Can we trust observational data and models of the state of the Baltic Sea?

Anders Grimvall

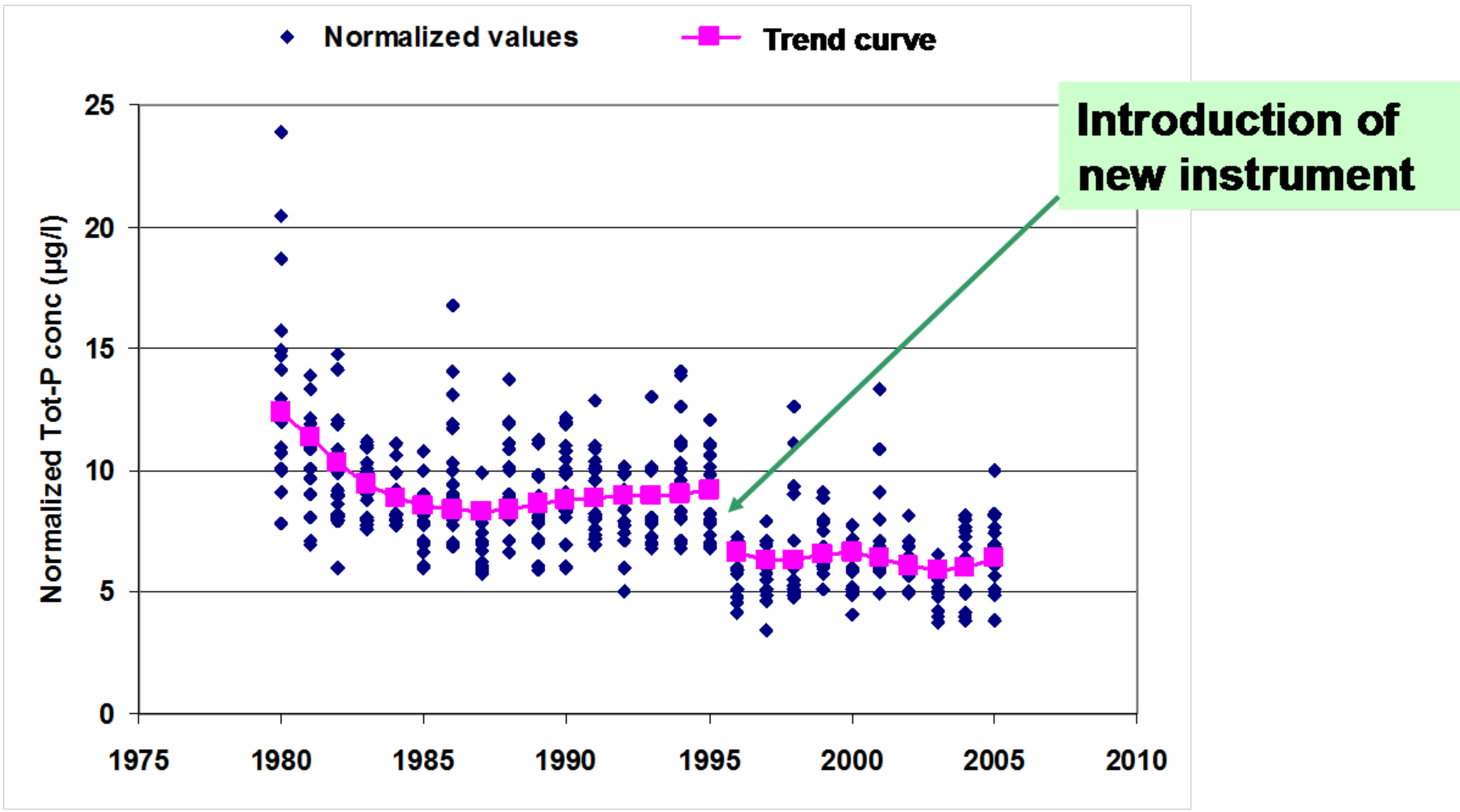
Division of Statistics, Department of Computer and Information Science
Linköping University, SE-58183 Linköping, Sweden

E-mail: anders.grimvall@liu.se

Outline

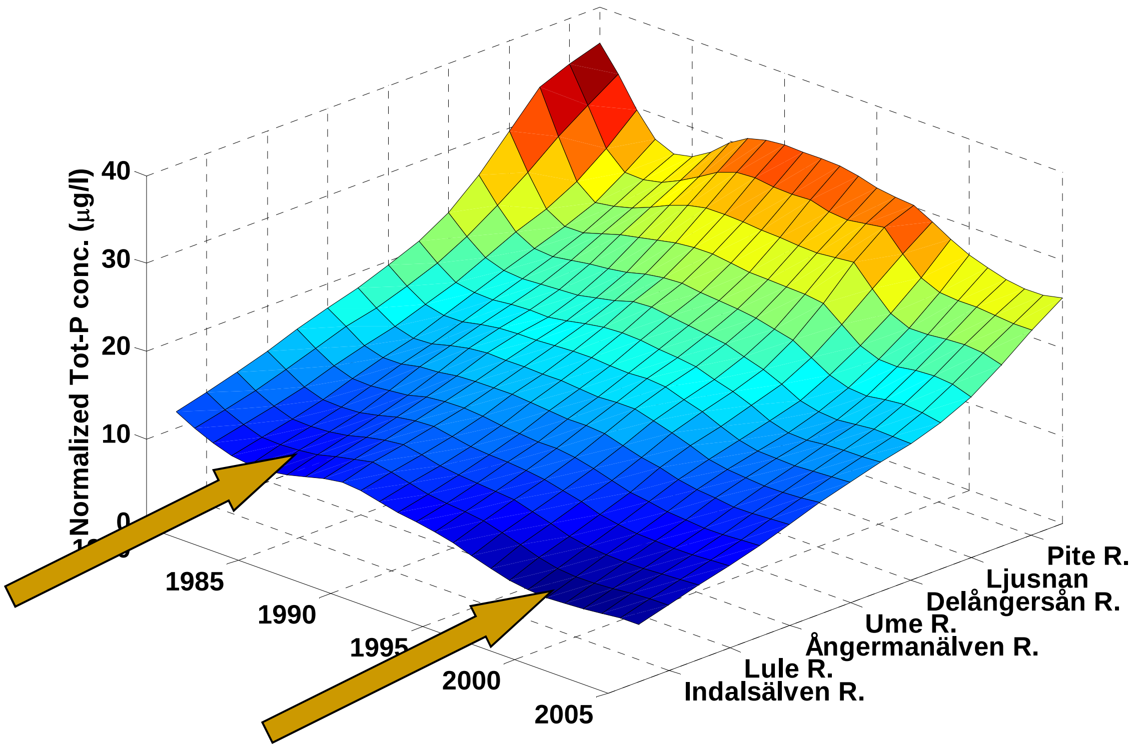
- **Some examples of low data quality**
- **A model of acidification of the Baltic Sea**
- **Observational pH and alkalinity data from the Baltic Proper**
- **General comments on tools for change-point detection and trend assessment**

Trend curve for temperature-normalized concentrations of phosphorus along with measured concentrations at different depths (0.5 – 70 m) at Dagskärsgrundet in Lake Vänern



Trend surface for normalized concentrations of total phosphorus in riverine input to the Gulf of Bothnia

