Bias-correcting individual inputs prior to combined calibration leads to more skillful forecasts of reference crop evapotranspiration

Qichun Yang[1], Quan J Wang[1], Kirsti Hakala[1], Yating Tang[1]

[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Australia

5   *Correspondence to*: Qichun Yang (qichun.yang@unimelb.edu.au)

Abstract. Reference crop evapotranspiration (ETo) is calculated using a standard formula with temperature, vapor pressure, solar radiation, and wind speed as input variables. ETo forecasts can be produced when forecasts of these input variables from numerical weather prediction (NWP) models are available. As raw ETo forecasts are often subjective to systematic errors, calibration is necessary for improving forecast quality. The most straightforward and widely used approach is to directly

10   calibrate raw ETo forecasts constructed with the raw forecasts of input variables. However, the potential predictability of ETo may not be fully explored by this approach, which ignores the non-linear interactions of input variables in constructing ETo forecasts. We hypothesize that reducing errors in individual inputs as a precursor to ETo forecast calibration will lead to more skillful ETo forecasts. To test this hypothesis, we evaluate two calibration strategies, including i) calibration directly applied to raw ETo forecasts constructed with raw forecasts of input variables, and ii) bias-correcting input variables first, and then

15   calibrating the ETo forecasts constructed with bias-corrected input variables. We calibrate ETo forecasts based on weather forecasts of the Australian Community Climate and Earth System Simulator G2 version (ACCESS-G2). Calibrated ETo forecasts with bias-corrected input variables (strategy ii) demonstrate lower bias, higher correlation coefficient, and higher skills than the calibration based on raw input variables (strategy i). This investigation indicates that improving raw forecasts of input variables could enhance ETo forecast calibration and produce more skillful ETo forecasts. This calibration strategy is

20   expected to enhance future NWP-based ETo forecasting.

1 Introduction

As a variable measuring the evaporative demand of the atmosphere, reference crop evapotranspiration (ETo) has been widely used to estimate potential water loss from the land surface to the atmosphere (Hopson and Webster, 2009; Liu et al., 2019; Renard et al., 2010). Quantification of ETo has been increasingly performed to support efficient water use and water

25   management (Mushtaq et al., 2019; Perera et al., 2016). Forecasts of short-term ETo (days to weeks) are highly valuable for real-time decision-making on farming activities and water allocation to competing users (Djaman et al., 2018; Kumar et al., 2012).

A plethora of methods with divergent statistical assumptions, dependence on observations, and requirements of weather forecasts have been developed to predict future ETo (Bachour et al., 2016; Ballesteros et al., 2016; Karbasi, 2018; Mariito et

30   al., 1993). Since ETo is impacted jointly by temperature, vapor pressure, solar radiation, and wind speed (Bachour et al., 2016;
      Luo et al., 2014), the inclusion of these input variables in statistical forecasting models could enable better representations of
      physical processes regulating ETo and often improves ETo predictions (Torres et al., 2011). The increasing availability of
      weather and climate forecasts based on numerical models has opened up new opportunities for ETo forecasting (Cai et al.,
      2007; Srivastava et al., 2013; Tian and Martinez, 2014; Zhao et al., 2019a). Forecasts of temperature, vapor pressure, solar

35   radiation, and wind speed from numerical weather prediction (NWP) models/global climate models (GCMs) could be
      conveniently translated into ETo forecasts using the Food and Agriculture Organization (FAO) ETo equation (Allen et al.,
      1998; Cai et al., 2007).

      Despite the advantages in modeling physical processes of the atmosphere, flexibility in temporal and spatial scales (Pelosi et
      al., 2016), and high data availability (Er-Raki et al., 2010), NWP/GCM-based raw ETo forecasts often demonstrate systematic

40   errors (Turco et al., 2017). Limitations in schemes of NWP/GCM models, parameterization, and data assimilation often lead
      to biases in raw forecasts of weather variables (Lim and Park, 2019; Vogel et al., 2018), with forecast skills only limited to a
      few days in the future. As a result, raw ETo forecasts calculated directly with the raw forecasts of input weather variables (e.g.,
      temperature, vapor pressure, solar radiation, and wind speed) typically demonstrate substantial inconsistencies with
      observations (Medina and Tian, 2020; Zhao et al., 2019a), and need to be post-processed to improve forecast quality.

45   Unlike other weather variables directly inheriting errors from NWP models, errors in raw ETo forecasts are contributed jointly
      by multiple input variables and their complex interactions (Traore et al., 2016; Yang et al., 2019). Effective calibration aims
      to effectively correct these errors, and provide unbiased, reliable, and skillful calibrated forecasts. Theoretically, there are two
      different calibration strategies that could be adopted to achieve this goal. The first strategy is to construct raw ETo forecasts
      directly with the raw forecasts of the input variables and then calibrate the derived ETo forecasts with a post-processing model.

50   This strategy lumps errors from the input variables together in the raw ETo forecasts and then corrects combined errors directly
      (Tian and Martinez, 2014; Zhao et al., 2019a). This strategy is straightforward and thus has been adopted by most existing
      calibrations of NWP/GCM-based ETo forecasts. For example, Medina et al. (2018) used a linear regression bias-correction
      method to calibrate ETo forecasts from three NWP models and achieved significant improvements in the quality of ETo
      forecasts. Medina and Tian (2020) employed three probabilistic-based calibration methods to improve ETo forecasts of

55   multiple NWP models, and generated more skillful and reliable calibrated ETo forecasts than the calibration with a simple
      regression bias-correction model. Another probabilistic post-processing method, the Bayesian Joint Probability (BJP) model,
      was used to calibrate ETo forecasts across multiple sites in Australia and produced ETo forecasts markedly superior to raw
      forecasts in accuracy and skills (Zhao et al., 2019a, 2019b).

      Alternatively, the ETo forecast calibration could start with dealing with the errors in the raw forecasts of the input variables.

60   In this strategy, raw forecasts of input variables could be calibrated first, and raw ETo forecasts could then be constructed with
      the improved input variables. After that, the derived raw ETo forecasts could be further improved through calibration. This
      strategy is more computationally demanding than the one using uncorrected input variables. With the improved input variables,
      errors in the constructed raw ETo forecasts could be significantly reduced (Nouri and Homaee, 2018; Perera et al., 2014).

However, there is no conclusion on whether improvements to raw ETo forecasts by calibrating input variables will eventually
add additional skills to calibrated ETo forecasts, when compared with the calibration using raw forecasts of input variables to
derive raw ETo forecasts (Medina and Tian, 2020).

Which calibration strategy produces more skillful calibrated forecasts is a critical question in NWP-based ETo forecasting, but
the answer remains unclear. There is a pressing need to investigate the necessity of improving raw forecasts of the input
variables as part of ETo forecast calibration. We hypothesize that reducing errors in individual inputs as a precursor to ETo
forecast calibration will lead to more skillful ETo forecasts. To test this hypothesis, we compare two calibration strategies,
including i) directly calibrating raw ETo forecasts constructed with raw forecasts of input variables, and ii) bias-correcting raw
forecasts of each input variable first, and then calibrating the ETo forecasts constructed with bias-corrected variables. NWP
forecasts from the Australian Community Climate and Earth System Simulator G2 version (ACCESS-G2) model across
Australia are used for this investigation. Objectives of this study are to i) Calibrate ETo forecasts across Australia based on
two different strategies, ii) evaluate the contribution of bias-correcting input variables to NWP-based ETo forecasting.

