Stochastic disaggregation of time series
an example in hydrology

Paul Torfs

Wageningen University
Paul.Torfs@wur.nl

Stochastic disaggregation . . .

- there are many scale problems in hydrology

Stochastic disaggregation . . .

- there are many scale problems in hydrology
- for some of those cases, proposition to defend:

sometimes it is better to
adapt the data to the model
then the model to the data

Stochastic disaggregation . . .

- there are many scale problems in hydrology
- for some of those cases, proposition to defend:

sometimes it is better to
adapt the data to the model
then the model to the data

- defend this with one man-made example

Outline

- the example

Outline

- the example (and this should be **very** representative)

Outline

- the example (and this should be **very** representative)
- the scale problem

Outline

- the example (and this should be **very** representative)
- the scale problem
- the “usual solutions” (adapt the model)

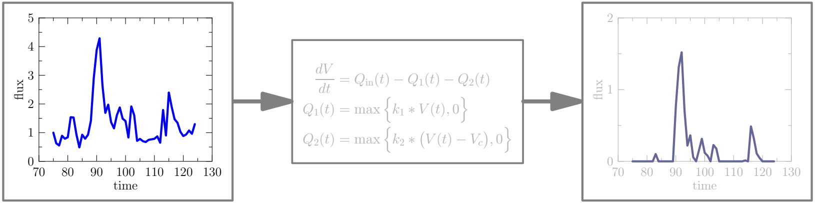
Outline

- the example (and this should be **very** representative)
- the scale problem
- the "usual solutions" (adapt the model)
- the "other solution": adapt the data

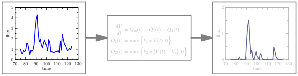
Outline

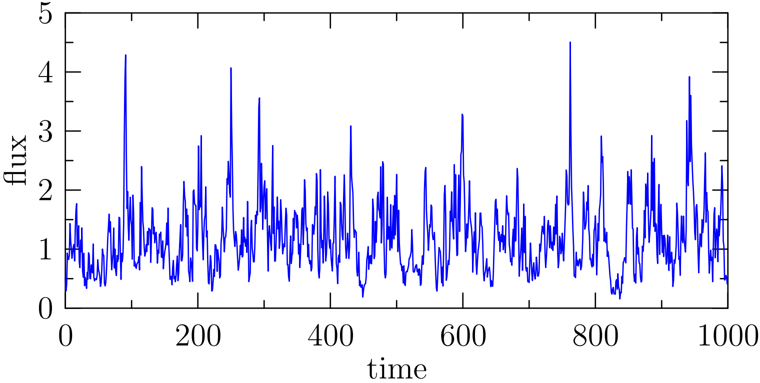
- the example (and this should be **very** representative)
- the scale problem
- the "usual solutions" (adapt the model)
- the "other solution": adapt the data
- technique of stochastic disaggregation

The example: input

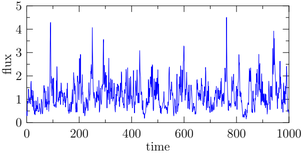


The example: input = time series



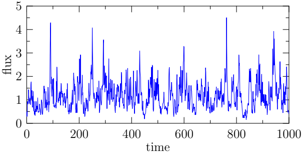


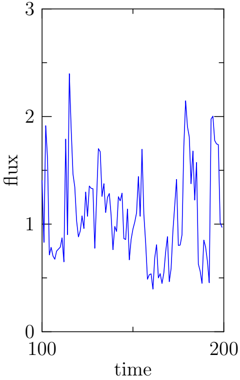
The example: input



- not continuous

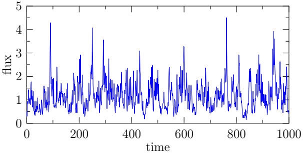
The example: input

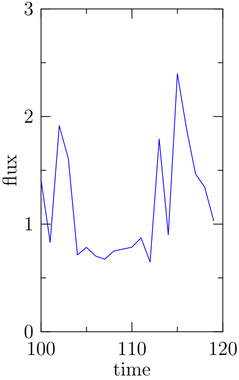




- not continuous but discrete

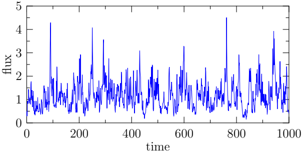
The example: input

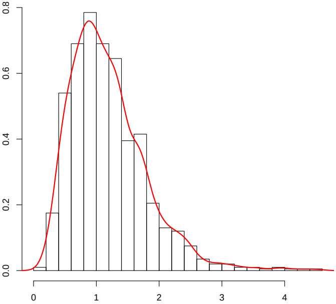




- not continuous but discrete
  resolution = finest scale

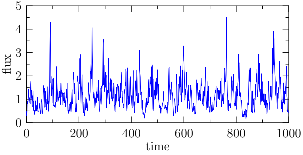
The example: input

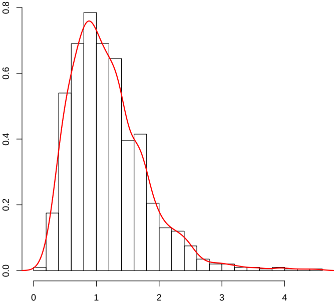




- not continuous but discrete resolution = finest scale
- non-gaussian

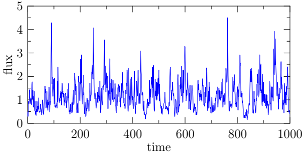
The example: input

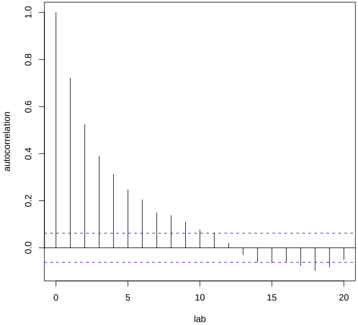




- not continuous but discrete resolution = finest scale
- non-gaussian
- positive

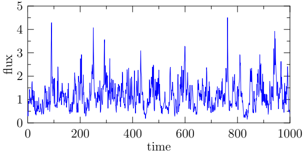
The example: input

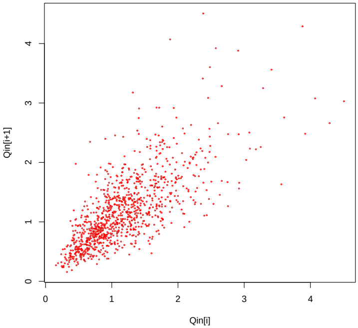




- not continuous but discrete resolution = finest scale
- non-gaussian
- positive
- dependency

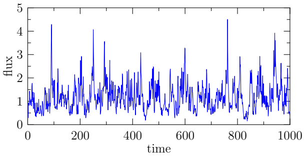
The example: input





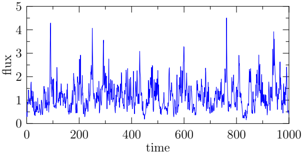
- not continuous but discrete
  resolution = finest scale
- non-gaussian
- positive
- dependency
  non-linear

The example: input



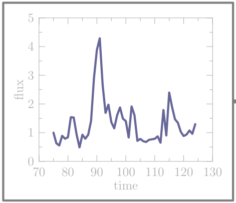
- not continuous but discrete
  resolution = finest scale
- non-gaussian
- positive
- dependency
  non-linear
- periodicity

The example: input

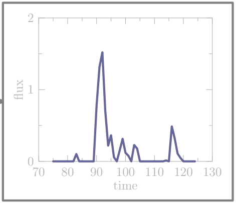


- not continuous but discrete
  resolution = finest scale
- non-gaussian
- positive
- dependency
  non-linear
- periodicity
- zero-ness

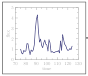
The example: the model



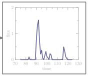
$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{k_1 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{k_2 * \left(V(t) - V_c\right), 0\right\}$$



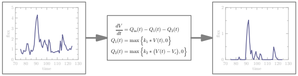
The example: the model

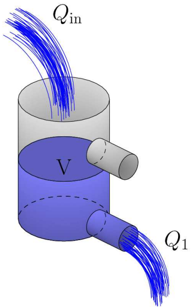


$$\frac{dV}{dt} = Q_{\mathrm{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{k_1 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{k_2 * \left(V(t) - V_c\right), 0\right\}$$



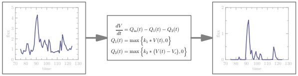
The example: the model

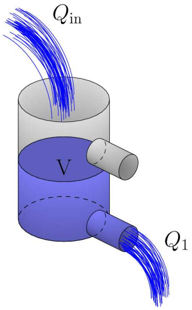


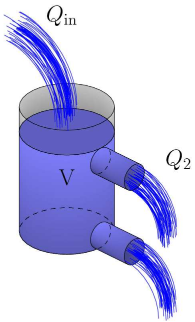


$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

The example: the model



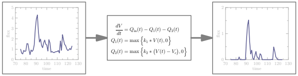


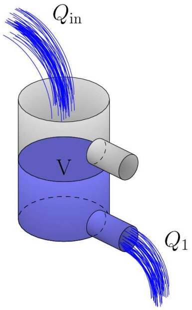


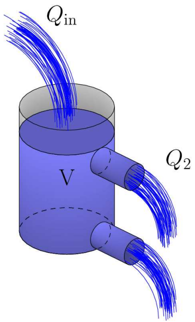
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * \left(V(n) - V_c\right), 0\right\}$$

