Time series analysis

Suzanne Aigrain
University of Oxford

Philosophy

- Astronomy = **state of the art (sometimes)**
  - highly sensitive instruments, huge datasets
  - exciting science

Philosophy

- Astronomy = **state of the art (sometimes)**
  - highly sensitive instruments, huge datasets
  - exciting science
- Astronomers = **dinosaurs (often)**
  - no formal statistics training
  - often use archaic or inappropriate methods

Philosophy

- Astronomy = **state of the art (sometimes)**
  - highly sensitive instruments, huge datasets
  - exciting science
- Astronomers = **dinosaurs (often)**
  - no formal statistics training
  - often use archaic or inappropriate methods
- No-one dies if we get it wrong

Philosophy

- Astronomy = **state of the art (sometimes)**
  - highly sensitive instruments, huge datasets
  - exciting science
- Astronomers = **dinosaurs (often)**
  - no formal statistics training
  - often use archaic or inappropriate methods
- No-one dies if we get it wrong

- In these lectures
  - recap basic/common concepts
    - cautionary notes on usage/implementation
  - pointers to some interesting methods not commonly used in astrophysics
    - outside my comfort zone!

Textbooks

- *Bayesian spectrum analyis and parameter estimation*, Bretthorst, Springer, 1988

  `bayes.wustl.edu/glb/book.pdf`

- *Introduction to time series and forecasting*, Brockwell & Davis, Springer, 2002 (2nd edition)

- *Time series analysis and its applications*, Shumway & Stoffer, Springer, 2006 (2nd edition)

- *Pattern recognition & machine learning*, Bishop, Springer, 2006

- *Gaussian processes for machine learning*, Rasmussen & Williams, MIT, 2006

  `www.gaussianprocess.org/gpml/chapters/RW.pdf`

- For additional reference material, data and source code see

  `http://camd08.ast.cam.ac.uk/Greatwiki/GreatStats11/TSA`

What is so special about time-series?

- A time series is an ordered sequence of observations of one or more variables
  - Uncertainty on time of observations usually (but not always) extremely well known (or at least much better than dependent variables)
  - Interval between consecutive observations (time sampling) may be constant (regular sampling) or not.

What is so special about time-series?

- A time series is an ordered sequence of observations of one or more variables
    - Uncertainty on time of observations usually (but not always) extremely well known (or at least much better than dependent variables)
    - Interval between consecutive observations (time sampling) may be constant (regular sampling) or not.
- Causality implies autocorrelation in time-series.

What is so special about time-series?

- A time series is an ordered sequence of observations of one or more variables
  - Uncertainty on time of observations usually (but not always) extremely well known (or at least much better than dependent variables)
  - Interval between consecutive observations (time sampling) may be constant (regular sampling) or not.
- Causality implies autocorrelation in time-series.
- What are we trying to do?
  - Detect / extract signal buried in noise
  - Learn about underlying physical processes
  - Predict observations
  - Make decisions about what to do next

Time-series modelling

- "Direct" modelling
  - $y_t = f(t, \boldsymbol{\theta}) + \varepsilon_t$
    - $f \equiv$ model function
    - $\boldsymbol{\theta} \equiv$ parameters
    - $\varepsilon_t \equiv$ typically IID noise
  - spectral analysis
    - $f(t, \{\mathrm{a}\}, \omega) = a_1 \sin(\omega t) + a_2 \cos(\omega t)$

Time-series modelling

- "Direct" modelling D. Hogg's lectures
  - $y_t = f(t, \boldsymbol{\theta}) + \varepsilon_t$
    - $f \equiv$ model function
    - $\boldsymbol{\theta} \equiv$ parameters
    - $\varepsilon_t \equiv$ typically IID noise
  - spectral analysis
    - $f(t, \{a\}, \omega) = a_1 \sin(\omega t) + a_2 \cos(\omega t)$

Time-series modelling

- "Direct" modelling D. Hogg's lectures
  - $y_t = f(t, \boldsymbol{\theta}) + \varepsilon_t$
    - $f \equiv$ model function
    - $\boldsymbol{\theta} \equiv$ parameters
    - $\varepsilon_t \equiv$ typically IID noise
  - spectral analysis
    - $f(t, \{\mathrm{a}\}, \omega) = a_1 \sin(\omega t) + a_2 \cos(\omega t)$

- ARMA-type models
  - $y_n = \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{prev}} + \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\varepsilon}_{\mathrm{prev}} + \varepsilon_n$
  - specific covariance properties
  - can be quasi-periodic

Time-series modelling

- "Direct" modelling — D. Hogg's lectures
  - $y_t = f(t, \boldsymbol{\theta}) + \varepsilon_t$
    - $f \equiv$ model function
    - $\boldsymbol{\theta} \equiv$ parameters
    - $\varepsilon_t \equiv$ typically IID noise
  - spectral analysis
    - $f(t, \{\mathrm{a}\}, \omega) = a_1 \sin(\omega t) + a_2 \cos(\omega t)$

- ARMA-type models
  - $y_n = \boldsymbol{\varphi}^{\mathrm{T}}\, \boldsymbol{y}_{\mathrm{prev}} + \boldsymbol{\theta}^{\mathrm{T}}\, \boldsymbol{\varepsilon}_{\mathrm{prev}} + \varepsilon_n$
  - specific covariance properties
  - can be quasi-periodic

- Gaussian Processes
  - $f \sim \mathcal{GP}(m, k)$ or $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}, \mathbf{K})$
  - $m(t, \boldsymbol{\theta}) \equiv$ mean function
  - $\mathrm{K}_{ij} = k(t_i, t_j) + \delta_{ij}\, \beta$
    - $k \equiv$ covariance function / kernel
    - $\delta_{ij}\, \beta \equiv$ white noise term

Time-series modelling

- "Direct" modelling — D. Hogg's lectures
  - $y_t = f(t, \boldsymbol{\theta}) + \varepsilon_t$
    - $f \equiv$ model function
    - $\boldsymbol{\theta} \equiv$ parameters
    - $\varepsilon_t \equiv$ typically IID noise
  - spectral analysis
    - $f(t, \{\mathrm{a}\}, \omega) = a_1 \sin(\omega t) + a_2 \cos(\omega t)$
- ARMA-type models
  - $y_n = \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{prev}} + \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\varepsilon}_{\mathrm{prev}} + \varepsilon_n$
  - specific covariance properties
  - can be quasi-periodic
- Gaussian Processes
  - $f \sim \mathcal{GP}(m, k)$ or $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}, \mathbf{K})$
  - $m(t, \boldsymbol{\theta}) \equiv$ mean function
  - $\mathrm{K}_{ij} = k(t_i, t_j) + \delta_{ij}\,\beta$
    - $k \equiv$ covariance function / kernel
    - $\delta_{ij}\,\beta \equiv$ white noise term
- State-space models
  - $y_i = \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}_i + \varepsilon_i \qquad (i = 1, \ldots, N)$
  - $\boldsymbol{x}_{ij} = \mathbf{g}_j^{\mathrm{T}} \boldsymbol{x}_{j,\mathrm{prev}} + w_{ij} \qquad (j = 1, \ldots, M)$
    - $\boldsymbol{x} \equiv$ state of system
    - $\boldsymbol{w}_i \equiv$ process noise, typically IID

Some basic concepts

Mean, variance and covariance

Let $x(t)$ represent an observation of a process X at time t.

The **variance** of X is

$$\mathrm{var}\left[X\right](t) \equiv \mathbb{E}\left[x^2(t)\right],$$

where $\mathbb{E}\left[k\right]$ represents the expectation of the quantity k.

We restrict ourselves to time-series with finite variance for all t.

Mean, variance and covariance

Let $x(t)$ represent an observation of a process X at time t.

The **variance** of X is

$$\mathrm{var}\left[X\right](t) \equiv \mathbb{E}\left[x^2(t)\right],$$

where $\mathbb{E}\left[k\right]$ represents the expectation of the quantity k.

We restrict ourselves to time-series with finite variance for all t.

The **mean function** of X is

$$\mu_X(t) = \mathbb{E}\left[x(t)\right]$$

and the **covariance function** of X is

$$\gamma_X(t', t) = \mathrm{cov}\left[x_{t'}, x_t\right] = \mathbb{E}\left[\{x_{t'} - \mu_X(t')\}\{x_t - \mu_X(t)\}\right]$$

for all t and t'.

Stationarity

X is **(weakly) stationary** if $\mu_X(t)$ is independent of t, and $\gamma_X(t+\tau, t)$ is independent of t for all τ.

Difference signals with long-term trends to achieve stationarity: $y_t = x_t - x_{t-1}$.

Autocovariance and autocorrelation functions

If X is stationary, the **autocovariance function** (ACV) of X at lag τ is

$$\gamma_X(\tau) = \operatorname{cov}\left[X(t+\tau), X(t)\right],$$

where $\operatorname{cov}[x, y]$ denotes the covariance of two random variables x and y, and the **autocorrelation function** (ACF) of X at lag τ is

$$\rho_X(\tau) = \frac{\gamma_X(\tau)}{\gamma_X(0)}.$$

Sample mean, autocovariance and autocorrelation

Let $\mathsf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ represent a set of N observations of a process X. The **sample mean** of x is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

Sample mean, autocovariance and autocorrelation

Let $\mathsf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ represent a set of N observations of a process X. The **sample mean** of x is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

the **sample autocovariance function** of x is

$$\hat{\gamma}_h = \frac{1}{N} \sum_{t=1}^{N-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

Sample mean, autocovariance and autocorrelation

Let $\mathsf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ represent a set of N observations of a process X. The **sample mean** of x is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

the **sample autocovariance function** of x is

$$\hat{\gamma}_h = \frac{1}{N} \sum_{t=1}^{N-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

and the **sample autocorrelation function** of x is

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}.$$

Both $\hat{\gamma}_h$ and $\hat{\rho}_h$ are defined for integer lags h, over the interval $-N < h < N$.

Notes on using the ACF

- Sample ACF becomes unreliable as $h \to N$
  - Stick to $h \leq N/4$

- `lags, acf, lines, axis = pylab.acorr(a, maxlags = N/4)`

Notes on using the ACF

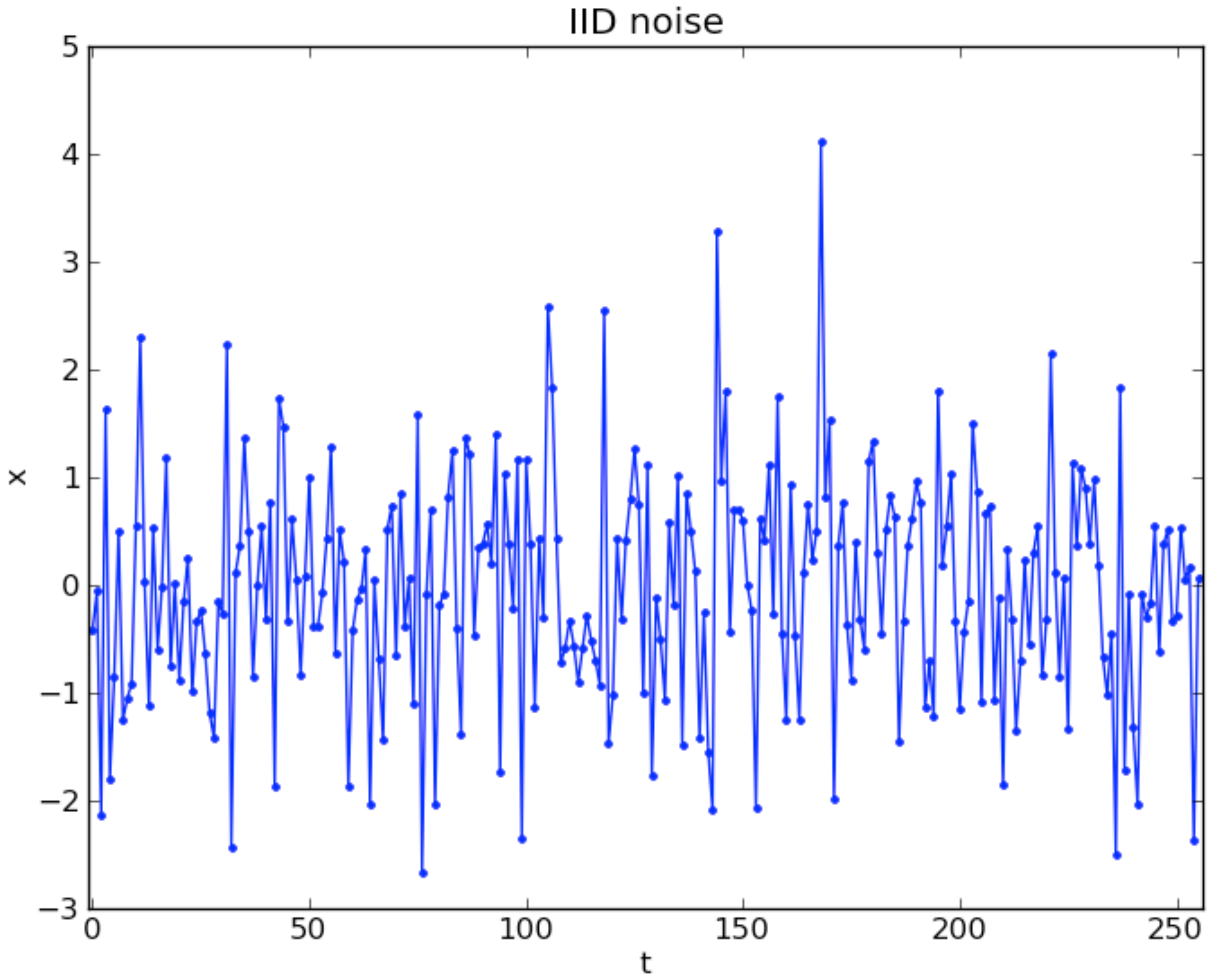
- Sample ACF becomes unreliable as $h \to N$
  - Stick to $h \leq N/4$

- `lags, acf, lines, axis = pylab.acorr(a, maxlags = N/4)`

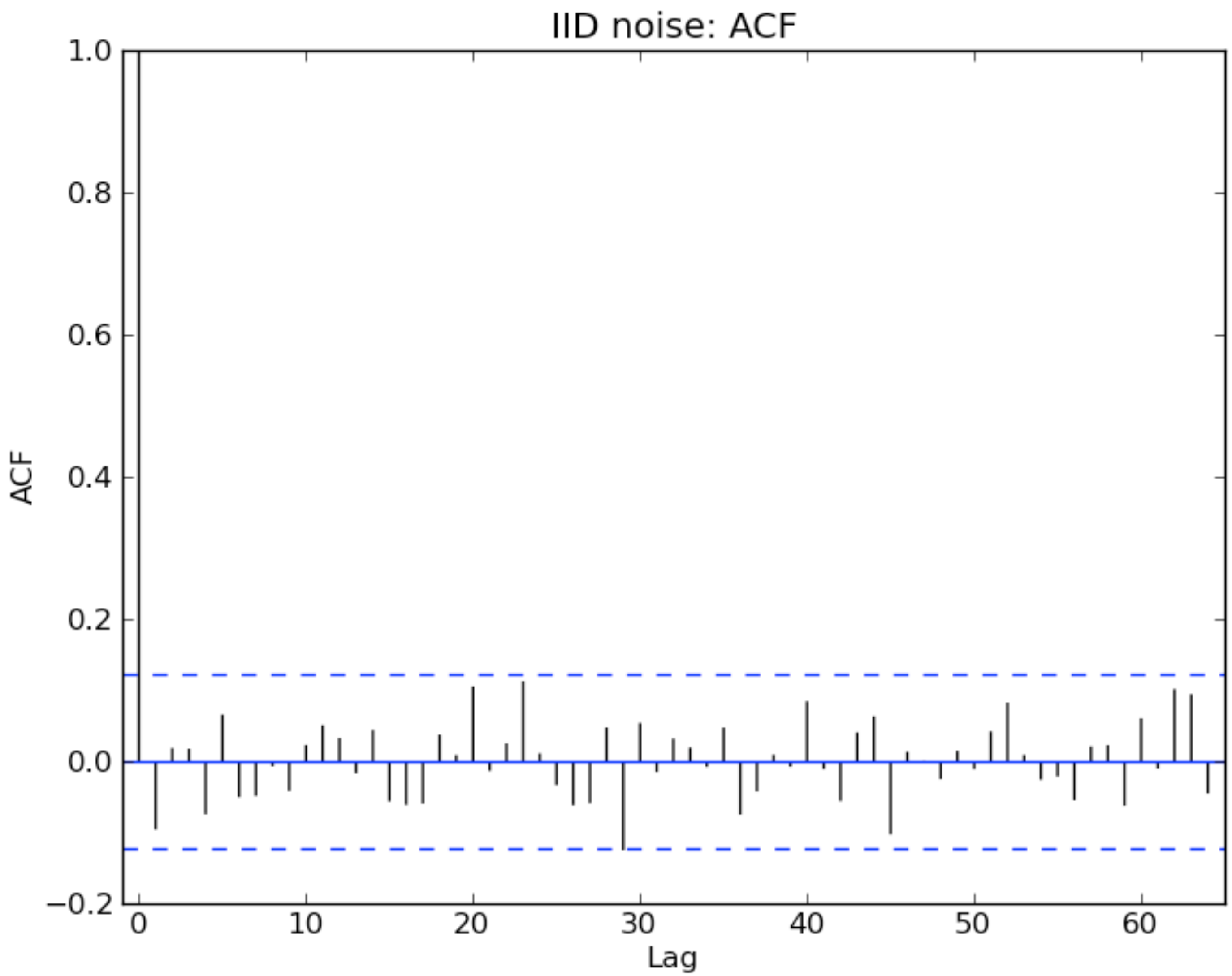
- Testing the IID hypothesis
  - For non-zero lag, ACF of IID noise is $\mathcal{N}(0, N^{-1})$
  - 95% confidence interval: $\pm\ 1.96\ N^{-1/2}$
    - if > 2 of the first 40 sample ACF values lie outside these bounds, reject IID hypothesis

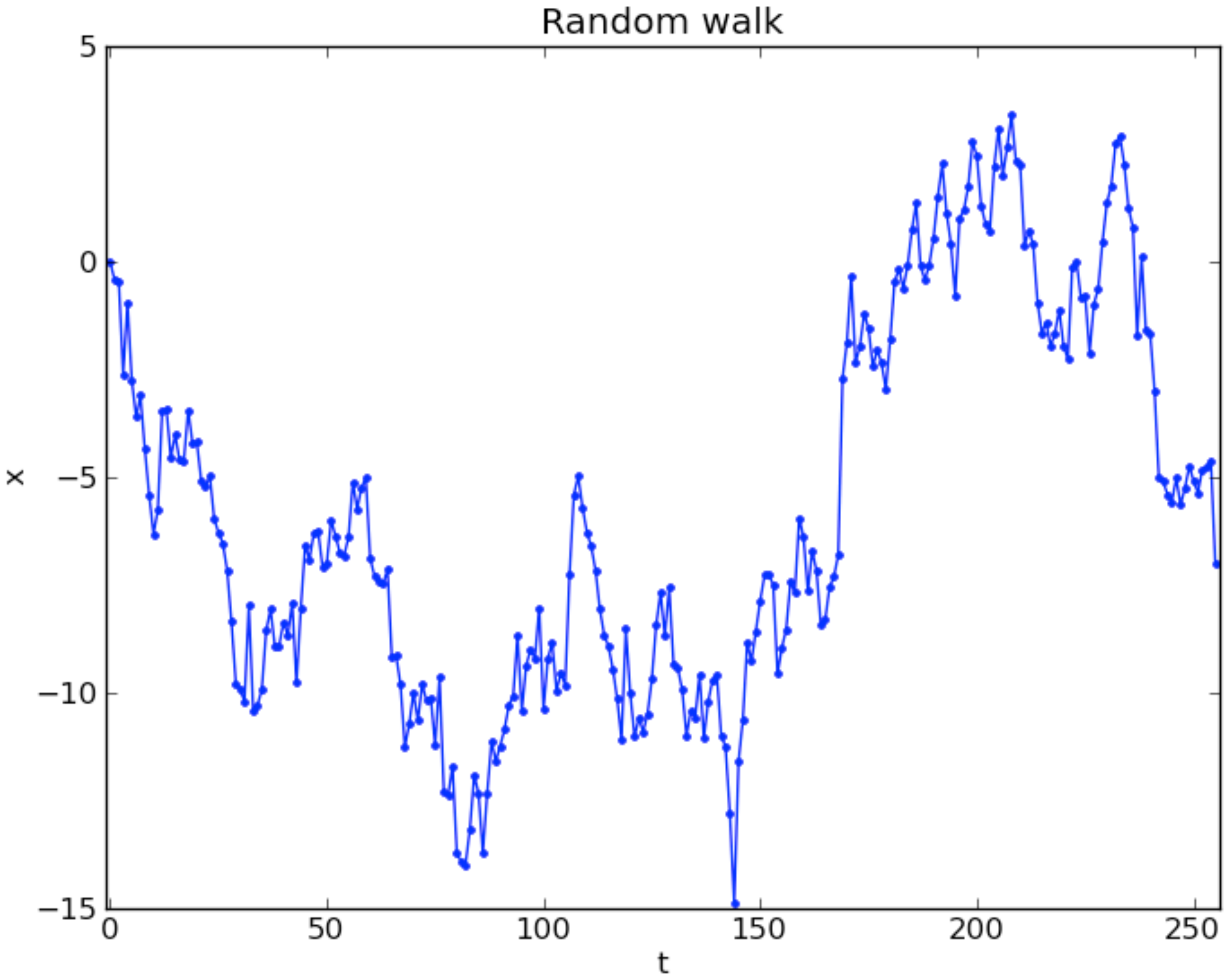
Simple examples 1: ACF

(see script `examples_1.py`)

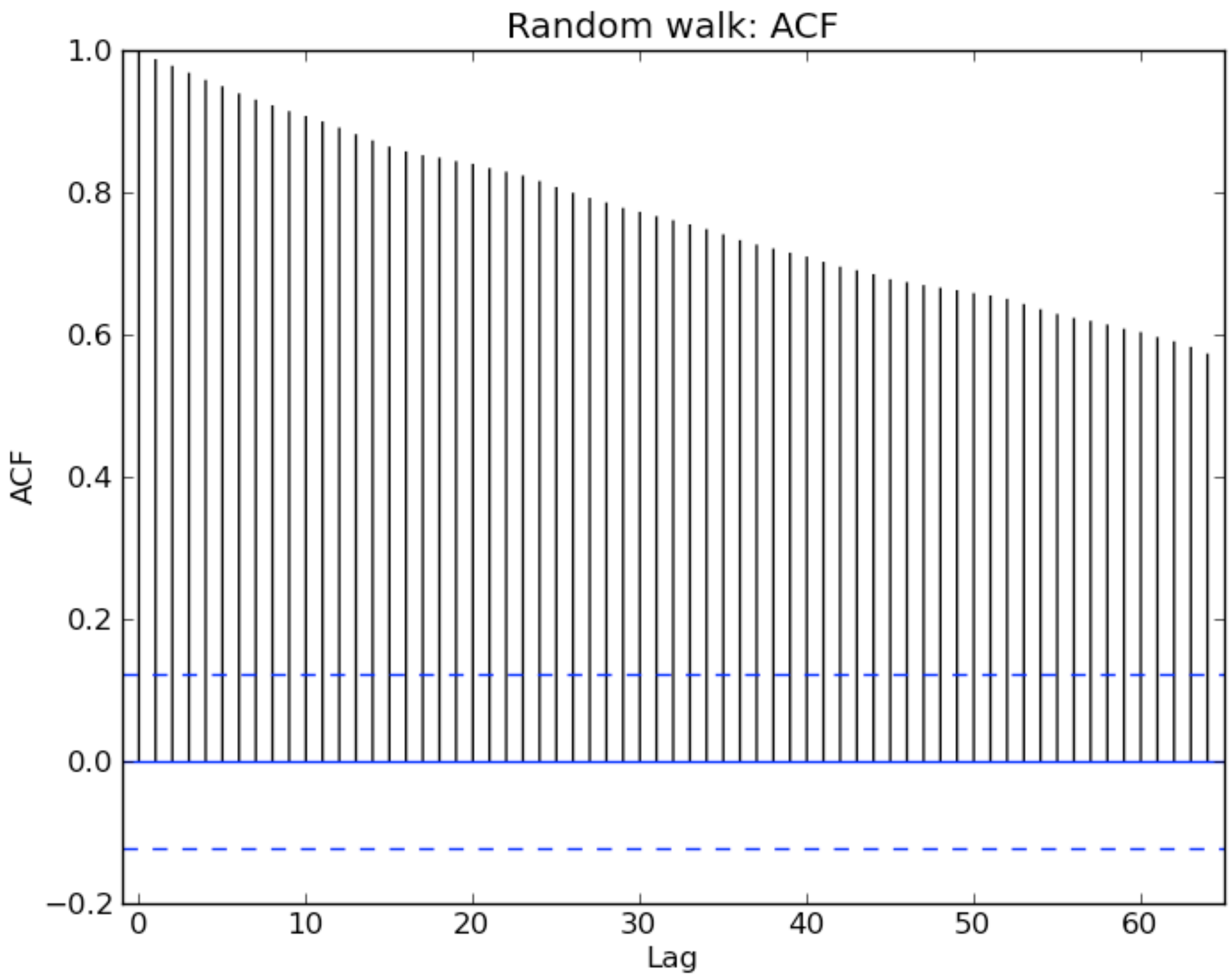


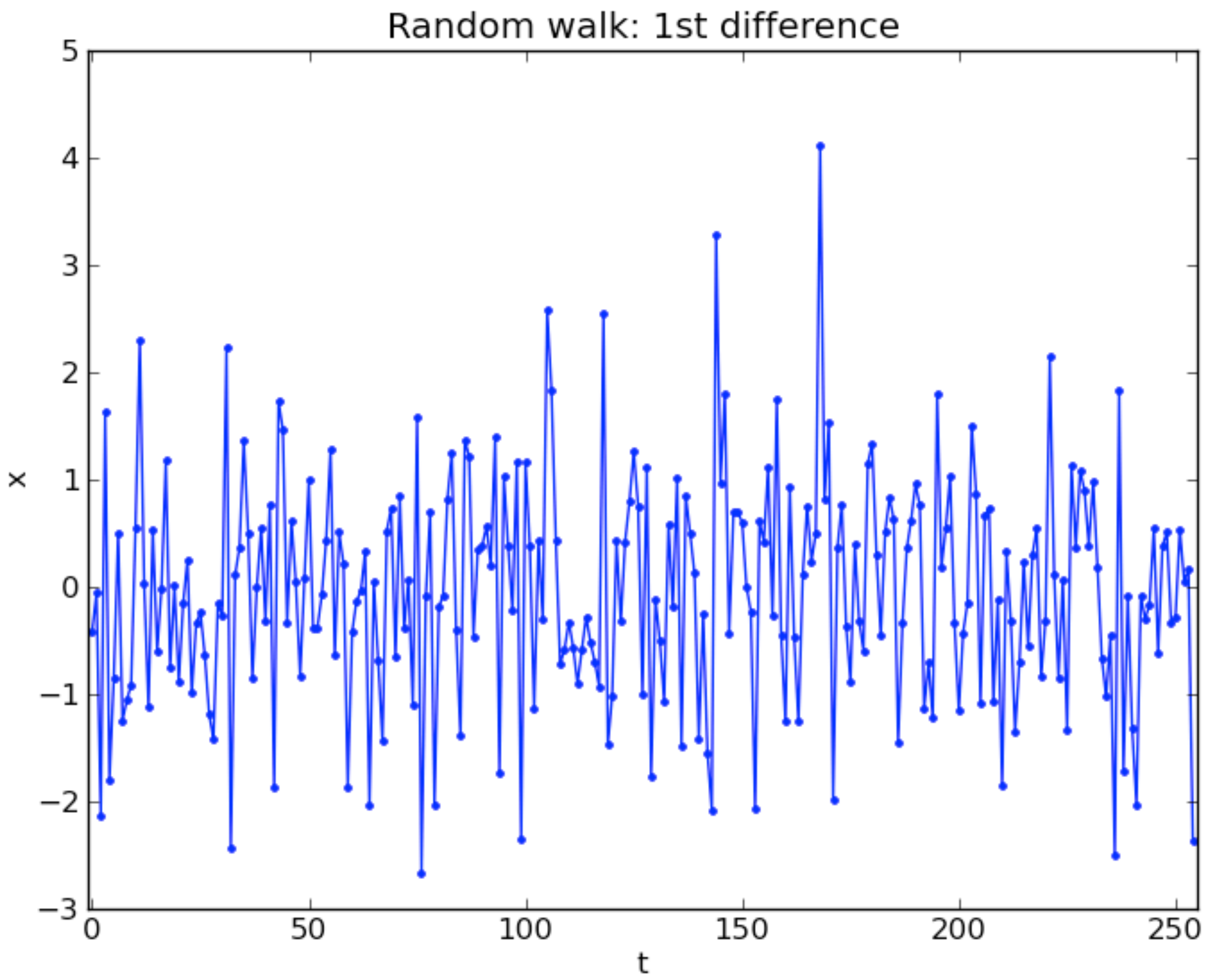
IID noise
x
t
5
4
3
2
1
0
−1
−2
−3
0
50
100
150
200
250

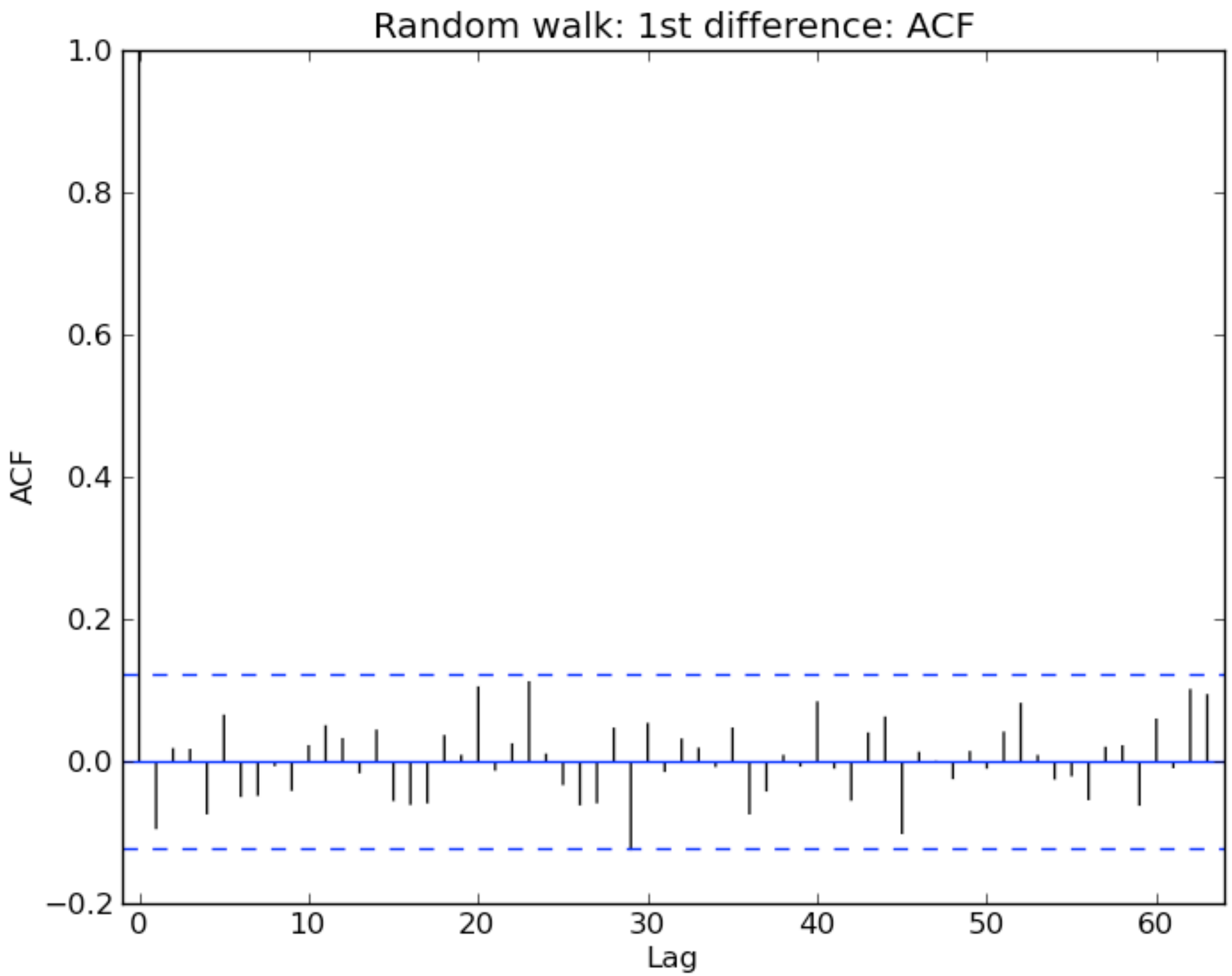

IID noise: ACF
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60



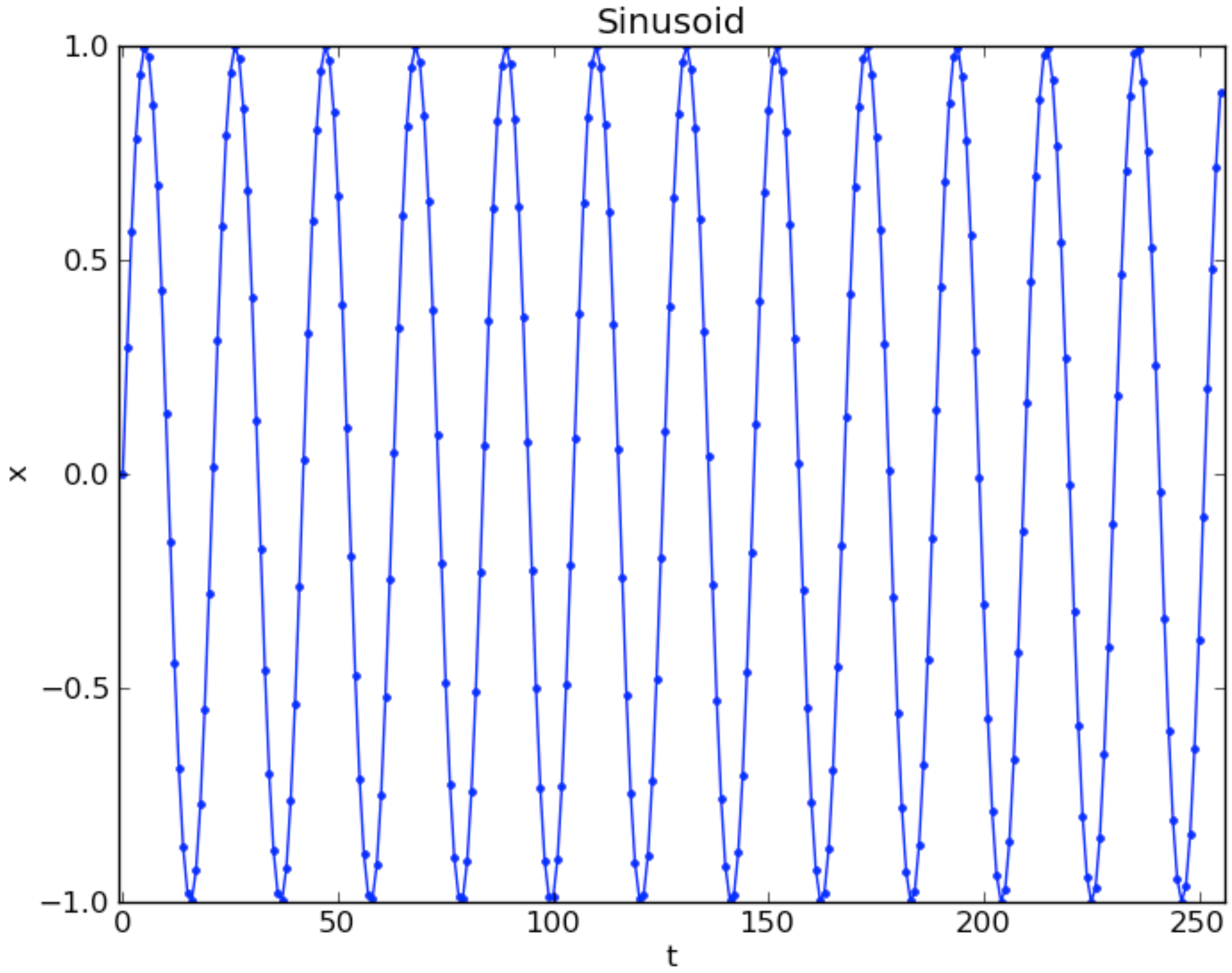
Random walk
5
0
−5
−10
−15
x
0
50
100
150
200
250
t

