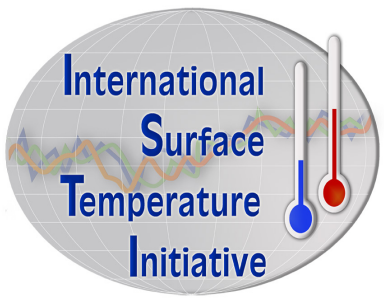
Benchmarking the Performance of Daily Temperature Homogenisation Algorithms

Rachel Warren

rw307@exeter.ac.uk

Supervised by Professor Trevor Bailey, Professor Ian Jolliffe and Dr Kate Willett



Overview

- Introduction and definitions
- Modelling
- Using the results
- Future work
- Summary

Why do we need benchmarks and what are they?

Temperature data contain inhomogeneities, but quantifying and correcting for these in the real world is very difficult where the underlying truth is not known.

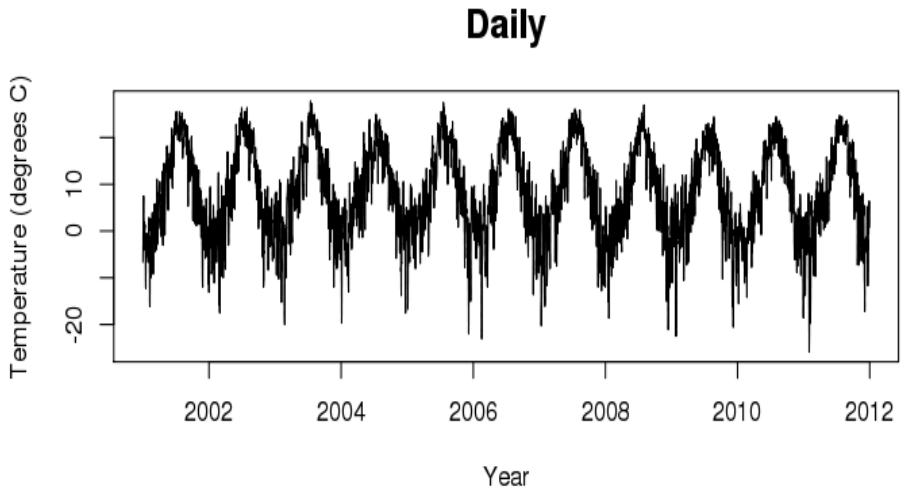
Benchmarking is a method by which homogenisation algorithms can be assessed without needing to know the true state of the real world.

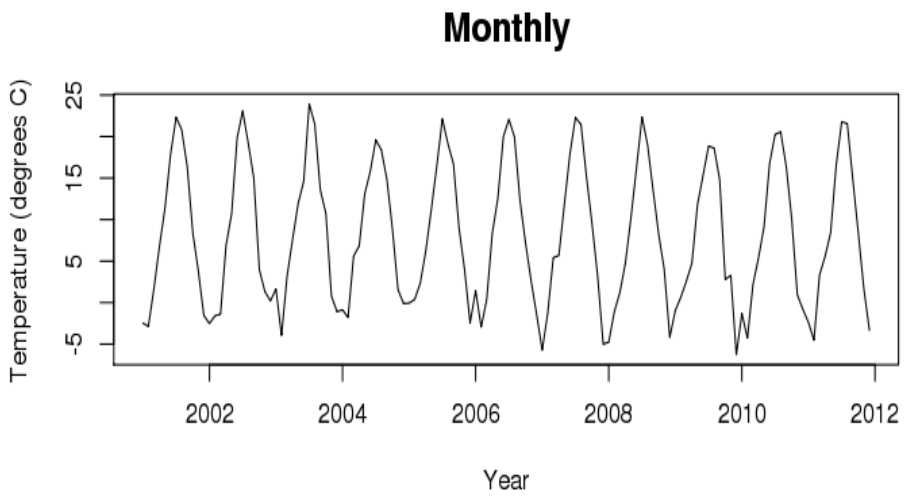
1. Clean worlds are produced free from inhomogeneities
2. Error structures that mimic real-world inhomogeneities are added on to these worlds to create error worlds
3. These error worlds will be released to the scientific community to allow homogenisation algorithms to be run on them
4. The returned series will be compared with the original clean world (the benchmark) using pre-defined validation measures.
5. The findings of this assessment will be released to allow the further development of homogenisation algorithms and assist the quantification of remaining uncertainty in the data.

