

# 1 Comparison of probabilistic post-processing approaches for 2 improving NWP-based daily and weekly reference evapotranspiration 3 forecasts

4 Hanoi Medina, Di Tian

5 Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849

6 Correspondence to: Di Tian ([tiandi@auburn.edu](mailto:tiandi@auburn.edu))

7 **Abstract:** Reference evapotranspiration (ET<sub>o</sub>) forecasts play an important role in agricultural, environmental, and water  
8 management. This study evaluated probabilistic post-processing approaches, including the nonhomogeneous Gaussian  
9 regression (NGR), affine kernel dressing (AKD), and Bayesian model averaging (BMA) techniques, for improving daily and  
10 weekly ET<sub>o</sub> forecasting based on single or multiple numerical weather predictions (NWP) from The International Grand  
11 Global Ensemble (TIGGE), including the European Centre for Medium-Range Weather Forecasts (ECMWF), the National  
12 Centers for Environmental Prediction Global Forecast System (NCEP), and the United Kingdom Meteorological Office  
13 forecasts ( UKMO). The approaches were examined for the forecasting of summer ET<sub>o</sub> at 101 U.S. Regional Climate  
14 Reference Network stations distributed all over the contiguous United States (CONUS). We found that the NGR, the AKD  
15 and the BMA methods greatly improved the skill and reliability of the ET<sub>o</sub> forecasts compared to a linear regression bias  
16 correction method, due to the considerable adjustments on the spread of ensemble forecasts. The methods were especially  
17 effective when applied over the raw NCEP forecasts, followed by the raw UKMO forecasts, because of their low skill compared  
18 to that of the raw ECMWF forecasts. The post-processed weekly forecasts had much lower rRMSE (between 8-11%) than the  
19 persistence-based weekly forecasts (22%), and the post-processed daily forecasts (13-20%). Compared with the single model  
20 ensemble ET<sub>o</sub> forecasts based on ECMWF, multi-model ensemble ET<sub>o</sub> forecasts showed higher skill at short lead times (1 or  
21 2 days) and over the southern and western regions of the United States. The improvement was higher at the daily timescale  
22 than at the weekly timescale. The NGR and AKD methods performed the best, but unlike the AKD method, the NGR method  
23 can post-process multi-model forecasts and it is easier to interpret than the other methods. In summary, the study demonstrated  
24 that the three probabilistic approaches generally outperform conventional procedures based on the simple bias correction of  
25 single model forecasts, with the NGR post-processing of the ECMWF and ECMWF-UKMO forecasts providing the most cost-  
26 effective ET<sub>o</sub> forecasting.

## 27 Introduction

28 Reference crop evapotranspiration (ET<sub>o</sub>) represents the weather driven component of the water transfer from plants and soils  
29 to the atmosphere. It plays a fundamental role in estimating mass and energy balance over land surface as well as in agronomic,

30 forestry, and water resources management. In particular, ETo forecasting is important for aiding water management decision  
31 making (such as irrigation scheduling, reservoir operation, etc.) under uncertainty by identifying the range of future plausible  
32 water stress and demand (Pelosi et al., 2016; Chirico et al., 2018). While ETo forecasts have been mostly focused on the daily  
33 timescale (e.g. Perera et al., 2014; Medina et al., 2018), weekly ETo forecasts are also important for users. Studies show that  
34 both daily and weekly forecasts have increasing influence on the decision makers in agriculture (Prokopy et al., 2013; Mase  
35 and Prokopy, 2014) and water resource management (Hobbins et al., 2017). For example, irrigation is commonly scheduled  
36 considering both daily and weekly basis, while weekly evapotranspiration forecasts are useful for planning water allocation  
37 from reservoirs, especially in cases of shortages. Weekly ETo anomalies can also be useful to provide warnings of wild-fires  
38 (Castro et al., 2003) and evolving flash drought conditions (Hobbins et al., 2017).

39 However, ETo forecasting is highly uncertain due to the chaotic nature of weather systems. In addition, ETo estimation requires  
40 full sets of meteorological data which are usually not easy to obtain. Due to the improvement of numerical weather predictions  
41 (NWP), studies have been recently emerged to forecast ETo using outputs of NWP over different regions of the world (Silva  
42 et al., 2010; Tian and Martinez, 2012 a, 2012b, and 2014; Perera et al., 2014; Pelosi et al., 2016; Chirico et al., 2018; Medina  
43 et al., 2018). Operationally, experimental ETo forecast products are being developed, such as Forecast Reference  
44 EvapoTranspiration (FRET) product (<https://digital.weather.gov/>), as part of the U.S. National Weather Service (NWS)  
45 National Digital Forecast Database (NDFD) (Glahn and Ruth, 2003), and the Australian Bureau of Meteorology's Water and  
46 Land website (<http://www.bom.gov.au/watl>), which provides current and forecasted ETo at the continental scale.

47 The improved performance of NWP during recent years is largely due to the improvement of physical, statistical  
48 representations of the major processes in the models, and the use of ensemble forecasting (Hamill et al., 2013, Bauer et al.,  
49 2015). Nevertheless, the NWP forecasts still commonly show systematic inconsistencies with measurements, which are often  
50 caused by inherent errors of NWP or local land-atmospheric variability which is not well resolved in the models. Post-  
51 processing methods, defined as any form of adjustment to the model outputs in order to get better predictions (eg., Hagedorn  
52 et al., 2012), are highly recommended to attenuate, or even eliminate, those inconsistencies (Wilks, 2006). Until a few years  
53 ago, most post-processing applications only considered single-model predictions (i.e., predictions generated by a single NWP  
54 model), and addressed errors in the mean of the forecast distribution while ignored those in the forecast variance (Gneiting,  
55 2014). These procedures regularly adopted some form of model output statistics (MOS, Glahn and Lowry, 1972; Klein and  
56 Glahn, 1974) methods, focusing on correcting current ensemble forecasts based on the bias in the historical forecasts.

57 As no forecast is complete without an accurate description of its uncertainty (National Research Council of the National  
58 Academies 2006), the dispersion of the forecast ensemble often misrepresent the true density distribution of the forecast  
59 uncertainty (Krzysztofowicz 2001; Smith 2001; Hansen 2002). The ensemble forecasts are, for example, commonly under-  
60 dispersed (e.g. Buizza et al. 2005; Leutbecher and Palmer, 2008), which make the probabilistic predictions overconfident  
61 (Wilks 2011). Therefore, another generation of probabilistic techniques was proposed to also address dispersion errors of the  
62 ensembles (Hamill and Colucci 1997; Buizza et al., 2005, Pelosi et al., 2017), in some cases through the manipulation of multi-  
63 model weather forecasts.

64 The nonhomogeneous Gaussian regression (NGR, Gneiting et al., 2005), the Bayesian model averaging, (BMA, Raftery et al.,  
65 2005; Fraley et al., 2010), the extended logistic regression (ELR, Wilks et al., 2009; Whan and Schmeits, 2018), the quantile  
66 mapping (Verkade et al., 2013) and the family of kernel dressing (Roulston and Smith 2003; Wang and Bishop 2005), such  
67 as the affine kernel dressing (AKD, Brocker and Smith 2008), are state of art probabilistic techniques (Gneiting, 2014).  
68 However, the ELR has been reported to fall short in using the information contained in the ensemble spread in efficient way  
69 (Messner et al., 2014), while the quantile mapping method have been found to degrade rather than improve the forecast  
70 performance in some circumstances (Madadgar et al., 2014). The NGR, AKD and BMA are sometimes considered as variants  
71 of dressing methods (Brocker and Smith 2008), as they produce a continuous forecast probability distribution function (pdf)  
72 based on the original ensemble. This property makes them particularly useful for the decision making (Gneiting, 2014),  
73 compared to the methods that provide post-processed ensembles. Another common advantage is that they perform commonly  
74 well with relatively short training datasets (Geiting et al., 2005; Raftery et al., 2005; Wilks and Hamill, 2007). A limitation of  
75 the NGR, compared to the AKD and BMA methods, is that the resulting forecast pdf is invariably Gaussian, while a limitation  
76 of the AKD is that it only considers single model ensembles. Instead, the NGR and AKD methods provide more flexible  
77 mechanisms for the simultaneous adjustments in the forecast mean and spread-skill (Brocker and Smith, 2008).  
78 Studies suggest that the post-processing of NWP-based ETo forecasts are crucial for informing decision making (e.g. Ishak et  
79 al., 2010). Medina et al. (2018) compared single and multi-model NWP-based ensemble ETo forecasts and the results showed  
80 that the performance of the multi-model ensemble ETo forecasts is considerably improved through a simple bias-correction  
81 post-processing, and that the bias-corrected multi-model ensemble forecasts were in general better than the single model  
82 ensemble forecasts. In reality, while most applications for the ETo forecasting have involved some form of post-processing,  
83 these have been often limited to simple MOS procedures of single-model ensembles (e.g. Silva et al., 2010; Perera et al., 2014).  
84 Poor treatments of uncertainty and variability is considered as a main issue affecting users' perceptions and adoptions of  
85 weather forecasts (Mase and Prokopy, 2014). The appropriate representation of the second and higher moments of the ETo  
86 forecast probability density is especially important to predict extreme values, as shown by Williams et al. (2014). Therefore,  
87 the use of probabilistic post-processing techniques such as the NGR, the AKD and BMA, may greatly enhance the overall  
88 performance of the ETo forecasts compared to the simple MOS procedures.

89 Only a few studies have considered probabilistic methods for post-processing of ETo forecasts. These include the works of  
90 Tian and Martinez (2012a, 2012b, and 2014), and more recently Zhao et al (2019). The former authors showed the Analog  
91 Forecast (AF) method to be useful for the post-processing ETo forecasts based on Global Forecast System (GFS, Hamill et al.,  
92 2006) and Global Ensemble Forecast System (GEFS, Hamill et al., 2013) reforecasts. Tian and Martinez (2014) found that  
93 water deficit forecasts produced with the post-processed ETo forecasts had higher accuracy than those produced with  
94 climatology. On other hand, Zhao et al. (2019) improved the skill and the reliability of the Australian BoM model using a  
95 Bayesian joint probability (BJP) post-processing approach, which is based on the parametric modelling of the joint probability  
96 distribution between forecast ensemble means and observations. However, a main disadvantage of the BJP method compared  
97 to the aforementioned state of art probabilistic approaches is that, while they transform the spread of the ensembles, they rely

98 on the mean of retrospective reforecasts, thus neglecting information about their dispersion. The AF approach has the  
99 disadvantages that requires long time series of retrospective forecasts, and may be unsuitable for extreme events forecasting  
100 (e.g. Medina et al., 2019). The use of new ETo forecasting strategies relying on the postprocessing of single and multi-model  
101 ensemble forecasts with the NGR, AKD and the BMA probabilistic techniques provide good opportunities for improving the  
102 predictions.

103 In this paper, we are addressing several scientific questions which have not been adequately studied in previous literature,  
104 including, how effective are the state of art probabilistic post-processing methods compared with the traditional MOS bias  
105 correction methods for post-processing ETo forecasts? Is it worth implementing the probabilistic post-processing for multi-  
106 model rather than single-model ensemble forecasting? For the first time, this work aims to evaluate and compare multiple  
107 strategies for post-processing both daily and weekly ETo forecasts using the NGR, AKD and BMA approaches. The study  
108 represents a major step forward with respect to Medina et al. (2018), which evaluated the performance of raw and linear  
109 regression bias corrected daily ETo forecasts produced with single and multi-model ensemble forecasts. It provides a broad  
110 characterization of the performance for different probabilistic post-processing strategies but also diagnoses the causes of high  
111 and low performance.

## 112 **2 Methods and Datasets**

### 113 **2.1 The probabilistic methods**

114 The NGR, AKD and BMA techniques follow a common strategy: they yield a predictive probability density function (PDF)  
115 of the post-processed forecasts  $y$  given the raw forecasts  $x$  and some fitting parameters  $\theta$  ( $p(y|x, \theta)$ ). The parameters  $\theta$  are  
116 fitted using a training dataset of ensemble forecasts and observations, as in the MOS techniques. Below is a brief description  
117 of each technique.

#### 118 **2.1.1 Non-Homogeneous Gaussian Regression**

119 The NGR (Gneiting et al., 2005) produces a Gaussian predictive (PDF) based on the current ensemble (of typically multi-  
120 model) forecasts. If  $x_{ij}$  denote the  $j^{\text{th}}$  ( $j = 1, \dots, m_i$ ) ensemble forecast member of model  $i$  ( $i = 1, \dots, n$ ), then  
121  $p(y|x, \theta) \sim \mathcal{N}(\mu, v)$ , where the mean

$$122 \mu = a + \sum_{i=1}^n b_i \bar{x}_i \tag{1}$$

123 is a linear combination of the mean ensemble forecasts  $\bar{x}_i$  and the variance

$$124 v = c + dS^2 \tag{2}$$

125 is a linear function of the ensemble variance  $S^2$ . The fitting parameters  $a$ ,  $b_i$ ,  $c$  and  $d$  are determined by minimizing the  
126 continuous rank probability score (CRPS) using the training set of forecasts and observations. Notice that parameters  $a$ ,  $c$  and  
127  $d$  are indistinguishable among members; therefore the  $b_i$  can be seen as a weighting parameters that reflect the better or worse

128 performance of one model compared to the others. The NGR technique is implemented in R (R Core Team) using the packages  
 129 ensembleMOS (Yuen et al., 2018),

### 130 2.1.2. Affine Kernel Dressing

131 The affine kernel dressing method (Bröcker and Smith, 2008) only considers single model ensemble forecasts. It  
 132 estimates  $p(y|x, \theta)$  using a mixture of normally distributed variables

$$133 \quad p(y|x, \theta) = \frac{1}{m\sigma} \sum_{j=1}^m K\left(\frac{y-z_j}{\sigma}\right) \quad (3)$$

134 where  $K$  represents a standard normal density kernel ( $K(\xi) = 1/\sqrt{2\pi} \exp(-1/2\xi^2)$ ), centered at  $z_j$ , such that

$$135 \quad z_j = ax_j + r_1 + r_2\bar{x} \quad (4)$$

136 and,

$$137 \quad \sigma^2 = h_s^2(s_1 + s_2u(\mathbf{z})) \quad (5)$$

138 where  $h_s$  is the Silversman's factor (Bröcker and Smith, 2008),  $u(\mathbf{z})$  is the variance of  $\mathbf{z}$  and  $a, r_1, r_2, s_1, s_2$  are fitting  
 139 parameters obtained by minimizing the mean Ignorance score. For clarity we use the same nomenclature for the parameters as  
 140 in the original study. From Eqs. 4 and 5 we can obtain that the predictive variance  $v$  is a function of the ensemble variance  $S^2$   
 141 (Brocker and Smith, 2008)

$$142 \quad v = h_s^2s_1 + a^2(1 + h_s^2s_2)S^2 = c^* + d^*S^2 \quad (6)$$

143 Here,  $S^2$  represents the variance of the ensemble of exchangeable members.

144 The AKD technique is implemented through the SpecsVerification R package (Siegert, 2017).

### 145 2.1.3 Bayesian Model Averaging

146 The BMA method (Raftery et al. 2005, Fraley et al., 2010) also produces a mixture of normally distributed variables, as the  
 147 AKD method, but based on multi-model ensemble forecasts. In this case the predictive PDF is given by a weighted sum of  
 148 component PDFs,  $g_i(y|x_{i,j}; \theta_i)$ , one per each member:

$$149 \quad p(y|x, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} w_i g_i(y|x_{i,j}, \theta_i) \quad (7)$$

150 such that the weights and the parameters are invariable among members of the same model and

$$154 \quad \sum_{i=1}^n m_i w_i = 1$$

151 In the study the component PDFs are assumed normal as for the affine kernel dressing method. Estimates of  $w_i$ s and  $\theta_i$ s are  
 152 produced by maximizing the likelihood function using an Expectation Maximization algorithm (Casella and Berger, 2002).

153 The BMA technique is implemented through the ensembleBMA R package (Fraley et al., 2016).

## 155 **2.2 Measurement and forecast datasets**

156 ETo observations and forecasts were computed with the FAO-56 PM equation (Allen et al., 1998), from daily meteorological  
157 data as inputs. They covered the same period, between May and August from 2014 to 2016. The observations used daily  
158 measurements of minimum and maximum temperature, minimum and maximum relative humidity, wind speed, and surface  
159 incoming solar radiation from 101 U.S. Climate Reference Network (USCRN) weather stations. The USCRN stations are  
160 distributed over nine climatologically consistent regions in CONUS (Fig. 1). The ETo forecasts used daily maximum and  
161 minimum temperature, solar radiation, wind speed, and dew point temperature reforecasts of European Centre for Medium-  
162 Range Weather Forecasts model (ECMWF) outputs, United Kingdom Meteorological office model (UKMO) outputs, and  
163 National Centers for Environmental Prediction model (NCEP) from The International Grand Global Ensemble (TIGGE;  
164 Swinbank et al. 2016) database at each of these stations, considering a maximum lead time of 7 days. We used the same models  
165 as Medina et al. (2018) for comparison purposes, and because they are considered among the most skillful globally (e.g.  
166 Hagedorn et al., 2012). The forecasts were interpolated to the same  $0.5^\circ \times 0.5^\circ$  grid using the TIGGE data portal. The weekly  
167 forecasts accounted for the sum of the daily predictions generated at a specific day of each week, and the weekly observations  
168 considered the sum of the daily observations over the corresponding forecasting days, such that the weekly observations were  
169 independent from each other. In the study, we used the nearest neighbor approach to interpolate the forecasts to the USCRN  
170 stations, which does not account for the effects of elevation. While the use of interpolation techniques considering the effects  
171 of elevation (e.g. van Osnabrugge et al., 2019) may correct part of the forecasts errors before the post-processing, it could also  
172 affect the multivariate dependence of the weather variables. Hagedorn et al. (2012) showed that the post-processing can not  
173 only address the discrepancies related to the model's spatial resolution, but also serve as a means of downscaling the forecasts.

## 174 **2.3 Post-processing schemes**

### 175 **2.3.1 Training and verification periods**

176 The training data for the daily post-processing comprised the pairs of daily forecasts and corresponding observations from 30  
177 days prior to the forecast initial day, as in Medina et al. (2018). Instead, the training data for the weekly post-processing  
178 included all the other pairs of weekly forecasts and observations available for the forecast location, similarly as in the case of  
179 a leave one out cross validation framework. In the study both the daily and weekly forecasts were verified for events over  
180 June-August, 2014-2016.

### 181 **2.3.2 Baseline approaches**

182 Linear regression bias correction (BC) of the ECMWF forecast was used as a baseline approach for measuring the effectiveness  
183 of the NGR, the AKD and the BMA methods considering both daily and weekly forecasts. Here, the current forecasts bias is  
184 estimated as a linear function of the forecasts mean, and the members of the ensemble are shifted accordingly. The function is  
185 calibrated using the forecasts mean and the actual biases based on the same training periods as for the other post-processing

186 methods. Persistence is also used as a baseline approach for weekly forecasts, considering its applicability in productive  
 187 systems. In this case the ETo for a current week is estimated as the observed ETo during the previous week.

### 188 2.3.3 Forecasting Experiments

189 Table 1 summarizes the daily and weekly NWP-based ETo forecasting experiments based on different post-processing  
 190 methods and model combinations. The analyses of the daily forecasts put more emphasis on the differences among post-  
 191 processing methods. They include an examination of the effect of the duration of the training period on the forecasts  
 192 assessments as well as the regression weights from the tested post-processing methods. Whereas, the weekly forecasts put  
 193 more emphasis on the differences among the several single and multi-model ETo forecasts under baseline and probabilistic  
 194 post-processing.

### 195 2.4 Forecast verification metrics

196 In this study we use several metrics to evaluate deterministic and probabilistic forecast performance of the post-processed ETo  
 197 forecasts. For consistency purposes, the metrics of the tested methods were assessed using 50 random samples, i.e., same  
 198 number as members in the bias corrected ECMWF forecasts. Deterministic ETo forecast was produced by taking the average  
 199 of the ensemble members. The deterministic forecast performance was assessed using the bias or mean error (ME) and relative  
 200 ME (rME), the root mean square error (RMSE) and the relative RMSE (rRMSE), and the correlation ( $\rho$ ), which are common  
 201 measures of agreement in many studies. The absolute bias and relative bias are calculated and reported.

202 The ME and rME were computed as

$$203 \text{ ME} = \frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i) \quad (8)$$

$$204 \text{ rME} = \frac{\sum_{i=1}^n (\bar{f}_i - \sigma_i)}{n\bar{\sigma}} \quad (9)$$

205 where  $\bar{f}_i$  represents the average ensemble forecast for the event  $i$  ( $i = 1 \dots n$ ),  $\sigma_i$  is the corresponding observation, and  $\bar{\sigma}$  is the  
 206 mean observed data.

207 The RMSE and the rRMSE were computed as

$$208 \text{ RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i)^2} \quad (10)$$

$$209 \text{ rRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i)^2}}{\bar{\sigma}} \quad (11)$$

210 The correlation was obtained as

$$211 \rho = \frac{\sum_{i=1}^n (\bar{f}_i - \bar{f})(\sigma_i - \bar{\sigma})}{s_{\bar{f}} s_{\sigma}} \quad (12)$$

212 where  $\bar{f}$  is the mean of the average ensemble forecast and  $s_{\bar{f}}$  and  $s_{\sigma}$  are the standard deviation of the average forecasts and the  
 213 observations, respectively.

214 The probabilistic forecast performance was assessed using range histogram, the spread-skill relationship (see Wilks, 2011) and  
 215 the forecast coverage as measures of the forecast reliability, the Brier Skill Score (BSS) as a measure of the skill, and the  
 216 continuous rank probability score (CRPS), for providing an overall view of the performance (Hersbach, 2000), as it is sensitive  
 217 to both errors in location and spread simultaneously.

218 Reliability here refers to the statistical consistency (as in Toth et al. 2003), which is met when the observations are statistically  
 219 indistinguishable from the forecast ensembles (Wilks, 2011). To obtain the rank histogram, we get the rank of the observation  
 220 when merged into the ordered ensemble of ETo forecasts and then we plot the ranks histogram. The spread-skill relationships  
 221 are represented as binned-type plots (e.g. Pelosi et al., 2017), accounting for the mean of the ensemble standard deviation  
 222 deciles (as an indication of the ensemble spread) against the mean RMSE of the forecasts in each decile over the verification  
 223 period. The plots include the correlation between these two quantities. Calibrated ensembles should show a 1:1 relationship  
 224 between the standard deviations and the RMSE. If the forecasts are unbiased and the spread is small compared to the RMSE,  
 225 then the ensembles tend to be under-dispersive. The inverse of the spread provides an indication of sharpness, which is the  
 226 level of “compactness” of the ensemble (Wilks, 2011).

227 In addition to the spread skill relationship, we also report the ratio between the observed and nominal coverage (hereinafter  
 228 referred as coverage ratio). The coverage of a  $(1 - \alpha)100\%$ ,  $\alpha \in (0, 1)$ , central prediction interval is the fraction of  
 229 observations from the verification data set lying between  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the predictive distribution. It is  
 230 empirically assessed by considering the observations lying between the extreme values of the ensembles. The nominal or  
 231 theoretical coverage of a calibrated predictive distribution is  $(1 - \alpha)100\%$ . A calibrated forecast of  $m$  ensemble members  
 232 provides a nominal coverage of about  $(m - 1)/(m + 1) 100\%$  central prediction interval (e.g. Beran and Hall, 1993). For  
 233 example, an ensemble of 50 members provides 96% central prediction interval. The ratio between the observed and nominal  
 234 coverages provides a quantitative indicator of the quality of the forecasts dispersion under unbiasedness: a ratio lower (larger)  
 235 than 1 suggest that the forecasts tend to be under (over) dispersive.

236 The BSS is computed as

$$237 \text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{clim}}} \quad (13)$$

238 where BS is the Brier score of the forecast

$$239 \text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad (14)$$

240  $p$  is the forecast probability  $p$  of the event, which is estimated based on the ensemble, and  $o$  is equal to 1 if the event occurs  
 241 and 0 otherwise.

242  $\text{BS}_{\text{clim}}$  in Eq. 8 represents the Brier Score of the sample climatology, computed as (Wilks, 2010)

$$243 \text{BS}_{\text{clim}} = \bar{o}(1 - \bar{o}) \quad (15)$$

244 where  $\bar{o}$  is the sample climatology computed as the mean of the binary observations  $o_i$  in the verification dataset.

245 In this study we compute the BSS associated to the tercile events of the ETo forecasts (upper or 1st, middle or 2nd, and lower  
 246 or 3rd terciles). Therefore, the sample climatology is equal to  $0.3\bar{3}$  and  $\text{BS}_{\text{clim}} = 0.2\bar{2}$ .



247 The CRPS was computed as

$$248 \text{ CRPS} = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \left( F_i^f(h) - F_i^o(h) \right)^2 dh \quad (16)$$

249 where  $F^f$  and  $F^o$  are the cumulative distribution function of the forecast and the observations, respectively, and  $h$  represents  
250 the threshold value.  $F_i^o(h) = H(h - \sigma_i)$ ,  $H$  representing the Heaviside function, which is 0 for  $h < \sigma_i$  and 1 for  $h \geq \sigma_i$ .

## 251 **3 Results**

### 252 **3.1 Comparing the NGR, AKD and BMA methods at daily scale**

#### 253 **3.1.1 Deterministic forecast performance**

254 Figure 2 shows the rME and rRMSE as well as the correlation of the forecasts post-processed using different approaches over  
255 the southeast (SE) and northwest (NW) regions. These regions are representative of the Eastern and Western zones, which  
256 tended to provide the worse and best rRMSE and correlations, respectively. In general, the probabilistic post-processing  
257 methods add no additional skill to the deterministic forecast performance compared to the simple bias correction. While the  
258 rRMSE are relatively high, the rME are very low, which indicates that the errors are mostly random. The BMA and the simple  
259 linear regression methods provided lower bias than the NGR and AKD methods. Instead, the BMA method provided higher  
260 rRMSE and lower correlations than the other three methods at long lead times. The rRMSE and the correlations tended to be  
261 more variable among lead times and regions than among post-processing methods, while for the rME was the opposite. In  
262 addition, the changes in rRMSE and correlation with lead time tended to be larger over the Eastern regions.

#### 263 **3.1.2 Probabilistic forecast performance**

264 Figure 3 shows the spread skill relationship and the rank histograms using all pairs of forecasts and observations for lead days  
265 1 and 7. The spread-skill relationship shows that the probabilistic post-processing methods considerably improved the  
266 reliability of the ETo forecasts compared with the linear regression bias correction. The former methods tend to correct evident  
267 shortcomings of the ensemble raw forecasts which are unresolved by the simple post-processing, i.e., the considerable under-  
268 dispersion at short lead times, and the poor consistency between the ensemble spread and the RMSE at longer lead times. The  
269 adjustments had a low cost in terms of sharpness, judging by the range of ensemble spreads for the different line plots, but  
270 seemed slightly insufficient. The correlations between the ensemble standard deviation and the RMSE are fairly low,  
271 suggesting a limited predictive ability of the spread (Wilks, 2011). Nonetheless, they were consistently higher for probabilistic  
272 post-processing methods, compared to the linear regression method, and at short lead times, compared to the long lead times.  
273 The rank histograms in Figure 3 show that the probabilistic methods provided better calibration than the linear regression  
274 approach both at 1 and 7 days, but the improvements were considerably larger at 1 day. At the short lead time, the three  
275 methods slightly over-forecasted ETo, suggesting that the departures from the predictive mean has a negative skew, but in  
276 general they were fairly confident. In this case all the methods provided almost the same result. At the long lead time, there is

277 also an overestimation and then a positive bias, but also a slight U-shaped pattern, associated to some underdispersion for the  
278 range of the low and medium observations, which is coherent with the spread skill relationships. These issues are more  
279 pronounced using the BMA method and less pronounced using the AKD methods. Scheuerer and Büermann (2014) reported  
280 similar issues when post-processing ensemble forecasts of temperatures with the NGR method and a version of the BMA  
281 method. On the other hand, the calibration was affected little by the choice of a single or multi-model strategy for a given  
282 post-processing method. Nevertheless, the probabilistic methods provided a coverage ratio close to 100% independently of the  
283 lead time (see Table 2) and the region (not shown). The simple bias correction method instead provided coverage ratios much  
284 lower and more variable among regions (see Table 2) and lead times.

285 The NGR and AFK methods provided better Brier skill score (BSS) than the BC method for the three categories of ETo values,  
286 with improvements being higher for the middle tercile, than for the lower and upper terciles (Figure 4). The BMA based skill  
287 scores tended to decrease with lead time. On west regions (SW, W and NW) and at short lead days the multi-model ensemble  
288 forecasts post-processed with the NGR were the most skillful; in the other cases the ECMWF forecasts post-processed with  
289 the NGR and the AKD methods tended to be best. The differences of BSS among regions were larger at longer lead times  
290 because the skill decreased more sharply over the Eastern regions. This issue is slightly addressed by the NGR and AKD  
291 methods based on the ECMWF.

