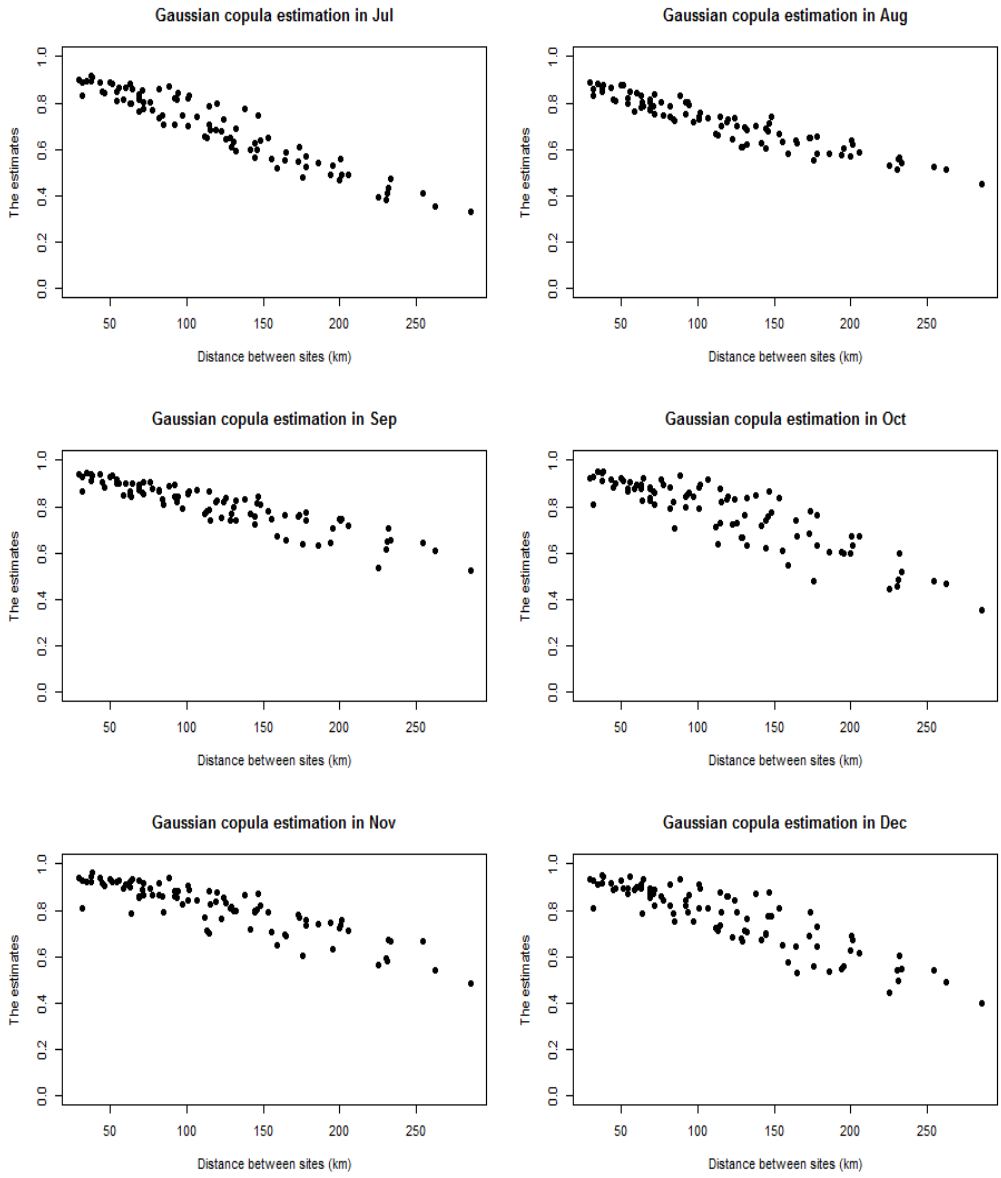


그림 88 7월-12월에 대한 Gaussian copula 추정 결과. 지점 s_{j_1}과 s_{j_2}사이의 거리(km)에 따른 $\sigma(j_1,j_2)$의 추정치를 나타낸다. 대체적으로 지점간 거리가 멀수록 상관관계가 작아진다.

