Contract n° 018486

NORMAN

Network of reference laboratories and related organisations for monitoring and bio-monitoring of emerging environmental pollutants

Co-ordination action

Priority 6.3 – Global Change and Ecosystems

Deliverable V4.1

Protocol for the validation of chemical and biological monitoring methods

- Improved version -

For the sake of user-friendliness the three protocols are combined into one single document

Due date of deliverable: June 2008
Actual submission date: February 2009 (Version 2.0)

Start date of the project: 1st September 2005 **Duration: 3 years**

Work package leaders: WP V1: Pim Leonards (IVM - NL)
WP V2, V3 and V4: David Schwesig (IWW - DE)

Subproject leader: IWW (DE)

Contact person: David Schwesig

Tel +49 (0) 208 40 30 32 17
Fax +49 (0) 208 40 30 38 0
E-mail d.schwesig@iww-online.de

Contributions from:
D. Schwesig (IWW – DE), U. Borchers (IWW - DE), L. Chancerelle (INERIS – FR), A. Duffek (UBA – DE), U. Eriksson (ITM – SE), A. Goksøyr (Biosense – NO), M. Lamoree (IVM – NL), P. Lepom (UBA – DE), P. Leonards (IVM – NL), D. Leverett (UKEA – UK), M. McLachlan (ITM – SE), V. Poulsen (INERIS, now AFSSA - FR), R. Robinson (NPL – UK), K. Silharova (SK), P. Tolgyessy (SK), JW Wegener (IVM – NL), D. Westwood (UKEA – UK).

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Project funded by the European Commission within the 6th FP (2002-2006)
Sub-priority 1.1.6.3 Global Change and Ecosystems
Co-ordinator : INERIS B.P. n°2 - 60550 - Verneuil-en-Halatte, France
Internet : www.ineris.fr Tel. +33(0)3 44 55 66 77 fax +33(0)3 44 55 66 99

CONTENTS

1 Preface

This document describes a framework to enable the validation of methods used for measuring emerging pollutants or assessing their toxicity. Emerging pollutants are usually substances that have not been included in routine monitoring programmes as required by European legislation. These substances are often potential candidates for future legislation (depending on research on their ecological and [eco-]toxicological relevance), and may be included in a range of requirements for subsequent monitoring purposes. Comparability and reliability of monitoring data are essential for any meaningful assessment and for the management of environmental risks.

For emerging pollutants, there is concern about the comparability of data at the European level. Methods used for the monitoring of emerging pollutants have often not been properly validated either in-house (i.e. within a single laboratory) or at the international level. Such methods are often not well established in the scientific community, and are therefore far from being harmonised or standardised. In addition, those methods developed by different institutions and organisations may only be applicable to specific conditions (matrix, organism, concentration), which may further complicate data comparability.

These issues may be addressed by adopting a harmonised approach towards method development and validation. The main objective of this document is to provide a common European approach to the validation of both chemical and biological methods for the respective monitoring and bio-monitoring of emerging pollutants (or their effects) in a broad range of matrices.

This guidance takes into account the different requirements for the level of method maturity and validation at different stages of the investigation or regulation of emerging pollutants.

The guidance in this document addresses three different validation approaches, in increasing order of complexity. These are:

(1) method development and validation at the level of research laboratories.
(2) method validation at the level of expert/reference laboratories.
(3) method validation at the level of routine laboratories.

The concept of these three approaches is strictly hierarchical, i.e. a method shall fulfil all criteria of the lower level before it can enter the validation protocol of a higher level.

2 Aims and Scope

In the case of a specific monitoring task, this protocol will guide the user through the following steps:

- classification of existing methods with respect to their status of validation, and the selection of the appropriate validation approach;
- in cases where more than one potential method for a specific purpose exists, the protocol will provide guidance to users for selecting those appropriate methods with respect to their potential for further development and validation. Methods applicable for use within routine laboratories (or with the potential to be standardised in the longer term) shall be highlighted by the selection process;
- development of a method so as to extend its application; for example, if a method for determining a required target compound in a particular matrix is available, but is not suitable for the same compound in a different matrix of interest;

- the validation procedures to be undertaken in order to effectively demonstrate the validation status of a selected method according to the three approaches adopted.

The intended scope of this protocol is to cover a broad range of quantitative and qualitative biological and chemical test methods for the analysis of water (including inland and marine waters, groundwaters, waste waters, and sediment), air, soil and biota.

Optimisation in terms of guidance on specific technical steps to modify a method to improve its performance towards a specific validation criterion cannot be covered by these validation protocols. The technical measures to be taken to improve a method in this way are far too diverse and method-specific to be treated in a general protocol. Nevertheless, a more general concept of optimisation is addressed as an integral part of the validation process, by providing a framework to increase the level of validation maturity of a method towards its intended purpose. For example this may entail running a method through the procedures outlined in the protocol on validation at the research level to identify the QA/QC requirements of the method. These requirements may then need to be modified in order to enable the method to satisfy the requirements of the protocol on validation at the expert laboratory and, subsequently, the routine level.

3 Introduction

3.1 What is validation?

In this document, the term 'validation' is used according to the following definition:

Method validation is the process of verifying that a method is fit for its intended purpose, i.e. to provide data suitable for use in solving a particular problem or answering a particular question. This process includes:

- establishing the performance characteristics, advantages and limitations of a method and the identification of the influences which may change these characteristics, and the extent of such changes;
- a comprehensive evaluation of the outcome of this process with respect to the fitness for purpose of the method.

3.2 The concept of three validation levels

The requirements for methods used for monitoring and bio-monitoring of emerging pollutants depend on

a) the extent of the intended or requested monitoring activity and
b) the potential of the available methods that may be used for monitoring a specific emerging pollutant.

In some cases, fully developed methods used by routine laboratories may already exist. More frequently, in the case of emerging pollutants or newly developing methods there will be a lack of information on the extent to which the methods have been fully developed and validated. It may be the case that there are few methods available, possibly developed in research or academic institutions, and which have been developed and validated for specific matrices or organisms rather than for those under investigation. In order to cover most eventualities, three distinct (and hierarchical) levels of method validation are described in this document:

Validation 1
The first (and lowest) validation protocol (described in Chapter 7) addresses method development (in terms of extending its application to new matrices) and method validation at the level of research laboratories. The endpoint of Validation 1 is a method with a complete internal validation for the intended purpose at the level of a single research laboratory. The endpoint of Validation 1 is identical with the starting point of Validation 2.

Validation 2
The middle ranking protocol Validation 2 (Chapter 8) addresses method validation at the level of expert or reference laboratories. The main issue is to demonstrate the transferability of the method. This means that the method can successfully be transferred to another laboratory possessing sufficient expertise and experience. The endpoint of Validation 2 is identical with the starting point of Validation 3.

Validation 3
The third and highest protocol Validation 3 (see Chapter 9) addresses method validation at the level of routine laboratories. The main issue is to demonstrate that the method possesses sufficient inter-laboratory performance and is applicable for use at the level of routine laboratories. This also comprises the development and control of key aspects of method documentation and method usability.

Having successfully satisfied the Validation 3 procedures, a method should be fit for standardisation at the European level.

3.3 Guiding principles and main elements of the document

The starting point for any validation activity is usually to demonstrate the applicability of the method to the intended purpose for which it is to be used. In order to find a method that can be used to generate reliable and comparable data (probably for future use in a regulatory context), evidence of the fitness-for-purpose of the applied method is essential. This comprises a number of general principles and criteria that are applicable to most test methods. These principles and criteria have been organised in a number of modules. In this chapter and its sub-sections, a short overview of the main validation modules and approaches will be given, followed by a description of other core elements.

3.3.1 Main validation modules

As outlined in Section 3.2, this document follows a hierarchical approach with respect to the three validation levels. This means that at a certain validation level all requirements for criteria of the lower level or levels shall be fulfilled. Nevertheless, in some cases the same criteria need to be checked again at the higher level of validation, using a slightly different approach. Therefore, certain aspects of some of the following modules are addressed in all three protocols, but sometimes with a different emphasis placed on the criteria.

3.3.1.1 Module A: Test method definition, documentation and general requirements

It is essential that an unambiguous definition of the method, its scope of application, its scientific basis, its (regulatory) purpose and a statement about the need for the test method, and criteria for acceptable method performance be documented. This should also include, if known, a description of the (correlative) relationship (mechanistic or empirical) between the measured quantity or effect and the phenomenon in, or property of, the investigated system, matrix or organism. Furthermore, a detailed documented protocol for the method should be available. This should include a description of all materials and reagents, a description of

what is to be measured, and how this is to be carried out. In addition, guidance on the necessary steps of data handling should also be included (e.g. treatment of raw data, procedures of calculation and data analysis etc). Depending on the level of validation, the protocol should also contain guidance on necessary QA/QC procedures and basic safety information.

3.3.1.2 Module B: Applicability domain and pre-validation

Test methods may be developed, optimised and validated for measuring a compound or an effect within a specified scope. This scope may include restrictions on the media sampled (including cell type or organism), the matrix, the concentration range of the measurand, and other limits of the scope of application. These limitations of the scope of a method need to be investigated and described properly. A prerequisite for any extension of the scope of a test method is the undertaking of a comprehensive investigation demonstrating the suitability of the method for the new scope. In many cases, this investigation should reveal, if appropriate, the need for a modification of the test method to ensure its fitness for the new purpose (and possibly consequential need for re-validation). When extending the scope of an existing method, additional testing on other influences (such as new interferences and confounding factors) may also be required. In the context of this protocol, this process is called pre-validation and is addressed in the Protocol V1.

3.3.1.3 Module C: Intra-laboratory performance

The basic performance characteristics of the test method within one laboratory should be determined in order to evaluate its reliability, relevance and limitations. These performance characteristics may include (among others) parameters such as precision, trueness, selectivity and traceability. There are differences between the various types of methods with respect to the exact set of criteria and the most appropriate tools to determine their quality or values. This guidance document aims to cover the most important scenarios.

