Development of a Pressure–Precipitation Transmitter

MASARU INATSU

Faculty of Science, and Center for Natural Hazards Research, Hokkaido University, Sapporo, Japan

TAMAKI SUEMATSU

Atmosphere Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan

YUTA TAMAKI

Obihiro Weather Station, Japan Meteorological Agency, Obihiro, Japan

NAOTO NAKANO

Institute for Liberal Arts and Sciences, Kyoto University, Kyoto, Japan

KAO MIZUSHIMA AND MIZUKI SHINOHARA

Tokio Marine and Nichido Risk Consulting Co., Ltd., Tokyo, Japan

(Manuscript received 19 March 2019, in final form 25 July 2019)

ABSTRACT

A novel method is proposed to create very long term daily precipitation data for the extreme statistics by computing very long term daily sea level pressure (SLP) with the SLP emulator (a statistical multilevel regression model) and then converting the SLP into precipitation by combining statistical downscaling methods of the analog ensemble and singular value decomposition (SVD). After a review of the SLP emulator, we present a multilevel regression model constructed for each month that is based on a time series of 1000 principal components of SLPs on global reanalysis data. Simple integration of the SLP emulator provides 100-yr daily SLP data, which are temporally interpolated into a 6-h interval. Next, the pressure–precipitation transmitter (PPT) is developed to convert 6-hourly SLP to daily precipitation. The PPT makes its first-guess estimate from a composite of time frames with analogous SLP transition patterns in the learning period. The departure of SLPs from the analog ensemble is then corrected with an SVD relationship between SLPs and precipitation. The final product showed a fairly realistic precipitation pattern, displaying temporal and spatial continuity. The annual-maximum precipitation of the estimated 100-yr data extended the tail of probability distribution of the 8-yr learning data.

1. Introduction

The extreme-value statistics have been required in various sectors such as hydrological management, erosion control, traffic engineering, and disaster insurance. In most extreme-event assessments, the return period of extreme precipitation event is computed from annual-maximum daily precipitation based on a very long term daily precipitation data and becomes an input parameter for each-sector modeling. Climate simulation studies have

recently provided physically consistent atmosphere and ocean data at a length on the order of 1000 years in the total of all the ensembles. Phase 5 of the Coupled Model Intercomparison Project (Taylor et al. 2012) joined more than 50 climate models, each of which provided several ensemble experiments, with a historical record before 2005 and some with representative trajectories of greenhouse gas concentrations during the twenty-first century. Moreover, a Japanese team realized a set of massive ensemble experiments called the d4PDF using an atmospheric general circulation model with 60-km grid spacing. They further conducted a dynamical downscaling (DDS) with 20-km grid resolution over an area

Corresponding author: Prof. Masaru Inatsu, inaz@sci.hokudai.ac.jp

covering Japan (Mizuta et al. 2017). The d4PDF's DDS data length is 3000 years in total, which was applied to the risk assessment of extreme rainfall and flash flood over the Tokachi and Tokoro Rivers in Hokkaido, Japan (Yamada et al. 2018).

In the field of disaster insurance, it is desirable that much infrequent extreme events be evaluated from precipitation data with 10 000 years or longer. However, the computational cost of an atmospheric general circulation model experiment and a subsequent DDS makes integration of 10 000 years infeasible. This kind of requirement can be met with a stochastic weather generator (SWG), which is a computationally inexpensive model designed to generate synthetic weather time series by replicating the statistics of observed precipitation at a specific location. According to a brief review of SWG in Ailliot et al. (2015), SWG for rainfall has two components, a model for determining precipitation days using a two-value Markov chain, and a model for calculating daily precipitation amount with a parametric (Richardson 1981) or nonparametric (Sharma and Lall 1999) fitting to the observed data. Extending a single-site SWG, Wilks (1998) and Wilks (1999) introduced a multisite generalization model with a parametric SWG that describes spatial correlation in precipitation amount between sites. Moreover, Wilby et al. (2003) introduced a multisite nonparametric SWG. Despite devotions of many studies to improve SWGs to match many aspects of the observed precipitation data (Baigorria and Jones 2010; Smith et al. 2018), the reproduction of extreme and long-lasting events of precipitation with SWG still remains to be a problem (Ailliot et al. 2015).

On the other hand, a multilevel regression (MLR) model was constructed to reproduce statistics that are similar to the observed statistics of midtropospheric streamfunction in a quasigeostrophic framework (Kravtsov et al. 2005; Kondrashov et al. 2006). Kravtsov et al. (2016, hereinafter K16) extended their idea to sea level pressure (SLP) data, which was named the SLP emulator. The forecast data with the SLP emulator produces independent realizations of SLP variability, quite similar to that observed, with regard to several statistical measures including spatial patterns of bandpass- and low-pass-filtered variability as well as extratropical cyclone tracks. This model also has the strength of computational inexpensiveness; for example, it accomplishes a 10 000-yr integration within 24 h on an ordinary computer server. This motivated us to take advantage of the SLP emulator to create a very long term precipitation data. However, a mere replacement of SLP to precipitation in the SLP emulator is unlikely to work well, because the precipitation data will be decomposed into a large number of modes to describe fine,

local patterns. Moreover, such replacement may suffer from reproducing negative precipitation values in the process of pattern reconstruction. Therefore, we introduced a statistical downscaling (SDS) method, which was expected to suitably convert the SLP data to precipitation data.

In the midlatitudes where a precipitation pattern is strongly related to a weather chart, the analog method, originally used in the field of forecasting (Lorenz 1969; Barnett and Preisendorfer 1978; Branstator et al. 2012; Ichikawa and Inatsu 2017), efficiently reconstructs the precipitation field from a weather pattern with comparable performance to other more complicated SDS (Zorita and von Storch 1999). Following their idea, Delle Monache et al. (2013) explored an analog ensemble (AnEn) that estimated the probability distribution of the climatic variable with a lead time from a set of past observations that correspond to the best analogs of a deterministic numerical weather prediction. For example, Blanchet et al. (2018) related large rainfall accumulation to a multiday sequence of atmospheric circulation patterns. Other SDS methods are mostly based on multivariate regression (Hellström et al. 2001), singular value decomposition (SVD; Kuno and Inatsu 2014), or canonical correlation analysis (Busuioc et al. 2001). See Fowler et al. (2007) for a conceptual classification of SDS methods.

