Guidelines for statistical analysis of hydrological data according to the requirements of the Groundwater Directive

Juan Grima Olmedo, Juan Antonio Luque Espinar

Abstract: The Groundwater Directive 2006/118/EC has been developed in response to the requirements of Article 17 of the Water Framework Directive. It aims to protect groundwater by preventing or limiting the input of polluting substances. As groundwater is an essential part of the hydrologic cycle, its deterioration may directly affects dependent aquatic and terrestrial ecosystems. It is necessary, therefore, to assess the current groundwater status by means of sound scientific methods to obtain comparable results. The Directive suggests linear regression as a standard methodology, although a number of assumptions must be checked before applying any parametric technique to avoid erroneous results. Even if the underlying assumptions of linear regression are met, a careful analysis of data must be done to find patterns or trends in data. The starting point of the evaluation is the calculation of a baseline, defined as the mean in values for reference years 2007 and 2008. To start the assessment, the data must be visualized graphically and exploratory data analysis techniques applied. A temporal period where the behaviour of groundwater is homogeneous can be chosen as the reference period for the elaboration of an extended or updated baseline. If no historical data are available, a decision must be made about how to update the baseline. In a second stage, a hypothesis test must be conducted to detect deterioration of groundwater quality. The ability of statistical intervals to detect small changes in quality depends on the size of the adopted baseline and, if required, on the future statistic to be evaluated. As a general rule, a minimum of eight measurements are required to keep uncertainty due to baseline itself within reasonable limits. The elaboration of confidence limit or prediction limits for future means or medians, in combination with trend analysis, is the recommended procedure for groundwater bodies' management. Finally, if results show a significant statistical impact on groundwater quality, exogenous factors must be analyzed to account for natural variability of data before declaring a groundwater body in bad ecological status.

Keywords: Groundwater body, trends, linear regression, baseline, Groundwater Directive, water quality

Juan GRIMA OLMEDO, **Juan Antonio LUQUE ESPINAR**
Dpto de Investigación y Prospectiva Geocientífica
IGME (Geological Survey of Spain)
http://www.igme.es

Juan GRIMA OLMEDO ✉
Cirilo Amorós, 42, 46004 Valencia, España
Tel.: +34 963943474 - Fax: +34 963944436
Email: j.grima@igme.es

Juan Antonio LUQUE ESPINAR
Urb. Alcázar del Genil, 4-Edif. Zulema, Bajo, 18006 Granada, España
Tel.: +34 958183143 - Fax: +34 958122990
Email: ja.luque@igme.es

Received: 25 may 2010 / Accepted: 30 may 2010

Riassunto: *La Direttiva sulle Acque sotterranee 2006/118/EC è stata elaborata in risposta alle richieste (esigenze) dell'Articolo 17 del Water Framework Directive dove si dice di proteggere le acque sotterranee impedendo o limitando l'immissione di sostanze inquinanti. Siccome le acque sono una parte essenziale del ciclo idrologico, il loro deterioramento può avere effetti direttamente a carico degli ecosistemi delle acque e terrestre. E' necessario, tuttavia, valutare l'attuale stato delle acque sotterranee attraverso validi metodi scientifici per ottenere dei risultati comparabili. La regressione lineare viene suggerita dalla Direttiva come una metodologia standard, sebbene un numero di presupposti devono essere controllati, prima di applicare ciascuna tecnica parametrica, per evitare risultati errati. Perfino se si arriva a conoscere i presupposti che sono alla base della regressione lineare, deve essere fatta una analisi curata dei dati, per trovarVI pattern o trends (nei dati). Il punto di partenza della valutazione è il calcolo della linea di delimitazione, definita come la media dei valori per gli anni di riferimento 2007 e 2008. Per iniziare con la valutazione dei dati li devo visualizzare graficamente ed applicare tecniche esplorative per l'analisi dei dati. Un periodo temporale nel quale il comportamento delle acque sotterranee è omogeneo può essere scelto come periodo di riferimento per l'elaborazione di una estesa o aggiornata linea di delimitazione linea di riferimento. Se non sono disponibili dati storici, deve essere presa una decisione su come aggiornare la linea di delimitazione. In una seconda fase deve essere elaborato un test delle ipotesi per individuare il deterioramento della qualità delle acque sotterranee. La capacità degli intervalli statistici nel trovare piccoli cambiamenti nella qualità dipende dalle dimensioni o misura della linea di delimitazione linea di riferimento e, se richiesto, dalla statistica futura per essere valutati. Come regola generale, sono richieste un minimo di otto misurazioni per mantenere l'incertezza relativa alla linea di riferimento stessa entro limiti ragionevoli. L'elaborazione del limite di confidenza o di previsione per le medie o mediane future in combinazione con l'analisi dei trend è una procedura raccomandata per la gestione degli acquiferi sotterranei. Finalmente, se i risultati mostrano un significativo impatto statistico sulla qualità delle acque sotterranee, fattori esogeni devono essere analizzati per giustificare la naturale variabilità dei dati prima di dichiarare un corpo idrico sotterraneo in un cattivo stato ecologico.*

Introduction

General Provisions for the protection and conservation of groundwater are set out in Directive 2000/60/EC. As a requirement, measures to prevent and control groundwater pollution should be adopted, including criteria for assessing good groundwater chemical status, and for the identification of significant and sustained upward trends in the concentration of substances which occur both naturally and as a result of human activities. Baseline levels are the starting point for the definition of such trends. Data collected before the starting of the monitoring period can be used when available.

DOI 10.4409/Am-001-10-0003

The Directive calls for the application of standardised statistical techniques, such as regression analysis, in order to obtain comparable results and to ensure equivalent scientific quality.

Application of regression techniques to large samples and complete data sets provides a powerful tool for analysis. Unfortunately, groundwater data have distinctive features, like positive skewness and non-normal distribution (Helsel and Hirsch, 2002) and, in practice, small samples and incomplete data sets are handled. In addition, other factors (like seasonal effects and multiple trends) may influence the response variable and cause statistical tests to provide erroneous results. From this perspective, in order to gain better understanding of the underlying structure and test assumptions, a variety of preliminary data analysis must precede the selection and performance of any hypothesis test.

