Daily Land Surface Temperature Estimation at ~100-Meter Resolution Using a Deep Neural Network

Shaerdan Shataer, Niall McCarroll, Christopher Merchant

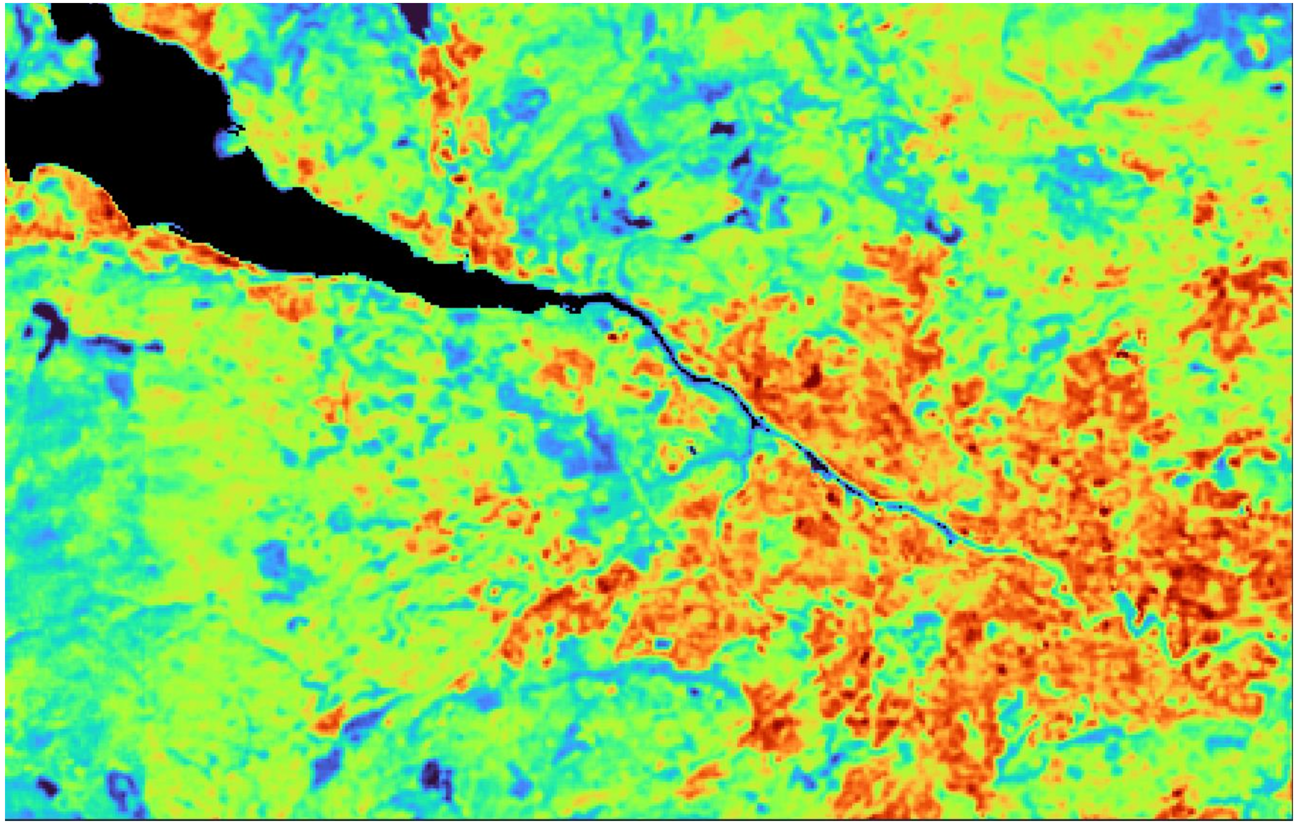
NCEO, University of Reading

JASMIN is operated by the Science and Technology Facilities Council on behalf of the Natural Environment Research Council.

Outline

- Model Objectives
- Input / Target Data
- Model Design
- Model Training
- Data Product
- Model Evaluation
- Discussion

Model 100m LST estimates – Glasgow & the Clyde, 2022-07-19



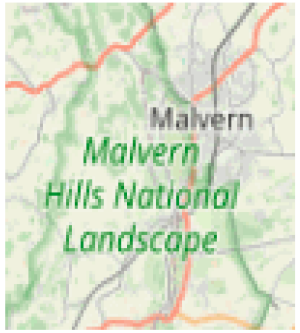
Model Objectives

- We aim to build a model that can infer daily land surface temperature at ~100m resolution, emulating Landsat 8/9.
- Predictors are ~9km resolution ERA5 Skin Temperature (for dynamical input), and static/semi static pattern indicators such as: Land Cover Class (Asphalt, Rock, Grassland, Water etc), Elevation, Albedo Monthly Climatology Mean.
- Day of year (doy) positioning is built in by introducing sine and cosine of doy as two additional input-channels.
- A customized and enhanced U-NET is the architecture.