The purpose of this study is to develop a new method to create a very long term daily precipitation data for the extreme statistics. Motivated by the SLP emulator and simple SDS methods such as AnEn, we propose a combination of their methods for our purpose. The strategy is to first compute 100-yr global SLP data, closely following K16, and then interpolate them to 6-hourly intervals. The precipitation is composited over analog dates when the SLP transition pattern is similar to the reference data within the target area of 120°–150°E, 22.5°–47.5°N including the whole Japanese territory except for Marcus Island. The key idea of our method to extend the tail of probability distribution from short-length data is to estimate the first guess of daily precipitation by taking the sum of four 6-hourly precipitation analogs. Empirical SVD modes between SLPs and precipitation over the target domain are also utilized to correct the fitting error in the AnEn. This proposed procedure including the analog composite and the SVD correction is hereinafter named the pressure–precipitation transmitter (PPT).

The rest of this paper is organized as follows: section 2 introduces the used analysis data; section 3 and the appendixes provide a self-contained description of the simplified SLP emulator. We propose the PPT with a simple test to check its performance in section 4, and we

describe the statistics of daily precipitation data from the PPT, based on 100-yr SLP data computed with the SLP emulator, in [section 5](#). [Section 6](#) gives a summary of the paper.

2. Data

We utilized the SLP dataset from January 1950 to December 2017 from Japanese reanalysis dataset JRA-55 ([Kobayashi et al. 2015](#)), over the Northern Hemisphere at $1.25^\circ \times 1.25^\circ$ grid spacing, averaged to daily means from the original 6-hourly output. The SLP data have 288×72 spatial grid points and the temporal length of 21 185 days. The daily climatology was defined in this study as the year-to-year average of the 31-day running mean of the original SLP ([Fig. 1a](#)). The SLP anomaly was created by subtracting the daily climatology from the daily value. We also used the precipitation dataset from January 2010 to December 2017 from the mesoscale model (MSM) analysis provided by the Japan Meteorological Agency (JMA), over the Japanese territory ([Saito et al. 2006](#)), averaged to daily mean from the original 3-hourly output. The resolution of the horizontal grid mesh of the original data is aggregated to 20 km from approximately 5 km. The precipitation data had 133×101 spatial grid points and a temporal length of 2922 days. We reconstructed precipitation data on 1023 land grid points within the Japanese territory, including small islands.

3. A simplified SLP emulator

As the first step of the construction of a simplified SLP emulator proposed in [K16](#), we computed the empirical orthogonal function (EOF) patterns of daily SLP anomaly throughout the period from 1960 to 2017, regardless of the seasons. The gravest EOF modes were teleconnection patterns identified in many previous studies ([Figs. 2a–c](#); [Wallace and Gutzler 1981](#); [Barnston and Livezey 1987](#)). The first mode had a high projection onto the North Atlantic Oscillation ([Hurrell et al. 2003](#)), and the third and fifth modes resembled the Pacific North American pattern. The SLP anomaly data can be precisely reconstructed by the 1000 leading modes of the EOF decomposition, which explain 99.87% of the total variance of the SLP anomaly. Each principal component (PC) mode was calculated by projecting daily SLP anomaly data onto each EOF pattern. Here the PCs were normalized so that the temporal variance is unity, and the EOF patterns are dimensional.

Second, following [K16](#), a three-level MLR model was empirically estimated separately for each month with 1000-mode truncation, described as

(a) SLP climatology on 1 July

(b) PDF [120–150E, 22.5–47.5N]

FIG. 1. (a) Climatology of SLP on 1 Jul, with a contour interval of 4 hPa. (b) PDF of observed (thick lines) and simulated (thin lines) SLPs by the SLP emulator, in the target area of 120°–150°E, 22.5°–47.5°N.

$$d\mathbf{x} = \mathbf{x}^T \mathbf{A}^{(0)} + \mathbf{e}^{(1)}, \quad (1)$$

$$d\mathbf{e}^{(1)} = [\mathbf{e}^{(1)} \mathbf{x}]^T \mathbf{A}^{(1)} + \mathbf{e}^{(2)}, \quad \text{and} \quad (2)$$

$$d\mathbf{e}^{(2)} = [\mathbf{e}^{(2)} \mathbf{e}^{(1)} \mathbf{x}]^T \mathbf{A}^{(2)} + \mathbf{e}^{(3)}, \quad (3)$$

where the superscript T denotes transpose. Here $\mathbf{x}$ is the 1000-dimensional vector that stores the normalized PCs; $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$, and $\mathbf{e}^{(3)}$ are the residual of the first-, second-, and third-level MLR models, respectively; and the left-hand sides of the equations denote the daily increments

FIG. 2. (a) First, (b) second, and (c) third dimensional EOF patterns for SLP in the Northern Hemisphere. The contour interval is 0.5 hPa, with negative contours dashed and the zero contour omitted. (d) The explained variance for each of the 1000 leading EOFs (%).

of the corresponding variables, for example, $dx = \mathbf{x}(t + 1 \text{ day}) - \mathbf{x}(t)$. This three-level MLR estimates the propagators of $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, and $\mathbf{A}^{(2)}$ from the time series of 1000 modes of SLP anomaly PCs. See [appendix A](#) for the partial least squares (PLS) method of MLR that effectively avoids multicollinearity. After the estimation of the propagator $\mathbf{A}^{(0)}$, which is a 1000×1000 matrix, $\mathbf{e}^{(1)}$, which is the residual of the MLR [Eq. (1)], was obtained from the data. At the second level, the predictor becomes the normalized PCs $\mathbf{x}$ and the first-level residual $\mathbf{e}^{(1)}$, and the size of propagator $\mathbf{A}^{(1)}$ becomes 2000×1000 . Similarly, the MLR [Eq. (2)] allows us to obtain the residual $\mathbf{e}^{(2)}$ using the learning data. At the third level, the predictors are the normalized PCs and the residuals in the first level and the second levels. We estimated the propagator $\mathbf{A}^{(2)}$ with the size 3000×1000 . This completes the construction of the three-level MLR model. The details are described in [appendix B](#).

