# Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam

T. Trinh[1,2,3] · N. Do[4] · V. T. Nguyen[5] · K. Carr[1]

## Abstract

Modeling of large rainfall events plays an important role in water resources and floodplain management. Rainfall is resulted from complex interactions between climate factors (air moisture, temperature, wind speed, etc.) and land surface (topography, soil, land cover, etc.). Therefore, deriving accurate areal rainfall is not only relied on atmospheric boundary conditions, but also on the reliability and availability of soils, topography, and vegetation data. Consequently, uncertainties in both atmospheric and land surface conditions contributes to rainfall model errors. In this study, a blended technique combining dynamical and statistical downscaling has been explored. The proposed downscaling approach uses input provided from three different global reanalysis data sets including ERA-Interim, ERA20C, and CFSR. These reanalysis atmospheric data are hybridly downscaled by means of the Weather Research and Forecasting (WRF) model, which is followed by the application of an artificial neural network (ANN) model to further downscale the WRF output to a finer resolution over the studied region. The proposed technique has been applied to the third largest river basin in Vietnam, the Sai Gon–Dong Nai Rivers Basin; and the calibration and validation show the simulation results agreed well with observation data. Results of this study suggest that the proposed approach can improve the accuracy of simulated data, as it merges model simulations with observations over the modeled region. Another highlight of this approach is inexpensive computational demand on both computation times and output storage.

✉ V. T. Nguyen
   vnguyen@snu.ac.kr

   T. Trinh
   tqtrinh@ucdavis.edu

   N. Do
   namdh@vawr.org.vn

   K. Carr
   kjcarr@ucdavis.edu

1  Hydrologic Research Laboratory, Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA

2  Institute for Computational Science and Technology, SBI Building, Quang Trung Software City, Ho Chi Minh City 700000, Vietnam

3  Institute of Ecology and Works Protection, Vietnam Academy for Water Resources, Hanoi 116830, Vietnam

4  Vietnam Academy for Water Resources, Hanoi 116830, Vietnam

5  Department of Civil and Environmental Engineering, Seoul National University, Seoul 151-742, Republic of Korea

## 1 Introduction

The modeling of large rainfall events is a fundamental and challenging topic in water resources and floodplain management. Rainfall results from complex interactions between climate factors, such as air moisture, temperature and wind speed; and the land surface, such as topography, soil, and land cover conditions. Therefore, deriving an accurate areal rainfall is not only relied on the atmospheric boundary conditions but also on the reliability and availability of soils, topography, and vegetation data. Consequently, the uncertainties of both atmospheric and land surface conditions contribute to rainfall model errors (Gebregiorgis and Hossain 2012; Shepherd 2014; Reichler and Kim 2008). Wherein, Gebregiorgis and Hossain (2012) explored uncertainties of three satellite rainfall products relating to unreliable topography and climate conditions. Shepherd (2014) found sources of uncertainty coming from climate boundary conditions and resulting in atmospheric model errors. Reichler and Kim

(2008) identified errors and uncertainties associated with the different reanalysis data sets by comparing them with a wide range of observations. These errors suggest the need for a holistic approach considering both high-resolution topography and land surface distribution to aid in the generation of realistic rainfall information.

Recently, there have been attempts to model rainfall events by means of global atmospheric models (GCMs) (Krishnamurti et al. 1997; Compo et al. 2006; Lledó et al. 2013; Fuka et al. 2014). Such GCMs consider various aspects of climate and the effect of the land surface on receiving surface rainfall. However, their spatial resolution, typically 100 km, is too coarse for use in analyzing water resources at the watershed or regional scale. One recommendation is the use of downscaling technologies to refine coarse grid resolution data to desired finer spatial grid resolutions. Commonly, there are two different approaches, statistical or stochastic downscaling (SD) and dynamical downscaling (DD). The SD refers to empirical relationships between large-scale modeled atmospheric variables and local-scale meteorological variables. The empirical relationships and inexpensive computational demands enable SD to be popular and widely used in many regional atmospheric studies (Burlando and Rosso 2002; Fowler et al. 2007; Goyal and Ojha 2011; Hashmi et al. 2011a, b, 2013; Pilling and Jones 2002; Raje and Mujumdar 2011; Wilby and Wigley 1997; Yang et al. 2011, 2012). Because SD methods rely on the assumption of an unchanged statistical relationship, they require long historical climate observation data for validation, which is not always available for every region. DD is an alternative to SD for empirical climate downscaling that can overcome the drawbacks of SD methods. DD works by employing a regional climate model (RCM), which is based on the same principles as a GCM, but has a higher resolution. A RCM uses large-scale GCMs' outputs for initial and lateral boundary conditions to generate much finer meteorological variables with incorporated high-resolution topography and land-sea distribution. This allows dynamic interaction between the atmosphere and land surface, thereby accounting for the impact of heterogeneity in the topography, vegetation, and soil on the local climate. DD is known as the most suitable technology for modeling climate information with complex topography at regional scales (Kavvas et al. 2013; Kjellström et al. 2016; Jang and Kavvas 2015; Jang et al. 2017). In spite of recent developments in DD making them easily accessible, this method still requires expensive computational demand on both long computation times and large output storage.

In order to overcome the limitations in both DD and SD approaches, a blended technique combining dynamical and statistical downscaling has been explored. Recently, Liu and Fan (2014), Tran and Taniguchi (2018), and Walton et al. (2015) have applied a hybrid dynamical-statistical downscaling approach by incorporating a regional climate model (RCM) with a statistical downscaling technique to some regions in China, Vietnam, and the Western United State. However, before coupling a RCM with a statistical model, both models need to be calibrated and validated in order to verify their capability and reliability for further downscaling applications. Hence, ignoring the calibration and validation of the RCM, Liu and Fan (2014) and Tran and Taniguchi (2018) may obtain unreliable downscaled data for the estimation of atmospheric variables; particularly, in mountainous regions. Furthermore, the temporal downscaling data obtained from Liu and Fan (2014), Tran and Taniguchi (2018) and Walton et al. (2015) is mainly focused on monthly scale, which are inappropriate for the analysis of floods and large rainfall events.

In this context, this study applied a regional climate model (RCM) coupled with machine learning algorithms to model and reconstruct rainfall data. This new technique, called hybrid downscaling (HD), first uses large-scale atmospheric conditions as determined by a GCM for its lateral boundary conditions before being downscaled by a RCM model, then applies ANN model to further downscale from selected RCM outputs to a finer spatial resolution. The HD also includes the influences of terrain factors and physical interactions between atmosphere and land surface conditions. Another highlight of this technology is that it improves the accuracy of simulated data as it merges model simulations with observations over the modeled region. The proposed downscaling technique uses input provided from three different global reanalysis datasets; ECMWF—Atmospheric Reanalysis coarse climate data of the twentieth century (ERA-20C, https://rda.ucar.edu/datasets/ds626.0) (Poli et al. 2013, 2016), ECMWF—Reanalysis Interim (ERA-Interim, https://rda.ucar.edu/datasets/ds627.0) (Berrisford et al. 2009; Dee et al. 2011), and Climate Forecast System Reanalysis (CFSR, https://rda.ucar.edu/datasets/ds093.0) (Saha et al. 2010; Wang et al. 2011). These three datasets provide three-dimensional data and uniformly cover the globe at a spatial resolution of 1.25° (ERA20C), 0.75° (ERA-Interim), and 0.5° (CFSR). These coarse scale atmospheric data are hybrid downscaled by means of the Weather Research and Forecasting model (WRF, Skamarock et al. 2005), then followed by the application of an artificial neural network (ANN) model to further downscale from the WRF output to a finer resolution over the studied watershed. First, the WRF and ANN models are calibrated and validated against existing ground observation data, then hybrid method is evaluated through time series and spatial analyses. The Sai Gon–Dong Nai Rivers Basin is selected as a case study for the application of the hybrid technique. Due to its important location and complicated physical processes causing severe rainfall in this area, it is necessary to apply advanced technologies to

investigate severe rainfall processes, and model realistic historical rainfall events for this region.

## 2 Description of the study region

The selected watershed, the Sai Gon–Dong Nai (SG–DN) Rivers, ranks third-largest in country after the Mekong and Red River water systems, but it is the largest inland river in Vietnam. The SG–DN Rivers have become an important source of hydropower, with many hydropower plants and large amounts of water resources used for all southern provinces of Vietnam. Natural impacts from meteorological factors have caused many difficulties for socio-economic development activities in the basin. The SG–DN Rivers Basin has a complex terrain system including mountainous and delta regions with tropical heavy rainfall experienced from summer monsoon (SMS) and tropical cyclone (TC) systems (Nguyen-Thi et al. 2012; Yokoi and Matsumoto 2008).

The SG–DN Rivers Basin shown in Fig. 1 covers the provinces of Lam Dong, Binh Phuoc, Binh Duong, Dong Nai, Dak Nong, Long An, Tay Ninh, and Ho Chi Minh City, and parts of Ninh Thuan, Binh Thuan, and Ba Ria-Vung Tau with a total catchment area of about 44,500 km². SG–DN Rivers Basin includes the two main river system including Sai Gon and Dong Nai Rivers. This area is a complex terrain region including mountainous and delta regions with elevations from 2 to 2291 m. Along with an important source of hydropower, the SG–DN Rivers Basin also include a number of important industrial zones. The region's atmospheric condition falls in a tropical monsoon climate experiencing a wet summer from late May through early November with an average annual rainfall of about 1800 mm, and humidity of 78–82%. The land use condition of the watershed is various land types including agricultural, forested, and urban areas.

## 3 Methodology and implementation

This study introduces a blended technique to model rainfall events by coupling physically based numerical atmospheric and machine learning models. The required atmospheric data used to set up the initial and boundary conditions in WRF simulations over SG–DN basin are taken from the three reanalysis datasets, including ERA-20C, ERA-Interim, and CFSR. These datasets were selected because they provide three-dimensional data at 6-h time increments for the required atmospheric and surface variables. They are also long enough to be reliable in a statistical sense and consistently cover the entire globe uniformly (Rossi et al. 2007). The WRF model is utilized as the physically based numerical atmospheric
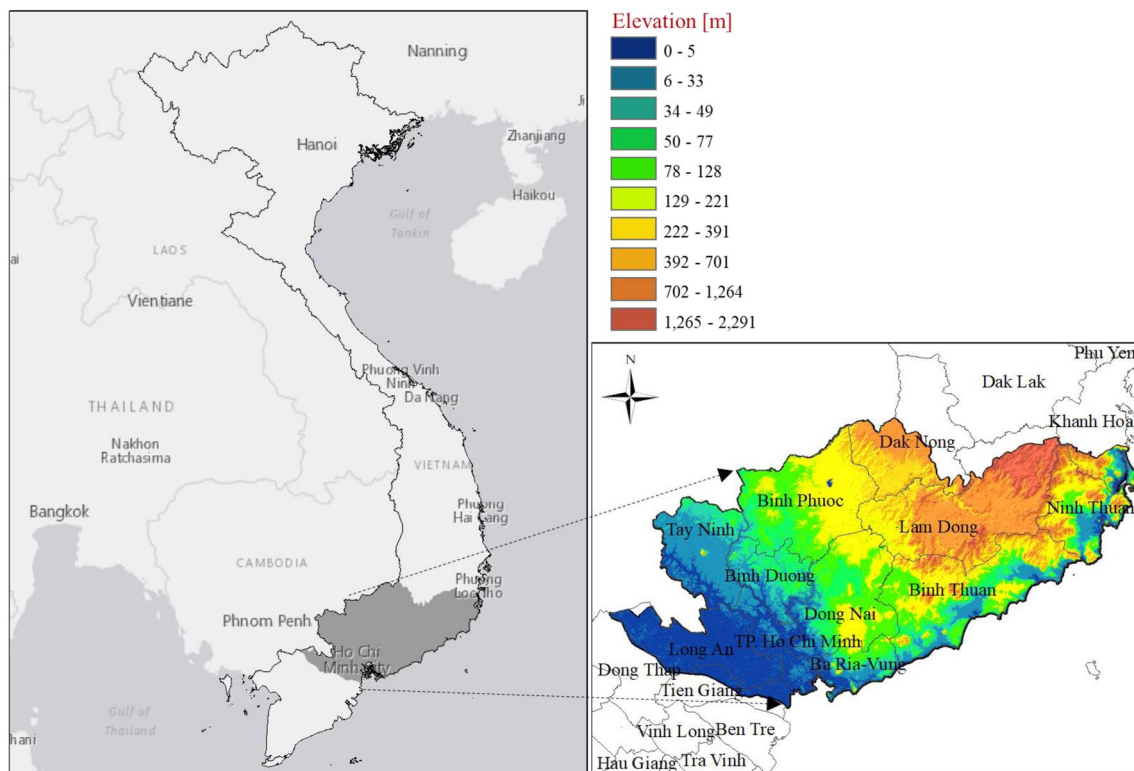


**Fig. 1** Plan view of the Sai Gon–Dong Nai Rivers Basin in Vietnam

model, while the ANN model is selected as the machine learning model, as shown in Fig. 2. There are five main steps in developing this hybrid rainfall model:

1. Implementation of the physically based numerical atmospheric model, WRF, over the target watershed for the three different reanalysis datasets.
2. Calibration and validation of the WRF model over the target watershed for the three different reanalysis datasets.
3. Implementation of the ANN model with its input provided from WRF's outputs.
4. Training and validation of the ANN model over the target watershed for the three different reanalysis datasets.
5. Provision of hybrid downscaling model for the target watershed.

In-depth description of each steps is presented in the following sections.

