Diffusive and Fast Filter for Trend Removal

Yih-Nen Jeng Department of Aeronautics and Astronautics, National Cheng Kung University, Taiwan 第二屆時頻分析與地球科學研討會 Time-Frequency Analysis and its Application in Geo-Science



NCKU

Content

- **1. Motivation**
- **2. Theoretical Development**
- **3. Results & Discussions**
- 4. Conclusions & Future Works



NCKU





(March 21, 1768 - May 16, 1830) was a French mathematician and physicist who is best known for initiating the investigation of Fourier series and their application to problems of heat flow. The Fourier transform is also named in his honor.

On the Propagation of Heat in Solid Bodies







- On the Propagation of Heat in Solid Bodies (1807)
- Committee: consisting of Lagrange, Laplace, Monge and Lacroix.
- Two objections:
- All these are written with such exemplary clarity from a logical as opposed to calligraphic point of view - that their inability to persuade <u>Laplace</u> and <u>Lagrange</u> ... provides a good index of the originality of Fourier's views.
- Biot against Fourier's derivation of the equations of transfer of heat. Fourier had not made reference to Biot's 1804 paper on this topic but Biot's paper is certainly incorrect. Laplace, and later Poisson, had similar objections.



- The Propagation of Heat in Solid Bodies (1811)
- Report of the Paris Institute about the 1811 mathematics award.
- Fourier submitted his 1807 memoir together with additional work on the cooling of infinite solids and terrestrial and radiant heat. Only one other entry was received and the committee set up to decide on the award of the prize, <u>Lagrange</u>, <u>Laplace</u>, <u>Malus</u>, Haüy and <u>Legendre</u>, awarded Fourier the prize.



- The report was not however completely favourable and states:
- ... the manner in which the author arrives at these equations is not exempt of difficulties and that his analysis to integrate them still leaves something to be desired on the score of generality and even rigour.
- Fourier won the prize but the paper was not accepted for publishing!



- Fourier was elected to the <u>Académie des Sciences</u> in 1817.
- In 1822, Fourier became Secretary, and the <u>Académie</u> published his prize winning essay *Théorie analytique de la chaleur* in 1822.
- Why Laplace & Lagrange did not agree with Fourier's point of view?
- 主要是當時的數學環境還無法完全証明 Fourier 級數理論的嚴格性,因此 Lagrange、Laplace 一直持保留態度,這個混亂的情況到1811年,Fourier 以 擴增的論文獲得數學大獎後,仍然未能解決,也造成得獎論文不能發表的怪 事。事實上這場論戰,要經過 Poisson、Cauchy,一直到 Dirichlet 登場(weak projection studies became the main stream of mathematics in the subsequent decades),才真正落幕。

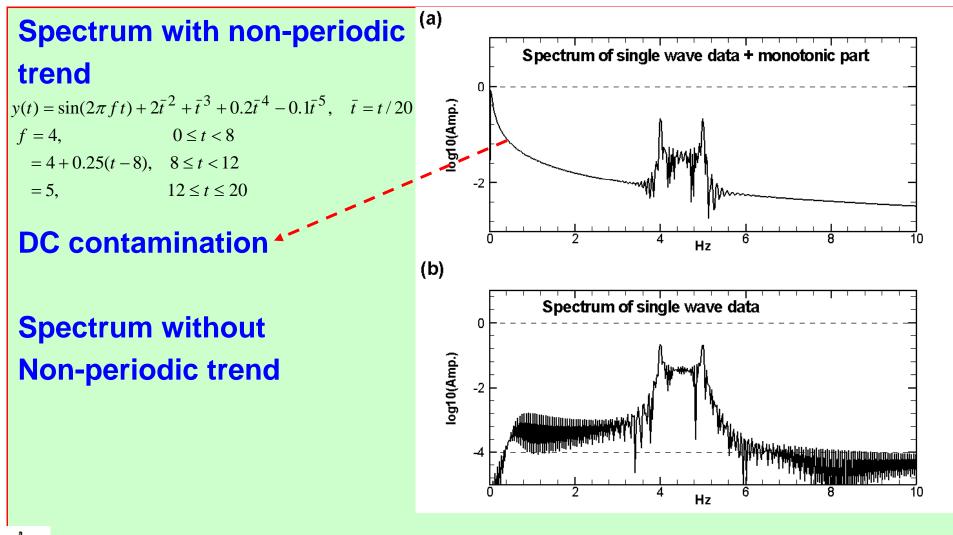


- Why Laplace & Lagrange did not agree with Fourier's point of view?
- Fourier Series use infinitely differentiable functions. Why use them to represent a simple discrete function (non-periodic)?
- 主要是當時的數學環境還無法完全証明 Fourier 級數理論的嚴格性,因此 Lagrange、Laplace 一直持保留態度,這個混亂的情況到1811年,Fourier 以 擴增的論文獲得數學大獎後,仍然未能解決,也造成得獎論文不能發表的怪 事。事實上這場論戰,要經過 Poisson、Cauchy,一直到 Dirichlet 登場(weak projection studies became the main stream of mathematics in the subsequent decades),才真正落幕。



 The non-periodic trend of a data string always introduces the Direct Current (DC) contamination to almost the whole spectrum •







- We had developed a diffusive Gaussian filter to deal with this problem •
- However, the transition zone of the Gaussian filter is too wide and a huge computing resource is necessary for a narrow transition zone.



Merit of Using a Diffusive Filter

 No unknown dispersive error (phase error) is introduced – important for precise data analysis.



Theoretical Development



Gaussian Smoothing

• For a data string (x_j, y_j) Gaussian smoothing gives

$$\overline{y}_{j} = \frac{1}{k} \sum_{i=-\infty}^{\infty} e^{-(i-j)^{2} (\Delta x)^{2}/2\sigma^{2}} y_{i}$$
$$k = \sum_{i=-\infty}^{\infty} e^{-(i-j)^{2} (\Delta x)^{2}/2\sigma^{2}} \approx \sqrt{2\pi\sigma} / \Delta x$$

• This is an approximately diffusive low-passed filter.



CKU

A diffusive filter but the transition zone is too wide

• Original data

$$y = \sin \frac{2\pi x}{\lambda_n}$$

• After smoothing, no phase error is introduced.

$$\overline{y} = a(\sigma, \lambda) \sin \frac{2\pi x}{\lambda_n} \qquad 0 \le a \le 1, \ a \approx \exp[-2\pi^2 \sigma^2 / \lambda^2]$$
$$a \to \pm |\varepsilon|, \text{ if } \lambda \le 0.6\sigma \quad (\varepsilon = \text{the machine error})$$
$$\to c, \text{ if } 0.6\sigma \le \lambda \le 40\sigma \text{ wide transition zone}$$
$$\to 1, \text{ if } 40\sigma \le \lambda$$