In the above procedure, we obtained the residual time series of $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$, and $\mathbf{e}^{(3)}$ from January 1960 to

December 2017, from the regression equations for each month that are based on the PC time series $\mathbf{x}$. The SLP emulator [Eqs. (1)–(3)] contains the three-level MLR equation with forcing by $\mathbf{e}^{(3)}$. Following [K16](#), we stored the residual time series $\mathbf{e}^{(3)}$ of the entire learning data period as the noise library. The reforecast experiment was then performed to assess model performance, in which the daily increment of $\mathbf{x}$, $\mathbf{e}^{(1)}$, and $\mathbf{e}^{(2)}$ at a date was computed with the simultaneous forcing of $\mathbf{e}^{(3)}$. The reforecast experiment started from 1 January 1960 successfully traced the PC based on the analysis data ([Fig. 3a](#)). Another experiment started from the same initial date was forced by the residual at the 1-yr leading date ([Fig. 3b](#)). The result shows that the effect of the initial date lingers during the first 10 days but that the forcing is dominant in pulling back the reforecast to the given date of the forcing. As stated in [K16](#), continuous use of the residual data should be avoided in the forecast experiments. The forecast system employs residual forcing for the first 15-day integration

FIG. 3. (a) The first PC of reforecast (dots) and analysis (open circles) in January 1960. (b) The first PC of the SLP emulator experiment (dots) from 1 Jan 1960 driven by the forcing library from January 1961. Also shown in (b) are the first mode of analysis in January 1960 (black open circles) and January 1961 (gray open circles).

using a 15-day-long snippet of residual forcing from the noise library that starts from the date on which the Euclidean distance between reforecast state and the observed state on library dates is smallest in the phase space spanned by the 10 leading modes. The next 15-day integration was also forced by a snippet of the residual, sampled from the noise library, starting from the nearest neighboring date in EOF phase space but avoiding the period from 7 days before to 7 days after the initial date of the snippet of the previous 15-day integration.

Much of the procedure for the SLP emulator construction followed that of K16. However, we have excluded the process of conversion from daily to 6-hourly data for simplification; instead, a linear temporal interpolation was applied to produce the 6-hourly data necessary for PPT. The stability of the SLP emulator, which was not explicitly stated in K16, is briefly checked here; we achieved a 100-yr integration with the SLP emulator without computational blowups (Fig. 4a). The daily SLP data forecast by the SLP emulator became the input to the PPT described in the next section. The centennial trend ranged approximately ± 0.1 for each

FIG. 4. (a) The first PC of a forecast for 100 years. (b) The centennial trend of each PC time series, calculated as the difference between the first 20-yr average and the last one $[(100 \text{ yr})^{-1}]$.

PC (Fig. 4b), which confirmed that this system runs without a climate drift. Hence the SLP emulator is a globally stable dynamic system attracted around the climatological state in the 1000-dimensional phase space, notwithstanding the forcing by the multiplicative noise (Sura et al. 2005; Inatsu et al. 2013).

K16 demonstrated that the results provided by the SLP emulator were almost indistinguishable from the observed features in terms of spatial patterns of bandpass and low-pass-filtered variability and diverse characteristics of midlatitude cyclone tracks. Here, for the specific purpose of this study, we compare the probability density functions (PDFs) of observed and simulated SLPs around the targeted area around Japan (Fig. 1b). The observed PDF based on 6-hourly JRA-55 reanalysis data from January 1960 to December 2017 featured nonnormality with a long tail below 1000 hPa, probably because of frequent passages of developing cyclones through the region throughout the year. The peak was near 1010 hPa, and SLPs mostly ranged between 960 and 1050 hPa. The PDF based on 100-yr daily SLP patterns reconstructed by adding on SLP anomaly simulated by the emulator to climatology showed a distribution that was very similar to

that of the observations. There were slight differences in the PDF, with underestimation of extreme low pressures and overestimation of moderate low pressures.

4. Pressure–precipitation transmitter

First, we have prepared a very long term, daily SLP time series as a form of the EOF mode decomposition with the SLP emulator in section 3. The daily SLP-mode data were spatially reconstructed and clipped to the target domain $\Xi \ni \xi$. The prepared data are a function of ξ and time t , which is written as

$$p_f(\xi, t) = \sum_{m=1}^{1000} p_m(\xi) x_m(t). \quad (4)$$

The daily data were then temporally interpolated to 6-h time intervals. Here, $p_m(\xi)$ is the m th EOF pattern (Figs. 2a–c) and $x_m(t)$ is the m th PC time series again. In preparation for the PPT, SVD analysis with the mode truncation at K was performed separately for each month to identify a spatial pattern with a high correlation between precipitation and three temporally adjacent SLP snapshots. It is noted that the use of three consecutive days follows Blanchet et al. (2018) who found the relationship between large rainfall accumulation and a multiday sequence of atmospheric circulation. A set of three temporally adjacent SLP anomalies p'_a was then decomposed into the left singular vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K\}$ normalized as

$$\mathbf{r}_a(t) = \sum_{k=1}^K a_k(t) \mathbf{u}_k, \quad (5)$$

where

$$\mathbf{r}_a(t) = \begin{bmatrix} p'_{a,1}(t-6\text{ h}) \\ \vdots \\ p'_{a,I}(t-6\text{ h}) \\ p'_{a,1}(t) \\ \vdots \\ p'_{a,I}(t) \\ p'_{a,1}(t+6\text{ h}) \\ \vdots \\ p'_{a,I}(t+6\text{ h}) \end{bmatrix} \in M_{3I,1}, \quad (6)$$

just lumping the three consecutive SLP snapshots together, and I is the total number of grid points in a single SLP snapshot. The precipitation anomaly p'_a was also decomposed into the right singular vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K\}$ normalized as

$$\mathbf{R}_a(t) = \sum_{k=1}^K b_k(t) \mathbf{v}_k^T, \quad (7)$$

where

$$\mathbf{R}_a(t) = \begin{bmatrix} P'_{a,1}(t) \\ \vdots \\ P'_{a,J}(t) \end{bmatrix} \in M_{J,1} \quad (8)$$

and J is the total number of grid points in a single precipitation snapshot.

