

“Experimental Data Processing”

Topic 2 "Quasi-optimal approximation under uncertainty”

Tatiana Podladchikova
Term 1, September-October 2020
t.podladchikova@skoltech.ru

The basis of statistical analysis

```
graph TD; A[The basis of statistical analysis] --> B[Least-square method and linear regression]; B --> C[The LSM method leads to divergence and loses its practical value when a model is inadequate or unknown]; B --> D[Linear regression doesn't provide reliable long-term forecasting];
```

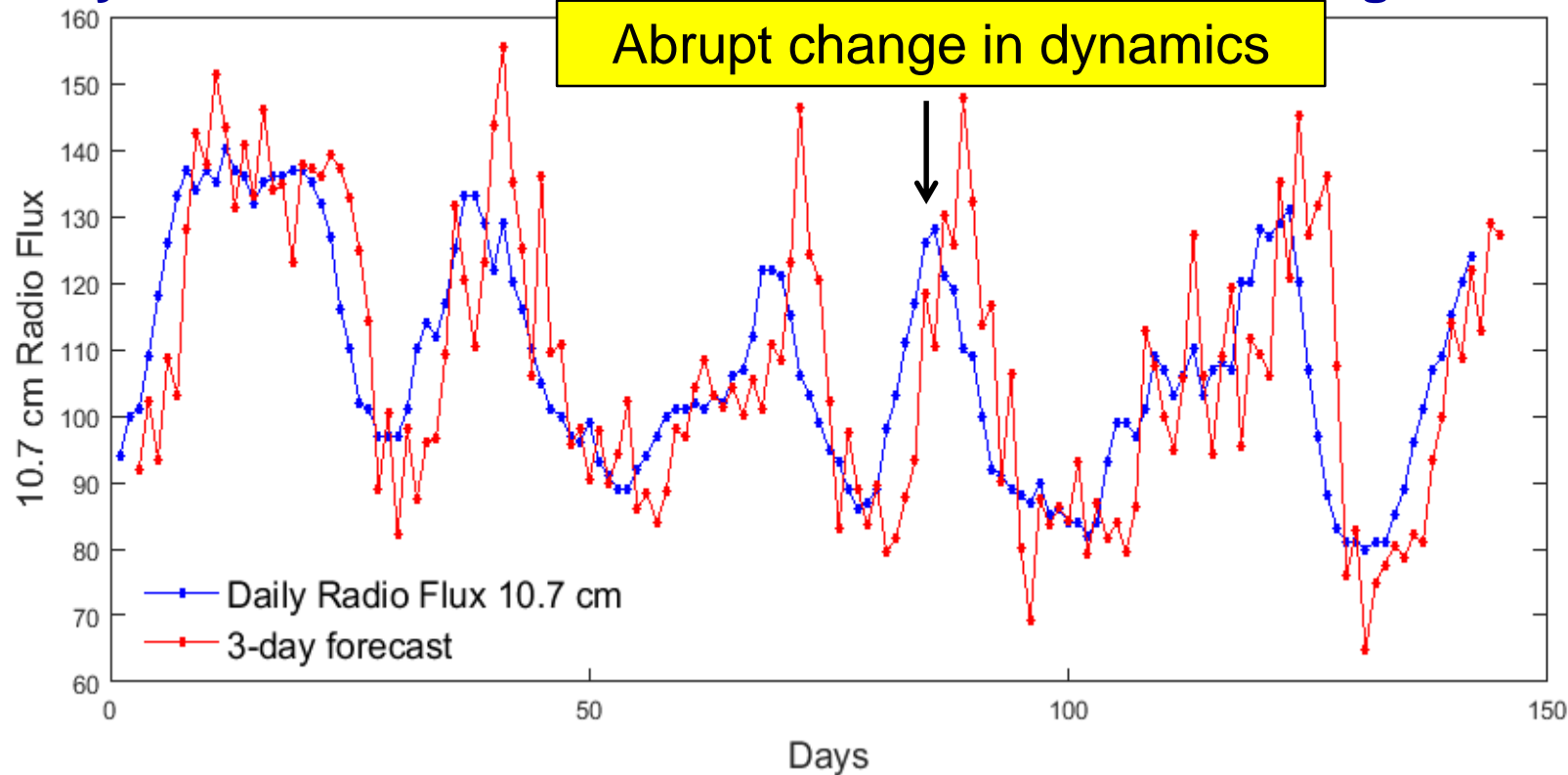
**Least-square method
and linear regression**

**The LSM method leads
to divergence and loses its
practical value when a model
is inadequate or unknown**

**Linear regression doesn't
provide reliable long-term
forecasting**

Linear regression doesn't provide long-term forecasting

3-day 10.7 cm radio flux forecast based on a linear regression



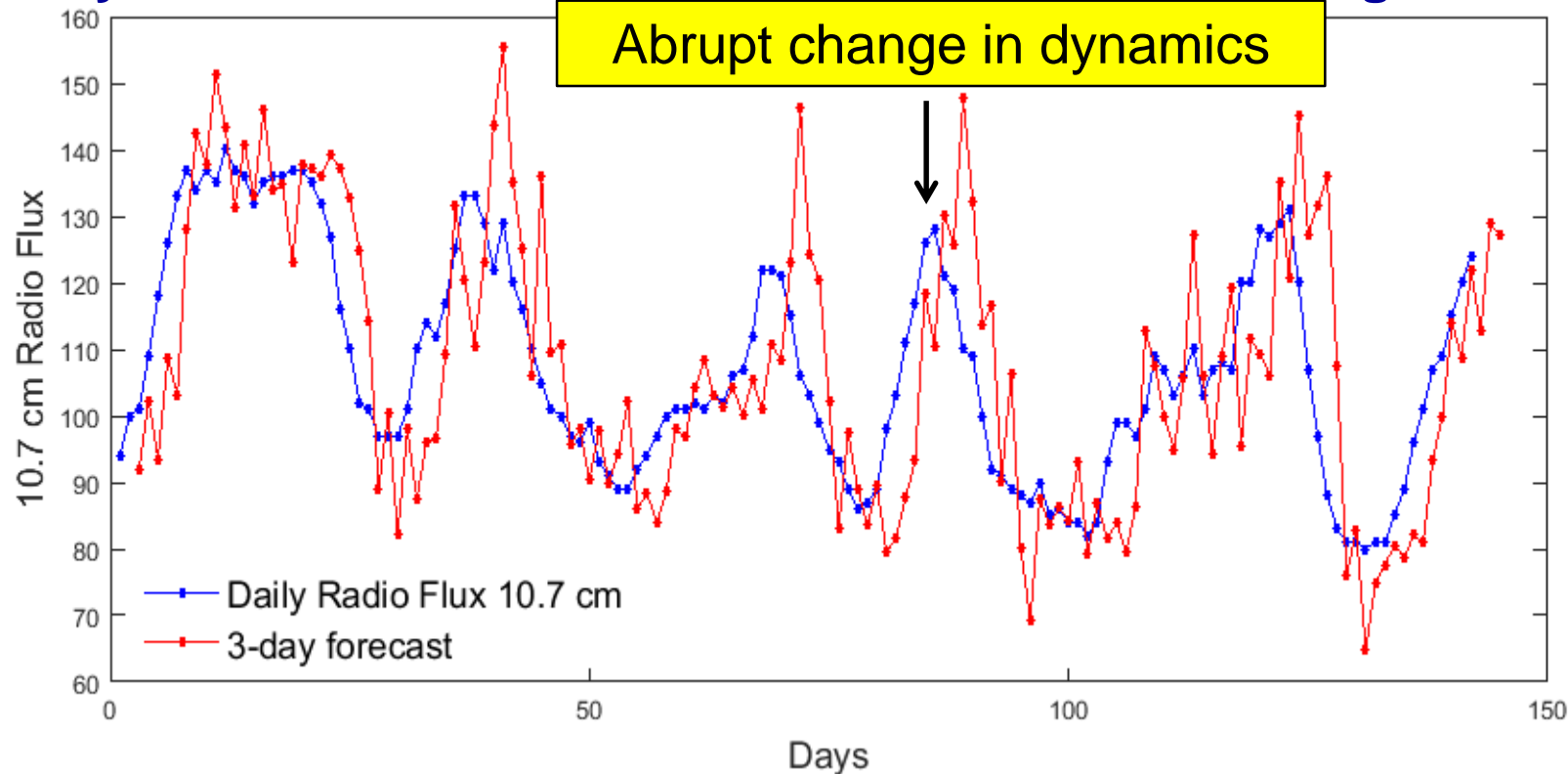
How would you characterize this forecast?

www.menti.com

65 94 13

Linear regression doesn't provide long-term forecasting

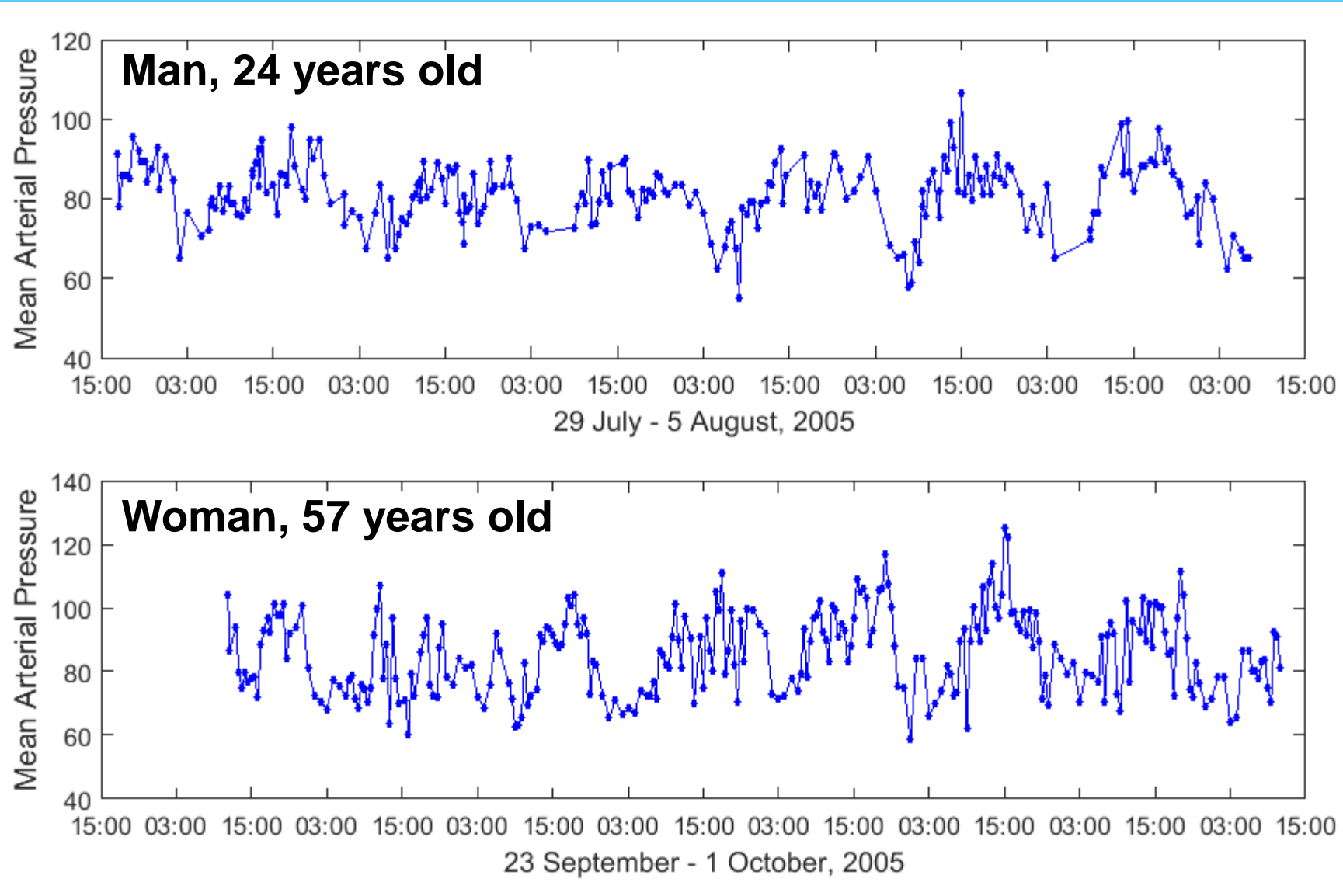
3-day 10.7 cm radio flux forecast based on a linear regression



**Changes in dynamics
of a process leads to great
increase of forecasting errors**

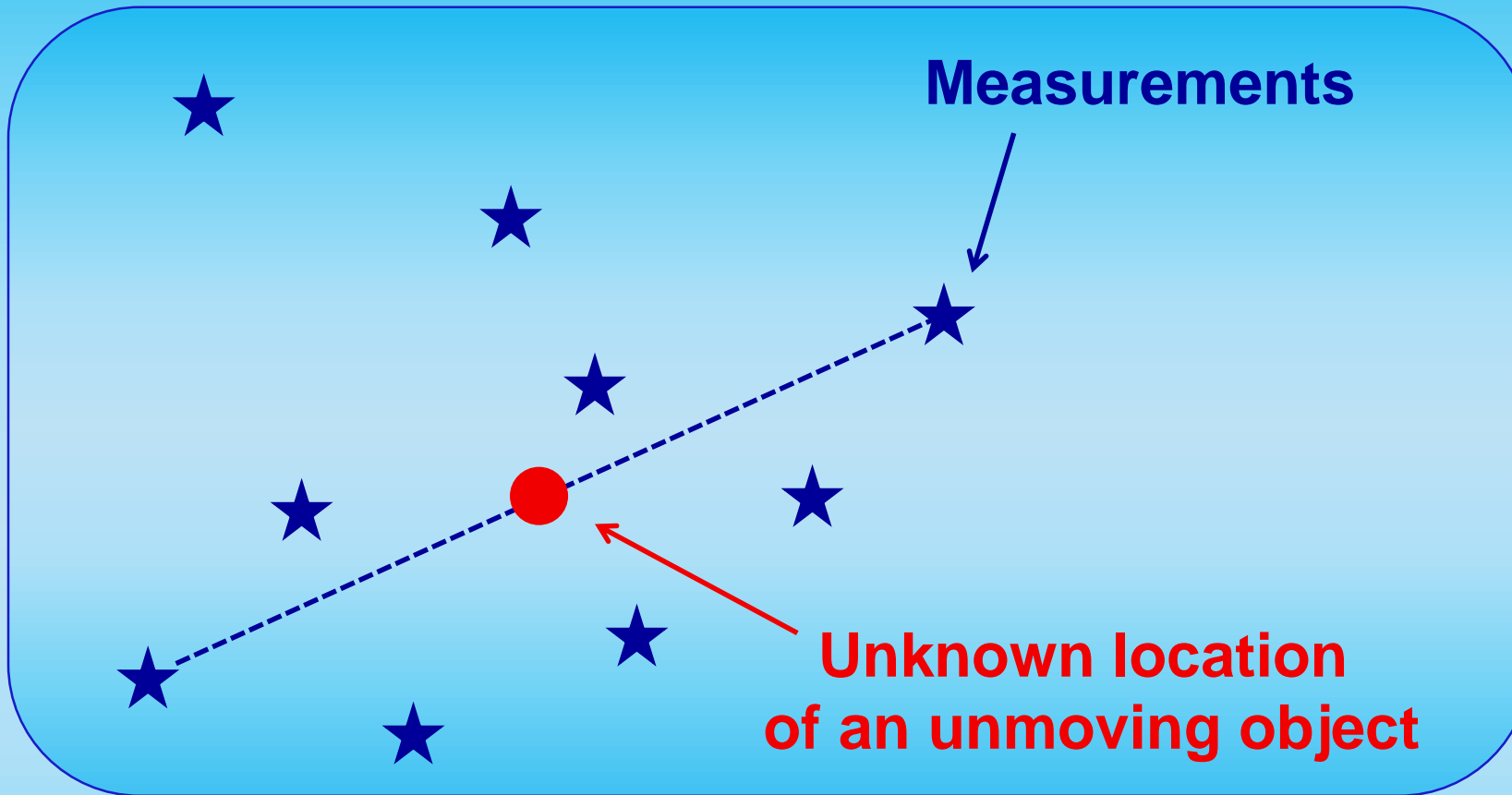
**To extract regularities that will
allow long-term forecasting
we need to smooth data**

Which regularities can you extract from measurements of mean arterial pressure?



Source of data: Medical Department, Russian University of Friendship of Nations

Estimate the location of an unmoving object



Smoothing is weighted averaging of noisy data.
Fluctuation components are self compensated.

