Response to Referee Comment 1 on **"Multivariate bias corrections of climate simulations: Which benefits for which losses?"** ***by*** **Bastien François et al.**

Jakob Zscheischler (Referee)

Comment:

This is a timely paper providing an overview about the plethora of newly emerging multivariate bias correction approaches that have been developed over the recent years. The authors provide recommendation about which approach should be used under which conditions. The paper has the potential to become a key reference for multivariate bias correction approaches. It has an easy to follow clear structure, is well written and falls into the scope of ESD. I have a few minor recommendations which should help to improve its accessibility and impact.

Response:

We would like to thank Dr. Zscheischler for his very positive comments. We also thank him for the detailed remarks that we will try to include in the updated manuscript. All the comments and our point-by-point responses are given below.

Comment:

Introduction: I miss some strong arguments why and in which situations we need MBC. For many impacts, univariate BC is (probably) enough and MBC does not provide are large boost in performance. Indeed a number of studies have argued over the last years that for their application domain MBC does not outperform univariate BC (Yang et al., 2015; Casanueva et al., 2018; Räty et al., 2018). However, I would argue that these results cannot be generalized. One particularly relevant field of application where MBC should be highly beneficial is the area of compound events, where multiple climate drivers result in a large impact (Zscheischler et al., 2018). Arguably, a bias in the dependence structure of the drivers can result in unknown biases of the modelled impacts, which may even be aggravated by univariate BC (Zscheischler et al., 2019).

Response: We agree with this comment and propose the following corrections (in blue) in the introduction from L40 of the initially submitted article:

"Although univariate distribution features are adjusted according to references, it can generate inappropriate multivariate situations where the dependence structure between variables and sites is not corrected from the model ***and misrepresented*** *(Maraun, 2013)****, or even modified****. Ignoring the observed inter-variable and inter-site dependencies in the correction procedure can result in obtaining corrected outputs with inappropriate physical laws, and thereby distorting the results of impact studies* ***(Zscheischler et al, 2019)****. It is therefore of paramount importance to adjust the dependence structures of climate simulations, in addition to 1d-characteristics, before using it in subsequent studies.*

These methodological issues have led up to the recent development of a few multivariate bias correction (MBC) methods. ***Not only do these methods adjust univariate distribution features, they are also aimed at correcting the dependence structure of climate simulations. Recent studies have shown that univariate BC methods can already provide adequate results for certain specific regional impact studies (Yang et al., 2015 ; Casanueva et al., 2018), and that using MBC methods does not necessarily present substantial benefits (Räty et al., 2018). However, this does not call into question the interest of MBC methods as these specific results cannot be generalized to each method and application. In particular, MBC methods could be valuable in larger-scale impact modelling frameworks such as compound events, where the combination of physical processes across multiple spatial and temporal scales leads to significant impacts (Zscheischler et al, 2018).*** *As mentioned by Vrac (2018), and* *completed by Robin et al. (2019),* ***MBC methods*** *may be grouped into three main categories of approaches: the "marginal/dependence" correction approach, the "successive conditional" correction approach, and the "all-in-one" correction approach."*

Comment:

I'm not sure I entirely agree with the interpretation of Section 5.5.2 and figure 7. As I understand it, Wd only measures a distance. Hence if one obtains a similar value >0 it is unclear whether the change goes into the same direction. One might obtain a similar value for Wd but very different changes in the underlying distributions (though I admit that this would be coincidence and might not be very likely). I think this caveat should be mentioned.

