

PACE-2021

International Congress on the Phenomenological Aspects of Civil Engineering

Research Article

20-23 June 2021

Coupling regional climate model and machine learning to model high-resolution precipitation

T. Trinh^{1,2,3}, H. Tran⁴.

¹Hydrologic Research Laboratory, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, CA

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

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

⁴Center for Training and International Cooperation, Vietnam Academy for Water Resources, Vietnam

Corresponding Author E-mail: tqtrinh@ucdavis.edu

Corresponding Author ORCID: 0000-0002-9113-1975

Keywords

Artificial neural network (ANN),
Weather research and forecasting (WRF),
ERA-Interim,
ERA20C,
CFSR.

Abstract

Accurate areal rainfall stems not only from atmospheric boundary conditions, but also from the quality and availability of soil, topography, and vegetation data. As a result, rainfall model errors are exacerbated by uncertainties in both atmospheric and land surface conditions. Hybrid technique combining dynamical and statistical downscaling is investigated in this research. The proposed downscaling method incorporates information from three global reanalysis data sets: ERA-Interim, ERA20C, and CFSR. The Weather Research and Forecasting (WRF) model is used to hybrid downscale this reanalysis of atmospheric data, which is then followed by the use of an artificial neural network (ANN) model to further downscale the WRF performance to a finer resolution over the studied area. The findings of this study indicate that the proposed method, which combines model simulations with observations over the modeled area, will improve the accuracy of simulated data. Another advantage of this method is the low cost of computation, both in terms of computation time, and performance storage.

1. Introduction

Rainfall is the product of complex interactions between climate variables, and land surface, land cover conditions. As a result, rainfall model errors are influenced by both atmospheric and land surface uncertainties.

Attempts to model rainfall events using global atmospheric models (GCMs) have recently been made (Krishnamurti et al. 1997; Compo et al. 2006; Lledó et al. 2013; Fuka et al. 2014) [1-4]. However, resolution of GCMs are too coarse for studying water supplies at the watershed scale. Then downscaling technologies are used to refine coarse grid resolution data. Statistical or stochastic downscaling (SD) and dynamical downscaling (DD) are the two most popular approaches.

A hybrid technique incorporating dynamical and statistical downscaling has been studied to address the limitations of both DD and SD approaches (Liu and Fan, 2014; Tran and Taniguchi, 2018; Walton et al., 2015) [5-7]. This new technique, known as hybrid downscaling (HD), downscale RCM outputs to a finer spatial resolution by first using large-scale atmospheric conditions calculated by a GCM for its lateral boundary conditions before being downscaled by an RCM model. The HD also focuses on the effects of terrain influences and the physical relations between the atmosphere and the land surface. Despite retaining a short simulation period, data resolution is increased. Another advantage of this technology is that it enhances simulated data accuracy by combining model simulations with observations over the modeled area.

This research used the WRF model in conjunction with a machine learning algorithm to simulate and recreate rainfall data. Three global reanalysis datasets are considered in the proposed downscaling technique: ECMWF - Atmospheric Reanalysis coarse climate data of the 20th century (ERA-20C) (Poli et al. 2013, 2016), ECMWF - Reanalysis Interim (ERA-Interim) (Berrisford et al., 2009; Dee et al., 2011), and Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010; Wang et al., 2011) [8-13]. The WRF is used to hybrid downscale these coarse scale atmospheric data, which is then accompanied by an artificial neural network (ANN) model to further downscale the WRF output to a finer resolution over the studied watershed. After calibrating and validating the WRF and ANN models using existing ground observation data, the hybrid approach is analyzed using time series and spatial analyses.

The Sai Gon – Dong Nai Rivers Basin has been determined in this experiment for the hybrid technique's implementation. Because of its strategic position and the complex physical processes that cause intense rainfall in this area, advanced technologies must be used to investigate severe rainfall processes, and model realistic historical rainfall events.

2. Methodology

2.1. Implementation of computational atmospheric model based on physical principle.

With inputs from the three reanalysis datasets, the WRF model was used for dynamical downscaling. With several physics options for moisture dynamics, microphysics processes, cumulus cloud parameterizations, planetary boundary layer (PBL) schemes, radiation schemes, and surface schemes, the WRF model can simulate vertical

and horizontal air motions. A sequence of three nested domains for WRF simulations are implemented in this report. The largest domain (D1) has a spatial resolution of 81 km (21×18 horizontal grid points) and covers the southern half of Vietnam as well as parts of Thailand, Laos, Cambodia, and Malaysia. D2 is the second domain with a resolution of 27 km (27×24 horizontal grid points), and D3 is the smallest domain with spatial resolution of 9 km (48×33 horizontal grid points). It is noted that WRF is based on all 3 domains only for ERA20C data, while the ERA-Interim and CFSR were used only on D2 and D3.

