Direct and Indirect Application of Univariate and Multivariate Bias Corrections on Heat-stress Indices based on Multi-RCM Simulations

Liying Qiu[1*], Eun-Soon Im[1,2*], Seung-Ki Min[3], Yeon-Hee Kim[3], Dong-Hyun Cha[4], Seok-Woo Shin[4], Joong-Bae Ahn[5], Eun-Chul Chang[6], and Young-Hwa Byun[7]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[3]Division of Environmental Science and Engineering, Pohang University of Science and Technology, South Korea
[4]Ulsan National Institute of Science and Technology, South Korea
[5]Pusan National University, South Korea
[6]Kongju National University, South Korea
[7]National Institute of Meteorological Science, South Korea

Correspondence to: Eun-Soon Im (ceim@ust.hk) and Liying Qiu (liying.qiu@connect.ust.hk)

Abstract. Statistical bias correction (BC) is a widely used tool to post-process climate model biases ~~for~~in heat-stress impact studies, which are often based on the indices calculated from multiple dependent variables. This study compares five bias correction methods (four univariate and one multivariate) with two applying strategies (direct and indirect) for correcting two heat-stress indices with different dependencies on temperature and relative humidity, using multiple Regional Climate Model simulations over South Korea. It would be helpful for reducing the ambiguity involved in the practical application of BC for climate modeling ~~as well as~~and end-user communities. Our results demonstrate that the multivariate approach can improve the corrected inter-variable dependence, which ~~and therefore~~ benefits the indirect correction of heat-stress indices ~~that~~depending on the adjustment of individual components, especially those indices relying equally on multiple drivers. On the other hand, the direct correction of multivariate indices using the Quantile Delta Mapping univariate approach can also produce a comparable performance in the corrected heat-stress indices. However, our results also indicate that attention should be paid to the non-stationarity of bias brought by climate sensitivity in the modeled data, which may affect the bias-corrected results unsystematically. Careful interpretation of the correction process is required for an accurate heat-stress impact assessment.

1 Introduction

Climate models unavoidably produce biased representations of the simulated variables, and it is more problematic not to know how these biases translate into the modeled response to external forcings such as the CO_2 concentration, which is known to be responsible for global warming. Therefore, statistical bias correction (BC) of climate model outputs has been progressively adopted as a standard procedure to improve their performance, in particular when feeding them into various climate change impact assessments (e.g., G. Kim et al., 2020; K. B. Kim et al., 2022; Masaki et al., 2015; Qiu et al., 2022; Schwingshackl et

al., 2021). Indeed, the visible benefits archived by adjusting simple statistics (e.g., mean, variance) have led to the wide application of BC. A significant body of research demonstrated that the systematic biases observed in the long-term pattern of the current climate can be well eliminated even when using a very simple technique (e.g., linear scaling). However, the effectiveness of BC methods and their improper assumptions (e.g., statistical stationarity) remain a topic for debate (Maraun et al., 2017). For example, the nonstationary model bias and the large monthly/seasonal correction factor can potentially degrade the BC's performance, particularly with respect to misleading interpretations of extremes (Chen et al., 2021; Lee et al., 2019a, b). ~~On the other hand~~Meanwhile, the choice of BC approaches in different contexts (e.g., heat-stress impact study, hydrological impact study, adjustment of boundary conditions in downscaling) needs careful assessments case by case (Ehret et al., 2012; K. B. Kim et al., 2022; Rocheta et al., 2017; Zscheischler et al., 2019).

A variety of BC methods with different levels of complexity and performance have been developed and implemented for both global and regional climate simulations (François et al., 2020; Y.-T. Kim et al., 2021; Teutschbein & Seibert, 2012). Generally, their aim is to correct certain features in the target's distribution, such as the simple statistics of the mean (Linear Scaling, LS, Teutschbein & Seibert, 2012) and variance (Variance Scaling, VA, Chen & Dudhia, 2001), or the more advanced quantiles (Quantile Mapping, QM) for adjusting the entire distribution by parametric (PQM) or empirical (EQM) transformation (Gudmundsson et al., 2012; Switanek et al., 2017). Continuous efforts have also been made to eliminate the drawbacks of existing BC approaches. Quantile Delta Mapping (QDM, Cannon et al., 2015), for example, is designed to explicitly preserve the long-term trend that may be artificially distorted in QM. Nonetheless, all the approaches described above correct bias in a univariate context. They cannot adjust the inter-variable dependencies, which are important for representing physical processes and estimating compound hazards. It was not until quite recently that the multivariate BC technique was considered and proposed (e.g., Bárdossy & Pegram, 2012; Cannon, 2018; Mehrotra & Sharma, 2015, 2016; Robin et al., 2019; Vrac, 2018), and they have been applied to various climate change impact studies (Dieng et al., 2022; Meyer et al., 2019; Qiu et al., 2022; Zscheischler et al., 2019). Although it is intuitively recognized that multivariate BC could be more suitable for dealing with climate variables characterized by a strong physical linkage in nature, an unambiguous assessment of univariate and multivariate BC methods is essential to understand the potential limitations of individual methods and to avoid misleading application.

~~On the other hand, d~~Despite the BC method used, when correcting the multivariate indices representing compound hazards, the index can also either be directly adjusted using BC techniques, as in the majority of studies (Coffel et al., 2017; Kang et al., 2019; Schwingshackl et al., 2021), or be indirectly corrected that its components are individually corrected prior to the index calculation (Casanueva et al., 2019; Zscheischler et al., 2019). In this regard, there have been few systematic comparisons of how the direct and indirect use of univariate and multivariate BC methods, respectively, affect the multivariate indices adjustment. Only Casanueva et al. (2018) tested the direct and indirect use of EQM in correcting the multivariate fire danger index, while several studies compared the indirect use of univariate and multivariate BC methods in impact assessments (e.g., Cannon, 2018; François et al., 2020; Zscheischler et al., 2019). Although Casanueva et al. (2018) pointed out that the direct

65   application of EQM outperforms the indirect one, how it compares with the multivariate BC method remains unknown. Therefore, there is room for a more comprehensive assessment of the effects of univariate and multivariate BC under direct and indirect applying strategies, which may vary along with the dependence structure of the multivariate indices and may affect correction efficiency since the multivariate approach has a higher computation cost.

In this study, we investigate the effects of different BC methods (univariate vs. multivariate) applied with different strategies
70   (direct vs. indirect) on the statistical adjustment of heat-stress indices that represent the combined effect of human exposure to temperature (T) and relative humidity (RH), using regionally-tailored, fine-scale climate information in Korea from multiple Regional Climate Models (RCMs). The extreme heat is one of the most critical impacts of climate change and we adopt two heat-stress indices with different sensitivities to humidity (Sherwood, 2018), namely, web-bulb globe temperature (WBGT) and apparent temperature (AT). A comparative assessment of the two indices derived from different BC methods and different
75   strategies will provide valuable insights into understanding how the relationship between heat-stress index and its drivers (e.g., T and RH) affects the performance of univariate and multivariate BC for modeled heat stress. This study will be helpful for reducing the ambiguity involved in the practical application of BC for climate modeling as well as end-user communities.

2 Data and Methods

2.1 Data

80   The 3-hourly data used for BC is the near-surface T and RH during the historical period (1979-2014) generated by five RCMs (Table S1) over the CORDEX-East Asia domain (Lee et al., 2020). It is the dynamical downscaling product of the UK Earth System Model (UKESM) in Coupled Model Intercomparison Project Phase 6 (CMIP6). The same variables from ERA5 reanalysis (Hersbach et al., 2018) during the same period are adopted as the observation for BC and validation procedures. For consistency, the variables from all RCMs are first interpolated spatially onto the $0.25° \times 0.25°$ regular latitude-longitude grid
85   of ERA5 and temporally interpolated onto a standard Gregorian calendar. The analysis focuses only on the land area in South Korea.