Response:

We agree with this comment and want to thank Dr. Zscheischler for this remark. We suggest to add the following sentences (in blue) in the paragraph starting at L489 of sub-section 5.5.2 (Results / Analysis of change in spatial correlations) of the initially submitted article:

"*In particular, computing Wd using ranks instead of raw values allows removing the change in the univariate distributions from that in spatial and inter-variable relationships.* ***However, comparing Wd values of climate datasets must be made with caution. Indeed, similar values of Wd for different climate datasets do not necessarily imply that their changes of spatial structure are similar.*** *Results for the three Wasserstein distances on ranks are displayed in Fig. 7 for both France and Brittany. Additional results for Wd on raw values are displayed in Fig. S7 for information purposes only.*

For France (Fig. 7a), the three Wd are slightly higher for the reference than for the model data (represented by straight lines). ***Although the differences are quite small, it cannot be concluded directly that changes of spatial structure are identical, as there is no particular reason for this. For CDF-t outputs, similar Wd are obtained as those from the model. However, as the 1d-BC method does not modify (too much) rank sequence of temperature and precipitation time series, it can be deduced that CDF-t outputs globally reproduce/preserve*** *the spatial structure change of the model."*

We also suggest to correct the following paragraph starting at L512 in the same sub-section 5.5.2 (Results / Analysis of change in spatial correlations) of the initially submitted article as follows:

For both dOTC and MBCn outputs, Wd are higher than those from the model. Although the changes in spatial correlations derived by these two methods are too strong, it nevertheless highlights their ability to capture such a change from the model ***and to use it in their bias correction procedure. Moreover, as explained in subsection 5.4, dOTC and MBCn methods modify only slightly the rank structure of the initial simulations. It can then be deduced that the changes in spatial correlations measured for the two methods are (partially) in agreement with those from the model. However,*** *for MBCn, the three Wasserstein distances increase according to the number of dimensions considered in the bias correction, from 2d- to Full-versions. It can be linked with the deterioration of the quality of results already observed for spatial features for very high-dimensional bias correction.* ***Regarding MRec, and without speaking about its Full-version, similar observations can be made for 2d- and Spatial-outputs as well.*** *In a general way, the Wd associated to the different configurations for dOTC, MBCn and MRec are always above the Wasserstein distances for R2D2, illustrating somehow the assumptions made by these methods about the stationary or non-stationary copula functions.*

Comment:

L604: Other examples for changes in dependence that might be highly relevant for impacts are: - increases in the dependence between storm surge and heavy precipitation in US coasts in the historical period (Wahl et al., 2015): affects the risk of compound floods; - increase in the strength of dependence between seasonal summer temperature and precipitation of most land regions with increasing warming (Zscheischler & Seneviratne, 2017): affects the likelihood of compound hot and dry events with a large array of impacts

Response:

To account for this remark, we propose a change of the sentence starting at L604 in the Discussion and recommendations sub-section of the initial submission as following:

"In a general way, copula non-stationarity for future periods can be reasonably expected, e.g. as documented for rainfall spatial distributions (Wasko et al., 2016), for the dependence between storm surge and rainfall (Wahl et al., 2015) and the dependence between seasonal summer temperature and precipitation (Zscheischler & Seneviratne, 2017). However, on the contrary, it can be argued that inter-variable and spatial dependence structures can be assumed to be stable over time for specific regions, because, to some extent, they can be considered as imposed by physical regional constraints (Vrac, 2018)."

Comment:

L642: This is easier said than done. The largest challenge in evaluating impact modelling output is the availability of impact data. It will therefore be difficult to decide which BC approach is more appropriate. That said, I agree that creating an ensemble of different approaches might help to cover uncertainties that are not only related to the choice of the GCM and forcing scenario but also the choice of BC method.

