Building Bayesian models: Why bother if you don't believe?
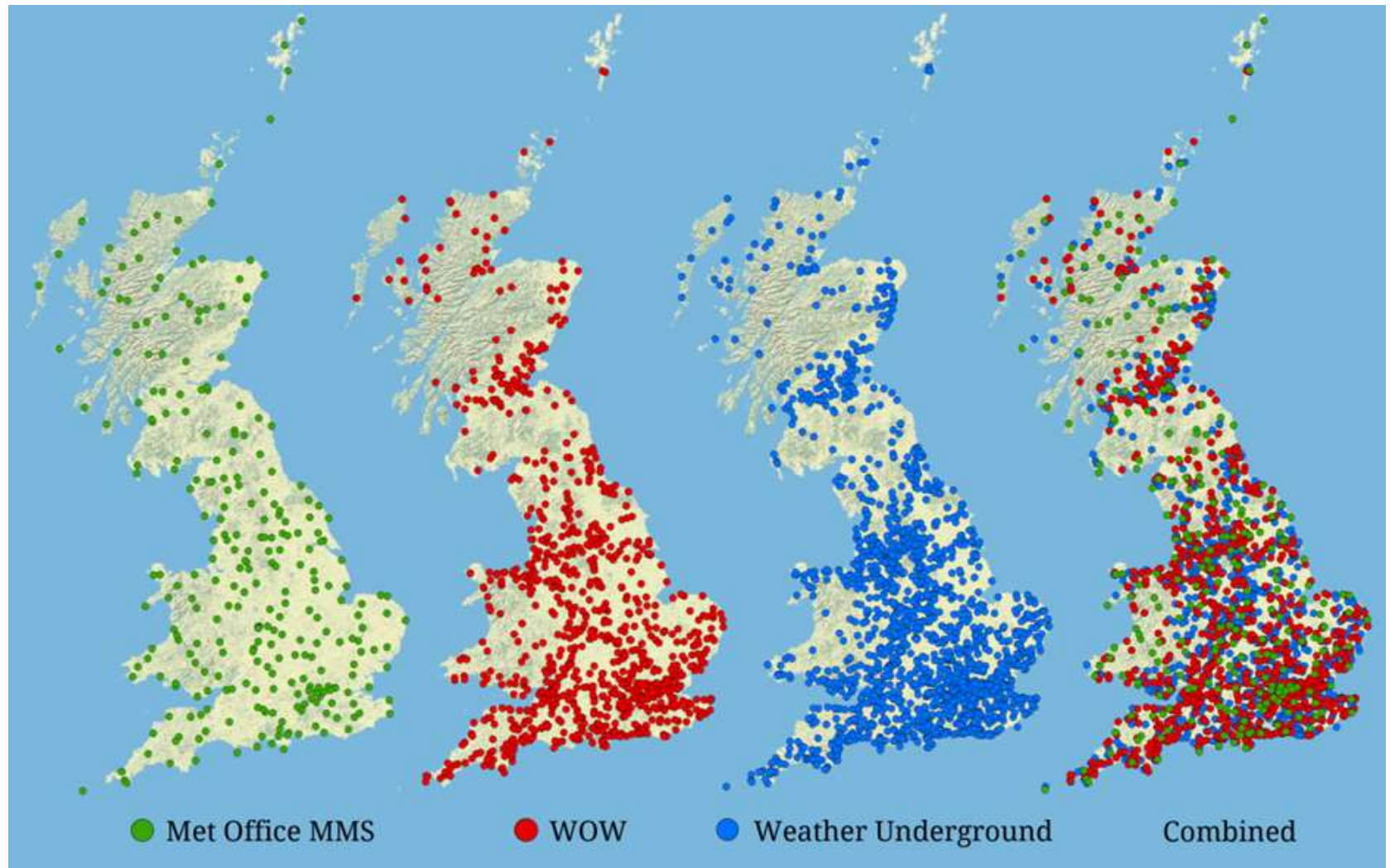
Dan Cornford, Simon Bell

Aston University

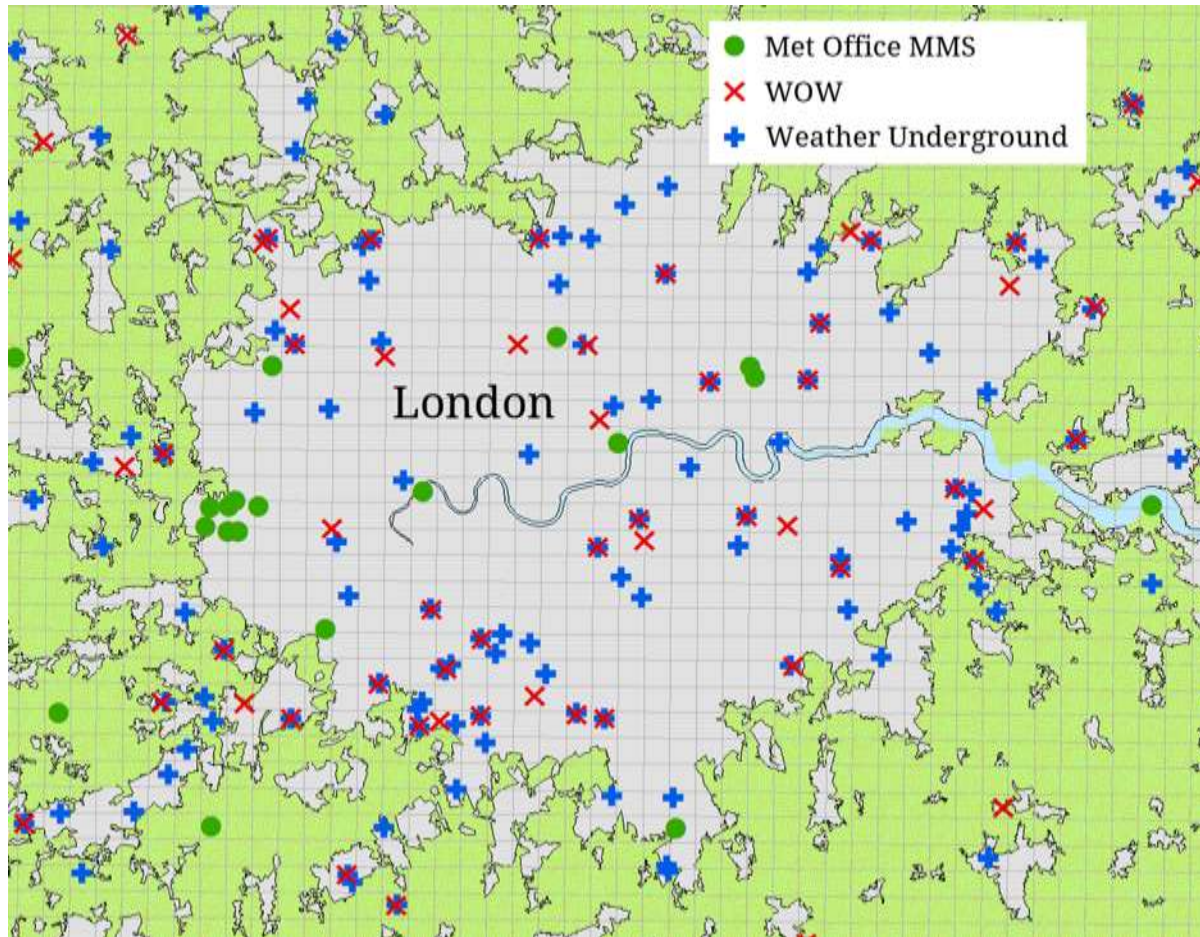# Spatial distribution of citizen weather stations

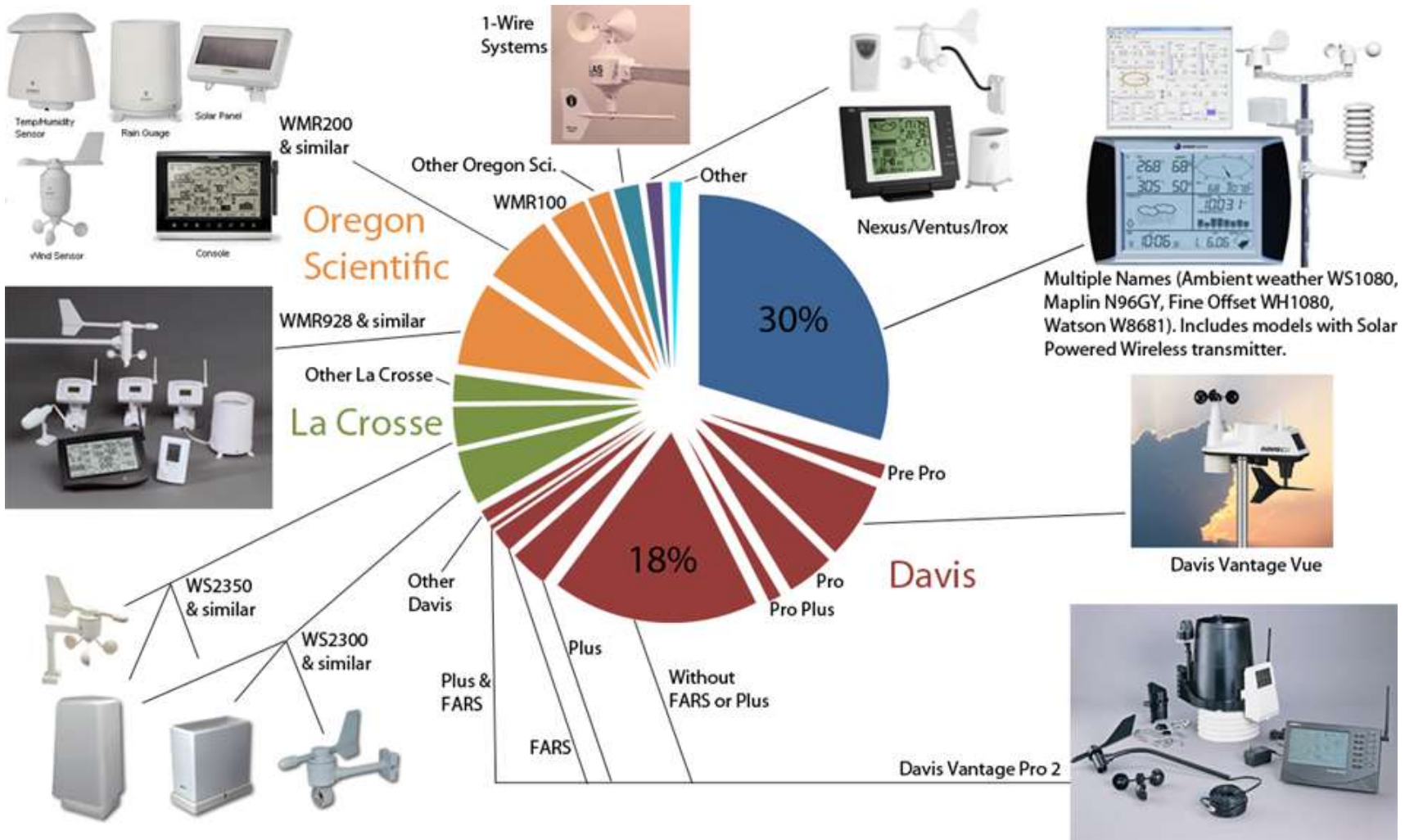Met Office MMS · WOW · Weather Underground · Combined

In more detail …

Growth of citizen weather data

Types of stations

Geophysical

Elevation

Coastality

Urbanisation

Met Office UKV NWP

MMS (tower/station)

MSG Satellite

Citizen Weather Stations

UKV Temperature forecast

MMS Temperature Data

MMS Radiation Data

Global Horizontal Irradiance Estimates

Hourly Infrared & Visible Satellite Images

CWS Temperature Data

CWS Metadata

Sample at CWS & MMS locations

Smoothing of synoptic variables?

Project images to British National Grid
Section **Error! Reference source not found**