## 3.1 Implementation of the physically based numerical atmospheric model

The WRF model was employed for dynamical downscaling with inputs from the three reanalysis datasets. The WRF model is able to simulate vertical and horizontal air motions with multiple physics options for moisture dynamics, microphysics processes, cumulus cloud parameterizations, planetary boundary layer (PBL) schemes, radiation schemes, and surface schemes. A number of studies successfully applied the WRF model for precipitation analysis on regions in Vietnam (Ho et al. 2019, 2020; Cuong and Toan, 2019; Raghavan et al. 2016; Minh et al. 2018) with encouraging performance when compared to the recent rainfall observation data. Thus, WRF is selected herein, although other numerical models can be implemented for regional atmospheric modeling. In this study, a series of three nested domains for WRF simulations are implemented, as shown in Fig. 3. The largest domain (D1) covers the southern half of Vietnam and parts of Thailand, Laos, Cambodia, and Malaysia, having a spatial resolution of 81 km ($21 \times 18$ horizontal grid points). D2 is the second largest domain with a resolution of 27 km ($27 \times 24$ horizontal grid points), and D3 is the innermost and smallest domain with spatial resolution of 9 km ($48 \times 33$
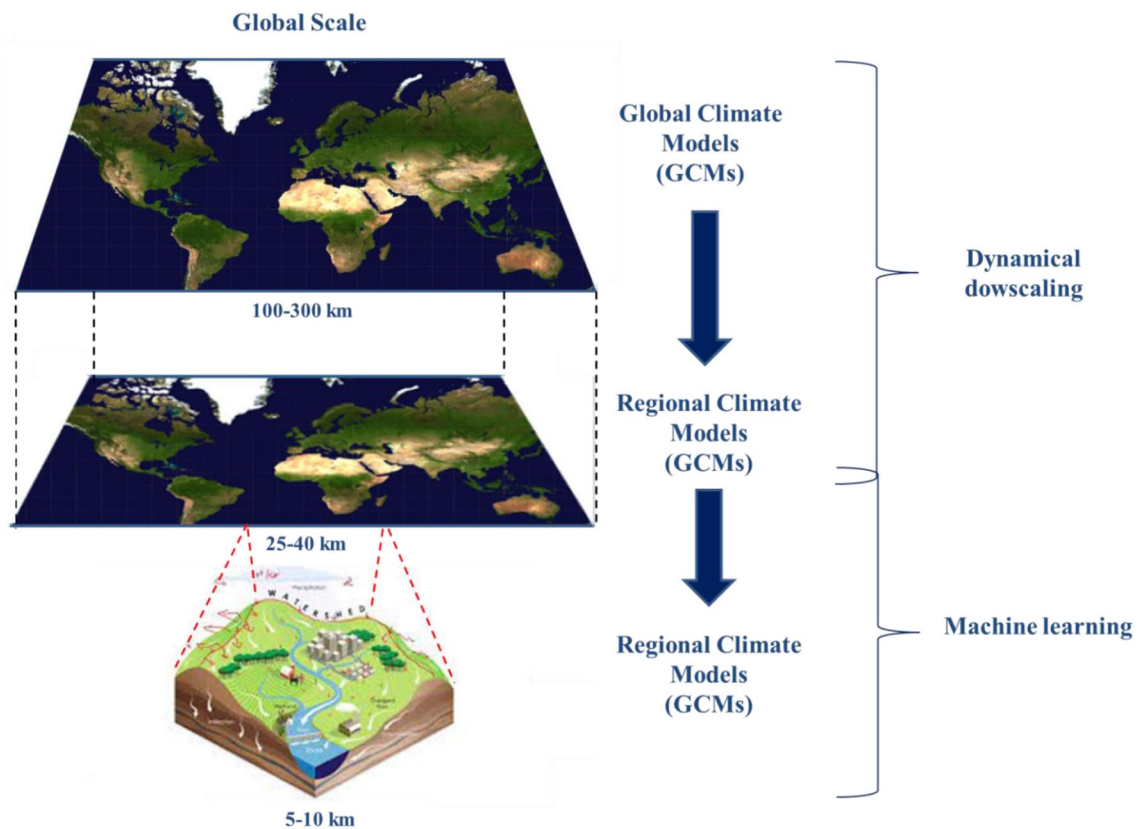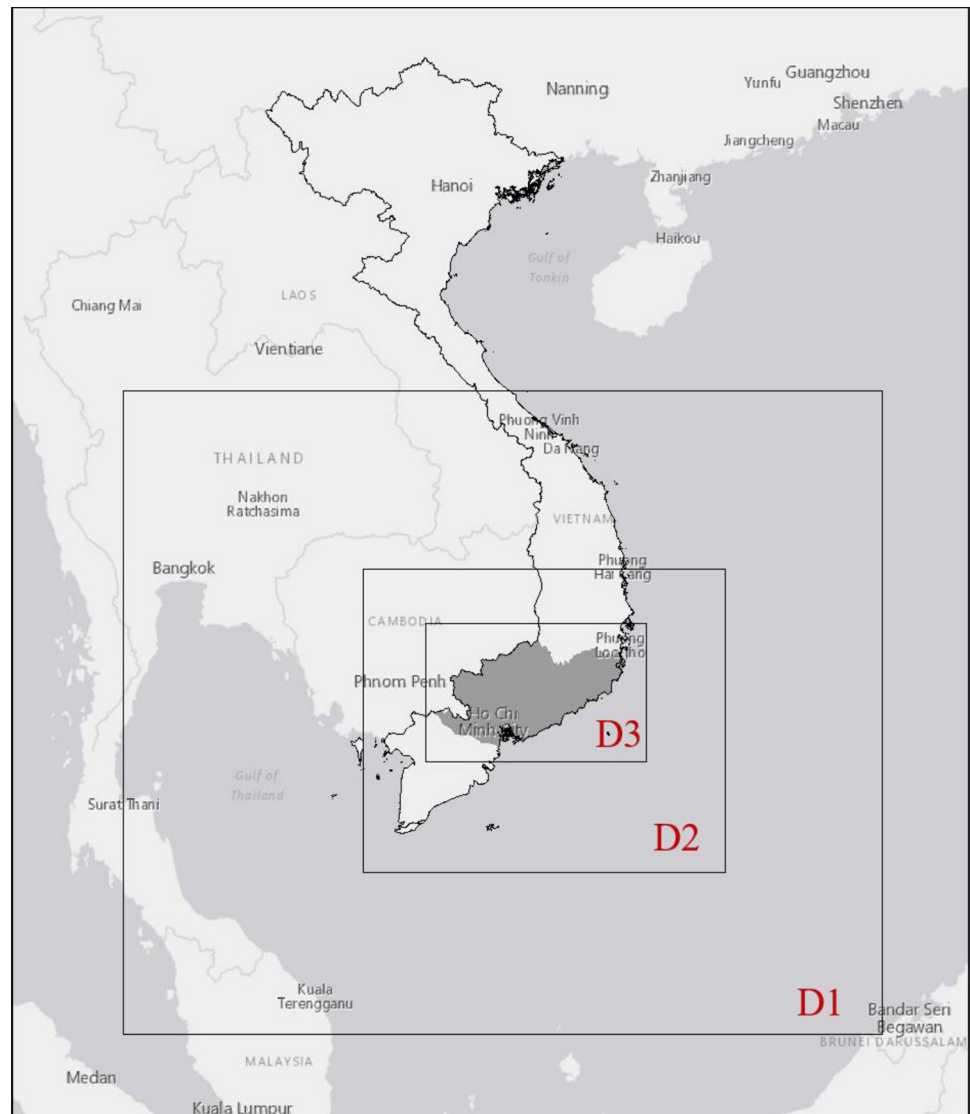


**Fig. 2** Methodology for hybrid modeling rainfall data

**Fig. 3** The three nested WRF domains for dynamical downscaling over DN-SG River Basin



horizontal grid points). It is noted that WRF is implemented based on all 3 domains only for ERA20C data, while the ERA-Interim and CFSR were used only on D2 and D3.

## 3.2 Calibration and validation of WRF model over the target watershed for the three different reanalysis datasets

After successful implementation of WRF for SG–DN, the model was calibrated and validated against the observation rainfall data. Recently, the Vietnam Gridded Precipitation (VnGP) dataset was developed, and has been widely used for reliable observation (Nguyen-Xuan et al. 2016). The VnGP is daily gridded rainfall dataset that was interpolated by means of the Sphere-map interpolation technique from 481 rain gauges. This dataset has the resolution of 0.1°, and covers the whole Vietnam (Nguyen-Xuan et al. 2016). The validation of VnGP was carried out by comparing with

gauge observations through correlations, mean absolute errors, root mean square errors, and spatial distribution. The validation results show that the VnGP is matched well with rainfall observation rather than different interpolation techniques. VnGP is currently available at the Data Integration and Analysis System (DIAS) (https://diasjp.net/en). The spatially-distributed daily rainfall data of VnGP are available from Jan 1980 to December 2010. This data was compared with the model's precipitation simulations over SG–DN. First, the WRF model's configurations are selected based on comparisons between downscaled rainfall data and the VnGP dataset. Table 1 shows 12 combinations of parameterization schemes based on previous studies in Vietnam (Ho et al. 2020; Ho et al. 2019; Cuong et al. 2019b; Raghavan et al. 2016; Minh et al. 2018; Trinh et al. 2020). The best parameterization scheme was selected based on the correlation coefficient for simulated daily basin average precipitation and VnGP data between 1 January, 1994 and 31

**Table 1** Twelve combinations of physics parameterizations for WRF configuration

| Simulation no | Microphysics | Cumulus parameterization | Planetary boundary layer | Radiation physics |
|---|---|---|---|---|
| 1 | WSM3 | New SAS | BouLac | New Goddard |
| 2 | Eta (Ferrier) | New SAS | BouLac | New Goddard |
| 3 | Goddard | New SAS | BouLac | New Goddard |
| 4 | Milbrandt 2-mom | New SAS | BouLac | New Goddard |
| 5 | CAM 5.1 | New SAS | BouLac | New Goddard |
| 6 | SBU-YLin | New SAS | BouLac | New Goddard |
| 7 | WSM3 | Kain-Fritsch | BouLac | New Goddard |
| 8 | WSM3 | Grell-Freitas | BouLac | New Goddard |
| 9 | WSM3 | Grell-3 | BouLac | New Goddard |
| 10 | WSM3 | New SAS | MYNN2 | New Goddard |
| 11 | WSM3 | New SAS | YSU | New Goddard |
| 12 | WSM3 | New SAS | UW | New Goddard |

**Table 2** Correlation coefficient for basin averaged daily precipitation between WRF simulations and VnGP using three different reanalysis datasets during January 1, 1994 to December 31, 1995 (D3 is primarily used in these comparisons)

| Simulation no | ERA-Interim | ERA-20C | CFSR |
|---|---|---|---|
| 1 | **0.72** | **0.65** | 0.61 |
| 2 | 0.64 | 0.57 | 0.56 |
| 3 | 0.61 | 0.52 | 0.53 |
| 4 | 0.60 | 0.51 | 0.55 |
| 5 | 0.62 | 0.502 | 0.51 |
| 6 | 0.69 | 0.61 | **0.69** |
| 7 | 0.59 | 0.51 | 0.51 |
| 8 | 0.58 | 0.50 | 0.62 |
| 9 | 0.63 | 0.55 | 0.65 |
| 10 | 0.63 | 0.59 | 0.63 |
| 11 | 0.68 | 0.62 | 0.62 |
| 12 | 0.65 | 0.60 | 0.61 |

December, 1995. Water years 1994–1995 were selected for comparison due to their inclusion of historical extreme flood events. Note that D3 is primarily used in these comparisons.

The results shown in Table 2 indicate that simulation No. 1, which uses WSM3 Hong et al. (2004, MWR) as the microphysics schemes and New SAS (Han and Pan 2011) for cumulus parameterization, shows the highest correlation coefficient for the ERA-Interim and ERA-20C reanalysis datasets at the target watershed. Simulation No. 6 uses SBU-YLin, Lin and Colle (2011) as the microphysics schemes and New SAS Han and Pan (2011) for cumulus parameterization, and shows the highest correlation coefficient for the CFSR reanalysis dataset. The selected WRF's parameterization options for each reanalysis dataset are shown in Tables 3 and 4.

The selected configurations for ERA20C, ERA-Interim, and CFSR are applied to the WRF model for the validation period. Figures 4a, b, 5a, b, and 6a, b show the time series

**Table 3** WRF model configuration for ERA20C and ERA-Interim

| WRF model configuration | Selected option |
|---|---|
| Microphysics processes | WSM3 (Hong et al. 2004, MWR) |
| Cumulus parameterization | New SAS (Han and Pan 2011, Wea. Forecasting) |
| Planetary boundary layer scheme | BouLac scheme (Bougeault and Lacarrere 1989) |
| Radiation scheme | New Goddard scheme (Chou and Suarez 1999) |
| Surface scheme | RUC Land Surface Model (Benjamin et al. 2004) |

