Introduction to State-space Techniques for Analyzing CMB Anisotropy Data

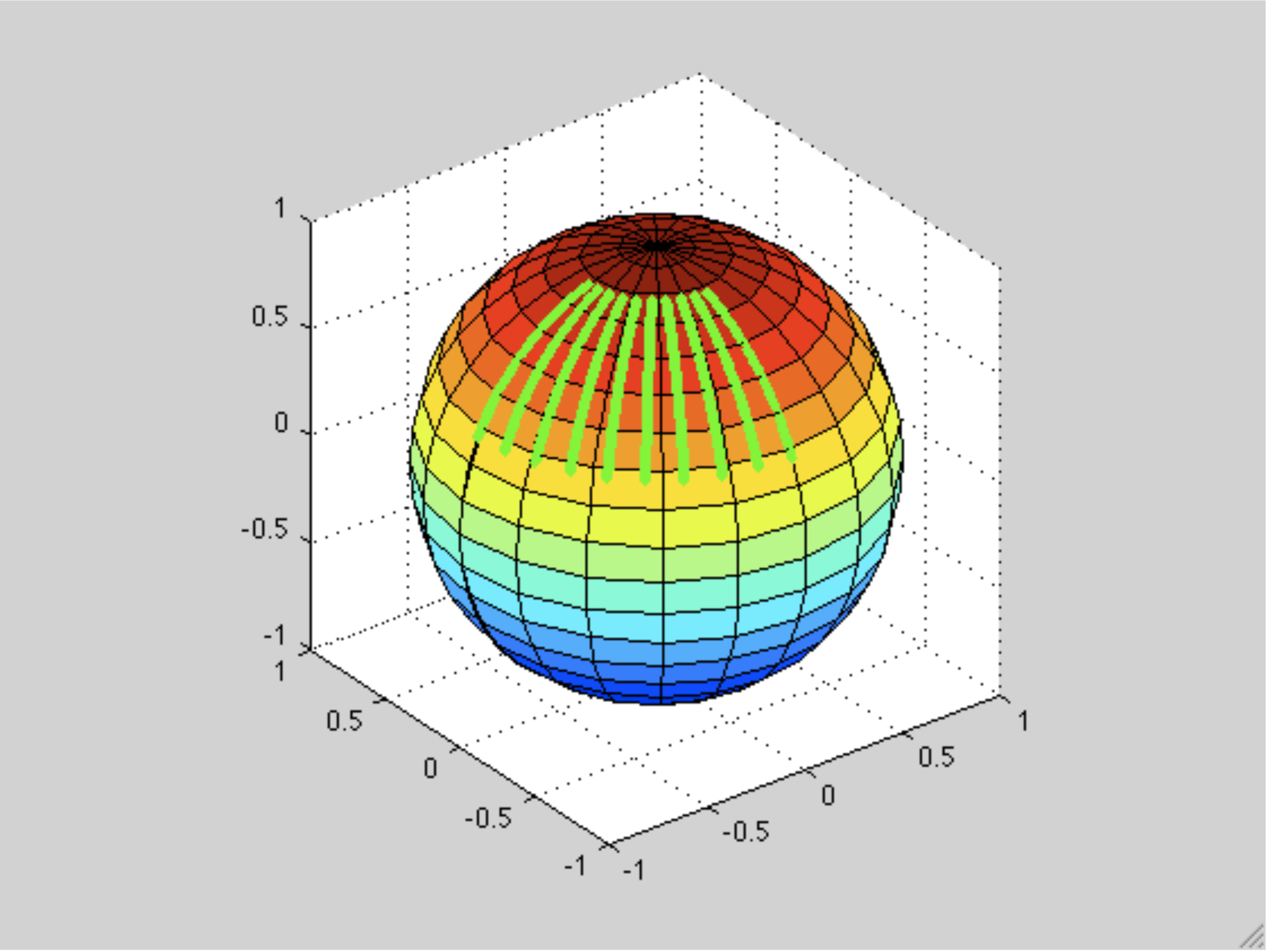
Jim White
October 21, 2009

james.v.white@gmail.com

Are CMB anisotropies homogeneous and isotropic?

