



**HAL**  
open science

# Should Multivariate Bias Corrections of Climate Simulations Account for Changes of Rank Correlation Over Time?

M. Vrac, S. Thao, Pascal Yiou

► **To cite this version:**

M. Vrac, S. Thao, Pascal Yiou. Should Multivariate Bias Corrections of Climate Simulations Account for Changes of Rank Correlation Over Time?. *Journal of Geophysical Research: Atmospheres*, 2022, 127 (14), 10.1029/2022jd036562 . hal-03738510

**HAL Id: hal-03738510**

**<https://hal.science/hal-03738510>**

Submitted on 26 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.




L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## RESEARCH ARTICLE

10.1029/2022JD036562

# Should Multivariate Bias Corrections of Climate Simulations Account for Changes of Rank Correlation Over Time?

M. Vrac<sup>1</sup> , S. Thao<sup>1</sup> , and P. Yiou<sup>1</sup> 

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), Centre d'Etudes de Saclay, CEA/CNRS/UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

### Key Points:

- Multivariate bias correction methods can account for changes in correlations in climate simulations or can consider that they are stationary
- A perfect model experiment is set up to evaluate consequences of these choices in terms of temperature versus precipitation rank correlations
- Results on two ensembles show that both approaches are meaningful, depending on the underlying confidence put on climate simulations

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

M. Vrac,  
mathieu.vrac@lsce.ipsl.fr

### Citation:

Vrac, M., Thao, S., & Yiou, P. (2022). Should multivariate bias corrections of climate simulations account for changes of rank correlation over time? *Journal of Geophysical Research: Atmospheres*, 127, e2022JD036562. <https://doi.org/10.1029/2022JD036562>

Received 27 JAN 2022  
Accepted 11 JUL 2022

### Author Contributions:

**Conceptualization:** M. Vrac  
**Data curation:** S. Thao  
**Formal analysis:** M. Vrac  
**Investigation:** M. Vrac, S. Thao, P. Yiou  
**Methodology:** M. Vrac, S. Thao, P. Yiou  
**Resources:** S. Thao  
**Software:** M. Vrac  
**Validation:** M. Vrac  
**Visualization:** M. Vrac  
**Writing – original draft:** M. Vrac  
**Writing – review & editing:** M. Vrac, S. Thao, P. Yiou

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Abstract** Inter-variable dependencies are key properties to characterize many climate phenomena—such as compound events—and their future changes. Yet, climate simulations often have statistical biases. Hence, univariate (1dBC) and multivariate bias correction (MBC) methods are regularly applied. Inter-variable properties (e.g., correlations) can be altered by BC corrections. Then, it is necessary to assess how underlying assumptions of BC methods on climate change affect the adjustments. This can lead to better choices of BC methods. Here, we investigate whether an MBC method should try reproducing, preserving or modifying the changes in rank correlations between daily temperature and precipitation over Europe. An original “perfect model experiment” is set up and applied to two different climate simulation ensembles over 2001–2100: 40 runs from the Community Earth System Model (CESM) global climate model and 11 runs from the Coupled Models Intercomparison Project (CMIP6) exercise. The results highlight that, within the multi-run single Global Climate Models ensemble (CESM), accounting for correlation changes brings valuable information for long-term projections but that a stationarity assumption provides less biased correlations, up to medium-term projections (2060). For the multi-model ensemble (CMIP6), the non-stationarity assumption provides larger biases than the stationarity approach, up to the end of the century. Not correcting the model rank correlations (1dBC) provides the worst results. Whenever an ensemble is available, the best results come from accounting for the “robust” part of the change signal (i.e., average change from different runs). This pleads for using ensembles and their robust information, in order to perform robust bias corrections.

## 1. Introduction

To investigate the manifold impacts of future climate changes, numerical simulations from Global and Regional Climate Models (GCM and RCM, respectively) remain essential (IPCC WGI, 2021). However, it is now well-known that these simulations can have statistical biases with respect to observational references (e.g., from measurements at weather stations, or reanalyses). Therefore, using such simulations directly as input in an impact model (e.g., in hydrology or agronomy) is not always relevant without “correcting” those biases (Doblas-Reyes et al., 2021). Various “bias correction” (BC) methods have thus been developed and extensively applied over the last few decades. Such methods transform the initial simulations to make the corrected data more similar to a reference data set in terms of specific statistical criteria such as mean, standard deviation or in terms of probability distribution. The transformation is defined based on calibration data—usually corresponding to references and simulations over a historical period—and is supposed valid for a different period (e.g., the future). It can then be applied to climate projections for a period of interest. Due to its coding facility, its speed of calculation and the fact that it globally preserves the main trends of the simulations (e.g., A. J. Cannon et al., 2015; Hempel et al., 2013), the “quantile-mapping” approach is certainly the most widely used BC method and has multiple variants (e.g., Déqué, 2007; Haddad & Rosenfeld, 1997; Kallache et al., 2011; Volosciuk et al., 2017; Vrac et al., 2012, 2016, among many others). However, it only corrects one variable at a time for one location at a time. This means that quantile-mapping only corrects the marginal distributions of the climate variables but not their copula function, characterizing their dependence structure. Therefore, the inter-variable properties—such as the rank correlation—after such a univariate correction are the same as in the initial (raw) simulations. Hence, if the dependence structure in the model is biased, the corrections will inherit this biased dependence (see e.g., Vrac, 2018). To overcome this issue and correct the inter-variable and/or spatial dependencies of the simulations in addition to their marginal distributions, multivariate bias correction (MBC) methods were designed. Three MBC categories can be considered, depending on how the corrections are made (François et al., 2020): the “successive conditional” methods where univariate BC is applied conditionally on previously corrected other

variables (e.g., Dekens et al., 2017; Piani & Haerter, 2012); the “marginal/dependence” methods correcting separately the marginals and the dependence before combining them (e.g., A. Cannon, 2017; Vrac & Thao, 2020; François et al., 2021); and the “all-in-one” methods correcting marginals and dependencies altogether (e.g., Robin et al., 2019; Robin & Vrac, 2021).

In most BC methods, it is common to impose or verify that the climate evolution—from one period to another—visible in the raw simulations (e.g., changes in mean temperature, or in its statistical moments, or in rainfall occurrence probabilities) are mostly kept by the corrected data (e.g., A. J. Cannon et al., 2015; Hempel et al., 2013): Even if climate simulations might have biases, the changes in the main properties are supposed to be driven by physical processes that are relevant and thus provide key information on climate changes. The climate change signal of single variables (such as temperature or precipitation separately) has been extensively investigated (e.g., Kendon et al., 2008; Matte et al., 2019, among many others). Climate change signal regarding multivariate properties (e.g., correlation or dependence between variables) is much less known, while it can be key for many studies. Indeed, it is an essential signal from the climate models, whose the robustness might have significant implications on the conclusions from impact studies.

Actually, multivariate properties and their potential future changes are closely related to “compound events,” a booming field of research (e.g., Ridder et al., 2021; Sadegh et al., 2018; Singh et al., 2021; Zscheischler et al., 2020, 2021, among many others). Such events are characterized by the occurrence of multiple meteorological events—either simultaneously or successively, spatially or with multiple variables, or both—whose impacts are stronger than those of the separate events (e.g., Zscheischler et al., 2020). The notions of dependencies and correlations between the events and between the variables are thus key in this context, and their climate change signal must then be investigated to understand the potential future changes in compound events (Vrac et al., 2021). Indeed, Hillier et al. (2020) showed that accounting for multivariate dependencies can increase or decrease the hazards of compound events. Recently, Abatzoglou et al. (2020) showed that, based on TerraClimate monthly reanalysis data (Abatzoglou et al., 2018), changes in multivariate climate departures have generally outpaced univariate departures in the recent decades. Moreover, Ridder et al. (2021) found that some CMIP6 models (not all) can be used to examine some compound events. Yet, Vrac et al. (2021) showed that climate models are not able to reproduce inter-variable temperature-precipitation rank correlations visible over Europe in the ERA5 reanalysis data (Hersbach et al., 2020), nor their changes in time. Nevertheless, Vrac et al. (2021) also showed that both multi-model (CMIP6) and multi-run (CESM) ensembles project significant changes of inter-variable rank correlations up to the end of the 21st century. This is in agreement with results from (Singh et al., 2021), who used a large ensemble of climate simulations and found that there is a strong non-stationarity in the dependence structure of temperature and precipitation under climate change that can play a significant role in future compound extremes. However, as these changes might show a strong variability among ensemble members and models, different from one season to another (Vrac et al., 2021), it is legitimate to wonder how this variability needs to be accounted for in practical applications and uses of climate simulations such as via MBC.