2 Method

2.1 Reference data and forecasts

In this study, we use the ETo data derived from the Australian Water Availability Project (AWAP)'s gridded data of
temperature, vapor pressure, and solar radiation (Jones et al., 2007, 2014), and wind speed data developed by Mcvicar et al.
(2008), as observations for ETo forecast calibration. Weather forecasts from the ACCESS-G2 model are compiled to generate
ETo forecasts. We modify the spatial resolution of ACCESS-G2 data to match the AWAP data's grid spacing. The 3-hourly
ACCESS-G2 forecasts during 4/2016-3/2019 are aggregated to the daily scale to match the timeframe of the original site
observations used to generate the AWAP data. AWAP ETo during 4/1999-3/2019 is used for parameter inference, and data
during 4/2016-3/2019 is selected for forecast calibration and evaluation.

2.2 Calculation of ETo

We calculate ETo forecasts and AWAP ETo using the FAO56 equation (Allen, et al., 1998):

$$ET_O = \frac{0.408\Delta(R_n-G)+\gamma\frac{900}{T+273}u_2(e_s-e_a)}{\Delta+\gamma(1+0.34u_2)} \tag{1}$$

where ET_O is the reference crop evapotranspiration (mm/day); Δ is the slope of the vapor pressure curve ($kPa\ °C^{-1}$); R_n is net
radiation at the crop surface ($MJ\ m^{-2}\ day^{-1}$); G is soil heat flux density ($MJ\ m^{-2}\ day^{-1}$); γ is the psychrometric constant
($kPa°C^{-1}$); T is average air temperature ($°C$); u_2 is the wind speed at the height of 2m ($m\ s^{-1}$); e_s and e_a are saturated and
actual vapor pressure (kPa), respectively.

AWAP and Mcvicar et al. (2008) have all the variables needed for developing ETo observations. For the ETo forecasts,
temperature and solar radiation are readily available from the ACCESS-G2 outputs. To obtain the wind speed forecasts, we

first use the forecasts of zonal velocity (*u*) and the meridional (*v*) components from ACCESS-G2 to calculate wind speed
95 forecasts at 10m, and then estimate wind speed at 2m using the equation recommended by FAO (Allen, et al., 1998). In addition,
we use the ACCESS-G2 forecasts of air pressure and specific humidity to obtain the vapor pressure forecasts.

2.3 Calibration of ETo forecasts

The calibration model used in this study is the newly developed Seasonally Coherent Calibration (SCC) model, which is found
to be effective in calibrating NWP-based ETo forecasts (Yang et al., 2021b). For the calibration across Australia with a spatial
100 resolution of 0.05 degree, we process 281,655 grid cells in total. We apply the SCC model for ETo forecast calibration to each
grid cell and lead time separately.

Table 1. Four sets of ETo forecast calibrations

Calibrations	Construction of raw ETo forecasts	Application of the SCC model
Calibration 1	Raw forecasts of input variables	SCC calibration based on anomaly and climatological mean
Calibration 2	Bias-corrected forecasts of input variables	SCC calibration based on anomaly and climatological mean
Calibration 3	Raw forecasts of input variables	The SCC model applied directly to ETo forecasts
Calibration 4	Bias-corrected forecasts of input variables	The SCC model applied directly to ETo forecasts

105

We conduct four sets of calibrations to evaluate how the two different calibration strategies would affect the calibrated ETo
forecasts. Our recent investigation suggests that ETo forecast calibration based on anomaly and climatological mean produces
more skillful calibrated forecasts than calibrating ETo forecasts directly (Yang et al., 2021b). As a result, in this study, we
primarily focus on calibrations based on ETo anomalies (Calibrations 1 and 2). The comparison between Calibrations 1 and 2
110 is to investigate whether the bias-correction of input variables would further improve ETo forecasts when working with ETo
anomalies. We also conduct additional calibrations applying the SCC model to ETo forecasts directly (Calibrations 3 and 4),
to test whether the contribution of improving the input variables to ETo forecast calibration, if there is any, will depend on
how ETo forecasts are calibrated (based on anomalies vs. based on original ETo forecasts). In this study, we primarily analyze
results from Calibrations 1 and 2, and only briefly introduce results from Calibrations 3 and 4 in the main text.

115 **2.3.1 Bias-correction of input variables**

For ETo forecast calibration employing strategy ii (Calibrations 2 and 4 in Table 1), we use a non-parametric quantile mapping method (QUANT) to correct raw forecasts of the input variables. This method has been widely used in hydrological and climatological investigations for correcting bias in raw data (Boe et al., 2007). To use QUANT, we first build up the empirical cumulative density function (CDF) of both raw forecasts and AWAP data for each variable. We then calculate the percentile

120 of each record in raw forecasts in their CDF. Next, these percentiles are used to search values in the corresponding AWAP data, which is then treated as the bias-corrected forecasts. Values in between the percentiles are approximated using a linear interpolation method (Boe et al., 2007).

2.3.2 Key steps of ETo forecast calibration using the SCC model based on ETo anomaly and climatological mean

After we construct the raw ETo forecasts, based on either raw (Calibrations 1 and 3) or bias-corrected (Calibration 2 and 4)

125 forecasts of the input variables, we employ the SCC model to further calibrate the ETo forecasts. For the calibrations (Calibrations 1 and 2 shown in Table 1) applying SCC to ETo anomaly, the first step is to derive the observed daily climatological mean using the 20-year AWAP ETo. Calibrations 3 and 4 skip this step and apply the SCC model to ETo forecasts directly. We use the method developed by Narapusetty et al. (2009) and adopt trigonometric functions and harmonics to simulate the annual cycle of AWAP ETo to derive the climatological mean:

130 $$y_{cm}(t) = a_0 + \sum_{j=1}^{H}[a_j \cos(w_j t) + b_j \sin(w_j t)] \tag{2}$$

where $y_{cm}(t)$ is the climatological mean at the daily scale; H is the number of harmonics. Here we use H=4 following Narapusetty et al. (2009); a_0, a_j, and b_j are coefficients, estimated through minimizing the mean squared differences between climatological mean and observations; w_j=2πj/P, and P is days in one year.

Once we obtain the daily climatological means of the observed ETo, we then remove them from both raw forecasts and AWAP

135 ETo to generate anomalies. We calibrate the derived anomalies of raw ETo forecasts against the anomalies of AWAP ETo using the SCC model, which is composed of four key components: i) a joint probability model to characterize the connection between raw forecasts and observations, ii) reconstruction of seasonal patterns in raw forecasts based on the long-term reference data, iii) reparameterization to obtain parameters for short-archived raw forecasts, and iv) generation of calibrated forecasts with the parameters and the joint model. The SCC model has been introduced in detail in our site- and continental-

140 scale calibration of NWP precipitation and ETo forecasts (Wang et al., 2019; Yang et al., 2021a, 2021b).

In this study, we use the Yeo-Johson transformation method to transform the anomalies of forecasts and reference data to approach a normal distribution (Yeo and Johnson, 2000):

$$\hat{x} = \begin{cases} (\lambda x + 1)^{\frac{1}{\lambda}} - 1, & (x \geq 0, \lambda \neq 0) \\ exp(x) - 1, & (x \geq 0, \lambda = 0) \\ -(\lambda - 2)x + 1)^{\frac{1}{2-\lambda}} + 1, & (x < 0, \lambda \neq 2) \\ -exp(-x) + 1, & (x < 0, \lambda = 2) \end{cases} \tag{3}$$

where λ is a transformation parameter; x refers to anomalies of daily raw ETo forecasts or AWAP ETo (mm day^{-1}); $\hat{x}$ is the transformed x through the Yeo-Johnson transformation.

We assume that the transformed anomalies of ETo forecasts ($f(t)$) and AWAP ETo ($o(t)$) are drawn from a bivariate normal distribution:

$$[f(t), o(t)] \sim BN(f(t), o(t)|\mu_f(m(t)), \sigma_f^2(m(t)), \mu_o(m(t)), \sigma_o^2(m(t)), \rho(m(t))) \tag{4}$$

where m(t) returns the month k (k=1 to 12) of daily forecasts or observations of day t; $\mu_f(m(t))$ and $\sigma_f(m(t))$ refer to the marginal distribution's mean and standard deviation of $f(t)$ in month m(t), respectively; $\mu_o(m(t))$ and $\sigma_o(m(t))$ are the mean and standard deviation of the marginal distribution of $o(t)$ in month m(t); ρ(m(t) is the correlation between $f(t)$ and $o(t)$ of month m(t).

