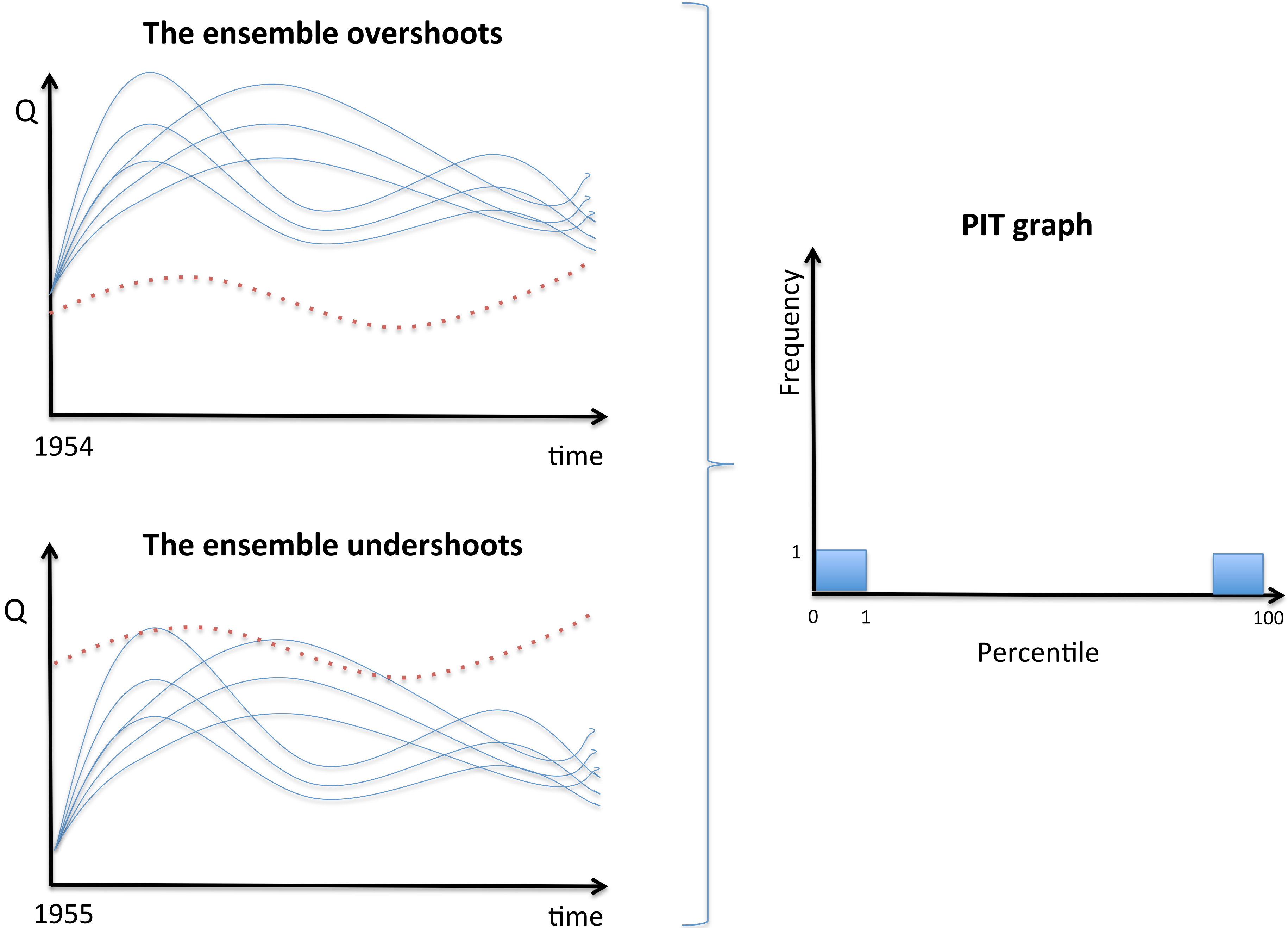
Eighth Montreal Industrial Problem Solving Workshop - Rio Tinto Presentation

K. A. Alahassa, S. Amraoui, R. Arsenault, N. Ayi, C. Budd, P. Duchesne, S. Ibrahim, D. Jovmir, S. H. Jun, M. Latraverse, T. Y. Lee, C.P. Liou, C. Poissant, A. Poterie, V. Rochon Montplaisir, L. Sarrazin

August 11, 2017


The ensemble overshoots
Q
1954
time
The ensemble undershoots
Q
1955
time
PIT graph
Frequency
1
0
1
100
Percentile

Presentation

Different approaches explored in this presentation:

1- Classification scenarios

2- Optimization in the initial state in summer

3- Time-series

4- Gaussian model

5- Filtering methods

1. Classification scenarios

Extension to Summer : Underground and Classification

Winter : ΔV method on the **Snow**

Extension to Summer : Underground and Classification

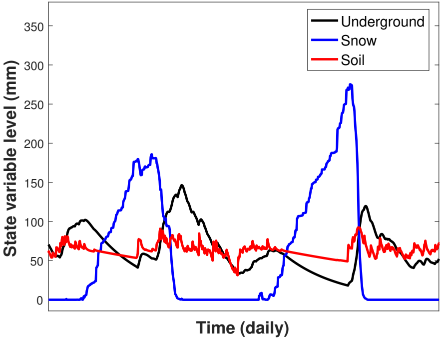
Winter : ΔV method on the **Snow**

Summer ?

Extension to Summer : Underground and Classification

Winter : ΔV method on the **Snow**

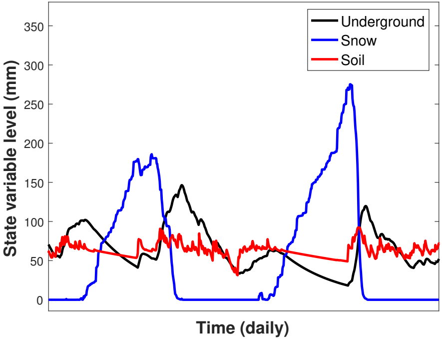
Summer ?



Extension to Summer : Underground and Classification

Winter : ΔV method on the **Snow**

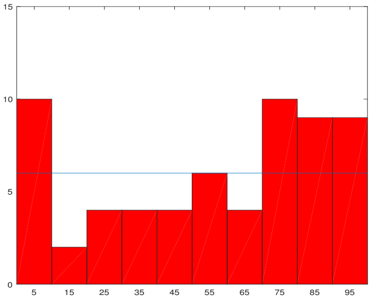
Summer ?



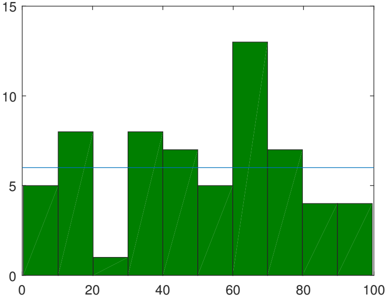
$\Rightarrow$ Summer : ΔV method on the **Underground**

Naive approach:

Not corrected (p_{value}=0.05)

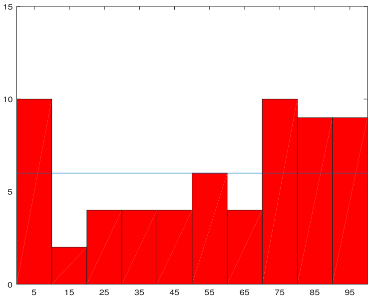


Corrected (p_{value}=0.34)

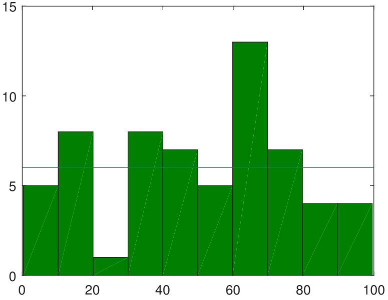


Naive approach:

Not corrected (p_{value}=0.05)



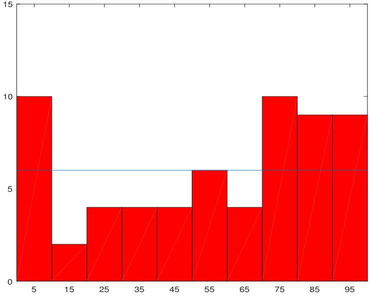
Corrected (p_{value}=0.34)



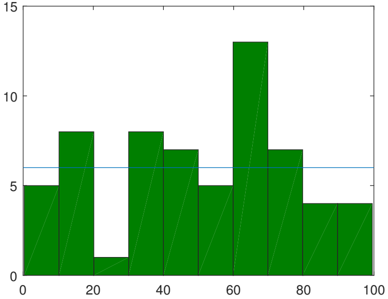
In winter, **only one** phenomenon = snow accumulation.

