Power transformation of variables for post-processing precipitation forecasts: regionally versus locally optimized parameter values

Yiliang Du*, Quan J. Wang, Wenyan Wu, Qichun Yang

Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Australia

*Corresponding author: yiliangd@student.unimelb.edu.au

This study is funded by an Australian Research Council Linkage Project (LP170100922).

The authors declare that they have no competing interests or personal relationships that could have appeared to influence the work reported in this paper.

Abstract

Short-term precipitation forecasts are mainly derived from numerical weather prediction (NWP) models. Raw NWP forecasts typically require post-processing to improve their accuracy and reliability through statistical calibration. For post-processing precipitation forecasts, several well-known calibration models employ power transformation to remove the positive skewness of precipitation. The most common practice is to use a pre-fixed transformation parameter value for both observation and forecast variables. Another approach is to allow the parameter values to differ for the two variables and locally optimize them at a spatial point to ensure the best performance of a calibration model. However, when calibrating forecasts across many grid points in a region, there is a considerable advantage in keeping the transformation parameter values constant, to enable spatial consistency in the calibration model so that model information transfers from gauged to ungauged locations can be easily achieved. For this reason, a third approach is proposed here to regionally optimize the parameter values for the two variables. A question then arises: how do the three different approaches affect forecast calibration performance? This question is answered in this study by evaluating the calibration performance of precipitation forecasts across 20 locations in Australia. The pre-fixed one parameter value approach is found to lead to poor performance. The regionally optimized two parameter value approach is almost as good as the locally optimized two parameter value approach, but has the additional advantage of spatial consistency for regional applications.

Keywords: data transformation, numerical weather prediction, post-processing, precipitation, forecast verification, Australia

1. Introduction

Operational short-term precipitation forecasts (e.g., up to ten days ahead) are mainly derived from numerical weather prediction (NWP) models in climate and weather centres worldwide. Due to the inherent imperfections of NWP models, model outputs tend to exhibit biases and, in the case of ensemble forecasts, fail to produce reliable spread. To make precipitation forecasts more informative, post-processing using statistical calibration models is required to improve the accuracy and reliability of raw NWP forecasts (Duan et al., 2019; Li et al., 2017; Vannitsem et al., 2018; Wilks, 2011). The development of these calibration models generally requires first to identify the relationship between the archived forecasts and the corresponding observed data, and then to utilize the relationship to adjust new NWP forecasts.

Precipitation data is often highly skewed and displays nonconstant variance, posing a challenge to statistical calibration. Various calibration models have been proposed by assuming the form of a probability distribution. For example, extended logistic regression implies the assumption of a censored logistic distribution (Wilks, 2009); the use of Bayesian model averaging is extended for precipitation by specifying a mixture of a discrete component at zero and a gamma distribution (Sloughter et al., 2007); nonhomogeneous Gaussian regression (also known as ensemble-MOS) assumes a censored generalized extreme value distribution (Scheuerer, 2014) or a censored, shifted gamma distribution (Scheuerer and Hamill, 2015).In addition, another type of calibration procedures is based on Gaussian models and therefore require data transformation. Examples are the Bayesian processor of forecast (Krzysztofowicz and Evans, 2008), the Ensemble Pre-Processor (Schaake et al., 2007; Wu et al., 2011), the Bayesian joint probability model (Robertson et al., 2013; Wang et al., 2009), and the Seasonally Coherent Calibration model (Wang et al., 2019a; Yang et al., 2021; Zhao et al., 2022, 2020), using various transformations such as the normal quantile transformation, the log-sinh transformation, and the power transformation. The effects of different transformations on calibration performance have been demonstrated to vary with the region of interest (Li et al., 2019).

In this study, we focus on power transformation for post-processing precipitation forecasts. There are two broad approaches to determining the transformation parameter values for observation and forecast variables. The first approach, referred to as the pre-fixed one parameter value approach in this study, is to pre-specify a fixed value for the power transformation parameter, and apply it to both observation and forecast variables (Duan et al., 2007; Hemri et al., 2015). In previous studies for the calibration of precipitation forecasts, the transformation parameter is typically set to 1/2 (Messner et al., 2014; Wilks, 2009), 1/3 (Berrocal et al., 2008; Kleiber et al., 2011; Sloughter et al., 2007), 1/4 (Ben Bouallègue, 2013; Hamill et al., 2004) or other values (Stauffer et al., 2017). There are also some studies that investigated the appropriate fixed transformation parameter value for precipitation in an area of interest through a series of experiments with different candidates (Cecinati et al., 2017; Erdin et al., 2012; Fu et al., 2010). For example, Fu et al. (2010) found that power transformation with 5/12 led to the “best” fitted truncated normal distribution for daily precipitation at 47% of stations in Australia.

The second approach, referred to as the locally optimized two parameter value approach in this study, is to allow the parameter values to differ for observation and forecast variables, and the specific values are optimally estimated at each location (Li et al., 2019; Schepen and Wang, 2013). The locally optimized two parameter value approach should theoretically lead to better calibration results than the pre-fixed one parameter value approach. However, the pre-fixed one parameter value approach has a significant advantage over the locally optimized two parameter value approach when calibrating forecasts across many grid points in a region. When the transformation parameter value is kept constant across space, the calibration models are directly comparable spatially in the transformed variable space. These calibration models at different gauged locations can be connected using for example geostatistics techniques (Cressie, 1993), facilitating model information transfer from gauged to ungauged locations (Berrocal et al., 2008; Kleiber et al., 2011; Stauffer et al., 2017).

