Predicting Rock Properties through ML Models

Buchanan Kerswell[1,2]*, Nestor G. Cerpa[2], Andréa Tommasi[2], Marguerite Godard[2], José Alberto Padrón-Navarta[3]

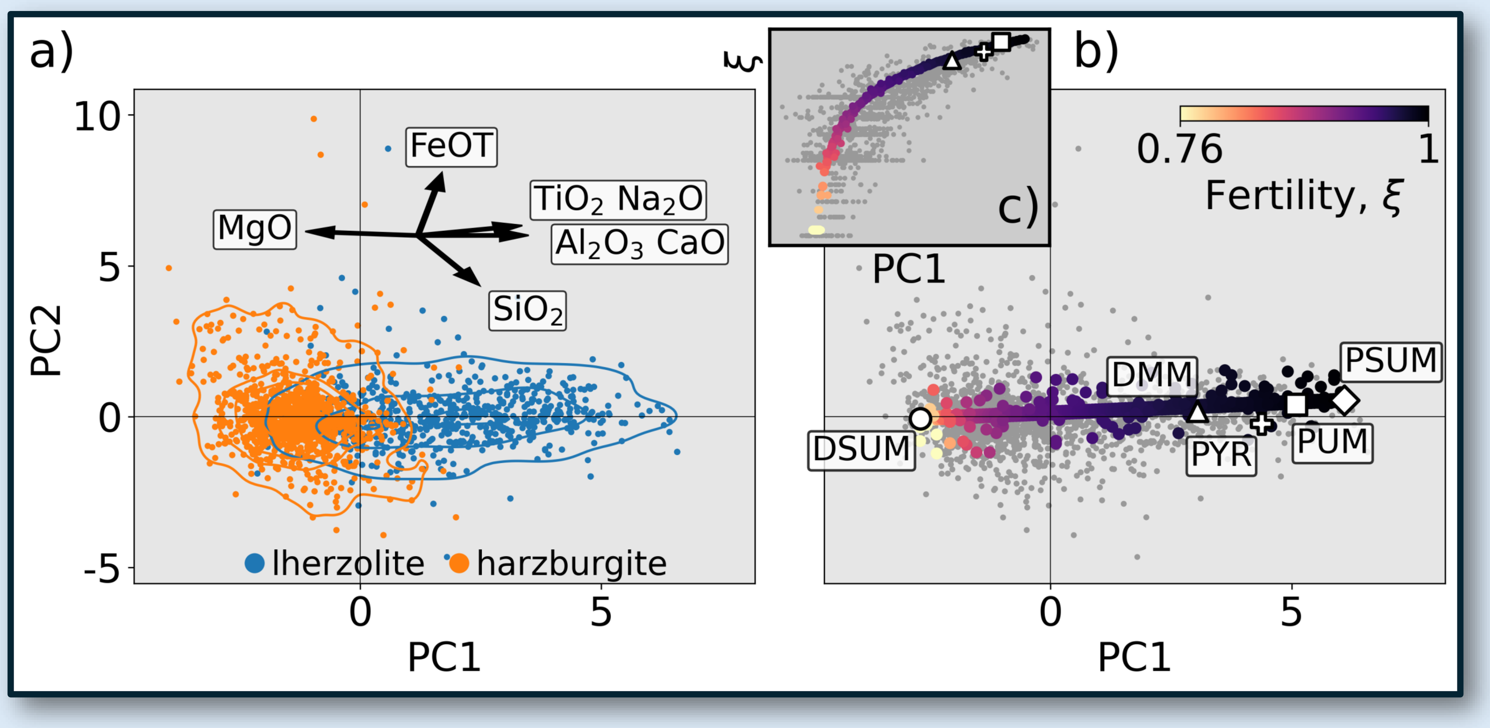
[1]University of Liverpool, UK; [2]Geosciences Montpellier, France; [3]Instituto Andaluz de Ciencias de la Tierra (IACT), Spain

