

January 10, 2025

Dear Editor,

We would be very grateful if you would consider this revision, entitled '*On the estimation of the number of components in multivariate functional principal component analysis*' for publication in *Communications in Statistics - Simulation and Computation*. We were able to address the comments made by the reviewers and strengthen our manuscript. A point-by-point reply is provided below.

We hope that this version matches reviewers' expectations, and we thank you in advance for your consideration.

Sincerely,

Steven Golovkine

(on behalf of the co-authors Edward Gunning, Andrew J. Simpkin and Norma Bargary)

Revisions to “On the estimation of the number of components in multivariate functional principal component analysis”

Steven Golovkine, Edward Gunning, Andrew J. Simpkin, and Norma Bargary.

The changes are in blue in the main text.

REVIEWER # 1

COMMENT # 1.1

Although I understand how s_1, \dots, s_p are obtained, the authors need to provide a more detailed explanation of how these values are generated using a Bernoulli distribution, for example.

Reply:

We have added the sentence “generated according to a Bernoulli distribution” to the text to clarify how these values are obtained.

COMMENT # 1.2

It would be helpful to simulate a case with irregular time points, where each function has different time points. Moreover, the time points for each function do not need to be regular. I recommend that the authors consider these situations.

Reply:

Thank you for your suggestion to include a case with irregular functional data. We have now performed the analysis with sparse irregular data. The sparsification consisted of randomly removing between 40% and 60% of the S sampling points for each curve. Figure 1 presents the results of the estimation of the eigenvalues. The case where $S = 25$ is not included because the number of points (after sparsification) is not enough to consistently estimate 10 univariate components for each univariate feature. The results are similar to the dense case (when all the sampling points are

observed on a regular grid). The estimation of the last multivariate eigencomponents is not good. We choose to not include the graph in the paper to not overload it but we display it below. We prefer to keep the message of the paper simple: even in simple situations (such as dense functional data), be careful when choosing the number of univariate components for the estimation of the multivariate components.

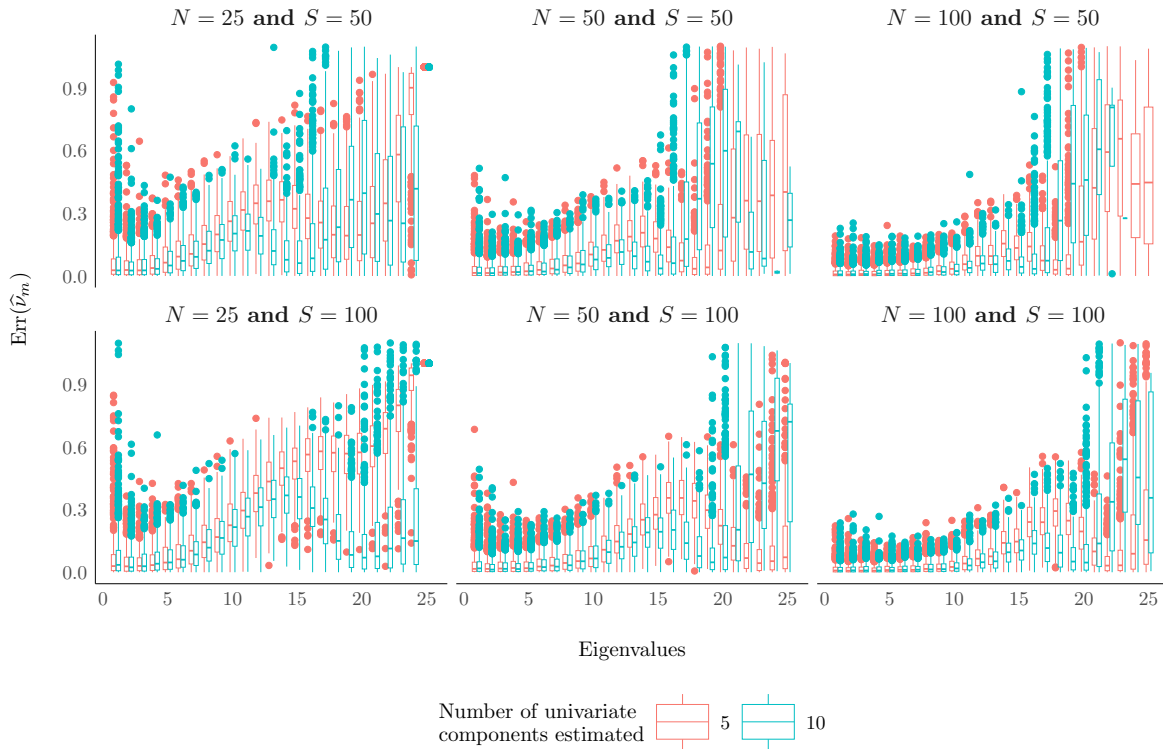


Figure 1: Boxplots of the estimation errors of the eigenvalues. We estimated $M_j = 5$ (red boxplots) and $M_j = 10$ (blue boxplots) univariate functional components for $j = 1, \dots, p$. The number of multivariate eigencomponents that are estimated is 25. N is the number of observations, S is the number of sampling points per curve before the sparsification. The sparsification is performed by randomly removing between 40% and 60% of the sampling points of each curve. We run 500 simulations.

