

Model-based objective methods for the estimation of groundwater geochemical background

Zoran Nakić, Kristijan Posavec and Jelena Parlov

Abstract: The importance of background levels is clearly stressed by the GWD for the reason that wherever elevated background levels of substances or ions or their indicators occur due to natural hydrogeological reasons, these background levels in the relevant body of groundwater shall be taken into account when establishing groundwater quality standards. However, to accurately determine background concentrations, free of anthropogenic influence, it is important to take into consideration original data quality for background analysis, usability of method(s), representation of the background – is it true or *ambient background values* of concern, any potential sources of contamination near the sampling points etc. Traditionally, geochemical background is often regarded as a fixed value (mean or median), that represents a hypothetical background concentration without taking into account natural variability. However, it was proved in many cases that background groundwater quality very often varies both in space and time as a result of the variations in the climate, rainfall composition, hydrogeological and hydrogeochemical processes, as well as due to the impact of human activities. In the paper it is demonstrated that background values of selected inorganic parameters should be determined on the scale of the particular study because these values are site specific. To treat adequately multimodal or skewed data distribution, resulting from more than one processes involved in the generation of the original data set, the model-based objective methods, the *iterative 2- σ technique* and the *calculated distribution function*, were used. Geochemical data from the groundwater of Samobor aquifer were subjected to the quantitative partitioning by which outliers were removed from the original data sets. Ambient background concentrations are then defined, resulting from the long term human impact on the groundwater quality in this area.

Keywords: groundwater geochemical background, groundwater threshold, normal distribution, iterative 2- σ technique, calculated distribution function, Samobor aquifer

Zoran NAKIĆ 

Kristijan POSAVEC · Jelena PARLOV

University of Zagreb, Faculty of mining, geology and petroleum engineering,
Pierottijeve 6, 10000 Zagreb, Croatia

Tel : (+385) 155 35 703

Fax : (+385) 148 36 053

Zoran NAKIĆ 

zoran.nakic@rgn.hr

Kristijan POSAVEC

kristijan.posavec@rgn.hr

Jelena PARLOV

jelena.parlov@rgn.hr

Received: 17 may 2010 / Accepted: 14 june 2010

Published online: 30 june 2010

© Scribo 2010

Riassunto: La recente Direttiva Quadro sulle Acque Sotterranee mette in evidenza la necessità di distinguere concentrazioni naturali di soluti, dalla concentrazione dei medesimi prodotti dagli effetti dell'antropizzazione. Livelli di "rumore di fondo" naturale, relativamente a sostanze, ioni o loro indicatori, devono essere prese in considerazione quando si stabiliscono gli standard di qualità. Una valutazione delle condizioni naturali di fondo è difficile a causa della lunga storia dell'influenza antropica, sia a causa della diretta immissione di inquinanti nelle acque sotterranee sia per l'effetto indiretto degli emungimenti che inducono un cambiamento nel sistema idrodinamico dell'acquifero. È importante capire che acque sotterranee prive degli effetti dell'influenza antropica di fatto non esistono più, principalmente negli acquiferi superficiali. Alternativamente potremmo definire dei valori ambientali di background tali da considerare gli effetti dell'immissione di lungo periodo di contaminanti antropogenici, che abbiano leggermente innalzato la concentrazione naturale nelle acque del sottosuolo. Sebbene il "rumore di fondo geochimico" sia spesso considerato come un valore fisso senza prenderne in considerazione la naturale variabilità, prevale, al giorno d'oggi, una nuova idea che riflette il fatto che la chimica di base delle acque sotterranee varia nello spazio e con la profondità in relazione a molti complessi fattori, clima, tipo di suolo, geologia, tempo di permanenza. Sono stati sviluppati i così detti metodi oggettivi basati su modello secondo il presupposto che la distribuzione di frequenza di una popolazione di background in un ambiente geologicamente omogeneo segua una distribuzione normale che risulta dalla sommatoria dei processi che hanno prodotto il background stesso. In questo articolo, è stata data enfasi a due metodi quantitativi per la partizione anomala ed i dati di background: la tecnica iterativa 2- σ e la funzione di distribuzione calcolata che mirano a definire background e soglia tendendo ad un range normale. Entrambi i metodi sono applicabili a distribuzioni asimmetriche e la funzione di distribuzione calcolata può essere applicata perfino quando la distribuzione è multimodale. La loro applicabilità è stata testata usando set di dati geochimici ottenuti dal campionamento delle acque sotterranee in pozzi di monitoraggio nell'acquifero di Samobor, che è situato vicino alla città di Zagabria, capitale della Croazia, in un'area caratterizzata da un'agricoltura intensiva. Lo sviluppo dell'industria e la rapida crescita della città di Zagabria e l'urbanizzazione dei dintorni hanno alterato la qualità e la quantità delle acque sotterranee dell'acquifero. Per le analisi dei valori di background e dei valori di soglia, sono stati selezionati cinque parametri chimici presenti naturalmente nelle acque sotterranee: nitrati, cloruri, solfati, ferro e manganese, che possono essere indicatori delle attività antropiche, ma possono pure costituire indicatori di cambiamento delle condizioni redox. I livelli dei valori ambientali di background ed i valori di soglia, determinati come alto limite superiore della normale fluttuazione dei valori di background, sono stati calcolati per tutti i parametri selezionati, ed il range normale dei valori di background è stato verificato mediante l'uso del test di Lilliefors. È stato dimostrato, in questo studio, che gli intervalli della concentrazione di background sono molto meno ristretti a confronto ai ranges dei valori di background negli acquiferi europei. Da ciò si conclude che i valori di background e di soglia di sostanze naturalmente presenti negli acquiferi debbano essere determinati a scala di studi di dettaglio sito-specifici.

Introduction

Groundwater, a highly sensitive and vulnerable medium, strives to accommodate its physical and chemical properties to continuous environmental perturbations driven by the interaction of complex natural, or geogenic, as well as non-natural, or anthropogenic influences. Groundwater quality is determined by the sum of soil-modified atmospheric inputs plus water-rock interactions taking place at the soil-bedrock interface, and from longer-term reactions taking place along flow paths in the saturated zone.