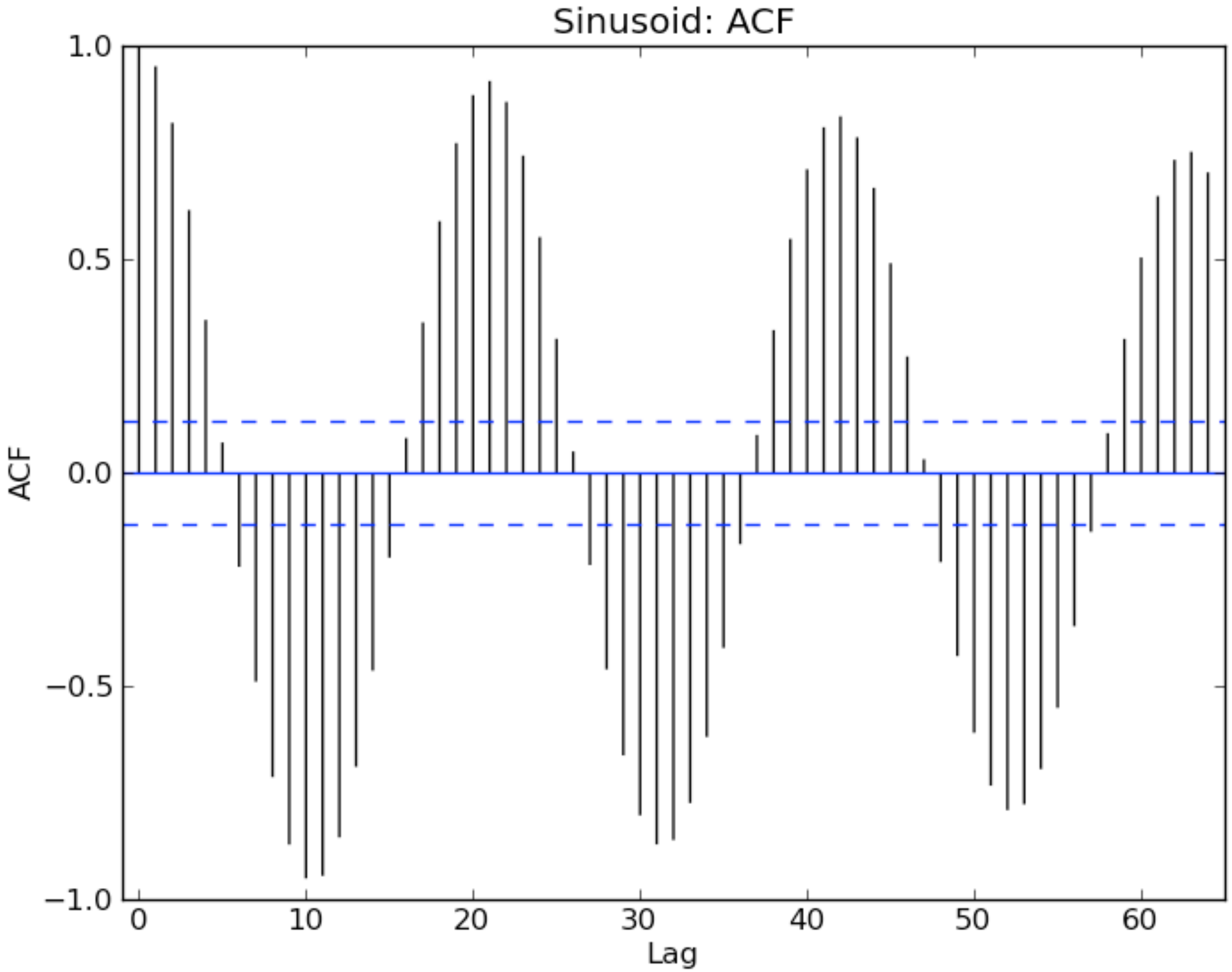

Random walk: ACF
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

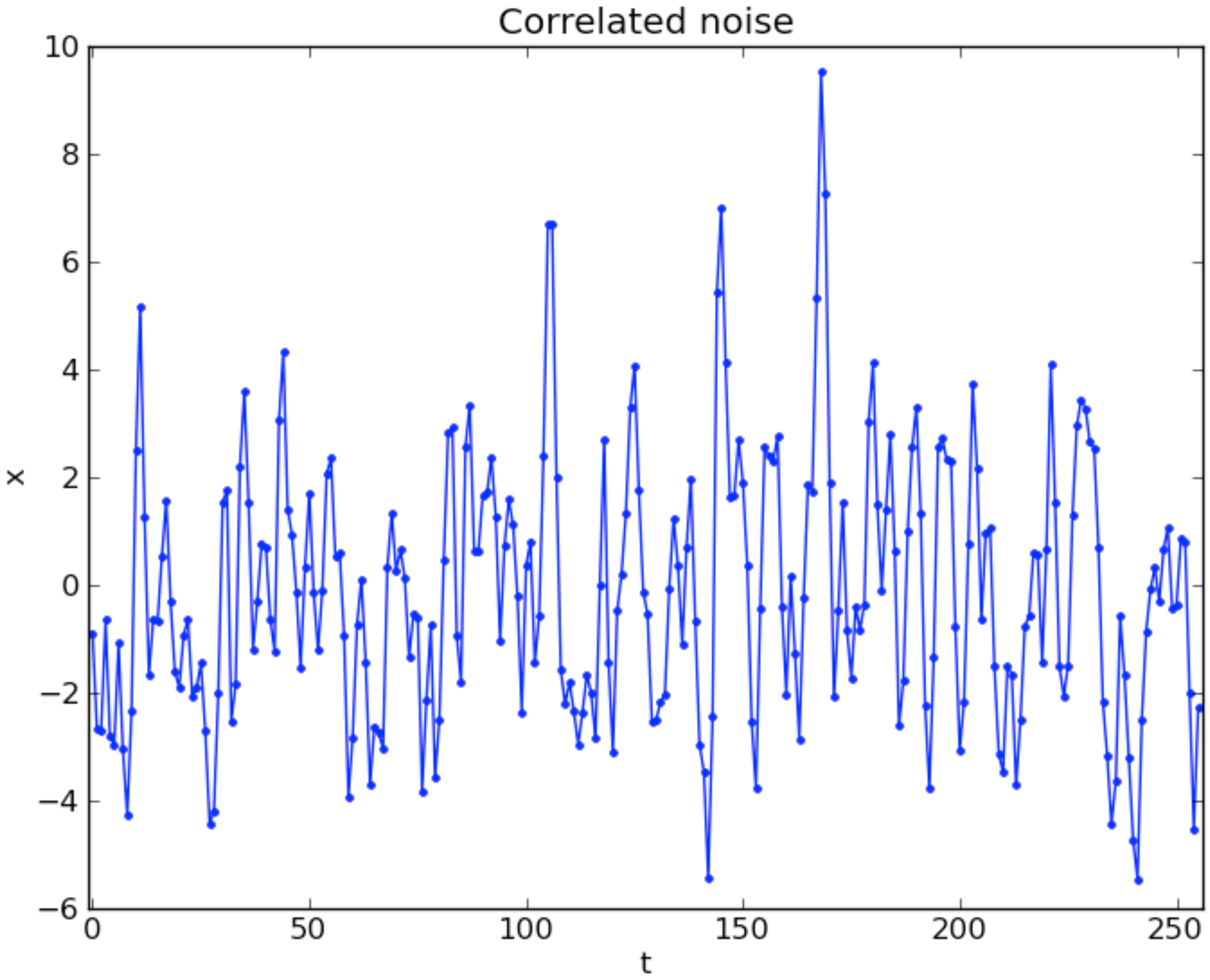

Random walk: 1st difference
x
t
5
4
3
2
1
0
−1
−2
−3
0
50
100
150
200
250

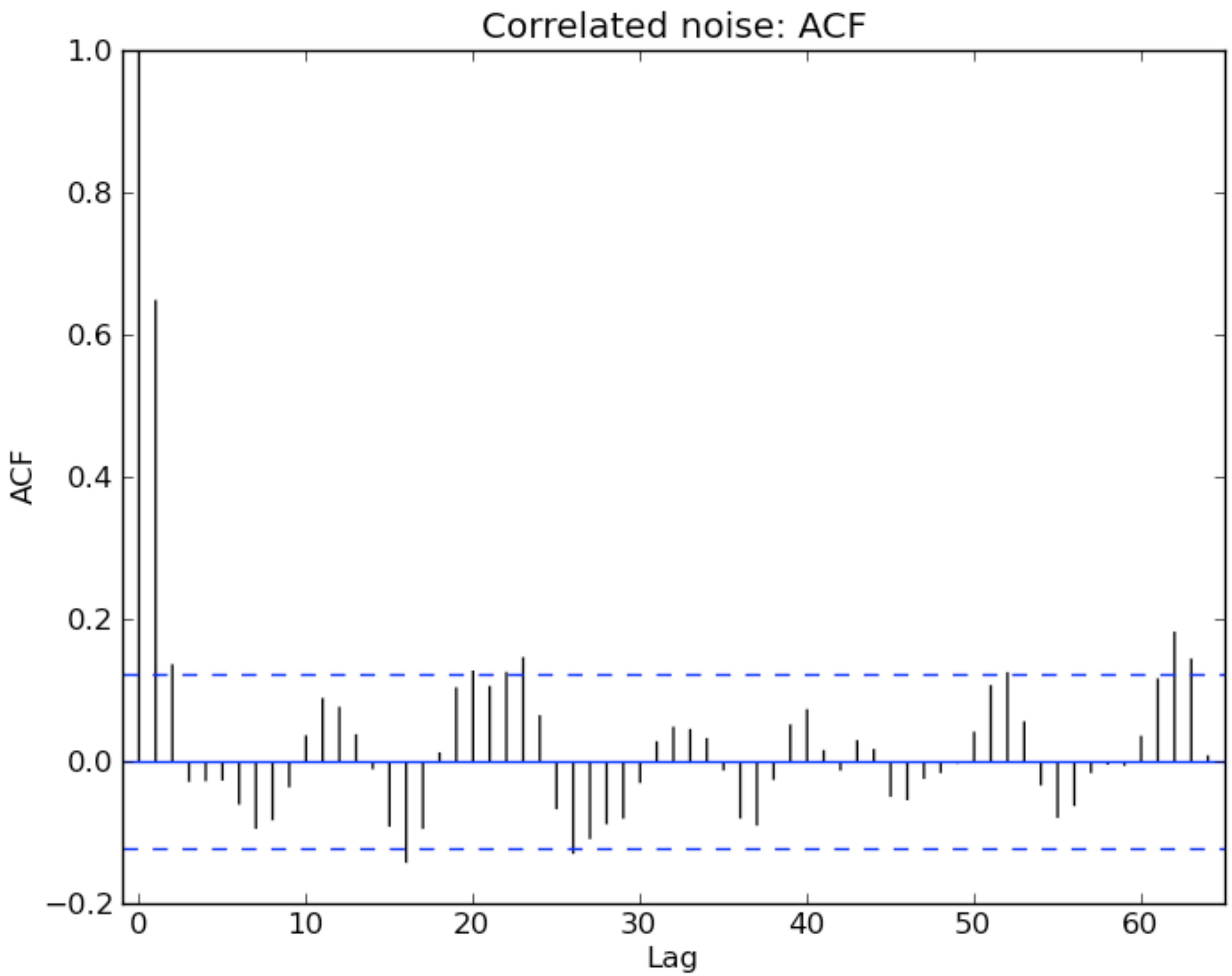


Random walk: 1st difference: ACF
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60


Sinusoid
1.0
0.5
0.0
−0.5
−1.0
x
0
50
100
150
200
250
t


Sinusoid: ACF
ACF
Lag
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
50
60

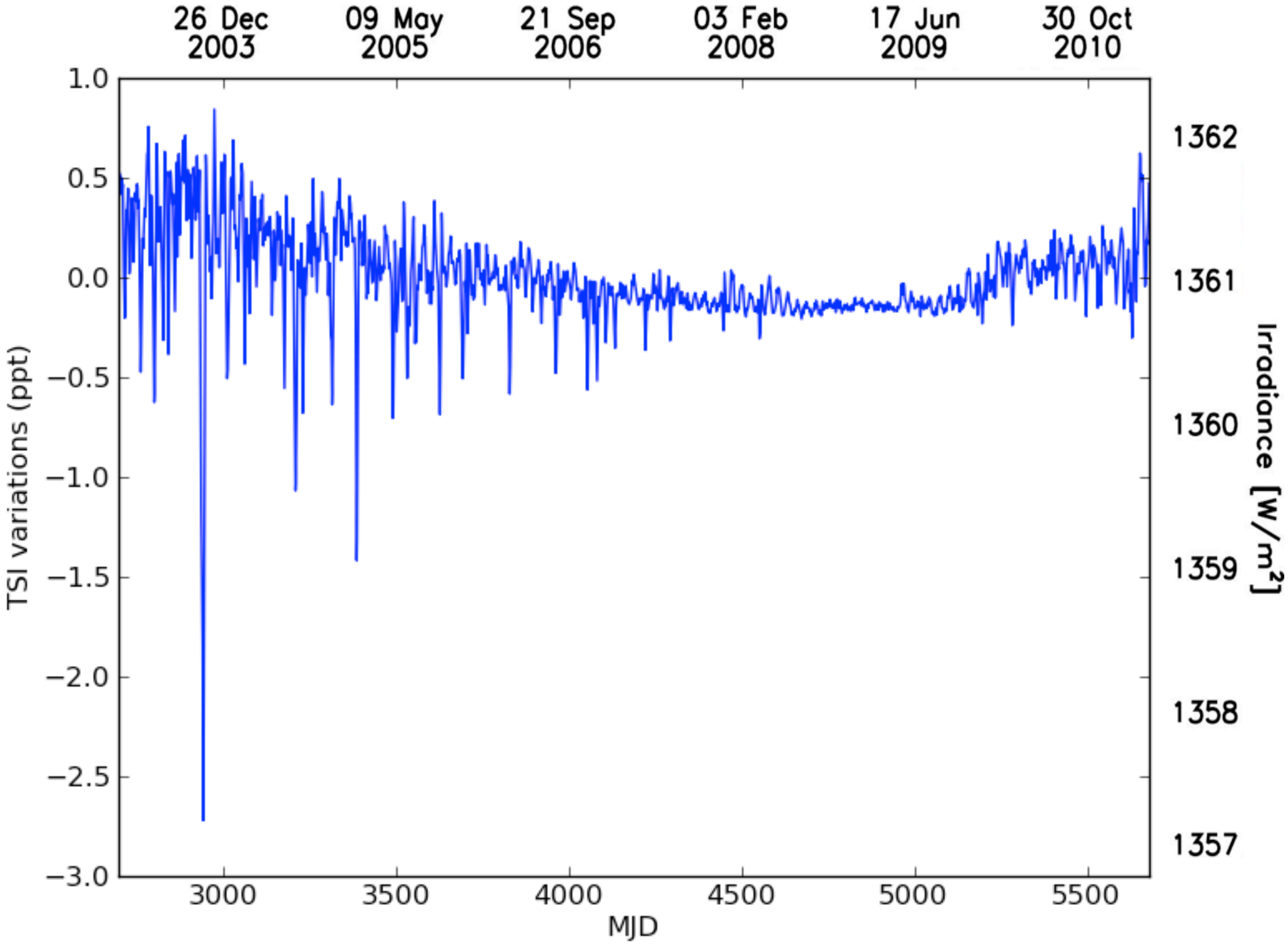

Correlated noise
x
t
10
8
6
4
2
0
−2
−4
−6
0
50
100
150
200
250



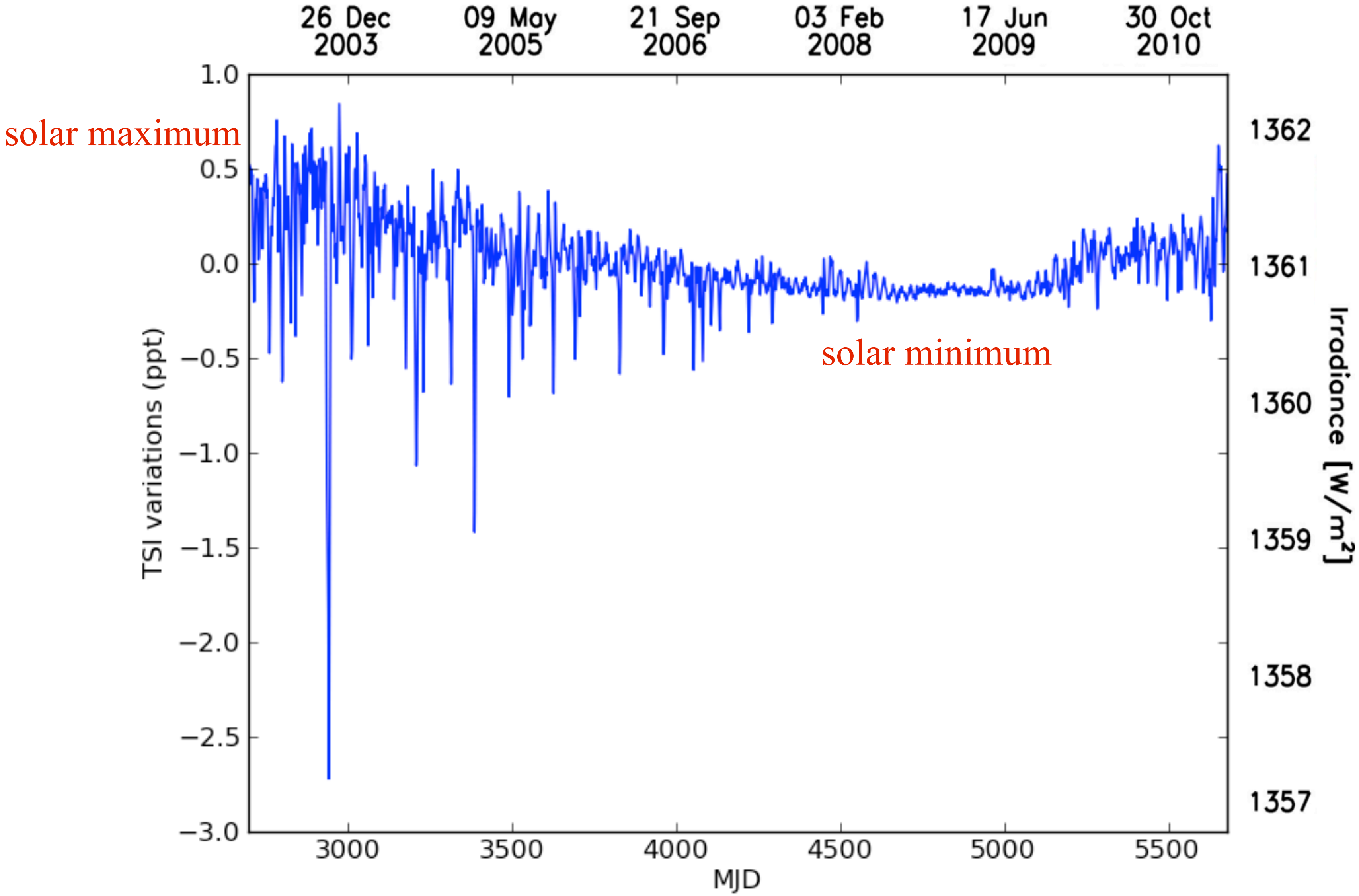
Correlated noise: ACF
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

Exercise 1: Total solar irradiance

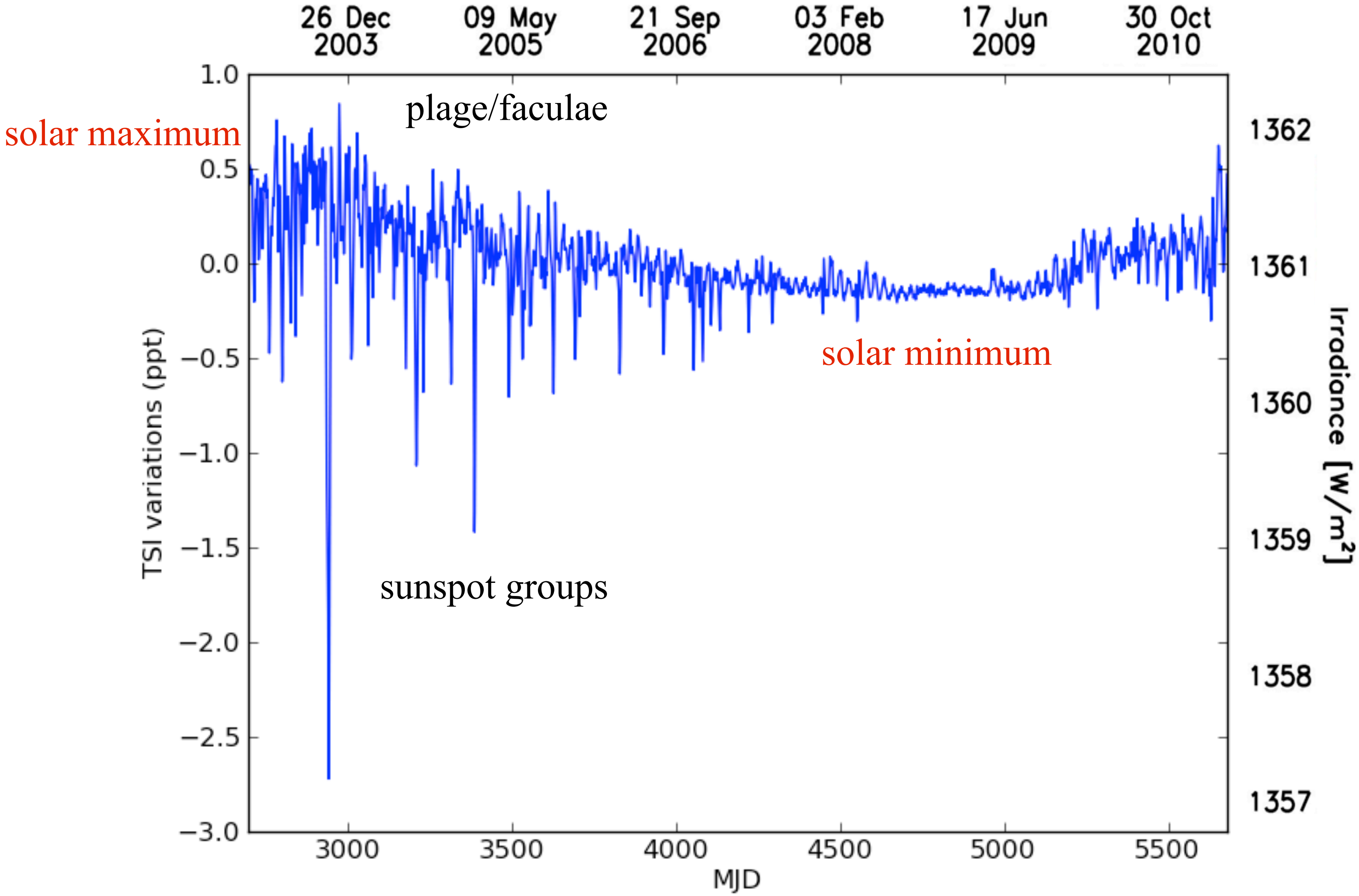
(see dataset `sorce_TSI_20110505.dat.gz`)



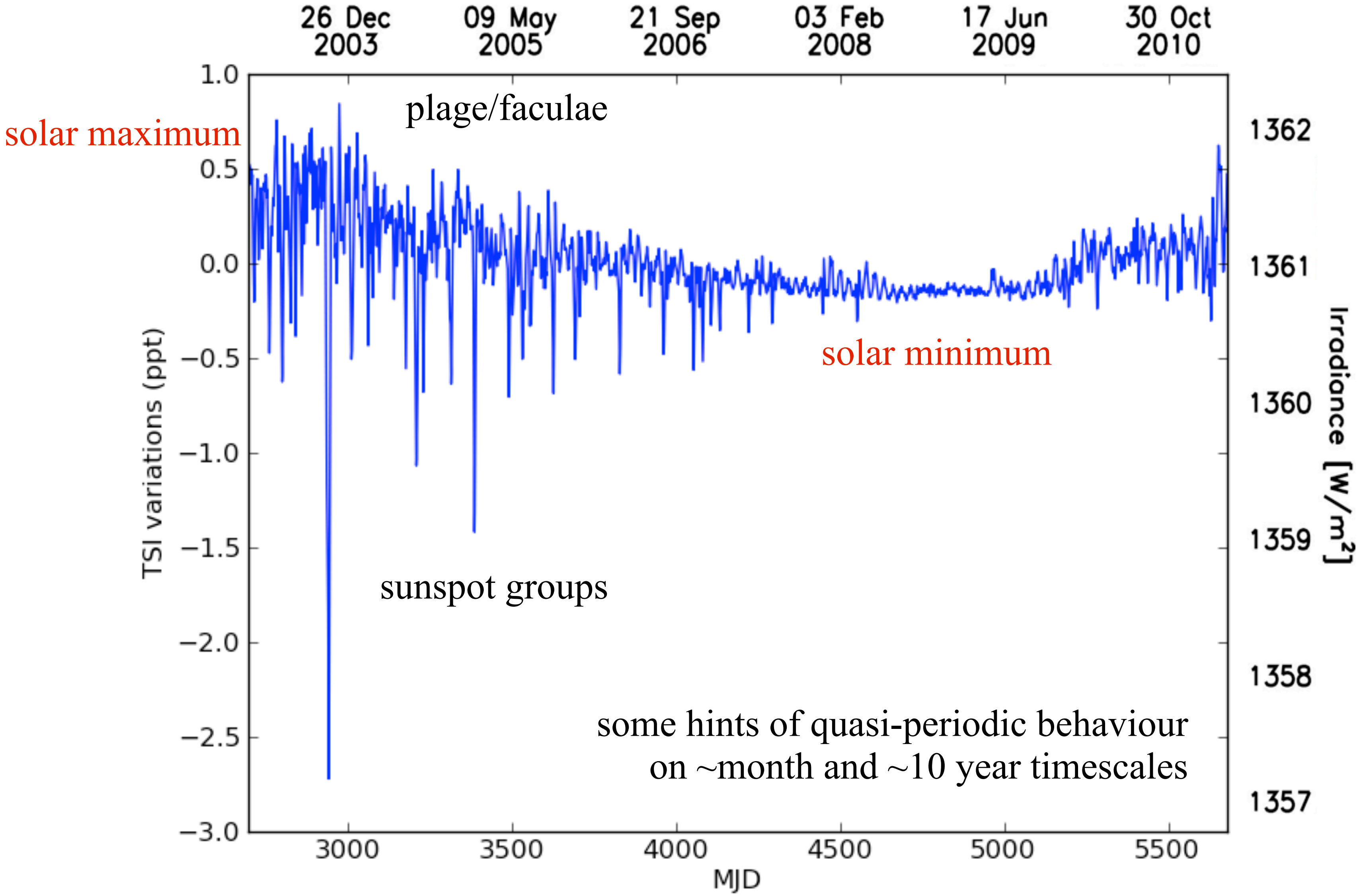
26 Dec 2003
09 May 2005
21 Sep 2006
03 Feb 2008
17 Jun 2009
30 Oct 2010
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
TSI variations (ppt)
3000
3500
4000
4500
5000
5500
MJD
1362
1361
1360
1359
1358
1357
Irradiance [W/m2]



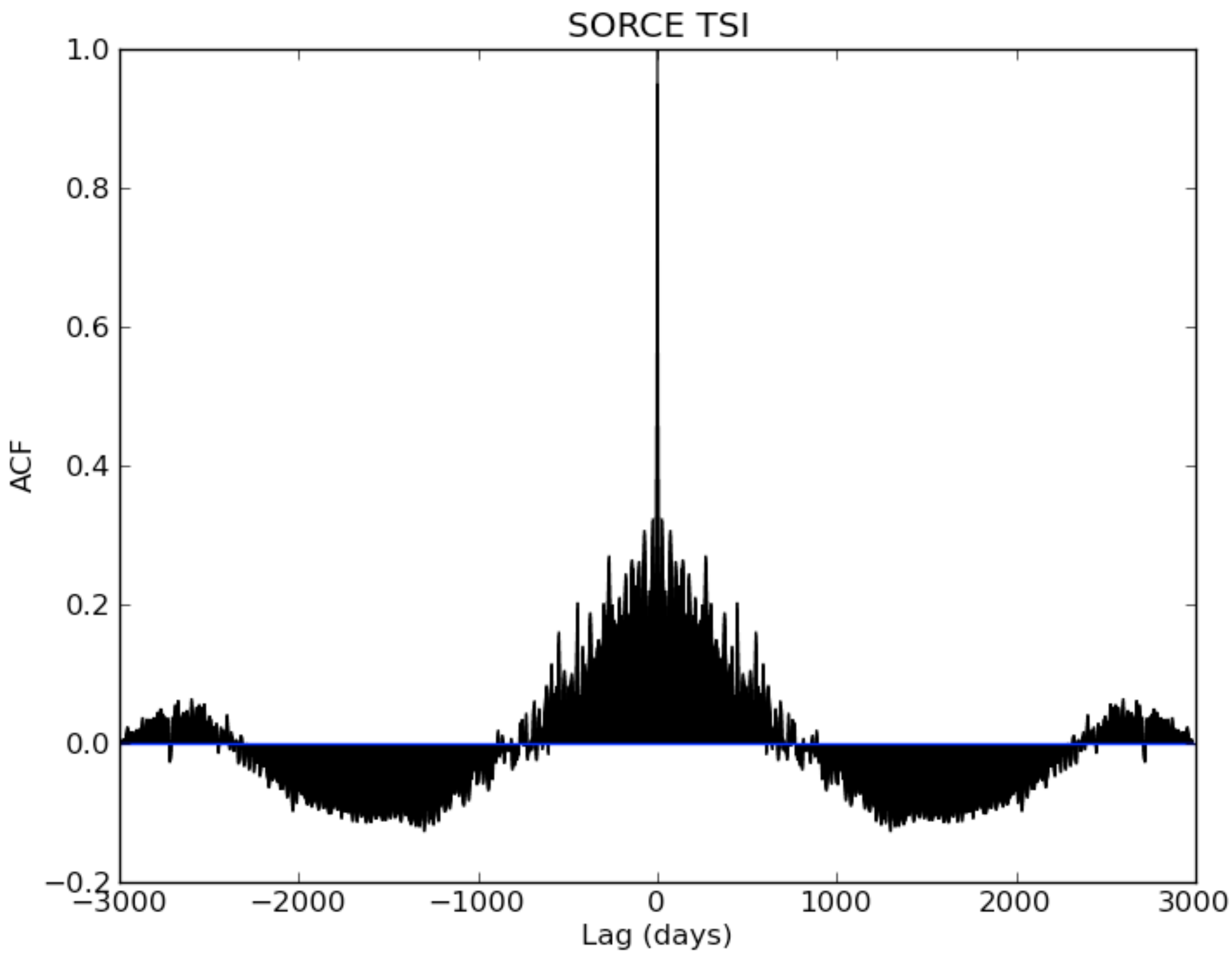
26 Dec 2003
09 May 2005
21 Sep 2006
03 Feb 2008
17 Jun 2009
30 Oct 2010
solar maximum
solar minimum
TSI variations (ppt)
Irradiance [W/m2]
MJD



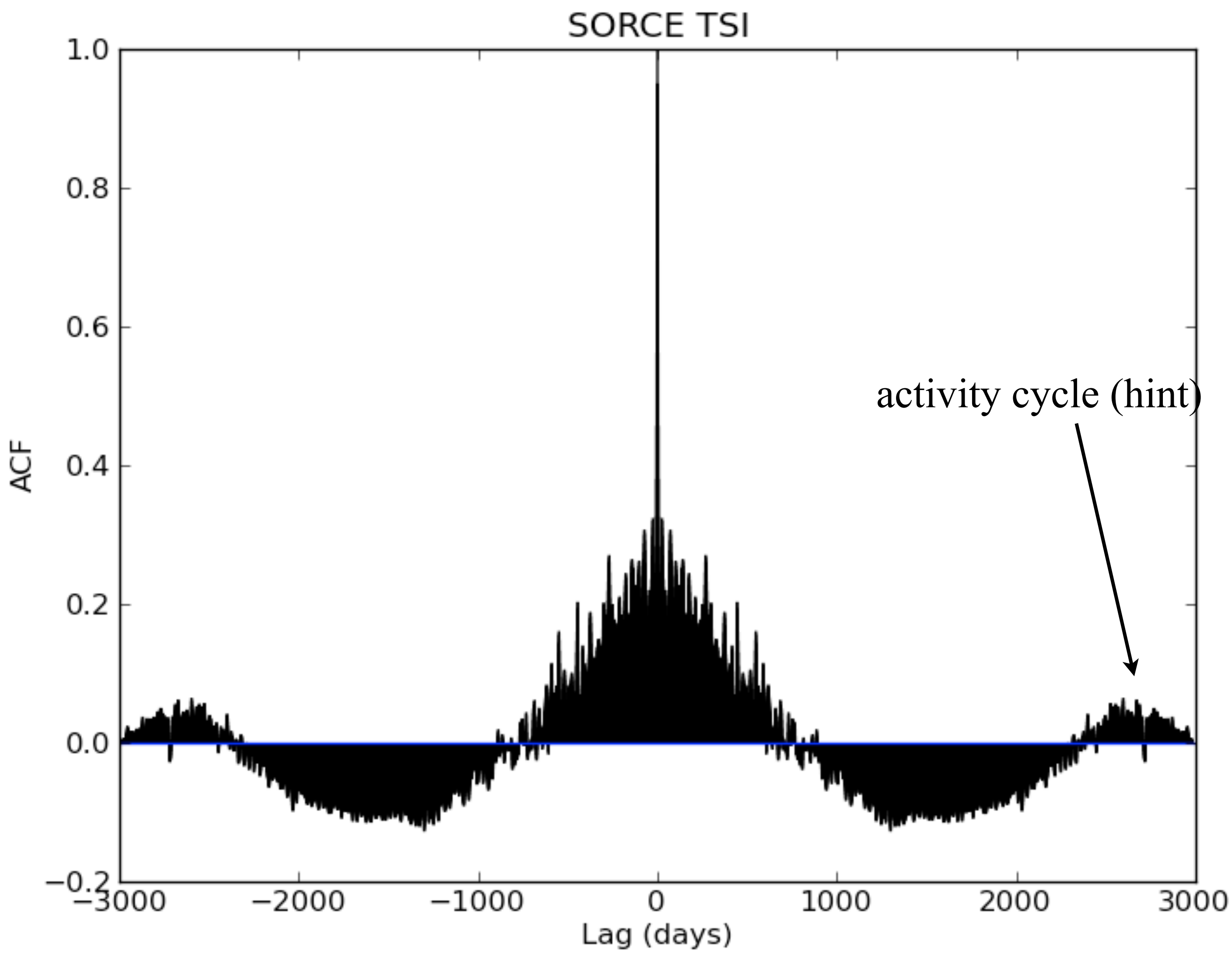
26 Dec 2003
09 May 2005
21 Sep 2006
03 Feb 2008
17 Jun 2009
30 Oct 2010
solar maximum
plage/faculae
solar minimum
sunspot groups
TSI variations (ppt)
Irradiance [W/m²]
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
1362
1361
1360
1359
1358
1357
3000
3500
4000
4500
5000
5500
MJD

