Final Report

"The EU Water Framework Directive:
Statistical aspects of the identification of groundwater pollution trends, and aggregation of monitoring results"

WFD-GW

www.wfdgw.net

December 2001

Financed

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

Project team at the Federal Environment Agency ltd. - Austria

Project leader/Co-ordination group

J. Grath, A. Scheidleder, K. Weber, F. Ott, Ch. Gmeiner

Authors

J. Grath, A. Scheidleder, S. Uhlig, K. Weber, M. Kralik, T. Keimel, D. Gruber

GIS/Web-GIS

D. Gruber, P. Aubrecht, K. Placer, M. Hadrbolec, G.Vincze

IT/Web-Support

W. Nagy, H. Kaisersberger, E. Knappitsch

Hydrogeology

M. Kralik, T. Keimel

Subcontractor/Statistics - quo data ltd. - Germany

S. Uhlig, N. Schick

Projectpartners/Contact persons

Johan Lermytte	AMINAL - Environment, Nature, Land and Water Administration, Water Division- Belgium
Jens Stockmarr	GEUS - Geological Survey of Denmark and Greenland - Denmark
Laurent Pavard	A.E.A.P. - Agence de l´eau Artois-Picardie - France
Georg Berthold	Hessisches Ministerium für Umwelt, Landwirtschaft und Forsten - Germany
Georgia Gioni, Costas Papadopoulos	IGME - Institute of Geology and Mineral Exploration - Greece
Paul Coleman	EPA - Environmental Protection Agency - Ireland
Leo Boumans	RIVM - National Institute of Public Health and the Environment – The Netherlands
Ana Rita Lopes, Rui Rodrigues	INAG - Instituto da Água - Portugal
Manuel Varela	Ministerio de Medio Ambiente - Spain
Rob Ward	Environment Agency of England and Wales – United Kingdom

Disclaimer

Examples and results presented in this report were elaborated on the basis of test data sets provided by project partners to develop proposals for algorithms on the basis of real data.

They serve for demonstration purposes only and do not anticipate assessments by the Member States

Material in this report reflects the discussion and conclusions of the working group and does not necessarily reflect the position of the EU and its Member States.

Copyright Notice

Reproduction is authorised, provided the source is acknowledged

Citation:

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. 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.

December, 2001

Content

1 EXECUTIVE SUMMARY

BACKGROUND

The "Directive of the European Parliament and of the Council 2000/60/EC establishing a framework for Community action in the field of water policy", the so-called EU Water Framework Directive (WFD), defines in Article 4 "Environmental objectives" for surface water, groundwater and protected areas.

In order to achieve the "Environmental Objectives" for groundwater (Article 4(1)(b)), the WFD requires that specific measures shall be adopted to prevent and control pollution of groundwater. Such measures shall be aimed at achieving the objective of good groundwater chemical status. The criteria for achieving good groundwater chemical status are defined in Annex V 2.3.2 and in particular require that concentrations of pollutants (in groundwater) do not exceed any quality standards applicable under other relevant Community legislation. Further there is a requirement to identify and reverse any significant and sustained upward trends in the concentration of pollutants. The Directive provides specifications on the identification of trends in pollutant concentrations originating from diffuse and/or point sources.

This project focused on the development of particular algorithms for the identification of trends in pollutants (Annex V 2.4.4) and a data aggregation method for interpretation and presentation of groundwater chemical status as defined in Annex V 2.4.5.

PROJECT TEAM

A consortium of partners from 11 EU Member States [Austria, Belgium, Denmark, France, Germany, Greece, Ireland, the Netherlands, Portugal, Spain and United Kingdom (England and Wales)] was formed under the leadership of the Federal Environment Agency ltd. – Austria (FEA). Thus it was ensured that the results of the project were based on data derived from a broad variety of different groundwater bodies within the EU.

Institutions from 5 countries (Finland, Hungary, Italy, Norway and Sweden) participated as observers in the project and in this capacity attended meetings and commented on the draft final report. Furthermore comments from the ECPA as stakeholder were received.

The project was commissioned and financed at approximately 1/3 by DG Environment of the European Commission and 2/3 by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management. In kind contributions from the partners constituted another important input.

Sub-contractor in the project and responsible for the development of the statistical concepts was the German company "quo data".

OBJECTIVES

The main goal of the project was to establish methods for the calculation of representative mean concentrations, for data aggregation and trend (reversal) assessment at the groundwater body level respectively for groups of groundwater bodies. The methods had to be suitable for Europe-wide application and implementation based on the provisions of the Water Framework Directive taking into account influences originating from diffuse and/or point sources.

The following main aspects were considered:

- Development of an appropriate data aggregation method for the assessment of groundwater quality at the groundwater body level respectively for groups of groundwater bodies including the determination of minimum requirements for calculation.
- Development of an appropriate statistical method for trend assessment and trend reversal including the determination of the minimum requirements for calculation.
- Concentrations below the limit of quantification.

TEST DATA SETS FOR DEVELOPMENT OF PROCEDURES

As the developed procedure shall be applicable to all types of GW-bodies (different hydrogeological characteristics, size, number of sampling sites, pressures etc) the test and discussion of the proposed procedure on the basis of test data sets was regarded to be of vital importance. This information was provided by the partners in the project. In total information (GW-body description and GW-quality data) on 21 GW-bodies in 9 countries was available. Apart from the proposed procedures the description and characterisation of the GW-bodies was an essential part of the project report. Quality data were available for nitrate, ammonium, electric conductivity, chloride, pH-value, pesticides, chlorinated hydrocarbons, metals etc.

MONITORING NETWORK

The working group agreed that the monitoring network should fulfil some minimum requirements. It was agreed that homogeneity (reflecting spatial representativity) of the network was a prerequisite and should be ensured to allow for sound statistical assessment in accordance with the requirements of the WFD. For assessing whether the distribution of sampling sites within a monitoring network is homogeneous or not, a representativity index was developed. If the GW-body is hydrogeologically heterogeneous and if a spatially homogeneous monitoring network is not feasible or sensible the monitoring network has to be developed to be hydrogeologically representative.

REQUIREMENTS ON QUALITY DATA, TREATMENT OF "LOWER THAN" VALUES

The sampling procedure itself and chemical analysis should ensure continuity in results. Relevant standards are to be applied. For several groups of substances provisions for the limit of quantification and the limit of detection are of vital importance for a sound data basis for the assessment. The discussion showed that there is an urgent need to provide sufficient information on both the limit of quantification (LOQ) and the limit of detection (LOD). This should be considered when defining monitoring requirements and analytical procedures. From the statistical aspect it is not recommended to perform the proposed aggregation and trend assessment if the LOQ exceeds 60 % of the limit value (if a limit value is available).

STATISTICS

Requirements on statistics

The working group agreed on the following general requirements on the statistical procedures

- statistical correctness,
- development of a pragmatic way,
- one data aggregation method suitable for small, large and groups of GW-bodies as well as for small GW-bodies with few sampling sites and
- applicability for all types of parameters.

Data aggregation

Provisions for data aggregation in the WFD (Annex V Section 2.4.5) are as follows:

> *In assessing status, the results of individual monitoring points within a groundwater body shall be aggregated for the body as a whole. Without prejudice to the Directives concerned, for good status to be achieved for a groundwater body, for those chemical parameters for which environmental quality standards have been set in Community legislation:*
>
> – *the mean value of the results of monitoring at each point in the groundwater body or group of bodies shall be calculated; and*
>
> – *in accordance with Article 17 these mean values shall be used to demonstrate compliance with good groundwater chemical status.*

For the calculation of a spatial mean a pragmatic way was proposed by the working group. In principle the selected aggregation method is the arithmetic mean (AM) and its 95 % upper confidence limit (CL_{AM}). Since under certain conditions (depending on the monitoring network, the GW-body characteristics etc.) the calculation of the AM is not applicable from a statistical point of view, the calculation of a weighted arithmetic mean and its CL 95 considering different GW-sub-bodies might be necessary. In this case the spatial mean is calculated as a weighted arithmetic mean (wAM) and its CL_{wAM}. In case of an exceeding of the limit value by the CL95 of the (w)AM it is regarded as permissible to verify the result by calculating an arithmetic mean weighted with regard to the area represented by the particular sampling site [Kriging mean (KM)] and its CL_{KM} for the estimation of the spatial mean.

The working group proposes the following procedure:

- Check whether the GW-body consists of several sub bodies with different sampling site densities.
  - If no, examine the monitoring network with regard to the network criterion (representativity index),
  - if yes, examine the monitoring networks within sub-bodies with regard to the network criterion.
- If the network criterion for the monitoring network(s) is not fulfilled, the monitoring network has to be adapted accordingly or the GW-body has to be subdivided into sub-bodies which fulfil the network criterion.
- If the GW-body or the sub-body is hydrogeologically heterogeneous and if a spatially homogenous network is not feasible or sensible, a hydrogeologically representative monitoring network has to be developed, and the spatial mean should be estimated with identical weights (AM).
- Use AM or the weighted AM (in case of several sub-bodies) to estimate the spatial mean (pragmatic approach).
- If the action limit is exceeded by CL_{AM}, CL_{KM} may be applied alternatively (which can be considerably smaller in case of spatial correlation and high variability of the concentration level).

In order to guarantee the required level of confidence for GW-bodies with only a few stations the agreed proposal is to use the upper confidence limit of arithmetic mean (respectively Kriging mean) instead of the mean values itself.

The upper confidence limit (CL) depends on the variability of the concentration level within the GW-body and on the number of stations. CL decreases with an increasing number of stations within the GW-body. The use of the CL allows to reduce the number of stations in GW-bodies with levels far below the limit value, and enforces a higher number of stations in GW-bodies with levels close to the limit value. To some extent it is therefore in the hands of the monitoring manager whether the CL will be below or above the limit value thus allowing an effective allocation of analytical resources.

From the statistical aspect the minimum number of sampling sites is 3 within a GW-body and 1 station per sub-body. For the treatment of "lower than LOQ" measurements a minimax approach (minimize maximum risk) was applied.

Trend and trend reversal assessment

Provisions for trend (reversal) assessment in the WFD (Annex V Section 2.4.4) are as follows:

> *Member States shall use data from both surveillance and operational monitoring in the identification of long term anthropogenically induced upward trends in pollutant concentrations and the reversal of such trends. The base year or period from which trend identification is to be calculated shall be identified. The calculation of trends shall be undertaken for a body or, where appropriate, group of bodies of groundwater. Reversal of a trend shall be demonstrated statistically and the level of confidence associated with the identification stated.*

The working group defined the following criteria for the selection of methods:

- applicability for all types of parameters,
- extensibility to potential adjustment factors,
- sufficient power for the detection of a trend/reversal,
- robustness was considered less important than power and extensibility (data validation will be responsibility of MS).

Trend analysis should be based on aggregated data from the whole GW-body (WFD, Annex V). Data aggregation for trend assessment consists of the same procedures (regularisation and spatial aggregation) as for quality status assessment.

With regard to extensibility and power the linear methods (based on a linear model) outperform non-parametric methods based on the test of Mann-Kendall, and therefore the decision was in favour of the linear methods. The consequence was a decision for the generalised linear regression test (ANOVA test) for the assessment of monotonic trends. For the assessment of a trend reversal, the consequence was a decision for the two sections model, due to its simple interpretability.

Trend assessment shall be performed with a constant LOQ in order to avoid induced trend phenomena. As LOQ values may change over time, there is a need of a consistent treatment of measurements (where the LOQ exceeds a given LOQ_{max}) in order to avoid induced trend phenomena. Provisions on the calculation of a constant LOQ_{max} and treatment of measurements where the LOQs exceed the minimum requirements were laid down.

Starting point of trend/reversal assessment

It was considered as important to detect an increase in pollutant concentration of 30 % with a power of 90 % or higher. The starting point for trend assessment is the same as for operational monitoring and shall allow for an "early warning function" of the trend detection. Therefore, it is proposed to start the trend analysis at a level where the **CL95 of the calculated mean** exceeds 75 % of the limit value.

Length of time series for trend/reversal assessment

For trend assessment, based on the WFD minimum requirement regarding the monitoring frequency, which is once a year, and on the requirement that an increase in pollutant concentration of 30% should be detected with a power of 90 % a **minimum length of time series of eight years** was derived. In case of half-yearly or more frequent sampling the minimum length can be **five years** (at least 10 respectively 15 values).

For trend reversal assessment the estimation of the required minimum length of time series the procedure was similar to the one described for trend assessment. The outcome was as follows: In case of annual data the **minimum length is 14 years** (14 values). In case of half-yearly or more frequent sampling the minimum length is **ten years** (at least 18 respectively 30 values).

Minimum number of sites, Network Criterion, Treatment of LOQ values

Data Aggregation

Regularisation

Spatial aggregation

- arithmetic mean and CL
- weighted arithmetic mean and CL

optional

- Kriging mean and CL

Trend Assessment

Regularisation

Spatial aggregation

Trend assessment

Starting point

Min. length of time series

Max. length of time series

Frequency of trend testing

Trend Reversal Assessment

Starting point

Min. length of time series

Max. length of time series

LINKS TO OTHER WORKING GROUPS

The working group of this project (Common Strategy on the Implementation of the WFD - Key activity 2: Development of guidance on technical issues, 2.8 Guidance on tools on assessment and classification of groundwater) is one of ten working groups initiated by the EC to develop guidance on specific issues of WFD implementation. Due to the integrated approach of the WFD, interaction between the working groups is required. The work of this group and the outcome of the project has to be seen as closely related to the work of the other working groups.

Topics on which it will be necessary to find a common understanding and to develop guidance are i.e. monitoring network design (e.g. site density), monitoring frequency, analytical requirements for LOQ and LOD, guidance for delimitation of GW-bodies, characterisation of GW-bodies, data exchange format (GW-body description, quality data), identification of risk, presentation of results, groundwater action values, …

PROVISION OF RESULTS

Algorithms and Software Tool

The outcome of the project comprises the algorithm and a software tool (*GWstat*) for both the proposed procedure for data aggregation and trend/reversal assessment. *GWstat* can be downloaded free of charge from the project web-site.

Side Products

Summary of current practice in Member States on data aggregation and the calculation of trends, monitoring strategy and network design; Web based form for the general characterisation of groundwater bodies, quality data exchange format.

Project web-site

All findings of the project and underlying documents are available on the project web-site http://www.wfdgw.net. For the presentation of groundwater bodies and sampling sites, land use etc. a Web-GIS was implemented where selected results and aggregated data from the project can be accessed. The Web GIS-site is linked to the project web-site.

2 INTRODUCTION

2.1 BACKGROUND

The "Directive of the European Parliament and of the Council 2000/60/EC establishing a framework for Community action in the field of water policy", the so-called EU Water Framework Directive (WFD), defines in Article 4 "Environmental objectives" for surface water, groundwater and protected areas.

In order to achieve the "Environmental Objectives" for groundwater (Article 4(1)(b)), the WFD requires that specific measures shall be adopted to prevent and control pollution of groundwater. Such measures shall be aimed at achieving the objective of good groundwater chemical status. The criteria for achieving good groundwater chemical status are defined in Annex V (Annex V 2.3.2) and in particular require that concentrations of pollutants (in groundwater) do not exceed any quality standards applicable under other relevant Community legislation. Further there is a requirement to identify and reverse any significant and sustained upward trends in the concentration of pollutants. The Directive provides specifications on the identification of trends in pollutant concentrations originating from diffuse and/or point sources (see chapter 3).

This project focuses on the development of particular algorithms for the identification of trends in pollutants (Annex V 2.4.4) and a data aggregation method for interpretations and presentation of groundwater chemical status as defined in Annex V 2.4.5.

The project is part of the "Common Strategy on the Implementation of the WFD", which was developed by the European Commission to achieve a common understanding and approach on WFD implementation. The working group of this project is one of ten working groups initiated by the EC to develop guidance on specific issues of WFD implementation. Due to the integrated approach of the WFD, interaction between the working groups is required. The work of this group and the outcome of the project has to be seen as closely related to the work of the other working groups (see chapter 7)
For co-ordination of the working groups and activities under the Common Strategy a EC Strategic Co-ordination group was set up.

2.2 PROJECT TEAM

A consortium of partners from 11 EU Member States (Austria, Belgium, Denmark, France, Germany, Greece, Ireland, the Netherlands, Portugal, Spain and United Kingdom (England and Wales)) was formed under the leadership of the Federal Environment Agency ltd. – Austria (FEA). Thus it was ensured that the results of the project were based on data derived from a broad variety of different groundwater bodies within the EU.

Institutions from 5 countries (Finland, Hungary, Italy, Norway and Sweden) participated as observers in the project and in this capacity attended meetings and commented on the draft final report. Furthermore comments from the ECPA as stakeholder were received.

The project was commissioned and financed at approximately 1/3 by DG Environment of the European Commission and 2/3 by the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management. In kind contributions from the partners constituted another important input.

Sub-contractor in the project and responsible for the development of the statistical concepts was the German company "quo data".