Gaussian Filter

An Iterative Filter basing on Gaussian smoothing

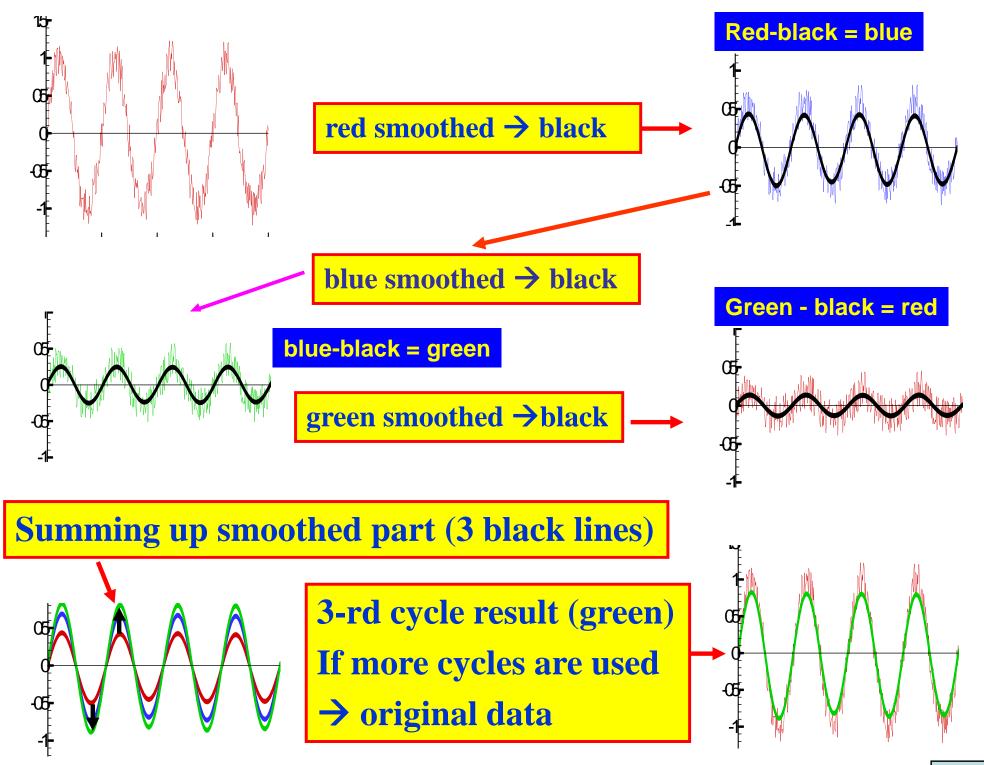
Jeng, Y. N., Huang, P. G., and Cheng, Y. C., "Decomposition of One-Dimensional Waveform Using Iterative Gaussian Diffusive Filtering Methods," Proc. Roy. Soc. A. (2008) vol.464, pp.1673–1695, doi:10.1098/rspa.2007.0031, Published online 13 March 2008.



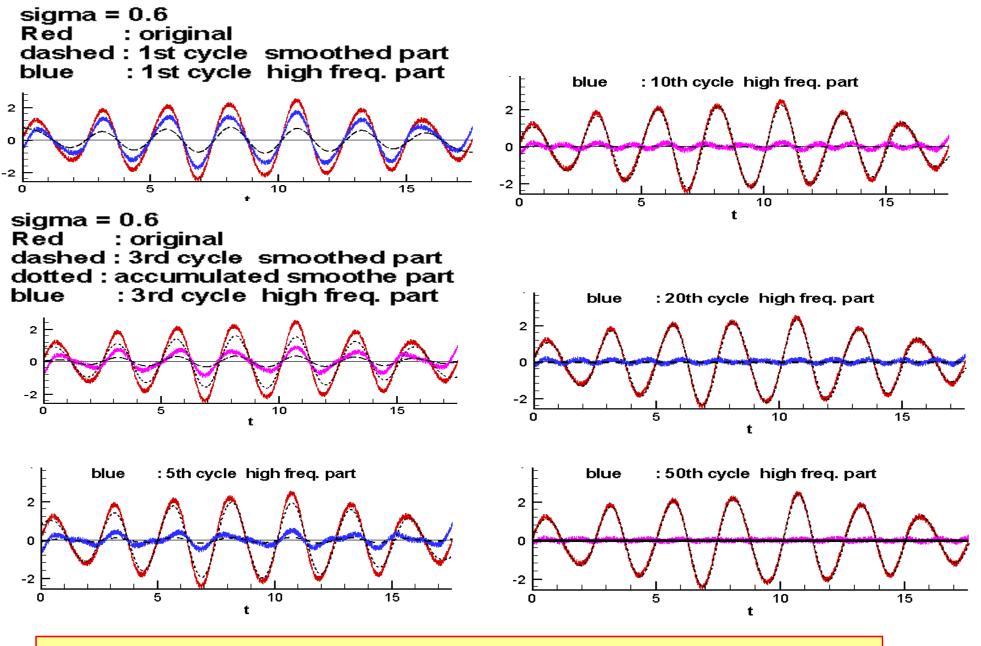
Gaussian Filter

 Repeatedly smooth the remaining high frequency part.





NCKU



1.Accumulated smooth part \rightarrow original data as iteration increases **?** High frequency part \rightarrow final high freq. part



Gaussian Filter

Original data

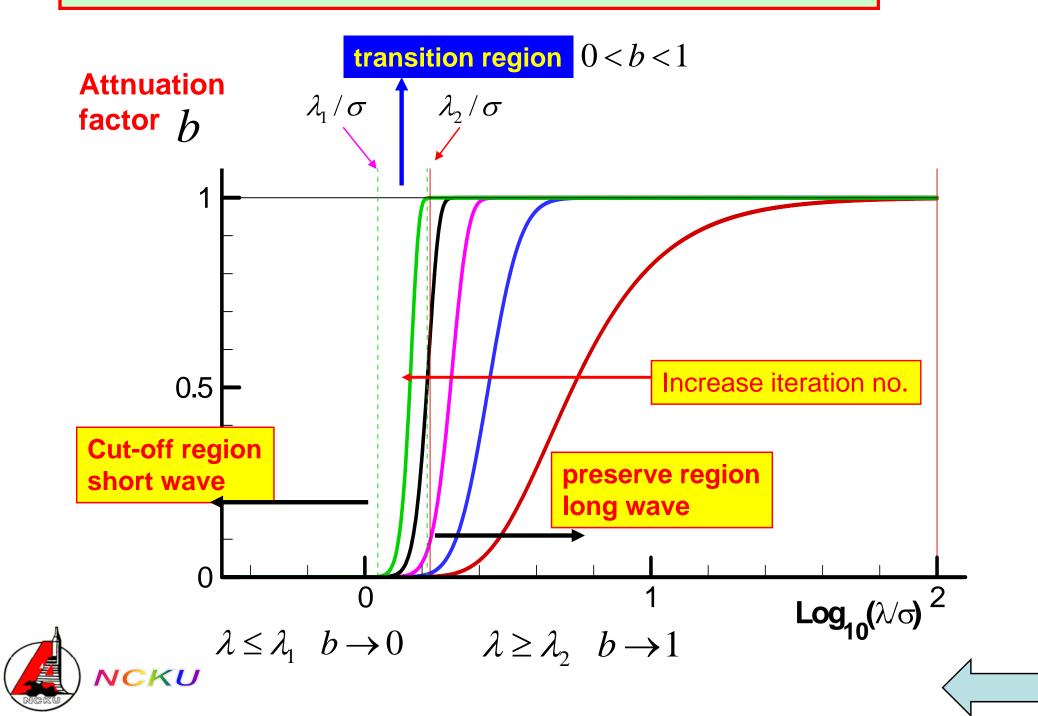
$$y = \sin \frac{2\pi x}{\lambda_n}$$

• Iterative filter with factor σ & iteration step number m, gives the accumulated smoothed part

$$\overline{y}_{m} = b(\sigma, \lambda) \sin \frac{2\pi \alpha}{\lambda_{n}}, \quad b \approx 1 - \{1 - a(\sigma, \lambda)\}^{m}$$
$$0 \le [1 - a(\sigma, \lambda)]^{m} \le 1 \pm m |\varepsilon| \qquad \pm |\varepsilon| \le a(\sigma, \lambda) \approx \exp[-2\pi^{2}\sigma^{2}/\lambda^{2}] \le 1$$

This iterative filter is also an approximately diffusive with narrow transition zone

Attenuation factor *b* vs. *m* & λ with fixed σ



Gaussian Filter

Given b1 & b2 \rightarrow solve m & σ from

$$b(\sigma/\lambda_{c1},m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c1}^2}\}]^m = b_1 \quad \Rightarrow 0.001$$

$$b(\sigma/\lambda_{c2},m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c2}^2}\}]^m = b_2 \quad \Rightarrow 0.999$$

