에너지인력양성사업

Cautionary remarks on the correlation analysis of non-Gaussian self-similar time series

Sung Yong Kim

Division of Ocean Systems Engineering

School of Mechanical, Aerospace & Systems Engineering

Korea Advanced Institute of Science and Technology (KAIST)

syongkim@kaist.ac.kr

http://imcool.kaist.ac.kr

Questions

- When we describe the variability of the system or characterize the system, we may use decorrelation scales. Could we apply this approach to any types of data?

- Can we apply the correlation analysis to self-similar time/spatial series (or non-Gaussian variables)?

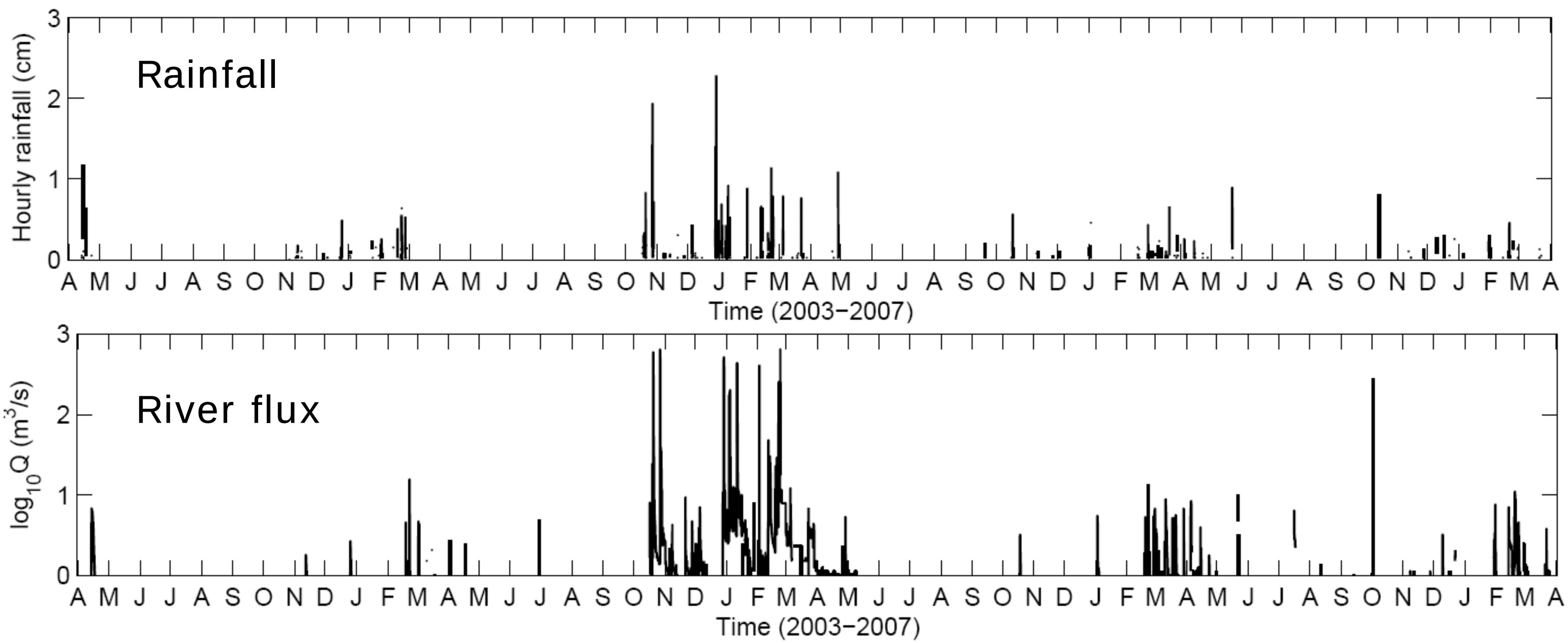
- Any other reasonable approaches to quantify the decorrelation scales?