Naive approach:

Not corrected (p_{value}=0.05)



Corrected (p_{value}=0.34)

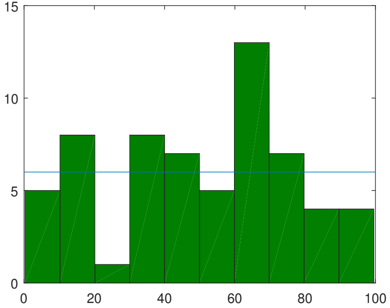


In winter, **only one** phenomenon = snow accumulation.

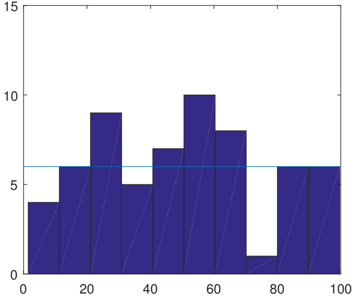
In summer, **more complicated** phenomena = it depends on the level of water in the soil and underground.

Idea= Separate the different categories for the ΔV method : **dry**, **medium**, **wet**. Choice of the threshold by an empirical method.

Corrected (p_{value}=0.34)

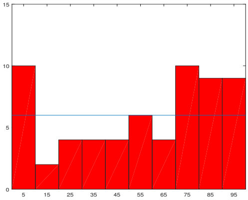


Conditioned corrected (p_{value}=0.17)

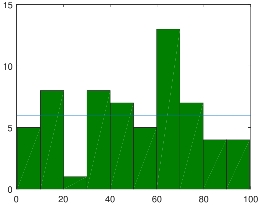


Idea= Separate the different categories for the ΔV method : **dry**, **medium**, **wet**. Choice of the threshold by an empirical method.

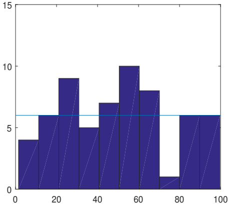
Not corrected



Corrected

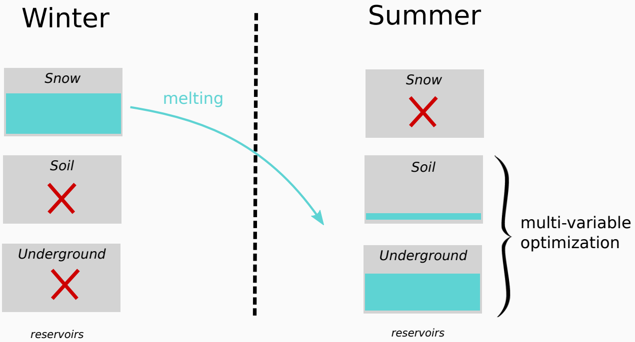


Conditioned corrected



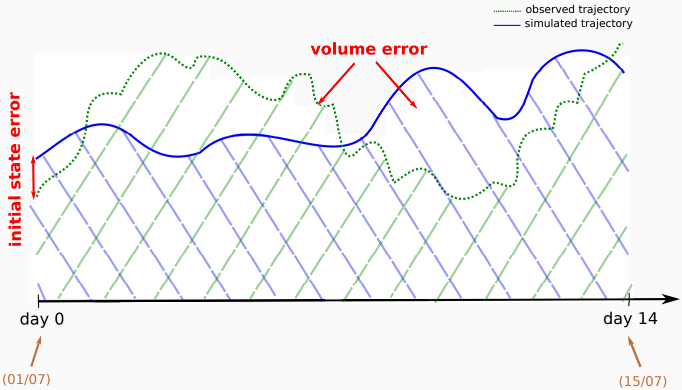
2. Optimization in the initial state in summer

Summer optimization problem



In **winter** the only variable in the optimization process that we can play with is **snow!**
In **summer**, we have to handle **soil AND underground water** and optimize both variables.

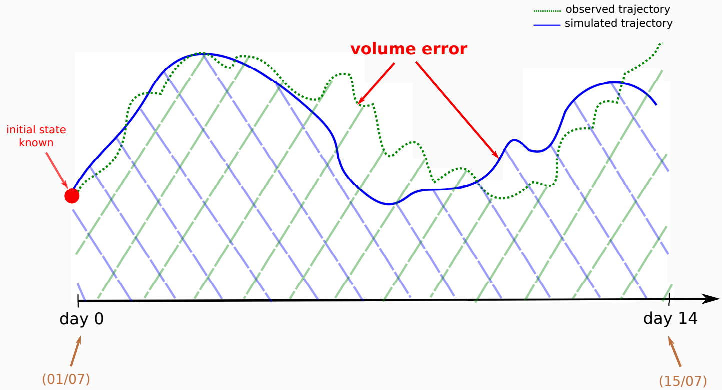
Summer optimization problem



Optimization problem :

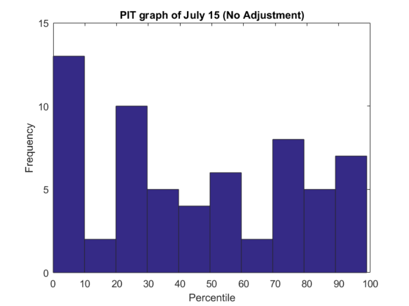
$$\min_{\color{red}{x_0}, \delta V} \left| \int_0^T \left(V_{observed}(t) - V_{simulated}(t)\right) dt \right|, \quad T > 0$$

Summer optimization problem

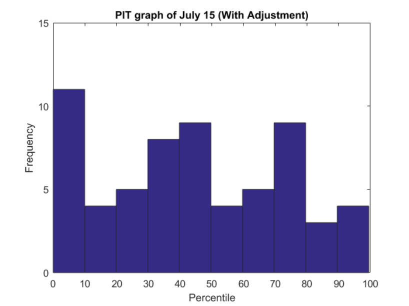


Optimization problem :

$$\min_{\delta V} \left| \int_0^T \left(V_{observed}(t) - V_{simulated}(t)\right) dt \right|, \quad T > 0$$

