Skoltech
Skolkovo Institute of Science and Technology

“Experimental Data Processing”

Topic 4
"Process reconstruction free from any constraints and assumptions”

Tatiana Podladchikova
Term 1B, October 2017
t.podladchikova@skoltech.ru

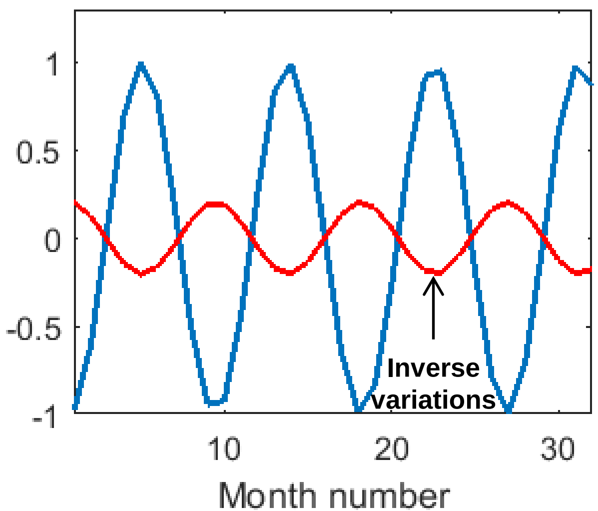
Analysis of running mean errors

Running mean may significantly distort the dynamics of the process

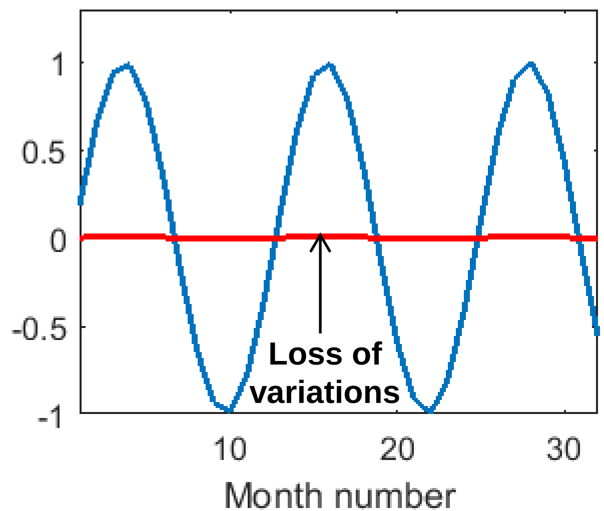