With the long-term (20-year) AWAP ETo data, we can directly estimate μ_o(m(t)) and σ_o(m(t)) based on a maximum likelihood optimization. Calculation of the mean (μ_f(m(t))) and standard deviation (σ_f(m(t))) using the short-archived raw ETo forecasts (3-year) is subjected to significant sampling errors. Instead, we indirectly estimate them using the following linearly regressions:

$$\mu_f(k) = a + b\mu_o(k) \tag{5}$$

$$\sigma_f(k) = c + d\sigma_o(k) \tag{6}$$

$$\rho(k) = r \tag{7}$$

where k is a month k (k= 1 to 12); a, b, c, and d are parameters characterizing the linear relationships; ρ (k) denotes the correlation coefficient between raw forecast anomaly and anomaly of AWAP ETo for each month; r is the correlation coefficient between anomalies of raw forecasts and observations in the transformed space (Wang et al., 2019).

Once we obtain all the parameters for the BN distribution (equation 4), a conditional distribution is established for $o(t)$ when a raw forecast ($f(t)$) is provided. From this conditional distribution, we randomly draw 100 values. These values are then treated as the calibrated ensemble forecasts for that raw forecast. Finally, the calibrated anomalies are back-transformed to their original space and added back to the climatological mean to produce calibrated ETo forecasts.

2.4 Evaluation of calibrated forecasts

170 We evaluate the performance of the ETo forecast calibration using a strict leave-one-month-out cross-validation strategy, in which each of the 36 months during 4/2016 to 3/2019 and the same month in the 20-year reference data (4/1999 to 3/2019) are left out in parameter inference. Optimized parameters are then used to calibrate raw forecasts of this specific month. This process is repeated until all 36 months are processed. This strategy is to make sure that calibrated forecasts are produced with data not used for parameter optimization.

175 We also produce climatology forecasts based on the monthly mean and standard deviation parameters of AWAP ETo listed in equation (4). The randomly sampled climatology is used as the baseline to evaluate the calibrated forecasts generated with the SCC model. We evaluate the calibrated forecasts by checking bias, reliability, temporal variability, and skill score in raw and calibrated forecasts. The evaluation metrics are further introduced in detail as follows.

2.4.1 Bias

180 We evaluate the accuracy of the raw and calibrated forecasts relative to AWAP-based ETo using the following equation:

$$Bias = \frac{1}{T}\sum_{t=1}^{T}(x(t) - y(t))$$ (8)

where *Bias* refers to the bias in average ETo (mm day^{-1}); T is total days during the 3-year validation period (4/2016-3/2019); $x(t)$ is raw or calibrated forecasts of ETo (mm day^{-1}), and $y(t)$ is the corresponding AWAP ETo of the same period. We also compare the absolute bias of calibrated forecasts from Calibrations 1 and 2 to evaluate how bias-correction of input variables affects the accuracy of calibrated ETo forecasts.

2.4.2 Reliability

To evaluate the reliability of calibrated forecasts, we calculate the probability integral transform (PIT) value using the following equation:

$$\pi(t) = F(t, x = y(t))$$ (9)

190 where $F(t,x)$ is the cumulative density function of the ensemble forecast, and $y(t)$ is the AWAP ETo. For reliable forecasts, $\pi(t)$ follows a uniform distribution. We use the alpha index (α) to summarize the reliability in each grid cell using the following equation to check the spatial patterns of forecast reliability (Renard et al., 2010):

$$\alpha = 1 - \frac{2}{n}\sum_{t=1}^{n}\left|\pi^*(t) - \frac{t}{n+1}\right|$$ (10)

where π* (t) is the sorted π(t), t=1,2,…,n, in ascending order, and n is the total number of days during 4/2016-3/2016. The α
195 index measures the total deviation of calibrated forecasts from the corresponding uniform quantile. Perfectly reliable forecasts should have an α index of 1, and forecasts with no reliability would have an α index of 0.

2.4.3 Temporal variability

We use the Pearson correlation coefficient (r) between calibrated forecasts and observations to evaluate their consistency in temporal variability. For the calibrated ensemble forecasts, we use the ensemble mean of the 100 ensemble members for the calculation of r:

$$r = \frac{\sum_{t=1}^{n}(x(t)-\bar{x})(y(t)-\bar{y})}{\sqrt{\sum_{t=1}^{n}(x(t)-\bar{x})^2}\sqrt{\sum_{t=1}^{n}(y(t)-\bar{y})^2}}$$ (11)

where $x(t)$ is raw or calibrated forecasts of ETo (mm/day), $\bar{x}$ is the average of $x(t)$ (mm day^{-1}); $y(t)$ is the corresponding AWAP ETo of the same period, and $\bar{y}$ is the average of $y(t)$ (mm day^{-1}).

2.4.4 Skills of the raw and calibrated forecasts

We use the continuous ranked probability score (CRPS) to measure skills in the raw and calibrated forecasts (Grimit et al., 2006):

$$CRPS(t) = \int\{F(t,x) - H(x - y(t))\}^2 dx$$ (12)

$$\overline{CRPS} = \frac{1}{n}\sum_{t=1}^{n}CRPS(t)$$ (13)

where $F(t,x)$ is the cumulative density function of an ensemble forecast, and $y(t)$ is the observation at time t; H is the Heaviside step function ($H = 1$ if $x - y(t) \geq 0$ and $H = 0$ otherwise); the overbar represents averaging across the n days. For deterministic forecasts, CRPS is reduced to absolute errors.

We further calculate the CRPS skill score ($CRPS_{SS}$) to measure the skills relative to climatology forecasts using the following equation:

$$CRPS_{SS} = \frac{CRPS_{reference} - CRPS_{forecasts}}{CRPS_{reference}} \times 100$$ (14)

where $CRPS_{reference}$ is the CRPS value of climatology forecasts; $CRPS_{forecasts}$ refers to CRPS value of raw or calibrated forecasts. For $CRPS_{SS}$ of Calibrations 1 and 2, climatology forecasts from Calibration 1 are used; For $CRPS_{SS}$ of Calibrations 3 and 4, climatology forecasts from Calibration 3 are used. Positive skill scores indicate better skills than the climatology forecasts and vice versa.

3 Results

3.1 Bias in input variables and raw ETo forecasts

Raw forecasts of the five input variables demonstrate significant inconsistencies with the corresponding AWAP data (Figures S1-S5). In most parts of Australia, daily maximum temperature (Tmax) forecasts are lower than AWAP data by 1-2 °C. Overpredictions in Tmax are only found in coastal areas of north-western Australia. The daily minimum temperature (Tmin) is underpredicted in western and central parts of Australia by the raw forecasts, but overpredicted in eastern and southern Australia. Forecasted wind speed is higher than the reference data by more than 1m/s in most parts of Australia. Similarly, raw

solar radiation forecasts are also higher than AWAP data across Australia. Vapor pressure is underpredicted in western and central regions by the raw forecasts, but demonstrates positive biases in coastal areas of south-eastern Australia. For each of the five variables, spatial patterns of biases in raw forecasts are consistent across the 9 lead times, demonstrating systematic errors in the raw NWP forecasts.

