New Approach to Multisite Downscaling of Precipitation by Identifying Different Set of Atmospheric Predictor Variables

Basu, Bidroha; Nogal, Maria; O'Connor, Alan

DOI
10.1061/(ASCE)HE.1943-5584.0001900

Publication date
2020

Document Version
Final published version

Published in
Journal of Hydrologic Engineering

Citation (APA)
Basu, B., Nogal, M., & O'Connor, A. (2020). New Approach to Multisite Downscaling of Precipitation by Identifying Different Set of Atmospheric Predictor Variables. *Journal of Hydrologic Engineering*, *25*(5), Article 04020013. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001900

ASCE

New Approach to Multisite Downscaling of Precipitation by Identifying Different Set of Atmospheric Predictor Variables

Bidroha Basu[1]; Maria Nogal[2]; and Alan O'Connor[3]

Abstract: Estimating reliable projections of precipitation considering climate change scenarios is important for hydrological studies. General circulation models provide future climate simulations at large scale in terms of large-scale atmospheric variables (LSAVs). Those LSAVs can be downscaled to finer special resolution using several downscaling approaches. This paper presents a support vector regression (SVR)-based downscaling approach to downscale rainfall at several locations in a study area. Because the rainfall generation mechanisms cannot be the same for all the sites in a study area, conventional multisite downscaling approaches that assume the same rainfall generation mechanism should not be used. Therefore, a new downscaling approach is proposed that (1) divides the study area in several climatological regions, and (2) develops different downscaling models for each of the climatological regions to obtain future projections of rainfall. The new approach was implemented on rainfall data obtained for Republic of Ireland to demonstrate the effectiveness of the approach compared with existing approaches. Future projections of rainfall were obtained for the period 2012–2050 corresponding to four Representative Concentration Pathway climate change scenarios. The performance of the SVR approach was compared with that of relevance vector machine– and deep learning–based downscaling approaches. **DOI: 10.1061/(ASCE)HE.1943-5584.0001900.**

Author keywords: Statistical downscaling; Rainfall; Support vector regression; Global *K*-means; Future rainfall projections.

Introduction

Precipitation is one of the most important hydrometeorological variables for hydrological modeling. Prediction of rainfall on a catchment scale due to the effect of climate change is a challenging problem. Predictors available at the large scale from general circulation models (GCMs) fail to capture the local- or catchment-scale variability present in the rainfall. To overcome this limitation, large-scale GCM outputs often are downscaled to finer spatial resolution. The downscaling approaches can be classified broadly into two categories, dynamic downscaling and statistical downscaling.

Dynamic downscaling considers the initial conditions, time-dependent lateral meteorological conditions, and surface boundary conditions along with the GCM data and transfers the large-scale GCM data to a higher-resolution regional scale by using regional climate models (RCMs). The advantage of those models is that they can account for the physical processes governing the rainfall and can simulate finer-scale atmospheric processes (e.g., orographic precipitation) better than the host GCMs (Fowler and Wilby 2010). The disadvantages of RCMs are that the models are complex and require significant time to produce outputs (Eden and Widmann 2014). In addition, RCMs provide future climatic conditions at specific grids, and hence it is not possible to directly obtain future rainfall simulations in a region located slightly outside the RCM grids (D'Onofrio et al. 2014). In statistical downscaling, the regional climate at the local scale is considered to be a function of the large-scale climate denoted by the large-scale atmospheric variables. To perform the downscaling, a statistical relationship is developed between those large-scale atmospheric variables (e.g., temperature at high elevations, atmospheric pressure, wind speed, humidity, and solar radiation) and the predictand variable (e.g., rainfall or temperature at the watershed scale) at the local scale. Statistical downscaling has gained wide popularity due to its low computational cost and simplicity compared with dynamic downscaling (Okkan and Fistikoglu 2014; Rashid et al. 2016; Sachindra et al. 2016). Wood et al. (2004) considered linear interpolation, spatial disaggregation, and bias correction and spatial disaggregation (BCSD)-based statistical downscaling methods to simulate meteorological variables from 1975 to 1995 and estimated land surface energy and water fluxes using a variable infiltration capacity model. They noted that the performance of BCSD was superior to that of the other two statistical downscaling methods, and the dynamic downscaling did not provide any improvement compared with statistical downscaling. Ahmed et al. (2013) also used a BCSD-based statistical downscaling approach to downscale precipitation and maximum and minimum temperature at the daily scale using six GCMs and compared it with four RCMs, and found similar results for both types of downscaling approaches. Application of multiple regression–based statistical downscaling and RCM dynamic downscaling of monthly precipitation at 42 stations in Sweden by Hellström et al. (2001) yielded similar performance in reproducing the seasonal precipitation cycle. Landman et al. (2009) noted that a model output

[1]Research Fellow, School of Architecture, Planning and Environmental Policy, Univ. College Dublin, Dublin D14 E099, Ireland; Adjunct Assistant Professor, Dept. of Civil, Structural and Environmental Engineering, Trinity College Dublin, Dublin D02, Ireland (corresponding author). ORCID: https://orcid.org/0000-0002-8822-7167. Email: bidroha.basu@ucd.ie; bbasu@tcd.ie

[2]Assistant Professor, Materials, Mechanics, Management, and Design, Delft Univ. of Technology, Delft 2628 CN, Netherlands. Email: m.nogal@tudelft.nl

[3]Professor, Dept. of Civil, Structural, and Environmental Engineering, Trinity College Dublin, Dublin D02, Ireland. Email: oconnoaj@tcd.ie

Note. This manuscript was submitted on December 8, 2018; approved on October 18, 2019; published online on February 27, 2020. Discussion period open until July 27, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

statistics-based statistical downscaling method was slightly better at simulating the rainfall at 963 stations in South Africa compared with the RCM dynamic downscaling method. Various other studies (Murphy 1999; Wilby et al. 2000; Diez et al. 2005) found that statistical and dynamic downscaling methods produce similar outcomes in simulating the present and future climate.

The statistical downscaling can be classified in three categories, weather generator, weather typing, and transfer function–based methods (Wilby et al. 1998; Nguyen et al. 2006; Fowler et al. 2007; Vrac and Naveau 2007). A weather generator (Wilks and Wilby 1999; Olsson et al. 2009) stochastically generates the future weather scenarios in terms of local-scale time series by using a statistical model (probability density functions) whose parameters are related to the large-scale data (e.g., Vrac and Naveau 2007), whereas weather typing generates synthetic sequences of weather patterns from past observations by conditioning the simulation of small-scale data on weather types over the region of interest (e.g., Vrac et al. 2007b). Neither of those two statistical downscaling methods is suitable for long-term rainfall projections. The third option is the transfer function–based downscaling approach, which translates the large-scale atmospheric information directly to the local-scale meteorological data using a statistical transfer function; it has been used by several researchers (Vrac et al. 2007a; Dibike et al. 2008; Olsson et al. 2004; Srinivas et al. 2014).

Multisite downscaling of rainfall is required for analyzing watersheds located in a complex terrain in which topography becomes a major factor in the generation of the rainfall process. Several studies in literature focused on multisite downscaling of rainfall (e.g., Fowler et al. 2005; Haylock et al. 2006; Vrac and Naveau 2007; Wetterhall et al. 2006). However, those studies assumed that the rainfall generation process was the same for the entire study area. Those studies considered the same set of large-scale atmospheric variables to develop a unique downscaling model for the entire study area to downscale precipitation. In a real-world scenario, the rainfall generation process is expected to change when the study area is considerably large in size. Few available approaches can address this issue. This study developed an alternative approach to identify different sets of large-scale atmospheric variables (LSAVs) at different regions of the study area and to use those LSAVs to develop different statistical downscaling models for each of those regions to obtain future projections of rainfall for the sites located in each region. The study used a global K-means (GKM) clustering algorithm to form clusters (climatological regions) in the study area and subsequently used a transfer function [support vector regression (SVR)]-based statistical downscaling approach to develop different downscaling models for each of the identified climatological regions. The performance of the proposed alternative approach was compared with that of the existing approaches which develop a unique downscaling model by assuming the same set of large-scale atmospheric variables for the entire study area. The advantages of the proposed method in downscaling rainfall were demonstrated by application to 464 rain gauges in the Republic of Ireland.

Description of Study Area and Data

The study considered 464 rain gauge stations having daily rainfall obtained from MET Eireann (Éireann 2009). The rain gauges were selected by ensuring that each gauge had at least 5 years of records between January 1979 and May 2016. The locations of the rain gauges are shown in Fig. 1 along with the elevation of the Republic of Ireland.

The goal of the study was to obtain future projections of rainfall considering the effect of climate change. For this purpose, European Centre for Medium-Range Weather Forecasts ReAnalysis (ERA) Interim data, which are available from January 1979, were used.

A set of LSAVs that can affect the rainfall events was downscaled at a $2° \times 2°$ grid scale covering the Republic of Ireland. The grids ranged from 51°N to 57°N and from 5°W to 11°W, and consisted of nine grids covering the Republic of Ireland. The three-dimensional LSAVs are available at different pressure levels. For this study, LSAV values at 17 pressure levels [100 (1,000), 92.5 (925), 85 (850), 70 (700), 60 (600), 50 (500), 40 (400), 30 (300), 25 (250), 20 (200), 15 (150), 10 (100), 7 (70), 5 (50), 3 (30), 2 (20), and 1 (10) kPa (mb)] were considered for analysis.

Proposed Methodology

This section presents the support vector regression–based rainfall downscaling approach. Subsequently, details of the GKM clustering used to identify climatological regions are provided. Following this, the theoretical background of the SVR approach and the procedure for obtaining future projections of rainfall is explained.

SVR Approach to Multisite Downscaling of Rainfall

The proposed methodology involves the following steps (Fig. 2):

1. The predictor variables are identified from large-scale atmospheric variables available in the observed/reanalysis data and GCM simulations, such that they are reasonably well correlated with historically observed rainfall (predictand) at all the target sites in the study area. In situations in which the study area is large, the sites in the study area should be delineated in clusters/climatological regions. The correlation vector between the historically observed rainfall data and the identified LSAVs corresponding to each site, along with the location indicators (latitude and longitude and elevation of the rain gauges), can be used as attributes to form climatological regions using a clustering technique. This study used partition-based global K-means clustering (Likas et al. 2003) to form the climatological regions, because the climatological regions must exhibit hard boundaries. Advantages of GKM are that, unlike conventional K-means clustering, the GKM is not sensitive to the initial conditions and it can identify the global optimal solution while optimizing the objective function to identify clusters. Other hard clustering techniques such as entropy-based clustering or hierarchical clustering can be an alternative option to form climatological regions. Most of the clustering techniques fail to form efficient clusters when the data used to form clusters are nonlinear and the regions/clusters cannot be separated by using linear planes. The objective in forming climatological regions is to increase the intersite correlation between the predictands in each region. In situations in which observations of LSAVs are unavailable, reanalysis data can be used as a surrogate for the analysis. To obtain future projections of rainfall (predictand), general circulation model simulations can be used, which provide the simulated values of LSAVs for the future in different climate change scenarios. The GCM data are available at various time scales, such as 6-hourly, daily, and monthly scales. However, GCM simulations available at finer time scales are not considered reliable (Prudhomme et al. 2002, p. 1,138; Brown et al. 2008, p. 20), and simulations at coarser time scales (e.g., monthly) are preferred. If the spatial resolution of reanalysis data and GCM data are different, the GCM data should be spatially interpolated to the resolution of the reanalysis data using software such as GrADS (Doty and Kinter 1993).

