Rejoinder for "A Spatial Modeling Approach for Linguistic Object Data: Analysing dialect sound variations across Great Britain"

Shahin Tavakoli,
Department of Statistics, University of Warwick, UK,

Davide Pigoli,
Department of Mathematics, King's College London, UK,

John A. D. Aston*†
Statistical Laboratory, University of Cambridge, UK,

and

John S. Coleman
Phonetics Laboratory, University of Oxford, UK

August 9, 2019

We thank the discussants for their thoughtful comments and suggestions for future research directions. They raise a number of interesting questions, which we address here. For ease of reference, we will write PM when referring to the discussion of Petersen & Müller, and DPS when referring to the discussion of Dryden, Preston & Severn (and Marron when referring to the discussion from Marron).

1. **Choice of the metric:** DPS suggest that the choice of the metric on the space of covariance matrices can give an extra level of tuning for the practitioner (PM also point this out), and propose to look at the family of power Euclidean metric between covariances. Denoting by $d_\alpha(A, B) = |||A^\alpha - B^\alpha|||$, DPS come up with a nice expression for the d_α-covariance, for αs of the form $\alpha = k/(2^m)$ for integers k, m (dyadic rationals). DPS propose to look at the case $\alpha = 3/4$, which could provide a reasonable compromise

*The authors gratefully acknowledge support from EPSRC grant EP/K021672/2
†Address for correspondence: Professor John Aston, Statistical laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, CB3 0WB Cambridge, United Kingdom. Email: j.aston@statslab.cam.ac.uk

between our (popular) choice $\alpha = 1/2$ and $\alpha = 1$ (which gives the usual covariance matrix). Choosing a good α is a difficult question, and depends on the application that is being considered. Petersen & Müller (2016) propose methods for choosing α in an adaptive way, by using a criterion-based approach. The criteria they give as examples are based on desired qualities of the interpolated (d_E-)covariance matrix (such as eigenvalue decay), and it is therefore not straightforward to extend their methods to the case of d_α-covariance matrices, for which eigenvalues do not have a clear interpretation through a Karhunen–Loève expansion.

2. **Local linear smoothing:** PM suggest the use of local linear regression (local Fréchet Regression in the terminology of Petersen et al. 2019), which should give smaller bias near the boundaries. We agree that local linear smoothers are better than local constant smoother: however the local linear smoother proposed by PM assumes that the metric on the predictor space (Great Britain in our case, denoted by $\mathcal{E}$) is the Euclidean metric, since $z \to \beta_0 + \beta_1(z - x)$ is a line only under the Euclidean metric. In our case our distance on Great Britain is the shortest inland distance, which we approximate by the graph distance $\mathrm{d_g}$. While this distance is equal (approximately) to the Euclidean distance for two points for which the (Euclidean) straight line connecting them lies within Great Britain, this is not true for points near the boundary (and in particular, the Euclidean metric is *not* adequate, not because Great Britain is not big, but because it is non-convex). Hence the method proposed is not directly applicable. Since only distances in the predictor space are available, we would need to use methods to perform local linear smoothing when only pairwise distances between the predictors are available. Baíllo & Grané (2009), Boj et al. (2010, 2016) propose such methods, however they only deal with scalar responses, and they would need to be extended to deal with a non-Euclidean metric in the response space. This is indeed an interesting avenue for future research.

3. **1 step procedure:** The comment of PM regarding the one-step estimation of the d-covariances is interesting. To obtain such a one-step estimation, one however needs to have an estimate $\hat{\mu}(x)$ of the mean field at $x \in \mathcal{E}$, which in the case of single observation at each sampling point requires an initial smoothing step to compute (we omit the time index t for ease of notation). Since we have multiple observations at each sampling points in $\mathcal{E}$, we can obtain an unbiased estimate of $\mu(X_l)$ by taking the sample average at $X_l \in \mathcal{E}$, $\overline{Y}_t$. Then, the method advocated in our paper results in non-parametric smoothing of $(X_l, \breve{\Omega}_l), l = 1, \ldots, L$. The Nadaraya–Watson estimator at $x \in \mathcal{E}$ is $\hat{\Omega}(x) = \left[\sum_{l=1}^{L} w_l(x)\sqrt{\breve{\Omega}_l}\right]^2$, where

$$w_l(x) = \tilde{w}_l(x) / \sum_{l'=1}^{L} \tilde{w}_{l'}(x) \quad \& \quad \tilde{w}_l(x) = K_h\left(\mathrm{d_g}(x, X_l)\right). \tag{0.1}$$

If one uses the pointwise d_S-covariances $\breve{\Omega}_{lj} = \sqrt{(Y_{li} - \overline{Y}_l)(Y_{li} - \overline{Y}_l)^\mathsf{T}}$ for the one-step procedure of PM, using a Nadaraya–Watson estimator, the resulting estimator $\hat{\Omega}^*(x)$

of $\Omega(x)$ would actually also result in a convex sum over the sample d_S-covariances $\breve{\Omega}_l$, namely $\hat{\Omega}^*(x) = \left[\sum_{l=1}^{L} w_l^*(x)\sqrt{\breve{\Omega}_l}\right]^2$, which we call the PM estimator, where

$$w_l^*(x) = \frac{n_l}{\sum_{k=1}^{L} w_k(x) n_k} w_l(x).$$

Compare this with (0.1). We see that the weights in the PM estimator are scaled by a factor $n_l/(\sum_{k=1}^{L} w_k(x) n_k)$, the ratio between the number of observations at X_l, and the Nadaraya–Watson estimator of the "number of observations" at x. Hence, in the PM estimator, the locations that have more observations (respectively less observations) than the estimated number of observations at x will have their weight increased (respectively decreased) compared to our estimator. Whether this is in general desired or not would require some theoretical investigation, and is left for future research. In our case, this does not seem desirable, due to the observational nature of the dataset.

