FUNDAMENTALS OF TIME SERIES ANALYSIS FOR TRANSFER FUNCTION ESTIMATION

XAVIER GARCIA

LABORATORY OF SEAFLOOR AND SUBSEAFLOOR GEOLOGICAL PROCESSES (LG3GP)

INSTITUTE OF MARINE SCIENCES, CSIC

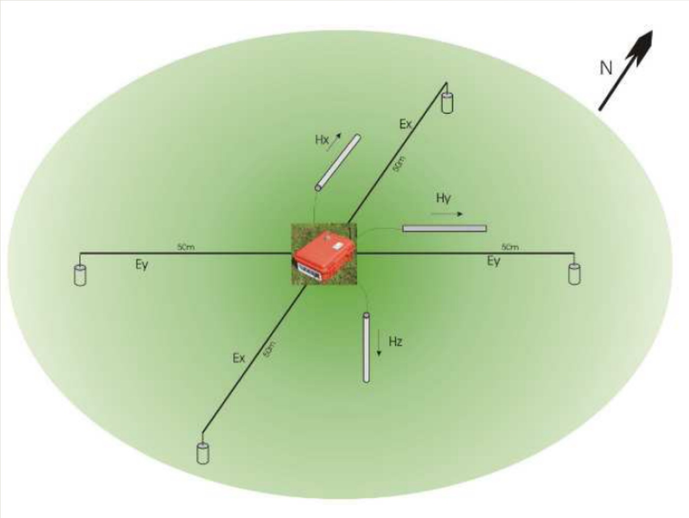
CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

LS3GP
Laboratory of Seafloor and SubSeafloor Geological Processes

Institut de Ciències del Mar

EMINAR 2020.12.02

FUNDAMENTAL PROBLEM

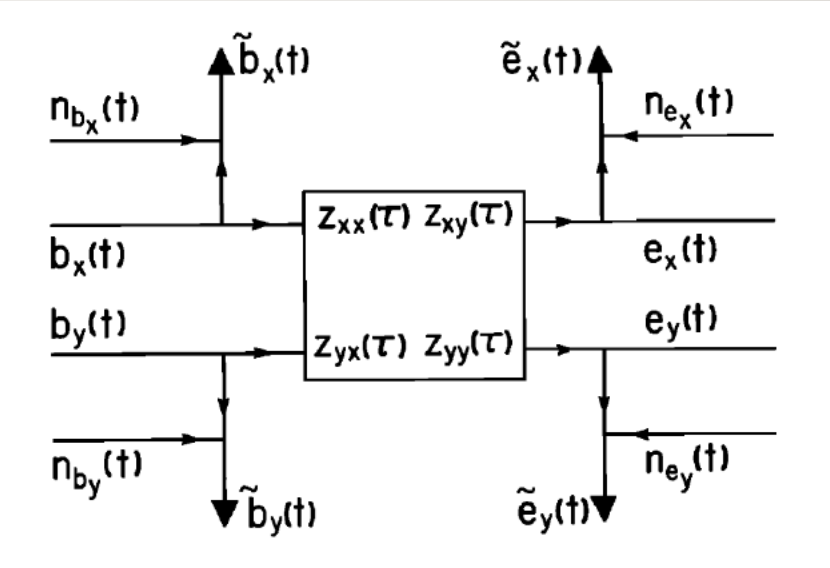


$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix}$$

↓

$$E_x = Z_{xx} \cdot B_x + Z_{xy} \cdot B_y$$
$$E_x = Z_{yx} \cdot B_x + Z_{yy} \cdot B_y$$

- electrical impedance *Z* describes subsurface conductivity structure
- data always contains source signal and noise

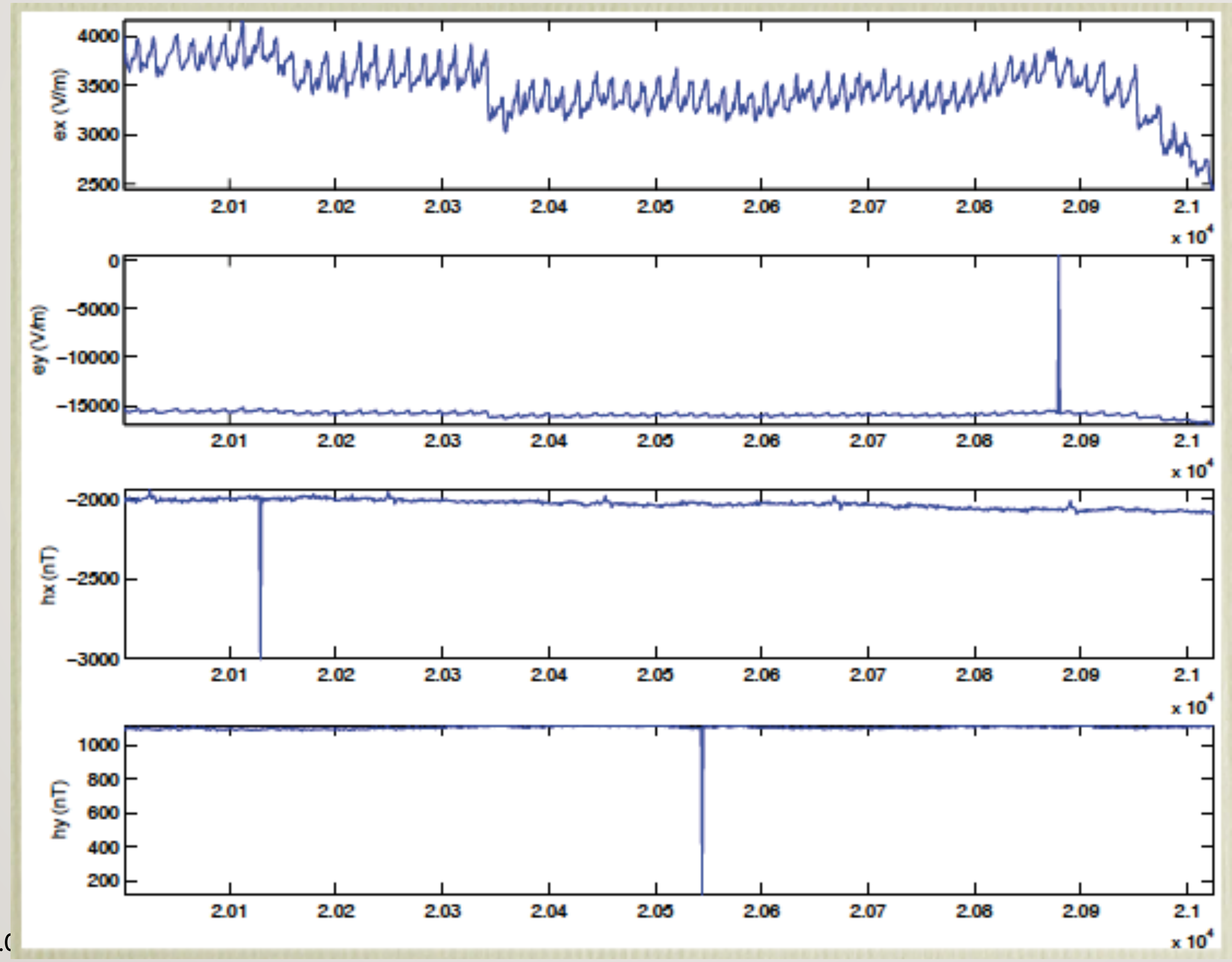


ROADMAP

1. Pre-processing
2. Spectral Estimation
3. Impedance Estimation
4. Uncertainty Estimation

1. PRE-PROCESSING

LOOK AT THE DATA
(GARCIA ET AL., 1999; WECKMANN ET AL., 2005)



DATA CONDITIONERS

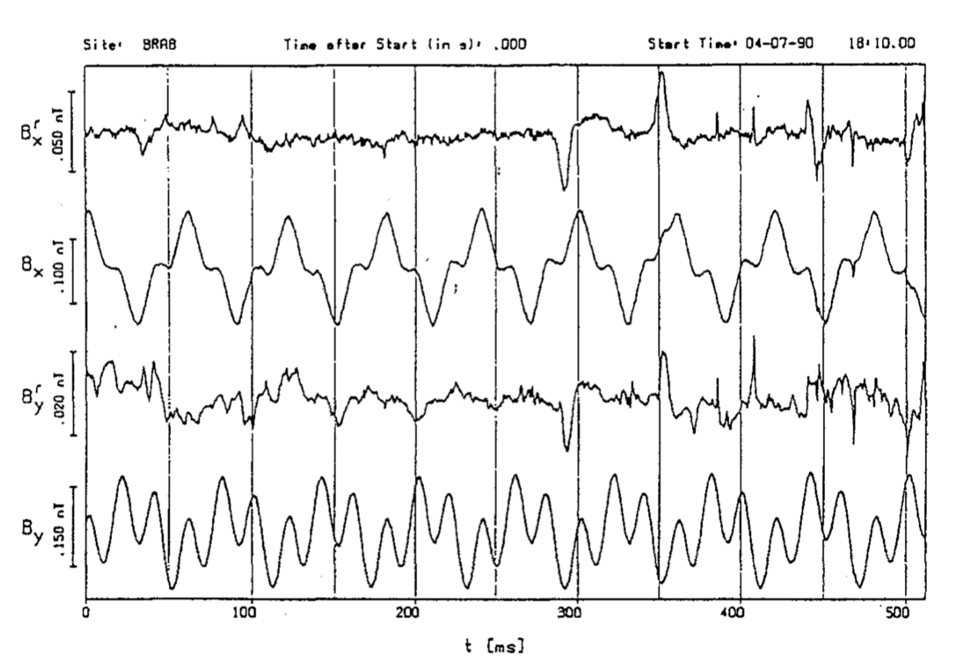
- Motion Inducted noise: adaptive cancellation filter
- Filters, including Notch filters
- Trend removal: AR filter
- Resampling, interpolation

DELAY LINE FILTER (SCHNEGG AND FISCHER, 1980; BRASSE, 1993)

Delay by one:

$$y_n = x_n - x_{n-1}$$

16 2/3 Hz & 50 Hz:



B_x after denoising

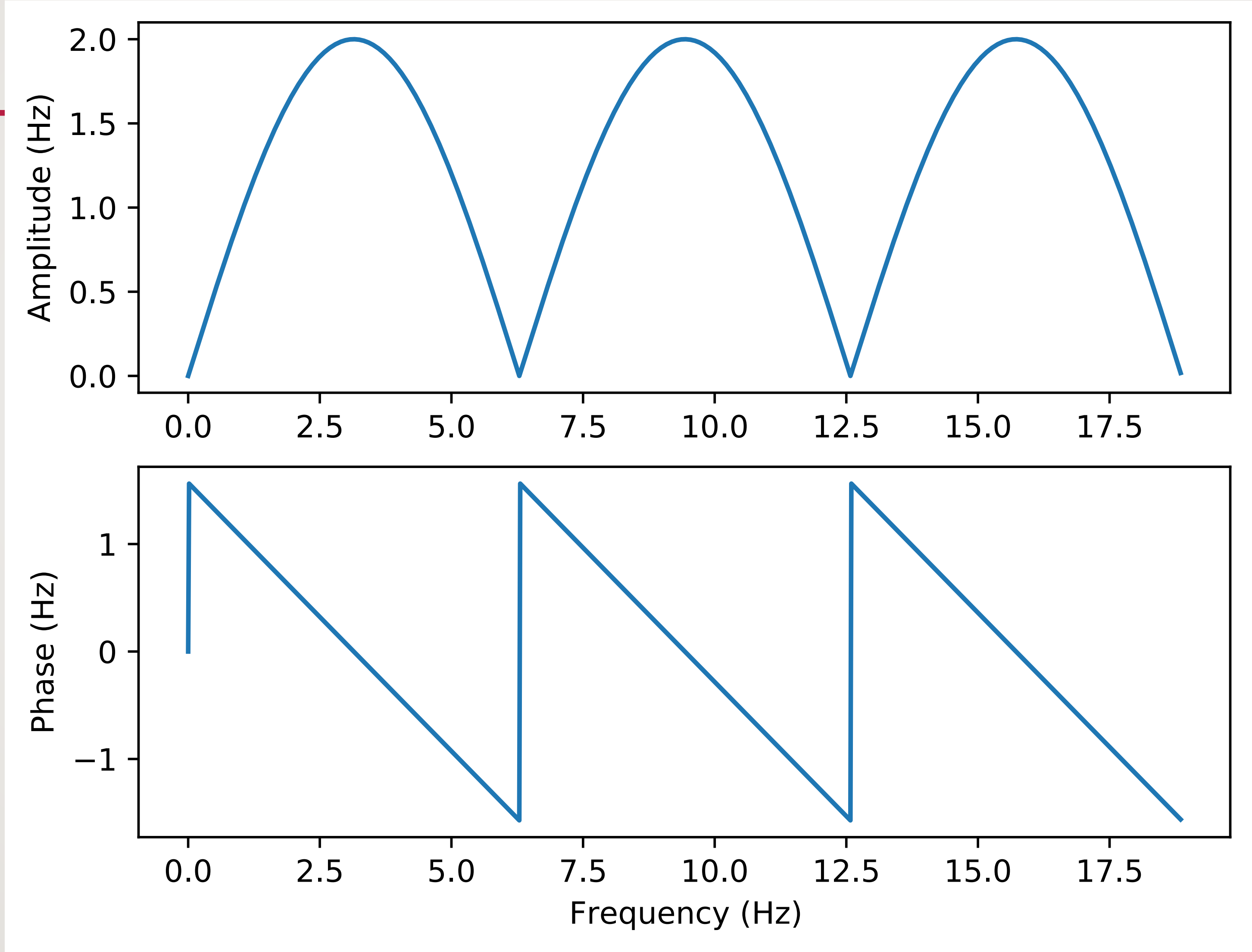
B_x before denoising

B_y after denoising

B_y before denoising

Transfer function:

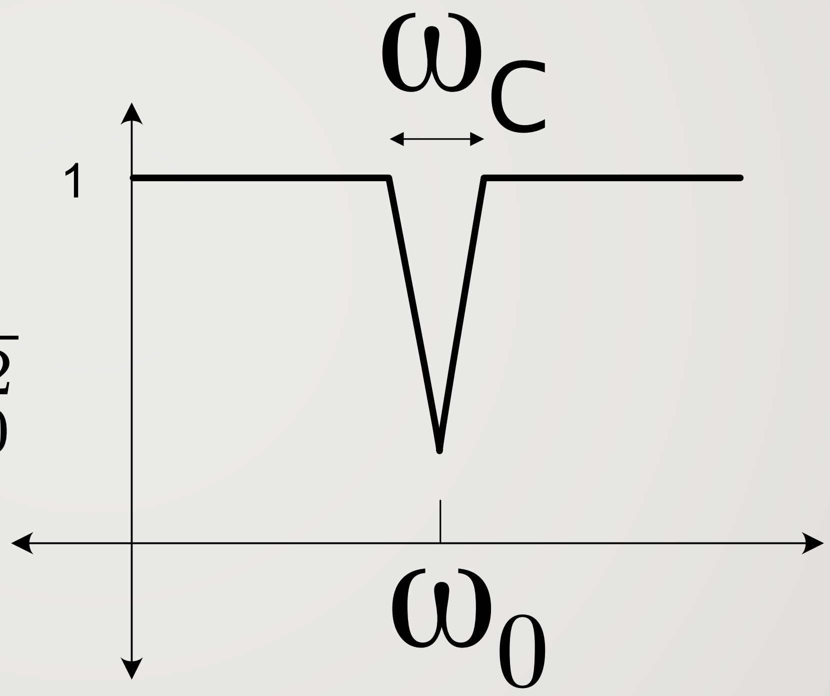
$$H(z) \stackrel{\text{def}}{=} \sum_{n=-\infty}^{\infty} h[n] \cdot z^{-n}$$



NOTCH FILTERS

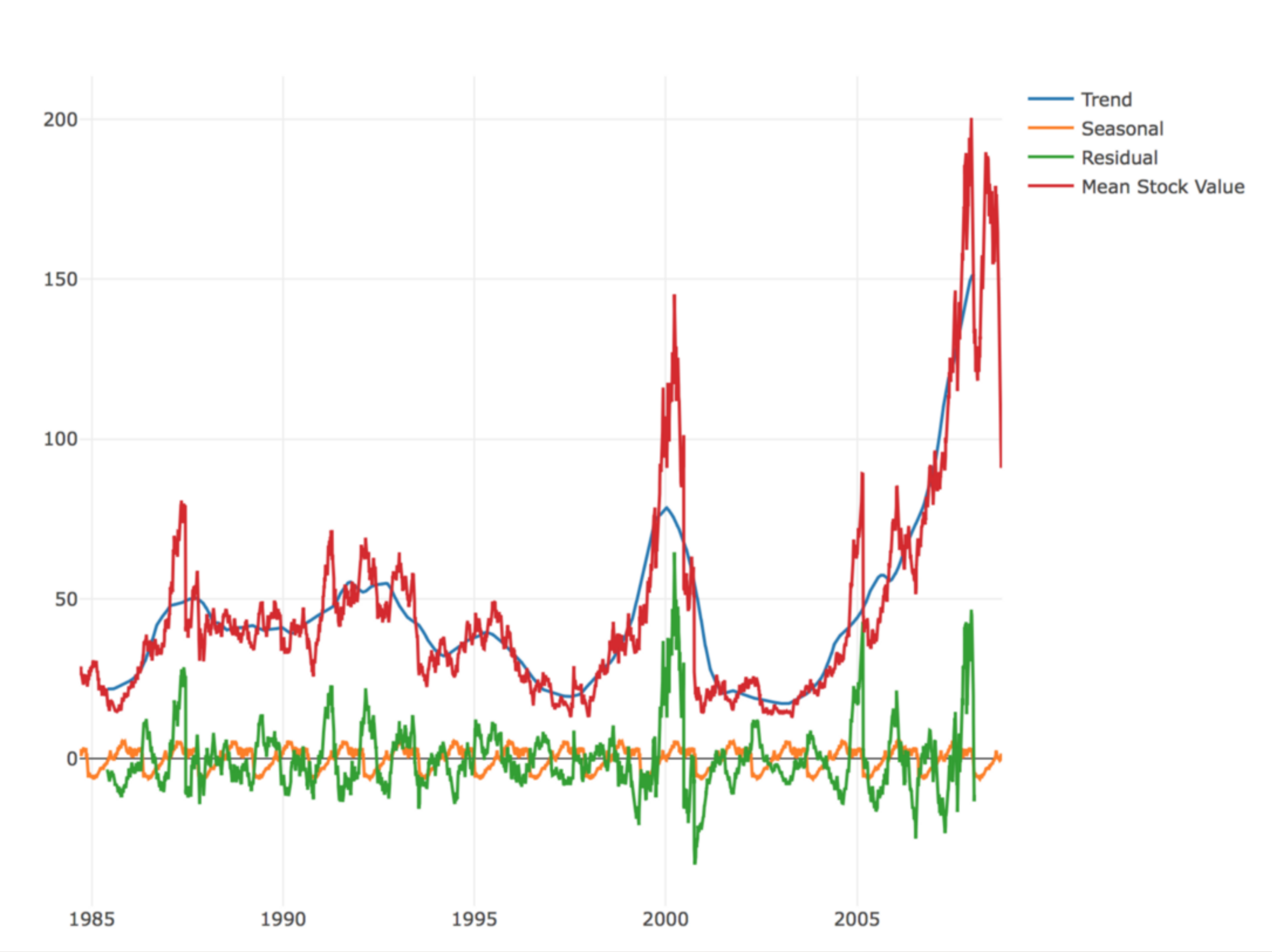
Transfer function:

$$H(z) = \frac{z^2 + \omega_0^2}{z^2 + \omega_c z + \omega_0^2}$$



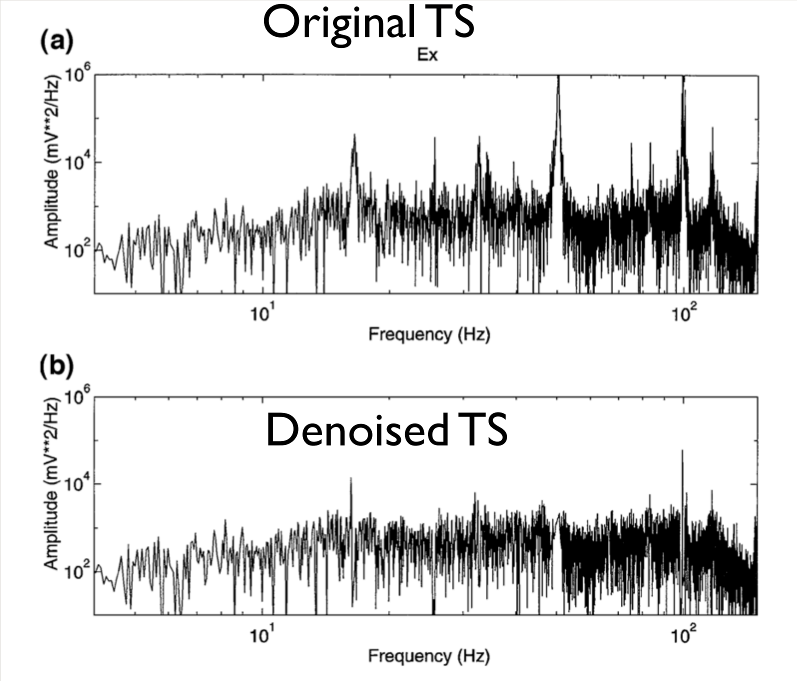
AR TREND REMOVAL (CHAVE, 2012)

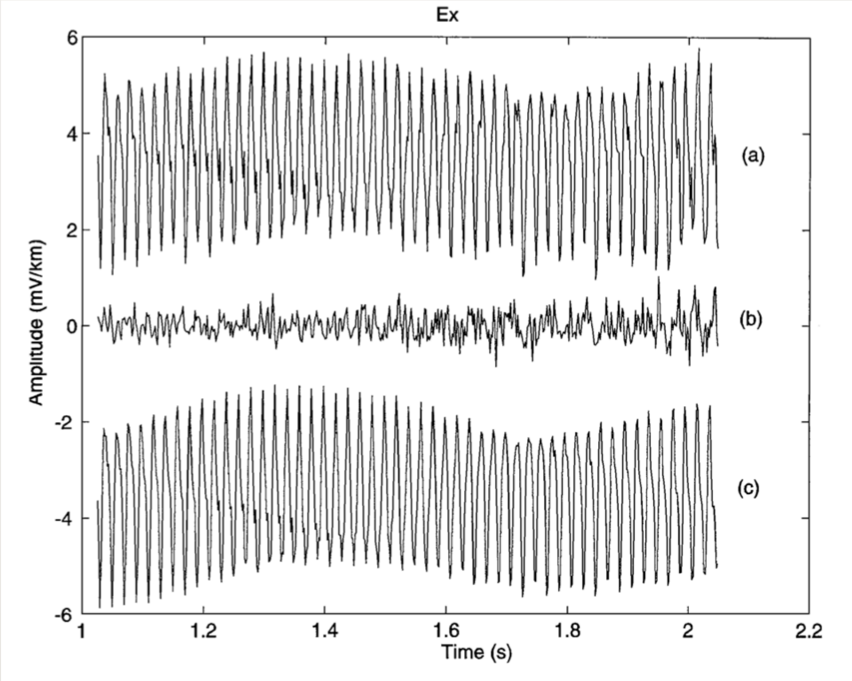
- AR (auto regression): is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature



WAVELET DENOISING
(TRAD & TRAVASSOS, 2000)

- Hypothesis: noise and signal can be sepparated at different scales
- Using Thomson weights, search for noisy segments and eliminate them





Signal + noise

Signal

Noise

USE OF ANALYTICAL SIGNAL (JONES AND SPRATT, 2002)

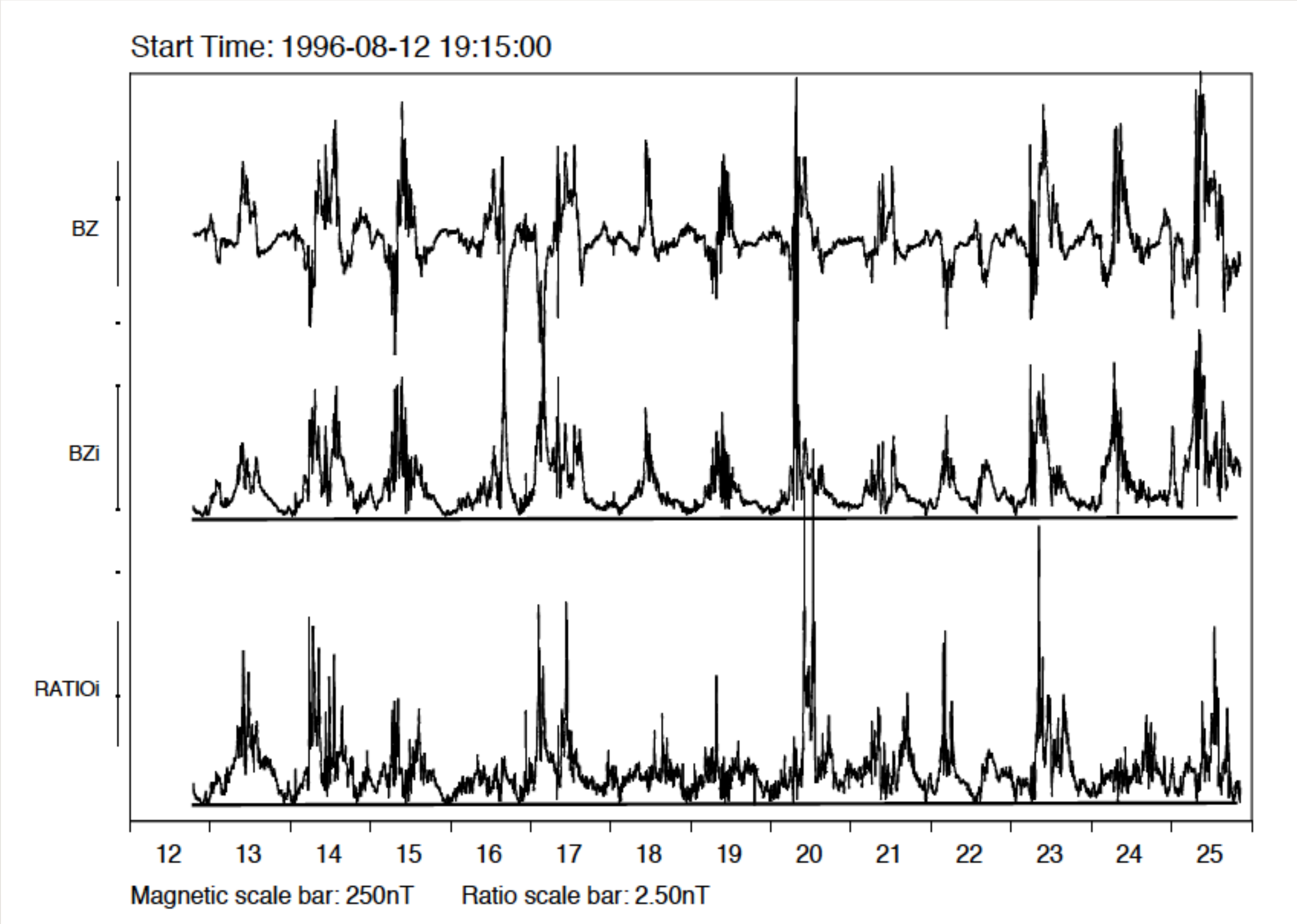
- Derive analytic signal:

$$z(t) = x(t) + i\mathcal{H}[x(t)]$$

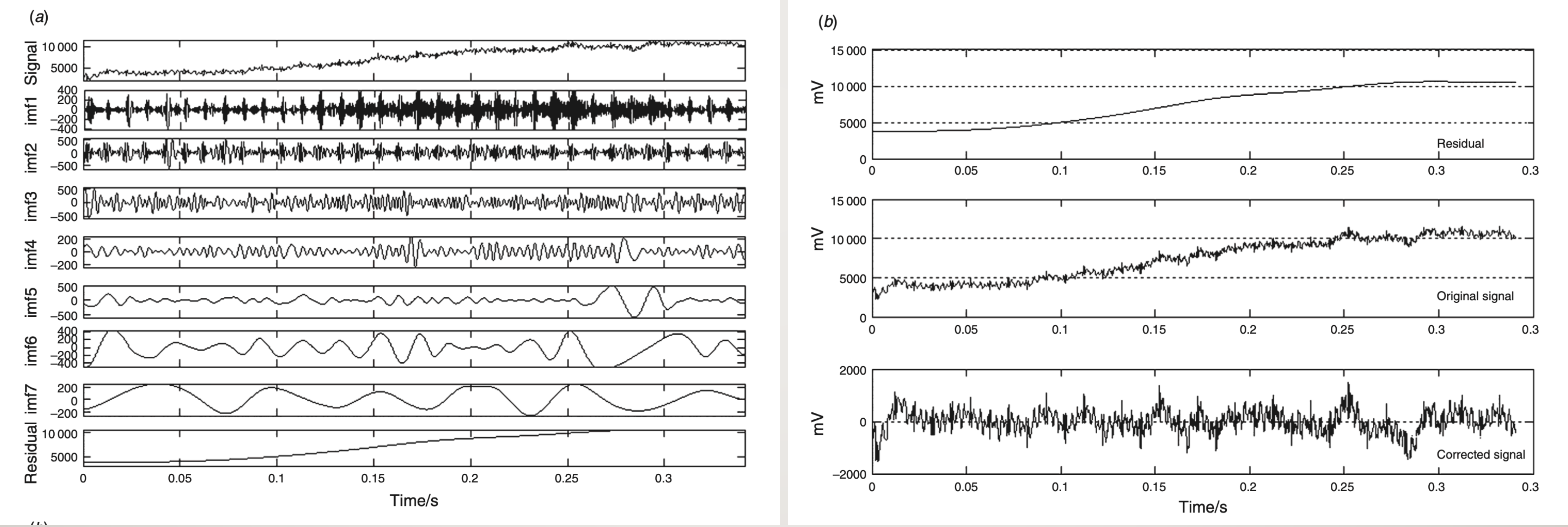
- Estimate ratio:

$$ratio = \frac{B_{zi}}{(B_{xi}^2 + B_{yi}^2)^{1/2}}$$

- Select segments
- Process TS

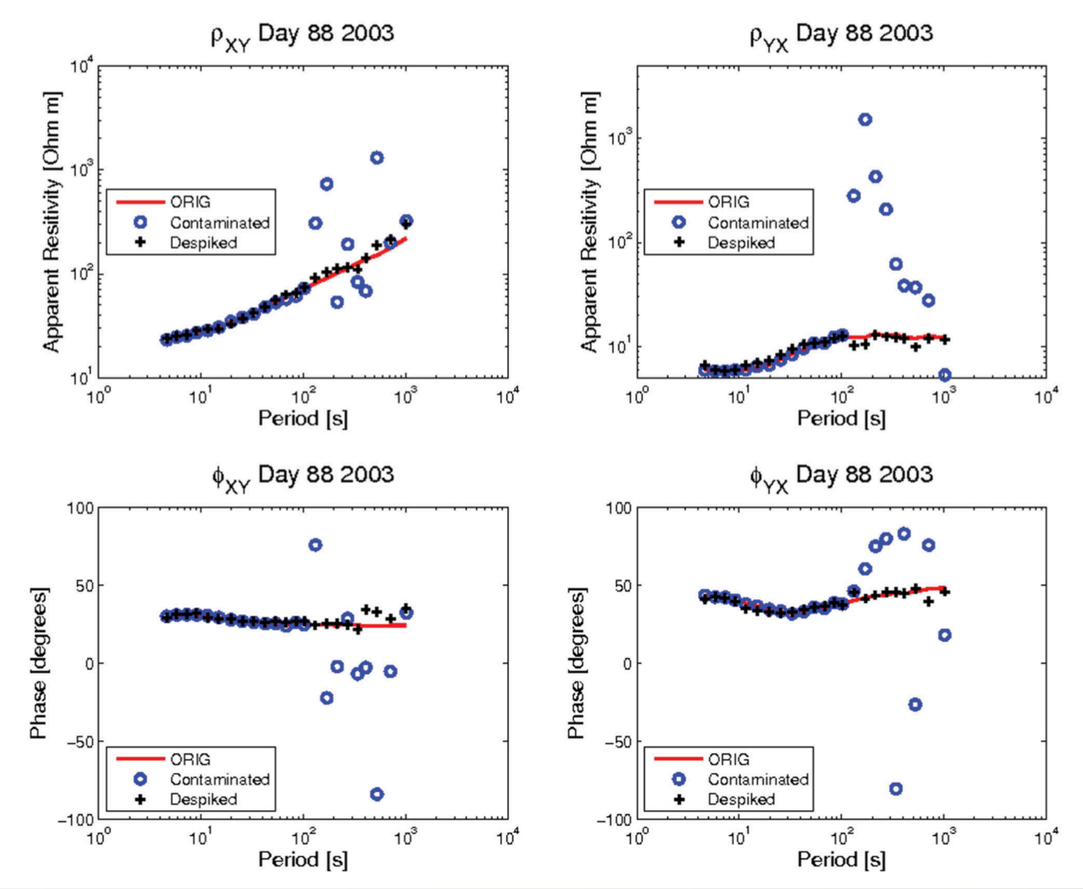


EMD DENOISING (CAI ET AL., 2019)



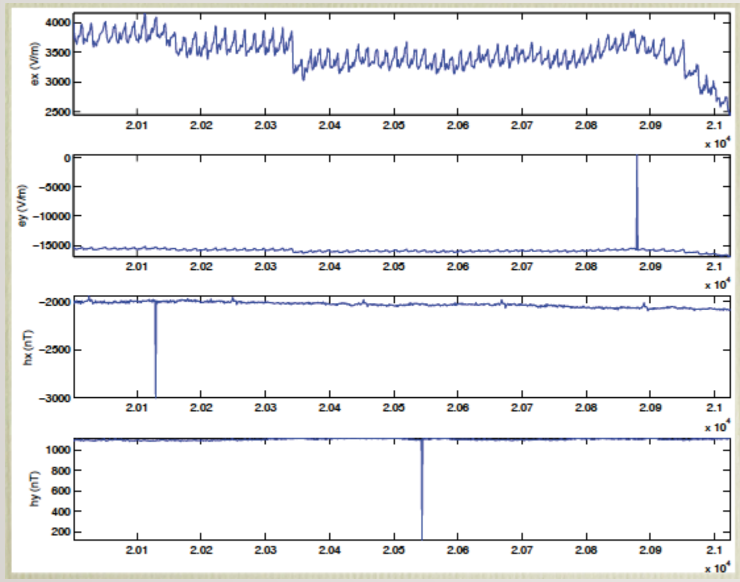
DATA VARIANCE TECHNIQUE FOR DESPIKING (KAPLER, 2012)

- Select clean segments
- Train a Wiener filter in clean segments with the aid of a remote reference
- Locate noisy sections
- Predict signal in noisy sections using Wiener filter and the remote reference
- Reconstruct signal and process

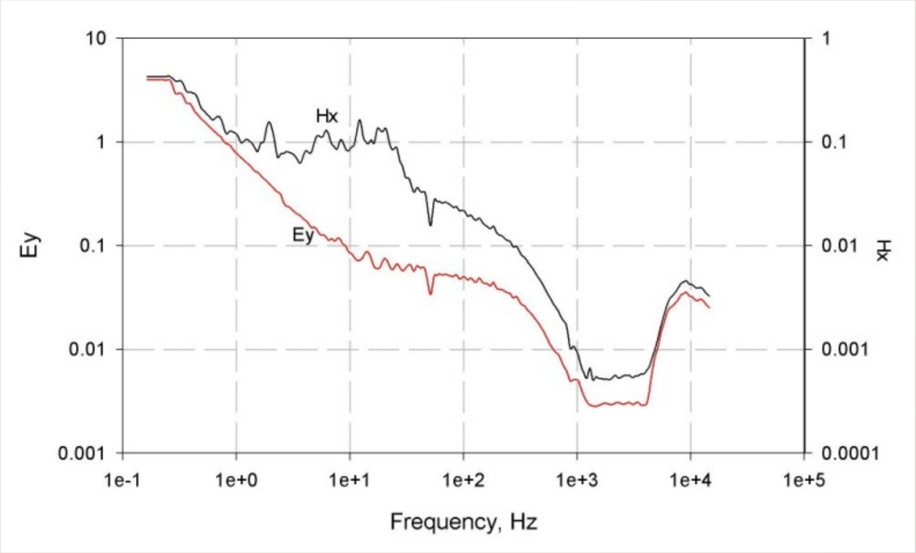


2. SPECTRAL ESTIMATION

It is desirable to be able to obtain a reasonable spectrum, with little or no bias and small uncertainties (Prieto et al., 2007)



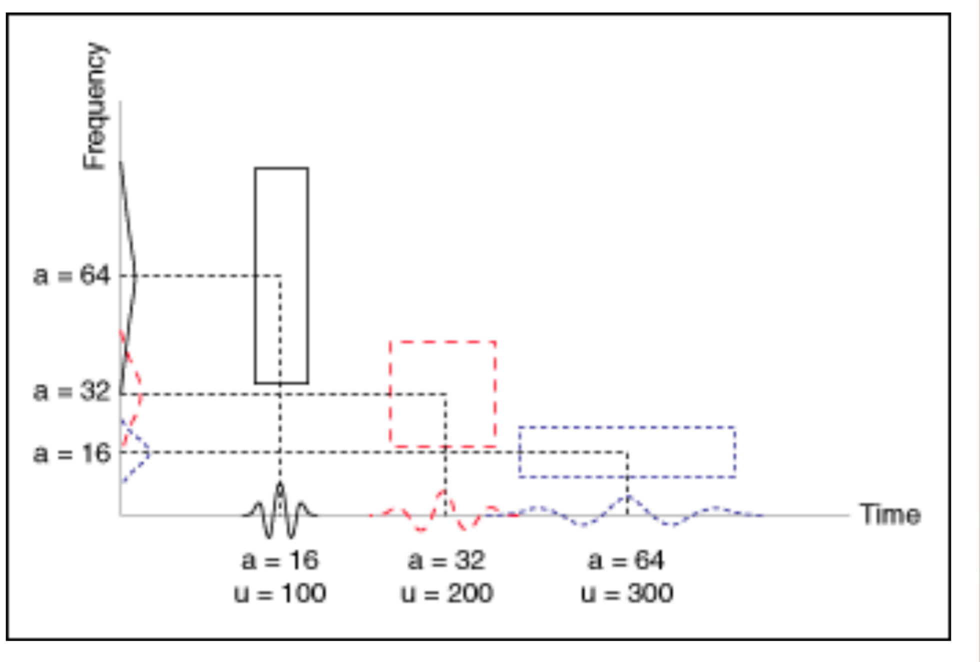
Fourier
Wavelet
EMD



NOTE

- The frequency information you obtain at one time point is a weighted sum of the frequency information at surrounding time points.
- In interpreting time-frequency results, each time point is an estimate of instantaneous activity influenced by neighboring activity.

$$\Delta t \Delta f \geq 1$$



FOURIER TECHNIQUES

- Conventional methods → Short-time Fourier Transform (STFT)

$$\mathbf{STFT}\{\mathbf{x}(\mathbf{t})\}(\tau,\omega) \equiv \mathbf{X}(\tau,\omega) = \int_{\infty}^{\infty} \mathbf{x}(\mathbf{t})\mathbf{w}(\mathbf{t}-\tau)\mathbf{e}^{-\mathbf{i}\omega\mathbf{t}}\mathbf{dt}$$

