University of Exeter

Centre for Environmental Intelligence

Generalisation of geospatial foundation models for wildfire burn scar detection in the UK

Remy Vandaele, Edward Pope, Geoffrey Dawson, Anne Jones, David Moffat, Hywel Williams, Claire Belcher, Chunbo Luo

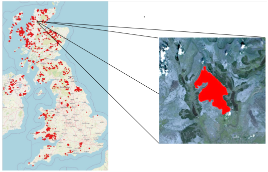
Centre for Environmental Intelligence, University of Exeter

Met Office IBM Research Plymouth Marine Laboratory

23 June 2026

Definition and motivation

The task is not active fire detection, but **burn scar** mapping **after** the fire.



Why this matters

Better burn scar maps support:

- Fire risk understanding
- Ecological impact assessment
- Early warning systems
- Land use decisions

Current problem

- UK Fire and Rescue Services attended about **32,000 wildfires per year** over 2009-2017.
- Operational databases such as EFFIS rely on MODIS/VIIRS products typically detecting wildfires larger than 30 ha
- In the UK, **99% of wildfires are smaller than 30 ha**.

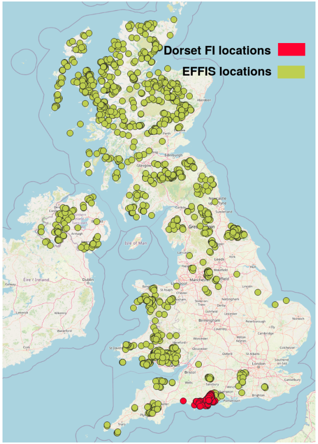
Research question

Using high resolution satellite data, can geospatial foundation models fine tuned over datasets of large (UK) wildfires generalise to more typical UK wildfire burn scars?

Main objectives

1. Build a dataset combining large (**EFFIS**) and small (**Dorset Fire Intervention - DFI**) wildfire polygons with high resolution satellite data (Sentinel-2 and Landsat-8)
2. Evaluate **Prithvi-EO-2.0** and **TerraMind** on UK wildfire burn scar detection
3. Test **transferability** across datasets
4. Compare against **baselines** (U-Net, Random Forest, NBR and BAI)

Dataset: EFFIS and DFI polygons + HLS imagery



EFFIS fires are distributed across the UK and DFI fires concentrate in Dorset.

	EFFIS	DFI
Source	Satellite product	Intervention records
Typical fire size	large	small
Original records	1,409	1,147
Mean area	115.11 ha	0.73 ha
Usable fires	908	113

- Harmonized Landsat and Sentinel (HLS) product: 30m resolution, 6 bands.
- Post-fire window: 10 days after fire start.
- Remove fires with no data / hidden by clouds

Method: fine-tune GFMs for segmentation

Geospatial Foundation Models

- **Prithvi-EO-2.0**: ViT, masked autoencoder pre-training (HLS data).
- **TerraMind**: multimodal EO foundation model with token and pixel-level learning, modality reconstruction pre-training

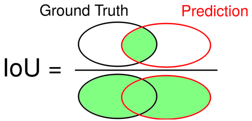
Baselines

- Prithvi and TerraMind without pre-training.
- U-Net and Random Forest models
- NBR and BAI spectral indices.
- Prithvi fine-tuned for wildfire burn scar detection in the USA.

Evaluation protocol

Metrics

- $\mathrm{IoU}_{\mathrm{Wf}}$: overlap between predicted and true wildfire pixels.
- MIoU: average of wildfire and background IoU.
- Det-0.1: percentage of images with $\mathrm{IoU}_{\mathrm{Wf}} > 0.1$.

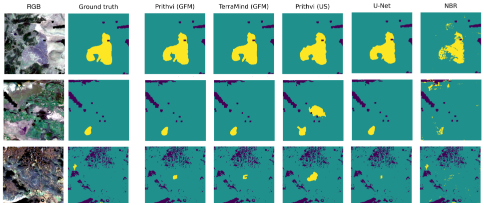


Training setups

- EFFIS only.
- DFI only.
- EFFIS + DFI mixed.

Dataset split: 5-fold cross-validation, with spatially co-located wildfire events kept in the same fold to avoid leakage.

Results: models trained/fine-tuned on EFFIS, evaluated on EFFIS

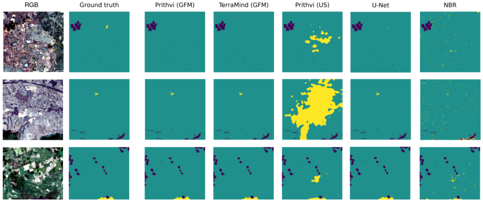


EFFIS examples: GFMs and U-Net localise burn scars; NBR creates more false positives.

Model	**IoU_{Wf}**	**Det-0.1**
GFM models		
Prithvi-EO-2.0	0.62 (0.60-0.64)	**82.62 (80.25-84.91)**
TerraMind	**0.70 (0.67-0.72)**	81.89 (79.44-84.32)
ML models, no GFM pre-training		
Prithvi-EO-2.0	0.59 (0.56-0.61)	78.43 (75.81-80.95)
TerraMind	0.68 (0.65-0.70)	81.41 (79.00-83.77)
U-Net	0.54 (0.51-0.56)	76.16 (73.64-78.71)
Random Forests	0.26 (0.24-0.27)	60.62 (58.11-64.06)
Remote sensing indices		
NBR	0.23 (0.22-0.24)	62.24 (59.80-64.72)
BAI	0.16 (0.15-0.18)	40.41 (37.41-43.43)

- Prithvi and TerraMind give the best segmentation quality.
- GFM pre-training does not have a strong influence.
- Spectral indices remain useful, but less precise.

