



### **NC Programme**

National Capabilities for UK Challenges (NC-UK)

### **Project Title**

Fine-tuning a state-of-the-art geospatial foundation model for UK land cover classification.

### **Supervisors**

Jeremy Carter - UKCEH

Clare Rowland – UKCEH

Michael Hollaway - UKCEH

### **UKCEH Site**

Lancaster

### **Project Overview**

This pilot project explores new approaches in Earth observation by testing whether the state-of-the-art Prithvi EO-2.0 geospatial foundation model can be fine-tuned to improve the production of the UKCEH Land Cover Map (LCM). At present, the UK LCM is created using supervised learning methods (specifically random forests) that depend heavily on high-quality labelled training data. Much of this data is inherited from previous LCM versions, which works well for existing land-cover classes. However, when new analytical needs arise, such as adding extra classes, there is a significant time and resource cost involved in generating sufficient new labelled data.

To address this, the project will explore transfer learning by fine-tuning the Prithvi EO-2.0 model, which has been pre-trained using self-supervised learning on very large volumes of multi-spectral satellite imagery by NASA and IBM. This pre-training allows the model to learn a compact, general representation of satellite data that captures important patterns while reducing noise. These learned representations can then be reused for downstream tasks such as land-cover classification, reducing the need for labelled data while improving efficiency and expected performance. This approach also makes it easier to adapt and extend the workflow to new use cases, including the exploration of ultra-high-resolution classification.

The intern will build on existing tools, frameworks and tutorial resources to validate performance against current LCM workflows using a test region (e.g. the Lake District). The findings will help assess whether geospatial foundation models can extend and accelerate future LCM production and support NC-UK's wider ambitions for rapid, national-scale

environmental monitoring. The work will also contribute a reproducible notebook to the NC-UK data science toolbox.

### **Key tasks**

- Run the Prithvi EO-2.0 model using tutorial code and demonstrate its ability to reconstruct masked satellite images.
- Test an existing fine-tuned Prithvi model (e.g. flood-detection) on example imagery.
- Download and prepare a Sentinel-2 dataset for a selected UK test area (e.g. the Lake District) and assess the model's reconstruction performance.
- Access the matching UKCEH Land Cover Map (LCM) data for the test region.
- Fine-tune the Prithvi model for land-cover classification using the LCM subset.
- Create training and test datasets, run experiments, and evaluate model performance against current methods.
- Identify any bottlenecks or limitations and make recommendations for future work.
- Work closely with the Digital Research and Land Use teams, using existing scripts and workflows.
- Adapt tasks as needed during the pilot study based on progress and findings.

### **Expected Outputs / Deliverables**

- A proof-of-concept analysis demonstrating whether a geo-spatial foundation model (Prithvi EO-2.0) can be fine-tuned to produce land-cover classifications for a small UK test region.
- A GitHub repository following UKCEH best-practice, containing clean, reproducible code using the TerraTorch package for fine-tuning Prithvi, plus saved configs and outputs.
- A clear, well-documented walk-through Jupyter Notebook showing how to load Prithvi, preprocess UK Sentinel-2 data, and run the fine-tuning workflow. This will be suitable for inclusion in the UKCEH Environmental Data Science Toolbox.
- A short feasibility and recommendations report summarising lessons learned, key bottlenecks (e.g., compute, data requirements), and strategic suggestions for scaling the approach to wider national or historical datasets.

### **Essential Skills**

- This internship is suitable for PhD students with an interest in environmental monitoring, machine learning, and geospatial analysis.
- Strong Python programming skills
- Familiarity with GitHub and version control
- Experience using VS Code or similar IDEs
- Experience with Jupyter Notebooks

- Interest in geospatial and environmental applications
- Organised, proactive, and able to document work clearly
- Good communication skills and experience collaborating in a team

#### Desirable Skills

- Familiarity with machine learning frameworks (e.g., PyTorch)
- Experience with classification tasks or similar ML workflows
- Interest in best coding practices, including structured GitHub workflows

#### Extra Information

The internship provides high-value training to the applicant to gain highly sought-after, industry- and research-relevant skills in state-of-the-art AI methods and transfer learning. The intern will also gain valuable networking opportunities being embedded within the Digital Research team at UKCEH and collaborating with researchers to improve a key national scale environmental data product.