Project co-ordination		Experts
	Federal Environment Agency ltd. – Austria	Johannes Grath
Sponsors		
	Federal Ministry of Agriculture, Forestry, Environment and Water Management - Austria	Karl Schwaiger
	European Commission - DG Environment	
Partners		
	AMINAL - Environment, Nature, Land and Water Administration, Water Division- Belgium	Johan Lermytte
	GEUS - Geological Survey of Denmark and Greenland - Denmark	Jens Stockmarr
	A.E.A.P. - Agence de l´eau Artois-Picardie - France	Laurent Pavard
	Hessisches Ministerium für Umwelt, Landwirtschaft und Forsten - Germany	Georg Berthold
	IGME - Institute of Geology and Mineral Exploration - Greece	Georgia Gioni, Costas Papadopoulos
	EPA - Environmental Protection Agency - Ireland	Paul Coleman
	RIVM - National Institute of Public Health and the Environment – The Netherlands	Leo Boumans
	INAG - Instituto da Água - Portugal	Ana Rita Lopes, Rui Rodrigues
	Ministerio de Medio Ambiente - Spain	Manuel Varela
	Environment Agency of England and Wales – United Kingdom	Rob Ward
Sub-contractor		
	quo data – Gesellschaft für Qualitätsmanagement und Statistik - Germany	Steffen Uhlig
Observer		
	Finnish Environment Institute - Finland	Reetta Waris
	Water Research Institute - Italy	Giuseppe Passarella
	Norwegian Water Resources and Energy Directorate - Norway	Panagiotis Dimakis
	Geological Survey of Sweden - Sweden	Magnus Åsman
	Water Resources Research Centre (VITUKI) - Hungary	Josef Deak

2.3 OBJECTIVES

The main goal of the project was to establish methods for the calculation of representative mean concentrations, for data aggregation and trend (reversal) assessment at the groundwater body level respectively for groups of groundwater bodies. The methods had to be suitable for Europe-wide application and implementation based on the provisions of the Water Framework Directive.

The following main aspects were considered (among other points):

- Development of an appropriate data aggregation method for the assessment of groundwater quality at the groundwater body level respectively for groups of groundwater bodies including the determination of minimum requirements for calculation.
- Development of an appropriate statistical method for trend assessment and trend reversal including the determination of the minimum requirements for calculation.
- Concentrations below the detection limit and groundwater pollution that is unevenly distributed within the groundwater body.
- Influences originating from diffuse and/or point sources.

In the discussion it was highlighted that a pragmatic way which can be implemented in different administration systems and applied for different hydrogeological conditions should be preferred as otherwise the proposed procedure could be of minor acceptance in the Member States. To allow for comparable assessment results throughout Europe it was agreed that one assessment method should be developed and proposed for each issue (data aggregation, trend and trend reversal assessment).

The methods and procedures proposed by this Working Group are related to several provisions which are subject of investigation in other EC Working Groups dealing with particular topics of WFD implementation. For example the delimitation of GW-bodies or groups of groundwater bodies and the selection of monitoring stations based on the provisions of the WFD were not subject of investigation in this study, however they will be of vital importance for groundwater quality monitoring and data assessment.

An enumeration of identified links to other Working Groups is given in chapter 7.

Note

Whenever in the report status assessment or good status is mentioned this refers to Annex V 2.4.5.

3 RELEVANT PROVISIONS OF THE WATER FRAMEWORK DIRECTIVE (2000/60/EC)

The groundwater provisions of the Water Framework Directive require both the achievement of particular standards applicable under other Community legislation on groundwater, and the identification and reversal of significant and sustained upward trends in the concentration of pollutants. The assessment of groundwater quality status is based on the following provisions:

- Initial characterisation to determine whether the body is at risk of failing to achieve the objectives set for it (Annex II Section 2.1). This includes information on both pressure and susceptibility. Further characterisation is carried out where required to refine this assessment under Section 2.1.
- For those bodies identified as being at risk, a further characterisation of the impact of human activity on the body of water is required (Annex II section 2.2).
- Surveillance monitoring of those bodies identified as being at risk to verify whether it in fact is at risk, and of bodies of water which cross international borders. For bodies at risk the parameters indicative of the relevant impacts are monitored. In addition, a set of core parameters (oxygen content, pH value, conductivity, nitrate and ammonium) are monitored at all bodies.
- Operational monitoring (at least once a year) for bodies confirmed as being at risk, sufficient to establish the chemical status of the water body, and establish the presence of any significant and sustained upward trend in concentration of any pollutant.

Clarification of the statistical aspects of the activities in the final indent is extremely important for a proper implementation of the Directive, and this study focuses on that.

3.1 ARTICLE 4 – ENVIRONMENTAL OBJECTIVES (EXCERPT WITH REGARD TO GROUNDWATER)

1. In making operational the programmes of measures specified in the River Basin Management Plans:

(b) for groundwater

(i) Member States shall implement the necessary measures to prevent or limit the input of pollutants into groundwater and to prevent the deterioration of the status of all bodies of groundwater, subject to the application of paragraphs 6 and 7 and without prejudice to paragraph 8 of this Article and subject to the application of Article 11(3)(j);

(ii) Member States shall protect, enhance and restore all bodies of groundwater, ensure a balance between abstraction and recharge of groundwater, with the aim of achieving good groundwater status at the latest 15 years after the date of entry into force of this Directive, in accordance with the provisions laid down in Annex V, subject to the application of extensions determined in accordance with paragraph 4 and to the application of paragraphs 5, 6 and 7 without prejudice to paragraph 8 of this Article and subject to the application of Article 11(3)(j);

(iii) Member States shall implement the necessary measures 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;

Measures to achieve trend reversal shall be implemented in accordance with paragraphs 2, 4 and 5 of Article 17, taking into account the applicable standards set out in relevant Community legislation, subject to the application of paragraphs 6 and 7 and without prejudice to paragraph 8.

3.2 ARTICLE 17 – STRATEGIES TO PREVENT AND CONTROL POLLUTION OF GROUNDWATER

1. The European Parliament and the Council shall adopt specific measures to prevent and control groundwater pollution. Such measures shall be aimed at achieving the objective of good groundwater chemical status in accordance with Article 4 (1) (b) and shall be adopted, acting on the proposal presented within two years after the entry into force of this Directive, by the Commission in accordance with the procedures laid down in the Treaty.

2. In proposing measures the Commission shall have regard to the analysis carried out according to Article 5 and Annex II. Such measures shall be proposed earlier if data are available and shall include:

a) criteria for assessing good groundwater chemical status, in accordance with Annex II.2.2 and Annex V 2.3.2 and 2.4.5;

b) criteria for the identification of significant and sustained upward trends and for the definition of starting points for trend reversals to be used in accordance with Annex V 2.4.4.

3. Measures resulting from the application of paragraph 1 shall be included in the programmes of measures required under Article 11.

4. In the absence of criteria adopted under paragraph 2 at Community level, Member States shall establish appropriate criteria at the latest five years after the date of entry into force of this Directive.

5. In the absence of criteria adopted under paragraph 4 at national level, trend reversal shall take as its starting point a maximum of 75% of the level of the quality standards set out in existing Community legislation applicable to groundwater.

3.3 CHARACTERISATION OF GW-BODIES (ANNEX II)

Annex II section 2.1 of the Directive provides the following specifications on the initial characterisation of the impact of human activity on the groundwater body:

Member States shall carry out an initial characterisation of all groundwater bodies to assess their uses and the degree to which they are at risk of failing to meet the objectives for each groundwater body under Article 4. Member States may group groundwater bodies together for the purposes of this initial characterisation. This analysis may employ existing hydrological, geological, pedological, land use, discharge, abstraction and other data but shall identify:

- *the location and boundaries of the groundwater body or bodies,*
- *the pressures to which the groundwater body or bodies are liable to be subject including:*
  - *diffuse sources of pollution*
  - *point sources of pollution*
  - *abstraction*
  - *artificial recharge,*
- *the general character of the overlying strata in the catchment area from which the groundwater body receives its recharge,*
- *those groundwater bodies for which there are directly dependent surface water ecosystems or terrestrial ecosystems.*

Annex II section 2.2 of the Directive provides the following specifications on the further characterisation of the impact of human activity on the groundwater body:

> *Following this initial characterisation, Member States shall carry out further characterisation of those groundwater bodies or groups of bodies which have been identified as being at risk in order to establish a more precise assessment of the significance of such risk and identification of any measures to be required under Article 11. Accordingly, this characterisation shall include relevant information on the impact of human activity and, where relevant information on:*
>
> – *geological characteristics of the groundwater body including the extent and type of geological units,*
>
> – *hydrogeological characteristics of the groundwater body including hydraulic conductivity, porosity and confinement,*
>
> – *characteristics of the superficial deposits and soils in the catchment from which the groundwater body receives its recharge, including the thickness, porosity, hydraulic conductivity, and absorptive properties of the deposits and soils,*
>
> – *stratification characteristics of the groundwater within the groundwater body,*
>
> – *an inventory of associated surface systems, including terrestrial ecosystems and bodies of surface water, with which the groundwater body is dynamically linked,*
>
> – *estimates of the directions and rates of exchange of water between the groundwater body and associated surface systems,*
>
> – *sufficient data to calculate the long term annual average rate of overall recharge,*
>
> – *characterisation of the chemical composition of the groundwater, including specification of the contributions from human activity. Member States may use typologies for groundwater characterisation when establishing natural background levels for these bodies of groundwater.*

3.4 QUALITY STATUS ASSESSMENT (ANNEX V)

Annex V section 2.4.5 of the Directive provides the following specifications for the interpretation of groundwater chemical status:

> *In assessing status, the results of individual monitoring points within a groundwater body shall be aggregated for the body as a whole. Without prejudice to the Directives concerned, for good status to be achieved for a groundwater body, for those chemical parameters for which environmental quality standards have been set in Community legislation:*
>
> – *the mean value of the results of monitoring at each point in the groundwater body or group of bodies shall be calculated; and*
>
> – *in accordance with Article 17 these mean values shall be used to demonstrate compliance with good groundwater chemical status.*

3.5 TREND/-REVERSAL ASSESSMENT (ANNEX V)

Annex V section 2.4.4 of the Directive provides the following specifications on the identification of trends in pollutant concentrations originating from diffuse and/or point sources:

> *Member States shall use data from both surveillance and operational monitoring in the identification of long term anthropogenically induced upward trends in pollutant concentrations and the reversal of such trends. The base year or period from which trend identification is to be calculated shall be identified. The calculation of trends shall be undertaken for a body or, where appropriate, group of bodies of groundwater. Reversal of a trend shall be demonstrated statistically and the level of confidence associated with the identification stated.*

3.6 TIMETABLE FOR WFD IMPLEMENTATION

year	WFD criteria (key words)	relevant Article or Annex
2000	WFD set into force	
2001		
2002	criteria for the assessment of good status, trend and trend reversal (Commission proposal)	Art. 17(2)a, b
2003		
2004	description of GW-bodies, human impacts etc.	Art. 5(1), Annex II
2005		
2006	establishment of monitoring programmes	Art. 8, Annex V
2007	interim overview of significant water management issues	Art. 14 (1) b
2008	production of river basin management plans - draft (involvement of interested parties)	Art. 14(1)a, c
2009	programme of measures; publication of river basin management plan	Art. 11(7); Art. 13(6)
2010		
2011		
2012	programme of measures operational	Art. 11(7)
2013		
2014		
2015	achievement of good status review and update of river basin management plan	Art. 4(1) Art. 13(7)
2016		
2017		
2018		
2019		
2020		
2021	review and update of river basin management plan	Art. 13(7)

4 GROUNDWATER BODIES SUBJECT TO THE STUDY

It was essential that the methods developed in this project are suitable for all groundwater bodies in Europe, which show a broad variety of size, pressures, hydrogeological conditions, level of pollution, monitoring network design and monitoring frequency.

Therefore, project partners were asked to provide general information and test data sets on groundwater quality of selected groundwater bodies from their country covering this broad variety for the testing of different statistical methods for data aggregation and trend (-reversal) assessment. Furthermore, the basis for the development of statistical procedures was the description of methods already applied in the EU Member States.

This section gives a brief summary of the general information provided on the groundwater bodies subject to the study and of the groundwater quality data for selected parameters. Furthermore, an overview of the methods for spatial and temporal analysis of the groundwater quality data applied in the Member States is given.

4.1 DATA COLLECTION/EXCHANGE

For the exchange of information within this project a web-site was implemented based on the CIRCA[1]-extranet tool developed by the EC. This system enables user-specific, password protected access to information and includes among other features a notification system for information on the upload of new files. With this system both partners in the project and contracting parties were continuously informed on the current state of work.

For the collection of information on the general description of the GW-bodies an on-line questionnaire was elaborated. Partners had direct, password protected access to a tailor-made data base.

For data storage and analyses of groundwater quality data the computer programme *WaterStat* was applied. This software was developed by "quo data" and adapted to the special requirements of the project. The requirements concerning the data transfer from the partners were determined by the database structure respectively the structure of the single tables of the database. For data import a flexible import module was developed, which allows transfer of data from Excel files and ASCII-files into the database.

As required in the contract the results of the project are published on a web site (http://www.wfdgw.net). With the exception of groundwater quality raw data all information gathered within the project as well as the findings and products of the project will be available on this dedicated web site.

[1] **C**ommunication **I**nformation **R**esource **C**entre **A**dministrator). CIRCA is an extranet tool, developed under the European Commission IDA programme, and tuned towards Public Administrations' needs. It enables a given community (e.g. committee, working group, project group etc.) geographically spread across Europe (and beyond) to maintain a private space on the Internet where they can share information, documents, participate in discussion fora and various other functionalities. This private space is called an '*Interest Group*'or '*User Group*'. The access and navigation in this virtual space is done via any Internet browser and Internet connection.

4.2 GROUNDWATER BODY DESCRIPTION

The general description of a GW-body provides an essential basis for the interpretation of its quality data. The information requirements for the general description of GW-bodies for the project were based on the provisions of the WFD (Annex II).

The general description was divided into

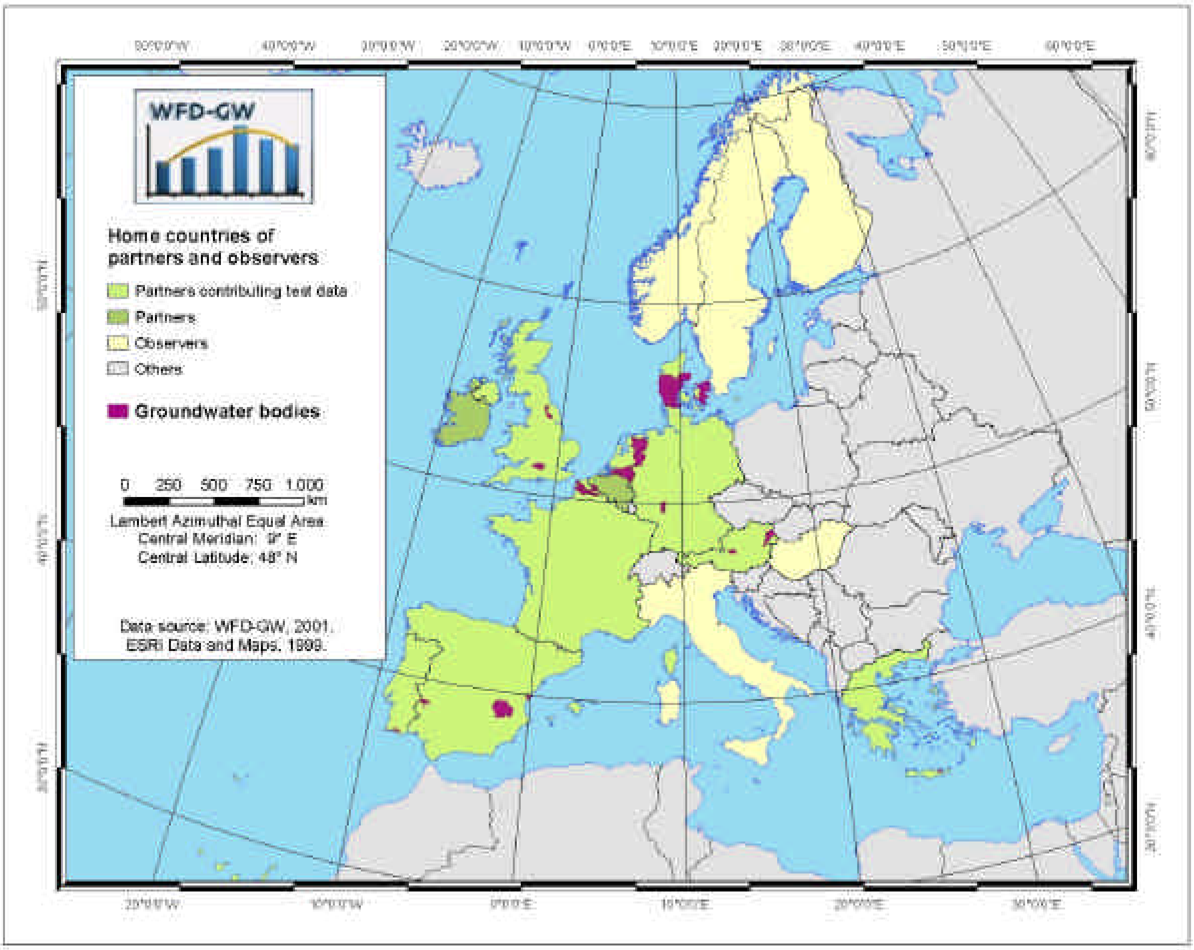
- a verbal description (1–2 pages) including a geological sketch/cross-section,
- a general characterisation of the groundwater body on the basis of a questionnaire, and
- a GIS map.

