



저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원 저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)



이학석사 학위논문

**Statistical Multisite Simulation of
Summertime Precipitation over South
Korea and Its Future Change
Based on Observational Data**

관측자료를 바탕으로 한 한국 여름철 강수량 모의와
모의된 결과에서의 미래 강수 변화

2013년 2월

서울대학교 대학원

지구환경과학부

김 지 원

Statistical Multisite Simulation of Summertime Precipitation over South Korea and Its Future Change Based on Observational Data

지도 교수 Kwang-Yul Kim

이 논문을 이학석사 학위논문으로 제출함
2013년 2월

서울대학교 대학원
지구환경과학부
김지원

김지원의 이학석사 학위논문을 인준함
2013년 2월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Abstract

Statistical Multisite Simulation of Summertime Precipitation over South Korea and Its Future Change Based on Observational Data

Ji-Won Kim

School of Earth and Environmental Sciences
The Graduate School
Seoul National University

In this study, a weather generator for summer (May 19 – September 15) precipitation over South Korea is developed. Precipitation data for 33 years (1979–2011) observed at 57 stations of Korea Meteorological Administration (KMA) are used to develop a new weather generator. Using the cyclostationary empirical orthogonal function (CSEOF) technique, the observed precipitation data is described as a linear combination of deterministic evolution patterns during summer and corresponding stochastic amplitude (principal component) time series. The deterministic evolution patterns repeat themselves every year while their amplitudes vary from one year to another. Then, an autoregressive moving average (ARMA) model was fit to each of the detrended stochastic amplitude time series. The resulting ARMA models are used to generate synthetic amplitude time series of arbitrary length such that they have

statistical properties similar to the original amplitude time series. Then, one hundred sets of new amplitude time series are constructed for the period of 1979-2061 (83 years). Based on these synthetic time series and the annually repeating evolution patterns, one hundred sets of synthetic summer precipitation records were generated.

Statistical characteristics of the synthetic datasets are examined in comparison with those of the KMA observational record for the period of the observational record. The synthetic datasets, particularly for a future period, are also compared with the precipitation from HadGEM3-RA regional climate model (RCM) under various climate change (representative concentration pathways: RCP) scenarios for further validation of the performance of the statistical approach developed in the present study.

The seasonal cycle in the synthetic precipitation is reproduced faithfully with typical bimodal peaks of summer precipitation although its amplitude is slightly overestimated. The RCM underestimates significantly the second peak of precipitation. The mean and the standard deviation of summer precipitation in the seasonal cycle also exhibit similar characteristics; the synthetic data reasonably reproduce the mean and the standard deviation of the seasonal cycle whereas the historical run data significantly underestimate them.

The frequency-intensity relationship of the synthetic precipitation also looks similar to that of the observational data. The frequency-intensity relationship of the model precipitation under the “historical run” significantly underestimates the occurrence of no-rain days and rainfall events with precipitation greater than 20 mm day⁻¹ whereas the number of rainfall events with precipitation between 0 and 20 mm

day^{-1} is seriously overestimated. The distribution of extreme events is delineated in terms of generalized Pareto distribution (GPD). The distribution of extreme events in the synthetic data is reasonably similar to that of the observational data whereas the model data under the historical run underestimates the occurrence of extreme events. The spatial correlation patterns of both the synthetic precipitation and the historical run precipitation are fairly similar to that of the observational data; both the statistical approach and the model reproduce the spatial characteristics of summer precipitation in Korea reasonably.

In the future period, precipitation amounts increases except in the precipitation range of $(0,10)$ mm day^{-1} with nearly no change in the frequency of no-rain days; frequency increase is particularly conspicuous in the range of $[100,500)$ mm day^{-1} . The model precipitation under the RCP 4.5 and RCP 8.5 scenarios both exhibit increased frequency of heavy precipitation greater than 100 mm day^{-1} compared with the historical run. The relative frequency of heavy precipitation clearly increases according to the GPD distributions of RCP 4.5 and RCP 8.5 precipitation. Nonetheless, the future frequency of heavy precipitation in the range of $[100,500)$ mm day^{-1} in the model datasets is still generally below that of the observational record. On the other hand, heavy precipitation events exceeding 500 mm day^{-1} are clearly more abundant than in the observational record.

Keywords: synthetic precipitation, weather generator, CSEOF, ARMA, RCM

Student Number : 2011-20362

Contents

Abstract	i
Contents	iv
List of Tables	v
List of Figures	vi
1. Introduction	1
2. Observational Data and Model Data	7
3. Method of Analysis	11
3.1. CSEOF analysis	11
3.2. The k -Means Clustering	12
3.3. Autoregressive-Moving Average (ARMA) Modeling	13
3.4. Generating Synthetic Data for Non-Extreme Precipitations	17
3.5. Generating Synthetic Data for Extreme Precipitations	18
4. Comparison of KMA Data with the RCM Datasets	20
5. Comparison of KMA Data with the Synthetic Datasets	51
6. Summary and Concluding Remarks	76
References	82
국문 초록	87

List of Tables

Table 1. The orders of ARMA model and the parameters of the skew student- <i>t</i> distribution for each PC time series of the KMA observational record from 1979-2011.	16
Table 2. Occurrence frequency per year in unit of days as a function of precipitation amount in the RCP 4.5 and the RCP 8.5 simulations for the period of 2030-2060 in comparison with the observational record (1979-2011) and the historical run (1950-2005).	39
Table 3. Occurrence frequency per year in unit of days as a function of precipitation amount in the RCP 4.5 and the RCP 8.5 simulations for the period of 2070-2100 in comparison with the observational record (1979-2011) and the historical run (1950-2005).	40
Table 4. The scale and the shape parameters of the GPD distribution for the observational record, the historical run, the RCP 4.5 run, and the RCP 8.5 run. The two scale and the shape parameters represent the period of 2030-2060 and the period of 2070-2100, respectively.	41
Table 5. Occurrence frequency per year in unit of days as a function of precipitation amount in the synthetic datasets during the observational period (1979-2011: SYN 33P) and in the future period (2029-2061: SYN 33F) in comparison with the observational record (1979-2011: OBS 33P). The occurrence frequencies below 100 mm day ⁻¹ in the observational data are based on the 20-mode reconstruction of precipitation.	73

List of Figures

Figure 1. The 57 KMA stations over the Korean Peninsula and the 5 groups of clusters of daily KMA precipitations for the period of 1979-2011.	9
Figure 2. The spatial domain of HadGEM3-RA (red box) with the horizontal resolution of 12.5 km, which derives from the coupled atmosphere-ocean general circulation model HadGEM2-AO with the horizontal resolution 135 km. The blue box denotes the domain for CSEOF analysis in the present study.	10
Figure 3. Summer-mean (May 19-September 15) patterns of the 1st CSLV of precipitation (mm day^{-1}): (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.	22
Figure 4. Patterns of standard deviation of summer (May 19-September 15) precipitation (mm day^{-1}) of the 1st CSLV: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.	23
Figure 5. The locations of HadGEM3-RA grid points closest to the 57 KMA stations over the Korean Peninsula.	26
Figure 6. The seasonal cycle of summer (May 19-September 15) precipitation averaged at 57 stations over the Korean Peninsula: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run. The red dotted curve denotes low-pass filtered seasonal cycle with a cut-off period of 30 days. (e) Comparison of the seasonal cycles in (a)-(d).	27
Figure 7. The low-passed seasonal cycle with the cutoff period of 30 days averaged over each cluster group defined in Figure 1: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.	30
Figure 8. The band-pass filtered CSLV of the seasonal cycle with a cut-off periods of 10-30 days for each cluster group: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.	31

Figure 9.	The standard deviation of the high-frequency (period less than 10 days) component of the seasonal cycle for each cluster group.	32
Figure 10.	The PC time series of the first CSEOF mode (the seasonal cycle) from the KMA observational data (red), the historical run (orange), the RCP 4.5 run (green), and the RCP 8.5 run (blue). All the PC time series are normalized. In panels (b) and (c), the PC time series of the RCP 4.5 data and that of the RCP 8.5 data were simply connected with that of the historical run since RCP 4.5 and RCP 8.5 runs start from the historical run.	34
Figure 11.	The AR spectra of the PC time series in Figure 10: observational data (red), historical run (orange), RCP 4.5 run (green), and RCP 8.5 run (blue).	35
Figure 12.	Histogram of precipitation for the amount less than 100 mm day ⁻¹ (left column) and for more than 100 mm day ⁻¹ (right column). The histograms based on the RCP 4.5 and RCP 8.5 runs represent the near future (2030-2060) in the upper panels and the far future (2070-2100) in the lower panels.	38
Figure 13.	The generalized Pareto distribution (GPD) of the extreme events with precipitation amount exceeding 100 mm day ⁻¹ . The RCP 4.5 and RCP 8.5 runs represent the near future (2030-2060) in the left column and the far future (2070-2100) in the right column.	42
Figure 14.	The occurrence frequency of dry days (upper panel) and precipitation events of (0,10) mm day ⁻¹ (lower panel) for the eight subintervals of 15 days each during summer. The red bars denote the observational data, and the yellow bars denote the RCM historical run.	45
Figure 15.	The occurrence frequency of precipitation for the amount of [10,100) mm day ⁻¹ (upper panel), for the amount of [100,500) mm day ⁻¹ (middle panel), and for the amount over 500 mm day ⁻¹ (bottom panel). The red bars denote the observational data, the yellow bars the historical run, the green bars the RCP 4.5 run, and	

the blue bars the RCP 8.5 run. The RCP 4.5 and RCP 8.5 results are based on the data in the near future (2030-2060). 46

Figure 16. The occurrence frequency of precipitation for the amount of [10,100] mm day-1 (upper panel), for the amount of [100,500] mm day-1 (middle panel), and for the amount over 500 mm day-1 (bottom panel). The red bars denote the observational data, the yellow bars the historical run, the green bars the RCP 4.5 run, and the blue bars the RCP 8.5 run. The RCP 4.5 and RCP 8.5 results are based on the data in the far future (2070-2100). 47

Figure 17. (left column) Spatial correlation function, $C(r, r')$, of the observational data with extreme events at 57 KMA stations (upper panel), and spatial correlation function of the historical run at 57 grid points in Figure 5 (bottom panel). (right column) Spatial correlation function with respect to the Seoul station. 49

Figure 18. (left column) Spatial correlation function of the RCP 4.5 run at the 57 grid points in Figure 5 (upper panel), and spatial correlation function of the RCP 8.5 run at 57 grid points (bottom panel). (right column) Spatial correlation function with respect to the Seoul station. 50

Figure 19. (top) Comparison of the first PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the first PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) in comparison with the standard deviation of the first PC time series from the observational record (red). 52

Figure 20. (top) Comparison of the second PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the second PC time

series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with the their mean (cyan) in comparison with the standard deviation of the second PC time series from the observational record (red). 53

Figure 21. (top) Comparison of the third PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the third PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with the their mean (cyan) in comparison with the standard deviation of the third PC time series from the observational record (red). ... 54

Figure 22. Histograms of the first eight PC time series of the observational data (red), and those of the 100 synthetic PC time series (black) with their mean histogram (yellow) for the observational period (1979-2011). 55

Figure 23. (top) Comparison of the first PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the first PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with the their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the first PC time series from the observational record (red). 57

Figure 24. (top) Comparison of the second PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the second

PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the second PC time series from the observational record (red). 58

Figure 25. (top) Comparison of the third PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the third PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the third PC time series from the observational record (red). 59

Figure 26. Histograms of the first eight PC time series of the observational data (red), and those of the 100 synthetic PC time series (black) with their mean histogram (yellow) for the future period (2029-2061). 60

Figure 27. (left column) Spatial correlation function of the observational data without any extreme events at 57 KMA stations (top), spatial correlation function of the 20-mode truncated observational data without any extreme events at 57 KMA stations (middle), and spatial correlation function of the 100 synthetic datasets at 57 stations in Figure 5 (bottom). (right column) Spatial correlation function with respect to the Seoul station. 62

Figure 28. Mean (left column) and standard deviation (right column) of the seasonal cycle derived from the observational data with extreme events (top), mean of the seasonal cycle based on the 100 synthetic datasets over the same period of the observational data (middle), and mean of the seasonal cycle based on the 100 synthetic datasets over the period of 2029-2061. 63

Figure 29. The temporal evolution of the seasonal cycle in (a) the observational data, (b) the 100 synthetic datasets over the same

period of the observational record (1979-2011), (c) the 100 synthetic datasets over the period of 2029-2061, and (d) all three low-passed seasonal cycles in (a)-(c). The cyan curves in (b) and (c) are the averages of all the synthetic seasonal cycles while the red dashed curve is the seasonal component (period > 30 days) of the average seasonal cycle. 66

Figure 30. The temporal evolution of the low-passed seasonal cycle with the cutoff period of 30 days in each of the five cluster groups in (a) the observational record, (b) the 100 synthetic datasets over the same period of the observational record (1979-2011), and (c) the 100 synthetic datasets over the period of 2029-2061. 67

Figure 31. The intraseasonal component (period of 10-30 days) of the seasonal cycle for each of the five cluster groups in (a) the observational record, (b) the 100 synthetic datasets over the same period of the observational record (1979-2011), and (c) the 100 synthetic datasets over the period of 2029-2061. 68

Figure 32. Standard deviation of the high-frequency component with period less than 10 days of the seasonal cycle for each of the five cluster groups in the observational data (red), the 100 synthetic datasets over the same period of the observational record (green), and the 100 synthetic datasets over the period of 2029-2061 (blue). 69

Figure 33. Histogram of precipitation for the amount less than 100 mm day^{-1} (left column) and for more than 100 mm day^{-1} (right column) in the observational period (upper panels) and in the period of 2029-2061 (lower panels). The red curve denotes the observational data, the black curves the 100 synthetic datasets, and the cyan curve the mean of the 100 black curves. 72

Figure 34. The occurrence frequency of dry days (upper panel) and precipitation events of $(0,10) \text{ mm day}^{-1}$ (lower panel) for the eight subintervals of 15 days each during summer. The red bars denote the observational data, the green bars the 100 synthetic datasets over the same period of the observational record, and the blue bars the 100 synthetic datasets over the period of 2029-2061.

.... 74

Figure 35. The occurrence frequency of precipitation for the amount of $[10,100] \text{ mm day}^{-1}$ (upper panel), for the amount of $[100,500] \text{ mm day}^{-1}$ (middle panel), and for the amount over 500 mm day^{-1} (bottom panel). The red bars denote the observational data, the green bars the 100 synthetic datasets over the same period of the observational record, and the blue bars the 100 synthetic datasets over the period of 2029-2061. 75

1. Introduction

A stochastic weather generator (WG) is a statistical model generating synthetic time series of weather data at a given location. A synthetic time series in a WG is generated typically based on the statistical characteristic of the observed weather at the specified location. Such weather variables as daily precipitation, maximum temperature, minimum temperature, solar radiation, humidity, and wind speed are produced by a WG. Conventionally, a WG first generates precipitation and the remaining variables are determined with the occurrence of precipitation events in consideration [Wilks and Wilby, 1999]. One of the utilities of a WG is to construct a long synthetic time series or a large number of realizations so that statistical inferences can be made on the variability of weather variables, which otherwise would be difficult based on a relatively short observational record.

In conventional WGs, the occurrence probability of precipitation is calculated on a given day based on occurrence or non-occurrence of precipitation on the previous day or on several consecutive days prior to the given day [Gabriel and Neumann, 1962; Richardson, 1981; Richardson and Wright, 1984; Stern and Coe, 1984; Wilks, 1999]. Such an approach is conceptually based on the assumption that occurrence of precipitation can be modeled by a Markov process [Durrett, 2010]. The amount of precipitation for a wet day, then, is randomly generated from a distribution function—typically one of the parametric distribution functions such as exponential distribution, gamma distribution, and mixed exponential distribution—fitted to observation data [Buishand, 1977; Chen et al., 2010; Dubrovsky et al., 2004; Katz, 1977; Richardson,

1981; Stern and Coe, 1984; Thom, 1958; Wilby 1994; Wilks, 1989, 1992; Woolhiser and Pegram, 1979].

In order to account for the seasonal cycle of precipitation, parameters of the distribution function are estimated for each month and daily precipitation for each month is independently generated from the monthly distribution function. Therefore, any continuity of daily precipitation between two consecutive months is ignored. In fact, the summertime rainfall in Korea exhibits two rainy spells separated by a period of weakened precipitation [Chen *et al.*, 2003; Ho and Kang, 1988; Kim *et al.*, 2010; Lee *et al.*, 2010; Lim *et al.*, 2002; Qian *et al.*, 2002; Roh *et al.*, 2012; Wang *et al.*, 2007]. Although the initiation and duration of the rainy spells are somewhat flexible, the climatological peak of the first rain spell occurs in mid-July while the peak of the second rain spell occurs in late August. Since the daily precipitation associated with the seasonal cycle evolves strongly in time, synthetic time series based on the monthly distribution functions cannot accurately describe the seasonal evolution of precipitation. It can be attempted to generate synthetic precipitation anomalies by removing the climatological seasonal cycle derived from a composite analysis; it is emphasized that the amplitude of the seasonal cycle varies strongly from one year to another. Therefore, the resulting synthetic precipitation dataset bounds to overestimate or underestimate the seasonal cycle.

A conventional weather generator is designed based on statistical properties of observational data at a given station; typically spatial correlations of weather variables between two neighboring stations are not considered in the development of a weather

generator. Therefore, synthetic time series at two adjacent stations do not exhibit an accurate spatial correlation structure. In a more advanced approach, synthetic time series are constructed based on a multivariate normal distribution in order to explain appropriately the spatial correlation of precipitation among multiple stations [Wilks, 1998]. Covariance matrix of a multivariate normal variable, however, is not necessarily definite positive [Wilks, 1998], in which case, the construction of synthetic time series at multiple stations is not straightforward. *Lee et al.* [2010] improved the multivariate approach by using the alternating projection (AP) algorithm [*Monserrat et al.*, 2007]. *Lee et al.* [2010] generated synthetic precipitation over South Korea from the improved weather generator. Although the spatial correlation of precipitation was improved compared with the algorithm of Wilks [1998], the regional scale of high correlation greater than ~0.6 was underestimated compared with that of the observation record. Also, lagged cross correlation between two stations tends to be underestimated; lagged cross correlation is an important characteristic of the precipitation field since the weather system often is observed to migrate from one region to another in the course of time.

The primary goal of this study to develop a method of synthetically generating daily summer precipitation, which account for approximately the three-fourths of the annual precipitation and is considered a main water resource for industrial and agricultural use in Korea [*Kim et al.*, 2010; *Roh et al.*, 2012]. A statistical approach will be developed based on cyclostationary empirical orthogonal function (CSEOF) analysis [*Kim et al.*, 1996; *Kim and North*, 1997] of the observed precipitations at the

Korea Meteorological Administration (KMA) stations. The new algorithm is founded on a sound conceptual and physical basis, since CSEOF analysis extracts physical evolutions together with their stochastic amplitude time series from a given dataset. As mentioned above, spatial correlation and lagged cross correlation, not to mention mean and variance, are important characteristics of summer precipitation in Korea. Therefore, it is essential for a weather generator to be able to reproduce the physical processes of precipitation in a reasonable manner. In this sense, CSEOFs constitute an excellent basis set for developing an accurate weather generator.

While numerical models are frequently used in place of statistical algorithms to generate synthetic precipitation, a large number of uncertainties are inevitable in making statistical and physical inferences on the future characteristics of precipitation. Therefore, it is a sensible idea to compare precipitation from a model with synthetic precipitation based on a statistical approach. Likewise, statistical approaches should be assessed for their physical and statistical validity against numerical models. For these reasons, precipitation from a regional climate model (RCM) is also analyzed and compared with the observational data in terms of their physical and statistical characteristics. This procedure, in turn, helps us appraise the performance of the statistical weather generator to be developed in the present study. In particular, statistical and physical properties of model precipitations under future climate scenarios are important baseline solutions, against which to compare synthetic precipitations.

In summary, there are four specific objectives in the present study:

- (1) Develop a statistical weather generator based on the CSEOFs of the KMA observations to generate synthetic precipitation over Korea with accurate physical and statistical properties.
- (2) Test the performance of the statistical weather generator developed in the present study by investigating the physical and statistical properties of synthetic precipitation in comparison with those of the observational data.
- (3) Analyze precipitation from a regional climate model in comparison with the observational data and the synthetic precipitation generated from the statistical weather generator.
- (4) Derive physical and statistical inferences on the future characteristics of precipitation over Korea based on the model precipitation and the synthetic precipitation.

It is a new and novel attempt of this study to employ a bilateral approach in generating synthetic precipitation based on numerical and statistical models.

The observational precipitation data and the model data used in the present study are addressed in Section 2. A detailed description and explanation of the methods used in this study—CSEOF analysis, autoregressive-moving average modeling, and the k -means clustering technique—appears in Section 3. Then the RCM’s ability to generate precipitation over Korea in the present and the future climate conditions will be evaluated in Section 4 in terms of the statistical properties of modeled and observed precipitations with an emphasis on future characteristics of precipitation. Then, synthetic precipitation data generated from the weather generator

developed in the present study are evaluated by comparing their statistical properties with those of the observational data in Section 5. Statistical inferences on the future characteristics of precipitation will be made based on the synthetic precipitation data extended until 2061. Summary and concluding remarks follow in Section 6.