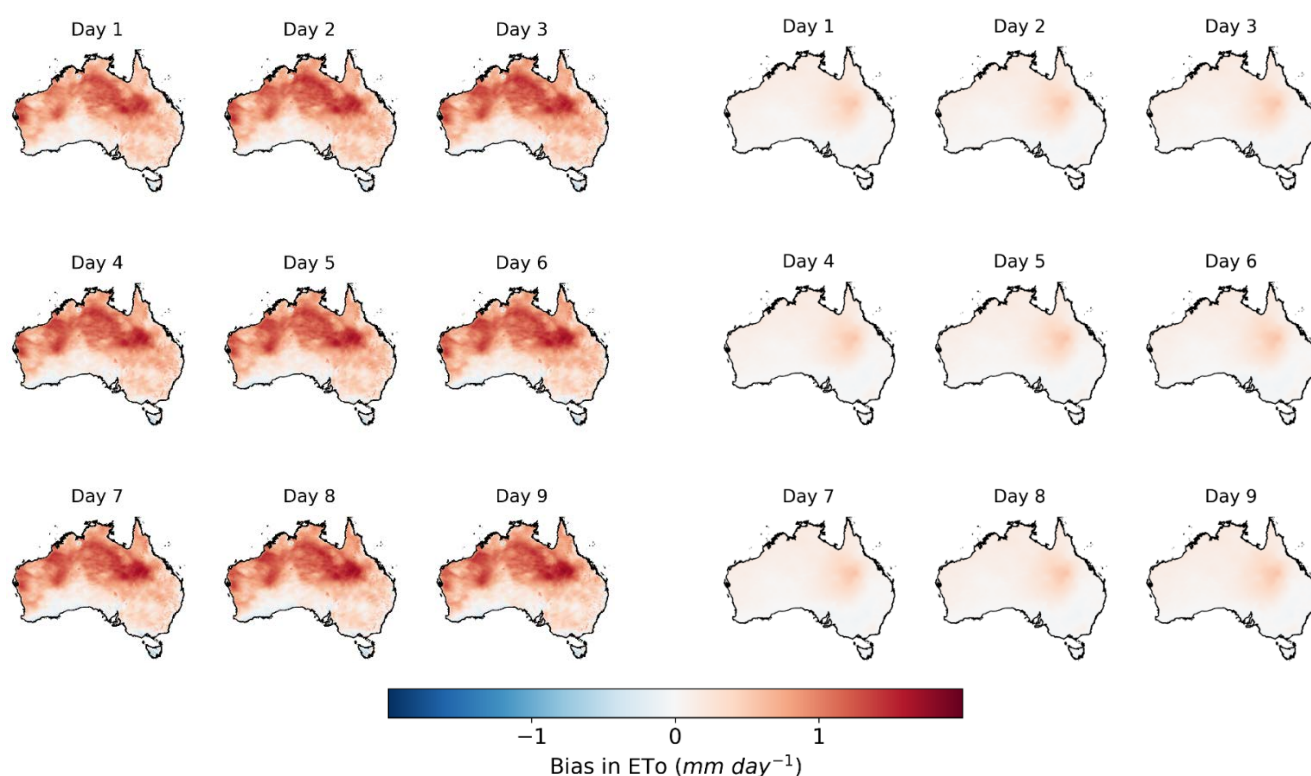
230   The quantile mapping effectively corrects biases in raw forecasts of the input variables. Through the bias correction, the significant overpredictions and underpredictions in raw forecasts of the five variables are effectively corrected, resulting in biases close to zero for all 9 lead times across Australia (Figures S1-S5).

235

Figure 1: Bias in (left three panels) raw ETo forecasts constructed with raw forecasts of input variables and (right three panels) raw ETo forecasts constructed with bias-corrected input variables.

Raw ETo forecasts constructed with the bias-corrected input variables are much more accurate than those calculated with raw

240   forecasts of these variables (Figure 1). When raw ETo forecasts are constructed with raw forecasts of input variables, biases in input variables are translated into errors of the raw ETo forecasts, which demonstrate substantial positive biases in the central and northern regions of the country. Correcting biases in the input variables leads to more accurate estimations of ETo. In the raw ETo forecasts calculated with bias-corrected input variables, biases are close to zero in most parts of Australia,

except for inland regions of Queensland, where biases in ETo are close to 0.3 mm/day. However, the remaining biases are

245   significantly lower than those in the raw ETo forecasts constructed with uncorrected input variables.

3.2 Bias in calibrated ETo forecasts

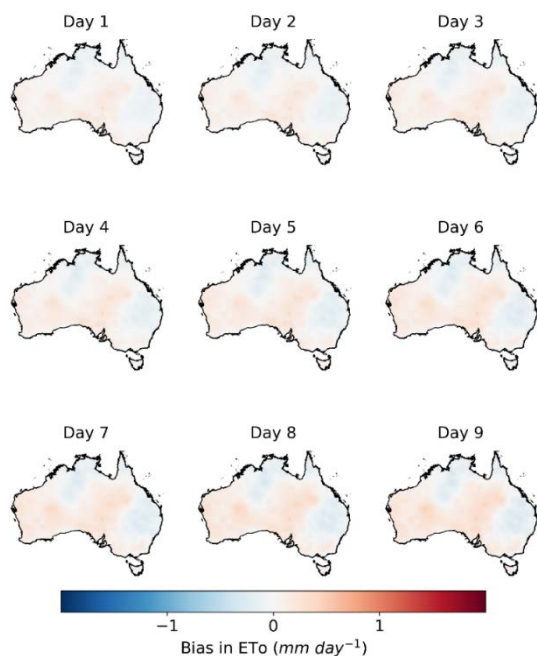



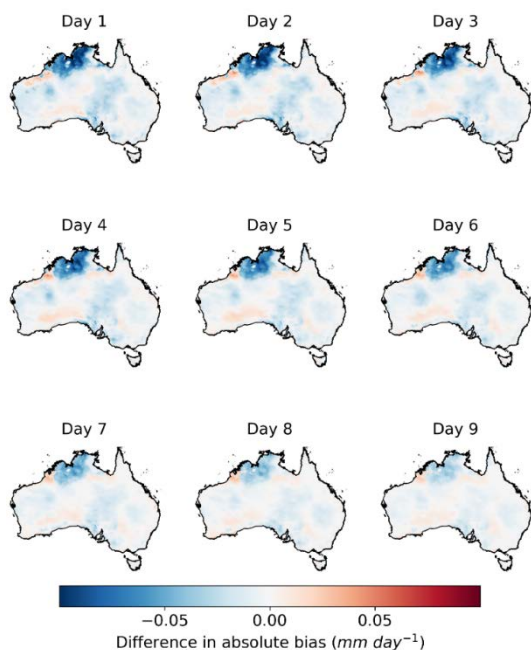
Figure 2: Bias in calibrated ETo forecasts from Calibration 2, which constructs raw ETo forecasts with bias-corrected input variables.

250

Figure 3: Differences in absolute bias between calibrated ETo forecasts from Calibration 2 with those from Calibration 1.

The calibration with the SCC model further reduces biases in ETo forecasts (Figure 2). The calibrated ETo forecasts from
255 Calibration 2 show low biases close to zero across all grid cells and lead times. Overpredictions in Queensland in the raw ETo
forecasts calculated with the bias-corrected input variables are effectively corrected (Figures 1 and 2). Compared with the
calibration constructing raw ETo forecasts with raw forecasts of input variables (Calibration 1), the post-processing based on
bias-corrected input variables (Calibration 2) produces more accurate calibrated ETo forecasts (Figure 3). Specifically,
calibrated ETo forecasts from Calibration 2 demonstrate smaller absolute biases than those of Calibration 1 across large areas
260 of northern Australia, particularly in coastal regions of the Northern Territory. The improvements are more pronounced for
short lead times (Days 1-3) than long lead times (Days 7-9).

