Estimating artificial level shifts in the presence of smooth trends

K. Wahlin,[1] A. Grimvall,[1] S. Sirisack[2]

[1]Department of Computer and Information Science, Linköping University, SE-58183 Linköping, Sweden

[2]Department of Mathematics, National University of Laos, Vientiane, Laos

e-mail: angri@ida.liu.se

Abstract

Changes in observational data over time can be severely distorted by errors in measurements, sampling, or reporting. Here, we show how smooth trends in vector time series can be separated from one or two abrupt level shifts that occur simultaneously in all coordinates. Trends are modelled nonparametrically, whereas abrupt changes and the impact of covariates are modelled parametrically. The model is estimated using a back-fitting algorithm in which estimation of smooth trends is alternated with estimation of regression coefficients for covariates and assessment of sudden level shifts. The proposed method is adaptive in the sense that the degree of smoothing over time and across coordinates is controlled by a roughness penalty and cross-validation procedure that automatically identifies the interdependence of the analysed data. Furthermore, it uses a resampling technique that can accommodate correlated error terms in the assessment of the uncertainty of both smooth trends and discontinuities. The method is applied to water quality data from Swedish national monitoring programmes to illustrate how known discontinuities can be quantified and how previously unrecognized discontinuities can be detected.

Introduction

The concern about global warming and other long-term changes in the environment has led to increased interest in long time series of environmental data, and there is also a growing awareness of issues related to data quality. A variety of guidance documents for environmental monitoring have been prepared, and substantial efforts have been made to assess and assure the quality of the reported data (e.g., Aguilar *et al*. 2003; Grath *et al.* 2007). Even so, trends in observational data can be severely contaminated by errors in measurements, sampling, or reporting. Compilations of regional and global climate datasets have revealed numerous examples of abrupt changes caused by things like new instrumentation or relocation of sampling sites (e.g., Jones 1995; Klein Tank *et al*. 2002). Our own investigations of air and water quality have indicated that, in such data, artificial level shifts can be a substantial problem even if state-of-the-art quality assurance is applied (Libiseller *et al*. 2005; Wahlin and Grimvall 2008a). Accordingly, there is a strong need for statistical methods that can help detect and estimate discontinuities and other inhomogeneities, and rescue information from old monitoring data.

So far, the most systematic attempts to assess the homogeneity of time series of environmental data have been undertaken by climate scientists. The statistical methods used in that field have their roots in a likelihood ratio test for shift in level at some unknown instant (Hawkins 1977; Worsley 1979), as well as a multivariate extension of that test (Sristava and Worsley 1986). In his pioneering work, Alexandersson (1986) embedded the mentioned type of tests in a procedure in which the climate signal of a candidate series is first removed by subtracting a reference series that is known to be homogeneous. More recent methods aim to detect inhomogeneities without specifying a priori that some series are more reliable than others. Szentimrey (1997) developed algorithms in which each series is compared with an optimally weighted mean of the other series, and Caussinus and Mestre (2004) designed a decision algorithm based on pairwise comparisons of data series from neighbouring sites. Picard and co-workers (2007) showed more generally how numerical algorithms for maximum likelihood estimators in mixed linear models can be employed to simultaneously estimate an arbitrary number of change points in a data matrix.

Outside the climate sector, efforts to detect artificial level shifts in environmental data have been less systematic. We have recently emphasized that joint analysis of multiple time series is needed to efficiently detect and estimate inhomogeneities (Wahlin and Grimvall 2008b). However, the tools developed by climatologists are far from ideal for estimating abrupt level shifts in vector time series in which the coordinates have different trends. Therefore, the aim of the current study was to reduce that deficiency by developing methods for joint analysis of smooth trends and synchronous discontinuities in multiple data time series. In this article, we first describe our model for detection and estimation of inhomogeneities. Thereafter, we show how the model parameters can be estimated, and, finally, we apply our technique to surface and groundwater quality data from Swedish national monitoring programmes.

Models and algorithms

A model class for level shifts in the presence of smooth trends

Let us consider an m-dimensional vector time series

$$\boldsymbol{y}_t = (y_t^{(1)}, ..., y_t^{(m)})^T, \quad t = 1, ..., n$$

representing observations made at m sites at n equidistant time points. Our model can then be written

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

where $\boldsymbol{\alpha}_t$, t = 1, …, n, is a deterministic trend surface, $\boldsymbol{x}_{kt}$, t = 1, …, n, k = 1, …, p, a set of p vector time series of covariates, $\boldsymbol{\gamma}_t$, t = 1, …, n, a sequence of vectors that are stepwise constant in each coordinate, and $\boldsymbol{\varepsilon}_t$, t = 1, …, n, a sequence of random vectors with mean zero and constant covariance matrix. As is customary, the symbols $\bar{\boldsymbol{x}}_{\boldsymbol{k}}$, k = 1, …, p represent sample means, and

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_p^T) = (\beta_1^{(1)}, ..., \beta_1^{(m)}, ..., \beta_p^{(1)}, ..., \beta_p^{(m)})^T$$

denotes a time-independent vector of regression coefficients.

The sequence γ_t, t = 1, …, n, can be parameterized in different manners in different applications. If observed data have a single change point common to all coordinates, we can set

$$\gamma_t^{(j)} = \begin{cases} \mu, \text{if } t \le t_1 \\ \mu + \theta(j), \text{if } t > t_1 \end{cases}$$

where $\theta(j)$ denotes the level shift in the jth coordinate between time t_1 and t_1+1. The parameter μ, which is unidentifiable in the presence of the vectors α_t, t = 1, …, n, is normally selected so that the sum of all $\gamma_t^{(j)}$ is zero. Furthermore, we can introduce simple parametric forms of the level shifts, such as

$$\theta(j) = \theta_0, \quad j = 1, ..., m$$

or

$$\theta(j) = \theta_0 + \theta_1 j, \quad j = 1, ..., m$$

Functions with two or more change points are defined analogously. If it is suspected that the measured data have been biased during a certain time period, it may be of interest to use the following parameterization:

$$\gamma_t^{(j)} = \begin{cases} \mu, \text{if } t \le t_1 \text{ or } t > t_2 \\ \mu + \theta(j), \text{if } t_1 < t \le t_2 \end{cases}$$

Expressions of the form

$$\gamma_t^{(j)} = \begin{cases} \mu, \text{if } t \le t_1 \\ \mu + \delta\,\theta(j), \text{if } t = t_1 + 1 \\ \mu + \theta(j), \text{if } t > t_1 + 1 \end{cases}$$

where $0 < \delta < 1$ can be useful if a level shift takes place in two consecutive steps.

Point estimation of parameters and selection of smoothing factors

Because the models presented above are over-parameterized, it is necessary to introduce some constraints or regularization functions when the parameters are estimated.

