



Quantifying climate internal variability using an hourly ensemble generator over South Korea

Jongho Kim^{1,2} · Muhammad Ejaz Tanveer² · Deg-Hyo Bae²

Published online: 3 October 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Identifying climate internal variability (CIV) generated by non-linear interactions and feedbacks among many components of the climate system is essential and challenging because of its irreducible and unpredictable characteristics. A range of studies have addressed this issue; however, these studies focused on the first order moments of few representative climate variables at relatively larger spatial and temporal scales. To investigate the magnitude and the spatial pattern of CIV relevant at finer spatial (point) and temporal (hourly) scales, CIV is assessed over a 30-year period in South Korea by analyzing 100-member ensemble generated using an hourly weather generator and a bootstrapping approach. Statistics addressing the first and second order moments, occurrences, and extremes are successfully verified at various temporal scales. The CIV is then estimated by the ‘detrended’ and ‘differenced’ methods for the four metrics proposed at different scales that signify rainfall volume, maxima, and occurrence. Consequently, the implications of this study are the following: (1) the estimation of CIV using bootstrapped ensembles often fails to represent the proper uncertainty range, resulting in high chances of underestimating extreme statistics, such as the maximum rainfall depth; (2) regardless of which of the two methods is used, no significant difference in the CIV estimation is observed; and (3) a temporal scale-dependency is observed for the proposed metrics used to identify the magnitude and the seasonal pattern of the CIV—the utility of an hourly time series and its associated extreme properties deserves significant attention. Ultimately, the spatial mapping and grouping of CIV will provide valuable information to identify which regions have high variability compared to climatological norms and thus are more vulnerable to extremes, and will serve as a guide for planning adaptation and mitigation measures against future extreme events.

Keywords Climate internal variability · Ensemble generation · Hourly weather generator · Bootstrapping approach

1 Introduction

Climate internal variability (CIV), an often used term and an essential part of the climate system, refers to the natural or stochastic variations generated by non-linear interactions and feedbacks among the many components of

climate system (IPCC 2013). Recognizing the consequence of the CIV and quantifying its uncertainty play a vital role when interpreting climate projections for the future as well as present because such an uncertainty does not ultimately diminish as time scales increase but remains unpredictable, even if the predictability of the climate models are enhanced or the uncertainty of the external conditions (e.g., emission scenarios) is well identified (Addor and Fischer 2015; Deser et al. 2012b; Hawkins and Sutton 2009).

Although there is no perfect consensus on how to quantify the climate internal variability, the uncertainty of the CIV has been extensively estimated by constructing ensemble members and calculating the degree of spread among their results. Depending on how the ensemble member is generated, several competing approaches can be categorized into four different groups. The widespread approach named ‘dynamic’ involves the use of

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00477-018-1607-0>) contains supplementary material, which is available to authorized users.

✉ Deg-Hyo Bae
dhbae@sejong.ac.kr

¹ School of Civil and Environmental Engineering, University of Ulsan, Ulsan, South Korea

² Department of Civil and Environmental Engineering, Sejong University, Seoul, South Korea

deterministic global or regional climate models in generating a large number of climate simulations, which are designed by slightly varying the initial conditions of atmospheric component of climate system but with the other conditions remaining entirely identical (Deser et al. 2012a, b). While extensively used at global and multi-decadal scales, this approach has limitations in capturing extremes at finer temporal scales (e.g., subdaily) for local impact studies because the outputs are inherently based on coarser spatial resolutions and low temporal resolutions (Chen et al. 2011; Maraun et al. 2010), and in securing as many ensemble members as possible because of its high demand for computational resource—the frequently-used number of ensemble members was in the range from 10 to 50. The second approach estimates the uncertainty of the CIV from a simple analytical model (Thompson et al. 2015). The analytical model can present a reasonable range of estimates for the CIV; however, it requires several key assumptions: the internal variability follows a roughly Gaussian distribution, the signature of climate change is estimated from linear trends, and the two statistics used in the model are stationary in time (Thompson et al. 2015). The last approaches are named ‘statistical’ (Addor and Fischer 2015; Prudhomme and Davies 2009) and ‘stochastic’ (Brisson et al. 2015; Fatichi et al. 2013; Ng et al. 2017). Both of these approaches produce ensembles from the statistical or stochastic characteristics extracted from any type of observed data. The differences depend on whether a realization that has not been observed in the historic record is fully explored. Because many features of the predicted climate variable are inherited from the historic data, the statistical approach intrinsically has drawbacks in that variability can be underestimated (Hingray and Said 2014). In contrast, the stochastic approach can reproduce the stochastic nature of climate variables with varying random terms in the stochastic components of climate models to the extent that the model captures the relevant stochastic processes. In this regard, some CIV studies build a framework of a weather generator to produce climate realizations using various statistics of the observed data (Brisson et al. 2015; Fatichi et al. 2016; Kim et al. 2016b).

The IPCC (2013) describes the climate variability on the basis of variations in the statistics of the climate on all temporal and spatial scales—the relevant statistics could be the mean state, standard deviation, the occurrence of extremes, etc. (Khalili et al. 2016; Singh and Goyal 2017). However, a range of studies addressing the climate internal variability emphasizes not only the first order moments of few representative climate variables (e.g., surface air temperature, mean precipitation) (Addor and Fischer 2015; Deser et al. 2012a, 2014; Thompson et al. 2015; Yao et al. 2016), with the exception of a few studies (e.g., Fischer

et al. 2013) but also on relatively larger spatial and temporal scales (e.g., regional or interannual) with relatively coarser resolutions (e.g., several hundred kilometers in degrees or daily) (Deser et al. 2012a, 2014; Fischer et al. 2013; Yao et al. 2016), with the exception of a few studies (e.g., Fatichi et al. 2016). Investigating the role of climate internal variability and its hydrogeomorphic implications that are particularly relevant at the regional to local scale (Mohammed et al. 2016) is challenging because of the nonlinear and non-unique characteristics of hydrogeomorphic processes (Kim et al. 2016a; Kim and Ivanov 2014) and increasing uncertainty by propagations (Coulthard et al. 2012; Kim and Ivanov 2015; Kim et al. 2016c); Thus, understanding the response of precipitation to internal forcing has become a topic of great emphasis in the climate and water resources communities (Kim et al. 2016d).

In this study, internal variability over 30 years periods in South Korea is assessed by analyzing 100-member ensembles (i.e., 30-year, 100 ensembles for each gauge location) generated in two different ways: using an hourly weather generator and a bootstrapping approach. The domain is selected because the rainfall characteristics of the Korean Peninsula belonging to the Far East Asian Monsoon region are very complex in time and space. Its internal variability is expected to be large, but this variability has not been completely investigated to date, especially via models with high temporal resolutions. To the best of our knowledge, this is the first study in which the internal variability is identified over the whole country by using alternative climate realizations with the feature being simulated at an hourly temporal scale. First, statistics addressing the first and second order moments, occurrences, and extremes are computed and verified at various temporal scales. Next, the climate internal variability is estimated by the ‘detrended’ and ‘differenced’ methods for the various metrics proposed at different scales representing rainfall volume, maxima, and occurrence. Finally, the results provide information regarding whether an area belongs to a region with higher variability and thus can be vulnerable to the variability.

2 Methodology of stochastic ensemble generation and CIV estimation

A brief outline of the methodology adopted for the generation of ensemble members and the estimation of climate internal variability is provided in the following section using a stepwise description; an associated flowchart is illustrated in Fig. 1.