*b.kerswell@liverpool.ac.uk

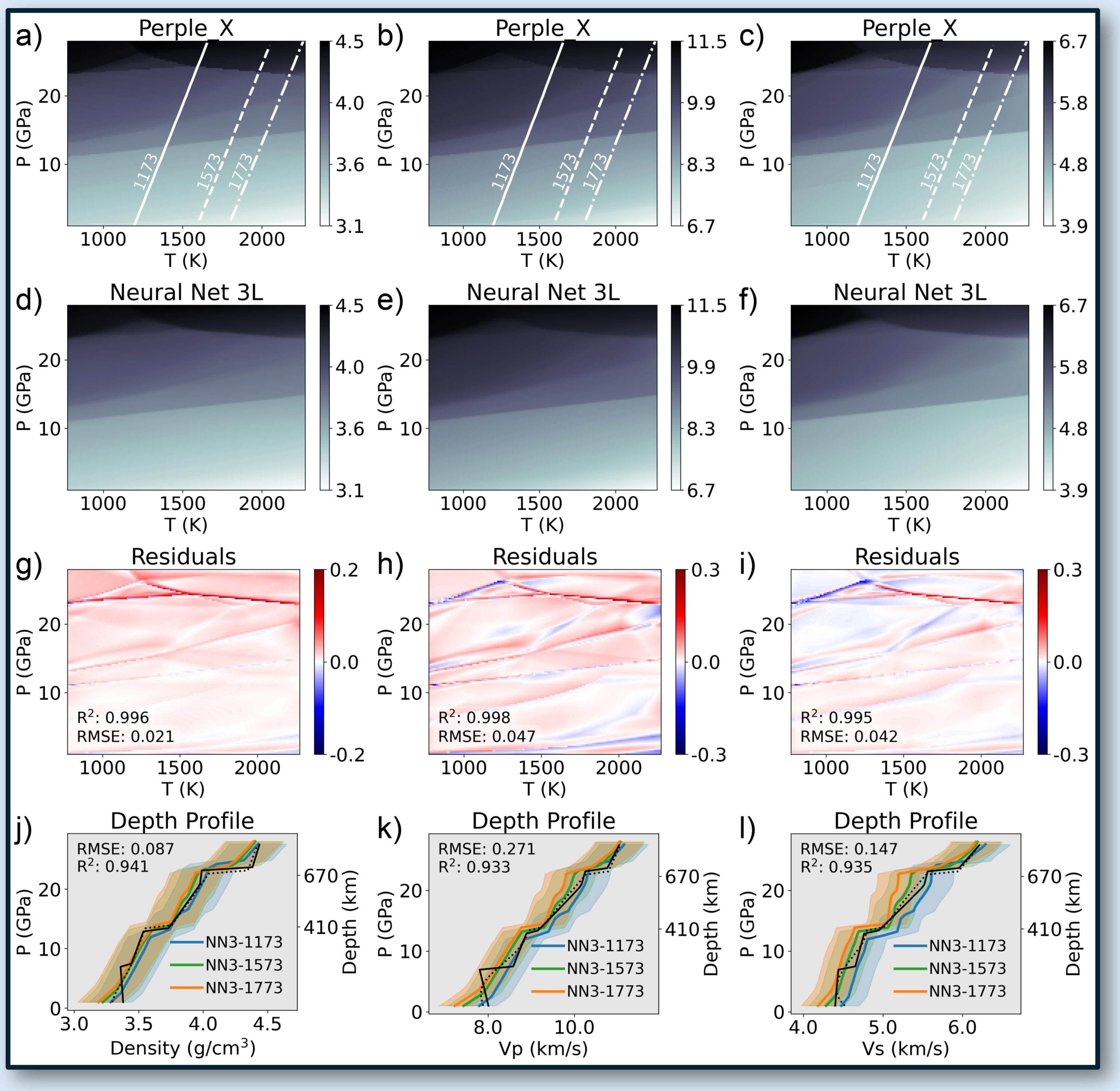
⇐ Problem ⇒

Lookup Tables/Perple_X are too slow for practical use in hi-resolution geodynamic simulations

