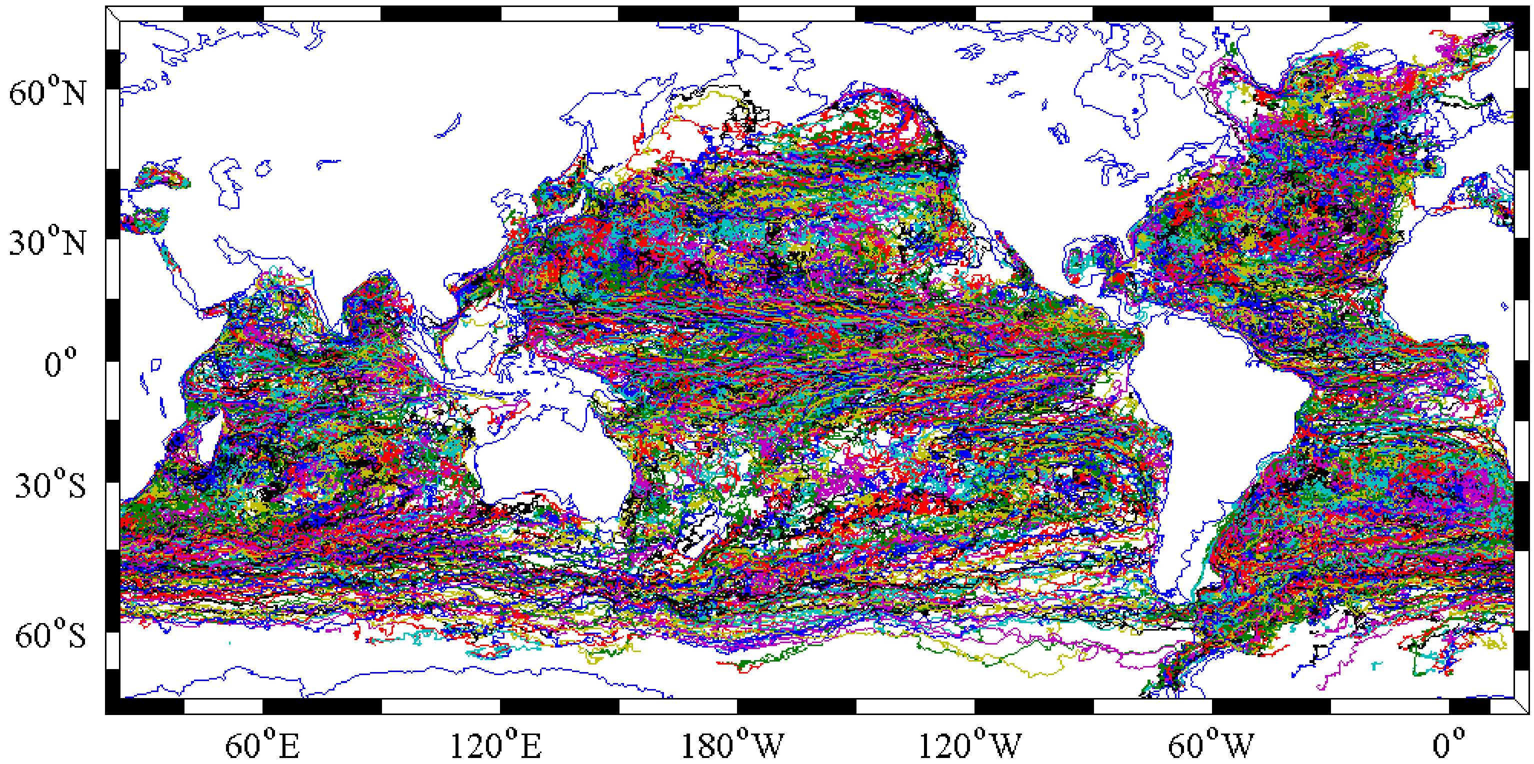
LOCALLY-STATIONARY MODELLING OF OCEANOGRAPHIC SPATIOTEMPORAL DATA

Adam M. Sykulski
Department of Statistical Science
University College London

Joint work with:
Sofia C. Olhede (University College London)
Jonathan M. Lilly and Jeffrey Early (NorthWest Research Associates)
Rick Lumpkin (AOML-NOAA, Miami)