Input and Target Data

Landsat 8/9 Imagery: Courtesy of USGS

Visible

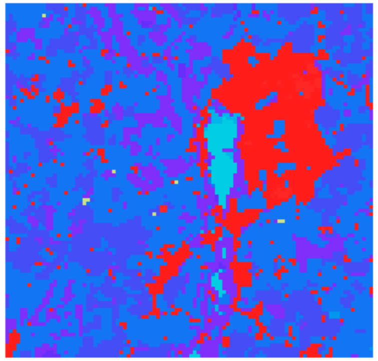


© Open Street Map

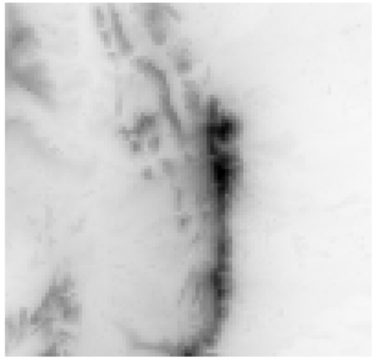
(1) Divide UK up into 40km boxes, randomly assign to test and train
(2) Collect data on 40km boxes, sub-divide 10km (100x100) tiles with <5% cloud
This example training tile covers the town of Malvern, 2019-09-18

Model Inputs

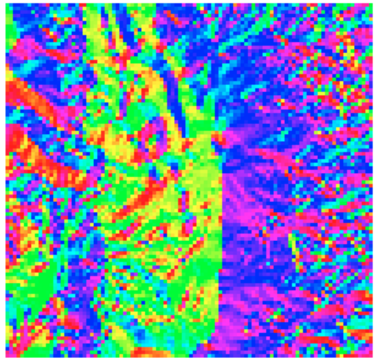
Land cover



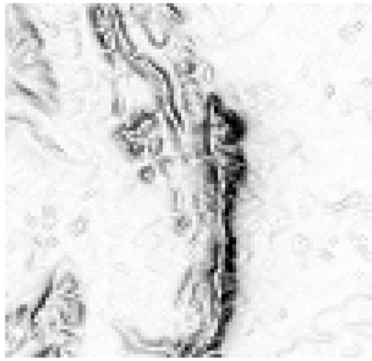
Elevation