A clear distinction between geogenic and anthropogenic influences in groundwater quality has always been of great importance in environmental investigations, with objectives involving decisions on boundary concentrations, to function as a separation criterion between the two. This need to distinguish between natural and man-made concentrations in groundwater is particularly reflected in the new Groundwater Directive (2006/118/EC), which introduces the term background level as: “the concentration of a substance or the value of an indicator in a body of groundwater corresponding to no, or only very minor, anthropogenic alterations to undisturbed conditions”. However, Hawkes and Webb (1962) introduced it into scientific terminology much earlier as the normal abundance of an element in barren earth material.

An assessment of the natural background condition is difficult because of the long history of anthropogenic influence through the direct input of pollutants. However, groundwater quality changes are not only due directly to human impacts but may also be brought about indirectly by pumping-changing the aquifer system hydrodynamics, conditioning displacement of saline and redox fronts, seawater encroaching, up-coning from depth, artificial recharge variations, and mixing (Edmunds and Shand, 2008). Furthermore, elevated levels of substances, ions, or their indicators might occur due to entirely natural processes, the result of geochemical conditions existing in aquifer, or due to the specific geology in the area. From this point, it is evident that groundwater geochemistry may be defined as a mixture of element populations with variable concentration ranges, reflecting dynamics of geogenic and anthropogenic processes. The emphasis on the pattern-process interaction as one of the inherent characteristics of open natural systems is vital, implying that a relationship between the background and non-background values can be observed through the lens of system response to external stimuli (Peh et al., 2010). This reaction may be linear, equilibrium-centred and thus self-stabilizing, which is reflected by the small disturbances of the element concentrations in the natural processes of weathering, transport and deposition of the materials along the flow paths. As a result, element concentrations fluctuate more or less uniformly around some central value. As opposed to the steady-state equilibrium between input and output phases in the open system, non-linear reaction is far removed from equilibrium in both spatial and temporal domain. It is evidenced by the major changes of groundwater quality, brought about by anthropogenic influences, which drive the system out of balance beyond the stability threshold.

The nature of linear reaction, expressed as uniform fluctuations of concentrations around central values, reflects the idea of geochemical background, which revolves around the normally distributed set of values. This concept was recognized in early days of regional geochemistry, when the formula $MEAN \pm 2 \cdot SD$ was frequently used to identify a certain proportion of “unusually high” values in a data set for further inspection (Reimann et al., 2009; Hawkes and Webb, 1962). In recent studies, the $MEAN \pm 2 \cdot SD$ rule has been extensively criticized in that it cannot provide a relevant background estimate (see Matschullat et al., 2000; Reimann and Filzmoser, 2000; Reimann et

al., 2005). It was argued that outliers, values belonging to a different population and originating from another process or source (Hampel et al., 1986; Barnett and Lewis, 1994; Reimann et al., 2005), could occur anywhere within a background data distribution, not only at the extremes of normal distribution. However, it is absolutely clear that geochemical backgrounds cannot be regarded as a fixed value (mean or median) that represents a hypothetical background concentration, without taking into account natural variability. The background chemistry of groundwater varies both spatially and with depth, due to many complex factors, including climate, soil, type, geology and residence time (Shand and Edmunds, 2008). It is more realistic to view it as a range of values rather than as an absolute value.

As announced in the title, the primary objective of this paper is to use model-based objective statistical methods to estimate the geochemical background range of selected inorganic substances in groundwater sampled in an alluvial aquifer in northern Croatia. The theory behind these methods refers to the use of a quantitative approach for the partitioning background data in the conceptual geochemical model framework, which recognizes that a background population has a characteristic probability density function that results from the summation of processes that have produced the background substrate.

Methods of geochemical background estimation

Many approaches can be used in defining background chemistry of groundwater, but essentially two types of methods are most often applied in evaluation of the geochemical background irrespective of the medium under survey.

Very often, particularly in the environmental and ecological studies, the so-called geochemical methods are employed, which make use of various geochemical data normalization procedure (Peh et al., 2010). The geochemical methods are based on an interpretation of the background value of an element or compound from individual samples or sample profiles. It is also important to know additional parameters, such as: pH value, carbon and sulfur content, and content of individual isotopes. Such an approach demands an expertise in geochemical behaviour of the researched element under the fully-determined natural conditions, including the paleo-environmental conditions. Geochemical modelling can give the overall picture of the groundwater quality evolution, in view of quantitative understanding of the geochemical processes, and their coupling to groundwater flow and transport modelling; however, very detailed data and knowledge on hydrodynamics, geological and hydrogeological properties along the flow paths are needed. The background value is most often determined as a fixed value (mean value or median) that represents a hypothetical background concentration without taking into account the natural variability (Matschullat et al., 2000). Rentier et al. (2006), who applied a quality assessment system that considered the patrimonial status of groundwater, have followed this approach, for qualifying the general status of groundwater body, in accordance to the requirements of the EU Water Framework Directive (WFD). They estimated fixed natural geochemical backgrounds as reference values for a selected number of parameters, from analytical detection limits (organics), as an expert judgment value (nitrates) or according to the local Soil Conservation Law (minerals and metals).

However, in scientific circles of late growing, importance is being given to statistical methods as a powerful tool for assessing more realistic numerical values (ranges and limits) (Peh et al., 2010). The use of statistics in background assessment is supported by the fact that the background chemistry of groundwater varies both spatially and with depth due to many complex factors including climate, soil,

type, geology and residence time. As a result, groundwater geochemistry is represented by a mixture of element populations with variable concentration ranges, which reflect dynamics of geogenic and anthropogenic processes. At the practical level, the background needs to relate to a specific water body with values related to the controlling geochemical processes. A subdivision of the water body (e.g. into confined and unconfined, oxidising or reducing) may be needed in order to adequately define background values, which change both regionally with the basic geology, as well as locally with the type and genesis of the overburden. In specific circumstances, in order to interpret heterogeneities within groundwater bodies and to define range of background concentrations due to natural variability of groundwater chemistry, a combination of approaches, e.g. statistical methods used in combination with geochemical modelling along the flow paths, have to be considered.

In environmental pollution problems, the threshold value - the boundary between background variation and extreme values in a normal distribution - is very often needed for defining an action level or clean-up goal. In "selection of the best statistics" for threshold and background assessment, the following issues should be taken into consideration: original data quality for background analysis, usability of method(s), scale of the particular study, and any potential sources of contamination near the sampling points. Furthermore, because of the inherent spatial component of applied geochemical and environmental data, and the difference between "extreme values" and "outliers", the estimated threshold values separating background from outliers will usually change when size and/or location of the survey area are changed (Reimann et al., 2009). It is absolutely clear that methods that do not consider the spatial data component or form of the data distribution will not be able to provide a reliable estimate of threshold and background; however, some of the most-widely-used statistics are highly subjective and non-reliable. Stanley and Sinclair (1989) proposed a classification of threshold selection techniques, based on three main categories: 1) experiential; 2) model-based-subjective; 3) model-based objective.