Daily Data

Daily temperature data pose new problems compared to monthly or annual data:

- Variability
- Quantity
- Correctly adjusting for inhomogeneities even after they've been found



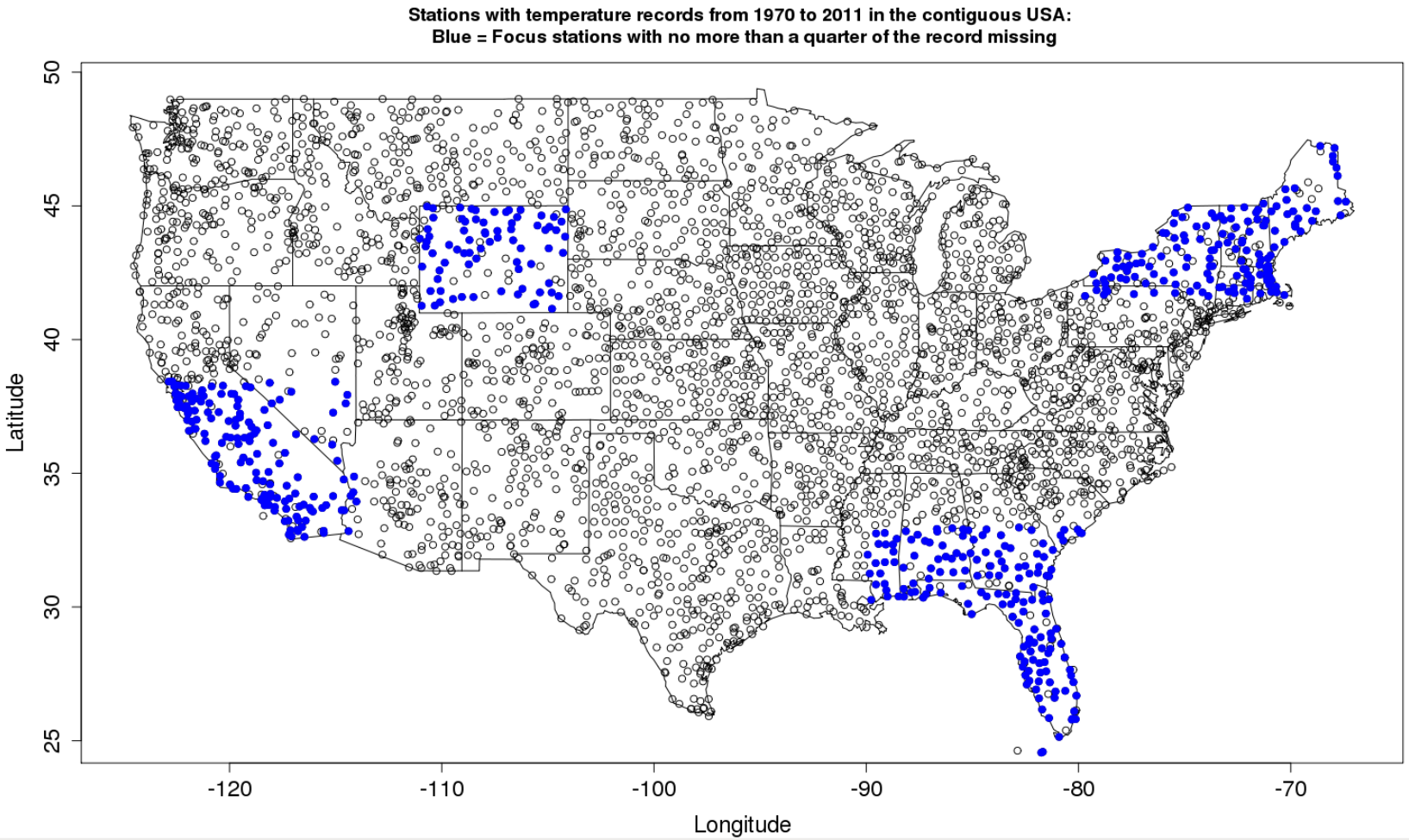


Data Sources

- Global Historical Climatology Network Daily (GHCND) Database
- NOAA's 20th Century Re-Analysis
- Australian Bureau of Meteorology