More precisely, the robustness of the multivariate properties from climate simulations can have major implications on the way MBC methods must be designed and applied. If the signal of change in multivariate dependence properties (e.g., in terms of rank correlations) in the raw simulations is trustworthy, MBCs have to respect it and generate multivariate corrected data with equivalent changes. If the change in dependence properties provided by the raw simulations is not robust enough, MBC data should better take a stationarity assumption regarding the dependence structure: the multivariate properties, such as the rank correlations, should not evolve and stay similar to the reference (and then not reproduce the changes in the raw simulations dependence) to avoid providing non-reliable multivariate projections. Either explicitly or implicitly, all MBC methods already incorporate one of these two assumptions. For example, the evolution of the multivariate dependencies is (mostly) reproduced by methods such as the “MBCn” (A. Cannon, 2017) or “dynamical Optimal Transport Correction” (“dOTC,” Robin et al., 2019) methods; while the assumption of stationary rank dependence features is made in the “Rank Resampling for Distribution and Dependency” (“R2D2,” Vrac, 2018) correction method and its extension (“R2D2v2,” Vrac & Thao, 2020). Knowing the robustness of the changes in dependencies simulated by climate models is thus also crucial to choose the proper assumption (stationarity or non-stationarity) regarding changes in multivariate properties, and therefore the proper MBC methods to use in climate change context. A follow up question is to know how these stationarity or non-stationarity assumptions compare to the approach consisting in keeping both the raw rank correlation values and changes given by the climate simulations. Such an approach is typically what

is done when a univariate BC method is applied. Indeed, a 1d-BC method does not adjust the copula function (i.e., function containing the dependence linking statistically the different variables of the climate simulations), and thus does not modify the rank correlations of the climate model (Vrac, 2018). Moreover, when considering an ensemble of climate simulations, it is common to average the changes in univariate properties of the simulations—via mean-model means or multi-run means—to get the most robust part of the climate change signal (see, e.g., Tebaldi & Knutti, 2007; Knutti et al., 2010). Such an approach has not been tested so far for changes in rank correlations, while such changes in the dependence structure (e.g., between temperature and precipitation) can play a significant role in future compound extremes (Singh et al., 2021).

Therefore, in the present article, we investigate whether or not MBC methods should try reproducing, preserving or modifying the change in inter-variable correlations. To do so, we do not perform any (univariate or multivariate) BC per se. Indeed, no time series will be adjusted. Instead, the main idea is to rely only on estimations of the evolution of the rank correlations, as proxies of results given by MBCs: these estimations will depend on the chosen assumption for accounting of the rank correlation changes. Hence, we do not evaluate specific methods and their details, but rather their main underlying philosophies. To perform this evaluation, we propose a “perfect model experiment” (PME) setting, using model simulations as pseudo-observations (e.g., de Elía et al., 2002; Krinner & Flanner, 2018; Vrac et al., 2007). Although our PME setting can be applied to other couples of climate variables and other statistical properties, in the present article it will allow us to estimate the biases in terms of Temperature (T) versus Precipitation (PR) rank correlations brought by the four different assumptions or approaches:

1. “Non-stationarity” (NSt) assumption: MBC should preserve change in correlations, that is, this approach seeks to maintain the signal of the model;
2. “Stationarity” (St) assumption: MBC should have stationary correlations, which are estimated from the reference data over the calibration period;
3. “No correction” (Raw) assumption: BC should not modify correlation values and changes;
4. “Multi-Model Mean Climate Change” (CC) approach: MBC should account for multi-model mean correlation changes, that is, this approach constrains the corrections to reproduce an ensemble average change in correlations.

Hence, we will compare their robustness with respect to the rank correlation change signal provided by current state-of-art climate models. Note that, all along this article, the focus is put on the inter-variable rank correlation between T and PR and its changes, and not directly on the whole dependence structure that can be characterized by the copula function linking T and PR marginal distributions. The correspondence between copula and rank correlation is obviously not bijective. For examples: two different copulas can induce the same rank correlation; or complicated copula structures representing non-trivial inter-variable dependencies may result in close-to zero rank correlations. The later example is however obvious as, in statistics, it is well-known that if independent variables induce a zero correlation, the opposite is not true (i.e., a zero correlation does not imply that the two variables are independent, they can even have complex relationships). Yet, the Spearman rank correlation remains an easily interpretable and useable proxy of the global dependence between two variables of interest, and is thus employed in this study as the inter-variable property of level 0.

The rest of this article is structured as follows: Section 2 describes the climate simulations used in this study. Then, the various possible (stationarity or non-stationarity) assumptions of the dependence structure are detailed in Section 3, as well as our PME to test the consequences of these assumptions on the dependence structure. The results are given and described in Section 4. Finally, conclusions and discussions are provided in Section 5.

## 2. Data

Two ensembles of climate model simulations are considered. The first one is a multi-model ensemble constituted of 11 Global Climate Models (GCMs) contributing to the sixth exercise of the “Coupled Models Intercomparison Project” (CMIP6, Eyring et al., 2016). The list of the GCMs is provided in Table 1.

The second ensemble is constituted by 40 members (i.e., runs) from a single GCM, the “Community Earth System Model” (CESM, Kay et al., 2015) developed by the National Center for Atmospheric Research and the University Corporation for Atmospheric Research (NCAR/UCAR, USA), at approximately 1° horizontal resolution. The use

**Table 1**  
List of CMIP6 Simulations Used in This Study, Their Run, Approximate Horizontal Resolution and References

Simulation name	Run	Atmospheric resolution	Data reference
BCC-CSM2-MR	r1i1p1f1	100 km	Wu et al. (2018)
CanESM5	r10i1p1f1	500 km	Swart et al. (2019)
CNRM-CM6-1-HR	r1i1p1f2	100 km	Voltaire (2019)
CNRM-CM6-1	r1i1p1f2	250 km	Voltaire (2018)
CNRM-ESM2-1	r1i1p1f2	250 km	Seferian (2018)
INM-CM4-8	r1i1p1f1	100 km	Volodin et al. (2019)
INM-CM5-0	r1i1p1f1	100 km	Volodin et al. (2019)
IPSL-CM6A-LR	r14i1p1f1	250 km	Boucher et al. (2018)
MIROC6	r1i1p1f1	250 km	Shiogama et al. (2019)
MRI-ESM2-0	r1i1p1f1	100 km	Yukimoto et al. (2019)
UKESM1-0-LL	r1i1p1f2	250 km	Tang et al. (2019)

of these multi-two ensembles (model or multi-run) allows us to distinguish inter-model variability from internal variability in our investigations about changes in correlations.

From each of these two ensembles, daily surface temperature (hereafter T) and precipitation (PR) time series have been extracted over the western Europe domain, defined as  $[10^{\circ}\text{W}, 30^{\circ}\text{E}] \times [30^{\circ}\text{N}, 70^{\circ}\text{N}]$ . Historical runs were used over the 2001–2014 period and the shared socioeconomic pathways 585 (SSP585) scenario (Riahi et al., 2017) over the 2015–2100 period. Hence, for each run of each ensemble, we consider continuous simulations from 2001 to 2100, which we separate into five 20-year periods: 2001–2020 as historical period and 2021–2040, 2041–2060, 2061–2080, 2081–2100 for future periods. To ease comparisons, all temperature and precipitation fields have been regridded to a common spatial resolution of  $1^{\circ} \times 1^{\circ}$ .

### 3. “Perfect Model Experiment” Design and Evaluation Tools

To investigate how changes in T-PR rank correlation should be handled in a MBC context, a “perfect model with turning reference” (PMTR) experiment is now set up. The PMTR setting assumes that models are statistically

indistinguishable from the truth (i.e., real climate). This means that the truth and the models are supposed to be generated from the same underlying probability distribution (Ribes et al., 2016; Thao et al., 2021). In particular, the distribution of the differences between the truth and the models is supposed to be the same as the distribution of the differences among models. Within this paradigm, it is sensible to consider one model as a possible truth (i.e., reference) to evaluate the Stationarity (St), non-stationarity (NSt), no-correction (Raw) or multi-model mean climate change (CC) assumptions. Hence, our PMTR consists in taking one model (or run) as reference and evaluate the four assumptions on the other models (resp. runs) with respect to the reference one. The same procedure is repeated for another reference model until all models (resp. runs) have each served once as reference.

For the various tested assumptions, the following common notations are used: For a given grid cell of the domain, let  $\rho_{ref,i}$  be the T versus PR Spearman (rank) correlation of this grid cell from the reference (ref) model and for the period  $i$ , where  $i = 0$  indicates 2001–2020, up to  $i = 4$  corresponding to 2081–2100;  $\rho_{mod,i}$  is the Spearman correlation from another (i.e., non reference) model or run (belonging to the same ensemble). Our PME setting will allow us to compare the relevance of the different assumptions regarding the modification of the rank correlations  $\rho_{mod,i}$  of the models.

#### 3.1. No-Correction (Raw) Assumption

First, before applying a modification (correction) of the T versus PR rank correlation present in the climate simulations, it is legitimate to wonder whether these model correlations really have to be corrected. Indeed, when applying a univariate BC method, the copula function linking statistically the different variables of the climate simulations is mostly kept untouched, that is, uncorrected, and thus so is their rank correlation (Vrac, 2018). Therefore, the “no-correction” (hereafter “Raw”) assumption does not modify either the initial correlation value  $\rho_{mod,0}$  (i.e., over the calibration time period), neither the correlation values  $\rho_{mod,i}$  at any other (future) time period  $i$  (and thus neither the change in correlation from period 0 to period  $i$ ). This Raw approach can then serve as a proxy of the results given in terms of Spearman correlation by a traditional univariate BC, such as a quantile-mapping method (Déqué, 2007). Then, the Raw assumption is tested, for each 20-year period  $i$ , by computing  $B_{Raw}$ , the absolute bias of  $\rho_{mod,i}$  with respect to  $\rho_{ref,i}$ , the Spearman correlation of the reference model:

$$B_{Raw}(\text{mod}, i) = |\rho_{ref,i} - \rho_{mod,i}|. \quad (1)$$

Indeed, the rank correlation of the model to be evaluated is not modified at all and can thus be directly compared to the reference correlation.