The PPT contains two parts: the first guess by AnEn, and the correction with the SVD. In the first guess, by searching the analysis SLP data throughout the learning period, we find the N 's nearest neighbors at $t_1, t_2, \dots, t_N$ to the SLP emulator output around a particular time t_0 , measured by the total square difference over the target domain. The first guess of precipitation at t_0 is the geometric mean of the learning precipitation data $P_a(\xi, t)$ over the set of times as

$$\ln \widetilde{P}_f(\xi, t_0) = \frac{1}{N} \sum_{n=1}^N \ln P_a(\xi, t_n), \quad (9)$$

whereas the corresponding SLP is the arithmetic mean of the three adjacent SLP snapshots as

$$\begin{bmatrix} \widetilde{p}_f(\xi, t_0 - 6\text{ h}) \\ \widetilde{p}_f(\xi, t_0) \\ \widetilde{p}_f(\xi, t_0 + 6\text{ h}) \end{bmatrix} = \frac{1}{N} \sum_{n=1}^N \begin{bmatrix} p_a(\xi, t_n - 6\text{ h}) \\ p_a(\xi, t_n) \\ p_a(\xi, t_n + 6\text{ h}) \end{bmatrix}. \quad (10)$$

The difference between original SLP emulator output and the first guess just above is discretized and forms the $3I$ -dimensional vector of

$$\mathbf{r}'_f(t_0) = \begin{bmatrix} p_{f,1}(t_0 - 6\text{ h}) - \widetilde{p}_{f,1}(t_0 - 6\text{ h}) \\ \vdots \\ p_{f,I}(t_0 - 6\text{ h}) - \widetilde{p}_{f,I}(t_0 - 6\text{ h}) \\ p_{f,1}(t_0) - \widetilde{p}_{f,1}(t_0) \\ \vdots \\ p_{f,I}(t_0) - \widetilde{p}_{f,I}(t_0) \\ p_{f,1}(t_0 + 6\text{ h}) - \widetilde{p}_{f,1}(t_0 + 6\text{ h}) \\ \vdots \\ p_{f,I}(t_0 + 6\text{ h}) - \widetilde{p}_{f,I}(t_0 + 6\text{ h}) \end{bmatrix} \in M_{3I,1}. \quad (11)$$

The orthonormality of the left SVD vectors allows us to find the coefficients corresponding to $\mathbf{r}'_f(t)$ as

FIG. 5. Illustration of the usage of the SLP emulator and PPT.

$$a'_k = \mathbf{r}'_f(t_0) \mathbf{u}_k^T. \quad (12)$$

The precipitation anomaly correlated with the difference between original SLP emulator output and the first guess is hence given as

$$\mathbf{R}'_f(t_0) = \sum_{k=1}^K a'_k \mathbf{v}_k^T \left[\int b_k^2(\tau) d\tau / \int a_k^2(\tau) d\tau \right]^{1/2}, \quad (13)$$

which can be reconstructed as $P'_f(\xi, t_0)$. The PPT finally provides the precipitation estimate corresponding to the SLP emulator output $\tilde{p}_f(\xi, t_0)$ as the sum of the AnEn's first guess and the SVD correction: The final product is then written as

$$\widehat{P}_f(\xi, t_0) = \widetilde{P}_f(\xi, t_0) + P'_f(\xi, t_0). \quad (14)$$

The final goal of this paper is to create the statistics of annual-maximum daily precipitation from long-term SLP data. It is more likely that AnEn averages out extreme data. Therefore, even if the SVD correction effectively worked to compensate for the averaging effect of AnEn, we cannot expect to extend the statistics for extreme precipitation from the learning period based on long-term daily SLP data. To get around this problem, we interpolated the learning SLP data to 6-hourly data. Since a PPT procedure picks up an analog date from the learning period every 6 h, we repeated this procedure 4 times daily and took their sum to estimate daily precipitation. The observed annual-maximum value of 6-h precipitation in the learning period was often chosen repeatedly within a day in PPT because of the similarity between the SLP pattern of the learning data and the emulator data. Therefore, the estimated

daily precipitation may exceed the maximum of the 6-hourly observed precipitation. We will provide an example to discuss the annual-maximum daily precipitation at sites from subtropical to subarctic climates in [section 6b](#).

The SLP emulator creates a long-term SLP dataset as long as one desires and the PPT, which is a function of three consecutive SLP data, produces a single precipitation pattern. Summarizing the procedure of producing precipitation pattern with PPT from simulated SLP ([Fig. 5](#)), the long-term SLP data created by the SLP emulator becomes the input and a sequential use of the PPT produces long-term precipitation data of the same length as the output. Although JRA-55 reanalysis data in the full period from 1960 to 2017 were utilized in the MLR estimate, data only from 2010 to 2017 were used to compute the SVD pattern correlated with MSM's precipitation data in the target region. We also used the data from 2010 to 2017 in the AnEn process, in which the search for the closest-neighbor state to the three consecutive simulated SLP patterns was conducted.

5. Setup and performance

In the first procedure of the PPT, we computed the SVD between precipitation and three temporally adjacent SLPs, in the target area of 120°–150°E, 22.5°–47.5°N, separately for each month. The heterogeneous regression map of an SVD mode was then composed from a panel for SLP and precipitation at a reference time, and two panels for SLP 6 h before and after the reference time ([Fig. 6](#)). The first SVD mode, with the opposite sign of [Figs. 6a–c](#), for January indicated that

FIG. 6. Heterogeneous regression map for the first SVD mode between precipitation and three adjacent SLPs in (top) January and (bottom) July: (a),(d) SLP with 6-h lag, (b),(e) SLP and precipitation without lead/lag, and (c),(f) SLP with 6-h lead. The contour interval is 0.5 hPa, and the shading is as per the reference at the bottom (mm day^{-1}).

the precipitation on the side of the Sea of Japan corresponds to eastward-moving extratropical cyclones located to the east of Japanese islands. The second SVD mode for January signified that the precipitation along the Pacific Ocean side was correlated with extratropical cyclones passing along the southern coast of Japan (not shown). The third mode was related to the precipitation induced by extratropical cyclones in northern Japan (not shown). These gravest modes in winter-related regional heavy precipitations in Japan to passages of extratropical cyclones. By contrast, the first SVD mode for July suggested that the persistent low pressure anomaly was correlated with the precipitation over Japan, which was mainly related to the baiu front including effects from tropical cyclones (Figs. 6d–f). The second mode was a pattern that showed a high correlation between precipitation and tropical cyclones in the Ryukyu Islands (not shown).

The third mode indicated a northward shift of the baiu front and an extension of the Bonin high toward the west (not shown). Thus, these gravest modes in summer characterized the precipitation in Japan with influences from Asian monsoon and tropical cyclones.