Considering the advantage and disadvantage of the above two approaches, we propose a third approach: a regionally optimized two parameter value approach. Under this approach, transformation parameter values are allowed to have different values for the observation variable and the forecast variable, but with each kept constant across a region. The regional optimum parameter values are found by considering overall fitting across a region. This third approach is more flexible than the pre-fixed one parameter value approach and better at bringing about spatial consistency in the calibration model than the locally optimized two parameter value approach. However, this approach offers less flexibility than the locally optimized two parameter value approach and therefore may lead to worse forecast calibration performance.

In this study, we present the three approaches and evaluate their performances in the context of statistical calibration of NWP model forecasts of daily precipitation. Numerical experiments are conducted for 20 locations across different climate zones in Australia. Insights are drawn from the results of the experiments, and implications are discussed.

2. Methods

In this section, we first introduce the three approaches to data normalization using power transformation. We then briefly introduce a statistical calibration model, the seasonally coherent calibration (SCC) model, which will be used to evaluate the impacts of the transformation approaches on forecast calibration. A workflow of this study is given in Fig. 1. Three transformation approaches are separately used to normalize both the observation variable and forecast variable, and then the SCC model is fitted in the transformed space. Once the SCC model is well-established, calibrated forecasts in original space are obtained by converting a conditional distribution derived from the fitted SCC model with the inverse of transformation. The calibrated forecasts are systematically evaluated and compared between transformation methods. Details on evaluation metrics used are also described in this section.

2.1. Three approaches to power transformation for data normalization

The power transformation is defined as follows:

$$\hat{y} = y^{\lambda} \tag{1}$$

where $\hat{y}$ is the transformed data and λ is the power parameter. This power transformation may also be expressed in the form of the widely used Box-Cox transformation. For the purpose of this study, we have adopted the simple form (Eq. 1) and should lead to the same findings as using the Box-Cox transformation. Three approaches to power transformation are investigated in this study:

(i) The pre-fixed one parameter value approach pre-specifies the same transformation parameter value for observation and forecast variables. The most commonly used parameter values identified from literature review (Section 1) are included in this study. These are 1/2, 5/12, 1/3 and 1/4.

(ii) The locally optimized two parameter value approach uses different parameter values for different variables and locations. The case-dependent parameter value is estimated together with the parameters of the underlying normal distribution using the maximum likelihood method. The likelihood function for variable *v* and location *s* can be written as:

$$L(\lambda_{v,s}, \mu_{v,s}, \sigma_{v,s}) = \prod_{t=1}^{T} l_{v,s}(t) \tag{2}$$

where $\mu_{v,s}$, $\sigma_{v,s}$ are the mean and standard deviation of the assumed normal distribution of the transformed variable $\hat{y}_{v,s}(t), t = 1, 2, \dots, T$. The calculation of likelihood $l_{v,s}(t)$

will be presented later considering that the intermittency and continuity of precipitation process need to be treated specifically.

(iii) The regionally optimized two parameter value approach specifies two different parameter values for observation and forecast variables, respectively. For each variable, the optimal value is estimated using the data of all locations together to maximize the likelihood function, which can be written as:

$$L(\lambda_v, \mu_{v,s}, \sigma_{v,s}, s = 1, \dots, S) = \prod_{s=1}^{S} \prod_{t=1}^{T} l_{v,s}(t) \quad (3)$$

The normal distribution should be continuous in its variable value, but data here is daily precipitation with a physical limit at zero. To overcome this issue, we follow the commonly used technique, which treats records below a threshold as left-censored data (Wang et al., 2020; Wang and Robertson, 2011). Given a precipitation amount *y* at day *t*, the likelihood $l_{v,s}(t)$ can be written as (subscripts *v* and *s* omitted here for simplicity and clarity):

$$l(t) = \begin{cases} \mathrm{pdf_N}\left([y(t)]^{\lambda};\ \mu, \sigma^2\right) \times \lambda y(t)^{\lambda-1} & \text{if} \quad y(t) > y_c \\ \mathrm{cdf_N}\left({y_c}^{\lambda};\ \mu, \sigma^2\right) & \text{if} \quad y(t) \leq y_c \end{cases} \quad (4)$$

where y_c is the censoring threshold; $\mathrm{pdf_N}$ is the probability density function and $\mathrm{cdf_N}$ the cumulative distribution function of a normal distribution; the term $\lambda y(t)^{\lambda-1}$ is the Jacobian of the power transformation. Specifically, we use the censoring thresholds of 0.2 mm/day for observations and 0.01 mm/day for forecasts following the recommendation by Robertson et al. (2013). This censoring treatment is applied to data transformation here, as well as to calibration model fitting and forecast evaluation later.

Normality of the transformed distribution is evaluated using a modified version of Anderson-Darling (A-D) test for left-censored data (Thode, 2002). The A-D test statistic measures deviation of the sample data from an assumed normal distribution. A smaller value of the test statistic indicates the assumption is more acceptable. Details for the method are given in Appendix A

2.2. The seasonally coherent calibration (SCC) model

Once the observation variable and forecast variable are transformed using the above approaches, we apply the seasonally coherent calibration (SCC) model to calibrate deterministic NWP forecasts of precipitation at each location and lead time. The SCC model was developed for working effectively with limited archived NWP data to produce unbiased, skillful and reliable calibrated ensemble forecasts that are consistent in seasonally-varying climatology with long-term observations (Wang et al., 2019a; Yang et al., 2021; Zhao et al., 2020). By using training data (archived NWP data and historical observations of a longer period), the SCC model forms a collection of bivariate normal distributions, representing the seasonally-varying relationships between raw NWP forecasts and observations. Once the model is established, it can take a new raw forecast after transformation as input and draw a large number of random samples (say 1000) from the conditional bivariate normal distribution of a specific month. Finally, the random samples are back-transformed to the original space using the inverse of Eq. 1 to obtain an ensemble of calibrated forecasts. For details, see Appendix B and Wang et al. (2019a).

