COMMON IMPLEMENTATION STRATEGY FOR THE WATER FRAMEWORK DIRECTIVE AND THE FLOODS DIRECTIVE

TECHNICAL REPORT ON GROUNDWATER QUALITY TREND AND TREND REVERSAL ASSESSMENT

Procedures applied by Member States for the first RBMP cycle

- June 2019 -

Disclaimer

This technical document has been developed through a collaborative framework (the Common Implementation Strategy) involving the Member States, EFTA countries, and other stakeholders including the European Commission. The document reflects the informal consensus position on best practice agreed by all partners. However, the document does not necessarily represent the position of any of the partners.

To the extent that the European Commission's services provided input to this technical document, such input does not necessarily reflect the views of the European Commission.

The technical document is intended to facilitate the implementation of Directive 2000/60/EC and is not legally binding. Any authoritative reading of the law should only be derived from Directive 2000/60/EC itself and other applicable legal texts or principles. Only the Court of Justice of the European Union is competent to authoritatively interpret Union legislation.

CIS WG GW – Voluntary Group “Trend in Groundwater”

Group leaders

- Laurence Gourcy (BRGM - French Geological Survey)
- Benjamin Lopez (BRGM - French Geological Survey)

Group members

- Ruxandra Balaet (MEWF)
- Serge Brouyere (ULG)
- Mária Bubeníková (SK Water Research Institute)
- Robert Chriaštel (SK Hydrometeorological Institute)
- Jacqueline Claessens (NL Institute for Public Health and the Environment – RIVM)
- Marina Gacin (Slovenian Environment Agency)
- Réka Gaul (BM)
- Falk Hilliges (UBA)
- Anna Kostka (PGI)
- Anna Kuczyńska (Polish Geological Institute-National Research Centre)
- Alexandra Lequien (Ministère de l'écologie, du développement durable et de l'énergie - DGALN/DEB/GR1)
- Dirk-Ingmar Müller-Wohlfeil (Ministry of Environment and Food of Denmark)
- Rudolf Philippitsch (Federal Ministry of Agriculture, Forestry, Environment and Water Management)
- Katarína Pilátová (SK Water Research Institute)
- Hana Prchalová (CZ Water Research Institute)
- Elisabetta Preziosi (Water Research Institute (CNR-IRSA))
- Inga Retike (LVGMC)
- Oliver Ries (Ministry of Environment and Food of Denmark)
- Susie Roy (Amec fw)
- Andreas Scheidleder (Umweltbundesamt)
- Jonathan Smith / Matthijs Bonte (Shell representing CONCAWE)
- LærkeThorling Sørensen (Geological Survey of Denmark and Greenland - GEUS)
- Robert Ward (British Geological Survey - BGS)

Contents

List of illustrations

List of appendices

1. Preamble

1.1. INTRODUCTION – CONTEXT OF THE STUDY

The identification, monitoring, assessment and need to implement any necessary measures to reverse significant and sustained upward trends in groundwater pollutant concentrations in bodies or groups of bodies of groundwater are integral requirements of the Water Framework Directive (WFD) (2000/60/EC) and the Groundwater Directive (GWD) (2006/118/EC).

In 2010, the Commission Staff Working Document - European Overview Accompanying the document "Report from the Commission to the European Parliament and the Council on the Implementation of the Water Framework Directive (2000/60/EC) River Basin Management Plans)" indicated that, in 68 River Basin Management Plans (RBMPs), 17 Member States (MS) had performed quality trend assessments for groundwater. The working document revealed that a majority of MS has chosen to apply statistical methods but the methods used vary considerably between RBMPs dependent on the length of the time series considered or trend reversal methodology applied.

To complement this report, in October 2014, the CIS Groundwater Working Group (WGGW) launched a voluntary process to exchange information on MS trend assessment methodologies. The aim was to share experiences, best practices on the statistical tests used, the assessment tools that have been developed, and solutions found to tackle main challenges. The process was based on the voluntary action of MS who were asked to respond to a questionnaire. EU Water Directors took note and acknowledged this report in June 2019 in Constanta, Romania.

This report is solely focusing on the assessment of trends and trend reversal in pollutant concentrations.

1.2. LEGISLATIVE BACKGROUND

The WFD and GWD require that trends in pollutant concentrations in groundwater are identified and are assessed to determine whether they are environmentally significant. Such trends are referenced several times in the WFD:

- Recital 26: "For groundwater, in addition to the requirements of good status, any significant and sustained upward trend in the concentration of any pollutant should be identified and reversed."
- Recital 28: The time lag for improvement should be taken into account when establishing measures to reverse trends.
- Article 4.1.b (iii) – (Environmental objectives) : "Member States shall implement the measures necessary to reverse any significant and sustained upward trend in the concentration of any pollutant resulting from the impact of human activity in order progressively to reduce pollution of groundwater. "
- Articles 17.2 (b) and 17.5 – criteria to be developed for the identification of significant and sustained upward trends and for the definition of starting points for trend reversals. These requirements are further specified by the GWD.

- Annex V: 2.4.1 (groundwater monitoring network) 2.4.4 (surveillance monitoring) 2.4.3 (operational monitoring), 2.4.4 (identification of trends in pollutants), 2.4.5 (interpretation and presentation of groundwater chemical status) and 2.5 (presentation of groundwater status).

In the GWD trends are cited in:

- Recitals 6 and 11,
- Article 1(b) (Purpose) – to fulfill the requirements noted under WFD Art 17 above;
- Article 2.3 - Definition of significant and sustained upward trend;
- Article 5 - Identification of such trends and the definition of starting points for trend reversals;
- Article 6.1.b - Measures to limit inputs of non-hazardous pollutants into groundwater to avoid such trends;
- Annex IV - Identification of such trends and the starting point for trend reversals. Reversal of significant and sustained upward trends.

In addition to the above, in the context of protecting drinking water sources, WFD Article 7.3 requires that MS ensure the necessary protection of water bodies with the aim of avoiding deterioration in quality. The GWD makes this a condition of good groundwater chemical status. By implication this requires the identification and assessment of deteriorating trends in groundwater quality.

Where significant upward trends exist they must be reversed through the application of programmes of measures to ensure that there will not be future failures of environmental objectives. The GWD starting point for trend reversal must be defined as a proportion of the threshold value (TV) or groundwater quality standard (GWQS) for the substance in question (75% by default).

The following CIS documents provide further explanation and guidance on meeting the requirements of the WFD and GWD:

- Technical Report No. 1 (Grath and al., 2001): Statistical aspects of the identification of groundwater pollution trends and aggregation of monitoring results – WG 2.8 Statistics (2001)

- Guidance Document No. 18 : Guidance on groundwater status and trend assessment, in which several relevant topics are detailed:
  - Conceptual model and trend assessment;
  - Concentrations below the limit of quantification and trend assessment;
  - Reporting and trend assessment; and
  - Trend and trend reversal assessment.

1.3. DEFINITION

Trend: In the mathematical sense, trend corresponds to a certain orientation assumed by the values of a data series over time, that is, a variation in process observed over time.

The trend appears when the data move away from stationarity. If the structure of the chronological data series remains the same, the process is stationary. In other words, a stationary record will always have the same pattern, whether on date t or on date t+k.

For example, a record whose mean, variance and autocorrelation does not change with time is stationary. The causes of non-stationarity can be:

⇨ either a gradual modification of the series over time, which shows up as a decreasing trend (or an increasing trend) of the values in the series,

⇨ or one (or several) break(s) occurring on a certain date, and the characteristics of the series being no longer the same after the date of the break,

⇨ or a change in the probability distribution of a characteristic variable beginning on a given date.

Trends can take several different forms. In the regulatory context in which this guide is written, only monotonic trends in pieces are considered, that is to say time series that remain always increasing or decreasing, the direction of variation remaining constant over time.

1.4. GENERAL PRINCIPLES

Temporal variability in groundwater quality can be attributed to numerous factors: variations in climatic and hydrogeochemical conditions, (variations in) groundwater extractions, variations in environmental emissions of substances to soil and groundwater, intrinsic characteristics of each element or even changes/improvements in chemical analysis, etc.

It is possible to distinguish several components in a concentration time series that are created as a result of various factors:

- A seasonal component shown as intra-annual fluctuations that repeats from year to year. This component can be generated by the seasonality of the climate or even by practices that are repeated regularly each year (withdrawals for agriculture, pesticide treatments, etc.)
- A pluri-annual cyclical component shown by a succession of increases or decreases over a period of time greater than a year, and often not perfectly regular. This component can be related initially to large-period climatic oscillations such as the North-Atlantic Oscillation (Lopez et al., 2015).
- A random component representing unpredictable fluctuations (or noise) that do not depend on time. This component is linked particularly to sample bias and measurement and analytical error, or to other unknown explanatory variables.
- A trend component shown by a gradual change that is observable over a long time period. This component can be the result of human activities (withdrawals, releases of chemical elements) or long term natural changes.

The analysis of records of groundwater variations conducted within the framework of the WFD is primarily oriented to the detection of the trend component and notably to variations explained by anthropogenic activities, because this is the component on which corrective measures can have an effect. The trends of interest are significant and sustaining increases in pollutant concentrations linked to releases from human activities. The implementation of this work is exposed to potential pitfalls that must be identified so that they can be avoided.

- **How to tell the difference between a pluri-annual cyclic variation and a long-term monotonic trend**

A pluri-annual cyclic variation can be defined as an alternation of rising and falling trends; it is a mistake to look for monotonic trends on such a signal. This is particularly true when the monitoring history is less than 10 years, a duration that corresponds to a half-period of the longest North-Atlantic oscillation. Thus, to be able to tell the difference between pluri-annual cycles and long-term monotonic trends, it is necessary to use very long records, or to evaluate the risk that the record analyzed is subject to pluri-annual cyclic variations (analysis of the context, analogy with nearby records that are longer, etc.).

- **How to distinguish trends linked to anthropic activity from natural trends**

Within the framework of the implementation of WFD, only trends linked to human activities must be identified. Thus, it is sufficient to be sure that an observed significant trend is not linked to a change in environmental variables (change in effective rain, particularly). Methods for studying the correlation between two series are recommended for distinguishing the types of variations. They are presented in the guidebook along with statistical methods for identifying the trends.

The WFD monitoring (surveillance and operational) allows for identifying the above-mentioned patterns. Further details are already outlined in the CIS Guidance No 15: Guidance on Groundwater monitoring

- **A surveillance monitoring network** to: (a) supplement and validate the Article 5 characterisation and risk assessment procedure with respect to the risks of failing to achieve good groundwater chemical status; (b) provide information for use in the assessment of long term trends in natural conditions and in pollutant concentrations resulting for human activity and; (c) to establish, in conjunction with the risk assessment the need for operational monitoring.

- **An operational monitoring network** to: (a) establish the status of all groundwater bodies, or groups of bodies, determined as being ‘at risk’, and (b) establish the presence of significant and sustained upward trends in the concentration of pollutants.

- **Applying statistical methods appropriate for the characteristics of raw environmental data**

The choice of statistical methods vary, depending on the objective of assessment, on the characteristics of data, on the pretreatment considered and on the level of measurements involved in the data (categorical, ordinal, interval and ratio).

Statistical tests for trend assessment are classified in two main groups: parametric tests and non-parametric tests. Parametric methods assume that the data are derived from a population with a normal (or another specific) distribution. Due to the characteristics of groundwater quality time series (such as abnormal data, autocorrelation, often not distributed normally, etc.) the use of parametric tests require pretreatment of raw data, such as the suppression and/or correction of anomalous data, normalization of data and residues, or decomposition of autocorrelated records.

Contrary to parametric tests, non-parametric tests do not require a hypothesis on the law of distribution of the variable under study. In particular, they can be applied to small samples where the condition of normality is difficult to verify. They also depend for the most part on the simple principle using classification by observation and are thus not very sensitive to extreme values

- **Describing the past rather than predicting the future**

It can be tempting to extend a trend in time after having identified it, particularly to estimate future changes. But trend identification by application of statistical methods as advocated by the regulations consists of the analysis of historical raw data. No other data from context or explanatory data is taken into account in trend calculation, unlike the creation of a conceptual or mathematical model. The goal is in fact not to predict but to characterize a variation. A trend detected as significant will therefore be representative only of the period for which the data are available. To extend a trend is to make the hypothesis that the variations in the set of processes that generated it and sustained it during the period of analysis will continue in an identical way, which is generally not the case. Models are developed with the goal of predicting the future, under different scenarios of variation of explanatory parameters.

2. Overview of procedures applied by Member States for the first RBMP cycle

2.1. DATA GATHERING

To gather information on the practice of assessing trends in pollutant concentrations for the first RBMP cycle in the European Member States, a questionnaire was distributed in 2014-2015 within WG GW.

The twelve questions included in the questionnaire (listed below) asked MS to provide general information on different elements of trend and trend reversal assessment with the aim of making it possible to compare the different approaches. Ten MS replied to the questionnaire (Appendix 2): Austria, Czech Republic, Denmark, France, Netherlands, Poland, Romania, Slovak Republic, United Kingdom, and Hungary.

List of Questions:

I. Which GWB and which substances?

II. Which data?

III. First determine trend per monitoring point, then per groundwater body? Or lump all data, then determine trend? From monitoring point to GWB level.

IV. Which statistical method used for trend determination? And its confidence

V. How environmental significance was defined for trend?

VI. Number of years of monitoring results required to determine trends

VII. Methods used in case of missing data or too few data

VIII. Monitoring density (spatial) for trends at GWB level?

IX. Results below Limit of Quantification

X. Can you use expert judgement for trend determination?

XI. Consideration of baseline level of concentration?

XII. Methodology for trend reversal assessment?

2.2. SYNTHESIS OF MS TREND PROCEDURES

The compilation of the 10 Member States responses to the questionnaire gives an overview of how trends in groundwater quality have been assessed during the first cycle of River Basin Management (RBMP) preparation. Some items were treated in the same way by each MS while some issues reveal differences in the approach.

Trends assessment was carried out by the 10 volunteer MS for all groundwater bodies (GWBs), and for at least the groundwater quality standard and pollutants listed in Annex I and II respectively of the GWD. The majority of the 10 MS, covered all relevant pollutants responsible for downgrading GWBs "at risk". Some MS performed trend assessment for all analysed substances with long enough time series.

All 10 MS used data from national monitoring networks (surveillance and operational monitoring), sometimes complemented with additional data from other networks (raw groundwater used for drinking water in particular). Several MS applied a filter to exclude data from trend assessment due to, for example, insufficient accuracy in the limit of quantification (LOQ) or too many “less than LOQ” values in the time series. The length of the considered time series varies considerably, starting mainly from 1995 up to 2008.

In seven of ten MS, statistical methods were applied to identify trends. Methods used could be divided in 2 groups:

- Parametric test ANOVA, in some cases combined with the LOESS smoother, used by 3 MS;
- Non-parametric Mann-Kendall test (and derivate Seasonal Kendall and Regional Kendall tests) used by 4 MS.

One MS applied the parametric linear regression test while 2 MS did not perform regression.

The statistical significance of trend assessments is directly provided by the p-value of the applied statistical tests. Environmental significance is mainly estimated based on exceedance of a TV or a percentage of a TV for a predicted concentration. The date for the prediction is often defined as the date when the trend line exceeds the starting point for trend reversal. These dates vary between MS from 2017 to 2021 or 2027.

All MS except 1 defined a minimum length and/or volume of data to determine trends. 3 MS applied recommendations stated in the WFD CIS Technical Report No. 1 (8 years minimum for annual monitoring and 6 years minimum for half-annual or quarterly frequencies). Other MS fixed criteria regarding the optimal conditions (e.g. number of data, statistical distribution of data) for the application of the statistical method chosen. Procedures applied for missing data vary a lot, from the exclusion of all monitoring points for which at least 1 value is missing, to the prediction of missing value applying a Likelihood method.

One half of the MS fixed a minimum number of sites per GWBs for trend assessment at the GWB level (at least 3 sites for 5 MS and at least 5 sites for 1 MS). MS indicate that the number of monitoring points per GWB strongly depends on the size and the heterogeneity of the GWB. 2 MS used the spatial density of monitoring sites not as criteria for trend assessment but for determining the chemical status reliability.

A majority of MS used the substitution procedure in case of “Less than LOQ” values. Values reported below LOQ are most often replaced by half of LOQ (some MS specified LOQmax). 2 MS proposed another value function of the parameter considered (organics, pesticides, metals, etc.). Only one MS considered the concentration of the Natural Baseline Level (NBL) for the TV determination process. Other MS did not specifically take into account the NBL in trend assessment as they carried out trends for parameters of possible natural occurrence in the same manner as for the other substances.

Trend reversal assessment firstly needs the assessment of a trend. Therefore sufficiently long time series will be necessary. This explains why only 2 MS reported a trend assessment reversal. 6 others MS specified that even if trend reversal was not assessed for the 1st RBMP cycle (upward trends to be determined first) the methods are available and ready to be applied. A two-section regression model is often cited as the procedure that would be applied.

3. Procedures for assessing trends

3.1. GROUNDWATER – ASSESSMENT OF SIGNIFICANT AND SUSTAINED UPWARD TREND IN POLLUTANT CONCENTRATION

3.1.1. Objective

The Article 5.1 of the GWD requires that "Member States shall identify any significant and sustained trend in concentrations of pollutants, groups of pollutants or indicators found in bodies or groups of bodies of groundwater identified as being at risk and define the starting point for reversing that trend, in accordance with Annex IV".

The assessment of a trend in groundwater quality variation is necessary for the evaluation of the chemical condition of GW bodies, for evaluation of significant and long-term trends, for evaluation of the risk of non-achievement of good status and for evaluation of variations in pollutant plumes originating from point sources.

The trend assessment can also be useful for characterizing the variation of chemical concentrations throughout the GW-bodies and Group of GW-bodies. The application perimeter of the method is limited, however, to the study of trends related to anthropic activity.

