

Response to the Reviewers

We would like to thank both reviewers for their insights and observations. We have made extensive modifications to the manuscript based on their recommendations. In the following, we directly address their comments.

Reviewer 1

This paper develops a variational Bayesian functional principal components analysis using Karhunen-Loève decomposition without smoothing and estimating a covariance surface. More specifically, they incorporate the notion of variational message passing(VMP) over a factor graph because it removes the need for rederiving approximate posterior density functions if there is a change in the model. The performance of the proposed model is investigated by means of simulation studies and an application to temperature data collected from various United States weather stations. Overall, the paper is well written, but there are some concerns further to be considered to better understand the proposed approach with VMP. My concern is that the proposed approach and numerical results here, although somewhat useful, are not compelling from the perspective of novelty or application. Some novel features of the proposed approach can be emphasized and clearly explained in the paper. In the following, I have some specific suggestions/concerns to improve the paper for revision.

- In Section 2, each function is represented by the sum of a mean function and the sum of the infinite linear combination of the orthonormal eigenfunctions. However, in practice, this infinite sum is truncated by the first L orthonormal eigenfunctions, as the authors say. In the real dataset, naturally we don't know the value of L , thus to select this

L is also a important problem. Therefore, it would be helpful to explain some criteria for selecting the number of first eigenfunctions for truncation. Additionally, in the first subsection of the Section 2, they simply say that they approximate the unknown functions to splines using O’Sullivan basis without justification. Since there are many other Basis expansions such as wavelets basis or orthogonal basis functions and each expansion has its own desirable theoretical properties, at least some motivations/explanations would be helpful for readers to see why they consider O’Sullivan penalized splines

In the application in Section 7, we have included a procedure for determining an “optimal” value for L . In short, we progressively increase L until the proportion of explained variance for additional eigenfunctions is only incremental.

I have not made any modifications to justify our choice of O’Sullivan penalized splines. I overlooked this issue when modifying our manuscript. I have, however, included a brief statement in the Closing Remarks explaining that experimentation with other spline or wavelet families was beyond the scope of our article.

- They briefly comment that the method using Karhunen-Lo  ve decomposition has several advantages over the other class of methods using covariance decompositions. Although they provide why these advantages exist shortly, I think more justifications will be better. For instant, it will be better to provide why the method using Karhunen-Lo  ve decomposition is more appropriate handling sparse or irregular functional data over the conventional methods. Also, in the first subsection where the model construction proceeds, the prior specifications are not complete since the covariance matrices and the scale parameter of half-Cauchy distribution are not specified. Thus, it is desired to make a complete description of the hyperparameter specification.

In the Introduction, we explain that using bivariate smoothing to estimate the covariance function can be difficult or impossible for dense functional data

and for sparse, irregular functional data. Therefore, the applicability of frequentist approaches, which rely on an eigendecomposition of the estimated covariance function, is limited. We then explain that the covariance function is not estimated in the Bayesian approach since eigenfunctions and eigenvalues are estimated directly as part of a hierarchical Bayesian model. We justify our choice of a Bayesian approach by explaining that we do not need to compute and store large matrices, representing estimates of the covariance function.

I now realise that I read the second part of this comment incorrectly. I thought that the reviewer was asking us to explain how the iterated inverse- χ^2 hierarchical prior is related to the half-Cauchy prior. That is

$$\sigma^2|a \sim \text{Inverse} - \chi^2(1, 1/a), \quad a \sim \text{Inverse} - \chi^2(1, 1/A^2)$$

implies that

$$\sigma \sim \text{Half-Cauchy}(A).$$

I have now included this explanation in the manuscript. However, it looks as though they are simply asking for the exact values of the hyperparameters σ_β^2 and A that we used in the computations. Unfortunately, I have not included this.