3.3.1.4 Module D: Inter-laboratory transferability

An important step towards implementation of a method for regulatory purposes is the demonstration of its transferability to another laboratory. The transferability of a method depends on many factors. Some of these factors are intrinsic to the test method itself, e.g. the need for specific and dedicated equipment which may not be available to other laboratories, or poor robustness of the method. Other factors are not intrinsic to the test method itself, and depend on the expertise and experience available within different laboratories. It is also important to consider 'external' factors. For example, the quality of the description or protocol of the method should be questioned (Is it detailed and complete, is it unambiguous or is it open to misinterpretation?). These factors may be essential for the transferability of a test method. A test method may be regarded as being transferable if at least one other laboratory can produce similar or better results to the one that undertook the initial development (and successful internal validation). The measure for the similarity of results between the laboratories and the level of acceptability can differ from case to case (depending on the type of method and measurand), and may subsequently be prescribed by the regulator. The validation protocols in this guidance cover most relevant scenarios.

3.3.1.5 Module E: Inter-laboratory performance

In order to be applicable for large-scale European monitoring programmes, a test method for emerging pollutants shall show sufficient inter-laboratory performance at the level of routine laboratories (preferably across Europe). For satisfactory performance in a large number of routine laboratories, it is essential that a test method should show a high degree of robustness and usability. Furthermore, the completeness and clarity of the method description (protocol) are more important than when testing the transferability of a method between expert laboratories. The protocol should contain detailed QA/QC procedures and should not be open to misinterpretation.

The principal tool used to evaluate the inter-laboratory performance of a method is an inter-laboratory comparison involving the analysis of identical test items across all participating

laboratories. A collaborative study to evaluate inter-laboratory performance requires a considerably higher number of participating laboratories (and broader geographical coverage) than the investigation of inter-laboratory transferability. These inter-laboratory studies are usually designed with the aim of minimising the effect of intra-laboratory variation on the measures used to characterise and evaluate the performance characteristics of the test method. The tools and procedures to establish the value of measures for inter-laboratory performance criteria may differ depending on the type of method and measurand. Validation studies carried out as a part of European standardisation activities are examples of this level of validation, i.e. European inter-laboratory comparison.

3.3.2 Method classification and method selection

3.3.2.1 Method classification

In order to select the appropriate validation protocol, potential candidate methods need to be classified according to their level of maturity and validation. This may be considered as an abridged version of a retrospective validation study. Guidance on method classification with respect to the three validation levels is given in chapter 4. However, this should only be used to provide a quick and approximate estimate of the maturity and validation status of the method.

In cases where numerous potential methods exist for a specific purpose, it may not be practicable to perform a detailed retrospective assessment of the status of all methods within a reasonable timeframe. If one of the classified methods has been selected (see section 3.3.2.2 and chapter 6), and subject to the appropriate validation protocol, detailed work on either retrospective or prospective validation studies may reveal some inadequacies within some of the methods. In this case it may be necessary to investigate certain modules of a lower validation level. This process may eventually lead to a downgrading of the method by one (or even two) validation levels. In a worst case scenario, this may lead to the decision to reject the method and to repeat the process of method classification and selection with another method.

3.3.2.2 Method selection

If more than one method is available which may be suitable for the intended purpose, a selection step is usually necessary. It is one of the main objectives of this guidance to facilitate, support and foster the selection of those methods, which, in the longer term, have a high potential for full validation. In order to support the European harmonisation of measurement methods, there should, at an early stage, also be control (even at the lower validation levels) of methods that show high potential to be harmonised or standardised in the future. Therefore, this guidance contains a separate section on method selection (chapter 6). However, the proposed selection approach is intended to assist in the method selection process and is not to be used in a way that absolves users from exercising their expert knowledge on case-by-case considerations about specific tasks and the framework within which the method is to be used. Fitness for purpose is the main driver for methods to become accepted for European monitoring tasks, and the objective of this selection approach is to focus available resources (which are usually limited) on the most promising methods, rather than to prescribe or restrict the use of certain methods. Furthermore, the method selection approach is designed to be objective, i.e. to prevent the selection of methods based on personal preferences, without some objective demonstration of their effectiveness over other methods that are equivalent.

The application of the method selection process outlined in this document does not mean that the selection process will highlight a single method (although this may be the most likely outcome). It may also be possible to highlight more than one method, provided sufficient resources are available to adequately perform the necessary validation work for all selected methods.

3.4 Visualisation of the workflow and its options

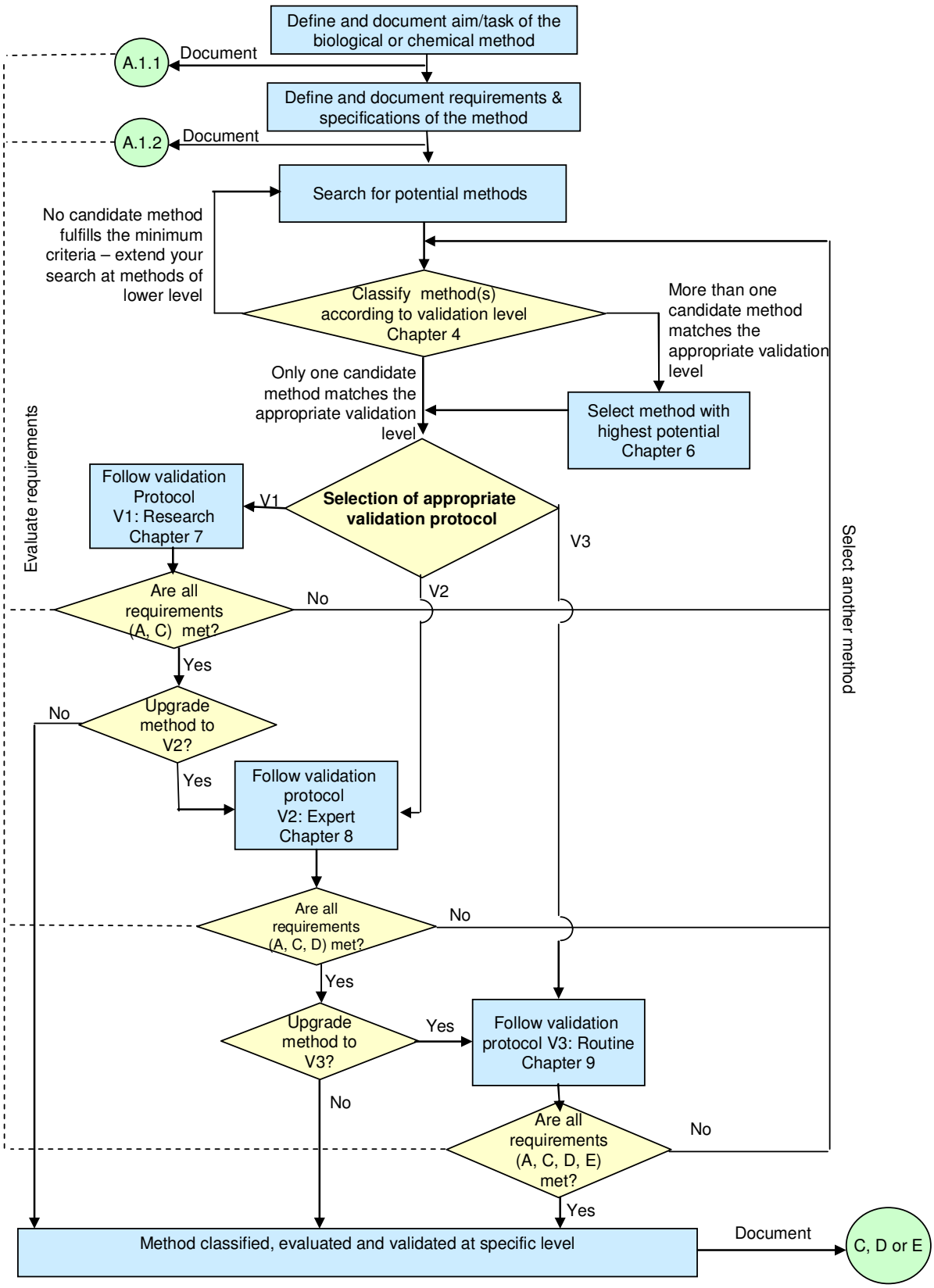

Figure 1 Visualisation of the Workflow

4 Method classification with respect to the level of validation maturity

This chapter provides guidance on how to classify existing methods with respect to the three levels of validation. As a result of the classification, the user should be directed to the appropriate validation protocol. The validation modules outlined in Section 3.3.1 should be used to identify the criteria that shall be fulfilled at the endpoint of each validation protocol. If a method fails to fulfil one or more mandatory criteria assigned to the modules of the respective validation level, the method should be placed in the next lower level of validation maturity. In Table 1 '+' indicates that the respective criterion must be fulfilled by the candidate method in order to be considered as validated at the respective level, and '(+)' means that the fulfilment of this criterion is not mandatory, but is, at least, highly recommended. For the lower Validation 1 protocol, minimum mandatory requirements for methods to enter the validation procedure are underlined. This classification scheme therefore acts as an input filter to the whole validation process.

Table 1 Method classification - Requirements for the three validation levels

Criteria (if applicable to the type of test method which has to be classified)	***Required at Endpoint of Validation level***		
	1	***2***	***3***
Module A – Test method definition & documentation			
Definition of need	+	+	+
Purpose			
Development of knowledge	+	+	+
Regulatory purpose			(+)
Scientific basis			
Defined mechanism/effect	+	+	+
Scientific proof of relationship between a measured signal or effect and a phenomenon in or property of the investigated system	+	+	+
Documentation (Protocol)			
With sufficient information for a researcher with special expertise to use the method	+	+	+
With detailed information sufficient for a trained analyst		+	+
According to ISO 78-2 (standard-like)		(+)	+
With detailed QA/QC procedures and performance criteria		(+)	+
Statistics available (record of performance characteristics)		(+)	+
Dissemination			
Grey literature	+	+	+
Peer-reviewed publication		(+)	+
National, European or International Standard			(+)
Module B – Applicability domain			
Applicability			
To the compound (class) or effect of interest	+	+	+
To the matrix of interest	+	+	+
To the environmental compartment of interest	+	+	+
To the organism or cell type of interest	+	+	+
Modules C to E – Intra- and Interlaboratory Performance			
Matching the performance characteristics required (e.g. from the regulator or other 'ordering' party)			
shown by one (research) laboratory only	+	+	+
shown by comparison study with at least 2 laboratories		+	+
By routine laboratories (proven by inter-laboratory study)			+

5 Documentation of the validation process

All validation steps need to be documented in a proper way. In order to facilitate this process and to ensure a common documentation format, templates for documentation (in the form of tables) are presented in the respective validation protocols, e.g. see Table 7 to Table 9in the V1 protocol (Chapter 7). A harmonised set of documentation templates may help to ensure that the documentation of the validation process is comprehensible and traceable.
Such templates also enable a quick evaluation of the validation status of the method (e.g. according to the method classification scheme given in Chapter 4), or the identification of gaps that need to be bridged.

