QBS
Vol. 41, No. 1, pp. 1-7 (2022)
https://doi.org/10.22283/qbs.2022.41.1.1
pISSN 2288-1344 eISSN 2508-7185 **Original Article**

Quantitative Bio-Science
Official Journal of the Institute of Natural Science at Keimyung University
Available Online at https://qbs.kmu.ac.kr:442/

Bias-correction of Reanalysis Data and Missing Value Estimation for Extreme Precipitation Across Asia

Hyebin Park[1], Juyoung Hong[1], Jeong-Soo Park[1,*]

[1]Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea

(Received January 11, 2022; Revised May 9, 2022; Accepted May 9, 2022)

ABSTRACT

So-called reanalysis data are useful for climatic data analysis. However, some reanalysis data, such as APHRODITE (Asia Precipitation Highly Resolved Observational Data Integration Towards Evaluation), have biases in their mean and variance, especially for extreme rainfall. Undeniably, it needs further analysis for improvement. This study applies a multivariate bias-correction method to Korea, southern China, Philippines, and Thailand to realize this improvement. Our focus in this study is the annual maximum daily precipitation (AMP1). After correcting APHRODITE data over grids, we construct a time series of the AMP1 observations for some locations with no observational records by applying the so-called back-interpolation technique. This technique is also used to estimate missing values in the AMP1 series. This method is illustrated in some cities in North Korea and the Philippines.

Key words : Barnes interpolation, Climate observations, Covariance matrix, Grid data, Heavy rainfall, Missing vale estimation, Spatial dependency

1. Introduction

In climatic data analysis, the observations and so-called reanalysis data are used. The weather stations are irregular in the sense that starting years are different, spatial distribution is not on grid, and the location is sometimes moved over years. Spatial resolutions for each nation are different. Reanalysis data are developed to overcome this disadvantage of the observations. APHRODITE (Asia Precipitation Highly Resolved Observational Data Integration Towards Evaluation) is a reanalysis data for a daily gridded rainfall covering a period of more than 57 years collected based on rain gauge information across Asia [1]. The spatial resolution is high as $0.5° \times 0.5°$. It has been widely used for the evaluation of the numerical model and for assessing climatic change [2-5]. However, some works have indicated that it significantly underestimates the extreme rainfall for several regions compared to the observations based on weather stations [6,7]. Major reasons of this bias are due to quality control of data and different interpolation method [7]. In this study, we first check the differences of extreme rainfall between APHRODITE and the observations across Korea, souther China, Philippines, and Thailand. Then we apply a multivariate bias-correction method to APHRODITE to reduce the bias and to estimate missing values of extreme rainfall. Our main focus in this study is the annual maximum daily precipitation (AMP1).

Fig. 1 shows 72 observation stations and 41 grid points over Korean peninsula. Note that there are spacious gap in North Korea. For the interpolation in North Korea, we have used some data in nearby stations of China.

*Correspondence should be addressed to Jeong-Soo Park, Professor, Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea. Tel: +82-62-530-3445, Fax: +82-62-530-3449, E-mail: jspark@jnu.ac.kr
Hyebin Park and Juyoung Hong are graduate students, Department of Mathematics andStatistics, Chonnam National University, Gwangju 61186, Korea.