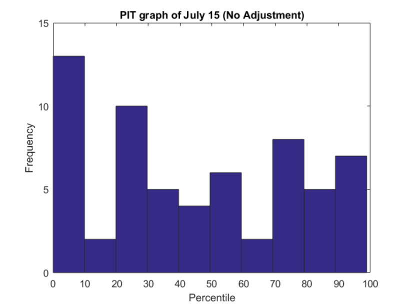


(a) Original Without Adjustment

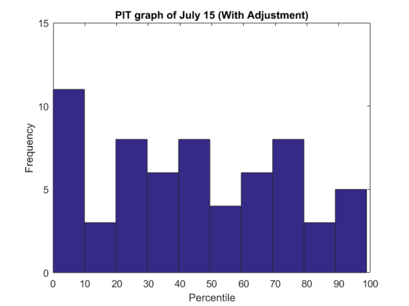


(b) Original With Adjustment

Summer optimization problem

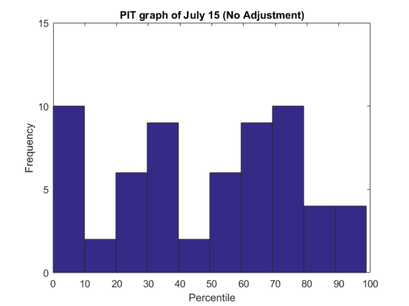


(c) Mixed Without Adjustment

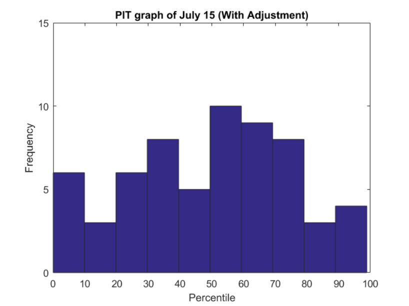


(d) Mixed With Adjustment

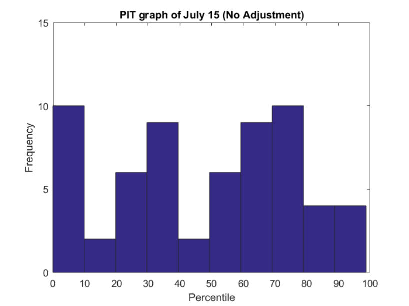
Summer optimization problem



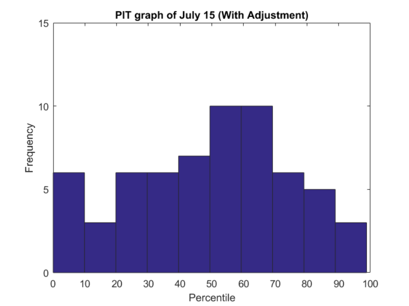
(e) Plus Without Adjustment



(f) Plus With Adjustment

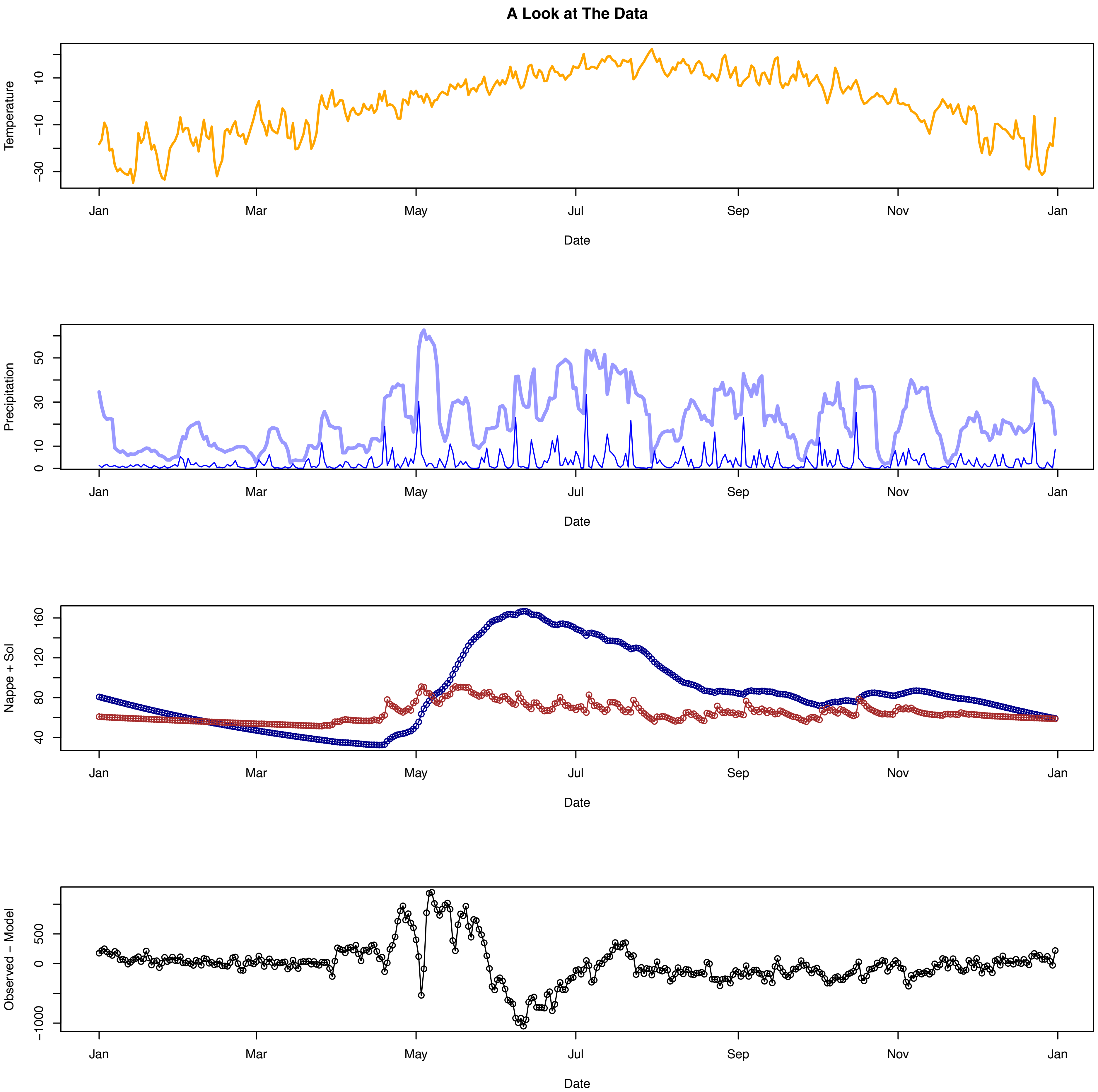


(g) Mixed Plus Without Adjustment



(h) Mixed Plus With Adjustment

3. Time-series


A Look at The Data
Temperature
10
−10
−30
Jan
Mar
May
Jul
Sep
Nov
Jan
Date
Precipitation
50
30
10
0
Nappe + Sol
160
120
80
40
Observed − Model
500
0
−1000