Five different templates are used for documentation of the validation process. These five templates correspond to the five validation modules A, B, C, D and E which are defined and described in Chapter 3.3.1. Therefore, the extent of documentation and the number of templates to be completed depends on the level of validation a method has passed (see Table 2).

Table 2 Method documentation - Requirements for the three validation levels

Documentation required of	***Method validated at level***		
	V1	***V2***	***V3***
Template A	+	+	+
Template B	+	+	+
Template C	+	+	+
Template D		+	+
Template E			+

Templates A and B shall contain general information on the method (e.g. its definition, and its applicability domain), whereas Templates C, D and E correspond to the specific validation tasks carried out at the level of V1, V2 and V3, respectively. The documentation templates (at least those corresponding to modules C, D and E) can therefore also be used as a preview of the validation tasks which have to be carried out at the respective validation level.

The documentation of the method validation process should not be confused with the method description, although some of the information in the two types of documents may be similar or even identical. Information from Templates A and B can be used to compile the information for the method description. At the V1 level, the information given in Templates A and B, together with an appropriate reference to the (scientific) literature, may be sufficient as method description, but at the higher levels the requirements for the description of the method successively increase. Therefore, more comprehensive level-specific sets of criteria for the method description have been compiled for the V2 and V3 level, and should be followed in the preparation of the method description.

If a method enters a higher validation level, information in Templates A and B may need to be updated, because more information has been or needs to be gathered on specific requirements or abilities of the method, or requirements for the method and its performance characteristics may change. Therefore, a method successfully validated up to the V3 level will usually be accompanied by a set of templates recording the history of the validation process of this particular method.

6 Method selection

6.1 General aspects

6.1.1 Method selection approach

If, during the search for a method for a specific purpose (and the classification of the methods), more than one method is highlighted at the requested validation level, there will be a need for method selection. This is to identify the method showing the greatest potential for progression through the validation procedures, for transferring the method to the routine level and for harmonising or even standardising the method in the longer term.

At a particular validation level, n, only methods which can be regarded as being validated according to the requirements of the validation level n-1, shall be considered as candidate methods for this selection procedure. Furthermore, this method selection procedure shall only be used for methods which generate an equivalent output, e.g. the measurement of the concentration of a specific compound, or the detection of the same well-defined effect in a certain biological system. The comparative selection procedure cannot be applied to methods which generate different outputs.

In the following chapter and its sub-sections, criteria are provided which can be used to compare potential methods, and select the one with the greatest potential for fulfilling the requirements outlined in Chapter 3.3.2.2. In general, the selection approach is based on generic criteria which are applicable to most types of methods (Chapter 6.2). This enables the comparison of a range of different types of methods, however diverse (e.g. chemical methods versus biological methods), provided the methods detect or measure the same compound or class of compounds. Some criteria combine complementary aspects applicable to different types of methods: for example, some criteria may be more applicable to chemical methods, whereas other criteria are more applicable to biological methods. Depending on the type of method and the required validation level, the selection criteria can have differing levels of significance. This is taken into account by introducing an indirect weighting approach using a level-specific aggregation of criteria in consecutive tiers (Chapter 6.3). The selection procedure is therefore based on the step-by-step approach visualised in Figure 2.

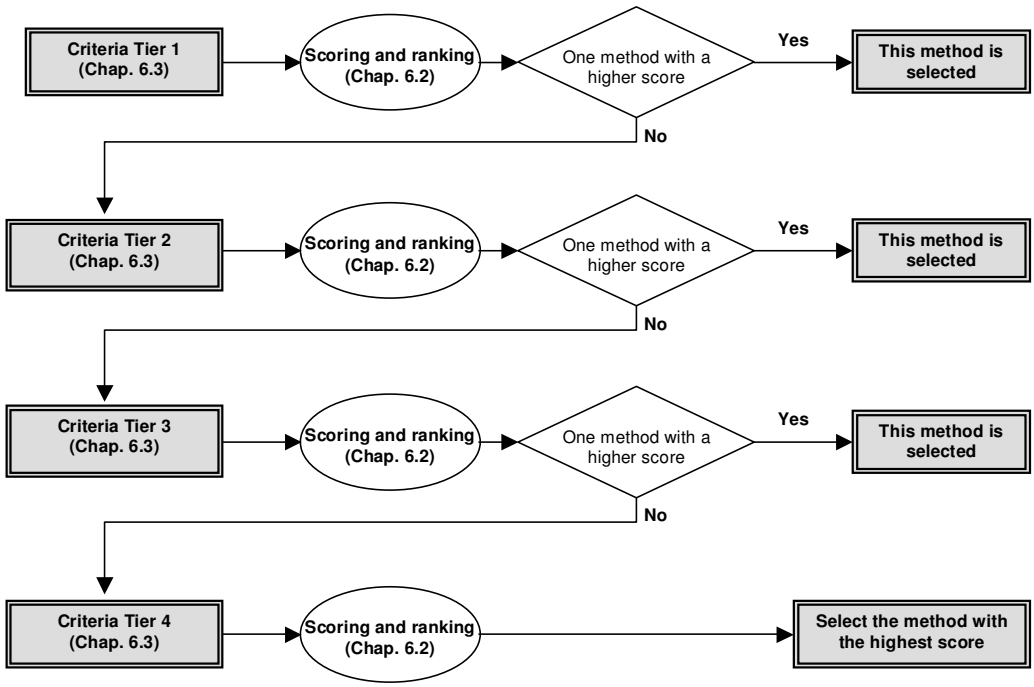

Figure 2 Method selection scheme

6.1.2 Important aspects for the selection of biological methods

The assessment of biological parameters is the only way to evaluate the biological effects of chemicals or mixtures of chemicals. Bioassays, biotests, biomarkers and bio-indicators are thus important elements of programmes that aim to assess the quality of the environment in a biological context.

Different levels of biological organisation, from molecular to individual, population and community, are addressed by these biological tools, and the measured endpoints can thus range from biochemical signals to a loss of mobility or other sub-lethal effects, through to death and/or viability failure of a population. Biological parameters often integrate several kinds of stress and can be used to assess the presence of chemicals or combinations of chemicals. They can be useful to:

- detect toxicants which may not have been previously identified as being of concern
- assess exposure to compounds for which analytical methods are either not currently available or are too expensive to be incorporated into a large monitoring programme.

They also help to identify regions of decreased environmental quality (USGS 2000 and JAMP 2003).

Bioassays and toxicity tests on environmental samples can be performed *in vitro*, with cells or tissues from a variety of organisms, or *in vivo* with whole organisms ranging from bacteria to vertebrates. The tests can provide direct evidence of cumulative contaminant effects on the survival, growth, behaviour or reproduction of living organisms, while controlling for extraneous confounding factors. Those tests conducted with whole organisms are typically quite general with respect to the contaminant eliciting the response.

The tests may also provide more specific information on the nature of the compound involved. For example, when multiple tests are conducted with organisms that exhibit different susceptibilities to specific contaminants (Ingersoll et al. 1992) or when combined with a reductionist approach such as Toxicity Identification Evaluation (TIE) or by selectively sampling or fractioning the test medium either prior to or after testing.

Biomarkers include biochemical, physiological, morphological or histopathological responses of organisms signifying chemical exposure (Melancon 1995). Although some biomarkers are specific, many are quite general. Responses can be unique to one contaminant or a relatively small group of structurally similar chemicals, or they may be general indicators of organism or population health that respond to a wide variety of chemicals and other stressors.

Biomarkers (i.e. sub-organismic changes) are useful tools for early detection of some changes in the chemical environment of autochthonous populations before any effects are observed at higher levels of organisation, because the response to a chemical is caused by the interaction between the chemical and a cellular or extra-cellular component. Nevertheless, as is the case for both chemical and biological tools, biomarkers possess limitations in the context of environmental monitoring and especially for ecological risk assessment. Lack of knowledge of the environmental factors likely to modify their response (when measured on autochthonous organisms in field), can impair their use and interpretation (giving rise to false-positive or false–negative responses). Moreover, the links between biomarker changes and higher biological level effects are not always established. However, several international bio-monitoring programmes such as MedPol ICES, UNESCO-JOC Black Sea Mussel Watch and RAMOGE, have used the biomarker approach to monitor the health status of aquatic organisms, such as mussels and fish, in European waters for several years. A thorough validation programme for biomarkers may help to discern powerful from not-so-useful parameters.

6.2 Selection criteria, scoring and ranking

A number of objective and generic criteria are used to evaluate the potential of biological and chemical methods in addressing the following objectives:

- to be suitable for European monitoring and bio-monitoring of emerging pollutants
- to successfully achieve the desired validation level
- to become applicable at the level of expert or routine laboratories in the longer term
- to become widely disseminated, harmonised or even standardised.

Any potential method should be scored with respect to the criteria defined for the respective validation level and relative to other methods against which it is being compared. It shall be performed on the basis of a maximum score approach: i.e. the method is scored against each criterion using integer values between 1 and x, with a score of 1 indicating the lowest and x the highest potential of a method to achieve the intended application. The maximum value, x, is limited by the number of methods which are to be compared, e.g. if four methods are to be compared, the maximum value of x is four, and only scores from 1 to 4 should be assigned. Furthermore, with respect to a single criterion, every effort should be made to assign each score only once, i.e. to one method. Nevertheless, it is possible to assign the same score to more than one method in cases where methods are regarded as genuinely indistinguishable with regard to the respective criterion. If there are insufficient data with which to evaluate a certain method against a specific criterion, a zero score should be assigned to the method for this criterion.