⇐ Training Data ⇒



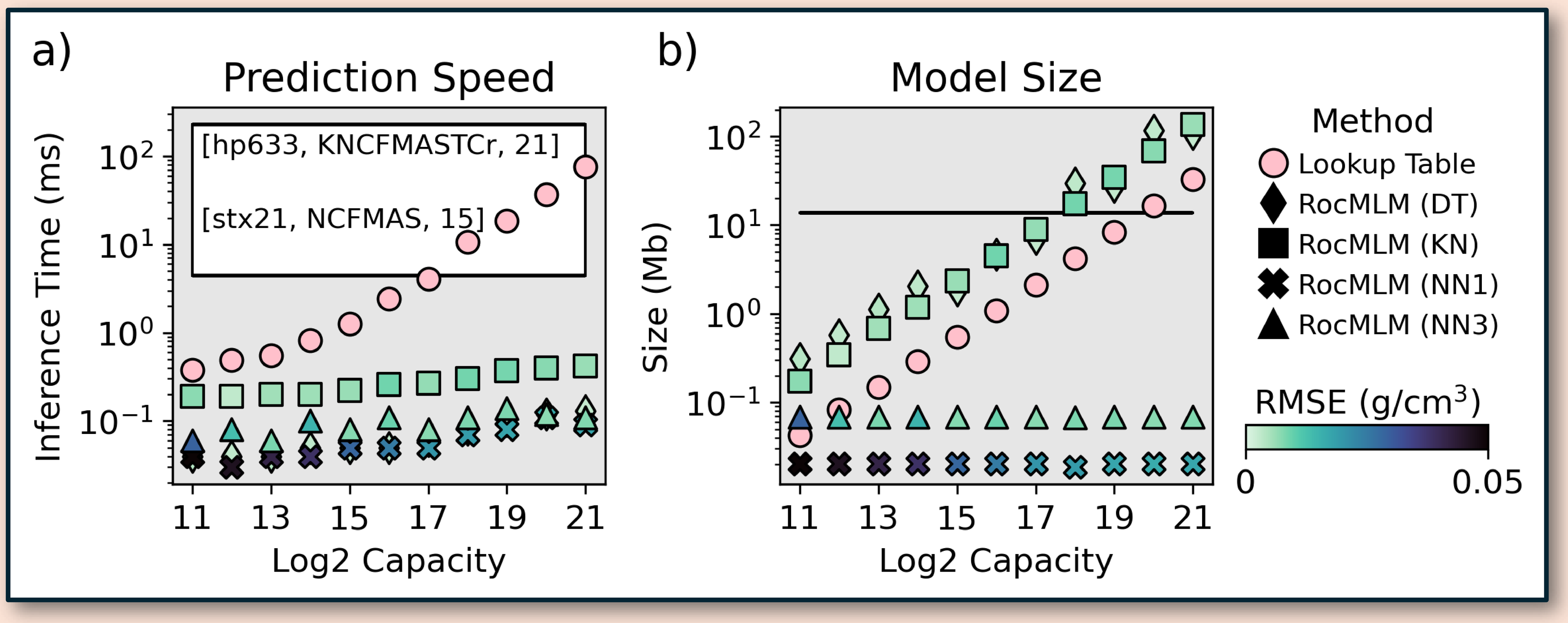
PC1-PC2 diagrams showing the standardized geochemical dataset of natural peridotite samples (a) and a mixing array between hypothetical end-member mantle compositions Primitive Synthetic Upper Mantle (PSUM) and Depleted Synthetic Upper Mantle (DSUM, b). Black arrows in (a) indicate PCA loading vectors. Colored data points in (b) are the synthetic mantle compositions used to train RocMLMs, which were sampled independently from the natural peridotite samples (gray data points). The inset (c) shows how the Fertility Index (ξ) changes nonlinearly with PC1. DMM, PUM, and PYR are Workman and Hart (2005), Sun and McDonough (1989), and Green (1979), respectively.