Monitoring data from groundwater bodies defined at risk of not achieving good ecological status have been analyzed extensively in Spain in the framework of a specific agreement between the Ministry of the Environment and Rural and Marine Affairs and the Spanish Geological Survey (IGME).

The main output of the project has been the definition of a methodology to analyze such data. Procedures for the selection of a reference standard (or background) to detect statistically significant impacts on groundwater quality are provided. Moreover, techniques for determination of trends and its nature have been analyzed. In order to illustrate the procedure and highlight the main results, the “Plana de Sagunto” (Júcar Pilot River Basin) groundwater body has been selected, and the methodology applied. The reason for the selection of this groundwater bodies has been the length of the series with relatively equally spaced values of time. The parameter analyzed is nitrate concentration.

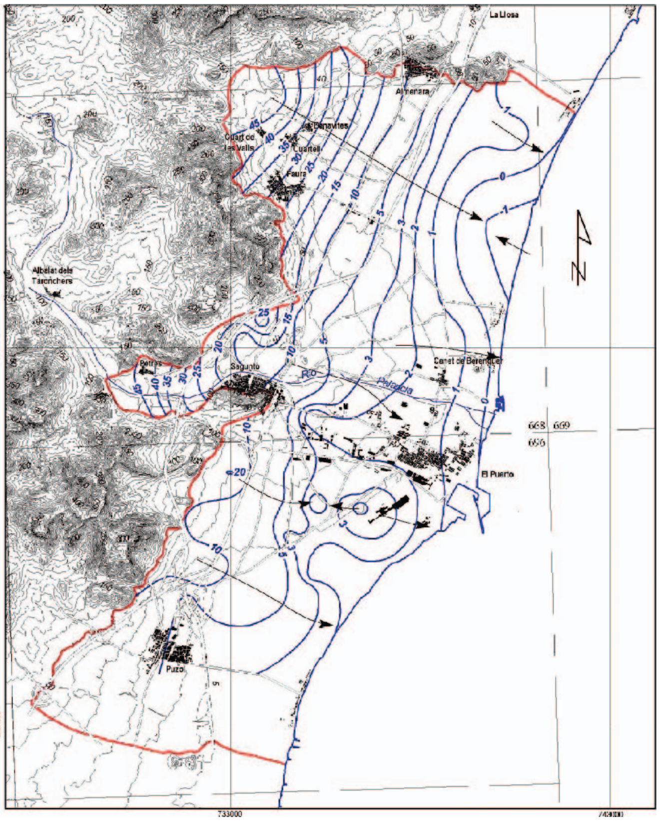


Fig. 1: *Plana de Sagunto Groundwater Body* .

Definition of reference levels

Groundwater bodies are complex systems presenting a high variability not only in geometry and hydrodynamic characteristics of its geological media, but in its physical-chemical properties (Ballesteros et al, 2001). It is then clear that from a scientific point of view it does not make sense to define one single concentration for an aquifer as a whole (Blum et al, 2009). However, referring to the European Directives requirements, the use of average concentrations is helpful.

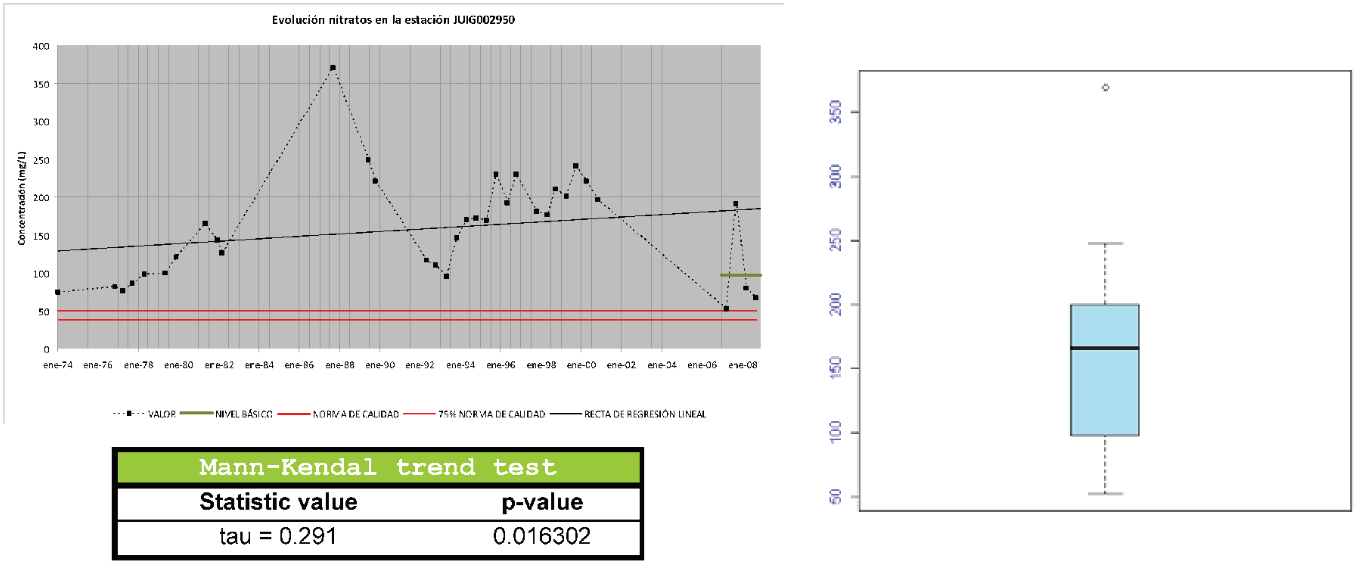
From a regulatory perspective (Groundwater Daughter Directive), “background level” means the concentration of a substance or the value of an indicator in a body of groundwater corresponding to no, or only minor, anthropogenic alterations to undisturbed conditions. “Sensu stricto” it refers to unpolluted conditions in pre-industrial times, which does not seem to be realistic (at least for substances with both natural and anthropic origin). For substances without a natural origin, the background should be set equal to zero (referenced to the Limit of Detection), while, for naturally occurring substances, such a level has to be considered as a concentration range (Müller et al, 2006).