Experiential methods of threshold selection depend on the experience of researchers and include the widely used techniques of evaluating tabulations of data or visual evaluation of histograms (Sinclair, 1991). These methods are not adequate for general use where decisions must be made based on comparisons (e.g. anomaly contrast), or where it is of critical importance to classify each individual sample as being either anomalous or background. A combination of exploratory data analysis tools (e.g. histogram, box plot, 1D and 2D scattergram) are still indispensable in environmental research and are always recommended as a first step in analysis in order to give an insight into the data structure and possible data errors.

Model-based subjective methods apply some type of formal statistical or mathematical model (including the "Mean \pm 2SD" rule) to a set of selected geochemical values in order to define background and threshold values. However, applying these methods, relevant background and threshold estimates cannot be provided because, under the assumption that data follow normal distribution, it is wrongly assumed that the upper 2.5 percent of the data are anomalous and belong to outliers. In reality, this percentage rule is of limited use because the real percentage of outliers, which are generally observations resulting from a secondary process, and not extreme values from the background distribution, could be very difficult to assess, particularly if multiple background (or anthropogenic) sources give rise to very complex data distribution pattern. If there is no assumption about the form of the distribution, then the value of mean plus two standard deviations has no meaning and is simply some small,

variable proportion of highest values of data set (Sinclair, 1991). To accommodate the special properties of groundwater chemical data, it is possible to use robust-to-outlier statistical methods, like the "Median \pm 2 MAD" procedure or box-plot upper inner fence calculation. These methods, which are based on median as the least biased measure of central tendency and are independent on assumptions of normality, are recently often used in exploratory data analysis to estimate the thresholds of background variation and to detect outliers (see Peh et al., 2010; Reimann et al., 2005).

Model-based objective methods differ from subjective methods only in that the thresholds are defined by the data themselves rather than by an arbitrary decision of the researcher (Sinclair, 1991). As an underlying assumption, conceptual geochemical models of the area under investigation must be developed, which recognize that a background population in a homogeneous geological environment has a characteristic probability density function that results from the summation of the processes that have produced the background substrate. Cumulative probability plots are particularly useful because chemical elements, such as metal concentrations in water sample, can be divided into two or more populations separated by inflection points on plots. Separation (partition) of the background and anomalous population is based on the assumption of normality regarding the form of the probability density functions, and the concentrations at the inflection points are defined as threshold values. In general, it is expected that the probability density functions for particular elements are different in anomalous and background samples. The problem is to define that difference with confidence, but it can be increased by a high degree of understanding of data and the underlying processes. Although cumulative frequency plots are favoured in background studies to discriminate between background and anomalous concentrations resulting from human impacts, care must be taken in interpretation of results, because the outliers may occur as a consequence of the anomalous natural concentration (e.g. due to mixing of the water of different origin, salt water intrusion etc.) (see Edmunds and Shand, 2008). Furthermore, redox reactions, sorption of solutes on solid mineral or organic phases and/or precipitation may alter data distribution by removing or limiting concentrations in solution.

The quantitative partitioning of anomalous values from data sets has the advantage in treating multimodal or skewed data distribution, in comparison with cumulative probability plots, which belong to graphical tools. The reason is that even probability plot approaches contain elements of subjectivity, in that threshold selection is subject to manual procedure and depends also to some extent on the experience of the researcher. Sinclair (1991) stated that the determination of the threshold must be viewed as an estimation procedure in the statistical sense, subject to random and systematic error. Further important limitations regarding the probability plot approaches are that a minimum of ~ 100 values is needed for the calculations (Panno et al., 2006). As opposed to the different graphical tools, quantitative methods for partitioning data remove outliers in ways that guarantee the normal distribution of the remaining data subset. The approach to the normally distributed set of data is iterative, with inspection and testing if the number of anomalous samples is greater than or equal to background samples. *Trimming method*, as described by Peh et al. (2010), is based on truncating the upper tail of the positively skewed distributions of variables with simultaneous testing of the remaining subset for normal distribution. The process is repeated until the distribution conforms to Gaussian law while the highest value after the assumption of normality is accepted as the upper limit for the background.

In this paper the emphasis will be given to the two quantitative methods for partitioning data i.e. the *iterative 2- σ technique* and the

calculated distribution function. Both methods aim at defining the background and threshold by approaching a normal range. These methods take an actual set of measured data for further partitioning until the final sub collective density distribution approaches normal distribution. As a rule of applying these methods, the anomalous population or populations, which are expected to be represented by high values in the original data set, need to be represented by a small proportion of data. Conversely, background values need to amount to a large proportion of the total data set.

The *iterative 2- σ technique* constructs an approximated normal distribution around the mode value of the original data set (Matschullat et al., 2000). Threshold value is calculated as the outer limit of background variation, aiming to determine the outliers above as well as below the lower limit of normal background fluctuation for a particular variable. This method can be applied to the often-ignored situation where negative anomalies result, e.g. in case of low values of dissolved oxygen that indicates extreme oxygen consumption in an aquifer. It is applicable to unimodal and skewed distributions, but its disadvantage is that it cannot be applied when the distribution is multimodal. If the distribution is, for example, bimodal, the mean might fall between the two peaks of the distribution and the large value of the standard deviation causes an overestimation of the background range (Nakić et al., 2007).

The *calculated distribution function* specifically aims at defining the upper limit of normal background concentrations. It is convenient for use if human activity may lead to changes in water quality caused by pollutants, causing positive anomalies shown in a distribution function as positive asymmetry of a normal curve. The lower values should thus be free from anthropogenic influence (Matschullat et al. 2000). An advantage of this method over the *iterative 2- σ technique* method is its applicability to the overlapping multimodal distribution if the data in the background population, which represents lower values in the original data set distribution, are dominant.