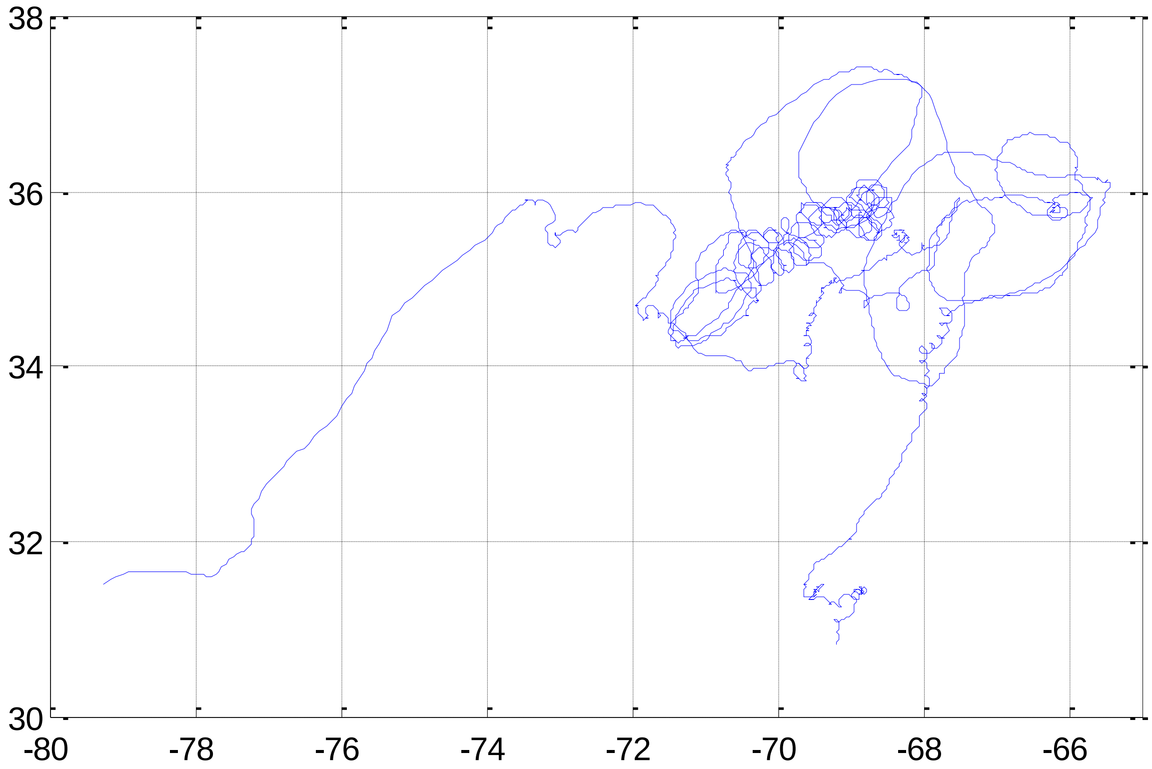
Global Drifter Program



- 10,000+ drifters
- Data going back to 1979
- Over 13 million data points

Objective → Useful Summary Statistics

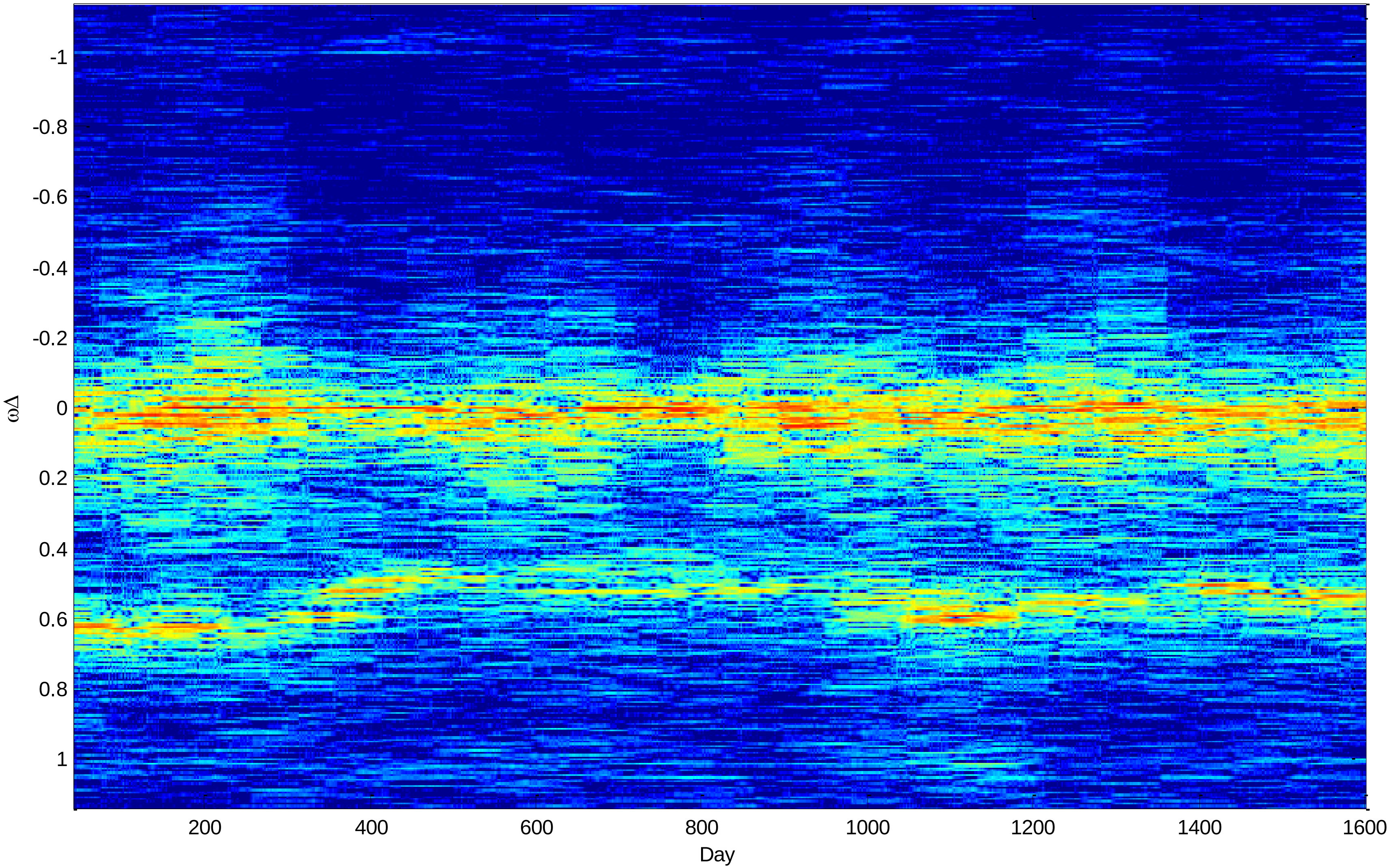
Single drifter analysis



Model Lagrangian trajectories as

complex-valued velocity time series: $z_t = x_t + iy_t$

Single drifter analysis



Spectrogram of velocity time series

Statistical Modelling

Locally-stationary process

- We combine two stochastic processes:
  - High-frequency oscillations modelled as a complex-valued OU process (Arató, 1962)
  - Low-frequency background process modelled as a Matérn process (Stein, 1999)
- The locally-stationary model then has a power spectrum given by:

$$S(\omega;\theta) = \frac{A^2}{(\omega-\omega_0)^2+\lambda^2} + \frac{B^2}{(\omega^2+h^2)^\delta}$$

- A maximum of 6 parameters need to be estimated meaning we can estimate these parameters over fine timescales

Sykulski, Olhede, Lilly and Danioux (2013)
Lagrangian Time Series Models for Ocean Surface Drifter Trajectories
arXiv preprint, arXiv:1312:2923

Statistical Estimation

The Whittle Likelihood

- We can estimate parameters for a given dataset by using the Whittle Likelihood (Whittle, 1953):

$$l(\theta) = -\sum_{\omega}\left(\frac{\hat{S}(\omega;\{Z_t\})}{S(\omega;\theta)} + \log S(\omega;\theta)\right)$$

$$\hat{\theta} = \max_{\theta}(l(\theta))$$