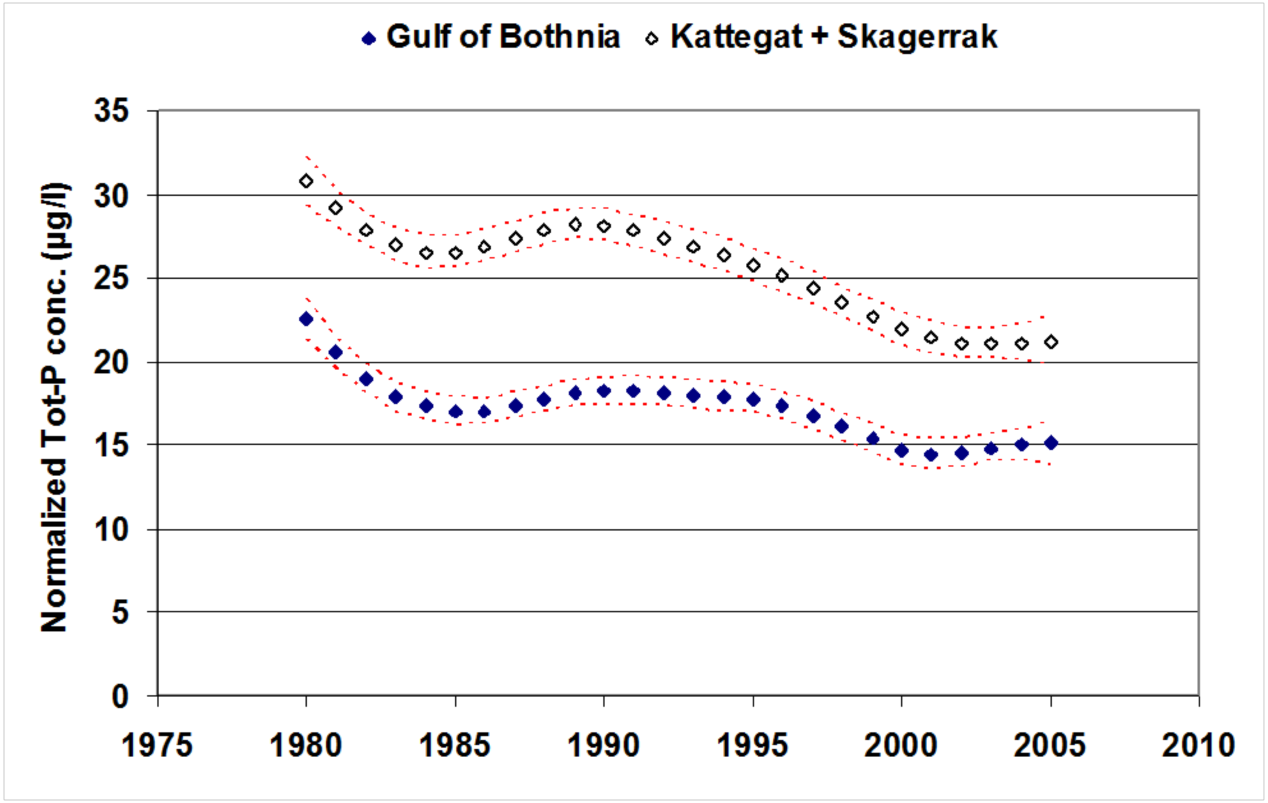
Normalization with respect to discharge and transport of particulate matter



LINKÖPINGS UNIVERSITET

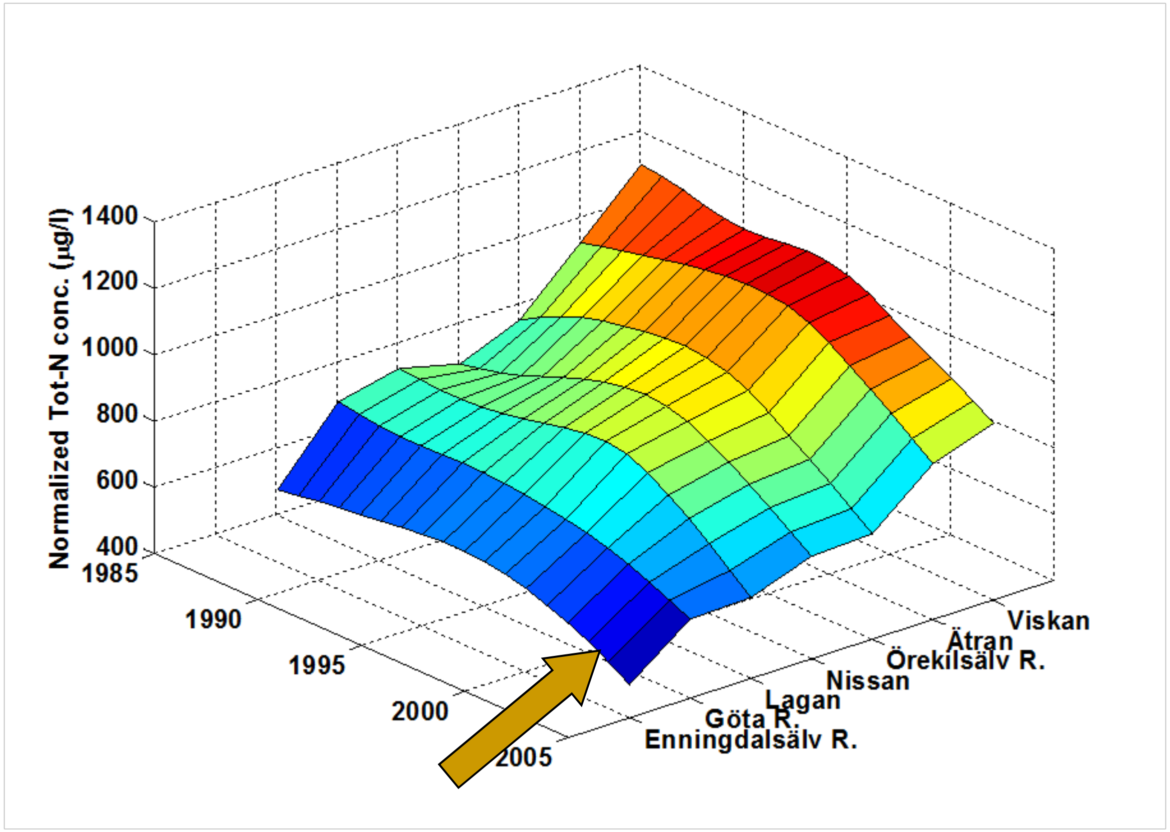
Trend curves for the arithmetic mean of normalized concentrations of phosphorus in riverine input to the Kattegat/Skagerrak and the Gulf of Bothnia

Normalization with respect to discharge and transport of particulate matter



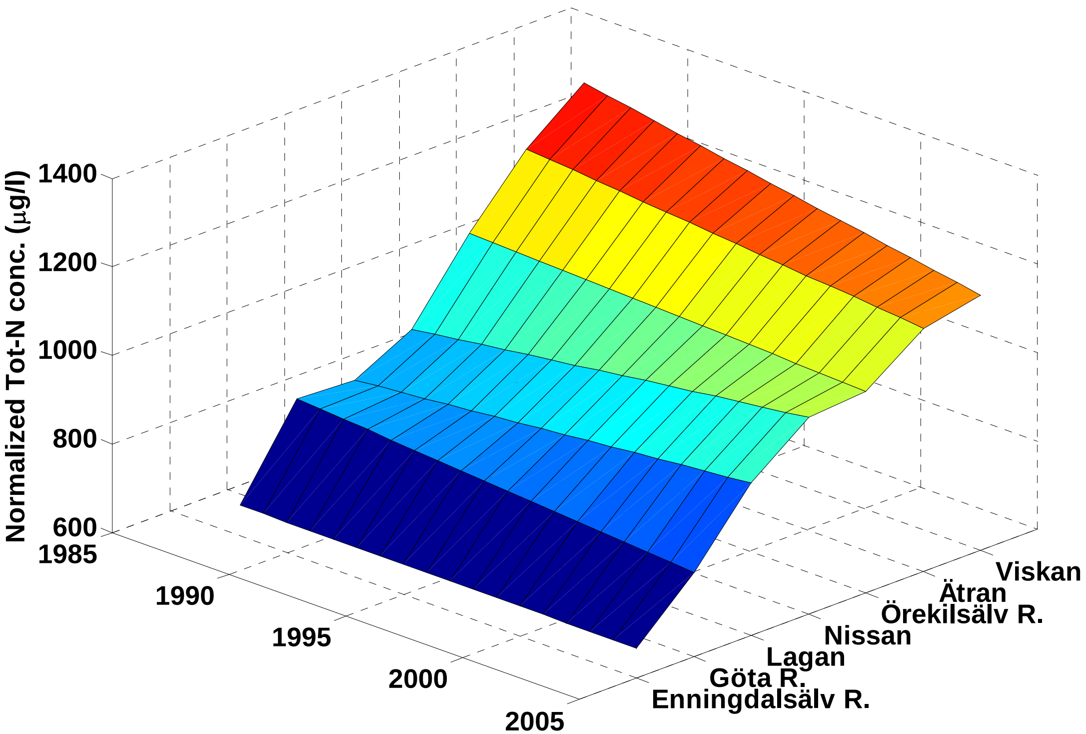
Trend surface for flow-normalized concentrations of total nitrogen in riverine input to the Kattegat and Skagerrak