Before the main result, we will illustrate the performance of the PPT from case studies. A wintertime example on 22 January 2018, outside the learning period from 2010 to 2017, is a typical case with a passing extratropical cyclone and precipitation along the southern coast of Japan (Fig. 7a), which brought traffic accidents on a metropolitan expressway as a result of unusual snowfall. Analog dates are next sought from the learning SLP data in January from 2010 to 2017. The precipitation estimated as the geometric mean over the analog dates (Fig. 7b) almost reproduced the precipitation at the example date (Fig. 7a). The SVD

FIG. 7. (a),(d) Daily mean SLP (contours) and daily precipitation (shading); (b),(e) the composite of the analog dates from 2010 to 2017; and (c),(f) the estimate with the PPT for (top) 22 Jan and (bottom) 5 Jul 2018. The contour interval is 4 hPa and the shading is as per the reference at the bottom (mm day^{-1}).

correction slightly modified the pattern (Fig. 7c), by adding slightly more rainfall in the east of Japan. The estimated precipitation reproduced well the observed precipitation, including mesoscale features, probably because a similar weather pattern existed during the learning period. Another example on 5 July 2018 is also a precipitation event in summer with a massive amount of rainfall in western Japan, which brought disastrous flash floods and landslides with 224 casualties; warm moist air intruded along the western fringe of the Bonin high to the south of Japan, and a subtropical rain belt formed in the convergence zone (Fig. 7d). Analog dates were also sought for this event from the learning SLP data, as with the winter example. The AnEn composite shows a similar spatial pattern in SLPs and precipitation (Fig. 7e). The SVD correction slightly compensated for the difference between the analysis and analog composite from the Korea area to Japan, although this correction did not seem to be effective (Fig. 7f). The estimated precipitation was much less than the observed

precipitation where it peaked in western Japan. This is probably because the case on 5 July 2018 involved record-breaking precipitation at many observation sites in western Japan.

6. Results

a. Climatology, PDF, and autocorrelation function

The PPT converts a 100-yr daily SLP dataset from the SLP emulator into a 100-yr daily precipitation dataset. As in section 5, we have prepared 6-hourly SLP and precipitation from 2010 to 2017 as an 8-yr learning period, and we have temporally interpolated the 100-yr daily SLP dataset from the emulator to a 6-h interval. The PPT matches the emulated SLP with the learning SLP to find two analog dates and takes their average for SLP and precipitation. The SVD correction was made to compensate for the difference between emulator and AnEn estimate SLP. As the result of these PPT procedures, a 6-hourly precipitation dataset

corresponding to the 6-hourly emulated SLP was obtained. A daily precipitation was then estimated from the daily sum of 6-hourly precipitation.

Comparing the precipitation from the PPT scheme with the observation (Fig. 8), the annual-mean precipitation had a small relative error over Japanese territory. The PPT scheme in this study tended to overestimate precipitation, particularly over the continent and over the Pacific (Fig. 8a). This overestimation tendency was also found over Japanese land grid points in all intensity ranges of the simulated precipitation probability, slightly exceeding the observation by 1 mm day^{-1} (Fig. 8b). Although we did not apply any bias correction here, this bias may be readily corrected with standard bias-correction techniques such as shifting and scaling (Leander and Buishand 2007) and quantile mapping (Piani et al. 2010). Despite the bias in it, our scheme, categorized as a statistical downscaling technique, has a strength to capture extreme daily precipitation with a comparable probability in the long-term simulation. However, one weakness is the redness of precipitation time series. The observed precipitation had a short memory of typically less than 1 day (Fig. 8c), whereas the simulated one had a memory of approximately 2 days. This is perhaps because of the short period of the reference data to seek an analog pattern by the metric of similarity of SLP. This problem could be relieved if one used a longer reference dataset or added artificial white noise in the postprocessing.

We checked the spatiotemporal continuity of the estimated precipitation. The estimated precipitation attains the maximum value at Tokyo on 27 October of the 83rd model year (Fig. 9). The SLP emulator calculates a cyclone moving northeastward to the south of Japan on 26 October. On 27 October, the central pressure deepened to lower than 992 hPa, and the daily precipitation exceeded 30 mm along the Pacific side of Japan. This is consistent with the observation, related to abundant moisture brought by southerlies. The cyclone moved to the far east of Japan and the precipitation became moderate in eastern Japan, perhaps corresponding to the easterlies blowing back from the cyclone. The transitions during this three-day period in SLP and precipitation around the precipitation extreme was therefore fairly realistic, with continuity of SLP and precipitation in time and space over the target area.

b. Annual-maximum precipitation estimate

The final product of the annual-maximum daily precipitation is displayed in normal quantile–quantile (Q–Q) map (Fig. 10). In the 8-yr learning data, the annual-maximum daily precipitation ranged from 100 to 300 mm at Naha in the subtropical climate zone, and it

FIG. 8. (a) Relative error of 100-yr mean annual precipitation in the PPT scheme to 8-yr mean annual precipitation in the MSM analysis (shading, with its color reference on the right). The black and blue contours respectively denote 1000 and 2000 mm of observed precipitation. (b) Probability of daily precipitation over Japanese land grid points of the MSM analysis (black bars) and that obtained from the PPT scheme (red bars). The leftmost bin includes nonprecipitation days. (c) Lag autocorrelation of daily precipitation of the MSM analysis (black) and that obtained from the PPT scheme (red).

ranged from 50 to 100 mm at Sapporo in the subarctic climate zone. The annual-maximum daily precipitation estimated from the PPT, plotted on the Q–Q map, ranges from 50 to 500 mm at Naha (Fig. 10a), 40 to 200 mm at Fukuoka (Fig. 10b) and Tokyo (Fig. 10c), and 30 to 200 mm at Sapporo (Fig. 10d). Hence PPT

FIG. 9. Sea level pressure (contours) and daily precipitation estimated by the PPT (shading) on (a) 26, (b) 27, and (c) 28 Oct of the 83rd model year. The contour interval is 4 hPa, and the shading is as per the reference at the bottom (mm).

successfully extended the probability distribution tail, doubling its maximum value at Naha, in Fukuoka, and in Sapporo. For example, at Naha, the 10-yr return-period (RP10) precipitation derived from the 8-yr learning data was about 250 mm day^{-1} , whereas the RP100 precipitation derived from the precipitation estimated from the PPT was about 500 mm day^{-1} (Fig. 10a). By contrast, the estimated RP100 precipitation did not exceed the maximum precipitation in the learning data in Tokyo (Fig. 10c), which is obviously attributed to the smoothing effect of AnEn in PPT to estimate the first guess. The SVD correction did not affect much on the overall feature of the Q-Q plot (Fig. 10). It only smoothed out bumps in the Q-Q plot those were probably produced by sets of analog dates that took similar values.