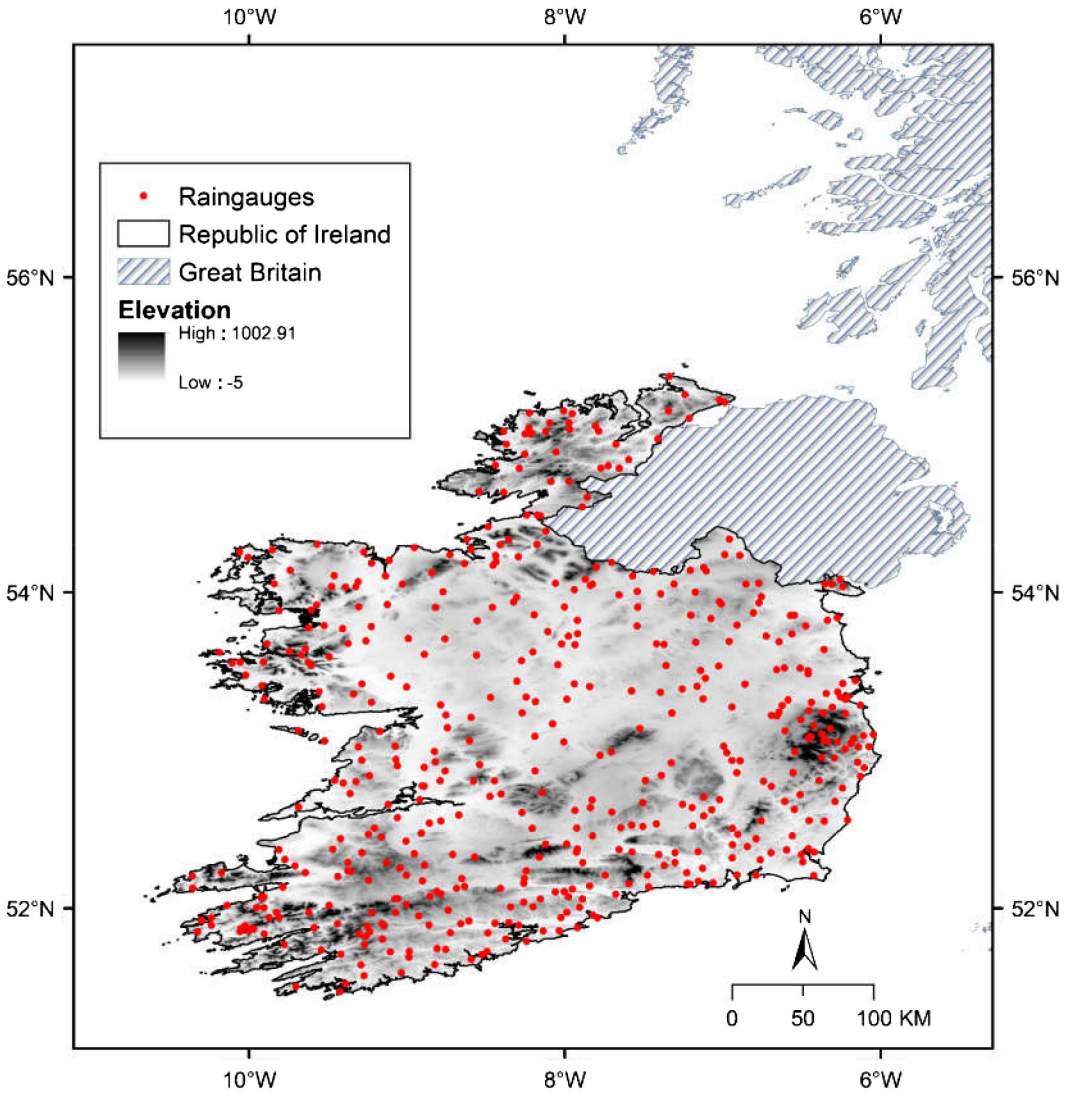

Fig. 1. Location of 464 rain gauges in the Republic of Ireland.

2. Once the clusters are formed, sets of predictor variables for each cluster are identified using the correlation between the LSAVs and the rainfall data. The identified predictor variables henceforth are referred to as large-scale atmospheric predictor variables (LSAPVs). The LSAPVs for separate climatological regions can be different from each other.
3. To develop the SVR-based downscaling model, the available historical data are divided into two subgroups. The first subgroup is called the calibration set, and the second is called the validation set. The SVR model is developed using data from the calibration set, and the performance of the model is tested based on data from the validation set.
4. Each of the identified monthly LSAPV data sets from the calibration set corresponding to each of the regions is standardized by subtracting the respective mean and dividing by the respective standard deviation. Standardization of the LSAPV is necessary to nullify the effect of differences in magnitude, range, and variance of values corresponding to the LSAPVs. Reduction of dimensionality of LSAPVs and intercorrelation between the LSAPVs can be achieved by preforming principal component analysis. However, because some information will be lost due to omission of some of the principal components, it was considered in this study.
5. Development of the SVR relationship is established for every site in the region between the LSAPVs and the observed monthly rainfall data. Because the optimal parameters of the SVR model are not known a priori, a grid search procedure (Gestel et al. 2004) is used to identify the optimal values of those parameters for the cluster. The parameters for which the model output (downscaled monthly values of rainfall for all sites in the region) is the closest [quantified in terms of RMS error (RMSE) and Nash–Sutcliffe error (NSE)] to the contemporaneous observed monthly rainfall at the corresponding target sites are considered to be the optimal. The selected optimal parameters are the same for all the sites in a region, and the developed SVR model is called the regional SVR downscaling model. The equations to estimate RMSE and NSE are as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{1}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{2}$$

where $y_i i = 1, \ldots, n)$ = observed monthly mean rainfall; n = number of observations; $\bar{y}_i$ = mean of y_i observations; and

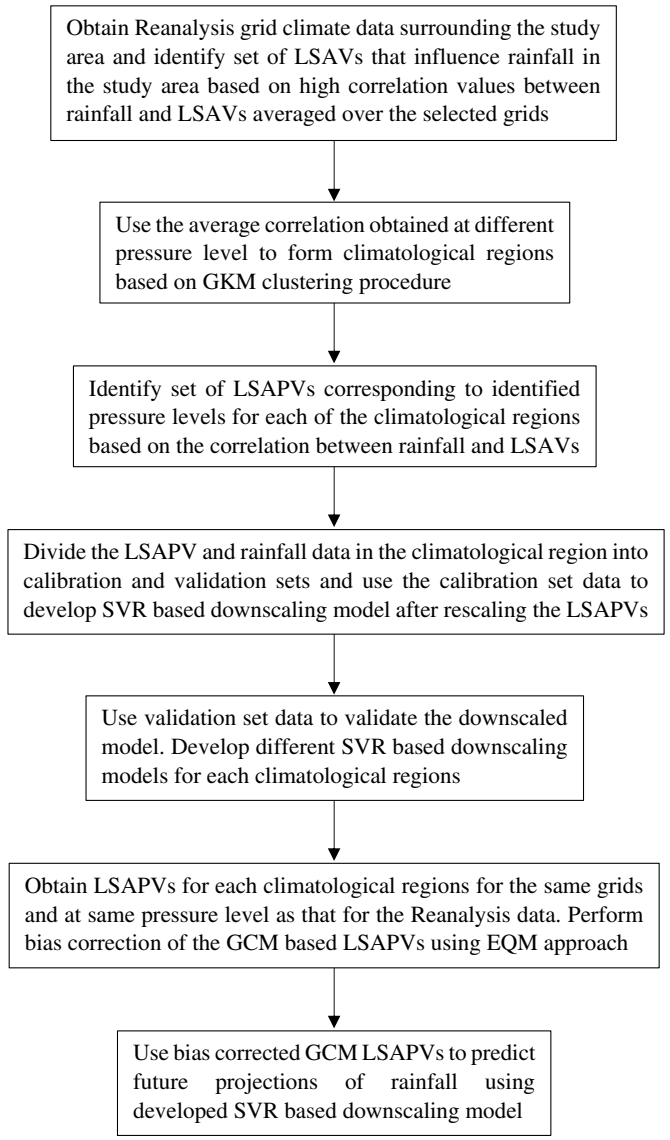

Fig. 2. Flowchart of the proposed methodology for multisite downscaling of rainfall.

$\hat{y}_i$ = predicted monthly mean rainfall based on the downscaling model.

6. The developed regional SVR downscaling model then is validated by using data from the validation set. For this purpose, each of the LSAPVs from the validation set is standardized by subtracting the respective mean and dividing by the respective standard deviation. The standardized LSAPVs then are provided to the regional SVR downscaling model to obtain downscaled values of monthly rainfall outputs. The modeled outputs of monthly rainfall for each site in a region are compared with the corresponding values of observed rainfall for the validation period, and the performance of the modeled output is quantified in terms of the two performance measures, RMSE and NSE. In situations in which the model outputs are close to the observed rainfall, the RMSE should be small and the NSE should be close to unity. When the model-predicted rainfall for all sites in the study area is close to the observed rainfall values, the developed model can be used to obtain future projections of rainfall for those rain gauge sites. In situations in which the performance is poor, it can be concluded that the selected LSAPVs cannot model the rainfall process and a new set of LSAPVs need to be identified for developing an effective downscaling model.
7. Once the model has been developed and validated, it can be used to obtain future projections of rainfall for all the sites in a cluster/climatological region by using the developed downscaling model for that region. The steps to obtain future projections of mean monthly rainfall are as follows:
   a. The selected LSAPVs extracted from the GCM simulations for the historical period are collated. Because the values in GCM have inherent biases, it is necessary to perform a bias correction before the future projections of LSAPVs from the GCM simulations can be used to obtain future projections of rainfall. To perform bias correction, the equidistant quantile matching (EQM) bias correction approach (Li et al. 2010) is used. EQM-based bias correction explicitly considers the changes in the distribution of the future climate, including the upper tail of the distribution involving extreme rainfall events (Li et al. 2010). Furthermore, EQM bias correction is applicable to smaller as well as larger spatial domain, whereas most of the existing bias correction approaches were developed based on statistical moments or regression equations and are constrained to local-scale application.
   b. The bias-corrected future projections of GCM-simulated LSAPVs then are standardized and considered as an input to the developed SVR-based downscaling model to obtain future projections of monthly mean rainfall for each site in the cluster/climatological region.