3.3 Reliability of calibrated ETo forecasts

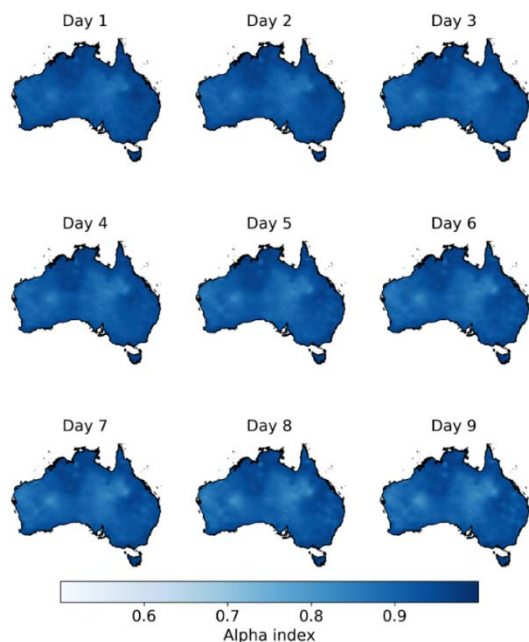



Figure 4: Alpha index of calibrated ETo ensemble forecasts from Calibration 2.

265

The calibrated ensemble ETo forecasts from Calibration 2 demonstrate high reliability (Figure 4). In addition to correcting bias, the calibration model adopted in this study converts deterministic raw forecasts to ensemble forecasts, with 100 ensemble members for each forecast. The ensemble forecasts provide forecast users with uncertainty information of future ETo. Figure 4 demonstrates highly reliable ensemble spreads in calibrated forecasts across all 9 lead times. In most grid cells, the α index

270 is over 0.9, indicating reasonable representations of ETo uncertainties by the ensemble spread, which is neither too narrow nor too wide (Figure 4).

Calibrated forecasts from Calibration 1, which uses uncorrected input variables, demonstrate similar reliability as those from the calibration with bias-corrected input variables (Calibration 2). Differences in α index of the calibrated ETo forecasts from Calibrations 1 and 2 are almost negligible (Figure S6), indicating that bias-correcting raw forecasts of input variables does not

275 lead to significant changes in the reliability of calibrated ETo forecasts.

3.4 Correlation between calibrated ETo forecasts and AWAP ETo

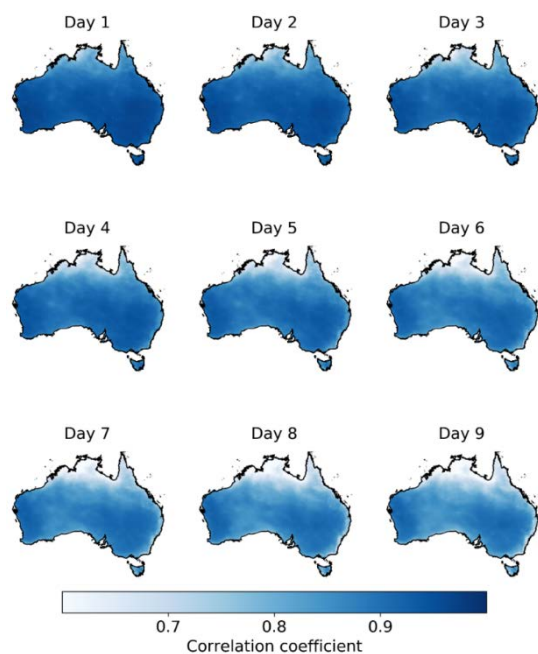



Figure 5: The correlation coefficient between calibrated ETo forecasts from Calibration 2 and AWAP ETo data.

280   We further examine the representation of temporal variability of ETo by calibrated forecasts. The correlation coefficient (*r*) between calibrated ETo forecasts from Calibration 2 and AWAP ETo demonstrates high consistency in temporal variability (Figure 5). The correlation coefficient is mainly above 0.85 across Australia for the first three lead times. With increases in lead time, *r* values tend to decrease but remain above 0.7 in most parts of the country, even at lead time 9. Coastal areas of northern Australia have lower *r* values than other regions of the country, demonstrating higher uncertainties in ETo forecasts

285   in regions with a dynamic tropical/subtropical climate system.

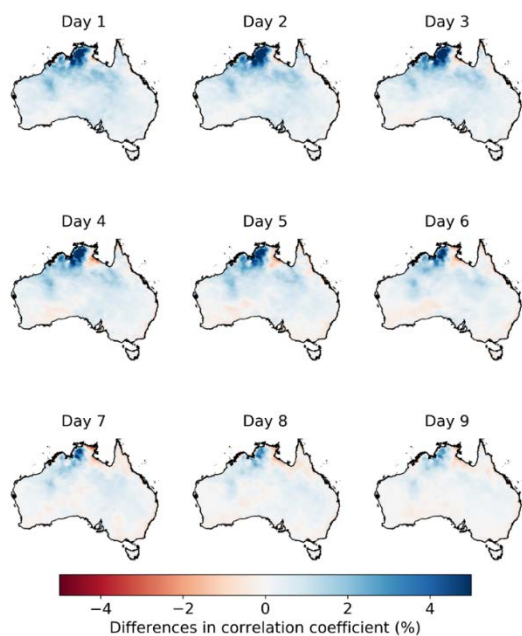



Figure 6: Differences in the correlation coefficient (calibrated forecasts vs. observations) between Calibrations 2 and 1.

290 The adoption of bias-correction to raw forecasts of input variables results in better representation of ETo variability in calibrated ETo forecasts (Figure 6). The comparison of r between AWAP ETo and calibrated ETo forecasts from Calibration 2 with that from Calibration 1 shows increases in r, after applying bias-correction to input variables. Increases in r are more significant for short lead times. Specifically, for the first three lead times, increases in r are mainly around 2% in central and northern parts of Australia, with more pronounced (>4%) increases found in coastal regions of the Northern Territory. For lead

295 times 4 to 6, increases in r values are mainly above 1%. For the remaining 3 lead times (7 to 9), we can still see slight increases in r, mainly in Northern Territory. Spatial patterns of r increases from Calibration 1 to Calibration 2 are consistent across the 9 lead times, with more significant increases occurring in northern parts of Australia.

