Spatio-Temporal Statistics with R

Chapter Two: Exploring Spatio-Temporal Data

Spatio-Temporal Data

Spatio-Temporal Data

- **Geostatistical**: continuous spatial index
- **Areal (lattice):** defined on finite/countable subset in space
- **Point process:** randomly located spatial processes

NOAA Daily Weather

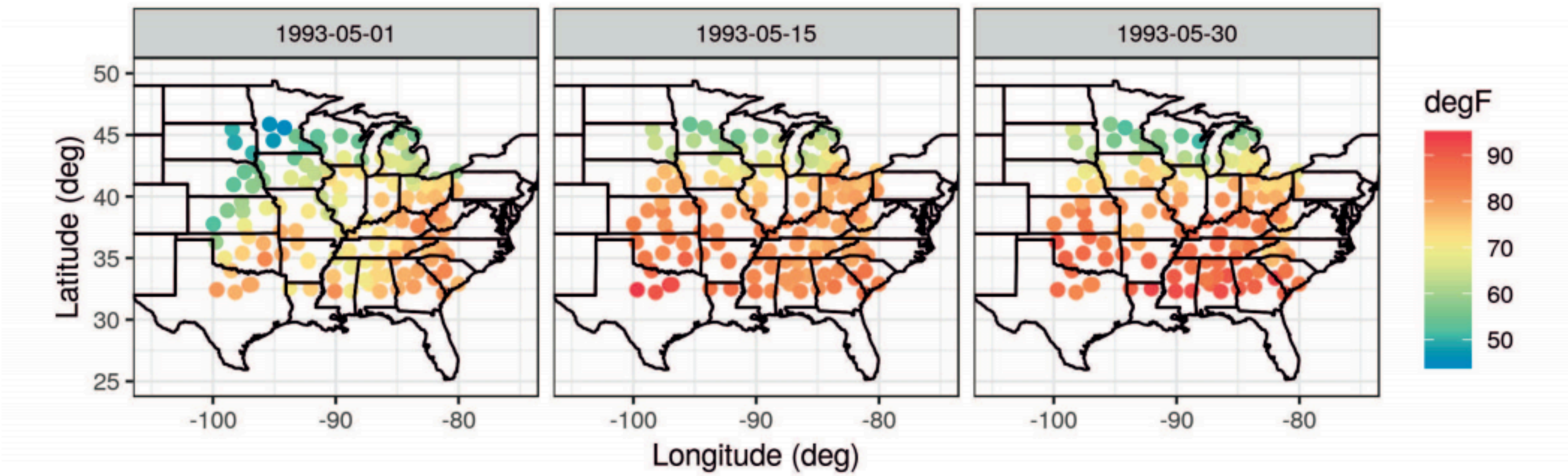


Figure 2.1: Maximum temperature (`Tmax`) in °F from the NOAA data set on 01, 15, and 30 May 1993.

Sea surface temperature

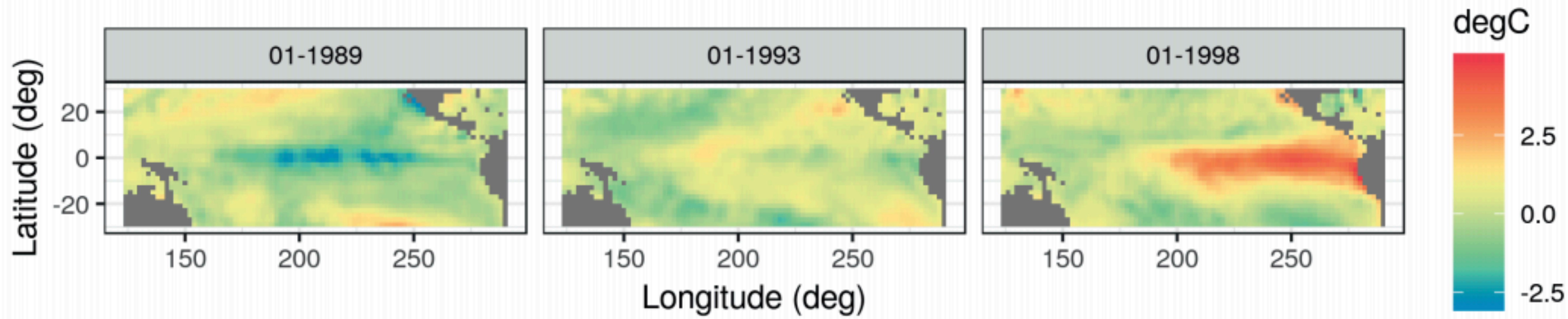


Figure 2.2: Sea-surface temperature anomalies in °C for the month of January in the years 1989, 1993, and 1998. The year 1989 experienced a La Niña event (colder than normal temperatures) while the year 1998 experienced an El Niño event (warmer than normal temperatures).