1. The first step relates to the collection of the hourly data available from and the completion of quality control

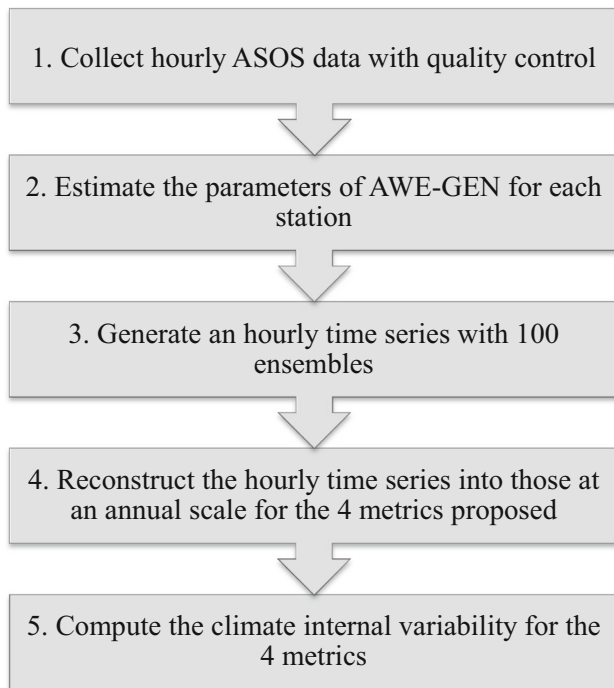


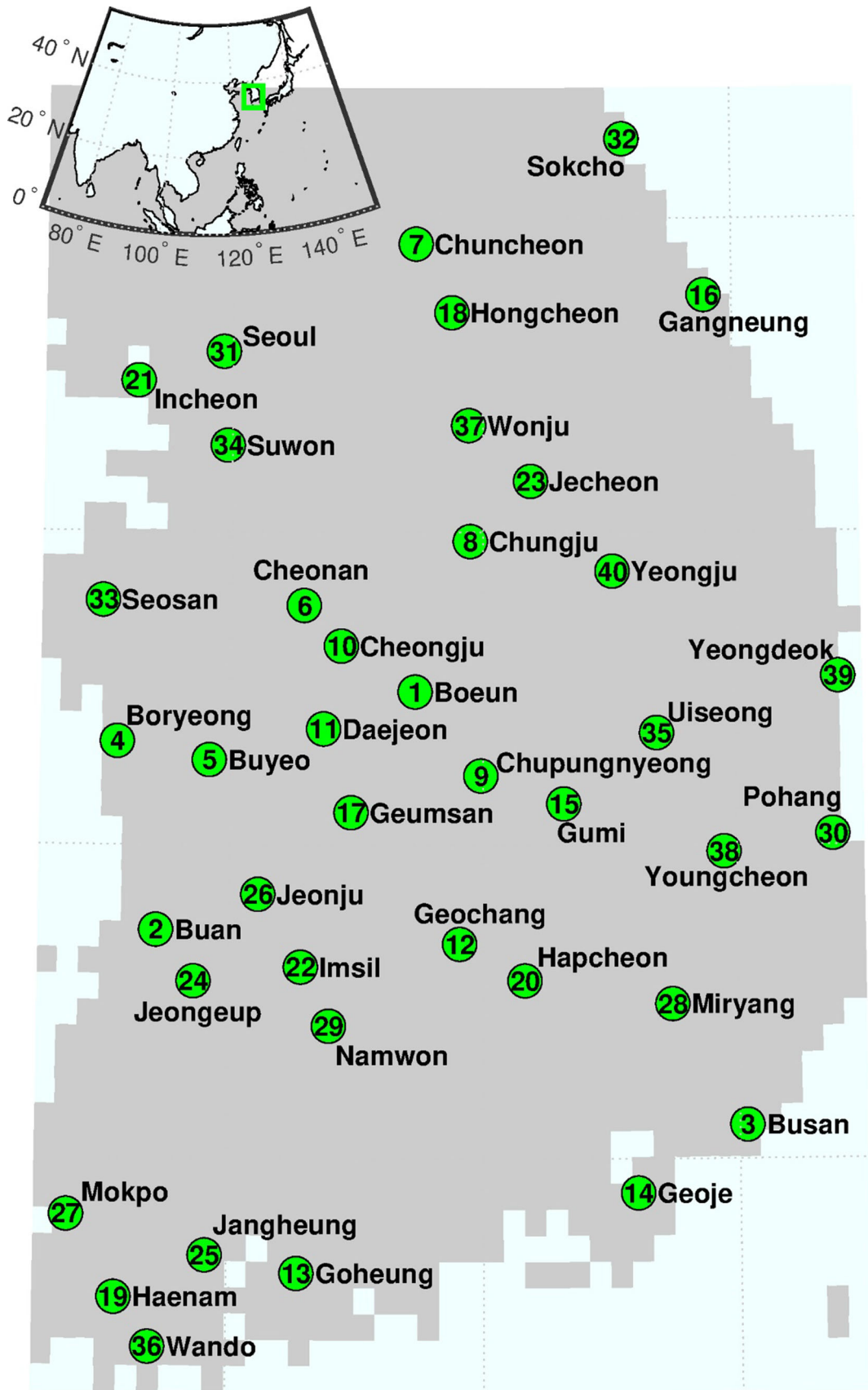
Fig. 1 Flow chart of the adopted procedure and methodology

for gauge stations over South Korea. Estimating the parameters of a stochastic ensemble generator described in the next step requires observation data such as precipitation, cloud cover, pressure, direct and diffusive radiation, temperature, dew point temperature, wind speed, and relative humidity at an hourly scale. The appropriate data is found in the automatic surface observation network operated by the Korea Metrological Administration (KMA), i.e., the Automated Surface Observing System (ASOS) network for synoptic surface observation. The 40 stations used in this study were chosen based on the criteria of data availability: at least 30 years hourly observation is recorded from the KMA website for 14 weather variables (<https://data.kma.go.kr/data/grnd/selectAsosList.do?pgmNo=34>). Before using the original data, one must verify the quality of the ASOS hourly data to prevent potential errors. Although very few in number, the subhourly or duplicate values are removed. In addition, missing values are inserted through a linear interpolation process. The percentage of the missing or duplicate hourly data from the total time series for all the stations is less than 0.01%, and most of the missing values correspond to times when it was not raining (simply filled with zero values). Thus, it should have minimal impact, and the data is considered reliable. The location and information of all the selected stations are shown in Fig. 2 and Table 1.

- In this study, a stochastic weather generator, the Advanced WEather GENerator (AWE-GEN) (Fatichi et al. 2011), followed by a substantial evolution of the model presented by Ivanov et al. (2007), is employed to generate the ensemble members. An overall outline of the stochastic weather generator can also be found in Fatichi et al. (2013) and Kim et al. (2016b); only a brief summary is presented here. The AWE-GEN is ultimately designed to produce ‘hourly’ time series of several weather variables for a given stationary climate that can be used to generate a stochastically varying, theoretically infinite time series. The latter is therefore appropriate in making many ensemble members and in representing the CIV, which satisfies the scope of this study. In particular, the model has an elevated capacity to reproduce the statistical properties of precipitation at multiple temporal scales, ranging from subdaily to inter-annual. The high-frequency component of precipitation is simulated using one of the Poisson process models, i.e., the Neyman–Scott Rectangular Pulse (NSRP) model, with statistics over different aggregation periods, and the low-frequency inter-annual variability is addressed using an autoregressive order-one model, AR(1), with the modification for skewness (Fatichi et al. 2011).

The parameters adopted to generate the internal structure of precipitation in the NSRP model are the storm time origin, occurring as a Poisson process with the rate λ (h^{-1}); a random number of cells generated for each storm following a geometrical distribution with the mean μ_c (–); the cell displacement from the storm origin that is exponentially distributed with the mean β^{-1} (h); a rectangular pulse associated with each precipitation cell assumed to be an exponentially distributed with the mean η^{-1} (h) and intensity (mm h^{-1}); and the intensity distributed as a Gamma distribution with the parameters α and θ . The parameters used in the AR(1) process are the average annual precipitation, the standard deviation, the lag-1 autocorrelation, and the random deviate of the process (Fatichi et al. 2011). In total, 76 parameters (6 NSRP parameters for each month plus 4 AR(1) parameters for annual observations) are estimated during the generation of precipitation (Fatichi et al. 2011).

The in situ point-scale observations over the 30-year period (1981–2010) obtained in step (1) are then applied to the estimation of its parameters. This estimation follows the procedure of Cowpertwait et al. (2007); this procedure employs an objective function comprising the summation of the relative squares of the following four statistics between the observed data and the simulated results: the coefficient of variation, the lag-1 autocorrelation, the skewness,



◀ **Fig. 2** Overview of the selected 40 locations in South Korea**Table 1** Detailed information of the 40 stations

No.	Station name	Elevation	Latitude	Longitude
1	Boeun	175	36.4875	127.7342
2	Buan	11.96	35.7265	126.7166
3	Busan	69.6	35.1047	129.0319
4	Boryeong	16.9	36.3272	126.5574
5	Buyeo	11.79	36.2694	126.9208
6	Cheonan	81.5	36.7624	127.2927
7	Chuncheon	75.64	37.9165	127.7357
8	Chungju	116.29	36.9676	127.9525
9	Chupungnyeong	243.7	36.2203	127.9944
10	Cheongju	58.7	36.6333	127.4406
11	Daejeon	68.9	36.3691	127.3721
12	Geochang	230.2	35.6833	127.9100
13	Goheung	53.76	34.6318	127.2757
14	Geoje	45.4	34.8881	128.6047
15	Gumi	48.88	36.1306	128.3205
16	Gangneung	26.04	37.7515	128.8910
17	Geumsan	170.34	36.1027	127.4817
18	Hongcheon	140.91	37.6975	127.8804
19	Haenam	13.21	34.5536	126.5690
20	Hapcheon	32	35.5664	128.1661
21	Incheon	71.43	37.4777	126.6249
22	Imsil	247.11	35.6092	127.2856
23	Jecheon	259.8	37.1592	128.1944
24	Jeongeup	44.61	35.5632	126.8661
25	Jangheung	45.02	34.6888	126.9195
26	Jeonju	61.4	35.8408	127.1172
27	Mokpo	37.99	34.8169	126.3812
28	Miryang	11.2	35.4914	128.7442
29	Namwon	132.52	35.4213	127.3965
30	Pohang	2.91	36.0320	129.3800
31	Seoul	85.8	37.5714	126.9658
32	Sokcho	18.1	38.2509	128.5647
33	Seosan	28.91	36.7766	126.4939
34	Suwon	34.1	37.2723	126.9853
35	Uiseong	81.8	36.3561	128.6886
36	Wando	35.447	34.3959	126.7018
37	Wonju	148.6	37.3376	127.9466
38	Yongcheon	93.86	35.9774	128.9514
39	Yeongdeok	40.61	36.5333	129.4094
40	Yeongju	210.61	36.8718	128.5169

and the probability that an arbitrary interval is dry (Fatichi et al. 2011). For each month, the statistics containing the six parameters of NSRP are computed at