Initial Quality Control
Section **Error! Reference source not found.**

Radiation Interpolation Model
Section **Error! Reference source not found**

Station Exposure & UCZ Classifier
Section **Error! Reference source not found.**

Station Design Classifier
Section **Error! Reference source not**

Radiation Estimates at CWS locations

CWS Station Exposure and UCZ Class

CWS Station Design Class

Interpolation Model
Section **Error! Reference source not found.**

Temperature Estimates at CWS locations

Bias Correction Model
Section **Error! Reference source not found.**

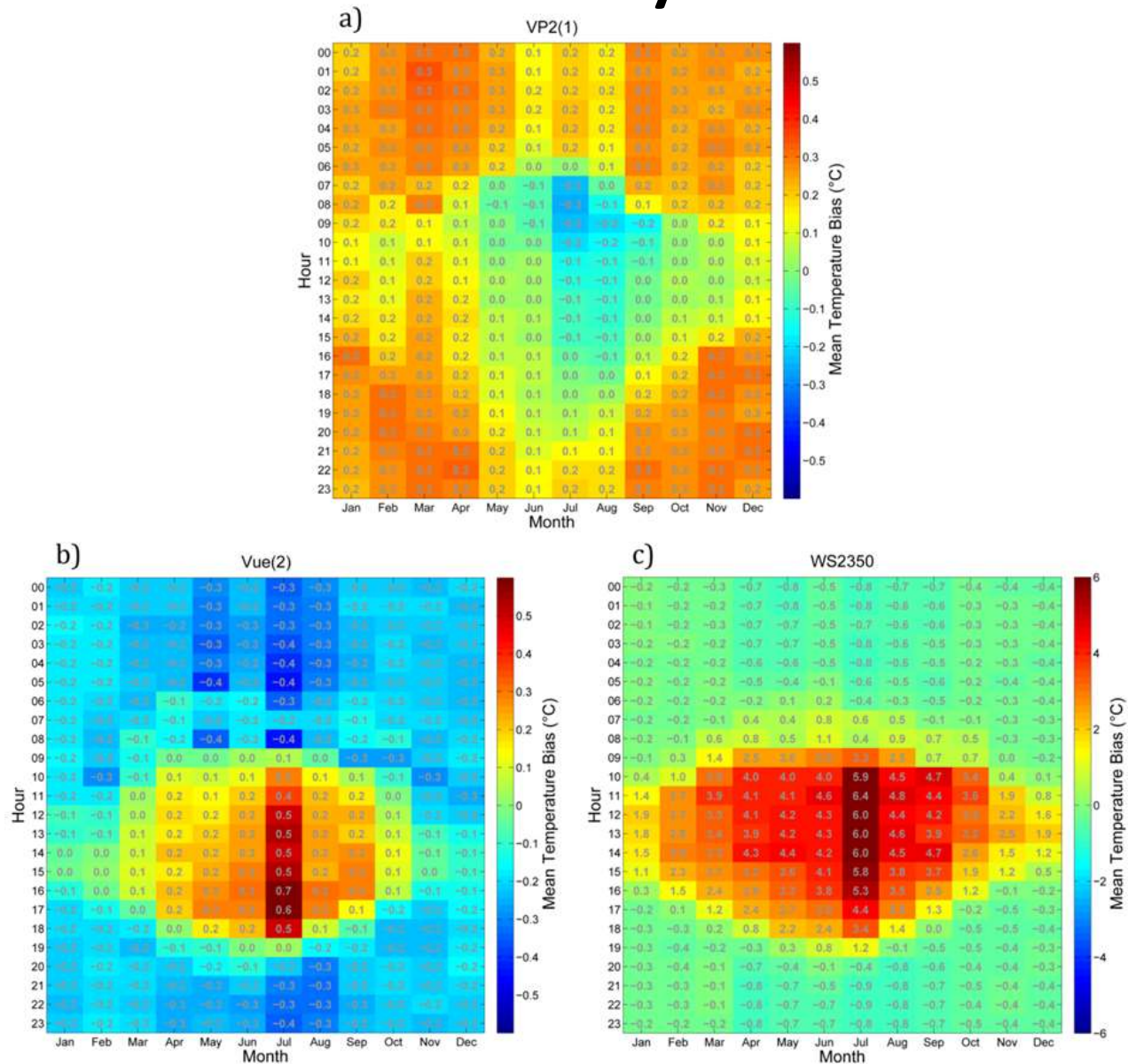Bias-Corrected CWS data with Uncertainty Estimates
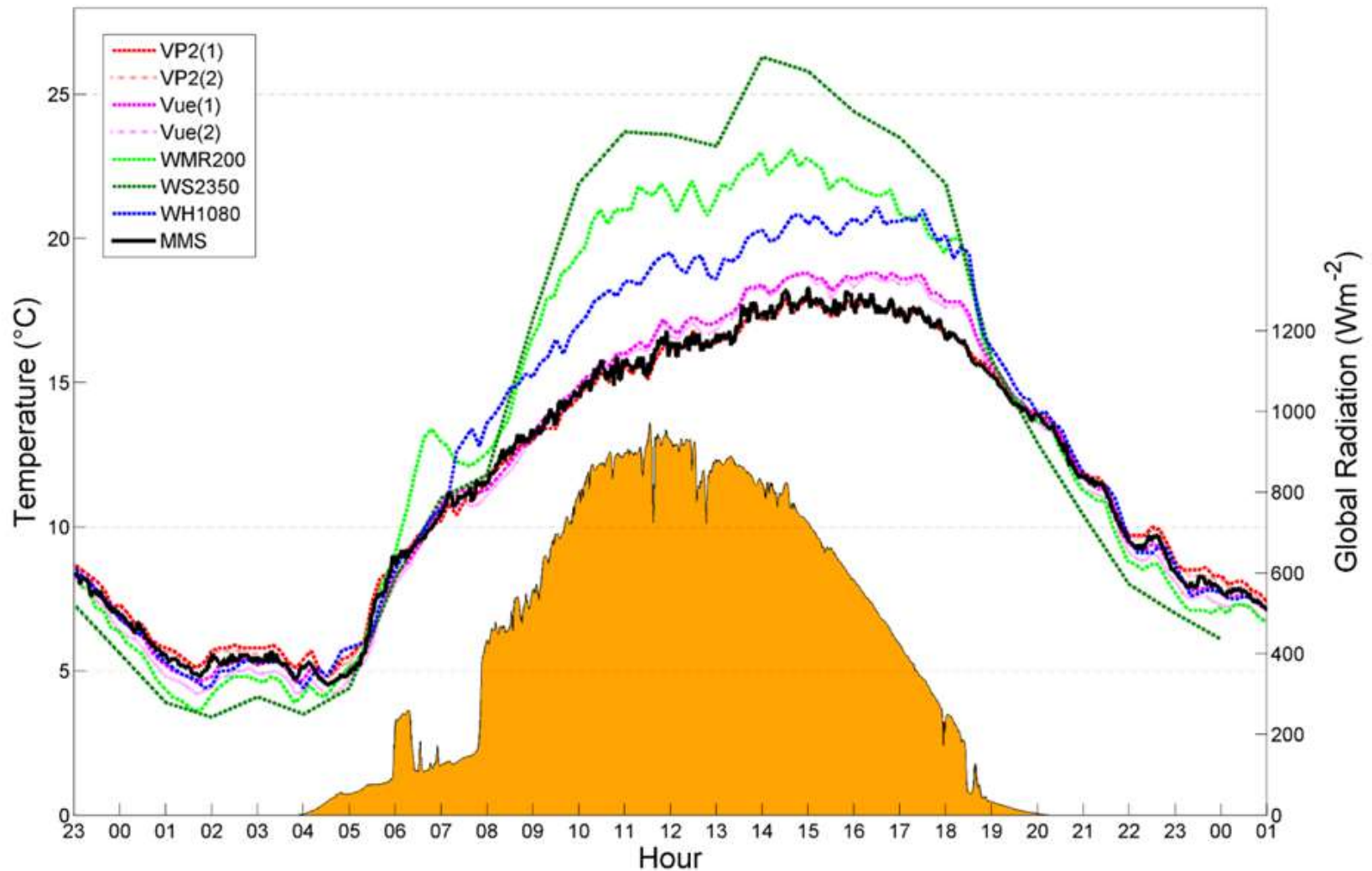
Data

Preprocessing

Model or Process

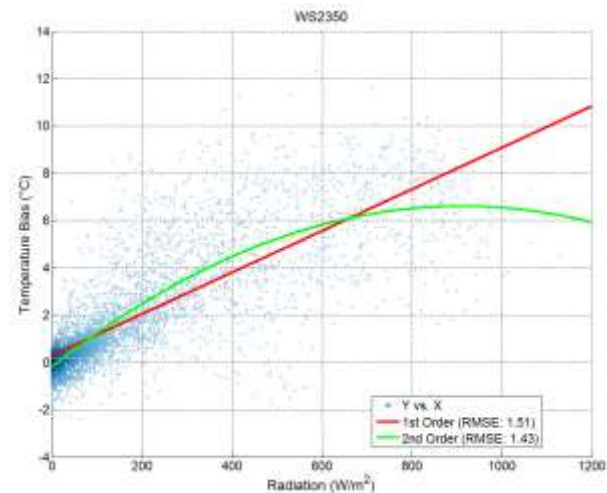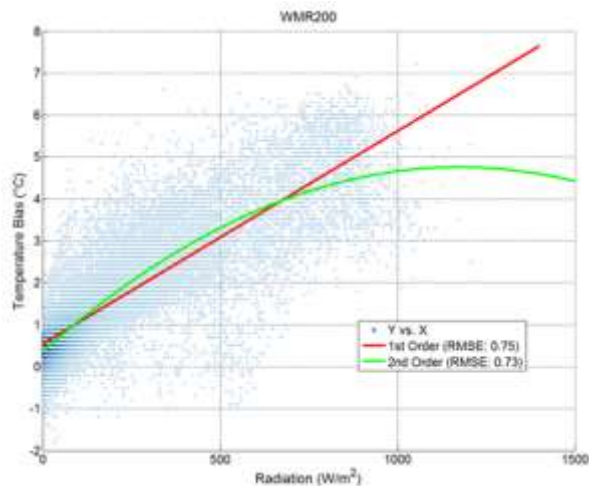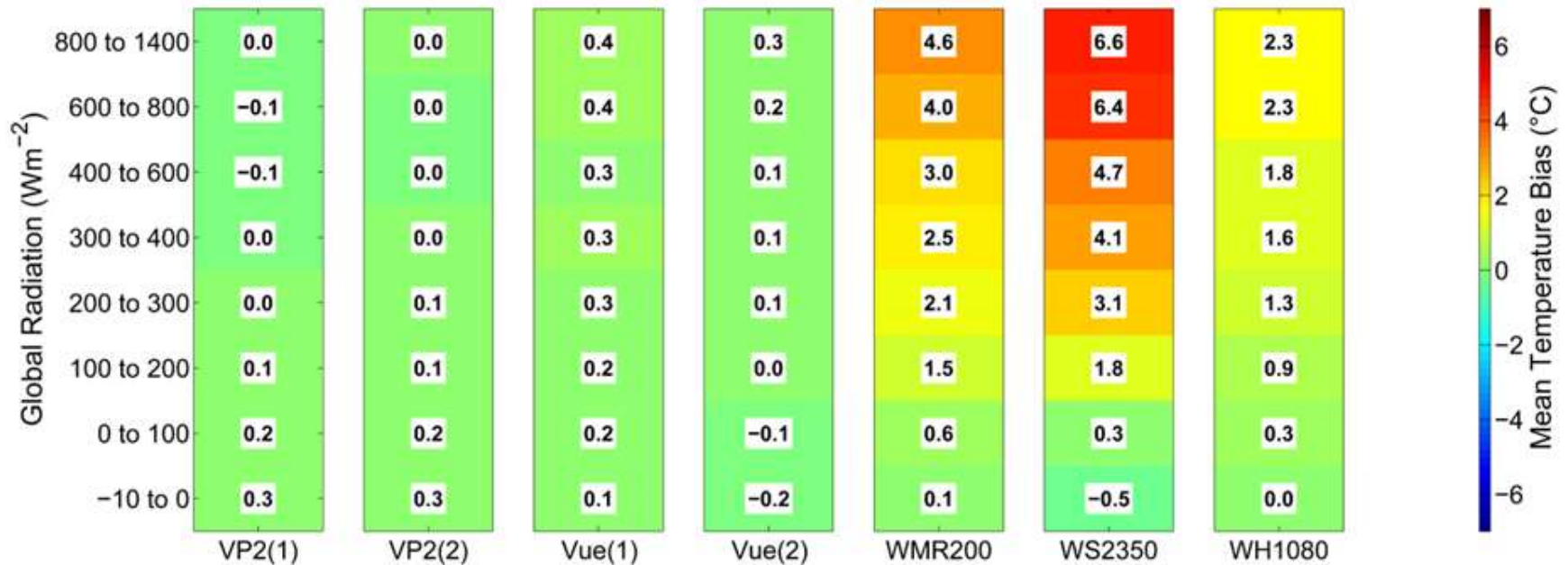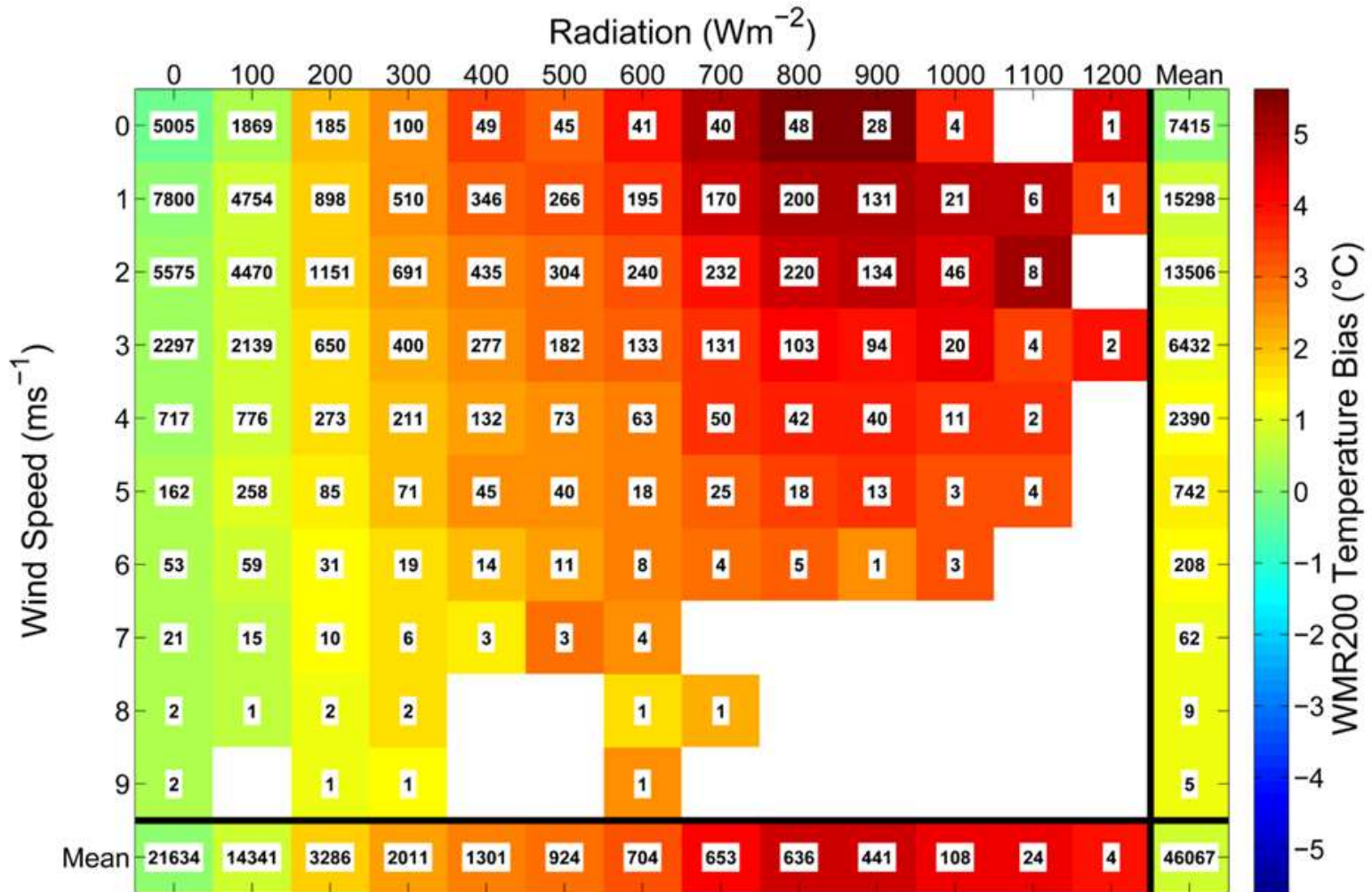# Data quality issues