Outline

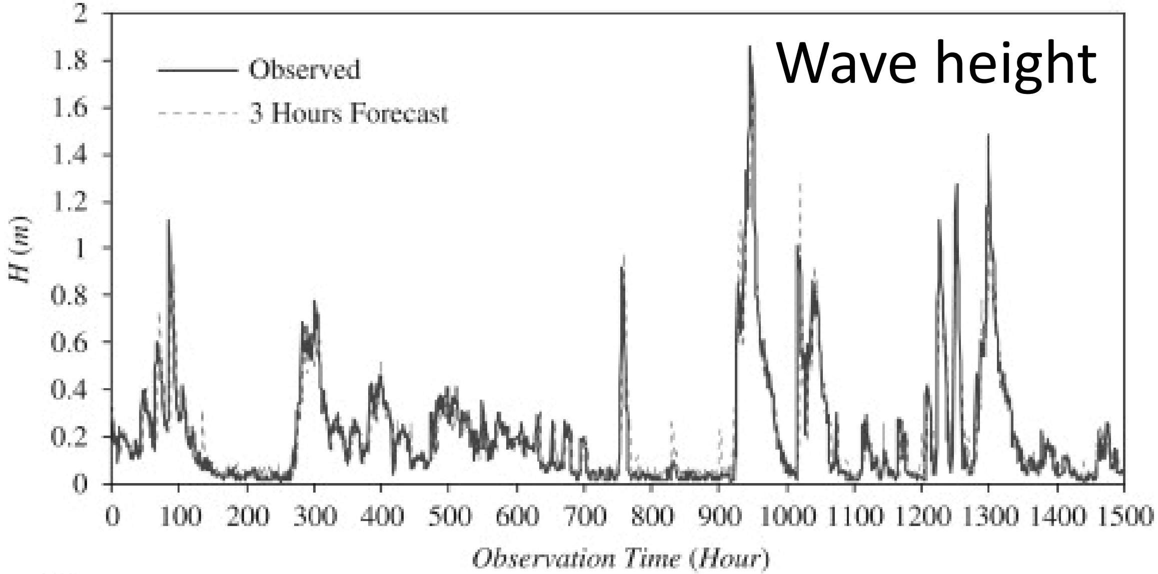
- Self-similar time series and correlation analysis
  - Examples of self-similar time series
  - Definition of correlation analysis
- Investigation with synthetic datasets
  - Generation of self-similar time series with exponential, Gaussian, and linearly decay patterns (single-sided and double-sided pulses)
  - Cross-correlation analysis to quantify the decay scales
  - Another approaches?
- Summary

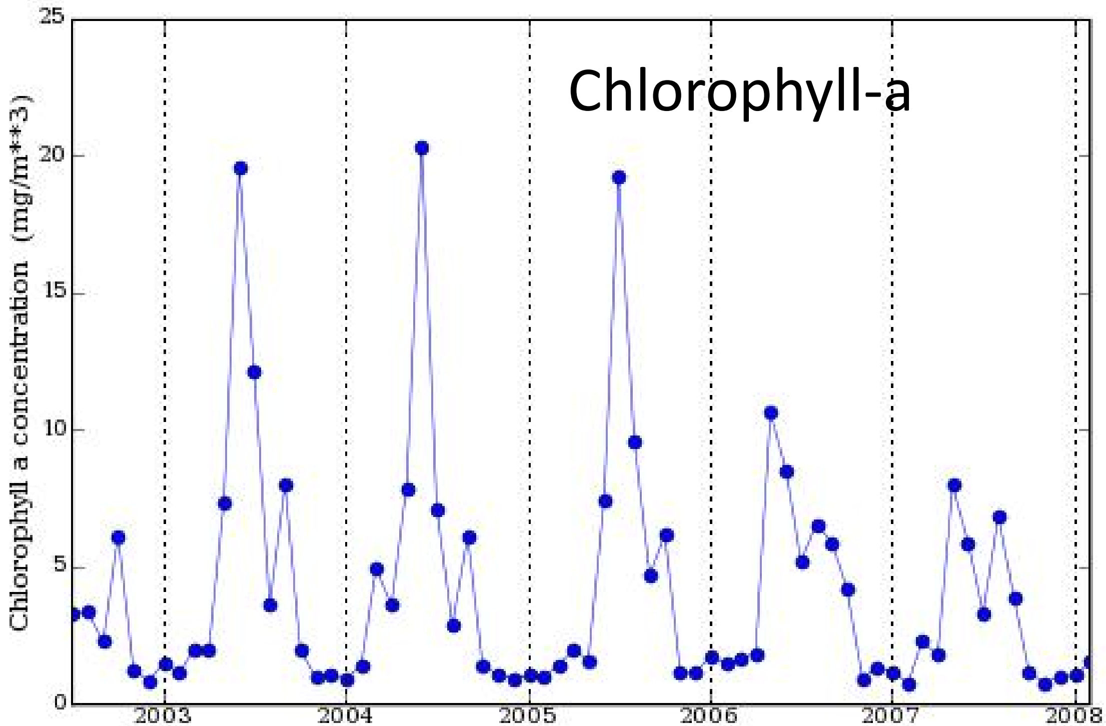
Self-similar time series?

