EarthPT

Aspia Space:
MJ Smith, L Fleming, JE Geach

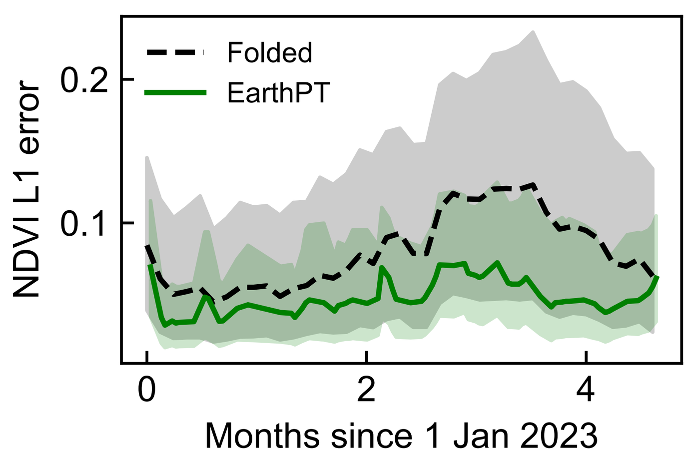
From LLMs to `LOMs'

Inspired by the recent successes of large language models, we asked ourselves: **'would similar models also work for earth observation time series data and if they do, would they perform as well as LLMs at scale?'**.

To answer this question we trained a 700 million parameter decoding transformer neural network model (a 'Large Observation Network') autoregressively on 1.5 billion tokens of ClearSky Sentinel-2 data.

Forecasting

We found that EarthPT is an effective forecasting model, and can predict future observations with a good degree of accuracy, given a conditioning timeseries. We compared EarthPT to a baseline model that predicts the next observation via a simple phasefolding method, where we folded the timeseries by year:



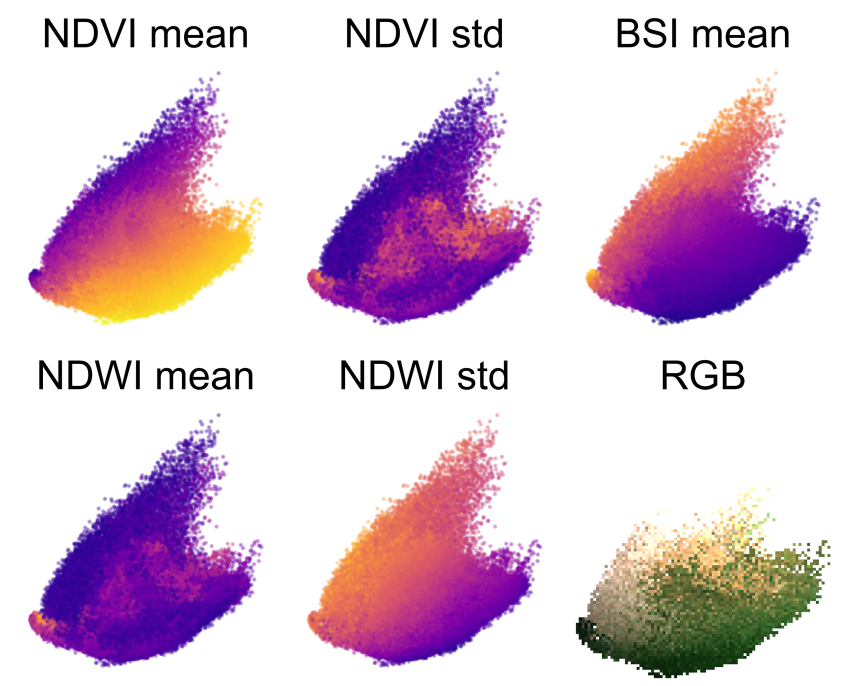
The centre of this poster shows further forecasting results for EarthPT, where we can see that it is able to predict the future with a good degree of accuracy. The EarthPT forecasts are in white, and the conditioning ground truth is in colour.

A **foundation model** trained on **1.5 billion tokens** of EO data can **zero-shot predict the future.**

Funded by UK Government
CORNWALL COUNCIL one and all • onen hag oll
Council of the ISLES OF SCILLY
GOOD GROWTH CORNWALL & ISLES OF SCILLY SHARED PROSPERITY FUND
LEVELLING UP

Meaningful embeddings

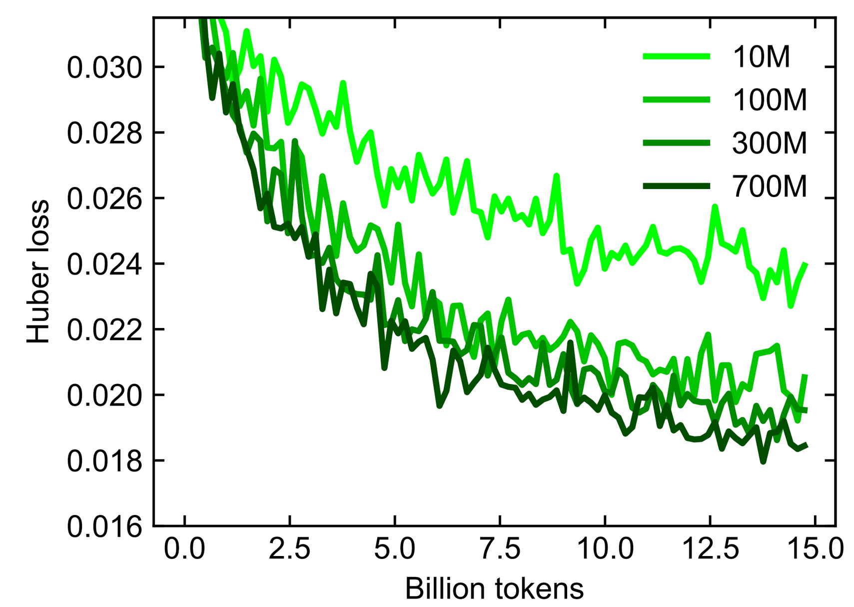
We also find that EarthPT emergently learns remote sensing indices such as the NDVI. We can see this in our model embeddings:



This suggests that EarthPT learns to represent the underlying physical environmental processes, and so these embeddings could be used as a basis for downstream tasks such as crop yield or drought predictions.

Scaling tests

We trained a series of models of differing sizes to see how EarthPT's performance scales with model size. We found that its performance scales similarly to LLMs, suggesting that there is still much to be squeezed out of this architecture.



Excitingly, the number of tokens available for training is of order 10^{15}, so we are not currently data constrained. If neural scaling laws hold, then improving EarthPT and similar large observation models is a simple matter of scaling data and compute.