**LCAT
Data pipeline overview
Simon Kirby**

Contents

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- Climate data
  - Overview
  - Processing
  - Edge cases

Introduction

University
of Exeter

LCAT serves CHESS-SCAPE climate data for user-selected geographic regions.

How?
- Take a geographic region, overlay it on our climate data, take average of the climate data encompassed by region, and give data back to the user.

User flow
- User opens tool, selects a boundary type, such as UK Counties.
- They select one or more of these regions with mouse.
- Climate data for these regions are averaged, and shown to the user.
- The user can change settings, such as RCP, season, and future decade.
- Updated data is shown to the user based on these settings.

Technical objectives:
- Serve climate data of a single (or multiple) selected regions.
- Do so in a timely, responsive manner.

Introduction

University
of Exeter

Data requirements:

- 1. Open source boundary files, in a known geographic projection
- 2. Climate data, which is also properly geo-referenced

Boundary data

- Introduction
  - How does LCAT work?

- **Boundary data**
  - **Overview**
  - Processing

- Climate data
  - Overview
  - Processing
  - Edge cases

- These are the regions users can click on/select in the LCAT tool.

Boundary name & shortland name	Year	Source projection	File type	Source
UK counties and Unitary Authorities (UK Counties)	2023	EPSG 27700	Shapefile	ONS: geoportal.statistics.gov.uk
Local Authority Districts (LA Districts)	2023	EPSG 27700	Shapefile	ONS: geoportal.statistics.gov.uk
Parishes	2023	EPSG 27700	Shapefile	ONS: geoportal.statistics.gov.uk
Lower Layer Super Output Areas (LSOAs)	2021	EPSG 27700	Shapefile	ONS: geoportal.statistics.gov.uk
Middle Layer Super Output Areas (MSOAs)	2021	EPSG 27700	Shapefile	ONS: geoportal.statistics.gov.uk
Scotland Data Zones	2011	EPSG 27700	Shapefile	spatialdata.gov.scot
Northern Ireland Data Zones	2021	EPSG 29902	Shapefile	NISRA
Isle of Man	2015	EPSG 4326	Shapefile	Stanford

Boundary data

- We perform very little processing on these files.
- We reproject from source projection to EPSG 27700 to store in the database.
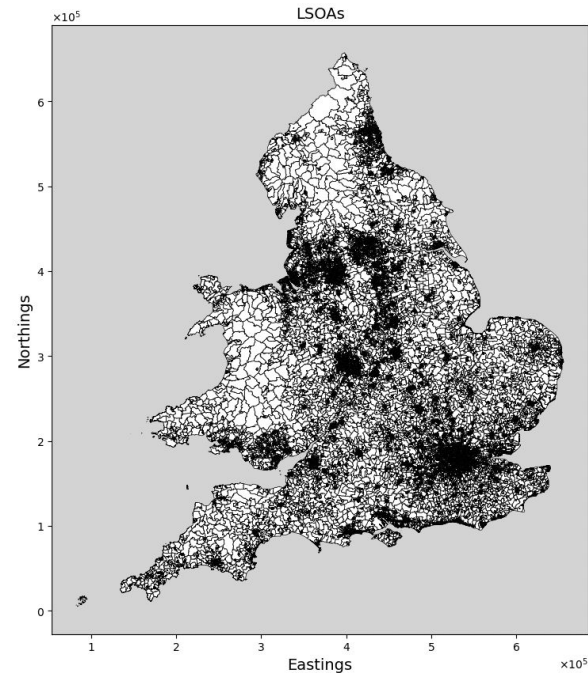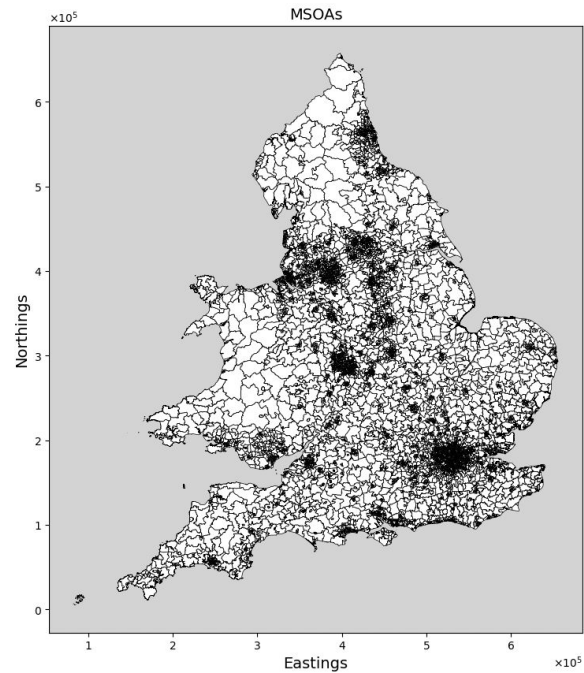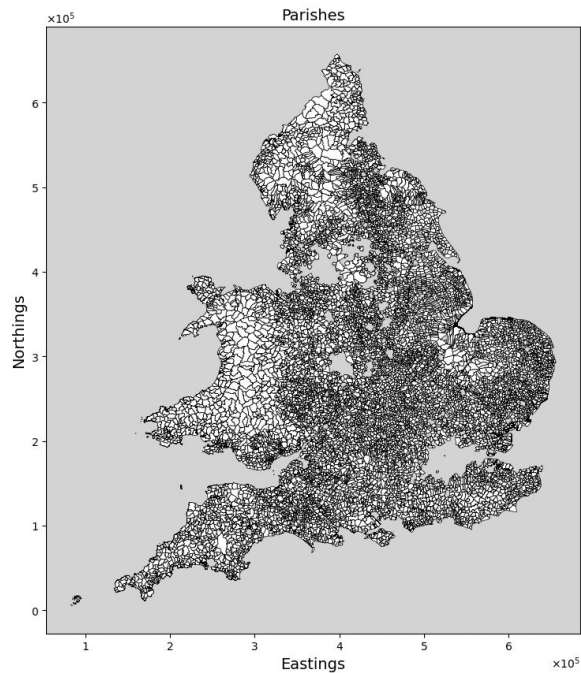- Finally we reproject to EPSG 4326 when visualising the data in the tool.

Boundary data – visualised

- UK Counties, LA Districts & Isle of Man
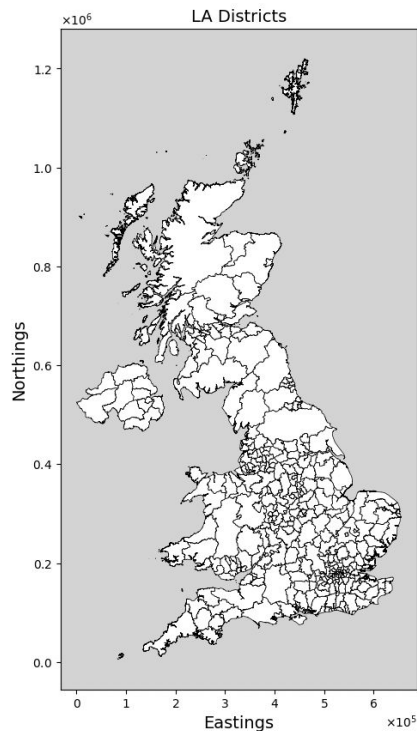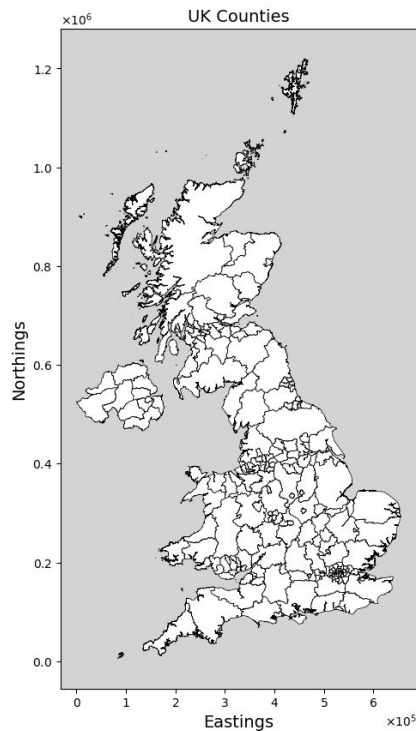
Boundary data - visualised

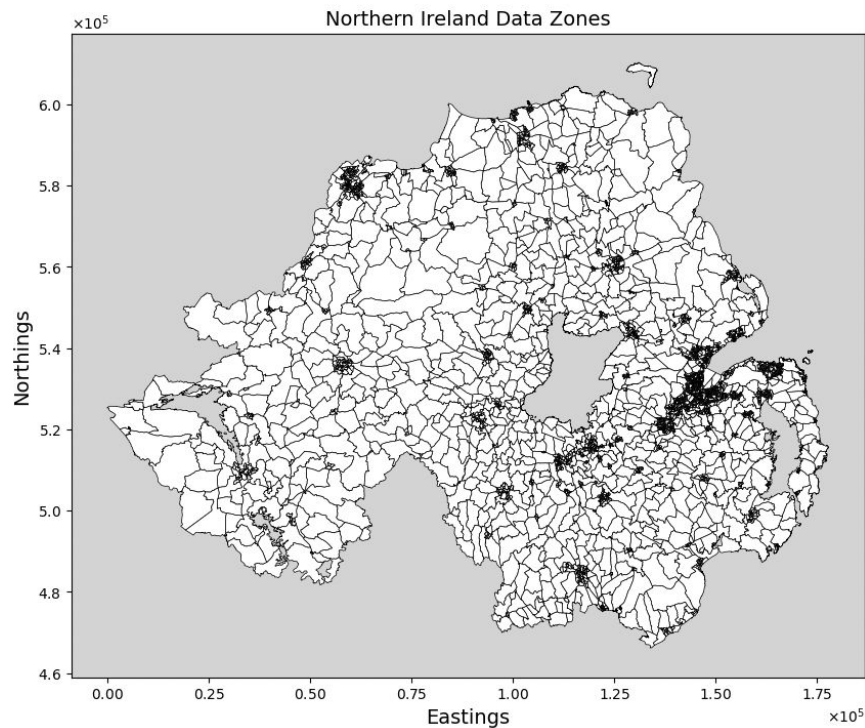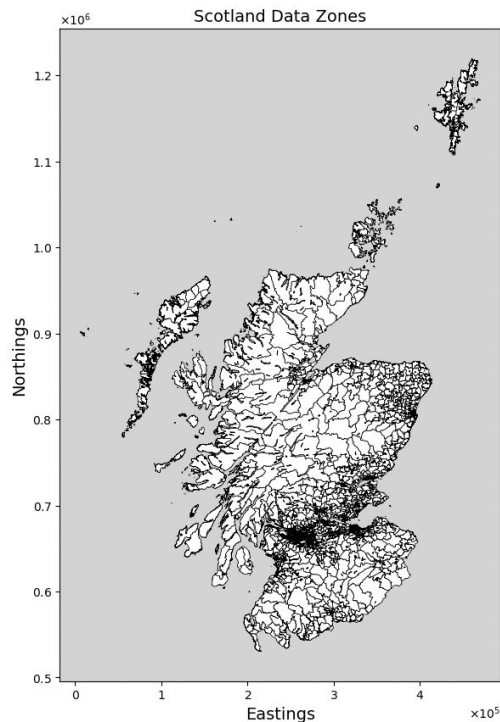- Parishes, MSOAs, LSOAs