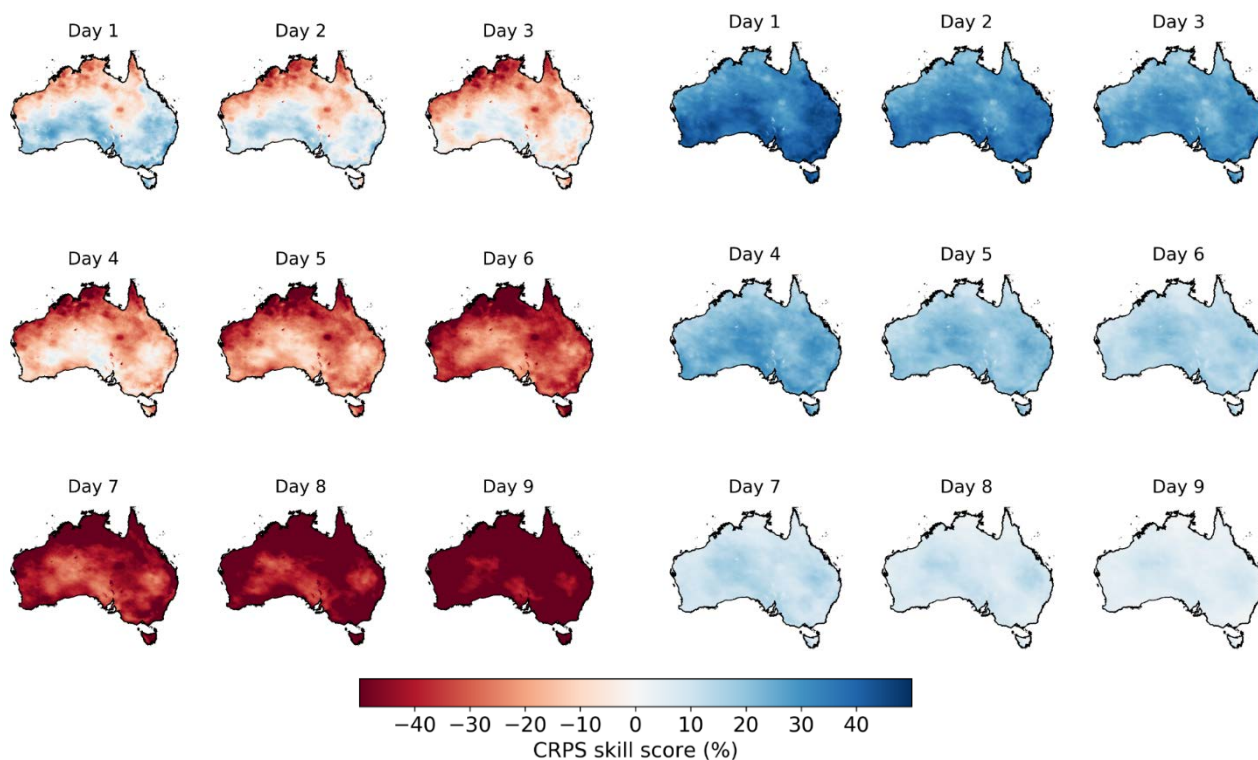
3.5 Improvements in forecast skills

300

Figure 7: CRPS skill score in the (left three panels) raw ETo forecasts calculated with bias-corrected input variables and (right three panels) calibrated forecasts from Calibration 2.

The calibration of ETo forecasts with the SCC model significantly improves forecast skills. The raw ETo forecasts calculated

305 with bias-corrected input variables demonstrate low skills, even at short lead times (Figure 7). Specifically, for the first two lead times, central and southern Australia show skills better than the climatology forecasts by 10% to 20%. However, in most parts of northern Australia, raw forecasts are worse than randomly sampled climatology. Skills in raw ETo forecasts decrease quickly with lead time. Regions with positive skills shrink substantially at lead times 3 and 4, and disappear at longer lead times. At lead time 9, skills of raw forecasts are mainly below -40%.

310 The calibration significantly improves forecast skills for all 9 lead times. Results from Calibration 2 show that the CRPS skill score of calibrated ETo forecasts is mainly above 40% at lead time 1 across Australia, and lead times 2 and 3 show skills generally above 30%. Since ETo forecasts have been widely used to inform real-time decision making for irrigation, high skills in calibrated ETo forecasts for the short lead times are expected to be highly valuable for activities such as determining the timing and frequency of irrigation. Although skills of calibrated forecasts also decrease with lead time, they remain above zero

315 for the remaining lead times (Figure 7).

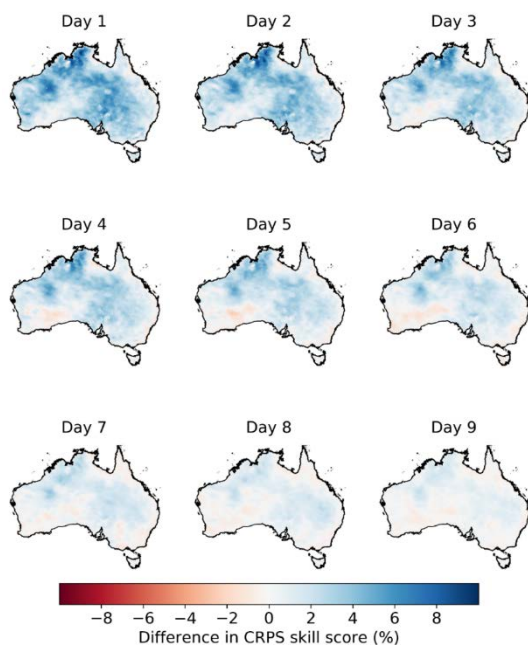



Figure 8. Differences in CRPS skill scores between the calibrated forecast from Calibration 2 with those from Calibration 1.

320   We further compare skills of calibrated ETo forecasts between Calibrations 2 and 1. The differences in forecast skills between the two sets of calibrated forecasts clearly demonstrate the importance and necessity of improving input variables (Figure 8). We find significant increases in CRPS skill scores with the adoption of bias-correction to input variables, particularly for short lead times. For the first three lead times, CRPS skill score is increased by more than 6% when forecasts of input variables are bias-corrected. For lead times 4 to 6, the CRPS skill scores are increased by 4% in many regions across Australia, with more

325   significant increases found in the central and northern parts of the country. Although the differences become less noticeable at lead times 7 to 9, they are generally above zero in most parts of Australia. The differences in CRPS skill score between Calibration 2 and 1 across the 9 lead times are in line with that of the absolute bias (Figure 3) and correlation coefficient (Figure 6), which all show more significant improvements at short lead times than long lead times.

   We also compare the CRPS skill score of calibrated forecasts from Calibrations 3 and 4, to evaluate whether we can obtain

330   similar improvements through the bias-correction of input variables if we conduct the ETo forecast calibration in a different way (without using climatological mean and anomalies). Results show similar improvements in CRPS skill score in magnitude, spatial patterns, and trend along the lead times, when input variables are bias-corrected before combined calibration (Figure S7). This additional comparison further confirms the necessity of improving forecasts of input variables as part of the calibration of ETo forecasts.

4 Discussion

4.1 Importance of improving forecasts of input variables for calibrating ETo forecasts

Forecasting of future ETo based on NWP outputs is gaining popularity as a result of high data availability, flexible forecast horizons, and timeliness (Perera et al., 2014, 2016; Silva et al., 2010; Yang et al., 2019). As for forecasts of other weather variables, statistical calibration plays an important role in improving the quality of raw ETo forecasts (Medina and Tian, 2019). Although existing investigations using raw forecasts of input variables to construct the raw ETo forecasts for calibration often achieved significant improvements in skills, it is still unclear whether improving forecasts of input variables could further enhance ETo forecast calibration (Medina and Tian, 2020). How the raw ETo forecasts should be constructed represents an important knowledge gap in the area of NWP-based ETo forecasting.

Results of this investigation provide strong evidence for the necessity of improving raw forecasts of input variables prior to constructing raw ETo forecasts. Each of the input variables demonstrates high non-linearity and non-stationary (Paredes et al., 2018). When a bias-correction method is applied to the input variables, the final calibrated ETo forecasts demonstrate higher accuracy and skills than the calibration based on uncorrected input variables. The results suggest that complex interactions of these input variables may lead to errors in raw ETo forecasts that could not be effectively corrected through statistical calibration. Bias-correction of input variables could help prohibit the propagation of errors from input variables to ETo forecasts (Zappa et al., 2010). More importantly, a further evaluation based on a different way of implementing the calibration model demonstrates similar improvements in calibrated ETo forecasts with the adoption of bias-correction to input variables (Calibrations 3 and 4). This additional analysis suggests that additional skills have been added to ETo forecasts before they are calibrated through the bias-correction of input variables. Consequently, we recommend that the future NWP-based ETo forecasting could take this extra step into account to improve the quality of calibrated ETo forecasts.

4.2 Implications for forecasting of integrated variables and future work

This investigation also provides valuable implications for the forecasting integrated variables, which need to be calculated with multiple NWP/GCM variables. Variables such as drought index (Zhang et al., 2017), bushfire danger index (Sharples et al., 2009), and severe weather index (Rabbani et al., 2020), are often derived from multiple NWP/GCM variables. Our investigation suggests that improving the input variables may help correct errors that could not be fixed when calibrating the integrated variables directly. This extra step could help improve the potential predictability of integrated variables.

We anticipate that the ETo forecasts could be further improved if a more sophisticated calibration model is applied to raw forecasts of the input variables. In this study, we adopt a simple bias-correction method to the input variables rather than using a sophisticated calibration model. Limitations of quantile-mapping have been reported (Schepen et al., 2020; Zhao et al., 2017). Our analyses also demonstrate that the raw ETo forecasts calculated with the bias-corrected input variables still contain biases and have low forecast skills, particularly at long lead times (Figure 7). As a result, the adoption of a more sophisticated

calibration method to correct errors in input variables, is expected to further improve forecasts of these input variables, resulting in more significant improvements in the final calibrated ETo forecasts. Moreover, calibration of the input variables with a sophisticated model will benefit other forecast users. Well-calibrated forecasts of input variables, such as temperature, vapor pressure, solar radiation, and wind speed, could be useful for forecast users such as crop modelers and bushfire managers.

