Signal Processing for Data Perception

Yetmen Wang
AnCAD, Inc.

June 2nd , 2009

Contents

- Composition of signal: trend, periodical, jump/discontinuity, and stochastic
- Definition of trend signal and its removal
- Identification removing of discontinuity from signal
- High resolution spectrum analysis
- Example 1: spectrum analysis of signal of ground water level
- Example 2: time-frequency analysis for strain meter signal

AnCAD

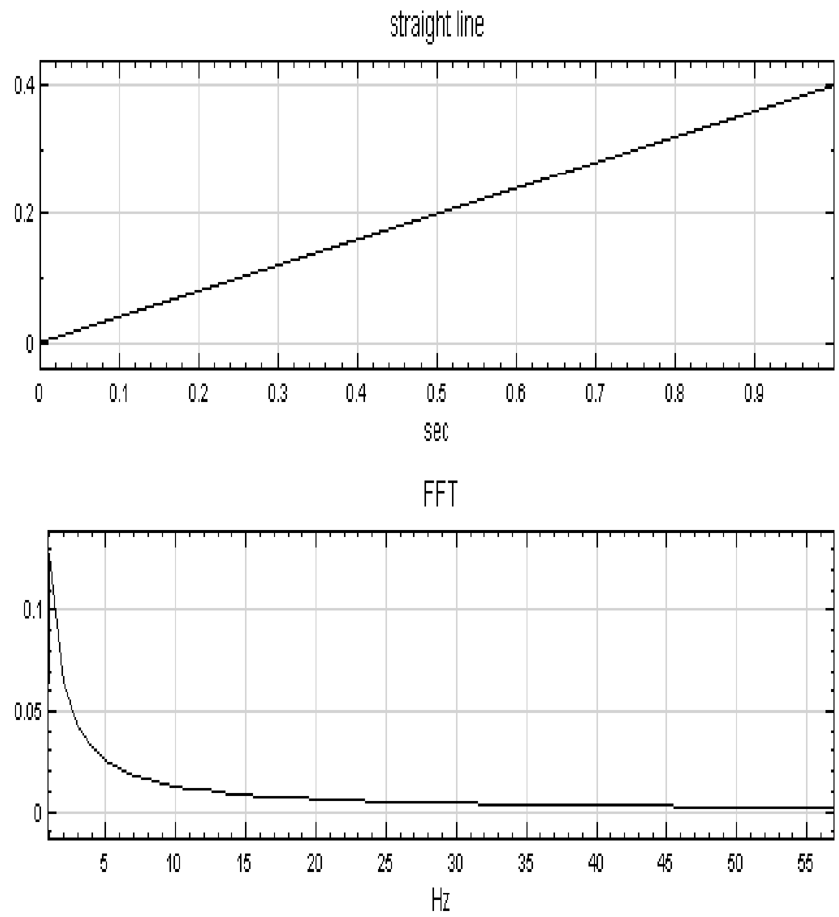
What do we want from frequency? Data perception

- Signal is normally composed of four parts: trend, periodical, discontinuity(jump, end effect), and stochastic.
- Trend is best perceived in time domain.
- Discontinuity which might result in infinitely many components in Fourier analysis is better eliminated before processing.
- Stochastic part currently is not involved in this discussion.
- Periodical signal can be well perceived in **time-frequency-energy representation** (e.g. spectrogram, scalogram, Hilbert Spectrum, etc.).
- Without time resolution, **spectrum analysis** provides highest frequency resolution.

Trend Removal and Iterative Gaussian Filter

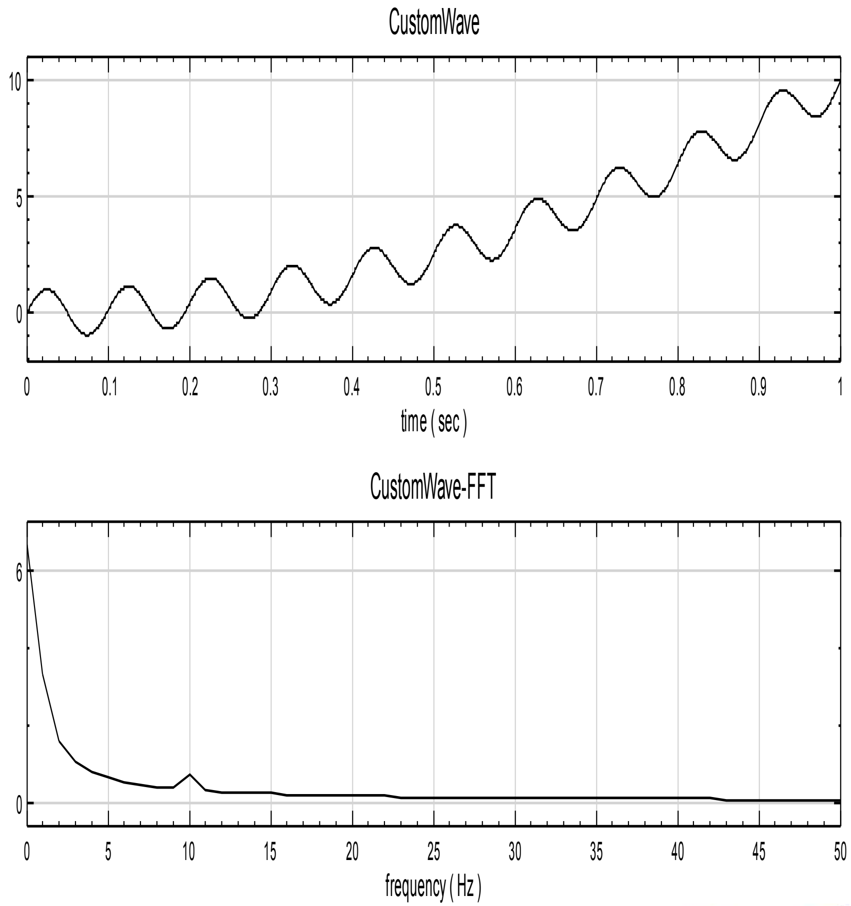
Is “linear” signal periodical?

- Does a straight has a frequency?
- Putting on Fourier glasses, we see so many frequencies from a straight line.
- Intuitively linear signal is not periodical, but composed of many signals of different peroids.

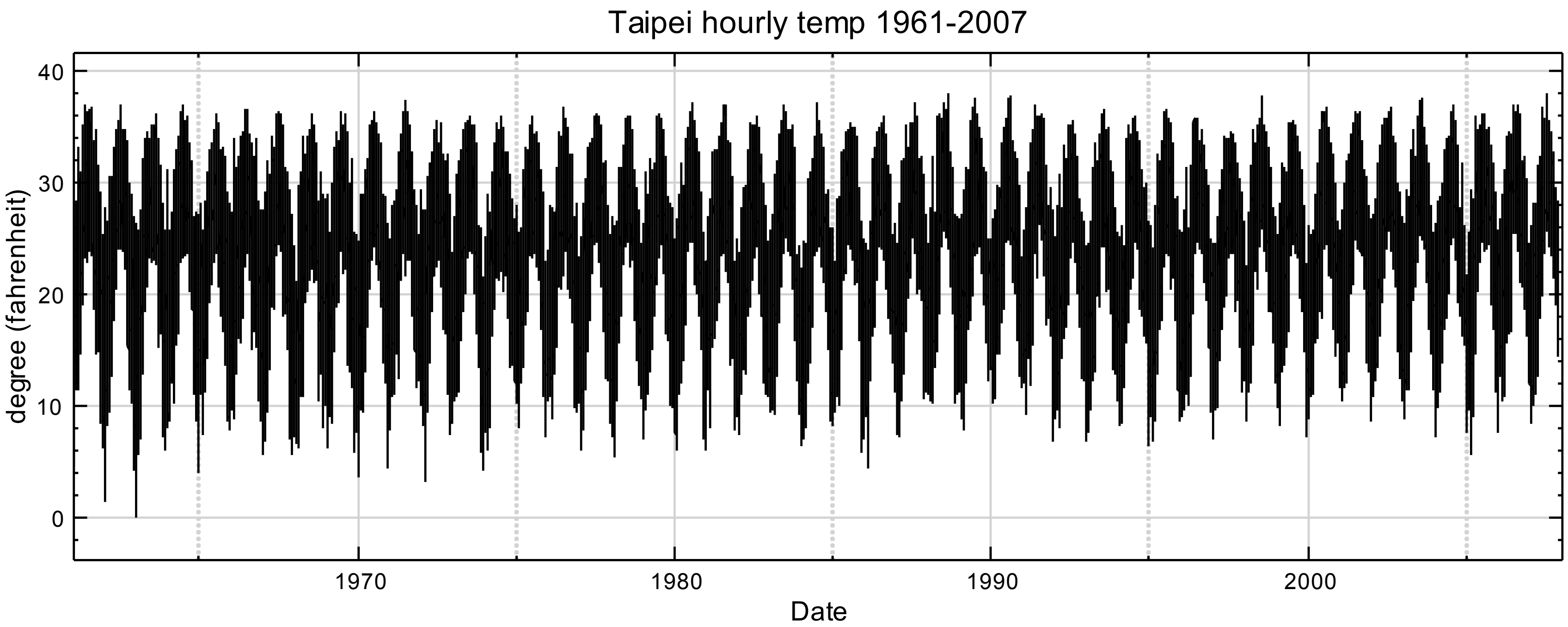


Embedded trend: signal with low frequency noise?