This approach enforces an equidistant ranking of the methods for each of the criteria, which are defined in the following subchapters.

6.2.1 Scientific basis and defined mechanism

In order to be considered as a potential method for European monitoring or bio-monitoring of emerging pollutants, a method shall fulfil the generally accepted minimum standards of scientific rigour and common sense. Methods based on effects without any scientifically sound description or based on highly speculative or dubious concepts shall not be considered for validation studies. The evaluation of a method against this criterion can also comprise the level to which a method has been demonstrated to operate as designed, expected or required. This may be achieved through published, peer-reviewed papers or other means (e.g., internal technical reports). The more detailed and convincing the description of the scientific basis and the mechanism of the method, the higher shall be its ranking with respect to this criterion.

6.2.2 Degree of dissemination and reputation

This criterion relates to the level of acceptance already achieved by a candidate method.
Measures for the degree of dissemination and reputation can be (in decreasing relevance):

1. existing use of the method in a national perspective by one or more member state
2. existing national or international standards based on the same method (but probably for investigation of a different matrix)
3. the number of (preferably peer-reviewed) publications based on the use of the method.

6.2.3 Target compound or effect

This criterion shall be used to describe whether the scope of the method encompasses the compound that is to be analysed or the effect that is to be monitored. A low score shall be given to a method that has been developed for a compound or effect other than that required. A high score shall be given to a method which has been developed (and at least

partly validated) for exactly the compound or effect in question. If the target compound is not just one single (chemical) compound but a class of compounds, consideration should also be given to whether the method is suitable to detect or quantify the whole class of compounds or only a subset of this compound class.

6.2.4 Target matrix or organism

This criterion relates to the suitability of the method for the target matrix or organism that is to be investigated. A low score shall be assigned to a method that has been developed for a target matrix or organism other than that required for the intended monitoring purpose. A high score shall be given to a method which has been developed (and at least partly validated) for exactly the matrix or organism in question.

6.2.5 Application Range and Sensitivity

This criterion relates to the suitability of the method to detect an emerging pollutant (or its effect) at the target concentration level. Depending on the type of method, there may be different approaches to score the method with respect to this criterion (relative to the other methods with which it is being compared).

For quantitative chemical methods in particular, a scoring of the method shall be based on the relation between:

- the lower limit of application (LLOA) of the method as documented in the method description or determined by use of the method
- the requirements for the LLOA of the anticipated monitoring purpose

If no requirements for the LLOA have been defined by the regulator (or another client requesting the conduct of the method selection and validation procedure), a (pragmatic) default approach shall be applied by assigning the highest score to the method with the lowest LLOA, and a similar ranking of the other candidate method(s).

With methods where an LLOA-like measure is not suitable for scoring, a measure of sensitivity may be more appropriate to score the method. The sensitivity of a method is usually represented by one or more measures characterising the relationship between the quantity or property of a compound and the signal or effect obtained.

The sensitivity of many biological methods is represented by the concentration of a compound which is required to elicit a prescribed response. In toxicity tests, for example, these responses are usually represented as the median inhibitory, effective or lethal concentration (IC_{50}, EC_x, LC_{50}), the no observed effect concentration (NOEC), and the lowest observed effect concentration (LOEC). The sensitivity of a particular test will usually vary markedly between different chemical classes due to their different modes of toxic action. The sensitivity of methods with the same category of endpoint (e.g., reproduction) can be compared, and for a given substance, the method with the lowest EC_x (or IC_{50}, LC_{50}) will be regarded as the most sensitive method. Sensitivity may, however, be expressed differently in other types of biological methods (e.g. biomarkers). However, methods selected for the analysis of emerging pollutants are likely to be specific to a single compound or group of compounds. In this context, sensitivity is likely to be represented by

i) the limit of detection (LOD)
ii) the limit of quantification (LOQ)
iii) the application range, i.e. the range of concentration of a chemical over which the method can be expected to respond. The score assigned to methods based on the comparison of application ranges will depend on the objective of the monitoring requirement for the compound of interest. Higher scores may be applied to those methods with the lowest threshold or those with the widest range of application.

6.2.6 Trueness

Score ranking of methods against this criterion shall reflect the amount and quality of information on trueness, and the degree to which the method fulfils the respective requirements of the intended purpose. In the context of this protocol, the term 'trueness' is used according to ISO terminology (cf. ISO 3534-1 and ISO 5725 series), and should not be confused with the term 'accuracy', which encompasses both, trueness and precision (for details see glossary in Chapter 12.1).

In chemical analysis, trueness usually represents the proximity of the average value obtained from a large series of test results and an accepted reference or 'true' value. However, trueness can also be defined for individual test results in relation to an accepted reference value (which may change depending on the type of trueness assessment being undertaken). In the application of biological methodologies the 'true' or expected value may be less evident than in chemical methodologies. However, they can usually be represented by a robustly derived (and generally accepted) reference value (for example the mean value [response] obtained over a series of measurements with a known concentration of test chemical). In all cases, it is the measurement of proximity between actual and reference values that is to be assessed.

In addition, and in the context of this protocol, trueness may also include an element of comparability of the results generated by a method with other well-established methods with similar mechanisms of operation (if available). This comparison will not necessarily be against methods which are used for the analysis of the same compound (for these are likely to be the other methods against which the score will be assigned) but methods which are comparable in terms of mode of action or response type. Potential methods which achieve a trueness measurement which is close to the trueness demonstrated in similar methods will score higher using this criterion, especially where no reference value can be derived with which to assess the criterion directly.

6.2.7 Precision

This criterion relates to the closeness of agreement between independent test results obtained under stipulated conditions. Depending on the exact stipulated conditions, there are several distinct quantitative measures to evaluate the precision of a method. Depending on the desired validation level, different measures of precision are of particular interest. At the lower validation levels, measures of intra-laboratory precision (such as repeatability) are of primary interest, whereas at higher validation levels measures of inter-laboratory precision (such as reproducibility) are increasingly important. Detailed information on the degree to which other factors (e.g. temporal, spatial or biological variability) affect the precision measures of a method should be regarded as a bonus in the ranking of a method (relative to the other methods with which it is being compared).

6.2.8 Calibration and Traceability

Method calibration is the underpinning process by which inter-comparable results can be achieved. The traceability in measurement results obtained from well founded calibration procedures can ensure that results obtained from different laboratories (using the same or different methods) can be compared and, if appropriate, combined. In addition, a complete traceability chain (if applicable for the particular type of method) linking the method calibration back to fundamental realisations of SI units provides a key component in understanding the uncertainty of the results using the method. A traceable calibration is an important characteristic of a method. In ranking methods against this criterion it has to be considered that for different types of methods, different approaches and degrees of traceability can be achieved. A method with a well-described calibration function is to be preferred over a method that lacks information about the calibration function. A calibration function can be a curve, a formula or a table showing the relationship between raw output

data of the method (i.e. the pure signal, e.g. light intensity, absorbance, counts, mV) and the concentration of a working standard (often a solution of the target compound). The working standard should be well described and obtained from a recognised source. It should preferably be certified, or at least have a known value. Methods based on working standards prepared in-house that have not been tested for purity will score low, methods with descriptions that lack any information on the working standard source and purity score lowest.

6.2.9 Selectivity/Specificity and Confounding Factors (Interferences)

This criterion compares equivalent methods with respect to their ability to determine the concentration of interest, and the degree of understanding of the mechanisms that generate the measured results including any confounding factors or interferences which may affect or complicate the interpretation of results.

When assessing selectivity / specificity, consideration should be given to the ability of the method to detect or respond to the target compound rather than the degree to which the method has been developed / designed for the target compound. In addition, the degree to which the method can actually detect the target compound in the relevant sample matrix (i.e. in a mixture of compounds) should also be considered.

Confounding factors or interferences will range from technical factors affecting the performance of the method (e.g. temperature, pH, retention time, presence of non-target compounds) or factors affecting the specified matrix or organism, to those concerned with interpretation of the measured effect.

Biological methods may be particularly susceptible to multiple interferences such as the 'state' (nutritional / reproductive), 'history' (genetic / exposure) or distribution of the exposed / sampled organism. In biomarker or *in vitro* biochemical methods, the interpretation of the results with respect to the exposure to chemical mixtures and the associated exposure time should also be accounted for, as well as an understanding of the quantitative correlation of concentration and response (if any). In scoring against this criterion, higher scores will be assigned to methods which are most specific / selective within the required sample matrix and those with the best understood, described and controllable confounding factors / interferences. A well-designed bioassay that has been tested for response to compounds other than the target compound(s), often expressed in percentage cross-reaction, will score higher than a method for which such a test has not been performed. A higher score can also be assigned for a method showing low levels of tested cross-reactions.

6.2.10 Robustness

Robustness in the context of this protocol can be defined as the ability of a method to provide a consistent response under changing external conditions. This is related to 'Ease of Use' (see below) and 'Precision', but differs in that it describes the degree to which a method provides meaningful results over repeat measurements under varied external conditions rather than the proximity of the repeat values themselves. Thus, a method that has been tested under (deliberately) varied experimental or environmental conditions (such as e.g., different staff, laboratory temperature, extraction or incubation time and temperature, solvent pH) and has the corresponding variation of the results expressed in percentage change, will score higher than a method that has not been subjected to such testing. If more than one method has been tested in this way, it may not necessarily be the method showing the lowest variation due to changing conditions which is assigned the highest score (for example, this may be because the range of variation of a particular external factor maybe unrealistic or at least not relevant under real laboratory conditions). Therefore, the following aspects should be considered in comparing the effect (in terms of variation) of varied experimental or environmental changes:

- the type of the external conditions which have been varied - only changes in those conditions which are relevant and likely to occur in the practical use of the method should be considered;
- the variation range (amplitude) of the deliberately varied external conditions - only the effect on the results caused by a comparable variation range of the external conditions should be evaluated.

