Berkeley Earth Methodology

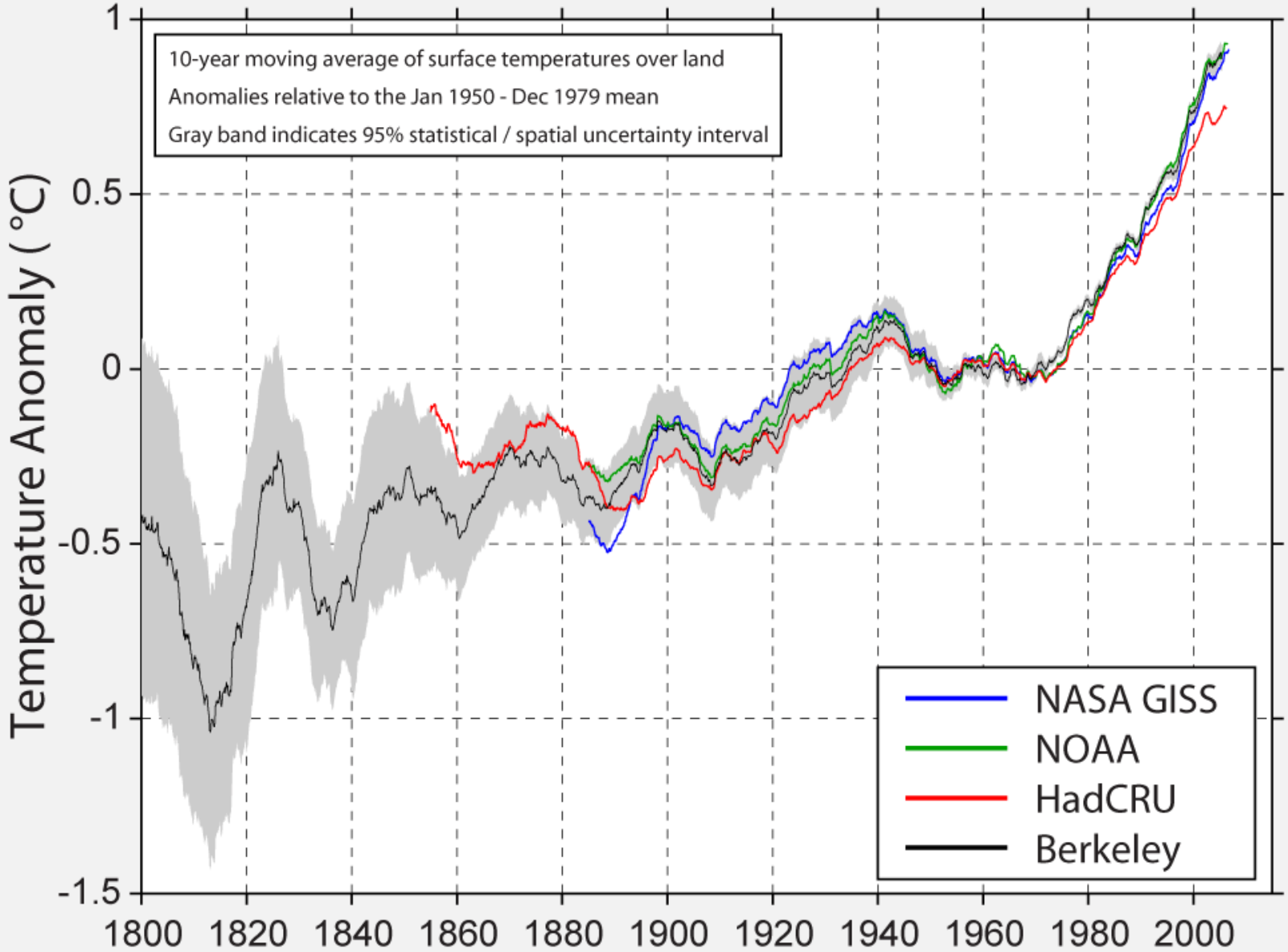
Robert Rohde, Judith Curry, Don Groom, Robert Jacobsen, Richard Muller (chair), Saul Perlmutter, Arthur Rosenfeld, Charlotte Wickham, Jonathan Wurtele

Third Santa Fe Conference on Regional and Global Climate Change

11/1/2011

www.BerkeleyEarth.org

Decadal Land-Surface Average Temperature



Innovations

- More Data
- Using Kriging for spatial relationships
- “Scalpel” as a tool for bias handling
- Simultaneous baseline determination
- Integrated outlier and reliability assessment
- Top-down approach to uncertainty quantification

Database

Monthly

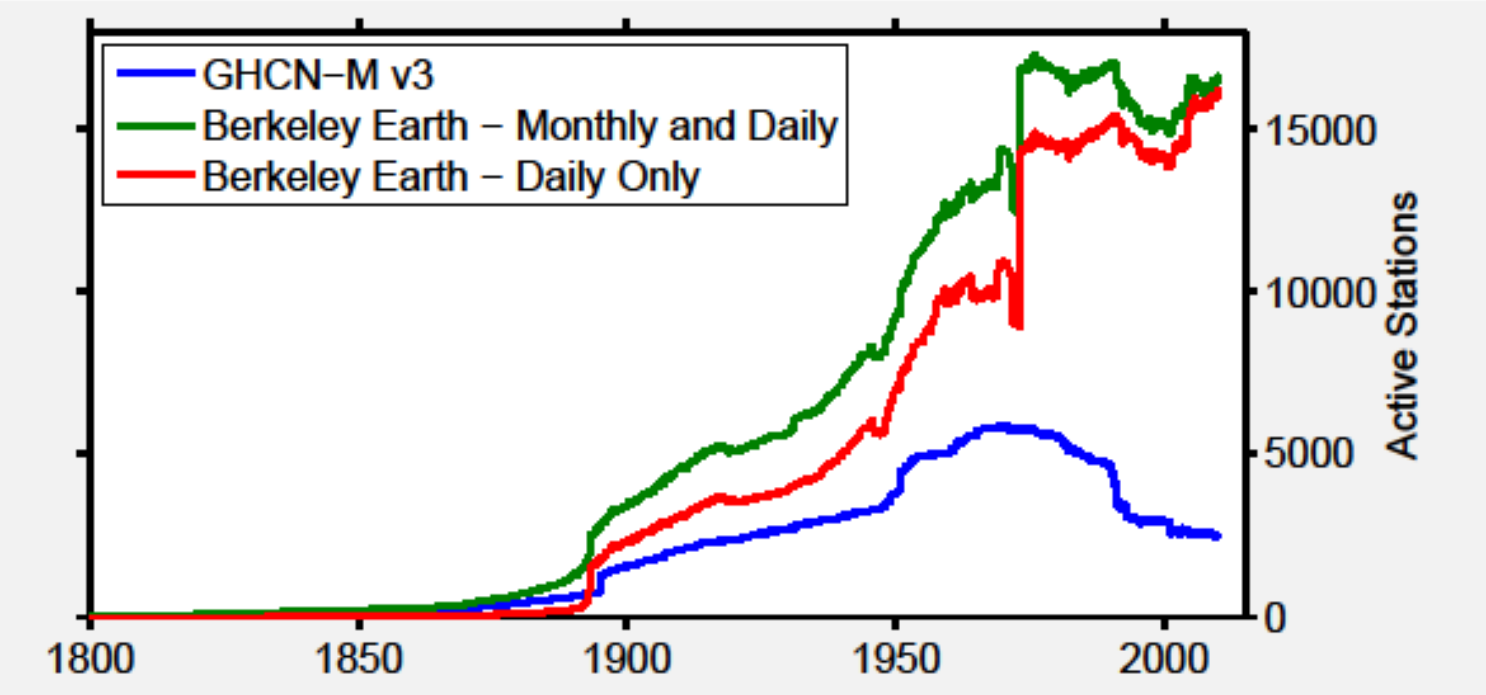
- Global Historical Climatology Network (GHCN 3)
- US Historical Climatology Network (USHCN)
- Monthly Climatic Data of the World (MCDW)
- World Monthly Surface Station Climatology (WMSSC)
- HadCRU Data Set Release (HadCRU)
- US Summary of the Month (USSOM)
- Scientific Committee on Antarctic Research (SCAR)
- GSN Monthly Archive (GSNMON)
- GCOS monthly summaries created by DWD (GCOS)
- World Weather Records (WWR)
- Colonial Era Weather Archive (CE)

Daily

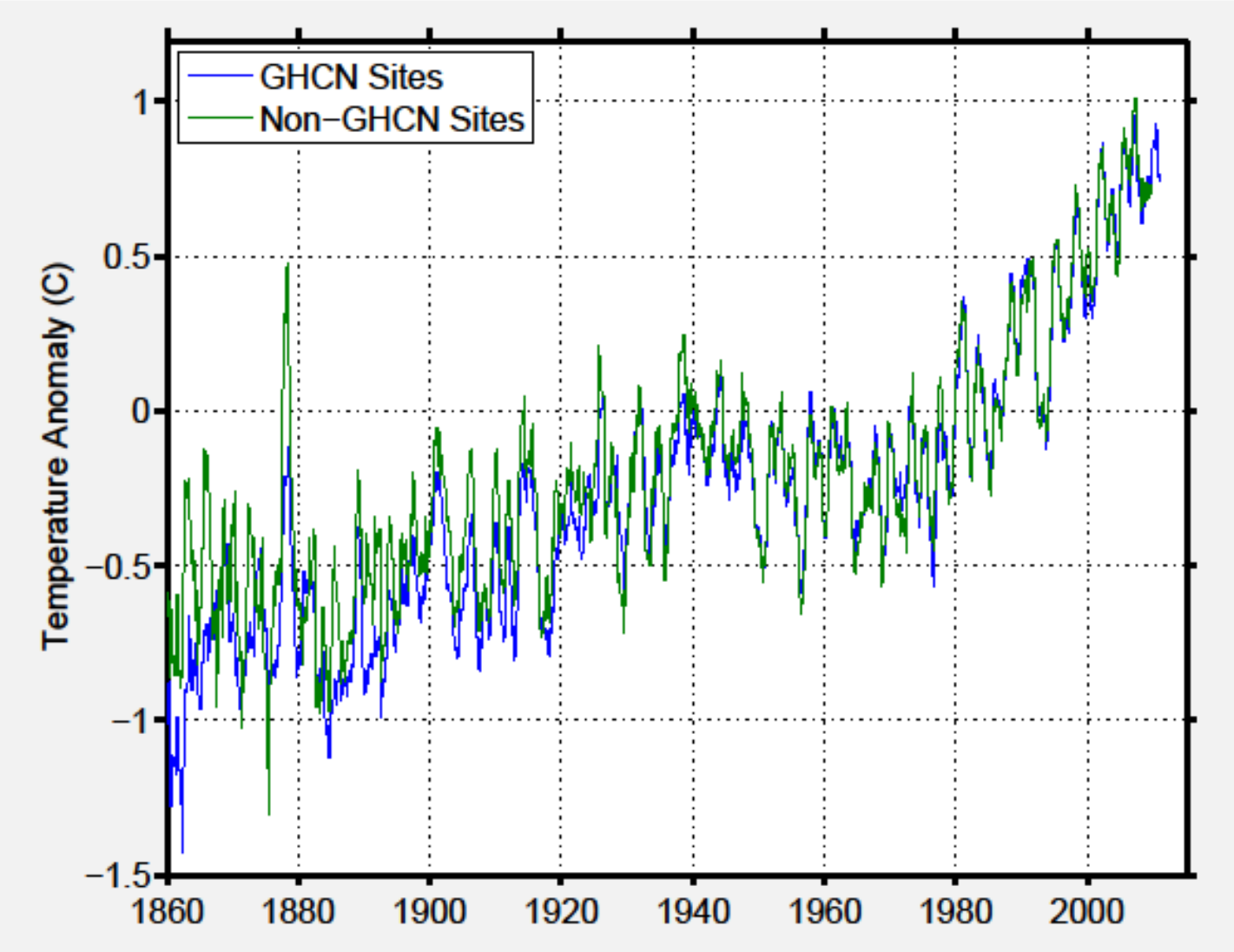
- GHCN - Daily (GHCN-D)
- US Cooperative Summary of the Day (USSOD-C)
- US First Order Summary of the Day (USSOD-FO)
- Global Summary of the Day (GSOD)

