Time series analysis

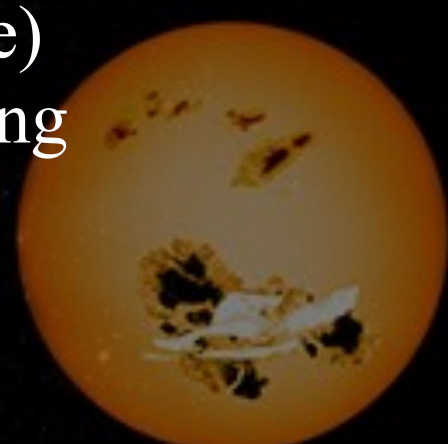
Maria Süveges

Department of Astronomy
University of Geneva
Switzerland

Time-dependent phenomena in astronomy

Deterministic change (+ noise):

- **periodic variations**
  - binarity, planetary systems
  - pulsations
  - rotation (spots)
- **non-periodic variations**
  - eruptive phenomena (flares, cataclysmic variability, supernovae)
  - microlensing



Stochastic process (+ noise):

- **jets, accretion**
  - quasar, AGN, blazar activity
  - X-ray binaries
  - young stellar objects
  - evolved binary systems
- **stellar surface phenomena**
  - granulation
- **scintillation,…**

…or any combination of these

Time-dependent phenomena in astronomy

Deterministic change (+ noise):

- **periodic variations**
  - Fourier analysis
  - least squares-based and other periodicity detection methods
- **non-periodic variations**
  - wavelet analysis
  - pattern recognition
  - matching pursuit, etc.
- **noise structure — crucial!**
  - distribution
  - dependence/correlation in it
  - can be time series itself

Stochastic process (+ noise):

- random walks, Markov processes
- autoregressive—moving average (ARMA) and derived processes
- continuous-time ARMA and variants
- state-space and hierarchical models

What is a time series?

Series of observations

For a statistician:

$$Y_1, Y_2, \ldots, Y_N$$

most often, they don't even mention time (because regularly sampled)

For an astronomer:

$$(t_1, Y_1),\ (t_2, Y_2),\ \ldots,\ (t_N, Y_N)$$

usually, needs the time (very frequently, regular sampling is impossible)

For both, the essential is that

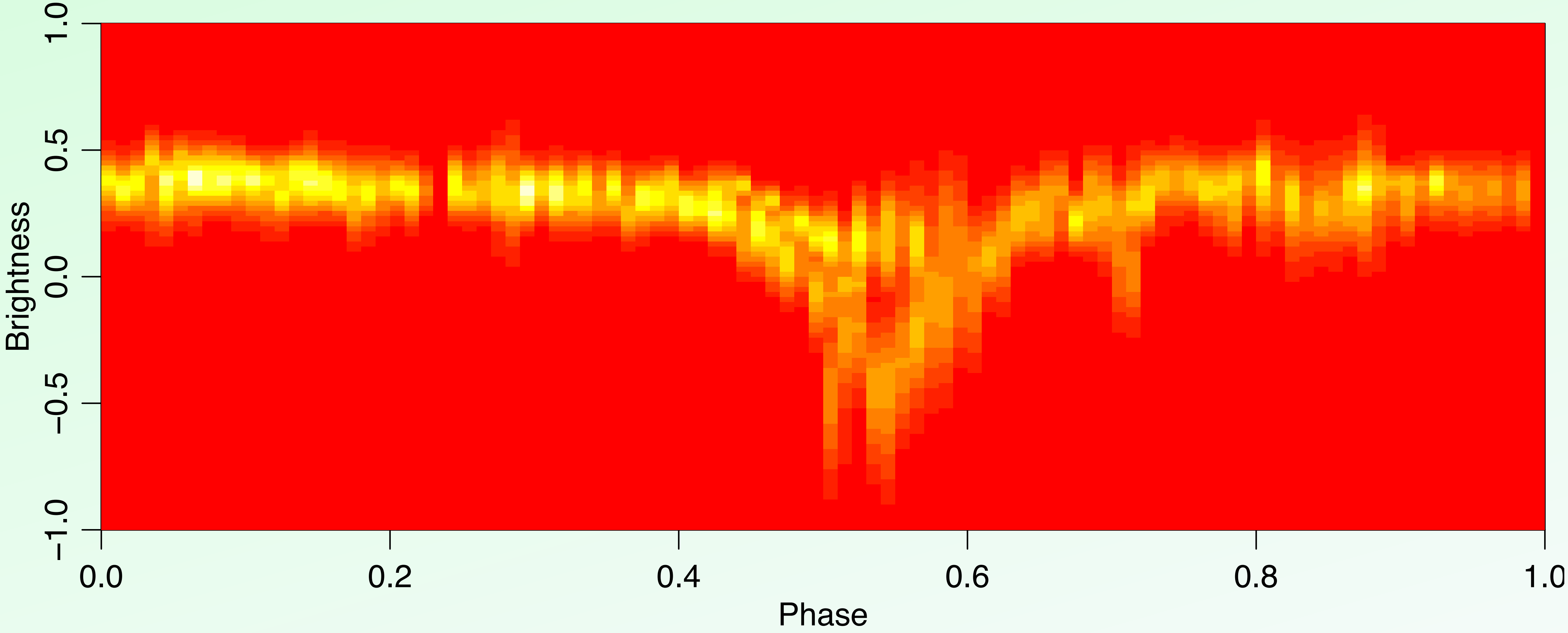
the distribution of Y_k depends on $Y_{k-1}, Y_{k-2}, \ldots$

What is a time series?

Simple case: deterministic light curve shape + independent errors

$$Y \mid T = t_i \overset{\text{ind}}{\sim} \mathcal{N}(g(t_i), \sigma_i^2), \quad \text{where } \sigma_i \text{ is known}$$

$$E(Y \mid T) = g(T)$$

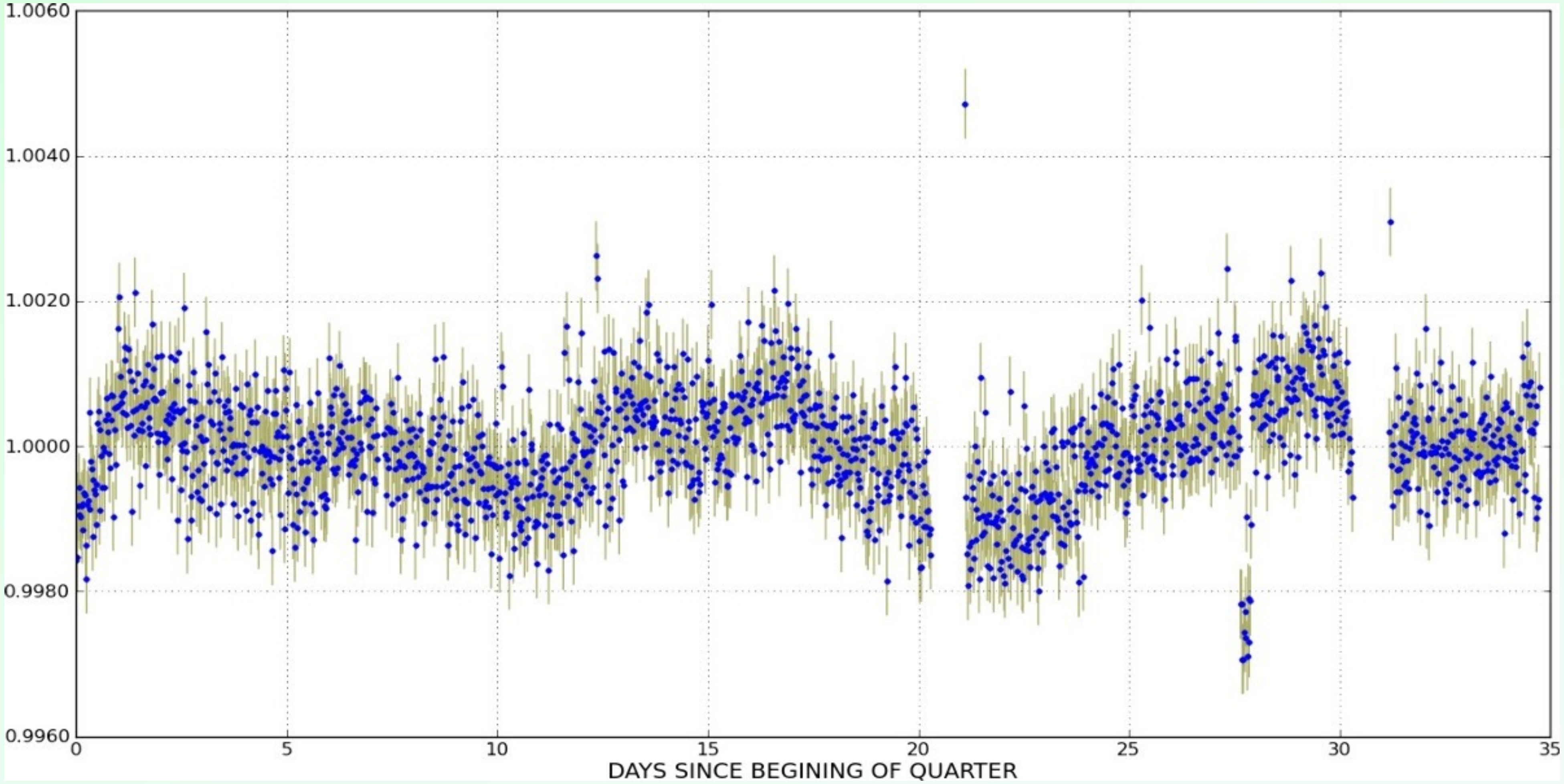


Example: eclipsing binary from LINEAR

What is a time series?

More generally: deterministic light curve shape + dependent errors

$$Y \mid T = t_i \sim \mathcal{F}(g(t_i), \boldsymbol{\theta}) \quad \text{where } \boldsymbol{\theta} \text{ may be partly known or restricted}$$
$$E(Y \mid T) = g(T)$$



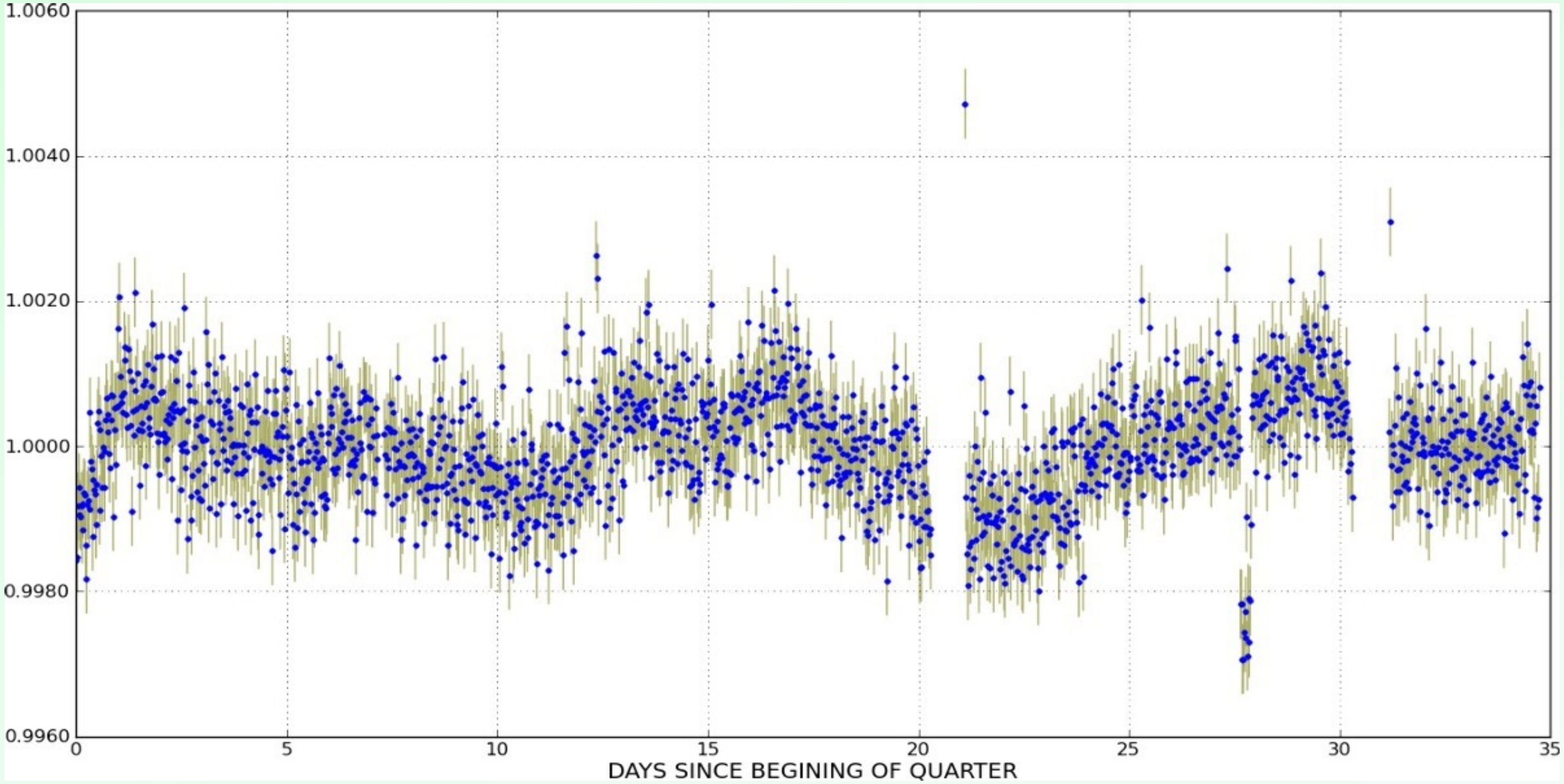
Example: planet candidate from Kepler, where correlation between consecutive residuals cannot be ignored (an improved model: Gaussian processes)

Image: http://keplerlightcurves.blogspot.ch/2012/09/the-eleventh-planethunters-habitable.html

What is a time series?

Stochastic case: the light curve variations are intrinsically stochastic

Very broad range of possibilities! ARMA, ARIMA, fractionally integrated ARIMA, random walks, renewal and branching processes, Markov processes…



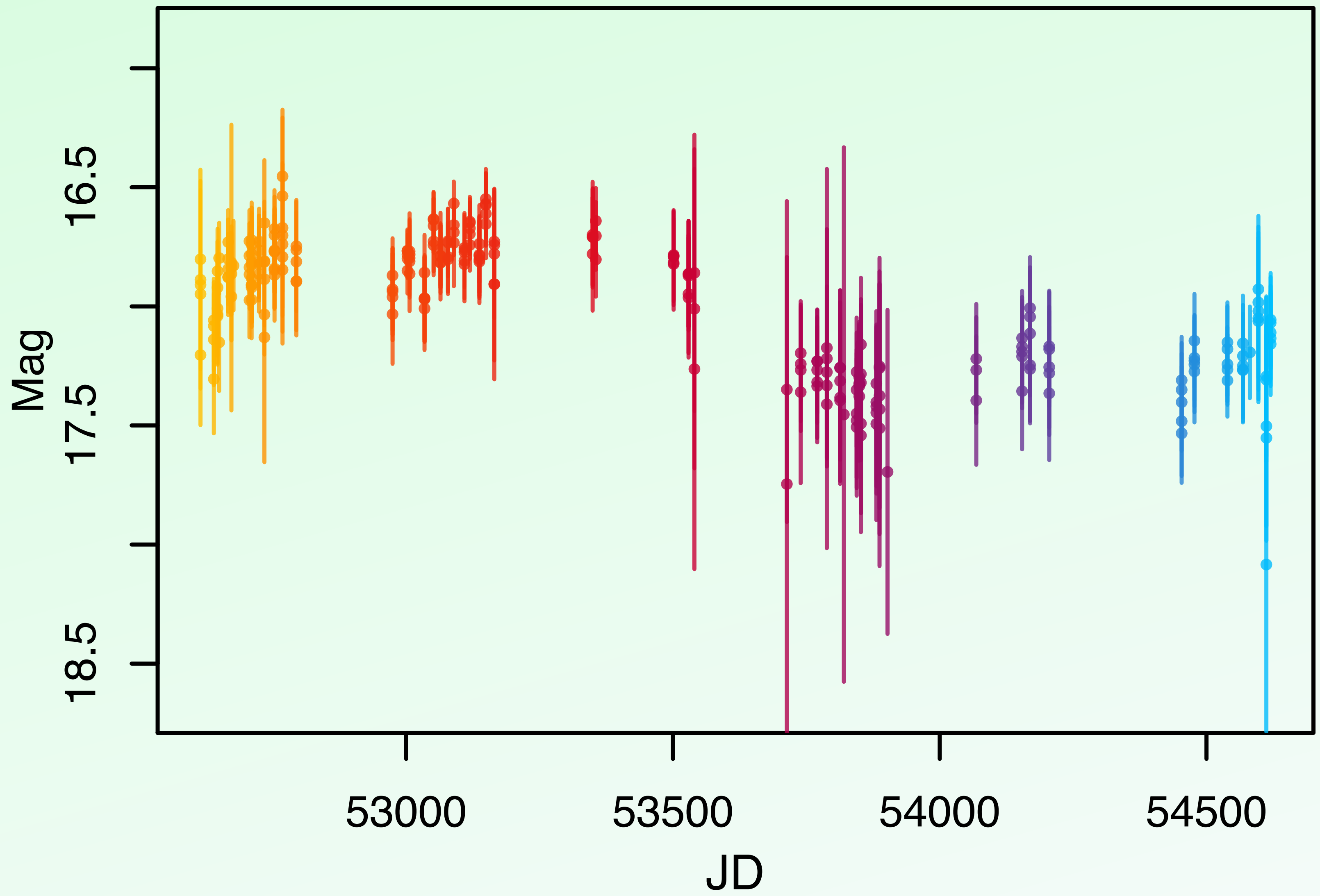
Example: planet candidate from Kepler, where the star signal seems to have irregular variation

Image: http://keplerlightcurves.blogspot.ch/2012/09/the-eleventh-planethunters-habitable.html

What is a time series?

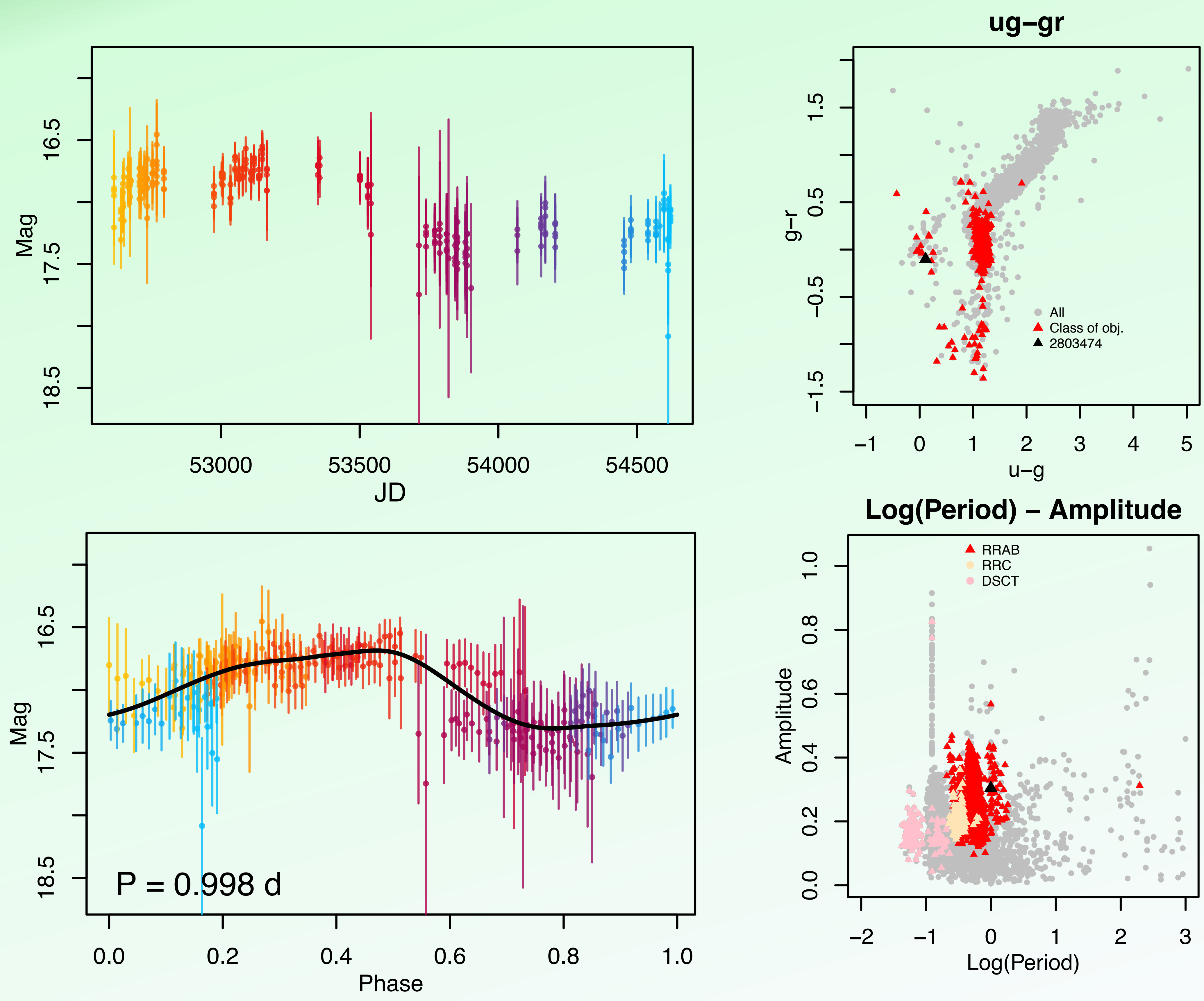
Stochastic case: the light curve variations are intrinsically stochastic
A model applied for quasars: Ornstein-Uhlenbeck process

$$Y_i = \phi(t_i, t_{i-1}) Y_{i-1} + \epsilon_i, \quad \text{where } \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2) \text{ and } \phi(t_i, t_{i-1}) = \mathrm{e}^{-\frac{|t_i - t_{i-1}|}{\tau}}$$



A likely quasar from LINEAR

The challenge



Autocorrelation

Let $Y_1, \ldots, Y_N$ be an equally-spaced time series.