Mediterranean winds

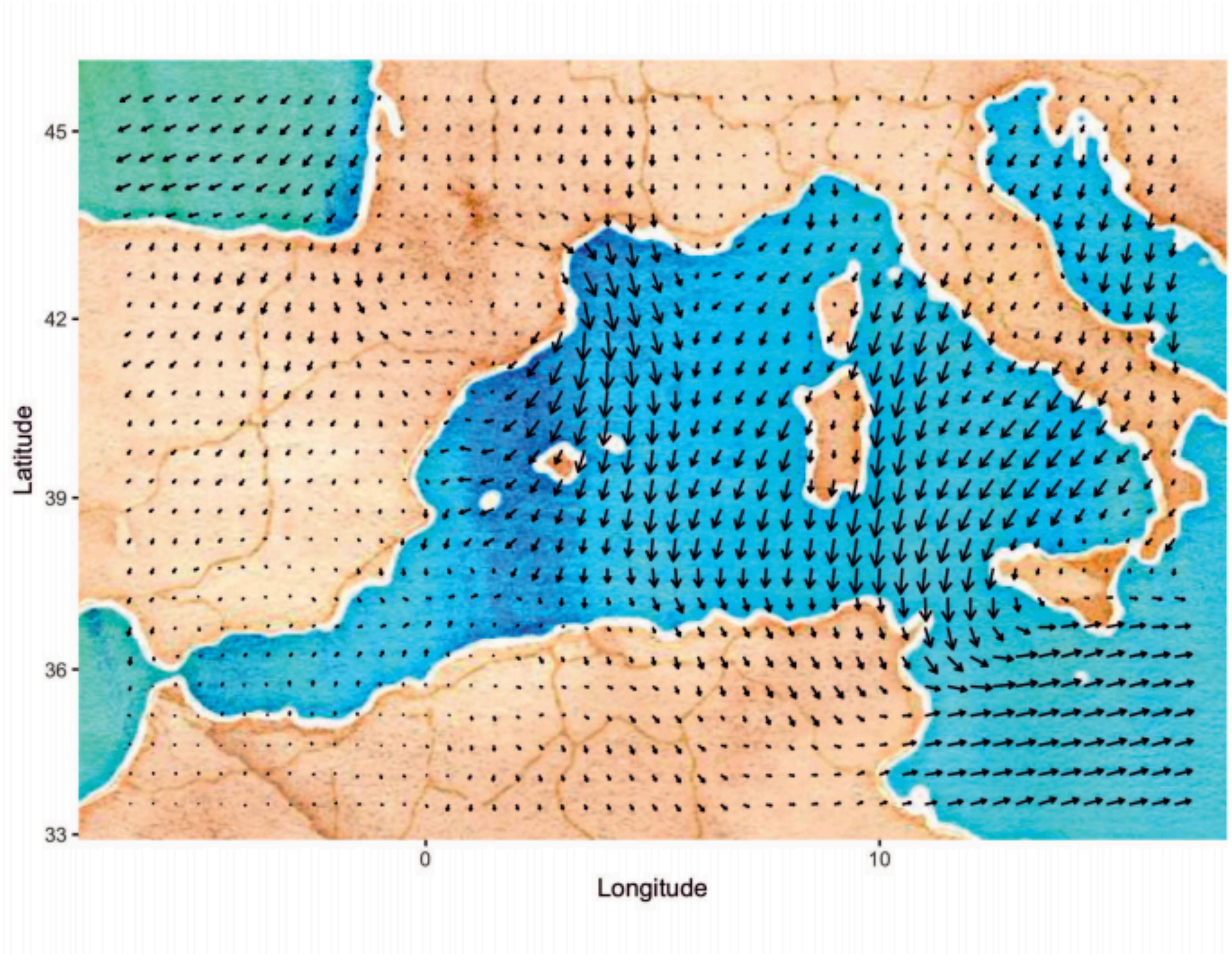


Figure 2.6: ECMWF wind vector observations over the Mediterranean region for 06:00 UTC on 01 February 2005.

Spatio-Temporal Data … in R

- spacetime package in R extends definitions from sp and xts
- How to represent spatio-temporal data?
    - Time-wide tables
    - Space-wide tables
    - Long format

Visualization

Spatial Plots

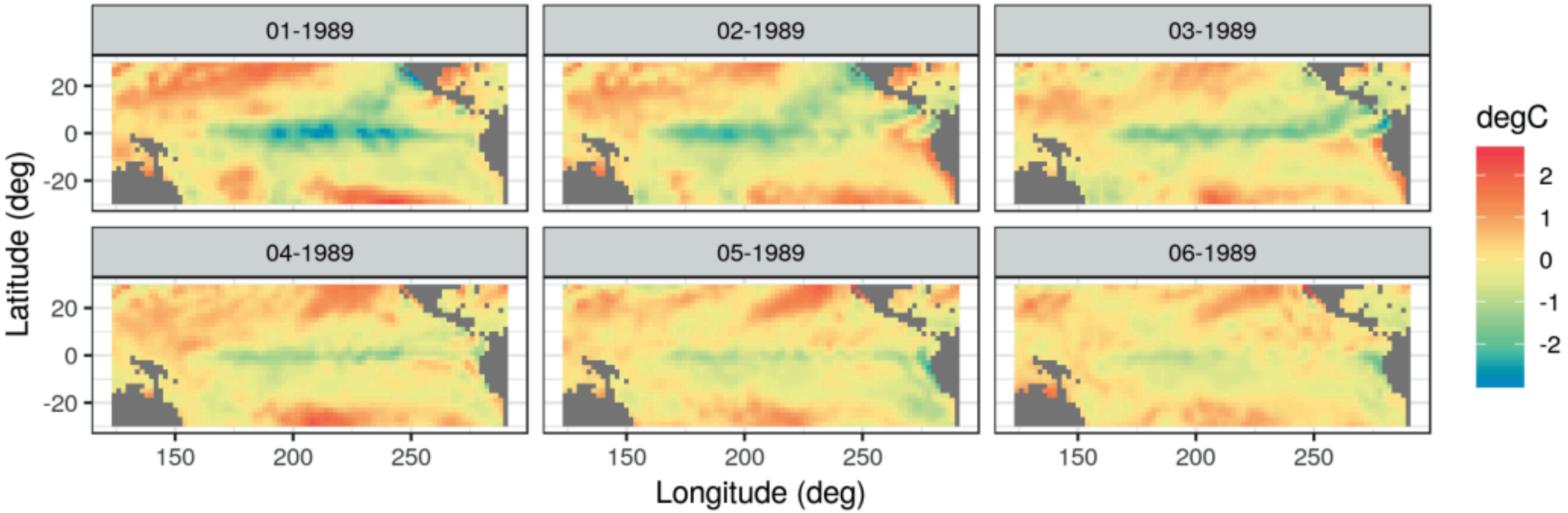


Figure 2.8: Sea-surface temperature anomalies (in °C) for January–June 1989.