On the other hand, “baseline level” means the average value measured at least during the reference years 2007 and 2008 on the basis of monitoring programs implemented under Article 8 of Directive 2000/60/EC or, in the case of substances identified after these reference years, during the first period for which a representative period of monitoring data is available. The baseline is defined when relevant risks have been identified during the process of characterization as set out by the Water Framework Directive. It means that human pressure does exist, and therefore, natural background levels are certainly difficult to achieve in most cases.

According to the Groundwater Directive, Member States may exempt from the measures to prevent or limit the input of pollutants into groundwater when disproportionately costly measures are required. It does not mean that nothing must be done to avoid deterioration of quality of groundwater bodies. A useful tool to keep track of the evolution of the quality of the groundwater is by means of measurements from the monitoring network. Even if it is not possible to reach natural background levels, it is feasible to evaluate statistically significant impacts on water quality from the beginning of the monitoring period. With this aim, all the existing measurements can be used to elaborate an “extended or updated baseline level”. The aggregation of additional measurements to the existing baseline level has the advantage that the power of the statistical tests is increased, more accurate prediction intervals can be elaborated, and natural variation in the concentration of the pollutant is better accounted for.

Preliminary data analysis

Before any statistical analysis is applied, a thorough investigation of data is advisable, in order to look for structures before formulating mathematical hypothesis about data distribution. Graphical representation of data is a convenient way of obtaining additional information to that which is provided by formal hypothesis tests. As an example, Shapiro-Wilk test provides a base to accept or reject normality in a set of data, while histograms or normal probability plots can indicate the causes of normality departure, like skewness or a single outlier.



Mann-Kendal trend test	
Statistic value	**p-value**
tau = 0.291	0.016302

Fig. 2: *Visual presentation of monitoring data and trend analysis.*

Following the indications of the Groundwater Directive, an ordinary linear regression can be performed over the whole data set. Results can be achieved by the simple application of a standard spreadsheet. The estimates from linear regression (and especially the slope of the regression line) are valid whenever the standard requirements are met. One of the initial hypotheses of linear regression is that the residuals are independent and identically distributed random variables, which, in practice, means that normality of residuals should be checked. In a case in which the residuals that form linear regression indicate non-normality, a non-parametric Mann-Kendall trend test can be done instead (Luque et al, 2010).

As can be seen in figure 2, a statistically significant upward trend is detected in "Plana de Sagunto" data by means of linear regression. A non-parametric Mann-Kendall trend test confirms the result (Shapiro-Wilk test rejects normality of data). Nevertheless, the high variability of the time series makes it difficult to get a clear picture of the evolution of data over time (Grima et al, 2010).

For reducing the noise due to random variation, an often used technique is smoothing. Smoothing techniques (when properly applied) are a simple and efficient tool to reveal more clearly underlying patterns and seasonal and cyclic components. The lowess is a data smoothing algorithm (Cleveland et al., 1979) that uses a moving window superimposed over a set of data. No model is assumed, which means that it is a non-parametric technique. In contrast to parametric procedures, it doesn't fit a straight line through the data. For example, a statistically significant trend may not be identified by linear regression, while the lowess line may detect an increase in concentration during part of the time frame, followed by a decrease.

That is the precisely the situation observed in "Plana de Sagunto" (figure 3) where two different intervals can be differentiated over the monitoring period. The first one (up to year 1994) exhibits an upward trend, while the second one shows stabilization or even a downward trend. Before making a decision about the selection of a reference period (i.e. discard some data), a careful analysis must be done in order to find real facts to explain the statistical findings. In this case, a thorough search to determine the causes of such an evolution was carried out, and the main search result was a Regional Government Decree for the modernization of the irrigation structures in the Autonomous Community of Valencia (as of May 1994). The effect of this piece of legislation was that a number of irrigation systems were changed, implementing drip irrigation systems in most of Valencia, and specifically in Plana de Sagunto system.

The protection of waters against pollution caused by nitrates from agricultural sources was adopted on 12 December 1991. Article 10 of the Nitrates Directive requires that Member States identify on their territory groundwater affected or liable to be affected by pollution (in particular when nitrate concentrations in groundwater or surface waters exceed 50 mg/l). The first outcome of interest of the preliminary data analysis is that nitrate concentrations in Plana de Sagunto are quite far away from this objective (figure 3). Subsequently, the reduction of concentration up to 50 mg/l may not be cost effective.

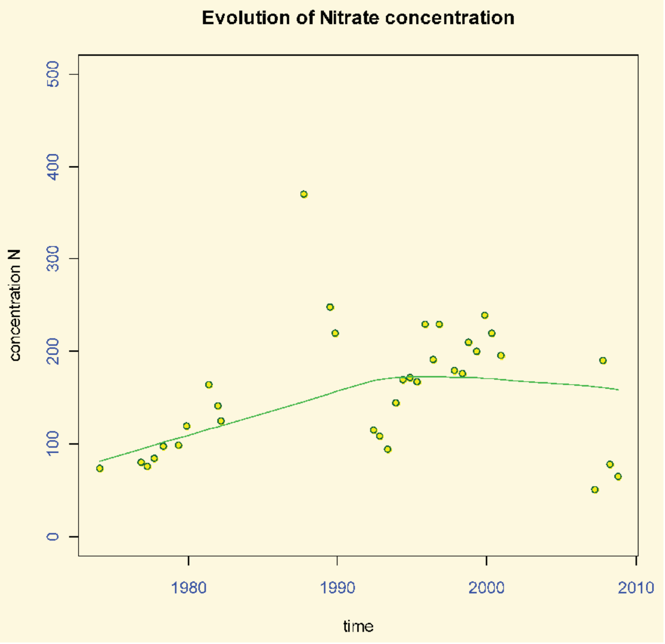


Fig. 3: *LOWESS adjustment for "Plana de Sagunto" data.*

However, the Directive has the objective of reducing water pollution caused or induced by nitrates from agricultural sources and preventing such pollution. It implies the obligation to avoid further deterioration, and the necessity to implement control measures in data from monitoring networks.