If so, statistical temperature correlations along different tracks that avoid the galactic foreground should be the same

- In any direction
- In any hemisphere or quadrant

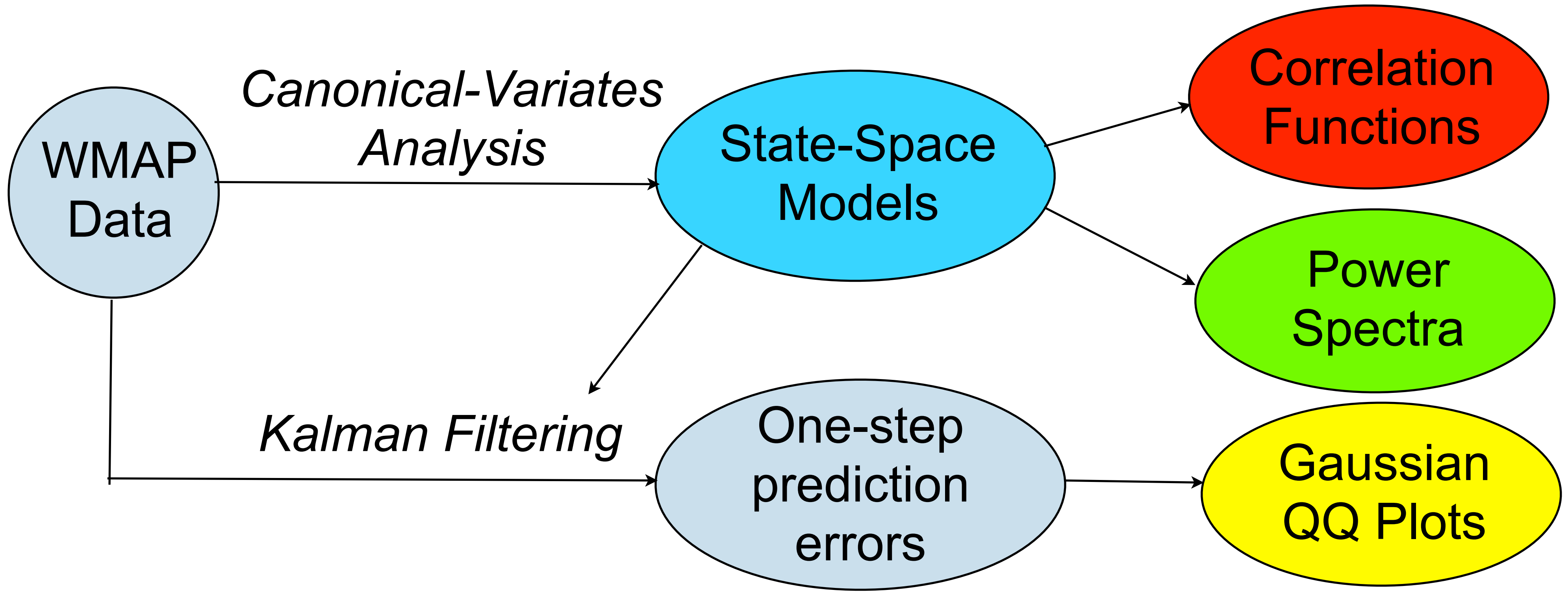


James.V.White@gmail.com

Are WMAP temperature anisotropies well-modeled as a homogeneous Gaussian field?

If so, established time-series techniques for Gaussian random processes can be used to model WMAP data along tracks

- Canonical variates analyses
- State-space models
- Kalman filters



James.V.White@gmail.com

How do empirical correlations and spectra compare with CMB theory?

- To answer this question, empirical correlations and spectra must be determined for subsets of the sphere without the benefit of CMB theory.

- The State-Space Canonical Variates (CV) algorithm provides a rigorous basis for modeling CMB data along tracks, independently of CMB theory.

James.V.White@gmail.com

Relevant Prior Experience

Anisotropies of earth's gravity field have already been modeled using state-space techniques

Both CMB and geodetic anisotropies

- Have the same fractional order of magnitude (10^{-5})
- Are defined on a sphere
- Are usefully modeled as 2-D Gaussian random fields and 1-D random processes along tracks

With Navy support, I have already developed and applied state-space techniques to determine correlations and power spectra of geodetic anisotropies

