Using information entropy to optimize and communicate certainty of continental scale tectonic models

Tobias Stål[1,2,3], Anya M. Reading[1,2,3], Matthew J. Cracknell[4], Jörg Ebbing[5], Jacqueline A. Halpin[2,3], Ian D. Kelly[1,3], Emma J. MacKie[6], Mohamed Sobh[7], Ross J. Turner[1], and Joanne M. Whittaker[2]

1, Physics, School of Natural Sciences, University of Tasmania, Australia; 2, Institute for Marine and Antarctic Studies, University of Tasmania, Australia; 3, The Australian Centre for Excellence in Antarctic Science; 4, CODES, School of Natural Sciences, University of Tasmania, Australia; 5, Institute of Geosciences, Kiel University, Germany; 6, Geological Sciences, University of Florida, USA; 7, Leibniz Institute for Applied Geophysics, Germany

Motivation

We use fundamental concepts of information theory to develop empirical tectonic models of the Antarctic continent. The Antarctic subglacial geology impact geothermal heat, subglacial sedimentation, and glacial isostatic adjustment; critical parameters for predicting the ice sheet's response to warming oceans. However, the tectonic architecture of the Antarctic interior is unresolved, with results dependent on datasets or extrapolation used. Most existing deterministic suggestions are derived from qualitative observations and often presented as robust results, hiding possible alternative interpretations.

Similarity Detection

We construct an empirical tectonic model by linking observed geology in other continents to Antarctica by multiple observables (22, in this presentation), including seismic tomography, potential field data such as gravimetry and magnetometry, and geological observations of volcanoes and rock samples. We also include derived properties such as curvature and relations between observables (Fig. 2). The observables have good cover and validity over the Antarctic continent. Each used observable reduces information entropy for the resulting models (Stål et al., 2021).
We detect the degree of similarity between each **target** location in Antarctica to each **reference** location in the rest of the world for each observable (Fig. 2):

$$S = \exp\left(-\frac{(o_R - o_T)^2}{2 \times (\sigma_R + \sigma_T)^2}\right), \tag{1}$$

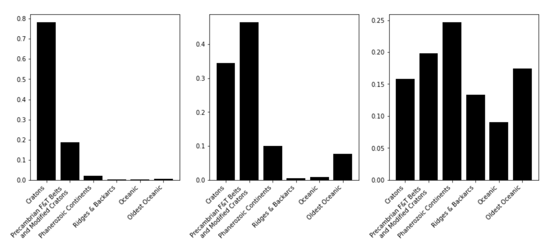
o_R is the reference value and o_T is the target value. The accepted range, σ, is derived from provided uncertainty range for the observable or optimised using a Monte Carlo approach (Stål et al., 2021).
The sum of similarities for all observables are stacked:

$$N_{sim} = \sum_{i=1}^{n_{obs}} S_{obs} \times w_{obs}, \tag{2}$$

where each observable can be weighted, w_{obs}, and used in an exponential similarity function:

$$w_i = K^{N_{sim}} \tag{3}$$

The w_i weighting is used to sample tectonic segmentation models as the sum of weight for each reference class included in the model. The K–parameter controls the focus of the similarity analysis and is analysed further using information entropy. A low K (e.g. below 5) produces a generally robust but less precise discrete distribution; a value of K over 10 provides low information entropy for the similarity of the distribution sampled for each target, but the results are less robust and hide possible alternative interpretations.



Examples of discrete distributions of tectonic class (Schaeffer and Lebedev, 2015) detected in this study. The left example is the best-case scenario, the centre is a typical distribution, and the right example is the worse case (Fig. 6).

