

Utilizing Probability Estimates from Machine Learning and Pollen to Understand the Depositional Influences on Branched GDGT in Wetlands, Peatlands, and Lakes

5 Amy Cromartie¹, Cindy De Jonge², Guillemette Ménot³, Mary Robles^{4,5}, Lucas Dugerdil^{3,5}, Odile Peyron⁵, Marta Rodrigo-Gámiz⁶, Jon Camuera⁷, Maria Jose Ramos-Roman⁸, Gonzalo Jiménez-Moreno⁶, Claude Colombié⁵, Lilit Sahakyan⁹, Sébastien Joannin⁵

Supplementary Figures:

10

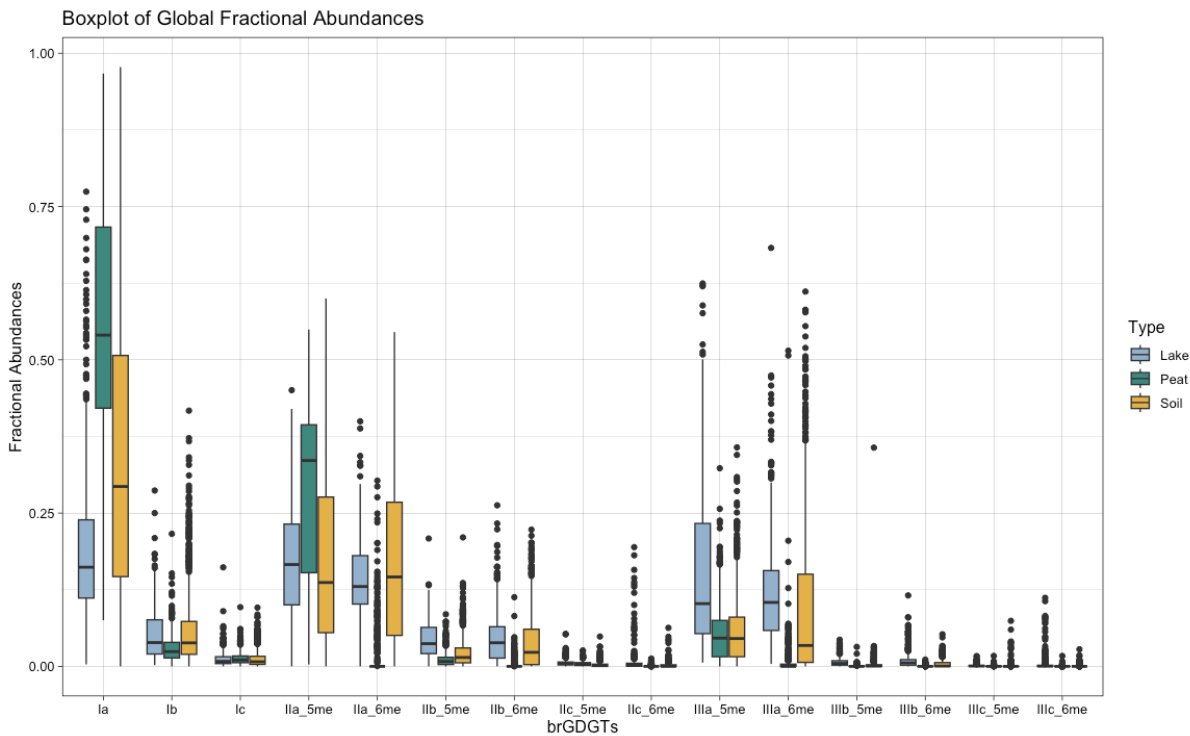
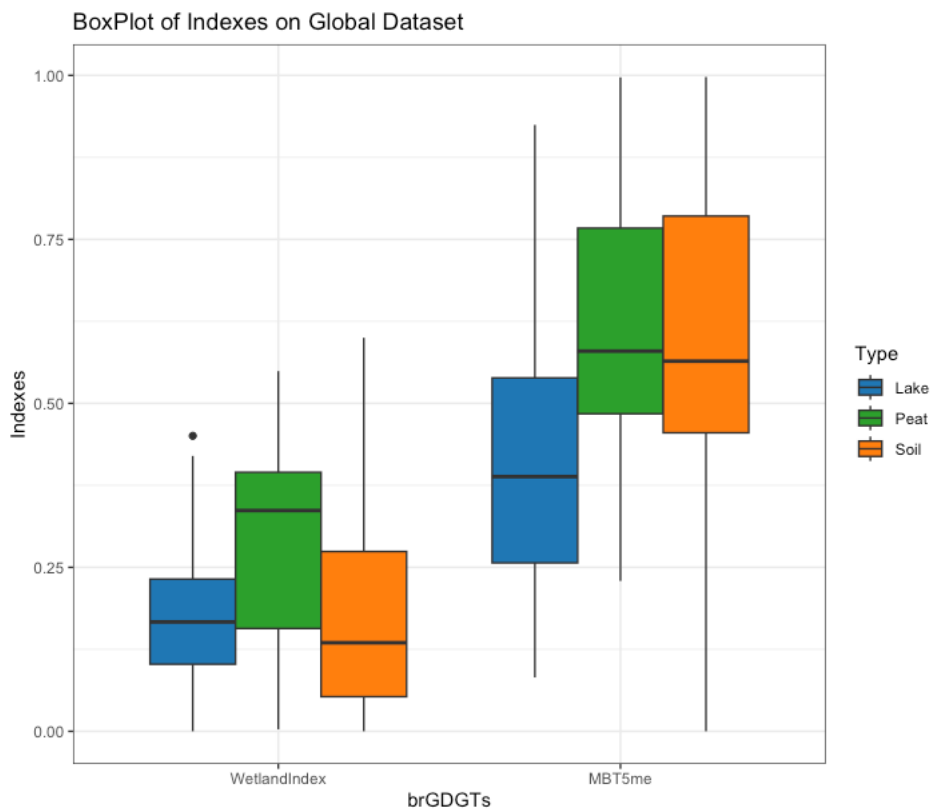


