Los Alamos
NATIONAL LABORATORY

The Transparent Earth: A Multimodal Foundation Model for the Earth's Subsurface

Daniel O’Malley
Earth & Environmental Sciences Division

March 23, 2026

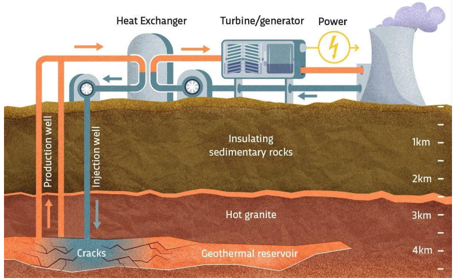
The Earth’s subsurface is more important than you think

Groundwater

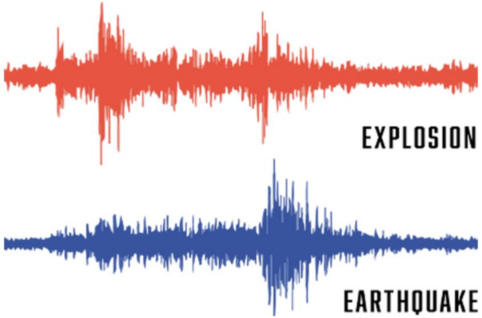
Fossil Fuels

Earthquakes

Critical Minerals

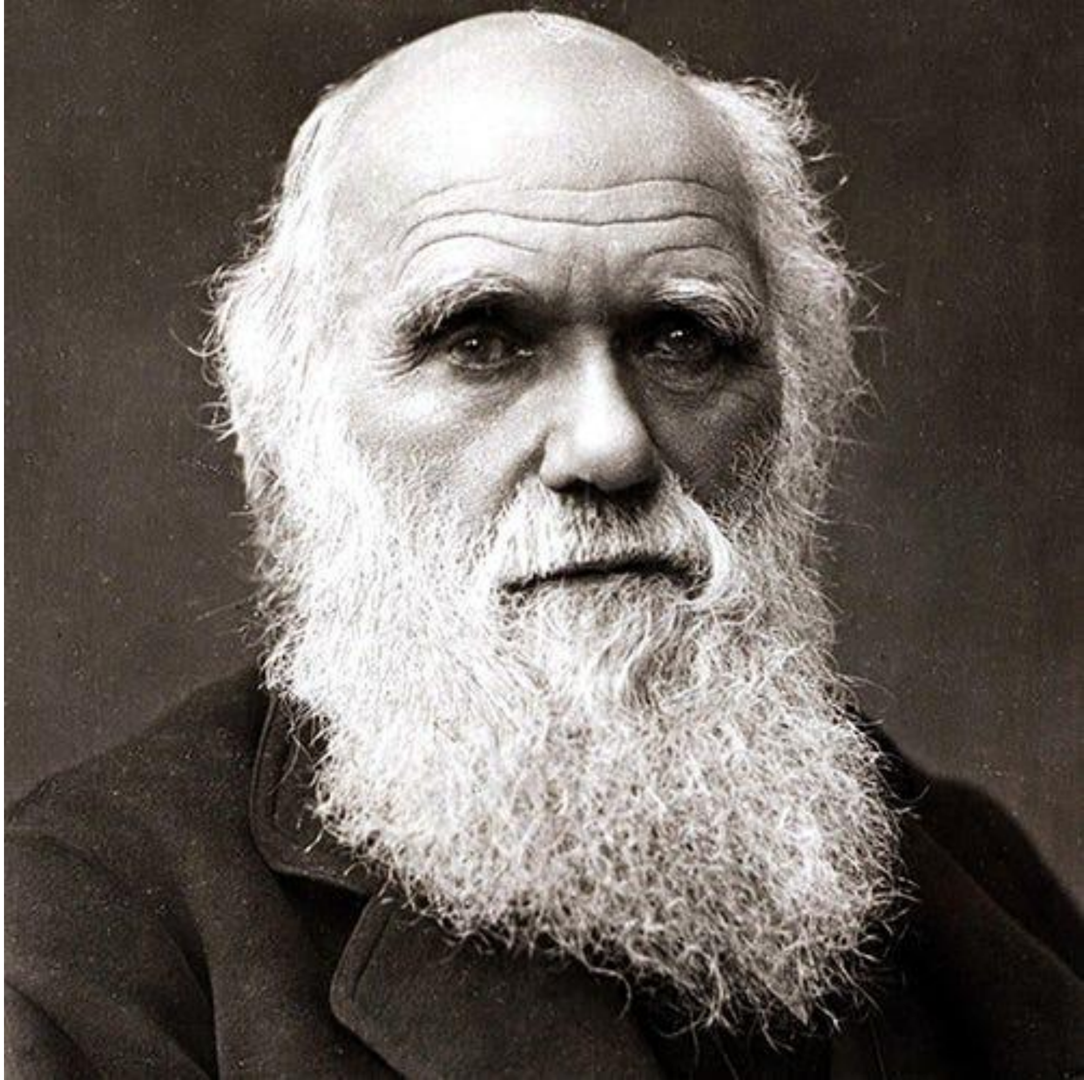


Geothermal Energy



Nuclear Nonproliferation

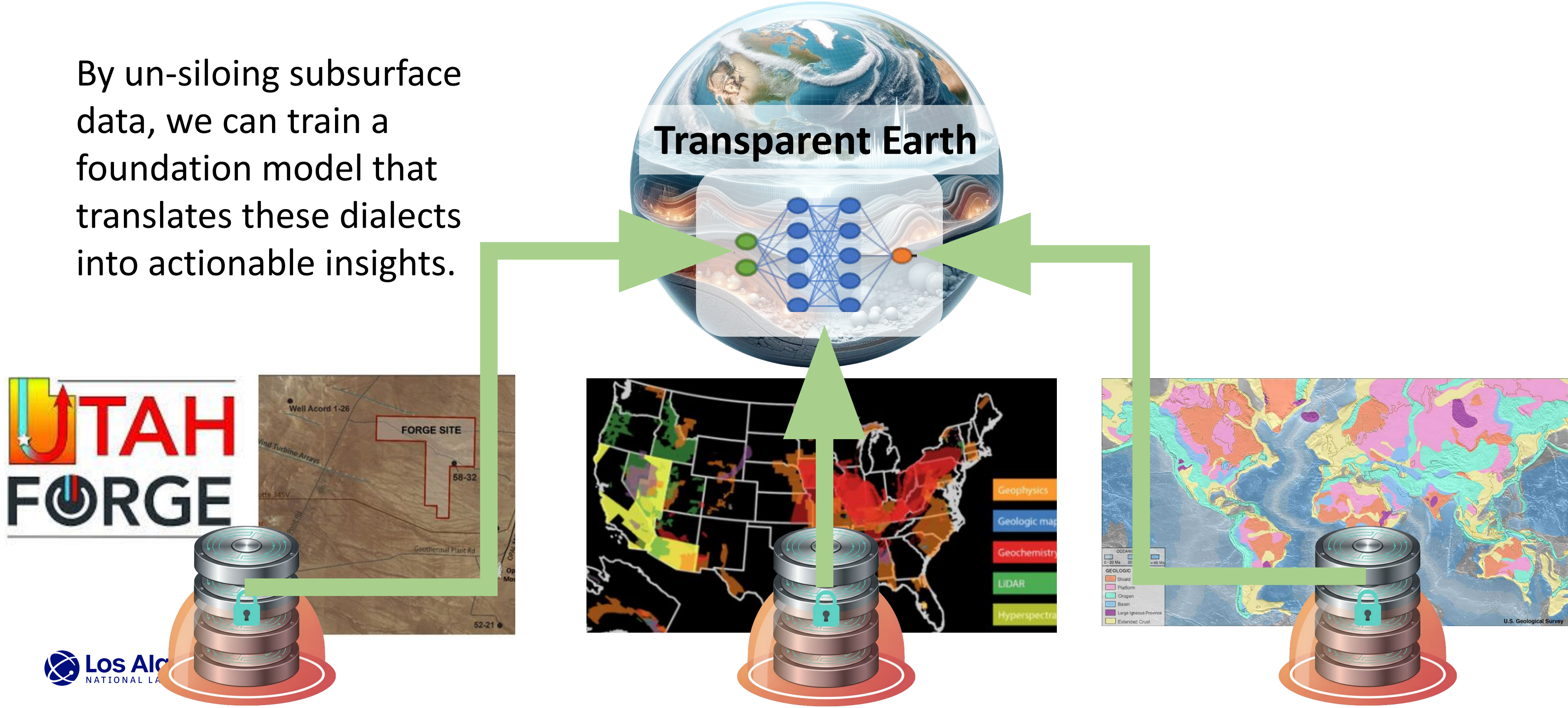
“I look at the geological record as a history of the world, imperfectly kept and written in a changing dialect”
– Charles Darwin