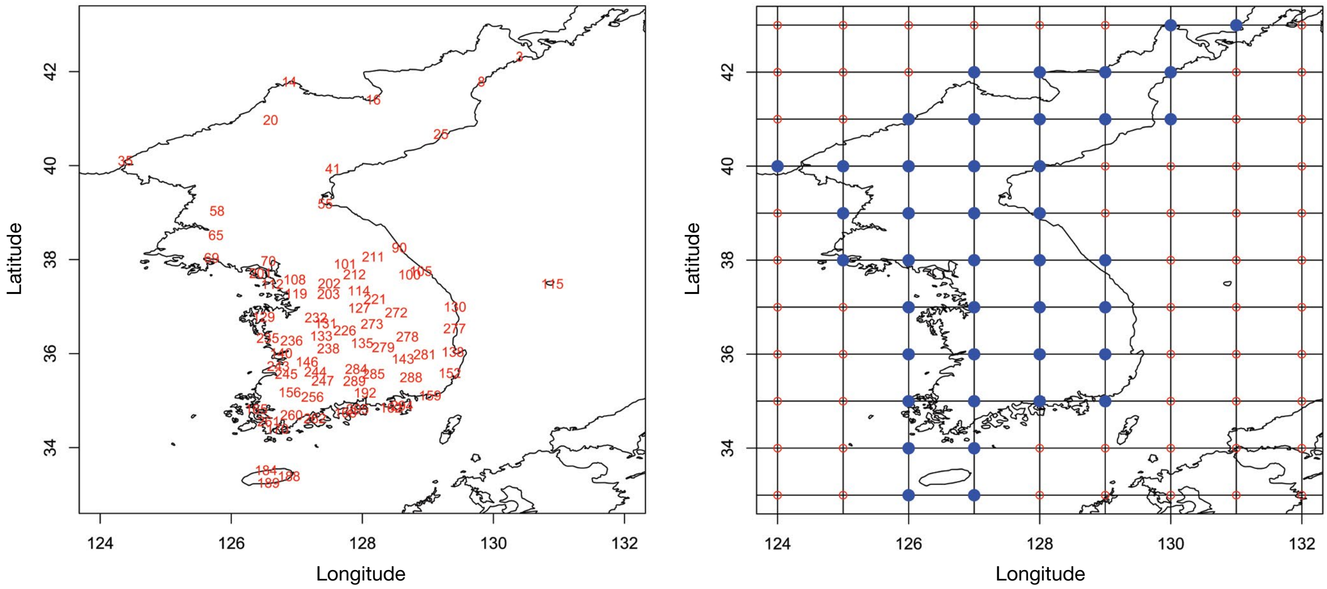

Fig. 1. Map of the Korean peninsula with 72 weather stations in the automated synoptic observing system (ASOS) available in this study (left panel) and with 41 grids of 1° × 1° on which maximum rainfall for dotted grids are analyzed (right panel).

2. Bias Correction

2.1 Difference between the observations and APHRODITE

The AMP1 in APHRODITE has serious bias in its mean and variance, as shown in Fig. 2. It shows examples of time series plots of the AMP1 from the observations (OBS, blue line) and APHRODITE data (green line) for some observational stations near the grid points over Korea, southern China, Philippines, and Thailand. From this figure, we see APHRODITE underestimates the extreme observations for almost locations. The variations of APHRODITE are smaller than those of the observations. Thus an improvement is needed for extreme rainfall of APHRODITE for further analysis such as for assessing climatic change.

The observations in Korea, in southern China, in Philippines, and in Thailand were obtained respectively from the Korea Meteorological Administration (KMA) [8], from Regional Climate Group (University of Gothenburg) [9], from the Philippines Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) [10], and from the Thai Meteorological Department [11].

2.2 Barnes interpolation

To predict the precipitation of grid points based on weather stations, an interpolation technique is usually used. Among many interpolation methods, we used the iterative Barnes interpolation scheme [12,13]. The Barnes technique produces a rainfall field on a regular grid from irregularly distributed rainfall observation stations [3,7].

For given data, $y(x_1), y(x_2), \cdots, y(x_n)$ at sites $x_1, x_2, \cdots, x_n$, we want to interpolate y for a new site (or grid) x. The first iteration of the Barnes scheme is

$$g_0(x) = \sum_{k=1}^{n} w_k y(x_k) \sum_{k=1}^{n} w_k, \tag{1}$$

where $w_k = exp(-\frac{d(x,x_k)^2}{k_0})$ is the weight and $k_0 = 5 \times (2\delta_n \pi)$ for the average distance between locations δ_n. The second iteration updates the weight to $w_k^{(1)} = exp(-\frac{d(x,x_k)^2}{\gamma \times k_0})$ for $0 < \gamma < 1$ with 12 default. Then a new interpolation is

$$g_1(x) = g_0(x) + \sum_{k=1}^{n} w_k^{(1)} [y(x_k) - g_0(x)] \sum_{k=1}^{n} w_k^{(1)}. \tag{2}$$

Note that the last term in (2) is a correction based on the residual $y(x_k) - g_0(x)$. The third iteration is

$$g_2(x) = g_1(x) + \sum_{k=1}^{n} w_k^{(2)} [y(x_k) - g_1(x)] \sum_{k=1}^{n} w_k^{(2)}, \tag{3}$$

where $w_k^{(2)} = exp(-\frac{d(x,x_k)^2}{\gamma \times k_1})$ is again the updated weight with $k_1 = \gamma k_0$. These iterations continue until convergence.