(Tot-N concentrations determined by persulphate digestion)



Trend surface for flow-normalized concentrations of total nitrogen in riverine input to the Kattegat and Skagerrak

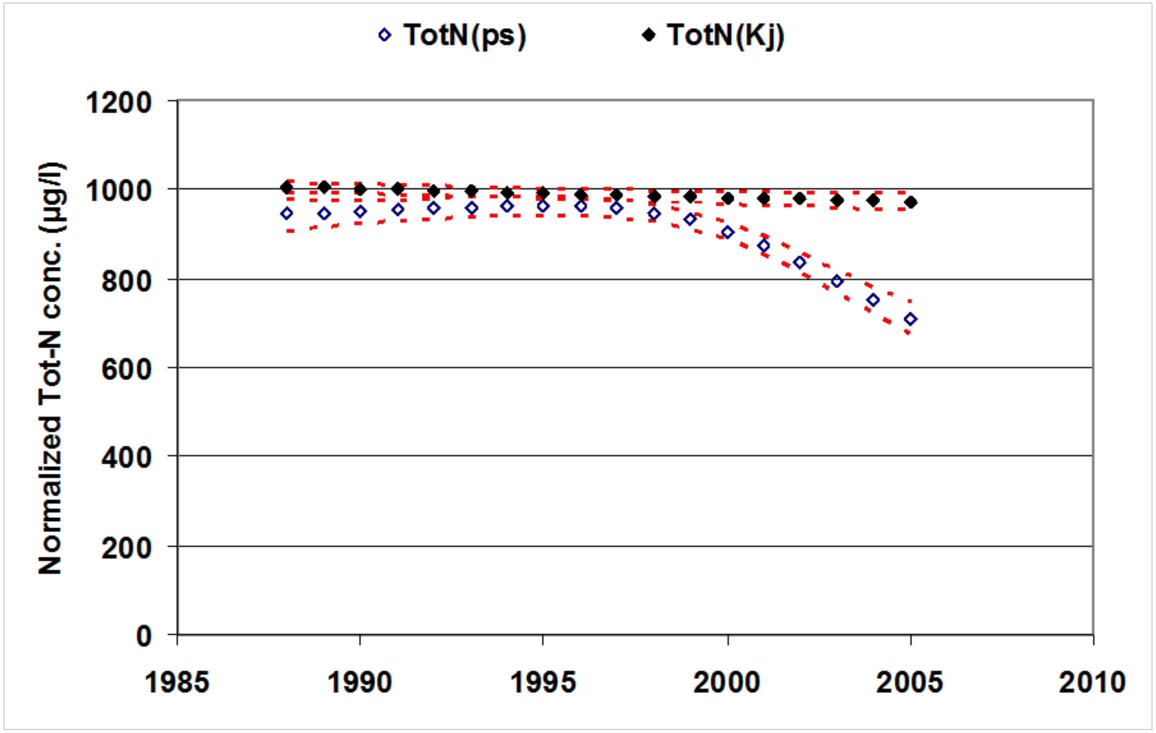
(Tot-N concentrations determined as the sum of nitrite/nitrate nitrogen and Kjeldahl nitrogen)



Trend curves for the arithmetic mean of flow-normalized concentrations of total nitrogen in rivers discharging into the Kattegat and Skagerrak

Tot-N (ps) = total nitrogen determined by persulphate digestion

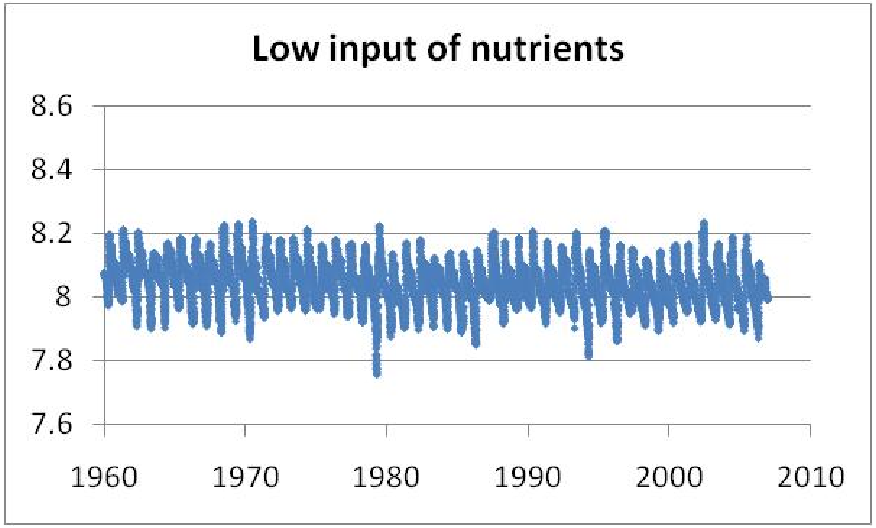
Tot-N (Kj) = sum of nitrite/nitrate nitrogen and Kjeldahl nitrogen

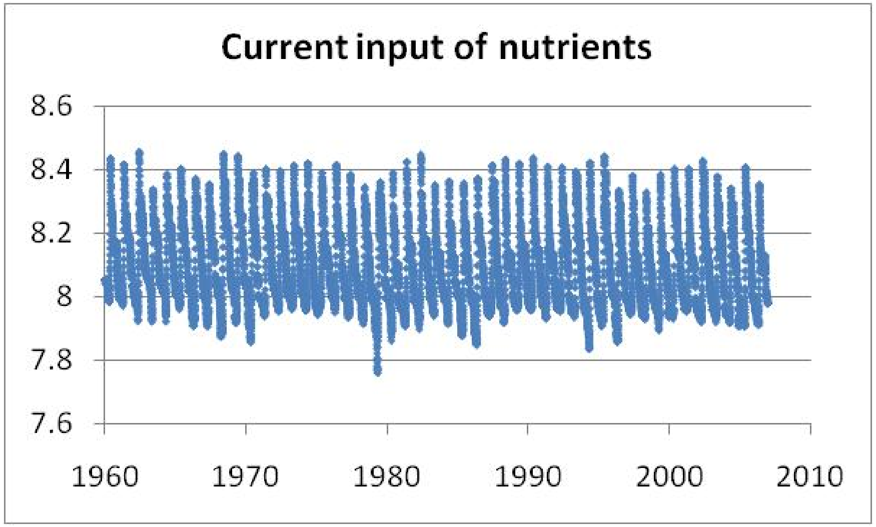


Acidification of the Baltic Sea

- Increased CO_2 levels in the atmosphere can lower the pH of weakly buffered waters
- It has been stated that the pH of sea water has decreased 0.1 units during the past century
- The effect of increasing CO_2 levels can be modified by changes in the input of nutrients to the Baltic Sea