Methodological and practical considerations about the baseline

After preliminary data analysis, the selection of the reference period for baseline determination must be accomplished. As set by the Groundwater Directive, existing data before the start of the monitoring period (historical data) can be used.

As previously explained, in order to make a decision about the number of measurements to be selected, a careful analysis of the real facts behind some existing pattern or change in trend must be done. As baseline levels are defined as the mean value of years 2007 and 2008, values from that date should exist at any monitoring station (at least for groundwater bodies at risk). When additional information exists, it can be used to provide additional information regarding variability in data over time, while in case no other measurements are available, the existing ones must be used, although some limitations must be taken into consideration.

If the data are normal or can be normalized, parametric approaches can be used. Otherwise, non-parametric alternatives must be employed. In both cases, a number of statistical assumptions must be met, like statistical independence, temporal and spatial stationarity and lack of statistical outliers. Moreover, when a parametric approach is to be used (which implies more statistical power), assumptions about distribution must be correct.

When just a few values are available, non-parametric methods are advised, because normality tests do not have enough power to tell whether or not a small sample of data comes from a Gaussian distribution, unless departure from normality are very high. Then, the problem of including outliers in groundwater monitoring data can be striking, because prediction limits based on small data samples can provide intervals of disproportionate length that are not protective of the environment.

To illustrate it with an example, let’s go back to the “Plana de Sagunto” data. In this case, values from year 1994 have been selected as the reference period. To estimate the slope, a linear regression is then performed over such period.

The box plot of the reference period exhibits symmetry across the middle 50% of the distribution, but a few outliers can be identified, so normality must be checked. In fact, if a Shapiro-Wilk test is applied on the data itself, normality is not accepted.

Nevertheless, the same test on the residuals of the regression passes the test, proving the existence of a slightly negative slope, what means that there is a small downward trend, although statistically significant (p-value of 0.003). Values for the mean or the median can be calculated, producing results of 174.77 and 190.6 mg/l of nitrates in ground water.

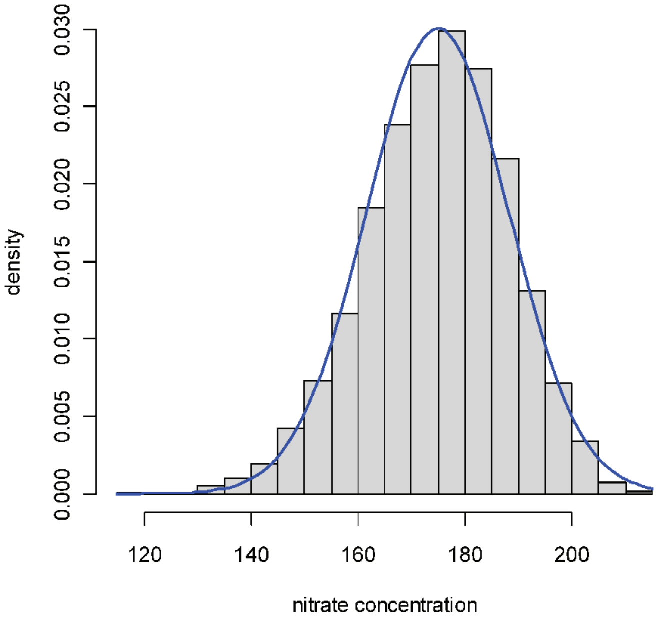
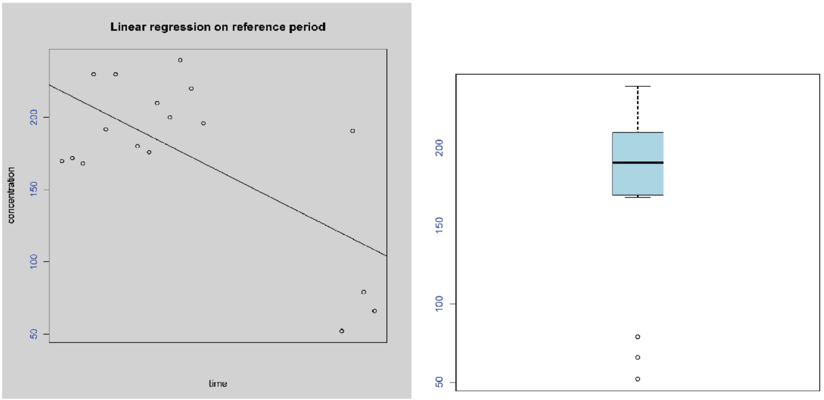


Fig. 5: *Histogram of bootstrapped mean for updated baseline.*



Normality of residuals Shapiro-Wilk test	
Statistic value	**p-value**
W = 0.9467	0.4068

Fig. 4: *Linear regression on the reference period.*

At first glance these values (mean or median) seem a straightforward comparison to check data from monitoring stations and obtain an early warning of deterioration of ground water quality. They have, however, the disadvantage that the mean represents a central tendency value, what implies that data from the same distribution will exceed such a value with at least a probability of 0.5.

To obtain a confidence interval for the mean or the median, a bootstrap method can be applied to the extended baseline, providing the following results in the example:

Tab. 1: *Summary of confidence limits for the mean.*

Confidence intervals for Bootstrapped mean			
Baseline		***Updated baseline***	
Upper	Lower	**Upper**	Lower
134.76 mg/l	58.82 mg/l	**195.59 mg/l**	151.64 mg/l

As can be seen from the table, an upper confidence limit about the mean of the updated baseline provides a value of nearly 196 mg/l in nitrate concentration, while the upper limit for the baseline calculation is just 134.76 mg/l. On the other hand, it may make little or no sense to construct a confidence interval around a baseline level with just a few measurements. In some extreme cases just as few of them may be available.