Response:

We agree with this comment and propose the modification of the sentences starting at L641 (Subsection 6.3/Future Work) as follows:

"Moreover, as mentioned in the introduction section, bias adjusted simulations are particularly valuable for impact studies. ***Despite the challenge of missing impact data****, evaluating how the quality of multivariate bias-corrected data influences the results of complex impact models is an important perspective.* ***Providing*** *such an analysis will be useful for the scientific community working on climate change impacts, e.g., in hydrology, agronomy or ecology."*

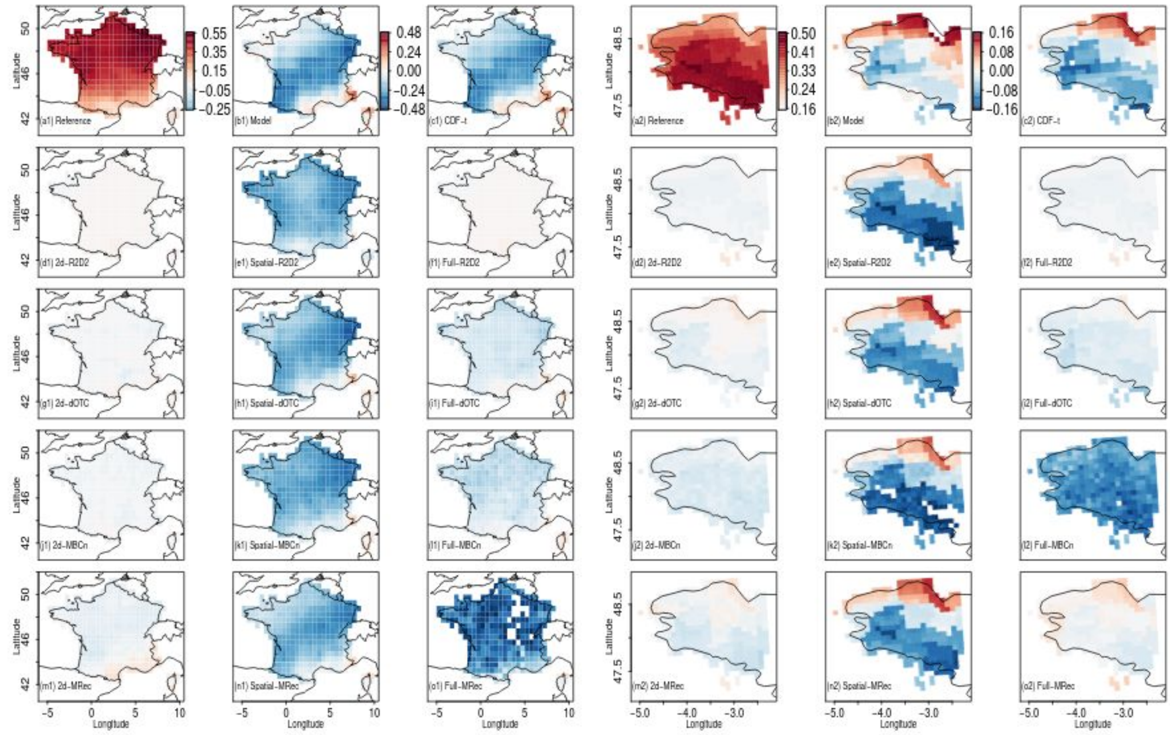
Comment:

Figure 2 and 4: The correlations could be plotted as difference to the reference to highlight the differences