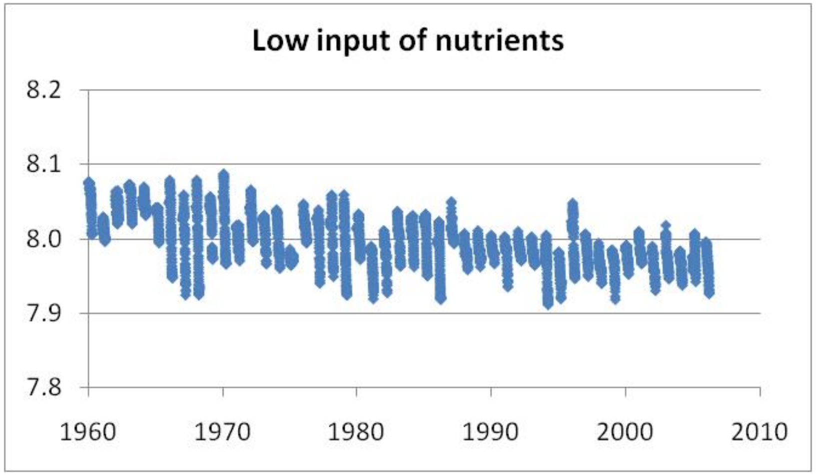
Modelled pH in the Baltic Proper 1960-2006, surface water (0-10 m)

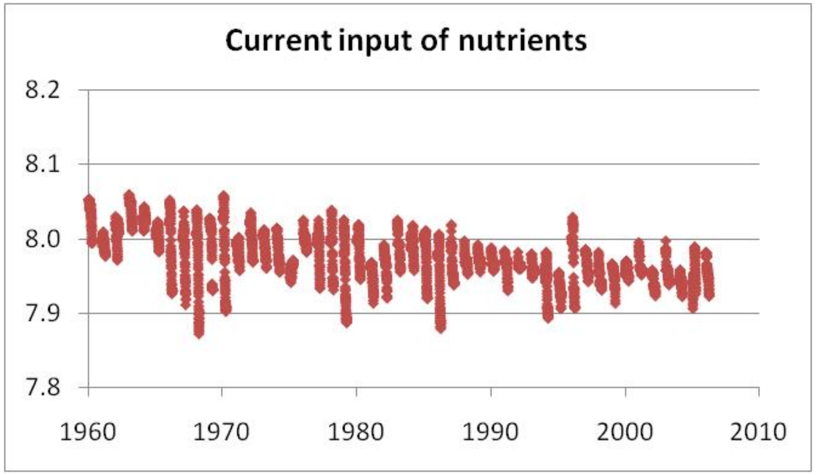


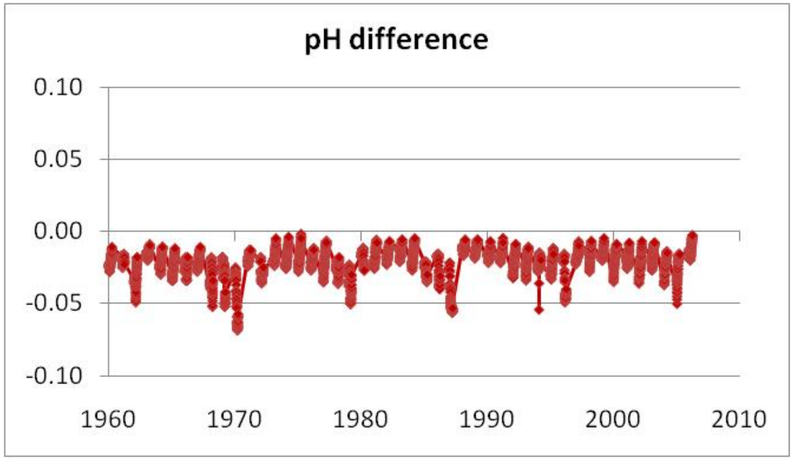


Anders Omstedt, GU

Modelled pH in the Baltic Proper
1960-2006, surface water, winter data (Jan-Mar)







Anders Omstedt, GU

Mann-Kendall tests of trends in modelled pH in the Baltic Proper

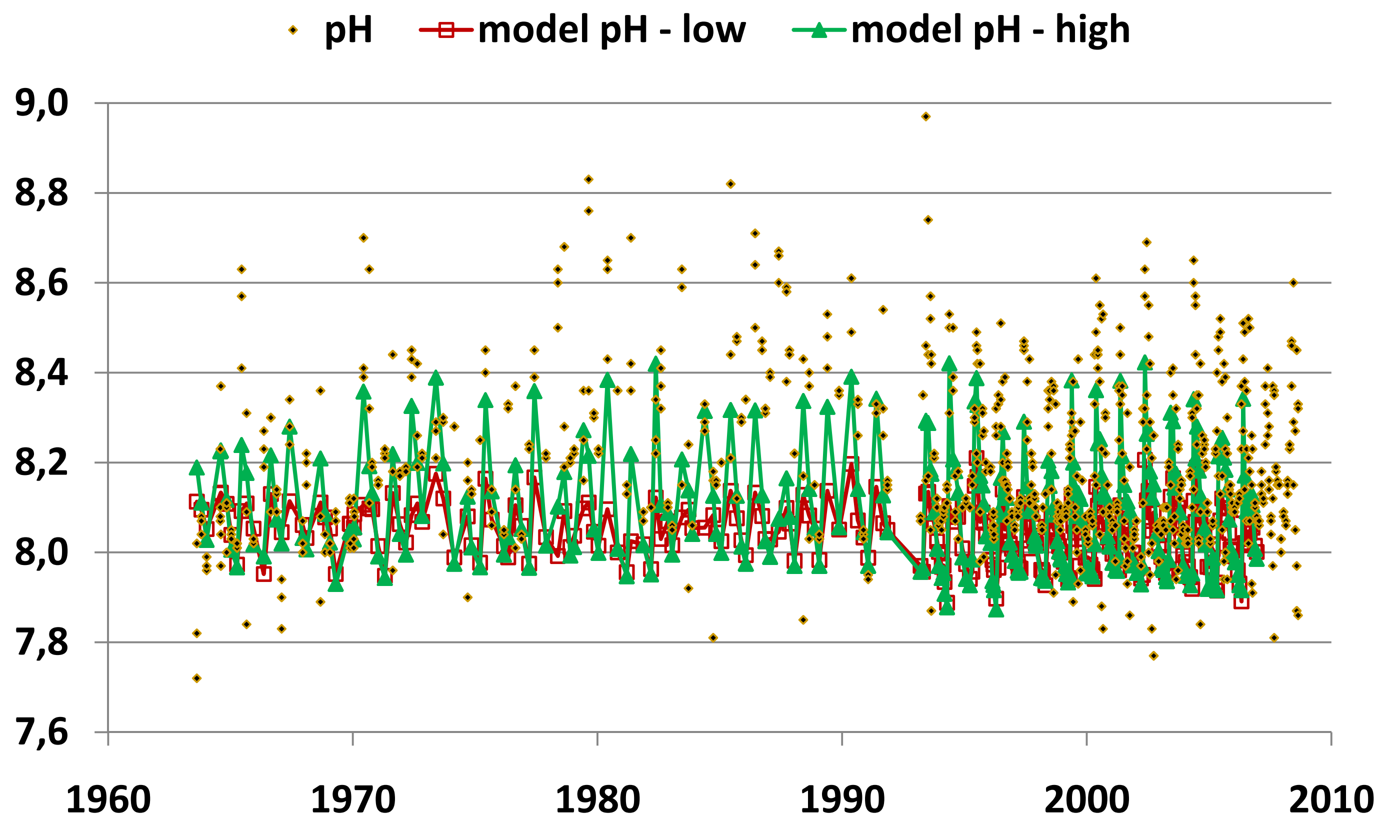
1960-2006

Month	MK statistic	p-value (twosided)	Significance code	Slope (change/yr)	Median
1	-698	0.0000	---	-0.0018	7.992
2	-646	0.0000	---	-0.0015	7.98
3	-337	0.0020	--	-0.0011	7.955
4	-178	0.1026		-0.0006	7.947
5	18	0.8689		0.0001	8.282
6	-293	0.0072	--	-0.0012	8.2545
7	-277	0.0111	-	-0.0016	8.233
8	-395	0.0003	---	-0.0018	8.173
9	-464	0.0000	---	-0.0014	8.153
10	-458	0.0000	---	-0.0013	8.106
11	-448	0.0000	---	-0.0014	8.037
12	-560	0.0000	---	-0.0013	8.002