### 292 **3.1.3 Summary of average performance for daily forecast**

293 Table 2 shows the average performance for the lead days 1 and 7, by weighting the values of each metric according to the  
294 number of stations in each region. The ECMWF- UKMO forecasts post-processed with the NGR method were best at short  
295 lead times (1-2 days), while the ECMWF forecasts post-processed with the AKD and the NGR methods were the first and  
296 second best at the longer lead times. The BMA method performed well at short lead times but poorly at long times, while the  
297 simple bias correction method performed well for deterministic forecasts, but poorly for the probabilistic forecasts. The  
298 forecast performance across climate regions is also associated with the choice of the ECMWF ensemble forecasts or the multi-  
299 model ensemble forecasts (Table A1, ANEX). The single model ECMWF forecasts performed better over northern climate  
300 regions than the multi-model ensemble forecasts, while the multi-model did better than any single model forecast over the  
301 western regions. The performance over the other regions was more variable among strategies. The performance of the  
302 ECMWF- UKMO forecasts was generally better than that of the ECMWF-NCEP- UKMO forecasts (see Table A1, and Figs.  
303 2 and 4). Unlike other performance metrics, the coverage was mostly better for the ECMWF ensemble forecasts than for the  
304 multi-model ensemble forecasts. Our CRPS values is comparable with those reported by Osnabrugge (2019) based on the  
305 ECMWF ensemble forecasts of potential evapotranspiration over the Rhine basin, in Europe.

### 306 **3.1.3 Effect of the length of training period**

307 The choice of an “optimum” training period is an important issue related to the operational use of post-processing techniques  
308 for ETo forecasts. Here we compared the performance of different forecasts post-processed with NGR and AKD techniques

309 using 45 and 30 training days. The results suggest that the payoff from using 45 days is practically minimal. Table A2 (Anex)  
310 shows the percentage differences the forecasting performance of using 45 and 30 training days for post-processing. While  
311 there are generally some minor improvements for using 45 days than 30 days, which tend to be higher at longer lead times  
312 than shorter times, these improvements usually represent less than 3 percent of original statistics. The largest percentage  
313 difference, accounting for the BSS at the middle tercile, actually represented a negligible gain in absolute terms since they  
314 were affected by the close-to-zero range of the variable. The improvements were a bit higher for multi-model ensemble  
315 forecasts than for single model forecasts. Notice that, while testing two different periods may be limited to evaluate the  
316 methods' sensitivity to the training period, they comprised the range for which methods such as the NGR and BMA have been  
317 reported to provide stable results (Gneiting et al., 2005; Raftery et al., 2005).

### 318 3.1.4 Weighting coefficients

319 The weighting coefficients reflect both the performance of the ensemble models and the performance the post-processing  
320 techniques relative to their counterparts. Figure 5 shows the mean  $b_i$  (Eq. 1) weighting coefficients of the NGR technique and  
321  $w_i$  (Eq. 7) weighting coefficient of the BMA techniques for each region and lead time for the post-processed ECMWF-NCEP-  
322 UKMO, respectively. The coefficients for the NGR and BMA techniques exhibited some common patterns of variability across  
323 regions and lead times. Both methods show that the weights of the ECMWF forecasts are at overall the highest, with a clear  
324 maximum at medium lead times. The weights of the UKMO model are the highest at 1 and 2 days, but sharply decreases with  
325 the lead time, while the weights of the NCEP model are in general the lowest, although they consistently increase with lead  
326 time, most likely because of the stronger decrease of performance with lead time by the other two models. It explains well the  
327 most outstanding features of the performance assessments, in relation to the role of each model, and the dependence on regions  
328 and lead times. Compared to the NGR method, the BMA method gives the UKMO forecasts a higher relative weight, at the  
329 expense of the ECMWF forecast weights. For example, the weighting coefficients of the BMA method over the west regions  
330 are consistently higher for the UKMO forecasts than for the ECMWF forecasts. It suggests that the lower performance of the  
331 BMA post-processing relative to the NGR and the AKD methods may be related to a misrepresentation of the model weights  
332 on the performance. This in turn may be caused by convergence problems during the parameter optimization with the  
333 expectation-maximization algorithm (Vrugt et al., 2008).

334 We observed considerable similarities on the distribution of variance coefficients for the NGR method (Eq. 2) and the AKD  
335 (Eq. 6) method after post-processing the ECMWF forecasts. The two methods also provide very similar adjustments on the  
336 mean forecast because, unlike the BMA method, they independently bias correct the mean and optimize the spread-skill  
337 relationship, (Bröcker and Smith, 2008). However, in the experiment the NGR method was about 60 faster than the AKD  
338 method. The BMA method was also faster than the AKD method, but still considerably slower than the NGR method.  
339 Considering the effectiveness of the NGR method, and its versatility to post-process both single and multi-model ensemble  
340 forecasts, we applied this probabilistic technique to weekly ETo forecasts based on single model and multi-model ensembles.

## 341 3.2 Assessing NGR method for post-processing weekly ETo forecasts

### 342 3.2.1 Deterministic forecast assessments

343 As for the daily predictions, the bias, the RMSE and the correlation of the weekly forecasts post-processed with the NGR  
344 method and the linear regression methods were similar (Fig. 6). However, while the RMSE of daily forecasts based on ECMWF  
345 model varies between 12 and 20 % of the total ETo (Fig. 2), the RMSE for any of weekly forecasting strategies commonly  
346 varies between 8 and 11%, which is lower than for daily forecasts, making it more useful for operational purpose. The post-  
347 processed forecasts showed much lower RMSE and twice higher correlation than the predictions based on persistence, with  
348 the weekly predictions based on ECMWF forecasts being generally better, followed by the predictions based on the UKMO  
349 forecasts.

### 350 3.2.2 Probabilistic forecast assessments

351 Both the skill and the reliability of the weekly forecasts considerably improved through the NGR post-processing compared  
352 with the bias correction post-processing (Table 3). The improvements were different among ETo forecast models. In most  
353 cases, the better the forecasts performance, the lower the improvements are. The adjustments in the coverage ratio and the  
354 Brier skill score were about 2.5 and 5 times larger for the UKMO and the NCEP forecasts, respectively, than for the ECMWF  
355 forecasts. The bias corrected ECMWF forecasts are generally better than both the UKMO and NCEP forecasts post-processed  
356 with the NGR method. We found that the post-processing of the NCEP forecasts with methods like the NGR is almost  
357 mandatory to get reasonable probabilistic weekly forecasts of ETo. For example, the coverage ratio of the bias corrected  
358 forecasts on the West region was only 29%, because of the considerable under-dispersion. However, it is notable that, once  
359 they were post-processed with the NGR technique, they performed almost comparably to the UKMO forecasts post-processed  
360 with the same method, increasing the coverage ratio to 98.4%. Table 3 also shows that the multi-model ECMWF- UKMO  
361 weekly forecasts are commonly the best among all of those post-processed using the NGR method, followed by the ECMWF  
362 and the ECMWF-NCEP-UKMO forecasts.

363 The improvements in the reliability came through substantial adjustments both in the ensemble spread and spread-skill  
364 relationship of the raw forecasts (Fig. 7). The correlations between the standard deviation of the ensembles and the RMSE  
365 were more than twice larger through the NGR post-processing than through the linear regression bias correction. The  
366 adjustments seemed even slightly more effective than those resulting from the probabilistic post-processing of the daily  
367 forecasts (Fig. 3), although at the expense of a greater loss of sharpness. The contrasts in the post-processing effectiveness are  
368 probably associated with the differences in the training strategies.

369 In the case of the probabilistic forecast skill (Fig. 8), the improvements were larger for the middle tercile than for the other two  
370 terciles, similarly as with daily forecasts. Unlike the bias corrected forecasts, any of the probabilistically post-processed  
371 forecasts outperform climatology for practically any tercile and at any region. Maybe more importantly, the Brier scores for  
372 the lower and upper tercile events of the forecasts that have been post-processed with the NGR method is in most cases over

373 30% better than the scores of climatology. In the coast regions, from the South to the Northwest the score is commonly over  
374 50% better, similarly as for the daily forecasts. Finally, the improvements resulting from the use of multi-model ensemble  
375 forecasts compared to the single model ensemble forecasts were generally small, except for the Southwest region.

## 376 **4. Discussion**

### 377 **4.1 Effects of probabilistic post-processing on ETo forecasting performance**

378 This study showed that NGR, AKD and BMA post-processing schemes considerably improved the probabilistic forecast  
379 performance (coverage ratio, calibration, spread-skill, BSS, CRPS) of the daily and weekly ETo forecasts compared with the  
380 simple (i.e., using linear regression based on ensemble mean) bias correction method. While sharpness is a wished quality of  
381 any forecast, the daily and weekly bias corrected ETo forecasts from NWP are spuriously sharp, which leads to a poor  
382 consistency between the range of the ETo forecasts and the true values, and ultimately undermine the confidence on those  
383 forecasts. They also exhibit a poor consistency in that the variance of the ensembles are commonly insensitive to the size of  
384 the forecast error. The probabilistic post-processed methods provided a much better reliability, with a coverage which is close  
385 to the nominal value, and at a low cost on sharpness. Therefore, they lead to a much better agreement between the forecasted  
386 probability of having an ETo event between certain thresholds and the proportions of times that the event occurs (see Gneiting  
387 et al., 2005).

388 In the case of the weekly ETo forecasts, the rate of the improvements are considerably smaller for the ECMWF forecasts, than  
389 for the UKMO, and especially the NCEP forecasts. This seems to be largely due to the better performance of the ECMWF raw  
390 forecasts compared to the other forecasting systems. The probabilistic post-processing of the weekly NCEP forecasts seemed  
391 practically mandatory to produce reasonable predictions, but once implemented it provided performance assessments almost  
392 comparable to those based on the UKMO forecasts. These results have important implications for operational ETo forecasts,  
393 such as the U.S. national digital forecast database, one of the few operational products of its type, which are based on the  
394 NCEP forecasts.

395 Unlike the probabilistic forecast metrics, the deterministic metrics (ME, RMSE and correlation of the ensemble mean) are low  
396 sensitive to the form (deterministic or probabilistic) of post-processing. In particular, the RMSE and correlation seemed more  
397 affected by the choice of the single or multi-model ensemble forecast strategy than the choice between the NGR, the AKD or  
398 the simple bias correction as post-processing method. Whereas, RMSE and correlation provided by the BMA method are  
399 consistently worse at long lead times. The daily errors under any post-processing were relatively large, but mostly random,  
400 and therefore tend to cancel out at weekly scales. Therefore, while the RMSE varied between 12% and 20% of the daily totals,  
401 it represented between 8% and 11% of the weekly totals. The RMSE for weekly ETo forecasts were in all cases more than  
402 100% lower than for the persistence-based ETo forecasts, and potentially more skillful than the forecasts that exploit the  
403 temporal persistence of the ETo timeseries (e.g. Landaras et al., 2009; Mohan and Arumugam, 2009).

## 404 **4.2 Comparing the three probabilistic post-processing methods**

405 The NGR and AKD based post-processing methods for the ECMWF forecasts produced comparable results, indicating that  
406 the simple Gaussian predictive distribution from the NGR method represents fairly well the uncertainty of the ETo predictions.  
407 The methods led to similar distribution of the first two moments of the predictive probability function and similar performance  
408 statistics (with the AKD based forecasts being just slightly better). However, the NGR method is more versatile since it can  
409 be applied to correct both single model and multi-model ensemble forecasts, while the AKD method can only be applied to  
410 correct single model forecast. The NGR based predictive distribution function is also easier to interpret than the AKD based  
411 predictive distribution, which is given by an averaged sum of standard Gaussians.

412 The BMA method performed slightly less desirable compared to the NGR and AKD presumably due to issues with the  
413 parameter identifiability. The implemented method uses the Expectation-Maximization (EM) algorithm to produce maximum  
414 likelihood estimates of the fitting coefficients, which is susceptible to converge to local minima, especially when dealing with  
415 multi-model ensemble forecasts with very different ensemble sizes (Vrugt et al., 2008). Archambeau et al. (2003) demonstrated  
416 that, in presence of outliers or repeated values, this algorithm tends to identify local maximums of the likelihood of the  
417 parameters of a Gaussian mixture model. Tian X. et al. (2012) found that adjusted BMA coefficients using both a quasi-  
418 Newtonian limited memory algorithm and the Markov Chain Monte Carlo were more accurate than those fitted with the EM  
419 algorithm, a procedure that is worth testing in future studies.

## 420 **4.3 Multi-model ensemble versus single model ensemble forecasts**

421 Daily multi-model ensemble forecasts performed better (in terms of ME, RMSE, correlation, CRPS and BSS) than daily  
422 ECMWF forecasts at short lead times (1-2 days) and over the western and southern regions, while the ECMWF forecasts are  
423 better over the northeastern regions for longer lead times. For other region/lead time combinations the performance of single  
424 and multi-model ensemble forecasts did not differ much. We observed similar patterns for the raw and simple bias corrected  
425 forecasts (Medina et al., 2018). Whereas, the weekly multi-model ensemble forecast where consistently better than the weekly  
426 single-model forecasts only in the Southwest region, seemingly because the weekly forecasts logically involve both short and  
427 long lead time assessments, and the effectiveness of the multi-models is degraded for long lead times. The observed behavior  
428 is associated with the performance of the ECMWF forecasts relative to the UKMO forecasts. While the ECMWF forecasts are  
429 in general better than the UKMO and NCEP forecasts, they are much better over the northeastern regions for medium lead  
430 times (4-6 days). The UKMO forecasts are in many cases the best at 1 and 2 lead days, but tend to be the worst at the longest  
431 times (6-7 days), especially over these regions. The NCEP forecasts had a small contribution with respect to the ECMWF and  
432 UKMO forecasts at short lead times. These forecasts are comparatively better at longer lead times, but still keep a minor role  
433 with regard to the ECMWF forecasts.

434 When considering daily forecasts we adopted a length of the training period of 30 days and showed that by increasing the  
435 length to 45 days the improvements were small (commonly lower than three percent). This seems a plausible range for future

436 works and represents an obvious advantage upon methods such as the analog forecast, which provide similar performance  
437 (Tian and Martinez 2012 a, b, 2014) but require long training datasets. Gneiting et al. (2005) and Wilson (2007) found that  
438 lengths between 30 and 40 days provided good and almost constant performance assessments of sea level pressure forecasts  
439 post-processed with the NGR method, and temperatures forecasts post-processed with the BMA method, respectively.

#### 440 **4.4. Post-processing the individual inputs versus post-processing ETo**

441 While in this study we considered the post-processing of ETo ensembles produced with raw NWP forecasts, a question is if  
442 by post-processing the forcing variables such as temperature, radiation and wind speed first, and then computing the ETo, we  
443 might have better predictions. The NGR method has been shown to be successful for the post-processing of surface  
444 temperatures (e.g. Wilks and Hamill, 2007), whose distribution is fairly Gaussian. For example, Hagedorn (2008) and  
445 Hagedorn et al. (2008) showed gains in lead time between two days and four days, with the gains being larger over areas where  
446 the raw forecast showed poor skill. Kann et al., (2009) and Kann et al., (2011), used the NGR method for improving short  
447 range ensemble forecasts of 2m-temperature. Recently, Scheuerer and Büermann (2014) provided a generalization of the  
448 original approach of Gneiting et al. (2005) that produces spatially calibrated probabilistic temperature forecasts. The wind-  
449 speed forecasts have been commonly post-processed with the use of quantile regression method (e.g. Bremnes 2004; Pinson  
450 et al. 2007; Møller et al., 2008). More recently Sloughter et al. (2010) extended the original BMA method of Raftery et al.  
451 (2005) for wind speed, by considering a gamma distribution for modeling the distribution of every member of the ensemble,  
452 which considerably improved the CRPS, the absolute errors and the coverage. Whereas, Vanvyve et al., (2015) and Zhang et  
453 al. (2015) used the analog method following the methodology of Delle Monache (2013). The accurate solar radiation  
454 forecasting is particularly challenging because it requires detailed representation of the cloud fields (Verzijlbergh et al., 2015),  
455 which is usually not well resolved by the NWP models. Davò et al. (2016) used artificial neural networks (ANN) and the  
456 analog method approaches for the post-processing of both wind speed and solar radiation ensemble forecasts, which  
457 outperformed a simple bias correction approach. However, the post-processing of meteorological forecasts for producing ETo  
458 ensemble forecasts may require accounting for the multivariate dependence among those forcing, which is often difficult (e.g.  
459 Wilks, 2015). Kang et al (2010) found that post-processing of the streamflow forecasts provided more accurate predictions  
460 than post-processing the forcing alone, while Vekade et al (2013) showed that the improvements in precipitation and  
461 temperature through the post-processing hardly benefited the streamflow forecasts. Lewis et al., 2014 showed that the  
462 performance of the ETo forecasts can largely surpass that of the individual input variables. Therefore, it is unclear if we can  
463 have any benefit by using the post-processed inputs, instead of the raw forecasts, to construct ETo forecasts.

#### 464 **4.5. Future outlook**

465 It is worth noting that, while the ETo forecasts are produced for being used in agriculture, they were tested over USCRN  
466 stations, which are not representative of agricultural settings. In real applications, the bias between the forecasts with no post-  
467 processing and the measurements based on agricultural stations could be higher than the bias resolved in this study. A question

468 that should be addressed in the future studies is to what extent the improvements of the predictive distribution of the ETo  
469 forecasts can be translated into a more reliable representation of the crop water use in agricultural lands and, ultimately, in  
470 water savings and economic gains. Since the ETo estimations can have remarkable impacts on the soil moisture estimations  
471 (Rodriguez-Iturbe et al., 1999), we envision that new studies relying on the combination of rainfall and ETo forecasts post-  
472 processed with probabilistic methods will lead to considerable reductions on the uncertainty of soil moisture forecasts. New  
473 attempts should also investigate the role of the state of art probabilistic post-processing techniques on ETo forecasts produced  
474 from regional numerical weather prediction models, which have had improved spatial resolution and already been used in  
475 different meteorological services (e.g. Baldauf et al. 2011; Seity et al. 2011; Hong and Dudhia, 2012; Bentzien and Friederichs,  
476 2012).

## 477 **5. Conclusions**

478 This study for the first time evaluated probabilistic methods based on NGR, AKD, and BMA techniques for post-processing  
479 daily and weekly ETo forecasts derived from single or multi-model ensemble numerical weather predictions. The different  
480 ETo post-processing methods were compared against the simple linear regression bias correction method using both daily and  
481 weekly forecasts, and also against persistence in the case of weekly forecasts. The probabilistic post-processing techniques  
482 largely modified the spread of the original ETo forecasts, with very favorably impacts on the probabilistic forecast  
483 performance. They corrected the notable under-dispersion and the poor consistency between the spread of the ETo forecasts  
484 and the dimension of the errors, leading to better BSS, reliability (both coverage ratio and spread-skill) and CRPS. The  
485 adjustments were crucial on the performance of the weekly NCEP forecasts, followed by the weekly UKMO forecasts, whose  
486 bias corrected versions show a clear disadvantage compared with simply post-processed ECMWF forecasts.

487 The deterministic performance based on the NGR, AKD and BMA methods were comparable to the performance based on the  
488 linear regression bias correction for both daily and weekly forecasts, and the skill is about 100% higher than those based on  
489 persistence in the case of the weekly forecasts. The rRMSE are between 12 and 20% for the daily totals and 8 and 11% for the  
490 weekly totals. The NGR and AKD provided similar estimates of the first and second order moments of the predictive density  
491 distribution; they showed similar effectiveness, but the NGR method has the advantage that can post-process both single and  
492 multi-model ensemble forecasts. Both NGR and AKD post-processing methods outperformed the BMA method when  
493 considering daily forecasts at long lead times.

494 The multi-model ensemble forecasting provided benefits at daily scales compared to the ECMWF ensemble forecasting, while  
495 the benefits were marginal at weekly scales. The multi-model ensemble forecasting seems a better choice when the UKMO  
496 forecasts are comparable or slightly better than the ECMWF forecasts, such as at short (1-2 days) lead times and over the  
497 southern and western regions. Post-processing single model forecast is a better choice than post-processing multi-model  
498 ensemble forecast in the circumstances where the ECMWF forecasts perform considerably better than the UKMO and NCEP,  
499 such as at mid and long lead times, especially over the northeastern regions. While we considered a length of the training



500 period of 30 days for daily post-processing, the increase of the training period to 45 days only led to minimal improvements.  
501 In conclusion, our results suggest that the NGR post-processing of ETo forecasts generated from the ECMWF or ECMWF-  
502 UKMO predictions is the most plausible strategy among those being evaluated, and is recommended for operational  
503 implementations, because accuracy and reliability requirements for practical applications have not been discussed.

#### 504 **Acknowledgement**

505 This research was supported in part by the Alabama Agricultural Experiment Station and the Hatch program of the National  
506 Institute of Food and Agriculture (NIFA), U.S. Department of Agriculture (USDA, Access No. 1012578), by the Auburn  
507 University Intramural Grant Program, by the Auburn University Presidential Awards for Interdisciplinary Research, and by  
508 the USDA-NIFA Agriculture and Food Research Initiative (AFRI) competitive grant. The authors want to thank the very  
509 helpful comments of the reviewers

#### 510 **Code/Data availability**

511 Request for materials should be addressed to Di Tian.

#### 512 **Author contributions**

513 Hanoi Medina and Di Tian designed and conceptualized the research. Hanoi Medina implemented the design, performed data  
514 curation, analysis, validation, visualization, and wrote the original draft. Di Tian supervised the research, contributed by advice,  
515 and reviewed and edited the manuscript.

#### 516 **Competing interests**

517 The authors declare that they have no conflict of interest.

#### 518 **References**

- 519 Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water  
520 requirements-FAO, Irrigation and drainage paper 56, Fao, Rome, 300(9), p.D05109, 1998.
- 521 Archambeau, C., Lee, J. A. and Verleysen, M.: On Convergence Problems of the EM Algorithm for Finite Gaussian Mixtures,  
522 In ESANN (Vol. 3, pp. 99-106), 2003.

523 Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T.: Operational convective-scale  
524 numerical weather prediction with the COSMO model: Description and sensitivities, *Monthly Weather Review*, 139(12),  
525 pp.3887-3905, 2011.

526 Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction, *Nature*, 525(7567): 47-55, 2015.

527 Bentzien, S. and Friederichs, P.: Generating and calibrating probabilistic quantitative precipitation forecasts from the high-  
528 resolution NWP model COSMO-DE. *Weather and Forecasting*, 27(4), pp.988-1002, 2012.

529 Beran, R. and Hall, P.: Interpolated nonparametric prediction intervals and confidence intervals, *Journal of the Royal Statistical*  
530 *Society, Series B (Methodological)*, pp.643-652, 1993.

531 Bremnes, J. B.: Probabilistic Wind Power Forecasts Using Local Quantile Regression, *Wind Energy*, 7, 47–54, 2004.

532 Bröcker, J. and Smith, L. A.: From ensemble forecasts to predictive distribution functions, *Tellus A: Dynamic Meteorology*  
533 *and Oceanography*, 60(4), pp.663-678, 2008.

534 Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M.: A comparison of the ECMWF, MSC, and NCEP  
535 global ensemble prediction systems, *Monthly Weather Review*, 133(5), pp.1076-1097, 2005.

536 Casella, G. and Berger, R. L. : *Statistical inference (Vol. 2)*. Pacific Grove, CA: Duxbury, 2002.

537 Castro, F. X., Tudela, A. and Sebastià, M. T.: Modeling moisture content in shrubs to predict fire risk in Catalonia (Spain).  
538 *Agricultural and Forest Meteorology*, 116(1-2), pp.49-59, 2003.

539 Chirico, G. B., Pelosi, A., De Michele, C., Bolognesi, S. F. and D'Urso, G.: Forecasting potential evapotranspiration by  
540 combining numerical weather predictions and visible and near-infrared satellite images: an application in southern Italy, *The*  
541 *Journal of Agricultural Science*, pp.1-9. <https://doi.org/10.1017/S0021859618000084>, 2018.

542 Davò, F., Alessandrini, S., Sperati, S., Delle Monache, L., Airolidi, D. and Vespucci, M. T.: Post-processing techniques and  
543 principal component analysis for regional wind power and solar irradiance forecasting, *Solar Energy*, 134, pp.327-338, 2016

544 Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B. and Searight, K.: Probabilistic weather prediction with an analog  
545 ensemble, *Monthly Weather Review*, 141(10), pp.3498-3516, 2013.

546 Fraley, C., Raftery, A. E. and Gneiting, T.: Calibrating multimodelmulti-model forecast ensembles with exchangeable and  
547 missing members using Bayesian model averaging, *Monthly Weather Review*, 138(1), pp.190-202, 2010.

548 Fraley, C., Raftery, A. E., Slougher, J. M., Gneiting T.: EnsembleBMA: Probabilistic Forecasting using Ensembles and  
549 Bayesian Model Averaging. R package version 5.1.3. <https://CRAN.R-project.org/package=ensembleBMA>, 2016.

550 Glahn, H. R. and Lowry, D. A.: The use of model output statistics (MOS) in objective weather forecasting. *J Appl Meteorol*,  
551 11(8): 1203-1211, 1972.

552 Glahn, H. R. and Ruth, D. P.: The new digital forecast database of the National Weather Service, *Bulletin of the American*  
553 *Meteorological Society*, 84(2), pp.195-202, 2003.

554 Gneiting, T., Raftery, A. E., Westveld III, A. H. and Goldman, T.: Calibrated probabilistic forecasting using ensemble model  
555 output statistics and minimum CRPS estimation, *Monthly Weather Review*, 133(5), 1098-1118., 2005.

556 Gneiting, T.: Calibration of medium-range weather forecasts, European Centre for Medium-Range Weather Forecasts,  
557 Technical Memorandum No. 71, 2014.

558 Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M. and Palmer, T. N.: Comparing TIGGE multimodelmulti-model  
559 forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Q J Roy Meteor Soc*, 138(668): 1814-1827, 2012.

560 Hagedorn, R., Hamill, T. M. and Whitaker, J. S.: Probabilistic forecast calibration using ECMWF and GFS ensemble  
561 reforecasts. Part I: Two-meter temperatures. *Monthly Weather Review*, 136, 2608–2619, 2008.

562 Hagedorn, R.: Using the ECMWF reforecast data set to calibrate EPS forecasts. *ECMWF Newsletter*, 117, 8–13, 2008.

563 Hamill, T. M. and Colucci, S. J.: Verification of Eta–RSM short-range ensemble forecasts, *Monthly Weather Review*, 125(6),  
564 pp.1312-1327, 1997.

565 Hamill, T. M. and Whitaker, J. S.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and  
566 application, *Mon Weather Rev*, 134(11): 3209-3229, 2006.

567 Hamill, T. M. et al.: Noaa's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *B Am Meteorol Soc*,  
568 94(10): 1553-1565, 2013.

569 Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and*  
570 *Forecasting*, 15(5), pp.559-570, 2000.

571 Hobbins, M., McEvoy, D. and Hain, C.: Evapotranspiration, evaporative demand, and drought, *Drought and Water Crises:*  
572 *Science, Technology, and Management Issues*, pp.259-288, 2017.

573 Hong, S. Y. and Dudhia, J.: Next-generation numerical weather prediction: Bridging parameterization, explicit clouds, and  
574 large eddies, *Bulletin of the American Meteorological Society*, 93(1), pp.ES6-ES9., 2012.

575 Ishak, A. M., Bray, M., Remesan, R. and Han, D.: Estimating reference evapotranspiration using numerical weather modelling,  
576 *Hydrological processes*, 24(24), pp.3490-3509, 2010.

577 Kang, T. H., Kim, Y. O. and Hong, I. P.: Comparison of pre - and post - processors for ensemble streamflow prediction,  
578 *Atmospheric Science Letters*, 11(2), pp.153-159, 2010.

579 Kann, A., Haiden, T. and Wittmann, C.: Combining 2-m temperature nowcasting and short-range ensemble forecasting,  
580 *Nonlinear Processes in Geophysics*, 18, 903–910, 2011.

581 Kann, A., Wittmann, C., Wang, Y. and Ma, X.: Calibrating 2-m temperature of limited-area ensemble forecasts using high-  
582 resolution analysis. *Monthly Weather Review*, 137, 3373–3387, 2009.

583 Klein, W. H. and Glahn, H. R.: Forecasting local weather by means of model output statistics, *Bulletin of the American*  
584 *Meteorological Society*, 55(10), pp.1217-1227, 1974.

585 Landaras, G., Ortiz-Barredo, A. and López, J. J.: Forecasting weekly evapotranspiration with ARIMA and artificial neural  
586 network models, *Journal of irrigation and drainage engineering*, 135(3), pp.323-334, 2009.

587 Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, *Journal of Computational Physics*, 227(7), pp.3515-3539, 2008.

