

---

# Addressing the Impacts of Collinearity on Interpretations of Variable Importance

Jesper B. Pedersen<sup>\*1</sup>, Felix Riede<sup>2,3</sup>, and Peter M. Yaworsky<sup>2,3,4</sup>

<sup>1</sup>Ro.C.E.E.H., University of Tübingen – Germany

<sup>2</sup>Department of Culture and Heritage studies, Aarhus University – Denmark

<sup>3</sup>Center for Ecological Dynamics in a Novel Biosphere, Aarhus University – Denmark

<sup>4</sup>Center of Molecular Ecology and Evolution, University of Copenhagen – Denmark

## Abstract

Predictor variable selection and interpretation is a core component of building correlative ecological niche models (ENMs), but how we select predictor variables, handle collinearity, and interpret variable importance measures is relatively poorly defined. The issue is particularly relevant for archaeological applications of ENMs due to the rather limited availability of paleoclimate predictor variables, which primarily capture aspects of temperature, precipitation, and environmental productivity (Beyer et al., 2020; Brown et al., 2018; Karger et al., 2023; Krapp et al., 2021). As important as they are, the rather limited scope of paleoclimatic variables results in collinearity among many of the predictor variables and is a central issue when using these data. In addition, thresholding methods for eliminating correlated variables (Dormann et al., 2013) results in variable selection which may not accurately reflect causal mechanisms. Both issues require that we carefully consider what environmental variables are really capturing and whether variable importance measures are capturing causal mechanisms or simply minor differences in model performance as a function of correlated predictor variables. Here, we outline the trade-offs between using another method, dimensional reduction with Principal Components Analysis, to remove collinearity from predictor variables, and how it provides a new look at how we might better interpret measures of variable importance and identify shifts in the fundamental niche space (Jackson & Overpeck, 2000).

## References:

Beyer, R. M., Krapp, M., & Manica, A. (2020). High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Scientific Data*, 7(1), 236. <https://doi.org/10.1038/s41597-020-0552-1>

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., & Haywood, A. M. (2018). PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data*, 5(1), 180254. <https://doi.org/10.1038/sdata.2018.254>

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013).

---

\*Speaker

Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>

Jackson, S. T., & Overpeck, J. T. (2000). Responses of Plant Populations and Communities to Environmental Changes of the Late Quaternary. *Paleobiology*, 26(4), 194–220.

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2023). CHELSA-TraCE21k – high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past*, 19(2), 439–456. <https://doi.org/10.5194/cp-19-439-2023>

Krapp, M., Beyer, R. M., Edmundson, S. L., Valdes, P. J., & Manica, A. (2021). A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. *Scientific Data*, 8(1), 228. <https://doi.org/10.1038/s41597-021-01009-3>