Definition: stationarity of $\{Y_t\}$

Strong: $\{Y_{j_1}, \ldots, Y_{j_K}\}$ and $\{Y_{j_1+h}, \ldots, Y_{j_K+h}\}$ have the same joint distribution for all index sets $j_1, \ldots, j_K$.

Weak: the first and second-order moments (mean, variance, covariance) do not depend on *t*.

Definition: autocovariance function of the (weakly) stationary sequence $\{Y_t\}$

$$\gamma(h) = \mathrm{Cov}(Y_t, Y_{t+h}), \quad h = 1, \ldots, H$$

Definition: autocorrelation function (ACF)

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \mathrm{Corr}(Y_t, Y_{t+h}), \quad h = 1, \ldots, H$$

Spectral analysis: periodogram

Aim: to discover of periodicities in data

$y_1, \ldots, y_N$: equally-spaced time series; $t_0, t_1, \ldots, t_N = 0, 1, \ldots, N$

Discrete Fourier transform:

$$J_Y(f) = \sum_{j=0}^{N-1} y_j \mathrm{e}^{-i\,2\pi f\, t_j}$$

Usual definition of DFT:

only at the Fourier frequencies k/N and $t_j = j$

Classical periodogram:

$$I_Y(f) = \frac{1}{N} \left| \sum_{j=0}^{N-1} y_j \mathrm{e}^{-i\,2\pi f\, t_j} \right|^2$$

The classical periodogram estimates the spectral density of the time series, which is the Fourier transform of the autocovariance function:

$$I_Y(f) = \sum_{h=-N}^{N} \hat{\gamma}(h) \mathrm{e}^{-i\,2\pi f h}$$

(taking the spacing of the times as time unit.)

Basic features of the Fourier transform

Inverse Fourier transform:

$$y_j = \frac{1}{N}\sum_{k=0}^{N-1} J_Y(f_k)\mathrm{e}^{i\,2\pi f_k t_j}, \quad \text{where } f_k = \frac{k}{N} \text{ and } t_j = j$$

Fourier frequencies

Convolution theorem:

the convolution of two functions *a* and *b* is defined as

$$(a * b)(t) = \int_{-\infty}^{\infty} a(u)b(t-u)du$$

(example: kernel smoothing).

Then, if A(f) and B(f) are the Fourier transforms of $a(t)$ and $b(t)$, the Fourier transform of $h = a * b$ is

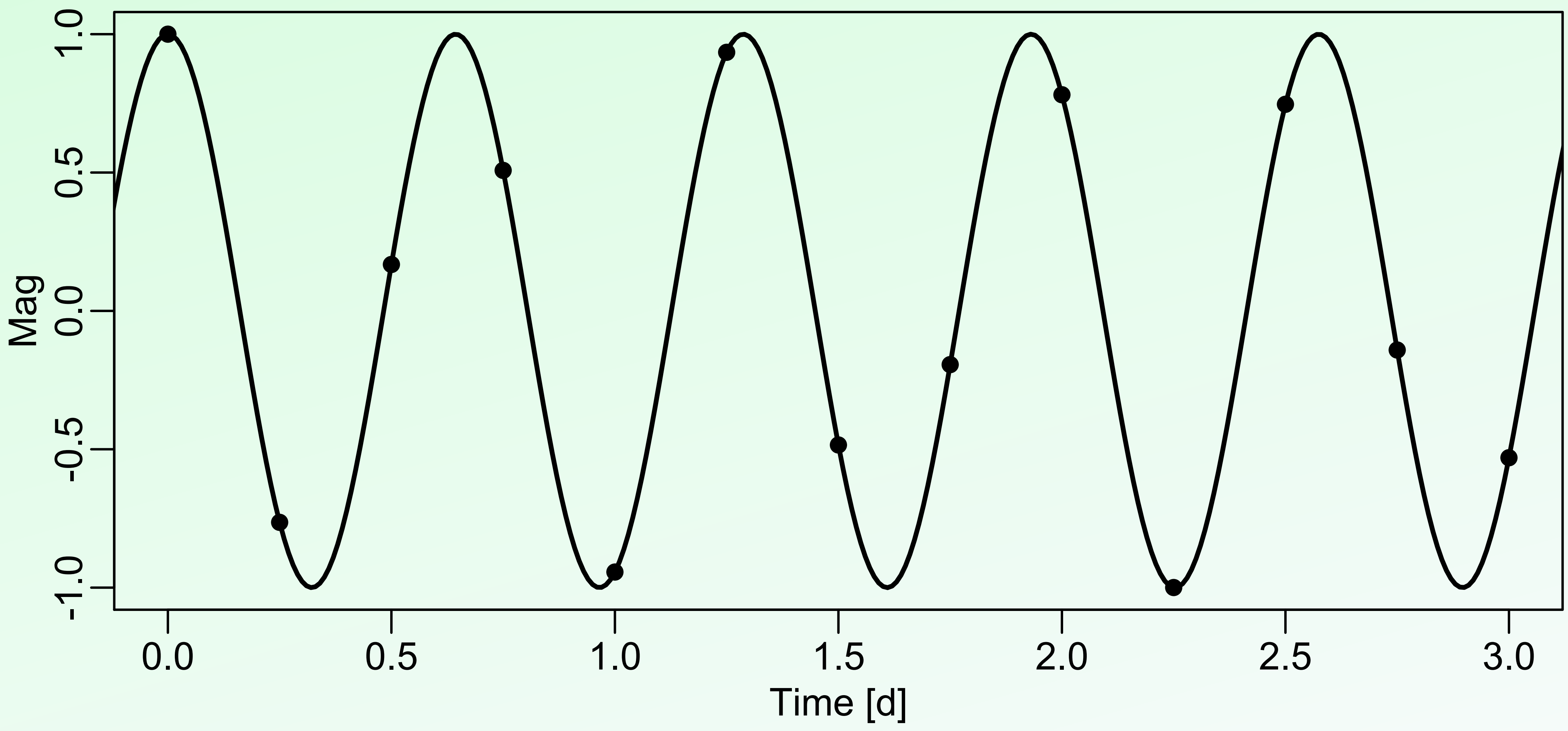
$$H(f) = A(f)B(f)$$

This works the other way round too: a product in time domain becomes convolution in the Fourier domain (example: the discrete time sampling).

Spectral analysis: the Nyquist frequency

For a given sampling rate f_s, the highest detectable frequency in an observed time series is $f_s/2$. This is called the **Nyquist frequency**.

Intuitively: it is necessary to have at least two observations in each cycle of the signal in order to be able to determine its frequency.

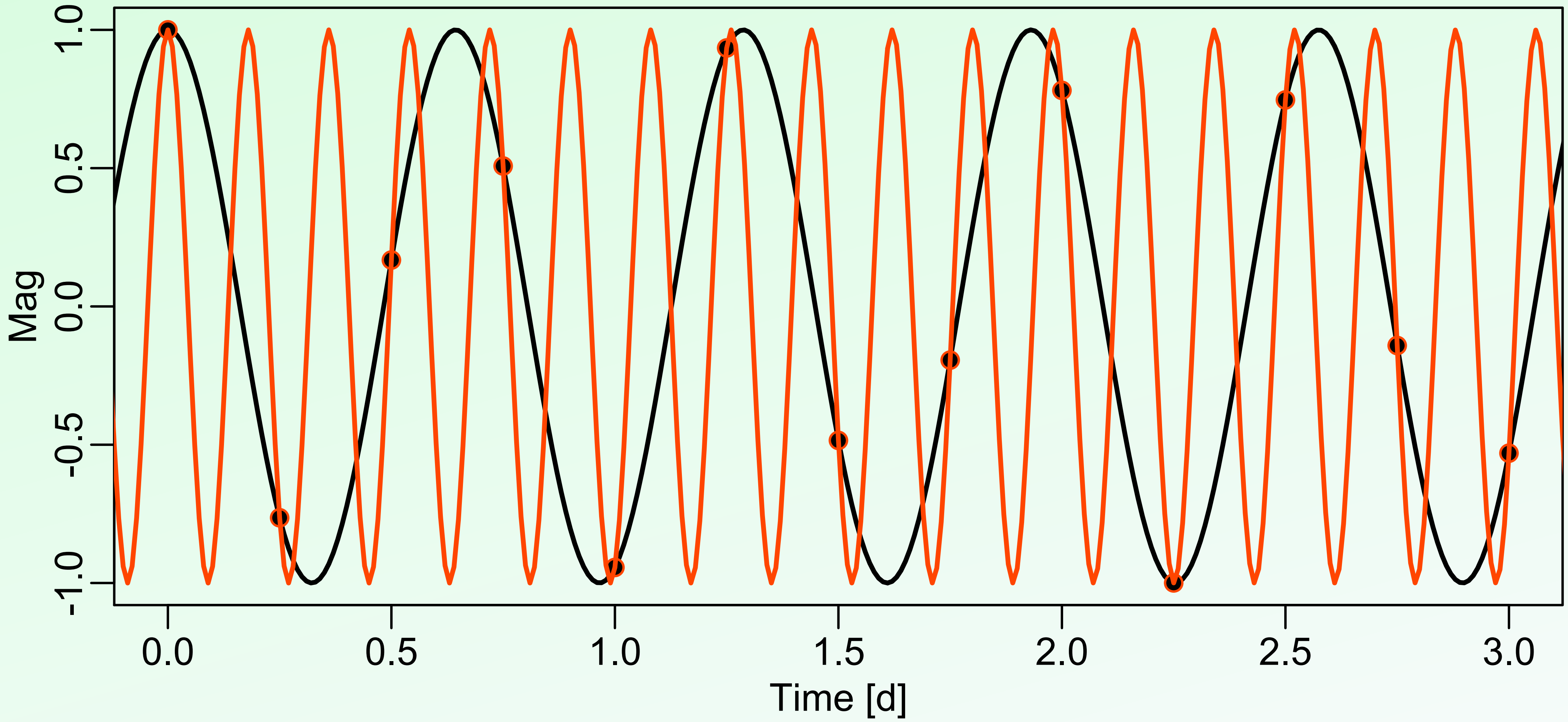


Sampling rate: 4 c/d (Gaia spin rate), Nyquist frequency: 2 c/d, $f_{\mathrm{signal}} = 1.55$ c/d

Spectral analysis: aliasing

For a given sampling rate f_s, the highest detectable frequency in an observed time series is $f_s/2$. This is called the **Nyquist frequency**.

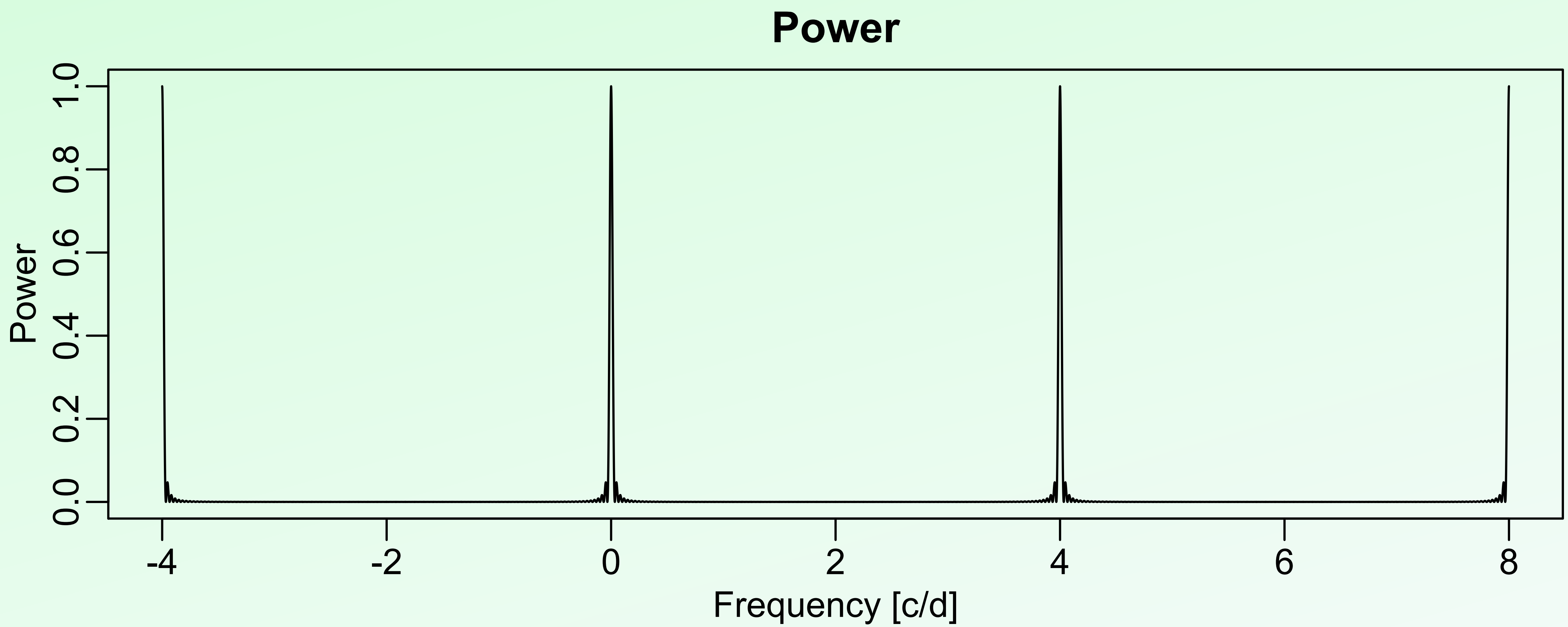
Signals with frequency higher than the Nyquist frequency are **aliased**: they look identical to signals of frequency $|f \pm n f_s|$, where $f \pm n f_s \in [0, f_s/2]$.



Sampling rate: 4 c/d (Gaia spin rate), Nyquist frequency: 2 c/d, $f_{\mathrm{signal}} = 1.55$ c/d
$f_{\mathrm{highfreq}} = 5.55$ c/d

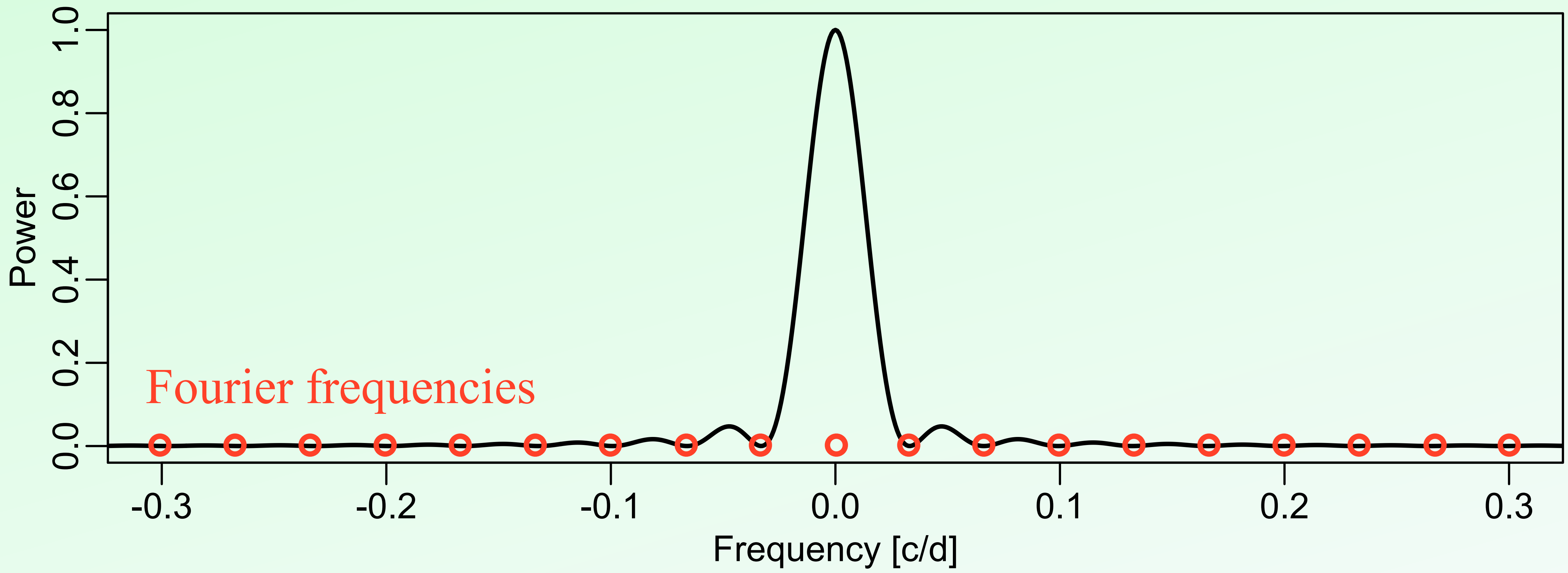
Spectral analysis: the spectral window

Definition: $$I_{\mathrm{s}}(f) = \frac{1}{N}\left|\sum_{j=1}^{N} \mathrm{e}^{-i\,2\pi f\,t_j}\right|^2$$



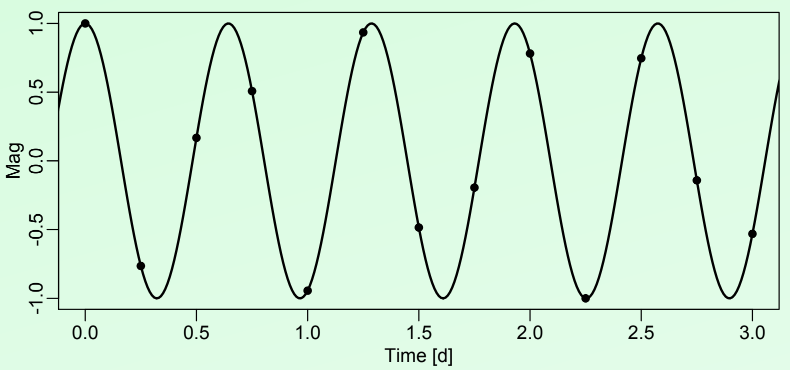
Spectral analysis: the spectral window

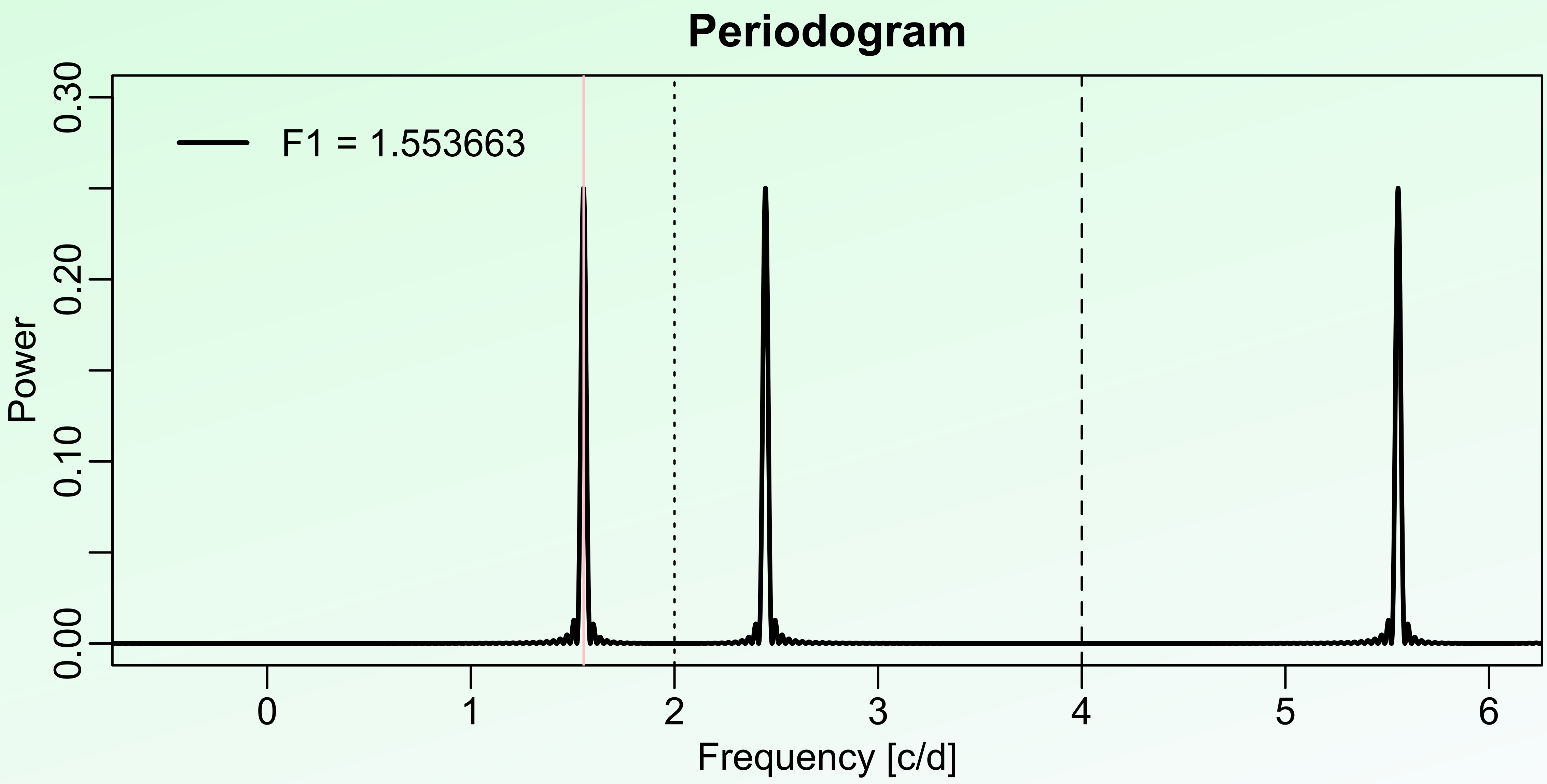
Definition: $$I_\mathrm{s}(f) = \frac{1}{N}\left|\sum_{j=1}^{N} \mathrm{e}^{-i\,2\pi f\, t_j}\right|^2$$