COMMENT # 1.3

The paragraph in the end of Page 2 and Page 3 requires a complete rewrite as it lacks the necessary rigor in its explanations and methodology.

Reply:

We have rewritten the paragraph to include more details on the methodology developed in Happ and Greven, 2018 which we comment on.

COMMENT # 1.4

The simulation study and application sections of the paper need to be significantly revised. The current writing provides very limited explanation, which makes it difficult for readers to fully understand the intuition behind the results, the significance of the findings, and how the method performs in practice. The authors should offer more detailed descriptions of the simulation settings including the rationale behind the chosen parameters and models. Moreover, the application section should clearly describe the dataset, the variables used, and how the results from the method apply to real-world scenarios. More emphasis should be placed on interpreting the application results and their implications which help readers understand the practical value of the proposed method.

Reply:

For the simulations, we have added details on the results and their implications for practitioners. For the application, we have added a figure to visually present the dataset and included a new paragraph that provides a detailed interpretation of the eigencomponents. This paragraph highlights how the interpretation of the results differs between the two scenarios and explains their practical implications.

COMMENT # 1.5

The authors classify the number of simulations based on the size of the circles, which is very confusing. Please present the results using a clearer visualization method.

Reply:

We have replaced the plot with tables that present the results more clearly to enhance the readability and interpretability of the findings.

COMMENT # 1.6

The authors use B-spline bases for the Canadian weather dataset. I recommend that they also use a Fourier basis and compare the results. This can demonstrate the robustness of the method, as Fourier bases typically provide a better fit for weather data.

Reply:

Thank you for your suggestion to include a comparison using a Fourier basis. As our primary aim is to provide commentary on the estimation of the number of principal components utilising the methodology proposed in Happ and Greven, 2018 and not finding the best representation of the data, we choose to not include the comparison into the main manuscript. We have included here the estimated eigenfunctions using 10 Fourier basis functions for smoothing the Canadian weather dataset (see Fig. 2). The estimated eigenfunctions obtained with the Fourier basis are consistent with those derived using the B-spline basis, demonstrating that the choice of basis does not affect the results in this case.

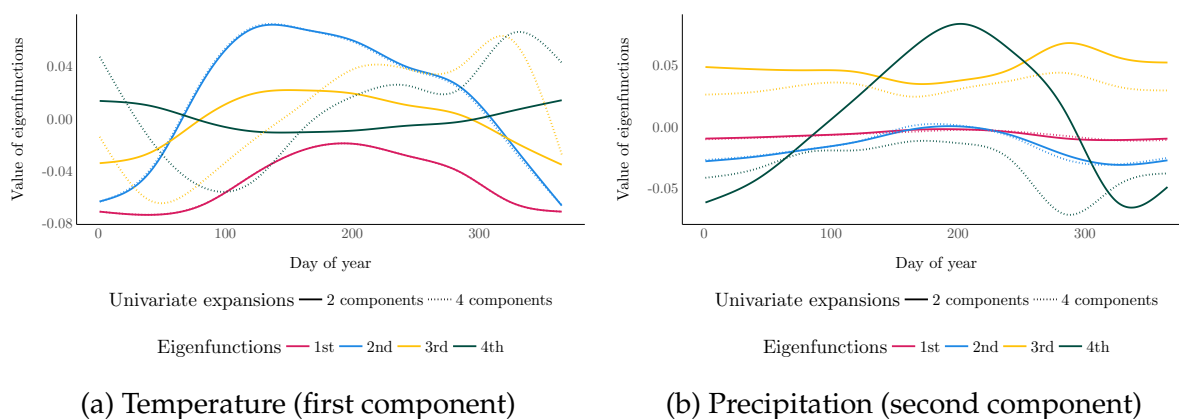


Figure 2: Estimation of the first four eigenfunctions of the Canadian weather dataset using using two and four univariate components for the univariate expansions after Fourier basis smoothing.

COMMENT # 1.7

Figure 3 is missing a y-axis.

Reply:

We have updated Figure 3 to include the y-axis as “Value of eigenfunctions”.

REVIEWER # 2

COMMENT # 2.1

The manuscript needs English proofreading.

Reply:

The manuscript has been carefully revised by the three authors who are native English speakers.

COMMENT # 2.2

What is the limitation of the proposed approach?

Reply:

We would like to clarify that the primary aim of our manuscript is to “provide commentary on the estimation of the number of principal components, utilising the methodology proposed in Happ and Greven, 2018”. We are not proposing a new methodology in this work but instead aim to critically evaluate and highlight the limitations of the approach introduced in Happ and Greven, 2018.

REFERENCES

Happ, C., & Greven, S. (2018). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, 113(522), 649–659.