Time-Series Plots

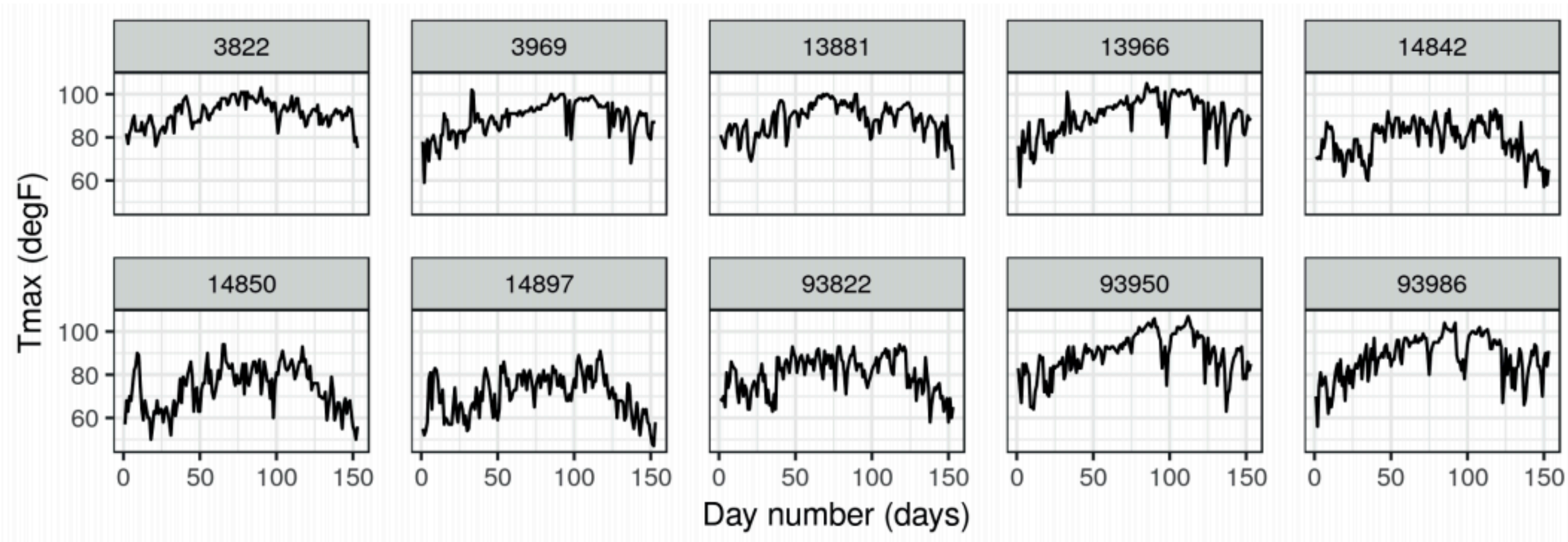


Figure 2.9: Maximum temperature (°F) for ten stations chosen from the NOAA data set at random, as a function of the day number, with the first day denoting 01 May 1993 and the last day denoting 30 September 1993. The number in the grey heading of each plot denotes the station ID.