370 Although we've conducted thorough analyses on the contribution of improving input variables to ETo forecast calibration, further investigation is needed to validate the findings of this investigation in the future. First, the two calibration strategies should be tested using other NWP models. In this study, we use one NWP model to investigate a critical knowledge gap in NWP-based ETo forecasting. Additional investigations based on other NWP models are necessary to further enrich our understanding of the impacts of improving input variables on NWP-based ETo forecasting. Second, further investigations

375 based on other calibration models are needed to validate the conclusions of this investigation. Our analyses based on two different methods (based on ETo anomalies vs. based on original ETo) find similar improvements in calibrated ETo forecasts by correcting input variables. Additional investigations using other calibration models will help clarify whether the improvements will hold for other calibration models.

5 Conclusions

380 NWP outputs have been increasingly used for ETo forecasting. As for other weather variables, statistical calibration plays an essential role in improving the quality of raw ETo forecasts. However, it is unclear whether improving raw forecasts of input variables is necessary for the calibration of ETo forecasts. We aim to fill this knowledge gap through a thorough comparison of calibrations with and without the adoption of bias-correction to input variables.

This investigation clearly suggests the necessity of improving the input variables as part of the NWP-based ETo forecasting.

385 With this extra step, the bias, correlation coefficient, and skills the calibrated ETo forecasts are all improved, particularly for the short-lead-time forecasts. Further investigation indicates that the contribution of improving input variables to the ETo forecasting tends to be independent of the calibration method applied to raw ETo forecasts. As a result, we anticipate that future calibration of NWP-based ETo forecasts could benefit from adopting this extra step to produce skillful calibrated ETo forecasts.

390 Short-term ETo forecasts are highly valuable for real-time decision making for water resource management and planning of farming activities. Forecasting of the highly variable ETo is often challenging. Our investigation provides a feasible method for advancing calibrations of the NWP-based ETo forecasts. This strategy is also expected to be useful in enhancing the forecasting of other integrated variables that are calculated using multiple NWP/GCM variables.

395

Author contributions:

Q. Yang and Q. J. Wang conceived this study. Q. J. Wang developed the calibration model. Q. Yang took the lead in writing the manuscript. All authors discussed the results and contributed to improving the manuscript**.**

400

Acknowledgments:

This study has been supported by a collaborative research project (TP707466) between the University of Melbourne and the Australian Bureau of Meteorology and an ARC Linkage Project (LP170100922). Computations of this research were undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is

405 supported by the Australian Government. This research was supported by the Sustaining and strengthening merit-based access to National Computational Infrastructure (NCI) LIEF Grant (LE190100021) and facilitated by The University of Melbourne. The authors declare that there is no conflict of interest regarding the publication of this article.

References:

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: FAO Irrigation and drainage paper No.56, Crop evapotranspiration:

410 guidelines for computing crop water requirements., 1998.

Bachour, R., Maslova, I., Ticlavilca, A. M., Walker, W. R. and Mckee, M.: Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration, Stoch. Environ. Res. Risk Assess., 30, 103–117, doi:10.1007/s00477-015-1039-z, 2016.

Ballesteros, R., Ortega, F. and Angel, M.: FORETo : New software for reference evapotranspiration forecasting, J. Arid

415 Environ., 124, 128–141, doi:10.1016/j.jaridenv.2015.08.006, 2016.

Boe, J., Terray, L., Habets, F. and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydrometeorological studies, Int. J. Clim., 27(January), 1463–1655, doi:10.1002/joc, 2007.

Cai, J., Liu, Y., Lei, T. and Pereira, S. L.: Estimating reference evapotranspiration with the FAO Penman – Monteith equation using daily weather forecast messages, Agric. For. Meteorol., 145, 22–35, doi:10.1016/j.agrformet.2007.04.012, 2007.

420 Djaman, K., Neill, M. O., Owen, C. K., Smeal, D., Koudahe, K., West, M., Allen, S., Lombard, K. and Irmak, S.: Crop Evapotranspiration , Irrigation Water Requirement and Water Productivity of Maize from Meteorological Data under Semiarid Climate, Water, 10(405), 1–17, doi:10.3390/w10040405, 2018.

Er-Raki, S., Chehbouni, A., Khabba, S., Simonneaux, V., Jarlan, L., Ouldbba, A., Rodriguez, J. C. and Allen, R.: Assessment of reference evapotranspiration methods in semi-arid regions: Can weather forecast data be used as alternate of ground

425   meteorological parameters?, J. Arid Environ., 74(12), 1587–1596, doi:10.1016/j.jaridenv.2010.07.002, 2010.

Grimit, E. P., Gneiting, T., Berrocal, V. J. and Johnson, N. A.: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, Q. J. R. Meteorol. Soc., 132, 2925–2942, doi:10.1256/qj.05.235, 2006.

Hopson, T. M. and Webster, P. J.: A 1-10-Day Ensemble Forecasting Scheme for the Major River Basins of Bangladesh :
430   Forecasting Severe Floods of 2003 – 07, J. Hydrometeorol., 11(July 2004), 618–641, doi:10.1175/2009JHM1006.1, 2009.

Jones, D. A., Wang, W. and Fawcett, R.: Climate Data for the Australian Water Availability Project, Australian Bureau of Meteorology,   Melbourne,   Australia.   [online]   Available   from: https://trove.nla.gov.au/work/17765777?q&versionId=20839991, 2007.

Jones, D. A., Wang, W. and Fawcett, R.: Australian Water Availability Project Daily Gridded Rainfall, [online] Available
435   from: http://www.bom.gov.au/jsp/awap/rain/index.jsp, 2014.

Karbasi, M.: Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet- Gaussian Process Regression Model, Water Resour. Manag., 32, 1035–1052, 2018.

Kumar, R., Jat, M. K. and Shankar, V.: Methods to estimate irrigated reference crop evapotranspiration - A review, Water Sci. Technol., (June), doi:10.2166/wst.2012.191, 2012.

440   Lim, J. and Park, H.: H Filtering for Bias Correction in Post-Processing of Numerical Weather Prediction, J. Meteorol. Soc. Japan, 97(3), 773–782, doi:10.2151/jmsj.2019-041, 2019.

Liu, Y. J., Chen, J. and Pang, T.: Analysis of Changes in Reference Evapotranspiration , Pan Evaporation, and Actual Evapotranspiration and Their Influencing Factors in the North China Plain During 1998 – 2005, Earth Sp. Sci., 6, 1366–1377, doi:10.1029/2019EA000626, 2019.

445   Luo, Y., Chang, X., Peng, S., Khan, S., Wang, W., Zheng, Q. and Xueliang, C.: Short-term forecasting of daily reference evapotranspiration using the Hargreaves – Samani model and temperature forecasts, Agric. Water Manag., 136, 42–51, doi:10.1016/j.agwat.2014.01.006, 2014.

Mariito, M. A., Tracy, J. C. and Taghavv, S. A.: Forecasting of reference crop evapotranspiration, Agric. Water Manag., 24, 163–187, 1993.

450   Mcvicar, T. R., Niel, T. G. Van, Li, L. T., Roderick, M. L., Rayner, D. P., Ricciardulli, L. and Donohue, R. J.: Wind speed climatology and trends for Australia , 1975 – 2006 : Capturing the stilling phenomenon and comparison with near-surface reanalysis output, Geophys. Res. Lett., 35, 1–6, doi:10.1029/2008GL035627, 2008.