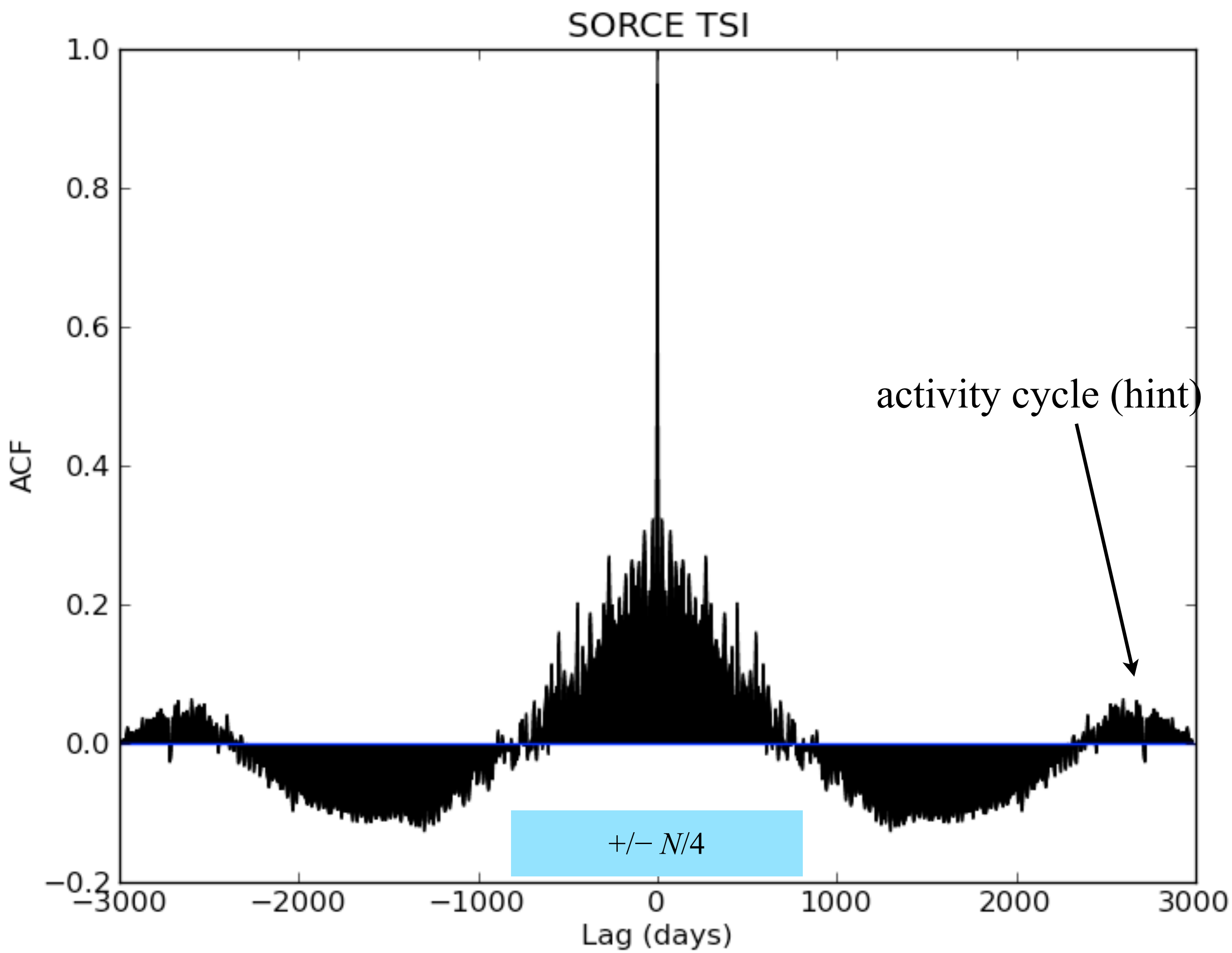


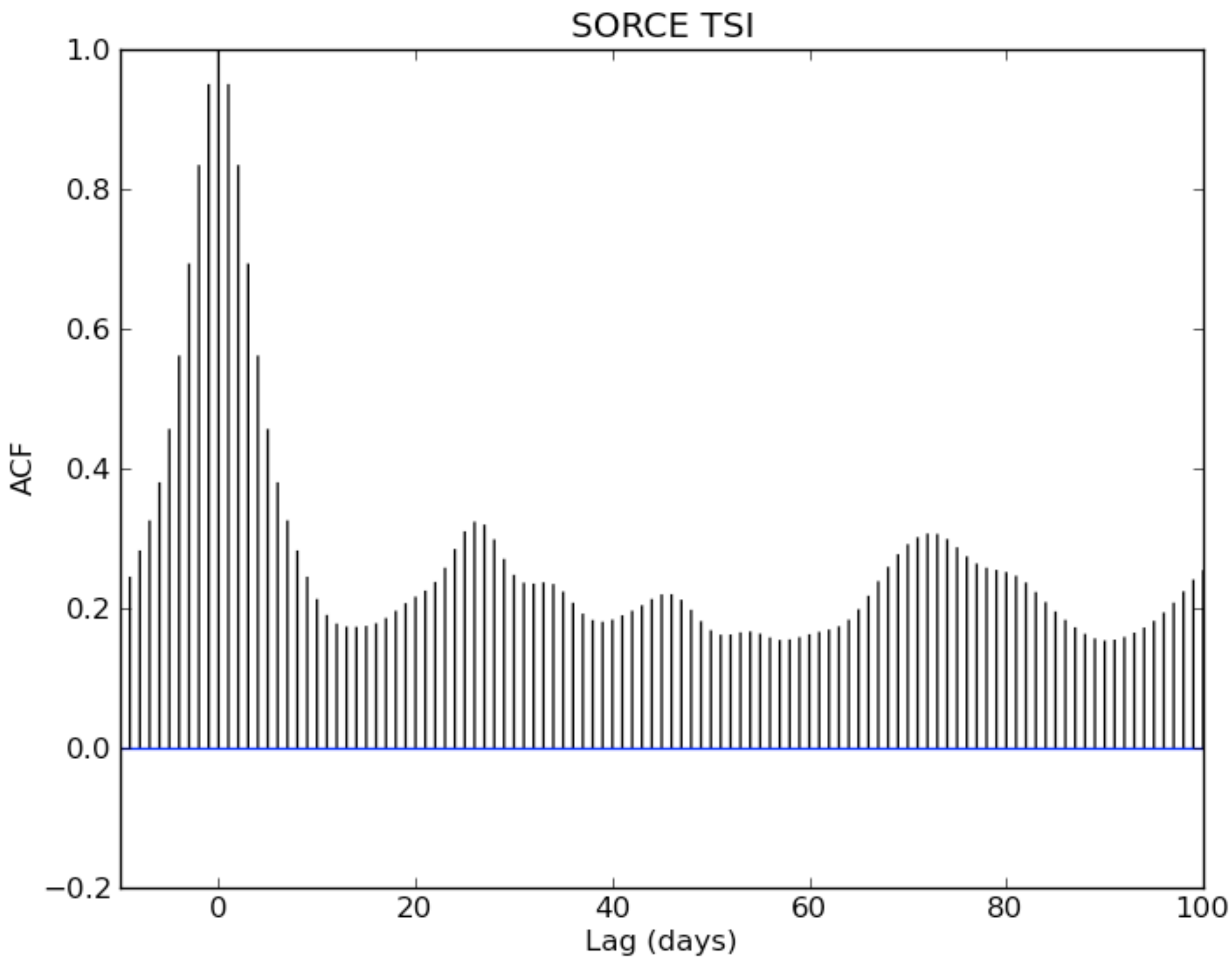
26 Dec 2003
09 May 2005
21 Sep 2006
03 Feb 2008
17 Jun 2009
30 Oct 2010
solar maximum
plage/faculae
solar minimum
sunspot groups
some hints of quasi-periodic behaviour
on ~month and ~10 year timescales
TSI variations (ppt)
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
3000
3500
4000
4500
5000
5500
MJD
1362
1361
1360
1359
1358
1357
Irradiance [W/m2]



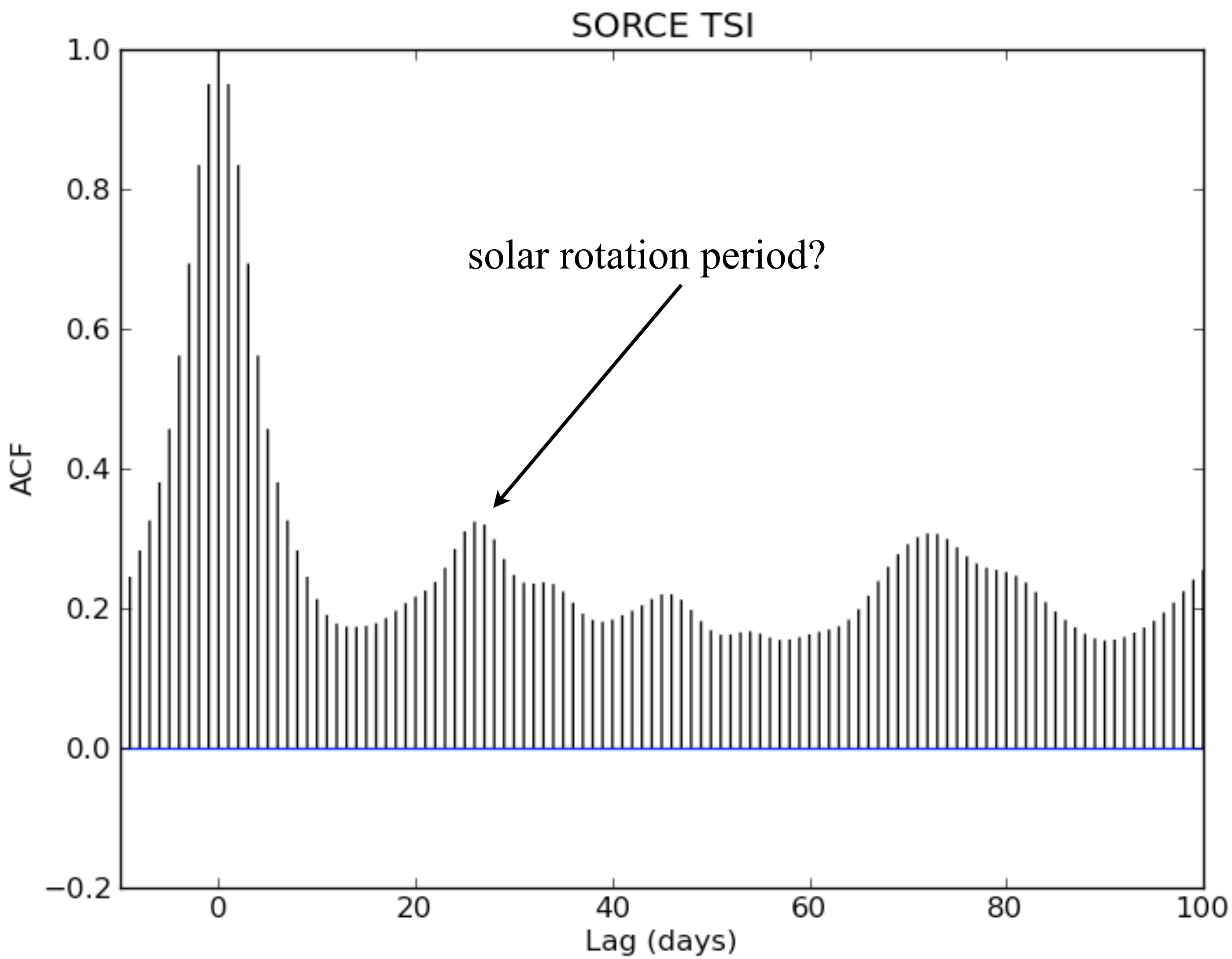
SORCE TSI
ACF
Lag (days)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−3000
−2000
−1000
0
1000
2000
3000

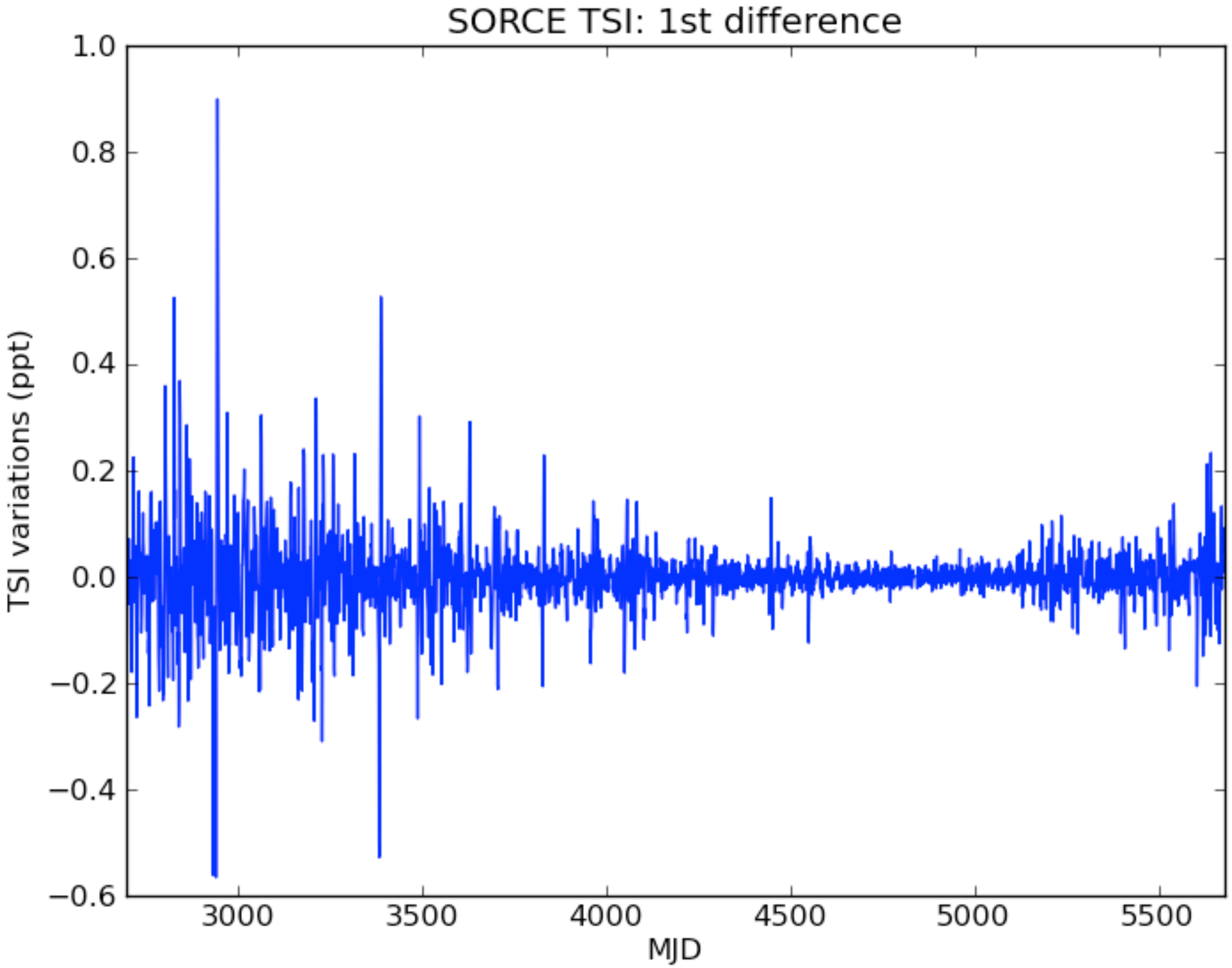

SORCE TSI
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−3000
−2000
−1000
0
1000
2000
3000
Lag (days)
activity cycle (hint)

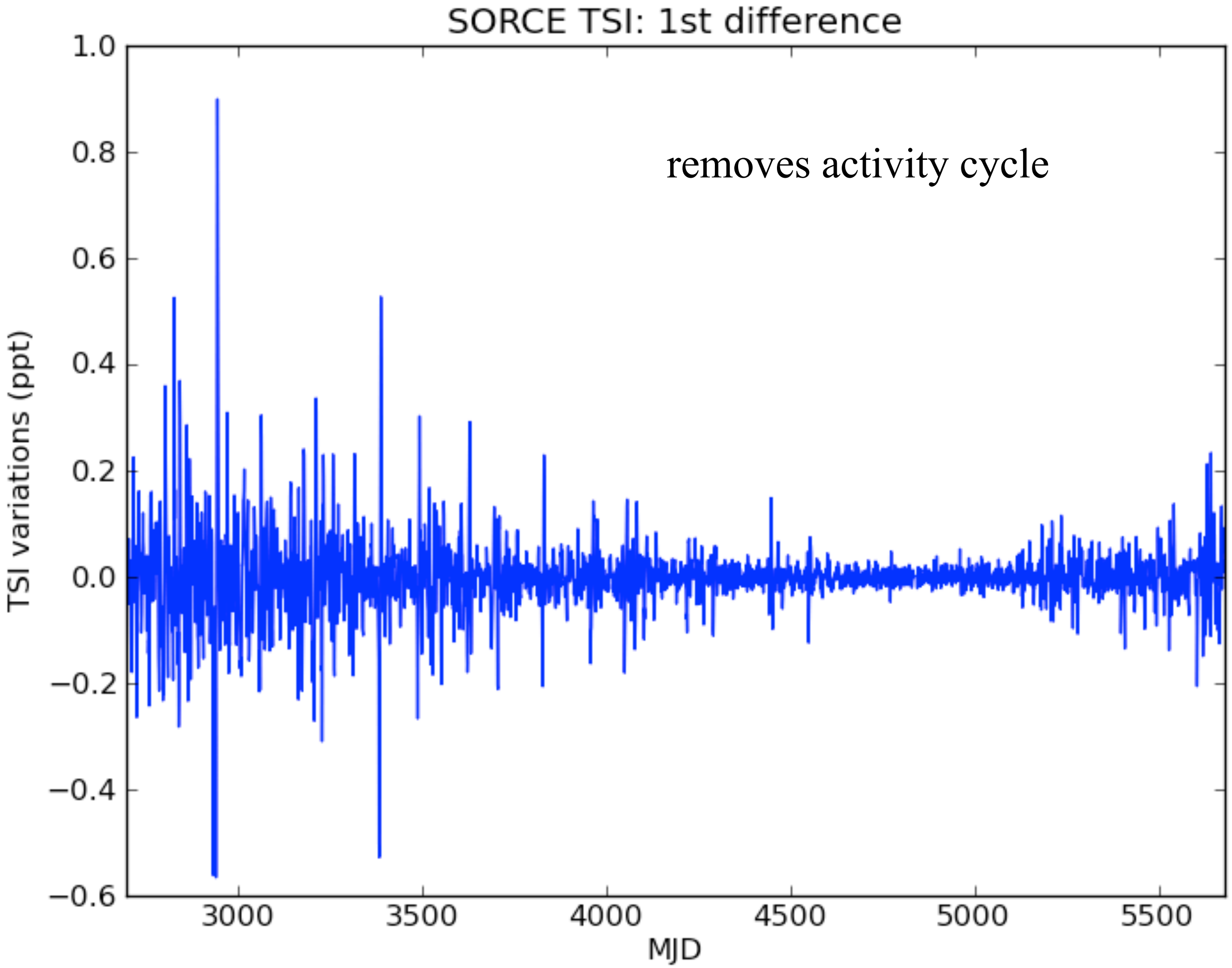

SORCE TSI
ACF
Lag (days)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−3000
−2000
−1000
0
1000
2000
3000
activity cycle (hint)
+/− N/4

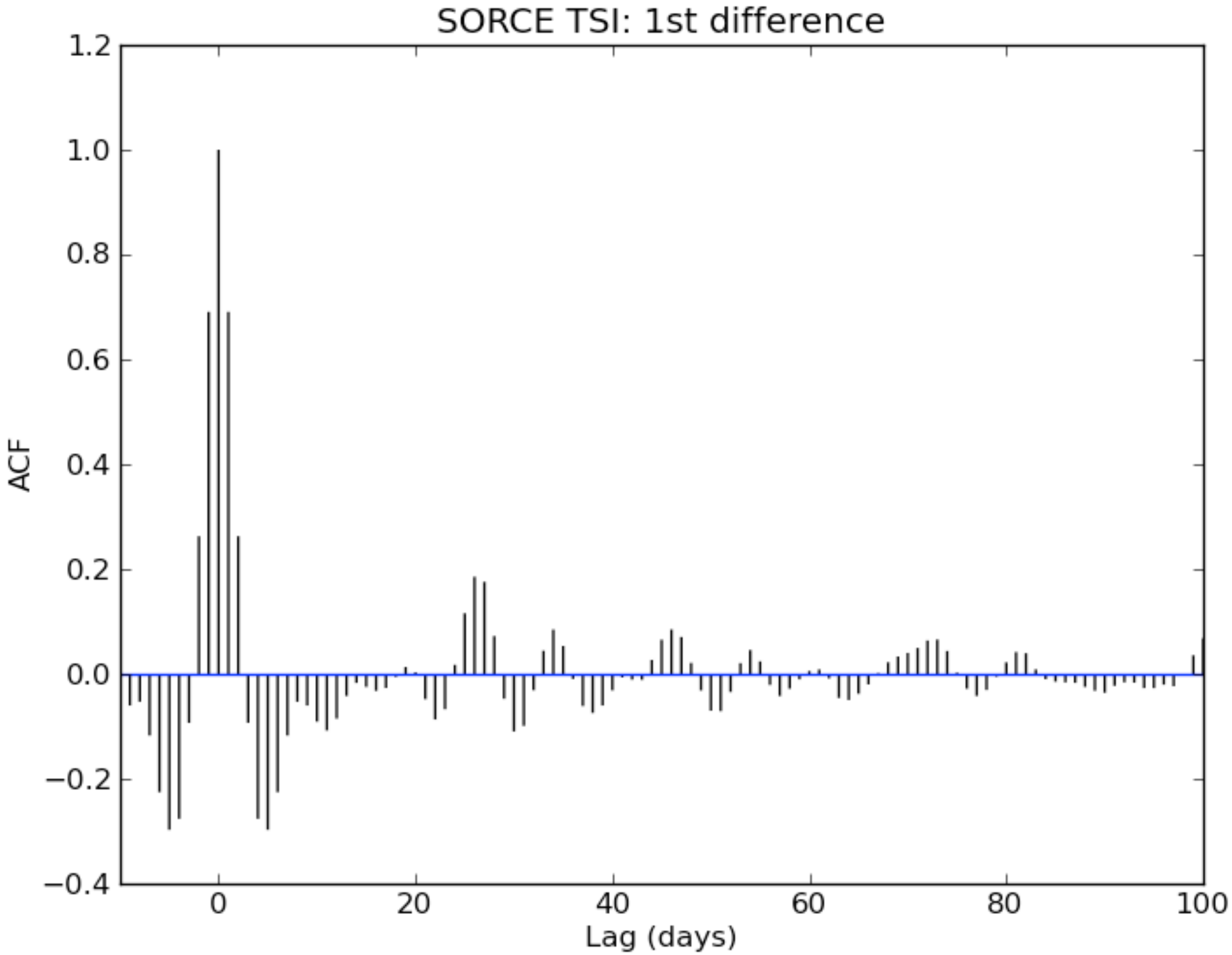


SORCE TSI
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
20
40
60
80
100
Lag (days)

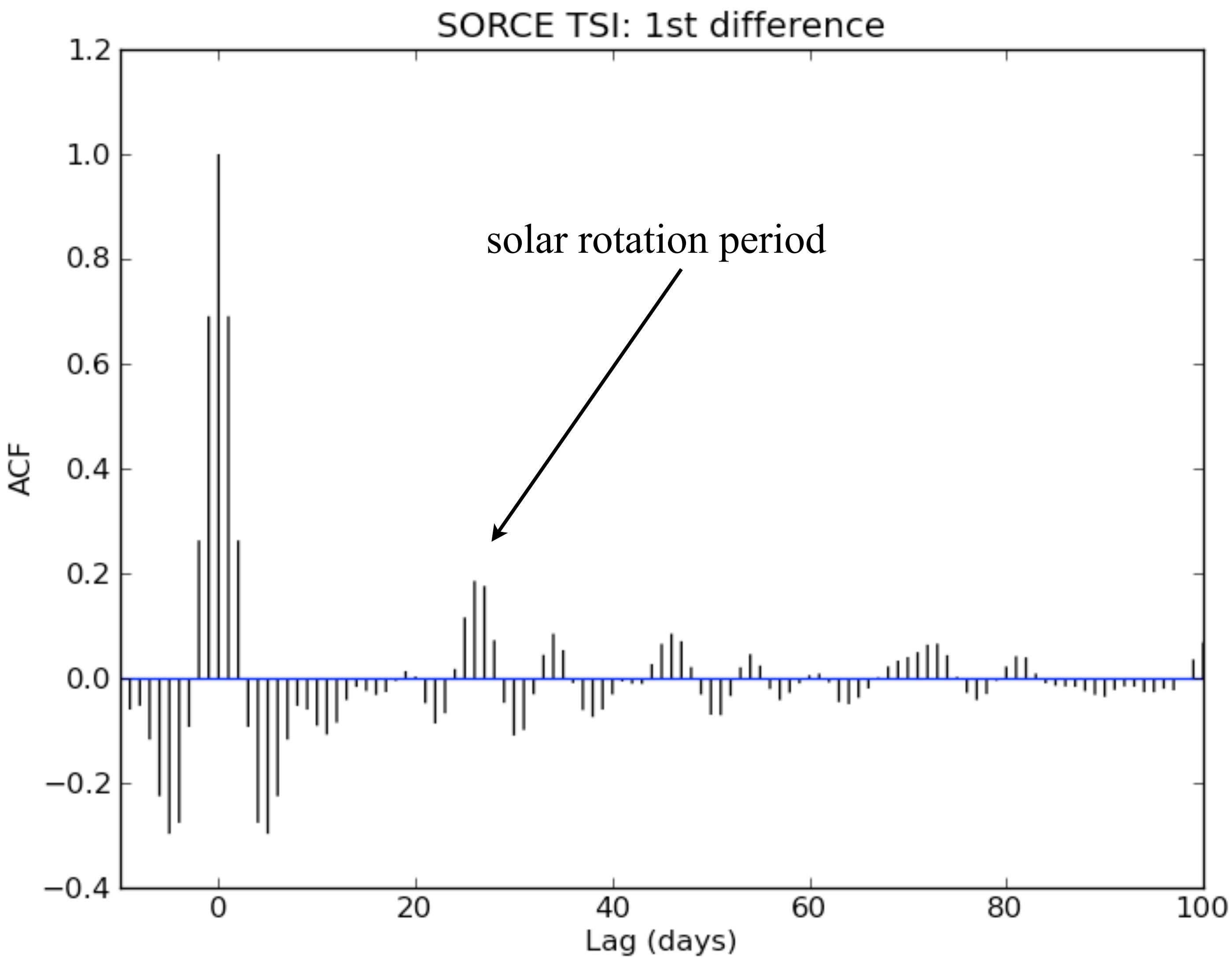

SORCE TSI
solar rotation period?
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
20
40
60
80
100
Lag (days)

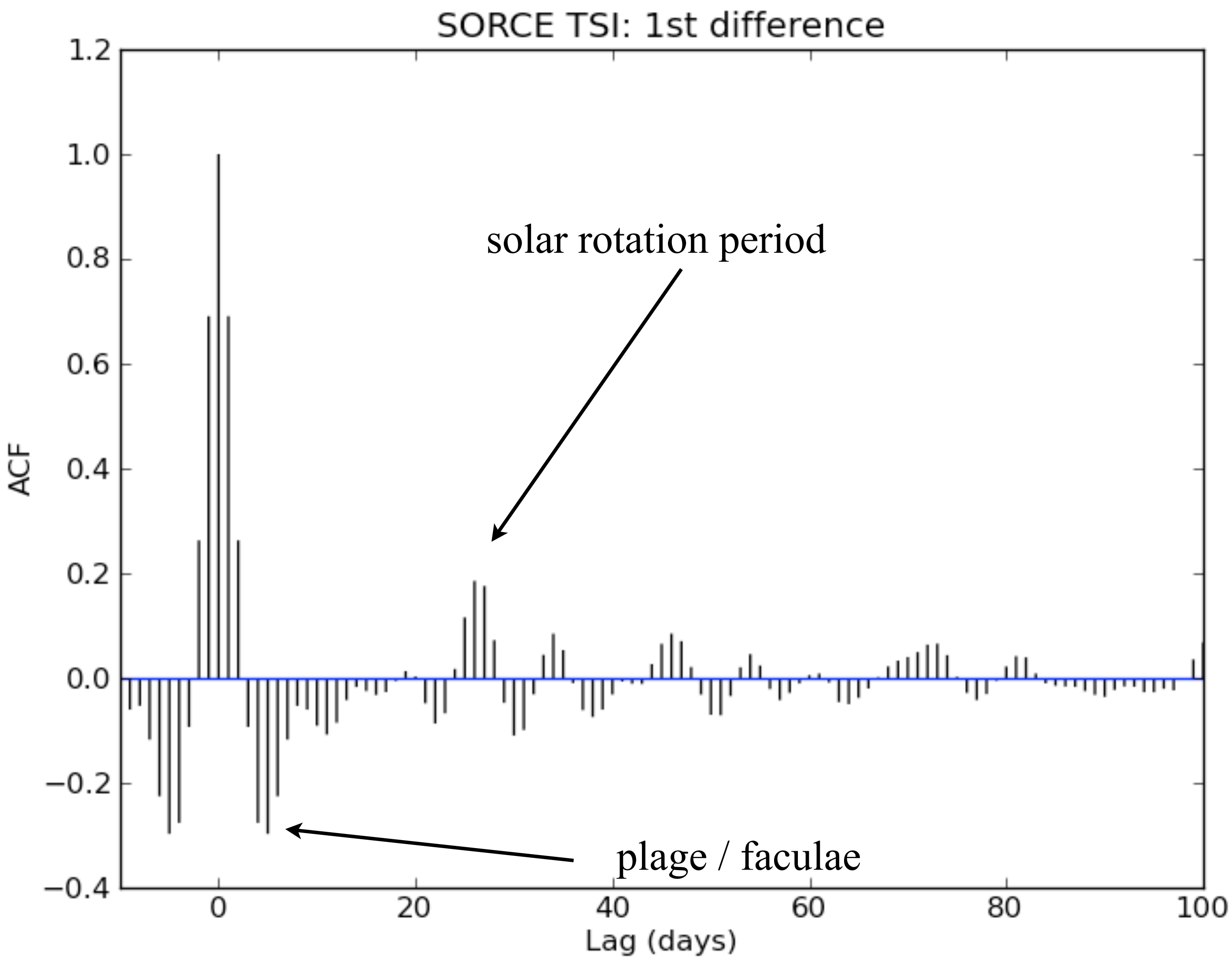

SORCE TSI: 1st difference
TSI variations (ppt)
MJD
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
3000
3500
4000
4500
5000
5500


SORCE TSI: 1st difference
removes activity cycle
TSI variations (ppt)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
3000
3500
4000
4500
5000
5500
MJD



SORCE TSI: 1st difference
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
ACF
0
20
40
60
80
100
Lag (days)


SORCE TSI: 1st difference
solar rotation period
ACF
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
20
40
60
80
100
Lag (days)


SORCE TSI: 1st difference
solar rotation period
plage / faculae
ACF
Lag (days)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
20
40
60
80
100

Spectral analysis

Much of this section follows Bretthorst (1988).

General linear basis model

- $y_i = f(t_i) + \varepsilon_i$ where
  - $i = 0, \ldots, N-1$, N = number of data points
  - $f(t) = \Sigma_j \, a_j \, g_j(t, \boldsymbol{\omega})$
    - $j = 0, \ldots, M-1$, M = number of basis functions
    - $\boldsymbol{a}$ = basis function weight(s) or amplitude(s)
    - $\boldsymbol{\omega}$ non-linear parameter(s)
  - $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$
    - σ_i^2 = noise variance associated with data point i
    - for simplicity assume $\sigma_i = \sigma$ here

General linear basis model

General linear basis model

- Likelihood $\mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma) = \mathrm{P}(\boldsymbol{\omega},\boldsymbol{a},\sigma|\mathrm{D},\mathrm{I}) \propto \Pi_\mathrm{i}\ \sigma^{-1} \exp\{-[y_i - f(t_i)]^2 / 2\sigma^2\}$
    - D = data, I = prior information

General linear basis model

- Likelihood $\mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma) = \mathrm{P}(\boldsymbol{\omega},\boldsymbol{a},\sigma|\mathrm{D},\mathrm{I}) \propto \Pi_{\mathrm{i}}\ \sigma^{-1} \exp\{-[y_i - f(t_i)]^2 / 2\sigma^2\}$
  - D = data, I = prior information
- Marginalise over **a**, σ: $\mathcal{L}(\boldsymbol{\omega}) \propto \int \mathrm{d}\boldsymbol{a} \int \mathrm{d}\sigma\ \mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma)\ \mathrm{P}(\boldsymbol{a}|\mathrm{I})\ \mathrm{P}(\sigma|\mathrm{I})$
  - Uninformative priors ~ setting parameters to maximum likelihood value
  - Much of this can be done analytically

General linear basis model

- Likelihood $\mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma) = \mathrm{P}(\boldsymbol{\omega},\boldsymbol{a},\sigma|\mathrm{D},\mathrm{I}) \propto \Pi_i\ \sigma^{-1} \exp\{-[y_i - f(t_i)]^2 / 2\sigma^2\}$
  - D = data, I = prior information
- Marginalise over **a**, σ: $\mathcal{L}(\boldsymbol{\omega}) \propto \int \mathrm{d}\boldsymbol{a} \int \mathrm{d}\sigma\ \mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma)\ \mathrm{P}(\boldsymbol{a}|\mathrm{I})\ \mathrm{P}(\sigma|\mathrm{I})$
  - Uninformative priors ~ setting parameters to maximum likelihood value
  - Much of this can be done analytically
- Scan through nonlinear parameter(s) $\boldsymbol{\omega}$, estimate $\mathrm{P}(\boldsymbol{\omega}|\mathrm{D},\mathrm{I}) \propto \mathcal{L}(\boldsymbol{\omega})\ \mathrm{P}(\boldsymbol{\omega}|\mathrm{I})$
  - grid search, sampling approach, ...