588 Madadgar, S., Moradkhani, H. and Garen, D.: Towards improved post-processing of hydrologic forecast ensembles.  
589 *Hydrological Processes*, 28(1), pp.104-122, 2014.

590 Mase, A. S. and Prokopy, L. S.: Unrealized potential: A review of perceptions and use of weather and climate information in  
591 agricultural decision making, *Weather, Climate, and Society*, 6(1), pp.47-61, 2014.

592 Medina, H., Tian, D., Marin, F. R. and Chirico, G. B.: Comparing GEFs, ECMWF, and Postprocessing Methods for Ensemble  
593 Precipitation Forecasts over Brazil, *Journal of Hydrometeorology*, 20(4), pp.773-790, 2019.

594 Medina, H., Tian, D., Srivastava, P., Pelosi, A. and Chirico, G. B.: Medium-range reference evapotranspiration forecasts for  
595 the contiguous United States based on multimodelmulti-model numerical weather predictions, *Journal of Hydrology*, 562,  
596 pp.502-517, 2018.

597 Messner, J. W., Mayr, G. J., Zeileis, A. and Wilks, D. S.: Heteroscedastic Extended Logistic Regression for Postprocessing of  
598 Ensemble Guidance. *Mon. Wea. Rev.*, 142, 448–456, <https://doi.org/10.1175/MWR-D-13-00271.1>, 2014.

599 Mohan, S. and Arumugam, N.: Forecasting weekly reference crop evapotranspiration series, *Hydrological sciences journal*,  
600 40(6), pp.689-702, 1995.

601 Møller, J. K., Nielsen, H. A., and Madsen, H.: Time-Adaptive Quantile Regression, *Computational Statistics & Data Analysis*,  
602 52, 1292–1303, 2008.

603 National Research Council of the National Academies: Completing the Forecast: Characterizing and Communicating  
604 Uncertainty for Better Decisions Using Weather and Climate Forecasts, The National Academies Press, 124 pp, 2006.

605 Osnabrugge, B. V., Uijlenhoet, R. and Weerts, A.: Contribution of potential evaporation forecasts to 10-day streamflow  
606 forecast skill for the Rhine River. *Hydrology and Earth System Sciences*, 23(3), pp.1453-1467, 2019.:

607 Pelosi, A., Medina, H., Van den Bergh, J., Vannitsem, S., and Chirico, G. B.: Adaptive Kalman filtering for post-processing  
608 ensemble numerical weather predictions, *Mon Weather Rev*, [doi.org/10.1175/MWR-D-17-0084.1](https://doi.org/10.1175/MWR-D-17-0084.1), 2017.

609 Pelosi, A., Medina, H., Villani, P., D’Urso, G. and Chirico, G. B.: Probabilistic forecasting of reference evapotranspiration  
610 with a limited area ensemble prediction system, *Agricultural water management*, 178, pp.106-118, 2016.

611 Perera, K. C., Western, A. W., Nawarathna, B. and George, B.: Forecasting daily reference evapotranspiration for Australia  
612 using numerical weather prediction outputs, *Agr Forest Meteorol*, 194: 50-63, 2014.

613 Pinson, P., and Madsen, H.: Ensemble-Based Probabilistic Forecasting at Horns Rev, *Wind Energy*, 12, 137–155, 2009.

614 Prokopy, L. S., Haigh, T., Mase, A. S., Angel, J., Hart, C., Knutson, C., Lemos, M. C., Lo, Y. J., McGuire, J., Morton, L. W.  
615 and Perron, J.: Agricultural advisors: a receptive audience for weather and climate information?, *Weather, Climate, and*  
616 *Society*, 5(2), pp.162-167, 2013.

617 R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna,  
618 Austria, <http://www.R-project.org/>, 2014.

619 Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian model averaging to calibrate forecast  
620 ensembles, *Monthly Weather Review*, 133(5), pp.1155-1174, 2005.

621 Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V. and Cox, D. R.: Probabilistic modelling of water balance at a point:  
622 the role of climate, soil and vegetation, *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and*  
623 *Engineering Sciences*, 455(1990), pp.3789-3805, 1999.

624 Roulston, M. S. and Smith, L. A.: Combining dynamical and statistical ensembles, *Tellus A: Dynamic Meteorology and*  
625 *Oceanography*, 55(1), pp.16-30, 2003.

626 Scheuerer, M. and Büermann, L.: Spatially adaptive post - processing of ensemble forecasts for temperature, *Journal of the*  
627 *Royal Statistical Society: Series C (Applied Statistics)*, 63(3), pp.405-422, 2014.

628 Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V.: The AROME-France  
629 convective-scale operational model, *Monthly Weather Review*, 139(3), pp.976-991, 2011.

630 Siegert, S.: SpecsVerification: Forecast Verification Routines for Ensemble Forecasts of Weather and Climate. R package  
631 version 0.5-2. <https://CRAN.R-project.org/package=SpecsVerificatio>, 2017.

632 Silva, D., Meza, F. J. and Varas, E.: Estimating reference evapotranspiration (ET<sub>o</sub>) using numerical weather forecast data in  
633 central Chile, *Journal of hydrology*, 382(1-4), pp.64-71, 2010.

634 Sloughter, J. M., Gneiting, T. and Raftery, A. E.: Probabilistic wind speed forecasting using ensembles and Bayesian model  
635 averaging, *Journal of the american statistical association*, 105(489), pp.25-35, 2010.

636 Swinbank, R. et al.: The Tigge Project and Its Achievements. *B Am Meteorol Soc*, 97(1): 49-67, 2016.

637 Tian, D. and Martinez, C. J.: Comparison of two analog-based downscaling methods for regional reference evapotranspiration  
638 forecasts, *J Hydrol*, 475: 350-364, 2012a

639 Tian, D. and Martinez, C. J.: Forecasting Reference Evapotranspiration Using Retrospective Forecast Analogs in the  
640 Southeastern United States, *J Hydrometeorol*, 13(6): 1874-1892, 2012b

641 Tian, D. and Martinez, C. J.: The GEFS-based daily reference evapotranspiration (ET<sub>o</sub>) forecast and its implication for water  
642 management in the southeastern United States. *J Hydrometeorol*, 15(3): 1152-1165, 2014.

643 Tian, X., Xie, Z., Wang, A. and Yang, X.: A new approach for Bayesian model averaging, *Science China Earth Sciences*,  
644 55(8), 1336-1344, 2012.

645 Toth, Z., Talagrand, O., Candille, G. and Zhu, Y.: Probability and ensemble forecasts, *Forecast Verification: A Practitioner's*  
646 *Guide in Atmospheric Science*, pp.137-163, 2003.

647 Vanvyve, E., Delle Monache, L., Monaghan, A.J. and Pinto, J.O.: Wind resource estimates with an analog ensemble approach,  
648 *Renewable Energy*, 74, pp.761-773, 2015.

649 Verkade, J. S., Brown, J. D., Reggiani, P. and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble  
650 reforecasts for operational hydrologic forecasting at various spatial scales, *Journal of Hydrology*, Volume 501,2013, Pages 73-  
651 91,<http://dx.doi.org/10.1016/j.jhydrol.2013.07.039>, 2013.

652 Verkade, J. S., Brown, J. D., Reggiani, P. and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble  
653 reforecasts for operational hydrologic forecasting at various spatial scales, *Journal of Hydrology*, 501, pp.73-91, 2013.

654 Verzijlbergh, R. A., Heijnen, P. W., de Roode, S. R., Los, A. and Jonker, H. J.: Improved model output statistics of numerical  
655 weather prediction based irradiance forecasts for solar power applications, *Solar Energy*, 118, pp.634-645, 2015

656 Vrugt, J. A., Diks, C. G. and Clark, M. P.: Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling,  
657 *Environmental fluid mechanics*, 8(5-6), pp.579-595, 2008.

658 Wang, X. and Bishop, C. H.: Improvement of ensemble reliability with a new dressing kernel, Quarterly Journal of the Royal  
659 Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131(607),  
660 pp.965-986, 2005.

661 Whan, K. and Schmeits, M: Comparing Area Probability Forecasts of (Extreme) Local Precipitation Using Parametric and  
662 Machine Learning Statistical Postprocessing Methods. Mon. Wea. Rev., 146, 3651–3673, [https://doi.org/10.1175/MWR-D-](https://doi.org/10.1175/MWR-D-17-0290.1)  
663 [17-0290.1](https://doi.org/10.1175/MWR-D-17-0290.1), 2018.

664 Wilks, D. S. and Hamill, T. M.: Comparison of ensemble-MOS methods using GFS reforecasts, Monthly Weather Review,  
665 135(6), pp.2379-2390, 2007.

666 Wilks, D. S.: Comparison of ensemble-MOS methods in the Lorenz'96 setting, Meteorological Applications, 13(3), pp.243-  
667 256, 2006.

668 Wilks, D. S.: Extending logistic regression to provide full probability distribution MOS forecasts, Meteorological Applications:  
669 A journal of forecasting, practical applications, training techniques and modelling, 16(3), pp.361-368, 2009.

670 Wilks, D. S.: Multivariate ensemble Model Output Statistics using empirical copulas, Quarterly Journal of the Royal  
671 Meteorological Society, 141(688), pp.945-952, 2015.

672 Wilks, D.S.: Sampling distributions of the Brier score and Brier skill score under serial dependence, Q J Roy Meteor Soc,  
673 136(653): 2109-2118, 2010.

674 Williams, R. M., Ferro, C. A. T. and Kwasniok, F.: A comparison of ensemble post-processing methods for extreme events.  
675 Quarterly Journal of the Royal Meteorological Society, 140(680), pp.1112-1120, 2014.

676 Wilson, L. J., Bearegard, S., Raftery, A. E. and Verret, R.: Calibrated surface temperature forecasts from the Canadian  
677 ensemble prediction system using Bayesian model averaging, Monthly Weather Review, 135(4), pp.1364-1385, 2007.

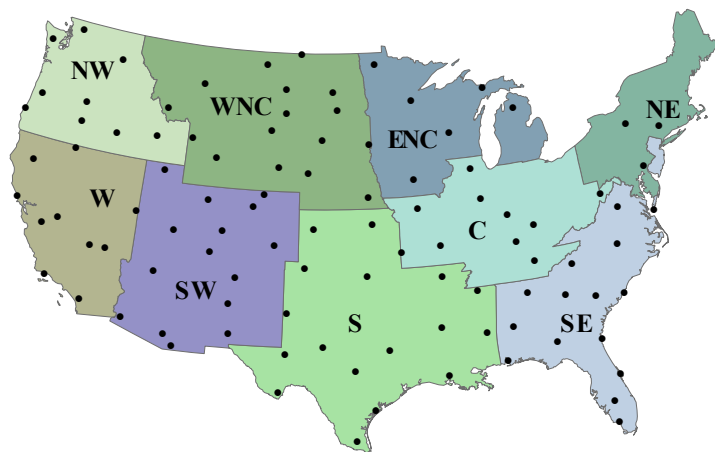
678 Yuen, R., Baran, S., Fraley, C., Gneiting, T., Lerch, S., Scheuerer, M., and Thorarinsdottir, T.: ensembleMOS: Ensemble  
679 Model Output Statistics. R package version 0.8.2. <https://CRAN.R-project.org/package=ensembleMOS>, 2018.

680 Zhang, J., Draxl, C., Hopson, T., Delle Monache, L., Vanvyve, E. and Hodge, B. M.: Comparison of numerical weather  
681 prediction based deterministic and probabilistic wind resource assessment methods, Applied Energy, 156, pp.528-541, 2015.

682 Zhao, T., Wang, Q. J. and Schepen, A.: A Bayesian modelling approach to forecasting short-term reference crop  
683 evapotranspiration from GCM outputs, Agricultural and Forest Meteorology, 269, pp.88-101, 2019.

684

685



686

687 Figure 1. U.S. climate regions: NW (North West), WNC (West North Central), ENC (East North Central), NE (North East),  
688 C (Central), SE (South East), C (Central), S (South), SW (South West), W (West). The circles represent the sampled USCRN  
689 stations in the experiment.

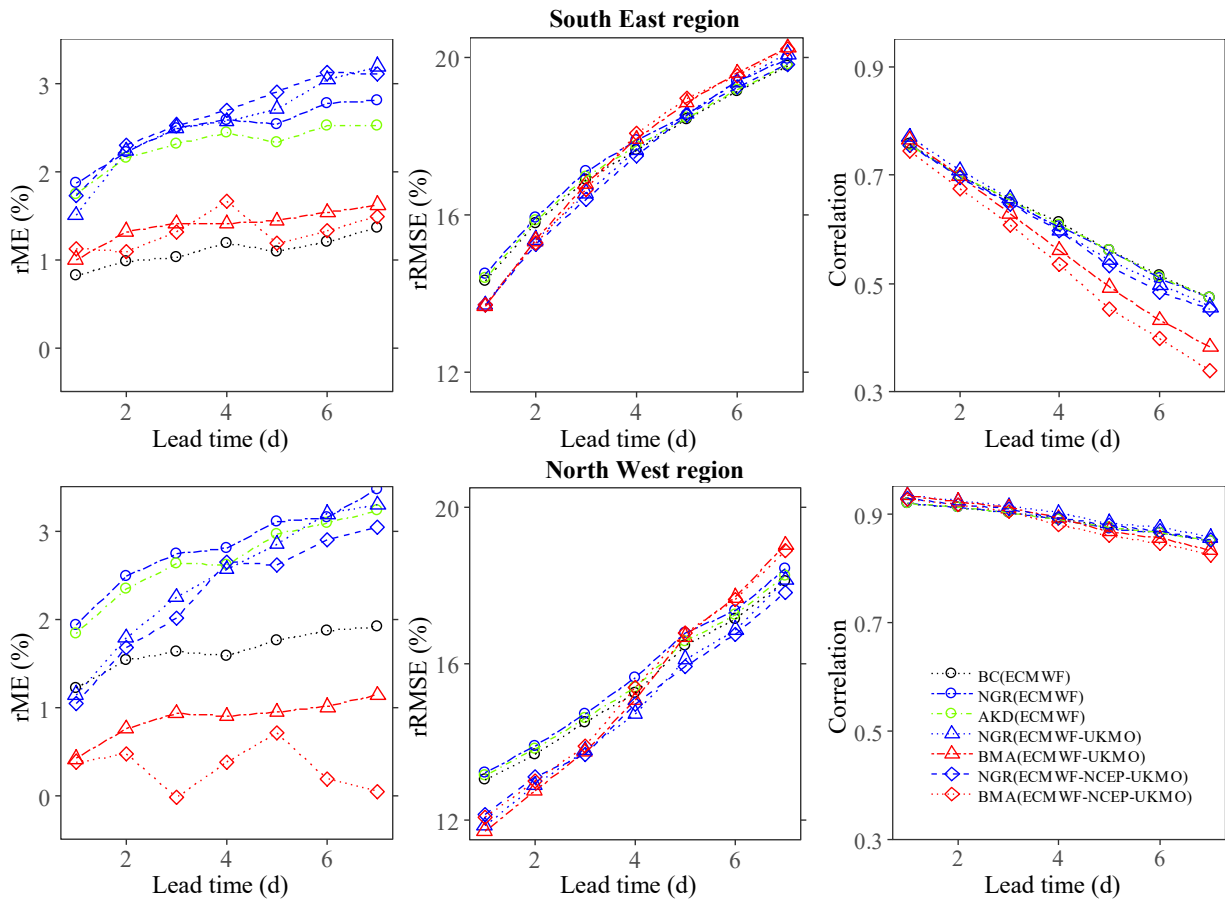


Figure 2. Relative mean error (rME), relative root mean square error (rRMSE), and correlation considering daily forecasts for different lead times over the SE and NW regions.



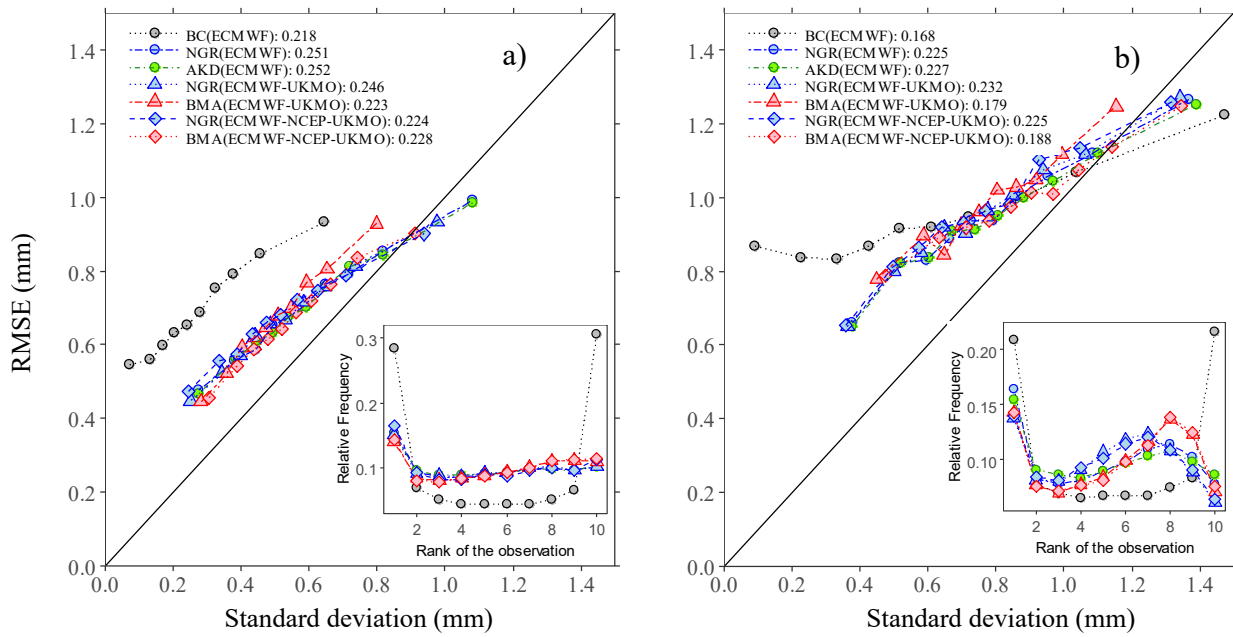


Figure 3. Binned spread-skill plots accounting for the mean of the ensemble standard deviation deciles against the mean RMSE of the forecasts in each decile over the verification period based on all pairs of forecasts and observations at a) 1-day and b) 7-day lead. The panel in the right and the bottom shows the corresponding rank histograms. The correlation between the standard deviations and the absolute errors is reported after the colon. The solid line represents the 1:1 relationship.

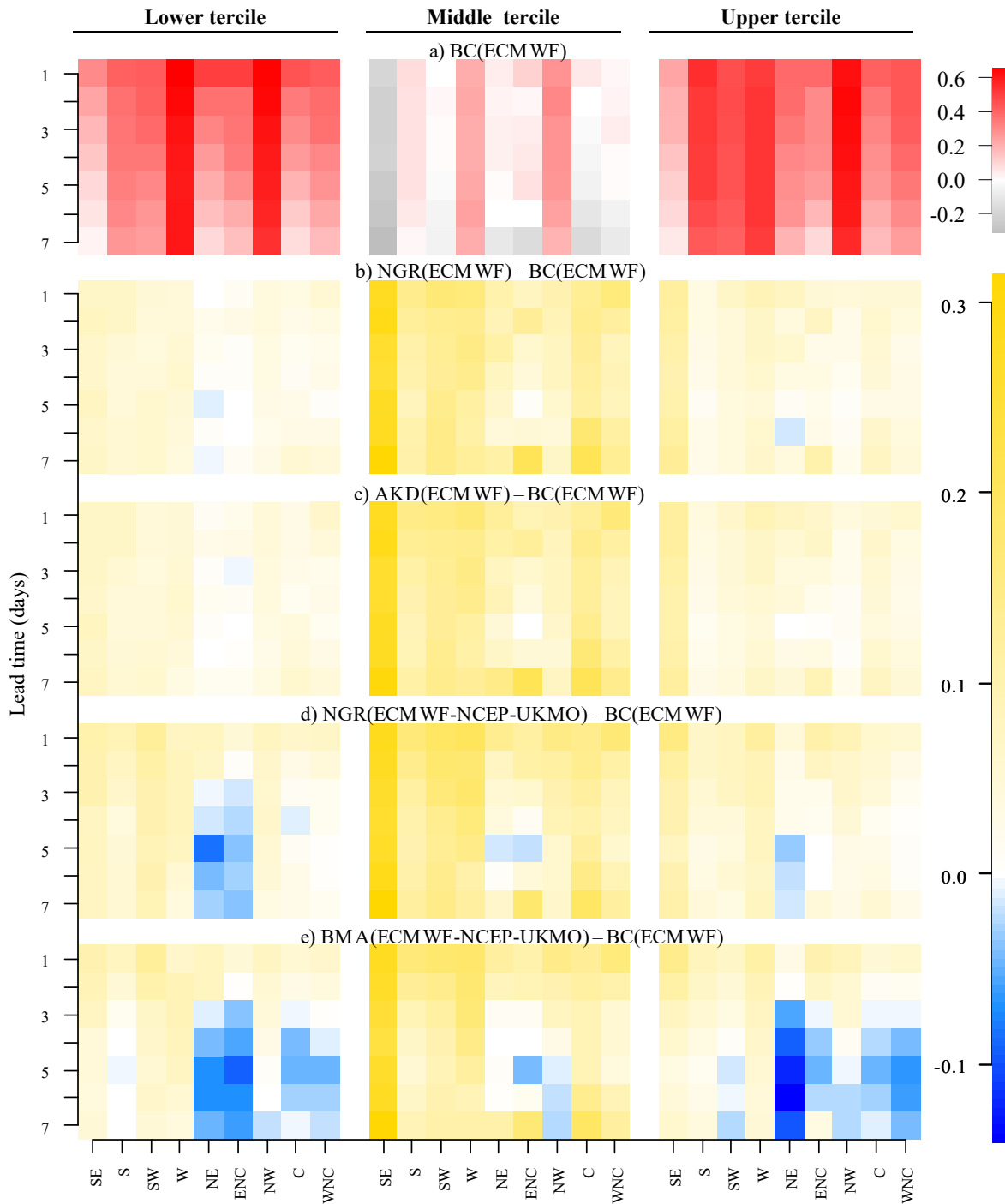


Figure 4. a) BSS for every region and lead time of the daily ECMWF forecasts post-processed using simple bias correction (used as reference BSS values) and b-e) differences between the BSS of the daily ECMWF forecasts post-processed with the b) NGR and c) AKD methods and the daily ECMWF-NCEP-UKMO forecasts post-processed with the d) NGR and e) BMA methods and the reference BSS.

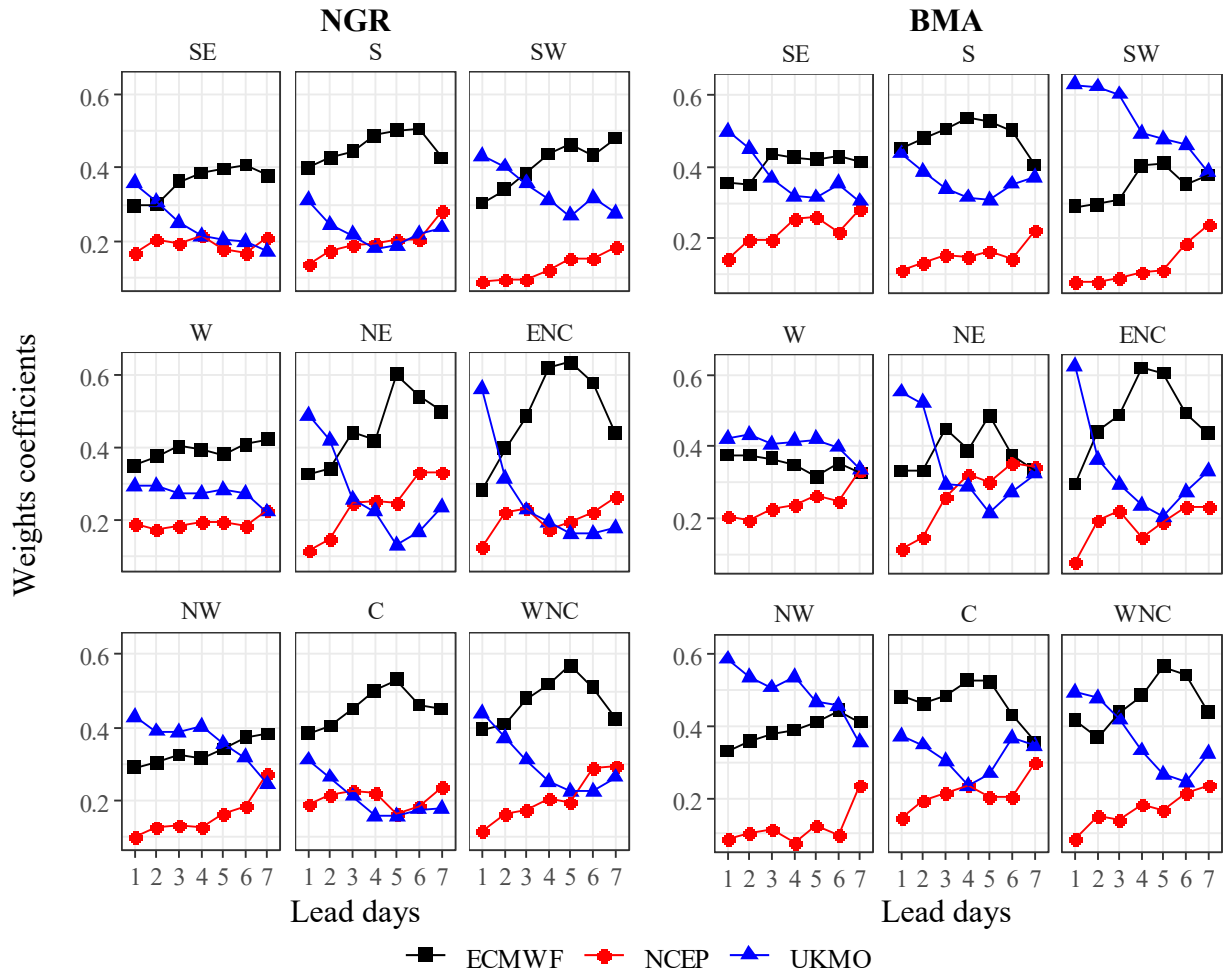


Figure 5. Regional mean weight coefficient  $b$  of the NGR technique (left panel) and the weight coefficient  $w$  of the BMA technique (right panel) for the post-processed daily ECMWF-NCEP-UKMO forecasts at different lead days.

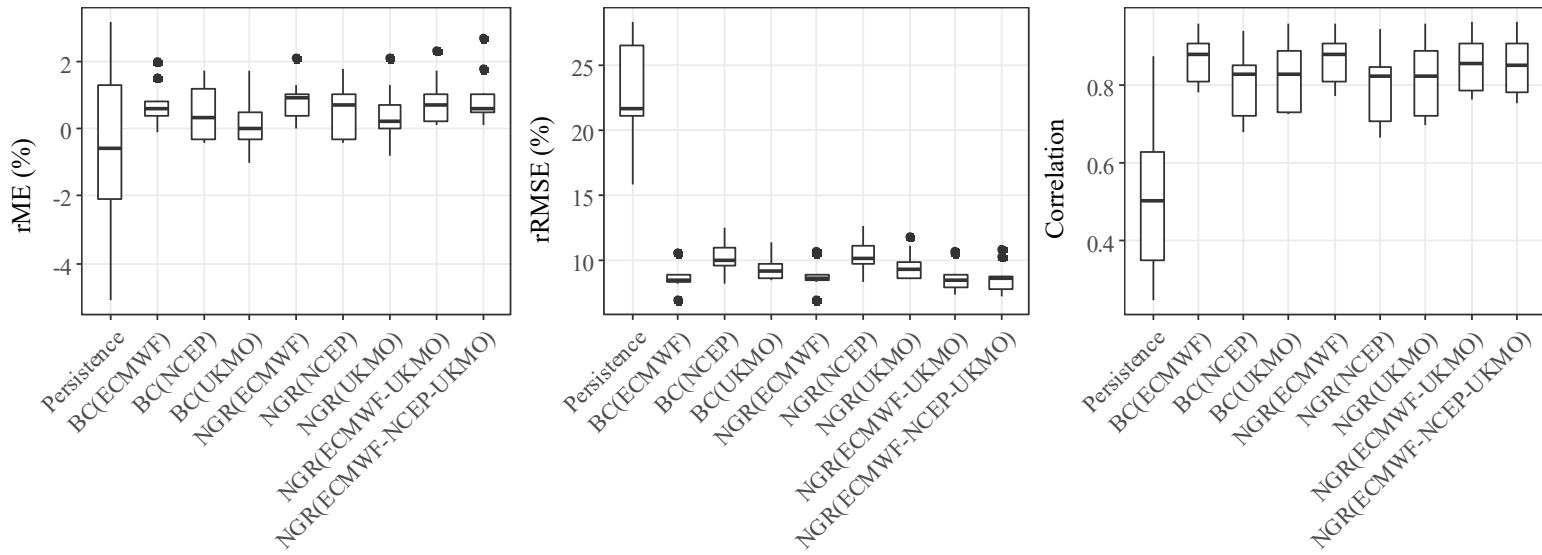


Figure 6. Whisker plot with the 2.5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 97.5<sup>th</sup> percentile of the distribution of the rME, rRMSE and correlation of weekly forecasts across different regions.

