Responses to referees' comments about the manuscript
"Multivariate bias adjustment of high-dimensional climate simulations: The "Rank Resampling for Distributions and Dependences" (R2D2) Bias Correction"
by Mathieu Vrac

First of all, I would like to thank the two referees for their careful reading and constructive comments. I tried to take all of their remarks into account. I think that based on the induced modifications, the manuscript is improved.
A point-by-point response to the reviewers' comments follows, where comments are in black and responses in blue.

Anonymous Referee #1

This paper proposes a novel Multivariate bias correction methodology. The author presents this methodology, referred to as R2D2, as the extension of a formerly developed and published methodology (EC-BC). R2D2 is meant to ameliorate some identified weaknesses in EC-BC. These are the "excessive constraint on temporal properties" and the lack of "stochasticity" in the adjustment. These are both laudable endeavours. The results are very interesting and, as I see it, deserving of publication. I have one major comment that, I believe, should be addressed by the author along with a series of minor and/or typographical comments that the author may wish to consider when editing the manuscript.

Response: I would like to thank the anonymous reviewer. I strongly appreciate the positive feedback. All the referee's comments have been carefully taken into account since they allowed clarifying and improving the manuscript. My responses and modifications brought to the manuscript are indicated below.

Major comment 1:
The author tests R2D2 on ERA-interim vs SAFRAN. These are both reanalysis datasets. This may be very interesting per se but is it really a good measure of how R2D2 will perform when applied to RCM data vs reanalysis or observations? I understand that the author does apply the new BC methodology to RCM data but not in a cross validation setting. I would like to understand why he has chosen not to apply R2D2 to RCM data, using SAFRAN as observations, calibrating with one part of the reanalysis data while cross-validating using another. Much like he did with ERA-Interim. In other words, why was ERA-Interim used at all? I find the tests using reanalysis data encouraging to be sure but not as satisfying as a test with RCM data would have been.

Response:
The ERA-Interim reanalyses were primarily used as data to be corrected because they ensure some consistency with the SAFRAN reference dataset. Indeed, when employing RCM data in a cross-validation context (or more generally when applying a BC method to RCM data over a projection time period different from the calibration time period), the changes in statistical properties (e.g., mean, variance, etc.) from the calibration to the projection time periods can be different for the reference and for the RCM data. Hence, when evaluating the results of a BC method over the projection period, it may be difficult to assess which remaining biases come from the disagreement between reference and RCM changes, and which come from the BC method itself. Using reanalysis data ensures more consistency with the reference and is therefore more appropriate for initial evaluation of a BC method. This is now discussed in the updated manuscript from line 31 of page 4 to line 6 of page 5.

However, the question of the cross-validation for the correction of the RCM simulations is highly relevant. Consequently, the same cross-validation framework as for ERA-I data has also been applied to RCM simulations. For illustration, the inter-variable Spearman correlation maps are presented in the figure below, showing the reference SAFRAN correlations (panel a), those from WRF (b), from the

1D-BC data (c) and finally from the 3012d-R^2D^2 (d), both calculated over the "evaluation" period (1995-2009) with a calibration on 1980-1994.
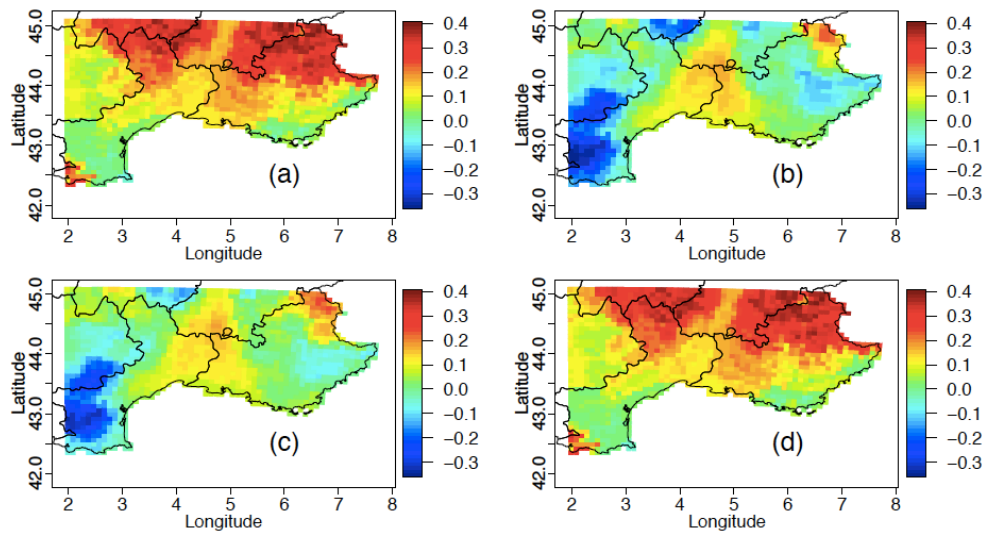

Figure: Inter-variable Spearman correlation maps in winter over the evaluation period (1995-2009) from: (a) SAFRAN; (b) WRF; (c) 1d-BC (CDF-t); (d) 3012d-R^2D^2.