Measurements

Smoothed, window size =13

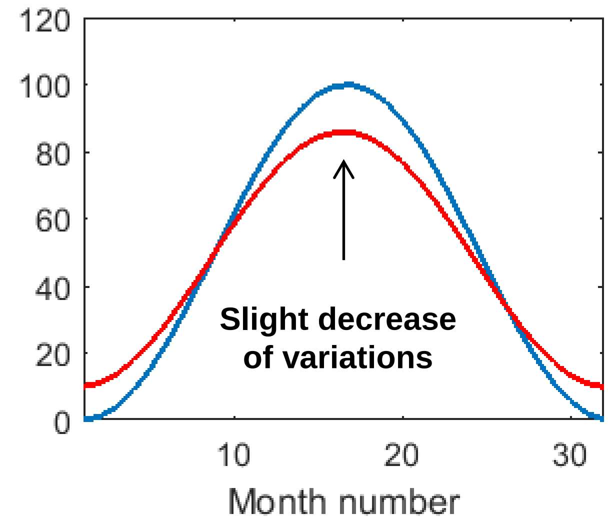
9-months variation



13-months variation



32-months variation

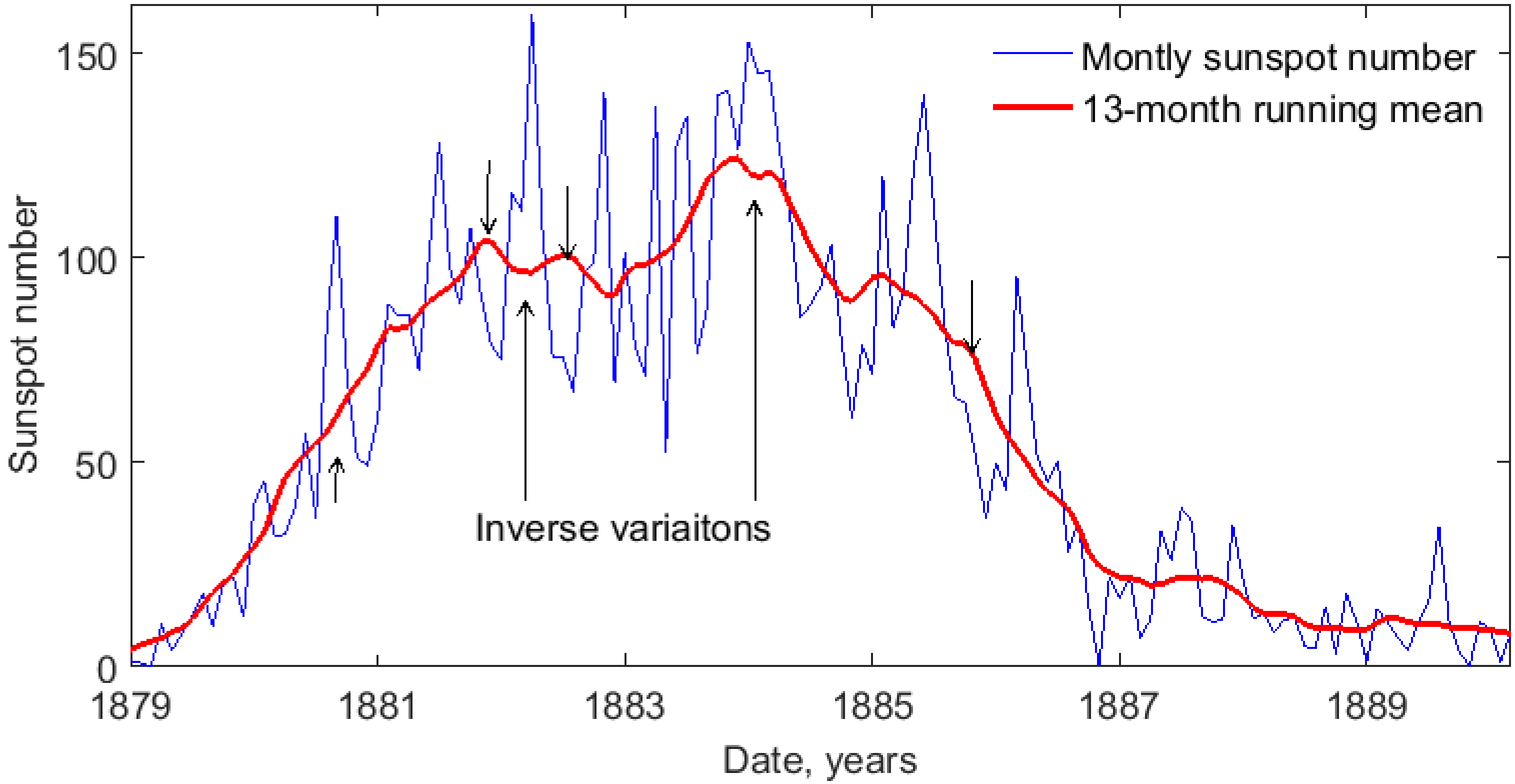


Inverse variations with periods from 6 to 12 months. Convex curve is replaced by concave curve and vice versa

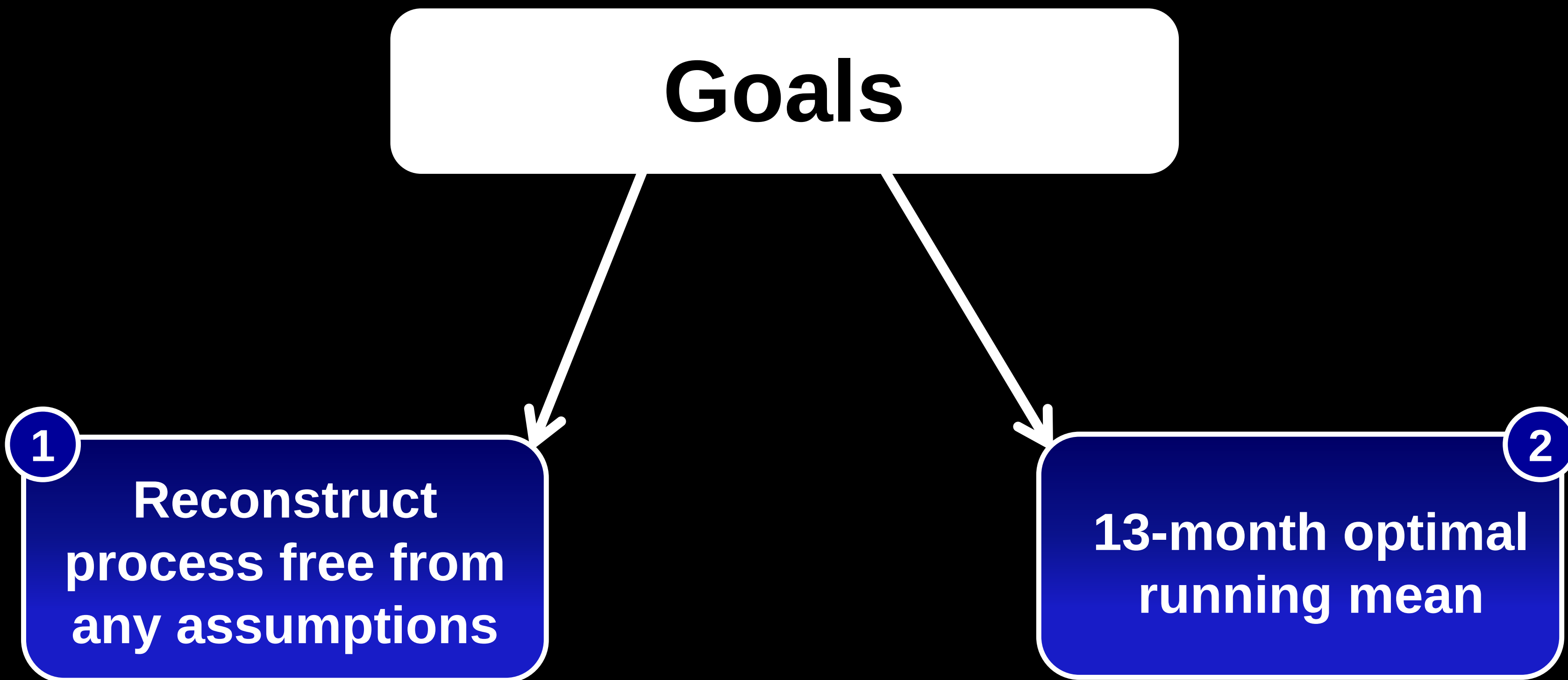
Total loss of 6- and 12-month variations decreasing them to zero