- four aggregation time intervals of 1, 6, 24, and 72 h. In total, 16 values (4 statistics \times 4 intervals) are used in minimizing the objective function for each month.
- After estimating the parameters for each station, AWE-GEN generates a set of hourly weather variables (e.g., precipitation, cloud cover, air temperature, shortwave incoming radiation, vapor pressure, wind speed and atmospheric pressure); however, we focus only on precipitation in this study. The time series generated by the model represents one of any potential stochastic realizations for a given stationary climate condition over the period of interest. Considering the implication of Kim et al. (2016b) that states that the adequate number of ensemble members required to reflect at least 90–110% of stochastic variation for several climate statistics is 100. Therefore, 100 ensemble members of the 30-year time series are produced in this study—note that the members are all plausible outcomes of the same set of parameters.
 - The fourth step of this methodology is to reconstruct the hourly time series into those at an annual scale for the 4 metrics computed over both the whole year and each month. The metrics are the total precipitation (abbreviated hereafter as ‘totPr’), daily maximum precipitation (‘maxDa’), hourly maximum precipitation (‘maxHr’), and non-precipitation days (‘nonPr’). This results in 13 reconstructed time series at an annual scale for each metric for each ensemble member. The ensemble for these reconstructed time series is generated in two ways. First, the ensemble for the weather generator (WG) is constructed by calculating the 100 hourly time series attained in Step (3). Second, the ensemble for bootstrapping (BS) is obtained by randomly resampling 100 times with replacements of the years of the reconstructed time series computed from the observed record. The metrics are designed to address extreme values of rainfall maxima at both hourly and daily scales, the occurrences of dry periods, and the total volume of rainfall. Values less than a threshold of 0.1 mm in the hourly time series are disregarded and set to zero.
 - One of the two methods of quantifying the climate internal variability is the most common “detrended” approach, which separates the forced signal and the internal variability (Frankcombe et al. 2015; Giorgi and Mearns 2002; Moise and Hudson 2008). When a long term forced signal representing the climate response to the external forcing exists and can be estimated, such a forced signal is subtracted from the time series, resulting in the fluctuations of the detrended time series (i.e., residuals) being regarded as the internal variability. The forced signal could be approximated with a linear trend (Frankcombe et al.

2015; Moise and Hudson 2008; Zhang and Wang 2013) or a more complicated form (e.g., the fourth order polynomial model) (Addor and Fischer 2015; Hawkins and Sutton 2009). The climate internal variability is finally estimated by computing either the standard deviation (e.g., Frankcombe et al. 2015) or the range (e.g., Moise and Hudson 2008) of the residual time series. In this study, we use the standard deviation of the residuals as an estimation of the CIV.

Another method named the “differenced” approach is built with a large ensemble of climate simulations (Frankcombe et al. 2015). Based on the assumption that the ensemble members are generated independently for the internal components, averaging over the ensemble members signifies that the component of internal variability is canceled out; as a result, the average over the ensemble becomes the forcing component (i.e., forced signal) representing the model response to the external forcing. Similar to the above method, the internal variability is then estimated with the standard deviation or the range of the residuals attained by subtracting the forcing component from time series of each ensemble. To investigate the internal (stochastic) variability over South Korea, the above steps are repeated over the selected 40 locations.

3 Results and analysis

3.1 Evaluation of AWE-GEN results

Before the ensemble generated is applied to estimate the CIV, the verification of the results by the AWE-GEN is performed on the statistics computed at the hourly and daily scales. The statistics entail extreme properties, such as maximum rainfall depth and consecutive dry or wet days, as well as the first and second moment properties, such as the mean and standard deviation, as demonstrated in Figs. 3 and 4, respectively. First, the statistics of the mean, standard deviation, and frequency of non-precipitation are evaluated for each month with respect to those from the ASOS data. Among the 100 ensemble members, their median is compared with the observed statistics in Fig. 3, which reveals a good model performance for all the locations. Note that the value of R^2 ranges from 0.55 to 0.99. In contrast with the excellent performance of mean and standard deviation, the simulated results for the frequency of non-precipitation show a slight overestimation that it is less likely to rain at an hourly scale. Because other statistics corresponding to the means and extremes are constrained within a comparatively satisfactory range, the difference can be due to the treatment of smaller

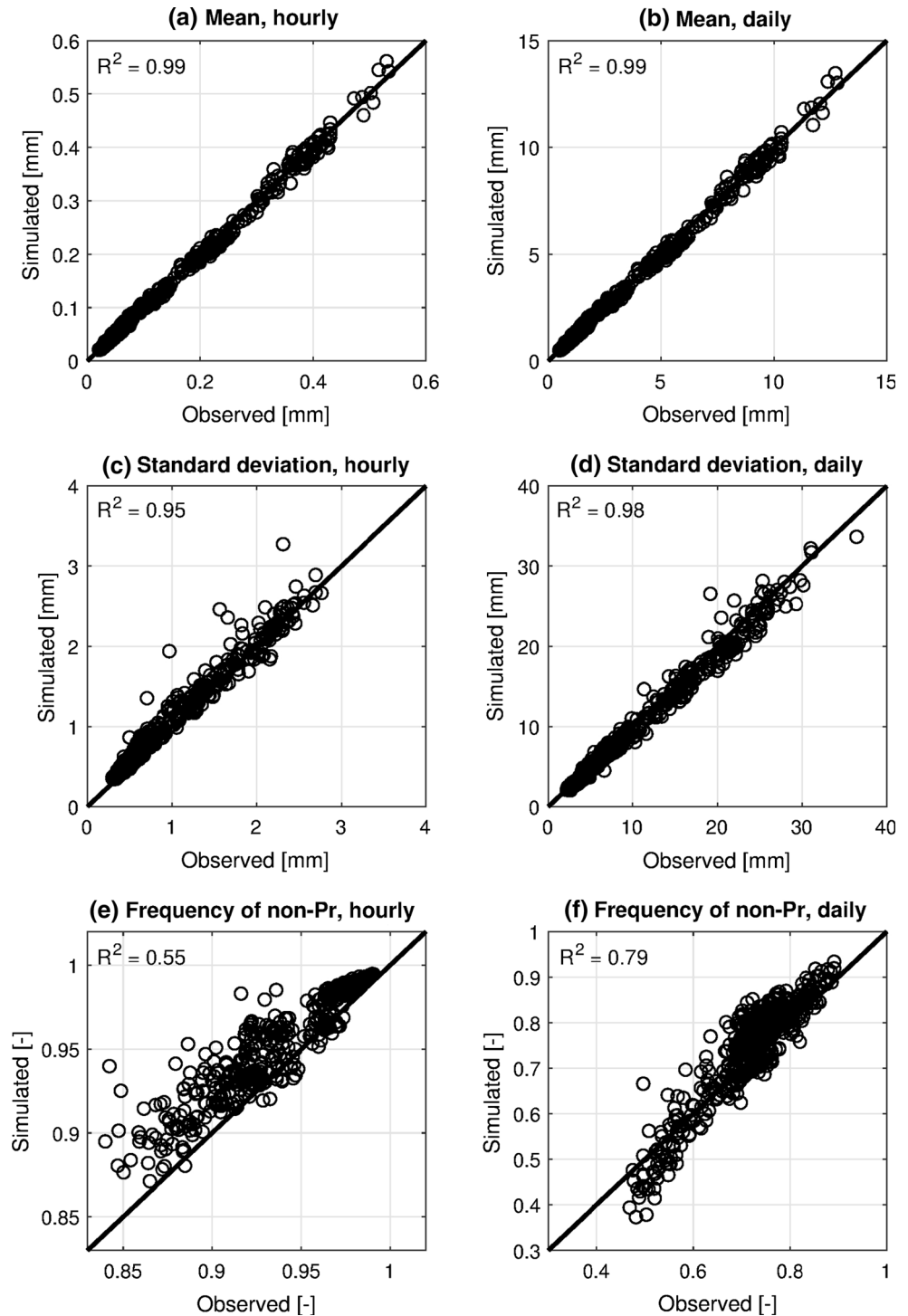
precipitation, i.e., if the amount of rain is very small (less than a threshold of 0.1 mm in this study), post-processing does not force rain, and the same could occur for the measurements. Next, the metrics for extreme precipitation are computed and compared in Fig. 4. The metrics involved are the maximum amount of precipitation during an hour and a day as well as the longest consecutive days with and without rain. For all 40 locations, the results by simulation and observation are comparable overall, and the results of underestimating the number of days that rain continuously are observed. The latter can also be related to the above issue of treatment, indicating that studies addressing a heavy rainfall and an associated flood phenomenon can use these results with high confidence, whereas those handling a rainfall occurrence should use these results carefully and further investigate the use of these results.