A successful application of both methods does not require normally or log-normally distributed total (combined) data and they can be applied to relatively small data sets ($n > 30$) (Nakić et al., 2007). This threshold of sample size separates small-sample statistics from large-sample statistics, where normal distribution can be used as an approximation (Davis, 2002). Variation devised by Lilliefors (1967) to the use of the Kolmogorov-Smirnov procedure for testing the fit of background data to a normal distribution can be applied if the number of the remaining data (background data), after processing the original data set, equals or exceeds four. The mean and variance of the sample are found and the empirical cumulative distribution is determined from ranked observations as in the ordinary Kolmogorov-Smirnov procedure (see Davis, 2002). A normal cumulative probability distribution is then calculated that has the same mean and variance as the sample. The largest absolute difference between the normal cumulative distribution function and the empirical distribution function is found: this is the test statistic, T .

The test statistic is then compared to critical values of T using statistical values of the Lilliefors test statistic, T , for testing goodness-of-fit to a normal distribution. It is important to emphasize that both methods, the *iterative 2- σ technique* and the *calculated distribution function*, can be also applied to scattered distributions, if the Lilliefors test statistic T is lower than the critical value of T . However, if the number of anomalous samples is greater than or equal to background samples, the Lilliefors test statistic T is greater than the critical value of T and the obtained background range will be overestimated. The main advantages of these techniques over other methods are that they calculate the normal range of background values (true value of background is within

Mean $\pm 2\sigma$ range) with 95-percent confidence (Nakić et al. 2007).

In this paper, the Visual Basic macro BACKGROUND, created by Nakić et al. (2007), was used to calculate background values of selected inorganic substances in groundwater sampled in alluvial aquifer system in northern Croatia. This macro provides information on the normal background concentrations with little effort, using a widely accessible platform (i.e. MS Excel), which also allows data to be plotted and visualized easily. Macro uses algorithms that incorporate both the *iterative 2- σ technique* and the *calculated distribution function* (see Nakić et al. 2007). The first computational step common for both methods is data preparation that consists of storing the data set in a new spreadsheet. In the next steps, calculation procedure differs between two methods. Algorithms based on the *iterative 2- σ technique* are used to calculate mean and standard deviation from the prepared data set. Then the mean $\pm 2\sigma$ range is calculated. The data set is inspected to verify that all values lie within the calculated range. All values that lie outside the range mean $\pm 2\sigma$ are cleared. These steps are repeated until all values lie within the mean $\pm 2\sigma$ range.

Algorithms based on the *calculated distribution function* are used to calculate the median value for the data set, and all values greater than the median are then cleared. In the next step, the macro mirrors all remaining values against the calculated median value in the following manner: *Mirrored value = Median - Data set value + Median*. The macro then calculates mean and standard deviation from the prepared data set and the Mean $\pm 2\sigma$ range. After calculating the Mean $\pm 2\sigma$ range, all values that lie outside the calculated range are cleared. In the next step, common for both algorithms, macro calculates normal cumulative distribution for the calculated mean and standard deviation.

Goodness-of-fit of a background data to a normal distribution, by using Lilliefors test for level of significance $\alpha = 0.05$, is visualized by automated generation of Lilliefors test statistic values (T_{crit} and T) into the spreadsheet. Finally, macro produces a chart showing histograms for original and calculated background values, cumulative distribution for original and calculated background values and normal cumulative distribution.

Study area

Hydrogeological setting

The Samobor aquifer is situated in the western part of the alluvial Zagreb aquifer system (Fig. 1), which is the only, and therefore exceptionally important, source of poTab water for the Croatian capital with a population that has almost reached one million. The aquifer system is of Quaternary age, deposited during the Middle and Upper Pleistocene and Holocene. Quaternary deposits are divided into three basic units: an aquifer system overburden built of clay and silt; shallow Holocene aquifer built of medium-grain gravel mixed with sands; and deeper aquifers from Middle and Upper Pleistocene, with frequent lateral and vertical alterations of gravel, sand, and clay. During the Middle and Upper Pleistocene, this area was under lakes and wetland, while the neighbouring mountain ranges (Mt. Medvednica, Marijagorička Brda and Žumberačko Gorje hills) were the main landmasses susceptible to intensive erosion and denudation. The weathered material was washed down by streams and deposited in lakes and marshes. In Holocene, the Sava River cut its course, and transport of materials from the Alpine regions began (Velić et al., 1999). The material transport was of varying intensities because of intensive climate changes. During the warm and humid periods it was intensive, and the intensity lowered during dry and cold periods. In addition to the climate changes, the deposit processes are affected by tectonic

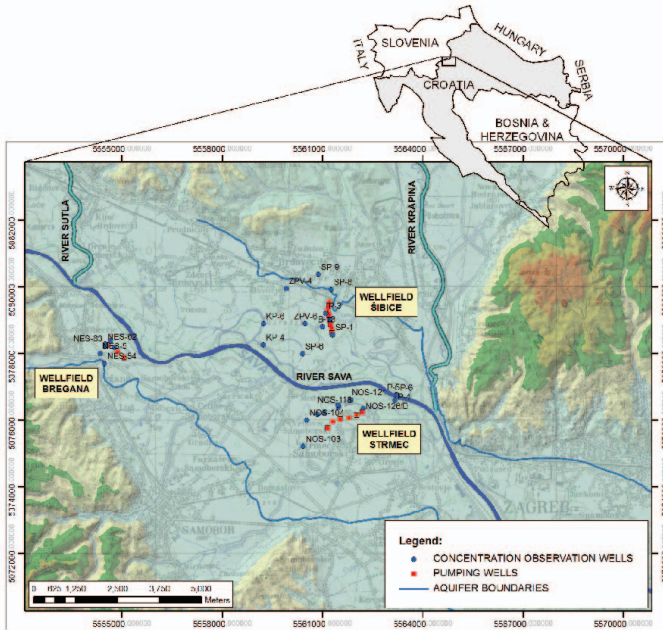


Fig. 1 Map of Samobor aquifer showing concentration observation well locations.

movements that have been present in this area until recently. The consequence of such depositing conditions is a pronounced heterogeneity and anisotropy of the water bearing system. Total depth of the gravelly–sandy Samobor aquifer starts at several meters near the edges and, moving towards the central part, thickness increases to a maximum of 40 m. This aquifer accommodates two important water supply well fields, Strmec and Šibice, located on the right and left bank of the Sava River, respectively (Fig. 1). The overlying clayey–silty aquitard of the aquifer, 0–3 m thick, provides virtually no protection to the aquifer from surface influences. The values of hydraulic conductivity are up to 1,500 m/day (Vlahović et al., 2008).