Database

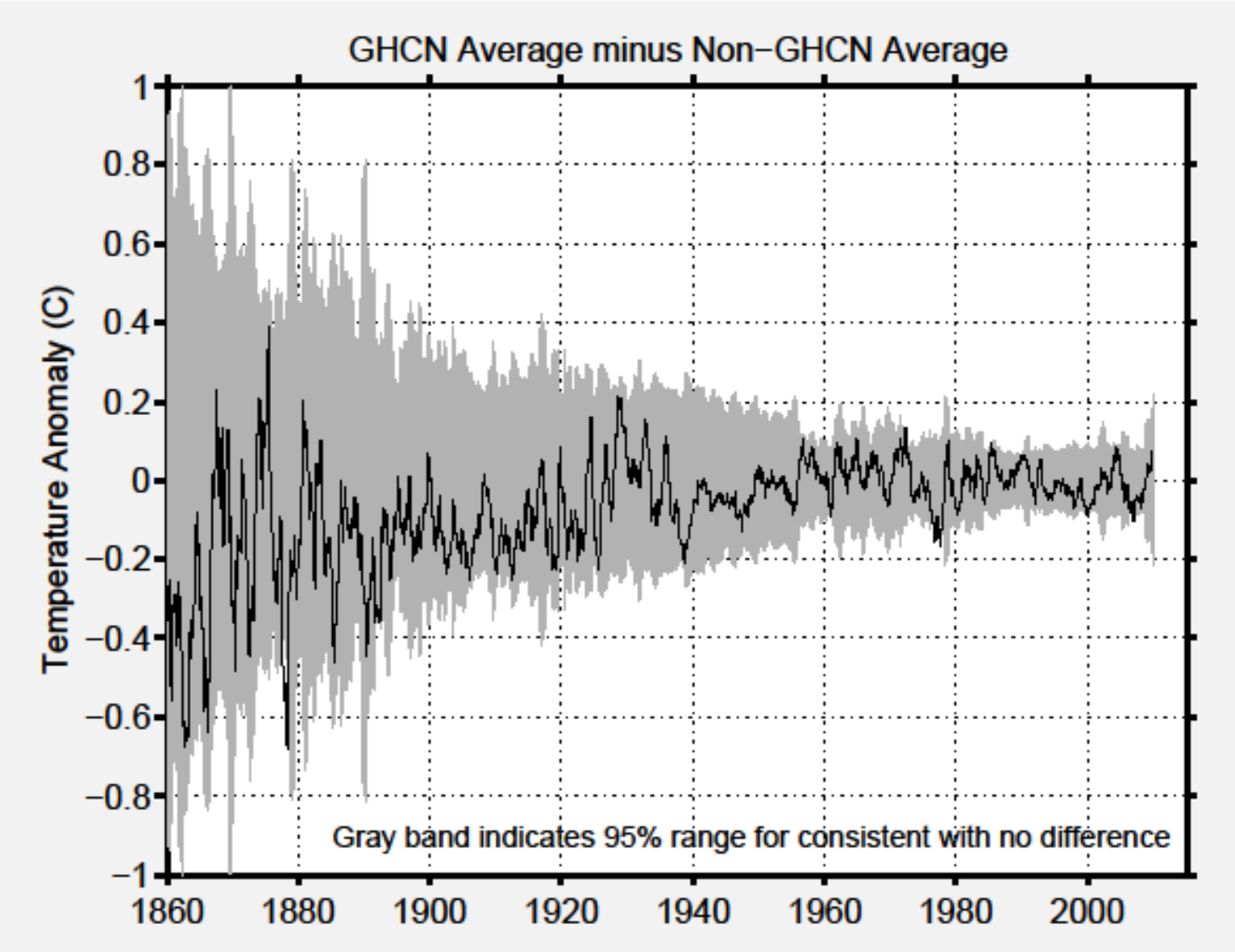
Dataset	Berkeley Earth	GHCN-M v3
Number of Stations	39,028	7,280
Monthly Observations	14,786,160	5,150,496
Median Record Length	25.8 Years	58.5 Years



Analysis of GHCN to Non-GHCN sites separately



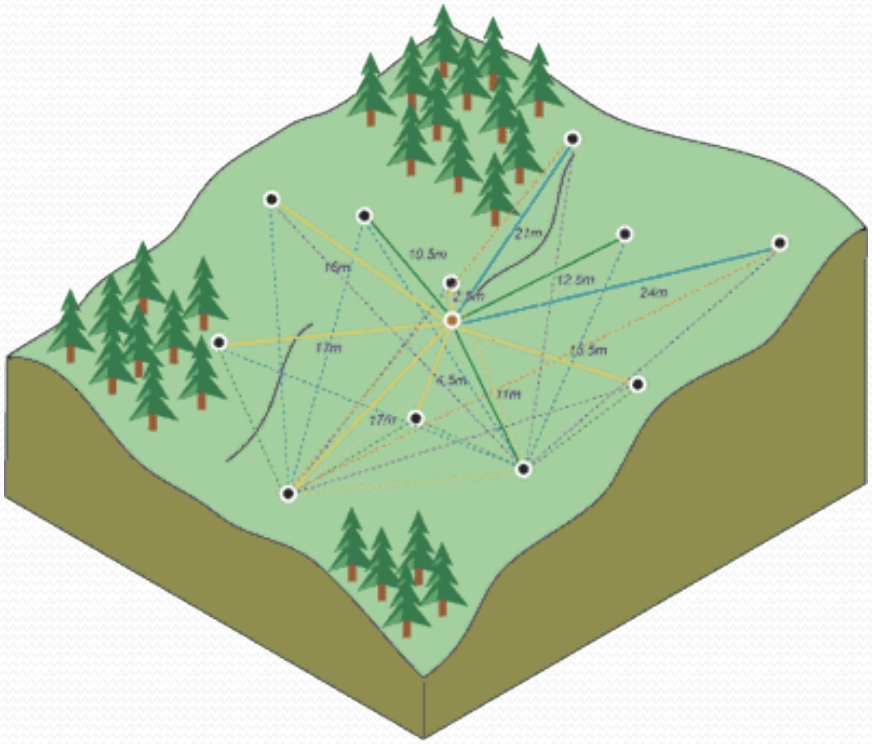
Analysis of GHCN to Non-GHCN sites separately



Spatial Interpolation (Kriging)

- Can create an interpolated field everywhere based on discrete measurements
- Best linear unbiased estimator is known as Kriging