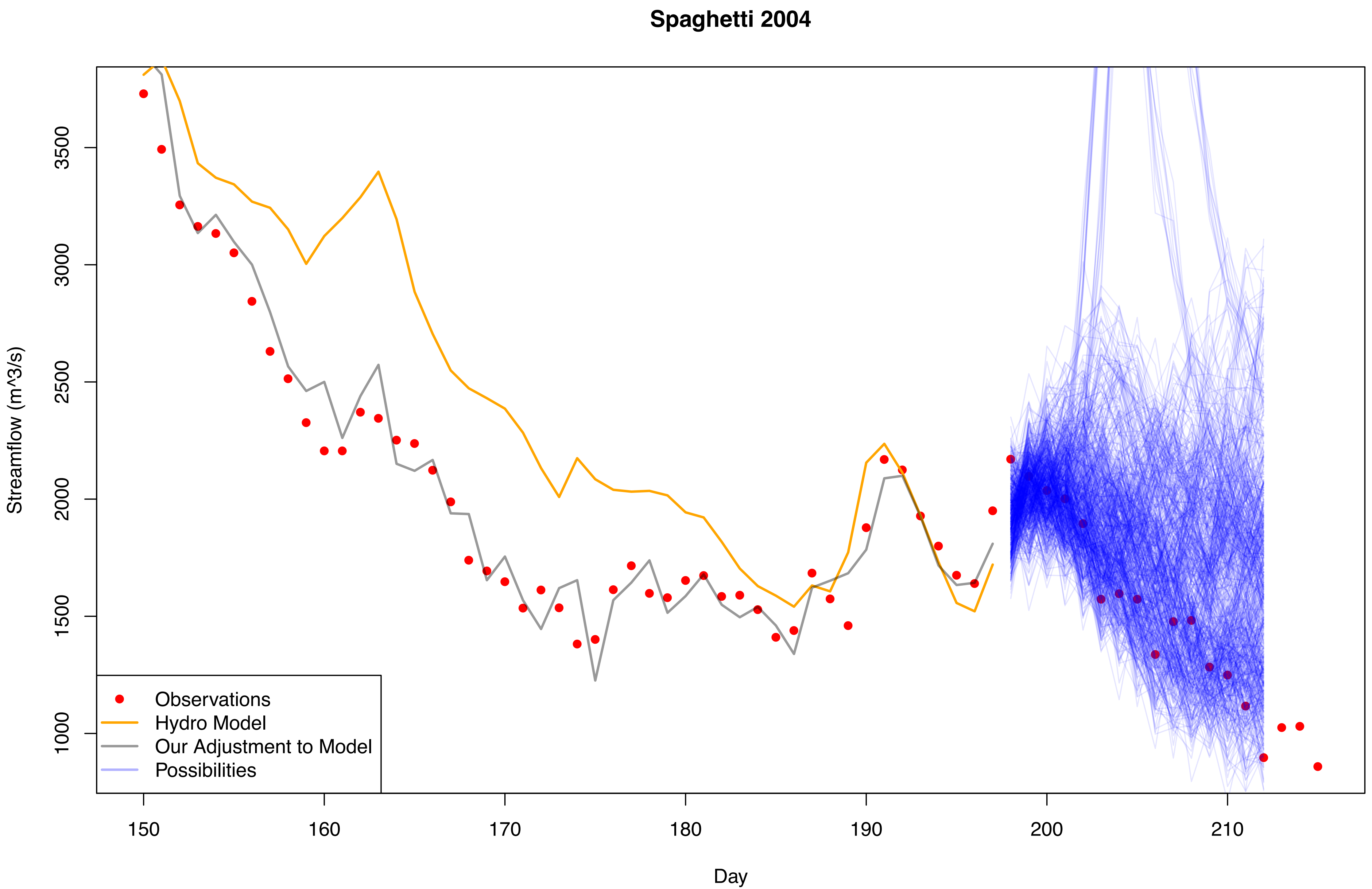
Time-series

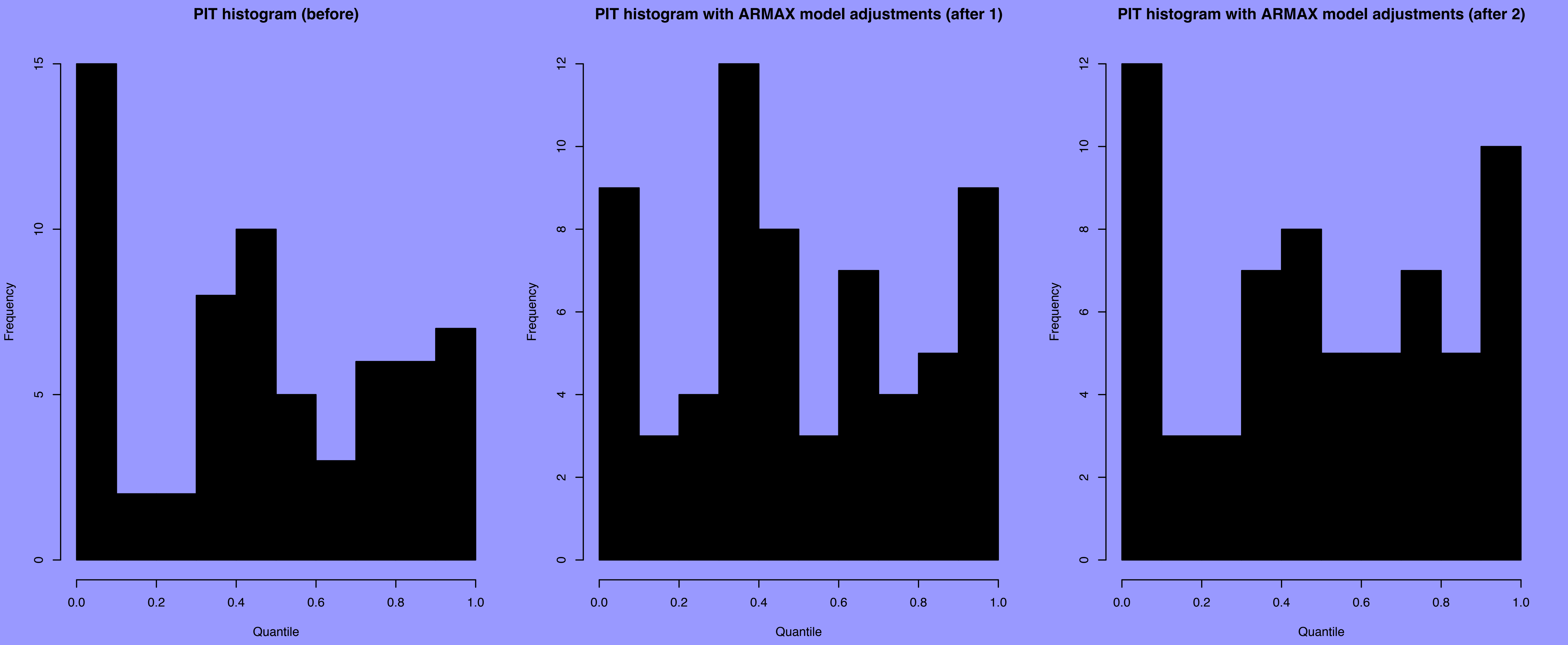
Daily data from 1954 to 2017. Let

$$
\begin{aligned}
Q_{t,sim} &= \text{streamflow at time } t\text{, based on the hydrological model.}\\
Q_{t,obs} &= \text{observed streamflow at time } t.\\
P_t &= \text{observed precipitations at time } t.\\
D_t &= Q_{t,obs} - Q_{t,sim}
\end{aligned}
$$

For a given year, data from January 1 to July 15. Calculated predictions for July 16 to July 30. For every year, we studied ARMAX type models for D_t (autoregressive moving average models with exogenous variables). We included as exogenous information P_t, and also P_{t-j}, $j = 0, \ldots, 7$ and sums of lag precipitations.
Predictions for D_t can be calculated.

To calculate prediction intervals, we considered a Monte Carlo approach.
For each year, we simulated values of $Q_{t,sim}$ have been used. We also used the historical values of precipitations.
Calculated so-called PIT histograms.
Three pictures: basis PIT graphic; one based only on precipitations, the other based on a more complex ARMAX model.


Spaghetti 2004
Streamflow (m^3/s)
Day
1000
1500
2000
2500
3000
3500
150
160
170
180
190
200
210
Observations
Hydro Model
Our Adjustment to Model
Possibilities



PIT histogram (before)
PIT histogram with ARMAX model adjustments (after 1)
PIT histogram with ARMAX model adjustments (after 2)
Frequency
Quantile
0
5
10
15
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0

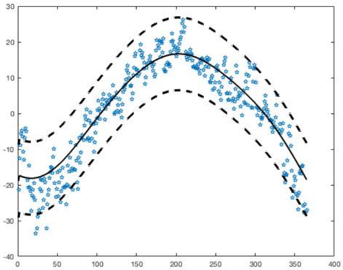
4. Gaussian model

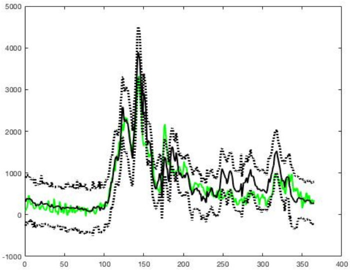
Statistical Approach: let the data talk.