Data Coverage

- Focusing on 4 regions of North America
- From 1970 to 2011



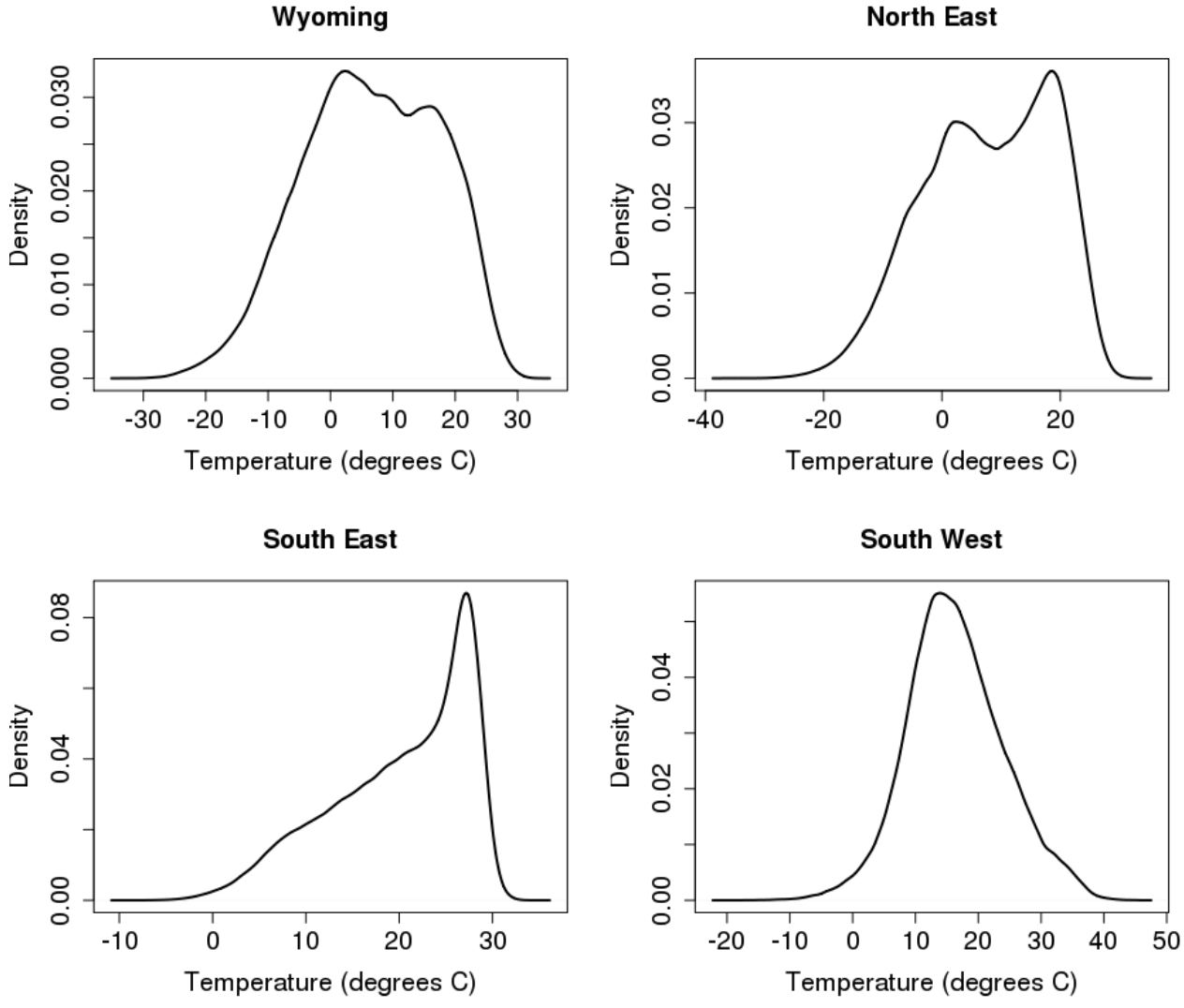
Modelling

- No 'how to' guide for this project
- Using Generalised Additive Models (GAMs)
- These models are flexible:
  - A choice of how to include explanatory variables
  - A choice of statistical distributions for the variability
  - More features can be modelled instead of filtered out

The Model

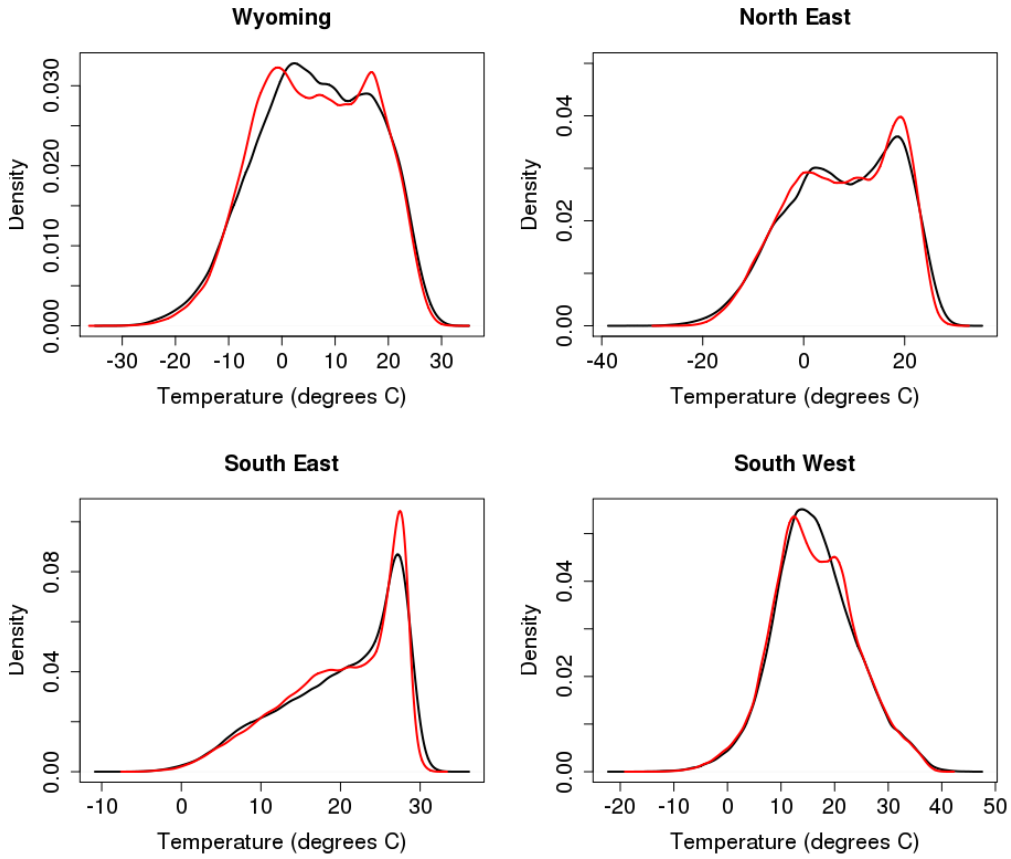
- Mean temperature is the response - GHCND
- Modelled using:
  - Location, Altitude - GHCND
  - Sea Level Pressure, Eastward and Northward wind components, Precipitation, Precipitable Water Content, Temperature, Downward Solar Radiation Flux - 20CR
  - Southern Oscillation Index - Australian Bureau of Meteorology
  - Time and Day of the Year
- Could keep adding more variables