Hovmöller Plots

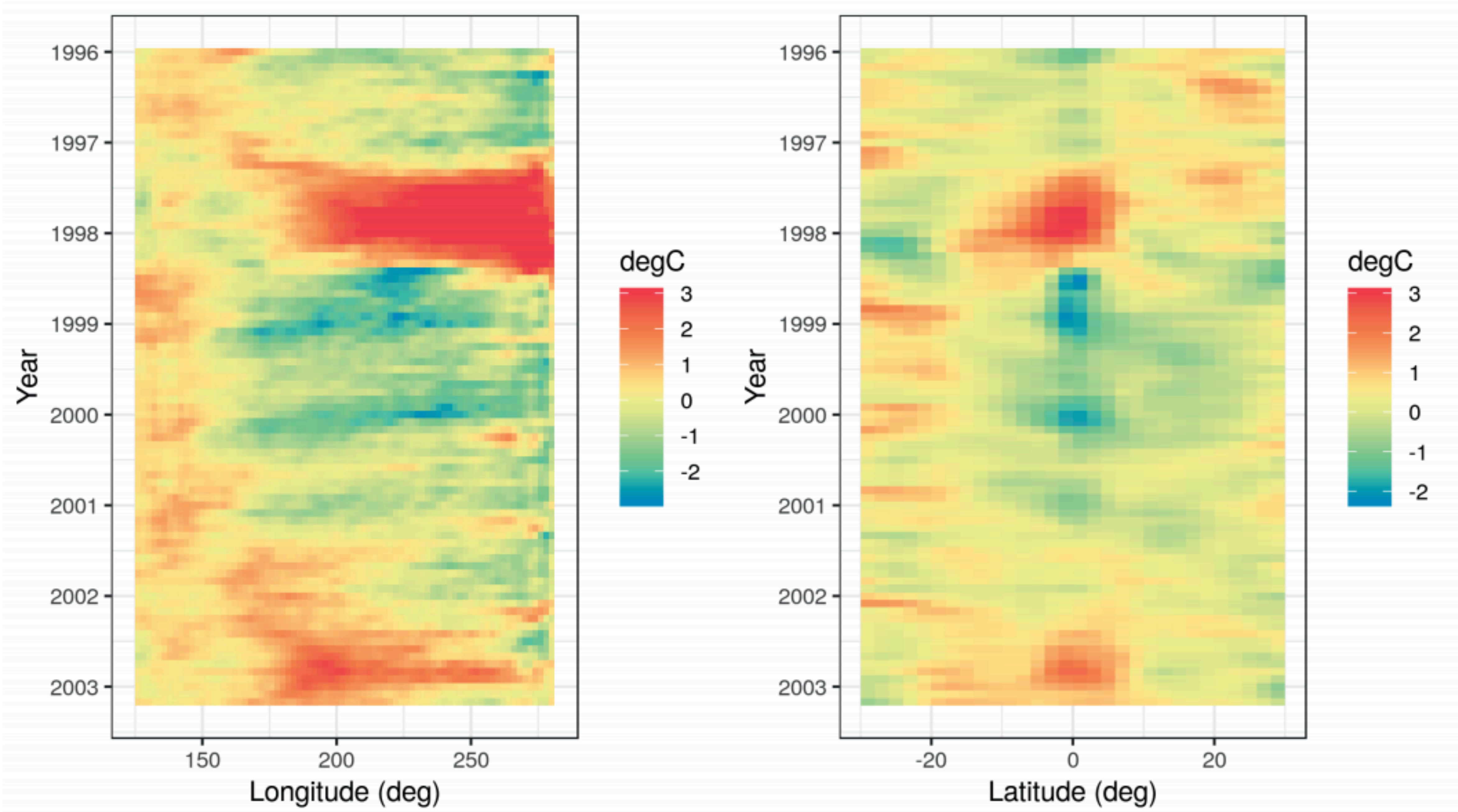


Figure 2.10: Hovmöller plots for both the longitude (left) and latitude (right) coordinates for the SST data set. The color denotes the temperature anomaly in °C.

Hovmöller Plots

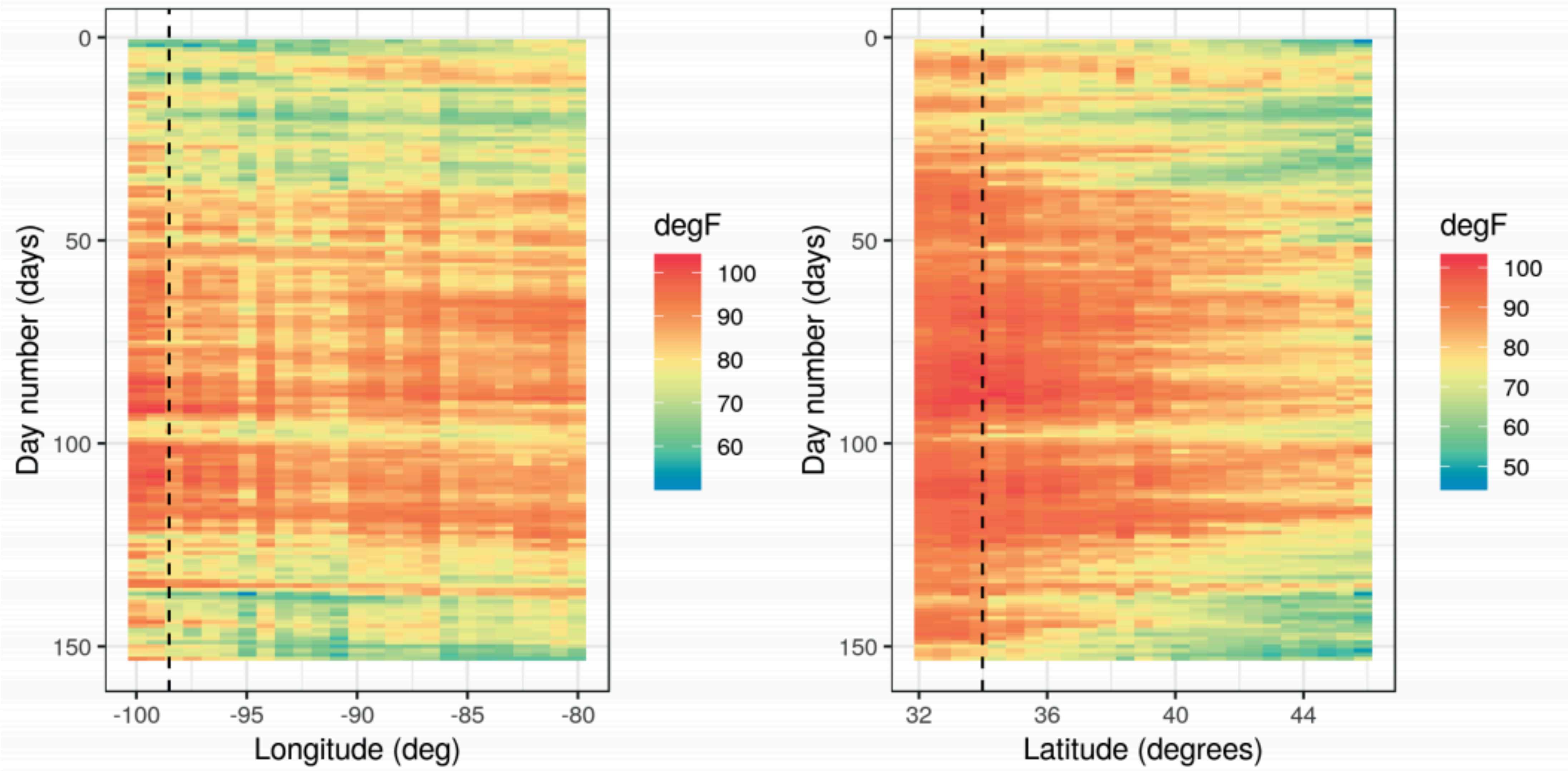


Figure 2.11: Hovmöller plots for both the longitude (left) and latitude (right) coordinates for the `Tmax` variable in the NOAA data set between 01 May 1993 and 30 September 1993, where the data are interpolated as described in Lab 2.2. The color denotes the maximum temperature in °F. The dashed lines correspond to the longitude and latitude coordinates of station 13966 (compare to Figure 2.9).