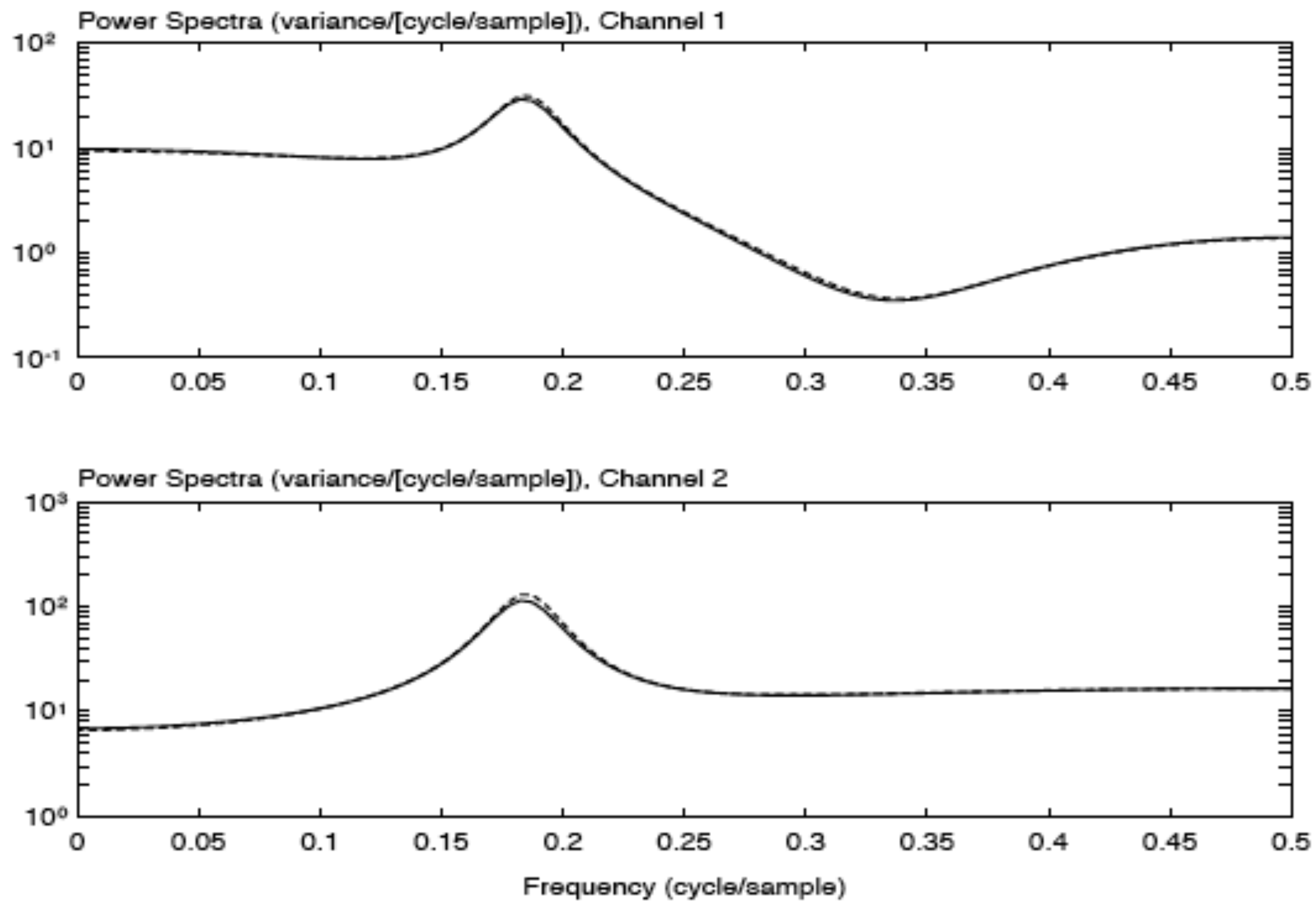
- *Scalar* gravity anomalies (analogous to temperature variations of CMB)
- *Vector* data (deflections of the vertical, 2-D like CMB polarization data)
- *Tensor* data (*3x3* gravity gradients, analogous to CMB ?)