6.2.11 Ease of use

This criterion relates to:

- The time taken to initiate the method and to achieve meaningful and robust results using the method;
- The degree (and number of areas) of expert knowledge needed to adequately perform the operational steps of the method (e.g. maintaining cultures, sample preparation, extraction and derivatisation steps, operating the measurement instruments), and all necessary data treatment steps (e.g. calculation and interpretation) in order to obtain meaningful results;
- The degree of training required (e.g. does it require a user to learn just a new way to apply existing skills in a new way and combination, or are completely new skills required?)
- The degree of effort needed to cope with the susceptibility of the method to produce false, biased or otherwise unreliable or unacceptable results. This is related to Robustness, Trueness and Precision, but differs in that it describes the effort and expertise needed to achieve a performance of the method compliant with the requirements of the specific task;
- The degree to which QA/QC measures for the method can be formalised or standardised in order to allow a quick and easy control / judgement on the reliability of the measurements on a routine basis.

A method that can easily be established in a routine laboratory, with laboratory staff being able to perform all operational and computational steps on a routine basis, shall rank higher than a method requiring large establishment efforts and case-by-case expert judgement from staff with specific (academic) expertise or training in order to produce robust and reliable results.

6.2.12 Cost of a method

The score ranking of a method against this criterion shall consider all expenditure associated with a method including those for:

1) implementing the procedure, which includes the purchase of equipment and the staff costs involved in setting up the method;
2) conducting tests, which includes staff costs incurred in carrying out the method, obtaining the data and analysing the data statistically, and the costs of any materials (e.g. apparatus, organisms, reagents). This cost heading may also include an allocation of the costs incurred in maintaining cultured test organisms (where applicable).

It is important to distinguish between tests with high establishment costs but low costs per unit test and those with low establishment costs but high costs per unit test. For different tests there will usually be different costs depending on the number of tests carried out in a specified period. The highest score should be assigned to the method with the lowest costs.

6.2.13 Rapidity of a method

This criterion relates to the total duration of a method from initiation to the collation of the final dataset. In biological methods, the sensitivity of a method may increase with longer exposure periods. However, methods of shorter duration may be advantageous for test substances that are unstable or that are likely to degrade. This criterion is somewhat related to the criteria 'Ease of use' and 'Cost of a Method', but should nevertheless be treated separately.

6.2.14 Availability of instrumental equipment

This criterion relates to the availability of all technical components needed for carrying out the method, e.g. equipment needed for sample treatment, processing and handling of sample extracts, measurement of the signal and calculation of the result. Instruments may be tailor-made or based on novel technologies. In this case, they are usually not available or not suitable to other laboratories. More well-developed instruments may be available, but subject to patent limitations, which also limits their availability to laboratories. On the other hand, well-established instruments, produced by a variety of suppliers, often with method development support, are readily available and often lower cost. The highest score shall be assigned to the method with the easiest available instrumental equipment.

6.2.15 Availability of materials

This criterion refers to the availability of all types of materials needed for carrying out the method, including reagents, test substrates and organisms. Methods using materials (such as extraction column packing materials and reagents) that are commercially available from a number of suppliers are preferred over methods using materials or reagents that can be purchased only from a single supplier. However, methods that use materials or reagents purchased from a single supplier are preferred over methods using materials or reagents that have solely been prepared by the laboratory that has developed or published the method.

In the case of methods requiring test organisms, the following aspects should be considered:

- temporal variability of the availability of the organisms or life stages of the organism (throughout the year). The temporal variability of the response (as biomarkers) should be considered as a confounding factor. Biological material may be available, but not suitable for a particular method at given periods during the year
- the possibility of maintaining test species in the laboratory
- the availability of organisms from a supplier or the environment when required
- legal regulations restricting the use of the particular organism
- ethical issues related to the use of the test organism.

6.2.16 Environmental and safety Aspects

This criterion relates to the precautions and implications with respect to environmental & safety aspects which are linked to the application of the method.
The following aspects should be considered

- the need for persistent, toxic or bio-accumulating chemicals (e.g. as reagents or solvents)
- procedures which are subject to specific safety regulations
- the need for test organisms which are currently (or will, in the near future, be) subject to specific protection measures.

For example, methods which lead to the production of large amounts of highly toxic wastes or require the use of large amounts of chemicals with a known adverse environmental effect

shall rank lower than methods using small amounts of less harmful or easily recyclable waste.

A more objective way to evaluate and rank methods with respect to this criterion may be a formalised risk assessment. An example of such an approach is given in Table 3.

Table 3 Risk assessment for health and environmental risks from chemicals and equipment

Risk value	Frequency of Use	Chemicals		Equipment
		Hazard symbol	*Amount used on each occasion*	
1	monthly or less often	no hazard class label	up to and including 100 g (or ml)	Slight harm superficial injuries such as minor cuts and bruises
2	weekly or fortnightly	harmful, irritant, flammable;	between 100 g (or ml) and 1 kg (or l)	Moderate harm more serious superficial injuries such as cuts with prolonged bleeding or severe bruising
3	daily or more often	toxic, very toxic, corrosive, explosive, dangerous for the environment	1 kg (or l) and more	Considerable harm minor fractures and ill health requiring up to a week away from work
4	NOTE: potential injury risk values of 4 and 5 are just present for completeness, and should only be used in risk assessments when staff involved have had specialist formal training, and then only where no other option is available	-	-	Serious harm serious fractures or other injuries causing minor permanent disabilities, or ill health resulting in prolonged effects.
5		-	-	Extreme harm Death, severe injuries causing profound permanent disabilities or ill health with permanent effects

For the evaluation of risks due to chemicals: multiply together each number scored for hazard, frequency and amount to achieve the risk rating for the use of each chemical.

For the evaluation of risks due to equipment: multiply together the numbers scored for potential injury and frequency of use to achieve the risk rating for the use of each item of equipment.

Risk rating of a method shall be done by adding up the numbers of all 'partial' risk ratings.

The method with the lowest score in the risk rating shall get the highest score in the ranking of methods with respect to this selection criterion.

6.2.17 Method description

This criterion relates to the availability of a detailed and unambiguous description of the method or standard operating procedure (SOP) which can be used by laboratories conducting the methods. The existence of a detailed method description will ensure consistent use of the method by different laboratories. Ranking of a method against this criterion (relative to the other methods with which it is being compared) shall reflect the degree of detail and comprehensiveness of the protocol. A comprehensive method

description should provide guidance on QA/QC measures and information on limitations, interferences and disturbances as well as a prescriptive procedural instruction of method performance.

6.3 Selection procedure

The selection criteria defined in Chapter 6.2 have been grouped into specific sets for each validation level (Table 4 to Table 6). Within each set, the criteria have been arranged in (up to four) consecutive tiers.

The selection process starts at the first tier of criteria (Tier 1). For each method, the scores of all criteria in Tier 1 are added together. The method with the highest total score in Tier 1 shall be selected. If no decision can be made at the first tier, i.e. the scores of two or more methods are identically, the scores at the second tier shall be evaluated. In a similar way, subsequent tiers shall only be evaluated if no decision can be made based on the tier-specific results of the methods at the preceding tier. Only those methods receiving identical scores at a certain tier should proceed to the subsequent tier, regardless of the initial number of methods compared at Tier 1. It may also be advisable to confirm the selection by evaluation of a subsequent tier in cases where the differences between methods are very small; e.g. if in the case of ten potential methods, the highest and next-highest scoring methods in Tier 1 differ only by one or two points, it is recommended that these two methods be considered equivalent in Tier 1, and both should pass to the next tier, i.e. Tier 2.

It may not always be possible, or necessary, to complete the selection process to obtain just a single method. In such cases, it is essential that sufficient resources and laboratories are available to perform the required validation steps for all selected methods.

The fact that certain criteria appear at later tiers at the V2 or V3 level does not mean that these criteria are unimportant. In several cases (e.g. for the criteria target compound, trueness), potential methods shall fulfil the (pre-set) requirements with respect to these criteria to be considered as potential methods for the higher validation levels at all. But even among those methods that fulfil the pre-set requirements, there may be different degrees of performance with respect to these criteria, and therefore these criteria can still be used as selection criteria, but usually not at the first tier.

Table 4 Tiers of criteria for method selection at the research level

	Criteria for method selection at the research level (V1)				
Tier 1	Scientific basis	Target compound or effect	Target matrix or organism	Selectivity, Specificity and Confounding Factors	Application Range & Sensitivity
Tier 2	Trueness	Precision			
Tier 3	Calibration & Traceability	Robustness	Environmental & Safety Aspects	Availability of instrumental equipment	Availability of materials

Table 5 Tiers of criteria for method selection at the expert level

Criteria for method selection at the expert level (V2)				
Tier 1	Target compound or effect	Target matrix or organism	Selectivity, Specificity and Confounding Factors	Application Range & Sensitivity
Tier 2	Trueness	Precision	Calibration & Traceability	Robustness
Tier 3	Availability of instrumental equipment	Availability of materials	Ease of use	
Tier 4	Dissemination & Reputation	Environmental & Safety Aspects	Method description	

Table 6 Tiers of Criteria for Method Selection at the Routine Level

Criteria for method selection at the routine level (V3)				
Tier 1	Ease of use	Availability of instrumental equipment	Availability of materials	Robustness
Tier 2	Calibration & Traceability	Selectivity, Specificity and Confounding Factors	Method description	Dissemination & Reputation
Tier 3	Application Range & Sensitivity	Trueness	Precision	
Tier 4	Environmental & Safety Aspects	Cost of the method	Rapidity of the method	

7 Protocol V1 – Within-Laboratory Validation (Research Level)

The Validation V1 protocol covers the scenario where for a given (group of) emerging substance(s) a method is available and is selected according to the procedure described in Chapter 6, but

- is either not applicable to the matrices, compartments or organisms of interest (pre-validation) or
- its suitability for the intended purpose with respect to certain performance criteria has not been sufficiently tested and proven.

“Of interest” means that there is a need for European monitoring or preliminary screenings or similar investigations with the aim of assessing the need for methods of a given compound or end-point for a given matrix.

Protocol V1 describes guidance for the within-laboratory validation of methods, parameters and criteria that are needed to establish a chemical or biological method at the research level (Section 3.2).