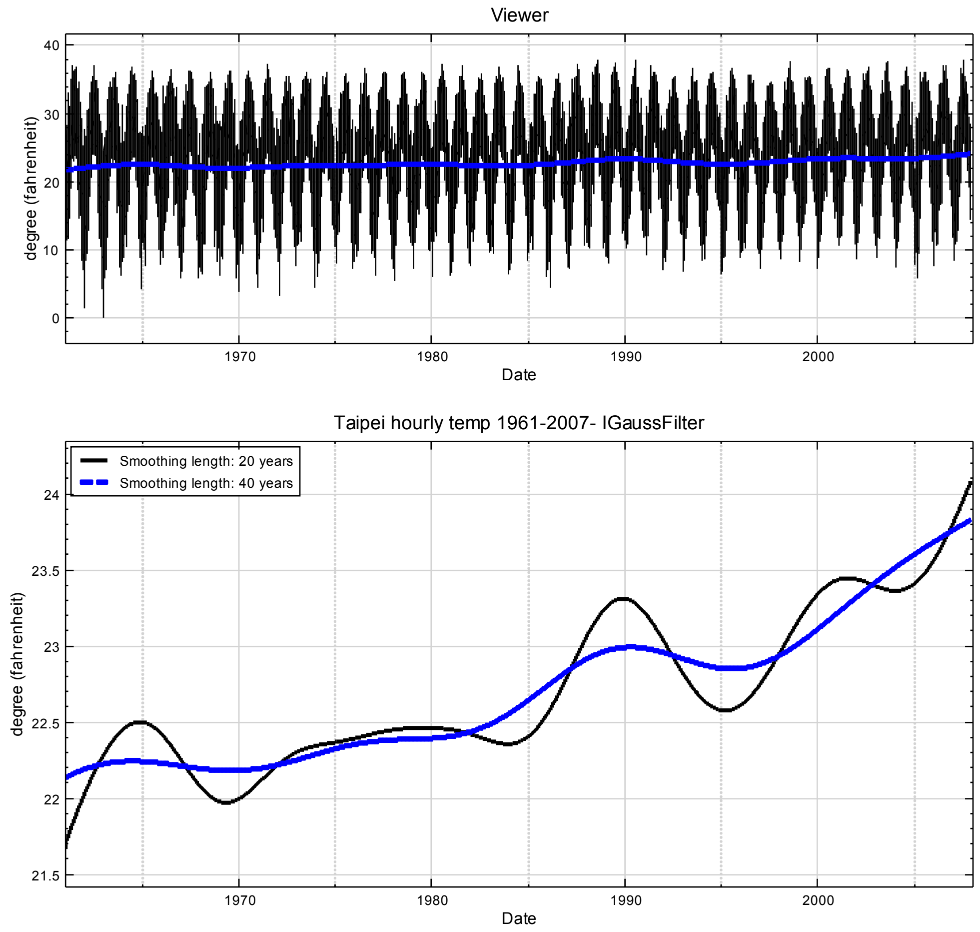
- Trend signal is mistakenly considered to be of low frequency.
- Removal of low frequency part results in high frequency contamination.



Is Taipei getting warmer?

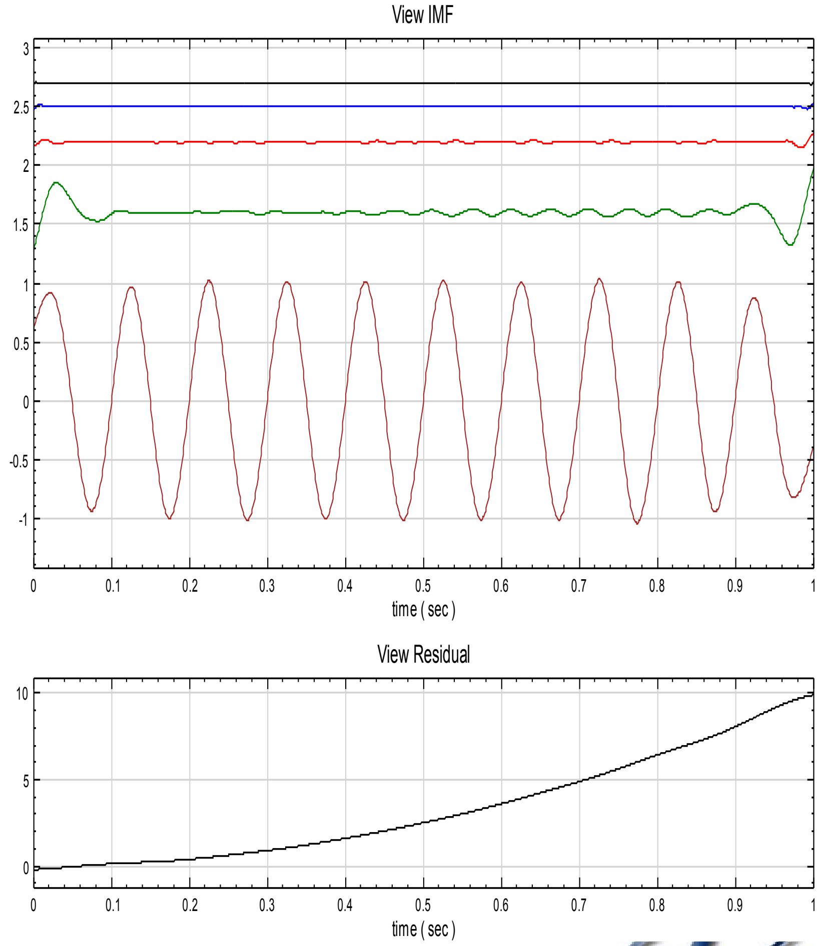


Trend Removal

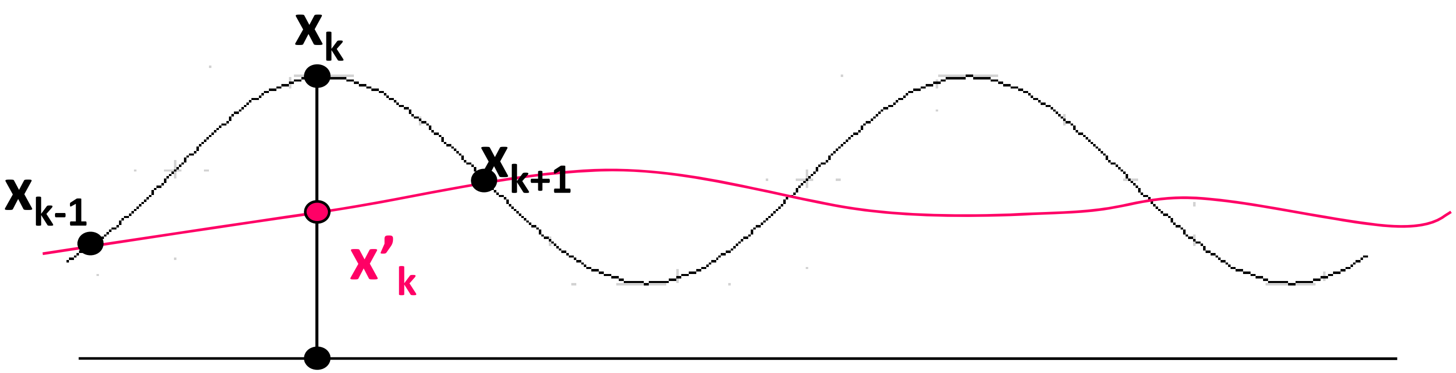


EMD for trend removal

- EMD is also a good way to separate trend signal.



Moving Average



Smooth curve can be obtained by moving average:

$$x'_k = \left(x_{k-1} + x_k + x_{k+1}\right)/\,3$$

In general number of point average can be many with Gaussian weighting.

$$x'_k = A\sum_j x_{k-j}\ \exp\left(-\frac{(k-j)^2}{2\sigma^2}\right)$$

Gaussian Filter and Heat Diffusion

For continuous signal, Gaussian filter is written as

$$u'(t) = A\int u(\tau)\exp(-\frac{(t-\tau)^2}{2\sigma^2})d\tau$$

It is the solution of the following heat equation

$$\frac{\partial u}{\partial t} = \alpha\frac{\partial^2 u}{\partial x^2} + \delta(t)$$

The analogy suggests that diffusion effect results in smoothing of signal (temperature).