2. Observational and Model Data

Daily precipitation data observed at 57 stations of Korea Meteorological Administration (KMA) during summer (May 19–September 15) for 33 years (1979 - 2011) were used in this study (Figure 1). The 57 KMA stations are dispersed over the Korean Peninsula [$\sim 126^{\circ}$ - 130° E \times $\sim 34^{\circ}$ - 38.5° N]. Four of the 61 KMA stations were excluded in the present study since one station (Andong station) has an incomplete record and the others represent precipitation over islands (Jeju Island and Ulleung Island) remote from the peninsula.

Daily precipitation data of regional climate model (RCM: HadGEM3-RA, [Walters *et al.*, 2011]) are used to examine the present and future prospects of precipitation during summer over South Korea. The climate scenarios adopted in this study represent the Representative Concentration Pathways (RCPs) reported in the Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate Change (IPCC; <http://www.ipcc.ch/activities/activities.shtml>). To simulate regional precipitations, National Institute of Meteorological Research (NIMR)/KMA ran the regional climate model based on the atmospheric component of the Earth System Model, HadGEM3, of the United Kingdom (UK) Met Office Hadley Centre. Three different climate scenarios were employed in the present study: historical run from 1950-2005, RCP 4.5 and RCP 8.5 runs from 2006-2100. The spatial domain of HadGEM3-RA includes the eastern part of China and most of Japan around the Korea peninsula [$\sim 111.0^{\circ}$ - 144.625° E \times $\sim 26.5^{\circ}$ - 48.875° N] with a horizontal resolution of 12.5 km (Figure 2). Global climate data obtained from the global coupled atmosphere-ocean general

circulation model, HadGEM2-AO [Collin *et al.*, 2008], with the horizontal resolution 135 km for each scenario were used as a boundary forcing for HadGEM3-RA.

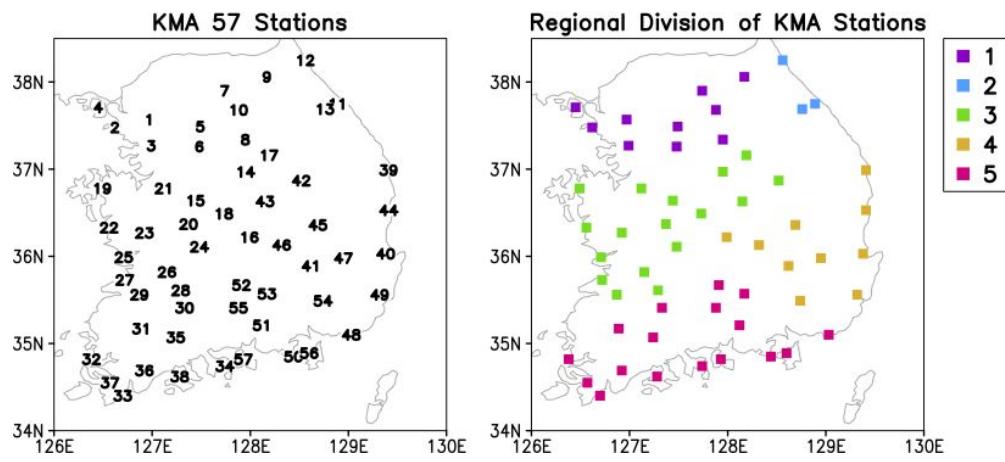


Figure 1. The 57 KMA stations over the Korean Peninsula and the 5 groups of clusters of daily KMA precipitations for the period of 1979-2011.

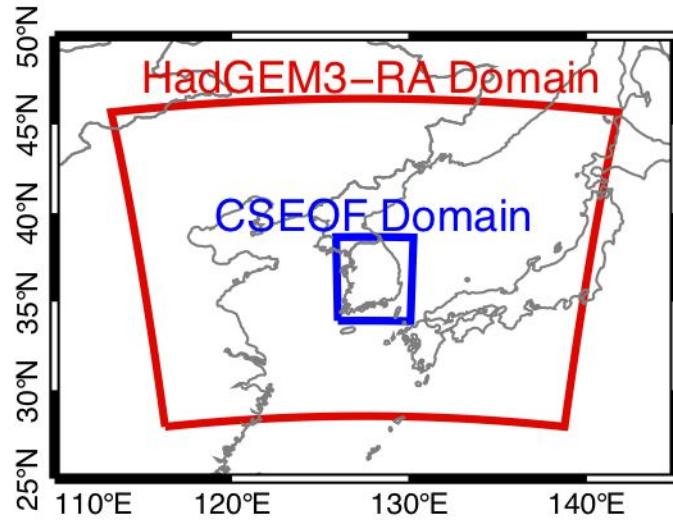


Figure 2. The spatial domain of HadGEM3-RA (red box) with the horizontal resolution of 12.5 km, which derives from the coupled atmosphere-ocean general circulation model HadGEM2-AO with the horizontal resolution 135 km. The blue box denotes the domain for CSEOF analysis in the present study.

3. Methods of Analysis

3.1. CSEOF Analysis

We used cyclostationary empirical orthogonal function (CSEOF) analysis to decompose space-time precipitation data, $P(r,t)$, into deterministic evolution patterns and corresponding stochastic amplitude time series [Kim *et al.*, 1996; Kim and North, 1997]. In CSEOF analysis, space-time precipitation data are written as:

$$P(r,t) = \sum_n B_n(r,t) T_n(t), \quad t \in D \quad (1)$$

where $B_n(r,t)$ is cyclostationary loading vector (CSLV) and $T_n(t)$ is principal component (PC) time series and D is the record length of the data. The CSLVs are periodic, that is,

$$B_n(r,t) = B_n(r,t+d), \quad (2)$$

where d is called the nested period; in this study $d=120$ days since each year consists of 120 summer days. Each CSLV evolves in time with the given nested period and describes a physical evolution of precipitation during summer in South Korea. In particular, the first CSEOF mode, the seasonal cycle of precipitation, is examined in detail in the present study.

To reduce nonstationarity of the PC time series generated by the extreme events in the observational data, extreme precipitations more than 100 mm day^{-1} were removed from the observational data. To avoid abrupt discontinuity in the record with precipitations less than or equal to 100 mm day^{-1} , linear interpolation was carried out at the points where extreme precipitations were removed; the resulting precipitation

dataset is called “no-extreme” precipitation record. In order to avoid negative precipitation in reconstruction and synthetic data based on a limited number of CSEOF modes, natural logarithmic transformation was conducted after resetting the daily minimum precipitation to 1 mm day^{-1} in the KMA observational record; thus, any days with precipitation less than or equal to 1 mm day^{-1} are considered no-rain days. This procedure does not introduce any significant error or bias. Then the log-transformed “no-extreme” precipitation data were subject to CSEOF analysis. The “extreme” precipitation data that were separated will be dealt based on a different approach, which will be discussed below.

3.2. The k -Means Clustering

Applying the k -means clustering on the first CSLV of the KMA observation data, the observational data were classified into 5 clusters (Figure 1). The k -means clustering separates data at n points into k groups in such a way that the sum of squared differences from the cluster center to the point in each group is minimized as follows:

$$Var = \sum_{i=1}^k \sum_{j \in S_i} \|x_j - \mu_i\|^2, \quad (3)$$

where x_j is the d measurements at the j th data point, S_i denotes a set of points belonging to the i th cluster and μ_i is d -dimensional center of the i th cluster, or equivalently, the mean of points in S_i . In this study, $n=57$ stations, $d=120$ days, and $k=5$ clusters. Thus, x_j represents the 1st CSLV at the j th stations for 120 summer days.

The detailed procedures of the k -means clustering are as follow:

- Step 1: Any k points are randomly assigned as the centers of k clusters.
- Step 2: Distances of each point from the assumed cluster centers are computed and each point is assigned to the cluster that has the smallest distance from the given point.
- Step 3: Mean vectors for the resulting clusters are computed as the new cluster centers.
- Step 4: Repeat procedures in Steps 2 and 3 to find new cluster centers until the assignments of points do not change further.

3.3. Autoregressive-Moving Average (ARMA) Modeling

New sets of artificial PC time series can be generated by using an autoregressive-moving average model fitted to the PC time series of the observational precipitation data. An ARMA(p, q) model for PC time series $T_n(t)$ is written as:

$$T_n(t) = \sum_{j=1}^p \alpha_j^{(n)} T_n(t-j) + \sum_{k=1}^q \beta_k^{(n)} \varepsilon^{(n)}(t-k) + \varepsilon^{(n)}(t), \quad (4)$$

where p the autoregressive (AR) order, $\{\alpha_j^{(n)} | j=1, \dots, p\}$ are AR coefficients, q is moving-average (MA) order, $\{\beta_k^{(n)} | k=1, \dots, q\}$ are MA coefficients, and $\varepsilon^{(n)}(t)$ is white noise time series with $SkewT(x; \nu, \xi)$ distribution [Fernández and Steel, 1998]. The probability distribution function of the $SkewT(x; \nu, \xi)$ distribution is defined as

$$f(x) = \frac{2\xi}{1+\xi^2} [t_\nu(\xi x)I(x < 0) + t_\nu(x/\xi)I(x \geq 0)], \quad (5)$$

where

$$t_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}} \quad (6)$$

is the student's t distribution with ν degrees of freedom (also called the shape parameter of the skew t distribution), the indicator function $I(A)$ is defined by

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is not true} \end{cases}, \quad (7)$$

and ξ is the skewness parameter of the $SkewT(x; \nu, \xi)$ distribution. The random variable x is formed by subtracting mean, μ , from the white noise time series and then diving the zero-mean time series by the respective standard deviation σ . That is,

$$x = \frac{\varepsilon^{(n)}(t) - \mu_n}{\sigma_n}, \quad (8)$$

where μ_n and σ_n are called the location parameter (mean) and the scale parameter (standard deviation). The superscript (n) denotes that the parameters are for the n th PC time series. Fitting an ARMA(p, q) model to a PC time series means to determine AR and MA orders, AR and MA coefficients and the parameters of white noise time series (Table 1). A best model is determined based on the AIC (Akaike's Information Criterion) measure [Newton, 1988] and the significance of AR and MA

coefficients. Note, in particular, that any trend in a PC time series is removed by a simple least square error method prior to fitting an ARMA model.

Table 1. The orders of ARMA model and the parameters of the skew student- t distribution for each PC time series of the KMA observational record from 1979-2011.

Mode	ARMA order		Estimated parameters of skew student- t distribution			
	p	q	μ	σ	ν	ξ
1	4	1	1.423E-04	1.210E-03	4	1
2	5	0	6.298E-06	1.313E-03	4	1
3	7	0	-6.770E-05	1.220E-03	4	1
4	4	4	-4.161E-07	1.056E-03	4	1
5	3	3	3.036E-05	1.271E-03	4	1
6	5	1	4.133E-05	9.927E-04	4	1
7	8	0	-3.440E-05	1.052E-03	4	1
8	3	3	-1.305E-05	8.853E-04	4	1
9	8	0	1.661E-05	9.847E-04	4	1
10	3	1	-7.436E-06	1.069E-03	4	1
11	3	1	-1.330E-06	8.416E-04	4	1
12	7	0	7.520E-06	8.283E-04	4	1
13	7	3	1.046E-05	1.085E-03	4	1
14	3	3	5.347E-06	9.905E-04	4	1
15	3	3	9.135E-06	7.867E-04	4	1
16	5	1	3.121E-05	6.883E-04	4	1
17	5	1	-1.948E-05	9.936E-04	4	1
18	3	4	7.946E-06	9.324E-04	4	1
19	3	3	3.681E-06	6.853E-04	4	1
20	3	3	-2.760E-05	8.309E-04	4	1

3.4. Generating Synthetic Data for Non-Extreme Precipitations

Once an optimal ARMA(p, q) is determined for each PC time series, synthetic PC time series is generated via:

$$\hat{T}_n(t) = \sum_{j=1}^p \alpha_j^{(n)} \hat{T}_n(t-j) + \sum_{k=1}^q \beta_k^{(n)} \hat{\varepsilon}^{(n)}(t-k) + \hat{\varepsilon}^{(n)}(t), \quad (9)$$

where $\hat{\varepsilon}^{(n)}(t)$ is a synthetic white noise time series that follows the $SkewT(x; \nu, \xi)$ distribution [Aas and Haff, 2006]. By using new realizations of the white noise time series, as many new synthetic PC time series of arbitrary length can be generated as needed. The hat sign denotes that the time series is synthetic. In this way, new sets of PC time series of log-transformed “no-extreme” precipitation can be produced. In order to avoid unrealistic values of precipitation upon exponentiation of the synthetic data based on the PC time series of the log-transformed precipitation, any precipitation exceeding 100 mm day⁻¹ was linearly interpolated by two surrounding precipitation values.

Then, synthetic datasets of Korean summertime precipitation, $\hat{P}(r, t)$, can be generated from:

$$\hat{P}(r, t) = \sum_{n=1}^N B_n(r, t) \hat{T}_n(t), \quad t \in D + R \quad (10)$$

where $\hat{T}_n(t)$ are the synthetic PC time series generated from the ARMA models of observed PC time series as described in (9) and R denotes future time. Note that the synthetic precipitation data can be extended into the future time since the periodic function $B_n(r, t)$ can be extended indefinitely. Note also that the linear slope

removed from each PC time series prior to ARMA model fitting can be added back in (9). In this way, 100 synthetic datasets of “no-extreme” precipitation were constructed for the period of 83 years (33 years of observation plus 50 years of forecast) from 1979 till 2061.

3.5. Generating Synthetic Data for Extreme Precipitations

Extreme precipitation events, in theory, can be generated from the distribution function of extreme precipitations at each station, which is fitted to the generalized Pareto distribution [GPD; *Hosking and Wallis*, 1987]. This approach, on the other hand, may generate spatially uncorrelated extreme events unless synthetic extreme events are generated with the consideration of neighboring events. We simplified this approach by assuming that extreme events occur simultaneously at all stations in one cluster. The amount of extreme precipitation, however, is randomly generated from the distribution function of extreme precipitation at each station.

Occurrence of extreme event is assumed to follow the Bernoulli distribution:

$$f(x; p) = p^x (1-p)^{1-x} \quad \text{for } x \in \{0, 1\}, \quad (11)$$

where x assumes two values (1: occurrence of an extreme event, 0: no occurrence of an extreme event) and p is the occurrence probability of an extreme event in a given cluster. In order to consider the occurrence or non-occurrence of extreme precipitation event on the day before, conditional probability of an extreme event should be considered. That is, the probability of an extreme event on day i is given by

$$p(x_i = 1) = p(x_i = 1 | x_{i-1} = 1)P(x_{i-1} = 1) + p(x_i = 1 | x_{i-1} = 0)P(x_{i-1} = 0) \quad (12)$$

It is simply assumed that the probability of extreme precipitation on a certain day depends only on the occurrence or non-occurrence of an extreme event on the day before.

To consider the frequency change of extreme events, occurrence probability and conditional probability of extreme events in a cluster were examined for each year. Then, slopes in the occurrence probability and the conditional probability were determined by fitting a line in a least square error sense. Then, occurrence probability of an extreme event was extrapolated into the future by using the slopes in the yearly probability of extreme events. In order to account for the temporal variation of the frequency of extreme events in summer, the 120-day summer period was divided into 8 15-day subintervals, for which the probability of an extreme events was separately calculated. In order to avoid unrealistically large values of precipitation, extreme values not exceeding 99.9% of the cumulative GPD were considered.

4. Comparison of KMA Data with the RCM Datasets

Before addressing the statistical aspects of precipitation in future climates, the accuracy of HadGEM3-RA in simulating precipitation in the present climate system is investigated. CSEOF analysis was conducted on the historical run (1950-2005) of HadGEM3-RA as well as the KMA observational data (1979-2011). As a reference, CSEOF analysis was conducted on the RCP 4.5 run (2006-2100), and the RCP 8.5 run (2006-2100) as well. The seasonal cycle of precipitation dominates the variability and CSEOF analysis extracts the seasonal cycle as the first mode in all the datasets above.

Figure 3 shows the mean of the first CSLV, the seasonal cycle, over the Korean Peninsula. In order to compare the magnitude of the seasonal cycle directly, each CSLV was multiplied by the square root of the respective eigenvalue (standard deviation) while, at the same time, the corresponding PC time series was normalized to have the standard deviation of one (see Eq. (1)). As can be seen, summer mean precipitation in the seasonal cycle was significantly underestimated in the model; this may be an expected result considering that the KMA precipitations are (point) measurements at stations while model precipitation at a grid point represents averages over a grid box. Gross patterns of mean precipitation in the model, on the other hand, are reasonably similar to that of the observational data. Relatively significant precipitation is seen in the northwestern part (region 1), northeastern part (region 2), and southern part (region 5) of the Korean Peninsula while relatively weak precipitation is seen in the central eastern part (region 4) in all three model datasets

although the details (such as exact location and distribution of local maximum precipitation) differ between the observations and the models. The mean precipitation increases substantially in the climate change scenarios; slightly larger increase is seen in the southern part of the peninsula in RCP 8.5 run but RCP 4.5 shows a slightly larger increase in the western part of the peninsula.

Figure 4 shows the standard deviation of summer precipitation as explained by the first CSLV (the seasonal cycle). Again, the magnitude of standard deviation is seriously underestimated by the model but the gross pattern is similar to that of the observational data in all three scenarios. The regions of relatively large standard deviation are identified with the regions of relatively large mean precipitation and vice versa. The standard deviation of the seasonal cycle increases in future scenarios; RCP 8.5 run exhibits stronger sensitivity in region 5 whereas RCP 4.5 run in regions 1 and 3.

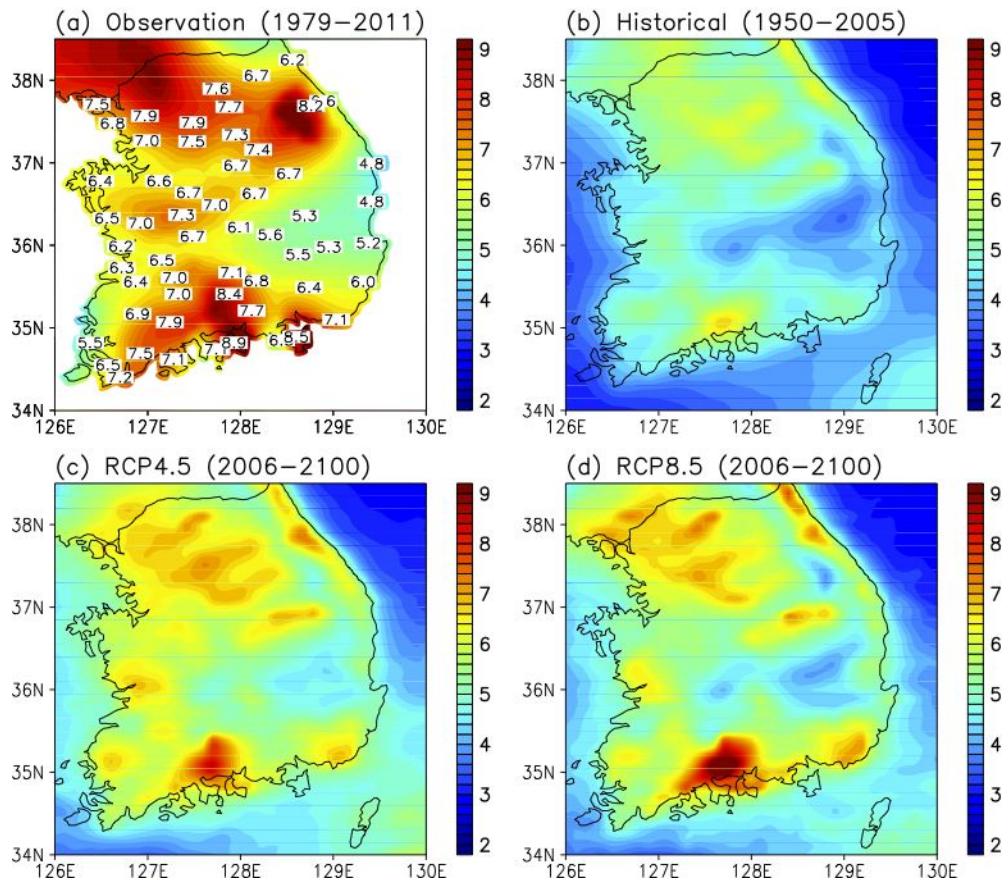


Figure 3. Summer-mean (May 19–September 15) patterns of the 1st CSLV of precipitation (mm day^{-1}): (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.