### 3.2. Stationarity (St) Assumption

The stationarity assumption relies on the hypotheses that (a) a MBC method will correctly adjust the model rank correlation  $\rho_{\text{mod},0}$  over the calibration period (i.e.,  $\rho_{\text{mod},0}$  is corrected to  $\rho_{\text{ref},0}$ ) and (b) that the correlation  $\rho_{\text{ref},0}$  does not change for other time periods. Hence, through the “St” assumption, an estimation of the future rank correlation,  $\tilde{\rho}_{\text{mod},i}^{\text{St}}$ , over period  $i$ , is given by:

$$\tilde{\rho}_{\text{mod},i}^{\text{St}} = \rho_{\text{ref},0}. \quad (2)$$

As previously, a bias  $B_{\text{St}}$  is then defined to test the stationarity assumption, by computing the absolute bias of  $\tilde{\rho}_{\text{mod},i}^{\text{St}}$  with respect to  $\rho_{\text{ref},i}$ :

$$B_{\text{St}}(\text{mod}, i) = |\rho_{\text{ref},i} - \tilde{\rho}_{\text{mod},i}^{\text{St}}| = |\rho_{\text{ref},i} - \rho_{\text{ref},0}|. \quad (3)$$

### 3.3. Non-Stationarity (NSt) Assumption

The NSt assumption relies on the hypotheses (a) that an MBC method will correctly adjust the model rank correlation  $\rho_{\text{mod},0}$  to  $\rho_{\text{ref},0}$  over the calibration period and (b) that the corrected future rank correlation (hereafter  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$ ) evolves from  $\rho_{\text{ref},0}$  in a same manner as  $\rho_{\text{mod},i}$  evolves from  $\rho_{\text{mod},0}$ . Accounting for the changes in correlation,  $\Delta_{\rho,i}$ , from model mod means accounting for the difference between  $\rho_{\text{mod},i}$  (for  $i \geq 1$ ) and  $\rho_{\text{mod},0}$ , that is, its change from period  $i = 0$  (2001–2020) to the future period  $i \geq 1$ :

$$\Delta_{\rho,i} = \rho_{\text{mod},i} - \rho_{\text{mod},0}. \quad (4)$$

To get rid of the initial bias in  $\rho_{\text{mod},0}$ , this change must start from the “real” reference  $\rho_{\text{ref},0}$  value. It is then needed to calculate the correlation ( $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$ ) resulting from the appropriate evolution (i.e., from  $\rho_{\text{mod},0}$  to  $\rho_{\text{mod},i}$ ) but starting from  $\rho_{\text{ref},0}$  instead of  $\rho_{\text{mod},0}$ . However, simple additive or multiplicative factors applied to  $\rho_{\text{ref},0}$ —such as  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}} = \rho_{\text{ref},0} + (\rho_{\text{mod},i} - \rho_{\text{mod},0})$ , or  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}} = \rho_{\text{ref},0} \times (\rho_{\text{mod},i} / \rho_{\text{mod},0})$  respectively, or other similar transformations—could result in a  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$  value outside the  $[0, 1]$  interval and are thus not appropriate. To avoid this issue while accounting for the evolution, the  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$  correlation is defined as:

$$\tilde{\rho}_{\text{mod},i}^{\text{NSt}} = \begin{cases} \rho_{\text{ref},0} + (\Delta_{\rho,i} / (1 - \rho_{\text{mod},0})) \times (1 - \rho_{\text{ref},0}), & \text{if } \Delta_{\rho,i} > 0, \\ \rho_{\text{ref},0} + (\Delta_{\rho,i} / (\rho_{\text{mod},0} + 1)) \times (\rho_{\text{ref},0} + 1), & \text{if } \Delta_{\rho,i} \leq 0. \end{cases} \quad (5)$$

With this definition,  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$  is constrained to the  $[-1, 1]$  interval and is the result of the evolution of  $\rho_{\text{ref},0}$  in the same manner as  $\rho_{\text{mod},i}$  results from the evolution of  $\rho_{\text{mod},0}$ . Indeed, when applying this transformation to  $\rho_{\text{mod},0}$  (instead of  $\rho_{\text{ref},0}$ ), the result corresponds exactly to  $\rho_{\text{mod},i}$  as expected. Although Equation 5 might not be proper from a physical point-of-view, it is statistically more appropriate than the additive or multiplicative approach, since the resulting correlations evolve consistently with the model simulations while staying in the  $[-1, 1]$  range. Then, the non-stationarity assumption is tested, for each 20-year period  $i$ , by computing  $B_{\text{NSt}}$ , the absolute bias of  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$  with respect to  $\rho_{\text{ref},i}$ , the Spearman correlation of the reference model:

$$B_{\text{NSt}}(\text{mod}, i) = |\rho_{\text{ref},i} - \tilde{\rho}_{\text{mod},i}^{\text{NSt}}|. \quad (6)$$

### 3.4. Multi-Model Mean Climate Change (CC) Assumption

The CC assumption is specifically designed to handle and bias correct correlations from an ensemble of climate simulations. As in the NSt assumption, the MBC procedure is supposed to correctly transform the correlation of any simulation over the calibration period. Hence, for any model or run  $m$ , its correlation  $\rho_{\text{mod},0}$  is corrected to  $\rho_{\text{ref},0}$ . This CC assumption basically works the same way as the Non-stationarity assumption but accounting for the multi-model mean changes of correlations provided by the ensemble, instead of the single simulation correlation change. Hence, the change  $\Delta_{\rho,i}$  in correlation provided by a model  $m$  is not used alone (as in Equation 5 for the “Non-stationarity” assumption) to generate a future rank correlation  $\tilde{\rho}_{\text{mod},i}^{\text{NSt}}$ . Instead, Equation 5 is employed with a multi-model mean changes of correlations,  $\overline{\Delta_{\rho,i}}$ , defined as the mean of the correlation changes (from

period 0 to period  $i$ , i.e., mean of the  $\Delta_{\rho,i}$  for a given  $i$ ) from all the simulations in the ensemble, except that used for reference. As for any model mod the resulting corrected correlation over the calibration period 0 is the same ( $\rho_{ref,0}$ ) and the same common evolution  $\overline{\Delta_{\rho,i}}$  is used to generate  $\tilde{\rho}_{mod,i}^{cc}$ , for a given time period  $i$ , the consequence is that, for a given reference, all simulations end up with the same rank correlations with the CC assumption.

The bias of the correlation  $\tilde{\rho}_{mod,i}^{cc}$  obtained from the “multi-model mean CC” assumption for model mod, over period  $i$ , is noted as  $B_{CC}(mod, i)$  and is defined as:

$$B_{CC}(mod, i) = |\rho_{ref,i} - \tilde{\rho}_{mod,i}^{cc}|. \quad (7)$$

### 3.5. Spatial or Temporal Aggregation of the Biases

If  $N$  is the total number of models (or runs) in the considered ensemble (CMIP6 or CESM), each bias ( $B_{NSr}$ ,  $B_{Sr}$ ,  $B_{CC}$  and  $B_{Raw}$ ) is then calculated  $N - 1$  times for each grid cell for a given reference model and a given period  $i$ ; and  $N(N - 1)$  times for each grid cell and for a given period  $i$ , when all models (or runs) have served once as reference.

These bias criteria  $B_{NSr}$ ,  $B_{Sr}$ ,  $B_{CC}$  and  $B_{Raw}$  are calculated for each grid cell over the four climatological seasons (DJF, MAM, JJA, SON) and then spatially averaged (in order to get  $N(N - 1)$  bias values for the whole domain and for each season). The obtained spatially averaged values are hereafter referred to as  $SB_{NSr}$ ,  $SB_{Sr}$ ,  $SB_{CC}$  and  $SB_{Raw}$ .

Moreover, in order to have a spatial visualization of the results, the seasonal bias criteria can also be averaged, for each grid cell separately, over the  $N$  runs or models of a given ensemble. These locally averaged bias values are hereafter referred to as  $LB_{NSr}$ ,  $LB_{Sr}$ ,  $LB_{CC}$  and  $LB_{Raw}$ .