Iterative Gaussian Filter

- For continuous signal, its Fourier representation is

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{j\omega t}d\omega$$

Gaussian filter is the weighted average with Gaussian as the weighting function

$$f_L(t) = \frac{1}{c(\sigma)}\int_{-\infty}^{+\infty} e^{-\frac{(t-\tau)^2}{\sigma^2}} f(\tau)d\tau$$

where $$c(\sigma) = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{\sigma^2}}dt = \sigma\sqrt{\pi}$$

Iterative Gaussian Filter

- After m times of iteration, filtered signal is represented as

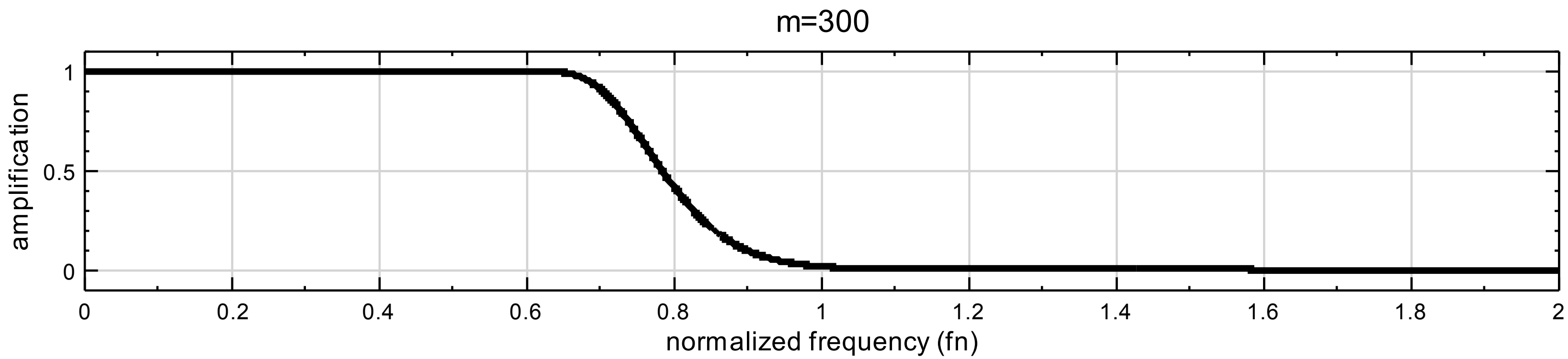
$$f_L^m(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[1-\left(1-e^{-\frac{1}{4}\omega^2\sigma^2}\right)^m\right]F(\omega)e^{j\omega t}\,d\omega$$

T(ω)=

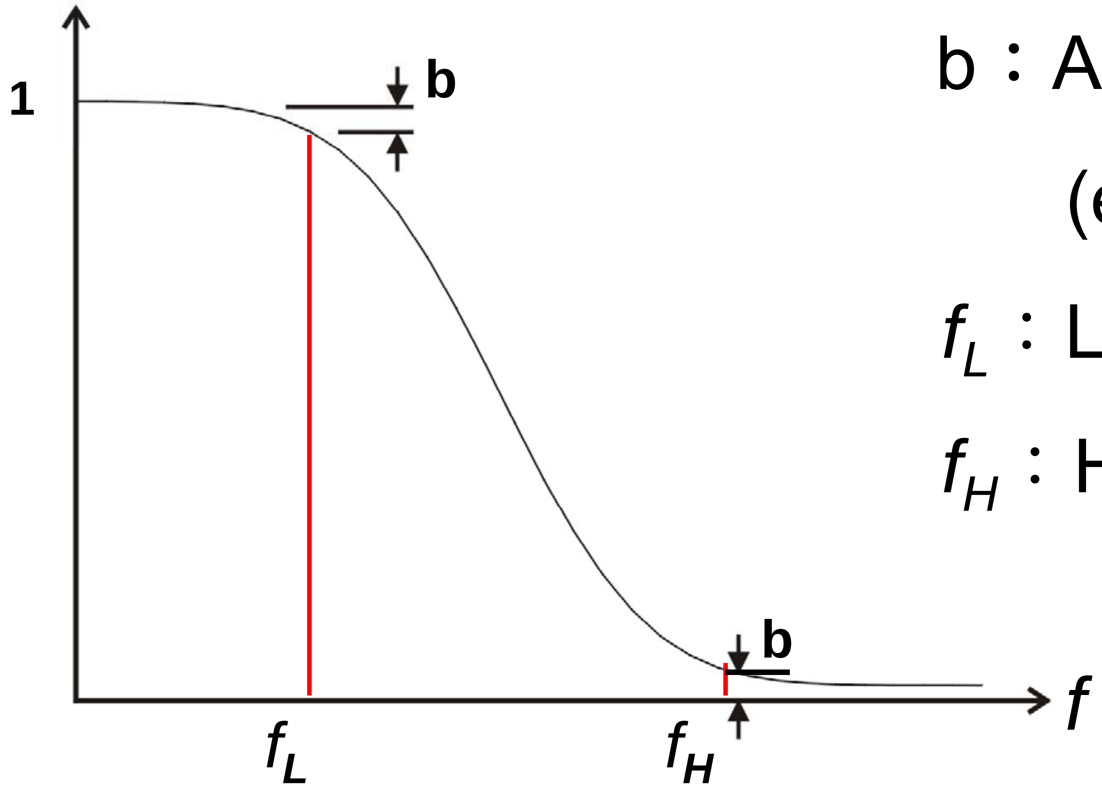
Iterative Gaussian Filter

After m times of iteration, the transfer function becomes:

$$T(\omega) = 1 - \left(1 - e^{-\frac{1}{4}\omega^2\sigma^2}\right)^m$$



Iterative Gaussian Filter



b：Attenuation factor

(e.g. b=0.01)

f_L：Low frequency

f_H：High frequency

Iterative Gaussian Filter

- The Gaussian factor ,σ, and number of iteration, *m*, are determined via solving

$$1-\left[1-e^{-2\pi^2(\sigma f_L)^2}\right]^m = b$$

$$1-\left[1-e^{-2\pi^2(\sigma f_H)^2}\right]^m = 1-b$$

Trend Estimator

T_C (Trend period) ：For low pass filter, period lower than the value will be annihilated. In practice Tc is approximately the length of "bump" (half wavelength) under which will be eliminated.

Cutoff frequency

$$f_C = \frac{2}{T_C}$$

$$f_L = \frac{2}{T_C}$$

$$f_H = 2 f_L$$

Number of iteration $m = 33$

Trend vs. signal of finite duration

- If a signal is absolute integrable in [-∞, +∞], there is no such concept as trend signal. For signal of finite length, we can regard the signal as it is defined in infinite domain. The trend signal which can be imagined in finite domain can be viewed as fraction of very low frequency in infinite domain.
- Discrete signal of finite length with periodical assumption results in discrete sampling in finite spectrum domain. Frequency higher than the maximum representable falls into low frequency due to aliasing mechanism. Whereas frequency lower than representable (trend signal) is also “aliased” into higher frequency in order to have energy conservation. Such mechanism accounts for why trend signal is composed of not only low frequencies.

AnCAD

Gaussian Filter

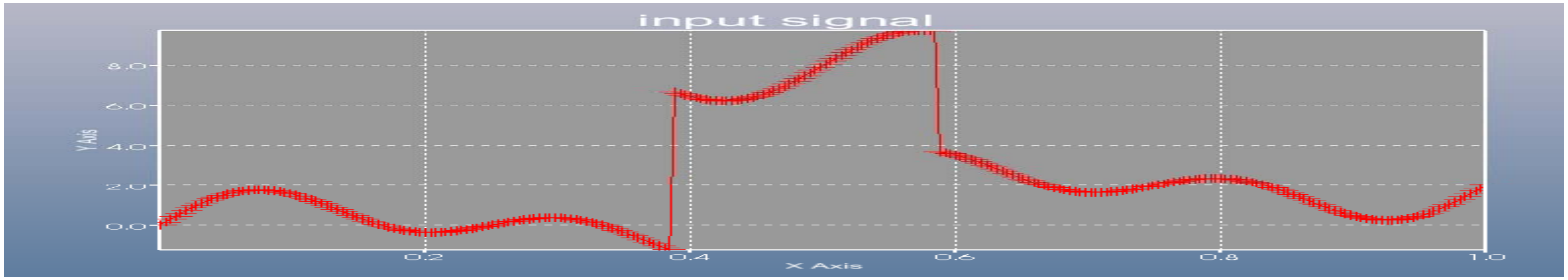
- Iterative moving least square method
- Zero phase error
- Fast algorithm developed by Prof. Cheng of NCKU.
- Continuous signal with trend can be written in the form:

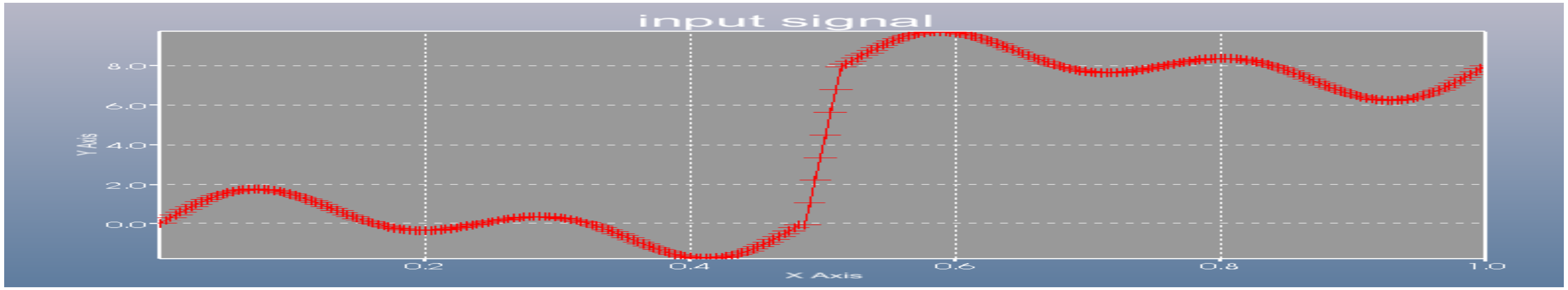
$$x(t) = \sum_{k=1}^{n} a_k \cos \omega_k t + b_k \sin \omega_k t + \sum_{i=0}^{m} \alpha_i t^i$$