Following the ideas previously outlined by our group (Grimvall *et al*. 2008), we used a penalized least squares technique in which we minimized an expression of the form

$$S(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \sum_{t=1}^{n}\sum_{j=1}^{m}(y_t^{(j)} - \hat{y}_t^{(j)})^2 + \lambda_1 L_1(W_1, \boldsymbol{\alpha}) + \lambda_2 L_2(W_2, \boldsymbol{\alpha})$$

where

$$L_1(W_1, \boldsymbol{\alpha}) = \sum_{(t_1, j_1, t_2, j_2, t_3, j_3) \in W_1} \left(\alpha_{t_1}^{(j_1)} - \frac{\alpha_{t_2}^{(j_2)} + \alpha_{t_3}^{(j_3)}}{2} \right)^2$$

$$L_2(W_2, \boldsymbol{\alpha}) = \sum_{(t_1, j_1, t_2, j_2, t_3, j_3) \in W_2} \left(\alpha_{t_1}^{(j_1)} - \frac{\alpha_{t_2}^{(j_2)} + \alpha_{t_3}^{(j_3)}}{2} \right)^2$$

and $\hat{y}_t^{(j)}$ denotes a prediction of $y_t^{(j)}$ based on data for all time points $s \neq t$. The vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ contains two nonnegative factors controlling the smoothness of the trend surface, and W_1 and W_2 denote two different smoothing patterns. Normally, W_1 is set to

$$W_1 = \{(t,\ j, t-1,\ j, t+1,\ j), \quad t = 2, ..., n-1 \quad j = 1, ..., m \quad \}$$

in order to generate horizontal smoothing (smoothing over time), whereas W_2 is used to impose a vertical smoothing pattern (smoothing across coordinates). However, the model can accommodate any user-defined smoothing patterns W_1 and W_2 (Grimvall *et al*. 2008). If both λ_1 and λ_2 are large, the fitted smooth trend surface will be almost a plane in $\mathbf{R}^3$. If both λ_1 and λ_2 are small, the smoothing of observed values will be practically negligible.

We achieved global minimization of $S(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ by systematically searching for λ-values that made the prediction error sum of squares (*press*) as small as possible when we left out one year of observations at a time, estimated the model parameters using the remaining data, and summed the squared prediction errors for the observations that were left out. Furthermore, it can be noted that the minimization of $S(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ for given smoothing factors was accomplished by employing a back-fitting algorithm alternating between estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, respectively. Written as pseudocode, our algorithm to determine *press* for given λ-values had the following structure:

Back-fitting algorithm for joint estimation of smooth trends and discontinuities

---

1. Initialize $\boldsymbol{\gamma} \equiv 0$
2. Initialize $\boldsymbol{\beta}$ by using a multiple linear regression model with intercept to regress $\boldsymbol{y}$ on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$
3. Initilize *press* = 0
4. Initialize $s = 1$
5. Repeat
6. $T = \{1, \ldots, s\text{-}1, s\text{+}1, \ldots, n\}$

   Cycle

   $$u_t^{(j)} = y_t^{(j)} - \sum_{k=1}^{p} \beta_k^{(j)} (x_{kt}^{(j)} - \bar{x}_k^{(j)}) - \gamma_k^{(j)}, \quad t \in T, \quad j = 1, \ldots, m$$

   $$\arg\min_{\boldsymbol{\alpha}} \left[\sum_{t \in T} \sum_{j=1}^{m} (u_t^{(j)} - \alpha_t^{(j)})^2 + \lambda_1 L_1(W_1, \boldsymbol{\alpha}) + \lambda_2 L_2(W_2, \boldsymbol{\alpha}) \right]$$

   $$u_t^{(j)} = y_t^{(j)} - \alpha_t^{(j)} - \gamma_t^{(j)}, \quad t \in T, \quad j = 1, \ldots, m$$

   $$\arg\min_{\boldsymbol{\beta}} \left[\sum_{t \in T} \sum_{j=1}^{m} (u_t^{(j)} - \beta_0 - \sum_{k=1}^{p} \beta_k^{(j)} (x_{kt}^{(j)} - \bar{x}_k^{(j)}))^2 \right]$$

   $$u_t^{(j)} = y_t^{(j)} - \alpha_t^{(j)} - \sum_{k=1}^{p} \beta_k^{(j)} (x_{kt}^{(j)} - \bar{x}_k^{(j)}), \quad t \in T, \quad j = 1, \ldots, m$$

   $$\arg\min_{\boldsymbol{\gamma}} \left[\sum_{t \in T} \sum_{j=1}^{m} (u_t^{(j)} - \gamma_t^{(j)})^2 \right]$$

   $$\gamma_t^{(j)} \leftarrow \gamma_t^{(j)} - mean(\sum_{t \in T} \sum_{j=1}^{m} \gamma_t^{(j)})$$

   until the relative change in the penalized sum of squares on T is below a pre-specified threshold

   $$press \leftarrow press + \sum_{j=1}^{m} (y_s^{(j)} - \alpha_s^{(j)} - \sum_{k=1}^{p} \beta_k^{(j)} (x_{ks}^{(j)} - \bar{x}_k^{(j)}) - \gamma_s^{(j)})^2$$

   $s \leftarrow s+1$
7. until $s = n$

If the number of observations varies from cell to cell in the matrix defined by time and site, it is easy to modify the formulae indicated above. Further details are given elsewhere (Grimvall *et al.* 2008).

Uncertainty assessment

The uncertainty of the fitted trend surface and the estimated level shifts was assessed using a residual resampling technique introduced by Grimvall *et al.* (2008). As in ordinary residual resampling in regression models with non-random design, the covariates $\boldsymbol{x}_{kt}$, $t = 1, \ldots, n$, $k = 1, \ldots, p$, were kept fixed, and new response values were generated by setting

$$y_t^{*(j)} = y_t^{(j)} - e_t^{(j)} + e_t^{*(j)}, \quad t = 1, \ldots, n \quad j = 1, \ldots, m$$

where $e_t^{*(j)}$ denotes a resampled residual, and the same symbol without an asterisk denotes the original residual (Mammen 2000). However, after selecting new residuals by sampling with replacement, pairs of resampled residuals were swapped until the correlation pattern was similar to that of the original residuals. Further details are available elsewhere (Grimvall *et al*. 2008).

Computational aspects

The algorithm presented above has been implemented as a VisualBasic macro for Excel (LiU 2008). Extensive experiments in which the macro was tested on simulated data with known level shifts showed that the back-fitting invariably converged to solutions that were coherent with the true data model. In addition, our runs with real water quality data did not reveal any convergence problems. The computational burden varied strongly with the mode in which the algorithm was run. Fitting a model with given smoothing factors and without resampling for uncertainty assessments took less than a second for the datasets presented in this article, and this was achieved mainly by exploiting the band matrix structure of the system of *mn* linear equations providing estimates of the α-parameters (Hussian *et al*. 2004; Stålnacke and Grimvall 2001). Moreover, the cross-validation was non-problematic, because the exact levels of the selected smoothing factors did not have a significant impact on the results obtained. The computational effort in the resampling is more substantial. When we generated 200 sample replicates and allowed 100,000 residual swaps for each replicate, the computational time varied from less than a minute to almost an hour when our datasets were analysed on a standard PC.

However, if the error components are only weakly correlated, the number of swaps, and hence also the total computational time, can be substantially reduced.

Case studies of change-point detection

Observational data

We tested the methods and algorithms on surface and groundwater data from national monitoring programmes conducted in Sweden. The surface water data represented one site in Lake Vänern (Dagskärsgrundet) and samples collected close to the mouths of fifteen major rivers in the northern part of the country (Table 1). The statistical analysis focused on total phosphorus, TOC (total organic carbon), and COD (chemical oxygen demand) measured as permanganate consumption. Datasets and further information can be obtained from the Swedish University of Agricultural Sciences (SLU 2008).