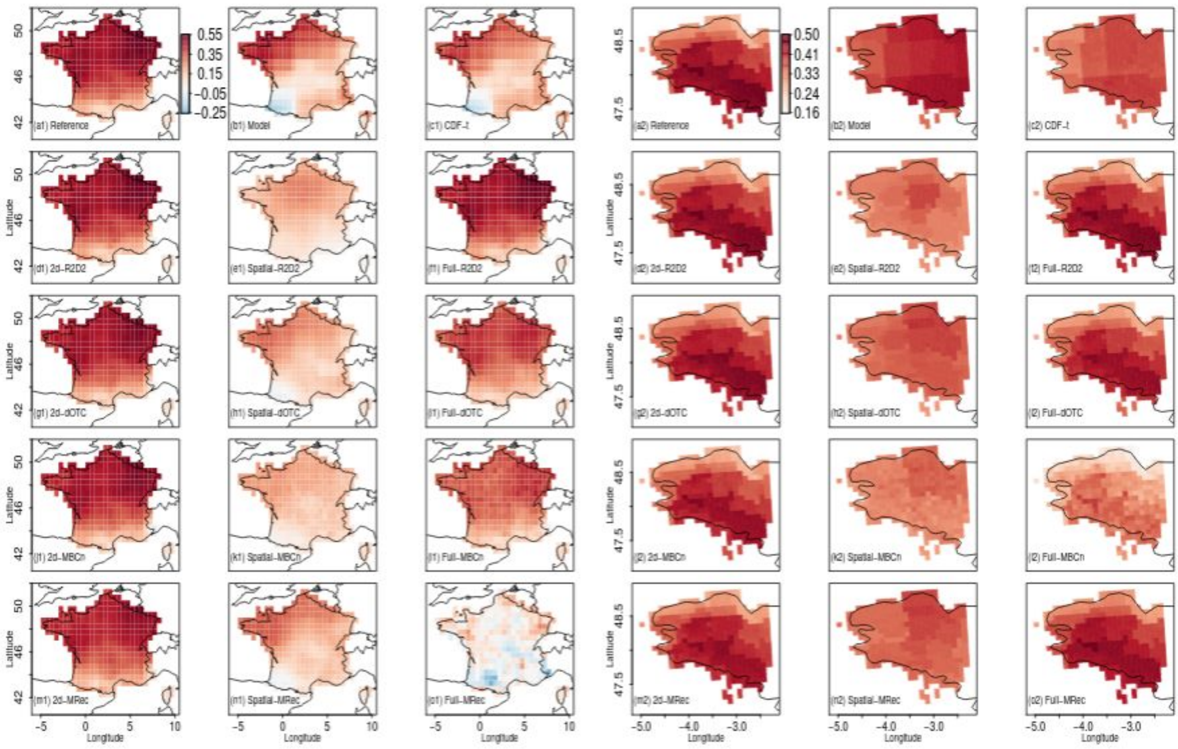
Response:

After careful consideration, we decided to change Figures 2 and 4 as suggested in this comment, by plotting differences with respect to the reference for Spearman correlations (Fig. 2) and order 1 Pearson autocorrelation for temperature (Fig. 4). However, in order to be consistent, we think that these changes imply to change Figure S2 and S4 in the supplementary materials. Captions and numbering of figures will be changed accordingly. Conclusions of the analysis based on Figure 2 and 4 do not change but few modifications of the text are needed to match the updated format of the new figures.

- Figure S2 previously corresponded to relative differences (in %) of Spearman correlations. We now propose to replace previous Fig. S2 by the initial Figure 2 to provide equivalent information to readers.



Proposed new Figure 2: Differences of temperature vs. precipitation Spearman correlations computed at each grid cell for BC methods using WFDEI reference (a1-o1) and SAFRAN reference (a2-o2) during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R2D2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between (a1-o1) and (a2-o2) are not the same to better emphasize intensities of values of the two regions.