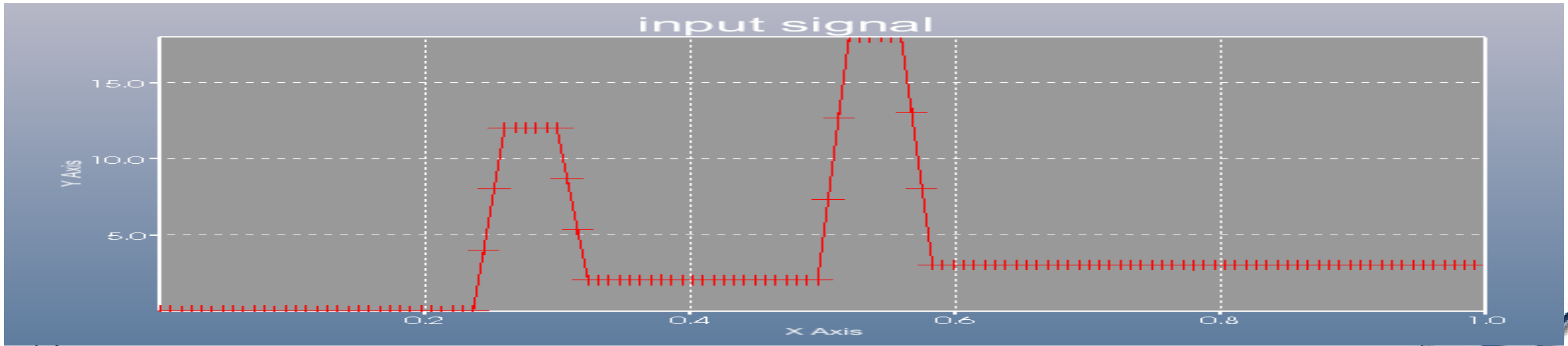
  For each iteration of Gaussian filter, polynomial order m is reduced by 2.
- Gaussian function is the fundamental solution to heat equation. The idea of smoothing is similar to the mechanism of heat spreading, regarding signal as unbalanced temperature distribution to begin with.

On Discontinuity: Bump Removal

Types of discontinuity



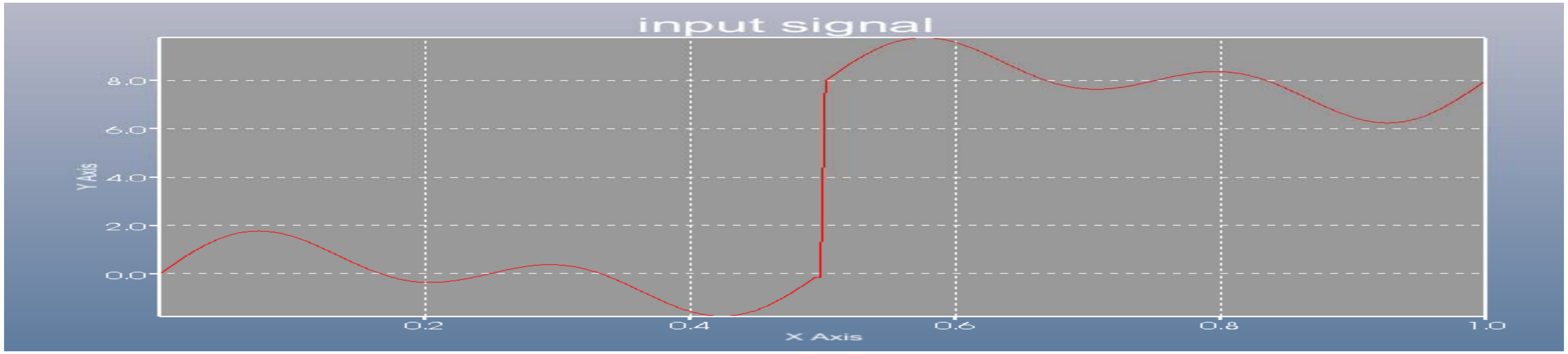


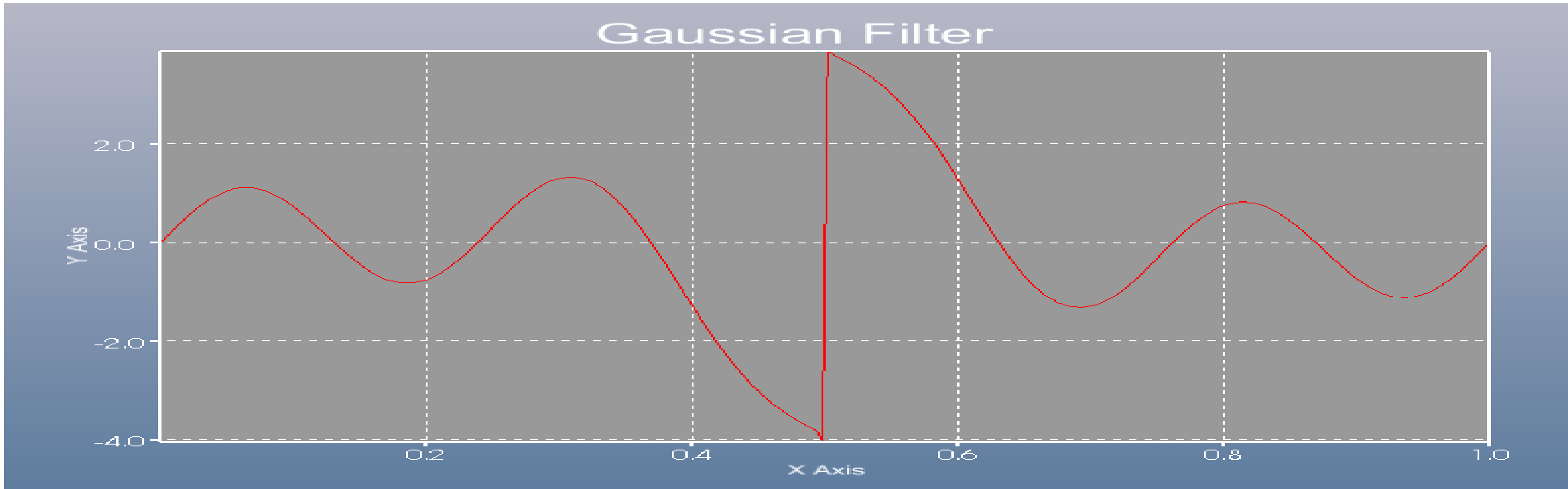


Method (1)

- Gaussian filter :The discontinuity can be removed partially by applying Gaussian Filter, but parts of the unwanted frequency components still exist in the spectrogram.

Example 1:
Apply Gaussian Filter to extract high freq. signal
Discontinuity exists!





Method (2)

- The better way to solve the problem:

- Apply our **new technology – bump removal** to "trim" the data first, then apply Gaussian Filter.

  Then the spectrogram will behave as if the discontinuity does not exist ,and has much better quality in the sense of spectrogram.

Example 2

- Given

  y = sin(4*pi*x) + sin(8*pi*x)

  Goal: filter out high frequency part sin(8*pi*x)

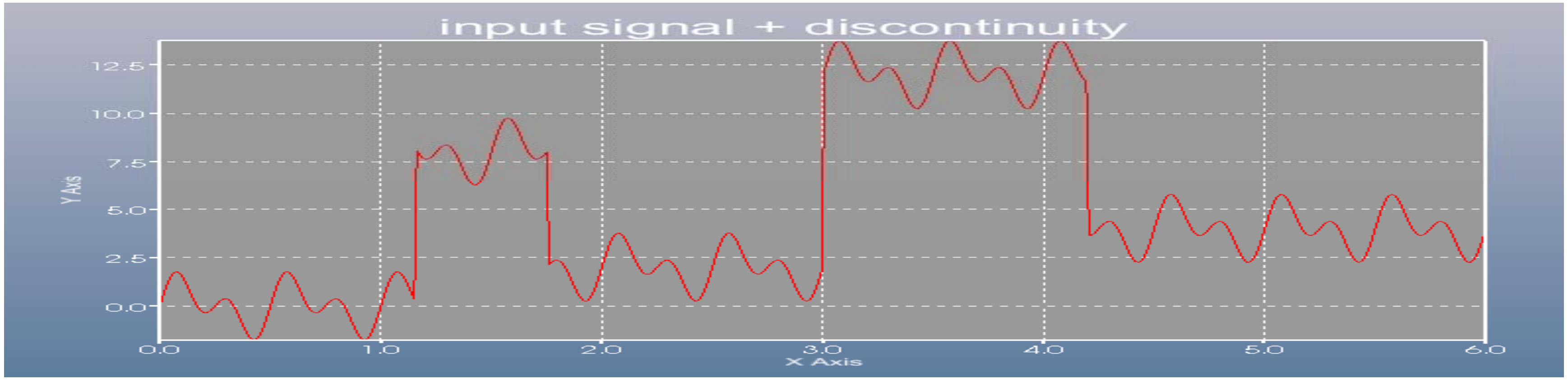
  (or extract sin(4*pi*x))