After specifying $\lambda_{c1} \& \lambda_{c2}$, apply the iterative Gaussian smoothing (with a fixed σ) m iterations \rightarrow to obtain the smoothing & high frequency parts NCKU

$$b(\sigma / \lambda_{c1}, m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c1}^2}\}]^m = 0.001$$

$$b(\sigma / \lambda_{c1}, m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c1}^2} \left(\frac{\lambda_{c1}}{\lambda_{c2}}\right)^2\}]^m = 0.999$$

Solve for σ / λ_{c1} & m

$$b(\sigma / \lambda_{c1}, m) = 0.011 \ \& 0.999$$

$$m = 2,3 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999$$

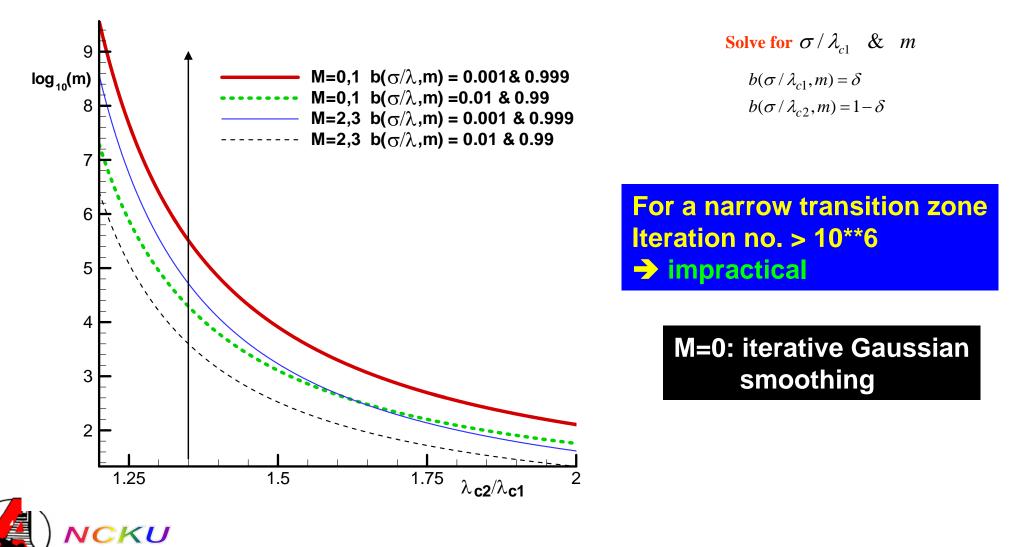
$$m = 2,3 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999$$

M = 0.1 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999
M = 0.1 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999
M = 0.1 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999
M = 0.1 \ b(\sigma / \lambda, m) = 0.001 \ \& 0.999
M = 0.1 \ b(\sigma / \lambda_{c2}, m) = 1 - \delta
M = 0: iterative Gaussian smoothing



$$b(\sigma/\lambda_{c1},m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c1}^2}\}]^m = 0.001$$

$$b(\sigma/\lambda_{c1},m) = 1 - [1 - \exp\{-\frac{2\pi^2 \sigma^2}{\lambda_{c1}^2} \left(\frac{\lambda_{c1}}{\lambda_{c2}}\right)^2\}]^m = 0.999$$



Required iteration no. & smoothing factor w.r.t. accuracy criteria

- δ 0.01, 0.001, 0.0001, 0.00001, 0.000001
- *m_g* 33 127 410 1199 3306
- Iteration no. increases as δ decreases
- • σ_g / λ_1 0.64 0.7716 0.8783 0.9708 1.0538

$$b(\sigma / \lambda_{c1}, m) = \delta$$

$$b(\sigma / \lambda_{c2}, m) = 1 - \delta \quad \text{for } \lambda_{c2} / \lambda_{c1} = 2$$



Effect of Gaussian filter upon a polynomial

$$y = \sum_{n=0}^{N} a_n t^n$$

$$y'(t) = y(t) - \overline{y}(t) \approx \sqrt{2\pi}\sigma^3(a_2 + 3a_3t) + \sqrt{2\pi}a_4[6t^2\sigma^3 + 3\sigma^5]$$

$$+ \sqrt{2\pi}a_5[10t^3\sigma^3 + 15t\sigma^5] +$$

$$\sum_{n=0}^{N} a_n \sum_{k=1}^{\left[\frac{n}{2}\right]} \left[\binom{n}{2k} t^{n-2k} \sqrt{2\pi}(2k-1)!!\sigma^{2k+1}\right] + O(\Delta t^2) \approx \sum_{n=0}^{N-2} b_n t^n$$

a1, a2, & highest two powers are removed



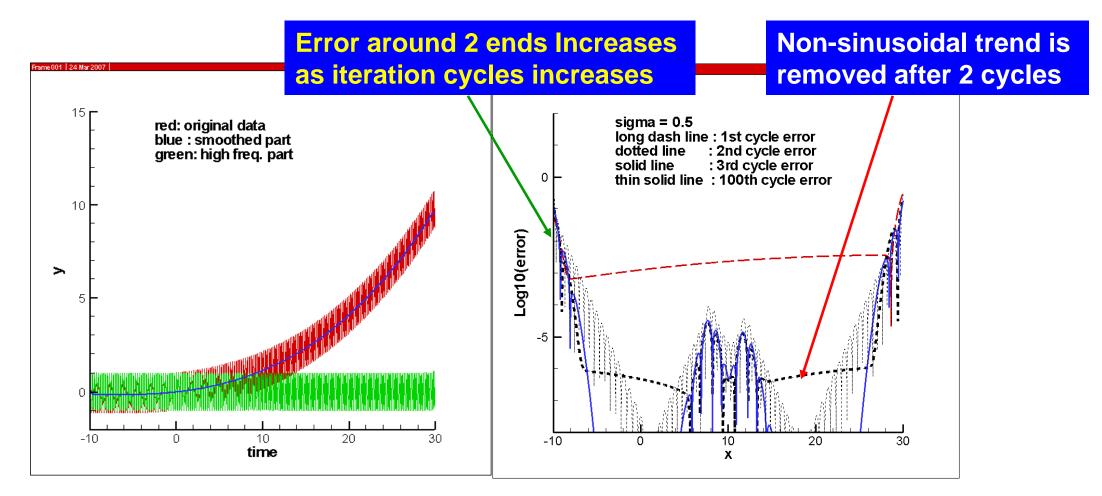
Effect of Gaussian filter upon a polynomial

$$y = \sum_{n=0}^{N} a_n t^n$$
$$y_m'(t) \approx O(\Delta t^2) \text{ for } m > N / 2$$

For a sufficient large iteration cycles, the non-periodic trend will be ultimately removed.



Error of high frequency part



$$y(t) = \sin(2\pi f t) + 2\bar{t}^2 + \bar{t}^3 + 0.2\bar{t}^4 - 0.1\bar{t}^5, \quad \bar{t} = t/20$$

$$f = 4, \qquad 0 \le t < 8$$

$$= 4 + 0.25(t-8), \quad 8 \le t < 12$$

$$= 5, \qquad 12 \le t \le 20$$



Iterative Gaussian Smoothing Methods for Finite Data String

Upper bound of error penetration distance

$$\approx k \cdot \log_{10}(m)$$





★ Consider a data string of the form

$$y_j = \sum_{l=0}^{\infty} [b_l \cos \frac{2\pi x_j}{\lambda_l} + c_l \sin \frac{2\pi x_j}{\lambda_l}] + \sum_{n=0}^{N} a_n x_j^n$$

The polynomial designs the non-periodic trend.



