Half-Spectral Space-Time Models

Joe Guinness
Cornell University Statistics

(Open with Acrobat to Display Embedded Videos)
Gaussian Process Models

Gaussian Process: Model for samples of continuous fields
- temperatures measured at weather stations
- soil moisture content at sampled locations
- windspeeds inferred from satellite observations

Common applications:
- interpolation of spatial-temporal data
- forecasting
- accounting for dependence in linear regressions

Use of the model is computationally demanding
- filling and storing covariance matrix
- manipulating the matrix (factoring, solving systems, etc.)
Structure in Spatial-Temporal Data

Dense in Space, Sparse in Time

Nicolas Hoepffner, European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy
Structure in Spatial-Temporal Data

GOES 16 and 17 satellites

- up to 500m and 30 second resolution, 16 spectral bands

Source: Colorado State CIRA
Structure in Spatial-Temporal Data

Roving Monitors

Horrell and Stein (2015), Statistica Sinica
“A covariance parameter estimation method for polar-orbiting satellite data”
Structure in Spatial-Temporal Data

Long time series at few spatial locations

\( n \) spatial locations, \( T \) time points, \( T >> n \)
Stationary Time Series Primer

$Y(t)$: time series

Stationary Model:

$$\text{Cov}(Y(t), Y(t + h)) = K(h)$$

Spectral representation

$$Y(t) = \frac{1}{\sqrt{2\pi}} \int_{[0,2\pi]} A(\omega)e^{i\omega t}dZ(\omega) \quad \leftarrow \quad Z(\omega) \text{ brownian motion}$$

Discrete Approximation for $t = 1, \ldots, T$ \quad $(\omega_j = 2\pi j/T)$

$$Y(t) \approx \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j)e^{i\omega_j t} \hat{Z}_j \quad \leftarrow \quad Z_1, \ldots, Z_T \text{ unc., mean 0, var. 1}$$

Discrete Fourier transform decorrelates $(Y(1), \ldots, Y(T))$

$$A(\omega_j)Z_j = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{-i\omega_j t}Y(t)$$
Stationary Space-Time Series Primer

Space-time Data: \( Y(x, t) \)

Stationary Model: \( \text{Cov}(Y(x, t), Y(x + s, t + h)) = K(s, h) \)

Spectral Representation:

\[
Y(x, t) = \int_{[0,2\pi]} \int_{\mathbb{R}^2} B(\omega, \nu)e^{i\omega t + i\nu'x} dX(\omega, \nu) \leftarrow \text{BM in } \omega \text{ and } \nu
\]

Integrate over spatial frequency (Stein, 2005)

\[
Y(x, t) = \int_{[0,2\pi]} A(\omega)e^{i\omega t} dZ(\omega, x) \leftarrow \text{BM in } \omega, \text{ correlated over } x
\]

DFT over time decorrelates

\[
A(\omega_j)Z_j(x) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{-i\omega_j t} Y(x, t)
\]

\( Z_j(x) \) stationary over space, uncorrelated over \( j \)

\[
\text{Cov}(Z_j(x), Z_j(x + s)) = C(s, \omega_j)
\]
Computational Savings

Dominant computational cost: storing covariance matrices

Original Data: \( Y(x, t) \) at \( x_1, \ldots, x_n, t = 1, \ldots, T \)
- Everything correlated with everything else
- Dense \((nT) \times (nT)\) covariance matrix

DFT over time: \( Z_j(x) \) at \( x_1, \ldots, x_n, j = 1, \ldots, T \)
- Correlation only over space \( x \), within frequency \( j \)
- \( T \ (n) \times (n) \) dense matrices (one for each frequency)
- Storage cost reduced by factor of \( T \)
Modeling Considerations: Separable vs Nonseparable

Coherence function: \( \text{Cov}(Z_j(x), Z_j(x + s)) = C(s, \omega_j) \)

Separable Model: \( C(h, \omega_j) = C(h) \implies K(s, h) = K_1(s)K_2(h) \)
Statistical Compression of Climate Model Data

National Center for Atmospheric Research Supercomputing Budgets:

- Yellowstone: 20% for storage hardware (now retired)
- Cheyenne: 50% for storage hardware (current)
- Cheyenne: 52.7 Petabytes (tape library is 320 PB)
- For reference: 52.7 PB on Google Cloud = $369K to $1.37M per month
Statistical Compression of Climate Model Data