Global K-Means Clustering Algorithm

Consider that there are N rain gauge stations in the study area. Furthermore, assume that the number of LSAVs considered is equal to l. Corresponding to every pressure level for every LSAV, the average correlation between rainfall and nine LSAV grids surrounding the study area are estimated. Those correlation values along with the three location details (latitude, longitude, and elevation) are considered to create feature vector for each rain gauge station. Denote the feature vector as $\{\boldsymbol{z}'_s = [z'_{s,1}, \ldots, z'_{s,L}], s = 1, \ldots, N\}$, where L is the dimension of the feature vector. The feature vectors are rescaled to $\{\boldsymbol{z}_s = [z_{s,1}, \ldots, z_{s,L}], s = 1, \ldots, N\}$ as

$$z_{s,j} = (z'_{s,j} - \bar{z}'_j)/\sigma_j \quad \text{for } 1 \le j \le L, 1 \le s \le N \tag{3}$$

where $\bar{z}'_j = \sum_{s=1}^{N} z'_{s,j}/N$ and $\sigma_j = \sqrt{\sum_{s=1}^{N}(z'_{s,j} - \bar{z}'_j)/(N-1)}$. Rescaling of the attributes is necessary to nullify influences associated with the magnitude and variance of their values in cluster formation.

The rescaled feature vectors $Z = [\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N]^{\mathrm{T}}$ are partitioned into K_{opt} clusters (climatological regions) using global K-means clustering algorithm. Details of the GKM steps that were adapted from Basu and Srinivas (2016) are provided in the Appendix.

Support Vector Regression

This section describes the procedure to develop the SVR relationship (Vapnik 1995) in terms of input vector (predictor) $\boldsymbol{x}_t = [x_{t,1}, \ldots, x_{t,m}] \in \Re^m$, where m denotes the number of predictors, and output (predictand) $y_t \in \Re$, corresponding to time $t = \{1, 2, \ldots, n\}$.

The relationship between $\boldsymbol{x}_t$ and y_t can be expressed as

$$y_t = f(\boldsymbol{x}_t) + \varepsilon_t \tag{4}$$

where $\{f(\cdot); \Re^m \rightarrow \Re\}$ = nonlinear transformation function; and ε_t = white noise whose expected value $E[\varepsilon_t]$ is zero.

Let the function $\phi(\cdot)$ map $\boldsymbol{x}_t$ at a higher p-dimensional space, where a linear relationship exists between $\phi(\boldsymbol{x}_t)$ and y_t. The linear relationship can be expressed as

$$[y_t]_{1\times 1} = [\phi(\boldsymbol{x}_t)]_{1\times p}[\boldsymbol{w}]_{p\times 1} + [b]_{1\times 1} \tag{5}$$

where weights

$$[\boldsymbol{w}]_{p\times 1} = [\, w_1 \quad w_2 \quad \cdots \quad w_p \,]^{\mathrm{T}} \tag{6}$$

The parameters $w_1, w_2, \ldots, w_p$, and b can be estimated by optimizing the following objective function:

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{e}) = \min_{\boldsymbol{w},b,\boldsymbol{e}} \left[\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\frac{1}{2}\sum_{t=1}^{n} e_t^2 \right] \tag{7}$$

where error

$$e_t = y_t - \phi(\boldsymbol{x}_t)\boldsymbol{w} - b \tag{8}$$

$[\boldsymbol{e}]_{n\times 1} = [\, e_1 \quad e_2 \quad \cdots \quad e_n \,]^{\mathrm{T}}$; and C = SVR parameter considered to be the weight associated itho the sum of squares of error e_t at a given time. Choosing high values of the parameter reduces (increases) the error in model prediction by overfitting (underfitting) the model.

Minimization of the term $(1/2)\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$ ensures that the model is not overfitted, whereas minimization of the term $C(1/2)\sum_{t=1}^{n} e_t^2$ ensures that the model prediction error is not significantly high.

Introducing Lagrange multipliers, the optimization problem becomes

$$\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha}) = \mathcal{L}(\boldsymbol{w}, \boldsymbol{e}) - \sum_{t=1}^{n} \alpha_t [\phi(\boldsymbol{x}_t)\boldsymbol{w} + b + e_t - y_t] \tag{9}$$

where α_t = Lagrange multipliers; and $[\boldsymbol{\alpha}]_{n\times 1} = [\, \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n \,]^{\mathrm{T}}$. The optimal solution can be obtained by optimizing the function $\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha})$ using the following conditions:

$$\frac{\partial \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha})}{\partial \boldsymbol{w}} = 0; \Rightarrow \boldsymbol{w} = \sum_{t=1}^{n} \alpha_t \phi(\boldsymbol{x}_t) \tag{10}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha})}{\partial b} = 0; \Rightarrow \sum_{t=1}^{n} \alpha_t = 0 \tag{11}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha})}{\partial e_t} = 0; \Rightarrow \alpha_t = C \times e_t \quad \forall\ t \tag{12}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{e}; \boldsymbol{\alpha})}{\partial \alpha_t} = 0; \Rightarrow \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_t) + b + e_t - y_t = 0 \quad \forall\ t \tag{13}$$

Based on the conditions in Eqs. (10)–(13), and eliminating $\boldsymbol{w}$ and $\boldsymbol{e}$

$$\sum_{t=1}^{n} \alpha_t = 0 \tag{14}$$

$$\sum_{t'=1}^{n} \alpha_{t'} \phi(\boldsymbol{x}_{t'})^{\mathrm{T}} \cdot \phi(\boldsymbol{x}_t) + b + \frac{\alpha_t}{C} = y_t; \quad \text{for} \quad \forall \quad t = 1, \ldots, n \tag{15}$$

where $\phi(\boldsymbol{x}_t)^{\mathrm{T}} \cdot \phi(\boldsymbol{x}_t)$ = dot product of input vector in high-dimensional transformed space. The dot product can be expressed by a kernel function (Vapnik 1995) given as

$$\phi(\boldsymbol{x}_i)^{\mathrm{T}} \cdot \phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{16}$$

This study considered a Gaussian radial basis kernel (RBF) function for analysis, which can be expressed as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}, \qquad \gamma > 0 \tag{17}$$

The RBF function was chosen instead of a polynomial because the RBF function gives access to any analytic functions, and it is parsimonious. For chosen values of γ and C, values of $\{\alpha_t, t = 1, \ldots, n\}$ and b are estimated by solving Eqs. (9)–(12). Subsequently, future projections of the predictand y_{t_f} can be obtained as

$$y_{t_f} = \sum_{t=1}^{n} \alpha_t \phi(\boldsymbol{x}_t)^{\mathrm{T}} \cdot \phi(\boldsymbol{x}_{t_f}) + b = \sum_{t=1}^{n} \alpha_t K(\boldsymbol{x}_t, \boldsymbol{x}_{t_f}) + b \tag{18}$$

where $\boldsymbol{x}_{t_f}$ = future projection of predictor vector corresponding to time t_f. From Eq. (12) it can be noted that $\alpha_t/c = e_t$, where e_t is the error, the expected value of which is assumed to be zero.

Results and Discussion

To identify the proper sets of LSAVs affecting rainfall for all the rain gauge stations in the study area, the correlations of rainfall data with each of the selected LSAVs at 17 pressure levels were estimated. In this study, 10 LSAVs were considered (Table 1). The average correlation between the rainfall and the surrounding 9 LSAV grids corresponding to each of the 10 variables and pressure levels are plotted in Figs. 3(a–j).

A set of variables that were highly correlated with the rainfall data for the 464 rain gauges at their corresponding pressure levels were identified (Table 1). Those identified LSAVs henceforth are called large-scale atmospheric predictor variables (LSAPVs).

One drawback of this approach to identify LSAPVs is that the same set of attributes was considered for entire study area (Republic of Ireland). In a real-world scenario, it is unlikely that the rainfall is influenced by the same climatological processes at every location in the entire country. To address this issue, the correlations between the LSAVs with the rainfall data along with three location indicators (latitude, longitude, and altitude) were considered as attributes to form climatological regions in the Republic of Ireland. The global K-means clustering technique (Basu and Srinivas 2016; Bharath et al. 2016) was used to identify those regions, and the analysis showed that the study area can be divided into six clusters/climatological regions (Fig. 4). The first cluster is in the windward

Table 1. List of LSAPVs considered for 464 rain gauges

No. V and $F(\cdot)$	LSAV	−ve	+ve
1	Specific humidity, q	200	100
2	Relative humidity, r	250	600
3	Temperature, t	700	200
4	U component of wind, u	—	150
5	V component of wind, v	—	925
6	Vertical velocity, w	500	—
7	Geopotential, z	1,000	—
8	Total column water, tcw	—	—
9	Mean sea level pressure, msl	1,000	—
10	Instantaneous surface sensible heat flux, ishf	—	—

Note: –ve (+ve) denote rainfall negatively (positively) correlated to the LSAPV.

direction of the Wicklow Mountains located in the eastern part of the country, where the wind predominantly arrives from the southwest direction. The second cluster is located in the western part of the country near the Atlantic Ocean, at medium to high altitude (Fig. 1). Cluster 3 is located in the central part of the country at low to medium altitude, Cluster 4 is located in the southern part at low altitude, and Cluster 5 is located in the northern part of the country, where the altitude is relatively high. Cluster 6 ranges along the southeastern coastline of the country at low altitude. The number of stations and average record length in each of the clusters were as follows: Cluster 1—82 stations, 22.7 years; Cluster 2—97 stations, 28.5 years; Cluster 3—98 stations, 26.1 years; Cluster 4—33