## 4. Results

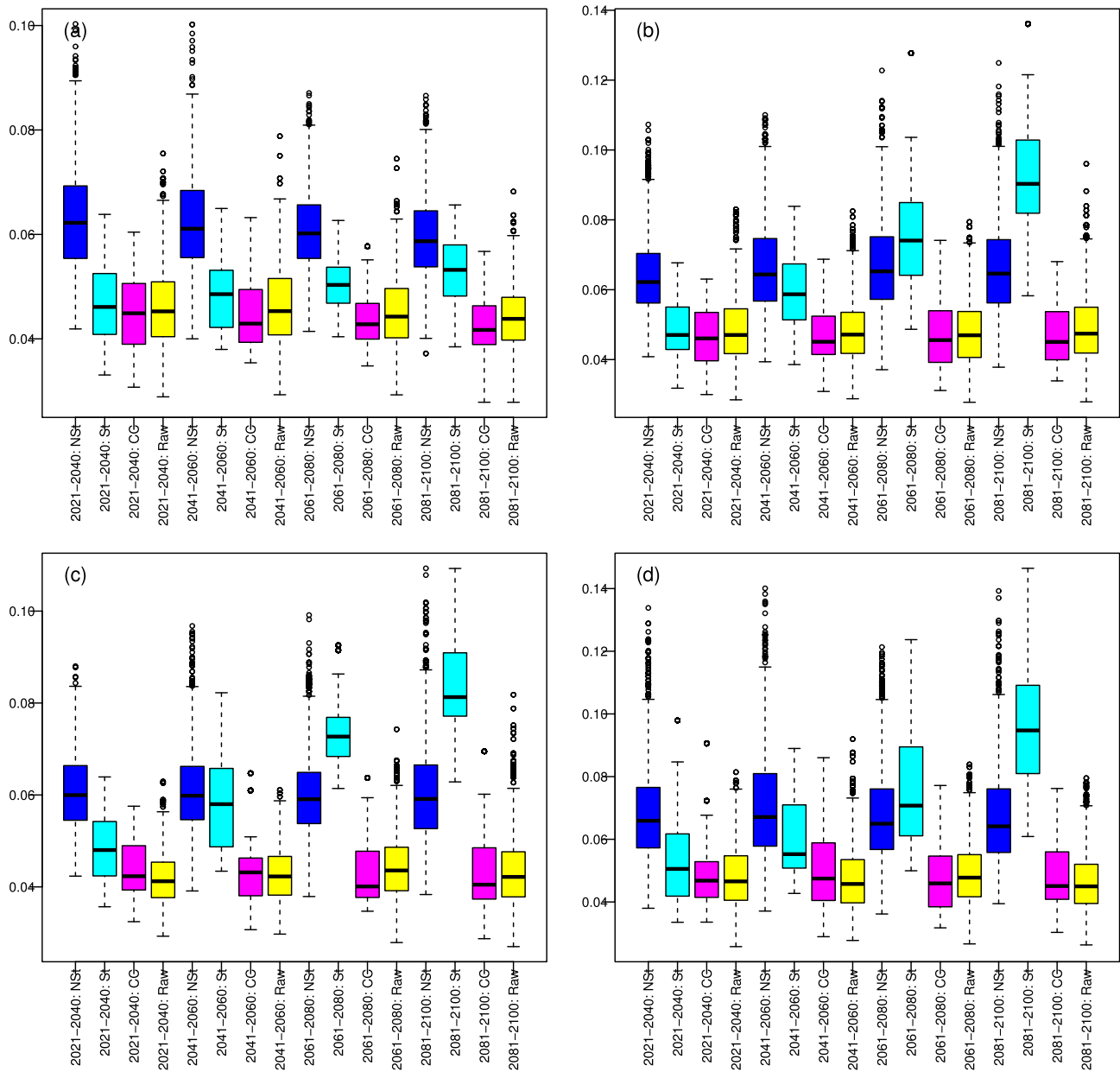
### 4.1. Spatially Averaged Biases

First, the spatially averaged  $SB_{NSr}$ ,  $SB_{Sr}$ ,  $SB_M$  and  $SB_{Raw}$  biases are presented as boxplots in Figure 1 for CESM and in Figure 2 for CMIP6, for the four seasons.

For CESM (Figure 1), except for the winter season (1.a) where the  $B_{Sr}$  values are generally lower than the  $B_{NSr}$  values whatever the future time period, the biases  $B_{Sr}$  induced by the stationarity assumption on the three other seasons (1.b-d) are lower or equivalent to those induced by the non-stationarity assumption up to about 2060 (i.e., the first two 20-year periods) but are larger afterward, for the last two periods, 2061 and on. This means that, for CESM runs, after 2060, the change in Spearman correlations becomes larger than the variability of the correlations among the different runs over the reference 2001–2020 period. This can be explained by the fact that all runs are made from a single climate model. Therefore, the change in correlations is consistent between the different runs and the variability of their time evolution is rather weak. In such a case and for long-term projections, the non-stationarity assumption has to be favored over the stationarity one. However, when looking at the CMIP6 PMTR results (Figure 2), the conclusions are quite different. Here, for all seasons and all periods in the future, the biases induced by the stationarity experiment are constantly equivalent to or lower than those induced by the non-stationarity test. This is due to the high variability of changes in correlations from one CMIP6 model to another. Contrary to the CESM ensemble, here the simulations are not generated by a single model. This implies a large inter-model uncertainty in the correlations. In such a case, the use of the non-stationarity assumption, that is, accounting for the change in correlations simulated by the different models, is not recommended and the stationarity assumption (i.e., considering  $\rho_{ref,0}$  as an approximation for the Spearman correlation in future periods) has to be favored as it reduces dependence biases.

When looking at the other assumptions (i.e., “CC” and “Raw”) for CESM (Figure 1), they appear quite equivalent from each other, for all the seasons. Unsurprisingly, these two approaches provide low  $SB_{CC}$  and  $SB_{Raw}$  values, indicating good estimates of the future correlations, for all time periods  $i$ . This is once more explained by the weak variability, among the CESM runs, of the correlation values and correlation changes over time.

The picture is not the same for CC and Raw with the CMIP6 ensemble (Figure 2). Here, the Raw assumption induces major biases that are stronger than with any other approach. Indeed, the variability of the correlation



**Figure 1.** Community Earth System Model Results of the perfect model with turning reference (PMTR) experiment presented as boxplots for the four seasons: (a) winter, (b) spring, (c) summer and (d) fall. For each season, the biases in rank correlation (no units) from the non-stationarity assumption (in dark blue), the stationarity assumption (light blue), the “CC” assumption (pink) and the “Raw” assumption (yellow) are given for 4 future periods (2021–2040, 2041–2060, 2061–2080, 2081–2100).

values and of the correlation changes over time is much higher within CMIP6 than within CESM. Therefore, not performing any initial adjustment of the modeled correlation values preserves this high variability, associated with high  $SB_{Raw}$  bias values. Yet, the CC approach appears as the most robust one for CMIP6: even with the high CMIP6 variability of correlations, accounting for the multi-model mean change of correlations allows to estimate the inter-variable dependence evolution in an efficient way, more appropriate than considering evolutions from single models separately as in the NS approach.