Trend analysis

Monitoring networks provide observations of a random variable (concentration of a pollutant) over a period of time. Trend analysis consists of the application of statistical techniques for making statements about the behaviour of a data set. In particular, the detection of statistically significant upward or downward trends in a number of measurements it is of interest.

The most widely used mechanisms for trend identification are simple linear regression and Mann-Kendall trends tests. The former is used for the identification of linear trends, and the main requirement is related to the residuals of the adjustment. They must be normal or at least reasonably symmetric. When investigating the increase in concentration of a given pollutant in a groundwater body, the relationship between the response or dependent variable (concentration) and time must be investigated. If such a relationship exists, its pattern may be linear or not. Because of that reason, the application of simple linear regression techniques with no previous check can produce erroneous results.

Both techniques have been applied to "Plana de Sagunto" data set. The linear regression produces an estimate of -0.02 for the value of the trend, while the Mann-Kendall non-parametric test doesn't provide evidence to reject the hypothesis of no trend.

Following the requirements of the linear regression, a Shapiro-Wilk test has been performed over the residuals of the regression, being the normality of the residuals accepted.

Tab. 2: *Summary of estimates for the slope.*

Estimates for trend in "Plana de Sagunto" data			
Linear regression		***Mann-Kendall***	
Value	p-value	**Value**	p-value
-0.02	0.003	**-0.096**	0.62

It highlights the fact that parametric tests applied over a series of measurements provide more statistical power than its non-parametric counterpart over the same data set. On one hand, non-parametric tests have the advantage that no assumptions are required about the shape of the underlying distribution. However, it comes with a price, this is the reason why parametric tests do provide more power with a fewer number of measurements.

Strategies to identify deterioration of the groundwater quality

The primary goal when measuring the ecological status of a groundwater body is the identification of upward trends and the detection of significant impacts on groundwater quality. In some environmental programs, it is recommended that any test performed should have the ability to detect increases between three and four standard deviations above the background data. The baseline corresponds to the expected value of a data set, whereas comparisons for assessment have to rely on middle or typical values.

Strategies to detect increases of pollutants are based typically in comparing groundwater-monitoring data to numerical data fixed as a groundwater protection standard. In Annex II of the Groundwater Directive, threshold value is defined as a groundwater quality standard set by Member States in accordance with Article 3. They must be established for all pollutants and indicators of pollution which, pursuant to the characterisation performed in accordance with Article 5 of the Directive, characterise bodies or groups of bodies of groundwater as being at risk of failing to achieve good groundwater chemical status. The determination of threshold levels must be based on the extent of interactions between groundwater and associated aquatic and dependent terrestrial ecosystems, actual or future uses of groundwater and its hydro-geological characteristics.

Several approaches are used to make comparisons between groundwater-monitoring data and a threshold value. The most commonly used is the confidence interval criterion (intervals estimates of a population parameter). The mean of the population is then estimated from the sample, but instead of estimating the parameter by a single value, an interval likely to include the mean is given. They provide an indication of the reliability of the estimate, and are characterised by a particular confidence level (the probability that the true parameter lies within the end points of the confidence interval).

The calculation of confidence intervals around a mean generally requires the assumption that the distribution from which the sample came is normal. A two-way confidence interval around a normally-distributed mean is given by the formula:

$$LCL_{1-\alpha} = \bar{x} - t_{1-\alpha,n-1}\frac{s}{\sqrt{n}} \qquad UCL_{1-\alpha} = \bar{x} + t_{1-\alpha,n-1}\frac{s}{\sqrt{n}}$$

The lower confidence limit is used in compliance monitoring (to verify values that exceed the threshold value), while the upper confidence limit is used during corrective action (trend reversal, for example).

As environmental data are typically intrinsically positive and often highly skewed, the lognormal distribution is a common choice in statistical analysis. In this particular case, confidence intervals can be calculated on the transformed data and then applied to transform the limits of the original concentration scale. The confidence interval obtained in this way is not an interval estimate of the mean, but the median instead.

Other methods to make decisions about ecological status in relation to thresholds are the Shewart-CUSUM control charts. They are parametric methods and no non-parametric alternatives exist, so they must be used only when the normality of data has been checked.

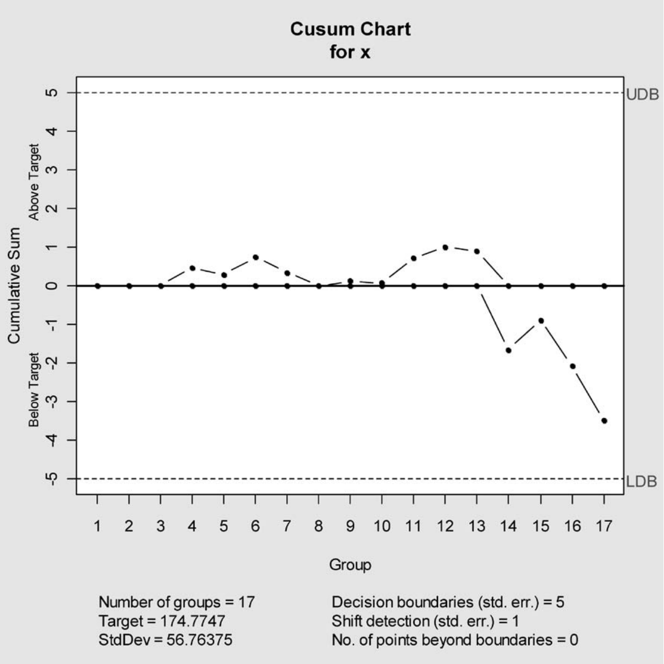


Fig. 6: *Cusum chart for "Plana de Sagunto" data.*

In case data can not be adjusted to a normal distribution or normalized, non-parametric confidence intervals around the median can be calculated. As illustrated in figure 5, bootstrap and resampling methods can be used to derive confidence intervals around the mean. Despite the fact that they are computer intensive methods, the advances of modern computing power make statistical inference a simple task. In addition, they provide more accurate results in some situations than traditional statistics. While statisticians prefer the bias-corrected and accelerated (BCa) bootstrap (Efron, 1987), a parametric bootstrap approach might be preferred for small data sets.