3.2 Evaluation of a stochastic ensemble generation

One reason for building an ensemble is that a realization is only one of many possibilities that could be occurring now or in the future. Therefore, generating the ensemble appropriately is one of the best ways to predict uncertainty associated with the randomness. As described in the Introduction, two of methods are used to generate ensemble members experiencing the same climate external conditions. Here, we attempt to evaluate the ensemble results generated by a stochastic weather generator (WG) with those by a well-known statistical approach, namely, the bootstrapping (BS) technique, which resamples the data points of a time series with repetitions. The latter method is adopted in a number of studies to investigate climate internal variability (Addor and Fischer 2015; Prudhomme and Davies 2009).

The first question is: given an external forcing condition, how much randomness can be reproduced by the ensemble generated? Therefore, we compared the hourly and daily maximum precipitation values corresponding to 95% of the 100 ensemble members with those of the 100- or 500-year return period. The latter values are borrowed from a ‘Korea Precipitation Frequency Data Server’ operated by the Korean government (<http://www.k-idf.re.kr/>) and estimated using appropriate data and methods. From the definition of the bootstrapping method, the values of annual maximum precipitation for all ensemble members should be theoretically the same as or less than those for the 30-year observation. As expected, Fig. 5 shows that most of the bootstrapped (BS) ensemble results predicted from the observed period are smaller than those of the 100-year return period. This tendency is more pronounced at an hourly scale. Alternatively, the ensemble results generated

Fig. 3 Comparison of the median of 100 ensembles simulated by the weather generator with the gauge observation at (left) hourly and (right) daily scales. The circles of all the subplots consist of 480 data points, which correspond to 12 months and 40 locations

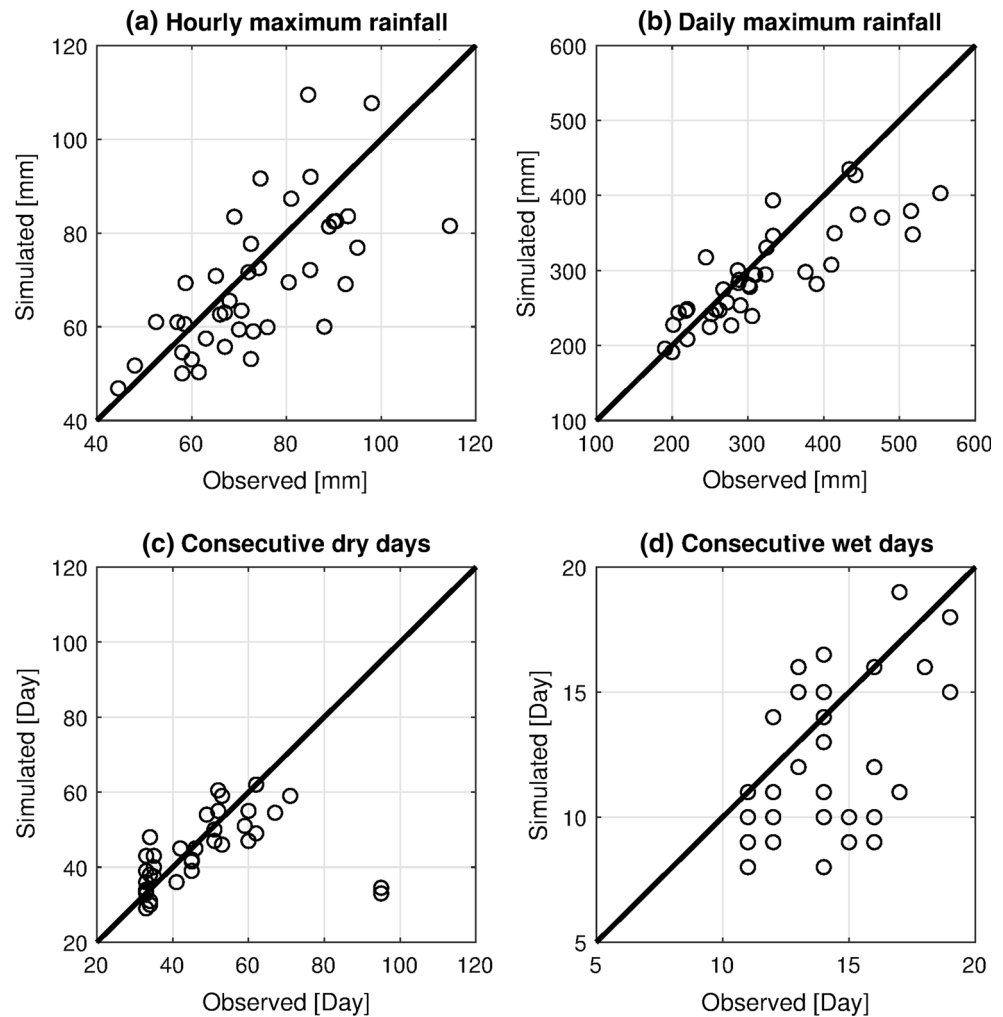


by the weather generator (WG) can be reproduced with the values of similar to and more than 100- or 500-year return period, even though they are generated from the same observations corresponding to 30 years.

Other than the maximum rainfall corresponding to the 95th percentile of 100 ensembles, one must investigate the distribution of the CIVs of the ensemble members. Here,

we simply compute the difference between the maximum and minimum (i.e., the range) of CIV values for each month and for the whole year over the 40 locations. Figure 6 shows a more detailed inspection of the data, involving a comparison of the range of CIVs estimated over the 100 ensemble generated by both the WG and BS methods. The ranges of the computed CIV values vary for

Fig. 4 Comparison of the median of 100 ensembles simulated by the weather generator with the gauge observation for the extreme precipitation metrics: **a** hourly maximum rainfall depth, **b** daily maximum rainfall depth, **c** the longest consecutive dry days, and **d** the longest consecutive wet days. The circles of all the subplots consist of 40 data points, which correspond to 40 locations

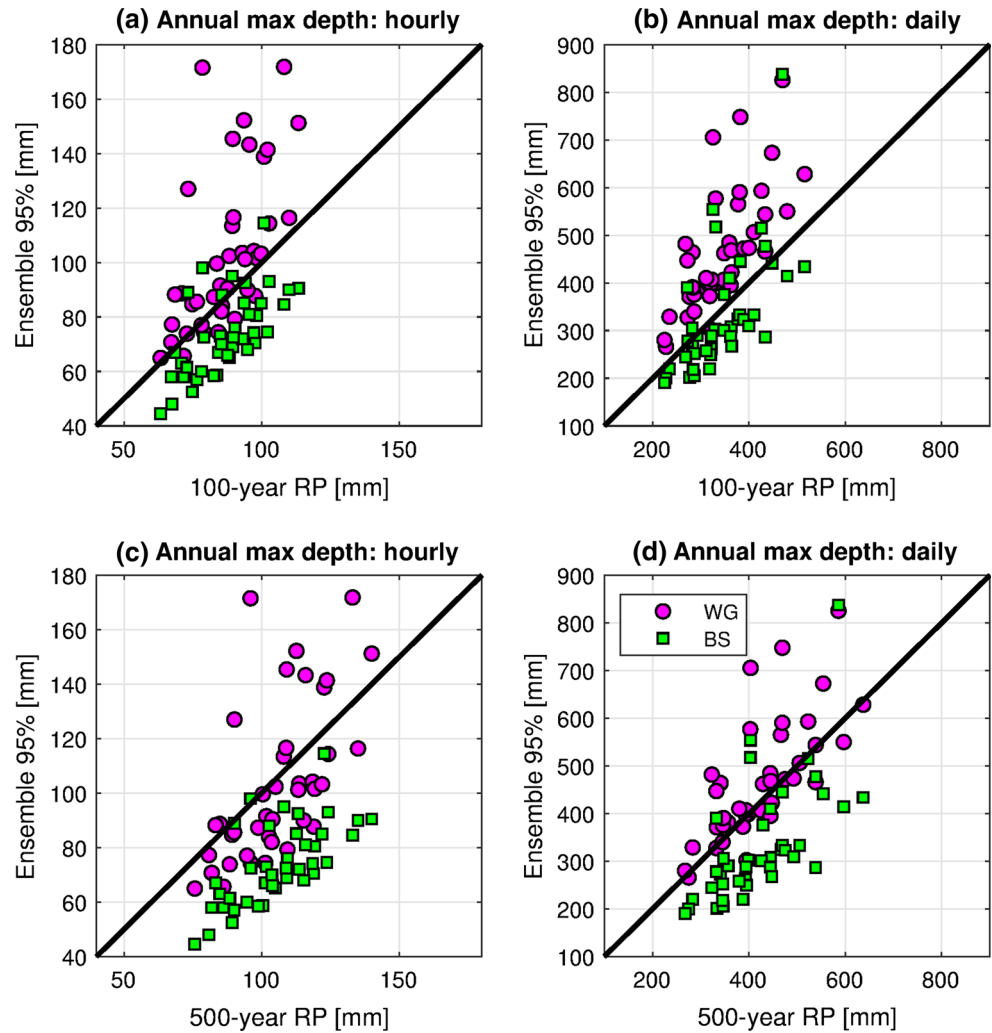


the four metrics: some of the ranges in the WG ensemble are large, whereas others are not. Such a mixed result can be further explicated when using the ratio of WG to BS range results of CIV (see the bottom plots of Fig. 6). If the ratio is greater than 1 (i.e., the right part of the kernel distribution), then the CIV range of WG is greater than that of BS—note that this portion fluctuates from 55.4% at the minimum for the ‘nonPr’ metric to 82.7% at the maximum for the ‘maxDa’ metric. How well does the range of CIV values by WG indicate wider variability? The first moment of the probability distribution of the ratio can answer this question. Little difference ($\sim 8\%$) is found between the CIV ranges estimated by two methods for the ‘nonPr’ metric, whereas in the metrics of ‘maxDa’ and ‘maxHr’, a difference of up to 60% is observed (see the bottom plots of Fig. 6). The latter indicates that the range of CIV values by WG is 60% wider than that of BS.