Period greater than running window size (13 months). The process in general is not distorted

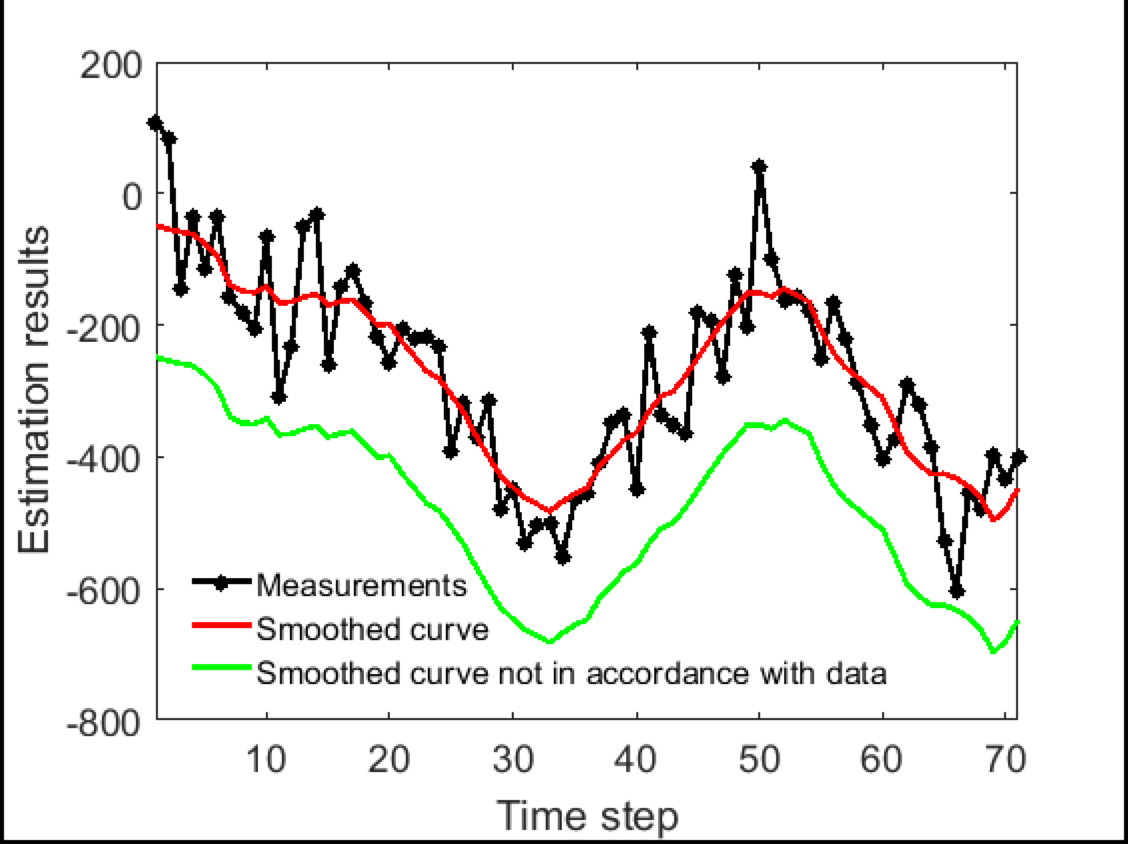
Distortion of physics in sunspot cycle 12



Performed analysis allows us to anticipate the errors of smoothing and getting false conclusions


Goals
1
Reconstruct process free from any assumptions
2
13-month optimal running mean

How reconstruct a process free from any assumptions?



Approximating curve $\widehat{X}_i$ has to be close to measurements z_i

1 **Deviation indicator**

$$I_d = \sum_{i=1}^{N} \left(z_i - \widehat{X}_i\right)^2$$

z_i - measurements
$\widehat{X}_i$ - estimation

Goal 1

How reconstruct a process free from any assumptions?