The extent of the distribution of the PPT procedure depends on the extreme statistics of the learning data. The standard deviation of the logarithm of annual-maximum daily precipitation is 0.16 at Naha, 0.07 at Fukuoka, 0.21 at Tokyo, and 0.13 at Sapporo. Since the possible maximum of estimated annual-maximum daily precipitation by applying the AnEn procedure four times to the 6-hourly dataset is fourfold of the maximum daily precipitation in the learning period, the logarithm of annual-maximum daily precipitation can extend to $\log 4 \cong 0.6$. Here we conduct a thought experiment for a case in which the PPT doubles the estimated maximum precipitation from that of the learning data at a site where the standard deviation of the logarithm of annual-maximum daily precipitation is 0.3. As the doubling is

equal to the addition of 0.301 in the logarithm, the estimation extends the tail of the Gaussian distribution to 2 times that of the learning data. If the data follow the normalized Gaussian distribution, then RP8, approximately the top 8.3% in the median method, corresponds to a Z value of 1.382. Doubling the maximum precipitation by addition of 0.301 to the Z value results in an estimated Z value of 2.386. This corresponds to RP of 81 yr because only 0.85% of the data will exceed this value (Fig. 11). This simple thought experiment suggests that, although very sensitive to the standard deviation of the learning data, the PPT method can estimate RP80 based on only 8-yr learning data. It can be emphasized, with the same thought experiment, that RP300 can be estimated from 20-yr data and RP1000 can be estimated from 50-yr data, assuming that the PPT doubles the maximum of the learning data (not shown).

7. Summary and discussion

We have proposed a new method to create a very long term daily precipitation data for the extreme statistics, which is often required in various sectors such as hydrological management, erosion control, traffic engineering, and disaster insurance. As has been described in the Introduction, we propose a method combining a statistical MLR model called the SLP emulator from K16 and SDS methods such as AnEn and SVD to produce very long term precipitation data to meet the purpose of this paper. In this paper, after a brief review

FIG. 10. Normal quantile–quantile map of annual-maximum daily precipitation (mm day^{-1}) at (a) Naha, (b) Fukuoka, (c) Tokyo, and (d) Sapporo. Black dots and red symbols denote the annual-maximum daily precipitation based on the PPT estimation with and without SVD correction, respectively; blue open circles denote that of the observation from 2010 to 2017.

of the SLP emulator with a self-contained description including [appendixes A and B](#), we constructed the MLR model for each month based on the time series of 1000 PCs from 1960 to 2017. We obtained 100-yr, daily SLP data, by simply integrating this statistical forecast model, and temporally interpolated the data to a 6-h interval. In the next procedure, we developed the method to convert 6-hourly SLP to precipitation, the PPT. The daily precipitation is estimated mainly from a composite of time frames with similar SLP transition patterns in the learning period and is then corrected with an SVD relation between SLP and precipitation. Using the analysis data from 2010 to 2017 as the learning data, we estimated daily precipitation based on the 100-yr SLP emulator output. We checked that the estimated daily precipitation was fairly realistic, maintaining temporal and spatial continuity. The annual-maximum precipitation in the estimated 100-yr data extended the tail of probability distribution of the 8-yr learning data.

This study presents a sequential combination of the SLP emulator to generate a surrogate time series on the

synoptic field and the PPT, a kind of SDS methods, to convert synoptic-scale pressure field to finer-scale precipitation field. In the former, there is a choice in orthogonal functions for expansion. Here we followed [K16](#) and chose the SLP's EOF-based model. [Crommelin and Majda \(2004\)](#) demonstrated that EOF-based empirical reduced model were unable to reproduce behaviors involving quick transitions of the original model, even when the EOFs represented up to 99.8% of the variance. They recommended another set of orthogonal functions called principal interaction patterns ([Kwasniok 1996](#)), despite the upper limit of time integration being a sensitive parameter. Though their discussion was confined to an idealized dynamic system, their suggestion seems consistent with the underestimation of the extreme of low pressures over the target domain ([Fig. 1b](#)). This implies that frequency of rapidly intensifying extratropical cyclones frequently found in winter ([Sanders and Gyakum 1980](#); [Yoshida and Asuma 2004](#)) and tropical cyclones typically passing through Japan from summer to autumn ([Yumoto and Matsuura 2001](#)) might

FIG. 11. Return-period (RP) distribution in two-parameter space, with the contours at RP20, RP50, RP100, RP200, RP500, ..., RP500 000, and RP1 000 000. The horizontal axis denotes the standard deviation of the logarithm of annual-maximum precipitation of the 10-yr learning data. The vertical axis denotes the estimate-to-learning ratio of the maximum value of annual-maximum precipitation.

be underestimated. However, this drawback was partly compensated by the analog estimate process in the PPT, because three consecutive SLP patterns must be chosen from a learning observed dataset.

Acknowledgments. We thank three anonymous reviewers who gave us insightful comments to improve the original paper. The meteorological observation data were provided by the JMA. Author MI is supported by JSPS KAKENHI Grants 18K03734, 18H03819, and 19H00963, the Social Implementation Program for Climate Change Adaptation Technology funded by Ministry of Education, Culture, Sports, Science, and Technology, and is partly supported by the Environment Research and Technology Development Fund 2-1905 of the Environmental Restoration and Conservation Agency of Japan, Research Field of Hokkaido Weather Forecast and Technology Development (Endowed by Hokkaido Weather Technology Center Co., Ltd.), and a collaborative study with the Tokio Marine and Nichido Risk Consulting Co., Ltd. Figures were drawn using the Grid Analysis and Display System.

APPENDIX A

PLS Method

The PLS method is utilized as the MLR algorithm (Wold et al. 2001). We now estimate the MLR relation of

$$\mathbf{y} = \mathbf{X}\hat{\mathbf{b}} \quad (\text{A1})$$

for multiple predictors

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \in M_{nm} \quad (\text{A2})$$

and a single predictand

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in M_{n1}, \quad (\text{A3})$$

where n is the data length in time and m is the number of predictors. Note that M_{nm} denotes the elements of an $n \times m$ matrix. The first left SVD $\mathbf{w}_1$ and the first latent vector $\mathbf{t}_1$ are defined respectively as

$$\mathbf{w}_1 = \mathbf{X}^T \mathbf{y} \quad \text{and} \quad \mathbf{t}_1 = \mathbf{X} \mathbf{w}_1. \quad (\text{A4})$$

Predictors and predictand are projected onto the first latent vector as

$$\mathbf{p}_1 = \frac{\mathbf{X}^T \mathbf{t}_1}{\|\mathbf{t}_1\|^2} \quad \text{and} \quad q_1 = \frac{\mathbf{y}^T \mathbf{t}_1}{\|\mathbf{t}_1\|^2}. \quad (\text{A5})$$

