A Comparative Verification of Raw and Bias-Corrected ECMWF Seasonal Ensemble Precipitation Reforecasts in Java (Indonesia)

DIAN NUR RATRI

Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands, and Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG), Jakarta, Indonesia

KIRIEN WHAN AND MAURICE SCHMEITS

Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

(Manuscript received 10 August 2018, in final form 23 May 2019)

ABSTRACT

Dynamical seasonal forecasts are afflicted with biases, including seasonal ensemble precipitation forecasts from the new ECMWF seasonal forecast system 5 (SEAS5). In this study, biases have been corrected using empirical quantile mapping (EQM) bias correction (BC). We bias correct SEAS5 24-h rainfall accumulations at seven monthly lead times over the period 1981–2010 in Java, Indonesia. For the observations, we have used a new high-resolution (0.25°) land-only gridded rainfall dataset [Southeast Asia observations (SA-OBS)]. A comparative verification of both raw and bias-corrected reforecasts is performed using several verification metrics. In this verification, the daily rainfall data were aggregated to monthly accumulated rainfall. We focus on July, August, and September because these are agriculturally important months; if the rainfall accumulation exceeds 100 mm, farmers may decide to grow a third rice crop. For these months, the first 2-month lead times show improved and mostly positive continuous ranked probability skill scores after BC. According to the Brier skill score (BSS), the BC reforecasts improve upon the raw reforecasts for the lower precipitation thresholds at the 1-month lead time. Reliability diagrams show that the BC reforecasts have good reliability for events exceeding the agriculturally relevant 100-mm threshold. A cost/loss analysis, comparing the potential economic value of the raw and BC reforecasts for this same threshold, shows that the value of the BC reforecasts is larger than that of the raw ones, and that the BC reforecasts have value for a wider range of users at 1- to 7-month lead times.

1. Introduction

Seasonal forecasts of precipitation are becoming an increasingly important element in decision-making systems in Indonesia, especially in the agriculture and hydrological sectors. In these two sectors, stakeholders and decision-makers need seasonal forecasts to assist them in their planning strategy. Predicting when above- or below-average rainfall might occur can be valuable for agriculture because such information can help farmers to decide the type of crop that they will plant during that season. For example, farmers may plant a crop that requires less water if they have been informed in advance that the forecast rainfall will be below average. The seasonal forecast is also important in an irrigated system regarding the availability of irrigation water as well as

the timing of high/low river stream flows. Currently, demands on seasonal forecasting are getting higher: users require seasonal rainfall forecasts that are skillful, statistically reliable, and free of bias (e.g., [Schepen et al. 2014](#)).

Global and regional circulation models that can forecast seasonal atmospheric and oceanic conditions are afflicted with biases to a degree that prevents their direct use for hydrological purposes (e.g., [Ehret et al. 2012](#)). Ensemble forecasts of seasonal precipitation from the European Centre for Medium-Range Weather Forecasts (ECMWF) also exhibit these systematic biases ([Johnson et al. 2019](#)).

Uncorrected meteorological forecasts are not suitable as direct input for quantitative models, such as those used in agriculture and water management ([Schepen et al. 2016](#)). The bias should be corrected because it can lead to significant errors in impact assessments ([Murphy 1999](#)).

Corresponding author: Maurice Schmeits, schmeits@knmi.nl

DOI: 10.1175/JAMC-D-18-0210.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

The seasonal rainfall forecast is a critical input for predicting the planting date, and the type of the rice crop that will be planted depends on either the expected duration of the rainy season or the forecast amount of rainfall in the dry season. It is expected that more skillful seasonal forecast information will lead to more successful agricultural activities and more skillful warnings for possible floods and droughts.

Apart from biases, ensemble forecasts can also have either a too-narrow or too-wide spread. So, before ECMWF or other GCM forecasts can be practically applied, statistical postprocessing is needed to correct these biases and errors in dispersion (e.g., Wilks 2011).

There are several existing statistical postprocessing methods, ranging from basic bias correction (BC) to more advanced methods (e.g., Piani et al. 2010; Salvi et al. 2011; Iizumi et al. 2011; Lafon et al. 2013; Teutschbein and Seibert 2010). Recently, a number of studies have assessed the skill of ECMWF ensemble forecast system S4 and its applicability in different contexts (Weisheimer and Palmer 2014; Kim et al. 2012; Di Giuseppe et al. 2013; Trambauer et al. 2015; Wetterhall et al. 2015). Schepen et al. (2014, 2016) have successfully applied calibration methods to 1-month temperature and precipitation forecasts in Australia from another seasonal forecasting system, the Predictive Ocean Atmosphere Model for Australia (POAMA). Two popular BC methods in seasonal forecasting are distribution mapping (e.g., quantile mapping) and linear scaling (Crochemore et al. 2016).

Amengual et al. (2012) used the empirical quantile mapping (EQM) method as a quantile–quantile calibration method based on a nonparametric function that corrects biases in the cumulative distribution functions (CDFs) of climatic variables. Reichle and Koster (2004), Boé et al. (2007), Déqué (2007), and Amengual et al. (2012) highlighted that quantile mapping is a flexible method that is widely used for correcting biases in meteorological variables. The quantile mapping method can be used to bias correct forecasts that can then be used to feed crop and hydrological models. Another advantage of applying that method is that it can be applied to various climate variables and various orographies (Wilcke et al. 2013). Quantile mapping also performs particularly well over complex terrain, similar to that found in Indonesia. As an example, Wilcke et al. (2013) showed that it exhibits good performance in Austria where there are high mountains in the west and flat regions in the east.

In Indonesia, seasonal forecasts are one of the main products of the Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG). They are released twice a year: for the wet and dry seasons (BMKG 2018).