The key performance parameters that require attention during the within-laboratory validation vary according to the measurement requirement and method. Nevertheless, commonly important parameters are listed in tables 7, 8 and 9. These are based on the earlier described validation modules A (test method definition, documentation and general requirements), B (applicability domain and pre-validation), and C (intra-laboratory performance).

Module A focuses on the requirements of the method and the information about the method which is needed. These requirements are compared to the application domain of the method, which is described in Module B, and with the intra-laboratory performance characteristics described in Module C.

In the next sections more details on the information needed for each module are described.

7.1 Module A: Test method definition, documentation and general requirements

In this module (Table 7) general information on the methods should be provided such as:

1. External requirements
2. Title of the method
3. Beginning and end of validation procedure
4. Responsible party
5. Scientific basis of the method
6. Method definition
7. Requirements on devices, reagents, organisms, experimental conditions

The focus of the documentation should be on those capabilities of the method that were covered by the actual validation rather than the overall capabilities of a method.

Most of the parameters listed are easy to understand and short descriptions of the terms are provided. Some parameters need more attention, and these are discussed in more detail in the following sections.

Table 7 Requirements for test method definition, documentation and general requirements as part of a within-laboratory validation

	Module A - Test method definition, documentation and general requirements	
A.1	**External requirements**	
A.1.1	Aim and task	Specify the (pre-set) objectives of the method (measurement application for which the method is being considered)
A.1.2	Requirements and specifications	Documentation of the pre-set requirements e.g., in terms of target values for method performance: • target compound, organism or end-point • application range • matrix • measurement uncertainty If no requirements are pre-set, a brief description of how sensible ad-hoc requirements might be derived should be given.
A.2	**Title of the method**	Brief but unambiguous title, e.g. "Determination of volatile aliphatic and aromatic hydrocarbons in the range C_6 – C_{10} in waste water by pentane extraction using GC-FID").
A.3	**Beginning and end of validation procedure**	Start and end date
A.4	**Responsible party**	Institute or person, including full contact data
A.5	**Scientific basis of the method**	Description of the reaction and/or detection principle(s), if necessary supported by reaction equations and separation principles; for biological procedures: description of the physiological principles or effects; endpoint. Indicate whether a similar method exists and is used as a starting point (e.g., if the adaptation of a biomarker measurement to another species is to be validated).
A.6	**Method definition**	
A.6.1	Method description / SOP	(Bibliographic) reference (if applicable) or source where the detailed method description (with a degree of detail according to the respective validation level) can be obtained.
A.6.2.	Experimental setup	Requires a brief description, with no duplication of the method description or SOP
A.6.3.	Sample preparation and pre-treatment	Indicate whether a specific pre-treatment of the environmental sample is needed (e.g., sieving, centrifugation, filtration)
A.6.4	Sample measurement	Give a brief description of the sample measurement technique
A.6.5	Endpoint measurement	Give a brief description of the endpoint that is measured

	Module A - Test method definition, documentation and general requirements	
A.7	**Requirements on devices, reagents, organisms, experimental conditions**	
A.7.1	Instruments/devices	Type of measurement devices; specify any requirements on certain materials (e.g. specific separation phases) and instruments (e.g. resolution, sensitivity...)
A.7.2	Environmental conditions	Environmental conditions under which the test should be conducted (temperature, light, moisture), if this is of relevance to the method.
A.7.3	Test organisms	Restrictions on the application to specific organisms where relevant to the method, e.g. • food, food quality & other quality criteria for the test organisms (e.g., sex, maturity, age, weight, size); • time intervals / development stages at which organisms have been (or can be) used; • light regime.
A.7.4	Reagents	• Purity of applied reagents • Have specific requirements for the purity of reagents been identified? • Are in-house purification procedures for reagents necessary? • Influence on blank values
A.7.5	Medium / Matrix	The sample medium in which the (biological) test is conducted (e.g. water, sediment, soil) and whether an artificial medium is needed (e.g., reconstituted water, artificial sediment or soil). In this case, the composition of the medium should be described. In all cases, the physico-chemical characteristics of the medium are to be indicated (e.g., pH, hardness, water retention capacity of soils).
A.8	**Health and Safety**	
A.8.1	Health and Safety Information	Information on Health and Safety aspects of the method should be provided (if required).

A.1 – External requirements

External requirements are used to evaluate if the validation of the method is successful. A clear description of pre-set requirements and specifications of the method are therefore needed. Information on the aim of the method shall be provided, e.g. which compound, organism or end-point is measured at which concentration level, in which matrix. Furthermore, the application range of the methods shall cover the expected range of interest, and any requirements on the measurement uncertainties of data produced by the method shall be documented (if any exist).

A.5 – Scientific basis of the method

A short description of the detection principle, and for biological methods a description of the physiological principle, or end-points shall be provided. Reference to literature can also be used to provide more detailed information. If another method is used as the starting point for the validation, a reference to the method shall be given. For instance, if an existing biomarker method, is validated for a particular species, and is now validated for another species. When necessary for biological methods (i.e. for biomarker measurements), the definition shall include a description of the mechanistic basis of the test, including a short list of known

confounding factors/interferences and related effects (more details on this issue are provided in Module C). Where the method provides a measurement of a surrogate parameter, the relationship to the required analyte in the environmental context, shall be described.

A.6 – Method definition

In the method definition a bibliographic reference (if appropriate) or source of the detailed method shall be provided (A 6.1). In the following sections (A 6.2 to A.6.5), a brief description of the experimental setup, the main pre-treatment, sample treatment and analysis steps shall be given (e.g. sieving, extraction, end-point etc.).

A.7 – Requirements on the devices, reagents, organisms, and experimental conditions

This section shall provide information on the physical requirements of the method such as instruments, reagents and medium needed for the determination of an emerging pollutant. In addition, details of the organism or cell line needed for the determination shall also be given. The quality of reagents used is an important issue in relation to background levels of the target compounds. If certain reagents of a specific source or quality should not be used this shall also be mentioned. Reagents used in the collection of the sample (for example sorbents used for sampling the air compartment) shall be included in this Section.

7.2 Module B: Applicability domain and pre-validation

In this module the focus centres on the applicability of the validated method, and covers the target parameters, matrix and samples, and sampling. An overview of the information needed on the applicability domain is shown in Table 8.

Table 8 Requirements on "applicability" for a within-laboratory validation

	Module B – Applicability domain and pre-validation	
B.1	**Target parameters**	Detailed information on the parameters covered by the method (analyte/s or endpoint), information on additional parameters or excluded parameters (if necessary). Indicate whether all requirements from A1 are met/covered
B.2	**Matrix and samples**	
B.2.1	Type of matrix	Indicate the matrices for which the validation has been successfully performed, provide information on specific limitations of the method with respect to the matrix composition (e.g. "method only applicable to soils with less than 20% organic carbon"). Refer to A1. This may be updated depending of the results from module C.
B.2.2	Sampling	Refer to specific sampling procedures and precautions which have been applied to obtain the sample materials used in the validation process, including information on material of containers and sources of error, etc.
B.2.3.	Sample characteristics	Provide information on origin and main composition of samples used for the validation procedure(s); e.g. amount of organic carbon, fat, suspended particulate matter.
B.2.4	Sample stability and preservation, including transport	Describe all measures taken to stabilise/preserve the samples, give advice on suitable/unsuitable techniques, influence of sample storage, specific

	Module B – Applicability domain and pre-validation	
		requirements. If specific requirements are needed for sample transport, e.g. cooling of samples, these should be defined here.
B.2.5	Availability of the organisms (if relevant to the method)	• Season or period when the organisms are available for the test and/or measurement. • Possibility of breeding for biological tests, or from which supplier the organisms have been taken.
B.3	**Expandability of the method (optional)**	Indicate expected future fields of application (extending the applicability to other matrices or working ranges). Refer to A.1 if appropriate

B.1 - Target parameters

Detailed information is needed on the target parameters that are covered by the method. For a chemical method the chemical name (not the brand name) and if possible a CAS number shall be provided. For biological methods the end-point, effect or marker covered by the method should be provided. If information on additional parameters is needed, e.g. level of proteins, this information should also be given.

B.2.1 - Type of matrix

The type of matrix for which the method is validated shall be described, including information on the limitations of the method; e.g. matrix composition, if a method is only suitable for soils with less than 20% organic carbon, or the method is only suitable for drinking water.

B.2.2 - **Sampling**

Describe specific sampling procedures or precautions that should have been carried out to obtain the sample materials used in the validation process. Information on the material of containers used and sources of contamination, e.g. some target compounds may also be present in the sample containers or sampling equipment. The use of field and laboratory sample blanks shall be described if required by the method.

B.2.3 - Sample characteristics

Information on the origin and main composition of the samples used for the validation procedure(s) shall be provided.

B.2.4 - Sample stability and preservation, including transport

Describe all measures taken to stabilise/preserve the samples, if necessary. If available give advice on suitable/unsuitable techniques, the influence of sample storage, and specific requirements. Describe how samples should be transported. As most biological analyses (like biochemical measurements) cannot be determined on site, tissue or organ samples from fish or macroinvertebrates shall be kept frozen or preserved (e.g. ethanol, formalin), until analyzed (ISO 23893-1, EN 27828, EN 13946, EN 27828, EN ISO 16665). Issues of possible contamination and sample integrity should be documented.

B.2.5 - Availability of the organisms

For biological methods, the availability of organisms shall be described (if this is relevant to the method). Describe how to obtain the organisms (from laboratory stock or from a supplier). Information on the season or period of sampling of specific organism is important. Details of the breeding of organism for biological tests, or obtaining organisms from particular suppliers shall be provided where appropriate.

