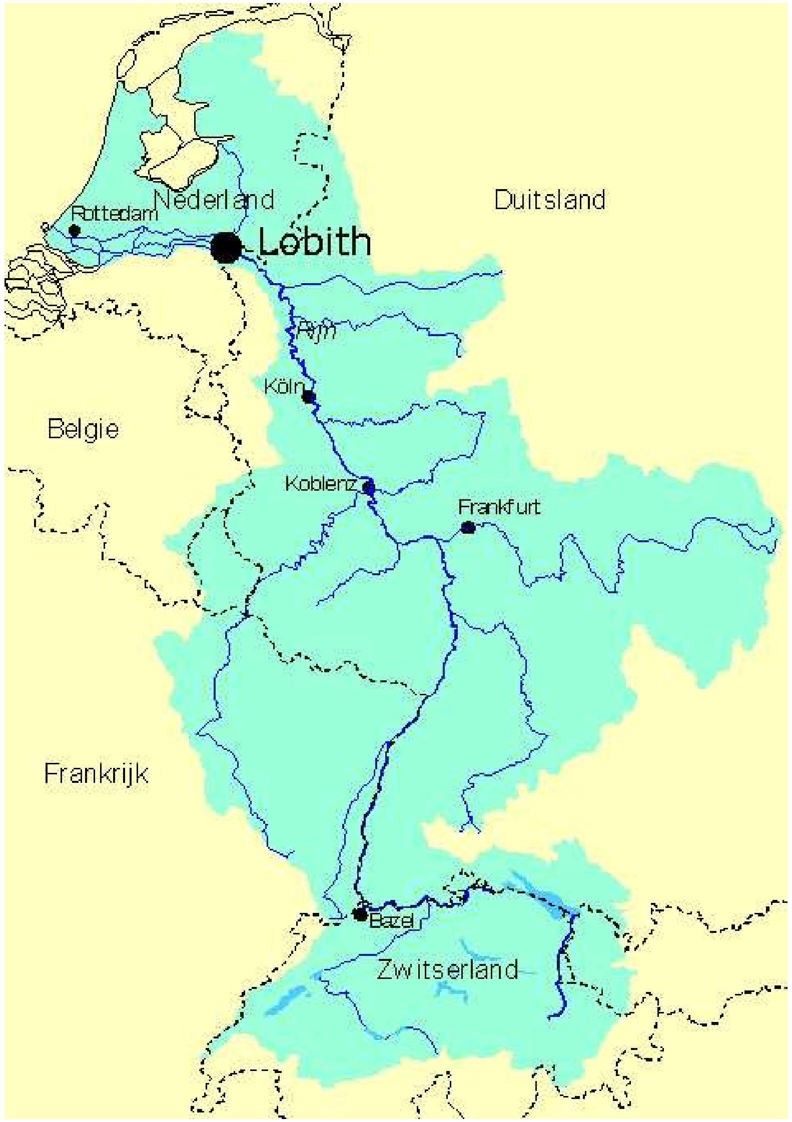
Predicting Lobith

black-box forecasting and uncertainty

Outline

- Linear regression?
- Linear regression model easy?
- Non linear regression model?
- When do new techniques pay off?

Predicting Lobith

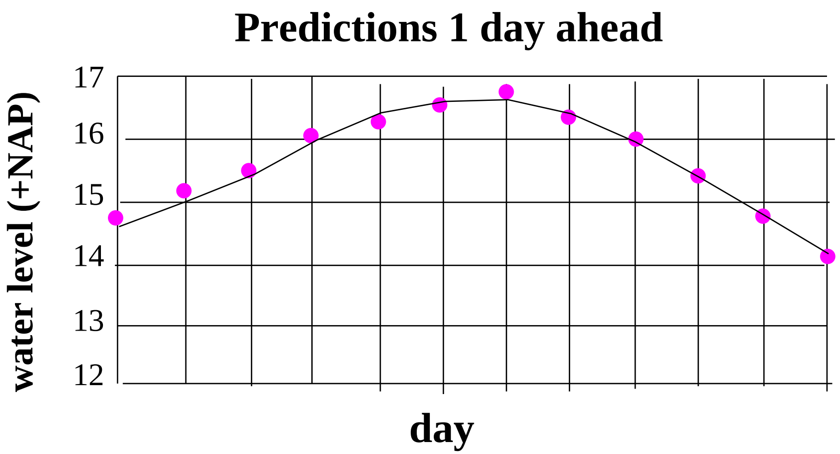


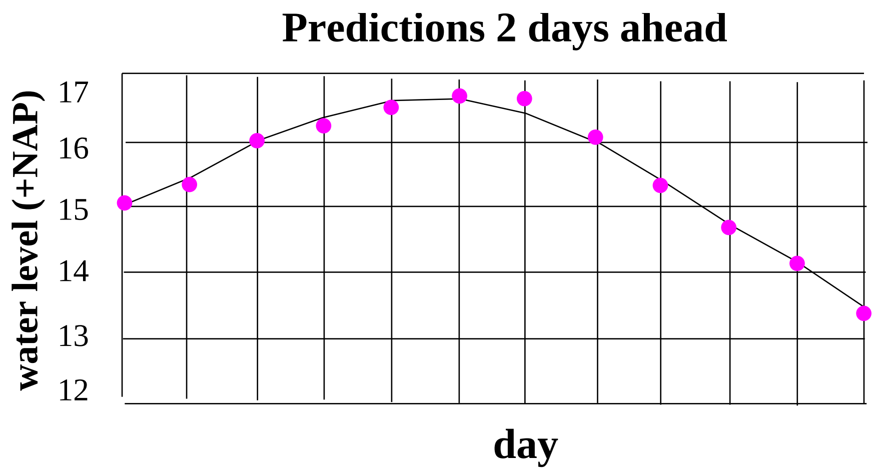
- no hydraulics, no hydrology (see e.g. FLORIJN)
- only black-box
- only one output: levels Lobith
- only 1 day, 2 days, . . . ahead

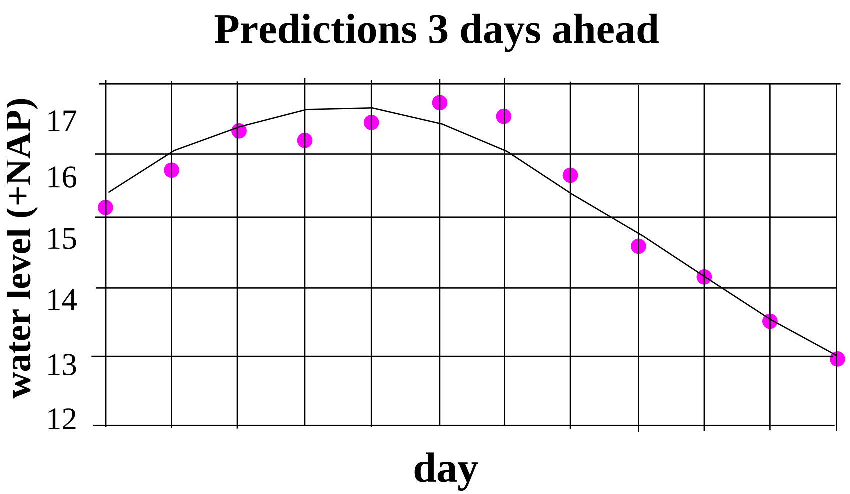
Flood 1995

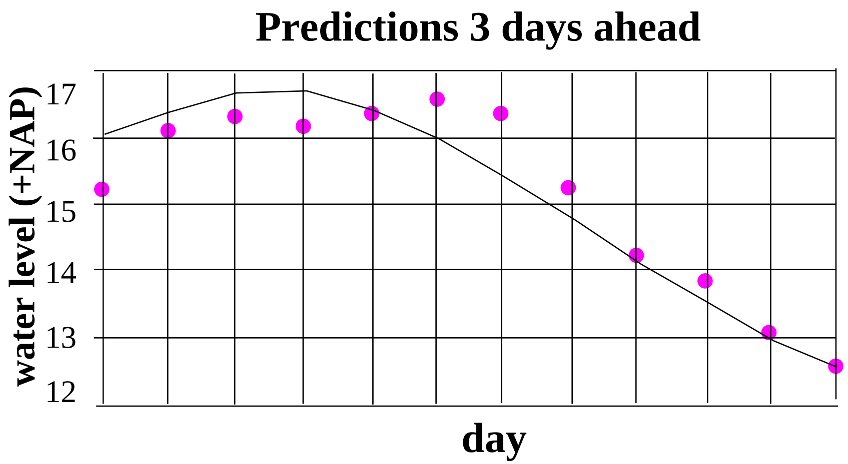
- floods! (1926, 1993, 1995)
- prediction needed:
  1995: 250000 people evacuated
  (decision based on prediction)
  evacuation took several days