$$\widehat{M}(\vec{x}) = \sum_{i=1}^{N} K_i(\vec{x})\, M_i$$



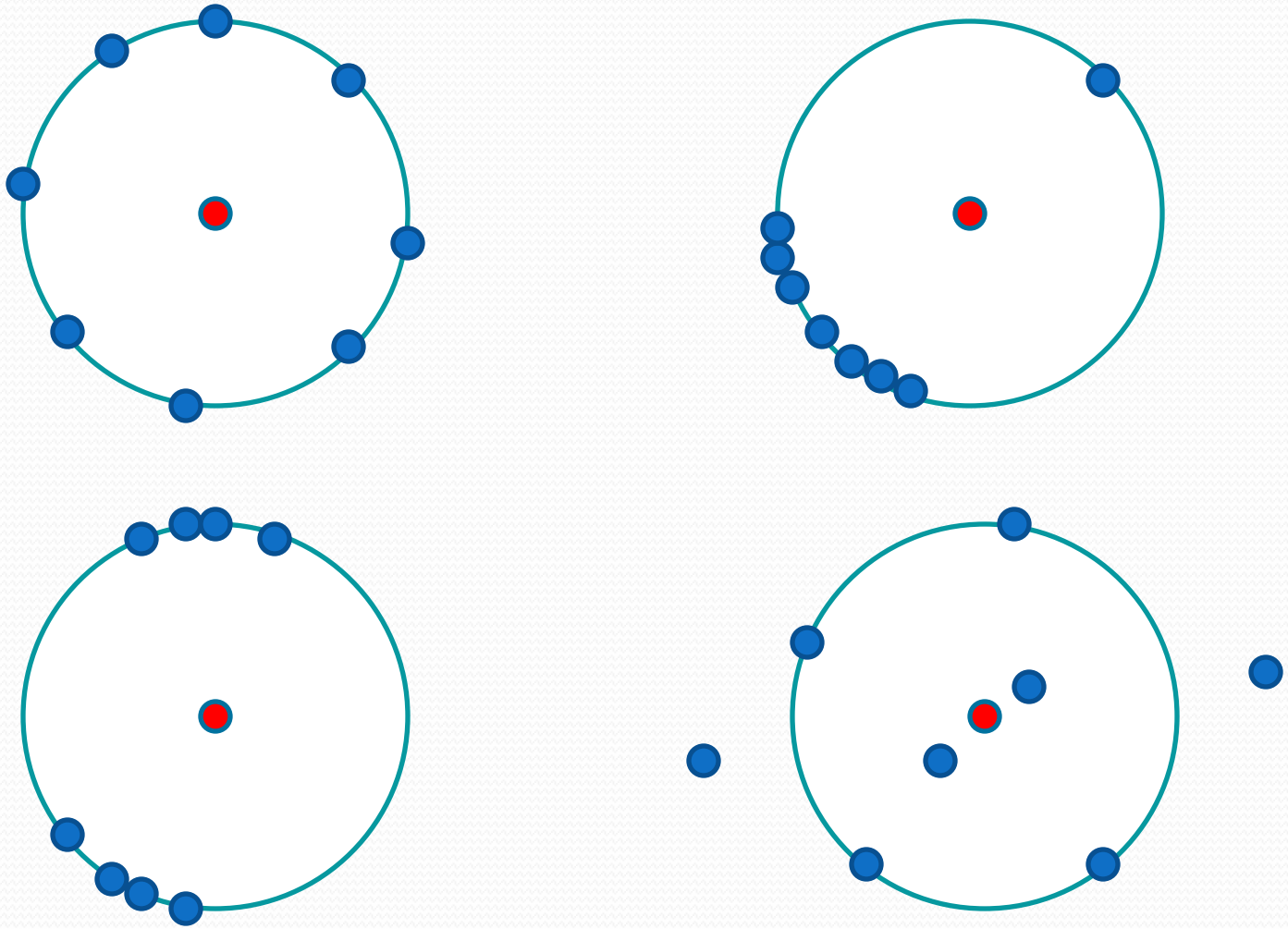
Uses Correlation Information

$$\widehat{M}(\vec{x}) = \sum_{i=1}^{N} K_i(\vec{x})\, M_i$$

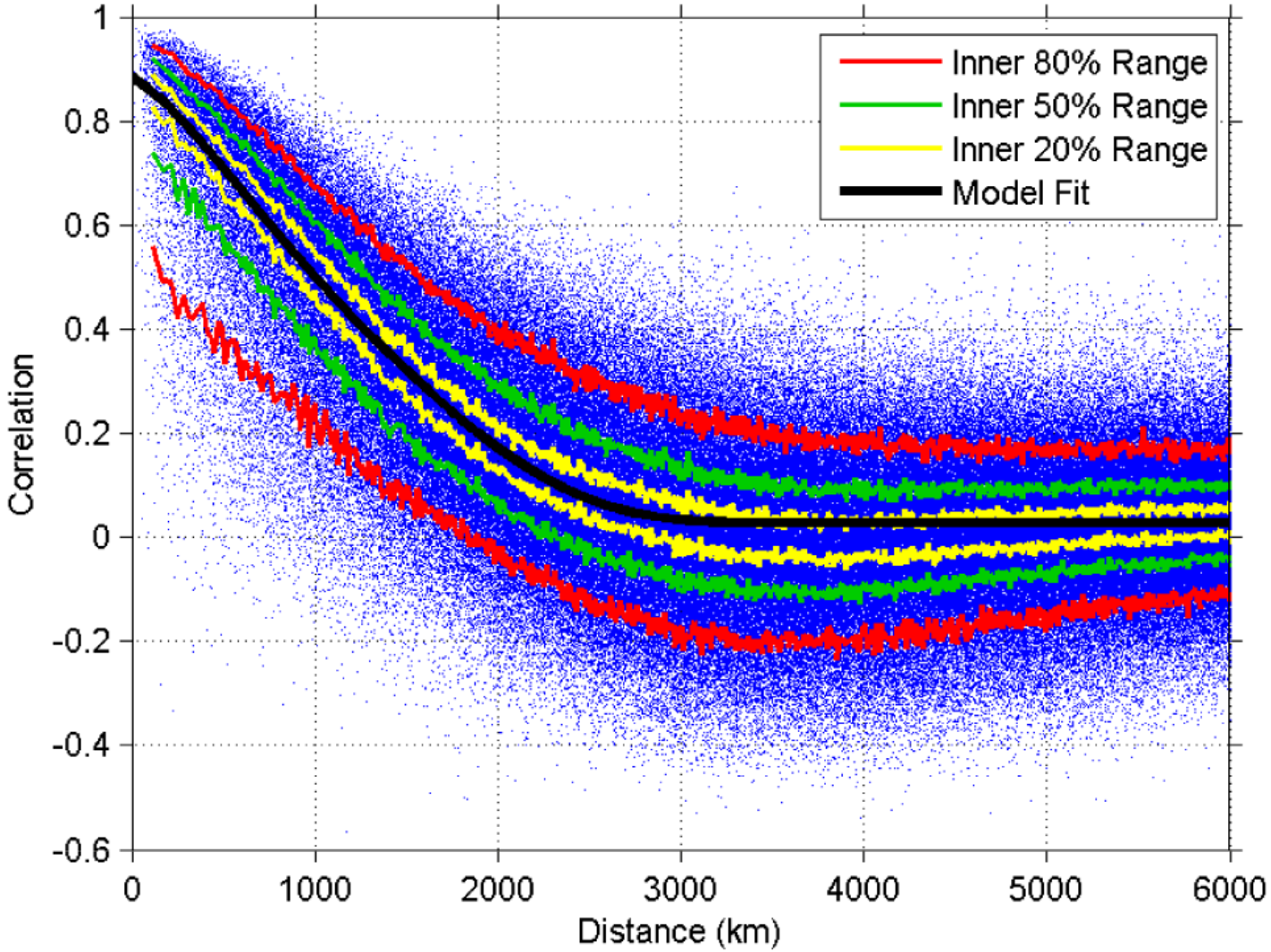
$$\begin{pmatrix} K_1(\vec{x}) \\ \vdots \\ K_N(\vec{x}) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \mathrm{Cov}(\vec{x}_1,\vec{x}_2) & \cdots & \mathrm{Cov}(\vec{x}_1,\vec{x}_N) \\ \mathrm{Cov}(\vec{x}_2,\vec{x}_1) & \sigma_2^2 & & \mathrm{Cov}(\vec{x}_2,\vec{x}_N) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(\vec{x}_N,\vec{x}_1) & \mathrm{Cov}(\vec{x}_N,\vec{x}_2) & \cdots & \sigma_N^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{Cov}(\vec{x},\vec{x}_1) \\ \vdots \\ \mathrm{Cov}(\vec{x},\vec{x}_N) \end{pmatrix}$$

$$\approx \begin{pmatrix} 1 & \mathrm{R}(\vec{x}_1,\vec{x}_2) & \cdots & \mathrm{R}(\vec{x}_1,\vec{x}_N) \\ \mathrm{R}(\vec{x}_2,\vec{x}_1) & 1 & & \mathrm{R}(\vec{x}_2,\vec{x}_N) \\ \vdots & & \ddots & \vdots \\ \mathrm{R}(\vec{x}_N,\vec{x}_1) & \mathrm{R}(\vec{x}_N,\vec{x}_2) & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{R}(\vec{x},\vec{x}_1) \\ \vdots \\ \mathrm{R}(\vec{x},\vec{x}_N) \end{pmatrix}$$

Kriging captures spatial relationships



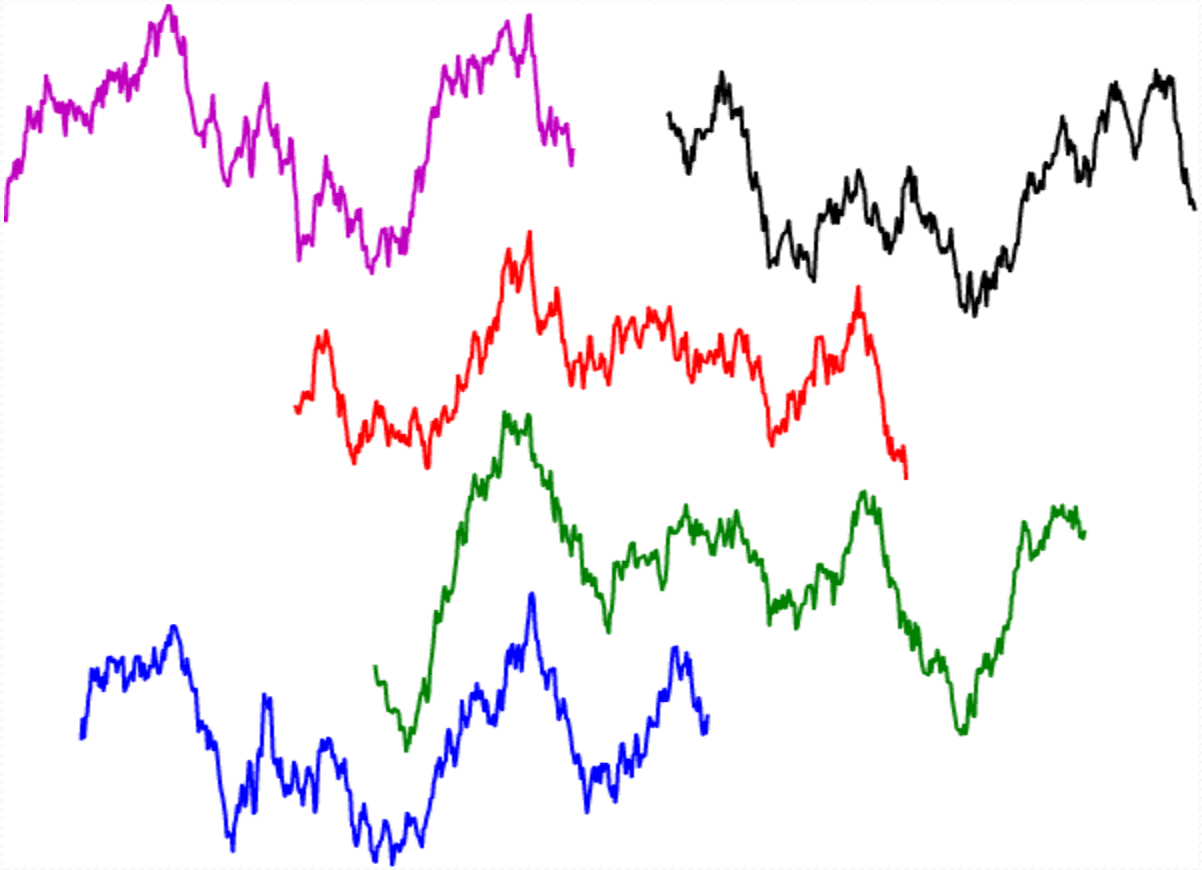
Uses Correlation Information



500,000 pairwise comparisons

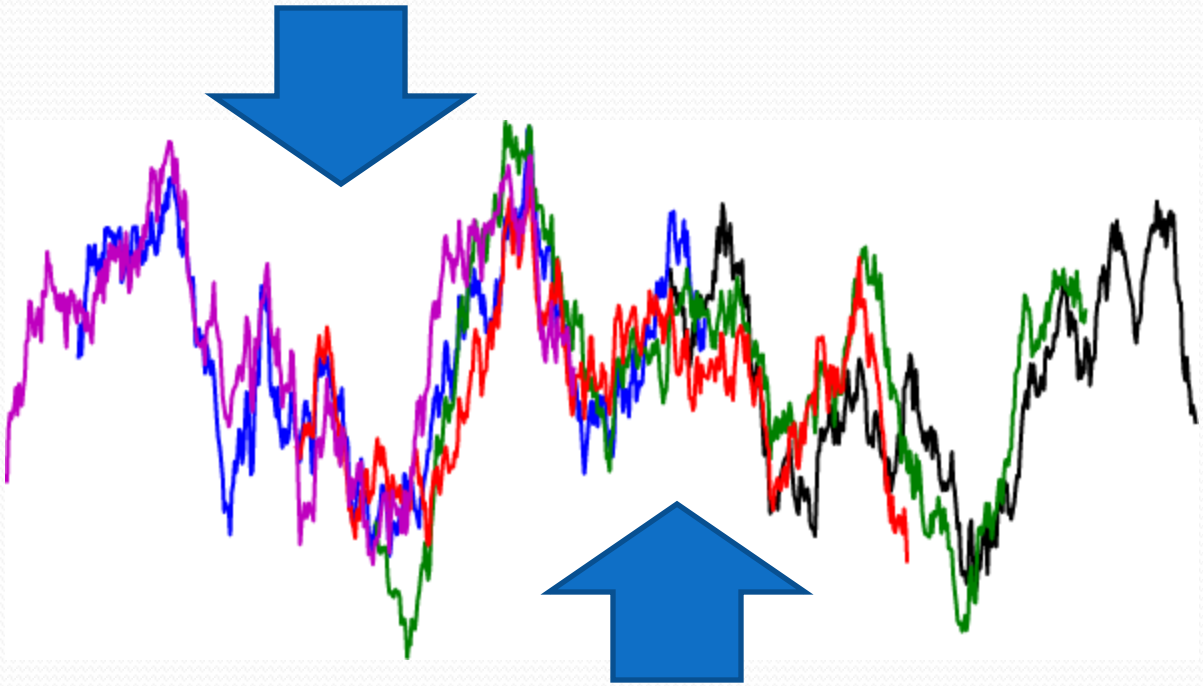
Assembling the Record

Local factors such as latitude and elevation have a very large effect on the mean climate at each site



Assembling the Record

Local factors such as latitude and elevation have a very large effect on the mean climate at each site. In order to detect common long-term behavior we need to compensate for these effects.



The “Scalpel”

- Rather than “correcting” bias shifts, our default position is to split the record into multiple pieces at each inhomogeneity.
- These fragments are assembled along with all other data in the normal way, and that assembly determines the relative alignments.