James.V.White@gmail.com

Demonstration of State-space Techniques

Using Synthetic Data from Specified Random Processes

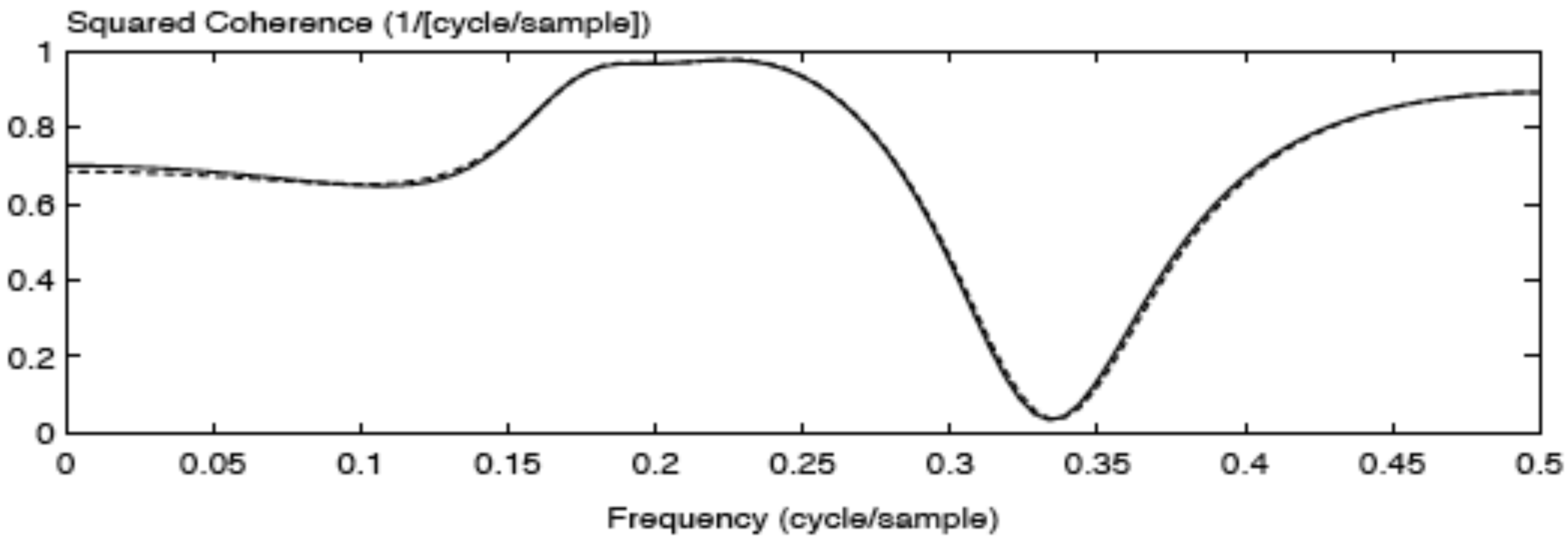
Comparison of State-space Estimated and True Power Spectra



- Two-channel vector sequence (like polarization data) containing 4,096 sample vectors
- Solid curves are state-space estimates generated from empirical data
- Dashed curves are the true spectra of the process that generated the data