① Deviation indicator

$$I_d = \sum_{i=1}^{N} \left(z_i - \hat{X}_i\right)^2$$

z_i - measurements
$\hat{X}_i$ - estimation

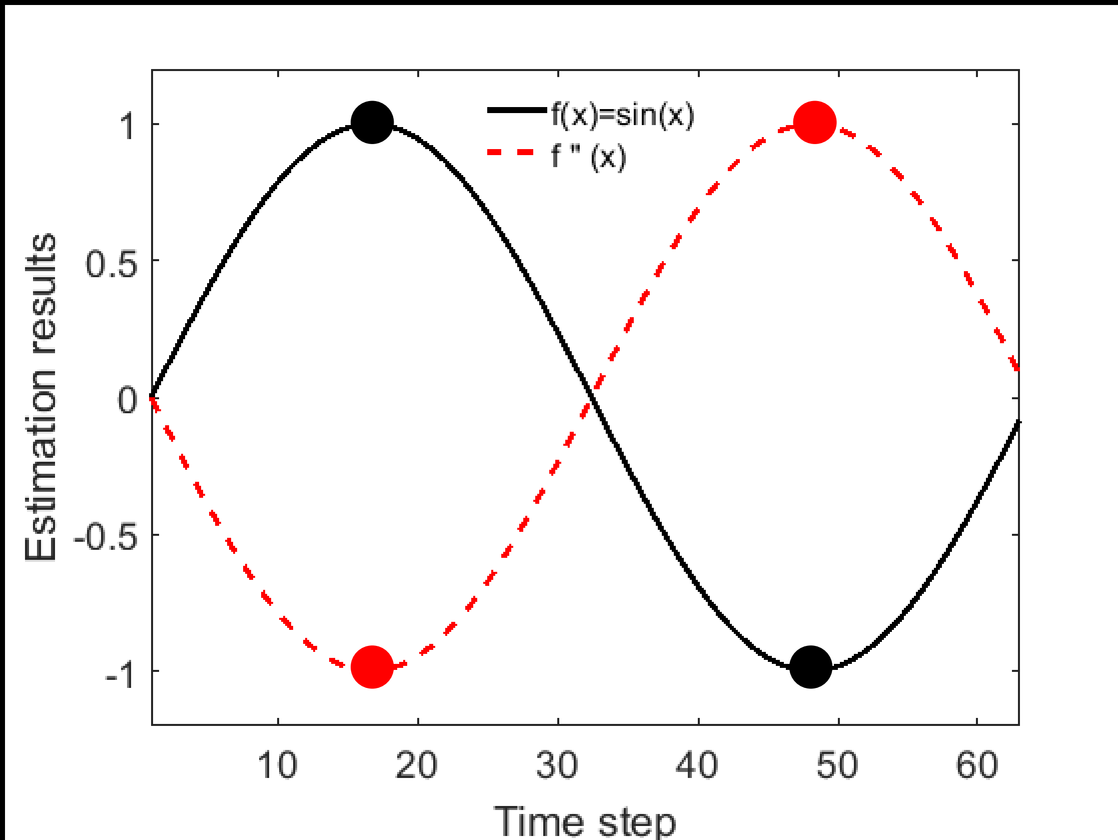
$I_d = 0$

No filtration of measurement noise
$\hat{X}_i = z_i$

**Not enough to use only deviation indicator.
Additional criterion is needed**

Goal 1

How reconstruct a process free from any assumptions?



Absolute value of second derivative is maximal at points of the greatest "variability" of approximating curve

Maximal rate of change of the process

② **Variability indicator**

$$I_v = \sum_{i=1}^{N-2} \left(\widehat{X}_{j+2} - 2\widehat{X}_{j+1} + \widehat{X}_j\right)^2$$

$\widehat{X}_i$ - estimation

Goal 1

Optimality criteria to find best approximation

z_i - measurements, $\widehat{X}_i$ - estimation

① Deviation indicator → $I_d = \sum_{i=1}^{N} (z_i - \widehat{X}_i)^2$ → $I_d \to min$

② Variability indicator → $I_v = \sum_{i=1}^{N-2} (\widehat{X}_{j+2} - 2\widehat{X}_{j+1} + \widehat{X}_j)^2$ → $I_v \to min$

Optimality criterion combining I_d and I_v?

Goal 1

Optimality criteria to find best approximation

z_i - measurements, $\widehat{X}_i$ - estimation

① Deviation indicator → $I_d = \sum_{i=1}^{N} \left(z_i - \widehat{X}_i\right)^2$ → $I_d \to min$

② Variability indicator → $I_v = \sum_{i=1}^{N-2} \left(\widehat{X}_{j+2} - 2\widehat{X}_{j+1} + \widehat{X}_j\right)^2$ → $I_v \to min$

Optimality criterion → $J = \beta I_d + I_v$ → $J \to min$

β → Smoothing coefficient → Greater β – Estimation $\widehat{X}_i$ closer to measurements z_i

Goal 1

Optimality criteria to find best approximation

$$J = \beta \sum_{i=1}^{N} \left(z_i - \hat{X}_i\right)^2 + \sum_{i=1}^{N-2} \left(\hat{X}_{j+2} - 2\hat{X}_{j+1} + \hat{X}_j\right)^2$$

Deviation indicator I_d

Variability indicator I_v

$J \to min$

Fidelity to measurements z_i

Smoothness of $\hat{X}_i$

Balance

G. Bohlmann, 1899
E.T. Whittaker, 1923

Goal 1

Minimization of functional *J* to find best estimation

1 Let's minimize functional *J*

$$J = \beta I_d + I_v$$

I_d → Deviation indicator

I_v → Variability indicator

β → Smoothing constant

Goal 1

Minimization of functional J to find best estimation

1 Let's minimize functional J

$$J = \beta I_d + I_v$$

I_d → Deviation indicator

I_v → Variability indicator

β → Smoothing constant

2 First derivative in respect to $\hat{X}$

$$\frac{dJ}{d\hat{X}_i} = 0$$

$\hat{X}_i$ → Estimation

Goal 1

Minimization of functional J to find best estimation

1. Let's minimize functional J

$$J = \beta I_d + I_v$$

- I_d → Deviation indicator
- I_v → Variability indicator
- β → Smoothing constant

2. First derivative in respect to $\hat{X}$

$$\frac{dJ}{d\hat{X}_i} = 0$$

- $\hat{X}_i$ → Estimation

3. System of equations

$$A\hat{X} = \beta Z$$

- A → $n \times n$ - matrix
- n → Observation interval
- Z → Measurement vector $Z = |z_1, z_2, \dots, z_n|^T$