Reasons for Breaks

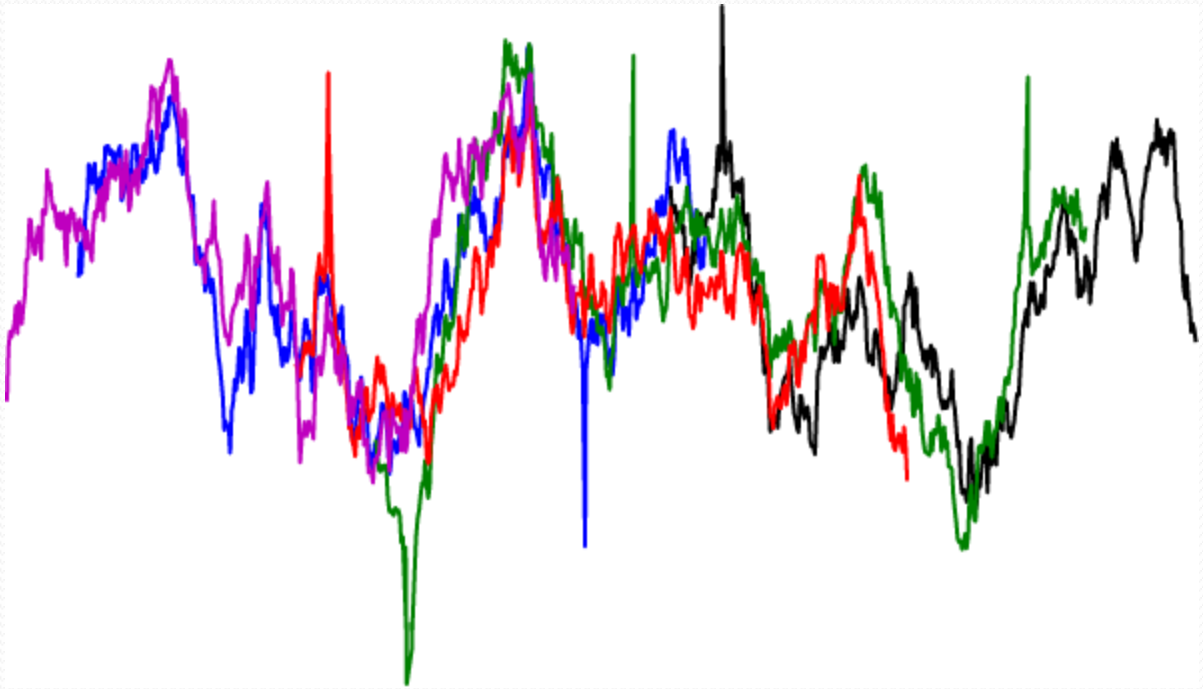
- Record gaps (> 1 year)
- Station Moves
- Station Time of Observation Changes
- Apparent changes in baseline greater than 4 standard deviations (as compared with local neighbors)

Scalpel Results

- One empirical break for every ~12 years of record on average
- After scalpel the mean record fragment is about 7.1 years (median: 5.1 years)
- Whole analysis process was designed to work with small pieces.

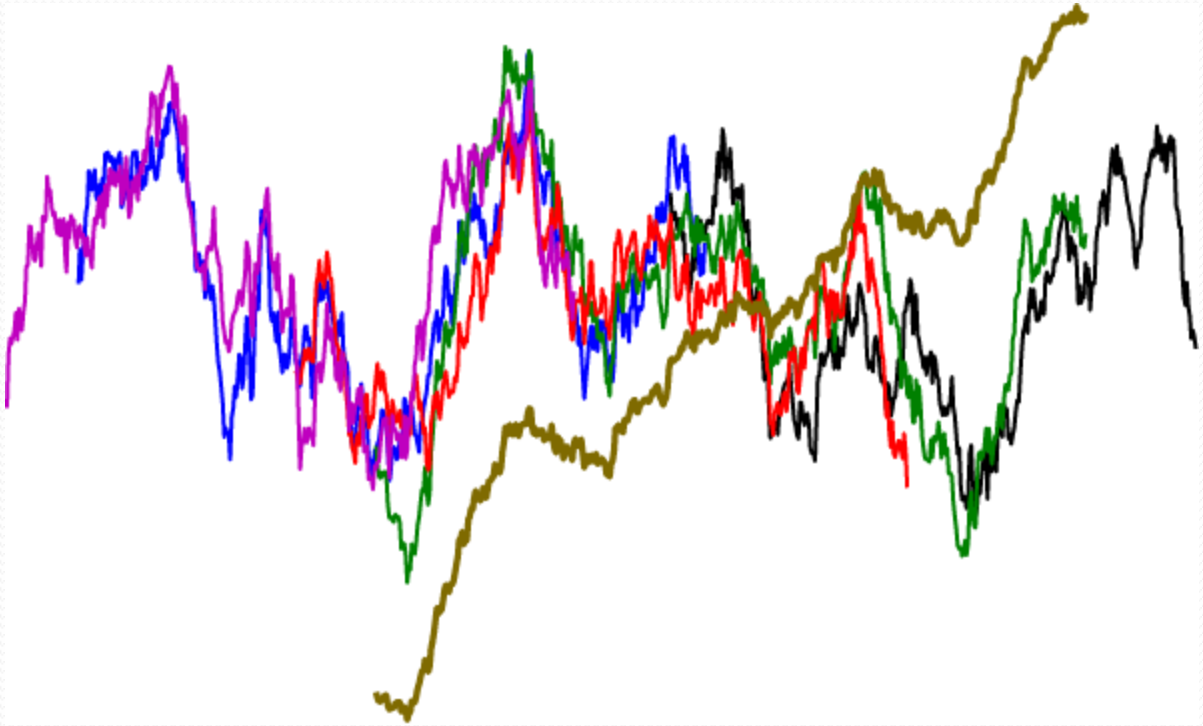
Integrated Outlier / Reliability Assessment

- Local outlier points are deweighted via an iterative process



Integrated Outlier / Reliability Assessment

- Highly anomalous series are also deweighted

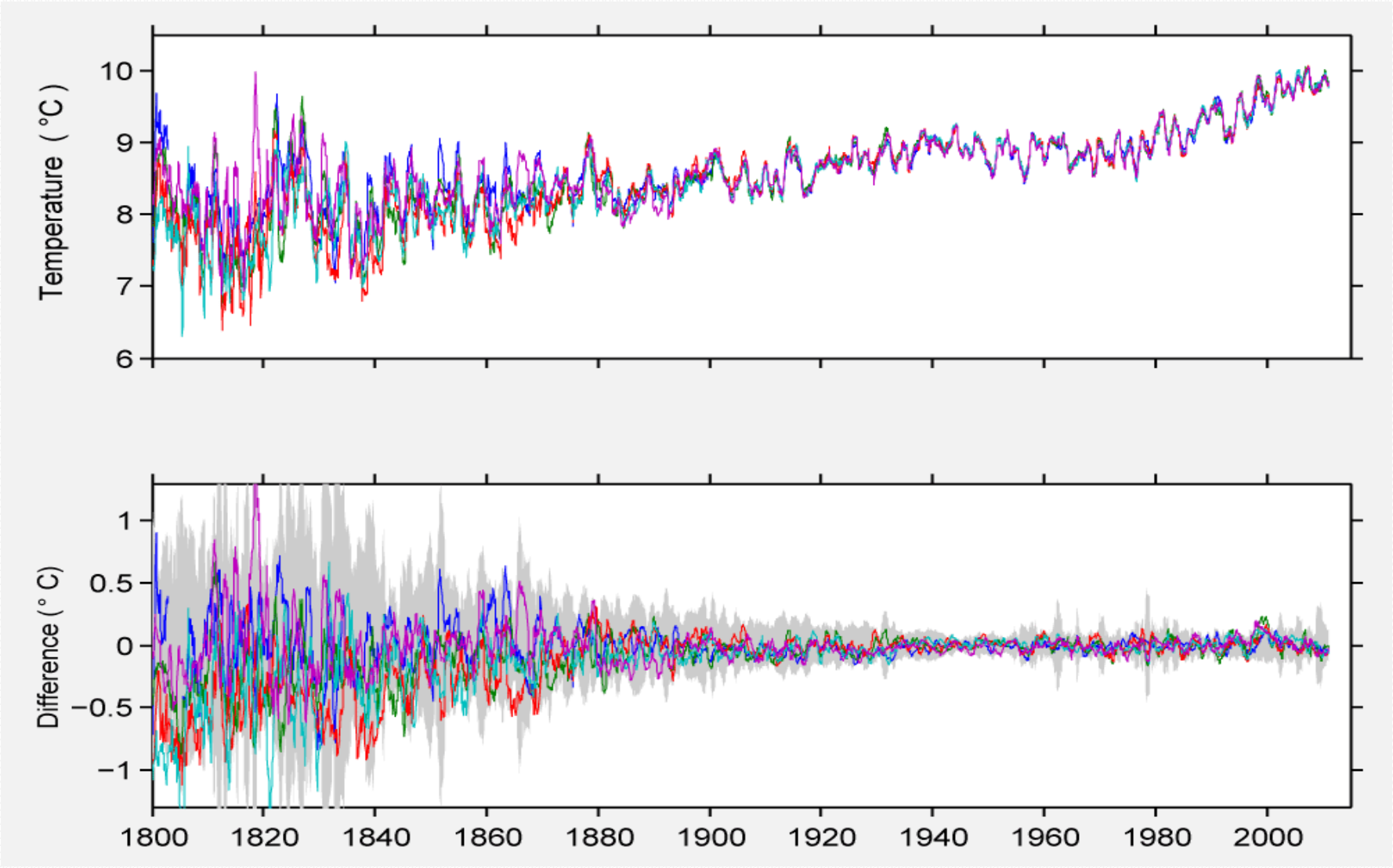


Top-down Assessment of Uncertainty

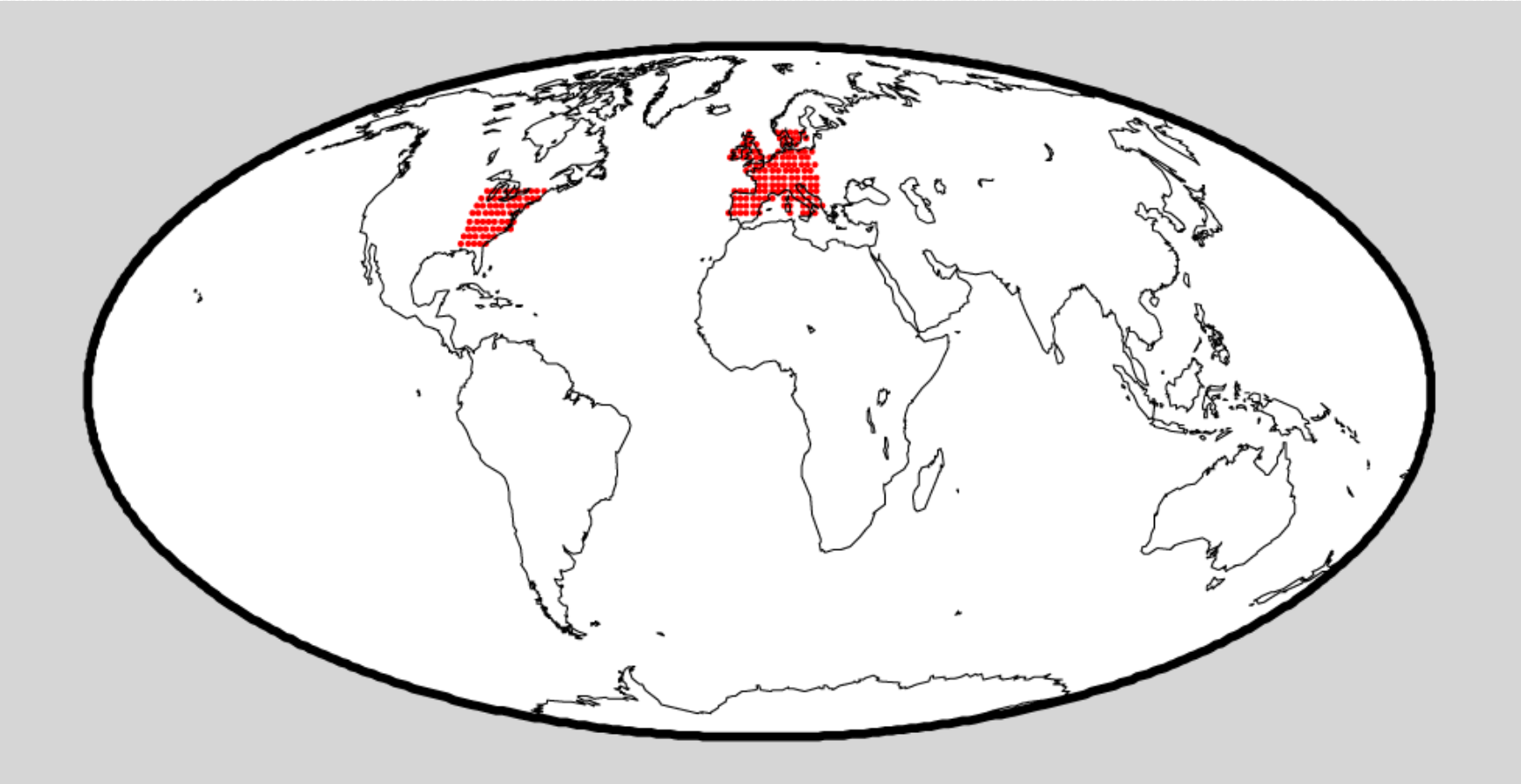
- Rather than assigning an error bar to each datum and each analysis step, we measure the statistical consistency of the data and the spatial completeness.