In order to optimise K without reducing the robustness, we observe the gradient of the total entropy of the discrete similarity distributions with K. We treat the K value as a cost function and identify the lowest value of K that meets the condition of a flatten out gradient, defined as $K/H > 0.075$.
The information gain with increasing value K varies; locally computed thresholds differ from the total entropy. We map the Kullback–Leibler divergence, comparing the local optimised K with the global value of K, using the distributions:

$$D_{\mathrm{KL}}(P_{K_{\mathrm{total}}} \parallel P_{K_{\mathrm{local}}}) = \sum_{x \in \mathcal{X}} P_{K_{\mathrm{total}}}(x) \log\left(\frac{P_{K_{\mathrm{total}}}(x)}{P_{K_{\mathrm{local}}}(x)}\right). \tag{4}$$

The divergence map shows how robust the data is for each given location in Antarctica. We suggest an interpretation where the Kullback–Leibler divergence is a measure of representation, that the target location can be assigned a class from a low number of reference locations where the observables are sufficient for classification.

Communicating Uncertainty

The similarity method provides a transparent classification algorithm that allows us to investigate how the results depend on the data and parameter choices. This is of particular importance for applications where the datasets are associated with large uncertainties or poor spatial consistency, e.g. due to the logistical challenges in remote regions.
The use of information entropy provides us with an unbiased and transparent metric to communicate the ambiguities from the uncertainties of qualitative classifications.
Given the complex geology and the limited reference dataset, Earth only, it is not feasible to generate an unambiguous final segmentation model; however, in some regions, the model is robust. We map robustness by determining the lowest value of K where the local entropy decreases at a low rate (Fig. 3).
One of the main challenges is the semantics of communicating the results; existing models are often deterministic. Maps showing the parameters of the sum of similarities, information entropy for class distribution, locally optimised K, and Kullback–Leibler divergence between global and locally optimised K are provided to guide the use of the model.

Key References

Cracknell, M. J., and Reading, A. M. (2015). Spatial-Contextual Supervised Classifiers Explored: A Challenging Example of Lithostratigraphy Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(3), 1371–1384.
Schaeffer, A. J., and Lebedev, S. (2015). Global heterogeneity of the lithosphere and underlying mantle: A seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization. In The Earth's Heterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective (pp. 3–46). Springer International Publishing.
Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(3), 379–423.
Stål, T., Reading, A. M., Halpin, J. A., and Whittaker, J. M. (2020). Antarctic geothermal heat flow model: Aq1. Geochemistry, Geophysics, Geosystems, 22(2), 1–22.
Stål, T., Reading, A. M., Halpin, J. A., Phipps, S. J., Whittaker, J. M., Steven J, P., Whittaker, J. M., Phipps, S. J., and Whittaker, J. M. (2020). The Antarctic Crust and Upper Mantle: A Flexible 3D Model and Software Framework for Interdisciplinary Research. Frontiers in Earth Science, 8(November), 1–19.

Contact

tobias.staal@utas.edu.au
@TobiasStl @comp_antarctic
TobbeTripitaka

The use of **information theory** to develop and evaluate large scale tectonic models