Regarding hydraulic boundaries of the aquifer, the western and south-western boundary forms the input, and the eastern boundary forms the output. In the north, there is the boundary of the set potential: the Sava River, while in the south there is boundary of zero discharge, which means that the system here transforms into impermeable layers. The Sava River is hydraulically connected with the aquifer, and thus represents the dominant source of groundwater recharge. It infiltrates the system in periods of high and medium water levels and drains it at times of low water levels.

Although groundwater quality in the Republic of Croatia, particularly in the Zagreb area, has been monitored from the very beginning of the public water supply, no systematic hydrogeochemical investigations have been carried out (Vlahović et al., 2008). However, since coming into force of the Water Framework Directive (2000/60/EC), several investigations have been carried out, aiming to characterize the dominant geochemical processes and sources influencing the groundwater chemistry of the Samobor aquifer. Nakić (2001) and Bačani et al. (2002) have used statistical methods-trend analysis and multivariate statistical analyses-in order to discriminate between natural and man-made concentrations on the qualitative basis. Brkić et al. (2003) studied the intrinsic vulnerability of the whole Zagreb aquifer system, while Nakić et al. (2004) isolated the hydrogeochemically homogeneous areas by multivariate statistical analyses and geochemical modelling tools, with the results supporting the stratigraphic evidence showing that it is possible to differentiate a shallow Holocene alluvial

aquifer from the deeper Middle/Upper Pleistocene lacustrine-marsh aquifers. Visual Basic macro Background was used in 2007 (Nakić et al.) and in 2008 (Nakić et al.) in order to calculate geochemical background values of several chemical parameters in the eastern part of the Zagreb aquifer system. However, no attempts have been made so far to use quantitative approaches for the partitioning of the background groundwater chemical data of the Samobor aquifer.

Threats and anthropogenic impacts to groundwater quantity and quality

The Samobor aquifer is situated in the suburban area of the Croatian capital. In this area, intensive agriculture dominates, even near the municipal well fields. Developments of industry and fast growth of the City of Zagreb and small towns in the region have considerably affected quality of groundwater in the aquifer. Quantity of groundwater is diminishing continuously due to negative trends of groundwater levels. Lowering of the groundwater levels is the consequence of excessive pumping at the municipal and industrial well fields, riverbed erosion and prolonged drought periods. Groundwater levels have already reached a minimum on some well fields, causing water scarcity during droughts. Increasingly progressive groundwater pollution in the heterogeneous aquifer has been observed for the last fifteen years (Nakić, 2001; Bačani et al., 2002). The most significant pollution sources are leaky sewerage, a Trebež waste disposal site, located upstream of the well field Strmec, agriculture, illegal waste depositories, illegal gravel pits, and also industrial facilities. High concentrations of pollutants like nitrates, atrazine and heavy metals in groundwater confirm the impacts of pollution sources on groundwater quality.

Field and analytical procedures

As a result of the human influence on the groundwater quality, it can be argued that natural background concentrations of chemical substances probably no longer exist in this area. An option, suggested by Reimann and Garrett (2005), would be to define *ambient background values* under slightly altered conditions, when elevated levels of element concentrations in soil or water are no longer natural and result from long term human impact. A similar approach has been presented by Panno et al. (2006), who introduced the concept of *present day background* in order to include concentrations of human-related contaminants such as NO_3 and Cl that are elevated above pre-settlement concentrations. However, in order to apply the quantitative methods for partitioning ambient background population, special care was taken to select data from observation wells which are as far distant as possible from the point pollution sources, in order to be sure that background values amount to a large proportion of the total data set. Those observation wells that are near to known pollution sources have been excluded from further analysis. In total, 30 observation wells were selected for background analysis from those that were sampled in June 2006.

The groundwater chemical data (June 2006) used in this paper (Tab. 1) have been compiled from the Water Supply Company and Croatian Waters groundwater quality database for the Samobor aquifer. Chemical indicators are analyzed at the laboratories of the Water Supply Company and Institute of Public Health “Andrija Štampar”, by the application of standard methods for water quality determination. Thus, basic cations and anions were determined by ion chromatography and heavy metals by ICP-OES. The following inorganic parameters were selected for further analysis: nitrate, chloride, sulphate, iron and manganese.

Tab. 1 Original data set of selected chemical parameters.

Observation wells	Nitrate (mgN/l)	Chloride (mgCl/l)	Sulphate (mgSO ₄ /l)	Iron (µgFe/l)	Manganese (µgMn/l)
NOS-125/2	0.00	9.60	50.00	681.70	145.80
NOS-125/3	3.19	18.40	47.80	138.60	10.70
NOS-126/D	1.13	8.70	30.60	44.80	0.00
NES-54	5.72	18.60	32.40	4.30	2.60
NES-61	3.00	14.80	39.60	3.30	0.40
NES-62	3.84	20.20	44.20	15.10	1.20
NES-63	4.20	23.60	43.30	8.30	0.30
NES-5	2.41	13.80	50.90	0.00	0.00
NES-14	4.83	26.50	41.50	0.00	0.10
NOS-29A	0.09	1.00	4.10	37.20	76.30
B-13	2.10	16.70	29.60	20.00	5.00
BP-4/1	8.00	15.10	31.10	20.00	5.00
KP-6	2.10	16.40	35.40	20.00	9.70
P-3	1.50	11.80	20.30	838.00	10.00
SP-6	3.00	15.60	42.70	20.00	5.00
SP-8	6.40	20.10	27.60	283.00	5.00
SP-9	0.10	13.00	32.50	311.00	57.90
P-4	6.80	28.90	29.30	20.00	5.00
P-5	7.00	25.30	30.90	20.00	5.00
P-6	5.10	26.30	31.10	20.00	5.00
NOS-118	0.00	7.70	37.50	302.40	92.50
NOS-121	1.94	11.10	31.10	1.40	0.00
SP-1	4.80	22.10	35.00	20.00	5.00
ZPV-4	7.30	16.90	32.20	20.00	5.00
ZPV-6	3.90	13.50	42.00	4030.00	8.50
KP-4	0.30	11.30	56.40	20.00	67.50
NOS-101	0.00	3.90	21.40	1021.00	149.00
NOS-103	4.40	15.10	31.30	3.40	0.10
NOS-104	3.46	26.20	44.90	1.10	0.10
NOS-125/1	0.00	9.90	49.40	591.60	141.80

These naturally occurring chemical elements and compounds in groundwater are indicators of human impacts from agriculture and industry, but can also indicate a change in redox conditions. This is due to, for example, high organic content in water that originates from permeable sewer system, which then infiltrates into groundwater, resulting in the reduction of oxygen or nitrate content due to its consumption for organic matter oxidation, as well as in the dissolution of manganese and iron, or even other metal ions (i.e. lead co-precipitated with iron and manganese oxide).