Transparent Earth: foundation model for the subsurface

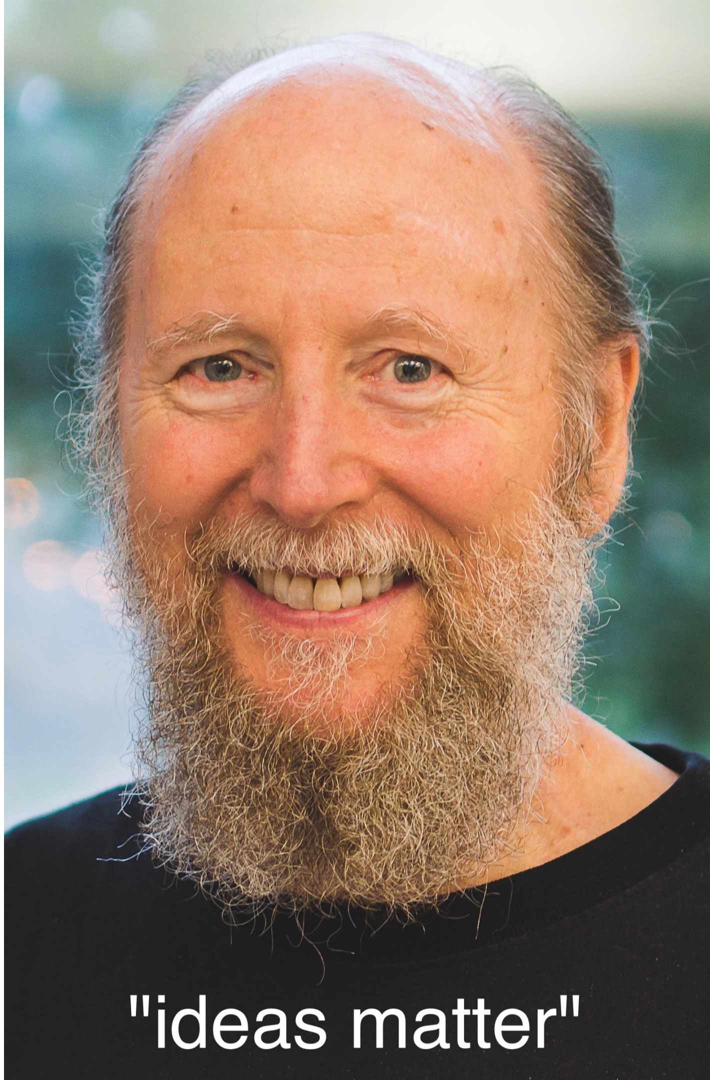
Predict any property of the Earth's subsurface anywhere

By un-siloing subsurface data, we can train a foundation model that translates these dialects into actionable insights.

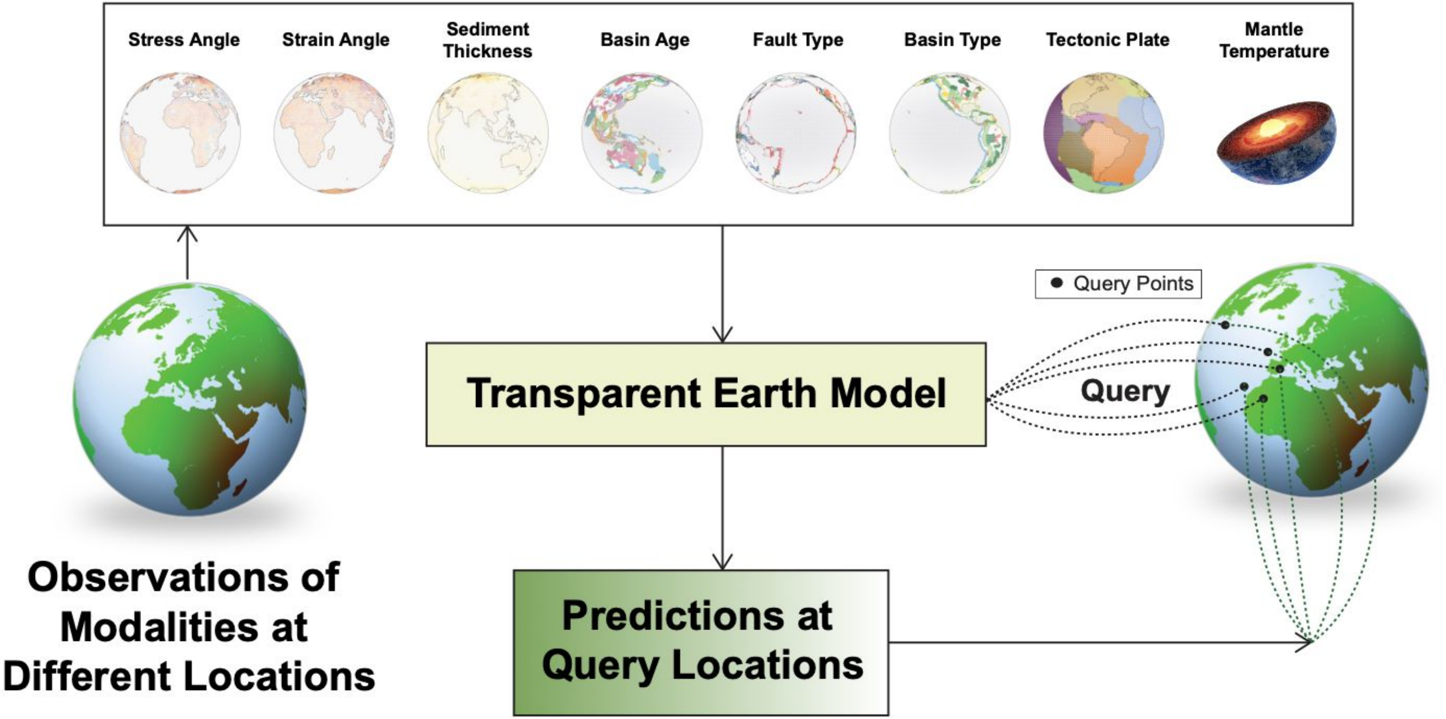


Richard Sutton’s bitter lesson

- “The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin.”
- “We have to learn the bitter lesson that building in how we think does not work in the long run. The bitter lesson is based on the historical observations that
  1. AI researchers have often tried to build knowledge into their agents
  2. this always helps in the short term, and is personally satisfying to the researcher, but
  3. in the long run it plateaus and even inhibits further progress, and
  4. breakthrough progress eventually arrives by an opposing approach based on scaling computation by search and learning.”