NCKU

Assume the error functional as

$$I_{k} = \sum_{i} \exp[-\frac{(x_{i} - x_{k})^{2}}{2\sigma^{2}}][y_{i} - \sum_{l=0}^{M} A_{l,k} (x_{i} - x_{k})^{l}]^{2}$$

The least squares approximation gives

$$(y')^{M} = \sum_{n=0}^{N-2M} g^{M}{}_{n}x^{n} + \sum_{l=0}^{\infty} [1 - a_{M}(\sigma, \lambda_{l})][b_{l}\cos(\frac{2\pi x_{i}}{\lambda_{l}}) + c_{l}\sin(\frac{2\pi x_{i}}{\lambda_{l}})]$$

$$a_{M}(\sigma,\lambda_{l}) = [1 + B + \frac{B^{2}}{2!} + \dots + \frac{B^{M}}{M!}]e^{-B}, \quad B = -2\pi^{2}\sigma^{2}/\lambda_{l}^{2}$$



NCKU

★ To prove

$$(y')^{M} = \sum_{n=0}^{N-2M} g^{M} {}_{n}x^{n} + \sum_{l=0}^{\infty} [1 - a_{M}(\sigma, \lambda_{l})] [b_{l} \cos(\frac{2\pi x_{l}}{\lambda_{l}}) + c_{l} \sin(\frac{2\pi x_{l}}{\lambda_{l}})]$$

$$a_{M}(\sigma, \lambda_{l}) = [1 + B + \frac{B^{2}}{2!} + ... + \frac{B^{M}}{M!}] e^{-B}, B = -2\pi^{2}\sigma^{2}/\lambda_{l}^{2}$$
★ For $M < 11$, use hand or Mathematica.
★ Otherwise, properly split the procedure and can prove it up to $M = 300$.
★ For a still high order, high computing device and algorithm are necessary.

Repeatedly apply the moving least squares method for *m* cycles gives

$$(y')_{m}^{M} \approx \sum_{n=0}^{N-2mM} g_{m_{n}}^{M} x^{n} + \sum_{l=0}^{\infty} [1 - a_{M}(\sigma, \lambda_{l})]^{m} [b_{l} \cos(\frac{2\pi x_{i}}{\lambda_{l}}) + c_{l} \sin(\frac{2\pi x_{i}}{\lambda_{l}})]$$

The embedded trend will be ultimately removed if 2mM > N.



★ The transition zone is determined by

$$\begin{split} & \left[1-a_{M}\left(\sigma,\lambda_{1}\right)\right]^{m}\approx1-\delta\\ & \left[1-a_{M}\left(\sigma,\lambda_{2}\right)\right]^{m}\approx\delta \end{split}$$

3 parameters: $M, m \& \sigma$

2 equations

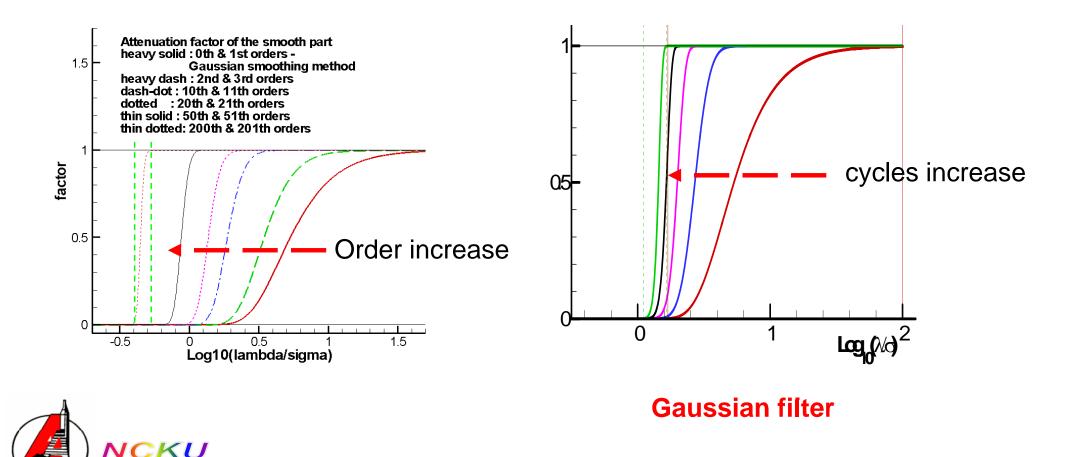
$$a_{M}(\sigma,\lambda_{l}) = [1+B+\frac{B^{2}}{2!}+...+\frac{B^{M}}{M!}]e^{-B}, B = -2\pi^{2}\sigma^{2}/\lambda_{l}^{2}$$

\bigstar For a given order *M* **solve for** *m* & σ **.**



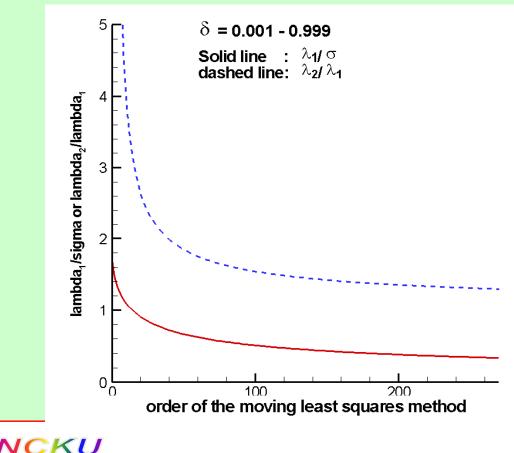
Increase order, decrease transition zone's width

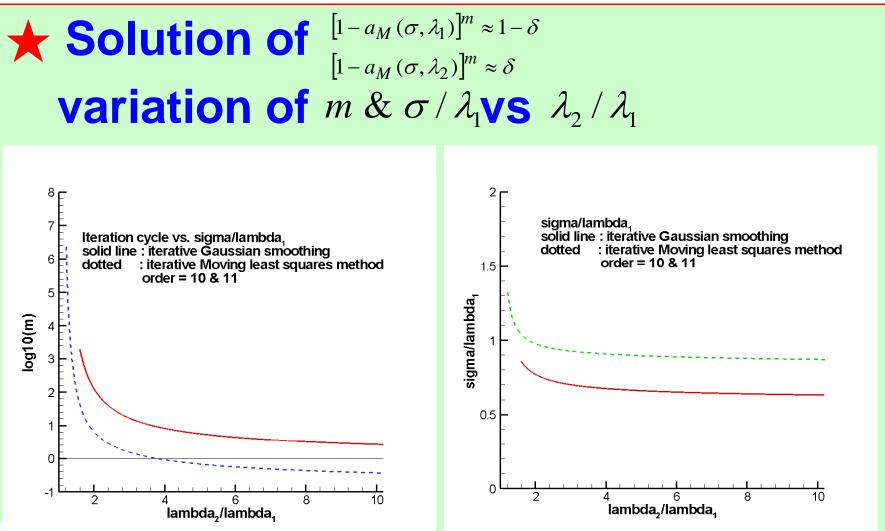
Increase iteration cycles, decrease transition zone's width





t Increase order, decrease λ_2 / λ_1 increase σ

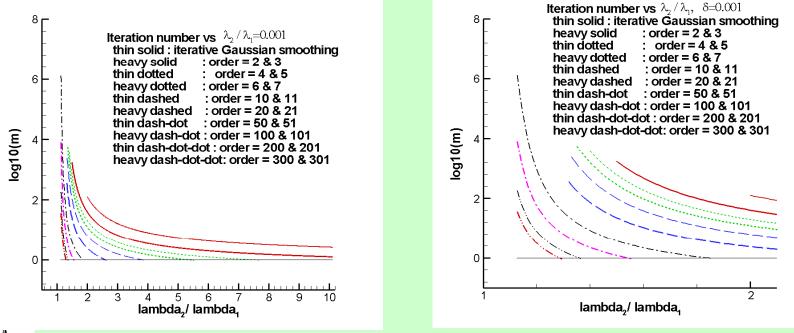






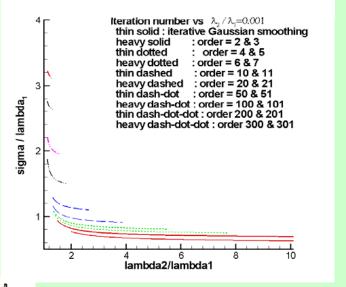
VCKU

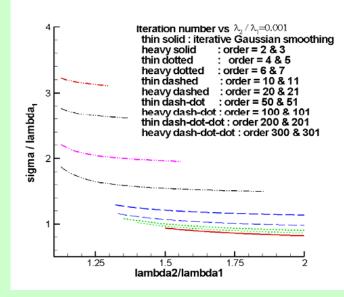
★ For small λ_2 / λ_1 , large *m* cycles is required. As order *M* increase *m* is reduced.