- In Section 3, they mention that if there is a change to the model, the parameter vector updates must be rederived in the MFVB method, while this kind of problem does not arise in the VMP method. However, since VMP method is somewhat foreign to the statisticians, it is better to illustrate why this difference happens between the MFVB and VMP method. In Section 5, there are some proposals for this work to be more persuasive. Firstly, it will be better to compare the results in terms of accuracy and computation speed with that of conventional methods in Section 1. Although this work's own contribution is to address a functional data model via variational message passing and it works well considering the results in Section 5, as a newly proposed

methodology corresponding for FPCA model, it will be also indicative to show what aspects get better compared to the conventional FPCA methods. Along with this context, I think results under more complex settings should be given.

The major modification that we made to this submission was to extend the Bayesian FPCA model to the multilevel setting. In doing so, we were able to explicitly show the advantage of using the VMP approach. In this extension, we only had to derive the message passing updates for one new fragment. In fact, the *multiple Gaussian penalization fragment* that was introduced for the Bayesian FPCA model was recycled into the multilevel model, emphasizing the utility of the VMP approach for model extensions.

The major purpose for variational inference is as a fast and computationally cheap alternative to Markov chain Monte Carlo methods. Indeed, there are conventional methods for applying FPCA, but their limitations have been made clear in the Introduction. Therefore, the main goal in our simulations is to show the speed gains offered by variational inference with high accuracy in comparison to MCMC methods.

- In the simulation setup, all functions are just the linear combinations of a mean function and two eigenfunctions, while real functional data would have more complex structures. That is, the simulation settings used in this section seems to be too simple, considering the structure of real functional dataset. Thus, additional experiments and simulation studies under more diverse settings would be expected, such as with more/less observational points or more eigenfunctions. Plus, the estimation and computational costs seem to heavily depend on the number of basis of O’Sullivan spline. If the number of basis is increasing, the precision of estimation measured in terms of ISE will be stronger while the computational cost gets more burdensome and vice versa. This is because the volume of design matrix depends on the number of basis. Therefore, it will be also meaningful to find the optimal number of basis in terms of ISE and computational speed.

We have enhanced the complexity of our simulations by increasing the number of eigenfunctions for the Bayesian FPCA model to four and setting three eigenfunctions at each level of the multilevel FPCA model.

Determining the optimal number of O’sullivan spline basis functions, K , must be done on a case-by-case basis. From preliminary simulations, we determined that it was not the case that increasing K would improve the ISE results. When K was too high, we had issues with overfitting, thus reducing the ISE. In addition, we have used sinusoidal curves in our simulations to accommodate the similar types of curves observed in the US temperature dataset. However, other applications may not have sinusoidal eigenfunctions, meaning that the optimal value for K in these applications may be quite different from our own simulations.

- Additionally, convergence of the algorithm would be an issue since VMP algorithm is one of the deterministic algorithms. Therefore, showing concrete evidence that the algorithm successfully converges would be better for the validity of this algorithm. In Section 6, I think it would be also available to expand the proposed model to a multilevel one. With this expansion, analysis on the yearly data on each state, not the averaged data over 25 years seems to be possible. Naturally, it would be more meaningful to analyze the yearly data rather the averaged one since this averaged one has little variation. Also, analyzing the yearly data, we could catch some tendency as time goes on. Thus, if it is available, expansion to the multilevel model should be considered and new analysis with this expanded one will be desirable.

In our computational speed comparisons, we have included the median number of VMP iterations, with median absolute deviation, prior to convergence of the Kullback-Leibler divergence. We explain the convergence process in more detail in Section I.1 of the supplementary material.

We would like to especially thank the review for their recommendation of extending the Bayesian FPCA model to the multilevel setting. We have included this extension in the manuscript, with simulations and an applica-

tion to the US temperature data. It has been especially useful in explaining the utility of VMP algorithms for arbitrary model extensions.

- Finally, as the computational technicalities associated with the proposed VMP algorithm are quite challenging and important for fast computation and scalability, it would be useful to provide code or make it accessible in the github for public use, so that implementation and reproduction of the authors work can be useful to practitioners and other statisticians.

We have added all code to github at https://github.com/tui-nolan/vmp_fpca.