We used a R function "interpBarnes" in OCE library [14]. The default interaction is 2, which means it stops after the third iteration. For the interpolation on a grid, we used 5 nearest neighbor stations surrounding the grid point in actual com-

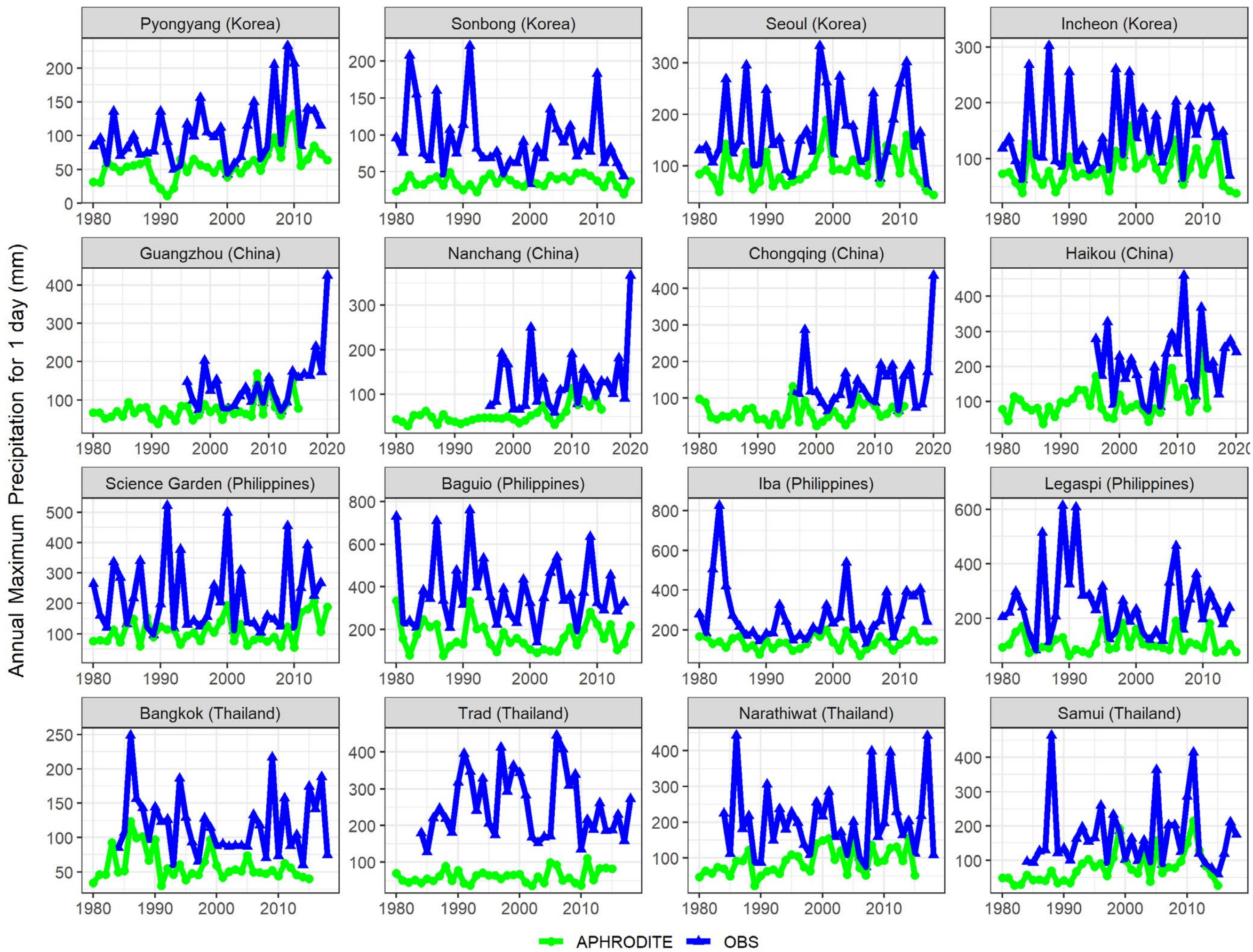

Fig. 2. Examples of time series plots of the observations (OBS, blue line) and APHRODITE data (green line) for some observational stations near the grid points over Korea, southern China, Philippines, and Thailand.

putation. Instead of using 5 nearest neighbors, one can apply a threshold of distance from the grid to determine the number of nearest stations, even though we did not try it in this study. It introduces a problem of selecting the threshold, but may be a good way to choose nearest stations.