ACEAS
Australian Centre for Excellence in Antarctic Science

UNIVERSITY of TASMANIA

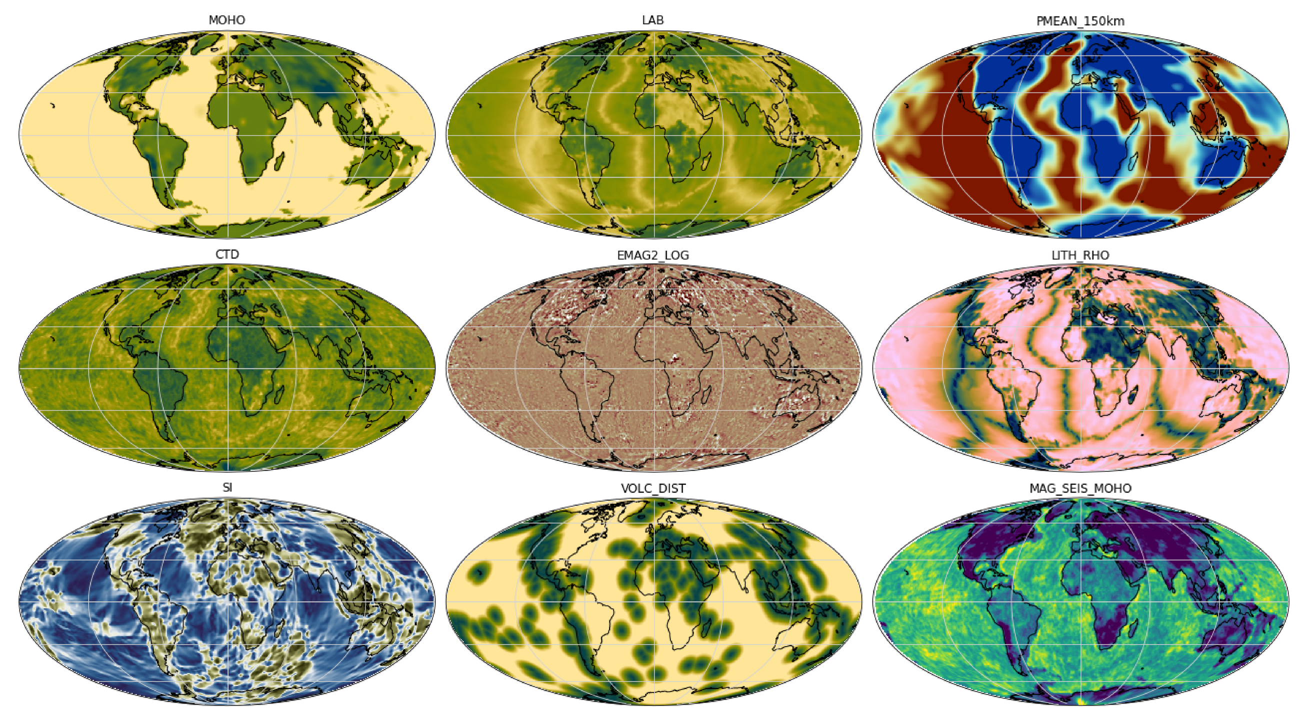


Fig. 1. Examples of observables used. In this presentation, we use 22 observables, including Moho depth, free air gravity, magnetic susceptibility, volcanoes, and hypsometry. Some observables are included as provided from the original studies, other observables are computed for this study. Reference tectonic classifications are subtracted from Schaeffer, A. J., and Lebedev, S. (2015). The Antarctic continent and Southern Ocean south of 60°S are excluded from the reference observables.

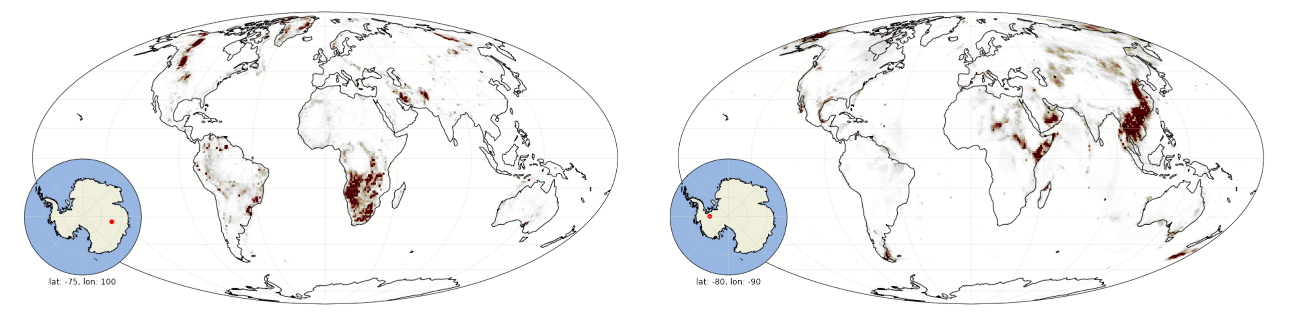


Fig. 2 Examples of the match between target and references. Using those observables, we can detect the degree of similarity between each location in Antarctica and the rest of the world. K=10