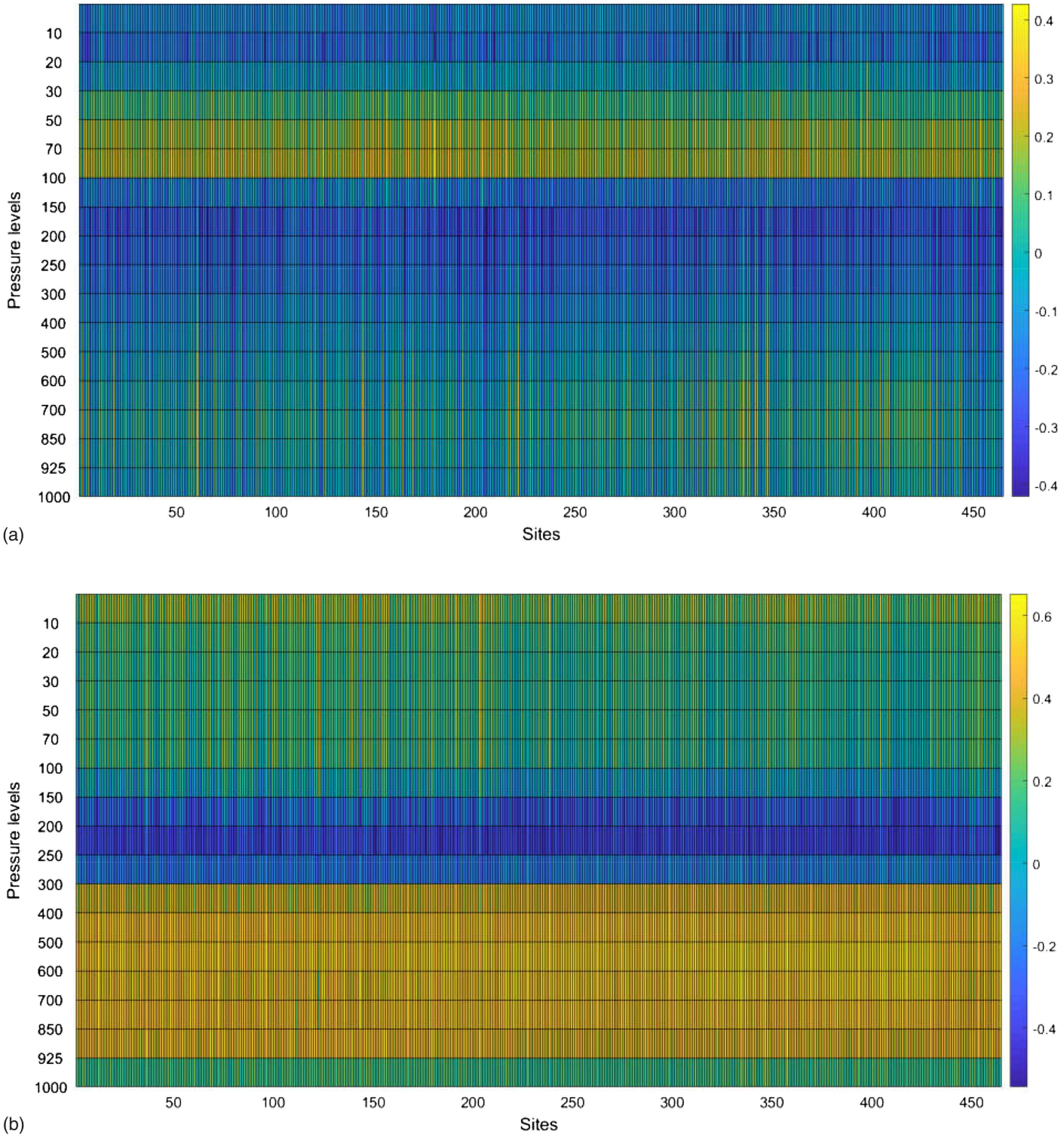

Fig. 3. Correlation plots between rainfall data for each of the 464 sites and LSAVs: (a) specific humidity (q); (b) relative humidity (r); (c) temperature (t); (d) U component of wind (u); (e) V component of wind (v); (f) vertical velocity (w); (g) geopotential (z); (h) total column water (tcw); (i) mean sea level pressure (msl); and (j) instantaneous surface sensible heat flux (ishf) corresponding to different pressure levels.

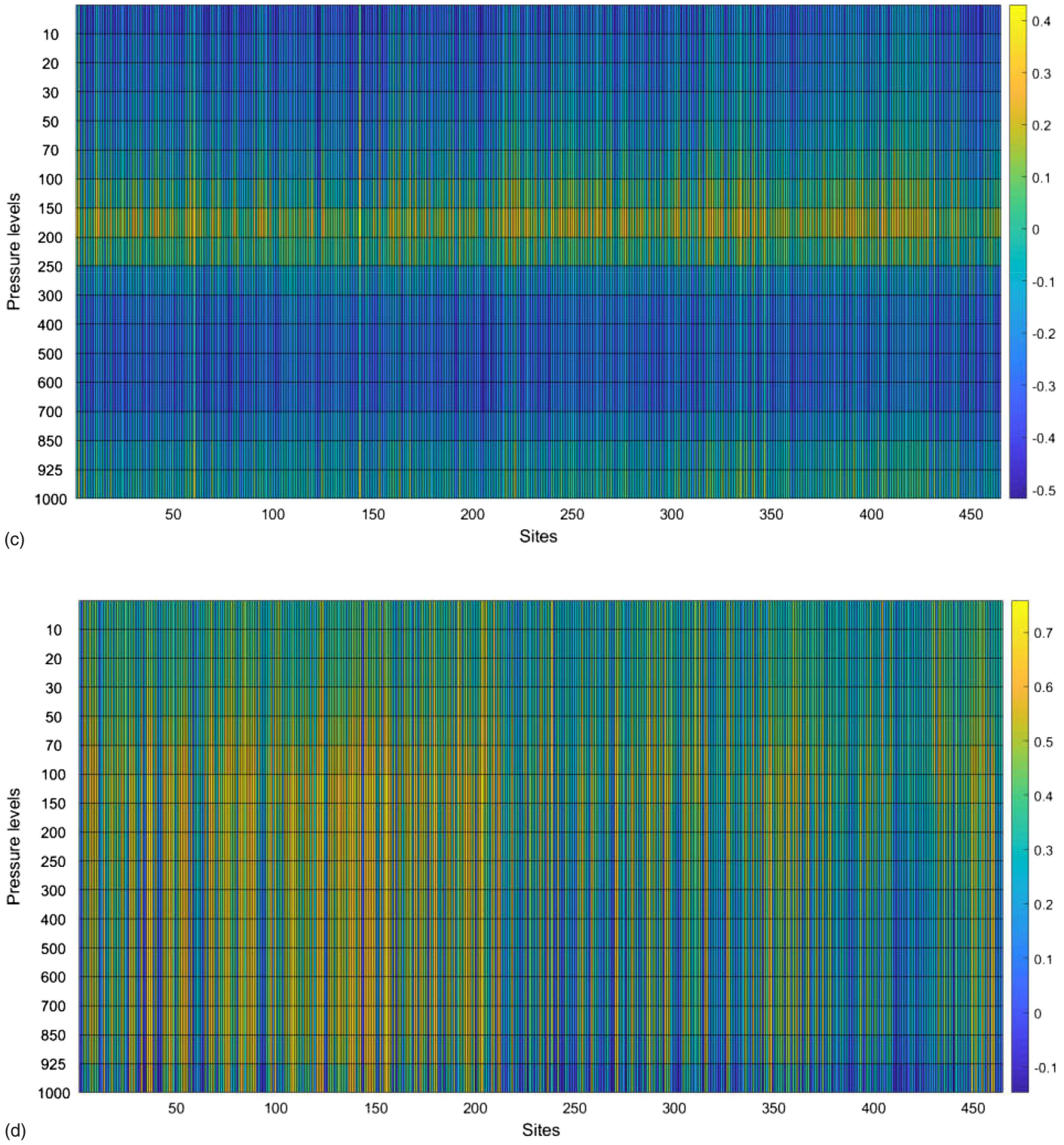

Fig. 3. *(Continued)*

stations, 12.2 years; Cluster 5—63 stations, 27.7 years; and Cluster 6—91 stations, 30 years. Except for Cluster 4, which had the least number of stations, the average record of each climatological region had precipitation records covering more than 22 years.

Intersite cross correlation for the calibration and validation data set for 464 stations was prepared (Fig. S1). Comparison of the correlations indicated that the correlations between the rain gauge stations changed considerably from the calibration period to the validation period due to the effect of climate change. The figures also indicate that the correlation structure between rain gauge stations in the study area varied considerably, justifying the requirement to develop different downscaling models in the study area. To demonstrate the advantage of forming climatological regions, correlation between sites in each of the six regions were plotted (Fig. S2). The figures in the calibration set indicate that most of the stations in those climatological regions had nearby correlations among themselves. The differences in intersite correlation between

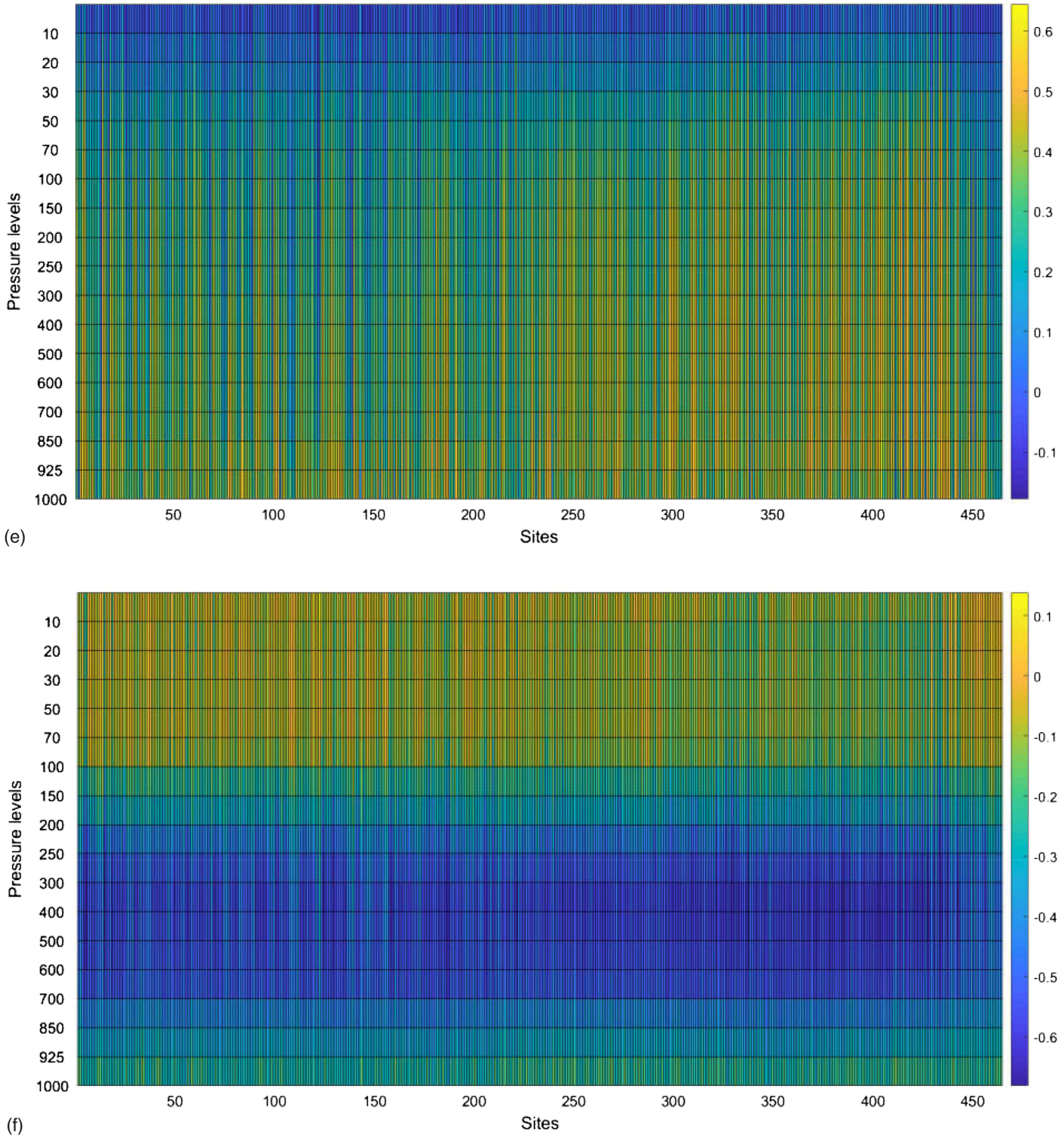

Fig. 3. *(Continued)*

rain gauge stations for the validation period was found to be higher, which might be due to the effect of climate change or to other natural or anthropogenic factors.

The large-scale atmospheric variables for each of those clusters then were identified by investigating the average correlations between the rainfall and 9 grids for each of the 17 pressure levels. The identified LSAPVs for each of the clusters are provided in Table 2. The LSAPVs corresponding to each of the clusters (climatological regions) were different from each other.