- Observed inflow $y_{obs}(x)$ is a stochastic process depending on weather, precipitation, and uncertain state of the earth (i.e., underground water, soil water, snow).
- Simulated inflow $y_{sim}(x) = f(x)$ is a deterministic process (x is the current state of the earth).
- **Problem 1**: We don't know how $y_{obs}(x)$ looks like.
- **Problem 2**: Stochastic quantity is being modelled using something deterministic.
- **Solution**: Model what the deterministic process cannot capture using stochastic process. Approximate the observation process.

We have decades of data. Let's see what we can get by letting the data do the talking.

- Model the residual: $r(x) = y_{obs} - y_{sim}(x)$ using *Gaussian process*.
- Gaussian process is a distribution over continuous (and smooth) functions: $r(x) \sim GP(\mu(x), K(x, x'))$.
- $\mu(x) = \beta' x$; $K(x, x') = \sigma^2 \exp\left(\frac{\|x - x'\|^2}{2\ell^2}\right)$.
- Estimate β, σ^2 using the decades of data we have.





Final comments

Why statistical approach?

- Uncertainty is captured by the confidence interval.
- Example: We can say with 95% confidence that the true inflow will be between $(\hat{\mu}(x) - 2\hat{\sigma}, \hat{\mu}(x) + 2\hat{\sigma})$.

Simulating random states

- Simulate temperature using the time dependent variance. Use precipitation forecast. Denote by I.
- $x_{t+1}^n = M(x_t^n, I_{t+1}^n)$.
- $\hat{y}(x_{t+1}^n) = f(x_{t+1}^n) + r(x_{t+1}^n)$.
- Prediction: $\bar{y}_{t+1} = \sum_{n=1}^{N} \hat{y}(x_{t+1}^n)$. Compute empirical variance.

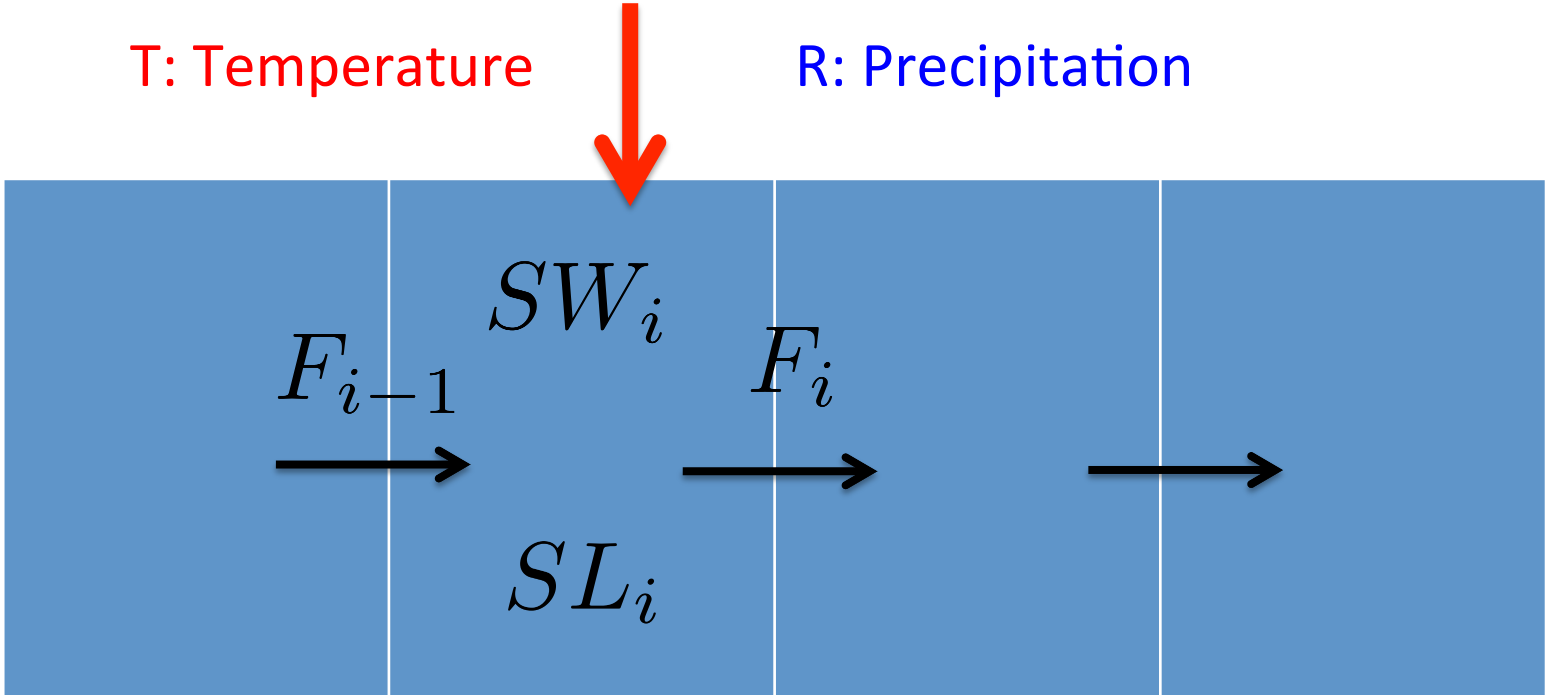
5. Filtering methods

One dimensional model of a hydrological system

IDEA: Produce a simple, but realistic, model, which will allow us to test various data assimilation methods

T: Temperature

R: Precipitation



SW : Snow SL: Water in soil F: Free water

T_i^n, R_i^n Measured meteorological data at time n in cell i

Free water is a function of the water in the soil

$$F_i^n = f(SL_i^n)$$

Physics:

Positive T: Snow melts as temperature rises and adds to SL, precipitation falls as rain and adds to soil water

Negative T: Precipitation falls as snow and adds to SW

Mathematical model

$$T_i^n > 0$$

$$SW_i^{n+1} = SW_i^n - \alpha T\, SW_i^n$$

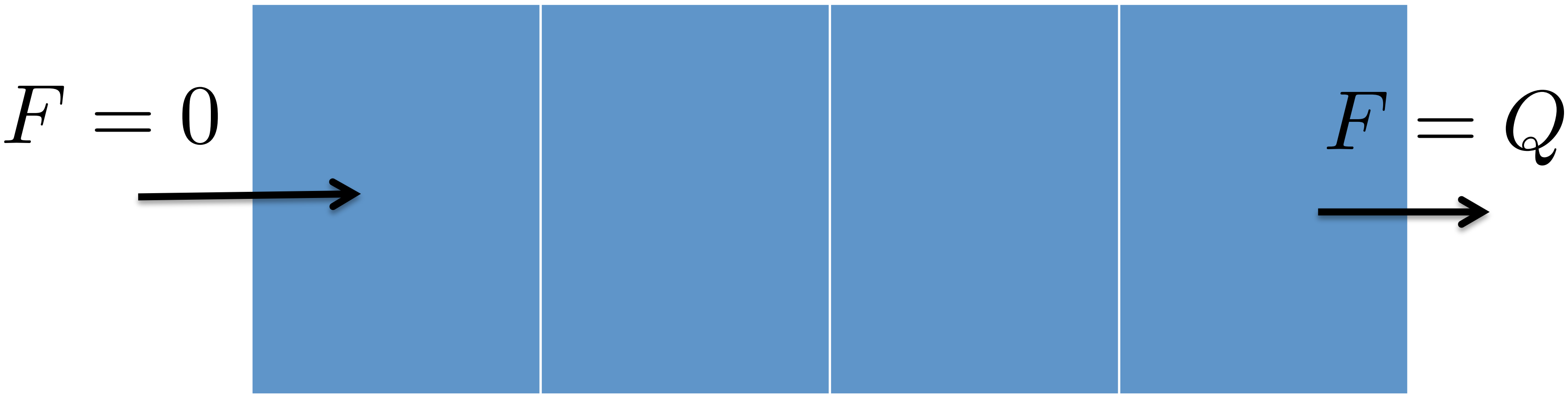
$$SL_i^{n+1} = SL_i^n + \alpha T\, SW_i^n + \beta R_i^n + F_{i-1}^n - F_i^n$$