3.3 Estimation of climate internal variability

The main objective of the study is to estimate the climate internal variability over all of South Korea and visualize it on a spatial map with separated zones, thereby enabling one to effortlessly judge which region is more fluctuating than the climatological norm or in which region the probability of occurrence of the extreme event is high. As described in Sect. 2, the standard deviation of the residual is employed for estimating the climate internal variability in both ‘detrended’ and ‘differenced’ approaches. Depending on how the residual of time series is calculated and how the ensemble members are involved, the CIV results obtained using the two methods can vary. One might be interested in the degree to which the results of the two approaches and the results of the two ensemble generation methods are different, given the 4 metrics used in this study. Figure 7 illustrates a comparison of the CIV estimates computed by the two methods for both ensemble members generated by the weather generator (WG) and the bootstrapping (BS) for the metric of ‘totPr’. Over the 40

Fig. 5 Comparison of **a, c** the hourly and **b, d** the daily maximum precipitations corresponding to **a, b** the 100-year return period (RP) and **c, d** the 500-year return period with those corresponding to the 95th percentile of 100 ensemble obtained from using the weather generator (WG) and bootstrapping (BS). The circles of the subplots refer to the results of 40 locations



locations, the comparison is almost perfectly matched, with R^2 ranging from 0.97 to 0.99. For other metrics of ‘maxDa’, ‘maxHr’, and ‘nonPr’ (see Supplementary Material), the same conclusion is made: the R^2 values are close to 1, especially for ‘maxDa’ and ‘maxHr’ whereas they range from 0.78 to 0.97 for ‘nonPr’. In other words, the method used to calculate the CIV has no significance to the results.

Examining the CIV values, the internal variability of annual total precipitation (‘totPr’) is 313.20 mm on average and ranges from 239.73 to 560.35 mm over the country when the observed ASOS data is used, and its internal variability can be enlarged into 368.03 mm, ranging from 292.29 to 597.50 mm if the time series corresponding to the 95th percentile of ensemble members is used. Refer to Table 2 for the exact minimum, maximum, and mean values of the 40 locations that correspond to the 16 subplots of Fig. 8. From Table 2, the spatial mean of the CIVs predicted from the observations in ‘nonPr’ are comparable to that of CIVs corresponding to the 95th percentile value

among 100 ensembles; however, the difference between them in ‘totPr’, ‘maxDa’, and ‘maxHr’ is approximately 20, 40, and 30%, respectively. Obviously, this difference varies (is even larger) according to location. Furthermore, the month in which the internal variability is maximized is dissimilar, depending on the metrics at different scales. For example, the monthly ‘OBS’ values of ‘totPr’ and ‘maxDa’ are found to be largest in August, whereas the largest values of ‘maxHr’ are in September (see Supplementary Material).

If we analyze the spatial pattern of the calculated CIV values with the locations and geographical features of S. Korea (e.g., latitude and elevation), no particular pattern is identified (not shown). Therefore, an unsupervised, clustering technique is introduced to distribute the domain into distinctive zones. In particular, the *K*-means cluster algorithm is applied to the CIV values of 40 locations for each month and for the whole year that will be classified into the *K* mutually exclusive clusters. According to the Elbow method that can determine the number of clusters, the

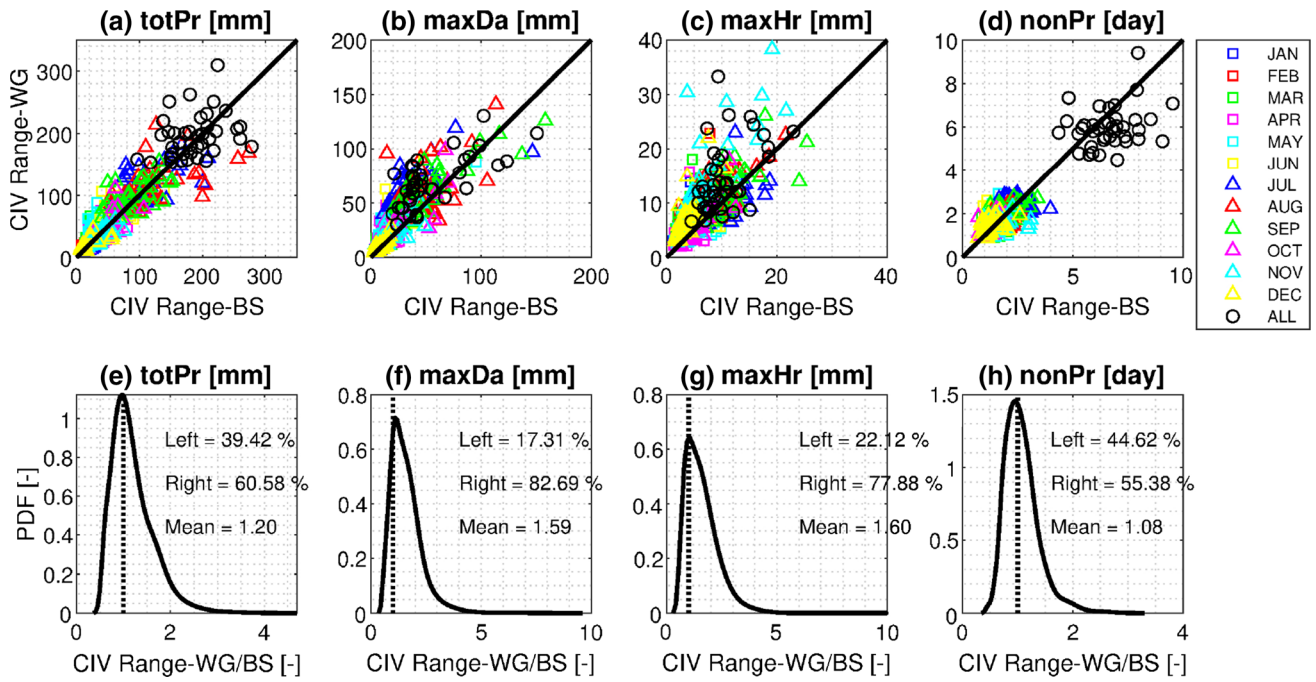


Fig. 6 Comparison for the range (difference between maximum and minimum) of CIVs computed using an ensemble of bootstrapping (BS) and weather generator (WG) for the four metrics. The subplots **a–d** include the data points of 520 = 13 (12 months + year) times 40

(locations). The bottom subplots represent the non-parametric kernel distributions for the ratio of the CIV range values. The ‘detrended’ approach is used