★ For a wide λ_2 / λ_1 , only lower order *M* is possible.







★ For a $\lambda_2 / \lambda_1 \le 1.1$, the solution is critical because the evaluation of

$$a_{M}(\sigma,\lambda_{l}) = [1+B+\frac{B^{2}}{2!}+...+\frac{B^{M}}{M!}]e^{-B}, B = -2\pi^{2}\sigma^{2}/\lambda_{l}^{2}$$

It needs a very high accuracy devices and algorithm.



Moving Least Squares Filters Procedure

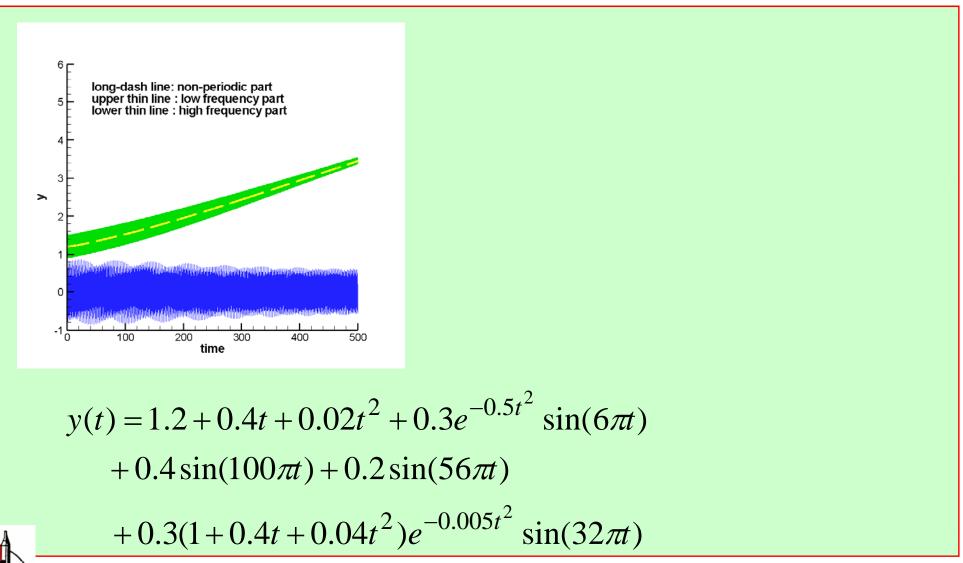
± Evaluate the Fourier spectrum via FFT **\bigstar** Determine the transition zone $\lambda_2 \& \lambda_1$ **\bigstar Find suitable** $M, m, \& \sigma$ **Multiple each mode with** $[1-a_M(\sigma,\lambda_l)]^m$ $a_{M}(\sigma,\lambda_{l}) = [1+B+\frac{B^{2}}{2!}+...+\frac{B^{M}}{M!}]e^{-B}, B = -2\pi^{2}\sigma^{2}/\lambda_{l}^{2}$ **+** Perform inverse FFT of the resulting spectrum -> desired high freq. part Required CPU = 2 * FFT's CPU + extra

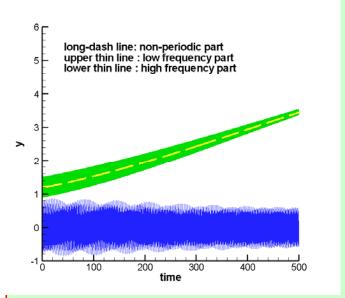
Results & Discussions



Moving Least Squares Filters Test Case

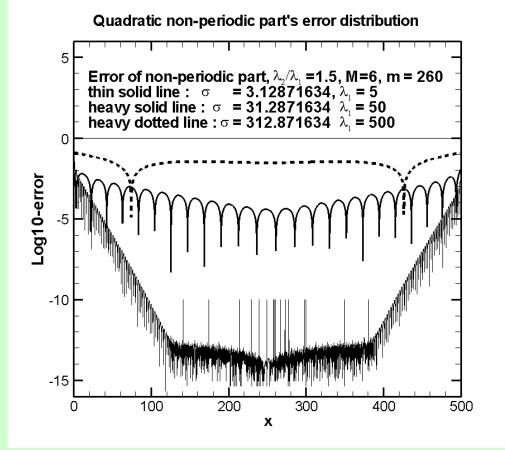




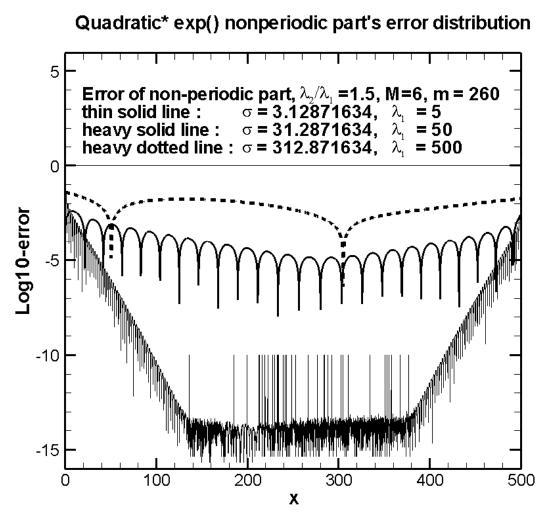


To extract the trend

- 1. Small σ gives good interior accuracy.
- 2. For large σ , the error is around 10^{-2}



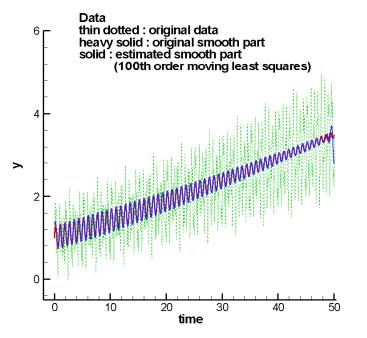




If the trend is made complicated, the result is similar.