2.2 Heat-stress Indices

Two popular heat-stress indices are evaluated in this study: WBGT (ACSM, 1984) and AT (Steadman, 1984). There are several different formulations for both indices, and we employ the versions only using T and RH as input variables (i.e., the simplified
90   WBGT and the AT without wind effect, Eq. 1-3). Although both indices are calculated as a function of T and RH, their T/RH dependences are different (Fig. 1). WBGT is more evenly dependent on T and RH, whereas AT relies mostly on T. Also, each index has strengths and limits in evaluating heat-stress impacts (Sherwood 2018; Schwingshackl et al., 2021); thus, they are selected for a more comprehensive evaluation of BC techniques' applicability.

$$WBGT = 0.567T + 0.393e + 3.94 \ (1)$$
95
$$AT = 0.92T + 0.22e - 1.3 \ (2)$$

T is the near-surface temperature in °C and RH is the near-surface relative humidity in %; e is the vapor pressure (hPa) that can be calculated by

$$e = \left(\frac{RH}{100}\right) \times 6.105 \exp\left(\frac{17.27T}{237.7+T}\right) \quad (3)$$

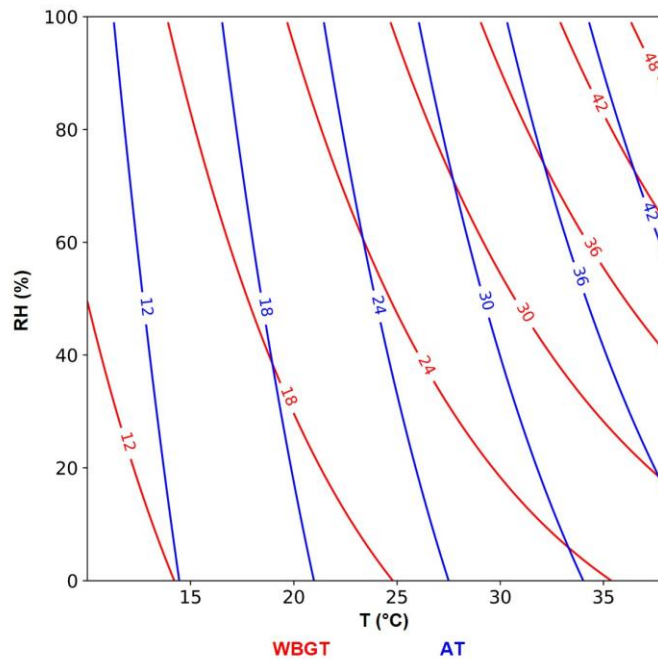
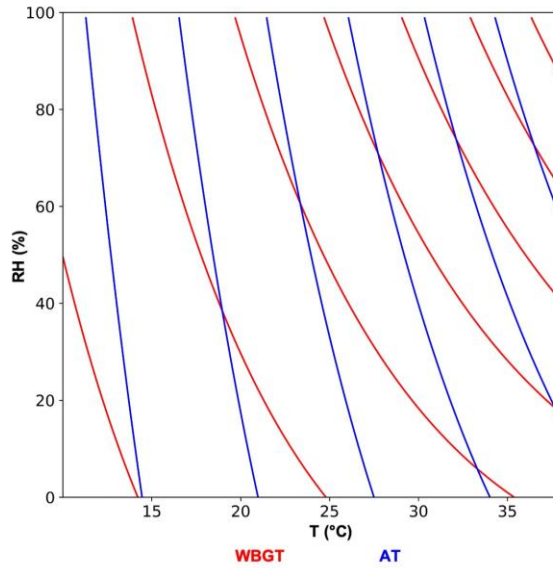



Figure 1 Contours lines of equal-level heat stress indicators: WBGT (red) and AT (blue)

All 3-hourly data are used for the BC procedure, but the daily maximum of WBGT/AT during summer (June-July-August, JJA), together with the T and RH at the corresponding time, are selected for analysis in order to facilitate the use of heat-stress impact studies.

105   2.3 Bias Correction

The principle of BC is to use observations to calibrate the simulated output (e.g., climate model output). In this study, ~~five~~ four BC methods are applied, including LS, VA, ~~EQM,~~ QDM, and a multivariate BC algorithm with an N-dimensional probability density function (MBCn). Information on each BC approach is provided in the Supplement. The five transformation algorithms cover a varying range of complexity, with MBCn being selected as an example of multivariate correction methods and the trend-preserving QDM being a more "advanced" member of the QM family. There are several different multivariate BC methods developed recently based on different statistical techniques and/or assumptions (e.g., Rank Resampling For Distributions and Dependences (R^2D^2, Vrac, 2018), Matrix recorrelation (MRec, Bárdossy & Pegram, 2012). Different multivariate methods have their pros and cons, depending on the varying perspectives considered (François et al., 2020). The MBCn adopted here is based on an image processing algorithm that repeatedly rotates the multivariate matrices and applies QDM correction on individual variables, until the multivariate distribution is matched to observation. It is selected in this study not only due to its wide application in various kinds of climate studies; more importantly, it facilitates the comparison with the univariate QDM as it is built on the latter.

During the BC process, univariate BC methods are applied to T, RH, and WBGT/AT, respectively, after WBGT/AT has been calculated from the original RCM output (ORI). For MBCn, the multivariate approach is applied simultaneously to T, RH, and WBGT (or T, RH, and AT). As the 3-hourly data is adopted, BCs are applied separately to each 3-hour interval in each calendar month (e.g., June00UTC). The direct correction of heat-stress levels is defined as WBGT/AT directly adjusted by BC, while~~;~~ ~~on the other hand,~~ the levels calculated from the bias-corrected T and RH are treated as an indirect correction of the heat-stress indices (marked as WBGT'/TW'). ENS is the unweighted ensemble mean of the five RCM models.

As an illustrative example, Fig. 2 provides the Quantile-Quantile plots of the WBGT corrected using various approaches for one grid point from one RCM during 1979-1996. ORI shows a cold bias inherited from the driving GCM (M.-K. Kim et al., 2020), leading to a notable underestimation over the entire distribution compared to ERA5. For the direct correction of WBGT, LS reduces the cold bias, but with a non-negligible overestimation, especially in the range of 30-32.5°C. This is due to the asymmetric distribution of T being corrected with a single correction coefficient taken only from the monthly mean. VA, on the other hand, provides a significant improvement by additionally taking the variance into account. ~~EQM and~~ QDM, ~~which~~

130   ~~are the same in this case~~equivalent to EQM for the calibration period, manage~~s~~ to show a perfect match with ERA5 across all the quantiles since the empirical distribution is designed to fit the observation. However, moving to the WBGT' obtained from the corrected T and RH, all univariate BC approaches show a degraded performance while only MBCn retains

a qualified correction output. The MBCn's algorithm ensures that the observed multivariate relations (e.g., the T-RH-WBGT pairwise dependency) are reflected in the corrected distribution, resulting in a better indirect correction outcome.

135 For cross-validation of the BC methods, we use a historical period of 1979-2014 and adopt the "jack-knifing" split-sample test, that first splits the historical period into two halves and uses one part for calibration and the other for validation, and then we reverse the two parts systematically (Refsgaard et al. 2014). Specifically, the 18-year period of first 18 years (1979-1996) isare first set as the calibration periodpart with the period of 1997-2014 as the validation part,; then, the periods are swapped using 1997-2014 for calibration and 1979-1996 for validation. For each test, during whichthe -ERA5 data in the corresponding

140 calibration period is used to obtain the correcting algorithms that are then applied to , and the second 18 years (1997-2014) are the validation period to which the correcting algorithms from the calibration period are applied. To distinguish the two tests, the one using 1997-2014 for calibration is all marked with a letter "r" standing for "reverse" and the default is the one using 1979-1996 for calibration. The statistical metrics used for validation evaluation are noted in the Supplement.