7.3 Module C: Intra-laboratory performance

In this module the intra-laboratory performance characteristics of the method are provided (Table 9), which are based on ISO 5725 and ISO 11843 standards and include:

- Trueness and bias
- Precision
- Calibration and traceability
- Linearity and Sensitivity
- Limits and application range
- Selectivity and specificity and interferences
- Uncertainty of measurement
- Robustness

Table 9 Requirements for intra-laboratory performance validation

	Module C - Intra-laboratory performance	
C.1	**Trueness and bias**	Describe the approach used to check trueness and bias of the method, and provide the result(s);
C.1.1	Reference materials	State the type of reference material(s) used; in-house or commercially available material (manufacturer); details on spiking solutions and spiked sample matrices, requirements on the uncertainty of the reference materials.
C.1.2	Reference substance(s)	For biological tests, give the name(s) of the reference substance(s) used as positive and negative controls
C.1.3	Recovery rates	How have recovery rates been determined? For what types of samples/matrices? What is the relation between concentration range and recovery rate?
C.1.4	Comparability with other methods	Provide results obtained with this method compared to another one (if there is one with the same endpoint), in order to compare the sensitivity of the developed/validated method.
C.2	**Precision**	Describe the approach used to check precision; provide results on intra-laboratory precision measures
C.2.1	Type of samples used for validation	For example real samples, reference materials, "synthetic" samples, reference substance(s)
C.2.2	Statistical evaluation	Describe the statistical tests were used to determine precision, trueness etc. For example range, mean and standard deviation of repeatability (control chart), identification of and treatment of outliers, check for normal distribution.
C.3	**Calibration**	
C.3.1	Type of calibration	The type of calibration used in the validation process shall be explained and justified (e.g. standard addition, internal/external standards...) The scope of the calibration should be described, ie which parts of the method (if any) are not covered by the calibration.

	Module C - Intra-laboratory performance	
C.3.2	Calibration substances	Give details on type, composition, origin and quality of substances used for calibration
C.3.3	Calibration data and function	Description of the handling of the raw data; How have the raw data been treated, e.g. - Evaluation of a calibration function according to ISO 8466-1 or -2? - Has homogeneity of variances been checked? - What type of calibration function has been used (linear, logarithmic, polynomial)?
C.3.4	Calibration stability	How has the stability of the calibration been checked? What are the results? Can recommendations be given on recalibration frequency?
C.4	**Traceability**	Are the calibration (or calibration standards) or spiking solutions traceable to national or international standards? if yes: provide details of the traceability chain. Describe how this relates to the traceability of the method as a whole. If not, describe the source of calibration.
C.5	**Limits and application range**	What are the lower (and probably upper) limits of application? How have they been determined? Where required express lower limits as quantification limits and detection limits.
C.6	**Selectivity, specificity and interferences, discriminative ability**	Check for interfering compounds and cross-reactivity for biological methods. Check for discriminative ability (if applicable to the method).
C.7	**Robustness**	Has robustness been checked, e.g. by systematic experiments (deliberate variation of specific parameters)? If yes, what have been the results? Indicate the most sensitive parameters and their impact (can be either qualitative, semi-quantitative of quantitative).
C.8	**Uncertainty of measurement**	How has the uncertainty of measurement been calculated? Which approach has been used? Provide the results.
C.9	**Final evaluation**	Are all requirements defined in A.1 met?

C.1 - Trueness and bias

In the context of this protocol, the term ‘trueness’ is used according to ISO terminology (cf. ISO 3534-1 and ISO 5725 series), and should not be confused with the term ‘accuracy’, which encompasses trueness and precision (for details see glossary in Chapter 12.1). In order to evaluate a method with respect to its trueness or bias, an accepted reference value (often referred to as “true value”) is essential. Ideally, the accepted reference value is established for a so-called certified reference material (CRM). For emerging pollutants that are to be monitored using methods validated at the V1 level, the availability of CRMs is unlikely. In the absence of a suitable CRM, consensus mean or median values of ring test samples are often used as an estimate of accepted reference values. At the V1-level, neither

CRMs nor ring test results may be available. As an alternative, spiking a sample with a known amount of analyte and analysing the sample before and after spiking offers a means of determining recovery. The recovery is then calculated as the difference between the measured concentrations in the spiked sample and in the unspiked sample related to the amount added to the sample. Be aware that other parameters in the matrix may combine with the added spike and produce a larger effect, or the reverse may occur and a smaller effect be noted (synergistic effect). These effects may be concentration dependent. The spiking level may influence the bias of the method when using this approach. Lower spike concentrations will give a larger bias and lower the trueness. Some guidance on typical recoveries as a function of the analyte concentrations is given below based on Huber (1998).

Analyte concentration or content in %	Typical recovery in %
0.01	90-107
0.001	80-110
0.0001	80-110
0.00001	80-110
0.000001	60-115
0.0000001	40-120

In general, 5 to 10 repeats of the spiked and unspiked samples should be measured to determine the recovery. A homogeneous batch of spiked sample material - a so-called internal reference material – can be prepared that can also be used for other steps in the validation process. The sample matrix should be very similar to or mimic the matrix type of that used when the method was developed; the spiking of tap-water to establish the trueness or bias of a method used for waste water would not be appropriate. Further guidance can be found elsewhere in the ICH documents (ICH 1995, 1996a,b).

Another way to estimate trueness is to compare the new method with a well-characterised reference method. As the V1 protocol focuses on methods at the research level this alternative approach is likely to be less useful.

The procedure outlined above is applicable to all chemical methods, i.e., for methods that do not use a biological effects of the analyte on a particular organism. In some biochemical methods, the trueness or bias of the method can be estimated. For example, for the estrogenic effect of an analyte (as measured with an assay) the calibration may be carried out using a solvent spiked with the analyte. Using this spiked solvent the relation between the measured effect and the amount of analyte can be determined (assuming the solvent plays no part in the measured effect). The trueness or bias can then be established by spiking a true sample with the analyte and measuring the effect. However, other parameters in the matrix may combine with the added spiked analyte and affect the measured response, produce a synergistic effect. These effects may be concentration-dependent. Additionally, for biological methods it is important to use positive and negative controls in parallel to the tested substance.

In biological systems, however, if whole organisms are used and measurements made of either individual or population related effects (e.g., mortality, growth, reproduction), the actual true or expected value may be difficult to determine (Johnson, 1994).

C.2 - Precision

Precision can be divided into repeatability (for example conditions like the same reagents, sample, analyst, laboratory being constant) and reproducibility (for example conditions like reagents, analysts etc being different). The latter can be subdivided into within-laboratory and between-laboratory reproducibility. At the V1 level, only repeatability and within-laboratory reproducibility is appropriate.

Precision can be estimated following repeated analysis of samples, preferably at different concentrations levels. In practice, the spiked sample used for the estimation of the trueness/bias (see C1) can be used for the precision determination, the average value of the outcome being used for the estimation of the trueness (or bias) and the variation for the estimation of the precision. A minimum of 3 repeats per concentration level is generally used. The repeatability standard deviation (s_r) and relative standard deviation (RSD_r) are determined. The repeatability precision (r) can be calculated by $r = 2.8 \times s_r$. (Taverniers et al. 2004). The calculated repeatability can be compared with existing methods, however, for emerging chemicals these are often not available. Therefore, the target value for the relative repeatability standard deviation (RSD_{target}, in %) can be calculated by using e.g. the modified Horwitz function:

$RSD_{target} = \exp(1 - 0.5 \log C)$

where C is the concentration of the analyte (in %).

Test results should, ideally, be independent. Very often the calibration is not independent. Ideally, a new calibration solution should be prepared from a different batch of the calibration standard used previously, in order to take into account variations in calibrant purity, weighing and diluting errors, etc.
Precision should also be established for biochemical methods, basically in the same way as for the chemical methods. Any method, chemical or biological, should have the agreement (precision) between repeated tests established, and expressed quantitatively.

C.3 - Calibration

In biological tests, uncertainties in the result can be observed between replicates due to the use of biological material. The method should then, as far as possible, specify biological factors that can have an impact on the measurement, for example factors such as fish size, weight or sex and species. At the V1 level these parameters should be listed in order to be taken into account at the next steps.
In biotests, comparison to a reference material should be used to evaluate sensibility level of the tested organism to the substance, as well as to control of temporal trends in the sensitivity of the tested organism to a reference substance.

C.3.3 - Calibration data and function – Linearity and sensitivity

Linearity
Under ideal conditions, linearity is a constant factor observed between the method response and the analyte concentration. For biological tests linearity is not strictly applicable as most of the biological responses are not linear (often sigmoid curves between concentration and response are observed) but a graduation of response should be established. Tests should have graduated response, rather than a total response or zero response, for the determination of ECx, ICx, LCx or index values. To date, calibration calculations are generally carried out using computer programmes that can manage both linear and non-linear functions. The linearity or working range should be established, mathematically.

One of the approaches used to determine linearity for chemical methods is to plot the response (e.g. signal divided by concentration) as a function of the concentration, on a log scale. The observed line should be horizontal. Often, a positive deviation for high concentrations and a negative deviation for low concentrations is observed. The linear range is between e.g. 95% and 105% of the horizontal response line. Linearity can be different for different matrices as the matrix can interfere with the detection system. Therefore, the linearity should be determined between the analytical standard calibration and also with sample calibration. It is more important to show reproducible curves rather than to show a wide linear range, as also non-linear functions can be fitted to the data, for example, as is the case with many biological tests.

When assessing the linearity of instrumental methods, the parameter related to linearity which is used in subsequent uncertainty calculations is the lack of fit. This is determined from the residuals of the fit of the calibration data to the calibration curve. See EN 14181 for a methodology to determine lack of fit.

The linearity or working range can be established by analysing analyte solutions possessing a wide range of concentration levels. This applies to chemical methods as much as to biological methods. For biological systems, dose-response curves often reach a plateau when the maximum effect is obtained.

Sensitivity
The sensitivity of a method is the change in the response of a measurand divided by the corresponding change in the stimulus (see Glossary). Stimulus may for example be the amount of the measurand present. Sensitivity is effectively the gradient of the response curve, i.e. the change in instrument response which corresponds to a change in analyte concentration. Where the response has been established as linear with respect to concentration, i.e. within the linear range of the method, and the intercept of the response curve has been determined, sensitivity is a useful parameter to calculate and be used in formulae for quantification.

Depending on the type of calibration function, sensitivity can be either a constant value or a more complex function of the analyte concentration.

C.3.4 - Calibration stability

The susceptibility of a method for instabilities of the calibration should be investigated. The level and frequency of re-calibration should be predefined and meet acceptable criteria. Results of robustness tests (see Section C.7) should also be considered when setting up the frequency of re-calibration. In many cases, the need for a re-calibration can be identified by periodic measurement of a limited number of standards or CRMs. In this section, it should be described how the stability of the calibration has been checked, and which QA/QC measures need to be taken to control calibration stability when applying the method.