- Preferable (asymptotically) to other spectral methods such as least-squares on the spectral slopes
- Time domain methods typically require inversion of covariance matrices which is $O(N^3)$ as opposed to $O(N \log N)$ for Whittle

Statistical Estimation

Sampling effects and semi-parametric estimation

Problem: Data is interpolated onto a regularly grid, which affects the estimate of $\hat{S}(\omega;\{Z_t\})$, particularly at high frequencies – also affected by rounding and measurement uncertainty

Solution: Remove the high frequencies from likelihood estimation – our model is now semi-parametric

$$l(\theta) = \sum_{\omega \epsilon \Omega^{(w)}} \left(\frac{\hat{S}(\omega;\{Z_t\})}{S(\omega;\theta)} + \log S(\omega;\theta) \right)$$

- We use a reduced range of frequencies, $\Omega^{(w)}$, in the estimation
  - We cut out high frequencies due to the sampling effects
  - We only use one side of the spectrum (the side of the inertial oscillation) to ignore correlations between positive and negative frequencies

Statistical Estimation

Sampling effects and semi-parametric estimation

- We still need to estimate $S(\omega;\theta)$ correctly in $\Omega^{(w)}$
- The periodogram is often used as an estimate of the spectrum, it is the magnitude of the discrete Fourier Transform squared:

$$\hat{S}(\omega;\{Z_t\}) = \frac{\Delta}{N}\left|\sum_{t=1}^{N}(Z_t-\mu)e^{-it\omega}\right|^2$$

Problem: $\mathrm{E}(\hat{S}(\omega;\{Z_t\})) \neq S(\omega;\theta_0)$!!!