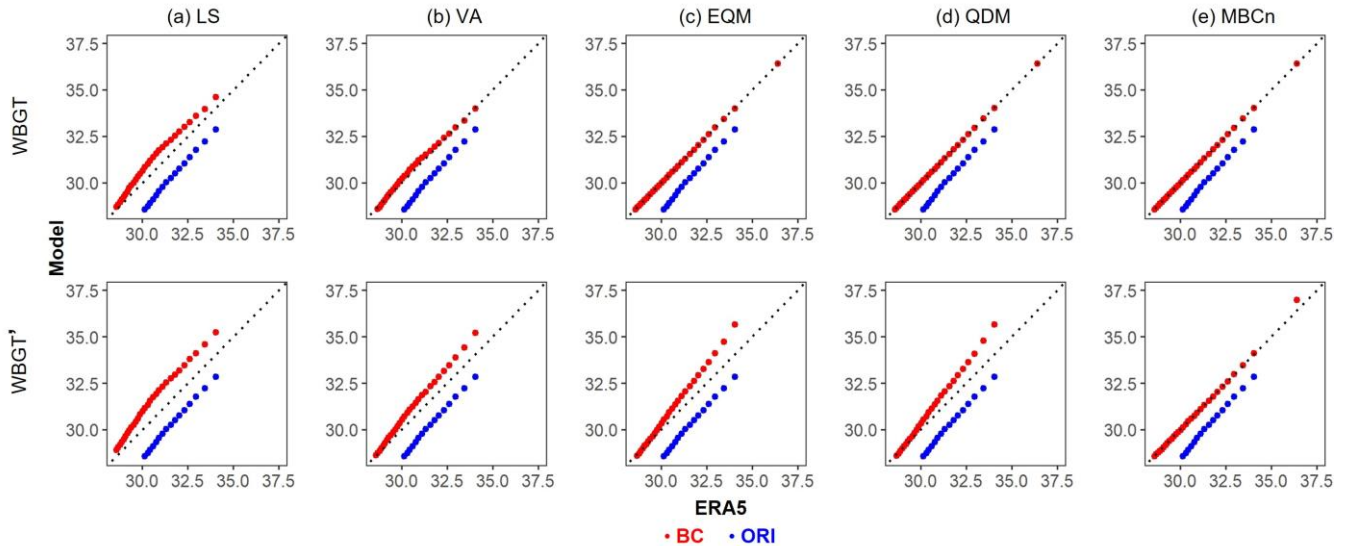
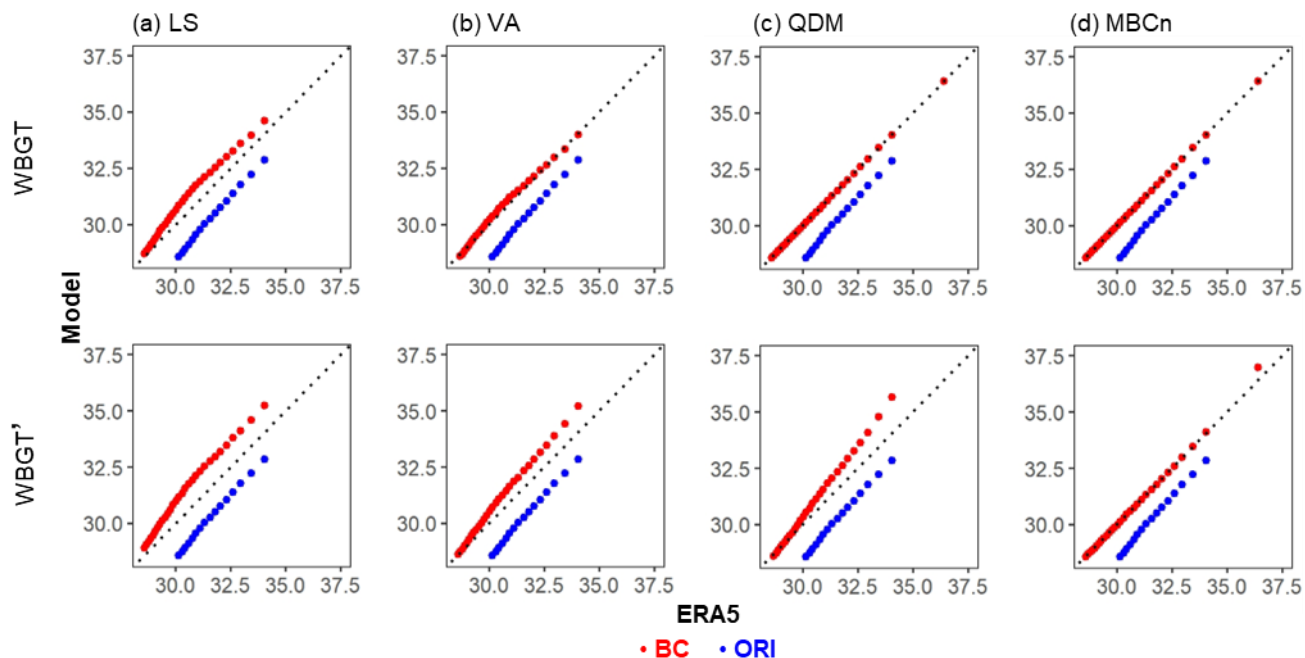



Figure 2: The Quantile-Quantile plots of ORI (blue) and data after BC (red) adjusted by (a) LS, (b) VA, ~~(e) EQM,~~ (c~~d~~) QDM, and (~~e~~d) MBCn. The X-axis is for quantiles from ERA5, and Y-axis is for quantiles from model simulations; the unit is °C. Row 1 is WBGT from direct correction, and Row 2 is WBGT' from indirect correction (calculated from directed T and RH). The data is from one point in GRIMs over South Korea land during the calibration period.

3 Results

Figure 3 presents the performance of WBGT and AT in ORI simulations. Substantial bias can be seen across the entire distribution of the heat-stress indices. For ~~the calibration period~~1979-1996, both WBGT and AT generally exhibit a cold bias covering the whole domain. There is more bias in the bottom and top 15% of the distribution, but the bias of WBGT is more skewed to the left tail, whereas that of AT is more skewed to the right. Taking the 90th percentile (90p) as an indicator representing heat events, Fig. 3b and 3c show a greater cold bias in the low-elevation regions (e.g., basins in southeastern Korea), where an RCM with a spatial resolution of around-20km is highly unlikely to capture the local high temperatures owing to an inadequate representation of topography (Qiu et al., 2020). For ~~the validation period~~1997-2014, however, i.e., the next 18 years within the historical period, the cold bias is systematically reduced, with certain area even displaying a slight warm bias. This can be explained by the high climate sensitivity in the driving GCM (i.e., UKESM; Zelinka et al., 2020), leading to a different level of warming between the simulations and ERA5 during this historical period. According to Fig. 3d and 3e, the model shows around 0.5°C more warming than ERA5 between the two periods, which could in turn "compensate" for the models' cold bias and result in a reduced bias in ~~the validation period~~1997-2014. However, while the biased presentation of the heat-stress indices emphasizes the necessity of BC application, the difference in bias between the two historical periods

underscores the need for caution when using and interpreting BC output in climate models, since BC is built on the fundamental
165 assumption of stationary bias (Teutschbein & Seibert, 2012). In particular, the combined bias from climate representation and
the long-term trend may amplify the non-stationarity of model biases, thereby causing potential problems in the BC output.