Based on the selected LSAPVs, support vector regression–based downscaling models were developed for each of the clusters in the study area. Because the model parameters γ and C were not known a priori, a grid search procedure was used to obtain the optimal parameters for the study. For this purpose, the first 75% of the data set corresponding to each rain gauge for the historical period was used for calibration, and the remaining 25% of the data set was used for validation of the model. The number of data points (in months) for the calibration and validation sets for each of the

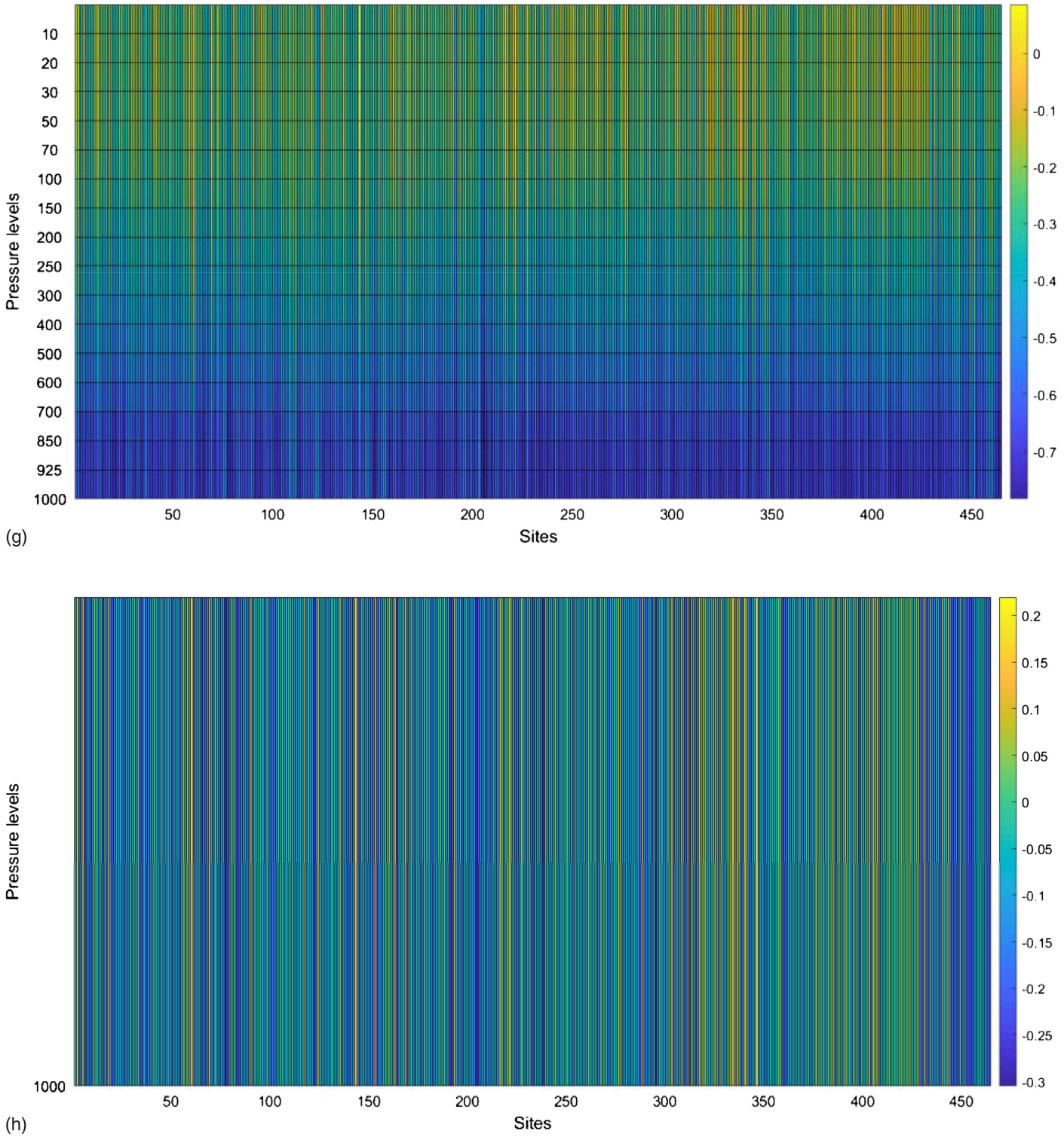

Fig. 3. *(Continued)*

464 stations was plotted as boxplot (Fig. S3). The majority of the stations had more than 150 months of data points in the calibration set and more than 50 data points in the validation set. Furthermore, the average records for the calibration and validation sets for stations corresponding to each of the climatological regions were 204.6 and 67.7 months (Cluster 1), 256.9 and 85.2 months (Cluster 2), 235.1 and 77.8 months (Cluster 3), 109.7 and 36.1 months (Cluster 4), 249.3 and 82.7 months (Cluster 5), and 270 and 90 months (Cluster 6), respectively. This indicates that, except for Cluster 4, which consisted of only 33 stations, the majority of stations had adequate number of records for developing the SVR-based downscaling model. Both SVR parameters were varied from 10 to 1,000 with an interval of 10 units, and the difference between the observed and model-predicted rainfall was measured in terms of two performance measures (NSE and RMSE). The optimal parameters of the SVR model were selected when the model yielded reasonable performance, i.e., NSE closer to unity and RMSE close to zero. All the rain gauges in a region are not

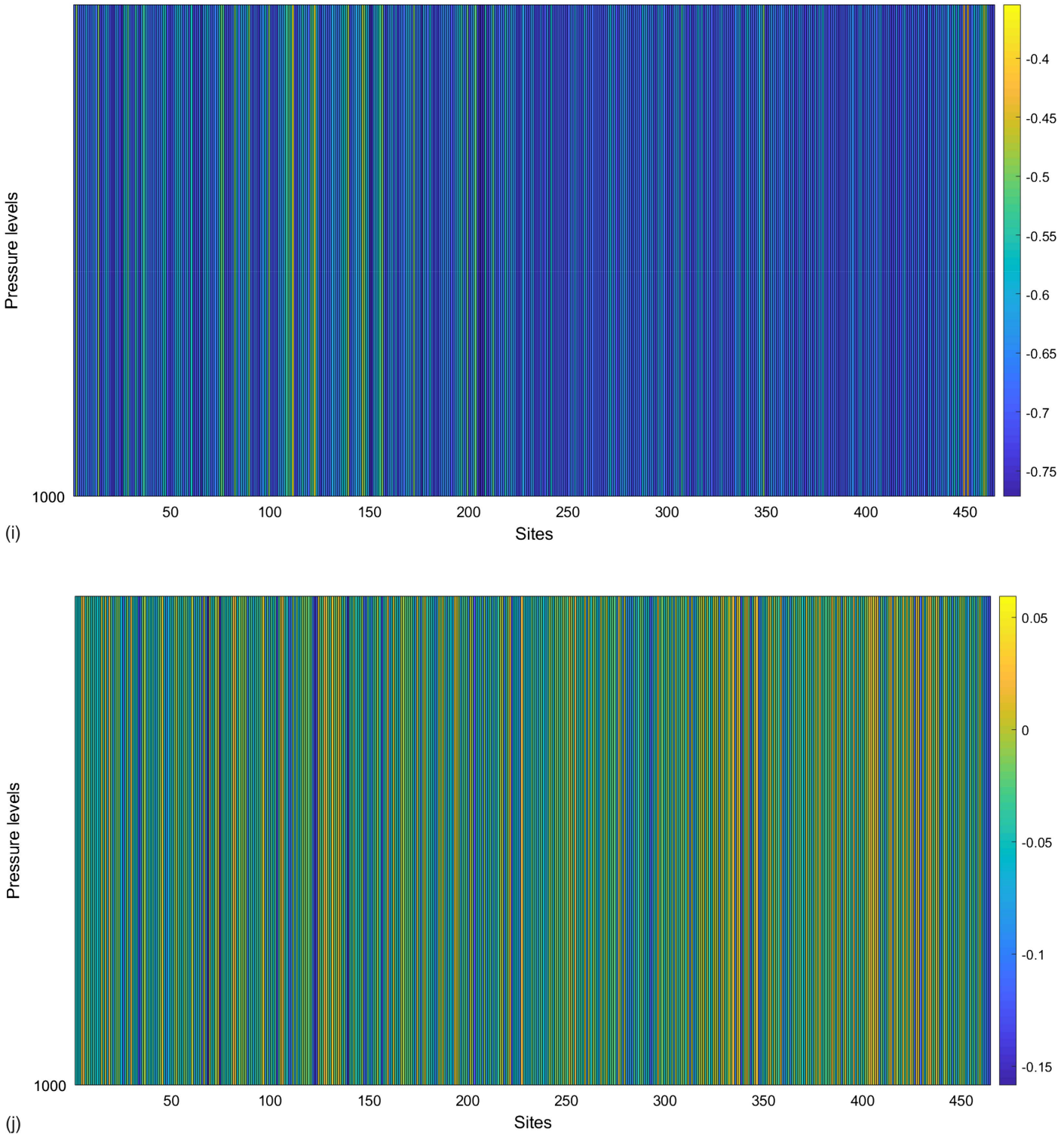

Fig. 3. *(Continued)*

expected to have the lowest error corresponding to a given parameter combination of SVR model. For this purpose, the parameter combination that was selected by the majority of the rain gauges in a climatological region was considered to be the optimal parameters for a region. The values of the optimal parameters are provided in Table 3, and the developed model is called Model 1. The performance measures (NSE and RMSE) for each of the 464 rain gauges in the calibration and validation periods for Model 1 are plotted in Figs. 5(a and b), respectively.

For comparison purposes, the experiment was repeated considering (1) the same set of LSAPVs for each of the climatological regions, but with different SVR parameters for each region (Model 2), and (2) the same set of LSAPVs and the same SVR model for each region (Model 3). The performance measures in the calibration and validation periods for Models 2 and 3 are provided as boxplots in Figs. 5(a and b) along with those for Model 1. The RMSE was the lowest and the NSE was closest to unity when different LSAPVs were considered and different SVR models were developed for each