James.V.White@gmail.com

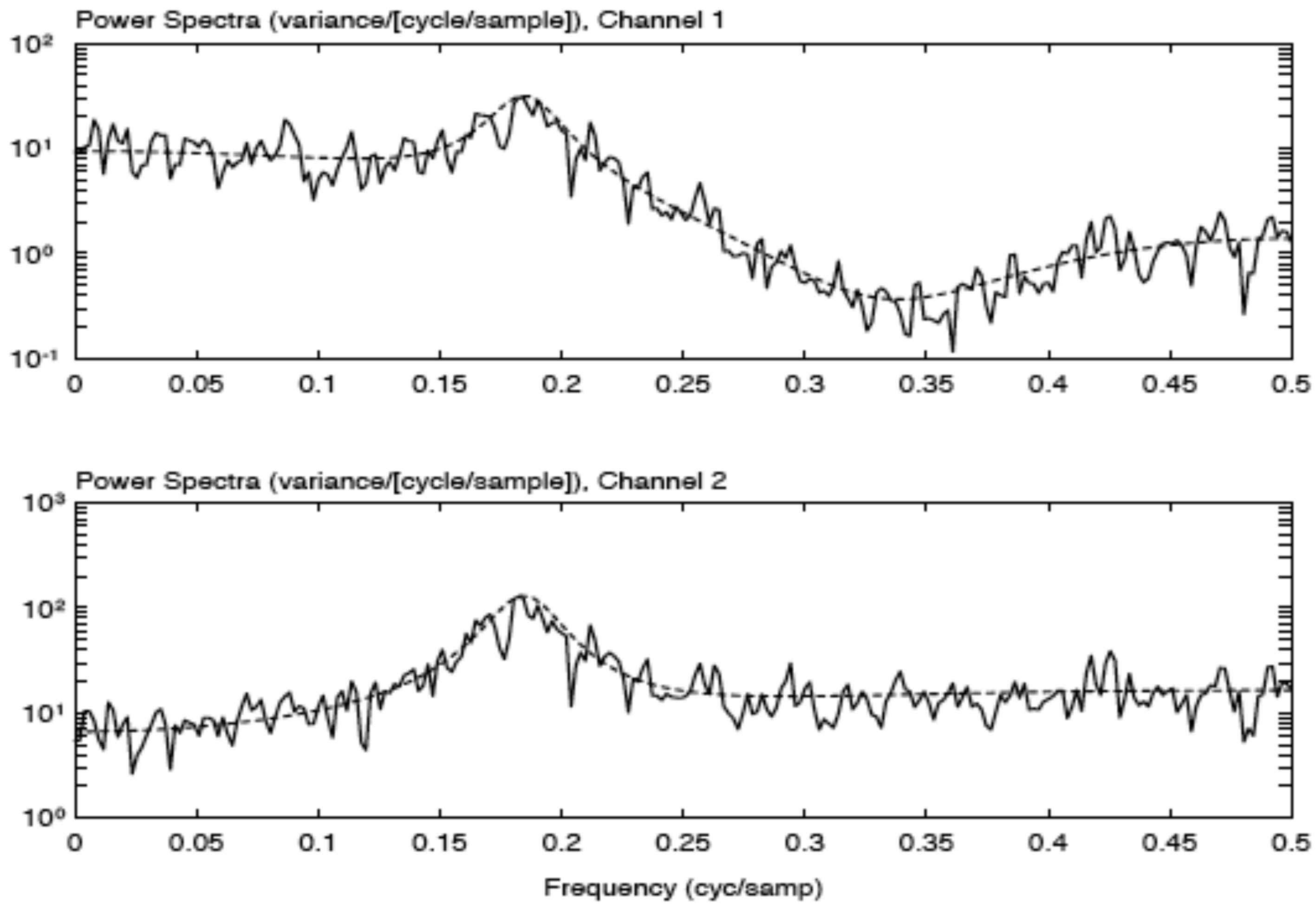
Comparison of State-space Estimated and True Spectral Coherence



- Solid curve is state-space estimate of spectral coherence between two channels based on the empirical data
- Dashed curve is the true coherence function of the generating process

James.V.White@gmail.com

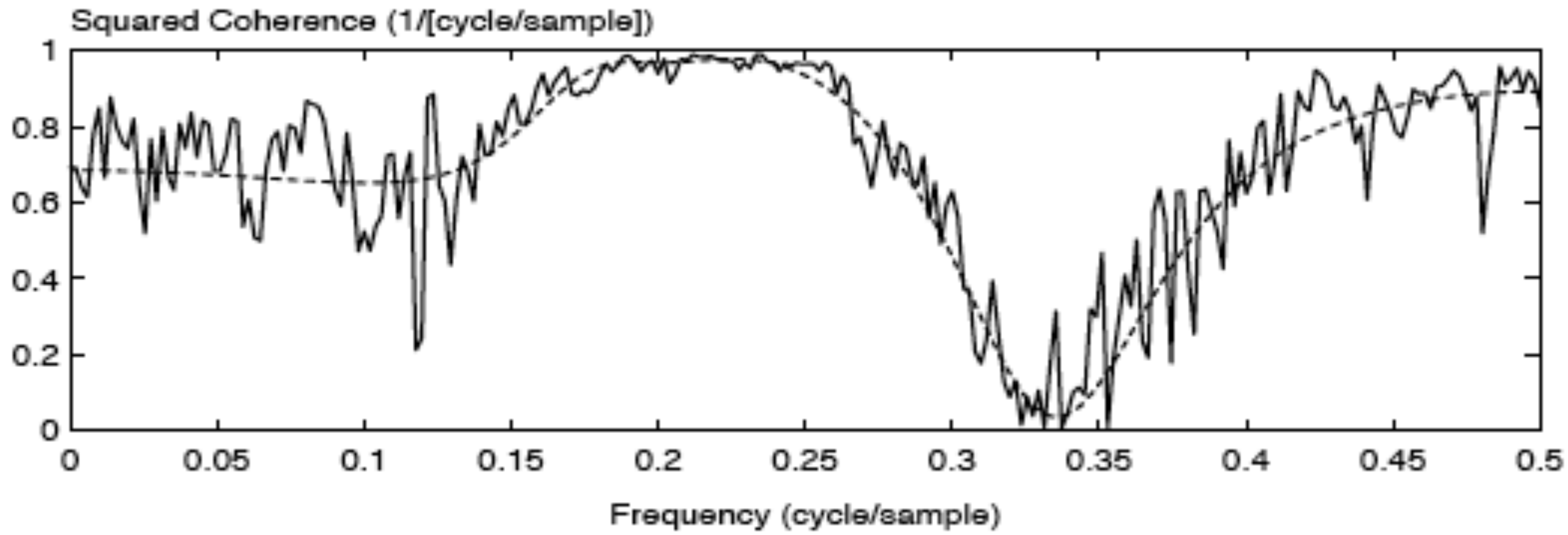
Comparison of Welch-estimated and True Power Spectra