General linear basis model

- Likelihood $\mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma) = \mathrm{P}(\boldsymbol{\omega},\boldsymbol{a},\sigma|\mathrm{D},\mathrm{I}) \propto \Pi_\mathrm{i}\, \sigma^{-1} \exp\{-[y_i - f(t_i)]^2 / 2\sigma^2\}$
    - D = data, I = prior information
- Marginalise over **a**, σ: $\mathcal{L}(\boldsymbol{\omega}) \propto \int \mathrm{d}\boldsymbol{a} \int \mathrm{d}\sigma\, \mathcal{L}(\boldsymbol{\omega},\boldsymbol{a},\sigma)\, \mathrm{P}(\boldsymbol{a}|\mathrm{I})\, \mathrm{P}(\sigma|\mathrm{I})$
    - Uninformative priors ~ setting parameters to maximum likelihood value
    - Much of this can be done analytically
- Scan through nonlinear parameter(s) $\boldsymbol{\omega}$, estimate $\mathrm{P}(\boldsymbol{\omega}|\mathrm{D},\mathrm{I}) \propto \mathcal{L}(\boldsymbol{\omega})\, \mathrm{P}(\boldsymbol{\omega}|\mathrm{I})$
    - grid search, sampling approach, ...
- Compare model families (different M and/or $\boldsymbol{g}$) using posterior odds ratio
    - $O = \mathrm{P}(f_j|\mathrm{D},\mathrm{I}) / \mathrm{P}(f_k|\mathrm{D},\mathrm{I}) = \mathrm{P}(f_j|\mathrm{I})\, \mathrm{P}(\mathrm{D}|f_j,\mathrm{I}) / [\mathrm{P}(f_k|\mathrm{I})\, \mathrm{P}(\mathrm{D}|f_k|\mathrm{I})]$

Single sinusoidal oscillation plus noise

- Work with $\delta y_i = y_i - \Sigma_i\, y_i \,/\, N$
- $f(t_i) = [a_1 \cos(\omega t_i) + a_2 \sin(\omega t_i)]$

Single sinusoidal oscillation plus IID noise

- Work with $\delta y_i = y_i - \Sigma_i\, y_i \,/\, N$
- $f(t_i) = [a_1 \cos(\omega t_i) + a_2 \sin(\omega t_i)]$
- Assume regular sampling ($\delta t = 1$)
    - Basis functions are almost orthogonal
        - $\Sigma_i\, f^2(t_i) \approx N\,[(a_1)^2 + (a_2)^2] \,/\, 2$

Single sinusoidal oscillation plus IID noise

- Work with $\delta y_i = y_i - \Sigma_i\, y_i \,/\, N$
- $f(t_i) = [a_1 \cos(\omega t_i) + a_2 \sin(\omega t_i)]$
- Assume regular sampling, ignore departure from orthogonality
- $\mathcal{L}(\omega, \boldsymbol{a}, \sigma) \propto \sigma^{-N} \exp(-NQ/2\sigma^2)$ where
  - $Q = \hat{Y} - 2\,[a_1 R(\omega) + a_2 I(\omega)] \,/\, N + [(a_1)^2 + (a_2)^2] \,/\, 2$
  - $\hat{Y} = \Sigma_i\, (y_i)^2 \,/\, N$
  - $R(\omega) = \Sigma_i\, y_i \cos(\omega t_i)$
  - $I(\omega) = \Sigma_i\, y_i \sin(\omega t_i)$

Single sinusoidal oscillation plus IID noise

- Work with $\delta y_i = y_i - \Sigma_i \, y_i \, / \, N$
- $f(t_i) = [a_1 \cos(\omega t_i) + a_2 \sin(\omega t_i)]$
- Assume regular sampling
- $\mathcal{L}(\omega, \boldsymbol{a}, \sigma) \propto \sigma^{-N} \exp(-NQ/2\sigma^2)$ where
  - $Q = \hat{Y} - 2\,[a_1 R(\omega) + a_2 I(\omega)] \,/\, N + [(a_1)^2 + (a_2)^2] \,/\, 2$
- Marginalise out $\boldsymbol{a}$
  - Appropriate uninformative prior for amplitude is infinite-width Gaussian
  - Result is the same as setting $\boldsymbol{a}$ to maximum likelihood values
    - $a_{1,\mathrm{ML}} = 2\,R(\omega) \,/\, N$
    - $a_{2,\mathrm{ML}} = 2\,I(\omega) \,/\, N$
  - $\mathcal{L}(\omega, \sigma) \propto \sigma^{2-N} \exp(-\,N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] \,/\, N\sigma^2\}$

Single sinusoidal oscillation plus IID noise

- $f(t_i) = [a_1 \cos(\omega t_i) + a_2 \sin(\omega t_i)]$
  - Work with $\delta y_i = y_i - \Sigma_i\, y_i \,/\, N$
- $\mathcal{L}(\omega, \boldsymbol{a}, \sigma) \propto \sigma^{-N} \exp(-NQ/2\sigma^2)$ where
  - $Q = \hat{Y} - 2\,[a_1 R(\omega) + a_2 I(\omega)] \,/\, N + [(a_1)^2 + (a_2)^2] \,/\, 2$
- Marginalise out $\boldsymbol{a}$
- Marginalise out σ
  - Appropriate uninformative prior for scale factor is Jeffrey's prior $\propto 1/\sigma$
  - $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] \,/\, N\hat{Y}\}^{(2-N)/2}$

Relation to the discrete Fourier transform

Relation to the discrete Fourier transform

- Discrete Fourier transform (DFT)
  - $y_j = \Sigma_k \, x_k \exp(-2\pi ijk/N)$ $\quad\quad\quad\quad$ $(k = 0, \ldots, N-1)$

Relation to the discrete Fourier transform

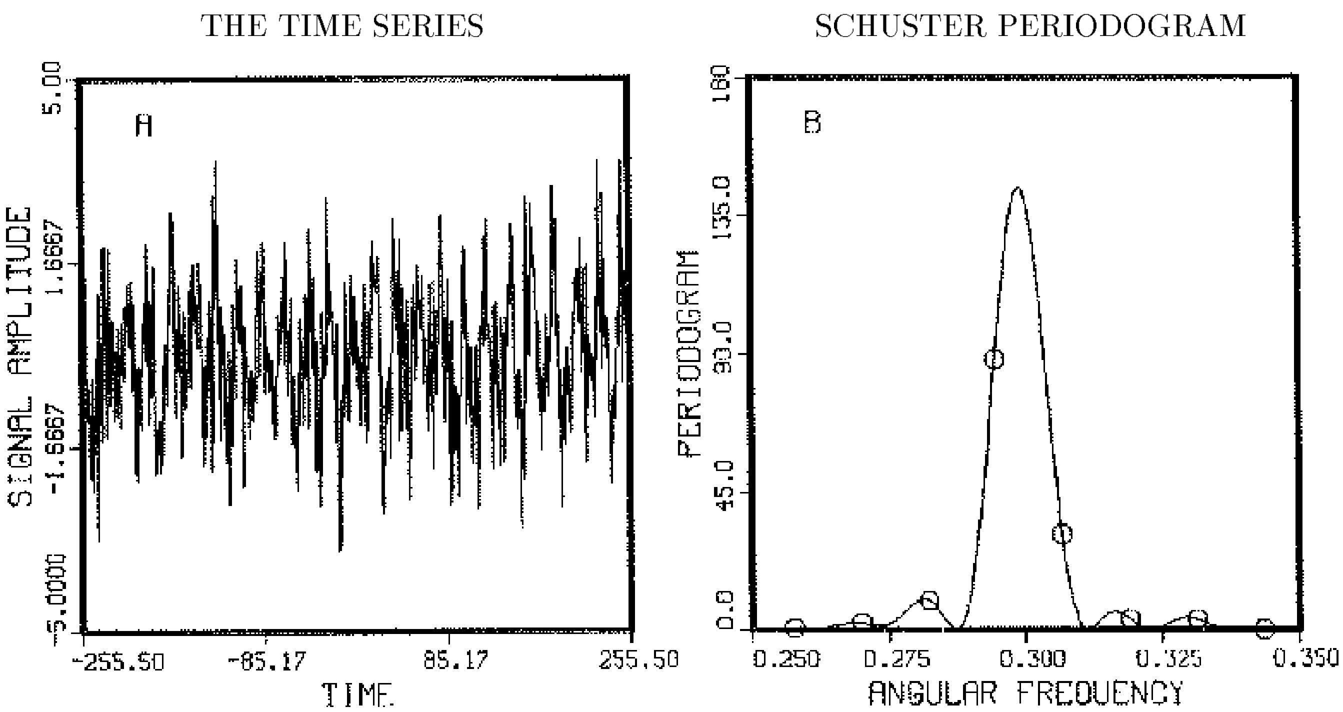
- Discrete Fourier transform (DFT)
    - $y_j = \Sigma_k\, x_k \exp(-2\pi ijk/N)$ $\quad$ $(k = 0, ..., N-1)$
- Inverse DFT (IDFT)
    - $x_k = [\Sigma_j\, y_j \exp(2\pi ijk/N)] / N$ $\quad$ $(n = 0, ..., N-1)$

Relation to the discrete Fourier transform

- Discrete Fourier transform (DFT)
  - $y_j = \Sigma_k x_k \exp(-2\pi ijk/N)$ $(k = 0, ..., N-1)$
- Inverse DFT (IDFT)
  - $x_k = [\Sigma_j y_j \exp(2\pi ijk/N)] / N$ $(n = 0, ..., N-1)$
- DFT power spectrum $|x_k|^2$
  - $\omega_k = 2\pi k/N\delta t$, $t_j = j\delta t$
  - $|x_k|^2 = [R(\omega_k)^2 + I(\omega_k)^2] / N^2 \equiv N\, C(\omega_k)$

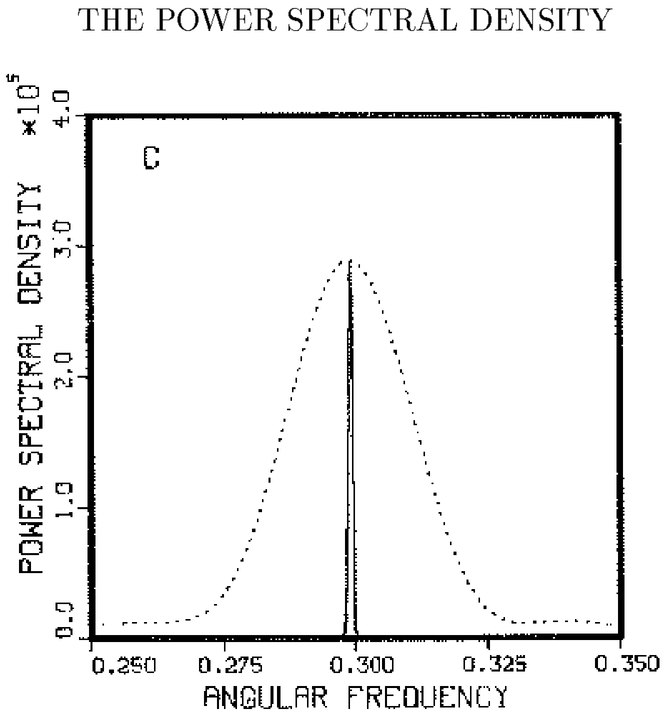
Relation to the discrete Fourier transform

- Discrete Fourier transform (DFT)
  - $y_j = \Sigma_k\, x_k \exp(-2\pi ijk/N)$ $(k = 0, ..., N-1)$
- Inverse DFT (IDFT)
  - $x_k = [\Sigma_j\, y_j \exp(2\pi ijk/N)] / N$ $(n = 0, ..., N-1)$
- DFT power spectrum $|x_k|^2$
  - $\omega_k = 2\pi k/N\delta t$, $t_j = j\delta t$
  - $|x_k|^2 = [R(\omega_k)^2 + I(\omega_k)^2] / N^2 \equiv N\, C(\omega_k)$
- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] / N\hat{Y}\}^{(2-N)/2}$
  - Define $C(\omega)$ such that $C(\omega_k) \equiv N\, |x_k|^2 = [R(\omega_k)^2 + I(\omega_k)^2] / N$
  - $\mathcal{L}(\omega) \propto \{1 - 2C(\omega) / \hat{Y}\}^{(2-N)/2}$
  - $\mathcal{L}(\omega)$ is much more sharply peaked than the DFT



Bretthorst (1988)

The data (A) contain a single harmonic frequency plus noise; $N = 512$, and $S/N \approx 1$. The Schuster periodogram, (B) solid curve, and the fast Fourier transform, open circles, clearly show a sharp peak plus side lobes. These side lobes do not show up in the power spectral density, (C), because the posterior probability is very sharply peaked around the maximum of the periodogram. The dotted line in (C) is a Blackman-Tukey spectrum with a Hanning window and 256 lag coefficients.



Using the DFT for the “single sinusoid” problem

Using the DFT for the "single sinusoid" problem

- DFT is applicable to the single sinusoid plus IID noise problem when
  - the time-sampling is regular
  - N is large, $\omega >> 2\pi N/\delta t$ (orthogonality)
  - the signal is stationary ($\boldsymbol{a}$, σ constant)
  - the noise is IID (noise = anything that the model can't reproduce)

Using the DFT for the “single sinusoid” problem

- DFT is applicable to the single sinusoid plus IID noise problem when
  - the time-sampling is regular
  - N is large, $\omega >> 2\pi N/\delta t$ (orthogonality)
  - the signal is stationary ($\boldsymbol{a}$, σ constant)
  - the noise is IID (noise = anything that the model can’t reproduce)
- Can also use DFT for multiple oscillatory signals if
  - $\delta\omega >> 2\pi N/\delta t$
  - must use multiple sinusoid basis to estimate residual noise variance

Using the DFT for the “single sinusoid” problem

- DFT is applicable to the single sinusoid plus IID noise problem when
  - the time-sampling is regular
  - N is large, $\omega >> 2\pi N/\delta t$ (orthogonality)
  - the signal is stationary ($\boldsymbol{a}$, σ constant)
  - the noise is IID (noise = anything that the model can’t reproduce)
- Can also use DFT for multiple oscillatory signals if
  - $\delta\omega >> 2\pi N/\delta t$
  - must use multiple sinusoid basis to estimate residual noise variance
- OK to ignore departure from orthogonality if the sampling is almost regular
  - Schuster periodogram: $C(\omega) = [R(\omega)^2 + I(\omega)^2] / N$ (FFT algorithm no longer applies)

DFT with Python

- Data in array `y`
- `x = numpy.fft.rfft(y)` is the DFT of `y` ("`r`" is for "real")[1]
- `C = abs(x)**2/N` is the power spectrum of `y` estimated at the frequencies...
- `freq = numpy.arange(N/2+1)/float(N*dt)`

[1] `rfft` computes the DFT at positive frequencies only.
For real input the DFT is Hermitian, so there is no extra information in the negative frequency part of the DFT

DFT with Python

- Data in array `y`
- `x = numpy.fft.rfft(y)` is the DFT of `y` ("`r`" is for "real")[1]
- `C = abs(x)**2/N` is the power spectrum of `y` estimated at the frequencies...
- `freq = numpy.arange(N/2+1)/float(N*dt)`

- If $y_i = \sin(\omega t_i)$:
  - IDFT is $x(\omega) = i\ \delta(\omega) / 2$, DFT power spectrum is $C(\omega) = \delta(\omega) / 4$
  - In the literature, power spectrum is often "normalised" by multiplying by a factor 4.
    - This suggests a one-to-one correspondence between peak amplitude in the power spectrum and oscillation amplitude, which is not necessarily true.

[1] `rfft` computes the DFT at positive frequencies only.
For real input the DFT is Hermitian, so there is no extra information in the negative frequency part of the DFT

Single sine − detectability and uncertainties

Bretthorst (1988).

Single sine − detectability and uncertainties

- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] / N\hat{Y}\}^{(2-N)/2}$
  - or $\mathcal{L}(\omega,\sigma) \propto \sigma^{2-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] / N\sigma^2\}$ if σ known a priori

Bretthorst (1988).

Single sine − detectability and uncertainties

- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] \,/\, N\hat{Y}\}^{(2-N)/2}$
    - or $\mathcal{L}(\omega,\sigma) \propto \sigma^{2-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] \,/\, N\sigma^2\}$ if σ known a priori
- Detectability of individual signals
    - $\mathrm{C}(\omega_{\mathrm{max}}) \approx N\,|\boldsymbol{a}|^2/4$
    - $\mathcal{L}(\omega_{\mathrm{max}}) \propto \exp(|N|\boldsymbol{a}|^2/4\sigma^2)$

Bretthorst (1988).

Single sine − detectability and uncertainties

- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] / N\hat{Y}\}^{(2-N)/2}$
  - or $\mathcal{L}(\omega,\sigma) \propto \sigma^{2-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] / N\sigma^2\}$ if σ known a priori
- Detectability of individual signals
  - $C(\omega_{\max}) \approx N\,|\boldsymbol{a}|^2/4$
  - $\mathcal{L}(\omega_{\max}) \propto \exp(|N|\boldsymbol{a}|^2/4\sigma^2)$
- Uncertainty on ω
  - derived from *shape* of peak in $\mathcal{L}(\omega)$
  - $\sigma_\omega = \sigma\,|\boldsymbol{a}|^{-1}\,(48/N^3)^{1/2}$

Bretthorst (1988).

Single sine − detectability and uncertainties

- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] / N\hat{Y}\}^{(2-N)/2}$
  - or $\mathcal{L}(\omega,\sigma) \propto \sigma^{2-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] / N\sigma^2\}$ if σ known a priori
- Detectability of individual signals
  - $C(\omega_{\max}) \approx N\,|\boldsymbol{a}|^2/4$
  - $\mathcal{L}(\omega_{\max}) \propto \exp(|N|\boldsymbol{a}|^2/4\sigma^2)$
- Uncertainty on ω
  - derived from *shape* of peak in $\mathcal{L}(\omega)$
  - $\sigma_\omega = \sigma\,|\boldsymbol{a}|^{-1}\,(48/N^3)^{1/2}$
- Uncertainty on $\boldsymbol{a}$: $(\sigma_{|\boldsymbol{a}|})^2 = \langle a_1 a_2\rangle - \langle a_1\rangle\langle a_2\rangle = \sigma^2/N$

Bretthorst (1988).

Single sine − detectability and uncertainties

- $\mathcal{L}(\omega) \propto \{1 - 2[R(\omega)^2 + I(\omega)^2] / N\hat{Y}\}^{(2-N)/2}$
    - or $\mathcal{L}(\omega,\sigma) \propto \sigma^{2-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp\{[R(\omega)^2 + I(\omega)^2] / N\sigma^2\}$ if σ known a priori
- Detectability of individual signals
    - $\mathrm{C}(\omega_{\mathrm{max}}) \approx N\,|\boldsymbol{a}|^2/4$
    - $\mathcal{L}(\omega_{\mathrm{max}}) \propto \exp(|N|\boldsymbol{a}|^2/4\sigma^2)$
- Uncertainty on ω
    - derived from *shape* of peak in $\mathcal{L}(\omega)$
    - $\sigma_\omega = \sigma\,|\boldsymbol{a}|^{-1}\,(48/N^3)^{1/2}$
- Uncertainty on $\boldsymbol{a}$: $(\sigma_{|\boldsymbol{a}|})^2 = \langle a_1 a_2\rangle - \langle a_1\rangle\langle a_2\rangle = \sigma^2/N$
- Estimated noise variance: $\langle\sigma^2\rangle = \Sigma_i\,[\omega_{\mathrm{max}}\,y_i - f(\omega_{\mathrm{max}},\langle\boldsymbol{a}\rangle,t_i)]^2 / (N-4)$

Bretthorst (1988).

Violating the assumptions

Violating the assumptions

- Bayesian approach ensures that if our assumptions are wrong, our conclusions are not misleading

Violating the assumptions

- Bayesian approach ensures that if our assumptions are wrong, our conclusions are not misleading
- For example (see Bretthorst 1989 for details),
  - Non-stationary noise
  - Non-white noise
  - Variable amplitude of signal
  - ...

Violating the assumptions

- Bayesian approach ensures that if our assumptions are wrong, our conclusions are not misleading
- For example (see Bretthorst 1989 for details),
  - Non-stationary noise
  - Non-white noise
  - Variable amplitude of signal
  - ...
- All of these lead to increased estimates of the parameter uncertainties, because even the best model doesn’t match the data well

Violating the assumptions

- Bayesian approach ensures that if our assumptions are wrong, our conclusions are not misleading
- For example (see Bretthorst 1989 for details),
  - Non-stationary noise
  - Non-white noise
  - Variable amplitude of signal
  - ...
- All of these lead to increased estimates of the parameter uncertainties, because even the best model doesn’t match the data well
- It is particularly important to treat measurement uncertainties with caution
  - Often we don’t know them as well as we think we do

Aliasing, harmonics and beating

Aliasing, harmonics and beating

- Beating
  - modulation of envelope when signal at two close frequencies is present
  - $f_{\text{beat}} = |f_1 - f_2|$

Aliasing, harmonics and beating

- Beating
  - modulation of envelope when signal at two close frequencies is present
  - $f_{\text{beat}} = |f_1 - f_2|$
- Aliasing
  - signal at $f > f_{\text{Nyquist}}$ appears at $f_{\text{alias}} = |f - k\, f_{\text{Nyquist}}|$
    - k is integer such that $0 < f \leq f_{\text{Nyquist}}$

Aliasing, harmonics and beating

- Beating
  - modulation of envelope when signal at two close frequencies is present
  - $f_{\text{beat}} = |f_1 - f_2|$
- Aliasing
  - signal at $f > f_{\text{Nyquist}}$ appears at $f_{\text{alias}} = |f - k\, f_{\text{Nyquist}}|$
    - k is integer such that $0 < f \leq f_{\text{Nyquist}}$
- Harmonics
  - the harmonics of fundamental frequency f are $2f$, $3f$, ...
  - non-sinusoidal (non-harmonic) periodic signals decompose into multiple harmonics

Aliasing, harmonics and beating

- Beating
  - modulation of envelope when signal at two close frequencies is present
  - $f_{\text{beat}} = |f_1 - f_2|$
- Aliasing
  - signal at $f > f_{\text{Nyquist}}$ appears at $f_{\text{alias}} = |f - k\, f_{\text{Nyquist}}|$
    - k is integer such that $0 < f \leq f_{\text{Nyquist}}$
- Harmonics
  - the harmonics of fundamental frequency f are $2f$, $3f$, ...
  - non-sinusoidal (non-harmonic) periodic signals decompose into multiple harmonics
- In Bayesian PSD, only main frequency matters
  - side lobes, aliases and harmonics heavily suppressed

Beyond sinusoids

Beyond sinusoids

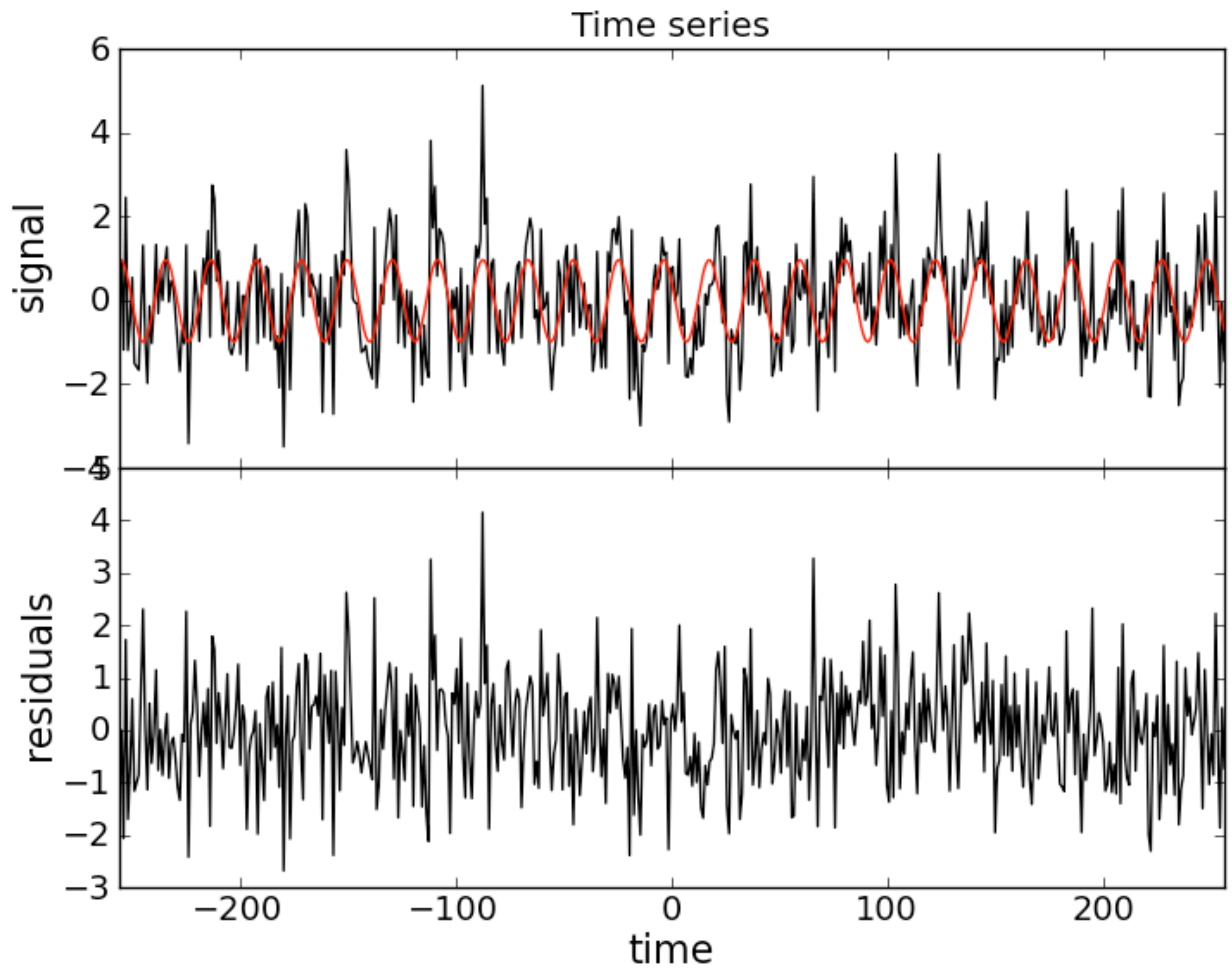
- The analysis outlined so far applies to any linear basis model
    - E.g. multiple sinusoids

Beyond sinusoids

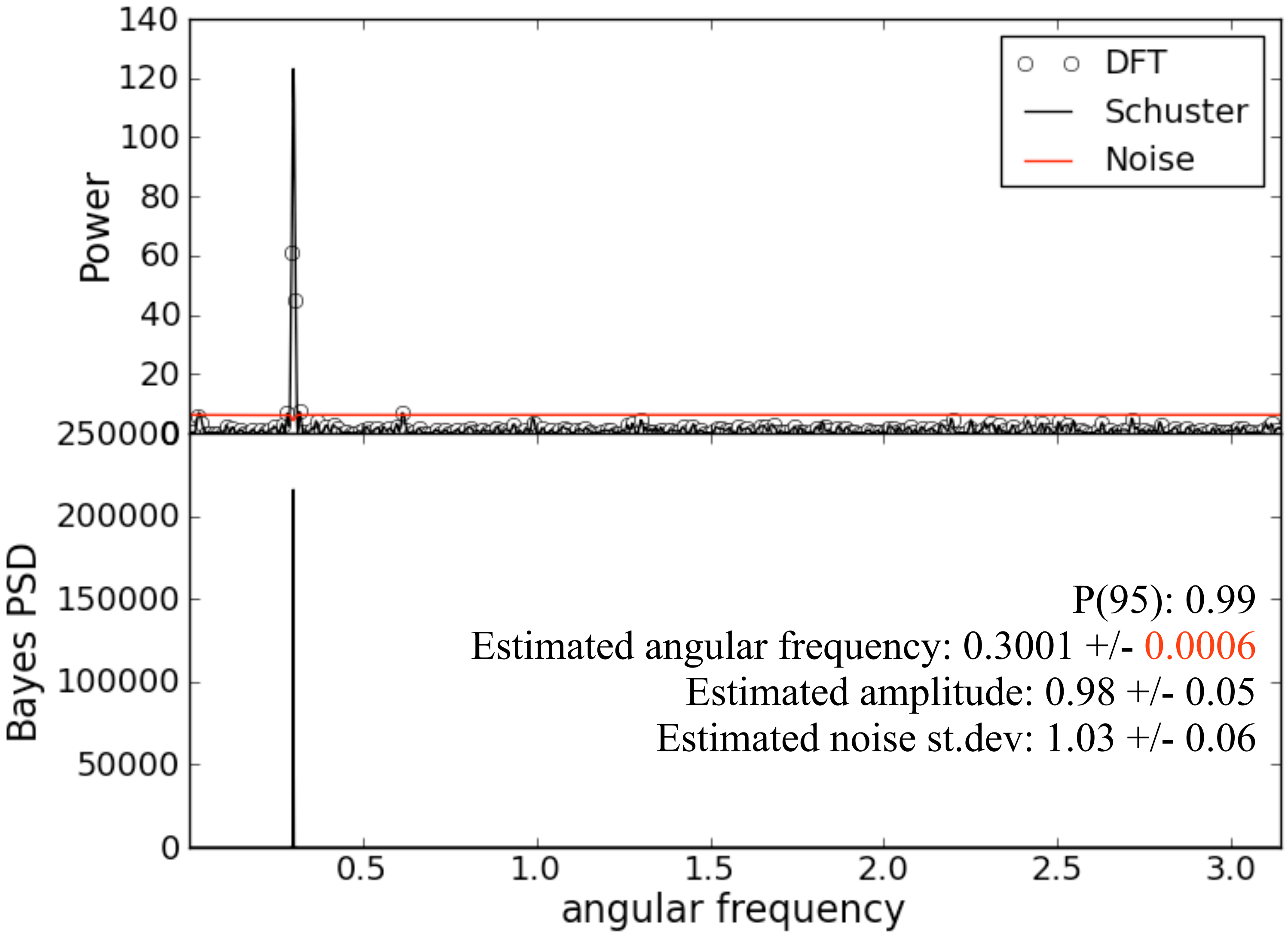
- The analysis outlined so far applies to any linear basis model
  - E.g. multiple sinusoids
- Can add extra w’s to represent decay / growth / other nonlinear parameters
  - growing / decaying signals
  - keplerian orbits
  - ...

Simple examples 2: spectral analysis (regular)
(see script `examples_2.py`)