"ideas matter"

Transparent Earth uses in-context learning



Transparent Earth Model

Observations of Modalities at Different Locations

Predictions at Query Locations

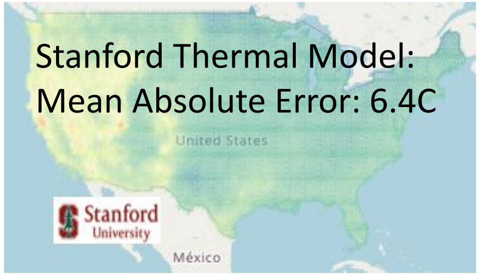
A. Mazumder, J.E. Santos, N. Hobbs, M. Mehana, and D. O'Malley. The Transparent Earth: A Multimodal Foundation Model for the Earth's Subsurface. *NeurIPS Workshop on Artificial Intelligence for Science*. 2025

Transparent Earth learned to accurately predict 8 disparate properties of the Earth’s subsurface

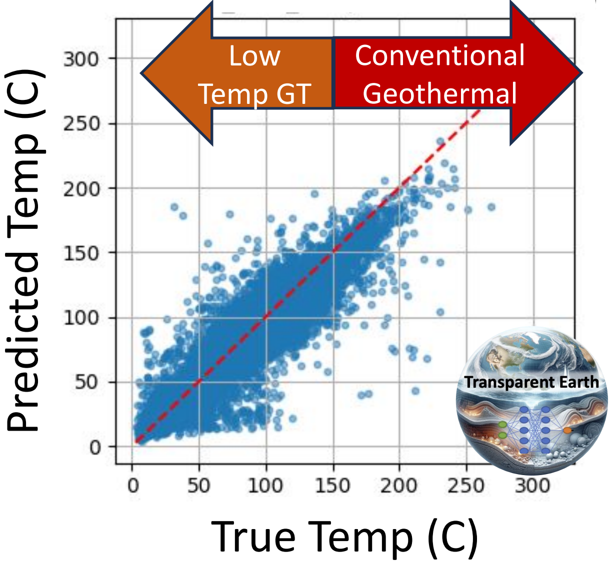
A. Mazumder, J.E. Santos, N. Hobbs, M. Mehana, and D. O'Malley. The Transparent Earth: A Multimodal Foundation Model for the Earth's Subsurface. *NeurIPS Workshop on Artificial Intelligence for Science*. 2025

New state-of-the-art for predicting geothermal temperature

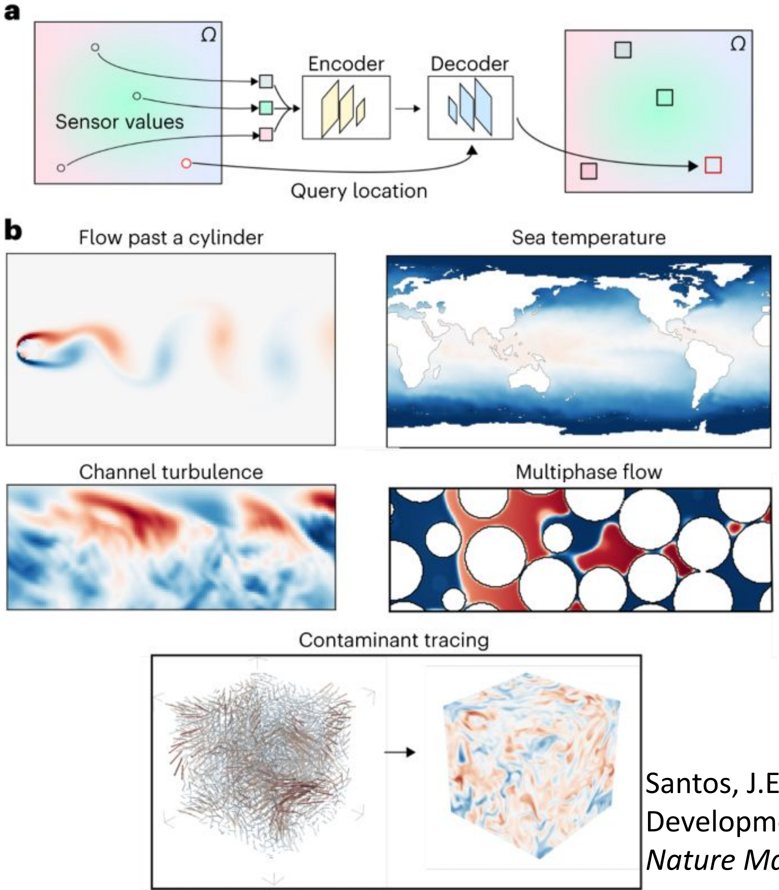
- There are not a lot of models that we can compare Transparent Earth to, but a group out of Stanford has a state-of-the-art model for predicting subsurface temperature
- This utilizes a dataset of 400k temperature observations from wells drilled as deep as 7km into the the Continental US
- Stanford's model is limited to the continental US and gets an error of 6.4C. Our model gets 5.8C, setting a new state of the art
- Our model also does reasonably well in the UK (9C) and Alberta, Canada (15C) despite having no training data for temperature there
    - But, it performs poorly in Australia (30C) – more on that later


Stanford Thermal Model:
Mean Absolute Error: 6.4C

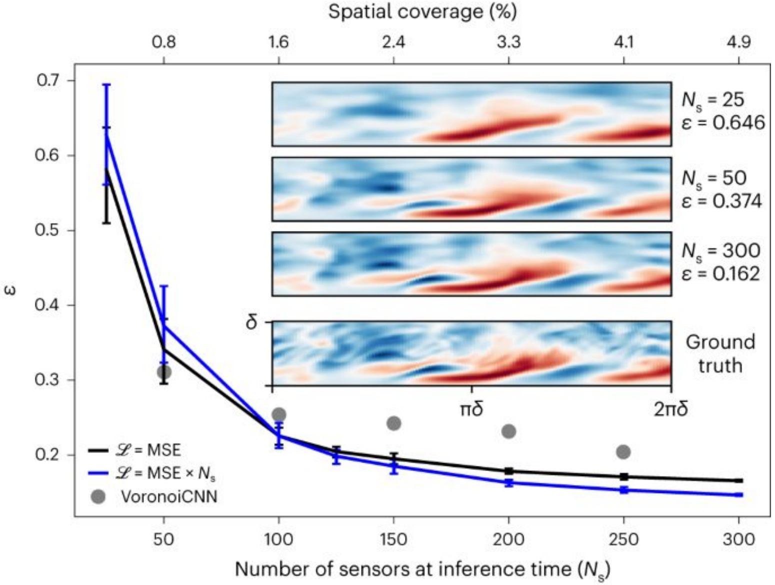
Ours: Mean Absolute Error: 5.8C



Senseiver: the core of the Transparent Earth



Senseiver achieved state-of-the-art performance on mapping sparse observations to global fields