The most popular methods
of quasi-optimal estimation

1

Running mean

2

Exponential smoothing



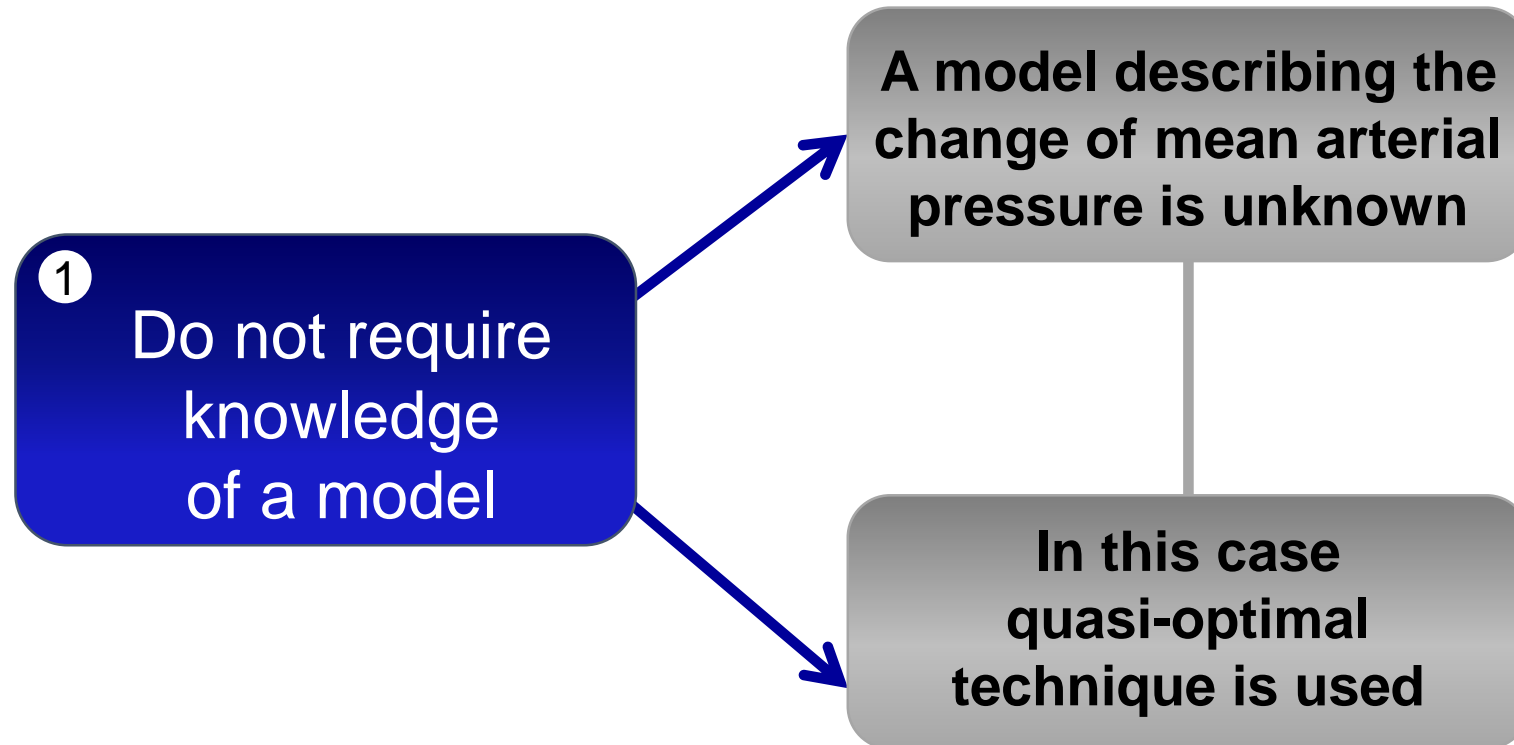
Advantages



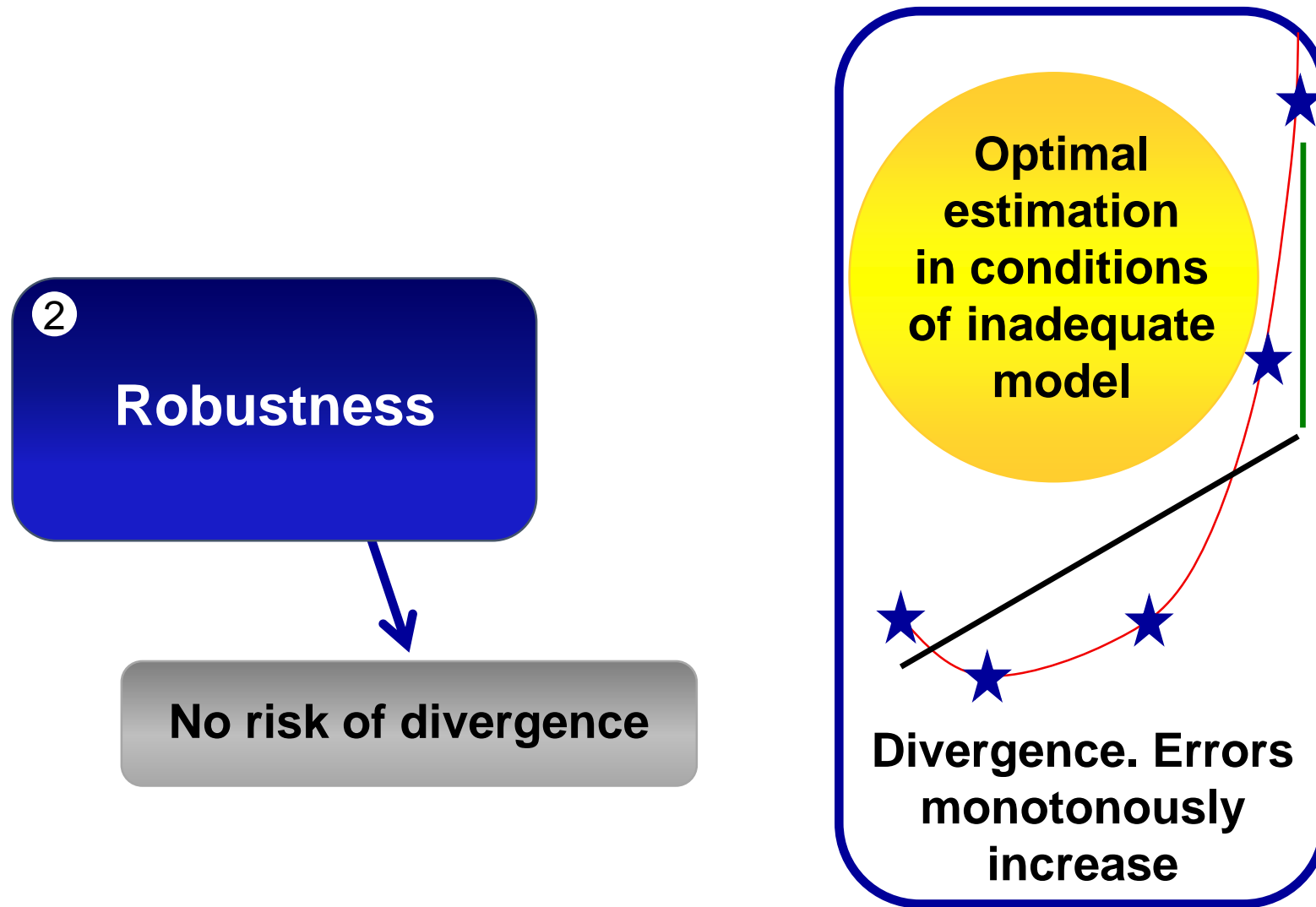
Limitations



Advantages of quasi-optimal estimation methods

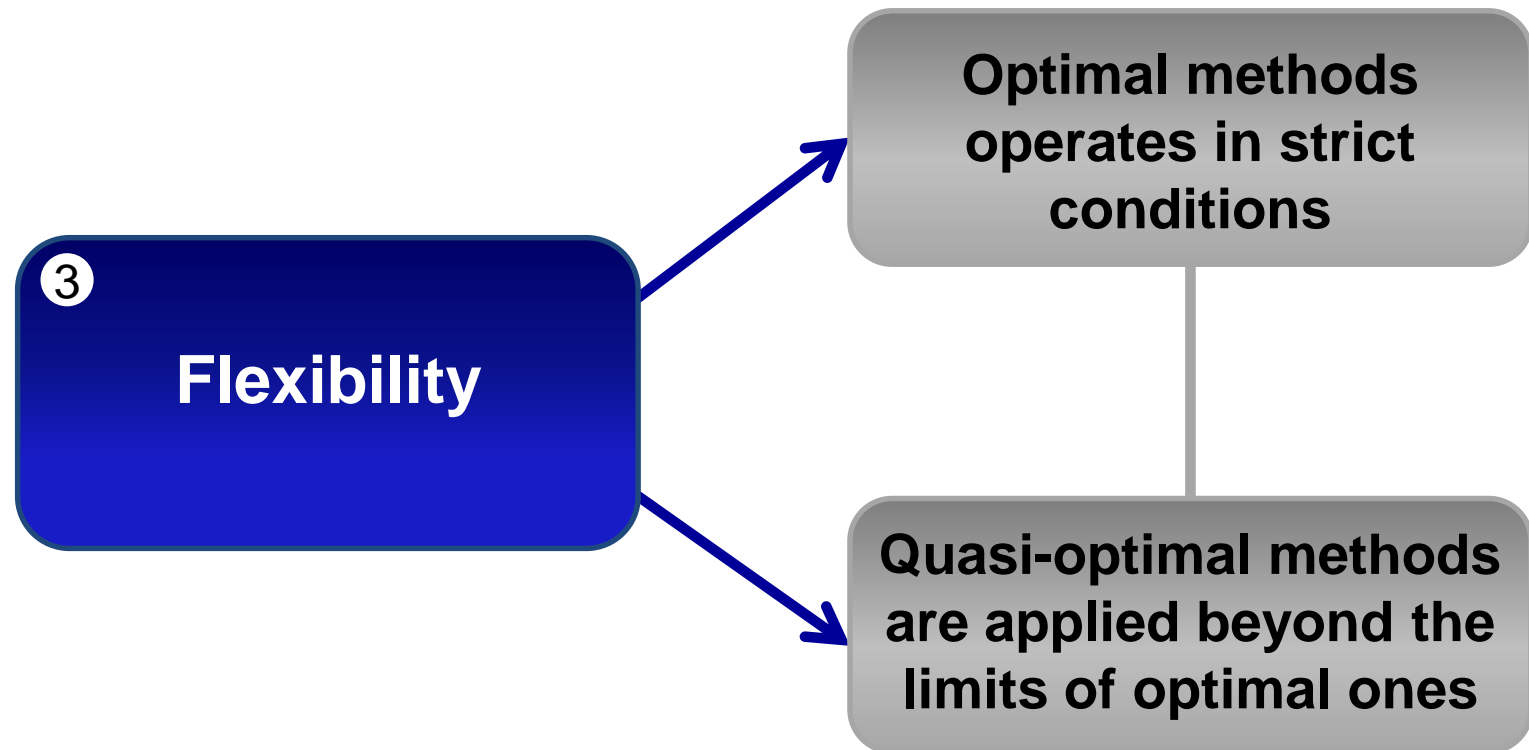


✓ Advantages of quasi-optimal estimation methods



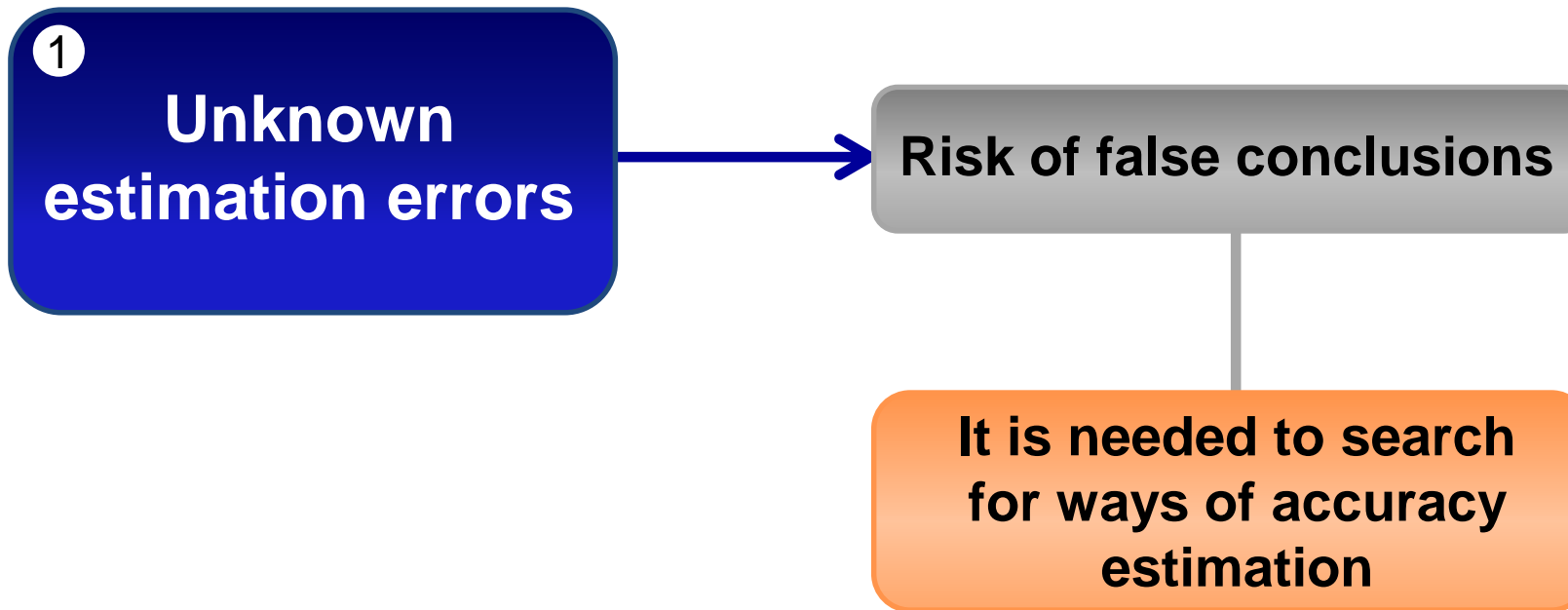


Advantages of quasi-optimal estimation methods





Disadvantages of quasi-optimal estimation methods



Quasi-optimal approximation
under uncertainty

```
graph TD; A[Quasi-optimal approximation under uncertainty] --> B[Learning goals]; B --> C[Analyze conditions for which methods provide effective solution and conditions under which they break down.]; B --> D[Chose the most effective method in conditions of uncertainty];
```

Learning goals

**Analyze conditions for which
methods provide effective
solution and conditions under
which they break down.**

**Chose the most effective
method in conditions
of uncertainty**

① Running mean

$$z_i = X_i + \eta_i$$

z_j
Measurements

X_i
True process
to be
estimated

η_i
Uncorrelated
unbiased noise
with variance
 σ_η^2

① Running mean

$$z_i = X_i + \eta_i$$

z_j
Measurements

X_i
True process
to be
estimated

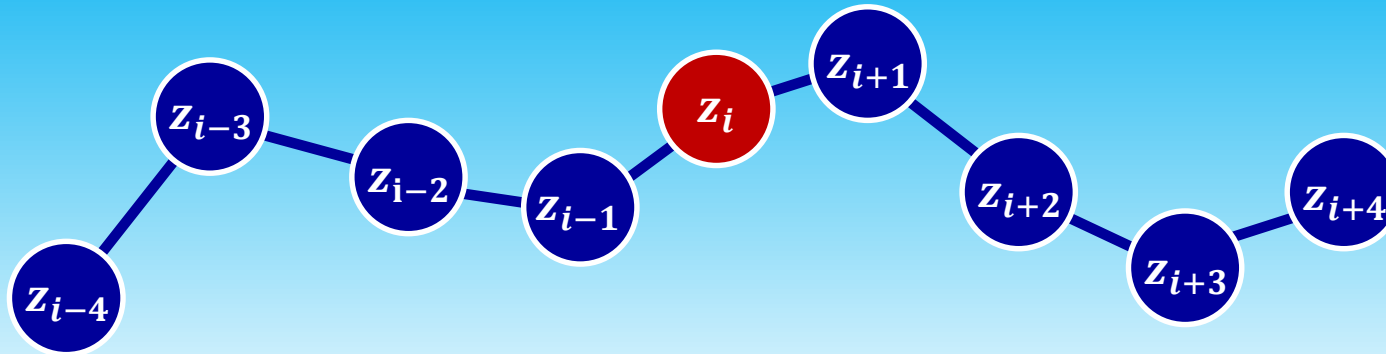
η_i
Uncorrelated
unbiased noise
with variance
 σ_η^2

Our goal

To reconstruct the dynamics of process X_i
using available measurements z_j when
the dynamical model is unknown