The following chapters summarized the approaches taken by each individual contributing country. Each chapter is introduced by a brief summary of the requirements of the WFD and GWD and the recommendations given in the WFD CIS Technical Report No. 1 and the WFD CIS Technical Report No. 18.

3.1.2. Data used

Methodology for trends and trend reversal assessment used was based on the WFD CIS Technical Report No. 1 and the WFD CIS Technical Report No. 18. According to WFS CIS No. 1 and CIS No. 18, the length of time series required for the assessment depends on data availability, specifically the frequency of monitoring. If data are aggregated as annual means, the minimum required period of observations is 8 years. If data are collected with higher frequencies (half-annual or quarterly), a minimum of 6 years of data are required.

Within trend assessment individual pollutant concentrations (or values) below LOQ should be replaced by half of the value of the highest LOQ occurring in the time series being analysed. (CIS No. 18, chapter 2.4.2.)

The WFD and GWD state that significant and sustained upward trends shall be identified in concentrations of pollutants, groups of pollutants or indicators of pollution found in groundwater bodies or groups of groundwater bodies at risk. Unlike for chemical status assessment, neither directive explicitly states which parameters shall be subject to this assessment.[…]. It is therefore considered that trend assessment […] should be undertaken for parameters that are posing a risk to the groundwater body. Trend assessment may also be undertaken for any other parameters (natural) which may occur as a result of human activity across the groundwater body. (CIS No.18, chapter 6.2.1).

Trend […] assessment should be based on surveillance and operational monitoring data from individual monitoring sites. The first identification of trends should be carried out by 2009 where possible, taking into account any data that were collected prior to the current river basin planning cycle in order to enable reliable trend assessment and reporting on trends within the first RBMP […]. Where available, Member States are free to include additional representative monitoring data in the assessment where they can contribute to an improved confidence in the assessment. (CIS No.18, chapter 6.2.3).

The length of time series that should be considered in trend assessment depends on how the groundwater body reacts to changes in practices at the land surface […], on the power of the trend test method in detecting trends and on the quality of the data […].The minimum length of time series to be used, in terms of the number of regularised measurements and the minimum number of considered years, depends on the monitoring frequency, the statistical method, […] and on the power of the method. (CIS No.18, chapter 6.2.5).

Case study: Austria

Trend assessment is performed for all GWBs and substances where the starting point for trend reversal is exceeded by the 3-year-average (which is the appraisal period) of annual mean values in at least 30% of all monitoring points within a GWB.All data from the national water monitoring network (WFD operational and surveillance monitoring) are considered in trend. Monitoring started in 1991. Length of time series considered depends on monitoring frequency.

The length of time series depends on the aggregation period and for annual aggregations it is 8 years and for half-annual and quarterly it is 6 years.

If more than 1 value is missing in the time series of a monitoring point (but not at the beginning and not at the end of the time series), the monitoring point is excluded from further analysis.

All values <LOQ are replaced by half of LOQmax. LOQmax is identified over the length of the analyzed time series (6 or 8 years for trends and 10 or 14 years for trend reversal – depending on annual, half-yearly or quarterly aggregation).

Case study: Czech Republic

All GWBs were considered for trend assessment. All assessed substances (except pesticides and their metabolites – too short time series, most of results below limit of quantification), but significant upward trend was found for 16 substances only – metals, selected PAHs, nitrates, nitrites, orthophosphates, chlorides and sulphates.

All data from national monitoring network (boreholes and springs) were used for trend assessment. Although the monitored period is longer for part of the stations, data from 2000–2012 were used for most of the stations due to different quality of analyses or too high limits of quantification.

Monitoring points with more than 9 results and at least 6 years could have significant trends; monitoring points with at least 4 years data were also assessed, but the results are not yet significant.

All values <LOQ in specific monitoring point were replaced by half of LOQmin.

Case study: Denmark

All GWBs for which chemical analyses have been available for at least 5 monitoring points for each of the two periods 2000-2006 and 2007-2013. Nitrate, Cadmium, Lead, Mercury, Sulphate, Pesticides, Chlorinated solvents, BTEX were considered.

Based on the same data that have been used for status assessment and that are available from the National borehole database (JUPITER), operated by the National Geological Survey (GEUS). Data from the period 2000-2013 have been used. Data have been subdivided into 2 periods (2000-2006 and 2007-2013). The assessment was based on a comparison of the chemical status between these 2 periods.

The minimum criterion has not been related to the number of years. Rather, it is required, that chemical analyses have been available for each of the two periods 2000-2006 and 2007-2013.

Trend calculations have been performed only for GWBs, for which at least 5 monitoring points have been available for each of the two periods 2000-2006 and 2007-2013.

The level (limit) of detection (LOD) was used. As no regression analysis was performed, no risk of false trends from low values was expected, and all LOD << TVs.

Case study: France

Nitrate on all French GWBs. All substances on all GWBs from Guadeloupe RBD. Sulfate, conductivity, chloride, ammonium and arsenic on all GWBs from La Réunion RBD. Locally, parameters placing GWBs at risk like chloride (though not always possible/easy for all parameters causing risk: e.g. pesticides).

No trend assessment performed for the chemical substances that have been determined in previous studies as being of potentially high concentration due to the natural geochemical background.

All data available (WFD operational and surveillance monitoring, Drinking Water monitoring) were used for the period 1996-2013 (1996 = first RBMPs, laboratories started to get better, enough data to assess trends). At least 10 data points are required to run the Mann-Kendall statistical trend test as it is implemented in the dedicated tool. So there is not a minimum number of years of monitoring results required but a minimum number of monitoring data available after the year 1996.

Trend is not assessed at monitoring point level in case of not enough available data but data at these monitoring points are used to estimate trends at the GWB level via the Kendall-regional trend test.

Regional trends at GWB level have been calculated when at least 3 monitoring points with data after 1996 are available.

As the Kendall Regional trend test is a non-parametric statistical test, results reported below LOQ are taken into account for estimating trend. For calculating slope, the Sen

Slope test has been used after substituting results reported below LOQ by the LOQ/2 value.

Given the wide variability of time steps and concentrations in the records of groundwater quality variation, the trend calculation for chemical variation can be done on raw data.

The trends will be sought on the period of calculation - determined by regulation and indicated in the guide for the evaluation of quality conditions if need be or according to the goal of the study - except if the hydrogeological functioning at the level of the sample point shows pluri-annual cyclic variation: to be sure that an observed increasing concentration trend is not a trend linked to the observation of a part of a pluri-annual cycle, the period over which the trend will be sought will be extended to at least 3 cycles (the method for determining cycles is set forth in the following paragraphs).

The presence of some gaps is tolerable if they are short. Thus, it is considered that the search for a trend can be conducted when data are available for at least 80% of the years.

A trend is calculated only on records for which at least 10 data are available (Lopez et al., 2011).

In the event that a record contains several quantification limits, all values (quantified or not) below the highest quantification limit will be considered as equal to the maximum quantification limit (cf. Figure 1 B).

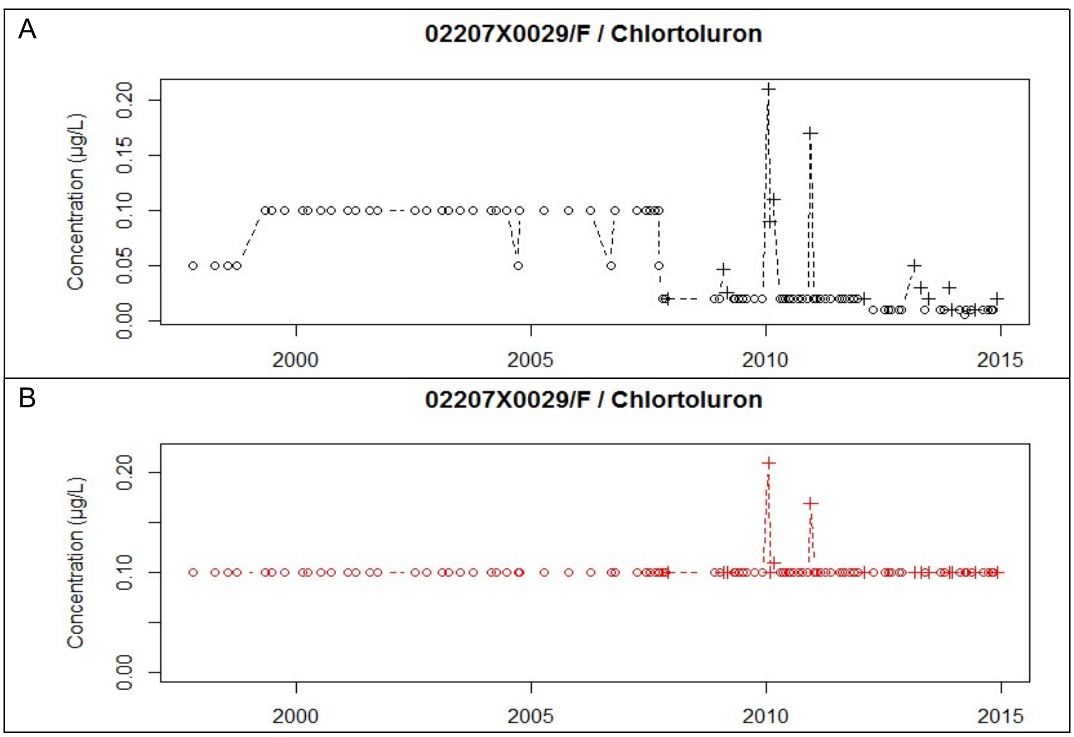


Figure 1: Chlortoluron concentrations at well 02207X0029/F from 1997 to 2015. Circle: quantification limits. Cross: measured concentrations. A - Raw results. B - All results below the maximum quantification limit in the record (0.1µg/L) are replaced by this value.

Case study: Hungary

All GWBs were considered using NH_4, NO_3, Cl, SO_4, EC, Hg, Cd, Pb, DO, pH, atrazine, AOX, pesticides total, simazine, terbuthylazine, terbutrine, trichloro- and tetrachloroethylene.

Data of the WFD surveillance and operative monitoring network for the time period 2000-2012 were used. In addition, the time period 2006-2012 was examined separately. Monitoring data is aggregated to yearly averages at each monitoring point. Monitoring points in the vicinity of polluted sites were not taken into consideration for GWB trend assessment.

At least 5 annual data per monitoring point is required. The applied non-parametric statistical methods allow for missing data. Time series with more than 80% of values below the LOQ are excluded from the trend assessment.

A certain amount of missing data is allowed by the non-parametric statistical trend test. However, when less than 5 annual data are available at a monitoring point the trend status assessment is considered ‘not applicable’. Trend assessment is not carried out on datasets with more than 80% of values below the LOQ.

Requirements for trend assessment are at least 3 sites per GWB. In cases, where an individual time series is considered representative of the GWB, less than 3 sites are also accepted.

Results below LOQ are replaced by half of LOQ except for:

- Pb, Cd and Hg where these data are replaced by 0.005 µg/l;
- organic pollutants and pesticides data by 0.

Case study: Italy:

A national guideline for trend and trend reversal assessment has been published on July 2017 (ISPRA - IRSA.CNR 2017) also on the basis of the indications provided by the activity of WG Groundwater ‘Trend assessment’. According to national regulations, River basin district authorities, Regions and Autonomous Provinces (Trento and Bolzano) are in charge of the assessment of sustained and significant upward trends in groundwater bodies classified at risk and can apply the procedure described in the guideline. The trend and trend reversal assessment is currently ongoing.

Trend assessment is performed for GWBs defined “at risk”, for the substances that pose the risk and for other substances that are critical for the maintenance/reaching of good groundwater chemical status. All data from the operational and surveillance monitoring are considered.

A minimum of 8 (annual) values is required for trend assessment, for a range of 8 to 15 years. For trend reversal calculation at least 14 annual data are required for a range of 14 to 30 years of monitoring. Years with no data are acceptable within the time series, provided that there are at least 8 years with data for trend calculation (14 for trend reversal). Datasets with missing data for more than one consecutive year should be discarded.

All values <LOQ are replaced by half of LOQ. When different LOQ are present in the dataset, the half of LOQmax is considered.

As for groundwater bodies not at risk, a simplified procedure is suggested by the national guideline due to the lower frequency of the surveillance monitoring. In such cases, only the compounds that have shown a value exceeding the 75% of its Environmental Quality Standard/Threshold Value in the preceding 6 years planning period, are considered for trend assessment. The average values of the next 6–years planning period are compared to the average values of the previous one.

Case study: The Netherlands

All GWBs were considered for substances mentioned in GWD Annex I (Nitrate and pesticides) plus substances with TVs (chloride, nickel, arsenic, cadmium, lead and phosphate).

All data that are used for status assessment are also used for trend assessment.

Monitoring for status assessment occurs at depths of 10 and 25 m. Data for each depth are assessed separately. Focus is on the 10 m data because trends are expected there earlier.

There is no minimum defined. Instead, the number of years for which data are available determines the relevance that should be ascribed to a trend. Example: if data from only 2 years are available, an upward trend should trigger continuation of the monitoring of that particular substance. A data series containing 8 to 15 years of data usually allows statements with sufficient statistical power about the presence or absence of upward trends.

In case of missing data, The REML-method is used (Residual Maximum Likelihood; Welham et al., 2004).

Theoretically, three aspects are distinguished for monitoring density:

1) Frequency: depends on groundwater flow rate. Interval between measurements should be longer than replacement rate (otherwise you measure the ‘same’ water).

2) Depth: should depend on local conditions and should be derived from the conceptual model.

3) Number of monitoring points per GWB: should depend on the heterogeneity of the GWB, e.g. in terms of soil-land use combination. Weighting can be applied to correct for non-proportionalities.

Ideally, raw data should be used, including values below LOQ (even negative values). These are often not reported by the analysing labs. In that case, the best is to: 1) replace by LOQ; 2) replace by zero; and 3) to compare 1 and 2 to investigate the sensitivity of the replacement method chosen.

Case study: Poland

All GWBs for which water quality data are available. All analysed substances (pollutants and natural parameters, except organic substances – too short time series, most of results below limit of quantification - LOQ).

All data available from 2005–2012 period collected within the WFD Operational and Surveillance Monitoring Programmes and supplemented by data from other monitoring programmes and borehole testing carried on by the Polish Hydrogeological Survey.

Data are aggregated to annual means, therefore the trend analysis was carried out for 8-year long time series.

Two missing values were allowed within time series. Where there was not enough available data, the trend was not estimated.

Results below LOQ were replaced by half of LOQ.

Case study: Romania

All GWBs for which chemical analyses have been available for a time series of at least 8 years (64 GWBs from the total of 143) were considered. Trend assessment was carried out for all pollutants/indicators that are causing risks for GWBs, identified through the risk assessment process, also for the GWBs which are not found at risk. These substances are the nitrogen compounds: Nitrates, Nitrites and Ammonia (NO_3, NO_2 and NH_4).

All monitoring data available from 2000–2013 period, provided by WFD Operational and Surveillance Monitoring Programmes carried on by the National Administration "Romanian Waters".

Trend analysis was carried out for more than 8 years, in the case of annual aggregated data. Trend is not assessed where there is not enough data available.

Trends at GWB level have been calculated where data series for a minimum of 3 monitoring points are available. The number of monitoring points per GWB depends on heterogeneity and areal size of the GWB.

Results below LOQ were replaced by half of LOQmax. LOQmax is defined as the largest LOQ which does not exceed twice the median of all LOQs.

Case study: Slovak Republic

All GWBs, for which chemical analysis is available for at least 6 years were taken into account for all substances stated in GWD Annex I and II plus substances with TVs relevant for GWB.

All data obtained from groundwater surveillance and operational monitoring network fulfilling following criteria for each time series:

- Minimum length = 6 years;
- Maximum length = 10 years;
- Start of monitoring in 2004 or later;

- End of time series in 2012 or 2013;
- Maximum gap in time series - 1 year;
- Percentage of non-detects < 50%;
- Percentage of unique aggregated values ≥ 50% (related to the count of all aggregated values evaluated in each time series).

Trend analyses are primarily based on monitoring results from the period 2004 - 2013. If data are not available for the whole period, a shorter period can be accepted if its minimum length is at least 6 years and the last measurement was carried out in 2012 or 2013.

If a gap in the monitoring data for the assessed parameter is more than 1 year, the monitoring point is excluded from the trend assessment.

Trend analyses at GWB level are performed, if time series with the same length are available for at least 3 monitoring points.

LOQmax is identified individually for each time series. Measurements below LOQ and measurements below LOQmax are replaced by LOQmax.

Case study: Slovenia

In accordance with national regulations, it is necessary to assess trends of pollutants in groundwater. In Slovenia, trends are identified at individual monitoring points for parameters with a sufficiently long datasets (at least 6 years).

Cased study: United Kingdom

Trend assessment was carried out on all GWBs at risk of failing to achieve the good chemical status objective or which were at poor status, for the purposes of meeting the trend assessment objective.

Additional trend assessment was applied as part of groundwater chemical status classification for two tests - the Saline Intrusion Test and the Drinking Water Protected Area Test (DWPA) Test.

Trend assessment was carried out for all pollutants/indicators that were putting the GWB at risk, identified through the risk assessment/characterisation process.

All monitoring data used for status assessment were considered for trend assessment where relevant. This included data from both Surveillance Monitoring (SM) and Operational Monitoring (OM). The data from SM were also be used to assess natural trends in groundwater.

The SM and OM networks form the national strategic groundwater quality monitoring network. Data for the period 1 January 1997 to 31 December 2006 (10 Years) were used.

Additional data from water companies was also used as supporting evidence. This information (Water Company WFD Article 7 data) recorded qualitative evidence of upward trends at drinking water abstractions.

For the assessment of anthropogenically induced upward trends in pollutant concentrations, monitoring data for a period of between 6 and 10 years leading up to the date at which assessment is being carried out, were used for identifying trends.