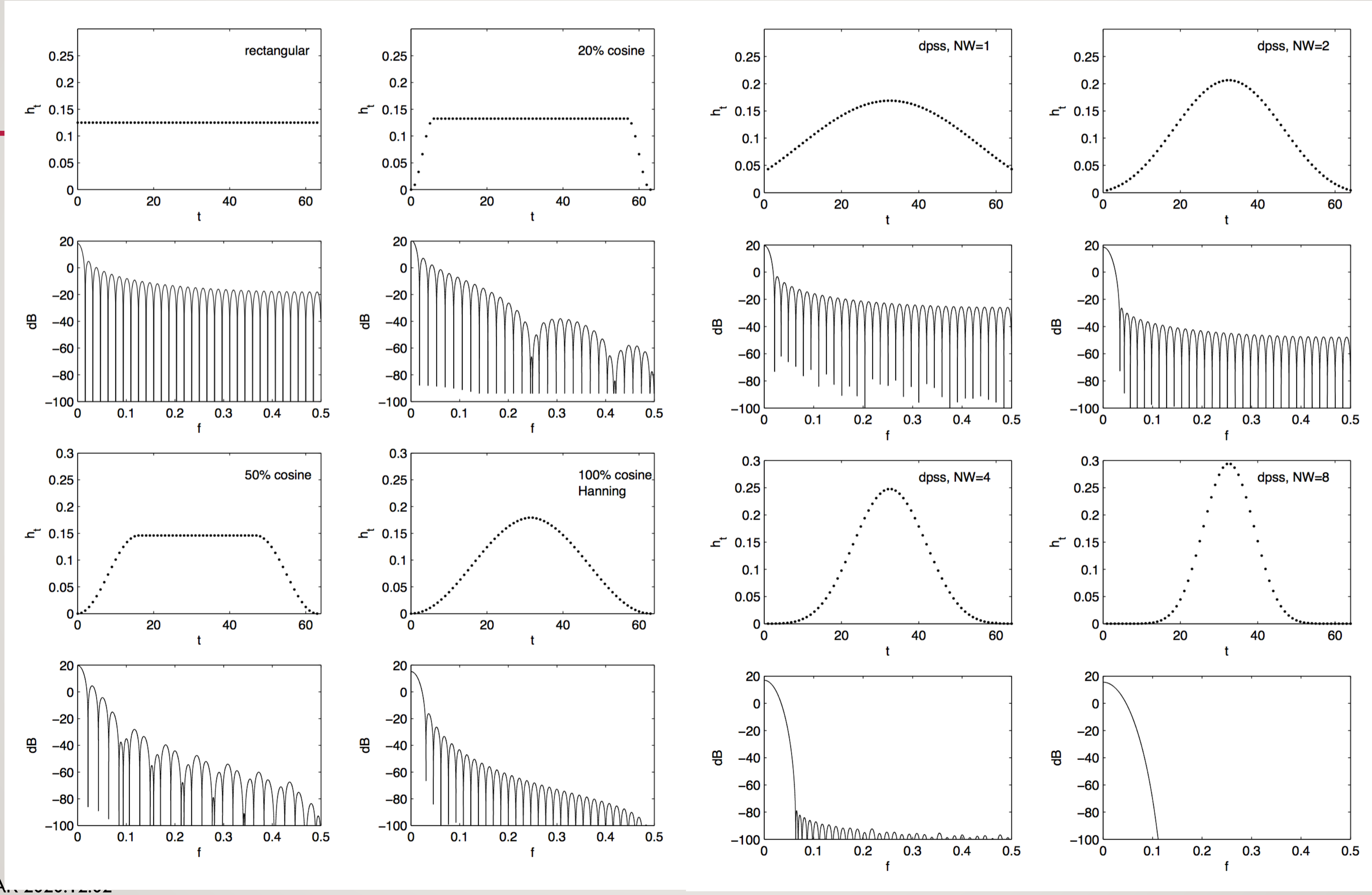
- Non-conventional methods → Cascade decimation (Wight and Bostick, 1979)
  - Use segments of 32 (64 or 128 …) points
  - Pre-condition with Hanning window
  - Use DFT to estimate spectra and select the 6th and 8th harmonic
  - Low pass data and decimate (by 2), start all over

FOURIER: WINDOWING (TAPER)

Reduce spectral leakage and bias caused by limited TS or by spectral peaks

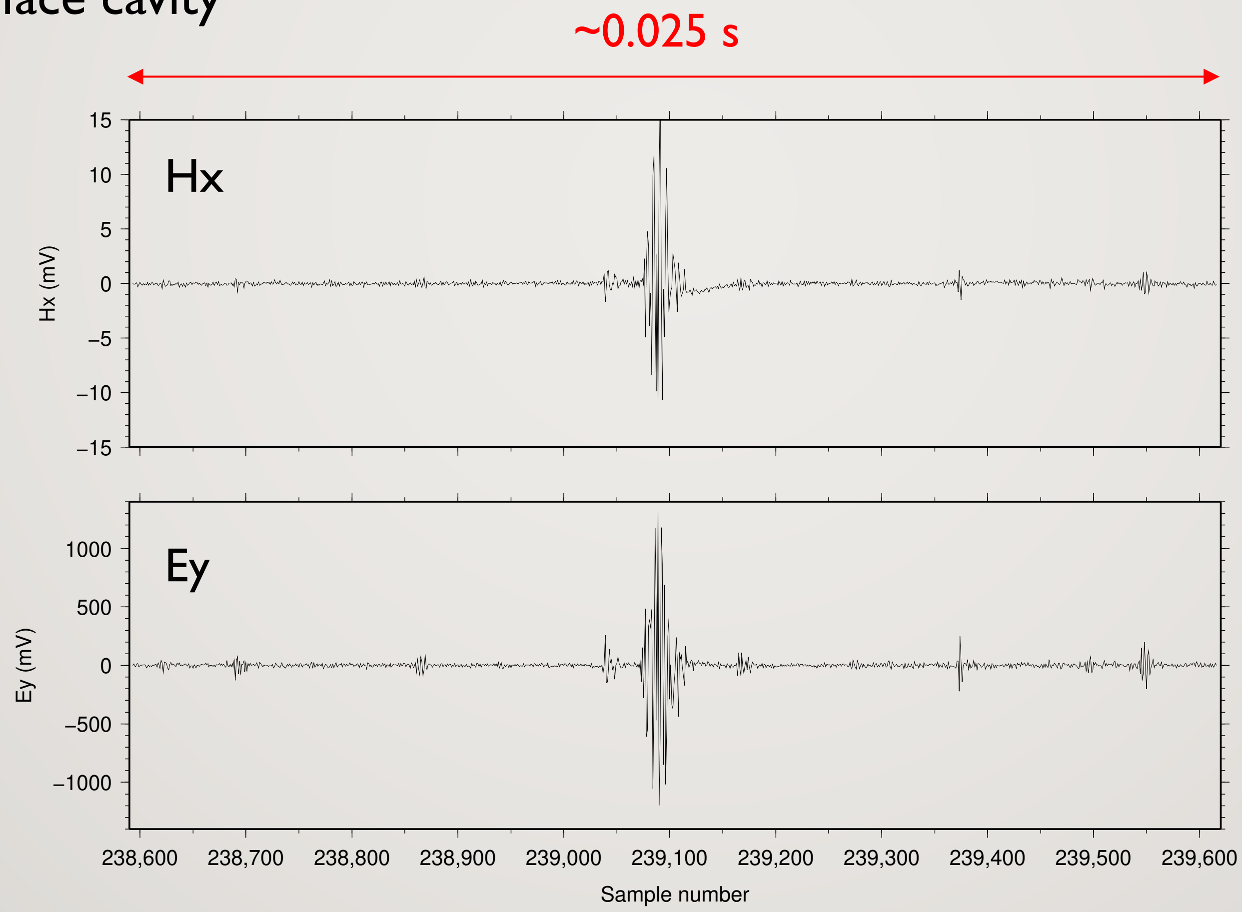
- Single window: Hann, Hamming, Parzen, Welch, Blackman, … (see wikipedia for more info)
- Multitaper: Discrete Prolate Spectral Sequence

TAPER


rectangular
20% cosine
dpss, NW=1
dpss, NW=2
50% cosine
100% cosine
Hanning
dpss, NW=4
dpss, NW=8
h_t
dB
t
f

WAVELET SPECTRAL ESTIMATION (GARCIA AND JONES, 2008)

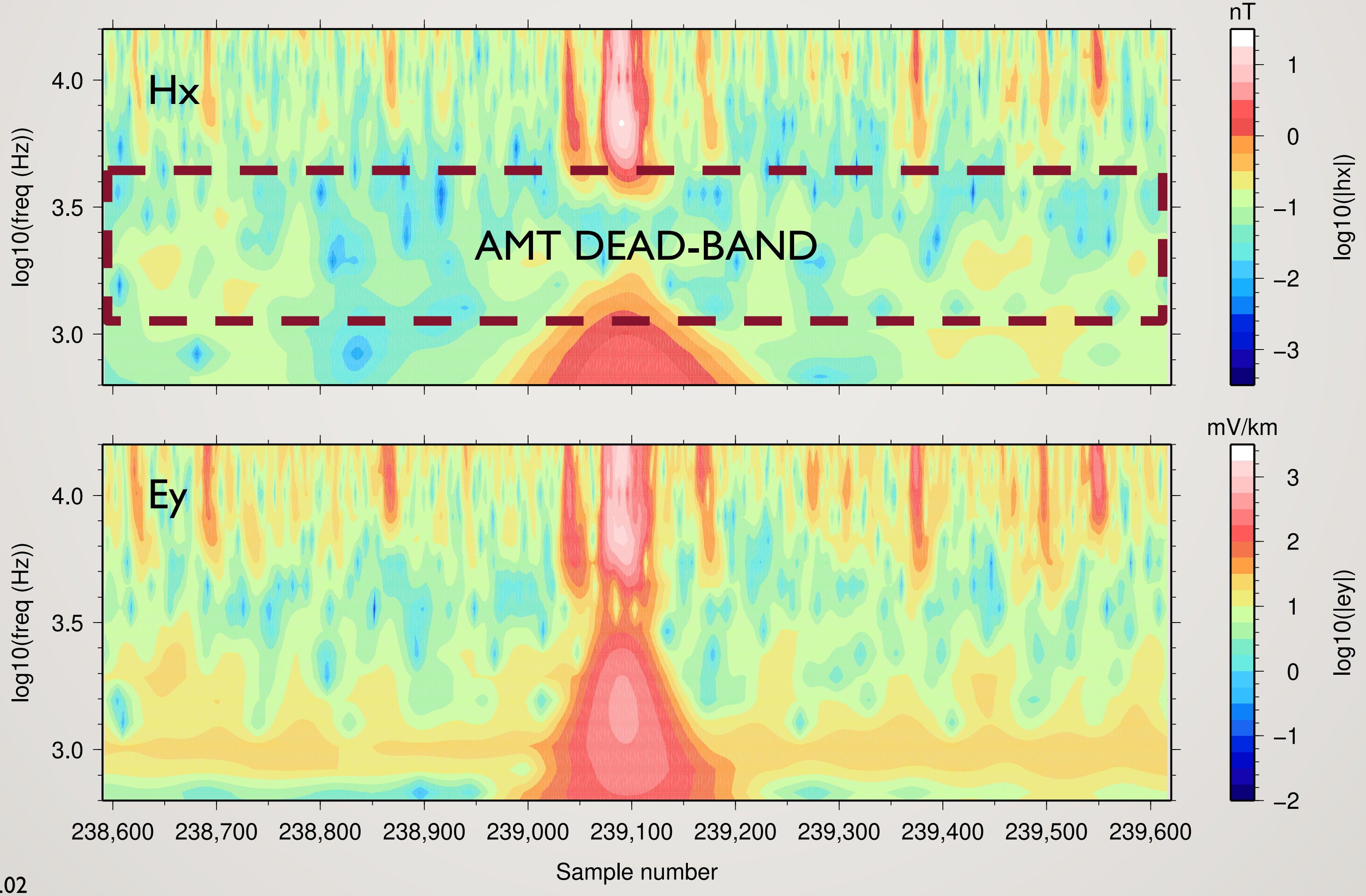
- AMT signal: global lightning discharges create plane waves that travel on the ionosphere-Earth's surface cavity



AMT SPECTRA

Blue: low amplitude

Red: high amplitude



CONTINUOUS WAVELET TRANSFORM

- Goal: locate high energy events
- Varying parameters allow for Time-Frequency tradeoffs in precision
- Similar to decomposition with sine waves of different frequencies in FFT: Time-Frequency decomposition involves wavelets of different frequencies
- To make a family you change the frequency of the sine wave while leaving other parameters unchanged

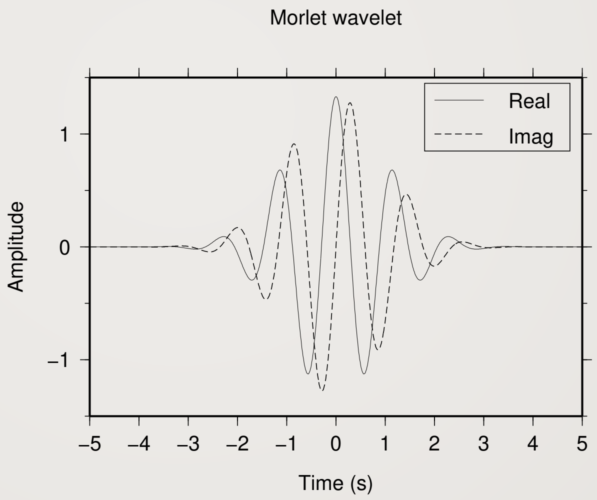
MORLET WAVELET

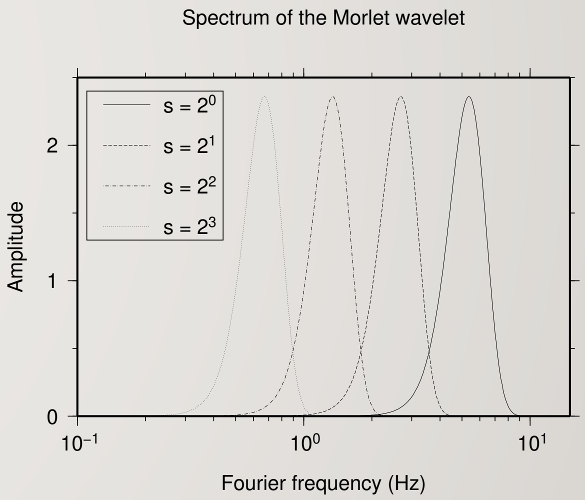
$$\psi_0(s) = \pi^{1/4} e^{i\omega_0 s} e^{-s^2/2}$$

$$\omega_0 = \pi\sqrt{2/\log(2)}$$

Period definition:

$$T = \frac{4\pi s}{\omega_0 + \sqrt{2 + \omega_0}}$$





STATIONARITY OF MT TS

- Uman and Rakov (2007), lightning discharges:
  - fast and transient leader-return stroke sequences
  - slow and quasi-stationary continuing currents
  - **perturbations and surges on the continuing currents**