Figure S1. Box plot of the distribution of fractional abundances on the global brGDGT database classified by type (lake, peat, soil)



15

Figure S2. Boxplot of indexes utilized in paper on the global modern database

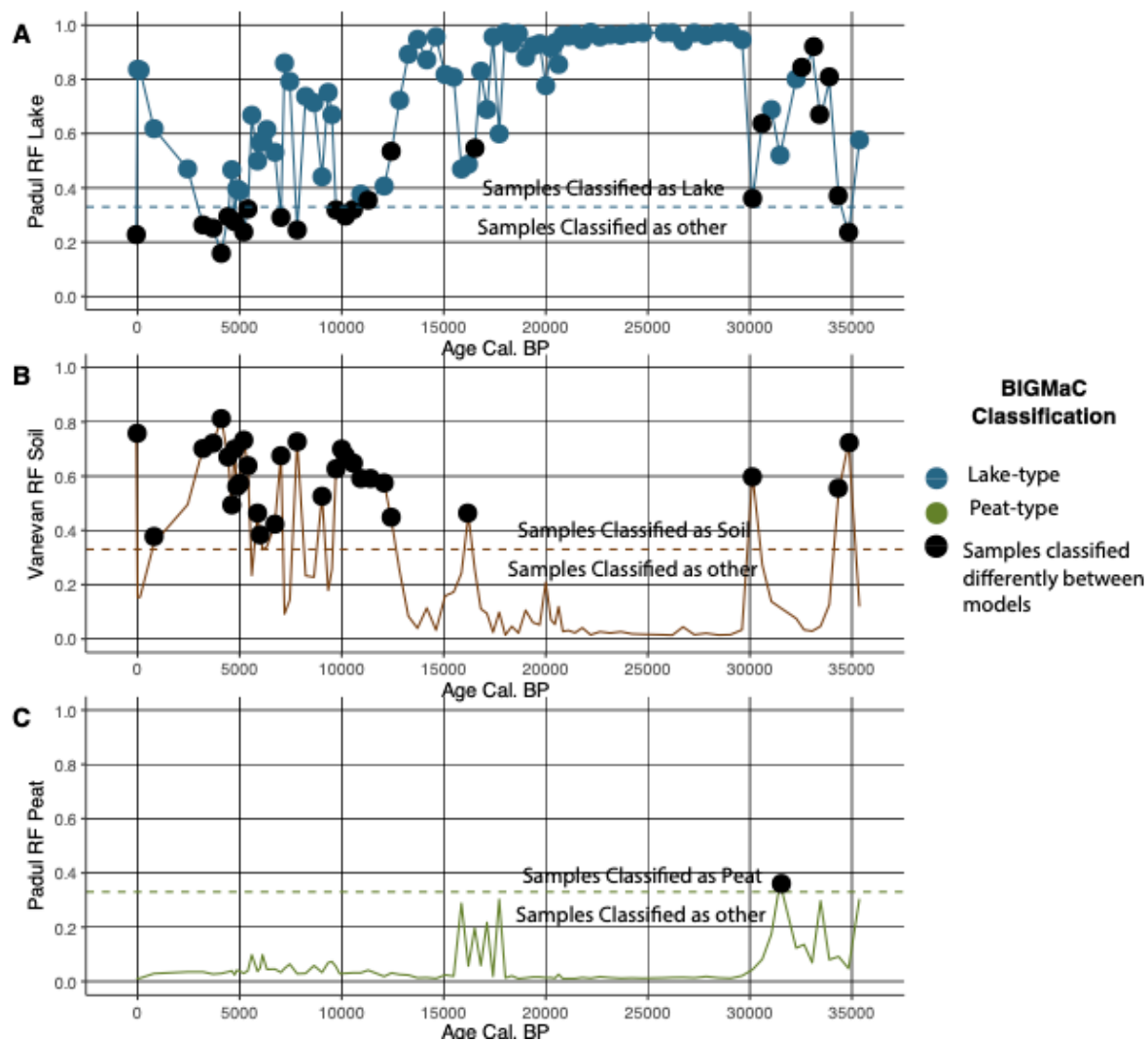


Figure S3. Probability estimates of brGDGTs from the Padul record from this article compared to the BIGMaC classification model of Martínez-Sosa et al. (2023). Lines are the probability estimates and the dots are the classification from the BIGMaC. Colors are blue = lake, brown = soil, and green = peat. Black dots are samples that are not in agreement between models.