The Model Performance



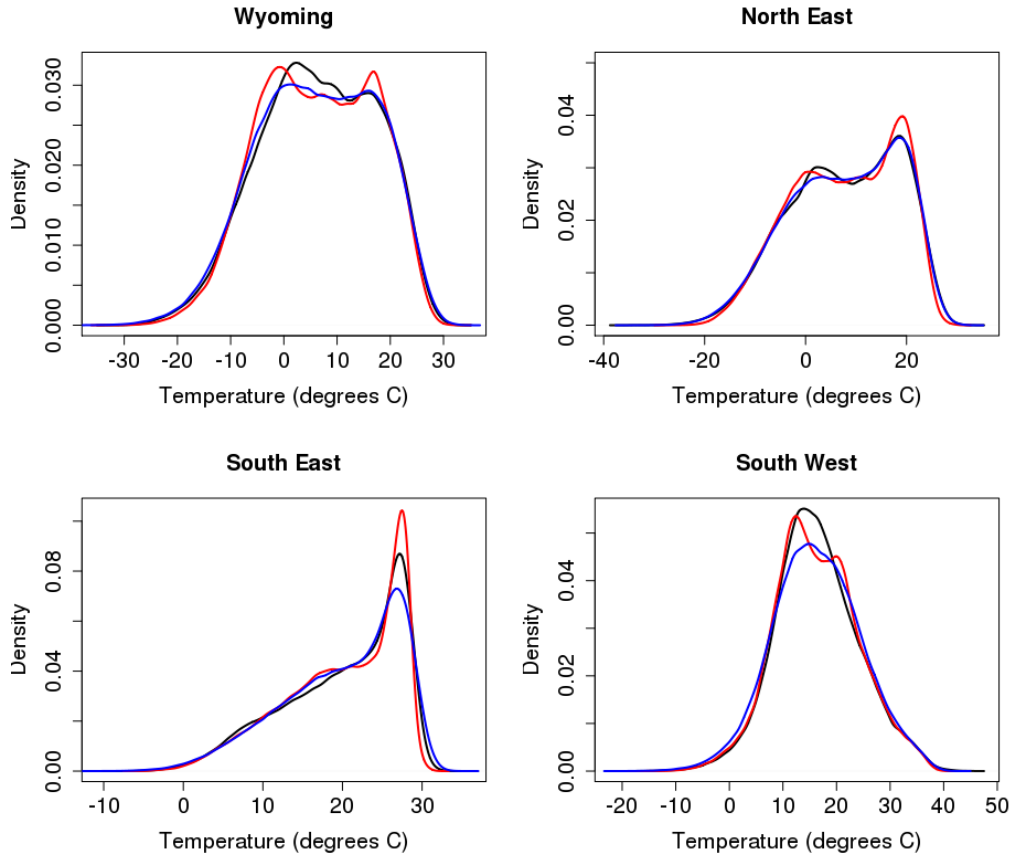
- Density plots showing temperature distributions in the four focus regions