Current prediction model works: 1995







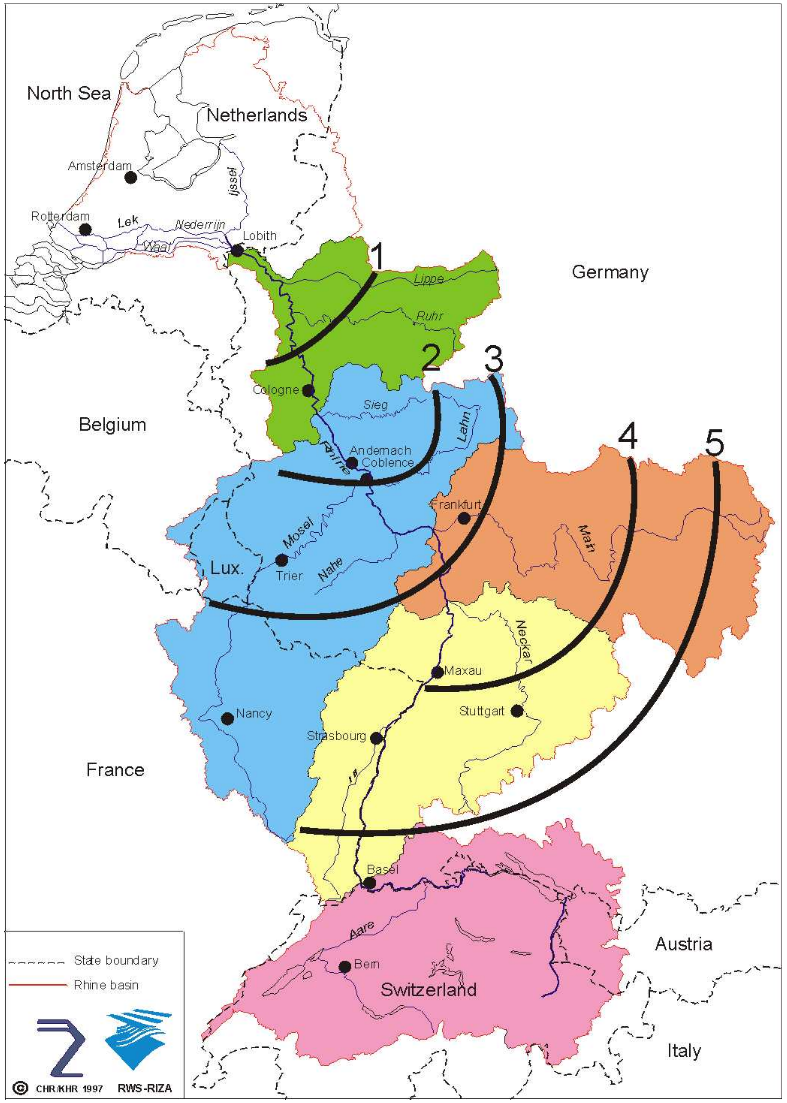


Current Lobith prediction model

operational predicting model = M(ulti)L(inear)R(egression)

$$
\begin{aligned}
H_{\text{Lobith}}(t+1) = \quad & 0.854\ H_{\text{Lobith}}(t) \\
& + 0.472\ H_{\text{Köln}}(t) - 0.333\ H_{\text{Köln}}(t-2) \\
& - 0.031\ H_{\text{Plochingen}}(t-2) \\
& + 0.172\ H_{\text{Hattingen}}(t) - 0.146\ H_{\text{Hattingen}}(t-1) \\
& + 0.250\ P_{\text{Düsseldorf}}(t-1)
\end{aligned}
$$

Predicting Lobith: black-box?

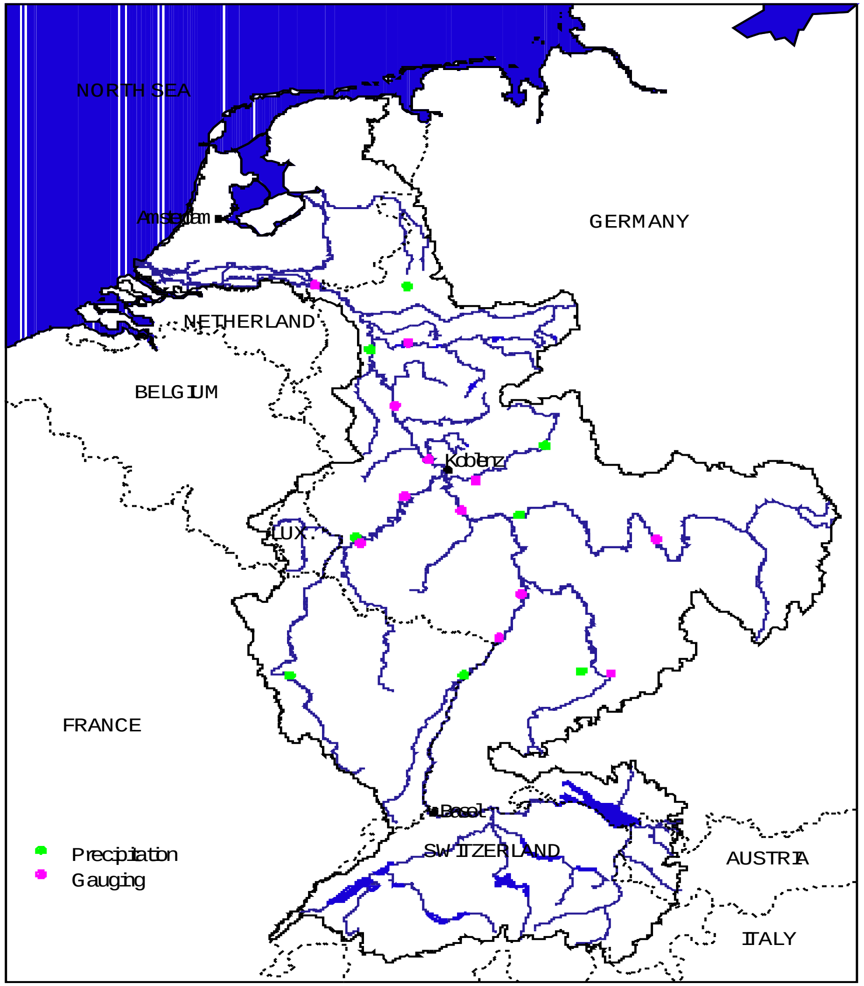


- lag times $\overset{??}{\Longleftarrow}$ travel times
- neg. coeff $\overset{??}{\Longleftarrow}$ level rise

Predicting Lobith: black-box?

$$
\begin{aligned}
H_{\text{Lobith}}(t+2) \quad = & -0.122\, H_{\text{Maxau}}(t-3) & +0.157\, H_{\text{Worms}}(t) \\
& +0.493\, H_{\text{Kaub}}(t) & -0.208\, H_{\text{Andernach}}(t-3) \\
& -0.254\, H_{\text{Köln}}(t-1) & +0.810\, H_{\text{Lobith}}(t) \\
& -0.159\, H_{\text{Plochingen}}(t-2) & +0.200\, H_{\text{Trier}}(t) \\
& +0.243\, H_{\text{Kalkofen}}(t) & -0.147\, H_{\text{Kalkofen}}(t-1) \\
& +0.296\, H_{\text{Hattingen}}(t) & -0.250\, H_{\text{Hattingen}}(t-1)
\end{aligned}
$$