In the original SCC model, the log-sinh transformation is used to normalize data (Wang et al., 2012). While the log-sinh transformation has been found effective for transforming precipitation data in a number of studies (Li et al., 2020; Schepen et al., 2018; Schepen and Wang, 2014; Wang et al., 2019b), we have focused on the power transformation in this study for two reasons: (i) Power transformation is widely used, and (ii) Power transformation has only one parameter, making it easier to regionalize than the log-sinh transformation with two parameters whose sample values may interact with each other. For completeness, however, we have included the evaluation of calibrated forecasts with the log-sinh transformation applied under the second approach of the locally optimized transformation parameter values. In fact, a systematic comparison of the power transformation with the log-sinh transformation can be interesting in its own right. Further information on the log-sinh transformation is given in appendix C.

2.3. Evaluation metrics

Leave-one-month-out cross-validation is applied to evaluate the calibration performance for each location and forecast lead time. The whole validation period is defined as the period when both observed and forecast data are available. For each round of cross-validation, the observed and forecast data outside the one-month validation period is used as the training data to infer the parameters of the SCC model. The raw NWP forecasts in the selected month are treated as the test data, and calibrated using the established SCC model. This procedure is repeated multiple rounds until the calibration of all raw NWP forecasts is completed. The calibrated ensemble forecasts generated from each round are pooled together to compare with corresponding observations to calculate the evaluation metrics, including bias, sharpness, reliability and forecast skills (overall skills and skills for heavy precipitation).

2.3.1 Bias

We first calculate bias in the ensemble mean of calibrated forecasts. Bias can be estimated as:

$$\text{Bias} = \frac{1}{T}\sum_{t=1}^{T}(\bar{y}_{cali}(t) - y_{obs}(t)) \quad (5)$$

where $\bar{y}_{cali}(t)$ is the mean of the calibrated ensemble at day *t*; $y_{obs}(t)$ is the corresponding observation; and *T* denotes the total number of days in the validation period. Considering bias cannot be measured accurately when referring to censored data, we set ensemble members below the censoring threshold of 0.2 mm/day as zero. For consistency, we also set observations below the same threshold as zero for this calculation.

2.3.2 Sharpness

We assess the forecast sharpness by measuring the average width of the ensemble forecast for the 95% interval, which can be written as:

$$\text{Sharpness} = \frac{1}{T}\sum_{t=1}^{T}\left(y_{cali,q_2}(t) - y_{cali,q_1}(t)\right) \quad (6)$$

where $y_{cali,q_1}(t)$ and $y_{cali,q_2}(t)$ denote the 2.5% and 97.5% quantile of ensemble forecast at day *t*, respectively. A sharper forecast should be more valuable for forecasters, as long as the forecast is unbiased and reliable in uncertainty spread.

2.3.3 Reliability

We assess the reliability of ensemble forecasts by calculating the probability integral transformations (PIT) (Gneiting et al., 2007). Reliability refers to the statistical consistency between the distribution of ensemble forecast and the observation frequency. Reliable ensemble forecasts are capable of correctly representing the probability of event occurrences. The PIT for the ensemble forecast and the corresponding observation at day *t* is defined as:

$$\pi(t) = F_t[y_{obs}(t)] \quad (7)$$

where $F_t(\cdot)$ is the cumulative distribution function of ensemble forecast at day *t*. To calculate the PIT for censored observations, we randomly generate pseudo PIT values from a uniform distribution with a range $[0, F_t(0.2\ \text{mm/day})]$, following the work of Wang and Robertson (2011). The real and pseudo PIT values are subsequently used to construct the PIT uniform probability plot. Ensemble forecasts of perfectly reliable uncertainty spread should yield PIT values that are uniformly distributed between 0 and 1. Here we use the α-index to measure the deviation of PIT values from the theoretical standard uniform distribution:

$$\alpha = 1 - \frac{2}{T}\sum_{t=1}^{T}\left|\pi^*(t) - \frac{t}{T+1}\right| \quad (8)$$

where $\pi^*(t)$ is the sorted $\pi(t)$ in increasing order. The α-index ranges from 0 (worst reliability) to 1 (perfect reliability).

2.3.4 Forecast skills

We assess the overall forecast errors using the continuous ranked probability score (CRPS) (Hersbach, 2000). The CRPS at day *t* is defined as:

$$\text{CRPS}(t) = \int_{-\infty}^{+\infty} \{F_t[y_{cali}(t)] - H_t[y_{cali}(t) - y_{obs}(t)]\}^2 dy_{cali} \tag{9}$$

where H is the Heaviside step function (H equals 0 if $y_{cali}(t) < y_{obs}(t)$ and equals 1 otherwise). The CRPS represents a summary difference between the ensemble forecast and the paired observation, by combining the Brier scores used for assessing two-category probabilistic forecasts over all possible category thresholds. A low CRPS tends to be found in an accurate ensemble forecast with a reliable spread.

We also assess the forecast errors for heavy precipitation using an extension of the CRPS - the threshold-weighted CRPS (twCRPS) (Gneiting and Ranjan, 2011). Given a threshold of heavy precipitation, the twCRPS in this study is defined as:

$$\text{twCRPS}(t) = \int_{y_{obs,q}}^{+\infty} \{F_t[y_{cali}(t)] - H_t[y_{cali}(t) - y_{obs}(t)]\}^2 dy_{cali} \tag{10}$$

where $y_{obs,q}$ denotes the threshold used, corresponding to the *q* quantile of observations at each location. In this study, we set *q* to 95%.