Spectral analysis: periodogram of a pure cosine

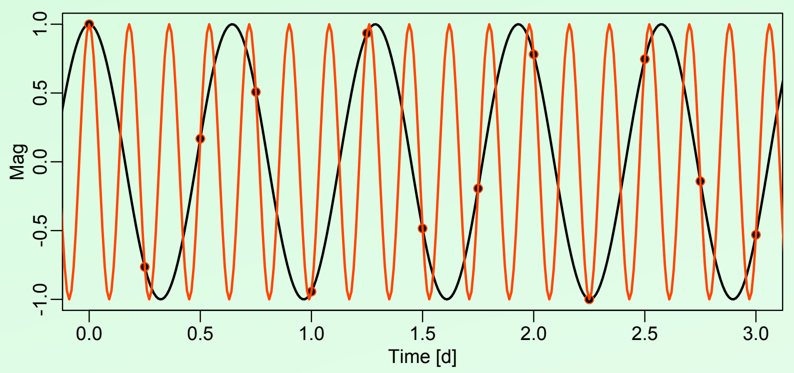
The former cosine at $f_{\mathrm{signal}} = 1.55$ c/d:

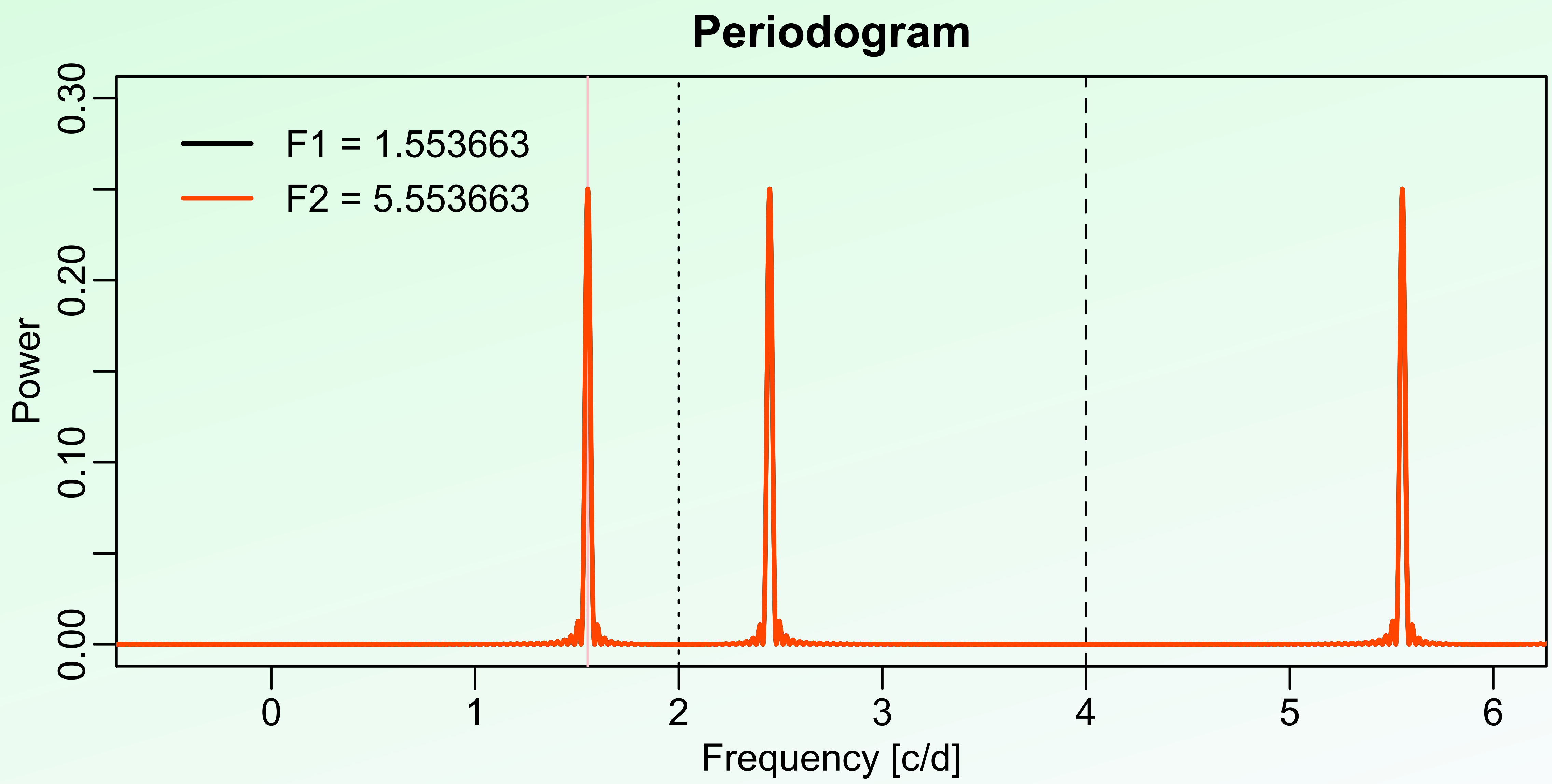




Spectral analysis: periodogram of a pure cosine

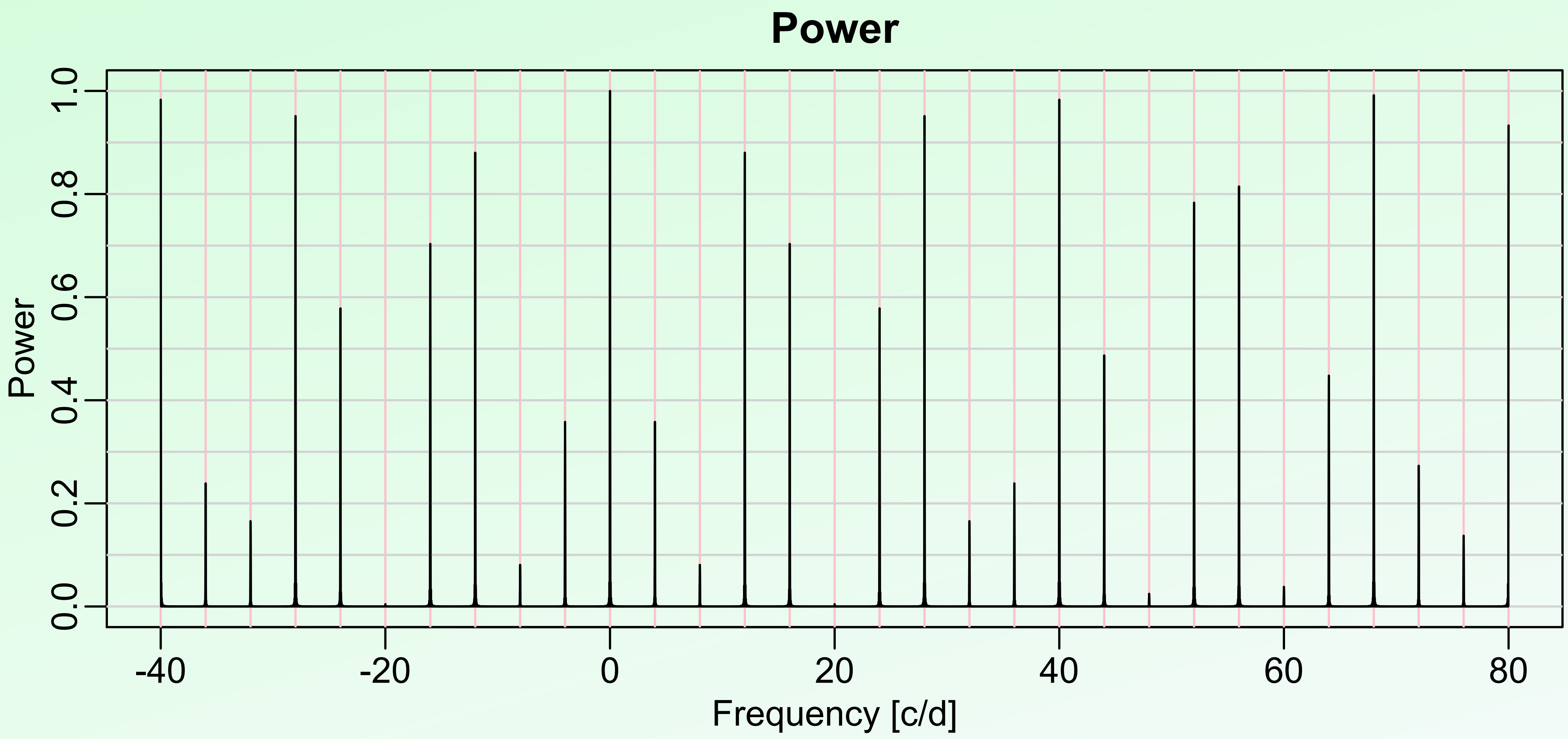
The two cosines at $f_{\text{signal}} = 1.55$ c/d
and at $f_{\text{highfreq}} = 5.55$ c/d:





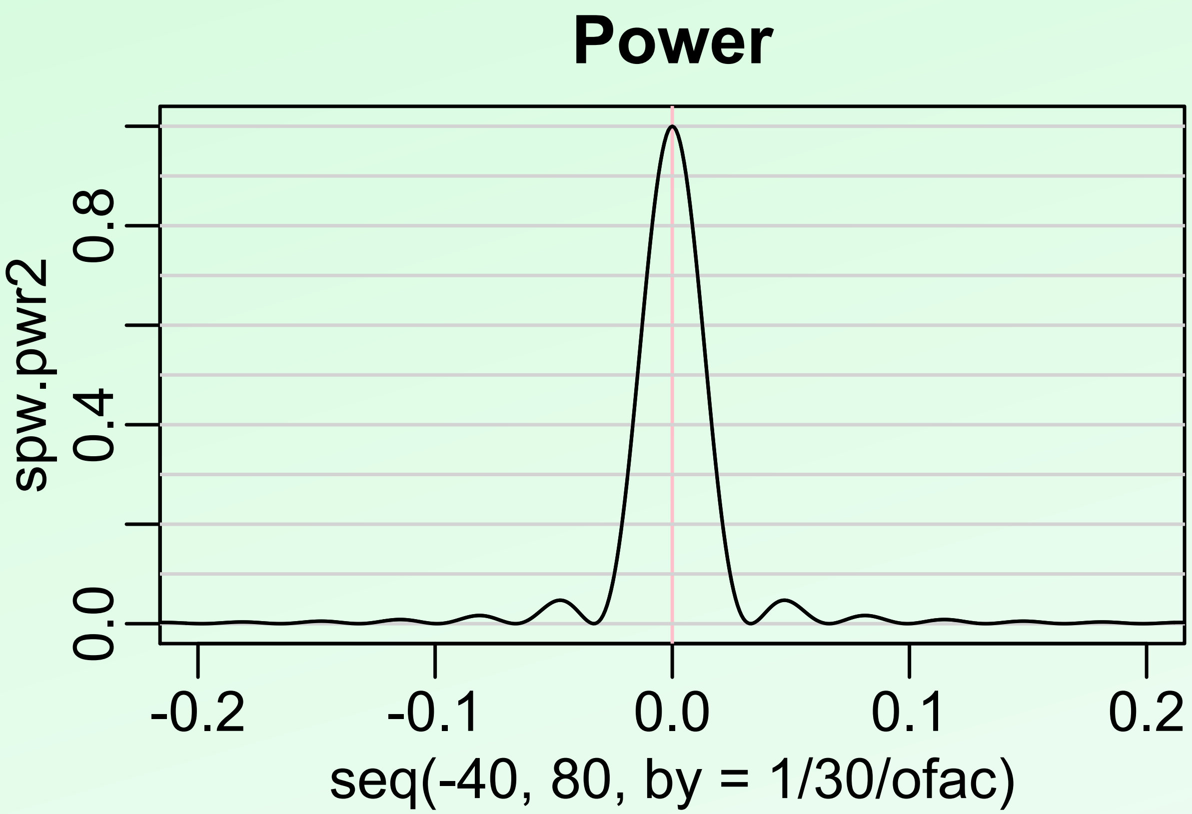
Uneven sampling

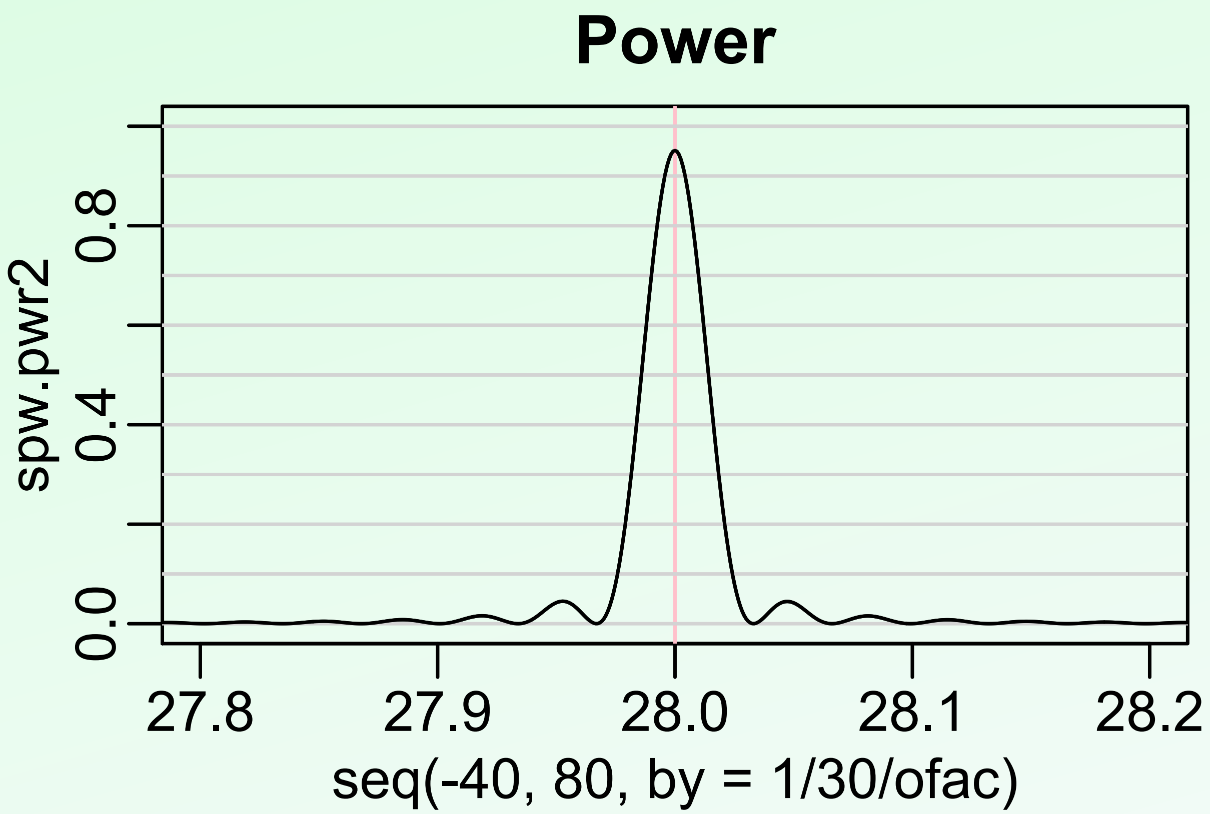
A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (basic Gaia pattern)



Uneven sampling

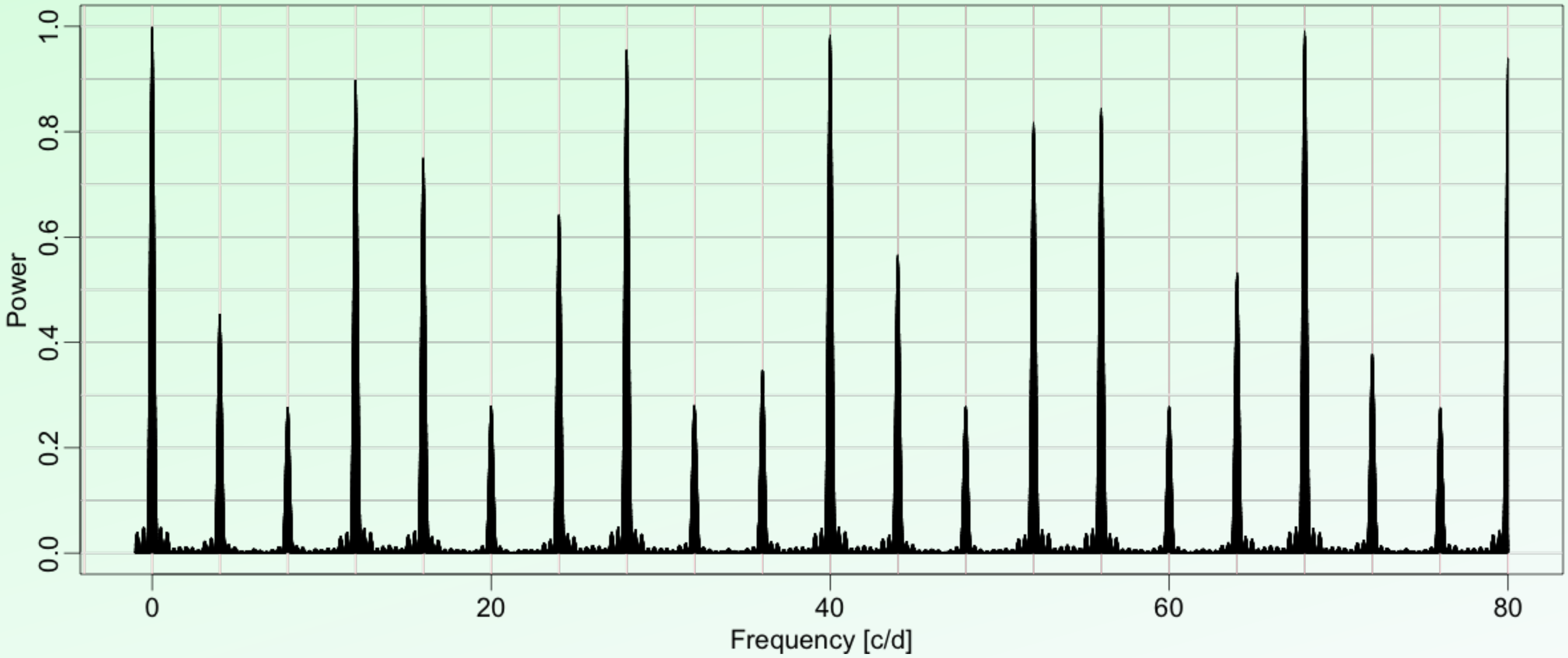
A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (basic Gaia pattern)





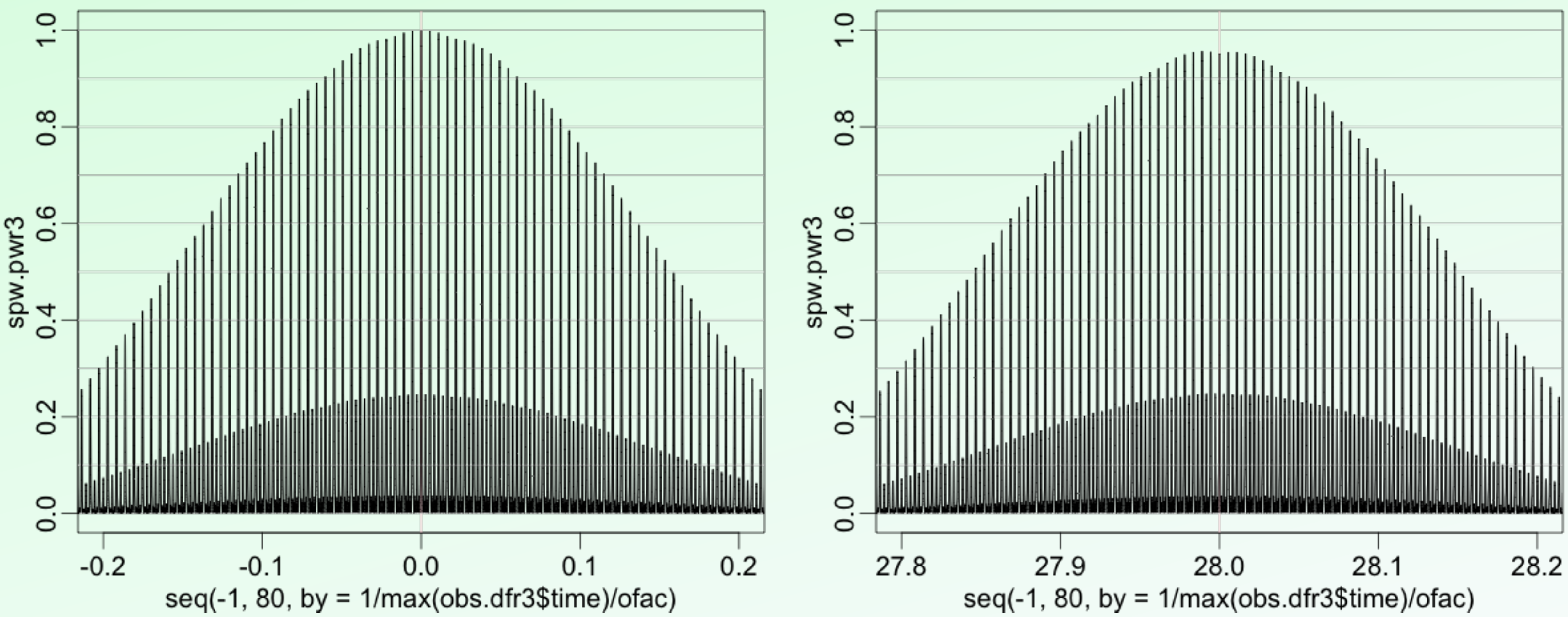
Uneven sampling

A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (basic Gaia pattern), omit observations according to Gaia scanning law…