2.3 Multivariate bias correction

Lee et al. [7] indicated this problem in South Korea and suggested to improve it using a quantile mapping (QM) bias-correction (BC) method [15,16]. For the same purpose, in this study, we employ the multivariate bias correction (MBC) method which may be better than QM method [17]. The MBC is a multivariate generalization of quantile mapping (QM). The MBC is applicable not only for several climate variables but also for spatial observations of one variable. It deal with spatial dependency (by a covariance matrix) among the observations of nearby stations [18]. In the MBC, an image processing technique designed to transfer color information from one image to another is adapted. In each iteration of MBC method, univariate QM is first applied separately to each variable. Then a linear multivariate BC [15] is applied by rescaling the multivariate anomalies based on Cholesky decomposition of the covariance matrix. The algorithm ends when both the corrected marginals and the dependence structure are sufficiently close to their observed counter parts. The MBC algorithm consists of a random orthogonal rotation of multivariate input data, a univariate quantile delta mapping on the rotated fields, and the inverse rotation, in each iteration. In this study, the multivariate input is consisted of nearest neighbor spatial reanalysis data around a grid point. Thus for each grid point, six-variate BC was performed because five nearby stations are co-operated with the one grid point, even though the climate

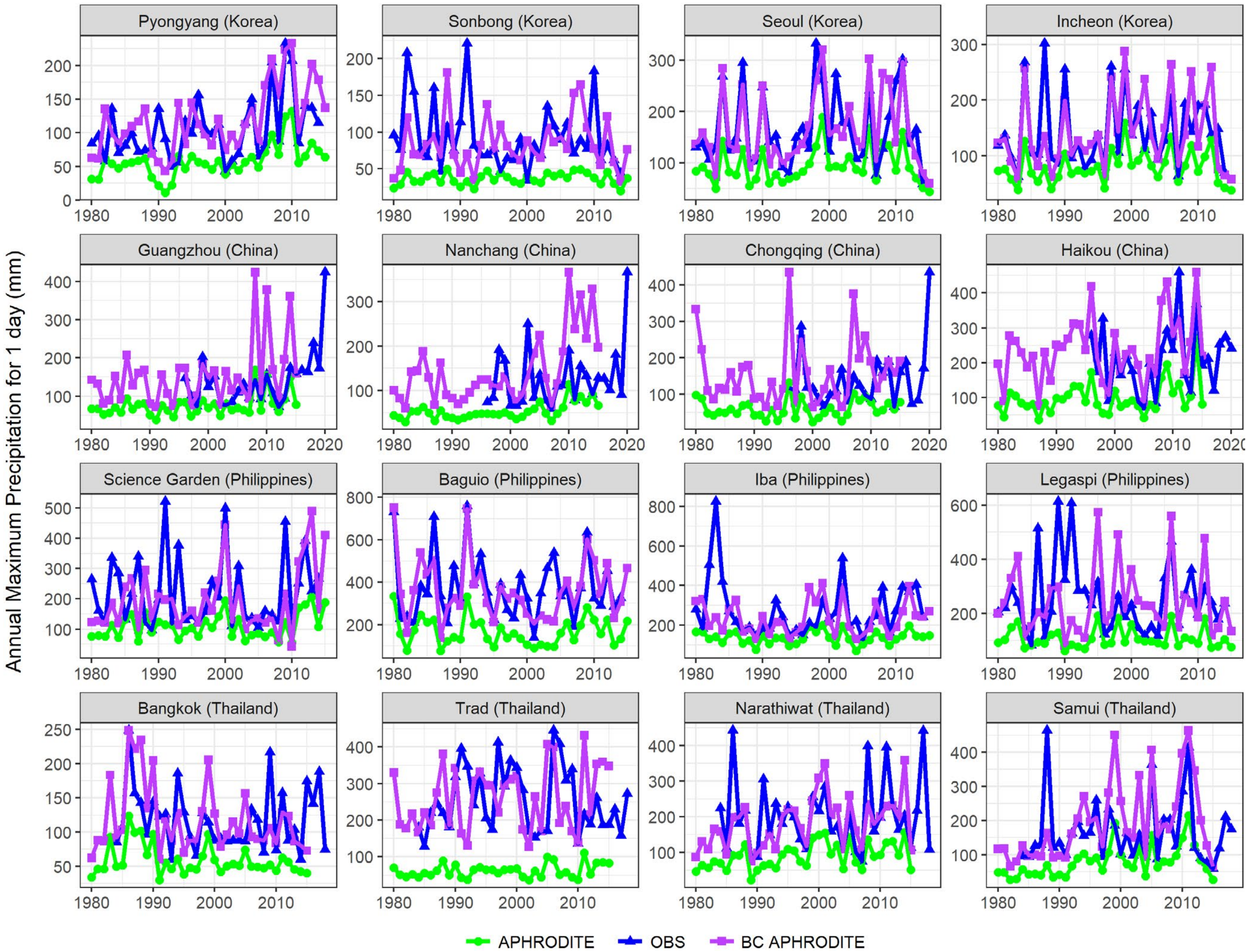

Fig. 3. Same as Fig. 2 but the bias-corrected values (purple line) are added.

variable is only the extreme precipitation. See [17] for more details.