- Continuous time series with similar shapes of disturbances, pluses, or amplitudes
- May be governed by non-Gaussian statistics
  - e.g., River flows, rain fall, wind speed, wave height, concentration of Chlorophyll, and inertial amplitudes

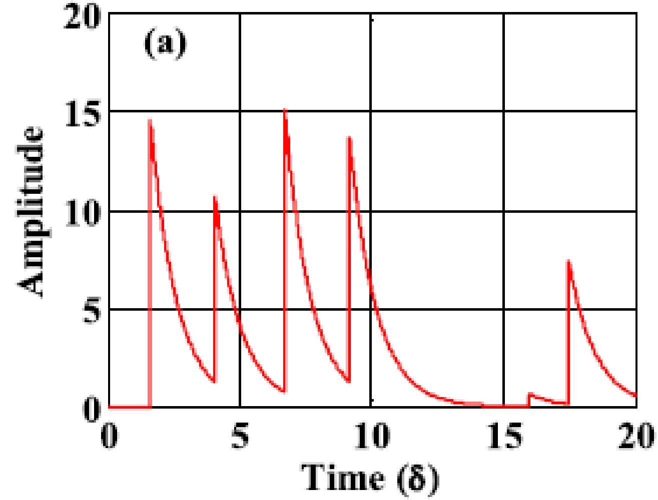


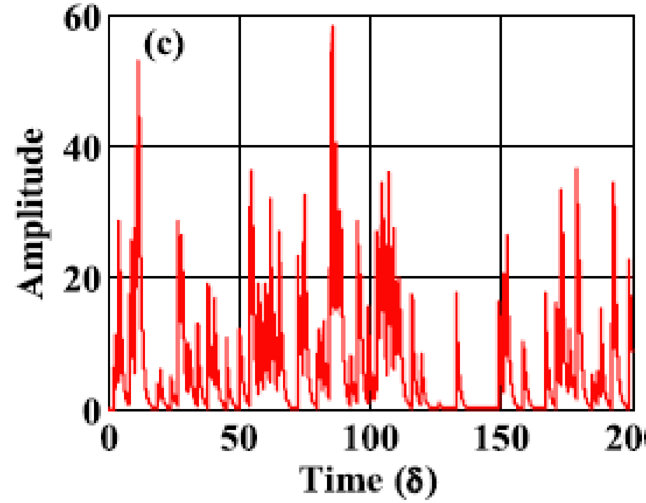
Examples of self-similar time series





Inertial-amplitudes





Decorrelation scales (or e-folding scales)

- Provide a convenient description on the structure of variables and a unique value for the system identification.
  - e.g.) decorrelation time scale is 3 days….
    E-folding length scales are 4 km and 10 km in the x- and y-directions.