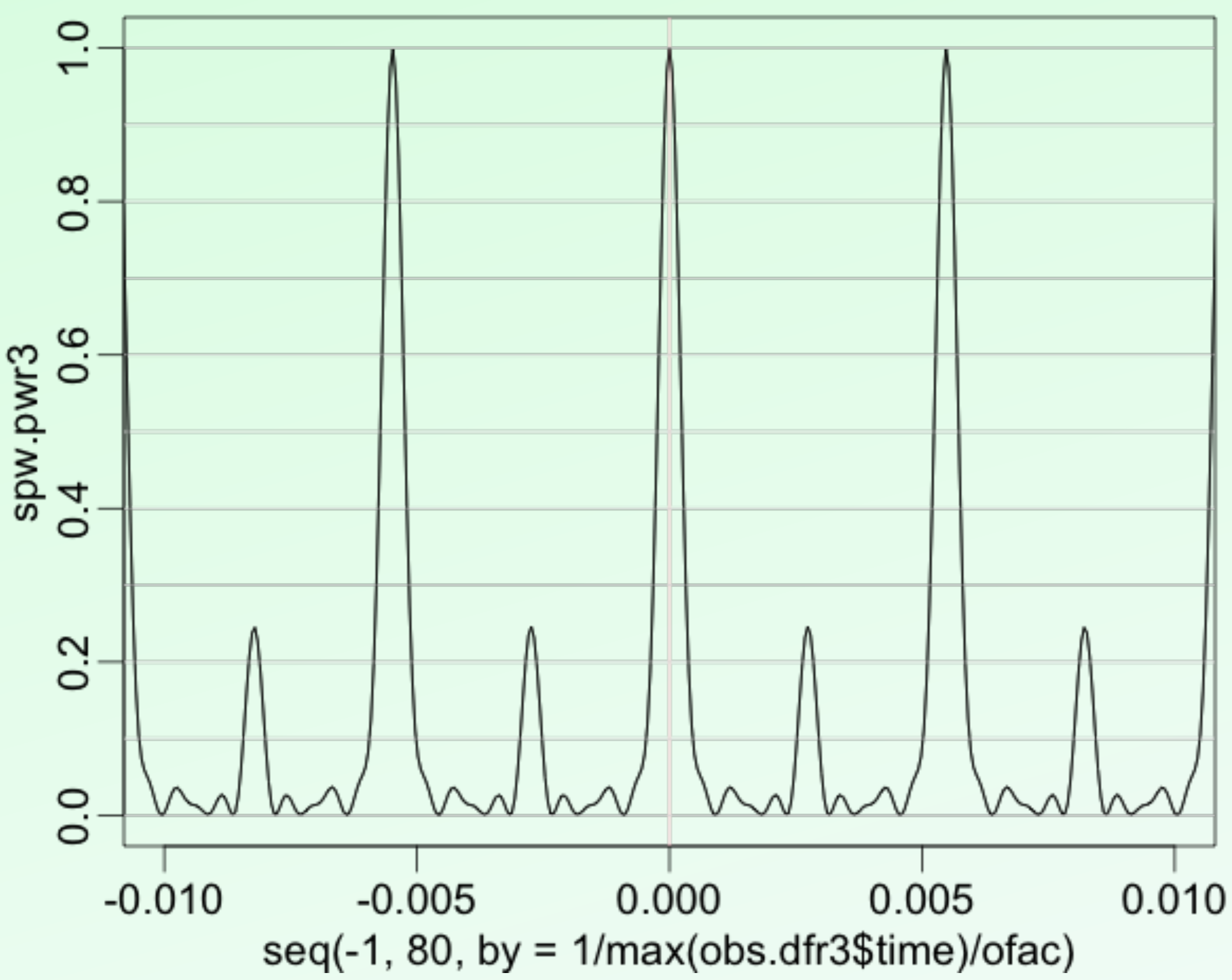
Uneven sampling

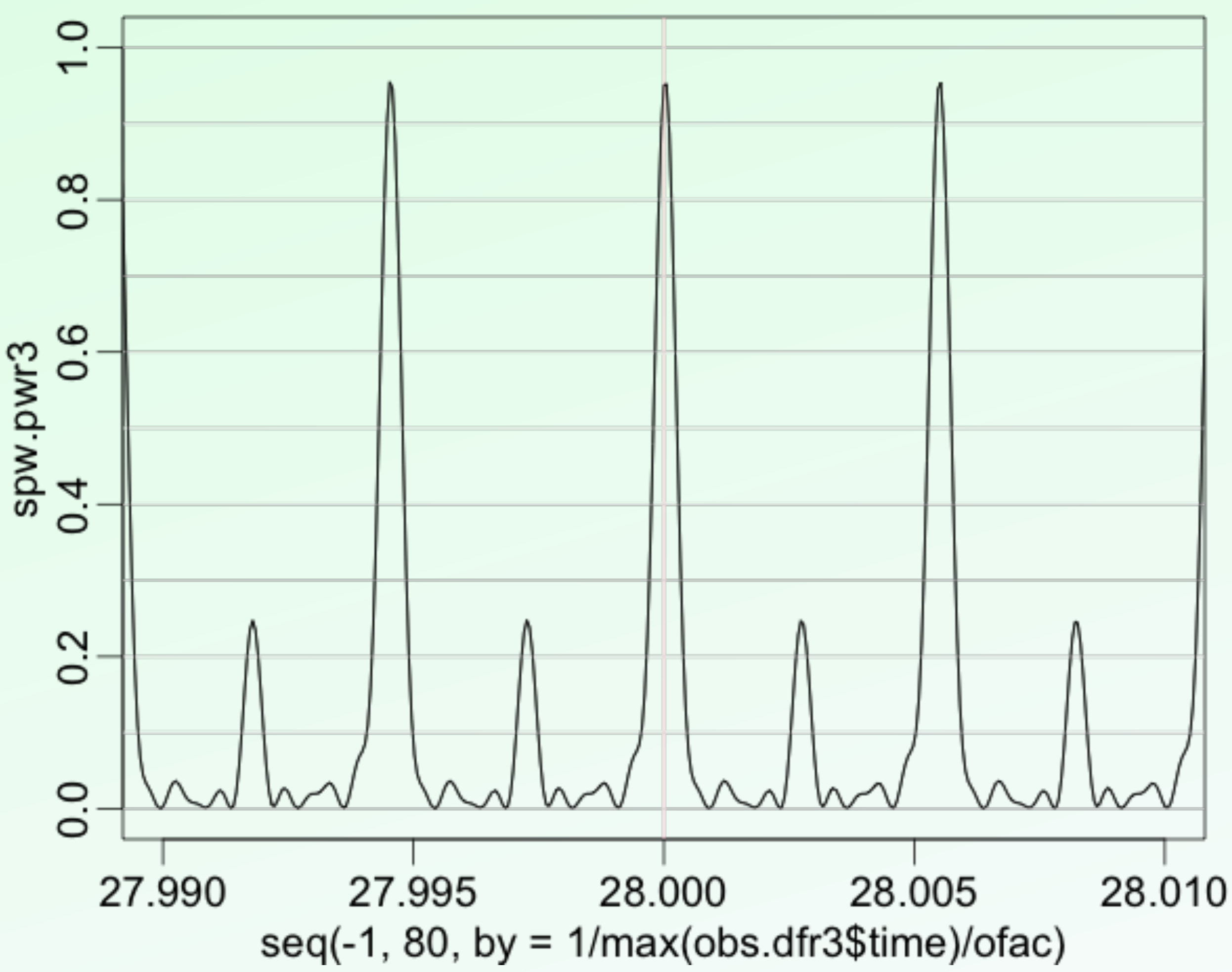
A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (same as that of the two Gaia telescopes), omit observations according to Gaia scanning law…



Uneven sampling

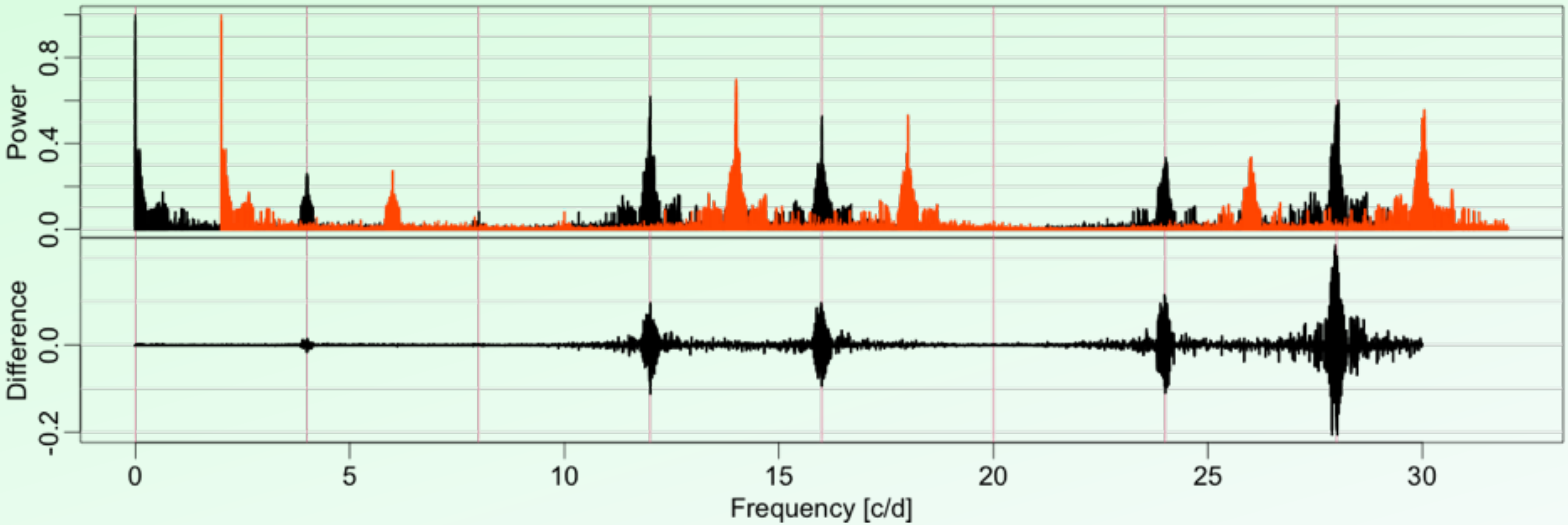
A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (same as that of the two Gaia telescopes), omit observations according to Gaia scanning law…





Uneven sampling

A simple case: two intercalated even samplings of 4 c/d, with 106.5 minute offset (same as that of the two Gaia telescopes), omit observations according to Gaia scanning law, and add a few further corrections including one for barycentric time…



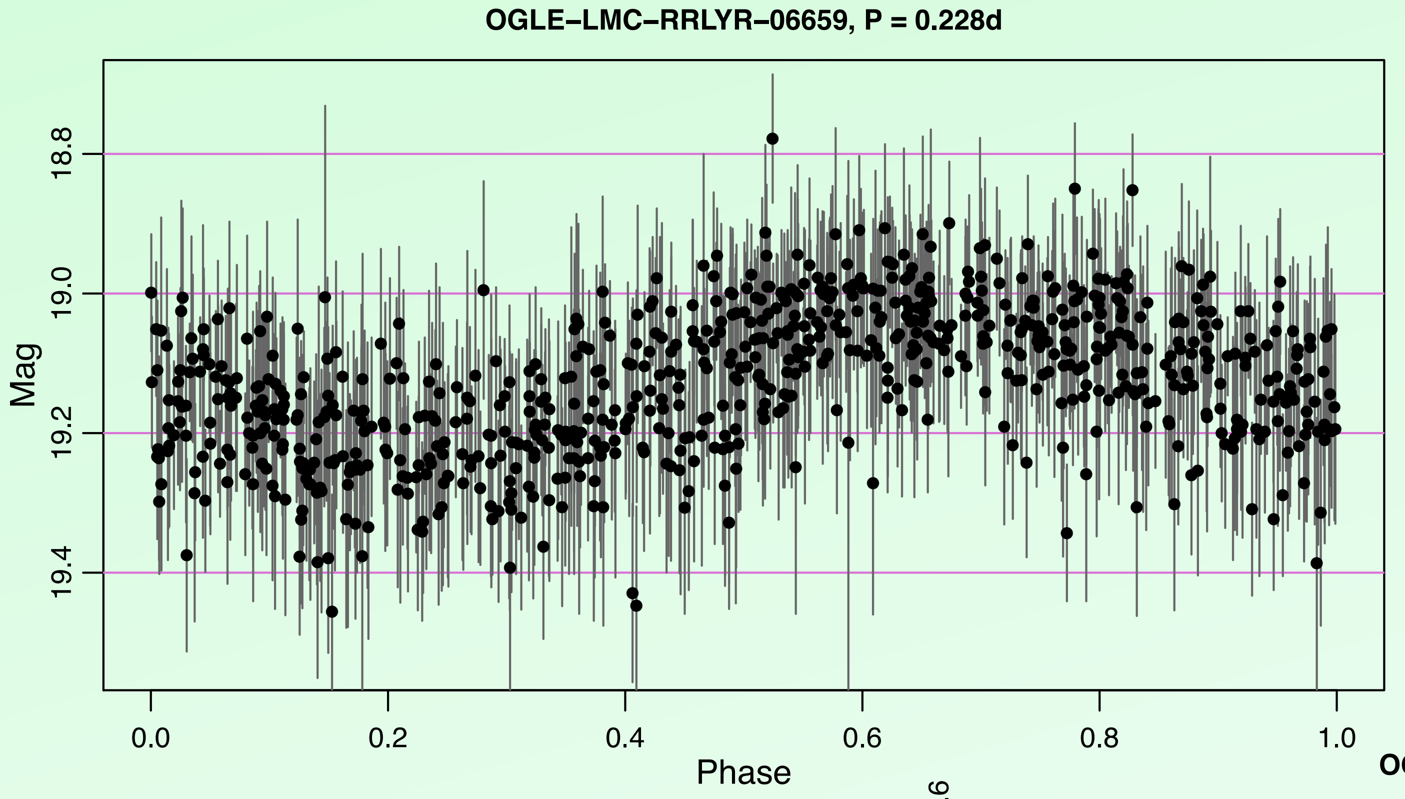
Black line, upper panel: before barycentric time correction

Orange line, upper panel: after barycentric time correction, offset by 2 c/d for visibility

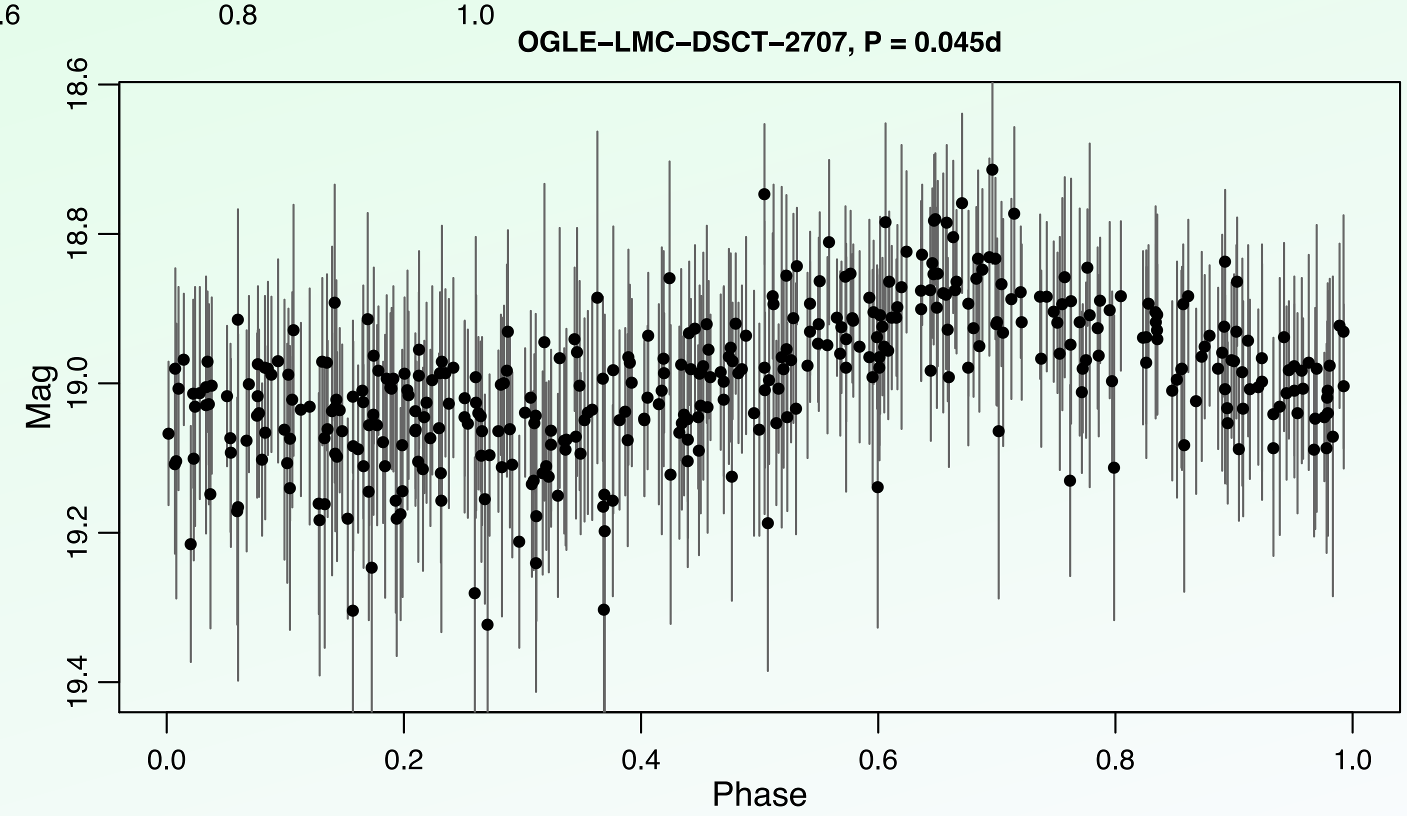
- Ex. 1: spectral window

Uneven sampling

For real stars from OGLE-III:



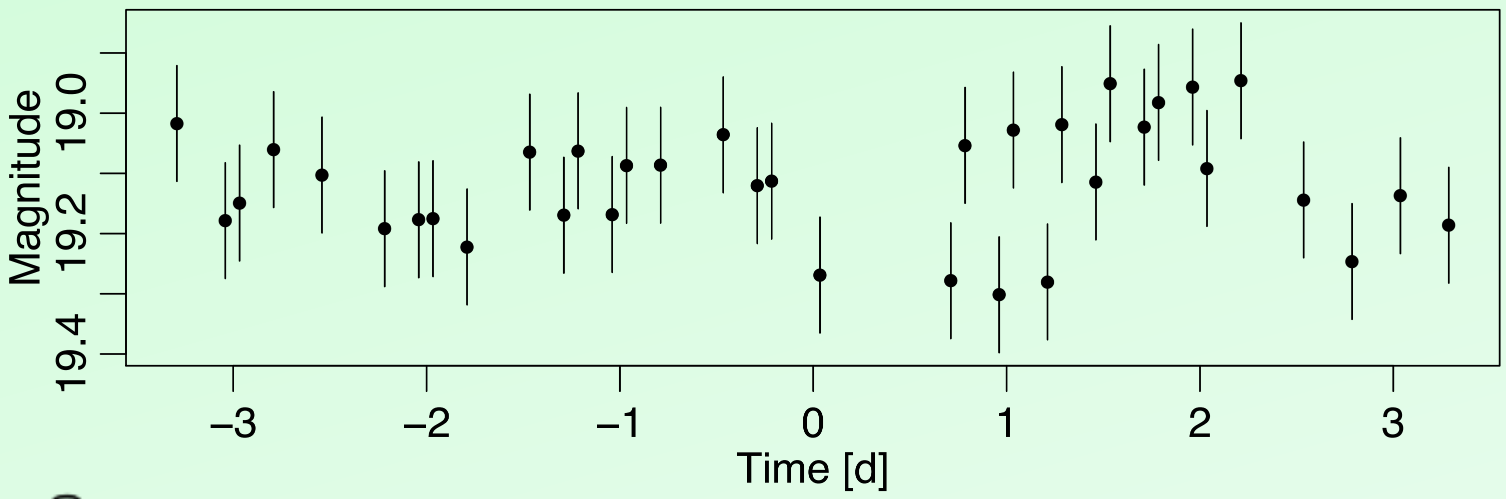
RRe; F = 4.39 c/d



DSCT; F = 22.05 c/d

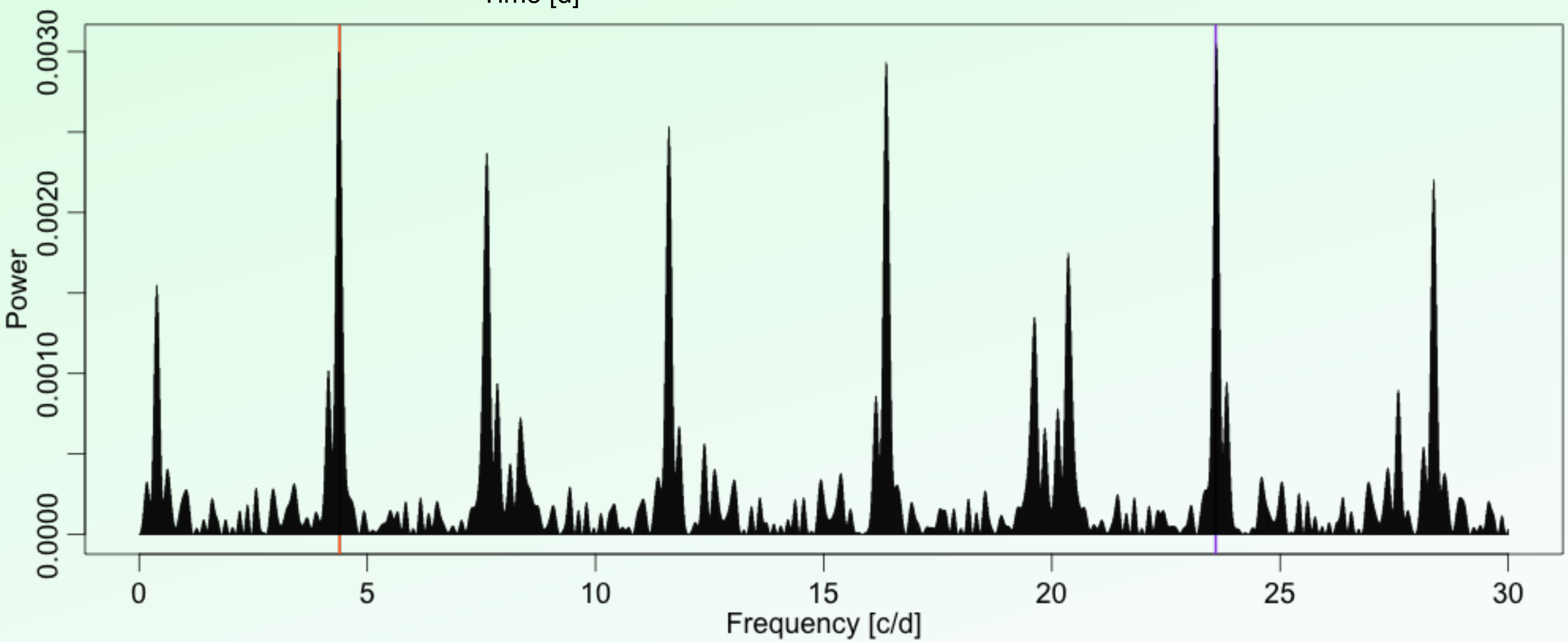
Uneven sampling

The (simulated) RRe with Gaia EPSL observations:



RRe; F = 4.39 c/d

Found: F = 23.59 c/d



Period search methods for uneven sampling

- **Deeming:** direct generalization of the Fourier transform to uneven sampling (Deeming 1975)
- **Least squares methods:**
  - ⋆ *Generalized Least Squares (GLS)*: fitting a single-harmonic + constant model at each trial frequency (Lomb 1976, Scargle 1982, …, Zechmeister & Kürster 2009)
  - ⋆ *Fast χ^2*: fitting a fixed-order harmonic + constant model at each trial frequency (Palmer 2009)
  - ⋆ *Phase Dispersion Minimization*: at each trial frequency, data is binned in phase, and the intra-bin variation is compared to the inter-bin variation (Stellingwerf 1972)
  - ⋆ *Box Least Squares*: fit a box-shaped light curve (for detection of planetary transits; Kovács et al. 2002)
- **Bayesian methods:** GLS put into a Bayesian context (Gregory & Loredo 1992,)
- **Other:**
  - ⋆ *Multi-band GLS*: generalization of GLS for multi-band data (VanderPlas & Ivezić 2015)
  - ⋆ *FAMOUS*: fast pre-selection of interesting frequency neighbourhoods using Deeming, then local GLS (Mignard, unpublished)
  - ⋆ *String length*: related to the Abbe statistics / Durbin-Watson statistics (Dworetsky 1983)
  - ⋆ *Rayleigh tests*: for Poisson-distributed data (Leahy et al. 1983, Paltani 2004)

Periodic time series with noise

Generalized Least Squares

Fits the following model at each test frequency *f*:

$$y_i = a + b\sin(2\pi f t_i) + c\cos(2\pi f t_i) + \epsilon_i,$$

$$\text{where}\ \ \epsilon_i \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma_i^2)$$

Returns a measure of goodness-of-fit *Z*(*f*) as a function of *f* ("periodogram"):

$$Z(f) = \frac{\chi_0^2 - \chi_f^2}{\chi_0^2},$$

where χ_0^2 is the residual sum of squares after fitting a constant model $y_i = a + \epsilon_i$,
and χ_f^2 is the residual sum of squares from the above model at frequency *f*.
With this choice of goodness-of-fit statistic, for Gaussian white noise:

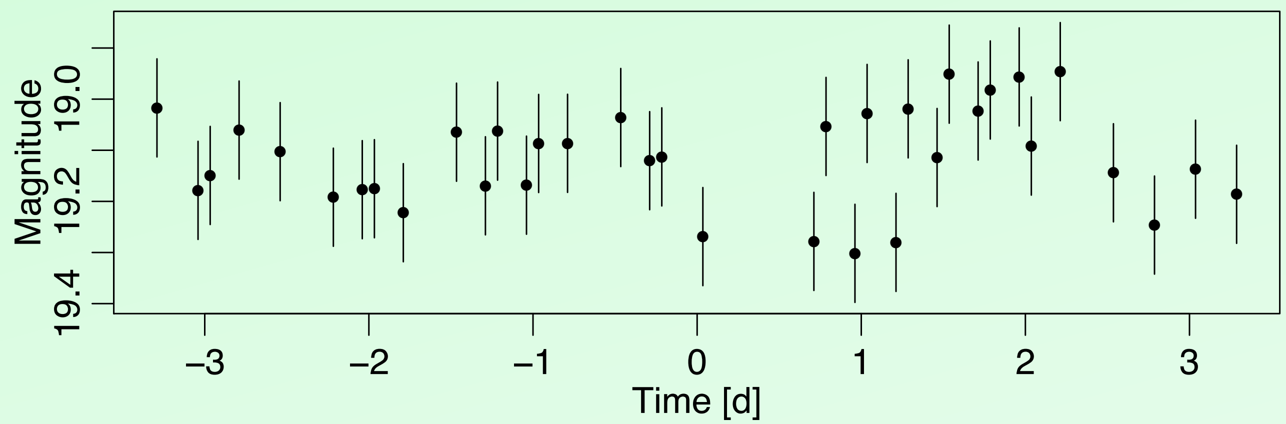
$$Z(f) \sim \text{Beta}\left(1, \frac{N-3}{2}\right).$$

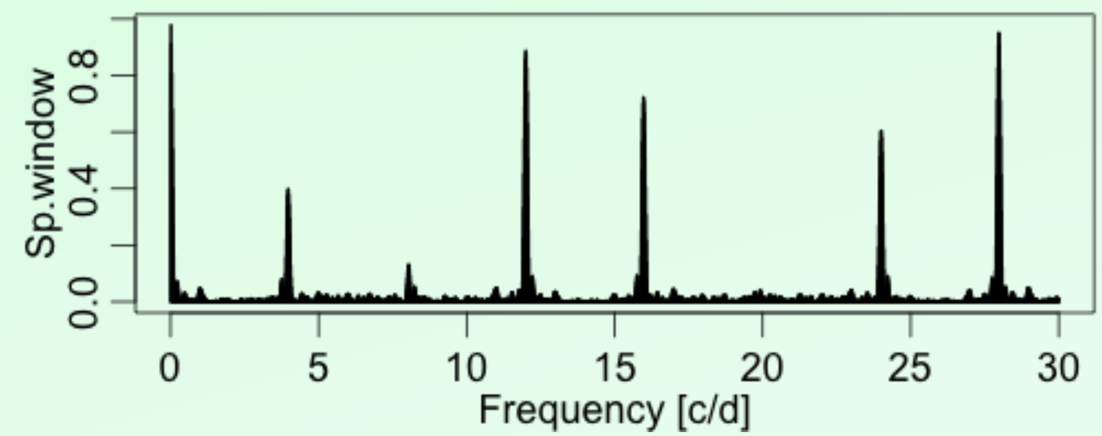
✳ Ex. 2: period search

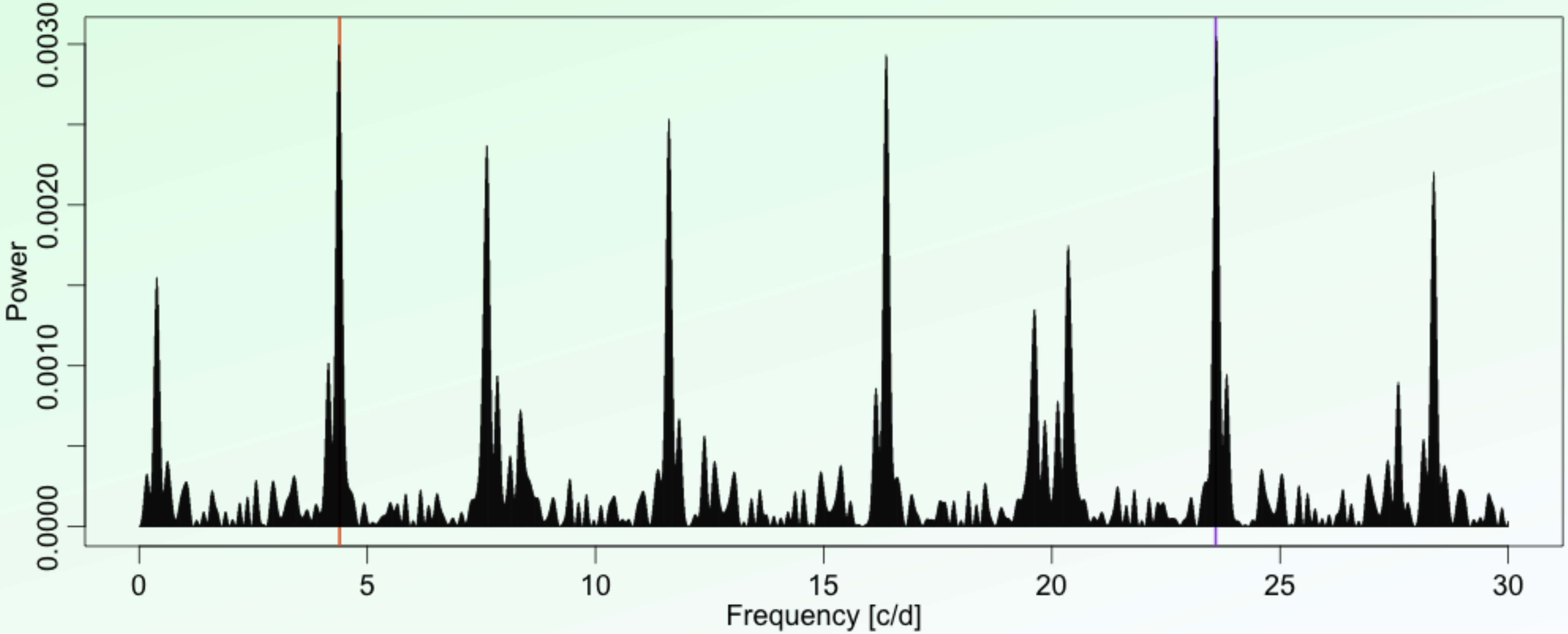
Periodic time series with noise

Generalized Least Squares:

N observations, periodogram at *n* frequencies with $n >> N$

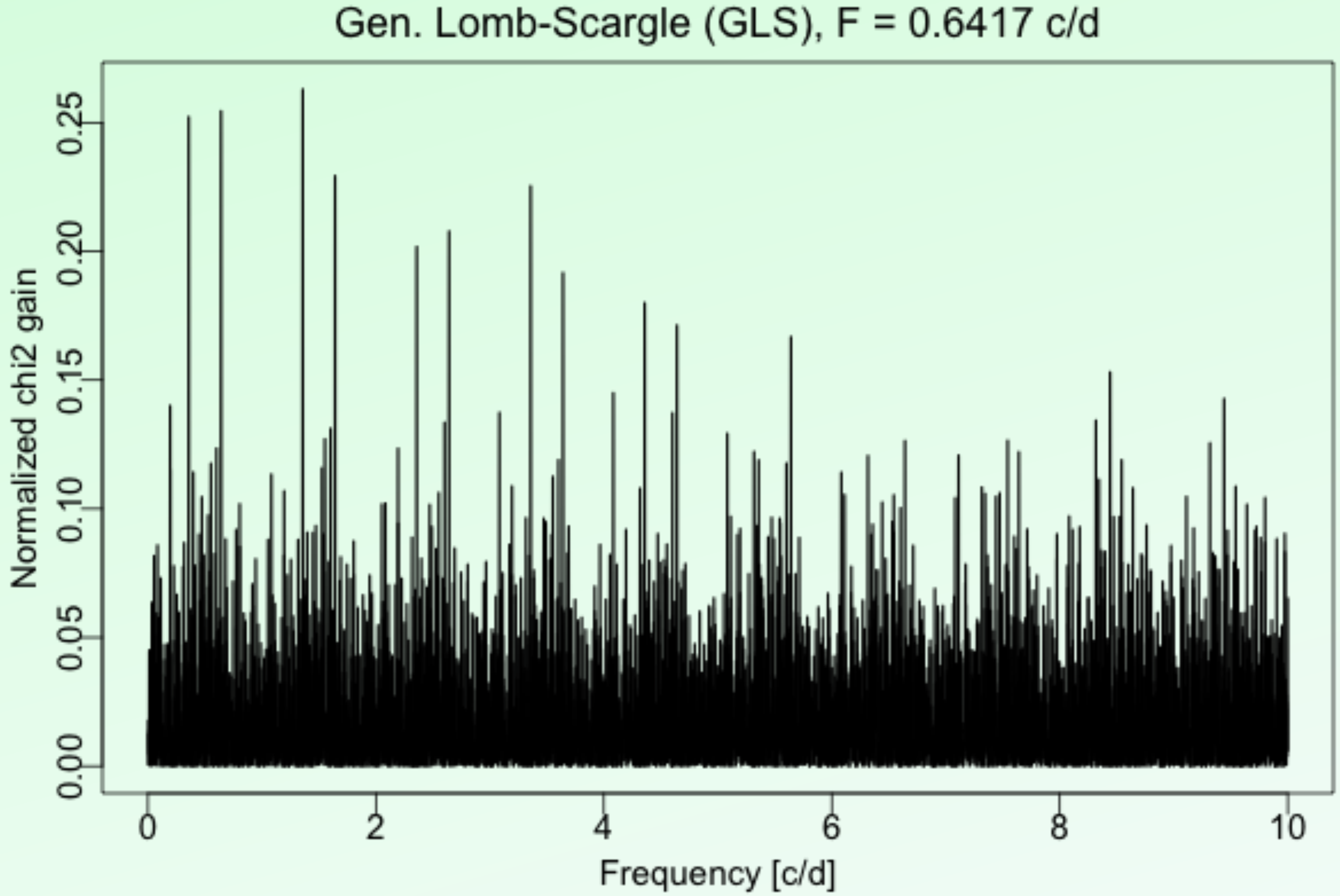






Periodic time series with noise

Hypothesis testing: “Is my variability really there?”



Is this source periodic?

H_0 : the time series does not contain any periodic signal (most often: white noise)

H_1 : the time series contains a periodic component

Periodic time series with noise

Main ingredients for a hypothesis testing (cf. Introduction to Statistics):

- State $\mathbf{H_0}$ and $\mathbf{H_1}$
- select a **test statistic**: here $z_{\max}$, the maximum of the periodogram
- its **distribution** $G(z_{\max})$ under $\mathbf{H_0}$

Result: the false alarm probability **FAP…**

… the probability that white noise produces an equal or higher periodogram maximum than the observed one:

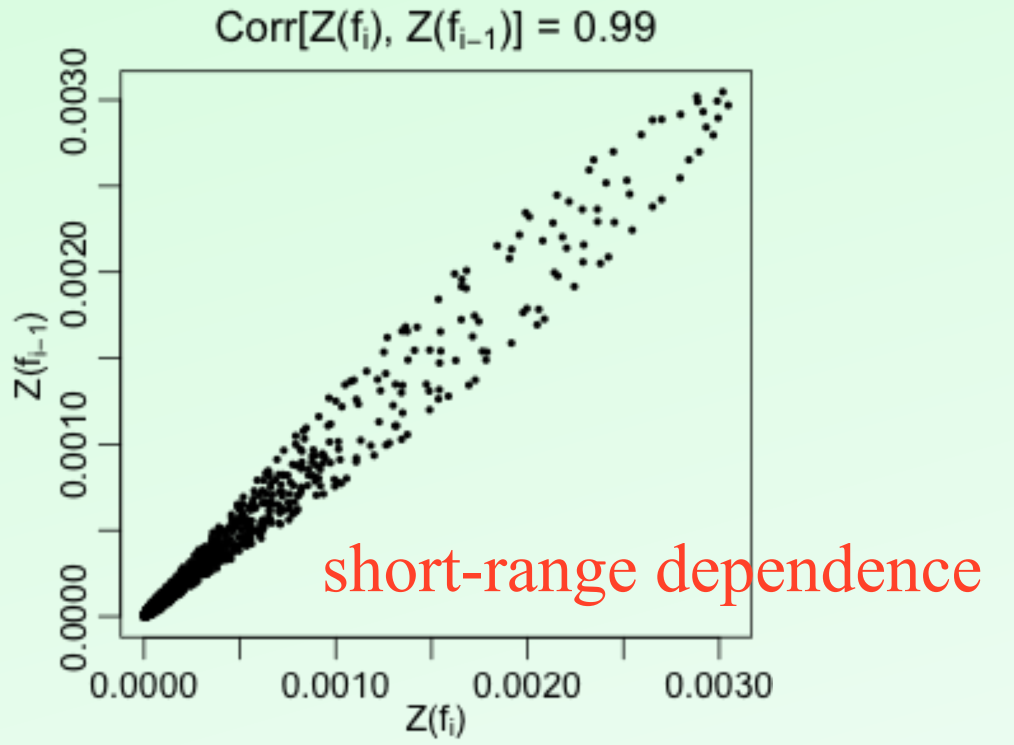
$$\mathrm{FAP} = 1 - G(z_{\max})$$

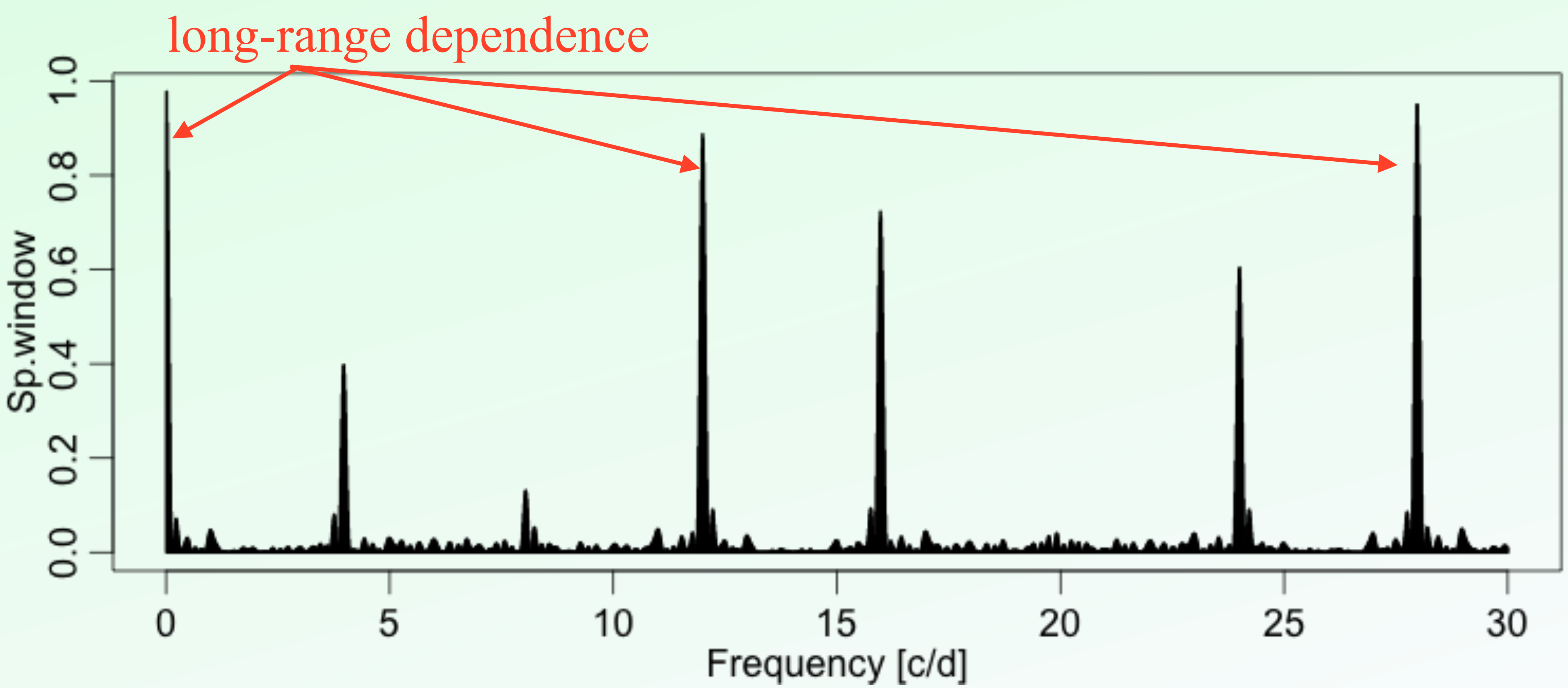
Periodic time series with noise

The distribution $G(z_{max})$ is non-trivial:

though the marginal distribution at each frequency is known, the periodogram is not a collection of independent variables!