- Can lead to badly biased parameter estimates
- This is because of aliasing and leakage
- The periodogram is in fact the convolution of the true spectrum with the Fejér Kernel (Percival and Walden, 1993)

Statistical Estimation

Sampling effects and semi-parametric estimation

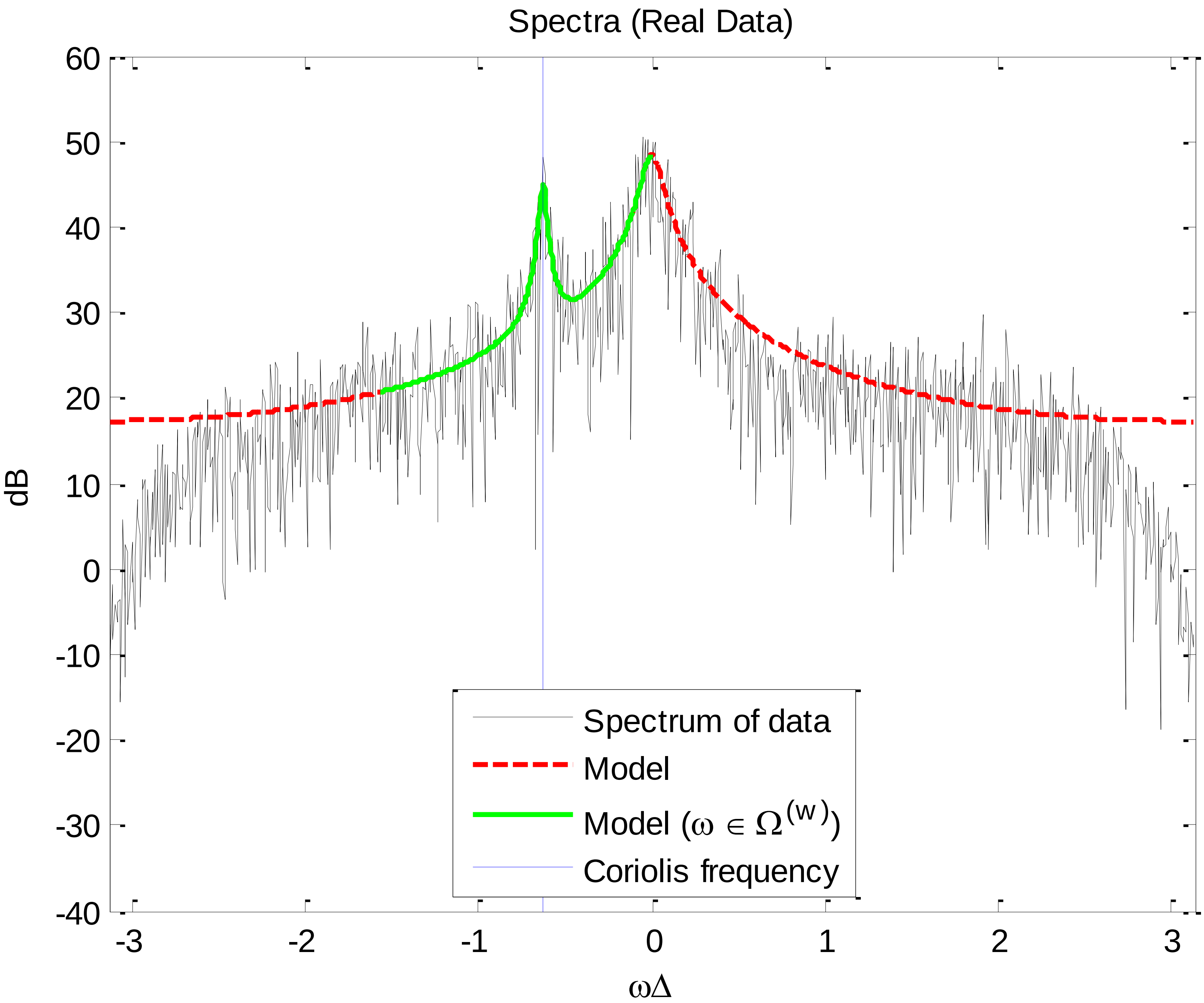
Solution: Avoid tapering and infinite summations by making use of the fact that we know the autocovariance sequence!

$$\mathrm{E}\left(\hat{S}(\omega;\{Z_t\})\right) = \sum_{\tau=-(N-1)}^{N-1} E(\hat{s}_\tau)e^{-i\omega\tau} = \sum_{\tau=-(N-1)}^{N-1} \left(1-\frac{|\tau|}{N}\right) s_\tau e^{-i\omega\tau}$$

and then:

$$l(\theta) = \sum_{\omega \epsilon \Omega^{(w)}} \left(\frac{\hat{S}(\omega;\{Z_t\})}{\mathrm{E}\left(\hat{S}(\omega;\{Z_t\})\right)} + \log\left(\mathrm{E}\left(\hat{S}(\omega;\{Z_t\})\right)\right)\right)$$

Sykulski, Olhede, Lilly and Early (2013)
The Whittle Likelihood for Complex-Valued Time Series
arXiv preprint, arXiv:1306:5993