- System design and analysis, sampling techniques and optimization

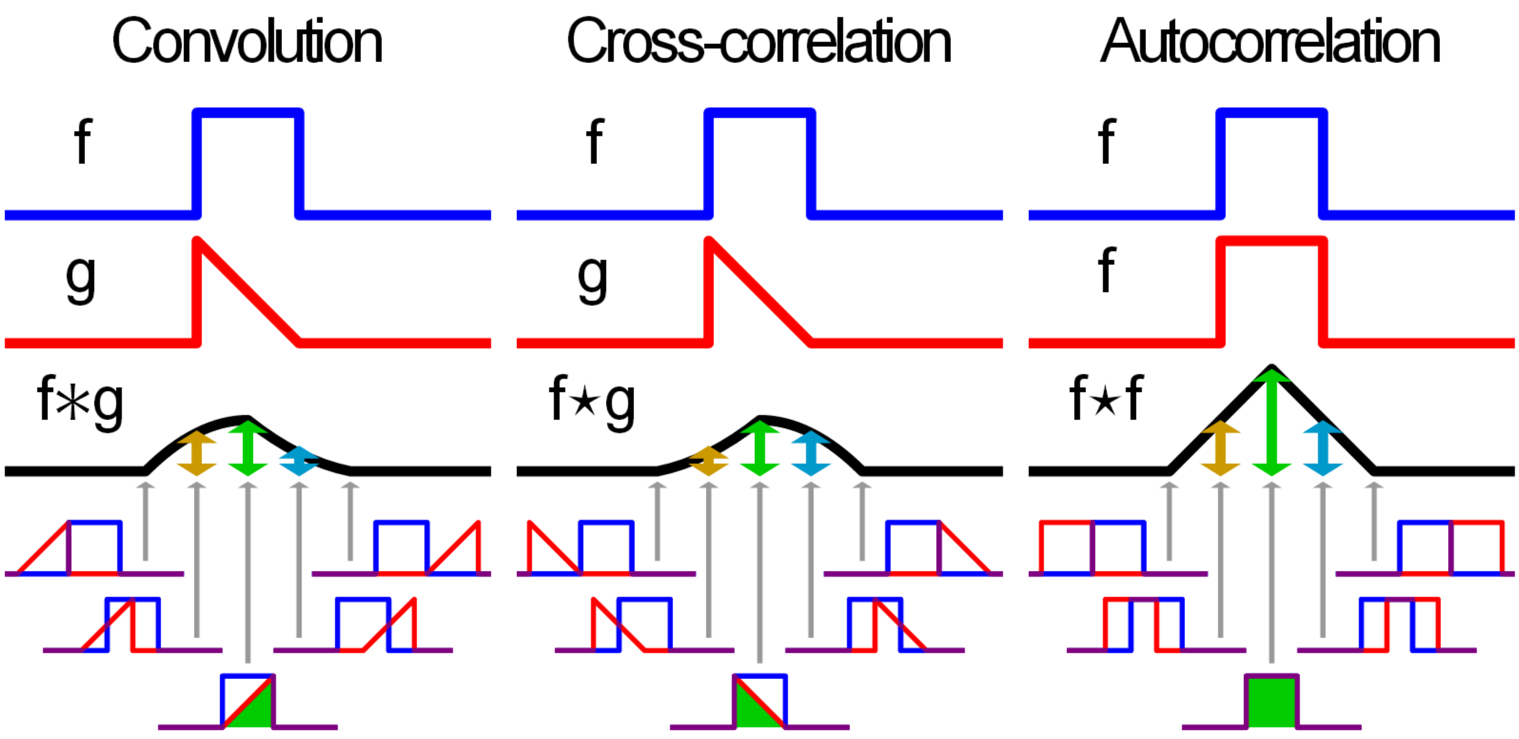
- How to quantify these scales?

Cross-correlations

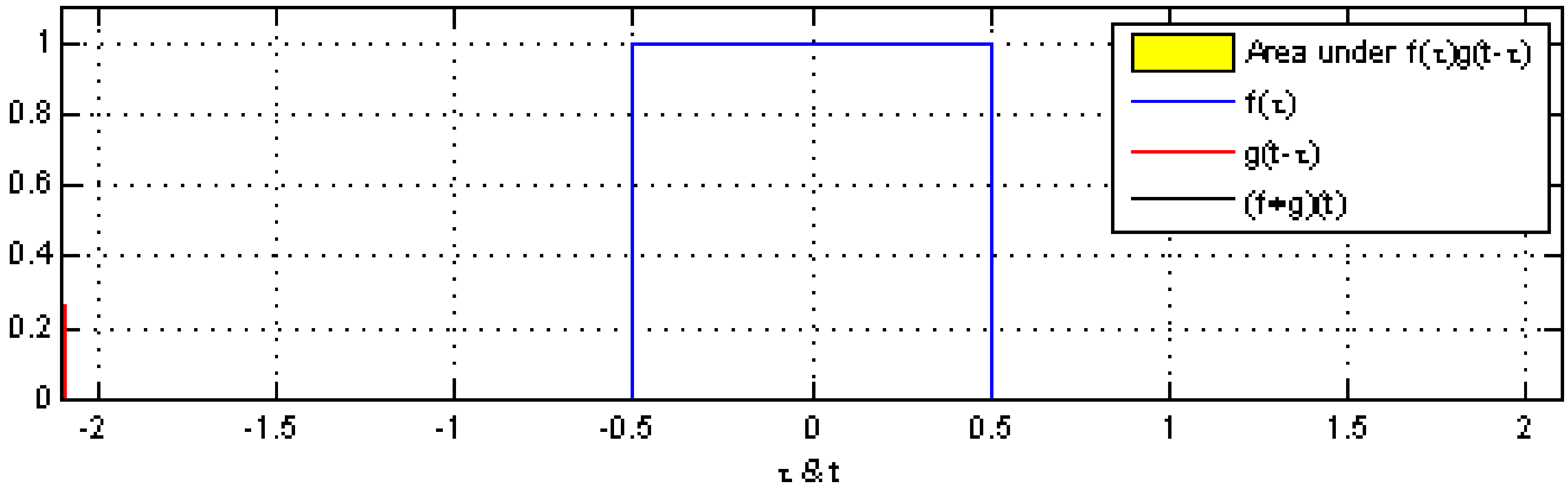
- For the finite and evenly spaced data [d(t)]:

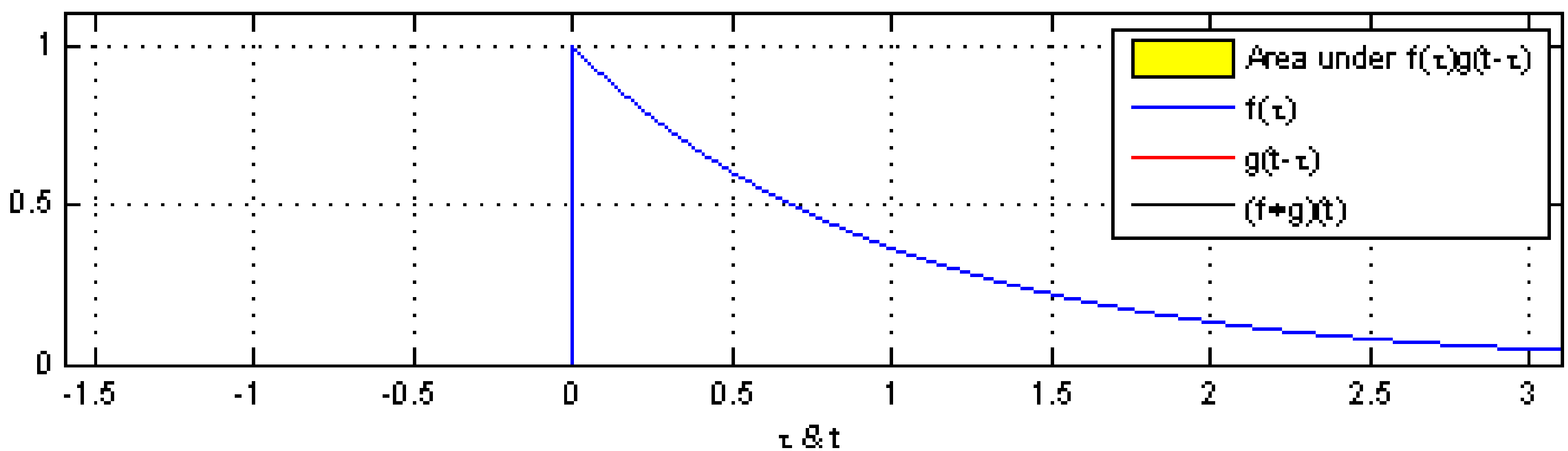
$$\rho(n\Delta t) = \frac{c(n\Delta t)}{\sqrt{\langle d^2(t)\rangle}\sqrt{\langle d^2(t+n\Delta t)\rangle}},$$

$$c(n\Delta t) = \langle d(t)d(t+n\Delta t)^{\dagger}\rangle = \frac{d(t)d(t+n\Delta t)^{\dagger}}{N}$$



Cross-correlation (vs. convolution)





Issues of correlation analysis of self-similar time series

- However, the auto-correlation of self-similar times series produces spurious structure which is not relevant the raw time series.

- Let’s take a look at what we examined with synthetic data sets:

Time series with multiple pulses

$$d^e(t) = \sum_{n=1}^{N} a_n^e b_n^e(t) = \sum_{n=1}^{N} a_n^e \exp\left[-\frac{|t-t_n|}{\lambda_n}\right],$$

$$d^g(t) = \sum_{n=1}^{N} a_n^g b_n^g(t) = \sum_{n=1}^{N} a_n^g \exp\left[-\frac{(t-t_n)^2}{\lambda_n^2}\right],$$