Assessing station quality

That is totally biased
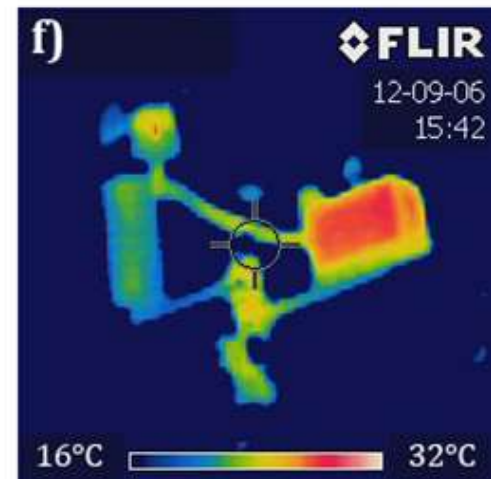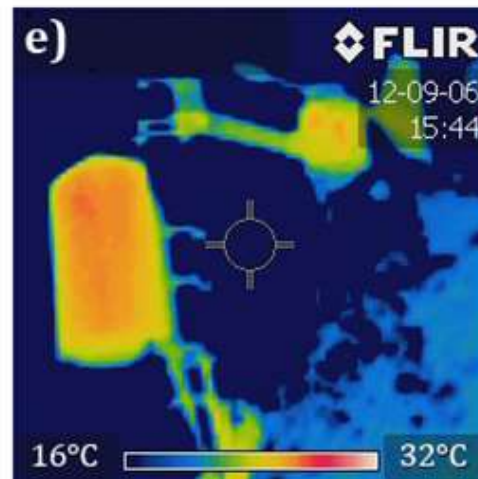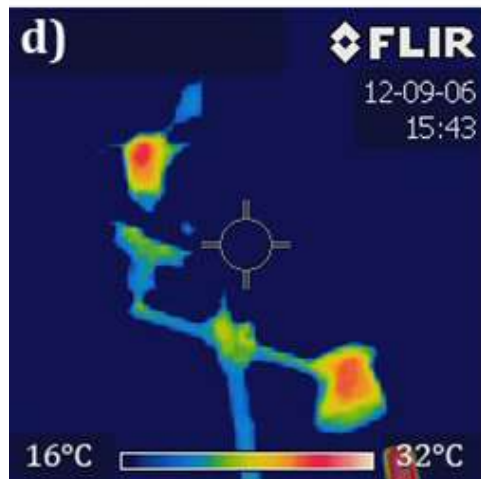
Exploring why – the role of physics and graphics