2.3. Implementation of ANN architecture with back-propagation algorithm

For downscaling precipitation simulated by the WRF model, one of the most common and simplest artificial neural network (ANN) architectures is considered in this research. In statistical downscaling, the ANN model is widely used.

The desired ANN architecture consists of three layers (input layer, hidden layer, and output layer) linked by synapse weights. The hidden layer's nodes were chosen from a range of $(2n+1)$ to $(2n0.5+m)$, where n is the number of input nodes and m is the number of output nodes (Fletcher and Goss, 1993) [14].

Approximation used for the weight change is given in Eq. (1) by the delta rule.

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

Where η is the learning rate parameter, w is the weights, and E^2 is the squared error (Brierley, 1998) [15].

2.4. Training and validation of ANN model over the target watershed for the three different reanalysis datasets

The ANN model is trained and validated using a gridded of 0.1° daily precipitation dataset (VnGP) in this analysis. It is better to split the entire data duration into preparation, validation, and testing components. Two-thirds of the data period is used for preparation, and the remaining time is used for calibration in most statistical downscaling exercises.

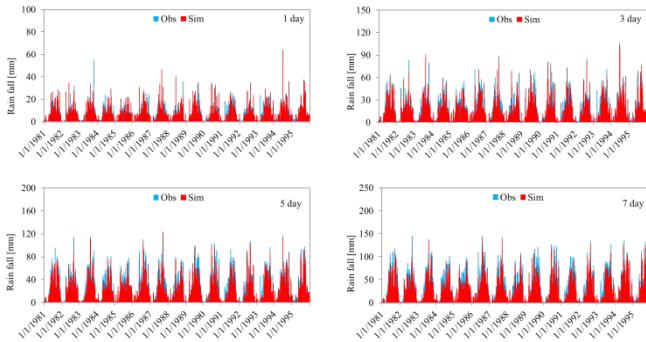


Figure 1. Time series of ANN calibrations using CFSR at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1981–1995 over the DN-SG.

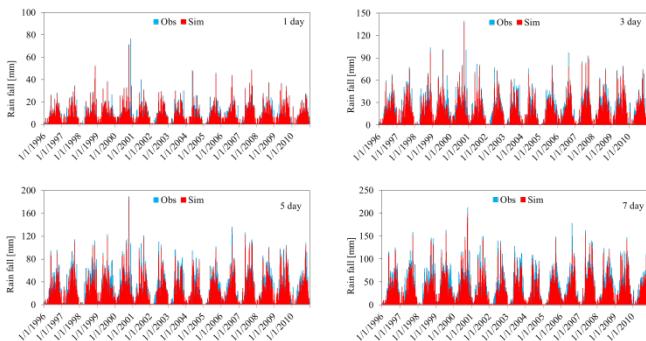


Figure 2. Time series of ANN validations using CFSR at D2 and VNGP for 1-, 3-, 5-, and 7-day basin-average precipitation during 1996–2010 over the DN-SG.

In general, the simulation results and observations responded well after calibration and validation. Statistical parameters that support the simulation results' alignment with VnGP data. The simulation output for daily precipitation is in the acceptable range, according to statistical parameters such as the correlation and Nash Sutcliffe efficiency coefficients ($0.7 \leq R^2 \leq 0.9$ and $0.5 \leq NSE \leq 0.78$).

3. Results and Discussion

The ERA-Interim dataset reported the best calibration and validation results out of the three reanalysis datasets. During the validation phase with the ERA-Interim dataset, the HD technique produced quite effective correlation coefficients from 0.89 to 0.90, and NSE from 0.76–0.78. The HD technique simulations using ERA-20C also obtained good validation results, with correlation coefficients ranging from 0.65–0.86. The ERA-Interim and ERA-20C datasets provided model calibration and validation results that were closer to the VnGP dataset than the CFSR dataset.