It is clear that the conclusions are exactly the same as those provided either without cross-validation with the RCMs, or with cross-validation with ERA-I: WRF inter-variable correlations (b) are quite different from SAFRAN's; 1d-BC basically reproduce the WRF correlation maps; and R^2D^2 (both in its 2d and 3012d-versions but only shown for 3012d here) is very close to the SAFRAN correlation map. This is also the case for the EOF analyses in temperature or precipitation performed on the R^2D^2 bias corrected dataset with the same cross-validation. Those analyses display the good spatial behaviour of R2D2, i.e., with its first EOFs (in T2 or PR) very similar to those from SAFRAN, as illustrated in the two figures below.
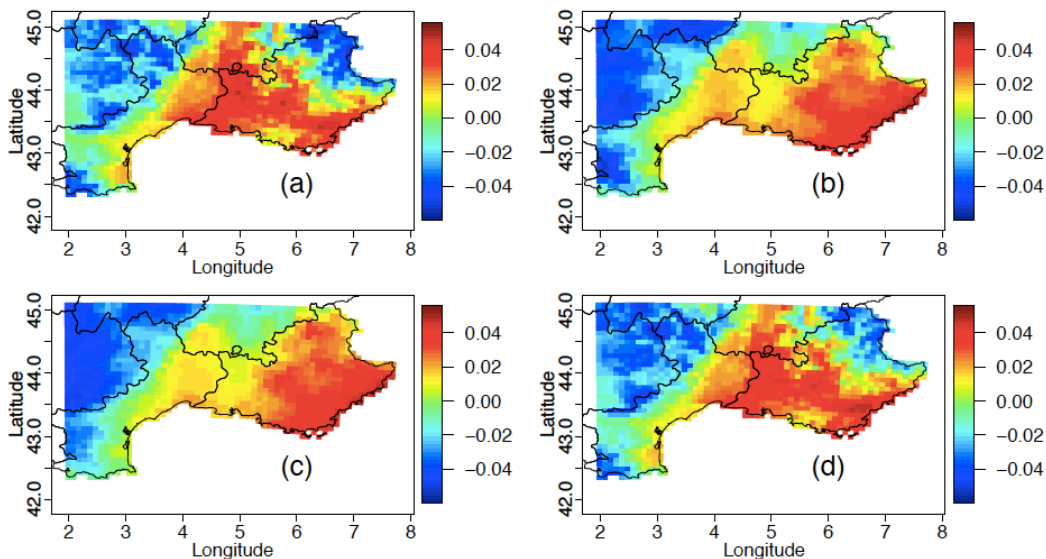

Figure: Maps of first **temperature** EOFs in winter over the evaluation period (1995-2009) from: (a) SAFRAN; (b) WRF; (c) 1d-BC (CDF-t); (d) 3012d-R^2D^2.
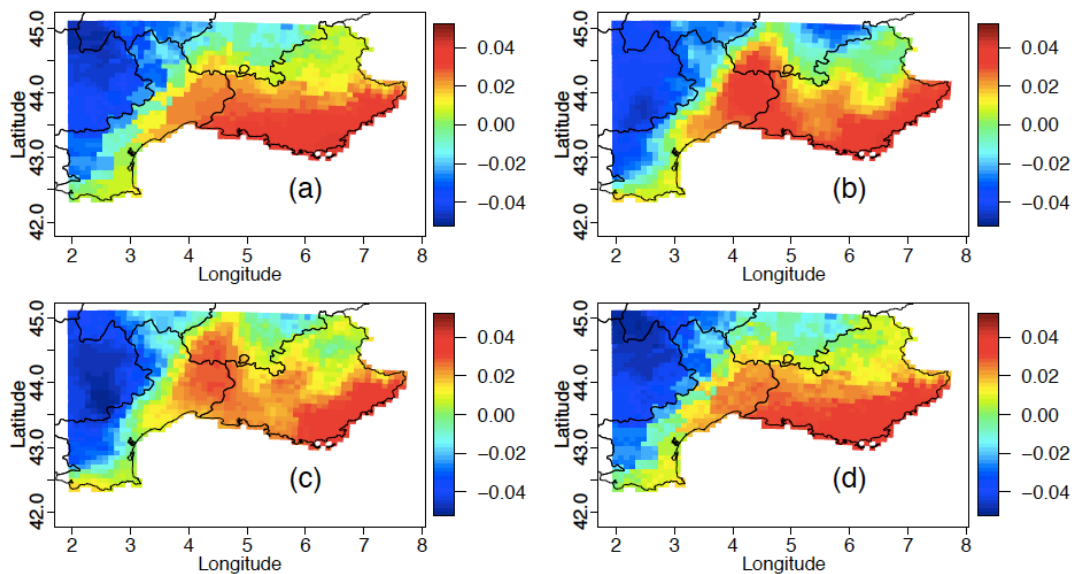
Figure: Maps of first **precipitation** EOFs in winter over the evaluation period (1995-2009)
from: (a) SAFRAN; (b) WRF; (c) 1d-BC (CDF-t); (d) 3012d-R^2D^2.

