Evaluation of total groundwater abstraction from public waterworks in Denmark using principal component analysis

Brian Lyngby Sørensen and Rasmus Rønde Møller

In Denmark water abstraction data have been collected since the late 1970s. Initially the purpose was to monitor and assess the groundwater resources available for future local water abstraction. For this reason, abstraction data were collected not only from waterworks, but also from irrigation, industry etc. Today water abstraction data are used for several purposes, for instance in water -balance calculations to estimate the available resource to wetlands, streams and lakes or to calculate the flow of chemical substances in the water environment.

The role of climatic changes in the future hydrological cycle is subject to increasing attention. Apart from a small reserve of surface water, all drinking water in Denmark comes from groundwater. When precipitation changes in the future the amount of groundwater available for abstraction will also change. Hence, for reasons of security of supply and environmental impact, it is important to know the amount and trend of abstraction each year.

At national level, it is a statutory objective to abstract groundwater in a way that does not obstruct the general water-environmental objectives outlined in the European Union's Water Framework Directive (The European Parliament and the Council of the European Union 2000). The purpose of this paper is to present a method to evaluate the errors in the overall national groundwater abstraction dataset and describe how to correct erroneous data. For the sake of overview the national data are typically presented as an overall sum in million cubic metres per year (e.g. Thorling *et al*. 2012).

Public groundwater abstraction in Denmark

Drinking water in Denmark comes from approximately 2500 waterworks, abstracting about 400 million m^3 of groundwater per year. There is a pronounced decentralised water supply structure with many small waterworks spread across the country. Approximately 72% of the waterworks each abstract less than 0.1 million m^3 water per year, amounting to a total of 56.5 million m^3 per year. At the other end of the

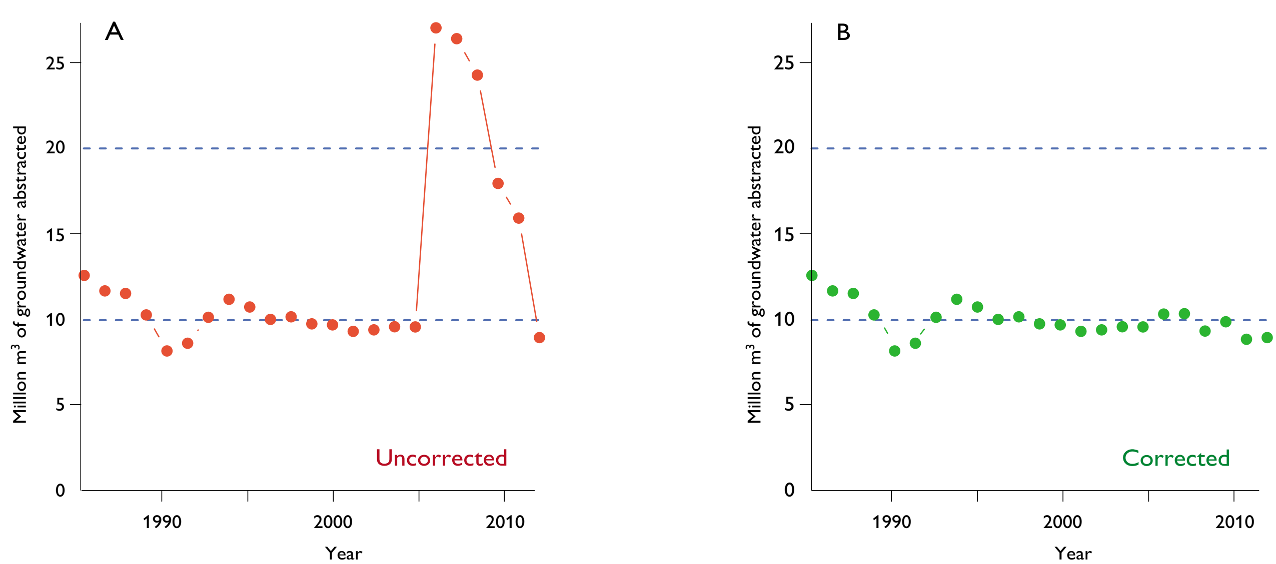

Fig. 1. An example of a time series for a specific municipality before (**A**) and after (**B**) correction of the abstraction data. Data from 2011 are included in the graph for clarity.