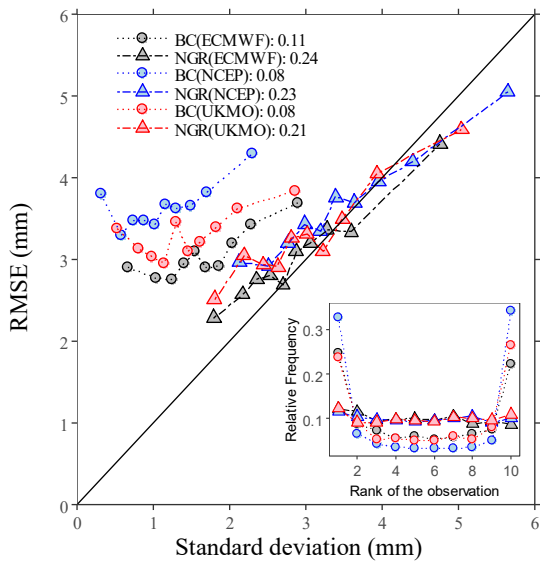


Figure 7. Binned spread-skill plots for the weekly forecasts accounting for the mean of the ensemble standard deviation deciles against the mean RMSE of the forecasts in each decile over the verification period using all pairs of forecasts and observations. The panel in the right and the bottom shows the corresponding rank histograms. The correlation between the standard deviations and the absolute errors is included in the legend. The solid line represents the 1:1 relationship.

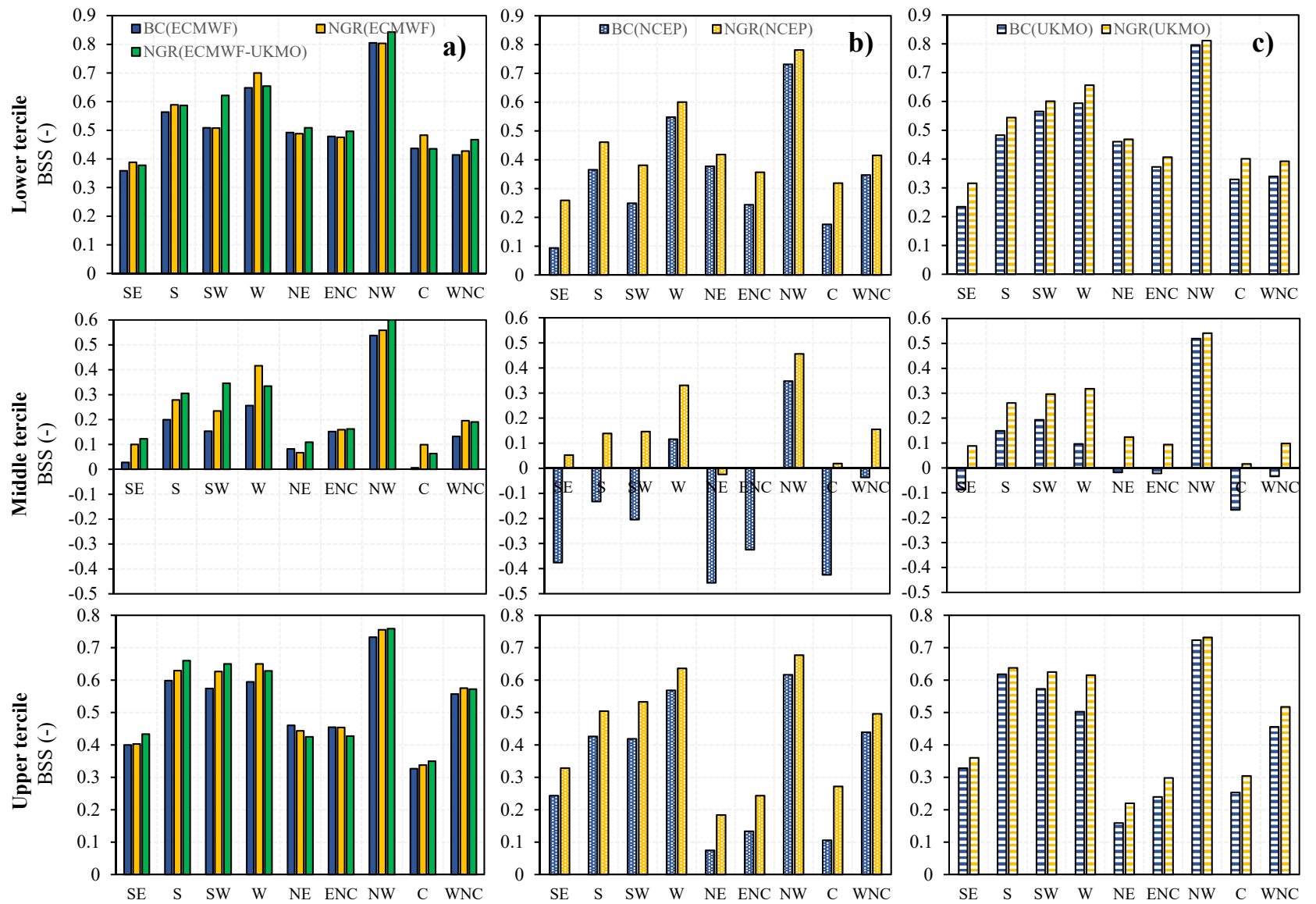


Figure 8. Comparison between BC and NGR based Brier Skill Scores considering a) ECMWF and ECMWF-UKMO forecasts, b) NCEP, and c) UKMO forecasts across the different climate regions.

Table 1. Evaluated schemes for daily and weekly ETo ensemble forecasts with different post-processing methods: BC (simple bias correction), NGR (nonhomogeneous Gaussian regression), AKD (affine kernel dressing), and BMA (Bayesian model averaging), and different model and ensemble schemes: ECMWF (European Centre for Medium-Range Weather Forecasts model), NCEP (National Centers for Environmental Prediction model), and UKMO (United Kingdom Meteorological office model) ensemble forecasts, as well as ECMWF-UKMO (ensembles of ECMWF and UKMO) and ECMWF-NCEP-UKMO (ensembles of ECMWF, NCEP, and UKMO) ensemble forecasts.

	Persistence	BC			NGR			AKD	BMA		
		ECMWF	NCEP	UKM	ECMWF	NCEP	UKM	ECMWF	ECMWF-	ECMWF-NCEP-	
		F	P	O	F	P	O	F	UKMO	UKMO	
Daily		✓			✓			✓		✓	✓
Weekly	✓	✓	✓	✓	✓	✓	✓	✓			

5

Table 2. Spatial weighted average values of daily forecast metrics over all climate regions for different methods at lead days 1 and 7. See the caption of Table 1 for explanations of the methods acronyms. Numbers in bold indicate the best performance for each lead day.

	BC		NGR		AKF		NGR		BMA		NGR		BMA	
	ECMWF		ECMWF		ECMWF		ECMWF-UKMO		ECMWF-UKMO		ECMWF-NCEP-UKMO		ECMWF-NCEP-UKMO	
	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days
rME (%)	0.822	1.203	1.695	2.682	1.626	2.419	1.327	2.735	0.632	0.939	1.394	2.778	<b>0.490</b>	<b>0.626</b>
rRMSE (%)	14.38	<b>19.64</b>	14.59	19.88	14.47	19.76	13.68	19.67	13.65	20.15	<b>13.59</b>	19.67	13.67	20.28
ME (mm day <sup>-1</sup> )	0.038	0.057	0.080	0.128	0.077	0.115	0.063	0.131	0.029	0.046	0.067	0.134	0.005	0.006
RMSE (mm day <sup>-1</sup> )	0.708	0.950	0.718	0.961	0.716	0.958	0.682	0.965	0.681	0.990	0.681	0.971	0.685	1.002
Correlation	0.832	<b>0.652</b>	0.829	0.649	0.830	0.649	<b>0.843</b>	0.639	0.841	0.586	0.841	0.635	0.832	0.560
Coverage ratio	64.54	79.40	95.63	95.44	95.93	96.10	94.24	94.73	<b>96.51</b>	96.56	93.52	94.57	96.47	<b>97.24</b>
CRPS (mm)	0.432	0.555	0.395	0.526	0.394	<b>0.525</b>	<b>0.374</b>	0.529	0.374	0.547	0.375	0.534	0.377	0.557
BSS_1st	0.442	0.232	0.492	0.279	0.492	<b>0.282</b>	<b>0.525</b>	0.274	0.519	0.240	0.521	0.271	0.513	0.225
BSS_2nd	0.042	-0.062	0.201	0.101	0.202	<b>0.101</b>	<b>0.224</b>	0.095	0.214	0.074	0.217	0.089	0.200	0.059
BSS_3rd	0.433	0.300	0.496	<b>0.359</b>	0.499	0.358	<b>0.519</b>	0.350	0.515	0.305	0.512	0.338	0.494	0.277



Table 3. Spatial weighted average values of weekly forecast metrics over all climate regions. See the caption of Table 1 for explanations of the methods acronyms.

	Persistence	BC			NGR				
		ECMWF	NCEP	UKMO	ECMWF	NCEP	UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO
rME (%)	-0.288	0.683	0.296	<b>0.097</b>	0.846	0.496	0.305	0.764	0.814
rRMSE (%)	22.108	8.872	10.453	9.460	8.952	10.571	9.599	8.753	<b>8.661</b>
ME (mm week <sup>-1</sup> )	-0.086	0.217	0.077	<b>0.007</b>	0.277	0.145	0.080	0.246	0.268
RMSE (mm week <sup>-1</sup> )	7.541	3.059	3.634	3.306	3.086	3.675	3.353	<b>3.059</b>	3.064
Correlation	0.530	<b>0.872</b>	0.806	0.835	0.870	0.801	0.829	0.863	0.856
Coverage ratio(%)		78.40	48.07	62.92	<b>99.29</b>	98.58	98.13	97.74	97.40
CRPS (mm)		1.836	2.406	2.072	1.727	2.071	1.884	<b>1.708</b>	1.715
BSS_1st		0.508	0.326	0.448	0.529	0.430	0.501	<b>0.547</b>	0.506
BSS_2nd		0.164	-0.147	0.069	0.238	0.150	0.204	<b>0.255</b>	0.225
BSS_3nd		0.528	0.371	0.468	0.553	0.461	0.515	<b>0.558</b>	0.550

## ANNEX

Table A1. Percentage differences (averaged over all lead times) of the ECMWF-UKMO and ECMWF-NCEP-UKMO forecast performance with the ECMWF forecast performance, after post-processing with the non-homogeneous Gaussian regression (NGR) method. See the caption of Table 1 for explanations of the forecast models acronyms.

	Western climate regions						Northern climate regions					
	SW		W		NW		NE		ENC		WNC	
	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO
ME	-26.75	-30.83	-9.11	9.42	-13.91	-18.80	-4.27	25.05	-2.15	-1.45	-10.12	0.76
RMSE	-4.68	-4.01	-3.46	-2.51	-3.97	-2.84	1.90	4.33	1.46	2.00	-1.31	-0.92
Correlation	1.76	0.63	0.95	0.71	1.20	0.61	-4.18	-4.60	-3.28	-3.14	-2.31	-2.06
Cov. ratio	-1.39	-2.09	-0.98	-1.19	-1.02	-1.14	-0.84	-1.66	-0.85	-0.99	-0.84	-1.40
CRPS	-4.84	-3.89	-3.42	-1.99	-3.90	-2.81	1.41	4.02	1.58	2.45	-1.00	-0.27
BSS_1st	12.02	7.48	3.22	2.85	3.55	4.24	-12.00	-9.68	-9.64	-9.38	-3.68	-5.18
BSS_2nd	8.99	-6.50	5.79	9.04	4.98	3.96	-112.95	-93.09	-19.09	-13.64	-15.73	-27.95
BSS_3nd	2.30	-1.81	3.58	6.56	4.20	2.37	-9.11	-8.99	-6.42	-10.61	-4.60	-5.84

Table A2. Percentage differences (averaged over regions) of forecast performance of using 45 days training period with using 30 days training period for lead days 1 and 7. See the caption of Table 1 for explanations of the methods acronyms.

	NGR(ECMWF)		AKD(ECMWF)		NGR(ECMWF-UKMO)		NGR(ECMWF-NCEP-UKMO)	
	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days
ME	16.57	18.73	21.65	22.86	4.71	10.09	-0.50	7.07
RMSE	-0.70	-2.64	-1.01	-3.12	-0.40	-3.72	-0.05	-4.74
Correlation	-0.16	0.53	-0.14	0.61	-0.10	1.33	-0.47	0.74
Cov. Ratio	1.28	0.95	1.62	1.26	1.70	1.50	1.94	1.34
CRPS (mm)	-0.77	-3.00	-1.22	-3.51	-0.92	-3.89	-0.01	-4.53
BSS_1st	-0.88	2.18	-1.16	2.76	-0.21	5.06	-2.60	6.28
BSS_2nd	-1.26	2.76	-1.28	5.68	3.61	8.96	-2.29	5.56
BSS_3rd	-0.38	-1.59	-0.90	-0.21	-1.34	2.63	-1.63	0.24

# 1 Comparison of probabilistic post-processing approaches for 2 improving NWP-based daily and weekly reference evapotranspiration 3 forecasts

4 Hanoi Medina, Di Tian

5 Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849

6 Correspondence to: Di Tian ([tiandi@auburn.edu](mailto:tiandi@auburn.edu))

7 **Abstract:** Reference evapotranspiration (ET<sub>o</sub>) forecasts play an important role in agricultural, environmental, and water  
8 management. This study evaluated probabilistic post-processing approaches, including the nonhomogeneous Gaussian  
9 regression (NGR), affine kernel dressing (AKD), and Bayesian model averaging (BMA) techniques, for improving daily and  
10 weekly ET<sub>o</sub> forecasting based on single or multiple numerical weather predictions (NWP) from The International Grand  
11 Global Ensemble (TIGGE), including the European Centre for Medium-Range Weather Forecasts (ECMWF), the National  
12 Centers for Environmental Prediction Global Forecast System (NCEP), and the United Kingdom Meteorological Office  
13 forecasts (UKMO). The approaches were examined for the forecasting of summer ET<sub>o</sub> at 101 U.S. Regional Climate  
14 Reference Network stations distributed all over the contiguous United States (CONUS). We found that the NGR, the AKD  
15 and the BMA methods greatly improved the skill and reliability of the ET<sub>o</sub> forecasts compared to a linear regression bias  
16 correction method, due to the considerable adjustments on the spread of ensemble forecasts. The methods were especially  
17 effective when applied over the raw weekly-NCEP forecasts, followed by the raw-UKMO forecasts, because of their low skill  
18 compared to that of the raw ECMWF forecasts. The post-processed weekly forecasts had much lower rRMSE (between 8-  
19 11%) than the persistence-based weekly forecasts (22%), and the post-processed daily forecasts (13-20%). Compared with the  
20 single model ensemble ET<sub>o</sub> forecasts based on ECMWF, multi-model ensemble ET<sub>o</sub> forecasts showed higher skill at short  
21 lead times (1 or 2 days) and over the southern and western regions of the United States. The improvement was higher at the  
22 daily timescale than at the weekly timescale. The NGR and AKD methods performed the best, but unlike the AKD method,  
23 the NGR method can post-process multi-model forecasts is more flexible and it is easier to interpret computationally efficient  
24 than the other methods. In summary, the study demonstrated that the three probabilistic approaches generally outperform  
25 conventional procedures based on the simple bias correction of single model forecasts, with the NGR post-processing of the  
26 ECMWF and ECMWF-UKMO forecasts providing the most efficient-cost-effective ET<sub>o</sub> forecasting.

Formatted: Highlight

Formatted: Highlight

## 27 Introduction

28 Reference crop evapotranspiration (ET<sub>o</sub>) represents the weather driven component of the water transfer from plants and soils  
29 to the atmosphere. It plays a fundamental role in estimating mass and energy balance over land surface as well as in agronomic,

30 forestry, and water resources management. In particular, ETo forecasting is important for aiding water management decision  
31 making (such as irrigation scheduling, reservoir operation, etc.) under uncertainty by identifying the range of future plausible  
32 water stress and demand (Pelosi et al., 2016; Chirico et al., 2018). While ETo forecasts based on global medium range NWP  
33 have been mostly focused on the daily timescale (e.g. Perera et al., 2014; Silva et al., 2010; Tian and Martinez, 2012a, b, 2014;  
34 Medina et al., 2018), weekly ETo forecasts are also important for users. Studies show that both daily and weekly forecasts  
35 have increasing influence on the decision makers in agriculture (Prokopy et al., 2013; Mase and Prokopy, 2014) and water  
36 resource management (Hobbins et al., 2017). For example, irrigation is commonly scheduled considering both daily and  
37 weekly basis, while weekly evapotranspiration forecasts are useful for planning water allocation from reservoirs, especially in  
38 cases of shortages. Weekly ETo anomalies can also be useful to provide warnings of wild-fires (Castro et al., 2003) and  
39 evolving flash drought conditions (Hobbins et al., 2017). Therefore, accounting for the post-processing of both daily and  
40 weekly ETo predictions provides a more comprehensive view of the capabilities of these forecasting approaches than  
41 considering only daily predictions while better fits the user's actual needs.

42  
43 However, ETo forecasting is highly uncertain due to the chaotic nature of weather systems. In addition, ETo estimation requires  
44 full sets of meteorological data which ~~is-are~~ usually not easy to obtain. Due to the improvement of numerical weather  
45 predictions (NWP), studies have been recently emerged to forecast ETo using outputs of NWP over different regions of the  
46 world (Silva et al., 2010; Tian and Martinez, 2012 a, 2012b, and 2014; Perera et al., 2014; Pelosi et al., 2016; Chirico et al.,  
47 2018; Medina et al., 2018). Operationally, experimental ETo forecast products are being developed, such as Forecast Reference  
48 EvapoTranspiration (FRET) product (<https://digital.weather.gov/>), as part of the U.S. National Weather Service (NWS)  
49 National Digital Forecast Database (NDFD) (Glahn and Ruth, 2003), and the Australian Bureau of Meteorology's Water and  
50 Land website (<http://www.bom.gov.au/wat/>), which provides current and forecasted ETo at the continental scale.

51 The improved performance of NWP during recent years is largely due to the improvement of physical, statistical  
52 representations of the major processes in the models, and the use of ensemble forecasting (Hamill et al., 2013, Bauer et al.,  
53 2015). Nevertheless, the NWP forecasts still commonly show systematic inconsistencies with measurements, which are often  
54 caused by inherent errors of NWP or local land-atmospheric variability which is not well resolved in the models. Post-  
55 processing methods, defined as any form of adjustment to the model outputs in order to get better predictions (eg., Hagedorn  
56 et al., 2012), ~~is-are~~ highly recommended to attenuate, or even eliminate, those inconsistencies (Wilks, 2006; Gneiting et al.,  
57 2005; Raftery et al., 2005). ~~However, Until a few years ago,~~ most post-processing ~~procedures-applications~~ only considered  
58 single-model predictions (i.e., predictions generated by a single NWP model), and addressed errors in the mean of the forecast  
59 distribution while ignored those in the forecast variance (Gneiting, 2014). These procedures regularly adopted some form of  
60 model output statistics (MOS, Glahn and Lowry, 1972; Klein and Glahn, 1974) methods, focusing on correcting current  
61 ensemble forecasts based on the bias in the historical forecasts.

62 As no forecast is complete without an accurate description of its uncertainty (National Research Council of the National  
63 Academies 2006), the dispersion of the forecast ensemble often misrepresent the true density distribution of the forecast

Formatted: None

64 uncertainty (Krzysztofowicz 2001; Smith 2001; Hansen 2002). The ensemble forecasts are, for example, commonly under-  
65 dispersed (e.g. Buizza et al. 2005; Leutbecher and Palmer, 2008), which make the probabilistic predictions overconfident  
66 (Wilks 2011). Therefore, ~~another new-generation of probabilistic techniques has been~~ proposed to also address dispersion  
67 errors of the ensembles (Hamill and Colucci 1997; Buizza et al., 2005, Pelosi et al., 2017), in some cases through the  
68 manipulation of multi-model weather forecasts.

69 The nonhomogeneous Gaussian regression (NGR, Gneiting et al., 2005), the Bayesian model averaging, (BMA, Raftery et al.,  
70 2005; Fraley et al., 2010), ~~the extended logistic regression (ELR, Wilks et al., 2009; Whan and Schmeits, 2018), the quantile~~  
71 ~~mapping (Verkade et al., 2013)~~ and the family of kernel dressing (Roulston and Smith 2003; Wang and Bishop 2005), such  
72 as the affine kernel dressing (AKD, Brocker and Smith 2008), are ~~emerging-state of art~~ probabilistic techniques (~~Gneiting,~~  
73 ~~2014~~). ~~However, the ELR has been reported to fall short in using the information contained in the ensemble spread in efficient~~  
74 ~~way (Messner et al., 2014), while the quantile mapping method have been found to degrade rather than improve the forecast~~  
75 ~~performance in some circumstances (Madadgar et al., 2014).~~ ~~The with the-NGR, AKD and -and the-BMA are sometimes~~  
76 ~~considered as variants of dressing methods (Brocker and Smith 2008), as they produce a continuous forecast probability~~  
77 ~~distribution function (pdf) based on the original ensemble. This property makes them particularly useful for the decision~~  
78 ~~making (Gneiting, 2014), compared to the methods that provide post-processed ensembles. Another common advantage is that~~  
79 ~~they perform commonly well with relatively short training datasets (Geiting et al., 2005; Raftery et al., 2005; Wilks and Hamill,~~  
80 ~~2007). A limitation of the NGR, compared to the AKD and BMA methods, is that the resulting forecast pdf is invariably~~  
81 ~~Gaussian, while a limitation of the AKD is that it only considers single model ensembles. Instead, the NGR and AKD methods~~  
82 ~~provide more flexible mechanisms for the simultaneous adjustments in the forecast mean and spread-skill (Brocker and Smith,~~  
83 ~~2008).~~

84 ~~The AKD, NGR and BMA methods produce continuous predictive density distributions, which may be useful for the decision~~  
85 ~~making (Gneiting, 2014), and perform commonly well with relatively short training datasets (Geiting et al., 2005; Raftery et~~  
86 ~~al., 2005; Wilks and Hamill, 2007).methods being especially designed for multi-model post-processing.~~

87 Studies suggest that the post-processing of NWP-based ETo forecasts are crucial for informing decision making (e.g. Ishak et  
88 al., 2010). Medina et al. (2018) compared single and multi-model NWP-based ensemble ETo forecasts and the results showed  
89 that the performance of the multi-model ensemble ETo forecasts is considerably improved through a simple bias-correction  
90 post-processing, and that the bias-corrected multi-model ensemble forecasts were in general better than the single model  
91 ensemble forecasts. In reality, while most applications for the ETo forecasting have involved some form of post-processing,  
92 these have been often limited to simple MOS procedures of single-model ensembles (e.g., Silva et al., 2010; Perera et al.,  
93 2014). Poor treatments of uncertainty and variability is considered as a main issue affecting users' perceptions and adoptions  
94 of weather forecasts (Mase and Prokopy, 2014). The appropriate representation of the second and higher moments of the ETo  
95 forecast probability density is especially important to predict extreme values, as shown by Williams et al. (2014). Therefore,  
96 the use of probabilistic post-processing techniques such as the NGR, the AKD and BMA, may greatly enhance the overall  
97 performance of the ETo forecasts compared to the simple MOS procedures.

Formatted: Highlight

98 Only a few studies have considered probabilistic methods for post-processing of ETo forecasts. These include the works of  
99 Tian and Martinez (2012a, 2012b, and 2014), and more recently Zhao et al (2019). The former authors showed the Analog  
100 Forecast (AF) method to be useful for the post-processing ETo forecasts based on Global Forecast System (GFS, Hamill et al.,  
101 2006) and Global Ensemble Forecast System (GEFS, Hamill et al., 2013) reforecasts. Tian and Martinez (2014) found that  
102 water deficit forecasts produced with the post-processed ETo forecasts had higher accuracy than those produced with  
103 climatology. On other hand, Zhao et al. (2019) improved the skill and the reliability of the Australian BoM model using a  
104 Bayesian joint probability (BJP) post-processing approach, which is based on the parametric modelling of the joint probability  
105 distribution between forecast ensemble means and observations. However, a main disadvantage of both the AF and the BJP  
106 methods compared to the aforementioned emerging-state of art probabilistic approaches is that, while they transform the spread  
107 of the ensembles, they rely on the mean of retrospective reforecasts, thus neglecting information about their dispersion. The  
108 AF approach also has the disadvantages that requires long time series of retrospective forecasts, and may be unsuitable for  
109 extreme events forecasting (e.g., Medina et al., 2019). The AKD, NGR and BMA methods produce continuous predictive  
110 density distributions, which may be useful for the decision making (Gneiting, 2014), and perform commonly well with  
111 relatively short training datasets (Geiting et al., 2005; Raftery et al., 2005; Wilks and Hamill, 2007). The use of new ETo novel  
112 forecasting strategies relying on the postprocessing of single and multi-model ensemble forecasts with these emerging the  
113 NGR, AKD and the BMA probabilistic techniques provide good opportunities for improving the ETo predictions.  
114 While ETo forecasts based on global medium range NWP have been mostly focused on the daily timescale (Perera et al., 2014;  
115 Silva et al., 2010; Tian and Martinez, 2012a, b, 2014; Medina et al., 2018), weekly ETo forecasts are also important for users.  
116 Studies show that both daily and weekly forecasts have increasing influence on the decision makers in agriculture (Prokopy et  
117 al., 2013; Mase and Prokopy, 2014) and water resource management (Hobbins et al., 2017). For example, irrigation is  
118 commonly scheduled considering both daily and weekly basis while weekly evapotranspiration forecasts are useful for  
119 planning water allocation from reservoirs, especially in cases of shortages. Weekly ETo anomalies can also be useful to provide  
120 warnings of wild fires (Castro et al., 2003) and evolving flash drought conditions (Hobbins et al., 2017). Therefore, accounting  
121 for the post-processing of both daily and weekly ETo predictions provides a more comprehensive view of the capabilities of  
122 these forecasting approaches than considering only daily predictions while better fits the user's actual needs.  
123 In this paper, we are addressing several scientific questions which have not been adequately studied in previous literature,  
124 including, how effective are the new-state of art probabilistic post-processing methods compared with the traditional MOS bias  
125 correction methods for post-processing ETo forecasts? Is it worth implementing the probabilistic post-processing for multi-  
126 model rather than single-model ensemble forecasting? For the first time, this work aims to evaluate and compare multiple  
127 novel strategies for post-processing both daily and weekly ETo forecasts using the emerging NGR, AKD and  
128 BMA probabilistic approaches. The study represents a major step forward with respect to Medina et al. (2018), which evaluated  
129 the performance of raw and linear regression bias corrected daily ETo forecasts produced with single and multi-model  
130 ensemble forecasts. It provides a broad characterization of the performance for different probabilistic post-processing strategies  
131 but also diagnoses the causes of high and low performance.

## 132 2 Methods and Datasets

### 133 2.1 The probabilistic methods

134 The NGR, AKD and BMA techniques follow a common strategy: they yield a predictive probability density function (PDF)  
135 of the post-processed forecasts  $y$  given the raw forecasts  $x$  and some fitting parameters  $\theta$  ( $p(y|x, \theta)$ ). The parameters  $\theta$  are  
136 fitted using a training dataset of ensemble forecasts and observations, as in the MOS techniques. Below is a brief description  
137 of each technique.