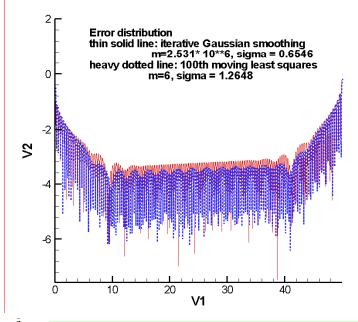
★ For this case the required transition zone width is $\lambda_2 / \lambda_1 \approx 1.3$, parameters are $m = 2.531 \times 10^6$, $\sigma = 0.6546$ for M = 0; m = 6, $\sigma = 1.26$..., M = 100



 $y(t) = 1.2 + 0.4t + 0.02t^{2} + 0.3e^{-0.5t^{2}} \sin(6\pi t)$ + 0.4 sin(100\pi t) + 0.2 sin(56\pi t) + 0.3(1 + 0.4t + 0.04t^{2})e^{-0.005t^{2}} sin(32\pi t)



★ For this case the required transition zone width is $\delta = 0.001$, $\lambda_2 / \lambda_1 \approx 1.3$, parameter are $m = 2.531 \times 10^6$, $\sigma = 0.6546$, M = 0; m = 6, $\sigma = 1.26$..., M = 100



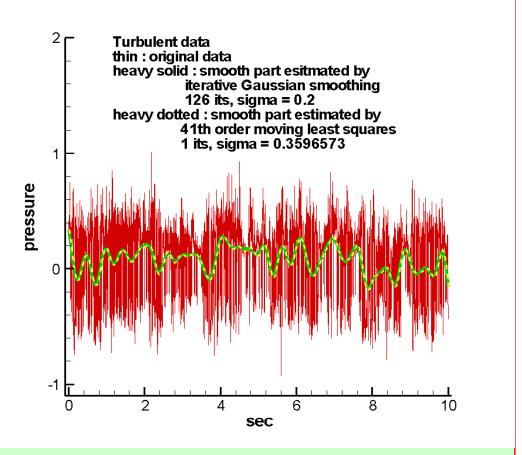
 $y(t) = 1.2 + 0.4t + 0.02t^{2} + 0.3e^{-0.5t^{2}} \sin(6\pi t)$ $+ 0.4 \sin(100\pi t) + 0.2 \sin(56\pi t)$ $+ 0.3(1 + 0.4t + 0.04t^{2})e^{-0.005t^{2}} \sin(32\pi t)$

Error at two ends are similar for 1. Gaussian smoothing 2. 100 th order moving least squares Error bond are $\approx k \cdot \log_{10}(m)$



Turbulent Data

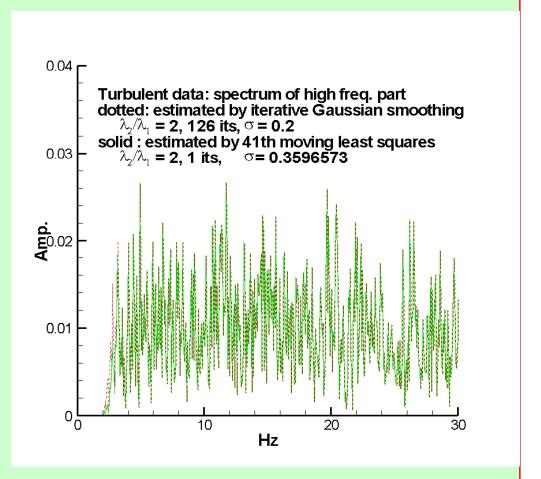
Both Gaussian filter
 (126 its) & moving
 least squares filter
 (41th order 1 its),
 trend & lower freq.
 part consist with
 each other as shown.





Turbulent Data

Both Gaussian filter
 (126 its) & moving
 least squares filter
 (41th order 1 its),
 trend & lower freq.
 part consist with
 each other as shown.





Applications



Sharp Filter Procedure

- Use the proposed fast and sharp filter to remove the non-periodic trend + extreme low frequency components.
- ★ The high frequency part and it's spectrum are obtained simultaneously.
- ★ Any filter or infinitely sharp filter upon the spectrum is straight forward.



Sharp Filter Application

- ★ It has been proven that a Gaussian window imposing upon the spectrum centered at certain freq. is the spectrogram coefficient of the freq..
- The high frequency part's spectrum is ready simultaneously.
- ★ A spectrogram with certain window is straight forward.

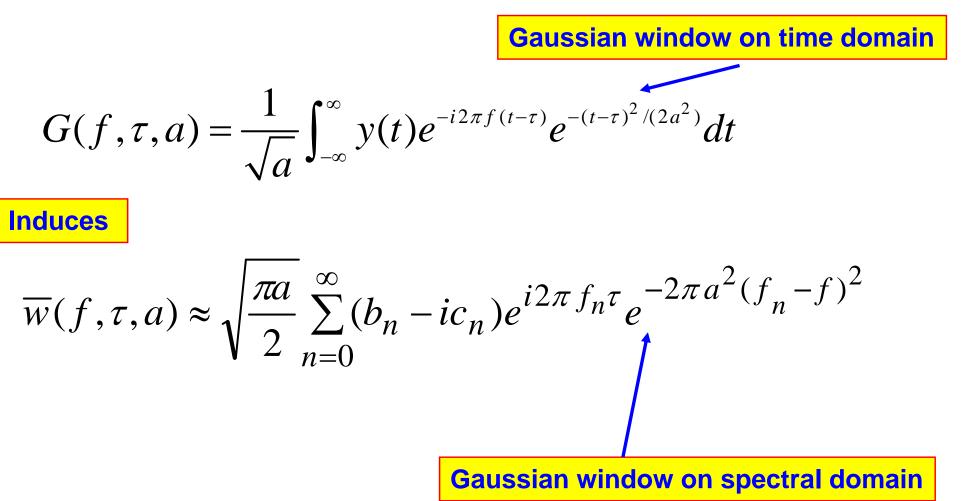


Characteristics of Morlet Transform (Continuous Wavelet Transform)

$$W(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(t) e^{-i6(t-\tau)/a} e^{-(t-\tau)^2/(2a^2)} dt$$

Gaussian window on time domain
 $f = 3/(a\pi)$
Induces (for short wave modes)
$$W(a,\tau) \approx \sqrt{\frac{\pi a}{2}} \left\{ \sum_{n=0}^{\infty} \exp\left(-\frac{a^2}{2} \left[\frac{2\pi}{\lambda_n} - \frac{6}{a}\right]^2\right]$$
Gaussian window
on spectral domain
$$\times \left[b_n \exp\left[\frac{i2\pi\tau}{\lambda_n}\right] + c_n \exp\left[-\frac{i2\pi\tau}{\lambda_n}\right] \right] \right\}$$

Gabor Transform (short time Fourier transform)





Application (2) SOI vs. CTI



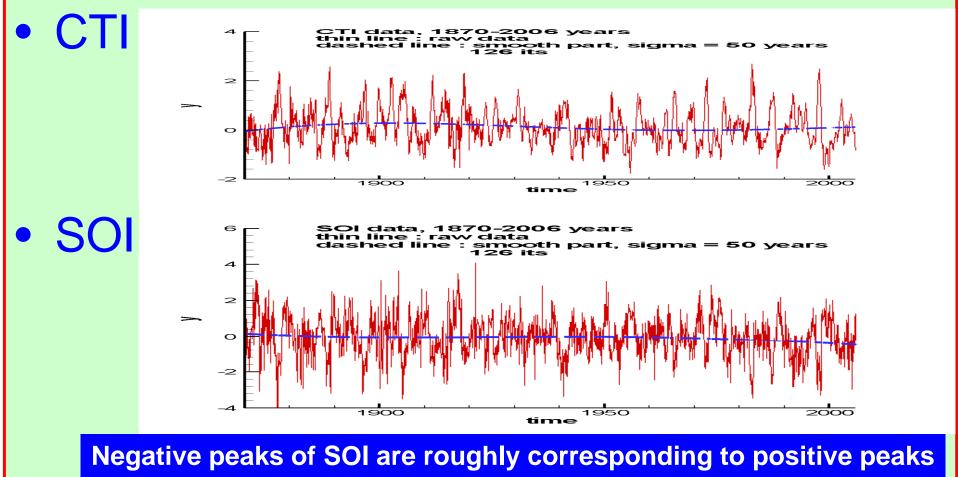
SOI vs. CTI

- CTI : normalized monthly sea level pressure index based on the pressure records collected in Darwin, Australia and Tahiti Island in the eastern tropical Pacific.
- SOI : average large year-to-year sea surface temperature anomaly fluctuations over 6°N-6°S, 180-90°W → An index of El Nino
- SOI : its negative peak often occurs with a 2 to 7 year period, corresponds to a strong EI Niño (global warm) event .