Gaussian 공간 극단값 모형 추정

본 소절에서는 연최대강수량을 대신해서 강수량 peak를 이용해 Gaussian 공간극단값모형을 추정하는 방법을 제시한다. 이를 통해 더 많은 관측자료를 사용할 수 있기 에 효율성을 기대할 수 있다. 더 나아가 제시하고자 하는 방법은 지점의 강수량분포추정 단계를 거치지 않는다.

추정법을 제시하기 앞서, 관련된 tail coefficient를 소개한다. $j_1 \neq j_2$인 $j_1, j_2 \in \{1,\cdots,d\}$에 대해서,

$$\begin{aligned} \mathrm{TDC}(j_1,j_2) &:= \lim_{\mathrm{b}\to 1}(1-\mathrm{b})^{-1}\Pr\left[F_{j_1}\left(\mathrm{R}\left(s_{j_1}\right)\middle|\mathrm{w}\right)>\mathrm{b},\, F_{j_2}\left(\mathrm{R}\left(s_{j_2}\right)\middle|\mathrm{w}\right)>\mathrm{b}\middle|\text{wet day}\right], \\ \mathrm{EC}(j_1,j_2) &:= 2-\mathrm{TDC}(j_1,j_2), \end{aligned}$$

라고 정의하며, 각각 tail dependence coefficient, extremal coefficient라고 부른다. 이 두 계수를 b을 적절히 1에 가깝게 선택하여,

$$\begin{aligned} \widehat{\mathrm{TDC}}(j_1,j_2) &:= \frac{1}{n(1-\mathrm{b})}\sum_{i=1}^{n}\mathrm{I}\left[F_{j_1}\left(\mathrm{R}\left(s_{j_1},t_i\right)\middle|\mathrm{w}\right)>\mathrm{b},\, F_{j_2}\left(\mathrm{R}\left(s_{j_2},t_i\right)\middle|\mathrm{w}\right)>\mathrm{b}\right], \\ \widehat{\mathrm{EC}}(j_1,j_2) &:= 2-\widehat{\mathrm{TDC}}(j_1,j_2), \end{aligned}$$

으로 추정한다. 한편, 모형 (16) 하에서,

$$\mathrm{EC}(j_1,j_2,\Sigma) = 2\Phi\left[\frac{1}{2}\left\{(s_{j_1}-s_{j_2})'\Sigma(s_{j_1}-s_{j_2})\right\}^{1/2}\right]$$

이다. 그래서, Σ의 추정값 $\hat{\Sigma}$을 목적함수

$$\ell(\Sigma) = \sum_{j_1=1}^{d-1}\sum_{j_2=j_1+1}^{d}\mathrm{w}(j_1,j_2)\left\{\widehat{\mathrm{EC}}(j_1,j_2)-\mathrm{EC}(j_1,j_2,\Sigma)\right\}^2$$

를 최소화하여 구한다. (Smith (1990) 비교.) 본 보고서에서는 $\mathrm{b}=0.95$, $\mathrm{w}(j_1,j_2)=1/\widehat{\mathrm{TDC}}(j_1,j_2)$으로 정하였다.

그림 89은 우리나라 호우기(6월–9월) 낙동강 유역의 강수에 대해 Gaussian 공간극단값 모형을 적합한 결과이다. 그림에서 추정된 extremal coefficient $\mathrm{EC}(j_1,j_2,\hat{\Sigma})$를 지점 s_{j_1}와 s_{j_2}간 거리에 따라 표시하였다. Extremal coefficient는 항상 그 값이

1과 2사이에 있으며, 1에 가까울수록 두 지점의 극한강우가 상관성이 큰 것이고 반대로 2에 가까울수록 독립적이다. 우리의 직관과 일치하게 대체로 지점 간 거리가 멀수록 상관성이 떨어진다. 그런데 그 정도는 월별로 다른데, 7-8월에 극한강우의 공간 상관성이 약해진다. 즉, 국지성 호우의 발생가능성이 큰 것으로 해석할 수 있다.

