ARPHA Conference Abstracts 8: e156762
doi: 10.3897/aca.8.e156762

OPEN ACCESS

Conference Abstract

Explainable machine learning for modeling of net ecosystem exchange in boreal forests

Topi Laanti[‡], Ekaterina Ezhova[§], Anna Lintunen[§], Pasi Kolari[‡], Tuomo Nieminen[‡], Ivan Mammarella[‡], Keijo Heljanko[‡], Markku Kulmala[§]

‡ University of Helsinki, Helsinki, Finland
§ INAR Physics, Helsinki, Finland

Corresponding author: Topi Laanti (topi.m.laanti@helsinki.fi)

Received: 23 Apr 2025 | Published: 02 Jun 2025

Citation: Laanti T, Ezhova E, Lintunen A, Kolari P, Nieminen T, Mammarella I, Heljanko K, Kulmala M (2025) Explainable machine learning for modeling of net ecosystem exchange in boreal forests. ARPHA Conference Abstracts 8: e156762. https://doi.org/10.3897/aca.8.e156762

Abstract

Accurate modeling of net ecosystem exchange (NEE) in boreal forests is crucial for understanding the global carbon cycle and predicting how forest ecosystems respond to a warming climate. In this work, we applied four machine learning (ML) methods—cubist, random forest, averaged neural networks, and linear regression—to predict subhourly NEE at two Finnish boreal forest sites with contrasting latitudes (SMEAR II Hyytiälä in the south and SMEAR I Värriö in the north). We modelled separately NEE during the peak growing season and the entire year.

To gain insight into the drivers of NEE, we employed two explainable artificial intelligence (XAI) methods called Permutation Feature Importance and ALE plots. During the peak growing season, photosynthetically active radiation, diffuse radiation, and vapor pressure deficit (or air temperature) were identified as the variables with the largest explaining power. When modeling NEE over the full year, vapor pressure deficit was replaced by soil temperature as a key driver. Combining data from the two sites into one set revealed soil water content as an influential factor; however, its effect diminished upon adding site-identifying input variables. Notably, we observed that some models exhibit contradictory dependencies on interdependent variables such as air and soil temperature, underscoring the importance of carefully interpreting ML results, particularly for extrapolations to future climate scenarios.

© Laanti T et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

These findings highlight that high accuracy alone does not ensure reliable predictions under changing environmental conditions. We emphasize the value of XAI techniques to dissect how ML models handle complex input variables, providing interpretability that traditional “black-box” approaches often lack. Moreover, the results underscore the need for robust model evaluation across spatiotemporal scales—especially when data from multiple sites and climatic regimes are combined—and call for a careful assessment of ML-based predictions under conditions that diverge substantially from the training data.

Presenting author

Topi Laanti

Presented at

POSTER

Author contributions

Ekaterina Ezhova 1, Topi Laanti 2, Anna Lintunen 1, Pasi Kolari 1, Tuomo Nieminen1, Ivan Mammarella 1, Keijo Heljanko 2,3, and Markku Kulmala 11 INAR Physics, University of Helsinki, Helsinki, Finland2 Department of Computer Science, University of Helsinki, Helsinki, Finland3 Helsinki Institute for Information Technology (HIIT), Helsinki, Finland

Conflicts of interest

The authors have declared that no competing interests exist.