Table 1. The investigated rivers and their recipients: the Bothnian Sea (BS) and the Bothnian Bay (BB)

River	Recipient	River	Recipient
Torne	BB	Ångermanälven	BS
Kalix	BB	Indalsälven	BS
Råne	BB	Ljungan	BS
Lule	BB	Delångersån	BS
Pite	BB	Ljusnan	BS
Ume	BS	Gavleån	BS
Öre	BS	Dalälven	BS
Gide	BS		

The groundwater represented a total of 77 sites. Special attention was paid to reported levels of potassium and alkalinity, and records of acid neutralizing capacity (ANC) computed according to

$$ANC = \left[Ca^{2+}\right] + \left[Mg^{2+}\right] + \left[Na^{+}\right] + \left[K^{+}\right] + \left[NH_4^{+}\right] - \left[Cl^{-}\right] - \left[SO_4^{2-}\right] - \left[NO_3^{-}\right]$$

Datasets and further information about the monitoring programme can be obtained from the Geological Survey of Sweden (SGU 2008).

Level shifts at known instants

Visual inspection of Figure 1 indicates that a level shift in the reported phosphorus concentrations took place in 1996, after the procedure to correct for the blank level of the chemical analysis was altered. As expected, it was impossible to achieve a good fit to this dataset when our model was run with large smoothing factors and without any discontinuities.

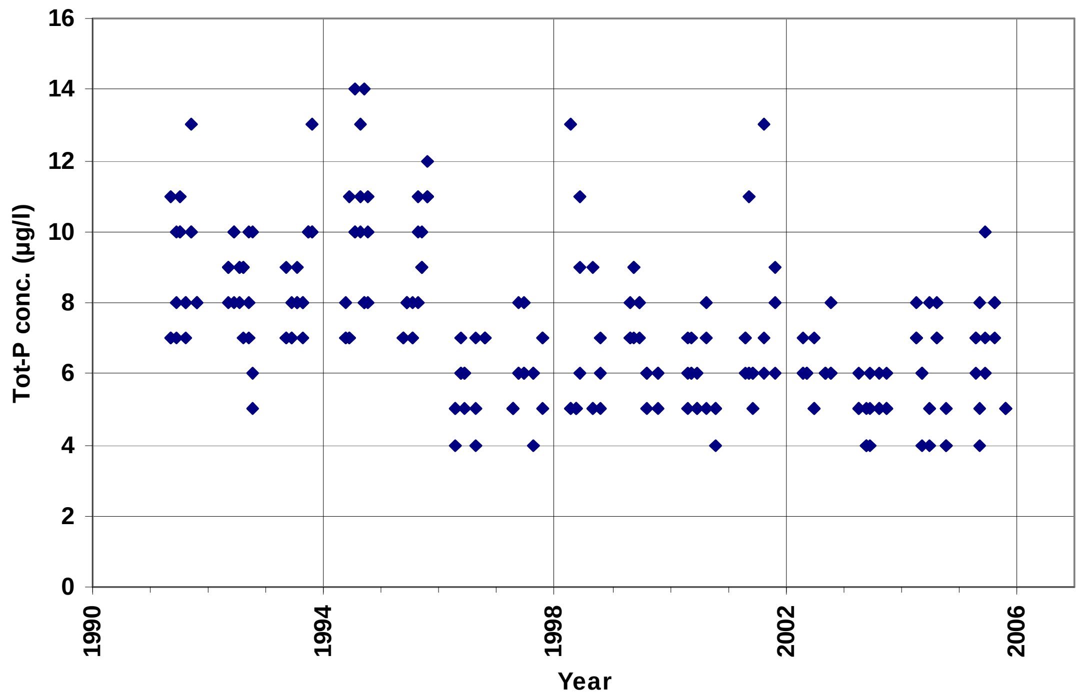

Figure 1. Total phosphorus (Tot-P) levels in surface water at Dagskärsgrund in Lake Vänern, 1991–2005. Samples were collected on 4–6 occasions per year from April to October at depths of 0.5, 10, and 20 m.

This was also the case when water temperature was incorporated as a covariate, and Figure 2 illustrates the rather rough trend surface that was selected by our algorithm. We subsequently augmented our model with a discontinuity between 1995 and 1996, and the level shift was assumed to be of the same size at all sampling depths. This modification substantially improved the fit to reported data. Moreover, the cross-validation then indicated that the highest predictivity of the model was obtained for large values of the smoothing factors (Fig. 3). The size of the discontinuity was estimated to 3.1 µg/l, and residual resampling showed that the standard error of the estimated level shift was considerably smaller (0.44 µg/l).

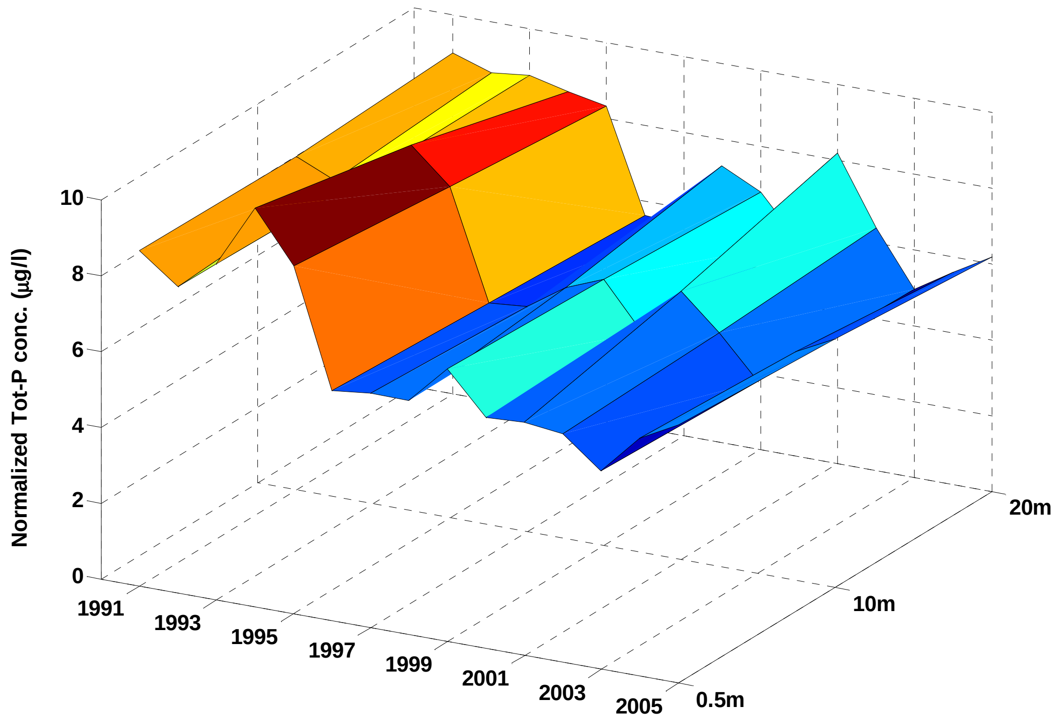

Figure 2. Trend surface without discontinuities for the total phosphorus (Tot-P) levels shown in Figure 1. Cross-validation indicated that the optimal smoothing factors were λ_1 = 0.16 and λ_2 = 32.