Reviewer 2

Summary. This work studies functional principle component analysis (FPCA), a technique that, analogous to classical PCA, can be used to reduce dimensionality and identify modes of variation but, unlike classical PCA, is used when the dataset is composed of curves that are independent realizations from a stochastic process. A Bayesian hierarchical model for estimating the eigenvalues and eigenfunctions of the principle components approximation is proposed and these values are estimated through a variational message passing (VMP) algorithm. In particular, this work proposes an FPCA extension to the VMP framework for variational Bayesian inference introduced in (Wand, 2017). Compared to MCMC, variational Bayes for FPCA is much more computationally efficient, and generally Bayesian methods are more efficient than standard approaches that smooth and estimate a covariance surface.

Review. I believe the work to be sufficiently novel for publication and of interest to the Bayesian Analysis readership. The paper is very clearly written and the mathematical details all appear to be correct. I have a few comments about how the authors could improve the paper prior to publication. One weakness of the paper is that the scope of the paper appears a bit

limited in that the work details an implementation of VMP for Bayesian inference on a specific hierarchical model for FPCA, instead of a more general VMP framework. I comment in more detail below about how I believe that the authors should include some additional discussions around how easy (or not) it would be to extend the applicability of the proposed ideas.

Detailed Comments. I enumerate my comments in the following.

- **Choice of L .** Throughout, the authors assume that the number of principle components to be used in the approximation, L , is determined beforehand. In practice, how should this be done? Would a practitioner run the full procedure for many choices of L and afterwards compare the estimates in some way? Is it clear that if the procedure is run for both L and $L + 1$ principle components that the top L eigenvalues and eigenfunctions found in both implementations would be approximately the same? I presume that the complexity analysis in Section 5.2 is for a fixed L , but how does the relative computational efficiency of VMP vs. MCMC depend on L ? Presumably VMP would outperform MCMC more as L grows, and it may be worth commenting on this if it is true.

As explained for the first reviewer, we have included a procedure for determining an “optimal” value for L in Section 7. In short, we progressively increase L until the proportion of explained variance for additional eigenfunctions is only incremental.

- **Generality of results.** The VMP algorithm proposed is specifically approximating posteriors from the full Bayesian hierarchical model specified in Equation (2.6). Could the authors comment on how easy it would be to apply these results if one wanted to make reasonable modeling changes to (2.6). For example, how dependent are the results on the use of Gaussian and inverse- χ^2 distributions or how easy would it be to extend the results if one added additional levels in the hierarchy? The authors include some small comments on this in Section 2.1 about

how the specific construction facilitates arbitrary non-informative priors, but I think a larger discussion about the assumptions for the specific model in (2.6) and the limitations that come with it, would be useful.

In this new manuscript, we have extended the VMP approach to accommodate multilevel functional data. This extension only required the introduction of one new fragment, while using most of the fragments from the Bayesian FPCA model, including the *multiple Gaussian penalization fragment*, which was specifically introduced for Bayesian FPCA. We believe that this sufficiently exemplifies the ease with which VMP algorithms can be extended for more complex FPCA models. Other prior specifications can be handled in a similar fashion by simply replacing the existing Gaussian prior fragments or iterated inverse- χ^2 fragments.

- **Convergence.** It is assumed in the article that the message passing algorithm will converge. Perhaps the authors could add a few comments on why that is reasonable to assume.

In our computational speed comparisons, we have included the median number of VMP iterations, with median absolute deviation, prior to convergence of the Kullback-Leibler divergence. We explain the convergence process in more detail in Section I.1 of the supplementary material.

- **Code and software.** Do the authors plan to share the code for their simulations and/or provide software for the proposed method? Surely, this would be useful for the community.

We have added all code to github at <https://github.com/tui-nolan/vmp-fpca>.

- **Some minor comments.** Do the authors plan to share the code for their simulations and/or provide software for the proposed method? Surely, this would be useful for the community.
 - If it is easy to do, I would suggest making Figure 2 and Figure 3 a bit larger.
 - There is a typo on page 18: "Each vector of principle component scores WAS simulated..."
 - I think something is wrong with the Section 7 sentence, "This study could be extended to other functional data models, such as function on scalar or vector regression models, that are yet to be treated under a VMP-based mean field variational Bayes approach."

We have addressed and corrected these issues in the new manuscript.