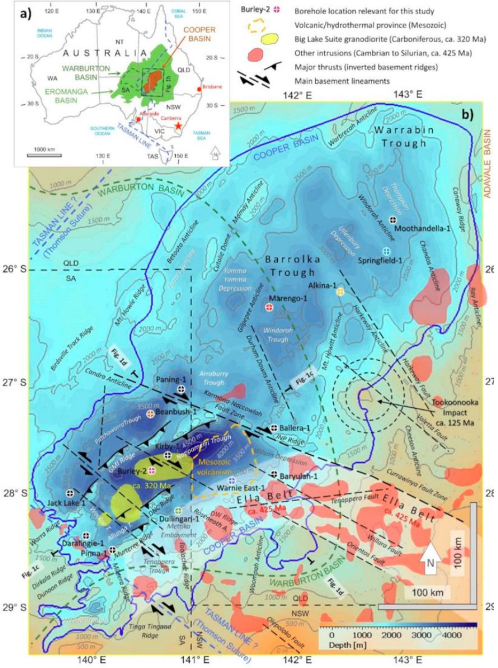


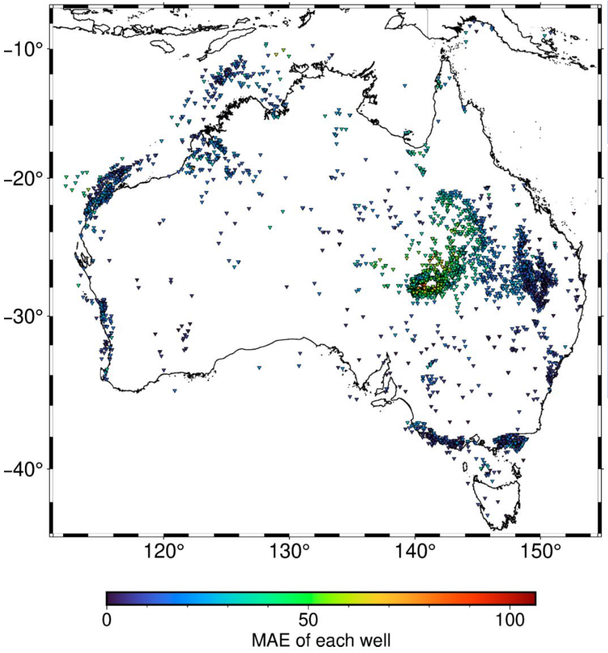
Santos, J.E., Fox, Z.R., Mohan, A., O'Malley, D., Viswanathan, H. and Lubbers, N., 2023. Development of the Senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, *5*(11), pp.1317-1325.

Towards Transparent Earth 2.0

- Senseiver has many desirable properties (e.g., in-context learning, powerful transformer architecture) that form a good foundation, but…
- Senseiver was designed for dynamic problems
  - The task is to use sparse observations to identify a snapshot in time. We are initially focused on relatively static properties of the subsurface
- Senseiver was designed to always predict a global field with observations dispersed in a relatively uniform fashion
  - Earth's subsurface is multiscale and often you want to stay zoomed-in on a local region
- We have begun to explore approaches that can overcome these limitations
- We initially focus on geothermal temperature modeling since there is a good baseline for this problem (Stanford Thermal Model)

Back to Australia





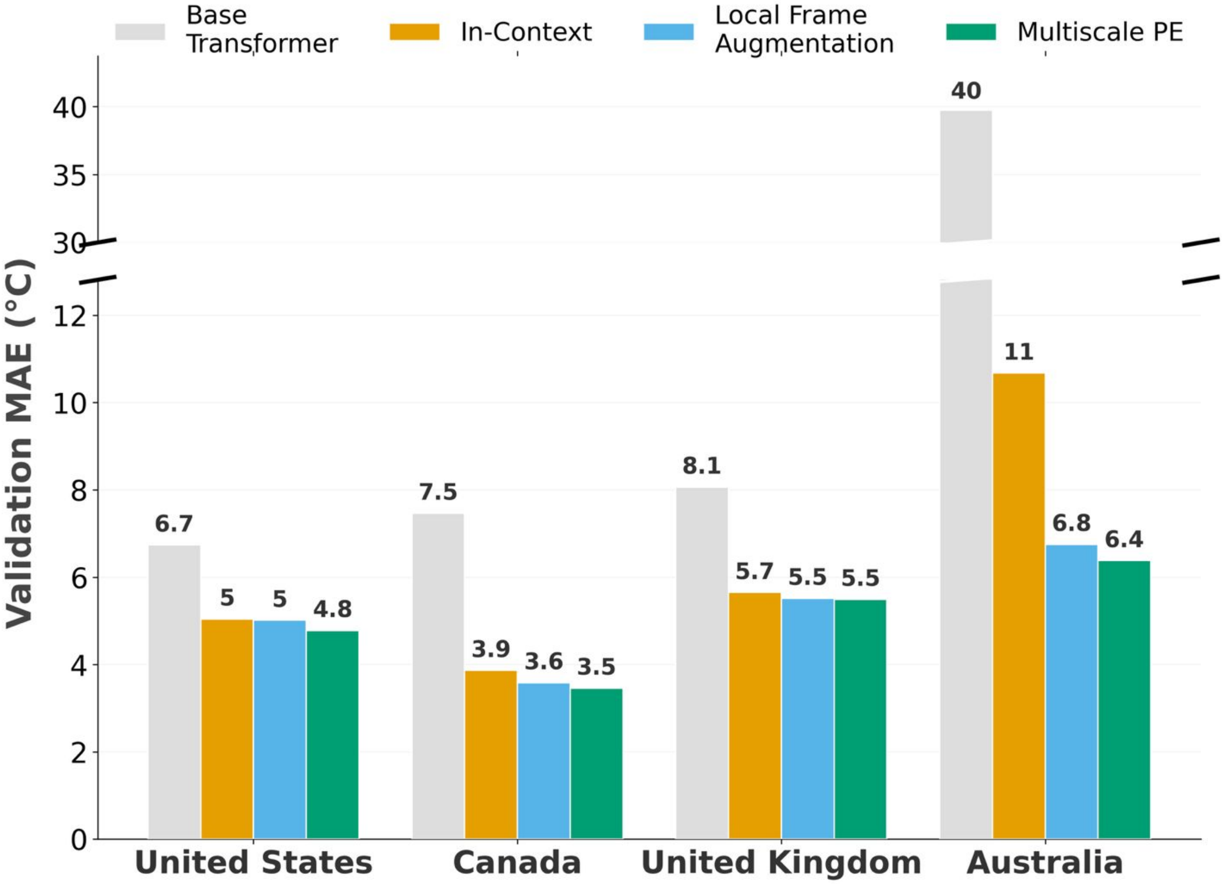
Model	USA	Alberta	UK	Australia
Stanford Thermal Model	6.4C	N/A	N/A	N/A
Transparent Earth 1	5.8C	15.0C	9.0C	30.1C
Transparent Earth 2 preview	**4.8C**	**3.5C**	**5.5C**	**6.4C**

Transparent Earth 2 Preview significantly reduces errors everywhere, especially in unseen, complex regions like Australia