- How much would the result change if some data were excluded?
- How much uncertainty occurs because some regions aren't sampled?
- Look at impact of other possible biases (Urban Heating, Station Quality, …)?

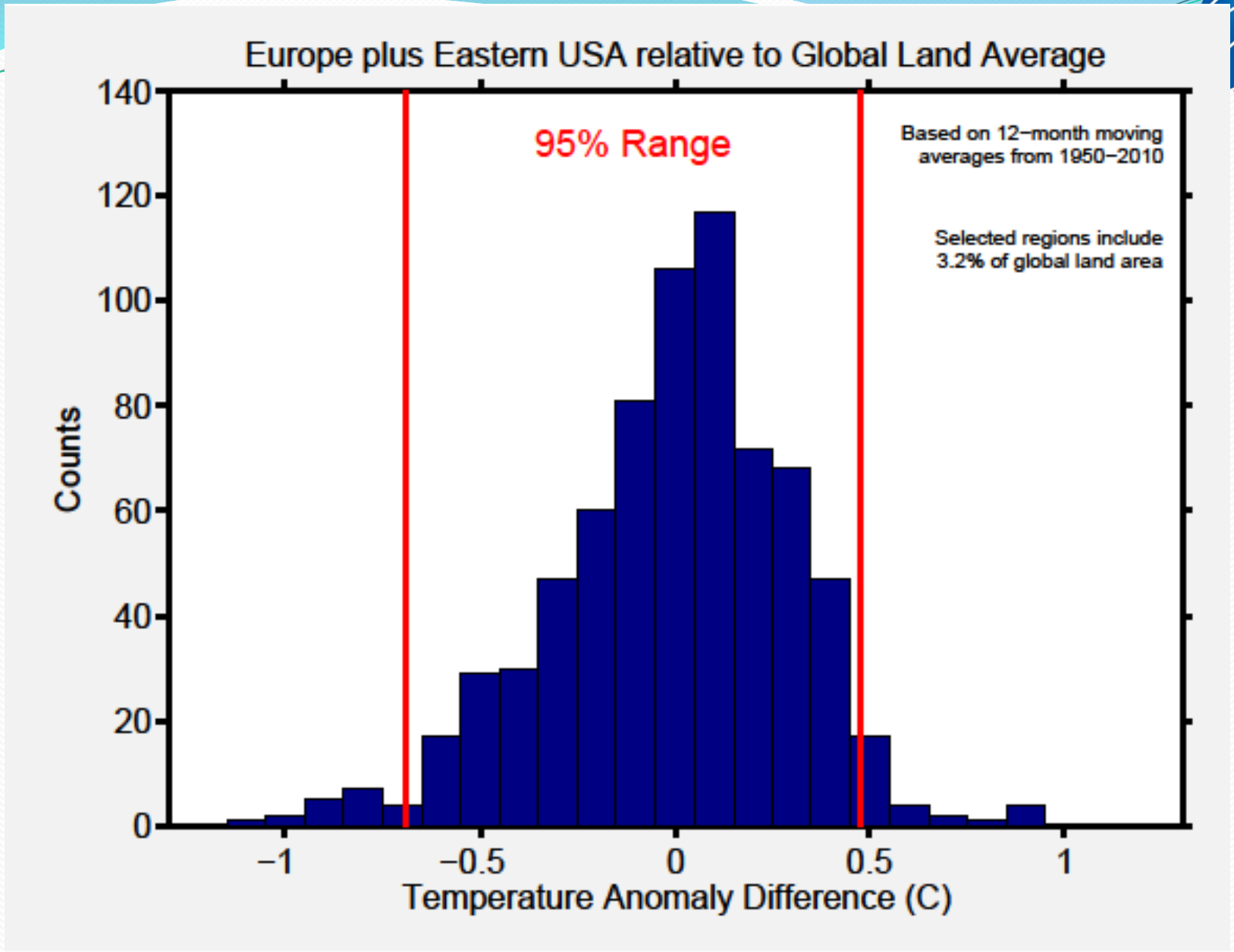
Five Independent 20% Samples



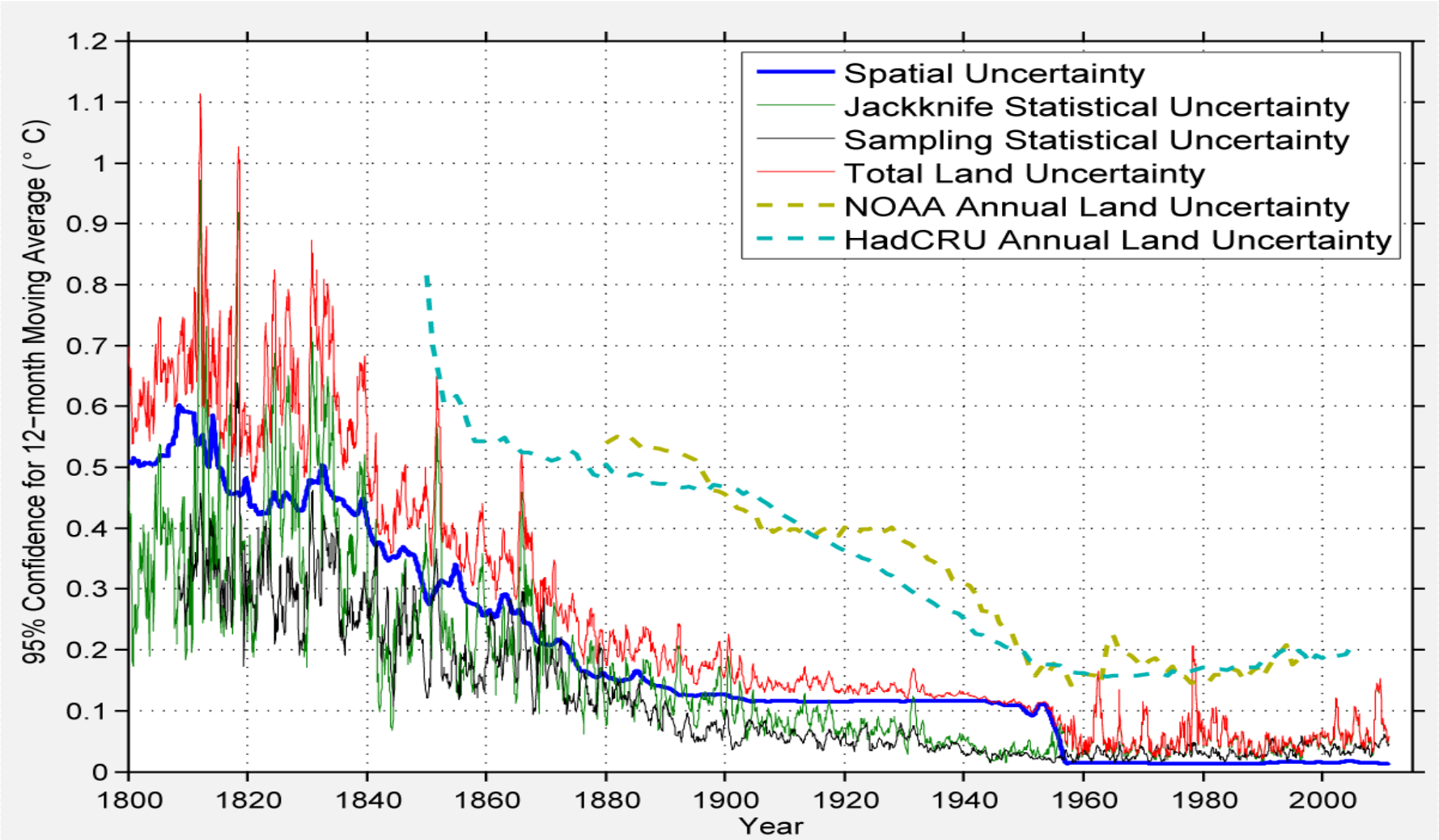
Spatial Incompleteness



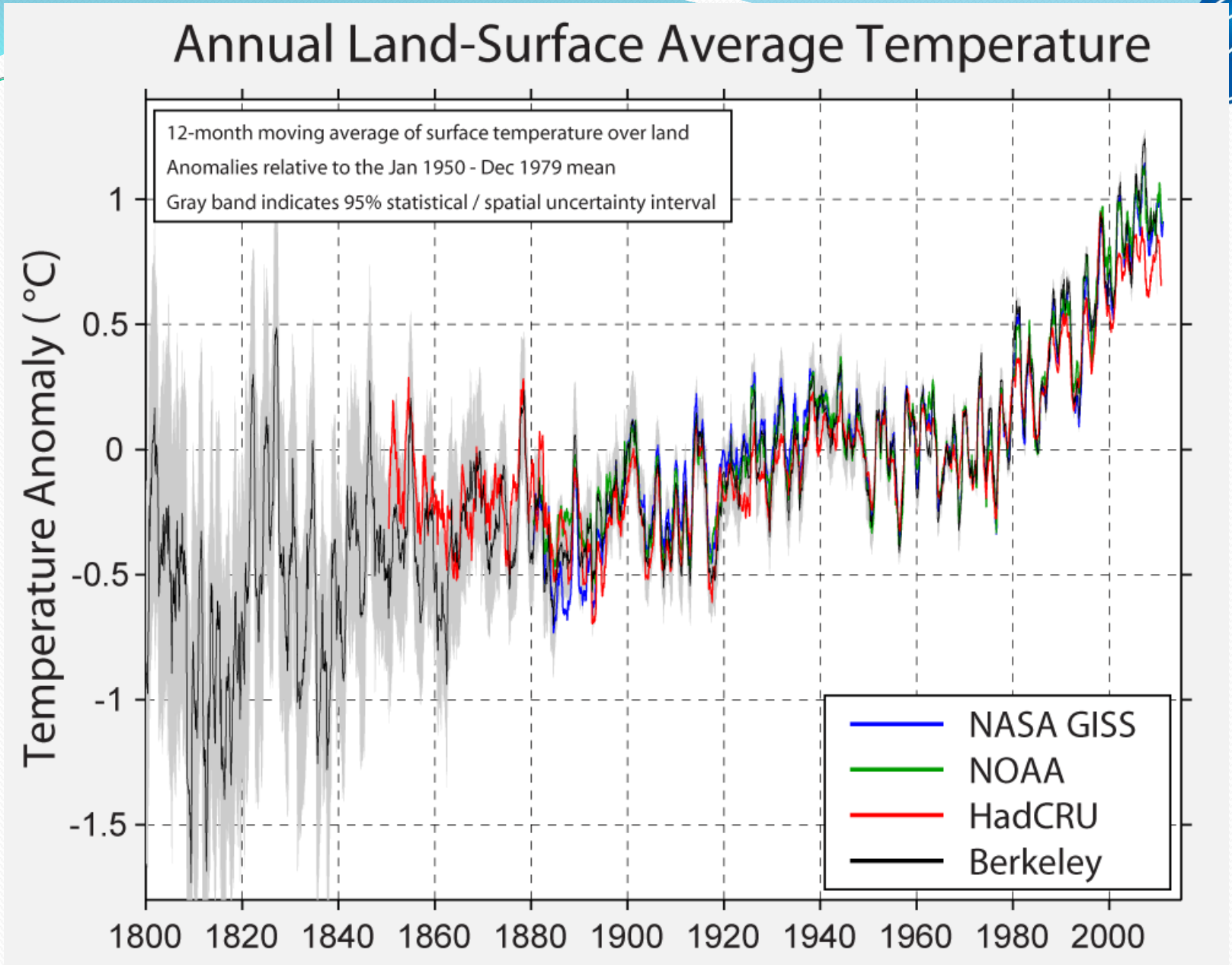
3.2% of Land Surface Highlighted


Europe plus Eastern USA relative to Global Land Average
95% Range
Based on 12−month moving averages from 1950−2010
Selected regions include 3.2% of global land area
Counts
140
120
100
80
60
40
20
0
−1
−0.5
0
0.5
1
Temperature Anomaly Difference (C)