4.2.1 Geographical coverage

Figure 1 shows the 16 countries involved in the project as partners (11) or as observers (5) and gives an impression of the geographical distribution of the 21 GW-bodies on which information was provided.

The geographical information on the GW-bodies included their boundaries and the co-ordinates of the sampling sites. In addition to identifying the location of groundwater bodies on a map for the information enabled the assessment of land use on the basis of CORINE Landcover data and an assessment of the monitoring network design for each GW-body. Furthermore, the geographical data were used for analysing groundwater quality data by spatial weighting methods (e.g. kriging).

Figure 1: Partners and observers in the project and GW-bodies on which information was provided



4.2.2 Verbal description

Each GW-body subject to the project was briefly introduced by a verbal description. This supported the interpretation of the general description (provided in an on-line questionnaire) and the assessment of the results of the statistical analyses. The verbal descriptions can be found in Annex 3 and comprise for most of the GW-bodies:

- name and location,
- information on the importance of the groundwater body (why was it chosen for the study),
- a description of the geological situation (including a geological profile),
- information on the monitoring strategy,
- details on quality problems and the pressure situation, and
- whether information provided for the general description was mainly based on measurements or on estimates.

4.2.3 General characterisation

More detailed information on the general characteristics of each GW-body was collected via an on-line questionnaire. This collection mode improved the comparability of information and avoided manipulation before the subsequent computerised assessment.

The contents and the definitions of the questionnaire (as well as the supporting glossary and help texts) were developed with special focus on the objectives of the project and were discussed and agreed within the project team. The detailed questions were based on the provisions of the WFD, Annex II (including initial characterisation and main elements of the further description) and comprised the following issues:

- general information,
- meteorological characterisation,
- hydrogeology and
- human impacts.

It must be clearly stated that the information obtained by this method cannot be considered exhaustive. Detailed analyses of the validated collected information can be found in Annex 2. Summarised information on the variety of the GW-bodies according to the general characterisation is given in the following chapter.

4.2.4 Summary characterisation – variety of groundwater bodies

As it is essential, that the methods developed in this project fit for all groundwater bodies in Europe, partners in the project were asked to provide information on groundwater bodies with different hydrogeological settings. This resulted in a collection of 21 groundwater bodies with the following variety of characteristics:

The size of the groundwater bodies subject to the project varies between 8 km^2 and approximately 10 600 km^2. The karstic groundwater bodies have an extent of less than 1 000 km^2 (except for the Spanish Mancha Oriental hydrogeological unit (ES0829) and the Danish groundwater body Zealand (DK300)) smaller than the groundwater bodies in porous media. The Danish groundwater bodies were the largest and range from about 5 800 to 10 600 km^2.

The meteorological data indicate a broad range of climatic conditions. In semiarid regions from Portugal, Spain and Eastern Austria mean annual precipitation is lowest with values below

600 mm/year. In all other regions mean precipitation ranges from 600 to 1 000 mm/year. An exception is the only alpine groundwater body in this study, the Austrian Dachstein massif (AT154) with 1 800 mm/year, due to its high average elevation of approximately 1 800 m above sea level.

The Dachstein is also by far the coldest place with an average annual air temperature of only 2 °C. The mean temperatures measured at the Western European sites range all between 7.5–10.5 °C, while the groundwater bodies in Portugal, Spain and Greece had temperatures above 12 °C representative of a temperate climate.

The groundwater bodies also represent a broad range of hydrogeological settings with all major types of aquifers represented in this study. The porous type from Quaternary alluvial deposits is most common, followed by karstic aquifers. Also GW-bodies in fractured aquifers from the British and French chalk deposits and one GW-body in lithified porous bedrock (sandstone) were contributed. All GW-bodies are unconfined except the French Calcaire carbonifère paleokarst (FR202) which is covered by impermeable secondary layers and the Danish GW-bodies DK200 and DK300.

The average protective cover above the groundwater bodies, which is of major importance for the vulnerability of the groundwater, varies from few decimetres topsoil to a coverage of low permeable layers with a total thickness of up to 40 m.

Whereas the mean depth to groundwater does not exceed 40 m in porous and fractured aquifers, it is higher in karstic aquifers.

A wide range of values was also provided for the hydraulic conductivity on which depend percolation time and transmissivity of an aquifer.

The highest hydraulic conductivity is found at the karstic conduit system of the Austrian Dachstein massif with an estimated velocity of 1.0 E-2 m/s, followed by porous GW-bodies with sand and gravel deposits and other karst areas. The fractured chalk and especially the British Sherwood Sandstone Group groundwater body (UK006) with an average hydraulic conductivity of only 5.6 E-6 m/s represent low permeable aquifers.

Diversity in land use is also obvious in the different areas. In most GW-bodies agricultural areas dominate land use. Their shares range from about 10 % up to 90 %. Forests and semi-natural areas dominate land use in several karst areas in Austria, Greece and Portugal.

The areas selected also show considerable variability with regard to nature and extent of pressures affecting groundwater bodies. Whereas in mountainous regions like the Austrian Dachstein (AT154), the French Calcaire carbonifère (FR202) or the Agios Nikolaos karst GW-body in Greece (GR100) very few pressures exist, groundwater problems caused by water abstraction and agriculture exist in most of the groundwater bodies located in sedimentary basins. Pressures from artificial recharge, influencing infrastructures, industrial plants and contaminated sites are subordinate in number but show also a divers potential for point source- and non-point source pollution.

In summary it can be stated that the 21 groundwater bodies chosen for this project include all major types of GW-bodies in Europe and show a broad variability with regard to hydrogeology, meteorology, land use and pressure situations.

4.3 QUALITY DATA

The requirements for quality data depended highly on the objectives of the project. The selection of parameters was based on the requirements of the WFD and the national monitoring situations.

The general requirements for the quality data exchange were specified by the FEA and "quo data" after discussion and agreement within the project team. Special emphasis was placed on the importance of long time series and a representative choice of the groundwater bodies. Another important item of discussion was the treatment of data below the detection and quantification limit.

4.3.1 Selected Parameters

Based on the provisions of the WFD and presentations on the national monitoring programmes from the partners the following range of parameters was chosen as the subject of the project. These were pH-value, electric conductivity, dissolved oxygen, nitrate, ammonium, chloride, pesticides and chlorinated hydrocarbons. Whereas the first five are explicitly mentioned in the WFD, chloride is included implicitly as indicator for seawater intrusion (it can also be an indicator for several other impacts). Pesticides and nitrate are of relevance as an indicator for diffuse sources of pollution. Chlorinated hydrocarbons were used as indicator for point sources of pollution.

It was agreed to use - as far as available - the provisions of the Drinking Water Directive for the determination of parameters and units.

For the collection of quality data several minimum requirements were defined (e.g. sampling site identification (unique code), location, sampling date, detection limit and limit of quantification). Other information was optional but of added value for the interpretation of the analysis (e.g. type and material of the sampling site, sampling depth etc). These specifications might provide information on a possible bias of monitoring results e.g. a predominance of drinking water wells could give an impression of better water quality than data from other types of sampling sites.

The structure required for storage of this information was determined by the database design and the elaborated data exchange format.

4.3.2 Data provided

Groundwater quality data for 69 parameters were submitted by the project partners. Amongst them were a number of pesticides, heavy metals and chlorinated hydrocarbons. Of these 6 core parameters (pH-value, electric conductivity, dissolved oxygen, nitrate, ammonium and chloride) and 9 additional parameters were selected for further statistical treatment. The selected pesticides, heavy metals and chlorinated hydrocarbons cover a broad range of concentrations including values below the limit of quantification and areas representative of the different kinds of pressures. Table 1 shows the 15 analysed parameters and the availability of data for each groundwater body.

Table 1: Analysed parameters

Parameter	AT154	AT224	AT250	DE001	DK100	DK200	DK300	ES0409	ES0812	ES0829	FR001	FR202	GR100	NL002	NL004	NL005	PTA2	PTM2	PTM5	UK002	UK006
pH-value		x	x					x	x	x	x	x		x	x	x	x	x	x		
El. conductivity	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x		
Ammonium(NH_4)	x	x	x	x				x	x	x	x	x		x	x	x	x	x	x	x	x
Nitrate (NO_3)	x	x	x	x				x	x	x	x	x		x	x	x	x	x	x	x	x
Dissolved oxygen	x	x	x	x							x	x									
Chloride (Cl)	x	x	x	x				x	x	x			x				x	x	x		
Nitrite (NO_2)																				x	x
Atrazine	x	x	x	x	x	x	x														
Tetrachloroethen	x	x	x																		
Cadmium (Cd)					x	x	x														
2-6-dichlorbenzamid					x	x	x														
Nickel (Ni)					x	x	x														
Lead (Pb)					x	x	x														
Selenium (Se)					x	x	x														
Vanadium (Va)					x	x	x														

4.4 STATISTICAL ASSESSMENTS IN PARTNER COUNTRIES

The statistical methods applied for data aggregation and trend assessment were described and presented by the partners and a summary table was set up. This inventory also includes information on the monitoring design, sampling procedure etc. (for further details see Annex 12).

4.4.1 Chemical status assessment

The following methods are applied in the Member States for assessing the chemical status:

- Median is used by UK, AT, PT and DK (median of annual median concentrations),
- the arithmetic mean (partly with confidence range) is used by NL, ES, AT,
- the mean based on the log-normal distribution (with confidence range) is used by BE,
- the percentage of sites with good quality (partly with tolerance intervals) is applied by NL and DE,
- maximum and minimum values are reported by almost all MS.

It should be noted that maximum and minimum values are considered as accompanying parameters, but not as parameters reflecting the overall GW-body status. Furthermore, the treatment of measurements below the limit of quantification (LOQ), the limit of determination (LOD) and the treatment of unequally distributed sites is not specifically addressed by the Member States.

The working hypotheses on the spatial aggregation methods which are applied in the Member States are summarised in Table 2. The assessment is split into Very good, Good, Fair and Poor. Further examinations and demonstrations are to be found in Annex 4.

Table 2: Aggregation methods applied in Member States (assessment of the working hypotheses - in Very good, Good, Fair and Poor)

	Arithmetic Mean	Mean based on log-normal distribution	Median	% of sites with good quality	Maximum / minimum
reflects the overall status of the GW-body (sites evenly distribution)	Very good	Very good –in case that the coefficient of variation is less than 80 %	Fair–good – in case that the coefficient of variation is less than 80 %	Fair-good – in case there are sites with good and with bad quality	Poor
reflects the status which is not exceeded in more than 50 % of the area (sites evenly distrib.)	Fair–good	Fair–good	Very good	Poor	Poor
reflects the impact of hot spots	Fair-good	Poor	Poor	Fair-good	Very good
Outlier-sensitivity	Poor	Poor	Very good	Good	Poor
applicability for measurements below LOQ, LOD	replacement of - values below LOQ/LOD by substitute values may introduce some bias	share of values below LOQ/LOD <80%: applicable	share of values below LOQ/LOD <50 %: applicable	If the LOQ/LOD is below the limit value for good quality: applicable	Very good

4.4.2 Trend assessment

For trend analysis the following methods for aggregating raw measurement results were reported from the Member States:

- Raw concentration data are used by UK, GE, GR, FR and AT,
- arithmetic mean is used by AT and NL,
- median of the annual median concentration of sites all over the country is used by DK.

It is concluded that, apart from NL, AT and DK, trend analysis is not based on spatial aggregation of data. DK uses the median, whereas NL and AT use the arithmetic mean.

Furthermore, it should be noted that the treatment of values below the limit of quantification/detection and the treatment of unequally distributed sites is not specifically addressed by the Member States.

The following statistical methods for temporal trend determination were reported from the MS:

- Regression analysis (simple linear regression) is applied by GE, GR, NL, FR and AT,
- the non parametric trend test of Mann Kendall is used by DK,
- a piecewise regression method for estimating a non linear trend in nitrate concentrations has been developed by UK,
- the F-test for a comparison of two levels is applied by GR.

Some Member States do not apply temporal trend detection methods on a routine basis, but focus their statistical analyses on spatial aspects and assessments of the current level.

Several things must be considered when choosing a method of testing the statistical significance of a measured trend:

- is the method relevant to the objectives of the assessment,
- are the assumptions underlying the method valid, and
- is the method sufficiently powerful.

If the assumptions of the method include specific requirements for the distribution of the data, e.g. normality and homogeneity of the error distribution, and these are unlikely to be met because of potential outliers, there will be a further requirement that:

- the method is robust.

In the context of trend assessment, relevance means that the method is sensitive to the kinds of changes of concern in the assessment. Not all tests are equally effective at detecting all patterns of change. For a very focused test, it may be a disadvantage if all patterns of change are of interest, or an advantage if the focus is put on patterns of interest. Robustness in the current context refers to the degree of sensitivity with regard to outliers.

Four groupings of patterns of change are of interest

	Patterns of change			
	linear trend	**non-linear monotonic**	**1st upward trend 2nd downward trend**	**1st downward trend 2nd upward trend**
linear trend (upward or downward)	x			
monotonic trend (upward or downward)	x	x		
systematic change	x	x	x	x
trend reversal			x	x

x…denotes patterns of change covered by the corresponding grouping

Each of the four statistical methods reported by the MS and listed above fulfils at least one of the specific functions and tasks of a trend analysis. A detailed description and assessment of the specific functions of a trend analysis can be found in Annex 4

The following table gives a very rough summary of the characteristics of the trend tests applied in the Member States with regard to their power to detect different types of patterns of change and their robustness. The assessment is split into: Very good, Good, Fair and Poor.

	Power (under Normality)				Robust
	Linear trend	**Monotonic trend**	**Systematic trend**	**Trend reversal**	
Mann-Kendall	Very good	Fair-good	Poor	Not applicable	Yes
Linear Regression	Very good (slightly better than Mann-Kendall)	Poor	Poor	Not applicable	No
2-sample comparison (F-test)	Fair-good	Fair	Poor	Not applicable	No

5 STATISTICAL METHODS AND PROCEDURE

Several methods for data aggregation, trend and trend reversal assessment, which were generally applicable, were implemented in a software tool and tested with the provided test data sets. The results were then discussed within the working group.

The investigated statistical methods and the proposed procedures are briefly described in the following subsections. Detailed explanations of the statistical background and example calculations can be found in Annex 4 and Annex 6.

The proposed statistical methods are based on several provisions and certain requirements as outlined below. For the application of each particular statistical method certain restrictions and requirements have to be considered.

5.1 PROVISIONS

The statistical methods were developed in accordance with the following provisions:

- the provisions of the WFD (see chapter 3),
- the provisions of the contract,
- the requirements on the statistical procedure and the monitoring network as agreed by the working group,
- consideration of the statistical methods applied in different Member States.

The principal basis for the development of the methods were the provisions of the WFD. Aggregated groundwater quality data are needed at the following stages of WFD implementation:

- identification of a GW-body as being at risk of failing to meet objectives under Art. 4 based on surveillance monitoring data,
- status assessment procedure(according to the project objectives) and/or identification as being at risk of failing to meet objectives under Art. 4 based on operational monitoring data,
- trend (reversal) assessment,
- definition of the starting point for trend assessment.

Identification of long term anthropogenically induced upward trends and reversal of such trends is mentioned in Annex V and Art. 4 of the WFD (see chapter 3).

5.2 REQUIREMENTS

5.2.1 Requirements on statistics

The working group agreed on the following general requirements on the statistical methods for data aggregation and trend assessment:

- statistical correctness,
- pragmatic solution,
- establishment of one method (suitable for small, large and groups of GW-bodies as well as for small GW-bodies with a small number of sampling sites),

- applicable for all types of parameters, and
- ability to accommodate uneven distribution of pollution caused by local or diffuse sources, observed at some points in a GW-body which show higher concentrations than the rest of the GW-body.

In particular for trend assessment and trend reversal assessment the following requirements had to be met:

- extensible to potential adjustment factors,
- sufficient power for the detection of trend/reversal,
- robustness was considered less important than power and extensibility (data validation will be responsibility of MS).

5.2.2 Requirements on the monitoring network

The definition of GW-bodies, sub-bodies and groups of GW-bodies is a prerequisite for the designation of a network for GW-monitoring. The monitoring network design shall be homogenous in order to guarantee spatial representativity. Homogeneity implies furthermore that there is no local accumulation of sites. Representativity with regard to anthropogenic and natural factors was also regarded as important.