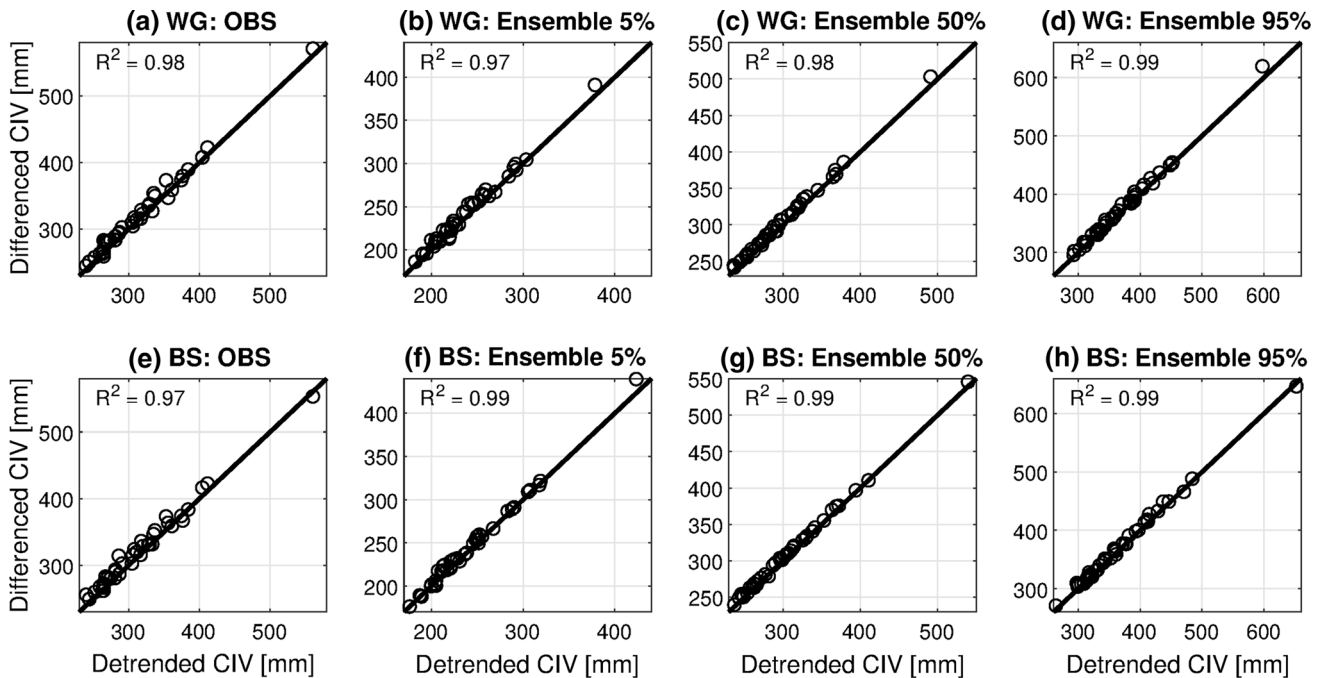


Fig. 7 Comparison of the CIV computed by the ‘detrended’ and ‘differenced’ approaches. The CIV is estimated for the metric of ‘totPr’. The ensembles are generated from the weather generator

(WG) and the bootstrapping (BS). The circles of the subplots refer to the results of 40 locations. The ‘OBS’ refers to observations

optimum “*K*” (elbow) ranges from 3 to 6 depending on the 208 cases (i.e., 16 subplots of 13 figures—Fig. 8 and Fig. S4 to S15), but 4 or 5 are the most frequent numbers

for the *K*. Thus, we choose the number of clusters as 4 (*K* = 4) in this study. Then, the following results are obtained. (1) The internal variability is relatively higher for

Table 2 The minimum, maximum (in the first line of each cell), and mean (in the second line) CIV values over the 40 locations

	OBS	Ensemble 5%	Ensemble 50%	Ensemble 95%
totPr (mm)	239.73–560.35	182.37–378.37	236.67–490.85	292.29–597.50
	313.20	236.71	299.50	368.03
maxDa (mm)	38.35–148.05	31.21–71.02	39.86–99.94	52.04–165.54
	65.73	48.07	63.11	91.92
maxHr (mm)	9.92–23.11	6.99–17.34	9.40–24.53	12.59–34.64
	15.14	11.15	14.37	19.69
nonPr (day)	9.11–12.58	6.07–7.82	7.47–10.01	9.12–12.58
	11.11	7.03	8.84	10.83

Each cell corresponds to the values of 16 subplots in Fig. 8

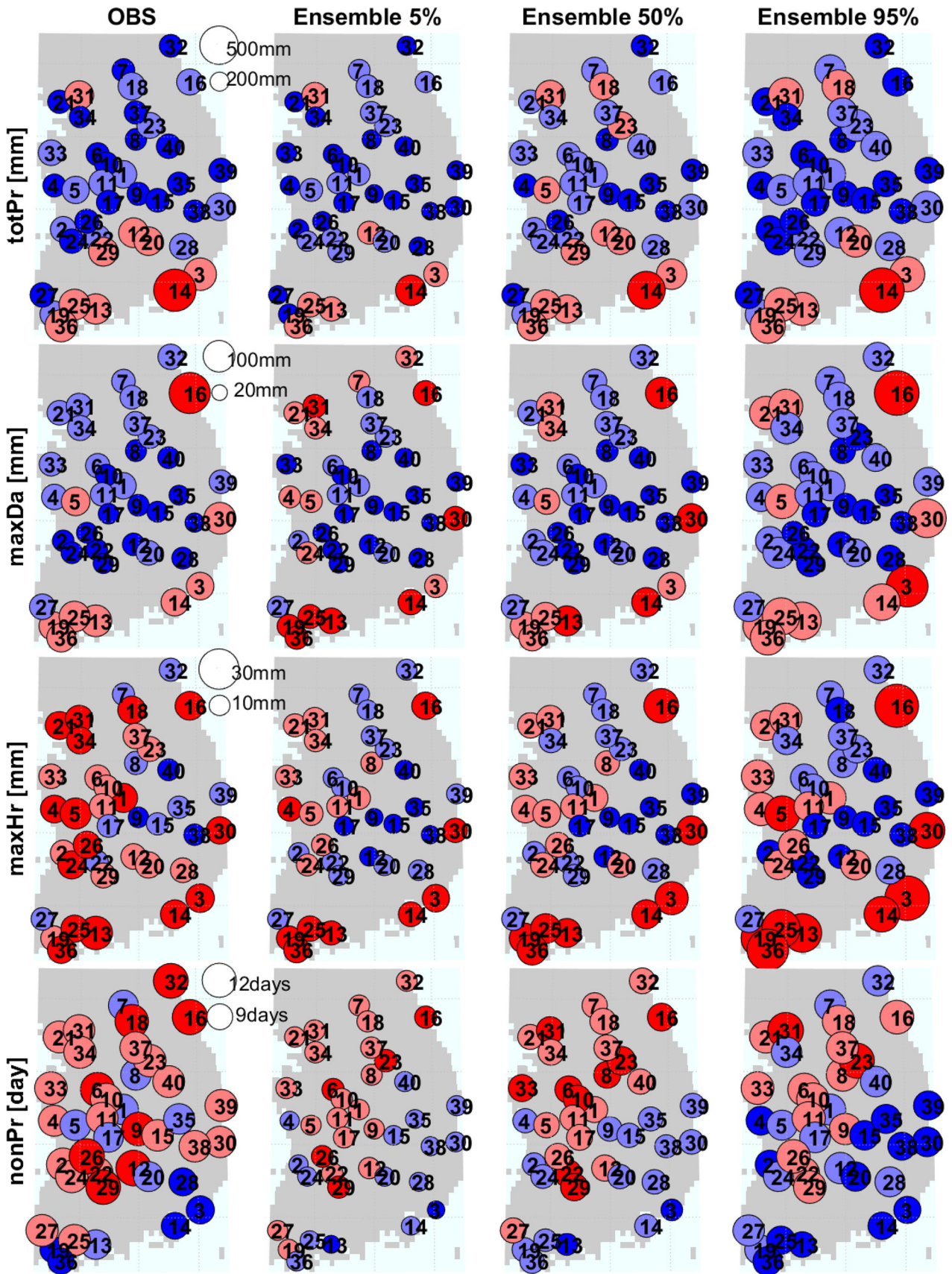
the stations located in the southern and eastern coastal regions of the country in the annual metrics of ‘totPr’, ‘maxDa’, and ‘maxHr’, whereas the variability is higher in the central inland province of S. Korea for ‘nonPr’ (see Fig. 8). (2) For the rainy summer season from July to September, the spatial zone representing higher CIV is changed into the southern and western coast region for ‘maxHr’. In addition, the zoning shows a mixed pattern for each month of the rainy season for ‘maxDa’ and ‘totPr’. For example, in August, when the largest internal variability is observed, the spatial variation of CIV among the locations in ‘maxDa’ is relatively homogeneous. (3) Last, examining the CIV of ‘nonPr’ over the recent severe spring drought period (e.g., May and June), one can see that the CIV is relatively high in the northern and western part of the country (specifically, adjacent to the southwestern island and ‘Chungnam’, ‘Gyeonggi’, and ‘Gangwon’ province).

4 Discussions and conclusions

This work attempted to assess the climate internal variability (CIV) over South Korea using an hourly weather generator, i.e., the Advanced WEather GENerator (AWE-GEN)—100 ensembles of 30 years time series are generated for 40 gauge stations. Before the ensemble members are applied to estimate the CIV, the results of the AWE-GEN are evaluated with those of the Automated Surface Observing System (ASOS) to check whether the model reproduces the equivalent precipitation statistics for the first and second moment properties, the occurrences, and the extreme values over the stations. The statistics of the mean, standard deviation and maximum rainfall computed at hourly and daily scales are in good agreement, whereas those related to the occurrence of rainfall are not as satisfactory as the former statistics. Because of the concurrence in the comparisons of the former statistics, the reason for this observation could be how one treats the occurrence of smaller precipitation, i.e., if the rain is very small (less than

a threshold of 0.1 mm in this study), post-processing does not force rain, and the same issue could occur during measurement. Another reason for this observation could be the inability of the optimization method used to match all the 16 values required to calculate the NSRP parameters at the same time. Because perfect agreement is nearly impossible, a certain error in a statistical property is unavoidable.