For farmers in Southeast Asia, including Indonesia, a common problem is a false start of the rainy season (Marjuki et al. 2014), which is an isolated rainfall event preceding the expected monsoon onset date but followed by a dry spell. This is one factor that makes more accurate forecasting a challenge. Currently, seasonal forecasts that are provided by BMKG are mainly based on the mean of the output from four statistical models using observational rainfall data from stations (Komalasari et al. 2016). It is expected that the new ECMWF seasonal prediction system 5 (SEAS5) will become the main seasonal forecast system in Indonesia, so probabilistic forecasts from a dynamical model can be implemented to provide seasonal forecasts. There is ample evidence that probabilistic forecasts are able to improve decisions, if decision-makers use the information correctly (Rayner et al. 2005; Joslyn and LeClerc 2012; Ramos et al. 2013). In this study EQM is applied to correct for biases and other systematic errors in the precipitation forecasts of SEAS5. Empirical quantile mapping has been largely applied. However, our research provides a number of important insights, first, the seasonal forecast in Indonesia has not yet been bias corrected, to our knowledge. This is very important because Indonesian users from many sectors use the seasonal forecast information to guide them in making decisions about their future plans. To our knowledge, empirical quantile mapping is mostly applied to monthly data in seasonal forecasting, but here we show that improved skill can also be gained by bias-correcting daily forecasts. The bias correction of daily forecasts is necessary for the use of bias-corrected data in a modeling chain, for example, where bias-corrected meteorological outputs are used as inputs to a hydrological and/or crop model.

This study will focus on the driest months, that is, July, August, and September. This period is important because farmers have to decide whether it is possible to grow a third rice crop based on the seasonal rainfall forecasts for these months. About 70% of the lowland rice area in Indonesia produces two crops per year. A skillful forecast for July, August, and September precipitation could give farmers an indication of the total field area that can be planted with rice for the third crop. A third rice crop is possible in this period if the season has a number of consecutive months with precipitation above $100 \text{ mm month}^{-1}$ (Oldeman 1980). However, late-season planting patterns could change if the precipitation during this period increases or decreases significantly (Naylor et al. 2007). A successful outcome of this study will be increased forecast skill of SEAS5 dry season precipitation forecasts, especially for the period of July, August, and September, which will promote the

use of corrected ECMWF SEAS5 output for seasonal forecasting in Indonesia. Reliable, skillful, and, especially, valuable forecasts bring benefits to many sectors. Therefore, we will also compute the potential economic value (PEV; Richardson 2000; Wilks 2001) of both the raw and bias-corrected SEAS5 precipitation forecasts, along with traditional verification measures (e.g., Brier skill score and continuous ranked probability skill score). This study is motivated by the needs of BMKG, who have been collaborating with the Ministry of Agriculture to develop seasonal forecasts that can be used in a dynamic cropping calendar (using a crop model), with a focus on rice. Rice plants can grow well in a humid month that has rainfall between 100 and 200 mm month⁻¹ (Ministry of Agriculture 2006). The motivation for this range of rainfall levels, identified by Oldeman (1980), is based on the amount of water demanded.

In section 2 of this article the EQM method is described. In section 3, the verification methods are presented. This is followed by a description of the data and study area in section 4. The results are presented in section 5 before the article concludes with a discussion of the findings in section 6.

2. Empirical quantile mapping

BC is one of the approaches that is commonly used to postprocess the direct output of a numerical model. EQM is a popular BC method and works with empirical probability density functions (PDFs) or CDFs for both the forecasts and the observations. This is a statistical approach based on the application of quantile–quantile adjustment: estimating quantiles for both the forecast and the observation dataset, then forming a transfer function by using corresponding quantile values. This aims to match the quantile of the forecast value to the observed value at the same quantile. To achieve this purpose, each predicted quantile is substituted by the corresponding observed quantile by means of their empirical cumulative distribution functions (ECDFs). Furthermore, this transfer function is then applied to the forecast data series:

$$Y_{\text{fc(bc)}} = \text{ECDF}_{\text{obs}}^{-1}[\text{ECDF}_{\text{fc}}(Y_{\text{fc}})], \quad (1)$$

where Y_{fc} is the raw precipitation forecast from SEAS5, and $Y_{\text{fc(bc)}}$ is the bias-corrected precipitation reforecast. $\text{ECDF}_{\text{obs}}^{-1}$ is the inverse ECDF of the observations, and ECDF_{fc} is the ECDF of the forecast values (for each ensemble member separately), both determined for the training period.

The number of quantiles is a free parameter. This method differs from parametric quantile mapping, which

uses the assumption that both the observed and the simulated precipitation distribution are well approximated by, for example, a gamma distribution (Piani et al. 2010). In our study, there is no fitting of a theoretical probability distribution to the data in creating the CDFs and no a priori assumptions are made (Gudmundsson et al. 2012). Reichle and Koster (2004) and Déqué (2007) already used this method. Reichle and Koster (2004) corrected surface soil moisture data from different sources such as satellite, ground measurements, and land modeling. Besides, Déqué (2007) corrected the extreme cold temperatures and summer heavy precipitation that are underestimated by the Météo-France atmospheric model ARPEGE. The major advantage of the EQM method is that it adjusts all moments, such that the entire distribution matches that of the observations for the training period, while maintaining the rank correlation between models and observations (Li et al. 2010). In this technique, the quantiles are determined by sorting model output and observations for the same historical base period, and constructing CDFs for each. Because daily input is needed to feed hydrological and crop models, for each month, lead month, and grid cell separately, we apply the BC at the daily level. We use a leave-three-years-out cross-validation technique to estimate the forecast error.