1990-2006

Month	MK statistic	p-value (twosided)	Significance code	Slope (change/yr)	Median
1	-28	0.2471		-0.0011	7.964
2	-45	0.0631		-0.0010	7.96
3	-15	0.5356		-0.0007	7.948
4	-38	0.1169		-0.0023	7.941
5	-37	0.1272		-0.0048	8.306
6	-38	0.1175		-0.0030	8.241
7	2	0.9343		0.0002	8.233
8	-28	0.2487		-0.0017	8.145
9	-25	0.3027		-0.0011	8.1335
10	3	0.9016		0.0001	8.092
11	-24	0.3228		-0.0012	8.015
12	-33	0.1737		-0.0013	7.985

Comparison of model outputs and observational data 1963-2006



Mann-Kendall tests of trends in observed pH levels at BY15 in the Baltic Proper

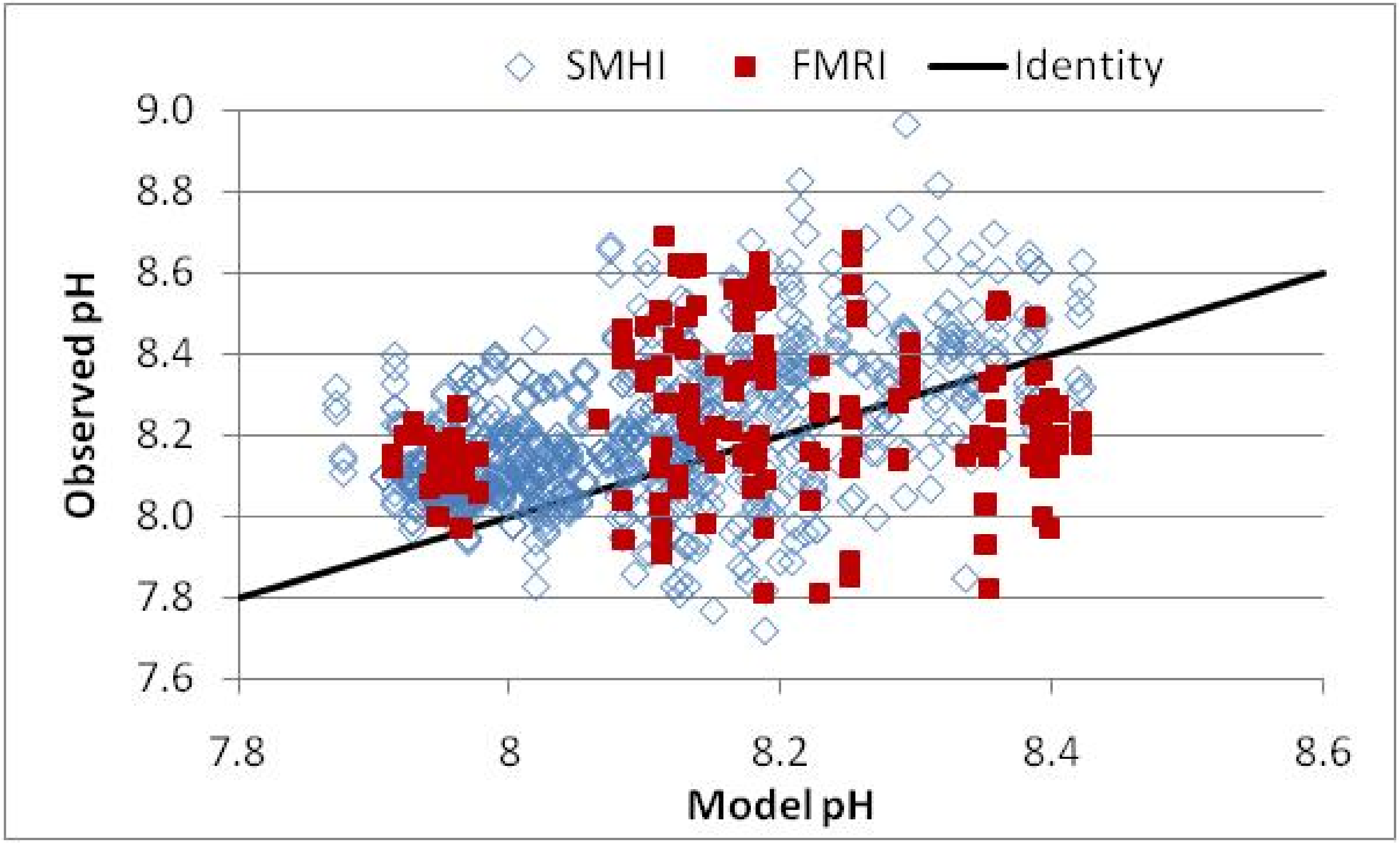
1963-2009

Institute	Month	p-value (twosided)	Significance code	Slope (change/yr)	Median
SMHI	1	0.6345		-0.0006	8.07
SMHI	2	0.5383		-0.0021	8.07
SMHI	3	0.5439		-0.0012	8.1
SMHI	4	0.0384	+	0.0050	8.23
SMHI	5	0.4911		0.0012	8.43
SMHI	6	0.2230		-0.0029	8.4
SMHI	7	0.0102	-	-0.0200	8.225
SMHI	8	0.7817		-0.0007	8.16
SMHI	9	0.6153		-0.0010	8.23
SMHI	10	0.5892		-0.0011	8.19
SMHI	11	0.8003		0.0000	8.13
SMHI	12	0.8561		0.0000	8.07
SMHI	All categ	0.6060		-0.0003	8.17

1990-2009

Institute	Month	p-value (twosided)	Significance code	Slope (change/yr)	Median
SMHI	1	0.9601		0.0000	8.07
SMHI	2	0.7845		0.0000	8.065
SMHI	3	0.1489		0.0036	8.1
SMHI	4	0.4155		0.0050	8.265
SMHI	5	0.8046		-0.0017	8.44
SMHI	6	0.2943		-0.0086	8.35
SMHI	7	0.0102	-	-0.0200	8.225
SMHI	8	0.7720		-0.0008	8.13
SMHI	9	0.3272		-0.0062	8.225
SMHI	10	0.3889		-0.0036	8.19
SMHI	11	0.5536		0.0010	8.12
SMHI	12	0.7578		-0.0008	8.07
SMHI	All cate	0.4165		-0.0011	8.15