The climatic characteristics involved with stochastic randomness necessitates the use of ensembles to quantify the degrees of the random fluctuation. We generate 100 ensembles through two popular approaches: a stochastic weather generator (WG) and a statistical bootstrapping (BS) technique. Both approaches are suited for generating as many samples as possible; however, the latter approach, which resamples an original data with repetitions, theoretically has a limitation that any value estimated from the ensemble is always the same as or less than that from the observed data (Costa et al. 2015). In this regard, Figs. 5 and 6 are used to examine how much randomness is reproduced by the ensemble generated by the two approaches. The ensemble of the maximum precipitation depths of the 30-year time series for the two durations of an hour and a day are compared with precipitation estimates of frequency corresponding to 100- and 500-year return periods (Fig. 5). Most of the bootstrapped (BS) results corresponding to the 95% of the 100 ensemble members are smaller than those of the 100-year return period—this tendency is more pronounced at an hourly scale. In contrast, the ensemble results generated by the weather generator (WG) are mostly comparable to or can be occasionally larger than the precipitation frequency estimates of the 500-year return period, even though they are generated from the same observations corresponding to 30 years. Likewise, a consistent conclusion can be drawn by examining the differences between the maximum and minimum values of the CIVs over the ensemble members and by computing the ratio of the CIV ranges of the WG to BS methods. The value of the ratio greater than 1 indicates that the CIV range of the WG is larger than that of the BS:



◀ **Fig. 8** Spatial distribution of the CIV values over South Korea for the 4 metrics computed over the whole year. Four clusters [from blue (as the smallest cluster), light blue, light red to red (as the largest)] are identified after applying the *K*-means algorithm. The number within each circle represents the station number summarized in Table 1, and the number near white circles is the legend to show the relative magnitude of the metrics. The ensemble is generated from the weather generator, and the ‘detrended’ method is used

for most (e.g., from 55.4% at the minimum for the ‘nonPr’ metric to 82.7% in the ‘maxDa’ metric) of the months and the locations, the ensemble by WG has higher chances of representing more diverse stochastic realizations and thus broader CIV ranges than the BS results. The question of how much the range of CIV values by WG signifies wider variability is answered from the first moment of the probability distribution of the ratio: there is little difference ($\sim 8\%$) between the CIV ranges for the ‘nonPr’ metric, whereas in the metrics of ‘maxDa’ and ‘maxHr’, a significant difference of up to 60% occurs (see the bottom plots of Fig. 6). Here, an important implication is that the use of the bootstrapped technique in estimating CIV can be beneficial in partial situations as an alternative estimator, unless other alternatives exist. However, note that the BS estimation of CIV fails to sufficiently represent the proper uncertain range, i.e., the CIV can be underestimated for extreme statistics, such as the maximum rainfall depth.

In addition to the issue that the degree of variability in BS is not fully investigated, the performance of bootstrapped samples can be far from satisfactory for time series data with serial correlations. Thus, it is worth checking whether the time series used in this study are serially correlated by computing the lag-1 autocorrelation. Approximately 46.6% of the 2080 data points ($40 \times 13 \times 4$) belong to the range within ± 0.1 of the lag-1 autocorrelation, whereas approximately 4.8% are greater than ± 0.3 ; and the independence tendency is more pronounced in the metrics of ‘maxDa’ and ‘maxHr’, whereas the tendency of correlation is more pronounced in the ‘nonPr’ metric (Fig. 9). Although most data show values of the lag-1 autocorrelation that are not large, this does not guarantee that the effect of serial correlation is inconsequential. Although bootstrapping is a very popular technique, one should keep in mind that this correlation effect exists when sampling the time series repeatedly. In contrast, WG is a technique for generating a time series by combining random numbers for the five probability distributions. Therefore, it can be assumed that the generated ensembles are independent from each other and a more stable predictive metric value can be provided.

The spatial mean of CIVs predicted from the 30-year observations in ‘nonPr’ is almost comparable to that of CIVs corresponding to the 95th percentile value among

100 ensembles; however, the differences between them in ‘totPr’, ‘maxDa’, and ‘maxHr’ extend up to approximately 20, 40, and 30%, respectively, whereas they approach zero when compared to the 50th percentile value (Table 2). Obviously, this difference varies (is even larger) according to the location. Additionally, seasonal variations of the spatial pattern are found to depend on the metrics and temporal scales: the month in which the internal variability is maximized is dissimilar. For example, the monthly ‘OBS’ values of ‘totPr’ and ‘maxDa’ are greatest in August, whereas those of ‘maxHr’ are highest in September (see Supplementary Material). Consequently, such a scale-dependent phenomenon requires careful attention when identifying the magnitude and the seasonal pattern of the internal variability before the use of an hourly time series and the associated extreme properties.

After applying an unsupervised *K*-means cluster algorithm, the locations over South Korea are sorted into four distinct zones according to the magnitude of their CIVs. Spatial patterns are identified regarding which regions belong to a greater variability zone and how they shift monthly: (1) the internal variability is relatively prominent for the stations located in the southern and eastern coastal regions of the country in the annual metrics of ‘totPr’, ‘maxDa’, and ‘maxHr’, and the variability is high in the central inland province of S. Korea for ‘nonPr’ (see Fig. 8). (2) For the rainy season from July to September, the spatial zones representing higher CIV become changed into southern and western coast regions for ‘maxHr’. Conversely, the zoning shows a mixed pattern for each month of the rainy season in ‘maxDa’ and ‘totPr’. For example, in August, when the largest internal variability is observed, the spatial variation of CIV among the locations in ‘maxDa’ is relatively homogeneous. (3) Last, examining the CIV of ‘nonPr’ over the recent severe spring drought period (e.g., May and June), the CIV is found to be relatively high in the northern and western part of the country.

Regarding the projection of signals for the future associated with external forcings, there is room for further improvement and development. It has however been reported that among many uncertainty sources in climate change studies, the uncertainty bound of climate internal variability is only irreducible, making future climate projections more difficult and thus making their predictability poorer (Addor and Fischer 2015; Deser et al. 2012b; Hawkins and Sutton 2009). Therefore, one of the pragmatic solutions is to estimate the noise related to internal variability efficiently using a stochastic ensemble, following our proposed approach, and the signal is assuredly projected from the results using either global or regional climate models. Because of a recent implication that the uncertainty bounds of CIV for the present and for the future can be comparable (Fatichi et al. 2016; Kim et al. 2016b),

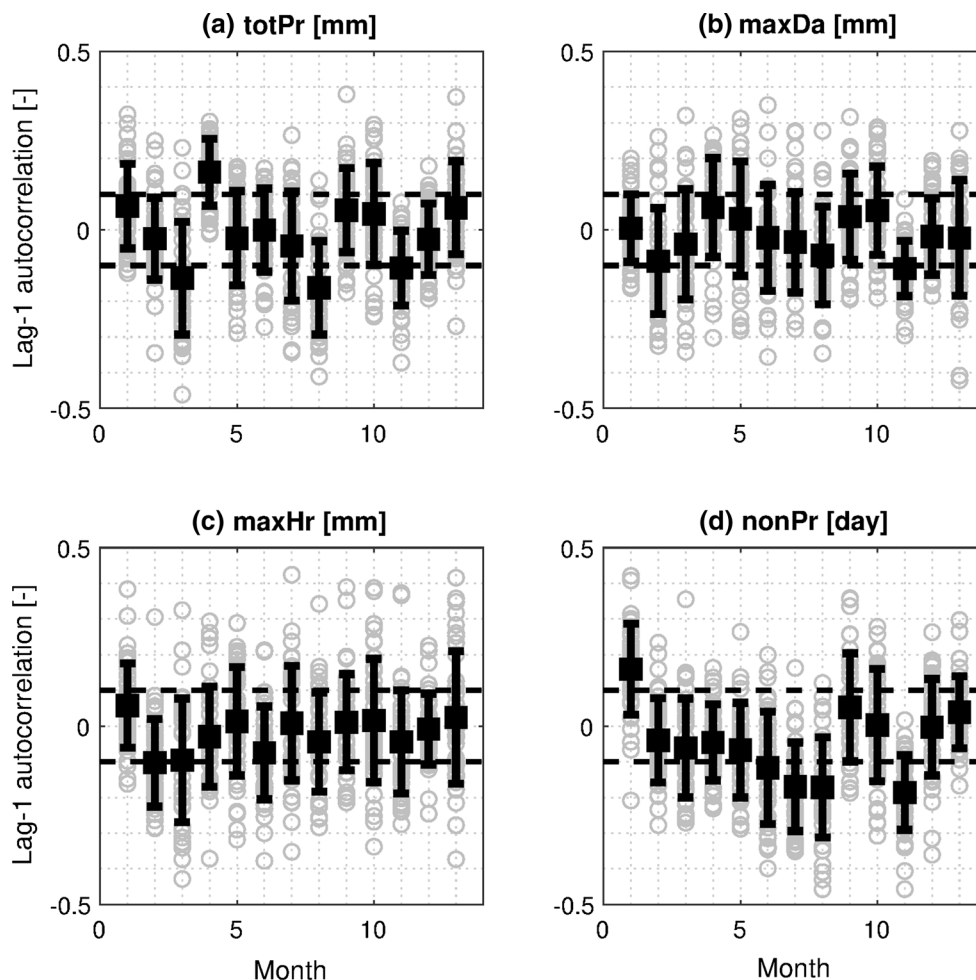


Fig. 9 The lag-1 autocorrelation of **a** totPr, **b** maxDa, **c** maxHr, and **d** nonPr for 12 months and the whole year. The latter corresponds to the “13” on the *x*-axis. The light gray circles represent the lag-1

autocorrelation for 40 locations, and the black error bars show the mean and standard deviation of the 40 values. In each plot, 520 data points (40×13) are illustrated

the investigation of CIV for the present is commendable because it can be employed as a proxy estimator of the future. In summary, our findings regarding the CIV will be valuable for identifying which regions have high variability compared to climatological norms and thus are more vulnerable to extreme conditions influenced by internal structures as well as external factors. Such information will be ultimately helpful for planning future adaptation and mitigation measures against extreme events.