The Model Performance



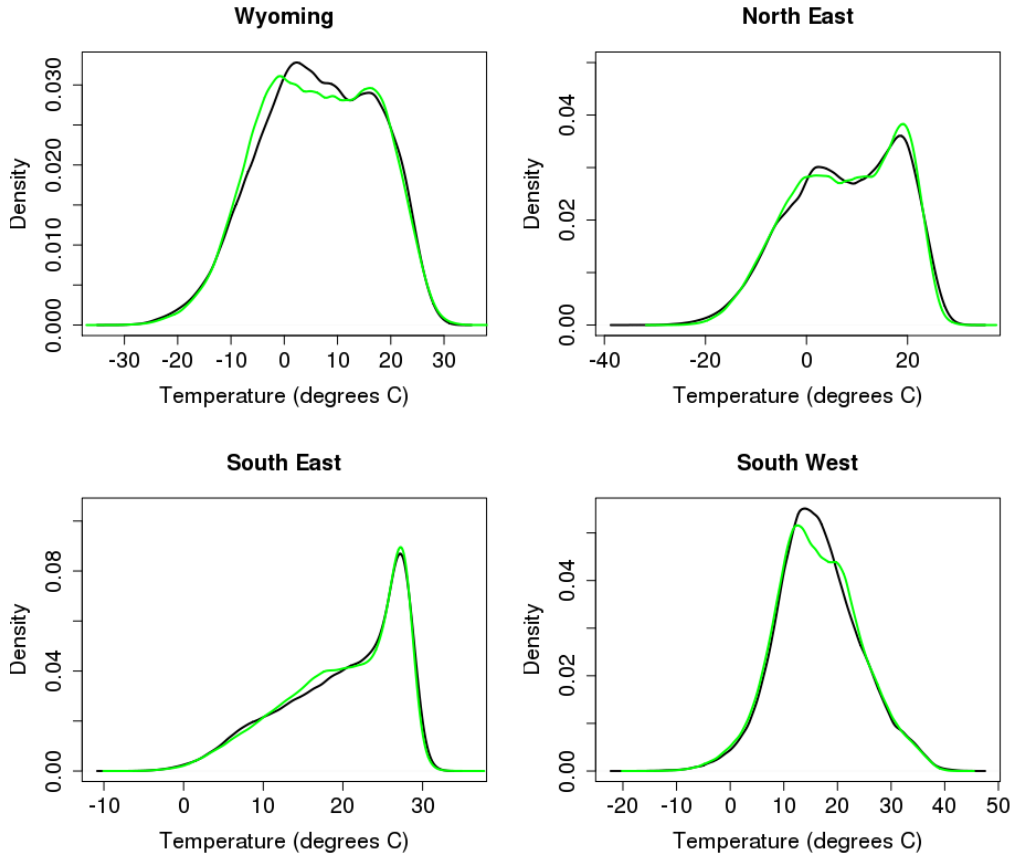
- Density plots showing temperature distributions in the four focus regions
- Red lines = Density of predictions, inter-station correlations too high, extremes not represented

The Model Performance



- Density plots showing temperature distributions in the four focus regions
- Red lines = Density of predictions
- Blue lines = Density of predictions with added variability, inter-station correlations too low

The Model Performance



- Density plots showing temperature distributions in the four focus regions
- Green lines = Density of predictions with added smoothed variability, reasonable reproduction of variability and inter-station correlations

Corrupting the Series

- Many inhomogeneities affect temperature series
- Three of the most common are:
  - Station Relocations
  - Shelter Changes
  - Urbanisation
- These can be mimicked by perturbing input values to the models

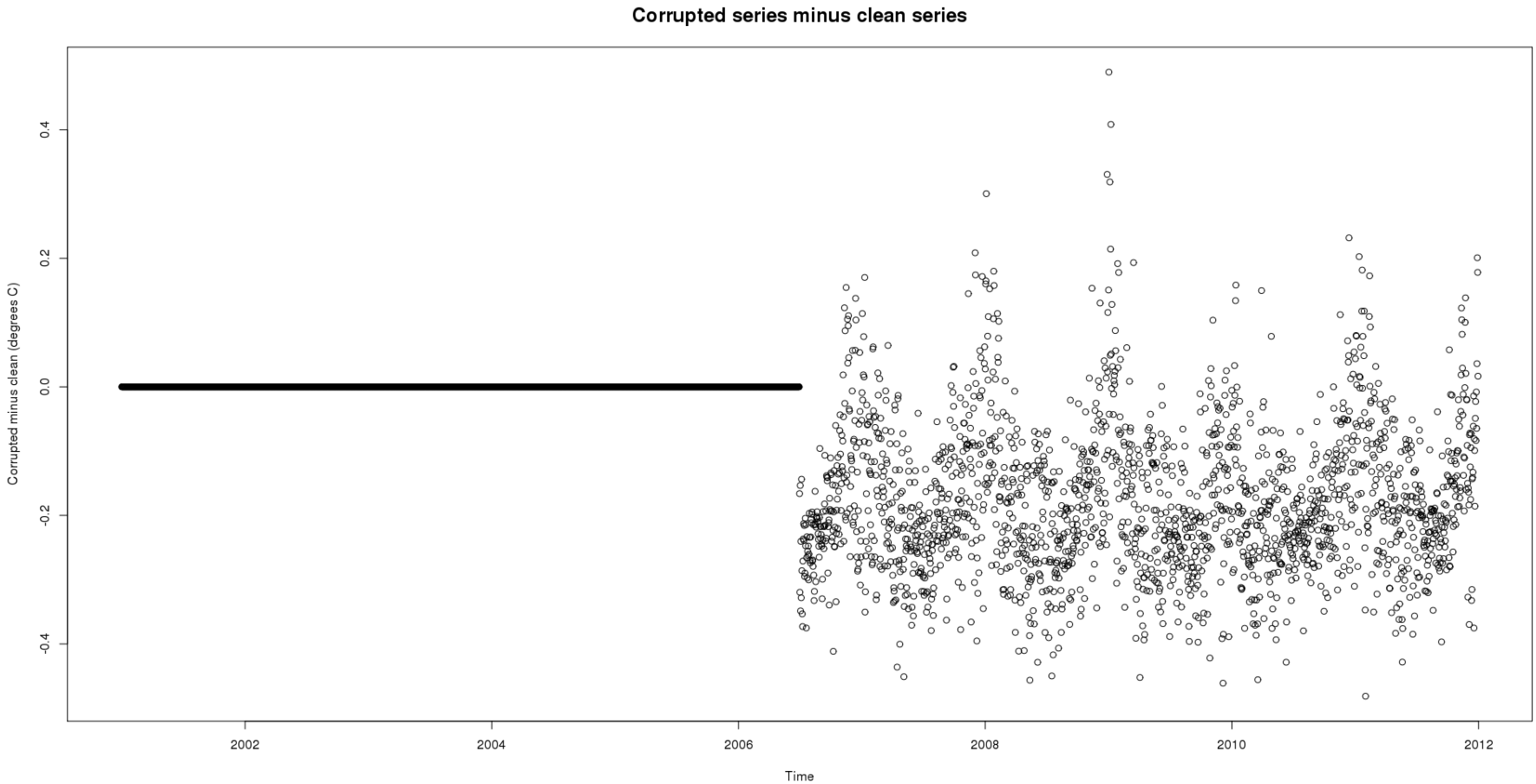
Endless Possibilities...