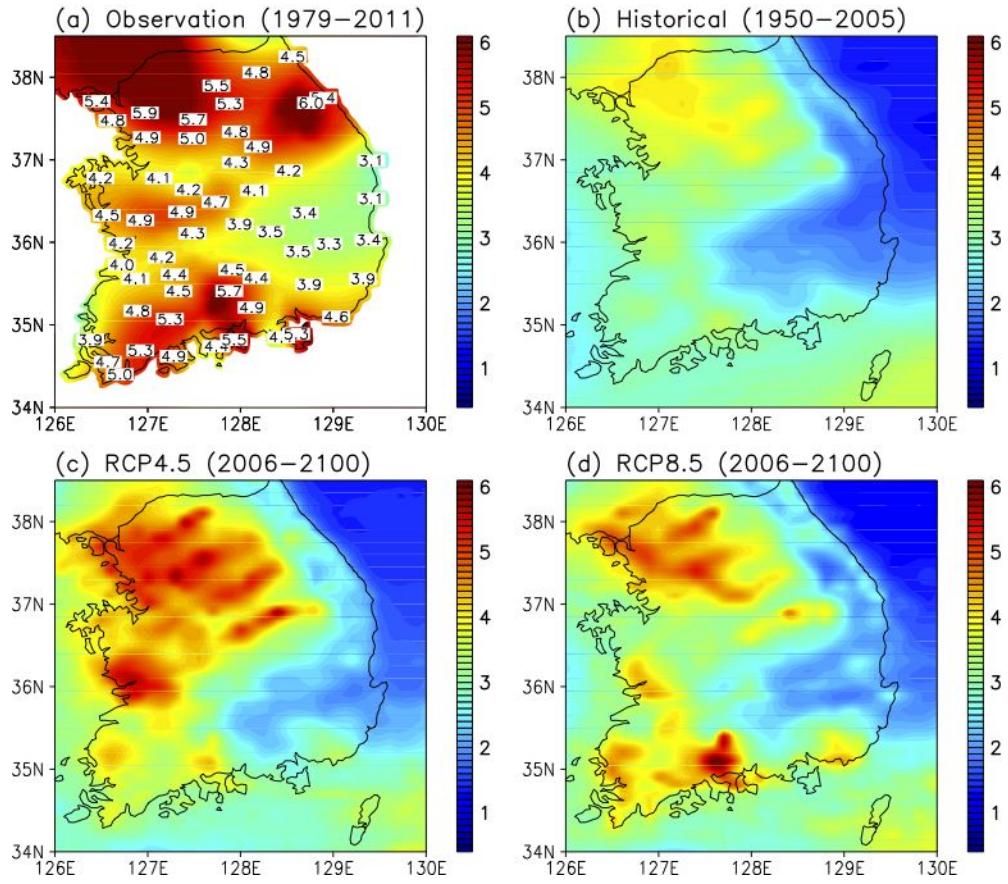


Figure 4. Patterns of standard deviation of summer (May 19–September 15) precipitation (mm day^{-1}) of the 1st CSLV: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.

In order to examine the temporal evolution of the seasonal cycle, RCM grid points closest to the KMA stations were identified as in Figure 5. Then, the seasonal cycle of precipitation averaged at the 57 KMA stations were compared with those of the HadGEM3-RA simulations (Figure 6). The red dotted lines denote the seasonal cycle with the high-frequency components with periods shorter than 30 days removed. The KMA observations show the first rainy spell (called the first Changma) in mid-July followed by the second rainy spell (called the second Changma or the Post Changma) in late August with a period of weakened precipitation between the two peaks (Figure 6a). This bimodal shape of the seasonal cycle is typical in Korea and in Japan [Lim *et al.*, 2002]. The temporal evolution of the seasonal cycle in the model datasets, on the other hand, does not exhibit the second spell of precipitation (Figures 6b-d). The first peak is underestimated and begins a little earlier in the historical run than in the observational record (Figure 6e).

In the RCP 4.5 and the RCP 8.5 simulations, the magnitude of the first peak increased substantially but the timing of the peak did not change appreciably if not delayed slightly. Roh *et al.* [2012] reported that the decadal change in the seasonal cycle of the KMA observational record was seen in the increased magnitude of the peak by ~25% and the earlier occurrence of the first peak; the decadal change in the observational record was not seen in the RCM simulations of future climate scenarios. Although the first peak of precipitation is stronger in the RCP 4.5 and RCP 8.5 simulations than in the observational record, the summer mean precipitation in the

seasonal cycle as simulated by HadGEM3-RA under the climate change scenarios is still less than that of the observational record as already discussed in Figure 3.

The second spell of precipitation is often assumed to be associated with a southward retreat of the rain band. *Lim et al.* [2002] explained the second spell of precipitation in regard to the development of positive pressure anomaly over the continent and negative pressure anomaly over the northwestern Pacific and the resulting northeasterly along the continental boundary in late August. It appears that HadGEM3-RA does not simulate the second precipitation spell accurately whatever the physical mechanism of the second precipitation spell may be. Also the Korean Peninsula is frequently visited by typhoons in July-September; specifically, the frequency of typhoon is highest during the last ten days of August. Therefore, it might be tempting to say that the second precipitation spell was not well reproduced due to the lack of RCM's ability to produce powerful typhoons. Precipitation produced by typhoons is only very weakly phase locked to calendar days. Therefore, the effect of typhoons is not significant in the seasonal cycle.

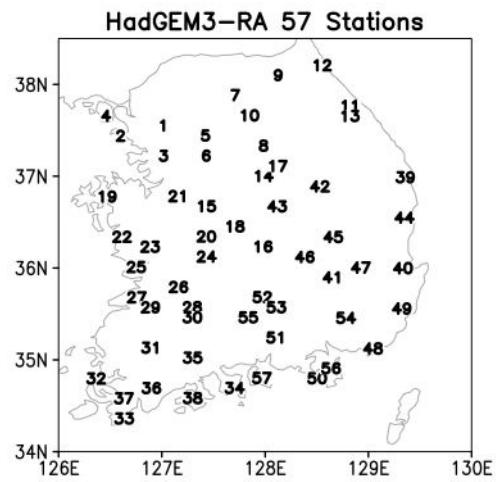


Figure 5. The locations of HadGEM3-RA grid points closest to the 57 KMA stations over the Korean Peninsula.

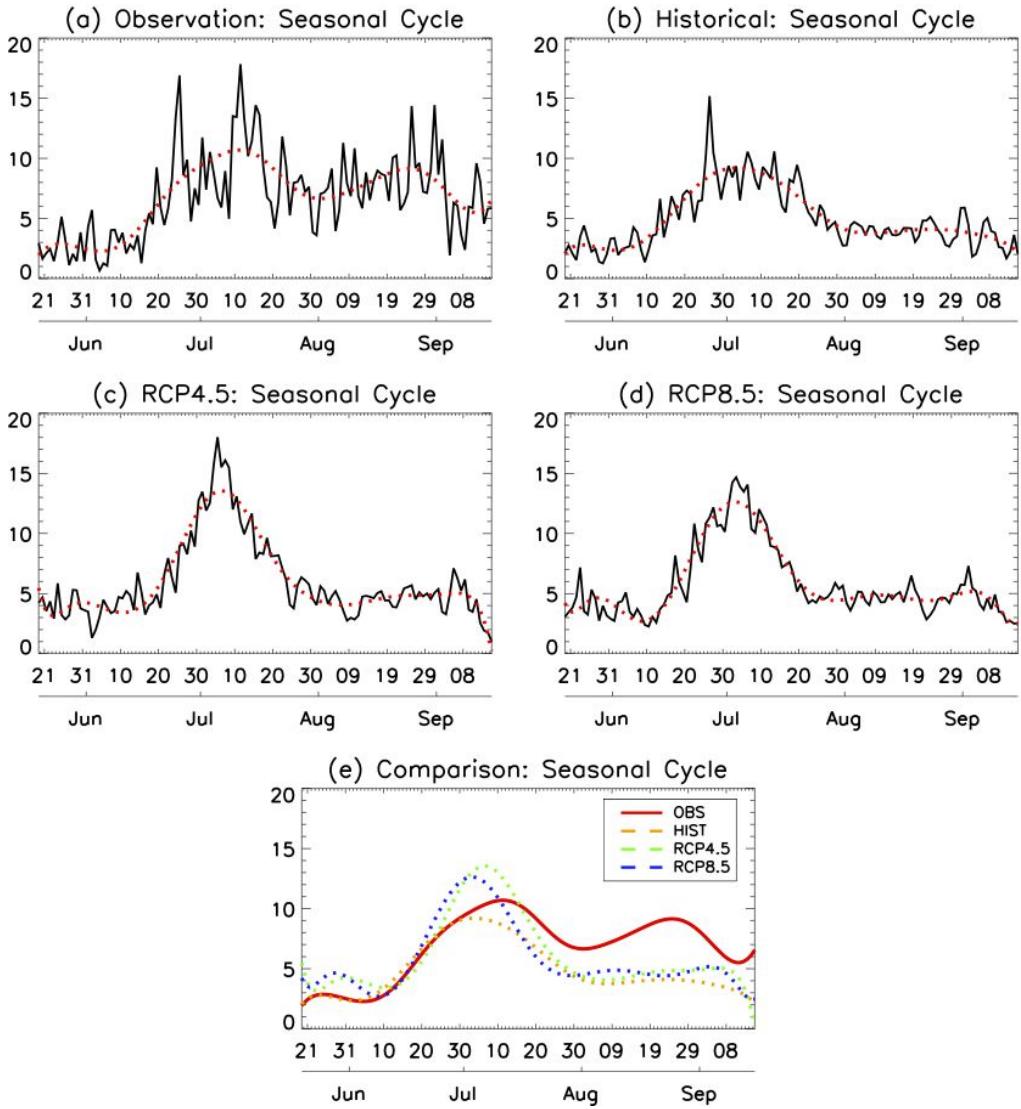


Figure 6. The seasonal cycle of summer (May 19–September 15) precipitation averaged at 57 stations over the Korean Peninsula: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run. The red dotted curve denotes low-pass filtered seasonal cycle with a cut-off period of 30 days. (e) Comparison of the seasonal cycles in (a)-(d).

Figure 7 shows the evolution of the low-passed seasonal cycle with the cutoff period of 30 days for each cluster group. The first rain band migrates northward for the observational data and the RCM datasets. It appears that the first rain spell starts later and the intensity of precipitation is stronger in the future scenarios than in the historical run (Figures 7b-7d). Also, the northward migration of the rain band from region 5 to region 1 is slightly faster in the future scenarios. The intensity of precipitation during the first rain spell is stronger in RCP 4.5 run than in RCP 8.5 run except in region 5, which is the southern part of the Korean Peninsula (Figures 7c and 7d). The first rain peak is larger than the second peak in the observational data and the RCM datasets except at region 2, which is the northeastern part of the peninsula. Figure 7 clearly shows that the seasonal cycle obtained from HadGEM3-RA is not quite realistic in terms of the temporal evolution structure of the seasonal cycle. Further, the seasonal cycle is significantly underestimated in the historical run.

Figure 8 shows the subseasonal component of the seasonal cycle for each region; the subseasonal component was obtained by conducting band-pass filtering with the cut-off periods of 10 and 30 days. The subseasonal component has an average period of \sim 15 days. The RCM model seriously underestimates the subseasonal component of the seasonal cycle although the average period of oscillations is quite compatible with that of the observational record. The strength of the subseasonal component increases in the future scenarios but there is no noticeable change in the period of oscillations.

The variance of the high-frequency component of the seasonal cycle with periods less than 10 days is shown for each region in Figure 9. The high-frequency component is strongly associated with baroclinic instability [Kim *et al.*, 2010]. High-frequency variability is strongest in region 2, which is a mountainous region in the northeastern part of the peninsula, for both the observational data and the RCM data. The RCM model, however, underestimates the high-frequency variability significantly. There is no significant increase in the variability of the high-frequency component in future climate.

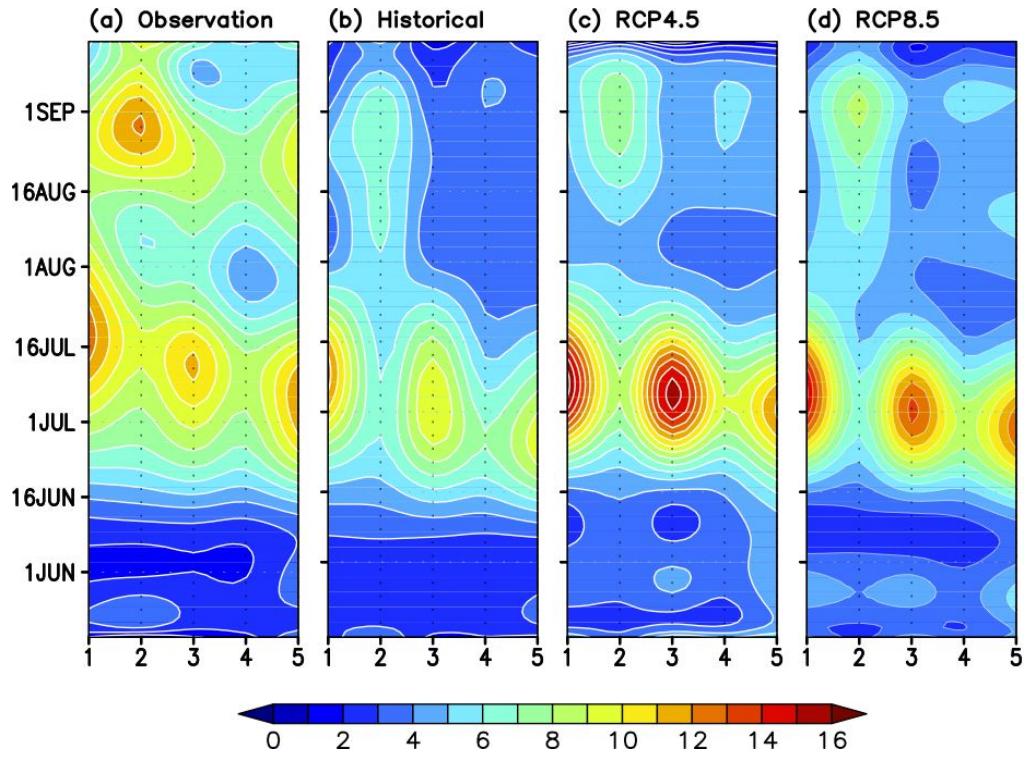


Figure 7. The low-passed seasonal cycle with the cutoff period of 30 days averaged over each cluster group defined in Figure 1: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.

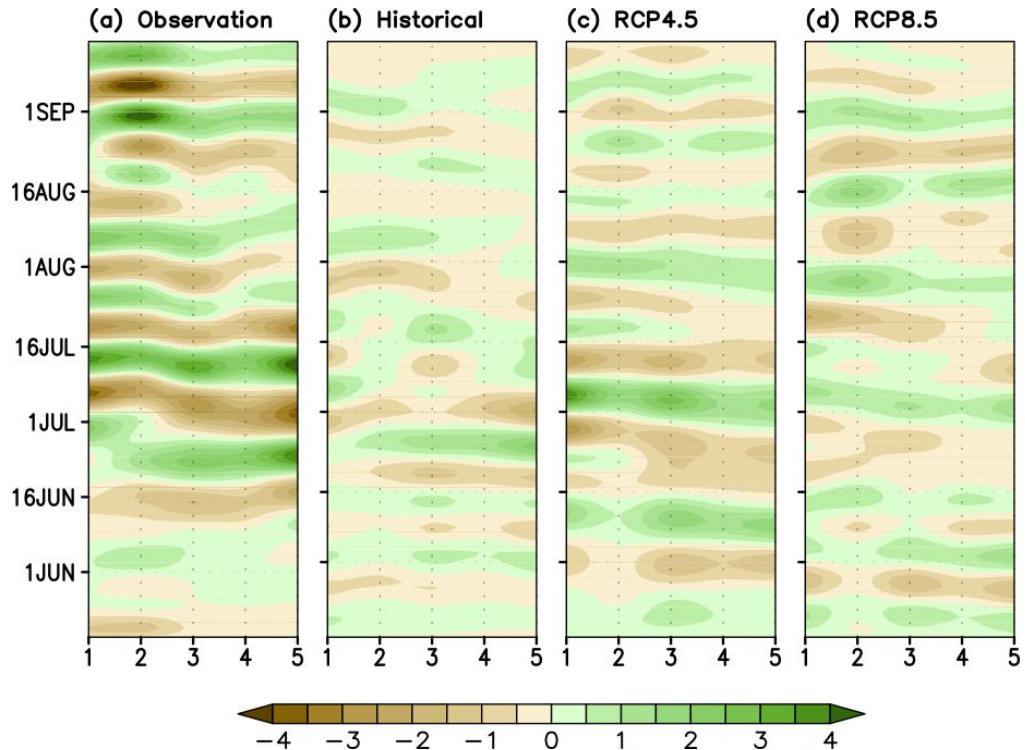


Figure 8. The band-pass filtered CSLV of the seasonal cycle with a cut-off periods of 10-30 days for each cluster group: (a) observation, (b) historical run, (c) RCP4.5 run, and (d) RCP8.5 run.

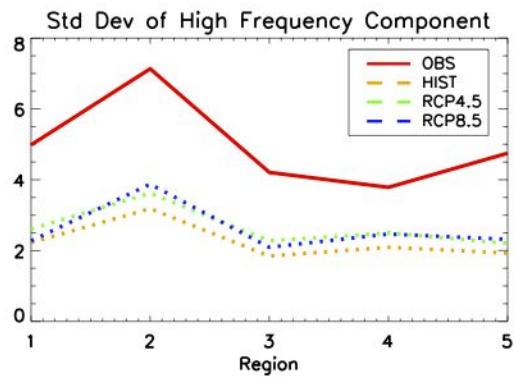


Figure 9. The standard deviation of the high-frequency (period less than 10 days) component of the seasonal cycle for each cluster group.

Figure 10 shows the PC time series of the first CSEOF mode (the seasonal cycle) derived from the observational data and the RCM datasets. In the figure, the PC time series of the seasonal cycle in the RCP 4.5 and RCP 8.5 datasets are put together with the PC time series of the historical run without any attempt to adjust the mean of the PC time series; there is a small difference in the mean of the PC time series. The PC time series of the historical run appears to exhibit similar spectral composition to that of the observational data. Likewise, the PC time series from the RCP 4.5 and RCP 8.5 also exhibit similar spectral compositions. In fact, the AR spectra of the PC time series from the RCM datasets are fairly similar to that of the observational data (Figure 11). There is a fairly broad peak at the period of ~3-4 years. The PC time series from the RCP 4.5 and RCP 8.5 runs indicate that the amplitude fluctuation of the seasonal cycle increases in future climate; standard deviation of the PC time series increases from 0.28 in the historical run to 0.34 in the RCP 4.5 run and 0.36 in the RCP 8.5 run.

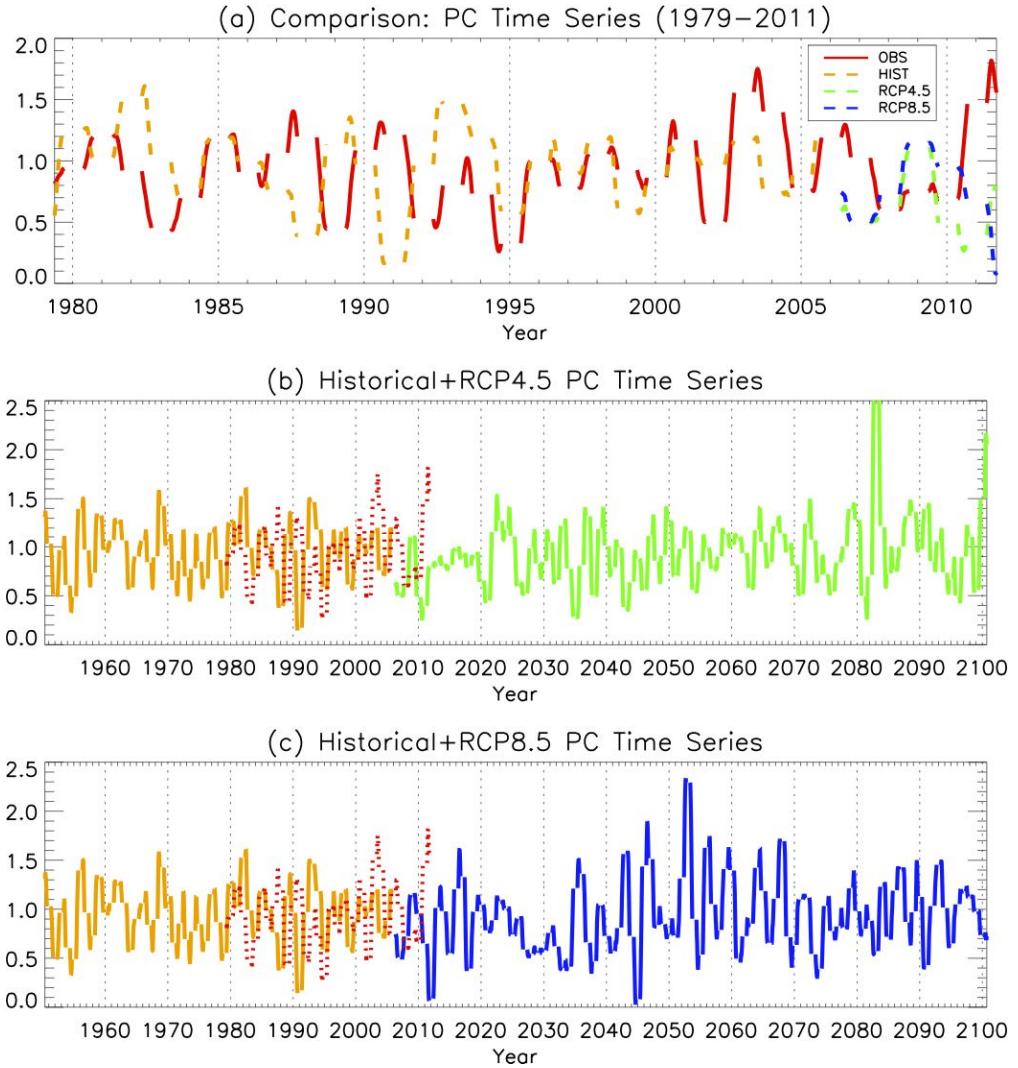


Figure 10. The PC time series of the first CSEOF mode (the seasonal cycle) from the KMA observational data (red), the historical run (orange), the RCP 4.5 run (green), and the RCP 8.5 run (blue). All the PC time series are normalized. In panels (b) and (c), the PC time series of the RCP 4.5 data and that of the RCP 8.5 data were simply connected with that of the historical run since RCP 4.5 and RCP 8.5 runs start from the historical run.