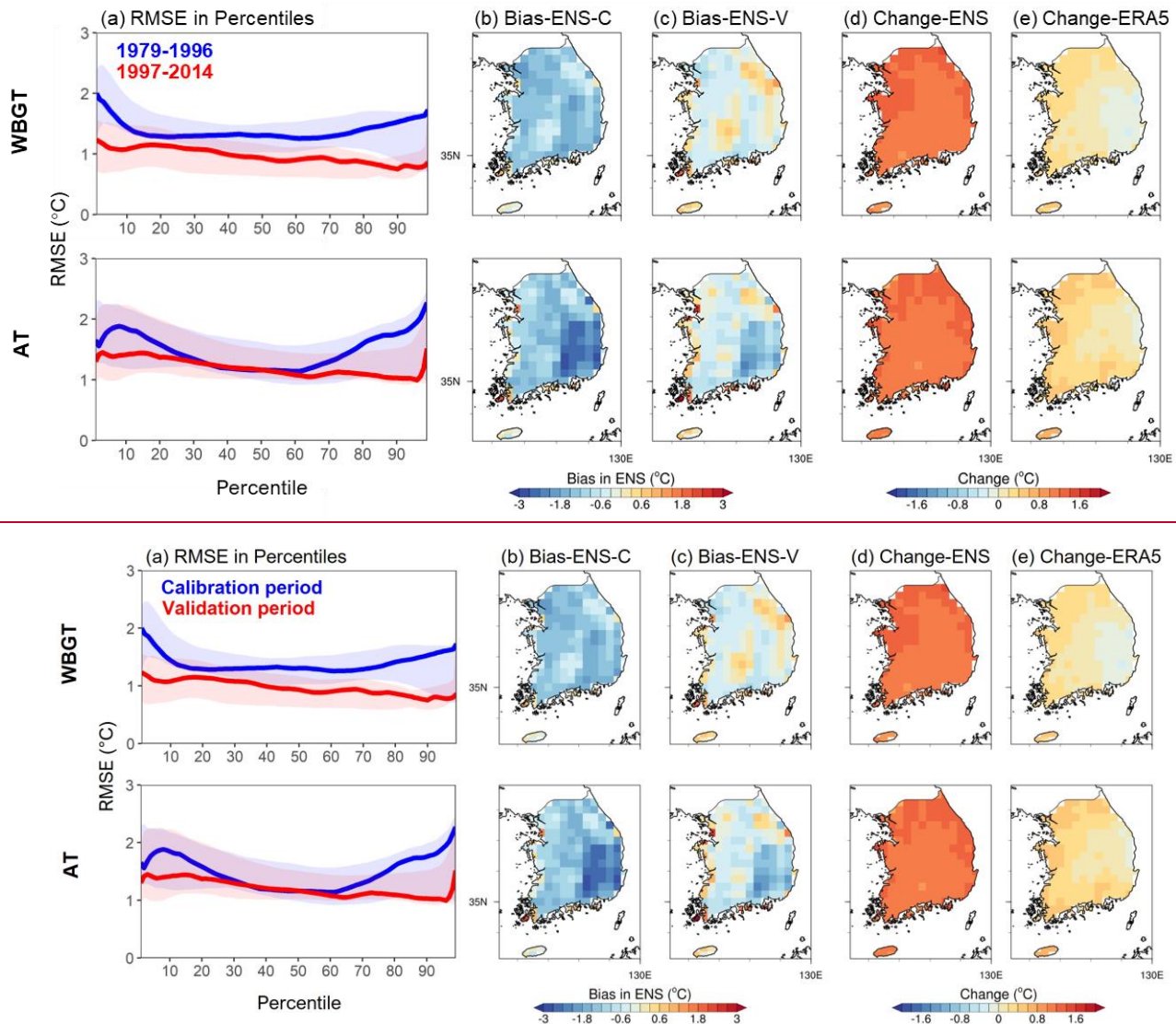



Figure 3: (a) Root-mean-square-error (RMSE, Eq. B1) over the land area of South Korea in percentiles 1-99 during ~~the~~
170 **~~calibration~~1979-1996 (blue) and ~~validation~~1997-2014 (red) ~~period~~. The lines and shading indicate the median and the range,**
respectively, of the five RCMs. (b, c) Spatial map of the bias in the 90p from ENS during the calibration (C, 1979-1996) and validation
(V, 1997-2014) period, respectively. (d, e) The difference between the validation and the calibration period in 90p from ENS and
ERA5, respectively. The upper row is for WBGT, and the lower row is for AT.

Figure 4 shows the median absolute error (MAE, Eq.~~. B2~~S2) over South Korea (land only) in all RCMs after BC using different
175 methods. Two indicators—the 90p and the mean of monthly maximum (MMX)—are selected to represent extreme heat events.

The diamonds standing for ENS are marked for ease of comparison. During the calibration period, LS, as the simplest BC approach used in this study, shows the largest bias among the five methods. For direct correction of WBGT, all other four methods have a reasonable MAE of less than 0.25°C in the 90p and less than 0.5°C in the MMX for ENS, with ~~EQM and~~ QDM slightly outperforming the VA and MBCn approaches. For the indirect correction, ~~on the other hand~~however, there is

180 more variability among the methods and a larger bias than the direct correction. In this case, while LS still shows the worst performance, ~~EQM and~~ QDM presents a degraded performance, with the MAE for WBGT' reaching 0.6°C and 1.2°C in the 90p and MMX, respectively. ~~Although QDM is supposed to differ from EQM in terms of considering the trend between the periods, the limited difference between the two historical periods restricts their difference in this study, so we use QDM to represent the QM family in the below discussion.~~ Surprisingly, VA outperforms the more-advanced QM methods in ENS,

185 indicating the complexity of using univariate approaches to apply an indirect correction for multivariate hazards. In this case, the multivariate approach, i.e., MBCn, clearly demonstrates its strengths in such indirect correction, regardless of the indicators or periods considered. MBCn performs comparably to the direct correction of QDM during the calibration period; however, for the validation period, MBCn surpasses direct correction with an MAE of roughly 0.5°C for both the 90p and MMX. In addition, MBCn shows less variability among the RCMs in WBGT'. For example, the range of MAE for WBGT' during the

190 calibration period as corrected by QDM is 0.38-1.23°C, while that corrected by MBCn is 0.12-0.14°C.

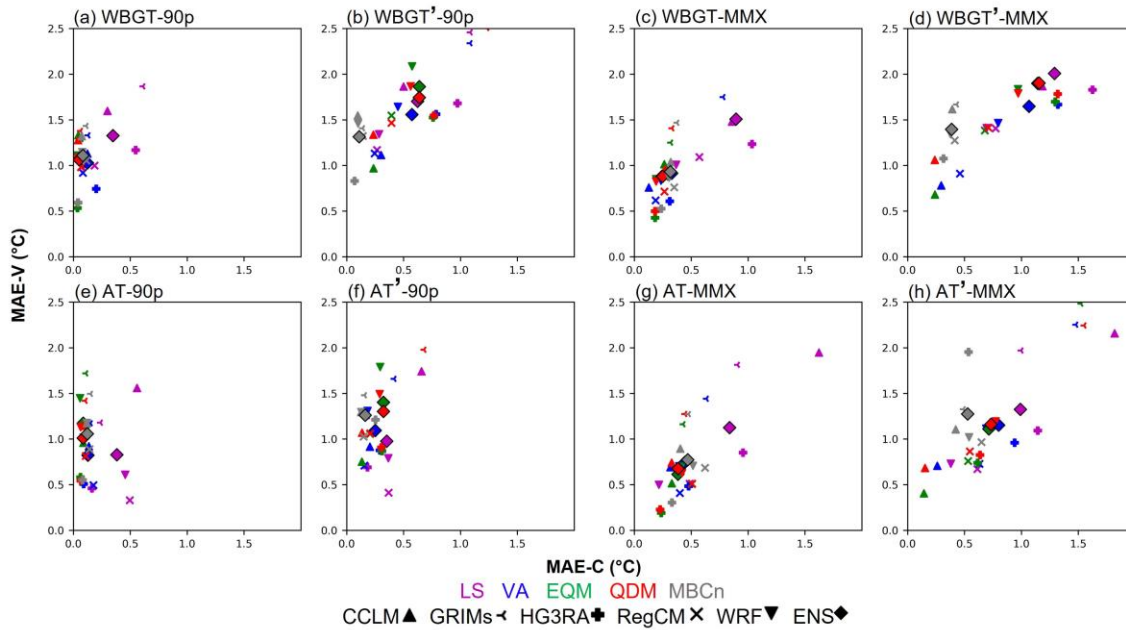
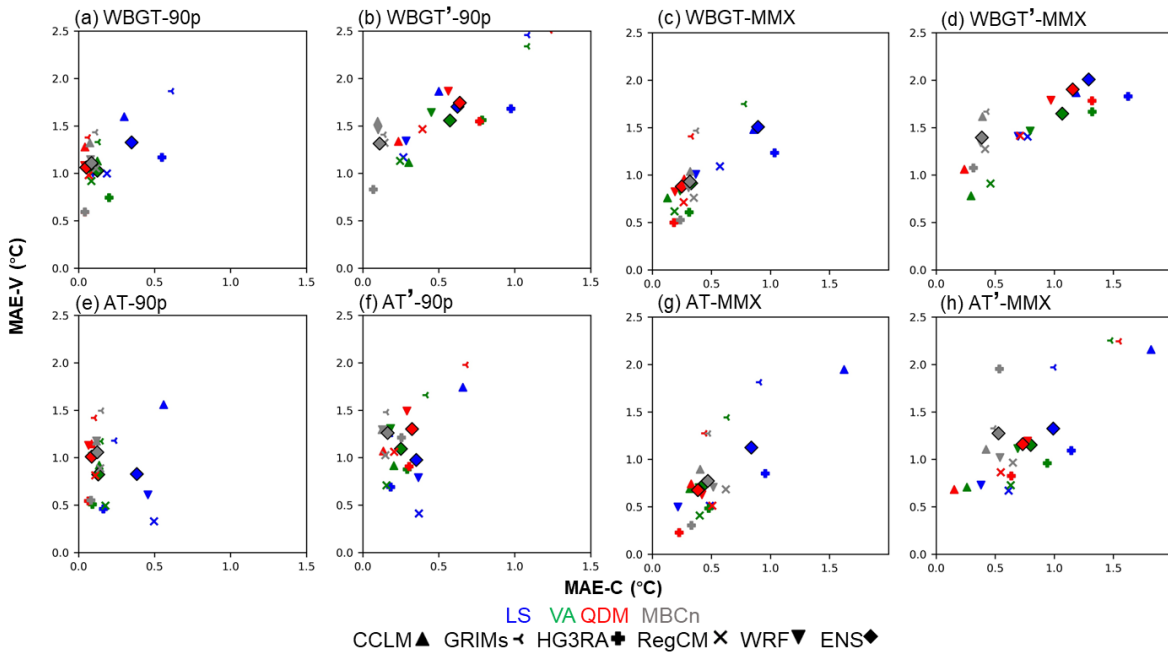



Figure 4: The MAE over South Korea (land only) for the calibration period (1979-1996, x-axis) and validation period (1997-2014, y-axis) in terms of the (a, b, e, f) 90p, and (c, d, g, h) MMX from (a, c) WBGT, (c, d) WBGT', (e, f) TW, and (g, h) TW'. The different colors stand for different BC methods, and the different markers stand for different RCMs. Some markers are overlapped.