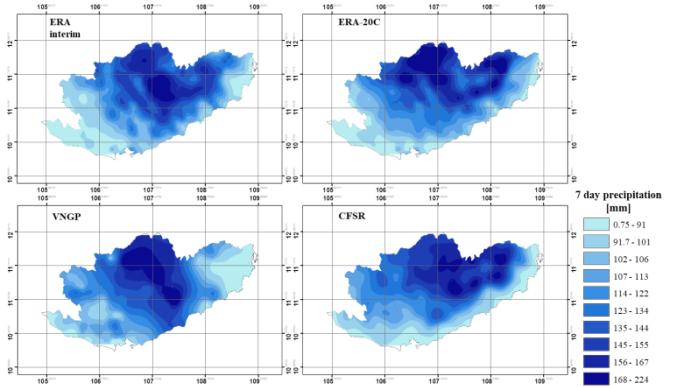


Figure 3. Spatial maps of the largest 7-day precipitation by means of simulation (WRF) and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP).

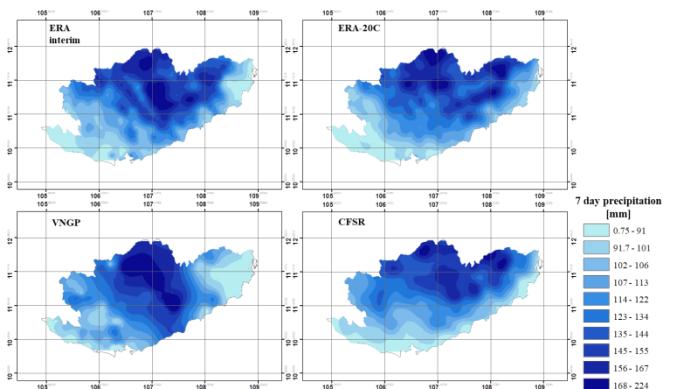


Figure 4. Spatial maps of the largest 7-day precipitation by means of ANN simulation and observation (VnGP) data during 1980–2010, top (ERA-Interim and VnGP; ERA-20C and VnGP), bottom (CFSR and VnGP).

Figure 3 presents the spatial distribution map of the largest difference in 7-day precipitation between the WRF model results and the VnGP dataset during the period from 1980–2010. Figure 4 presents the spatial distribution map of the largest 7-day precipitation between the HD model and VnGP dataset during the period from 1980–2010. When compared to the WRF data, the HD technique significantly improved

the precipitation spatial distribution. The forecast precipitation is more relevant to the observation values after applying the ANN simulations as shown in Figure 4.

It is possible to simulate precipitation data in both time series and spatial map over the selected watershed using a combination of dynamical and statistical downscaling. Not only can this technique be used to predict precipitation, but it can also be used to measure other significant atmospheric variables including temperature, wind speed, humidity, pressure, and radiation.

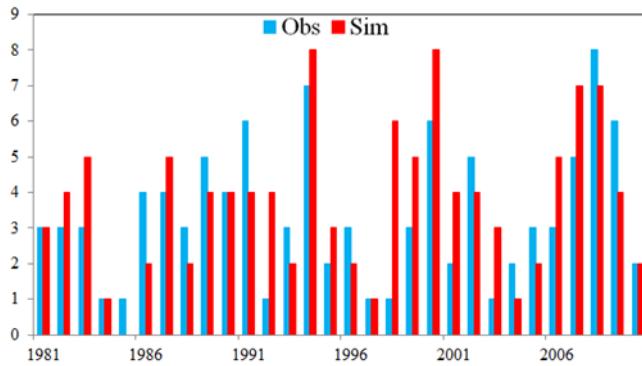


Figure 5. A comparison between the numbers of day having rainfall larger than 99% quartile in a year.

Figure 5 illustrates the difference of the numbers of day which have rainfall larger than 99% quartile in a year between simulation and observation data. When compared to VnGP dataset from 1980 to 2010, the simulation method gives the data relatively close to observation.

This approach is suggested to reconstruct and forecast weather risks, such as floods and droughts, because the results are produced at different time resolutions (Trinh et al., 2016; Trinh et al., 2017)[16, 17] (i.e., hourly, daily, and monthly).