Building a model

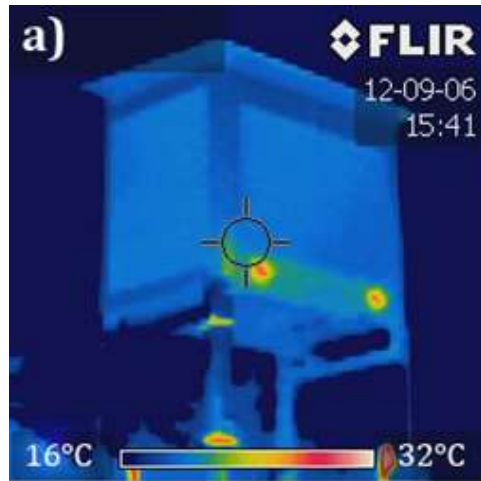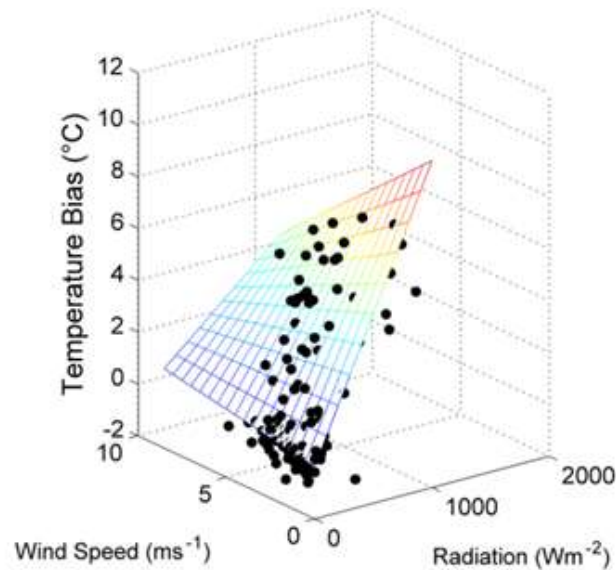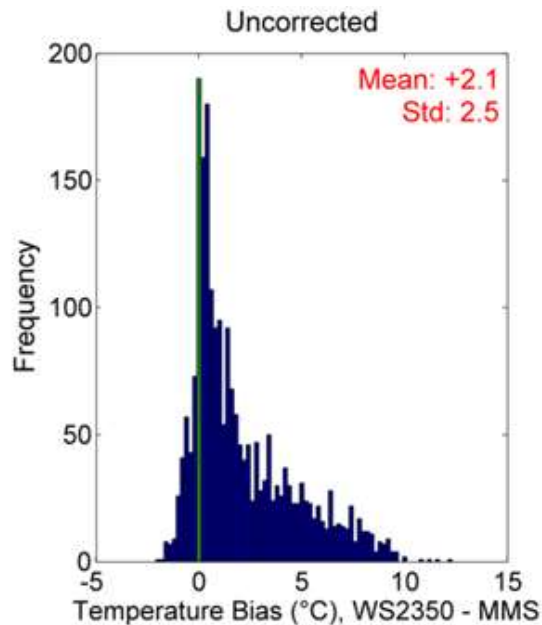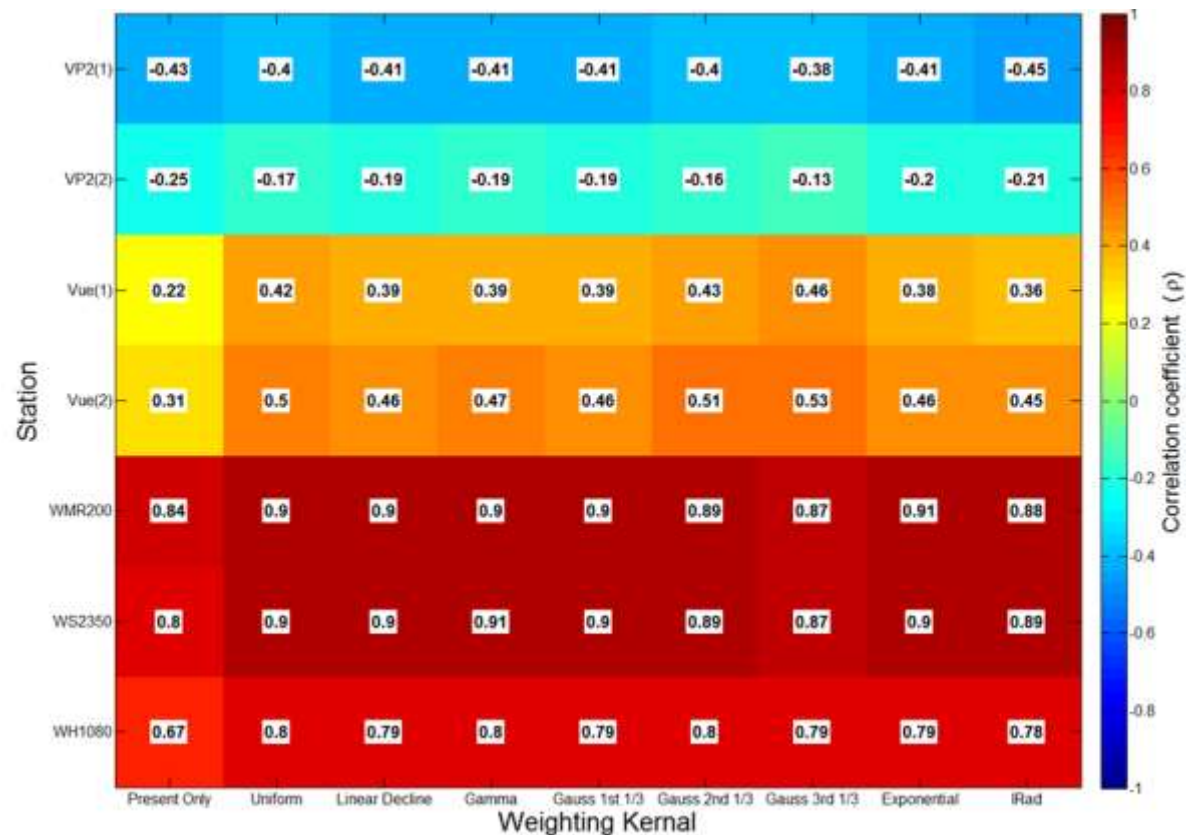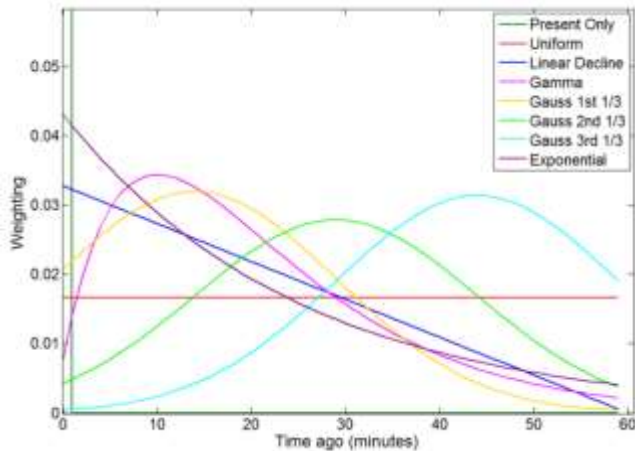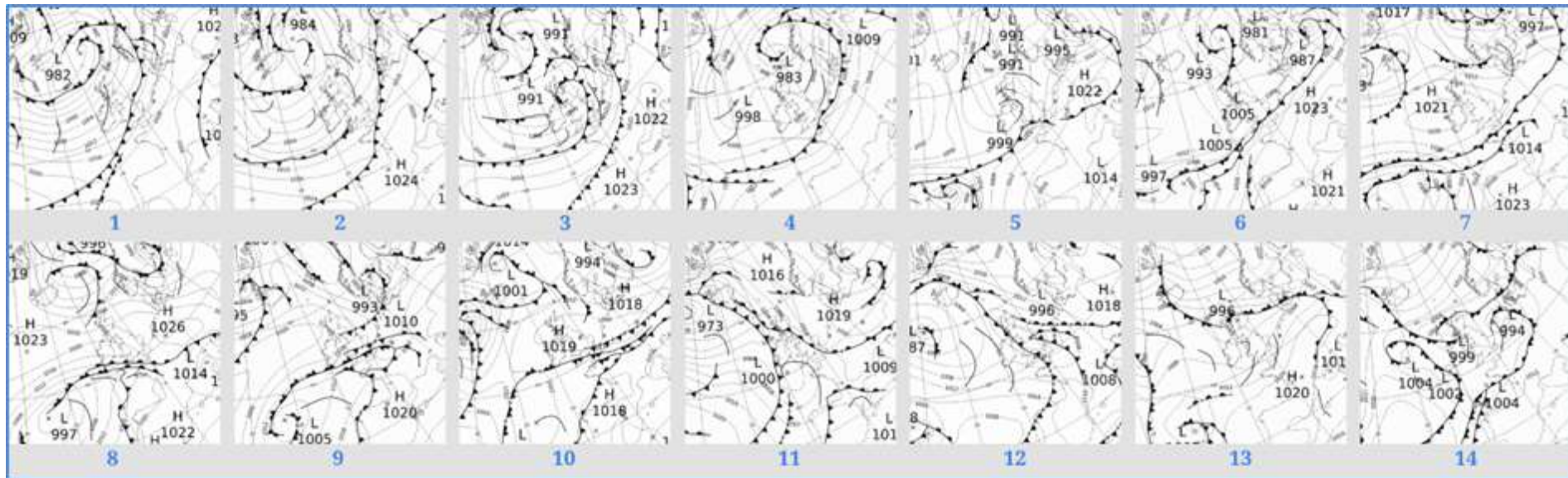# Other explanations?
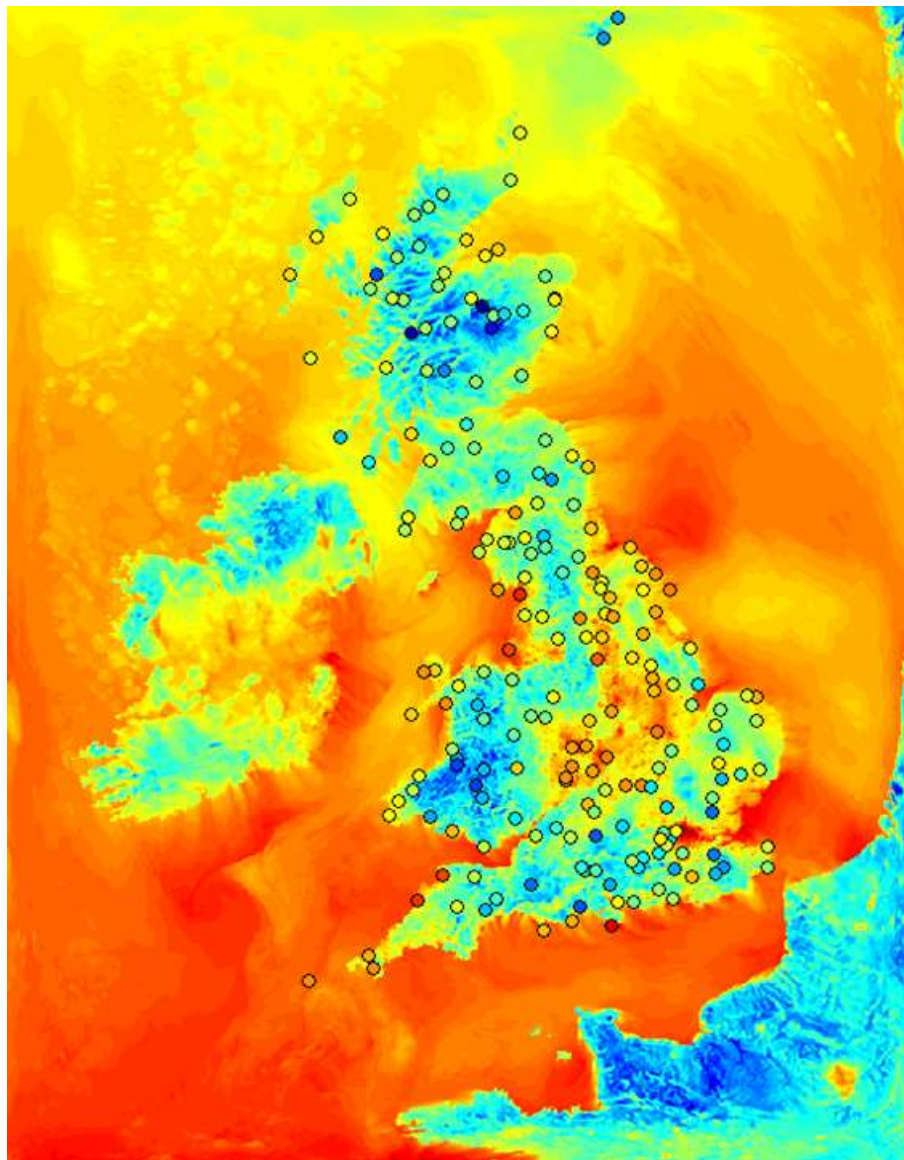
Confirmation – a mechanism
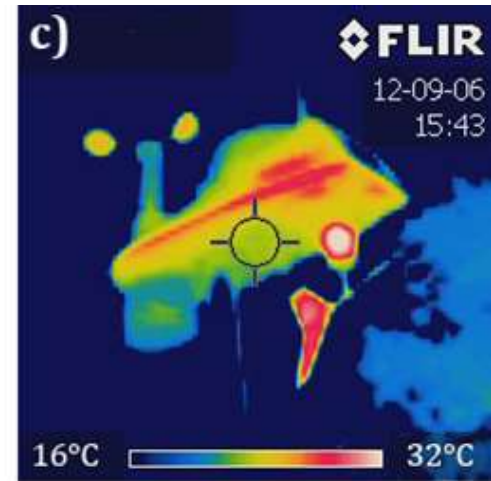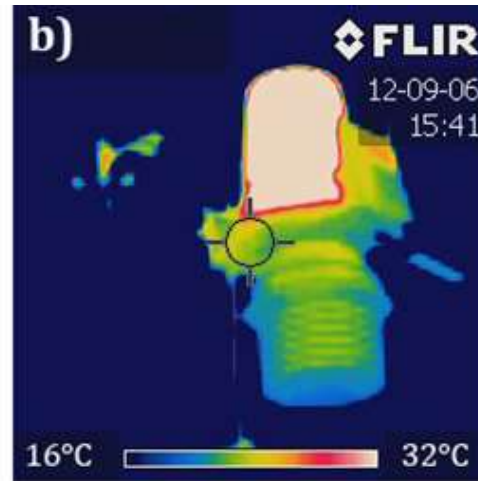
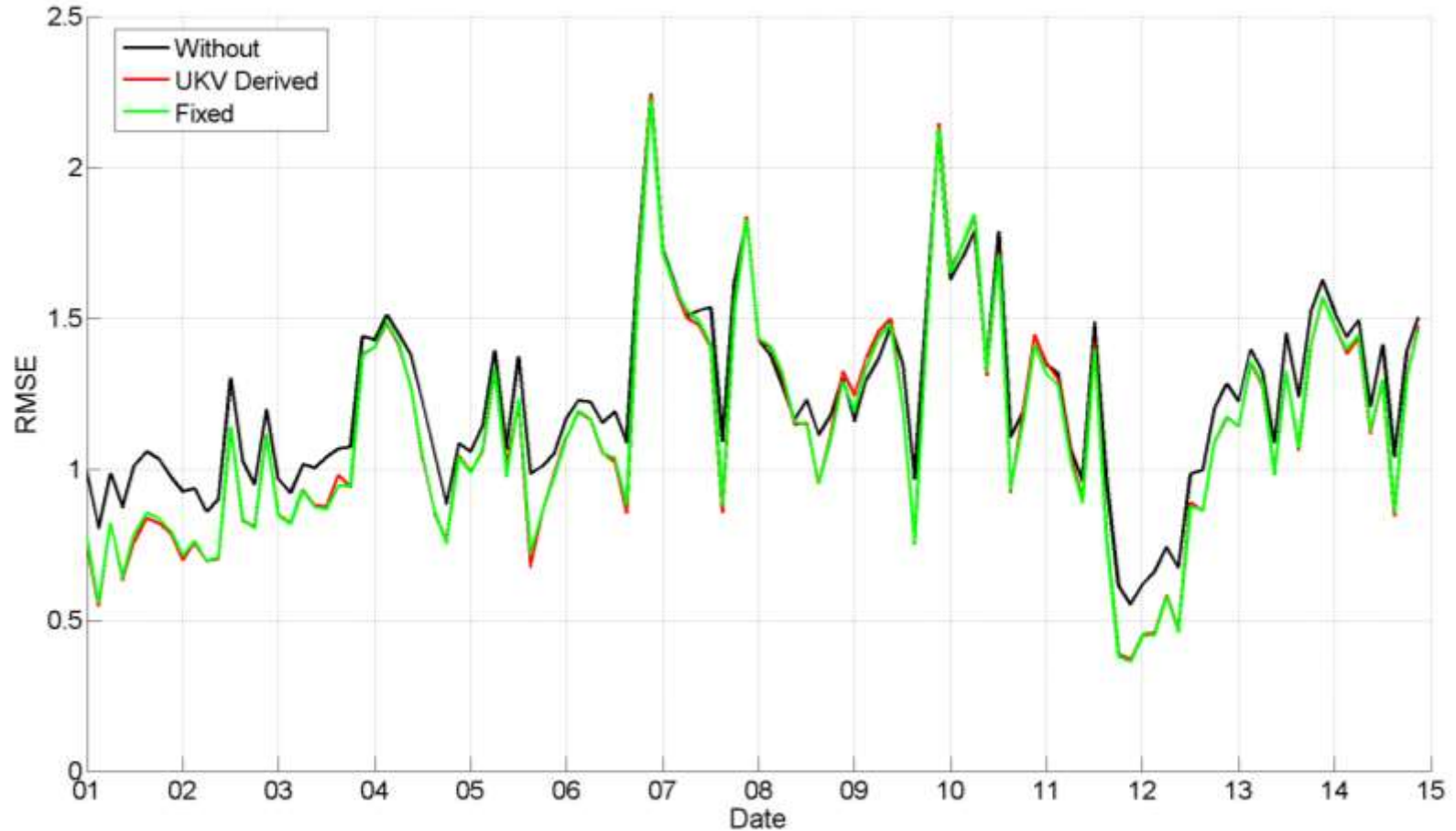# Simple models can go a long way
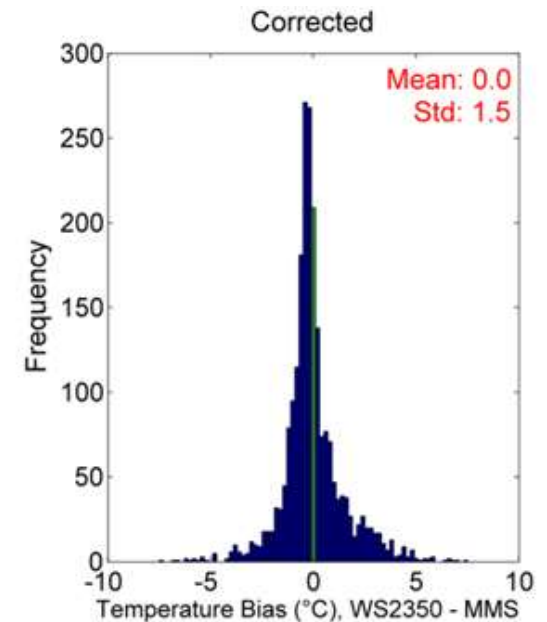
Marginal improvements, better physics?
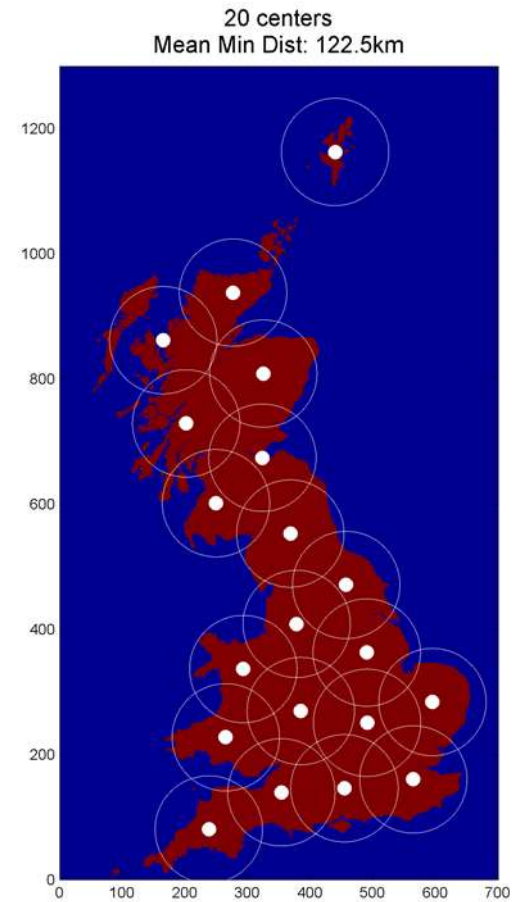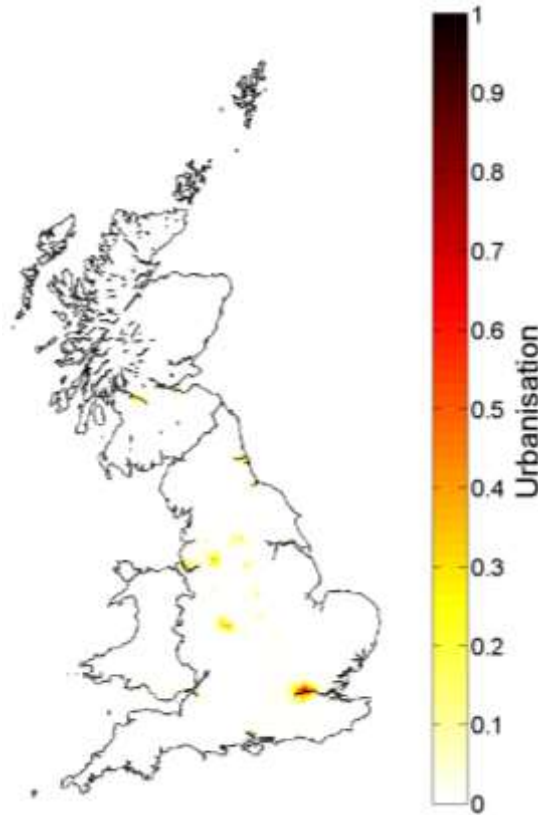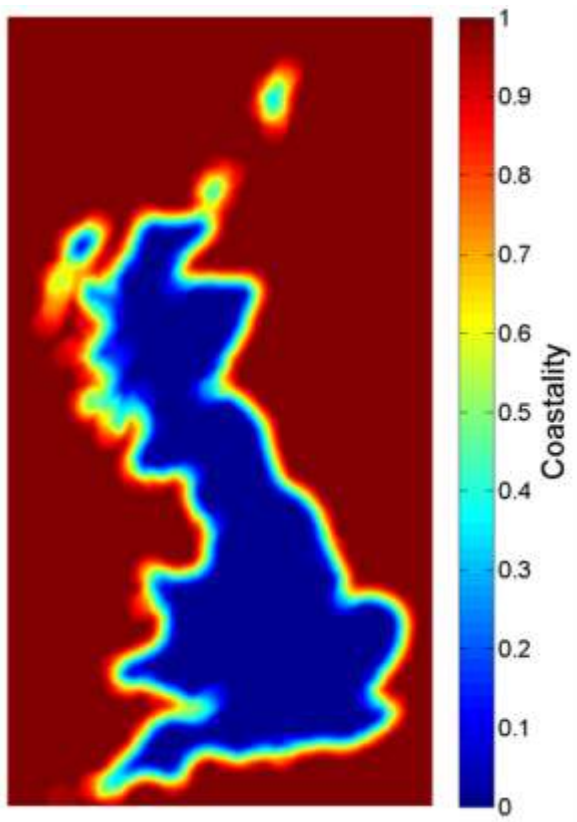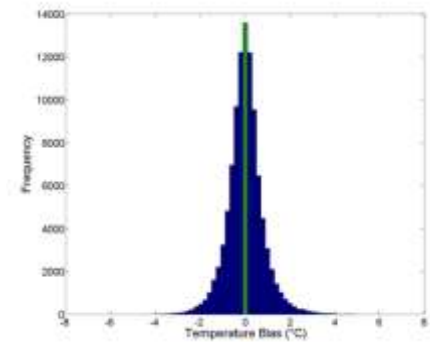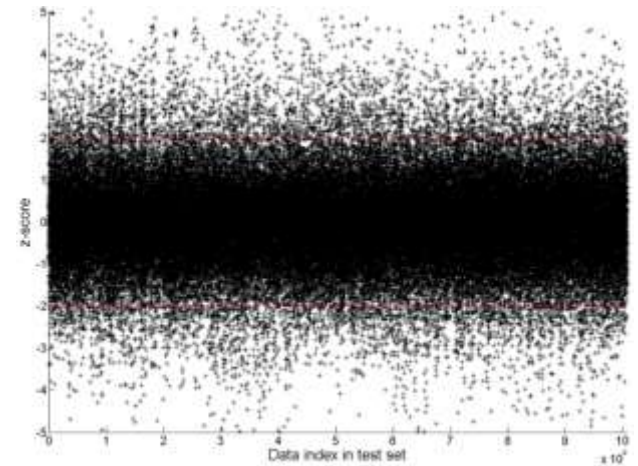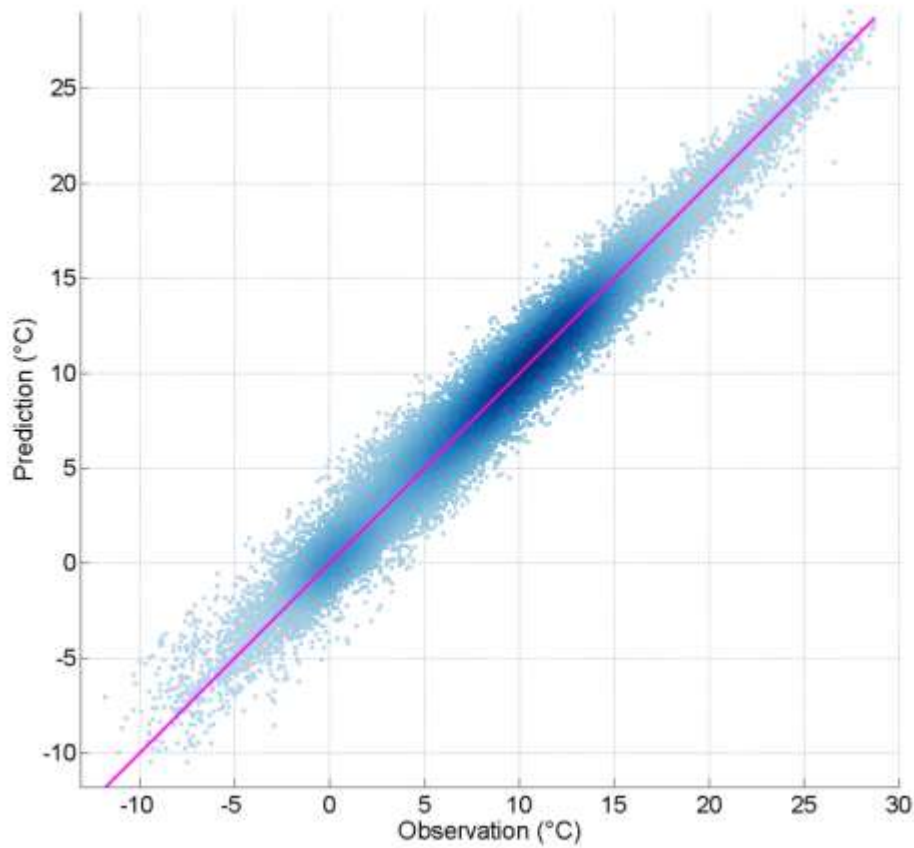
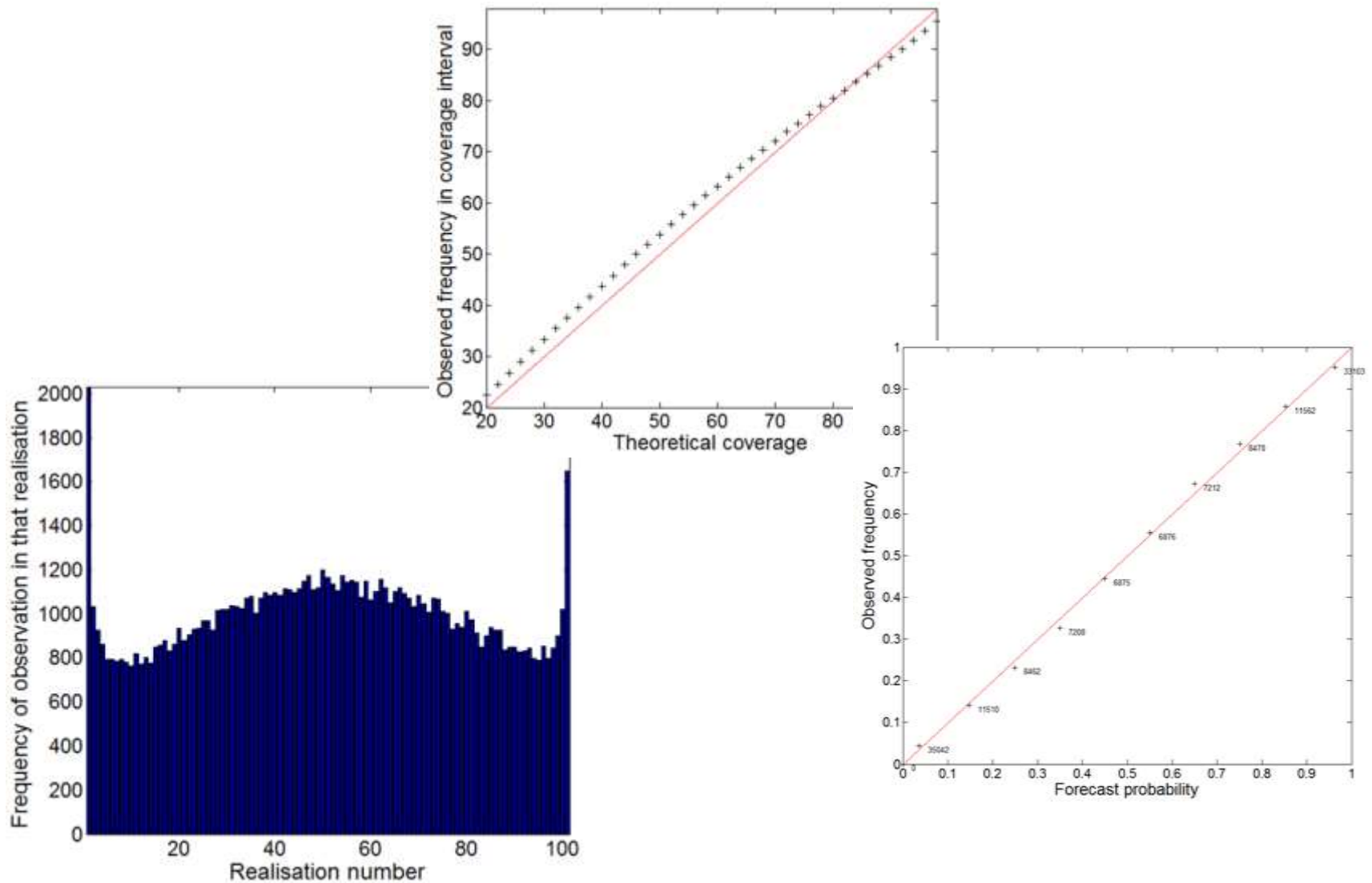# What is a case study?