In a recent study (Wahlin and Grimvall 2008a), we found that there were also abrupt level shifts in other phosphorus records from the same laboratory. Figure 4 shows the measured concentrations of phosphorus in fifteen major rivers in northern Sweden. When we reanalysed that dataset using the algorithm presented here, we found that the level shift in 1996 was statistically significant, and that discontinuity emerged even more clearly when the analysis was restricted to the four rivers with the lowest frequency of outliers. Figure 5 shows the fitted trend surface with the estimated discontinuity. Inasmuch as the change in the laboratory practice took place in the middle of 1996, we used a model in which the discontinuity was split between two consecutive years. Furthermore, we used water discharge as a covariate and allowed the size of the discontinuity to vary with the average phosphorus concentration in the analysed river. Table 2 illustrates the estimated level shifts and their standard errors. In particular, it can be noted that level shifts also occurred in rivers where measured phosphorus concentrations were far above the detection limit of the analytical procedure employed.

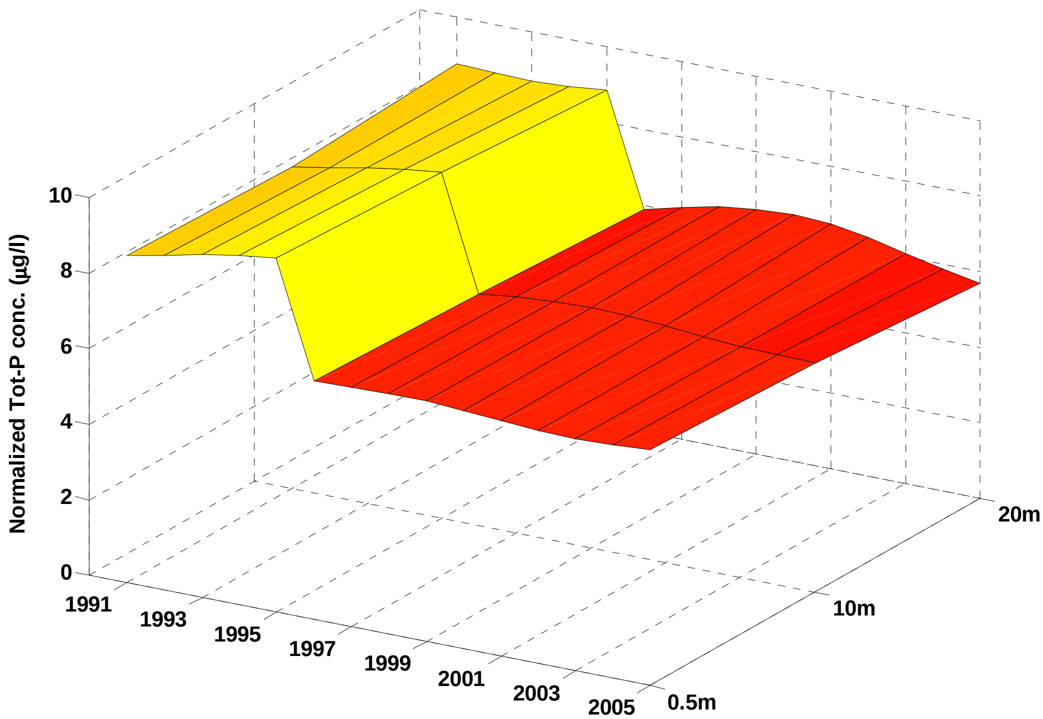

Figure 3. Smooth trend surface augmented with a discontinuity between 1995 and 1996. The underlying data were the same as in Figures 1 and 2, and cross-validation indicated that the optimal smoothing factors were λ_1 = 10240 and λ_2 = 16.

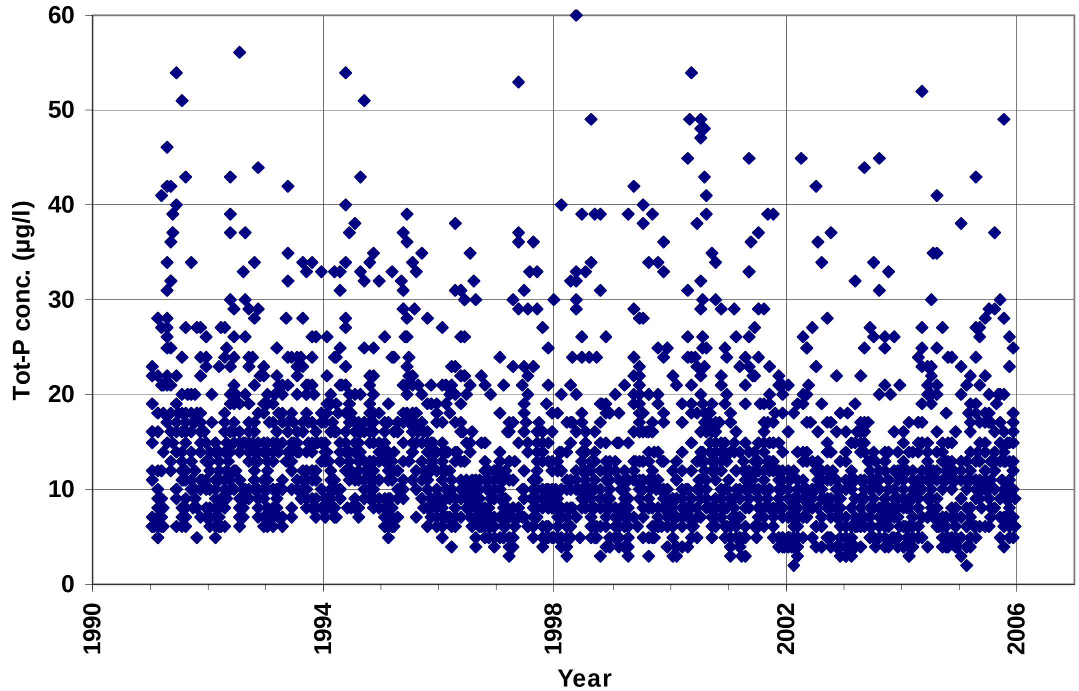

Figure 4. Total phosphorus (Tot-P) levels recorded at the mouths of fifteen major rivers in northern Sweden. Monthly sampling was done in all rivers throughout the investigated period.