The example: the model



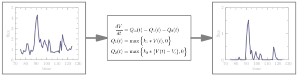


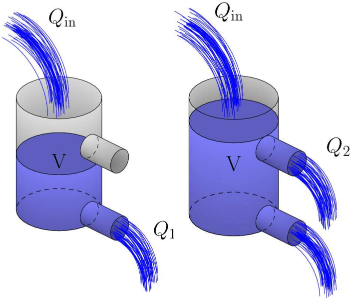


$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\} \qquad Q_2(n) = \max\left\{k_2 * (V(n) - V_c), 0\right\}$$

$$V(n+1) = V(n) + Q_{\mathrm{in}}(n) - Q_1(n) - Q_2(n)$$

The example: the model



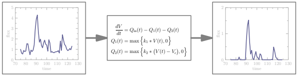


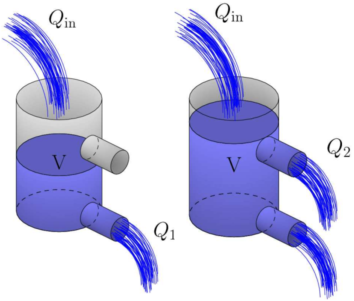
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * \left(V(n) - V_c\right), 0\right\}$$

$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

The example: the model





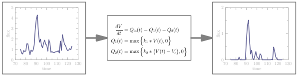
- based on mass balance

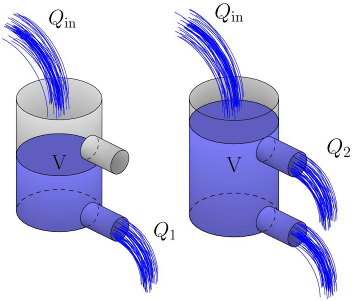
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * \left(V(n) - V_c\right), 0\right\}$$

$$V(n+1) = V(n) + Q_{in}(n) - Q_1(n) - Q_2(n)$$

The example: the model





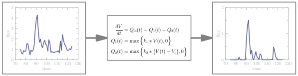
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

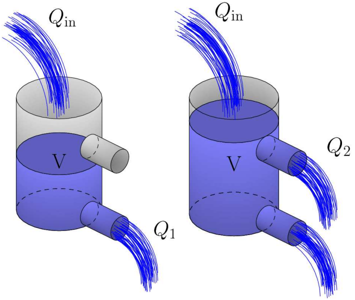
$$Q_2(n) = \max\left\{k_2 * (V(n) - V_c), 0\right\}$$

$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

- based on mass balance
- storage generates memory (=time dependency)

The example: the model





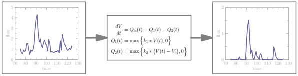
- based on mass balance
- storage generates memory (=time dependency)
- non-linear

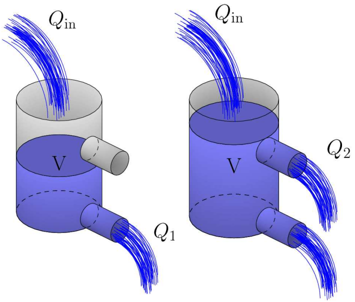
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * \left(V(n) - V_c\right), 0\right\}$$

$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

The example: the model

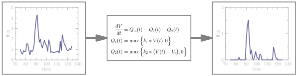


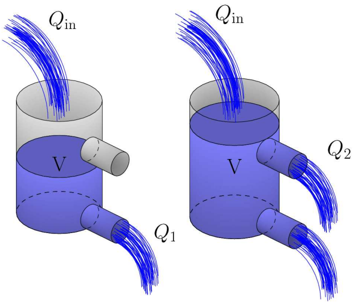


$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$
$$Q_2(n) = \max\left\{k_2 * (V(n) - V_c), 0\right\}$$
$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

- based on mass balance
- storage generates memory (=time dependency)
- non-linear
- well known structure, no model error (lab)

The example: the model





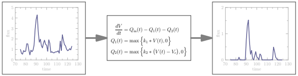
- based on mass balance
- storage generates memory (=time dependency)
- non-linear
- well known structure, no model error (lab)
- determined by parameters

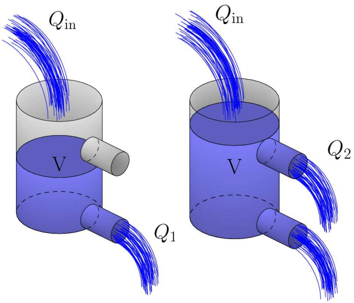
$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * (V(n) - V_c), 0\right\}$$

$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

The example: the model





- based on mass balance
- storage generates memory (=time dependency)
- non-linear
- well known structure, no model error (lab)
- determined by parameters (assumed known):

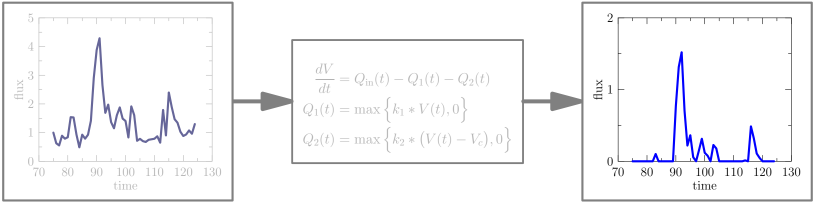
$$k_1 = 0.5 \quad k_2 = 0.5 \quad V_c = 2.5$$

$$Q_1(n) = \max\left\{k_1 * V(n), 0\right\}$$

$$Q_2(n) = \max\left\{k_2 * (V(n) - V_c), 0\right\}$$

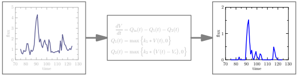
$$V(n+1) = V(n) + Q_{\text{in}}(n) - Q_1(n) - Q_2(n)$$

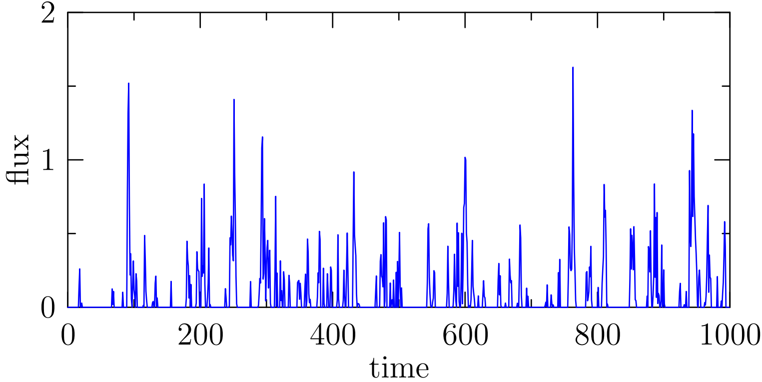
The example: output



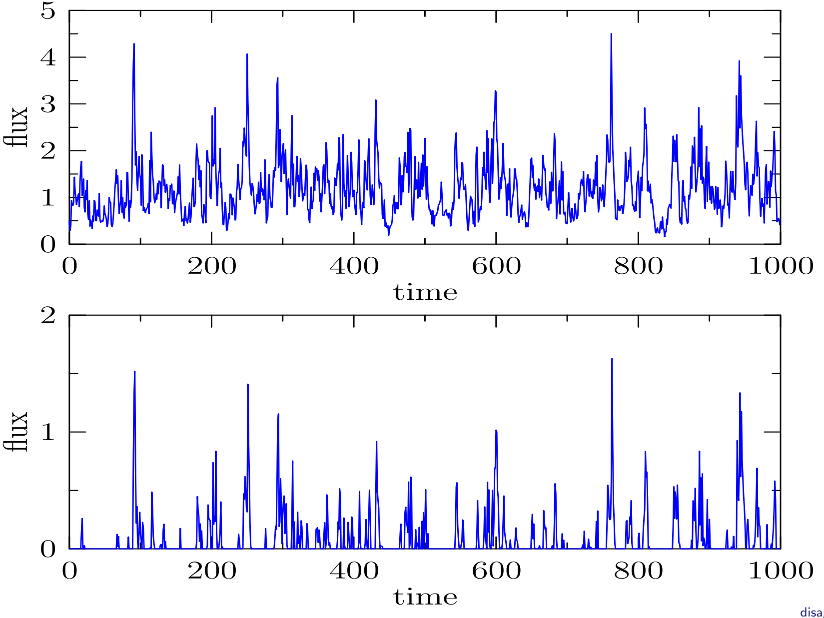
only Q_2 (top outlet) is observed

The example: output = time series

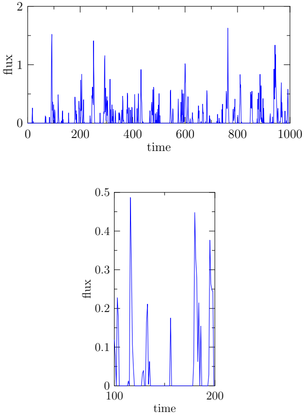




The example: input-output

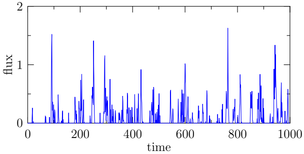


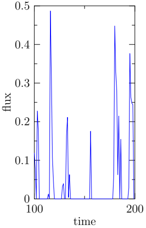
The example: output



- complex combination model+input

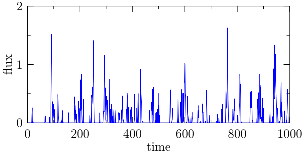
The example: output

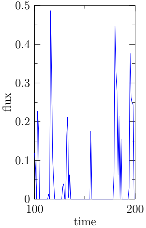




- complex combination model+input
- only partial information on the process is observed