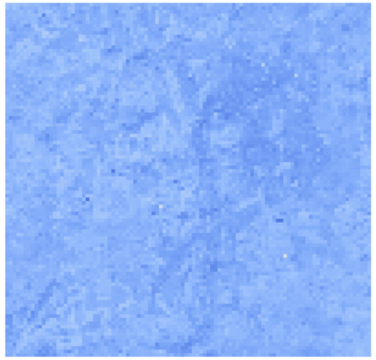


Slope direction

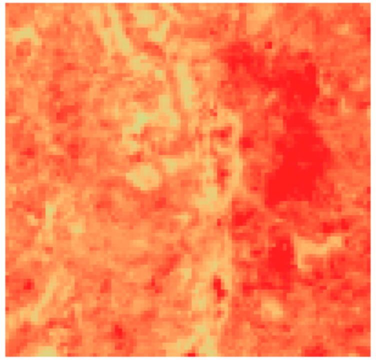


Slope magnitude

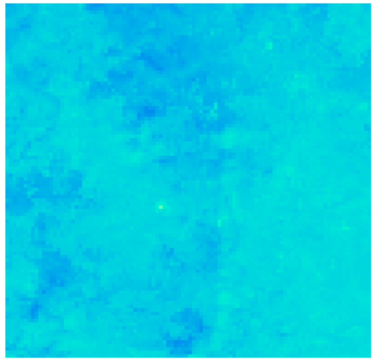




Winter typical Landsat 8/9 LST



Summer typical Landsat 8/9 LST

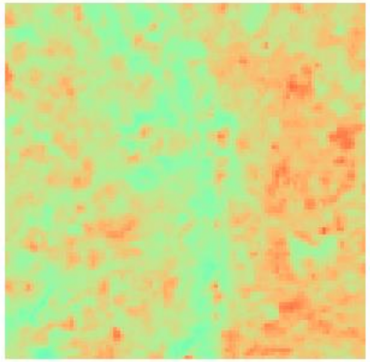


Albedo mean September LST

Era5 SKT: 293.44
Day of Year: 261

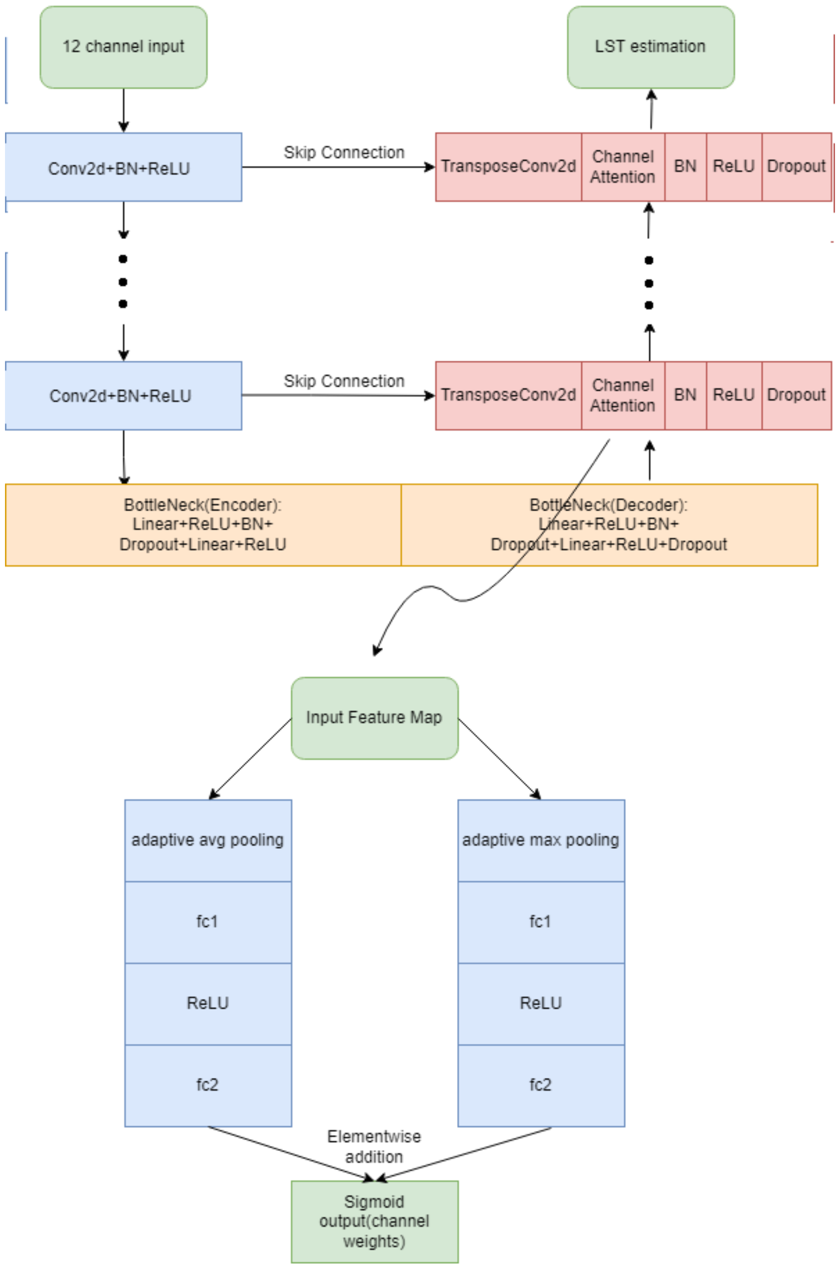
Time-varying Input

Model Target



Landsat 8/9 L2SP LST

Model Design (Architecture)



Model Training

Loss function and regularization:

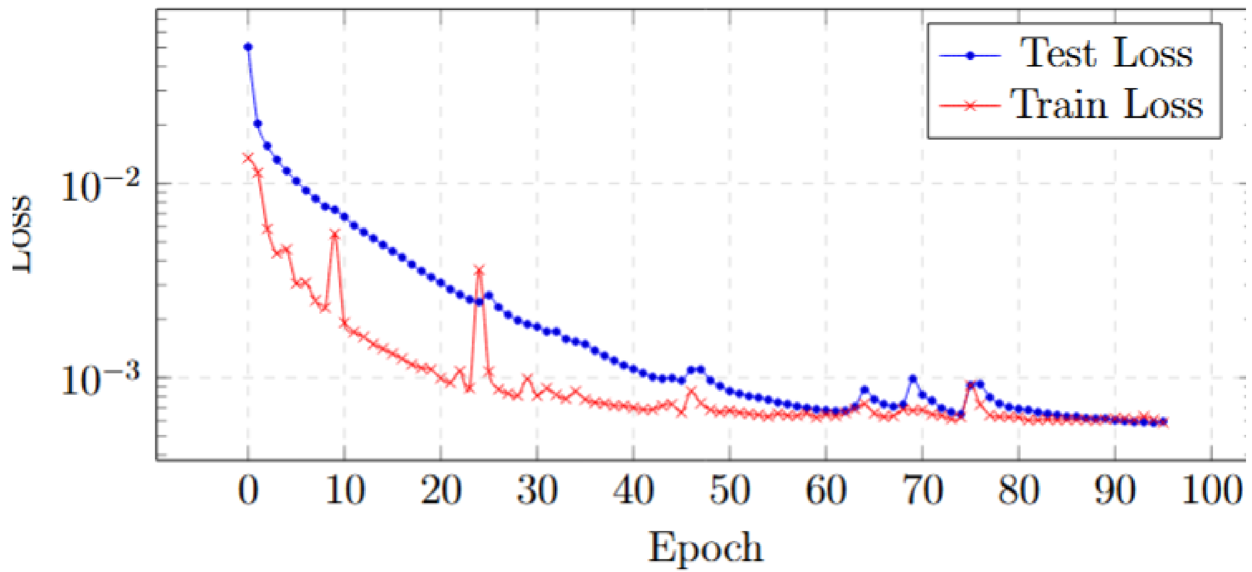
- Spatial pattern related loss: SSIM and Pearson spatial correlation loss
- MSE loss and Huber loss
- Dropout at FC layers (bottleneck); Weight decay AdamW; BN (Batch Normalization)

Data enhancement

Data filtering:

- Screening process to reduce cloud pixels

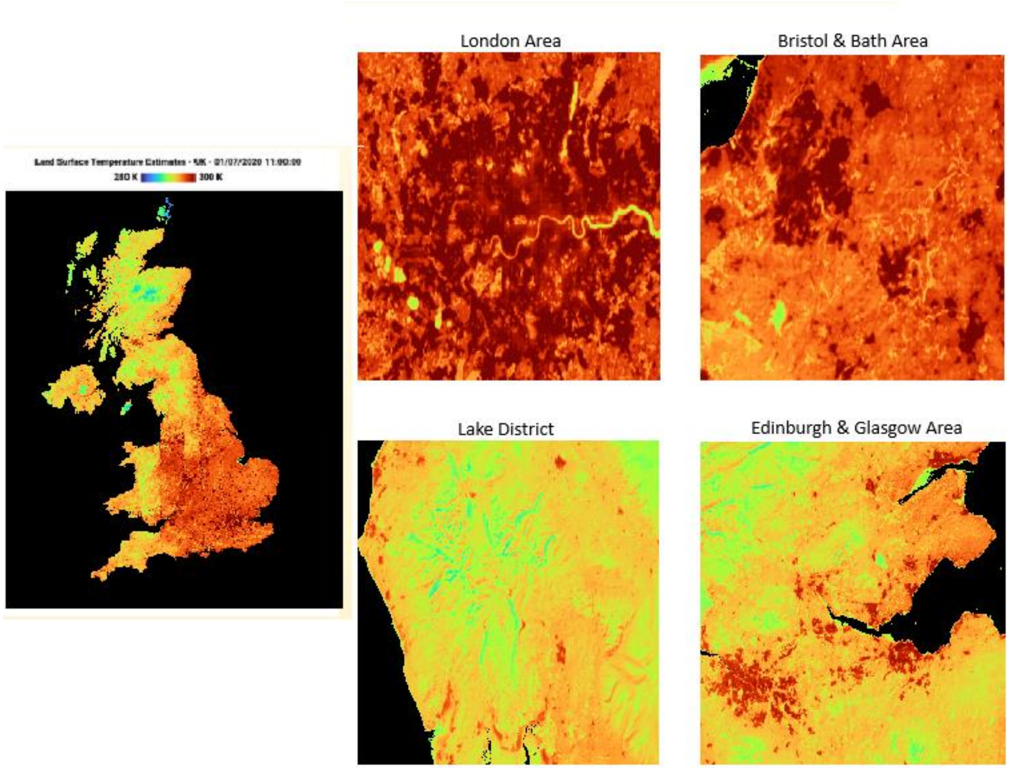
Training:

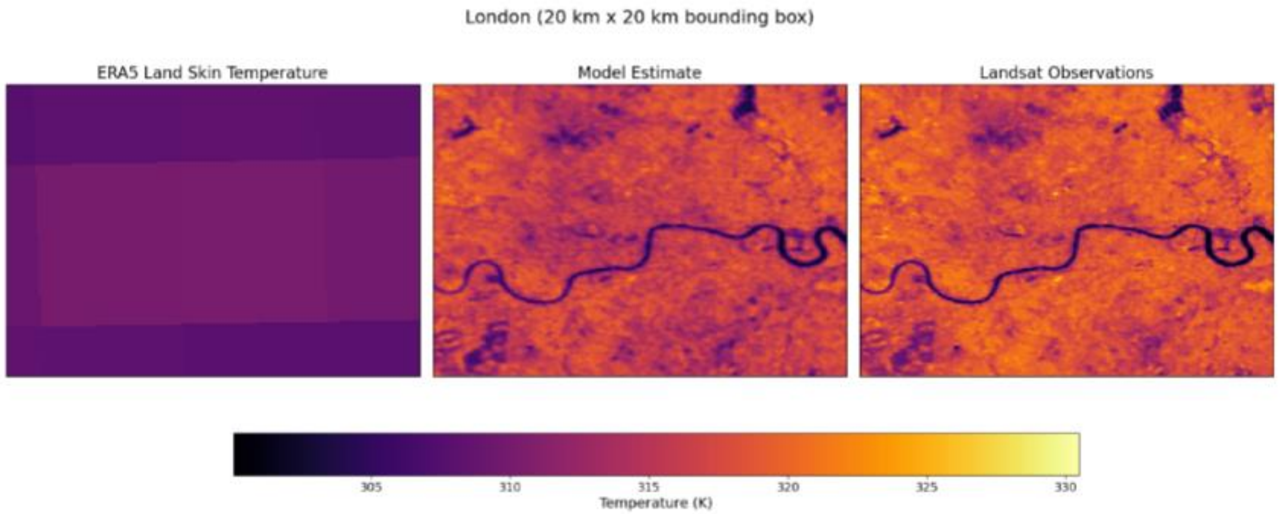


Major Issues:

- Low temporal resolution of ground truth
  - Landsat-9's narrow, shifting swath over the UK limits revisit frequency.
- Persistent cloud contamination
  - Inadequate Landsat-9 cloud mask allows noise to dominate the signal.
- Noisy input predictors
  - Unphysical albedo values and other artefacts degrade model inputs.
- Challenges in data cleaning
  - Noise often lacks a clear, separable signature, making systematic filtering difficult.
- Uneven data coverage
  - Seasonal scarcity (e.g. winter cloudiness), spatial gaps from swath shifts, north–south disparity, and scene clustering (when data exist) all lead to sparse, unbalanced sampling across the UK.
- Per-scene mean temperature estimation generalizes less well than pattern estimation
  - Training MSE is typically three to four times lower than test MSE, whereas pattern-based MSE is nearly identical between training and testing.