- Plenty of things to investigate - even with only three inhomogeneities to focus on:
  - Regional differences
  - Size of inhomogeneities
  - Frequency of inhomogeneities
  - Station density
  - Missing data
- 3 worlds investigating different combinations of these options

Implementing the Ideas

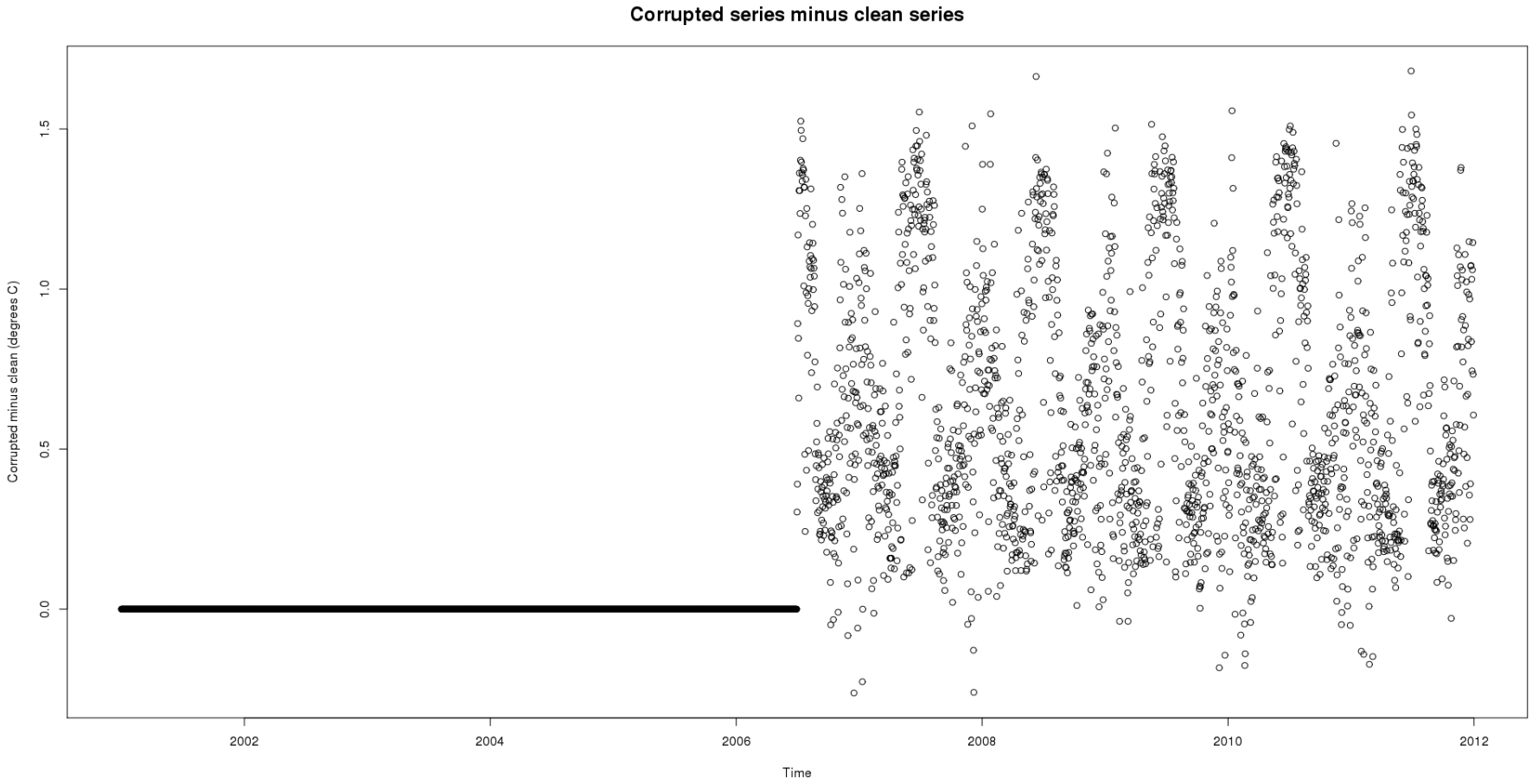
- Step by step process
  1. Poisson process to decide the location of the inhomogeneities
  2. Random number generators to decide type and size
  3. Implementation using explanatory variables
- Inhomogeneities could be amplified/ damped using the addition of constants