Measures of performance are evaluated using an independent sample that was not included in the training dataset. The total length of the record N is 30 years. In m -fold cross validation, m years are repeatedly removed, in this case $m = 3$. Subsequently, the EQM method is trained on each 27-yr period and applied to the three independent years. Finally, all independent years can be used in the verification procedure, so $N = 30$ years for the verification sample. To check robustness of the results to a different choice of m , we tested using a leave-one-year-out cross validation ($m = 1$), which showed almost identical results, probably because EQM is applied to daily accumulated precipitation. All results in section 5 are shown for $m = 3$.

3. Verification methods

Forecast skill refers to the relative accuracy of a set of forecasts with respect to some standard reference forecast (Wilks 2011). This study aims to investigate whether the EQM method improves the skill of the raw ECMWF seasonal precipitation forecasts in Java compared to a climatological forecast.

The precipitation forecast takes values on a continuous scale. In addition, it comes in the form of an ensemble forecast. Therefore, an appropriate and proper verification metric is the continuous ranked probability

skill score (CRPSS; e.g., Wilks 2011). The CRPSS compares the continuous ranked probability score (CRPS) of a (raw or BC) forecast to that of a reference forecast. Both forecasts and observation are expressed by CDFs.

The CRPS compares the distribution of the ensemble members with the observation. Briefly, this score is the total area between the CDF of the forecast and the CDF of the observation. In the computation of the CRPSS, the observation climatology is used as a reference forecast. The CRPS is defined as follows (e.g., Wilks 2011):

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dy, \quad (2)$$

where

$$F_o(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases}.$$

$F(y)$ denotes the forecast distribution expressed as CDF and $F_o(y)$ is a step function CDF that jumps from 0 to 1 at the observed value. Equation (2) defines the CRPS for a single forecast, but CRPS values for multiple cases are often averaged (Hersbach 2000; Wilks 2011). A perfect CRPS is 0.

The associated skill score (CRPSS) is defined as follows (e.g., Wilks 2011):

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{\text{ref}}}, \quad (3)$$

where CRPS_{ref} stands for the CRPS of the reference, that is, climatology. The range of CRPSS is $-\infty$ to 1. A positive skill score ($\text{CRPSS} > 0$) indicates that the forecast improves on the climatological forecast, while it is worse than climatology when $\text{CRPSS} < 0$. The CRPSS is used to verify both the raw and bias-corrected SEAS5 forecasts. In this verification, the daily rainfall data were aggregated to monthly accumulated rainfall. The CRPSS values in this study are averaged over all 30 years.

Another verification tool is the reliability (or attributes) diagram for assessing reliability and resolution. Reliability diagrams are common verification tools to illustrate the properties of probabilistic forecast systems. They are graphs of the observed frequency of an event plotted against the forecast probability of an event (e.g., Wilks 2011). From these diagrams, users can simply understand how often each forecast probability actually occurred. For example, if the forecast probability of an event equals 50% then, for perfect reliability, the event should occur on 50% of occasions on which the forecast is made. Reliability and resolution of probabilistic forecasts are separate terms in the decomposition of the Brier score. Resolution measures the ability of a forecast

to distinguish situations with distinctly different frequencies of occurrence (Wilks 2011). The reliability diagrams in this study are shown for events that exceed the 100-mm threshold; observed frequencies are plotted as a function of forecast probabilities for all model grid cells over Java over the 30-yr verification period. If the reliability curves have positive slopes, it indicates that the observed frequency of the event increases as a function of the forecast probability. In case of perfect reliability the diagram shows a reliability curve equal to the diagonal line.

We also evaluate forecast skill using the Brier skill score (BSS), based on the Brier score (BS):

$$\text{BS} = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2, \quad (4)$$

where N is the number of events, f_t is the forecast probability, and o_t is the binary observation (1 if the event occurred, and 0 if it did not). The BS is converted to a skill score, the BSS, using the BS of a reference forecast (i.e., climatology):

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}}. \quad (5)$$

BSS can range from $-\infty$ to 1, where 1 is a perfect forecast and all negative values indicate a forecast with negative skill compared to climatology.

Further, investigation of the PEV of the SEAS5 precipitation forecasts is also conducted in this study, estimated using basic cost/loss ratio (C/L) situations (Richardson 2000; Wilks 2001). The PEV of a forecasting system can be interpreted as the economic gain (relative to climatology) obtained by performing an action or nonaction, depending on the forecast. Decision-makers have a number of courses of action to choose from, and the choice is to some extent influenced by the forecast. A cost is associated with a preventive action. If no action is taken, there might be a loss if the adverse event occurs (Richardson 2000). Decision-makers can alter their actions based on forecast information. In probability forecast cases, users have to decide on the probability threshold at which to take action. In the simplest cost/loss model, users should take action if the forecast probability $> C/L$. The use of probabilistic forecasts in the cost/lost decision framework has the advantage that it has the potential to improve farmers' decision-making more than deterministic forecasts because probabilistic forecasts are able to provide users with more complete information about future scenarios. The cost/loss model gives a simple illustration of the way that probability forecasts can be

TABLE 1. Contingency table showing the matrix of hits (H), misses (M), false alarms (FA), and correct rejections (CR), and their associated costs (C) and losses (L).

		Forecast/action	
		Yes	No
Observed	Yes	Hit (H) $\rightarrow C$	Miss (M) $\rightarrow L$
	No	False alarm (FA) $\rightarrow C$	Correct rejection (CR)

used in such situations and the benefits that may be obtained. Table 1 shows a contingency table that indicates the costs and losses accrued by the use of forecasts, depending on forecast and observed events. PEV is defined as follows:

$$\text{PEV} = \begin{cases} \frac{(C/L)(H + \text{FA} - 1) + M}{(C/L)(\bar{o} - 1)}, & \text{if } C/L < \bar{o} \\ \frac{(C/L)(H + \text{FA}) + M - \bar{o}}{\bar{o}[(C/L) - 1]}, & \text{if } C/L \geq \bar{o} \end{cases}, \quad (6)$$

where H = hits, FA = false alarms, M = misses, and $\bar{o}$ = climatological frequency.