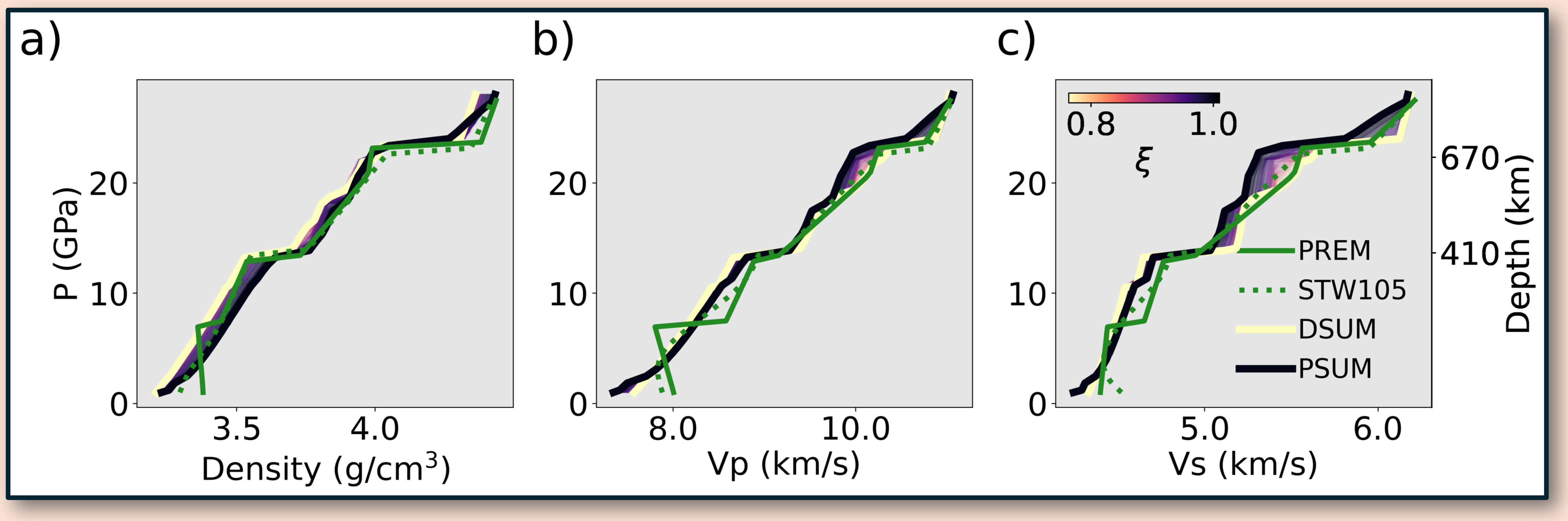
Computational efficiency of various approaches in terms of prediction speed (a) and model size (b). "Capacity" (x-axis) is a proxy for the petrological "knowledge", or predictive capabilities, of Lookup Tables and RocMLMs. The white region in (a) indicates GFEM prediction speed for different Perple_X configurations (thermodynamic dataset, chemical system, and number of solution phases are indicated in square brackets). GFEM model size is constant (bold black line). stx21: Stixrude and Lithgow-Bertelloni (2022), hp633: Holland and Powell (2011) updated in Holland et al. (2018). Perple_X was run without multilevel grid refinement. RMSE is measured using kfold cross-validation with k=5.

⇐ ML Model Predictions ⇒

ML models predictions are accurate and up to 10^1–10^3 faster than Lookup Tables or Perple_X



Computational efficiency of various approaches in terms of prediction speed (a) and model size (b). "Capacity" (x-axis) is a proxy for the petrological "knowledge", or predictive capabilities, of Lookup Tables and RocMLMs. The white region in (a) indicates GFEM prediction speed for different Perple_X configurations (thermodynamic dataset, chemical system, and number of solution phases are indicated in square brackets). GFEM model size is constant (bold black line). stx21: Stixrude and Lithgow-Bertelloni (2022), hp633: Holland and Powell (2011) updated in Holland et al. (2018). Perple_X was run without multilevel grid refinement. RMSE is measured using kfold cross-validation with k=5.



Depth profiles of RocMLM training data along a 1573 K mantle adiabat showing the sensitivities of thermodynamic estimates of density (a), Vp (b), and Vs (c) to changes in bulk mantle composition (as represented by the Fertility Index, ξ). Geophysical profiles PREM and STW105 (green lines) and the profiles of synthetic mantle end-member compositions PSUM and DSUM (thick colored lines) are shown for reference. Thin colored lines show profiles for the entire range of RocMLM training data.

⇐ Take-Away ⇒

Provided a sufficiently large and diverse training dataset, ML models can effectively implement complex physics into geodynamic simulations with minimal additional computational costs (at inference)

Open Research

All data, code, and relevant information for reproducing this work are published at github.com/buchanankerswell/kerswell_et_al_rocmlm and archived on the Open Science Framework data repository doi.org/10.17605/OSF.IO/K23TB. All code is MIT Licensed and free for use and distribution (see license details). Reference models PREM and STW105 are freely available from the Incorporated Research Institutions for Seismology Earth Model Collaboration (Trabant et al., 2012). All computations were made using CPUs of a Macbook Pro (2022; M2 chip) with macOS 14.5 and using Python 3.12.3.

Acknowledgements

This work was supported by the Tremplin-ERC grant LEARNING awarded to Nestor Cerpa by the I-SITE excellence program at the Université de Montpellier. We thank Dr. Zhou Zhang, one anonymous reviewer, Maurine Montagnat, Fernando Carazo, Nicolas Berlie, and others at Géosciences Montpellier for their thoughtful feedback. We gratefully acknowledge additional support from the European Research Council (ERC) under the European Union Horizon 2020 Research and Innovation program grant agreement No. 882450 (ERC RhEoVOLUTION) awarded to Andréa Tommasi.