$$T_i^n \leq 0$$

$$SW_i^{n+1} = SW_i^n + \gamma R_i^n$$

$$SL_i^{n+1} = SL_i^n + F_{i-1}^n - F_i^n$$

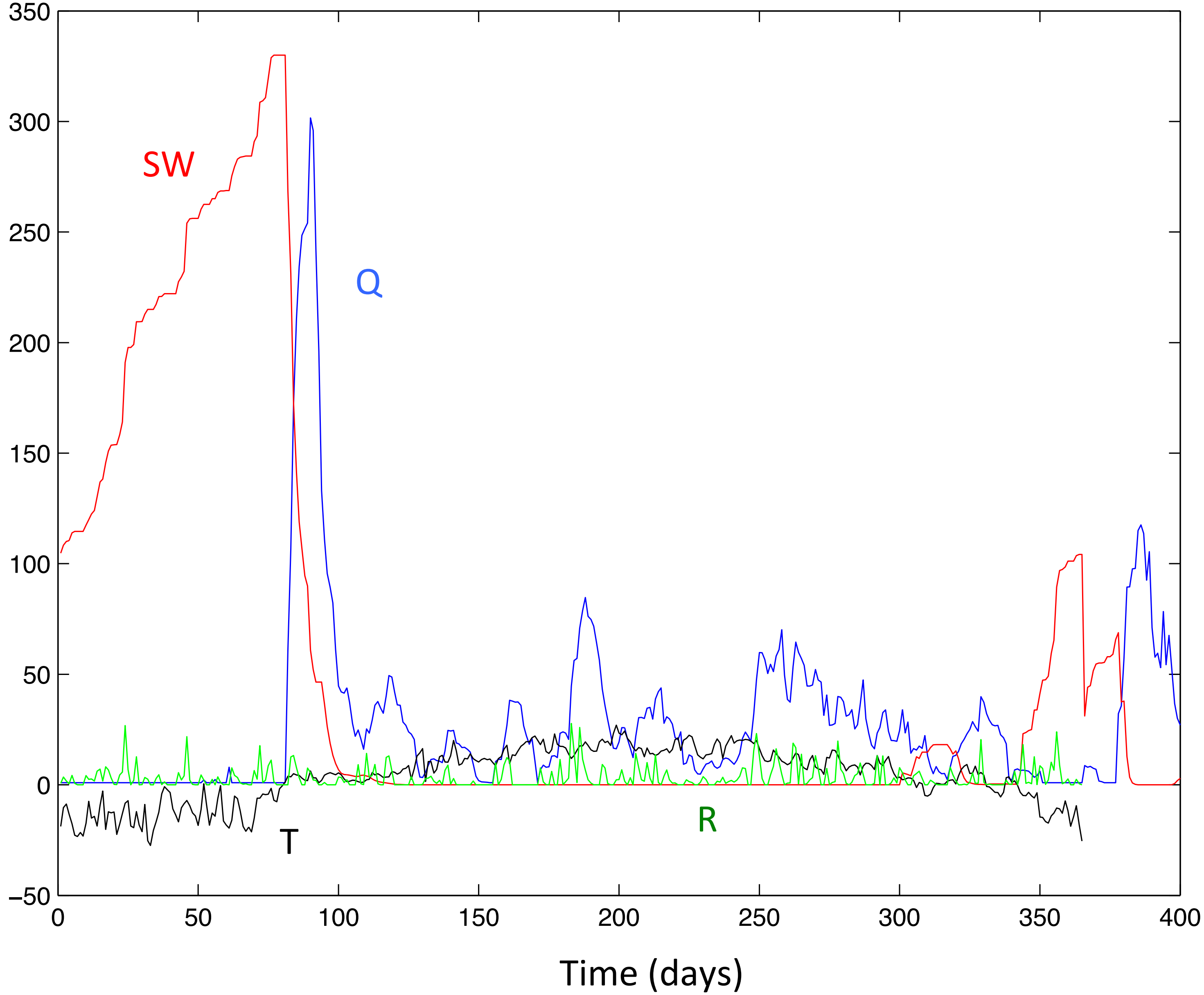
Boundary and Other Conditions

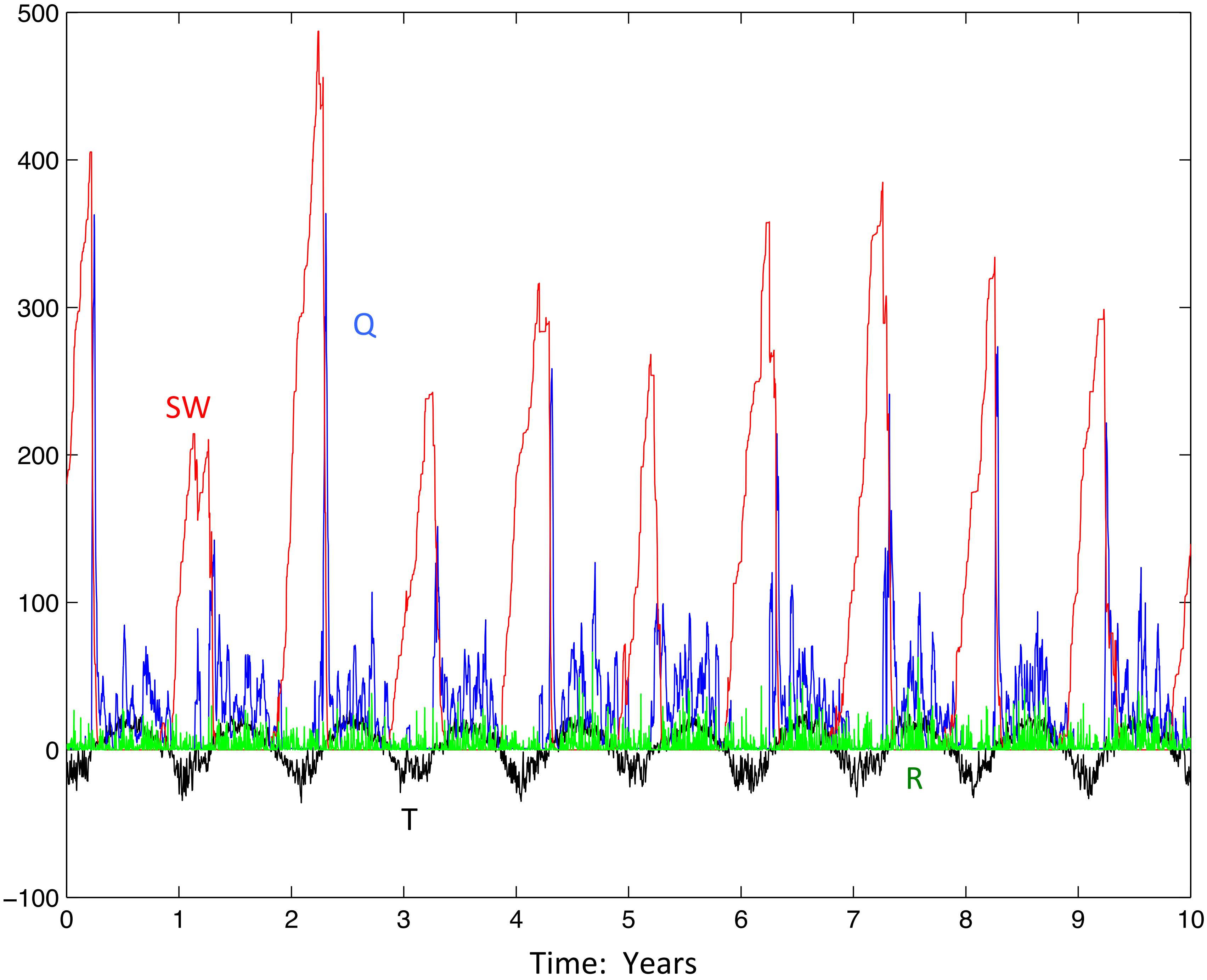


Q: Measured stream flow

Take: 10 cells, $\alpha = 1/15, \beta = \gamma = 1$, f(S) = S/4

Results of simulations





500
400
300
200
100
0
−100
0
1
2
3