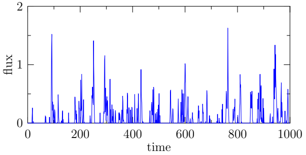
The example: output

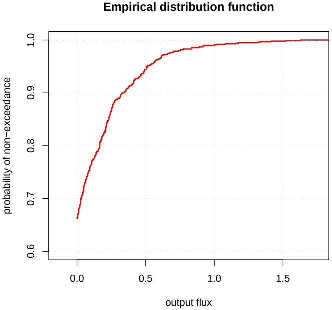




- complex combination model+input
- only partial information on the process is observed
- contains enough (hidden) information to identify parameters
  any change in k_1, k_2 or V_c results in different output

The example: output





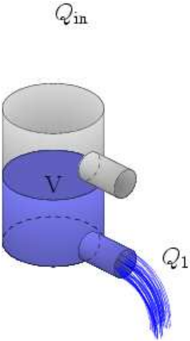
- complex combination model+input
- only partial information on the process is observed
- contains enough (hidden) information to identify parameters
  any change in k_1, k_2 or V_c results in different output
- in many cases: more interest in the statistics then in the exact timing

The problem

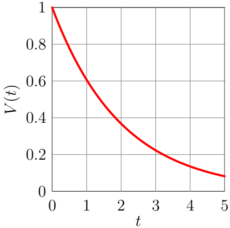
sometimes
resolution of available input
is to coarse compared to the
resolution needed by model

Resolution of the model=1

$$k_1 = 0.5 \Rightarrow \text{time scale} = 2$$



timescale = 2



$$Q_1(t) = 0.5\, V(t) = \frac{1}{2}\, V(t)$$

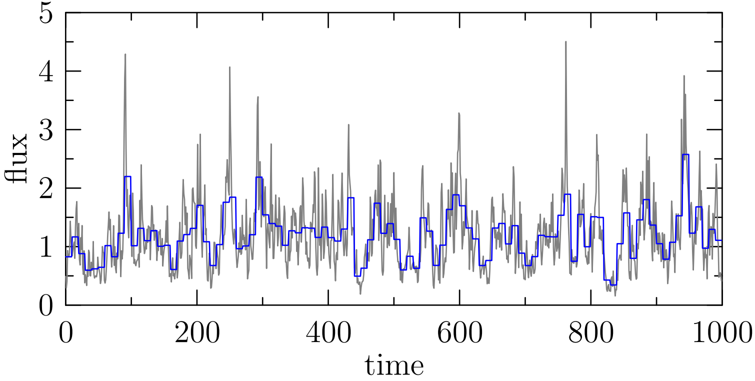
$$V(t) = V(0)\, e^{-\frac{t}{2}}$$

Resolution of the input=10

in many cases:

input data only available aggregated on lower resolution

in our example: aggregated at level=10

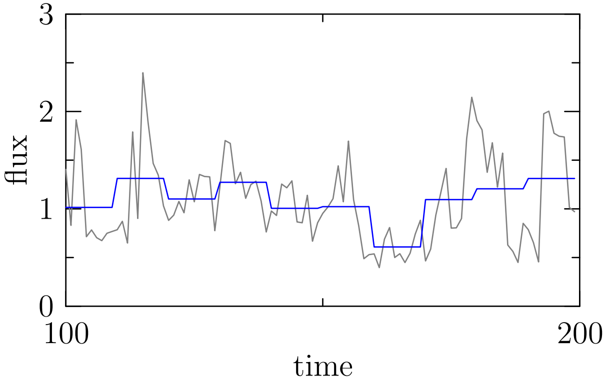


Resolution of the input=10

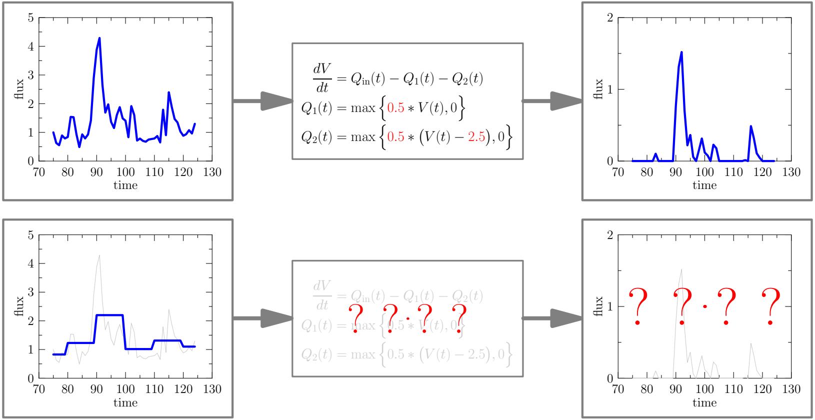
in many cases:

input data only available aggregated on lower resolution

in our example: aggregated at level=10

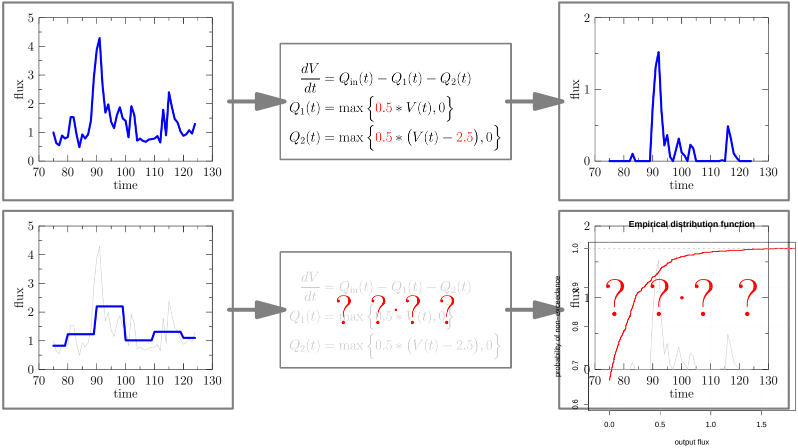


The problem

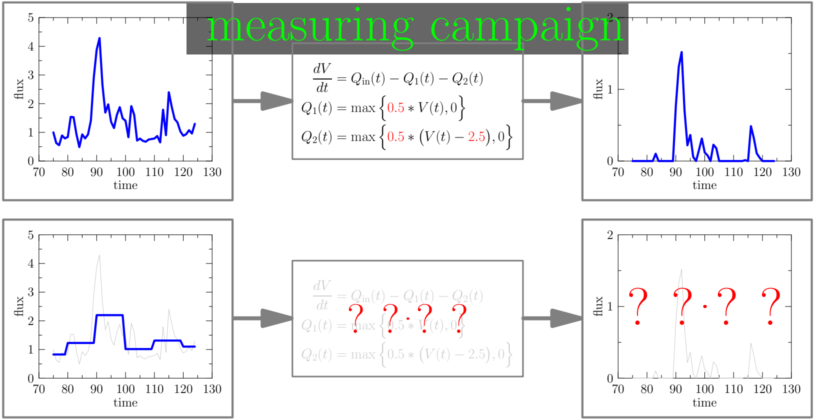


flux
time
$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$
$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$
$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$
? ? ? ?
? ? ? ?

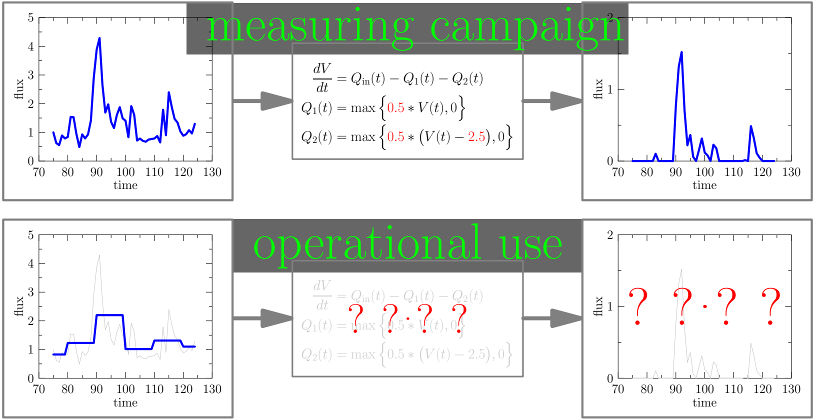
The problem


flux
time
$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$
$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$
$Q_2(t) = \max\left\{0.5 * \left(V(t) - 2.5\right), 0\right\}$
? ? · ? ?
Empirical distribution function
probability of non−exceedance
output flux

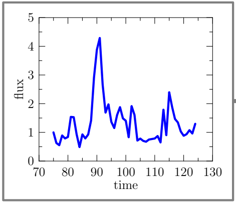
The problem


measuring campaign
flux
time
$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$
$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$
$Q_2(t) = \max\left\{0.5 * \left(V(t) - 2.5\right), 0\right\}$
? ? ? ?
? ? ? ?

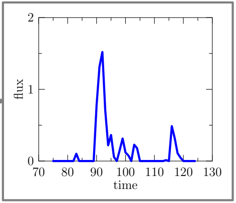
The problem

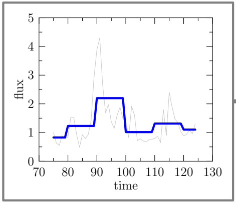

measuring campaign
operational use
flux
time
dV/dt = Q_in(t) − Q_1(t) − Q_2(t)
Q_1(t) = max {0.5 ∗ V(t), 0}
Q_2(t) = max {0.5 ∗ (V(t) − 2.5), 0}
?? · ??
?? · ??