Boundary data – visualised

- Scotland and Northern Ireland Data Zones

Climate data

- CHESS-SCAPE data contain records & predictions of gridded meteorological variables over the UK at 1km resolution (1980-2080), built from a number of other datasets.

- Data details <u>here</u> & <u>here</u>.

- The NetCDF files themselves can be found <u>here</u>.

- In LCAT we are using the bias and non-bias corrected data for:
  - Temperature (incl. min & max)
  - Precipitation
  - Wind speed
  - Downward shortwave radiation

- We are using ensemble member 1 only, at RCP6.0 and RCP8.5.

- Finally, we are using the annual, and seasonal means from the dataset, averaged by decade. This is around 10GB of the 11TB(!) dataset.

Climate data in LCAT

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
  - Processing
  - Edge cases

Variable name	Shorthand name	Source units	LCAT units	Climate variable it represents
Precipitation	pr	kg/m2/s	mm/day	Rainfall
Wind speed	sfcWind	m/s	m/s	Windiness
Downward shortwave radiation	rsds	W/m^2	W/m^2	Cloudiness (inverse)
Air temperature	tas	K	deg Celsius	Temperature
Max air temperature	tasmax	K	deg Celsius	Max temperature
Min air temperature	tasmin	K	deg Celsius	Min temperature

RCPs	Seasons	Decades	Variables	Number of data tables we need to store
6.0, 8.5 (2 RCPs)	Annual, summer, winter (3 seasons)	1980 – 2070 (10 decades)	6 variables	6 tables each with ~60 columns

Grid cell data in LCAT

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
  - Processing
  - Edge cases

- CHESS-SCAPE climate data are provided as variables on 1km grid cells, i.e. locations.
- The grid cells are properly geo-located, so will act as our link between the boundary data and the climate data.

	Grid cell count	Grid cell area	Source projection
Bias corrected	230573	1 km^2	BNG EPSG 27700
Non-bias corrected	14439	1 km^2	BNG EPSG 27700
Total cells in LCAT	245012	1 km^2	BNG EPSG 27700

CHESS-SCAPE decade means

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
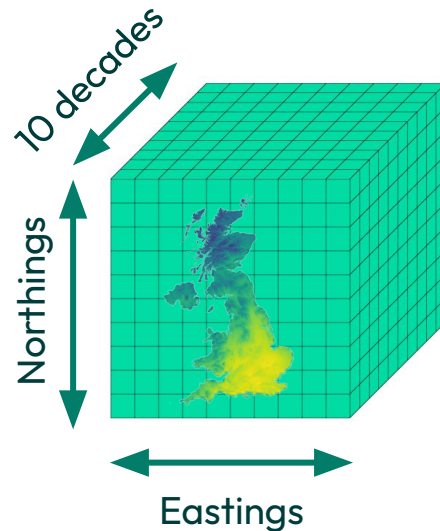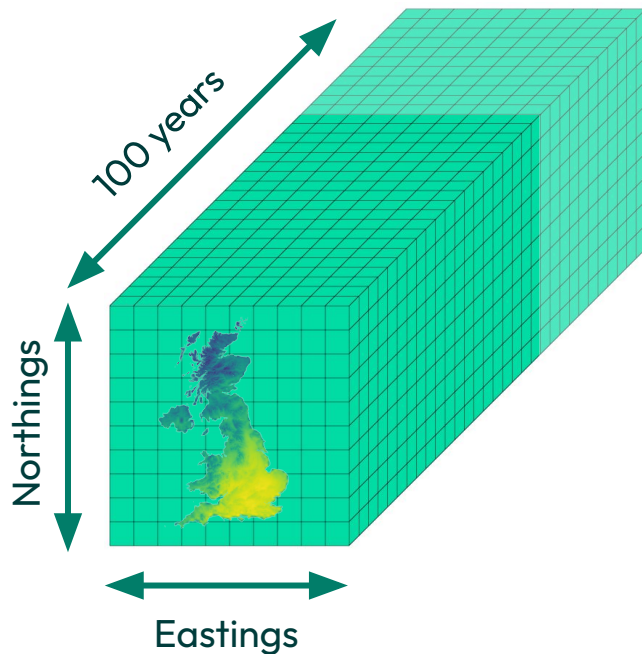 - Processing
  - Edge cases

CHESS-SCAPE is a large data set.
- For each of our 6 variables, 2 RCPs, & 3 seasons we get 100 years of predictions.
- First thing we do is downscale to decade means.

CHESS-SCAPE visualised

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
  - Processing
  - Edge cases

For each decade, we have found the mean values of each grid cell.

We have moved from 100 yearly data points to 10 yearly data points, i.e. we downscale in the time dimension.

This means we can now plot some climate data!
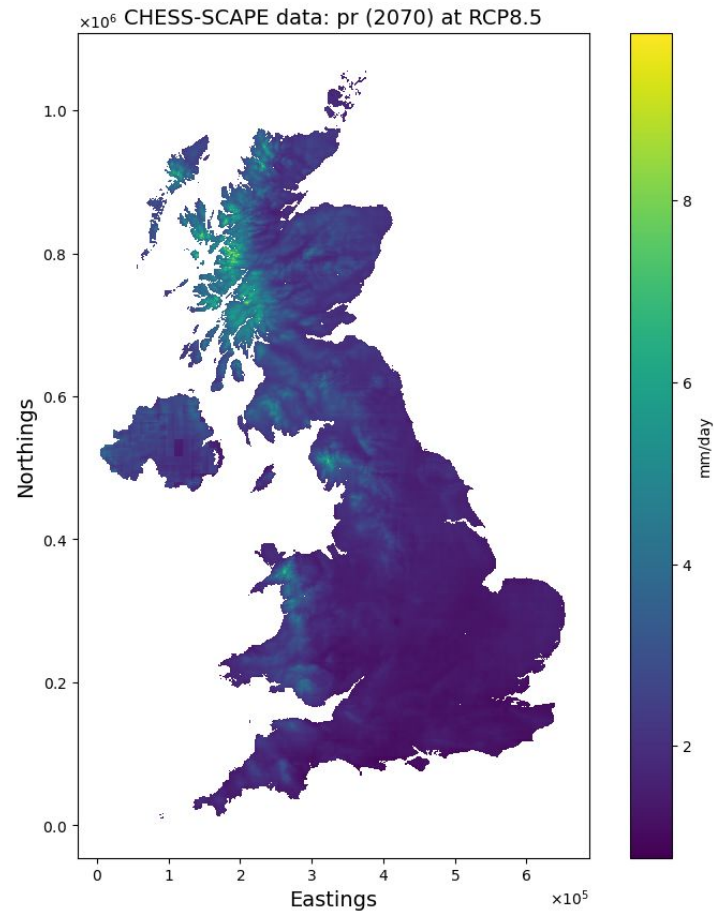
I am going to quickly show data for our 6 variables for:
- One season: **summer**
- One RCP: **RCP 8.5**
- One decade: **2070 – 2080**