Results: models trained/fine-tuned on DFI, evaluated on DFI



DFI examples: small burn scars can be subtle;
US-trained Prithvi often over-predicts.

Model	$\mathrm{IoU_{Wf}}$	Det-0.1
GFM models		
Prithvi-EO-2.0	0.34 (0.26-0.42)	39.82 (30.97-48.67)
TerraMind	0.33 (0.25-0.41)	36.28 (27.43-45.13)
ML models, no GFM pre-training		
Prithvi-EO-2.0	0.31 (0.24-0.39)	36.28 (27.73-45.43)
TerraMind	0.33 (0.25-0.41)	36.58 (28.02-45.13)
U-Net	**0.35 (0.27-0.44)**	**40.12 (31.27-48.67)**
Random Forests	0.13 (0.09-0.16)	36.28 (27.73-45.13)
Remote sensing indices		
NBR	0.04 (0.02-0.05)	12.15 (6.78-18.08)
BAI	0.03 (0.02-0.06)	8.19 (3.95-12.99)

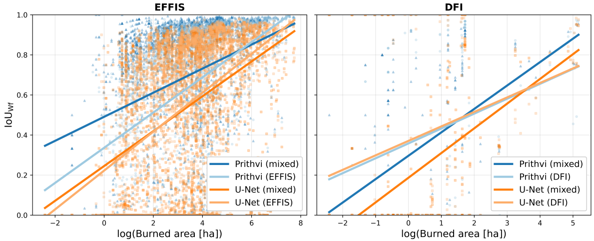
- DFI results are lower across all methods.
- U-Net is slightly ahead, but confidence intervals overlap with GFMs.
- NBR/BAI are clearly insufficient for these fires.

Transferability experiment

Model	EFFIS IoU_{Wf}	EFFIS Det-0.1	DFI IoU_{Wf}	DFI Det-0.1
Fine-tuned on EFFIS only				
Prithvi-EO-2.0	0.62 (0.60-0.64)	**82.62 (80.25-84.91)**	0.10 (0.06-0.14)	18.88 (12.39-25.96)
TerraMind	**0.70 (0.67-0.72)**	81.89 (79.44-84.32)	0.09 (0.06-0.13)	18.29 (11.80-25.08)
U-Net	0.54 (0.51-0.56)	76.16 (73.64-78.71)	**0.11 (0.07-0.15)**	**20.35 (13.57-27.73)**
Random Forests	0.26 (0.24-0.27)	61.02 (58.11-64.06)	0.06 (0.03-0.09)	15.93 (10.03-22.42)
Fine-tuned on DFI only				
Prithvi-EO-2.0	0.15 (0.14-0.17)	31.89 (29.04-34.73)	0.34 (0.26-0.42)	39.82 (30.97-48.67)
TerraMind	0.11 (0.10-0.12)	23.25 (20.70-25.84)	0.33 (0.25-0.41)	36.28 (27.43-45.13)
U-Net	0.15 (0.13-0.17)	30.27 (27.50-33.11)	**0.35 (0.27-0.44)**	**40.12 (31.27-48.67)**
Random Forests	**0.23 (0.21-0.24)**	**57.20 (54.3-60.43)**	0.13 (0.09-0.16)	36.28 (27.73-45.13)
Fine-tuned on USA data				
Prithvi-EO-2.0	0.16 (0.15-0.18)	31.15 (28.63-33.55)	0.02 (0-0.04)	2.95 (0.59-6.19)
Mixing both (EFFIS and DFI) datasets				
Prithvi-EO-2.0	**0.69 (0.66-0.71)**	**84.68 (82.49-86.97)**	**0.27 (0.20-0.33)**	**39.53 (31.27-48.38)**
TerraMind	0.67 (0.64-0.69)	81.74 (79.30-84.07)	**0.27 (0.20-0.34)**	37.76 (28.91-46.31)
U-Net	0.53 (0.50-0.55)	76.19 (73.60-78.75)	0.16 (0.11-0.21)	28.02 (20.35-35.99)
Random Forests	0.29 (0.27-0.31)	66.57 (63.69-69.42)	0.10 (0.07-0.13)	25.37 (18.29-33.33)

- Models trained on EFFIS transfer poorly to DFI, and vice versa.
- USA Prithvi wildfire model performs poorly, especially on DFI.
- Mixed fine-tuning is the most robust route to detect DFI and EFFIS fires.
- GFMs are the only models able to take advantadge of the mixed training!

Error analysis



Detection IoU increases with burned area; DFI remains harder at similar sizes.

- Smaller fires are consistently harder to segment.
- Mixing EFFIS + DFI improves robustness for some size ranges, but creates trade-offs.
- Open landscapes such as semi-natural grassland are easier.
- Coniferous woodland and low-contrast vegetation structures are harder.

Take-home message

Main conclusion

Geospatial foundation models are promising for UK wildfire burn scar detection, but their generalisation depends on **diverse fine-tuning data** that captures fire size, severity, morphology, and land-cover variation.

What worked

- GFMs improve robustness with mixed fine-tuning.
- Strong EFFIS performance.
- Better false-positive control than simple indices.

What remains open

- Small, low-intensity DFI fires.
- Higher resolution imagery
- New wildfire datasets

Diverse wildfire datasets are key to generalisable wildfire mapping!

Thank you

Generalisation of geospatial foundation models for UK wildfire burn scar detection

Remy Vandaele
Centre for Environmental Intelligence
University of Exeter

Want to know more?

Centre for Environmental Intelligence

environmentalintelligence.org.uk