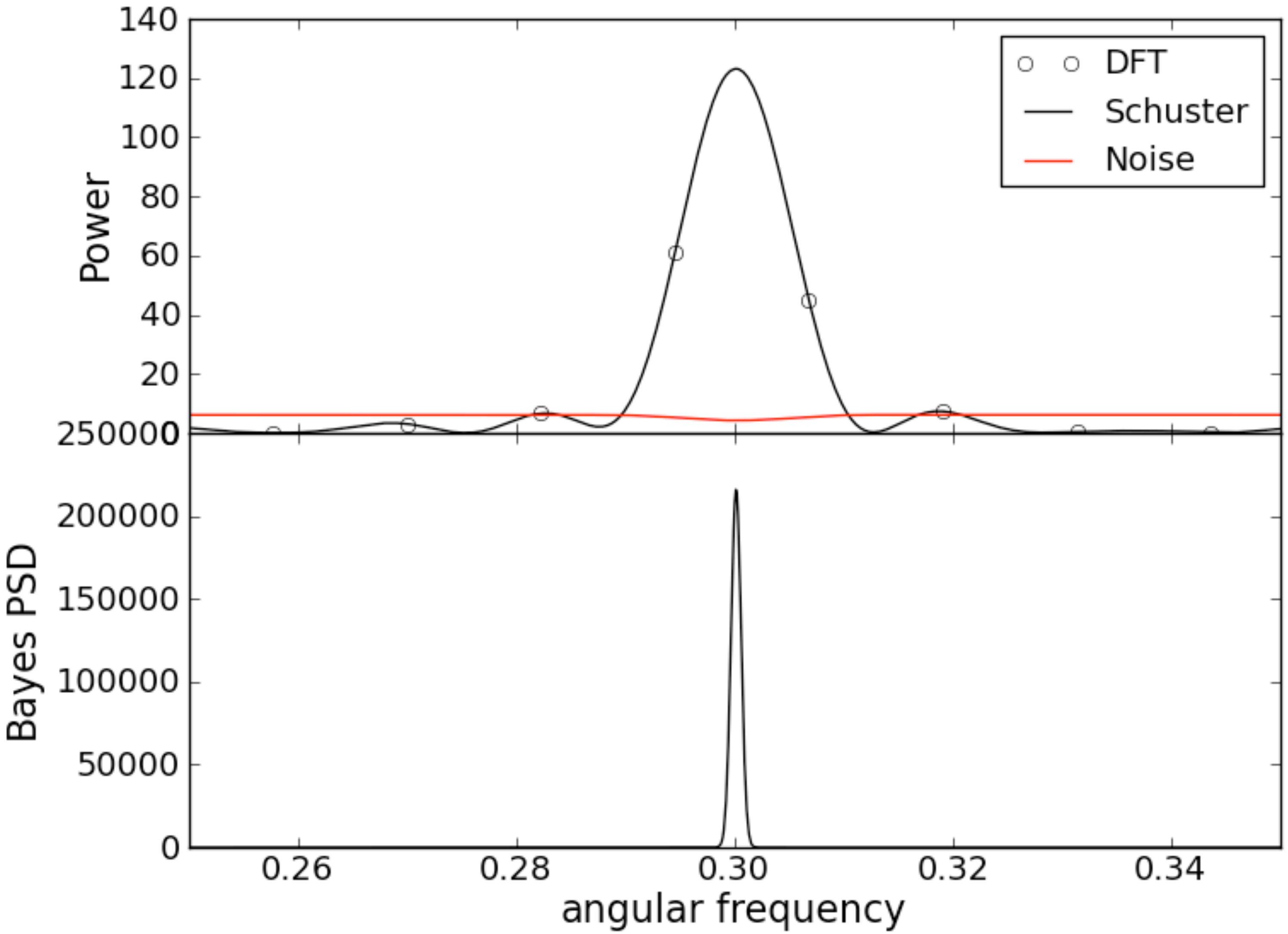
Single sinusoid + noise



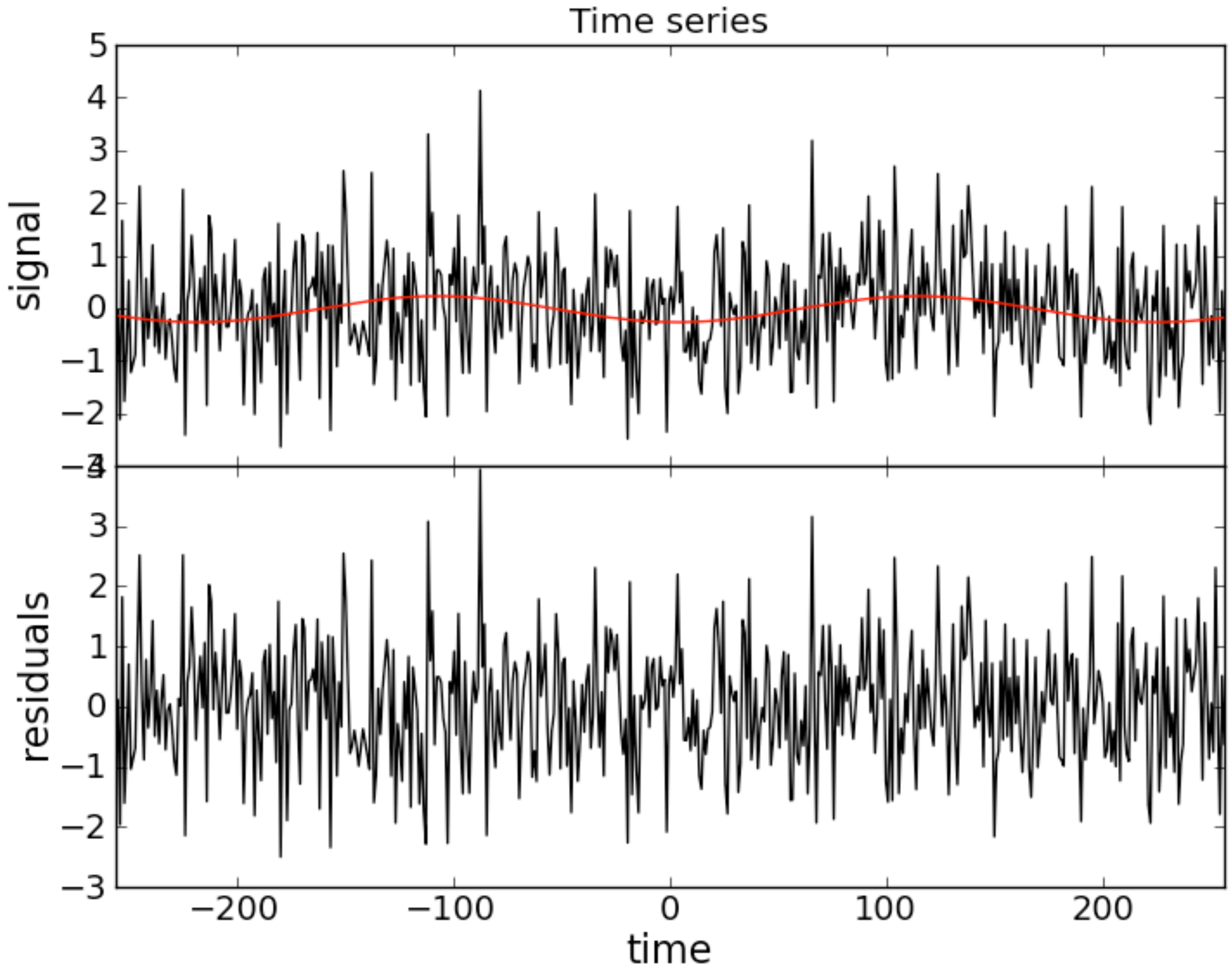
Single sinusoid + noise



Single sinusoid + noise

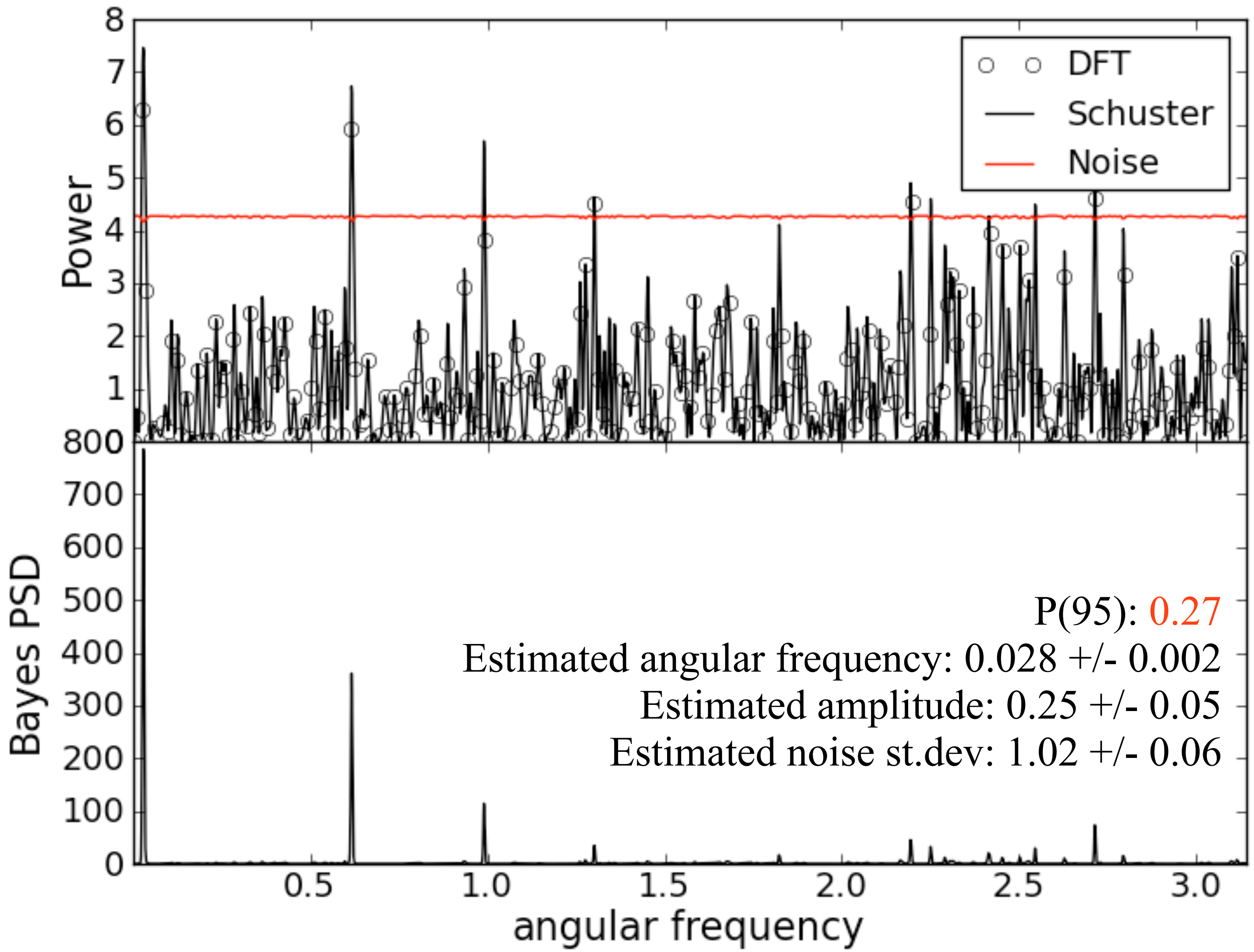


Noise only



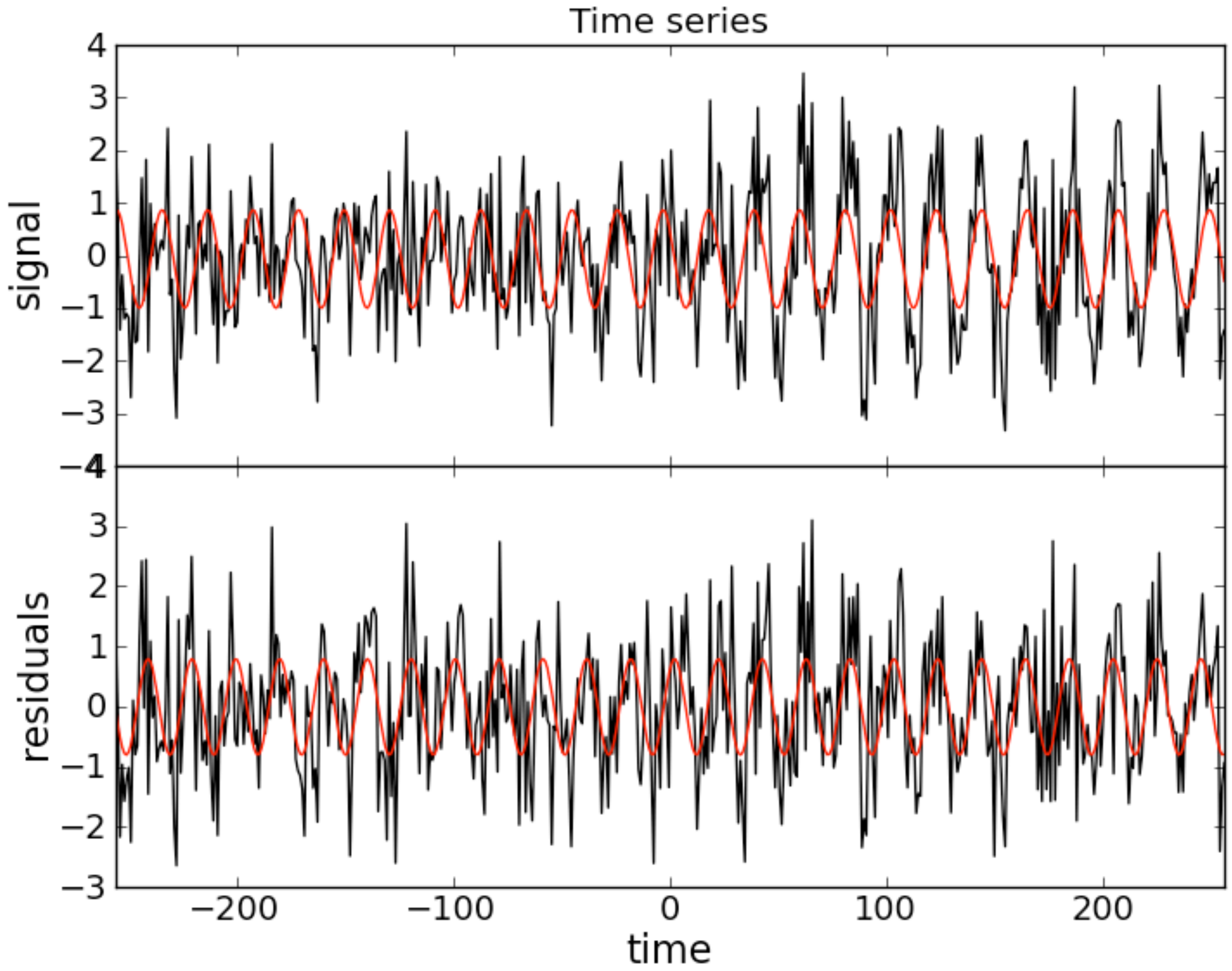
Time series
signal
residuals
time
5
4
3
2
1
0
−1
−2
−3
−200
−100
0
100
200

Noise only

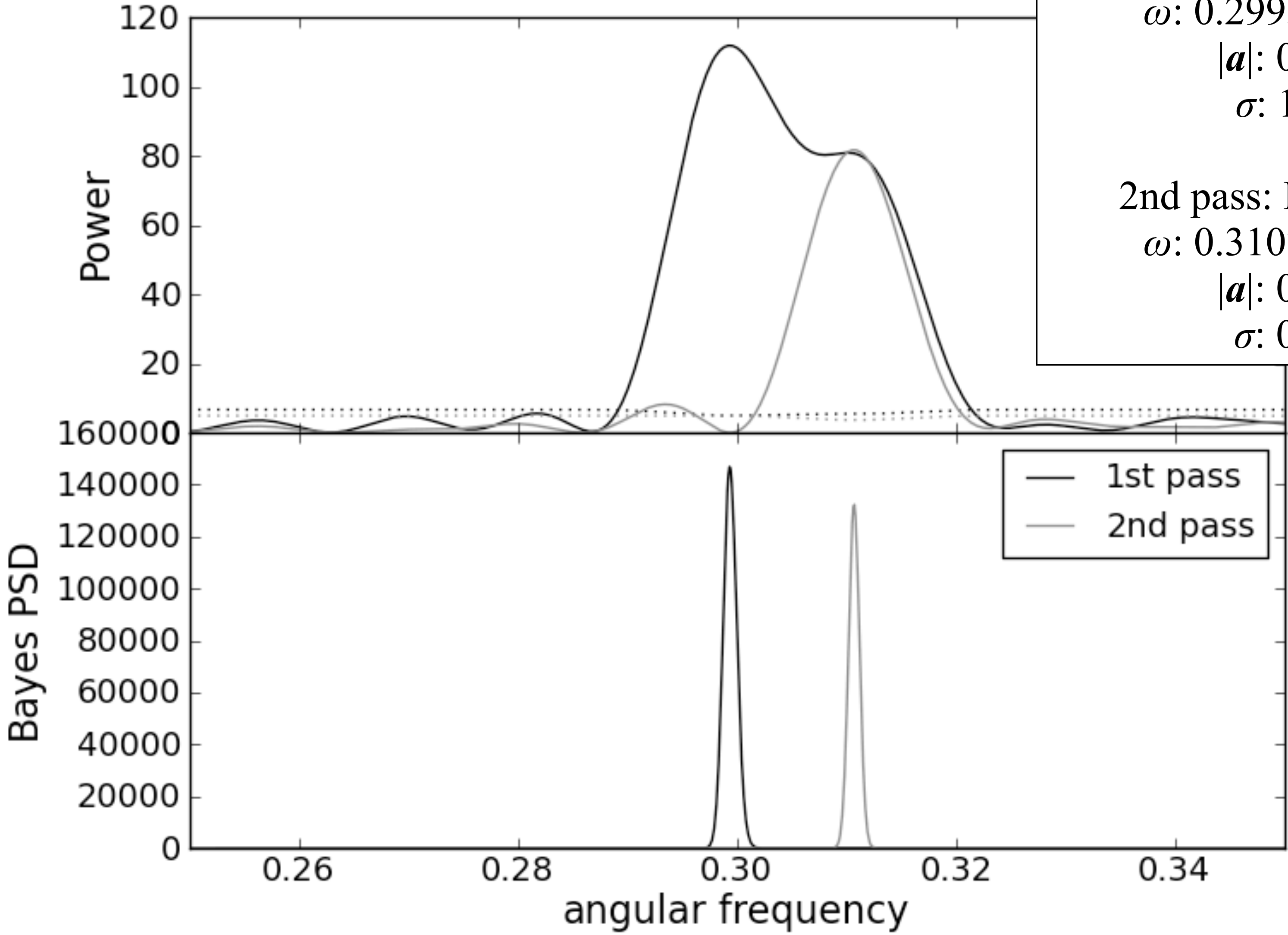


8
7
6
5
4
3
2
1
0
Power
DFT
Schuster
Noise
800
700
600
500
400
300
200
100
0
Bayes PSD
P(95): 0.27
Estimated angular frequency: 0.028 +/- 0.002
Estimated amplitude: 0.25 +/- 0.05
Estimated noise st.dev: 1.02 +/- 0.06
0.5
1.0
1.5
2.0
2.5
3.0
angular frequency

Double sinusoid + noise



Double sinusoid + noise

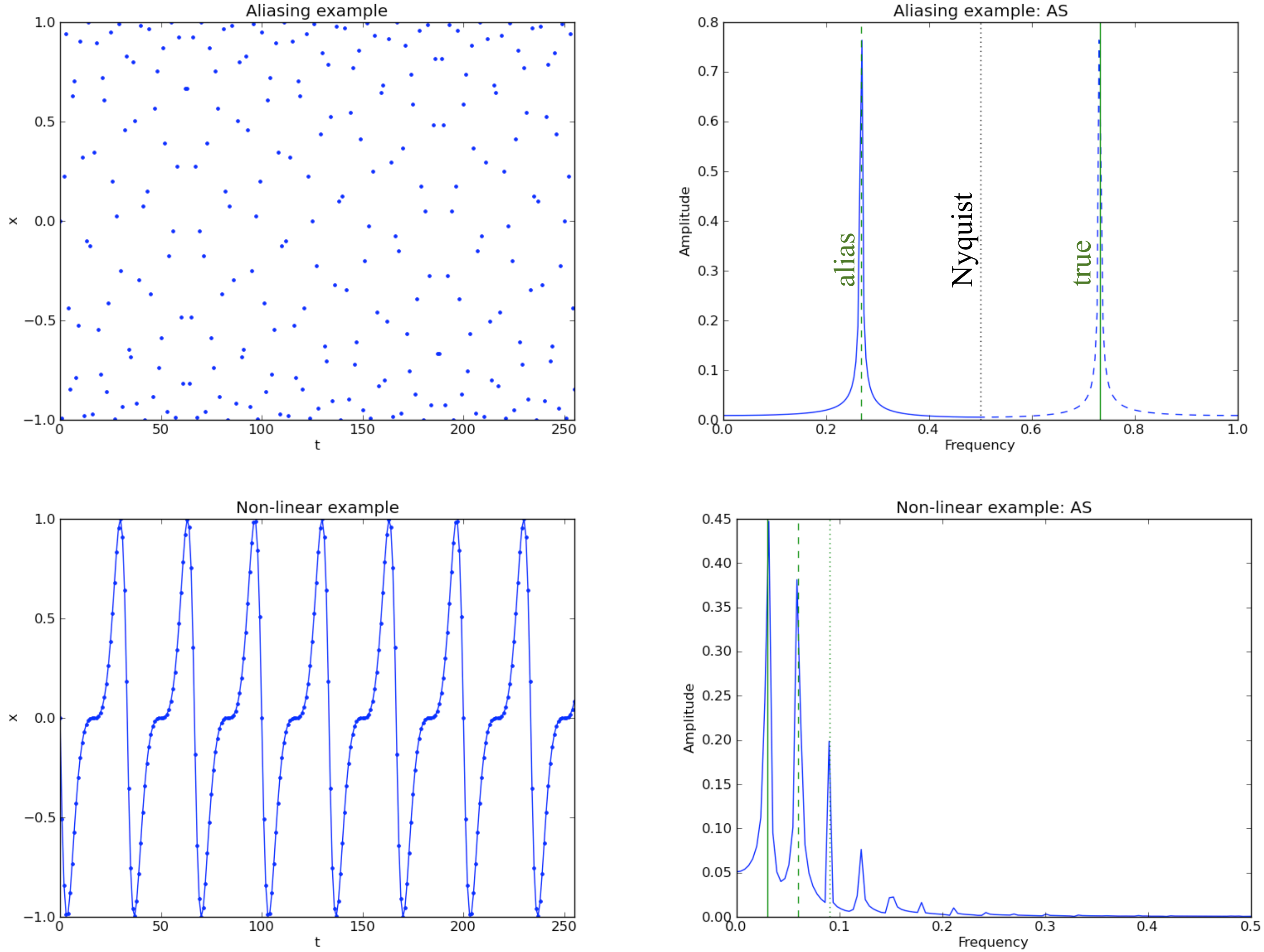


1st pass: P(95): 0.978
ω: 0.2993 +/- 0.0007
$|\boldsymbol{a}|$: 0.93 +/- 0.05
σ: 1.11 +/- 0.06

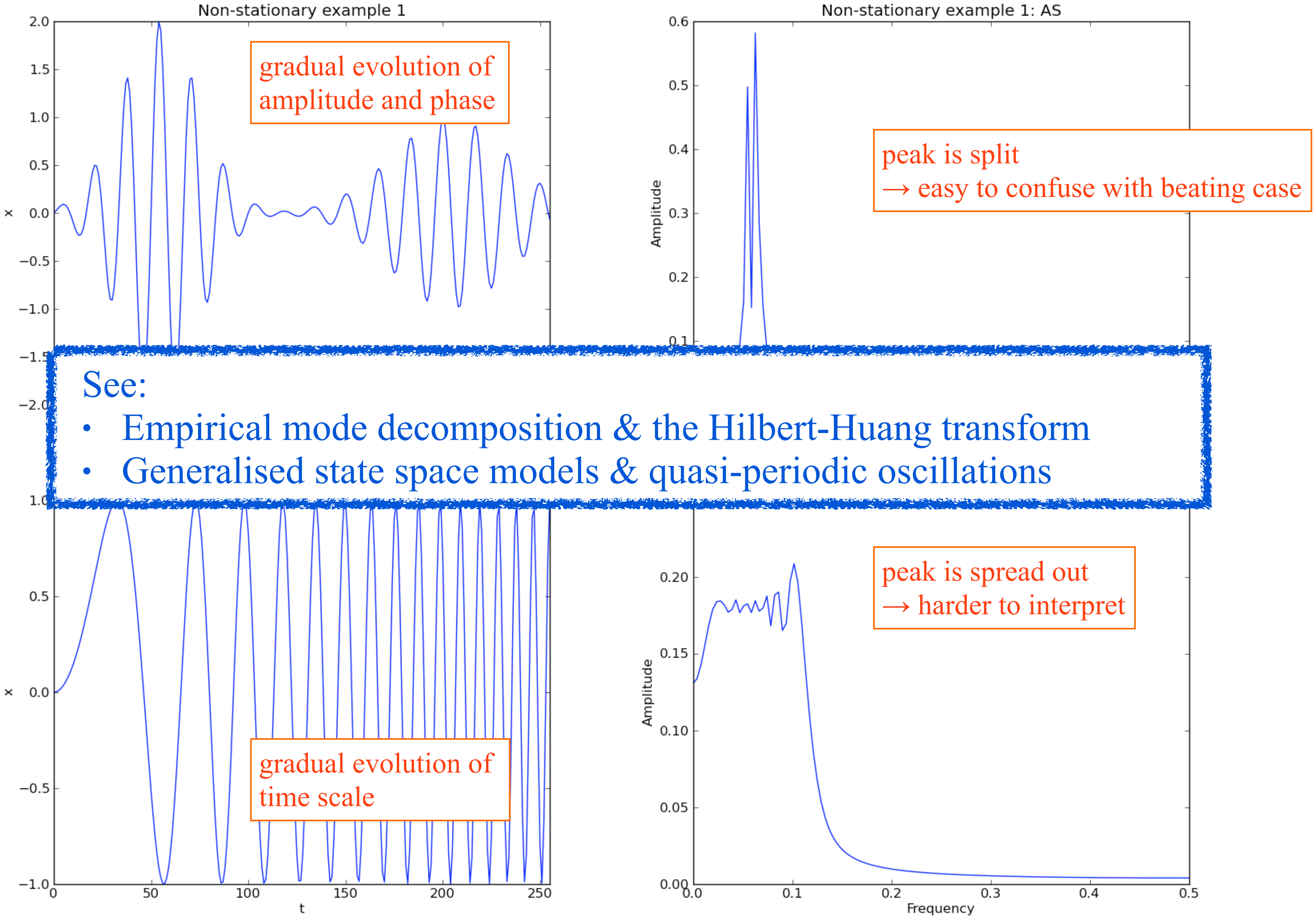
2nd pass: P(95): 0.996
ω: 0.3107 +/- 0.0007
$|\boldsymbol{a}|$: 0.80 +/- 0.04
σ: 0.95 +/- 0.06

Simple examples 3 - exploratory spectral analysis

Aliasing / harmonics example



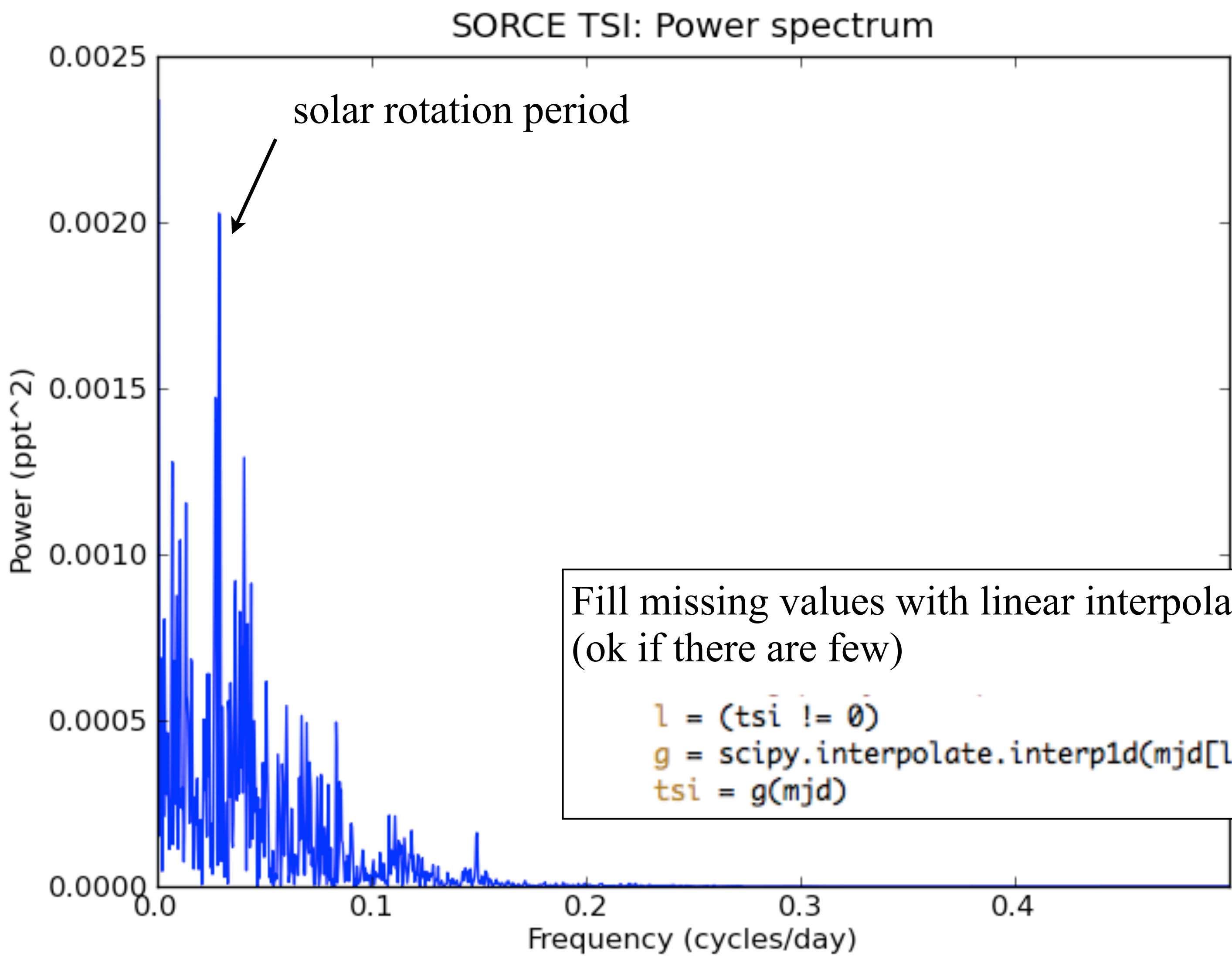
What if the signal isn't stationary?



See:

- Empirical mode decomposition & the Hilbert-Huang transform
- Generalised state space models & quasi-periodic oscillations

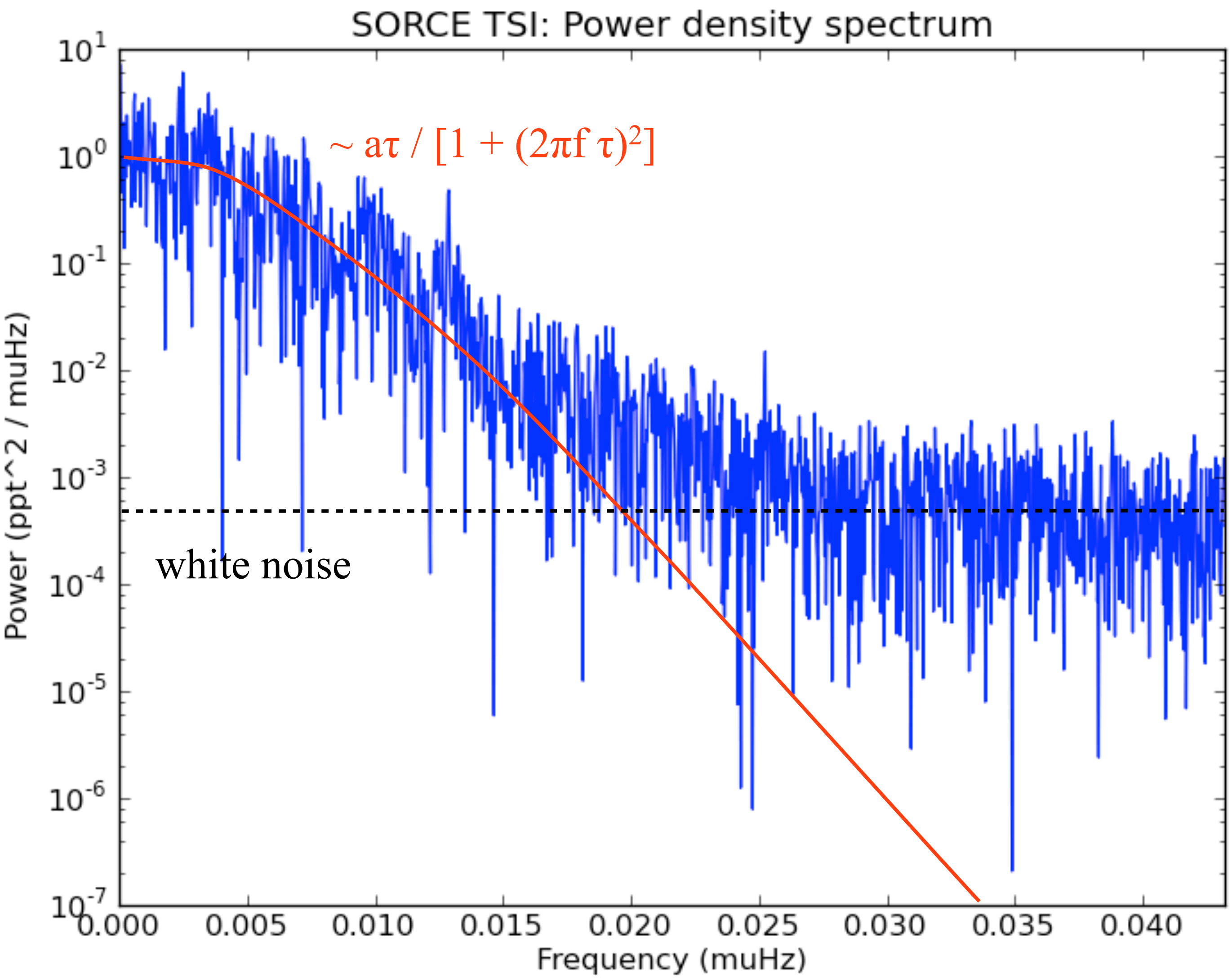
TSI – power spectrum



Fill missing values with linear interpolation
(ok if there are few)

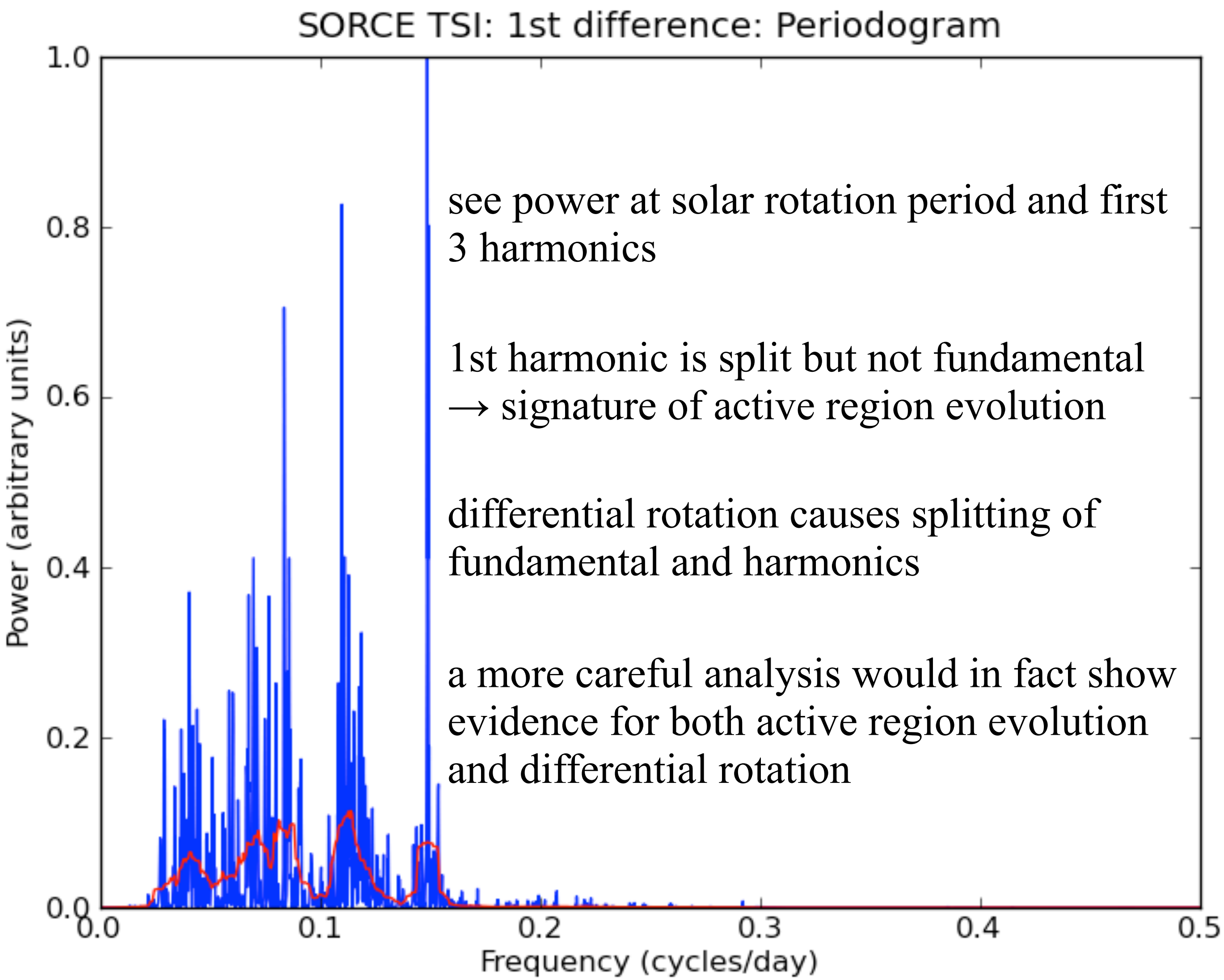
```
l = (tsi != 0)
g = scipy.interpolate.interp1d(mjd[l], tsi[l])
tsi = g(mjd)
```

TSI – power density spectrum



Spectral density

spectral density ≡ power spectrum of ACF
smoothed spectral density ≡ spectral plot



see power at solar rotation period and first 3 harmonics

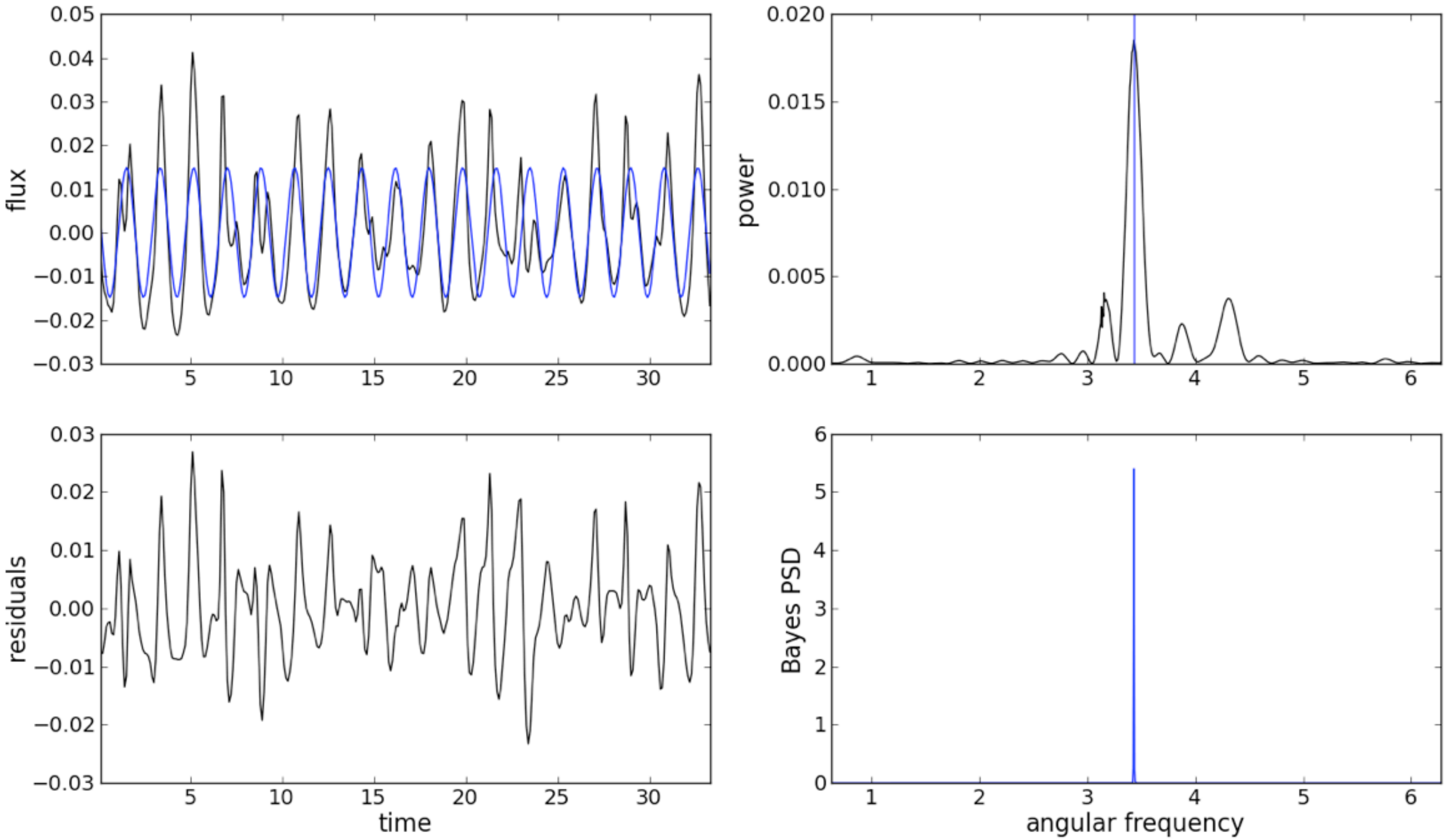
1st harmonic is split but not fundamental → signature of active region evolution

differential rotation causes splitting of fundamental and harmonics

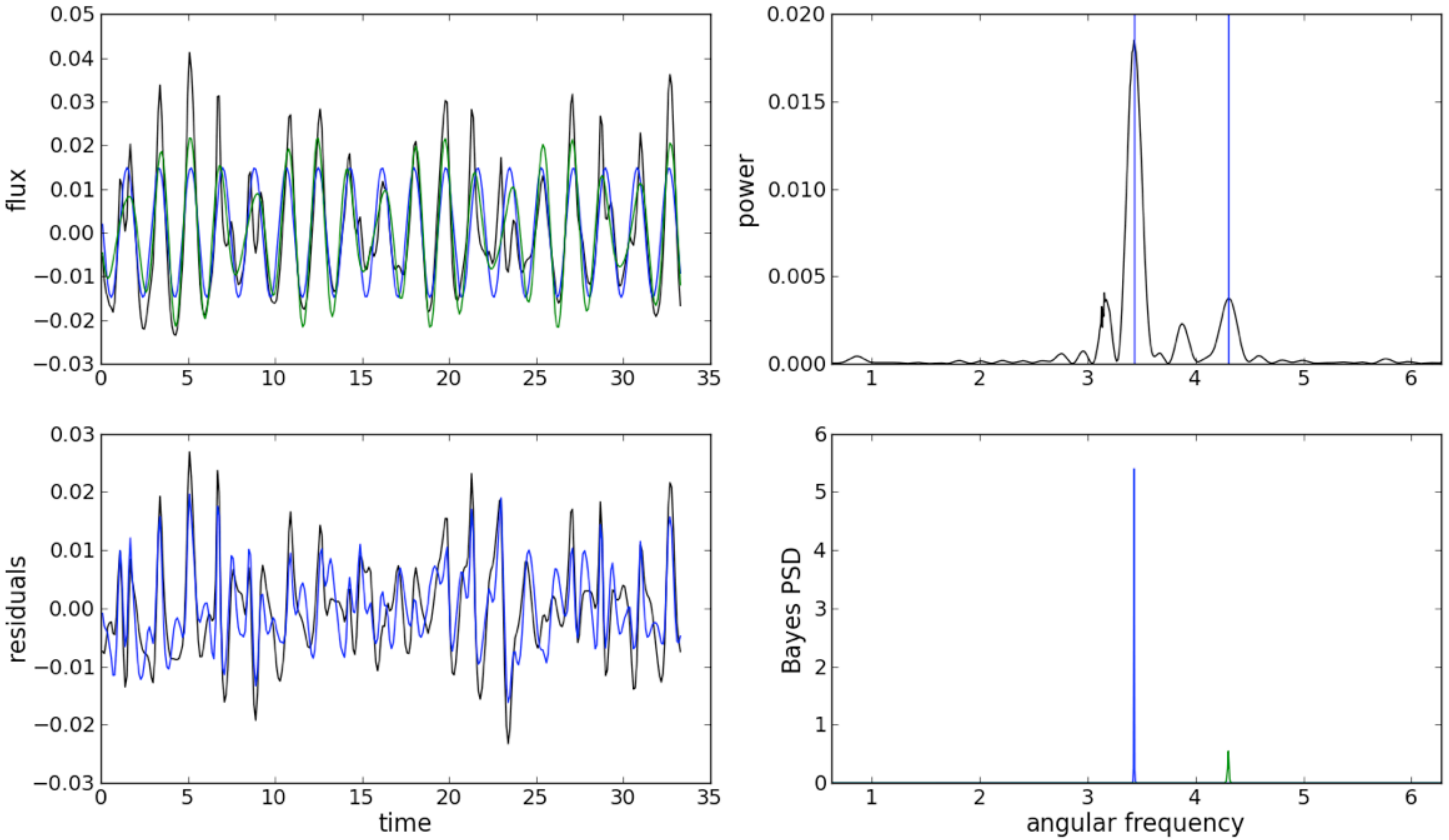
a more careful analysis would in fact show evidence for both active region evolution and differential rotation