National Center for Atmospheric Research Supercomputing Budgets:

- Yellowstone: 20% for storage hardware (now retired)
- Cheyenne: 50% for storage hardware (current)
- Cheyenne: 52.7 Petabytes (tape library is 320 PB)
- For reference: 52.7 PB on Google Cloud = $369K to $1.37M per month

We propose lossy statistical compression algorithms

- store a low-dimensional summary + conditional distribution of data
- decompress using conditional distributions (expectations, simulations)
Statistical Compression of Climate Model Data

National Center for Atmospheric Research Supercomputing Budgets:

- Yellowstone: 20% for storage hardware (now retired)
- Cheyenne: 50% for storage hardware (current)
- Cheyenne: 52.7 Petabytes (tape library is 320 PB)
- For reference: 52.7 PB on Google Cloud = $369K to $1.37M per month

We propose lossy statistical compression algorithms

- store a low-dimensional summary + conditional distribution of data
- decompress using conditional distributions (expectations, simulations)

Low Dimensional Summary: Subset of (transformed) data

Statistical Model: Nonstationary Gaussian Process Model
Climate Model Output

Global Climate Model Data

$T = 365 \text{ days, } n = 54,720 \text{ pixels}$
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency
Maps of Fourier coefficients $A(\omega_j)Z_j(x)$

Variation concentrated on low frequencies

Spatial correlation within each frequency

Real Part, frequency 180
Imaginary Part, frequency 180
Lower-Dimensional Summary

Our idea - store a subset of Fourier coefficients

Real Part, frequency 3
Locations of Saved Coefficients, frequency 3

Gaussian process model for the remaining coefficients

- Can interpolate (with uncertainties) remaining coefficients
Half-Spectral Space-Time Models

Stationary Models (Stein, 2005)

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]
Half-Spectral Space-Time Models

Stationary Models (Stein, 2005)

\[ Y(\mathbf{x}, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]

Allow Transfer functions to vary in space, but retain stationary \( Z_j(x) \)

\[ Y(\mathbf{x}, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j, x) e^{i\omega_j t} Z_j(x) \]

\[ \text{Cov}(Z_j(x), Z_j(x')) = C(x - x'; \omega_j) \]
Half-Spectral Space-Time Models

Stationary Models (Stein, 2005)

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]

Allow Transfer functions to vary in space, but retain stationary \( Z_j(x) \)

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j, x) e^{i\omega_j t} Z_j(x) \]

\[ \text{Cov}(Z_j(x), Z_j(x')) = C(x - x' ; \omega_j) \]

OR Nonstationary \( Z_j(x) \)

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]

\[ \text{Cov}(Z_j(x), Z_j(x')) = C(x, x' ; \omega_j) \]
Spatial Correlation in Fourier Coefficients

Under half-spectral model:

$Z_j(x)$ approximately uncorrelated across $\omega$, variance 1.

Expect to see correlation across $x$ in $Z_j(x)$
Fitted Spatial Coherence Parameters

Spatial Covariance Model: Matern, smoothness 1, inverse range $\kappa$

\[
\text{Cov}(Z_j(x), Z_j(x')) = M_1(\kappa \|x - x'\|)
\]
Original and Decompressed Time Series
Original and Decompressed Time Series

10:1 Decompressed Data

Temperature (Celsius)

![Graph showing original and decompressed time series data.](image-url)
Conditional Simulations

![Map Showing Conditional Simulations with Original Data and 10:1 Decompressed Data]
Temperature Observations

Elevations and average temperatures at ARM extended facilities

Degrees longitude west

Degrees latitude north

Temperatures at extended facilities, October 1–30, 2005

Site no.

Days

Degrees Celsius
Nonstationary in Time

Stationary Models

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]
Nonstationary in Time

Stationary Models

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j) e^{i\omega_j t} Z_j(x) \]

Allow transfer function to depend on time

\[ Y(x, t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} A(\omega_j, t) e^{i\omega_j t} Z_j(x) \]

Inverse transformation (to get \( Z_j(x) \)) no longer DFT

Generalized Fourier transform: simply inverse of above

- Can be computed efficiently in some situations
Selected References