Based on CRPS or twCRPS, a skill score can be given by calculating the percentage of reduction in errors of forecasts from reference forecasts:

$$\text{Skill Score} = \frac{\bar{S} - \overline{S_{\text{ref}}}}{0 - \overline{S_{\text{ref}}}} \times 100\% \tag{11}$$

where *S* denotes CRPS or twCRPS of the forecasts; S_{ref} denotes the corresponding CRPS or twCRPS of the reference forecasts; and overbars represent averaging over the validation period.

In this study, ensemble climatology forecasts are used as reference forecasts. More specifically, the SCC fitted climatology established in the locally optimized power-transformed space is used to generate the reference forecasts. A negative skill score indicates the forecast quality is poorer than the reference forecasts and vice versa. A skill score of 100% indicates that the forecasts perfectly match the observations.

3. Case study

Twenty gauged locations covering a wide range of geographical and climate zones in Australia are selected for this study. The gauged locations and the climate zones defined by Peel et al. (2007) are presented in Fig. 2. The basic information, such as the elevation information and precipitation characteristics for the selected locations is provided in Table S1 in the supplementary material, highlighting the climatologically diverse choice of locations. Daily observations of precipitation are obtained from the SILO (Scientific Information for Land Owners) dataset (Jeffrey et al., 2001). We use the SILO point data (01/04/1999 – 31/03/2019) for each selected location with long term observational records, noting that the quality of the actual observations has been checked by the Australian Bureau of Meteorology. Deterministic precipitation forecasts are derived from the global version of Australian Community Climate and Earth-System Simulator (ACCESS-G2) model, which was used operationally by the Bureau of Meteorology in Australia. The model was run four times a day. For each run, hourly forecasts were produced at a spatial resolution of 0.35°(longitude) × 0.23°(latitude) out to a lead time of 240 hours. In this study, the precipitation forecasts issued at 1200 UTC are accumulated to 24-hour daily sums over the subsequent nine days, and bilinearly interpolated to the selected locations. The calibration experiments and evaluation are focused on the forecast period of 01/04/2016 – 31/03/2019.

For normalizing precipitation, we apply the different transformation approaches presented in section 2.1 to both the observed data and the forecast data. Then we apply the SCC model to the transformed data for each location and each forecast lead time. The establishment and evaluation of the SCC model is conducted using leave-one-month-out cross-validation as introduced in section 2.3. Specifically, for the 3-yr period of 04/2016-03/2019 used in this study, each of the 36 months is targeted as a validation month successively.

4. Results

In this section, the transformation parameter values applied in the locally and regionally optimized two parameter value approaches are first presented in section 4.1. Results of normalization effects of the three transformation approaches are presented in section 4.2. Then, detailed calibration results of each approach are presented in section 4.3. In section 4.4, the performance of power transformation is compared to that of the log-sinh transformation under the locally optimized two parameter value approach. A summary of the overall differences in the calibration performances among the approaches is presented in section 4. 5. Finally, spatial analysis results of the regionally optimized two parameter value approach is presented in section 4.6.

4.1. Locally and regionally optimized parameter values

Boxplots of locally optimized parameter values for the observation variable and the forecast variable of different lead times are shown in Fig. 3. The locally optimized parameter values for the observation variable vary between 0.33 and 0.85, while the locally optimized parameter values for the forecast variable are overall lower than those of the observation variable and largely consistent among different lead times, ranging from 0.07 to 0.49. It is unclear why the NWP forecasts are more skewed (i.e., of lower transformation parameter values) than observations. One possibility is that NWP models tend to produce precipitation in small amounts on too many days but occasionally forecast very large events.

For both observation and forecast variables, the regionally optimized parameter values display good correspondences to the medians of the locally optimized values across the 20 locations, with an approximation of 0.47 for the observation variable and 0.20 for the forecast variable irrespective of the lead times. Given the previously shown large ranges of the locally optimized parameter values, it is of concern to us whether using regionally optimized parameter values that are strictly constrained to constants will adversely affect forecast calibration. We investigate this by examining the effectiveness of normalization transformation and validation results of forecast calibration next.

4.2. Comparison of normalization effect with different power transformation approaches

The effect of normalization from each transformation approach for observation and forecast variables is evaluated using the Anderson-Darling (A-D) normality test statistic both before and after data transformations. The results show that observation and forecast variables are clearly far from being normally distributed before transformations. The A-D statistic is much reduced after all the transformations. However, on a closer inspection of the results after transformation, using the pre-fixed approaches cannot achieve acceptable normalization for either the observation or forecast variables or both. The locally optimized two parameter value approach achieves the best normalization results, followed closely by the regionally optimized two parameter value approach. The results for the 20 locations are shown in Fig. S1 in the supplementary material.

4.3. Comparison of forecast calibration with different power transformation approaches

Forecasts calibrated using different transformation approaches are compared by examining the bias, sharpness, reliability, CRPS skill score and twCRPS skill score. The locally optimized two parameter value approach is used as the benchmark for the evaluation. For each evaluation metric, the results from the other two approaches are separately plotted against the results from the benchmark, as shown in Fig. 4 to Fig. 7. Only the results for lead days of 1, 2, 5 and 9 are shown to make the presentations concise.

As can be seen in Fig. 4, the mean of calibrated ensemble forecast across the majority of locations demonstrates positive bias. There is no obvious trend that the bias changes over the forecast lead time. The pre-fixed one parameter value approach exhibits a larger percentage of bias than the locally optimized two parameter value approach. The forecast bias from the pre-fixed one parameter value approach significantly increases as the parameter value decreases from 1/2 to 1/4. The bias from the regionally optimized two parameter value approach is similar to, and sometimes slightly smaller than, that from the locally optimized two parameter value approach. Large bias mostly results from tails of predictive distributions, which are very sensitive to transformation parameter values. The transformation parameter values derived from the locally optimized two parameter value approach is more likely to be subject to sampling errors than the regionally optimized two parameter value approach, leading to the tails of the fitted distributions occasionally deviate significantly from sample values.