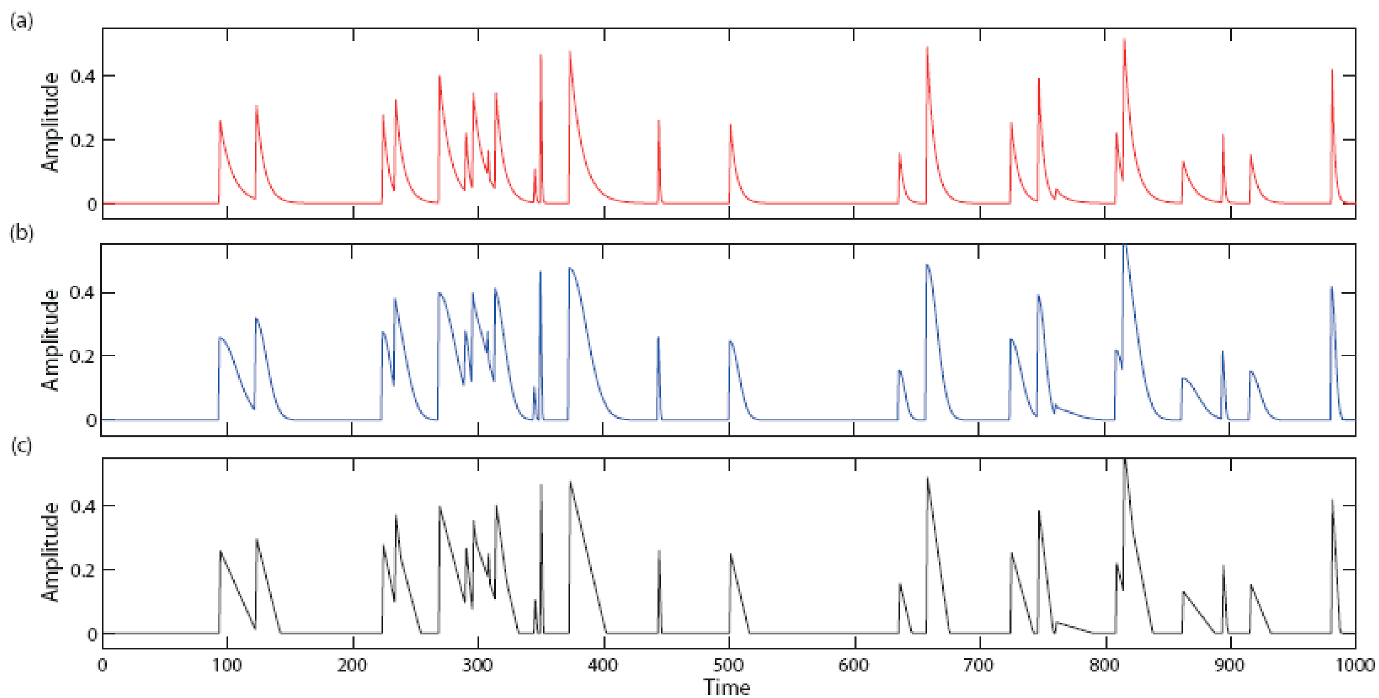
$$d^l(t) = \sum_{n=1}^{N} a_n^l b_n^l(t) = \sum_{n=1}^{N} a_n^l \left[\frac{t-t_n}{t_n-\beta}+1\right],$$

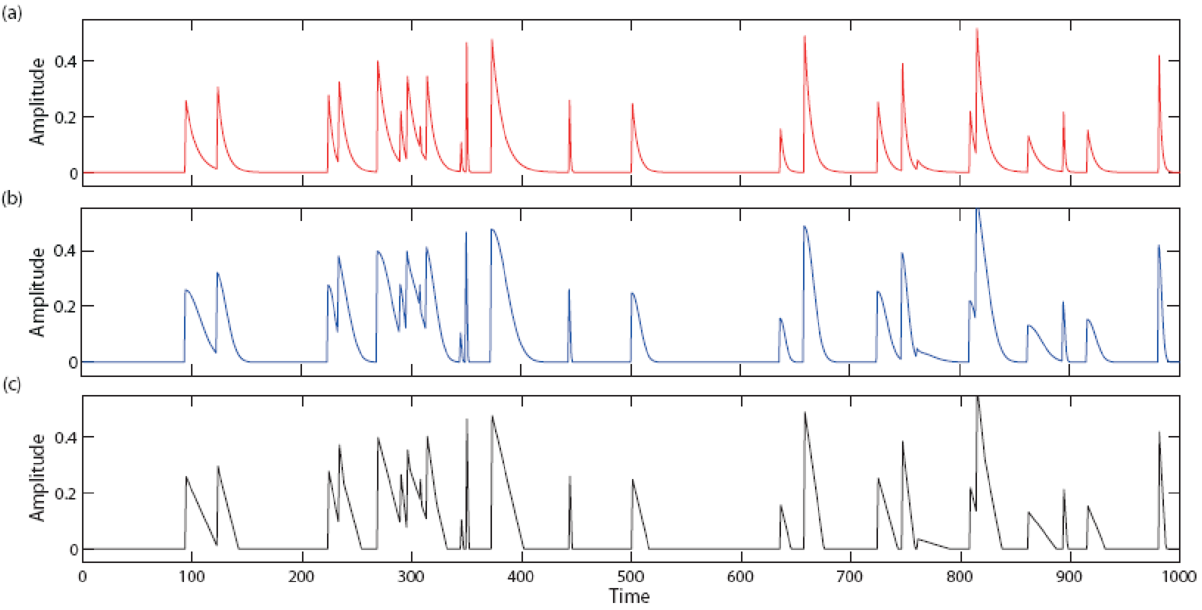
- a_n: Amplitude of each pulse
- b_n: Shape of decay of each pulse.