CHESS-SCAPE data

- Mean prediction for summer season
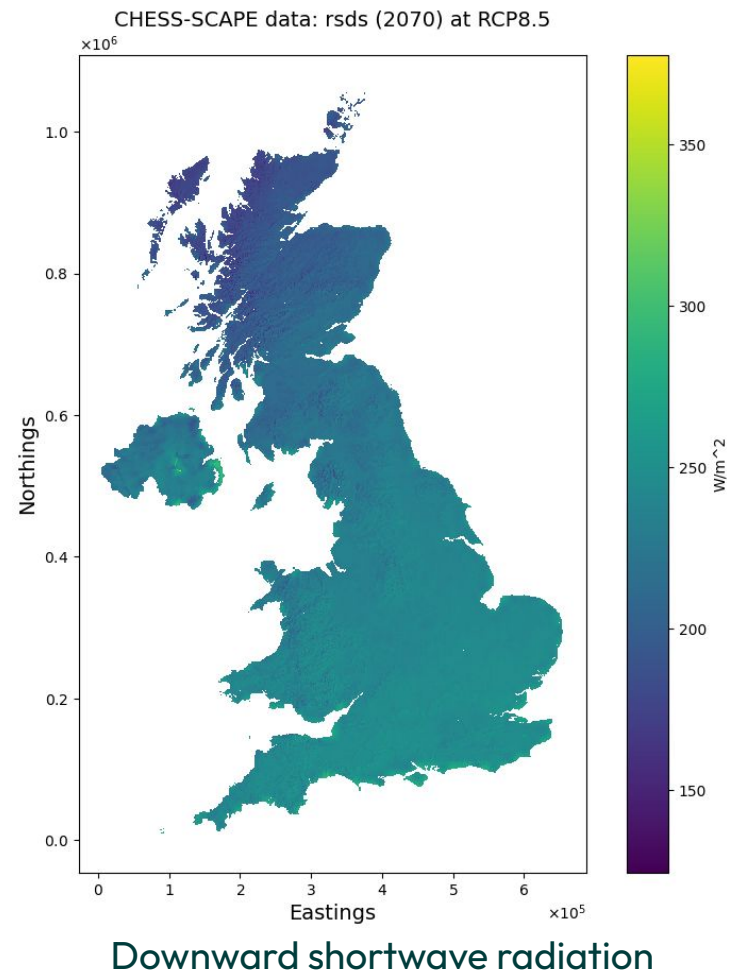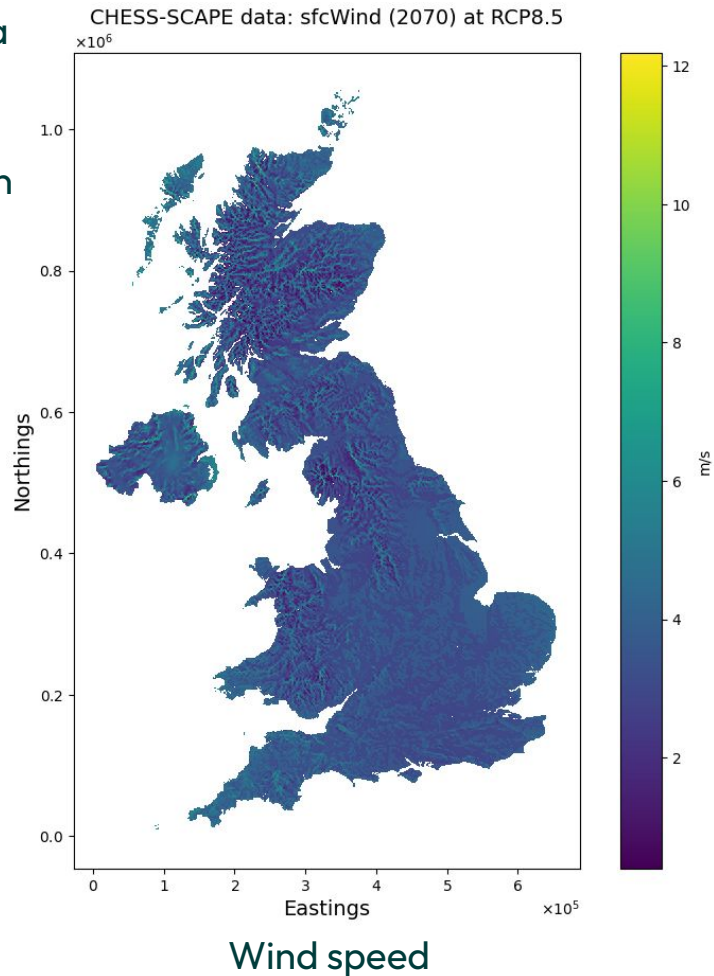- Across decade 2070 – 2080
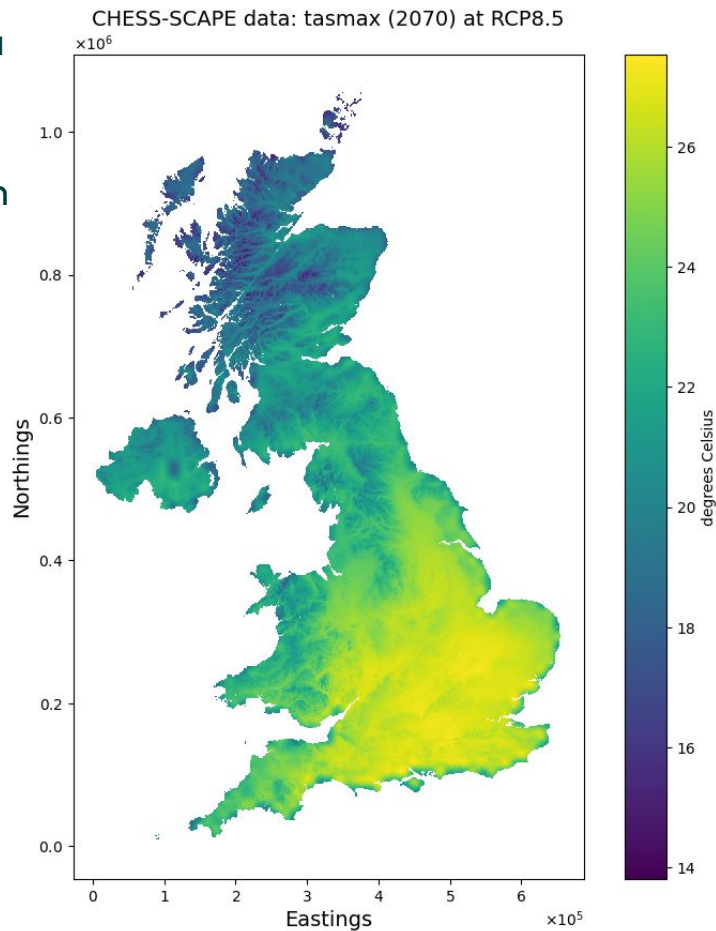- For RCP 8.5
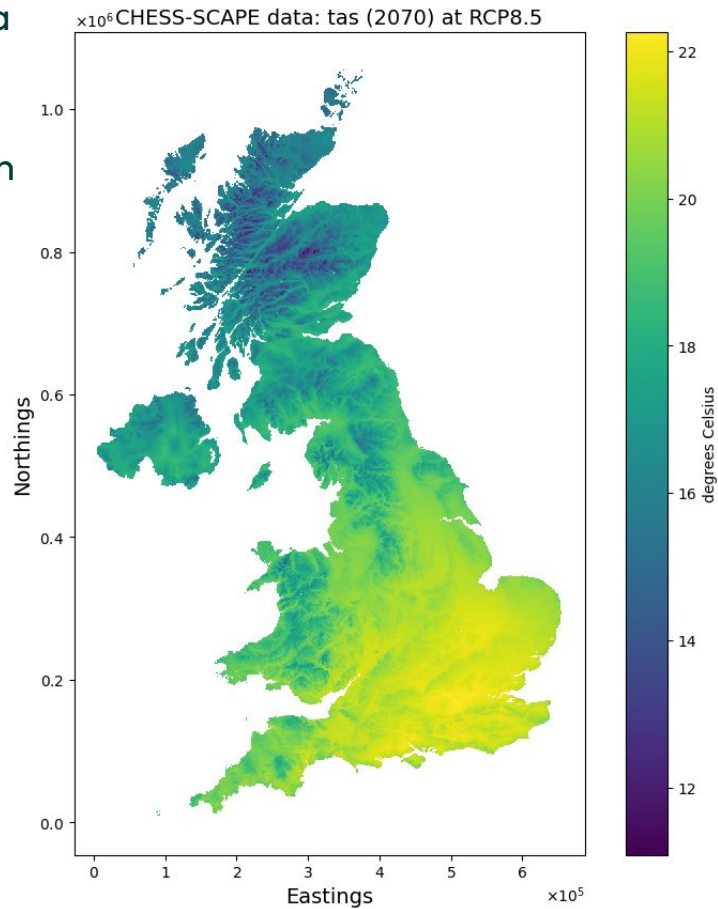
Temperature

Precipitation

CHESS–SCAPE data

- Mean prediction for summer season
- Across decade 2070 – 2080
- For RCP 8.5

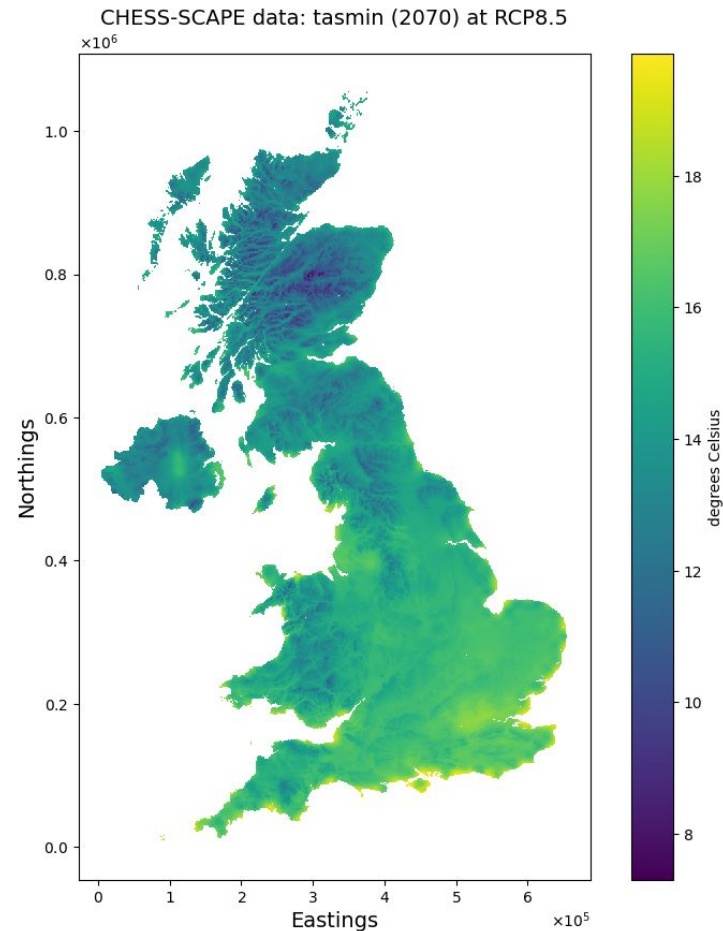CHESS-SCAPE data: sfcWind (2070) at RCP8.5