Spectra (Real Data)
dB
60
50
40
30
20
10
0
-10
-20
-30
-40
-3
-2
-1
0
1
2
3
ωΔ
Spectrum of data
Model
Model (ω ∈ Ω(w))
Coriolis frequency

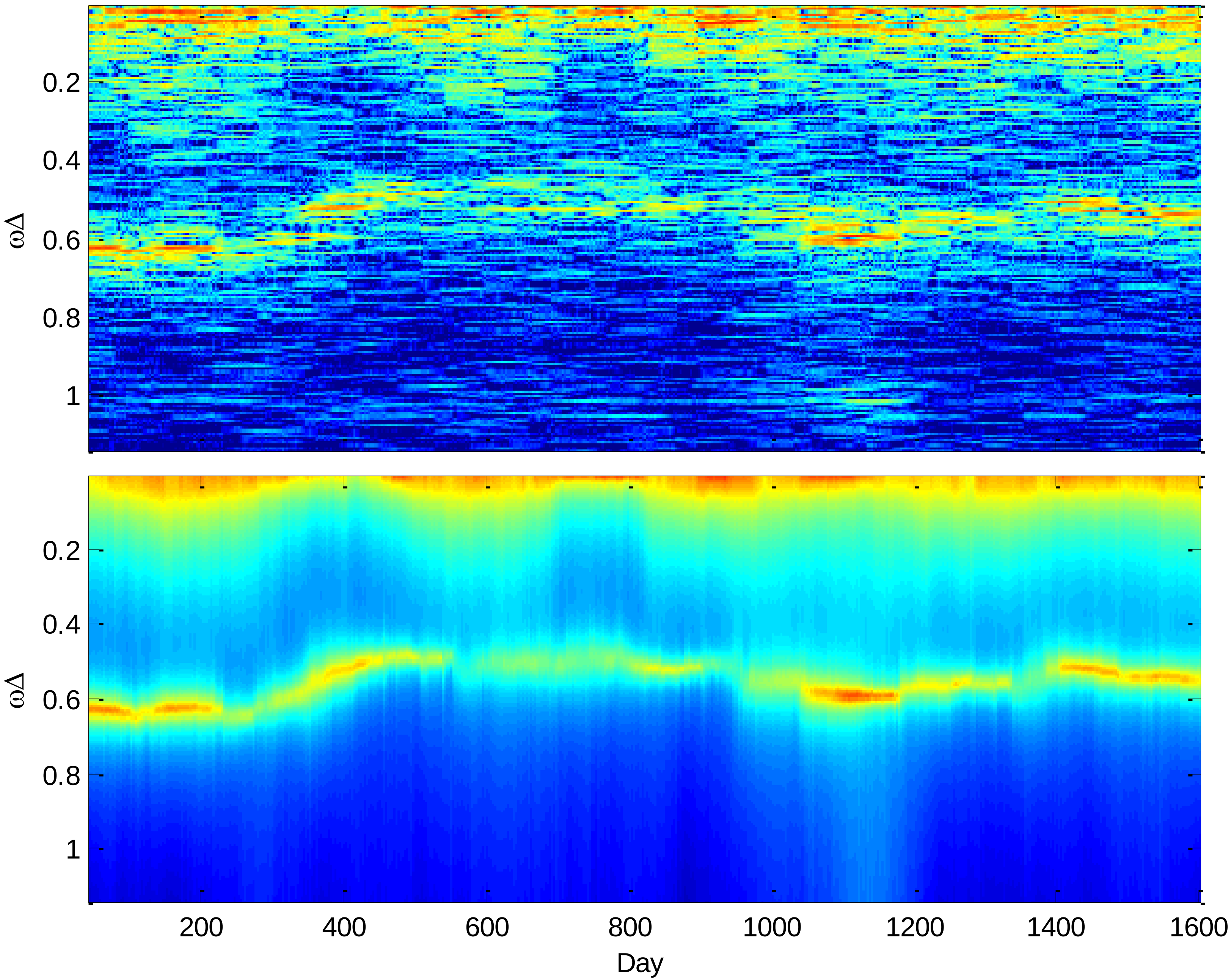
Statistical Modelling