- nb. of obs. << nb. of frequencies ⇒ short-range dependence
- time sampling patterns induce aliasing ⇒ long-range dependence





Periodic time series with noise

Alternatives: based on…

- an equivalent independent frequency system: method $\boldsymbol{F^M}$

  To estimate: ***M***

  Paltani (2004) A&A, **420**, 789; Schwarzenberg-Czerny (2012) IAU Symposium, **285**, 81

- the theory of excursions of stochastic processes to extreme levels: method **Baluev**

  Nothing to estimate

  Baluev (2008) MNRAS **385**, 1279 and later publications

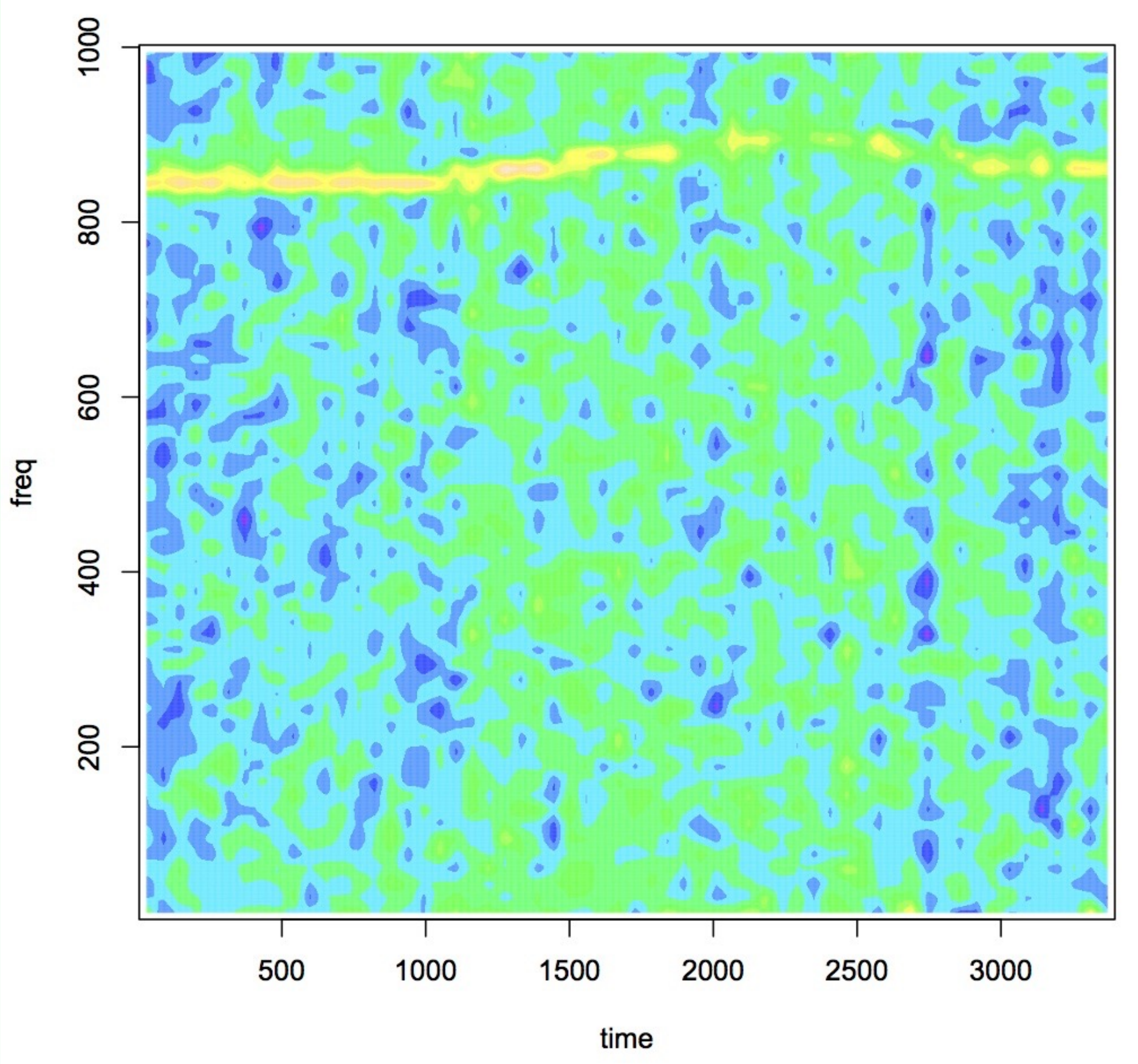
- the basic Central Limit Theorem-like result of extreme-value statistics: method **GEV**

  To estimate: **the three parameters of the GEV**

  Süveges (2014) MNRAS, **440**, 2099

Towards nonperiodicity: time-resolved methods

Fourier/harmonic analysis in time segments, if you have enough data



High-frequency quasi-periodic oscillation in the X-ray binary ….

Towards nonperiodicity: time-resolved methods

Wavelet analysis: use temporally localized waveforms instead of sine functions.

- In original form, developed for even sampling, and heavily builds on orthogonality of the basis functions.
- Wavelets are constructed step-by-step from a scaling function (father wavelet), halving the scale at each step, and describing more details.
- All wavelets are fitted simultaneously in a large model. Then based on the assumption of sparsity (only a few are important in the signal), thresholding or other sparse methods are applied in order to select the relevant components.

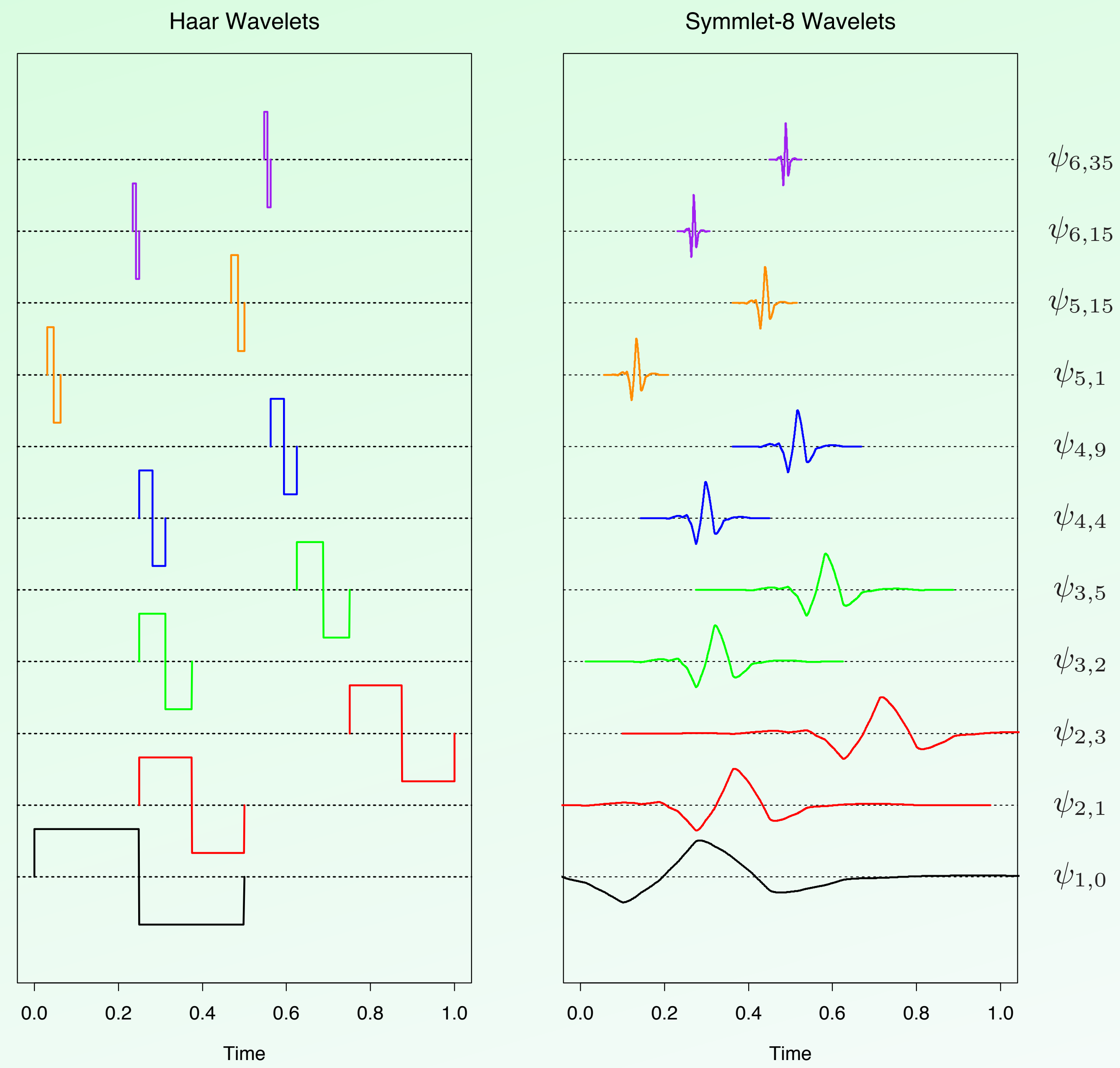


FIGURE 5.16. *Some selected wavelets at different translations and dilations for the Haar and symmlet families. The functions have been scaled to suit the display.*

Hastie, Tibshirani and Friedman. *Elements of Statistical Learning*, Springer, 2009

Towards nonperiodicity: time-resolved methods

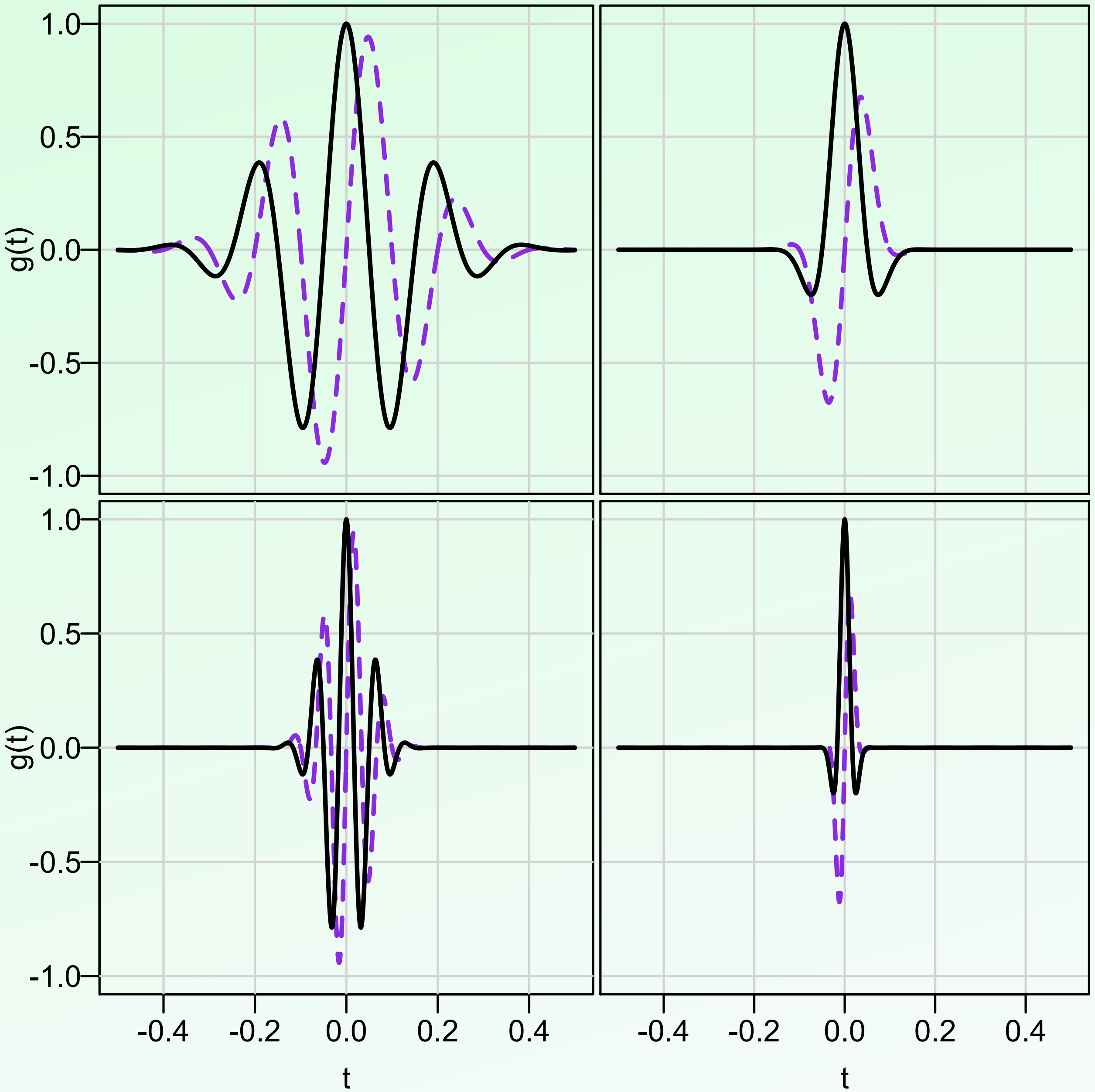
Wavelet analysis: use temporally localized waveforms instead of sine functions.

- In original form, developed for even sampling, and heavily builds on orthogonality of the basis functions.
- In implementations for uneven sampling, orthogonality is lost, and time-resolved real phenomena are tough to disentangle from the effects of the time sampling, similar to aliasing (see Foster (1996), AJ **112**, 1709)
- A specific wavelet: Morlet wavelet

$$g(t) = A \mathrm{e}^{i\,2\pi f_0 (t-t_0)} \mathrm{e}^{-f_0^2 (t-t_0)^2/\tau^2}$$

Wavelet transform of signal $h(t)$:

$$H(t_0, f_0, \tau) = \int_{-\infty}^{\infty} h(t) g(t; t_0, f_0, \tau) dt$$

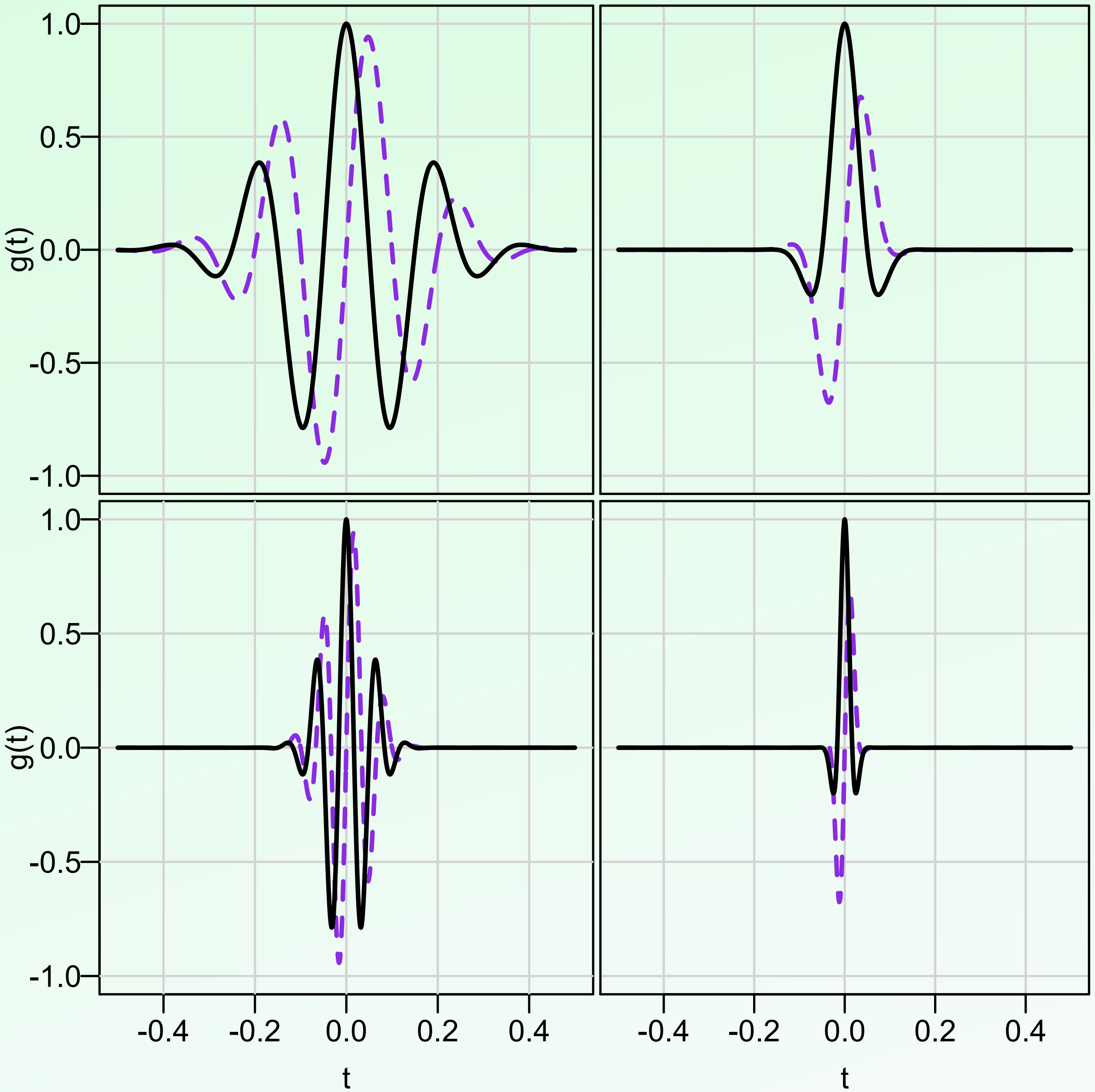


solid black line: Re$\{g(t)\}$
broken violet line: Im$\{g(t)\}$

Towards nonperiodicity: time-resolved methods

Matching pursuit: iterative fitting/prewhitening the data with Gabor atoms (= Morlet wavelets)

- Uses a redundant dictionary (of Gabor atoms/Morlet wavelets plus sinusoids) (most often, millions).
- Looks for the best-matched atom, removes it from the data, continues on the residuals, etc.
- Similar to successive frequency searches when multiperiodicity is suspected, but also applicable for time-resolved frequency analysis



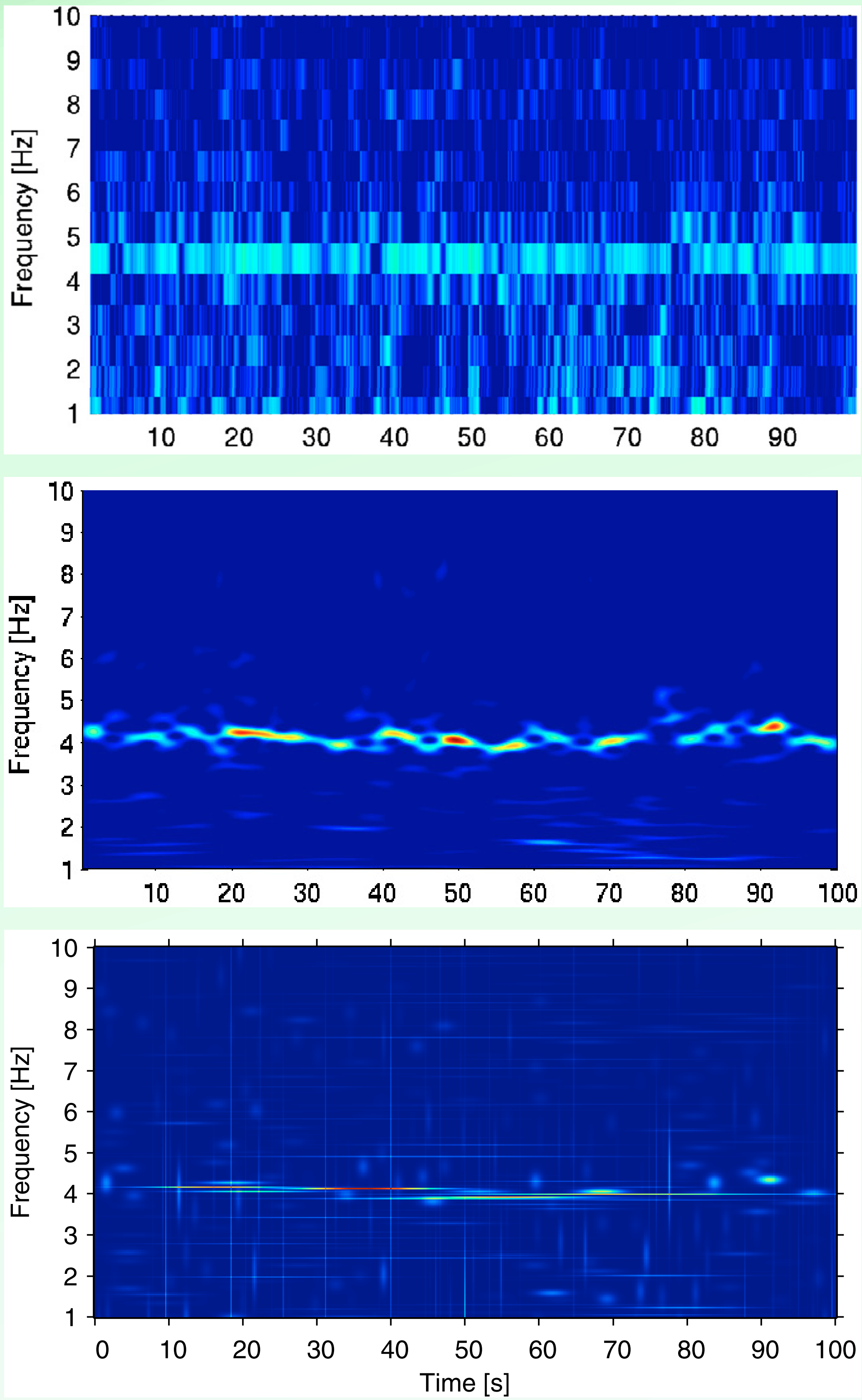
solid black line: Re{g(t)}
broken violet line: Im{g(t)}

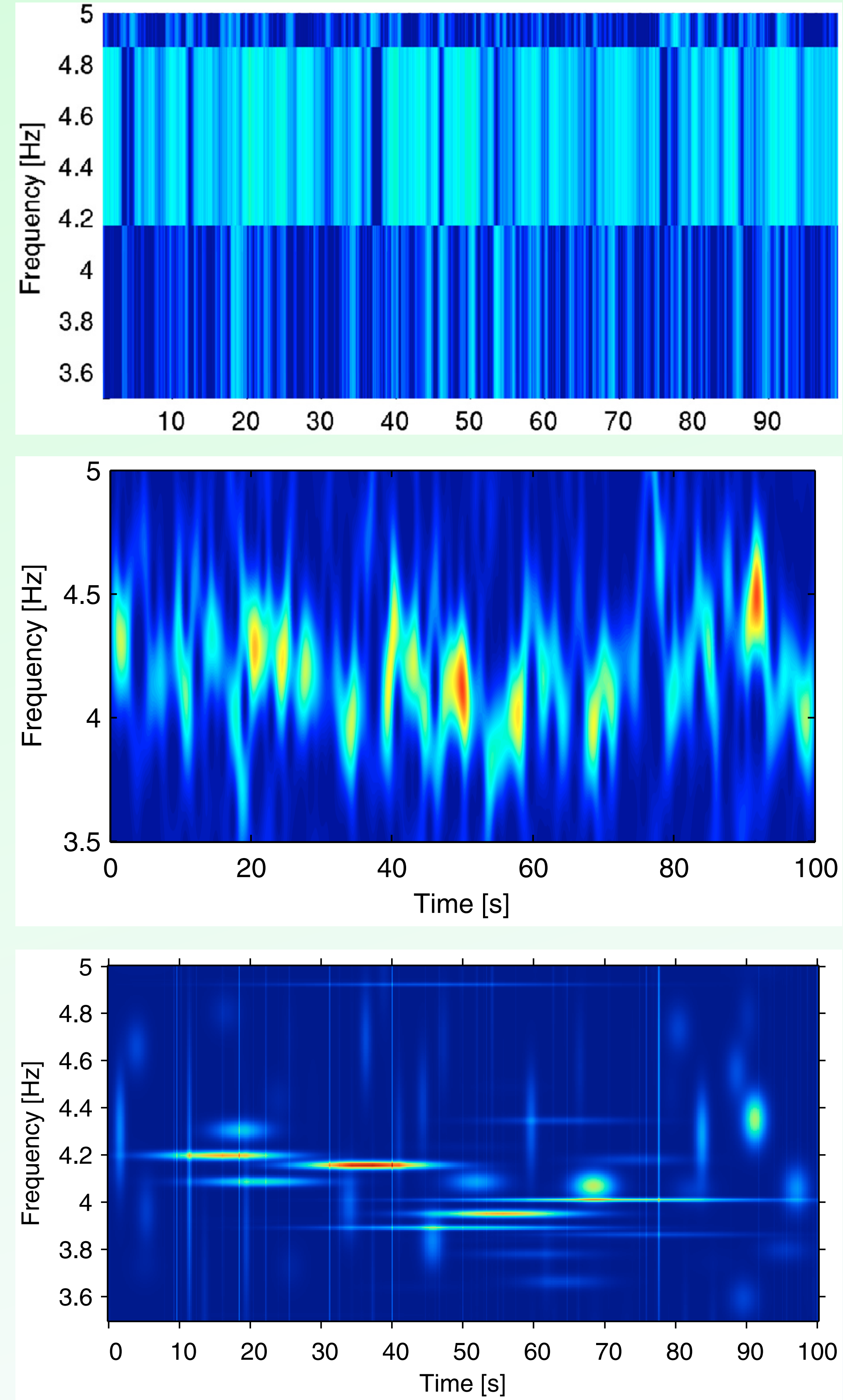
Towards nonperiodicity: time-resolved methods

Segmented Fourier analysis

Wavelet analysis

Matching pursuit





Lachowitz & Done (2010) A&A **515**, A65

XTE J1550–564

✳ Ex. 3: time-rsvd, Cep, XRB

Stochastic processes

A stochastic process is given by:

- a probability space $(\Omega, \mathcal{F}, P)$;
- a state space (S, Σ) which is indexed by an (ordered) set T

Then the collection of S-valued random variables $\{X_t : \Omega \to S, \ \forall \ t \in T\}$ is a stochastic process.