Exercise 2: Spectral analysis of Kepler light curve (see dataset `Kepler_Q1_mod24_out4.mat`)

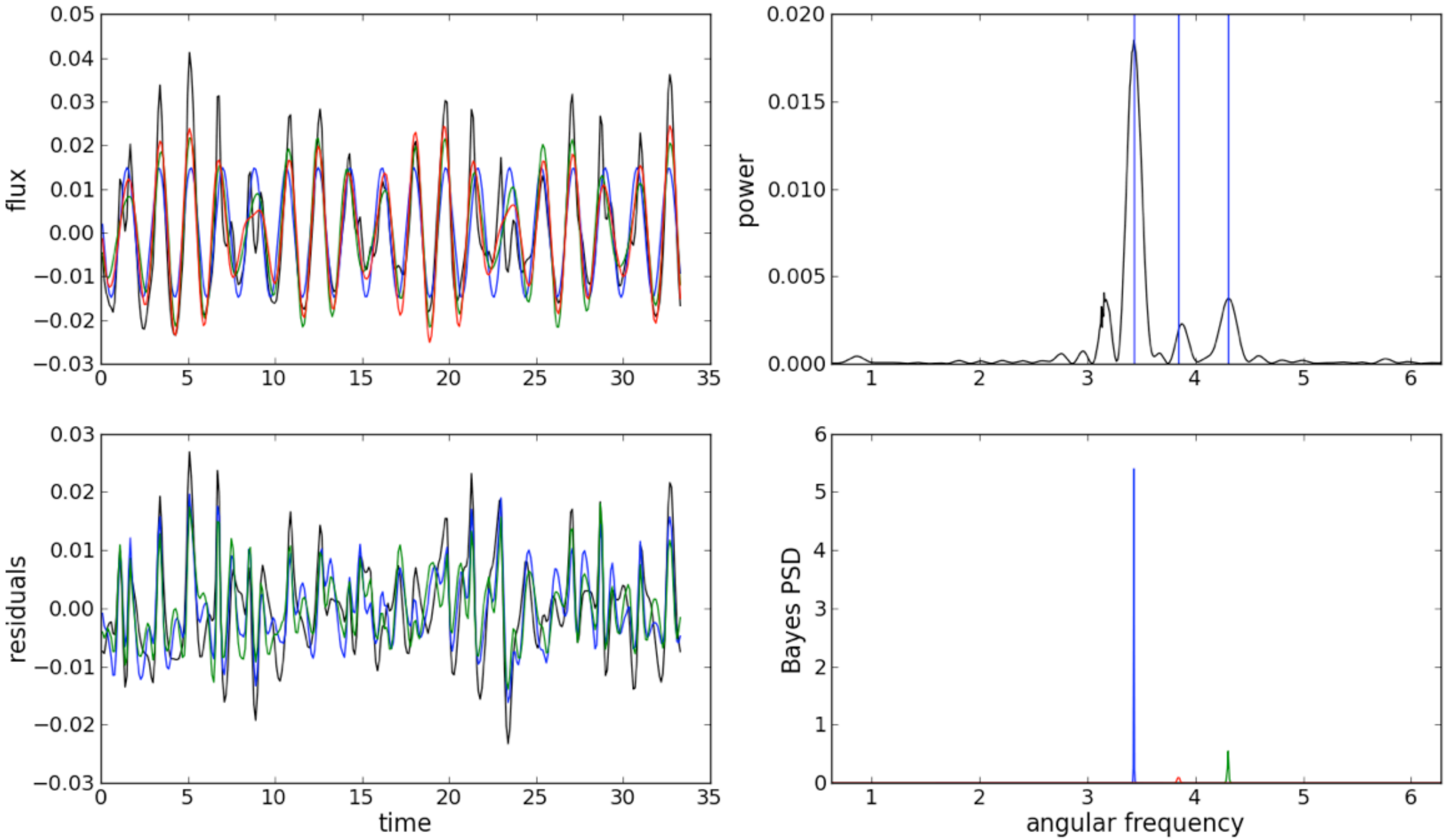
Pick the 19th light curve: multi-periodic pulsator
Rebin to $\delta t = 0.1$ to avoid overflow



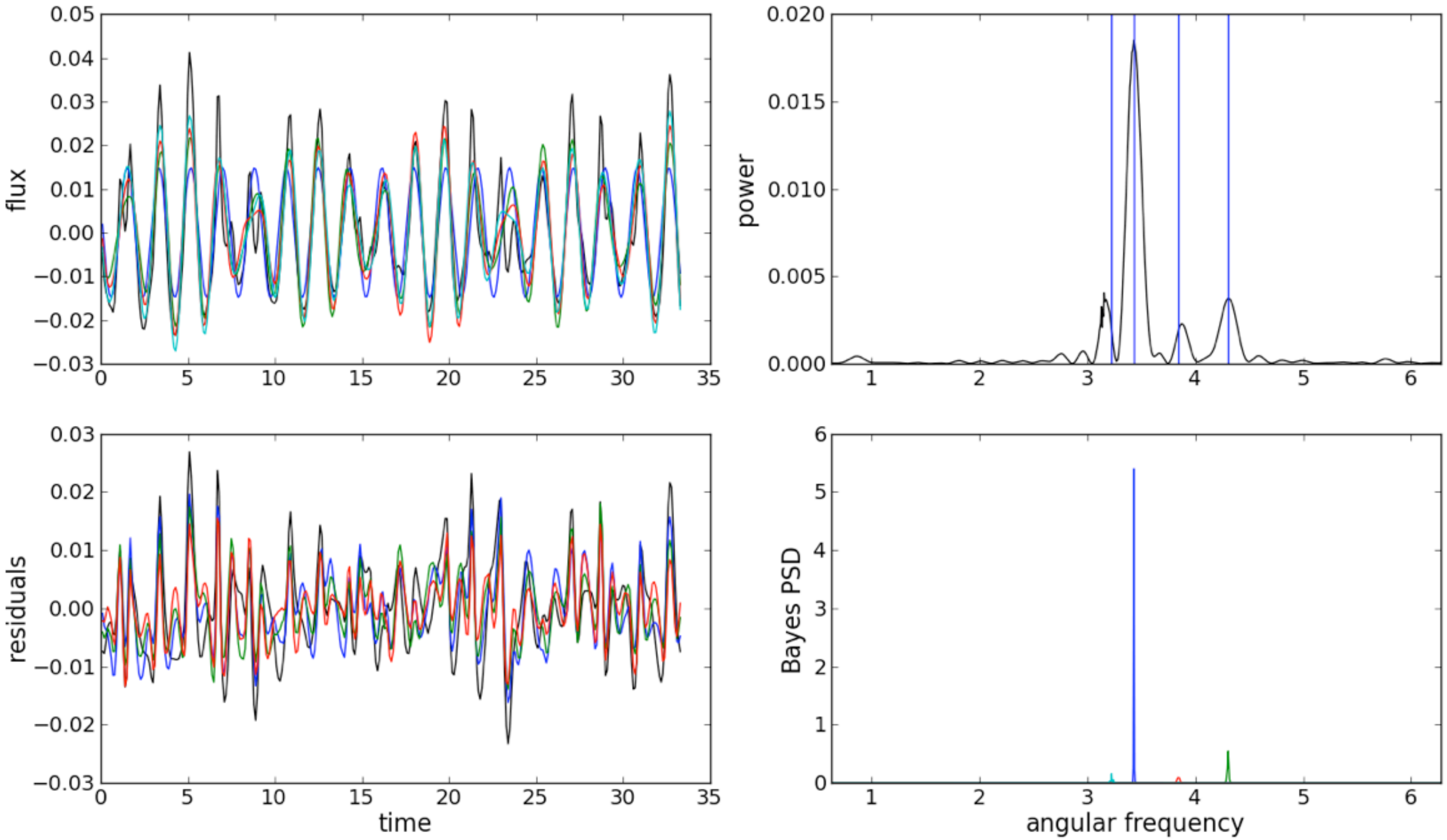
flux
time
residuals
power
Bayes PSD
angular frequency



flux
time
residuals
power
Bayes PSD
angular frequency



flux
time
residuals
power
Bayes PSD
angular frequency



flux
time
residuals
power
Bayes PSD
angular frequency

What about irregularly sampled data?

Making the basis orthonormal

- Not only simplifies likelihood but ensures uncorrelated a_j's

Making the basis orthonormal

- Not only simplifies likelihood but ensures uncorrelated a_j's
- General linear basis model: $f(t) = \Sigma_j\, a_j\, g_j(t, \boldsymbol{\omega})$ $\quad (j = 1, ..., M)$
    - Rewrite as $f(t) = \Sigma_k\, b_k\, h_k(t)$ such that $\Sigma_i\, h_j(t_i)\, h_k(t_i) = \delta_{jk}$
    - Form matrix $G_{jk} = \Sigma_i\, g_j(t_i)\, g_k(t_i)$
    - Define e_{jk} as j^{th} component of its k^{th} normalised eigenvector: $\Sigma_k\, G_{jk}\, e_{kl\ lk} / \lambda_l\, e_{lj}$
    - Orthogonality achieved if $h_k(t) = \Sigma_{kj}\, e_{kj}\, g_j(t) / (\lambda_k)^{1/2}$
    - Then $a_j = \Sigma_k\, [b_k\, e_{kj} / (\lambda_j)^{1/2}]$

Making the basis orthonormal

- Not only simplifies likelihood but ensures uncorrelated a_j's
- General linear basis model: $f(t) = \Sigma_j\, a_j\, g_j(t,\boldsymbol{\omega})$ $(j = 1, ..., M)$
  - Rewrite as $f(t) = \Sigma_k\, b_k\, h_k(t)$ such that $\Sigma_i\, h_j(t_i)\, h_k(t_i) = \delta_{jk}$
- Single sinusoid in regularly sampled data is only approximately orthogonal
  - $h_k = (c_1)^{-1/2} \cos(\omega t) + (c_2)^{-1/2} \sin(\omega t)$ where $\boldsymbol{c} = (N/2) \pm \sin(N\omega) / 2 \sin(\omega)$
  - $C'(\omega) = [R(\omega)^2 / c_1 + I(\omega)^2 / c_2] / N$

Making the basis orthonormal

- Not only simplifies likelihood but ensures uncorrelated a_j's
- General linear basis model: $f(t) = \Sigma_j\, a_j\, g_j(t,\boldsymbol{\omega})$ $\quad\quad$ $(j = 1, ..., M)$
    - Rewrite as $f(t) = \Sigma_k\, b_k\, h_k(t)$ such that $\Sigma_i\, h_j(t_i)\, h_k(t_i) = \delta_{jk}$
- Single sinusoid in regularly sampled data
    - $C'(\omega) = [R(\omega)^2 / c_1 + I(\omega)^2 / c_2] / N$
- Generally, estimate numerically
    - `eigval, eigvec = numpy.linalg.eig(a)`

Making the basis orthonormal

- Not only simplifies likelihood but ensures uncorrelated a_j's
- General linear basis model: $f(t) = \Sigma_j\, a_j\, g_j(t,\boldsymbol{\omega})$ $(j = 1, ..., M)$
  - Rewrite as $f(t) = \Sigma_k\, b_k\, h_k(t)$ such that $\Sigma_i\, h_j(t_i)\, h_k(t_i) = \delta_{jk}$
- Single sinusoid in regularly sampled data
  - $C'(\omega) = [R(\omega)^2 / c_1 + I(\omega)^2 / c_2] / N$
- Generally, estimate numerically
  - `eigval, eigvec = numpy.linalg.eig(a)`
- Rest of analysis proceeds as before
  - Obtain $\mathcal{L}(\boldsymbol{\omega}) \propto [1 - (\Sigma_k\, \hat{h}_k^2 / N\hat{Y})]^{(m-N)/2}$ where $\hat{h}_k = \Sigma_i\, \delta y_i\, h_j(t_i)$
  - or $\mathcal{L}(\boldsymbol{\omega}\sigma) \propto \sigma^{m-N} \exp(-N\hat{Y}/2\sigma^2) \times \exp(\Sigma_k\, \hat{h}_k^2 / 2\sigma^2)$ if σ is known

The Lomb-Scargle periodogram

The Lomb-Scargle periodogram

- Scargle (1972), Horne & Baliunas (1989)

The Lomb-Scargle periodogram

- Scargle (1972), Horne & Baliunas (1989)
- Bayesian spectral analysis for:
  - single sinusoid + white noise model
  - **mildly** irregular sampling

The Lomb-Scargle periodogram

- Scargle (1972), Horne & Baliunas (1989)
- Bayesian spectral analysis for:
  - single sinusoid + white noise model
  - **mildly** irregular sampling
- Beware
  - noise is assumed known, white, and constant over dataset
  - if the sampling is strongly irregular
    - trick to achieve orthogonality breaks down
    - tests against null hypothesis (“significance”, “false alarm probability”) break down

The generalised least-squares periodogram

The generalised least-squares periodogram

- Schuster & .... [though widely used long before]

The generalised least-squares periodogram

- Schuster & .... [though widely used long before]
- **Maximum likelihood** for
  - single sinusoid + offset + white noise
  - irregular time sampling
  - variable measurement uncertainties

The generalised least-squares periodogram

- Schuster & .... [though widely used long before]
- **Maximum likelihood** for
  - single sinusoid + offset + white noise
  - irregular time sampling
  - variable measurement uncertainties
- Beware
  - noise is assumed known and white
  - no recipe for testing agains null hypothesis provided

In practice...

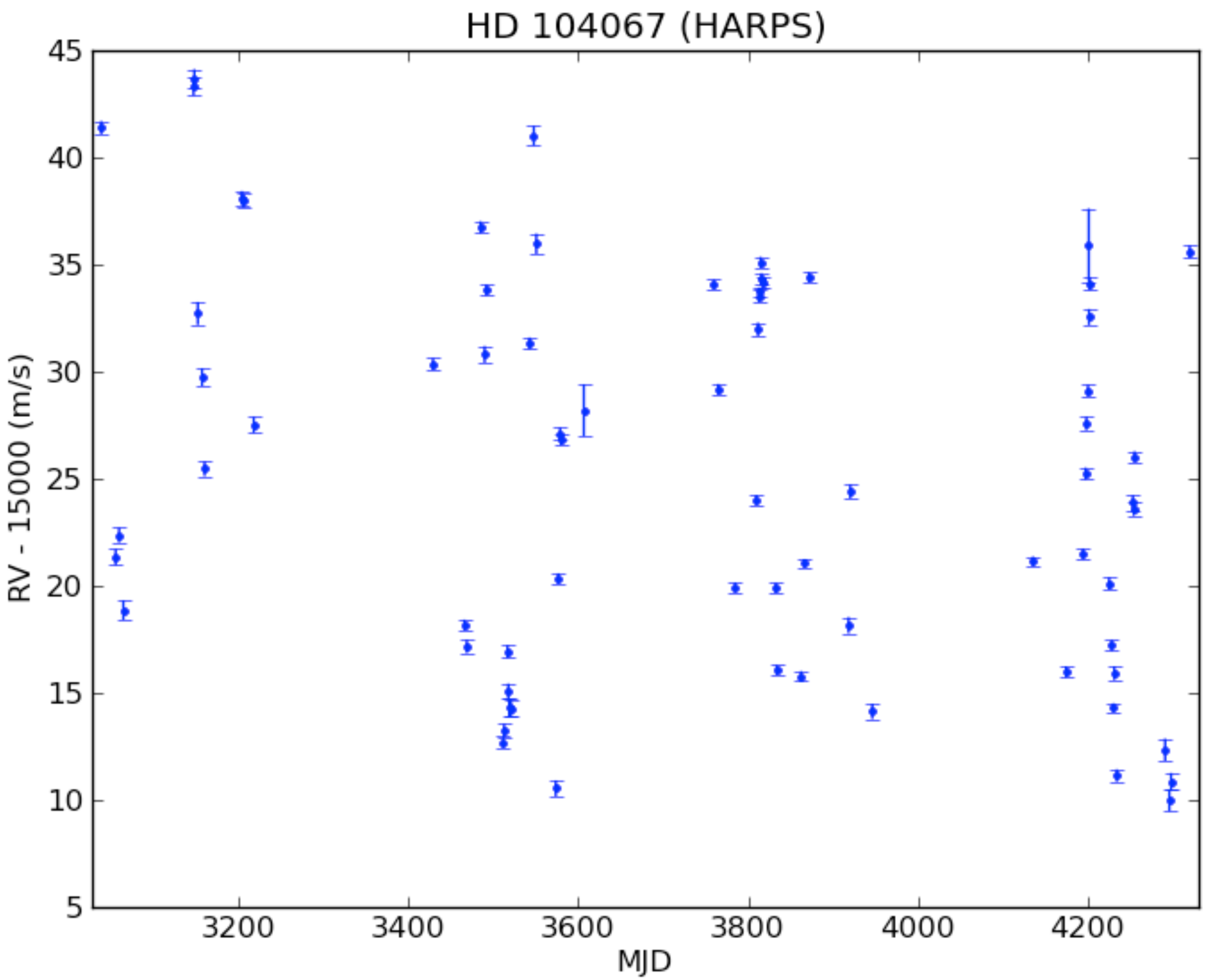
In practice...

- Use Lomb-Scargle / GLS periodogram as exploratory tool

In practice...

- Use Lomb-Scargle / GLS periodogram as exploratory tool
- Final analysis should always be fully Bayesian
  - Break down into linear / nonlinear parts
  - Apply principles described above for linear part
  - (Clever) sampling methods needed for nonlinear parameters

Exercise 3: HARPS RV data of HD 104067
(see dataset `HD104067.dat`)



HD 104067 (HARPS)
RV - 15000 (m/s)
MJD
45
40
35
30
25
20
15
10
5
3200
3400
3600
3800
4000
4200

Analysing the HARPS data of HD104067

- Fairly obvious periodic modulation
  - Try it, module `gls.py`
- Some hint of trend in residuals
- Questions:
  - If the modulation is due to a planet, what are its parameters?
    - Keplerian orbit: module `orbit.py`
  - Can we say much about any other signal in the data?
    - This data is not published, is it because the HARPS team are waiting for a 2^{nd} planet to become significant?
  - How realistic are the measurement uncertainties?
    - Measurement uncertainties for RVs are notoriously hard to measure
    - Particularly sensitive to stellar rotation rate (broad lines) and signal-to-noise of spectrum (weather)

Stochastic processes: ARMA models

Auto-regressive models

The AR(p) process is defined by

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-1} + \epsilon_t$$

where $\boldsymbol{\epsilon} = \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$. This process is stationary if the roots of the polynomial

$$z^p - \sum_{i=1}^{p} \phi_i z^{p-1}$$

lie inside the unit circle.

Auto-regressive models

http://en.wikipedia.org/wiki/Autoregressive_model

ACF and spectral density of AR(p) processes

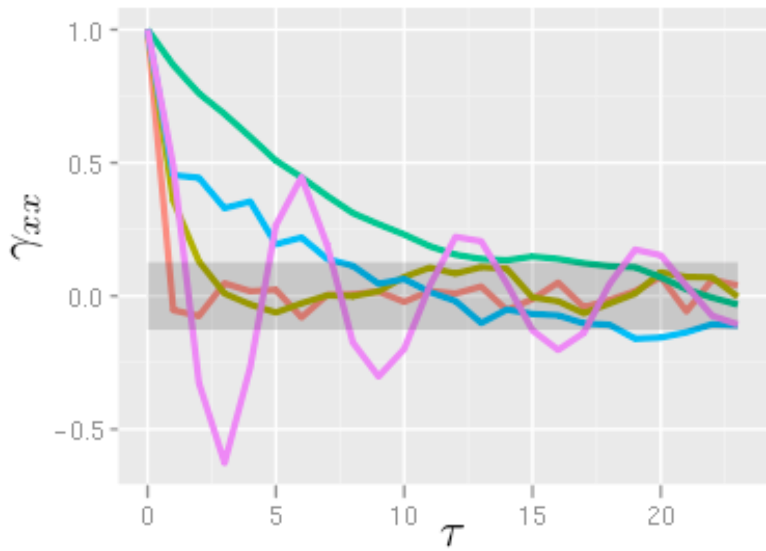
The ACF of an AR(p) process has the form

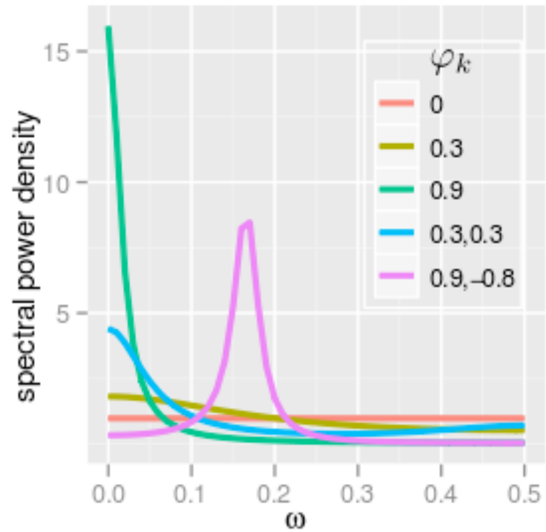
$$\rho(h) = \sum_{k=1}^{p} a_k y_k^{-|h|}$$

where the y_k are the roots of the polynomial $z^p - \sum_{i=1}^{p} \phi_i z^{p-1}$.

The spectral density of an AR(p) process is

$$S(f) \propto \frac{1}{|1 - \sum_{k=1}^{p} \phi_k e^{-2\pi i k f}|^2}$$





http://en.wikipedia.org/wiki/Autoregressive_model

Spectral density of an AR(1) process

The AR(1) process $x_t = c + \phi x_{t-1} + \epsilon_t$ has ACV

$$\gamma_n = \frac{\sigma_\epsilon^2}{1 - \phi^2} \phi^{|n|}$$

which decays exponentially with decay constant $\tau = -1 \ \ln \phi$.

Provided $\tau \gg 1$, i.e. the sampling interval is much shorter than the decay time, we can treat the ACF as continuous. The spectral density is then

$$S(f) \propto \int_{-\infty}^{\infty} \gamma(t) e^{-2\pi i f} \mathrm{d}t \propto \frac{\tau}{1 + (2\pi f \tau)^2}$$

Spectral density of an AR(1) process

The AR(1) process $x_t = c + \phi x_{t-1} + \epsilon_t$ has ACV

$$\gamma_n = \frac{\sigma_\epsilon^2}{1-\phi^2}\phi^{|n|}$$

which decays exponentially with decay constant $\tau = -1 \;\; \ln\phi$.

Provided $\tau \gg 1$, i.e. the sampling interval is much shorter than the decay time, we can treat the ACF as continuous. The spectral density is then

$$S(f) \propto \int_{-\infty}^{\infty} \gamma(t) e^{-2\pi i f} \mathrm{d}t \propto \frac{\tau}{1 + (2\pi f \tau)^2}$$

Recall TSI...

Finding the coefficients of an AR(p) model

- Fit the data: $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$
  - model each X_t as a linear combination of the previous p observations
  - get $n-p$ simultaneous equations
- But
  - complicated for $p>1$
  - how do you chose p anyway?

Fitting the ACV: Yule-Walker equations

$$\mathrm{E}[X_t X_{t-m}] = \mathrm{E}\left[\sum_{i=1}^{p} \varphi_i\, X_{t-i} X_{t-m}\right] + \mathrm{E}[\varepsilon_t X_{t-m}].$$

$= \sigma_\varepsilon^2$ if m=0, 0 otherwise

$$= \sum_{i=1}^{p} \varphi_i\, \mathrm{E}[X_t X_{t-m+i}] = \sum_{i=1}^{p} \varphi_i\, \gamma_{m-i},$$

so that $$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

- In matrix form, for *m*>1

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \end{bmatrix}$$ or $$\mathbf{R}\mathbf{\Phi} = \mathbf{r}.$$

- This is well posed and the matrix **R** is always invertible.

But what about choosing p?

- Given the sample ACF up to $n_{\max} \leq n/4$
  - Compute the sample partial ACF (PACF)
    - Estimate $\boldsymbol{r}$ and $\mathbf{R}$ in the YW equations
    - keep the first k rows only
    - solve for the AR coefficients $\boldsymbol{\varphi}$
    - PACF(k) = φ_k
- PACF(k) measures the amount of correlation not accounted for by an AR(k−−1) model
- The PACF of an AR(p) model is (indistinguishable from) 0 for lags > p

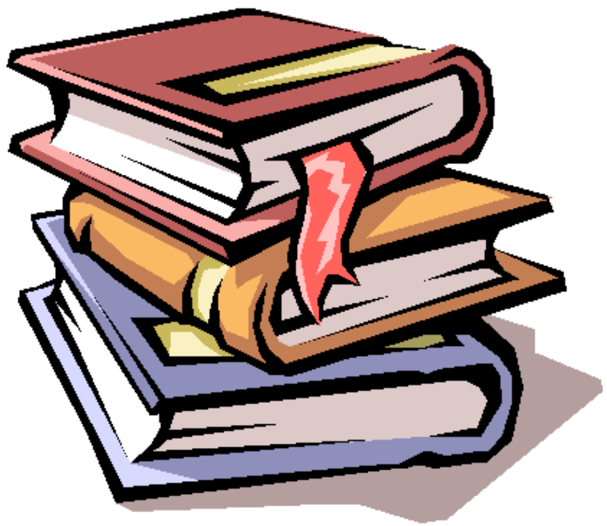
Moving average & ARMA models

The MA(q) process is defined by $x_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-1} + \epsilon_t$ where $\boldsymbol{\epsilon} = \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$.

The ACF of an MA(p) process cuts off after lag $> q$.

The PACF of an MA decays exponentially, with or without oscillations, much like the ACF of an AR process.

The ARMA(p, q) process is defined by $x_t = \sum_{i=1}^{q} \phi_i x_{t-1} + \sum_{i=1}^{p} \theta_i \epsilon_{t-1} + \epsilon_t$ where $\boldsymbol{\epsilon} = \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$.



Before applying AN ARMA model...

- Plot x_t vs t
    - Is there a trend?
    - If yes, plots the first difference
    - repeat d times until stationary
    - Apply ARMA model to differenced data = ARIMA(p,d,q)
- Plot the spectral density
    - Is there a significant peak (seasonality)?
    - If so, compute the seasonally differenced data $x_t - x_{t-s}$
    - Use a seasonal ARMA model
        - relates x_t to x_{t-si} rather than to x_{t-i}

Box-Jenkins model selection

- Plot ACF & PACF
    - ACF gradually decays, PACF cuts of after p lags
        - AR(p) model
    - ACF cuts off after q lags, PACF gradually decays
        - MA(q) model
    - Both ACF and PACF decay gradually, starting after a few lags
        - Mixed ARMA model
- Fit model coefficients
    - No python module to my knowledge, but they exist in R & Matlab.
- Test & compare models with different p, q

Distribution of sample ACF

- The fitting of ARMA models relies largely on modelling the sample ACF.
- To estimate the goodness of fit, we need some knowledge of the distribution of the sample ACF. This is usually intractable, but...

For large n, the sample ACF is approximately normally distributed:

$$\hat{\boldsymbol{\rho}} \simeq \mathcal{N}(\boldsymbol{\rho}, n^{-1}\boldsymbol{W})$$

where $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(h_{\max}))^T$ and the elements of the covariance matrix $\boldsymbol{W}$ are given by

$$w_{ij} = \sum_{k=1}^{\infty} \{[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \times [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)]\}\,.$$

ARMA-like models in astrophysics

ARMA-like models in astrophysics

- ARMA models and relative are most widely used for prediction
    - rarely what we need to do in astrophysics

ARMA-like models in astrophysics

- ARMA models and relative are most widely used for prediction
    - rarely what we need to do in astrophysics
- They do give us a framework for understanding stochastic processes

ARMA-like models in astrophysics

- ARMA models and relative are most widely used for prediction
  - rarely what we need to do in astrophysics
- They do give us a framework for understanding stochastic processes
- They can be built into state-space models, providing a means to study quasi-periodic oscillations

A few things before we start

- All the documents, code and the PDF of yesterday's lecture are on the wiki
  - Use them to go back over anything that wizzed past a bit too fast yesterday...
  - One code I haven't supplied: full Bayes spectral analysis for irregular data.
- Today: newer methods, more conceptual, less math
  - relevant papers and textbooks are available on the wiki if you want to go further
- Time for "exercises"
  - getting ready for David Hogg's model selection workshop
  - anything else you want to ask me about the 3 datasets and code I've supplied

Advertising

- IAU Syposium 285: New Horizons in Time Domain Astronomy, 19-23 September 2011, St Catherine College, Oxford
  - http://www.physics.ox.ac.uk/IAUS285/
  - talks in the morning, hands-on workshops in the afternoons

- SFTC graduate school “Exoplanets and their host stars”, 12-16 March 2012, St Anne’s College, Oxford
  - http://www.physics.ex.ac.uk/EAHS12
  - contact me for more information

Stochastic processes: Gaussian processes

Most plots in this section are from Rasmussen & Williams (2009).

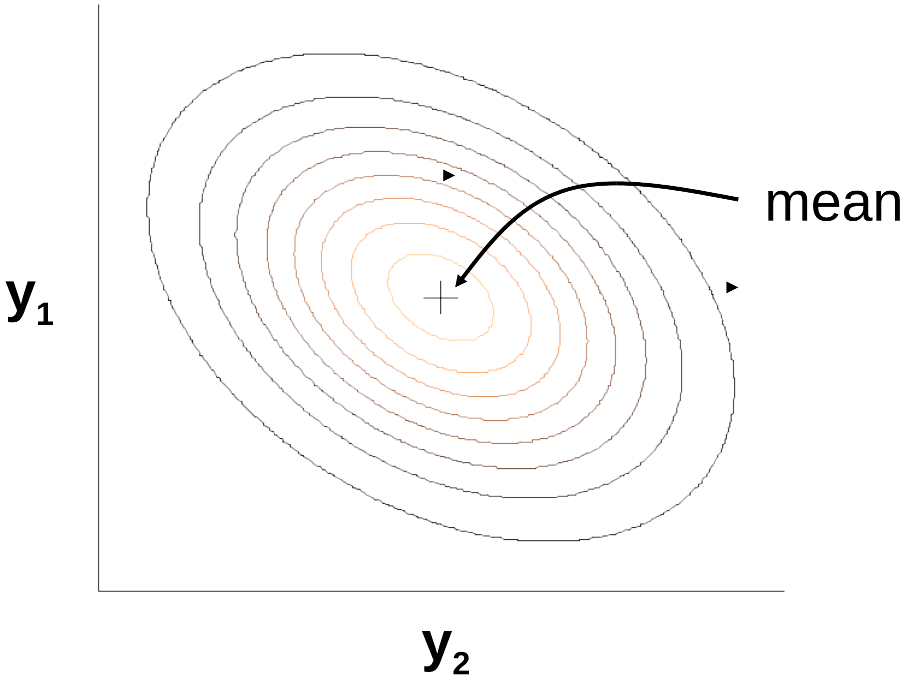
Motivation

- Typically model time-series as $y_i = f(t_i, \boldsymbol{\theta}) + \varepsilon_i$ where
  - $f(t, \boldsymbol{\theta})$ is a parametric model
  - ε_i is IID noise
- But
  - IID noise assumption almost never holds
  - Process of interest may be stochastic
  - May have additional information (housekeeping data) but not sure how to tie it in
- We would like to incorporate **random functions** (with certain properties) into our model
- Gaussian processes are probability distributions over random functions
  - Generalisation of Bayesian linear regression to random functions (via kernel trick)

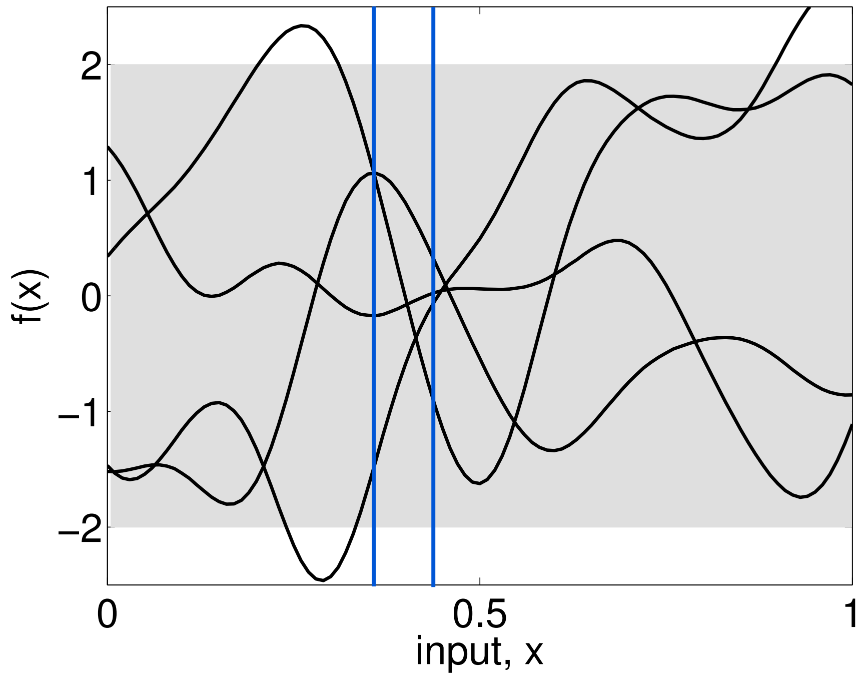
What is a Gaussian process?