#### 138 2.1.1 Non-Homogeneous Gaussian Regression

139 The NGR (Gneiting et al., 2005) produces a Gaussian predictive (PDF) based on the current ensemble (of typically multi-  
140 model) forecasts. If  $x_{ij}$  denote the  $j^{\text{th}}$  ( $j = 1, \dots, m_i$ ) ensemble forecast member of model  $i$  ( $i = 1, \dots, n$ ), then  
141  $p(y|x, \theta) \sim \mathcal{N}(\mu, v)$ , where the mean:

$$142 \mu = a + \sum_{i=1}^n b_i \bar{x}_i \quad (1)$$

143 is a linear combination of the mean ensemble forecasts  $\bar{x}_i$  and the variance:

$$144 v = c + dS^2 \quad (2)$$

145 is a linear function of the ensemble variance  $S^2$ . The fitting parameters  $a$ ,  $b_i$ ,  $c$  and  $d$  are determined by minimizing the  
146 continuous rank probability score (CRPS) using the training set of forecasts and observations. Notice that parameters  $a$ ,  $c$  and  
147  $d$  are indistinguishable among **exchangeable** members; therefore the  $b_i$  can be seen as a weighting parameters that reflect the  
148 better or worse performance of one model compared to the others. The NGR technique is implemented in R (R Core Team)  
149 using the packages ensembleMOS (Yuen et al., 2018),

#### 150 2.1.2. Affine Kernel Dressing

151 The affine kernel dressing method (Bröcker and Smith, 2008) only considers single model ensemble forecasts. It  
152 estimates  $p(y|x, \theta)$  using a mixture of normally distributed variables:

$$153 p(y|x, \theta) = \frac{1}{m\sigma} \sum_{j=1}^m K\left(\frac{y-z_j}{\sigma}\right) \quad (3)$$

154 where  $K$  represents a standard normal density kernel ( $K(\xi) = 1/\sqrt{2\pi} \exp(-1/2\xi^2)$ ), centered at  $z_j$ , such that:

$$155 z_j = ax_j + r_1 + r_2 \bar{x} \quad (4)$$

156 and,

$$157 \sigma^2 = h_s^2 (s_1 + s_2 u(\mathbf{z})) \quad (5)$$

158 where  $h_s$  is the Silversman's factor (Bröcker and Smith, 2008),  $u(\mathbf{z})$  is the variance of  $\mathbf{z}$  and  $a$ ,  $r_1$ ,  $r_2$ ,  $s_1$ ,  $s_2$  are fitting  
159 parameters obtained by minimizing the mean Ignorance score. For clarity we use the same nomenclature for the parameters as  
160 in the original study. From Eqs. 4 and 5 we can obtain that the predictive variance  $v$  is a function of the ensemble variance  $S^2$   
161 (Brocker and Smith, 2008):



162  $v = h_1^2 s_1 + a^2(1 + h_2^2 s_2)S^2 = c^* + d^*S^2$  (6)

163 Here,  $S^2$  represents the variance of the ensemble of exchangeable members.

164 The AKD technique is implemented through the SpecsVerification R package (Siegert, 2017).

165 **2.1.3 Bayesian Model Averaging**

166 The BMA method (Raftery et al. 2005, Fraley et al., 2010) also produces a mixture of normally distributed variables, as the  
 167 AKD method, but based on ~~multi-model-fore~~multi-model ensemble forecasts. In this case the predictive PDF is given by a  
 168 weighted sum of component PDFs,  $g_i(y|x_{i,j}; \theta_i)$ , one per each member:

169  $p(y|x, \theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} w_i g_i(y|x_{i,j}, \theta_i)$  (7)

170 such that the weights and the parameters are invariable among members of the same model and

174  $\sum_{i=1}^n m_i w_i = 1$

171 In the study the component PDFs are assumed normal as for the affine kernel dressing method. Estimates of  $w_i$ s and  $\theta_i$ s are  
 172 produced by maximizing the likelihood function using an Expectation Maximization algorithm (Casella and Berger, 2002).

173 The BMA technique is implemented through the ensembleBMA R package (Fraley et al., 2016).

175 **2.2 Measurement and forecast datasets**

176 ETo observations and forecasts were computed with the FAO-56 PM equation (Allen et al., 1998), from daily meteorological  
 177 data as inputs. They covered the same period, between May and August from 2014 to 2016. The observations used daily  
 178 measurements of minimum and maximum temperature, minimum and maximum relative humidity, wind speed, and surface  
 179 incoming solar radiation from 101 U.S. Climate Reference Network (USCRN) weather stations. The USCRN stations are  
 180 distributed over nine climatologically consistent regions in CONUS (Fig. 1). The ETo forecasts used daily maximum and  
 181 minimum temperature, solar radiation, wind speed, and dew point temperature reforecasts of European Centre for Medium-  
 182 Range Weather Forecasts model (ECMWF) outputs, United Kingdom Meteorological office model (UKMO) outputs, and  
 183 National Centers for Environmental Prediction model (NCEP) from The International Grand Global Ensemble (TIGGE;  
 184 Swinbank et al. 2016) database at each of these stations, considering a maximum lead time of 7 days. We used the same models  
 185 as Medina et al. (2018) for comparison purposes, and because they are considered among the most skillful globally (e.g.  
 186 Hagedorn et al., 2012). The forecasts were interpolated to the same  $0.5^\circ \times 0.5^\circ$  grid using the TIGGE data portal. The weekly  
 187 forecasts accounted for the sum of the daily predictions generated at a specific day of each week, and the weekly observations  
 188 considered the sum of the daily observations over the corresponding forecasting days, such that the weekly observations were  
 189 independent from each other. In the study, we used the nearest neighbor approach to interpolate the forecasts to the USCRN  
 190 stations, which does not account for the effects of elevation. While the use of interpolation techniques considering the effects  
 191 of elevation (e.g. van Osnabrugge et al., 2019) may correct part of the forecasts errors before the post-processing, it could also

Formatted: Superscript  
 Formatted: Superscript

192 affect the multivariate dependence of the weather variables. Hagedorn et al. (2012) showed that the post-processing can not  
193 only address the discrepancies related to the model's spatial resolution, but also serve as a means of downscaling the forecasts.

## 194 2.3 Post-processing schemes

### 195 2.3.1 Training and verification periods

196 The training data for the daily post-processing ~~comprehended~~ comprised the pairs of daily forecasts and corresponding  
197 observations corresponding from 30 days prior to the forecast initial day, as in Medina et al. (2018). Instead, the training data  
198 for the weekly post-processing ~~comprehended~~ included all the other pairs of weekly forecasts and observations available for  
199 the forecast location, similarly as in the case of a leave one out cross validation framework. In the study both the daily and  
200 weekly forecasts were verified for events over June-August, 2014-2016.

### 201 2.3.2 Baseline approaches

202 Linear regression bias correction (BC) of the ECMWF forecast was used as a baseline approach for measuring the effectiveness  
203 of the NGR, the AKD and the BMA methods considering both daily and weekly forecasts. Here, the current forecasts bias is  
204 estimated as a linear function of the forecasts mean, and the members of the ensemble are shifted accordingly. The function is  
205 calibrated using the forecasts mean and the actual biases ~~from a retrospective based on the same training periods as for the~~  
206 ~~other post-processing methods set of forecasts and observations~~. Persistence is also used as a baseline approach for weekly  
207 forecasts, considering its applicability in productive systems. In this case the ETo for a current week is estimated as the  
208 observed ETo during the previous week.

### 209 2.3.3 Forecasting Experiments

210 Table 1 summarizes the daily and weekly NWP-based ETo forecasting experiments based on different post-processing  
211 methods and model combinations. The analyses of the daily forecasts ~~make~~ put more emphasis on the differences among post-  
212 processing methods. They include an examination of the effect of the duration of the training period on the forecasts  
213 assessments as well as the regression weights from the tested post-processing methods. Whereas, the weekly forecasts ~~make~~  
214 put more emphasis on the differences among the several single and multi-model ETo forecasts under baseline and probabilistic  
215 post-processing.

## 216 2.4 Forecast verification metrics

217 In this study we use several metrics to evaluate deterministic and probabilistic forecast performance of the post-processed ETo  
218 forecasts. For consistency purposes, the metrics of the tested methods were assessed using 50 random samples, i.e., same  
219 number as members in the bias corrected ECMWF forecasts. Deterministic ETo forecast was produced by taking the average  
220 of the ensemble members. The deterministic forecast performance was assessed using the bias or mean error (ME) and relative

221 ~~bias~~ME (rME) ~~or mean error~~, the root mean square error (RMSE) and the relative RMSE (rRMSE), and the correlation ( $\rho$ ),  
 222 which are common measures of agreement in many studies. The ~~Both the relative and the absolute bias and relative -RMSE~~  
 223 bias are calculated and reported.

224 The ME and rME were computed as

$$225 \text{ME} = \frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i) \quad (8)$$

$$226 \text{rME} = \frac{\sum_{i=1}^n (\bar{f}_i - \sigma_i)}{n\bar{\sigma}} \quad (9)$$

227 where  $\bar{f}_i$  represents the average ensemble forecast for the event  $i$  ( $i = 1 \dots n$ ),  $\sigma_i$  is the corresponding observation, and  $\bar{\sigma}$  is the  
 228 mean observed data.

229 The RMSE and the rRMSE were computed as

$$230 \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i)^2} \quad (10)$$

$$231 \text{rRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{f}_i - \sigma_i)^2}}{\bar{\sigma}} \quad (11)$$

232 The correlation was obtained as

$$233 \rho = \frac{\sum_{i=1}^n (\bar{f}_i - \bar{f})(\sigma_i - \bar{\sigma})}{s_f s_\sigma} \quad (12)$$

234 where  $\bar{f}$  is the mean of the average ensemble forecast and  $s_f$  and  $s_\sigma$  are the standard deviation of the average forecasts and the  
 235 observations, respectively.

236 The probabilistic forecast performance was assessed using range histogram, the spread-skill relationship (see Wilks, 2011) and  
 237 the forecast coverage as measurements of the forecast reliability, and the Brier Skill Score (BSS) as a measure of the skill,  
 238 and the continuous rank probability score (CRPS), for providing an overall view of the performance (Hersbach, 2000), as it is  
 239 sensitive to both errors in location and spread simultaneously as a measurement of the skill.

240 Reliability here refers to the statistical consistency (as in Toth et al. 2003), which is met when the observations are statistically  
 241 indistinguishable from the forecast ensembles (Wilks, 2011). To obtain the rank histogram, we get the rank of the observation  
 242 when merged into the ordered ensemble of ETo forecasts and then we plot the ranks histogram. The spread-skill relationships  
 243 are represented as binned-type plots (e.g.; Pelosi et al., 2017), accounting for the mean of the ensemble standard deviation  
 244 deciles (as an indication of the ensemble spread) against the mean RMSE of the forecasts in each decile over the verification  
 245 period. The plots include the correlation between these two quantities. Calibrated ensembles should show a 1:1 relationship  
 246 between the standard deviations and the RMSE. If the forecasts are unbiased and the spread is small compared to the RMSE,  
 247 then the ensembles tend to be under-dispersive. The inverse of the spread provides an indication of sharpness, which is the  
 248 level of “compactness” of the ensemble (Wilks, 2011).

249 In addition to the spread skill relationship, we also report the ratio between the observed and nominal coverage (hereinafter  
 250 referred as coverage ratio). The coverage of a  $(1 - \alpha)100\%$ ,  $\alpha \in (0, 1)$ , central prediction interval is the fraction of

251 observations from the verification data set lying between  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the predictive distribution. It is  
 252 empirically assessed by considering the observations lying between the extreme values of the ensembles. The nominal or  
 253 theoretical coverage of a calibrated predictive distribution is  $(1 - \alpha)100\%$ . A calibrated forecast of  $m$  ensemble members  
 254 provides a nominal coverage of about  $(m - 1)/(m + 1)100\%$  central prediction interval (e.g., Beran and Hall, 1993). For  
 255 example, an ensemble of 50 members provides 96% central prediction interval. The ratio between the observed and nominal  
 256 coverages provides a quantitative indicator of the quality of the forecasts dispersion under unbiasedness: a ratio lower (larger)  
 257 than 1 suggest that the forecasts tend to be under (over) dispersive.

258 The BSS is computed as

259 
$$BSS = 1 - \frac{BS}{BS_{clim}} \tag{13}$$

260 where BS is the Brier score of the forecast

261 
$$BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \tag{14}$$

262  $p$  is the forecast probability  $p$  of the event, which is estimated based on the ensemble, and  $p$  is equal to 1 if the event occurs  
 263 and 0 otherwise.

264  $BS_{clim}$  in Eq. 8 represents the Brier Score of the sample climatology, computed as (Wilks, 2010)

265 
$$BS_{clim} = \bar{o}(1 - \bar{o}) \tag{15}$$

266 where  $\bar{o}$  is the sample climatology computed as the mean of the binary observations  $o_i$  in the verification dataset.

267 ~~Finally, the BSS represents a traditional skill-score relationship that adopts the Brier score (Wilks, 2011), as the accuracy~~  
 268 ~~measure. In this study we compute the BSS associated to the tercile events of the ETo forecasts (upper or 1st, middle or 2nd,~~  
 269 ~~and lower or 3rd terciles). Therefore, the sample climatology is equal to 0.33 and  $BS_{clim} = 0.22$ , exactly as in Medina et al.~~  
 270 ~~(2018).~~

271 The CRPS was computed as

272 
$$CRPS = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} (F_i^f(h) - F_i^o(h))^2 dh \tag{16}$$

273 where  $F_i^f$  and  $F_i^o$  are the cumulative distribution function of the forecast and the observations, respectively, and  $h$  represents  
 274 the threshold value,  $F_i^o(h) = H(h - o_i)$ ,  $H$  representing the Heaviside function, which is 0 for  $h < o_i$  and 1 for  $h \geq o_i$ .

276 **3 Results**

277 **3.1 Comparing the NGR, AKD and BMA methods at daily scale**

278 **3.1.1 Deterministic forecast performance**

279 Figure 2 shows the relative bias, ME and rRMSE as well as the correlation of the forecasts post-processed using different  
 280 approaches over the southeast (SE) and northwest (NW) regions. These regions are representative of the Eastern and Western

Formatted: Font: Italic

Formatted ... [1]

Formatted ... [2]

Formatted: Font: Not Italic

Formatted ... [3]

Formatted: Font: 10 pt

Formatted ... [4]

Formatted: Font: Italic

Formatted: Font: 10 pt

Formatted ... [5]

Formatted ... [6]

Formatted ... [7]

Formatted: Highlight

Formatted ... [8]

281 zones, which tended to provide the worse and best rRMSE and correlations, respectively. In general, the probabilistic post-  
282 processing methods add no additional skill to the deterministic forecast performance compared to the simple bias correction.  
283 While the rRMSE are relatively high, the bias-rME is-are very low, which indicates that the errors are mostly random. The  
284 BMA and the simple linear regression methods provided lower bias than the NGR and AKD methods. Instead, the BMA  
285 method provided higher rRMSE and lower correlations than the other three methods at long lead times. The rRMSE and the  
286 correlations tended to be more variable among lead times and regions than among post-processing methods, while for the rME  
287 was the opposite. In addition, the changes in rRMSE and correlation with lead time tended to be larger over the Eastern regions.

Formatted: Highlight

Formatted: Highlight

### 288 3.1.2 Probabilistic forecast performance

289 Figure 3 shows the spread skill relationship and the rank histograms using all pairs of forecasts and observations for lead days  
290 1 and 7. The spread-skill relationship in Figure 3 shows that the probabilistic post-processing methods considerably improved  
291 the reliability of the ETo forecasts compared with the linear regression bias correction (Figure 3). The former methods tend to  
292 correct evident shortcomings of the ensemble raw forecasts which are unresolved by the simple post-processing, i.e., the  
293 considerably under-dispersion at short lead times, and the poor consistency between the ensemble spread and the RMSE at  
294 longer lead times. The adjustments had a low cost in terms of sharpness, judging by the range of ensemble spreads for the  
295 different line plots, but seemed slightly insufficient. The correlations between the ensemble standard deviation and the RMSE  
296 are fairly low, suggesting a limited predictive ability of the spread (Wilks, 2011). Nonetheless, they were consistently higher  
297 for probabilistic post-processing methods, compared to the linear regression method, and at short lead times, compared to the  
298 long lead times. The performance was low sensitive to the type of probabilistic post-processing, independent of the single or  
299 multi-model forecasts strategy, although the BMA post-processing provided slightly lower correlations, especially for longer  
300 lead times. The rank histograms in Figure 3 show that the probabilistic methods provided better calibration than the linear  
301 regression approach both at 1 and 7 days, but the improvements were considerably larger at 1 day. At the short lead time, the  
302 three methods slightly over-forecasted ETo, suggesting that the departures from the predictive mean has a negative skew, but  
303 in general they were fairly confident. In this case all the methods provided almost the same result. At the long lead time, there  
304 is also an overestimation and then a positive bias, but also a slight U-shaped pattern, associated to some underdispersion for  
305 the range of the low and medium observations, which is coherent with the spread skill relationships. These issues are more  
306 pronounced using the BMA method and less pronounced using the AKD methods. Scheuerer and Büerger (2014) reported  
307 similar issues when post-processing ensemble forecasts of temperatures with the NGR method and a version of the BMA  
308 method. On the other hand, the calibration was affected little by the choice of a single or multi-model strategy for a given  
309 post-processing method. Nevertheless, the probabilistic methods provided a coverage ratio close to 100% independently of the  
310 lead time (see Table 2) and the region (not shown). The simple bias correction method instead provided coverage ratios much  
311 lower and more variable among regions (see Table 2) and lead times.

Formatted: Highlight

Formatted: Highlight

312 The coverage ratios in table 2 provides quantitative insights about the forecasts under dispersion for the different strategies.  
313 The simple bias corrected ECMWF forecasts provided a mean coverage ratio of 77%, but it can be as low as the 50%. The

314 other forecasts provided coverage ratios of over 91%. The ratios were slightly better (i.e., closer to one) using the BMA  
315 method than with the NGR and the AKD methods, and using single ECMWF forecasts than with the ECMWF-UKMO and  
316 the ECMWF-NCEP-UKMO forecasts.

317 The NGR and AFK methods provided better Brier skill score (BSS) than the BC method for the three categories of ETo values,  
318 with improvements being higher for the middle tercile, than for the lower and upper terciles (Figure 4). The BMA based skill  
319 scores tended to decrease with lead time. On west regions (SW, W and NW) and at short lead days the ~~multi-model fore~~multi-  
320 model ensemble forecasts post-processed with the NGR were the most skillful; in the other cases the ECMWF forecasts post-  
321 processed with the NGR and the AKD methods tended to be best. The differences of BSS among regions were larger at longer  
322 lead times because the skill decreased more sharply over the Eastern regions. This issue is slightly addressed by the NGR and  
323 AKD methods based on the ECMWF.

Formatted: Highlight

### 324 3.1.3 Summary of average performance for daily forecast

325 Table 23 shows the average performance for the lead days 1 and 7, by weighting the values of each metric according to the  
326 number of stations in each region. The ECMWF-UKMO forecasts post-processed with the NGR method were best at short  
327 lead times (1-2 days), while the ECMWF forecasts post-processed with the AKD and the NGR methods were the first and  
328 second best at the longer lead times. The BMA method performed well at short lead times but poorly at long times, while the  
329 simple bias correction method performed well for deterministic forecasts, but poorly for the probabilistic forecasts. The  
330 forecast performance across climate regions is also associated with the choice of the ECMWF ensemble forecasts or the ~~multi-~~  
331 ~~model fore~~multi-model ensemble forecasts (Table A1, ANEX4). The single model ECMWF forecasts performed better over  
332 northern climate regions than the multi-model ensemble forecasts, while the multi-model did better than any single model  
333 forecast over the western regions. The performance over the other regions was more variable among strategies. The  
334 performance of the ECMWF-UKMO forecasts was generally better than that of the ECMWF-NCEP-UKMO forecasts (see  
335 Table 4A1, and Figs. 2 and 4). Unlike other performance metrics, the coverage was mostly better for the ECMWF ensemble  
336 forecasts than for the ~~multi-model fore~~multi-model ensemble forecasts. Our CRPS values seems comparable with those  
337 reported by Osnabrugge (2019) based on the ECMWF ensemble forecasts of potential evapotranspiration over the Rhine basin,  
338 in Europe.

### 339 3.1.3 Effect of the length of training period

340 The choice of an “optimum” training period is an important issue related to the operational use of post-processing techniques  
341 for ETo forecasts. Here we compared the performance of different forecasts post-processed with NGR and AKD techniques  
342 using 45 and 30 training days. The results suggest that the payoff from using 45 days is practically minimal. Table 5-A2 (Anex)  
343 shows the percentage differences the forecasting performance of using 45 and 30 training days for post-processing. While  
344 there are generally some minor improvements for using 45 days than 30 days, which tend to be higher at longer lead times  
345 than shorter times, these improvements usually represent less than 3 percent of original statistics. The largest percentage

346 difference, accounting for the BSS at the middle tercile, actually represented a negligible gain in absolute terms since they  
347 were affected by the close-to-zero range of the variable. The improvements were a bit higher for ~~multi-model fore~~multi-model  
348 ensemble forecasts than for single model forecasts. Notice that, while testing two different periods may be limited to evaluate  
349 the methods' sensitivity to the training period, they comprised the range for which methods such as the NGR and BMA have  
350 been reported to provide stable results (Gneiting et al., 2005; Raftery et al., 2005).

Formatted: Highlight

### 351 3.1.4 Weighting coefficients

352 The weighting coefficients reflect both the performance of the ensemble models and the performance the post-processing  
353 techniques relative to their counterparts. Figure 5 shows the mean  $b_i$  (Eq. 1) weighting coefficients of the NGR technique and  
354  $w_i$  (Eq. 7) weighting coefficient of the BMA techniques for each region and lead time for the post-processed ECMWF-NCEP-  
355 UKMO, respectively. The coefficients for the NGR and BMA techniques exhibited some common patterns of variability across  
356 regions and lead times. Both methods show that the weights of the ECMWF forecasts are at overall the highest, with a clear  
357 maximum at medium lead times. ~~T~~the weights of the UKMO the UKMO model are the highest at 1 and 2 days, but sharply  
358 decreases with the lead time, while the weights of the ~~of the~~ NCEP model are in general the lowest, although they consistently  
359 increase with lead time, ~~most likely because of the stronger decrease of performance with lead time by the other two models.~~  
360 It explains well the most outstanding features of the performance assessments, in relation to the role of each model, and the  
361 dependence on regions and lead times. Compared to the NGR method, the BMA method gives ~~the UKMO~~ the UKMO forecasts  
362 a higher relative weight, at the expense of the ECMWF forecast weights. For example, the weighting coefficients of the BMA  
363 method over the west regions are consistently higher for ~~the UKMO~~ the UKMO forecasts than for the ECMWF forecasts. It  
364 suggests that the lower performance of the BMA post-processing relative to the NGR and the AKD methods may be related  
365 to a misrepresentation of the model weights on the performance. This in turn may be caused by convergence problems during  
366 the parameter optimization with the expectation-maximization algorithm (Vrugt et al., 2008).

367 We observed considerable similarities on the distribution of variance coefficients for the NGR method (Eq. 2) and the AKD  
368 (Eq. 6) method after post-processing the ECMWF forecasts. The two methods also provide very similar adjustments on the  
369 mean forecast because, unlike the BMA method, they independently bias correct the mean and optimize the spread-skill  
370 relationship, (Bröcker and Smith, 2008). ~~However,~~ ~~However~~ the computing speed using the ~~in the experiment the NGR~~ NGR  
371 method ~~is~~ ~~was~~ about 60 faster than the AKD method ~~about 60 times faster than using the AKD, which was perceived as the~~  
372 ~~main drawback of the AKD method.~~ The BMA method ~~is also more computationally demanding~~ was also faster than the AKD  
373 ~~method,~~ ~~than the NGR method~~ but ~~less than the AKD still considerably slower than the NGR method.~~ Considering the  
374 effectiveness of the NGR method, ~~computational efficiency~~ and its versatility to post-process both single and multi-model  
375 ~~ensemble forecasts of the NGR method,~~ we applied this probabilistic technique to weekly ETo forecasts based on single model  
376 and multi-model ensembles.

### 3.2 Assessing NGR methods for post-processing weekly ETo forecasts

#### 3.2.1 Deterministic forecast assessments

As for the daily predictions, the bias, the RMSE and the correlation of the weekly forecasts post-processed with the NGR method and the linear regression methods were similar (Fig. 6). However, while the RMSE of daily forecasts based on ECMWF model varies between 12 and 20 % of the total ETo (Fig. 2), the RMSE for any of weekly forecasting strategies commonly varies between 8 and 11%, which is lower than for daily forecasts, making it more useful for operational purpose. The post-processed forecasts showed much lower RMSE and twice higher correlation than the predictions based on persistence, with the weekly predictions based on ECMWF forecasts being generally better, followed by the predictions based on the UKMO forecasts.

#### 3.2.2 Probabilistic forecast assessments

Both the skill and the reliability of the weekly forecasts considerably improved through the NGR post-processing compared with the bias correction post-processing (Table 36). The improvements were different among ETo forecast models. In most cases, the better the forecasts performance, the lower the improvements are. The adjustments in the coverage ratio and the Brier skill score were about 2.5 and 5 times larger for the UKMO and the NCEP forecasts, respectively, than for the ECMWF forecasts. The bias corrected ECMWF forecasts are generally better than both the UKMO and NCEP forecasts post-processed with the NGR method. We found that the post-processing of the NCEP forecasts with methods like the NGR is almost mandatory to get reasonable probabilistic weekly forecasts of ETo. For example, the coverage ratio of the bias corrected forecasts on the West region was only 29%, because of the considerable under-dispersion. However, it is notable that, once they were post-processed with the NGR technique, they performed almost comparably to the UKMO forecasts post-processed with the same method, increasing the coverage ratio to 98.4%. Table 36 also shows that the multi-model ECMWF- UKMO weekly forecasts are commonly the best among all of those post-processed using the NGR method, followed by the ECMWF and the ECMWF-NCEP-UKMO forecasts.

The improvements in the reliability came through substantial adjustments both in the ensemble spread and spread-skill relationship of the raw forecasts (Fig. 7). The correlations between the standard deviation of the ensembles and the RMSE were more than twice larger through the NGR post-processing than through the linear regression bias correction. The adjustments seemed even slightly more effective than those resulting from the probabilistic post-processing of the daily forecasts (Fig. 3), although at the expense of a greater loss of sharpness. The contrasts in the post-processing effectiveness are probably associated with the differences in the training strategies.

In the case of the probabilistic forecast skill (Fig. 8), the improvements were larger for the middle tercile than for the other two terciles, similarly as with daily forecasts. Unlike the bias corrected forecasts, any of the probabilistically post-processed forecasts outperform climatology for practically any event-tercile and at any region. Maybe more importantly, the skill-Brier scores for the lower and upper tercile events of the forecasts that have been post-processed with the NGR method is in most



409 cases over 30% better than the ~~skill-scores~~ of climatology. In the coast regions, from the South to the Northwest the ~~skill-score~~  
410 is commonly over 50% better, similarly as for the daily forecasts. Finally, the improvements resulting from the use of ~~multi-~~  
411 ~~model-fore~~~~multi-model ensemble forecasts~~ compared to the single model ~~ensemble~~ forecasts were generally small, except for  
412 the Southwest region.

#### 413 4. Discussion

##### 414 4.1 Effects of probabilistic post-processing on ETo forecasting performance

415 This study showed that NGR, AKD and BMA post-processing schemes considerably improved the probabilistic forecast  
416 performance (~~coverage ratio~~, ~~calibration~~, ~~spread-skill~~, ~~BSS~~, ~~CRPS~~) of the daily and weekly ETo forecasts compared with the  
417 simple (~~i.e., using linear regression based on ensemble mean~~) bias correction method. While sharpness is a wished quality of  
418 any forecast, the daily and weekly bias corrected ETo forecasts from NWP are spuriously sharp, which leads to a poor  
419 consistency between the range of the ETo forecasts and the true values, and ultimately undermine the confidence on those  
420 forecasts. They also ~~experiment-exhibit~~ a poor consistency in that the variance of the ensembles are commonly insensitive to  
421 the size of the forecast error. The probabilistic post-processed methods provided a much better reliability, with a coverage  
422 which is close to the nominal value, and at a low cost on sharpness. Therefore, they lead to a much better agreement between  
423 the forecasted probability of having an ETo event between certain thresholds and the proportions of times that the event occurs  
424 (see Gneiting et al., 2005).

425 In the case of the weekly ETo forecasts, the rate of the improvements are considerably smaller for the ECMWF forecasts, than  
426 for the UKMO, and especially the NCEP forecasts. ~~This seems to be largely due to the better performance of the ECMWF raw~~  
427 ~~forecasts compared to the other forecasting systems~~. The probabilistic post-processing of the weekly NCEP forecasts seemed  
428 practically mandatory to produce reasonable predictions, but once implemented it provided performance assessments almost  
429 comparable to those based on ~~the UKMO~~~~the UKMO~~ forecasts. These results have important implications for operational ETo  
430 forecasts, such as the U.S. national digital forecast database, one of the few operational products of its type, which are based  
431 on the NCEP forecasts.

432 Unlike the probabilistic forecast metrics, the deterministic metrics (~~bias~~~~ME~~, RMSE and correlation ~~od the ensemble mean~~) are  
433 low sensitive to the form (deterministic or probabilistic) of post-processing. In particular, the RMSE and correlation seemed  
434 more affected by the choice of the single or ~~multi-model-fore~~~~multi-model ensemble forecast~~ strategy than the choice between  
435 the NGR, the AKD or the simple bias correction as post-processing method. Whereas, RMSE and correlation provided by the  
436 BMA method are consistently worse at long lead times. The daily errors under any post-processing were relatively large, but  
437 mostly random, and therefore tend to cancel out at weekly scales. Therefore, while the RMSE varied between 12% and 20%  
438 of the daily totals, it represented between 8% and 11% of the weekly totals. The RMSE for weekly ETo forecasts were in all  
439 cases more than 100% lower than for the persistence-based ETo forecasts, and potentially more skillful than the forecasts that

Formatted: Highlight

Formatted: Highlight

440 exploit the temporal ~~autocorrelation-persistence~~ of the ETo timeseries (e.g.; Landaras et al., 2009; Mohan and Arumugam,  
441 2009).