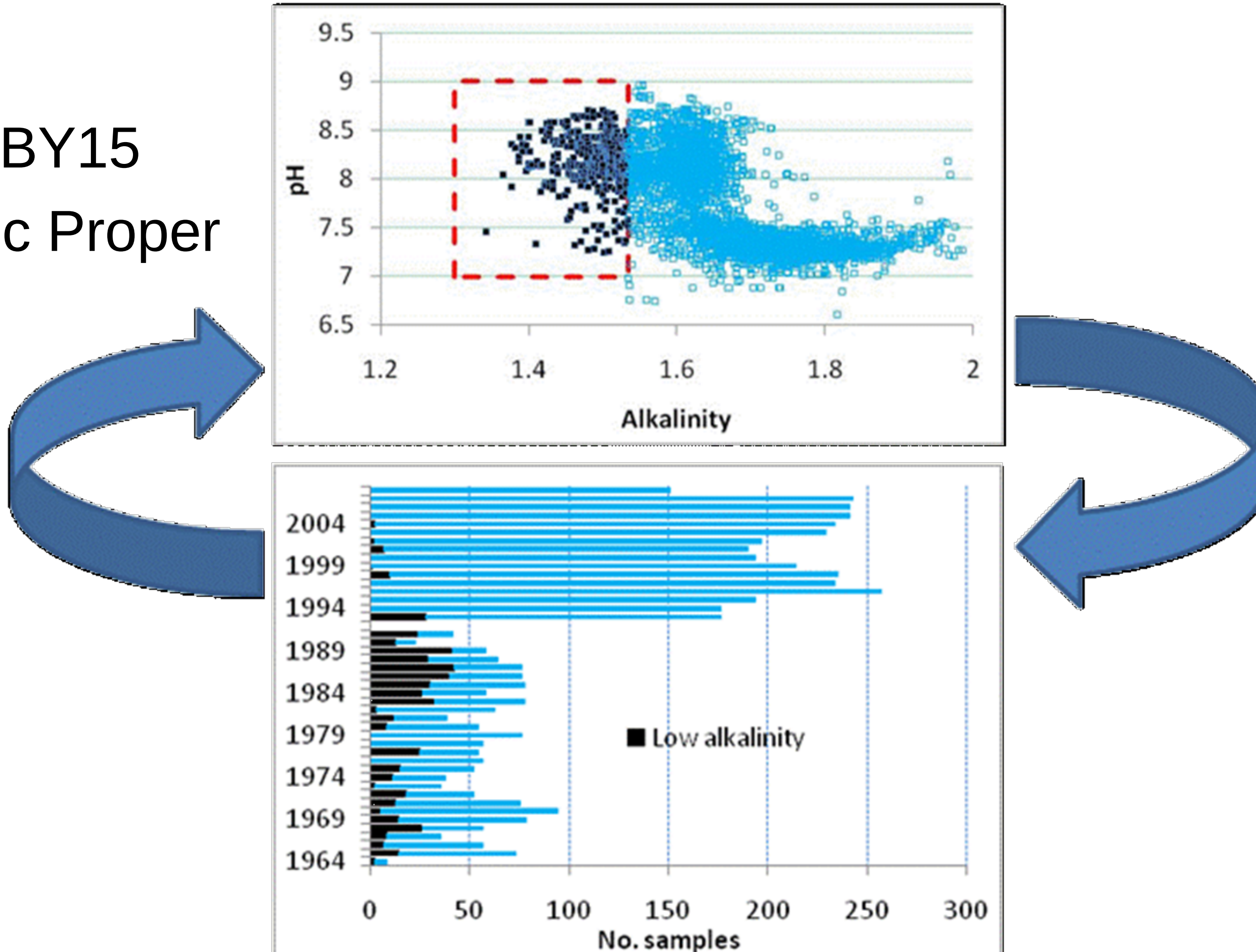
Comparison of observational and modelled pH data – BY15 and the Baltic Proper



Visual inspection of observational data

Data from BY15

In the Baltic Proper



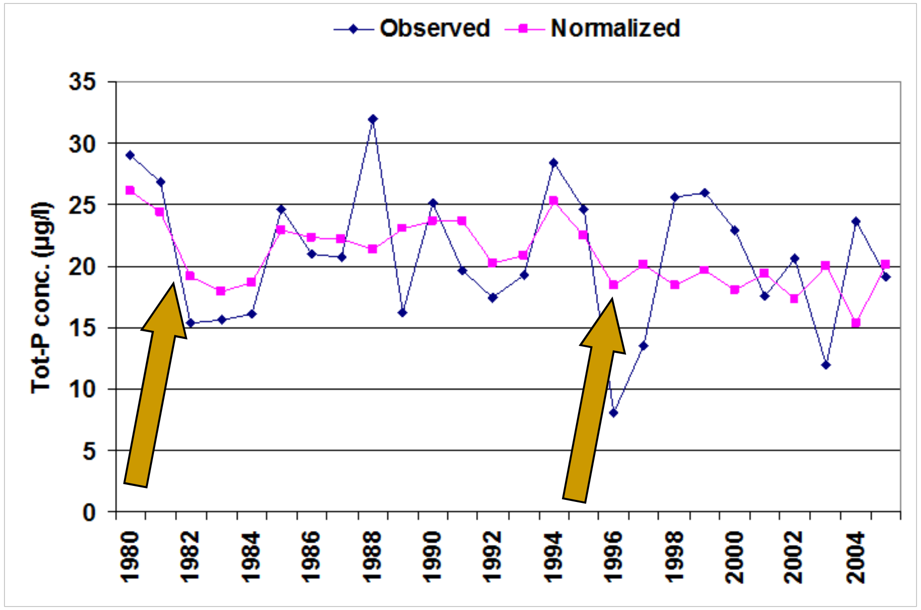
Major achievements in visual data analysis

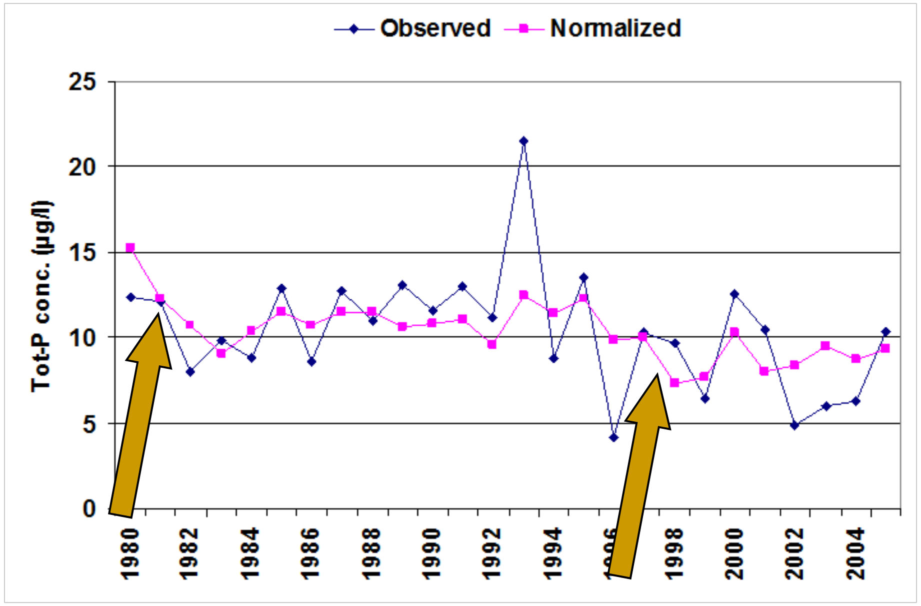
- Interactive linked diagrams (Becker & Cleveland, 1987; www.Ggobi.org)
- Motion charts (Rosling; www.gapminder.org; Google gadget; VB-macros for Excel)