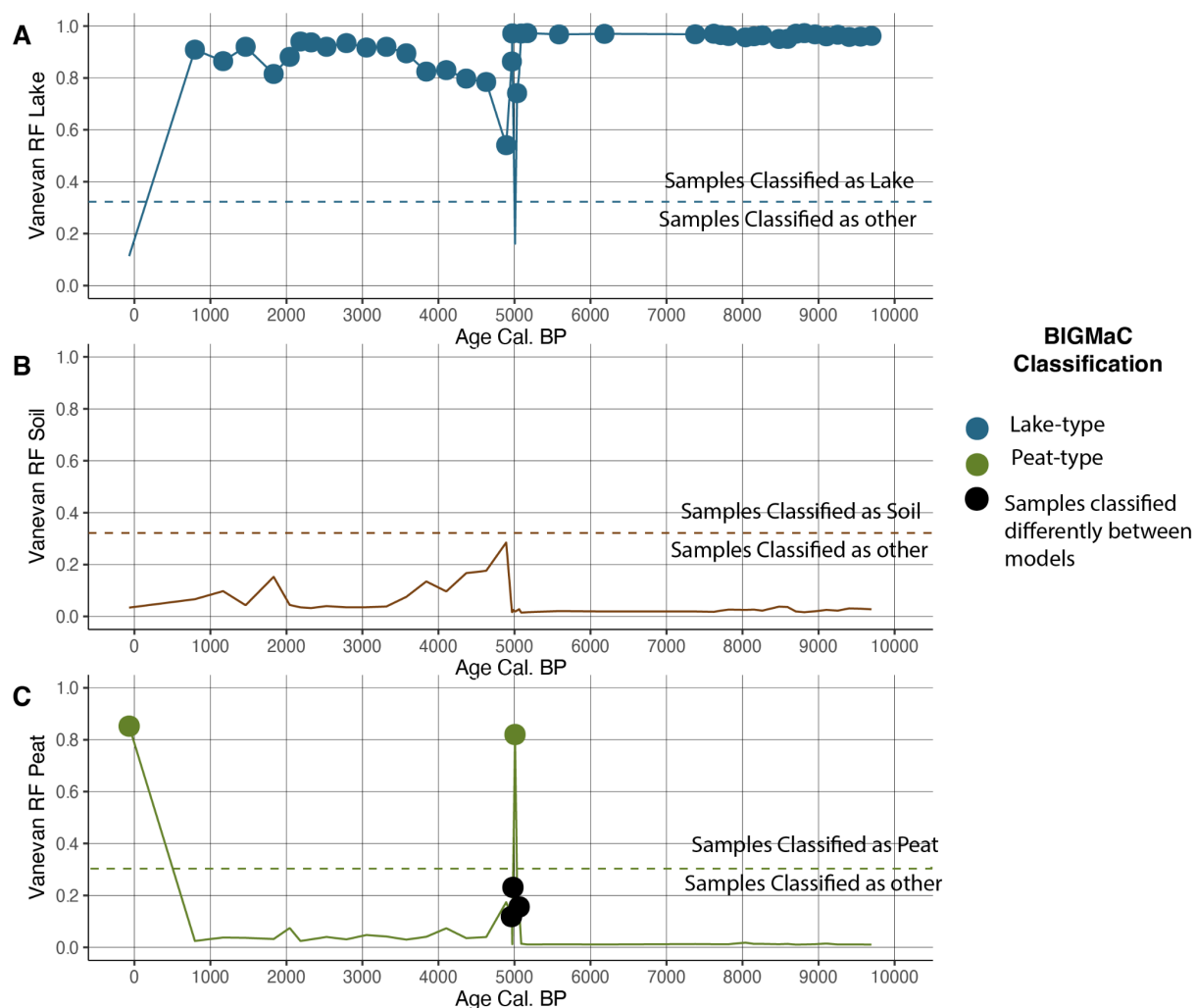