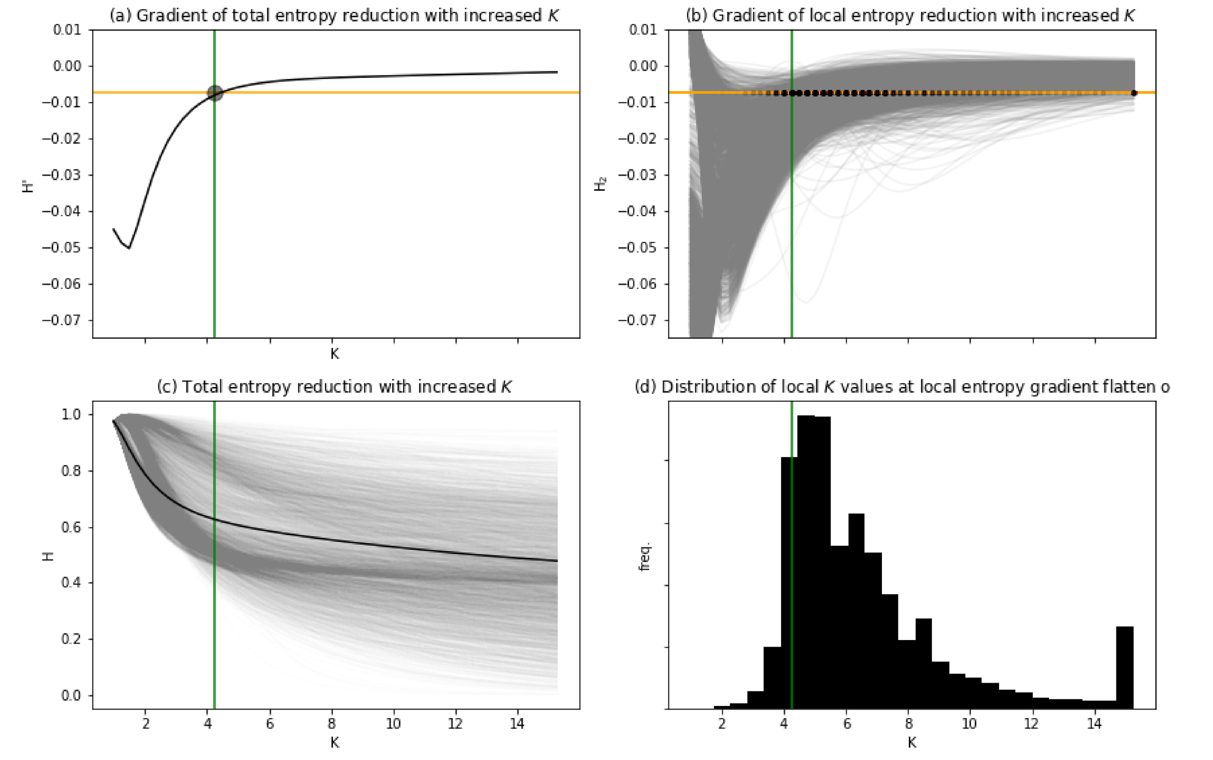


Fig. 3. The entropy decreases with increasing K, this follows from the weighting $w_i = K^{N_{sim}}$; however, the rate depends on the observables and location. We use a H/K gradient of -0.0075 to define stabilisation and compute the optimal K value for the entire model (black line), as well as for each grid cell (grey lines). (a) the gradient of the H/K relation; here showing the total entropy and the point where the entropy change flattens out. (b), the gradient of the H/K for the entire model and each local cell. (c), the relation between K values and normalised entropy. (d), distribution of local K-values, optimised from the information entropy.