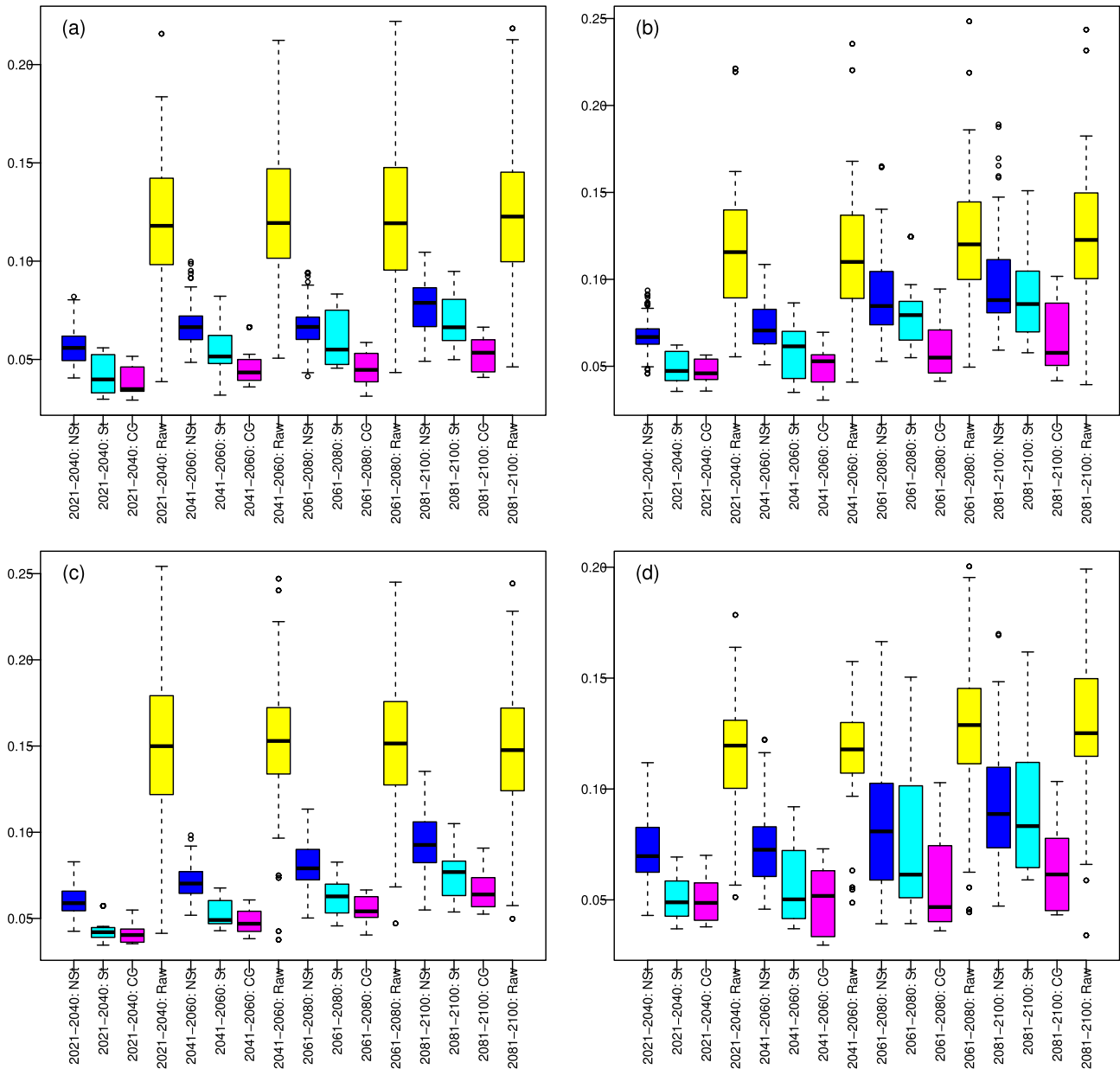
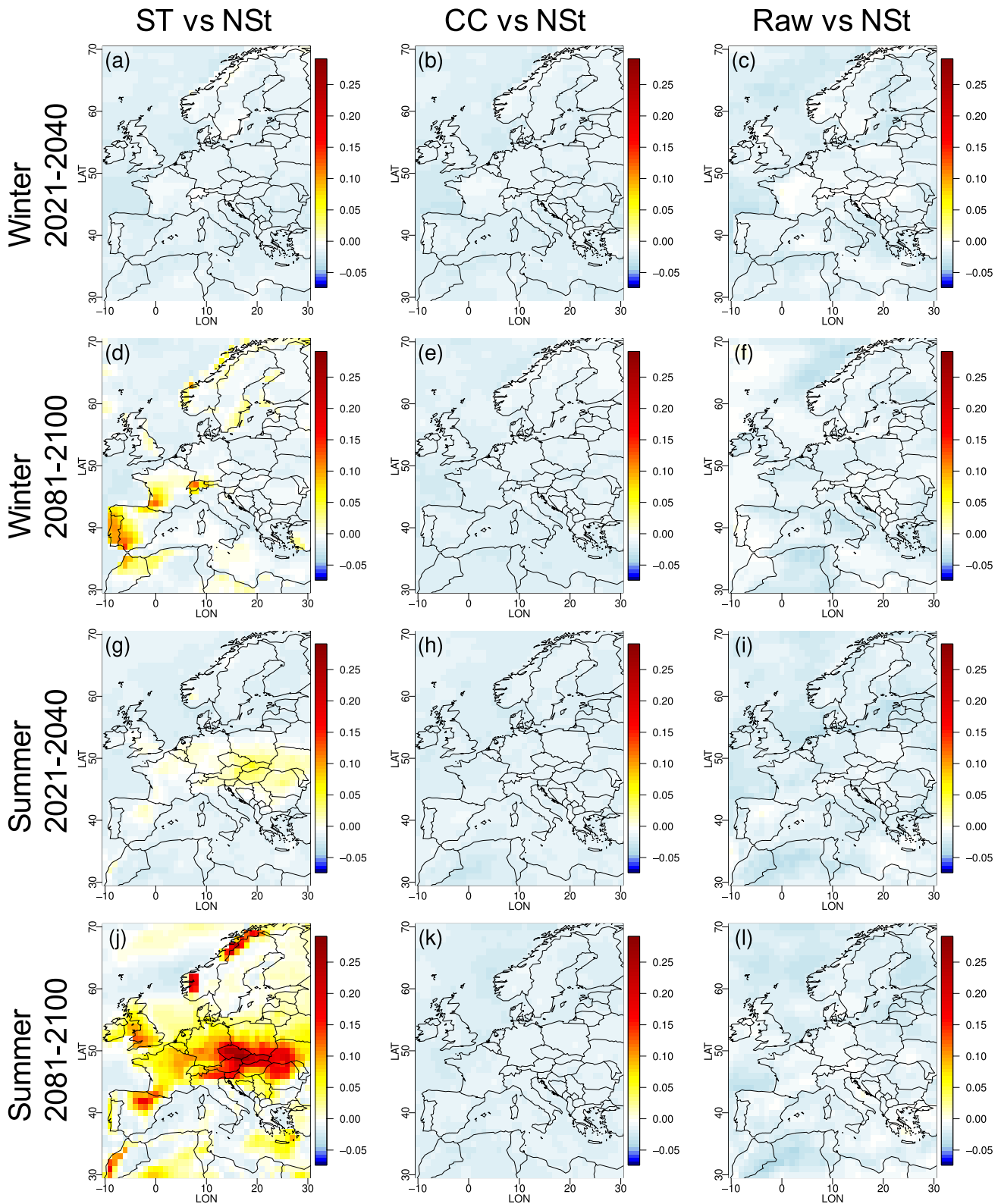


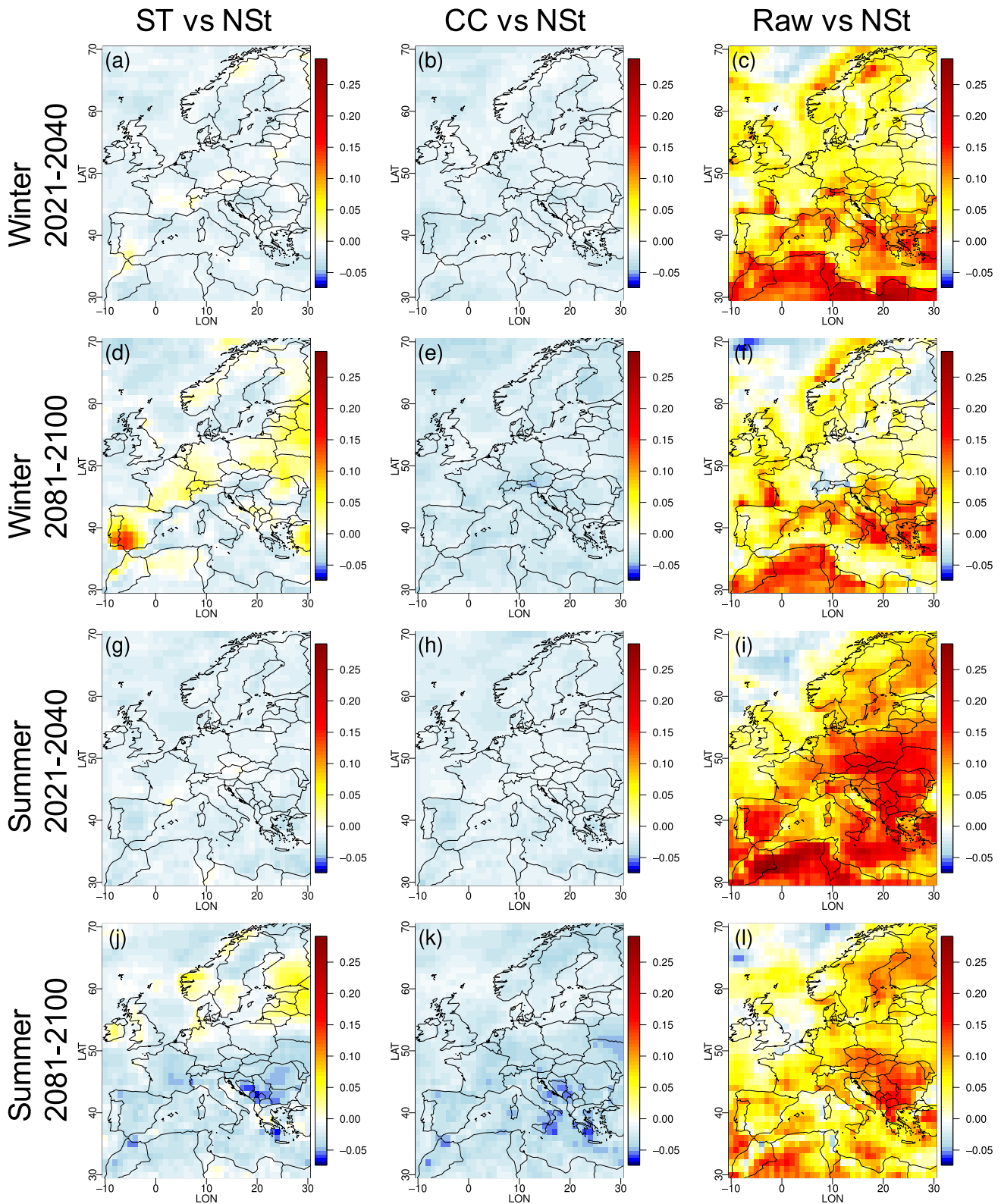
Figure 2. Same as Figure 1 but for Coupled Models Intercomparison Project6 models.