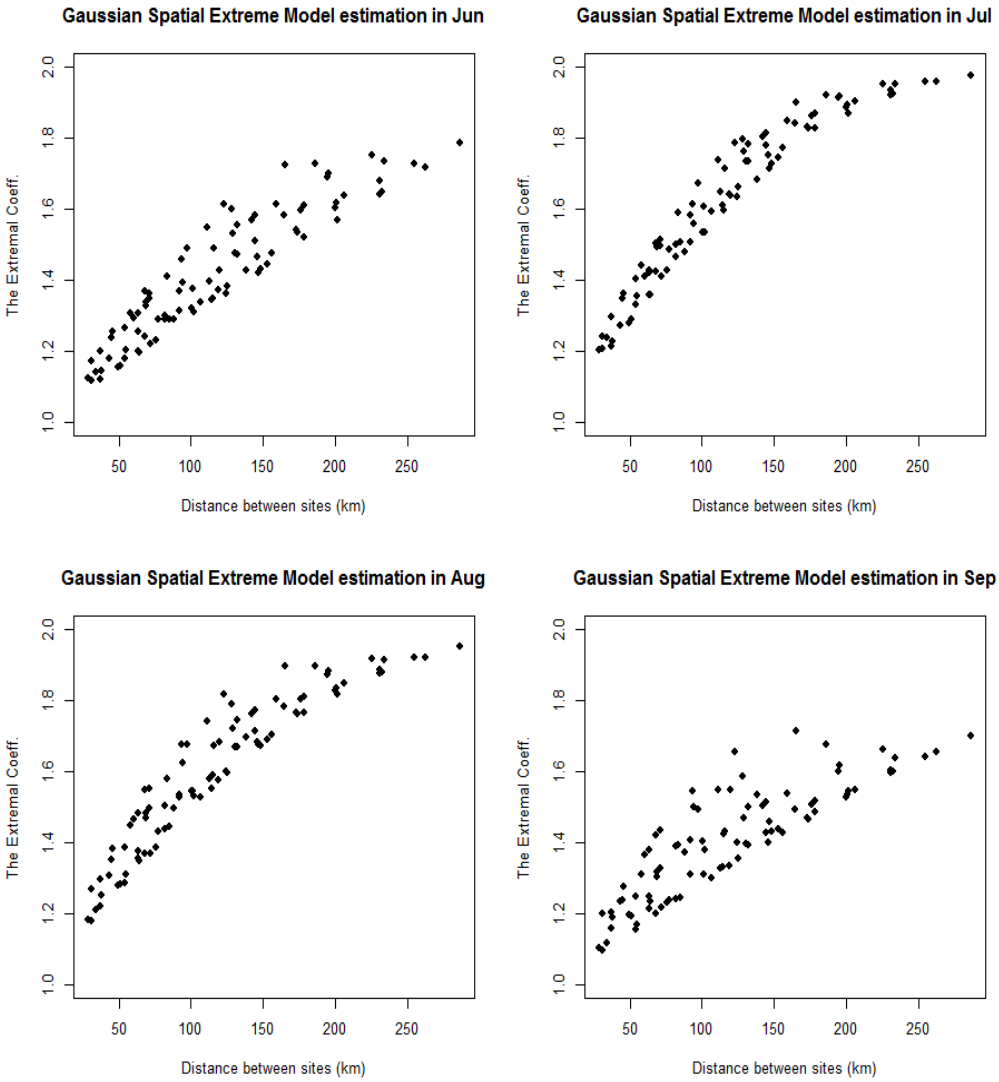


그림 89 우리나라 호우기 (6월-9월)에 대해 Gaussian 공간 극단값 모형 적합 결과. Extremal coefficient 추정값 $\mathrm{EC}(j_1, j_2, \hat{\Sigma})$을 지점 s_{j_1}과 s_{j_2}간 거리에 따라 표시하였다. Extremal coefficient는 항상 1과 2 사이에 있는데, 1에 가까우면 지점간 상관성이 큰 것이고, 반대로 2에 가까우면 독립적이다. 모든 월에서 지점간 거리가 증가할수록 독립적이나, 7월과 8월에는 상대적으로 지점간 상관성이 떨어진다.

연구보고서 2015-04

통계적 상세화를 위한 Weather Generator 개발

Development of Weather Generator for Statistical Downscaling

김무섭

APCC
APEC CLIMATE CENTER

APEC 기후센터
48058 부산광역시 해운대구 센텀7로12
Tel: 051-745-3900 Fax: 051-745-3949

www.apcc21.org

93550
9 791156 980940
ISBN 979-11-5698-094-0