PEV ranges between $-\infty$ to 1. A forecast that is better than the climatological forecast has positive PEV, and a perfect forecast has PEV = 1. Maximum PEV occurs when the cost/lost ratio is equal to the observed frequency of the event; then it equals the hit rate minus the false alarm rate, which is equal to the true skill statistic (TSS) and is defined as follows (e.g., Wilks 2011):

$$\text{TSS} = \frac{(H \times \text{CR} - M \times \text{FA})}{(H + M)(\text{FA} + \text{CR})}, \quad (7)$$

where CR = correct rejections. The TSS measures the ability of the forecast to distinguish between occurrences and nonoccurrences of the event. The score can vary between -1 and 1 , where -1 is the worst value.

As is common (e.g., Richardson 2000), we show the potential economic value as a function of the cost/loss ratio. This allows a specific user to see the potential economic value for their specific C/L ratio. Most users have low cost/loss ratios, as the costs of taking preventative action are generally relatively low, while the losses incurred can be relatively large.

4. Data

a. Study region

The present study focuses on Java in Indonesia. Figure 1 shows the real and model topography. The latter is based on geopotential data from ECMWF with a resolution of 0.4° that are divided by the acceleration of gravity to get the approximate elevation. There is substantial mismatch between the topography of the model and that of the real world (as captured by a 1-km digital elevation model), which is a result of the much lower model resolution. Variation in regional rainfall in Java is partly a consequence of the presence of very high volcanoes and mountains, with the highest mountain being Mahameru, or Great Mountain, whose height is 3676 m above sea level. Rice production is heavily concentrated on this island where rainfall and/or irrigation is very important for the growth of the plant.

b. Forecast data

This study uses seasonal precipitation reforecasts from SEAS5 for the period 1981–2010. SEAS5 was introduced in the autumn of 2017 replacing S4, which was released in 2011. SEAS5 issues a 7-month forecast, on the fifth of each month at 1200 UTC. The initial date of the forecast is the first of each month. The reforecast data used in this study are total precipitation (TP) in meters of equivalent water in the last 24 h, with a 36-km grid resolution. SEAS5 also consists of a set of

FIG. 1. (left) Real elevation and (right) approximate ECMWF SEAS5 model elevation (m) on Java. Data sources: (left) Digital Elevation Model (DEM) from The Indonesian National Coordinating Agency for Surveys and Mapping (BAKOSURTANAL) and (right) approximate ECMWF SEAS5 model elevation calculated from geopotential data.

FIG. 2. (a) Monthly area-averaged precipitation climatology (1981–2010) in Java based on ECMWF SEAS5 (1-month lead time; red) and SA-OBS (blue) and (b) scatterplot of SA-OBS vs ECMWF SEAS5 ensemble mean (1-month lead time) monthly area-averaged precipitation in Java for 1981–2010 with the red line indicating the diagonal.

retrospective seasonal forecasts for past dates, which are called reforecasts (also sometimes known as hindcasts) that can be compared to the historical record. The SEAS5 reforecast ensemble contains 25 members, while the operational SEAS5 ensemble contains 51 members.

ECMWF's Integrated Forecasting System (IFS) represents uncertainty both in the initial conditions and in the model physics, the latter by perturbing the physics tendencies. Stochastic kinetic energy backscatter (SKEB) and stochastically perturbed physical tendency (SPPT) schemes are applied to all ensemble members of SEAS5 (Johnson et al. 2019). More information about SEAS5 and its configuration can also be found in Johnson et al. (2019; see also <https://www.ecmwf.int/forecasts/documentation-and-support/long-range>).

c. Observations

To correct the bias, a suitable dataset of precipitation observations is required. This study uses daily precipitation amounts for the period 1981–2010 from a new high-resolution (0.25°) observational dataset, called the Southeast Asia observations (SA-OBS) dataset (Van den Besselaar et al. 2017). SA-OBS is a high-resolution

land-only gridded dataset for the Southeast Asian region for daily precipitation amount along with daily minimum, mean, and maximum temperature (Van den Besselaar et al. 2017). The gridded dataset SA-OBS is part of the Southeast Asian Climate Assessment and Dataset (SACA&D), that contains daily station and gridded observations for Southeast Asia (Van den Besselaar et al. 2017). The gridding method used to create this dataset involves kriging using a geographically independent variogram. This method solves a set of linear equations to minimize the variance of the observations around the interpolating surface. Further information about the gridding methodology and a comparison of the SA-OBS dataset to other observational datasets (both from rain gauges and satellites) can be found in Van den Besselaar et al. (2017). [The SA-OBS dataset can be downloaded from <http://sacad.database.bmkg.go.id/download/grid/download.php>. It is introduced by Van den Besselaar et al. (2017).] Matching the regular latitude-longitude SA-OBS grid to the ECMWF grid is applied before further analysis. The weighted average of all SA-OBS 0.25° grid cells in each corresponding ECMWF 36-km grid cell is computed. There are 18×47

FIG. 3. Monthly average observed precipitation (mm; 1981–2010 climatology) of (a) July, (b) August, and (c) September based on the SA-OBS dataset.

grid cells of SA-OBS and 14×40 grid cells of ECMWF. A weighting factor is applied where a weight is given to every single SA-OBS grid cell based on the fraction of it in the corresponding ECMWF grid cell.