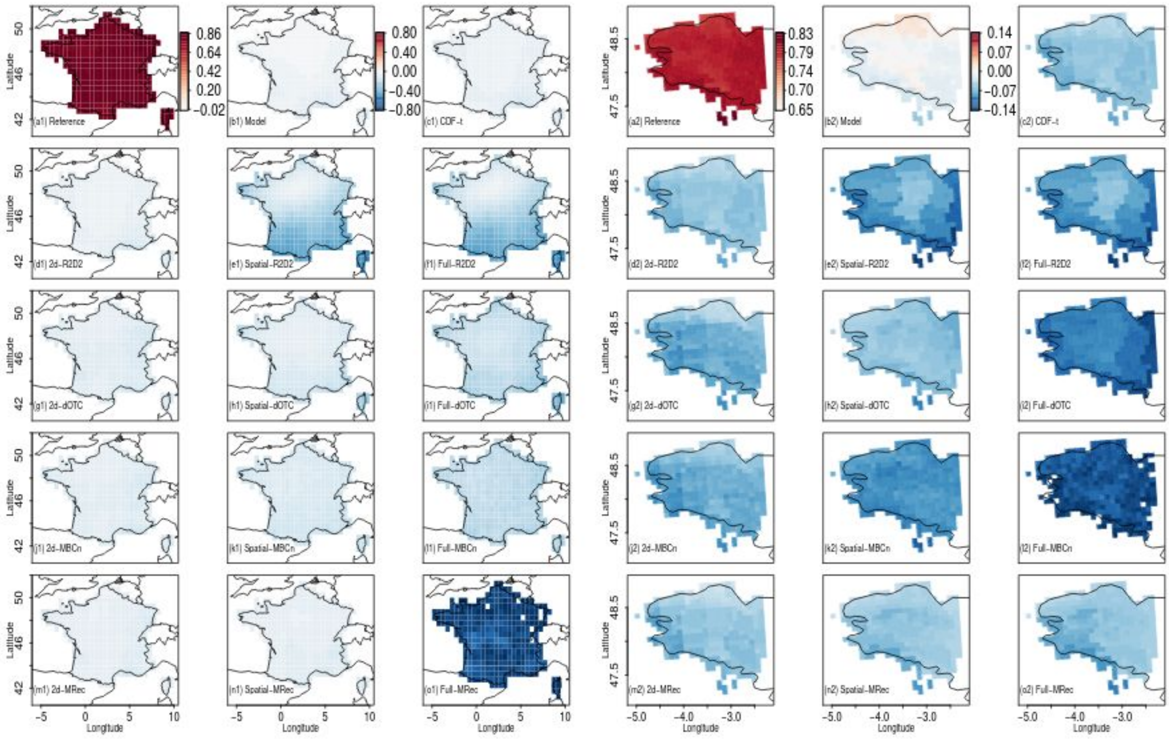
Proposed new Figure S2: Maps of temperature vs. precipitation Spearman correlation computed at each grid cell for BC methods using WFDEI reference (a1-o1) and SAFRAN reference (a2-o2) during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R2D2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between (a1-o1) and (a2-o2) are not the same to better emphasize intensities of values in the two regions.

With the new Figure 2, we propose the following corrections of the sentences starting at L297 in the subsection 5.2 (Inter-variable correlations):

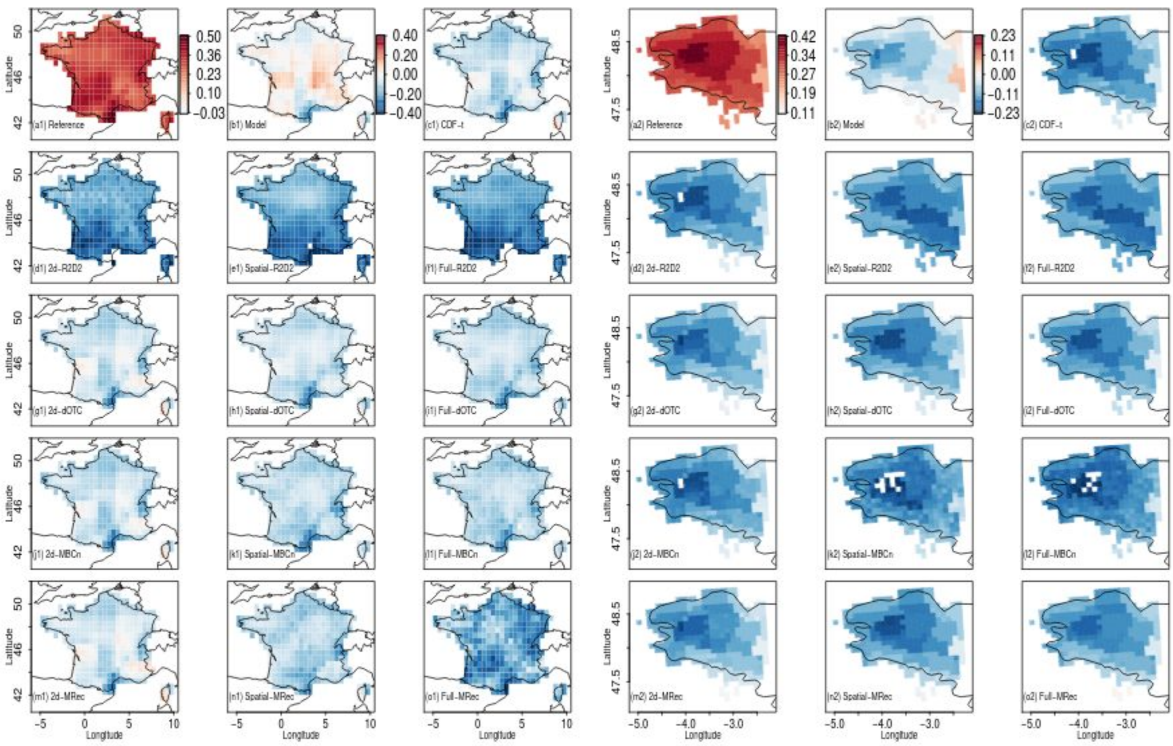
*"**The maps of the Spearman correlation differences with respect to the reference - for the IPSL model and the bias-corrected data -** are displayed in Fig. 2 for both France and Brittany. **Initial maps of Spearman correlations, i.e. without differences with respect to the reference, are also provided in Fig. S2**.*