When considering natural trends, data were considered for as long a time period as possible. As a minimum this was at least 6 years with at least one measurement per year. For England and Wales, trend analysis was only carried out for monitoring points where there were a minimum of 10 analyses (results) over a period of at least four years.

At least one measurement per year was required for trend assessment to be carried out. Where data were inadequate, then no trend assessment was carried out and an explanation was recorded.

Trends at the GWB level have been assessed using all data that are part of the Operational and Surveillance monitoring networks. It is considered that the number of sites and their distribution will enable a representative assessment of the status of groundwater and hence trends. Therefore for small GWBs this may be a single monitoring point whereas in larger GWB a larger number of points would be used.

Trend assessment was carried out on data from individual monitoring sites for individual parameters and for total pesticides. Where individual parameter concentrations (or values) were reported as below the LOQ, values were replaced by values with half of the reported LOQ, except for total pesticides. For example, a value reported as <0.1 µg/l would become 0.05 µg/l. Trend assessment was not carried out on datasets that comprised more than 80% of values below the LOQ.

Case study: Wallonia (Belgium)

Two specific trend assessment studies were undertaken at the scale of the Geer basin chalk aquifer for respectively nitrate (Batlle-Aguilar, J. et al 2007) and pesticides (Hakoun et al, 2017). For nitrate, long time series (more than ten years) were used to cover natural multi-annual fluctuations of nitrate concentrations due to groundwater level fluctuations. For pesticides, trends were assessed for 4 pesticides and 2 metabolites detected in groundwater based on time series covering almost 20 years.

Trend assessment and reversal was also carried out on all GWB of the Walloon Region for nitrate time series (D'Or D. and Allard. D., 2014) combining concentration data and soil crop modelling results:

- Temporally isolated data were discarded
- The maximum period considered in the analysis is the period 1971-2013. However, trend assessment was performed only on datasets containing at least than 10 data on nitrate concentrations.
- Outliers were detected using LOESS (Grath et al., 2001)

3.1.3. Trend calculation

The assessment must be based on a robust statistical method such as regression analysis (GWD, Annex IV A(2)(c)). Since “significant” relates to statistical significance (as well as environmental), the method chosen should also be able to test the statistical significance of a measured trend. […]. In order to distinguish between natural variation and trends with an adequate level of confidence and precision the trend test methodology should also be able to perform a test on seasonality where appropriate. (CIS No.18, chapter 6.2.6).

As the trend assessment is based on data from individual surveillance or operational monitoring points, a procedure (criteria) is needed to combine the results of the individual trend […] assessments at sampling points in order to assess the trend at the groundwater body level. To determine whether a trend is environmentally significant the same principles used when assessing chemical status can be applied. This means that the trend assessment should be applied at whatever scale is needed to test for significance i.e. trend assessment may need to be carried out at individual monitoring points, groups of monitoring points or by aggregating results across the whole body. (CIS No.18, chapter 6.3.1).

Case study: Austria

Trend assessment is performed at GWB level based on the aggregated values of all monitoring sites within a GWB. Following minimum criteria have to be met:

- At least 3 sites per GWB;
- At least 2/3 of all sites per GWB have to have sufficient data;
- Only 1 missing value in the time series of each monitoring is acceptable (but not in the beginning and not at the end of the time series); and
- At least 60% of the values are quantified (above LOQ).

Statistically significant trend of pollutant concentration is determined by the generalised linear regression test (ANOVA) based on the LOESS smoother and 5% significance level.

A trend is environmentally significant when the trend line exceeds the starting point for trend reversal.

Case study: Czech Republic

- Each monitoring point was assessed;
- Monitoring points with significant upward trend (but their mean concentrations are below TVs) selected;
- Aggregation on GWB level – for metals and polyaromatic hydrocarbons (PAHs) one monitoring point exceeded, for nitrates, nitrites, orthophosphates, chlorides and sulphates at least half of monitoring points exceeded.

The aggregation on GWB level is not so important, it is necessary to know specific monitoring station and pollutant.

The statistical methods used depend on the volume of data:

- two-section model was used for pollutants and monitoring points with more than 18 results and at least 10 years;
- linear regression for pollutants and monitoring points with more than 9 results and at least 6 years;
- simplified assessment (comparison of two means) was used for pollutants and monitoring points with at least 6 years – but the results were not identified as significant.

A predicted concentration was calculated as a result – for 2015 year and 2017 year. Monitoring points that exceeded TVs in 2015 were considered as having significant upward trends, monitoring points below TVs in 2015, but exceeded in 2017 were identified as having potentially significant trends.

Case study: Denmark

For each monitoring point and chemical parameter, a mean value was calculated for each of the periods, as an average of the respective annual means.

The state of the 1st and 2nd periods are compared and sorted by strongly increasing (STRI), slightly increasing (SLII), stable (STA), slightly decreasing (SLD) and strongly decreasing (STD), following the guidelines used for the reporting of monitoring results for the Nitrates Directive. All changes are calculated as a % of the relevant threshold value. The trend is assessed for differences of respectively: > 10% (STRI); 10% to 2% (SKLII); 2% to -2% (STA); -2% to -10% (SLD); and <-10% (STD). Assessment results where there is a strongly increasing (STRI) parameter value have been attributed to a "significant upward trend."

For each GWB the percentage of results in each of the 5 above mentioned groups was calculated for each parameter.

The trend for each parameter is designated as "non-adverse" development (1), unknown development (0) and worsened development (-1) according to the following criteria:

- "non-adverse development" (1): The parameter is found in the GWB, but there are no measurements greater than 75% of the TV or the parameter values are larger than 75% of the TV, but less than 20% of the monitoring points indicated a strongly increasing trend.
- Unknown development (0): The parameter was not found in the GWB.
- Undesired Development (-1): The parameter was found in concentrations of more than 75% of the TV and more than 20% of the monitoring points at the same time indicate a strong upward trend.

Based on the identification of individual parameter trends an overall trend for the GWB was calculated according to the following criteria:

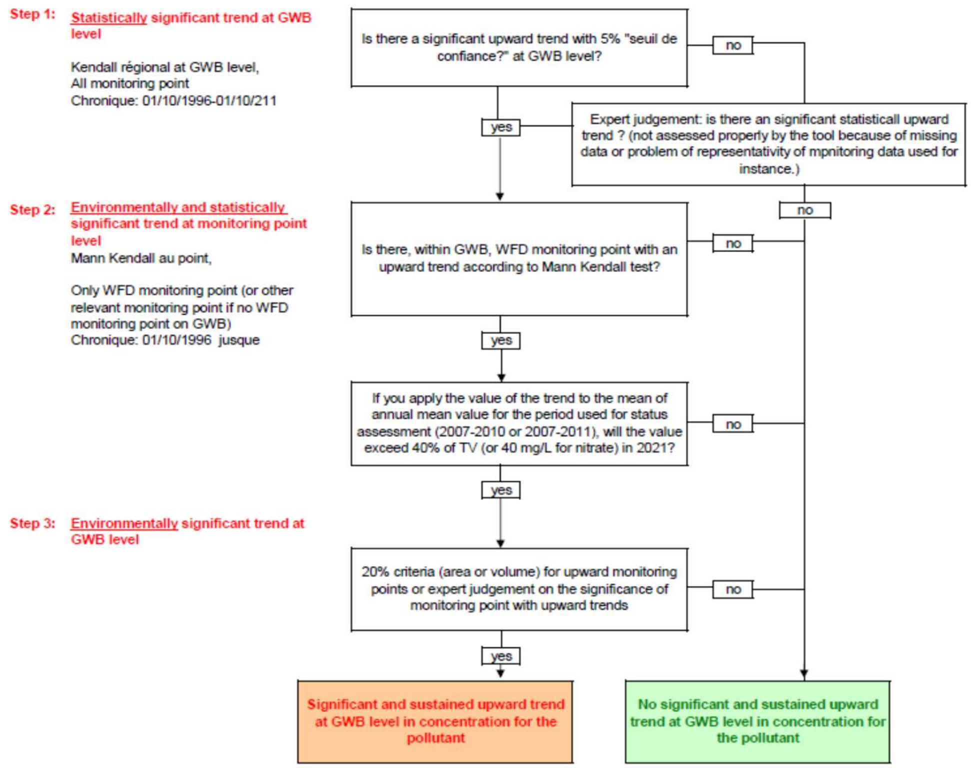
- Not-undesired development (1): At least one parameter has non-adverse development, and the remaining parameters have unknown development
- Unknown development (0): All parameters have unknown development
- Undesired Development (-1): At least one parameter has undesirable development.

A significant upward trend is designated, if a specific parameter is found in concentrations of more than 75% of the TV and if more than 20% of the wells at the same time had a strong upward trend.

Monitoring points with strongly increasing parameter values (> 10%) have been attributed to a "significant upward trend".

Case study: France

Combination of 2 methods: First per groundwater body, if there is a trend at GWB level with all data: determine trend per WFD monitoring point and apply it at the monitoring point to determine whether the concentration at the end of the cycle will be above 40% of the TV. If yes, there is a statistical and environmentally significant trend for this monitoring point. If these monitoring points with trends represent more than 20% of GWB, we consider there is a significant trend for the GWB.



- Kendall regional at GWB level (95% confidence chosen);
- Mann-Kendall at monitoring point (95% confidence chosen);
- Slopes have been calculated using the Sen's slope trend test.

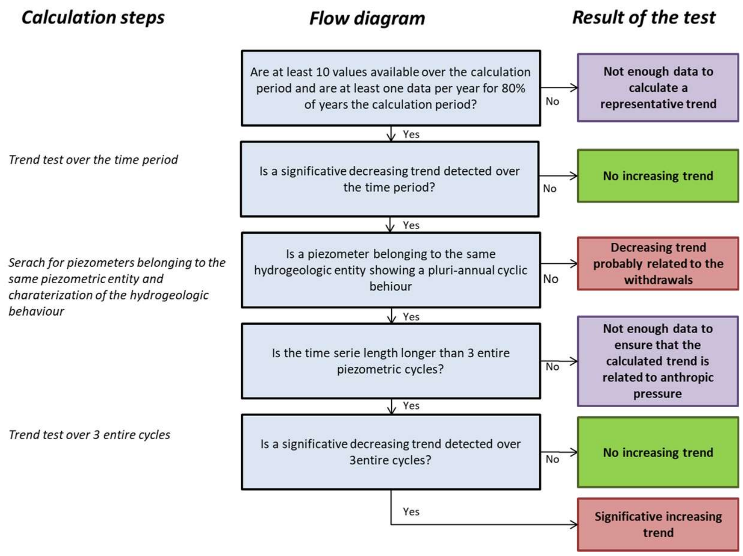


Figure 2: Diagram of the process for determining groundwater quality trends at the sample site scale in France.

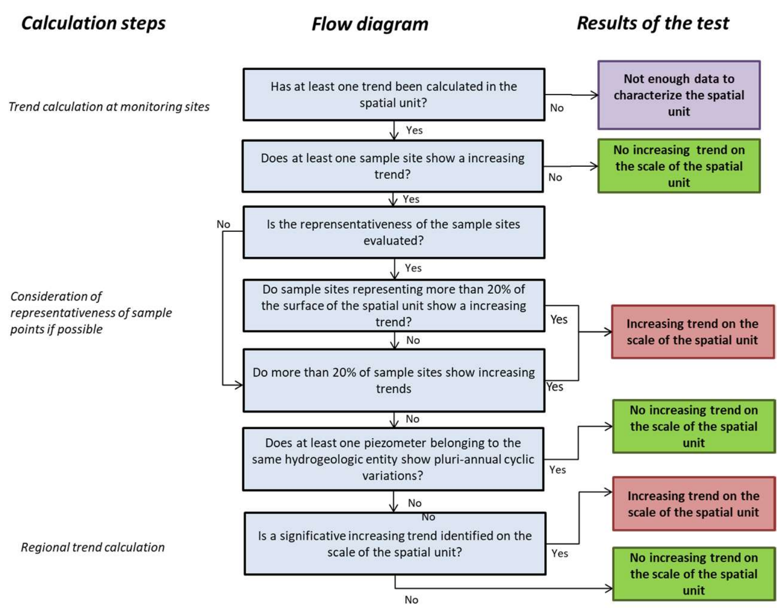


Figure 3: Diagram of the process for determining groundwater quality trends at the scale of a group of sample sites in France.

Case study: Hungary

Trend assessment is carried out at two levels, at the monitoring site level and at the GWB level.

Trend assessment at monitoring site level:

- Trend assessment, based on the non-parametric Mann-Kendall trend test and Sen's method, is carried out for monitoring data time series of each monitoring site at all GWBs (trends categorised as upward, downward or no trend, or not applicable if not enough data are available).

Trend assessment at GWB level:

- Trend assessment of GWBs (also based on the non-parametric Mann-Kendall trend test and Sen's method) is carried out on aggregated data of all monitoring sites within the GWB for the time periods 2000-2012 and 2006-2012:
  - monitoring data of all the monitoring sites aggregated into one annual average, then trend assessment for the whole GWB;
  - 3 levels of reliability:
    - high: min. 5 sites per GWB with time series longer than 6 years;
    - medium: number of sites less than 5, but with time series longer than 6 years; and
    - low: few sites and time series with gaps or inconsistent results of aggregation.

In addition to the trend assessment at GWB level, which is based on the aggregated annual averages of monitoring data of all individual monitoring sites, Sen's trend slope statistics are also determined for GWBs by the aggregation of the Sen's trend slope values of all individual monitoring objects with significant trends. Aggregation of all the individual Sen's trend slope parameters (median, minimum and maximum) is defined for information but is not used for the final assessment of GWB trend.

The statistical trend assessment of GWBs is based on aggregated annual data of all monitoring sites within the GWB. The final evaluation of significant and sustained upward GWB trends is based on those results of the 2000-2012 time period which did not contradict with the significant trends of the 2006-2012 time period. GWBs characterised by significant trends only in the 2006-2012 time period are also considered in the final trend evaluation.

Trend determination is based on the non-parametric Mann-Kendall trend test. Upward or downward trends are considered significant at either of the two confidence levels, 90% and 95%. The magnitude of the trend (trend slope) is obtained by Sen's method. Upper and lower confidence limits (95%) are also reported for the Sen's slope.

A trend is environmentally significant when the trend line exceeds the starting point for trend reversal within the examined time period of 2000-2012. A trend is also considered environmentally significant when the projected trend line exceeds the TV in the next two river basin cycles, that is, until 2027.

In general, 75% of the respective TV is considered as the starting point for trend reversal. In cases of protected GWBs (deep porous intergranular and karst GWBs), a stricter starting point, 30% of the respective TV, is applied according to the Hungarian legislation, but this was not examined for trend reversal. In these (30% of TV) cases particularly, definition of the starting point as a percentage of the TV is considered problematic, as natural background and reversal point values are often very close to each other (within error limits of measurements), and even background and TVs have similar values in some of the GWBs for certain parameters.

Case study: Italy

Trend assessment is performed first at the monitoring station level on annual time series. Monitoring data are first aggregated to annual averages; else, one seasonal value can be selected per year, depending on the seasonality of the data set.

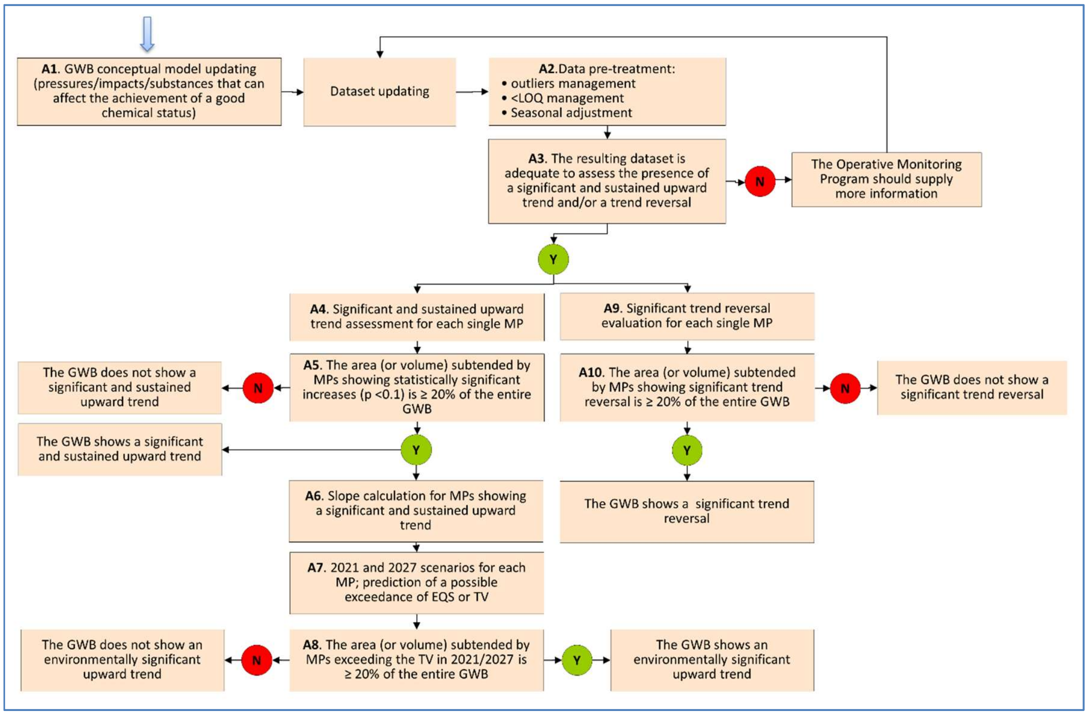


Figure 4 : Diagram of the process for determining groundwater quality trends at sample site and group of sample sites scale in Italy.

Statistically significant trend of pollution is determined by the Mann Kendall test (90% of significance). Sen's slope method is applied to calculate the slope of the trend.