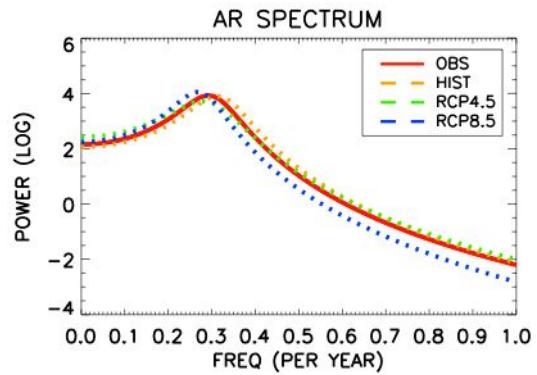


Figure 11. The AR spectra of the PC time series in Figure 10: observational data (red), historical run (orange), RCP 4.5 run (green), and RCP 8.5 run (blue).

Figure 12 shows the histogram of precipitation (frequency per year) as a function of precipitation intensity. The KMA observational data exhibit much more frequent occurrence of dry days than the RCM datasets (Table 2). On the other hand, precipitation intensity of (0,20) mm day⁻¹ in the observational data is much less frequent than in the RCM datasets (Tables 2 and 3; Figure 12). Precipitation intensity of [20,100) mm day⁻¹ is significantly underestimated in the model datasets; this is a major reason for the underestimation of the summer total precipitation by ~250 mm in the historical run. The frequency of extreme precipitation (over 100 mm day⁻¹) is also seriously underestimated although the occurrence of precipitation amount exceeding 500 mm day⁻¹ increased dramatically in future climates (Tables 2 and 3). On the other hand, no significant change is seen in the frequency of moderate rain events (less than 100 mm day⁻¹) in the future period.

The distribution of extreme precipitation exceeding 100 mm day⁻¹ was determined in terms of the generalized Pareto distribution (GPD):

$$f_{(\xi,\mu,\sigma)}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-(1+\xi)/\xi}, \quad (13)$$

where μ is the location parameter, σ is the scale parameter, and ξ is called the shape parameter. Table 4 summarizes the values of scale parameters and shaper parameters for the observational data and the model datasets. The location parameter defines the lower bound of extreme precipitations and is set to be 100 mm day⁻¹ in the present study. The scale parameter essentially defines the range of meaningful

extreme values. The change in the shape parameters as in Table 4 does not significantly alter the shape of distribution.

The scale parameter is slightly underestimated in the historical run; as a result, the middle section ($150\text{-}400 \text{ mm day}^{-1}$) of the distribution function is closer to the intensity (x) axis while the left section ($100\text{-}150 \text{ mm day}^{-1}$) is further away from the intensity axis than that of the observational data (Figure 13; Table 4). There is no significant difference between the GPD of the observational data and those of the model datasets in the near future (2030-2060). A significant increase in the scale parameters in the far future (2070-2100) is seen in the RCP 4.5 and RCP 8.5 runs; as a result, the middle section of the distribution function moves further away from the intensity axis. In other words, possibility of extreme events, particularly in the range of ($100,500$) mm day^{-1} , increases in future according to the RCM simulations of future climate with increased concentrations of greenhouse gases.

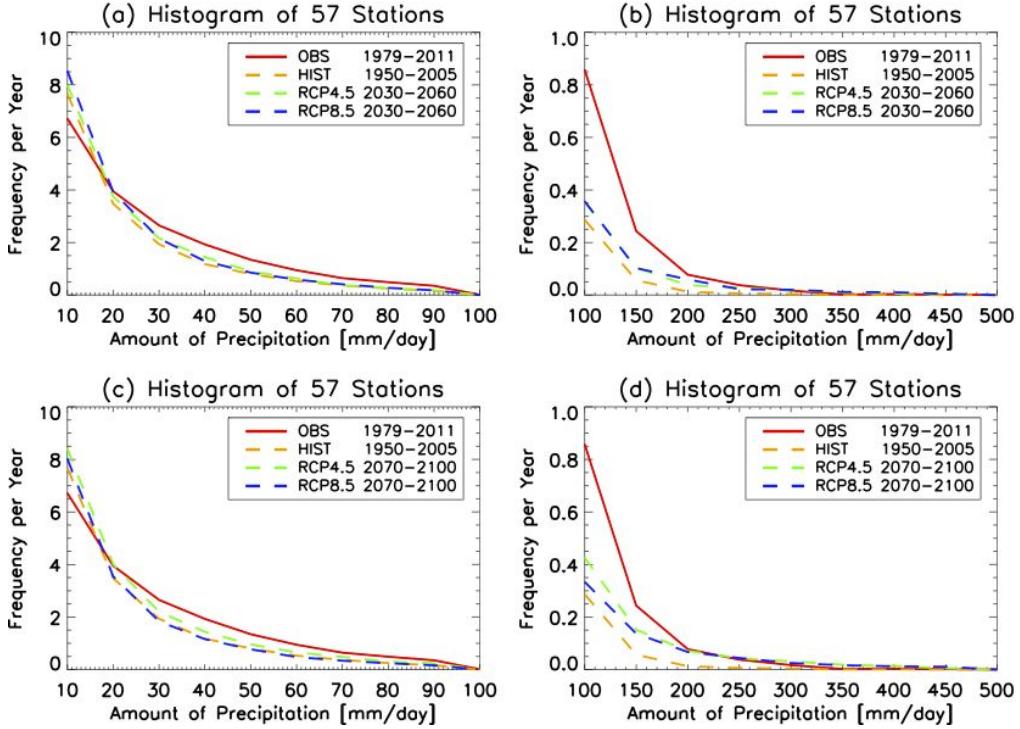


Figure 12. Histogram of precipitation for the amount less than 100 mm day^{-1} (left column) and for more than 100 mm day^{-1} (right column). The histograms based on the RCP 4.5 and RCP 8.5 runs represent the near future (2030-2060) in the upper panels and the far future (2070-2100) in the lower panels.

Table 2. Occurrence frequency per year in unit of days as a function of precipitation amount in the RCP 4.5 and the RCP 8.5 simulations for the period of 2030-2060 in comparison with the observational record (1979-2011) and the historical run (1950-2005).

	Near Future (2030-2060)				
	0	(0,10)	[10,100)	[100,500)	≥ 500
Observation	73.21	26.51	19.03	1.241	0.002
Historical run	46.54	56.72	16.38	0.367	0.000
RCP 4.5	45.90	55.82	17.71	0.562	0.009
RCP 8.5	46.57	54.61	18.22	0.589	0.016

Table 3. Occurrence frequency per year in unit of days as a function of precipitation amount in the RCP 4.5 and the RCP 8.5 simulations for the period of 2070-2100 in comparison with the observational record (1979-2011) and the historical run (1950-2005).

	Far Future (2070-2100)				
	0	(0,10)	[10,100)	[100,500)	≥ 500
Observation	73.21	26.51	19.03	1.241	0.002
Historical run	46.54	56.72	16.38	0.367	0.000
RCP 4.5	44.78	55.73	18.70	0.761	0.026
RCP 8.5	49.18	53.58	16.56	0.646	0.035

Table 4. The scale and the shape parameters of the GPD distribution for the observational record, the historical run, the RCP 4.5 run, and the RCP 8.5 run. The two scale and the shape parameters represent the period of 2030-2060 and the period of 2070-2100, respectively.

	Scale parameter		Shape parameter	
Observation	39.57		0.116	
Historical run	29.35		0.164	
RCP 4.5	43.39	57.88	0.290	0.186
RCP 8.5	43.77	66.28	0.343	0.149

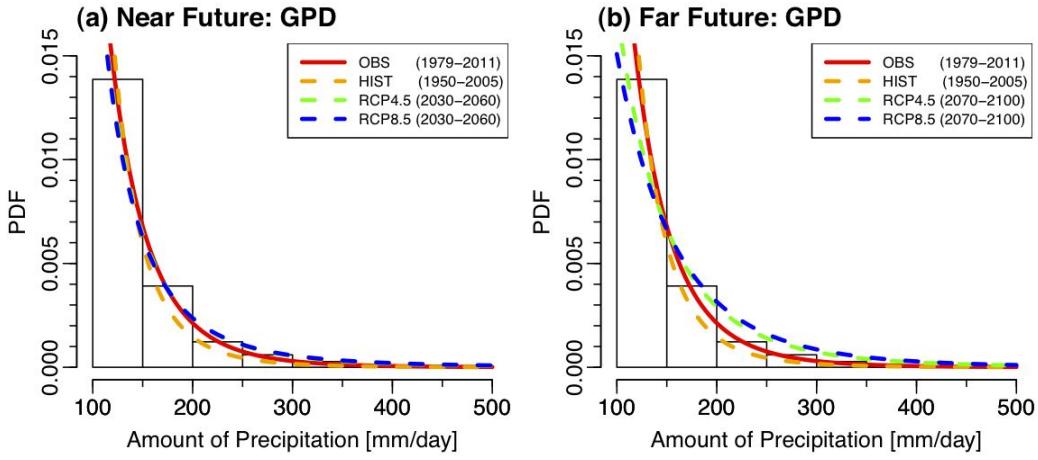


Figure 13. The generalized Pareto distribution (GPD) of the extreme events with precipitation amount exceeding 100 mm day^{-1} . The RCP 4.5 and RCP 8.5 runs represent the near future (2030–2060) in the left column and the far future (2070–2100) in the right column.

As already mentioned above, the frequency of rain day and that of precipitation in the range of $(0,10)$ mm day $^{-1}$ do not exhibit significant differences among the different scenarios. Figure 14 represents the frequency per year of dry days and precipitation events in the range of $(0,10)$ mm day $^{-1}$ for the observational data and the historical run. Since precipitation statistics are strongly time dependent, summer 120 days are divided into 8 periods of 15 days each. The frequency of dry days is higher prior to the first rain spell and after the second rain spell according to the observational data (Figure 14a). On average, there are \sim 7-11 no-rain days in each 15-day interval in the observational record whereas \sim 4-8 no-rain days in each interval in the historical run. The frequency of mild rain days with precipitation in the range of $(0,10)$ mm day $^{-1}$, on the other hand, is higher in the middle of summer (Figure 14b). While the temporal evolution patterns are similar, the frequency of rain days and that of mild rain events are significantly overestimated in the historical run.

The frequencies of more intense rain events for each of the 15-day interval are shown in Figures 15 and 16. While the maximum peak in the observational record occurs in the first half of July in consistence with the seasonal cycle, there is a secondary peak in the second half of August, resulting in a bimodal distribution structure both in the $[10,100]$ mm day $^{-1}$ and the $[100,500]$ mm day $^{-1}$ ranges. The historical run, on the other hand, shows that rain events dominate in July. In particular, there is a significant gap between the observational data and the historical run in the frequency of extreme precipitations of $[100,500]$ mm day $^{-1}$. It appears that the occurrence frequency of $[10,100]$ mm day $^{-1}$ precipitation events does not

significantly change (Figures 15a and 16a) whereas the frequency of [100,500) mm day⁻¹ precipitation events increases somewhat in future climates particularly in July (Figures 15b and 16b). The RCP 4.5 and RCP 8.5 simulations produce rain events, in which precipitation exceeds 500 mm day⁻¹; such an event does not exist in the historical run.

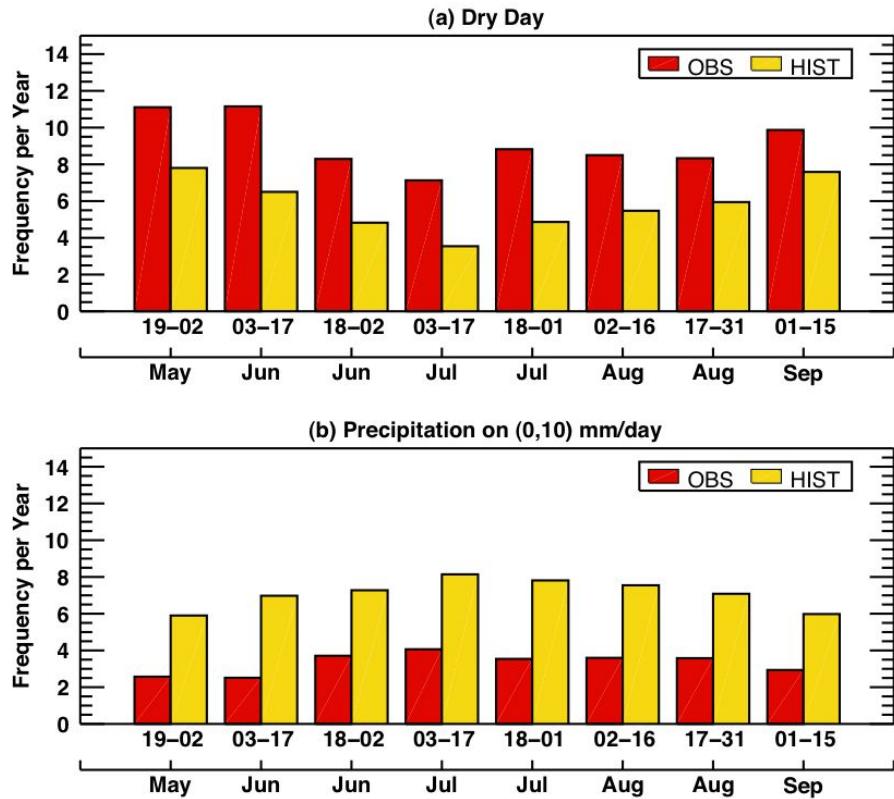


Figure 14. The occurrence frequency of dry days (upper panel) and precipitation events of (0,10) mm day⁻¹ (lower panel) for the eight subintervals of 15 days each during summer. The red bars denote the observational data, and the yellow bars denote the RCM historical run.

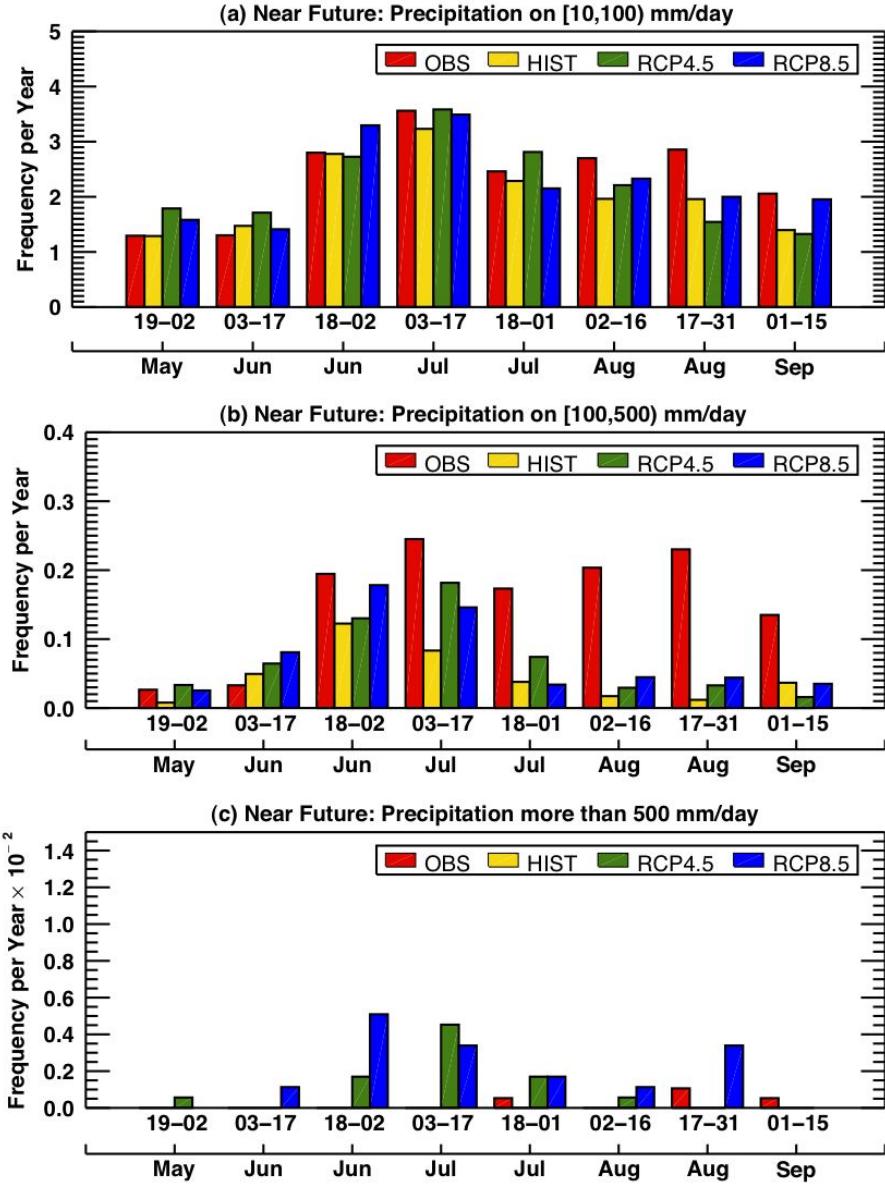


Figure 15. The occurrence frequency of precipitation for the amount of [10,100) mm day⁻¹ (upper panel), for the amount of [100,500) mm day⁻¹ (middle panel), and for the amount over 500 mm day⁻¹ (bottom panel). The red bars denote the observational data, the yellow bars the historical run, the green bars the RCP 4.5 run, and the blue bars the RCP 8.5 run. The RCP 4.5 and RCP 8.5 results are based on the data in the near future (2030-2060).

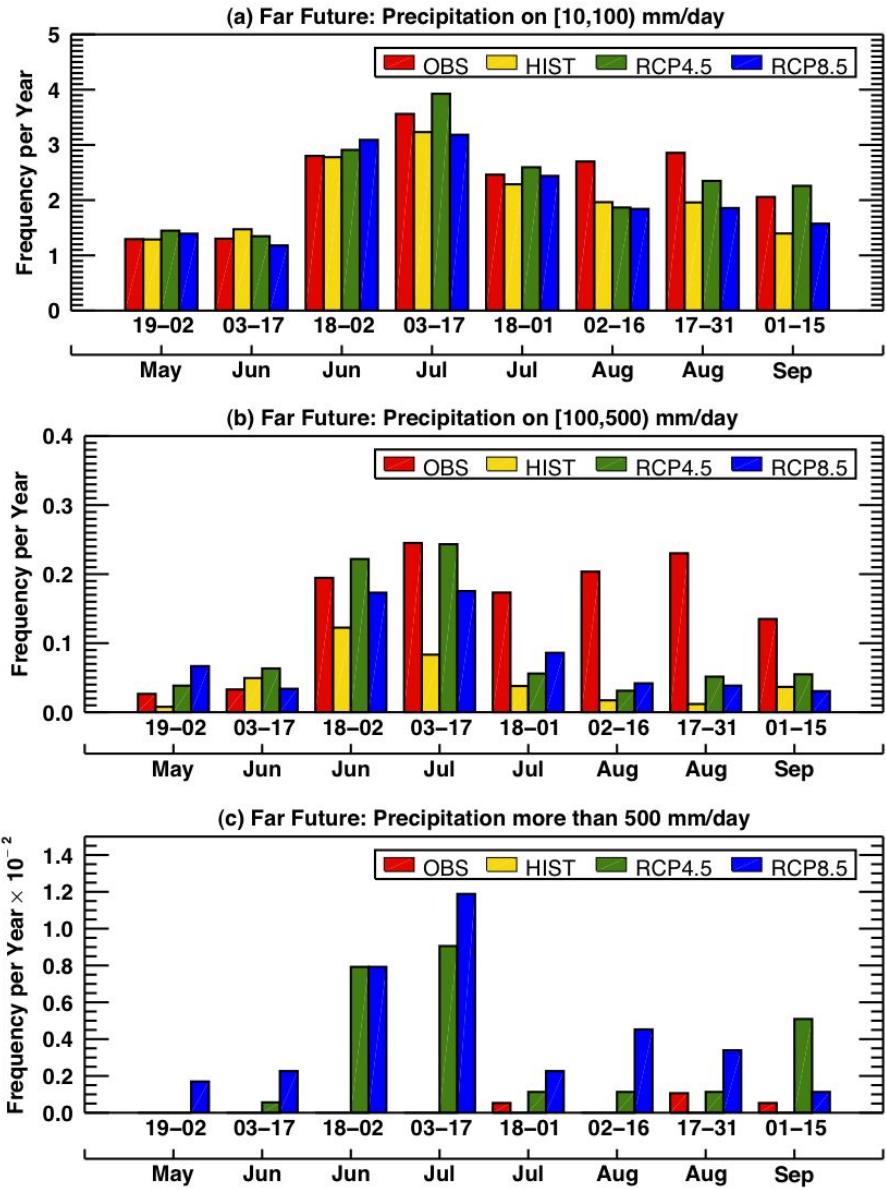


Figure 16. The occurrence frequency of precipitation for the amount of [10,100) mm day⁻¹ (upper panel), for the amount of [100,500) mm day⁻¹ (middle panel), and for the amount over 500 mm day⁻¹ (bottom panel). The red bars denote the observational data, the yellow bars the historical run, the green bars the RCP 4.5 run, and the blue bars the RCP 8.5 run. The RCP 4.5 and RCP 8.5 results are based on the data in the far future (2070-2100).