Table 1. Typical problems associated with registration of water abstraction data

Problem	Cause	Action
No data	No data were reported at all from the municipality	An expected average was calculated based on data from 1–2 years before and after the year with missing data.
Evidently missing data	No data from one or more waterworks. Typing errors	An expected average was calculated for the individual waterworks, or in case of typing errors a more probable value was estimated.
Evidently too high amount quoted	Double registration from one or more waterworks. Typing errors	Evident double registrations were subtracted from the sum. In case of typing errors a more probable value was estimated.
Other apparent error	Unidentified	No action taken.

scale, 3% of the waterworks each abstract more than 1 million m^3 per year, totalling 154 million m^3 per year.

According to Danish legislation it is mandatory for waterworks and other users abstracting groundwater to report the amount abstracted once a year to the municipalities. The municipalities check for mistyped data and forward them to the national Danish database on geology, groundwater and drinking water (the Jupiter database at the Geological Survey of Denmark and Greenland).

Municipal reform

In 2007, a major municipal reform took place in Denmark. Thirteen former counties (amter) were replaced by five so-called regions and most municipalities (kommuner) were merged into fewer and larger units, resulting in a drop from 271 to 98 municipalities. As part of this reform the new municipalities took over the responsibility to manage the water resources including abstraction licensing. This involved transferring employees from the former counties, new distribution of responsibilities and introduction of new computer systems and new procedures; all of which influenced the overall quality of the abstraction data. For instance, the new municipalities were responsible for submitting the 2006 water abstraction data to Jupiter, although they were not operative before 1 January 2007.

Data preparation

The water abstraction data used in this study were extracted from the Jupiter database for the period 1989–2010. Based on the extracted data, a date table was compiled with the sum of groundwater abstraction per year within each municipality. A time series for each municipality was plotted and visually inspected. At municipality level, small year-to-year changes and thus a smooth curve are expected, because

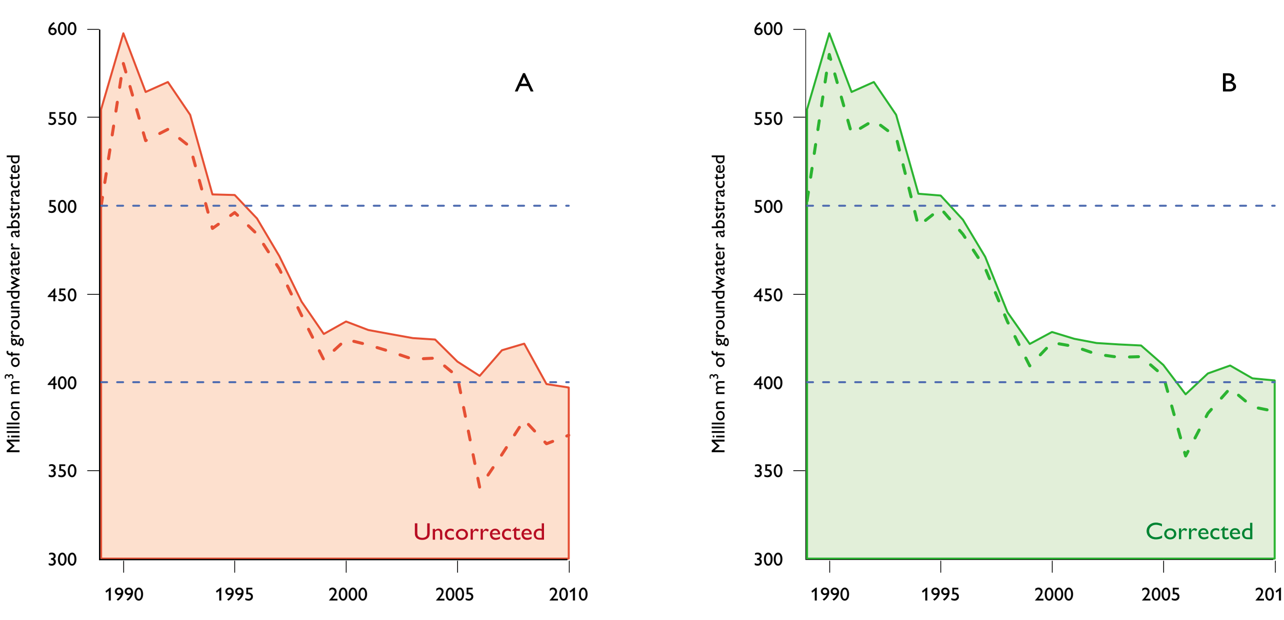

Fig. 2. Total water abstraction in Denmark for uncorrected (**A**) and corrected (**B**) data. The dashed lines show VAR_{exp} – the correlation between the PCA score of the first primary component (PC1) and the input data, expressed in million m^3 per year.