*For France, map for the IPSL simulations (Fig. 2b1) **indicates strong differences with respect to the WFDEI map (Fig. 2a1). As the univariate CDF-t** method does not modify rank sequence of temperature and precipitation time series, it globally conserves both the rank correlation intensities and structures of the IPSL model for each region and does not provide any correction of this aspect (Fig. 2c1). By construction, clear improvements of the inter-variable correlation structure are provided by 2d-versions (Fig. 2d1, 2g1, 2j1 and 2m1). This is also the case for most of the full-configuration of MBCs (respectively, Fig. 2f1, 2i1, 2l1) despite possible differences in intensities. Note that **maps of correlation differences for 2d-R2D2 (Fig. 2d1)** and Full-R2D2 (Fig. 2f1) are identical. Indeed, for the inter-variable aspect, 2d-version is nested within the Full-configuration (see Vrac, 2018), due to the use of the reordering technique in R2D2. Also, for R2D2, the choice of the reference dimension does not have any impact on results in the inter-variable context, as it only modifies the rank chronology of time series. **As expected from previous explanations, the map for the Full-version of MRec (2o1) indicates a strong deterioration of the inter-variable correlation structure.** It highlights again the inability of the method to work properly for France in this dimensional setting. Concerning Spatial-versions of MBCs (Fig. 2e1, 2h1, 2k1 and 2n1), as they adjust the whole simulated field of temperature and precipitation separately, they disregard inter-variable relationships. It results in BC outputs with strongly weakened inter-variable correlations structures.*
Regarding Brittany, the same conclusions can be drawn for R2D2 and dOTC, for which spatial resolution does not affect the results of inter-variable properties adjustment. As noted previously, Full-MRec over Brittany provides more satisfactory results than those obtained over France, and are in line with those obtained for R2D2 and dOTC. However, for MBCn outputs, a degrading effect from 2d- (Fig. 2j2) to Full- (Fig. 2l2) is observed, in providing a corrected correlations' structure but with underestimated intensities in the high-dimensional context."

- For Figure S4: As order 1 Pearson autocorrelations for precipitation were initially plotted in Figure S4, we propose to change this figure and replace it by the differences of order 1 Pearson autocorrelations for precipitation with respect to the reference. This change permits to be in line with the changes proposed for Figure 4.