As shown in Fig. 5, the different transformation approaches result in similar sharpness except for the pre-fixed 1/4 approach, which leads to notably wider ensemble spread. In addition, all approaches lead to wider ensemble spread as the lead time increases, demonstrating higher forecast uncertainty at the longer lead time.

In terms of the reliability evaluation, all three transformation approaches can yield statistically equivalently reliable ensemble forecasts. None of the α-index of PIT is below 0.95, indicating that the distributions of calibrated ensemble forecasts are consistent with the observation frequency (Fig. S2).

As shown in Fig. 6, the CRPS skill scores for the calibrated forecasts are mostly positive, indicating that the calibrated forecasts are more skillfull than the referenced climatology forecasts. The CRPS skill scores gradually reduce to zero over the increasing lead time, as the majority of skill scores resulted from the NWP models diminish with lead time. The pre-fixed one parameter value approach has consistently lower CRPS skill scores than the locally optimized two parameter value approach, especially for forecasts at lead time of day 1 and 2. The regionally optimized two parameter value approach reaches nearly the same scores as the locally optimized two parameter value approach.

In terms of the calibration performance for the heavy precipitation forecasts evaluated using the twCRPS skill scores (Fig. 7), the pre-fixed one parameter value approach is clearly inferior to the locally optimized two parameter value approach, while the regionally optimized two parameter value approach performs as well as the locally optimized two parameter value approach. A significant decrease of twCRPS skill scores over the increasing lead time is also noticed for each transformation approach.

A clear pattern has emerged from the CRPS and twCRPS skill score results. The pre-fixed one parameter value approach generates less skillful ensemble forecasts, especially for heavy precipitation, while the regionally optimized two parameter value approach leads to ensemble forecasts that are as skillful as the locally optimized two parameter value approach, regardless of heavy or non-heavy precipitation.

4.4. Comparison of forecast calibration with power transformation and log-sinh transformation

The last rows of Fig. 4 to Fig. 7 show the comparison between power transformation and log-sinh transformation used in the locally optimized two parameter value approach with respect to their corresponding calibration performances measured using bias, sharpness, CRPS skill scores and twCRPS skill scores. The calibrated ensemble forecasts using the two transformation methods possess nearly the same bias and spread, and similar skill scores. The slight decrease of twCRPS skill scores from the log-sinh transformation is only noticeable with the lead time of day 1. This is likely to be caused by the wider distribution tails of calibrated forecasts resulted from the log-sinh transformation. The finding of the very slight drop in twCRPS skill scores from the log-sinh transformation is consistent with Li et al. (2019).

4.5 Overall comparison between the transformation approaches

Beyond visual inspection of forecast calibration above, the performances of the three transformation approaches are further analysed by showing the cumulative distribution functions (CDFs) of the differences in evaluation metrics, using again the locally optimized two parameter value approach as the benchmark. The differences are first calculated for individual locations and lead times and then plotted as CDF curves, as shown in Fig. 8. The pre-fixed one parameter value approach mostly yields positive differences in bias and sharpness. But some of these differences may not be significant relative to the magnitude of bias and ensemble spead, as shown in Fig. 4 and 5. Differences in CRPS and twCRPS skill scores of the pre-fixed one parameter value approach are mostly negative, which indicate its worse performance than the locally optimized two parameter value approach. In contrast, the regionally optimized two parameter value approach yields mostly close to zero differences in bias, sharpness, and CRPS and twCRPS skill scores, with some positive and negative differences cancelling each other, indicating similar performance to the locally optimized two parameter value approach. In terms of the α-index of PIT, all approaches show close to zero differences, indicating that the three approaches are similarly reliable.

4.6 Spatial analysis of the regionally optimized two parameter value approach

To demonstrate the generalizability of the regionally optimized two parameter value approach, we first analyze the distribution of transformation parameter values and then we compare the calibration performances in different climate contexts. As shown in Fig. S3, the distributions of the locally optimized values for both observation and forecast variables do not follow a clear spatial pattern. However, it is evident that the parameter values for the observation variable are always much higher than those for the forecast variable. Nevertheless, maps of the CRPS skill scores and twCRPS skill scores shown in Fig. S4 and S5 respectively demonstrate that the regionally optimized two parameter value approach exhibits the same spatial pattern of calibration performances as the locally optimized two parameter value approach. Higher calibration skill scores are observed in eastsouth and westsouth coast region of Australian mainland and Tasimania island, which are all dominated by the temperate climate type, while the northern tropical region displays the lowest skill scores, followed by the central arid region.

5. Discussion

Overall, the results show that the pre-fixed one parameter value approach performs the worst and the regionally optimized two parameter value approach performs as well as the locally optimized two parameter value approach in the context of precipitation forecast calibration. The poorer calibration performance with the pre-fixed one parameter value approach is manifested as having larger bias, slightly wider ensemble spread and lower skills scores especially for heavy precipitation.

The calibration performance of each transformation approach is related to its effectiveness of normalization. The pre-fixed one parameter value approach results in poor calibration performance, and this is because the transformation value may not be ideally selected without considering the data characteristics. This is evident from Fig. 3 that the optimized power values for observation variable are overall higher than those for forecast variable. Therefore, using just one value for both variables as in the pre-fixed one parameter value approach is not able to normalize and homogenize the variables. In contrast, the regionally optimized two parameter value approach accommodates the high and low transformation parameter values for observation and forecast variables respectively, achieving very similar normalization effect to the locally optimized two parameter value approach. Therefore, we anticipate that the regionally optimized two parameter value approach proposed will also work well with other Gaussian-based models that require data transformation.