Moreover, when a fraction of the monitoring data falls below detection limits, two different approaches can be used. Each non-detect value can be substituted by half of its detection limit, or statistical techniques to handle this type of data can be used. Non parametric confidence intervals are a choice technique in these situations. If the non-detect fraction of the updated baseline is no more than 50%, robust regression on order statistics and Kaplan-Meier method can be utilized.

Nevertheless, in line with the Groundwater Directive, threshold values have to be defined for just a few substances (the minimum list of pollutants and their indicators for which Member States must establish threshold values in accordance with article 3). In case no threshold value has been defined, another strategy must be accounted for. Based on data selected from the baseline, prediction limits for future data or sample statistics generated from the background population with a pre-specified confidence level (1-α) can be elaborated. The procedure consists of collection of one or more data for testing compliance data with the baseline. Although different sampling strategies can be used, like 1-of-m prediction limits on future observations, a prediction limit strategy based on future means may be more environmentally protective than the former (U.S. EPA, 2009).

Limitations and uncertainty

When monitoring data are analyzed, i.e. contaminant concentration, some constraints must be explicitly taken into account. Measurements are not exact values, but the obtained value is composed of the measurement itself, the margin error and the confidence level. That is, we could assert that the concentration of a specific substance in a given groundwater body is 27µg plus or minus 0.1µg with 95% level of confidence (Helsel and Hirsch, 2002). It clearly indicates the necessity of incorporating uncertainty in data analysis, and calls for the application of tiered approaches. As stated by the Bridge Project (Müller et al, 2006), "A tiered approach allows the effort to be proportional to the risk involved," meaning: greater risks for greater effort. Thus a tiered approach supports a practical and cost-effective way forward.

Hydrogeochemistry is a very important source of variation within groundwater bodies, as interactions between water-bearing rocks and the water itself generate different water quality sectors, and this fact will determine the natural evolution of every particular system. The effect of interactions between hydraulically connected aquifers and surface water is to be considered when looking for a detailed description of heterogeneities. This factor is particularly important, for example, when dealing with coastal aquifers, where a freshwater and saltwater interface exists. Interactions between both types of waters and geological formations must be analyzed, as seawater intrusion is a dynamic and three-dimensional process that creates water quality variations on both horizontal and vertical scales. That implies more dense monitoring networks than required for continental aquifers.

Obviously, pollution processes are contributing factors to spatial and temporal heterogeneities within groundwater bodies. As a result, a change in mean levels in groundwater can be associated with natural variability or human activity.

To define particularities and heterogeneities of groundwater systems, monitoring networks are a key factor, along with information coming from well logs. Nevertheless, a proper design, according to the typology and characteristics of investigated systems, is needed. The importance of building a conceptual model coming from a detailed investigation must therefore be highlighted.

An important but frequently overlooked assumption of statistical methods like linear regression is that observations are independent and identically distributed [i.i.d.]. Unfortunately, due to the nature of the parameters being monitored, much of the data from one monitoring period are not independent of the preceding measurement. To tackle with this problem, many authors (ASTM, 2004) have recommended that sampling be conducted no more often than quarterly to avoid temporal dependence.

Another uncertainty factor is the size of the baseline. Depending on the monitoring frequency, the number of measurements to elaborate the baseline can fluctuate. In a common scheme of semi-annual sampling, four measurements are available from years 2007 and 2008, but in some other cases just two measurements are accessible. If this is the case, little can be done except waiting for additional data to become available. Then, the baseline value can be recalculated. In this way, the precision of estimation can be improved and the length of the confidence interval reduced (the wider the confidence interval the less the precision).

Once the baseline has been calculated (considering historical data when available), it is necessary to choose a statistic for compliance with its value. As explained before, if the baseline data set follows normal distribution, parametric prediction intervals can be executed,

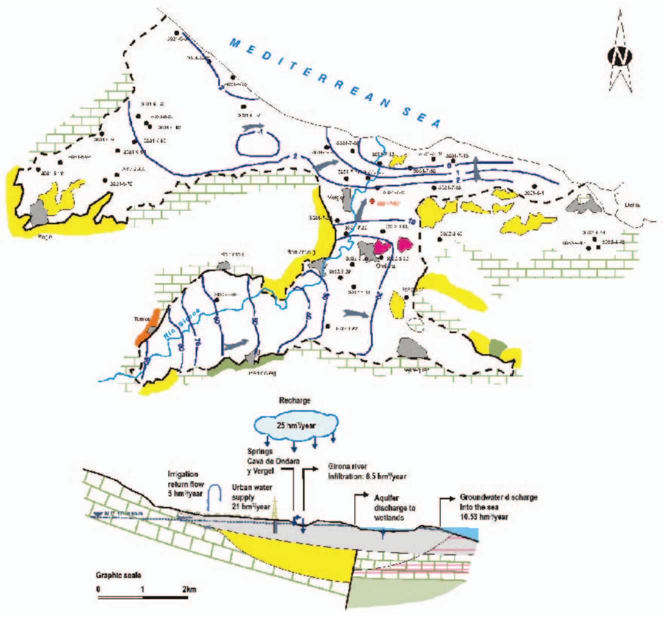


Fig. 7: *Conceptual model and heterogeneities of an aquifer at the Mediterranean coast (B. Ballesteros et al, 2001).*

with the usual advantages gained when the shape of the distribution is known, like power and number of measurements required. In this case, a minimum of eight measurements is desirable for an accurate analysis, and more if normality can not be proved.