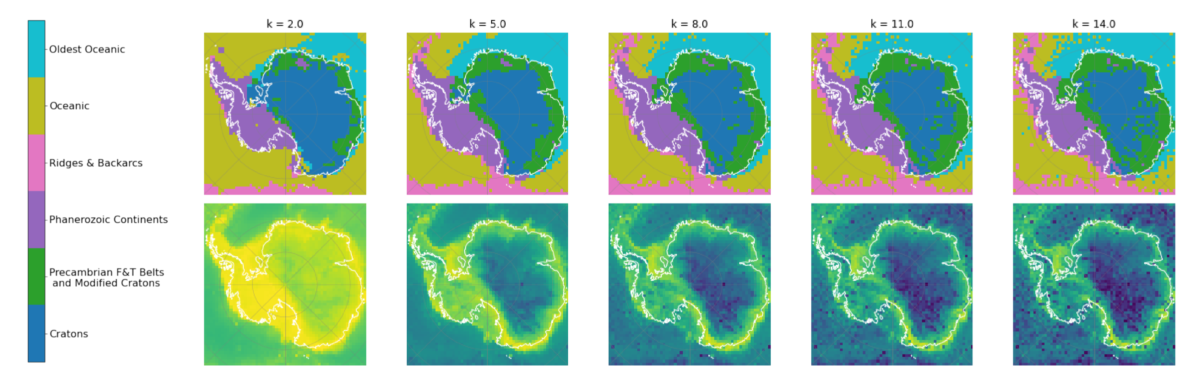


Fig. 4. The K parameter also impacts the most likely class for each cell (upper row), and for high values of K the model contains artefacts from overfitting. The optimal local K (lower row) is highest around the continental shelf and in West Antarctica. Classes adapted from Schaeffer, A. J., and Lebedev, S. (2015) indicated on the legend. The normalised entropy is shown as high entropy in a light colour.

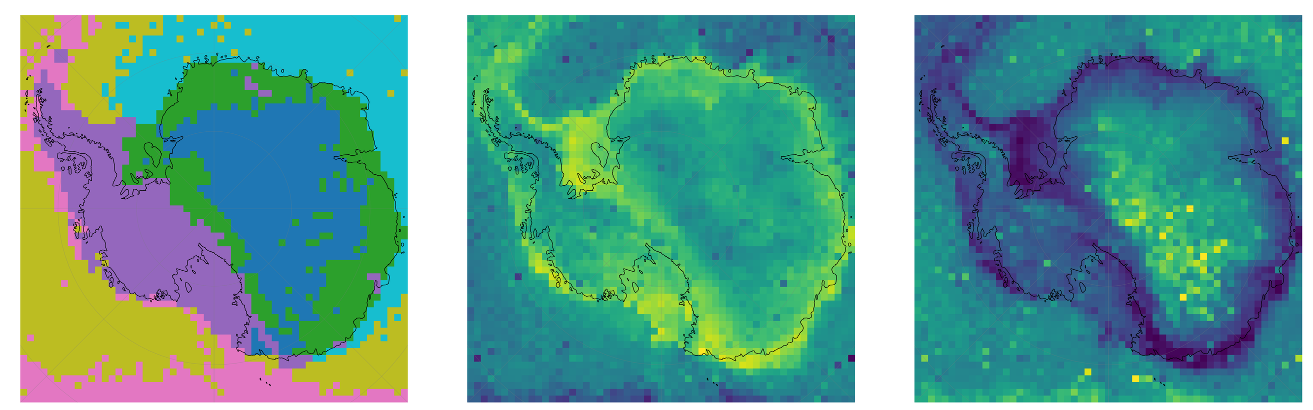


Fig. 5. Preliminary results. (a), The most likely class for each location in Antarctica. (b), The normalised information entropy using K=10. Lighter colour indicates higher information entropy. (c), the divergence between the normalised information entropy using K=10 and the local optimum K value.