Table 2. Estimated level shifts in total phosphorus data from four rivers in northern Sweden. The model had level shifts that were equally split between 1995–96 and 1996–97, and the size of the shifts was allowed to vary with the sampled river

River	Level shift (µg/l)	Standard error (µg/l)
Indalsälven	–2.90927	1.130746
Råne	–2.61134	0.774648
Dalälven	–3.26740	1.115243
Gide	–2.89887	1.348316
Average	–2.92172	0.875484

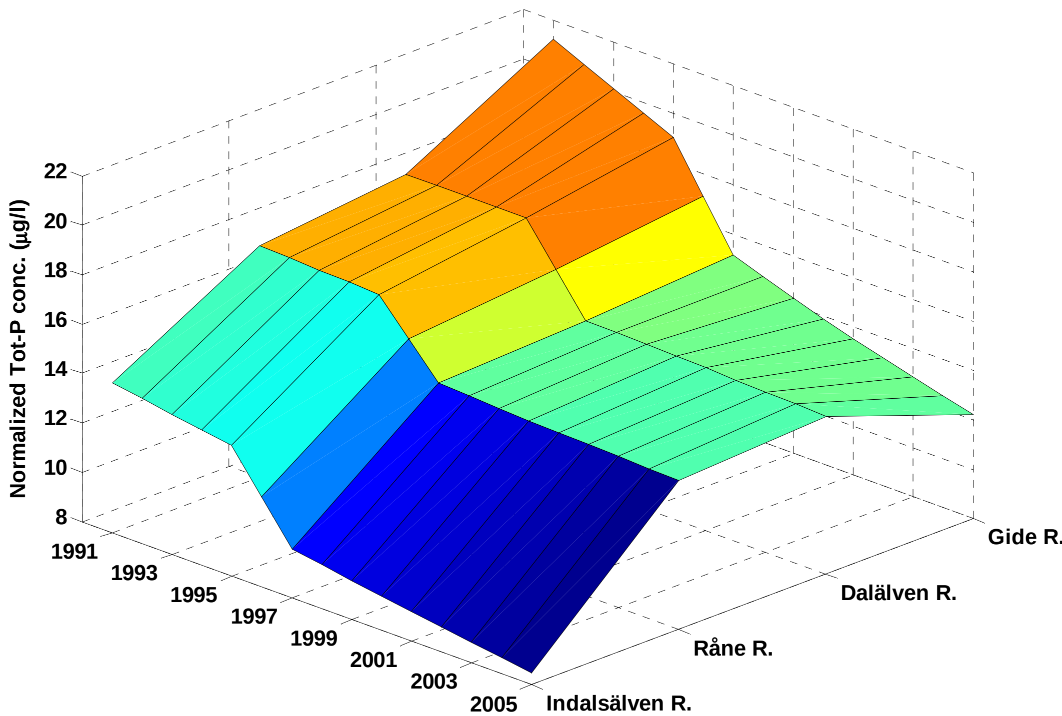

Figure 5. Trend surface with discontinuities fitted to total phosphorus (Tot-P) concentrations in four major rivers in northern Sweden. The statistical model and the sampled rivers were the same as in Table 2.

Level shifts at unknown instants

The response surfaces in Figures 3 and 5 were obtained with change points specified by the user. However, the results were identical when the algorithm was run with an unprejudiced search for level shifts, and that outcome was expected because the abrupt changes were quite evident. Figure 6 illustrates a dataset in which the presence and location of discontinuities is less obvious.

Since the measured potassium levels varied strongly between sampling occasions, and the potential discontinuities were relatively small, we focused our study on average level shifts. Figure 7 illustrates the annual means of the estimated trend levels when the model contained a discontinuity that was equally split between two consecutive years. The thick solid line with attached error margins (± 2 standard errors) contained two consecutive level shifts specified by the user to occur in 1990–1992, because the analytical procedure was altered in the middle of 1991. The thick dashed line was obtained in a purely data-driven search for the most significant discontinuity in the investigated time interval. As can be seen, these two curves differ slightly with respect to the timing of the discontinuity, whereas the size of the level shifts was practically the same in the two model runs. This was expected, considering that the timing can be strongly influenced by a relatively small number of observations that are temporally close to the true change point. The size of the level shift is less sensitive to small subsets of observations, provided that the smoothing factors are not too small.

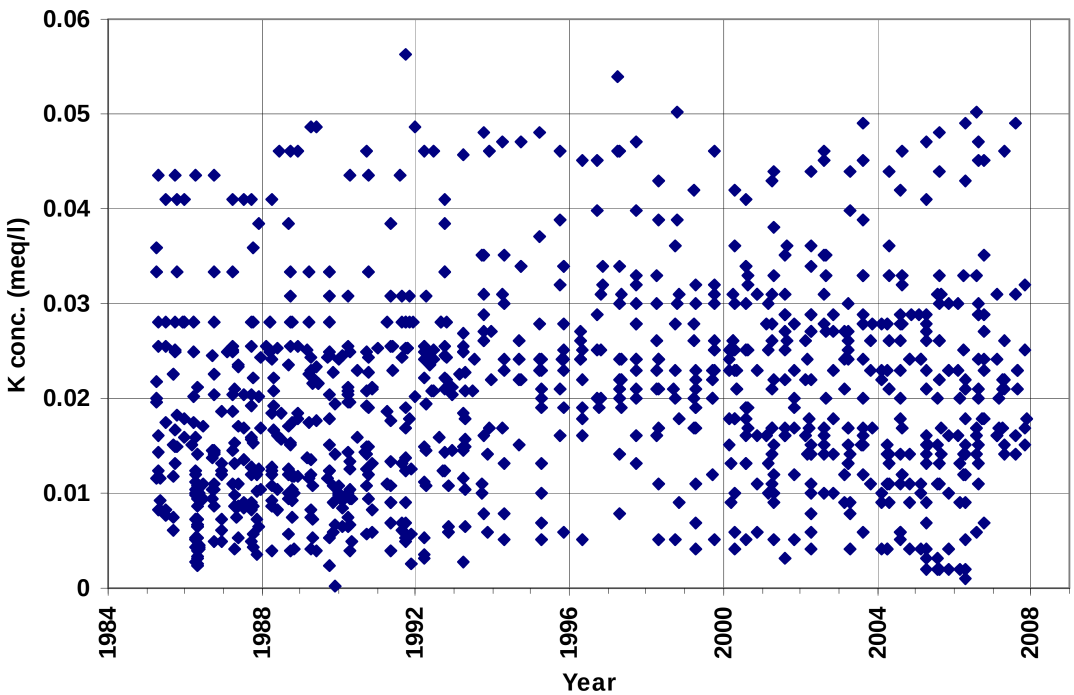

Figure 6. Potassium concentrations in groundwater sampled in 1985–2007 at 19 sites in the South Swedish Highlands. Samples were normally collected on 2–6 occasions per year at each site, although there were also some longer breaks in the dataset.