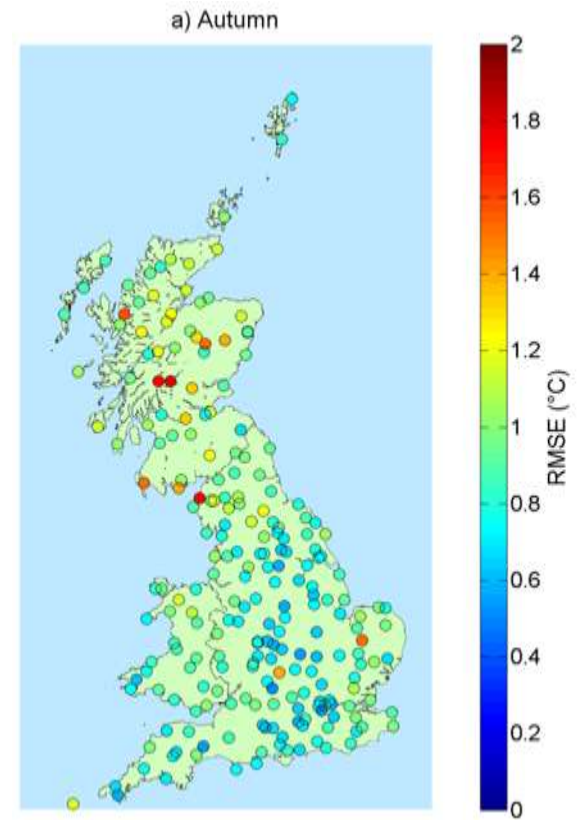# Really complex priors …
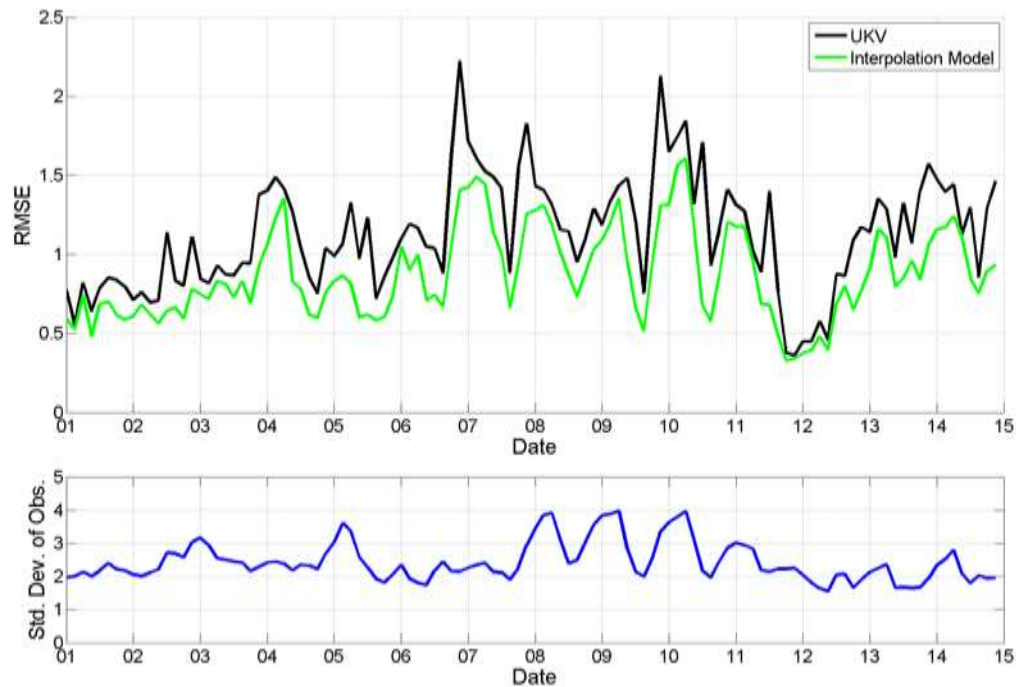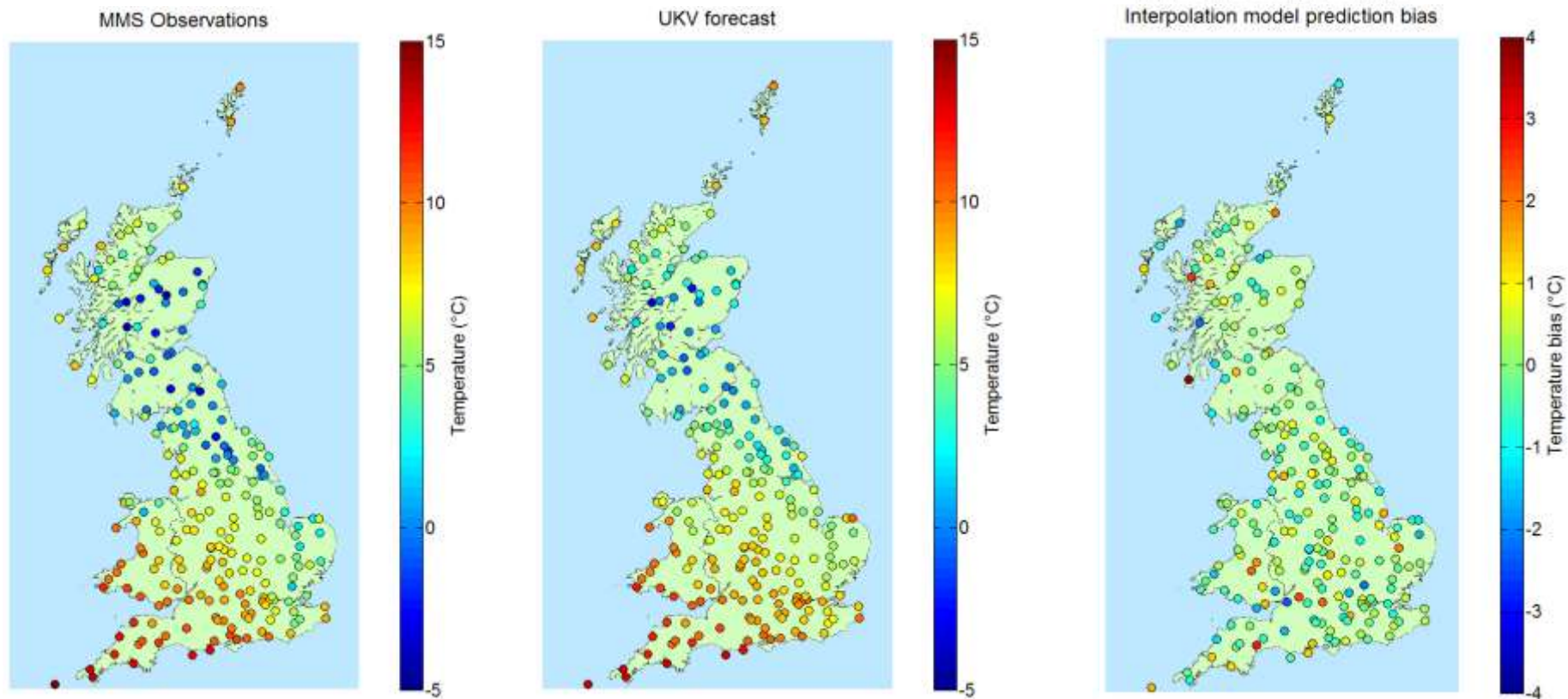
Simple corrections can help

Select your inputs carefully …

And validate

Carefully ...

Sir Plotalot

Seriously …

MMS Observations

UKV forecast

Interpolation model prediction bias

Understanding your model
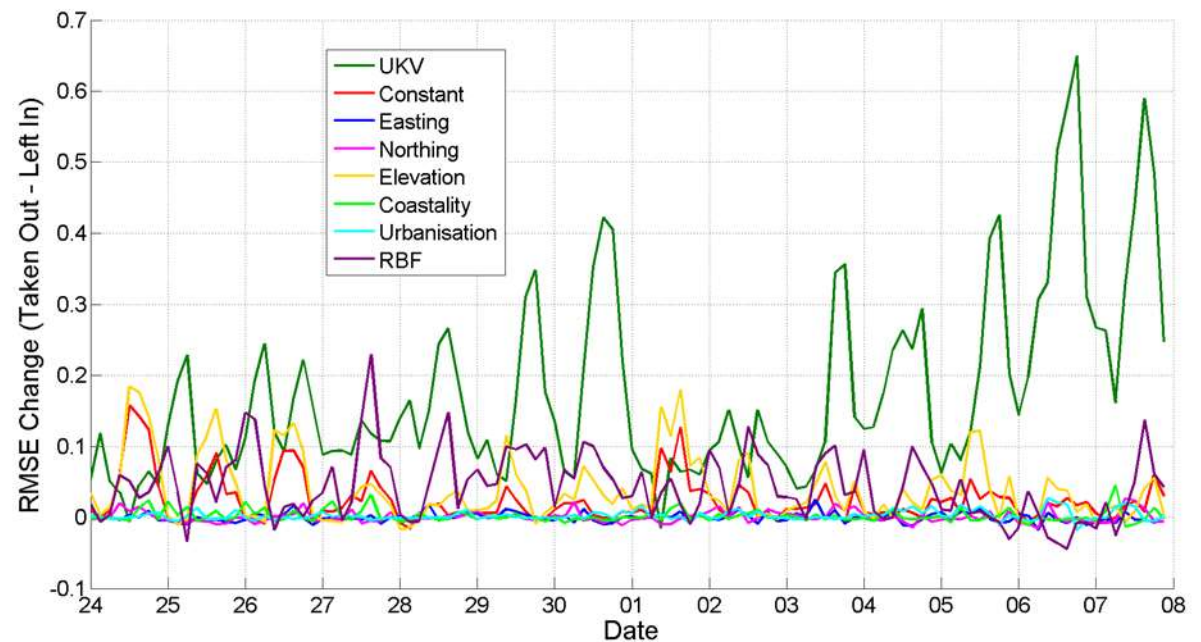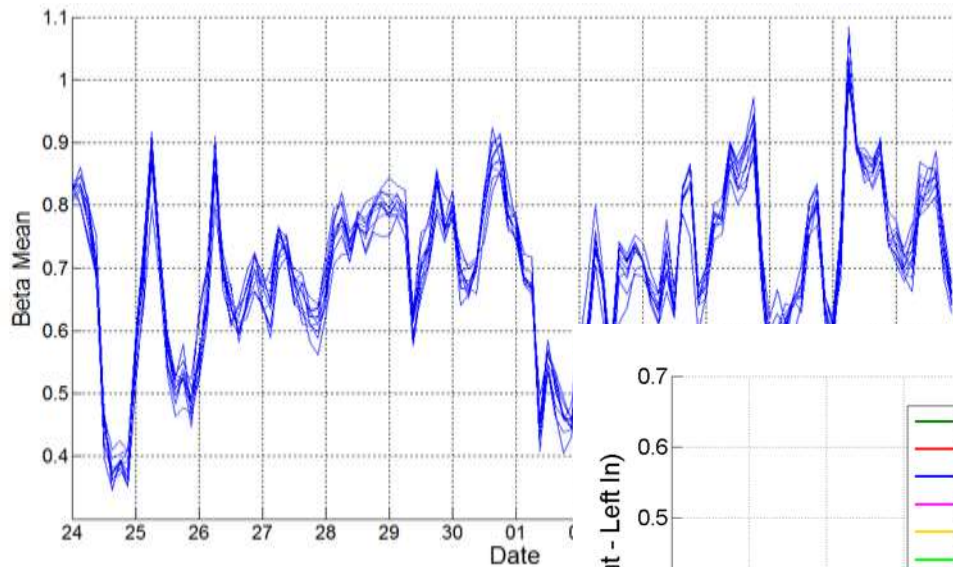
Using your knowledge is never easy
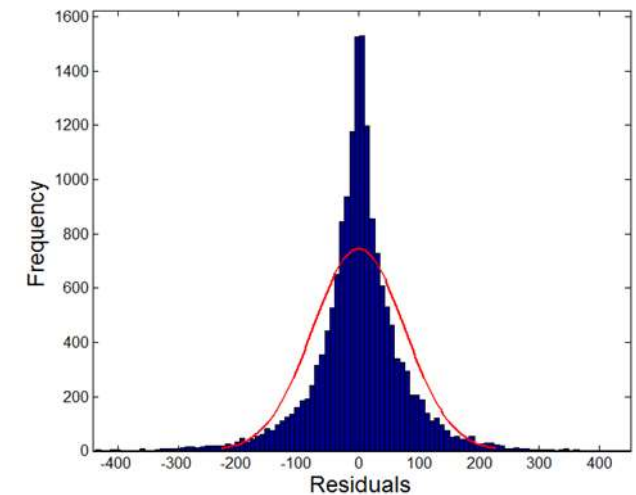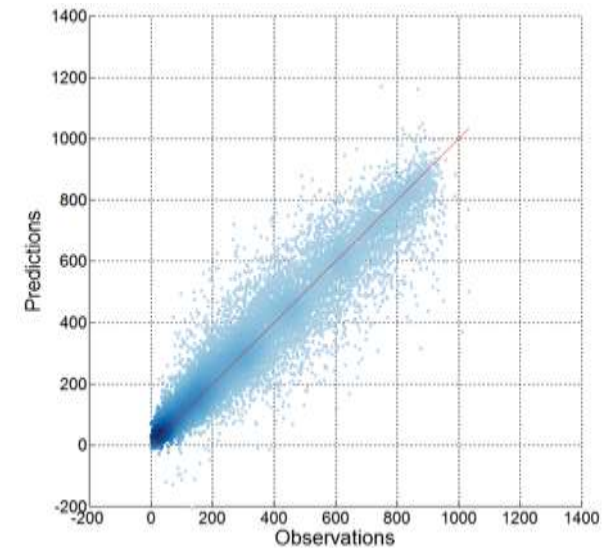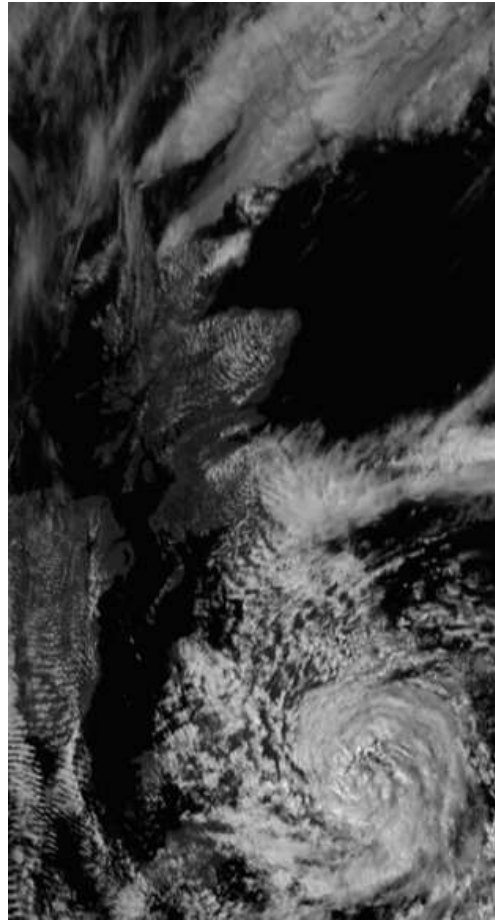
But always look for simple solutions