Goal 1

Minimization of functional J to find best estimation

System of equations → $A\widehat{X} = \beta Z$

n equations with n unknowns

$\widehat{X}_i$ → Estimation

n → Observation interval

Z → Measurement vector $Z = |z_1, z_2, \dots, z_n|^T$

β → Smoothing constant

$$A = \begin{vmatrix} 1+\beta & -2 & 1 & 0 & 0 & \dots & 0 \\ -2 & 5+\beta & -4 & 1 & 0 & \dots & 0 \\ 1 & -4 & 6+\beta & -4 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & -4 & 6+\beta & -4 & 1 \\ 0 & \dots & 0 & 1 & -4 & 5+\beta & -2 \\ 0 & \dots & 0 & 0 & 1 & -2 & 1+\beta \end{vmatrix}$$

A → $n \times n$ five diagonal banded matrix

Goal 1

Minimization of functional J to find best estimation

System of equations → $A\hat{X} = \beta Z$

n equations with n unknowns

$\hat{X}_i$ → Estimation

n → Observation interval

Z → Measurement vector $Z = |z_1, z_2, \dots, z_n|^T$

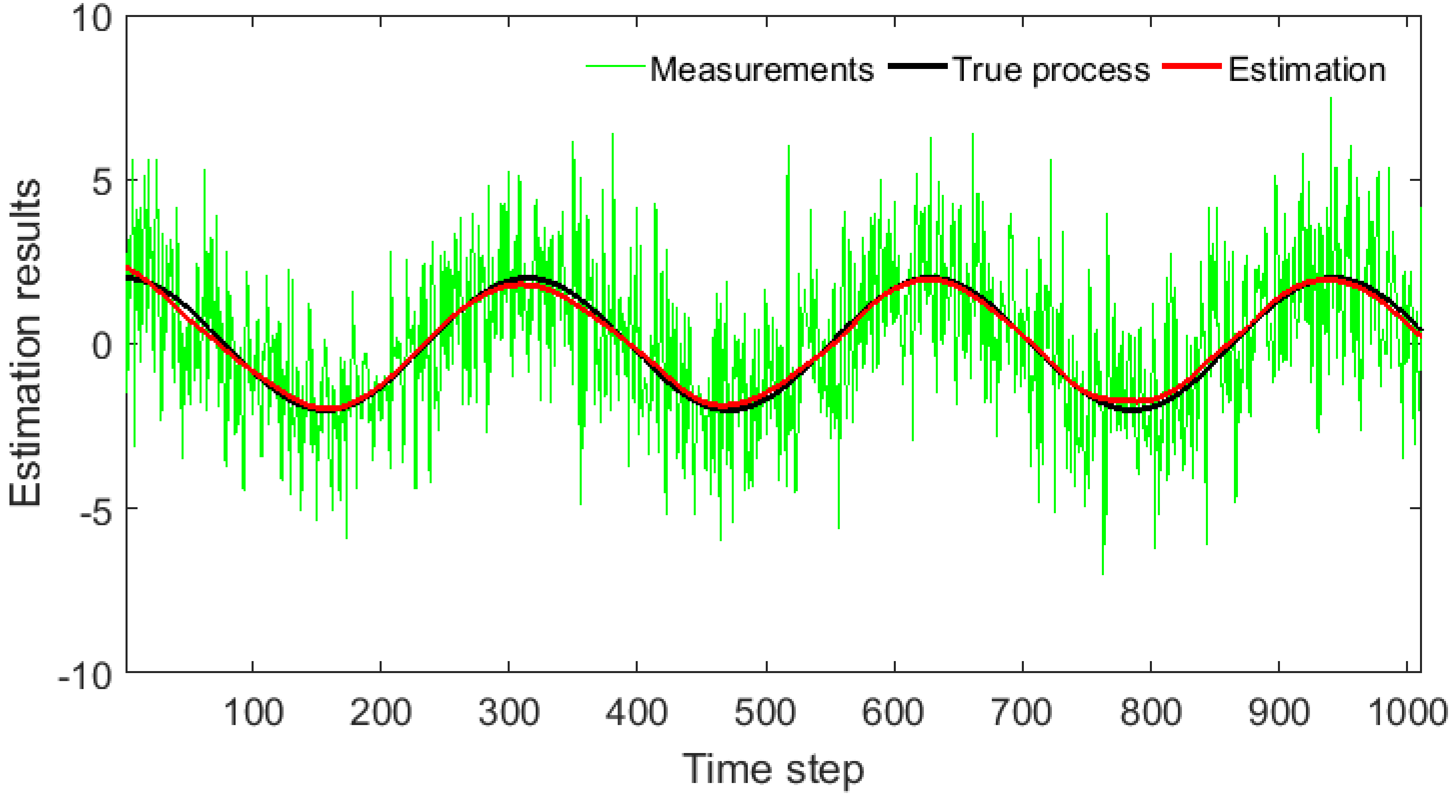
β → Smoothing constant

A → $n \times n$ five diagonal banded matrix

Solution of the system → $\hat{X} = \beta A^{-1} Z$ → Approximation $\hat{X}$ - weighted sum of measurements Z