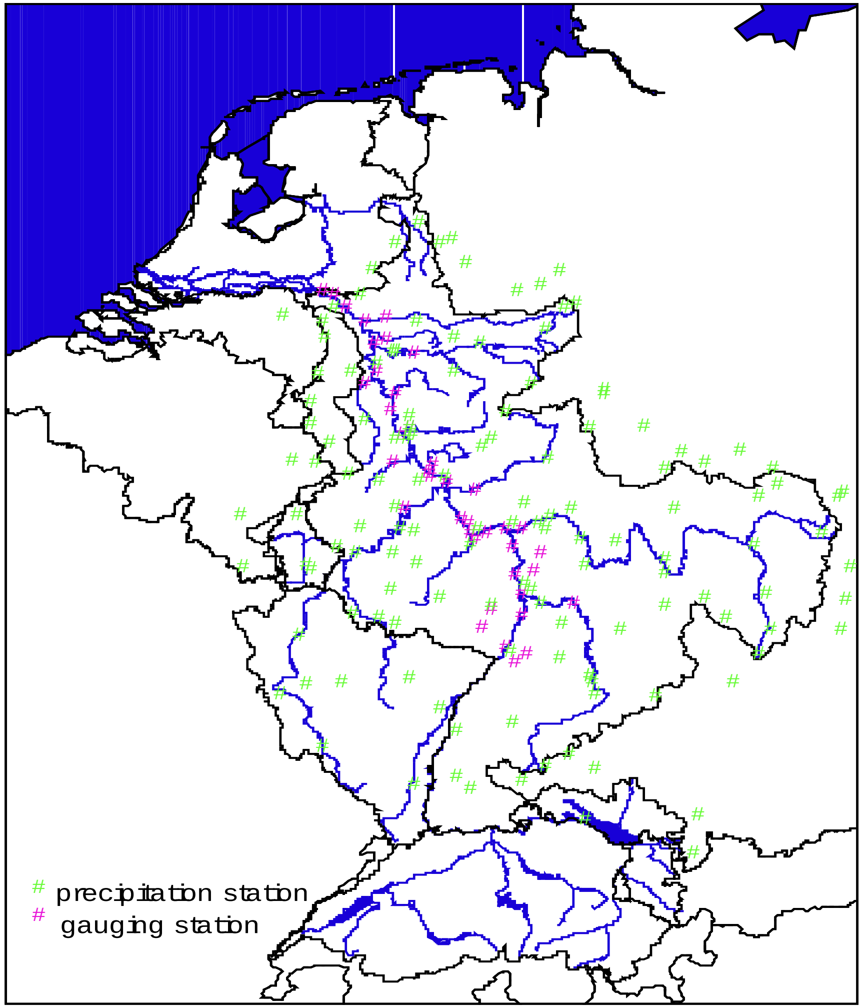
Input for prediction models



Input stations for the statistical forecasting model

- many (10. . . 20. . .) stations available
- this for many years
- with good quality (calibration!)
- levels, discharges

Input for prediction models



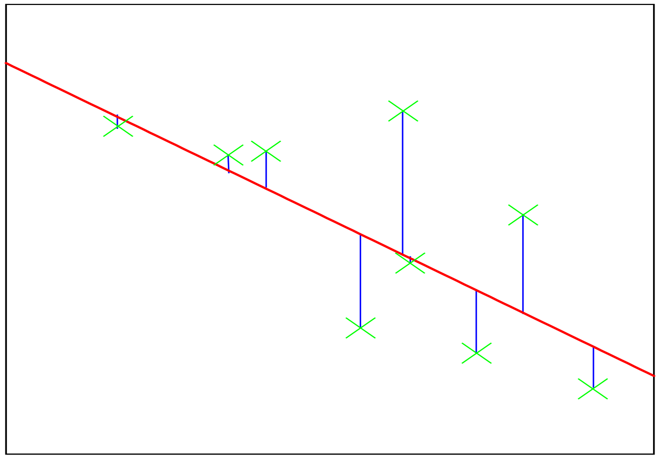
Input stations for the statistical FEWS model

- more stations become available
- at higher time resolution
- including rainfall and rainfall predictions
- and also other predictions and model results

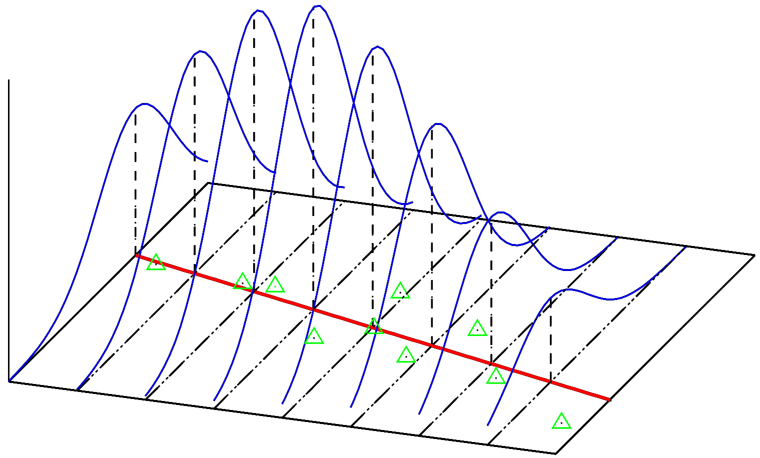
Views on linear regression

data $\{x_i, y_i\}$

Algebraic



Probabilistic



$y = a\ x + b$

$(a, b) \Leftarrow \text{Minimize}\left(\sum \epsilon_i{}^2\right)$

$y = a\ x + b = \mathbf{E}[Y|X = x]$

$(X, Y) = \text{Gaussian} \Leftarrow \text{ML}(\{x_i, y_i\})$

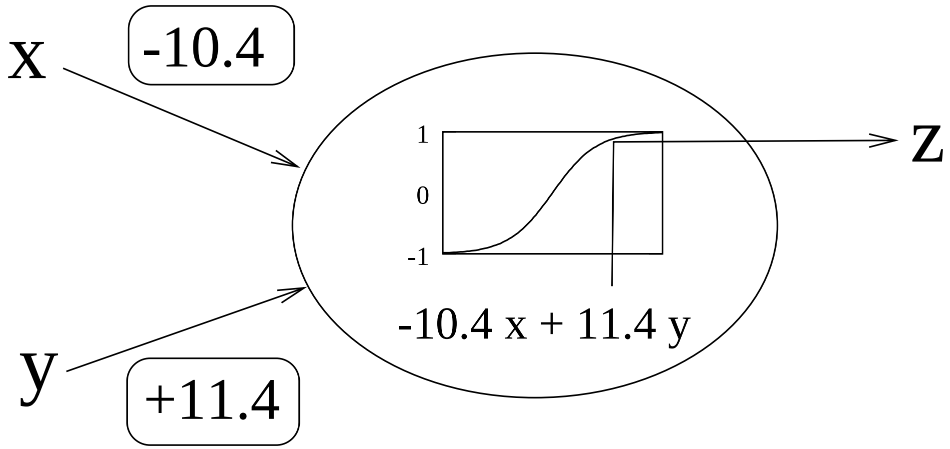
Non-linear black-box models

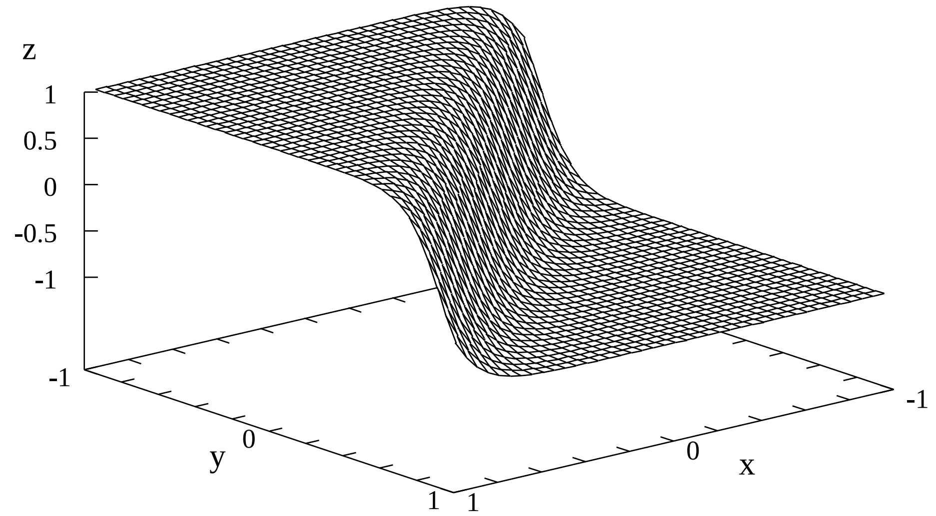
Nowadays many models non-linear black-box models available:

- Neural Networks
- MARS
- piecewise linear models
- ...
- Kernel Regression

both in theory and in practice (implementations available)