For a given grid point, we applied MBC to APHRODITE data based on the observations of nearest neighbor stations. We used 'MBC' package [19] in R for computation. The details on various multivariate BC methods are available in [20].

Fig. 3 shows the bias-corrected APHRODITE values (purple line) for some grids near the observational stations over Korea, southern China, Philippines, and Thailand.

2.4 Back-interpolation

When we obtained the bias-corrected APHRODITE values over grid points, we can apply Barnes interpolation again to some locations (not necessarily on grids) where there are no observational records. It is an interpolation from grids to uneven locations, which is a reverse action from the above subsection 2.2. We call it "back-interpolation". In this study, the bias corrected APHRODITE data and the observations at stations (if those are available near to the target station) were used together for a better back-interpolation. This technique is useful to construct a series of new data as illustrated in the next subsection for North Korea.

2.5 Application to North Korea

As seen in Fig. 1, there are spacious gap in North Korea. We firstly obtained the bias-corrected APHRODITE values over grid points, and then we constructed a time series of the AMP1 observations for some cities by the back-interpolation. Those data obtained by this approach are presented in Table 1 for 5 locations (Hoiryoung, Hoicheon, Guseong, Goksan, and Jangjin), for example.

Table 1. Time series of the annual maximum daily precipitation (AMP1) of the APHRODITE from 1973 to 2014 for five locations (Hoiryoung, Hoicheon, Guseong, Goksan, and Jangjin) in North Korea. The values in the parenthesis are constructed by a back-interpolation technique

