Using machine learning to improve resolution and bias in urban temperature projections

Risa Ueno[1,2], Dr. Scott Hosking[1], Dr. Alex Archibald[2], Dr. Emily Shuckburgh[3], Dr. Richard Turner[2]

[1]British Antarctic Survey, [2]University of Cambridge, [3]Cambridge Zero

British Antarctic Survey NATURAL ENVIRONMENT RESEARCH COUNCIL

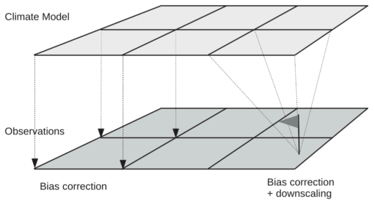
UNIVERSITY OF CAMBRIDGE

The Alan Turing Institute

Motivation

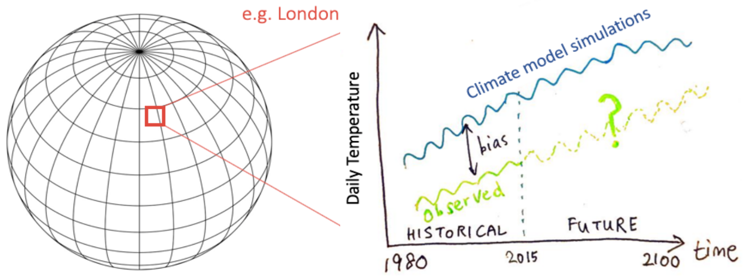
- Extreme heat is expected to become more frequent and severe in the future, and we need accurate projections to assess climate risk at specific locations.
- Global climate models are based on differential equations and our knowledge of physical climate processes. They can simulate the future, but are very coarse gridded (~200km^2) and biased.
- Our work builds on *statistical downscaling*, which uses observation data together with the simulations to obtain a more reliable projection at finer scales.

What is statistical downscaling?



This is a post-processing method used on climate model simulations to (1) correct bias and (2) to achieve projections at a higher resolution.

Our work is based on a more specific technique called *model output statistics,* which finds a mapping between the long-term distributions from model to observation, which is then applied to future simulations.



Proposed technique

1. Detrend the timeseries
(We preserve the simulated long-term trend as a credible climate change signal)

2. Train transfer function using raw values and corresponding quantiles in the historical period

$$T_{q(\text{observed})} = f\,(T_{q(\text{model})}, \text{quantile, day of year})$$

For a grid-point location, or for a gridded window:

$$T_{q(\text{observed})} = f\,(T_{q(\text{model})}, \text{quantile, day of year, longitude, latitude})$$

This is done using **Gaussian process regression** to quantify the output uncertainty

3. Apply transfer function to the distribution of future climate simulations

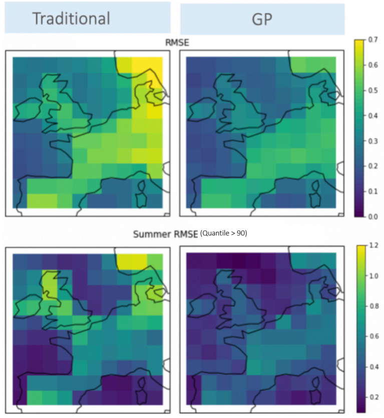
Compared to traditional mapping:

- ✔ No need to assume a specific distribution
- ✔ Sharing of information across neighbouring locations
- ✔ Uncertainty quantification
- ✔ Better generalisation to unseen periods
- X Computational cost
- X Traditional assumptions still present

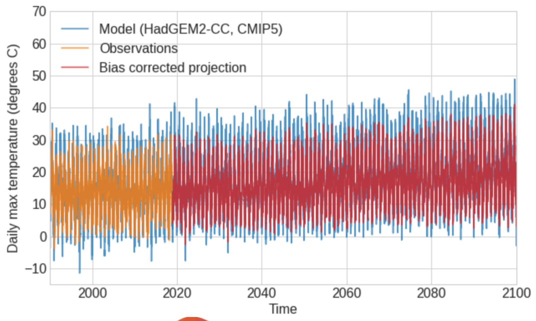
Experiment summary

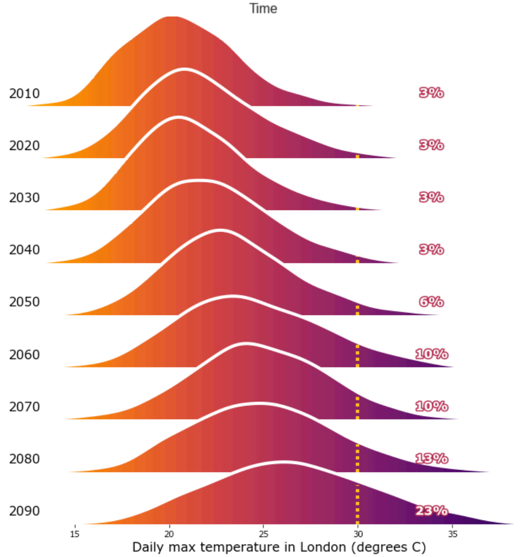
K-fold experiment on 200 grid locations, bias correcting 3 CMIP5 models with ERA5 as observation data:

- +28% RMSE improvement compared to traditional quantile mapping
- +32% for summer extremes (quantiles > 90)



Demo: daily temperature projection in London





Further work

- Multivariate bias correction, e.g. $f\,(T_{q(\text{model})}, \text{quantile}_T, \text{day of year}, H_{q(\text{model})}, \text{quantile}_H)$
- Selecting an optimal set of climate models to downscale