Wind speed

CHESS-SCAPE data: rsds (2070) at RCP8.5

Downward shortwave radiation

CHESS-SCAPE data

- Mean prediction for summer season
- Across decade 2070 – 2080
- For RCP 8.5

CHESS-SCAPE data: tasmax (2070) at RCP8.5

Max temperature

CHESS-SCAPE data: tasmin (2070) at RCP8.5
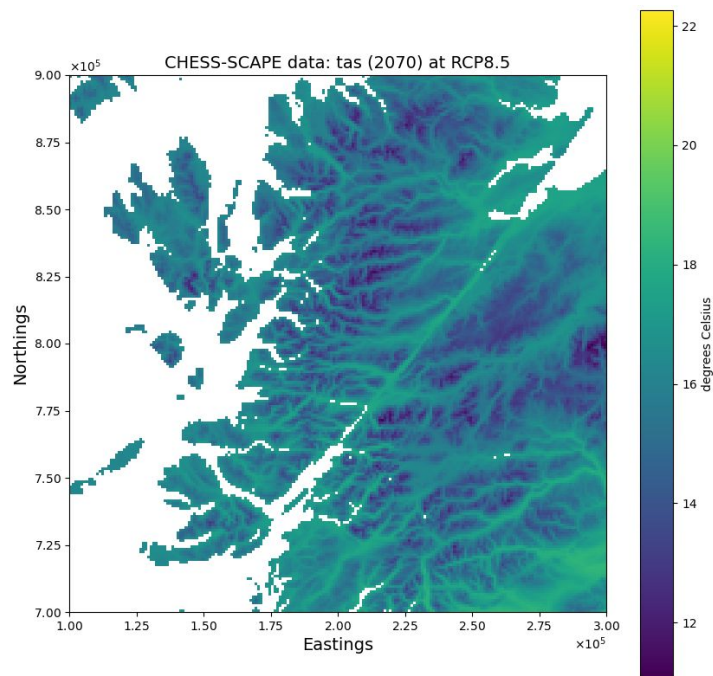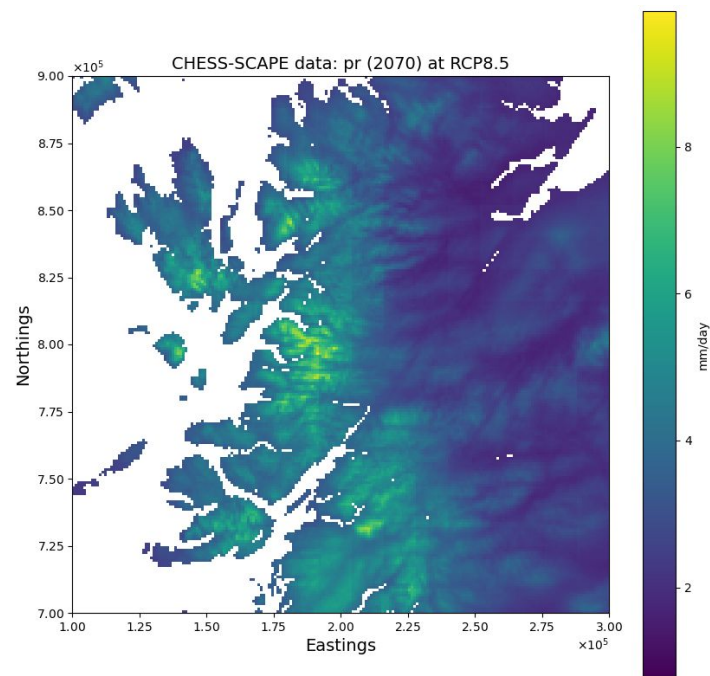
Min temperature

CHESS-SCAPE data

- Mean prediction for summer season
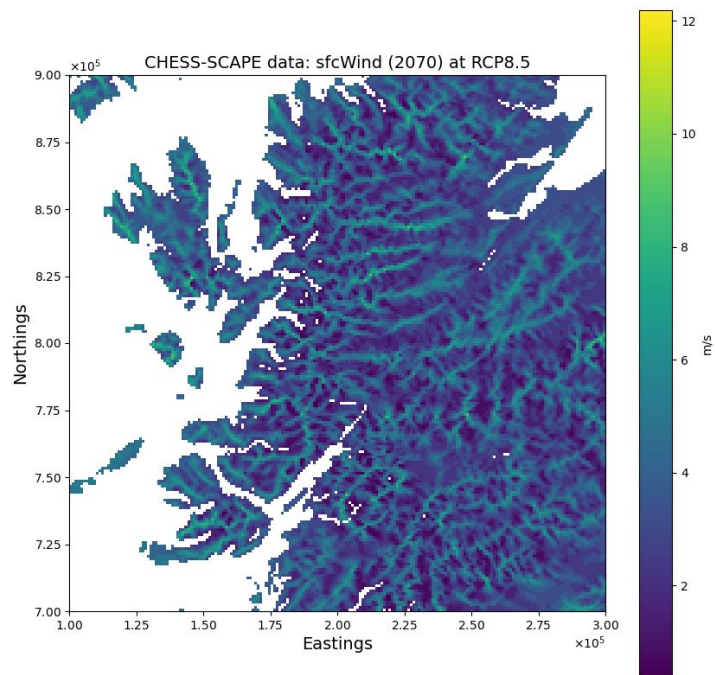- Across decade 2070 – 2080
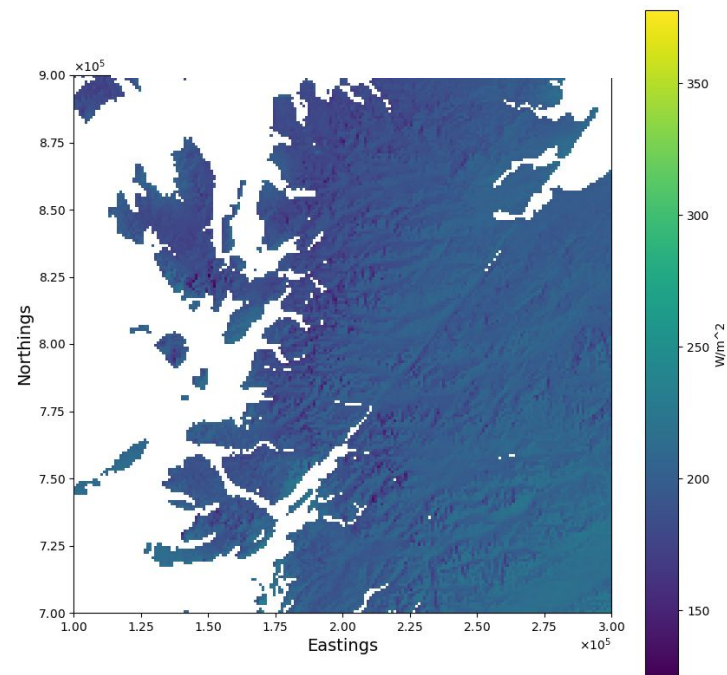- For RCP 8.5

Temperature
(zoomed in)

Precipitation
(zoomed in)

CHESS-SCAPE data

- Mean prediction for summer season
- Across decade 2070 – 2080
- For RCP 8.5

Wind speed
(zoomed in)

Downward shortwave radiation
(zoomed in)

Grid cell data in LCAT

- Introduction
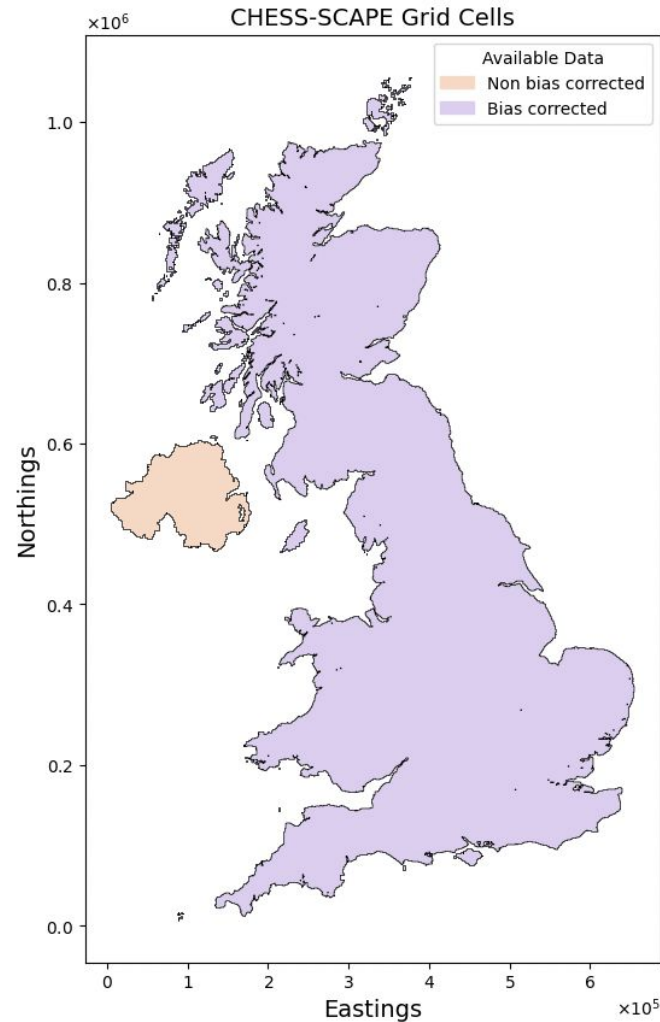 - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
  - Processing
  - Edge cases

On the right is a plot of the CHESS-SCAPE grid cells.

- These are the regions that we have climate data for.

- These have been coloured by availability (i.e. bias corrected vs non-bias corrected vs no data).