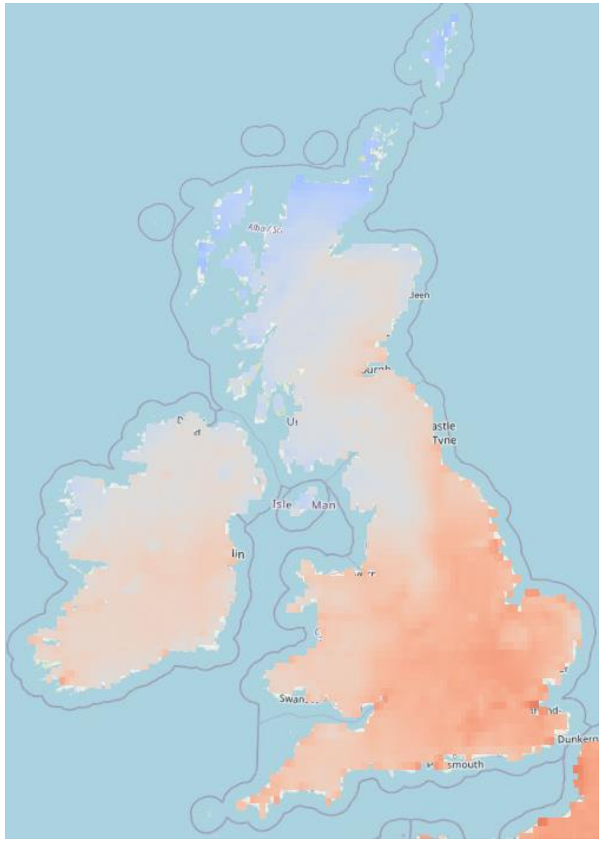
Visual Check

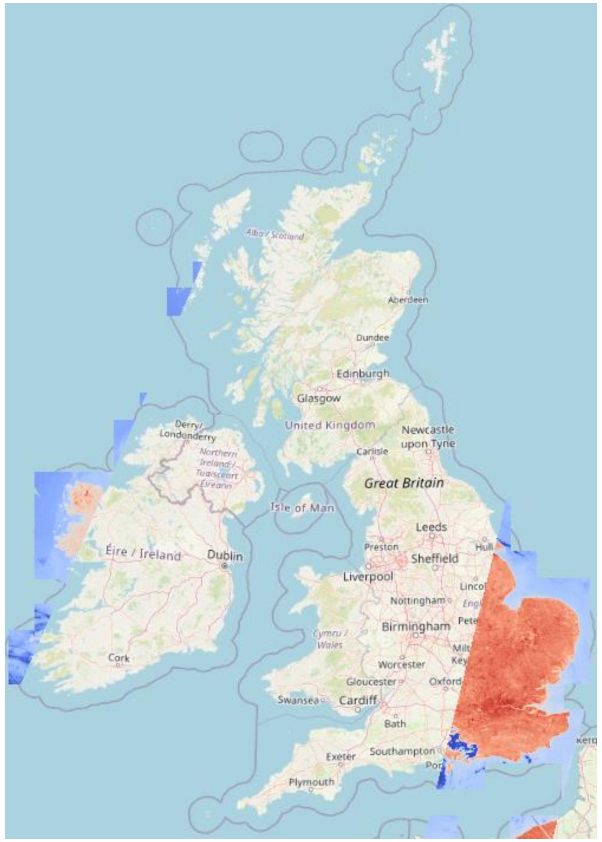


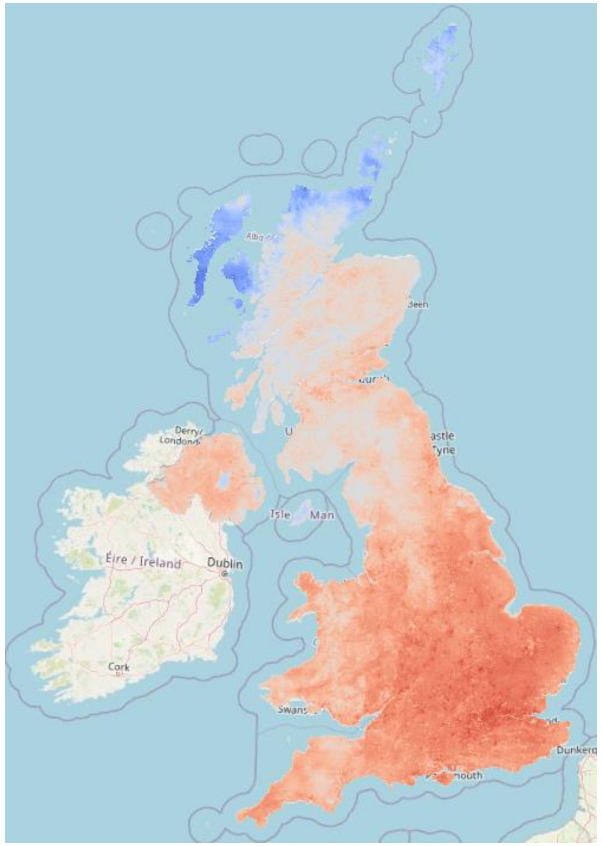


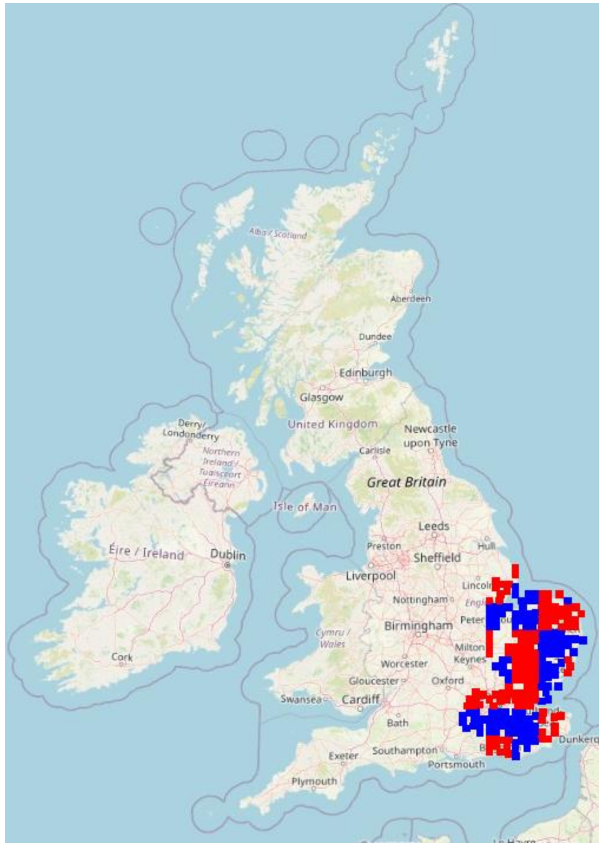
Data Product

- 19/07/2022 – UK, daily 100m gap-filled surface temperature estimates


ERA5 Skin Temperature


Landsat 8/9 Observations

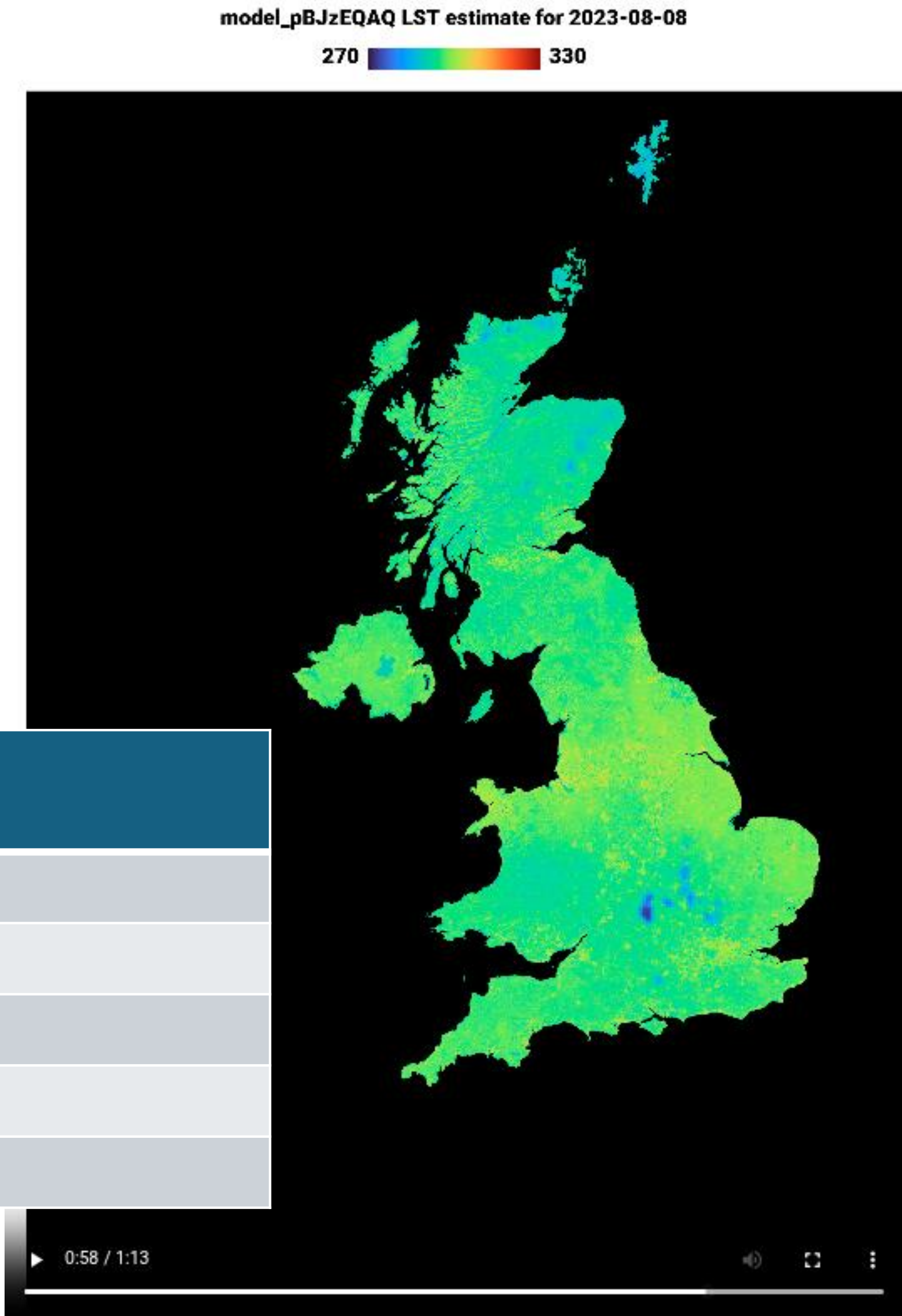

Model LST Estimates


Training/test data locations

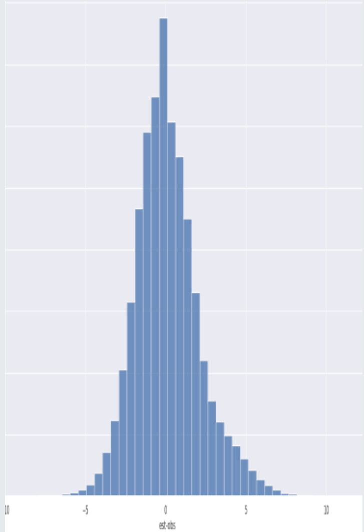
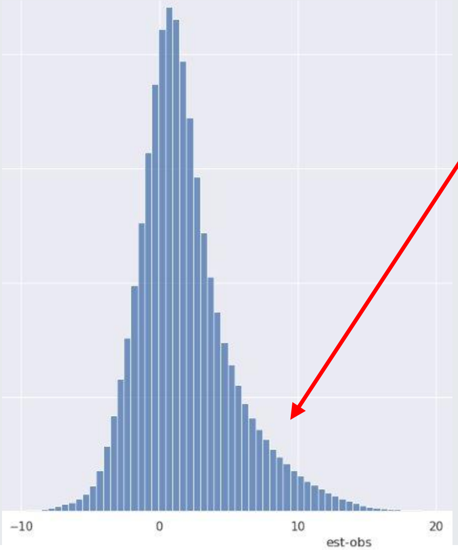
Model Evaluation

- Observations have spatial/temporal bias
- Data product designed for "bring your own evaluation"
- Overall Statistics:

UK / 2014-2023 / model_pBJzEQAQ		
Metric	**Clear sky (test data)**	**All sky**