**Table 4** WRF model configuration for CFSR

| WRF model configuration | Selected option |
|---|---|
| Microphysics processes | SBU-YLin (Lin and Colle 2011, MWR) |
| Cumulus parameterization | New SAS (Han and Pan 2011, Wea. Forecasting) |
| Planetary boundary layer scheme | BouLac scheme (Bougeault and Lacarrere 1989) |
| Radiation scheme | New Goddard scheme (Chou and Suarez 1999) |
| Surface scheme | RUC Land Surface Model (Benjamin et al. 2004) |
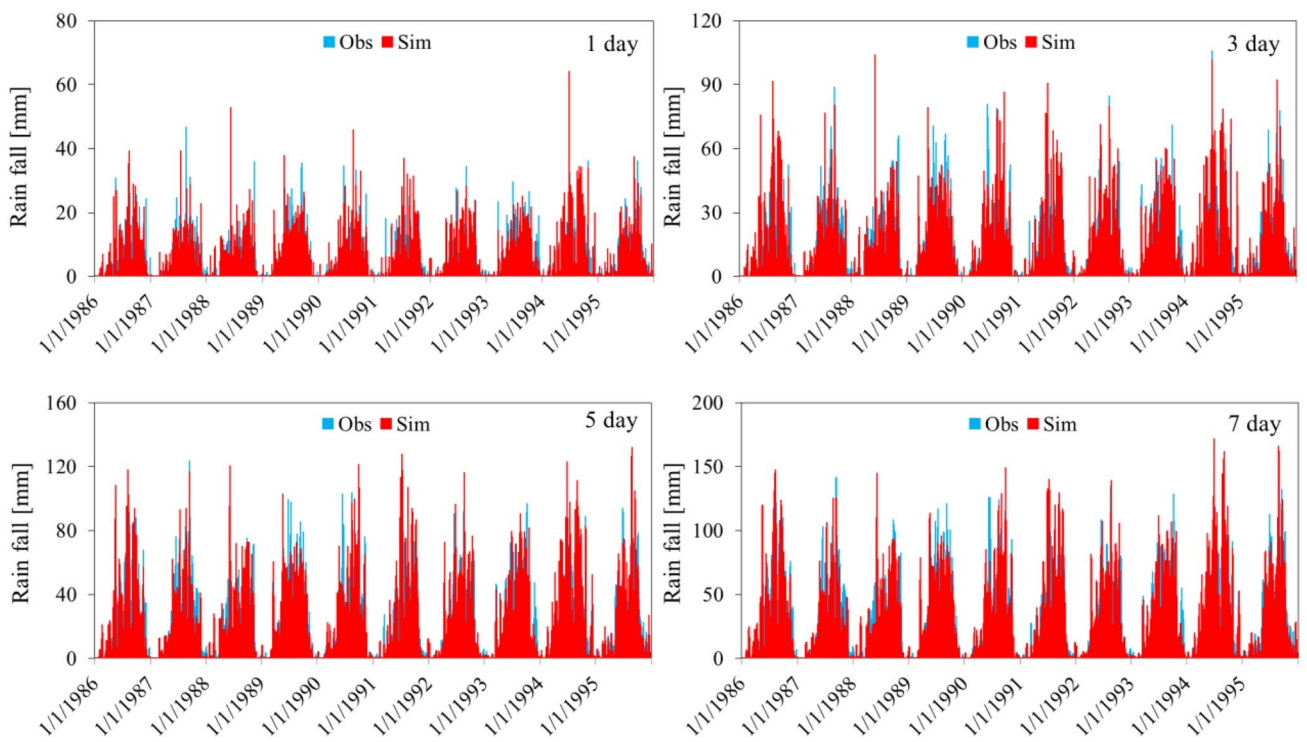
(a)
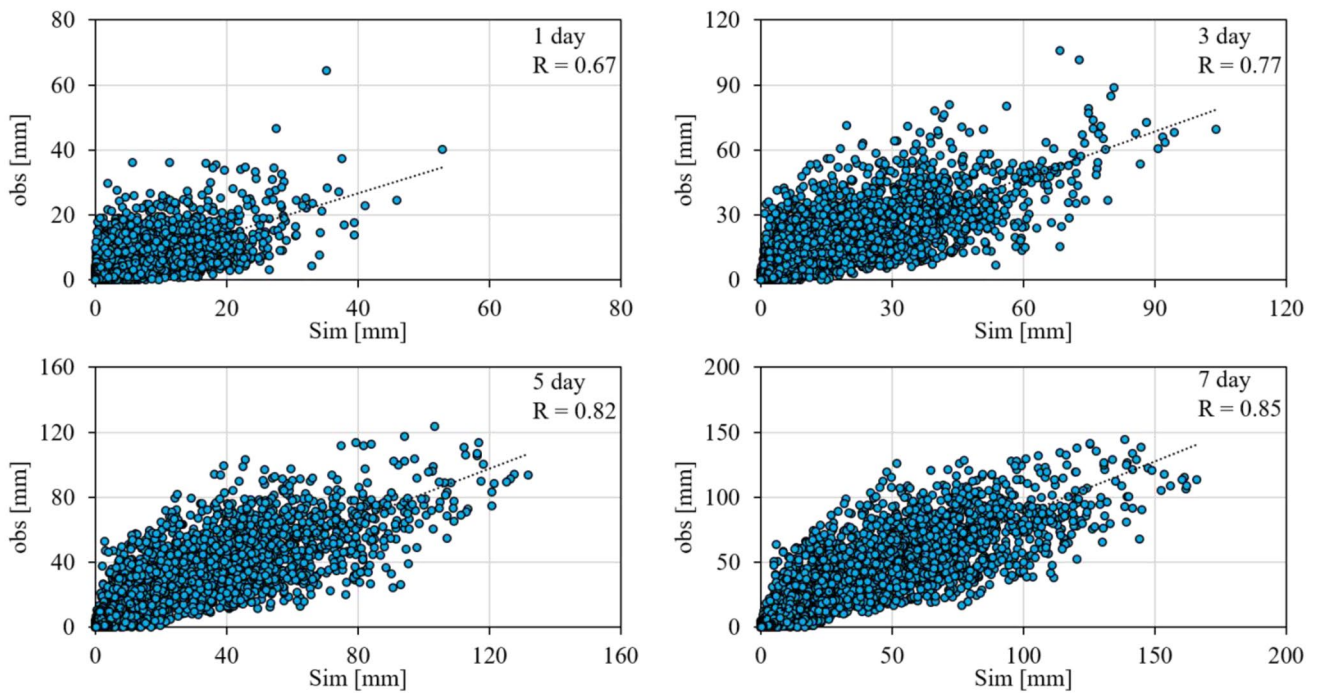


(b)

**Fig. 4** **a** Time series of WRF simulations using ERA-Interim at D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN. **b** Correlation coefficients between the WRF simulations using ERA-Interim in D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

**(a)**



**(b)**

**Fig. 5** **a** Time series of WRF simulations using ERA-20C in D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN. **b** Correlation coefficients between the WRF simulations using ERA-20C in D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN
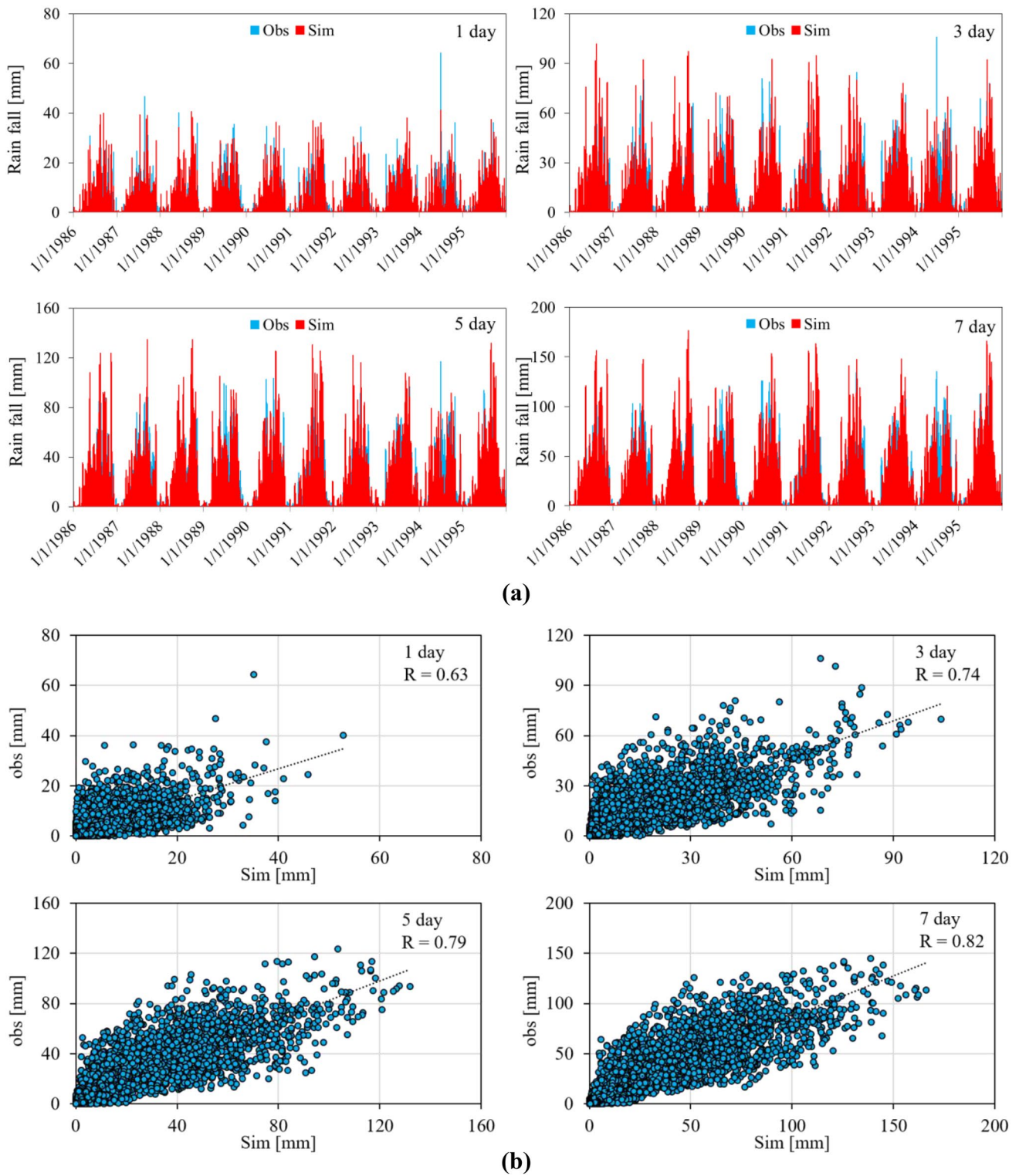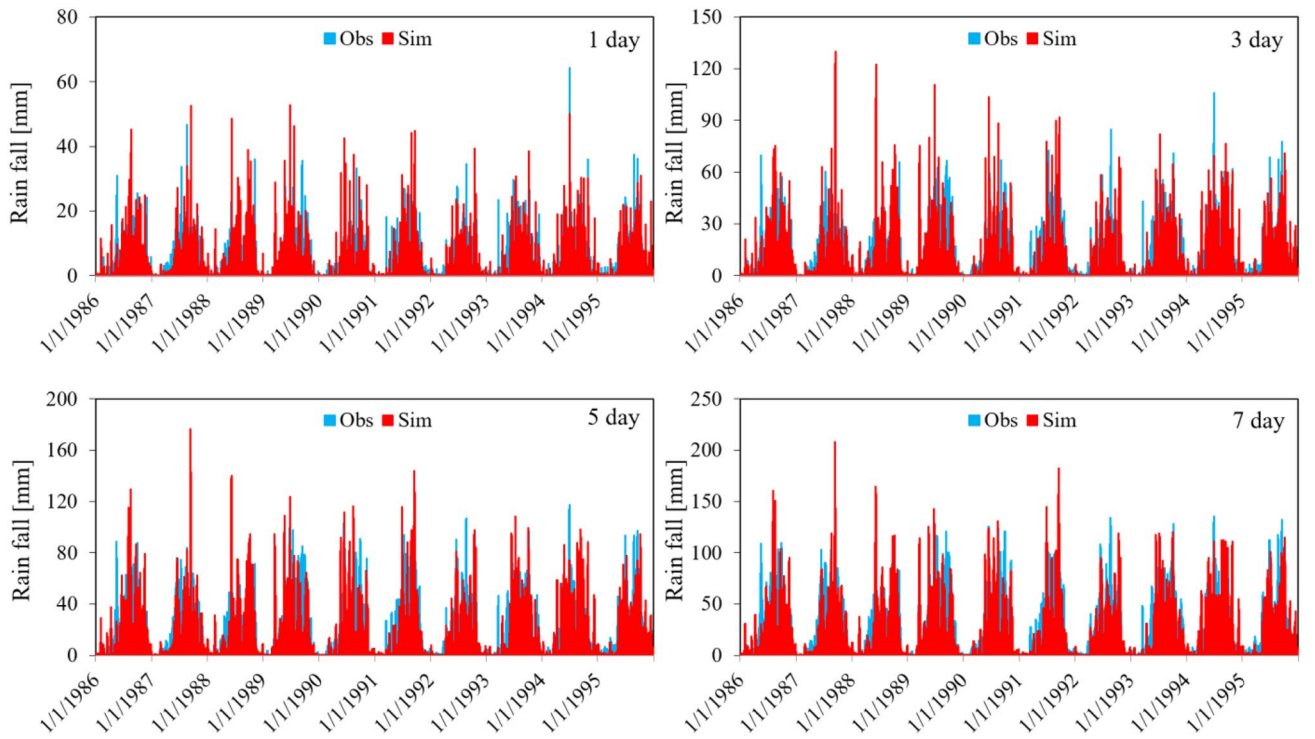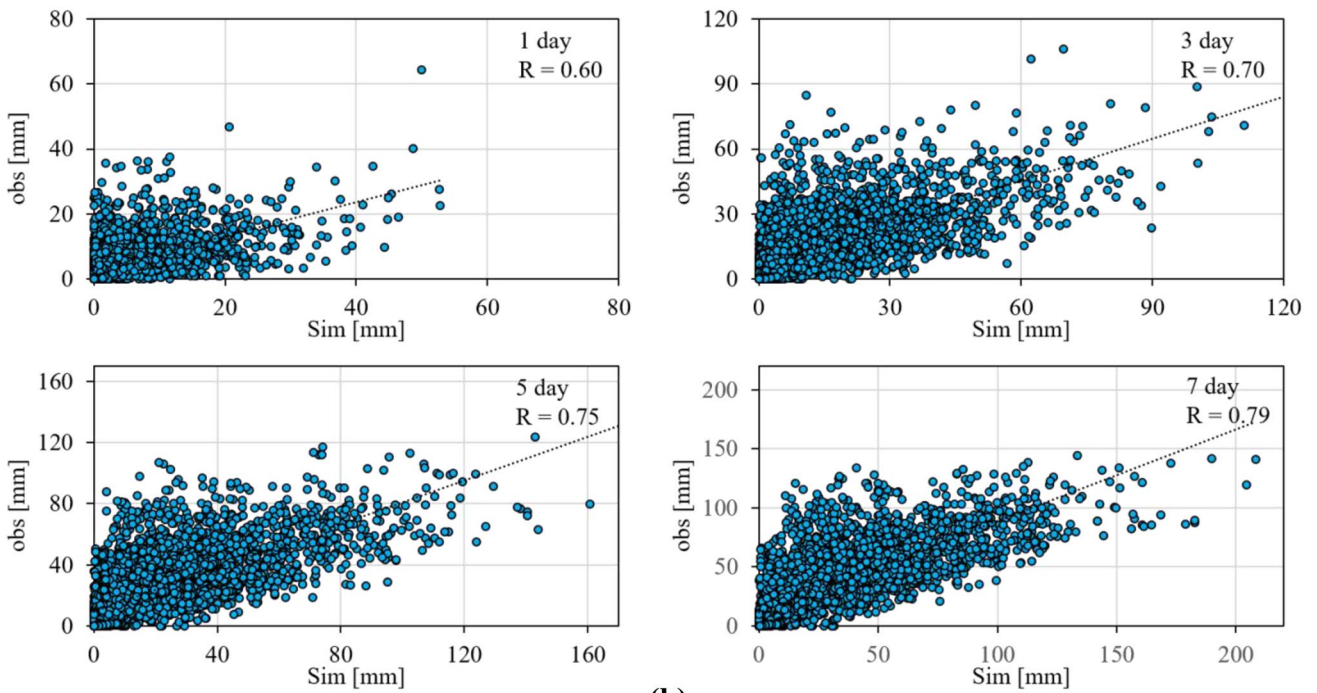
**(a)**



**(b)**

**Fig. 6** **a** Time series of WRF simulations using CFSR in D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN. **b** Correlation coefficients between the WRF simulations using CFSR in D3 and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

comparisons of ground observations and model-simulated 1-, 3-, 5-, and 7-day basin-averaged rainfall using the three different reanalysis datasets over SG–DN Basin during 1986–1995 (10-year period).

