

Responses are in red.

General Comments

This paper is about a Bayesian sparse regression technique and an application to statistical downscaling. The work seems to be interesting, and potentially novel. While it is quite valuable to have interdisciplinary submissions in this field, a significant caveat is that the main contribution needs to be identified in the submission. The downscaling application does not seem to be emphasized in the submission, e.g. it appears to be just an evaluation of their method. (cf. "We have evaluated our method both on synthetic and climate datasets."). If instead the contribution is on the technical (methodological/algorithmic/modeling) side, it needs to be more clearly stated in the exposition. In the Conclusion, the authors state "Our major contribution is to develop an efficient and scalable variational inference algorithm for inference in the fully Bayesian model." In order to facilitate evaluation of this claim, the authors would need to delineate and distinguish their contributions from past work, especially in the technical sections. Currently, sections 2 and 3 do not clearly distinguish between prior work in the technical area, including by the authors, and the claimed technical contributions in the submission. This makes it difficult to evaluate the novelty of the contribution. Finally, if the primary contribution is indeed on the methodological/algorithmic/modeling side, then it would seem much more appropriate to submit the manuscript to a data mining, statistics, AI, or machine learning publication, where it could be refereed by reviewers with the relevant expertise. This reviewer is not aware whether or not similar/parallel submissions have been made by the authors to such venues, but this should be clearly stated by the authors.

We agree with this reviewer that the main contribution in this paper is the development of a non-parametric Bayesian framework for finding multiple underlying sparse regression relationships that can potentially be used to select appropriate covariates for statistical downscaling of rainfall. To our knowledge, this is a novel method that combines Dirichlet Process mixture and Bayesian version of sparse regression to facilitate discovery of multiple sparse linear relationship within a complex dataset. The results on the climate dataset need to be viewed as mostly exploratory and preliminary in nature, which we have readily acknowledged both in the introductory (last paragraph) and in the concluding sections of the paper. This work by the authors is not currently under review in any other venues.

Specific comments:

It would be helpful to expand the related work discussion to a variety of Bayesian techniques for downscaling by Andrew Robertson (IRI, Columbia LDEO) and collaborators.

We thank the reviewer for pointing us to the work of Andrew Robertson and his collaborators. However, most of their work can be categorized under weather generator class of statistical downscaling methods, whereas our method can be classified under regression based methods. We have added a reference to the work of Greene *et. al.* in the second paragraph of the introduction section where we described related work.

Clarifications are needed in the technical section, e.g.:

- "stop when the probabilities $E[z]$ stop changing anymore"
- $E[z]$ is an expectation, not a probability.

In reality, the iterative algorithm stops when indicator variables \mathbf{Z} stop changing any more, not the expectation $E[\mathbf{Z}]$. This was a mistake in our write-up. We have fixed this error in the revised manuscript, and would like to thank the reviewer for pointing this out.

- How do you quantify that they "stop changing"? Is there a threshold? How do you set the threshold?

We did not use any threshold. Algorithm stops when indicator variables \mathbf{Z} does not change in 5 consecutive iterations.

- How is $5k$ chosen for the non-zero components?

We chose $5k$ (where k is the index of the cluster, $k=1,\dots,K$) for the non-zero components within the k -th cluster so that two clusters are distinctly identifiable in case the indices of non-zero components of the clusters are same. We will clarify it in the revised manuscript.

- It would be good to show experiments for various K values.

We have shown results of experiments for various values of K in Figure 2.

- How does your method compare to other clustering techniques, including non-generative techniques such as k-means++?

Our method is not comparable with parametric unsupervised clustering techniques such as k-means++ owing to at least two reasons:

A) Parametric methods require the number of clusters to be known beforehand. One of the principal advantages of our non-parametric technique is that it can automatically estimate the number of clusters from data.

B) Our method simultaneously learns a sparse regression model within each discovered cluster which is not possible for methods such as k-means++ that performs only clustering.

- Many figures are too small to read the axes.

We will increase the font size of the axes labels in the revised manuscript.

Technical corrections:

The submission needs extensive copy editing to fix numerous grammatical errors. E.g. there are many instances in which verbs do not agree with nouns (e.g. plurality). Due to these errors (along with typos, missing/repeated words), the submission is very tedious to read in its current form, in particular the introductory and non-technical sections. A few corrections are listed below; however the above should be done in a concerted way, e.g. by a copy-editor.

We will go through another pass and correct any grammatical errors we find in the revised manuscript.

GCM is defined twice with two definitions. "Global Climate Model" → "General Circulation Model."

Both definitions are occasionally used interchangeably in the literature, although since climate models are no longer just confined to the atmosphere, perhaps the use of "Global Climate Models" is more justified. We have changed "General Circulation Model" to "Global Climate Model" in the revised manuscript for uniformity.

"facing the mankind" → delete "the"

"Variational Bayes inferences" → "Variational Bayesian inference"

We made the suggested changes in the revised manuscript.