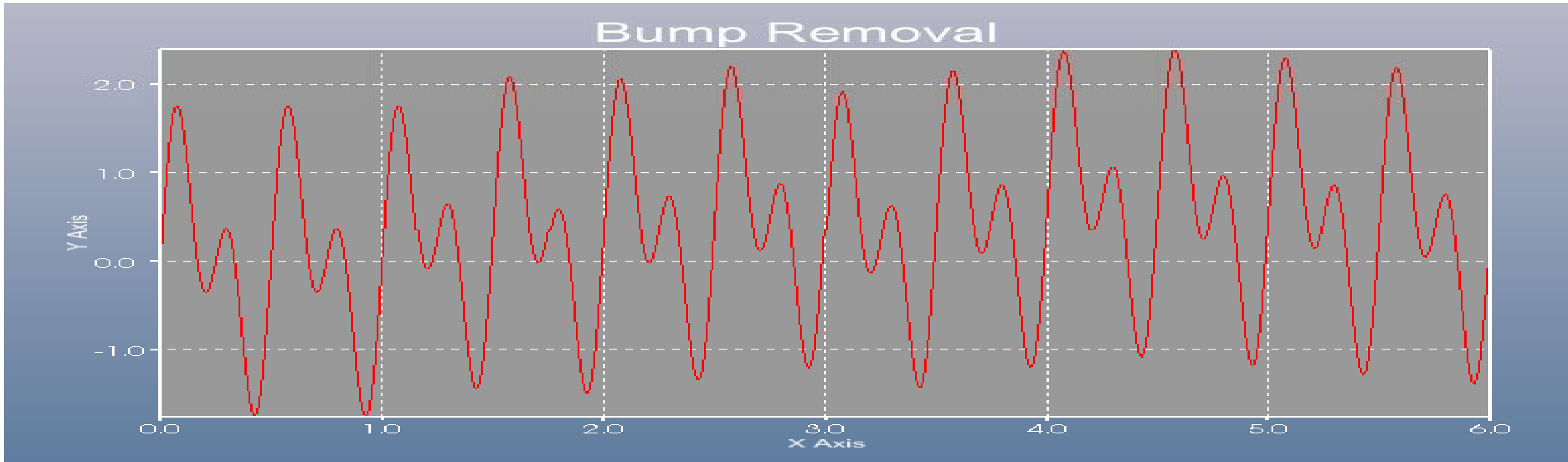
Suppose contaminant in data, input signal becomes

y = sin(4*pi*x) + sin(8*pi*x) + 8   for x1 <=x<= x2

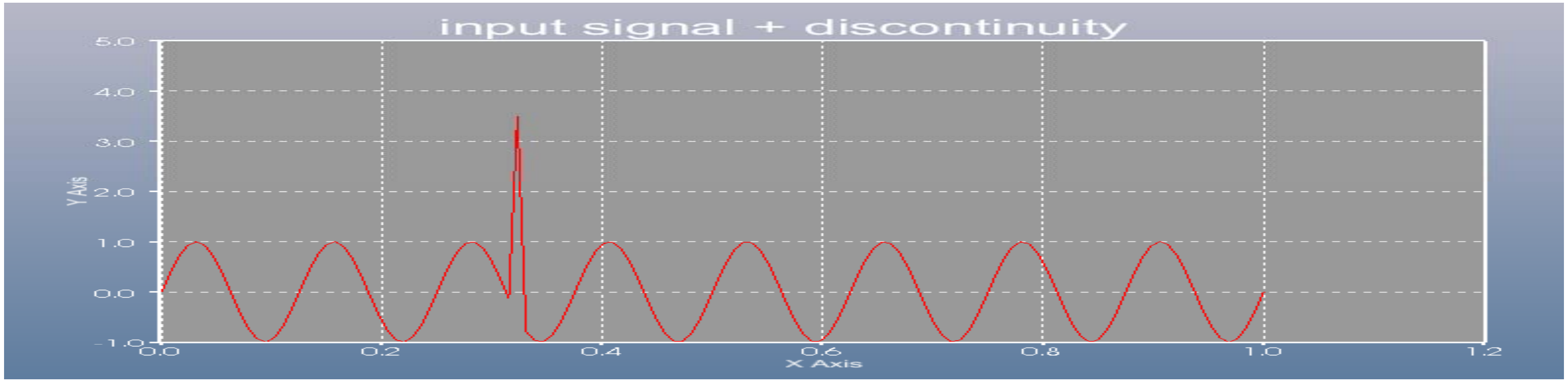
y = sin(4*pi*x) + sin(8*pi*x) + 2   for x2 <=x<= x3

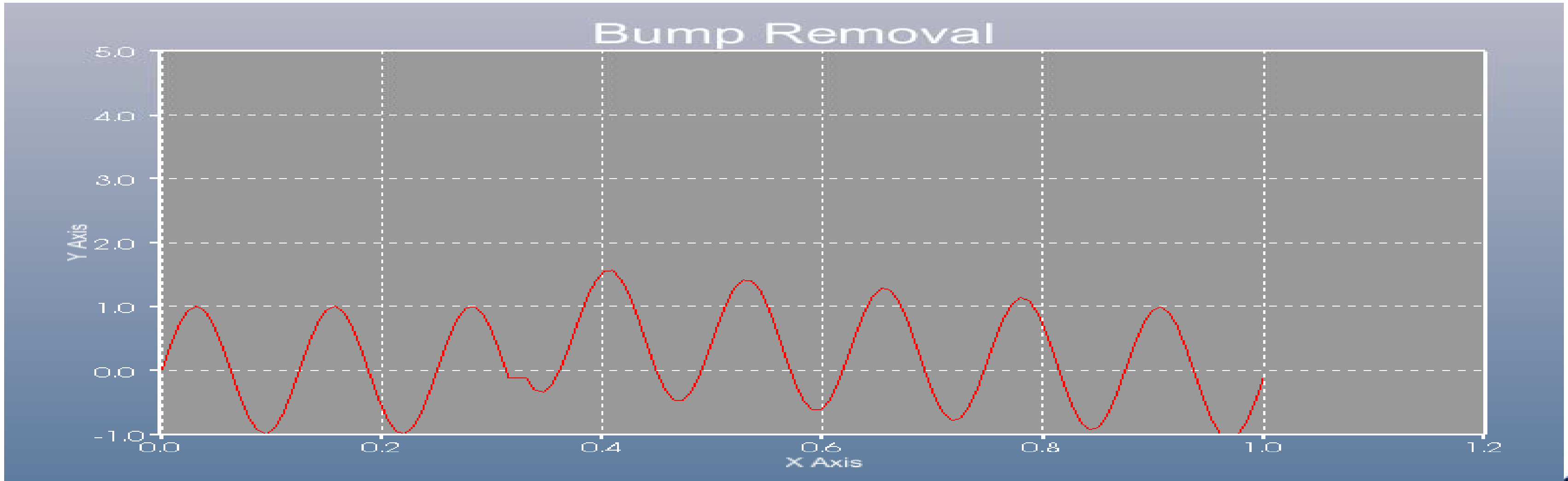
Example 2: Bump Removal





Example 3: spike removal





Example 4

- Given

  y = sin(4*pi*x) + sin(8*pi*x) -- fig. 4.1

  Goal: to Filter out low freq part sin(4*pi*x)