Visual comparison between the WRF simulation and corresponding observations in D3 show good agreement for the 1-, 3-, 5-, and 7-day basin-averaged comparisons for the three datasets. Although, the simulated peak discharges occasionally underestimate the observed data, the differences are not significant. Tables 5, 6 and 7 list the statistical test values of the WRF-simulated results versus the VnGP data with WRF's inputs from ERA-Interim, ERA-20C, and CFSR, respectively; whereby Goodness of fit of the modeled simulation to the corresponding observations is shown based on the mean, standard deviation, and correlation coefficient (R). A correlation coefficient (R)'s value of 1 corresponds to a perfect match of the modeled simulation to the observed data. A correlation coefficient larger than 0.6 can be considered acceptable for validation periods at daily time step.

## 3.3 Implementation of ANN architecture with back-propagation algorithm

In the early 1970s, the model output statistic (MOS) was an efficient technique widely used in atmospheric research to improve forecasted variables from the Numerical Weather Prediction (NWP) models (Glahn and Lowry 1972). MOS aims to establish an empirical relationship between the large-scale atmospheric predictors simulated from numerical models and local climatic variables (predictands). The formulation of empirical relationships is fundamentally based on either linear or nonlinear transfer functions. Empirical relationships based on linear transfer functions often apply multiple linear regressions or similar formulas. For example, the Statistical Downscaling Model (SDSM) is a hybrid statistical downscaling method incorporating the weather generator and the multiple linear regression techniques (Wilby et al. 2002). Nonlinear transfer functions are commonly applied through machine learning

methods that employ nonlinear transfer functions to connect the predictors and predictands. There are a number of nonlinear transfer functions, for instance, Genetic Programing (GP) that exhibits reasonable downscaling of daily extreme temperatures (Coulibaly 2004); Gene Expression Programming (Hashmi et al. 2011a, b) that is a variant of GP; and artificial neural networks. In this research, one of the most popular and simplest artificial neural network architectures, which is the feed-forward multilayer perceptron using error back-propagation weight update rule (hereafter referred to as the ANN model), is employed for downscaling precipitation simulated by the WRF model. The ANN model is commonly used in statistical downscaling. It was reported that the ANN model is generally observed to have a better learning ability than other regression-based downscaling techniques (Schoof and Pryor 2001).

The selected ANN architecture is comprised of three layers (input layer, hidden layer, and output layer) that are interconnected by synapse weights (see Fig. 7). The number of nodes of the hidden layer was selected ranging from $(2n + 1)$ to $(2n0.5 + m)$, where n is the number of input nodes and m is number of output nodes (Fletcher and Goss 1993).

The training phase of the ANN model serves to adjust the weights to minimize the difference between the

**Table 6** Statistical test values of WRF-simulated results based on ERA-20C and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | ERA-20C | | | |
|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.63 | 13.91 | 23.18 | 32.46 |
| Mean by simulation (mm) | 5.27 | 15.81 | 26.35 | 36.89 |
| Standard deviation by observation (mm) | 5.88 | 15.29 | 23.62 | 31.69 |
| Standard deviation by simulation (mm) | 6.84 | 17.87 | 27.41 | 36.37 |
| Correlation coefficient | 0.63 | 0.74 | 0.79 | 0.82 |

**Table 5** Statistical test values of WRF-simulated results based on ERA-Interim and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | ERA-INTERIM | | | |
|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.63 | 13.91 | 23.18 | 32.46 |
| Mean by simulation (mm) | 4.60 | 13.80 | 23.01 | 32.20 |
| Standard deviation by observation (mm) | 5.88 | 15.29 | 23.62 | 31.69 |
| Standard deviation by simulation (mm) | 6.36 | 16.49 | 25.17 | 33.31 |
| Correlation coefficient | 0.67 | 0.77 | 0.82 | 0.85 |

**Table 7** Statistical test values of WRF-simulated results based on CFSR and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | CFSR | | | |
|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.63 | 13.91 | 23.18 | 32.46 |
| Mean by simulation (mm) | 3.85 | 11.51 | 19.95 | 28.79 |
| Standard deviation by observation (mm) | 5.88 | 15.29 | 23.62 | 31.69 |
| Standard deviation by simulation (mm) | 6.38 | 16.17 | 24.25 | 31.63 |
| Correlation coefficient | 0.60 | 0.70 | 0.75 | 0.79 |

**Fig. 7** The sketch of ANN model



Neural network
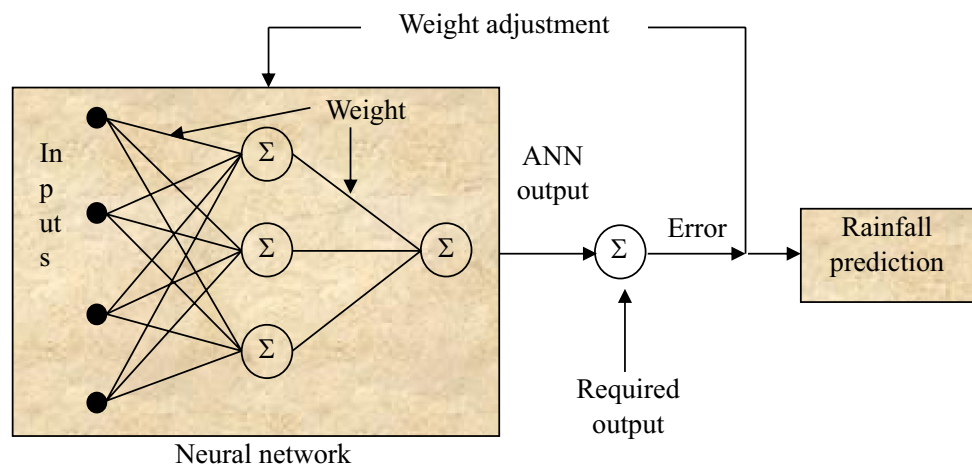
**Table 8** Candidate predictors for the input layer of the ANN model

| Variables[a] | Unit | Pressure layer (hPa) |
|---|---|---|
| Precipitation flux | (mm/day) | Surface |
| Meridional wind velocity | (m/s) | 700, 810, 910 |
| Zonal wind velocity | (m/s) | 700, 810, 910 |
| Vertical pressure velocity | (Pa/s) | 700, 810, 910 |
| Total | Variables | 10 |

[a]These are large-scale atmospheric variables simulated by the WRF model for domain D2

network outputs (predictands) and the expected outputs. For each node at a given layer, the outputs of n neurons in the previous layer provide the inputs to that node. These outputs are multiplied by the respective weights of connections between nodes and then the summation function adds together all these products to produce the value at that node. Approximation used for the weight change is given in Eq. (1) by the delta rule.

$$w_{new} = w_{old} - \eta \frac{\partial E^2}{\partial w} \tag{1}$$

where $\eta$ is the learning rate parameter, w is the weights, and $E^2$ is the squared error (Brierley 1998).

The selection of input variables is crucial for any data-driven methods. Based on that convention, the study area is characterized by the South-western monsoon regime, so that the low-level movement of the cloud mass is favorable for wet season over the region. Changes in wind fields (meridional and zonal wind velocities, vertical pressure velocity) at low altitudes (700, 810, and 910 hPa) together with the precipitation flux on the surface are candidate predictors for the input layer of the ANN model, as presented in Table 8 below. However, additional predictor screening "stepwise regression" was then applied to remove the less significant

predictor variables that may contribute to overfitting during ANN model learning processes.

## 3.4 Training and validation of ANN model over the target watershed for the three different reanalysis datasets

In this study, a gridded (0.1° daily precipitation dataset (VnGP) is used for the ANN model training and validation. It is ideal to divide the entire data duration into portions for training, validation, and testing. Most statistical downscaling exercises use two-thirds of the data duration for training and the remaining duration for calibration. Both training and calibration are examined at grid- and basin-average-scale for sensitivity assessment of spatial influences. A Fortran code is developed to train and calibrate the ANN model.

After successful implementation of the ANN model for the three downscaled datasets, the 10-year period from 1986 to 1995 of VnGP data is used for calibration and validation. Figure 8a–c show the calibration results for the period of (1986–1992) in comparing the basin-averaged daily precipitation dataset (VnGP) with the ANN simulations using the downscaled ERA-Interim, ERA-20C, and CFSR, respectively. Validation results for the period of (1993–1995) are exhibited in Fig. 9a–c. Generally, the calibration and validation showed a good agreement between the simulation results and observations. Statistical criteria supporting the agreement of the simulation results to VnGP data are shown in the Tables 9, 10 and 11. Statistical criteria, such as the correlation and Nash Sutcliffe efficiency coefficients, show that the simulation performance for the simulated precipitation is in the "satisfactory" range ($0.83 \leq R^2 \leq 0.9$ and $0.63 \leq NSE \leq 0.78$) based on the 7 day precipitation comparisons (Moriasi et al. 2015).

**(a)**



**(b)**

**Fig. 8** **a** Time series of ANN calibrations using ERA-Interim at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1992 over the SG–DN. **b** Time series of ANN calibrations using ERA-20C at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-aver-
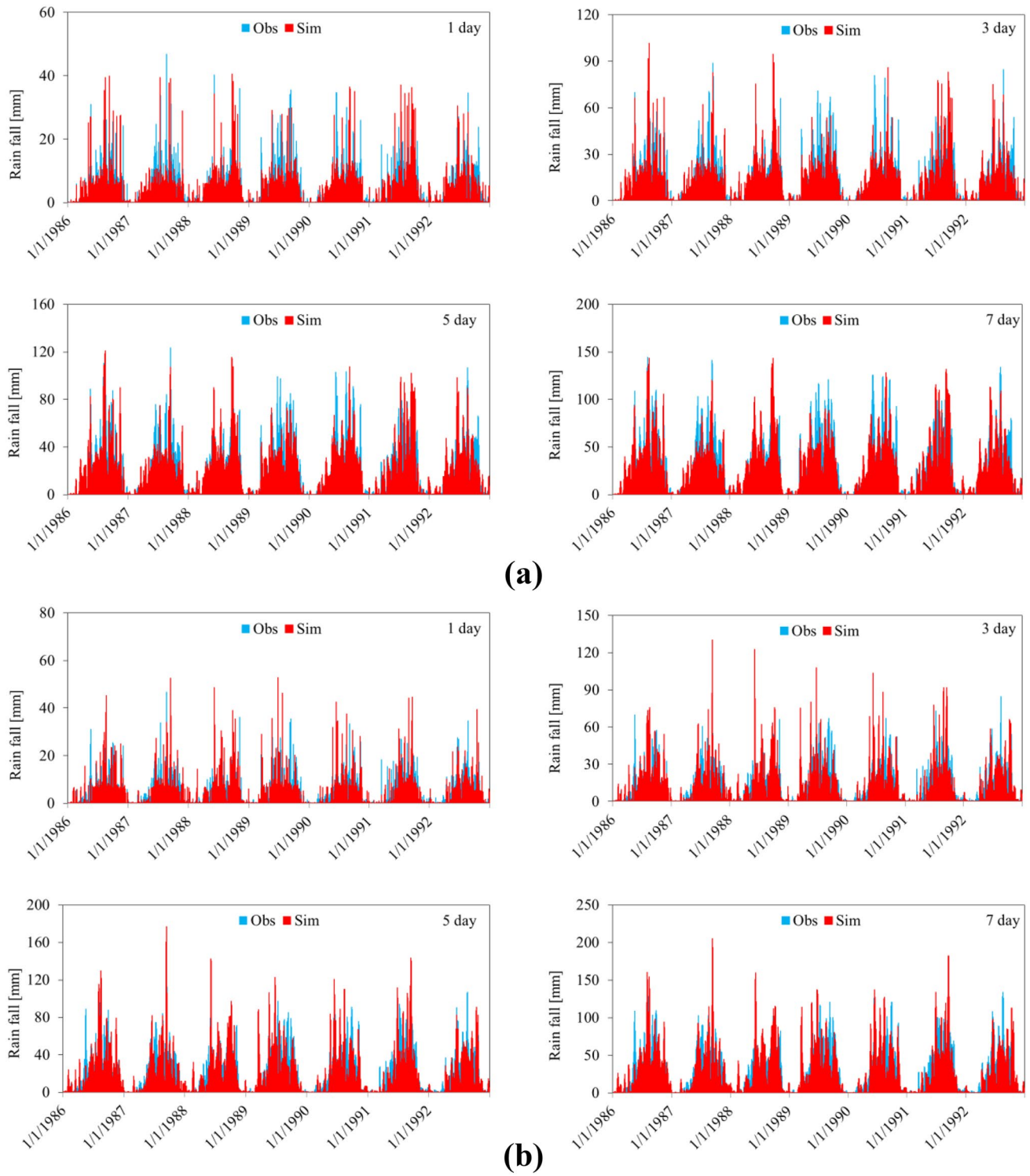
age precipitation during 1986–1992 over the SG–DN. **c** Time series of ANN calibrations using CFSR at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1992 over the SG–DN
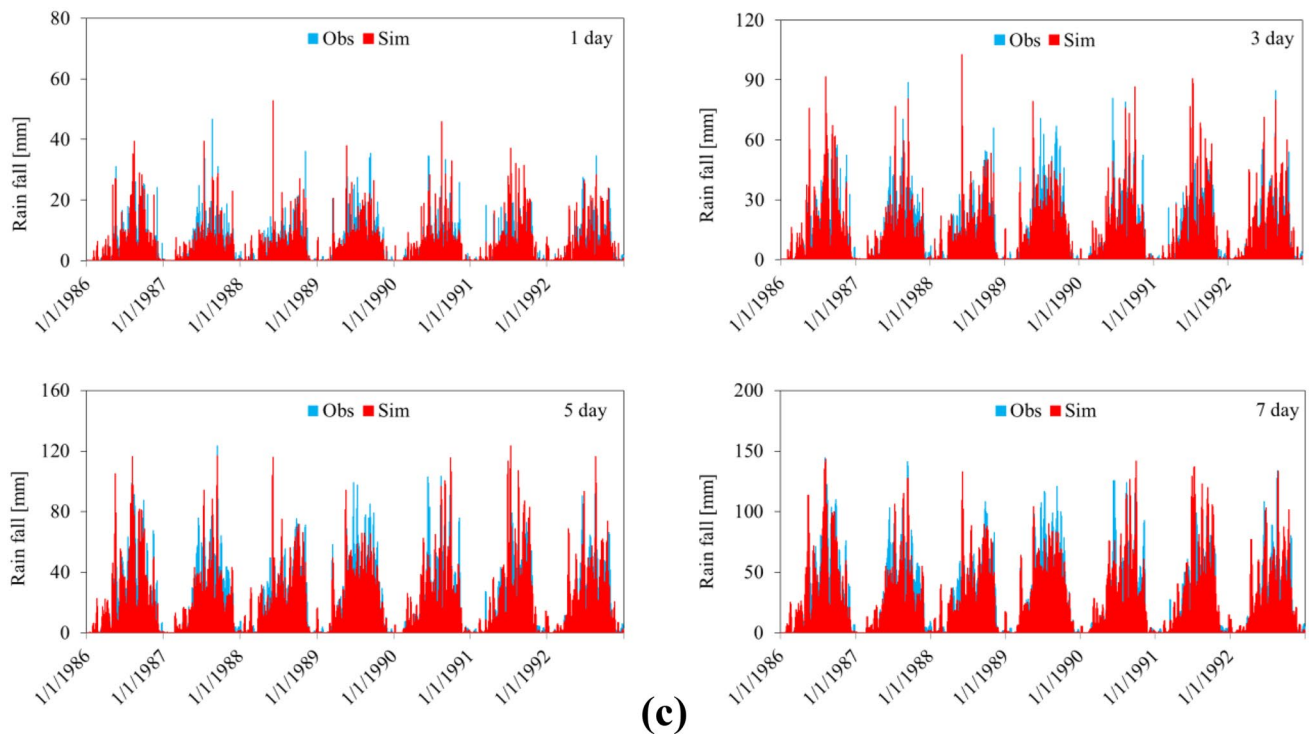
**Fig. 8** (continued)