Solution 1: just use the old model

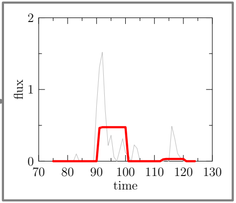


$$
\begin{aligned}
\frac{dV}{dt} &= Q_{\text{in}}(t) - Q_1(t) - Q_2(t) \\
Q_1(t) &= \max\left\{0.5 * V(t), 0\right\} \\
Q_2(t) &= \max\left\{0.5 * \left(V(t) - 2.5\right), 0\right\}
\end{aligned}
$$

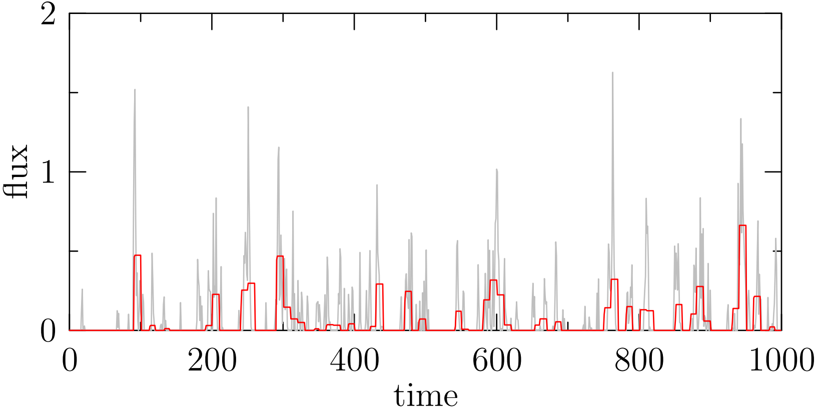




$$
\begin{aligned}
\frac{dV}{dt} &= Q_{\text{in}}(t) - Q_1(t) - Q_2(t) \\
Q_1(t) &= \max\left\{0.5 * V(t), 0\right\} \\
Q_2(t) &= \max\left\{0.5 * \left(V(t) - 2.5\right), 0\right\}
\end{aligned}
$$



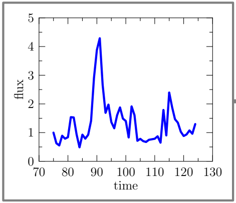
Solution 1: just use the old model



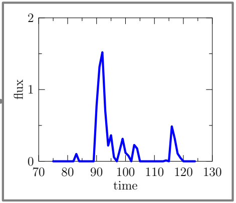
bad result due to non-linearity of the model

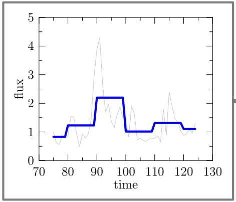
Solution 2: use the structure of old model

but with new parameters

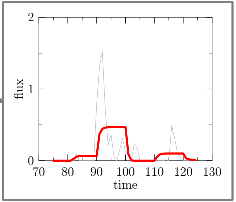


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$



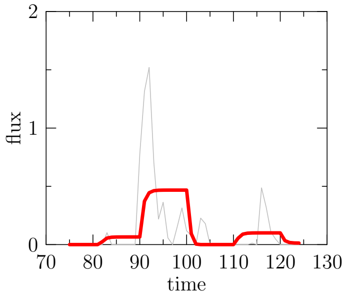


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.45 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.31 * \big(V(t) - 2.4\big), 0\right\}$$



Solution 2: use the structure of old model

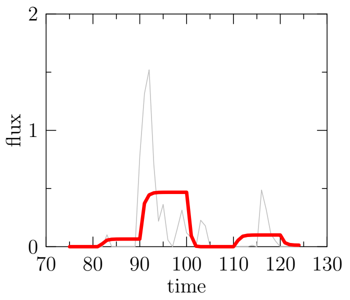
$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.45 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.31 * \left(V(t) - 2.4\right), 0\right\}$$



- parameters "fitted" (assumes some high resolution data available)

Solution 2: use the structure of old model

$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.45 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.31 * \left(V(t) - 2.4\right), 0\right\}$$

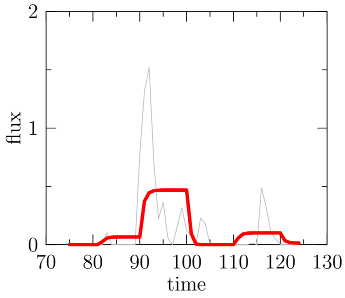


- parameters "fitted" (assumes some high resolution data available)
- new parameters are called effective

par	true	effective
k_1	0.5	0.45
k_2	0.5	0.31
V_c	2.5	2.4

Solution 2: use the structure of old model

$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.45 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.31 * \left(V(t) - 2.4\right), 0\right\}$$



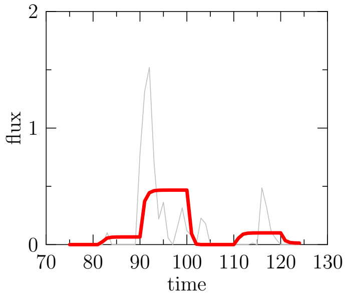
- parameters "fitted" (assumes some high resolution data available)
- new parameters are called effective

par	true	effective
k_1	0.5	0.45
k_2	0.5	0.31
V_c	2.5	2.4

- effective parameters = model + statistics input/output

Solution 2: use the structure of old model

$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.45 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.31 * \left(V(t) - 2.4\right), 0\right\}$$

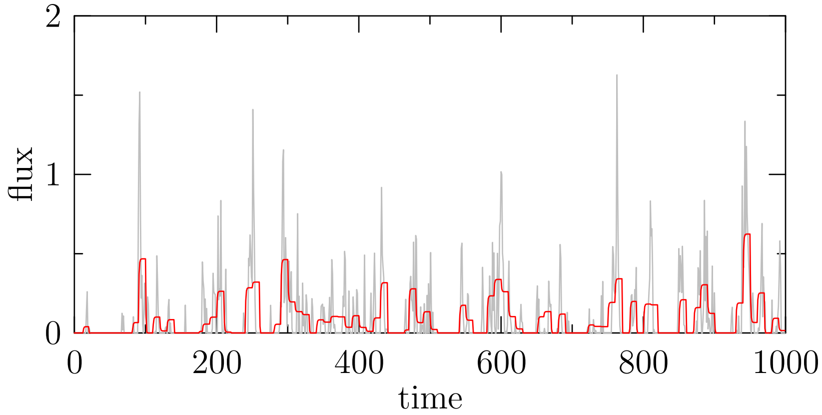


- parameters "fitted" (assumes some high resolution data available)
- new parameters are called effective

par	true	effective
k_1	0.5	0.45
k_2	0.5	0.31
V_c	2.5	2.4

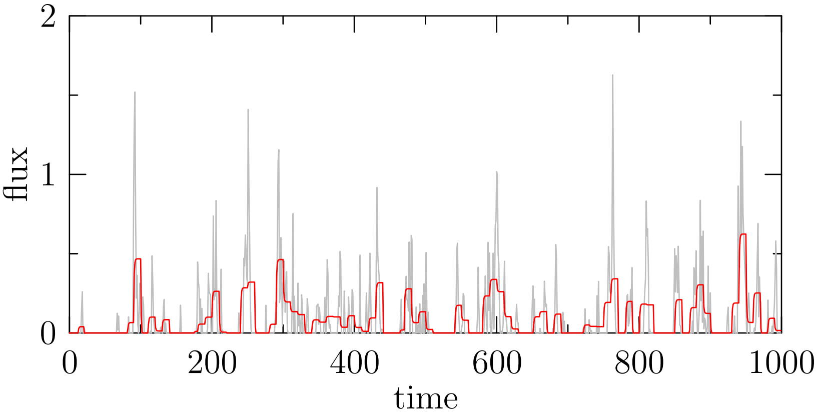
- effective parameters = model + statistics input/output
- can not be transfered to other cases

Solution 2: use the structure of old model



bad result due to non-linearity of the model

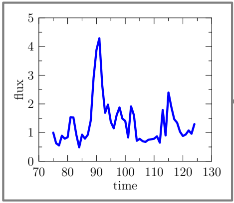
Solution 2: use the structure of old model



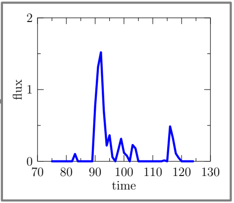
bad result due to non-linearity of the model
but best possible with given structure

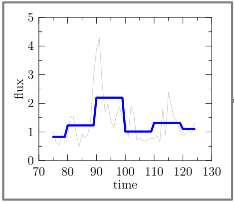
Solution 3: use new type of model

e.g.: linear regression = black-box model

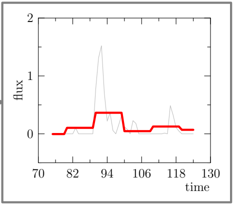


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \left(V(t) - 2.5\right), 0\right\}$$

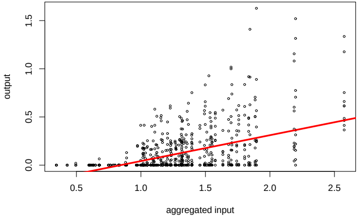


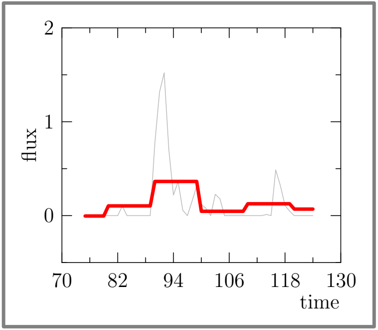


$$Q_2 = -0.23 + 0.27\, Q_{\text{in,low-res}}$$



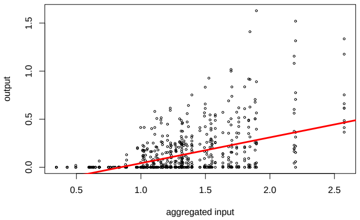
Solution 3: use new type of model

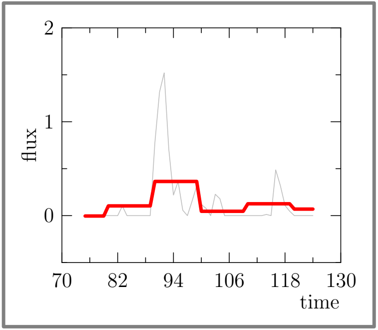




- needs to be "fitted" (assumes some high resolution data available)