5. Results

a. Rainfall patterns and characteristics over Java

The following factors affect rainfall variability in Indonesia (Aldrian and Susanto 2003): topography,

tropical cyclones, meridional (Hadley) circulation, zonal (Walker) circulation, El Niño–Southern Oscillation (ENSO; e.g., Hendon 2003), the Madden–Julian oscillation (MJO; Madden and Julian 1994), and the activity of the monsoon. The monsoon system is a very dominant factor that controls the climate of Java, as Aldrian and Susanto (2003) described. In fact, there are two monsoon systems that have significantly different precipitation characteristics influencing Java. From November to March (NDJFM) the climate in this area is influenced by the northwest monsoon, which leads to rainy conditions. The monthly area-averaged rainfall climatology during this season ranges from about 210 to 330 mm month^{−1} (Fig. 2a). Figure 2a shows the monthly area-averaged precipitation climatology of 1981–2010 based on the SA-OBS dataset and the SEAS5 reforecasts (1-month lead time) in Java. On the other hand, during the period from May to September (MJJAS), the climate of Java is influenced by the southeast monsoon, which leads to relatively dry conditions. The monthly area-averaged rainfall climatology in the dry period is only between 30 and 120 mm month^{−1} (Fig. 2a). The graph clearly shows the seasonal variations in rainfall over this area as a result of the monsoon, both in the observations and in the SEAS 5 reforecasts, but the latter has a tendency to overestimate the precipitation. This is confirmed by the scatterplot of SA-OBS versus ensemble mean SEAS5 monthly area-averaged precipitation in Fig. 2b.

Figure 3 shows precipitation climatology maps in July, August, and September. The patterns show that the average monthly precipitation of these months is mostly below 125 mm month^{−1}. The southwest part of Java is wetter than the northeast because of its mountainous geography (Fig. 1; Van der Eng 2010).

b. Verification of raw and bias-corrected seasonal precipitation reforecasts

As mentioned in section 2 the EQM method is applied to daily SEAS5 precipitation forecasts in each grid cell of the study area for each lead time, month, and ensemble member separately. Figure 4 shows quantile–quantile (Q–Q) plots of monthly accumulated precipitation of the raw and bias-corrected ensemble mean SEAS5 reforecasts against SA-OBS precipitation. The EQM method ensures that the distribution of SEAS5 matches that of the observation on the daily scale (the scale to which the correction is applied). As mentioned before, we chose to calibrate the daily values so that they can be used as inputs to a hydrological and/or crop model. However, it is clear from Fig. 4 that the correction at the daily scale does not ensure that the quantiles are equivalent on the monthly scale. For the 1-month lead time (Fig. 4a), the bias correction using

FIG. 4. Q–Q plots of SA–OBS vs raw (red) and bias-corrected (blue) ensemble mean SEAS5 monthly precipitation for all grid cells in Java and for (a) 1-month and (b) 7-month lead time.

EQM successfully corrects the raw SEAS5 forecasts for July and August, but for September there is a wet bias for higher values of the BC reforecasts. For the 7-month lead time (Fig. 4b), both the Q–Q plots of the raw and BC reforecasts show that they are generally systematically lower than the observations (dry bias), but the distribution of reforecasts is closer to that of the observations after bias correction for August and September.

Figure 5 presents the CRPSS values (over 95% of the grid cells) of the raw and BC reforecasts, for reforecasts valid in July, August, and September, for the seven lead times. The 95% range of CRPSS values (i.e., the 2.5th–97.5th percentile range of the distribution of CRPSS values) after BC (blue shaded area) for those three months is much smaller than for the raw reforecasts (pink shaded area). In July and August (September) the median CRPSS values of lead times ≤ 2 (3) months are positive after BC (shown by blue solid lines in Fig. 5), but they are still negative for longer lead times (lead

times of 5–7 months for July, 3–7 months for August, and 4–7 months for September).

EQM generally improves CRPSS with values up to about 0.5 for July in the 1-month lead time. More positive CRPSS values appear after BC for most grid cells, as can also be seen in Fig. 6, which displays the spatial pattern of the CRPSS for that lead time. In the central and eastern parts of Java, most CRPSS values have become positive after BC, but the CRPSS has decreased and become negative in a number of grid cells in the western part of Java. This might be a result of the fact that the dry season in the eastern part of Java is more consistently dry compared to the western part. The western part of Java has higher variability in rainfall during June, July, and August. Sometimes, during the dry season there is still rainfall over this region (Satyawardhana and Susandi 2015).

Figure 7 shows the reliability diagrams of raw and BC SEAS5 reforecasts for events that exceed the $100 \text{ mm month}^{-1}$ threshold in July, August, and September.

FIG. 5. CRPSS for raw (red) and bias-corrected (blue) forecasts valid in July, August, and September. The solid lines indicate the median and the shaded areas show the CRPSS spread of the grid cells between the 0.025 and 0.975 quantiles (dotted lines).

The diagonal line should be considered as a reference for perfect reliability. In this reliability diagram, the raw SEAS5 curves (red curves in Fig. 7) are located below the perfect reliability line, indicating that the raw SEAS5 overforecasts the event (i.e., forecasts the event with higher probability than the observed frequency). In July and August, the reliability curves are generally below the “no skill” line and exhibit little resolution. In September, the raw forecasts have better reliability.

For those three months the reliability curves of the BC reforecasts (blue curves in Fig. 7) show an improvement over the raw forecasts. The curves mostly lie within the skillful region and display more reliability than the raw forecasts, with very good reliability in July (Fig. 7a). Furthermore, for those three months, the reliability curves of the BC reforecasts for April–June, and October and November also improve over the raw forecasts (not shown).

BSS values (over 95% of the grid cells) of the raw and BC reforecasts as a function of the threshold are shown in Fig. 8a for 1-month lead time in July, August, and September. The 95% range of BSS values after BC (blue shaded area in Fig. 8) is generally smaller than for the raw forecasts (pink shaded area). In July (August) the median BSS values of thresholds less than 200 (100) mm are positive after BC (shown by blue solid lines in Fig. 8), but they are still negative for higher thresholds, some of them even showing worse skill than the raw forecasts (shown by the red solid lines in Fig. 8).

However, in September the median BSS values of the BC reforecasts are positive up to high thresholds.