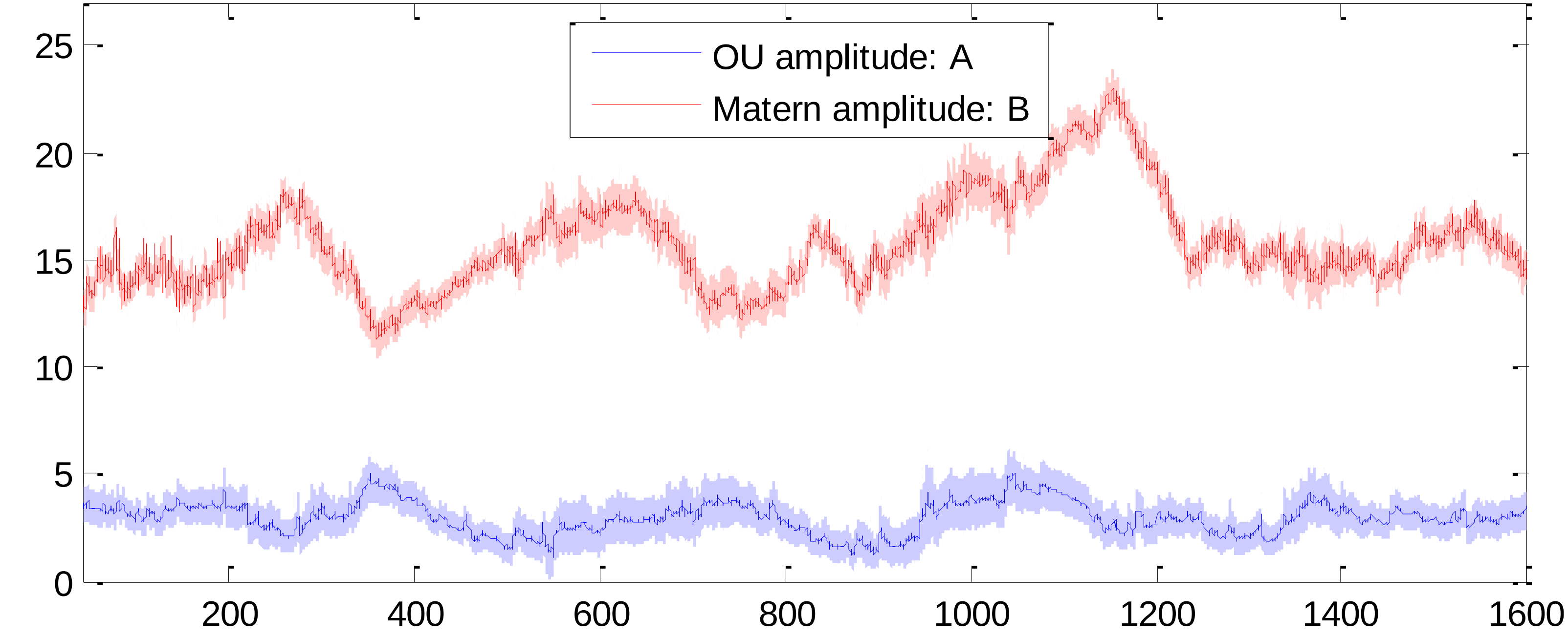
Time Variability

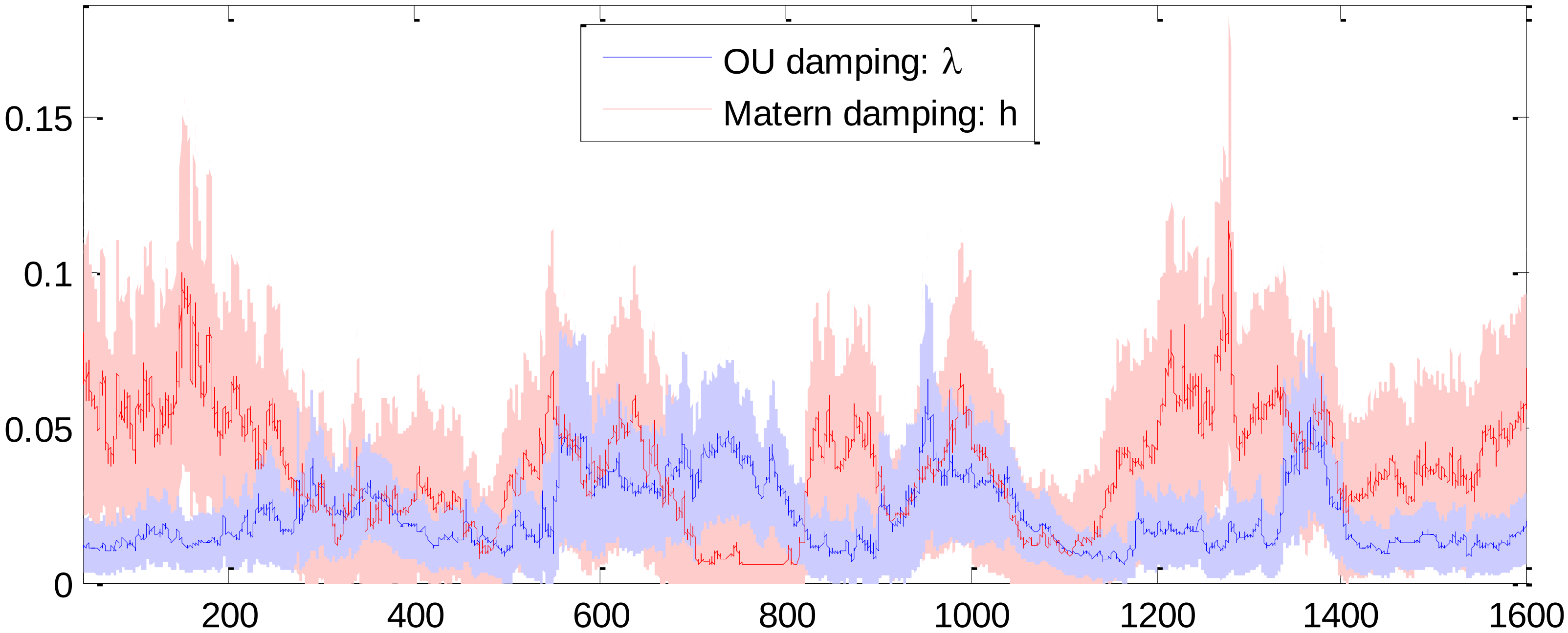
- As the drifter moves in space the characteristics change. Use time-varying model:

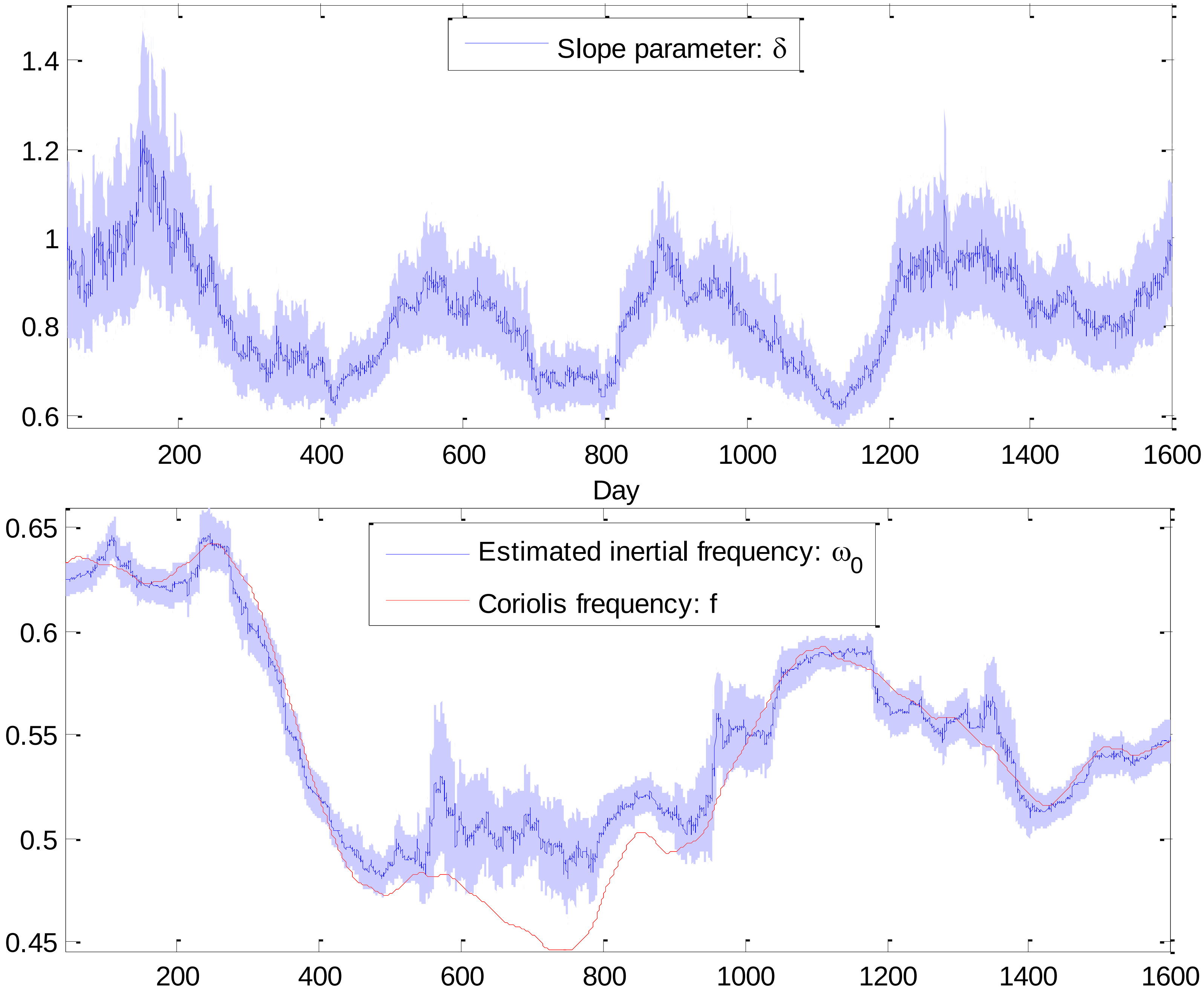
$$S(\omega;\theta(t)) = \frac{A(t)^2}{(\omega-\omega_0(t))^2+\lambda(t)^2} + \frac{B(t)^2}{(\omega^2+h(t)^2)^{\delta(t)}}$$