1 Running mean

Window size $M = 9$



Last 9 measurements z_i

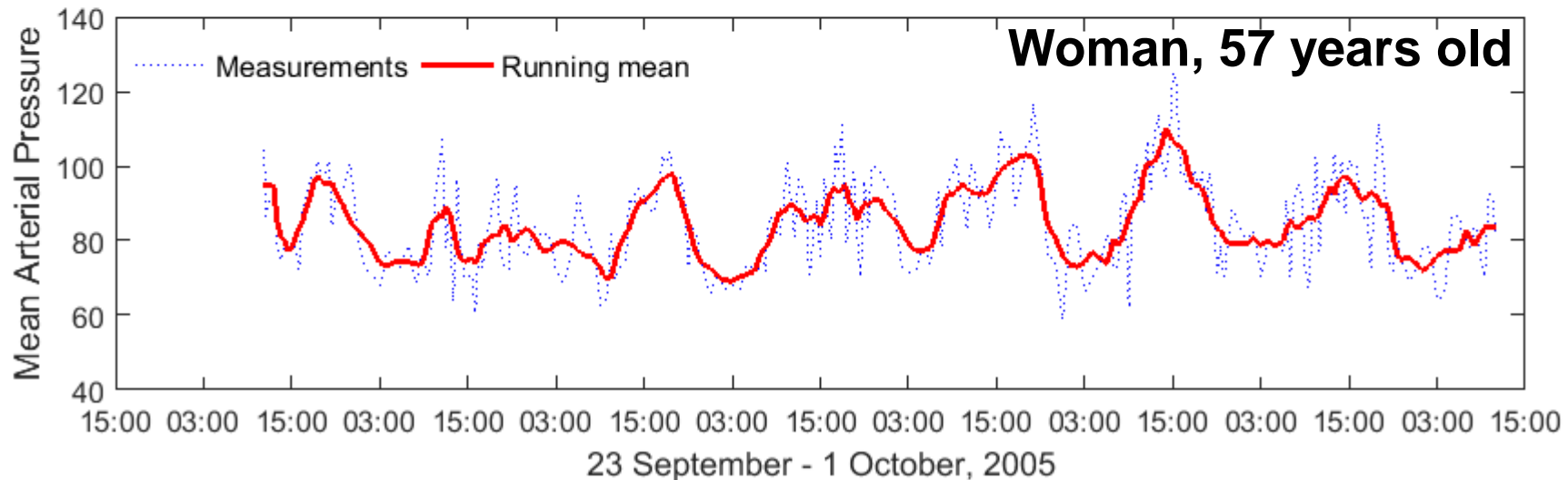
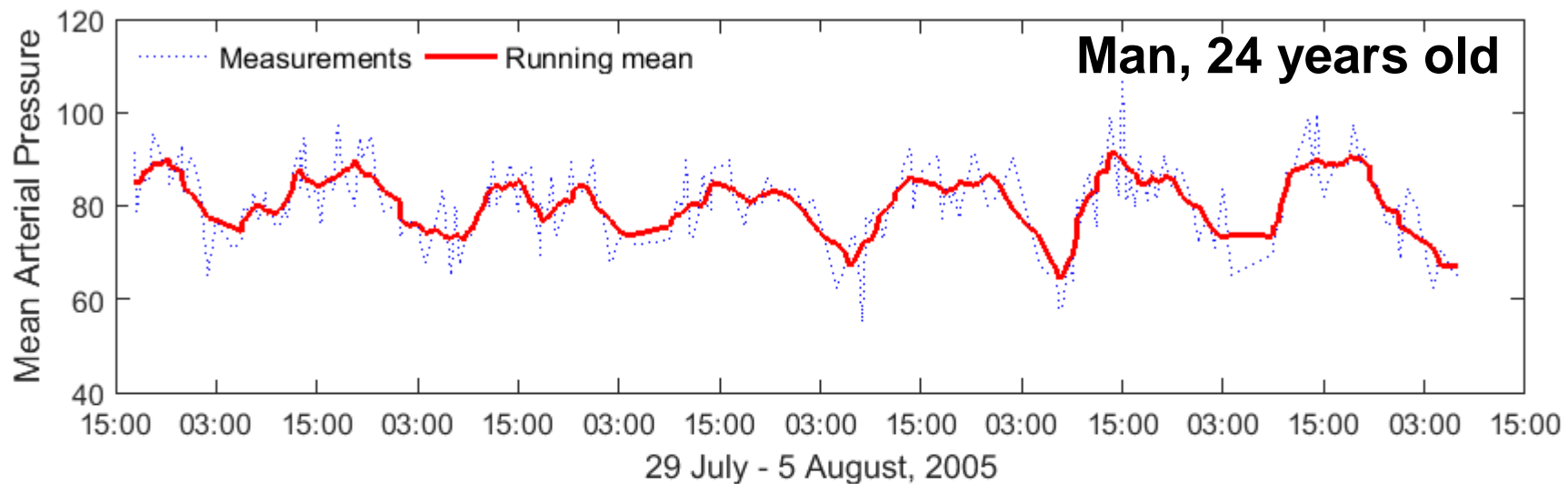
Estimation \hat{X}_i

$$\hat{X}_i = \frac{1}{9} \sum_{k=i-4}^{i+4} z_k$$

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

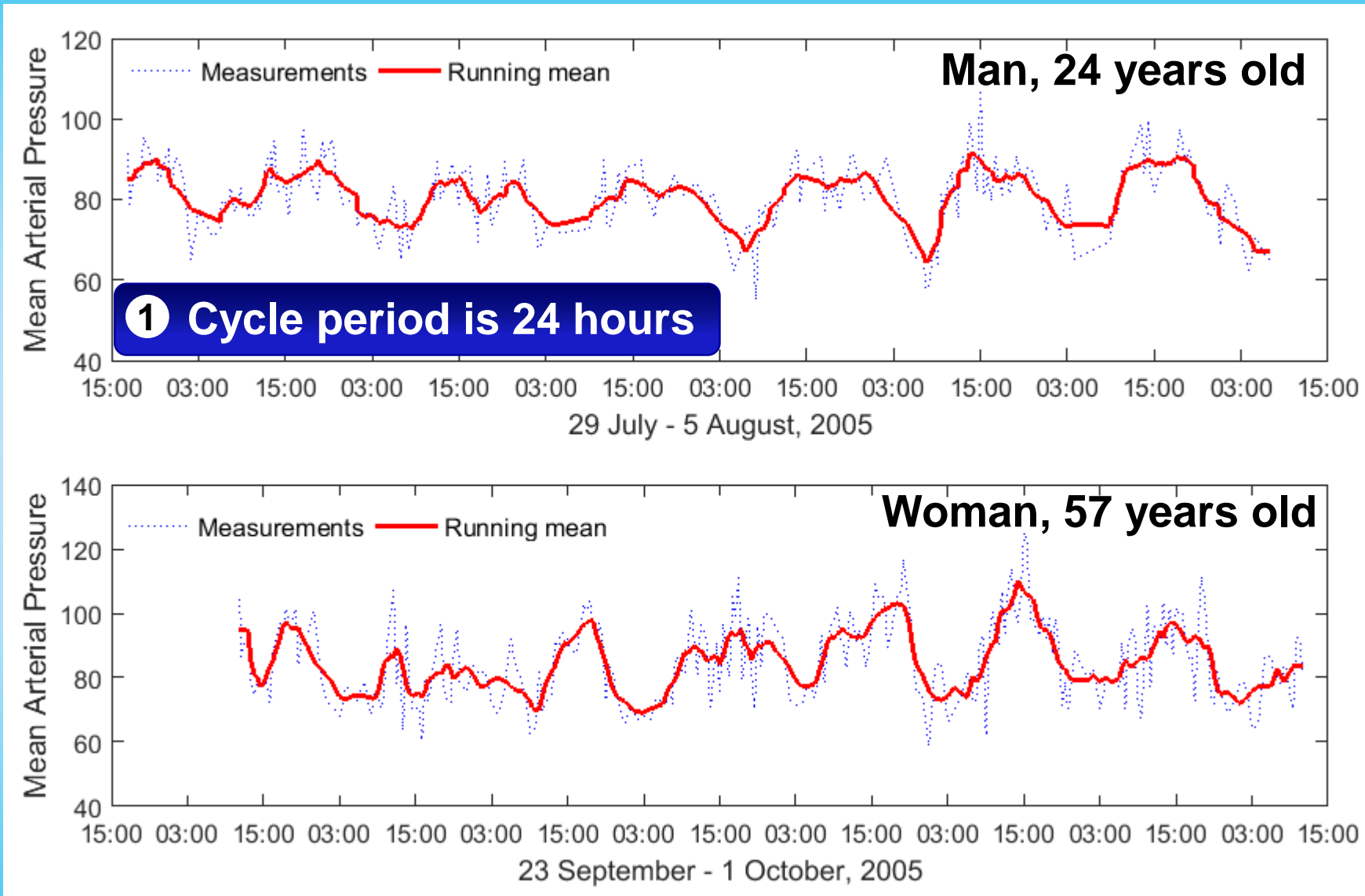
M - window size

Which regularities can you extract from smoothed measurements of mean arterial pressure?



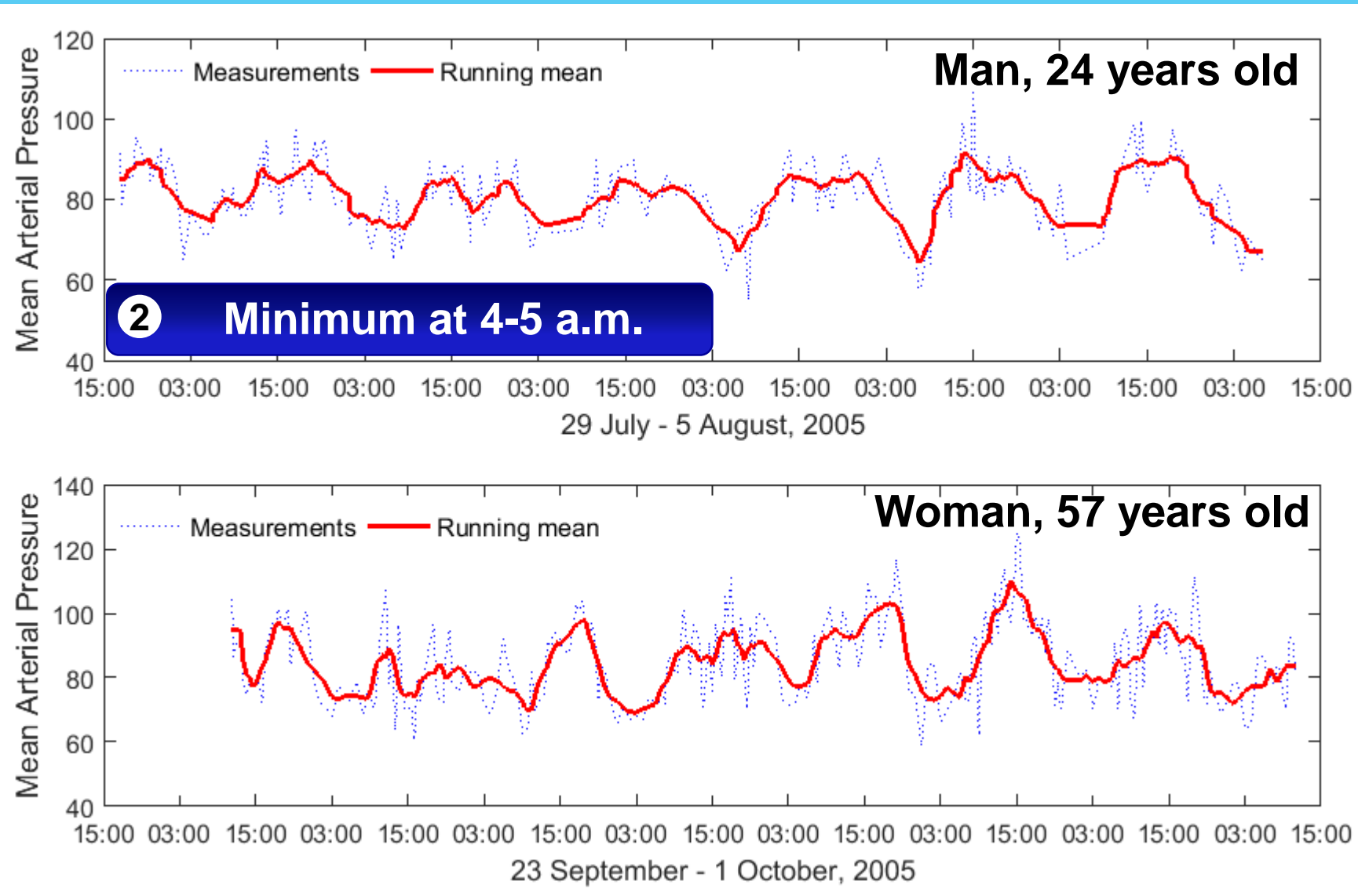
Running mean with window $M = 7$

Which regularities can you extract from smoothed measurements of mean arterial pressure?



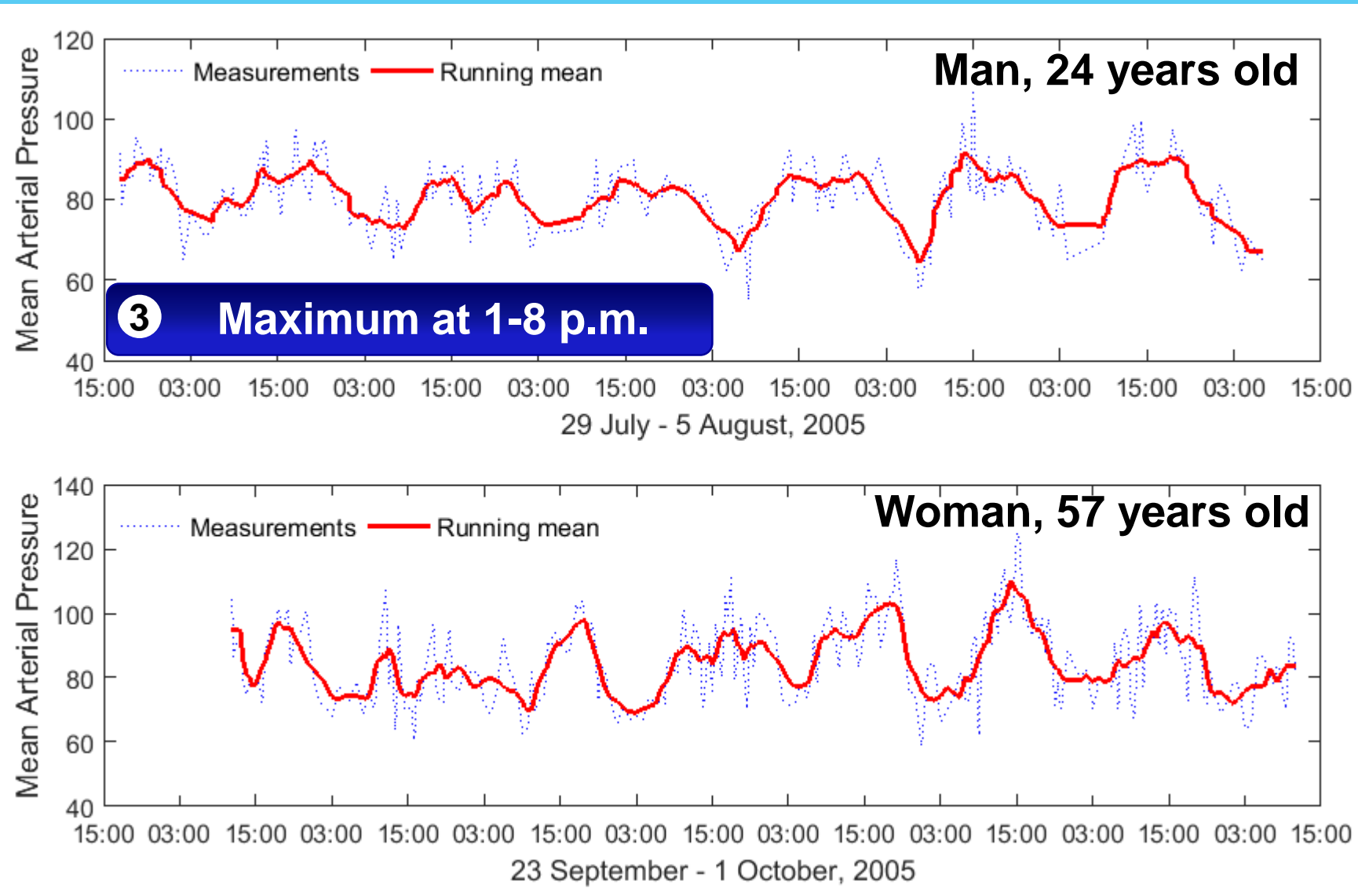
Running mean with window $M = 7$

Which regularities can you extract from smoothed measurements of mean arterial pressure?