May be viewed as being stationary on a section with some dynamic length confined by the recurrent transient strokes

- Liu and Fujimoto (2011):
  - magnetospheric current is nonlinearly driven by the dynamic solar wind but behaves in a static manner for high magnetospheric pressure conditions

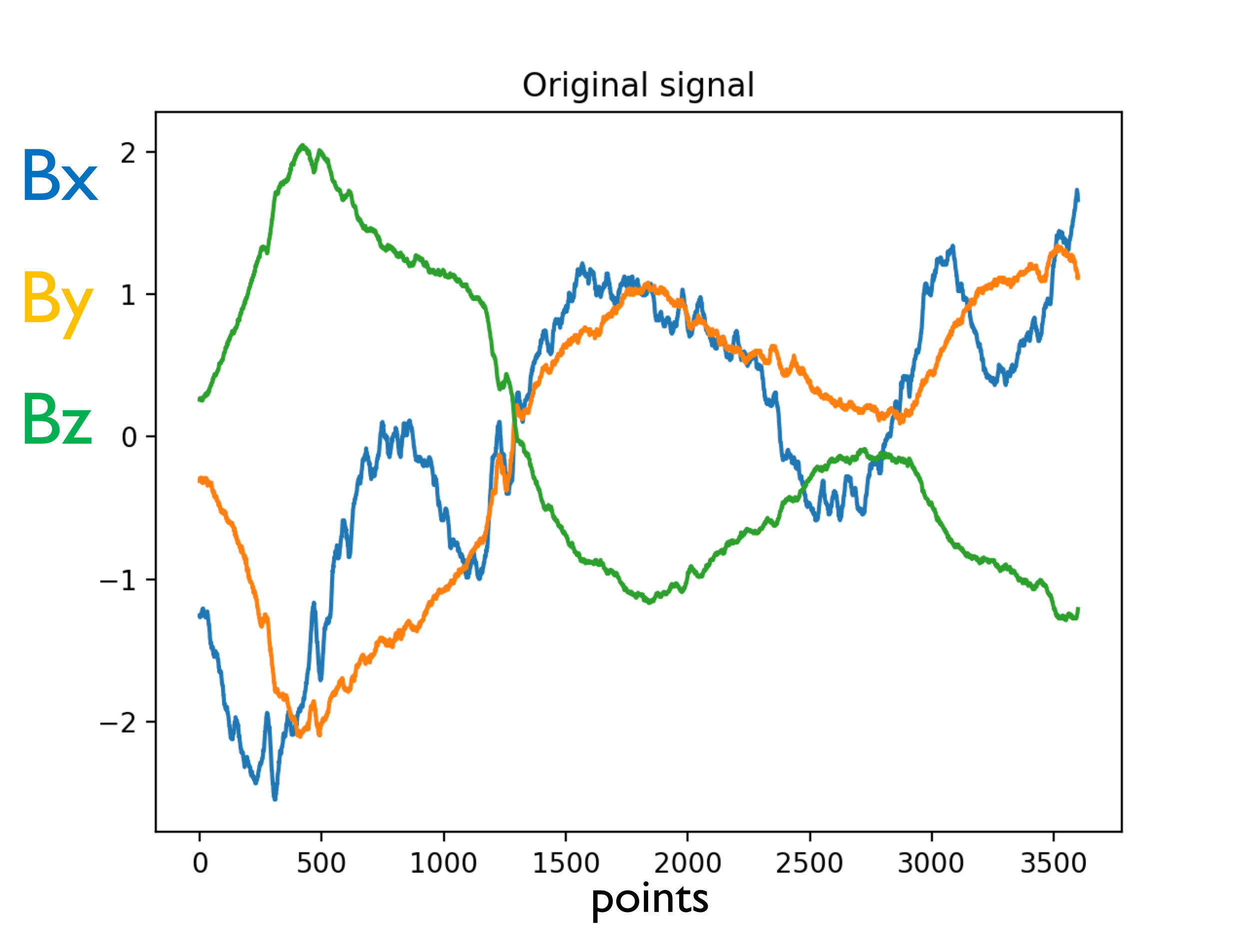
STATIONARITY OF MT TS (2) (NEUKIRCH AND GARCIA, 2014)

- EM sources are naturally nonstationary, since both, lightning strokes and magnetospheric pressure conditions, are very dynamic and thus strictly limit the duration of any stationary electromagnetic signal (Neukirch and Garcia, 2014)
- Noise sources do not need to be stationary, e.g., a station near a road (Adam et al., 1986)
- Try a tool that does not assume stationarity of TS

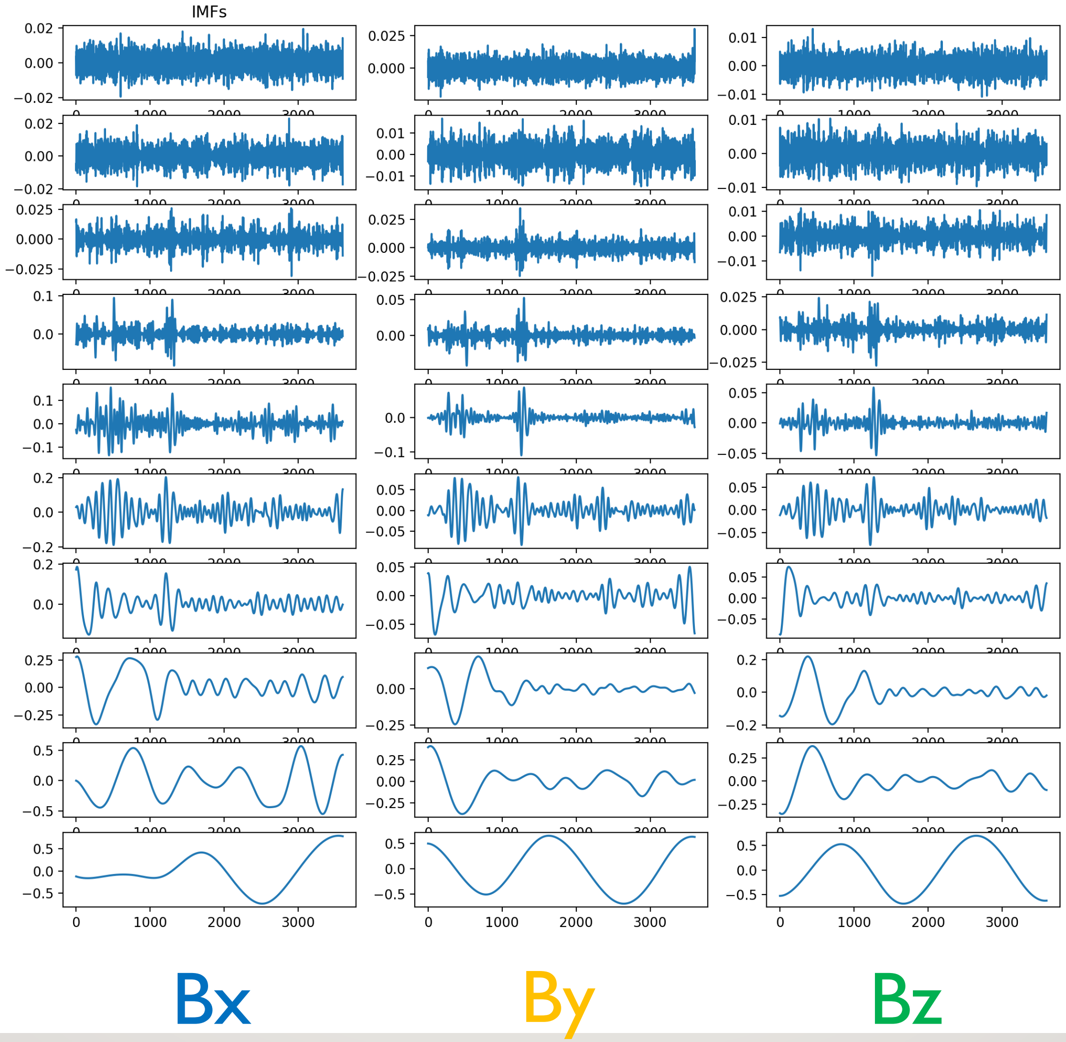
MEMD: MULTIEMPIRICAL MODE DECOMPOSITION

- Data based
- Does not make assumptions on data stationarity
- Based on the Hilbert-Huang transform
- Multivariate version of EMD
- Decompose TS into Intrinsic Mode Functions (IMF)
- IMF is a time series with a dynamic and locally narrow banded IF (*Flandrin and Rilling*, 2004)
- Offers great time-frequency localization
- BUT it suffers from high sensitivity to the signal-to-noise ratio
- EMD and MEMD spectra equivalent to Fourier spectra (*Neukirch and Garcia*, 2013)

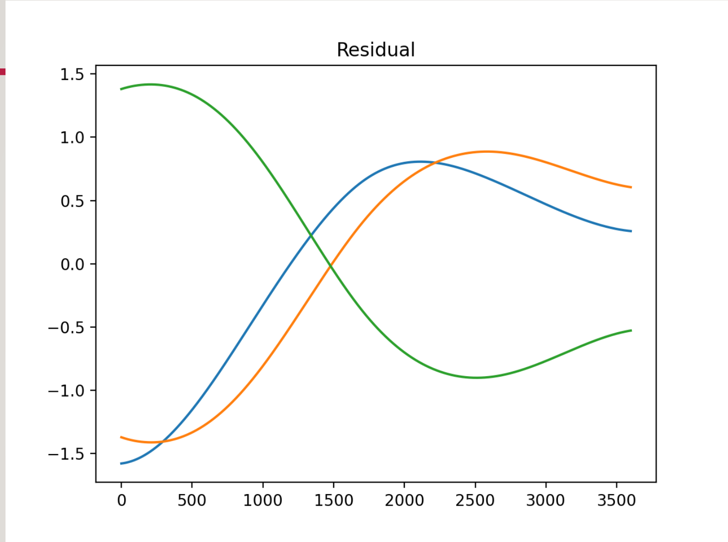
EXAMPLE: TIPPER DATA FROM LEMI INSTRUMENT MEASURED IN NATAL (BRAZIL)