If the data do not follow a normal distribution and can not be normalized via log-transformation, non parametric prediction intervals can be computed from the data. The minimum requirement is 4 measurements for the elaboration of a non-parametric prediction interval for a future median of order 3 (in this way the power of the test is maximized). First, we could select a high order statistic (like the maximum) to fix a limit. This, in turn, has the disadvantage that, in cases where the variability is very high, the maximum value can result in an unacceptable high value. On the other hand, the limited number of values restricts the capacity to obtain solid conclusions.

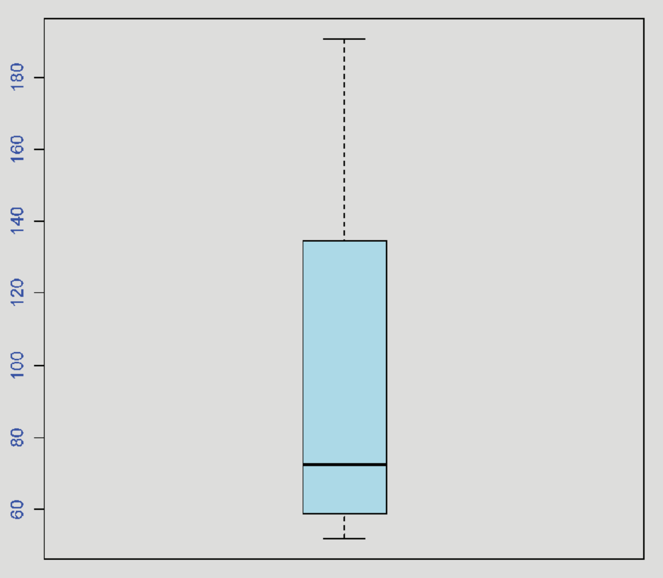


Fig. 8: *Box plot of base line values.*

As shown in the unified guidance for statistical analysis of groundwater monitoring data elaborated for the U.S. Environmental Protection Agency (U.S. EPA, 2009), if the order statistic selected for comparison is the maximum value, then the confidence level on a median of three future values is set to 0.875, which can be enough for a wide range of situations. For the Plana de Sagunto groundwater body, it means that, if the median of three future values is under 195,6 mg/l of nitrate concentration, a significant statistical impact can be declared with respect to baseline (assuming no threshold had been defined).

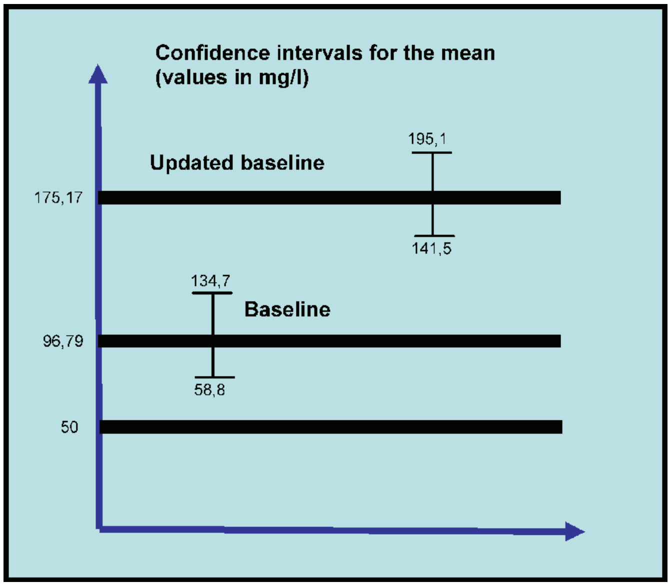


Fig. 9: *Summary of key statistics for "Plana de Sagunto" data.*

The term exogenous variable is used for factors whose value is independent from the other variables included in the model. The purpose of including the effect of exogenous variables into a particular model is to reduce noise due to external influences into the data being examined, and to also make the conducted tests more powerful.

Seasonality is one of the major sources of variation of the dependent variable (concentration) and its effect must be removed to better discover trends and improve power of tests. Techniques for dealing with seasonality fall into three major categories (Helsel and Hirsch, 2002). One is fully nonparametric, one is a mixed procedure, and the last is fully parametric.

The effect due to variability of precipitation or by the irrigation season has been proved to be the driving force behind seasonal changes in water quality. Water quality data from Plana de Sagunto can be differentiated into wet season and dry season (Figure 9), and review of time series plots depict regular patterns in the data that correspond to both seasons. A Wilcoxon rank sum test applied to Plana de Sagunto data shows a statistical significant difference between wet and dry seasons.

Wilcoxon rank sum test	
Statistic value	p-value
W =51	0,026

DOI 10.4409/Am-001-10-0003

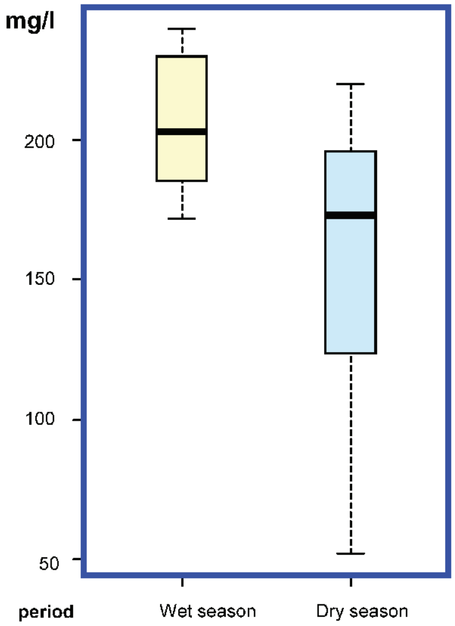


Fig. 10: *Seasonal effects for "Plana de Sagunto" concentration data.*

Discussion and conclusions

The fundamental goal of a monitoring program to assess ecological status of groundwater bodies must be based on the design of strategies for the detection of increases in concentration of the substances that could make the groundwater body declared at risk of not achieving good chemical status.