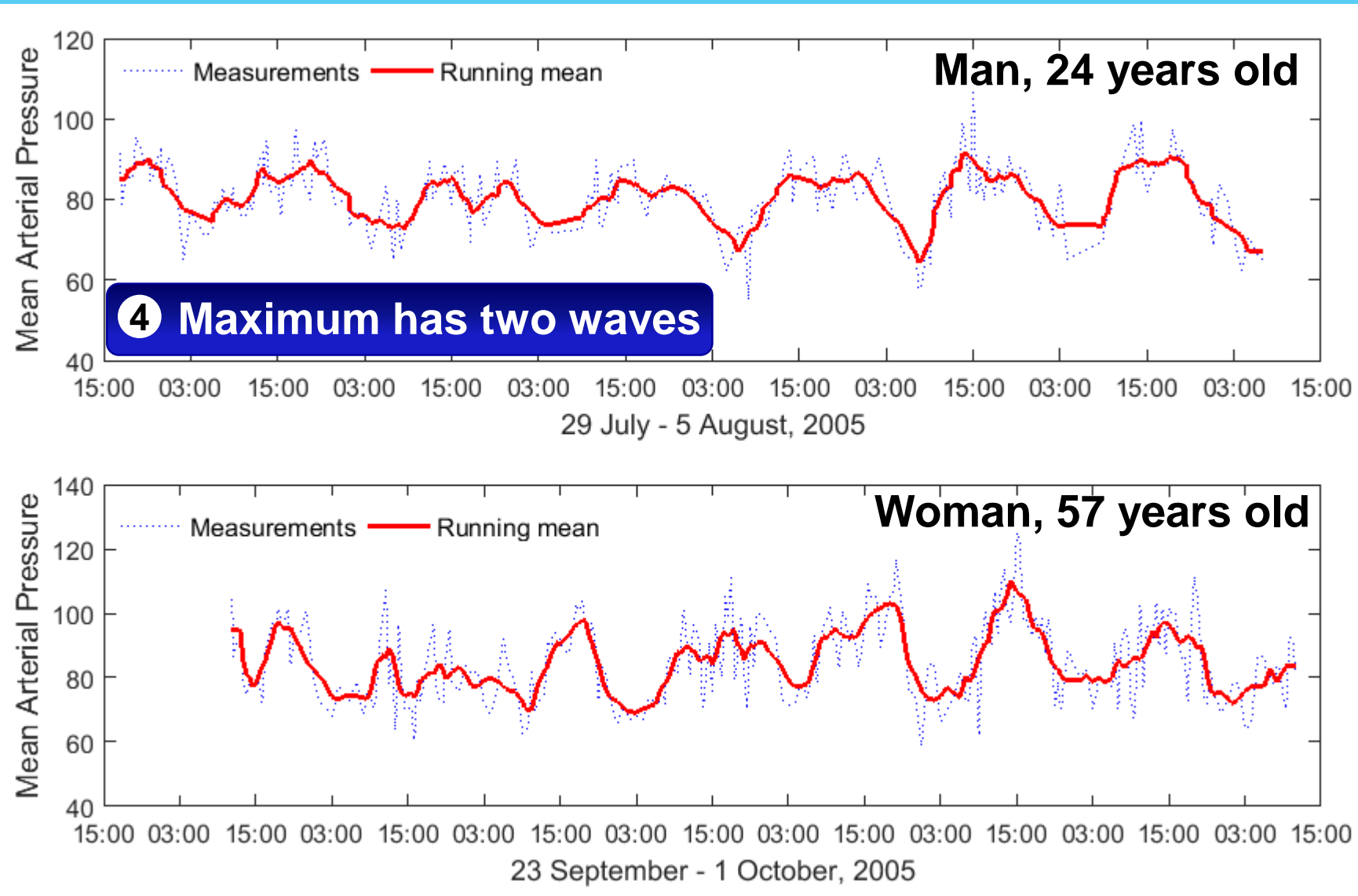
Running mean with window $M = 7$

Which regularities can you extract from smoothed measurements of mean arterial pressure?



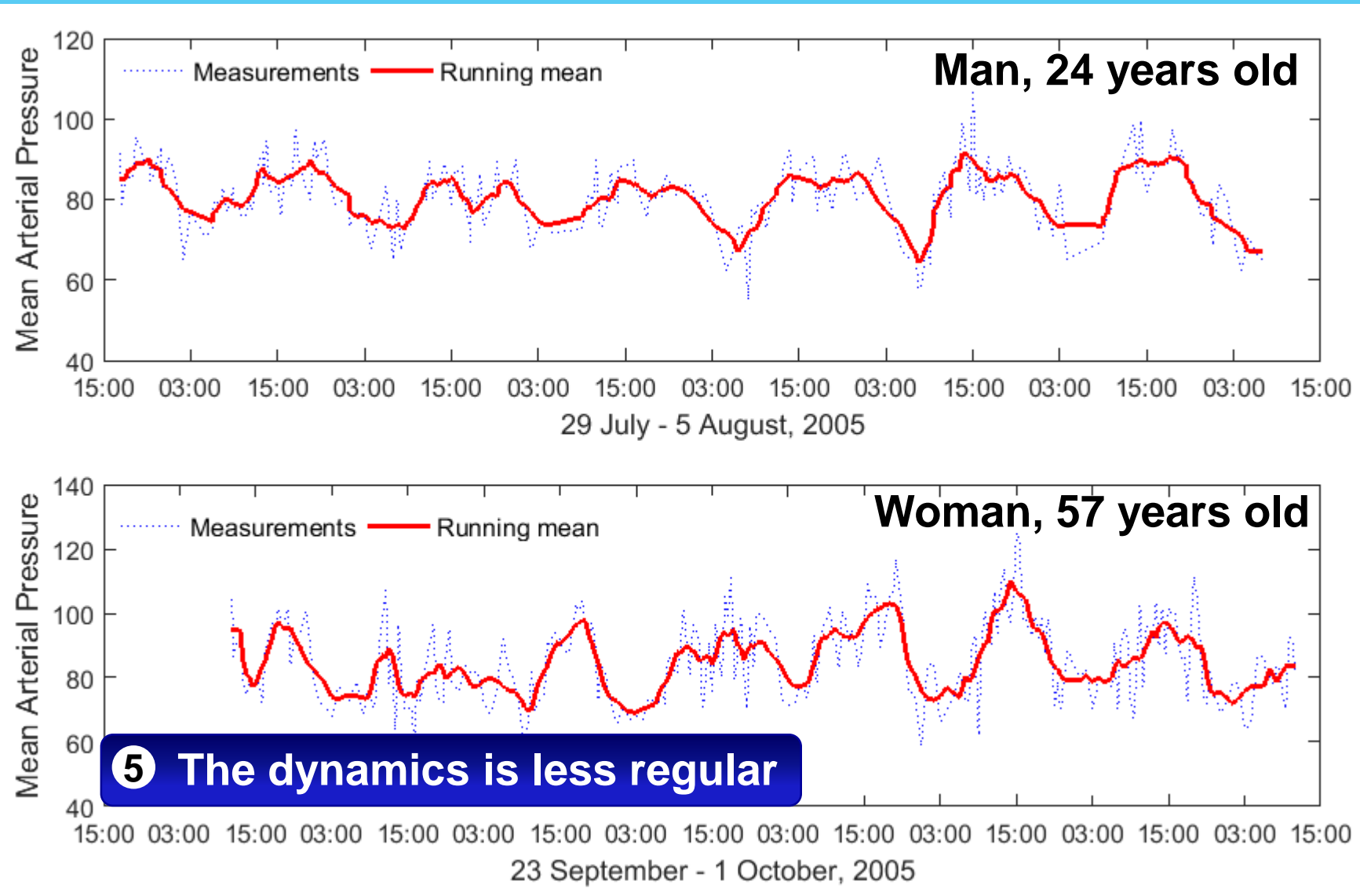
Running mean with window $M = 7$

Which regularities can you extract from smoothed measurements of mean arterial pressure?



Running mean with window $M = 7$

Which regularities can you extract from smoothed measurements of mean arterial pressure?



Running mean with window $M = 7$

① Running mean

**Dilemma
of setting
goal**

**Greater size of
running window**

**Smaller size of
running window**

?

?