Minimum requirements: There is no minimum number of sites (i.e. one site can be acceptable). At least 8 annual data per monitoring point are required. The applied non-parametric statistical methods allow for missing data; however, datasets with missing data for more than one consecutive year should be discarded. Time series with the last data older than 3 years before the assessment should be discarded.

When the monitoring sites showing statistically significant upward trends represent more than 20% of the GWB, the scenarios at 2021 and/or 2027 are examined. If the trend lines exceed the starting point for trend reversal, the upward trend is considered environmentally significant at the GWB level.

A simplified assessment (comparison of two means) is suggested for GWB not at risk and for those pollutants that do not put the GWB at risk of failing to achieve the good chemical status objective.

Case study: Poland

Trend assessment was carried out only at individual monitoring points; each monitoring point was assessed.

No assessment at GWB level has been undertaken yet due to limited number of monitoring points per GWB with an appropriate length of time series to carry out trend assessment.

Trend analysis was carried out using linear regression test.

An upward trend was classified as environmentally significant if a predicted concentration of a pollutant, calculated using a linear regression, exceeded the starting point for trend reversal by 2021.

Case study: Romania

For each monitoring point the annual mean value (arithmetic average) was calculated for each parameter (substance). Data from all monitoring points was spatially aggregated at the GWB scale.

For each groundwater body, depending on its surface, the GIS matrix was determined through the GISView program. Depending on the number of rows and columns of the GIS matrix, every side of which is 250 m., the relative coordinates were assigned to each monitoring point.

The GIS Matrix and the data files (in Excel 5.0/95), with annual average NO_3, NO_2 or NH4 concentrations in the monitoring points of the GWB were introduced in the computer program, which also estimates the network representativeness - Ru (%).

In case of a homogeneous network (Ru ≥ 80%), the spatial aggregation of data is based on the arithmetic average (AM) and of the confidence limit > 95% (CL (AM)).

In case of a heterogeneously network (Ru < 80%), the spatial aggregation of the data is based on the Kriging average (KM) and on the confidence limit > 95% (CL (KM)).

In this way, the annual mean value per GWB was calculated as mean of all mean values of all monitoring points, for each parameter. Afterwards, the trend for the whole GWB, for each parameter, was determined.

Length of time series considered was 14 years (yearly monitoring frequency).

Statistically significant trend of pollution is determined by the generalised linear regression test (ANOVA) based on the LOESS smoother (95% Confidence level).

No environmental significance was defined for trends (the trends were analysed for all GWBs which were continuously monitored in the 2000 – 2013 period).

Case study: Slovak Republic

In the first step, a trend assessment was carried out at the monitoring point level. Annual average values are assessed.

Subsequently, the trend assessment is performed only for those groundwater bodies, where at least one monitoring point showed a significant and sustained upward trend. Trend assessment is performed at GWB level based on the aggregated values (mean of yearly means) of all monitoring sites with the same length of time series.

For each time series goodness-of-fit tests for normal distribution (Lilliefors test, Shapiro-Wilk test) are calculated. The following methods are used for the identification of a statistically significant upward trend (5% significance level is used for all tests):

- Generalised linear regression test (ANOVA) - if data set follows normal distribution;
- Mann-Kendall test – each time series.

A significant and sustained upward trend occurs if a statistically significant upward trend was identified and:

- the mean of 2 last yearly mean concentrations ≥ 0.75 of the limit value (GW quality standard, TV) and/or
- the predicted concentration for 2021 ≥ limit value.

Case study: Slovenia

The statistical significant trends are determined with nonparametric Spearman's correlation coefficient (r'), with confidence level 0.05 (with 95% certainty). Spearman correlation is run in with STATISTICA software tool.

Spearman (r') can be thought as nonparametric version of Pearson product - moment correlation coefficient (Pearson r). Spearman (r') is a nonparametric measure of statistical dependence of two variables. It tells us how well this dependence is described by the monotone function.

As mentioned above, Spearman (r') assumes that the variables under consideration (one of the variable is time) are measured on at least an ordinal (rank order) scale; that is, the individual observations (cases) can be ranked into two ordered series. Instead of the variable values, Spearman (r') compares their ranks and, it is not sensitive to the normal data distribution and extreme values.

Spearman's classification correlation coefficient is calculated as:

$$r' = \frac{6 * \sum_{i=1}^{n} (R(x_i) - R(y_i))^2}{n(n^2 - 1)}$$

where:

- R(xi) - R(yi) is the difference between the two ranks of each observation
- n is the number of observations

Coefficient values (r') range from 0 (no correlation) to 1 or - 1 (complete correlation). The statistical significance of Spearman's correlation coefficient (r') is tested with the hypothesis that it is statistically significant and different from zero, at confidence level 0.05 (with 95% certainty).

Test of statistical assumption is carried out with zero and alternative hypothesis:

H_0: ρ' = 0 - there is no correlation

H_1: ρ' ≠ 0 - correlation exists

Calculated r' is compared with the tabulated critical value. When the tabulated critical value is exceeded with the higher value of the calculated r', H_0 hypothesis is rejected. The result of the test is a statistically significant correlation between the considered variables.

Case study: UK

a) Trend assessment at individual monitoring points:

Trend assessment was carried out at individual monitoring points to determine (for the 1st river basin management planning cycle):

- Upward trends in pollutant concentrations,
- Trends (upwards or downwards) in concentrations (or values) of natural parameters resulting from natural processes,

Trend assessment was carried out on data from individual monitoring sites for individual parameters and for total pesticides. Where individual parameter concentrations (or values) were reported as below the LOQ, values were replaced by values with half of the reported LOQ, except for total pesticides. For example, a value reported as <0.1 µg/l would become 0.05 µg/l. Trend assessment was not carried out on datasets that comprised more than 80% of values below the LOQ.

b) Trend assessment - Groundwater Body

To determine whether a GWB had a significant and sustained upward trend, the results from the analysis of trends at individual monitoring sites were assessed initially (as in (a)). Where one or more statistically significant trends is identified across a GWB, this trend must be tested for environmental significance, i.e. would lead to failure of one or more chemical status tests within 2 river basin cycles (12 years). Depending on the test(s) for environmental significance being applied (i.e. which status assessment criteria are applied); the number of monitoring points used may vary. For example, for TVs related to the DWPA test, an individual monitoring point (drinking water abstraction) that indicates an environmentally significant upward trend would lead to the GWB being reported as having environmentally significant upward trend. For TVs related to the general chemical status test an aggregation of the results of trend assessment from all relevant monitoring sites is required to test for environmental significance.

If environmentally and statistically significant upward trends are identified for one or more parameters (pollutants or indicators), the GWB was reported as having a significant and sustained upward trend.

Note: the presence of a statistically significant trend at an individual monitoring point will not on its own lead to a requirement to report that the GWB has an upward trend because the trend also needs to be environmentally significant.

In many groundwater systems there will be considerable seasonal variability in parameter concentrations. This variability may introduce problems in the trend analysis unless it can be corrected. Where there are sufficient data within a given year, the non-parametric Seasonal Kendall trend test was used. Where there were variable or insufficient data seasonality was removed by calculating the annual means, and then trend analysis performed on the annual means. In this case Sen's Method was used (Sen, 1968). Both the Seasonal Kendall test and Sen's Method are robust methods that allow for some missing data in the time series and are not badly affected by gross errors or outliers in the data series. Additionally linear regression analysis was carried as a further test for trends. For a trend to be statistically significant there should be at least 80% confidence. For England and Wales 90% confidence was used.

The test of environmental significance was determined by examining whether the trend(s) at the monitoring point would lead to the body failing one or more of its environmental objectives within the period covered by two river basin cycles (12 years). This was assessed by comparing predicted concentrations at the end of the two river basin cycles with the TVs and status assessment criteria relevant to the monitoring point.

Case study: Wallonia (Belgium)

In the framework of the EU FP6 IP AquaTerra project, the following procedure was established for trend detection and quantification for nitrate time series at individual monitoring stations (point by point analysis) in groundwater bodies of the Walloon region. A normality test was first applied on NO_3 datasets. If the dataset is normally distributed, linear regression is used for trend detection and quantification. When normality cannot be assumed, the non-parametric Mann-Kendall test is used for detection and Sen (1968) slope for quantification (also called Kendall-Theil). Both techniques were compared for trend detection in nitrate time series from the Geer basin aquifer and it was concluded that linear regression and Man-Kendall test and Sen slope analysis produce the same results for normally distributed data set. Man-Kendall test and Sen slope analysis can thus be used for both normally and non-normally distributed datasets.

A procedure was also developed for pesticides, considering non-detect values and applied at the scale of the Geer basin chalk aquifer. Chemical data on pesticides were aggregated at the scale of the unconfined part of the chalk aquifer. Non-parametric Mann-Kendall test and Sen slope was then used respectively for trend detection and quantification. Pesticides non-detect values were taken into account in trend assessment using the Regression on Order Statistics method.

3.2. GROUNDWATER TREND REVERSAL ASSESSMENT

3.2.1. Objective

According to Preamble 26 of the WFD: "*For groundwater, in addition to the requirements of good status, any significant and sustained upward trend in the concentration of any pollutant should be identified and reversed."* and Annex IV 2.4.4 requires that: "*Reversal of a trend shall be demonstrated statistically and the level of confidence associated with the identification stated."*

The assessment of trend reversal allows for identifying a reduction of the pollution of groundwater and whether the implemented measures are effective of breaking upward trends of pollutants concentrations resulting from the impact of anthropogenic activity.

Following CIS No.18 (chapter 2.1.3.), for the assessment of trends and trend reversal the conceptual model plays a key role as follows:

- when considering the physical and chemical temporal characteristics, including groundwater flow conditions, recharge rates and percolation time through soil or subsoil.

- when selecting monitoring location and frequencies to provide the information necessary to ensure that significant upward trends can be distinguished from natural variations with an adequate level of confidence and precision.

- when establishing starting points for trend reversal that are different to 75% of the groundwater quality standard (GW-QS) or TV which will depend on the characteristics, the aquifer and the ability to prevent most cost-efficiently any environmentally significant detrimental change in groundwater quality.

Member States must identify the starting point for trend reversal such that trends can be reversed in time to avoid a (future) failure of relevant environmental objectives (GWD Article 5(3) and Annex IV(B)). This starting point must be defined as a percentage of the level (or concentration) of the relevant groundwater quality standard or threshold value, and reported in the river basin management plan. (CIS No.18, chapter 6.2.).

When defining starting points for trend reversal, the length of time between the starting point and the point of exceedance of a GW-QS or TV should be sufficient for the trend assessment methodology to be able to detect a significant trend, i.e. there is time to detect that a trend is environmentally significant and take action to reverse the trend. The ability of a trend method to detect a given increase in concentrations of pollutants with a certain probability is called “power” of a trend test method. (CIS No.18, chapter 6.2.6).

The starting point needs to take into account the environmental risk(s) associated with the GWB, environmental objectives and the groundwater quality standard (GW-QS) and/or threshold values (TVs) established for the GWB. The starting point shall be a percentage of the GW-QS or TVs. By default, the starting point shall be when the concentration of the pollutant reaches 75% of the relevant GW-QS or TV. The starting point for implementing measures to reverse trends depends mainly on the characteristics of the groundwater body and its ability to respond to those measures […]. Once a starting point has been established for a trend, it should not be changed during the six-year cycle of the river basin management plan. (CIS No.18, chapter 6.2.8).

In the CIS Technical Report No. 1 a methodology for assessing trend reversal is described. It is based on a regression analysis, where each time series is analysed for whether there is a break in the trend. This is the case where an identified sustained and significant upward trend is followed by a significant downward trend. (CIS No.18, chapter 6.2.9).

3.2.2. Implementation

Case study: Austria

Trend reversal is following the method proposed in CIS Technical Report No. 1 and assessed by an extended linear regression model where the time series is divided into 2 sections (legs) and the inclinations are compared. The choice of time sections is optimised. The significance level is 5%.

The left leg needs a length of at least 4 years (without the break) and a positive slope and the right leg needs a minimum length of 2 years (without the break) with a negative slope for half-yearly and quarterly aggregations and 3 years for annual values.

Length of time series: 14 annual values or values from 10 years for half-yearly and quarterly aggregations.

Further criteria are identical with those for trend assessment:

- at least 3 sites per GWB;
- at least 2/3 of all sites per GWB have to have sufficient data;
- only 1 missing value in the time series of each monitoring is acceptable (but not at the beginning or end of the time series); and
- at least 60% of the values are quantified (above LOQ).

Case study: France

To search for a trend reversal on a groundwater quality record, the non-parametric Darken test is used. It is based on the comparison of the Kendall's Tau for 2 sections in the time series.

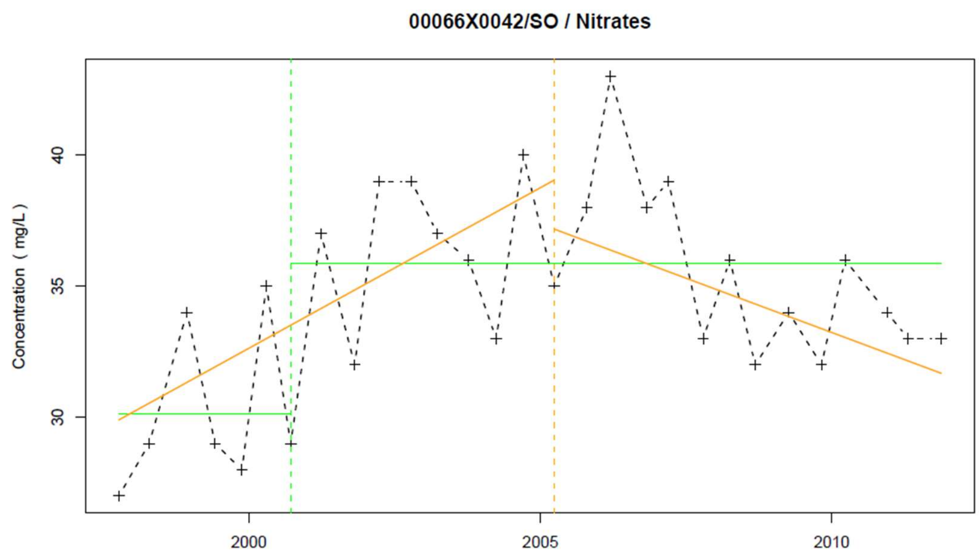


Figure 5 : Trend reversal Darken test on nitrate time series at a French groundwater site (00066X0042/SO).

Case study: Italy

The trend reversal analysis aims at statistically demonstrate the reversal of the trend. Hence, the reversal test should be applied to those monitoring sites which have shown a significant and sustained upward trend.

The national guideline suggests the Pettitt test (Pettitt, 1979), which is currently under test for the reversal analysis. This is able to identify the existence of a changing point in the dataset when a sub-period with a statistically significant trend is followed by period with no statistically significant trend. The level of confidence required is 90%. When the test rejects the null hypothesis at 90% confidence level, the time series is considered to be reversed.

At the GWB level, the sites where the reversal of the trend is demonstrated must represent at least 20% of the GWB. In addition, one should verify that there are no upward trends in other monitoring sites representing more than 20% of the GWB.
The minimum number of sites must be coherent with the number of sites used for the upward trend assessment. At least 14 annual data are required for a range of 14 to 30 years of monitoring. Some missing data in the time series are acceptable; however, datasets with missing data for more than one consecutive year should be discarded. Time series with the last data older than 3 years before the assessment should be discarded.

Case study: Romania

Trend reversal is assessed based on the assumption that time series can be characterized by 2 linear trends with a slope direction change within the analysed time period.

Trend reversal can be detected if in the first section the slope is ascending (positive) and in the second section the slope is descending (negative), applying 95% quantile of the distribution.

Method steps are: optimization of time sections selection; analysis of slope change significance for the model of simple linear regression; statistical test to verify that the model of the 2 sections is more significant than the model of simple linear regression.

Case study: UK

Trend reversal was not assessed in the 1st RBMP cycle as it was considered that upward trends had to be determined and reported first. Trend reversal, if achieved, would be reported in the 2nd RBMP cycle. The methodology that will be followed is:

- By default the starting point for trend reversal will be when pollutant concentrations reach 75% of any relevant TV. Alternative earlier or later starting points can be set if it can be demonstrated they are needed to avoid a future failure of WFD objectives or can be justified in terms of cost-effectiveness and technical feasibility.
- Reasons for selecting starting points have to be reported in the RBMP. Once a starting point has been reported in the RBMP it cannot be changed until the next river basin cycle.

The results of trend assessment should be taken into account when determining the starting point for trend reversal. Where a trend indicates that there is likely to be a failure of one or more environmental objectives, before the end of the river basin cycle then this failure should be used taken into account when defining a starting point. A starting point must be chosen to allow measures to be put in place to reverse trends in sufficient time to avoid failure of WFD environmental objectives.

Case study: Wallonia (Belgium)

In the scope of a regional project on NO_3 trend assessment and reversal for surface water and groundwater, the following model procedure was applied. Temporal outliers are discarded. If less than 10 data are available, the time series cannot be analyzed and the procedure is terminated. In a second step, statistical outliers are discarded. Three models are compared using a parametric analysis: M_0: a constant model (no change in time); M_1: a simple linear model and M_2: a trend reversal model imposing the continuity at the breakpoint (i.e. two linear regressions on both sides of the breakpoint imposing that their values are equal at the breakpoint).

The selected model is the one with the smallest value for the BIC criterion. If a M_2 model is selected but the difference in BIC value with the M_1 model is less than 2, or the two slopes are not significant, the simpler M_1 model is kept. If a M_1 model is selected but the slope is not significant, we fall back to the simpler M_0 model.

A normality test is performed on the residuals. If normality is not rejected, the analysis stops here. If the residuals are not normally distributed, a non-parametric analysis is performed: The statistical outliers are reincorporated. Non-parametric regressions are computed for the M_1 and M_2 models.-Non-parametric Mann-Kendall test is used for trend detection and Sen slope for trend quantification (also called Kendall-Theil).