Interactivity

- trelliscope package allows for exploration/visualization of large-scale data sets
    - Designed to visualize distributed date by processing *in parallel* and then recombining.

Exploratory Analysis

Spatial Means

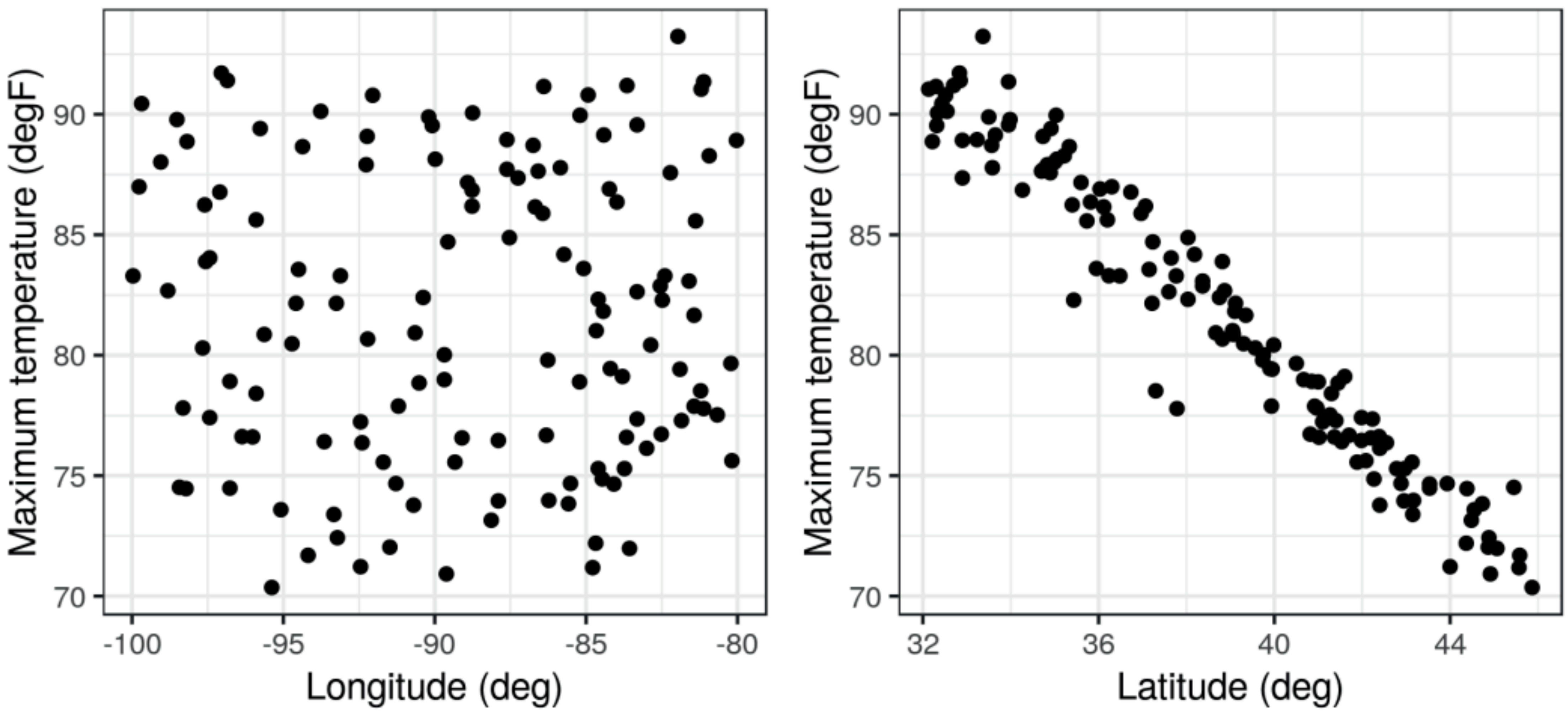


Figure 2.14: Empirical spatial mean, $\widehat{\mu}_{z,s}(\cdot)$, of Tmax (in °F) as a function of station longitude (left) and station latitude (right).

