Eawag
Überlandstrasse 133
8600 Dübendorf
Switzerland
Phone +41 (0)58 765 53 61
Fax +41 (0)58 765 53 75
info@eawag.ch
www.eawag.ch

Eawag Seminar Invitation

Scalability, accuracy and assumptions of joint species distribution models – a synthesis

Speaker **Prof. Dr. Florian Hartig,** *University of Regensburg, Germany*

When **November 19, 16.00 – 17.00**

Where **Online via Zoom, contact seminars@eawag.ch for access details.**

Abstract In recent years, interest in ecology has shifted from analyzing the association of single species with their environment (as in standard species distribution / niche models) to the analysis of entire communities. Community data allows, at least in principle, to simultaneously infer the importance of biotic and abiotic factors for species distributions, opening up many new avenues for testing ecological theories and making predictions.

Most current statistical methods approach this task through a multivariate GLMM structure known as a joint species distribution models (jSDMs). A jSDM is a GLMM with a species-species covariance term that is often further regularized through a latent variable approach.

While generally promising, an issue for most current jSDM software is that they scale poorly on large datasets, which limits their use for emerging novel data sources, such as large acoustic or eDNA datasets. A second question is the accuracy of the inference – to reduce the degrees of freedom in the covariance matrix, latent-variable models introduce an auxiliary latent-variable structure into the model, and it is still poorly understood to which extent this structure creates bias or limitations on the fitted species-species associations.

In my talk, I will first summarize the assumptions and theory for estimating jSDMs. I will then compare how current software solutions scale for large datasets and examine if latent-variable approaches create constraints on the species-species covariance matrix that limits their accuracy for inferring species associations. Finally, I will present an alternative and more flexible method for fast and scalable estimation of jSDMs with a multivariate probit structure.