www.menti.com
90 97 67 6

① Running mean

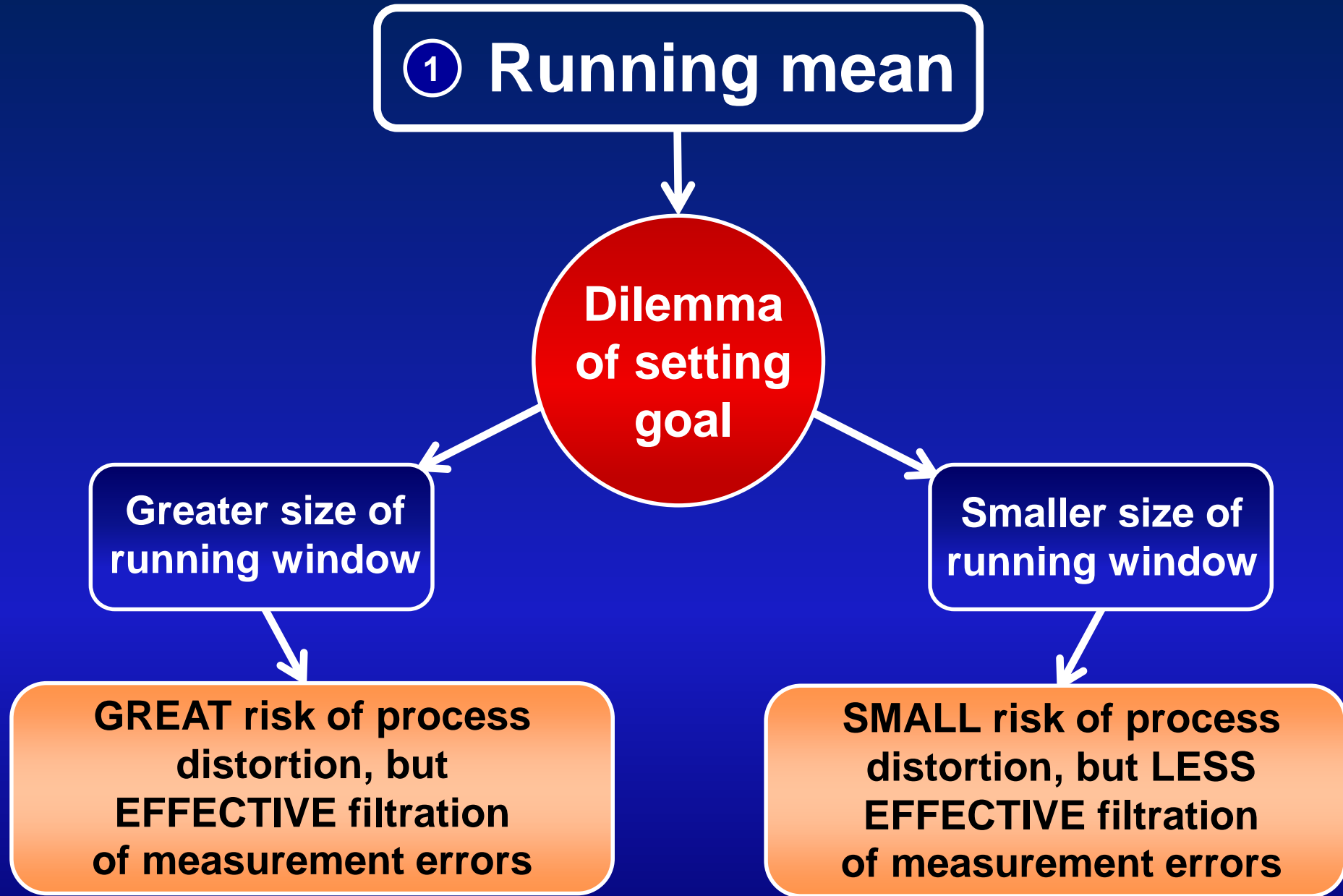
**Dilemma
of setting
goal**

**Greater size of
running window**

**GREAT risk of process
distortion, but
EFFECTIVE filtration
of measurement errors**

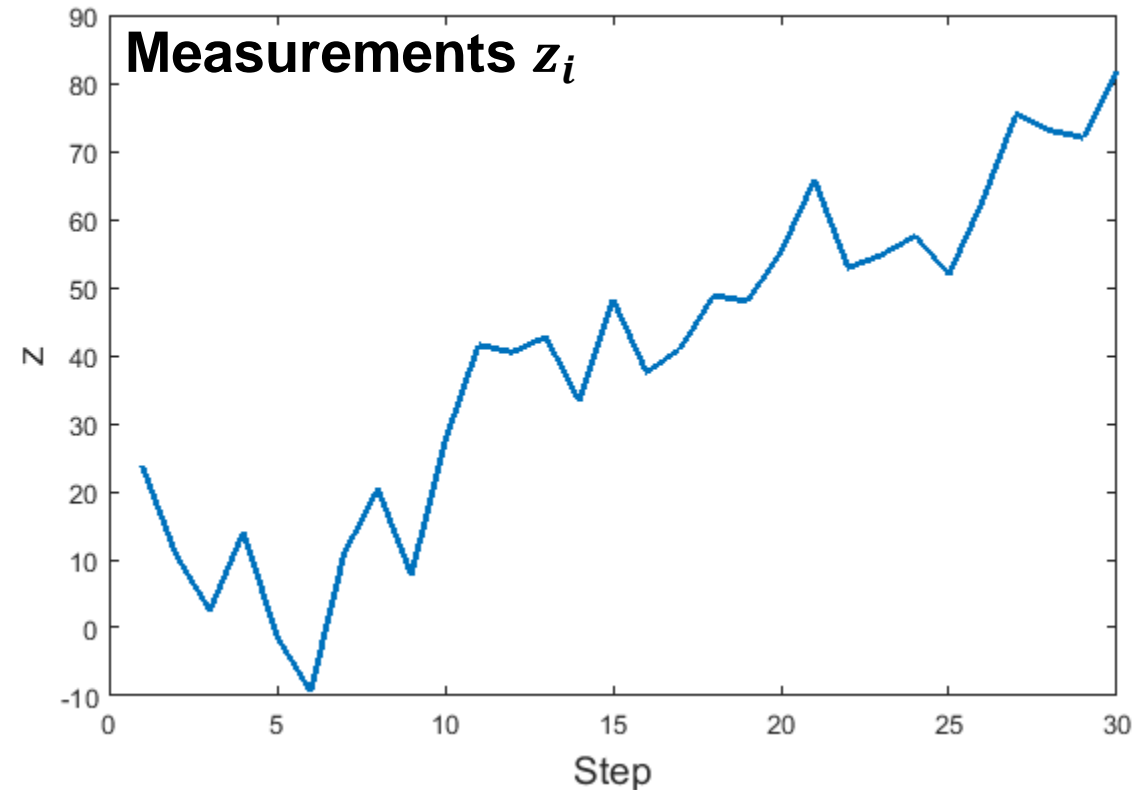
**Smaller size of
running window**

**SMALL risk of process
distortion, but LESS
EFFECTIVE filtration
of measurement errors**



② Exponential smoothing

Let's assume that values of process X are characterized by sudden change

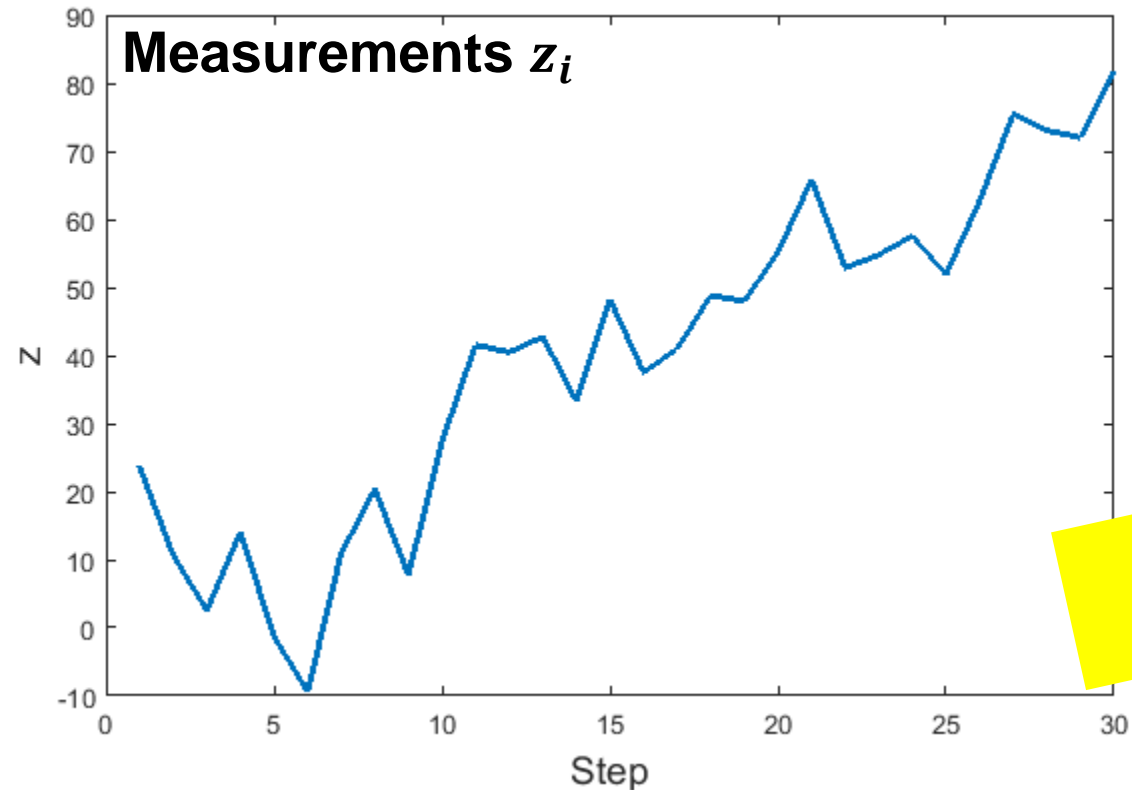


www.menti.com

60 11 34 7

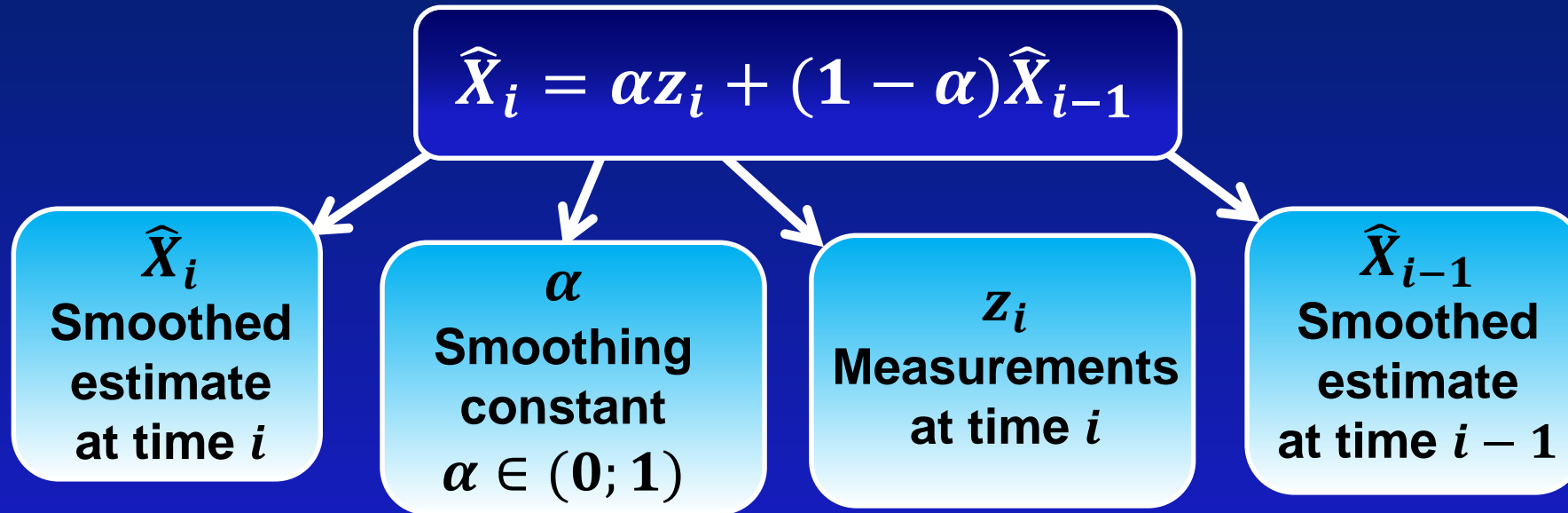
② Exponential smoothing

Let's assume that values of process X are characterized by sudden change



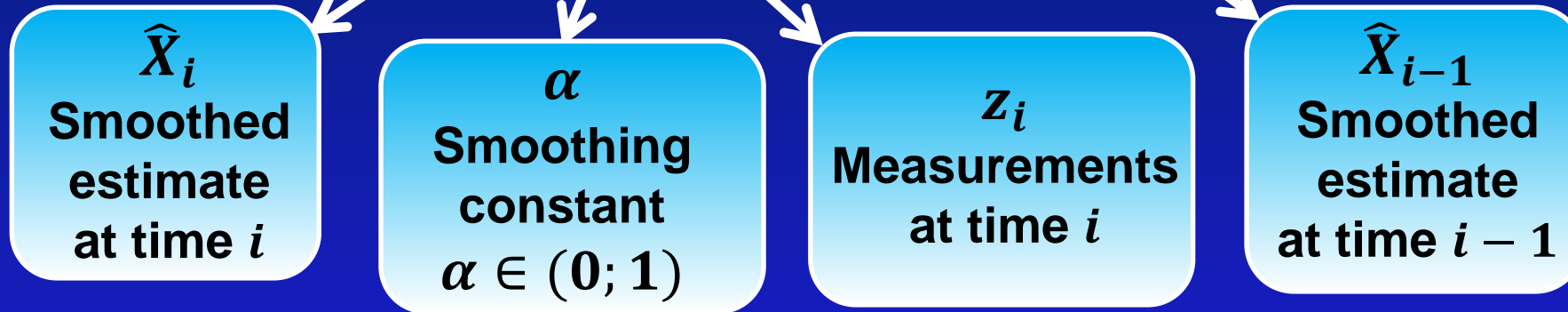
Latest measurements have more confidence

② Exponential smoothing



② Exponential smoothing

$$\hat{X}_i = \alpha z_i + (1 - \alpha)\hat{X}_{i-1}$$



$$\hat{X}_i = \alpha z_i + \alpha(1 - \alpha)z_{i-1} + \alpha(1 - \alpha)^2 z_{i-2} + \cdots + \alpha(1 - \alpha)^i z_0$$

The weight of measurements decreases according to geometric progression or exponential law

② Exponential smoothing: Dilemma of setting goal

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

\hat{X}_{i-1}
Previous estimate

$(z_i - \hat{X}_{i-1})$
Residual – mismatch between
measurement and previous estimate

2 Exponential smoothing: Dilemma of setting goal

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

$$\hat{X}_{i-1}$$

Previous estimate

$$(z_i - \hat{X}_{i-1})$$

Residual – mismatch between
measurement and previous estimate

**SMALLER α , GREATER
confidence to the latest
estimate, SLOWER
reaction to changes**

**But EFFECTIVE
filtration of
measurement errors**

**Choice
of α**

**GREATER α , GREATER
confidence to the latest
measurement, FASTER
reaction to changes**

**But less EFFECTIVE
filtration of
measurement errors**

Comparison of smoothing methods

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

Last M
measurements
are used

2 Exponential mean

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

All previous
measurements
are used

Comparison of smoothing methods

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

Equal weights of
measurements

2 Exponential mean

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

The weight of measurements
decreases according
to exponential law

Comparison of smoothing methods

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

Delay of estimation
on $\frac{M-1}{2}$ steps

2 Exponential mean

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

Estimation is
obtained at a last
available time moment

3

Comparison of smoothing methods

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

$$M = 1$$

2 Exponential mean

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

$$\alpha = 1$$

No filtration
of errors

4

Estimates of both
smoothing methods
are the same

Comparison of smoothing methods

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

$M \rightarrow \infty$

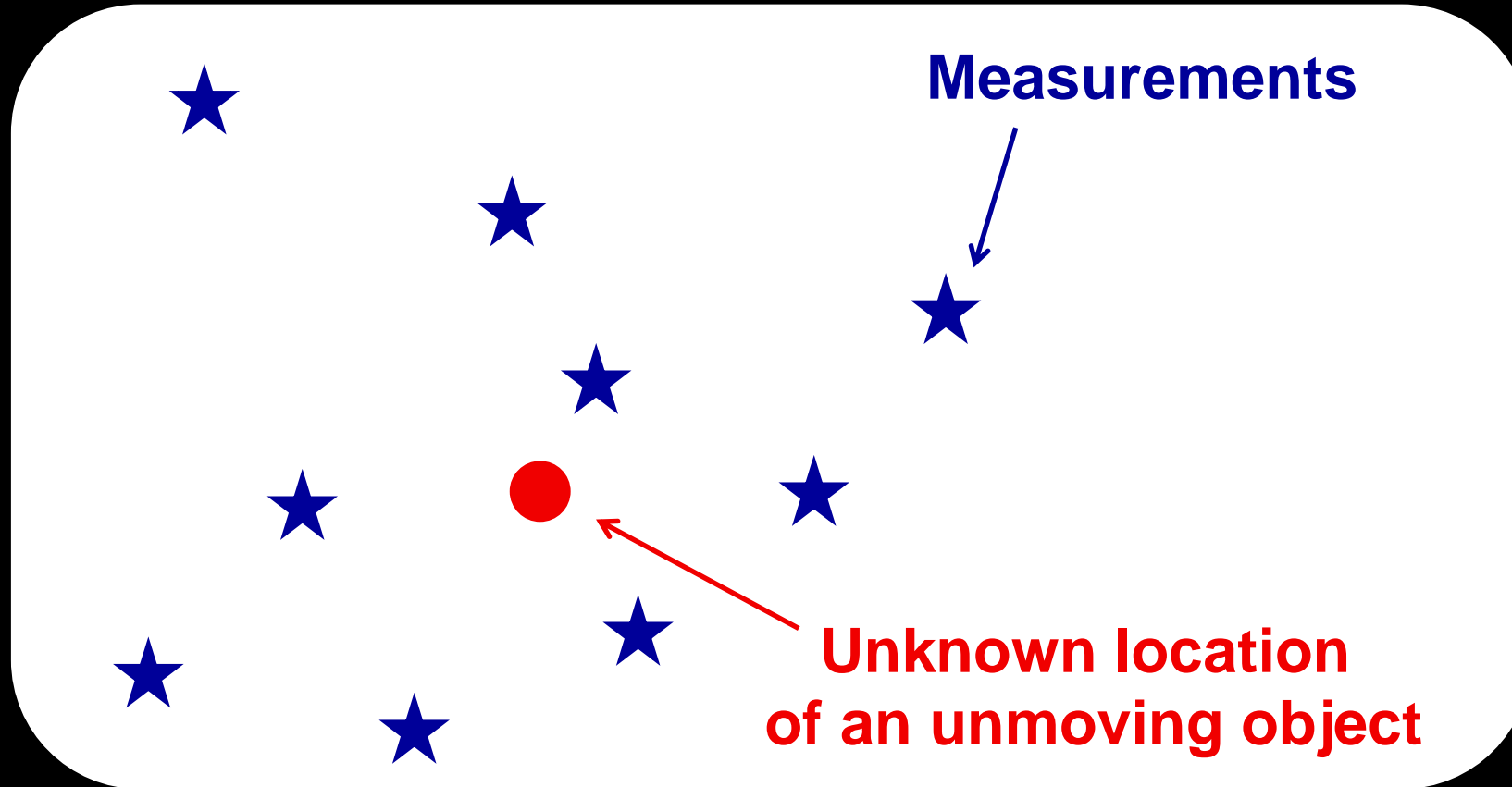
Effective filtration
of errors. But no
reaction to changes
in dynamics

2 Exponential mean

$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

$\alpha \rightarrow 0$

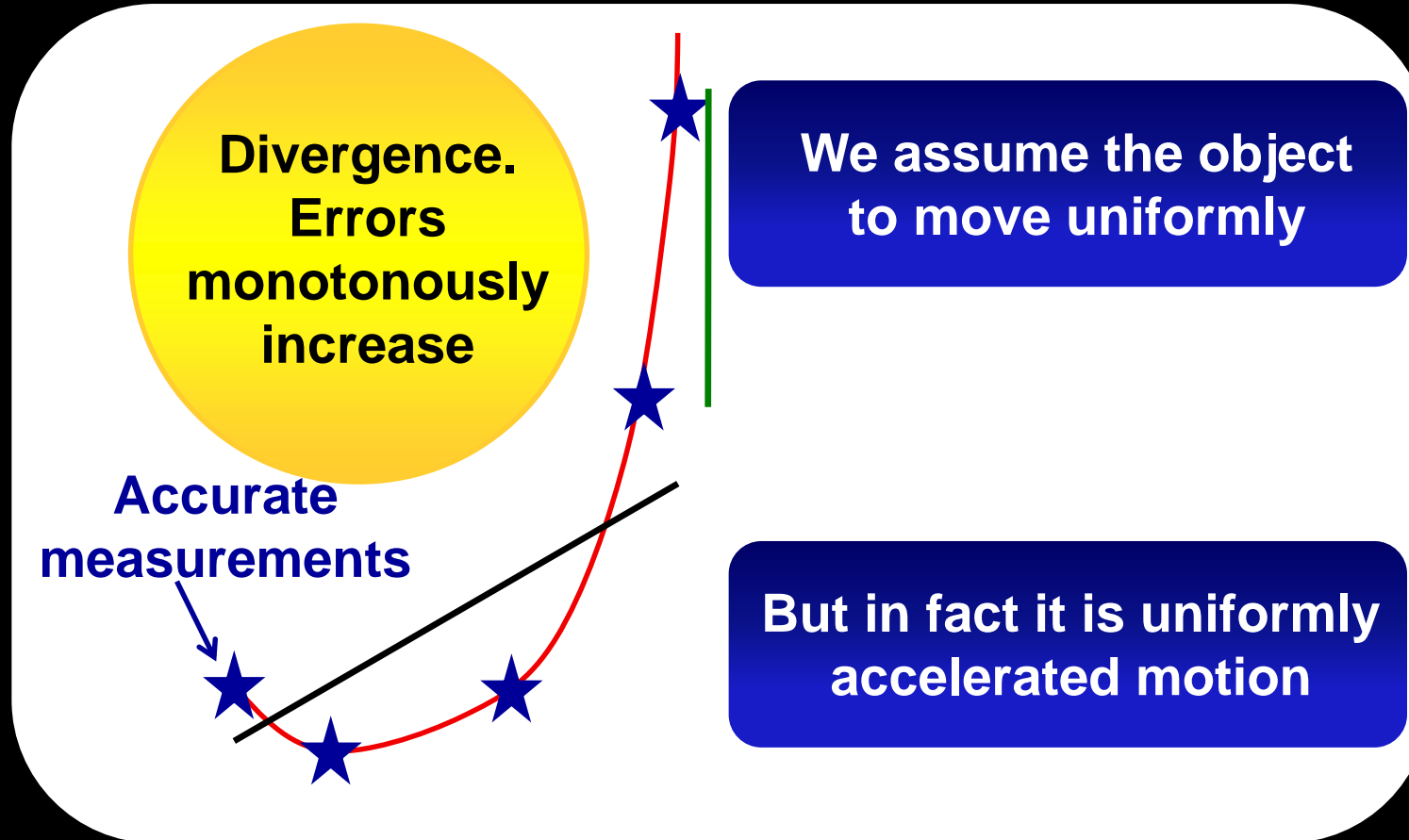
Sources of estimation errors



**Source 1:
Measurement
errors**

**Errors of estimation are related
with only measurements errors.
Model of motion is accurate**

Sources of estimation errors



**Source 2:
Methodical
errors**

**Errors of estimation are
related with errors of methods.
Model of motion is inaccurate.**

Source 1: Measurement errors

1 Running mean

$$\hat{X}_i = \frac{1}{M} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} z_k$$

$$\sigma_{\hat{X}}^2 = \frac{1}{M^2} \sum_{k=i-\frac{M-1}{2}}^{i+\frac{M-1}{2}} \sigma_{\eta}^2$$

$$\sigma_{\hat{X}}^2 = \frac{\sigma_{\eta}^2}{M}$$

2 Exponential mean

$$\hat{X}_i = \alpha \sum_{k=0}^{i-1} (1 - \alpha)^k z_{i-k} + (1 - \alpha)^i z_0$$

$$\lim_{i \rightarrow \infty} \sigma_{\hat{X}}^2 = \lim_{i \rightarrow \infty} \left(\alpha^2 \sigma_{\eta}^2 \sum_{k=0}^{i-1} (1 - \alpha)^{2k} \right)$$