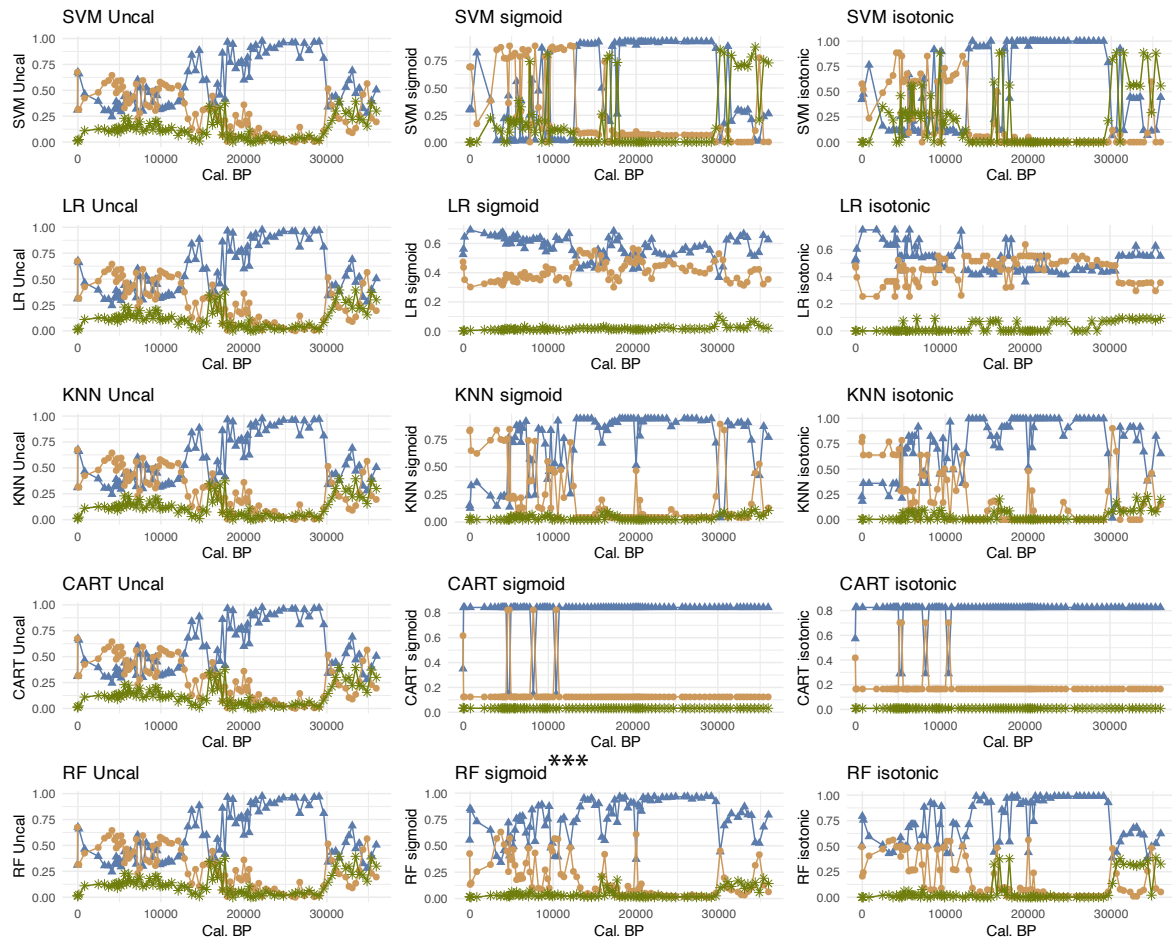