The selected model is the one with the smallest Sum of Squared Residuals (SSR) value, since no BIC nor p-values can be calculated. Again, in case of non-significant slopes in the selected model, the simplest model is kept, as exposed in the parametric approach.

Trend reversal was determined using linear regression (normally distributed residuals) and non-parametric regression models (non-normally distributed residuals) with breakpoints estimation (imposing continuity at the breakpoints). Trend reversal is achieved if the slopes on both sides of the breakpoint show a different sign, or when one slope is significant while the other is not.

To identify the breakpoint in the parametric case, the M_2 model is fitted for all the dates along the time series and the one with the smallest variance of residuals is selected.

In the non-parametric case, a robust estimation of the slopes is performed on both sides of the breakpoint using the Sen method. As this method only yields a slope, we search for the value at the origin minimizing the Sum of Squared Residuals (SSR) on both sides. Finally, the date of the breakpoint is searched by minimizing the total SSR.

4. Statistical tools

Several MS developed or adapted tools for statistical analysis of time series that make it possible to conduct the majority of tests presented in the technical report.

4.1. FRENCH GEOLOGICAL SURVEY (BRGM) HYPE TOOL

HYPE, a statistical tool for analysis of trends and breaks in groundwater quality records, was developed in 2013 as part of the ONEMA-BRGM agreement. This tool provides valuable assistance in understanding temporal variations in groundwater contaminants.

Several years of work on the applicability of statistical methods in the domain of groundwater quality and of tests on theoretical and real data have made it possible to construct a decision tree leading to the most appropriate statistical method for the study of the characteristics of the series. It makes it possible to avoid errors resulting from the use of statistical tests outside their conditions of applicability.

This tool is programmed in R language; it is very easy to use, does not require any skill in programming, and a manual is available for users (Croiset et al., 2013). The tool and manual are freely downloadable on the BRGM website.

HYPE software can be freely download here (in French): http://www.brgm.fr/production-scientifique/logiciels-scientifiques/formulaire-telechargement-logiciel-hype

4.2. PRAGUE INSTITUTE TREND TOOL

G. Masaryk Water Research Institute developed a dedicated statistical tool for trend assessment based on Excel software.

4.3. GWSTAT SOFTWARE – TREND ASSESSMENT

The underlying methodology is described in detail in the CIS Technical Report No. 1 and the accompanying Annexes (www.wfdgw.net).

ANOVA linear regression generalized test

- data adjustment (regularisation): calculation of the annual average of available concentration values for each point of the monitoring network (well, spring);

- input of annual average concentrations / annual review of the trend option (because a large number of monitoring sites had only one analysis per year);

Spatial aggregation of data for the whole GWB surface was necessary: GIS matrix through GISView Program

Trend Reversal Assessment

Model of the 2 sections

Based on the assumption that time series can be characterized by 2 linear trends with a slope direction change within the analysed time period.

Applying 95% quantile of the distribution, reversal can be detected if in the first section, the slope is ascending (positive) and in the second section, the slope is descending (negative).

Method steps:

1. optimization of time sections selection;

2. analyse of slope change significance for the model of simple linear regression;

3. statistical test to verify that the model of the 2 sections is more significant that the model of simple linear regression.

Practically, processing the introduced data series, the program can indicate an inversion in the trend slope, thus a trend reversal.

4.4. ITALIAN WATER RESEARCH INSTITUTE EXCEL SPREADSHEET "TRENDY"

An excel spreadsheet has been developed for the trend assessment of chemical data sets on a single monitoring station. It allows for identification of statistically significant trends in water quality records. Specific values of concentration threshold, LOQ and significance threshold for the Mann-Kendall test have to be entered in the spreadsheets by the user.

In the first sheet, the available monitoring data for a particular compound are averaged on a yearly basis; averaged values are automatically plotted onto a graph (concentration versus time). Values < LOQ are transformed into LOQ/2.

In the second sheet, the non parametric Mann Kendall test is applied on the annual values. Because the annual data set is always < 40, the S statistics from Hollander et al (1973) and Kendall (1975) for the one-tail test was implemented in the spreadsheet. The confidence for a one-tail distribution is calculated and the null hypothesis is verified against a 90% significance threshold.

Then, the slope of the trend is computed with the non-parametric Sen's method (Sen, 1968). The theoretical scenario of the concentration at 2021 and 2027 is calculated by a linear extrapolation of the slope. For those trends starting with lower than the threshold values, the theoretical date when the 75% and 100% of the threshold is reached is supplied.

Finally, the Pettitt test (Pettitt, 1979) can be applied to identify the existence of a changing point in the dataset, in order to distinguish a sub-period with a statistically significant trend from a sub-period characterized by no-trend. The same test can be also adopted to distinguish a period with no-trend from a period with a statistically significant trend. The level of confidence associated with the identification is also calculated. The procedure, still under evaluation, is implemented as an excel "macro" embedded in the spreadsheet.

4.5. GROUNDWATER SPATIO TEMPORAL DATA ANALYSIS TOOL (GWSDAT)

The GroundWater Spatiotemporal Data Analysis Tool (GWSDAT) is an open source software tool to analyze and report trends in groundwater quality monitoring data (Jones et al. 2015). GWSDAT is based in the open source statistical programming language R and Microsoft Excel. GWSDAT 3.0 is currently being developed which will run in a web browser, without the need to install R and Excel.

GWSDAT's primary use is for interrogation and interpretation of groundwater monitoring data derived from contaminated sites. The software can handle large data sets with multiple monitoring locations, variable sampling events, and differing chemical constituents. GWSDAT has been used extensively in the assessment of soil and groundwater conditions at numerous sites around the world.

The main functionalities of GWSDAT include trend analyses, data smoothing, spatiotemporal smoothing, and determination of contamination plume characteristics. Groundwater quality observation data for individual monitoring points can be fitted to a linear or log linear regression model, where the significance of a trend can be determined by using the Mann–Kendall approach, which is widely used for trend detection in groundwater and surface water studies. Groundwater concentration maps are developed by using a spatiotemporal solute concentration smoother derived from a penalized-splines nonparametric regression. The simultaneous statistical smoothing over space and time generally provides a more accurate, more consistent illustration of a contamination plume when compared against contour maps of individual sampling rounds (McLean et al. 2018).

Further details on the input and output features of GSWDAT can be found in the supporting information. Additional information on each of the methodologies (including descriptions of the R packages used, but excluding the plume metrics, which are added in the latest version of GWSDAT) can be found in Jones et al (2014).

5. General conclusions

The European Water Framework Directive (WFD) asks Member States to identify trends in contaminant concentrations in groundwater and to take measures to reach a good chemical status by 2015, 2021 and 2027.

The synthesis of the procedures applied by Member State (MS) to assess trend in groundwater quality for the first River Basin Management Plan reveals the high diversity of procedures and methods that can be applied to respond to a simple and unique question: what are the trends in groundwater quality? MS have thought a great deal about this not so simple question and have proposed a variety of solutions to assess trends in groundwater quality. This compilation exercise enables comparison of methodologies and highlights the need to go further in analysing all existing methodologies, in order to identify the best practices and to provide recommendations for groundwater quality assessment under the WFD

For the first RBMP cycle, a majority of MS has chosen to apply statistical methods to comply the trend assessment requirement. The methods used vary between RBMPs dependent on data, procedure or trend and trend reversal assessment methodology applied. Statistical methods used by MS to identify trends in contaminant concentrations in groundwater could be divided in 2 groups: i) parametric test ANOVA, based (or not) on the LOESS smoother; and/or ii) non-parametric Mann-Kendall test (and derivate Seasonal Kendall and Regional Kendall tests). However, non-parametric statistical tests generally are preferred to parametric tests when the analysis of environmental data is involved, particularly in the absence of pretreatment of raw data.

Environmental significance was mainly estimated based on exceedance of a TV or a percentage of a TV for a predicted concentration. The date for the prediction was often defined as the date when the trend line exceeds the starting point for trend reversal. These dates vary between MS from 2017 to 2021 or 2027.

Major issues remain difficult to tackle: the improvement of monitoring design for trend identification, values reported below the limit of quantification for micropollutants trend assessment, and the spatial distribution and scaling of trends from single monitoring sites to Groundwater Bodies (GWBs) scale.

Although the different tools developed make it possible to identify trends, they cannot be used alone. We must still seek to explain trends in order to be able to address the causes and attempt to reverse them if the trends are upward. This complex work often requires searching for and compiling local data, notably on anthropogenic pressure and climate changes, and a better understanding of local hydrogeology and hydrodynamics.

6. Bibliographic References

REGULATORY TEXTS / GUIDANCE

CIS (2001) – Statistical aspects of the identification of groundwater pollution trends and aggregation of monitoring results – Working Group 2.8 Statistics. Technical Report No. 1. ISBN 92-894-5639-6. European Communities, Luxembourg

CIS (2007) – Guidance on Groundwater monitoring. Guidance Document No. 15. ISBN 92-79-04558-X. European Communities, 2007.

CIS (2009) – Guidance on groundwater status and trend assessment, Guidance Document No 18. Technical Report - 2009 - 026. ISBN 978-92-79-11374-1. European Communities, Luxembourg

European Parliament and Council Directive 2000/60/EC of 23 October 2000 establishing a framework for Community action in the field of water policy, OJ L 327, 22.12.2000, pp1-72.

European Parliament and Council Directive 2006/118/EC of 12 December 2006 on the protection of groundwater against pollution and deterioration, OJ L372, 27.12.2006, pp19-31.

ISPRA - IRSA.CNR (2017) - Linee guida per La valutazione delle tendenze ascendenti e d'inversione degli inquinanti nelle acque sotterranee (DM 6 luglio 2016). (Guideline for the assessment of upward trends and trend reversal of pollutants in groundwater, Ministry Decree 6 July 2016). http://www.isprambiente.gov.it/it/pubblicazioni/manuali-e-linee-guida/linee-guida-per-la-valutazione-delle-tendenze-ascendenti-e-dinversione-degli-inquinanti-nelle-acque-sotterranee-dm-6-luglio-2016

TECHNICAL REPORTS

Austrian Federal Ministry of Agriculture and Forestry, Environment and Water Management (Ref.: 41.046/01-IV1/00 and GZ 16 2500/2-I/6/00), European Commission (Grant Agreement Ref.: Subv 99/130794), in kind contributions by project partners. Vienna.

Bloomfield J.B., Jackson C.R., Stuart M.E., (2013), Changes in groundwater levels, temperature and quality in the UK over the 20th century: an assessment of evidence of impacts from climate change. Working Technical Paper. http://www.lwec.org.uk/publications/water-climate-change-impacts-report-card/1-groundwater-temperature-quality, LWEC, 14pp.

Croiset N., Lopez B. (2013) *HYPE: Outil d'analyse statistique des séries temporelles d'évolution de la qualité des eaux souterraines – Manuel d'utilisation.* [Statistical analytical tool for time series variations in groundwater quality - User's Manual] Final report. BRGM/RP-63066-FR. 64 p., 33 fig.

D'Or D. and Allard. D., (2014), *Mise en évidence de tendances éventuelles sur les séries chronologiques présentées par les points du réseau de mesure wallon des eaux de surface et souterraines en ce qui concerne les nitrates*. Rapport final. Rapport Ephesia Consult RP DGO3 2014002 - Novembre 2014.

J. Grath, A. Scheidleder, S. Uhlig, K. Weber, M. Kralik, T. Keimel, D. Gruber (2001): "The EU Water Framework Directive: Statistical aspects of the identification of groundwater pollution trends, and aggregation of monitoring results". Final Report.

Lopez. B., Leynet A., Blum A., Baran N. (2011) *Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines.* [Assessment of pollutant concentration trends groundwater.] *Revue des méthodes statistiques existantes et recommandations pour la mise en œuvre de la DCE*. [Review of existing statistical methods and recommendations for the implementation of the WFD] Rapport final [Final report] BRGM/RP-59515-FR

Report from the commission to the European parliament and the council on the Implementation of the Water Framework Directive (2000/60/EC) River Basin Management Plans. (Text with EEA relevance). Brussels, 14.11.2012. COM(2012) 670 Final.

SCIENTIFIC LITERATURE

Batlle-Aguilar, J., Orban, Ph., Dassargues, A., Brouyère, S., (2007), Identification of Groundwater quality trends in a chalk aquifer threatened by intensive agriculture in Belgium, Hydrogeology Journal, 15(8), pp. 1615-1627.

Hakoun, V., Orban, Ph., Dassargues, A., Brouyère, S., (2017), Factors controlling spatial and temporal patterns of multiple pesticide compounds in groundwater (Hesbaye chalk aquifer, Belgium), Environmental Pollution, 223, pp. 185-199.

Helsel D.R., Hirsch R.M., (1992), Statistical method in water resources, Studies in Environmental Science 49, Elsevier, Amsterdam.

Hollander, M., Wolfe, D. A., & Chicken, E., (2013), *Nonparametric statistical methods*. John Wiley & Sons.

Jones, W.R., Spence, M.J. and Bonte, M. (2015) Analyzing Groundwater Quality Data and Contamination Plumes with GWSDAT. Groundwater 53(4), pp. 513-514.

Jones, W.R., M.J. Spence, A.W. Bowman, L. Evers, and D.A. Molinari. 2014. A software tool for the spatiotemporal analysis and reporting of groundwater monitoring data. Environmental Modelling & Software 55, May: pp. 242–249.

Kendall, M. G., (1975), *Rank Correlation Methods*, 4th ed. Charles Griffin, London.

Lopez, B., Baran, B., Bourgine, B., (2015), An innovative procedure to assess multi-scale temporal trends in groundwater quality: Example of the nitrate in the Seine–Normandy basin, France. Journal of Hydrology 522, pp. 1–10.

Mann, H. B., (1945), *Nonparametric tests against trend*. Econometrica: Journal of the Econometric Society, pp. 245-259.

McLean, M., Evers, L., Bowman, A., Bonte, M. and Jones, W.R. (2018) Statistical modelling of groundwater contamination monitoring data: A comparison of spatial and spatiotemporal methods Science of the Total Environment In press.

Pettitt, A.N., (1979), *A non-parametric approach to the change-point problem*."Appl. Statist., 28, pp. 126-135.

Sen P.K., (1968), - Estimates of the regression coefficient based on Kendall's tau", Journal of the American Statistical Association, 63 (324): pp. 1379–1389, doi:10.2307/2285891.

Visser, A., Dubus, I., Broers, H. P., Brouyère, S., Korez, M., Orban, P., Goderniaux, P., Batlle-Aguilar, J., Surdyk, N., Amraoui, N., Job, H., Pinault, J.-L., & Bierkens, M., (2009), Comparison of methods for the detection and extrapolation of trends in groundwater quality. Journal of Environmental Monitoring, 11, pp. 2030-2043.

Welham, S. J., Cullis, B. R., Gogel, B. J., Gilmour, A. R., and Thompson, R., (2004), Prediction in linear mixed models. Australian and New Zealand Journal of Statistics, 46, pp. 325–347.

Appendix 1: Technical description of some statistical methods

Mann-Kendall

•*Application conditions*

Applies to a sample of n **independent values** (not correlated among themselves) to detect **monotonic** trends.

•*Hypothesis tested (= Null hypothesis H_0)*

Stationarity of the series

•*Implementation of the test*

The test rests on the calculation of the following sum:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(x_j - x_k)$$

where x_j and x_k are two successive values in the series (j >k) and where:

$$\mathrm{sgn}(x_j - x_k) = \begin{cases} 1 \text{ si } x_j - x_k > 0 \\ 0 \text{ si } x_j - x_k = 0 \\ -1 \text{ si } x_j - x_k < 0 \end{cases}$$

A count is made of all couples (x_j , x_k) and the number of cases in which the 2nd value is greater than the 1st and the number of cases where the numbers are reversed, and the difference between these two numbers is calculated. If the series is stationary, S must be near 0.

The test does not require a hypothesis on the probability distribution of values (non-parametric test). If the series is stationary (Hypothesis H_0), then:

- the mean of S is: E(S) =0
- and its variance is: Var(S) = n(n-1)(2n+5)/18

If there are any equal values in the series, the variance Var (S) must be calculated as follows:

$$\mathrm{Var}(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} E_p (E_p - 1)(2E_p + 5) \right]$$

where Ep designates the number of equalities identified in the series involving p values.

If the number n of values in the series is greater than 10, implementation of the test causes the random variable Z to play a part; Z is defined as:

$$Z = \frac{S-1}{\sqrt{\mathrm{Var}(S)}} \text{ if } S > 0$$

$$Z = \frac{S+1}{\sqrt{\mathrm{Var}(S)}} \text{ if } S < 0$$

$$Z = 0 \text{ if } S = 0$$

A positive value of Z indicates an increasing trend (cf. remark below)

This variable follows (approximately) the standard normal distribution. The probability of observing a trend for which the value of Z is at least equal in absolute value to z_{obs} is thus:

$$P(|Z| \geq |z_{obs}|) = 2(1 - \Phi(|z_{obs}|))$$

where Φ is the normal distribution function.

The hypothesis of stationarity will be rejected at the significance level α if this calculated probability (p-value) is less than α (it is unlikely).

The result of the test, in terms of significance α, is also obtained by comparing the value z_{obs} obtained for Z to the theoretical value z_{th} of the cumulate distribution of the standard normal distribution (distribution function). For a two-tailed test aimed at the detection of a trend with a confidence level, α, H_0 is rejected if the absolute value of Z is higher than $Z_{1-\alpha/2}$.

if $|z_{obs}| > z_{th(1-\alpha/2)}$, dismissal of H_0 at confidence level α

Slope calculation: Sen method

A commonly used method is Sen's method (Sen, 1968); the slope is calculated as follows:

$$a_{Sen} = \text{Médiane}\left[\frac{(x_j - x_i)}{(j-i)}\right] \quad \forall\ j < i$$

x_j and x_i are respectively jth and the ith observation in the series.