Examples - Station Relocation



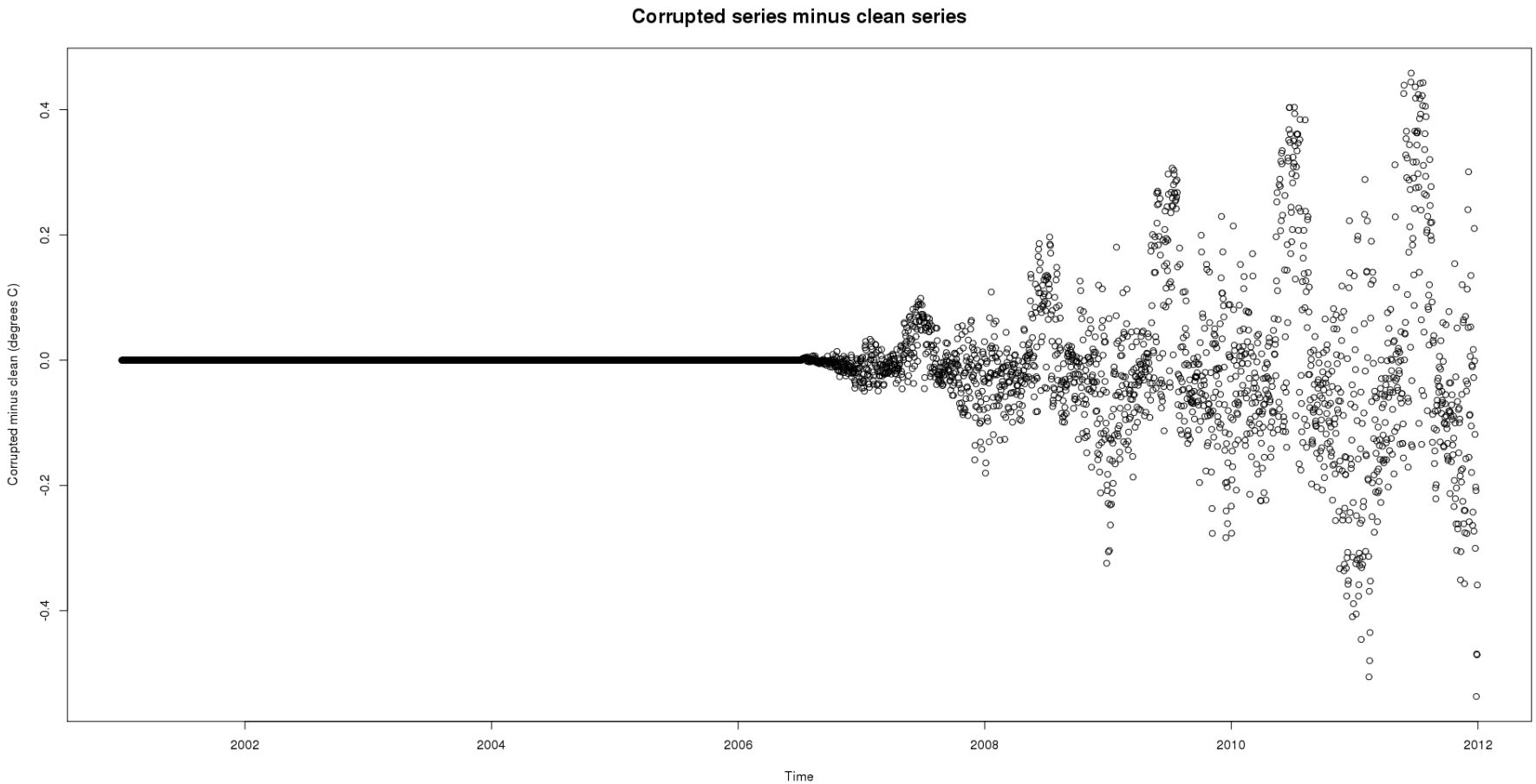
- Inhomogeneity causes an apparent cooling in summer and a less pronounced change in winter

Examples - Shelter Change



- Inhomogeneity causes an apparent warming

Examples - Urbanisation



- Inhomogeneity has an interesting seasonal cycle and this grows as urbanisation increases

Release and Validation

- Aim for release of corrupted series June/ July 2014
- Request results returned by end of November 2014
- The results will be evaluated according to measures such as:
  - False alarm rate
  - Hit rate
  - Similarity to original clean series
- The results of these findings will then be released, allowing the improvement of algorithms and in turn the improvement of real world temperature series