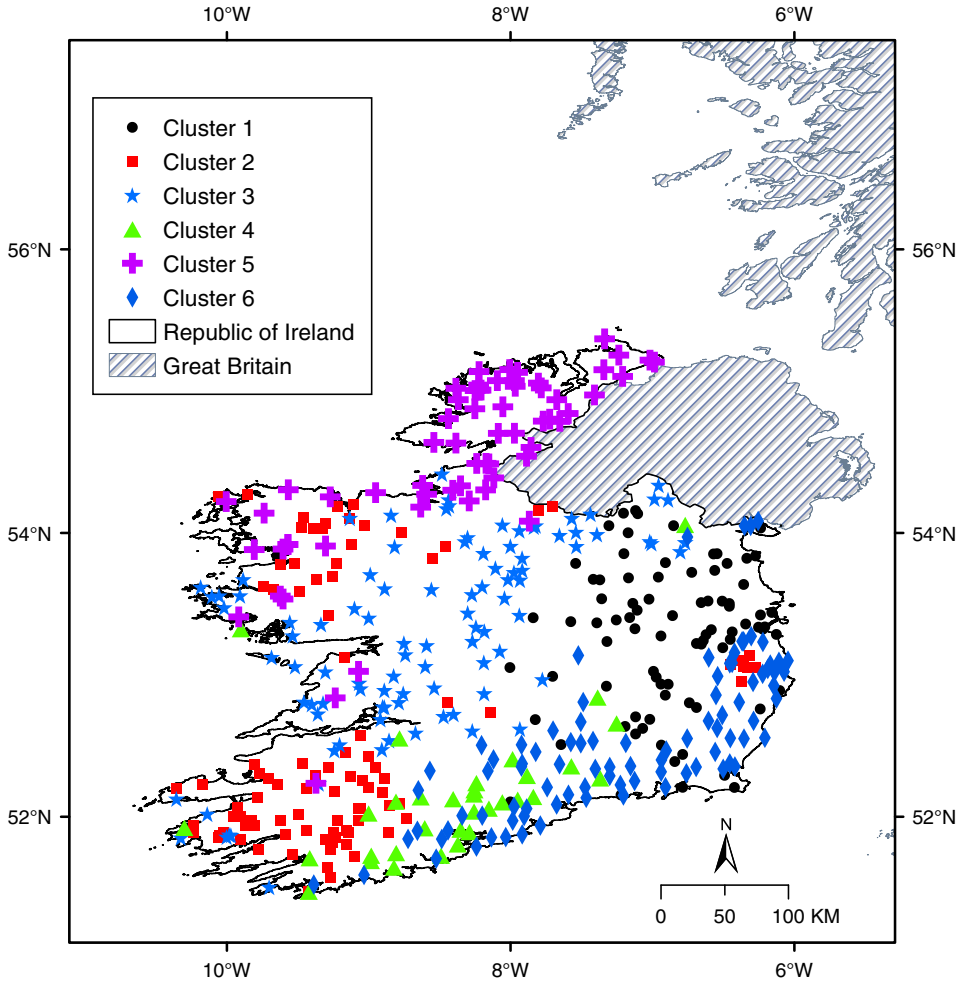

Fig. 4. Locations of climatological regions in Ireland.

Table 2. List of LSAPVs considered for rain gauges corresponding to each of the clusters

		Region 1		Region 2		Region 3		Region 4		Region 5		Region 6	
No.	Variable	−ve	+ve	−ve	+ve	−ve	+ve	−ve	+ve	−ve	+ve	−ve	+ve
1	q	150	850	200	100	200	100	200	100	200	100	200	100
2	r	200	700	250	500	250	600	250	600	250	925	250	600
3	t	700	200	600	—	700	200	500	200	700	—	700	200
4	u	—	200	—	150	—	150	—	100	—	150	—	150
5	v	—	500	—	925	—	1,000	—	925	—	1,000	—	850
6	w	400	—	500	—	400	—	400	—	700	—	400	—
7	z	1,000	—	925	—	1,000	—	1,000	—	1,000	—	1,000	—
8	tcw	—	—	1,000	—	—	—	—	—	—	—	—	—
9	msl	1,000	—	1,000	—	1,000	—	1,000	—	1,000	—	1,000	—
10	ishf	—	—	—	—	—	—	—	—	—	—	—	—

Note: −ve (+ve) denote rainfall negatively (positively) correlated to the LSAPV.

of the climatological regions. The performance decreased slightly (RMSE values increased and NSE values decreased) when the same set of LSAPVs was considered for all climatological regions but the SVR models were different for each region. The performance was inferior when the same set of LSAPVs was considered and the same SVR model was used for the entire study area. Better performance in the case of Model 1 can be attributed to the identification of climatological regions and the selection of appropriate attributes for each region. In addition, different SVR models were used to obtain future projections of rainfall for gauges located in each of those regions because the rainfall generation process for each climatological region was different. The performance decreased slightly in case of Model 2, in which different SVR models were selected but the LSAPVs were considered to be the same for the entire study area. The third approach was expected to provide inferior performance because the approach assumed that the same rainfall generation mechanism governs the rainfall for the entire study area.

Performance of the downscaled model (Model 1) in obtaining rainfall was compared with the high-resolution Copernicus Climate Change Service E-OBS v 19.0e gridded rainfall data obtained

Table 3. Optimal values of SVR parameters for each region using Model 1 and Model 2

Region	Model 1		Model 2	
	γ	C	γ	C
1	80	940	40	660
2	80	1,000	60	980
3	50	940	60	960
4	40	980	70	1,000
5	60	960	50	880
6	50	880	60	1,000

at 0.1° resolution. One representative station from each climatological region was selected, and the performance of the model predictions and the E-OBS gridded data was compared with the observed station rainfall data at the monthly scale. The performance of each station was plotted as a Taylor diagram (Taylor 2001) (Fig. S4). In the figure, the radial distance indicates the standard deviation of the monthly rainfall data, the azimuthal position provides the correlation coefficient value between the observed rainfall and the model-predicted/gridded rainfall, and the dotted line (shown in green) originating from the reference point (REF) provides the RMS distance (RMSD). In situations in which the predicted rainfall is closer to the observed rainfall, the standard deviation of the predicted rainfall is expected to be closer to that of the observed rainfall; the correlation should be closer to unity and the RMSD should be closer to zero. The figures indicate that the model-predicted rainfall had less error than that obtained from the gridded rainfall, except for the representative station located in Climatological region 5, for which the gridded rainfall was slightly better than the modeled value.

Once the SVR-based downscaling model is validated, it can be used to obtain future projections of mean monthly rainfall for all the sites in the study area. Because the performance of Model 1 was superior compared with that of the other two models, future projections of mean monthly rainfall were obtained based on this model only. For this purpose, simulations of National Aeronautics and Space Administration Goddard Institute for Space Studies General Circulation Model (NASA GISS GCM) coupled atmospheric–oceanic model data were used, corresponding to four Representative Concentration Pathways (RCPs). The resolution of the data was at a $2° \times 2.5°$ grid scale, and they were regridded to a $2° \times 2°$ grid scale that matched the ERA Interim reanalysis data spatial scale. The historical simulations of the GCM data from January 1979 to December 2005 were considered to perform a bias correction for the future projections of selected LSAPVs for the period January 2021–December 2050. The bias-corrected future projections of LSAPVs corresponding to each of the four RCP scenarios (RCP2.6, RCP4.5, RCP6.0, and RCP8.5) were used in the developed SVR-based downscaling model (Model 1) to obtain future projections of mean monthly rainfall at each of the 464 rain gauge sites. The change in the mean for three decades (2021–2030, 2031–2040, and 2041–2050) was estimated for each site in every cluster (climatological region) corresponding to each of the four RCP scenarios and were compared with the historical mean monthly rainfall. The results corresponding to each of the six climatological regions (clusters) are shown as boxplots in Figs. 6(a–f). For Clusters 1, 4, and 6 (located in the eastern and southeastern part of the Republic of Ireland), the mean monthly rainfall is expected to increase in the future according to the RCP2.6 and RCP8.5 scenarios, whereas based on scenarios RCP 4.5 and RCP 6.0, the mean monthly rainfall is expected to increase in the next two decades (2021–2040) and to decrease after 2040. For Cluster 3, which is located in the central part of the country, the mean rainfall is expected to increase in the future according to the RCP8.5 scenario. For other locations and corresponding to all scenarios, the mean monthly rainfall is

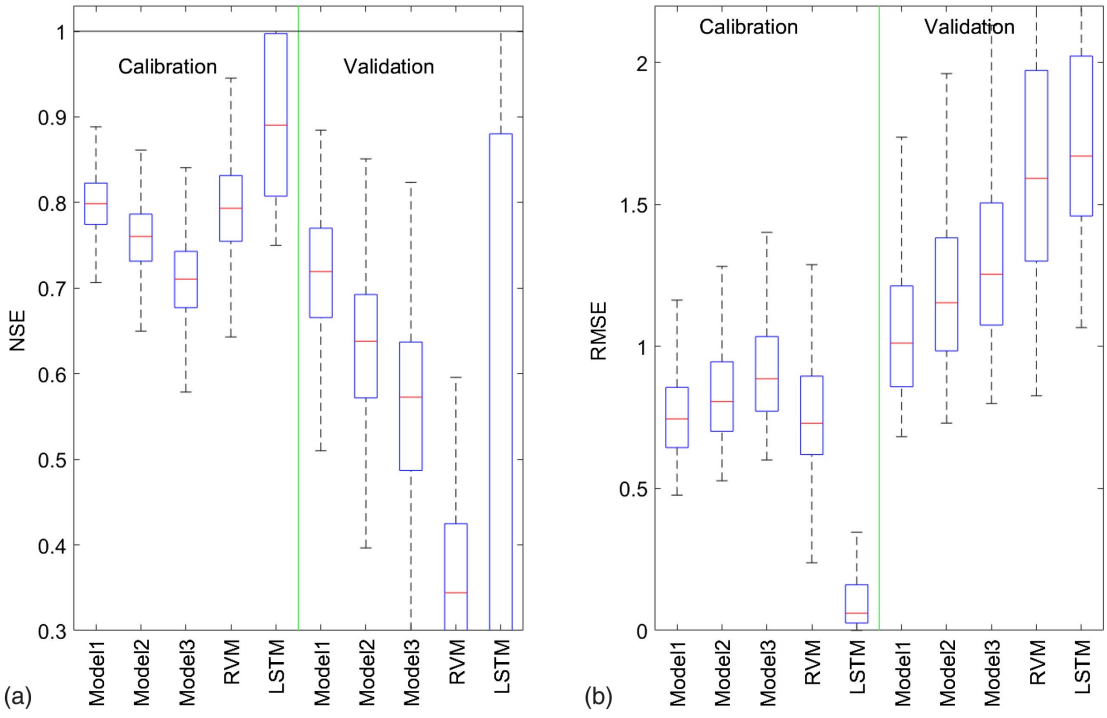

Fig. 5. Two performance measures obtained in calibration and validation by (1) identifying a set of attributes in each cluster and developing different SVR models for each cluster (Model 1); (2) considering the same set of attributes but developing different SVR models for each cluster (Model 2); (3) considering the same set of attributes and the same model for the entire study area (Model 3); (4) RVM-based downscaling considering different LSAPVs; and (5) LSTM-based downscaling considering different LSAPVs: (a) NSE; and (b) RMSE.