4
5
6
7
8
9
10
Time: Years
SW
Q
T
R

Data assimilation

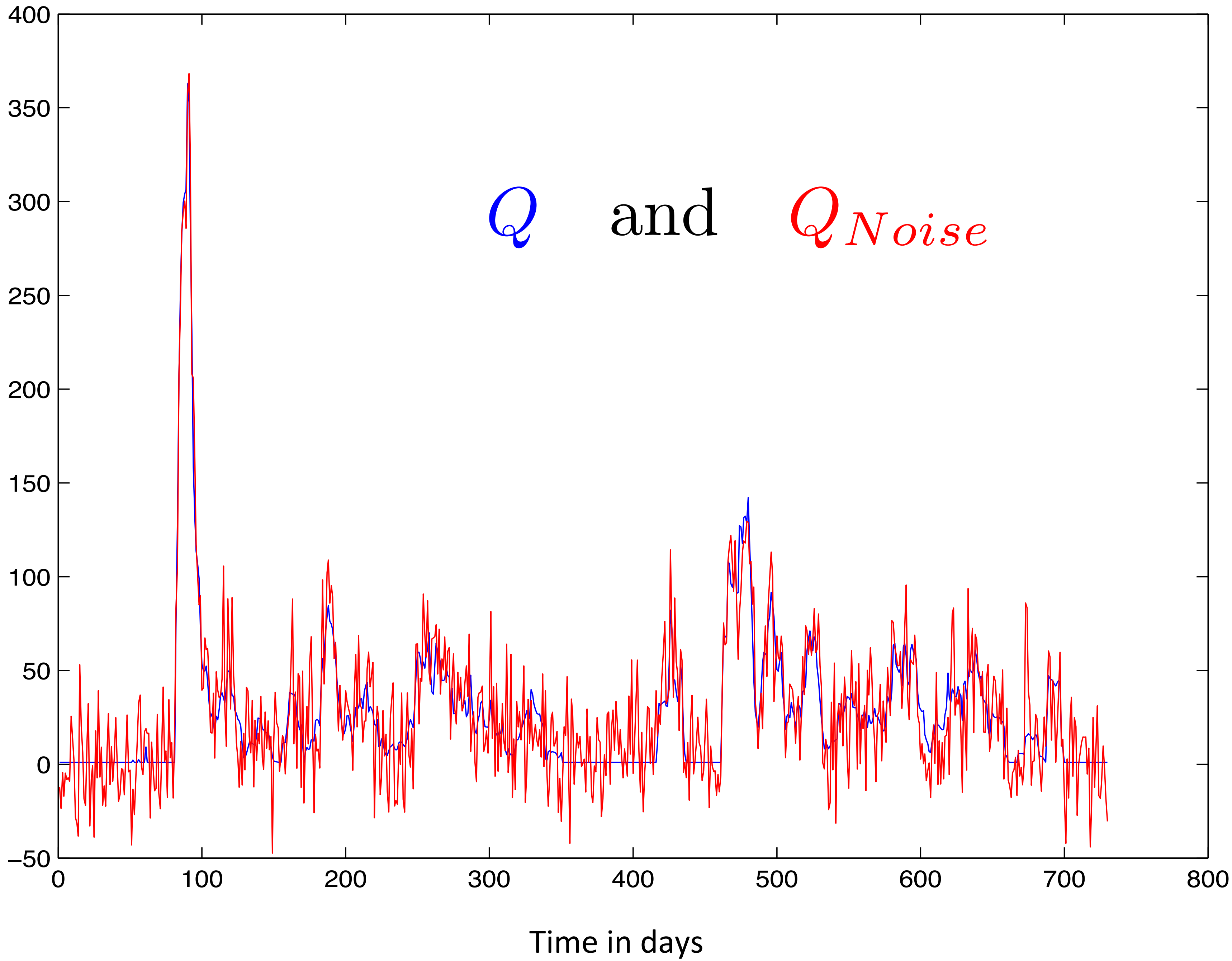
Do the following test:

- Take initial value of SW
- Generate two years values of the stream flux Q
- Add noise to Q to give Q_{Noise}
- Take shift $SW \to SW + \delta\, SW$
- Generate new stream flux Q_δ
- Find $\delta\, SW$ minimising the error between Q_{Noise} and Q_δ