For the assessment of the homogeneity of a monitoring network, criteria were developed within the project – see chapter 5.3.1and 5.4.1.

5.2.3 Requirements on the monitoring

Sampling techniques were regarded as important since considerable bias can be avoided by applying a sound sampling strategy. Consequently the quality of data can be improved. The following particular aspects were highlighted:

- The importance of continuity with regard to the monitored sampling sites. The replacement of sampling sites should be kept as low as possible. In case of changes of monitoring stations it should be assured that these changes do not the affect the outcome of the assessment.
- In a time series some observations may be missing, but the missing of two or more subsequent values should be avoided, as this would cause a risk of bias due to extrapolation.
- Samples should be taken within a certain period of a year to avoid bias by seasonal effects. In particular for yearly measurements it should be guaranteed that the measurements are taken in one and the same quarter or within a certain time period of the year. This is required to avoid a high random variation which reduces the power of the trend analysis.
- The sampling frequency should reflect the natural conditions and dynamics of the GW-body.

5.2.4 Requirements on the quality assurance

The sampling procedure itself and chemical analysis should ensure continuity in results. Relevant standards are to be applied. The importance of adequate treatment of samples was emphasised – e.g. conservation of samples, filtration – yes or no, immediately or in the laboratory, type of filtration, etc. It was regarded as important to record the applied analytical methods, to ensure comparability of results. For several groups of substances provisions for the limit of quantification and the limit of detection are of vital importance for a sound data basis for the assessment.

5.3 INVESTIGATED METHODS

The following chapters characterise **all methods investigated** for network assessment, treatment of LOQ/LOD, data aggregation, trend assessment and trend reversal assessment with regard to their applicability, interpretability and some statistical properties. All these methods were tested, examples were elaborated and results were discussed within the working group. Based on this, a procedure for data aggregation and trend (reversal) assessment is proposed (see chapter 5.5).

5.3.1 Network criteria

It was agreed that homogeneity (reflecting spatial representativity) of the network is a prerequisite and should be ensured to allow for sound statistical assessment in accordance with the requirements of the WFD. If the GW-body is hydrogeologically heterogeneous and if a spatially homogeneous monitoring network is not feasible or sensible the monitoring has to be developed in a hydrogeologically representative way.

Several network criteria for the assessment of the homogeneity of the network were developed and investigated:

- uniform distribution of sampling sites over the whole water body,
- no local accumulation of sites,
- the share of the GW-body represented by each site should be almost constant (1/n).

Investigated Representativity Indices

R_1 Minimum distance between any two sites, expressed as percentages of the minimum distance for an optimal network:

- depends very much on local accumulation of sites,
- can easily be improved by reducing the number of sites, so that networks with a small number of sites are favoured.

R_2 10% percentile of distances between any two sites, expressed as percentages of the minimum distance for an optimal network:

- less dependent on local accumulation of sites, but still favouring networks with reduced number of sites,
- not sufficiently sensitive with regard to "holes" in the network (holes are not measured, only the distances between the sites).

R_3 Maximum Kriging weight expressed as percentages of the Kriging weight for an optimal network with constant weights (inverse representation):

- reflects the characteristics of the network only locally (at the site with maximum weight).

R_4 Relative standard deviation of Kriging weights:

- reflects the characteristics of the network quite properly, but does not reflect holes in the network,
- is only relevant if the range of the spatial correlation is larger than the distances between sites: Networks with low sampling site density obtain better ranking than networks with high sampling site density (in other words: for networks with low sampling site density the network design does not matter). Hence, again networks with a smaller number of sites are favoured.

R_U Average minimum distance between any location in the area to the closest sampling site, expressed in percentages of the average distance for an optimal network (inverse presentation):

- R_u is highly correlated with R_4, but does not favour smaller networks and reflects holes in the network properly.

5.3.2 Treatment of LOQ/LOD values

With regard to LOQ and LOD, the following definitions apply: LOQ denotes the "Limit of Quantification" or "Determination limit", whereas LOD denotes the "Limit of Detection". Generally LOD is below LOQ, and there are three different scenarios, depending on analytical methods and standard operation procedures which may be different in different MS:

- Measurements below LOQ are not quantitatively reported, and LOD is not reported: This is the scenario underlying all calculation schemes of this project.
- Measurements below LOQ are not quantitatively reported, but it is reported whether measurements <LOQ are above or below LOD: This was not taken into account in the calculation schemes, since none of the partners reported such data. However, it is highly recommended to extend the calculation scheme accordingly. This would allow for a considerable reduction of the bias due to <LOQ measurements.
- All measurements above LOD are quantitatively reported: In this case in all calculation schemes "LOQ" can be replaced by "LOD", i. e. below LOD can be treated in the same way as data below LOQ.

Throughout the report it is assumed that measurements below the LOQ are not available as quantitative figures. However, all statements hold also if the LOQ is replaced by the LOD.

Note: From a statistical point of view, measurements below the LOQ can be taken into account, since the assessment is not based on single measured values (for which the LOQ is focused on), but on aggregated data. It would even improve the quality of the assessments if all measurements above the LOD would be available as quantitative figures and would be treated as regular measurement values. However, for practical and administrative reasons it could be preferable to make measured values only available if they exceed the LOQ.

In some cases measurements below the LOQ are not known quantitatively, but it is known whether or not they are above the LOD. It would be useful to incorporate this information as well.

However, information on the LOD is currently not available in the Member States. As information is only available for the LOQ all following considerations within the project refer to the treatment of the LOQ. The concept can be modified if the LOD is reported.

For the treatment of "lower than LOQ" measurements a minimax approach (minimize maximum risk) is applied. Several different percentages of replacement of measurements below the LOQ were investigated and biases were calculated. The shares range from the replacement by "0" up to the replacement by LOQ (i.e. 0 %–100 %).

Calculations based on the test data sets showed that for several parameters detailed assessments were hardly possible due to the high share of values lower than LOQ and/or relatively high LOQs in comparison to the limit value (LV). This problem occurred in particular for pesticides and for a few other parameters. Within the project the maximum impact of "lower than LOQ" measurements on the arithmetic mean as well as the behaviour of the confidence limit was considered.

As LOQ values may change over time, there is a need of a consistent treatment. This is of major interest with regard to trend (-reversal) assessment. Detailed assessments, considerations and examples can be found in Annex 4.

The discussion showed that there is an urgent need to provide sufficient information on both LOQ and LOD. This will have to be considered when defining monitoring requirements and analytical procedures. (see also chapter 7)

5.3.3 Data Aggregation

In general the procedure for data aggregation consists of a regularisation of the measured data for each sampling site and a spatial aggregation of the regularised data for a GW-body or a sub-body.

Assuming that chemical status assessment of a GW-body should be based on the annual arithmetic mean values at each site, the following methods were investigated (The classification of methods refers to chapter 4.4.1. The assessment is split into Very good, Good, Fair and Poor):

Regularisation

For each site the arithmetic mean of the concentration data shall be calculated. This provision for determining groundwater chemical status (according to the project objectives - see 2.3) is laid down in Annex V section 2.4.5 of the WFD.

For the treatment of values below LOQ a replacement method based on minimum and maximum calculations was applied.

Quantiles (Median, 70 % Percentile)

The quantiles (50 %, 70 %, 90 %) represent the concentration which is not exceeded at more than x % (50 %, 70 %, 90 %) of the stations (of the area - in case of evenly distributed sites).

The reflection of the overall status of the GW-body by quantiles is fair to good if the coefficient of variation is less than 80 % but they do neither reflect outliers nor the impact of uneven distribution of pollution caused by local or diffuse sources, observed at some points in a groundwater body which show higher concentrations than the rest of the groundwater body.

Methodologies of quantiles would require at least 10 sites for statistically sound results. The fact was raised that in several countries small GW-bodies exist, which are monitored by less than 10 sites and cannot be grouped due to different hydrogeological conditions,.

Therefore, a data aggregation method had to be provided for GW-bodies with less than 10 sampling stations. For that reason the methodologies with quantiles were excluded from further discussion.

Arithmetic Mean (AM) and Confidence Limit (CL_{AM})

The arithmetic mean reflects the overall status of the GW-body very well and its reflection of the status which is not exceeded in more than 50 % of the area (in case of evenly distributed sites) is fair to good. The outlier sensitivity of the arithmetic mean is poor and its reflection of the impact of uneven distribution of pollution caused by local or diffuse sources, observed at some points in a groundwater body which show higher concentrations than the rest of the groundwater body is fair to good. The replacement of values below LOQ by substitute values may introduce some bias.

Confidence limits for the mean are an interval estimate for the mean. Instead of a single estimate for the mean, a confidence interval generates a lower and upper limit for the mean. The interval estimate gives an indication of how much uncertainty there is in the estimate of the true mean. The narrower the interval, the more precise is the estimate.

Compliance with good groundwater chemical status at a given level of confidence can be demonstrated with a statistical test for the null hypothesis

- H_0: "GW-body is not in good status, i.e. true mean level above the limit value"

and the alternative hypothesis

- H_1: "GW-body is in good status, i.e. true mean level below the limit value"

H_1 may be considered as statistically proven at significance level alpha/2, if the corresponding upper CL at confidence level (1-alpha) (e.g. 95%) is below the limit. Value alpha denotes the probability of making a wrong decision for a good status (although the true, unknown mean exceeds the limit value); alpha might vary for different parameters[2].

The CL decreases with an increasing number of stations within the GW-body or a decreasing variability of concentration levels.

Even with 3 stations in a GW-body the CL is applicable, as it automatically considers the required confidence. For AM or percentiles the statistical minimum requirement would rather be 10 stations. Therefore the application of the CL was decided by the working group. Generally with 3 stations the CL is higher than with 10 stations, but under the (not really justified) assumption of a normal distribution and a not too extreme variation of concentration levels, it is still possible to achieve positive results. Further details on the normal distribution assumption can be found in Annex 4.

The use of the CL allows to reduce the number of stations in GW-bodies with levels far below the limit value, and enforces a higher number of stations in GW-bodies with levels close to the limit value, thus allowing an effective allocation of analytical resources.

Weighted Arithmetic Mean (wAM) and Confidence Limit (CL_{wAM})

The weighted arithmetic mean (wAM) was introduced for GW-bodies which can be divided into sub-bodies. It takes into account the share of the sub-bodies and the corresponding arithmetic means.

In order to be consistent with the AM, the calculation of the CL_{wAM} is performed under the same model assumption as with CL_{AM} , i.e. all measurements are assumed to be stochastically independent and identically distributed (In case of even distribution of sites both the AM and the wAM could be applied, and the results should not differ). Furthermore the 1-way random effect model should not be applied since the definition of the sub-bodies is not independent from the measured data. Hence it is assumed that there is no spatial correlation at all, and under this assumption calculation leads to the approximate upper confidence limit. If the sites are evenly distributed, calculation shows that $CL_{wAM} = CL_{AM}$.

Theoretical considerations on the derivation of the confidence limit for the wAM include the model 0 (no sub-body effects), model I (with fixed effects) and model II (with random effects) as attached in Annex 5.

Kriging Mean (KM) and Confidence Limit (CL_{KM})

The kriging mean represents the average concentration in the area of a GW-body and corrects for a heterogeneous distribution of stations.

It has to be clearly stated that the calculation of the kriging mean is used as an instrument for the calculation of a **weighted mean** and its **CL95**. This way of calculating a **spatial mean** must not be seen as a tool for modelling the regional level of concentration of pollutants in the GW-body.

Kriging reflects the spatial structure of the GW-quality data and also to a certain extent the impact of factors affecting the concentration level within an area as there are e.g. land use, hydrogeological conditions, etc. if these are spatially correlated.

In case of no spatial correlation of these factors, results of kriging will be similar to those of arithmetic mean calculations. In general the difference between the kriging mean and the arithmetic mean decreases if the monitoring network is homogeneously designed and representative with regard to human impacts and environmental conditions.

---

[2] It should be noted that due to deviations from the normal distribution the actual probability of making a wrong decision can be slightly larger than alpha/2. In the examples presented the confidence level is 95%, and the formal significance level for the one-sided test is therefore 2.5%. This guarantees that the actual probability of making a wrong decision will be in most cases below 5%, even in case of severe deviations from the normal distribution.

The extension of the model to explicitly include hydrogeological information, etc. would require a much more complicated statistical algorithm and this would not allow to meet the objective of the project to develop a pragmatic solution which can be implemented Europe-wide. However, it is the decision of the MS whether to perform more detailed assessments.

The kriging method and its advantages were demonstrated on several examples. Especially the effect of down-weighting the impact of clusters of stations and giving higher weights to solitary stations was shown. It was pointed out that there is a close relationship between the arithmetic mean and the kriging mean in case of evenly distributed sites or in absence of spatial auto-correlation.

The kriging analysis can also be used for calculating the upper confidence limit for the kriging mean and for the 70 % percentile.

Maximum likelihood (ML) approach based on the censored log normal distribution

The advantage of the Maximum Likelihood (ML) approach is its applicability if the share of measured values below LOQ is very high. Especially for parameters like pesticides and chlorinated hydrocarbons the share of "lower than" values may be quite high. The ML approach corrects for bias caused by "lower than" measurements under the assumption of log normal distribution. The ML approach can be used to obtain both an estimate for the mean and for the 70% percentile.

However, it was demonstrated by examples and theoretical considerations that the ML approach based on the log-normal distribution may produce artificial outliers if the data are not log-normally distributed. Problems appear especially if more than 80 % of the values are below the LOQ and in case of a high variation of the numeric values (causing a very large variability of the ML parameters). As this was frequently the case for the test data sets it was decided that this methodology would no longer be taken into account.

Percentages of sites with good quality approach

The method of percentages of sites with good quality reflects the overall status of the GW-body fair–good if there are sites with good and with bad quality. The status which is not exceeded in more than 50 % of the area (in case of evenly distributed sites) is reflected poorly. The outlier sensitivity of the percentage of sites method is good and its reflection of the impact of uneven distribution of pollution caused by local or diffuse sources, observed at some points in a groundwater body which show higher concentrations than the rest of the groundwater body is fair–good. The applicability of the method for measurements below LOQ depends on the limit value. The methodology of percentages of sites with good quality is applicable if the LOQ values are below the limit value for good quality.

Minimum, Maximum

Minimum and maximum reflect the impact of uneven distribution of pollution caused by local or diffuse sources, observed at some points in a groundwater body which show higher concentrations than the rest of the groundwater body very well and the applicability for measurements below LOQ is very good. The overall status of the GW-body as well as the status which is not exceeded in more than 50 % of the area is poorly reflected by minimum and maximum values. Furthermore, the method is not outlier sensitive.

5.3.4 Trend assessment

As already mentioned at the beginning of this chapter the following requirements on the statistical procedures were additionally considered for the selection of the trend assessment method:

- applicability for all types of parameters,
- extensibility to potential adjustment factors,
- sufficient power for the detection of a trend/reversal,
- robustness was considered less important than power and extensibility (data validation will be responsibility of MS).

Within the discussion it was highlighted that a pragmatic way which can be implemented in different administration systems should be preferred as otherwise the recommended way could be of minor acceptance in MS. This could lead to the development of separate methods in Europe and results and assessments would not be comparable.

It was not a goal of the project to provide tools for the calculation of cause-effect relationships. This will be left to the discretion of the MS. The applied trend calculation model should be extendable to an adjustment by factors such as precipitation, water level, nitrogen application, etc.

The power of a trend assessment method stands for the probability to detect a trend. The power depends on the intensity of the trend, the trend detection method, the significance level, the random variation of data from year to year and the length of the time series.

The general procedure of a trend assessment and trend reversal assessment method is based on the same procedures as applied for data aggregation (regularisation and spatial aggregation – see chapter 5.3.3).

Assuming that any trend analysis should be based on aggregated data from the whole GW-body, the following methods were discussed:

Regularisation

For each sampling site temporal regularisation of measured values is required in order to obtain:

- quarterly,
- half-yearly, or
- yearly concentration data.

For each site the arithmetic mean of the concentration data shall be calculated. In case of only one measured value for each period, the temporal aggregation means simply an assignment of raw data to the corresponding time period.

Measurements below the LOQ will be replaced by a fixed percentage, usually by 50 % LOQ. As LOQ values may change over the time, there is a need of a consistent treatment of measurements below LOQ in order to avoid induced trend phenomena.

Spatial aggregation

The trend assessment procedure shall be based on the same aggregation methodology which is applied for the assessment of the quality status. Trend assessment is carried out with the mean values but without the confidence limits.

Trend estimation methods

Theil slopes (in case of a linear trend)

The Theil slope is a non-parametric, outlier-resistant estimation method for linear trends. It fails, however, in case of a non-linear trend.

Linear and quadratic regression

The classical approach to assess trends is to fit a linear or quadratic regression model. This approach however is not sufficiently flexible with regard to the shape of the trend.

LOESS smoother

The LOESS smoother applies the linear regression method only locally and is much more flexible with regard to the shape of trend.

Tests of Trends

Test of Mann-Kendall

The widely used test of Mann-Kendall is a non-parametric method for the detection of trends. It is outlier-resistant but only applicable in case of monotonic trends.