# 4 Results and discussion

In comparison between three selected reanalysis datasets, the ERA-20C and ERA-Interim datasets are come from the same sources, the European Centre for Medium-Range Weather Forecasts (ECMWF); while the CFSR dataset is from National Center for Atmospheric Research (NCAR, USA). Regarding data resolution, the order ranging from coarse to fine among the three selected datasets are the ERA-20C (1.25°), then ERA-Interim (0.75°) and CFSR (0.5°). Since ERA-Interim has a better resolution than ERA-20C, so it is more reliable in connection to the surface measurements. In addition, the temperature and precipitation data of ERA-Interim are based on a reanalysis of precipitation fields generated with a meteorological model (Berrisford et al. 2009; Dee et al. 2011). On the other hand, ERA-Interim dataset is obtained by successive linearisations of the model and observation operator (Courtier et al. 1994; Veerse and Thepaut 1998). The ability of the observation operator to accurately model observations affects the quality of the analysis; errors or inaccuracies in the observation operator result in incorrect or suboptimal interpretation of the available data. While the temperature and precipitation data of CFSR (Saha et al. 2010) are based on meteorological model in combination with data from satellite-based observing systems and surface observation. The direct assimilation of observations represents one of the major improvements of the CFSR dataset. However, substantial biases exist when observations are compared to those simulated. These biases are complicated and relate to instrument calibration, data processing, and deficiencies in the radiative transfer model. Therefore, the combination with remote sensing data and surface observations may accumulate more errors in CFRS dataset than using ERA-Interim dataset. Eventually, the best calibration and validation results were obtained from the ERA-Interim dataset. Under the ERA-Interim dataset, the HD technique provided quite satisfactory correlation coefficients, ranging from 0.89 to 0.90, and Nash–Sutcliffe efficiencies, from 0.76 to 0.78 based on the 7-day precipitation (Table 9). The HD technique simulations using ERA-20C also gave good validation results with correlation coefficients ranging from 0.65 to 0.86. Model calibration and validation results under the ERA-Interim and ERA-20C were closer to the VnGP dataset than the CFSR results were.

In addition, the comparison between the simulation results obtained from the WRF model and HD technique with the VnGP dataset is carried out not only for temporal distribution, but also for spatial distribution.

Figure 10a–d presents the spatial distribution map of the largest difference in 1-, 3-, 5-, and 7-day precipitation between the WRF model results and the VnGP dataset during the period from 1980 to 2010. Figure 11a–d presents the spatial distribution map of the largest difference in 1-, 3-, 5-, and 7-day precipitation between the HD model and VnGP

**Fig. 9** **a** Time series of ANN validations using ERA-Interim at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1993–1995 over the SG–DN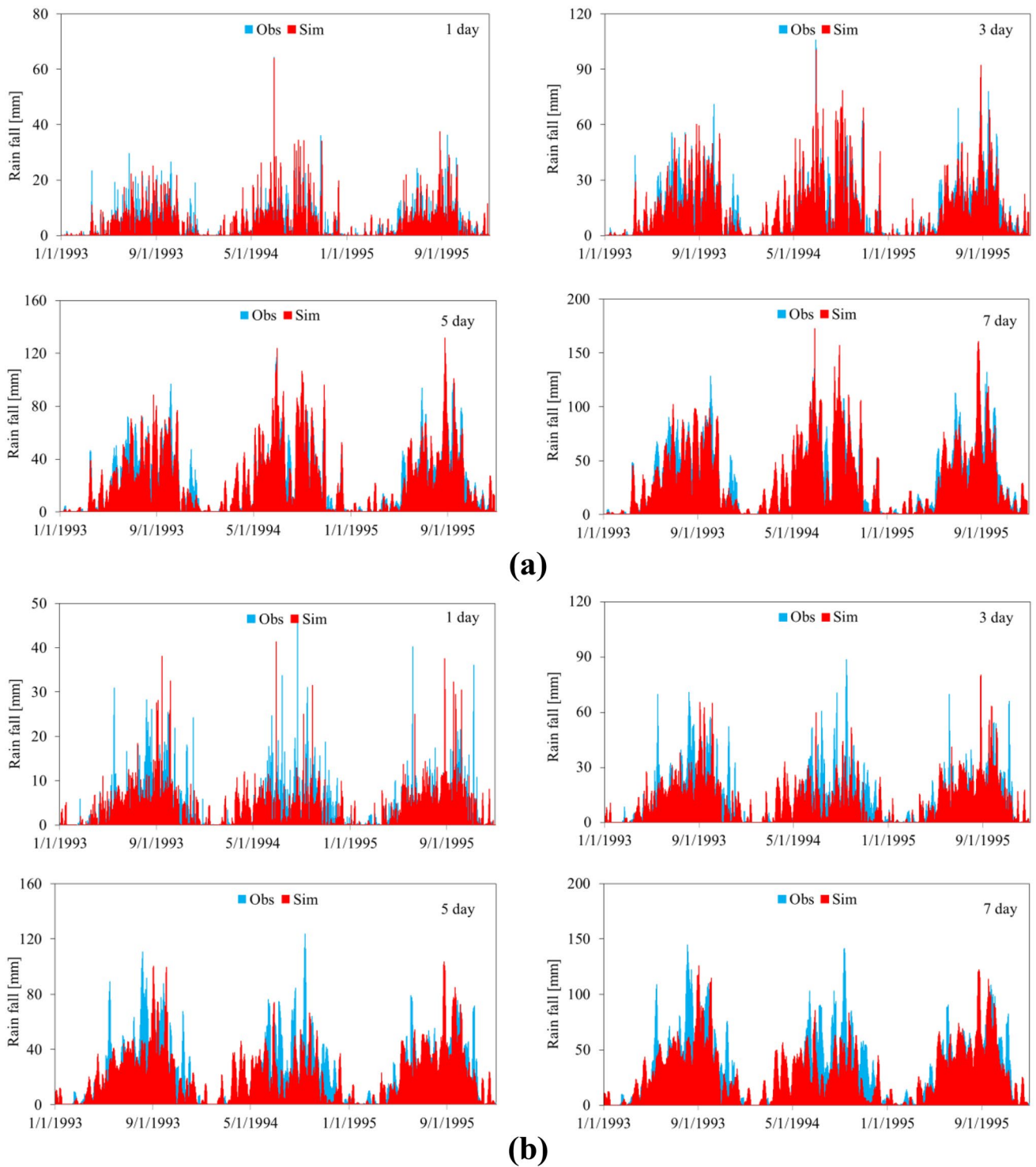. **b** Time series of ANN validations using ERA-20C at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-aver-age precipitation during 1993–1995 over the SG–DN. **c** Time series of ANN validations using CFSR at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1993–1995 over the SG–DN
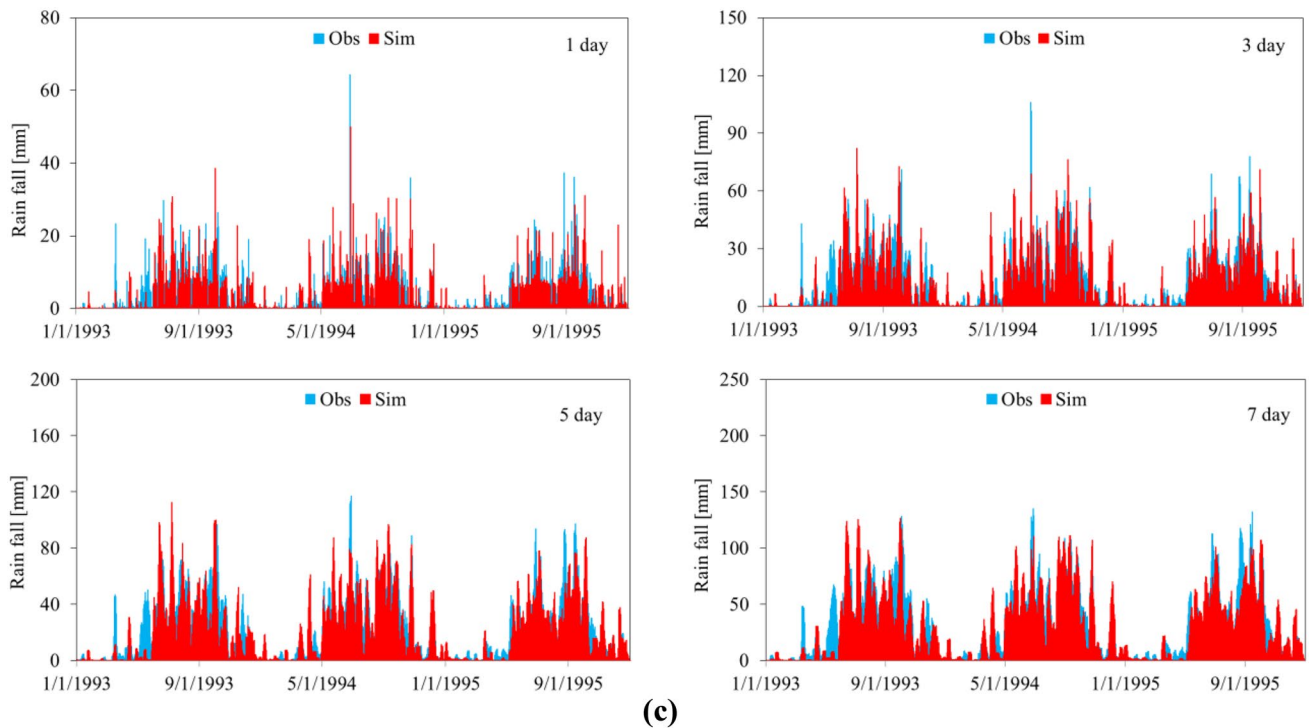
**(c)**

**Fig. 9** (continued)

**Table 9** Statistical test values of ANN-simulated results, based on ERA-Interim-simulation at D2, and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | ANN calibrations | | | | ANN validations | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.56 | 13.7 | 22.10 | 32.02 | 4.77 | 14.3 | 23.8 | 33.4 |
| Mean by simulation (mm) | 4.01 | 11.80 | 19.77 | 28.10 | 4.45 | 13.3 | 22.1 | 31.4 |
| Standard deviation by observation (mm) | 5.80 | 15.20 | 23.61 | 31.66 | 6.05 | 15.4 | 23.6 | 31.5 |
| Standard deviation by simulation (mm) | 5.01 | 13.69 | 20.11 | 26.21 | 5.61 | 14.2 | 21.9 | 29.6 |
| Correlation coefficient | 0.71 | 0.810 | 0.85 | 0.89 | 0.73 | 0.83 | 0.87 | 0.90 |
| Nash coefficient | 0.47 | 0.651 | 0.70 | 0.76 | 0.50 | 0.68 | 0.75 | 0.78 |