$$\sigma_{\hat{X}}^2 = \sigma_{\eta}^2 \frac{\alpha}{2 - \alpha}$$

Source 1: Measurement errors

1 Running mean

$$\sigma_{\hat{X}}^2 = \frac{\sigma_{\eta}^2}{M}$$

$$M = 1$$

No filtration
of errors

2 Exponential mean

$$\sigma_{\hat{X}}^2 = \sigma_{\eta}^2 \frac{\alpha}{2 - \alpha}$$

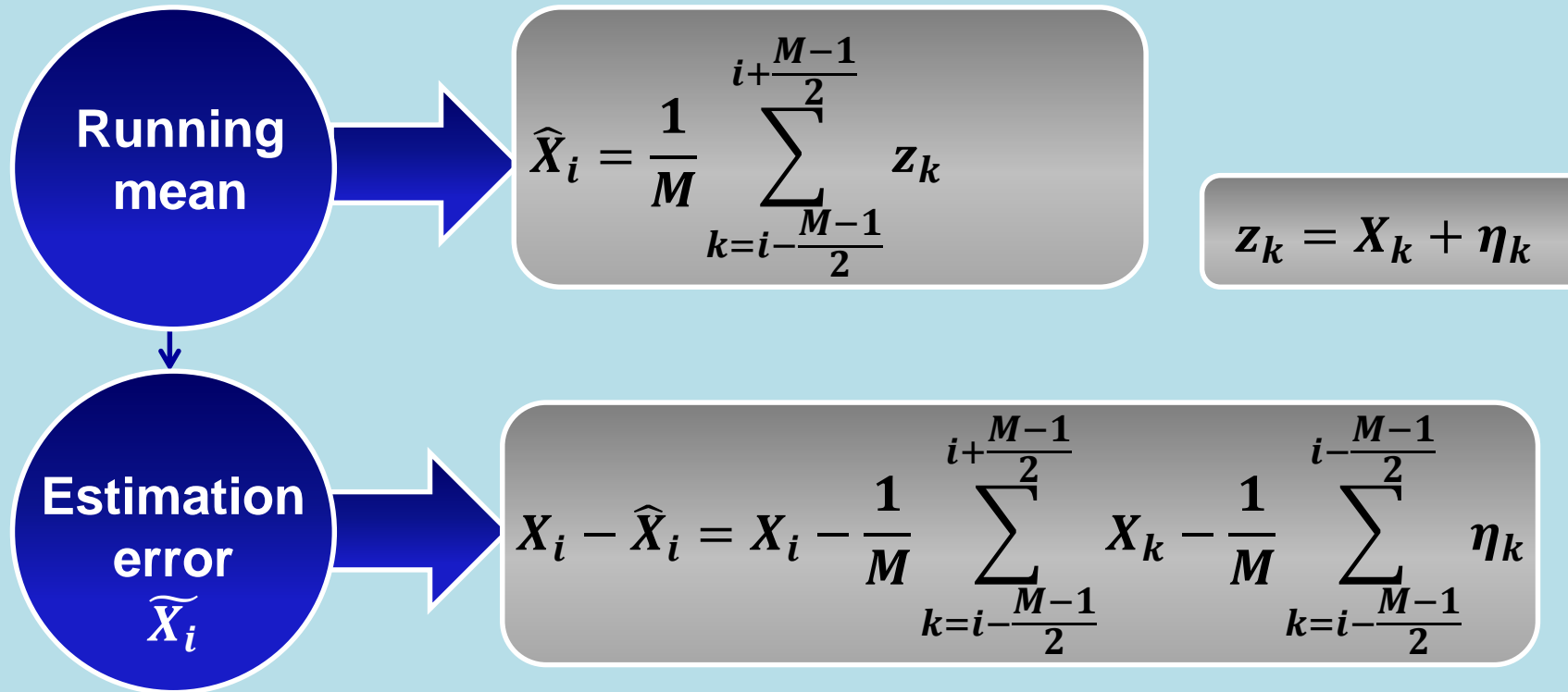
$$\alpha = 1$$

$$M \rightarrow \infty$$

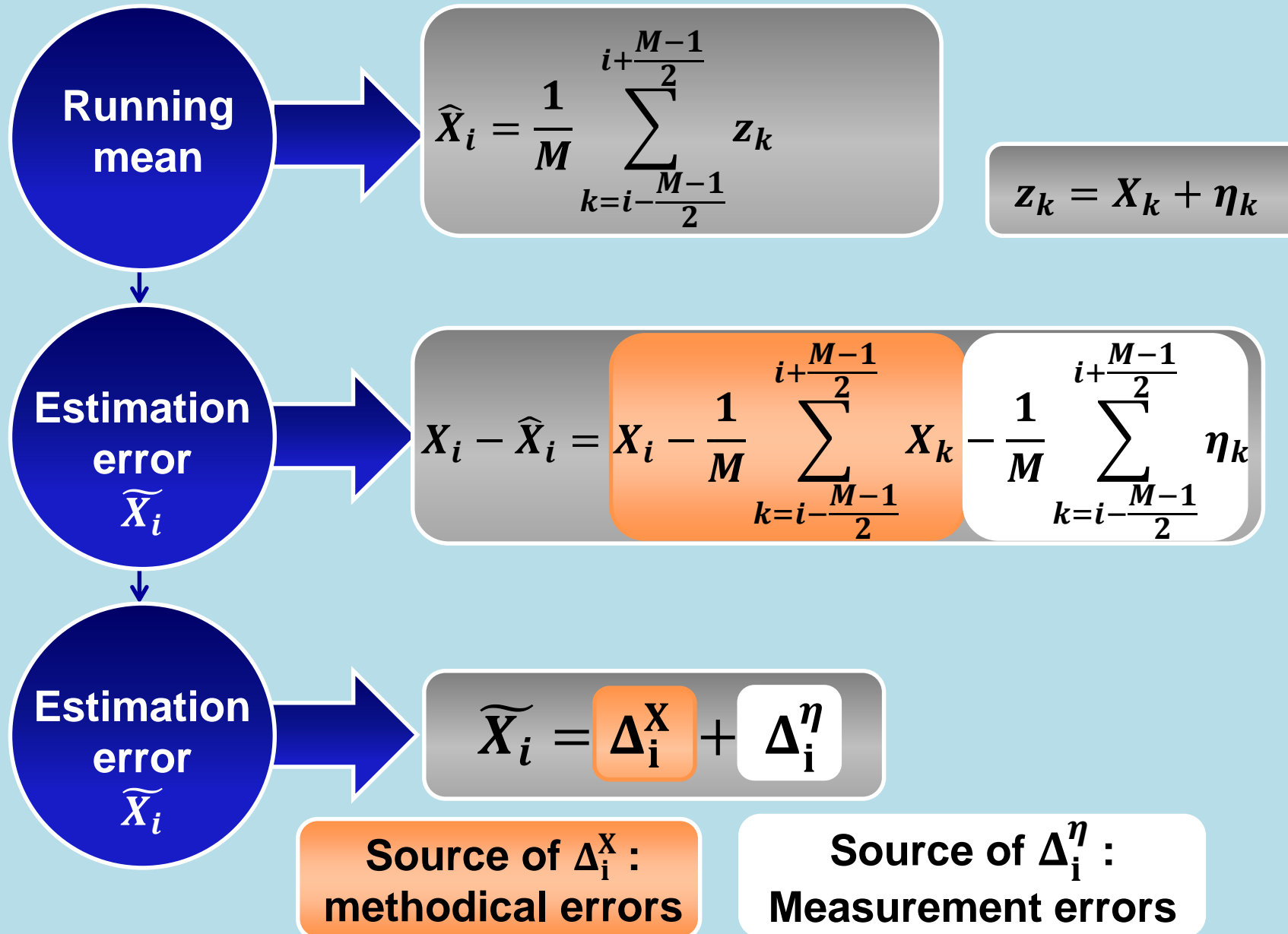
Effective filtration
of errors. But no
reaction to changes
in dynamics

$$\alpha \rightarrow 0$$

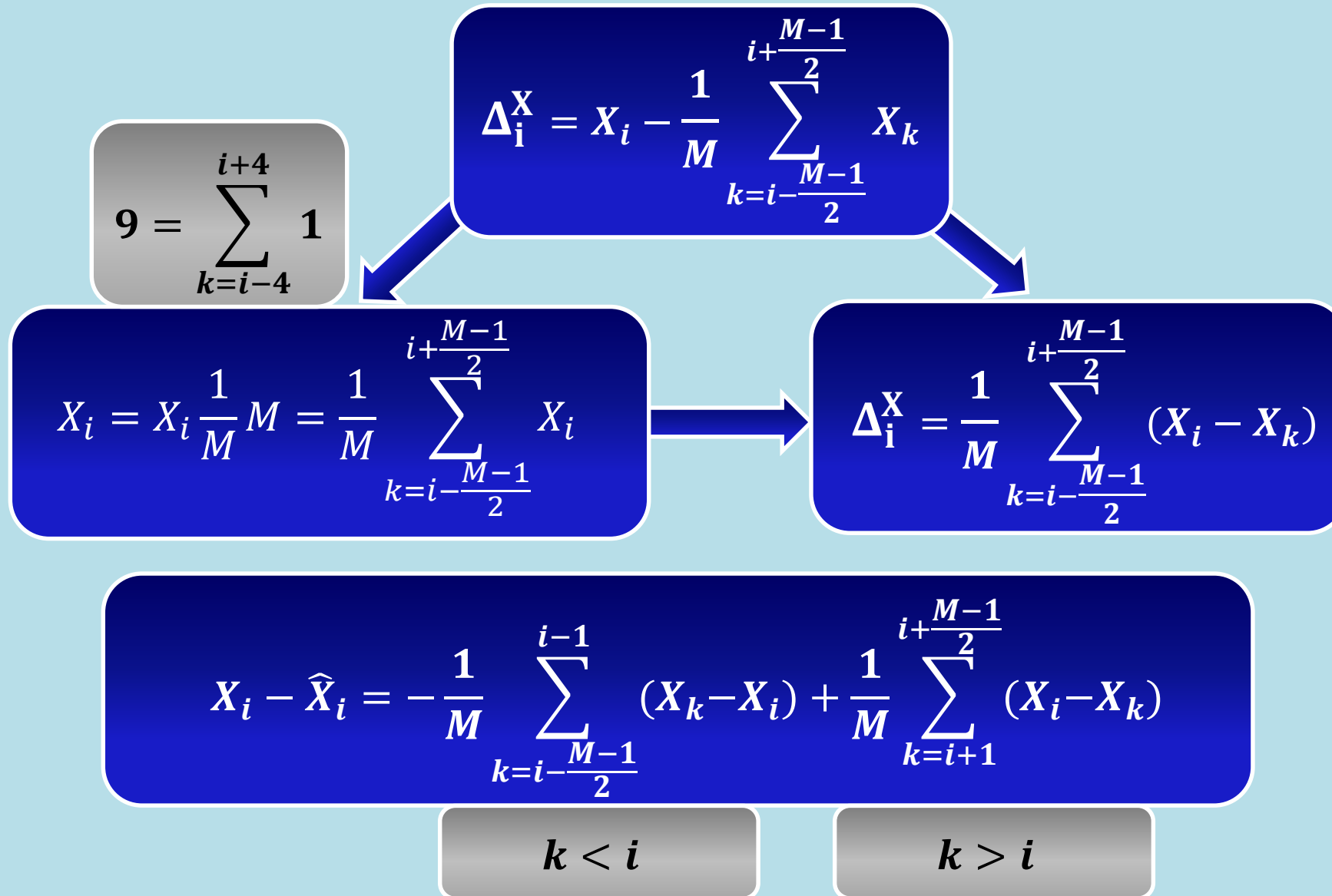
Source 2: Methodical errors of running mean



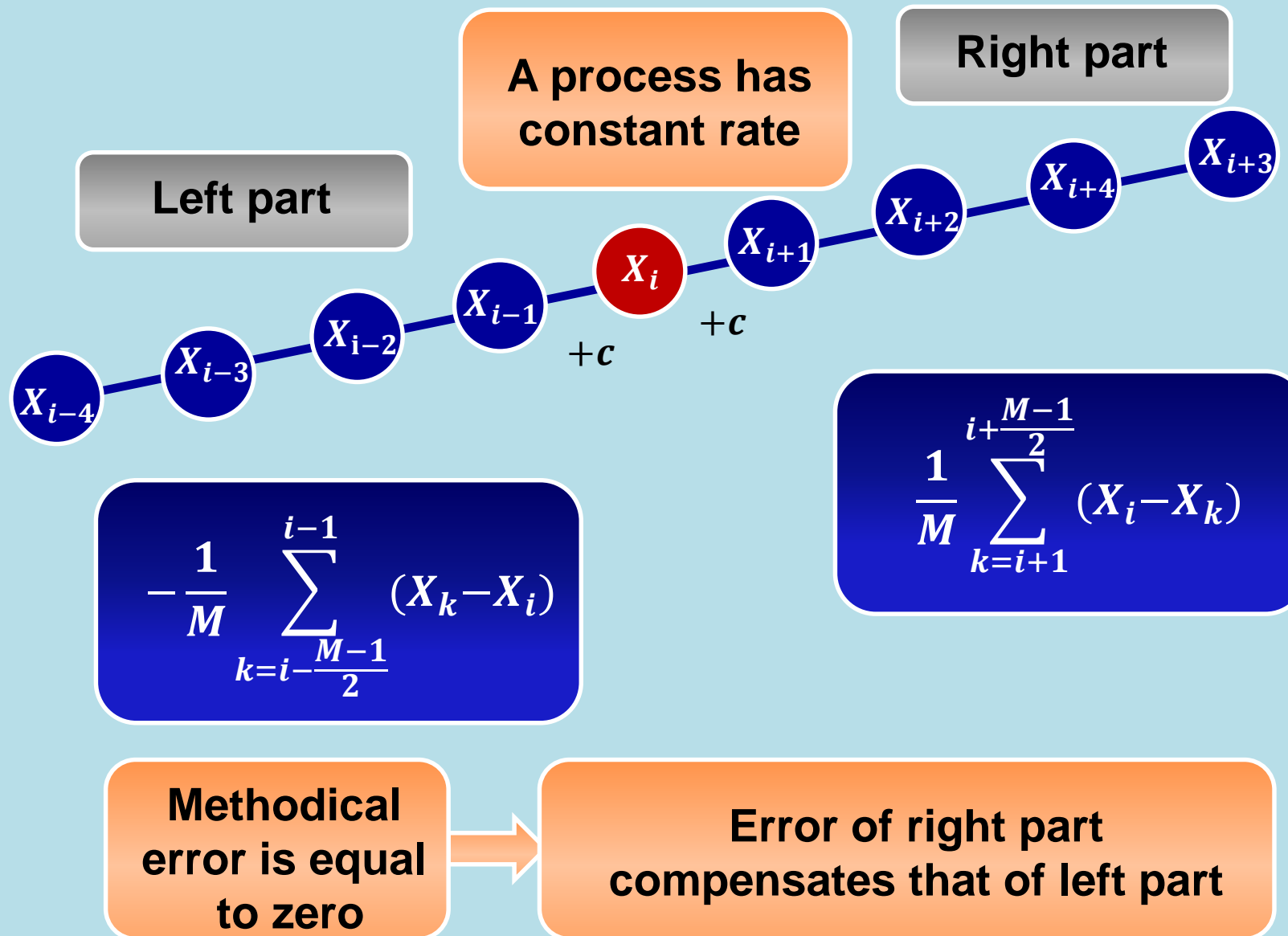
Source 2: Methodical errors of running mean



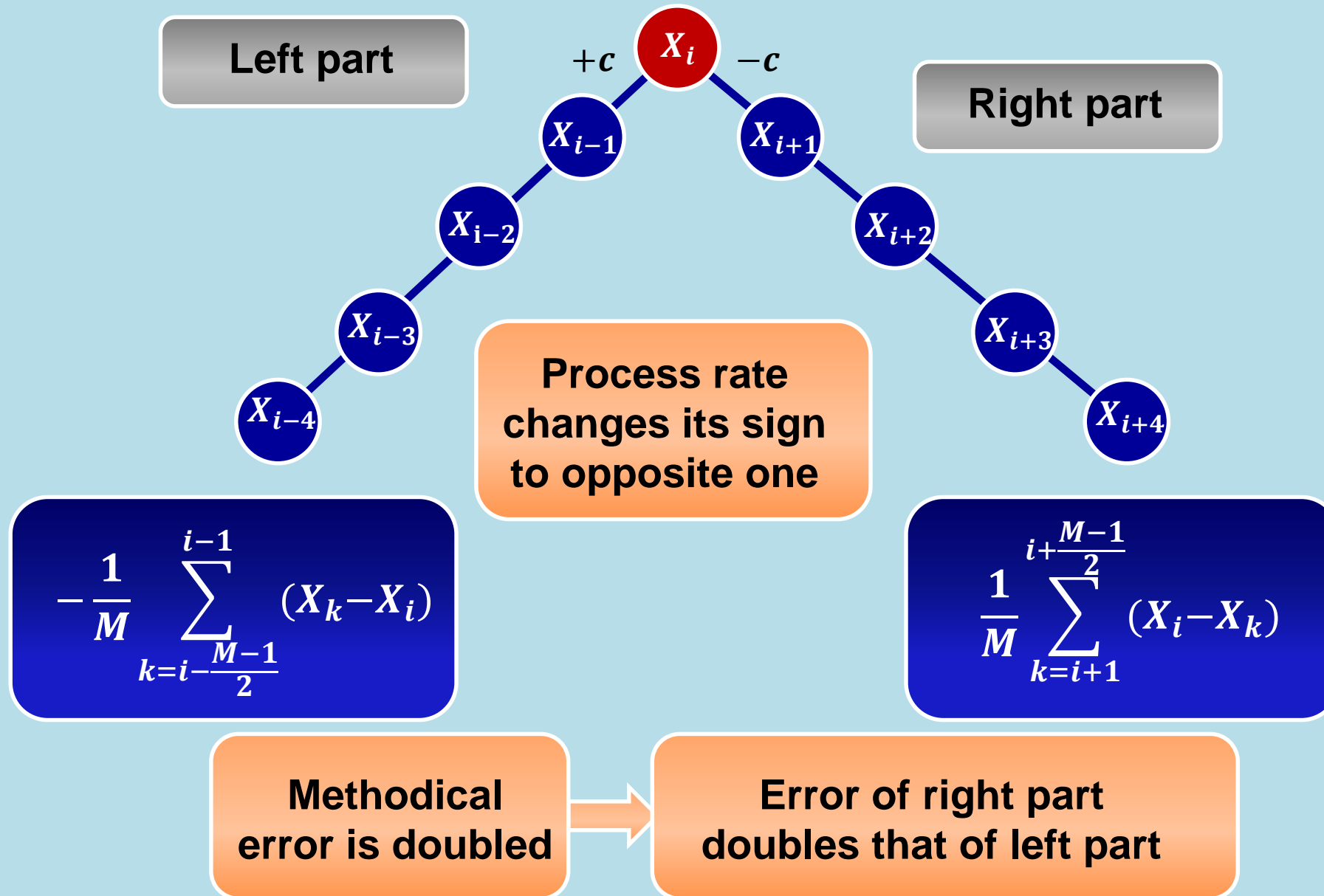
Source 2: Methodical errors of running mean



