

Imputation for Functional Data Analysis

Supervision: Marlène Baumeister and Markus Pauly jointly with Łukasz Smaga and Natalia Stefanska (Adam Mickiewicz University, Poznań)

Research environment: International collaboration within the WEAVE project by Łukasz Smaga and Markus Pauly

In some modern applications, the basic statistical unit is not a single number but an entire curve or function. Examples include trajectories measured over time, growth curves, longitudinal medical measurements, EEG or other biomedical signals, wearable-sensor data, spectroscopic curves, or measurements recorded over spatial domains. Functional Data Analysis (FDA) provides statistical methods that explicitly exploit this structure. In practice, however, such functional observations are often incomplete.

While imputation techniques like multiple imputation (MI) are well established for scalar data, their use in functional data analysis (FDA) are less developed. In particular, questions arise when the subsequent analysis involves functional regression or factorial designs with functional responses.

The aim here is to investigate methodological and/or computational aspects of multiple imputation for functional data. Depending on the student's interests and level, the thesis may focus on one of the following directions:

- 1. Simulation study for factorial designs with functional responses (B.Sc. or M.Sc. theses).**
Study how different imputation strategies affect estimation and inference in functional factorial designs. Relevant factors should include the proportion and mechanism of missingness (MCAR/MAR), sample size, covariance structure and type of missing functional observations (missing a complete function or only partial missings). Performance can be assessed in terms of bias, coverage probabilities, type I error and power and the concrete designs (more simple for B.Sc. Students) will be discussed based on interest.
- 2. Multiple imputation in functional regression models (M.Sc. thesis).**
Investigate MI when predictors are incomplete. Possible models include scalar-on-function, function-on-scalar or function-on-function regression. Here, we could also study estimation, inference or even prediction properties.
- 3. Pooling rules beyond approximately normal estimators (M. Sc. thesis).**
Rubin's classical combination rules for valid inference after MI are directly applicable to estimators that are approximately normally distributed. Many statistics relevant in FDA (like norms, suprema, quadratic-form or integral statistics) do not satisfy this. This more methodological thesis should investigate transformations, alternative pooling procedures, bootstrap-based approaches, or new adaptations of Rubin-type rules for such statistics.

The exact scope can be adapted to the student's background. A simulation-oriented project would

be particularly suitable for students interested in R/Python, statistical computing and experimental design, whereas the third direction offers opportunities for a more methodological/theoretical thesis.

Contact: Marlène Baumeister (marlene.baumeister@tu-dortmund.de) and Markus Pauly (pauly@statistik.tu-dortmund.de)

Some References

- Jang, J. H., Manatunga, A. K., Chang, C., & Long, Q. (2021). A Bayesian multiple imputation approach to bivariate functional data with missing components. *Statistics in medicine*, 40(22), 4772-4793.
- Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2, 321–359.
- Munko, M., Ditzhaus, M., Pauly, M., & Smaga, Ł. (2025). Multiple comparison procedures for simultaneous inference in functional MANOVA. *Electronic Journal of Statistics*, 19(1), 2637-2732.
- Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer.
- Rao, A. R. & Reimherr, M. (2021). Modern multiple imputation with functional data. *Stat*, 10, e331.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.