**Table 10** Statistical test values of ANN-simulated results, based on ERA20C-simulation at D2, and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | ANN calibrations | | | | ANN validations | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.56 | 13.7 | 22.10 | 32.02 | 4.77 | 14.3 | 23.8 | 33.4 |
| Mean by simulation (mm) | 4.01 | 12.11 | 20.11 | 28.77 | 4.15 | 12.5 | 18.5 | 28.7 |
| Standard deviation by observation (mm) | 5.80 | 15.20 | 23.61 | 31.66 | 6.05 | 15.4 | 23.6 | 31.5 |
| Standard deviation by simulation (mm) | 5.51 | 12.59 | 19.87 | 26.11 | 4.35 | 11.5 | 18.3 | 25.7 |
| Correlation coefficient | 0.65 | 0.79 | 0.83 | 0.86 | 0.65 | 0.78 | 0.79 | 0.83 |
| Nash coefficient | 0.40 | 0.57 | 0.66 | 0.71 | 0.42 | 0.56 | 0.61 | 0.65 |

dataset during the period from 1980 to 2010. It is noted that the HD technique significantly improved the precipitation spatial distribution compared to the WRF results. After applying the ANN simulations, the estimated precipitation is more representative of the observation values as shown in Figs. 10 and 11; hereby the details of the improvement can be seen through the color distribution. The light-yellow area, which indicates that the simulated data match closely to the VnGP high-resolution observation data, is larger for the HD technique results (Fig. 11) than for the WRF model results

**Table 11** Statistical test values of ANN-simulated results, based on CFSR-simulation at D2, and VnGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1986–1995 over the SG–DN

| Precipitation statistics | ANN calibrations | | | | ANN validations | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 day | 3 day | 5 day | 7 day | 1 day | 3 day | 5 day | 7 day |
| Mean by observation (mm) | 4.56 | 13.7 | 22.10 | 32.02 | 4.77 | 14.3 | 23.8 | 33.4 |
| Mean by simulation (mm) | 3.81 | 11.61 | 17.88 | 25.42 | 3.79 | 11.4 | 19.1 | 27.0 |
| Standard deviation by observation (mm) | 5.80 | 15.20 | 23.61 | 31.66 | 6.05 | 15.4 | 23.6 | 31.5 |
| Standard deviation by simulation (mm) | 4.88 | 12.49 | 19.22 | 26.71 | 4.89 | 11.5 | 18.8 | 24.8 |
| Correlation coefficient | 0.64 | 0.76 | 0.80 | 0.83 | 0.68 | 0.79 | 0.84 | 0.88 |
| Nash coefficient | 0.33 | 0.52 | 0.59 | 0.63 | 0.42 | 0.57 | 0.64 | 0.70 |

(Fig. 10). Tables 12, 13, 14, 15, 16, 17, 18 and 19 contain the percent of coverage area associated with the difference (mm) between the precipitation estimates and corresponding observations. The light-yellow areas of HD simulations under the CFSR dataset are 96.7%, 90.09%, 86.29%, and 80.19% for 1-, 3-, 5-, and 7-day precipitation, respectively; while the light-yellow areas of the WRF simulations under the CFSR dataset are smaller at 88.04%, 71.89%, 67.05%, and 65.29% for 1-, 3-, 5-, and 7-day precipitation, respectively. These results also confirm the improvement obtained after applying the proposed technique. Utilizing a combination of dynamical and statistical downscaling, it is possible to simulate and reconstruct precipitation data in both time series and spatial map over the selected watershed. This technique can apply not only for the simulation of precipitation, but also for other relevant atmospheric variables; such as temperature, wind speed, humidity, pressure, and radiation. Furthermore, this approach is useful to reconstruct and forecast weather risks, such as floods and droughts, because the results are produced at different time resolutions (Trinh et al. 2016, 2017) (i.e., hourly, daily, and monthly).

Eventually, the HD technique also is inexpensive computational demand with respect to computer resources and time consumption. Table 20 shows a comparison of computational resources and time consumption between the DD and HD techniques on a workstation with the same configuration of Intel® Xeon® Processor E5 v4 Family, Processor E5-2687WV4, 3.00 GHz, 24 cores. It shows the computation time and output storage are significantly reduced.

## 5 Conclusions

This study applies a new technique by means of coupling dynamical and statistical downscaling in order to overcome the limitations in both dynamic and statistical downscaling approaches. This new technique called hybrid downscaling (HD) not only incorporates the impacts of terrain factors and physical interactions between atmosphere and land surface conditions, but also improves the accuracy of simulated data, as it merges model simulations with observations over the modeled region. First, precipitation data were dynamical downscaled by a regional climate model, WRF, through the three domains (D1, D2, D3 for ERA-Interim and ERA-20C; D2, D3 for CFSR) with an inner domain (D3) of 9 km resolution under the three selected global reanalysis datasets (ERA-Interim, ERA-20C, and CFSR). After successful implementation and validation of the WRF model, the downscaled precipitation data were merged with the local observation data by means of the ANN model with back-propagation algorithm. The ANN utilizes the VnGP data from 1986 to 1992 for model training and calibration, and from 1993 to 1995 for model validation. Note that the calibration and validation are independent processes. This study demonstrated that the blended technique combining dynamical and statistical downscaling not only provides better data estimates in time series, but also in spatial distribution.

Among the three selected reanalysis datasets, the best calibration and validation results were obtained from the ERA-Interim dataset. Under the ERA-Interim dataset, the HD technique performance correlation coefficient (ranging from 0.89 to 0.90) and the Nash–Sutcliffe efficiency (0.76–0.78) are quite satisfactory (Table 9) based on the 7 day precipitation comparisons. These results are closer to the observation data than those using the CFSR dataset. However, the spatial difference of precipitation estimates using the CFSR dataset is lower than those under ECMWF—Atmospheric Reanalysis data (ERA-Interim and ERA-20C). One explanation is that the grid resolution of CFSR (0.5°), is finer than that of ERA-Interim (0.75°) and ERA-20C (1.25°).

Lastly, this technique can apply to simulate not only for precipitation but also for other relevant atmospheric variables; such as temperature, wind speed, humidity, pressure, and radiation. Furthermore, the new approach of this study can be applied widely in many parts of the world where the local observation data are available.

Future study will focus on modeling hydrologic conditions with inputs provided from the three hybrid downscaled datasets. Once a hydrologic model is implemented, it is possible to reconstruct and assess hydrologic conditions over the target region. In addition, the calibrated and validated WRF and ANN models for SG–DN can be utilized for the projection of future precipitation and stream
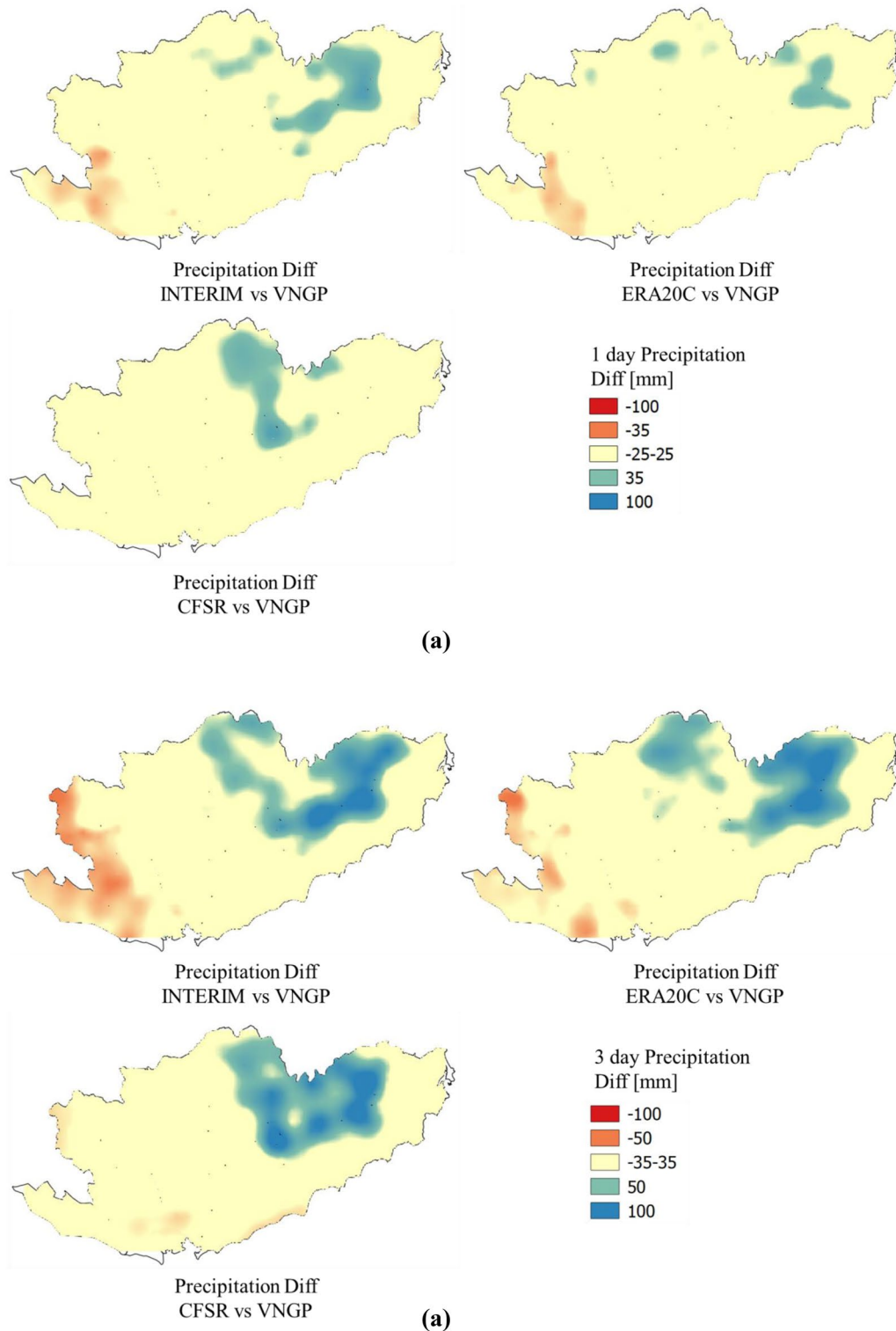
Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

1 day Precipitation
Diff [mm]
■ -100
■ -35
□ -25-25
■ 35
■ 100

Precipitation Diff
CFSR vs VNGP

(a)

Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

3 day Precipitation
Diff [mm]
■ -100
■ -50
□ -35-35
■ 50
■ 100

Precipitation Diff
CFSR vs VNGP

(a)

**Fig. 10 a** Spatial maps of the largest 1-day precipitation changes between simulation (WRF) and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **b** Spatial maps of the largest 3-day precipitation changes between simulation (WRF) and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **c** Spatial maps of the largest 5-day precipitation changes between simulation (WRF) and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **d** Spatial maps of the largest 7-day precipitation changes between simulation (WRF) and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnNGP), bottom (CFSR and VnGP)

Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

Precipitation Diff
CFSR vs VNGP

5 day Precipitation
Diff [mm]
-140
-75
-45-45
75
140

**(c)**



Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

Precipitation Diff
CFSR vs VNGP

7 day Precipitation
Diff [mm]
-160
-80
-50-50
80
160

**(d)**

**Fig. 10** (continued)

**Fig. 11** **a** Spatial maps of the largest 1-day precipitation changes between ANN simulation and observation (VnGP) data during 1980–2010, top (ERA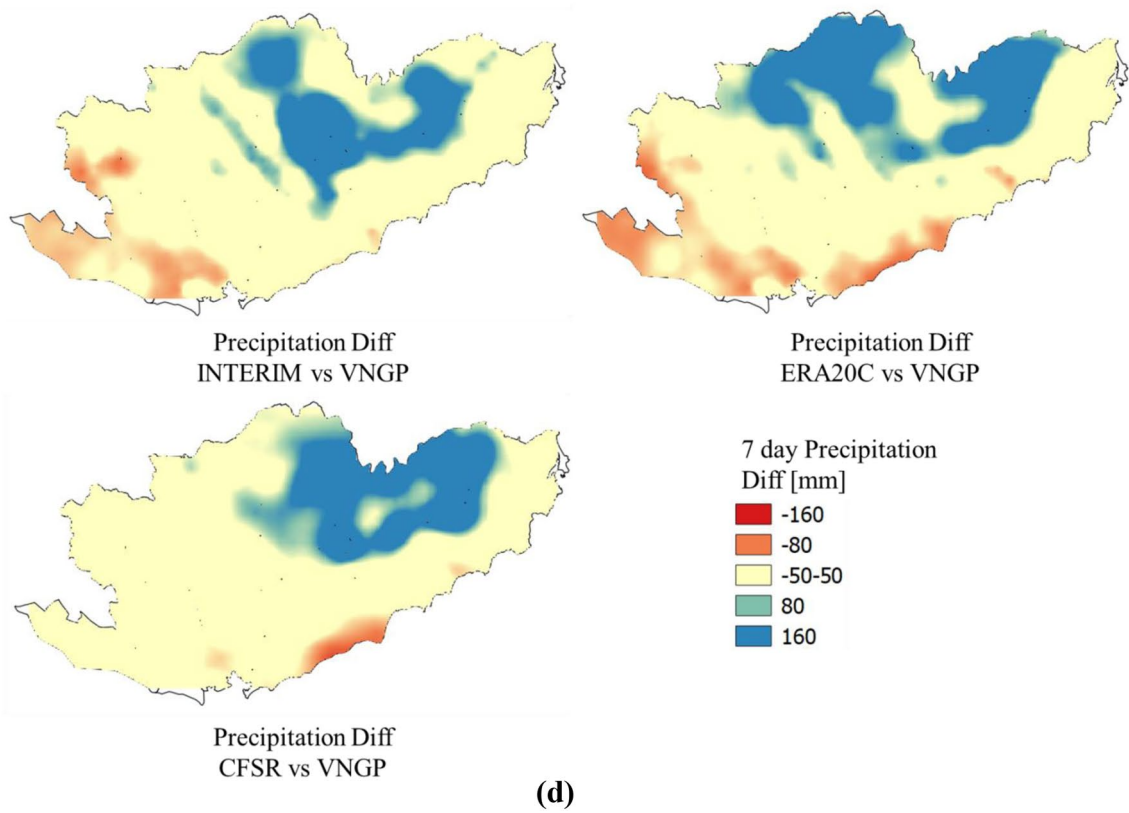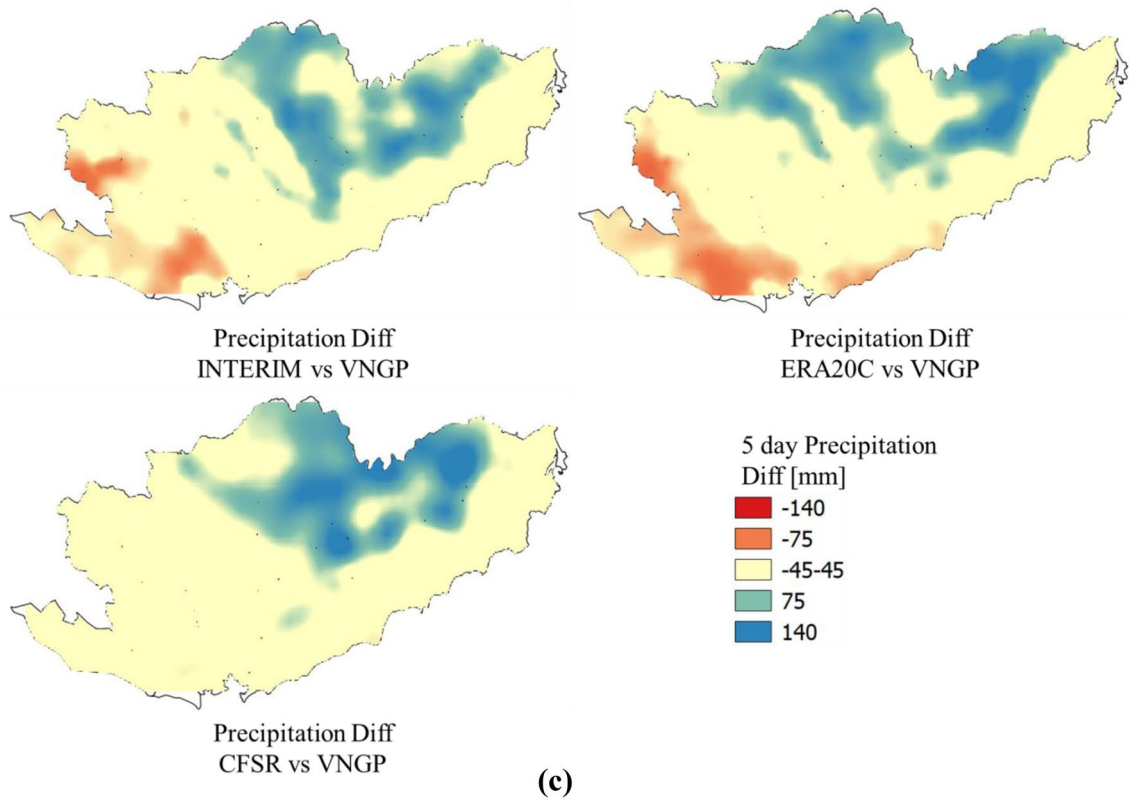-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **b** Spatial maps of the largest 3-day precipitation changes between ANN simulation and observation (VnGP) data during 1980–2010, top (ER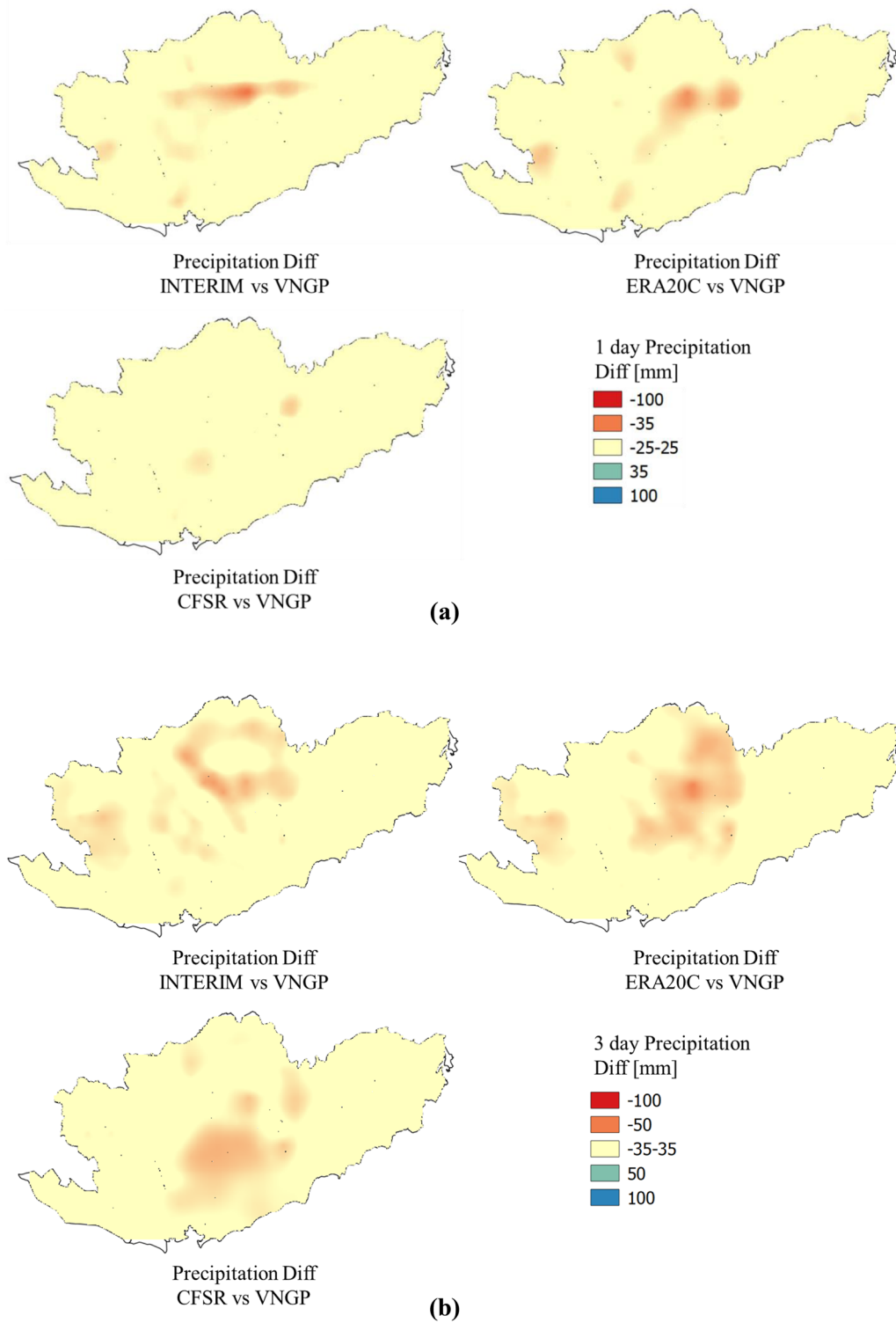A-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **c** Spatial maps of the largest 5-day pre-cipitation changes between ANN simulation and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP). **d** Spatial maps of the largest 7-day precipitation changes between ANN simulation and observa-tion (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP)

Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

Precipitation Diff
CFSR vs VNGP

5 day Precipitation
Diff [mm]
- -140
- -75
- -45-45
- 75
- 140

(c)



Precipitation Diff
INTERIM vs VNGP

Precipitation Diff
ERA20C vs VNGP

Precipitation Diff
CFSR vs VNGP

7 day Precipitation
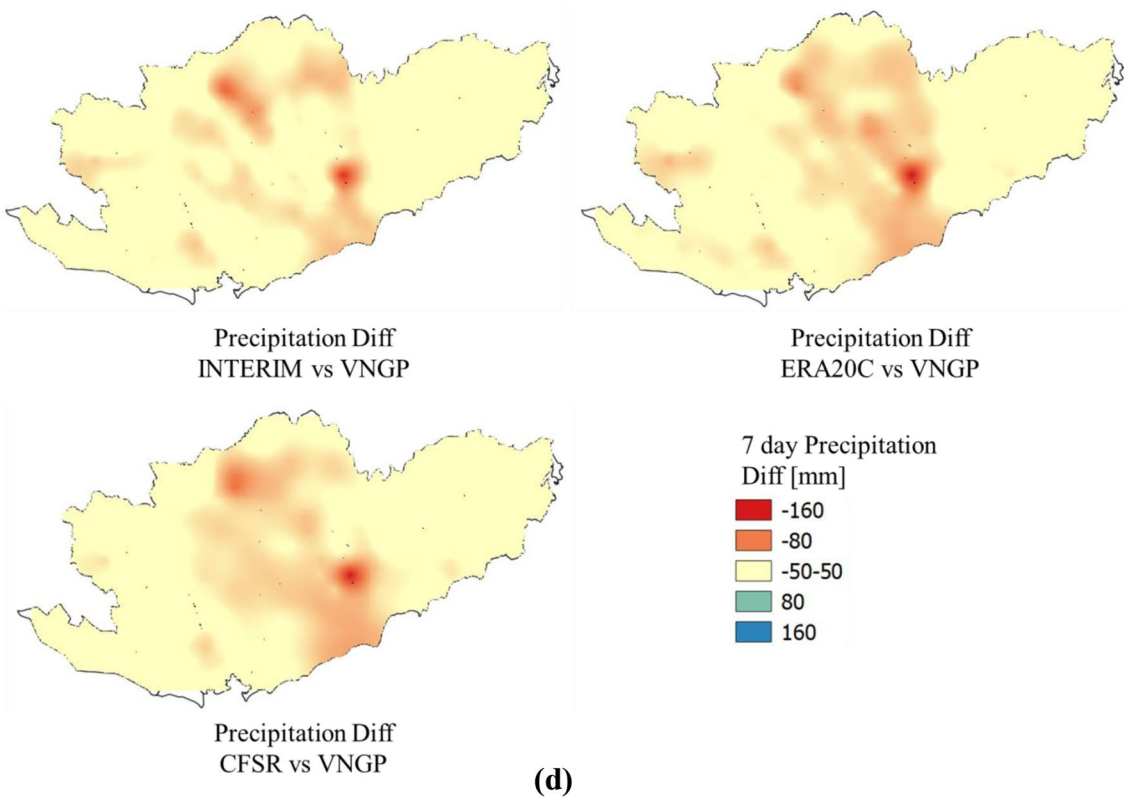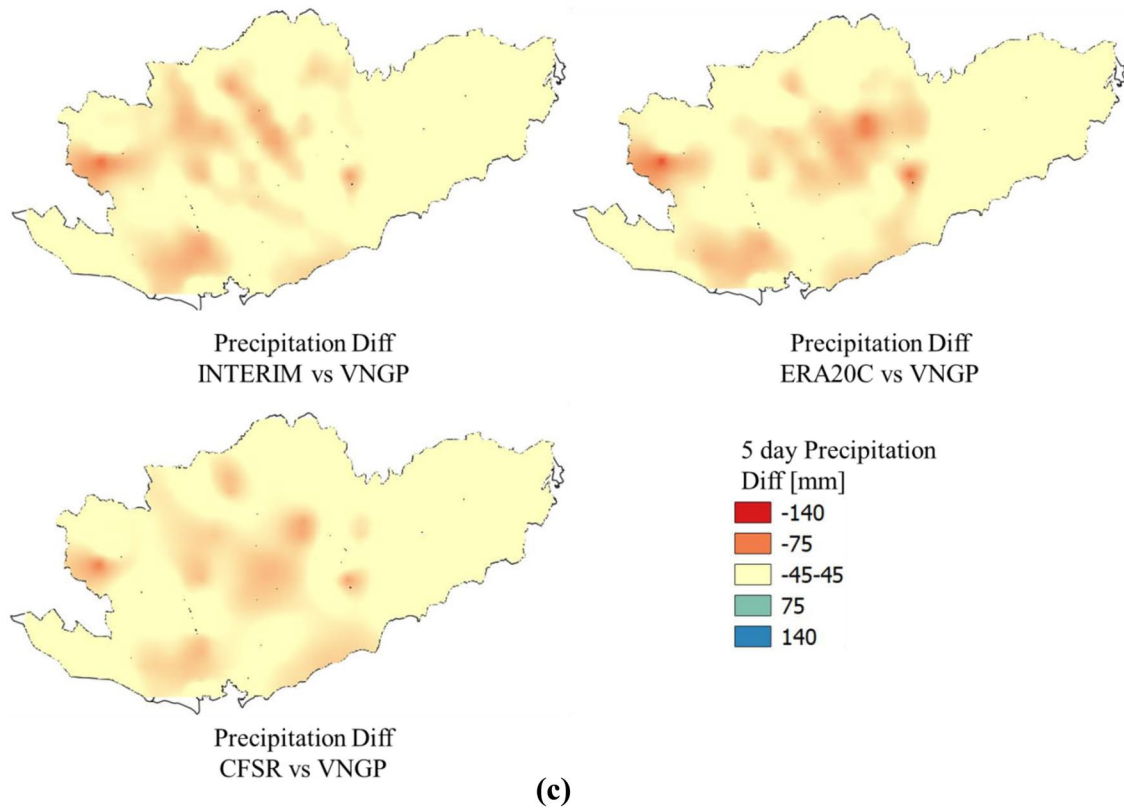Diff [mm]
- -160
- -80
- -50-50
- 80
- 160

(d)

**Fig. 11** (continued)

**Table 12** Percent coverage areas with respect to the difference of the largest 1-day precipitation estimates under the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR from the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 35 to − 25 | 7.48 | 4.76 | 0.20 |
| − 25 to 25 | 78.11 | 87.68 | 88.04 |
| 25 to 35 | 8.70 | 5.48 | 5.92 |
| 35 to 100 | 5.71 | 2.09 | 5.84 |

**Table 13** Percent coverage areas with respect to the difference of the largest 3-day precipitation estimates under the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR from the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 100 to − 50 | 0.22 | 0.23 | 0.00 |
| − 50 to − 35 | 14.88 | 9.88 | 4.71 |
| − 35 to 35 | 59.58 | 63.56 | 71.89 |
| 35 to 50 | 10.46 | 10.59 | 5.24 |
| 50 to 100 | 13.64 | 14.32 | 16.33 |
| > 100 | 1.22 | 1.42 | 1.82 |