ANOVA test based on the LOESS smoother

The ANOVA-test (ANalysis Of VAriance) based on the LOESS smoother is a linear method and less outlier-resistant, but allows to investigate both monotonic and non-monotonic trends.

With regard to extensibility and power, the linear methods (based on a linear model) outperform non-parametric methods based on the test of Mann-Kendall

5.3.5 Trend reversal Assessment

Test of Mann-Kendall applied to subsequent differences

Applying the test of Mann-Kendall to the subsequent differences allows to assess if there is a significant trend in the slopes. If there is such a trend in the slopes, a trend reversal can be stated although there is no trend in the original data. However, it turned out that the power of the method, i. e. the sensitivity for detecting a trend reversal, is poor.

Test of the quadratic trend component based on a quadratic regression model

A classical approach for the detection of a trend reversal is to fit a quadratic regression model and to examine whether there is a maximum in the fit and whether this maximum lies in the investigated time interval. Example calculations show, however, that this method is not sufficiently flexible with regard to the type of trend and not sensitive enough to detect a trend.

Two-sections test (based on a linear model in two-sections)

The 2-sections model is a linear method, based on an extended linear regression model fitting a linear trend with one break in the interval. This test was derived from generalised linear regression theory. Its advantages are its simple interpretability, flexibility and high sensitivity to detect a trend reversal. Details are explained in the Annex.

5.4 PROPOSED METHODS AND PROCEDURE

5.4.1 Monitoring Network

Due to the proposed application of the upper confidence limit of the mean value, the minimum number of sampling sites per GW-body is 3. In case of sub-bodies it is proposed to monitor at least 1 site per sub-body. If, however, the number of sub-bodies exceeds analytical resources, monitored sub-bodies should be selected by a real random process. This random selection process should be well documented.

Minimum number of sites

- = 3 sites per GW-body
- = 1 site per sub-body

Network criterion

It was agreed that homogeneity (reflecting spatial representativity) of the network was a prerequisite and should be ensured to allow for sound statistical assessment in accordance with the requirements of the WFD.

Network criterion

- Representativity Index = 80 %

The Representativity Index was developed as a tool for assessing the homogeneity of the network. A certain degree of homogeneity of the network is a statistical prerequisite for the admissibility of applying the arithmetic mean as aggregation method as proposed.

To assess the homogeneity of a monitoring network, i.e. the homogenous coverage of the whole GW-body area, the average minimum distance between any location in the area to the closest sampling site is to be calculated and expressed as percentage of the average minimum distance for an optimal network. For a theoretical network with an optimal triangular pattern of sites the Representativity Index will be 100 %. For sub-optimal (less homogeneous) networks the index will decrease. For considering a network to be homogeneous the Representativity Index should be 80 % or higher. A value of 80 % means that the average minimum distance is 25 % larger than it would be for an optimal network

The proposed algorithm is also applicable for GW-bodies with very few sampling sites and elongated size.

From the statistical point of view it is proposed to achieve a value of at least 80 % as with a value of less than 80 % the estimated spatial mean can be highly biased. It has to be pointed out that the 80 % limit is a compromise to allow for sound estimations on the one hand and to take into account practicable requirements on networks on the other hand.

If the GW-body is **hydrogeologically heterogeneous** and if a spatially homogeneous monitoring network is not feasible or sensible the monitoring has to be developed in a hydrogeologically representative way and the spatial mean should be estimated with identical weights (AM).

5.4.2 Treatment of LOQ values

For the treatment of "lower than LOQ" measurements a minimax approach (minimize maximum risk) is applied: In order to avoid artificial trends or large biases, it is recommended to assess the chemical status and to calculate trends based on AM50 (50 means that <LOQ values are replaced by 50 % LOQ) as long as AM0/AM100 = 0.6. Under these circumstances the maximum bias does not exceed 25 %. If a limit value (LV) is available, the LOQ should not exceed 60 % of the LV.

As a rule for **risk assessment** it is proposed to assess the risk (**surveillance monitoring**) only if AM0/AM100 = 0.6. However if AM0/AM100 < 0.6 and the CL95 based on the AM100 is below 75 %limit value it can be assumed that there is no risk. If the CL95 based on the AM100 exceeds 75 % of the limit value no conclusion can be drawn and it is recommended to repeat the assessment in the following year.

As a rule for **status assessment** it is proposed to assess the chemical status (**operational monitoring**) only if AM0/AM100 = 0.6. However if AM0/AM100 < 0.6 and the CL95 based on the AM100 is below the limit value it can be assumed that there is good status. If the CL95 based on the AM100 exceeds the limit value no conclusion can be drawn and it is further recommended to repeat the assessment in the following year.

In general if AM0/AM100 < 0.6 any trend assessment should be based on sampling site level.

In order to avoid bias (induced trend phenomena), the trend analysis should be performed with a constant LOQ_{max}. All measurements (above or below LOQ) where the LOQ exceeds LOQ_{max} should be eliminated. LOQ_{max} is defined as the largest LOQ which does not exceed twice the median of all LOQs or half of the limit value (if available), i.e. $LOQ_{max} = \max\{LOQ;\ LOQ \leq 2 * \text{median}(LOQ)\ \text{or}\ LOQ \leq 0.5 * LV\}$. The calculation of the median of the LOQs is based on the whole data set under consideration. All further trend calculations have to be performed with LOQ_{max}, considering only stations where the LOQs do not exceed LOQ_{max}.

Furthermore, it is proposed to perform the trend analysis (based on AM50) only if the average of AM0/AM100 (calculated for each year) is = 0.6.

LOQ values

- replacement of LOQ values with 50 % LOQ
- AM0/AM100 = 0.6: risk- and status assessment without restrictions
- AM0/AM100 < 0.6: limited statements possible!
- average of AM0/AM100 = 0.6, otherwise trend assessment is not possible
- LOQ = 60 % limit value
- trend analysis with constant LOQ_{max}
- omit all measurements where LOQ exceeds LOQ_{max}

5.4.3 Data Aggregation

It is proposed to assess the chemical status of a GW-body only if AM0/AM100 = 0.6. Otherwise the procedure as described in chapter 5.4.2 can be applied.

Regularisation

For each site the arithmetic mean of the concentration data shall be calculated. This provision for determining groundwater chemical status is laid down in Annex V section 2.4.5 of the WFD.

Measurements below the LOQ are replaced by 50 % LOQ.

Regularisation

- For each site: arithmetic mean of concentration data.
- Replacement of LOQ values with 50 % LOQ

Estimation of the spatial mean with required confidence

For the calculation of a spatial mean value a pragmatic method is proposed.

In principle the selected aggregation method is the arithmetic mean (AM) and its 95 % upper confidence limit (CL_{AM}). Since under certain conditions (depending on the monitoring network, the GW-body characteristics etc.) the calculation of the AM is not applicable from a statistical point of view, the calculation of a weighted arithmetic mean and its CL95 considering different GW-sub-bodies might be necessary. In this case the spatial mean is calculated as a weighted arithmetic mean (wAM) and its CL_{wAM}. In case of an exceeding of the limit value by the CL95 of the (w)AM it is regarded as permissible to verify the result by calculating an arithmetic mean weighted with regard to the area represented by the particular sampling site (kriging mean (KM)) and its CL_{KM} for the estimation of the spatial mean.

In order to guarantee the required level of confidence for GW-bodies with only a few stations the agreed proposal is to use the upper confidence limit (with a significance level alpha of 5 %) of the (weighted) aggregated mean (respectively the kriging mean) instead of the mean values itself. The calculation of the CL is performed under the normal distribution assumption (details see Annex 4).

The **upper confidence limit** depends on the variability of the concentration level within the GW-body and on the number of stations. CL decreases with an increasing number of stations within the GW-body or a decreasing variability of concentration levels. The use of the CL allows to reduce the number of stations in GW-bodies with levels far below the limit value, and enforces a higher number of stations in GW-bodies with levels close to the limit value. To some extent it is therefore in the hands of the monitoring manager whether the CL will be below or above the limit value thus allowing an effective allocation of analytical resources.

Spatial Aggregation

Arithmetic Mean (AM) and CL_{AM}

- GW-body or sub-body hydrogeologically homogeneous and Representativity Index = 80 %
- GW-body or sub-body hydrogeologically heterogeneous and monitoring network hydrogeologically representative

Weighted Arithmetic Mean (wAM) and CL_{wAM}

- GW-body consists of several sub-bodies

Kriging Mean (KM) and CL_{KM}

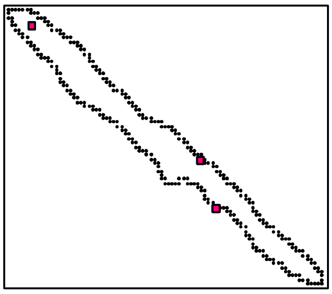
- alternatively, if an action limit is exceeded

Example

The comparison of results of two selected test GW-bodies illustrates the influence of the number of sampling sites on the ratio CL to AM. One GW-body is monitored by 3 sampling sites, one by 92 sampling sites. Due to the small number of sampling sites for GW-body PTA2 the CL_{AM} is 40–150 % higher than the AM. For GW-body AT224 the higher number of sites is the reason why the CL is just 10–20 % above the AM.

GW-body	PTA2
Parameter	**nitrate**
Area (km²):	7.7
Minimum:	6.20
Maximum:	19.00
No. of values:	40
No. of subst:	0
No. of sites :	3

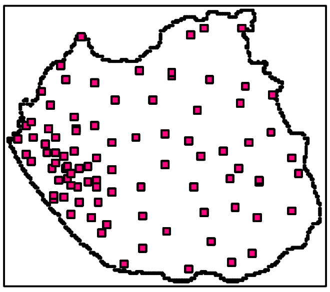
Body & stations



Begin	AM	CL_{AM}	CL_{AM}/AM
01.01.99	9.96	13.96	1.4
01.04.99	10.38	15.08	1.5
01.07.99	10.28	15.32	1.5
01.10.99	11.69	18.22	1.6
01.01.00	10.50	15.69	1.5
01.04.00	10.13	25.44	2.5
01.07.00	14.87	33.08	2.2

GW-body	AT224
Parameter	**nitrate**
Area (km²):	1018
Minimum:	0.062
Maximum:	206.00
No. of values:	1429
No. of subst:	85
No. of sites :	92

Body & stations



Begin	AM	CLAM	CL_{AM}/AM
01.01.95	52.84	61.41	1.2
01.01.96	57.45	65.61	1.1
01.01.97	58.77	67.58	1.1
01.01.98	55.41	63.71	1.1
01.01.99	52.58	59.68	1.1

For a GW-body which can be divided into several homogenous sub-bodies The **weighted arithmetic mean** (wAM) shall be calculated, taking into account the share of the sub-bodies and the corresponding AM.

In order to be consistent with the AM, the calculation of the CL_{wAM} is performed under the same model assumption as for CL_{AM} , i.e. all measurements are assumed to be stochastically independent and homogeneously distributed (in case of an even distribution of sites both the AM and the wAM could be applied, and the results should not differ). If the sites are evenly distributed, straightforward calculation shows that $CL_{wAM} = CL_{AM}$.

The confidence limit of AM or wAM does not take into account spatial correlation within the GW-body. In case of a sufficiently even distribution sampling sites the inclusion of spatial correlation may improve the statistical model and may lead to a reduction of the CL. However, it should be noted that the calculation of the spatial mean with the **Kriging** method must not be seen as a tool for modelling the regional level of concentration of pollutants in the GW-body. The Kriging method is solely applied to correct for uneven sampling site distribution, but not to estimate concentration levels for a GW-body to create a spatial map.

Proposed procedure for the estimation of a spatial mean

The proposed procedure for calculating a spatial mean is illustrated by a flow chart (see Figure 2) and can be described as follows:

- Check whether the GW-body consists of several sub-bodies with different sampling site densities,
  - if no, examine the monitoring network with regard to the network criterion (Representativity Index),
  - if yes, examine the monitoring networks within sub-bodies with regard to the network criterion.
- If the network criterion for the monitoring network(s) is not fulfilled, the monitoring network has to be adapted accordingly or the GW-body has to be subdivided into sub-bodies which fulfil the network criterion.
- If the GW-body or the sub-body is hydrogeologically heterogeneous and if a spatially homogenous network is not feasible or sensible, a hydrogeologically representative monitoring network has to be developed, and the spatial mean should be estimated with identical weights (AM).
- Use AM or the weighted AM (in case of several sub-bodies) to estimate the spatial mean (pragmatic approach).
- If the action limit is exceeded by CL_{AM}, it is permissible to calculate CL_{KM} alternatively (which can be considerably smaller in case of spatial correlation and high variability of the concentration level).

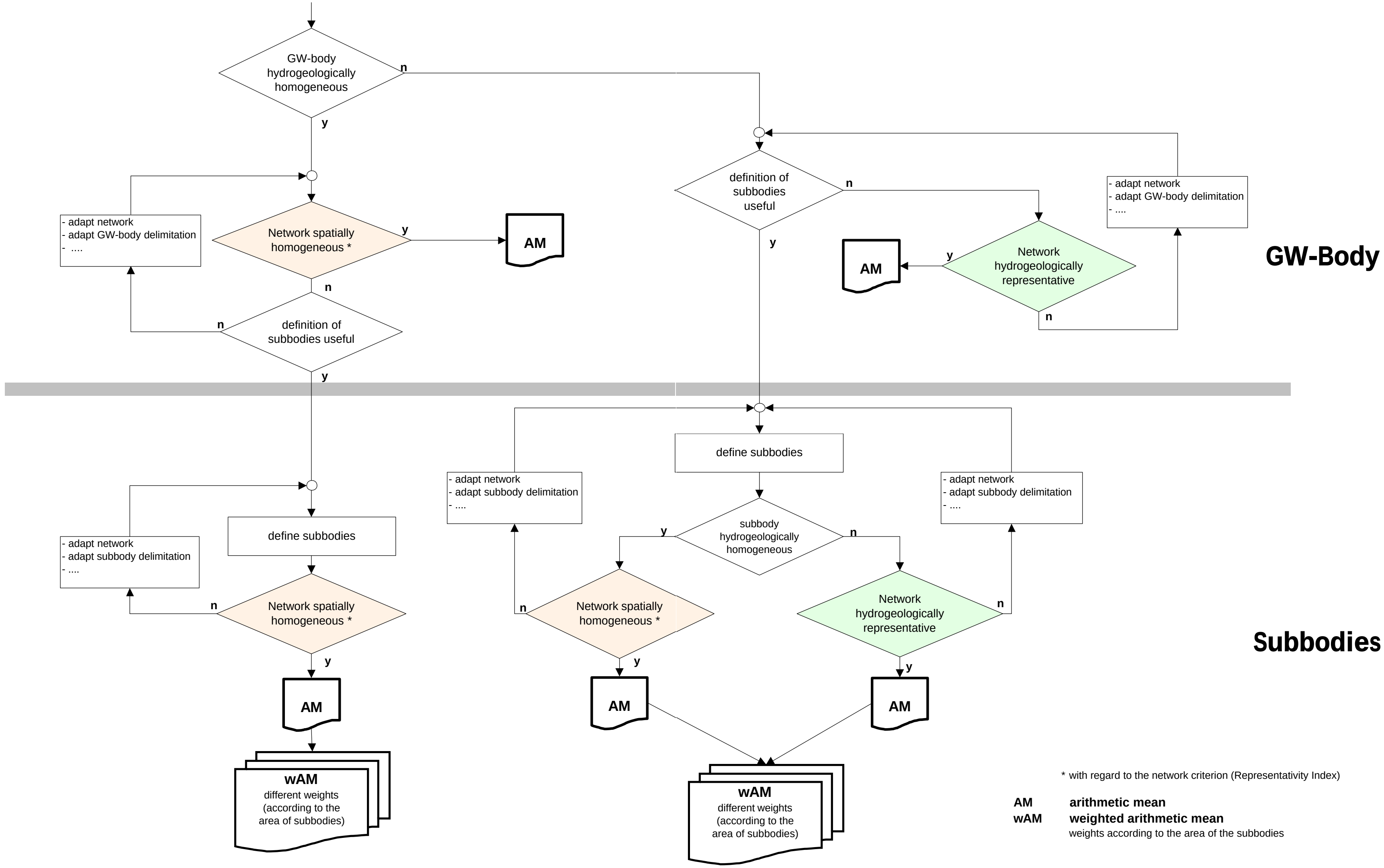


Figure 2: Proposed procedure for the estimation of a spatial mean with required confidence

5.4.4 Trend assessment

Trend analysis should be based on aggregated data from the whole GW-body (WFD, Annex V). Data aggregation for trend assessment consists of the same procedures (regularisation and spatial aggregation) as for quality status assessment. The following procedure is proposed:

Regularisation

For trend assessment provisions for the temporal regularisation of measurement values correspond to the aggregation of data for status assessment. Furthermore, it is possible to obtain:

- quarterly,
- half-yearly, or
- yearly concentration data.