This leads to the first approximation of predictors and predictand as $\mathbf{t}_1 \mathbf{p}_1^T$ and $\mathbf{t}_1 q_1$, respectively. The residuals of the PLS first approximation are redefined as $\mathbf{X}$ and $\mathbf{y}$, say,

$$\mathbf{X} := \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T \quad \text{and} \quad \mathbf{y} := \mathbf{y} - \mathbf{t}_1 q_1, \quad (\text{A6})$$

and the calculation goes back to Eq. (A4) to find the second left SVD $\mathbf{w}_2$ and the second latent vector $\mathbf{t}_2$. This procedure is iterated l ($=40$) times to obtain higher SVD modes $\mathbf{w}_k$ and the projection of predictors and predictand onto the higher latent vector $\mathbf{p}_k$ and q_k ($1 \leq k \leq l$). After the iteration to find the matrices $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_l) \in M_{ml}$ and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_l) \in M_{ml}$ and the vector $\mathbf{q} = (q_1, \dots, q_l)^T \in M_{l1}$, the MLR coefficient is estimated by the PLS method as

$$\hat{\mathbf{b}} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q}. \quad (\text{A7})$$

APPENDIX B

Details of MLR estimation in K16

In the first level, the propagator $\mathbf{A}^{(0)}$ has the size of 1000×1000 . However, the size is reduced so that the

FIG. B1. The (left) MLR estimation for January in 1960 and 1961 [predictor (thin solid line), regression estimate (thick solid line), and residual (dotted line)] and (right) lag-1 autocorrelation (%) residual of each mode of the MLR in January from 1960 to 2017 for the (a),(b) first, (c),(d) second, and (e),(f) third levels.

increment of each mode is related to the nearest 101 modes. Therefore, the first-level MLR for the k th mode is virtually performed for

$$\begin{aligned} dx_k &= \sum_{m=1}^{101} x_m a_{m,k}^{(0)} + e_k^{(1)} & (1 \leq k \leq 50) \\ dx_k &= \sum_{m=-50}^{50} x_{k+m} a_{k+m,k}^{(0)} + e_k^{(1)} & (51 \leq k \leq 950) \\ dx_k &= \sum_{m=900}^{1000} x_m a_{m,k}^{(0)} + e_k^{(1)} & (951 \leq k \leq 1000). \end{aligned} \quad (B1)$$

Based on the daily SLP anomaly time series, the residual time series $\mathbf{e}^{(1)}$ and propagator $\mathbf{A}^{(0)}$ are derived at the first level of the MLR. For example, from January 1960 to December 1961, the daily increment dx_1 can nearly reproduce the linear regression $\sum_{m=1}^{101} x_m a_{m,1}^{(0)}$. However, the residual $e_1^{(1)}$ is not negligible (Fig. B1a). At this level, because the lag-1 autocorrelation of the residual of the 10 leading modes is greater than 0.4 (Fig. B1b), the residuals seldom follow the Gaussian distribution.

The second-level MLR for the k th mode is virtually performed for

$$\begin{aligned} de_k^{(1)} &= \sum_{m=1}^{101} e_m^{(1)} b_{m,k}^{(1)} + x_m a_{m,k}^{(1)} + e_k^{(2)} & (1 \leq k \leq 50) \\ de_k^{(1)} &= \sum_{m=-50}^{50} e_{k+m}^{(1)} b_{k+m,k}^{(1)} + x_{k+m} a_{k+m,k}^{(1)} + e_k^{(2)} & (51 \leq k \leq 950) \\ de_k^{(1)} &= \sum_{m=900}^{1000} e_m^{(1)} b_{m,k}^{(1)} + x_m a_{m,k}^{(1)} + e_k^{(2)} & (951 \leq k \leq 1000), \end{aligned} \quad (B2)$$

with the unknown parameters reduced to 101×2 for a single mode estimation. For the same example, from January 1960 to December 1961, the daily increment $de_1^{(1)}$ can approximately reproduce the linear regression $\sum_{m=1}^{101} x_m a_{m,1}^{(1)} + e_m^{(1)} b_{m,1}^{(1)}$, and the residual $e_1^{(2)}$ is smaller

(Fig. B1c). However, the lag-1 autocorrelation of the residual for the 10 leading modes is close to 0.2 (Fig. B1d) and the residuals are still colored.

The third-level MLR for the k th mode is virtually performed for

$$\begin{aligned} de_k^{(2)} &= \sum_{m=1}^{101} e_m^{(2)} c_{m,k}^{(2)} + e_m^{(1)} b_{m,k}^{(2)} + x_m a_{m,k}^{(2)} + e_k^{(3)} & (1 \leq k \leq 50) \\ de_k^{(2)} &= \sum_{m=-50}^{50} e_{k+m}^{(2)} c_{k+m,k}^{(2)} + e_{k+m}^{(1)} b_{k+m,k}^{(2)} + x_{k+m} a_{k+m,k}^{(2)} + e_k^{(3)} & (51 \leq k \leq 950) \\ de_k^{(2)} &= \sum_{m=900}^{1000} e_m^{(2)} c_{m,k}^{(2)} + e_m^{(1)} b_{m,k}^{(2)} + x_m a_{m,k}^{(2)} + e_k^{(3)} & (951 \leq k \leq 1000). \end{aligned} \quad (B3)$$

For the same example, the daily increment $de_1^{(2)}$ can now well reproduce the linear regression $\sum_{m=1}^{101} x_m a_{m,1}^{(2)} + e_m^{(1)} b_{m,1}^{(2)} + e_m^{(2)} c_{m,1}^{(2)}$, and the residual $e_1^{(3)}$ is negligible (Fig. B1e) and can be regarded as white noise (Fig. B1f).