Source 2: Methodical errors of running mean



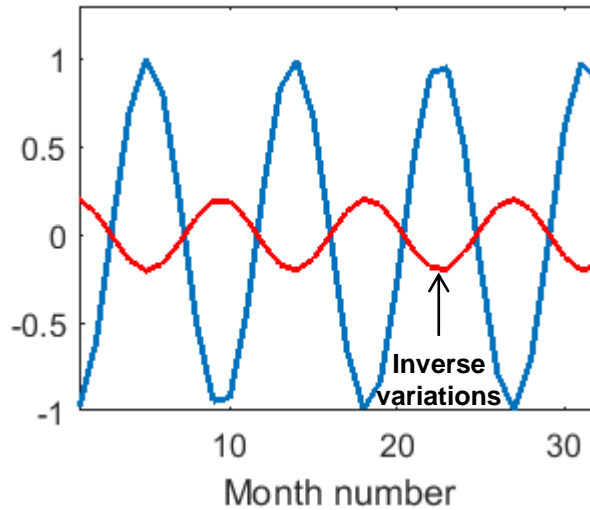
Source 2: Methodical errors of running mean



Analysis of running mean errors

Running mean may significantly distort the dynamics of the process

9-months variation

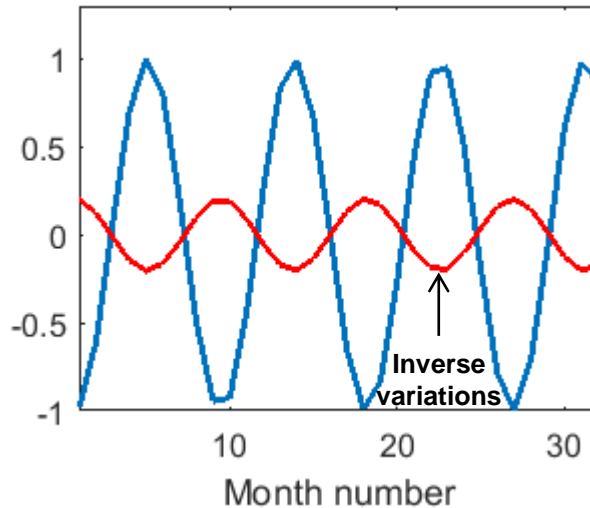


Inverse variations
with periods from
6 to 12 months.
Convex curve is
replaced by concave
curve and vice versa

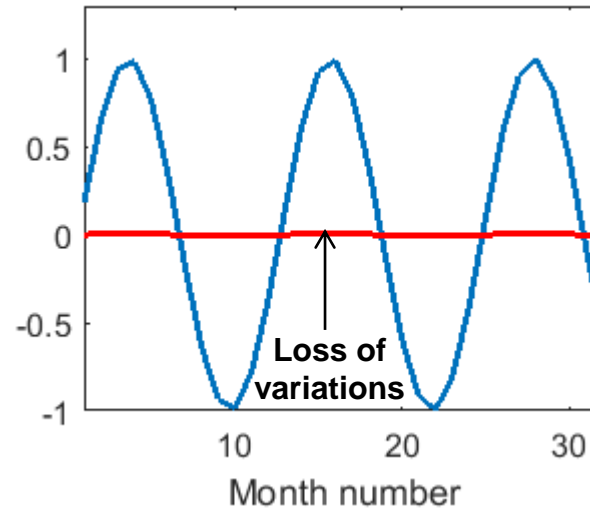
Analysis of running mean errors

Running mean may significantly distort the dynamics of the process

9-months variation



13-months variation



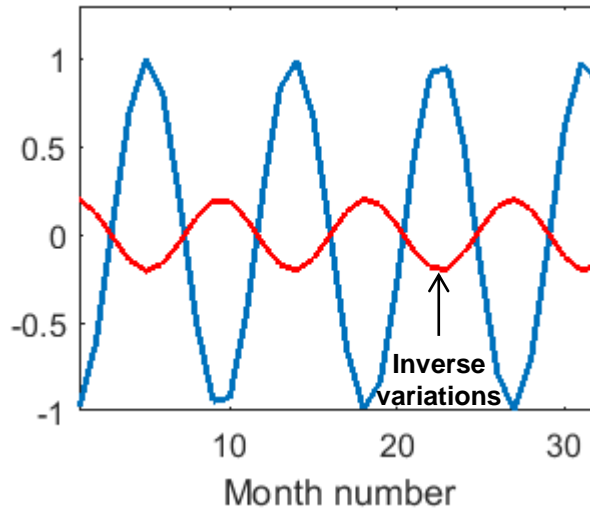
Inverse variations with periods from 6 to 12 months. Convex curve is replaced by concave curve and vice versa

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

Analysis of running mean errors

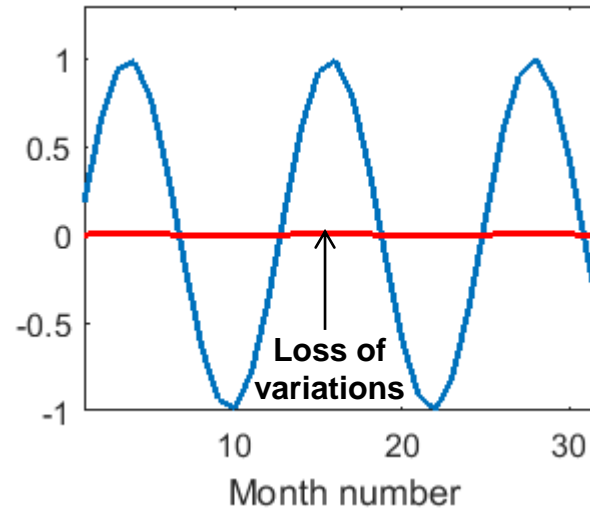
Running mean may significantly distort the dynamics of the process

9-months variation



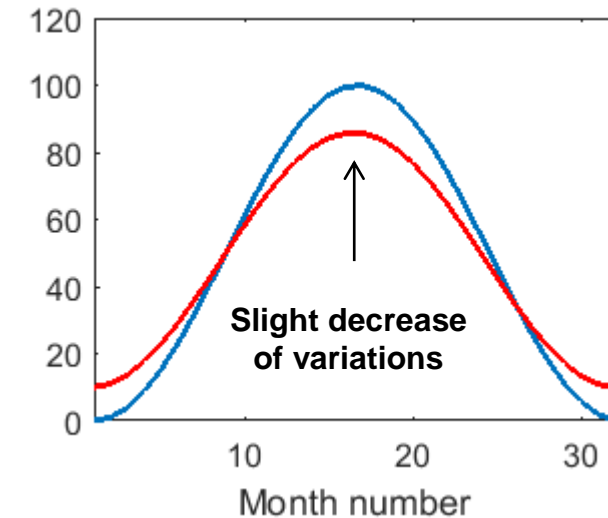
Inverse variations with periods from 6 to 12 months. Convex curve is replaced by concave curve and vice versa

13-months variation



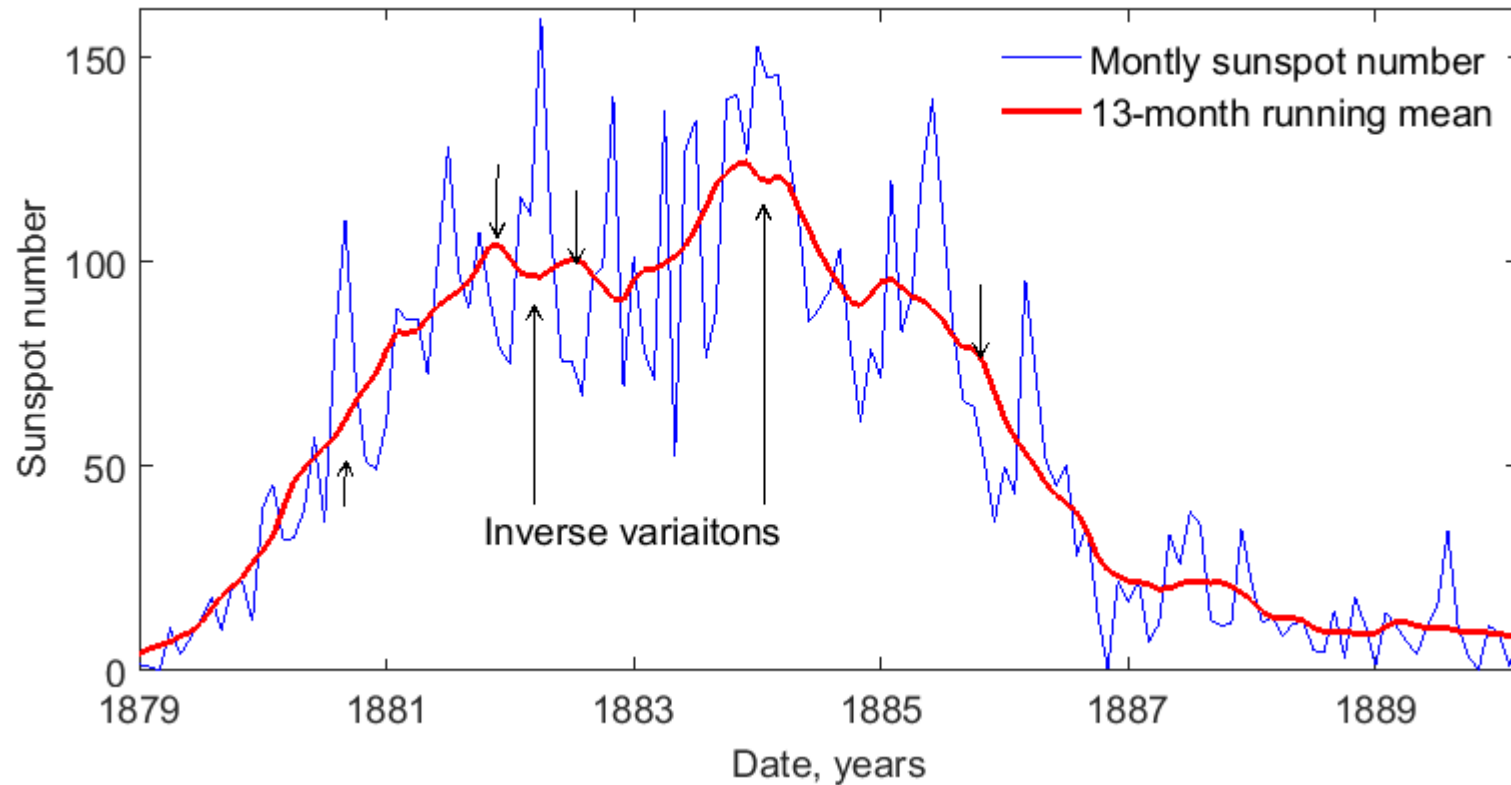
Total loss of 6- and 12-month variations decreasing them to zero

32-months variation



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

Conclusions

Don't apply methods in blind to not fall into the trap leading to false conclusions

Even if implementation is simple, the method itself requires careful analysis

Exponential smoothing

```
graph TD; A[Exponential smoothing] --> B["\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})"]; B --> C[Errors of exponential smoothing due to measurement errors]; C --> D["\sigma_{\hat{X}}^2 = \sigma_{\eta}^2 \frac{\alpha}{2 - \alpha}"];
```

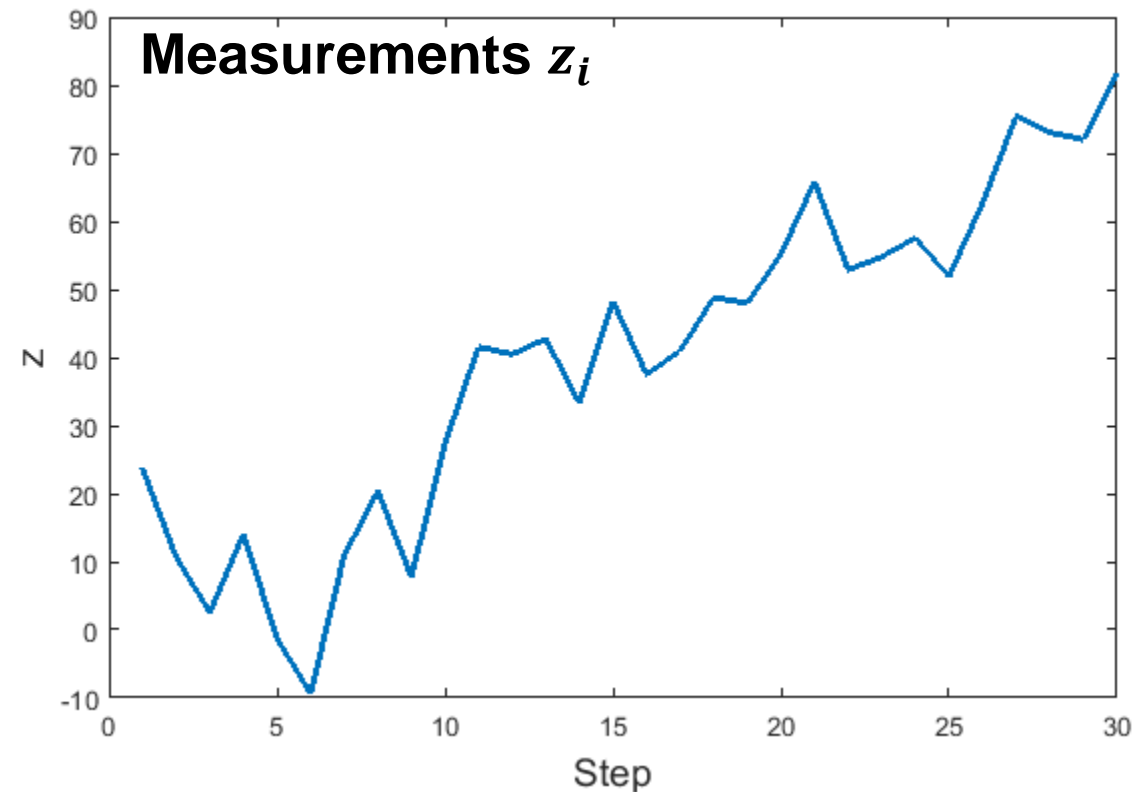
$$\hat{X}_i = \hat{X}_{i-1} + \alpha(z_i - \hat{X}_{i-1})$$