SOI vs. CTI

- Huang had employed the ensemble HHT to study the correlation between SOI and CTI •
- 3 modes 2.83, 5.23 and 20.0 years/cycle.
- Cross-correlation coeffi= -0.78, -0.79, and -0.75, respectively. → Confirm 2 to 7 year period of correlation.
- How about the detailed cross correlation coefficient distribution with respect to frequency variation?

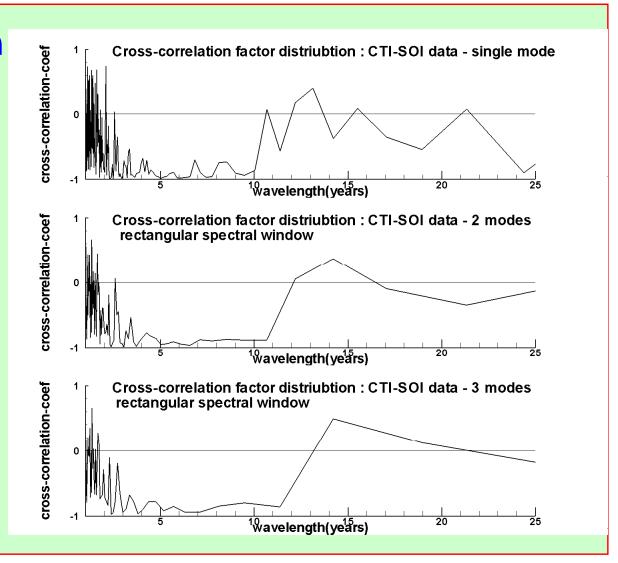
CTI & SOI raw data



of CTI.

SOI vs. CTI

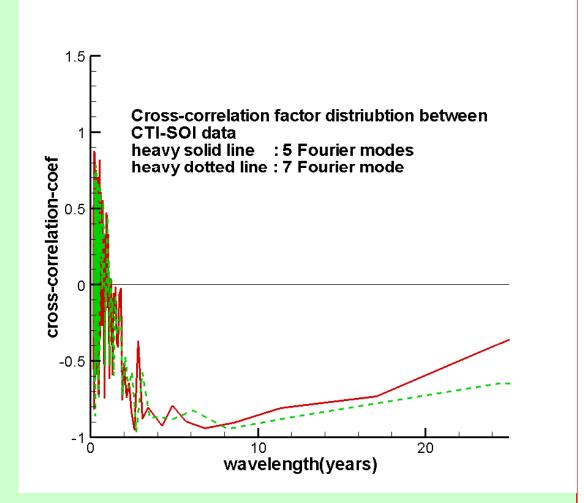
 Cross correlation factor for 1, 2, & 3 Fourier modes.
 2-7 years/cycle does have high correlation coeff.





SOI vs. CTI

Cross correlation
 factor for the 7
 Fourier modes:
 1.85-25 years/cycle
 does have high
 correlation coeff.





Application (3) ECG vs. ABP



ECG vs. ABP cross-correlation

★ Data base: on open domain (Web site) Moody GB, Mark RG. A database to support development and evaluation of intelligent intensive care monitoring. Computers in Cardiology 1996;23:657–660.

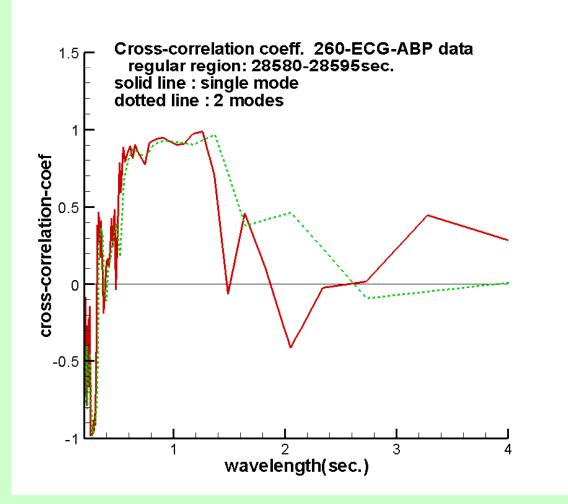
It consists of over 104 patient-days of real-time signals and accompanying annotations.

For Intensive care patients.



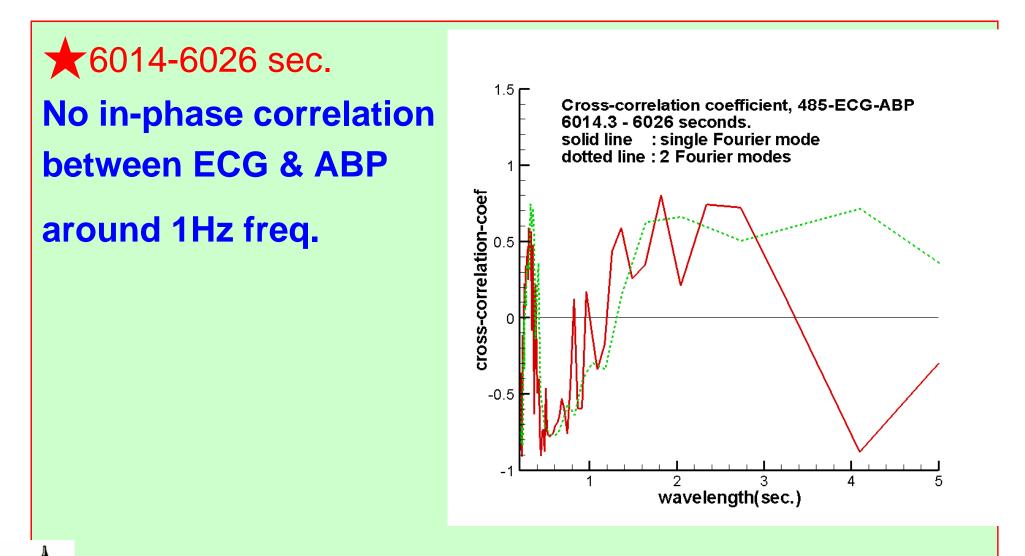
The false ventricular-related (VT) alarm case

*****28580-28595 sec. (with clear in-phase correlation factor around 1Hz (heart beating freq. in most intervals) → Need not special attention.