Neural Networks: One Neuron

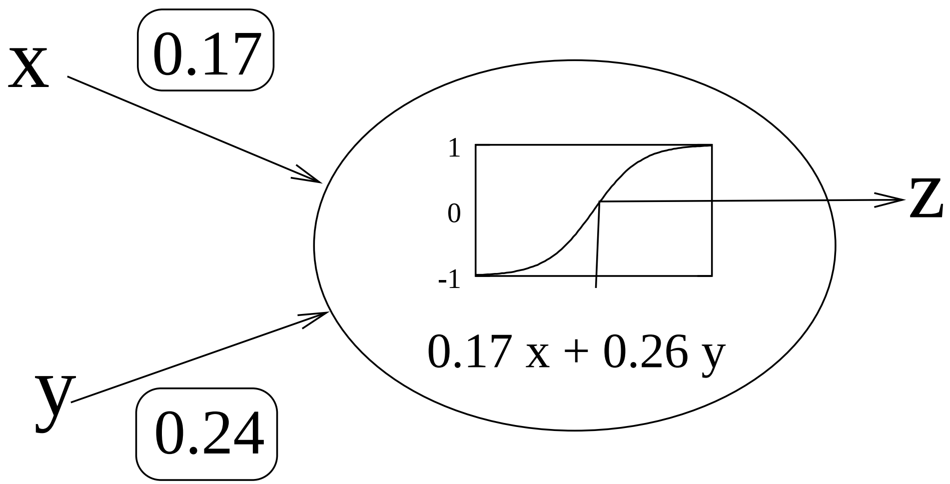


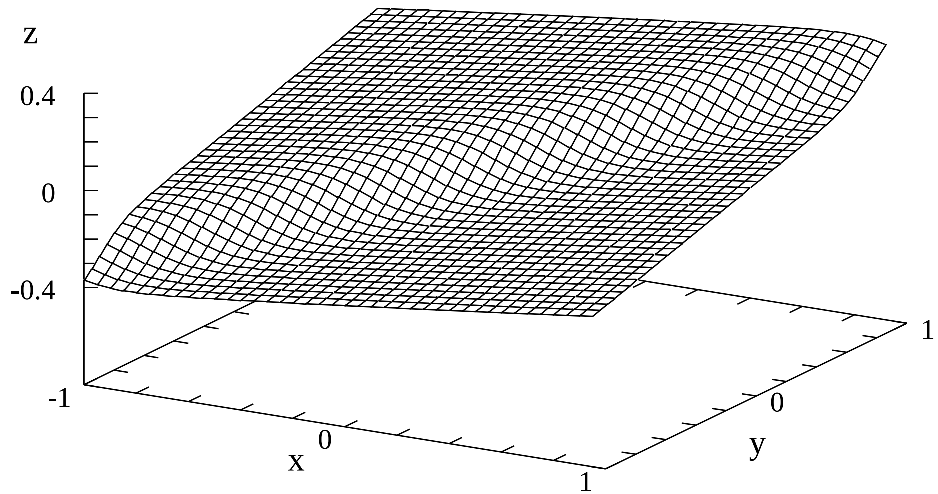


- inputs (x, y), outputs (z)
- weights = parameters $(-10.4, 11.4)$
- weighted input non-linearly transformed in output ($z = \tanh(-10.4 * x + 11.4 * y)$)

NN can model non-linear functions

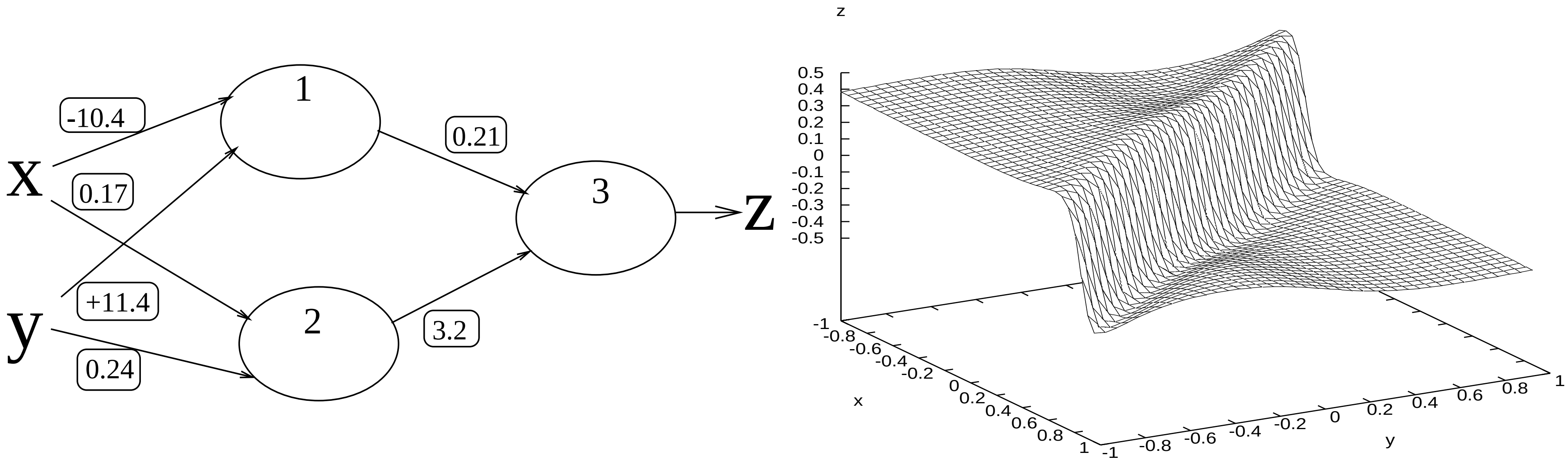
Neural Networks: an other Neuron





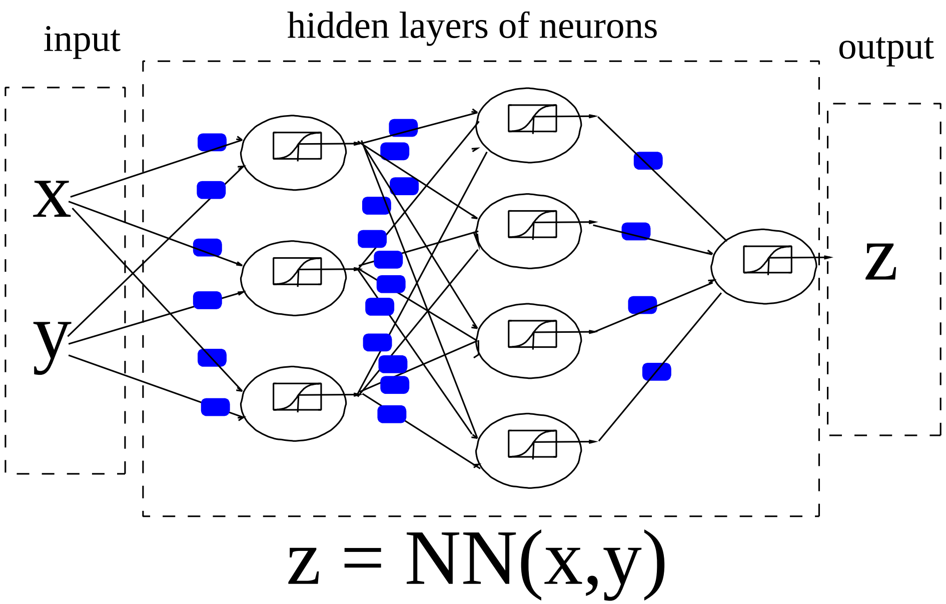
- weights determine character approximation
- NN generalize linear functions

Neural Networks: combination of Neurons



- outputs of Neurons can be input to other Neurons
- all combinations of linear and non-linear are possible