Results and discussion

Before calculating *ambient background values* by applying described methods, a combination of exploratory data analysis tools (histogram and cumulative distribution function) was used to give an insight into the data structure, possible data errors and outliers, which are the values belonging to a contaminant population. After screening through the data, several error or outlier values were omitted from further analysis. Even the cursory glance at the histograms and cumulative distribution functions of the original data sets (Fig. 2 to 6)

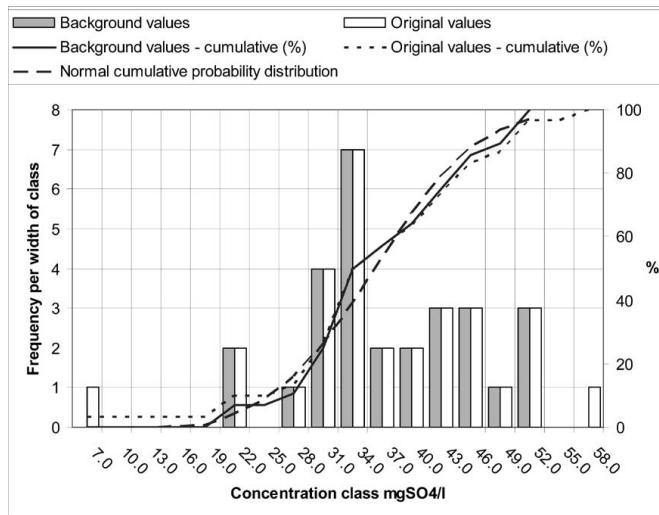


Fig. 2 Ambient background concentrations of sulphate, SO₄ (background levels: 19.6/52.9 mg/l; threshold limit: 52.9 mg/l); Lilliefors test statistic [$\alpha = 0.05$]: $T_{crit} = 0.173$, $T = 0.108$.

reveal the influence of contaminant populations, which is reflected in the positive skewed distribution for sulphate, iron and manganese and in the overlapping multimodal distribution for nitrate and chloride.

Ambient background levels (mean $\pm 2\sigma$ range) and threshold values determined as a higher limit of the normal background fluctuation (mean + 2σ) were calculated for all selected parameters. As stated before, *iterative 2- σ technique* is applicable for skewed distribution and was used to find background for sulphate, iron and manganese (Fig. 2, 3, 4). The *calculated distribution function*, which can be applied even when distribution is multimodal, was used for background calculations for nitrate and chloride (Fig. 5, 6). The range of background and threshold values for selected parameters is shown in Tab. 2.

Goodness of fit of a background data to a normal distribution is tested by using Lilliefors test for the level of significance $\alpha = 0.05$. The Lilliefors test statistic T in all cases, except for iron, was less than the critical value of the test statistic T_{crit} , revealing that calculated *ambient background values* distributions fit the normal distribution well. For the iron test statistic T is equal to the critical value of the

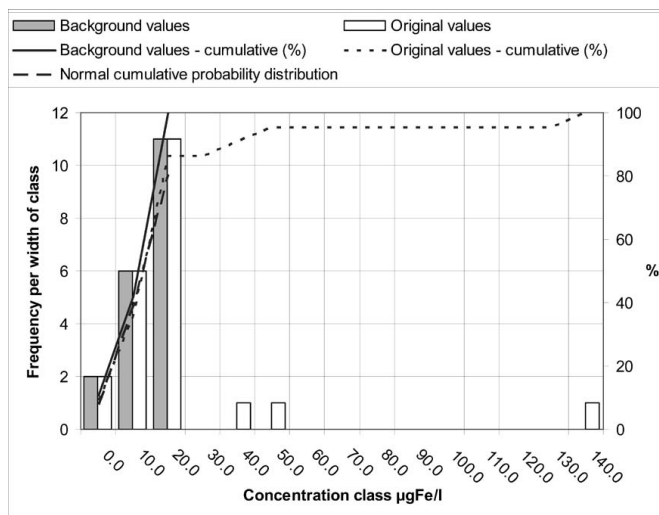


Fig. 3 Ambient background concentrations of iron, Fe (background levels: 0/30 µg/l; threshold limit: 30 µg/l); Lilliefors test statistic [$\alpha = 0.05$]: $T_{crit} = 0.195$, $T = 0.195$.

Tab. 2 Results of geochemical background calculations for selected parameters.

Parameters	Mean	Standard deviation	Background values	Threshold limits
Sulphate (mg/l SO ₄)	36.3	8.3	19.6 – 52.9	52.9
Iron (µg/l Fe)	12.5	8.8	0.0 – 30.0	30.0
Manganese (µg/l Mn)	3.9	3.5	0.0 – 10.9	10.9
Nitrate (mgN/l)	3.1	2.3	0.0 – 7.60	7.6
Chloride (mg/l Cl)	15.4	6.2	2.9 – 27.8	27.8

test statistic T_{crit} , indicating that the range of iron background values is calculated with less than 95 percent confidence.