The locally optimized two parameter value approach theoretically leads to the best results of calibration model for individual locations, but may inhibit the spatial connection of individual calibration models for regional applications. This inhibition can be overcome by using a pre-fixed power transformation. However, this study has proven the pre-fixed one parameter value approach severely affects the calibration performance. The regionally optimized two parameter value approach is proposed for taking advantage of keeping transformation value constant across space, and it still leads to nearly the same calibration performance as the locally optimized two parameter value approach.

The evaluation of the three transformation approaches is presented for 20 selected locations in this study. For the regionally optimized two parameter value approach, the regionally optimized values may change with the selection of sites, observation dataset, NWP model or region of interest. Using the same NWP model, we repeat our analysis with a much larger number of locations (over 1000) for Australia and obtain the same regionally optimized parameter values for observation and forecast variables. Results of the locally optimized parameter values at over 1000 locations are provided in Fig. S6. For applications elsewhere or with different NWP models, regional optimum values of the transformation parameters can be found by following the same procedure as outlined in this paper.

6. Conclusions

Transformation for data normalization is an integral part of the well-known Gaussian-based calibration models for post-processing precipitation forecasts from NWP models. In this study, we investigate three different approaches to data transformation, including the pre-fixed one parameter value approach, the locally optimized two parameter value approach and the regionally optimized two parameter value approach. The pre-fixed one parameter value approach is to use one pre-fixed transformation parameter value for both observation and forecast variables. The locally optimized two parameter value approach is to locally optimize the parameter values, allowing the values to differ for the two variables and the locations. The regionally optimized two parameter value approach is to optimally specify parameter values that are different for the two variables but constant in a region. The three approaches are systematically compared in terms of their impacts on forecast calibration using daily precipitation data from 20 locations across Australia.

Compared with the locally optimized two parameter value approach, the pre-fixed one parameter value approach is found to lead to calibrated forecasts that have higher bias, wider spread, and lower skill. With this approach, the positive bias and ensemble spread tend to increase when a smaller value of transformation parameter is applied. The regionally optimized two parameter value approach is found to achieve the same calibration performance as the locally optimized two parameter value approach. This is because the regionally optimized two parameter value approach is nearly as good as the locally optimized two parameter value approach in achieving variable normalization. In addition, since the regionally optimized two parameter value approach yields stable transformation values, its calibration performance is as skillful as the locally optimized two parameter value approach across climatologically diverse locations.

The regionally optimized two parameter value approach has significant advantages over the other two approaches. Compared with the pre-fixed one parameter value approach, the regionally optimized two parameter value approach can accommodate the high and low transformation parameter values required for normalizing observation and forecast variables respectively, leading to significant improvement of forecast calibration. It is also better at bringing about spatial consistency in the calibration model than the locally optimized two parameter value approach. In our future work, we will take the advantages of the regionally optimized two parameter value approach to calibrate NWP precipitation forecasts over a large number of grid cells in a region where observations for training are available only for gauged locations.

We recommend the use of the regionally optimized two parameter value approach, because of its strong performance in forecast calibration and its advantage for regional applications. The approach may also be relevant to other applications that involve spatial analysis of multivariate datasets and data transformations.

Acknowledgments

This study is undertaken as part of an ARC Linkage Project (LP170100922). We thank the Queensland Department of Environment and Science for the SILO meteorological data used in this study. We thank the Australian Bureau of Meteorology for providing access to the ACCESS-G2 data used in this study. Wenyan Wu acknowledges financial support provided by an ARC Discovery Early Career Researcher Award (DE210100117).

Appendix A

Given a sample of $y(t), t = 1, 2, \ldots, T$, where some of the values are only known as lower than a certain threshold, the Anderson-Darling (A-D) normality test statistic can be calculated using:

$$A^2 = -\frac{1}{T}\sum_{i=1}^{k}(2i-1)[\log(p(i) - \log(1-p(i))]] - 2\sum_{i=1}^{k}\log\big(1-p(i)\big) -$$

$$\frac{1}{T}\left[(k-T)^2\log\big(1-p(k)\big) - k^2\log\big(p(k)\big) + T^2p(k)\right] \quad \text{(A1)}$$

where $$p(i) = 1 - \mathrm{cdf}_{\mathrm{N}}\big(y_{(i)};\ \mu, \sigma^2\big) \quad \text{(A2)}$$

$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(k)}, (k < T)$ are observed values from $y(t)$, μ and σ are the mean and standard deviation of an underlying normal distribution of $y(t)$.

Appendix B

Given the daily transformed forecasts $f(t)$ and observations $o(t), t = 1, 2, \ldots, T$, the Seasonal Coherent Calibration (SCC) model can be described as:

$$[f(t), o(t)] \sim \mathrm{BN}\{f(t), o(t) | \mu_f[m(t)], \sigma_f^2[m(t)], \mu_o[m(t)], \sigma_o^2[m(t)], \rho[m(t)]\} \quad \text{(A1)}$$

where $m(t)$ is a function that marks the month of the year for day t, $m(t) = k \epsilon \{1, 2, \ldots, 12\}$; $\mu_f[m(t)]$ and $\sigma_f[m(t)]$ are the mean and standard deviation of the climatological distribution of $f(t)$; $\mu_o[m(t)]$ and $\sigma_o[m(t)]$ are the mean and standard deviation of the climatological distribution of $o(t)$; and $\rho[m(t)]$ is the correlation between $f(t)$ and $o(t)$. For the climatological distribution of $f(t)$ or $o(t)$, the statistical characteristics are assumed as uniform within one month but allowed to vary from month to reflect seasonal variation.