Figures 17 shows the spatial correlation function, $C(r, r')$, of the RCM historical run in comparison with that of the KMA observational data. The spatial correlation structure is rather similar although the spatial correlation scale is slightly overestimated in the historical run. As can be seen in the spatial correlation pattern with respect to the Seoul station, the historical run exhibits a larger spatial correlation scale than the observational data. Figure 18 shows the spatial correlation functions of the RCP 4.5 and RCP 8.5 runs. The spatial correlation function is noisier than that of the historical run and indicates the reduction of the spatial correlation scale. The spatial correlation pattern with respect to the Seoul station also exhibits the reduction of spatial correlation scale. This seems to indicate that extreme precipitation events in the RCP 4.5 and RCP 8.5 runs tend to be rather sporadic in space affecting only a few neighboring stations.

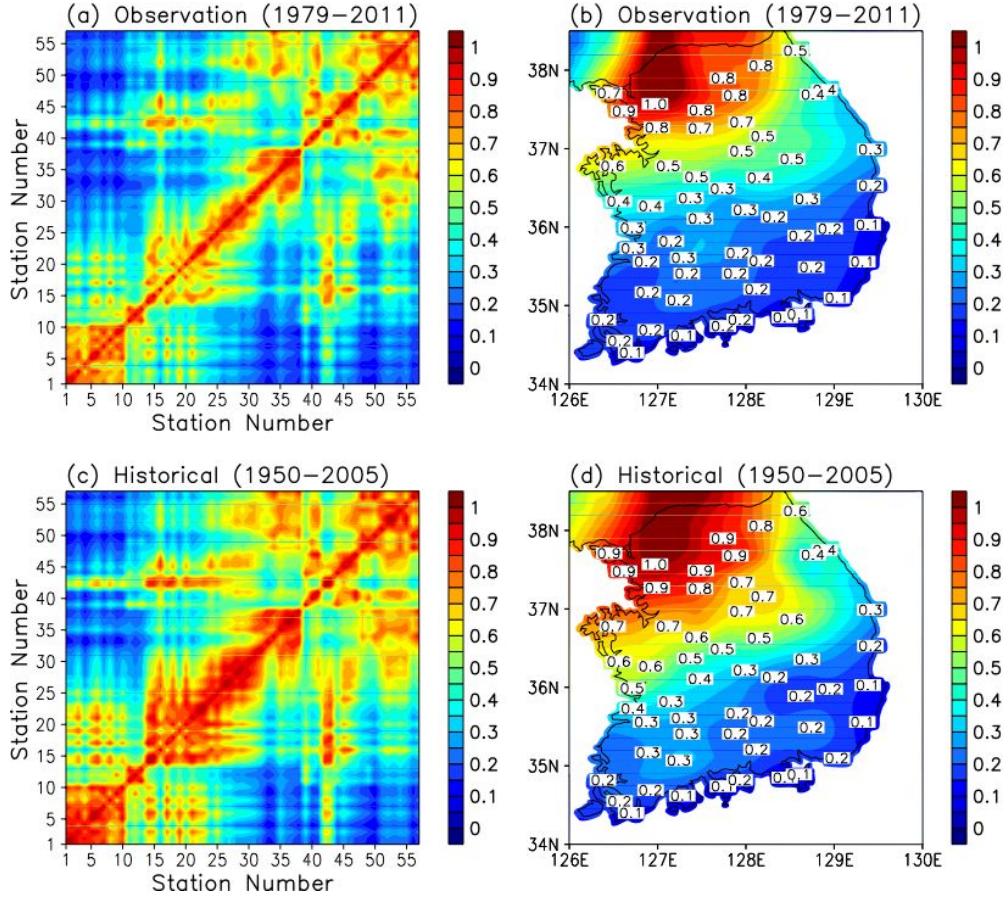


Figure 17. (left column) Spatial correlation function, $C(r, r')$, of the observational data with extreme events at 57 KMA stations (upper panel), and spatial correlation function of the historical run at 57 grid points in Figure 5 (bottom panel). (right column) Spatial correlation function with respect to the Seoul station.

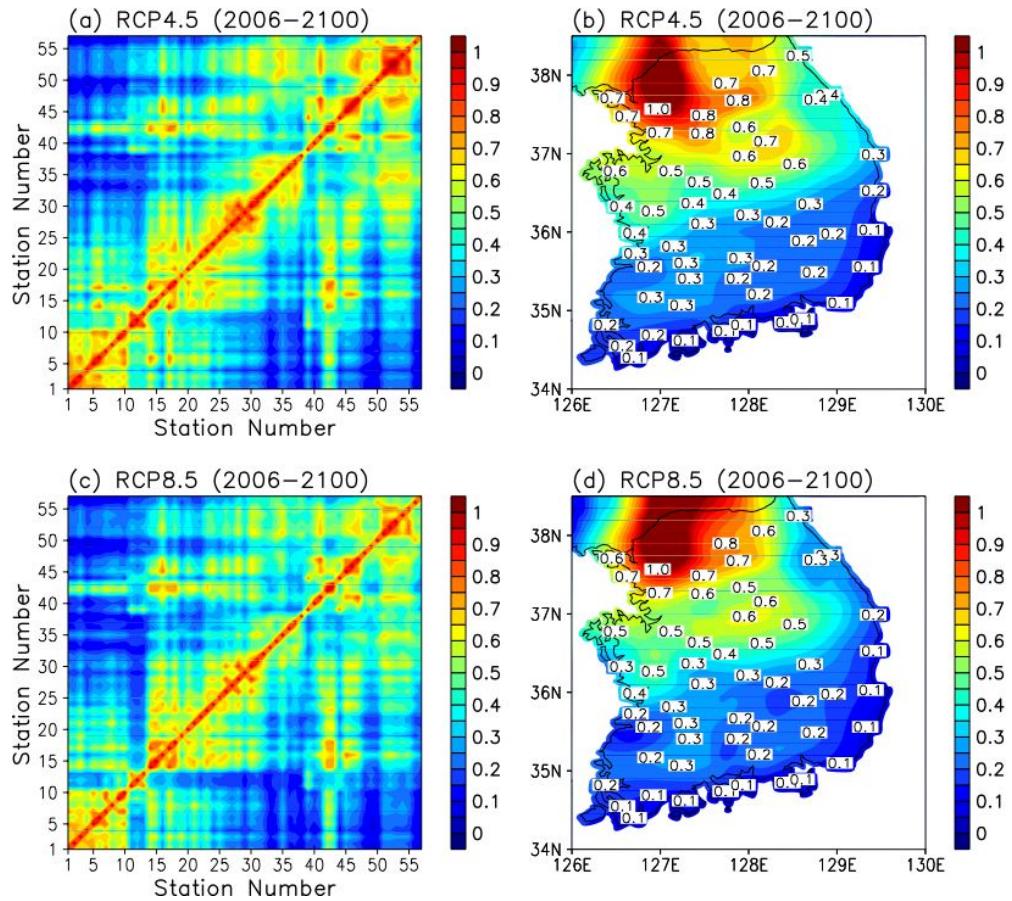


Figure 18. (left column) Spatial correlation function of the RCP 4.5 run at the 57 grid points in Figure 5 (upper panel), and spatial correlation function of the RCP 8.5 run at 57 grid points (bottom panel). (right column) Spatial correlation function with respect to the Seoul station.

5. Comparison of KMA Data with the Synthetic Datasets

Figure 19 shows the 100 synthetic PC time series generated from the corresponding ARMA model in comparison with the 1st PC time series of the KMA observational data during the observational period; gaps in the plot indicate that CSEOF analysis was conducted on the summertime precipitation. The amplitude of the PC time series of the observational data falls, in general, within the spread of the 100 synthetic PC time series. The synthetic PC time series slightly overestimate the mean amplitude of the seasonal cycle in the observational record. The standard deviations of the 100 synthetic PC time series match closely with that of the PC time series of the observational record.

Figures 20 and 21 show the synthetic PC time series and their means and standard deviations for the second and the third CSEOF mode, respectively. It appears that the second and the third PC time series were reasonably reproduced with realistic means and standard deviations based on the ARMA modeling. In this way, 100 sets of synthetic PC time series of the first 20 PC time series of the observational data were reproduced based on their ARMA models and their means and standard deviations were examined in comparison with those of the PC time series of the observational data. The histograms of the synthetic time series also appear to be reasonable as shown in Figure 22.

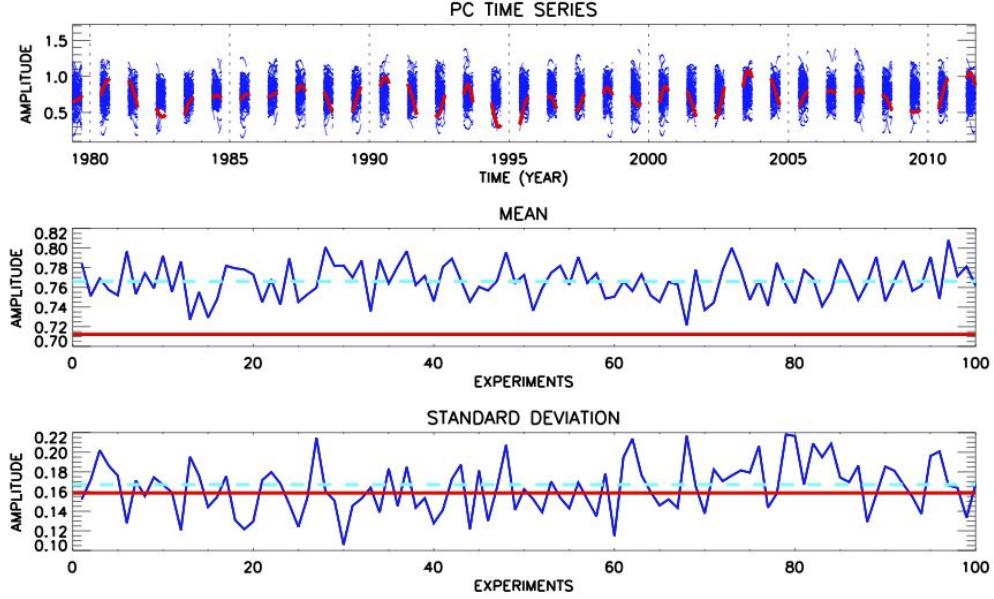


Figure 19. (top) Comparison of the first PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the first PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) in comparison with the standard deviation of the first PC time series from the observational record (red).

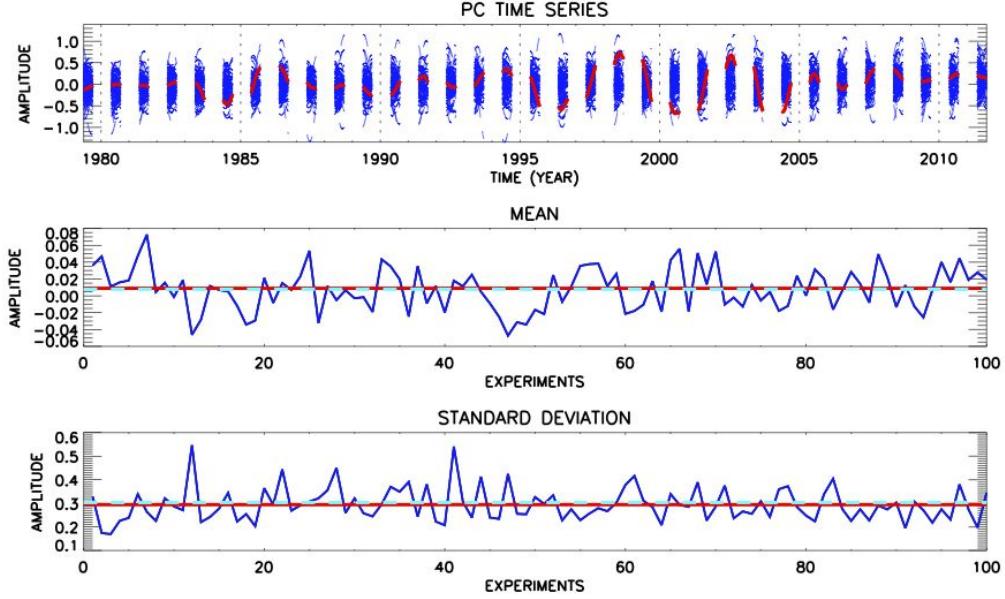


Figure 20. (top) Comparison of the second PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the second PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) in comparison with the standard deviation of the second PC time series from the observational record (red).

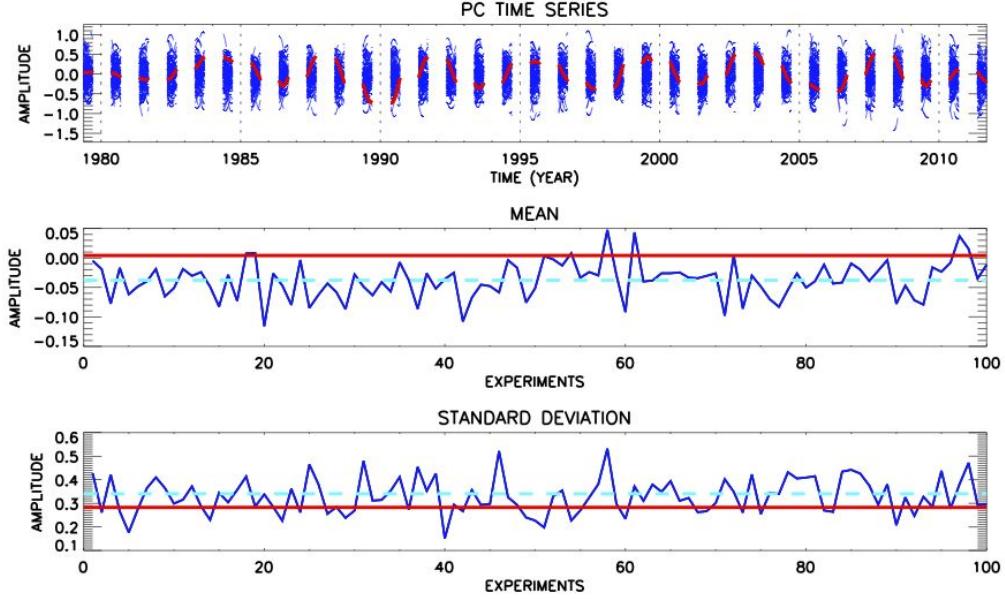


Figure 21. (top) Comparison of the third PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of the observational record. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series in comparison with the mean of the third PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) in comparison with the standard deviation of the third PC time series from the observational record (red).

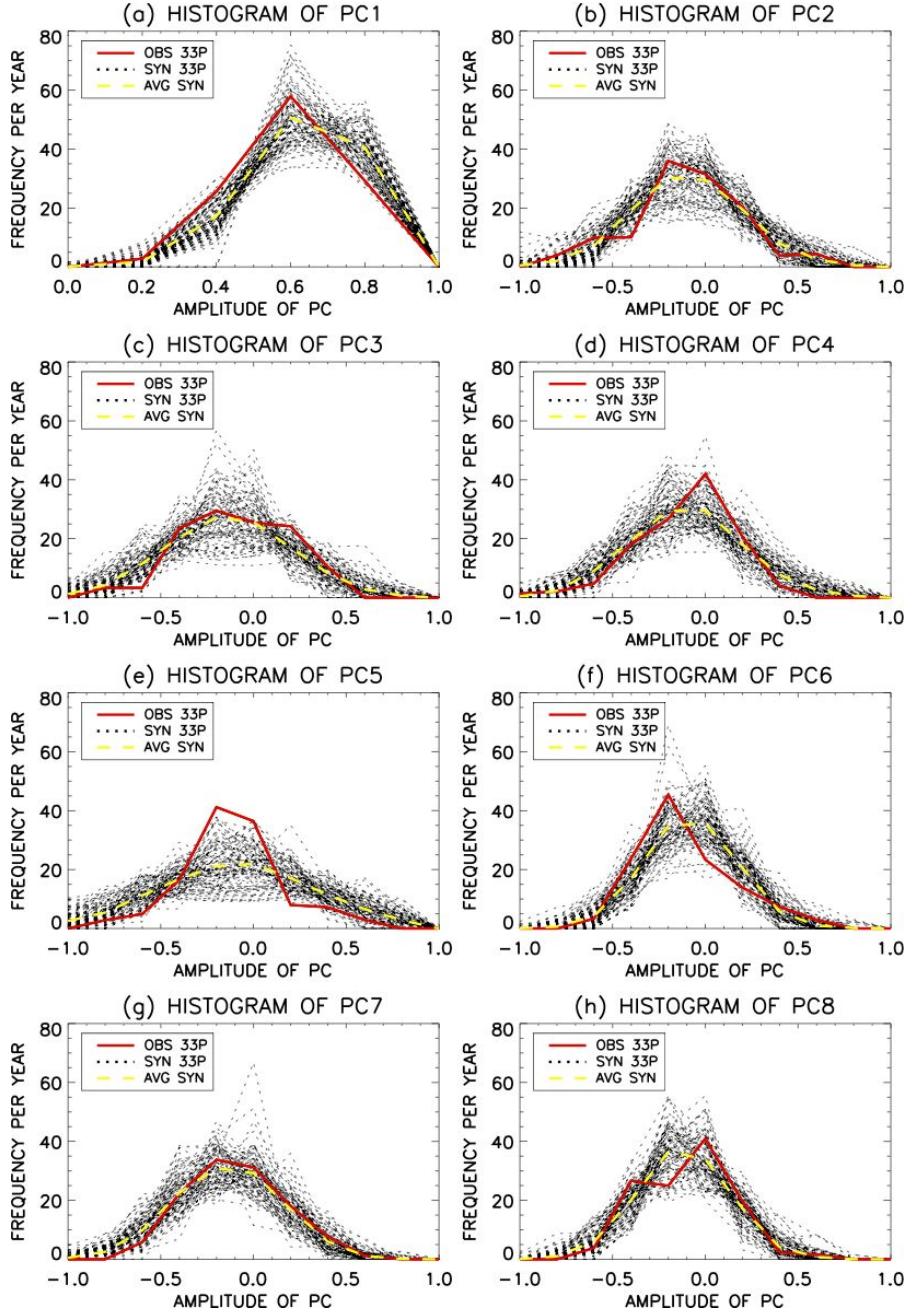


Figure 22. Histograms of the first eight PC time series of the observational data (red), and those of the 100 synthetic PC time series (black) with their mean histogram (yellow) for the observational period (1979-2011).

Figures 23-25 show the synthetic time series of the first three PC time series extended until 2061 and the means and the standard deviations over the period of 2029-2061. In future, the mean of the first PC time series increased by ~20% whereas the standard deviations remained nearly the same. The mean of the second PC time series increased significantly while the standard deviation remained almost the same. The mean of the third PC time series decreased significantly and the standard deviation increased slightly. These changes are primarily due to the linear trend in the PC time series of the observational data. Note that the linear trend in each PC time series is added back to the synthetic time series; the linear trend amplifies in time as can be seen in Figures 23 and 24. The histograms of the PC time series also change appreciably as can be seen in Figure 26. The lower ends of the histogram shift toward higher values for the first PC time series resulting in increased means of the synthetic PC time series in future climate. The histograms of the second PC time series shift toward higher values resulting in increased means of the synthetic time series in future.