4. Summary and Conclusion

To overcome the limitations of both dynamic and statistical downscaling methods, this study applied a hybrid approach that enhances dynamical and statistical downscaling. Hybrid downscaling (HD) is a modern technique that not only combines the effects of terrain variables and physical interactions between the environment and land surface conditions, but also increases the precision of simulations. First, under the three selected global gridded datasets, precipitation data were dynamically downscaled by a regional climate model, WRF, by three domains (D1, D2, D3 for ERA-Interim and ERA-20C; D2, D3 for CFSR) with an inner domain (D3) of 9 km resolution (ERA-Interim, ERA-20C, and CFSR). The downscaled precipitation data were combined with the local observation data using the ANN model with back-propagation algorithm after the WRF model was successfully implemented and validated. The ANN uses VnGP data for model training and calibration from 1986 to 1992, and model validation from 1993 to 1995. 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 0.89-0.90) and the NSE (0.76-0.78) are quite satisfactory in the validation period. These findings are more robust than those obtained with 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 reason is that the grid resolution of CFSR (0.5°), is finer than that of ERA-Interim (0.75°) and ERA-20C (1.25°).

Acknowledgements

This research was funded by Ho Chi Minh City's Department of Science and Technology (HCMC-DOST) and Institute for Computational Science and Technology (ICST) under the grant number 16/2020/HĐ-QPTKHCN.

Declaration of Conflict of Interests

The authors declares that there is no conflict of interest. We have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1.] Krishnamurti, T.N., Jha, B., Rasch, P.J. and Ramanathan, V.. A high resolution global reanalysis highlighting the winter monsoon. Part I, reanalysis fields. *Meteorology and Atmospheric Physics*, 64(3) (1997) 123-150.
- [2.] Compo, G.P., Whitaker, J.S. and Sardeshmukh, P.D., Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, 87(2) (2006) 175-190.
- [3.] Lledó, L., Lead, T. and Dubois, J., A study of wind speed variability using global reanalysis data, (2013).
- [4.] Fuka, D.R., Walter, M.T., MacAlister, C., Degaetano, A.T., Steenhuis, T.S. and Easton, Z.M., Using the Climate Forecast System Reanalysis as weather input data for watershed models. *Hydrological Processes*, 28(22) (2014) 5613-5623.
- [5.] Liu, Y., & Fan, K., An application of hybrid downscaling model to forecast summer precipitation at stations in China. *Atmospheric research*, 143 (2014) 17-30.
- [6.] Anh, Q.T. and Taniguchi, K., Coupling dynamical and statistical downscaling for high-resolution rainfall forecasting: Case study of the Red River Delta, Vietnam. *Progress in Earth and Planetary Science*, 5(1) (2018) 1-18.
- [7.] Walton, D. B., Sun, F., Hall, A., & Capps, S., A hybrid dynamical-statistical downscaling technique. Part I: Development and validation of the technique. *Journal of Climate*, 28(12) (2015) 4597-4617.
- [8.] Poli, P., Hersbach, H., Tan, D., Dee, D., Thepaut, J.N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P. and Dragani, R., 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*, (2013).
- [9.] Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G., Peubey, C., Thépaut, J.N. and Trémolet, Y., ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11) (2016) 4083-4097.
- [10.] Berrisford, P., Dee, D.P.K.F., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S. and Uppala, S., The ERA-interim archive. *ERA report series*, (1) (2009) pp.1-16.
- [11.] Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P. and Bechtold, P., The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656) (2011) 553-597.
- [12.] Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D. and Liu, H., The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8) (2010) 1015-1058.
- [13.] Wang, W., Xie, P., Yoo, S.H., Xue, Y., Kumar, A. and Wu, X., An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Climate dynamics*, 37(7) (2011) 1601-1620.
- [14.] Fletcher D S. and Goss E., Forecasting with neural networks: an application using bankruptcy data. *Information and Management*, 24 (1993) 159-167.
- [15.] Brierley P., Some practical application of neural networks in the electricity industry, Eng.D. Thesis, Cranfield University, UK, (1998).
- [16.] Trinh, T., Ishida, K., Fischer, I., Jang, S., Darama, Y., Nosacka, J., Brown, K. and Kavvas, M.L., New methodology to develop future flood frequency under changing climate by means of physically based numerical atmospheric-hydrologic modeling. *Journal of Hydrologic Engineering*, 21(4) (2016) 04016001.
- [17.] Trinh, T., Ishida, K., Kavvas, M.L., Ercan, A. and Carr, K., 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. *Science of the Total Environment*, 586 (2017) 197-205.