In summary, the cross-validation exercise performed on the bias correction method applied to the RCM simulations does lead to the exact same conclusions as the cross-validation performed on the ERA-Interim reanalyses. Therefore, as this is an important and relevant point, this is now mentioned in the updated manuscript on page 12, lines 25-29.
However, those additional figures are not included in the manuscript and only given here to the attention of the referee and editor.

Minor comments:
Page 1, Line 10, "allowing to deal" might be "making it possible to deal" or similar?
Response: Corrected.

Page 3, Line 15, "Whose the quality is not equivalent" could be " with different qualitative results". In any case the sentence might me rewritten.
Response: Corrected.

Line 25, I do not believe this is not the only problem with the Schaake Shuffle. The main problem is not that it is deterministic. The main problem, in my view, is that is unlikely to be robust in time. Why should the rank chronology be the same? This is why I think the authors should cross-validate with a RCM instead of reanalysis data.

Response:
I think, the reviewer and I agree. Indeed, this is more or less exactly what is written in this paragraph. I was NOT saying that the only or main problem is that it is deterministic (although it can be an issue). Around line 25 of page 3 of the initial submission, I wrote: "*...the price for this reproduction is that the temporal sequence of the ranks of the corrected data is exactly that of the reference data over the calibration time period, even for an adjustment performed over a future time period (or more generally over a projection/correction time period different from the calibration one)*". Indeed, there is no reason why the rank chronology should be the same (actually, there are plenty of reasons for which it should not be the same). This is now clarified in the updated manuscript on page 3, line 27.

Concerning the cross-validation with a RCM, see response to major comment.

Page 4, Line 6, Conclusion, perspective and discussion. The word "Perspective" does not translate identically from French :) . Perhaps "Conclusions, future work and discussion" or simply "Conclusions and discussions".

Response: Corrected, as well as in the title of the subsection 7.2.

Line 17, "co-located" could be "regridded" by simple association to nearest neighbour.
Response: Corrected.

Page 6, Line 2, "R2D2 looks for the time step t* in the calibration time period for which the rank of the reference dimension is the same as the current rank of the reference dimension" should add "Please refer to Appendix A for a detailed mathematical description of the R2D2 algorithm".
Response: To avoid making the description of this step heavier, the suggested mention "*Please refer to Appendix A for a detailed mathematical description of the R2D2 algorithm*" has been added just before starting the description of the different steps, on page 6, lines 7-8 of the updated manuscript.

Line 15, "wad" should be "was"
Response: Corrected.

Line 27, "needed" could be changed to "necessary"
Response: Corrected.

Line 30, If I understand correctly, this last step explains how restrictive the assumption of stationarity is in R2D2. In most BC methods "stationarity" means the "bias" stays the same. Here "stationarity" means the inter variable and inter site structure of the climate stays the same: : : This is a lot more restrictive I think. But these considerations do not matter if a cross-correlation with RCM were performed.
Response: This is true that this part explains that the dependence structure stays the same. However, as indicated in an earlier response, a cross-validation performed with the RCM data provided the exact same conclusions.

Page 7, Line 12, Is the parenthesis in "step 1.)" needed here? I found it a bit confusing. Also "others" might be "other"
Response: The parenthesis has been removed and "others" has been corrected to "other".

Page 10, Line 27, "Materials" might be "materials"
Response: Corrected.

Page 13, Line 31, The word "perspectives" here is a bit confusing. See above.
Response: Corrected.