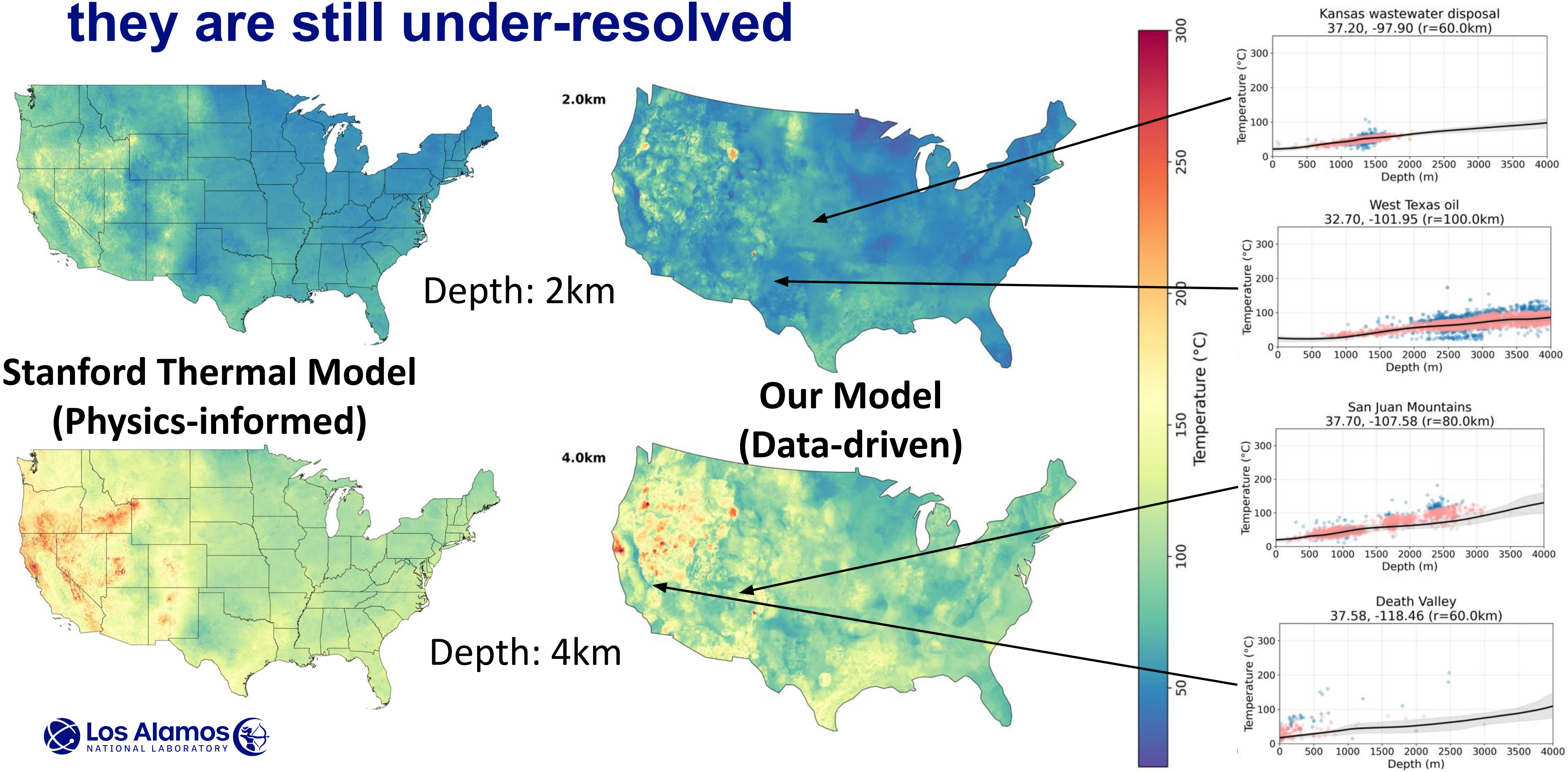
Right now, Transparent Earth 2 Preview uses only temperature data, so there is room for improvement by using multimodal relationships

In-context, data augmentation, and multiscale positional encodings are key

- In-context learning: TE1 has this, but now we provide local information in-context
- Local frame augmentation: the model doesn't know where on Earth the prediction is being made, just what the surrounding environment looks like
- Multiscale Positional Encodings: allows the model to understand small and large distances



We are resolving sharper temperature gradients, but they are still under-resolved



Depth: 2km

Stanford Thermal Model (Physics-informed)

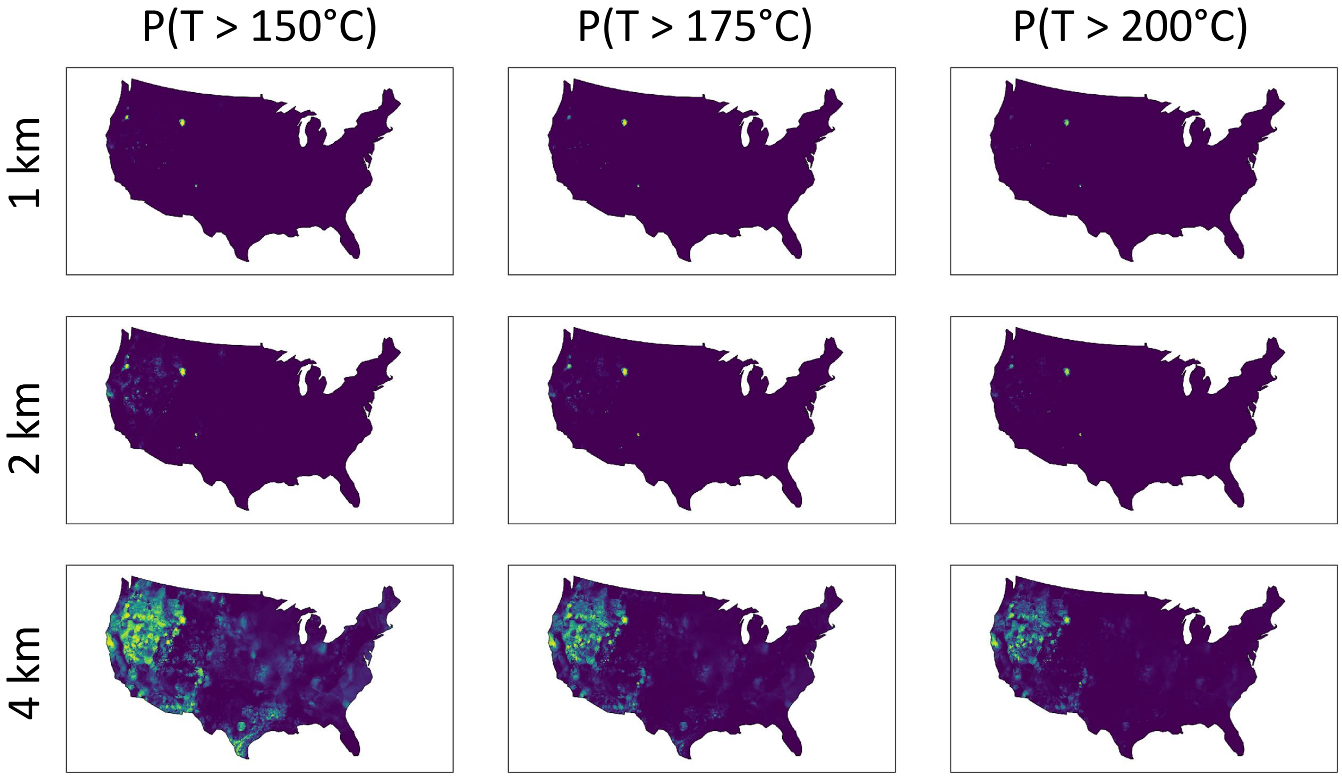
Our Model (Data-driven)