Ask the public ...

Geophysical

- Elevation
- Coastality
- Urbanisation

Met Office UKV NWP

MSG Satellite

Citizen Weather Stations

UKV Temperature forecast

MMS Temperature Data

MMS Radiation Data

Global Horizontal Irradiance Estimates

Hourly Infrared & Visible Satellite Images

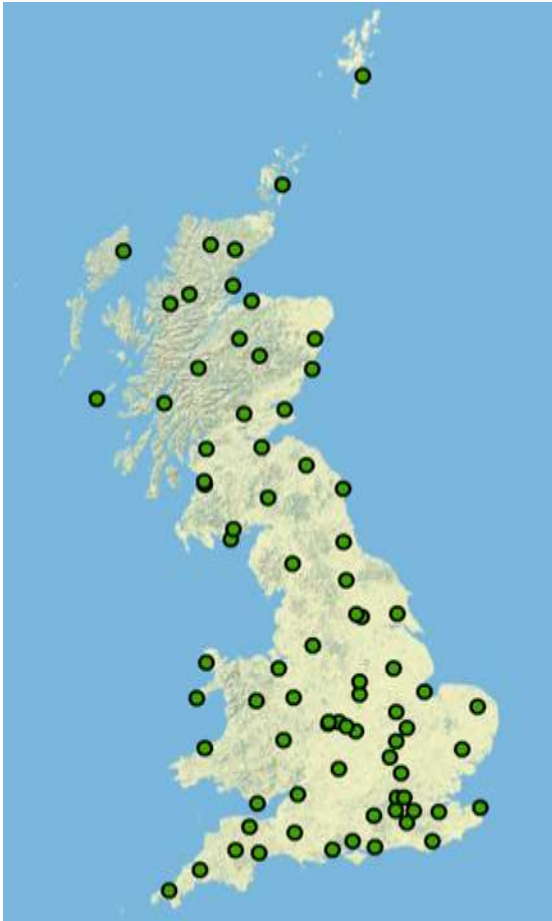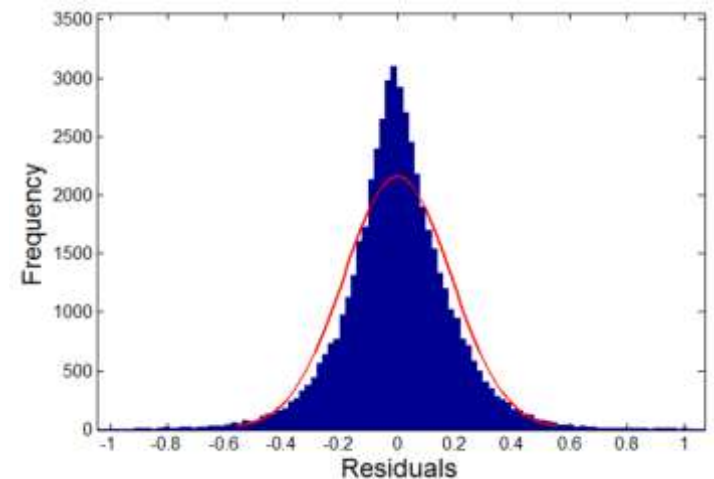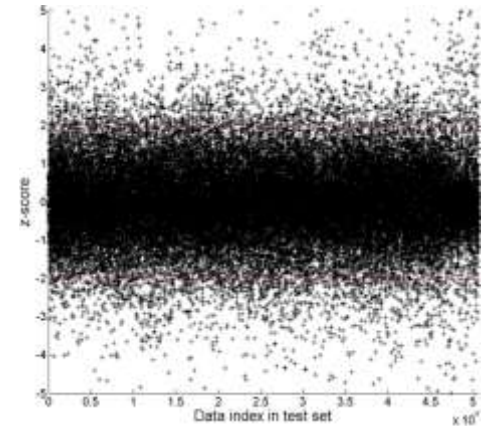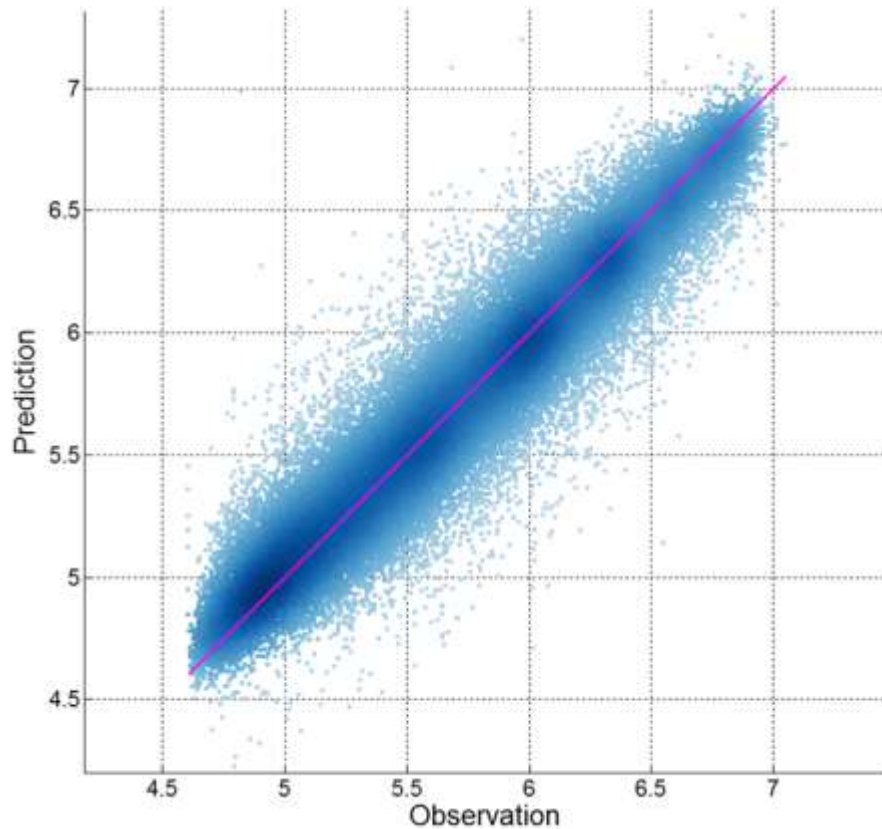CWS Temperature Data

CWS Metadata

Sample at CWS & MMS locations

Smoothing of synoptic variables?

Project images to British National Grid
Section **Error! Reference source not found**

Initial Quality Control
Section **Error! Reference source not found.**

Radiation Interpolation Model
Section **Error! Reference source not found**

Station Exposure & UCZ Classifier
Section **Error! Reference source not found.**

Station Design Classifier
Section **Error! Reference source not**

Radiation Estimates at CWS locations

CWS Station Exposure and UCZ Class

CWS Station Design Class

Interpolation Model
Section **Error! Reference source not found.**

Temperature Estimates at CWS locations

Bias Correction Model
Section **Error! Reference source not found.**

Bias-Corrected CWS data with Uncertainty Estimates
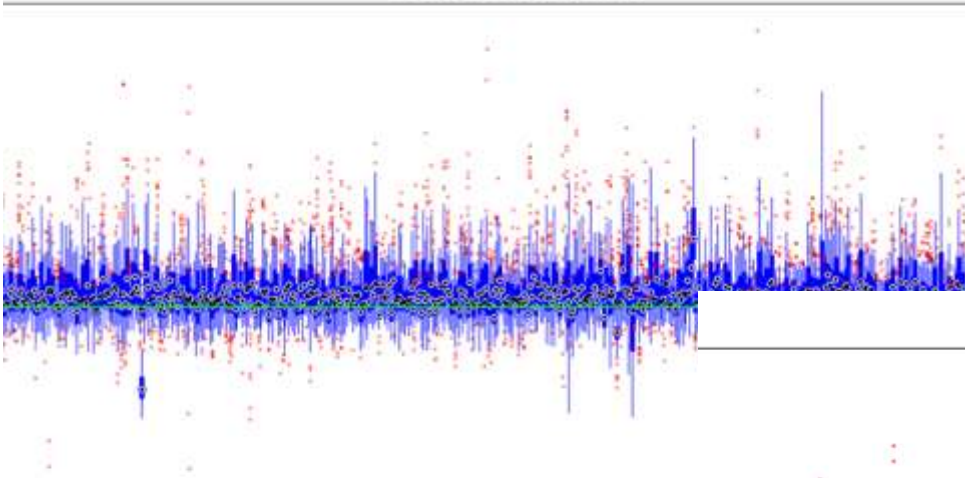
- Data
- Preprocessing
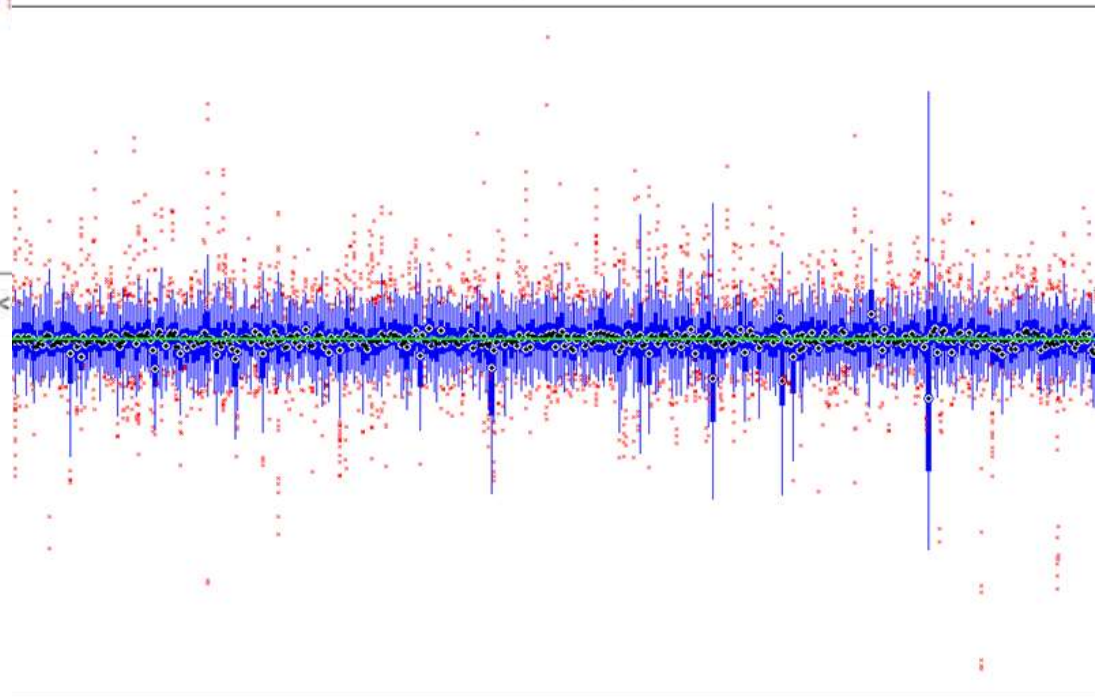- Model or Process
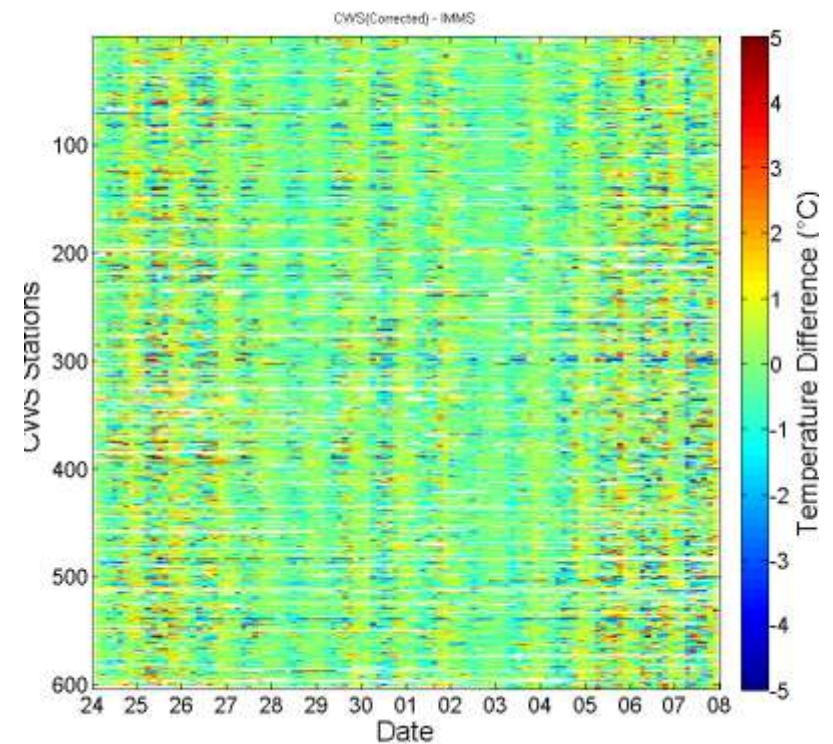
You can fixing things