4. **Sharp Changes:** PM mention the possibility of using a model that assumes that the mean and d_S-covariance field are piecewise constant. This is a very nice suggestion, that was hinted at by some colleagues. It is well-known by English speakers that certain regions of Great Britain have distinct accents, and therefore it would make sense to try to find sharp transitions (if they exist) between distinct dialect regions This opens the door to detecting and estimating the boundaries where the sharp changes occur, and testing whether the mean and d_S-covariance fields are constant within these boundaries, and the suggestions made by PM give some initial ideas for investigating this.

5. **Cross-validation scale:** Marron points out that in our cross-validation plots, the tuning parameter h should be displayed on the log scale, since it is a scaling parameter. It is true that this would have been preferable, but we do not believe that it would change the results of the cross-validation.

6. **Sonic interpretations of the principal component (PC) loadings:** Marron raises an excellent question when asking about the possibility of drawing some simple sonic interpretations of the PC loadings. We provide as supplementary materials sounds that exhibit the effect of the PC loadings. If we denote by μ the MFCC of the vowel sound of a good quality recording of the word "class", and let φ be the PC loading (MFCC) which we would like to interpret. Letting `MFCCToSound` denote the function that maps MFCCs to (playable) sound waves, we can listen to the effect of the PC loading φ by listening to `MFCCToSound`$(\mu + \lambda\varphi)$ for various values of $\lambda \in \mathbb{R}$. Negative values of λ will correspond to low PC scores (in yellow on the left-hand side subplot of Figure 4 of the paper) whereas positive values of λ will correspond to high PC scores (in red in the same Figure). In order to make the interpretation easier, we replace the vowel sound of the recording used to get μ by the generated vowel sounds `MFCCToSound`$(\mu + \lambda\varphi)$, and play the entire word sound. These are sequentially played for

$$\lambda \in \{-10, -7.78, -5.56, -3.33, -1.11, 1.11, 3.33, 5.56, 7.78, 10\},$$

for PCs 1 and 2, and

$$\lambda \in \{-4, -3.11, -2.22, -1.33, -0.44, 0.44, 1.33, 2.22, 3.11, 4\},$$

for PC 3. The sounds can be found in the supplementary materials, in the file `sounds/effect-pcX.wav`, where `X`$= 1, 2, 3$. Listening to these files, we can hear that negative λs for PCs 1, 2 and 3 seem to correspond to the short and open from vowel [a] (as in "pat"), whereas positive λs seem to correspond to the long back vowel [ɑ] ("aah"), similar to the vowel in "part". Interpreting the left-hand side color maps in Figure 4 of the paper reveals that we indeed capture with the PC 1 and PC 2 scores the expected [a]-[ɑ] differences, while the interpretation of the map of PC 3 scores is less straightforward. The sonic interpretation of the PC loadings requires however carefully listening to these generated sounds (with high-fidelity headphones or loudspeakers), and is also confounded by the rather low signal-to-noise ratio in the sound tokens used for the analysis. Furthermore, the range of values for λ considered for these sound reconstructions is greater than the ones found in the projected mean MFCC field: this reflects the great amount of noise in our non-parametric regression (indeed, the PC scores of the original vowel sound MFCCs are $[-8.09, 13.27]$ for PC 1, $[-8.69, 8.59]$ for PC 2, and $[-5.39, 6.00]$ for PC 3). The noise is due to possible confounding effect of sex or age on the MFCCs, the microphone and room reverberation effect, in addition to the low signal-to-noise in the recordings.

We would like to thank again the discussants for their critical comments. We also extend our thanks to the associate editor and the three referees that all provided valuable feedback on earlier versions of the paper.

References

Baíllo, A. & Grané, A. (2009), 'Local linear regression for functional predictor and scalar response', *Journal of Multivariate Analysis* **100**(1), 102–111.

Boj, E., Caballé, A., Delicado, P., Esteve, A. & Fortiana, J. (2016), 'Global and local distance-based generalized linear models', *Test* **25**(1), 170–195.

Boj, E., Delicado, P. & Fortiana, J. (2010), 'Distance-based local linear regression for functional predictors', *Computational Statistics & Data Analysis* **54**(2), 429–437.

Petersen, A. & Müller, H.-G. (2016), 'Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data', *Biometrika* **103**(1), 103–120.

Petersen, A., Müller, H.-G. et al. (2019), 'Fréchet regression for random objects with euclidean predictors', *The Annals of Statistics* **47**(2), 691–719.