Future Work

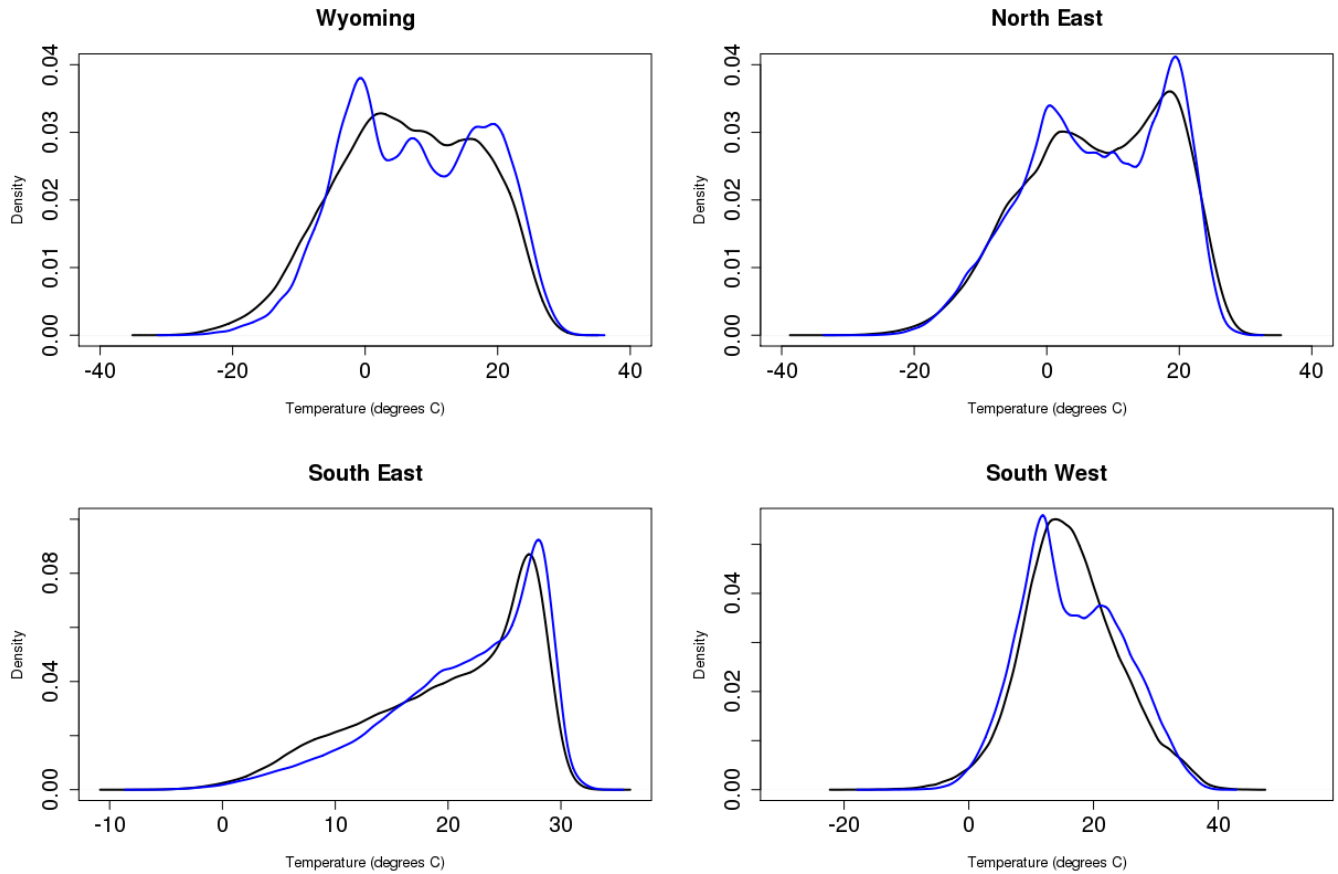
- More types of inhomogeneity to focus on
- More regions of the world available
- More variables to investigate

Summary

- The problem: We need homogeneous daily data with a well understood uncertainty estimate to robustly assess the impacts of climate change
- The method: Using statistical models to create synthetic temperature series where the truth is known a priori to allow the benchmark testing of daily homogenisation algorithms
- The intended outcomes: The assessment and improvement of daily homogenisation algorithms leading to more robust daily data sets

Re-analysis Temperatures

- Why not just use reanalysis temperatures if more can be assumed about their homogeneity?
- Scale - we can't get the same station level variability



The Actual Model Formulation

$$TMEAN60_{it} \sim Gamma(\mu_{it}, \phi)$$

where

$$\mu_{it} = \beta_0 + \beta_1 Altitude_{it} + \beta_2 Tempforecast_{it} + f_1(Dyear_{it}) + f_2(Time_{it}) + f_3(Lat_{it}) + f_4(Long_{it}) + f_5(Sun_{it}) + f_6(SOI_{it}) + f_7(UW_{it}) + f_8(VW_{it}) + f_9(Precip_{it}) + f_{10}(PWC_{it}) + f_{11}(SLP_{it}) + f_{12}(Coast_{it})$$

- You can also include smooth surfaces
- Have investigated smooth surfaces of Day of the Year and Eastward wind, Northward wind and Precipitation - Allowing a seasonally varying relationship between the variable in question and temperature

rw307@exeter.ac.uk