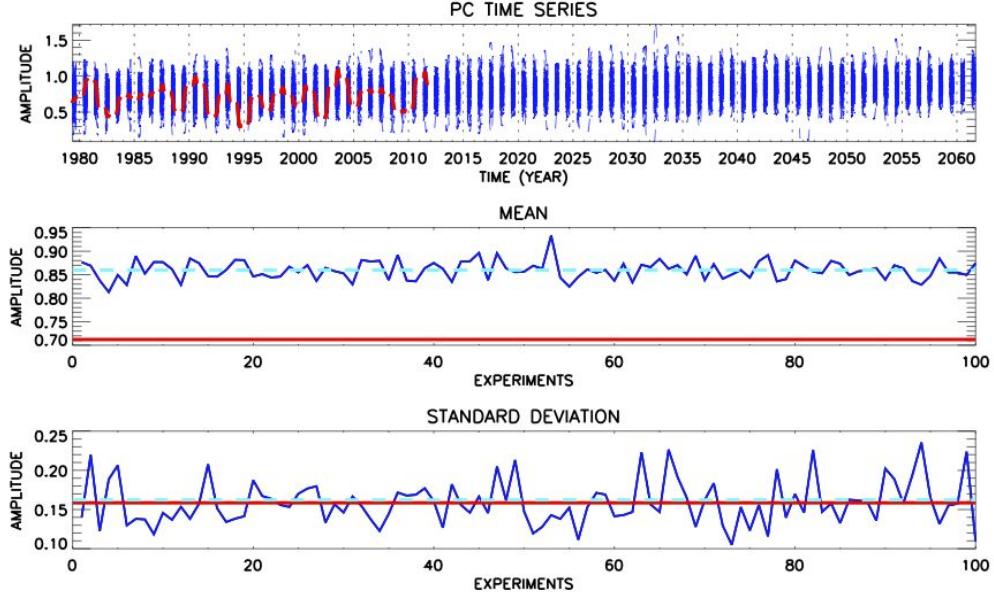


Figure 23. (top) Comparison of the first PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the first PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the first PC time series from the observational record (red).

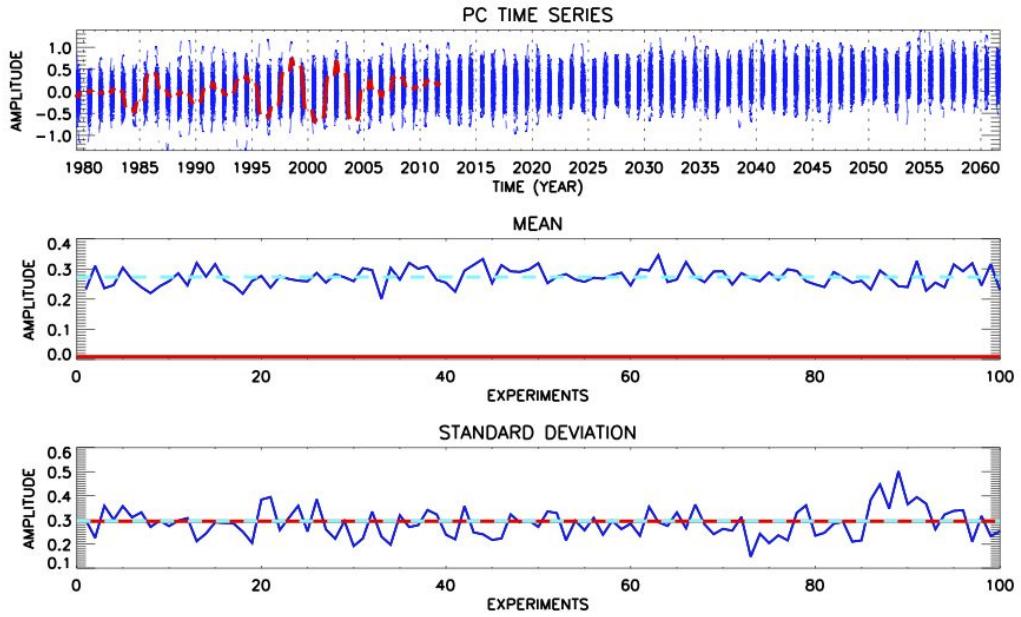


Figure 24. (top) Comparison of the second PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the second PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the second PC time series from the observational record (red).

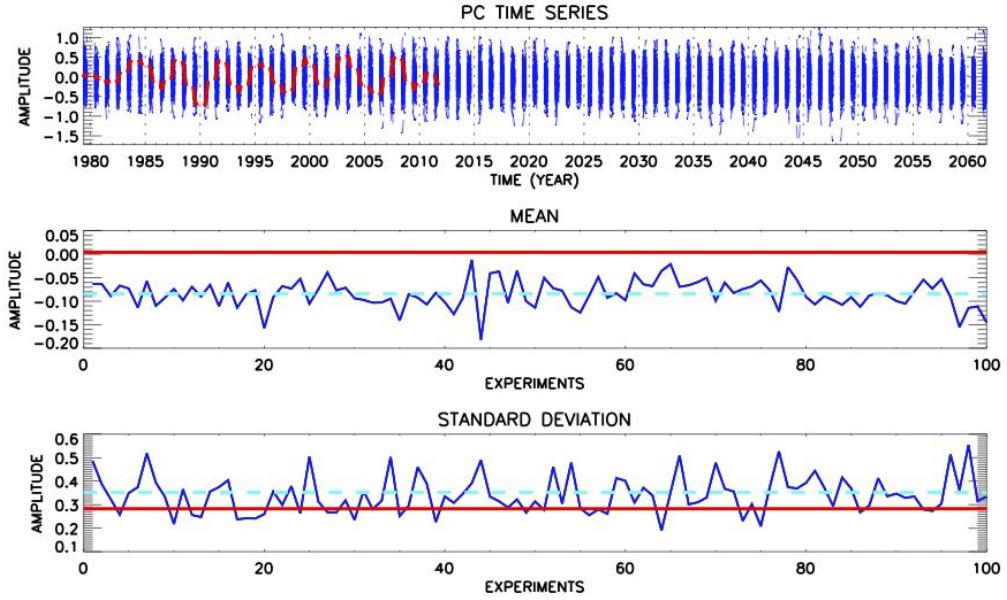


Figure 25. (top) Comparison of the third PC time series of the KMA observational data (red curve) and 100 synthetic time series (blue curves) generated as described in (9) over the period of 1979-2061. (middle) The mean of each synthetic time series (blue curve) and the mean (cyan) of the 100 synthetic PC time series over the period of 2029-2061 in comparison with the mean of the third PC time series from the observational record (red). (bottom) The standard deviation of each synthetic time series (blue curve) with their mean (cyan) over the period of 2029-2061 in comparison with the standard deviation of the third PC time series from the observational record (red).

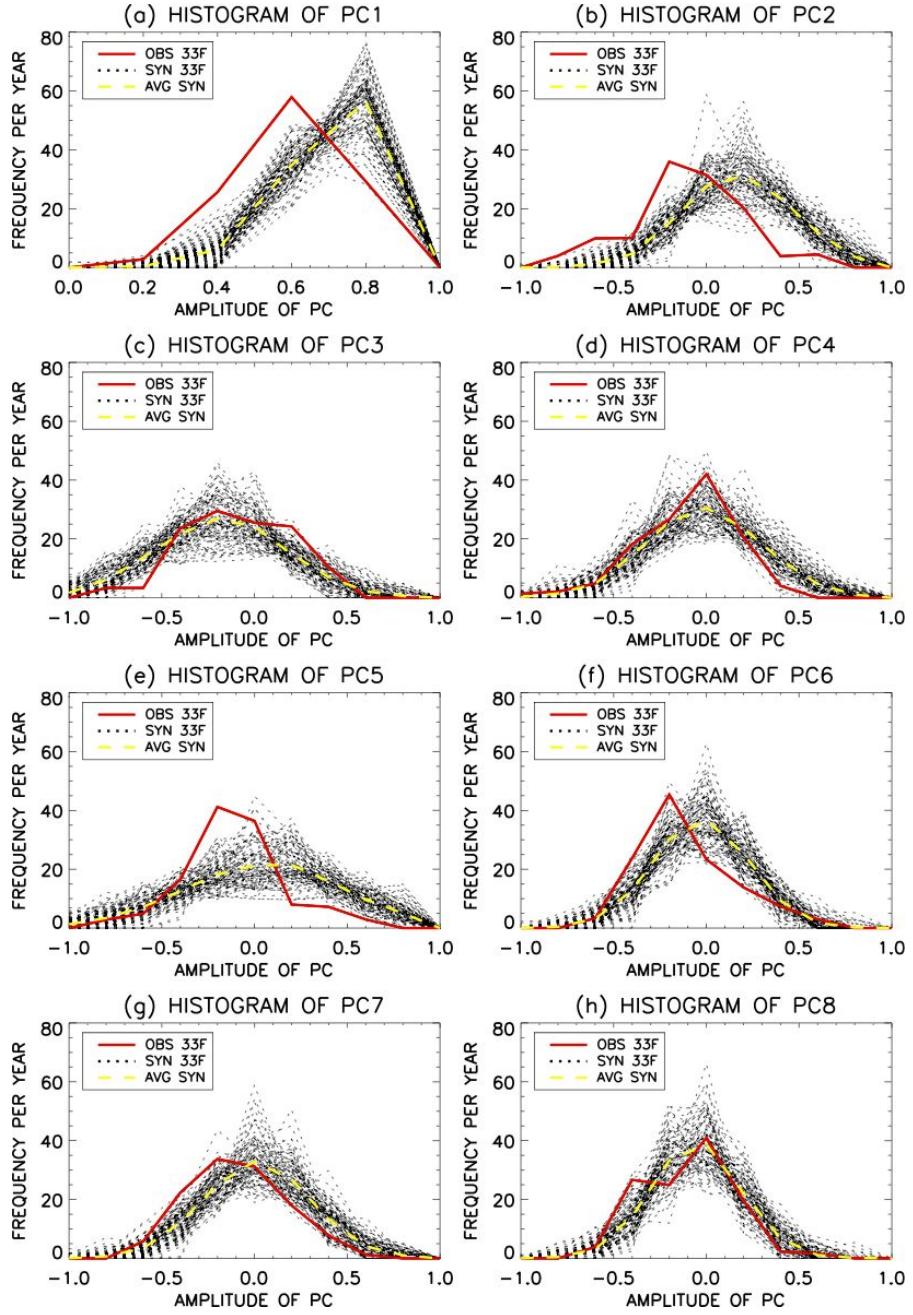


Figure 26. Histograms of the first eight PC time series of the observational data (red), and those of the 100 synthetic PC time series (black) with their mean histogram (yellow) for the future period (2029-2061).

Based on the synthetic PC time series, synthetic datasets of precipitation were generated according to (10). In constructing the synthetic datasets, it was assumed that CSLVs extracted from the observational data would not change in future; this, of course, is an important caveat in the present study. Based on the synthetic datasets of “no-extreme” precipitation, spatial correlation function was computed for the observational period as in Figure 27. Spatial correlation function based on the 20-mode reconstruction of synthetic precipitation matches well with that of the 20-mode truncated observational data. Spatial correlation function with respect to the Seoul station, for example, is accurately reproduced in the synthetic precipitation data. This map demonstrates that the CSEOF-based approach of generating synthetic precipitation is excellent in faithfully reproducing the spatial correlation function in the observational record.

Figure 28 shows the patterns of mean and standard deviation of the seasonal cycle derived from the observed precipitation and those of the 100 synthetic precipitations for the observational period (1979-2011) and for future (2029-2061). The model slightly underestimates or overestimates the mean and standard deviation of the seasonal cycle for the observational period. In future, both the mean and the standard deviation increase significantly at all stations over the Korean Peninsula; in particular, increase in the mean and the standard deviation is most conspicuous in the northwestern part of the peninsula.

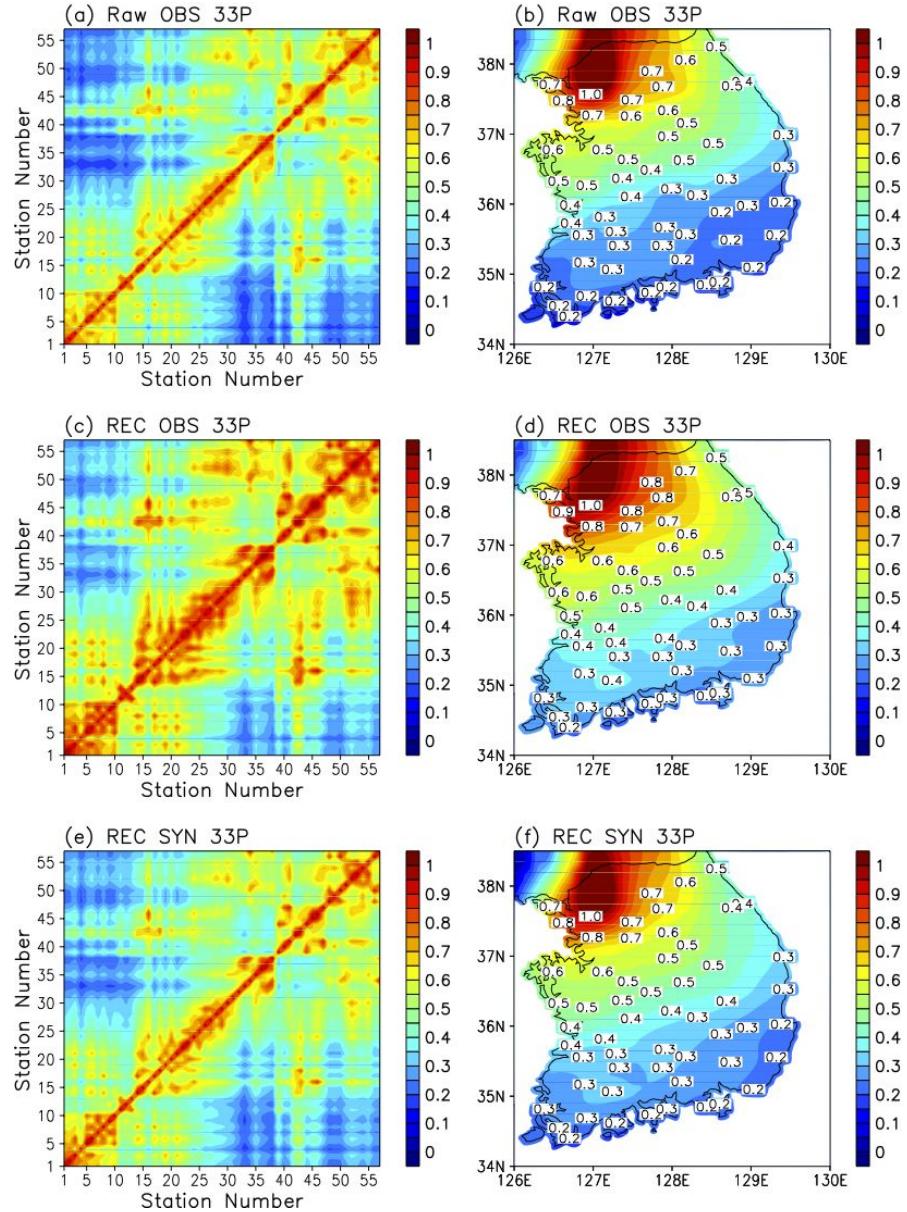


Figure 27. (left column) Spatial correlation function of the observational data without any extreme events at 57 KMA stations (top), spatial correlation function of the 20-mode truncated observational data without any extreme events at 57 KMA stations (middle), and spatial correlation function of the 100 synthetic datasets at 57 stations in Figure 5 (bottom). (right column) Spatial correlation function with respect to the Seoul station.

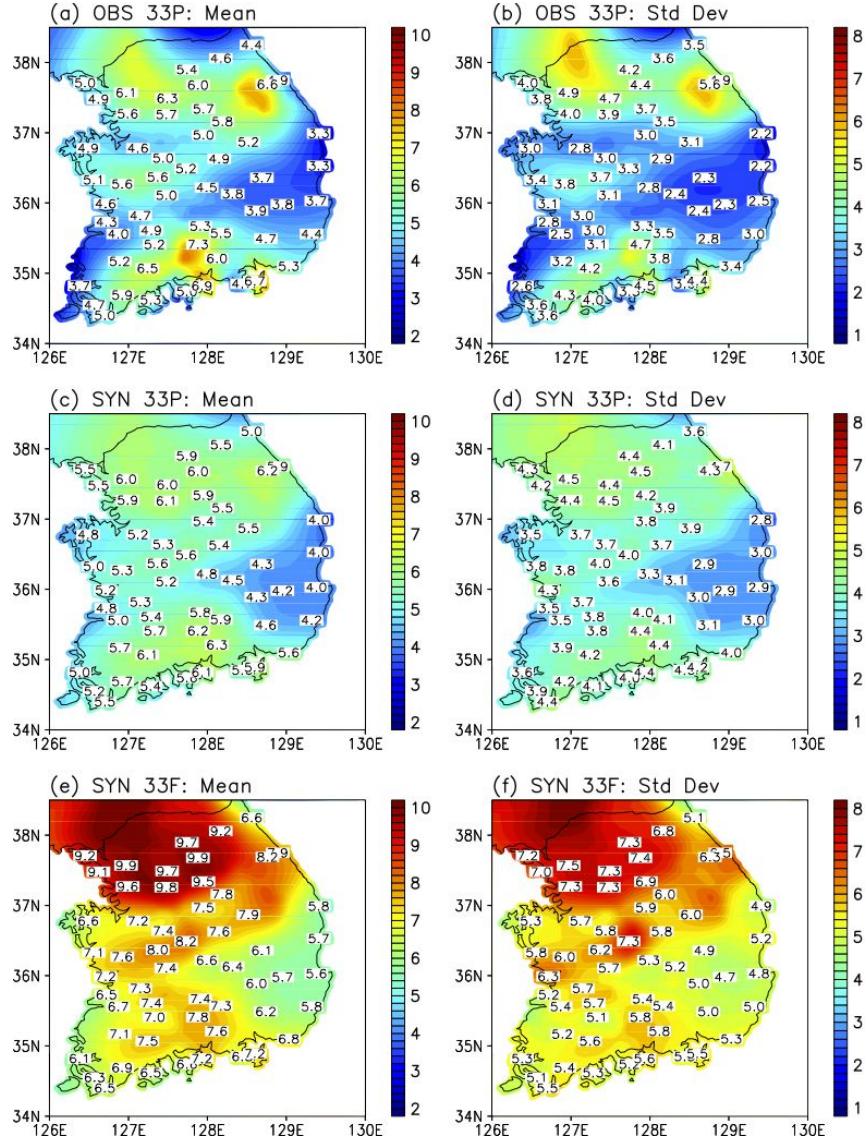


Figure 28. Mean (left column) and standard deviation (right column) of the seasonal cycle derived from the observational data with extreme events (top), mean of the seasonal cycle based on the 100 synthetic datasets over the same period of the observational data (middle), and mean of the seasonal cycle based on the 100 synthetic datasets over the period of 2029-2061.

Figure 29 shows the seasonal cycle averaged at the 57 KMA stations and those from the 100 synthetic datasets for the observational period (1979-2011) and for future (2029-2061). The seasonal cycles from the synthetic datasets for the observational period look very similar to that of the observational data. The amplitude of the seasonal cycle increases by more than 50% in future climate; the range of fluctuations also increases significantly. A comparison shows that both the first and the second peaks of precipitation amplify in future. In future, both the first and the second peaks of precipitation occur earlier than in the present record. This result is consistent with *Roh et al.* [2012] in that the first peak of precipitation amplified and appeared earlier in recent observational record; this result differs from those of *Roh et al.* [2012], who reported that the second peak weakens and are delayed according to the trend in the observational record.

The temporal evolution of the low-passed seasonal cycle with cutoff period of 30 days is depicted in Figure 30 for each of the five cluster groups. In the observational record, the first spell of precipitation occurs earlier in region 5, which is the southern part of the peninsula; it is well known that rain band (front) migrates northward during the early stage of East Asian summer monsoon reaching the southern part of the peninsula earlier. It is worthwhile to mention that the seasonal cycle is stronger in regions 1, 3 and 5, which represent the western part of the peninsula. On the other hand, the second peak is strongest in region 2, which is the northeastern part of the peninsula. The synthetic precipitation data show a reasonably similar pattern of the seasonal cycle although the second peak appears a little earlier than in the

observational data. In future climate, the seasonal cycle looks remarkably different from that in the present climate, in particular in the northern part of the peninsula. A period of weakened precipitation between the two spells of precipitation is obvious in all regions except for region 1. This result is consistent with Figure 9 in *Roh et al.* [2012], which shows that the period of the weakened precipitation is not clear in recent years at northern stations.

Figure 31 shows the intraseasonal component with periods of 10-30 days of the seasonal cycle. The synthetic datasets appear to underestimate the intraseasonal component of the seasonal cycle in the observational data. The intraseasonal component in the future period appears to be comparable in magnitude to that of the observational data. The high-frequency component with periods less than 10 days is shown in Figure 32. The synthetic datasets in the observational period slightly underestimates the standard deviation of the high-frequency variability particular in region 2. In the future period, on the other hand, the high frequency component of variability in the synthetic datasets is significantly higher than that of the observational data except for regions 2 and 5.