Figure S4. Probability estimates of brGDGTs from the Vanevan record from this article compared to the BIGMaC

25 classification model of Martínez-Sosa et al. (2023). Lines are the probability estimates and the dots are the classification from the BIGMaC. Colors are blue = lake, brown = soil, and green = peat. Black dots are samples that are not in agreement between models.

Padul Probabilities - Regular Dataset



30 Figure S5. Probability estimates for the Padul record on models tested in the paper on the regular unsampled dataset

Vanevan Probabilities - Regular Dataset

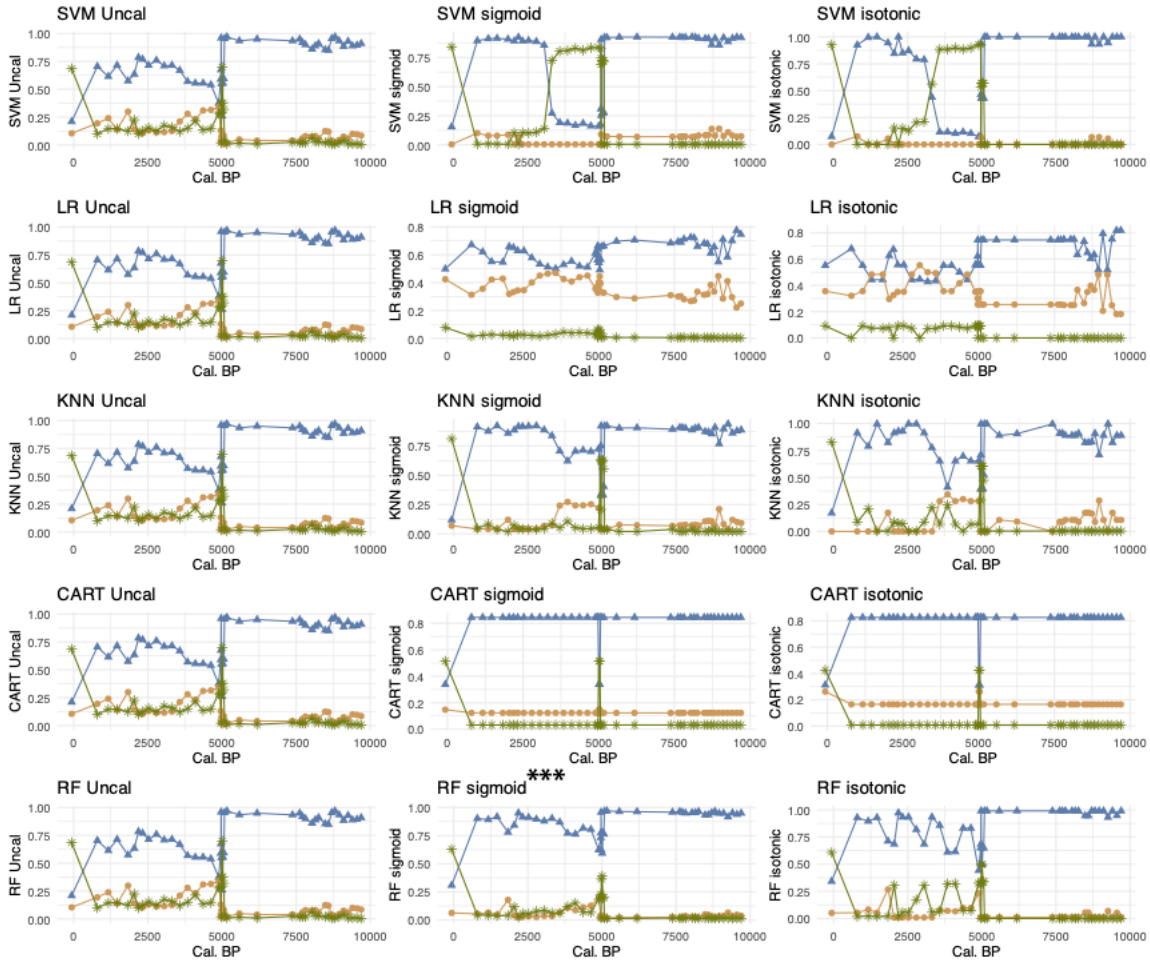


Figure S6. Probability estimates for the Vanevan record on models tested in the paper on the regular unsampled datasets.

Padul Probabilities - SMOTE Dataset

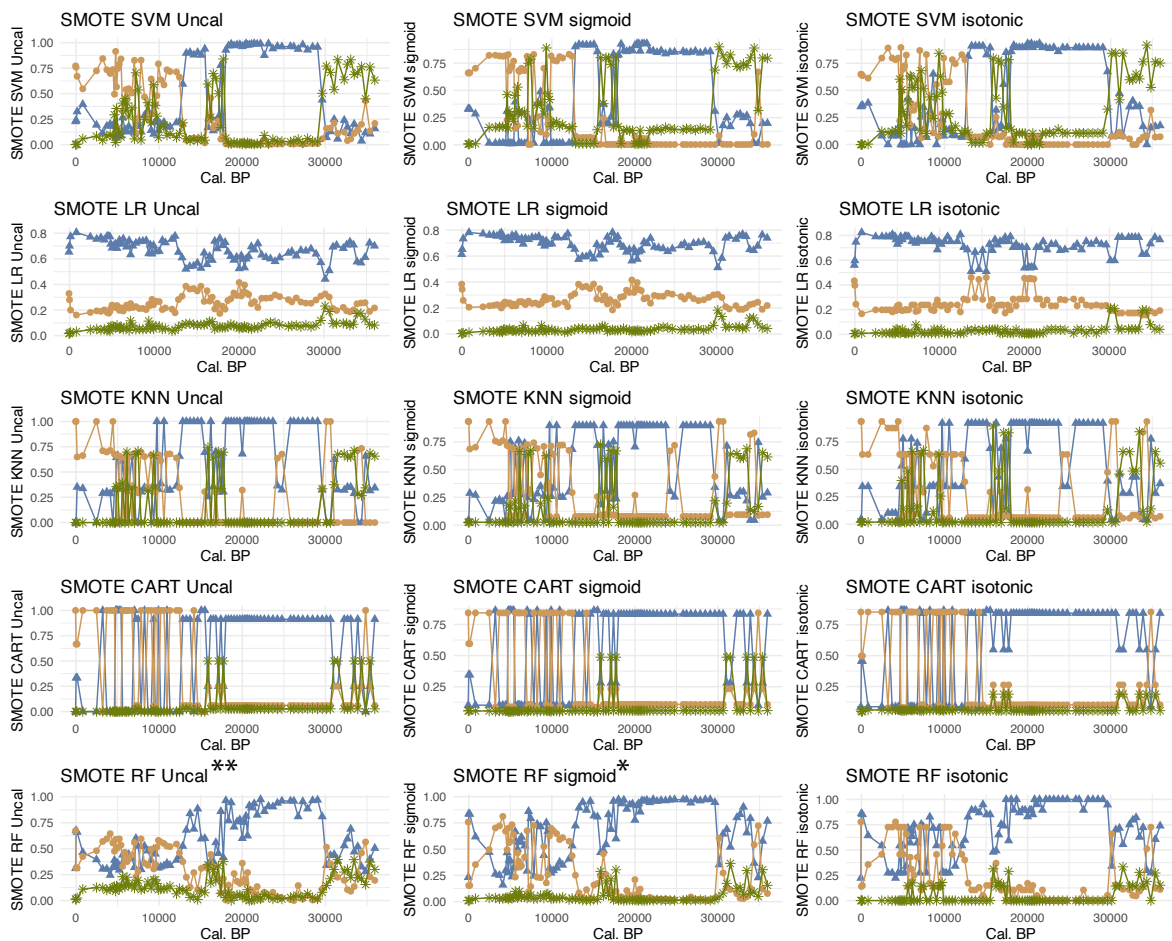


Figure S7. Probability estimates for the Padul record on models tested in the paper on the resampled datasets.