Practically any time series you can think of can be cast as a stochastic process.

Some kinds of stochastic processes:

- random walks
- Markov processes
- autoregressive—moving average (ARMA) and derived processes: ARIMA, fractionally integrated ARIMA, ARCH, GARCH, etc.
- continuous-time ARMA and variants
- state-space and hierarchical models

Ornstein-Uhlenbeck process

… i.e., damped random walk

Properties:

- (strictly) stationary;
- all marginal distributions are Gaussian
- Markovian (the present state depends on only the immediate past state(s))

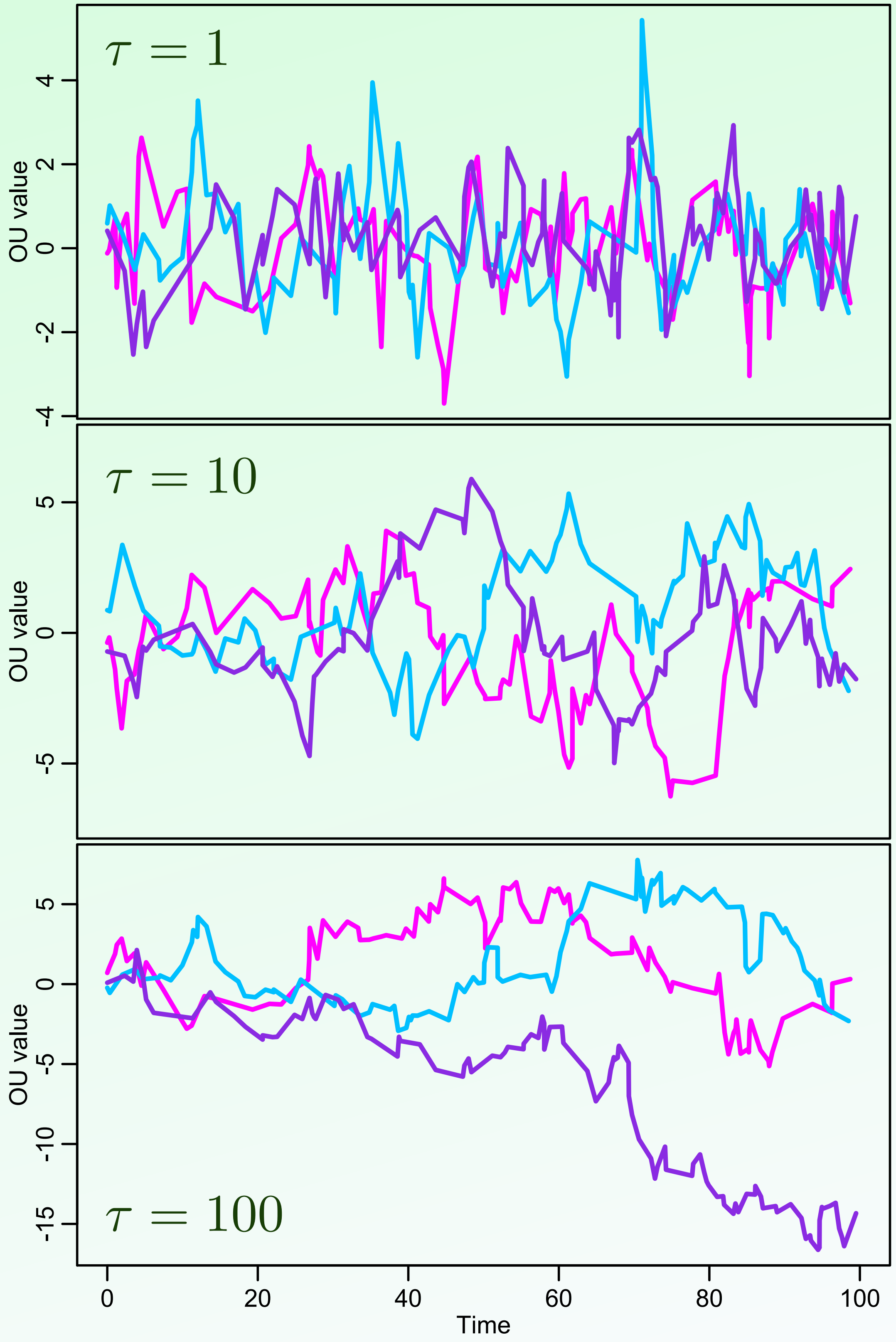
The OU process is the only stochastic process possessing all these properties.

Formulation:

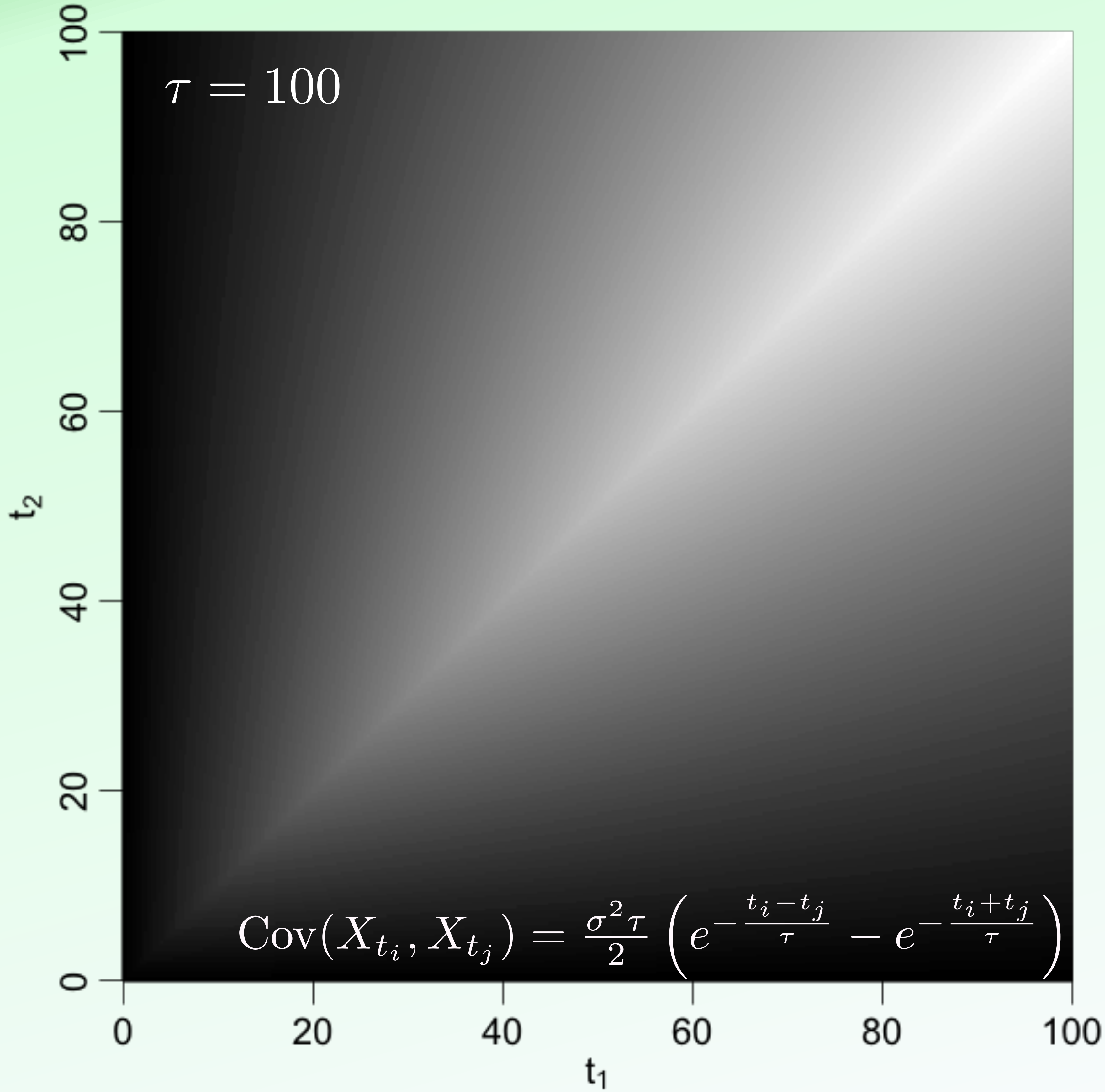
$$Y_i = \phi(t_i, t_{i-1}) Y_{i-1} + \epsilon_i,$$

with $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$

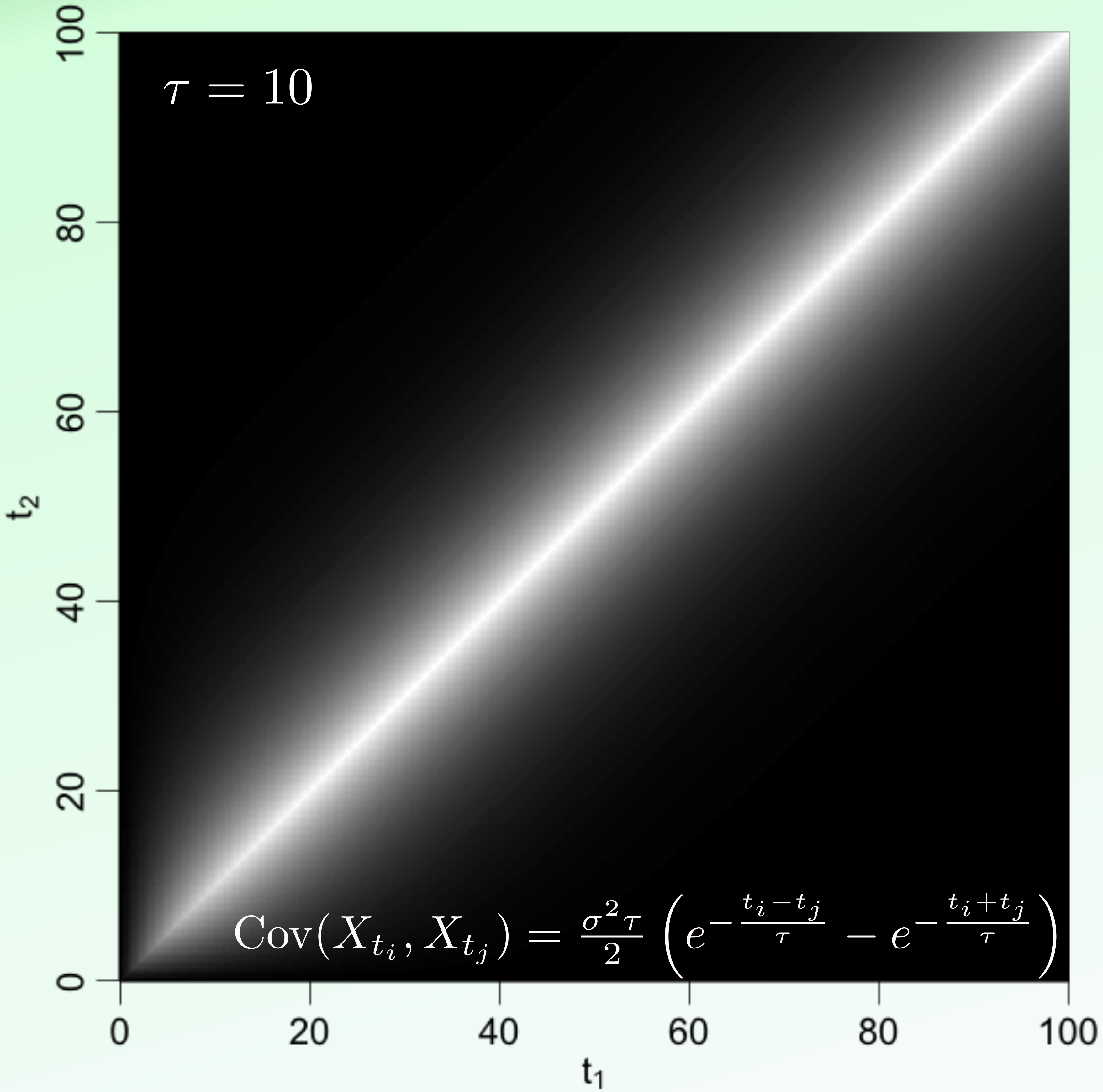
$$\phi(t_i, t_{i-1}) = \mathrm{e}^{-\frac{|t_i - t_{i-1}|}{\tau}}$$



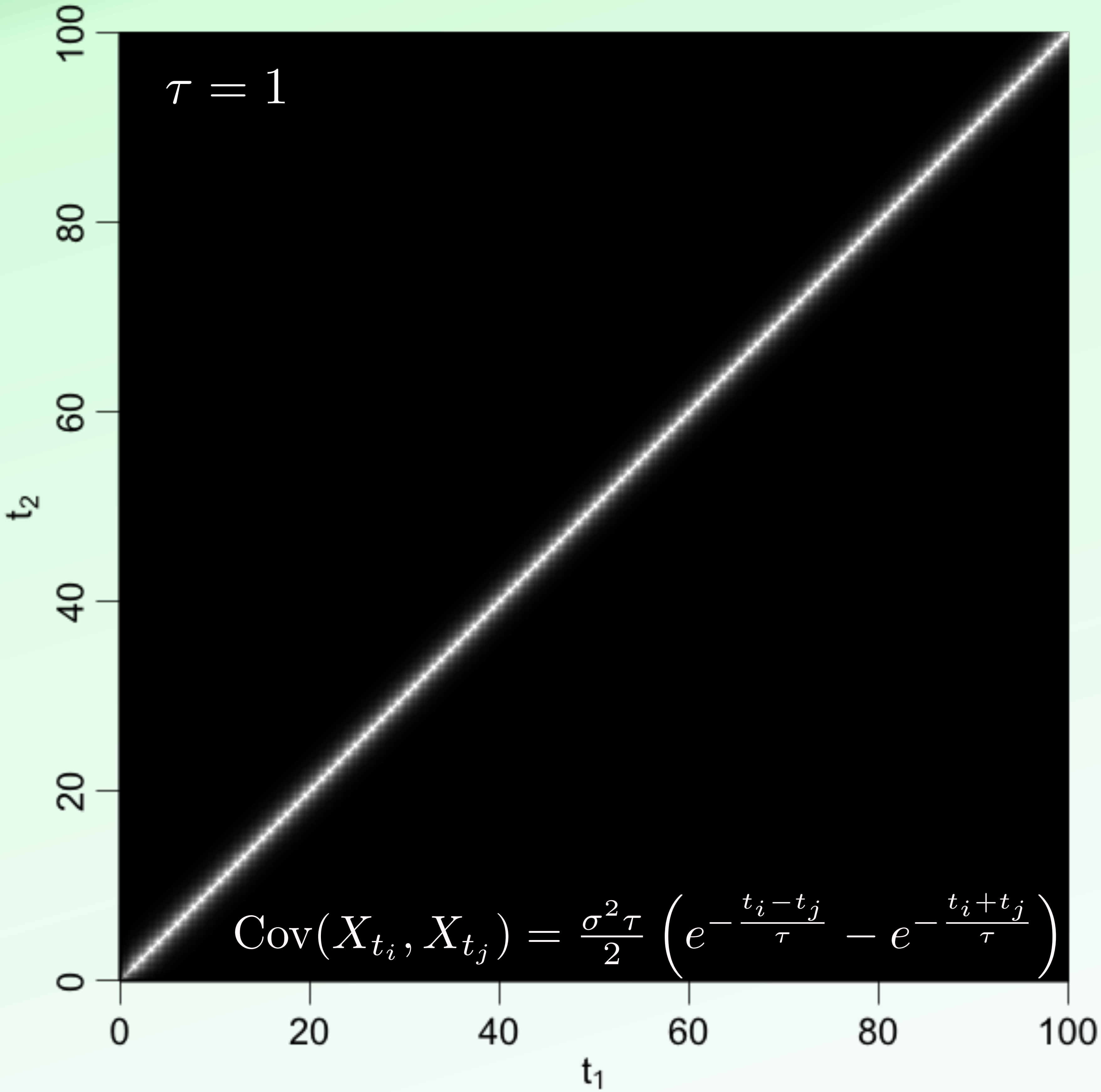
Ornstein-Uhlenbeck process



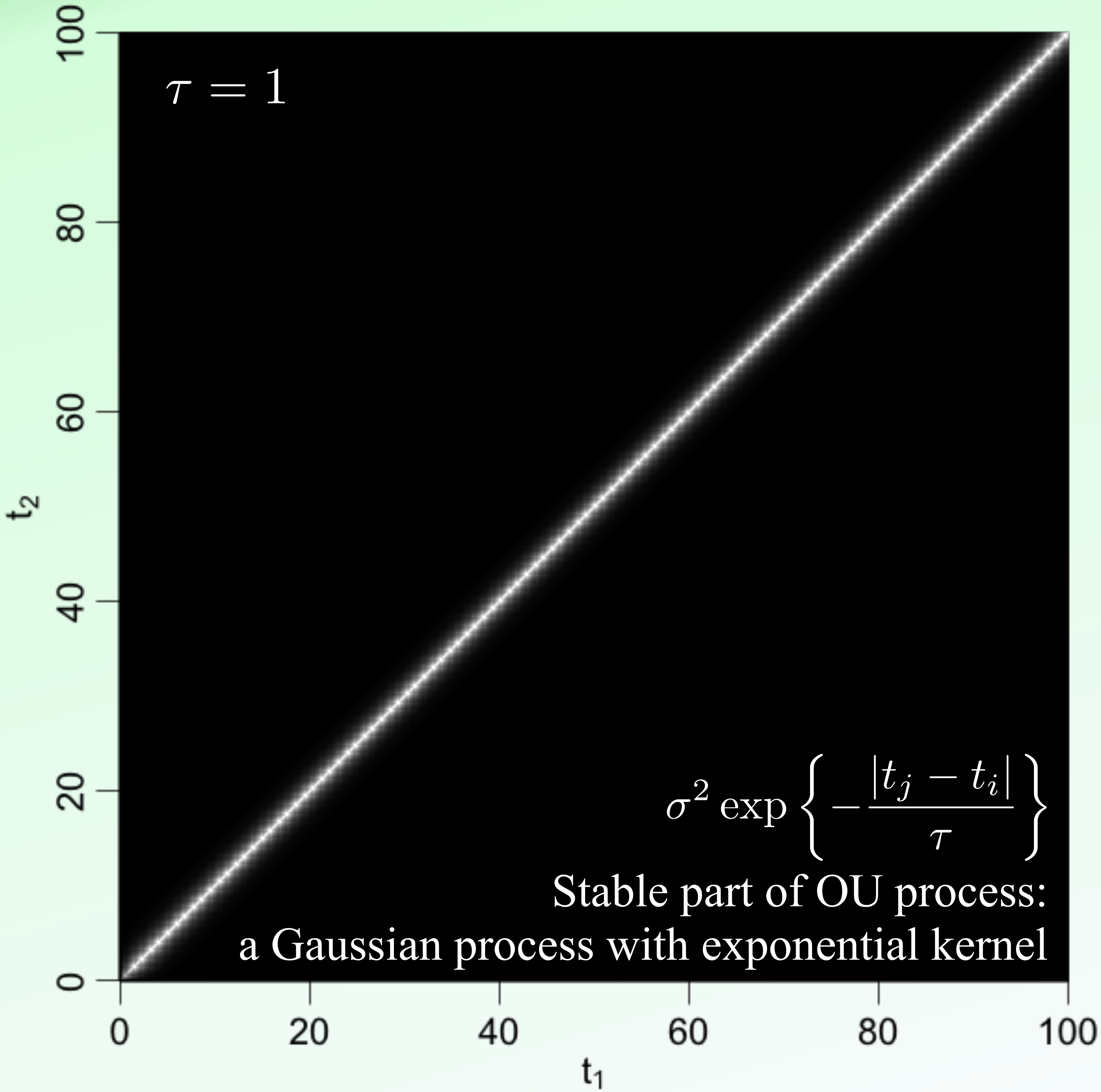
Ornstein-Uhlenbeck process



Ornstein-Uhlenbeck process



Gaussian process: exponential kernel



Gaussian process

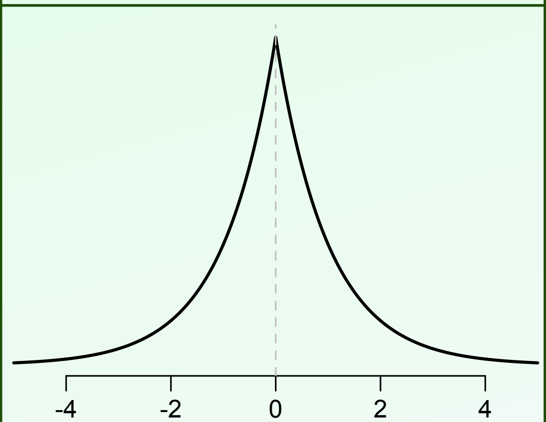
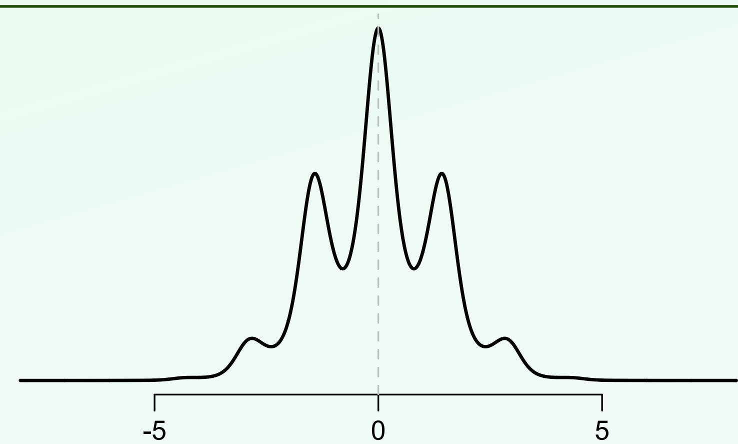
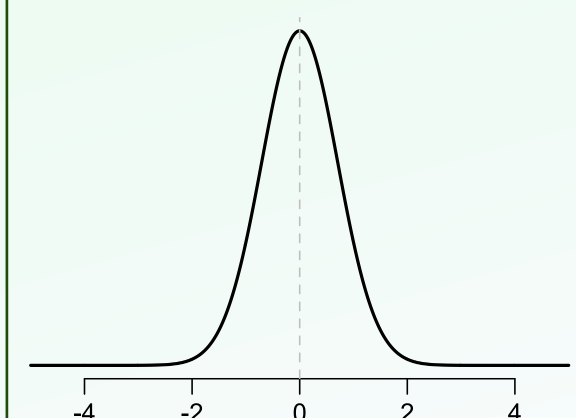
General Gaussian likelihood for model $m(\boldsymbol{t}; \boldsymbol{\theta})$, where $\boldsymbol{t} = \{t_1, \ldots t_N\}^T$:

$$\ell(\boldsymbol{\theta}; \boldsymbol{t}) = -\frac{1}{2}[\boldsymbol{y} - m(\boldsymbol{t}; \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{y} - m(\boldsymbol{t}; \boldsymbol{\theta})] - \frac{1}{2} \log \det \boldsymbol{\Sigma} + \text{cst.}$$

where

$$\boldsymbol{\Sigma}_{ij} = \mathrm{Cov}\{\epsilon(t_i), \epsilon(t_j)\} = k(t_i, t_j)$$

$k(t_i, t_j)$: **kernel function**. Some options below:

Kernel name	Exponential	Quasi-periodic (variant with always positive corr.)	Squared exponential
$k(t_i, t_j)$	$\sigma^2 \exp\left\{-\frac{\lvert t_j - t_i \rvert}{\tau}\right\}$	$\sigma^2 \exp\left\{-\frac{(t_j - t_i)^2}{\lambda^2} - \frac{\sin^2[2\pi\nu(t_j - t_i)]}{\lambda^2}\right\}$	$\sigma^2 \exp\left\{-\frac{(t_j - t_i)^2}{\tau^2}\right\}$
Plot			

Gaussian process

General Gaussian likelihood for model $m(\boldsymbol{t};\, \boldsymbol{\theta})$, where $\boldsymbol{t} = \{t_1, \ldots t_N\}^T$:

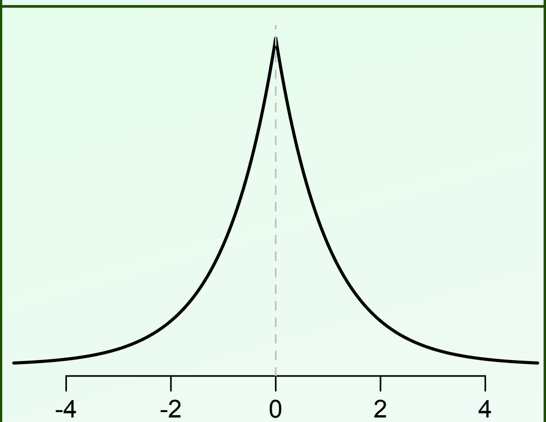
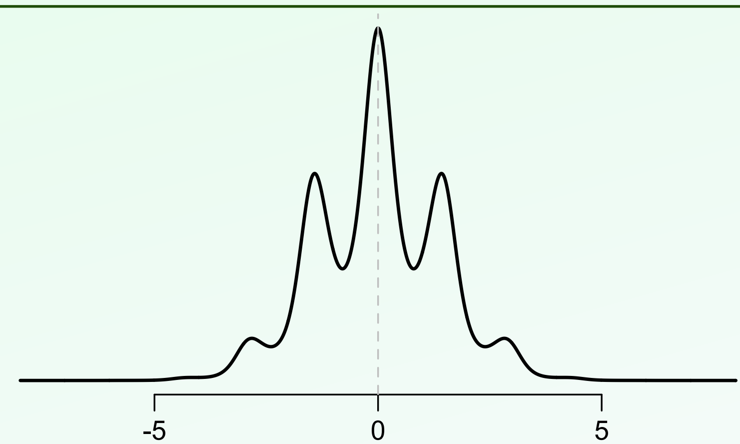
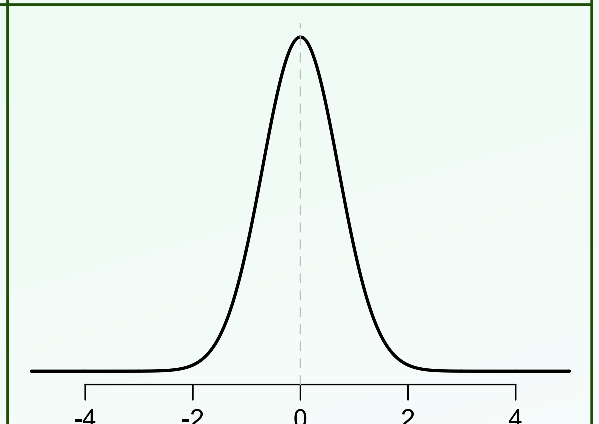
$$\ell(\boldsymbol{\theta}; \boldsymbol{t}) = -\frac{1}{2}[\boldsymbol{y} - m(\boldsymbol{t}; \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{y} - m(\boldsymbol{t}; \boldsymbol{\theta})] - \frac{1}{2} \log \det \boldsymbol{\Sigma} + \text{cst.}$$

where

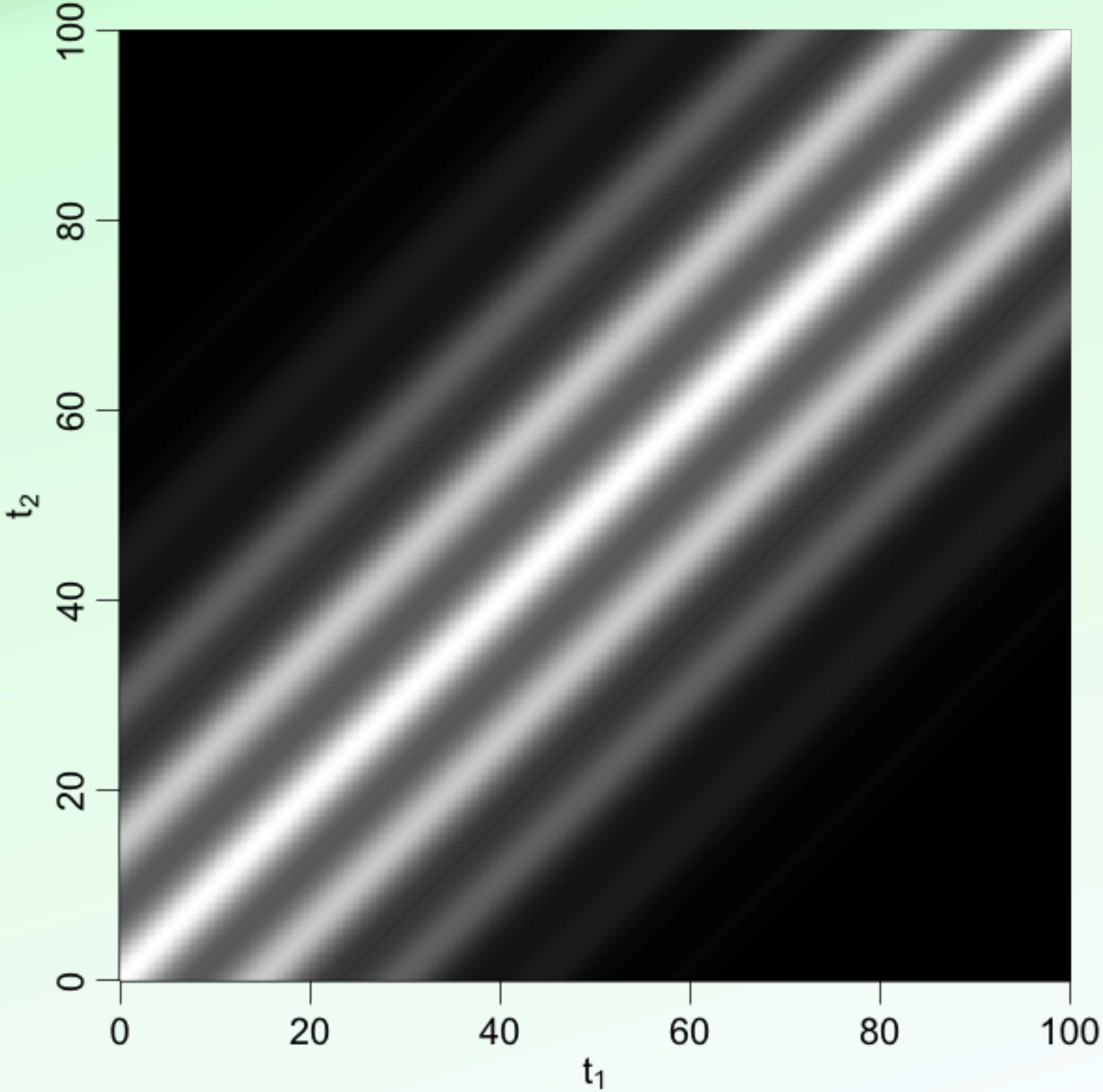
$$\boldsymbol{\Sigma}_{ij} = \text{Cov}\{\epsilon(t_i), \epsilon(t_j)\} = k(t_i, t_j) + \sigma_i^2 \delta_{ij}$$

"**nugget effect**": additional variance at each point

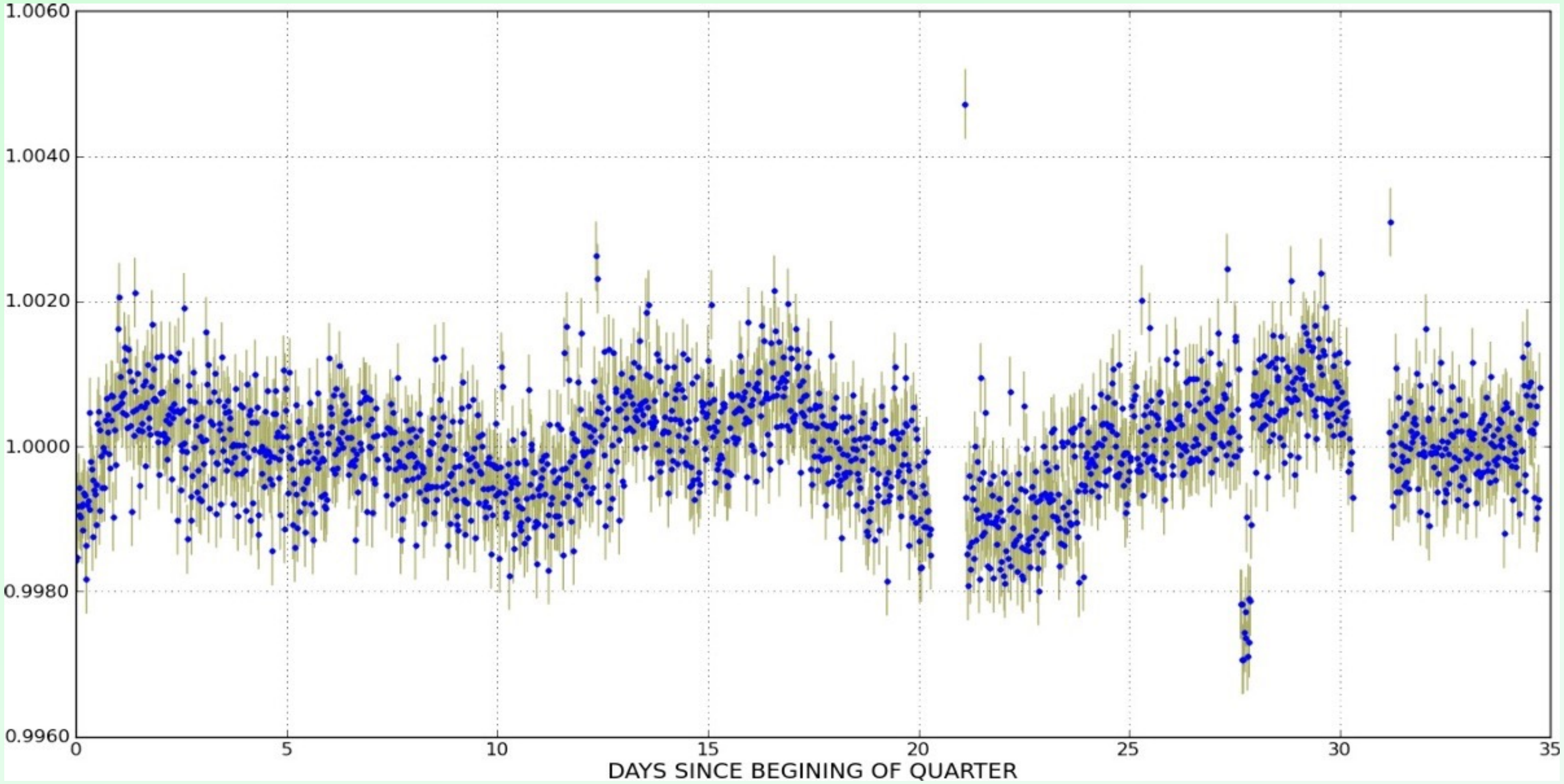
$k(t_i, t_j)$: **kernel function**. Some options below:

Kernel name	Exponential	Quasi-periodic (variant with always positive corr.)	Squared exponential
$k(t_i, t_j)$	$\sigma^2 \exp\left\{-\frac{\lvert t_j - t_i \rvert}{\tau}\right\}$	$\sigma^2 \exp\left\{-\frac{(t_j - t_i)^2}{\lambda^2} - \frac{\sin^2[2\pi\nu(t_j - t_i)]}{\lambda^2}\right\}$	$\sigma^2 \exp\left\{-\frac{(t_j - t_i)^2}{\tau^2}\right\}$
Plot			

Gaussian process: exponential-cosine kernel



Gaussian process

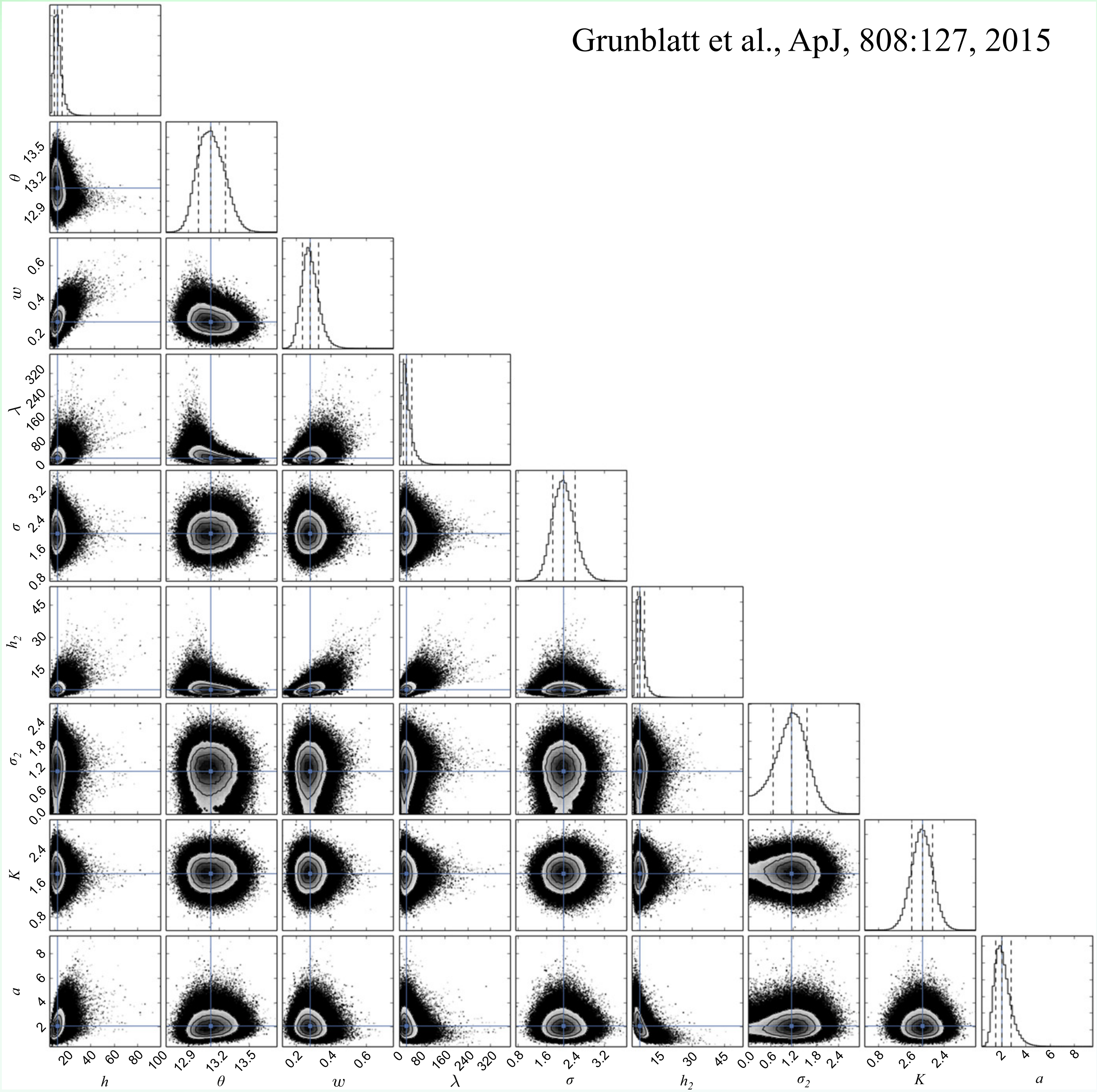


Example: planet candidate from Kepler, where correlation between consecutive residuals cannot be ignored (an improved model: Gaussian processes)

Image: http://keplerlightcurves.blogspot.ch/2012/09/the-eleventh-planethunters-habitable.html

Fitting them: usually, by Bayesian methods.

Gaussian process



✳ Ex. 4: gp

Useful references

Specifically for astronomers:

- Ž. Ivezić, A. Connolly, J. VanderPlas, A. Gray. *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press, 2014

General:

- P. Brockwell, R. Davis. *Introduction to Time Series and Forecasting.* Springer, 2002
- P. Brockwell, R. Davis. *Time Series: Theory and Methods*. Springer, 1991
- Shumway & Stoffer: *Time Series Analysis and its Applications:With R examples.* Springer, 2010

Spectral analysis:

- P. Bloomfield. *Fourier analysis of time series*. John Wiley, 2004
- Press, Teukolsky, Vetterling, Flannery. *Numerical Recipes in C.* Cambridge University Press, 1992

Time-resolved methods (wavelets & matching pursuit):

- G. Foster (1996), AJ 112, 1709
- Hastie, Tibshirani and Friedman. *Elements of Statistical Learning*, Springer, 2009
- S. Mallat, Z. Zhang (1992). IEEE Transactions on Signal Processing 41(12), 3397

Stochastic processes:

- S. Karlin & Taylor. *A first course on stochastic processes*.
- J. Honerkamp. *Stochastic dynamical systems.* John Wiley, 1994
- B. Hajek. *Random processes for engineers*. 2015, Cambridge University Press. Online course notes: `http://hajek.ece.illinois.edu/Papers/randomprocJuly14.pdf`
- Andrae, Kim & Bailer-Jones (2013), A&A 554, A137
- Bailer-Jones, C. A. L. 2012, A&A, 546, A89
- C. Rasmussen & C. Williams. *Gaussian Processes for Machine Learning*. 2006, MIT Press. Online: `http://www.gaussianprocess.org/gpml/`

State-space models:

- `http://www-stat.wharton.upenn.edu/~stine/stat910/lectures/14_state_space.pdf`

Thank you,
and good luck!