Goal 1

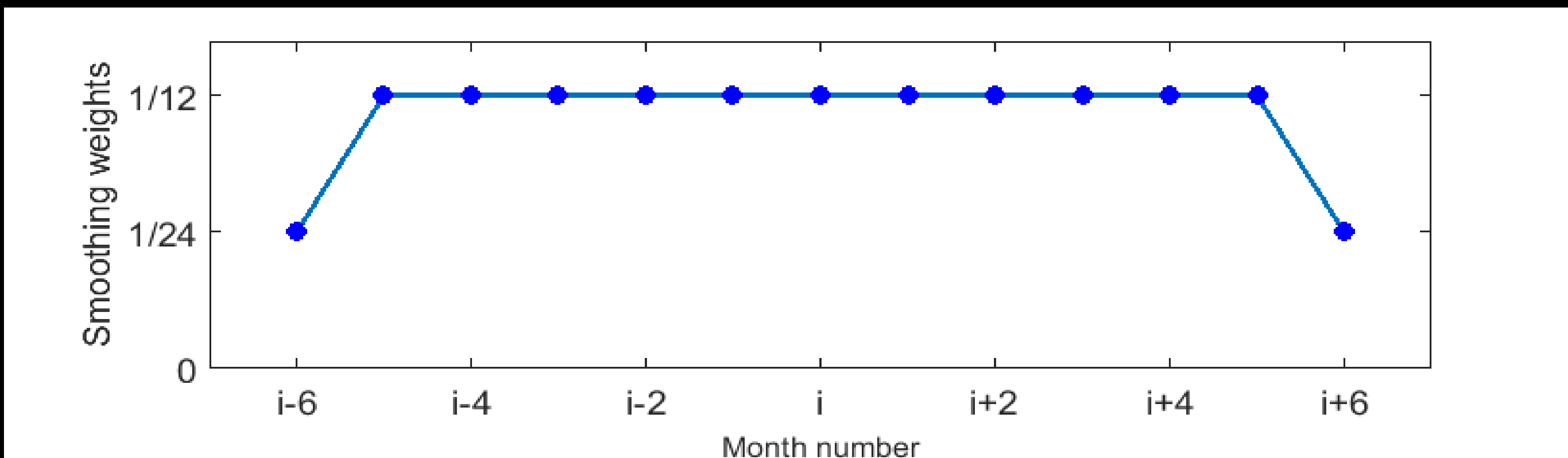
Reconstruction of a process free from any assumptions



13-month running mean

13-month sequent measurements Z

z_{i-6} z_{i-5} z_{i-4} z_{i-3} z_{i-2} z_{i-1} z_{i-6} z_{i+1} z_{i+2} z_{i+3} z_{i+4} z_{i+5} z_{i+6}



13-month running mean $\widehat{X}$

$$\widehat{X}_i = \frac{1}{24} z_{i-6} + \frac{1}{12}(z_{i-5} + z_{i-4} + \cdots + z_{i-1} + z_i + z_{i+1} + \cdots + z_{i+5}) + \frac{1}{24} z_{i+6}$$

Goal 2

13-month optimal running mean

13-month sequent Measurements Z

$$Z = \left| z_{i-6}, z_{i-5}, \ldots z_i, \ldots, z_{i+5}, z_{i+6} \right|^T$$

Goal 2

13-month optimal running mean

13-month sequent Measurements Z

$$Z = |z_{i-6}, z_{i-5}, \dots z_i, \dots, z_{i+5}, z_{i+6}|^T$$

13-month optimal running mean $\widehat{X}$

$$\widehat{X} = |x_{i-6}, x_{i-5}, \dots x_i, \dots, x_{i+5}, x_{i+6}|^T$$

?

Goal 2

13-month optimal running mean

13-month sequent Measurements Z

$Z = |z_{i-6}, z_{i-5}, \dots z_i, \dots, z_{i+5}, z_{i+6}|^T$

13-month optimal running mean $\hat{X}$

$\hat{X} = |x_{i-6}, x_{i-5}, \dots x_i, \dots, x_{i+5}, x_{i+6}|^T$

?

Optimal approximation $\hat{X}$ - weighted sum of measurements Z

$\hat{X} = \beta A^{-1} Z$

Goal 2

13-month optimal running mean

Optimal approximation $\widehat{X}$ - weighted sum of measurements Z → $\widehat{X} = \beta A^{-1} Z$

$$
\begin{bmatrix}
1+\beta & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2 & 5+\beta & -4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -4 & 6+\beta & -4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -4 & 5+\beta & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1+\beta
\end{bmatrix}
$$

A → 13×13 five diagonal banded matrix

Goal 2

13-month optimal running mean

Optimal approximation $\widehat{X}$ - weighted sum of measurements Z → $\widehat{X} = \beta A^{-1} Z$

$$\widehat{X} = \begin{vmatrix} x_{i-6} \\ x_{i-5} \\ \dots \\ x_i \\ \dots \\ x_{i+5} \\ x_{i+6} \end{vmatrix} = \beta A^{-1} \begin{vmatrix} z_{i-6} \\ z_{i-5} \\ \dots \\ z_i \\ \dots \\ z_{i+5} \\ z_{i+6} \end{vmatrix}$$