#### 4.2. Locally Averaged Biases

In order to see how the  $B_{NS}$ ,  $B_{St}$ ,  $B_{CC}$  and  $B_{Raw}$  values are distributed over the geographical domain, the locally averaged  $LB_{NS}$ ,  $LB_{St}$ ,  $LB_{CC}$  and  $LB_{Raw}$  values are used. To ease the visual assessment, for each grid cell, season and 20-year period, the difference  $DLB(A) = LB_A - LB_{NS}$  is computed, where  $A$  corresponds to one of the three assumptions (St, CC or Raw). In other words, the NS approach is used here as an arbitrary benchmark for the spatial evaluation. The resulting  $DLB$  maps for 2021–2040 ( $i = 1$ ) and 2081–2100 ( $i = 5$ ) and for winter and summer are presented in Figure 3 for CESM and Figure 4 for CMIP6. The maps for Spring and Fall are given as Supporting Information S1 for CESM and CMIP6 in Figures S1 and S2 in Supporting Information S1, respectively. A positive (yellow or red) difference indicates that the “non-stationarity” assumption implies smaller biases than the “A” one, while negative (light or dark blue) differences show locations where the “A” assumption implies smaller biases than the “non-stationarity” one.



**Figure 3.** Community Earth System Model Maps of the differences of  $LB$  values ( $DLB(A) = LB_A - LB_{NSt}$ , see text for details) computed for each grid cell, where assumption  $A$  is either “St” (first column), or “CC” (second column), or “Raw” (third column), in Winter (first and second lines) or summer (third and fourth lines), over the 2021–2040 period (first and third lines) or the 2081–2100 period (second and fourth lines). The equivalent maps for the other seasons (i.e., spring and fall) are provided as Supporting Information S1 in Figure S1 of Supporting Information S1.



**Figure 4.** Same as Figure 3 but for Coupled Models Intercomparison Project6. The equivalent maps for the other seasons (i.e., spring and fall) are provided as Supporting Information S1 in Figure S2 of Supporting Information S1.

The  $DLB(CC)$  results show very uniform maps of negative values for both CESM (panels 3.b,e,h,k) and CMIP6 (panels 4.b,e,h,k), all future periods and for all seasons. This indicates that the CC assumption performs uniformly better than the NSt approach.

This is also true for CESM  $DLB(Raw)$  maps (panels 3.c,f,i,l), showing only very weak spatial structures. However, the maps are not uniform for the CMIP6  $DLB(Raw)$  maps, where strongly positive spatial structures also change in time, for example, in summer from 2021 to 2040 to 2081–2100.

Regarding the St assumption over the near-future (2021–2040) period, as already shown in Figures 1 and 2, the results are rather equivalent for CESM (3.a) and CMIP6 (4.a) maps, with mostly close to 0 or negative  $DLB(St)$  differences all over the western Europe domain. However, spatial structures appears more and more when going through the different periods of the 21st century, as illustrated for CESM in panels 3.b and 3.j, and for CMIP6 in panels 4.b and 4.j, showing  $DLB(St)$  results over 2081–2100. For CESM, except for winter that shows weakly positive  $DLB(St)$  values over the south-western part of the domain, the other seasons present more pronounced positive  $DLB$  structures (i.e., yellow and red), especially in summer (Figure 3j) over central eastern Europe. For CMIP6, the spatial structures are much less pronounced and the major part of the domain shows only mild  $DLB(St)$  values, indicating that, even over the 2081–2100 period, the “St” and “NSt” assumptions do not distinguish much from each other and that, thus, the “stationarity” assumption remains reasonable up to the end of the 21st century. In general, it is interesting noting that, for CESM, the positive  $DLB(St)$  values—indicating smaller biases of rank correlation from the “NSt” assumption—mostly appear over lands, while negative  $DLB(St)$  values—that is, smaller biases of rank correlation from the “St” assumption—are over seas. However, although with much lower intensities than for CESM, the positive CMIP6  $DLB(St)$  values also mostly appear lands, except for summer (Figure 4j) for which most inland Europe presents negative  $DLB(St)$  values.

### 4.3. Inter-Run Biases Versus Inter-Model Biases

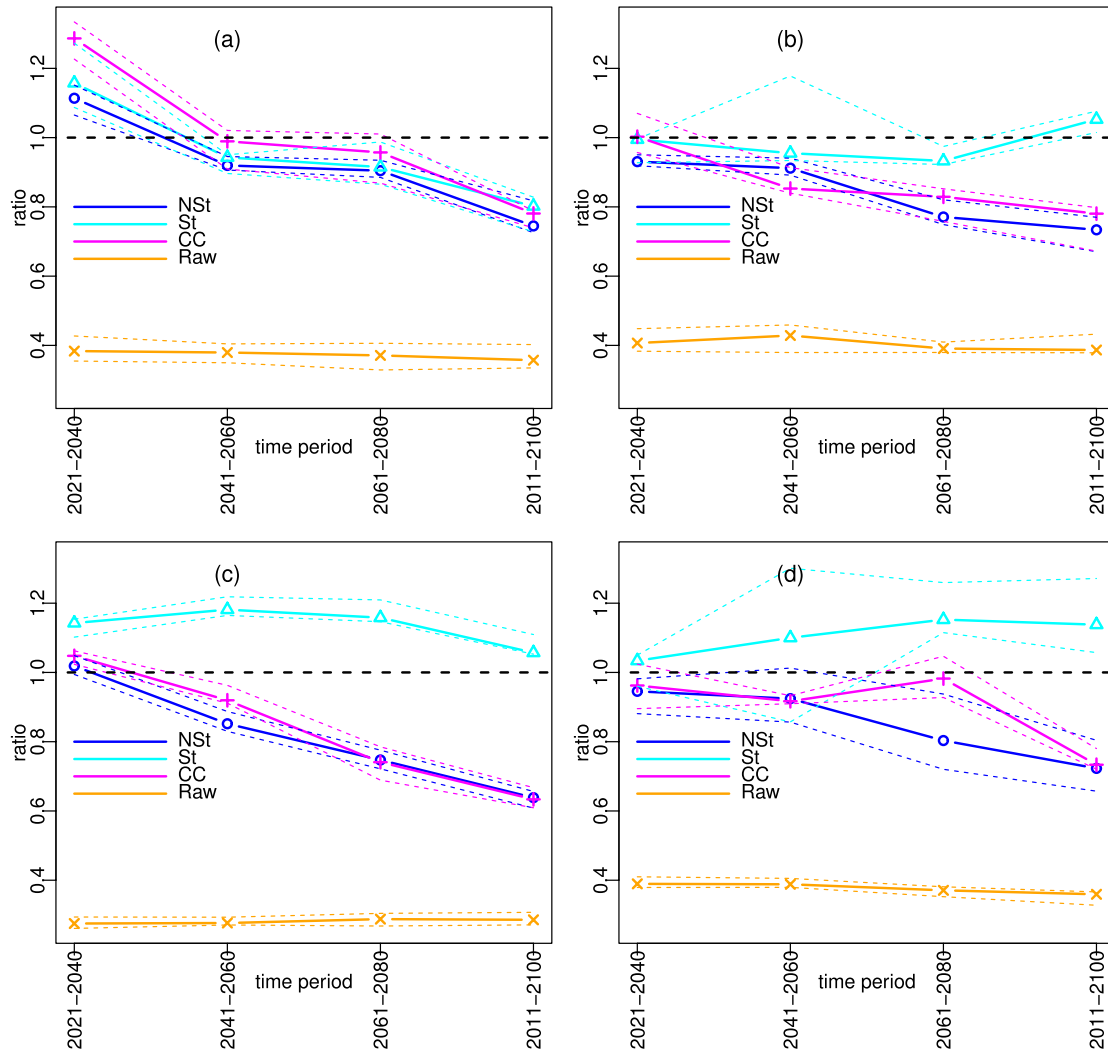
To compare the contribution of the biases from the multi-run PM experiment to the biases from the multi-model PM experiment, for each period and season, a ratio of bias,  $R_A$  is calculated for each assumption  $A$ , as the median bias from CESM (given in Figure 1) divided by the median bias from CMIP6 (Figure 2):

$$R_A = Q_{50}(SB_A^{CESM}) / Q_{50}(SB_A^{CMIP6}) \quad (8)$$

where  $Q_{50}$  is the function giving the median of a data set,  $A$  is one of the four assumptions and  $SB_A^{CESM}$  (respectively  $SB_A^{CMIP6}$ ) is the data set of the  $SB$  biases calculated for CESM (respectively CMIP6) from assumption  $A$ . By assuming that the CESM  $SB_A$  biases are representative of the  $SB_A$  biases from any single model multi-run ensemble in the CMIP6 ensemble,  $R_A$  allows to quantify the relative weights of the biases from inter-run or inter-model biases, based on assumption  $A$ . However, it is not possible to assume such a representativity of the CESM ensemble. Hence, more rigorously,  $R_A$  quantifies the relative weights of the biases from the CESM inter-run biases over the CMIP6 inter-model biases from assumption  $A$ . The  $R_A$  values are plotted in function of the time period in Figure 5 for the four assumptions and the four seasons. The 90% confidence interval of each ratio is also computed via a bootstrap of 75% of the  $SB_A^{CESM}$  and  $SB_A^{CMIP6}$  values, repeated 100 times. These intervals are given as dashed colored lines in Figure 5. Note that the intervals are generally small and relatively similar for one period to another and from one assumption to another. Yet, larger intervals are visible for the stationarity assumption (light blue) during the transition seasons (i.e., Spring 5.b and Fall 5.d) as well as for the Non-Stationarity assumption (dark blue) during Fall (5.d).