- In LCAT, we try to serve bias corrected data as a priority.
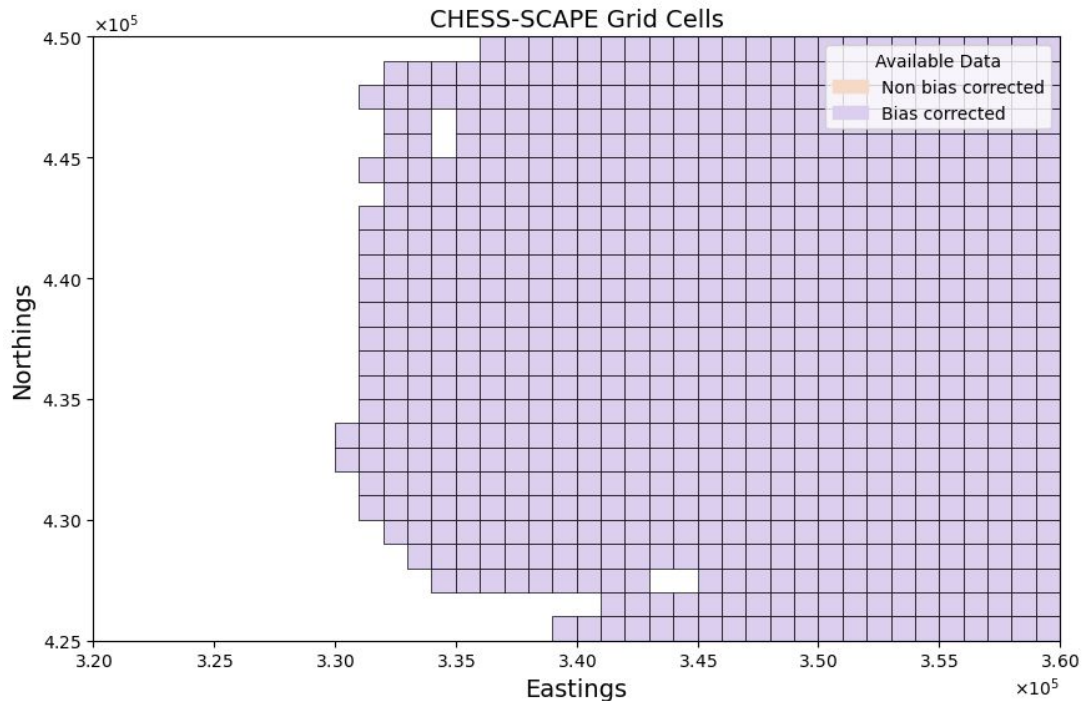

CHESS-SCAPE Grid Cells

Grid cell data in LCAT

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - **Overview**
  - Processing
  - Edge cases

- These are the same grid cells but zoomed in.

Climate data processing

What we want:
- User clicks a region, and selects i.e. RCP6.0, summer, 2040.
- The climate data for the region is calculated from CHESS-SCAPE.
- This data is presented to the user.

Two methods – depending on size of regions.
- Note that these perform the same computations and serve identical results for a single region.
- Method 1: Uses cache tables.
- Method 2: Selects and averages climate data on the fly.

Climate data processing

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - **Processing**
  - Edge cases

Both methods:
- We precompute overlapping grid cells of all boundary regions, get the climate data for these grid cells, and average it.

Method 1: cache method (UK Counties, LA Districts, IoM)
- For large regions that overlap many grid cells, we take the overlapping grid cells, select the climate data for these cells, average it, and cache this.
- When a user selects a single region, we serve this cached data.
- When a user selects multiple regions, we serve the average of the cached data. Hence, if the regions overlap, some cells are "counted" twice.
- For large regions that overlap many cells, this effect is small, compared to the cost of computing overlapping grid cells for large regions on the fly.

Method 2: cell method (all other boundaries, i.e. small regions)
- We take the overlapping grid cells, select and average the climate data for these cells on the fly. The server performs the computations, and doesn't rely on the cached climate tables in the database. This means we don't count cells twice, but at computational cost.

Climate data processing

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - **Processing**
  - Edge cases

However, for both methods, the following process is essentially the same.

Climate data processing

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - **Processing**
  - Edge cases

What we want:
- User clicks a region, and selects i.e. RCP6.0, summer, 2040.
- The climate data for the region is calculated from CHESS-SCAPE.
- This data is presented to the user.

Steps (for one boundary file, one RCP and one season):
- 1. Average the 6 climate data variables by decade. Then store these in the database indexed by grid cell.

- 2. For each region in a boundary, find the overlapping CHESS grid cells. Then store these overlaps in the database, indexed by region.

- 3. For each region, get the overlapping grid cells, and the relevant climate data for each overlapping grid cell. Then average the climate data for each cell, and store in the database.

Climate data processing

Given a region, how do we find the climate prediction?
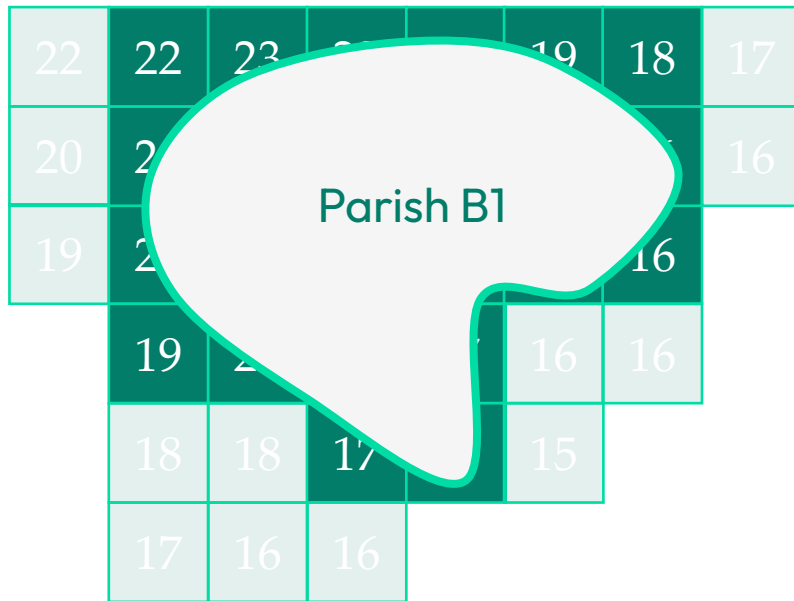
Parish B1

Climate data processing

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - **Processing**
  - Edge cases

Step 1: average and store our climate data cells.

22	22	23	22	21	19	18	17
20	21	23	20	19	17	16	16
19	21	21	19	18	17	16	
	19	20	18	17	16	16	
	18	18	17	16	15		
	17	16	16				

- Here we have some coastal grid cells with average temperature predictions stored in them.

Climate data processing

Step 2: find overlapping climate grid cells with a region

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - **Processing**
  - Edge cases

Climate data processing

Step 3: average these cells to find the climate prediction for the region.

22	23	22	21	19	18
21	23	20	19	17	16
21	21	19	18	17	16
19	20	18	17		
		17	16		

Parish B1: average = 19.17 deg

To perform this process, we need the boundary regions, the grid cell squares, and the climate data (for each variable, decade, RCP etc) in the database.

Regions & overlapping cells
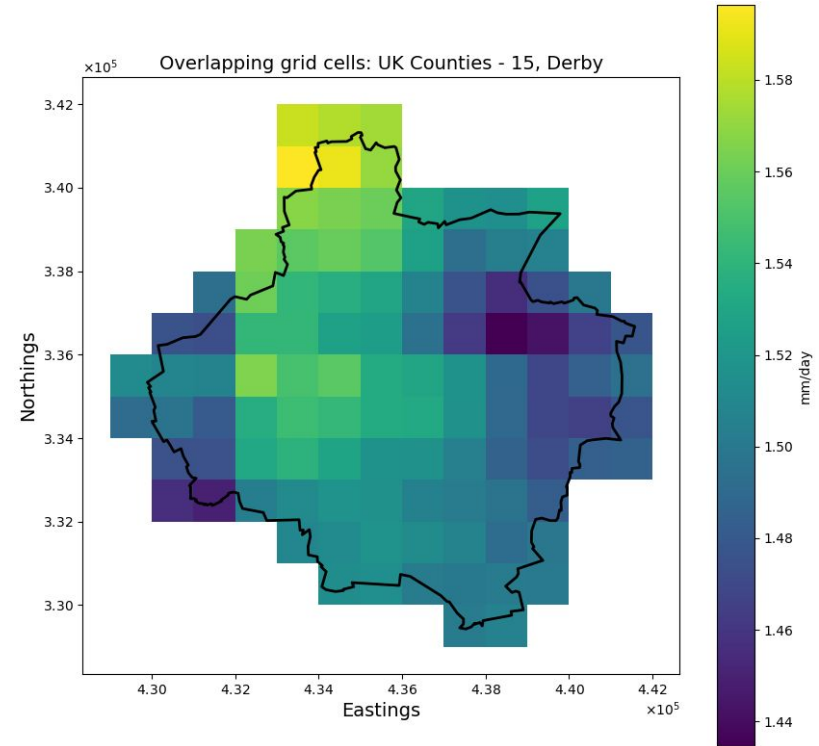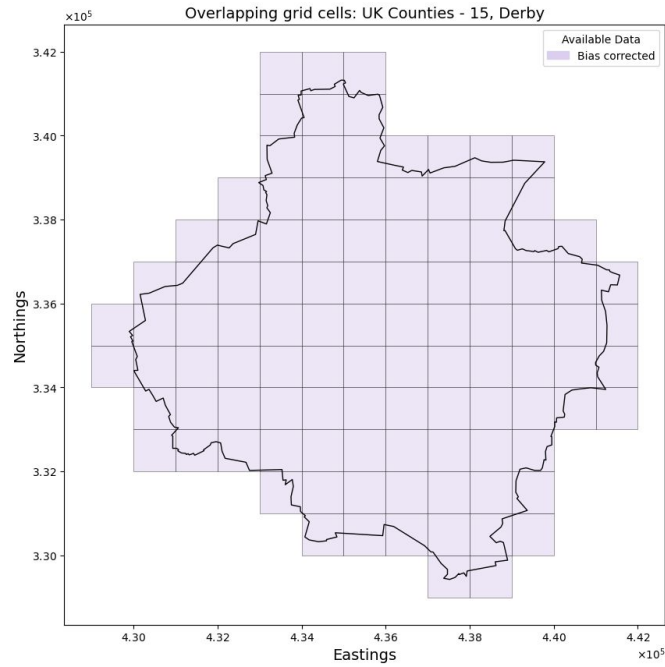
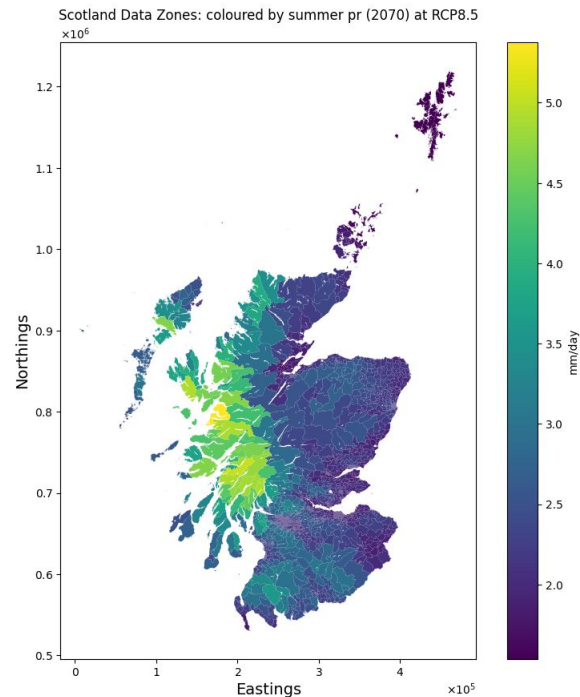• Mean prediction for Derby precipitation: summer season, across decade 2070 – 2080, for RCP 8.5