- Joint distribution of samples from a Gaussian process = multivariate Gaussian: $\mathcal{N}(m(\boldsymbol{t}), \mathbf{K})$
  - $m(t)$ is the mean function (can be parameteric
  - $\mathbf{K}$ is the covariance matrix (must be positive semi-definite)
  - $\mathbf{K}$ is the Gram matrix of some covariance function k: $K_{ij} = k(t_i, t_j)$

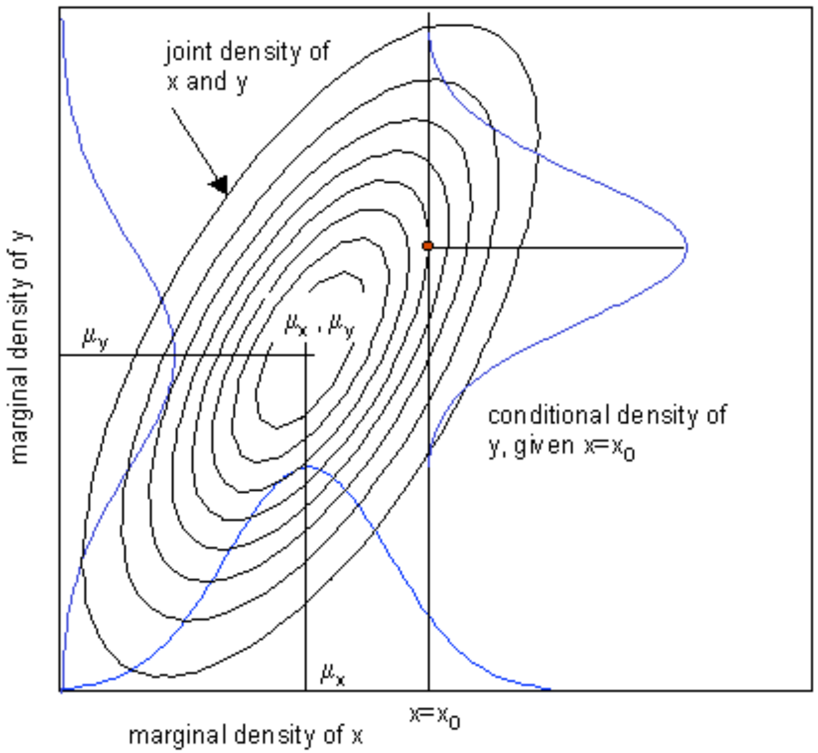
joint probability distribution



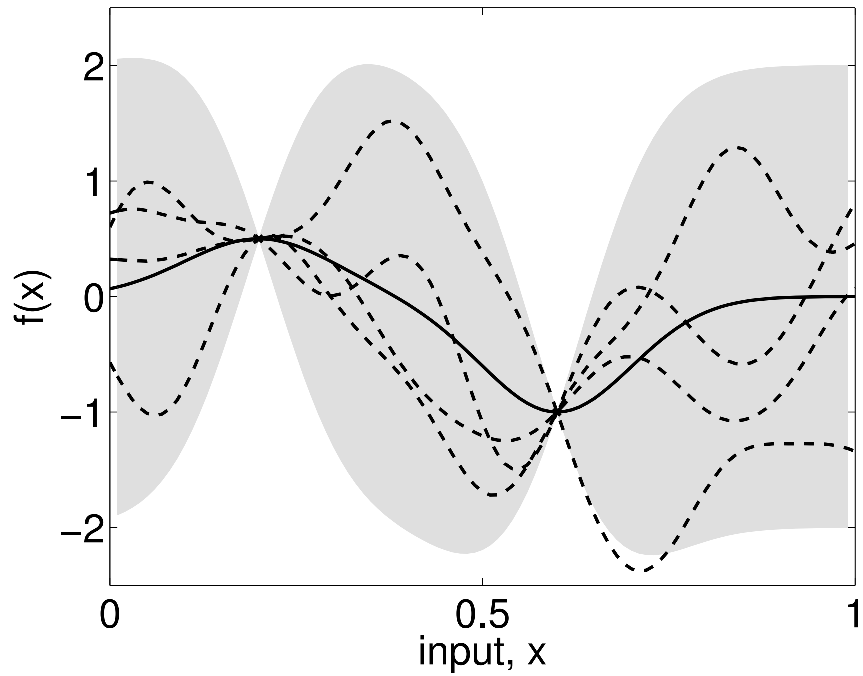
samples from the GP



conditional probability distribution

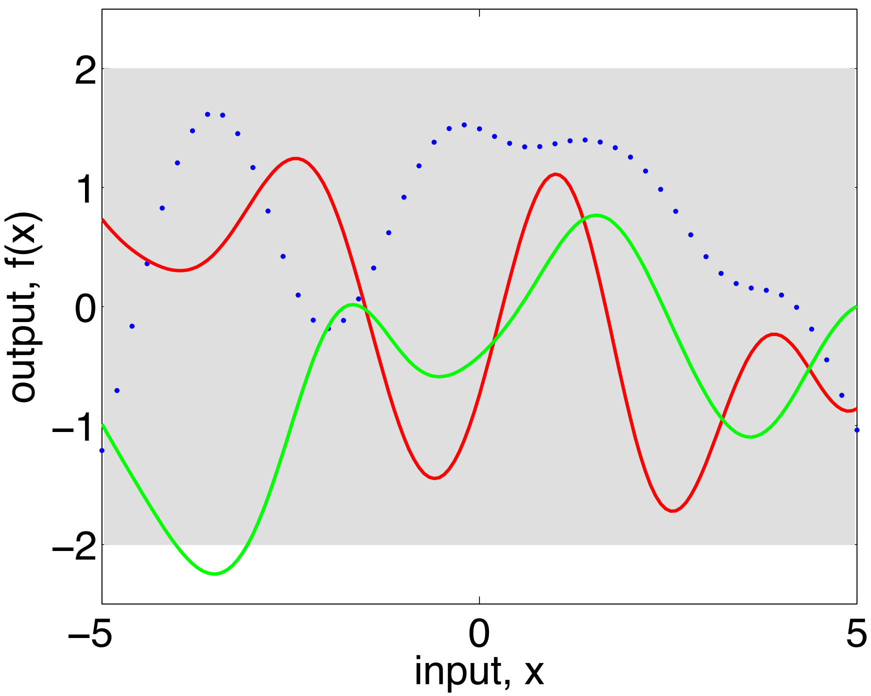


samples from conditioned GP

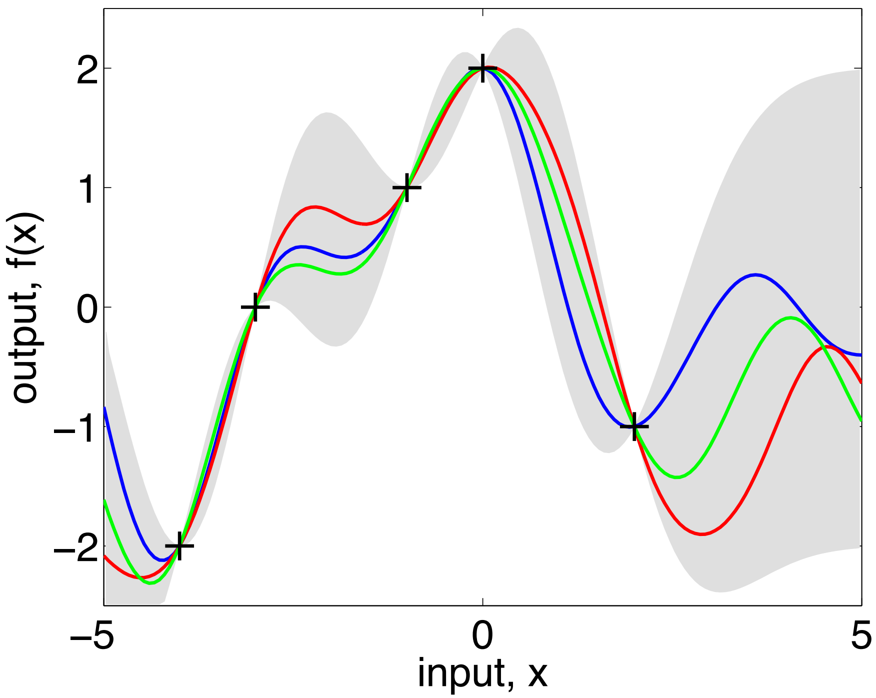


Modelling data with a GP

- Prior distribution (no data)
  - $P(\boldsymbol{y}_*|k,I) \sim \mathcal{N}(0, \mathbf{K})$, $\quad k(t,t') = \sigma_d^2 \exp[-(t-t')^2 / 2l^2]$
- Posterior distribution is conditioned on data



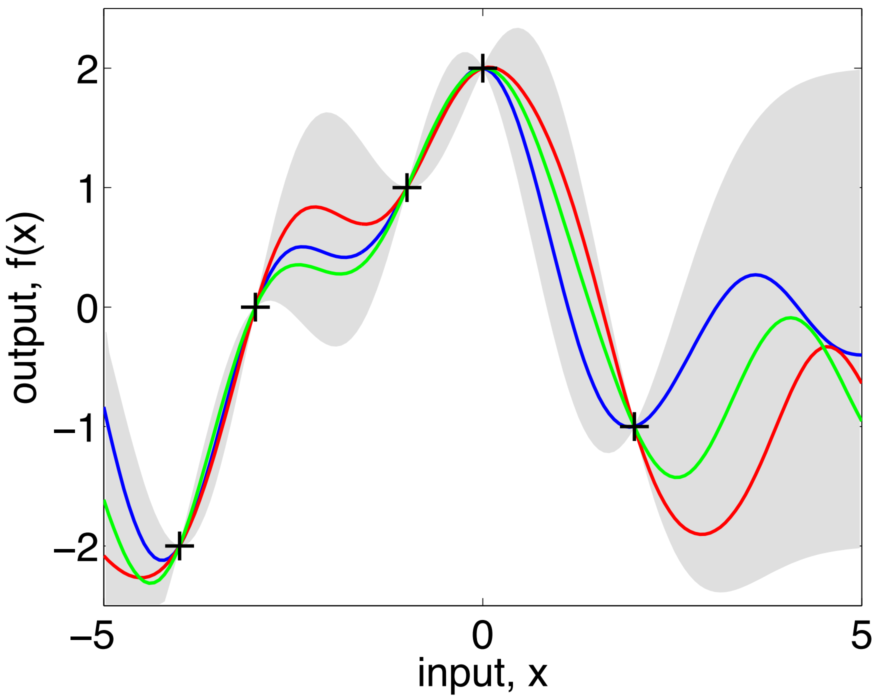
(a), prior



(b), posterior

Modelling data with a GP prior

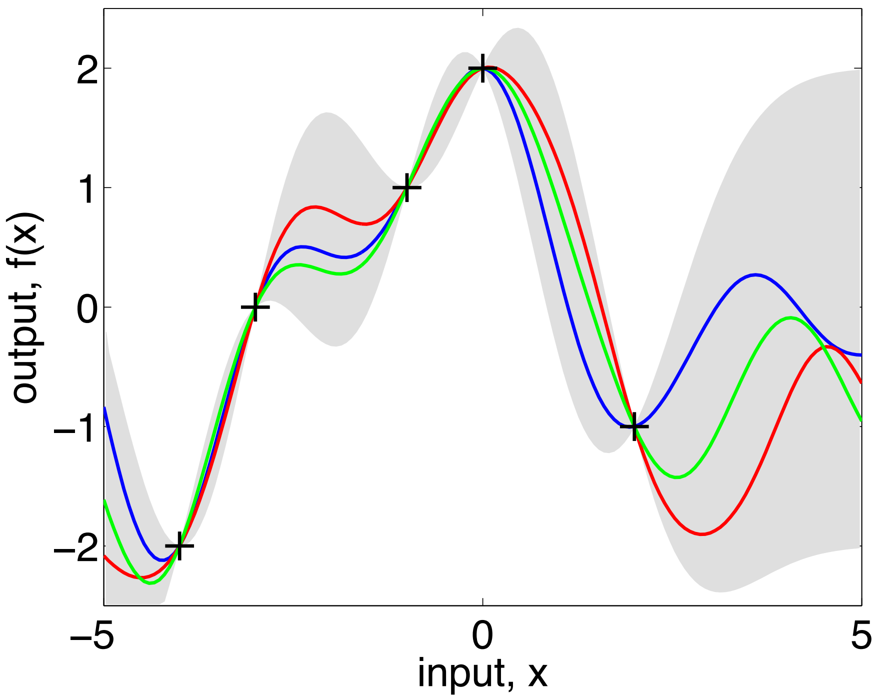
- Prior distribution (no data)
  - $P(\boldsymbol{y}_*|m,k,\mathrm{I}) \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$
- Posterior distribution is conditioned on data
  - $P(\boldsymbol{y}_*|m,k,\mathrm{D},\mathrm{I}) \sim \mathcal{N}(f_*,\mathrm{var}[f_*])$ where
    - $f_* = \boldsymbol{k}_*^{\mathrm{T}} (\mathbf{K} + \sigma_{\mathrm{n}}^2\mathbf{I})^{-1} \boldsymbol{y}$
    - $\mathrm{var}[f_*] = k(t_*,t_*) - \boldsymbol{k}_*^{\mathrm{T}} (\mathbf{K} + \sigma_{\mathrm{n}}^2\mathbf{I})^{-1} \boldsymbol{k}_*$
    - $\boldsymbol{k}_* = (k(t_1,t_*), \ldots, k(t_{\mathrm{N}},t_*))^{\mathrm{T}}$.



(b), posterior

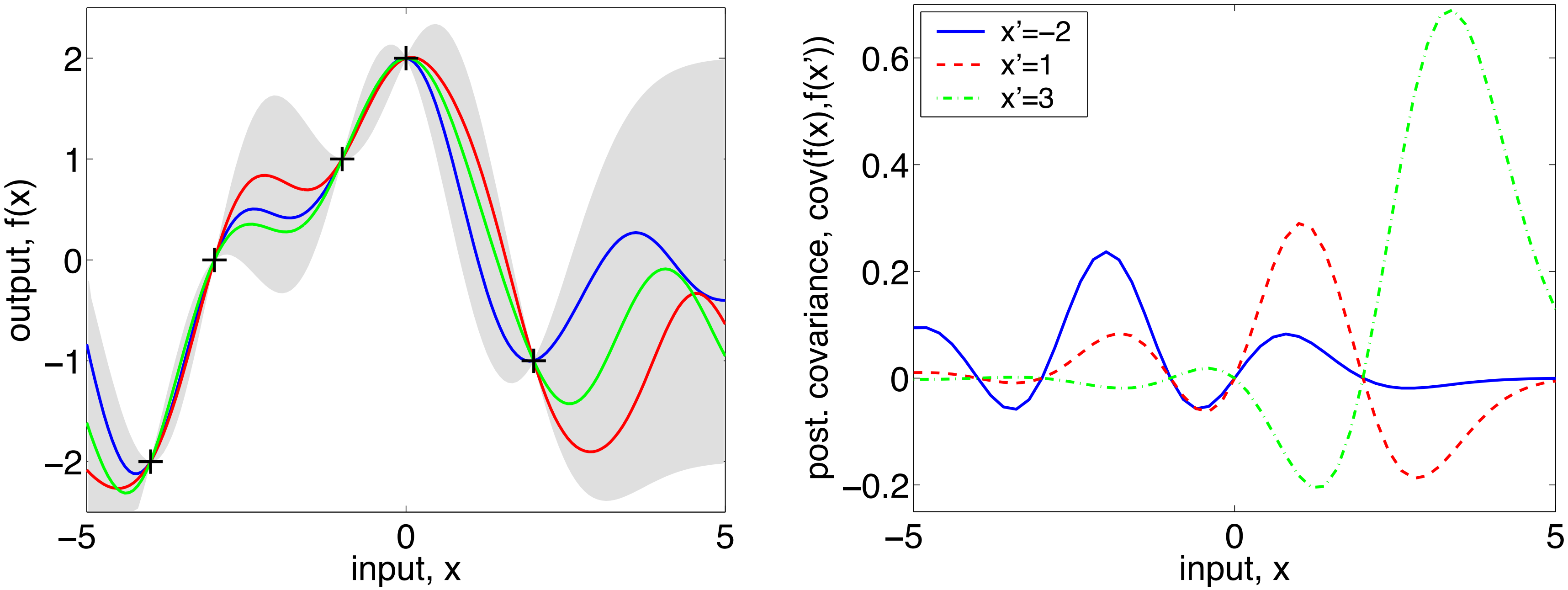
Modelling data with a GP prior

- Prior distribution (no data)
  - $P(\boldsymbol{y}_* | m, k, I) \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$
- Posterior distribution is conditioned on data
  - $P(\boldsymbol{y}_* | \boldsymbol{y}, m, k, I) \sim \mathcal{N}(f_*, \mathrm{var}[f_*])$ where
    - $f_* = \boldsymbol{k}_*^{\mathrm{T}} (\mathbf{K} + \sigma_{\mathrm{n}}^2 \mathbf{I})^{-1} \boldsymbol{y}$
    - $\mathrm{var}[f_*] = k(t_*, t_*) - \boldsymbol{k}_*^{\mathrm{T}} (\mathbf{K} + \sigma_{\mathrm{n}}^2 \mathbf{I})^{-1} \boldsymbol{k}_*$
    - $\boldsymbol{k}_* = (k(t_1, t_*), \ldots, k(t_{\mathrm{N}}, t_*))^{\mathrm{T}}$.
- If $m(t_*) \neq 0$, perform regression on residuals
- If observations are noise, add white noise variance to diagonal of **K**
- Parameters of mean and covariance functions and white noise variance are hyper-parameters of the GP



(b), posterior

Posterior mean and covariance



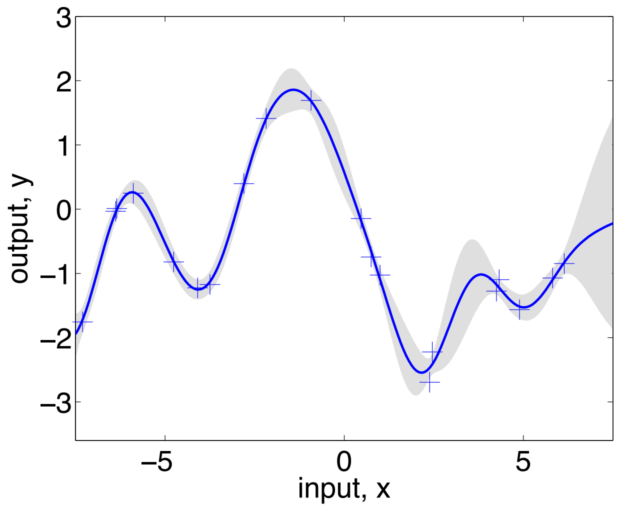
(a), posterior (b), posterior covariance

- P($\boldsymbol{y}_*$|m,k,I) is also multivariate Gaussian, with ≠ mean and covariance from prior
- Even if prior GP was stationary, posterior is not necessarily
- GPs can be used to model non-stationary data

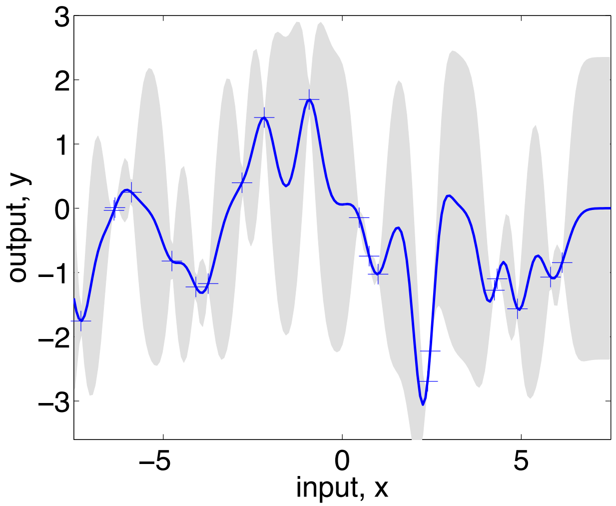
Impact of the hyper-parameters

parameters = values of mean vector and covariance matrix
hyper-parameters = parameters of covariance and mean function

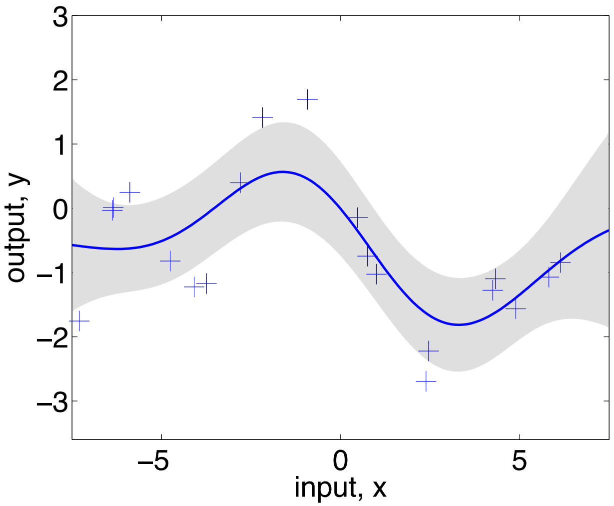
- $k(t,t') = \sigma_d^2 \exp[-(t-t')^2 / 2l^2]$
- Top panel:
  - data generated with $(l,\sigma_d,\sigma_n) = (1,1,1)$
  - posterior distribution for fixed $l = 1$
- Bottom panels: assumed other values of l.
- Upper panel has larger marginal likelihood $P(\boldsymbol{y}|m,k,D,I)$



(a), $\ell = 1$



(b), $\ell = 0.3$

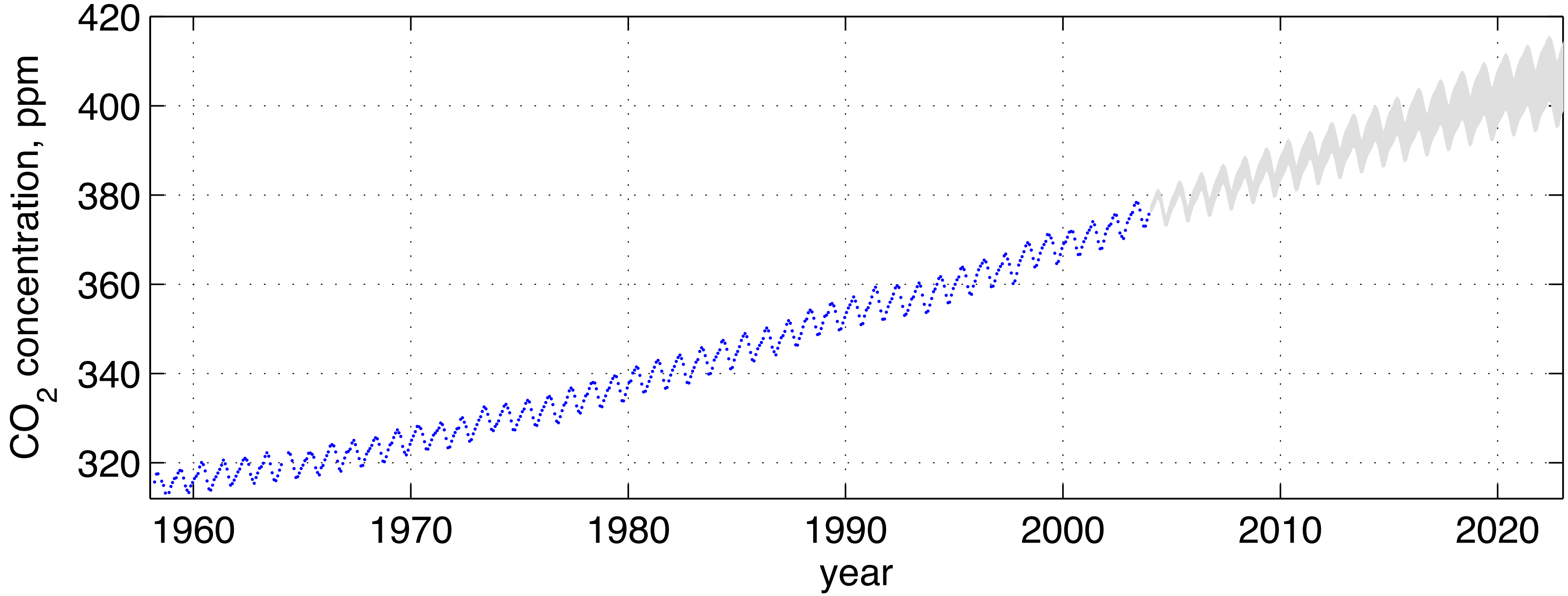


(c), $\ell = 3$

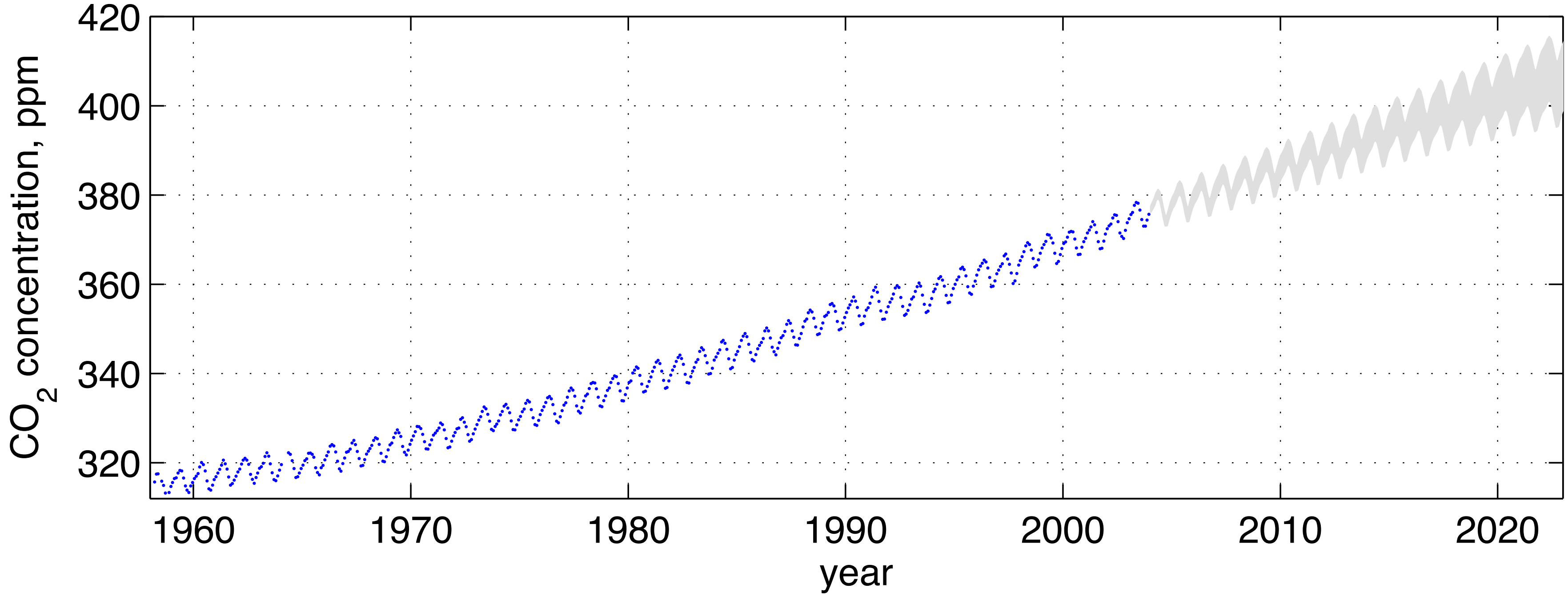
Fitting for the hyper-parameters

- Ideally: Use sampling techniques to
  - marginalise over the hyper-parameters we don’t care about
  - measure the posterior distribution for those which are physically relevant
  - expensive! Matrix inversion for each trial set of hyper-parameters
- Compromise: set some (or all) of them to their maximum likelihood values
  - find those using standard optimisation methods

Example: Mauna Kea CO_2



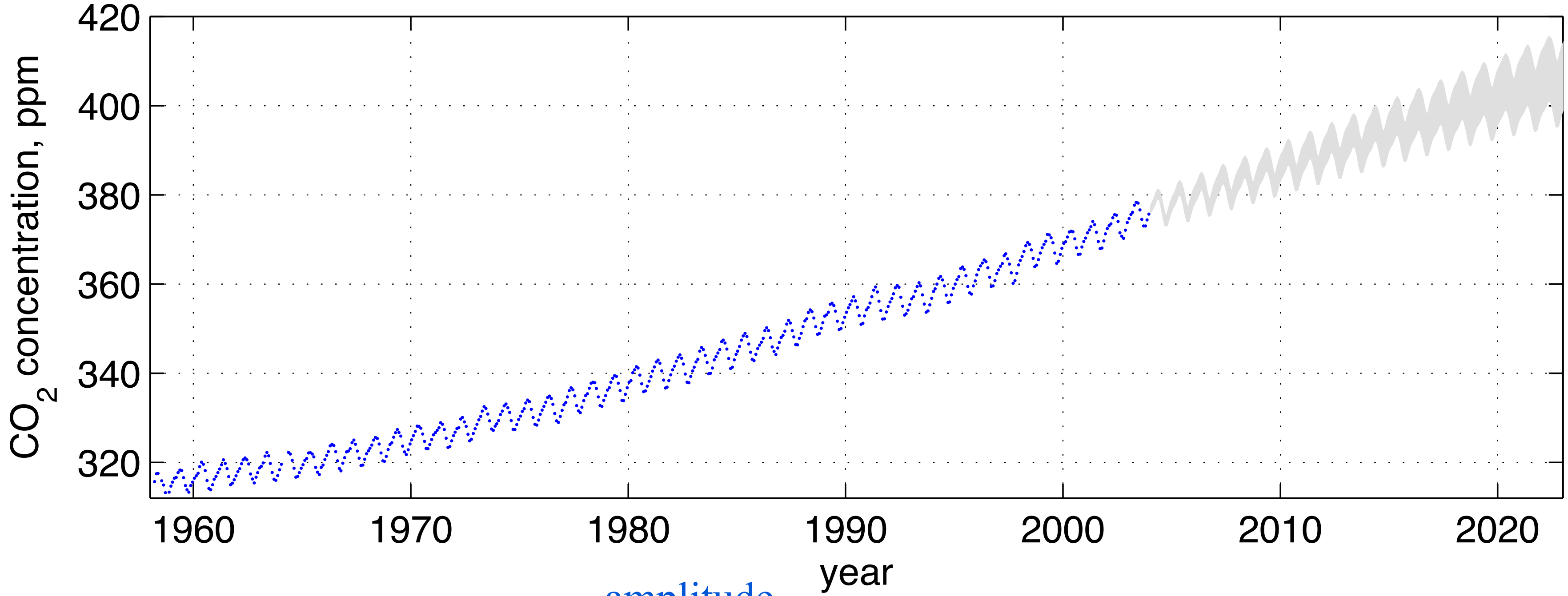
Example: Mauna Kea CO_2



long term trend

$$k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x - x')^2}{2\theta_2^2}\right).$$

Example: Mauna Kea CO_2



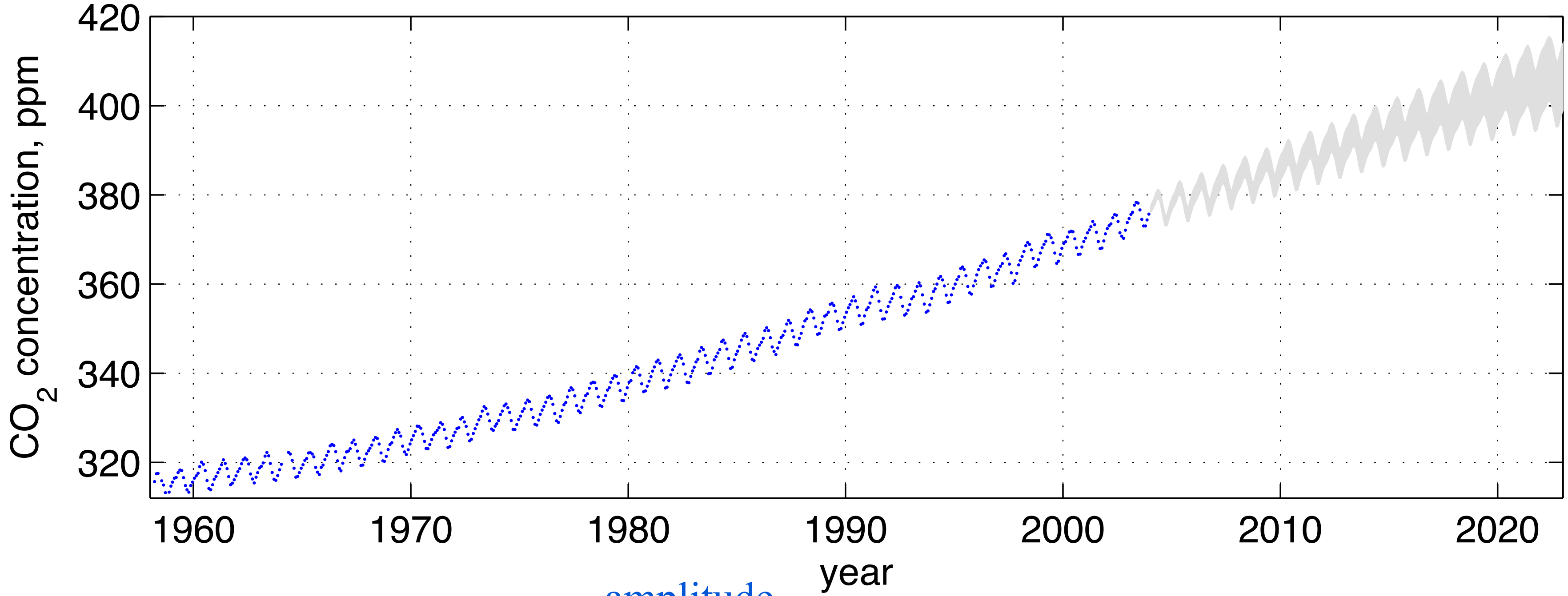
amplitude

long term trend

$$k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x - x')^2}{2\theta_2^2}\right).$$

time scale

Example: Mauna Kea CO_2

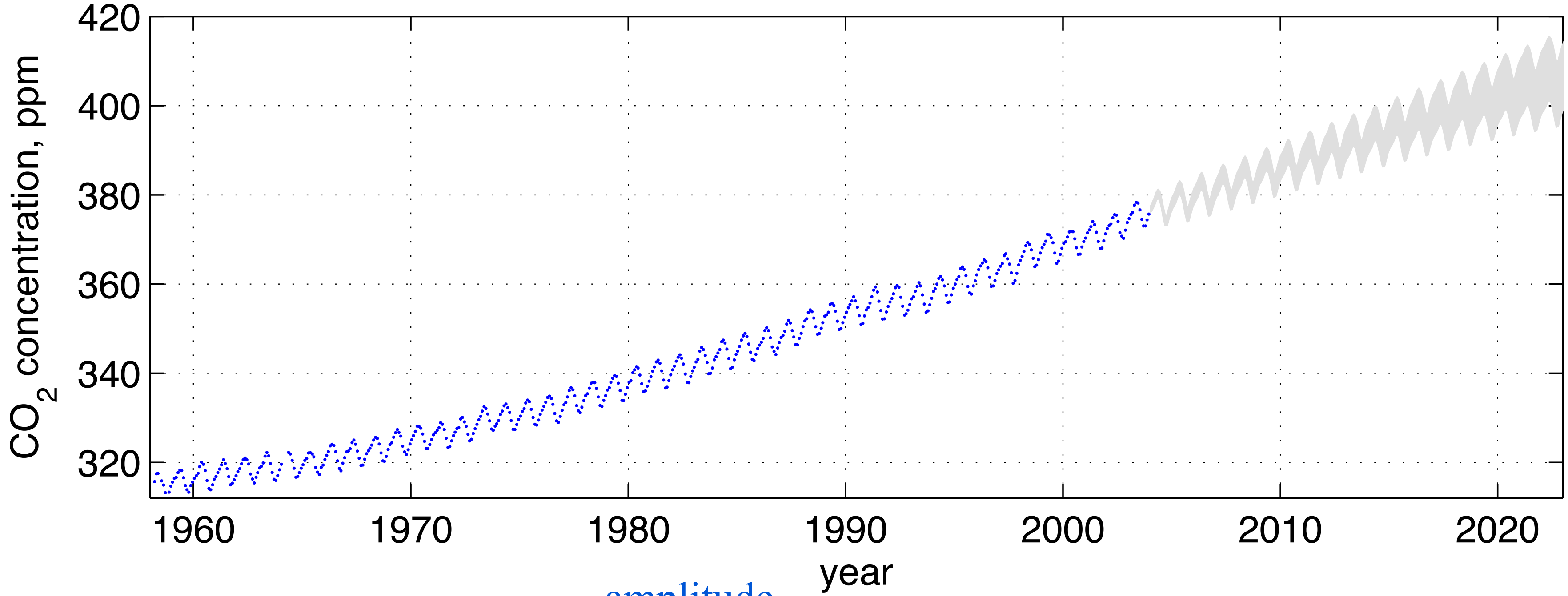


amplitude

long term trend $$k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right).$$ time scale

quasi-periodic oscillation $$k_2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$$