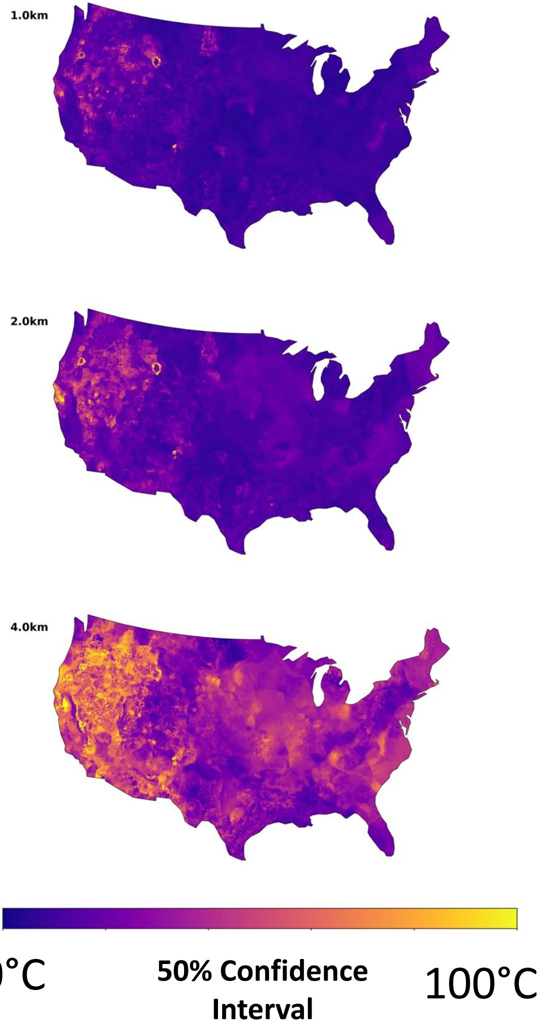
Depth: 4km

Data-rich examples

Adversarial examples

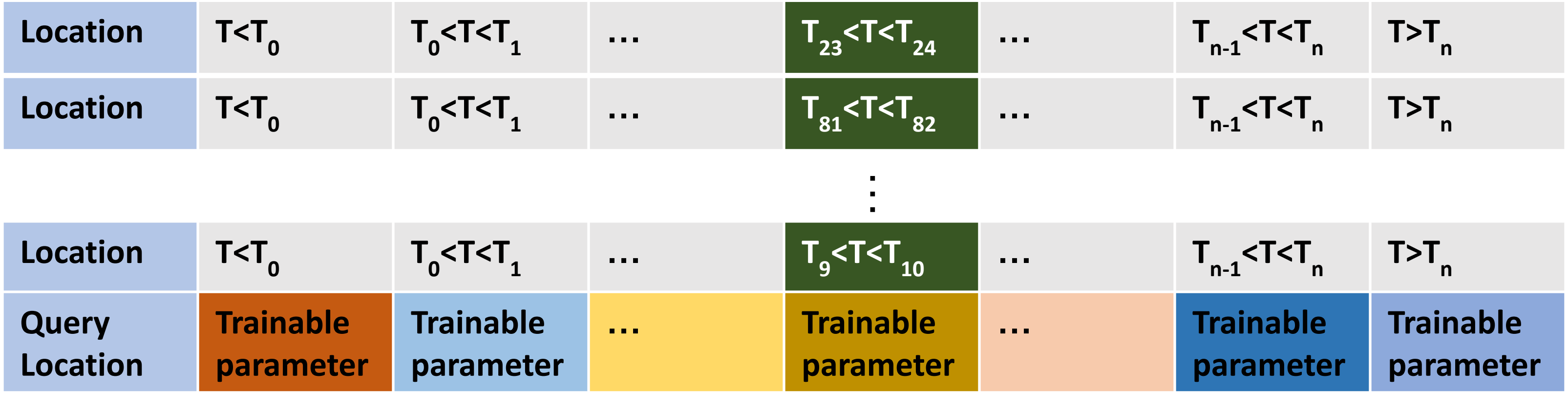
We can map uncertainty too





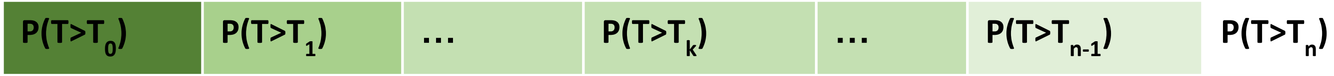
To enable UQ, we turn regression into classification

- The input to the model looks like this:



Location	$T<T_0$	$T_0<T<T_1$	…	$T_{23}<T<T_{24}$	…	$T_{n-1}<T<T_n$	$T>T_n$
Location	$T<T_0$	$T_0<T<T_1$	…	$T_{81}<T<T_{82}$	…	$T_{n-1}<T<T_n$	$T>T_n$
				⋮			
Location	$T<T_0$	$T_0<T<T_1$	…	$T_9<T<T_{10}$	…	$T_{n-1}<T<T_n$	$T>T_n$
Query Location	Trainable parameter	Trainable parameter	…	Trainable parameter	…	Trainable parameter	Trainable parameter

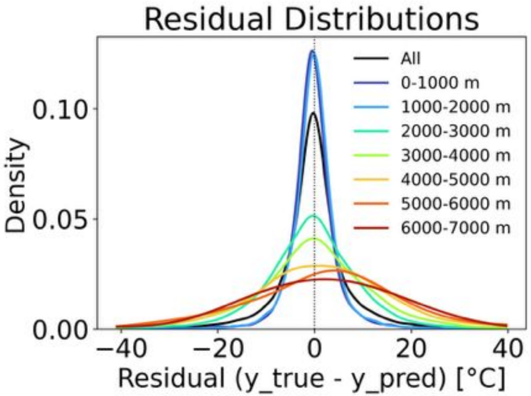
- The output looks like this



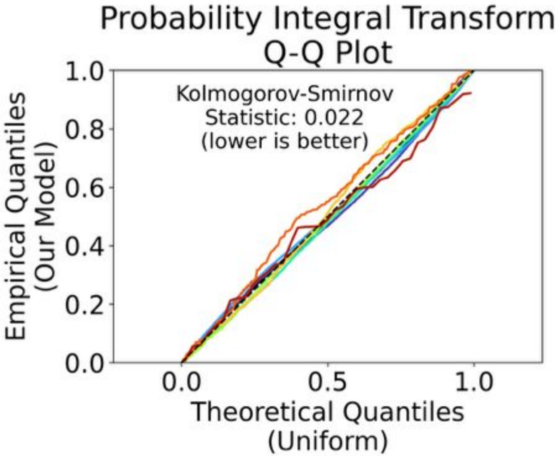
$P(T>T_0)$	$P(T>T_1)$	…	$P(T>T_k)$	…	$P(T>T_{n-1})$	$P(T>T_n)$

- Our loss is the sum of the binary cross-entropy for each of the classification tasks in the output (Is $T>T_k$ or not?)

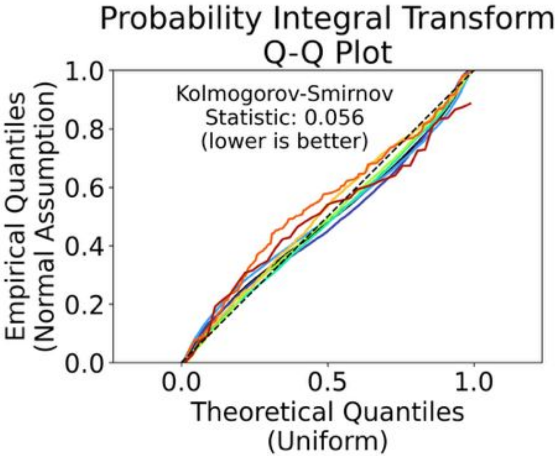
Our uncertainty quantification is accurate and validated



Residual Distributions



Probability Integral Transform Q-Q Plot

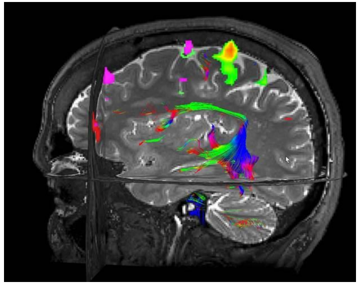


Probability Integral Transform Q-Q Plot

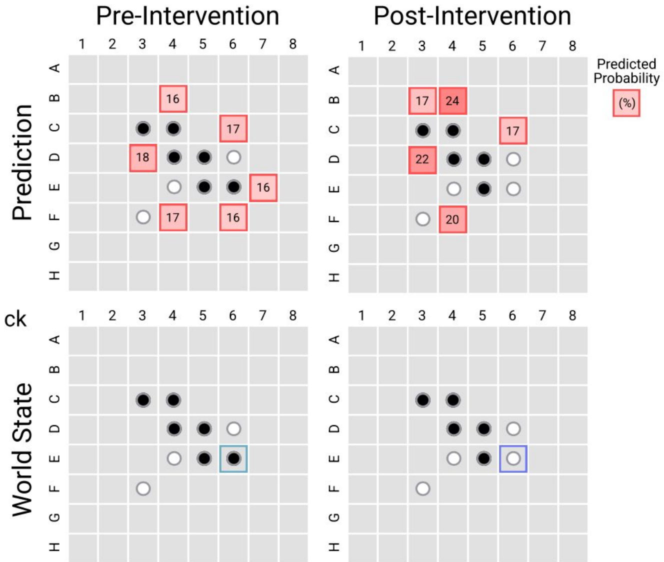
- TE2 Preview predicts the full distribution of the temperature rather than just a regression where one number is output – the prediction
- The model residuals are approximately Gaussian. Despite this, our nonparametric UQ approach outperforms a Gaussian approximation for the residual distribution