Errors of exponential smoothing
due to measurement errors

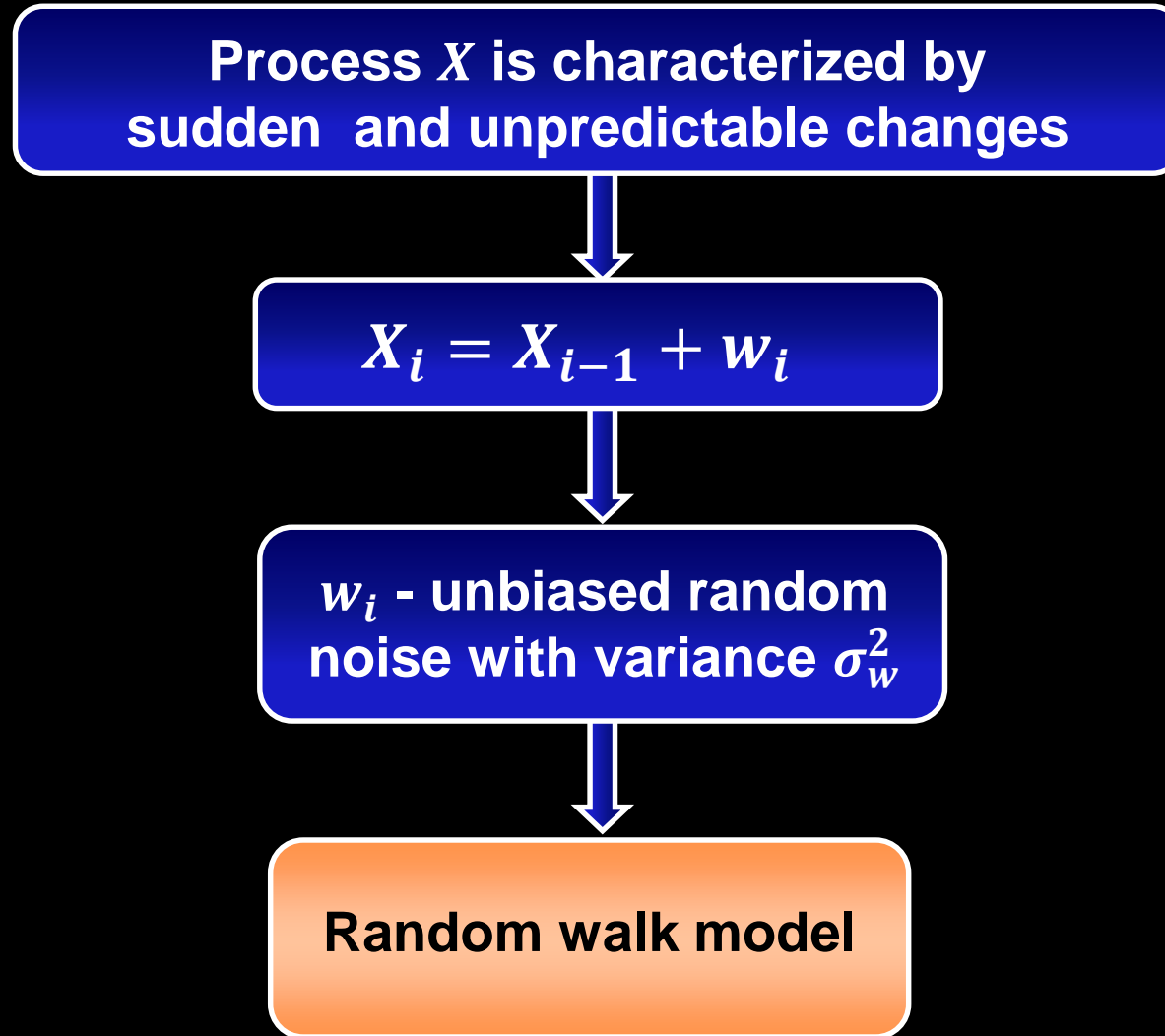
$$\sigma_{\hat{X}}^2 = \sigma_{\eta}^2 \frac{\alpha}{2 - \alpha}$$

Optimal choice of smoothing constant α

Process X is characterized by sudden and unpredictable changes



Optimal choice of smoothing constant α



Optimal choice of smoothing constant α

Optimal α
for random
walk model

$$\alpha = \frac{-\chi + \sqrt{\chi^2 + 4\chi}}{2}$$

$$\chi = \frac{\sigma_w^2}{\sigma_\eta^2}$$

σ_η^2 - variance
of measurement
noise

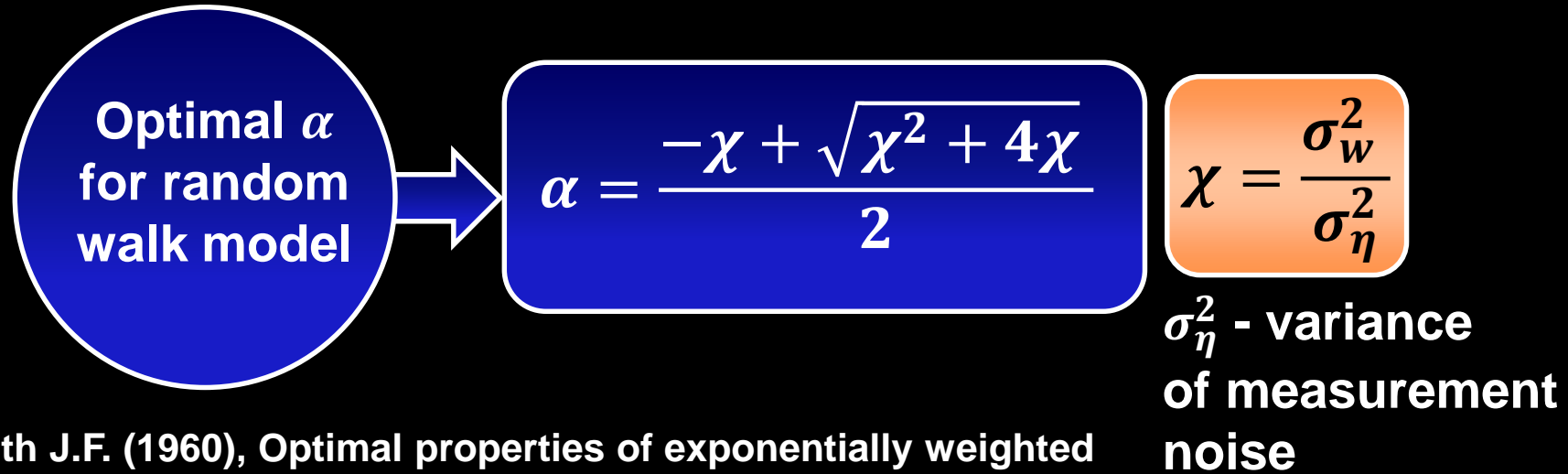
Muth J.F. (1960), Optimal properties of exponentially weighted forecasts of time series with permanent and transitory components, J.Amer. Statist. Ass. 01960.-Vol.55.-p.299.

Full error
of smoothing

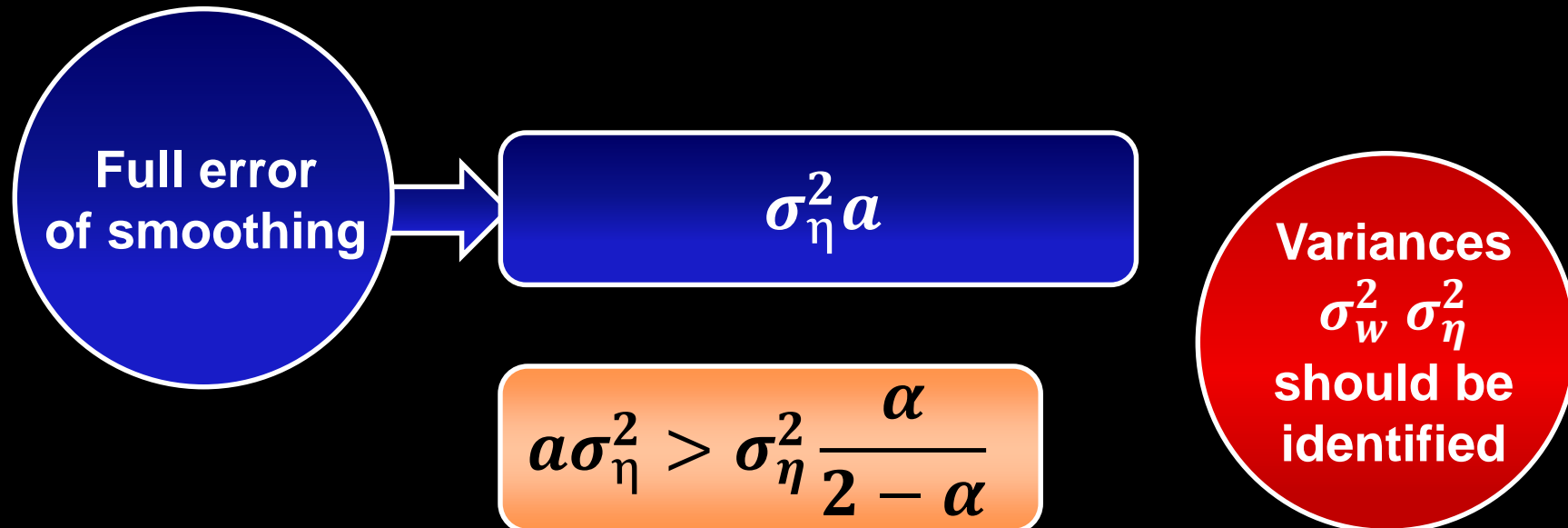
$$\sigma_\eta^2 \alpha$$

$$a\sigma_\eta^2 > \sigma_\eta^2 \frac{\alpha}{2 - \alpha}$$

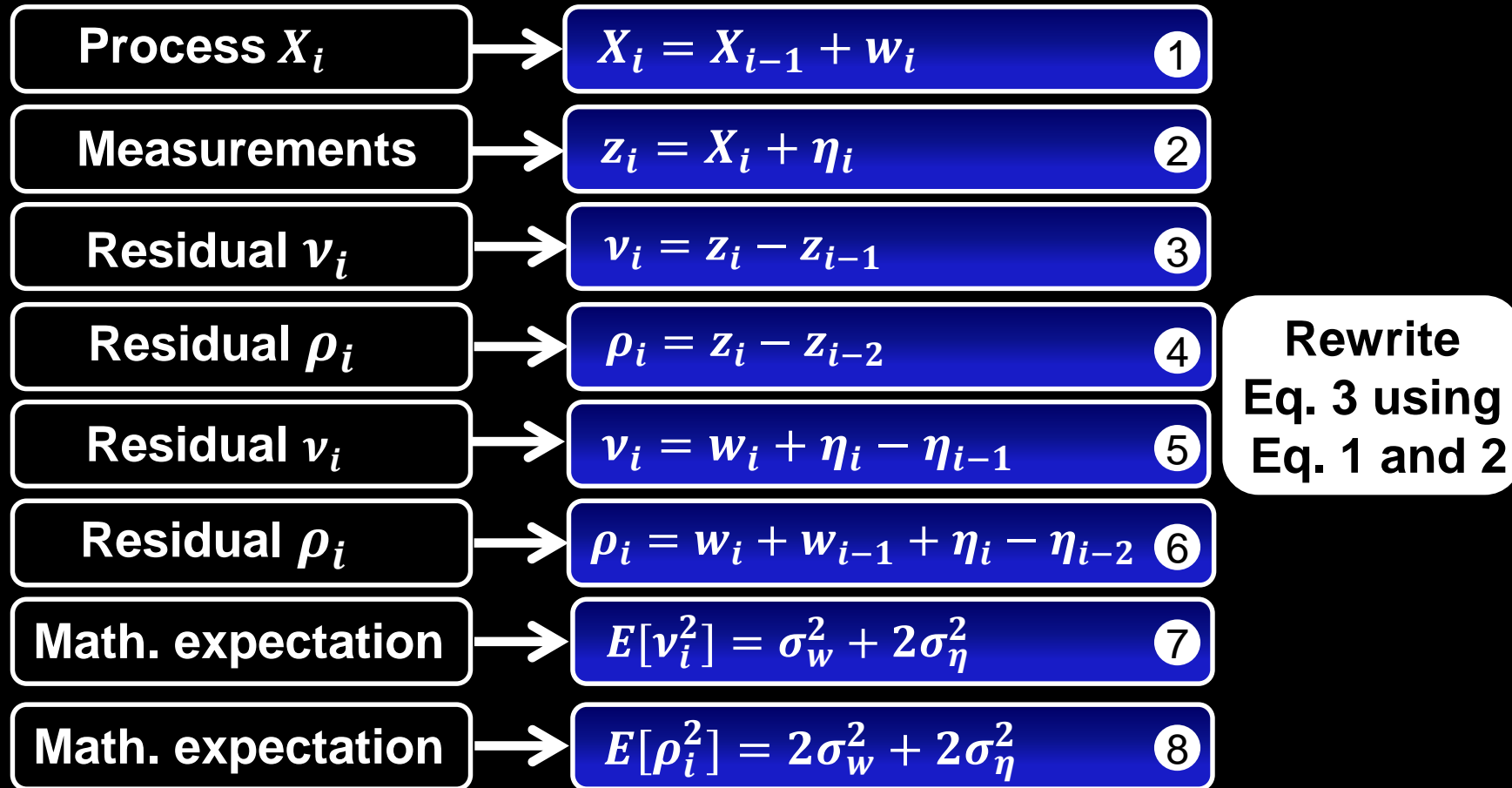
Optimal choice of smoothing constant α



Muth J.F. (1960), Optimal properties of exponentially weighted forecasts of time series with permanent and transitory components, J.Amer. Statist. Ass.01960.-Vol.55.-p.299.

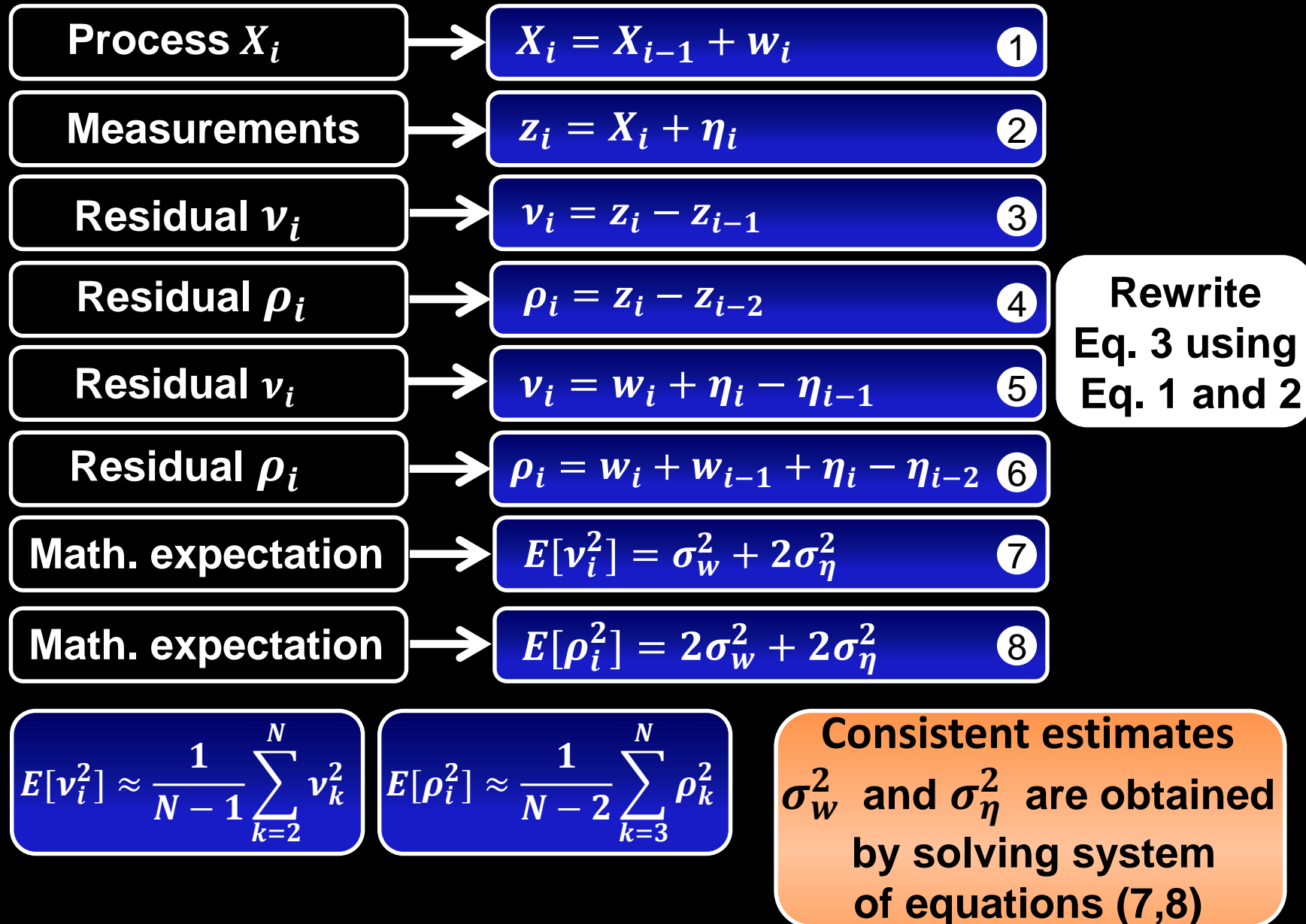


Identification of noise statistics σ_w^2 and σ_η^2



Anderson, W. N., G. B. Kleindorfer, P. R. Kleindorfer,
and M. B. Woodroffe (1969), Consistent estimates
of the parameters of a linear system, Ann. Math. Stat., 40(3), 2064–2075.

Identification of noise statistics σ_w^2 and σ_η^2



Thank you for your attention!

www.menti.com

56 00 12 5