Example: Mauna Kea CO_2



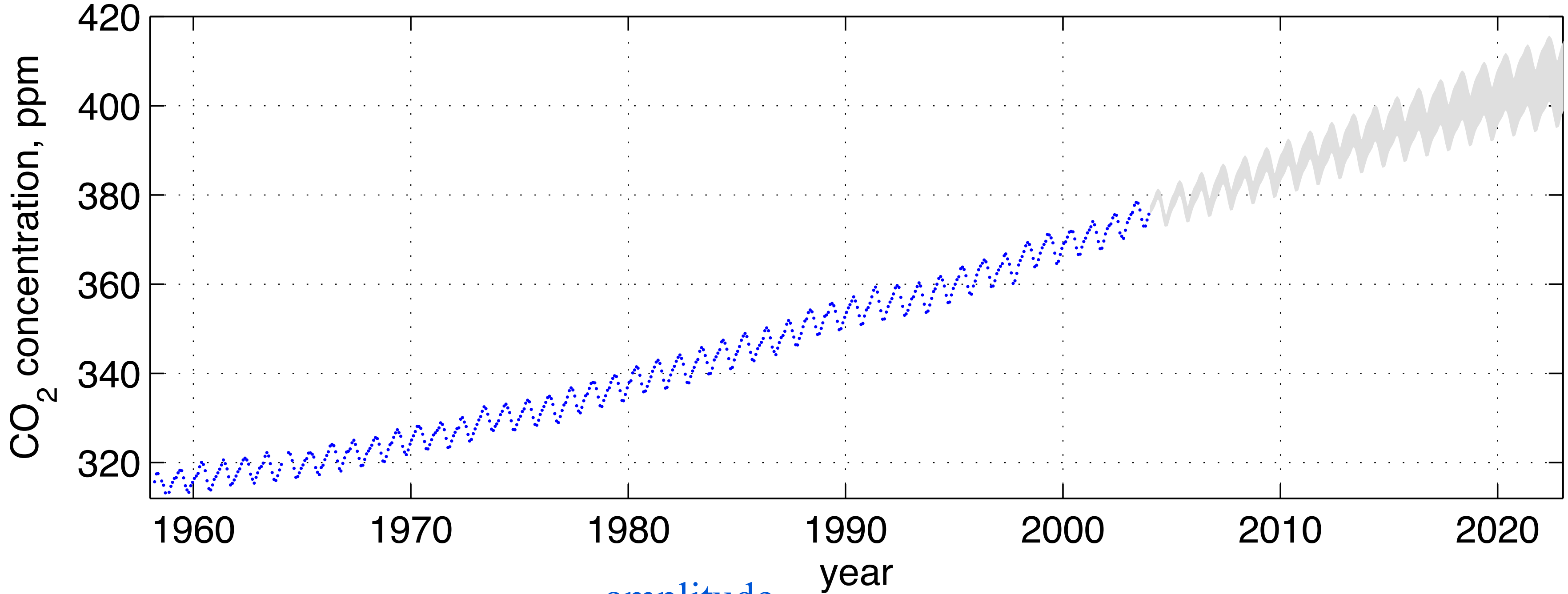
amplitude

long term trend $$k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right),$$ time scale

decay

quasi-periodic oscillation $$k_2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$$ periodic term (fixed period)

Example: Mauna Kea CO_2



amplitude

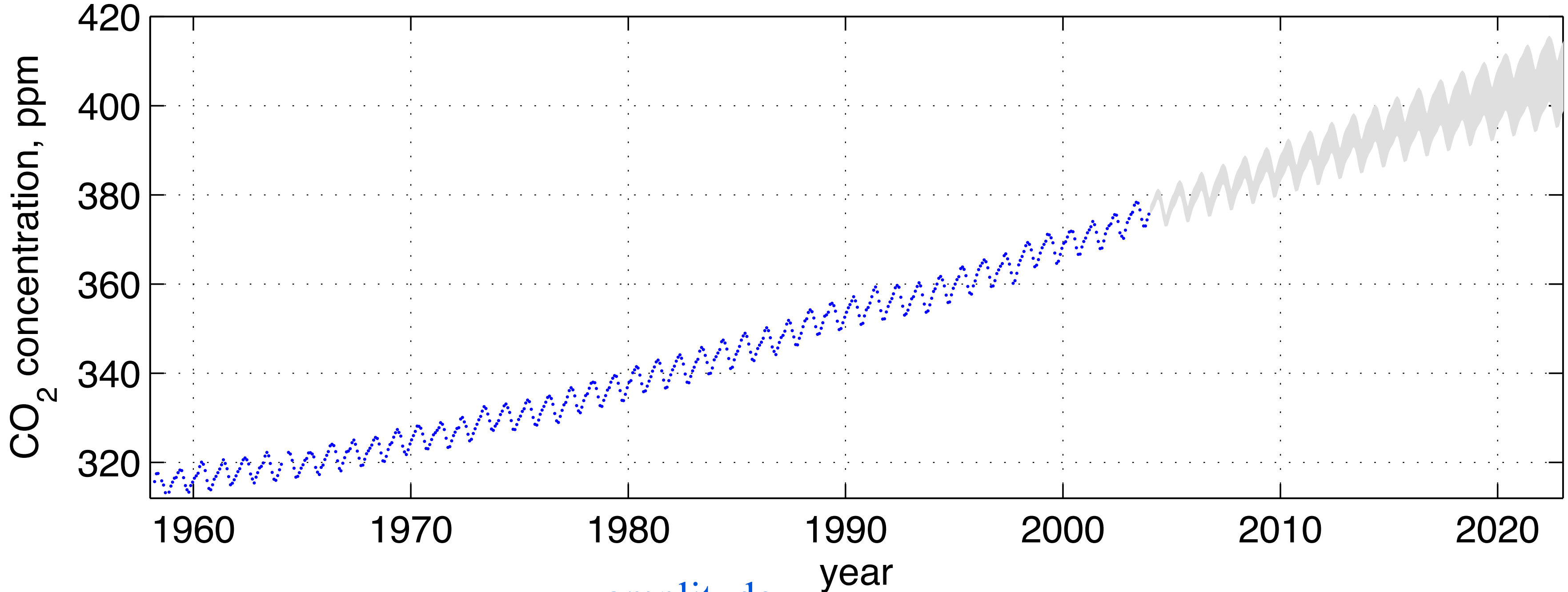
long term trend $$k_1(x,x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right).$$ time scale

decay

quasi-periodic oscillation $$k_2(x,x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$$ periodic term (fixed period)

medium-term irregularities $$k_3(x,x') = \theta_6^2\left(1 + \frac{(x-x')^2}{2\theta_8\theta_7^2}\right)^{-\theta_8},$$

Example: Mauna Kea CO_2



amplitude

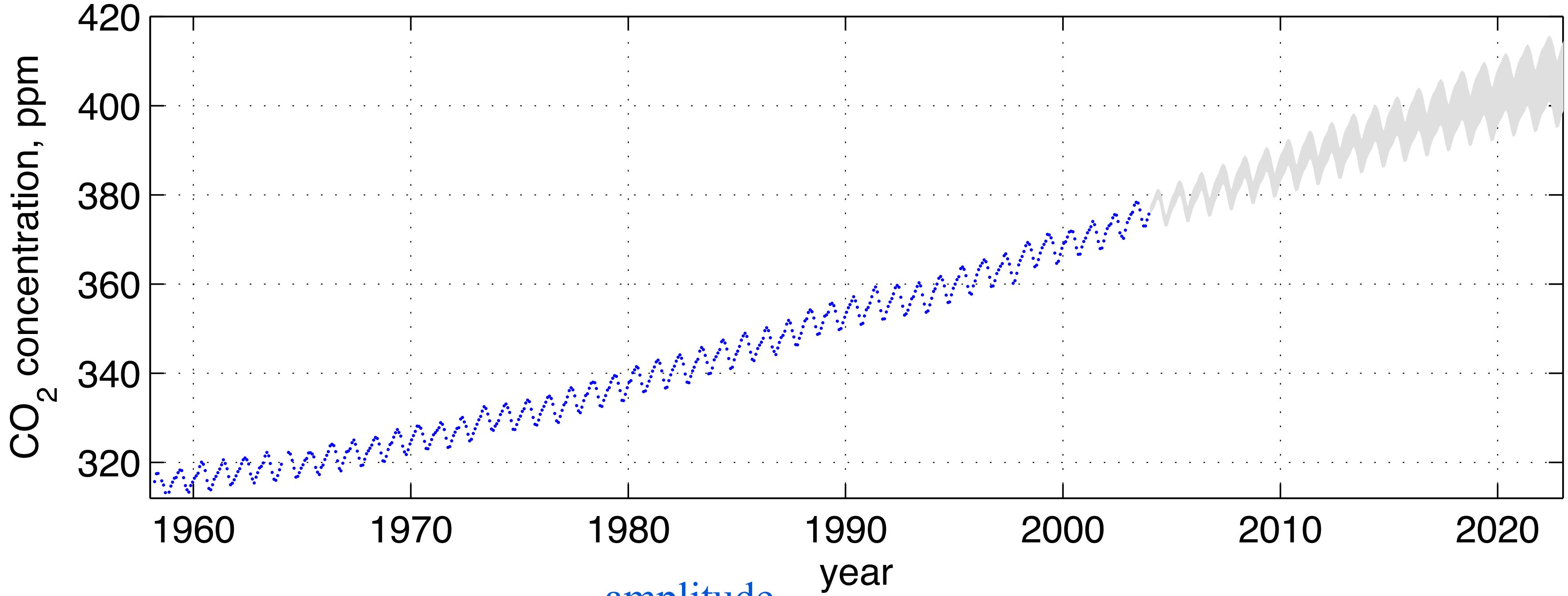
long term trend $k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right)$, time scale

decay

quasi-periodic oscillation $k_2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right)$, periodic term (fixed period)

medium-term irregularities $k_3(x, x') = \theta_6^2 \left(1 + \frac{(x-x')^2}{2\theta_8\theta_7^2}\right)^{-\theta_8}$, multiple time-scales

Example: Mauna Kea CO_2



amplitude

time scale

decay

periodic term (fixed period)

multiple time-scales

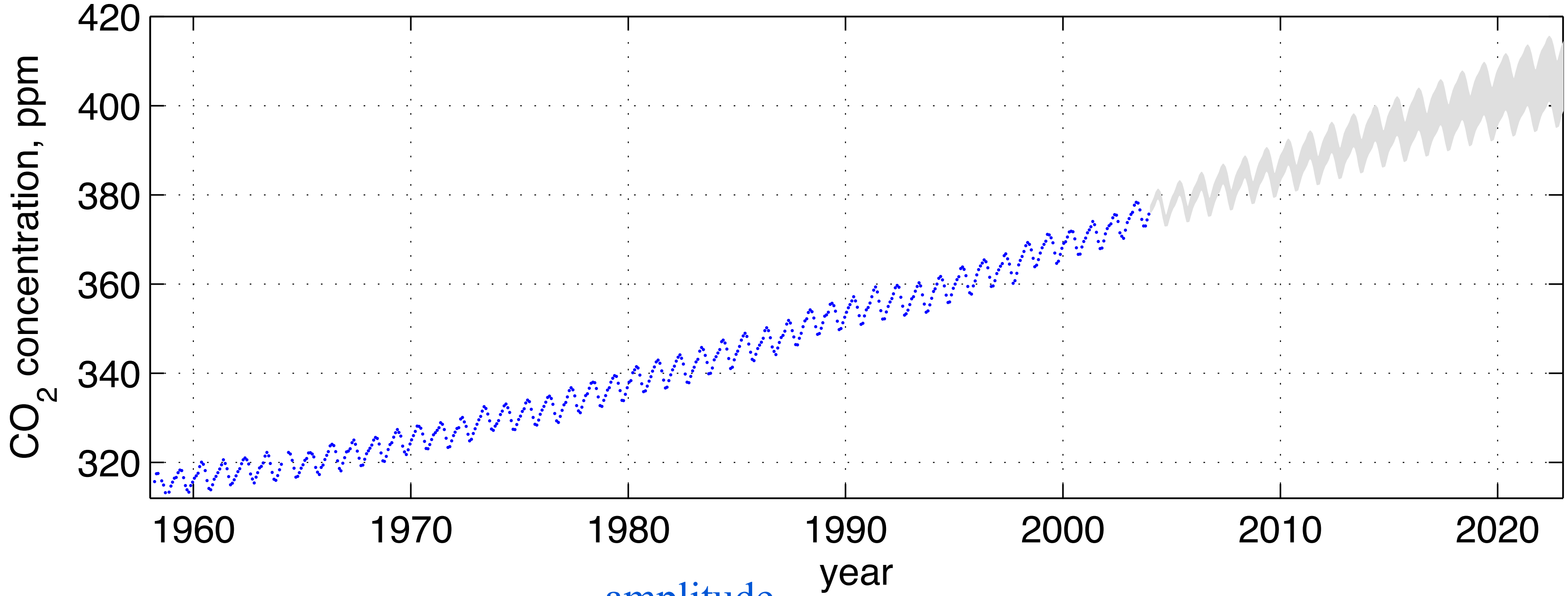
long term trend
$$k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right).$$

quasi-periodic oscillation
$$k_2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$$

medium-term irregularities
$$k_3(x, x') = \theta_6^2 \left(1 + \frac{(x-x')^2}{2\theta_8\theta_7^2}\right)^{-\theta_8},$$

noise
$$k_4(x_p, x_q) = \theta_9^2 \exp\left(-\frac{(x_p-x_q)^2}{2\theta_{10}^2}\right) + \theta_{11}^2 \delta_{pq},$$

Example: Mauna Kea CO_2



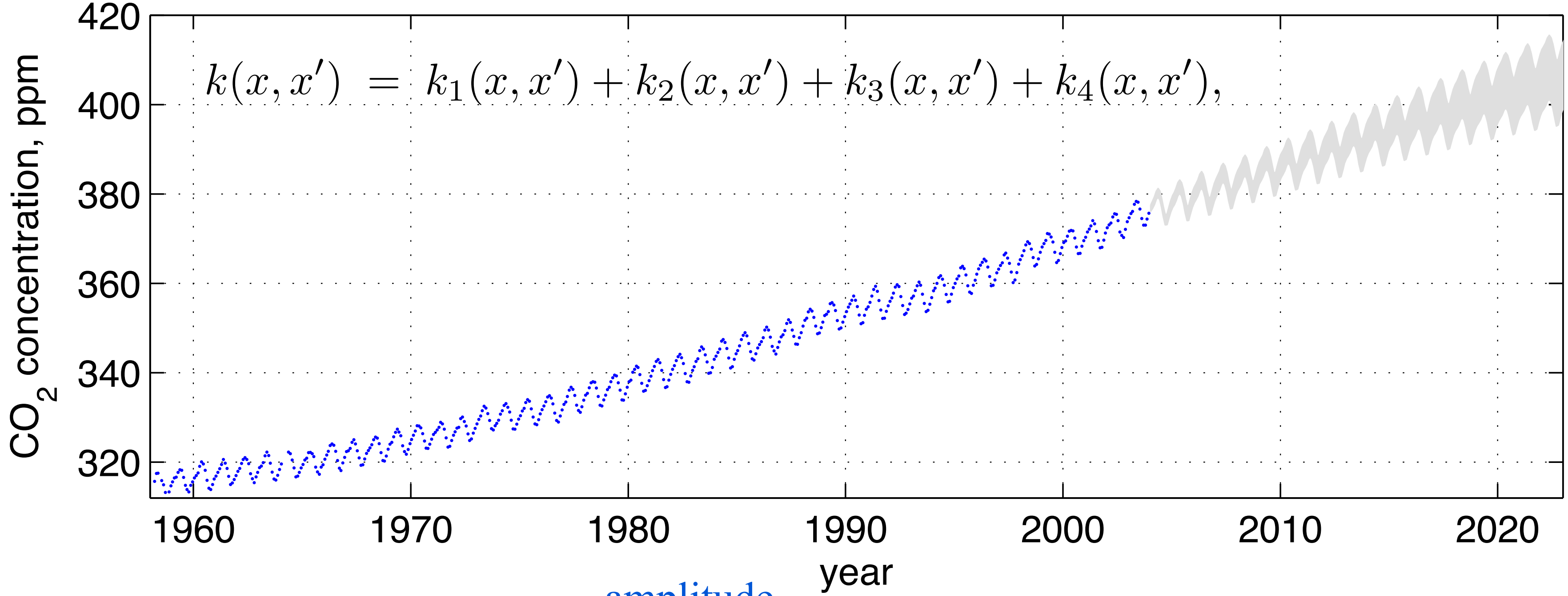
long term trend: $k_1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right),$ (amplitude: θ_1^2; time scale: θ_2^2)

quasi-periodic oscillation: $k_2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$ (decay; periodic term (fixed period))

medium-term irregularities: $k_3(x, x') = \theta_6^2 \left(1 + \frac{(x-x')^2}{2\theta_8\theta_7^2}\right)^{-\theta_8},$ (multiple time-scales: θ_8)

noise: $k_4(x_p, x_q) = \theta_9^2 \exp\left(-\frac{(x_p-x_q)^2}{2\theta_{10}^2}\right) + \theta_{11}^2\delta_{pq},$ (correlated; white)

Example: Mauna Kea CO_2



long term trend

$$k_1(x,x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right).$$

amplitude; time scale

quasi-periodic oscillation

$$k_2(x,x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right),$$

decay; periodic term (fixed period)

medium-term irregularities

$$k_3(x,x') = \theta_6^2\left(1 + \frac{(x-x')^2}{2\theta_8\theta_7^2}\right)^{-\theta_8},$$

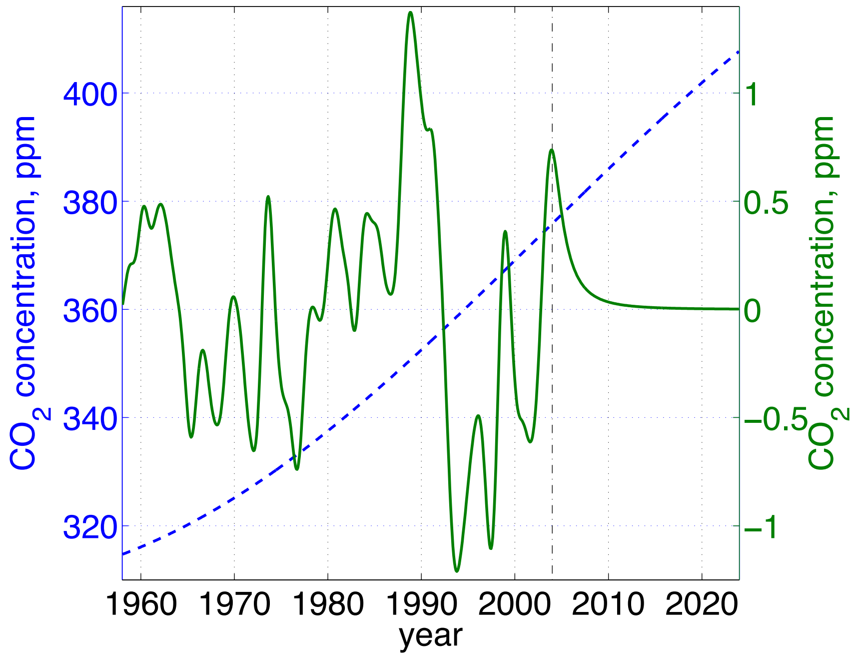
multiple time-scales

noise

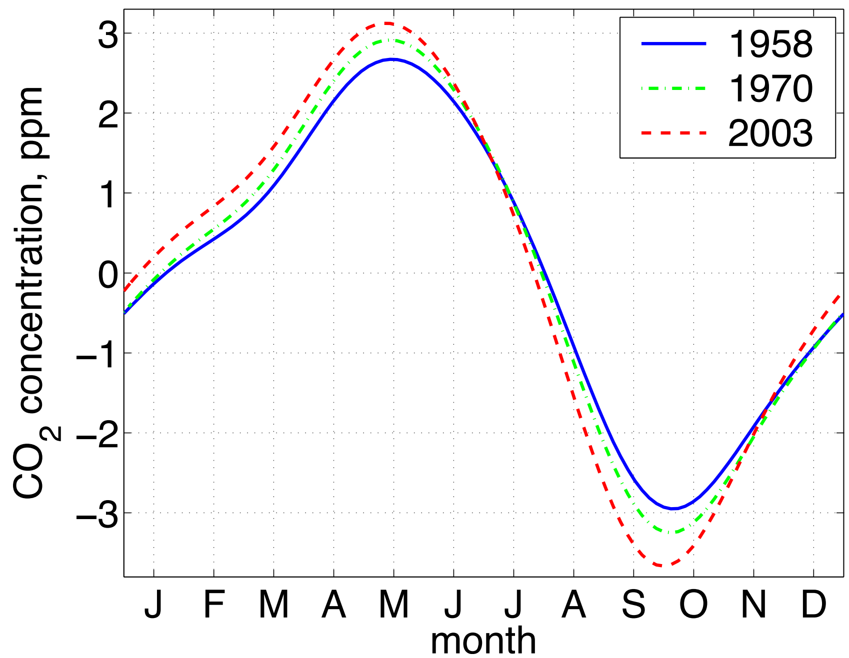
$$k_4(x_p,x_q) = \theta_9^2 \exp\left(-\frac{(x_p-x_q)^2}{2\theta_{10}^2}\right) + \theta_{11}^2\delta_{pq},$$

correlated; white

Example: Mauna Kea CO_2



(a)



(b)

AR(p) processes as GPs

- Matern class of covariance functions

$$k_{\mathrm{Matern}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\Big(\frac{\sqrt{2\nu}r}{\ell}\Big)^{\nu} K_{\nu}\Big(\frac{\sqrt{2\nu}r}{\ell}\Big),$$

- K_c = modified Bessel function, $\Gamma(v)$ = error function

- setting $v + 1/2 = p$ for integer p gives class of AR(p) processes
  - can check if by computing spectral density, i.e. FT of covariance function

- Almost ANYTHING is a special case of GP

Some cool things you can do

Some cool things you can do

- Multivariate time-series
    - both input (what I have so far called “time”) and output (“y”) can be multi-dimensional
    - an example of this will be shown tomorrow afternoon

Some cool things you can do

- Multivariate time-series
  - both input (what I have so far called “time”) and output (“y”) can be multi-dimensional
  - an example of this will be shown tomorrow afternoon
- Change-point detection
  - detect sudden change of variance, or time-scale of variation, or any other discontinuity
  - see paper by Garnet et al
  - an example tomorrow also

Some cool things you can do

- Multivariate time-series
  - both input (what I have so far called “time”) and output (“y”) can be multi-dimensional
  - an example of this will be shown tomorrow afternoon
- Change-point detection
  - detect sudden change of variance, or time-scale of variation, or any other discontinuity
  - see paper by Garnet et al
  - an example tomorrow also
- Decision making: when should I take my next observation

Some cool things you can do

- Multivariate time-series
  - both input (what I have so far called "time") and output ("y") can be multi-dimensional
  - an example of this will be shown tomorrow afternoon
- Change-point detection
  - detect sudden change of variance, or time-scale of variation, or any other discontinuity
  - see paper by Garnet et al
  - an example tomorrow also
- Decision making: when should I take my next observation
- Use GPs to model the probability distributions you're trying to estimate
  - known as "Bayesian quadrature"
  - allows you to chose where in the parameter space to take the next sample
  - you can estimate multivariate probability distribution with very few samples

GPs - python packages

- infpy (GP library developped in a systems biology context) by John Reid
    - http://sysbio.mrc-bsu.cam.ac.uk/group/index.php/Gaussian_processes_in_python
    - tested, works ok, but no hyper-parameter marginalisation

- pyXGPR (GP regression and relational GPs) by Marion Newmann
    - http://www-kd.iai.uni-bonn.de/index.php?page=software_details&id=19
    - not tested

- GPAstro: GP regression library currently under development in Oxford
    - basic GP regression in python
    - fully Bayesian treatment of hyperparameters using python-wrapped C code (will also supply MATLab wrappers)
    - optimized sampling methods

Some other things to try out

State space model

- **state model**: current state = combination of past states + process noise
  - linear combination of past states: linear dynamic system
- **observation model**: current observation = function of current state + observation noise

- **Kalman filter**: current best estimate of system state = weighted average of past estimate and latest observation

- See most time series text books

- Gives rise to powerful class of algorithms for quasi-periodic oscillations (see West 1995 and references to that)

Empirical mode decomposition

- Non stationary, non-harmonic signal
- Wish to attain instantaneous measure of frequency and energy
- Hilbert transform

$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} \, \mathrm{d}t',$$

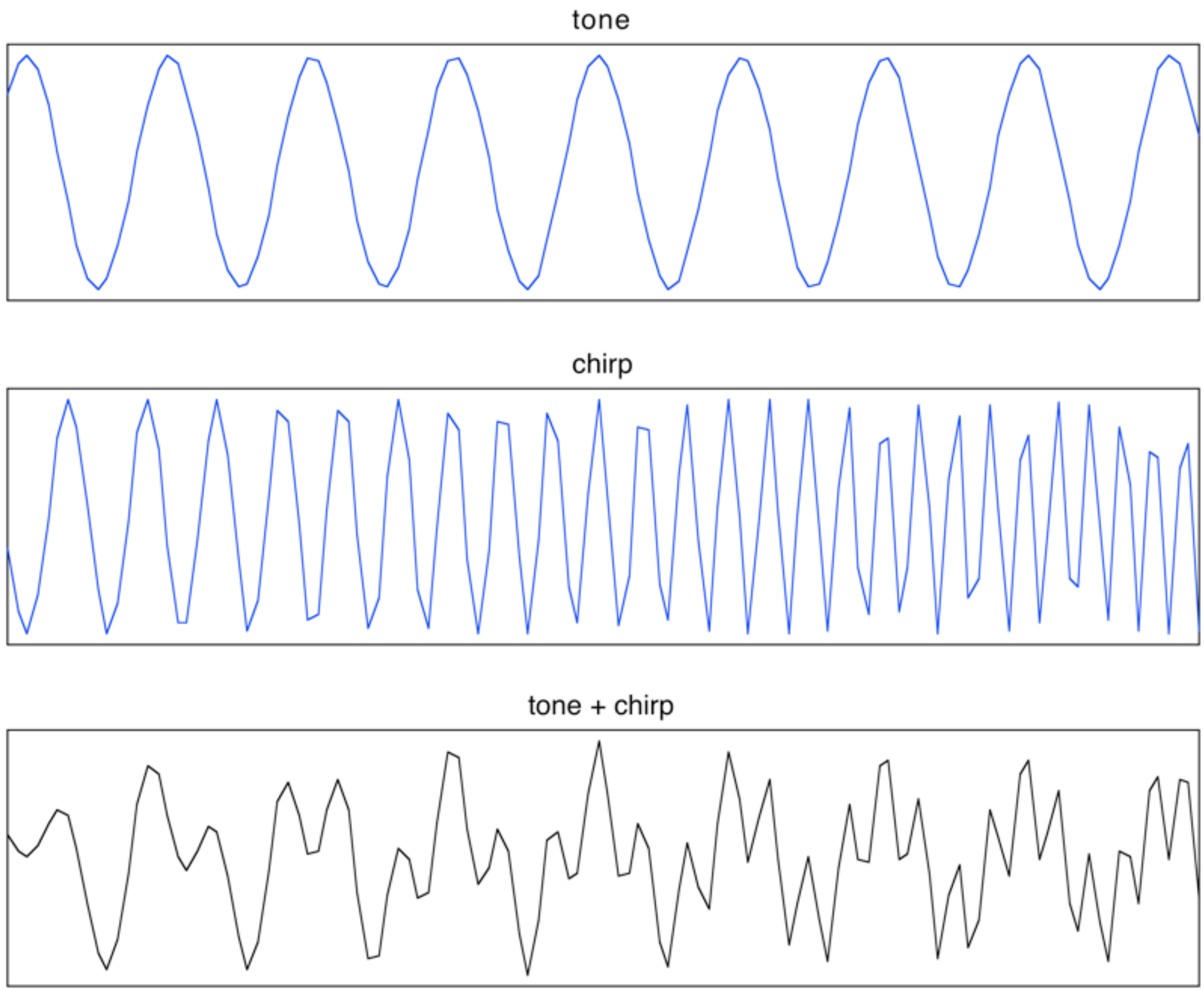
- Construct

$$Z(t) = X(t) + \mathrm{i}Y(t) = a(t)\mathrm{e}^{\mathrm{i}\theta(t)},$$

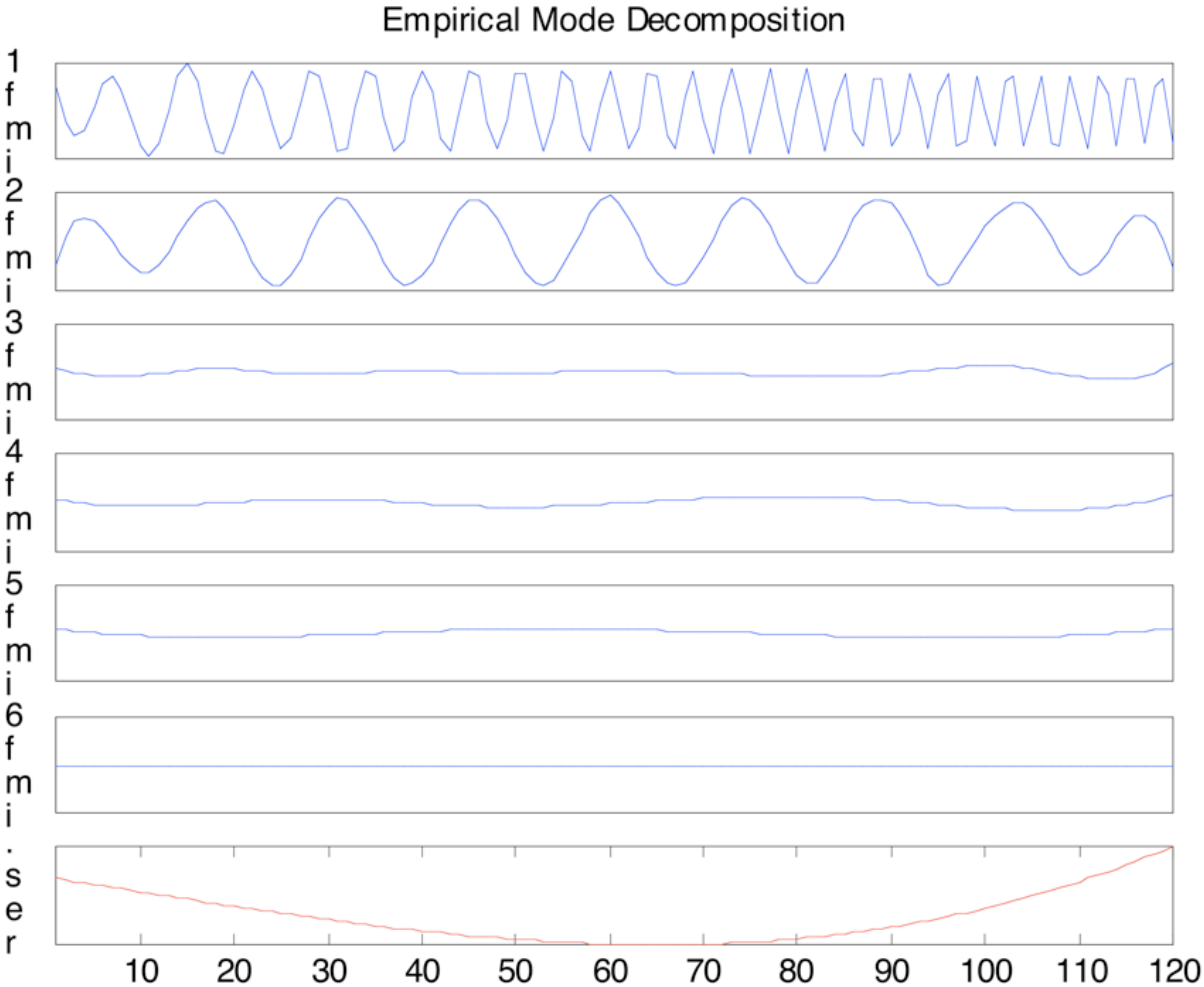
- then

$$\omega = \frac{\mathrm{d}\theta(t)}{\mathrm{d}t}.$$

- But, for this work, the X(t) must satisfy a number of conditions (same number of zero crossings and extrema)
- Empirical mode decomposition is a way of decomposing any signal into a linear combination of "intrinsic modes" which satisfy these conditions



slides from P. Flandrin



slides from P. Flandrin

EMD - further reading & software

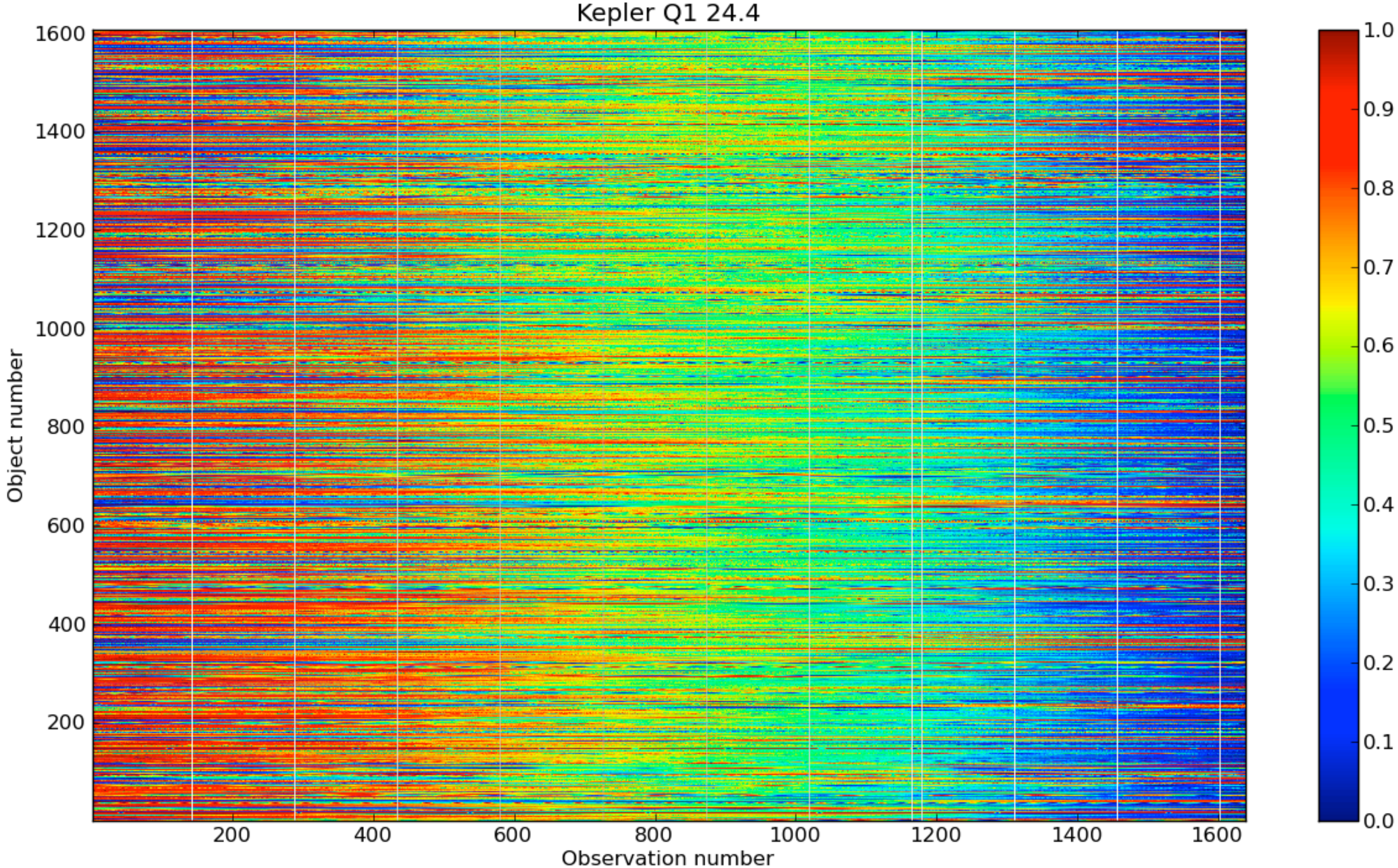
- Huang (1998)
- www.clear.rice.edu
- Patrick Flandrin website http://perso.ens-lyon.fr/patrick.flandrin/emd.html

- No python implementation that I know of yet
- Someone should do it!

- MATLab/C implementation by Patrick Flandrin
  - http://perso.ens-lyon.fr/patrick.flandrin/emd.html
  - tested, works ok

Systematics in ensembles of time series

Kepler Quarter 1 data (1 CCD)



The general idea

- Systematics = trends common to many light curves
- Model each light curve as linear combination of 1 or more systematic trends + intrinsic component (in this case, the “noise” is the intrinsic component)
- How to chose the basis? Must set some constraint
  - Require it to be orthogonal (convenient!): PCA, or extensions thereof
  - Equate it with ancillary observations: external parameter decorrelation
  - Equate it with a subset of your observation sequences (e.g. light curves): TDA
  - Some combination of the above
    - PCA of external parameter vectors

Example: the Kepler pipeline (~)

- start from:
  - the data matrix: flux versus observation number and star number
  - external information: pointing, detector temperature, background level vs time ("system_info")

- Perform PCA on the external parameters
- Linear decomposition of individual light curves onto principal components + "intrinsic" component

- Try it?

- For a different approach, see my talk tomorrow