Temporal Means

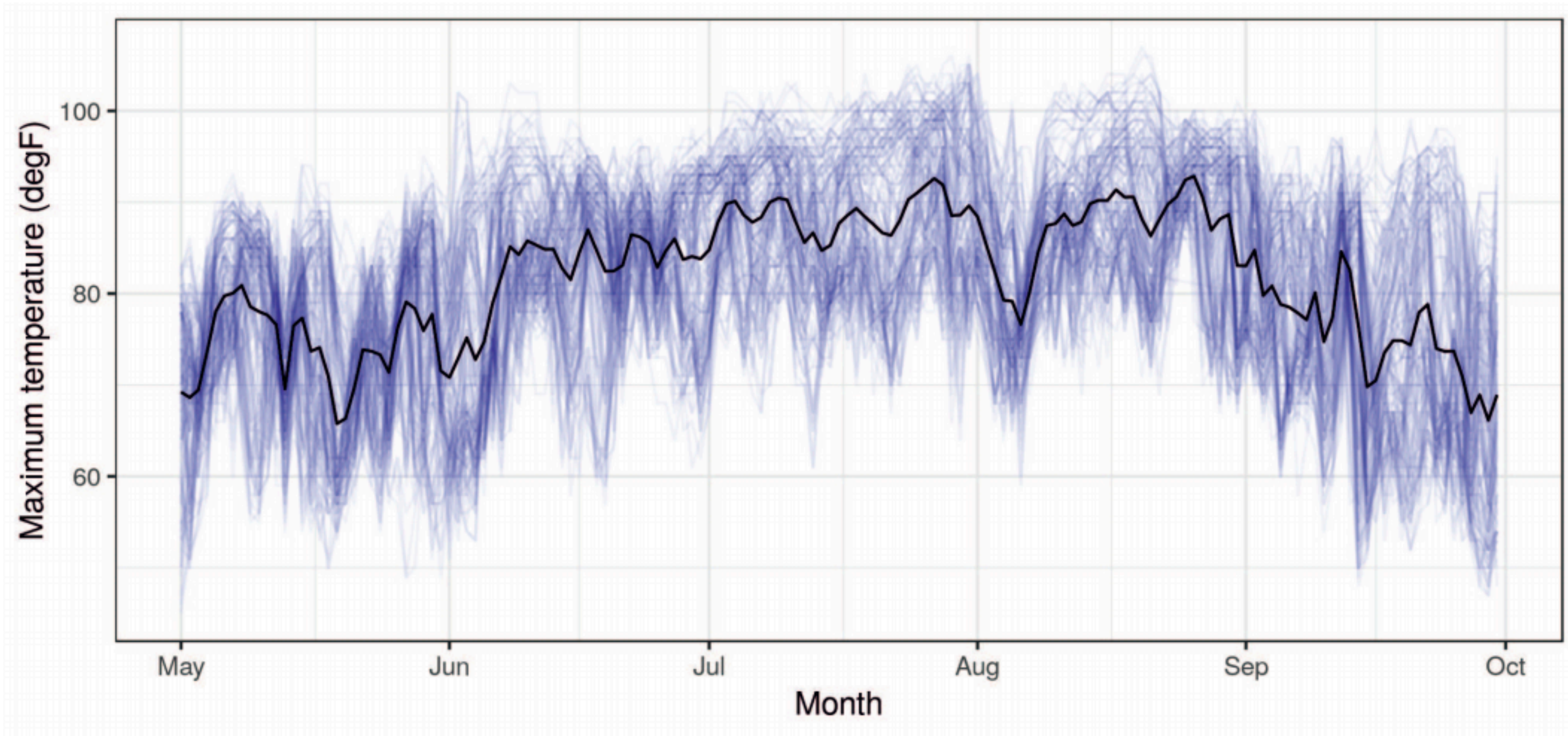


Figure 2.15: `Tmax` data (in °F), from the NOAA data set (blue lines, where each blue line corresponds to a station) and the empirical temporal mean $\widehat{\mu}_{z,t}(\cdot)$ (black line) computed from (2.2), and t is in units of days, ranging from 01 May 1993 to 30 September 1993.

Empirical spatial covariability

It is often useful to consider the empirical spatial covariability in the spatio-temporal data set. This covariability can be used to determine to what extent data points in the data set covary (behave similarly) as a function of space and/or time. In the context of the data described above, the empirical lag-τ covariance between spatial locations $\mathbf{s}_i$ and $\mathbf{s}_k$ is given by

$$\widehat{C}_z^{(\tau)}(\mathbf{s}_i, \mathbf{s}_k) \equiv \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (Z(\mathbf{s}_i; t_j) - \widehat{\mu}_{z,s}(\mathbf{s}_i))(Z(\mathbf{s}_k; t_j - \tau) - \widehat{\mu}_{z,s}(\mathbf{s}_k)), \qquad (2.3)$$