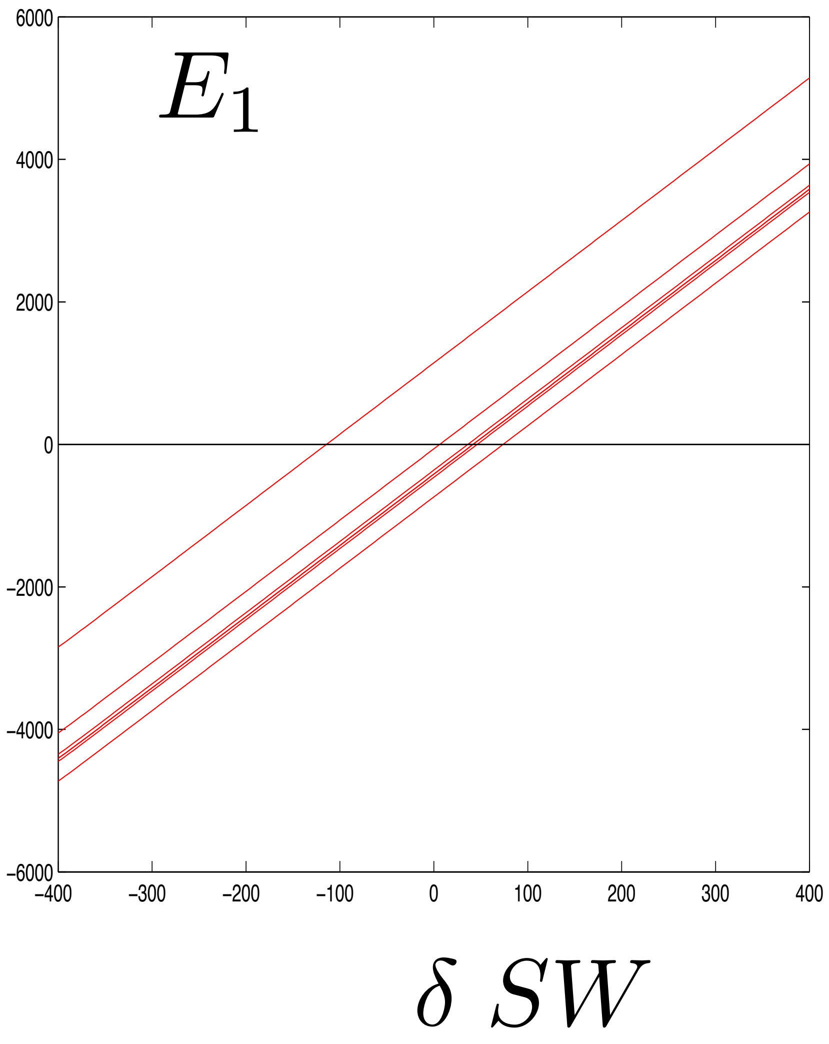
Error measures

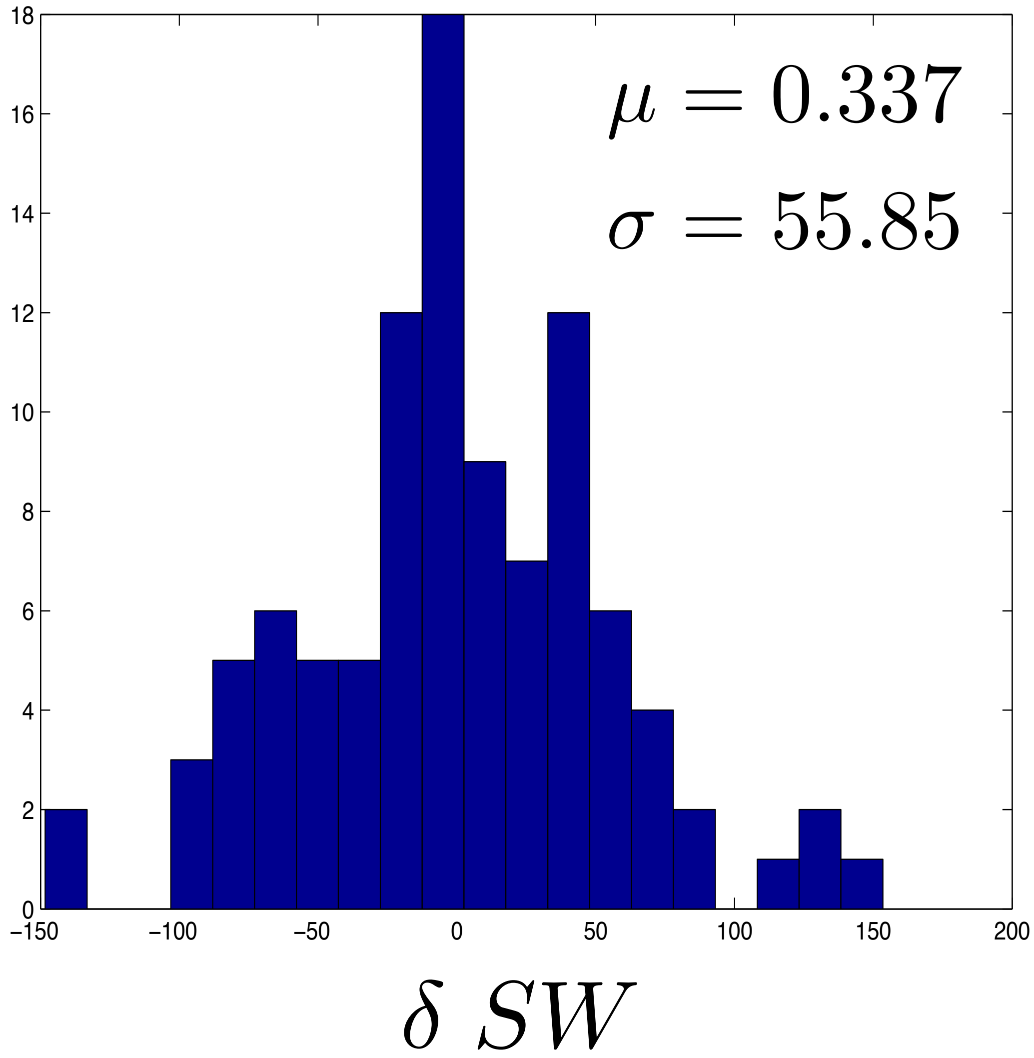
$$E_1 = \int \left(Q_{Noise} - Q_\delta\right) \, dt$$

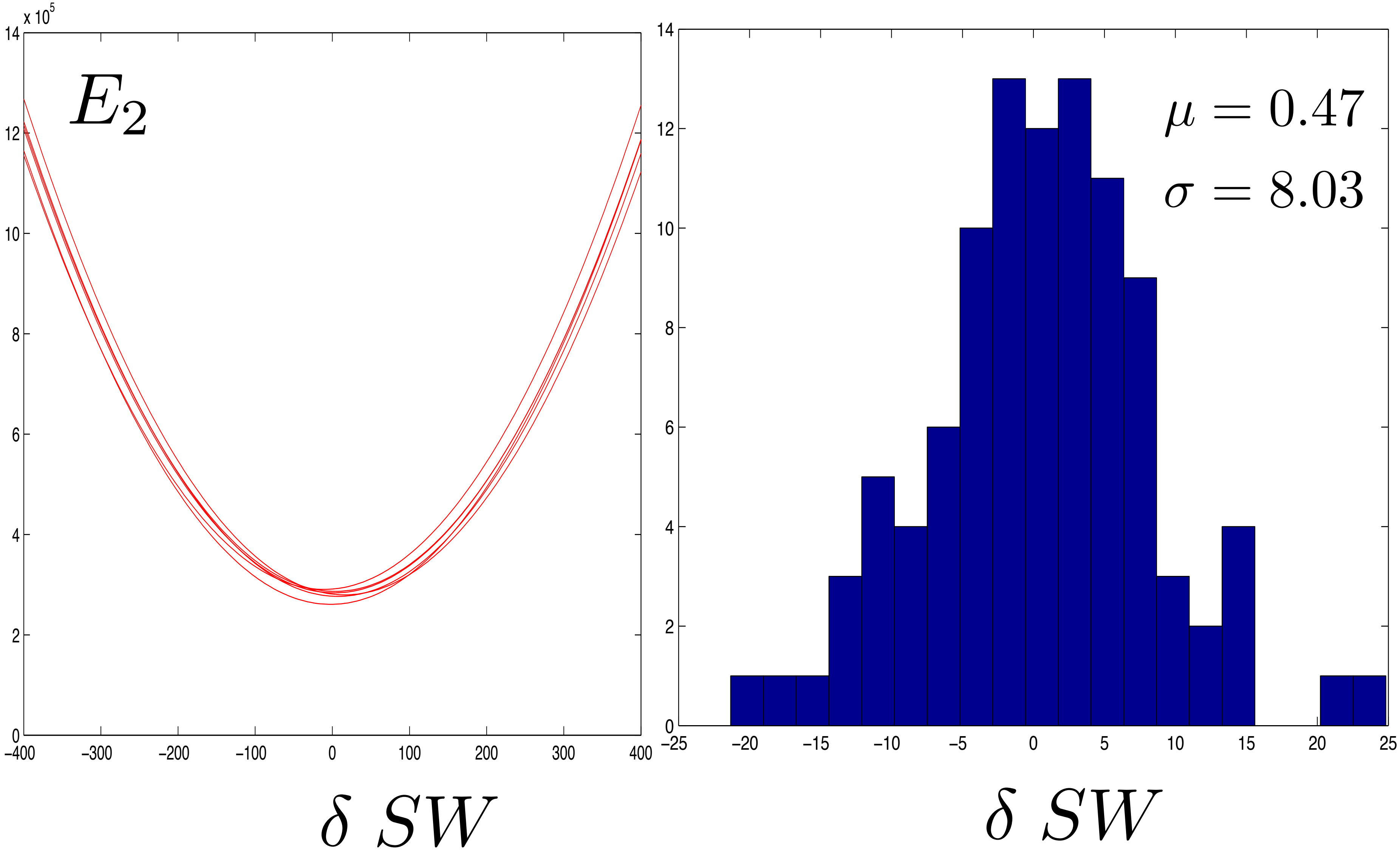
$$E_2 = \int \left(Q_{Noise} - Q_\delta\right)^2 \, dt$$


Q and Q_{Noise}
400
350
300
250
200
150
100
50
0
−50
0
100
200
300
400
500
600
700
800
Time in days

Do a number of noisy realisations






E_2
x 10^5
$\mu = 0.47$
$\sigma = 8.03$
$\delta\ SW$
$\delta\ SW$

Conclusion

Conclusion and Future work:

- **Several approaches** were tested.
- **Improvement** was **measured** for the problem.
- Some methods are still in **prototype phase** but are **promising**.

Thanks for your attention !