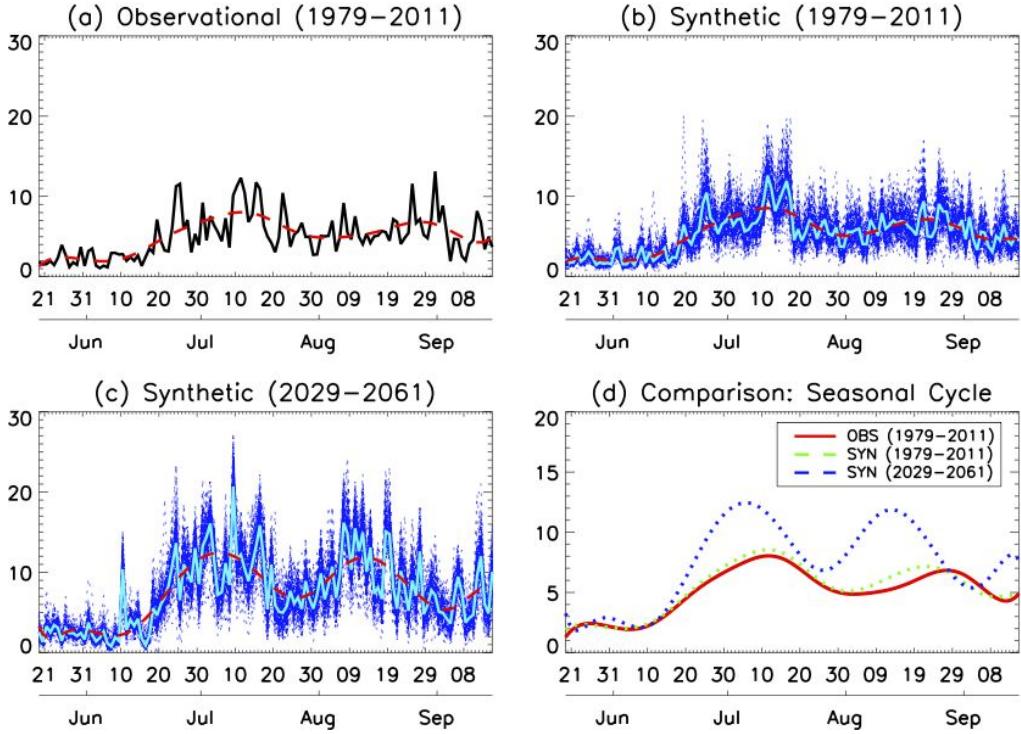


Figure 29. The temporal evolution of the seasonal cycle in (a) the observational data, (b) the 100 synthetic datasets over the same period of the observational record (1979–2011), (c) the 100 synthetic datasets over the period of 2029–2061, and (d) all three low-passed seasonal cycles in (a)–(c). The cyan curves in (b) and (c) are the averages of all the synthetic seasonal cycles while the red dashed curve is the seasonal component (period > 30 days) of the average seasonal cycle.

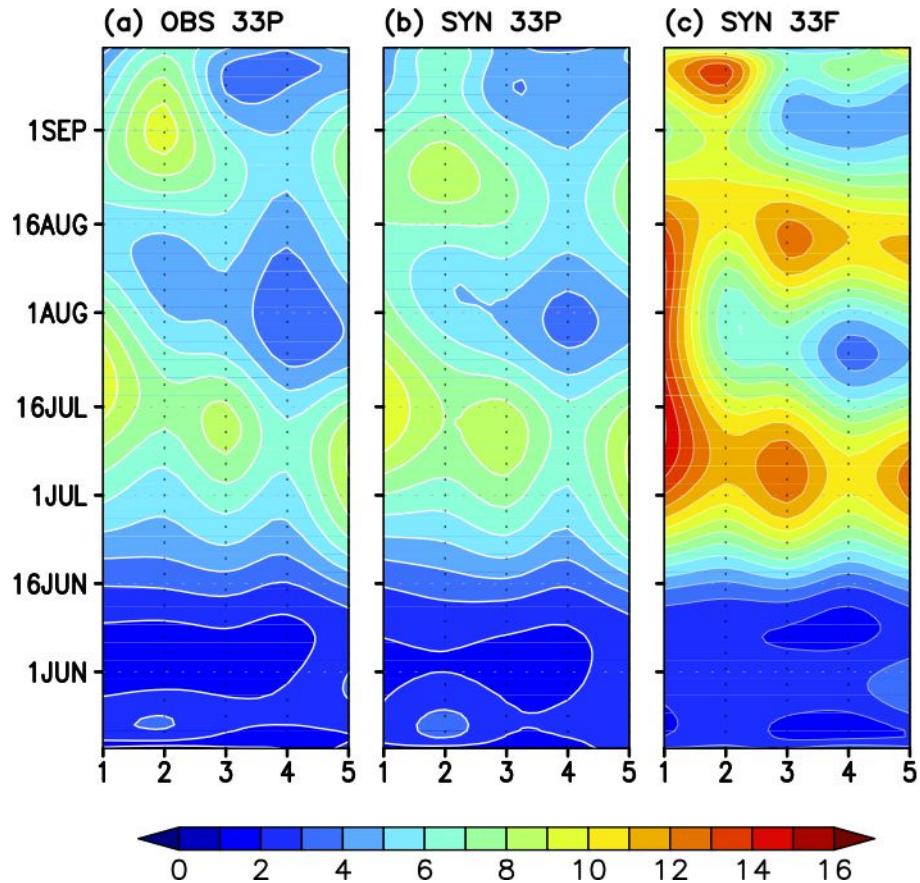


Figure 30. The temporal evolution of the low-passed seasonal cycle with the cutoff period of 30 days in each of the five cluster groups in (a) the observational record, (b) the 100 synthetic datasets over the same period of the observational record (1979-2011), and (c) the 100 synthetic datasets over the period of 2029-2061.

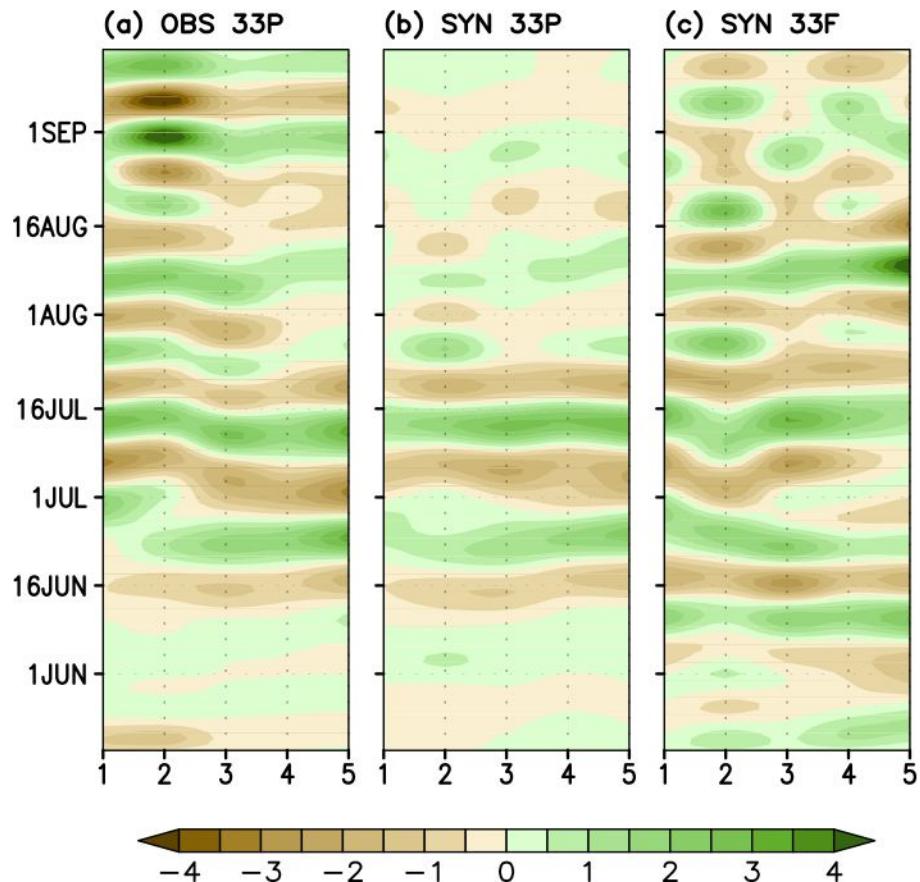


Figure 31. The intraseasonal component (period of 10-30 days) of the seasonal cycle for each of the five cluster groups in (a) the observational record, (b) the 100 synthetic datasets over the same period of the observational record (1979-2011), and (c) the 100 synthetic datasets over the period of 2029-2061.

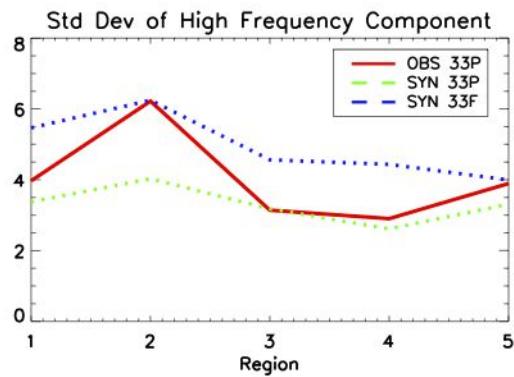


Figure 32. Standard deviation of the high-frequency component with period less than 10 days of the seasonal cycle for each of the five cluster groups in the observational data (red), the 100 synthetic datasets over the same period of the observational record (green), and the 100 synthetic datasets over the period of 2029–2061 (blue).

The histograms of synthetic precipitation data are shown in Figure 33 in comparison with those of the KMA observational data. In the observational period, the histograms of the synthetic data are fairly similar to that of the observational data both in the range of no severe rainfalls and in the range of extreme rainfalls; the dashed curve in cyan represents the averaged histograms from the 100 synthetic datasets and agrees well with the histogram from the observational data (red curve). In the future period (2029-2061), histograms of synthetic data depict more frequent occurrence of precipitation events greater than 10 mm day^{-1} than the observational data (see also Table 5). On the contrary, occurrence frequency of precipitation less than 10 mm day^{-1} decreases significantly in future.

Figure 34 shows the occurrence frequency of dry days and that of precipitation of $(0,10) \text{ mm day}^{-1}$ for the eight subintervals of 15 days each during summer. The occurrence frequency of dry days increases slightly in the synthetic datasets; there is not much difference between the observational period (1979-2011) and the future period (2029-2061). The occurrence frequency of precipitation of $(0,10) \text{ mm day}^{-1}$, on the other hand, decreased in the synthetic datasets particularly in the future period (Table 5). Figure 35 shows that the occurrence frequency of precipitation greater than 10 mm day^{-1} , in general, increases significantly in the future period particularly in July and August. Events with precipitation greater than 500 mm day^{-1} are rare and may not be significantly reliable. Figures 33 and 34 indicate that the number of moderate-to-heavy ($> 10 \text{ mm day}^{-1}$) rain events increases significantly in future while the number

of weak ($< 10 \text{ mm day}^{-1}$) precipitation events decreases significantly. The frequency of dry days remains nearly the same in the synthetic datasets.

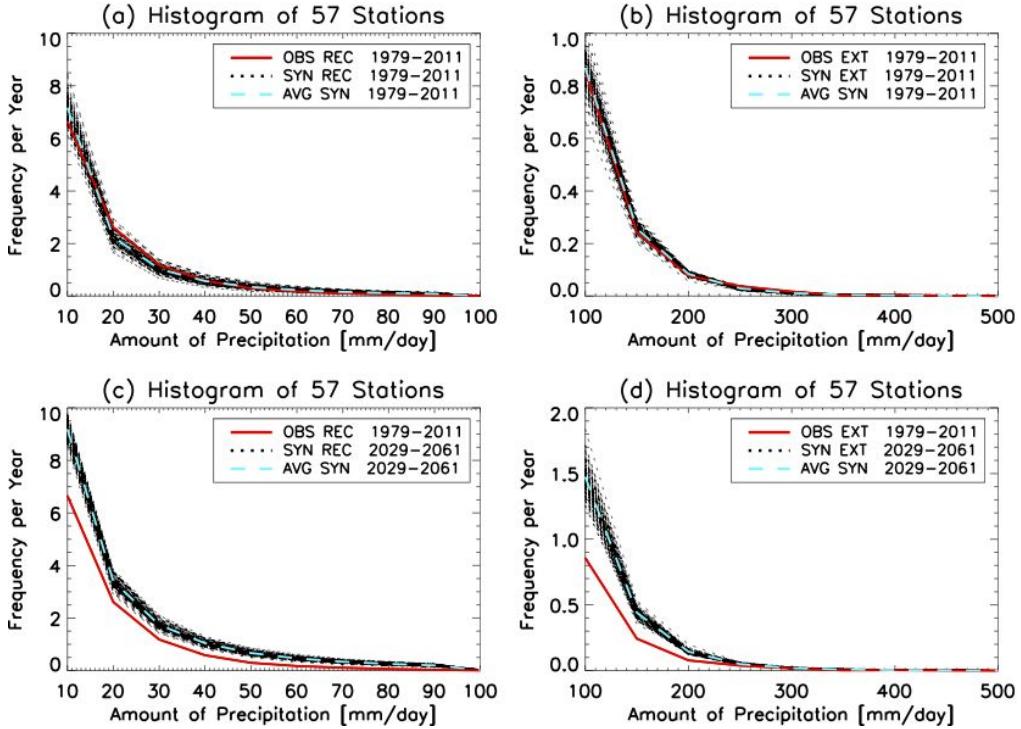


Figure 33. Histogram of precipitation for the amount less than 100 mm day^{-1} (left column) and for more than 100 mm day^{-1} (right column) in the observational period (upper panels) and in the period of 2029–2061 (lower panels). The red curve denotes the observational data, the black curves the 100 synthetic datasets, and the cyan curve the mean of the 100 black curves.

Table 5. Occurrence frequency per year in unit of days as a function of precipitation amount in the synthetic datasets during the observational period (1979-2011: SYN 33P) and in the future period (2029-2061: SYN 33F) in comparison with the observational record (1979-2011: OBS 33P). The occurrence frequencies below 100 mm day⁻¹ in the observational data are based on the 20-mode reconstruction of precipitation.

	Precipitation amounts				
	0	(0,10)	[10,100)	[100,500)	≥ 500
OBS 33P	32.23	74.83	11.70	1.241	0.002
SYN 33P	34.44	72.34	11.95	1.260	0.003
SYN 33F	34.92	65.64	17.29	2.147	0.004

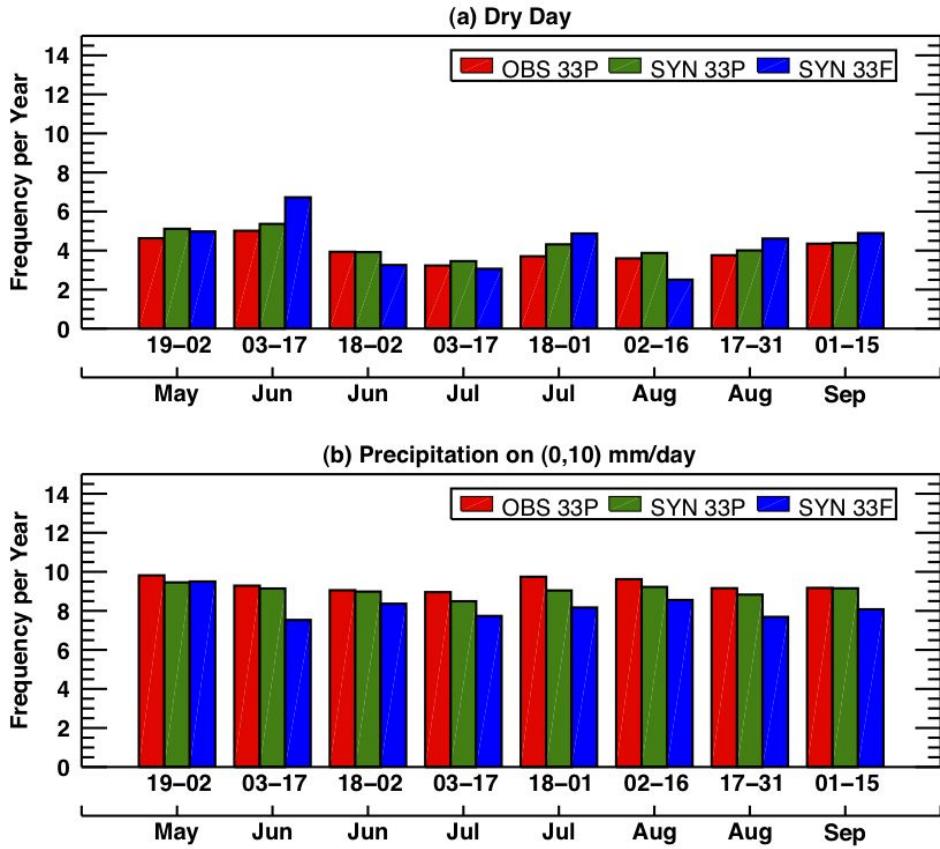


Figure 34. The occurrence frequency of dry days (upper panel) and precipitation events of $(0,10)$ mm day $^{-1}$ (lower panel) for the eight subintervals of 15 days each during summer. The red bars denote the observational data, the green bars the 100 synthetic datasets over the same period of the observational record, and the blue bars the 100 synthetic datasets over the period of 2029-2061.

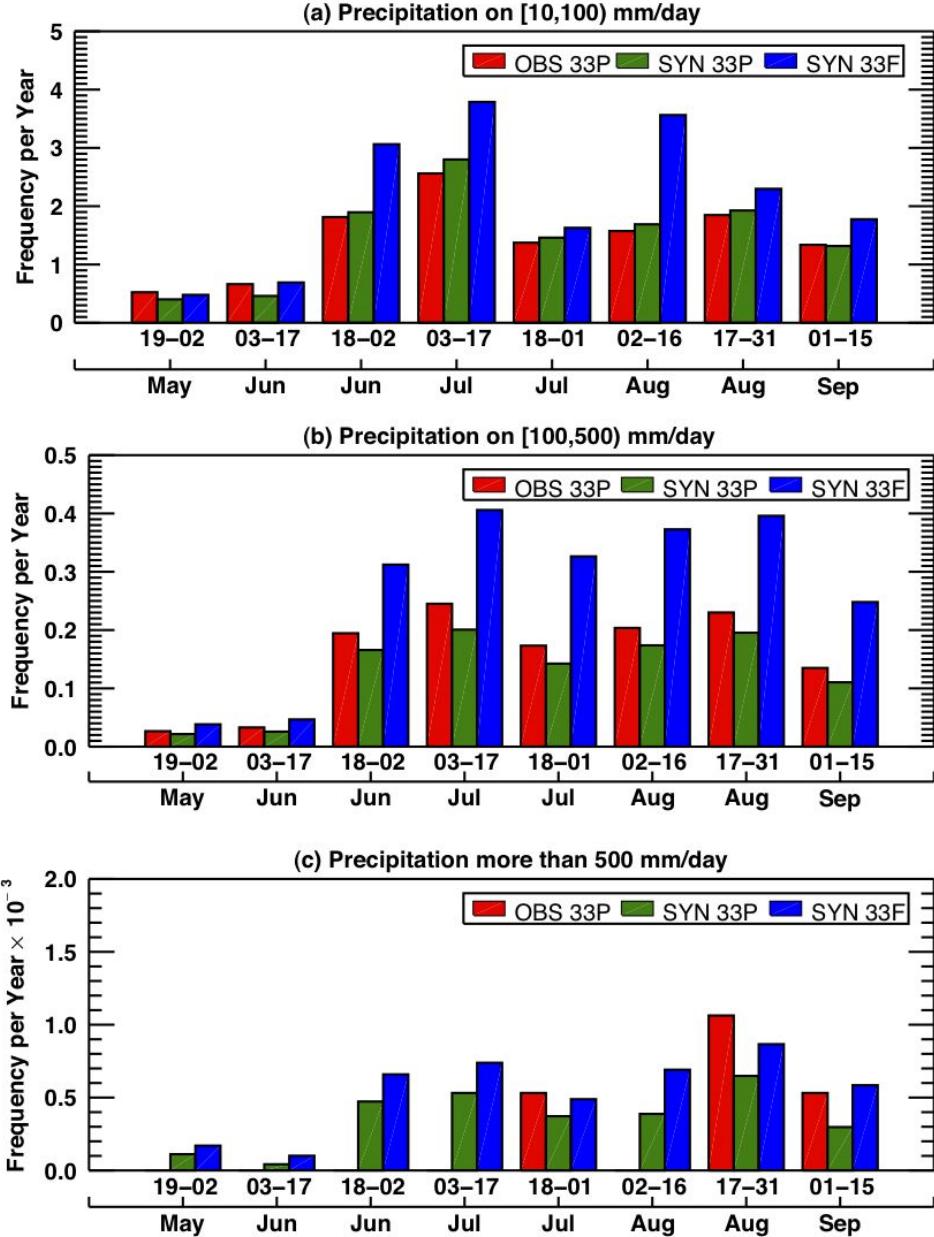


Figure 35. The occurrence frequency of precipitation for the amount of $[10,100]$ mm day $^{-1}$ (upper panel), for the amount of $[100,500]$ mm day $^{-1}$ (middle panel), and for the amount over 500 mm day $^{-1}$ (bottom panel). The red bars denote the observational data, the green bars the 100 synthetic datasets over the same period of the observational record, and the blue bars the 100 synthetic datasets over the period of 2029-2061.