Proposed new Figure 4: Differences of order 1 Pearson autocorrelation for temperature using WFDEI reference (a1-o1) and SAFRAN reference (a2-o2) during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R2D2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between (a1-o1) and (a2-o2) are not the same to better emphasize intensities of values of the two regions.



Proposed new Figure S4: Differences of order 1 Pearson autocorrelation for precipitation using WFDEI reference (a1-o1) and SAFRAN reference (a2-o2) during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R2D2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between (a1-o1) and (a2-o2) are not the same to better emphasize intensities of values of the two regions.

With the new Figure 4, we propose the following corrections of the sentences starting at L368 in the subsection 5.4 (Temporal structure):

"The different MBC methods implemented here are not intended to adjust temporal structures. Indeed, these multivariate procedures adjust multivariate distributions without accounting for any temporal information. However, although the temporal structures are not adjusted according to the reference, MBCs necessarily modify the rank sequences of the simulations (Vrac, 2018). This modification is not performed in the same way depending on the MBC or the dimensional configuration used, and remains therefore to evaluate. To do so, 1-day lag Pearson autocorrelations are computed at each grid cell for temperature and precipitation. ***The resulting maps of differences with respect to the reference for the different climate datasets are displayed in Fig. 4 (resp. Fig. S4) for temperature (resp. precipitation).***

For France, IPSL temperature autocorrelations differences (Fig. 4b1) are small, indicating a relative agreement of IPSL with WFDEI reference dataset (Fig. 4a1), showing equivalent high values. Similar differences map are provided by CDF-t outputs (Fig. 4c1). *It is however not the case for precipitation (Fig. S4c1), for which a decrease of autocorrelation values is observed over France with respect to the reference and to the model. Although not observed for temperature, it highlights that the univariate correction could have a non-negligible effect on Pearson autocorrelation. Interestingly, 2d-versions (Figs. 4d1, 4g1, 4j1 and 4m1) do not lead to a strong modification of temporal properties with respect to CDF-t. However, from one method to another, temporal structure modifications are not equivalent for Spatial- and Full-versions. For dOTC and MBCn (Figs. 4h1, 4i1, 4k1 and 4l1), as the number of dimensions increases, the temperature autocorrelations seem to be more and more modified, with intensities of values decreasing slightly from Spatial- to Full-versions. This result can also be seen for precipitation in Fig. S4. With regard to MRec, its Spatial-version (Fig. 4n1) presents similar results than those obtained from Spatial-dOTC and Spatial-MBCn. Also, and as expected, Full-MRec outputs (Fig. 4o1) do not provide sensible results due to the inability of the method to work properly over the whole France. Concerning R2D2, as the reference dimension driving the rank sequence is the same between Spatial- and Full-configurations,* ***same differences of autocorrelation maps*** *are obtained for these two versions (Figs. 4e1 and 4f1). Moreover, autocorrelation value in the grid cell of the reference dimension, i.e. located over Paris for France, is exactly equal to the corresponding one in the CDF-t outputs, by construction. Remarkably, as mentioned by Vrac (2018), autocorrelations of the CDF-t outputs are partially reproduced around the specific locations of the reference dimensions for Spatial-R2D2 and Full-R2D2 versions, as evidenced by the* ***lightly-shaded area around Paris****. This reflects the existing spatial correlations between the reference dimension and its local neighbourhood, which results in partially reproducing the temporal properties of the model over this area. However, for precipitation (Figs. S4e1 and S4f1), this result is not as clear-cut as it is for temperature, probably due to weaker spatial correlations around Paris for this physical variable.*

In a general way, the same conclusions can be drawn for Brittany, sometimes even better illustrated due to a narrower color scale. The results for Full-MRec are easier to interpret. They present results similar to those from 2d- and Spatial-MRec (Fig. 4o2). In particular, it indicates that, contrary to dOTC and MBCn, MRec does not present an increasing modification of temperature autocorrelations from 2d- to Full-versions."