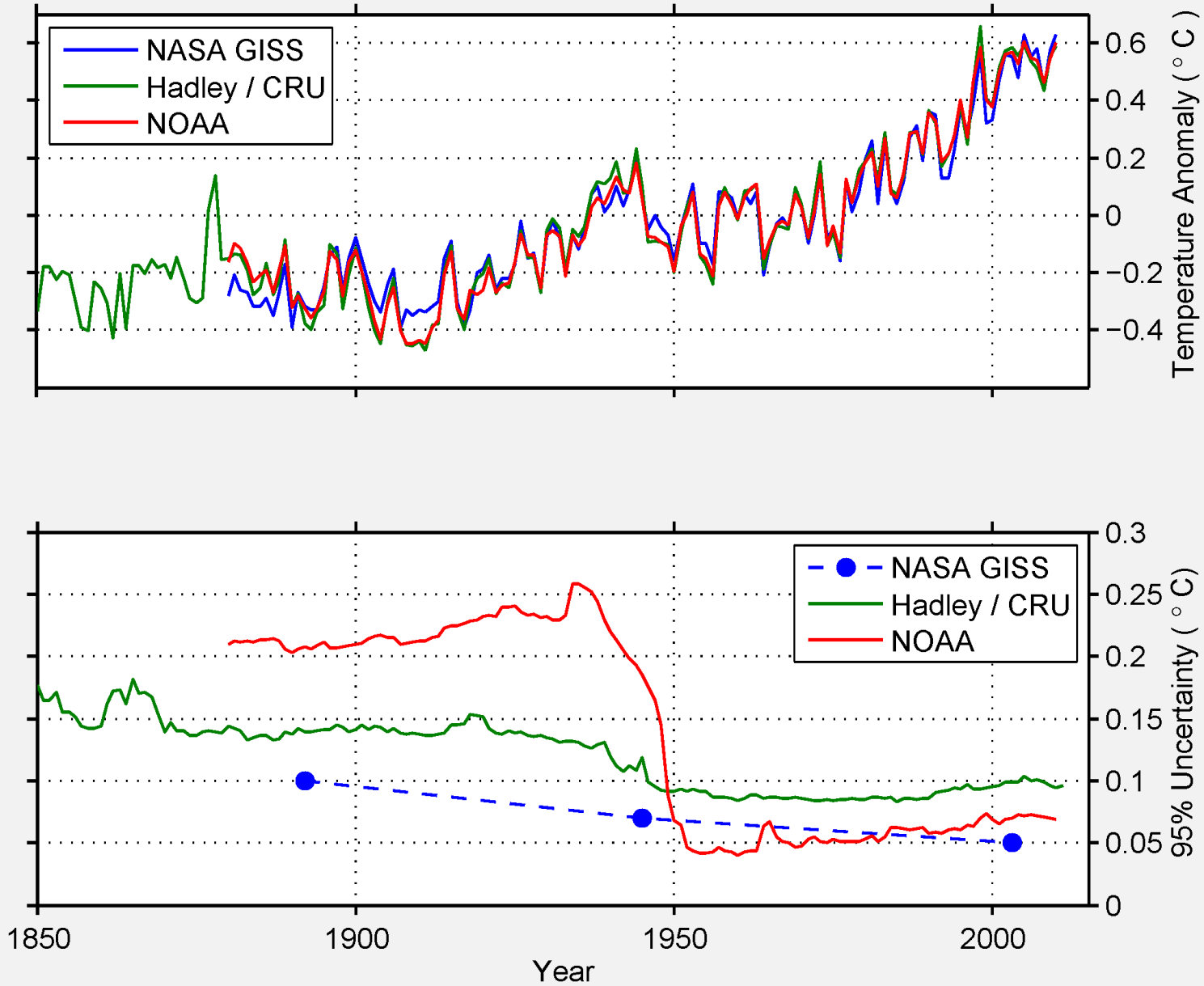
Comparison to Prior Uncertainties

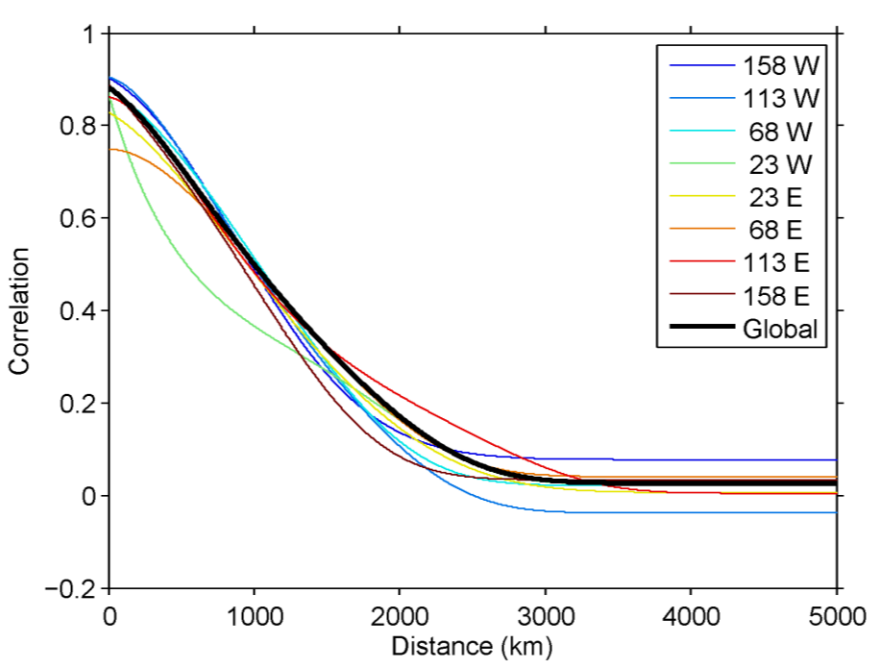


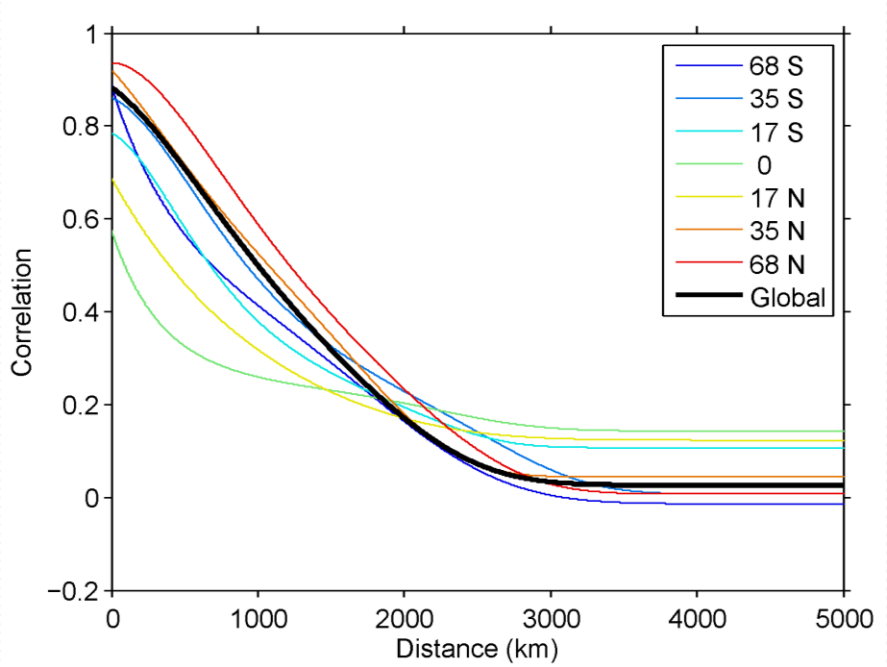
Based on Analysis of GHCN Data

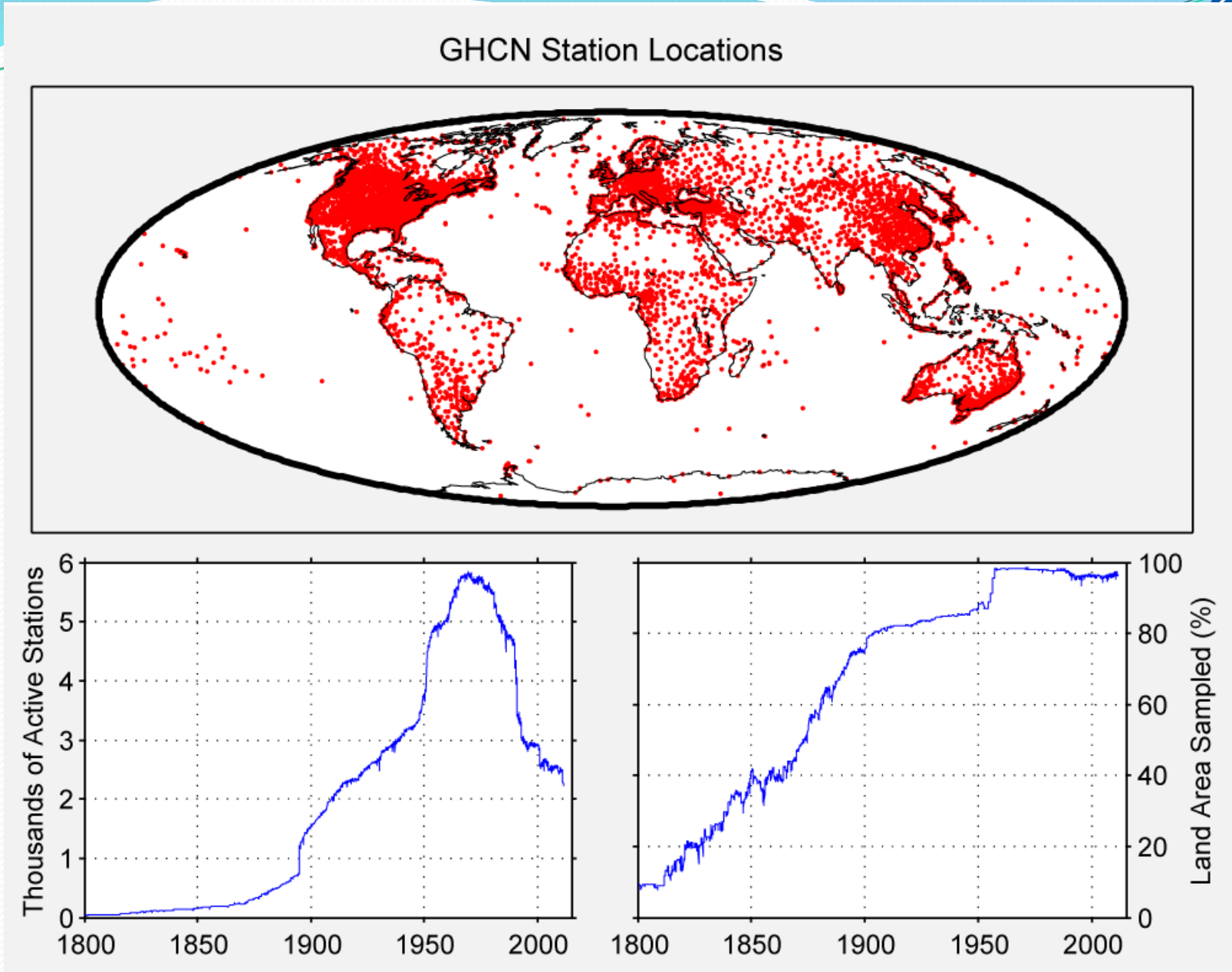


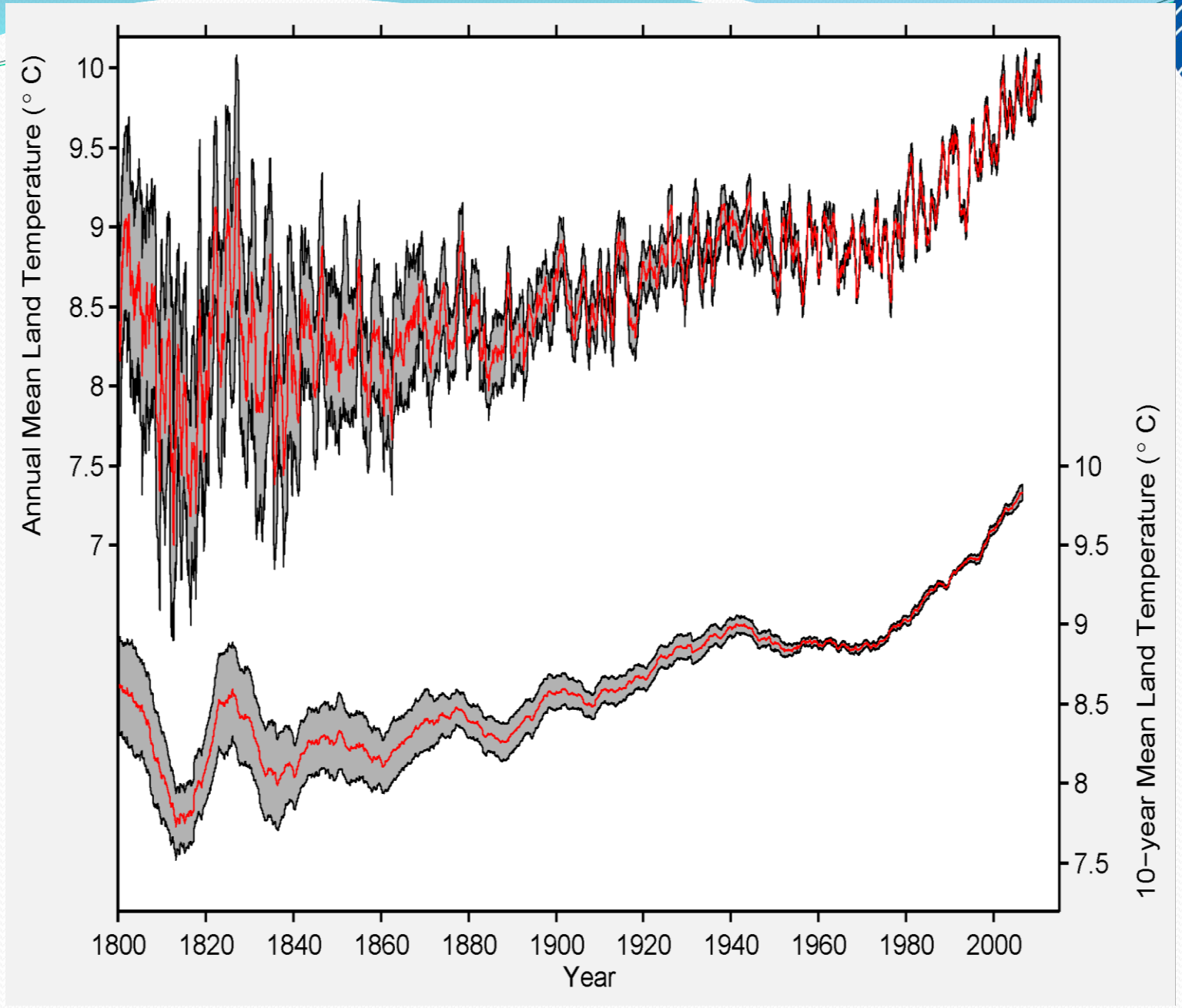