So what are we serving in LCAT?
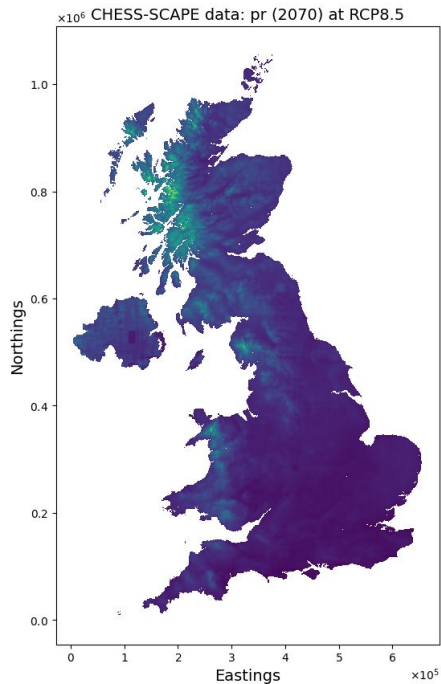
- CHESS–SCAPE precipitation data: summer season, across decade 2070 – 2080, for RCP 8.5
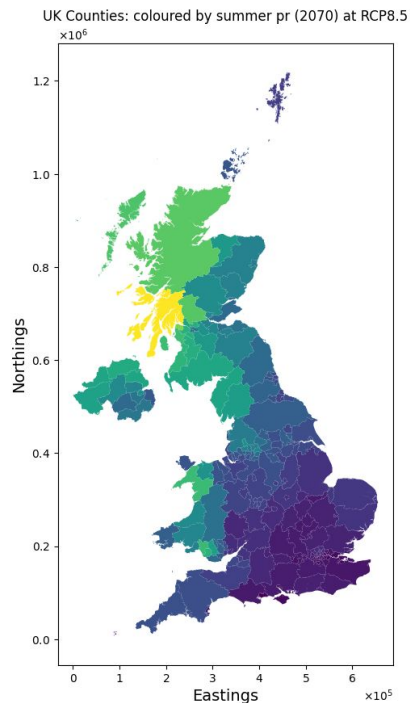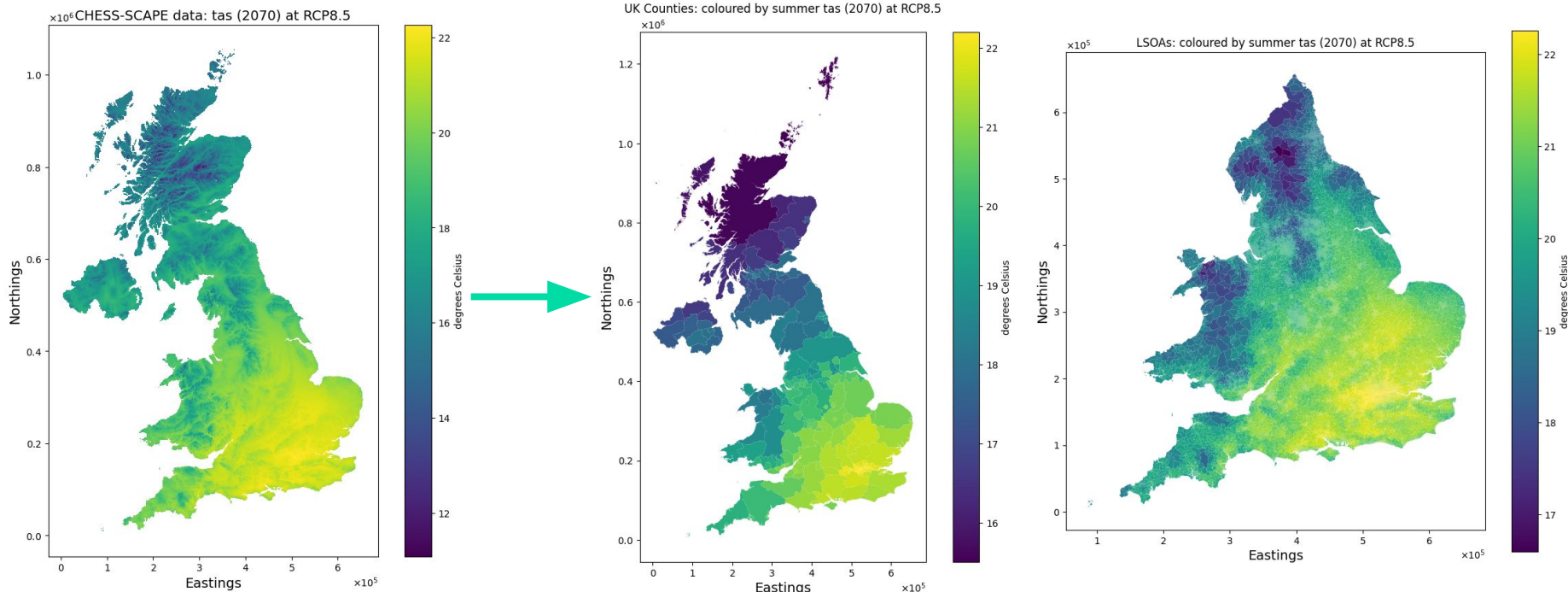
- Mean precipitation prediction for individual counties: summer season, across decade 2070 – 2080, for RCP 8.5

- Or, for Scotland data zones

So what are we serving in LCAT?

- CHESS–SCAPE temperature data: summer season, across decade 2070 – 2080, for RCP 8.5

- Mean temperature prediction for individual counties: summer season, across decade 2070 – 2080, for RCP 8.5
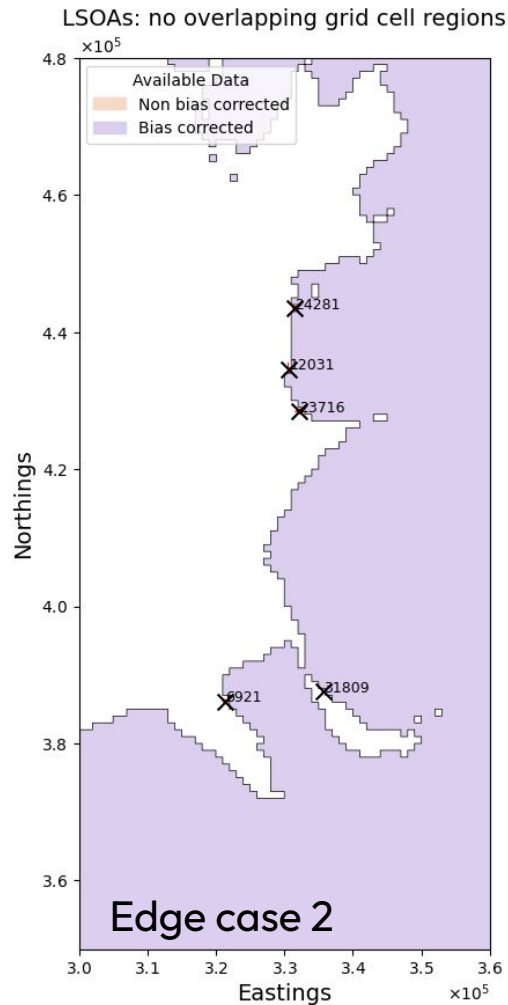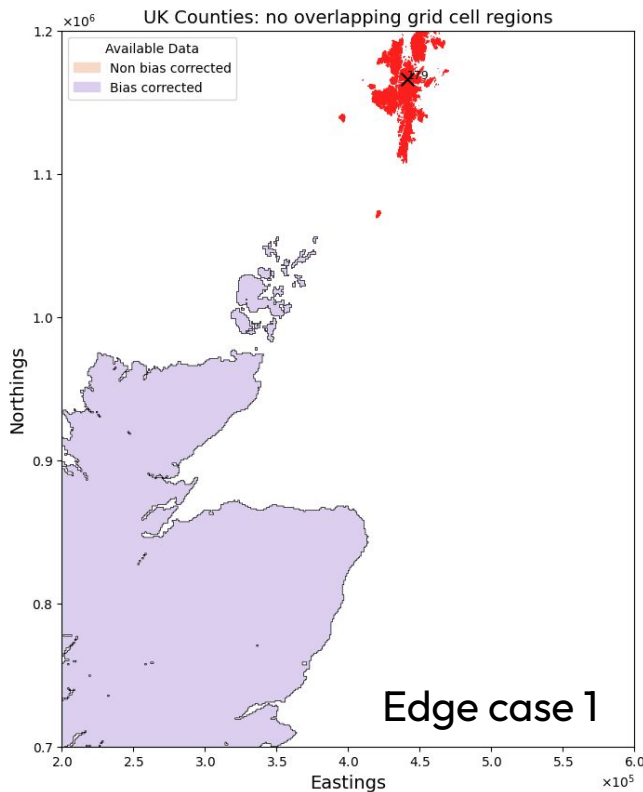
- Or, for LSOAs

Edge cases

We have a number of edge cases to handle. These are cells where there are no overlapping grid cells.