  or to Extract high freq part sin(8*pi*x) -- fig.4.2

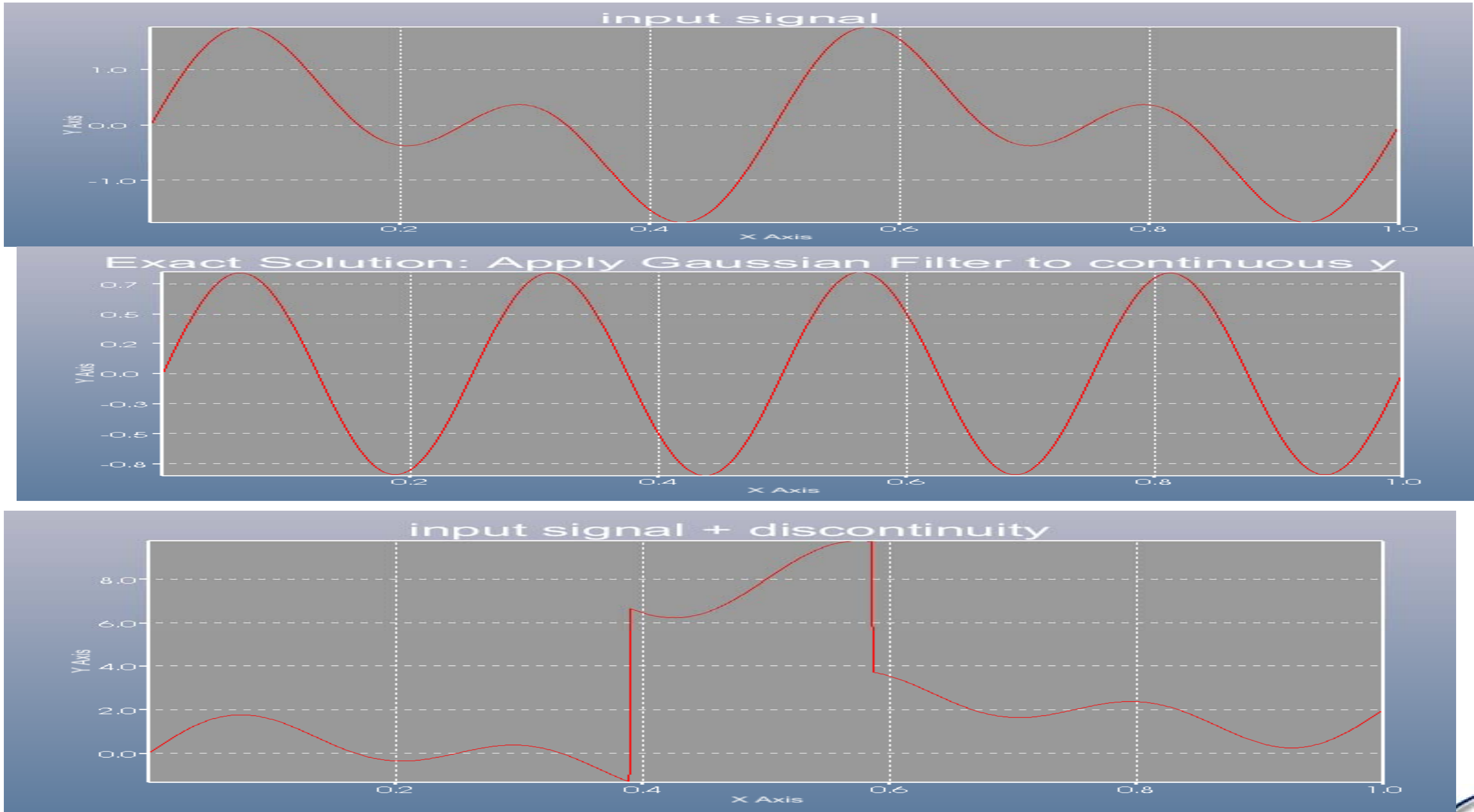
  Suppose

  contaminant in data, input signal becomes -- fig 4.3

  y = sin(4*pi*x) + sin(8*pi*x) + 8 for x1 <=x<= x2

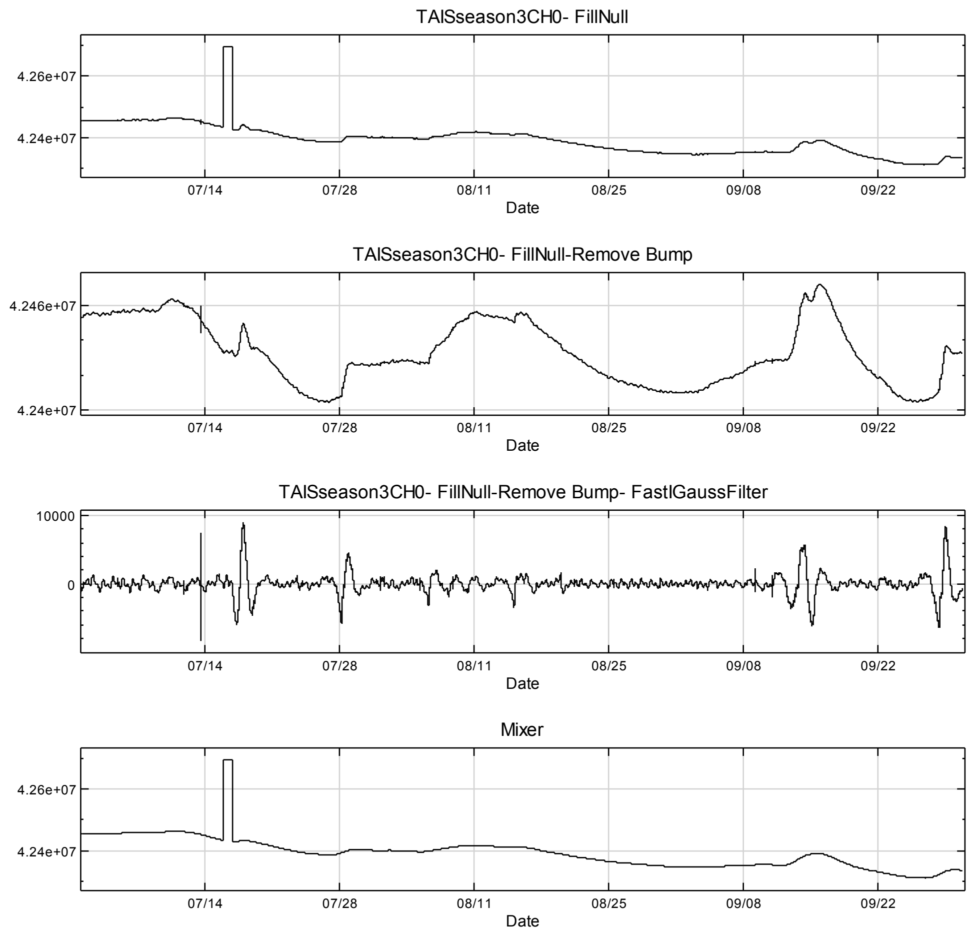
  y = sin(4*pi*x) + sin(8*pi*x) + 2 for x2 <=x<= x3

Example 4



input signal
Y Axis
1.0
0.0
-1.0
0.2
0.4
0.6
0.8
1.0
X Axis
Exact Solution: Apply Gaussian Filter to continuous y
Y Axis
0.7
0.5
0.2
0.0
-0.3
-0.5
-0.8
0.2
0.4
0.6
0.8
1.0
X Axis
input signal + discontinuity
Y Axis
8.0
6.0
4.0
2.0
0.0
0.2
0.4
0.6
0.8
1.0
X Axis

Removal Process for strain meter signal* (TAISeason3Ch0)



Raw Data

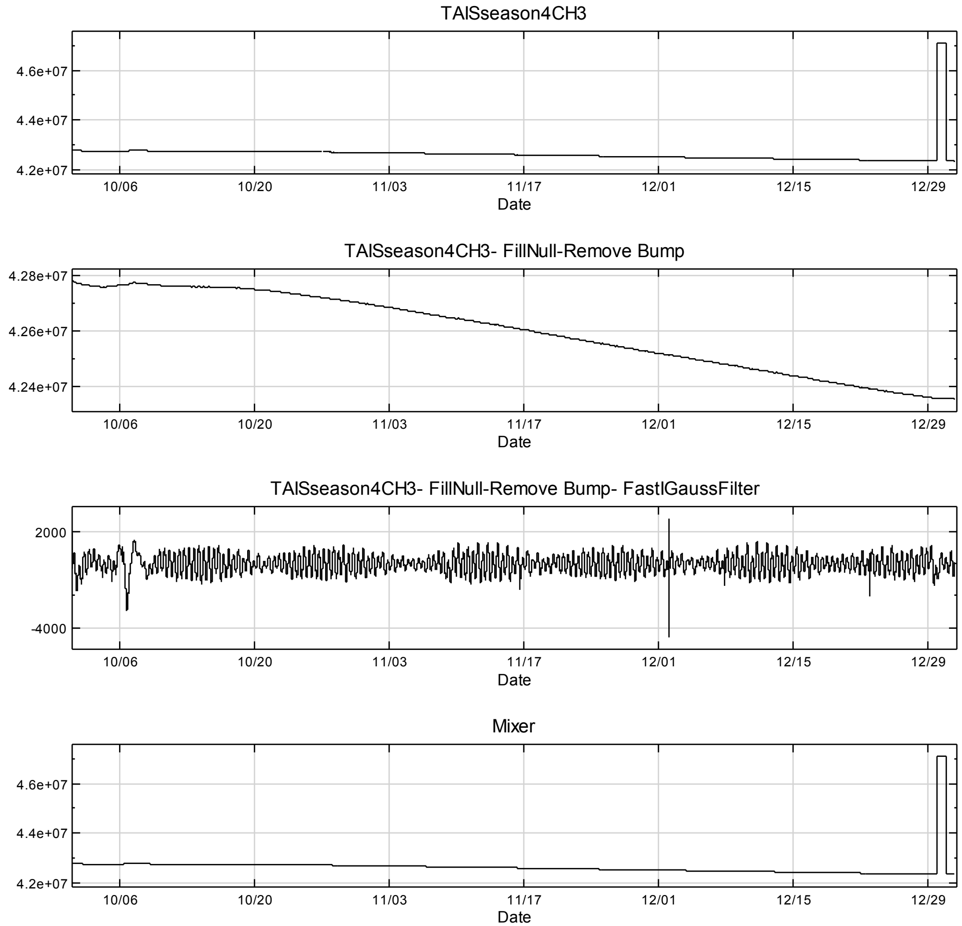
Bump Removal

Trend Removal

Trend + Discontinuity

*資料來源:地調所

Removal Process for strain meter signal* (TAISeason4Ch3)