For each site and for each aggregation period the arithmetic mean of the concentration data will be calculated.

Measurements below the LOQ are replaced by 50 % LOQ.

As LOQ values may change over the time, there is a need of a consistent treatment of measurements (where the LOQ exceeds a given LOQ_{max}) in order to avoid induced trend phenomena. Provisions on the calculation of a constant LOQ_{max} and treatment of measurements where the LOQs exceed the minimum requirements are laid down in chapter 5.4.2.

Regularisation

- for each site: arithmetic mean of concentration data.
- quarterly, half-yearly or yearly aggregation
- replacement of LOQ values by 50 % LOQ
- trend analysis with constant LOQ_{max}
- omit all measurements where LOQ exceeds LOQ_{max}

Spatial aggregation

The trend assessment procedure shall be based on the same aggregation methodology which is applied for the assessment of the quality status as already outlined in chapter 5.4.3.

Trend assessment is based on the mean value and not on the confidence limit.

Furthermore, it is proposed to perform the trend analysis (based on AM50) only if the average of AM0/AM100 (calculated for each year) is = 0.6.

Spatial aggregation

- calculation as outlined under Data Aggregation (chapter 5.4.3)
- average of AM0/AM100 = 0.6, otherwise trend assessment is not possible

Generalised linear regression test (ANOVA test) based on the LOESS smoother

With regard to extensibility and power the linear methods (based on a linear model) outperform non-parametric methods based on the test of Mann-Kendall, and therefore the decision was in favour of the linear methods. The consequence was a decision for the generalised linear regression test (ANOVA) for the assessment of monotonic trends. For the assessment of a trend reversal, the consequence was a decision for the two sections model, due to its simple interpretability.

Trend assessment is based on the mean value and not on the confidence limit.

Trend assessment

- ANOVA test based on the LOESS smoother
- assessment based on mean value and NOT on CL

Starting point

It was considered as important to detect an increase in pollutant concentration of 30 % with a power of 90 % or higher. The starting point for trend assessment is the same as for operational monitoring and shall allow for an "early warning function" of the trend detection Both starting points are defined as follows: if CL95 > 75 % LV the GW-body is identified as being at risk to fail the objectives in Art. 4 of the WFD, hence operational monitoring has to be carried out to have a sound data basis for the assessment of the status of the GW-body and for the trend assessment.

Starting point

- CL95 > 75 % limit value

Minimum length of time series for the detection of an upward trend

For the minimum length of time series for trend detection the power of trend detection, the timetable for WFD implementation as well as the minimum requirement regarding the monitoring frequency, which is once a year, were taken into account. As operational monitoring is likely to start in 2007 and in 2015 a review and update of river basin management plans is required, it is assumed that in 2015 data from 2007 to 2014 are available. This means a time series of 8 years with at least 8 values as a minimum.

As with less than 8 yearly measurements a statistical trend examination may be critical, it is recommended to perform a trend analysis with at least 8 measurements. In case of half-yearly measurements the total sample number should not be less than 10, in case of quarterly measurements it should not be less than 15. In each case the time span of measurements should be at least 5 years since short-term changes may distort the detection of long-term trends.

In case of yearly measurements it should be guaranteed that the measurements are undertaken in one and the same quarter or within a certain time period of the year. This is required to avoid a high random variation which reduces the power of the trend analysis. In the time series some observations may be missing, but the missing of two or more subsequent values should be avoided, as this would cause a bias due to extrapolation.

Minimum length of time series

- Annual data:
  = 8 years and = 8 values
- Half-yearly data:
  = 5 years and =10 values
- Quarterly data:
  = 5 years and =15 values

Example of admissible sample patterns

year	quarter	measurements		
		yearly	half-yearly	quarterly
2007	1	x	x	x
	2			x
	3		x	x
	4			x
2008	1	x	x	x
	2			x
	3		x	x
	4			x
2009	1	x	x	x
	2			x
	3		x	x
	4			x
2010	1	x	x	x
	2			x
	3		x	x
	4			x
2011	1	x	x	x
	2			x
	3		x	x
	4			
2012	1	x		
	2			
	3			
	4			
2013	1	x		
	2			
	3			
	4			

Maximum length of time series for the detection of an upward trend

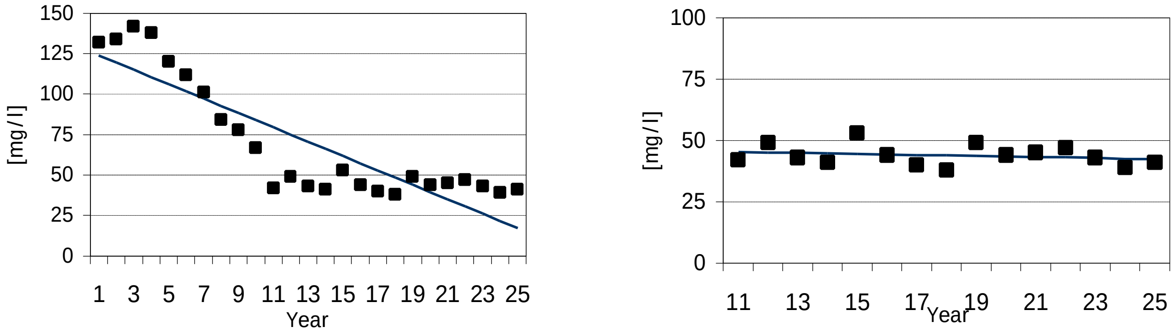
If long-term time series are assessed, there is a risk of obtaining trend results which are clearly affected by changes in the earlier years of the time series. For trend assessment purposes it is therefore proposed to restrict the time span of trend assessment to the last 15 years.

An alternative would be the application of an adaptive method to check whether there is a significant break of the (linear) trend (e.g. by a trend reversal method (two-section method)). If there is a significant break then the second section should be subject of the trend examination, otherwise the start year of the whole time series should be taken as starting point.

Maximum length of time series

- 15 years

Figure 3: Example - Influence of the length of time series on the detection of a trend



Frequency of trend assessment

Due to the problem of sequential or multiple testing it is recommended to restrict trend assessments to a maximum of three tests, after 8 years, 14 years and 20 years. More details and explanations on the bias and errors caused by multiple testing can be found in the Annex

Frequency of trend testing

- after 8 years
- after 14 years
- after 20 years

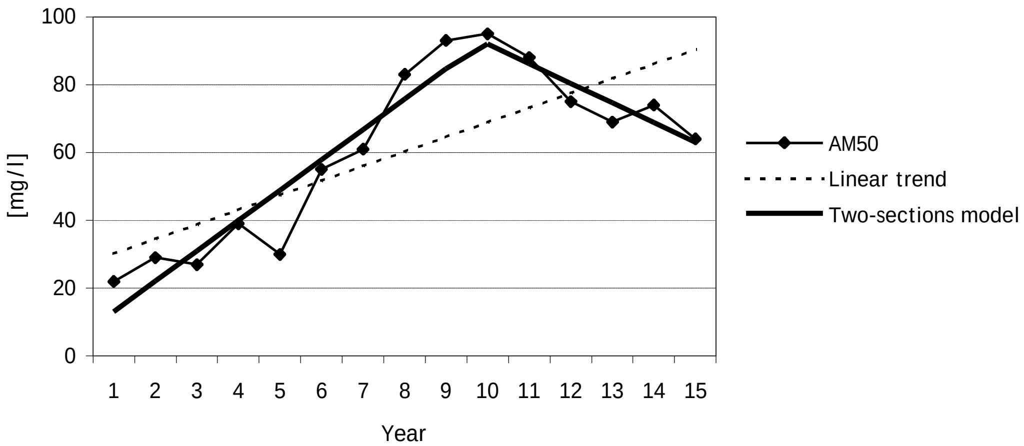
5.4.5 Trend Reversal Assessment

Two-sections test (based on a linear model in two-sections)

The 2-sections model is a linear method, based on an extended linear regression model fitting a linear trend with one break in the interval. It is preferred due to its simple interpretability, flexibility and high sensitivity to detect a trend reversal.

Trend reversal assessment

- Two sections model



Starting point for the detection of a trend reversal

Trend reversal may be the case if in the penultimate period an upward trend was detected and in the last period no more upward trend could be identified. Therefore, the starting point of the trend reversal assessment is to be equal to the starting point of the penultimate period of a trend assessment.

Starting point

- no upward trend in the last period but detection of an upward trend in the penultimate period
- starting point equal to the starting point of the penultimate period of a trend assessment

Minimum length of time series for the detection of a trend reversal

The procedure for the estimation of the minimum length of time series for trend reversal assessment was similar to the one described for trend assessment For the minimum length of time series for trend detection the power of trend detection, the timetable for WFD implementation as well as the minimum requirement regarding the monitoring frequency, which is once a year, were taken into account. As operational monitoring is likely to start in 2007 and in 2021 the second review and update of river basin management plans is required, it is assumed that in 2021 data from 2007 to 2020 are available. This means a time series of 14 years with at least 14 values as a minimum.

With yearly measurements, at least 14 measurements are recommended in order to guarantee a certain level of power for the detection of a trend reversal. With data on a half-yearly or quarterly basis 10 years are considered to be the minimum. For half-yearly measurements at least 18 values and for measurements on a quarterly basis at least 30 values would be necessary.

Minimum length of time series

- Annual data:
  = 14 years and = 14 values
- Half-yearly data:
  = 10 years and =18 values
- Quarterly data:
  = 10 years and =30 values

Maximum length of time series for the detection of a trend reversal

It is not recommended to apply the trend reversal test to a time span of more than 30 years. If even after 30 years no trend reversal can be detected, this has to be stated clearly.

The risk of obtaining such a result depends on e.g.

- the effectiveness of the pollution reduction measures taken,
- the effectiveness of the monitoring design (site selection and frequency of sampling),
- analytical quality of measurements,
- retention time, i.e. the (minimum) time a chemical needs to pass the unsaturated zone and to reach the groundwater.

Maximum length of time series

- 30 years

5.5 PROCEDURE OF IMPLEMENTATION

5.5.1 Status assessment

Monitoring is most likely to start with surveillance monitoring for *bodies identified as being at risk following the characterisation exercise undertaken in accordance with Annex II* and for *bodies which cross a Member State boundary* (Annex V, 2.4.2).

Based on the impact assessment of the groundwater body (Annex II) and on the assessment of monitoring results (of the surveillance monitoring) it has to be decided whether operational monitoring has to be carried out or not. *Operational monitoring shall be carried out for all those GW-bodies or groups of bodies which on the basis of both the impact assessment carried out in accordance with Annex II and surveillance monitoring are identified as being at risk of failing to meet objectives under Article 4* (Annex V, 2.4.3). *The results of this program* (surveillance monitoring) *shall be used to install an operational monitoring* (Annex V, 2.4.1).

Derived from the data requirements for trend assessment it is proposed to identify a GW-body as being at risk in terms of groundwater quality (and start operational monitoring) if the CL95 – which is the relevant figure for the assessment of GW-quality status – exceeds 75 % of the limit value (see Article 17 (5)).

Art. 14 (1) b requires *an interim overview of the significant water management issues identified in the river basin, at least two years before the beginning of the period to which the plan refers.* This means that in 2007 the first interim overview is required. As monitoring activities are likely to start in 2006 as surveillance monitoring and operational monitoring will start in 2007 it will not be possible to include status assessments from this year in the overview as sufficient data will not be available (for status assessment data from operational monitoring would be required). The first river basin management plan has to be produced in 2008 as a draft version for discussion and involvement of interested parties. Its publication is required in 2009. At this time monitoring data for an assessment of the groundwater quality status will (most probably) be available for the investigations in 2007 (operational monitoring).

In 2015 a review and update of the river basin management plan is required. Risk assessment will (most probably) refer to data from 2012 (surveillance or operational monitoring). If the CL95 (still) exceeds 75 % of the limit value, operational monitoring has to be started (continued). Status assessment for the river basin management plan in 2015 can then be based on data from operational monitoring in 2014.

5.5.2 Trend assessment

Starting point

The starting point for trend assessment is the same as for operational monitoring. Both are defined as follows: if the CL95 exceeds 75 % of the limit value the GW-body is identified as being at risk to fail the objectives in Art. 4 of the WFD, hence operational monitoring has to be carried out to have a sound data basis for the assessment of the status of the GW-body and for the trend assessment.

One of the findings during the data assessment phase was that a significant upward trend is to be detected with a power of 90 % (for most substances) if the increase in pollutant concentration is at least 30 % or even higher, depending on the type of pollutant. For that reason it is proposed to start operational monitoring if the CL95 exceeds 75 % of the limit value (Article 17(5)). In this case an increase of 33 % of the CL95 in pollutant concentration would mean that good status is failed. In this connection the importance of data from operational monitoring for trend assessment has to be pointed out, as otherwise insufficient data would not allow for trend calculations and an increase in pollutant concentration endangering good status could not be detected.

Length of time series

The first river basin management plan has to be produced in 2008 as a draft version for discussion and for the involvement of interested parties. Its publication is required in 2009. At this time monitoring data will (most probably) be available from investigations in 2006 (surveillance) and 2007 (operational monitoring), which is too short for a trend assessment.

In 2015 a review and update of the river basin management plan is required. It has to be assessed whether CL95 is still exceeding 75% of the limit value (the GW-body is at risk from the quality point of view) – if no, depending on the GW-body characterisation, monitoring can be reduced to surveillance monitoring, if yes operational monitoring has to be continued and trend calculation can be performed and tested for an upward trend.

Considering the time required for analysis, data control, data flow, data assessment, review and discussion of the management plan etc. and considering that the data from the surveillance monitoring cannot be included in the trend assessment if the number of monitoring sites of the surveillance monitoring is different from (lower than) that of the operational monitoring, it is assumed that in 2015 data for the time period from 2007 to 2014 are available. This means a time series of eight years with at least eight values. Trend calculations are carried out based on **mean values** for GW-bodies and **not on the CL95**!

The second review of river basin management plans is due in 2021. As described above, it has to be assessed whether CL95 is still exceeding 75% of the limit value – if not, depending on the GW-body characterisation, monitoring can be reduced to surveillance monitoring, if yes operational monitoring has to be continued and trend calculation can be performed and tested for an upward trend. If there is an upward trend in the first period and no upward trend in the subsequent period, data can be assessed for trend reversal. In order to test for trend reversal, data of the time period 2007 to 2020 have to be taken into account, which means a time series of 14 years with at least 14 annual values.

Figure 4: ***Proposed procedure for the implementation of the WFD regarding status assessment, trend and trend reversal assessment*** *(years refer to the first time it is required. In case of loops this information has to be neglected)*

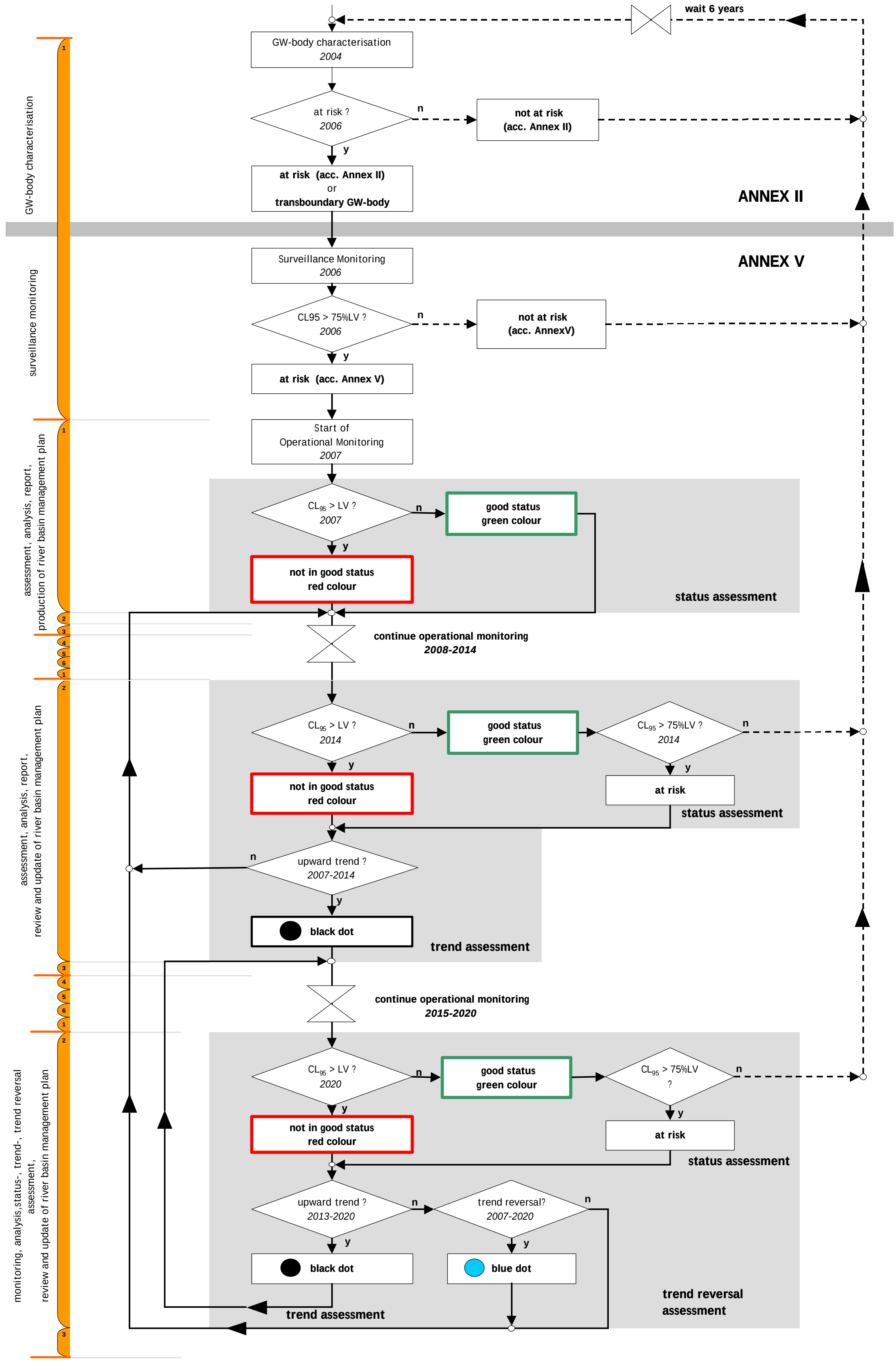
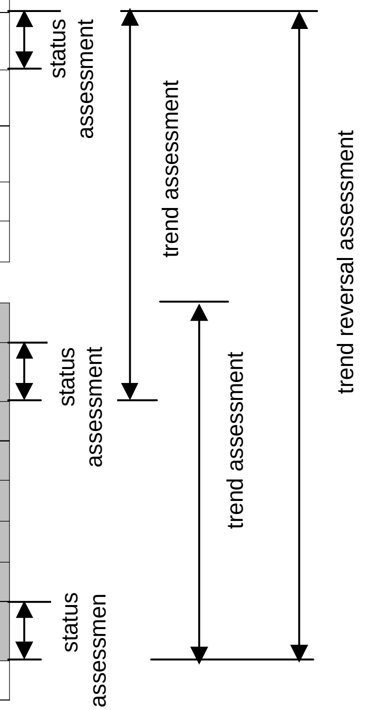