Statistical tools for assessing the quality of environmental data

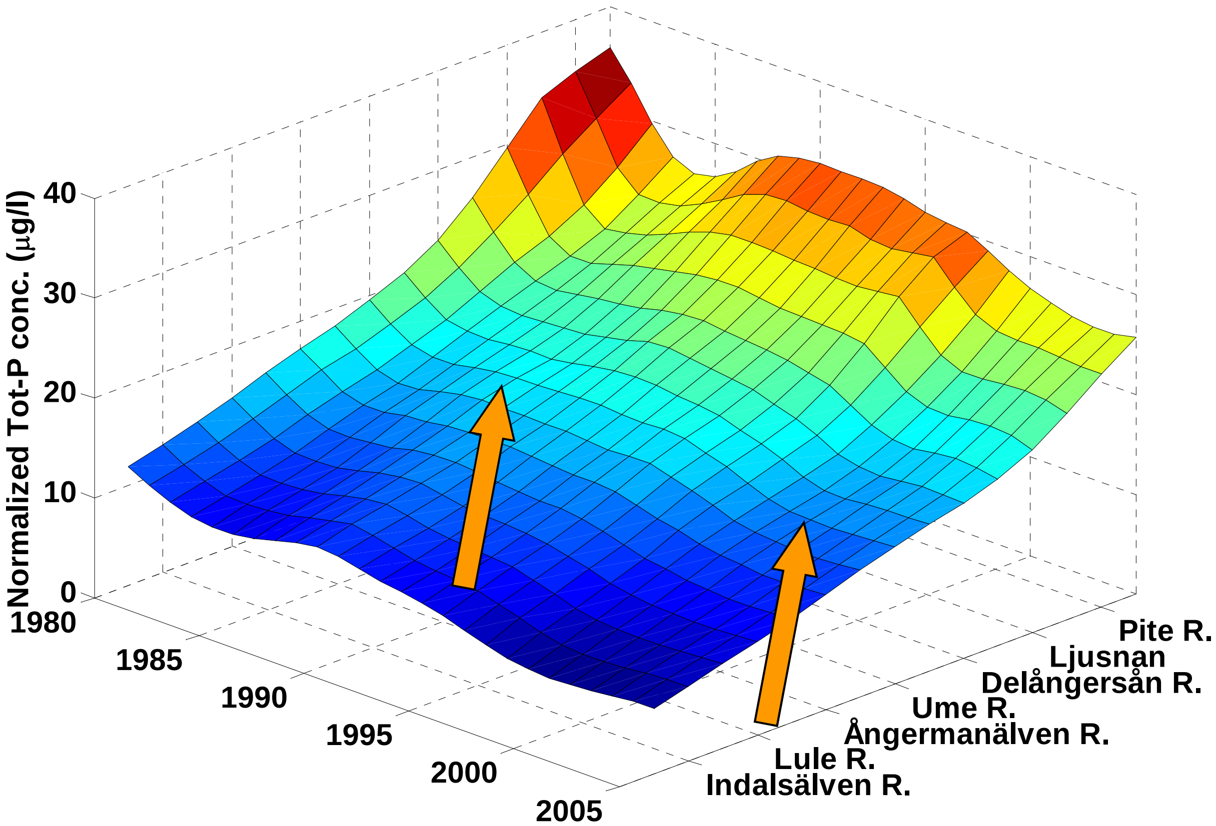
- Trend surface methodologies
- Change-point detection
- Detection of abrupt changes in the presence of smooth trends

Assessing time series of data one by one for the presence of abrupt level shifts

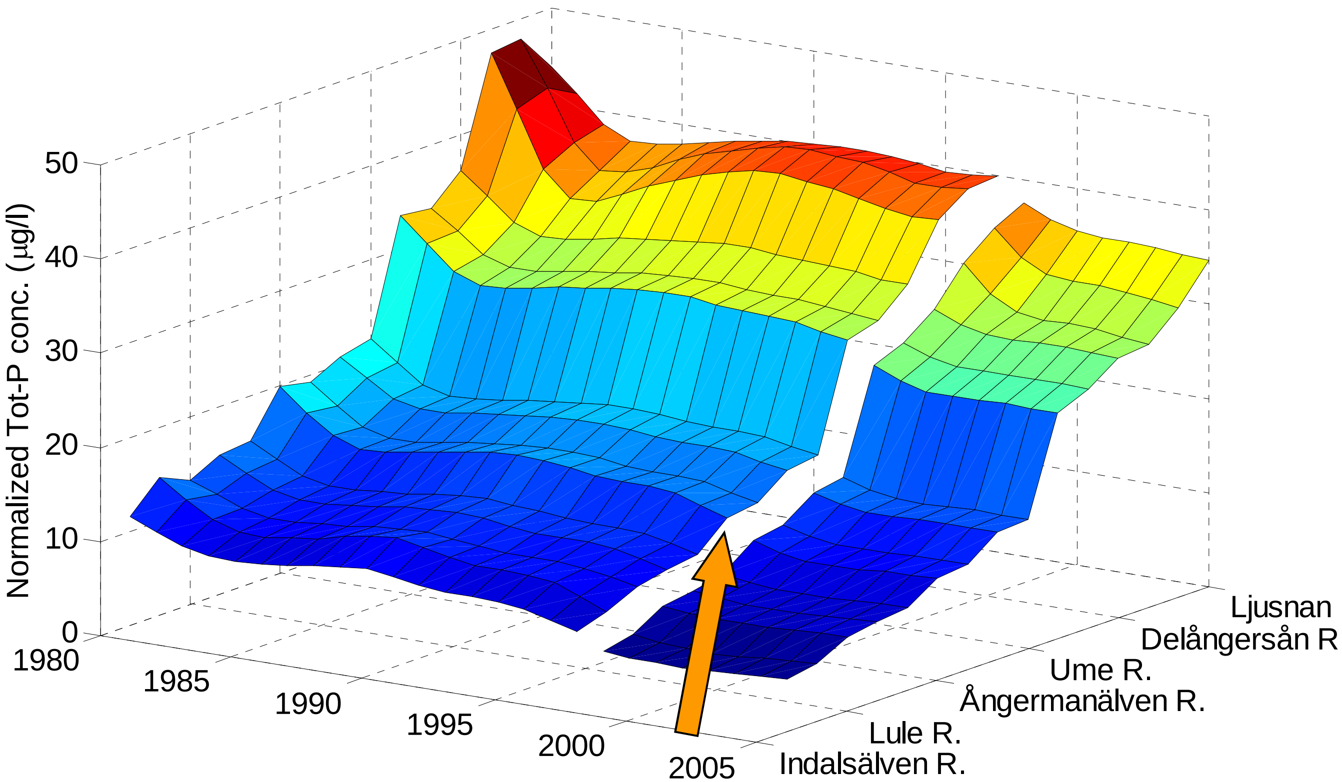




Assessing multiple time series of data for the presence of synchronous smooth changes



Assessing multiple time series of data for the presence of smooth changes and break points



A general semiparametric model for smoothing a vector of time series with similar trends and synchronous change-points

Let y_{tj} be the observed value in the jth series at time t

Assume that

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_t + \sum_{k=1}^{p} (\boldsymbol{x}_{kt} - \overline{\boldsymbol{x}}_{\boldsymbol{k}})\boldsymbol{\beta}_k + \boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n$$

Deterministic ”smooth” trend surface

Impact of covariate

Change-point

Error term

Impose constraints on the variability of the intercepts by penalizing nonlinear variation in these coefficients

Computational aspects

- Use a back-fitting algorithm that alternates between
  i. Estimation of a smooth trend surface in the presence of a given set of change-points
  ii. Detection and estimation of level shifts in the presence of a given smooth response surface

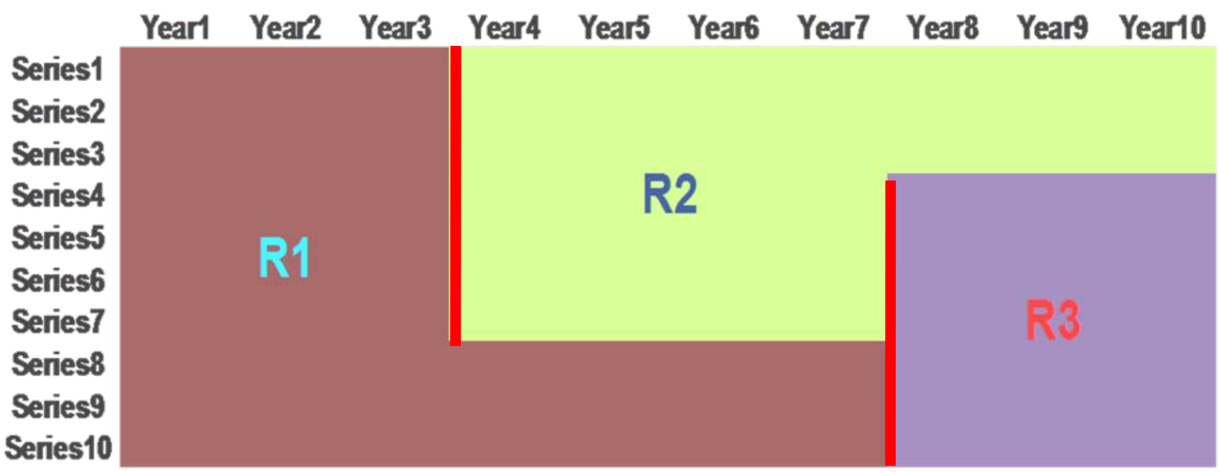
- Estimate the smooth trend using a roughness penalty technique allowing smoothing over time and across series