6. Summary and Concluding Remarks

This study aimed at developing an efficient and accurate statistical model of summer precipitation in Korea for the purpose of using it as a weather generator. A conventional weather generator tends to be inaccurate in reproducing the spatial covariance structure of a weather variable. Therefore, the prime goal of this study was to develop a new weather generation algorithm based on CSEOF analysis, which extracts realistic temporal evolution patterns and their amplitude time series from a given observational dataset.

A new weather generation algorithm was developed in the following manner:

- (1) The observed precipitation was divided into two parts: precipitation less than 100 mm day⁻¹ and extreme precipitation over 100 mm day⁻¹. This distinction is necessary since extreme precipitation events are not frequent enough and are difficult to describe in terms of a small number of CSEOF modes.
- (2) The “no-extreme” precipitation was subject to CSEOF analysis. A natural log transformation was conducted on the precipitation data prior to CSEOF analysis; any values less than one were changed into one before taking the log transformation.
- (3) Each PC time series from step (2) was detrended and was fitted to an ARMA model.
- (4) Based on the ARMA models identified in step (3), 100 new sets of PC time series were generated by using 100 different realizations of error time series in the ARMA models. The new PC time series are longer than the original PC time series by 50 years (1979-2061).

(5) After adding the removed trend back in the synthetic time series, 100 synthetic datasets of precipitation were constructed for the period of 1979-2061.

(6) Synthetic data for “extreme” precipitation were generated based on the probability distribution function of the observed “extreme” precipitation events. Note that the probability distribution exhibits a small trend, which has been accounted for in generating 100 synthetic datasets.

(7) The two products of synthetic precipitation (“no-extreme” and “extreme” precipitation) were added together to come up with the final product of synthetic summer precipitation in Korea.

The synthetic precipitation data were examined in comparison with the observational data in terms of the statistical characteristics. Over the observational period, the synthetic data exhibited statistical properties similar to those of the observational data. A bimodal structure in the seasonal cycle of summer precipitation was faithfully reproduced in the synthetic data. The mean and the standard deviation of the seasonal cycle of precipitation were reasonably similar to those of the observational data. The frequency-intensity relationship was similar to that of the observational data from “no-rain” events to “extreme” events of precipitation. The spatial correlation function was also reproduced well; this is a feature of significant improvement over conventional weather generators.

The model data based on the HadGEM3-RA historical run showed that precipitation was significantly underestimated except in the range of $(0, 20)$ mm day $^{-1}$. As a consequence, an arbitrary boost of model precipitation may be necessary in order

to make a quantitative inference on the summer precipitation in Korea. The seasonal cycle derived from the model data was fairly weak resulting in a significant underestimation of the mean and the standard deviation of precipitation associated with the seasonal cycle. Further, the second peak of summer precipitation in Korea was missing although the first peak was of reasonable magnitude. Therefore, it is not appropriate to scale up the model precipitation in order to match the observational record.

The synthetic precipitation from a statistical model indicated that future precipitation would increase significantly except in the precipitation range of (0,10) mm day⁻¹. Both the mean and the standard deviation of the seasonal cycle increased significantly in the future period (2029-2061). There was a sign of precipitation spells starting earlier in future with earlier precipitation peaks. The model precipitation data under the RCP 4.5 and RCP 8.5 scenarios also exhibited increased precipitation compared with the historical run; the frequency of precipitation events in the range of [100,500] mm day⁻¹ increased substantially. On the other hand, future precipitation was still less than the observational precipitation. The first peak of the seasonal cycle was larger in magnitude than that of the observational data but the second peak was still missing in the RCM runs.

A detailed comparison of the synthetic precipitation and model precipitation against the observational data reveals that the statistical approach developed in this study is fairly reasonable. Despite the simplicity of the statistical approach, synthetic precipitation exhibits desirable statistical properties of precipitation over the

observational record. On the other hand, the RCM model (HadGEM3-RA) is less reliable in terms of reproducing statistical properties of the observed precipitation in Korea. In regard to future characteristics of precipitation over Korea, the two approaches—statistical approach versus RCM model approach—suggest both similar and dissimilar features of future precipitation. Both synthetic and model precipitations increase in future in comparison with their counterparts in the observational record yielding higher occurrence frequency of extreme precipitation events. On the other hand, the model precipitation exhibits little sensitivity in the precipitation range of [10,100] mm day⁻¹ while the synthetic precipitation indicates a substantial increase in the frequency of precipitation in the same range; the effect of climate change is more obvious in the model precipitation exceeding 500 mm day⁻¹. Both approaches should be employed and compared with each other until a clear answer is obtained in regard to the superiority of one approach to another.

The CSEOF-based weather generator is clearly an improvement over the existing weather generation algorithms. A notable improvement is that the CSEOF approach yields reasonable spatial covariance structures of weather variables. The CSEOF approach is different from existing approaches in that space-time physical evolutions are faithfully rendered. Note that each CSEOF loading vector consists of a number of spatial patterns describing the evolution of a physical evolution for a given period of time (nested period). The new algorithm makes a specific use of CSEOF loading vectors. In other words, physically realistic space-time covariance statistics are incorporated into weather generation. While some weather generation algorithms

consider the spatial covariance structures of weather variables they generate, the existing algorithms do not match spatial covariance structure with temporal covariance structure in a physically reasonable manner. As a result, synthetic datasets may exhibit features that are physically different from those of the observational record.

Another advantage of the CSEOF-based weather generation algorithm is that physical relationships among several variables in the observational record can be reproduced with accuracy. This can be achieved via regression analysis in CSEOF space. Once CSEOF analysis is completed for different physical variables, temporal evolution patterns of a predictor variable can be made to be physically consistent with that of a target variable via regression analysis in CSEOF space:

$$T_n(t) = \sum_{m=1}^M \alpha_m^{(n)} P_m(t) + \varepsilon^{(n)}(t), \quad n = 1, 2, \dots, N, \quad (14)$$

where $\{\alpha_m^{(n)}\}$ are the regression coefficients, N is the number of target PC time series and M is the number of predictor PC time series. Equation (14) simply describes a regression analysis among PC time series of the target and predictor variables. Then, a new evolution pattern, $D_n(r, t)$, can be found via

$$D_n(r, t) = \sum_{m=1}^M \alpha_m^{(n)} C_m(r, t), \quad n = 1, 2, \dots, N, \quad (15)$$

where $C_m(r, t)$ are the CSEOF loading vectors of the predictor variable. Then, we can write the target and predictor time series as

$$(T(r, t); P(r, t)) = \sum_{n=1}^N \{B_n(r, t); D_n(r, t)\} T_n(t), \quad (16)$$

where $T_n(t)$ is the PC time series of the target variable. The terms in the curly brackets, $B_n(r,t)$ and $D_n(r,t)$, represent physically consistent evolutions of the target and the predictor variables. Therefore, generation of synthetic PC time series, $\hat{T}_n(t)$, allows the construction of not only synthetic target variable, $\hat{T}(r,t) = \sum_{n=1}^N B_n(r,t)\hat{T}_n(t)$, but also synthetic predictor variable, $\hat{P}(r,t) = \sum_{n=1}^N D_n(r,t)\hat{T}_n(t)$. In this way, synthetic data for different variables can be generated in a physically consistent manner. This idea was used in the reconstruction of sea level height based on sea surface temperature in *Hamlington et al.* [2012].

An important caveat in the present study is that the linear trend in each of the PC time series of the KMA precipitation will continue in the future. There is no guarantee that the trend in the PC time series is linear nor this linear trend persists in the future. Further, error occurs undoubtedly in the estimation of the linear trend, which may amplify in the future. The increasing trend of precipitation in the synthetic datasets, therefore, should be compared with those in the RCP 4.5 and the RCP 8.5 datasets. This can be accomplished by conducting CSEOF analysis on the RCP 4.5 and RCP 8.5 datasets. Despite the uncertainties, which are unavoidable in a statistical approach, the new weather generation algorithm appears to produce reasonable precipitation over the observational record and provide useful information on future precipitation change. Statistical approaches should accompany numerical approaches until numerical models are able to produce accurate precipitation field under various climate change scenarios.

References

- Aas, K., and I. H. Haff (2006), The generalized hyperbolic skew student's t-distribution, *Journal of Financial Econometrics*, 4, 275-309, doi:10.1093/jjfinec/nbj006.
- Buishand, T. A. (1977), Stochastic modeling of daily rainfall sequences, Ph.D. thesis, Communications Agricultural University, Wageningen, Netherlands.
- Chen, J., F. P. Brissette, and R. Leconte (2010), A daily stochastic weather generator for preserving low-frequency of climate variability, *J. Hydrol.*, 388, 480-490, doi:10.1016/j.jhydrol.2010.05.032.
- Chen, J., H. Uyeda, and D.-I. Lee (2003), A method using radar reflectivity data for the objective classification of precipitation during the Baiu season, *J. Meteorol. Soc. Jpn.*, 81(2), 229–249, doi:10.2151/jmsj.81.229.
- Collins, W. J., et al. (2008), Evaluation of HadGEM2 model, *Hadley Centre Technical Note 74*, Met Office, Exeter, U. K.
- Dubrovsky, M., J. Buchteke, and Z. Zalud (2004), High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modeling, *Clim. Change*, 63, 145–179, doi:10.1023/B:CLIM.0000018504.99914.60.
- Durrett, R. (2010), *Probability: Theory and Examples*, 4th ed., 438 pp., Cambridge Univ. Press, New York.
- Fernández, C., and M. F. J. Steel (1998), On Bayesian modeling of fat tails and skewness, *J. Am. Stat. Assoc.*, 93, 359–371, doi:10.1080/01621459.1998.10474117.

- Gabriel, K. R., and J. Neumann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv, *Q. J. R. Meteorol. Soc.*, 88, 90–95,
doi:10.1002/qj.49708837511.
- Hamlington, B. D., R. R. Leben, and K.-Y. Kim (2012), Improving sea level reconstructions using non-sea level measurements, *J. Geophys. Res.*, 117, C10025, doi:10.1029/2012JC008277.
- Ho, C.-H., and I.-S. Kang (1988), The variability of precipitation in Korea, *Journal of the Korean Meteorological Society*, 24(1), 38–48.
- Hosking, J. R. M., and J. R. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, 29(3), 339–349,
doi:10.1080/00401706.1987.10488243.
- Katz, R. W. (1977), Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, 16, 671–676, doi:10.1175/1520-0450(1977)016<0671:PAACDP>2.0.CO;2.
- Kim, K.-Y., and G. R. North (1997), EOFs of harmonizable cyclostationary processes, *J. Atmos. Sci.*, 54, 2416–2427.
- Kim, K.-Y., G. R. North, and J. Huang (1996), EOFs of one-dimensional cyclostationary time series: Computations, examples and stochastic modeling, *J. Atmos. Sci.*, 53, 1007–1017.
- Kim, K.-Y., J.-W. Roh, D.-K. Lee, and J.-G. Jhun (2010), Physical mechanisms of the seasonal, subseasonal, and high-frequency variability in the seasonal cycle of summer precipitation in Korea, *J. Geophys. Res.*, 115, D14110,
doi:10.1029/2009JD013561.

- Lee, D., H. An, Y. Lee, J. Lee, H.-S. Lee, and H.-S. Oh (2010), Improved multisite stochastic weather generation with applications to historical data in South Korea, *Asia-Pacific Journal Atmospheric Sciences*, 46, 497-504, doi:10.1007/s13143-010-0031-2.
- Lee, S.-S., P. N. Vinayashandran, K.-J. Ha, and J.-G. Jhun (2010), Shift of peak in summer monsoon rainfall over Korea and its association with El Niño–Southern Oscillation, *J. Geophys. Res.*, 115, D02111, doi:10.1029/2009JD011717.
- Lim, Y.-K., K.-Y. Kim, and H. S. Lee (2002), Temporal and spatial evolution of the Asian summer monsoon in the seasonal cycle of synoptic fields, *J. Clim.*, 15, 3630-3644, doi:10.1175/1520-0442(2002)015<3630:TASEOT>2.0.CO;2.
- Monserrat, J. F., R. Fraile, and L. Rubio (2007), Application of alternating projection method to ensure feasibility of shadowing cross-correlation models, *Electron. Lett.*, 43(13), 724-725, doi:10.1049/el:20070236.
- Newton, H. J. (1988), *Timeslab: A Time Series Analysis Laboratory*, 625 pp., Wadsworth, New York.
- Qian, W. H., H.-S. Kang, and D.-K. Lee (2002), Distribution of seasonal rainfall in East Asian monsoon region, *Theoretical and Applied Climatology*, 73, 151–168, doi:10.1007/s00704-002-0679-3.
- Richardson, C. W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, 17(1), 182–190, doi:10.1029/WR017i001p00182.

- Richardson, C. W., and D. A. Wright (1984), WGEN: a model for generating daily weather variables, *ARS-8*, 83pp., Agricultural Research Service, U.S. Department of Agriculture, Washington, D. C.
- Roh, J.-W., K.-Y. Kim, and J.-G. Jhun (2012), Decadal changes in the physical mechanisms of the seasonal cycle of summertime precipitation variability in Korea, *J. Geophys. Res.*, 117, D07115, doi:10.1029/2011JD017268.
- Stern, R. D., and R. Coe (1984), A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society Series A*, 147(1), 1–34.
- Thom, H. C. S. (1958), A note on the gamma distribution, *Mon. Weather Rev.*, 86(4), 117–122.
- Walters, D. N., et al. (2011), The Met office unified model global atmosphere 3.0/3.1 and JULES global land 3.0/3.1 configurations, *Geoscientific Model Development Discussions*, 4, 1213–1271, doi:10.5194/gmdd-4-1213-2011.
- Wang, B., J.-G. Jhun, and B.-K. Moon (2007), Variability and singularity of Seoul, South Korea, rainy season (1778–2004), *J. Clim.*, 20, 2572–2580, doi:10.1175/JCLI4123.1.
- Wilby, R. L. (1994), Stochastic weather type simulation for regional climate change impact assessment, *Water Resour. Res.*, 30(12), 3395–3403, doi:10.1029/94WR01840.
- Wilks, D. S. (1989), Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resour. Res.*, 25(6), 1429–1439, doi:10.1029/WR025i006p01429.

- Wilks, D. S. (1992), Adapting stochastic weather generation algorithms for climate change studies, *Clim. Change*, 22, 67–84, doi:10.1007/BF00143344.
- Wilks, D. S. (1998), Multisite generalizations of a daily stochastic precipitation generation model, *J. Hydrol.*, 210, 178–191.
- Wilks, D. S. (1999), Interannual variability and extreme value characteristics of several stochastic daily precipitation models, *Agricultural and Forest Meteorology*, 93, 153–169.
- Wilks, D. S., and R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Progress in Physical Geography*, 23(3), 329-357, doi: 10.1177/030913339902300302.
- Woolhiser, D.A., and G.G.S. Pegram (1979), Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models, *J. Appl. Meteorol.*, 18, 34–42, doi: 10.1175/1520-0450(1979)018<0034:MLEOFC>2.0.CO;2.

초 록

관측자료를 바탕으로 한 한국 여름철 강수량 모의와 모의된 결과에서의 미래 강수 변화

이 연구에서는 한국에서의 여름철(5월 19일–9월 15일) 강수량을 생성하는 날씨 생성기(weather generator)를 개발하였다. 기상청 57 지점에서 관측된 33년(1979–2011년) 동안의 강수량 자료가 새로운 날씨 생성기를 개발하는데 사용되었다. cyclostationary empirical orthogonal function (CSEOF) 기법을 사용하여 관측된 강수량 자료를 여름철 동안의 결정적 변천 패턴들과 그에 대응되는 확률적 크기 시계열의 선형조합으로서 표현하였다. 결정적 변천 패턴들은 해마다 반복되지만 그것의 크기는 해마다 달라질 수 있다. 그리고 나서 각각의 추세가 제거된 크기 시계열을 autoregressive moving average (ARMA) 모델에 적합을 한다. 적합된 ARMA 모델은 본래의 크기 시계열과 통계적 성질이 유사한 임의의 길이의 인공 크기 시계열을 생성시키는데 사용된다. 그리고 나서 1979–2061년 (83년)까지의 100개의 새로운 크기 시계열 자료셋을 생성하였다. 이러한 인공 크기 시계열과 해마다 반복되는 변천 패턴들을 바탕으로 100개의 인

공적인 여름철 강수량 자료셋을 생성하였다.

관측기간에서의 인공 강수량 자료셋의 통계적 성질과 기상청 관측 강수량 자료의 통계적 성질을 비교하였다. 현재 연구에서 발달된 날씨 생성 기의 수행능력을 좀더 면밀히 평가하기 위해 미래 기간에서의 인공 강수량 자료셋과 대표 농도 경로(representative concentration pathways: RCP) 시나리오에 따른 지역기후모델(regional climate model: RCM)로부터 산출된 강수량을 비교하였다.

인공 강수량의 seasonal cycle은 크기를 약간 과대추정하기는 하지만 여름철 강수량의 전형적인 두 개의 피크 형태를 잘 묘사한다. RCM은 강수의 두 번째 피크를 유의미하게 과소추정한다. Seasonal cycle의 평균과 표준편차에서도 인공 강수량 자료는 관측 강수량 자료의 seasonal cycle의 평균과 표준편차를 잘 나타나지만 RCM의 현재기후에서 강수량은 과소추정한다.

인공 강수량 자료의 강도 별 빈도수는 관측자료의 것과 유사하다. 현재기후에서의 RCM 강수량 자료의 강도 별 빈도수는 무강수 빈도수와 20 mm day^{-1} 이상의 강수 빈도수를 유의미하게 과소추정하는 반면 20 mm day^{-1} 미만의 강수의 빈도수는 크게 과대추정한다. 극한 강수의 분포를 generalized Pareto distribution (GPD)로 표현하였다. 인공 강수량 자료에 극한 강수의 분포는 관측자료의 것과 유사하지만 RCM의 현재기후에서의 강수량은 극한 강수의 빈도수를 과소추정한다. 인공 강수량 자료와

RCM의 현재기후에서의 강수량 자료 모두 관측 강수량 자료의 공간 상관성을 잘 나타낸다.

무강수의 빈도수는 거의 변하지 않은 채 10 mm day^{-1} 미만의 강수를 제외한 대부분의 강수 강도 별 발생빈도가 증가하였다. 특히 [100, 500) mm day^{-1} 의 범위에 강수 빈도수가 눈에 띄게 증가하였다. RCM의 RCP 4.5와 RCP 8.5 모두에서 현재기후에 비해 100 mm day^{-1} 이상의 강수 발생빈도가 증가하였다. RCM의 RCP 4.5와 RCP 8.5의 경우에 GPD를 보면 현재기후에 비해 극심한 강수의 상대적 빈도수가 증가하였다. 그럼에도 불구하고 RCM의 미래 시나리오에서 [100, 500) mm day^{-1} 의 강수 발생빈도는 관측 강수량 자료의 것보다 여전히 작다. 반면에 500 mm day^{-1} 를 초과하는 극한 강수 발생 빈도는 관측 강수량 자료의 것보다 더욱 빈번하게 발생한다.

주요 용어: 인공 강수량, 날씨 생성기, CSEOF, ARMA, 지역기후모델

학 번: 2011-20362

감사의 글

2년동안 부족한 저를 열정적으로 이끌어주신 김광열 교수님께 감사합니다.

바쁜 일정에도 불구하고 논문심사를 위해 시간을 내어주신 임규호 교수님,
전종갑 교수님께 감사합니다.

통계기후학 실험실에서 지내면서 저의 사소한 질문에도 친절하게 답변해주신 노준우 박사님, 나한나 박사님, 유진언니, 새림언니, 훈규오빠, 진수에게
감사 드립니다. 새로 입학하는 대학원 신입생 서연아 발표 들어줘서 고맙다.
자주 뵙지는 못하였지만 도움을 주려고 하셨던 지영선배 감사합니다.

제가 자주 들락거리면서 번거롭게 했음에도 불구하고 항상 즐겁게 받아주신
가까운 옆방 기후역학 실험실 선배님들에게도 감사합니다. 대학원 생활을
무료하지 않게 해준 대학원 동기 진수, 민섭오빠, 훈영, 광연, 평화, 소라,
민진, 정화, 지연언니, Rijana, 상원오빠, 재창오빠, 상무, 승규에게도 감사합
니다. weather generator 관련 자료들을 제공해주신 통계학과의 이동환 선
생님께 감사합니다.

항상 저를 응원해주신 부모님께 감사합니다.