Location	Time series of AMP1 in North Korea						
Hoiryoung	55 (54)	65 (49)	40 (79)	102 (67)	56 (56)	63 (51)	69 (57)
	62 (43)	46 (63)	198 (189)	78 (65)	213 (63)	57 (78)	185 (168)
	76 (77)	59 (160)	60 (86)	68 (62)	98 (47)	111 (53)	93 (55)
	50 (164)	83 (105)	75 (143)	47 (64)	79 (53)	76 (56)	61 (99)
	101 (109)	167 (77)	89 (109)	73 (71)	67 (82)	50 (82)	75 (91)
	96 (96)	58 (82)	144 (89)	55 (52)	106 (132)	46 (65)	81 (51)
Hoichen	257 (86)	107 (68)	140 (70)	26 (46)	202 (81)	175 (87)	45 (49)
	58 (60)	247 (118)	79 (74)	77 (107)	187 (92)	83 (95)	151 (73)
	134 (146)	104 (142)	176 (120)	97 (104)	140 (50)	68 (84)	89 (154)
	104 (117)	472 (181)	119 (108)	88 (93)	114 (96)	71 (134)	48 (70)
	88 (125)	42 (79)	91 (129)	86 (112)	51 (104)	83 (77)	141 (191)
	89 (105)	157 (146)	217 (300)	91 (210)	133 (252)	280 (443)	91 (104)
Guseong	266 (85)	117 (76)	121 (64)	66 (59)	190 (73)	123 (73)	137 (88)
	110 (75)	170 (103)	132 (91)	171 (106)	96 (104)	136 (161)	199 (90)
	110 (122)	99 (115)	160 (91)	182 (76)	113 (59)	55 (59)	126 (161)
	170 (105)	172 (127)	86 (98)	374 (85)	113 (107)	131 (94)	134 (92)
	142 (127)	118 (130)	106 (133)	100 (137)	74 (106)	95 (90)	111 (176)
	114 (94)	109 (107)	260 (260)	77 (125)	139 (157)	219 (171)	211 (111)
Goksan	77 (60)	68 (57)	156 (101)	86 (75)	133 (77)	176 (103)	101 (75)
	216 (116)	141 (124)	101 (115)	78 (98)	85 (84)	211 (83)	81 (112)
	118 (155)	101 (120)	123 (89)	206 (62)	75 (50)	63 (62)	143 (106)
	91 (96)	80 (100)	118 (192)	159 (83)	94 (140)	92 (129)	79 (105)
	135 (117)	88 (95)	173 (120)	101 (121)	143 (119)	67 (374)	183 (112)
	88 (133)	204 (363)	209 (184)	161 (179)	279 (124)	164 (136)	63 (85)
Jangjin	107 (56)	40 (31)	56 (35)	57 (40)	83 (31)	94 (59)	92 (43)
	44 (34)	70 (67)	75 (66)	78 (77)	64 (78)	66 (75)	80 (76)
	56 (82)	97 (71)	96 (56)	72 (41)	77 (30)	58 (64)	45 (79)
	99 (83)	72 (96)	35 (73)	52 (78)	29 (56)	39 (81)	62 (75)
	48 (86)	56 (62)	37 (70)	69 (81)	39 (67)	74 (58)	120 (91)
	56 (74)	118 (114)	132 (117)	83 (82)	98 (140)	101 (141)	77 (73)

3. Missing Value Estimation

There are sometimes missing values in the series of the AMP1. Missing values are usually incurred due to incorrect measurements or handling error. When we have the bias-corrected APHRODITE values, it can be used to estimate the missing values in the AMP1 of the station. That is, we employed the back-interpolation to estimate the missing values of the AMP1 from the bias-corrected APHRODITE values of nearest grids surrounding the station with missing. WE think this approach to estimate missing values in the observations is simple and easy to use, so be useful for climate data analysis.

3.1 Application to the Philippines

There are some missing values in the AMP1 in some stations in the Philippines. We thus estimated those missing values using back-interpolation. Table 2 indicates time series of the AMP1 from 1973 to 2014 for four locations (Butuan, Cotabato, Itbayat, and Virac Synop) in the Philippines constructed by a back-interpolation technique. The values in the parenthesis just after NA are the estimated missing values.

Fig. 4 shows how the missing values are estimated by the back-interpolation method for four stations (Butuan, Cotabato, Itbayat, and Virac Synop) in the Philippines. The blue line

Table 2. Time series of the annual maximum daily precipitation (AMP1) of the APHRODITE from 1973 to 2014 for four locations (Butuan, Cotabato, Itbayat, and Virac Synop) in the Philippines. The values in the parenthesis are constructed by a back-interpolation technique. The values in the parenthesis just after NA are the estimated missing values