Figure 5: Timetable on implementation of the WFD - with regard to GW-Quality monitoring and data assessment

			GW-quality monitoring activity						
year	WFD criteria (key words)	relevant Article or Annex	SM	OM	OM	OM			
2000	WFD set into force								
2001									
2002	criteria for the assessment of good status, trend and trend reversal (Commission proposal)	Art. 17(2)a, b							
2003									
2004	description of GW-bodies, human impacts etc.	Art. 5(1), Annex II							
2005									
2006	establishment of monitoring programmes	Art. 8, Annex V	X						
2007	interim overview of the significant water management issues	Art. 14 (1) b		X					
2008	production of river basin management plans - draft (involvement of interested parties)	Art. 14(1)a, c		X					
2009	programme of measures; publication of river basin management plan	Art. 11(7); Art. 13(6)		X					
2010				X					
2011				X					
2012	programmme of measures operational	Art. 11(7)	X	X					
2013					X				
2014					X				
2015	review and update of river basin management plan	Art. 13(7)			X				
2016					X				
2017					X				
2018			X		X				
2019						X			
2020						X			
2021	review and update of river basin management plan	Art. 13(7)				X			



SM surveillance monitoring
OM operational monitoring

X if CL95 > 75% of Limit Value => operational monitoring
X if CL95 <= 75% of Limit Value => surveillance monitoring

5.6 REMARKS AND RECOMMENDATIONS

Monitoring network, improving the Representativity Index

Monitoring networks with lower sampling site densities might fulfil the requirements of the Representativity Index (as a minimum requirement) easier than networks with high site densities. It has to be considered, however, that with fewer sampling sites the confidence limit will increase considerably, which as a consequence may require an earlier start of action (e.g. operational monitoring).

By establishing additional sampling sites at appropriate locations of a monitoring network it could be easier to achieve an improvement of the Representativity Index than by the closing of sites.

Status assessment

If the CL_{AM} is close to or above a limit value or an action value an increase of the number of sampling stations might lead to a considerable decrease of the CL and hence avoid a need of action.

In case of sufficiently evenly distributed sampling sites the inclusion of spatial correlation (calculation of a Kriging mean and CL_{KM}) may improve the statistical model and may lead to a reduction of the CL

Trend assessment

In order to allow for trend assessment at an earlier stage it is recommended to assess trends on the basis of quarterly data. This would enable to also set actions to achieve good status earlier.

If the amount of values below the LOQ is that large that the average of AM0/AM100 falls below 0.6, trend assessment is not possible. In that case it is recommended to analyse trends separately for each sampling site.

More sophisticated treatment of non-linear trends (Two-sections model)

The procedure proposed for the assessment of an upward trend does not perform satisfyingly in case of a sudden change in the trend slope (e. g. due to earlier pollution reduction measures). Prior to the examination of an upward trend it is therefore recommended to add the following steps :

1. Check whether there is a significant break of the (linear) trend, e. g. a trend reversal.
2. If yes, use the start year of the second section for the trend examination. If not, use the whole time series.

Note

As already mentioned the scheme is based on the assumption that monitoring systems will be operational in 2006 (as required in the WFD). In case of monitoring systems that have already existed before this time available data can be assessed by the proposed methods if they are in line with the requirements described above.

6 ALGORITHM AND COMPUTATION

In this chapter all algorithms of the methods of the proposed procedure are presented. These algorithms which allow for implementation in calculation tools. Furthermore, several examples are provided for further clarification.

6.1 NETWORK CRITERION

In order to allow for sound statistical assessment in accordance with the requirements of the WFD and especially with respect to the estimation of the spatial mean, homogeneity of the network is required. Homogeneity is measured by the Representativity Index $\mathbf{R_U}$, which represents the average minimum distance between any location in the area to the closest sampling site, expressed in percentages of the average distance for an optimal network (inverse presentation). $\mathbf{R_U}$ depends on the number of sites, denoted by *k*, on the average minimum distance between any location in the area to the closest sampling site, $dist_{ave}$, and on the size of the area, denoted by *Area*.

$$R_U = \frac{37.7}{dist_{ave}\sqrt{k / Area}} \text{ [\%]}$$

For a theoretical network with an optimal triangular pattern of sites the Representativity Index will be 100 %. For sub-optimal (less homogeneous) networks the index will decrease. For considering a network to be homogeneous the Representativity Index should be 80 % or higher. A value of 80 % means that the average minimum distance is 25 % larger than it would be for an optimal network. With a value of less than 80% the estimated spatial mean can be highly biased. It should be noted that this limit is result of a compromise between the aim to get an optimal estimation and practical requirements of networks.

If – for ground water bodies with few sampling sites - the squared maximum length of the groundwater body, $length^2$ is larger than the product of the size of the area, *Area*, and the number *k* of sites, i.e. if $length^2 > k \times Area$, then $\mathbf{R_U}$ can be replaced by

$$R_S = \frac{25\sqrt{\left(\frac{Area}{length}\right)^2 + \left(\frac{length}{k}\right)^2}}{dist_{ave}} \text{ [\%].}$$

6.2 TREATMENT OF LOQ VALUES

6.2.1 Calculation of LOQ_{max} (requirement of trend analysis)

In order to avoid bias (induced trend phenomena), the trend analysis should be performed with a constant LOQ_{max}. All measurements (above or below LOQ) where the LOQ exceeds LOQ_{max} should be eliminated.

LOQ_{max} is defined as the largest LOQ which does not exceed twice the median of all LOQs or half of the limit value (if available), i.e. $LOQ_{max} = \max\{LOQ;\ LOQ \leq 2 * \text{median}(LOQ) \text{ or } LOQ \leq 0.5 * LV\}$.

The calculation of the median of the LOQs is based on the whole data set under consideration. All further trend calculations have to be performed with LOQ_{max}, considering only stations where the LOQs do not exceed LOQ_{max}.

Note 1: If due to varying analytical quality the number of years or the number of concentration mean values entering the trend analysis is below the minimum requirements, the timeframe of operational monitoring has to be prolonged accordingly.

Note 2: If both the LOD and the LOQ are available the proposed procedure should be modified to include both values.

Example

A data set with 3 measurements below LOQ and a limit value (LV) of 100:

Year	Measurement	LOQ
1	<LOQ	100
1	122	100
1	145	100
2	110	50
2	<LOQ	50
2	89	50
3	72	30
3	65	30
3	80	30
4	44	30
4	<LOQ	30
4	59	30
5	65	30

(1) Calculate: median(LOQ) = 30

(2) Calculate: 2*median(LOQ) = 60 and 0.5*LV = 50

(3) The largest LOQ not exceeding 60 is 50. Hence LOQ_{max} = 50

Year	Measurement	LOQ
2	110	50
2	<LOQ	50
2	89	50
3	72	50
3	65	50
3	80	50
4	<LOQ	50
4	<LOQ	50
4	59	50
5	65	50

(4) Eliminate stations with LOQ > LOQ_{max} = 50 (stations with LOQ = 100), and replace LOQ < LOQ_{max} by LOQ_{max} (= 50)

Year	AM0	AM100	AM0/AM100
2	66.33	83.00	0.80
3	72.33	72.33	1.00
4	19.67	53.00	0.37
5	65.00	65.00	1.00

(5) Calculate: AM0/AM100 with LOQ = LOQ_{max} for each year

Year	AM50
2	74.67
3	72.33
4	36.33
5	65.00

(6) The average of AM0/AM100 is 0.72, which is above 0.6. Hence calculate AM50 for each year and perform trend analysis based on AM50

6.3 DATA AGGREGATION

In this section the results of all sites within a given water body for a fixed period *j* of year *i* are considered.

6.3.1 Regularisation - Calculation of AM50

Let *i* denote the year and *j* the period within the year (*j*=1,2,3,4 for quarterly data, *j*=1,2 for half year periods and *j*=1 for annual data). Then the lower and the upper mean at site *s*=1,…, *n* are calculated

$$\text{AM0} = x_{ijs}^{\min} = \frac{1}{n_{ijs} + p_{ijs}} \sum_t m_{ts}$$

$$\text{AM100} = x_{ijs}^{\max} = \frac{1}{n_{ijs} + p_{ijs}} \left(\sum_t m_{ts} + \sum_t l_{ts} \right),$$

where the sum is taken over all measurements at site *s* within period *j* of year *i*. n_{ijs} denotes the number of measured values m_{ts} at site *s* within period *j* of year *i*, and p_{ijs} denotes the respective number of measurements below the substitution value l_{ts} (LOQ limit of quantification).

If there is only one measurement in each period, these mean values at site *s* can be calculated

$$x_{ijs}^{\min} = \begin{cases} m_{ts} & \text{if measurement value available} \\ 0 & \text{if less than LOQ} \end{cases}$$

$$x_{ijs}^{\max} = \begin{cases} m_{ts} & \text{if measurement value available} \\ l_{ts} & \text{if less than LOQ} \end{cases}$$

Example

Values recorded		Values after substitution for <0.2	
<0.20	LOQ	**0.00**	**0.20**
<0.20	LOQ	**0.00**	**0.20**
<0.20	LOQ	**0.00**	**0.20**
0.22		0.22	0.22
0.25		0.25	0.25
0.29		0.29	0.29
0.31		0.31	0.31
0.42		0.42	0.42
0.54		0.54	0.54
	Total	2.03	2.63
	Mean	AM0 = 0.226	AM100 = 0.292

AM50 = (AM0 + AM100) / 2 **= 0.259**

6.3.2 Arithmetic Mean (AM)

The arithmetic mean within the water body for period *j* of year *i* is computed

$$\mathrm{AM} = \frac{1-w}{n}\sum_{s=1}^{n} x_{ijs}^{\min} + \frac{w}{n}\sum_{s=1}^{n} x_{ijs}^{\max}$$

where *w* denotes the weighting factor for the measurements below the detection limit or the determination limit. Typically, *w*=0.5, which means that measurements below LOQ are replaced by 0.5 x LOQ:

$$\mathrm{AM} = \frac{1-w}{n}\sum_{s=1}^{n}\left(\frac{1}{n_{ijs}+p_{ijs}}\sum_{t} m_{ts}\right) + \frac{w}{n}\sum_{s=1}^{n}\left(\frac{1}{n_{ijs}+p_{ijs}}\left(\sum_{t} m_{ts} + \sum_{t} l_{ts}\right)\right)$$

$$= \frac{1}{n}\sum_{s=1}^{n}\left(\frac{1}{n_{ijs}+p_{ijs}}\left(\sum_{t} m_{ts} + \sum_{t} w l_{ts}\right)\right).$$

In order to make the notation as simple as possible, AM50 denotes the arithmetic mean AM with *w*=0.5, AM0 with *w*=0, and AM100 with *w*=1.

6.3.3 Upper confidence limit of the arithmetic mean (CL_{AM})

Under the normal assumption the confidence limit for the arithmetic mean can be calculated

$$CL_{AM} = AM + t_{N-1,1-a/2}\, s/\sqrt{N}\ ,$$

where s denotes the standard deviation of the regularized site mean values

$$(1-w)x_{ij1}^{\min} + wx_{ij1}^{\max}, \ldots, (1-w)x_{ijn}^{\min} + wx_{ijn}^{\max}\ .$$

6.3.4 Weighted Arithmetic Mean (wAM)

$$wAM = \sum w_i AM_i$$

where w_i denotes the share of subbody *i*.

Example

sub-body	AM	part of GW-body	w * AM
a	31.97	0.09	2.83
b	24.42	0.15	3.59
c	21.25	0.14	3.10
d	42.92	0.08	3.56
e	36.54	0.15	5.36
f	54.65	0.10	5.56
g	33.50	0.09	3.01
h	75.71	0.09	6.48
i	18.44	0.07	1.31
j	53.31	0.04	2.18
		Total = wAM =	36.97

6.3.5 Confidence limit of the weighted Arithmetic Mean(CL_{wAM})

Assume that $Y_{ij} = N(m, s_0^2)$, let $AM_i = \frac{1}{n_i}\sum_j Y_{ij}$ and $wAM = \sum_i w_i AM_i$,

where w_i denotes the share of sub-body *i* and let *i=1,...,a*. The variance of wAM equals

$$\sum_i w_i^2 \frac{s_0^2}{n_i}.$$

In case of homogenous sampling site distribution weights are proportional to the number of sites, i.e. $w_i = n_i / N$. Then the variance can be written

$$\sum_i w_i^2 \frac{s_0^2}{n_i} = \sum_i \frac{n_i^2}{N^2}\frac{s_0^2}{n_i} = \frac{s_0^2}{N}.$$

This is the variance of AM, and hence with this model it is guaranteed that calculations are consistent with CL_{AM}.

s_0^2 can be estimated by the empirical variance of the whole data set, and the corresponding confidence limit equals

$$CL_{wAM} = wAM + t_{N-1,1-a/2}\, s\sqrt{\sum_i \frac{w_i^2}{n_i}}$$

6.4 TREND ASSESSMENT

6.4.1 LOESS smoother

Let y_i (i=1,…,n) denote the observation for period p_i . The smoother matrix S of the loess smoother is constructed from a series of weighted regressions constructed for each period p_t using any data that fall within a range $p_t \pm \Delta_t$. The weight of the observation y_i in the local regression for period p_t is determined by the distance between period p_i and period p_t ,

$$w_{ti} = \begin{cases} \left(1 - \left(\frac{|p_t - p_i|}{\Delta_t}\right)^3\right)^3 & \text{for } 0 \le \frac{|p_t - p_i|}{\Delta_t} \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

The width of Δ_t controls the amount of smoothing, and Nicholson and Fryer choose Δ_t to include data from a fixed span of periods, i.e.

$$\Delta_t = \max\left\{\frac{\text{span}+1}{2}, \text{span - min}\{\text{t - 1, n - t}\}\right\}$$

and, with a span of seven periods (span=7) ,

$$\Delta_t = \begin{cases} 4 & \text{for} \quad 4 \le t \le n-3 \\ 5 & \text{for} \quad t = 3 \text{ or } t = n-2 \\ 6 & \text{for} \quad t = 2 \text{ or } t = n-1 \\ 7 & \text{for} \quad t = 1 \text{ or } t = n \end{cases}.$$

Then, writing X for the design matrix of the simple linear regression model,

$$X = \begin{pmatrix} 1 & p_1 \\ 1 & p_2 \\ \vdots & \vdots \\ 1 & p_n \end{pmatrix},$$

Y for the vector of observations,

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and W_t for the corresponding diagonal matrix of weights,

$$W_t = \begin{pmatrix} w_{t1} & 0 & \cdots & 0 \\ 0 & w_{t2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_{tn} \end{pmatrix},$$

the value of the smoother in period p_t is given by

$$S_t^{'}Y ,$$

where $S_t^{'}$ is the t^{th} row of

$$X\left(X'W_t X\right)^{-1} X'W_t$$

In case of equidistant observations and with a span of 7 periods, the resulting smoother matrix equals

0.569	0.388	0.199	0.0278	-0.0766	-0.0822	-0.0253						
0.383	0.3	0.209	0.111	0.0246	-0.0178	-0.0104						
0.209	0.223	0.202	0.172	0.123	0.0595	0.0113						
0.0417	0.145	0.206	0.216	0.206	0.145	0.0417						
	0.0417	0.145	0.206	0.216	0.206	0.145	0.0417					
		0.0417	0.145	0.206	0.216	0.206	0.145	0.0417				
			0.0417	0.145	0.206	0.216	0.206	0.145	0.0417			
				.	.	.	.	.	.	.		
					0.0417	0.145	0.206	0.216	0.206	0.145	0.0417	
						0.0417	0.145	0.206	0.216	0.206	0.145	0.0417
						0.0113	0.0595	0.123	0.172	0.202	0.223	0.209
						-0.0104	-0.0178	0.0246	0.111	0.209	0.3	0.383
						-0.0253	-0.0822	-0.0766	0.0278	0.199	0.388	0.569

6.4.2 LOESS smoother with seasonality

For half-year and quarterly data, a LOESS smoother with seasonality can be obtained by extending the design matrix X for local regression with seasonal dummies s_t:

$$X = \begin{pmatrix} 1 & p_1 & s_1 \\ 1 & p_2 & s_2 \\ \vdots & \vdots & \vdots \\ 1 & p_n & s_n \end{pmatrix},$$

where in case of half year data

$$s_t = \begin{cases} 0 & t \text{ represents winter period} \\ 1 & t \text{ represents summer period} \end{cases}$$

and in case of quarterly data

$$s_t = \begin{cases} (0,0,0) & t \text{ represents winter} \\ (1,0,0) & t \text{ represents spring} \\ (0,1,0) & t \text{ represents summer} \\ (0,0,1) & t \text{ represents autumn} \end{cases}.$$

The other calculations are as in the preceding sub section.