- To estimate $\theta(t)$ we use local complex Whittle likelihood by using sliding windows over the data (Dahlhaus, 1997)
- This is a semi-parametric model
- Setting the right window length is non-trivial


0.2
0.4
ωΔ
0.6
0.8
1
0.2
0.4
ωΔ
0.6
0.8
1
200
400
600
800
1000
1200
1400
1600
Day


OU amplitude: A
Matern amplitude: B
0
5
10
15
20
25
200
400
600
800
1000
1200
1400
1600

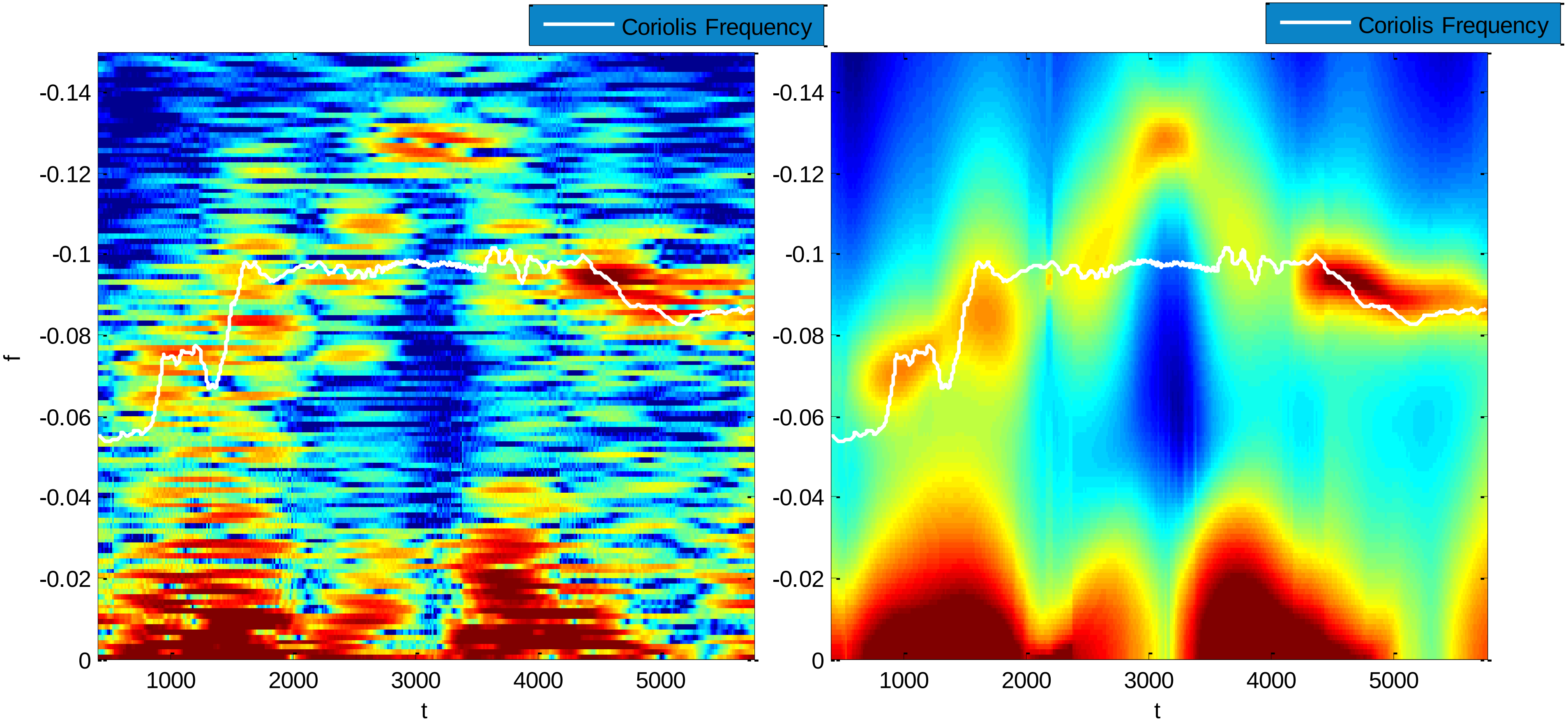

OU damping: λ
Matern damping: h
0
0.05
0.1
0.15
200
400
600
800
1000
1200
1400
1600



Slope parameter: δ
1.4
1.2
1
0.8
0.6
200
400
600
800
1000
1200
1400
1600
Day
Estimated inertial frequency: ω0
Coriolis frequency: f
0.65
0.6
0.55
0.5
0.45
200
400
600
800
1000
1200
1400
1600

Statistical Estimation

Time Variability



Real Data [See as Video]() Model