



## Uncertainty quantified trait predictions

Farideh Fazayeli (1), Jens Kattge (2), Arindam Banerjee (1), Franziska Schrodtr (2), and Peter Reich (3)

(1) Computer Science & Engineering, University of Minnesota, Twin Cities, US, (2) Max Planck Institute for Biogeochemistry, Jena, Germany, (3) Department of Forest Resources, University of Minnesota, Twin Cities, US

Functional traits of organisms are key to understanding and predicting biodiversity and ecological change, which motivates continuous collection of traits and their integration into global databases. Such composite trait matrices are inherently sparse, severely limiting their usefulness for further analyses. On the other hand, traits are characterized by the phylogenetic trait signal, trait-trait correlations and environmental constraints, all of which provide information that could be used to statistically fill gaps. We propose the application of probabilistic models which, for the first time, utilize all three characteristics to fill gaps in trait databases and predict trait values at larger spatial scales.

For this purpose we introduce BHPMF, a hierarchical Bayesian extension of Probabilistic Matrix Factorization (PMF). PMF is a machine learning technique which exploits the correlation structure of sparse matrices to impute missing entries. BHPMF additionally utilizes the taxonomic hierarchy for trait prediction. Implemented in the context of a Gibbs Sampler MCMC approach BHPMF provides uncertainty estimates for each trait prediction.

We present comprehensive experimental results on the problem of plant trait prediction using the largest database of plant traits, where BHPMF shows strong empirical performance in uncertainty quantified trait prediction, outperforming the state-of-the-art based on point estimates. Further, we show that BHPMF is more accurate when it is confident, whereas the error is high when the uncertainty is high.