In addition, Fig. 8b shows the BSS for the 100-mm threshold over lead time. In July and September, the median BSS values of the BC reforecasts are positive for the 1- to 4-month lead times, while in August they are positive only for the 1- and 2-month lead times. For the forecasts that are initialized in April (June), the skill of the corrected monthly forecasts for July (September) is positive for most of the grid cells.

c. Potential economic value

Figure 9 shows the PEV of the raw and bias-corrected SEAS5 forecasts for 1- and 7-month lead time for July, August, and September at the 100-mm threshold. Positive PEV indicates that there is more PEV than climatology. It is clear from Fig. 9 that the PEV of SEAS5 forecasts varies considerably for users with different cost/loss ratios, with the largest PEV found for low C/L ratios. This is good to see as the majority of users generally have low cost/loss ratios (i.e., the cost of taking preventative action is generally relatively small and the potential losses are relatively large). It is important to note that all users get more potential value from the bias-corrected SEAS5 forecasts compared to the raw SEAS5 forecasts. We can see that the bias-corrected SEAS5 forecasts have more value than the raw forecasts for a wider range of users at the 1-month lead time in all 3 months. For example, in the 1-month lead-time

FIG. 6. CRPSS maps of the (a)–(c) raw and (d)–(f) bias-corrected forecasts for the 1-month lead time, verifying in (from top to bottom) July, August, and September.

forecast for August, the raw forecast has more PEV than climatology (positive numbers) for users that have C/L ratios between 0.1 and 0.5, while the bias-corrected forecast provides positive PEV for users with a C/L ratio up to 0.8. This means that more users can make better decisions using the bias-corrected SEAS5 forecast compared to the raw forecast, using a simple cost/loss model.

Comparing the PEV for the 1-month lead time to that for the 7-month lead time, it is clear that the longer the

lead time, the lower the potential economic value, as expected. However, it is promising to see that there is positive PEV even at the longest lead time (7 months), and that there is still an increase in PEV after BC for cost/loss ratios around the climatological frequency. At the longest lead time, the bias-corrected forecasts are still able to provide positive PEV for a wider range of users, although the associated range of cost/loss ratios is more narrow than for the shortest lead time. When we

FIG. 7. Attributes (or reliability) diagrams of the raw (red) and bias-corrected (blue) forecasts for the 100-mm threshold at lead month +1 in (a) July, (b) August, and (c) September. The gray area in the reliability diagram indicates the region where the contribution to the BSS (based on a no-skill climatological reference) is positive. The dashed horizontal line is referred to as the “no resolution” line. Points from the reliability curve on this line have no resolution. In addition, a line halfway between the no resolution and perfect reliability (diagonal) lines is referred to as the “no-skill” line, because points on it do not add to the BSS. Numbers indicate the fraction of forecasts in each probability bin.

look at Fig. 9, the 7-month forecast has less value than the 1-month forecast up to relatively high cost/loss ratios but it becomes zero or only slightly negative for the highest ratios. This is contrary to the 1-month forecast, which has negative PEV for higher C/L ratios. Because the 7-month forecast has no skill the PEV is closer to climatology. On the other hand, the 1-month forecast has skill because of reliable and sharper probability forecasts but this apparently leads to negative PEV for higher C/L ratios.

6. Discussion and conclusions

The results in this study show the skill and potential economic value of SEAS5 raw and bias-corrected precipitation reforecasts for July–September in Java. For the dry season (May–November), the bias-corrected SEAS5 forecasts show positive CRPSS values (with climatology as a reference) for most grid cells for 1-month lead time and for July–October also 2-month lead time. In July and September, many grid cells are skillful to the 3-month lead time (Fig. 5). For the 1-month lead time, the CRPSS values of the bias-corrected forecasts for most of the grid cells located in the central and eastern part of Java are positive (Fig. 6). The western part of Java shows less positive or even negative CRPSS values probably because of its high rainfall variability. The rainfall variability is caused by several factors, including topography. Figure 1 shows that the topographical conditions of western Java vary, with low land plains and beaches in the north, and high land plains and a mountain range in the south. Note that both the raw and the bias-corrected SEAS5 forecasts are not skillful compared to climatology for most of the wet season (December–April; not shown in this paper).

The BSS values of the bias-corrected SEAS5 forecasts are lower in August compared to July and September, maybe because August is mostly the peak of the dry season. In July and August at the threshold of 50 mm, the median BSS reaches its maximum value (nearly 0.5 in July) and the range of BSS values is almost entirely positive. The bias-corrected forecasts still show negative BSS for some grid cells and especially the higher thresholds, but the BSS is generally higher than for the raw forecasts. In September the threshold with a range of almost entirely positive BSS values is close to 100 mm, where the median BSS is nearly 0.3.

In addition, BC by EQM is not only able to improve forecast skill (as defined by the BSS and CRPSS), but also enhances the potential economic value of the SEAS5 forecasts, especially in the dry season, which is an agriculturally important season. It is shown in Fig. 9

FIG. 8. BSS for raw (red) and bias-corrected (blue) forecasts valid in July, August, and September (a) as a function of threshold for 1-month lead time and (b) for probabilistic forecasts of precipitation $\geq 100 \text{ mm month}^{-1}$, over forecast lead months 1 to 7. The solid lines indicate the median and the shaded areas show the BSS spread of the grid cells between the 0.025 and 0.975 quantiles (dotted lines).

that the corrected SEAS5 forecasts have a higher PEV than the raw forecasts for all ranges of users; or, in other words, the bias-corrected SEAS5 forecasts are more valuable than the raw forecasts, even for the longest lead time of 7 months. During the peak of the wet season (December–February), however, the bias-corrected

SEAS5 forecasts cannot improve upon the negative PEV of the raw forecasts (not shown).