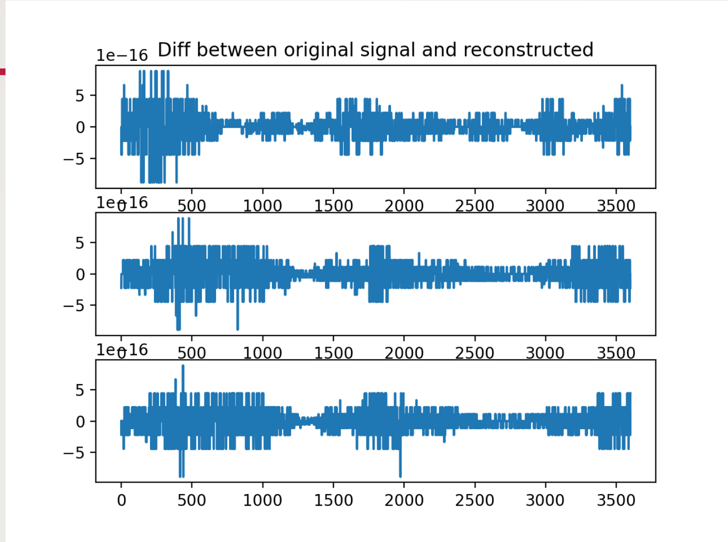


IMF DECOMPOSITION



RECONSTRUCTED SIGNAL …

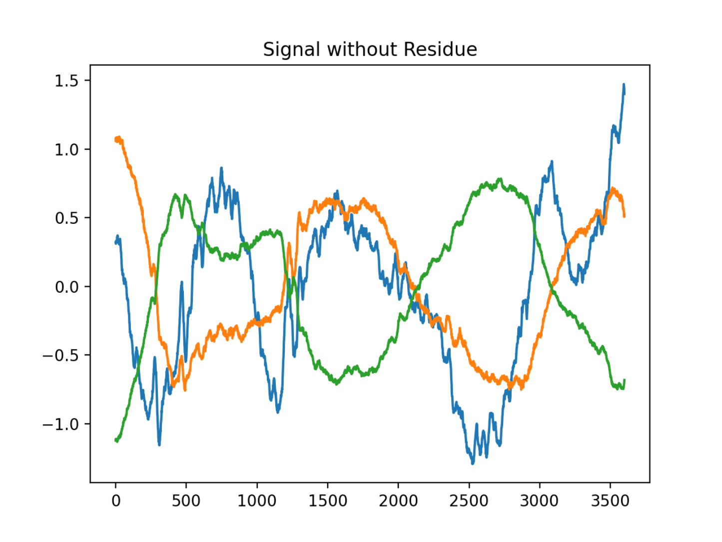




Bx

By

Bz



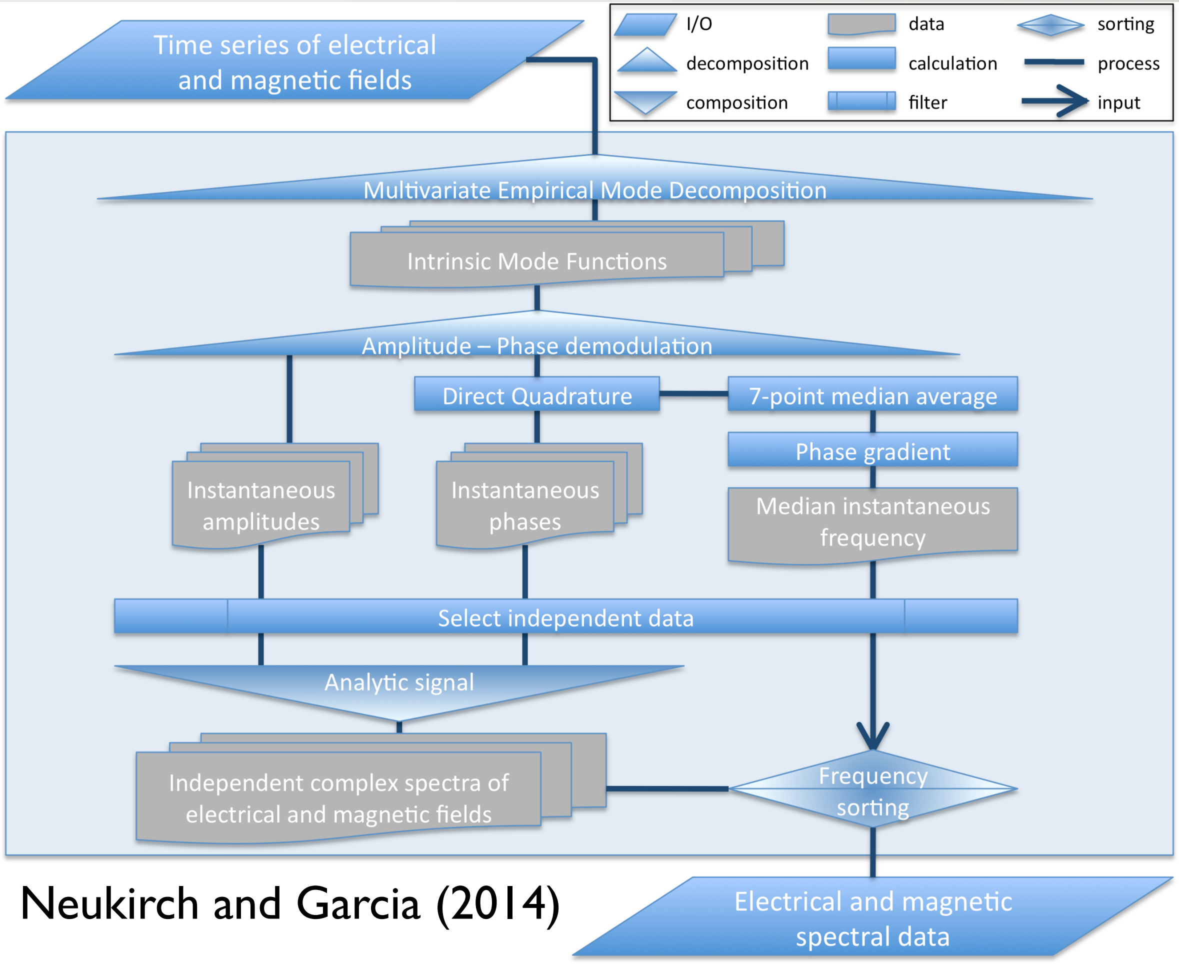
IMF → SIGNAL + FREQUENCY

EMD

- Chen and Jegen (2008) suggest to use the marginal spectra for each IMF
- Chen et al (2012) use instantaneous parameters (Huang et al, 2008): common Instantaneous Frequency in all channels

MEMD

- Neukirch and Garcia (2014) use multivariate EMD and then similar strategy as Chen et al (2012)



Neukirch and Garcia (2014)

3. IMPEDANCE ESTIMATION

- Tranformed TS into spectral domain, solve:

$$\mathbf{e} = \mathbf{z}\mathbf{b}$$

- Can be solved using OLS, converting it to a *linear regression model*

$$\mathbf{e_i} = \mathbf{b_i}\mathbf{z} + \epsilon_\mathbf{i}$$

- And solving the objective function *S*:

$$\hat{z} = \underset{z}{\arg\min}\, S(z)$$

$$S = \|\mathbf{e} - \mathbf{b}\mathbf{z}\|$$

OLS

- Minimize *S* by solving the normal equations

$$\left(\mathbf{b}^{\mathbf{T}}\mathbf{b}\right)\hat{\mathbf{z}} = \mathbf{b}^{\mathbf{T}}\mathbf{e}$$

$$\hat{\mathbf{z}} = \left(\mathbf{b}^{\mathbf{T}}\mathbf{b}\right)^{-1}\mathbf{b}^{\mathbf{T}}\mathbf{e}$$

- but $\boldsymbol{z}$ is correlated with ε, therefore OLS in inconsistent

REMOTE REFERENCE (AKA, INSTRUMENTAL VARIABLES METHOD)

- Generally used as presented by Gamble et al (1979) using magnetic channels
- Magnetic fields suffer less noise than electric ones and are less affected by local inhomogeneities
- **Requirements:**
  1. Number of RR has to be equal or larger than independent variables, for MT: $\geq$ **2**
  2. RR must be exogenous (valid):

$$Cov(\mathbf{r}, \epsilon) = \mathbf{0}$$

  3. RR must be relevant: must be correlated to each of the endogenous regressors ***b***

2SLS: 2-STAGE LEAST SQUARES

- Solve the RR problem by first solving the regression between local channels and the remotes (1st stage), then substituting these predicted values back into the normal equation (2nd stage)

$$\hat{\mathbf{b}}_i = \mathbf{r_i}\pi + \nu_\mathbf{i}$$

Since ***r· ε = 0*** and assuming any number of RR larger or equal than 2:

$$\mathbf{H_r} = \mathbf{r} \cdot (\mathbf{r^T} \cdot \mathbf{r})^{-1} \mathbf{r^T}$$

$$\breve{\mathbf{b}} = \mathbf{r} \cdot (\mathbf{r^T} \cdot \mathbf{r})^{-1} \mathbf{r^T} \cdot \mathbf{b}$$

$$\hat{\mathbf{z}}_{RR} = (\breve{\mathbf{b}}^\mathbf{T} \cdot \mathbf{b})^{-1} \breve{\mathbf{b}}^\mathbf{T} \cdot \mathbf{e}$$

WARNING

Chave and Thomson (1989):

The Remote Reference method can help in MT data processing but is not ***robust****, and in the presence of severe contamination or uncorrelated channels it will fail at providing reliable responses*

ROBUST LEAST SQUARES

- **ROBUST**: Identify and remove or damp the effect of outliers to make estimate insensitive to their presence

  →that means that it is not greatly affected by outliers and responds slowly to addition of more data

- Introduced by Huber in late 60s
- MT:
  - Jones and Jödicke (1984): LTS
  - Egbert and Booker (1986): regression M-estimator
  - Chave et al (1987): regression M-estimator
  - Larsen (1989): D+ time-series event picking

Iterative Reweighted Least Squares (IRWLS)

IRWLS – ITERATIVE REWEIGHTED LEAST SQUARES

- Recall that OLS minimizes:

$$\hat{z} = \arg\min_{z} S(z)$$

$$S = \|\mathbf{e} - \mathbf{bz}\|$$

- In IRWLS, the function to minimize is:

$$\arg\min_{z} \sum_{i=1}^{n} |y_i - f_i(z)|^p$$

- Solved iteratively as:

$$z^{t+1} = \arg\min_{z} \sum_{i=1}^{n} \omega(z^t)\, |y_i - f_i(z)|^2$$

IRWLS MATRIX FORM

- For no RR, only single site processing:

$$\hat{\mathbf{z}} = (\mathbf{b}^{\mathbf{T}} \cdot \omega \cdot \mathbf{b})^{-\mathbf{1}} \mathbf{b}^{\mathbf{T}} \cdot \omega \cdot \mathbf{e}$$

where ω is a diagonal matrix containing the Robust weights

IRWLS

- IRWLS is used to find the maximum likelihood of a generalized linear model, or what is the same, to find an M-estimator that minimizes the effect of outliers
- For MT, the solution would be:

$$z^{t+1} = \arg\min_{z} \sum_{i=1}^{n} \omega(z^t) \left| e_i - h_i z \right|^2 = \left(h^T W^{(t)} h \right)^{-1} h^T W^{(t)} e$$

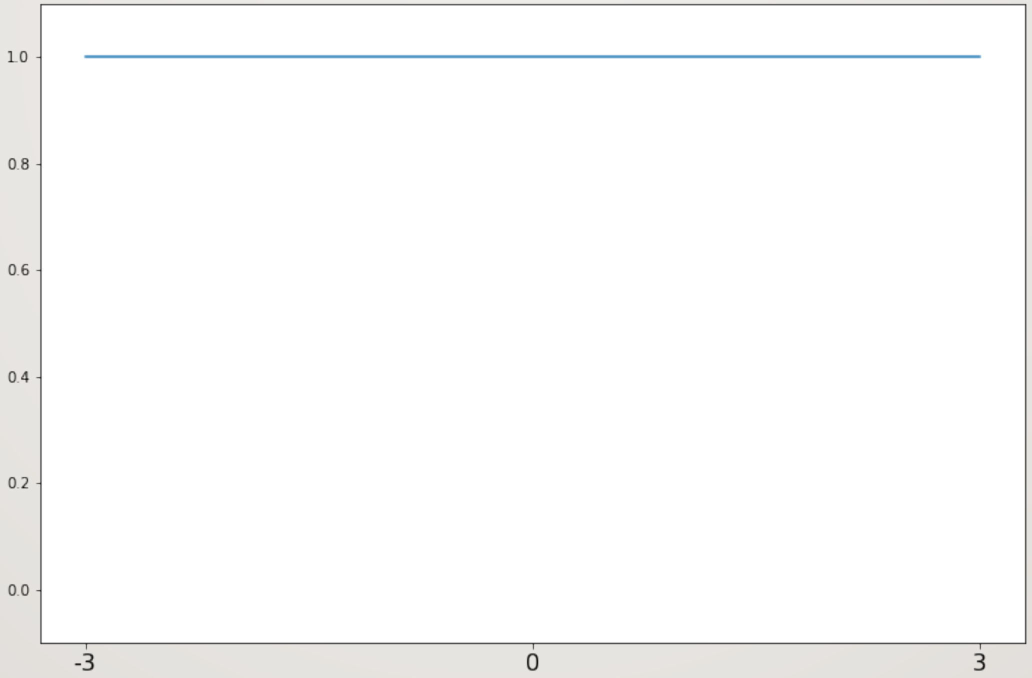
$$\omega^{(0)} = 1$$

$$\omega^{(t)} = \sum_i \rho \left(\frac{e_i}{s} \right)$$

ρ : weight function of choice

WEIGHT FUNCTIONS: OLS

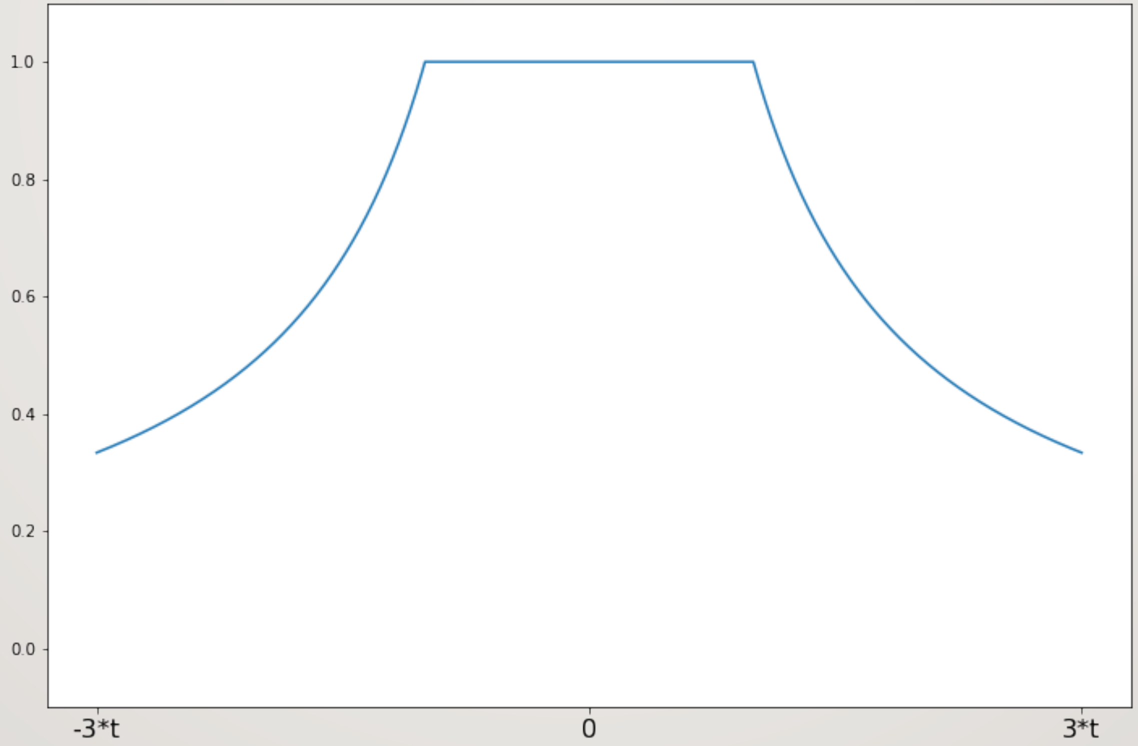
$$\omega = 1$$



WEIGHT FUNCTIONS: HUBER'S T

$$\omega^{(t+1)} = \begin{cases} 1 & |e^{(t)}| = t \\ \frac{t}{|e^{(t)}|} & |e^{(t)}| > t \end{cases}$$