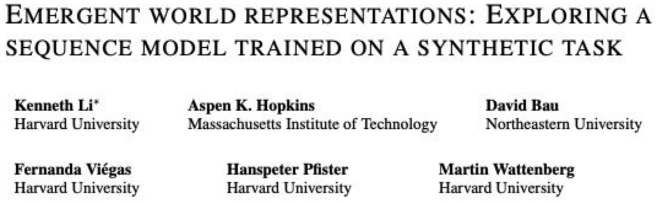
We looked inside the machine learning black box

- Researchers have been able to crudely reconstruct thoughts (e.g., language) and perceptions (e.g., images) from fMRI images of human brains
- We did something similar with our model to understand what it “perceives” about the Earth’s subsurface, then tricked it (“intervened”) to see how it uses the perception



With ML models, you don’t just get to see what signals in the “brain” are being activated, you can amplify or weaken signals





EMERGENT WORLD REPRESENTATIONS: EXPLORING A SEQUENCE MODEL TRAINED ON A SYNTHETIC TASK

Kenneth Li*
Harvard University

Aspen K. Hopkins
Massachusetts Institute of Technology

David Bau
Northeastern University

Fernanda Viégas
Harvard University

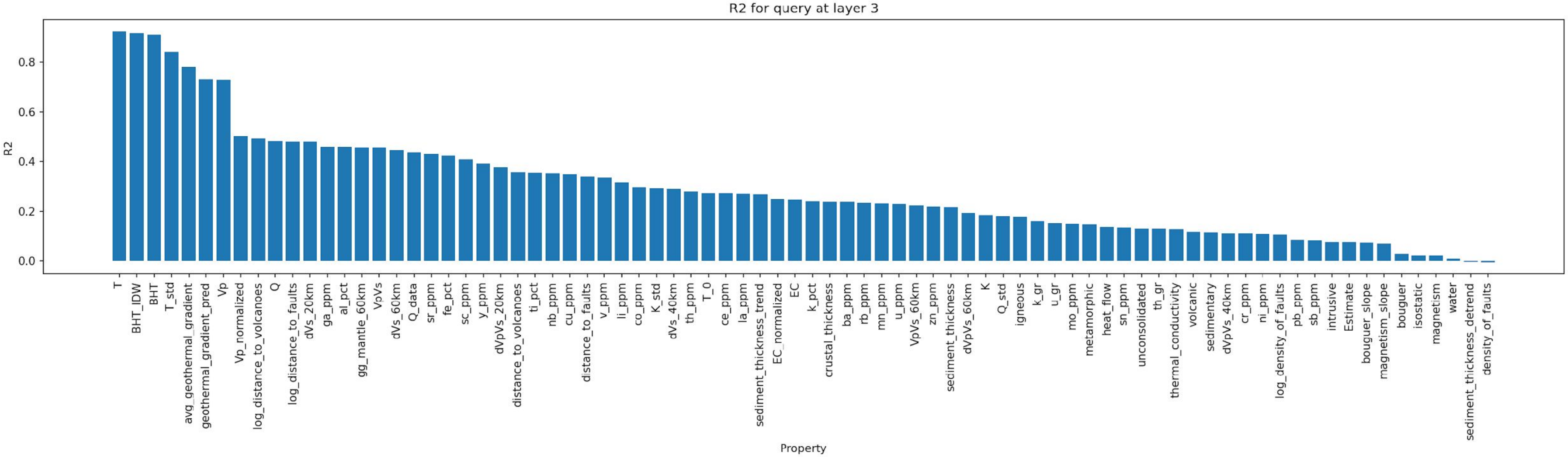
Hanspeter Pfister
Harvard University

Martin Wattenberg
Harvard University

ICLR 2023

Los Alamos
NATIONAL LABORATORY

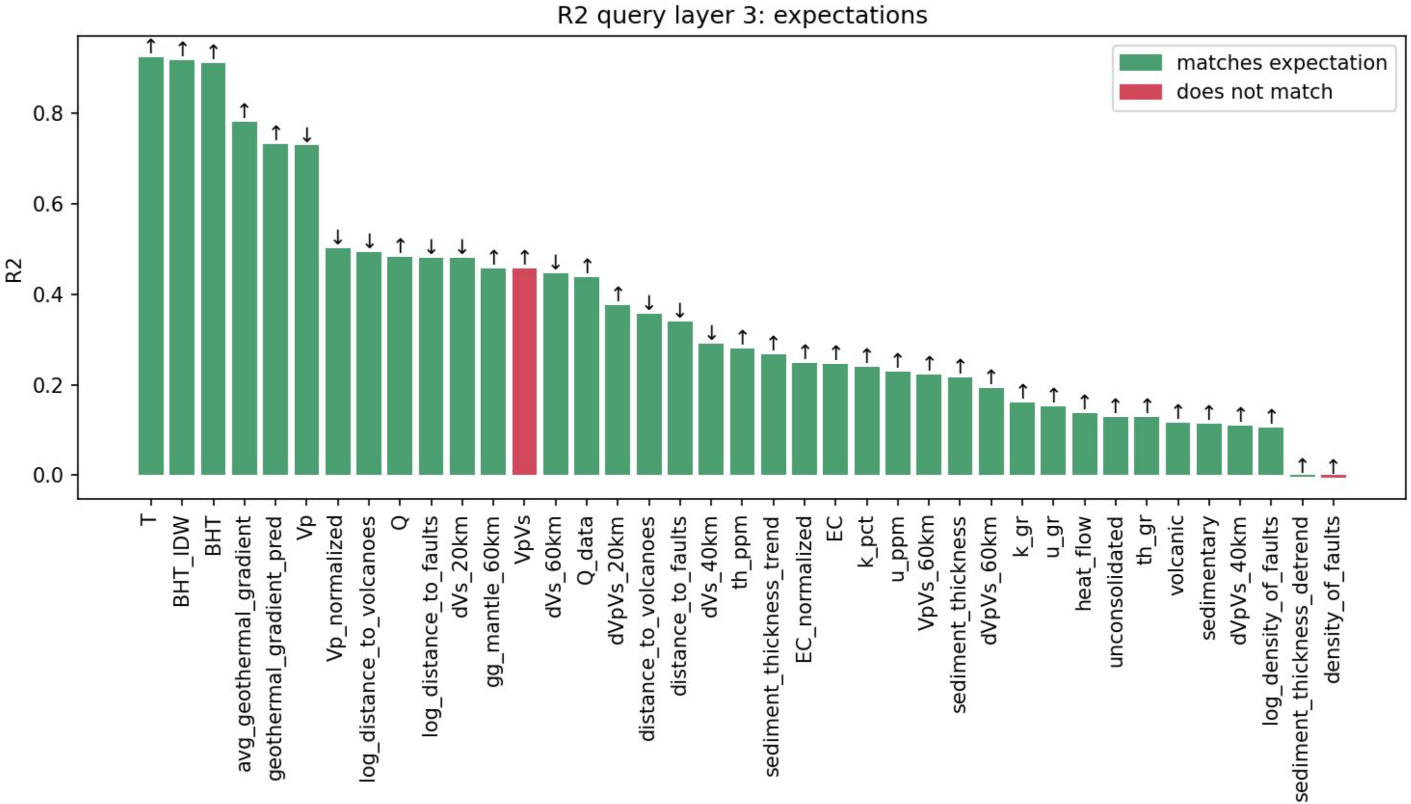
We found it has a crude representation of many subsurface properties