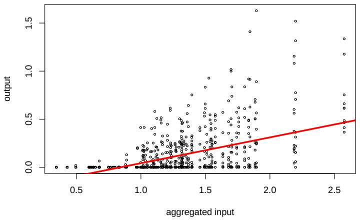
Solution 3: use new type of model

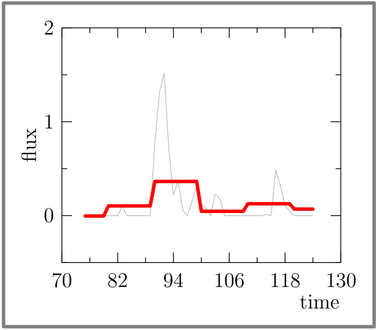




- needs to be "fitted" (assumes some high resolution data available)
- knowledge on structure not used "physics" $\rightarrow$ correlation

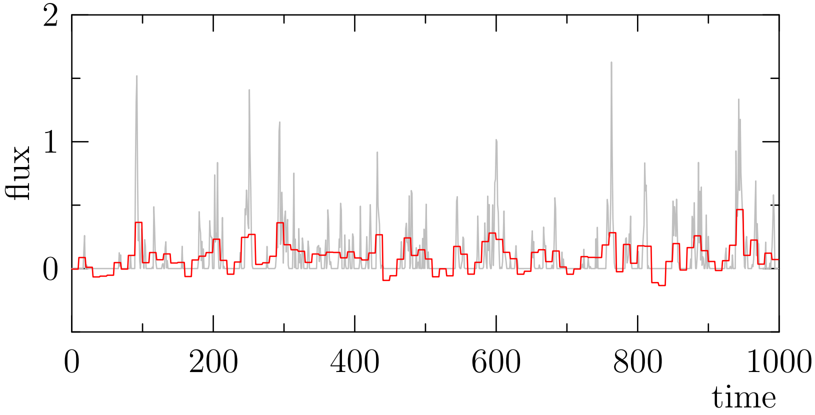
Solution 3: use new type of model





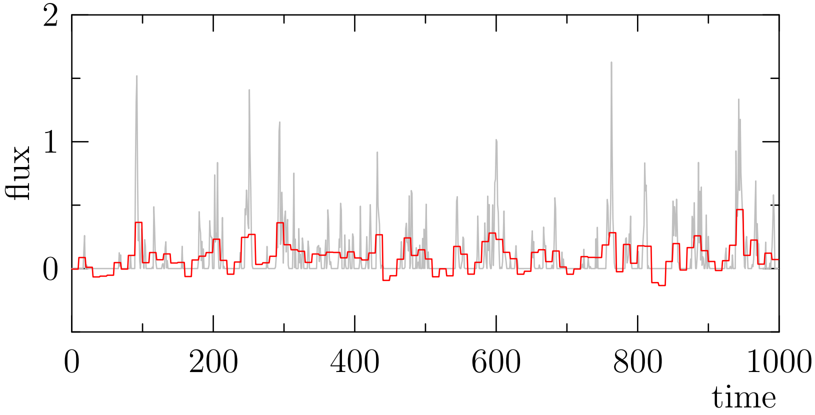
- needs to be "fitted" (assumes some high resolution data available)
- knowledge on structure not used "physics" $\rightarrow$ correlation
- can not be transfered to other cases

Solution 3: use new type of model



bad results (although not in least squares sense)

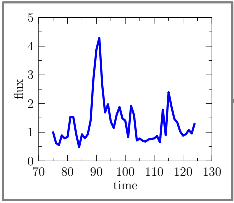
Solution 3: use new type of model



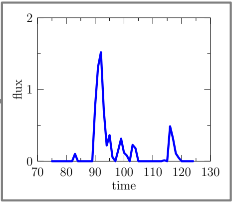
bad results (although not in least squares sense)
even non-physical results (negative values) due to black-box

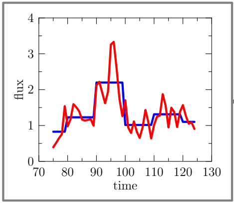
Solution 4: use new data with old model

new data = simulated high resolution data with given low resolution means

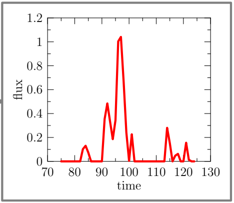


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$



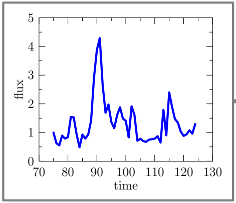


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$

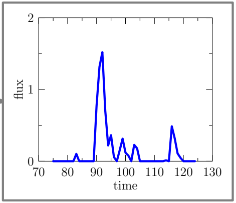


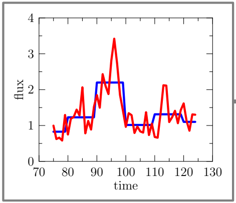
Solution 4: use new data with old model

new data = simulated high resolution data with given low resolution means

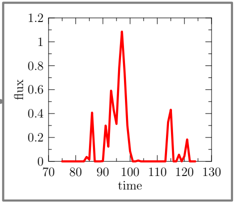


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$



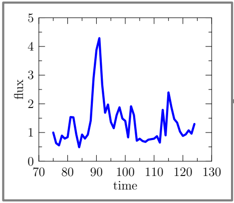


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$

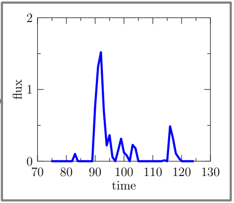


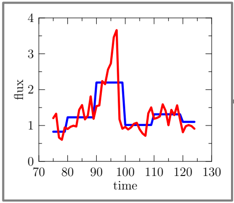
Solution 4: use new data with old model

new data = simulated high resolution data with given low resolution means

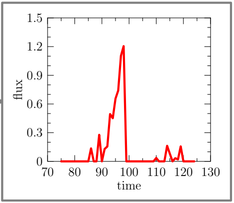


$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$

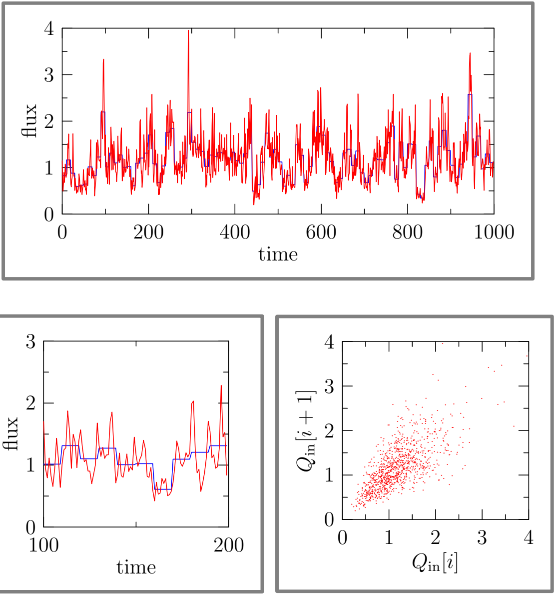




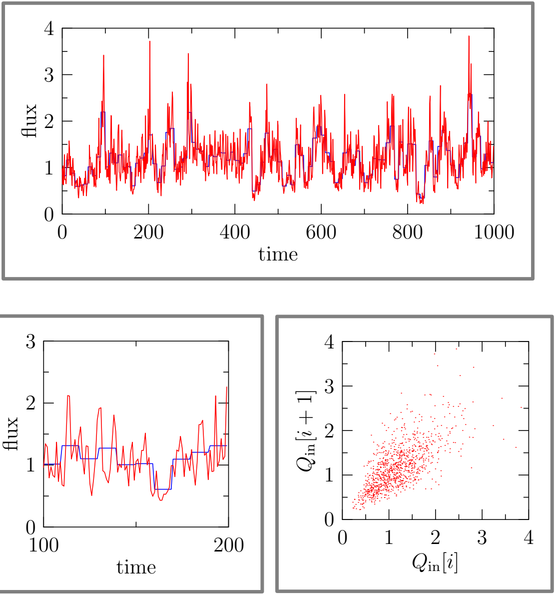
$$\frac{dV}{dt} = Q_{\text{in}}(t) - Q_1(t) - Q_2(t)$$
$$Q_1(t) = \max\left\{0.5 * V(t), 0\right\}$$
$$Q_2(t) = \max\left\{0.5 * \big(V(t) - 2.5\big), 0\right\}$$