REFERENCES

- Ailliot, P., D. Allard, V. Monbet, and P. Naveau, 2015: Stochastic weather generators: An overview of weather type models. *J. Soc. Fr. Statist.*, **156**, 101–113.
- Baigorria, G. A., and J. W. Jones, 2010: GiST: A stochastic model for generating spatially and temporally correlated daily rainfall data. *J. Climate*, **23**, 5990–6008, <https://doi.org/10.1175/2010JCLI3537.1>.
- Barnett, T., and R. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci.*, **35**, 1771–1787, [https://doi.org/10.1175/1520-0469\(1978\)035<1771:MAPOST>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<1771:MAPOST>2.0.CO;2).
- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).
- Blanchet, J., S. Stalla, and J.-D. Creutin, 2018: Analogy of multiday sequences of atmospheric circulation favoring large rainfall accumulation over the French Alps. *Atmos. Sci. Lett.*, **19**, e809, <https://doi.org/10.1002/asl.809>.
- Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827–1846, <https://doi.org/10.1175/JCLI-D-11-00227.1>.
- Busuioc, A., D. Chen, and C. Hellström, 2001: Performance of statistical downscaling models in GCM validation and regional climate change estimates: Application for Swedish precipitation. *Int. J. Climatol.*, **21**, 557–578, <https://doi.org/10.1002/joc.624>.
- Crommelin, D. T., and A. J. Majda, 2004: Strategies for model reduction: Comparing different optimal bases. *J. Atmos. Sci.*,

- 61, 2206–2217, [https://doi.org/10.1175/1520-0469\(2004\)061<2206:SFMRCDD>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<2206:SFMRCDD>2.0.CO;2).
- Delle Monache, L., F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, <https://doi.org/10.1175/MWR-D-12-00281.1>.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Hellström, C., D. Chen, C. Achberger, and J. Räisänen, 2001: Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Res.*, **19**, 45–55, <https://doi.org/10.3354/cr019045>.
- Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, Geophys. Monogr., Vol. 134, Amer. Geophys. Union, 1–36.
- Ichikawa, Y., and M. Inatsu, 2017: An alternative estimate of potential predictability on the Madden–Julian oscillation phase space using S2S models. *Atmosphere*, **8**, 150, <https://doi.org/10.3390/atmos8080150>.
- Inatsu, M., N. Nakano, and H. Mukougawa, 2013: Dynamics and practical predictability of extratropical wintertime low-frequency variability in a low-dimensional system. *J. Atmos. Sci.*, **70**, 939–952, <https://doi.org/10.1175/JAS-D-12-048.1>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Kondrashov, D., S. Kravtsov, and M. Ghil, 2006: Empirical mode reduction in a model of extratropical low-frequency variability. *J. Atmos. Sci.*, **63**, 1859–1877, <https://doi.org/10.1175/JAS3719.1>.
- Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multilevel regression modelling of nonlinear processes: Derivation and applications to climatic variability. *J. Climate*, **18**, 4404–4424, <https://doi.org/10.1175/JCLI3544.1>.
- , N. Tilinina, Y. Zyulyaeva, and S. K. Gulev, 2016: Empirical modelling and stochastic simulation of sea level pressure variability. *J. Appl. Meteor. Climatol.*, **55**, 1197–1219, <https://doi.org/10.1175/JAMC-D-15-0186.1>.
- Kuno, R., and M. Inatsu, 2014: Development of sampling downscaling: A case for wintertime precipitation in Hokkaido. *Climate Dyn.*, **43**, 375–387, <https://doi.org/10.1007/s00382-014-2060-5>.
- Kwasniok, F., 1996: The reduction of complex dynamical systems using principal interaction patterns. *Physica D*, **92**, 28–60, [https://doi.org/10.1016/0167-2789\(95\)00280-4](https://doi.org/10.1016/0167-2789(95)00280-4).
- Leander, R., and T. A. Buishand, 2007: Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrol.*, **332**, 487–496, <https://doi.org/10.1016/j.jhydrol.2006.08.006>.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogs. *J. Atmos. Sci.*, **26**, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Mizuta, R., and Coauthors, 2017: Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bull. Amer. Meteor. Soc.*, **98**, 1383–1398, <https://doi.org/10.1175/BAMS-D-16-0099.1>.
- Piani, C., J. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Richardson, C., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.*, **17**, 182–190, <https://doi.org/10.1029/WR017i001p00182>.
- Saito, K., and Coauthors, 2006: The operational JMA non-hydrostatic mesoscale model. *Mon. Wea. Rev.*, **134**, 1266–1298, <https://doi.org/10.1175/MWR3120.1>.
- Sanders, F., and J. R. Gyakum, 1980: Synoptic–dynamic climatology of the “bomb.” *Mon. Wea. Rev.*, **108**, 1589–1606, [https://doi.org/10.1175/1520-0493\(1980\)108<1589:SDCOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1589:SDCOT>2.0.CO;2).
- Sharma, A., and U. Lall, 1999: A nonparametric approach for daily rainfall simulation. *Math. Comput. Simul.*, **48**, 361–371, [https://doi.org/10.1016/S0378-4754\(99\)00016-6](https://doi.org/10.1016/S0378-4754(99)00016-6).
- Smith, K., C. Strong, and F. Rassoul-Agha, 2018: Multisite generalization of the SHARp weather generator. *J. Appl. Meteor. Climatol.*, **57**, 2113–2127, <https://doi.org/10.1175/JAMC-D-17-0236.1>.
- Sura, P., M. Newman, C. Penland, and P. Sardeshmukh, 2005: Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes? *J. Atmos. Sci.*, **62**, 1391–1409, <https://doi.org/10.1175/JAS3408.1>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wilby, R. L., O. J. Tomlinson, and C. W. Dawson, 2003: Multi-site simulation of precipitation by conditional resampling. *Climate Res.*, **23**, 183–194, <https://doi.org/10.3354/cr023183>.
- Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, [https://doi.org/10.1016/S0022-1694\(98\)00186-3](https://doi.org/10.1016/S0022-1694(98)00186-3).
- , 1999: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agric. For. Meteorol.*, **96**, 85–101, [https://doi.org/10.1016/S0168-1923\(99\)00037-4](https://doi.org/10.1016/S0168-1923(99)00037-4).
- Wold, S., M. Sjöström, and L. Eriksson, 2001: PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, **58**, 109–130, [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- Yamada, T. J., and Coauthors, 2018: The influence of climate change on flood risk in Hokkaido (in Japanese with English abstract). *J. Japan Soc. Civ. Eng.*, **24B**, 391–396.
- Yoshida, A., and Y. Asuma, 2004: Structure and environment of explosively developing extratropical cyclones in the north-western Pacific region. *Mon. Wea. Rev.*, **132**, 1121–1142, [https://doi.org/10.1175/1520-0493\(2004\)132<1121:SAEOED>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1121:SAEOED>2.0.CO;2).
- Yumoto, M., and T. Matsuura, 2001: Interdecadal variability of tropical cyclone activity in the western North Pacific. *J. Meteor. Soc. Japan*, **79**, 23–35, <https://doi.org/10.2151/JMSJ.79.23>.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).