t = 1.345 or 1.5

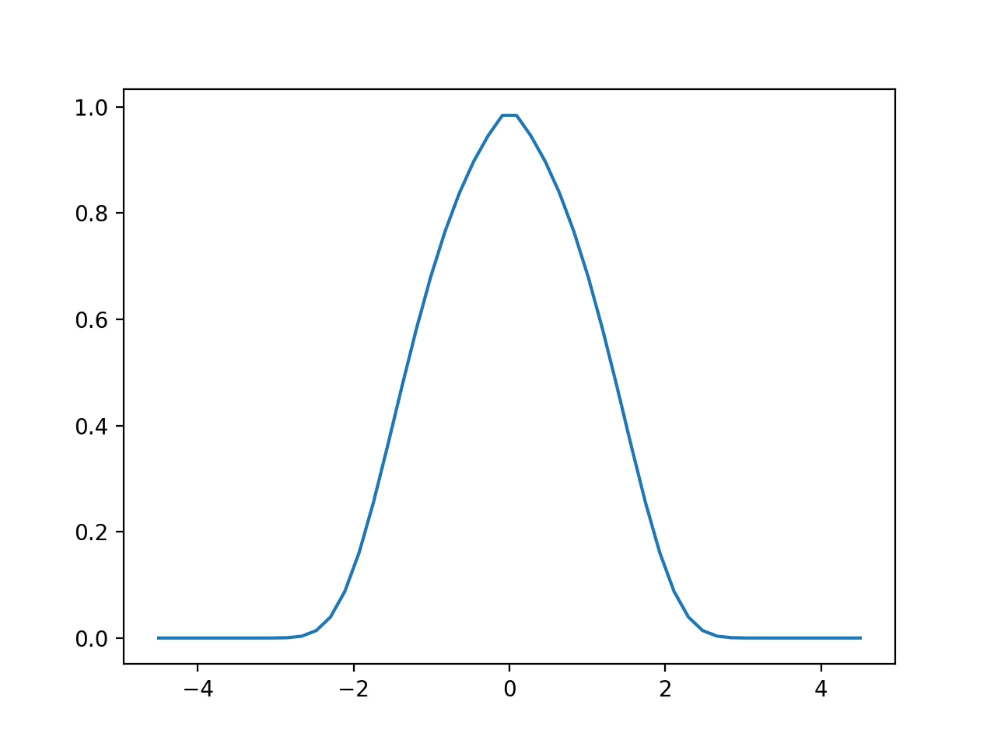


WEIGHT FUNCTION: THOMSON

$$\omega^{(t+1)} = \omega^{(t)} exp(e^{-\xi^2}) exp(-e^{\xi(|\epsilon^{(t)}|-\xi)})$$

$$\xi = \sqrt{2log(2N)}$$ N: *Nth quantile of the Rayleigh distribution*

$$\omega(e)^{(0)} = 1$$



IRWLS + RR

- IRWLS:

$$\hat{\mathbf{z}} = (\mathbf{b}^{\mathbf{T}} \cdot \omega \cdot \mathbf{b})^{-\mathbf{1}} \mathbf{b}^{\mathbf{T}} \cdot \omega \cdot \mathbf{e}$$

- RR:

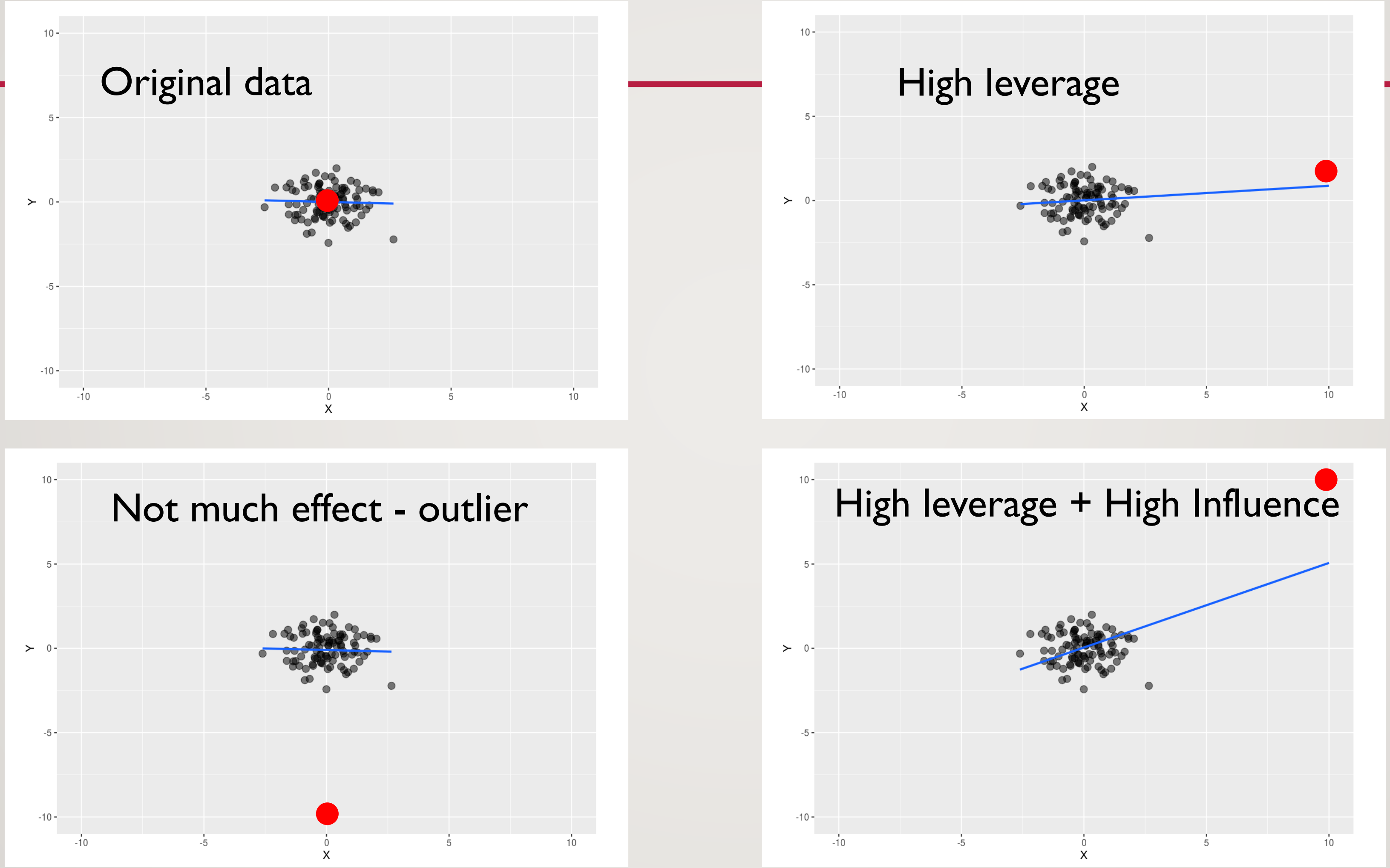
$$\hat{\mathbf{z}}_{RR} = (\breve{\mathbf{b}}^{\mathbf{T}} \cdot \mathbf{b})^{-\mathbf{1}} \breve{\mathbf{b}}^{\mathbf{T}} \cdot \mathbf{e}$$

- IRWLS + RR:

$$\hat{\mathbf{z}}_{RR} = (\breve{\mathbf{b}}^{\mathbf{T}} \cdot \omega \cdot \mathbf{b})^{-\mathbf{1}} \breve{\mathbf{b}}^{\mathbf{T}} \cdot \omega \cdot \mathbf{e}$$

$$\breve{\mathbf{b}} = \mathbf{r} \cdot (\mathbf{r}^{\mathbf{T}} \cdot \mathbf{r})^{-\mathbf{1}} \mathbf{r}^{\mathbf{T}} \cdot \mathbf{b}$$

LEVERAGE AND INFLUENCE



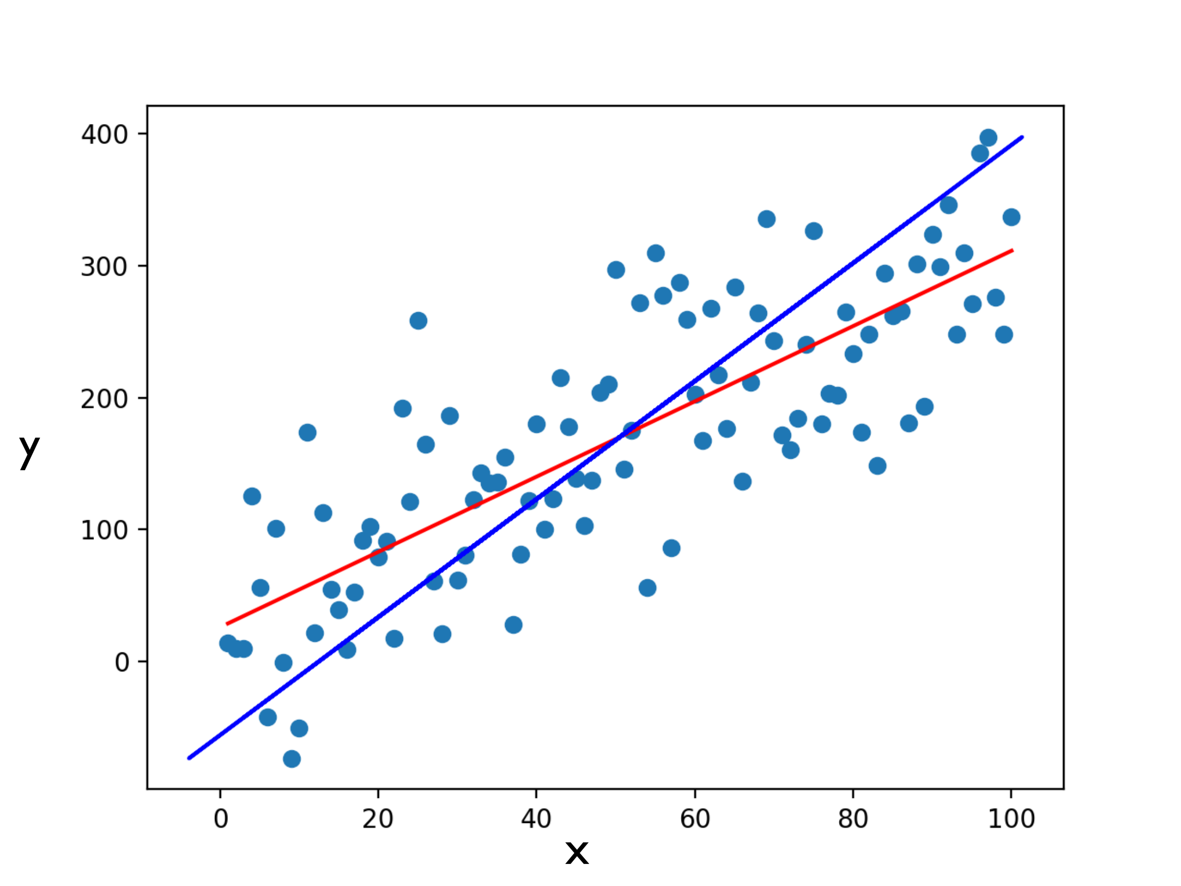
https://omaymas.github.io/InfluenceAnalysis/

WHAT ABOUT PCA?

(*EGBERT*, 1997, 2002; *SMIRNOV AND EGBERT*, 2012; NEUKIRCH AND GARCIA, 2014)

- *Egbert* (1997) shows that MT sources are well described by two electromagnetic field polarizations
- The entire data vector space of all channels in a data set can be represented by the combination of two polarization vectors
- Find the 2 PCs that better represent the source
- BUT the fields are a combination of signal and noise
- **Multivariate-Multisite** approach helps