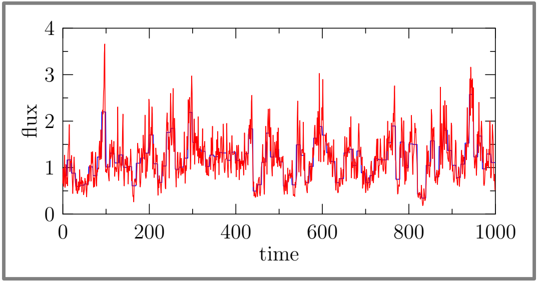
Solution 4: use new data with old model

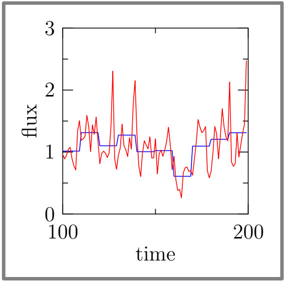


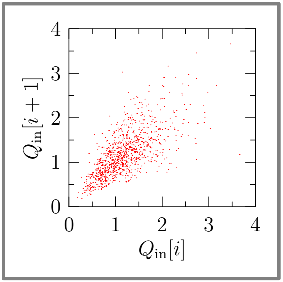
Solution 4: use new data with old model



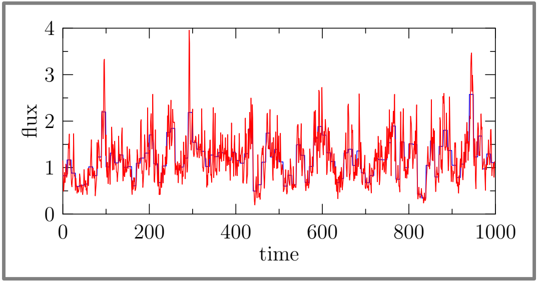
Solution 4: use new data with old model

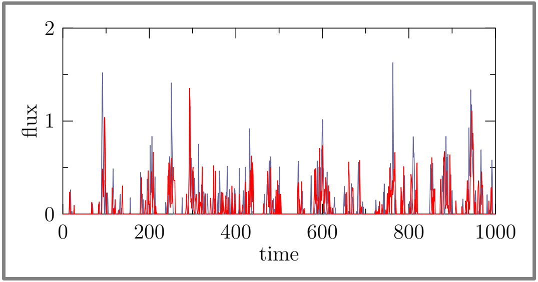




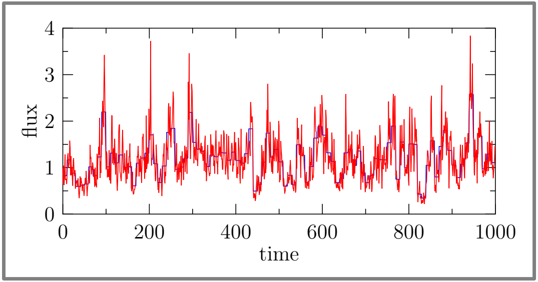


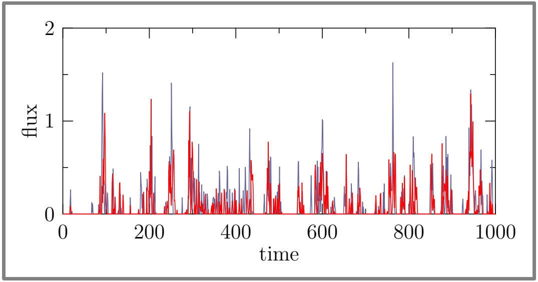
Solution 4: use new data with old model



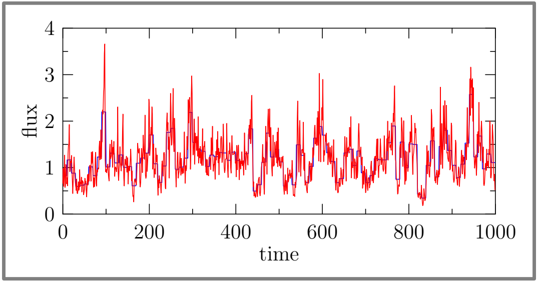


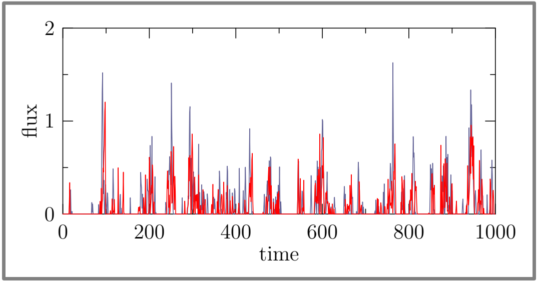
Solution 4: use new data with old model



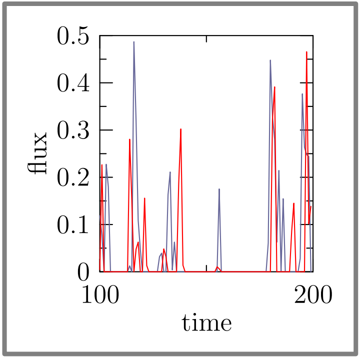


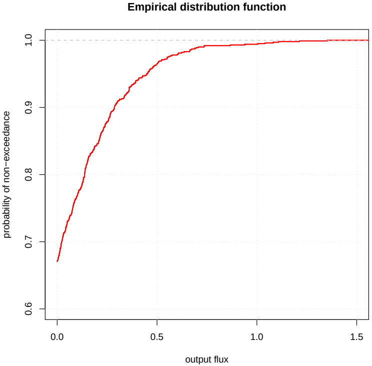
Solution 4: use new data with old model





Solution 4: use new data with old model

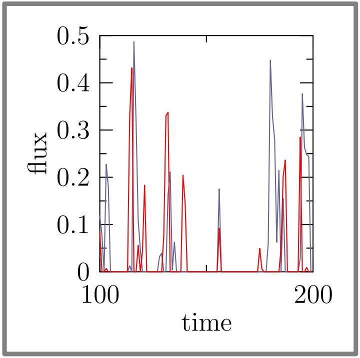


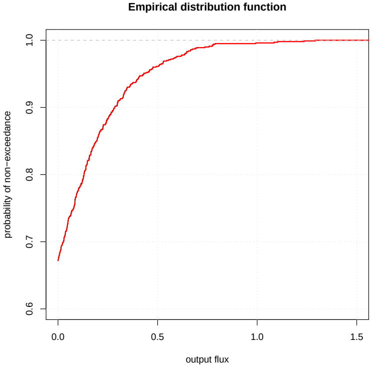


large differences in least squares

small differences in statistics

Solution 4: use new data with old model

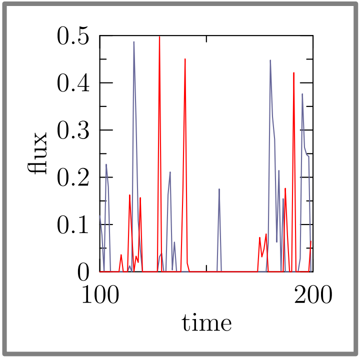


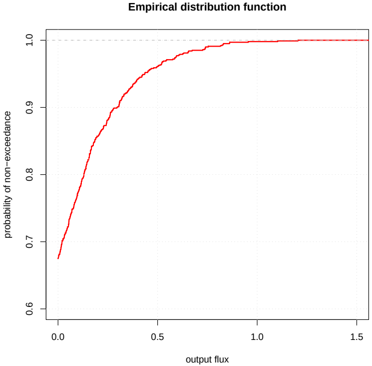


large differences in least squares

small differences in statistics

Solution 4: use new data with old model

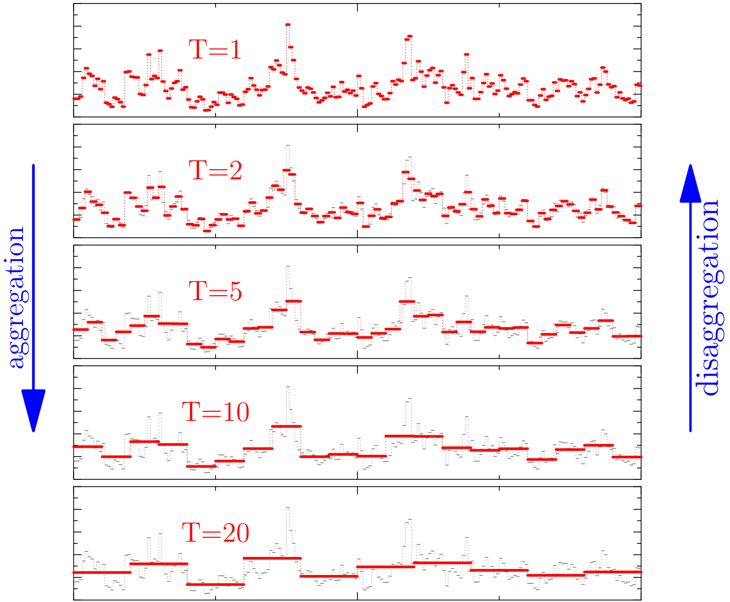




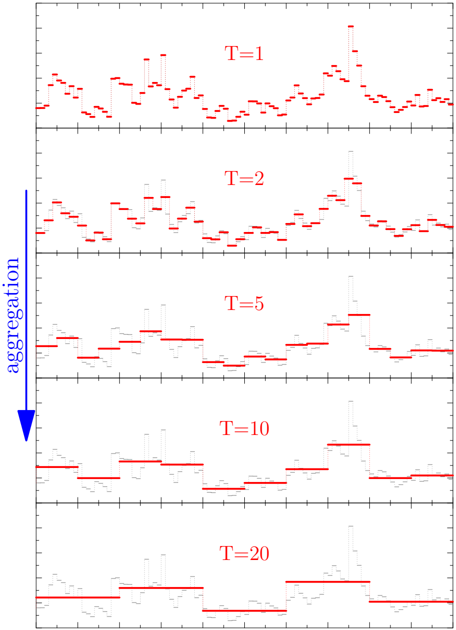
large differences in least squares

small differences in statistics

Technique: disaggregation

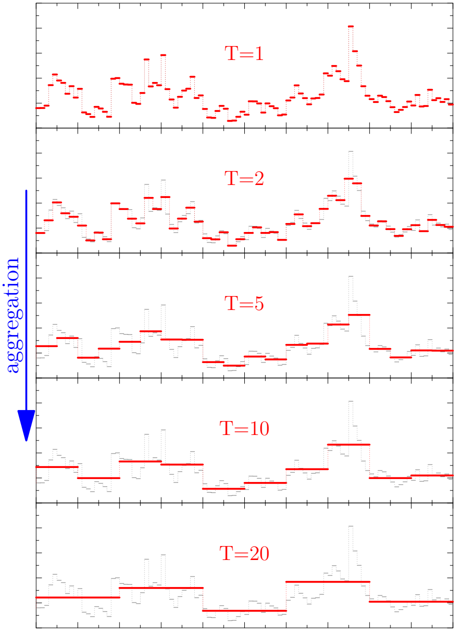


Technique: disaggregation



aggregation = simple averaging

Technique: disaggregation



aggregation = simple averaging

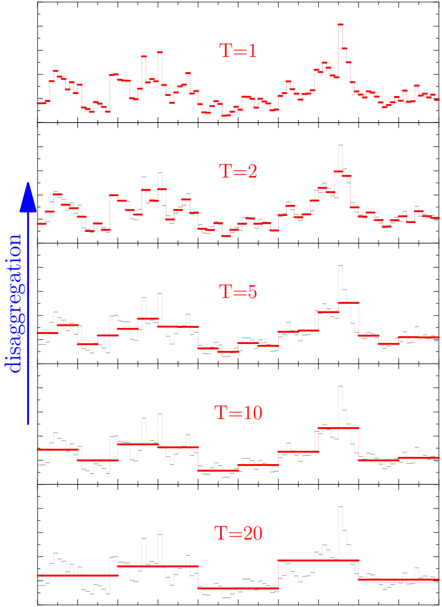
for $n\,T < t \leq (n+1)\,T$:

$$X_T(t) = \frac{1}{T} \int_{nT}^{(n+1)T} X(\tau)\, d\tau$$

