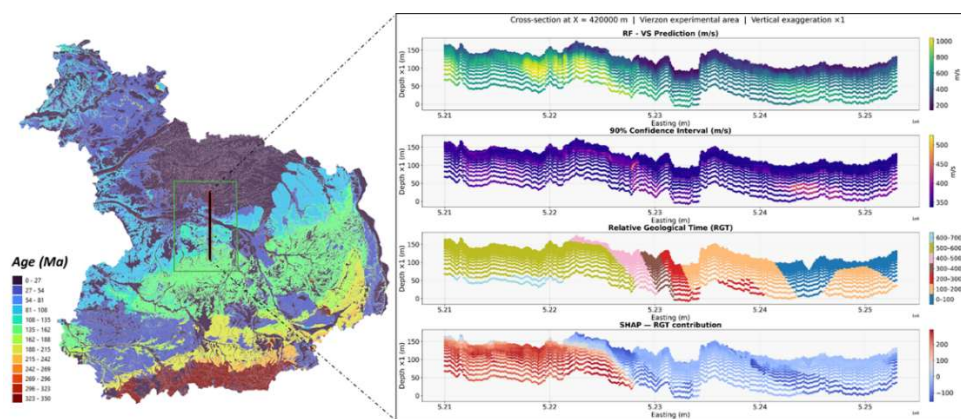


Uncertainty-Aware 3D Prediction of S-wave and P-wave Velocities from Multimodal Data

Roberto Cilli¹, Guillaume Caumon^{1,2}, Behshad Koohbor¹, Marianne Clausel³, Jihad Bourassi⁴, Pierre-Alexandre Reninger⁴, Aurélie Peyrefitte⁴

1 Université de Lorraine, CNRS, GeoRessources, Nancy. 2 Institut Universitaire de France. 3 Université de Lorraine, CNRS, CRAN, Nancy. 4 Bureau de Recherches Géologiques et Minières (BRGM), Orléans

Accurate characterization of the shallow subsurface is a prerequisite for natural hazard assessment, hydrogeological studies and land use planning. We present a multimodal machine learning framework for the 3D prediction of seismic S-wave and P-wave velocities in an experimental area of 500 km² in the southern part of Paris basin. The framework integrates four heterogeneous data sources: vertical resistivity profiles obtained from airborne electromagnetic data (SkyTEM [1]), interpolated using layer-by-layer 2D Ordinary Kriging; surface lithological descriptions from the BRGM converted into semantic embeddings via a pre-trained SentenceBERT model [2]; a 3D structural geological model built from nearly 10,000 stratigraphic contacts from ~600 borehole using LoopStructural [3], providing relative geological time (RGT); and Earth Observation embeddings (AlphaEarth) [4]. Sparse S-wave and P-wave velocity measurements from boreholes serve as ground truth. A key finding is that the inclusion of the structural geological model consistently improves predictive performance across all tested algorithms, with Random Forest [5] achieving the best coefficient of determination (R^2). Spatial generalization is evaluated using a Leave-One-Borehole-Out (LOBO) protocol with increasing spatial exclusion buffers (100m to 10km), designed to disentangle memorization from a fair spatial generalization. SHAP [6] analysis confirms that RGT is among the most influential features, providing interpretable evidence for the added value of geological knowledge integration into the pipeline. We finally discuss the relative contribution of each data modality, the limitations of uncertainty estimates using Gaussian processes [7] and conformal prediction [8] under covariate shift due to spatial correlation [9]. Representative cross-sections illustrate predicted velocities, associated uncertainties, and the influence of geological structure (RGT) on model behavior according to the treeSHAP algorithm [6]. Ongoing developments include the integration of surface geological contacts to improve the structural model, and the application of natural language processing techniques to extract lithological properties (grain size, chemistry) from detailed textual descriptions in explanatory notes of geological maps.



Left: Surface geological age map of the study area. **Right:** Cross-section at fixed longitude with, from top to bottom: RF-predicted S-wave velocities (m/s); 90% confidence interval width, evaluated with a spatial-aware cross-conformal calibration; Relative Geological Time (RGT) voxel model; SHAP values for the RGT feature, its contribution to the overall RF prediction.

- 1 Sørensen K.I., Auken E. A new system for airborne TEM surveying. *Exploration Geophysics*, 35(3), 191–199, 2004.
- 2 Reimers N., Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks (2019). arXiv:1908.10084.
- 3 Grose L., et al. LoopStructural 1.0: time-aware geological modelling. *Geosci. Model Dev.*, 14, 3915–3937, 2021. doi:10.5194/gmd-14-3915-2021
- 4 Brown, C. F., et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291.
- 5 Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). doi.org/10.1023/A:1010933404324
- 6 Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- 7 Williams, C., & Rasmussen, C. (1995). Gaussian processes for regression. *Advances in neural information processing systems*, 8.
- 8 Cordier, Thibault, et al. “Flexible and systematic uncertainty estimation with conformal prediction via the mapie library.” *Conformal and Probabilistic Prediction with Applications*. PMLR, 2023.
- 9 Tibshirani, R.J., Barber, R.F., Candès, E.J., & Ramdas, A. (2019). Conformal Prediction under Covariate Shift. *Neural Information Processing Systems*.