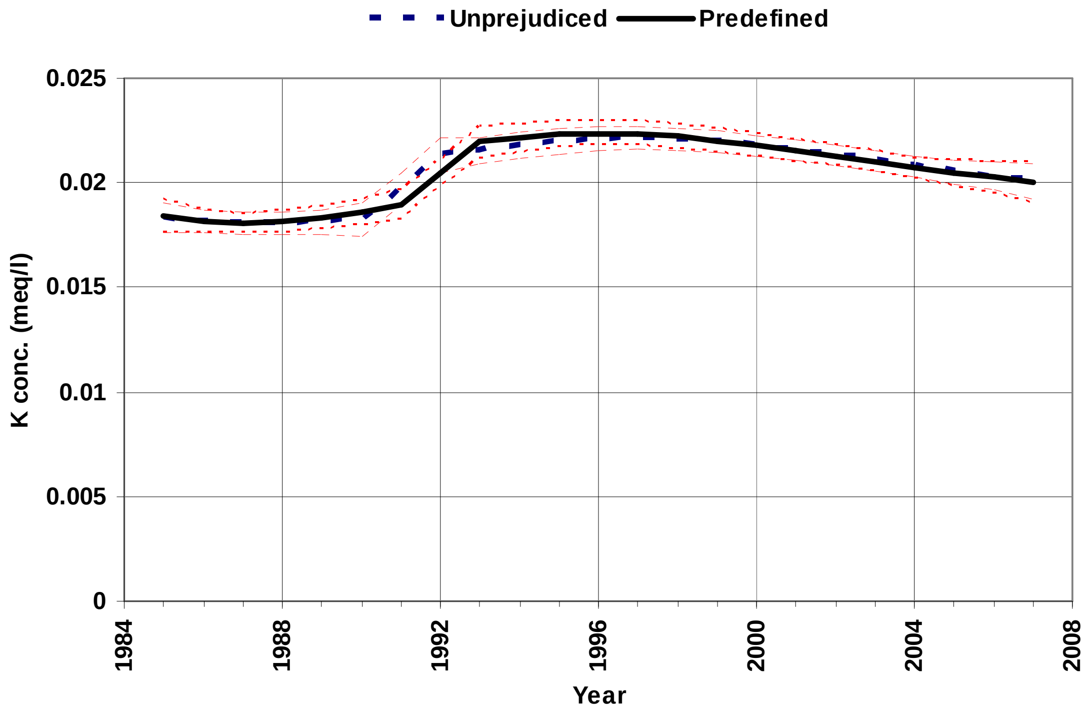

Figure 7. Annual means of potassium trends, including discontinuities, at 19 sites in the South Swedish Highlands. The thick solid and the thick dashed line represent two modes of the model runs: predefined change points and unprejudiced search for discontinuities, respectively.

Figure 8 shows the trend lines obtained after the estimated level shifts were removed. Apparently, there were only minor differences between the results obtained with a user-defined change point and those acquired in an unprejudiced search for discontinuities.

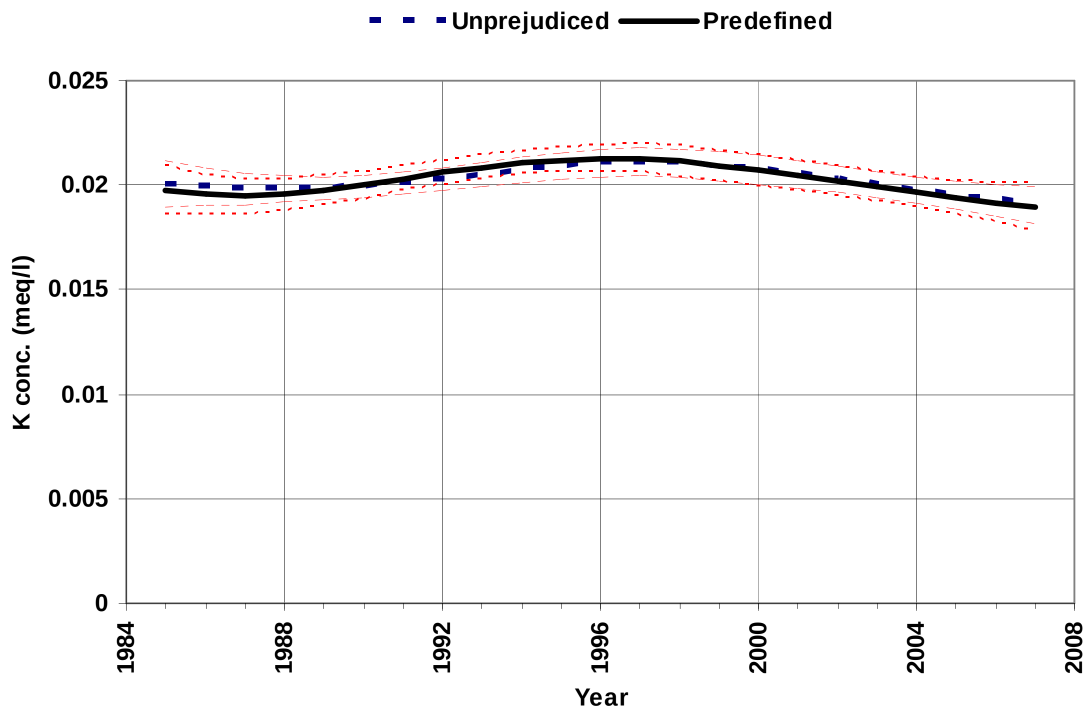

Figure 8. Annual means of potassium trends after removing the estimated level shifts. The thick solid and the thick dashed line represent two modes of the model runs: predefined change points and unprejudiced search for discontinuities, respectively.

Temporary bias

Since 1987, the amount of organic matter in Swedish surface waters has been measured as both TOC and COD (analysis of the latter using potassium permanganate as oxidant). Although there is no fixed relationship between the results obtained by the two methods, the data for each water body are normally strongly correlated, which makes it possible to identify time periods when the TOC or COD measurements have been biased. We chose to examine data from 1990 to 2005, because the first few years of TOC measurements were deemed to be less accurate.

Figure 9 illustrates the variation in TOC-to-COD ratios for fifteen major rivers in northern Sweden. This dataset was analysed using a model with two level shifts that were of the same size but had different signs. The timing of the level shifts was estimated from the data, and Figure 10 shows the sum of the estimated smooth trend surface and level shifts. As expected, the algorithm identified 1997 as a period during which the data deviated strongly, and closer analysis showed a level shift of 0.062 for that year, with a standard error of 0.0038.

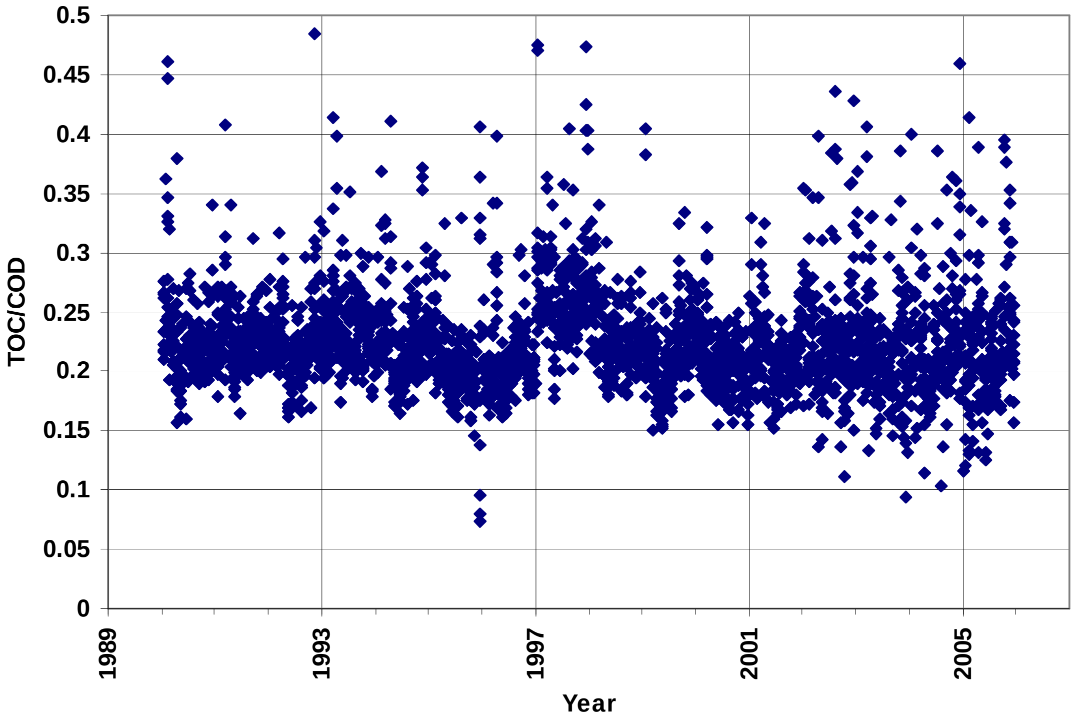

Figure 9. TOC-to-COD ratios recorded at the mouths of fifteen major rivers in northern Sweden. Monthly sampling was done in all the rivers throughout the investigated period.