For the Raw assumption,  $R_{Raw}$  values are constant for all periods, whatever the season, around 0.4, indicating that the CESM inter-run biases are always smaller than the CMIP6 inter-model biases. For the Stationarity assumption, the ratio of bias stays mostly constant for spring, summer and fall ( $R_{St}$ , between 1 and 1.2) but decreases with time in winter (from 1.2 in 2021–2040 to 0.8 in 2081–2100). However, tendencies to decrease with time are visible for  $R_{NSt}$  and  $R_{CC}$  for all seasons and overall for winter and summer.

For assumptions showing decreasing trends of  $R_A$  (NSt and CC in all seasons, as well as St in winter), the inter-model biases increase with time with respect to the inter-run biases. In this case, the  $R_A$  values quickly go down below 1. This indicates that inter-model correlation biases are rapidly getting predominant with respect to



**Figure 5.** Ratio of the median bias from Community Earth System Model internal variability over the median bias from Coupled Models Intercomparison Project6 inter-model variability for the 4 correlation change assumptions (“NST,” “St,” “CC,” “Raw”) in winter (a), spring (b), summer (c) and fall (d). The dashed lines give the 90% confidence intervals of the ratio values for each assumption and period.

the inter-run biases. For longer term projections, this tends to favor the selection of approaches that minimize the biases of the inter-model ensemble based on our PMTR experiment.

## 5. Conclusions and Discussion

Bias correction (BC) methods are now routinely applied to adjust climate simulations and then drive impact models. If univariate BC methods are generally well-understood and have been extensively studied, multivariate ones (MBC) are still in an expansion phase that requires to understand their main assumptions, differing from univariate BCs. Whereas 1d-BC methods mostly keep the copula dependence function (e.g., its Spearman correlation) of the climate model untouched (see Vrac, 2018, among others), MBCs can rely on various assumptions regarding the possible evolutions (i.e., changes over time) of the multivariate dependencies between climate variables, such as the inter-variable (rank) correlation. Some MBCs try reproducing—generally implicitly—the future correlation changes projected by the climate model (e.g., A. Cannon, 2017; Robin et al., 2019), while other MBCs assume a stationary dependence between variables, sticking to the observational copula function (e.g., Vrac, 2018; Vrac & Thao, 2020). In this study, without correcting any multivariate simulation, we have thus investigated what these “stationarity” and “non-stationarity” assumptions imply in terms of biases of the inter-variable Spearman (rank) correlation between temperature and precipitation. To do so, an original PMTR

experiment has been set up and applied based on two different climate simulation ensembles: 40 runs from the CESM global climate model and 11 runs from the CMIP6 exercise. A run is taken as reference and the St and NSt assumptions are tested on the other runs against this reference. In addition to the St and NSt experiments, two other assumptions were evaluated: one (“CC”) that makes the correlations evolve according to the multi-model change in correlations, hence trying to capture the most robust part of the signal; and one (“Raw”) that does not transform at all the correlations and that can be seen as a proxy for a univariate BC method. For each of the four assumptions, biases of rank correlations were defined with respect to the reference and averaged either spatially or locally. All runs served once as reference, allowing to get an estimation of the variability of the biases. In our PMTR setting, the goal was to evaluate the MBC methods and their underlying assumptions about how they consider changes in rank correlations. As, for each “turn,” the reference used for training the BC methods is also a simulation, it is possible to quantify the bias or quality of the corrections (here, in terms of correlations) in a climate change context. Relying on the assumption that the truth (i.e., the real climate) and the models are generated from the same underlying probability distribution, the conclusions obtained in the PME setting are supposed to be also valid, in a practical case, when the BC methods are trained on observations. More precisely, under the above assumption, it is not possible to directly infer, from the PMTR results, the performances of the BC methods when applied to observations as the location of the observations in the underlying probability distribution (especially in the future) is not known. However, we can expect that a method that performs well on all references in the PMTR setting, would have higher chances to work well with observations. The main results highlight that:

1. Within a multi-run single GCM ensemble (such as CESM), accounting for changes in correlation (with “NSt” or “CC” assumptions) can bring valuable information, especially for long-term projections, if the real climate is assumed to be physically part of this multi-run ensemble. However, a stationary (“St”) correlation assumption appears to provide less biased correlation results than the “NSt” one, up to medium-term projections (about 2060). This is due to the low variability of correlations and correlation changes among such a multi-run single GCM ensemble. It is also worth noting that the use of a multi-run single GCM to investigate climate changes (not only in changes of correlations but also for any other aspect) is highly questionable, especially in a PMTR experiment. Indeed, as stated in Section 3, a PMTR setting assumes that the truth (i.e., the real climate) and the models are generated from the same underlying probability distribution. Given that the internal variability of a single GCM is usually lower than the variability of a multi-model ensemble, the distribution (e.g., of correlation or correlation changes) underlying the runs of a single model might thus not contain the truth. More generally, based on the same reasoning, the use of a multi-run single GCM for investigating climate changes is not recommended, except for analyzing its internal variability and potentially compare it to a larger multi-model ensemble.
2. For multi-model ensembles (such as from CMIPs), as the inter-model uncertainty in the evolution of the correlations is quite large, the use of the non-stationarity (“NSt”) assumption is not recommended and the stationarity (“St”) assumption (i.e., considering the Spearman correlations of the reference as an approximation for the correlations in future periods) has to be favored. This has important consequences for studies relying on changes of inter-variable dependence, as well as for MBC methods designed either to keep the dependence structures stationary with respect to a reference (e.g., as in Vrac, 2018; Vrac & Thao, 2020) or to make the dependencies evolve in agreement with the changes provided by the biased climate model simulations to correct (e.g., as in A. Cannon, 2017; Robin et al., 2019). Based on the results of this study, both approaches can make sense, but their appropriate use clearly depend on the confidence the MBC user puts on the model simulations and on their changes in inter-variable correlations and dependencies.
3. When an MBC method has to be applied based on a single run, the Stationary approach is preferable, rather than a non-stationarity assumption. This is particularly true if the projection period is before 2060. For longer term projections, the choice (St or NSt) mostly depends on the confidence put in the climate model used. If a high variability in correlation changes is possible (as in the CMIP6 ensemble), the Stationary approach appears safer. If the correlation changes are thought to be weakly variable around that provided by the single model run, then a non-stationary approach can be more robust. Yet, in practice, if only one single run is used, it is difficult to quantify the confidence we can have in this run, as it is not possible to quantify the agreement between models and runs. This clearly also underlines the need to investigate the different statistical and physical features of the climate model of interest and the degree of trust that can be placed in them. Therefore, constraining models by observations to reduce uncertainties in projections, as done in Robin and Ribes (2020)

or Ribes et al. (2021) for non-stationary univariate extremes, but for correlation/dependence features is a relevant and useful perspective.

4. One can note that some regional differences and variability in the results are visible (Figures 3 and 4) when comparing the “St” and “NSt” assumptions. This is the case, for example, within the CMIP6 ensemble in winter 2081–2040 (Figure 4d), over South Spain, where the NSt assumption appears locally less biased than the St assumption. This implies that the BC strategy to apply might depend on the region and the season of interest. However, this is not the case when inserting the “CC” case, which clearly appears to be the most appropriate approach, for all regions and seasons. Hence, globally, when an ensemble is available, the best and more stable results were obtained from the “CC” approach, that allows accounting for the mean change in correlations, computed as the average change from the different models or runs. Thus, when it is possible to use an ensemble, the CC approach has to be favored over any of the three other discussed assumptions. More generally, this result also pleads for the use of ensemble—instead of a single model run—and the robust information about climate change that they can provide.

This study can, of course, be further extended in many ways. First, it is worth reminding that, here, we have not performed any (univariate or multivariate) BC per se. Indeed, no time series have been adjusted, as we only relied on ways to estimate the evolution of the correlations, as proxies of results given by MBCs. Hence, we did not evaluate specific methods and their details, but rather their main underlying philosophies. The consequence is that, to get a precise understanding of the suited MBC methods to use, the same PME protocol could be applied directly to the MBC methods of interest.

Second, GCM simulations have been used to constitute the ensembles of this study. Yet, ensembles for RCM simulations, with higher spatial resolution, could provide different complementary insights and results. This also raises the question on the different sources of variability in the correlation values and changes: from large-scale or from local-scale simulations? From inter-GCM or from inter-RCM simulations? Our proposed PME would then have to be adapted to tackle such questions and to be able to separate the various sources.

Moreover, if this study only looked at the T versus PR correlations, other couples of climate variables (e.g., wind and precipitation, or humidity and temperature) or other dependence metrics (e.g., Kendall's tau) can be investigated in the same way, depending on the specific interest. In the same type of idea, different types of dependencies, other than inter-variable, could also benefit from such a framework: spatial dependencies, or temporal dependencies, including cross-dependencies. In addition, extending the PME approach developed in this study to copula functions instead of rank correlations could help describing better the dependence between variables and the consequences from MBC methods. Indeed, as explained in the Introduction section, the correspondence between copula and rank correlation is not one-to-one. Hence, the conclusions brought by this study based on rank correlations can be deepened with an approach directly working on copulas. Yet this copula-based approach still has to be defined and is left for future work. With such systematic evaluations, it would then be possible to get clear pictures of the most robust ways to account for changes in various dependence properties with MBC methods.

### Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

### Data Availability Statement

All CMIP6 GCM simulations used in this article (Boucher et al., 2018; Seferian, 2018; Shiogama et al., 2019; Swart et al., 2019; Tang et al., 2019; Voltaire, 2018, 2019; Volodin et al., 2019; Wu et al., 2018; Yukimoto et al., 2019) and listed in Table 1 can be downloaded through the Earth System Grid Federation portals. Instructions to access the data are available at: <https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html>. The CESM Large Ensemble Simulations can be downloaded from the CESM Large Ensemble Community Project website: <https://www.cesm.ucar.edu/projects/community-projects/LENS/>.

**Acknowledgments**

We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their models outputs. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We acknowledge the CESM Large Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone for making the CESM Large Ensemble Simulations publicly available. MV and ST have been supported by the "COESION" project funded by the French National program LEFE (Les Enveloppes Fluides et l'Environnement), the Swiss national programm FNS "Combine" project, as well as the French National "Explore2" project funded by the French Ministry of Ecological Transition (MTE) and the French Office for Biodiversity (OFB).

**References**

Abatzoglou, J., Dobrowski, S., & Parks, S. (2020). Multivariate climate departures have outpaced univariate changes across global lands. *Scientific Reports*, 10(3891), 1–9. <https://doi.org/10.1038/s41598-020-60270-5>

Abatzoglou, J., Dobrowski, S., Parks, S., & Hegewisch, K. (2018). Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(170191), 1–12. <https://doi.org/10.1038/sdata.2017.191>

Boucher, O., Denvil, S., Levassasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., et al. (2018). IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1534>

Cannon, A. (2017). Multivariate quantile mapping bias correction: An n-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50(1–2), 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>

de Elia, R., Laprise, R., & Denis, B. (2002). Forecasting skill limits of nested, limited-area models: A perfect-model approach. *Monthly Weather Review*, 130(8), 2006–2023. [https://doi.org/10.1175/1520-0493\(2002\)130<2006:fslonl>2.0.co;2](https://doi.org/10.1175/1520-0493(2002)130<2006:fslonl>2.0.co;2)

Dekens, L., Parey, S., Grandjacques, M., & Dacunha-Castelle, D. (2017). Multivariate distribution correction of climate model outputs: A generalization of quantile mapping approaches. *Environmetrics*, 28(6), e2454. <https://doi.org/10.1002/env.2454>

Déqué, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, 57(1–2), 16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>

Doblas-Reyes, F., Sörensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., et al. (2021). Linking global to regional climate change. In V. Masson-Delmotte, P. Zhai, A. Pirani, et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>

François, B., Thao, S., & Vrac, M. (2021). Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. *Climate Dynamics*, 57(11), 3323–3353. <https://doi.org/10.21203/rs.3.rs-299929/v1>

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

Haddad, Z., & Rosenfeld, D. (1997). Optimality of empirical z-r relations. *Quarterly Journal of the Royal Meteorological Society*, 123(541), 1283–1293. <https://doi.org/10.1002/qj.49712354107>

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction – The isi-mip approach. *Earth System Dynamics*, 4(2), 219–236. <https://doi.org/10.5194/esd-4-219-2013>

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>

Hillier, J., Matthews, T., Wilby, R., & Murphy, C. (2020). Multi-hazard dependencies can increase or decrease risk. *Nature Climate Change*, 10(7), 595–598. <https://doi.org/10.1038/s41558-020-0832-y>

IPCC WGL. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*. In V. Masson-Delmotte, P. Zhai, et al. (Eds.), Cambridge University Press.

Kallache, M., Vrac, M., Naveau, P., & Michelangeli, P.-A. (2011). Non-stationary probabilistic downscaling of extreme precipitation. *Journal of Geophysical Research: Atmospheres*, 116(D5), D05113. <https://doi.org/10.1029/2010JD014892>

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community Earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8), 1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1>

Kendon, E. J., Rowell, D. P., Jones, R. G., & Buonomo, E. (2008). Robustness of future changes in local precipitation extremes. *Journal of Climate*, 21(17), 4280–4297. <https://doi.org/10.1175/2008JCLI2082.1>

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models [Journal Article]. *Journal of Climate*, 23(10), 2739–2758. <https://doi.org/10.1175/2009JCLI3361.1>

Krinner, G., & Flanner, M. G. (2018). Striking stationarity of large-scale climate model bias patterns under strong climate change. *Proceedings of the National Academy of Sciences*, 115(38), 9462–9466. <https://doi.org/10.1073/pnas.1807912115>

Matte, D., Larsen, M. A. D., Christensen, O. B., & Christensen, J. H. (2019). Robustness and scalability of regional climate projections over Europe. *Frontiers in Environmental Science*, 6, 163. <https://doi.org/10.3389/fenvs.2018.00163>

Piani, C., & Haerter, J. O. (2012). Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters*, 39(20), 2012GL053839. <https://doi.org/10.1029/2012GL053839>

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>

Ribes, A., Qasmi, S., & Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Science Advances*, 7(4), eabc0671. <https://doi.org/10.1126/sciadv.abc0671>

Ribes, A., Zwiers, F., Azais, J., & Naveau, P. (2016). A new statistical approach to climate change detection and attribution. *Climate Dynamics*, 48(1–2), 367–386. <https://doi.org/10.1038/sdata.2017.191>

Ridder, N. N., Pitman, A. J., & Ukkola, A. M. (2021). Do CMIP6 climate models simulate global or regional compound events skillfully? *Geophysical Research Letters*, 48(2), e2020GL091152. <https://doi.org/10.1029/2020GL091152>

Robin, Y., & Ribes, A. (2020). Nonstationary extreme value analysis for event attribution combining climate models and observations. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(2), 205–221. <https://doi.org/10.5194/ascmo-6-205-2020>

Robin, Y., & Vrac, M. (2021). Is time a variable like the others in multivariate statistical downscaling and bias correction? *Earth System Dynamics Discussions*, 2021, 1–32. <https://doi.org/10.5194/esd-2021-12>

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>

Sadegh, M., Mofatkhari, H., Gupta, H. V., Ragno, E., Mazdiyasi, O., Sanders, B., et al. (2018). Multihazard scenarios for analysis of compound extreme events. *Geophysical Research Letters*, 45(11), 5470–5480. <https://doi.org/10.1029/2018GL077317>

Seferian, R. (2018). CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1391>



- Shiogama, H., Abe, M., & Tatebe, H. (2019). MIROC MIROC6 model output prepared for CMIP6 ScenarioMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.898>
- Singh, H., Najafi, M., & Cannon, A. (2021). Characterizing non-stationary compound extreme events in a changing climate based on large-ensemble climate simulations. *Climate Dynamics*, 56(5–6), 1389–1405. <https://doi.org/10.1007/s00382-020-05538-2>
- Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). CCCma CanESM5 model output prepared for CMIP6 ScenarioMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1317>
- Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., et al. (2019). MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.6113>
- Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections [Journal Article]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 365(1857), 2053–2075. <https://doi.org/10.1098/rsta.2007.2076>
- Thao, S., Garvik, M., Mariétoz, G., & Vrac, M. (2021). Combining global climate models using graph cuts. *Climate Dynamics*. <https://doi.org/10.21203/rs.3.rs-586868/v1>
- Voltaire, A. (2018). CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1375>
- Voltaire, A. (2019). CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.1387>
- Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019). INM INM-CM5-0 model output prepared for CMIP6 CMIP abrupt-4xCO2 [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.4932>
- Volosciuk, C., Maraun, D., Vrac, M., & Widmann, M. (2017). A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrology and Earth System Sciences*, 21(3), 1693–1719. <https://doi.org/10.5194/hess-21-1693-2017>
- 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>
- Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., & Somot, S. (2012). Dynamical and statistical downscaling of the French Mediterranean climate: Uncertainty assessment. *Natural Hazards and Earth System Sciences*, 12(9), 2769–2784. <https://doi.org/10.5194/nhess-12-2769-2012>
- Vrac, M., Noël, T., & Vautard, R. (2016). Bias correction of precipitation through Singularity Stochastic Removal: Because occurrences matter. *Journal of Geophysical Research: Atmospheres*, 121(10), 5237–5258. <https://doi.org/10.1002/2015JD024511>
- Vrac, M., Stein, M. L., Hayhoe, K., & Liang, X.-Z. (2007). A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, 34(18), L18701. <https://doi.org/10.1029/2007GL030295>
- Vrac, M., & Thao, S. (2020).  $R^2d^2$  v2.0: Accounting for temporal dependences in multivariate bias correction via analogue rank resampling. *Geoscientific Model Development*, 13(11), 5367–5387. <https://doi.org/10.5194/gmd-13-5367-2020>
- Vrac, M., Thao, S., & Yiou, P. (2021). Changes in temperature-precipitation correlations over Europe: Are climate models reliable? *Climate Dynamics*. <https://doi.org/10.21203/rs.3.rs-1008080/v1>
- Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., et al. (2018). BCC BCC-CSM2MR model output prepared for CMIP6 CMIP piControl [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.3016>
- Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP [Dataset]. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.621>
- Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R., et al. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, 1(7), 333–347. <https://doi.org/10.1038/s43017-020-0060-z>
- Zscheischler, J., Naveau, P., Martius, O., Engelke, S., & Raible, C. C. (2021). Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics*, 12(1), 1–16. <https://doi.org/10.5194/esd-12-1-2021>