- Estimate level shifts by first using a tree algorithm to identify candidate change-points and then using dynamic programming for the final detection and estimation of level shifts

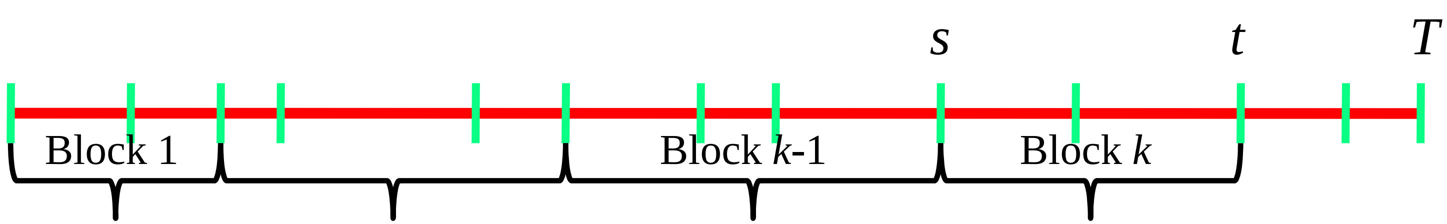
Two-stage search for change-points

- Identification of candidate change-points using a tree algorithm

- Final estimation of change-points (identification of a relevant subset of the candidate changepoints)



Dynamic programming for optimal segmentation of a univariate time series



- **Basic equation for dynamic programming**

$$c(1,t,k) = \min_{0 \le s < t} \{c(1,s,k-1) + h(s+1,t)\},\ 1 \le t \le T,\ 1 \le k \le K$$

Min. cost of joining segments1 *to t* into *k* blocks

Min. cost of joining segments1 *to s* into *k*-1 blocks

Heterogeneity of the last block

- **Determine** $c(1,t,k)$ **recursively for** $1 \le t \le T$ **and** $1 \le k \le K$

Optimal segmentation of a multivariate time series into K blocks

- **Basic equation for dynamic programming**

$$C(1,m,K) = \min_{1 \le k_m \le K-m+1} \{C(1,m-1,K-k_m) + C(m,m,k_m)]$$

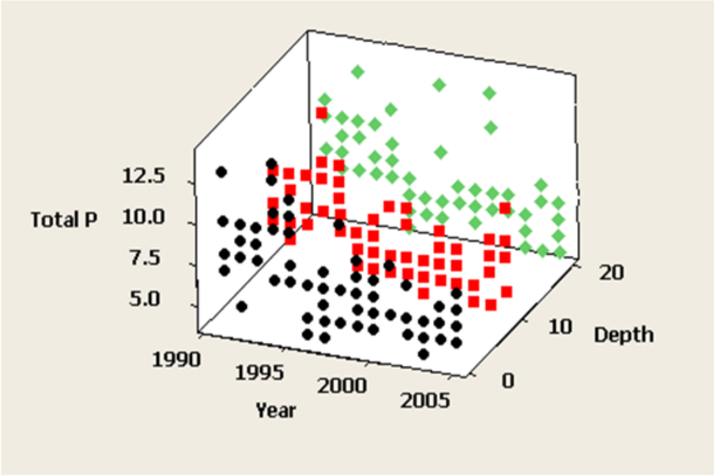
Min. segmentation cost for K blocks in series 1 to m

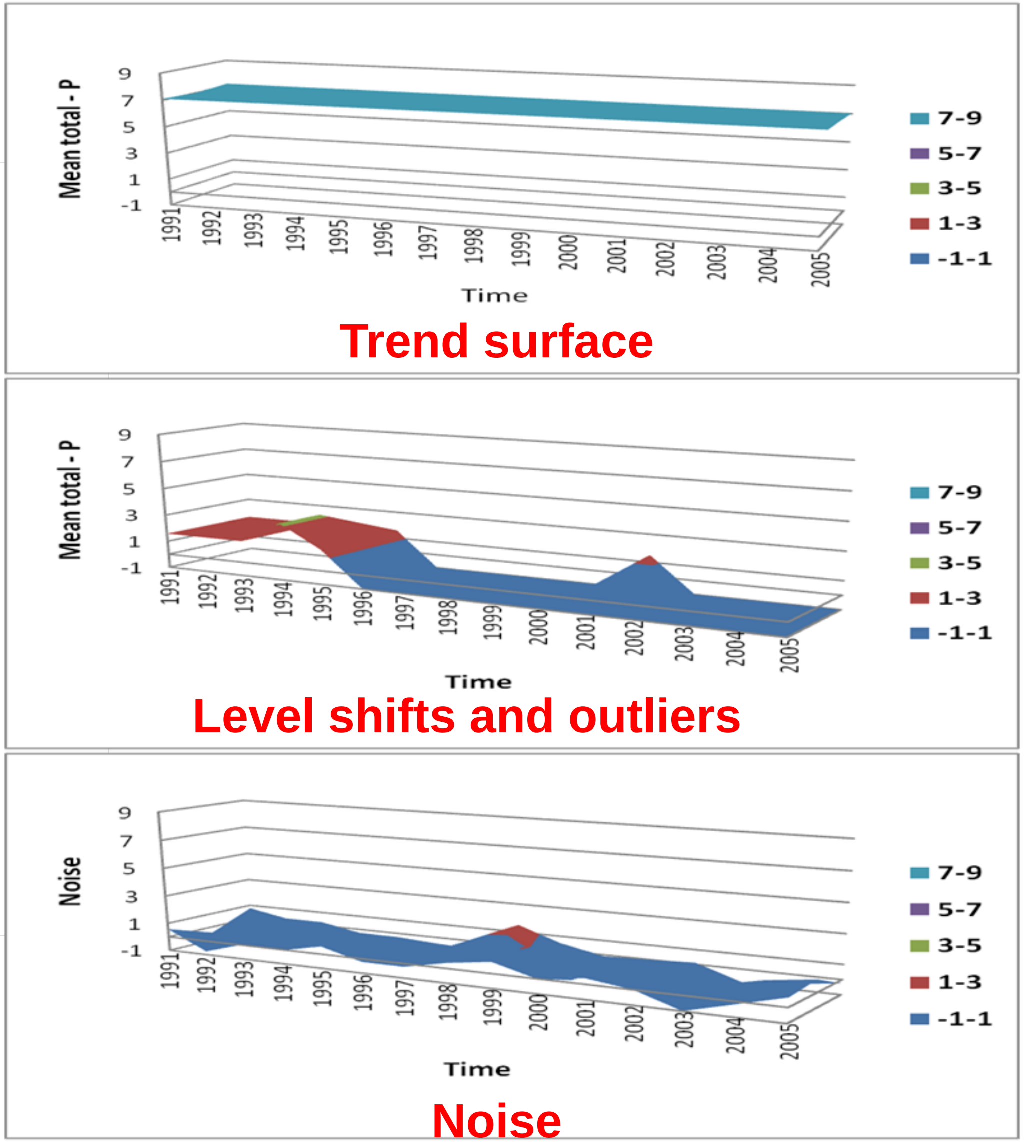
Min. segmentation cost for $K - k_m$ blocks in series 1 to m -1

Min. segmentation cost for k_m blocks in series m

- **Determine** $C(1,m,K)$ **recursively for** $1 \le m \le M$ **and** $1 \le K \le K_{max}$

Inputs and outputs of our integrated algorithm





Concluding remarks

- Interactive visualization techniques provide new indispensable tools for data analysis
- Semiparametric smoothing techniques can provide a useful basis for the analysis of multiple time series of environmental data
- Analysis of large datasets requires efficient algorithms