6.4.3 ANOVA tests based on the LOESS smoother

Tests for the linear trend component, the quadratic trend component and for the systematic trend of the time series may be obtained with the sums of squares

$$SS_{total} = \sum_{t=1}^{n} \left(y_t - \bar{y}\right)^2$$

$$SS_{linerror} = \min_{a,b} \sum_{t=1}^{n} \left(y_t - (a + bx_t)\right)^2$$

$$SS_{quaderror} = \min_{a,b,c} \left\{ \sum_{t=1}^{n} \left(y_t - \left(a + bp_t + cp_t^2\right)\right)^2 \right\}$$

$$SS_{error} = \sum_{t=1}^{n} \left(y_t - z(p_t)\right)^2$$

with degrees of freedom

$$df_{total} = n - 1$$

$$df_{linerror} = n - 2$$

$$df_{quaderror} = n - 3$$

$$df_{error} = n - tr(2S - SS'),$$

the linear regression coefficients *a*, *b*, *c* and the smoother estimate $z(p_t)$. df_{error} is determined by the mean value of SS_{error} under the assumption of independent and normally distributed random deviations with variance σ²:

$$E[SS_{error} / s^{2}] = E\left[(y'(I-S)'(I-S)y)/s^{2}\right] = tr[(I-S)(I-S')]$$
$$= tr(I) - 2tr(S) + tr(SS'),$$

where *y* denotes the vector of observed data and *I* denotes the identity matrix. It should be noted that SS_{error} / s^{2} is not exactly chi-squared distributed, but its distribution can be approximated by a chi-squared distribution with df_{error} degrees of freedom.

The total sum of squares SS_{total} can be split into four components

$$SS_{total} = SS_{lin} + SS_{nonlin} + SS_{nonquad} + SS_{error}$$

where

$$SS_{lin} = SS_{total} - SS_{1inerror}$$

with $df_{total} - df_{nonlin} = 1$ degree of freedom and

$$SS_{quad} = SS_{linerror} - SS_{quaderror}$$

with $df_{nonlin} - df_{nonquad} = 1$ degree of freedom and

$$SS_{nonquad} = SS_{quaderror} - SS_{error}$$

with $df_{quaderror} - df_{error} = tr(2S - SS') - 3$ degrees of freedom.

These sums of squares can be used to perform approximate tests for the hypotheses whether or not there is a linear, a quadratic or a systematic trend.

Test for seasonality

Comparing the residual sums of squares of the LOESS smoother with and without seasonality, and approximate ANOVA test can also be applied to perform a test for seasonality.

6.5 TREND REVERSAL ASSESSMENT

6.5.1 Two-sections test

If it can be assumed that the time series can be explained by two linear trends with a change of the slope within the time interval, a two sections test can be applied. This test consists of three steps:

(1) Divide the time interval under consideration into two time sections and estimate corresponding regression lines.

(2) Optimise the choice of time sections with regard to the fit of the resulting model.

(3) Perform a statistical test to check whether the two-sections model is significantly better than simple linear regression model.

Model

The underlying model is as follows:

$$y_t = a_{s(t)} + b_{s(t)} p_t + error\ ,\ t=1,\ldots,n$$

where $s(t) = \begin{cases} 1 & p_t \le b \\ 2 & p_t > b \end{cases}$,

and b denotes the time of the break in the trend.

It is assumed that the trend line is continuous in the break and therefore the continuity condition holds:

$$a_1 + b_1 b = a_2 + b_2 b\ .$$

In order to avoid complicated regression analyses with additional restrictions, in matrix notation the model

will be re-parameterized and extended to non-equidistant time periods:

$$Y = X_b \mathbf{b} + e$$

where

$$Y=(Y_1,\ldots,Y_n)^T\ ,$$

$$X_b = \begin{pmatrix} 1 & p_1 - b & 0 \\ \vdots & \vdots & \vdots \\ 1 & p_k - b & 0 \\ 1 & 0 & p_{k+1} - b \\ \vdots & \vdots & \vdots \\ 1 & 0 & p_n - b \end{pmatrix},$$

$$p_k \le b \le p_{k+1}\ ,$$

$$4 \le k \le n-4 \quad \text{and}$$

$$\mathbf{b} = (b_1, b_2, b_3)^T.$$

Algorithm

The proposed algorithm is as follows.

Step A: Selection of time intervals

Select b out of $\{p_4,\ldots, p_{n-3}\}$ so that the sum of squared residuals

$$SS_{2sections} = \left(Y - X_b\left(X_b^T X_b\right)^{-1} X_b^T Y\right)^T \left(Y - X_b\left(X_b^T X_b\right)^{-1} X_b^T Y\right)$$

attains its minimum.

Step B: Examination of the significance of the break

Calculate the sum of squared residuals

$$SS_{lin} = \left(Y - X\left(X^T X\right)^{-1} X^T Y\right)^T \left(Y - X\left(X^T X\right)^{-1} X^T Y\right)$$

for the simple linear regression model

$$X = \begin{pmatrix} 1 & p_1 \\ 1 & p_2 \\ \vdots & \vdots \\ 1 & p_n \end{pmatrix}$$

and the test statistic

$$F = \frac{(SS_{lin} - SS_{2sections})/2}{SS_{2sections}/(n-4)}.$$

Under the null hypothesis that there is no break in the trend, the test statistic F can be considered approximately F-distributed with $(2,n\text{-}4)$ degrees of freedom.

Hence if F exceeds the 95% quantile of the F distribution with $(2,n\text{-}4)$ degrees of freedom, it can be concluded that there is a break in the trend.

There is a trend reversal, if in the first section the slope of the trend line is positive, and in the second section negative.

6.5.2 Trend reversal with seasonality

For half-yearly and quarterly data, the examination of a trend reversal with seasonality can be obtained by extending the design matrix X_b and the design matrix for a linear trend with seasonal dummies s_t:

$$X_b = \begin{pmatrix} 1 & p_1 - b & 0 & s_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & p_k - b & 0 & s_k \\ 1 & 0 & p_{k+1} - b & s_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & p_n - b & s_n \end{pmatrix} \text{ and}$$

$$X = \begin{pmatrix} 1 & p_1 & s_1 \\ 1 & p_2 & s_2 \\ \vdots & \vdots & \vdots \\ 1 & p_n & s_n \end{pmatrix}$$

where in case of half year data

$$s_t = \begin{cases} 0 & t \text{ represents winter period} \\ 1 & t \text{ represents summer period} \end{cases}$$

and in case of quarterly data

$$s_t = \begin{cases} (0,0,0) & t \text{ represents winter} \\ (1,0,0) & t \text{ represents spring} \\ (0,1,0) & t \text{ represents summer} \\ (0,0,1) & t \text{ represents autumn} \end{cases}.$$

The determination of the time intervals is as described in the preceding sub section.

The test statistic can be calculated

$$F = \frac{(SS_{lin} - SS_{2\sec tions})/2}{SS_{2\sec tions}/(n-m)}.$$

where *m*=5 for half-yearly data and

m=7 for quarterly data.

Under the null hypothesis that there is no break in the trend, the test statistic *F* can be considered approximately F-distributed with (2,*n*-m) degrees of freedom.

Hence if *F* exceeds the 95% quantile of the F distribution with (2,*n*-m) degrees of freedom, it can be concluded that there is a break in the trend. There is a trend reversal, if in the first section the slope of the trend line is positive, and in the second section negative.

6.6 COMPUTATION

The outcome of the project comprises the algorithm and a software tool (*GWstat*) for both the proposed procedure for data aggregation and trend/reversal assessment.

The software tool developed by the subcontractor "quo data" allows for data aggregation and trend/reversal assessment and includes also the calculation of the network criterion "Representativity Index".

In line with the contract this software tool is available free of charge. This version of the tool does not comprise a database, for analysis data have to be provided in different files.

For the project the comprehensive statistics package *WaterStat* was applied. This package was adapted to the needs of the project and has been in use for the calculation of test data sets

A further adaptation of the *WaterStat* software package with regard to database, data management, reports and report formats and utilities for the network design etc. is possible. This would allow for a harmonised procedure within the EU with regard to the assessment of the chemical status and the trend analyses. The adaptations of the software could be performed by "quo data" and provided at self costs if there is an expressed interest by institutions of the Member States.

7 LINKS TO OTHER WORKING GROUPS

The project is part of the "Common Strategy on the Implementation of the WFD", which was developed by the European Commission to achieve a common understanding and approach on WFD implementation. The working group of this project (Common Strategy on the Implementation of the WFD - Key activity 2: Development of guidance on technical issues, 2.8 Guidance on tools on assessment and classification of groundwater) is one of ten working groups initiated by the EC to develop guidance on specific issues of WFD implementation. Due to the integrated approach of the WFD, interaction between the working groups is required. The work of this group and the outcome of the project has to be seen as closely related to the work of the other working groups.

Links to following groups were identified:

- WG 2.7 *Develop guidance on monitoring*
- WG 2.1 *Guidance on the analysis of pressures and impacts*
- WG 3.1 *Develop a shared Geographical Information System*
- WG 4.1 *Integrated testing of guidelines in pilot river basins*
- *Expert Advisory Forum for Groundwater EAF–GW*

Within the project basic requirements for the WFD implementation process with regard to groundwater became apparent and should be considered by the other Working Groups in order to ensure a harmonised approach and the applicability of the proposed procedure.

Several topics on which guidance should be elaborated are listed below to be further discussed within the Working Groups. The topics are separated into "findings" of the project to be taken into account and "open questions" to be further discussed.

Delimitation of GW-bodies

Open questions

- There is an expressed and urgent need, which arose several times during the project, to find a common understanding and a common definition of the term "GW-body". As the further implementation of groundwater aspects of the WFD within MS starts with the delimitation of GW-bodies, guidance on this topic would be crucial, especially in view of comparability.
  For the purpose of the project a common understanding and definition of the term "GW-body" was found (see Annex 8).

Initial and further characterisation

Open questions

- There is a need of guidelines on the initial and further characterisation of a GW-body.
  As the general description of a GW-body is an essential basis for the interpretation of the quality data, such information (based on the provisions of the WFD, Annex II) was collected on the basis of an online questionnaire. Furthermore, a verbal description and GIS-data were collected (see chapter 4.2)

Selection of parameters to be monitored and Criteria for surveillance monitoring and operational monitoring - identification of a risk

Findings

- In order to give the trend detection an "early warning function" the starting point was defined as follows: if CL95 > 75 % limit value the GW-body is identified as being at risk to fail the objectives in Art. 4 of the WFD, hence operational monitoring has to be carried out to have a sound data basis for the assessment of the status of the GW-body and for the trend assessment (see chapter 5.4.4).

Open questions

- Which parameters should be monitored? Which criteria should be applied if no limit values are available?
  Within the group it was even unclear if there is a basis for monitoring substances which might pollute groundwater without having any limit or action values. Operational monitoring of parameters starts with the detection of a risk as well as trend detection is closely connected to the identification of a risk. Identification of a risk and starting point for trend assessment are defined as functions of a limit value within the project. As there are only few limit values to which the WFD refers (only nitrate and pesticides), quality status is principally only to be assessed for very few parameters.
  The term "action values" was proposed to be applied for aggregated results at the GW-body level but not for individual points within a GW-body. In case of the exceeding of an action value, actions should be taken to improve GW-quality.
  In order to extend the applicability of the WFD to other parameters there is a need of explicit enumeration of additional parameters and the designation of "action values" if no limit values are available.

Monitoring network

Open question

- strategy for designing the surveillance and the operational monitoring network.

Findings

- A minimum number of 3 sampling sites per GW-body (1 site per sub-body) is required (see chapter 5.3.3 and 5.4.1).
- The importance of continuity with regard to selected sampling sites was highlighted.
  The replacement of sampling sites should be kept as low as possible. In case of changes of monitoring stations it should be assured that these changes do not affect the outcome of the assessment.
- The distribution of monitoring sites as well as the selected number and types of sites was highlighted as important with regard to the applicability of the proposed statistical methods and the comparability of the assessment.
  Homogeneity of the monitoring network is a prerequisite and can be checked by a representativity index which was elaborated. In case of hydrogeologic inhomogeneity the monitoring network has to be representative of these conditions. (see chapter 5.4.1).

Monitoring frequency

Findings

- The proposed procedures take into account of the minimum requirement (WFD) of one measurement per year.
- The sampling frequency should be in accordance with the natural conditions of the GW-body.
- In the time series some observations may be missing, but the missing of two or more subsequent values should be avoided for trend assessment, as this would cause a risk of bias due to extrapolation.
- In order to avoid bias by seasonal effects which reduces the power of the trend analyses and to avoid induced trend phenomena it is necessary to take care of the sampling time or period (see chapter 5.4.4). In case of yearly measurements it should be guaranteed that the measurements are taken in one and the same quarter or within a certain time period of the year.

- Seasonality effects might be induced due to different monitoring frequency from site to site.

Analytical requirements for LOQ and LOD

Open questions

- The discussion showed that there is an urgent need to provide sufficient information on both LOQ and LOD (see chapter 5.3.2 and 5.4.2)
- The sampling procedure itself and chemical analysis should ensure continuity in results. Relevant norms/standards are to be applied. (e.g. the treatment of samples was emphasised). It was regarded as important to characterise the applied analytical methods, to ensure comparability of results.

Findings

- If a limit value is available the LOQ shall not be higher than 60 % of the limit value. This minimum requirement on the LOQ in relation to a limit value was laid down in order to enable the proposed calculations (see 5.4.2).

Data exchange formats

Open questions

- In order to support the comparability and comprehensibility of results it is necessary to develop common data exchange formats.

Findings

- For the purpose of the project data exchange formats for the collection of test data were developed.
  The collection of information on the general characterisation of GW-bodies via an online questionnaire improved the comparability of information and avoided manipulation of data as a common source of mistakes (see Annex 10 and 11).
  For the collection of quality data a databases exchange format has been developed (see Annex 8 and 9).
- GIS data (location of GW-body borders, location of sampling sites and supportive information (e.g. projection, scale etc.)) were collected for the purpose of the project. Data were converted and merged.
- GIS data were not only used for presentation of results (via WebGIS) but also for calculation purposes (assessment of network criteria, land use via CORINE Landcover).

How to present results

Open questions

- In order to support the comparability of results it is necessary to develop guidelines on the presentation of results.

Findings

- For the project a WebGIS was implemented where selected results and aggregated data from the project can be accessed.

8 ANNEX

The following Annexes can be found on the CD-ROM attached.

Annex 1: *GWstat* Manual

Annex 2: Groundwater Body Characterisation

Annex 3: Verbal descriptions of Groundwater-Bodies

Annex 4: Statistical considerations

Annex 5: Proposed and Investigated Algorithms

Annex 6: Data sheets

Annex 6a: Sample calculation of the Representativity Index R_U

Annex 7: Histograms

Annex 8: Data exchange Guidance

Annex 9: Data exchange Specifications

Annex 10: Groundwater Body Description - Online Questionnaire

Annex 11: Groundwater Body Description - Glossary and Online Help

Annex 12: Monitoring & Statistics - Activities in Member States