BC using the EQM method has some limitations. For example, it cannot efficiently correct overconfidence in the raw ensemble forecasts (Zhao et al. 2017), and downscaling using quantile mapping can produce outputs

FIG. 9. Potential economic value (PEV) of raw (red) and bias-corrected (blue) forecasts as a function of cost/loss ratio for (a)–(c) 1-month and (d)–(f) 7-month lead times, 100-mm threshold, in July, August, and September. The dashed line indicates the climatological frequency.

that have larger variance than observations when the outputs are rescaled to the original resolution (Maraun 2013). Besides, forecasts with both positive and negative skill are generated because EQM does not consider the strength of the correlation between forecasts and observations (Zhao et al. 2017). Maraun (2013) also mentioned another limitation of quantile mapping: it does not introduce any small-scale variability. This means that the temporal structure is still that of the grid box and not of the local scale. Despite these limitations, a number of studies (e.g., Piani et al. 2010; Mehrotra and Sharma 2016) have shown that quantile mapping is effective for correcting the bias of climate projections. Crochemore et al. (2016) mention that quantile mapping can be useful if the raw forecasts are only biased, and neither under nor overdispersed. A benefit of the EQM method is that it makes no prior assumption about the shape of the probability density function. This makes the distribution of corrected values identical to that of the observed data for the training period.

Besides the limitations of the method, the reader should keep in mind that there is also uncertainty in the SA-OBS data (Van den Besselaar et al. 2017). Sources of the observational uncertainty stem from uncertainty in the data quality, variability in the number of stations, the interpolation technique, and uncertainty in the metadata, as it comes from different sources. Also, because of the variations in topography and the limited number of stations the large-scale patterns are probably well defined, but small-scale features will likely be more uncertain.

Finally, although every statistical postprocessing method has its limitation(s) in its capacity to correct systematic errors in ensemble forecast data, it is a computationally cheap and effective way to improve forecast skill. In a future study, more advanced statistical postprocessing methods will be applied to further improve seasonal forecast skill.

Acknowledgments. This study was part of the G4INDO project, which was mostly funded by the Netherlands Space Office (NSO). The authors are grateful to the Santander met group for developing and maintaining the downscaleR package (<https://github.com/SantanderMetGroup/downscaleR>), and to the three anonymous reviewers and Albert Klein Tank (WUR) for their comments on an earlier version of the manuscript, which have helped to improve it. We also thank Gerard van der Schrier (KNMI) and Robi Muharsyah (BMKG) for assistance with the observational data and Folmer Krikken (KNMI) for the geopotential ECMWF model data.

REFERENCES

- Aldrian, E., and R. D. Susanto, 2003: Identification of three dominant rainfall region within Indonesia and their relationship to sea surface temperature. *Int. J. Climatol.*, **23**, 1435–1452, <https://doi.org/10.1002/joc.950>.
- Amengual, A., V. Homar, R. Romero, S. Alonso, and C. Ramis, 2012: A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. *J. Climate*, **25**, 939–957, <https://doi.org/10.1175/JCLI-D-10-05024.1>.
- BMKG, 2018: Dry seasonal forecasting 2018 for Indonesia. BMKG, 138 pp., https://cdn.bmkg.go.id/web/Release_Buku_PMK_2018.pdf.
- Boé, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol.*, **27**, 1643–1655, <https://doi.org/10.1002/joc.1602>.
- Crochemore, L., M. Ramos, and F. Pappenberger, 2016: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecast. *Hydrol. Earth Syst. Sci.*, **20**, 3601–3618, <https://doi.org/10.5194/hess-20-3601-2016>.
- Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global Planet. Change*, **57**, 16–26, <https://doi.org/10.1016/j.gloplacha.2006.11.030>.
- Di Giuseppe, F., F. Molteni, and A. M. Tompkins, 2013: A rainfall calibration methodology for impacts modelling based on spatial mapping. *Quart. J. Roy. Meteor. Soc.*, **139**, <https://doi.org/10.1002/qj.2019>.
- Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, 2012: HESS opinions: “Should we apply bias correction to global and regional climate model data?” *Hydrol. Earth Syst. Sci.*, **16**, 3391–3404, <https://doi.org/10.5194/hess-16-3391-2012>.
- Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>.
- Hendon, H. H., 2003: Indonesian rainfall variability: Impacts of ENSO and local air–sea interaction. *J. Climate*, **16**, 1775–1790, [https://doi.org/10.1175/1520-0442\(2003\)016<1775:IRVIOE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<1775:IRVIOE>2.0.CO;2).
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Iizumi, T., M. Nishimori, K. Dairaku, S. A. Adachi, and M. Yokozawa, 2011: Evaluation and intercomparison of downscaled daily precipitation indices over Japan in present-day climate: Strengths and weaknesses of dynamical and bias correction-type statistical downscaling methods. *J. Geophys. Res.*, **116**, D01111, <https://doi.org/10.1029/2010JD014513>.
- Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>.
- Joslyn, S. L., and J. E. LeClerc, 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol. Appl.*, **18**, 126–140, <https://doi.org/10.1037/a0025185>.
- Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dyn.*, **39**, 2957–2973, <https://doi.org/10.1007/s00382-012-1364-6>.