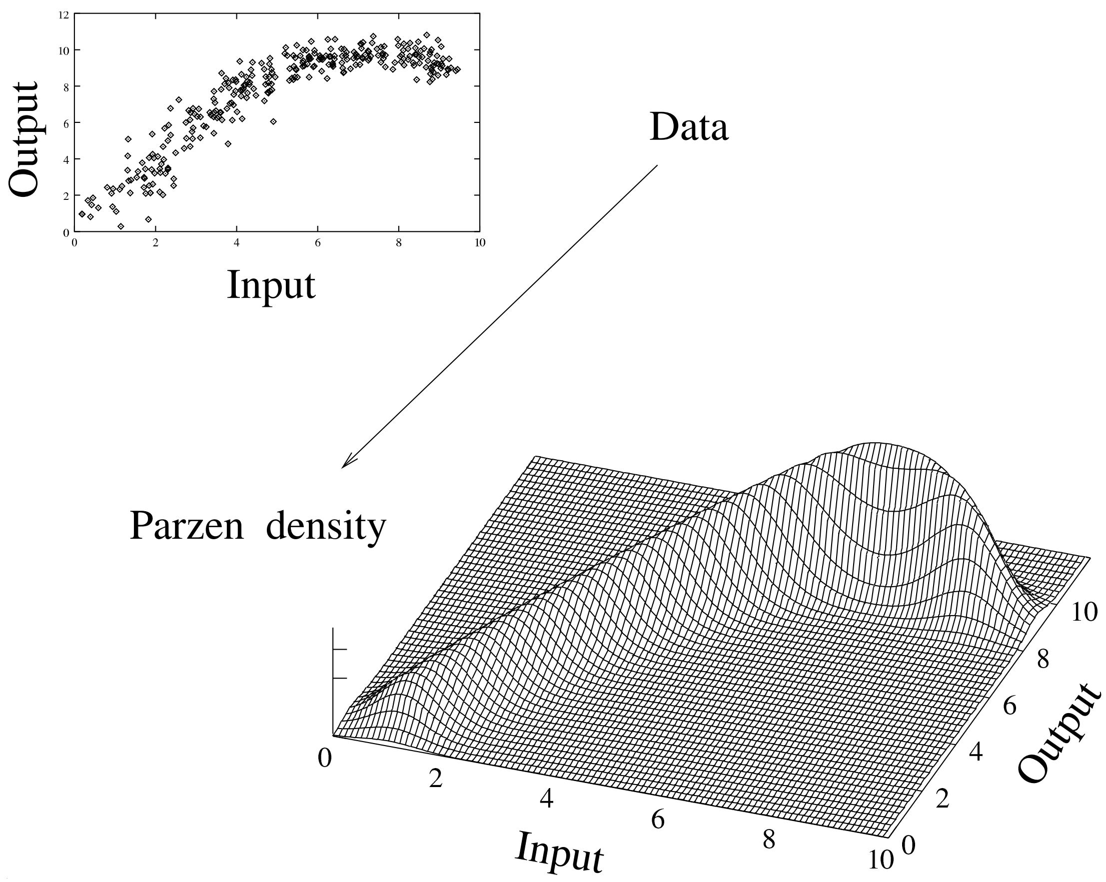
Neural Network



- can (in principle) model everything!
- many parameters
- many (local) minima
- fitting (supervised learning): $\{(x_i, y_i), (z_i)\}$ + least squares
- lengthy calibration

$$H_{\text{Lobith}}(t+1) \stackrel{?}{=} \text{NN}\left(H_{\text{Lobith}}(t), H_{\text{Köln}}(t), H_{\text{Köln}}(t-2), H_{\text{Plochingen}}(t-2), \ldots\right)$$

Probabilistic regression (1): fitting density

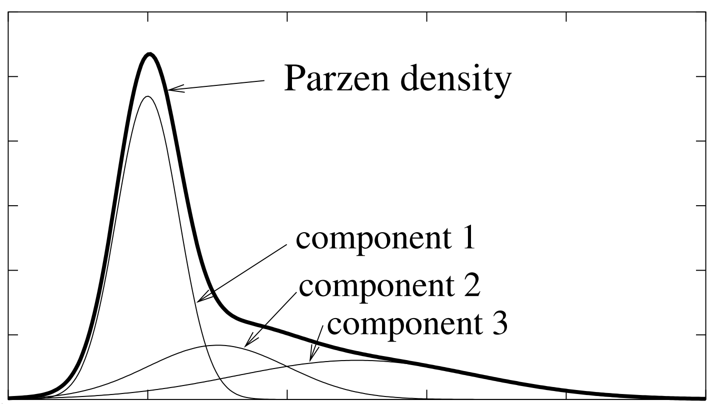


e.g. Parzen

or Mixtures of Gaussians:

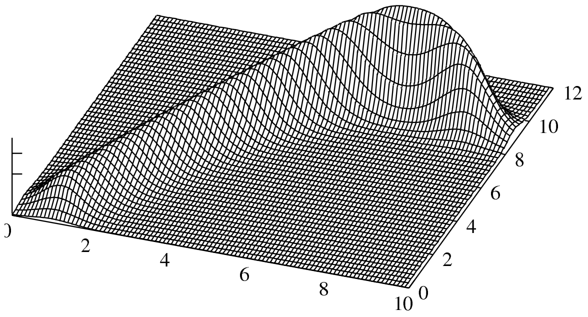
$$f(x) = \sum_{i=1}^{N_c} w_i \; f_{\mathcal{N}(\mu_i,\Sigma_i)}(x)$$

$$f_{\mathcal{N}(\mu,\boldsymbol{\Sigma})}(\mathbf{x}) = \text{Gaussian(mean=}\mu\text{, covariance=}\Sigma\text{)}$$