Sulphate concentrations, in the original data set, approach normal distribution (Fig. 2); however there is a slight positive skew, impacted by agricultural activities on the groundwater chemistry in this area. Range of normal background values for sulphate from *iterative 2-σ technique* is wide; however, when comparing it to the whole range of estimated background values from European aquifers (Shand and Edmunds, 2008) it is evident that it reflects specific natural groundwater chemistry in Samobor alluvial aquifer. Similar evidence can be provided for other elements and compounds under consideration. Iron and manganese distributions in the original data set are strongly skewed (Fig. 3 and 4), although these elements usually have low background concentrations in shallow parts of the aquifers where oxygen is present (see in Shand and Edmunds, 2008). Concentration ranges of these elements increase significantly across redox boundaries and may vary over four to five orders of magnitude across the individual aquifers. Distribution of original data sets for nitrate and chloride is multimodal (Fig. 5 and 6), revealing impacts of diffuse anthropogenic sources, like agricultural activities or atmospheric inputs due to emissions from the burning of fossil fuels. However, ranges of background values for both compounds are much narrower when compared to the ranges of background values of these compounds across European aquifers (see Shand and Edmunds, 2008). It is evident that statistically derived background values from pan-European scale would not be appropriate for the scale of Samobor aquifer. The experience from the EU BaSeLiNe project, funded under the Fifth Framework Programme (see Edmunds and Shand, 2008), showed that it is extremely difficult to discriminate between background

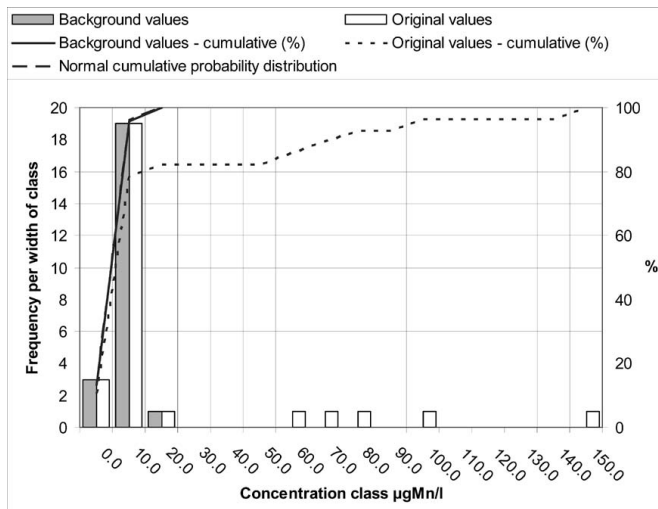


Fig. 4 Ambient background concentrations of manganese, Mn (background levels: 0/10.9 µg/l; threshold limit: 10.9 µg/l); Lilliefors test statistic [$\alpha = 0.05$]: $T_{crit} = 0.190$, $T = 0.006$.

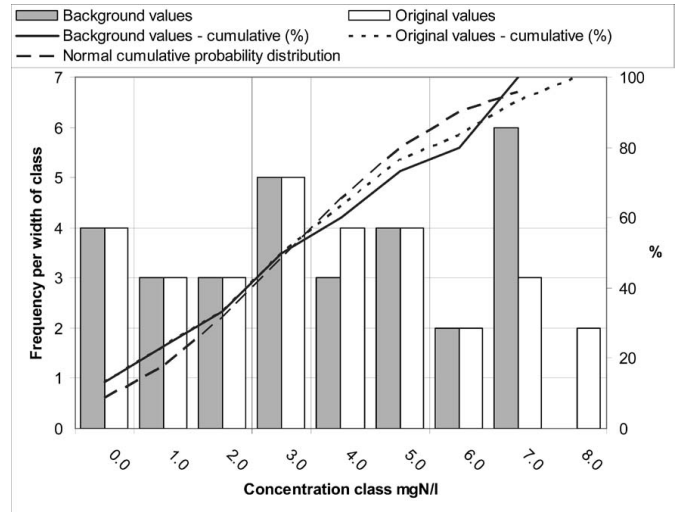


Fig. 5 Ambient background concentrations of nitrate, NO₃ (background levels: 0/7.6 mg/l; threshold limit: 7.6 mg/l); Lilliefors test statistic [$\alpha = 0.05$]: $T_{crit} = 0.161$, $T = 0.100$.

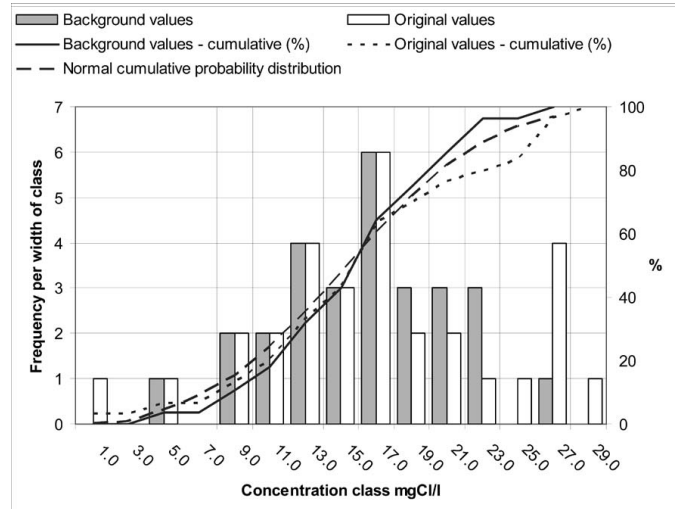


Fig. 6 Ambient background concentrations of chloride, Cl (background levels: 2.9/27.8 mg/l; threshold limit: 27.8 mg/l); Lilliefors test statistic [$\alpha = 0.05$]: $T_{crit} = 0.173$, $T = 0.074$.

groundwater chemistry in different aquifer types, due to the subtle differences in geological structure, lithological and geochemical facies changes, and mineralogy. For these reasons, background values should be determined on the scale of the particular study, because these values are not universal entities in three-dimensional space and may be site- and even depth- specific.

Conclusions

The model-based objective methods, the *iterative 2- σ technique* and the *calculated distribution function*, are proven to be a powerful tool to discriminate between anomalous and background populations. These methods, incorporated in the algorithms used by Visual Basic macro BACKGROUND, were successfully applied for calculating the ambient geochemical background values of five inorganic parameters in groundwater of the Samobor aquifer. It is found that the distribution of elements and compounds under consideration are influenced by more than one natural or pollution process, resulting in skewed or multi-modal distribution. However, geochemical background concentration ranges in site-specific conditions are proven to be much less narrower compared to the ranges of the background values across European aquifers. For this reason, attempting to apply background values determined on a regional or even nationwide scale to a relatively small study area is not a useful approach and may lead to erroneous management decisions. This investigation shows that, in principle, background and threshold values of naturally occurring substances need to be determined on the scale of particular study.

Acknowledgements: The study was supported by the European Community 7th Framework Project GENESIS (226536) on groundwater system.