Similar results are found in AT and AT' according to Fig. 4(e-h). However, for the indirect correction of AT', the weakness of QDM is less significant and the advantage of MBCn is also weakened compared to WBGT'. The ability to additionally

correct the multivariate dependency despite their individual distributions leads to a better result in the indirect correction of the heat-stress indices, which are functions of T and RH. In this case, since AT is more reliant than WBGT on T, the effect of correcting T-RH interdependence is less critical to its correction outcome. On the other hand, because T and RH both play important roles in WBGT, multivariate BC is more likely to demonstrate its importance in this case. Not surprisingly, the performances of different BC methods are retained in the reverse test, although with different magnitudes of MAE (Figure S1). MBCn shows an even better performance in this case, outperforming all other methods despite the heat indices and matrices considered.

To assess the quantitative differences in the marginal distributions corrected by different BC methods, Fig. 5 (a , b, e, f) presents the maximum differences calculated from the Kolmogorov–Smirnov (K-S) test (Eq. S3) between the observed (i.e., ERA5) and bias-corrected empirical cumulative distribution functions (CDFs). A smaller value stands for a better correction output. For the direct correction, QDM and MBCn show better performances than LS and VA across all the indices and matrices considered. However, for indirect correction, MBCn shows its unique advantage in the multivariate index depending unequally on the components (i.e., WBGT' in this study), that it can provide a similarly good result in either the direct or indirect correction. In this aspect, QDM shows the largest difference between the direct and indirect applications. Fig. 5 (c, d, g, f) is the D value calculated between outputs from direct and indirect applications of the same BC method, and a smaller value stands for more similar outputs. It clearly indicates a higher similarity seen in the multivariate method than the univariate methods in WBGT, as MBCn successfully retains the intervariable dependence during the correction procedure.

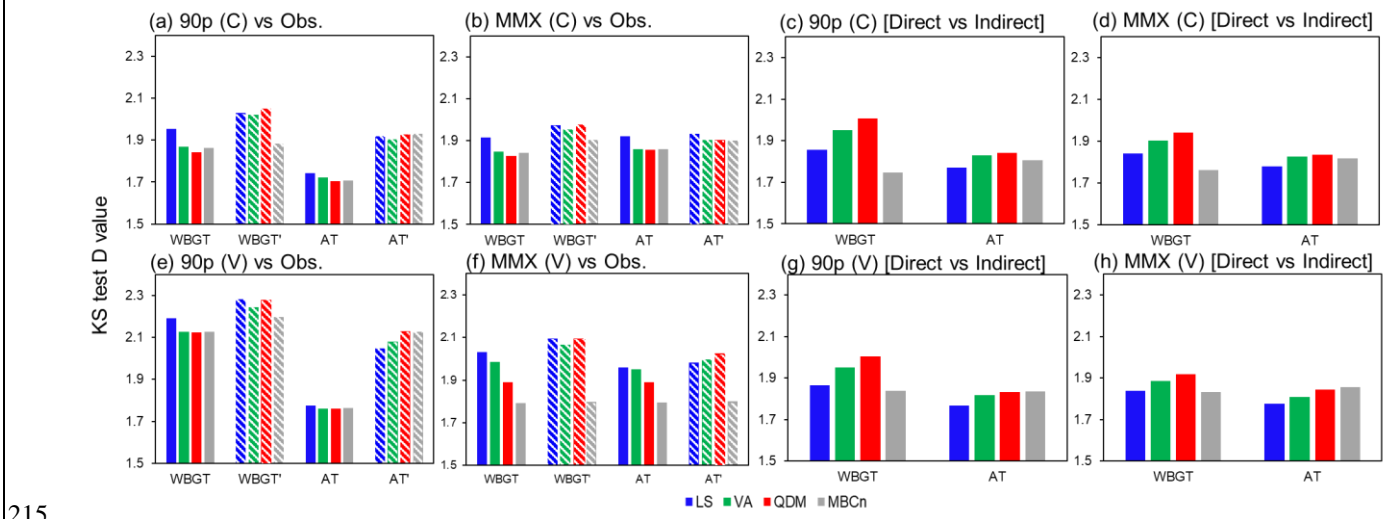



Figure 5: K-S test D value between bias-corrected output and observation for (a,e) 90p, and (b,f) MMX, and between direct and indirect corrected output for (c,g) 90p and (d,h) MMX. The D value is ensemble mean of 5 RCMs averaged over South Korea (land only). The different colors stand for different BC methods. The first row is for the Calibration period (C) and the second is for the Validation period (V). In (a, b, e, f), the solid and patterned fill is for the direct and indirect BC, respectively.

Figure 5 6 investigates the spatial distribution of bias in the QDM and MBCn corrections, using the 90p as an example for WBGT and AT. A similar pattern can also be seen in the case of MMX (Fig. 6S3). For the calibration period, the biases are well eliminated to less than 0.5 °C with only indirect correction by QDM shows showing a noticeable warm bias and the bias increases from the northwest to thein the southeast part. Specifically, the resultant bias magnitude from indirect QDM correction is even larger than in ORI (Fig. 3b) over southeastern Korea. The spatial pattern of the warm bias persists in the validation period, although with greater magnitude, which can be explained by the different bias magnitudes for the two periods in ORI simulations. This behavior is seen in both WBGT and AT but more strongly in WBGT which is more affected by the T-RH dependency. The overall cold bias in the model simulations during the calibration period must result in a positive correction coefficient (i.e., towards a warmer condition). However, as discussed above, a reduced cold bias in the RCMs is seen in the validation period because of overestimated warming in the models. Such a "trimmed" bias in the validation period may be over-corrected by the correction coefficient derived from the calibration period, even causing a larger bias than in ORI over the eastern part of the country with a warm bias in validation period. The results from the reverse test (Fig. 7 and Fig. S4) can further prove the impact of non-stationary bias on the result. In this case, the validation period of 1979-1996 retains a cold bias after BC for the reason that the correction coefficient derived in 1997-2014 is not large enough to compensate for its negative bias. Again, this warns us of the careful interpretation of bias-corrected climate data, especially in the context of future warming projections.

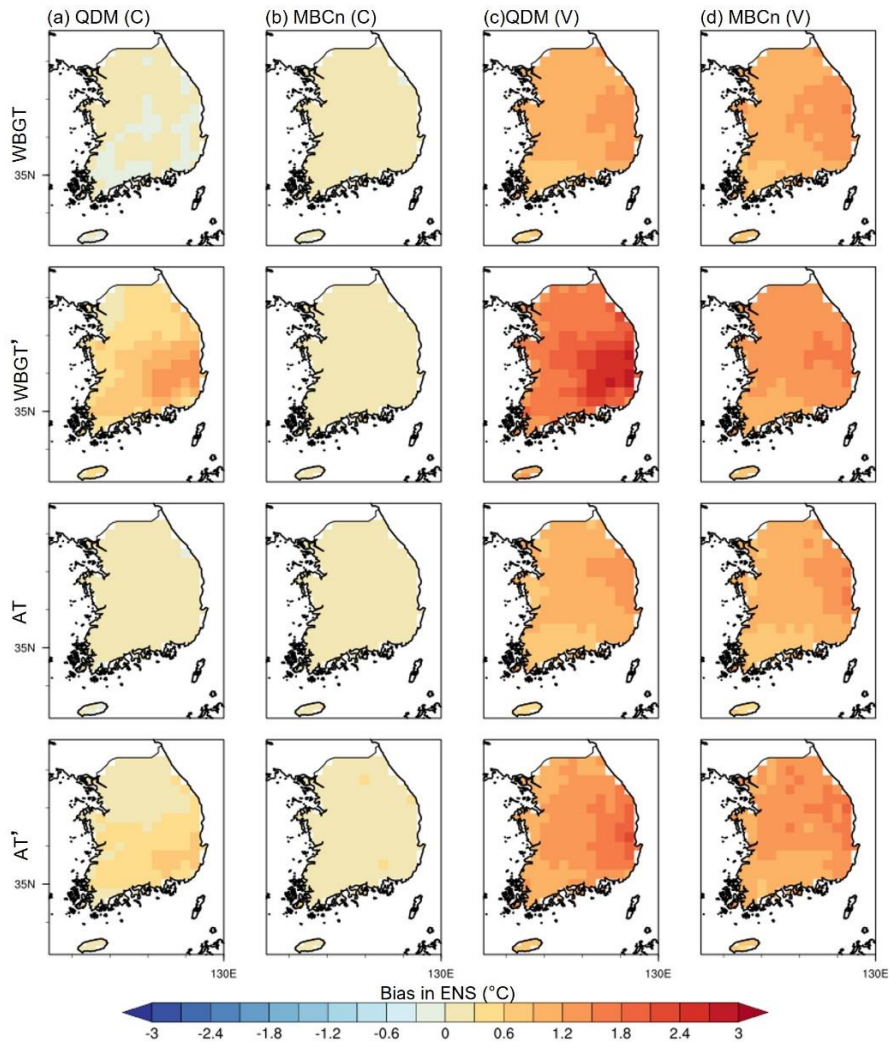



Figure 56: Spatial maps of the bias in the 90p during the calibration period (C) and validation (V) period corrected by QDM and MBCn in ENS. The first and third rows are the directedly corrected WBGT and AT. The second and fourth rows are the WBGT' and AT' calculated by the corrected T and RH.