for $n\,T < k \leq (n+1)\,T$:

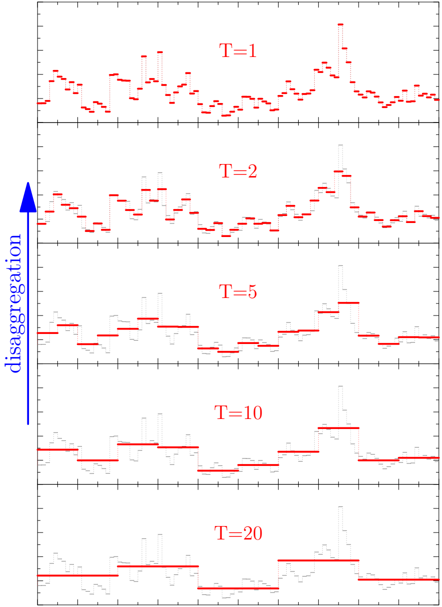
$$X_T(k) = \frac{1}{T} \sum_{i=nT}^{i=(n+1)T} X_1(i)$$

Technique: disaggregation



disaggregation
=
difficult statistical problem

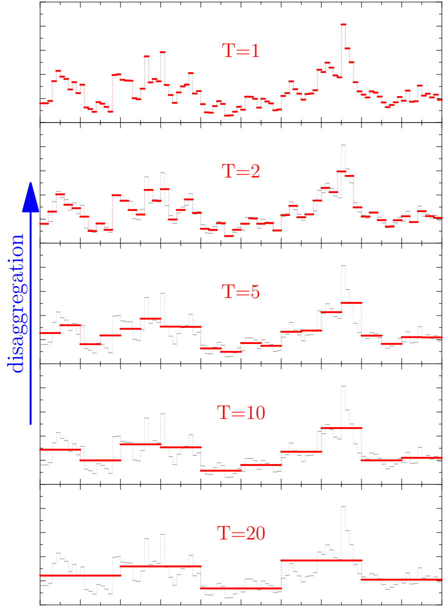
Technique: disaggregation



disaggregation
=
difficult statistical problem

given means
reconstruct variation
by simulation

Technique: disaggregation

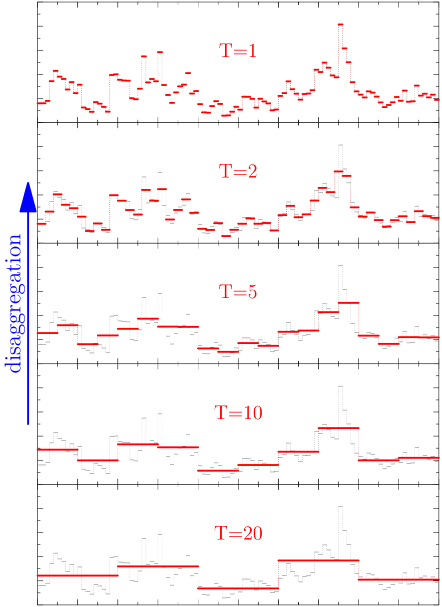


disaggregation
=
difficult statistical problem

given means
reconstruct variation
by simulation

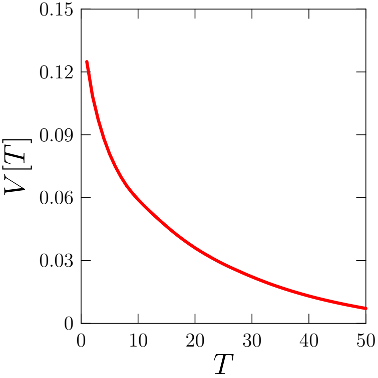
so we need a good statistical
description of stochastic variation
over the aggregation scales

Technique: disaggregation



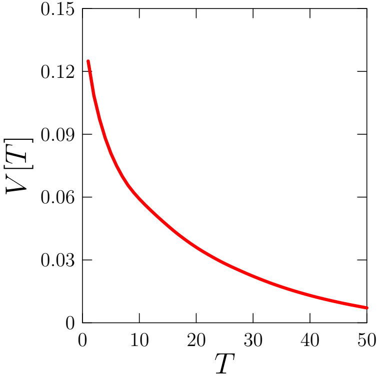
statistical description of stochastic variation over the aggregation scales

$$\mathtt{V}[T] = \text{variation at level } T$$



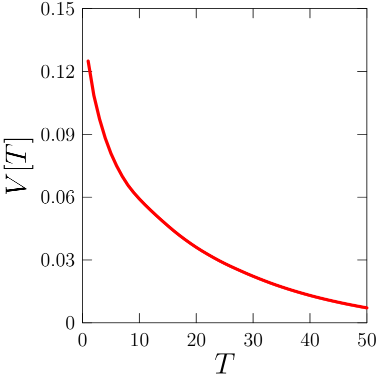
Technique: disaggregation

$\mathtt{V}[T] =$ variation at level T



Technique: disaggregation

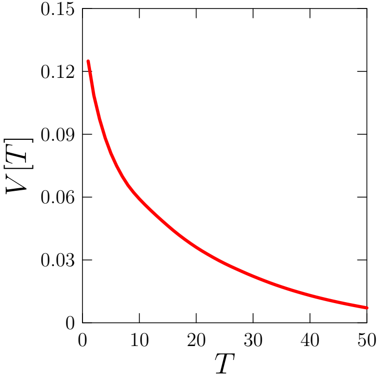
$\mathtt{V}[T] =$ variation at level T



variation reduction function

Technique: disaggregation

$\mathtt{V}[T]$ = variation at level T



variation reduction function

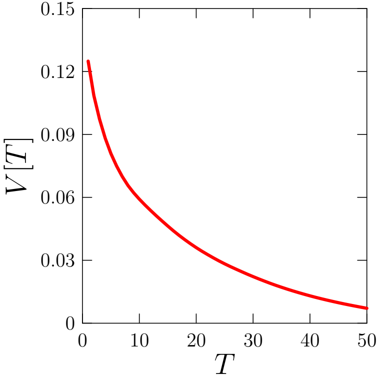
How to measure statistical variation?

1. for "standard" data: classical variance

$$\begin{aligned}\mathtt{V}[T] &= \mathtt{VAR}[X_T] \\ &= \mathtt{E}[X_T^2] - \Big(\mathtt{E}[X_T]\Big)^2\end{aligned}$$

Technique: disaggregation

$\mathtt{V}[T]$ = variation at level T



variation reduction function

How to measure statistical variation?