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - Processing
  - **Edge cases**

Edge case 1

Edge case 2

Edge cases

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
  - Processing

- **Climate data**
  - Overview
  - Processing
  - **Edge cases**

Edge case 1:
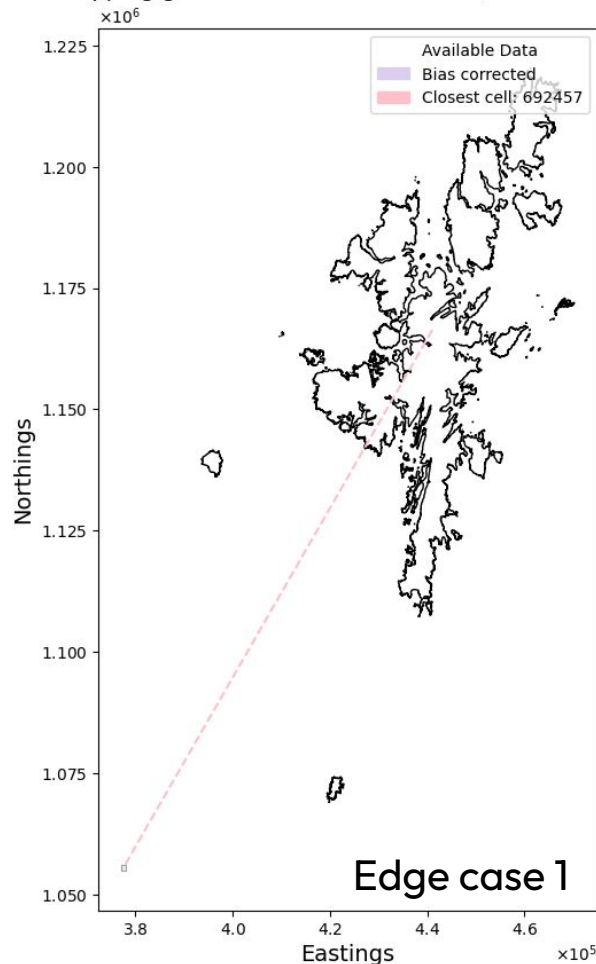- Region is large and has no overlapping grid cells at all.

Solution: find closest bias corrected cell and use the data from that.

- Process:
  - We get the bounding envelope of the shape, iteratively expand it around the centroid until we hit some grid cells.
  - We then perform a closest distance calculation for all these cells, and store the closest cell.
  - We serve the closest cells climate data for the region.

Shetland Islands only.

Overlapping grid cells: UK Counties - 179, Shetland Islands

Edge case 1

Edge cases

- Introduction
  - How does LCAT work?

- Boundary data
  - Overview
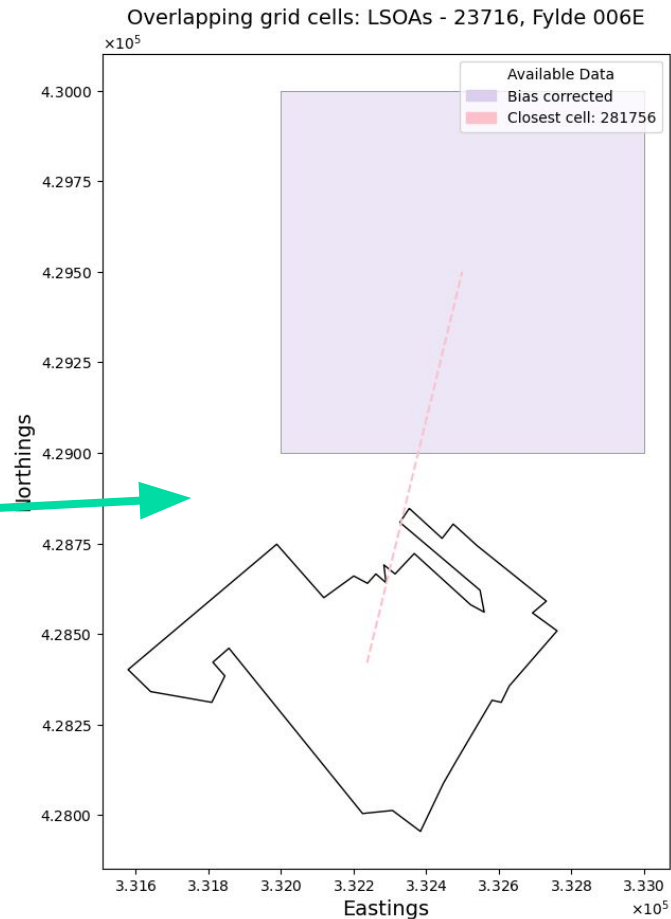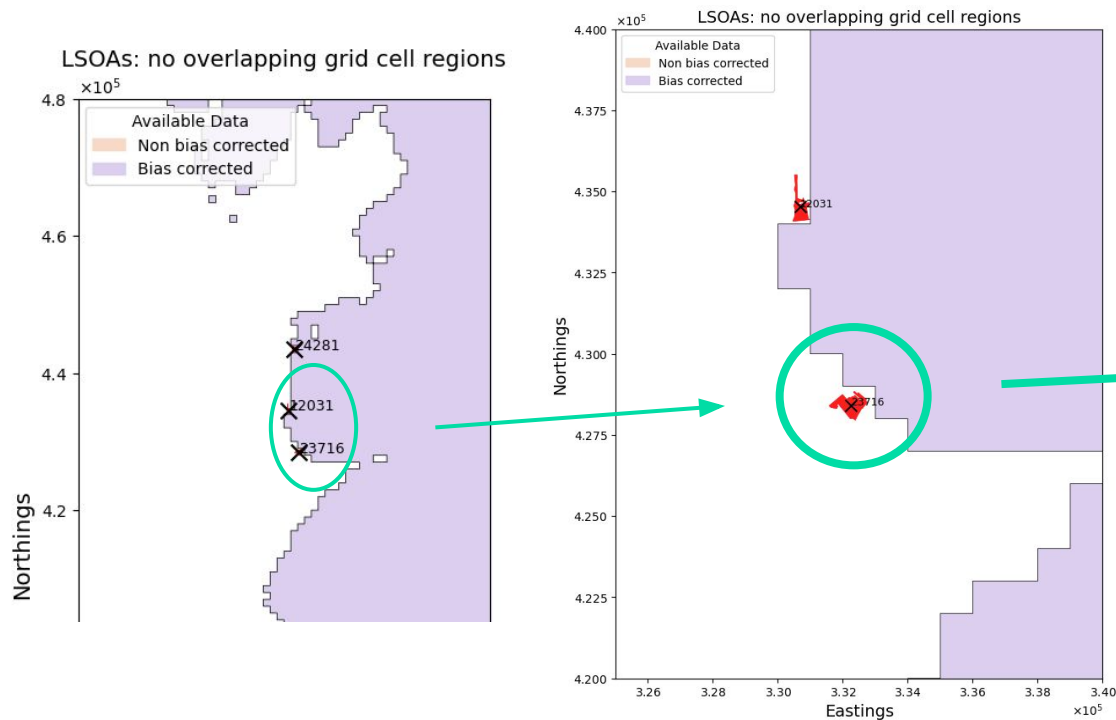 - Processing

- **Climate data**
  - Overview
 - Processing
  - **Edge cases**

Edge case 2:
- Region is small and (as before) has no overlapping cells.
- Two cases:
  - a. Region is smaller than a single grid cell, and an intersection calculation will not return a hit.
  - b. Region is close to a grid cell, but doesn't quite overlap as a result of the CHESS-SCAPE rasterisation they have performed when creating the data set.

Solution: in both cases, we can use the same method as before, i.e. make the region envelope iteratively larger until we hit cells. We then take the closest cells ID as being overlapping, and serve this cells climate data for the region.