Acknowledgements This work was supported by the Korean Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 17AWMP-B083066-04) and the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (No. 2011-0030040).

References

- Addor N, Fischer EM (2015) The influence of natural variability and interpolation errors on bias characterization in RCM simulations. *J Geophys Res Atmos* 120(19):10180–110195. <https://doi.org/10.1002/2014jd022824>
- Brisson E, Demuzere M, Willems P, van Lipzig NPM (2015) Assessment of natural climate variability using a weather generator. *Clim Dyn* 44(1):495–508. <https://doi.org/10.1007/s00382-014-2122-8>
- Chen J, Brissette FP, Leconte R (2011) Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J Hydrol* 401(3):190–202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>
- Costa V, Fernandes W, Naghettini M (2015) A Bayesian model for stochastic generation of daily precipitation using an upper-bounded distribution function. *Stoch Environ Res Risk Assess* 29(2):563–576. <https://doi.org/10.1007/s00477-014-0880-9>
- Coulthard TJ, Ramirez J, Fowler HJ, Glenis V (2012) Using the UKCP09 probabilistic scenarios to model the amplified impact of climate change on drainage basin sediment yield. *Hydrol Earth Syst Sci* 16(11):4401–4416. <https://doi.org/10.5194/hess-16-4401-2012>

- Cowpertwait P, Onof C, Isham V (2007) Point process models of rainfall: developments for fine-scale structure. *Proc R Soc Lond Ser A* 463(2086):2569–2587. <https://doi.org/10.1098/rspa.2007.1889>
- Deser C, Knutti R, Solomon S, Phillips AS (2012a) Communication of the role of natural variability in future North American climate. *Nat Clim Change* 2(11):775–779. <http://www.nature.com/nclimate/journal/v2/n11/abs/nclimate1562.html#supplementary-information>
- Deser C, Phillips A, Bourdette V, Teng H (2012b) Uncertainty in climate change projections: the role of internal variability. *Clim Dyn* 38(3–4):527–546. <https://doi.org/10.1007/s00382-010-0977-x>
- Deser C, Phillips AS, Alexander MA, Smoliak BV (2014) Projecting North American climate over the next 50 years: uncertainty due to internal variability. *J Clim* 27(6):2271–2296. <https://doi.org/10.1175/jcli-d-13-00451.1>
- Fatichi S, Ivanov VY, Caporali E (2011) Simulation of future climate scenarios with a weather generator. *Adv Water Resour* 34(4):448–467. <https://doi.org/10.1016/j.advwatres.2010.12.013>
- Fatichi S, Ivanov VY, Caporali E (2013) Assessment of a stochastic downscaling methodology in generating an ensemble of hourly future climate time series. *Clim Dyn* 40(7–8):1841–1861. <https://doi.org/10.1007/s00382-012-1627-2>
- Fatichi S, Ivanov VY, Paschalis A, Peleg N, Molnar P, Rimkus S, Kim J, Burlando P, Caporali E (2016) Uncertainty partition challenges the predictability of vital details of climate change. *Earths Future* 4(5):240–251. <https://doi.org/10.1002/2015ef000336>
- Fischer EM, Beyerle U, Knutti R (2013) Robust spatially aggregated projections of climate extremes. *Nat Clim Change* 3(12):1033–1038. <https://doi.org/10.1038/nclimate2051>. <http://www.nature.com/nclimate/journal/v3/n12/abs/nclimate2051.html#supplementary-information>
- Frankcombe LM, England MH, Mann ME, Steinman BA (2015) Separating internal variability from the externally forced climate response. *J Clim* 28(20):8184–8202. <https://doi.org/10.1175/jcli-d-15-0069.1>
- Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J Clim* 15(10):1141–1158. [https://doi.org/10.1175/1520-0442\(2002\)015%3c1141:coaura%3e2.0.co;2](https://doi.org/10.1175/1520-0442(2002)015%3c1141:coaura%3e2.0.co;2)
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90(8):1095–1107. <https://doi.org/10.1175/2009bams2607.1>
- Hingray B, Said M (2014) Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections. *J Clim* 27(17):6779–6798
- IPCC (2013) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, 1535 pp. Cambridge University Press, Cambridge. <https://doi.org/10.1017/cbo9781107415324>
- Ivanov VY, Bras RL, Curtis DC (2007) A weather generator for hydrological, ecological, and agricultural applications. *Water Resour Res* 43(10):21. <https://doi.org/10.1029/2006wr005364>
- Khalili K, Tahoudi MN, Mirabbasi R, Ahmadi F (2016) Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stoch Environ Res Risk Assess* 30(4):1205–1221. <https://doi.org/10.1007/s00477-015-1095-4>
- Kim J, Ivanov VY (2014) On the nonuniqueness of sediment yield at the catchment scale: the effects of soil antecedent conditions and surface shield. *Water Resour Res* 50(2):1025–1045. <https://doi.org/10.1002/2013wr014580>
- Kim J, Ivanov VY (2015) A holistic, multi-scale dynamic downscaling framework for climate impact assessments and challenges of addressing finer-scale watershed dynamics. *J Hydrol* 522:645–660. <https://doi.org/10.1016/j.jhydrol.2015.01.025>
- Kim J, Dwelle MC, Kampf SK, Fatichi S, Ivanov VY (2016a) On the non-uniqueness of the hydro-geomorphic responses in a zero-order catchment with respect to soil moisture. *Adv Water Resour* 92:73–89. <https://doi.org/10.1016/j.advwatres.2016.03.019>
- Kim J, Ivanov VY, Fatichi S (2016b) Climate change and uncertainty assessment over a hydroclimatic transect of Michigan. *Stoch Environ Res Risk Assess* 30(3):923–944. <https://doi.org/10.1007/s00477-015-1097-2>
- Kim J, Ivanov VY, Fatichi S (2016c) Environmental stochasticity controls soil erosion variability. *Sci Rep* 6:22065. <https://doi.org/10.1038/srep22065>. <http://www.nature.com/articles/srep22065#supplementary-information>
- Kim J, Ivanov VY, Fatichi S (2016d) Soil erosion assessment: mind the gap. *Geophys Res Lett* 43(24):12446–412456. <https://doi.org/10.1002/2016gl071480>
- Maraun D et al (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys*. <https://doi.org/10.1029/2009rg000314>
- Mohammed R, Scholz M, Nanekely MA, Mokhtari Y (2016) Assessment of models predicting anthropogenic interventions and climate variability on surface runoff of the Lower Zab River. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-016-1375-7>
- Moise AF, Hudson DA (2008) Probabilistic predictions of climate change for Australia and southern Africa using the reliability ensemble average of IPCC CMIP3 model simulations. *J Geophys Res Atmos*. <https://doi.org/10.1029/2007jd009250>
- Ng JL, Abd Aziz S, Huang YF, Wayayok A, Rowshon MK (2017) Stochastic modelling of seasonal and yearly rainfalls with low-frequency variability. *Stoch Environ Res Risk Assess* 31(9):2215–2233. <https://doi.org/10.1007/s00477-016-1373-9>
- Prudhomme C, Davies H (2009) Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. *Clim Change* 93(1):177–195. <https://doi.org/10.1007/s10584-008-9464-3>
- Singh V, Goyal MK (2017) Spatio-temporal heterogeneity and changes in extreme precipitation over eastern Himalayan catchments India. *Stoch Environ Res Risk Assess* 31(10):2527–2546. <https://doi.org/10.1007/s00477-016-1350-3>
- Thompson DW, Barnes EA, Deser C, Foust WE, Phillips AS (2015) Quantifying the role of internal climate variability in future climate trends. *J Clim* 28(16):6443–6456
- Yao S-L, Luo J-J, Huang G (2016) Internal variability-generated uncertainty in east asian climate projections estimated with 40 CCSM3 ensembles. *PLoS ONE* 11(3):e0149968. <https://doi.org/10.1371/journal.pone.0149968>
- Zhang L, Wang C (2013) Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. *J Geophys Res Oceans* 118(10):5772–5791. <https://doi.org/10.1002/jgrc.20390>