Empirical spatial covariability

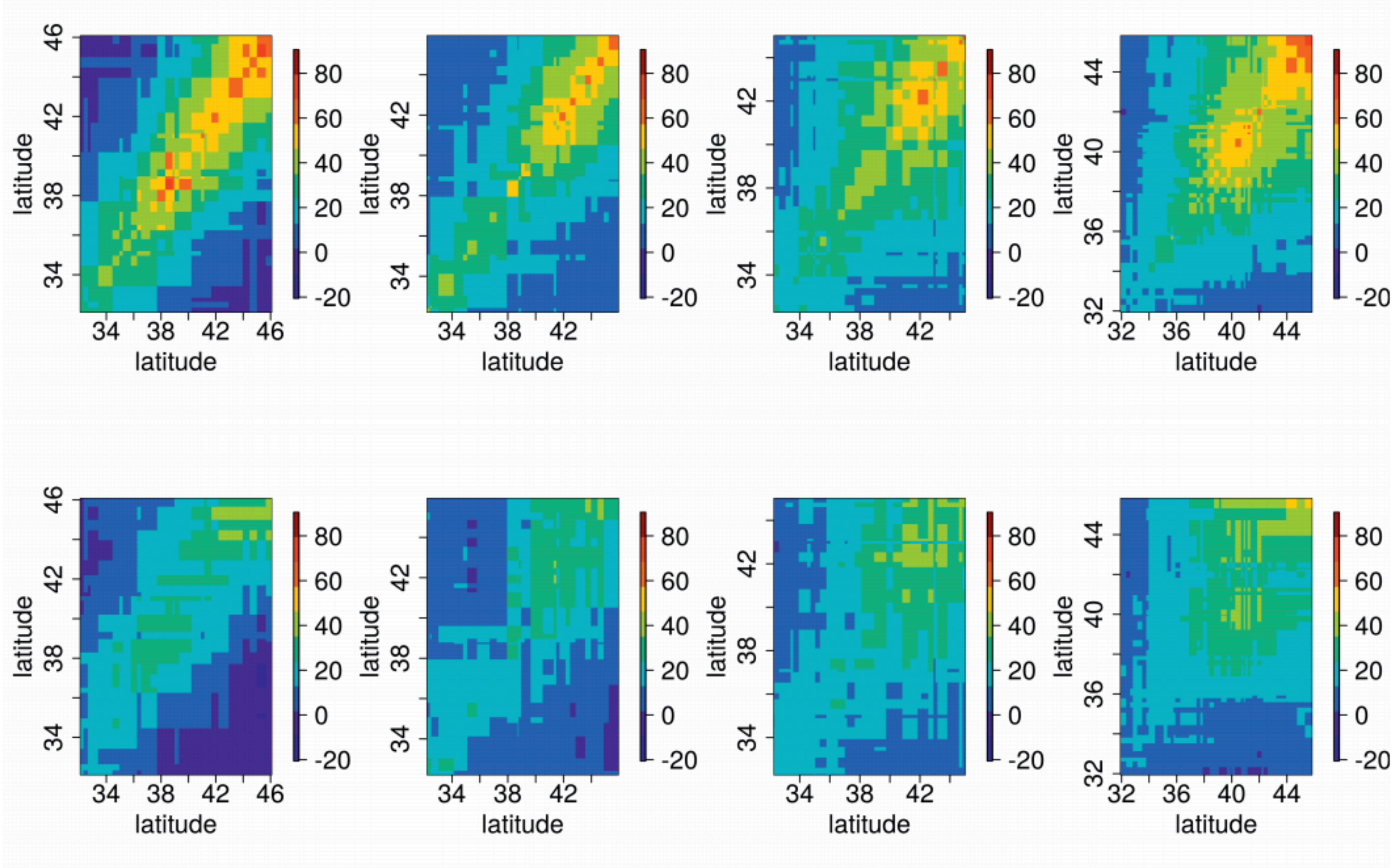


Figure 2.16: Maximum temperature lag-0 (top) and lag-1 (bottom) empirical spatial covariance plots for four longitudinal strips (from left to right, $[-100, -95), [-95, -90), [-90, -85), [-85, -80)$ degrees) in which the domain of interest is subdivided.

Spatio-Temporal Covariograms and Semivariograms

- The empirical spatiotemporal covariogram for spatial lag **h** and time lag τ is given by

$$\widehat{C}_z(\mathbf{h};\tau) = \frac{1}{|N_{\mathbf{s}}(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{\mathbf{s}_i,\mathbf{s}_k \in N_{\mathbf{s}}(\mathbf{h})}\sum_{t_j,t_\ell \in N_t(\tau)} (Z(\mathbf{s}_i;t_j) - \widehat{\mu}_{z,s}(\mathbf{s}_i))(Z(\mathbf{s}_k;t_\ell) - \widehat{\mu}_{z,s}(\mathbf{s}_k))$$

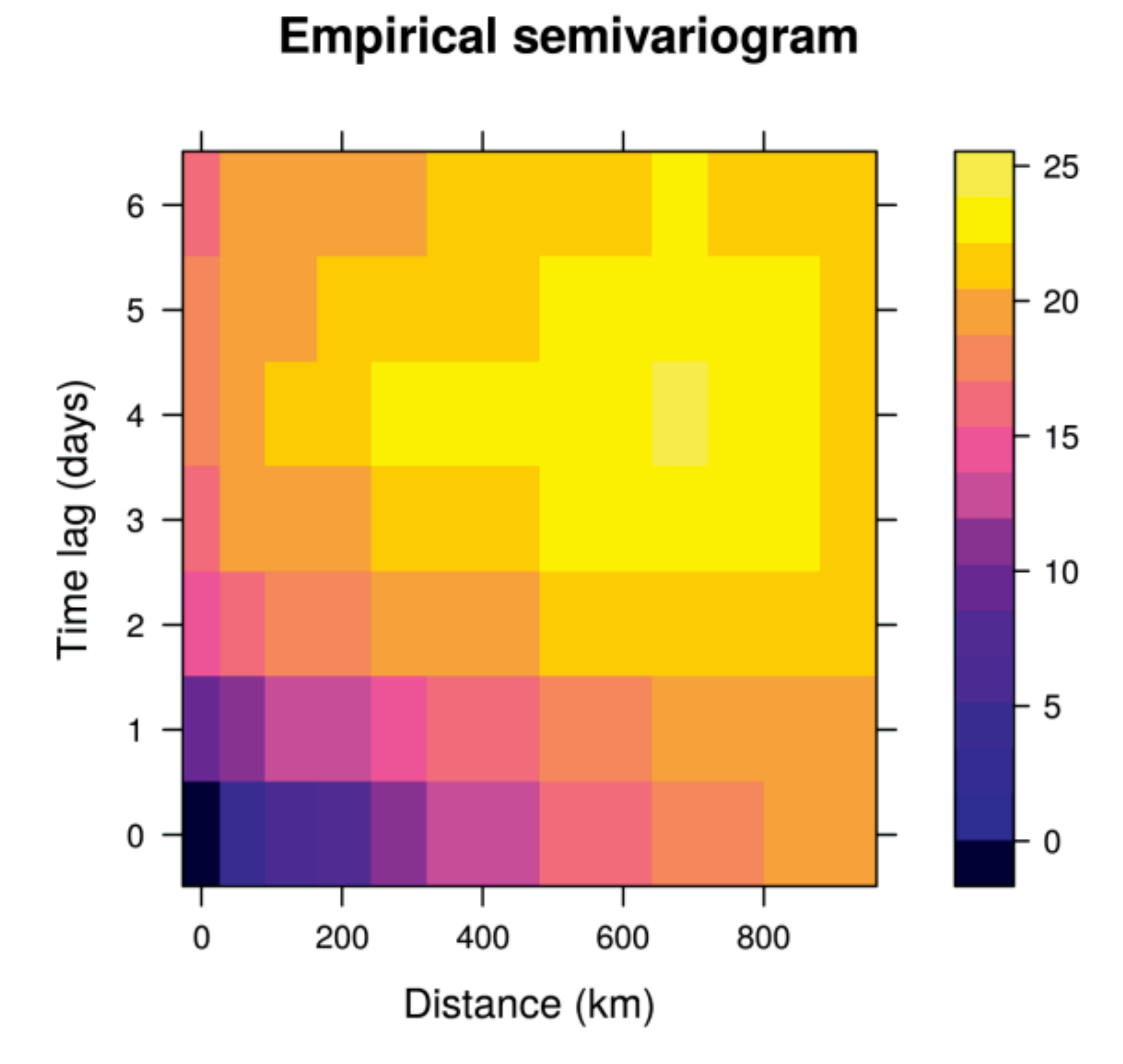
Spatio-Temporal Covariograms and Semivariograms