To estimate the above distribution parameters robustly, a two-step approach is implemented:

(i) $\mu_o(k)$ and $\sigma_o(k)$ are first estimated using the observations of a long-term period (say the last 20 years), rather than the common period when forecast data is available. This is because the observations during the common period are typically insufficient to infer a representative climatology, this step can make sure the SCC outputs are nudged towards climatology-like forecasts when the raw NWP forecast as input has little skill. To estimate $\mu_o(k)$ and $\sigma_o(k)$, the daily transformed observations in month k of each year are used to maximize the likelihood function of a normal distribution.

(ii) For estimating $\mu_f(k), \sigma_f(k)$ and $\rho(k)$, these 36 parameters are replaced with fewer parameters for working effectively with a short period (say 3 years) of forecast data. This reparameterization is achieved by establishing underlying relationships between the forecast and observed climatology:

$$\mu_f(k) = a + b\mu_o(k) \tag{B2}$$

$$\sigma_f(k) = c + d\sigma_o(k) \tag{B3}$$

$$\rho(k) = r \tag{B4}$$

where $b \geq 0$, $c \geq 0$, $d \geq 0$, and $0 \leq r \leq 1$. The Eq. B2 and Eq. B3 assume that the seasonal variation of forecast climatology is consistent with that of the observed climatology, so that the bivariate normal distributions of 12 months can be linked up for robust parameter estimation. Parameters a, b, c, d, and r are estimated using the method of maximum likelihood given the observed and forecast data during the common period. Once the parameters $\mu_o(k)$, $\sigma_o(k)$, a, b, c, d, and r are determined, the $\mu_f(k), \sigma_f(k)$ and $\rho(k)$ are available.

With all the parameter values in the Eq. A1 available, we can derive a conditional distribution for $o(t)$ as a calibrated probabilistic forecast when a new transformed forecast $f(t)$ is given.

Appendix C

The log-sinh transformation can be written as:

$$\hat{y} = \frac{1}{\lambda}\log[\sinh(\varepsilon + \lambda y)] \quad \text{(C1)}$$

where y and $\hat{y}$ are the original and transformed data, respectively; ε and λ are the transformation parameters. The maximum likelihood function is applied to estimate the transformation parameters and the normal distribution parameters. The likelihood function is as follows:

$$L(\varepsilon, \lambda, \mu, \sigma) = \prod_{t=1}^{T} l(t) \quad \text{(C2)}$$

$$l(t) = \begin{cases} \mathrm{pdf_N}(\hat{y}(t);\ \mu, \sigma^2) \times \coth[\varepsilon + \lambda y(t)] & \text{if} \quad y(t) > y_c \\ \mathrm{cdf_N}(\hat{y}_c;\ \mu, \sigma^2) & \text{if} \quad y(t) \le y_c \end{cases} \quad \text{(C3)}$$

The parameter values that maximize the likelihood function are found using the Nelder-Mead downhill simplex method.

The transformed $\hat{y}$ can be back-transformed to y by:

$$y = \frac{1}{\lambda}\{\mathrm{arcsinh}[exp(\lambda\hat{y})] - \varepsilon\} \quad \text{(C4)}$$

References

Ben Bouallègue, Z., 2013. Calibrated short-range ensemble precipitation forecasts using extended logistic regression with interaction terms. Weather Forecast. 28, 515–524. https://doi.org/10.1175/WAF-D-12-00062.1

Berrocal, V.J., Raftery, A.E., Gneiting, T., 2008. Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. Ann. Appl. Stat. 2, 1170–1193. https://doi.org/10.1214/08-AOAS203

Cecinati, F., Wani, O., Rico-Ramirez, M.A., 2017. Comparing Approaches to Deal With Non-Gaussianity of Rainfall Data in Kriging-Based Radar-Gauge Rainfall Merging. Water Resour. Res. 53, 8999–9018. https://doi.org/10.1002/2016WR020330

Cressie, N.A.., 1993. Statistics for Spatial Data, Wiley. ed. New York.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30, 1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014

Duan, Q., Pappenberger, F., Wood, A., Cloke, H.L., Schaake, J.C., 2019. Handbook of Hydrometeorological Ensemble Forecasting. Springer, Berlin.

Erdin, R., Frei, C., Künsch, H.R., 2012. Data transformation and uncertainty in geostatistical combination of radar and rain gauges. J. Hydrometeorol. 13, 1332–1346. https://doi.org/10.1175/JHM-D-11-096.1

Fu, G., Viney, N.R., Charles, S.P., 2010. Evaluation of various root transformations of daily precipitation amounts fitted with a normal distribution for Australia. Theor. Appl. Climatol. 99, 229–238. https://doi.org/10.1007/s00704-009-0137-6

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B 69, 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gneiting, T., Ranjan, R., 2011. Comparing density forecasts using threshold- and quantile-weighted scoring rules. J. Bus. Econ. Stat. 29, 411–422. https://doi.org/10.1198/jbes.2010.08110

Hamill, T.M., Whitaker, J.S., Wei, X., 2004. Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. Mon. Weather Rev. 132, 1434–1447. https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2

Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. Water Resour. Res. 51, 7436–7451.

https://doi.org/10.1002/2014WR016473.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast. 15, 559–570. https://doi.org/https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Model. Softw. 16, 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1

Kleiber, W., Raftery, A.E., Gneiting, T., 2011. Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. J. Am. Stat. Assoc. 106, 1291–1303. https://doi.org/10.1198/jasa.2011.ap10433

