Goldschmidt 2024 Abstract
https://doi.org/10.46247/gold2024.22931

Disentangling the sources of *n*-alkane in lacustrine sediments through a novel machine learning approach

ZHE WANG[1], YANG PANG[1], KAIFENG PANG[2], JIAWEI JIANG[3] AND BIN ZHOU[1]

[1]Nanjing University

[2]University of California, Los Angeles

[3]The University of Hong Kong

Presenting Author: xaviiniestamessi1@gmail.com

A significant challenge in reconstructing terrestrial vegetation using the long-chain *n*-alkane preserved in lake sediments is the notable contribution from non-emergent aquatic plants and algae, as observed in modern lakes. We drilled a 600-m core of lacustrine sediments spanning from ~2 Ma to present in the Weihe Basin on the southern margin of the Chinese Loess Plateau (CLP). However, to fully leverage the high-resolution and well-preserved *n*-alkane records in the core, in contrast to adjacent eolian deposits (i.e. loess), it is essential to resolve the latent interference. Previous research focused on the mixing model of long-chain $\delta^{13}C_{alkane}$ to quantify the relative contribution. However, we hypothesize that the difference between the compositions of *n*-alkane in distinctive depositional environments can be a more basic and rigorous trial. Here, we propose a machine learning classifier to categorize *n*-alkane from two distinct depositional environments: loess deposits and modern lake sediments. This classification is based on the fractional abundances of C_{17}–C_{33} *n*-alkane molecules, which serve as 17 input features for the model. We collected published data from loess sections (n = 144) and mid-latitude Asian lakes (n = 104). After model selection, hyperparameter tuning, and cross validation, we developed a random forest model that demonstrated outstanding performance (ROC-AUC > 0.9) on the test set. This highlights the intrinsic difference in *n*-alkane compositions between the two depositional environments, which can be intuitively understood that in contrast to loess, lake sediments always contain autochthonous, other than terrestrial constituents. We utilized the optimized model to ascertain the dominant sources of *n*-alkane within the core. The consistent identification of 'loess' type corresponds to the high contributions (~80%) of *n*-alkane from ambient loess deposits independently quantified by the MixSIAR model of long-chain $\delta^{13}C_{alkane}$ through the sequence. We assume that the high input of coeval eolian deposits, evidenced by high sedimentation rates in the core, veils the contributions from aquatic producers present in the lake then. All the evidence strongly suggests that *n*-alkane records in the core can compensate for the dearth of vegetation history on CLP.