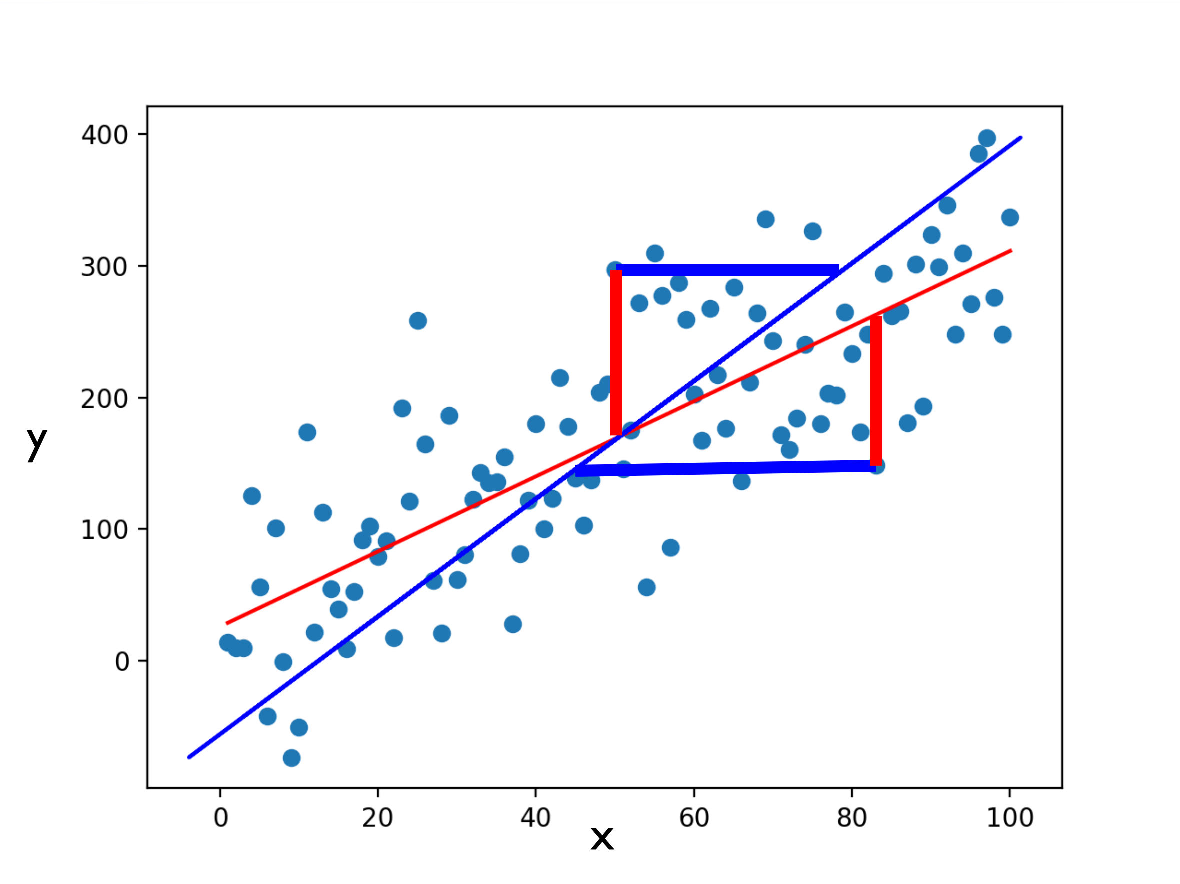
PCA VS OLS:
I – OLS



$$y = c \cdot x + m$$

$$x = d \cdot y + n$$

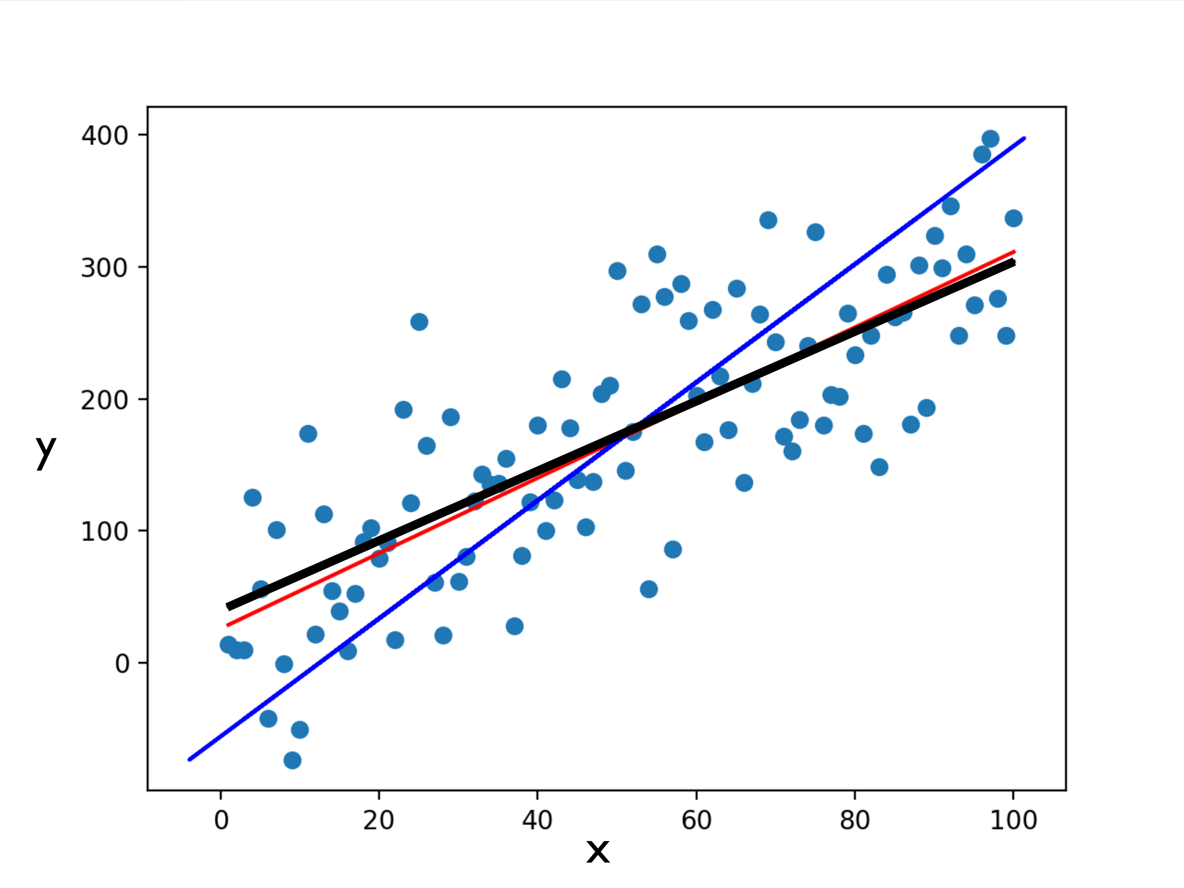
PCA VS OLS:
I – OLS (2)



$$y = c \cdot x + m$$

$$x = d \cdot y + n$$

PCA VS OLS:
2 – PCA



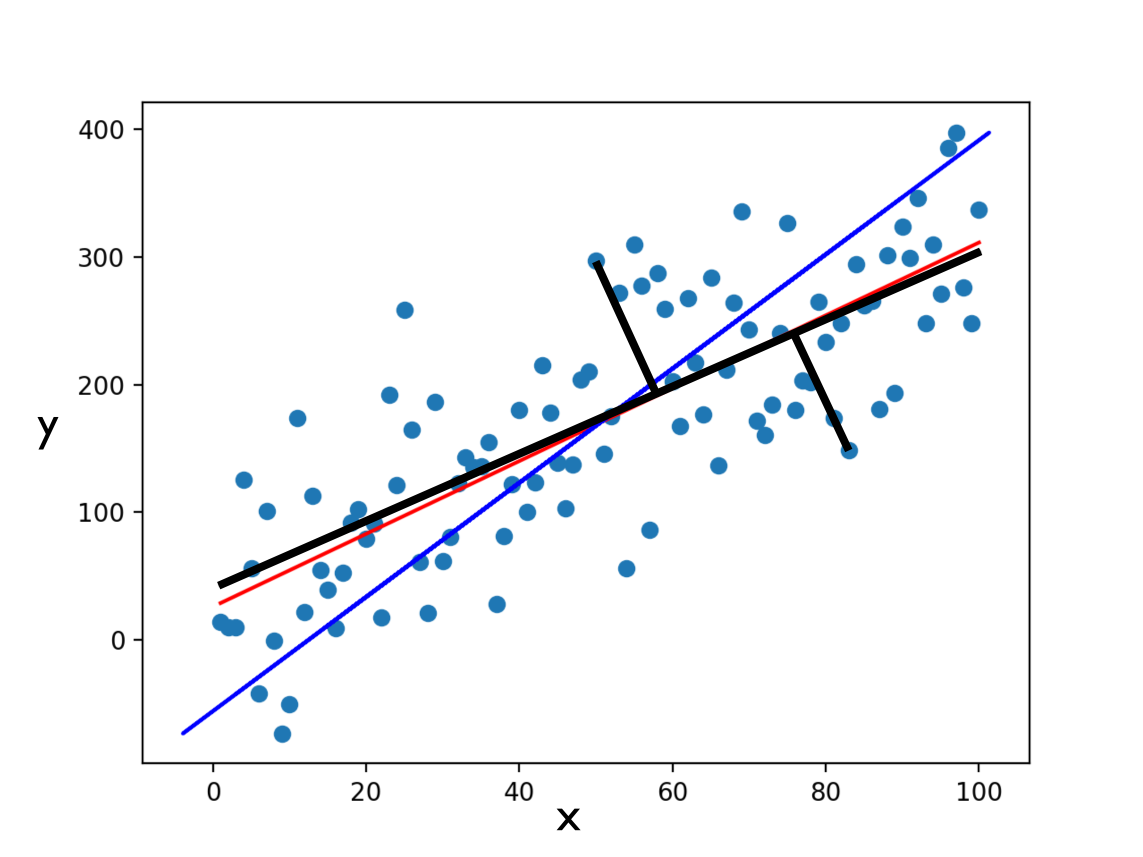
$$y = c \cdot x + m$$

$$x = d \cdot y + n$$

PCA regression:

- Scale input variables
- Estimate Covariance
- SVD on cov
- Choose largest eigenvalue
- Predict y

PCA VS OLS:
2 – PCA(2)



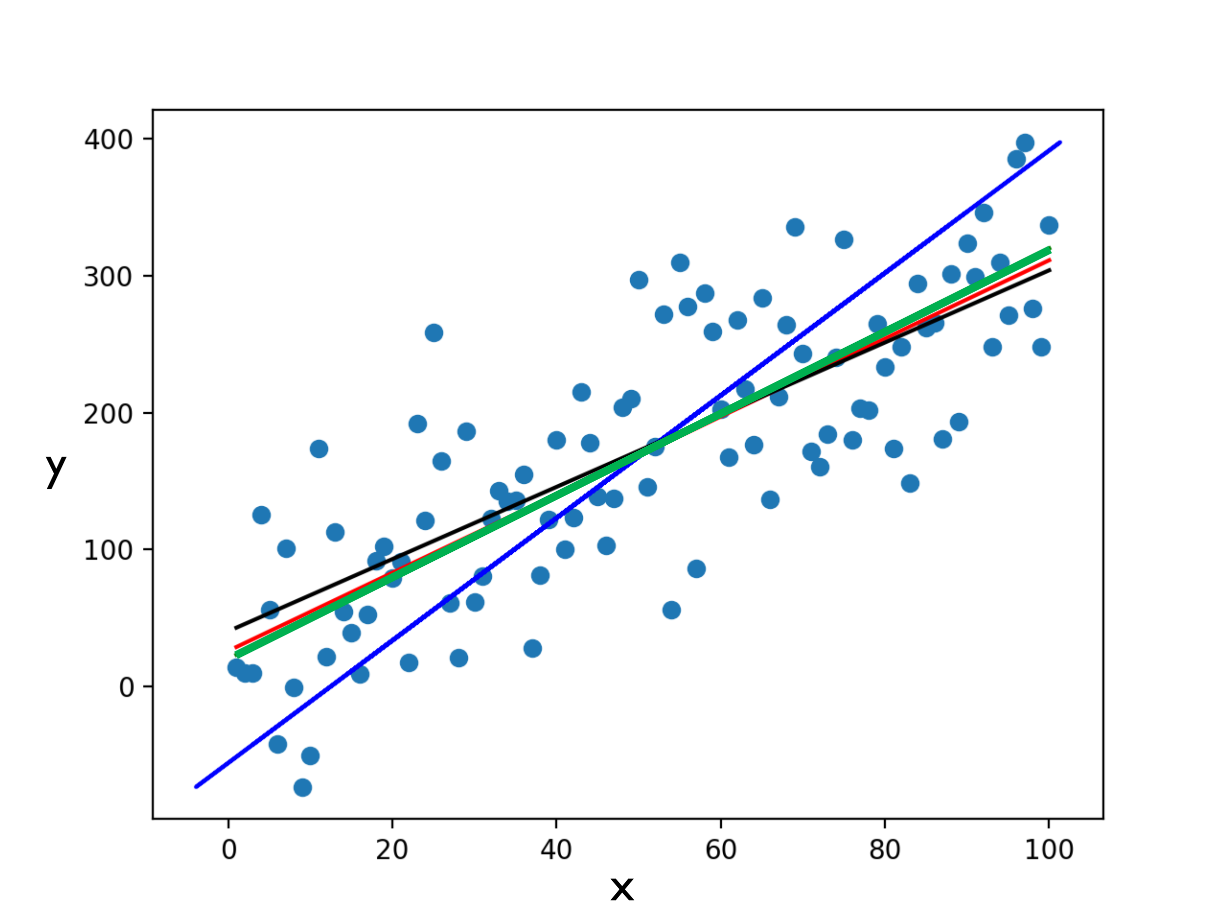
$$y = c \cdot x + m$$

$$x = d \cdot y + n$$

PCA regression:

- Scale input variables
- Estimate Covariance
- SVD on cov
- Choose largest eigenvalue
- Predict y

PCA VS OLS



TRUE CURVE

$$y = c \cdot x + m$$

$$x = d \cdot y + n$$

PCA regression:

- Scale input variables
- Estimate Covariance
- SVD on cov
- Choose largest eigenvalue
- Predict y

4. UNCERTAINTY ESTIMATION

- Parametric methods:
  - Huber, 1967: assymptotically normal, only valid for large number of regressors
  - Chave and Lezaeta, 2002: non-central Chi-square (2 dof) for rho, short-tailed Gaussian for phases
- Non-parametric methods:
  - Bootstrap
  - Jackknife

JACKKNIFE

- Let $\{x_i\}$ be an independent sample of size N
- Let ξ be a statistical parameter using the estimator Z:

$$\xi = Z(x_1, ..., x_N)$$

- Create N groups of N-1 samples, always leaving one of the samples from $\{x_i\}$ out:

$$\xi_{\setminus i} = Z(x_1, ..., x_{i-1}, x_{i+1}, ..., x_N)$$

- For example, estimate variance:

$$\sigma_J^2 = \frac{N-1}{N} \sum_{i=1}^{N} (\xi_{\setminus i} - \xi)^2$$

BOOTSTRAP

- Let $\{x_i\}$ be an independent sample of size N
- Let ξ be a statistical parameter using the estimator Z:

$$\xi = Z(x_1, ..., x_N)$$

- Create K groups of M samples, chosen randomly with replacement from $\{x_i\}$:

$$\xi_i = Z(M\,elements) \quad i = 1, ..., K$$

- For example, estimate variance:

$$\sigma_B^2 = \frac{K-1}{K} \sum_{i=1}^{K} (\xi_i - \xi)^2$$

NON-PARAMETRIC APPLIED TO MT

JACKKNIFE: BIRRP, Chave and Thomson, 2003, 2004:

$$\sigma_J^2 = \frac{N}{N-P}\sum_{i=1}^{N}(1-h_{ii})^2(z-z_{\setminus i})\cdot(z-z_{\setminus i})^T$$

BOOTSTRAP: EMT, Neukirch and Garcia, 2014:

$$\sigma_B^2 = \frac{K-1}{K}\sum_{i=1}^{K}(z-z_i)\cdot(z-z_i)^T$$

FINAL NOTES

- LOOK at the data
- There are plenty of freely available ROBUST processing codes
- Lots of developments in TS analysis and regression

THANK YOU!