To an extent

CWS(uncorrected) - IMMS

CWS(Corrected) - IMMS
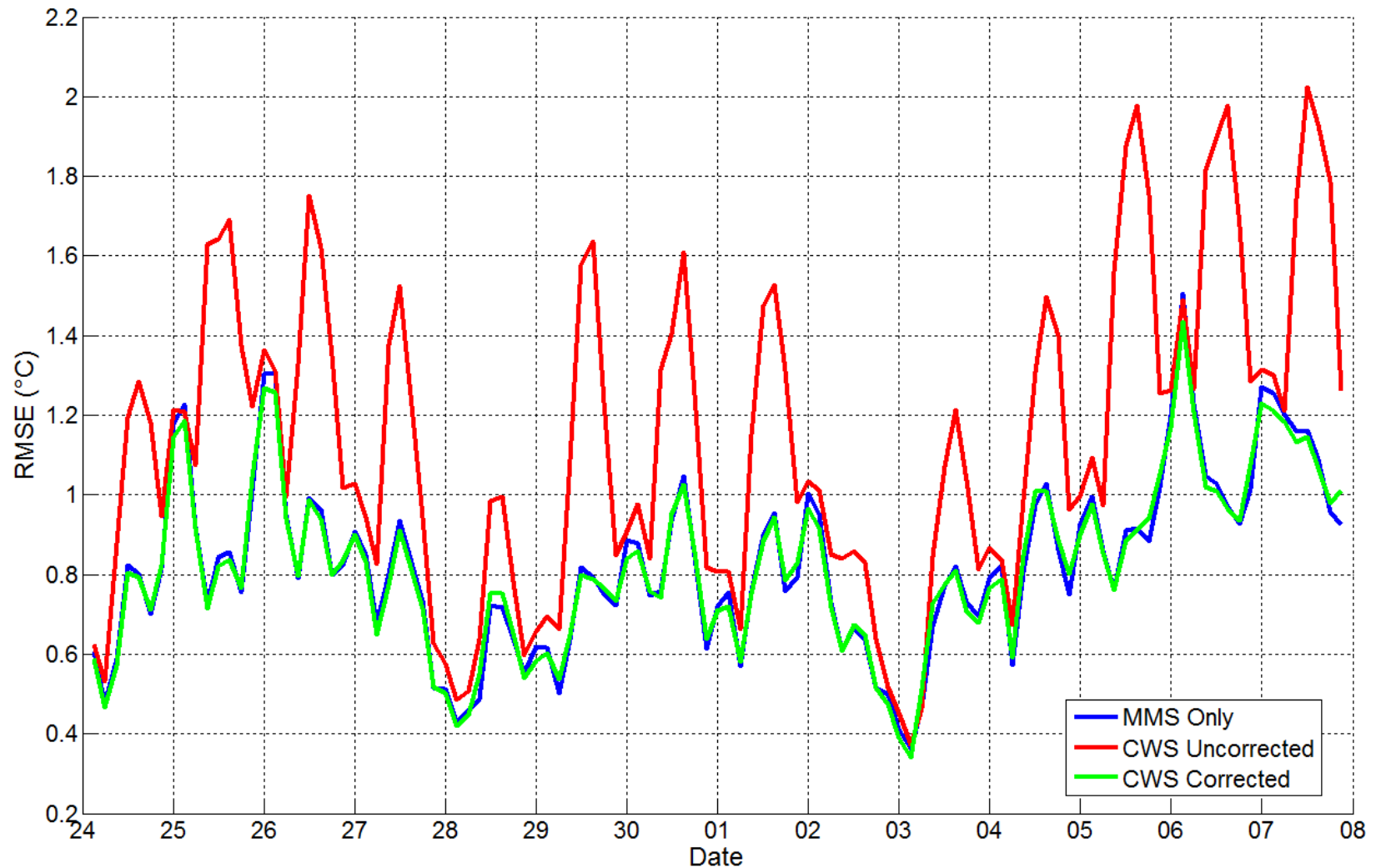
Stations - Ordered by elevation (Low <
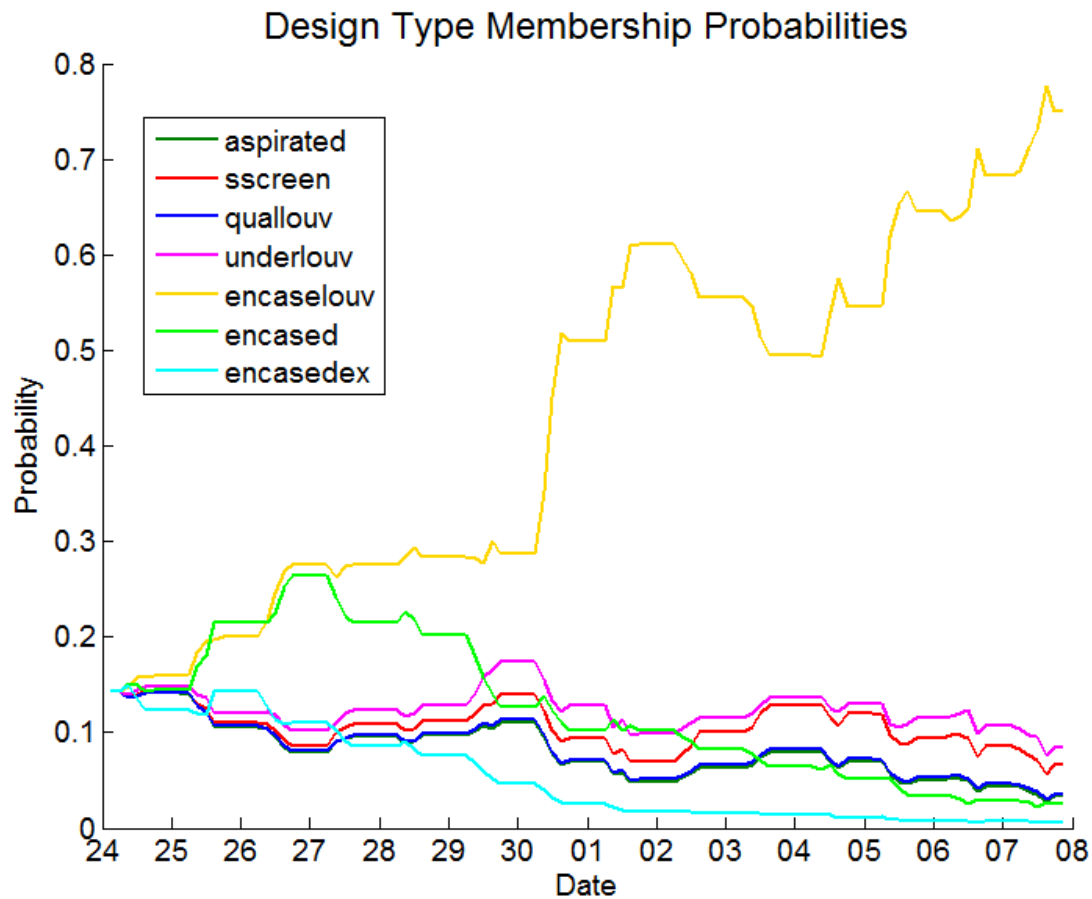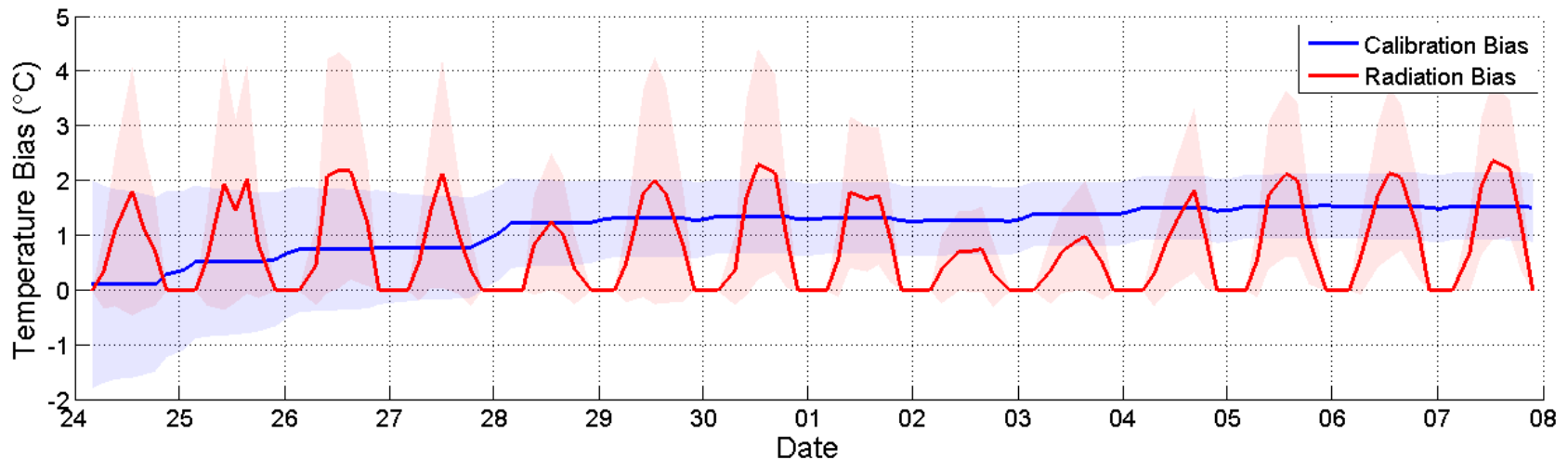
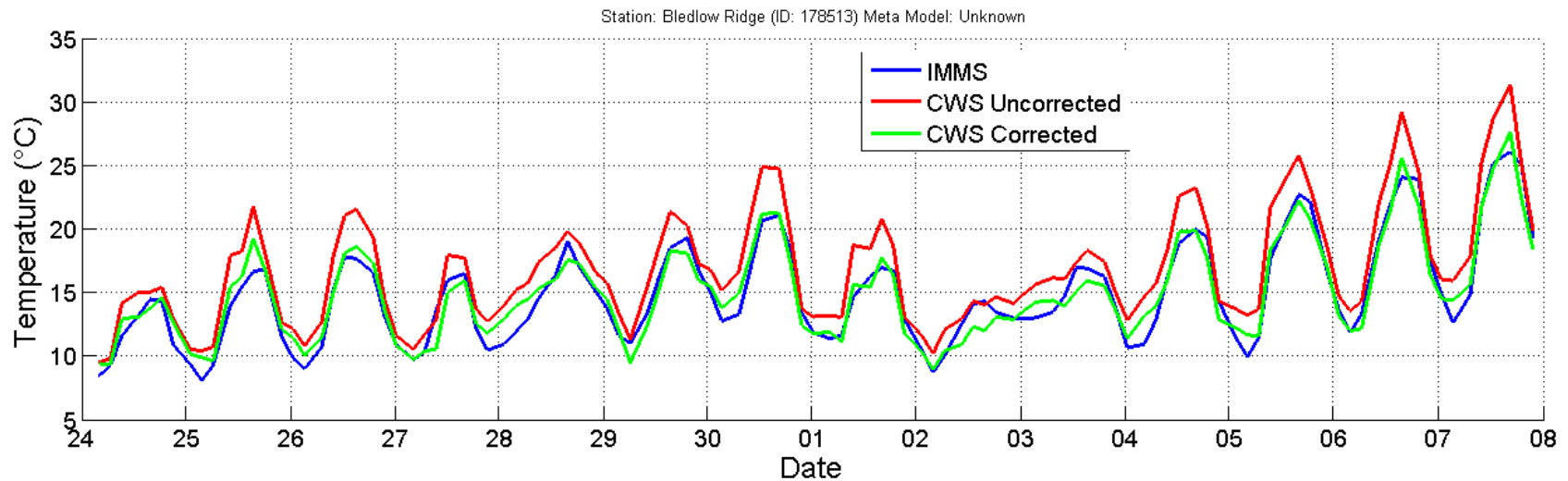Stations (ordered by elev)