**Table 14** Percent coverage areas with respect to the difference of the largest 5-day precipitation estimates under the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR from the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 140 to − 75 | 0.69 | 2.05 | 0.00 |
| − 75 to − 45 | 12.67 | 14.01 | 0.51 |
| − 45 to 45 | 55.08 | 50.08 | 67.05 |
| 45 to 75 | 14.46 | 13.47 | 12.84 |
| 75 to 140 | 17.09 | 18.83 | 17.48 |
| > 140 | 0.01 | 1.57 | 2.13 |

**Table 15** Percent coverage areas with respect to the difference of the largest 7-day precipitation estimates under the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR from the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 160 to − 80 | 0.04 | 0.33 | 0.49 |
| − 80 to − 50 | 12.44 | 17.72 | 3.55 |
| − 50 to 50 | 59.99 | 46.66 | 65.29 |
| 50 to 80 | 10.94 | 10.58 | 9.60 |
| 80 to 160 | 14.50 | 19.56 | 17.76 |
| > 160 | 2.09 | 5.14 | 3.32 |

**Table 16** Percent of coverage area, associated with difference magnitude (mm), of the largest 1-day precipitation estimates between the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR using ANN simulation and the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 100 to − 25 | 7.6 | 8.76 | 3.30 |
| − 25 to 25 | 92.4 | 91.24 | 96.7 |
| 25 to 35 | 0.00 | 0.00 | 0.00 |
| 35 to 100 | 0.00 | 0.00 | 0.00 |

**Table 17** Percent of coverage area, associated with difference magnitude (mm), of the largest 3-day precipitation estimates between the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR using ANN simulation and the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 100 to − 50 | 0.31 | 1.67 | 0.11 |
| − 50 to − 35 | 19.55 | 21.33 | 9.8 |
| − 35 to 35 | 80.14 | 77.00 | 90.09 |
| 35 to 50 | 0.00 | 0.00 | 0.00 |
| 50 to 100 | 0.00 | 0.00 | 0.00 |
| > 100 | 0.00 | 0.00 | 0.00 |

**Table 18** Percent of coverage area, associated with difference magnitude (mm), of the largest 5-day precipitation estimates between the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR using ANN simulation and the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 140 to − 75 | 1.33 | 2.11 | 1.15 |
| − 75 to − 45 | 25.45 | 28.56 | 16.56 |
| − 45 to 45 | 73.22 | 69.33 | 82.29 |
| 45 to 75 | 0.00 | 0.00 | 0.00 |
| 75 to 140 | 0.00 | 0.00 | 0.00 |
| > 140 | 0.00 | 0.00 | 0.00 |

**Table 19** Percent of coverage area, associated with difference magnitude (mm), of the largest 7-day precipitation estimates between the three different reanalysis datasets ERA-Interim, ERA-20C, CFSR using ANN simulation and the one corresponding to the VnGP during 1980–2010

| Difference (mm) | ERA-Interim (%) | ERA-20C (%) | CFSR (%) |
|---|---|---|---|
| − 160 to − 80 | 3.24 | 4.54 | 3.22 |
| − 80 to − 50 | 27.67 | 31.66 | 20.59 |
| − 50 to 50 | 69.09 | 63.8 | 76.19 |
| 50 to 80 | 0.00 | 0.00 | 0.00 |
| 80 to 160 | 0.00 | 0.00 | 0.00 |
| > 160 | 0.00 | 0.00 | 0.00 |

| **Table 20** A comparison of computational demand with respect to computer resources and time consumption between the DD and HD techniques (based on the same workstation) | Intel(R) Xeon(R) CPU E5-2687 W v4 @ 3.00 GHz×2 (24 cores, Memory: 64G, Storage: 3 T) | Dynamic downscaling method (DD) | Hybrid downscaling method (HD) |
|---|---|---|---|
| | Time consumption for 1 year simulation (using CFSR) | ~30 h | ~6 h |
| | Output storage for 1 year simulation (using CFSR) | 2.5 GB | 0.8 GB |

flow under future atmospheric inputs from the global climate models' future climate projections.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Benjamin SG, Dévényi D, Weygandt SS, Brundage KJ, Brown JM, Grell GA, Kim D, Schwartz BE, Smirnova TG, Smith TL, Manikin GS (2004) An hourly assimilation–forecast cycle: the RUC. Mon Weather Rev 132(2):495–518

Berrisford P, Dee DPKF, Fielding K, Fuentes M, Kallberg P, Kobayashi S, Uppala S (2009) The ERA-interim archive. ERA Rep Ser 1:1–16

Bougeault P, Lacarrere P (1989) Parameterization of orography-induced turbulence in a mesobeta-scale model. Mon Weather Rev 117(8):1872–1890

Brierley P (1998) Some practical application of neural networks in the electricity industry, Eng. D Thesis. Cranfield University, Cranfield

Burlando P, Rosso R (2002) Effects of transient climate change on basin hydrology. 1. Precipitation scenarios for the Arno River, central Italy. Hydrol Process 16(6):1151–1175

Chou MD, Suarez MJ (1999) A solar radiation parameterization (CLIRAD-SW) for atmospheric studies. NASA Tech. Memo NASA/TM-1999-104606, 40

Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100-year reanalysis using only surface pressure data. Bull Am Meteorol Soc 87(2):175–190

Coulibaly P (2004) Downscaling daily extreme temperatures with genetic programming. Geophys Res Lett 31:L16203. https://doi.org/10.1029/2004GL020075

Courtier P, Thepaut JN, Hollingsworth A (1994) A strategy for operational implementation of 4DVar using an incremental approach. Q J Roy Meteorol Soc 120:13671387

Cuong HV, Toan TQ (2019) Assessment of hydro-climatological drought conditions for Hong-Thai Binh river watershed in Vietnam using high-resolution model simulation. Vietnam J Sci Technol Eng 61(2):90–96

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer DP, Bechtold P (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Fletcher DS, Goss E (1993) Forecasting with neural networks: an application using bankruptcy data. Inf Manag 24:159–167

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27(12):1547–1578

Fuka DR, Walter MT, MacAlister C, Degaetano AT, Steenhuis TS, Easton ZM (2014) Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrol Process 28(22):5613–5623

Gebregiorgis AS, Hossain F (2012) Understanding the dependence of satellite rainfall uncertainty on topography and climate for hydrologic model simulation. IEEE Trans Geosci Remote Sens 51(1):704–718

Glahn HR, Lowry DA (1972) The use of model output statistics (MOS) in objective weather forecasting. J Appl Meteorol 11:1203–1211

Goyal MK, Ojha CSP (2011) Evaluation of linear regression methods as downscaling tools in temperature projections over the Pichola Lake Basin in India. Hydrol Process 25(9):1453–1465

Han J, Pan HL (2011) Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. Weather Forecast 26(4):520–533

Hashmi MZ, Shamseldin AY, Melville BW (2011a) Statistical downscaling of watershed precipitation using gene expression programming (gep). Environ Model Softw 26(12):1639–1646

Hashmi MZ, Shamseldin AY, Melville BW (2011b) Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stoch Environ Res Risk Assess 25(4):475–484

Hashmi MZ, Shamseldin AY, Melville BW (2013) Statistically downscaled probabilistic multi-model ensemble projections of precipitation change in a watershed. Hydrol Process 27(7):1021–1032

Ho C, Trinh T, Nguyen A, Nguyen Q, Ercan A, Kavvas ML (2019) Reconstruction and evaluation of changes in hydrologic conditions over a transboundary region by a regional climate model coupled with a physically-based hydrology model: application to Thao river watershed. Sci Total Environ 668:768–779

Ho C, Nguyen A, Ercan A, Kavvas ML, Nguyen V, Nguyen T (2020) Assessment of atmospheric conditions over the Hong Thai Binh river watershed by means of dynamically downscaled ERA-20C reanalysis data. J Water Clim Change 11(2):540–555

Hong SY, Dudhia J, Chen SH (2004) A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Weather Rev 132(1):103–120

Jang S, Kavvas ML (2015) Downscaling global climate simulations to regional scales: statistical downscaling versus dynamical downscaling. J Hydrol Eng 20(1):A4014006

Jang S, Kavvas ML, Ishida K, Trinh T, Ohara N, Kure S, Chen ZQ, Anderson ML, Matanga G, Carr KJ (2017) A performance evaluation of dynamical downscaling of precipitation over northern California. Sustainability 9(8):1457

Kavvas ML, Kure S, Chen ZQ, Ohara N, Jang S (2013) WEHY-HCM for modeling interactive atmospheric-hydrologic

processes at watershed scale. I: Model description. J Hydrol Eng 18(10):1262–1271

Kjellström E, Bärring L, Nikulin G, Nilsson C, Persson G, Strandberg G (2016) Production and use of regional climate model projections—a Swedish perspective on building climate services. Clim Serv 2:15–29

Krishnamurti TN, Jha B, Rasch PJ, Ramanathan V (1997) A high resolution global reanalysis highlighting the winter monsoon. Part I, reanalysis fields. Meteorol Atmos Phys 64(3):123–150

Lin Y, Colle BA (2011) A new bulk microphysical scheme that includes riming intensity and temperature-dependent ice characteristics. Mon Weather Rev 139(3):1013–1035

Liu Y, Fan K (2014) An application of hybrid downscaling model to forecast summer precipitation at stations in China. Atmos Res 143:17–30

Lledó L, Lead T, Dubois J (2013) A study of wind speed variability using global reanalysis data. AWS Truepower LLC, pp 1–12

Minh PT, Tuyet BT, Thao TTT (2018) Application of ensemble Kalman filter in WRF model to forecast rainfall on monsoon onset period in South Vietnam. Vietnam J Earth Sci 40(4):367–394

Moriasi DN, Gitau MW, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation. Trans ASBE 58:1763–1785

Nguyen-Thi HA, Matsumoto J, Ngo-Duc T, Endo N (2012) A climatological study of tropical cyclone rainfall in Vietnam. Sola 8:41–44

Nguyen-Xuan T, Ngo-Duc T, Kamimera H, Trinh-Tuan L, Matsumoto J, Inoue T, Phan-Van T (2016) The Vietnam gridded precipitation (VnGP) dataset: construction and validation. SOLA 12:291–296

Pilling CG, Jones JAA (2002) The impact of future climate change on seasonal discharge, hydrological processes and extreme flows in the Upper Wye experimental catchment, Mid-Wales. Hydrol Process 16(6):1201–1213

Poli P, Hersbach H, Tan D, Dee D, Thepaut JN, Simmons A, Peubey C, Laloyaux P, Komori T, Berrisford P, Dragani R (2013) The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). European Centre for Medium Range Weather Forecasts, Reading

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DG, Peubey C, Thépaut JN, Trémolet Y (2016) ERA-20C: an atmospheric reanalysis of the twentieth century. J Clim 29(11):4083–4097

Raghavan SV, Vu MT, Liong SY (2016) Regional climate simulations over Vietnam using the WRF model. Theor Appl Climatol 126(1–2):161–182

Raje D, Mujumdar PP (2011) A comparison of three methods for downscaling daily precipitation in the Punjab region. Hydrol Process 25(23):3575–3589

Reichler T, Kim J (2008) Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. J Geophys Res 113:D05106. https://doi.org/10.1029/2007JD009278

Rossi G, Vega T, Bonaccorso B (eds) (2007) Methods and tools for drought analysis and management, vol 62. Springer Science and Business Media, New York

Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H (2010) The NCEP climate forecast system reanalysis. Bull Am Meteorol Soc 91(8):1015–1058

Schoof JT, Pryor SC (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int J Climatol 21(7):773–790

Shepherd TG (2014) Atmospheric circulation as a source of uncertainty in climate change projections. Nat Geosci 7(10):703–708

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2005) A description of the advanced research WRF version 2 (No. NCAR/TN-468+STR). University Corporation for Atmospheric Research

Tran QA, Taniguchi K (2018) Coupling dynamical and statistical downscaling for high-resolution rainfall forecasting: Case study of the Red River Delta. Vietnam Prog Earth Planet Sci 5(1):1–18

Trinh T, Ishida K, Fischer I, Jang S, Darama Y, Nosacka J, Brown K, Kavvas ML (2016) New methodology to develop future flood frequency under changing climate by means of physically based numerical atmospheric-hydrologic modeling. J Hydrol Eng 21(4):04016001

Trinh T, Ishida K, Kavvas ML, Ercan A, Carr K (2017) Assessment of 21st century drought conditions at Shasta Dam based on dynamically projected water supply conditions by a regional climate model coupled with a physically-based hydrology model. Sci Total Environ 586:197–205

Trinh T, Ho C, Do N, Ercan A, Kavvas ML (2020) Development of high-resolution 72 h precipitation and hillslope flood maps over a tropical transboundary region by physically based numerical atmospheric–hydrologic modeling. J Water Clim Change 11(S1):387–406

Veerse F, Thepaut JN (1998, Part B) Multiple-truncation incremental approach for four-dimensional variational data assimilation. Q J R Meteorol Soc 124(550):1889–1908

Walton DB, Sun F, Hall A, Capps S (2015) A hybrid dynamical–statistical downscaling technique. Part I: development and validation of the technique. J Clim 28(12):4597–4617

Wang W, Xie P, Yoo SH, Xue Y, Kumar A, Wu X (2011) An assessment of the surface climate in the NCEP climate forecast system reanalysis. Clim Dyn 37(7):1601–1620

Wilby RL, Wigley TM (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21(4):530–548

Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. Environ Model Softw 17(2):147–159

Yang TC, Yu PS, Wei CM, Chen ST (2011) Projection of climate change for daily precipitation: a case study in Shih-Men reservoir catchment in Taiwan. Hydrol Process 25(8):1342–1354

Yang T, Li H, Wang W, Xu CY, Yu Z (2012) "Statistical downscaling of extreme daily precipitation, evaporation, and temperature and construction of future scenarios. Hydrol Process 26(23):3510–3523

Yokoi S, Matsumoto J (2008) Collaborative effects of cold surge and tropical depression-type disturbance on heavy rainfall in central Vietnam. Mon Weather Rev 136(9):3275–3287