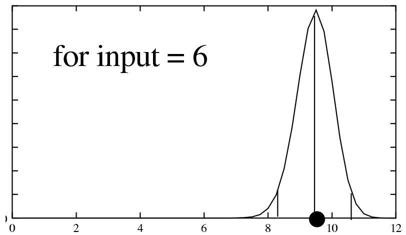


Probabilistic regression(2) conditionals

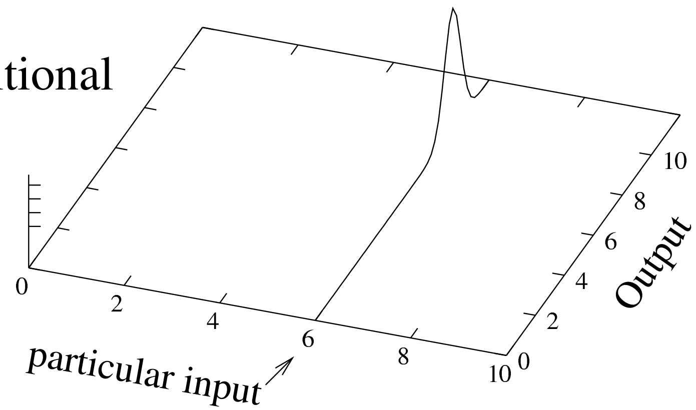
Taking conditionals

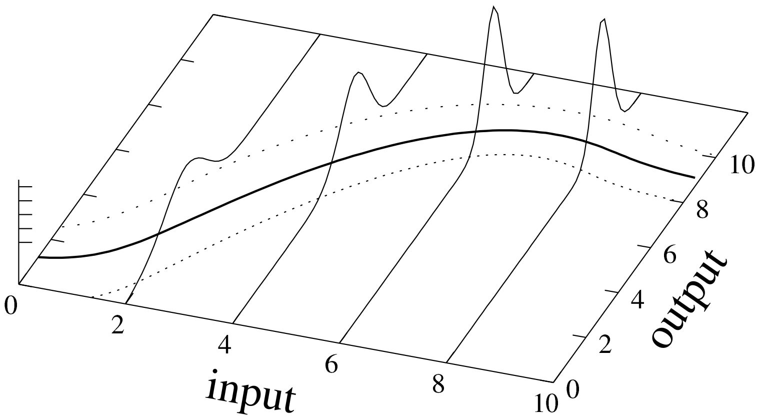


to regression

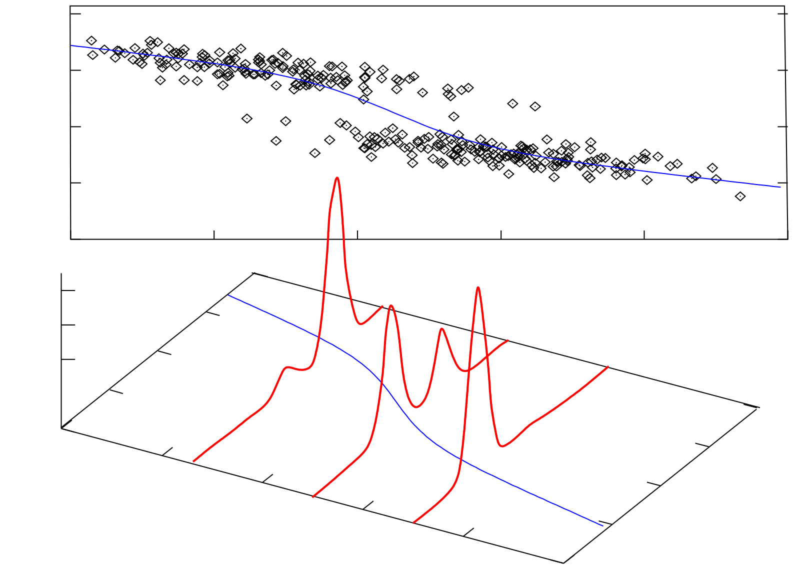


Conditional

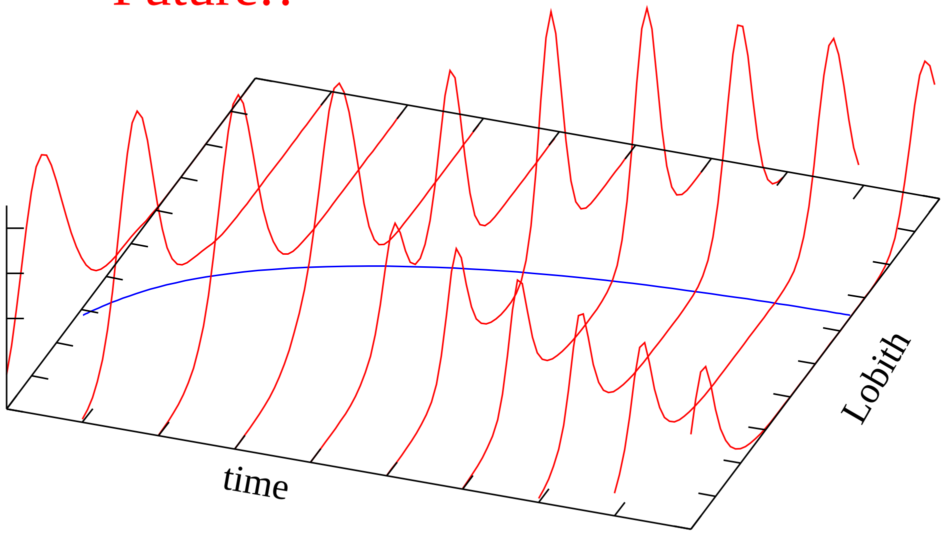




Probabilistic predicting Lobith?



Future:?



Potential:

- non linear features captured
- probabilistic regression has more output

Price:

- large calibration time
- performance depends on dimension
- new parameters

Improving predictions Lobith?

predicting Lobith:

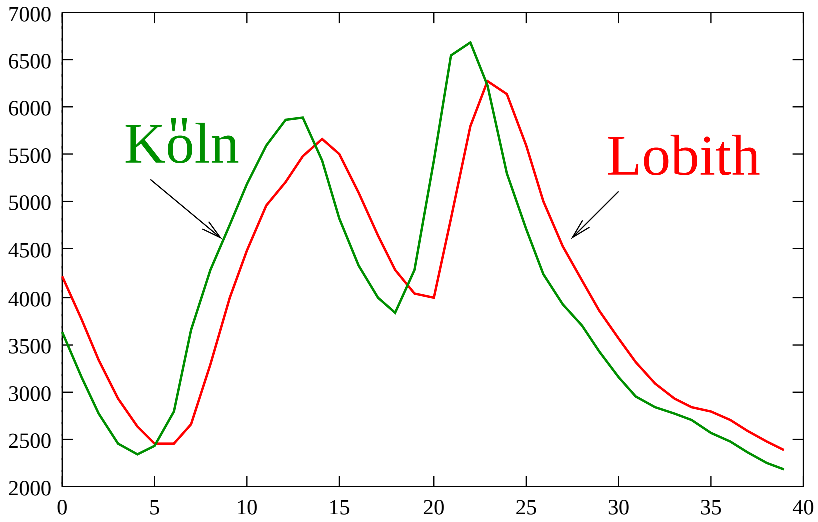
- linear regression: works
- non-linear regression techniques promise
  - improvement
  - more output

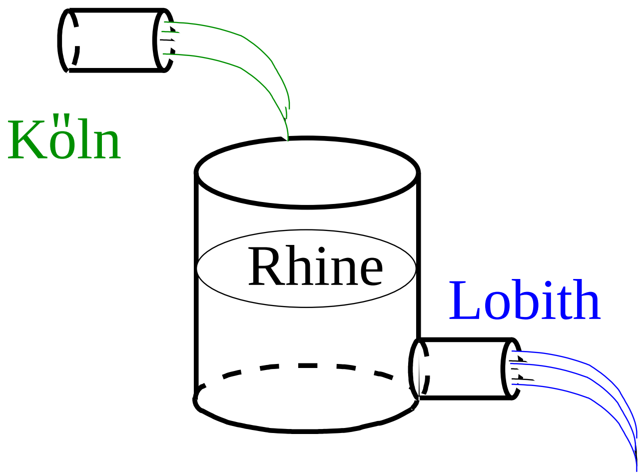
my experience:

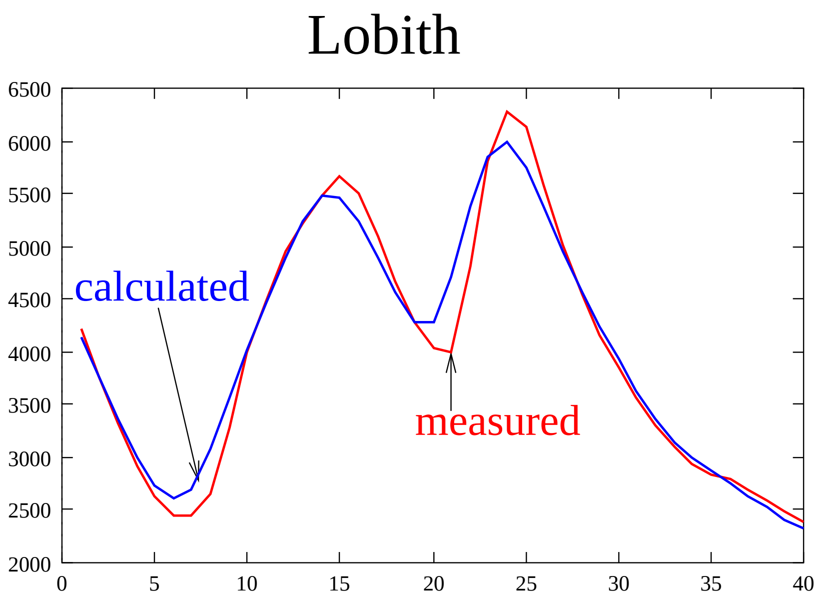
new techniques do not improve significantly linear results

WHY?

Flood wave of 1976: Rhine is linear?







Multilinear regression: an example

G	M	H	A
4.4	4.97	39.7	89.72
5	5.01	39.8	92.34
5.5	5.26	40.4	96.56
6.25	5.49	40.5	99.63
7	5.7	40.7	102.97
7.75	6.03	41.2	107.53
8.525	6.55	41.3	112.3
9.1	6.85	40.6	114.9
9.3	7.43	40.7	122.51
9.5	8.02	40.6	129.51
9.7	8.34	39.8	133.73

data: number of Golfers in the USA (1960-1970)

- "G" : number of golfers (in millions)
- "M" : median income
- "H" : average working hours/week
- "A" : average weekly earnings

Regression results: one input

I	M	H	A	R^2
-2.1533 ± 1.1791	1.5178 ± 0.1833			0.8840
-56.268 ± 43.973		1.574 ± 1.086		0.1892
-6.17867 ± 1.42173			0.12481 ± 0.01291	0.9122

summary:

$$A(0.9122) > M(0.8840) \gg H(\ 0.1892)$$

Regression results: two inputs

remember: A(0.9122) $>$ M(0.8840) $\gg$ H(0.1892)

I	M	H	A	R^2
-45.0852 $\pm$ 9.0547	1.4436 $\pm$ 0.1006	1.0721 $\pm$ 0.2256		0.9697
-40.974432 $\pm$ 9.275522		0.877762 $\pm$ 0.232869	0.118059 $\pm$ 0.008409	0.9684
-12.3271 $\pm$ 4.5712	-2.4797 $\pm$ 1.7615		0.3248 $\pm$ 0.1426	0.9296

summary: M+H $>$ H+A $\gg$ M+A

- "strong"+ "strong" may loose: dependency!; no added information
- "weak" inputs may re-appear: contain conditional extra information

Regression results: three inputs

I	M	H	A	R^2
-43.85402	1.01578	1.01394	0.03509	0.9699
± 10.77331	± 1.67805	± 0.33091	± 0.13735	

compare with previous:

I	M	H	A	R^2
-45.0852	1.4436	1.0721		0.9697
± 9.0547	± 0.1006	± 0.2256		

- M+H+A(0.9699) $\approx$ M+H(0.9697)
- errors on M and A coeff: very high!!
- conclusion : overfitted

Linear regression with many inputs

if there are many possible inputs (which is the case in predicting Lobith):

- the biggest problem is the input selection, not the fitting of the line;
- the more dependency there is in the input data, the more difficult this problem is
  in the Lobith case:
  - strong physical relation between discharges, levels etc, e.g. at conjunctions
  - including lags introduce new dependencies, e.g. $H_{\text{Lobith}}(t) \approx H_{\text{Köln}}(t-1)$
- one must select, to avoid overfitting;
- the more dependency, the greater the danger of overfitting:

curse of dimensionality

Input selection in the linear case

- for multilinear regression techniques, there do exist a number of selection procedures: sequential (hierarchical) techniques, setwise selection, stepwise regression, . . .
- all these techniques have a "try" aspect in them (require a lot of calculations)
- one of the keys why they can be efficiently implemented in computer programs, is that calculation of linear dependency by linear correlation is straightforward

This made the making of the prediction model for Lobith feasible.
Even then: new selection at re-calibration!