Fixing everything is not going to happen
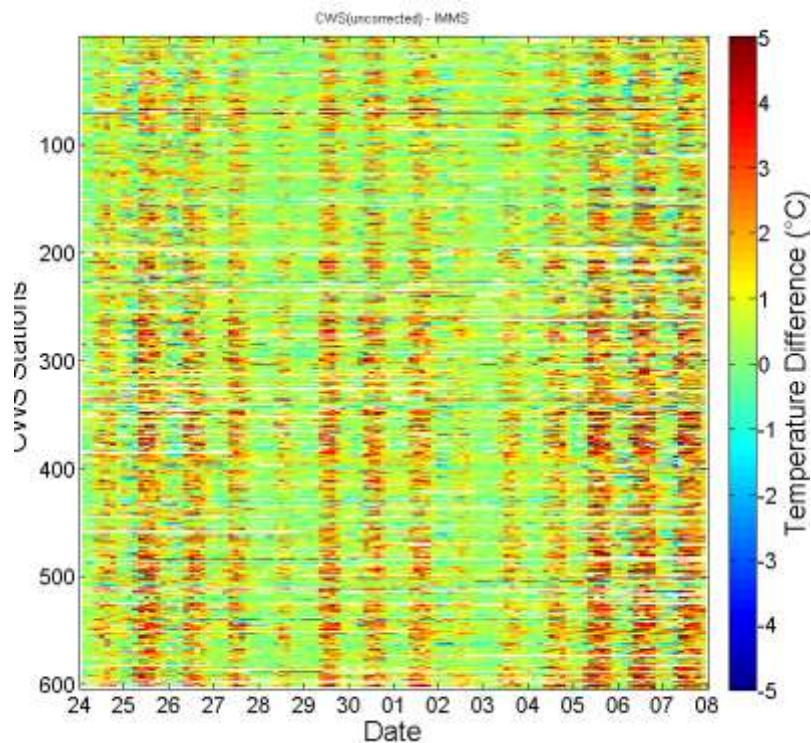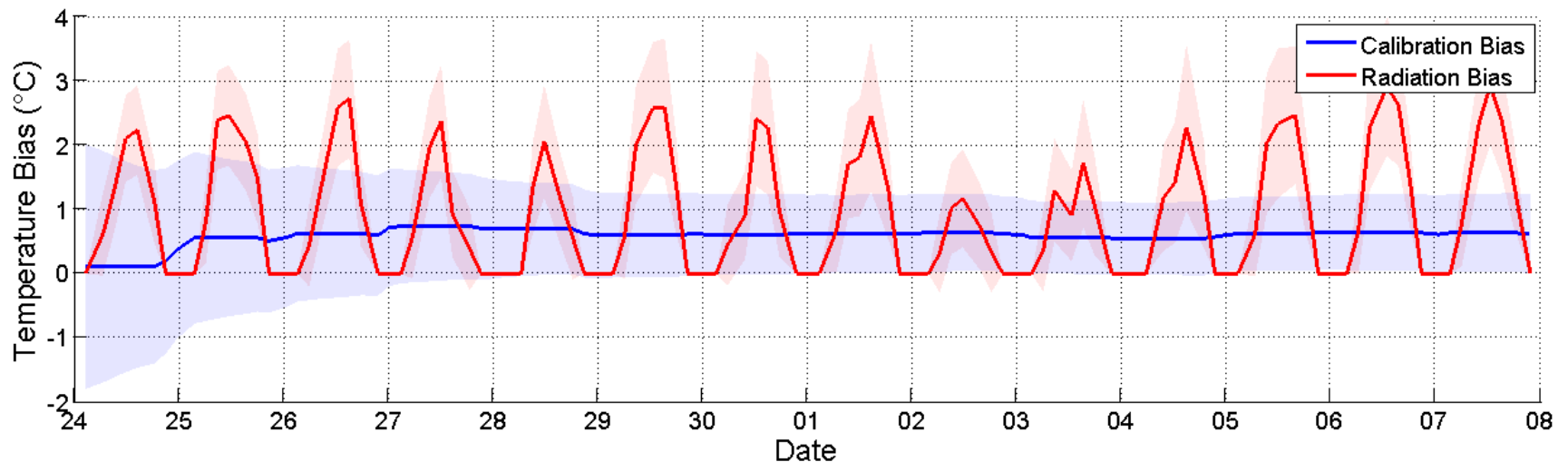
Do all the hard work and ...

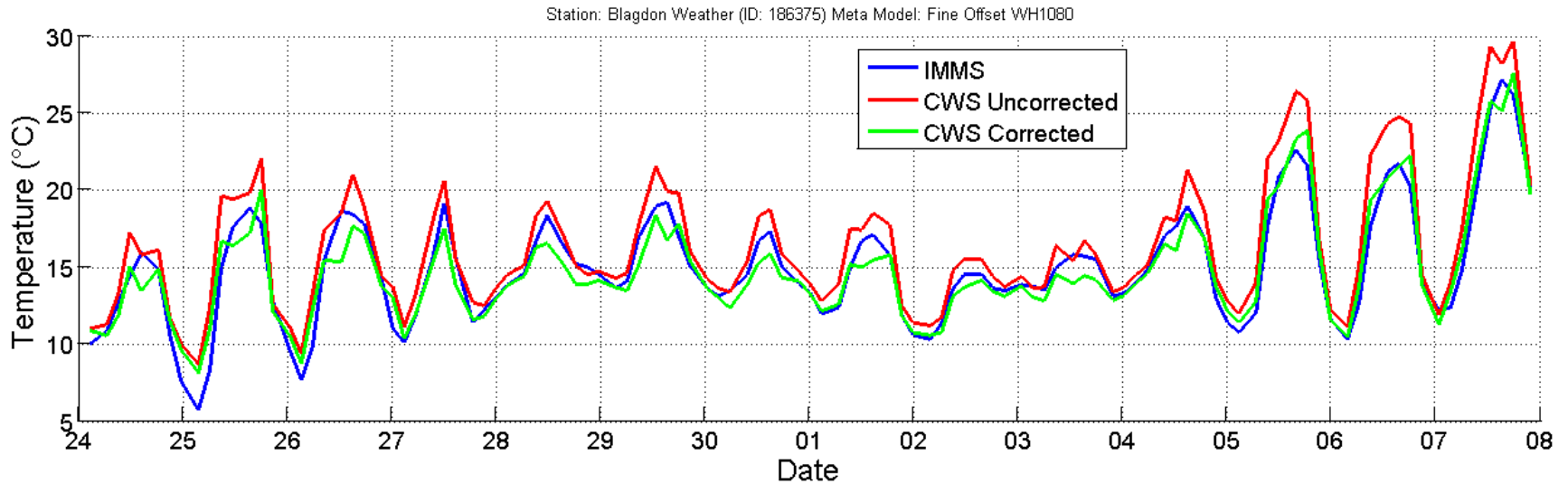# But we can learn a lot

Design Type Membership Probabilities

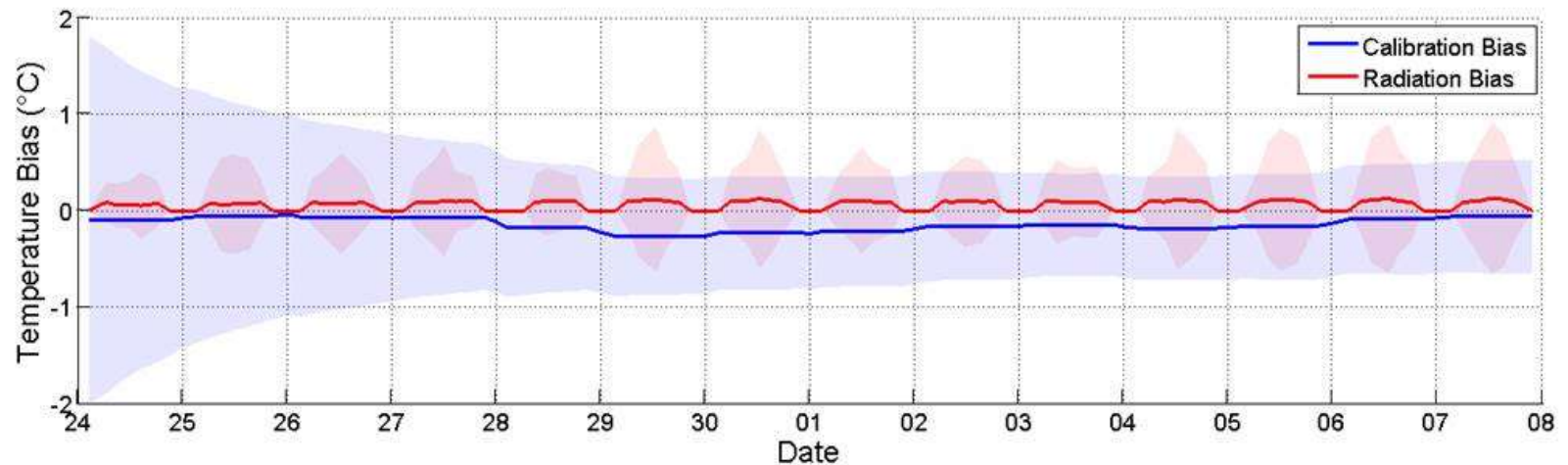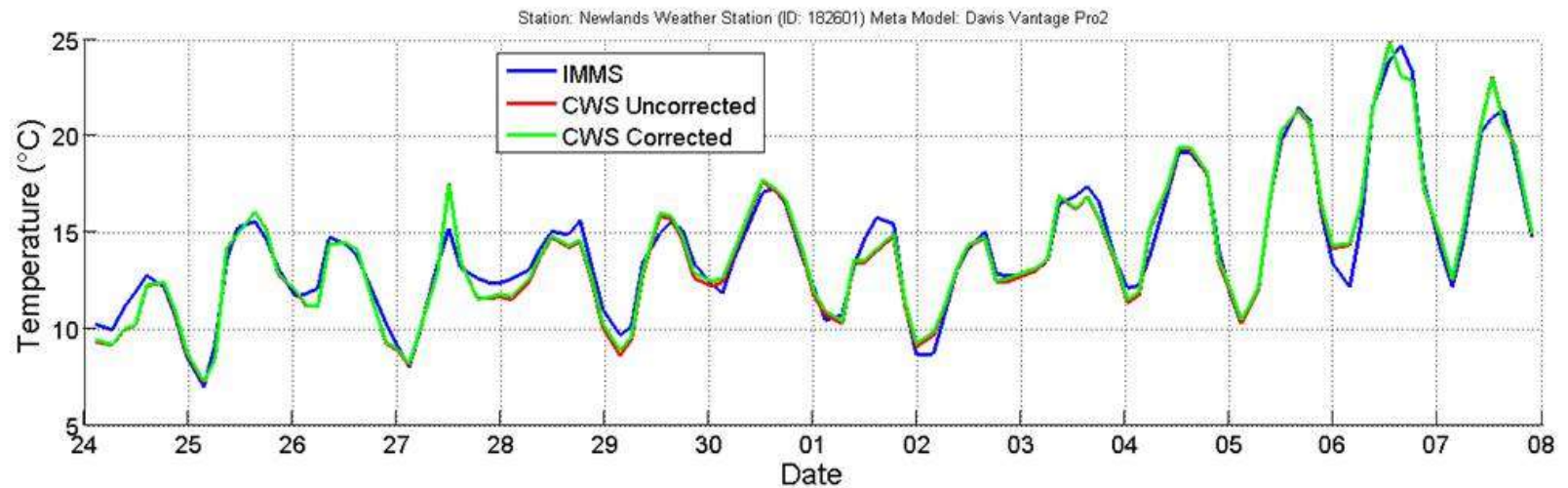# Mitigating some 'nasty' issues

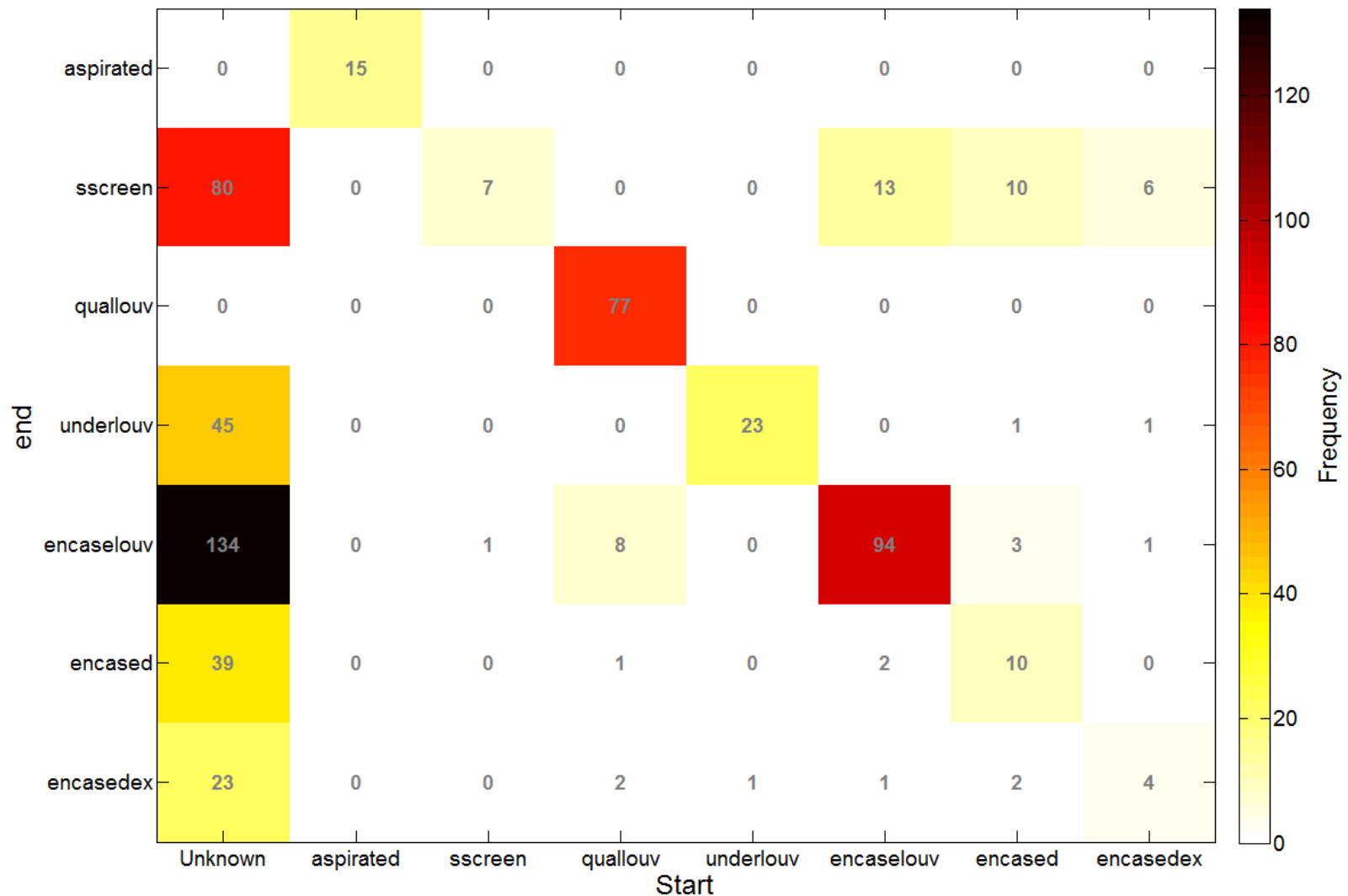Station: Bledlow Ridge (ID: 178513) Meta Model: Unknown

Or simply confirm what we see

Even for the 'best' stations

Simple models, but significant insight

Summary

- Don't try and do a purely graphical talk!
- Be clear on your aims …
- Validate probabilistic models probabilistically
  - if you care!
- Be prepared to spend 80% of your time exploring the mess that is your data
- I'd strongly prefer simple (and fast to run) models
  - allows prior sensitivity, model exploration and cross validation (without waiting)
- Be clear what is 'good enough'
  - your model will never be perfect on real data – remember it is a model
  - I've never seen real data that actually comes from a Gaussian process

- Thanks!