#### 442 4.2 Comparing the three probabilistic post-processing methods

443 The NGR and AKD based post-processing methods for the ECMWF forecasts produced comparable results, indicating that  
444 the simple Gaussian predictive distribution from the NGR method represents fairly well the uncertainty of the ETo predictions.  
445 The methods led to similar distribution of the first two moments of the predictive probability function and similar performance  
446 statistics (with the AKD based forecasts being just slightly better). However, the NGR method ~~requires less computing time~~  
447 ~~and~~ is more versatile since it can be applied to correct both single model and multi-model ensemble forecasts, while the AKD  
448 method can only be applied to correct single model forecast. The NGR based predictive distribution function is also easier to  
449 ~~manipulate and~~ interpret than the AKD based predictive distribution, which is given by an averaged sum of standard Gaussians.  
450 The BMA method performed slightly less desirable compared to the NGR and AKD presumably due to issues with the  
451 parameter identifiability. The implemented method uses the Expectation-Maximization (EM) algorithm to produce maximum  
452 likelihood estimates of the fitting coefficients, which is susceptible to converge to local minima, especially when dealing with  
453 ~~multi-model fore~~multi-model ensemble forecasts with very different ensemble sizes (Vrugt et al., 2008). Archambeau et al.  
454 (2003) demonstrated that, in presence of outliers or repeated values, this algorithm tends to identify local maximums of the  
455 likelihood of the parameters of a Gaussian mixture model. Tian X. et al. (2012) found that adjusted BMA coefficients using  
456 both a quasi-Newtonian limited memory algorithm and the Markov Chain Monte Carlo were more accurate than those fitted  
457 with the EM algorithm, a procedure that is worth testing in future studies.

#### 458 4.3 Multi-model ensemble versus single model ensemble forecasts

459 Daily multi-model ensemble forecasts performed better (in terms of ME, RMSE, correlation, CRPS and BSS) than daily  
460 ECMWF forecasts at short lead times (1-2 days) and over the western and southern regions, while the ECMWF forecasts are  
461 better over the northeastern regions for longer lead times. For other region/lead time combinations the performance of single  
462 and multi-model ensemble forecasts did not differ much. We observed similar patterns for the raw and simple bias corrected  
463 forecasts (Medina et al., 2018). Whereas, ~~the effect of the~~ weekly multi-model foremulti-model ensemble forecast where  
464 consistently better than the weekly single-model forecasts only in the Southwest region is generally inconsistent at weekly  
465 seales, seemingly because the weekly forecasts logically involve both short and long lead time assessments, and the  
466 effectiveness of the multi-models is degraded for long lead times~~due to the variable impact of the forecasting strategy with~~  
467 lead days. The observed behavior is associated with the performance of the ECMWF forecasts relative to ~~the -UKMO~~the  
468 UKMO forecasts. While the ECMWF forecasts are in general better than ~~the -UKMO~~the UKMO and NCEP forecasts, they  
469 are much better over the northeastern regions for medium lead times (4-6 days). The -UKMO forecasts are in many cases the  
470 best at 1 and 2 lead days, but tend to be the worst at the longest times (6-7 days), especially over these regions. The NCEP

471 forecasts had a small contribution with respect to the ECMWF and -UKMO forecasts at short lead times. These forecasts are  
472 comparatively better at longer lead times, but still keep a minor role with regard to the ECMWF forecasts.

473 When considering daily forecasts we adopted a length of the training period of 30 days and showed that by increasing the  
474 length to 45 days the improvements were small (commonly lower than three percent). This seems a plausible range for future  
475 works and represents an obvious advantage upon methods such as the analog forecast, which provide similar performance  
476 (Tian and Martinez 2012 a, b, 2014) but require long training datasets. Gneiting et al. (2005) and Wilson (2007) found that  
477 lengths between 30 and 40 days provided good and almost constant performance assessments of sea level pressure forecasts  
478 post-processed with the NGR method, and temperatures forecasts post-processed with the BMA method, respectively.

#### 479 **4.4. Post-processing the individual inputs versus against post-processing ETo**

480 While in this study we considered the post-processing of ETo ensembles produced with raw NWP forecasts, a question is if  
481 by post-processing the forcing variables such as temperature, radiation and wind speed first, and then computing the ETo, we  
482 might have better predictions. The NGR method has been shown to be successful for the post-processing of surface  
483 temperatures (e.g. Wilks and Hamill, 2007), whose distribution is fairly Gaussian. For example, Hagedorn (2008) and  
484 Hagedorn et al. (2008) showed gains in lead time between two days and four days, with the gains being larger over areas where  
485 the raw forecast showed poor skill. Kann et al., (2009) and Kann et al., (2011), used the NGR method for improving short  
486 range ensemble forecasts of 2m-temperature. Recently, Scheuerer and Büermann (2014) provided a generalization of the  
487 original approach of Gneiting et al. (2005) that produces spatially calibrated probabilistic temperature forecasts. The wind-  
488 speed forecasts have been commonly post-processed with the use of quantile regression method (e.g. Bremnes 2004; Pinson  
489 et al. 2007; Møller et al., 2008). More recently Slougher et al. (2010) extended the original BMA method of Raftery et al.  
490 (2005) for wind speed, by considering a gamma distribution for modeling the distribution of every member of the ensemble,  
491 which considerably improved the CRPS, the absolute errors and the coverage. Whereas, Vanvyve et al., (2015) and Zhang et  
492 al. (2015) used the analog method following the methodology of Delle Monache (2013). The accurate solar radiation  
493 forecasting is particularly challenging because it requires detailed representation of the cloud fields (Verzijlbergh et al., 2015),  
494 which is usually not well resolved by the NWP models. Davò et al. (2016) used artificial neural networks (ANN) and the  
495 analog method approaches for the post-processing of both wind speed and solar radiation ensemble forecasts, which  
496 outperformed a simple bias correction approach. However, the post-processing of meteorological forecasts for producing ETo  
497 ensemble forecasts may require accounting for the multivariate dependence among those forcing, which is often difficult (e.g.  
498 Wilks, 2015). Kang et al (2010) found that post-processing of the streamflow forecasts provided more accurate predictions  
499 than post-processing the forcing alone, while Vekade et al (2013) showed that the improvements in precipitation and  
500 temperature through the post-processing hardly benefited the streamflow forecasts. Lewis et al., 2014 showed that the  
501 performance of the ETo forecasts can largely surpass that of the individual input variables. Therefore, it is unclear if we can  
502 have any benefit by using the post-processed inputs, instead of the raw forecasts, to construct ETo forecasts.

Formatted: Heading 2, Left, Line spacing: single

#### 4.45. Future outlook

It is worth noting that, while the ETo forecasts are produced for being used in agriculture, they were tested over USCRN stations, which are not representative of agricultural settings. In real applications, the bias between the forecasts with no post-processing and the measurements based on agricultural stations could be higher than the bias resolved in this study. A question that should be addressed in the future studies is to what extent the improvements of the predictive distribution of the ETo forecasts can be translated into a more reliable representation of the crop water use in agricultural lands and, ultimately, in water savings and economic gains. Since the ETo estimations can have remarkable impacts on the soil moisture estimations (Rodriguez-Iturbe et al., 1999), we envision that new studies relying on the combination of rainfall and ETo forecasts post-processed with probabilistic methods will lead to considerable reductions on the uncertainty of soil moisture forecasts. New attempts should also investigate the role of the ~~emerging state of art~~ probabilistic post-processing techniques on ETo forecasts produced from regional numerical weather prediction models, which have had improved spatial resolution and already been used in different meteorological services (e.g., Baldauf et al. 2011; Seity et al. 2011; Hong and Dudhia, 2012; Bentzien and Friederichs, 2012).

#### 5. Conclusions

This study for the first time evaluated probabilistic methods based on NGR, AKD, and BMA techniques for post-processing daily and weekly ETo forecasts derived from single or multi-model ~~ensemble~~ numerical weather predictions. The different ~~ETo forecasting strategies~~ ~~post-processing methods~~ were compared against the simple linear regression bias correction method using both daily and weekly forecasts, and also against persistence in the case of weekly forecasts. The probabilistic post-processing techniques largely modified the spread of the original ETo forecasts, with very favorably impacts on the probabilistic forecast performance. They corrected the notable under-dispersion and the poor consistency between the spread of the ETo forecasts and the dimension of the errors, leading to better ~~skill~~ ~~BSS~~, ~~and~~ ~~reliability~~ ~~(both coverage ratio and spread-skill)~~ ~~and CRPS~~. The adjustments were crucial on the performance of the weekly NCEP forecasts, followed by the weekly UKMO forecasts, whose bias corrected versions show a clear disadvantage compared with ~~simply post-processed the strategies that include the~~ ECMWF forecasts.

The ~~deterministic forecast performance~~ ~~deterministic performance~~ based on the ~~probabilistic~~ ~~NGR, AKD and BMA~~ methods were comparable to ~~the performance based on~~ the linear regression bias correction for both daily and weekly forecasts, and the skill is about 100% higher than those based on persistence in the case of the weekly forecasts. The ~~r~~RMSE are between 12 and 20% for the daily totals and 8 and 11% for the weekly totals. The NGR and AKD provided similar estimates of the first and second order moments of the predictive density distribution; they showed similar effectiveness, but the NGR method ~~has the advantage that can post-process both single and multi-model ensemble forecasts~~ ~~exhibited higher flexibility and computational efficiency~~. Both NGR and AKD post-processing methods outperformed the BMA method when considering daily forecasts at long lead times.

535 The ~~multi-model fore~~multi-model ensemble forecasting provided benefits at daily scales compared to the ECMWF ensemble  
536 forecasting, while the benefits were marginal at weekly scales. The multi-model ensemble forecasting seems a better choice  
537 when ~~the UKMO~~the UKMO forecasts are comparable or slightly better than the ECMWF forecasts, such as at short (1-2 days)  
538 lead times and over the southern and western regions. Post-processing single model forecast is a better choice than post-  
539 processing multi-model ensemble forecast in the circumstances where the ECMWF forecasts perform considerably better than  
540 the UKMO and NCEP, such as at mid and long lead times, especially over the northeastern regions. While we considered a  
541 length of the training period of 30 days for daily post-processing, the increase of the training period to 45 days only led to  
542 minimal improvements. In conclusion, our results suggest that the NGR post-processing of ETo forecasts generated from the  
543 ECMWF or ECMWF-UKMO predictions is the most plausible strategy among those being evaluated, and is recommended  
544 for operational implementations, because accuracy and reliability requirements for practical applications have not been  
545 dicussed.

#### 546 Acknowledgement

547 This research was supported in part by the Alabama Agricultural Experiment Station and the Hatch program of the National  
548 Institute of Food and Agriculture (NIFA), U.S. Department of Agriculture (USDA, Access No. 1012578), by the Auburn  
549 University Intramural Grant Program, by the Auburn University Presidential Awards for Interdisciplinary Research, and by  
550 the USDA-NIFA Agriculture and Food Research Initiative (AFRI) competitive grant. The authors want to thank the very  
551 helpful comments of the reviewers

#### 552 Code/Data availability

553 Request for materials should be addressed to Di Tian.

#### 554 Author contributions

555 Hanoi Medina and Di Tian designed and conceptualized the research. Hanoi Medina implemented the design, performed data  
556 curation, analysis, validation, visualization, and wrote the original draft. Di Tian supervised the research, contributed by advice,  
557 and reviewed and edited the manuscript.

#### 558 Competing interests

559 The authors declare that they have no conflict of interest.

← -- Formatted: Heading 1, Left, Line spacing: single

← -- Formatted: Heading 1, Left, Line spacing: single

← -- Formatted: Heading 1, Left, Line spacing: single

560 **References**

- 561 [Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water](#)  
562 [requirements-FAO, Irrigation and drainage paper 56, Fao, Rome, 300\(9\), p.D05109, 1998.](#)
- 563 [Archambeau, C., Lee, J. A. and Verleysen, M.: On Convergence Problems of the EM Algorithm for Finite Gaussian Mixtures,](#)  
564 [In ESANN \(Vol. 3, pp. 99-106\), 2003.](#)
- 565 [Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T.: Operational convective-scale](#)  
566 [numerical weather prediction with the COSMO model: Description and sensitivities, Monthly Weather Review, 139\(12\),](#)  
567 [pp.3887-3905, 2011.](#)
- 568 [Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525\(7567\): 47-55, 2015.](#)
- 569 [Bentzien, S. and Friederichs, P.: Generating and calibrating probabilistic quantitative precipitation forecasts from the high-](#)  
570 [resolution NWP model COSMO-DE. Weather and Forecasting, 27\(4\), pp.988-1002, 2012.](#)
- 571 [Beran, R. and Hall, P.: Interpolated nonparametric prediction intervals and confidence intervals, Journal of the Royal Statistical](#)  
572 [Society, Series B \(Methodological\), pp.643-652, 1993.](#)
- 573 [Bremnes, J. B.: Probabilistic Wind Power Forecasts Using Local Quantile Regression, Wind Energy, 7, 47–54, 2004.](#)
- 574 [Bröcker, J. and Smith, L. A.: From ensemble forecasts to predictive distribution functions, Tellus A: Dynamic Meteorology](#)  
575 [and Oceanography, 60\(4\), pp.663-678, 2008.](#)
- 576 [Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M.: A comparison of the ECMWF, MSC, and NCEP](#)  
577 [global ensemble prediction systems, Monthly Weather Review, 133\(5\), pp.1076-1097, 2005.](#)
- 578 [Casella, G. and Berger, R. L. : Statistical inference \(Vol. 2\). Pacific Grove, CA: Duxbury, 2002.](#)
- 579 [Castro, F. X., Tudela, A. and Sebastià, M. T.: Modeling moisture content in shrubs to predict fire risk in Catalonia \(Spain\),](#)  
580 [Agricultural and Forest Meteorology, 116\(1-2\), pp.49-59, 2003.](#)
- 581 [Chirico, G. B., Pelosi, A., De Michele, C., Bolognesi, S. F. and D'Urso, G.: Forecasting potential evapotranspiration by](#)  
582 [combining numerical weather predictions and visible and near-infrared satellite images: an application in southern Italy, The](#)  
583 [Journal of Agricultural Science, pp.1-9. <https://doi.org/10.1017/S0021859618000084>, 2018.](#)
- 584 [Davò, F., Alessandrini, S., Sperati, S., Delle Monache, L., Airolidi, D. and Vespucci, M. T.: Post-processing techniques and](#)  
585 [principal component analysis for regional wind power and solar irradiance forecasting, Solar Energy, 134, pp.327-338, 2016](#)
- 586 [Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B. and Searight, K.: Probabilistic weather prediction with an analog](#)  
587 [ensemble, Monthly Weather Review, 141\(10\), pp.3498-3516, 2013.](#)
- 588 [Fraley, C., Raftery, A. E. and Gneiting, T.: Calibrating multimodelmulti-model forecast ensembles with exchangeable and](#)  
589 [missing members using Bayesian model averaging, Monthly Weather Review, 138\(1\), pp.190-202, 2010.](#)
- 590 [Fraley, C., Raftery, A. E., Slougher, J. M., Gneiting T.: EnsembleBMA: Probabilistic Forecasting using Ensembles and](#)  
591 [Bayesian Model Averaging. R package version 5.1.3. <https://CRAN.R-project.org/package=ensembleBMA>, 2016.](#)

592 [Glahn, H. R. and Lowry, D. A.: The use of model output statistics \(MOS\) in objective weather forecasting. J Appl Meteorol,](#)  
593 [11\(8\): 1203-1211, 1972.](#)  
594 [Glahn, H. R. and Ruth, D. P.: The new digital forecast database of the National Weather Service, Bulletin of the American](#)  
595 [Meteorological Society, 84\(2\), pp.195-202, 2003.](#)  
596 [Gneiting, T., Raftery, A. E., Westveld III, A. H. and Goldman, T.: Calibrated probabilistic forecasting using ensemble model](#)  
597 [output statistics and minimum CRPS estimation, Monthly Weather Review, 133\(5\), 1098-1118., 2005.](#)  
598 [Gneiting, T.: Calibration of medium-range weather forecasts, European Centre for Medium-Range Weather Forecasts,](#)  
599 [Technical Memorandum No. 71, 2014.](#)  
600 [Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M. and Palmer, T. N.: Comparing TIGGE multimodelmulti-model](#)  
601 [forecasts with reforecast-calibrated ECMWF ensemble forecasts. Q J Roy Meteor Soc, 138\(668\): 1814-1827, 2012.](#)  
602 [Hagedorn, R., Hamill, T. M. and Whitaker, J. S.: Probabilistic forecast calibration using ECMWF and GFS ensemble](#)  
603 [reforecasts. Part I: Two-meter temperatures. Monthly Weather Review, 136, 2608–2619, 2008.](#)  
604 [Hagedorn, R.: Using the ECMWF reforecast data set to calibrate EPS forecasts. ECMWF Newsletter, 117, 8–13, 2008.](#)  
605 [Hamill, T. M. and Colucci, S. J.: Verification of Eta–RSM short-range ensemble forecasts, Monthly Weather Review, 125\(6\),](#)  
606 [pp.1312-1327, 1997.](#)  
607 [Hamill, T. M. and Whitaker, J. S.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and](#)  
608 [application, Mon Weather Rev, 134\(11\): 3209-3229, 2006.](#)  
609 [Hamill, T. M. et al.: Noaa's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. B Am Meteorol Soc,](#)  
610 [94\(10\): 1553-1565, 2013.](#)  
611 [Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and](#)  
612 [Forecasting, 15\(5\), pp.559-570, 2000.](#)  
613 [Hobbins, M., McEvoy, D. and Hain, C.: Evapotranspiration, evaporative demand, and drought, Drought and Water Crises:](#)  
614 [Science, Technology, and Management Issues, pp.259-288, 2017.](#)  
615 [Hong, S. Y. and Dudhia, J.: Next-generation numerical weather prediction: Bridging parameterization, explicit clouds, and](#)  
616 [large eddies, Bulletin of the American Meteorological Society, 93\(1\), pp.ES6-ES9., 2012.](#)  
617 [Ishak, A. M., Bray, M., Remesan, R. and Han, D.: Estimating reference evapotranspiration using numerical weather modelling,](#)  
618 [Hydrological processes, 24\(24\), pp.3490-3509, 2010.](#)  
619 [Kang, T. H., Kim, Y. O. and Hong, I. P.: Comparison of pre - and post - processors for ensemble streamflow prediction,](#)  
620 [Atmospheric Science Letters, 11\(2\), pp.153-159, 2010.](#)  
621 [Kann, A., Haiden, T. and Wittmann, C.: Combining 2-m temperature nowcasting and short-range ensemble forecasting,](#)  
622 [Nonlinear Processes in Geophysics, 18, 903–910, 2011.](#)  
623 [Kann, A., Wittmann, C., Wang, Y. and Ma, X.: Calibrating 2-m temperature of limited-area ensemble forecasts using high-](#)  
624 [resolution analysis. Monthly Weather Review, 137, 3373–3387, 2009.](#)

625 [Klein, W. H. and Glahn, H. R.: Forecasting local weather by means of model output statistics, \*Bulletin of the American\*](#)  
626 [\*Meteorological Society\*, 55\(10\), pp.1217-1227, 1974.](#)

627 [Landeras, G., Ortiz-Barredo, A. and López, J. J.: Forecasting weekly evapotranspiration with ARIMA and artificial neural](#)  
628 [network models, \*Journal of irrigation and drainage engineering\*, 135\(3\), pp.323-334, 2009.](#)

629 [Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, \*Journal of Computational Physics\*, 227\(7\), pp.3515-3539, 2008.](#)

630 [Madadgar, S., Moradkhani, H. and Garen, D.: Towards improved post-processing of hydrologic forecast ensembles,](#)  
631 [\*Hydrological Processes\*, 28\(1\), pp.104-122, 2014.](#)

632 [Mase, A. S. and Prokopy, L. S.: Unrealized potential: A review of perceptions and use of weather and climate information in](#)  
633 [agricultural decision making, \*Weather, Climate, and Society\*, 6\(1\), pp.47-61, 2014.](#)

634 [Medina, H., Tian, D., Marin, F. R. and Chirico, G. B.: Comparing GEFS, ECMWF, and Postprocessing Methods for Ensemble](#)  
635 [Precipitation Forecasts over Brazil, \*Journal of Hydrometeorology\*, 20\(4\), pp.773-790, 2019.](#)

636 [Medina, H., Tian, D., Srivastava, P., Pelosi, A. and Chirico, G. B.: Medium-range reference evapotranspiration forecasts for](#)  
637 [the contiguous United States based on multimodelmulti-model numerical weather predictions, \*Journal of Hydrology\*, 562,](#)  
638 [pp.502-517, 2018.](#)

639 [Messner, J. W., Mayr, G. J., Zeileis, A. and Wilks, D. S.: Heteroscedastic Extended Logistic Regression for Postprocessing of](#)  
640 [Ensemble Guidance, \*Mon. Wea. Rev.\*, 142, 448–456, <https://doi.org/10.1175/MWR-D-13-00271.1>, 2014.](#)

641 [Mohan, S. and Arumugam, N.: Forecasting weekly reference crop evapotranspiration series, \*Hydrological sciences journal\*,](#)  
642 [40\(6\), pp.689-702, 1995.](#)

643 [Møller, J. K., Nielsen, H. A., and Madsen, H.: Time-Adaptive Quantile Regression, \*Computational Statistics & Data Analysis\*,](#)  
644 [52, 1292–1303, 2008.](#)

645 [National Research Council of the National Academies: Completing the Forecast: Characterizing and Communicating](#)  
646 [Uncertainty for Better Decisions Using Weather and Climate Forecasts, The National Academies Press, 124 pp, 2006.](#)

647 [Osnabrugge, B. V., Uijlenhoet, R. and Weerts, A.: Contribution of potential evaporation forecasts to 10-day streamflow](#)  
648 [forecast skill for the Rhine River, \*Hydrology and Earth System Sciences\*, 23\(3\), pp.1453-1467, 2019.:](#)

649 [Pelosi, A., Medina, H., Van den Bergh, J., Vannitsem, S., and Chirico, G. B.: Adaptive Kalman filtering for post-processing](#)  
650 [ensemble numerical weather predictions, \*Mon Weather Rev.\* \[doi.org/10.1175/MWR-D-17-0084.1\]\(https://doi.org/10.1175/MWR-D-17-0084.1\), 2017.](#)

651 [Pelosi, A., Medina, H., Villani, P., D'Urso, G. and Chirico, G. B.: Probabilistic forecasting of reference evapotranspiration](#)  
652 [with a limited area ensemble prediction system, \*Agricultural water management\*, 178, pp.106-118, 2016.](#)

653 [Perera, K. C., Western, A. W., Nawarathna, B. and George, B.: Forecasting daily reference evapotranspiration for Australia](#)  
654 [using numerical weather prediction outputs, \*Agr Forest Meteorol\*, 194: 50-63, 2014.](#)

655 [Pinson, P., and Madsen, H.: Ensemble-Based Probabilistic Forecasting at Horns Rev, \*Wind Energy\*, 12, 137–155, 2009.](#)

656 [Prokopy, L. S., Haigh, T., Mase, A. S., Angel, J., Hart, C., Knutson, C., Lemos, M. C., Lo, Y. J., McGuire, J., Morton, L. W,](#)  
657 [and Perron, J.: Agricultural advisors: a receptive audience for weather and climate information?, \*Weather, Climate, and\*](#)  
658 [\*Society\*, 5\(2\), pp.162-167, 2013.](#)



659 [R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna,](#)  
660 [Austria, http://www.R-project.org/, 2014.](http://www.R-project.org/)

661 [Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian model averaging to calibrate forecast](#)  
662 [ensembles, Monthly Weather Review, 133\(5\), pp.1155-1174, 2005.](#)

663 [Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V. and Cox, D. R.: Probabilistic modelling of water balance at a point:](#)  
664 [the role of climate, soil and vegetation, Proceedings of the Royal Society of London, Series A: Mathematical, Physical and](#)  
665 [Engineering Sciences, 455\(1990\), pp.3789-3805, 1999.](#)

666 [Roulston, M. S. and Smith, L. A.: Combining dynamical and statistical ensembles, Tellus A: Dynamic Meteorology and](#)  
667 [Oceanography, 55\(1\), pp.16-30, 2003.](#)

668 [Scheuerer, M. and Büermann, L.: Spatially adaptive post - processing of ensemble forecasts for temperature, Journal of the](#)  
669 [Royal Statistical Society: Series C \(Applied Statistics\), 63\(3\), pp.405-422, 2014.](#)

670 [Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V.: The AROME-France](#)  
671 [convective-scale operational model, Monthly Weather Review, 139\(3\), pp.976-991, 2011.](#)

672 [Siegert, S.: SpecsVerification: Forecast Verification Routines for Ensemble Forecasts of Weather and Climate. R package](#)  
673 [version 0.5-2. https://CRAN.R-project.org/package=SpecsVerificatio, 2017.](#)

674 [Silva, D., Meza, F. J. and Varas, E.: Estimating reference evapotranspiration \(ETo\) using numerical weather forecast data in](#)  
675 [central Chile, Journal of hydrology, 382\(1-4\), pp.64-71, 2010.](#)

676 [Slughter, J. M., Gneiting, T. and Raftery, A. E.: Probabilistic wind speed forecasting using ensembles and Bayesian model](#)  
677 [averaging, Journal of the american statistical association, 105\(489\), pp.25-35, 2010.](#)

678 [Swinbank, R. et al.: The Tigge Project and Its Achievements. B Am Meteorol Soc, 97\(1\): 49-67, 2016.](#)

679 [Tian, D. and Martinez, C. J.: Comparison of two analog-based downscaling methods for regional reference evapotranspiration](#)  
680 [forecasts, J Hydrol, 475: 350-364, 2012a](#)

681 [Tian, D. and Martinez, C. J.: Forecasting Reference Evapotranspiration Using Retrospective Forecast Analogs in the](#)  
682 [Southeastern United States, J Hydrometeorol, 13\(6\): 1874-1892, 2012b](#)

683 [Tian, D. and Martinez, C. J.: The GEFs-based daily reference evapotranspiration \(ETo\) forecast and its implication for water](#)  
684 [management in the southeastern United States. J Hydrometeorol, 15\(3\): 1152-1165, 2014.](#)

685 [Tian, X., Xie, Z., Wang, A. and Yang, X.: A new approach for Bayesian model averaging, Science China Earth Sciences,](#)  
686 [55\(8\), 1336-1344, 2012.](#)

687 [Toth, Z., Talagrand, O., Candille, G. and Zhu, Y.: Probability and ensemble forecasts, Forecast Verification: A Practitioner's](#)  
688 [Guide in Atmospheric Science, pp.137-163, 2003.](#)

689 [Vanvyve, E., Delle Monache, L., Monaghan, A.J. and Pinto, J.O.: Wind resource estimates with an analog ensemble approach,](#)  
690 [Renewable Energy, 74, pp.761-773, 2015.](#)

691 [Verkade, J. S., Brown, J. D., Reggiani, P. and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble](#)  
692 [reforecasts for operational hydrologic forecasting at various spatial scales, Journal of Hydrology, Volume 501,2013, Pages 73-](#)  
693 [91,http://dx.doi.org/10.1016/j.jhydrol.2013.07.039, 2013.](#)

694 [Verkade, J. S., Brown, J. D., Reggiani, P. and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble](#)  
695 [reforecasts for operational hydrologic forecasting at various spatial scales, Journal of Hydrology, 501, pp.73-91, 2013.](#)

696 [Verzijlbergh, R. A., Heijnen, P. W., de Roode, S. R., Los, A. and Jonker, H. J.: Improved model output statistics of numerical](#)  
697 [weather prediction based irradiance forecasts for solar power applications, Solar Energy, 118, pp.634-645, 2015](#)

698 [Vrugt, J. A., Diks, C. G. and Clark, M. P.: Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling,](#)  
699 [Environmental fluid mechanics, 8\(5-6\), pp.579-595, 2008.](#)

700 [Wang, X. and Bishop, C. H.: Improvement of ensemble reliability with a new dressing kernel, Quarterly Journal of the Royal](#)  
701 [Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131\(607\),](#)  
702 [pp.965-986, 2005.](#)

703 [Whan, K. and Schmeits, M: Comparing Area Probability Forecasts of \(Extreme\) Local Precipitation Using Parametric and](#)  
704 [Machine Learning Statistical Postprocessing Methods. Mon. Wea. Rev., 146, 3651–3673, https://doi.org/10.1175/MWR-D-](#)  
705 [17-0290.1., 2018.](#)

706 [Wilks, D. S. and Hamill, T. M.: Comparison of ensemble-MOS methods using GFS reforecasts, Monthly Weather Review,](#)  
707 [135\(6\), pp.2379-2390, 2007.](#)

708 [Wilks, D. S.: Comparison of ensemble-MOS methods in the Lorenz'96 setting, Meteorological Applications, 13\(3\), pp.243-](#)  
709 [256, 2006.](#)

710 [Wilks, D. S.: Extending logistic regression to provide full probability distribution MOS forecasts, Meteorological Applications:](#)  
711 [A journal of forecasting, practical applications, training techniques and modelling, 16\(3\), pp.361-368, 2009.](#)

712 [Wilks, D. S.: Multivariate ensemble Model Output Statistics using empirical copulas, Quarterly Journal of the Royal](#)  
713 [Meteorological Society, 141\(688\), pp.945-952, 2015.](#)

714 [Wilks, D.S.: Sampling distributions of the Brier score and Brier skill score under serial dependence, Q J Roy Meteor Soc,](#)  
715 [136\(653\): 2109-2118, 2010.](#)

716 [Williams, R. M., Ferro, C. A. T. and Kwasniok, F.: A comparison of ensemble post-processing methods for extreme events,](#)  
717 [Quarterly Journal of the Royal Meteorological Society, 140\(680\), pp.1112-1120, 2014.](#)

718 [Wilson, L. J., Bearegard, S., Raftery, A. E. and Verret, R.: Calibrated surface temperature forecasts from the Canadian](#)  
719 [ensemble prediction system using Bayesian model averaging, Monthly Weather Review, 135\(4\), pp.1364-1385, 2007.](#)

720 [Yuen, R., Baran, S., Fraley, C., Gneiting, T., Lerch, S., Scheuerer, M., and Thorarinsdottir, T.: ensembleMOS: Ensemble](#)  
721 [Model Output Statistics. R package version 0.8.2. https://CRAN.R-project.org/package=ensembleMOS, 2018.](#)

722 [Zhang, J., Draxl, C., Hopson, T., Delle Monache, L., Vanvyve, E. and Hodge, B. M.: Comparison of numerical weather](#)  
723 [prediction based deterministic and probabilistic wind resource assessment methods, Applied Energy, 156, pp.528-541, 2015.](#)

724 [Zhao, T., Wang, Q. J. and Schepen, A.: A Bayesian modelling approach to forecasting short-term reference crop](#)  
725 [evapotranspiration from GCM outputs, \*Agricultural and Forest Meteorology\*, 269, pp.88-101, 2019.](#) Allen, R. G., Pereira, L.,  
726 S., Raes, D. and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO, Irrigation and  
727 drainage paper 56, Fao, Rome, 300(9), p.D05109, 1998.