Input selection in non-linear case

- using inputs selected on their linear prediction power as input to a non-linear model is frustrating
- brute force is out-of-order
  - try every possible selection of 25 inputs, 1 second = calculation time needed of one particular choice, takes:

$$2^{25}/(60 * 60 * 24 * 365) \approx 1 \text{ year}$$

  - for 50 inputs:

$$2^{50}/(60 * 60 * 24 * 365 * 100) \approx 360\,000 \text{ centuries}$$

- non-linear dependency: Mutual Information is best candidate to replace correlation, but is tricky and costly to calculate in high dimensions!

Experience: non-linear Lobith prediction models

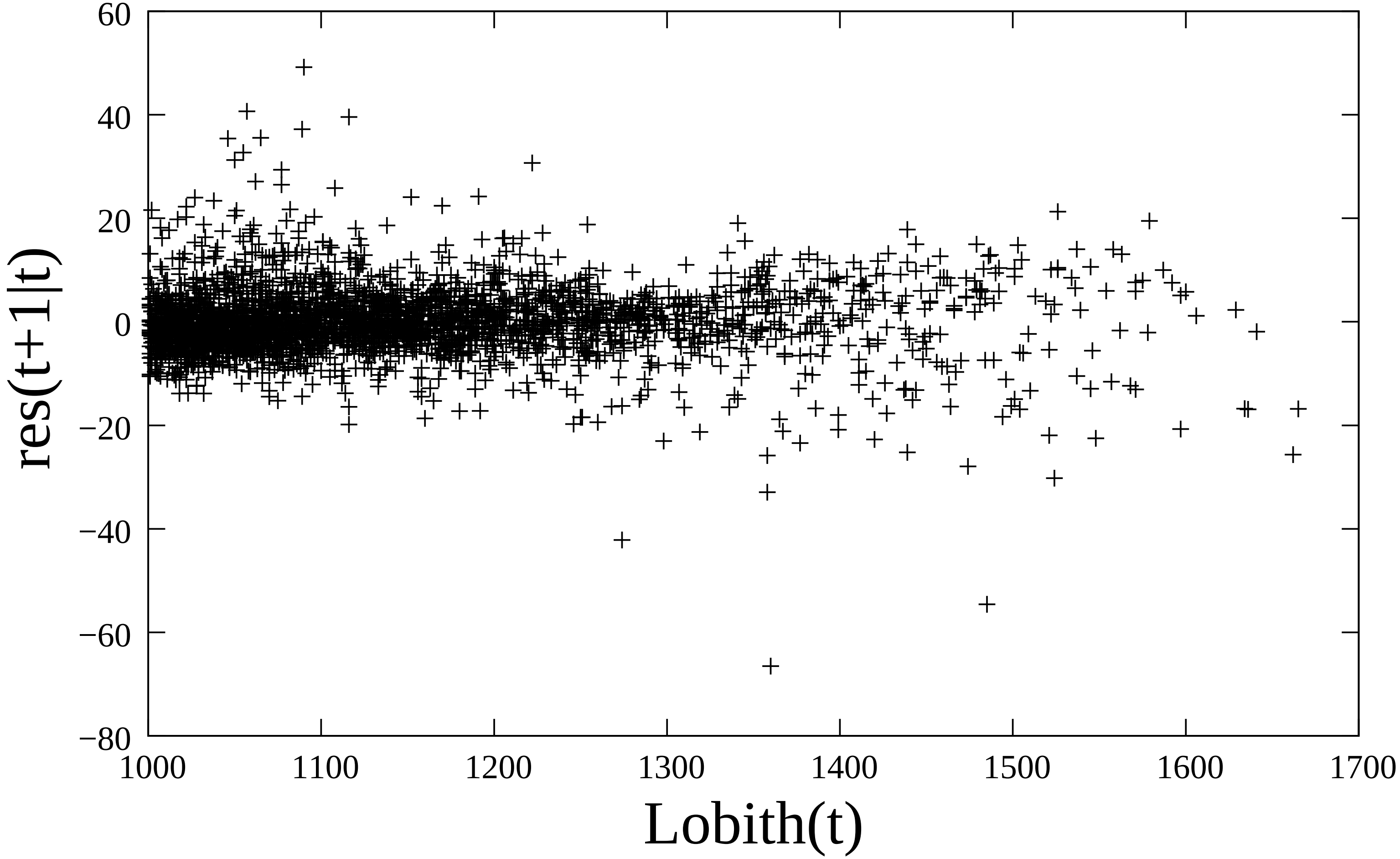
My experience with predicting Lobith with non-linear black box models:

- fitting non-linear models to the linear-selected inputs is feasible, but results are not significantly better
- input selection can not (yet) be tackled (in general)
- expect slow progress
(possible directions: hybrid, residuals, special techniques . . .)

My conclusions:

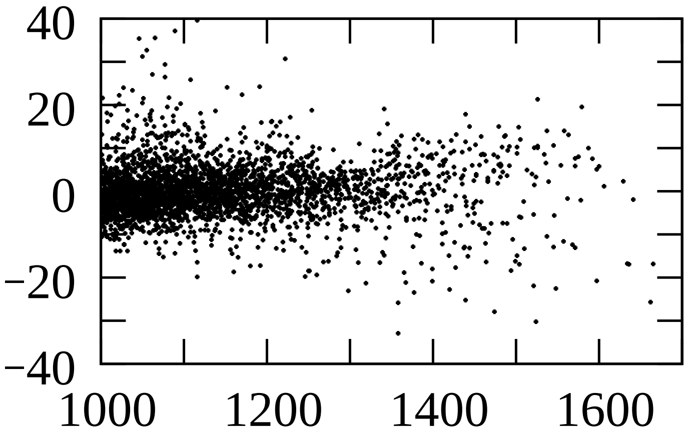
1. the effectiveness of a new technique (NN, Kernel Regression) depends heavily on the total of embedding framework
(analysis of variance, selection procedures, . . .);
2. brute force (try all . . .) does not work always;
3. as distributed modeling and remote sensing is growing, the problem addressed in this talk will become even more important