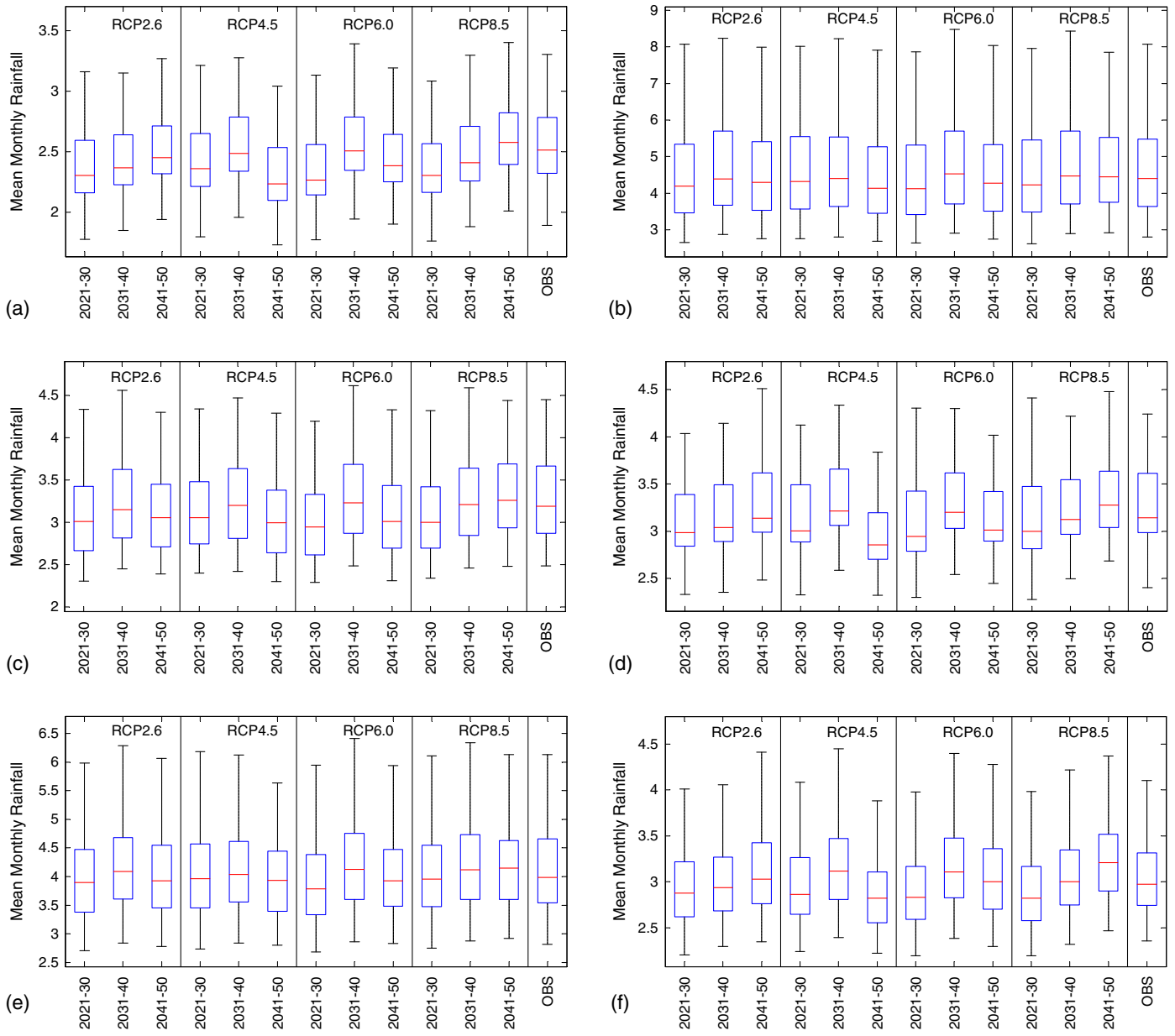

Fig. 6. Mean projected monthly rainfall for each cluster averaged over 1 decade, corresponding to four climate change scenarios: (a) Cluster 1; (b) Cluster 2; (c) Cluster 3; (d) Cluster 4; (e) Cluster 5; and (f) Cluster 6. OBS = mean monthly rainfall of historically observed rainfall data.

expected to decrease after 2040. RCP2.6 assumes that the global annual greenhouse gas (GHG) emissions (measured in equivalent CO_2) will peak between 2010–2020 and decrease after 2020; in the case of the RCP4.5 scenario, the emissions peak around 2040 and then decrease; for RCP6.0, the emissions are assumed to peak around 2080 and then decline; whereas RCP8.5 assumes a continuous increase in emissions that continue to rise throughout the 21st century (Meinshausen et al. 2011). Chandler and Wheater (2002) considered generalized linear models to relate the changes in rainfall pattern at stations located in western Ireland with the North Atlantic Oscillation (NAO) index. Nolan et al. (2017) used a RCM model to investigate the effect of climate change in the rainfall pattern across Ireland after the 2040s. They noted that the mean annual rainfall as well as rainfall in the summer and spring seasons are expected to decrease, whereas the frequency of heavy rainfall events in the winter months are supposed to increase.

Currently, with advancement of computational facilities, robust neural network–based techniques are gaining popularity for developing complex relationships between high-dimensional predictor and predictand data sets. The relevance vector machine (RVM)-based downscaling approach has been considered in a few studies (Ghosh and Mujumdar 2008; Okkan and Inan 2015; Joshi et al. 2015; Deo et al. 2016) to downscale different hydroclimatological variables. The statistical framework of RVM is the same as that of the support vector machine algorithm; however, RVM considers an alternative functional algorithm that develops a probabilistic regression between the predictors and predictant. Details of the RVM approach were given by Tipping (2001). Recently, deep learning–based techniques such as long short-term memory (LSTM) are becoming popular due to their better performance when trained with large amounts of data compared with other neural network–based approaches. This study further investigated the use of the LSTM algorithm in downscaling the rainfall data for the study area.

The LSTM algorithm was considered to have 100 hidden units, and the maximum iteration was set to 250. Technical details of LSTM algorithm were given by Hochreiter and Schmidhuber (1997).

The performance of the RVM- and LSTM-based downscaling approaches was compared with that of the SVR-based downscaling approach for the Model 1 case, in which different sets of LSAPVs were selected for each climatological region. Only the Model 1 case was considered because the other two models (Model 2 and Model 3) consider unrealistic assumptions, leading to poor performance in downscaling rainfall. The performance of the two new approaches is provided in terms of two performance measures (RMSE and NSE) in Figs. 5(a and b). Results indicate that the performance of LSTM downscaling was superior in the calibration period, followed by that of the RVM-based downscaling approach. However, the model performance of RVM and LSTM for the validation data set was found to be considerably inferior to that obtained using the SVR-based downscaling approach. Because the RVM considers the posterior distribution and LSTM can retrain the data set based on the model performance, prediction of rainfall for the calibration set using both approaches was better than that using the SVR-based approach. On the other hand, both RVM and LSTM require larger data sets for training the model, and because the monthly rainfall data for each station were limited, the performance of those models in the validation set was poor compared with that of the SVR-based model.

Conclusions

The paper presented a new approach to form climatological regions using a global K-means clustering algorithm where the rainfall generation mechanism can be considered to be similar. Subsequently, support vector regression–based statistical downscaling models were developed to obtain future projections of mean monthly rainfall at rain gauges located in each of those climatological regions. The effectiveness of the new approach was demonstrated through a case study of 464 rain gauges in the Republic of Ireland. The results indicated that the new approach provides better rainfall projections than the existing downscaling approaches which assume the same rainfall generation process for the entire study area. The newly developed downscaling model subsequently was used to obtain future projections of rainfall for the period 2021–2050 for four Representative Concentration Pathways scenarios. In general, the mean rainfall is expected to increase until 2040 and then to decrease for all scenarios in the western part of the country, whereas in the eastern part of the country the rainfall is expected to increase after 2040 according to two climate change scenarios. Detailed analysis is underway to attribute the changes in the rainfall pattern in the Republic of Ireland. The performance of the rainfall downscaling technique proposed in the study depends on the developed climatological regions as well as on the model used to develop regression relationships between LSAPVs and historical rainfall. Research is underway to evaluate effectiveness of the climatological regions and to explore other nested downscaling approaches that can account for the physical rainfall generation phenomena.

Appendix. Details of Global *K*-Means Algorithm

The method attempts to minimize the following objective function:

$$F(\boldsymbol{Z}, \boldsymbol{V}{:}K) = \sum_{s=1}^{N} \sum_{i=1}^{K} I(\boldsymbol{z}_s \in C_i) d^2(\boldsymbol{z}_s, \boldsymbol{v}_i) \tag{19}$$

where $C_i = i$th hard cluster; $I = 1$ if $\boldsymbol{z}_s \in C_i$ is true, and $I = 0$ otherwise; $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K]^{\mathrm{T}}$ is a matrix containing centroids of K clusters, such that $\boldsymbol{v}_i = [v_{i,1}, \ldots, v_{i,l}] \in \Re^l$; $d^2(\cdot, \cdot)$ = square of distance measure, which was considered to be Euclidean in this study; and $v_{i,j}$ = mean value of attribute j for cluster i.

1. Define $K_{\min}$ and $K_{\max}$ as the lower and the upper bounds, respectively, of the possible number of clusters. Initialize K to $K_{\min}$ (= 2).
2. Compute the mean of N feature vectors $\{\boldsymbol{z}_s, s = 1, \ldots, N\}$ [Eq. (20)] and choose it as the centroid of the first cluster. Set epoch to 1.

$$v_{i,j} = \sum_{s=1}^{N} z_{s,j}/N \quad j = 1, \ldots, l \tag{20}$$

3. Choose $\boldsymbol{z}_{\text{epoch}}$ as the centroid of the Kth cluster (i.e., $\boldsymbol{v}_K = \boldsymbol{z}_{\text{epoch}}$) and set iteration count to 1.
4. Determine the Euclidean distance of each feature vector $\boldsymbol{z}_s$ from the centroids of each K cluster and assign the vector to the cluster whose centroid is nearest to it.
5. Update the centroid of each cluster by computing the average of the feature vectors assigned to it. Then compute the value of the objective function $F(\cdot)$ for the current iteration. If iteration count = 1, increment the count by 1 and proceed to Step 4. Otherwise, compute the difference in the value of $F(\cdot)$ between the current and previous iterations. If the difference is sufficiently small (<0.001), store $\boldsymbol{V}$ and $F(\cdot)$ and proceed to Step 6. Otherwise, increment iteration count by 1. If the count is less than or equal to a prespecified upper bound on the number of iterations (e.g., 500), proceed to Step 4; otherwise store and proceed to Step 6.
6. Increment epoch by 1. If the resulting number is less than or equal to N, repeat Steps 3–5. Otherwise, proceed to Step 7.
7. Identify the epoch h that yielded the minimum value for the objective function $F(\cdot)$ and note the corresponding $\boldsymbol{V}$ as $\boldsymbol{V}^o(K) = [v_1^o, \ldots, v_K^o]^{\mathrm{T}}$ that denotes the optimal cluster centroid matrix for the case in which the number of clusters is equal to K.
8. If $K < K_{\max}$, increment K by 1 and consider the first $K - 1$ cluster centroids as $\boldsymbol{V}^o(K-1)$, set epoch to 1, and repeat Steps 3–7. Otherwise, proceed to Step 9.
9. From among the partitions resulting for all the K values ($2 \le K \le K_{\max}$), identify the optimal partition K_{opt} as that for which the value of the Davies–Bouldin cluster validity index (Davies and Bouldin 1979) is minimum. The index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation. It is computed as

$$\text{DBI}(K) = \frac{1}{K} \sum_{i=1}^{K} \max_{i, i \neq r} \left[\frac{\text{Scatter}_i + \text{Scatter}_r}{d_{i,r}} \right] \tag{21}$$

where Scatter_i and Scatter_r = within-cluster scatter for ith and rth clusters, respectively. In general, Scatter_i is computed as $\text{Scatter}_i = \sqrt{\sum_{z_s \in C_i} \|\boldsymbol{z}_s - \boldsymbol{v}_i^o\|^2 / N_i}$, $d_{i,r}$ is the Euclidean distance between the centroids of the ith and rth clusters, and N_i is the number of sites in cluster i.