728 [Arehambeau, C., Lee, J. A. and Verleysen, M.: On Convergence Problems of the EM Algorithm for Finite Gaussian Mixtures,](#)  
729 [In ESANN \(Vol. 3, pp. 99-106\), 2003.](#)

730 [Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T.: Operational convective-scale](#)  
731 [numerical weather prediction with the COSMO model: Description and sensitivities, \*Monthly Weather Review\*, 139\(12\),](#)  
732 [pp.3887-3905, 2011.](#)

733 [Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction, \*Nature\*, 525\(7567\): 47-55, 2015.](#)

734 [Bentzien, S. and Friederichs, P.: Generating and calibrating probabilistic quantitative precipitation forecasts from the high-](#)  
735 [resolution NWP model COSMO-DE. \*Weather and Forecasting\*, 27\(4\), pp.988-1002, 2012.](#)

736 [Beran, R. and Hall, P.: Interpolated nonparametric prediction intervals and confidence intervals, \*Journal of the Royal Statistical\*](#)  
737 [Society, Series B \(Methodological\), pp.643-652, 1993.](#)

738 [Bröcker, J. and Smith, L. A.: From ensemble forecasts to predictive distribution functions, \*Tellus A: Dynamic Meteorology\*](#)  
739 [and Oceanography, 60\(4\), pp.663-678, 2008.](#)

740 [Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y. and Wei, M.: A comparison of the ECMWF, MSC, and NCEP](#)  
741 [global ensemble prediction systems, \*Monthly Weather Review\*, 133\(5\), pp.1076-1097, 2005.](#)

742 [Casella, G. and Berger, R. L. : Statistical inference \(Vol. 2\). Pacific Grove, CA: Duxbury, 2002.](#)

743 [Castro, F. X., Tudela, A. and Sebastià, M. T.: Modeling moisture content in shrubs to predict fire risk in Catalonia \(Spain\).](#)  
744 [Agricultural and Forest Meteorology, 116\(1-2\), pp.49-59, 2003.](#)

745 [Chirico, G. B., Pelosi, A., De Michele, C., Bolognesi, S. F. and D'Urso, G.: Forecasting potential evapotranspiration by](#)  
746 [combining numerical weather predictions and visible and near-infrared satellite images: an application in southern Italy, \*The\*](#)  
747 [Journal of Agricultural Science](#), pp.1-9. <https://doi.org/10.1017/S0021859618000084>, 2018.

748 [Fraley, C., Raftery, A. E. and Gneiting, T.: Calibrating multimodel/multi-model forecast ensembles with exchangeable and](#)  
749 [missing members using Bayesian model averaging, \*Monthly Weather Review\*, 138\(1\), pp.190-202, 2010.](#)

750 [Fraley, C., Raftery, A. E., Slaughter, J. M., Gneiting T.: EnsembleBMA: Probabilistic Forecasting using Ensembles and](#)  
751 [Bayesian Model Averaging. R package version 5.1.3. <https://CRAN.R-project.org/package=ensembleBMA>, 2016.](#)

752 [Glahn, H. R. and Lowry, D. A.: The use of model output statistics \(MOS\) in objective weather forecasting, \*J Appl Meteorol\*,](#)  
753 [11\(8\): 1203-1211, 1972.](#)

754 [Glahn, H. R. and Ruth, D. P.: The new digital forecast database of the National Weather Service, \*Bulletin of the American\*](#)  
755 [Meteorological Society](#), 84(2), pp.195-202, 2003.

756 [Gneiting, T., Raftery, A. E., Westveld III, A. H. and Goldman, T.: Calibrated probabilistic forecasting using ensemble model](#)  
757 [output statistics and minimum CRPS estimation, \*Monthly Weather Review\*, 133\(5\), 1098-1118., 2005.](#)

758 Gneiting, T.: Calibration of medium-range weather forecasts, European Centre for Medium-Range Weather Forecasts,  
759 Technical Memorandum No. 71, 2014.

760 Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M. and Palmer, T. N.: Comparing TIGGE multimodelmulti-model  
761 forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Q J Roy Meteor Soc*, 138(668): 1814-1827, 2012.

762 Hamill, T. M. and Colucci, S. J.: Verification of Eta-RSM short-range ensemble forecasts, *Monthly Weather Review*, 125(6),  
763 pp.1312-1327, 1997.

764 Hamill, T. M. and Whitaker, J. S.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and  
765 application, *Mon Weather Rev*, 134(11): 3209-3229, 2006.

766 Hamill, T. M. et al.: NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *B Am Meteorol Soc*,  
767 94(10): 1553-1565, 2013.

768 Hobbins, M., McEvoy, D. and Hain, C.: Evapotranspiration, evaporative demand, and drought, *Drought and Water Crises: Science, Technology, and Management Issues*, pp.259-288, 2017.

769 Hong, S. Y. and Dudhia, J.: Next-generation numerical weather prediction: Bridging parameterization, explicit clouds, and  
770 large eddies, *Bulletin of the American Meteorological Society*, 93(1), pp.ES6-ES9., 2012.

771 Ishak, A. M., Bray, M., Remesan, R. and Han, D.: Estimating reference evapotranspiration using numerical weather modelling,  
772 *Hydrological processes*, 24(24), pp.3490-3509, 2010.

773 Klein, W. H. and Glahn, H. R.: Forecasting local weather by means of model output statistics, *Bulletin of the American  
774 Meteorological Society*, 55(10), pp.1217-1227, 1974.

775 Landeras, G., Ortiz-Barredo, A. and López, J. J.: Forecasting weekly evapotranspiration with ARIMA and artificial neural  
776 network models, *Journal of irrigation and drainage engineering*, 135(3), pp.323-334, 2009.

777 Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, *Journal of Computational Physics*, 227(7), pp.3515-3539, 2008.

778 Mase, A. S. and Prokopy, L. S.: Unrealized potential: A review of perceptions and use of weather and climate information in  
779 agricultural decision-making, *Weather, Climate, and Society*, 6(1), pp.47-61, 2014.

780 Medina, H., Tian, D., Marin, F. R. and Chirico, G. B.: Comparing GEFS, ECMWF, and Postprocessing Methods for Ensemble  
781 Precipitation Forecasts over Brazil, *Journal of Hydrometeorology*, 20(4), pp.773-790, 2019.

782 Medina, H., Tian, D., Srivastava, P., Pelosi, A. and Chirico, G. B.: Medium-range reference evapotranspiration forecasts for  
783 the contiguous United States based on multimodelmulti-model numerical weather predictions, *Journal of Hydrology*, 562,  
784 pp.502-517, 2018.

785 Mohan, S. and Arumugam, N.: Forecasting weekly reference crop evapotranspiration series, *Hydrological sciences journal*,  
786 40(6), pp.689-702, 1995.

787 National Research Council of the National Academies: Completing the Forecast: Characterizing and Communicating  
788 Uncertainty for Better Decisions Using Weather and Climate Forecasts, The National Academies Press, 124 pp, 2006.

789 Pelosi, A., Medina, H., Van den Bergh, J., Vannitsem, S., and Chirico, G. B.: Adaptive Kalman filtering for post-processing  
790 ensemble numerical weather predictions, *Mon Weather Rev*, doi.org/10.1175/MWR-D-17-0084., 2017.

791

792 Pelosi, A., Medina, H., Villani, P., D'Urso, G. and Chirico, G. B.: Probabilistic forecasting of reference evapotranspiration  
793 with a limited area ensemble prediction system, *Agricultural water management*, 178, pp.106-118, 2016.

794 Perera, K. C., Western, A. W., Nawarathna, B. and George, B.: Forecasting daily reference evapotranspiration for Australia  
795 using numerical weather prediction outputs, *Agr Forest Meteorol*, 194: 50-63, 2014.

796 Prokopy, L. S., Haigh, T., Mase, A. S., Angel, J., Hart, C., Knutson, C., Lemos, M. C., Lo, Y. J., McGuire, J., Morton, L. W.  
797 and Perron, J.: Agricultural advisors: a receptive audience for weather and climate information?, *Weather, Climate, and*  
798 *Society*, 5(2), pp.162-167, 2013.

799 R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna,  
800 Austria, <http://www.R-project.org/>, 2014.

801 Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian model averaging to calibrate forecast  
802 ensembles, *Monthly Weather Review*, 133(5), pp.1155-1174, 2005.

803 Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V. and Cox, D. R.: Probabilistic modelling of water balance at a point:  
804 the role of climate, soil and vegetation, *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and*  
805 *Engineering Sciences*, 455(1990), pp.3789-3805, 1999.

806 Roulston, M. S. and Smith, L. A.: Combining dynamical and statistical ensembles, *Tellus A: Dynamic Meteorology and*  
807 *Oceanography*, 55(1), pp.16-30, 2003.

808 Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V.: The AROME France  
809 convective scale operational model, *Monthly Weather Review*, 139(3), pp.976-991, 2011.

810 Siebert, S.: SpeesVerification: Forecast Verification Routines for Ensemble Forecasts of Weather and Climate. R package  
811 version 0.5-2. <https://CRAN.R-project.org/package=SpeesVerification>, 2017.

812 Silva, D., Meza, F. J. and Varas, E.: Estimating reference evapotranspiration (ET<sub>0</sub>) using numerical weather forecast data in  
813 central Chile, *Journal of Hydrology*, 382(1-4), pp.64-71, 2010.

814 Swinbank, R. et al.: The Tigge Project and Its Achievements. *B Am Meteorol Soc*, 97(1): 49-67, 2016.

815 Tian, D. and Martinez, C. J.: Comparison of two analog-based downscaling methods for regional reference evapotranspiration  
816 forecasts, *J Hydrol*, 475: 350-364, 2012a

817 Tian, D. and Martinez, C. J.: Forecasting Reference Evapotranspiration Using Retrospective Forecast Analogs in the  
818 Southeastern United States, *J Hydrometeorol*, 13(6): 1874-1892, 2012b

819 Tian, D. and Martinez, C. J.: The GEFS-based daily reference evapotranspiration (ET<sub>0</sub>) forecast and its implication for water  
820 management in the southeastern United States. *J Hydrometeorol*, 15(3): 1152-1165, 2014.

821 Tian, X., Xie, Z., Wang, A. and Yang, X.: A new approach for Bayesian model averaging, *Science China Earth Sciences*,  
822 55(8), 1336-1344, 2012.

823 Toth, Z., Talagrand, O., Candille, G. and Zhu, Y.: Probability and ensemble forecasts, *Forecast Verification: A Practitioner's*  
824 *Guide in Atmospheric Science*, pp.137-163, 2003.

825 Vrugt, J. A., Diks, C. G. and Clark, M. P.: Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling,  
 826 *Environmental fluid mechanics*, 8(5-6), pp.579-595, 2008.

827 Wang, X. and Bishop, C. H.: Improvement of ensemble reliability with a new dressing kernel, *Quarterly Journal of the Royal  
 828 Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(607),  
 829 pp.965-986, 2005.

830 Wilks, D. S. and Hamill, T. M.: Comparison of ensemble-MOS methods using GFS reforecasts, *Monthly Weather Review*,  
 831 135(6), pp.2379-2390, 2007.

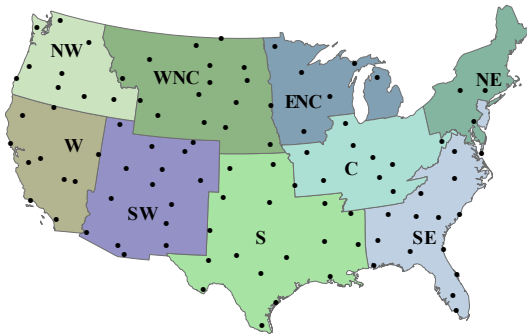
832 Wilks, D. S.: *Statistical methods in the atmospheric sciences*, (Vol 100), Academic press, 2011.

833 Wilson, L. J., Beaugard, S., Raftery, A. E. and Verret, R.: Calibrated surface temperature forecasts from the Canadian  
 834 ensemble prediction system using Bayesian model averaging, *Monthly Weather Review*, 135(4), pp.1364-1385, 2007.

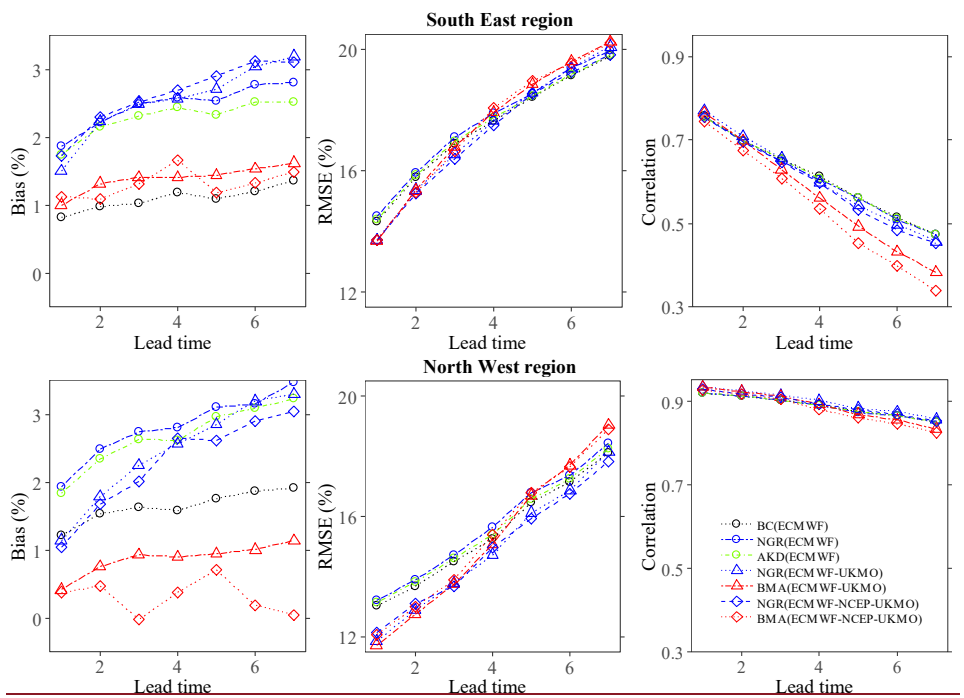
835 Yuen, R., Baran, S., Fraley, C., Gneiting, T., Lerch, S., Scheuerer, M., and Thorarindottir, T.: ensembleMOS: Ensemble  
 836 Model Output Statistics. R package version 0.8.2. <https://CRAN.R-project.org/package=ensembleMOS>, 2018.

837 Zhao, T., Wang, Q. J. and Schepen, A.: A Bayesian modelling approach to forecasting short-term reference crop  
 838 evapotranspiration from GCM outputs, *Agricultural and Forest Meteorology*, 269, pp.88-101, 2019.

839  
840



841  
 842 Figure 1. U.S. climate regions: NW (North West), WNC (West North Central), ENC (East North Central), NE (North East),  
 843 C (Central), SE (South East), C (Central), S (South), SW (South West), W (West). The circles represent the sampled USCRN  
 844 stations in the experiment.



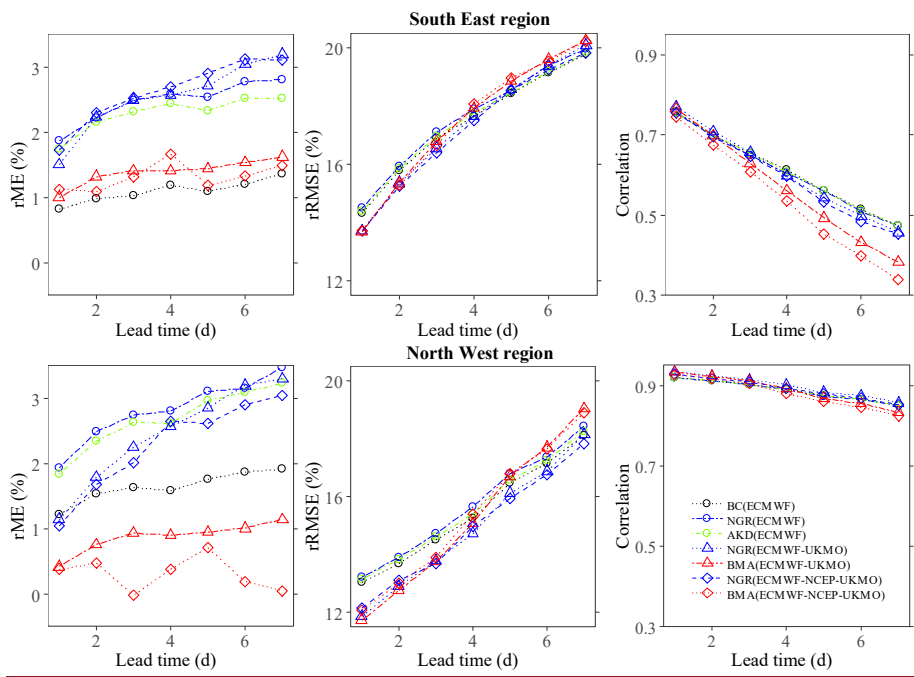


Figure 2. Relative mean error (rME), relative root mean square error (rRMSE), and correlation considering daily forecasts for different lead times over the SE and NW regions.

Formatted: English (United States)  
 Formatted: English (United States)



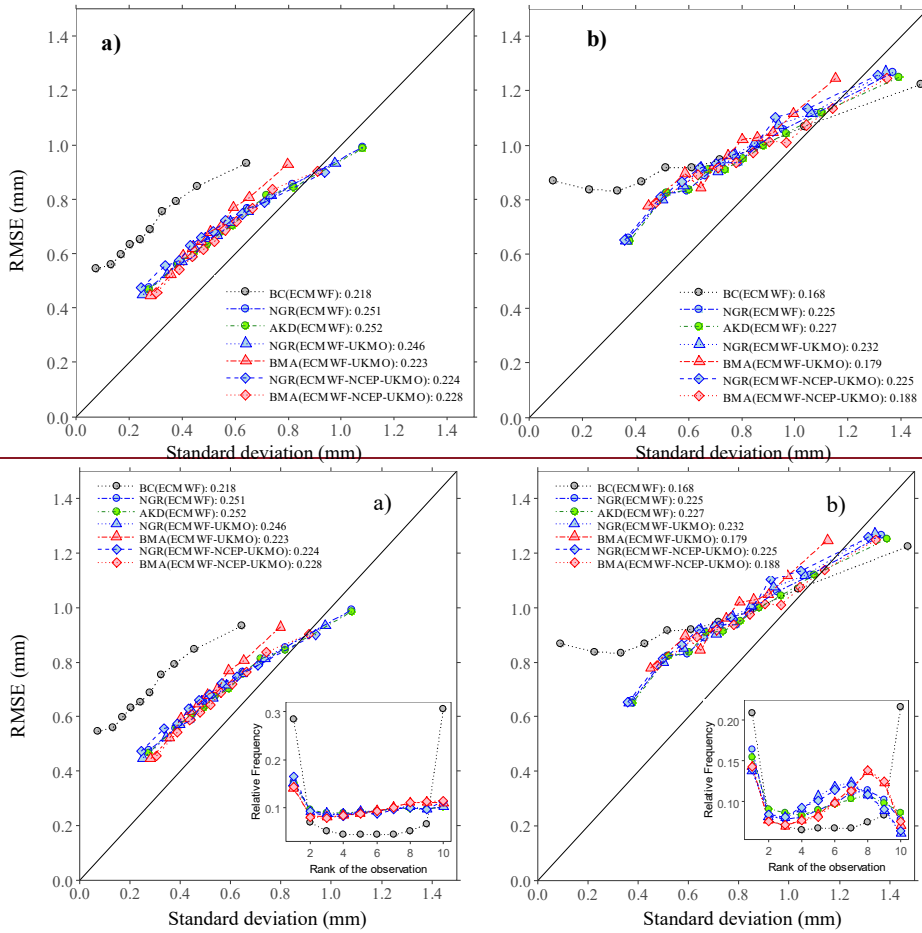


Figure 3. Binned spread-skill plots accounting for the mean of the ensemble standard deviation deciles against the mean RMSE of the forecasts in each decile over the verification period using based on all pairs of forecasts and observations at a) 1-day and b) 7-day lead. The panel in the right and the bottom shows the corresponding rank histograms. The correlation between the standard deviations and the absolute errors is reported after the colon. The solid line represents the 1:1 relationship.

- Formatted: Justified
- Formatted: English (United States)
- Formatted: English (United States)
- Formatted: English (United States)

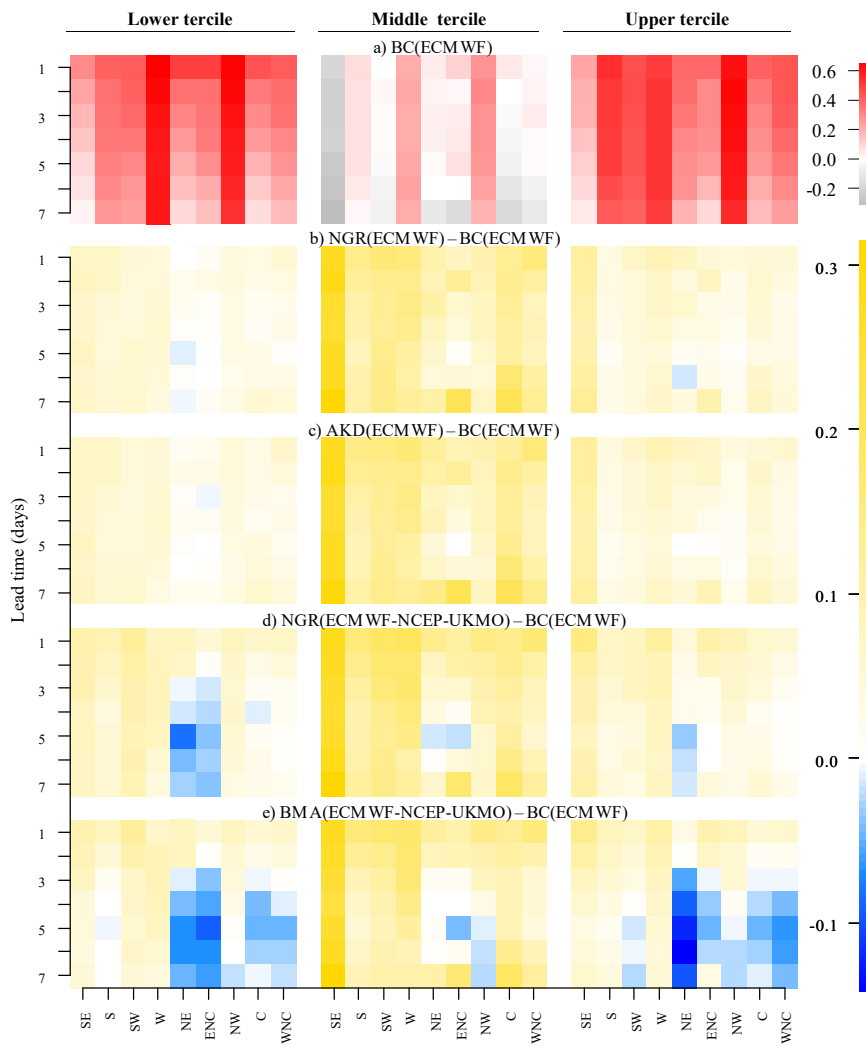


Figure 4. a) BSS for every region and lead time of the daily ECMWF forecasts post-processed using simple bias correction (used as reference BSS values) and b-e) differences between the BSS of the daily ECMWF forecasts post-processed with the b) NGR and c) AKD methods and the daily ECMWF-NCEP-UKMO forecasts post-processed with the d) NGR and e) BMA methods and the reference BSS.

- Formatted: English (United States)
- Formatted: English (United States)
- Formatted: English (United States)
- Formatted: English (United States)
- Formatted: English (United States)

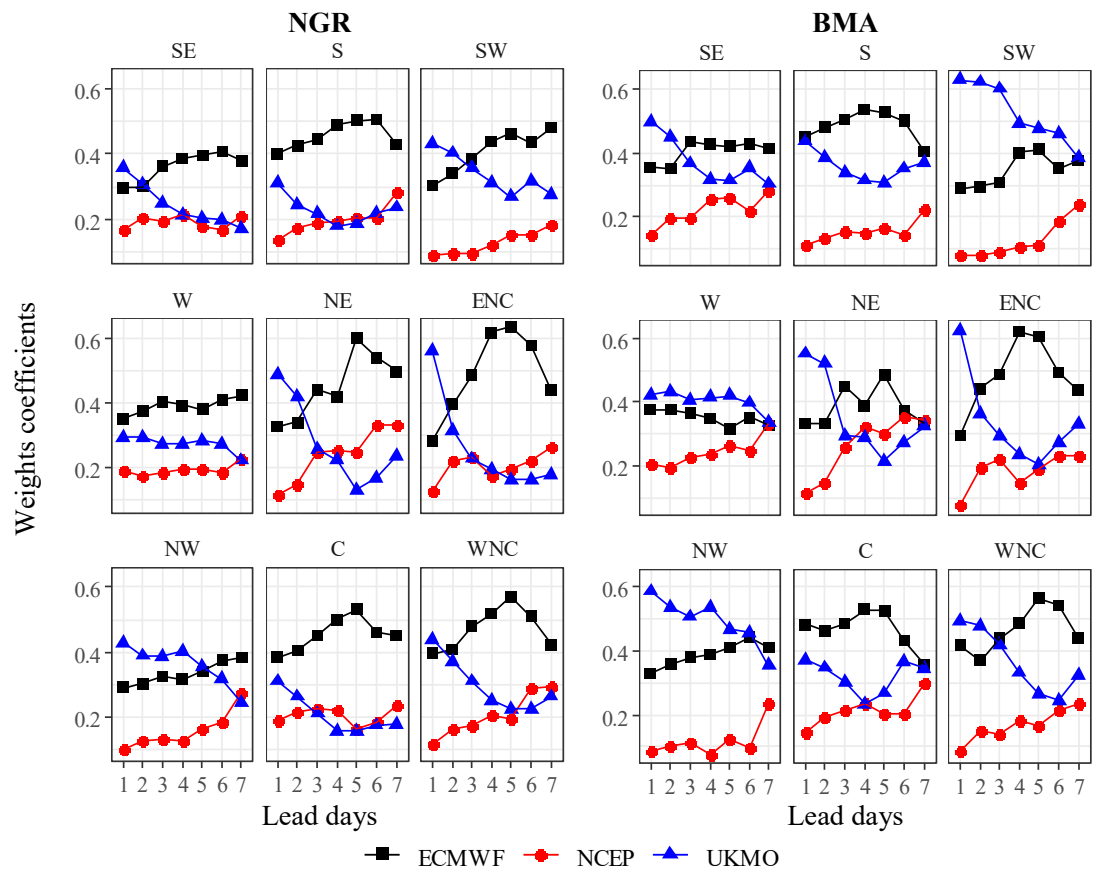


Figure 5. Regional mean weight coefficient  $b$  of the NGR technique (left panel) and the weight coefficient  $w$  of the BMA technique (right panel) for the post-processed daily ECMWF-NCEP-UKMO forecasts at different lead days.