- Komalasari, K. E., Y. Fajariana, T. A. Nuraini, and R. Anggraeni, 2016: Application of ensemble mean to improve reliability of HyBMKG forecast. *J. Meteor. Geophys.*, **17** (1), 47–52, <http://puslitbang.bmkg.go.id/jmg/index.php/jmg/article/view/381>.
- Lafon, T., S. Dadson, G. Buys, and C. Prudhomme, 2013: Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods. *Int. J. Climatol.*, **33**, 1367–1381, <https://doi.org/10.1002/joc.3518>.
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, <https://doi.org/10.1029/2009JD012882>.
- Madden, R. A., and P. R. Julian, 1994: Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.*, **122**, 814–837, [https://doi.org/10.1175/1520-0493\(1994\)122<0814:OOTD0>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0814:OOTD0>2.0.CO;2).
- Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>.
- Marjuki, G. van der Schrier, A. M. G. Klein Tank, E. J. M. van den Besselaar, Nurhayati, and Y. S. Swarinto, 2014: Observed climatic trends and variability relevant for crop yields in Southeast Asia. *Proc. Int. ASEAN SACA&D Conf. and Workshop 2014 (IASCW-2014): Past, Present and Future; A Display of Climate Science and Services in South East Asia*, Bogor, Indonesia, BMKG and KNMI, 22–23, http://sacad.database.bmkg.go.id/documents/IASCW_Proceedings.pdf.
- Mehrotra, R., and A. Sharma, 2016: A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: Implications for downscaling. *J. Climate*, **29**, 3519–3539, <https://doi.org/10.1175/JCLI-D-15-0356.1>.
- Ministry of Agriculture, 2006: Antisipasi produksi pangan dari ancaman kekeringan. *Warta Penelitian dan Pengembangan Pertanian*, **28** (24), 4–6, <http://203.190.37.42/publikasi/wr284062.pdf>.
- Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate*, **12**, 2256–2284, [https://doi.org/10.1175/1520-0442\(1999\)012<2256:AEOSAD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2).
- Naylor, R., D. S. Battisti, D. J. Vimont, W. P. Falcon, and M. B. Burke, 2007: Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proc. Natl. Acad. Sci. USA*, **104**, 7752–7757, <https://doi.org/10.1073/pnas.0701825104>.
- Oldeman, L. R., 1980: The agroclimatic classification of rice growing environments in Indonesia. *Proc. Symp. on the Agrometeorology of the Rice Crop*, Los Baños, Philippines, World Meteorological Organization and the International Rice Research Institute, 47–55.
- Piani, C., O. Haerter, and E. Corpola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Ramos, M. H., S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, **17**, 2219–2232, <https://doi.org/10.5194/hess-17-2219-2013>.
- Rayner, S., D. Lach, and H. Ingram, 2005: Weather forecasts are for wimps: Why water resource managers do not use climate forecasts. *Climatic Change*, **69**, 197–227, <https://doi.org/10.1007/s10584-005-3148-z>.
- Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, <https://doi.org/10.1029/2004GL020938>.
- Richardson, D. S., 2000: Skill and economic value of the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–668, <https://doi.org/10.1002/qj.49712656313>.
- Salvi, K., S. Kannan, and S. Ghosh, 2011: Statistical downscaling and bias correction for projections of Indian rainfall and temperature in climate change studies. *2011 International Conference on Environmental and Computer Science*, Vol. 19, IACSIT Press, 7–11.
- Satyawardhana, H., and H. Susandi, 2015: Season onset projection in Java based on Conformal Cubic Atmospheric Model (CCAM) downscaling output. *J. Sains Dirgantara*, **13** (1), 1–14, http://jurnal.lapan.go.id/index.php/jurnal_sains/article/view/2311/2062.
- Schepen, A., Q. J. Wang, and D. E. Robertson, 2014: Seasonal forecast of Australian rainfall through calibration and bridging of coupled GCM outputs. *Mon. Wea. Rev.*, **142**, 1758–1777, <https://doi.org/10.1175/MWR-D-13-00248.1>.
- , —, and Y. Everingham, 2016: Calibration, bridging and merging to improve GCM seasonal temperature forecasts in Australia. *Mon. Wea. Rev.*, **144**, 2421–2441, <https://doi.org/10.1175/MWR-D-15-0384.1>.
- Teutschbein, C., and J. Seibert, 2010: Regional climate models for hydrological impact studies at the catchment scale: A review of recent modeling strategies. *Geogr. Compass*, **4**, 834–838, <https://doi.org/10.1111/j.1749-8198.2010.00357.x>.
- Trambauer, P., M. Werner, H. C. Winsemius, S. Maskey, E. Dutra, and S. Uhlenbrook, 2015: Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa. *Hydrol. Earth Syst. Sci.*, **19**, 1695–1711, <https://doi.org/10.5194/hess-19-1695-2015>.
- Van den Besselaar, E. J., G. van der Schrier, R. C. Cornes, A. S. Iqbal, and A. M. Klein Tank, 2017: SA-OBS: A daily gridded surface temperature and precipitation dataset for Southeast Asia. *J. Climate*, **30**, 5151–5165, <https://doi.org/10.1175/JCLI-D-16-0575.1>.
- Van der Eng, P., 2010: Market responses to climate stress: Rice in Java in the 1930s. *Aust. Econ. Hist. Rev.*, **50**, 62–79, <https://doi.org/10.1111/j.1467-8446.2009.00272.x>.
- Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, <https://doi.org/10.1098/rsif.2013.1162>.
- Wetterhall, F., H. C. Winsemius, E. Dutra, M. Werner, and E. Pappenberger, 2015: Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin. *Hydrol. Earth Syst. Sci.*, **19**, 2577–2586, <https://doi.org/10.5194/hess-19-2577-2015>.
- Wilcke, R., T. Mendlik, and A. Gobiet, 2013: Multi-variable error correction of regional climate models. *Climatic Change*, **120**, 871–887, <https://doi.org/10.1007/s10584-013-0845-x>.
- Wilks, D. S., 2001: A skill score based on economic value for probability forecasts. *Meteor. Appl.*, **8**, 209–219, <https://doi.org/10.1017/S1350482701002092>.
- Zhao, T., J. C. Bennett, Q. J. Wang, A. Schepen, A. W. Wood, D. E. Robertson, and M. Ramos, 2017: How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J. Climate*, **30**, 3185–3196, <https://doi.org/10.1175/JCLI-D-16-0652.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.