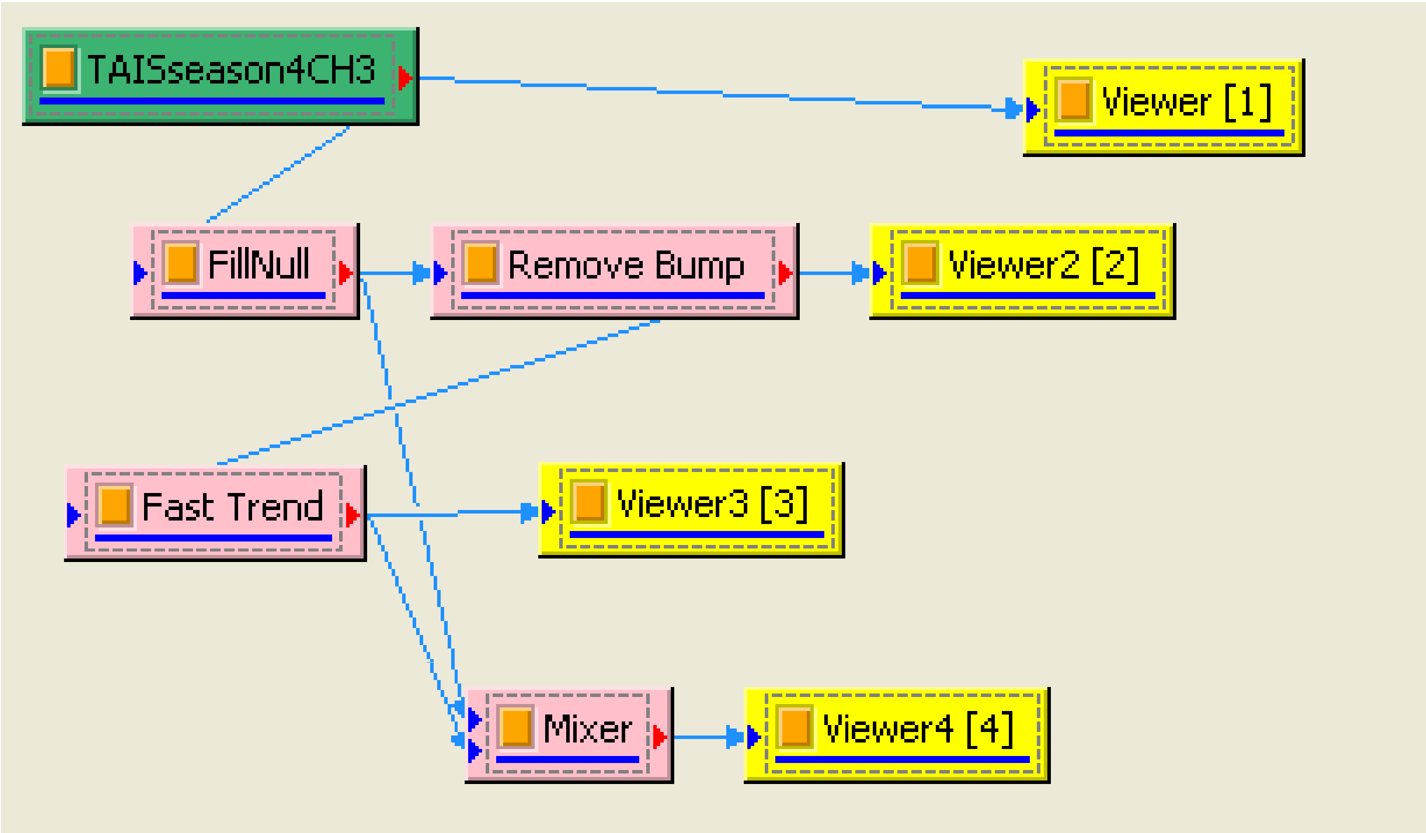
Raw Data

Bump Removal

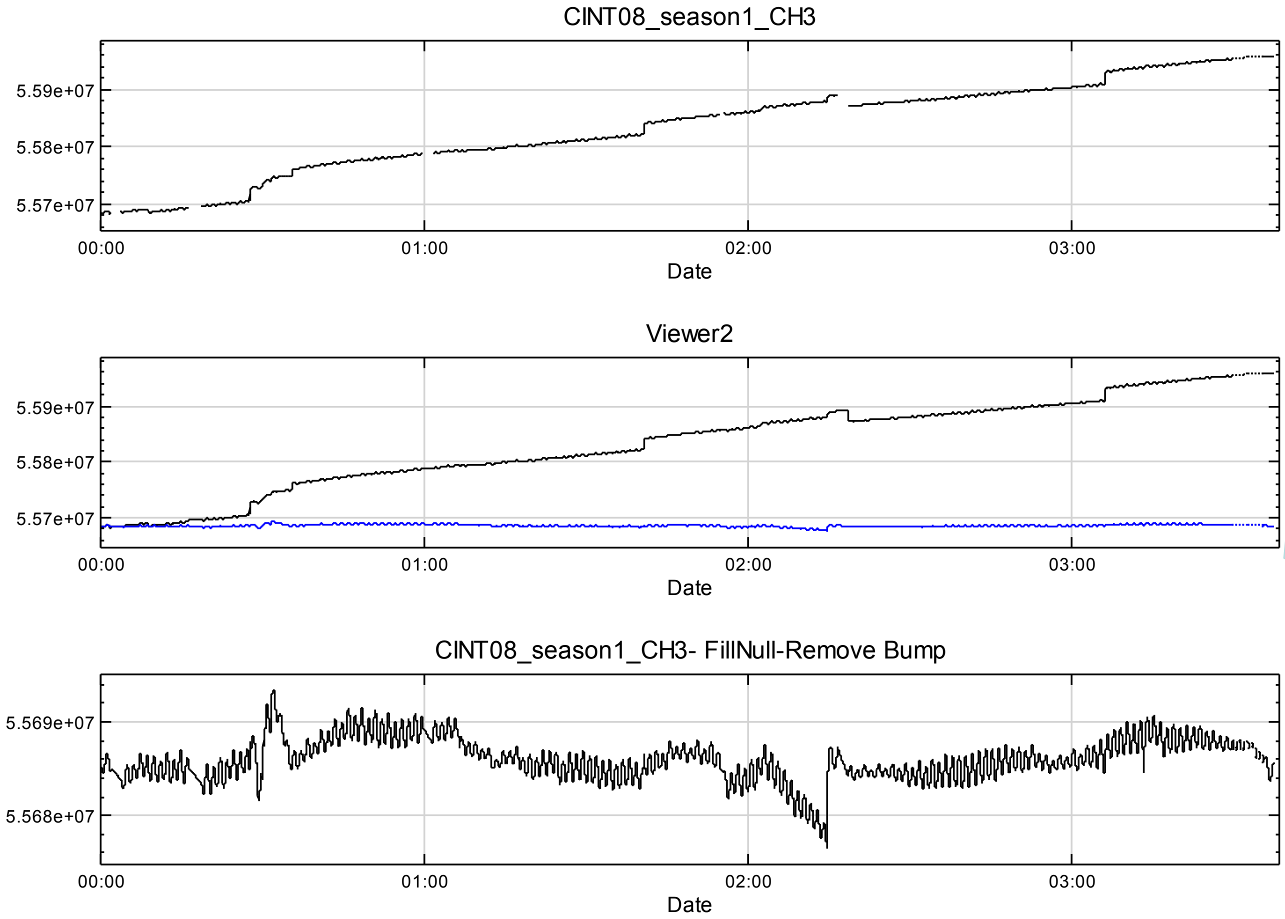
Trend Removal

Trend + Discontinuity

The process


TAISseason4CH3
Viewer [1]
FillNull
Remove Bump
Viewer2 [2]
Fast Trend
Viewer3 [3]
Mixer
Viewer4 [4]

Strain Meter (CINT08season1ch3)

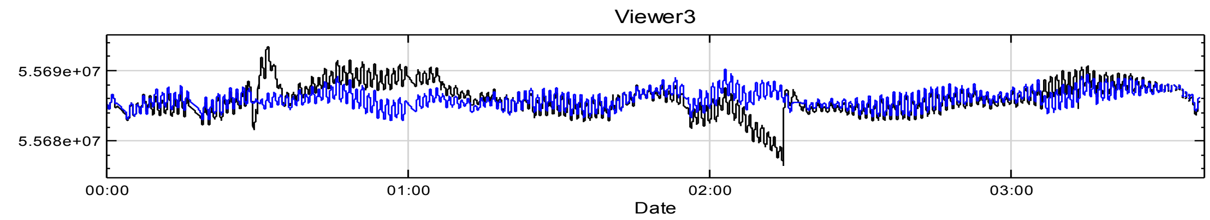


Raw data

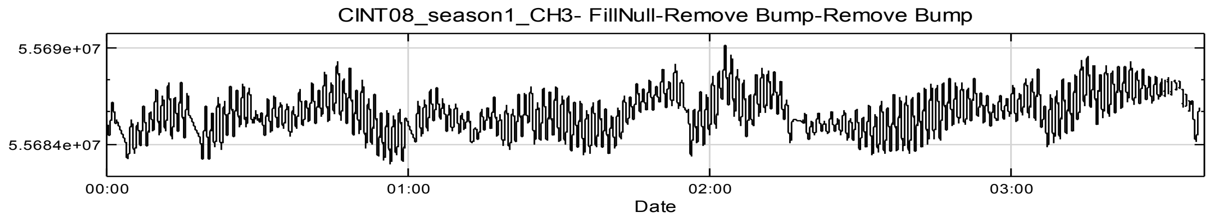
Fill Null
(previous value)