Exponential

Gaussian

Linear



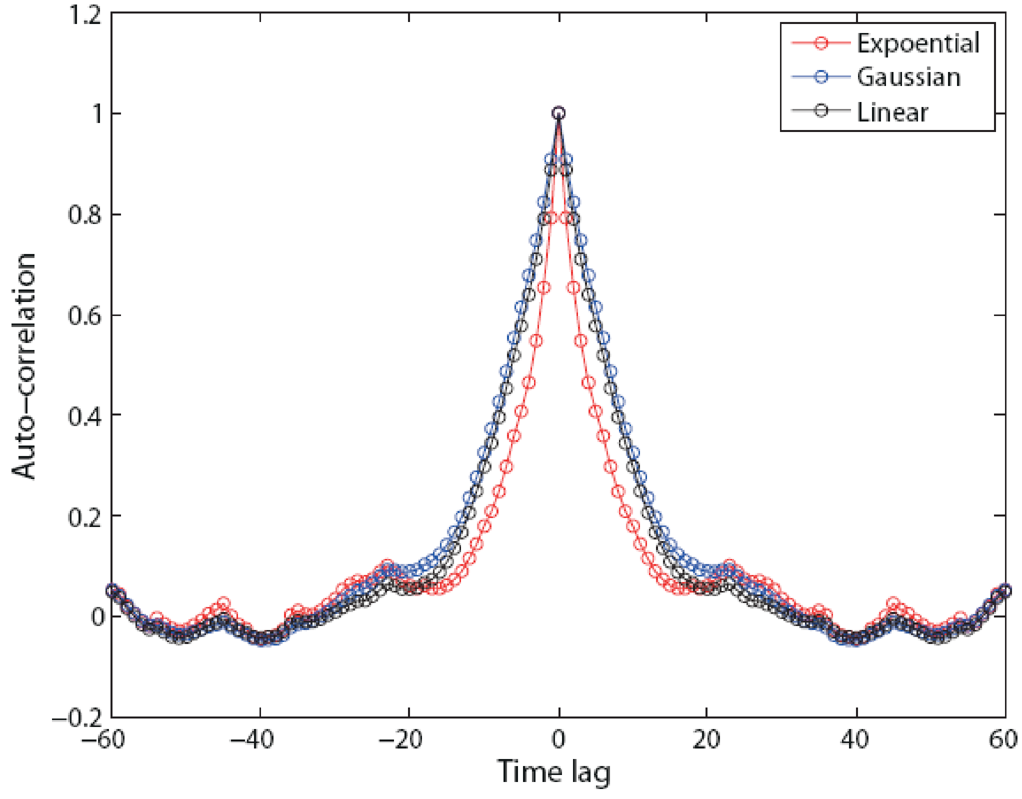


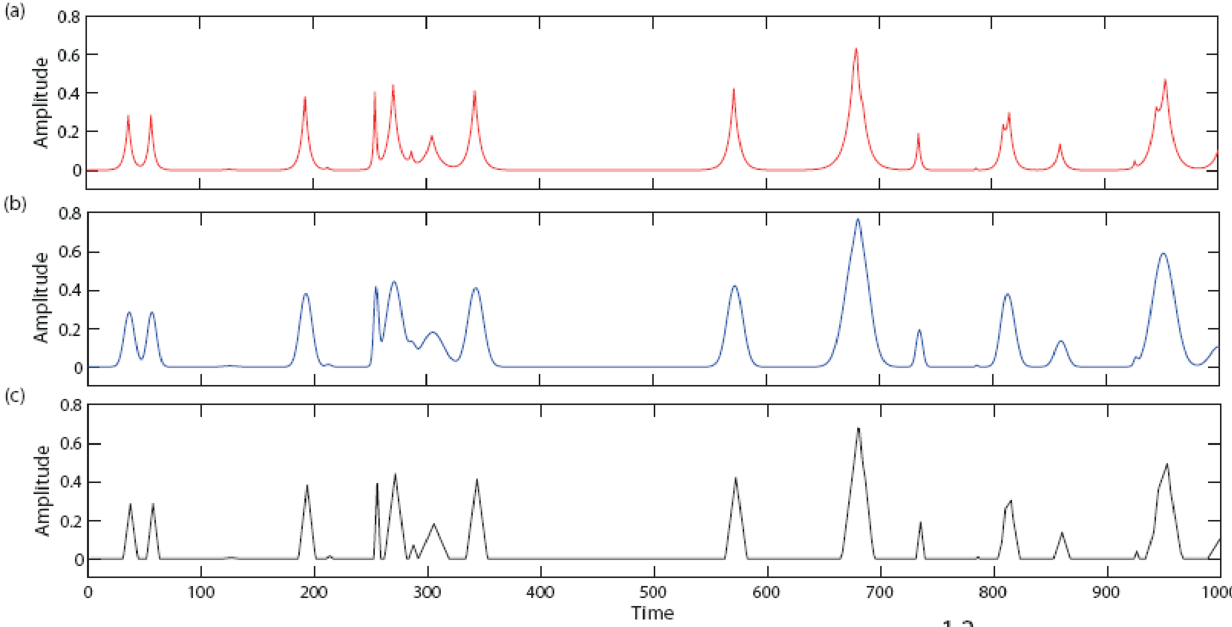
Exponential

Gaussian

Linear

Single-side pulses



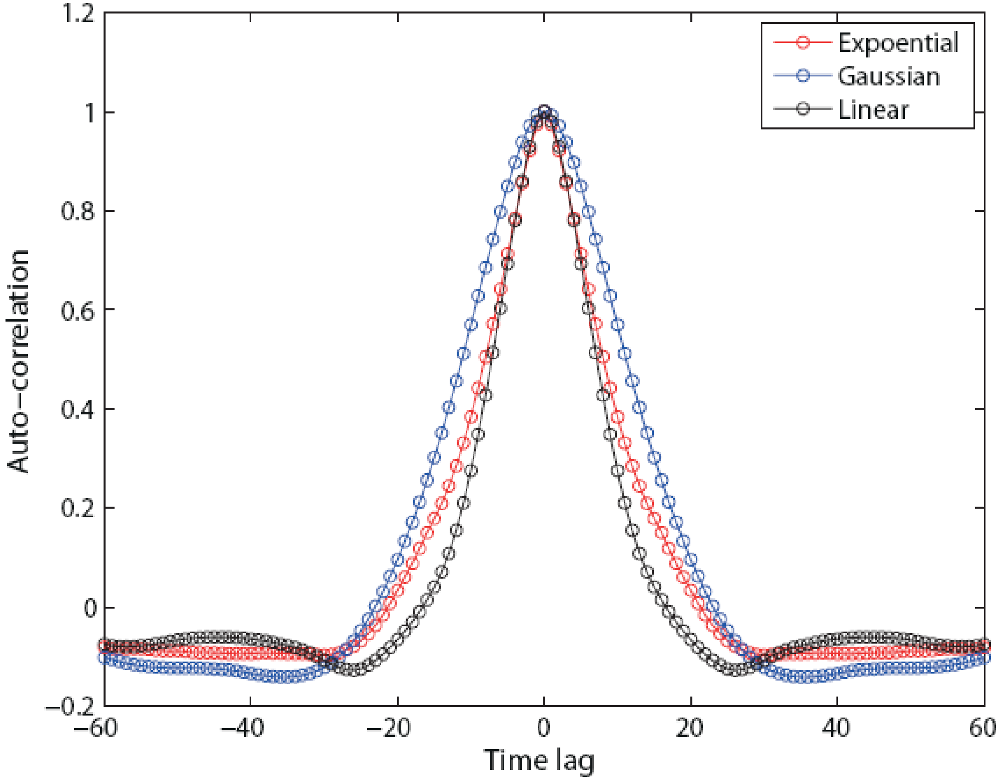


Exponential

Gaussian

Linear

Two-side pulses



Another approach to estimate decorrelation scales?

- Direct correlation analysis can mislead the estimate of decorrelaiton scales.
- Decorrelation scales of Individual pulses can be averaged or their statistics can identify the system.
- Dynamical constraints can be used.
  - e.g., near-inertial amplitudes can be addressed with the response function estimated from observations of wind stress and currents instead of conducting the correlation analysis on the (self-similar) time series of amplitudes of near-inertial currents.

Summary

- Cross-correlation analysis of self-similar time series may generate the spurious results in the estimate of decorrelation scales because convolving two time series does not guarantee their shape based on the original datasets.

- Using cross-correlation analysis the time lag having a maximum correlation can be found.

- Dynamical data analysis or composite mean of individual pulses in the self-similar time series can be used to identify the system or quantify the decorrelation scales.