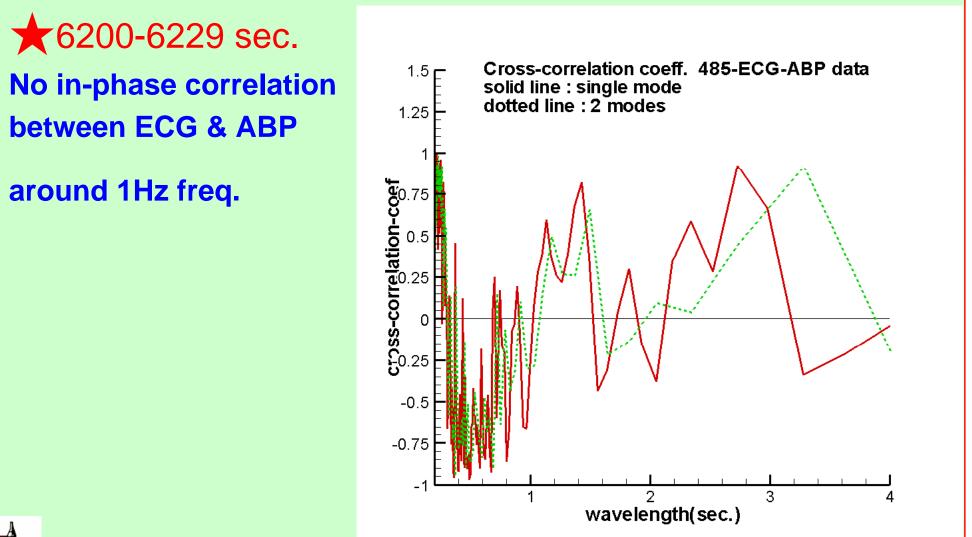


The true ventricular-related (VT) alarm case





The true ventricular-related alarm case





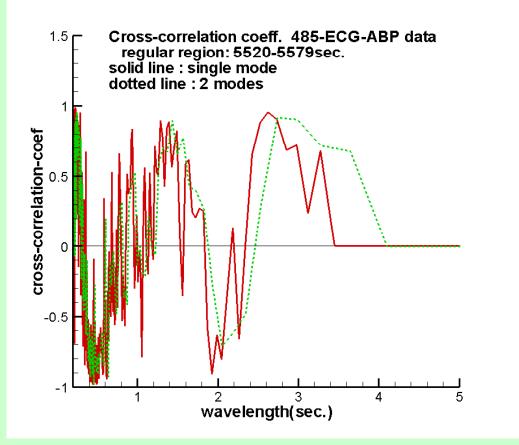
The true ventricular-related alarm case

★ 5520-5579 sec. No in-phase correlation between ECG & ABP around 1Hz freq.

In most data intervals, no in-phase correlation between ECG & ABP around 1Hz freq exists. → The patient is in critical

condition → need special care.

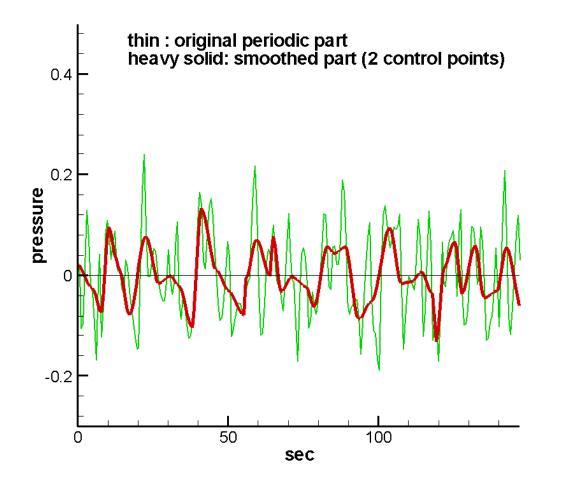




Enhanced IMF



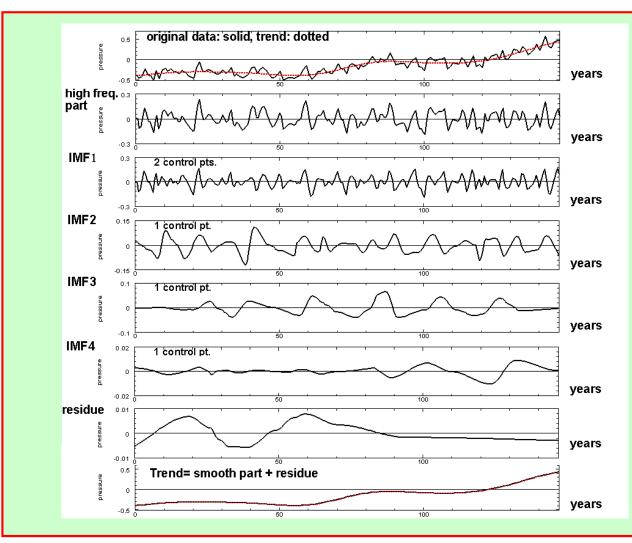
New IMF Generator via cubic spline polynomial + Least Squares method



Smooth response is not smooth enough Yet !



New IMF Generator via cubic spline polynomial

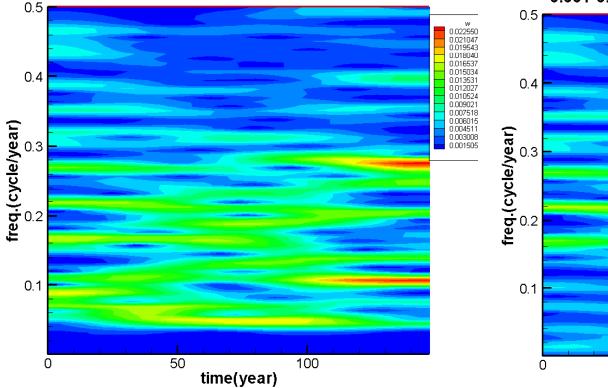


Original data is the GSTA (Global Surface air Temperature Anomaly)

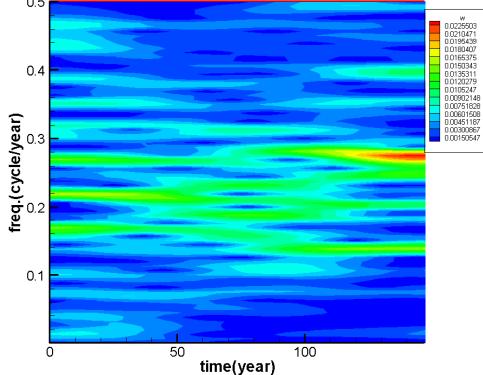


Original HHT (IMF1)

Gstal147 data, periodic part 0.001-0.501 cycle/year, 80 lines



Gstal147 data, IMF1-original HHT 0.001-0.501 cycle/year, 80 lines

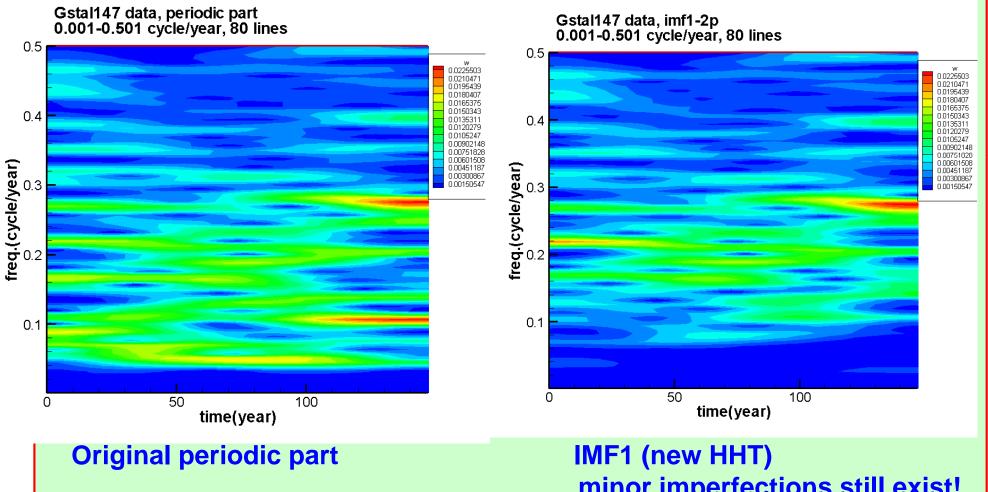


Original periodic part

IMF1 (original HHT) unknown extra part in low freq. regime is obvious!



Improved IMF1 (use spline polynomial + least squares)



minor imperfections still exist!



Conclusions

- The fast and sharp diffusive filter is proposed.
- The required CPU time is slightly longer than 2 times of the FFT.
- For a narrow transition zone (width < 1.125), the solution is unavailable now. It needs a high accuracy algorithm + high accuracy computing device.
- ★It is successfully applied to several cases.



Future Works



Future Works

To construct a table to show the best combination of parameters $M, m, \& \sigma$ for a given transition zone $\lambda_2 \& \lambda_1$. **Apply to science and engineering** problems. **The application is valuable because this** filter is a new tool to look into minor modes in the low frequency regime.



Thank You !