The slope thus calculated is a more robust estimation of the slope of the trend than that indicated from the regression where the existence of very high and very low values can bias the slope calculation. On the contrary, with Sen's method, the slope is not very affected by the presence of extreme values.

Regional Kendall

The regional Kendall method, proposed by Hirsch et al. (1982) makes it possible to estimate trends at the scale of a region of the study, including several wells. The principle of the test is to determine if a coherent trend can be demonstrated by different records.

The Kendall statistic S_k is calculated from the sum of the statistics for each well (Hirsch et al., 1982).

$S_k = \sum_{i=1}^{s} S_i$ where s is the number of wells and S_i are the statistics S of Mann-Kendall$S_i = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn\ [(y_{j_i} - y_{k_i})(x_{j_i} - x_{k_i})]$

And $\sigma_{S_k} = \sqrt{\sum_{i=1}^{S} n_i(n_i - 1)(2n_i + 5)/18}$ where n_i is the number of data for the well i.

The calculated statistic is:

$$Z_{S_k} = \frac{S_k}{\sigma_{S_k}}$$

If the product of the number of wells is greater than 25 years, the distribution of S_k can be approximated by a normal distribution and the variance equal to the sum of variances. $S_k$is standardized by subtraction of the mean (equal to 0) and division by its standard deviation. The result is evaluated by comparing it to a table of normal standard distribution.

The null hypothesis is to be rejected at a level of significance α if$|Z_{Sk}| > Z_{crit}$ where Z_{crit} is the normal distribution value with an exceedance probability of $\propto/2$

The Sen slope calculation and the ordinate at the origin are slightly modified compared to the Mann-Kendall test: the slope is the median of slopes calculated between two analyses **for the same well**, the calculated slopes between data from different well points are not taken into account.

Darken Test - trend reversal

Darken proposes in his thesis (Darken, 1999) a method to detect a sign change in the slope (and not a change in the magnitude of the trend without a change in sign) which is valid for non-normally-distributed data.

For one well, for each time t the Darken statistic can be calculated using the following formula:

$$Z_t = \frac{\tau_1 - \tau_2}{\sqrt{Var(\tau_1) + Var(\tau_2)}}$$

where τ_1, $Var(\tau_1)$, τ_2 and $Var(\tau_2)$ are respectively the *tau* of Kendall and its variance for period 1 starting from the initial date to t and a period 2 starting from t to the final date. The calculation of *tau* is described by Kendall (1975).

The most probable rupture date is the date for which Z is maximum.

The p-value of the test is then calculated by comparing the Z statistics for the rupture date identified to the quantiles of the standard normal distribution (Darken, 1999).

Autocorrelation calculation

Autocorrelation makes it possible to evaluate to what extent a measurement at a given time is correlated to a subsequent or preceding measurement over a given time interval. An autocorrelation coefficient is calculated for each interval tested. The closer the autocorrelation coefficients are to 1, the higher the correlation. Correlation increases as the autocorrelation coefficient approaches 1. A slow decrease in autocorrelation coefficients for increasing intervals is the marker of inertial functioning of the aquifer. When a series shows noticeable cyclic behavior, one can observe peaks relative to the period of the cycles (a positive significant coefficient marks the period or a multiple of the period and a significant coefficient marks a half-period).

Appendix 2: Raw compilation of the methods used by some Member States

CIS Working Group Groundwater

TREND ASSESSMENT:

Raw compilation of the methods used by some Member States

Data capture and assessment

The main issues that had been identified by WGGW were: the improvement of monitoring design for trend identification, how to deal with values reported below the limit of quantification, and the spatial distribution and scaling of trends from single monitoring sites to Groundwater Bodies (GWBs) scale.

A questionnaire with the following twelve questions was elaborated and distributed within WG GW to collate information on the currently applied EU MS approaches and methodologies in assessing trends and trend reversal of pollutants concentrations. This compilation aims at supporting the comparison of the different approaches and of exchanging experiences and best practice between MS. Ten MS replied to the questionnaire and provided information: Austria, Czech Republic, Denmark, France, Netherlands, Poland, Romania, Slovak Republic, United Kingdom and Hungary.

List of Questions:

I. Which GWB and which substances?
II. Which data?
III. First determine trend per monitoring point, then per groundwater body? Or lump all data, then determine trend? From monitoring point to GWB level.
IV. Which statistical method used for trend determination? And its confidence
V. How environmental significance was defined for trend?
VI. Number of years of monitoring results required to determine trends
VII. Methods used in case of missing data or too few data
VIII. Monitoring density (spatial) for trends at GWB level?
IX. Results below Limit of Quantification
X. Can you use expert judgement for trend determination?
XI. Consideration of baseline level of concentration?
XII. Methodology for trend reversal assessment?

Member States responses to questions

==

Which GWB and which substances?

Austria:

- Trend assessment is performed for all GWBs and substances where the starting point for trend reversal is exceeded by the 3-year-average (which is the appraisal period) of annual mean values in at least 30% of all monitoring points within a GWB.
- Starting points for trend reversal (and TVs) are laid down for all Annex I and (updated) Annex II substances and for benzene, B, Cr, 1,2-dichloroethane, Cu, Ni, sum of PAHs, hydrocarbon index, TRI+PER and trihalomethane (total).

Czech Republic:

- All GWBs
- All assessed substances (except pesticides and their metabolites – too short time series, most of results below limit of quantification), but significant upward trend was found for 16 substances only – metals, selected PAHs, nitrates, nitrites, orthophosphates, chlorides and sulphates.

Denmark:

- All GWB, for which chemical analyses have been available for at least 5 monitoring points for each of the two periods 2000-2006 and 2007-2013.
- Nitrate, Cadmium, Lead, Mercury, Sulphate, Pesticides, Chlorinated solvents, BTEX.

France:

- Nitrate on all French GWBs.
- All substances on all GWBs from Guadeloupe RBD.
- Sulphate, conductivity, chloride, ammonium and arsenic on all GWBs from La Réunion RBD.
- Locally, parameters placing GWBs at risk like chloride (though not always possible/easy for all parameters causing risk: e.g. pesticides...)
- No trend assessment performed for the chemical substances that have been determined in previous studies as being of potentially high concentration due to the natural geochemical background.

Netherlands:

- All GWBs.
- Substances mentioned in Annex I (Nitrate and pesticides) plus substances with TVs (chloride, nickel, arsenic, cadmium, lead and phosphate).

Poland:

- All GWBs for which water quality data have been available.
- All analysed substances (pollutants and natural parameters, except organic substances – too short time series, most of results below limit of quantification - LOQ).

Romania:

- All GWBs for which chemical analyses have been available for a time series of at least 8 years (64 GWBs from the total of 143);
- Trend assessment was carried out for all pollutants/indicators that are causing risks for GWBs, identified through the risk assessment process, also for the GWBs which are not found at risk. These substances are the nitrogen compounds: Nitrates, Nitrites and Ammonia (NO3, NO2 and NH4),

Slovak Republic:

- All GWBs, for which chemical analysis is available for at least 6 years;
- All substances stated in GWD (2006/118/EC) - Annex I and II plus substances with TVs relevant for GWB.

United Kingdom:

- Trend assessment was carried out on all GWBs at risk of failing to achieve the good chemical status objective or which were at poor status, for the purposes of meeting the trend assessment of objective.
- Additional trend assessment was applied as part of groundwater chemical status classification for two tests - the Saline Intrusion Test and the Drinking Water Protected Area Test (DWPA) Test.
- Trend assessment was carried out for all pollutants/indicators that were putting the GWB at risk, identified through the risk assessment/characterisation process.

Hungary:

- All GWBs;
- NH4, NO3, Cl, SO4, EC, Hg, Cd, Pb, DO, pH, atrazine, AOX, pesticides total, simazine, terbutil-azine, terbutrine, trichloro- and tetrachloroethylene.

===

Which data?

Austria:

All data from the national water monitoring network (WFD operational and surveillance monitoring) are considered in trend and trend reversal assessment. Monitoring started in 1991. Length of time series considered depends on monitoring frequency.

Czech Republic:

All data from national monitoring network (boreholes and springs) were used for trend assessment. Although monitored period is longer for part of the stations, data from 2000 – 2012 were used for most of stations due to different quality of analyses or too high limits of quantification.

Denmark:

Based on the same data that have been used for status assessment and that are available from the National borehole database (JUPITER), operated by the National Geological Survey (GEUS). Data from the period 2000-2013 have been used. Data have subdivided into 2 periods (2000-2007 and 2007-2013). The assessment was based on a comparison of the chemical status between these 2 periods.

France:

All data available (WFD operational and surveillance monitoring, Drinking Water monitoring...)

1996-today (1996 = first RBMPs, laboratories started to get better, enough data to assess trends...)

Netherlands:

All data that are used for status assessment are also used for trend assessment. To assess possible trends, a six-year period is compared with the previous six-year period. E.g. when trend assessment in 2006 was done, data from 2000-2005 were compared with data from the period 1994-1999.

Monitoring for status assessment occurs at 10 and 25 m deep. Data for each depth are assessed separately. Focus is on the 10 m data because trends are expected there earlier.

Poland:

All data available from 2005 – 2012 period collected within the WFD Operational and Surveillance Monitoring Programmes and supplemented by data from other monitoring programmes and borehole testing carried on by the Polish Hydrogeological Survey.

Romania:

All monitoring data available from 2000 – 2013 period, provided by WFD Operational and Surveillance Monitoring Programmes carried on by the National Administration "Romanian Waters"

Slovak Republic:

All data obtained from groundwater surveillance and operational monitoring network fulfilling following criteria for each time series:

- Minimum length = 6 years;
- Maximum length = 10 years;
- Start of monitoring in 2004 or later;
- End of time series in 2012 or 2013;
- Maximum gap in time series - 1 year;
- Percentage of non-detects < 50%;
- Percentage of unique aggregated values ≥ 50% (related to the count of all aggregated values evaluated in each time series).

UK:

All monitoring data used for status assessment were considered for trend assessment where relevant. This included data from both Surveillance Monitoring (SM) and Operational Monitoring (OM). The data from SM were also be used to assess natural trends in groundwater.

The SM and OM networks form the national strategic groundwater quality monitoring network. Data for the period 1 January 1997 to 31 December 2006 (10 Years) were used.

Additional data from water companies was also used as supporting evidence. This information (Water Company Article 7 data) recorded qualitative evidence of upward trends at drinking water abstractions.

Hungary:

- Data of the WFD surveillance and operative monitoring network for the time period 2000-2012 were used. In addition, the time period 2006-2012 was examined separately. Monitoring data is aggregated to yearly averages at each monitoring point. Monitoring points in the vicinity of polluted sites were not taken into consideration for GWB trend assessment.
-

==

First determine trend per monitoring point, then per groundwater body? Or lump all data, then determine trend? From monitoring point to GWB level

Austria:

Trend assessment is performed at GWB level based on the aggregated values of all monitoring sites within a GWB. Following minimum criteria have to be met:

- At least 3 sites per GWB;
- At least $^2/_3$ of all sites per GWB have to have sufficient data;
- Only 1 missing value in the time series of each monitoring is acceptable (but not in the beginning and not at the end of the time series); and
- At least 60% of the values are quantified (above LOQ).

Czech Republic:

- Each monitoring point was assessed;
- Monitoring points with significant upward trend (but their mean concentrations are below TVs) selected;
- Aggregation on GWB level – for metals and PAHs one monitoring point exceeded, for nitrates, nitrites, orthophosphates, chlorides and sulphates at least half of monitoring points exceeded.

The aggregation on GWB level is not so important, it is necessary to know specific monitoring station and pollutant.

Denmark:

For each monitoring point and chemical parameter, a mean value was calculated for each of the periods, as an average of the respective annual means.

The state of the 1st and 2nd periods are compared and sorted by strongly increasing (STRI), slightly increasing (SLII), stable (STA), slightly decreasing (SLD) and strongly decreasing (STD), following the guidelines used for the reporting of monitoring results for the Nitrates Directive. All changes are calculated as a % of the relevant threshold value. The trend is assessed for differences of respectively: > 10% (STRI); 10% to 2% (SKLII); 2% to -2% (STA); -2% to -10% (SLD); and <-10% (STD). Assessment results where there is a strongly increasing (STRI) parameter value have been attributed to a "significant upward trend."

For each GWB the percentage of results in each of the 5 above mentioned groups was calculated for each parameter.

The trend for each parameter is designated as "non-adverse" development (1), unknown development (0) and worsened development (-1) according to the following criteria:

- "non-adverse development" (1): The parameter is found in the GWB, but there are no measurements greater than 75% of the TV *or* the parameter values are larger than 75% of the TV, but less than 20% of the monitoring points indicated a strongly increasing trend.

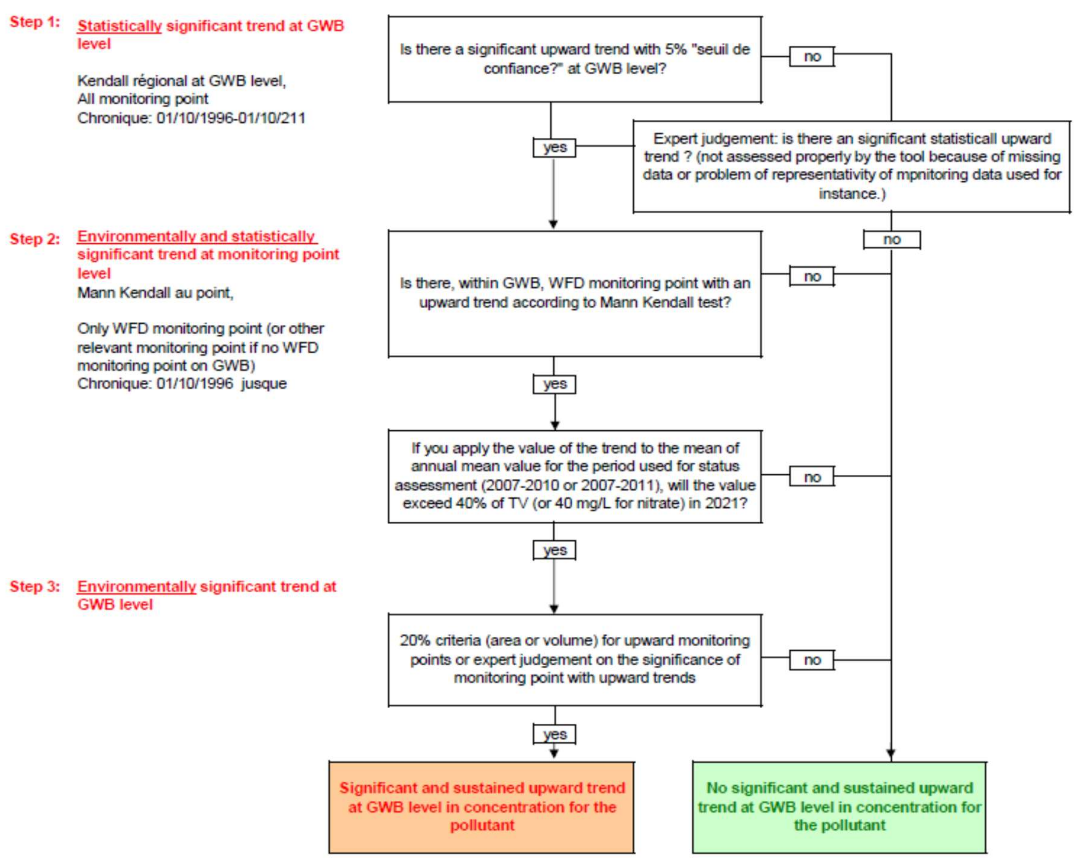
- Unknown development (0): The parameter was not found in the GWB.
- Undesired Development (-1): The parameter was found in concentrations of more than 75% of the TV and more than 20% of the monitoring points at the same time indicate a strong upward trend.

Based on the identification of individual parameter trends an overall trend for the GWB was calculated according to the following criteria:

- Not-undesired development (1): At least one parameter has non-adverse development, and the remaining parameters have unknown development
- Unknown development (0): All parameters have unknown development
- Undesired Development (-1): At least one parameter has undesirable development.

France:

Combination of the 2 methods: First per groundwater body, if there is a trend at GWB level with all data: determine trend per WFD monitoring point and apply it at the monitoring point to determine whether the concentration at the end of the cycle will be above 40% of the TV. If yes, there is a statistical and environmentally significant trend for this monitoring point. If these monitoring points with trends represent more than 20% of GWB, we consider there is a significant trend for the GWB.



Poland:

- Trend assessment was carried out only at individual monitoring points;
- Each monitoring point was assessed;
- No assessment at GWB level has been undertaken yet due to limited number of monitoring points per GWB with an appropriate length of time series to carry out trend assessment.

Romania:

For each monitoring point the annual mean value (arithmetic average) was calculated for each parameter (substance). Data from all monitoring points was spatially aggregated at the GWB scale.

For each groundwater body, depending on its surface, the GIS matrix was determined through the GISView program. Depending on the number of rows and columns of the GIS matrix, every side of which is 250 m., the relative coordinates were assigned to each monitoring point.

The GIS Matrix and the data files (in Excel 5.0/95), with annual average NO_3, NO_2 or NH_4 concentrations in the monitoring points of the GWB were introduced in the computer program, which also estimates the network representativeness - Ru (%).

In case of a homogeneous network (Ru ≥ 80%), the spatial aggregation of data is based on the arithmetic average (AM) and of the confidence limit > 95% (CL (AM)).

In case of a heterogeneously network (Ru < 80%), the spatial aggregation of the data is based on the Kriging average (KM) and on the confidence limit > 95% (CL (KM)).

In this way, the annual mean value per GWB was calculated as mean of all mean values of all monitoring points, for each parameter. Afterwards, the trend for the whole GWB, for each parameter, was determined.