Edge case 2: LSOA example – Fylde 006E
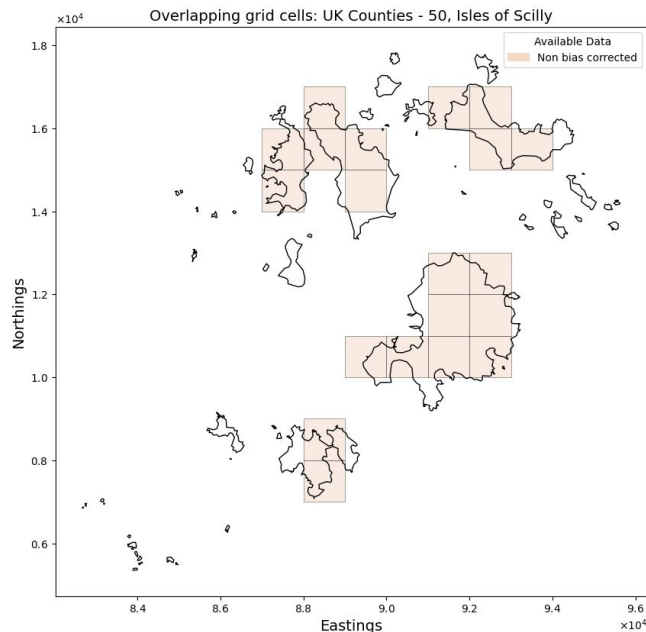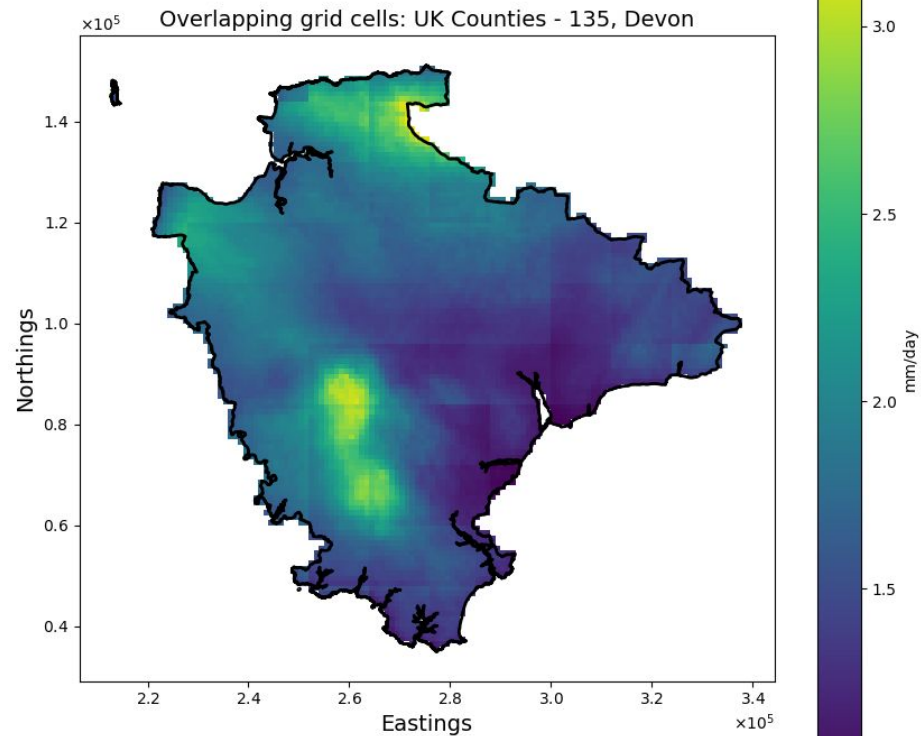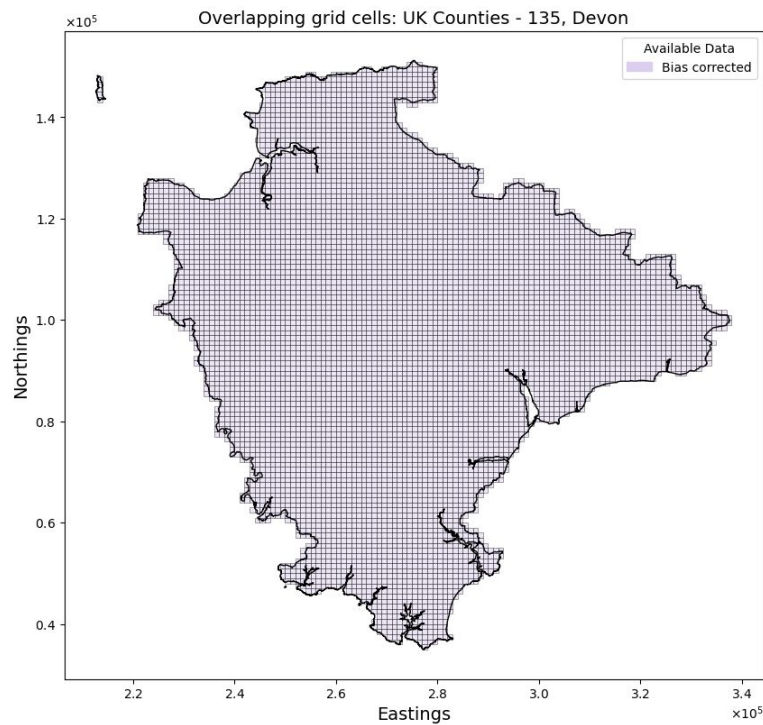
Edge cases

Edge case 3:
- Region has no overlapping bias corrected data cells, or "nearby" bias corrected cells.

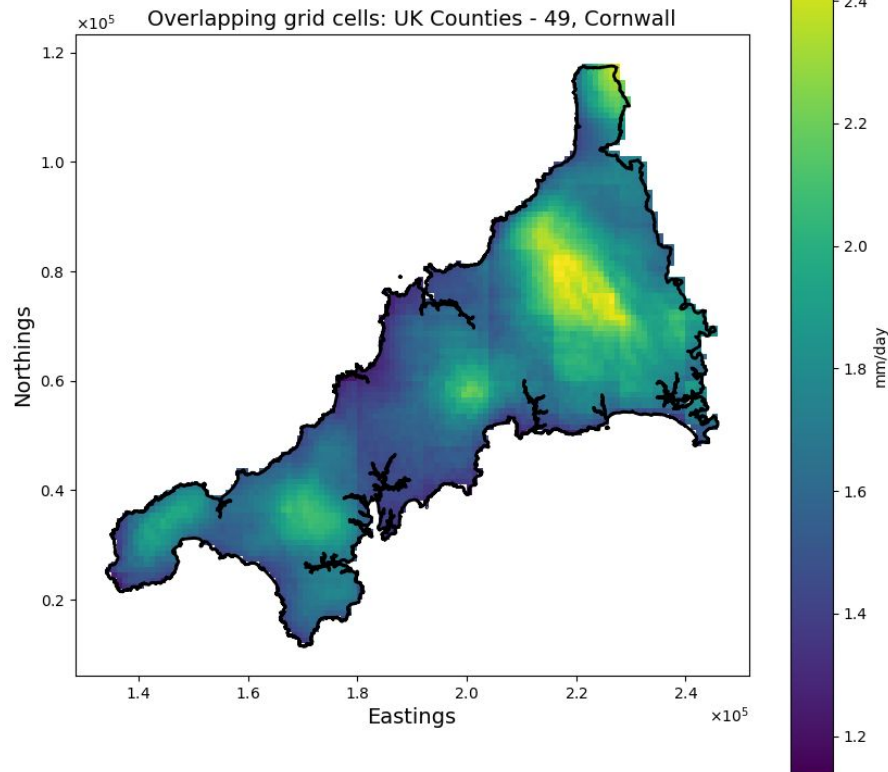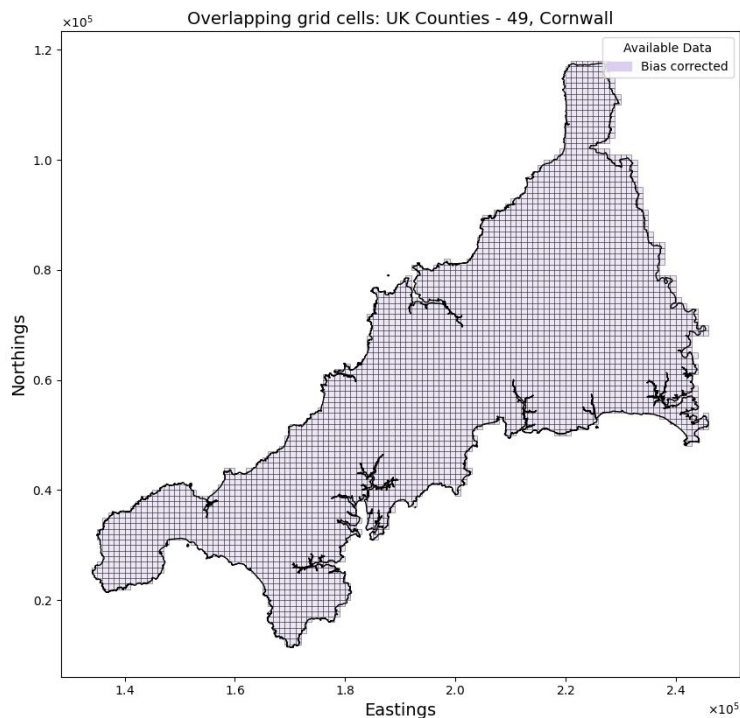**Solution**: look for overlapping non-bias corrected cells instead.
- All Northern Ireland data zones, and the Isles of Scilly

Regions & overlapping cells: Devon
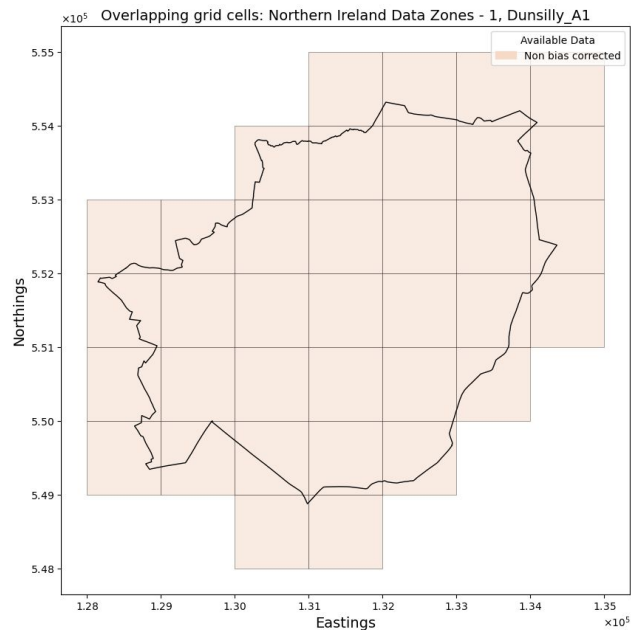
Overlapping grid cells: UK Counties - 135, Devon

Regions & overlapping cells: Cornwall

Changelog

- Version 1:
  - First version

- Version 2:
  - Add details about cache vs cell methods
  - Remove next steps slide
  - Update links to boundary datasets