- The Welch-FFT estimator is a well-known and widely used, classical, periodogram algorithm based on the Fast Fourier Transform (analogous to computing spherical harmonics)
- The Welch estimates are noisy because the number of parameters being estimated is large, given the 4096 samples of data

James.V.White@gmail.com

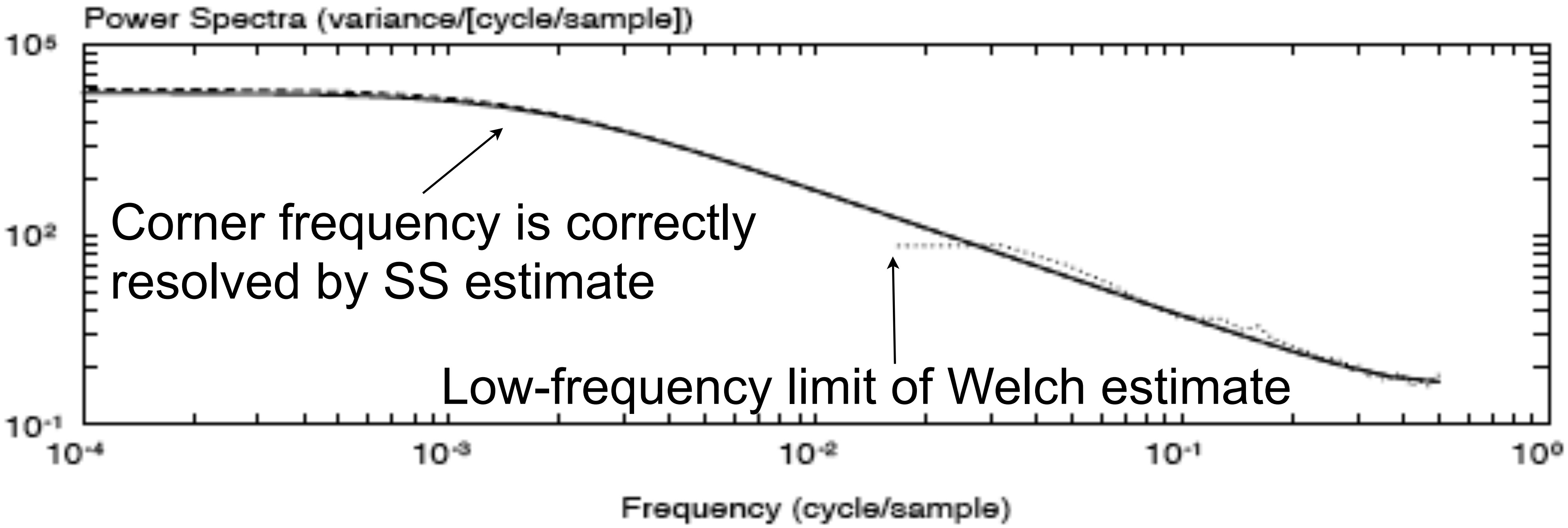
Comparison of Welch-estimated and True Coherence



- Welch estimation errors are much larger than state-space errors because the number of parameters being estimated is so much larger