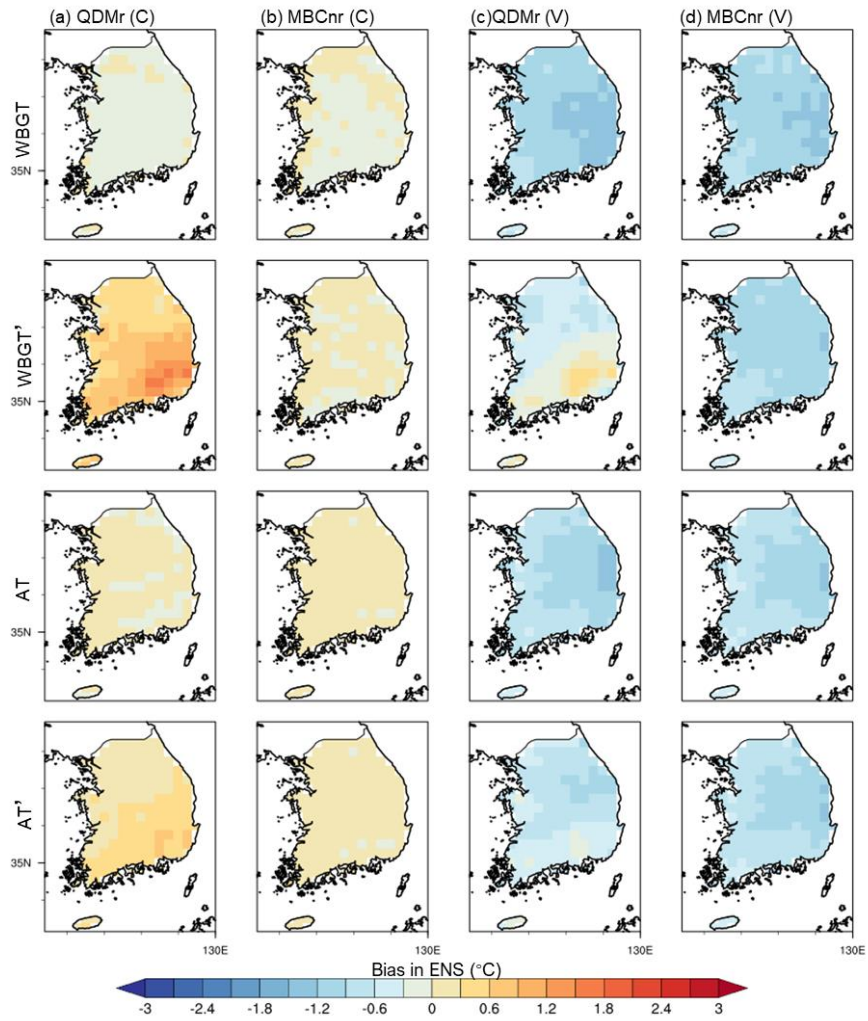
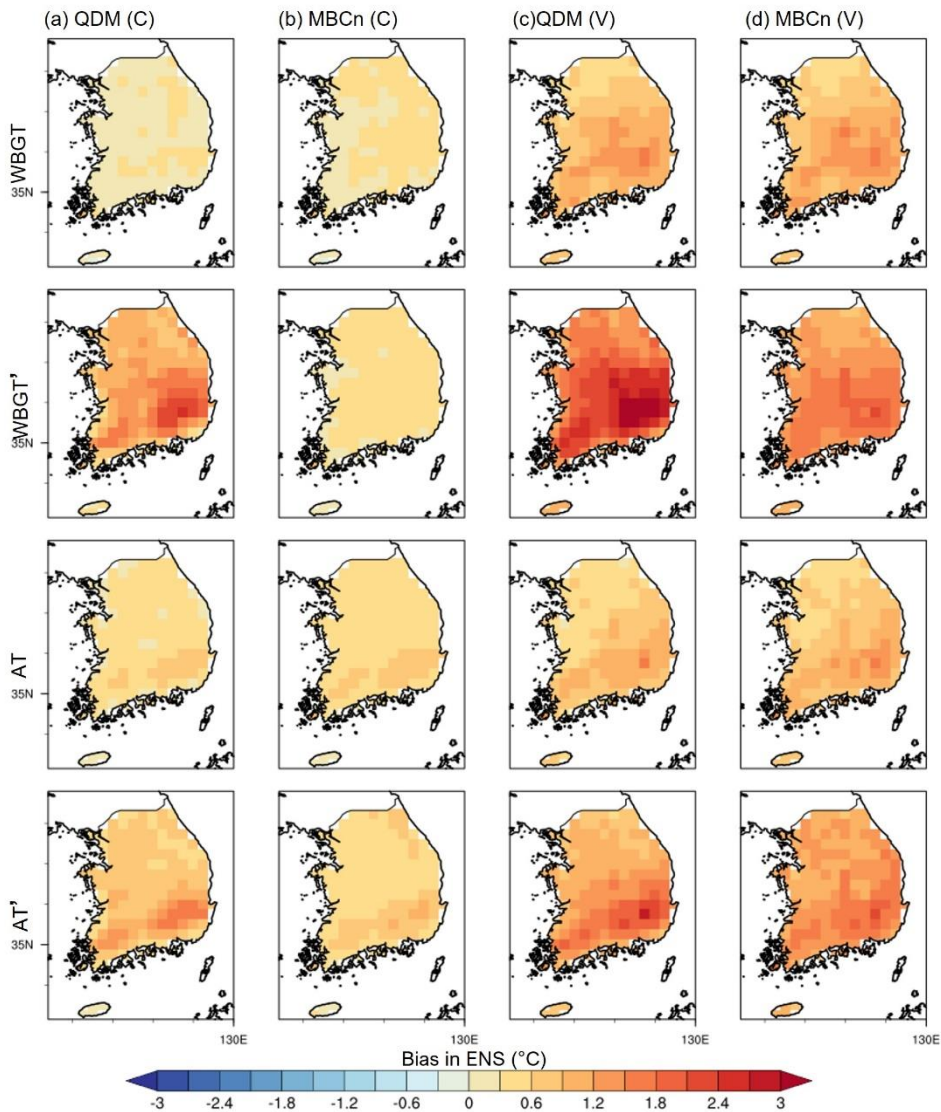



Figure 7: Same as Fig. 6 but for the reverse test.

On the other hand, the spatial maps of bias also clearly demonstrate the superiority of MBCn for the indirect correction of the heat-stress indices over the entire domain in both the calibration and validation periods. Since the heat-stress indices are functions of T and RH, we investigate the T vs. RH Spearman's rank correlation at a confidence interval of 99% (p value < 0.01) using daily T and RH at the time when the heat-stress indices reach their daily maxima (Fig. 78). ERA5 shows a negative correlation ranging from -0.4 to -0.6, that gradually increases from northeast to southwest. Comparatively, ORI has a significantly weaker negative correlation and does not adequately reflect the spatial gradient. The correction with QDM, even with the good outcome in the direct correction of the heat-stress indices, cannot properly present the T-RH relation. In fact, it even further weakens their correlation during the calibration period. On the contrary, MBCn calibrates the multivariate

dependency according to the observed correlation pattern, which explains why it significantly improves the correction of WBGT' and AT'. The correlation derived from the calibration period is also passed to the validation period by MBCn, which in this case shows no significant change between the two historical periods according to ERA5.