Location	Time series of AMP1						
Butuan	NA (139)	NA (150)	NA (102)	NA (122)	NA (109)	NA (208)	NA (98)
	NA (121)	176 (96)	110 (117)	109 (182)	128 (91)	272 (110)	147 (129)
	130 (117)	191 (158)	96 (109)	94 (100)	98 (100)	95 (86)	172 (132)
	168 (129)	132 (196)	128 (104)	98 (90)	86 (169)	111 (135)	168 (176)
	167 (177)	117 (110)	134 (150)	186 (191)	98 (179)	160 (165)	91 (117)
	120 (164)	201 (107)	118 (96)	207 (202)	153 (169)	118 (122)	129 (243)
Cotabato	NA (88)	NA (98)	NA (75)	NA (127)	NA (86)	NA (100)	NA (202)
	NA (88)	NA (118)	NA (141)	NA (82)	NA (102)	NA (73)	67 (76)
	69 (87)	104 (100)	142 (128)	142 (88)	96 (74)	82 (86)	61 (137)
	99 (113)	136 (150)	202 (176)	83 (139)	85 (82)	80 (76)	146 (69)
	89 (95)	78 (109)	101 (83)	97 (61)	83 (87)	130 (122)	115 (66)
	109 (117)	141 (91)	81 (68)	75 (140)	96 (82)	66 (79)	77 (130)
Itbayat	187 (144)	246 (230)	151 (182)	155 (168)	67 (177)	104 (252)	119 (56)
	202 (310)	340 (182)	450 (401)	1000 (143)	396 (58)	208 (236)	160 (194)
	838 (716)	43 (174)	366 (101)	NA (169)	NA (100)	NA (151)	23 (287)
	190 (127)	41 (275)	250 (202)	241 (299)	94 (121)	NA (256)	129 (397)
	241 (248)	194 (386)	297 (298)	223 (137)	174 (259)	303 (709)	129 (160)
	171 (87)	175 (190)	210 (116)	164 (211)	256 (154)	134 (209)	257 (235)
Virac Synop	NA (320)	NA (256)	NA (207)	NA (226)	NA (104)	NA (356)	NA (97)
	NA (138)	250 (213)	230 (195)	340 (108)	95 (155)	179 (125)	264 (126)
	340 (76)	122 (227)	143 (130)	95 (340)	110 (131)	76 (133)	207 (274)
	132 (228)	179 (171)	191 (236)	183 (206)	415 (127)	273 (140)	199 (300)
	149 (265)	126 (137)	302 (363)	220 (122)	114 (234)	172 (325)	343 (221)
	117 (217)	245 (165)	138 (127)	245 (284)	136 (124)	98 (173)	240 (289)

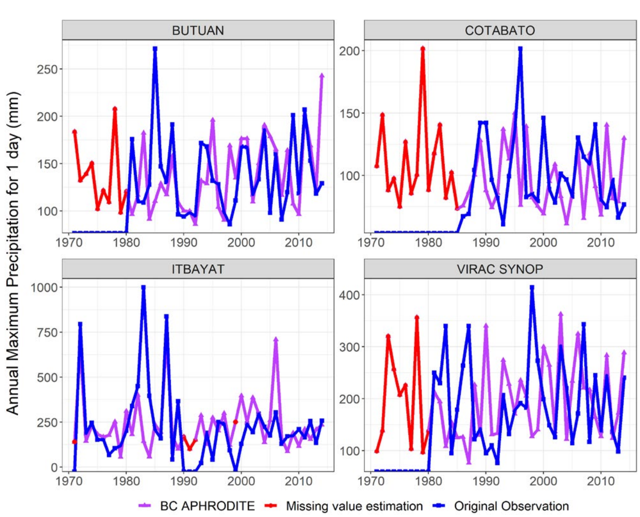

Fig. 4. The missing values estimates (red line) obtained by the back-interpolation method for four stations (Butuan, Cotabato, Itbayat, and Virac Synop) in the Philippines. The back-interpolated values (purple line) and the observations (blue line) are also presented.

stands for the observations, purple line and red line are back-interpolated values and estimated missing values, respectively.

4. Summay and Discussion

We found there are significant bias in APHRODITE reanalysis data, specially in heavy rainfall, in its mean and variance for such areas as in Korea, southern China, Philippines, and Thailand. We thus tried to obtain an improved data by reducing biases using a multivariate bias correction (MBC) method. When we had the bias-corrected reanalysis data on grids, so-called back-interpolation is applied to construct a time series of the AMP1 (annual maximum daily precipitation) on the locations where there are no observational records. The back-interpolation is also applied to estimate missing values in the observations. This approach is illustrated with some cities in North

Korea and in the Philippines. It would be nice to extend this approach to all grids over Asia. We leave this work for a future study.

For interpolation method in this study, we used the Barnes iterative technique. However, one can employ other methods such as Kriging or bilinear interpolation or thin-plate spline or machine learning methods [21,22]. The result will be different upon what method is employed for the interpolation.