References

- Bačani, A., Vlahović, T., Nakić, Z. (2002), Endangered groundwater quality of the Strmec well field. In: Ribeiro L (ed) Proceedings. Paper presented at the 3rd International Conference on future groundwater resources at risk, Lisbon, 25–27 June 2001, pp 329–336. ISBN 972-97711-3-8
- Barnett, V., Lewis, T. (1994), Outliers in statistical data, 3rd Ed. New York: John Wiley & Sons Inc. (Cited by Reimann et al. 2005). ISBN 0-471-93094-6
- Brkić, Ž., Kapelj, S., Larva O., Marković T., Vlahović T. (2003), Hydrogeology and hydrogeochemistry in the alluvial aquifer of the Zagreb area. RMZ-M&G Vol 50, No.1: 75–78. ISSN: 1408-7073
- Davis, J. C. (2002), Statistics and Data Analysis in Geology, Third Edition. New York: John Wiley & Sons Inc. ISBN 0-471-17275-8.
- Edmunds, W.M., Shand, P. (2008), Groundwater Baseline Quality. In: Natural Groundwater Quality (eds. Edmunds, W.M. and Shand, P), Blackwell Publishing Ltd. ISBN: 978-14051-5675-2
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel W. (1986), Robust statistics. The approach based on influence functions. New York: John Wiley & Sons. (Cited by Reimann et al., 2005). ISBN: 0-471-73577-9
- Hawkes, H.E., Webb, J.S. (1962), Geochemistry in Mineral Exploration. New York: Harper and Row.
- Lilliefors, H.W. (1967), On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Jour. American Statistical Assoc., vol. 62: 399-402 (Cited by Davis 2002). ISSN: 0162-1459
- Matschullat, J., Ottenstein, R., Reimann, C. (2000), Geochemical background - can we calculate it? Environmental Geology no. 39: 173-188. ISSN 0943-0105 (Print) 1432-0495 (Online) DOI 10.1007/s002549900084
- Nakić, Z., Bačani, A., Vlahović, T. (2001), Antropogeni utjecaj na podzemne vode crpilišta Strmec (Zagreb, Hrvatska) (Anthropogenic influence on groundwater of the Strmec well field (Zagreb, Hrvatska) (in Croatian). Min Geol Pet Eng Mag 13:25–38. ISSN: 0353-4529
- Nakić, Z., Posavec, K., Parlov, J. (2004), Hydrogeochemical model of the heterogeneous aquifer system: the case of Zagreb, Croatia. In: Martinez S, Hergt T (ed) GROUNDWATER FLOW UNDERSTANDING from local to regional scales (XXXIII IAH – 7 ALHSUD), IAH, ALHSUD, Zacatecas City. ISBN 970-32-1749-4
- Nakić, Z., Posavec, K., Bačani A. (2007), A Visual Basic spreadsheet macro for geochemical background analysis. Ground Water 45, no. 5: 642-647. Print ISSN: 0017-467X Online ISSN: 1745-6584 DOI: 10.1111/j.1745-6584.2007.00325.x
- Nakić, Z., Posavec, K., Kontrec, P., Bačani, A., Parlov, J. (2008), Geochemical background values of selected chemical parameters in groundwater of Zagreb aquifer system. In: Taniguchi, M., Yoshioka, R., Skinner, A., Aureli, A. (ed.) Integrating Groundwater Science and Human Well-being, Toyama, XXXVI IAH Congress.
- Panno, S.V., Kelly, W.R., Martinsek, A.T., Hackley, K.C. (2006), Estimating Background and Threshold Nitrate Concentrations Using Probability Graphs. Ground Water 44, no. 5: 697-709. Print ISSN: 0017-467X Online ISSN: 1745-6584 DOI: 10.1111/j.1745-6584.2006.00240.x
- Peh, Z., Miko, S., Hasan, O. (2010), Geochemical background in soils: a linear process domain? An example from Istria (Croatia). Environ Earth Sci (2010) 59: 1367-1383. ISSN 1866-6280 (Print) 1866-6299 (Online) DOI 10.1007/s12665-009-0125-2
- Reimann, C., Filzmoser, P. (2000), Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environmental Geology no. 39: 1001-1014. ISSN 0943-0105 (Print) 1432-0495 (Online) DOI 10.1007/s002549900081
- Reimann, C., Filzmoser, P., Garrett, R.G. (2005), Background and threshold: critical comparison of methods of determination. Science of the Total Environment no. 346: 1-16. ISSN 0048-9697 DOI:10.1016/j.scitotenv.2004.11.023
- Reimann, C., Garrett, R.G. (2005), Geochemical background - concept and reality. Science of the Total Environment no. 350: 12-27. ISSN 0048-9697 DOI:10.1016/j.scitotenv.2005.01.047
- Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R. (2009), Statistical Data Analysis Explained. West Sussex, England: John Wiley & Sons Inc. ISBN 978-0-470-98581-6
- Rentier, C., Delloye, F., Brouyre, S., Dassargues, A. (2006), A framework for an optimised groundwater monitoring network and aggregated indicators. Environ. Geol. ISSN (printed): 0943-0105. ISSN (electronic): 1432-0495 DOI 10.1007/s00254-006-0200-x.
- Shand, P., Edmunds, W.M. (2008), The Baseline Inorganic Chemistry of European Groundwaters. In: Natural Groundwater Quality (eds. Edmunds, W.M. and Shand, P), Blackwell Publishing Ltd. ISBN: 978-14051-5675-2
- Sinclair, A.J. (1991), A fundamental approach to threshold estimation in exploration geochemistry: probability plot revisited. J. Geochem. Explor. no. 41: 1-22. ISSN 0375-6742, DOI: 10.1016/0375-6742(91)90071-2
- Stanley, C.R., Sinclair, A.J. (1989), Comparison of probability plots and the gap statistics in the selection of thresholds for exploration geochemistry data. J. Geochem. Explor. Vol 32, Issues 1-3: 355-357. ISSN 0375-6742 DOI: 10.1016/0375-6742(89)90076-9
- Velić, J., Saftić, B., Malvić, T. (1999), Lithologic composition and stratigraphy of quaternary sediments in the area of the “Jakuševac” waste depository (Zagreb, Northern Croatia). Geol Croat 52(2):119–130. ISSN: 1330-030X
- Vlahović, T., Bačani, A., Posavec, K. (2008) Hydrogeochemical stratification of the unconfined Samobor aquifer (Zagreb, Croatia), Environ. Geol. Vol. 57, No. 8: 1707-1722. ISSN (printed): 0943-0105. ISSN (electronic): 1432-0495 DOI: 10.1007/s00254-008-1452-4