Medina, H. and Tian, D.: Comparison of probabilistic post-processing approaches for improving NWP-based daily and weekly reference evapotranspiration forecasts, Hydrol. Earth Syst. Sci. Discuss., (August), 1–32, 2019.

455   Medina, H. and Tian, D.: Comparison of probabilistic post-processing approaches for improving numerical weather prediction-based daily and weekly reference evapotranspiration forecasts, Hydrol. Earth Syst. Sci., 24, 1011–1030, 2020.

Medina, H., Tian, D., Srivastava, P., Pelosi, A. and Chirico, G. B.: Medium-range reference evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions, J. Hydrol., 562(April), 502–517,

doi:10.1016/j.jhydrol.2018.05.029, 2018.

460 Mushtaq, S., Reardon-smith, K., Kouadio, L., Attard, S., Cobon, D. and Stone, R.: Value of seasonal forecasting for sugarcane farm irrigation planning, Eur. J. Agron., 104, 37–48, doi:10.1016/j.eja.2019.01.005, 2019.

Narapusetty, B., Delsole, T. and Tippett, M. K.: Optimal estimation of the climatological mean, J. Clim., 22(18), 4845–4859, doi:10.1175/2009JCLI2944.1, 2009.

Nouri, M. and Homaee, M.: On modeling reference crop evapotranspiration under lack of reliable data over Iran, J. Hydrol.,
465 566(August), 705–718, doi:10.1016/j.jhydrol.2018.09.037, 2018.

Paredes, P., Fontes, J. C., Azevedo, E. B. and Pereira, L. S.: Daily reference crop evapotranspiration with reduced data sets in the humid environments of Azores islands using estimates of actual vapor pressure, solar radiation, and wind speed, Theor. Appl. Climatol. Appl., 134, 1115–1133, 2018.

Pelosi, A., Medina, H., Villani, P., D'Urso, G. and Chirico, G. B.: Probabilistic forecasting of reference evapotranspiration
470 with a limited area ensemble prediction system, Agric. Water Manag., 178, 106–118, doi:10.1016/j.agwat.2016.09.015, 2016.

Perera, K. C., Western, A. W., Nawarathna, B. and George, B.: Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs, Agric. For. Meteorol., 194, 50–63, doi:10.1016/j.agrformet.2014.03.014, 2014.

Perera, K. C., Western, A. W., Robertson, R. D., George, B. and Nawarathna, B.: Ensemble forecasting of short-term system scale irrigation demands using real-time flow data and numerical weather predictions, Water Resour. Res., 52, 4801–4822,
475 doi:10.1002/2015WR018532.Received, 2016.

Rabbani, G., Yazd, N. K., Reza, M. and Daneshvar, M.: Factors affecting severe weather threat index in urban areas of Turkey and Iran, Environ. Syst. Res., 9(9), 1–14, doi:10.1186/s40068-020-00173-6, 2020.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, 1–22, doi:10.1029/2009WR008328,
480 2010.

Schepen, A., Everingham, Y. and Wang, Q. J.: On the Joint Calibration of Multivariate Seasonal Climate Forecasts from GCMs, Mon. Weather Rev., 148, 437–456, doi:10.1175/MWR-D-19-0046.1, 2020.

Sharples, J. J., Mcrae, R. H. D., Weber, R. O. and Gill, A. M.: A simple index for assessing fire danger rating, Environ. Model. Softw., 24(6), 764–774, doi:10.1016/j.envsoft.2008.11.004, 2009.

485 Silva, D., Meza, F. J. and Varas, E.: Estimating reference evapotranspiration (ETo) using numerical weather forecast data in central Chile, J. Hydrol., 382, 64–71, doi:10.1016/j.jhydrol.2009.12.018, 2010.

Srivastava, P. K., Han, D., Ramirez, M. A. R. and Islam, T.: Comparative assessment of evapotranspiration derived from NCEP and ECMWF global datasets through Weather Research and Forecasting model, Atmos. Sci. Lett., 14, 118–125, doi:10.1002/asl2.427, 2013.

490 Tian, D. and Martinez, C. J.: The GEFS-Based Daily Reference Evapotranspiration (ETo) Forecast and Its Implication for Water Management in the Southeastern United States, J. Hydrometeorol., 15, 1152–1165, doi:10.1175/JHM-D-13-0119.1, 2014.

Torres, A. F., Walker, W. R. and Mckee, M.: Forecasting daily potential evapotranspiration using machine learning and limited climatic data, Agric. Water Manag., 98(4), 553–562, doi:10.1016/j.agwat.2010.10.012, 2011.

495   Traore, S., Luo, Y. and Fipps, G.: Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages, Agric. Water Manag., 163, 363–379, doi:10.1016/j.agwat.2015.10.009, 2016.

Turco, M., Ceglar, A., Prodhomme, C., Soret, A., Toreti, A. and Francisco, J. D.-R.: Summer drought predictability over Europe : empirical versus dynamical forecasts, Environ. Res. Lett., 12, 084006, 2017.

500   Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A. and Gneiting, T.: Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa, Weather Forecast., 33, 369–388, doi:10.1175/WAF-D-17-0127.1, 2018.

Wang, Q. J., Zhao, T., Yang, Q. and Robertson, D.: A Seasonally Coherent Calibration (SCC) Model for Postprocessing Numerical Weather Predictions, Mon. Weather Rev., 147, 3633–3647, doi:10.1175/MWR-D-19-0108.1, 2019.

Yang, Q., Wang, Q. J. and Hakala, K.: Achieving effective calibration of precipitation forecasts over a continental scale, Under

505   Rev., 2021a.

Yang, Q., Wang, Q. J. and Hakala, K.: Working with anomalies improves forecast calibration of daily reference crop evapotranspiration, , Under Revi, 2021b.

Yang, Y., Cui, Y., Bai, K., Luo, T., Dai, J., Wang, W. and Luo, Y.: Short-term forecasting of daily reference evapotranspiration using the reduced-set Penman-Monteith model and public weather forecasts, Agric. Water Manag., 211, 70–80,

510   doi:10.1016/j.agwat.2018.09.036, 2019.

Yeo, I. and Johnson, R. A.: A new family of power transformations to improve normality or symmetry, Biometrika, 87(4), 954–959, 2000.

Zappa, M., Beven, K. J., Bruen, M., Cofino, A. S., Kok, K., Martin, E., Nurmi, P., Orfila, B., Roulin, E., Schroter, K., Seed, A., Szturc, J., Vehvilainen, B., Germann, U. and Rossa, A.: Propagation of uncertainty from observing systems and NWP into

515   hydrological models : COST-731 Working Group 2, Atmos. Sci. Lett., 11(January), 83–91, doi:10.1002/asl.248, 2010.

Zhang, X., Tang, Q., Liu, X., Leng, G. and Li, Z.: Soil Moisture Drought Monitoring and Forecasting Using Satellite and Climate Model Data over Southwestern China, J. Hydrometeology, 18, 5–23, doi:10.1175/JHM-D-16-0045.1, 2017.

Zhao, T., Bennett, J., Q.J., W., Schepen, A., Wood, A., Robertson, D. E. and Ramos, M.-H.: How Suitable is Quantile Mapping For Postprocessing GCM Precipitation Forecasts?, J. Hydrol., 30, 3185–3196, doi:10.1175/JCLI-D-16-0652.1, 2017.

520   Zhao, T., Wang, Q. J. and Schepen, A.: A Bayesian modelling approach to forecasting short-term reference crop evapotranspiration from GCM outputs, Agric. For. Meteorol., 269–270(January), 88–101, doi:10.1016/j.agrformet.2019.02.003, 2019a.

Zhao, T., Wang, Q. J., Schepen, A. and Griffiths, M.: Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs, Agric. For. Meteorol., 264(August 2018), 114–124,

525   doi:10.1016/j.agrformet.2018.10.001, 2019b.