A → 13×13 five diagonal banded matrix

c_{ij} → Elements of matrix A^{-1}

Goal 2

13-month optimal running mean

Optimal approximation $\widehat{X}$ - weighted sum of measurements Z → $\widehat{X} = \beta A^{-1} Z$

$$\widehat{X} = \begin{vmatrix} x_{i-6} \\ x_{i-5} \\ \dots \\ x_i \\ \dots \\ x_{i+5} \\ x_{i+6} \end{vmatrix} = \beta \begin{vmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,7} & \dots & c_{1,12} & c_{1,13} \\ c_{2,1} & c_{2,2} & \dots & c_{2,7} & \dots & c_{2,12} & c_{2,13} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{7,1} & c_{7,2} & \dots & c_{7,7} & \dots & c_{7,12} & c_{7,13} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{12,1} & c_{12,2} & \dots & c_{12,7} & \dots & c_{12,12} & c_{12,13} \\ c_{13,1} & c_{13,2} & \dots & c_{13,7} & \dots & c_{13,12} & c_{13,13} \end{vmatrix} \begin{vmatrix} z_{i-6} \\ z_{i-5} \\ \dots \\ z_i \\ \dots \\ z_{i+5} \\ z_{i+6} \end{vmatrix}$$

A → 13×13 five diagonal banded matrix

c_{ij} → Elements of matrix A^{-1}

Goal 2

13-month optimal running mean

Optimal approximation $\widehat{X}$ - weighted sum of measurements Z → $\widehat{X} = \beta A^{-1} Z$

$$\widehat{X} = \begin{vmatrix} x_{i-6} \\ x_{i-5} \\ \cdots \\ x_i \\ \cdots \\ x_{i+5} \\ x_{i+6} \end{vmatrix} = \beta \begin{vmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,7} & \cdots & c_{1,12} & c_{1,13} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,7} & \cdots & c_{2,12} & c_{2,13} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{7,1} & c_{7,2} & \cdots & c_{7,7} & \cdots & c_{7,12} & c_{7,13} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{12,1} & c_{12,2} & \cdots & c_{12,7} & \cdots & c_{12,12} & c_{12,13} \\ c_{13,1} & c_{13,2} & \cdots & c_{13,7} & \cdots & c_{13,12} & c_{13,13} \end{vmatrix} \begin{vmatrix} z_{i-6} \\ z_{i-5} \\ \cdots \\ z_i \\ \cdots \\ z_{i+5} \\ z_{i+6} \end{vmatrix}$$

A → 13×13 five diagonal banded matrix

c_{ij} → Elements of matrix A^{-1}

How to get estimation x_i at time i?

Goal 2

13-month optimal running mean

Optimal approximation $\hat{X}$ - weighted sum of measurements Z → $\hat{X} = \beta A^{-1} Z$

$$\hat{X} = \begin{vmatrix} x_{i-6} \\ x_{i-5} \\ \dots \\ x_i \\ \dots \\ x_{i+5} \\ x_{i+6} \end{vmatrix} = \beta \begin{vmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,7} & \dots & c_{1,12} & c_{1,13} \\ c_{2,1} & c_{2,2} & \dots & c_{2,7} & \dots & c_{2,12} & c_{2,13} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{7,1} & c_{7,2} & \dots & c_{7,7} & \dots & c_{7,12} & c_{7,13} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{12,1} & c_{12,2} & \dots & c_{12,7} & \dots & c_{12,12} & c_{12,13} \\ c_{13,1} & c_{13,2} & \dots & c_{13,7} & \dots & c_{13,12} & c_{13,13} \end{vmatrix} \begin{vmatrix} z_{i-6} \\ z_{i-5} \\ \dots \\ z_i \\ \dots \\ z_{i+5} \\ z_{i+6} \end{vmatrix}$$

A → 13×13 five diagonal banded matrix

c_{ij} → Elements of matrix A^{-1}

$x_i = \beta C(\beta) Z$

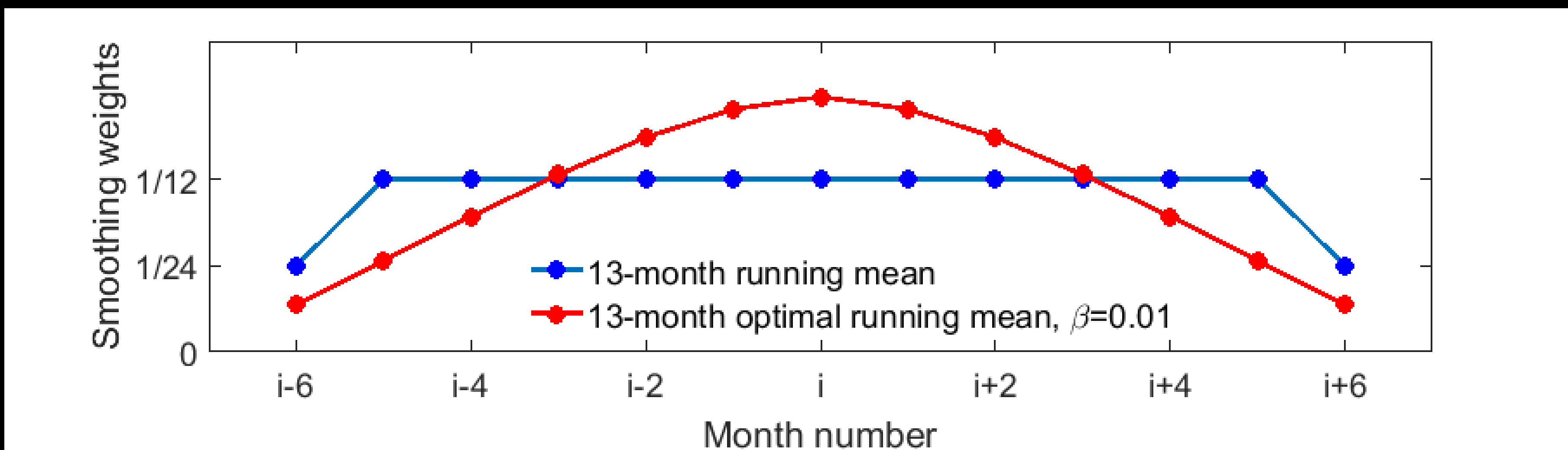
$C(\beta) = \left| c_{7,1}, c_{7,2}, \dots, c_{7,7}, \dots c_{7,12}, c_{7,13} \right|$