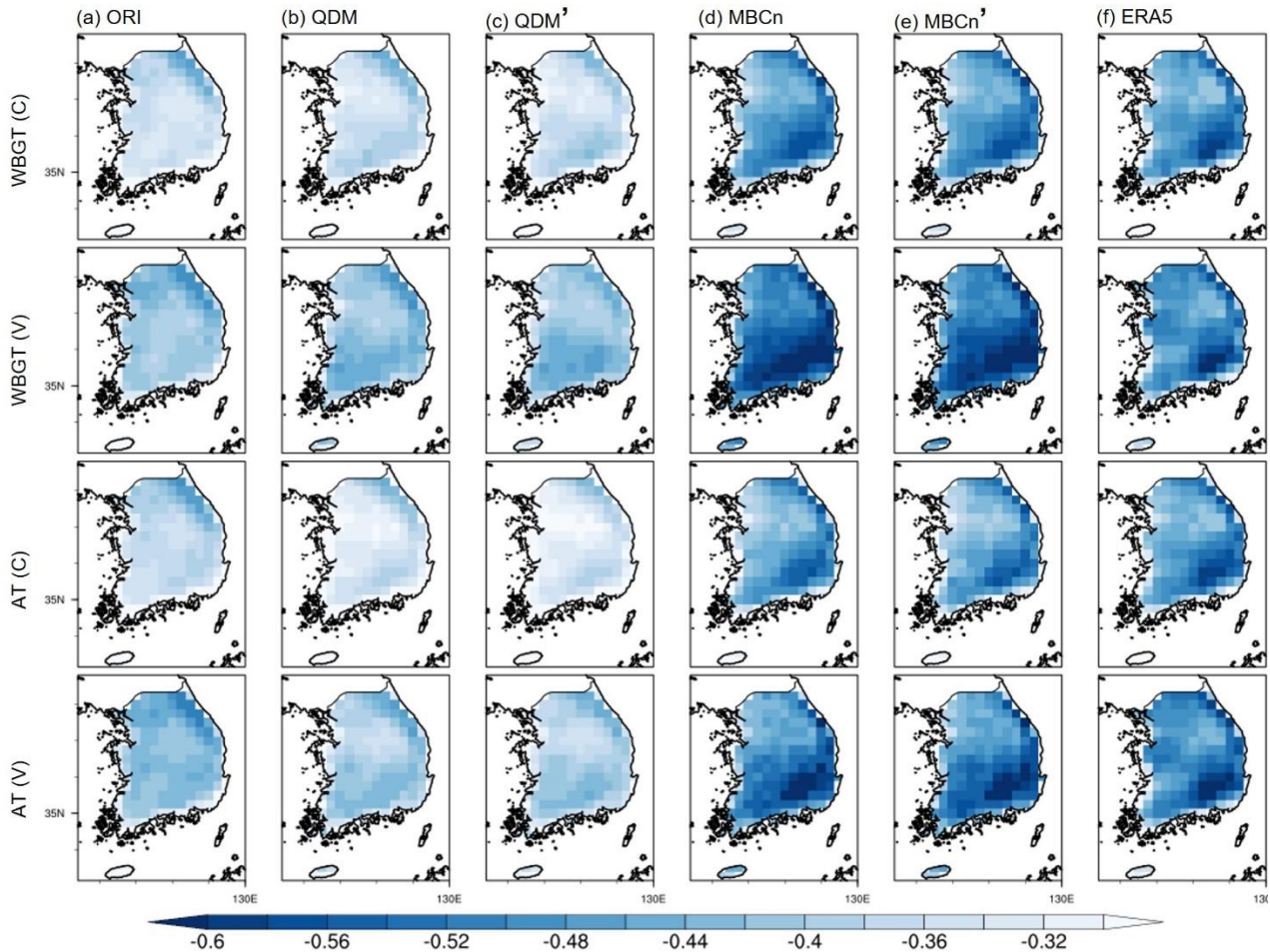



Figure 78: Spatial patterns of T vs. RH Spearman's rank correlation $(\alpha = 0.01)$ computed in each grid cell during the calibration (rows 1 and 3) and validation (rows 2 and 4) period. Column (a) shows the results from ORI simulations. Columns (b) and (d) are the heat-stress indices directly corrected by QDM and MBCn. Columns (c) and (e) are the heat-stress indices indirectly corrected by QDM and MBCn. Column (f) is from ERA5.

4 Discussion and Conclusion

Previous studies have challenged the applicability of univariate BC for adjusting individual components of multivariate hazard indicators and proved the benefit of multivariate BC in compound event evaluations (François et al., 2020; Zscheischler et al., 2019). Our study also demonstrates MBCn's advantage in correcting the interdependence of the relevant variables, which results in a substantial improvement in the indirect BC of heat-stress indices. Such an advantage is more prominent for the

index relying more equally on the composing variables (e.g., WBGT), which was also pointed out by Zscheischler et al. (2019) . However, to the best of our knowledge, no study has been conducted to compare the multivariate BC methods with the direct application of univariate BC on multivariate indices. Our results show that QDM applied directly to the multivariate indices can provide a similar result as MBCn in heat-stress assessments, while MBCn additionally provides a more reasonable underlying inter-variable dependence. In this regard, if only considering heat-stress indices, the more computationally-efficient QDM direct correction may be sufficient for the impact assessment. ~~On the other hand~~However, if the relationship between T/RH and the heat-related impacts is of interest, the multivariate BC is suggested for maintaining the physical linkage of the variables.

On the other hand, regarding the study of heat stress under future warming that is not evaluated in this study, more aspects should be considered. This study uses historical climate simulations comprising non-stationarity ~~Our results comparing the validation and calibration period~~ combined with two "jack-knifing" split-sample tests. It is found that ~~may offer certain insights for future studies:~~ the non-stationarity of bias in the modeled heat-stress indices, as combined effects of internal climate variability and climate model sensitivity, can significantly affect the BC output. Teutschbein & Seibert (2012) once suggested that the more advanced correction methods (e.g., QM) are more robust to a non-stationary bias compared to the simpler ones (e.g., LS), but our result shows no significant difference. In fact, lying under the fundamental assumption of stationary bias, current BC approaches may not be able to provide a suitable solution to this issue. Therefore, ~~For an appropriate impact assessment,~~ a case-by-case evaluation of BC approaches for ~~certain a~~a certain climate model and study area, as well as a clear understanding of the relevant processes including the uncertainties underlying original model data, is required for reliable data post-processing using BC methods.~~.~~

Meanwhile, for the continuous development in future projections of multivariate heat-stress indices, there are also potential problems worth investigating. For example, we may need to consider if there is any substantial change in the modeled multivariate dependence structure, which is also highly likely under global warming (Hao et al., 2019; Singh et al., 2021). Although both QDM and MBCn are supposed to preserve the simulated trend in the corrected variables, MBCn, as well as other multivariate BC methods ~~(e.g., R²D² developed by , TSQM developed by)~~, does not consider the change in the multivariate relationships. In this regard, the direct correction of QDM may outperform MBCn. However, as direct correction of QDM may discard the physical consistency in the input variables, both in terms of the variable representation and the projected change, it can hide the compensating bias (Schwingshackl et al., 2021) and thus introduce additional uncertainty in climate change signal (Casanueva et al., 2018) in the multivariate heat-stress indices. To solve these problems, a deeper understanding and continuous enhancement in climate models, particularly for the uncertainty and credibility of projections, may be prerequisites for better evaluation and application of the statistical procedures (i.e., BC approaches).

Code and Data Availability

Near-surface temperature and relative humidity data from the CORDEX-East domain downscaling product used in this study is archived in the institutional repository at https://doi.org/10.14711/dataset/GTXJVQ. ERA5 hourly data on single levels is downloaded from the Climate Data Store via https://doi.org/10.24381/cds.adbb2d47 (Hersbach et al., 2018). R package

300 "qmap" (https://CRAN.R-project.org/package=qmap, (Gudmundsson, 2016) is used for applying EQM and QDM, and R package "MBC" (https://CRAN.R-project.org/package=MBC, Cannon, 2020) is used for applying MBCn. Climate Data Operators (CDO) open-source package is used for 1) computations in LS and VA; 2) temporal and spatial correlation; and 3) statistical analysis.

Author Contribution

305 Im ES and Min SK conceptualized the study and acquired the funding. Qiu LY was responsible for investigation, formal analysis, methodology, software, visualization. Im ES supervised all Qiu LY's work and provided investigations. Qiu LY and Im ES wrote the original draft and Im ES and Min SK reviewed and edited it. Qiu L, Min SK, Kim YH, Cha DH, Shin SW, Ahn JB, Chang EC, and Byun YH created the data used in the study.

Competing Interests

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

Acknowledgments

This study was supported by the Korea Meteorological Administration Research and Development Program under Grant No. KMI2021-00912.