Length of time series considered was 14 years (yearly monitoring frequency).

Slovak Republic:

In the first step, a trend assessment is carried out at the monitoring point level. Annual average values are assessed.

Subsequently, the trend assessment is performed only for those groundwater bodies, where at least one monitoring point showed a significant and sustained upward trend. Trend assessment is performed at GWB level based on the aggregated values (mean of yearly means) of all monitoring sites with the same length of time series.

UK:

a) Trend assessment at individual monitoring points:

Trend assessment was carried out at individual monitoring points to determine (for the 1st river basin management planning cycle):

- Upward trends in pollutant concentrations,

• Trends (upwards or downwards) in concentrations (or values) of natural parameters resulting from natural processes,

Trend assessment was carried out on data from individual monitoring sites for individual parameters and for total pesticides. Where individual parameter concentrations (or values) were reported as below the LOQ, values were replaced by values with half of the reported LOQ, except for total pesticides. For example, a value reported as <0.1 µg/l would become 0.05 µg/l. Trend assessment was not carried out on datasets that comprised more than 80% of values below the LOQ.

b) Trend assessment - Groundwater Body

To determine whether a GWB had a significant and sustained upward trend, the results from the analysis of trends at individual monitoring sites were assessed initially (as in (a)). Where one or more statistically significant trends is identified across a GWB, this trend must be tested for environmental significance, i.e. would lead to failure of one or more chemical status tests within 2 river basin cycles (12 years). Depending on the test(s) for environmental significance being applied (i.e. which status assessment criteria are applied); the number of monitoring points used may vary. For example, for TVs related to the DWPA test, an individual monitoring point (drinking water abstraction) that indicates an environmentally significant upward trend would lead to the GWB being reported as having environmentally significant upward trend. For TVs related to the general chemical status test an aggregation of the results of trend assessment from all relevant monitoring sites is required to test for environmental significance.

If environmentally and statistically significant upward trends are identified for one or more parameters (pollutants or indicators), the GWB was reported as having a significant and sustained upward trend.

Note: the presence of a statistically significant trend at an individual monitoring point will not on its own lead to a requirement to report that the GWB has an upward trend because the trend also needs to be environmentally significant.

Hungary:

Trend assessment is carried out at two levels, at the monitoring site level and at the GWB level.

Trend assessment at monitoring site level:

- Trend assessment, based on the non-parametric Mann-Kendall trend test and Sen's method, is carried out for monitoring data time series of each monitoring site at all GWBs (trends categorised as upward, downward or no trend, or not applicable if not enough data are available).

Trend assessment at GWB level:

- Trend assessment of GWBs (also based on the non-parametric Mann-Kendall trend test and Sen's method) is carried out on aggregated data of all monitoring sites within the GWB for the time periods 2000-2012 and 2006-2012:
    - monitoring data of all the monitoring sites aggregated into one annual average, then trend assessment for the whole GWB;

- 3 levels of reliability:
  - high: min. 5 sites per GWB with time series longer than 6 years;
  - medium: number of sites less than 5, but with time series longer than 6 years; and
  - low: few sites and time series with gaps or inconsistent results of aggregation.

In addition to the trend assessment at GWB level, which is based on the aggregated annual averages of monitoring data of all individual monitoring sites, Sen's trend slope statistics are also determined for GWBs by the aggregation of the Sen's trend slope values of all individual monitoring objects with significant trends. Aggregation of all the individual Sen's trend slope parameters (median, minimum and maximum) is defined for information but is not used for the final assessment of GWB trend.

The statistical trend assessment of GWBs is based on aggregated annual data of all monitoring sites within the GWB. The final evaluation of significant and sustained upward GWB trends is based on those results of the 2000-2012 time period which did not contradict with the significant trends of the 2006-2012 time period. GWBs characterised by significant trends only in the 2006-2012 time period are also considered in the final trend evaluation.

==

Which statistical method used for trend determination? and its confidence

Austria:

Statistically significant trend of pollution is determined by the generalised linear regression test (ANOVA) based on the LOESS smoother and 5% significance level.

Czech Republic:

The statistical methods used depend on the volume of data:

- two-section model was used for pollutants and monitoring points with more than 18 results and at least 10 years;
- linear regression for pollutants and monitoring points with more than 9 results and at least 6 years;
- simplified assessment (comparison of two means) was used for pollutants and monitoring points with at least 6 years – but the results were not identified as significant.

Denmark:

Specified above: a significant upward trend is designated, if a specific parameter is found in concentrations of more than 75% of the TV and if more than 20% of the wells at the same time had a strong upward trend.

France:

Kendall regional at GWB level (95% confidence chosen);

Mann-Kendall at monitoring point (95% confidence chosen);

Slopes have been calculated using the Sen's slope trend test.

An open-source tool (R) has been developed to apply these methods to monitoring data in an easy and automatic way (format used by the tool is issued from the national groundwater database).

Poland:

Trend analysis was carried out using linear regression test.

Romania:

Statistically significant trend of pollution is determined by the generalised linear regression test (ANOVA) based on the LOESS smoother (95% Confidence level).

Slovak Republic:

For each time series goodness-of-fit tests for normal distribution (Lilliefors test, Shapiro-Wilk test) are calculated. The following methods are used for the identification of a statistically significant upward trend (5% significance level is used for all tests):

- Mann-Kendall test in each time series;
- Generalised linear regression test (ANOVA) - if data set follows normal distribution.

UK:

In many groundwater systems there will be considerable seasonal variability in parameter concentrations. This variability may introduce problems in the trend analysis unless it can be corrected. Where there are sufficient data within a given year, the non-parametric Seasonal Kendall trend test was used. Where there were variable or insufficient data seasonality was removed by calculating the annual means, and then trend analysis performed on the annual means. In this case Sen's Method was used (Sen, 1968). Both the Seasonal Kendall test and Sen's Method are robust methods that allow for some missing data in the time series and are not badly affected by gross errors or outliers in the data series. Additionally linear regression analysis was carried as a further test for trends. For a trend to be statistically significant there should be at least 80% confidence. For England and Wales 90% confidence was used.

Hungary:

Trend determination is based on the non-parametric Mann-Kendall trend test. Upward or downward trends are considered significant at either of the two confidence levels, 90% and 95%. The magnitude of the trend (trend slope) is obtained by Sen's method. Upper and lower confidence limits (95%) are also reported for the Sen's slope.

==

How environmental significance was defined for trend?

Austria:

A trend is environmentally significant when the trend line exceeds the starting point for trend reversal.

Czech Republic:

A predicted concentration was calculated as a result – for 2015 year and 2017 year. Monitoring points that exceeded TVs in 2015 were considered as having significant upward trends, monitoring points below TVs in 2015, but exceeded in 2017 were identified as having potentially significant trends.

Denmark:

Monitoring points with strongly increasing parameter values (> 10%) have been attributed to a "significant upward trend".

France:

If 40% of the TV ("risk" or value for trend reversal value) is reached at the end of the WFD cycle.

Poland:

An upward trend was classified as environmentally significant if a predicted concentration of a pollutant, calculated using a linear regression, exceeded the starting point for trend reversal by 2021.

Romania:

No environmental significance was defined for trends (the trends were analysed for all GWBs which were continuously monitored in the 2000 – 2013 period).

Slovak Republic

A significant and sustained upward trend occurs if a statistically significant upward trend was identified and:

- the mean of 2 last yearly mean concentrations ≥ 0.75 of the limit value (GW quality standard, TV) and/or
- the predicted concentration for 2021 ≥ limit value.

UK:

The test of environmental significance was determined by examining whether the trend(s) at the monitoring point would lead to the body failing one or more of its environmental objectives within the period covered by two river basin cycles (12 years). This was assessed by comparing predicted concentrations at the end of the two river basin cycles with the TVs and status assessment criteria relevant to the monitoring point.

Hungary:

A trend is environmentally significant when the trend line exceeds the starting point for trend reversal within the examined time period of 2000-2012. A trend is also considered environmentally significant when the projected trend line exceeds the TV in the next two river basin cycles, that is, until 2027.

In general, 75 % of the respective TV is considered as the starting point for trend reversal. In cases of protected GWBs (deep porous intergranular and karst GWBs), a stricter starting point, 30% of the respective TV, is applied according to the Hungarian legislation, but this was not examined for trend reversal. In these (30% of TV) cases particularly, definition of the starting point as a percentage of the TV is considered problematic, as natural background and reversal point values are often very close to each other (within error limits of measurements), and even background and TVs have similar values in some of the GWBs for certain parameters.

==

Number of years of monitoring results required to determine trends

Austria:

The length of time series depends on the aggregation period and for annual aggregations is 8 years and for half-annual and quarterly, 6 years.

Czech Republic:

Monitoring points with more than 9 results and at least 6 years could have significant trends; monitoring points with at least 4 years data were also assessed, but the results are not yet significant.

Denmark:

The minimum criterion has not been related to the number of years. Rather, it is required, that chemical analyses have been available for each of the two periods 2000-2006 and 2007-2013.

France:

At least 10 data points are required to run the Mann-Kendall statistical trend test as it is implemented in the dedicated tool. So there is not a minimum number of years of monitoring results required but a minimum number of monitoring data available after the year 1996.

Netherlands:

There is no minimum defined. Instead, the number of years for which data are available determines the relevance that should be ascribed to a trend. Example: if data from only 2 years are available, an upward trend should trigger continuation of the monitoring of that particular substance. A data series containing 8 to 15 years of data usually allows statements with sufficient statistical power about the presence or absence of upward trends.

Poland:

Methodology for trends and trend reversal assessment used was based on the WFD CIS Technical Report Nº 1. According to this report's recommendations, the length of time series required for the assessment depends on data availability, specifically the frequency of monitoring. If data are aggregated as annual means, the required period of observations is 8 years. If data are collected with higher frequencies (half-annual or quarterly), 6 years of data are required.

In Poland data are aggregated to annual means, therefore the trend analysis was carried out for 8-year long time series.

Romania:

The methodology used for trends and trend reversal assessment was based on WFD CIS Technical Report Nº 1 (as noted for Poland above).

Trend analysis was carried on for more 8 years, in the case of annual aggregated data.

Slovak Republic

Trend analyses are primarily based on monitoring results from the period 2004 - 2013. If data are not available for the whole period, a shorter period can be accepted if its minimum length is at least 6 years and the last measurement was carried out in 2012 or 2013.

UK:

For the assessment of anthropogenically induced upward trends in pollutant concentrations, monitoring data for a period of between 6 and 10 years leading up to the date at which assessment is being carried out, were used for identifying trends.

When considering natural trends, data were considered for as long a time period as possible. As a minimum this was at least 6 years with at least one measurement per year. For England and Wales, trend analysis was only carried out for monitoring points where there were a minimum of 10 analyses (results) over a period of at least four years.

Hungary:

At least 5 annual data per monitoring point is required. The applied non-parametric statistical methods allow for missing data. Time series with more than 80% of values below the LOQ are excluded from the trend assessment.

===

Methods used in case of missing data or too few data

Austria:

If more than 1 value is missing in the time series of a monitoring point (but not at the beginning and not at the end of the time series), the monitoring point is excluded from further analysis.

Czech Republic:

See question “statistical method used for trend determination”.

Denmark:

Trend calculations have been performed only for GWB, for which at least 5 monitoring points have been available for each of the two periods 2000-2006 and 2007-2013.

France:

Trend is not assessed at monitoring point level in case of not enough available data but data at these monitoring points are used to estimate trends at the GWB level via the Kendall-regional trend test.

Netherlands:

In case of missing data, The REML-method is used (Residual Maximum Likelihood; Welham et al., 2004).

Poland:

Two missing values were allowed within time series. Where there was not enough available data, the trend was not estimated.

Romania:

Trend is not assessed where there is not enough data available.

Slovak Republic:

If a gap in the monitoring data for the assessed parameter is more than 1 year, the monitoring point is excluded from the trend assessment.

UK:

At least one measurement per year was required for trend assessment to be carried out. Where data were inadequate, then no trend assessment was carried out and an explanation was recorded.

Hungary:

A certain amount of missing data is allowed by the non-parametric statistical trend test. However, when less than 5 annual data are available at a monitoring point the trend status assessment is considered ‘not applicable’. Trend assessment is not carried out on datasets with more than 80% of values below the LOQ.

==

Monitoring density (spatial) for trends at GWB level?

Austria:

Trend assessment is performed for the GWB as a whole. Minimum requirements are:

- at least 3 sites per GWB;
- At least $^2/_3$ of all sites per GWB have to have sufficient data;

Czech Republic:

Monitoring density was considered for chemical status reliability, no specific assessment was prepared for trend assessment.

Denmark:

Trend calculations have been performed only for GWB, for which at least 5 monitoring points have been available for each of the two periods 2000-2006 and 2007-2013.

France:

Regional trends at GWB level have been calculated when at least 3 monitoring points with data after 1996 are available.

Netherlands:

Theoretically, three aspects are distinguished:

1) Frequency: depends on groundwater flow rate. Interval between measurements should be longer than replacement rate (otherwise you measure the 'same' water).
2) Depth: should depend on local conditions and should be derived from the conceptual model.
3) Number of monitoring points per GWB: should depend on the heterogeneity of the GWB, e.g. in terms of soil-land use combination. Weighting can be applied to correct for non-proportionalities.

Poland:

Monitoring density was considered for chemical status reliability, no specific assessment was prepared for the trend assessment. The analysis undertaken so far has only been at individual monitoring points; no assessments at GWB level have been undertaken yet.

Romania:

Trends at GWB level have been calculated where data series for a minimum of 3 monitoring points are available. The number of monitoring points per GWB depends on heterogeneity and areal size of the GWB.

Slovak Republic:

Trend analyses at GWB level are performed, if time series with the same length are available for at least 3 monitoring points.

UK:

Trends at the GWB level have been assessed using all data that are part of the Operational and Surveillance monitoring networks. It is considered that the number of sites and their distribution will enable a representative assessment of the status of groundwater and hence trends. Therefore for small GWBs this may be a single monitoring point whereas in larger GWB a larger number of points would be used.

Hungary:

Requirements for trend assessment are at least 3 sites per GWB. In cases, where an individual time series is considered representative of the GWB, less than 3 sites are also accepted.

==

Results below Limit of Quantification

Austria:

All values <LOQ are replaced by half of LOQmax. LOQmax is identified over the length of the analyzed time series (6 or 8 years for trends and 10 or 14 years for trend reversal – depending on annual, half-yearly or quarterly aggregation).

Czech Republic:

All values <LOQ in specific monitoring point were replaced by half of LOQmin.

Denmark:

The level (limit) of detection (LOD) was used. As no regression analysis was performed, no risk of false trends from low values was expected, and all LOD << TVs.

France:

As the Kendall Regional trend test is a non-parametric statistical test, results reported below LOQ are taken into account for estimating trend. For calculating slope, the Sen Slope test has been used substituting results reported below LOQ by the LOQ/2 value.

Netherlands:

Ideally, raw data should be used, including below LOQ (even negative values). These are often not reported by the analysing labs. In that case, the best is to: 1) replace by LOQ; 2) replace by zero; and 3) to compare 1 and 2 to investigate the sensitivity of the replacement method chosen.

Poland:

Results below LOQ were replaced by half of LOQ.

Romania:

Results below LOQ were replaced by half of LOQmax. LOQmax is defined as the largest LOQ which does not exceed twice the median of all LOQs.

Slovak Republic:

LOQmax is identified individually for each time series. Measurements below LOQ and measurements below LOQmax are replaced by LOQmax.

UK:

Trend assessment was carried out on data from individual monitoring sites for individual parameters and for total pesticides. Where individual parameter concentrations (or values) were reported as below the LOQ, values were replaced by values with half of the reported LOQ, except for total pesticides. For example, a value reported as <0.1 µg/l would become 0.05 μg/l. Trend assessment was not carried out on datasets that comprised more than 80% of values below the LOQ.

Hungary:

Results below LOQ are replaced by half of LOQ except for:

- Pb, Cd and Hg where these data are replaced by 0.005 μg/l;
- organic pollutants and pesticides data by 0.

Trend assessment is not carried out on a dataset that comprises of more than 80% of values below the LOQ.

==

Can you use expert judgement for trend determination?

Austria:

No

Czech Republic:

Not yet.

France:

Yes

Netherlands:

Yes. A protocol has been set for trend assessment and trend reversal based on expert judgement.

Poland:

No. Expert judgement is helpful, but reliable results cannot be obtained without statistical analysis.

Romania:

No. Expert judgement is helpful, but without statistical methods we cannot obtain reliable results.

Slovak Republic:

Yes

UK:

No

Hungary:

No.

===

Consideration of baseline level of concentration?

Austria:

Not documented

Czech Republic:

Baseline level of concentration was considered by use of time series 2000 – 2012.

Denmark:

No.

France:

No.

Netherlands:

Not documented

Poland:

No. Statistical analysis of natural background concentrations was carried out for the purposes of establishing TVs.

Romania:

The annual mean value of the concentrations for the year 2000 aggregated on GWBs were considered as the baseline level for each parameter in the determination of trends process.

Natural background level was considered within the TV determination process.

Slovak Republic:

No

UK:

Trend analysis of natural background concentrations was also carried out for the purposes of characterisation but not as part of formal trend assessment/status assessment. Only trends which were statistically significant and the data quality (time series) was adequate were recorded.

Hungary:

No. Statistical analysis of natural background concentrations are carried out to establish TVs and trend reversal points, which are used in the assessment of significant and sustained upward trends.

==

Methodology for trend reversal assessment?

Austria:

Trend reversal is assessed by an extended linear regression model where the time series is divided into 2 sections (legs) and the inclinations are compared (as noted in CIS Technical Report No 1). The choice of time sections is optimised. The significance level is 5%.

The left leg needs a length of at least 4 years (without the break) and a positive slope and the right leg needs a minimum length of 2 years (without the break) with a negative slope for half-yearly and quarterly aggregations and 3 years for annual values.