Goal 2

Smoothing weights in 13-month running mean and 13-month optimal running mean

13-month sequent measurements Z

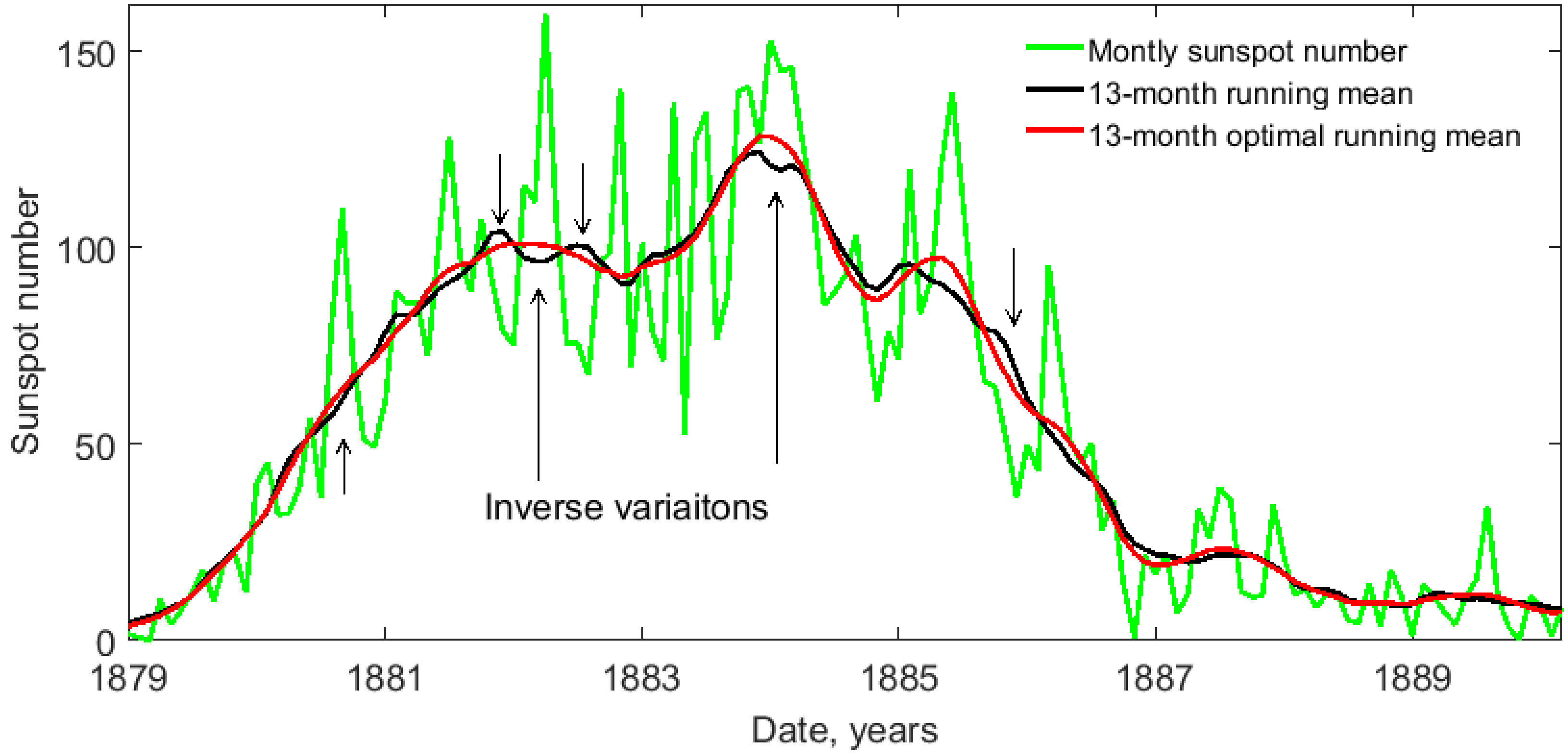
z_{i-6} z_{i-5} z_{i-4} z_{i-3} z_{i-2} z_{i-1} z_{i-6} z_{i+1} z_{i+2} z_{i+3} z_{i+4} z_{i+5} z_{i+6}



The optimal weights monotonously increase when measurements approach step i

Goal 2

Approximation of sunspot cycle 12



The optimal filter provides more adequate presentation of the sunspot cycle and doesn’t distort the short-term variations of sunspot numbers

Comparative analysis of optimal and non-optimal techniques

To analyze qualitatively advantages of optimal smoothing technique compared to the 13-month running mean

↓

Determine and compare deviation indicator I_d and variability indicator I_v