on average the waterworks abstract almost the same amount each year.
After initial inspection, 22 municipalities with unexpected data pattern were selected for detailed examination. Four types of main problems were identified (Table 1); the causes for three of the types could be identified and relevant action taken.

Correction of abstraction data for a single municipality

An example of a time series for a selected municipality is shown in Fig. 1A. The water abstraction from a specific waterworks was erroneously reported three times in the years 2006–2008 and twice in the years 2009–2010. Thus, the water abstraction in the municipality was overestimated by 16.7 and 7 million m^3, respectively, in the two periods. With the extra registrations removed, the time series shows a behaviour similar to what is expected (Fig. 1B). A similar inspection was made of the time series from the 21 other municipalities. Finally, a new data table was compiled by merging the corrected data with the data from the uncorrected time series from the remaining 76 municipalities.

Principal component analysis and Pearson's correlation coefficient

Principal component analysis is a mathematical procedure introduced by Pearson (1901) and widely used to visualise multivariate data by dimension reduction (Garcia & Filzmoser 2011). According to Garcia & Filzmoser, the main problems of multivariate data can be avoided by using the principal component analysis to transform "… the original variables into a smaller set of latent variables which are uncorrelated". Each new variable (principal component or PC) can then be interpreted independently.

There are several ways to perform principal component analysis, some of which are described in Wikipedia (2013). The method used here is singular value decomposition (SVD) using the 'prcomp' function of the base package of R (R Core Team 2012).

The time series for the individual municipalities were used as objects (rows) and the years were used as variables (columns). For each year the Pearson's correlation coefficient ρ between the scores of the first principal component (PC1) and the corrected and uncorrected datasets D, was calculated and expressed in terms of million m^3 (VAR_{exp}) using the formula:

$$VAR_{exp} = \rho \times T = cor(PC1, D) \times T$$

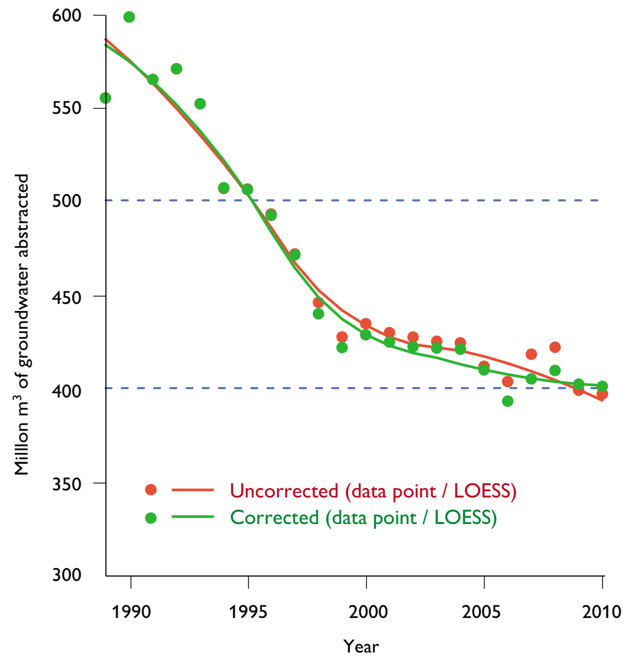

Fig. 3. Locally weighted average (LOESS) of uncorrected and corrected groundwater abstraction data.

where T is the total national abstraction. The correlation was done using the default settings of the 'cor' function of R (R Core Team 2012). The magnitude of ρ shows the strength of the linear dependence between the score of PC1 and D.