James.V.White@gmail.com

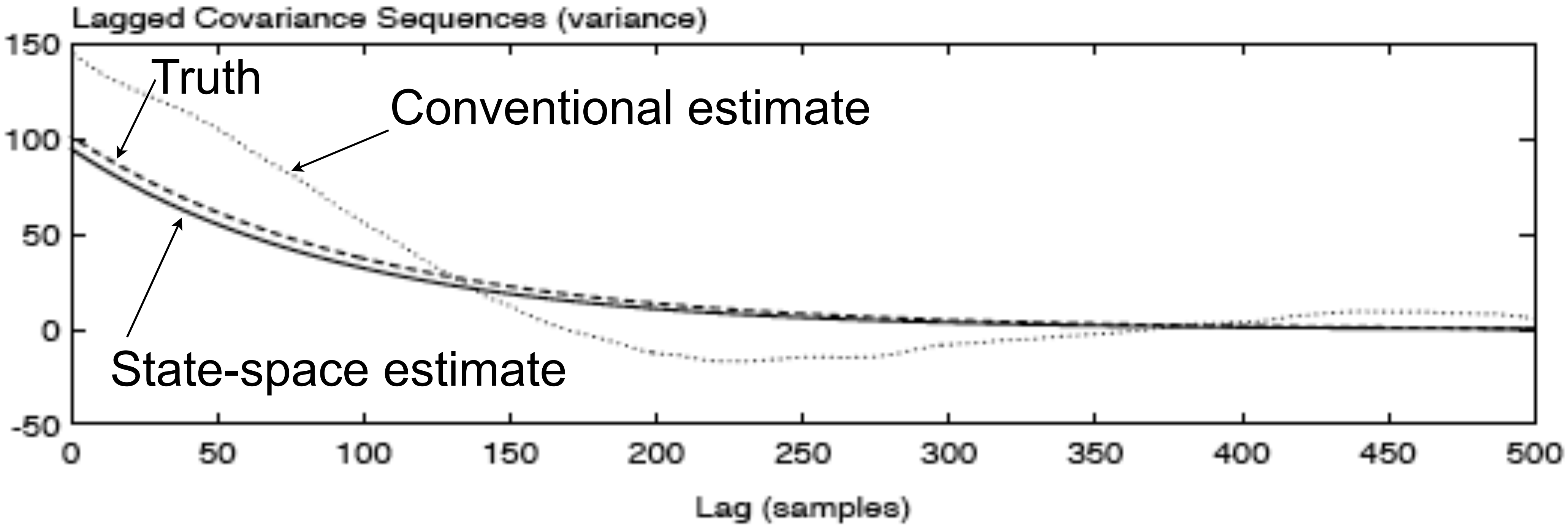
Comparison of Low-Frequency Resolutions



- All estimates are based on a sample path of 1500 samples from a first-order Markov process with a correlation length of 100 samples.
- The Welch method averages spectra based on 64-point data segments to reduce noise, given the 1500 samples of data. This smoothing limits the low-frequency resolution.
- The SS (state-space) estimate is based on a *pool of 30 data segments*, each 50 samples long. This is sufficient to model the low-frequency behavior of the Markov process.

James.V.White@gmail.com

Comparison of Estimated Correlation Functions



- Original data sequence: 1500 samples from a first-order Markov process having a correlation distance of 100 samples and a variance of 100
- Dashed curve is the true correlation function of the generating process
- Dotted curve is conventional sample correlation function based on 1500 samples
  - Highly inaccurate because data are correlated and the number of parameters being estimated is comparable to the number of data in the data sequence
- Solid curve is state-space estimate based on *pooling thirty short sequences*, each 50 samples long

James.V.White@gmail.com

Other Uses for State-space Analysis

- Efficient rigorous tests of the *Gaussian hypothesis*
  - Based on Kalman filtering and quantile-quantile plots of standardized innovation sequences
- Easy generation of *synthetic along-track data* that have known power spectra and autocorrelation functions
- Specifying *Attenuated White-noise Processes*, which are 2-D isotropic fields that are consistent with 1-D along-track spectra and correlations
  - Both flat-sky and round-sky models are supported
- *Cross-correlation analysis* of *n* data channels (sequences of *n*-vectors)
  - Optimal estimation of a common signal in multiple channels via a state-space model for all cross-correlations between the channels
- Design of *optimal matched filters* for detecting signals of known shape in colored noise
- Design of *optimal classifiers* for classifying different regions of the sphere according to spectrum or correlation criteria
  - Active vs benign regions
  - Regions with high power in specified frequency band

James.V.White@gmail.com