Krzysztofowicz, R., Evans, W.B., 2008. Probabilistic forecasts from the national digital forecasts database. Weather Forecast. 23, 270–289. https://doi.org/10.1175/2007WAF2007029.1

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. Wiley Interdiscip. Rev. Water 4, e1246. https://doi.org/10.1002/wat2.1246

Li, W., Duan, Q., Wang, Q.J., 2019. Factors influencing the performance of regression-based statistical postprocessing models for short-term precipitation forecasts. Weather Forecast. 34, 2067–2084. https://doi.org/10.1175/WAF-D-19-0121.1

Li, W., Wang, Q.J., Duan, Q., 2020. A variable-correlation model to characterize asymmetric dependence for postprocessing short-term precipitation forecasts. Mon. Weather Rev. 148, 241–257. https://doi.org/10.1175/MWR-D-19-0258.1

Messner, J.W., Mayr, G.J., Wilks, D.S., Zeileis, A., 2014. Extending extended logistic regression: Extended versus separate versus ordered versus censored. Mon. Weather Rev. 142, 3003–3014. https://doi.org/10.1175/MWR-D-13-00355.1

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Robertson, D.E., Shrestha, D.L., Wang, Q.J., 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. Hydrol. Earth Syst. Sci. 17, 3587–3603. https://doi.org/10.5194/hess-17-3587-2013

Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D.J., 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrol. Earth Syst. Sci. Discuss. 4, 655–717. https://doi.org/10.5194/hessd-4-655-2007

Schepen, A., Wang, Q.J., 2014. Ensemble forecasts of monthly catchment rainfall out to long lead times by post-processing coupled general circulation model output. J. Hydrol. 519, 2920–2931. https://doi.org/10.1016/j.jhydrol.2014.03.017

Schepen, A., Wang, Q.J., 2013. Toward accurate and reliable forecasts of Australian seasonal rainfall by calibrating and merging multiple coupled GCMs. Mon. Weather Rev. 141, 4554–4563. https://doi.org/10.1175/MWR-D-12-00253.1

Schepen, A., Zhao, T., Wang, Q.J., Robertson, D.E., 2018. A Bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments. Hydrol. Earth Syst. Sci. 22, 1615–1628. https://doi.org/10.5194/hess-22-1615-2018

Scheuerer, M., 2014. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. Q. J. R. Meteorol. Soc. 140, 1086–1096. https://doi.org/10.1002/qj.2183

Scheuerer, M., Hamill, T.M., 2015. Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. Mon. Weather Rev. 143, 4578–4596. https://doi.org/10.1175/MWR-D-15-0061.1

Sloughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Mon. Weather Rev. 135, 3209–3220. https://doi.org/10.1175/MWR3441.1

Stauffer, R., Umlauf, N., Messner, J.W., Mayr, G.J., Zeileis, A., 2017. Ensemble postprocessing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies. Mon. Weather Rev. 145, 955–969. https://doi.org/10.1175/MWR-D-16-0260.1

Thode, H.C., 2002. Tesing for Normality.

Vannitsem, S., Wilks, D.S., Messner, J.W., 2018. Statistical postprocessing of ensemble forecasts, Elsevier. Amsterdam.

Wang, Q.J., Bennett, J.C., Robertson, D.E., Li, M., 2020. A Data Censoring Approach for Predictive Error Modeling of Flow in Ephemeral Rivers. Water Resour. Res. 56, 1–19. https://doi.org/10.1029/2019WR026128

Wang, Q.J., Robertson, D.E., 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. Water Resour. Res. 47, 1–19. https://doi.org/10.1029/2010WR009333

Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resour. Res. 45, 1–18. https://doi.org/10.1029/2008WR007355

Wang, Q.J., Shao, Y., Song, Y., Schepen, A., Robertson, D.E., Ryu, D., Pappenberger, F., 2019b. An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. Environ. Model. Softw. 122, 104550. https://doi.org/10.1016/j.envsoft.2019.104550

Wang, Q.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. Water Resour. Res. 48, 1–7. https://doi.org/10.1029/2011WR010973

Wang, Q.J., Zhao, T., Yang, Q., Robertson, D., 2019a. A seasonally coherent calibration (SCC) model for postprocessing numerical weather predictions. Mon. Weather Rev. 147, 3633–3647. https://doi.org/10.1175/MWR-D-19-0108.1

Wilks, D.S., 2011. Statistical methods in the atmospheric sciences, Third. ed. Elsevier.

Wilks, D.S., 2009. Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteorol. Appl. 16, 361–368. https://doi.org/10.1002/met.134

Wu, L., Seo, D.J., Demargne, J., Brown, J.D., Cong, S., Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. J. Hydrol. 399, 281–298. https://doi.org/10.1016/j.jhydrol.2011.01.013

Yang, Q., Wang, Q.J., Hakala, K., 2021. Achieving effective calibration of precipitation forecasts over a continental scale. J. Hydrol. Reg. Stud. 35, 100818. https://doi.org/10.1016/j.ejrh.2021.100818

Zhao, P., Wang, Q.J., Wu, W., Yang, Q., 2022. Extending a joint probability modelling approach for post-processing ensemble precipitation forecasts from numerical weather prediction models. J. Hydrol. 605, 127285. https://doi.org/10.1016/j.jhydrol.2021.127285

Zhao, P., Wang, Q.J., Wu, W., Yang, Q., 2020. Which precipitation forecasts to use? Deterministic versus coarser-resolution ensemble NWP models. Q. J. R. Meteorol. Soc. 1–14. https://doi.org/10.1002/qj.3952