This is mainly accomplished by means of comparison between past and present data. The baseline value (as defined by the Groundwater Directive) is the starting point for such a comparison. The quality of the data used for the establishment of the baseline is then crucial for a successful statistical monitoring program.

On the other hand, exploratory data analysis techniques are a key tool to gain insight into data, and are the first step before formal statistical tests are used for interpreting experimental data. They are mostly based on data visualization, and must include smoothing techniques for exploring and understanding patterns and trends.

Based on the previous analysis, a reference period has to be chosen. The design of statistical methods for determining whether baseline concentrations (or updated baseline concentration when available) have been exceeded, are the basis for such a comparison. As the baseline will be used for statistical testing, the addition of historical data when possible is desirable in order to enlarge the baseline and incorporate natural variability that cannot be seen within just a few values.

Formal statistical tests have to be designed afterwards with the aim of providing comparable results (as required by the Groundwater Directive). Trend testing is the first issue to be checked, while confidence limits and prediction intervals are the techniques used to check compliance with threshold values and detect statistical significant impacts on groundwater quality.

The baseline represents a central tendency value, so assessment has to rely on the mean. The mean of the population is then estimated by confidence intervals. Its calculation generally requires the assumption that the distribution from which the sample came is normal. If data can not be normalized or the fraction of non-detects is very high, robust regression techniques or methods to handle censored data must be employed.

When no threshold value has been established, the approach must be based on the construction of prediction intervals and on the analysis of trends. The former is a suitable way to evaluate statistical significant impacts on groundwater quality. They can be based on future single values or preferably on future means (or medians), to be tested against baseline. The latter provides indication about sustained upward trends in the concentration of pollutants.

A relevant aspect, emphasized in the Nitrates Directive, is the time lag between measures to improve water quality and responses in quality of a groundwater body. The elaboration of a good conceptual model of the groundwater body at risk is a key issue. Findings based on an incorrect conceptual model, or on no model at all, can produce results in an opposite direction than expected.

The selection of an appropriate monitoring network and an adequate monitoring program are essential aspects in groundwater quality monitoring programs. In this sense, geostatistics is an optimum tool for the design of control networks.

In order to perform the analysis of data sets the statistical language, R has been selected as the reference tool. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form.

As a final conclusion, the use of strategies based in trend assessment combined with the calculation of confidence limits (when a threshold value has been defined) and prediction limits for future means or medians (when no threshold value is available) is a powerful tool for decision making when assessing the ecological status of a groundwater body.

Acknowledgements: We are grateful for the financial support and the data given by the Ministry of the Environment and Rural and Marine Affairs of Spain in the context of a specific agreement between the Ministry and the Spanish Geological Survey (IGME).

References

American Society for Testing and Materials (2005). Standard Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs (D6312-98[2005])

American Society for Testing and Materials (2004). D7048-04: Standard Guide for Applying Statistical Methods for Assessment and Corrective Action Environmental Monitoring Programs. West Conshohocken, PA: ASTM International.

Blum, A., Pauwels, H., Wenland, F. And Grifioen, J. (2009). Background levels under the Water Framework Directive. Groundwater Monitoring (Water quality measurement series). John Wiley and Sons, Chipelham, Wiltshire, p.p. 145-153. ISBN: 978-0470-77809-8.

B.J. Ballesteros, L. Rodríguez, J. López, J. Grima, O. García, J.R. Cobos, J.D.Gómez and J.A. De la Orden (2001). Análisis y ordenación de recursos hídricos de la Marina Alta (Alicante). Alternativas y Directrices. Vol II. Evaluación de recursos subterráneos. Instituto Geológico y Minero de España-Diputación Provincial de Alicante. Fondo documental IGME. 179 pp. Alicante.

Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey, (1983), Graphical Methods for Data Analysis, Wadsworth

Cleveland, W.S. (1979), \Robust locally weighted regression and smoothing scatter-plots", Journal of the American Statistical Association, V. 74, pp. 829-836.

Efron, B. (1987). "Better Bootstrap Confidence Intervals". Journal of the American Statistical Association 82: 171–185. JSTOR 2289144

Müller, D., Fouillac, A.M., Hart, A. and Quevauviller, Ph. (2006), European Groundwater Conference. Presentation Abstracts – Chapter 1.3.4

Gilbert RO (1987). Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold.

Grima, J.; Luque, J.A.; Varela, M.; Sánchez, I. (2010). Trend of contaminants in groundwater bodies of intercommunity Spanish basins. European Groundwater Conference, Madrid, 20-21 may, poster session and proceedings.

Grima, J., Chacón, E., Ballesteros, B., Rodríguez, R. and Mejía, J.A. (2009). Communication of groundwater realities base don assessment and monitoring data.Groundwater Monitoring (Water quality measurement series). John Wiley and Sons, Chipelham, Wiltshire, p.p. 395-407. ISBN: 978-0470-77809-8.

Helsel, D.R. and Hirsch (2002), R.M. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation. Statistical Methods in Water Resources. Environmental science and pollution research international, 14(5):297-307.

K. Hornik (2008). Frequently Asked Questions on R. {ISBN} 3-900051-08-9. http://CRAN.R-project.org/doc/FAQ/R-FAQ.html

Kaplan EL & Meier P (1958). Non-parametric estimation from incomplete observations. J. Amer. Stat. Assn, 53, 457-481.

Luque, J.A.; Grima, J.; Varela, M.; Sánchez, I. (2010). Searching for trends in data from groundwater monitoring networks through exploratory data analysis. European Groundwater Conference, Madrid, 20-21 may, poster session and proceedings.

Tukey, J. W. (1977) Exploratory Data Analysis. Addison-Wesley, Reading, MA.

U.S. Environmental Protection Agency (2009). Office of Resource Conservation and Recovery,. Unified guidance for statistical analysis of groundwater monitoring data at RCRA facilities

U.S. Environmental Protection Agency (1986). Resource Conservation and Recovery Act (RCRA) Ground-Water Monitoring Technical Enforcement Guidance Document. Washington, DC: USEPA. OSWER-9950.1.