First Bump
Removal

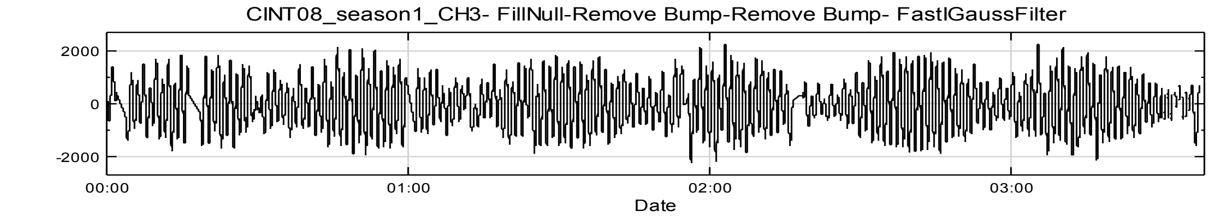
Strain Meter (CINT08season1ch3)



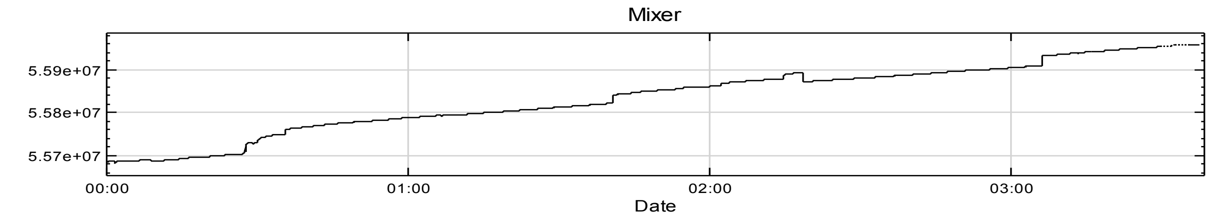
First Bump Removal



Second Bump Removal*



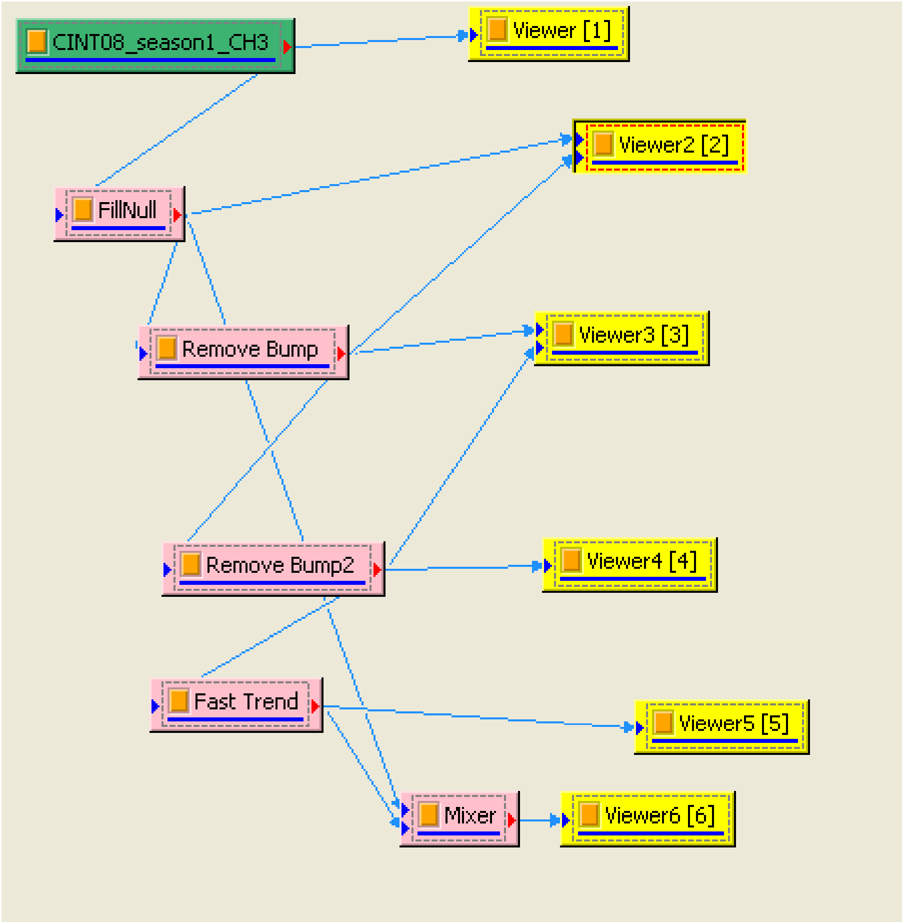
Trend removal



Trend + Discontinuity

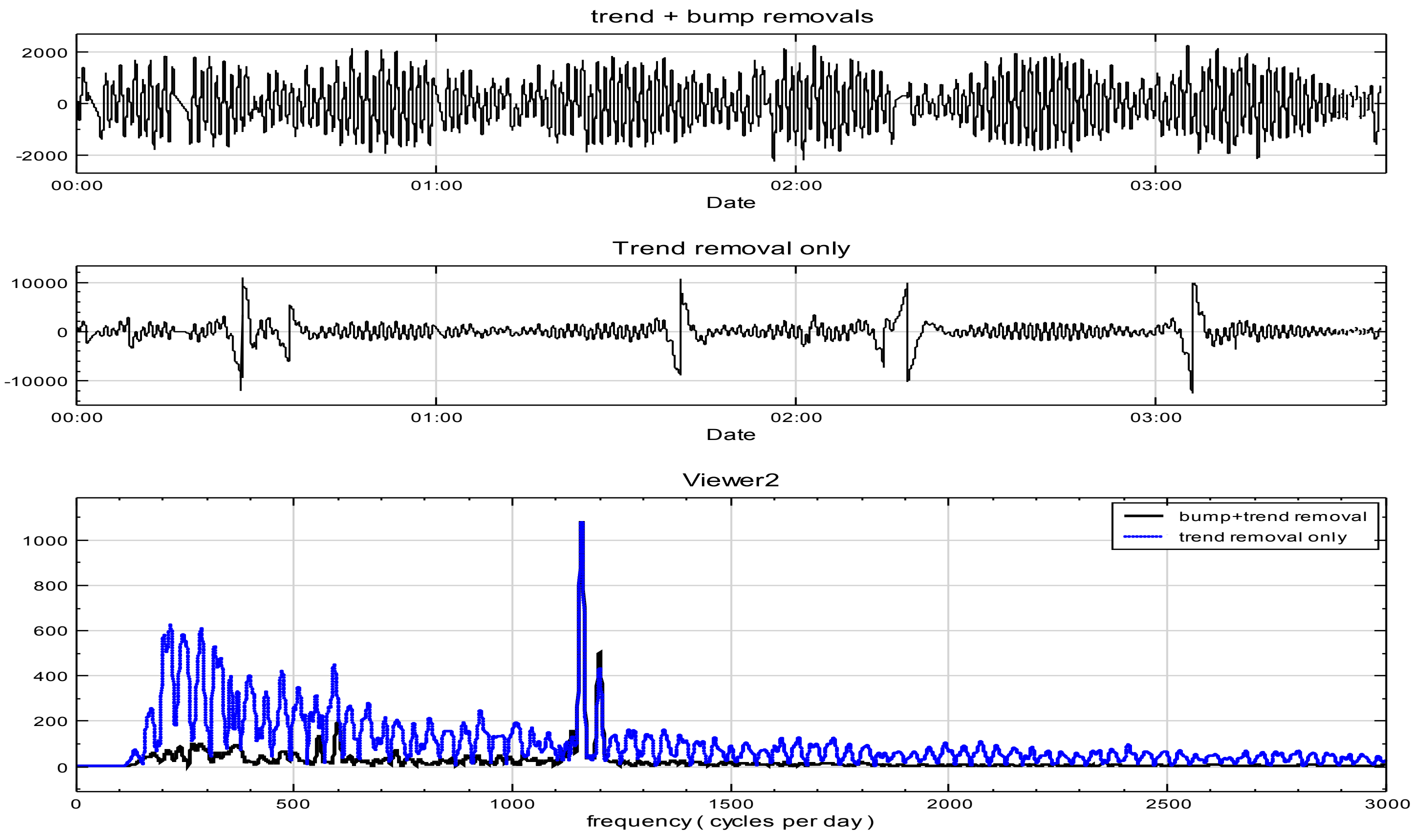
*The second bump removal uses less jump length (10% of the maximum) as threshold.

The process



CINT08_season1_CH3
Viewer [1]
Viewer2 [2]
FillNull
Remove Bump
Viewer3 [3]
Remove Bump2
Viewer4 [4]
Fast Trend
Viewer5 [5]
Mixer
Viewer6 [6]

Spectrum Improvement



Signal with both bump and trend removals has much better spectrum.

Thank you!