- If we assume covariance only depends on displacement in time/ space, we can estimate **semivariogram**:

$$\gamma_z(\mathbf{h};\tau) = \frac{1}{2}E\left(Z(\mathbf{s}+\mathbf{h};t+\tau) - Z(\mathbf{s};t)\right)^2,$$

$$\widehat{\gamma}_z(\mathbf{h};\tau) = \frac{1}{|N_{\mathbf{s}}(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{\mathbf{s}_i,\mathbf{s}_k\in N_{\mathbf{s}}(\mathbf{h})}\ \sum_{t_j,t_\ell\in N_t(\tau)}(Z(\mathbf{s}_i;t_j) - Z(\mathbf{s}_k;t_\ell))^2,$$

Spatio-Temporal Covariograms and Semivariograms



Empirical Orthogonal Functions

- Empirical orthogonal functions (EOFs) used for dimensionality reduction and for studying spatial structure
- Recall principal component analysis projects original data onto new coordinate space where the first coordinate aligns with the axis of largest variation, …
- EOFs are obtained by treating observations at different time points as "sample" and computing principal components

Empirical Orthogonal Functions

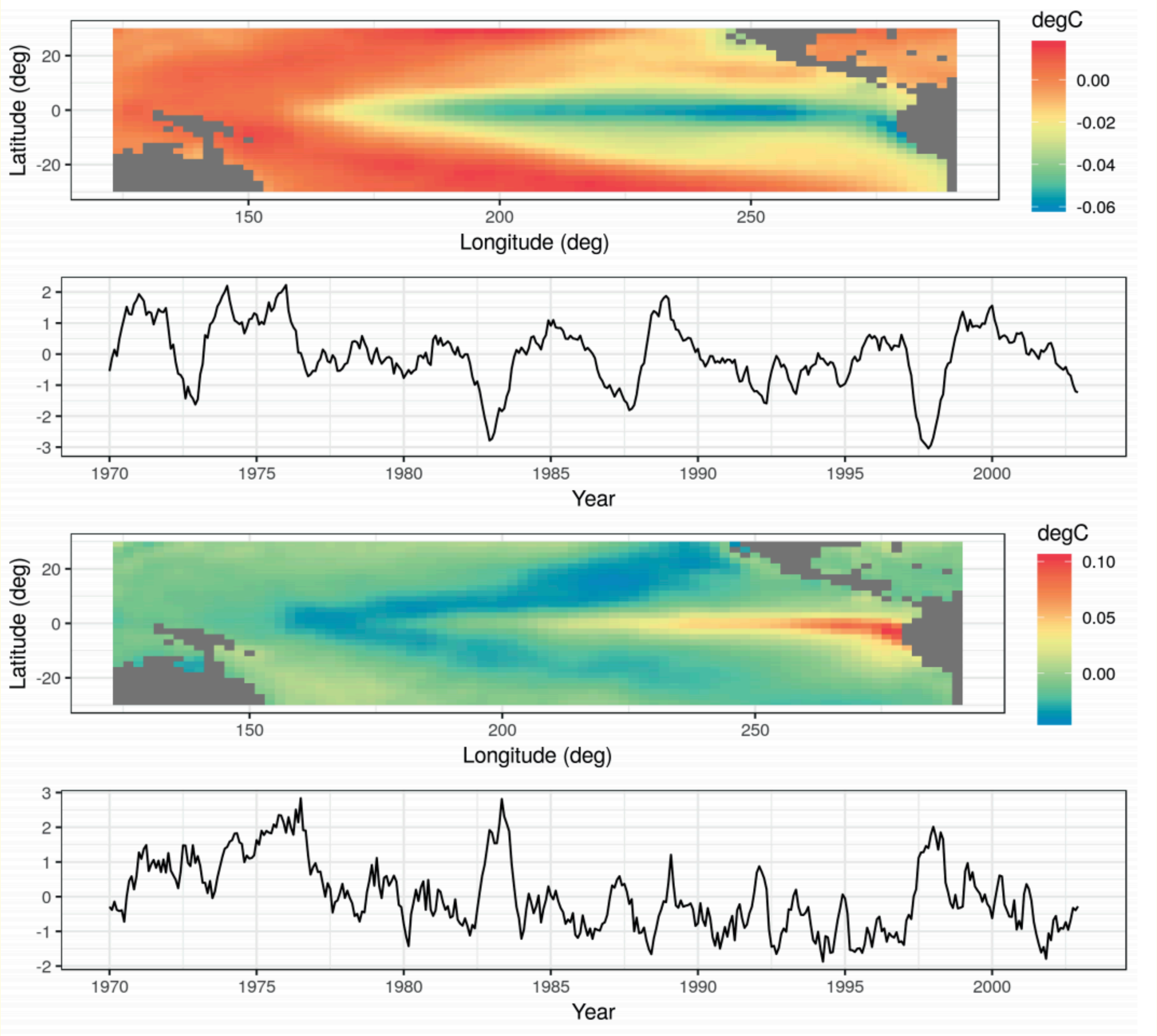


Figure 2.20: The first two empirical orthogonal functions and normalized principal-component time series for the SST data set obtained using an SVD of a space-wide matrix.

Lab 2.2: Visualization

Lab 2.3: Exploratory Data Analysis