The back-interpolation (BkI) technique may be useful to check any bias or uncertainty due to bias correction and interpolation methods. For a given location with good observations, we firstly apply Barnes interpolation (BrI) and MBC on nearby grid. Then, from the bias-corrected reanalysis data, we apply BkI to the given original location to check how much difference between the original observations and the back-interpolated values. If the difference is small, we can say the methods (BrI, MBC, and BkI) are sound and acceptable.

Heavy rainfall can have a significant effect on human life, infrastructure, agriculture and livestock, and natural ecosystems. Hence, in addressing the impact of frequent downpour events, governments and communities should prepare the proper systems and infrastructure more securely and carefully to prevent critical damage such as a loss of life from landslides and floods.

Acknowledgements

The authors thank the reviewers who have provided valuable comments to improve this paper. This study was supported by BK21 FOUR (NO. 5120200913674) funded by the Ministry of Education and National Research Foundation of Korea (NRF).

References

1. Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A. APHRODITE: Constructing a long-term daily griddied precipitation dataset for Asia based on a dense network of rain gauges. Bull Amer Meteorol Soc 2012;93:1401-1415.
2. Kishore P, Jyothi S, Basha G, Rao SVB, Rajeevan M, Velicogna I, et al. Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends. Clim Dynam 2016;46:541-556.
3. Shin Y, Shin Y, Hong J, Kim M-K, Byun Y-H, Boo K-O, et al. Future projections and uncertainty assessment of precipitation extremes in the Korean Peninsula from the CMIP6 ensemble with a statistical framework. Atmosphere 2021;12:97.
4. Supharatid S, Aribarg T, Nafung J. Bias-corrected CMIP6 climate model projection over Southeast Asia. Theor Appl Climat 2022;147:669-690.
5. Ohn I, Seo SB, Kim S, Kim Y-O, Kim Y. Uncertainty decomposition in climate-change impact assessments: a Bayesian perspective. CSAM 2020;27:109-128.
6. Ono K, Kazama S. Analysis of extreme daily rainfall in Southeast Asia with a gridded daily rainfall data set. IAHS publication 2011;34;169-175.
7. Lee G, Cha D-H, Park C-Y. Improvement of extreme summer precipitation over South Korea in APHRODITE data. Climate Research 2017;12:41-51.
8. Korean Meteorological Administration. https://data.kma.go.kr/cmmn/main.do.
9. Regional Climate Group (University of Gothenburg) Data Archive. http://rcg.gvc.gu.se/.
10. PAGASA (Philippines Atmospheric, Geophysical and Astronomical Services Administration). https://www.pagasa.dost.gov.ph (accessed on 27 December 2021).
11. Thai Meteorological Department (TMD) https://www.tmd.go.th. (accessed on 17 December 2021).
12. Barnes SL. A technique for maximizing details in numerical weather map analysis. Jour Appl Meteor 1964;3:396-409.
13. Koch SE, DesJardins M, Kocin PJ. An interactive Barnes objective map analysis scheme for use with satellite and conventional data. J Climate Appl Meteo 1983;22:1487-1503.
14. Kelly D. Package 'oce'. https://cran.rproject.org/web/packages/oce/oce.pdf.
15. Maraun D, Widmann M. Statistical downscaling and bias correction for climate research. New York: Cambridge Univ Press; 2018.
16. Hong J, Park JS. Multivariate bias correction for climate simulation data, with application to precipitation extremes in Korea. QBS 2018;38:121-130.
17. Cannon AJ. Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. Clim Dynam 2018;50:31-49.
18. Heo J, Lee JY, Kim W. Bayesian information criterion accounting for the number of covariance parameters in mixed effects models. CSAM 2020;27:301-311.
19. Cannon AJ. Package 'MBC' in R, version 0.10-4. https://cran.rproject.org/web/packages/MBC/MBC.pdf.
20. Francois B, Vrac M, Cannon AJ, Robin Y, Allard D. Multivariate bias corrections of climate simulations: which benefits for which losses?. Earth Syst Dynam 2020;11;537-562.
21. Hofstra N, Haylock M, New MG, Jones PD, Frei C. Comparison of six methods for the interpolation of daily European climate data. J Geophys Res-atmos 2008;113;D21110.
22. Rhee JT, Ahn WB, Oh KJ. Using machine learning algorithms to forecast the optimal bidding rate in apartment auctions. QBS 2021;40;31-37.