Mean Absolute Error	1.93 K	2.52 K
Median Absolute Error	1.3 K	1.6 K
RMSE	3.16 K	4.11 K
Pixel Count	104M	2323M



Model Evaluation – London - Summer

London Area / 2014-2023 - SUMMER (JJA) / model_pBJzEQAQ		
Metric	**Clear sky (test data)**	**All sky**
Mean Absolute Error	1.56 K	3.11 K
Median Absolute Error	1.2 K	2.0 K
RMSE	2.04 K	4.97 K
Pixel Count	0.3M	5.73M
Error Distribution (Estimate – Observation)		

"warm tail"

Discussion

- Model accuracy is highest for clear-sky scenes, achieving moderate MAE with acceptable pattern similarity.
- Validation under all-sky conditions is challenging due to inaccuracies in the cloud mask.
- All-sky pixels tend to be cooler (lower insolation), which the model underestimates because training data are predominantly clear-sky.
- Ongoing experiments to address this include:
  - Incorporating more cloudy pixels and using a masked-loss function
  - Adding extra filtering steps to mitigate cloud-mask errors
- Despite strong SSIM (~91% on the test set), the model struggles to generalize scenemeans: training MSE is 3–4× lower than test MSE, even though pattern loss is similar.
  - We are testing a bias correction head aiming to reduce the MSE gap
- We can generate historical, current, and future daily LST estimates whenever low-resolution skin-temperature measurements or forecasts (e.g., ERA5) are available.

Thank you
Questions?