Status of water abstraction and comparison of uncorrected and corrected data

Figure 2 shows the total groundwater abstraction from public waterworks in million m^3 per year from 1989 to 2010 with uncorrected and corrected data. Both diagrams show the Pearson's correlation coefficient expressed in million m^3 (VAR_{exp}, dashed lines), according to the formula above. The variance explained ranges between 90 and 98% of the total yearly water abstraction. The remaining 2–10% can be perceived as 'noise' in the sense that this part of the variance is due to errors, short-term but large extra deliveries of water, abrupt changes in water needs, new or closed down waterworks etc. Before the municipal reform (the period from 1989 to 2005) the unexplained variance on average corresponds to 16 million m^3 for the uncorrected data and 12.7 million m^3 for the corrected data. The improvement of the explained variance by correcting the data is thus 3.3 million m^3. After the reform (2006–2010) the unexplained variance on average corresponds to 45.3 million m^3 for the uncorrected data and 20.8 million m^3 for the corrected data, leading to an average improvement of 24.5 million m^3 by correcting the data.

Because of the errors mentioned above the amount of groundwater abstracted in Denmark by the waterworks is only known with some uncertainty. In Fig. 3 a locally weighted regression (LOESS) is calculated for corrected and uncorrected abstraction data in order to yield a 'best guess' of the total water abstraction. The curves show an overall trend with a large decline in the first half of the 1990s when abstraction decreased *c.* 20% from *c.* 550 million m^3 in 1990 to *c.* 460 million m^3 in 1996. Later, the abstraction dropped to just over 400 million m^3 in 2005. From Fig. 3 it is clear that when corrected data are used, the abstraction flattens out at around 400 million m^3 per year from 2005 onwards. If uncorrected data are used the abstraction level seems to decrease even further to below 400 million m^3 per year over the same period. Therefore the interpretation of trends depends to a large degree on whether the data are corrected or not. The main reasons for the large decline after 1989 are adoption of new legislation, increased water taxes and water saving campaigns (Stockmarr & Thomsen 2006).

Conclusions

After the municipal reform in 2007 water abstraction data reported to the Jupiter database show increased levels of errors due to changes in the way data are treated and reported. This means that national trends and levels are blurred which can lead to misinterpretations. By carefully examining data from the individual waterworks, it is often possible to determine the causes of errors and thereby correct them. The combined use of PCA and Pearson's correlation coefficient is a useful way to provide an overall check on how well the data are corrected. This study shows that after the municipal reform the improvement is on average equivalent to 24.5 million m^3 or *c.* 6%.

On regional and local scales the impact of erroneous data can be severe. The example in Fig. 1 shows that the abstraction can be overestimated by a factor 2.5 if no action is taken to investigate and correct erroneous data. It is crucial to correct and improve such data before they are used in water-balance calculations, hydrological modelling, abstraction licensing and projections of water use in Denmark.

References

Garcia, H. & Filzmoser, P. 2011: Multivariate statistical analysis using the R package chemometrics, 71 pp. Vienna: Vienna University of Technology, Department of Statistics and Probability Theory.

Thorling, L., Hansen, B., Langtofte, C., Brüsch, W., Møller, R.R. & Mielby, S. 2012: Grundvandsovervågning 2012 – Grundvand. Status og udvikling 1989–2011. Teknisk rapport, http://www.geus.dk/publications/grundvandsovervaagning/1989_2011.htm

Pearson, K. 1901: On lines and planes of closest fit to systems of points in space. Philosophical Magazine, series 6, **2**, 559–572.

R Core Team 2012: R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, *http://www.R-project.org/*

Stockmarr, J. & Thomsen, R. 2006: Water supply in Denmark. The Danish action plan for promotion of eco-efficient technologies – Danish lessons, 18 pp. Copenhagen: Miljøstyrelsen.

The European Parliament and the Council of the European Union 2000: Establishing a framework for community action in the field of water policy. *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0060:EN:HTML*

Wikipedia, the free encyclopedia 2013: Principal component analysis. Accessed 7 February 2013. *http://en.wikipedia.org/wiki/Principal_component_analysis*

Authors' addresses

B.L.S. & R.R.M.* *Geological Survey of Denmark and Greenland, Lyseng Allé 1, DK-8270 Højbjerg, Denmark.* E-mail: *bls@geus.dk*

* Present address: *Horsens Kommune, Rådhustorvet 4, 8700 Horsens, Denmark.*