Vanevan Probabilities - SMOTE Dataset

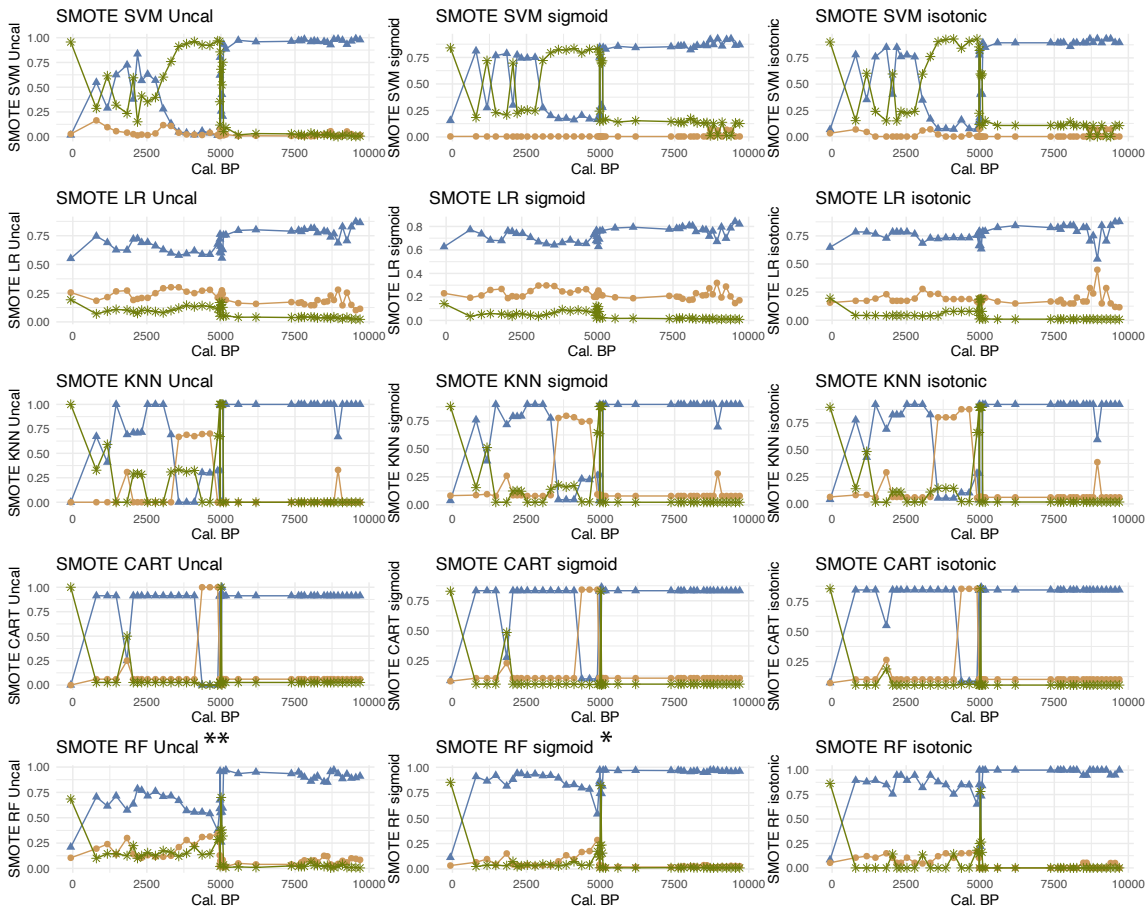


Figure S8. Probability estimates for the Vanevan record on models tested in the paper on the resampled datasets.

40

References:

45

Martínez-Sosa, Pablo, Jessica E. Tierney, Lina C. Pérez-Angel, Ioana C. Stefanescu, Jingjing Guo, Frédérique Kirkels, Julio Sepúlveda, Francien Peterse, Bryan N. Shuman, and Alberto V. Reyes. 2023. “Development and Application of the Branched and Isoprenoid GDGT Machine Learning Classification Algorithm (BIGMaC) for Paleoenvironmental Reconstruction.” *Paleoceanography and Paleoclimatology* 38 (7). <https://doi.org/10.1029/2023PA004611>.