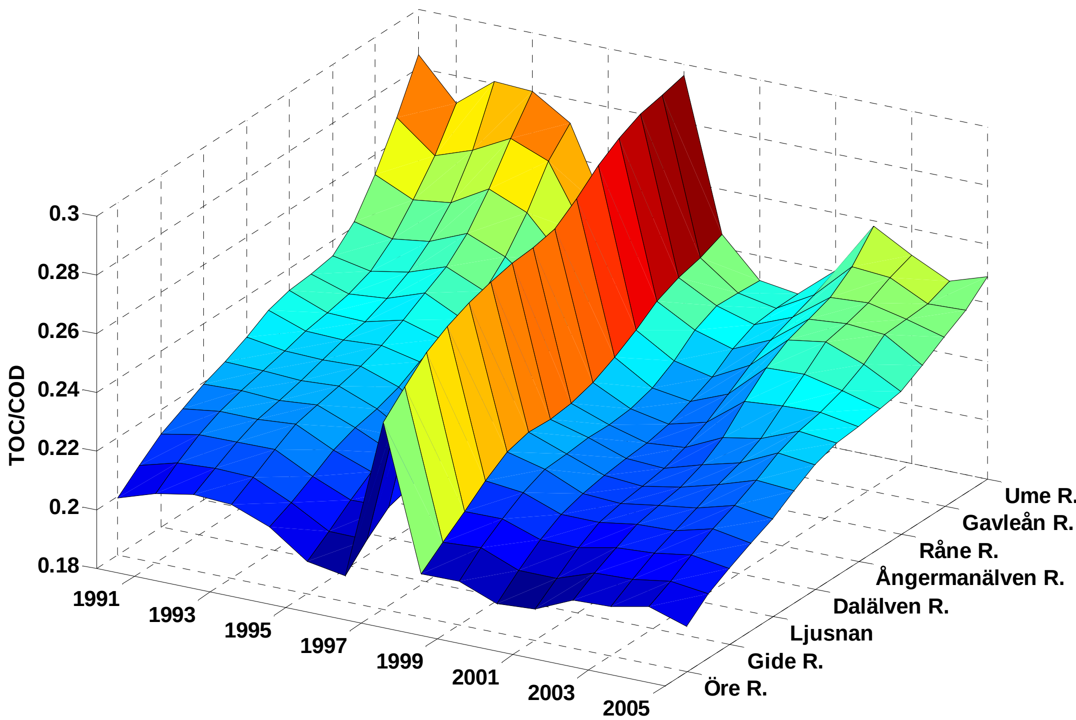

Figure 10. Trend surface with discontinuities fitted to the data given in Figure 9. Two level shifts of equal size but with different signs were assumed to be present during the period 1990–2005. The timing of the shifts was determined by an unprejudiced search.

In our previously cited article on data quality (Wahlin and Grimvall 2008a), we claimed that alkalinity trends in Swedish groundwaters were contaminated by systematic measurement errors in the early 1980s. We reanalysed that dataset in the present work. More specifically, we examined the difference between alkalinity and ANC in samples with low ANC levels (less than 0.3 but greater than 0.05 meq/l). We found that our algorithm, which can accommodate observations that are unevenly distributed in time and space, confirmed our previous suspicion. Figure 11 illustrates how introduction of a new analytical procedure stabilized the annual mean of the estimated trend surface (including discontinuities) after 1984.

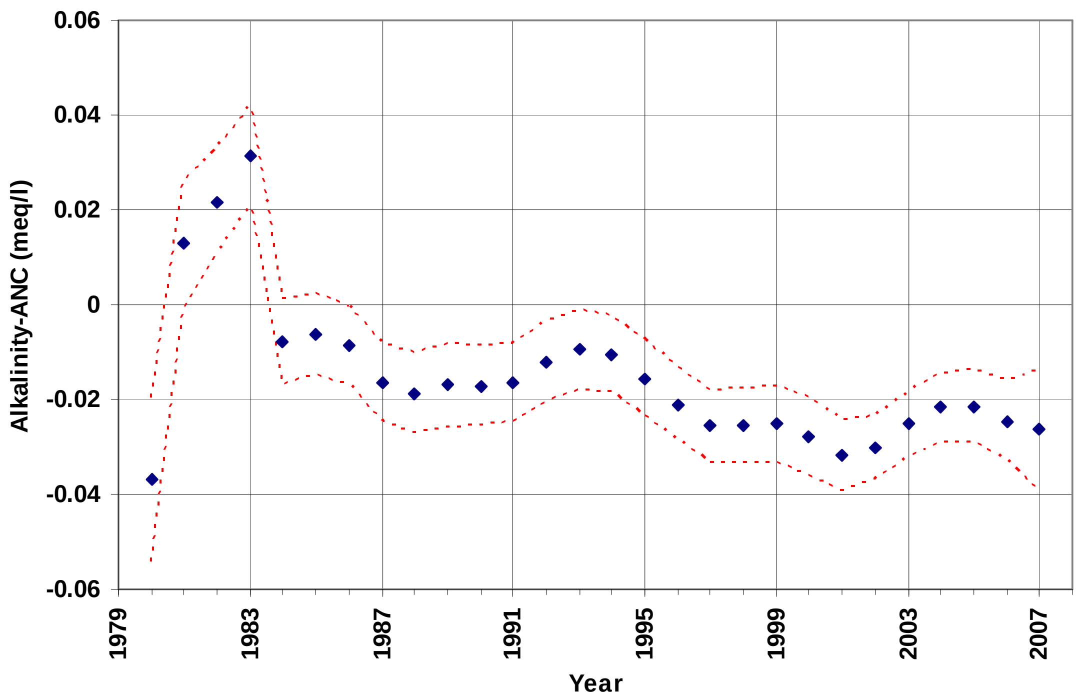

Figure 11. Annual means of trend levels (including discontinuities) fitted to differences between alkalinity and ANC in low-ANC samples from 77 Swedish groundwater sites.

Discussion

This article has demonstrated how smooth trends in a vector time series can be separated from abrupt level shifts that occur simultaneously in all coordinates. Such methods are obviously needed in environmental monitoring, but they can also be applied in almost any context in which several time series with similar trends are recorded. We developed our method primarily to facilitate unprejudiced searches for abrupt level shifts at unknown time points. However, our procedure is still applicable if we know when there has been some kind of major change, such as a switch in laboratory, personnel, analytical procedure, or sampling technique. More specifically, we can test the statistical significance of a level shift and examine whether the step size increases or decreases with the coordinate of the analysed vector time series.