References

315 ACSM. (1984). Prevention of thermal injuries during distance running: Postition stand. *The Medical Journal of Australia*, *141*, 876–879. https://doi.org/10.5694/j.1326-5377.1984.tb132981.x

Bárdossy, A., & Pegram, G. (2012). Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resources Research*, *48*(9). https://doi.org/https://doi.org/10.1029/2011WR011524

320 Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, *50*(1), 31–49. https://doi.org/10.1007/s00382-017-3580-6

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Casanueva, A., Bedia, J., Herrera, S., Fernández, J., & Gutiérrez, J. M. (2018). Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool. *Climatic Change*, *147*(3), 411–425. https://doi.org/10.1007/s10584-018-2167-5

Casanueva, A., Kotlarski, S., Herrera, S., Fischer, A. M., Kjellstrom, T., & Schwierz, C. (2019). Climate projections of a multivariate heat stress index: the role of downscaling and bias correction. *Geoscientific Model Development*, *12*(8), 3419–3438. https://doi.org/10.5194/gmd-12-3419-2019

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface – hydrology model with the Penn State – NCAR MM5 Modeling System. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2

Coffel, E. D., Horton, R. M., & de Sherbinin, A. (2017). Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century. *Environmental Research Letters*, *13*(1), 14001. https://doi.org/10.1088/1748-9326/aaa00e

Dieng, D., Cannon, A. J., Laux, P., Hald, C., Adeyeri, O., Rahimi, J., Srivastava, A. K., Mbaye, M. L., & Kunstmann, H. (2022). Multivariate Bias-Correction of High-Resolution Regional Climate Change Simulations for West Africa: Performance and Climate Change Implications. *Journal of Geophysical Research: Atmospheres*, *127*(5), e2021JD034836. https://doi.org/https://doi.org/10.1029/2021JD034836

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS Opinions "Should we apply bias correction to global and regional climate model data?" *Hydrology and Earth System Sciences*, *16*(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth Syst. Dynam.*, *11*(2), 537–562. https://doi.org/10.5194/esd-11-537-2020

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences*, *16*(9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Guo, Q., Chen, J., Zhang, X., Shen, M., Chen, H., & Guo, S. (2019). A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Climate Dynamics*, *53*(5), 3603–3623. https://doi.org/10.1007/s00382-019-04729-w

Hao, Z., Phillips, T. J., Hao, F., & Wu, X. (2019). Changes in the dependence between global precipitation and temperature from observations and model simulations. *International Journal of Climatology*, *39*(12), 4895–4906. https://doi.org/10.1002/joc.6111

Hersbach, H., Bell, B., Berrisford, P., & Biavati, G. (2018). *ERA5 hourly data on single levels from 1979 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 10.24381/cds.adbb2d47

Kang, S., Pal, J. S., & Eltahir, E. A. B. (2019). Future heat stress during Muslim Pilgrimage (Hajj) projected to exceed "Extreme Danger" levels. *Geophysical Research Letters*, *46*(16), 10094–10100.

Kim, G., Cha, D.-H., Lee, G., Park, C., Jin, C.-S., Lee, D.-K., Suh, M.-S., Ahn, J.-B., Min, S.-K., & Kim, J. (2020). Projection of future precipitation change over South Korea by regional climate models and bias correction methods. *Theoretical and Applied Climatology*, *141*(3), 1415–1429. https://doi.org/10.1007/s00704-020-03282-5

Kim, K. B., Kwon, H.-H., & Han, D. (2022). Intercomparison of joint bias correction methods for precipitation and flow from a hydrological perspective. *Journal of Hydrology*, *604*, 127261. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.127261

Kim, M.-K., Yu, D.-G., Oh, J.-S., Byun, Y.-H., Boo, K.-O., Chung, I.-U., Park, J.-S., Park, D.-S. R., Min, S.-K., & Sung, H. M. (2020). Performance Evaluation of CMIP5 and CMIP6 Models on Heatwaves in Korea and Associated Teleconnection Patterns. *Journal of Geophysical Research: Atmospheres*, *125*(23), e2020JD032583. https://doi.org/https://doi.org/10.1029/2020JD032583

Kim, Y.-T., Kwon, H.-H., Lima, C., & Sharma, A. (2021). A Novel Spatial Downscaling Approach for Climate Change Assessment in Regions With Sparse Ground Data Networks. *Geophysical Research Letters*, *48*(22), e2021GL095729. https://doi.org/https://doi.org/10.1029/2021GL095729

Lee, M., Cha, D.-H., Suh, M.-S., Chang, E.-C., Ahn, J.-B., Min, S.-K., & Byun, Y.-H. (2020). Comparison of tropical cyclone activities over the western North Pacific in CORDEX-East Asia phase I and II experiments. *Journal of Climate*, 1–36. https://doi.org/10.1175/JCLI-D-19-1014.1

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., & Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*(11), 764–773. https://doi.org/10.1038/nclimate3418

Masaki, Y., Hanasaki, N., Takahashi, K., & Hijioka, Y. (2015). Propagation of biases in humidity in the estimation of global irrigation water. *Earth System Dynamics*, *6*(2), 461–484. https://doi.org/10.5194/esd-6-461-2015

Mehrotra, R., & Sharma, A. (2015). Correcting for systematic biases in multiple raw GCM variables across a range of timescales. *Journal of Hydrology*, *520*, 214–223. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.11.037

Mehrotra, R., & Sharma, A. (2016). A Multivariate Quantile-Matching Bias Correction Approach with Auto- and Cross-Dependence across Multiple Time Scales: Implications for Downscaling. *Journal of Climate*, *29*(10), 3519–3539. https://doi.org/10.1175/JCLI-D-15-0356.1

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., & Cannon, A. J. (2019). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, *23*(3), 1339–1354. https://doi.org/10.5194/hess-23-1339-2019

Qiu, L., Im, E.-S., Hur, J., & Shim, K.-M. (2020). Added value of very high resolution climate simulations over South Korea using WRF modeling system. *Climate Dynamics*, *54*(1), 173–189. https://doi.org/10.1007/s00382-019-04992-x

Qiu, L., Kim, J.-B., Kim, S.-H., Choi, Y.-W., Im, E.-S., & Bae, D.-H. (2022). Reduction of the uncertainties in the hydrological projections in Korean river basins using dynamically downscaled climate projections. *Climate Dynamics*, *59*(7–8), 2151–2167. https://doi.org/10.1007/s00382-022-06201-8

Robin, Y., Vrac, M., Naveau, P., & Yiou, P. (2019). Multivariate stochastic bias corrections with optimal transport. *Hydrology and Earth System Sciences*, *23*(2), 773–786. https://doi.org/10.5194/hess-23-773-2019

Rocheta, E., Evans, J. P., & Sharma, A. (2017). Can Bias Correction of Regional Climate Model Lateral Boundary Conditions Improve Low-Frequency Rainfall Variability? *Journal of Climate*, *30*(24), 9785–9806. https://doi.org/10.1175/JCLI-D-16-0654.1

Schwingshackl, C., Sillmann, J., Vicedo-Cabrera, A. M., Sandstad, M., & Aunan, K. (2021). Hear stress indicators in CMIP6: Estimating future trends and exceedances of impact-relevant thresholds. *Earth's Future*, *9*(3), e2020EF001885. https://doi.org/10.1029/2020EF001885

Sherwood, S. C. (2018). How Important Is Humidity in Heat Stress? *Journal of Geophysical Research: Atmospheres*, *123*(21), 11, 808–811, 810. https://doi.org/https://doi.org/10.1029/2018JD028969

Singh, H., Najafi, M. R., & Cannon, A. J. (2021). Characterizing non-stationary compound extreme events in a changing climate based on large-ensemble climate simulations. *Climate Dynamics*, *56*(5), 1389–1405. https://doi.org/10.1007/s00382-020-05538-2

Steadman, R. G. (1984). A Universal Scale of Apparent Temperature. *Journal of Applied Meteorology and Climatology*, *23*(12), 1674–1687. https://doi.org/10.1175/1520-0450(1984)023<1674:AUSOAT>2.0.CO;2

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., & Demaria, E. M. C. (2017). Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences*, *21*(6), 2649–2666. https://doi.org/10.5194/hess-21-2649-2017

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456–457*, 12–29. https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.05.052

Vrac, M. (2018). Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R^2D^2) bias correction. *Hydrology and Earth System Sciences*, *22*(6), 3175–3196. https://doi.org/10.5194/hess-22-3175-2018

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/https://doi.org/10.1029/2019GL085782

Zscheischler, J., Fischer, E. M., & Lange, S. (2019). The effect of univariate bias adjustment on multivariate hazard estimates. *Earth System Dynamics*, *10*(1), 31–43. https://doi.org/10.5194/esd-10-31-2019