Acknowledgments

The authors gratefully acknowledge the support of the Risk Analysis of Infrastructure Networks (RAIN) Project, Grant No. 608166, which is funded by European Union's Seventh Framework Programme for Research, Technological Development and Demonstration Activities.

Supplemental Data

Figs. S1–S4 are available online in the ASCE Library (www.ascelibrary.org).

References

Ahmed, K. F., G. Wang, J. Silander, A. M. Wilson, J. M. Allen, R. Horton, and R. Anyah. 2013. "Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the US northeast." *Global Planet. Change* 100 (Jan): 320–332. https://doi.org/10.1016/j.gloplacha.2012.11.003.

Basu, B., and V. V. Srinivas. 2016. "Regional flood frequency analysis using entropy-based clustering approach." *J. Hydrol. Eng.* 21 (8): 04016020. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001351.

Bharath, R., V. V. Srinivas, and B. Basu. 2016. "Delineation of homogeneous temperature regions: A two-stage clustering approach." *Int. J. Climatol.* 36 (1): 165–187. https://doi.org/10.1002/joc.4335.

Brown, C., A. Greene, P. Block, and A. Giannini. 2008. *Review of downscaling methodologies for Africa climate applications.* IRI Technical Rep. No. 08-05. New York: Columbia Univ.

Chandler, R. E., and H. S. Wheater. 2002. "Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland." *Water Resour. Res.* 38 (10): 10-1. https://doi.org/10.1029/2001WR000906.

Davies, D. L., and D. W. Bouldin. 1979. "A cluster separation measure." *IEEE Trans. Pattern Anal. Mach. Intell.* 1 (2): 224–227.

Deo, R. C., P. Samui, and D. Kim. 2016. "Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models." *Stochastic Environ. Res. Risk Assess.* 30 (6): 1769–1784. https://doi.org/10.1007/s00477-015-1153-y.

Dibike, Y. B., P. Gachon, A. St-Hilaire, T. B. M. J. Ouarda, and V. T. V. Nguyen. 2008. "Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada." *Theor. Appl. Climatol.* 91 (1–4): 149–170. https://doi.org/10.1007/s00704-007-0299-z.

Diez, E., C. Primo, J. A. Garcia-Moya, J. M. Gutiérrez, and B. Orfila. 2005. "Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts." *Tellus A: Dyn. Meteorol. Oceanogr.* 57 (3): 409–423. https://doi.org/10.1111/j.1600-0870.2005.00130.x.

D'Onofrio, D., E. Palazzi, J. von Hardenberg, A. Provenzale, and S. Calmanti. 2014. "Stochastic rainfall downscaling of climate models." *J. Hydrometeorol.* 15 (2): 830–843. https://doi.org/10.1175/JHM-D-13-096.1.

Doty, B., and J. L. Kinter, III. 1993. "The grid analysis and display system (GrADS): A desktop tool for earth science visualization." In *Proc., American Geophysical Union 1993 Fall Meeting*, 6–10. Washington, DC: American Geophysical Union.

Eden, J. M., and M. Widmann. 2014. "Downscaling of GCM-simulated precipitation using model output statistics." *J. Clim.* 27 (1): 312–324. https://doi.org/10.1175/JCLI-D-13-00063.1.

Éireann, M. 2009. "Climate of Ireland." Met Eireann. Accessed September 1, 2009. https://www.met.ie/climate/available-data/historical-data.

Fowler, H. J., S. Blenkinsopa, and C. Tebaldi. 2007. "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modeling." *Int. J. Climatol.* 27 (12): 1547–1578. https://doi.org/10.1002/joc.1556.

Fowler, H. J., C. G. Kilsby, P. E. O'Connell, and A. Burton. 2005. "A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change." *J. Hydrol.* 308 (1–4): 50–66. https://doi.org/10.1016/j.jhydrol.2004.10.021.

Fowler, H. J., and R. L. Wilby. 2010. "Detecting changes in seasonal precipitation extremes using regional climate model projections: Implications for managing fluvial flood risk." *Water Resour. Res.* 46 (3): W03525. https://doi.org/10.1029/2008WR007636.

Gestel, T. V., J. A. K. Suykens, B. Baesens, S. Viaene, J. Vanthienen, G. Dedene, B. D. Moor, and J. Vandewalle. 2004. "Benchmarking least squares support vector machine classifiers." *Mach. Learn.* 54 (1): 5–32. https://doi.org/10.1023/B:MACH.0000008082.80494.e0.

Ghosh, S., and P. P. Mujumdar. 2008. "Statistical downscaling of GCM simulations to streamflow using relevance vector machine." *Adv. Water Resour.* 31 (1): 132–146. https://doi.org/10.1016/j.advwatres.2007.07.005.

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess. 2006. "Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios." *Int. J. Climatol.* 26 (10): 1397–1415. https://doi.org/10.1002/joc.1318.

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." *Clim. Res.* 19 (1): 45–55. https://doi.org/10.3354/cr019045.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short-term memory." *Neural Comput.* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Joshi, D., A. St-Hilaire, T. B. M. J. Ouarda, and A. Daigle. 2015. "Statistical downscaling of precipitation and temperature using sparse Bayesian learning, multiple linear regression and genetic programming frameworks." *Can. Water Resour. J.* 40 (4): 392–408. https://doi.org/10.1080/07011784.2015.1089191.

Landman, W. A., M. J. Kgatuke, M. Mbedzi, A. Beraki, A. Bartman, and A. D. Piesanie. 2009. "Performance comparison of some dynamical and empirical downscaling methods for South Africa from a seasonal climate modelling perspective." *Int. J. Climatol.: J. R. Meteorol. Soc.* 29 (11): 1535–1549. https://doi.org/10.1002/joc.1766.

Li, H., J. Sheffield, and E. F. Wood. 2010. "Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching." *J. Geophys. Res.* 115 (D10): 1–20. https://doi.org/10.1029/2009JD012882.

Likas, A., N. Vlassis, and J. J. Verbeek. 2003. "The global *k*-means clustering algorithm." *Pattern Recognit.* 36 (2): 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2.

Meinshausen, M., et al. 2011. "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300." *Clim. Change* 109 (1–2): 213–241. https://doi.org/10.1007/s10584-011-0156-z.

Murphy, J. 1999. "An evaluation of statistical and dynamical techniques for downscaling local climate." *J. Clim.* 12 (8): 2256–2284. https://doi.org/10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2.

Nguyen, V. T. V., T. D. Nguyen, and P. Gachon. 2006. "On the linkage of large-scale climate variability with local characteristics of daily precipitation and temperature extremes: An evaluation of statistical downscaling methods." In *Advances in geosciences: Volume 4: Hydrological science (HS)*, 1–9. Singapore: World Scientific.

Nolan, P., J. O'Sullivan, and R. McGrath. 2017. "Impacts of climate change on mid-twenty-first-century rainfall in Ireland: A high-resolution regional climate model ensemble approach." *Int. J. Climatol.* 37 (12): 4347–4363. https://doi.org/10.1002/joc.5091.

Okkan, U., and O. Fistikoglu. 2014. "Evaluating climate change effects on runoff by statistical downscaling and hydrological model GR2M." *Theor. Appl. Climatol.* 117 (1–2): 343–361. https://doi.org/10.1007/s00704-013-1005-y.

Okkan, U., and G. Inan. 2015. "Bayesian learning and relevance vector machines approach for downscaling of monthly precipitation." *J. Hydrol. Eng.* 20 (4): 04014051. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001024.

Olsson, J., K. Berggren, M. Olofsson, and M. Viklander. 2009. "Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden." *Atmos. Res.* 92 (3): 364–375. https://doi.org/10.1016/j.atmosres.2009.01.015.

Olsson, J., C. B. Uvo, K. Jinno, A. Kawamura, K. Nishiyama, N. Koreeda, T. Nakashima, and O. Morita. 2004. "Neural networks for rainfall forecasting by atmospheric downscaling." *J. Hydrol. Eng.* 9 (1): 1–12. https://doi.org/10.1061/(ASCE)1084-0699(2004)9:1(1).

Prudhomme, C., N. Reynard, and S. Crooks. 2002. "Downscaling of global climate models for flood frequency analysis: Where are

we now?" *Hydrol. Process.* 16 (6): 1137–1150. https://doi.org/10.1002/hyp.1054.
Rashid, M. M., S. Beecham, and R. K. Chowdhury. 2016. "Statistical downscaling of rainfall: A non-stationary and multi-resolution approach." *Theor. Appl. Climatol.* 124 (3–4): 919–933. https://doi.org/10.1007/s00704-015-1465-3.
Sachindra, D. A., A. W. M. Ng, S. Muthukumaran, and B. J. C. Perera. 2016. "Impact of climate change on urban heat island effect and extreme temperatures: A case-study." *Q. J. R. Meteorolog. Soc.* 142 (694): 172–186. https://doi.org/10.1002/qj.2642.
Srinivas, V. V., B. Basu, N. Nagesh Kumar, and S. K. Jain. 2014. "Multi-site downscaling of maximum and minimum temperature using support vector machine." *Int. J. Climatol.* 34 (5): 1538–1560. https://doi.org/10.1002/joc.3782.
Taylor, K. E. 2001. "Summarizing multiple aspects of model performance in a single diagram." *J. Geophys. Res.: Atmos.* 106 (D7): 7183–7192. https://doi.org/10.1029/2000JD900719.
Tipping, M. E. 2001. "Sparse Bayesian learning and the relevance vector machine." *J. Mach. Learn. Res.* 1 (Jun): 211–244.
Vapnik, V. 1995. *The nature of statistical learning theory.* New York: Springer.
Vrac, M., P. Marbaix, D. Peillard, and P. Naveau. 2007a. "Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe." *Clim. Past* 3 (4): 669–682. https://doi.org/10.5194/cp-3-669-2007.
Vrac, M., and P. Naveau. 2007. "Stochastic downscaling of precipitation: From dry events to heavy rainfalls." *Water Resour. Res.* 43: W07402. https://doi.org/10.1029/2006WR005308.
Vrac, M., M. Stein, and K. Hayhoe. 2007b. "Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing." *Clim. Res.* 34 (3): 169–184. https://doi.org/10.3354/cr00696.
Wetterhall, F., A. BÃrdossy, D. Chen, S. Halldin, and C.-Y. Xu. 2006. "Daily precipitation-downscaling techniques in three Chinese regions." *Water Resour. Res.* 42 (11): W11423. https://doi.org/10.1029/2005WR004573.
Wilby, R. L., L. E. Hay, W. J. Gutowski, Jr., R. W. Arritt, E. S. Takle, Z. Pan, G. H. Leavesle, and M. P. Clark. 2000. "Hydrological responses to dynamically and statistically downscaled climate model output." *Geophys. Res. Lett.* 27 (8): 1199–1202. https://doi.org/10.1029/1999GL006078.
Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks. 1998. "Statistical downscaling of general circulation model output: A comparison of methods." *Water Resour. Res.* 34 (11): 2995–3008. https://doi.org/10.1029/98WR02577.
Wilks, D. S., and R. L. Wilby. 1999. "The weather generation game: A review of stochastic weather models." *Prog. Phys. Geogr.* 23 (3): 329–357. https://doi.org/10.1177/030913339902300302.
Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier. 2004. "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs." *Clim. Change* 62 (1–3): 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.