www.BerkeleyEarth.org

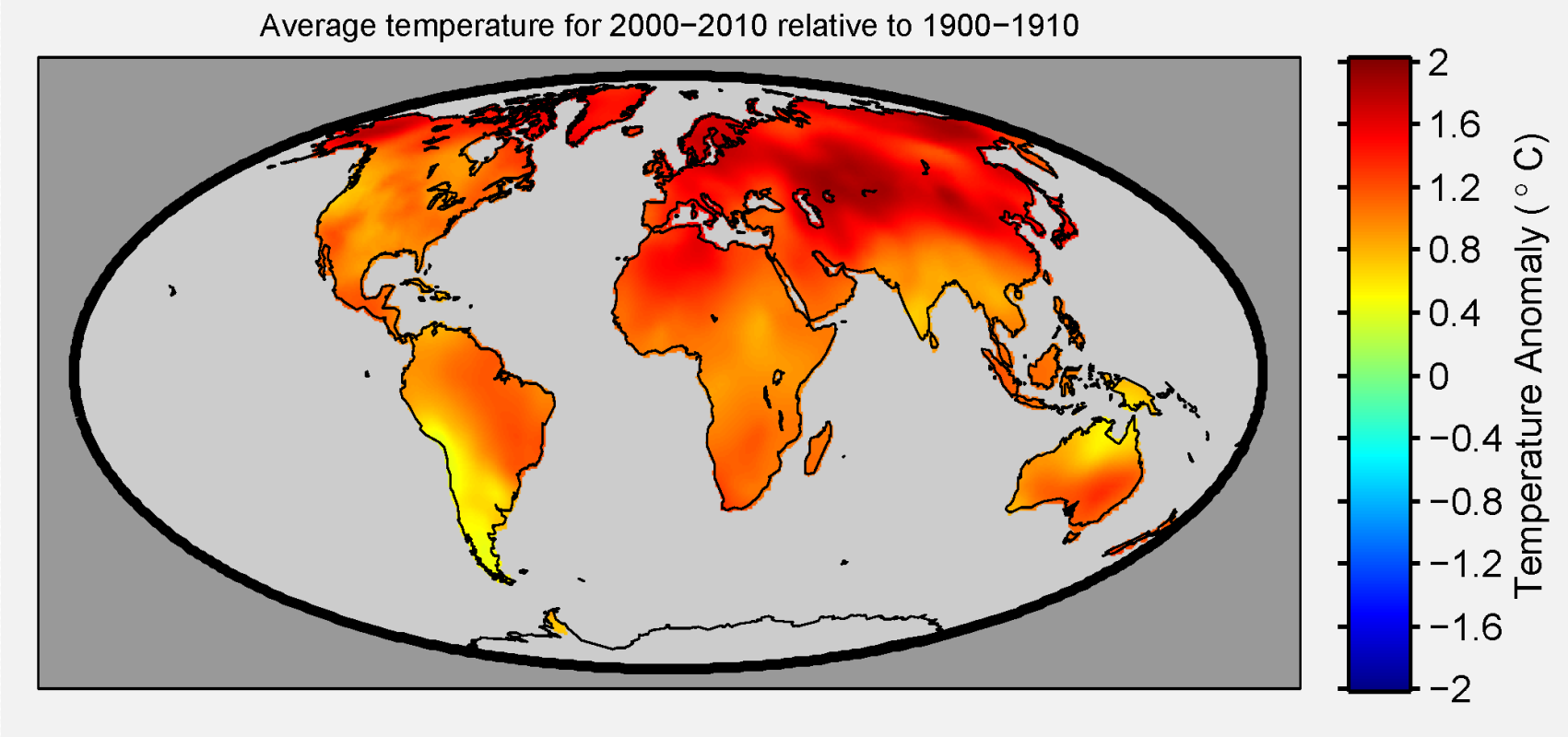

NASA GISS
Hadley / CRU
NOAA
0.6
0.4
0.2
0
−0.2
−0.4
Temperature Anomaly (° C)
NASA GISS
Hadley / CRU
NOAA
0.3
0.25
0.2
0.15
0.1
0.05
0
95% Uncertainty (° C)
1850
1900
1950
2000
Year