Length of time series: 14 annual values or values from 10 years for half-yearly and quarterly aggregations.

Further criteria are identical with those for trend assessment:

- at least 3 sites per GWB;
- at least $^{2}/_{3}$ of all sites per GWB have to have sufficient data;
- only 1 missing value in the time series of each monitoring is acceptable (but not at the beginning or end of the time series); and
- at least 60% of the values are quantified (above LOQ).

Czech Republic:

The methodology for trend reversal was established (by two-section model or linear regression – the same approach as for upward trend assessment). However, the trend reversal was not assessed, because the last monitoring results were from 2012, when measures should be implemented, and their effect could not be observed.

Denmark:

Not made.

France:

Not done for the 2nd cycle exercise. The tool is proposing to estimate the date of a trend reversal by Darken method. Also estimating the significance of variation between 2 periods is done by Pettitt and Buishand tests.

Netherlands:

A two-section model applying linear regression should be applied. Not done yet for first and second RBMP cycles.

Poland:

Not done yet.

Romania:

Trend reversal is assessed based on the assumption that time series can be characterized by 2 linear trends with a slope direction change within the analysed time period.

Trend reversal can be detected if in the first section the slope is ascending (positive) and in the second section the slope is descending (negative), applying 95% quantile of the distribution.

Method steps are: optimization of time sections selection; analysis of slope change significance for the model of simple linear regression; statistical test to verify that the model of the 2 sections is more significant than the model of simple linear regression.

Slovak Republic:

Not made yet. A two-section linear regression model will be applied.

UK:

Trend reversal was not assessed in the 1st RBMP cycle as it was considered that upward trends had to be determined and reported first. Trend reversal, if achieved, would be reported in the 2nd RBMP cycle. The methodology that will be followed is:

- By default the starting point for trend reversal will be when pollutant concentrations reach 75% of any relevant TV. Alternative earlier or later starting points can be set if it can be demonstrated they are needed to avoid a future failure of WFD objectives or can be justified in terms of cost-effectiveness and technical feasibility.
- Reasons for selecting starting points have to be reported in the RBMP. Once a starting point has been reported in the RBMP it cannot be changed until the next river basin cycle.

The results of trend assessment should be taken into account when determining the starting point for trend reversal. Where a trend indicates that there is likely to be a failure of one or more environmental objectives, before the end of the river basin cycle then this failure should be used taken into account when defining a starting point. A starting point must be chosen to allow measures to be put in place to reverse trends in sufficient time to avoid failure of WFD environmental objectives.

Hungary:

Trend reversal has not been carried out by a statistical methodology, as monitoring time series were too short. Only the possible indication of reversal is considered, such as identification of a downward trend with environmental significance.

Appendix 3: Example of application to a groundwater body - Italy

The Italian methodology for trend and trend reversal assessment is described in ISPRA - IRSA.CNR (2017). It was applied to the « Conoide Secchia » groundwater body, which is a monolayer phreatic aquifer in an Apennine alluvial fan. Surface : 85.7 km^2; monitoring sites: 11 ; monitoring density: 7.8 km^2 (Figure A1).

In this groundwater body the nitrate threshold of 50 mg/L has been exceeded at 5/11 monitoring sites in 2015 (Table A1). The trend assessment reveals that 4 of the 5 have an upward trend of the nitrate concentration which is statistically significant at 90% for the Mann Kendall test (Figure A2). If we assume that each monitoring site represent 1/11 of the groundwater body, these 4 sites represent 4/11*100=36% of the groundwater body. Hence, a significant and sustained upward trend for nitrate can be deduced for the whole groundwater body and the evaluation of the scenario in 2021 and 2027 is requested.

The scenario in 2021 confirms that this trend is environmentally significant at the groundwater body scale, because at least 4/11 monitoring sites representing more than 20% of the groundwater body, will continue to exceed the threshold, unless adequate measures are implemented to achieve trend reversal.

Thence, the existence of a reversal of the trend should be evaluated and eventually statistically demonstrated. The Pettitt test (Pettitt, 1979) is applied to identify the changing point in the dataset and the level of confidence associated. None of the sites affected by upward trends shows a statistically significant trend reversal.

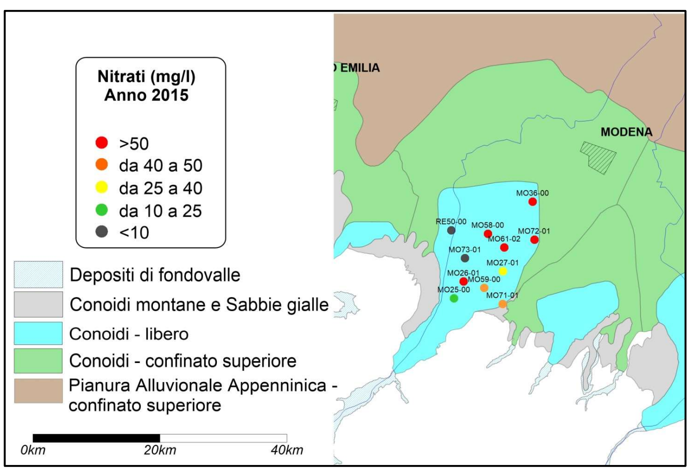


Figure A1 Case study location and monitoring sites with NO_3 concentration in 2015

Soglia di significatività:	**90**	%
Valore Tabellato:	**0.071**	(Hollander et al.)
Rigetto ipotesi Nulla:		
Conf. 2 tails	**85.8**	*trend non significativo*
Conf. 1 tail	**92.9**	*trend significativo*

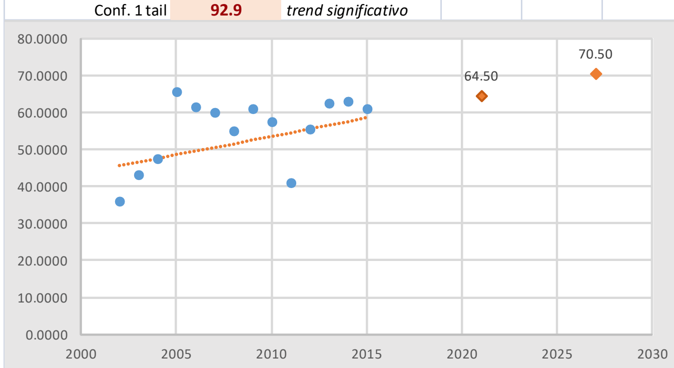


Figure A2 – Nitrate trend assessment between 2001 and 2015 for MO26-01 monitoring site (blue dots and orange dotted line). Sen's slope=1.00 mg/L/y. The orange diamonds show the theoretical scenario for 2021 and 2027.

Table A1 – Average nitrate concentration on 2015, trend, slope and scenario in 2021

Site	**Average nitrate 2015 (mg/l)**	**Trend**	**Slope (mg/l/y)**	**Scenario in 2021 (mg/l)**	**Trend reversal**
MO25-00	20	No			
MO26-01	61	Yes	1.00	64.5	No
MO27-01	36	No			
MO36-00	60.5	Yes	1.9286	70.21	No
MO58-00	53	Yes	1.20	77.6	No
MO59-00	42	No			
MO61-02	52.5	No			
MO71-01	42	No			
MO72-01	56.5	Yes	0.4825	61.36	No
MO73-01	8	No			
RE50-00	5.5	No			

Appendix 4: Trend case studies in Wallonia

Geer basin – Nitrate

Context and objective	Nitrate trends assessment – EUFP6 Aquaterra project
Scale	Geer basin aquifer
Studied parameters	Nitrate
Methods used	• Test of normality (Shapiro-Wilks and Shapiro-Francia test) • Linear regression for dataset normally distributed • Non parametric Mann-Kendall test for detection and Sen slope for quantification on all dataset (normally and non-normally distributed). • Point by point analysis
Main results	• Upward trend in nitrate concentrations measured in most of the points • Spatial variation of the slope in function of the hydrodynamic conditions prevailing in the aquifer
Main lessons	• Statistical trend analysis seems to be robust • Man-Kendall test and Sen slope analysis can be used for both normally and non-normally distributed datasets • Time series have to be long enough to cover natural fluctuations
References	**Batlle-Aguilar, J., Orban, Ph., Dassargues, A., Brouyère, S. (2007)**, Identification of Groundwater quality trends in a chalk aquifer threatened by intensive agriculture in Belgium, Hydrogeology Journal, 15(8), 1615-1627. **Visser, A., Dubus, I., Broers, H. P., Brouyère, S., Korez, M., Orban, P., Goderniaux, P., Batlle-Aguilar, J., Surdyk, N., Amraoui, N., Job, H., Pinault, J.-L., & Bierkens, M. (2009)**. Comparison of methods for the detection and extrapolation of trends in groundwater quality. Journal of Environmental Monitoring, 11, 2030-2043.

Geer basin – Pesticides

Context and objective	Pesticide trends assessment – EUFP7
Scale	Geer basin aquifer
Studied parameters	4 pesticides and two metabolites
Methods used	• Management of non-detect values using the Regression on Order Statistics method • Non parametric Mann-Kendall test for detection and Sen slope for quantification. • Analysis made on data aggregated at the scale of the unconfined part of the aquifer
Main results	• Contrasted results in function the pesticides • Link can be made with regulation (pesticides banned or regulated)
Main lessons	• Methodology developed taking into account non-detect values.
References	**Hakoun, V., Orban, Ph., Dassargues, A., Brouyère, S., (2017)**, Factors controlling spatial and temporal patterns of multiple pesticide compounds in groundwater (Hesbaye chalk aquifer, Belgium), Environmental Pollution, 223, 185-199.

Nitrate trends at the scale of the Walloon region

Context and objective	Nitrate trends assessment – SPW/DGO3 project
Scale	Walloon Region
Studied parameters	Nitrate
Methods used	• A complete and robust workflow is defined: o Identification and discarding of the temporally isolated data o Test on the number of available data o Outliers detection with LOESS o Parametric (least-square) model fitting for several models (see below), using water table depth as covariable, if available o Normality test on the residuals o If the residuals are non-Gaussian, non-parametric regression and tests o Aggregation at the basin level • Model selection between linear regression and regression models with breakpoints estimation (imposing continuity at the breakpoints) • When normality of the residuals cannot be assumed, non parametric Mann-Kendall test for detection and Sen (Sen, 1968) slope for quantification (also called Kendall-Theil). • Point by point analysis • 986 Time series for shallow and underground water, issued from measurements and from modelling (EPICgrid model, Sohier, 2011) • Classification of the points according to the trend behaviour
Main results	• Downward trend in nitrate concentrations observed in 2/3 of the points • Breakpoints identified in 2/3 of the points • Most basins show a favourable evolution
Main lessons	• Definition of a complete and robust workflow combining trend estimation and hypothesis testing • Efficient screening and classification of the time series according to their behaviour
References	**D'Or D., Allard. D., (2014),** Mise en évidence de tendances éventuelles sur les séries chronologiques présentées par les points du réseau de mesure wallon des eaux de surface et souterraines en ce qui concerne les nitrates. Rapport final. Rapport Ephesia Consult RP DGO3 2014002 - Novembre 2014.

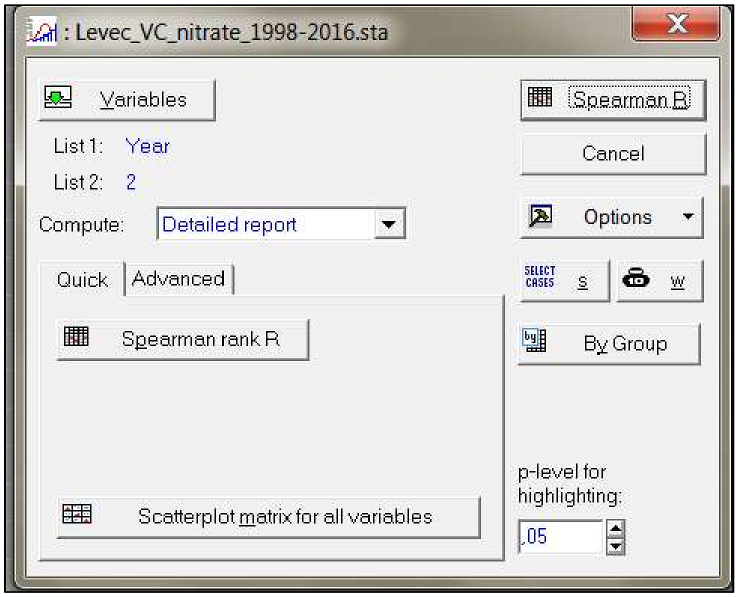
Appendix 5: Trend assessment with STATISTICA software tool in Slovenia

1. **The purpose of the case study:** We would like to find out whether, between the variable of time, for data set of 19 years and concentrations of nitrate (mg NO_3/l), measured at monitoring point Levec VČ 1772 from 1998 to 2016, there is statistically significant correlation.

Correlation between the variable of time for data set od 19 years and concentrations of nitrate (mg/l), measured at measuring point Levec VČ 1772 from 1998 to 2016

	1 Year	2 Levec VČ 1772 nitrate (mg/l) 1998-2016
1	1998	76,15
2	1999	89,25
3	2000	68,30
4	2001	69,40
5	2002	54,45
6	2003	49,80
7	2004	69,23
8	2005	66,27
9	2006	58,75
10	2007	49,00
11	2008	62,00
12	2009	58,00
13	2010	53,50
14	2011	55,50
15	2012	46,50
16	2013	57,50
17	2014	51,00
18	2015	38,50
19	2016	43,00

2. **Specifying the analysis:** *Nonparametric* module to calculate matrice of correlations, with two lists of variables is selected, default selection of *Detalied report is accepted.*



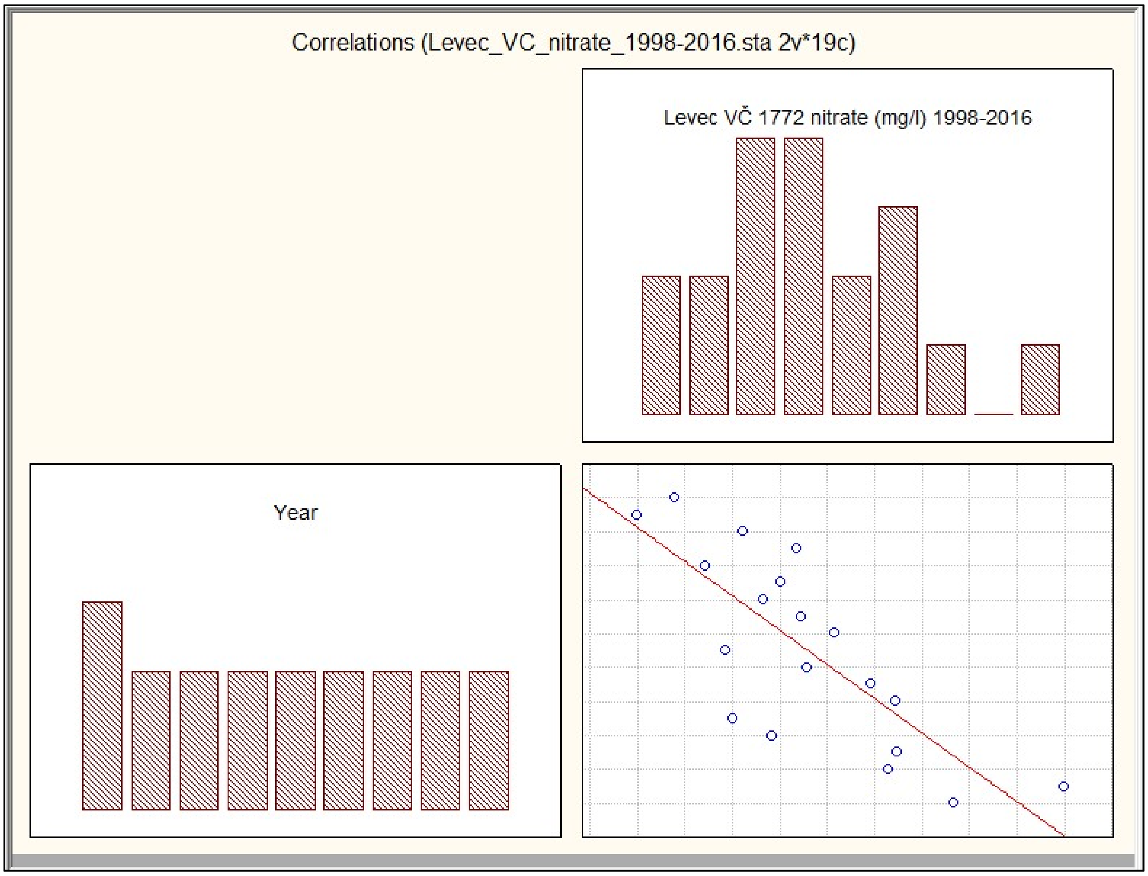
3. **Reviewing the results:** *Spearman rank R* button is selected - to display a spreadsheet with the results of the analysis.

Pair of Variables	Spearman Rank Order Correlations (Levec_VC_nitrate_1998-2016.sta) MD pairwise deleted Marked correlations are significant at p <,05000 Valid N	Spearman R	t(N-2)	p-level
Year & Levec VČ 1772 nitrate (mg/l) 1998-2016	19	-0,761404	-4,84259	0,000153

The statistical analysis shows that:

- correlation between two variables is statistically significant (red highlited result with p level<0,05)
- Spearman's correlation coefficient (r') is negative (- 0,76) which means that nitrate concentrations decrease with time.

4. **Vizualization of correlation:** The correlation is visualized by displaying the *Scatterplot matrix* for variables button. In the illustration below, the correlation plot and histograms of the variables are included.


Correlations (Levec_VC_nitrate_1998-2016.sta 2v*19c)
Levec VČ 1772 nitrate (mg/l) 1998-2016
Year