The greatest strength of our method is its adaptive character. If the analysed time series have different trends, it is unlikely that the mean function can be made stepwise constant by subtracting a suitable reference from each series. This implies that, in such cases, none of the existing methods mentioned in the introduction are applicable. It has been suggested that ordinary regression models in which the mean function is linear between the change points can serve as alternatives to models with stepwise constant means (Alexandersson and Moberg 1997; Easterling and Peterson 1995). However, that model class forces the user to choose between constant and discontinuous trend slopes. Our method is based on the more natural assumption that the trend slope (after removing the level shifts at the change points) varies smoothly over the entire study period, and the selection of smoothing factors by cross-validation automatically adapts the degree of smoothness to the analysed data.

The limitations of our method are also related to its adaptive character. In principle, our technique can be generalized to handle multiple change points that occur at different times in different coordinates of the studied vector time series. However, there are two major obstacles to such generalizations. First, it is difficult to distinguish between smooth changes in the trend surface and the combined effect of multiple discontinuities, which occur relatively close in time. In addition, there are computational obstacles to the handling of multiple change points. The model we propose is a three-step back-fitting algorithm in which the smooth trend surface, the regression coefficients of the covariates, and the discontinuities are estimated separately. In this type of algorithm, each step must be very fast, because it is repeated many times during the model fitting and an even larger number of times during the cross-validation and the analysis of resampled data. Consequently, it is not feasible to make unprejudiced searches for complex patterns of discontinuities in the presence of smooth trends that may vary from coordinate to coordinate.

Some comments should also be made about the resampling technique we used to assess the uncertainty of the detected level shifts. Our technique offers the important advantage of taking into account the correlation structure of the model residuals. Moreover, it is well coordinated with the smoothers used to extract the trend surface. However, like any other form of residual resampling, our method creates a new resampled dataset by adding resampled residuals to fitted response values. Consequently, it is tacitly assumed that the errors in the fitted responses are considerably smaller than the individual error terms. This assumption is reasonable as long as the fitted responses are influenced by a large number of observations, but it is less appropriate if there are only a few influential data points. In practice, this implies that the uncertainty estimates are reliable for models with relatively strongly regularized trend surfaces (large λ-values).

Acknowledgements

The authors are grateful for financial support from the Swedish Environmental Protection Agency. The Geological Survey of Sweden funded the case studies of groundwater quality.

References

Aguilar E., Auer I., Brunet M., Peterson T.C. and Wieringa J. (2003). Guidelines on climate metadata and homogenization. WMO TD No. 1186, World Meteorological Organization.

Alexandersson H. (1986). A homogeneity test applied to precipitation data. Journal of Climatology, 6, 661-675.

Alexandersson H. and Moberg A. (1997). Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. International Journal of Climatology, 17, 25-34.

Caussinus H. and Mestre O. (2004). Computation and analysis of artificial shifts in climate series. Journal of the Royal Statistical Society Series C, 53, 405-425.

Easterling D.R. and Peterson T.C. (1995). A new method for detecting and adjusting for undocumented discontinuities in climatological time series. International Journal of Climatology, 15, 369-377.

Grath J., Ward R. and Quevauviller P. (eds) (2007). Common implementation strategy for the water framework directive. Guidance on groundwater monitoring. Office for Official Publications of the European Communities: Luxembourg.

Grimvall A., Wahlin K., Hussian M. and Libiseller C. (2008). Semiparametric smoothers for trend assessment of multiple time series of environmental quality data. Submitted to Environmetrics.

Hawkins D.M. (1977). Testing a sequence of observations for a shift in location. Journal of the American Statistical Association, 68, 941-943.

Hussian M., Grimvall A. and Petersen W. (2004). Estimation of the human impact on nutrient loads carried by the Elbe River. Environmental Monitoring and Assessment, 96, 15-33.

Jones P.D. (1995). The instrumental data record: its accuracy and use in attempts to identify the "CO2" signal. In von Storch H. and Navarra H. (eds), Analysis of Climate Variability. Springer Verlag: Berlin.

Klein Tank A.M.G., Wijngaard J.B., Können G.P., Böhm R., Demarée G., Gocheva A., Mileta M., Paschiardis S., Hejkrlik L., Kern-Hansen C., Heino R., Bessemoulin P., Müller-Westermeier G., Tzanakou M., Szalai S., Pálsdóttir T., Fitzgerald D., Rubin S., Capaldo M., Maugeri M., Leitass A., Bukantis A., Aberfeld R., van Engelen A.F.V., Førland E., Mietus M., Coelho F., Mares C., Razuvaev V., Nieplova E., Cegnar T., López A.J., Dahlström B., Moberg A., Kirchhofer W., Ceylan A., Pachaliuk O., Alexander L.V. and Petrovic P. (2002). Daily dataset of 20th-century surface air temperature and precipitation observations for European Climate Assessment. International Journal of Climatology, 22, 1441–1453.

Libiseller C., Grimvall A., Waldén J. and Saari H. (2005). Meteorological normalisation and non-parametric smoothing for quality assessment and trend analysis of tropospheric ozone data. Environmental Monitoring and Assessment, 100, 33-52.

LiU (Linköping University) (2008). http://www.ida.liu.se/divisions/stat/research/. Accessed 2008-08-20.

Mammen E. (2000). Resampling methods for nonparametric regression. In Schimek M.G. (ed), Smoothing and Regression – Approaches, Computation, and Application. John Wiley & Sons: New York.

Picard F., Lebarbier E., Budinska E. and Robin S. (2007). Joint segmentation of multivariate Gaussian processes using mixed linear models. Statistics for Systems Biology Group Research Report No. 5, Jouy-en-Josas/Paris/Evry, INRA, France.

SGU (Geological Survey of Sweden) (2008). http://www.sgu.se/sgu/sv/samhalle/miljo/miljoovervakning/datavard-grundvatten.html. Accessed 2008-08-20.

SLU (Swedish University of Agricultural Sciences) (2008). http://www.ma.slu.se. Accessed 2008-08-20.

Sristava M.S. and Worsley K.J. (1986). Likelihood ratio test for a change in multivariate normal mean. Journal of the American Statistical Association, 81, 199-204.

Stålnacke P. and Grimvall A. (2001). Semiparametric approaches to flow-normalisation and source apportionment of substance transport in rivers. Environmetrics, 12, 233-250.

Szentimrey T. (1997). Statistical procedure for joint homogenization of climatic time series. Proceedings of the 1st Seminar on Homogenization of Surface Climatological Data, Budapest, Hungary, pp. 47-62.

Wahlin K. and Grimvall A. (2008a). Uncertainty in water quality data and its implications for trend detection: lessons from Swedish environmental data. Environmental Science and Policy, 11, 115-124.

Wahlin K. and Grimvall A. (2008b). Roadmap for assessing regional trends in groundwater quality. Submitted to Environmental Monitoring and Assessment.

Worsley K.J. (1979). On the likelihood ratio test for a shift in location of normal populations. Journal of the American Statistical Association, 74, 365-367.