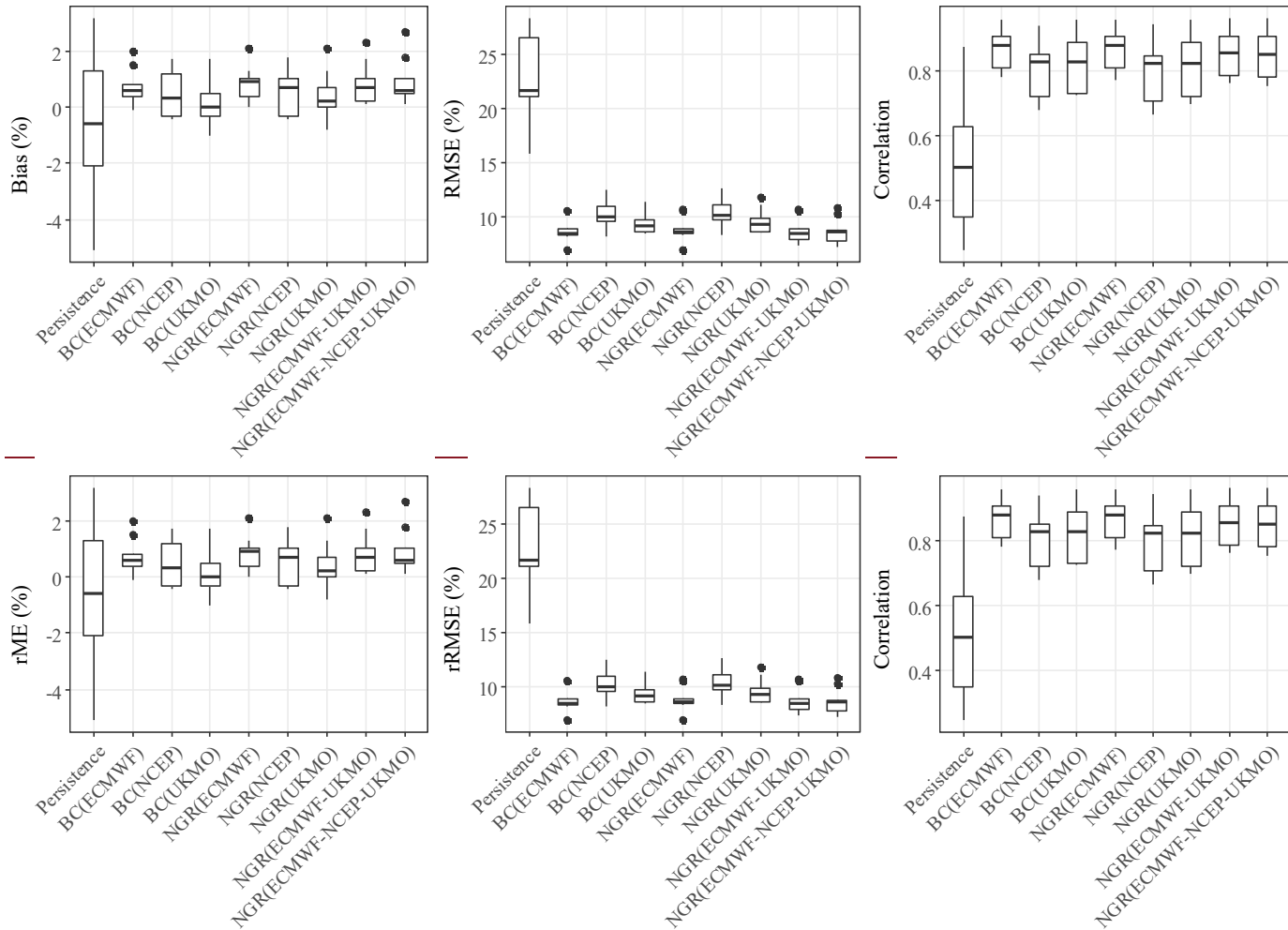


Figure 6. Whisker plot with the 2.5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 97.5<sup>th</sup> percentile of the distribution of the Relative bias, ME, relative-RMSE and correlation of weekly forecasts across different regions.

- Formatted: English (United States)
- Formatted: English (United States), Superscript
- Formatted: English (United States)
- Formatted: English (United States), Superscript
- Formatted: English (United States)
- Formatted: English (United States), Superscript
- Formatted: English (United States)
- Formatted: English (United States), Superscript
- Formatted: English (United States)
- Formatted: English (United States), Superscript
- Formatted: English (United States)
- Formatted: English (United States)

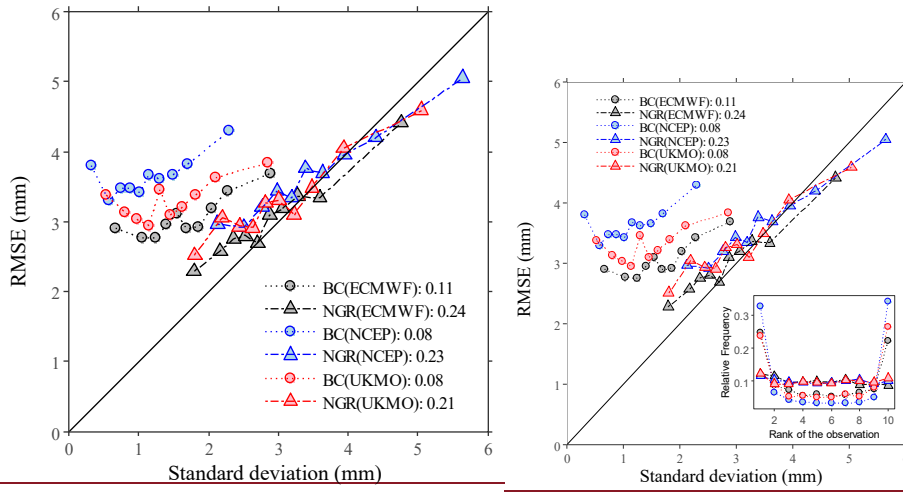


Figure 7. Binned spread-skill plots for the weekly forecasts accounting for the mean of the ensemble standard deviation deciles against the mean RMSE of the forecasts in each decile over the verification period. Binned spread-skill plot for the weekly forecasts using all pairs of forecasts and observations. The panel in the right and the bottom shows the corresponding rank histograms. The correlation between the standard deviations and the absolute errors is reported in the legend after the colon. The solid line represents the 1:1 relationship.

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

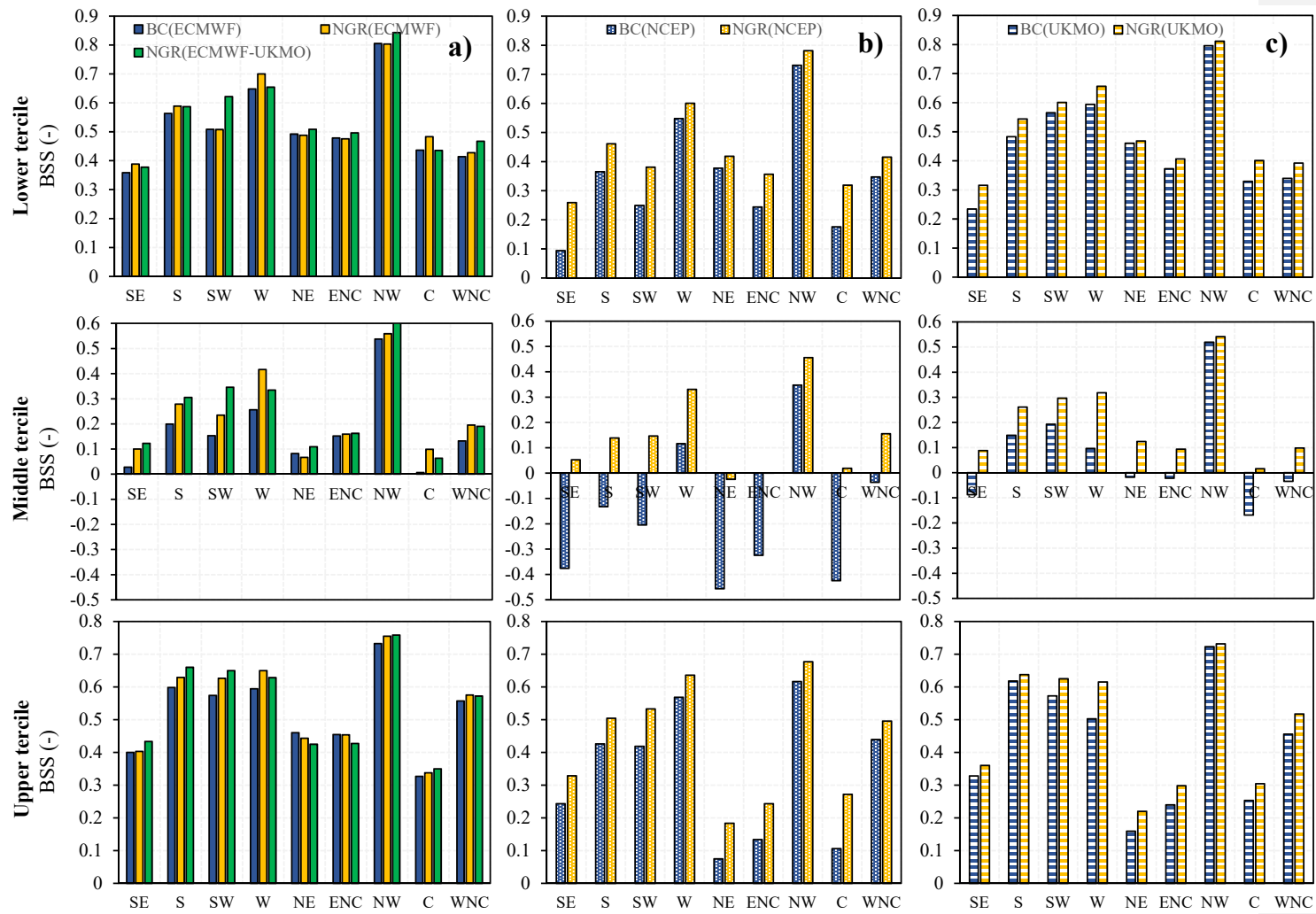


Figure 8. Comparison between BC and NGR based Brier Skill Scores considering a) ECMWF and ECMWF-UKMO forecasts, b) NCEP, and c) UKMO forecasts across the different climate regions.

Table 1. Evaluated schemes for daily and weekly ETo ensemble forecasts, with different post-processing methods: BC (simple bias correction), NGR (nonhomogeneous Gaussian regression), AKD (affine kernel dressing), and BMA (Bayesian model averaging), and different model and ensemble schemes: ECMWF (European Centre for Medium-Range Weather Forecasts model), NCEP (National Centers for Environmental Prediction model), and UKMO (United Kingdom Meteorological office model) ensemble forecasts, as well as ECMWF-UKMO (ensembles of ECMWF and UKMO) and ECMWF-NCEP-UKMO (ensembles of ECMWF, NCEP, and UKMO) ensemble forecasts.

Persistence	BC			NGR			AKD		BMA		
	ECMWF	NCEP	UKMO	ECMWF	NCEP	UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF	ECMWF-UKMO	ECMWF-NCEP-UKMO
Daily	✓			✓			✓	✓	✓	✓	✓
Weekly	✓	✓	✓	✓	✓	✓	✓	✓			

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted Table

Table 2. Minimum, mean and maximum coverage ratios over all the climate regions and lead times for different methods. See the caption of Table 1 for explanations of the methods acronyms.

	BC	NGR	AKD	NGR	BMA	NGR	BMA
-	ECMWF	ECMWF	ECMWF	ECMWF-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-NCEP-UKMO
Minimum coverage ratio	49.69	94.27	94.69	93.23	94.38	92.60	91.35
Mean coverage ratio	76.67	95.73	96.25	94.90	96.98	94.38	96.88
Maximum coverage ratio	93.13	98.02	98.33	97.29	99.38	96.56	99.58



Table 23. Spatial weighted average values of daily forecast metrics over all climate regions for different methods at lead days 1 and 7. See the caption of Table 1 for explanations of the methods acronyms. Numbers in bold indicate the best performance for each lead day.

	BC		NGR		AKF		NGR		BMA		NGR		BMA	
	ECMWF		ECMWF		ECMWF		ECMWF-UKMO		ECMWF-UKMO		ECMWF-NCEP-UKMO		ECMWF-NCEP-UKMO	
	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days
rBiasrME (%)	0.822	1.203	1.695	2.682	1.626	2.419	1.327	2.735	0.632	0.939	1.394	2.778	<b>0.490</b>	<b>0.626</b>
rRMSE (%)	14.38	<b>19.64</b>	14.59	19.88	14.47	19.76	13.68	19.67	13.65	20.15	<b>13.59</b>	19.67	13.67	20.28
Bias- <u>ME</u> (mm day <sup>-1</sup> )	0.038	0.057	0.080	0.128	0.077	0.115	0.063	0.131	0.029	0.046	0.067	0.134	0.005	0.006
RMSE (mm day <sup>-1</sup> )	0.708	0.950	0.718	0.961	0.716	0.958	0.682	0.965	0.681	0.990	0.681	0.971	0.685	1.002
Correlation	0.832	<b>0.652</b>	0.829	0.649	0.830	0.649	<b>0.843</b>	0.639	0.841	0.586	0.841	0.635	0.832	0.560
Coverage ratio	64.54	79.40	95.63	95.44	95.93	96.10	94.24	94.73	<b>96.51</b>	96.56	93.52	94.57	96.47	<b>97.24</b>
<u>CRPS</u> (mm)	<u>0.432</u>	<u>0.555</u>	<u>0.395</u>	<u>0.526</u>	<u>0.394</u>	<u>0.525</u>	<u>0.374</u>	<u>0.529</u>	<u>0.374</u>	<u>0.547</u>	<u>0.375</u>	<u>0.534</u>	<u>0.377</u>	<u>0.557</u>
BSS_1st	0.442	0.232	0.492	0.279	0.492	<b>0.282</b>	<b>0.525</b>	0.274	0.519	0.240	0.521	0.271	0.513	0.225
BSS_2nd	0.042	-0.062	0.201	0.101	0.202	<b>0.101</b>	<b>0.224</b>	0.095	0.214	0.074	0.217	0.089	0.200	0.059
BSS_3nd	0.433	0.300	0.496	<b>0.359</b>	0.499	0.358	<b>0.519</b>	0.350	0.515	0.305	0.512	0.338	0.494	0.277

Formatted Table

- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt
- Formatted: Font: 9 pt

Table 4. Percentage differences (averaged over all lead times) of the ECMWF-UKMO and ECMWF-NCEP-UKMO forecast performance with the ECMWF forecast performance, after post-processing with the non-homogeneous Gaussian regression (NGR) method. See the caption of Table 1 for explanations of the forecast models acronyms.

	Western climate regions						Northern climate regions					
	SW		W		NW		NE		ENC		WNC	
	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO
Bias	-26.753	-30.826	-9.111	9.421	-13.908	-18.800	-4.268	25.053	-2.149	-1.445	-10.119	0.761
RMSE	-4.682	-4.013	-3.455	-2.505	-3.973	-2.839	1.898	4.333	1.455	2.000	-1.313	-0.922
Correlation	1.760	0.627	0.947	0.707	1.197	0.607	-4.180	-4.600	-3.275	-3.137	-2.312	-2.062
Cov. ratio	-1.386	-2.094	-0.977	-1.194	-1.019	-1.144	-0.835	-1.656	-0.850	-0.986	-0.835	-1.402
BSS_1st	12.022	7.481	3.222	2.846	3.548	4.236	-11.999	-9.676	-9.643	-9.384	-3.680	-5.181
BSS_2nd	8.991	-6.504	5.792	9.044	4.984	3.961	-112.954	-93.092	-19.092	-13.642	-15.725	-27.949
BSS_3rd	2.295	-1.807	3.575	6.557	4.196	2.370	-9.105	-8.992	-6.420	-10.605	-4.595	-5.835

Table 5. Percentage differences (averaged over regions) of forecast performance of using 45 days training period with using 30 days training period for lead days 1 and 7. See the caption of Table 1 for explanations of the methods acronyms.

	NGR(ECMWF)		AKD(ECMWF)		NGR(ECMWF-UKMO)		NGR(ECMWF-NCEP-UKMO)	
	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days
Bias	16.569	18.732	21.654	22.859	4.714	10.089	-0.496	7.070
RMSE	-0.701	-2.641	-1.007	-3.121	-0.404	-3.720	-0.045	-4.742
Correlation	-0.157	0.525	-0.141	0.605	-0.099	1.332	-0.467	0.741
Cov. Ratio	1.276	0.954	1.615	1.257	1.701	1.495	1.938	1.338
BSS_1st	-0.884	2.183	-1.164	2.761	-0.212	5.062	-2.600	6.277
BSS_2nd	-1.259	2.764	-1.283	5.680	3.614	8.959	-2.293	5.562
BSS_3rd	-0.382	-1.589	-0.904	-0.212	-1.340	2.632	-1.625	0.240



**ANNEX**

Table A1. Percentage differences (averaged over all lead times) of the ECMWF-UKMO and ECMWF-NCEP-UKMO forecast performance with the ECMWF forecast performance, after post-processing with the non-homogeneous Gaussian regression (NGR) method. See the caption of Table 1 for explanations of the forecast models acronyms.

	Western climate regions						Northern climate regions					
	SW		W		NW		NE		ENC		WNC	
	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO	ECMWF-UKMO	ECMWF-NCEP-UKMO
ME	-26.75	-30.83	-9.11	9.42	-13.91	-18.80	-4.27	25.05	-2.15	-1.45	-10.12	0.75
RMSE	-4.68	-4.01	-3.46	-2.51	-3.97	-2.84	1.90	4.33	1.46	2.00	-1.31	-0.92
Correlation	1.76	0.63	0.95	0.71	1.20	0.61	-4.18	-4.60	-3.28	-3.14	-2.31	-2.06
Cov. ratio	-1.39	-2.09	-0.98	-1.19	-1.02	-1.14	-0.84	-1.66	-0.85	-0.99	-0.84	-1.40
CRPS	-4.84	-3.89	-3.42	-1.99	-3.90	-2.81	1.41	4.02	1.58	2.45	-1.00	-0.82
BSS_1st	12.02	7.48	3.22	2.85	3.55	4.24	-12.00	-9.68	-9.64	-9.38	-3.68	-5.12
BSS_2nd	8.99	-6.50	5.79	9.04	4.98	3.96	-112.95	-93.09	-19.09	-13.64	-15.73	-27.12
BSS_3rd	2.30	-1.81	3.58	6.56	4.20	2.37	-9.11	-8.99	-6.42	-10.61	-4.60	-5.12

- Formatted ... [9]
- Formatted ... [10]
- Formatted ... [11]
- Formatted ... [12]
- Formatted ... [13]
- Formatted ... [14]
- Formatted ... [15]
- Formatted ... [16]
- Formatted ... [17]
- Formatted ... [18]
- Formatted ... [19]
- Formatted ... [20]
- Formatted ... [21]
- Formatted ... [22]
- Formatted ... [23]
- Formatted ... [24]
- Formatted ... [25]
- Formatted ... [26]
- Formatted ... [27]
- Formatted ... [28]
- Formatted ... [29]
- Formatted ... [30]
- Formatted ... [31]
- Formatted ... [32]
- Formatted ... [33]
- Formatted ... [34]
- Formatted ... [35]
- Formatted ... [36]
- Formatted ... [37]
- Formatted ... [38]
- Formatted ... [39]
- Formatted ... [40]
- Formatted ... [41]
- Formatted ... [42]
- Formatted ... [43]
- Formatted ... [44]
- Formatted ... [45]

Table A2. Percentage differences (averaged over regions) of forecast performance of using 45 days training period with using 30 days training period for lead days 1 and 7. See the caption of Table 1 for explanations of the methods acronyms.

	NGR(ECMWF)		AKD(ECMWF)		NGR(ECMWF-UKMO)		NGR(ECMWF-NCEP-UKMO)	
	1 day	7 days	1 day	7 days	1 day	7 days	1 day	7 days
ME	16.57	18.73	21.65	22.86	4.71	10.09	-0.50	7.07
RMSE	-0.70	-2.64	-1.01	-3.12	-0.40	-3.72	-0.05	-4.74
Correlation	-0.16	0.53	-0.14	0.61	-0.10	1.33	-0.47	0.74
Cov. Ratio	1.28	0.95	1.62	1.26	1.70	1.50	1.94	1.34
CRPS (mm)	-0.77	-3.00	-1.22	-3.51	-0.92	-3.89	-0.01	-4.53
BSS_1st	-0.88	2.18	-1.16	2.76	-0.21	5.06	-2.60	6.28
BSS_2nd	-1.26	2.76	-1.28	5.68	-3.61	8.96	-2.29	5.56
BSS_3rd	-0.38	-1.59	-0.90	-0.21	-1.34	2.63	-1.63	0.24

Formatted	... [119]
Formatted	... [120]
Formatted	... [121]
Formatted	... [122]
Formatted	... [123]
Formatted	... [124]
Formatted	... [125]
Formatted	... [126]
Formatted Table	... [118]
Formatted	... [127]
Formatted	... [128]
Formatted	... [129]
Formatted	... [130]
Formatted	... [131]
Formatted	... [132]
Formatted	... [133]
Formatted	... [134]
Formatted	... [135]
Formatted	... [136]
Formatted	... [137]
Formatted	... [138]
Formatted	... [139]
Formatted	... [140]
Formatted	... [141]
Formatted	... [142]
Formatted	... [143]
Formatted	... [144]
Formatted	... [145]
Formatted	... [146]
Formatted	... [147]
Formatted	... [148]
Formatted	... [149]
Formatted	... [150]
Formatted	... [151]
Formatted	... [152]
Formatted	... [153]
Formatted	... [154]
Formatted	... [155]
Formatted	... [156]
Formatted	... [157]
Formatted	... [158]
Formatted	... [159]
Formatted	... [160]
Formatted	... [161]
Formatted	... [162]
Formatted	... [163]
Formatted	... [164]
Formatted	... [165]
Formatted	... [166]
Formatted	... [167]

<b>Page 9: [1] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:08:00 AM</b>
------------------------------	---------------------	------------------------------

Font: 10 pt

<b>Page 9: [1] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:08:00 AM</b>
------------------------------	---------------------	------------------------------

Font: 10 pt

<b>Page 9: [2] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:08:00 AM</b>
------------------------------	---------------------	------------------------------

Font: 10 pt

<b>Page 9: [2] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:08:00 AM</b>
------------------------------	---------------------	------------------------------

Font: 10 pt

<b>Page 9: [2] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:08:00 AM</b>
------------------------------	---------------------	------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [3] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [4] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [4] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt

<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
------------------------------	---------------------	-------------------------------

Font: 10 pt



<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
Font: 10 pt		
<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
Font: 10 pt		
<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
Font: 10 pt		
<b>Page 9: [5] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:14:00 PM</b>
Font: 10 pt		
<b>Page 9: [6] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
Font: 10 pt		
<b>Page 9: [6] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
Font: 10 pt		
<b>Page 9: [6] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
Font: 10 pt		
<b>Page 9: [7] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
Font: 10 pt		
<b>Page 9: [7] Formatted</b>	<b>Hanoi Medina</b>	<b>11/16/2019 10:07:00 PM</b>
Font: 10 pt		
<b>Page 9: [8] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 11:07:00 PM</b>
Highlight		
<b>Page 9: [8] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 11:07:00 PM</b>
Highlight		
<b>Page 9: [8] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 11:07:00 PM</b>
Highlight		
<b>Page 9: [8] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 11:07:00 PM</b>
Highlight		
<b>Page 42: [9] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:45:00 AM</b>
Font: Bold		

<b>Page 42: [9] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:45:00 AM</b>
Font: Bold		
<b>Page 42: [10] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [10] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [11] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [11] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [12] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [12] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [13] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [13] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [14] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [14] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [15] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [15] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [16] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [16] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [17] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [17] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [18] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [18] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [19] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [19] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [20] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [20] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [21] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [21] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [22] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [22] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [23] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [23] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [24] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [24] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [25] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [25] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [26] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [26] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [27] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [27] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [28] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [28] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [29] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [29] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [30] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [30] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [31] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [31] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [32] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [32] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [33] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [33] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [34] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [34] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [35] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [35] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [36] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [36] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [37] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [37] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [38] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [38] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [39] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [39] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [40] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [40] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [41] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [41] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [42] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [42] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [43] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [43] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [44] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

**Page 42: [44] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [45] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [45] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [46] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [46] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [47] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [47] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [48] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [48] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [49] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [49] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [50] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [50] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [51] Formatted Table**                      **Hanoi Medina**                      **11/18/2019 12:59:00 PM**

Formatted Table

<b>Page 42: [52] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [52] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [53] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [53] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [54] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [54] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [55] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [55] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [56] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [56] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [57] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [57] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [58] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [58] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		



**Page 42: [59] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [59] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [60] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [60] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [61] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [61] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [62] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [62] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [63] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [63] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [64] Formatted**                      **Hanoi Medina**                      **11/18/2019 12:58:00 PM**

Font: 9 pt

**Page 42: [65] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [65] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [66] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [66] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [67] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [67] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [68] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [68] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [69] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [69] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [70] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [70] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [71] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [71] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [72] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [72] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

**Page 42: [73] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman

<b>Page 42: [73] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [74] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [74] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [75] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [75] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [76] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [76] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 42: [77] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [77] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [78] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [78] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [79] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [79] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [80] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [80] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [81] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [81] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [82] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [82] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [83] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [83] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [84] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [84] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [85] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [85] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [86] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [86] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [87] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [87] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [88] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [88] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [89] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [89] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [90] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [90] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [91] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [91] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [92] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [92] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [93] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [93] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [94] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [94] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [95] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [95] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [96] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [96] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [97] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [97] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [98] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [98] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [99] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [99] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [100] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [100] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [101] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [101] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [102] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [102] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [103] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [103] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [104] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [104] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [105] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [105] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [106] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [106] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [107] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [107] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [108] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		

<b>Page 42: [108] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [109] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [109] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [110] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [110] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [111] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [111] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [112] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [112] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [113] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [113] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [114] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [114] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		
<b>Page 42: [115] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:06:00 PM</b>
Font: (Default) Times New Roman, 9 pt		



**Page 42: [115] Formatted** **Hanoi Medina** **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [116] Formatted** **Hanoi Medina** **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [116] Formatted** **Hanoi Medina** **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [117] Formatted** **Hanoi Medina** **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 42: [117] Formatted** **Hanoi Medina** **11/18/2019 6:06:00 PM**

Font: (Default) Times New Roman, 9 pt

**Page 43: [118] Formatted Table** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Formatted Table

**Page 43: [119] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [120] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [121] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [122] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [123] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [124] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [125] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [126] Formatted** **Hanoi Medina** **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [127] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [128] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [129] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [130] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [131] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [132] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [133] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [134] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [135] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [136] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [137] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [138] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [139] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [140] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

<b>Page 43: [141] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [142] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [143] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [144] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [145] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [146] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [147] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [148] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [149] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [150] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [151] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [152] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [153] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [154] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		

<b>Page 43: [155] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [156] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [157] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [158] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [159] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [160] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [161] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [162] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [163] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [164] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [165] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [166] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [167] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		
<b>Page 43: [168] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
Font: (Default) Times New Roman		

**Page 43: [169] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [170] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [171] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [172] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [173] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [174] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [175] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [176] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [177] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [178] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [179] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [180] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [181] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

**Page 43: [182] Formatted**                      **Hanoi Medina**                      **11/18/2019 6:05:00 PM**

Font: (Default) Times New Roman

<b>Page 43: [183] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
---------------------------------	---------------------	------------------------------

Font: (Default) Times New Roman

<b>Page 43: [184] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
---------------------------------	---------------------	------------------------------

Font: (Default) Times New Roman

<b>Page 43: [185] Formatted</b>	<b>Hanoi Medina</b>	<b>11/18/2019 6:05:00 PM</b>
---------------------------------	---------------------	------------------------------

Font: (Default) Times New Roman

<b>Page 43: [186] Formatted</b>	<b>Hanoi Medina</b>	<b>11/17/2019 9:32:00 AM</b>
---------------------------------	---------------------	------------------------------

Indent: Left: 0"