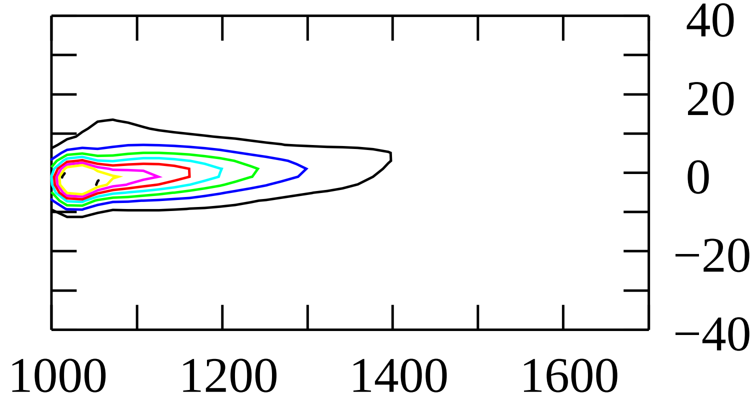
What can be done: an example

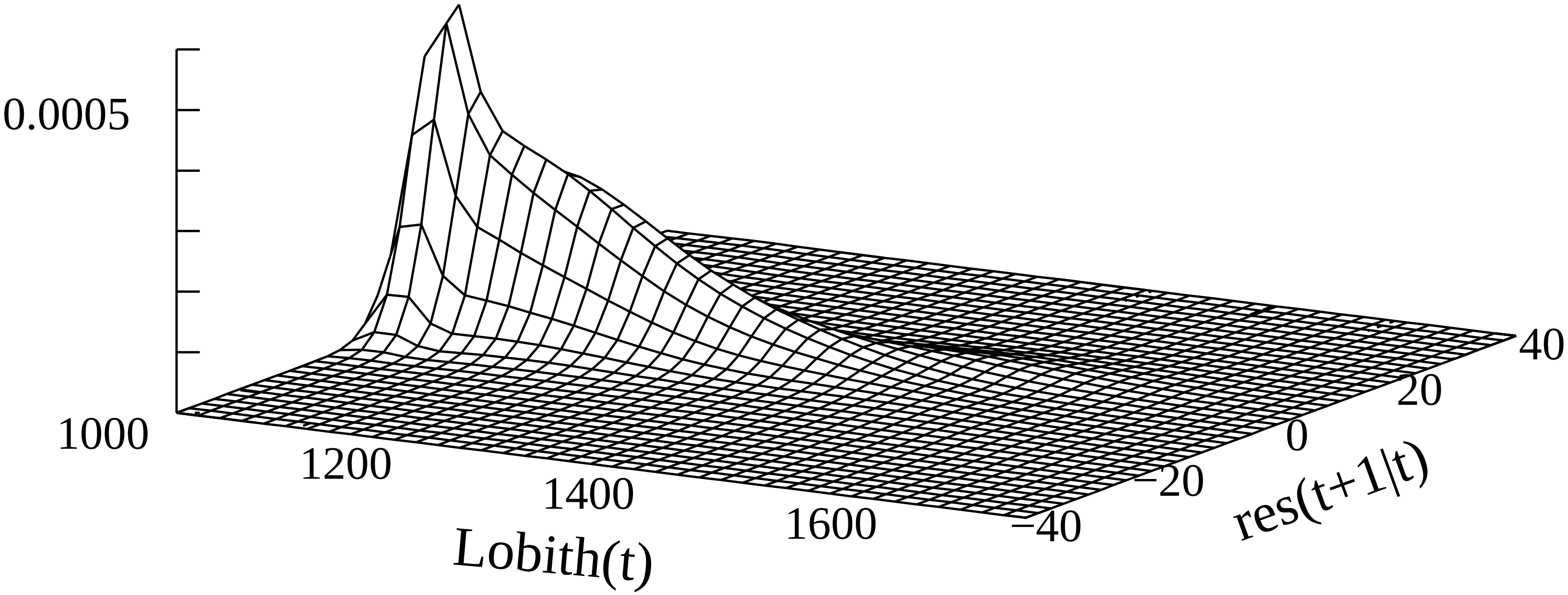


$\mathrm{res}(t + 1|t)$ = residual of predicted level one day ahead

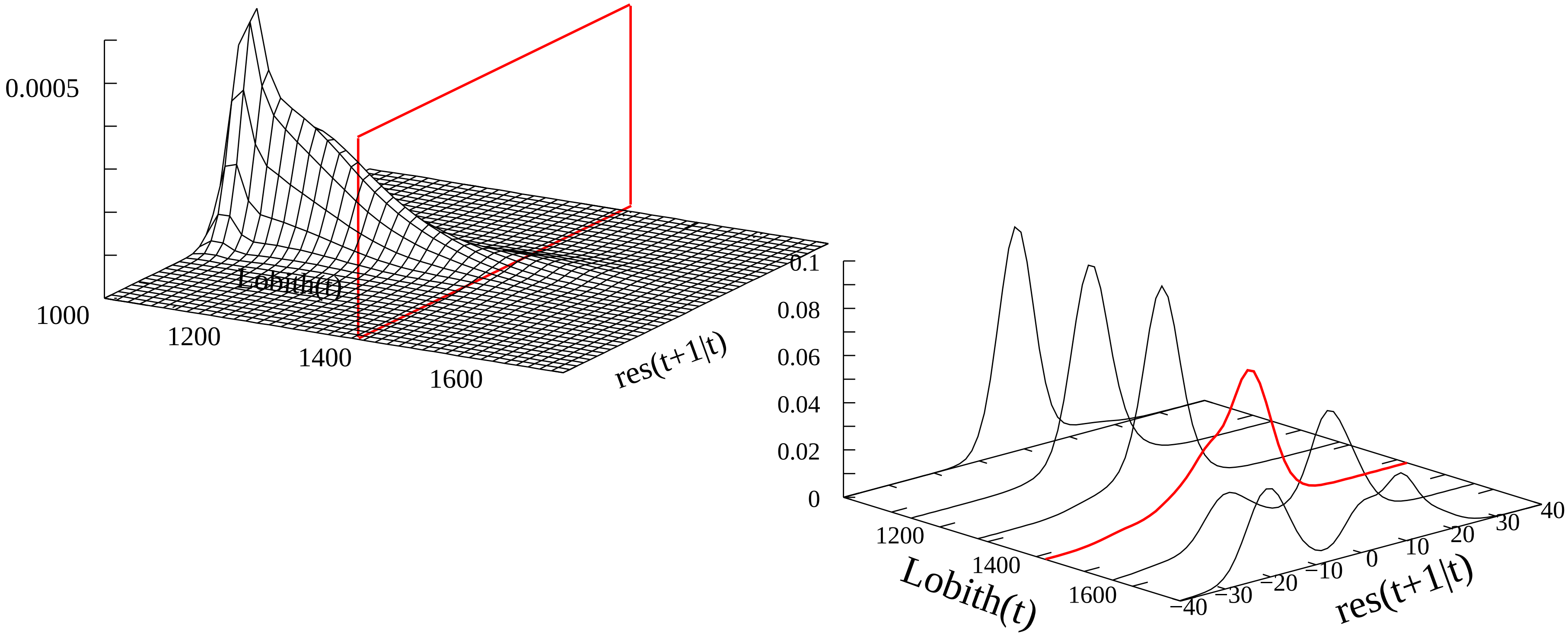
Fitting density





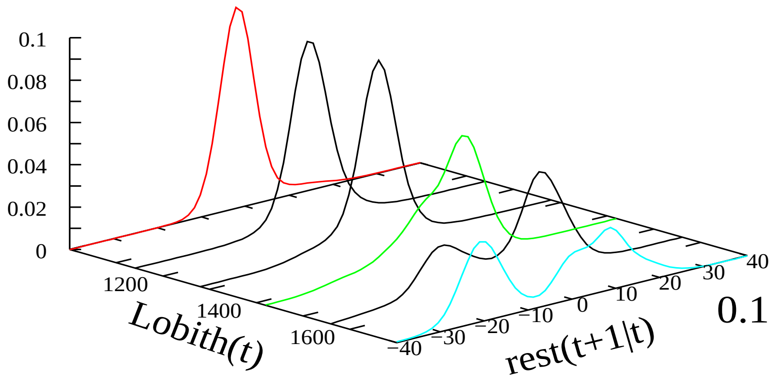


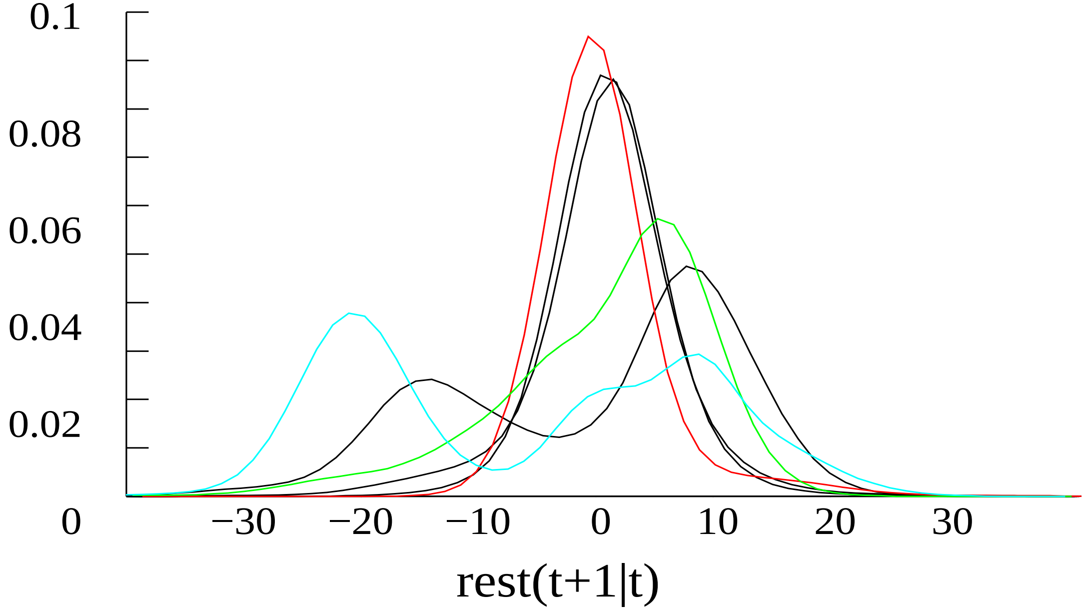
Taking conditionals



conditioning = renormalizing secions

Result: uncertainty information on prediction





uncertainty changes as function of Lobith(t)