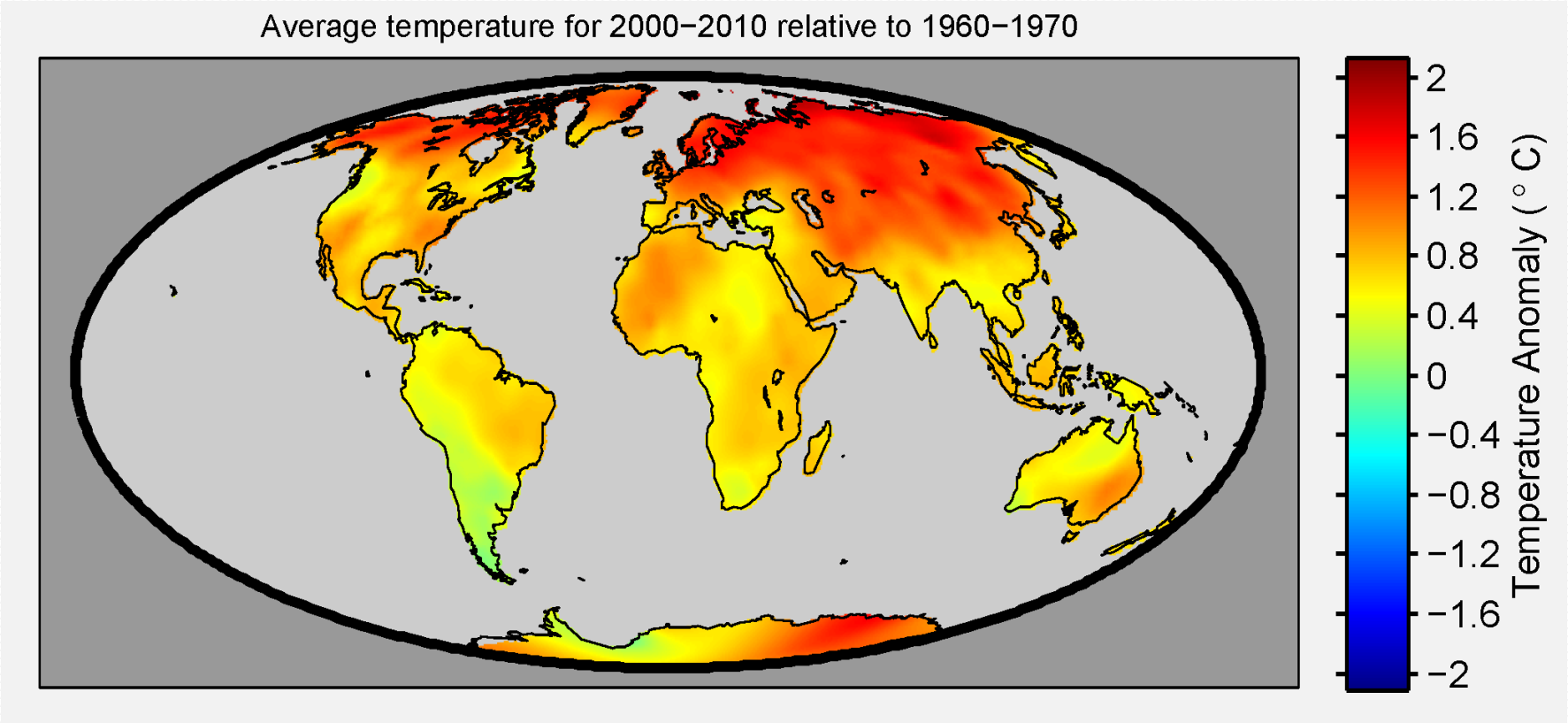

Correlation
Distance (km)
158 W
113 W
68 W
23 W
23 E
68 E
113 E
158 E
Global

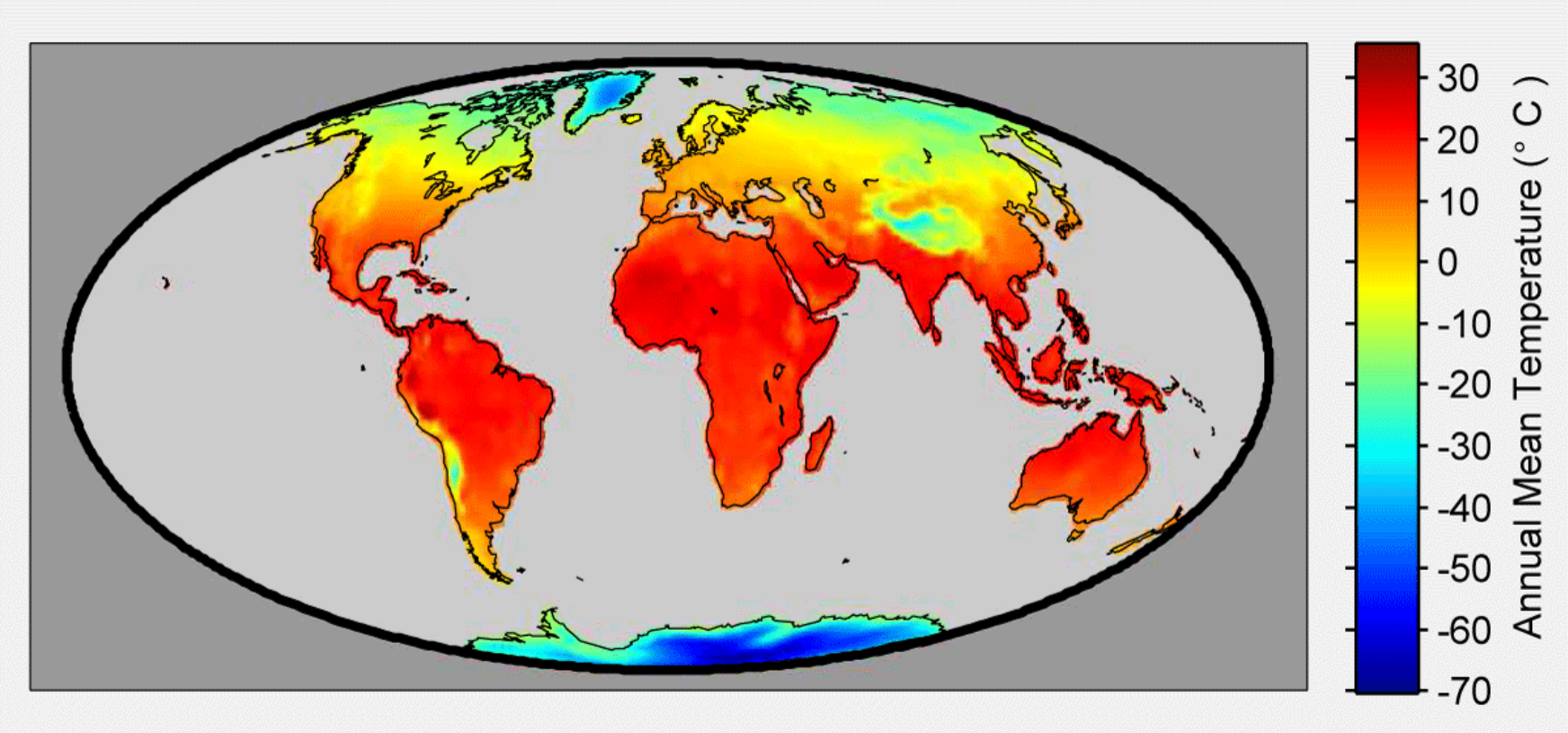

Correlation
Distance (km)
68 S
35 S
17 S
0
17 N
35 N
68 N
Global

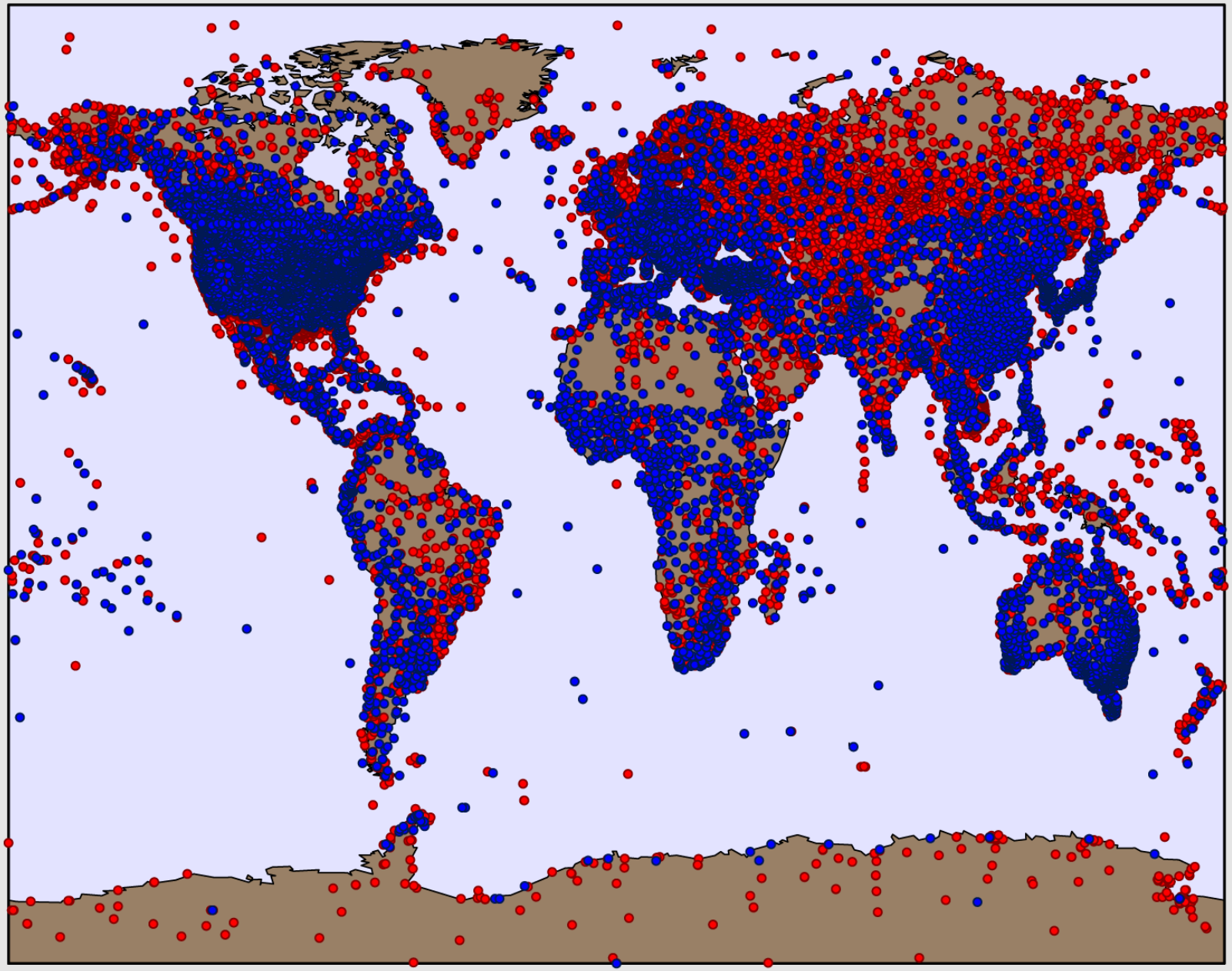

GHCN Station Locations
Thousands of Active Stations
0
1
2
3
4
5
6
1800
1850
1900
1950
2000
Land Area Sampled (%)
0
20
40
60
80
100
1800
1850
1900
1950
2000


Annual Mean Land Temperature (° C)
10
9.5
9
8.5
8
7.5
7
10−year Mean Land Temperature (° C)
10
9.5
9
8.5
8
7.5
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Year


Average temperature for 2000−2010 relative to 1900−1910
2
1.6
1.2
0.8
0.4
0
−0.4
−0.8
−1.2
−1.6
−2
Temperature Anomaly (° C)


Average temperature for 2000−2010 relative to 1960−1970
2
1.6
1.2
0.8
0.4
0
−0.4
−0.8
−1.2
−1.6
−2
Temperature Anomaly (° C)


30
20
10
0
-10
-20
-30
-40
-50
-60
-70
Annual Mean Temperature (° C)



Existing GHCN-M Stations (7,280; plotted in front)

Additional Berkeley Stations (32,110; plotted in back)

$$\hat{\theta}(t_j) + \hat{W}(\vec{x}, t_j) = \hat{\theta}(t_j) + \sum_i S_i(\vec{x}, t_j)\left(d_i(t_j) - \hat{b}_i - \hat{\theta}(t_j)\right)$$

$$= \left(1 - F(\vec{x}, t_j)\right)\hat{\theta}(t_j) + \sum_i S_i(\vec{x}, t_j)\left(d_i(t_j) - \hat{b}_i\right)$$

$$T(\vec{x},t) = \theta(t) + C(\vec{x}) + W(\vec{x},t)$$

$$\int_{\text{Earth's surface}} C(\vec{x})d\vec{x} = 0$$

$$\int_{\text{Earth's surface}} W(\vec{x},t)d\vec{x} = 0$$

$$\int_{\text{Earth's surface}} W(\vec{x},t)dt = 0$$

$$\widehat{W}(\vec{x}, t_j) = \sum_{i=1}^{N} S_{a_i}(\vec{x}, t_j)\left(d_{a_i}(t_j) - \hat{\theta}(t_j) - \hat{b}_{a_i}\right)$$