Emphasis on the crudeness. The R^2 is in the range of 0.1-0.5 for most properties, with a few that are higher, which are very closely related to the model’s temperature prediction

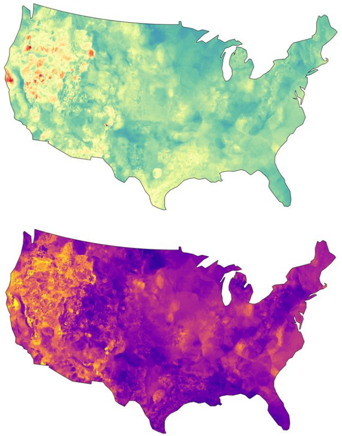
It doesn’t just have the representation. It uses the representation in an intuitive way

- For variables in the Stanford Thermal Model’s dataset, the use of the feature aligns with the physically intuitive use in 37/39 cases
- For example, increasing distance to the nearest volcano tends to decrease temperature
- In one of the two wrong uses (density of faults), the model doesn’t really have a representation



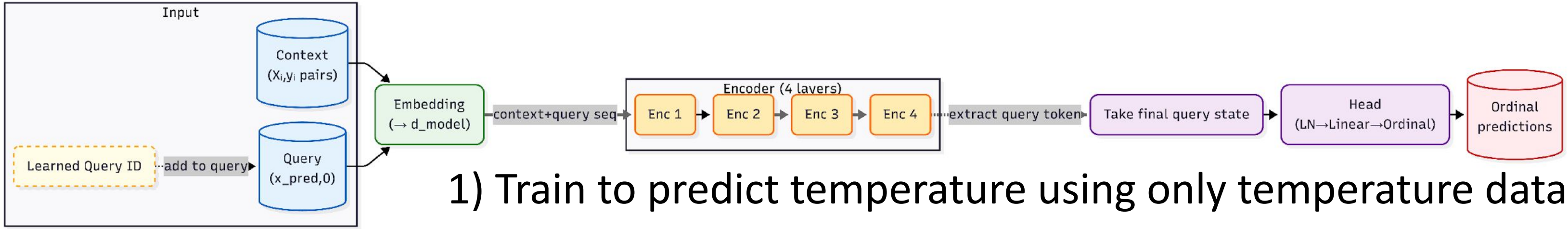
Conclusions

- There is an opportunity to build a ChatGPT for the subsurface that speaks all the dialects that Darwin mentioned and translates them into actionable insights
- We have built an initial model and are rapidly progressing toward Transparent Earth 2.0 with major performance improvements and UQ
- Bitter lesson: breakthrough progress eventually arrives by an opposing approach based on scaling computation by search and learning
  - Don’t encode your knowledge in your models!

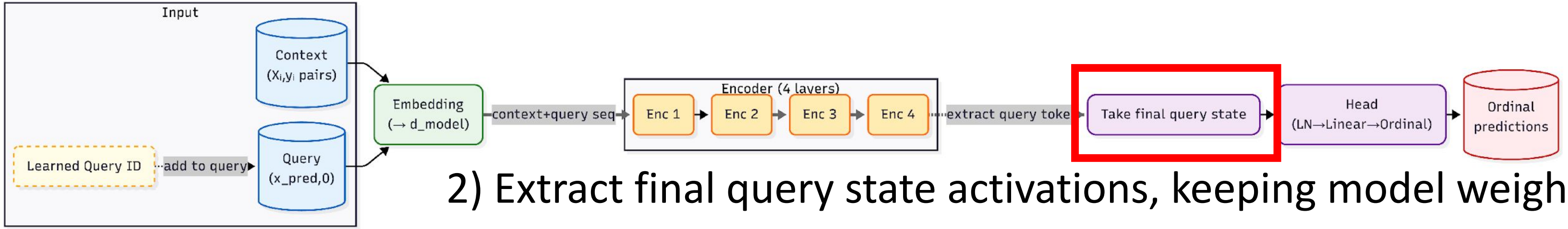


Model	USA	Alberta	UK	Australia
Stanford Thermal Model	6.4C	N/A	N/A	N/A
Transparent Earth 1	5.8C	15.0C	9.0C	30.1C
Transparent Earth 2 preview	**4.8C**	**3.5C**	**5.5C**	**6.4C**

We search for directions in activation-space to correspond to a variety of subsurface properties



1) Train to predict temperature using only temperature data



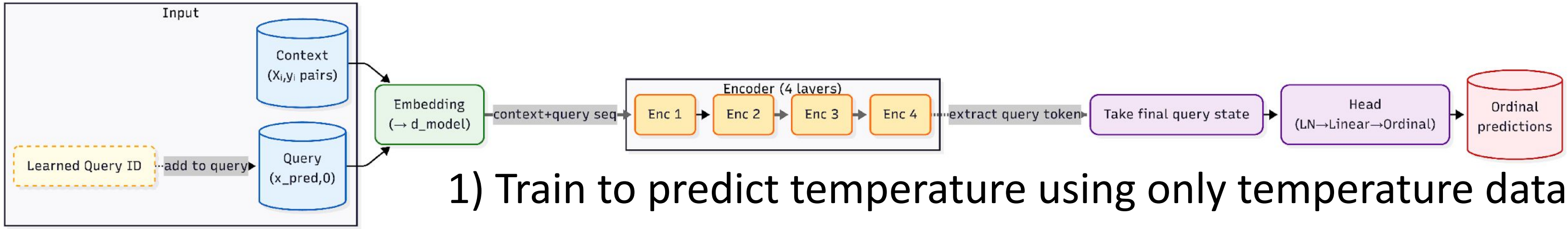
2) Extract final query state activations, keeping model weights frozen

3) Crustal thickness ≈ (final query state activations) · (trainable array) + trainable bias

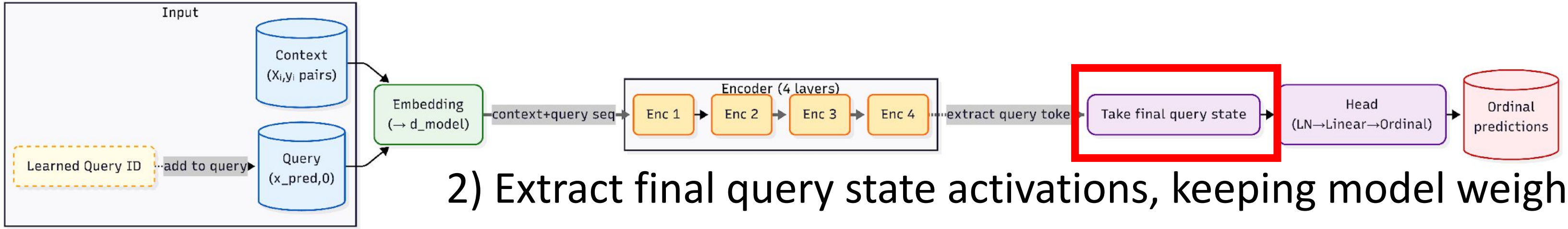
4) Repeat this for many properties besides crustal thickness

If this works, it shows the model has learned a crude representation of the subsurface that goes beyond temperature, like the Othello GPT paper showed it had a world model of the board

We search for directions in activation-space to correspond to a variety of subsurface properties



1) Train to predict temperature using only temperature data



2) Extract final query state activations, keeping model weights frozen

3) Crustal thickness ≈ (final query state activations) · (trainable array) + trainable bias

4) Intervention: (final query state activations) = (final query state activations) + α(trainable array)

If we make the crust thicker, it should cause the temperature to decrease. If the model does that, it would show it's using the crude representation of the subsurface in the right way.