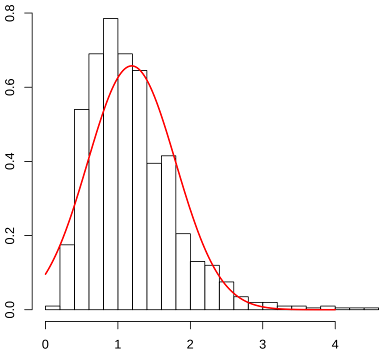
1. for "standard" data: classical variance

$$\mathtt{V}[T] = \mathtt{VAR}[X_T] = \mathtt{E}[X_T^2] - \Big(\mathtt{E}[X_T]\Big)^2$$

2. for "positive" data,
not symmetric around the mean:

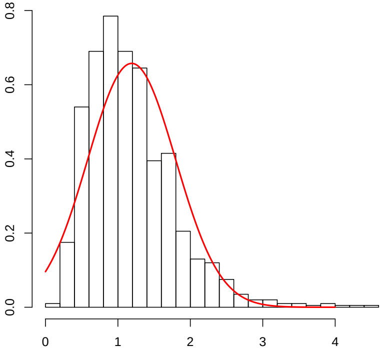
$$\mathtt{V}[T] = \mathtt{VAR}[X_T] = \mathtt{E}[-\log(X_T)] + \log\big(\mathtt{E}[X_T]\big)$$

Technique: disaggregation



measure of variation $\neq$ variance

Technique: disaggregation

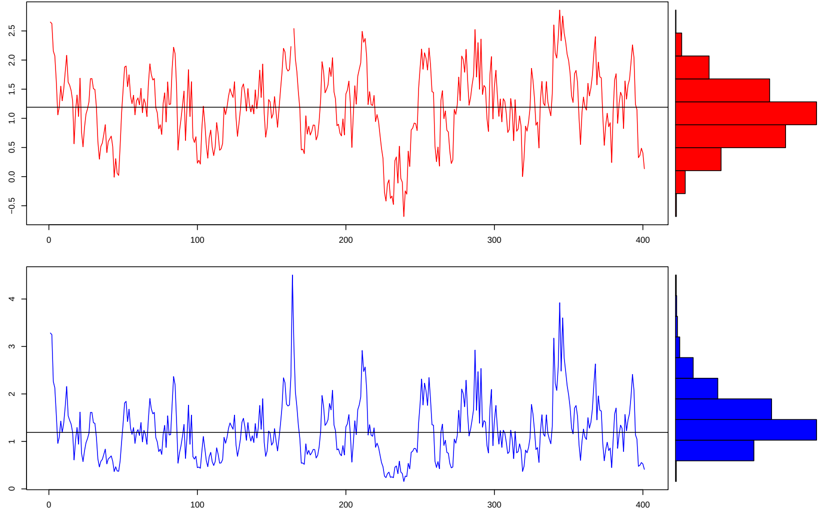


measure of variation $\neq$ variance

reason:

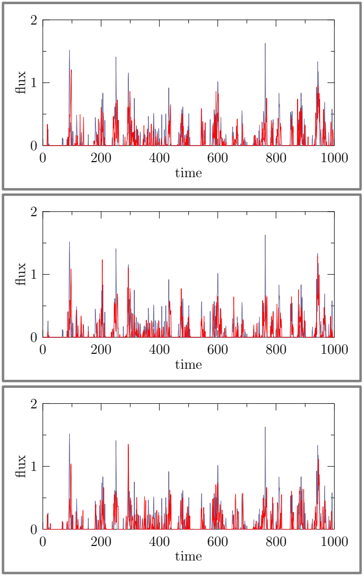
distribution is not gaussian
gaussian $=$ symmetrically around mean

Technique: disaggregation



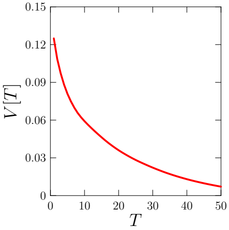
symmetrical disaggregation is not appropriate

Technique: disaggregation



disaggregation

- = stochastically reconstructing variation by simulation procedures
- variation is given by expectations (of logarithms)



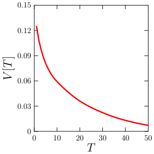
Technique: disaggregation

disaggregation = (mathematically spoken)

Technique: disaggregation

disaggregation = (mathematically spoken)

simulate $X(1)$, $X(2)$, $\ldots X(n) \ldots$ given:



in general

1. $\mathtt{E}[X]$
2. $\mathtt{V}[X_1]$
3. $\mathtt{V}[X_2]$
4. $\vdots$

in the example

1. $\mathtt{E}[X(i)]$
2. $\mathtt{E}[-\log(X(i))]$
3. $\mathtt{E}\Big[-\log\Big(\dfrac{X(i)+X(i+1)}{2}\Big)\Big]$
4. $\vdots$

Technique: feauture based modeling

Moment problem
Given functions $\phi_j(x)$ and corresponding numbers M_j, find a probability density f such that for $j = 1, \ldots, K$:

$$\int \mathrm{d}x \; f(x) \; \phi_j(x) = M_j$$

- $\phi_i(x)$: feature function
- M_j: given moment

Technique: feauture based modeling

Moment problem

Given functions $\phi_j(x)$ and corresponding numbers M_j, find a probability density f such that for $j = 1, \ldots, K$:

$$\int \mathsf{d}x\ f(x)\ \phi_j(x) = M_j$$

- $\phi_i(x)$: feature function
- M_j: given moment
- general statement, makes sense for all spaces, including time series

Technique: feauture based modeling

Moment problem

Given functions $\phi_j(x)$ and corresponding numbers M_j, find a probability density f such that for $j = 1, \ldots, K$:

$$\int \mathrm{d}x \; f(x) \; \phi_j(x) = M_j$$

- $\phi_i(x)$: feature function
- M_j: given moment
- general statement, makes sense for all spaces, including time series
- problem: does not define a unique probability density

Technique: feauture based modeling

Moment problem

Given functions $\phi_j(x)$ and corresponding numbers M_j, find a probability density f such that for $j = 1, \ldots, K$:

$$\int \mathrm{d}x\ f(x)\ \phi_j(x) = M_j$$

Technique: feauture based modeling

Moment problem

Given functions $\phi_j(x)$ and corresponding numbers M_j, find a probability density f such that for $j = 1, \ldots, K$:

$$\int \mathrm{d}x \; f(x) \; \phi_j(x) = M_j$$

Answer : Jaynes Theorem

there is a unique probability maximizing the entropy

with the following form :

$$f(x) = \frac{e^{-\sum_j \lambda_j \, \phi_j(x)}}{Z(\lambda_1, \lambda_2, \ldots, \lambda_K)}$$

where

$$Z(\lambda_1, \lambda_2, \ldots, \lambda_K) = \int \mathrm{d}x \; e^{-\sum_j \lambda_j \, \phi_j(x)}$$

Technique: feature based modeling: does it work?

Jaynes theorem = theoretical result:

$$f(x) = \frac{e^{-\sum_j \lambda_j \, \phi_j(x)}}{Z(\lambda_1, \lambda_2, \ldots, \lambda_K)} \quad Z(\lambda_1, \lambda_2, \ldots, \lambda_K) \;\; = \int \mathrm{d}x \; e^{-\sum_j \lambda_j \, \phi_j(x)}$$

But: does it work in practice?
In order to be practical, we need to be able to:

1. to fit to data
2. to simulate
3. to apply it to time series
4. to apply it to disaggregation problems

Technique: feature based modeling: does it work?

Probabilities as from Jaynes theorem:

$$f(x) = \frac{e^{-\sum_j \lambda_j \, \phi_j(x)}}{Z(\lambda_1, \lambda_2, \ldots, \lambda_K)} \quad Z(\lambda_1, \lambda_2, \ldots, \lambda_K) \;\; = \int \mathrm{d}x \; e^{-\sum_j \lambda_j \, \phi_j(x)}$$

have one **big** problem:

$Z(\lambda_1, \lambda_2, \ldots, \lambda_K)$ can almost never be calculated analytically

but also have one **large** advantage:

there does exist a simulation technique: Metropolis-Hastings

Technique: Metropolis-Hastings

To simulate from the density: $f(x) = \dfrac{e^{-E(x)}}{\int_{\mathcal{X}} e^{-E(\xi)}\, d\xi}$

we take *any* Markov transition kernel $q(x'|x)$, and construct a sequence $x(n)$ as follows:

- let x=x(n) be the previous in the sequence
- simulate a candidate x' out of the density $q(\bullet|x)$
- calculate the acceptance function by :

$$a(x', x) = min\left\{ \frac{e^{-E(x')}\ q(x|x')}{e^{-E(x}\ q(x'|x)}, 1 \right\}$$

- let $U \in [0, 1]$ uniformly randomly chosen;
- $$\begin{aligned} &\text{if } U < a \text{ then } x(n+1) = x' \\ &\text{if } U > a \text{ then } x(n+1) = x = x(n) \end{aligned}$$

Technique: Metropolis-Hastings

To simulate from the density: $f(x) = \frac{e^{-E(x)}}{\int_{\mathcal{X}} e^{-E(\xi)}\, d\xi}$

we take *any* Markov transition kernel $q(x'|x)$, and construct a sequence $x(n)$ as follows:

- let x=x(n) be the previous in the sequence
- simulate a candidate x' out of the density $q(\bullet|x)$
- calculate the acceptance function by :

$$a(x', x) = min\left\{ \frac{e^{-E(x')}\; q(x|x')}{e^{-E(x}\; q(x'|x)}, 1 \right\}$$

- let $U \in [0, 1]$ uniformly randomly chosen;
-

$$\text{if } U < a \text{ then } x(n+1) = x'$$
$$\text{if } U > a \text{ then } x(n+1) = x = x(n)$$

!!!evaluation red part not needed!!!

Conclusions

- as a general attitude: replace uncertainty/lack of knowledge by simulations and ensemble-thinking

Conclusions

- as a general attitude: replace uncertainty/lack of knowledge by simulations and ensemble-thinking (and: do not touch what you know)
- for hydrology and time scale problems = stochastic disaggregation

Conclusions

- as a general attitude: replace uncertainty/lack of knowledge by simulations and ensemble-thinking (and: do not touch what you know)
- for hydrology and time scale problems = stochastic disaggregation
- technically (statistics-theory, programming-practice) not yet finished and not-standard