C.4 - Traceability

Traceability is defined in VIM (2004) as the mechanism by which the result of a measurement may be related back to a primary reference through an unbroken series of calibrations, each of which has an assigned uncertainty.

C.5 - Limits and application range

For biological methods that express their results in terms of effects, the term 'concentration of analyte' can be replaced with 'level of effect'. Indeed, biological methods can not be used to determine a substance concentration, especially in the case of a mixture of pollutants, e.g. as might be the case in an effluent. For chemical methods, and for biochemical methods such as vitellogenin (vtg) detection, the limit of detection (LOD) can be defined according to ISO/DIS 13530 as three times the standard deviation of the blank sample (s_0):

$LOD = 3\ s_0$

Often the LOD can be measured by the determination of the analyte in a blank matrix, or by using a material containing a low-level concentration (near the expected LOD concentration). In this case, the expected LOD value may depend on the actual low-level concentration used.

Care should be taken when assessing results from techniques such as chromatography in which a low level discrimination threshold is built in to the method (i.e. when peaks below a certain size are not quantified). In such cases the standard deviation of the readings of a blank or zero sample will be artificially low, and will not relate to the actual LOD.

The limit of quantification (LOQ) can then be defined, more or less arbitrarily, as a fixed multiple of the limit of detection (LOD). This leads to a relation between the limits of detection and quantification (ISO/DIS 13530): LOQ is usually 3 times the LOD.

LOQ = 3 LOD

In order to verify the LOQ, spiked blank samples at this concentration level are often used. Establishing the limit of detection by analysing decreasing concentrations of pure analyte in solution until the signal disappears in the detector signal noise, is usually far too optimistic an approach and should not be used.

If no requirements on the LOD and LOQ values have been pre-set, at least the relationship between the concentration (or effect) level and its variance shall be established (and preferably be expressed in mathematical terms), which should enable future users of the method to decide up to which point the method can be deemed “fit for purpose”.

C.6 - Selectivity and specificity

Selectivity and specificity should be taken into consideration at the V1 level, albeit more in qualitative terms of the attention that has been paid to possible interferences than in terms of quantification.

Selectivity and specificity are both method performance characteristics that are difficult to quantify. It is necessary to establish that the signal produced by the measurement system, or other measured property, is actually attributed to the analyte, and not produced by accident or coincidence or due to the presence of chemically or physically similar compounds. The selectivity of a method can usually be investigated by studying its ability to measure the analyte of interest compared to specific interferences which have been introduced in the sample (i.e. those interferences thought likely to be present in samples or which from expert knowledge are known to be likely interferences for the method). Another indirect way is checking the trueness of a method (e.g. by analysing a CRM). However, often a CRM is not available for emerging chemicals, and a possible interference should also be present in the CRM. If an acceptable trueness or bias of the method can be demonstrated by analysis of CRMs that also contain representative amounts of interfering compounds, then the method should be specific and selective.

Where it is unclear whether interferences are present, the selectivity of the method can be investigated by studying and comparing this and other different, independent methods / techniques to measure the analyte concentration or effect. These definitions are less applicable for biological methods. One possibility may be to use spiked samples (see C1), and to add possible candidates for cross reactivity, and calculate the percentages of the active compound (as is used in some biochemical methods). If the exposure is a mixture of compounds, some biological methods are discriminating substances. An alternative way is to determine the percentage of false-positive observations for a minimum number of blank samples and/or to determine the percentage of false-negative observations for a minimum number of positive samples (e.g. samples known to contain the compound of interest).

Some chemical methods are more prone to interferences from matrix constituents than other methods. Some interferences will be known to the method developer, and the method should therefore be investigated by adding known amounts of suspected interfering compounds to samples comprising different matrices. Interference effects can lead to an increased response (indicating potential false-positive or increased results), due to signal enhancement, or may lead to a decreased response (indicating potential false-negative or decreased results), due to signal suppression.

Similar factors apply to biological methods that detect analytes. For biological methods that detect effects, the situation is more complex. If there is an effect enhancing or attenuating matrix component present, there is as much such an enhancement or attenuation in the sample as in the method response.

Some of the main factors leading to a misinterpretation of bio-marker tests (such as false-negative or false-positive results), and impairing a rigorous interpretation of biomarker measurement at higher biological level organisation (individual, population or community), are:

- Confounding non-chemical influences: temperature, nutritional state, reproductive condition etc. Such factors should be understood so that the response can be calibrated appropriately.
- Difference in biomarker response among the population of a species (due to geographical, genetic, exposure history, etc.) should be known or be minimal.
- The effect of a mixture of analytes on a biomarker response must be to correctly interpret the biomarker response, in terms of the exposure to or effect of the specific chemical or group of chemicals.
- The time dependence of the biomarker response with respect to the beginning of exposure should be known

C.7 - Robustness

The capacity of an analytical method to remain unaffected by small variations in environmental and/or operational conditions provides an indication of its reliability during normal usage. This can be tested using a systematic set of experiments that introduce small but deliberate changes to the experimental conditions of the method, and by observing (either in a qualitative or quantitative way) how these changes affect the final result by determining the relative standard deviations of e.g. the spike sample used in C1.

With regard to the deliberate changes that are introduced, these can be different instrumental settings, reagents, materials, amounts of sample material, exposure times, etc. Eventually, this approach should provide information about the most critical conditions that affect the performance and reliability of the method.

To examine the effect of the variation in the environmental conditions on the results, a “factorial design” approach could be applied as described in von Holst et al. (2001). The advantage of this approach is that information can be provided on which environmental / operational conditions significantly affect the results.

C.8 - Uncertainty of Measurement

Measurement uncertainty is an important parameter of a measurement, which is increasingly used as a comparable data quality objective. It aids the interpretation of results, allows the users of measurement to allowing comparison of different measurements, potentially obtained using different measurement methods. The concept of measurement uncertainty is closely linked to that of traceability in its metrological sense.

Strictly speaking, measurement uncertainty is a property of a measurement result – and should be calculated for each result. In practice, uncertainty calculations are often carried out for a method, and if these are undertaken with sufficient care and rigour then it may be assumed that future results, obtained using the method, providing the same care and rigour used previously are employed, will exhibit the same uncertainty. If this is the case, it is necessary to understand the scope and range of possible measurement scenarios for which the uncertainty calculation will be valid. This includes such parameters, amongst others, as the range of conditions (e.g., temperature and pressure) that may be encountered, the range of matrices, interfering substances, and range of analyte values.

Measurement uncertainty reflects the sum total of our understanding of how close that result may be expected to be with respect to the 'true' value of the measured quantity.

One key benefit of the evaluation of measurement uncertainty is that it can provide information on the key steps in the measurement procedure which have the most impact on the overall uncertainty of the result. Therefore, these key steps should be the focus of QA/QC efforts.

Uncertainty determinations can also provide important information for validation studies. An initial, preliminary uncertainty evaluation can identify those conditions and external parameters which should be varied during intra- and inter–laboratory studies in order to ensure that significant potential uncertainty sources are quantified and reduced in these validation studies.

It may therefore be worthwhile to carry out uncertainty analyses before and after the validation experiments, using the uncertainty evaluation to inform the design of the validation experiments, and then updating the uncertainty evaluation with results from the validation experiments. This cycle can continue, throughout the various levels of validation in this protocol, and indeed into the continuous use of the method, as QA/QC procedures (e.g., proficiency testing results) may be used to further refine an individual laboratory's uncertainty in the result it obtains using the method.

There are several guidelines on different approaches to estimate the uncertainty of measurement that may be applied in the context of this validation protocol, e.g. the ISO Guide 98 ('GUM'), Eurachem (2000) and Magnusson et al. (2003). One approach, which is mainly based on ISO/TS 21748, is described in the annex (see Chapter 12.2).

8 Protocol V2 – Basic External Validation (Expert Level)

The V2 validation protocol covers the scenario for which a method that has been successfully validated at the intra-laboratory level (V1 protocol) is to be applied at the level of expert laboratories. To this purpose, the method must be transferable to another laboratory. A test method may be regarded as being transferable if at least one other laboratory can produce similar results to the one that undertook the initial development (and successful internal validation).

The measure for the similarity of results between the laboratories and the level of acceptability can differ from case to case (depending on the type of method and measurand), and may also be prescribed by external authorities or stakeholders. The V2 validation protocol provides the tools and procedures necessary to demonstrate this basic transferability of a test method.

8.1 Method definition and description

At the V2 level, a more detailed method description is required than at the V1 level. The detailed and unambiguous method description should enable users with less expert knowledge than the research laboratory that has done the initial (internal) validation to perform the method in an appropriate way.

The method description from the V1 level (together with the information in the documentation templates A to C in sections 7.1 to 7.3) may be used as a starting point for the preparation of the method description at the V2 level.

Based on the outcome of the transferability study at the V2 level, the method description may be revised or refined, but only the requirements for V2 need to be applied. However, if the need or potential for a further development of the method to the V3 level is anticipated or foreseen, it may be appropriate to follow the instructions for method description for the V3-level.

Table 10 Requirements for the method description at the expert level

No.	Chapters, their content and the required degree of detail
1	**Title** The title shall express concisely and without ambiguity • the test objects to which the method can be applied, • the substances (analytes) or the effects to be measured, and • the nature or principle of the determination.
2	**Introduction** This is an optional element. Additional information (e.g. on the technical background or the reaction principle of the method) which might not be given in the title can be provided here. Furthermore, information on the "history" of the method with respect to its development can be provided.
3	**Warnings** If any of the reagents, samples or organisms used in the method is known to be dangerous to either human health or the environment, these hazards should be clearly identified here. Furthermore, appropriate precautions and safety measures should be described.
4	**Scope** This section should state succinctly the chemical or biological method and specify the matrices and test objects to which it applies

No.	Chapters, their content and the required degree of detail
5	**(normative) References** A list of references used in deriving, preparing or researching the method should be given. Furthermore, any standard methods which are referenced and to which the user is expected to have access should be listed.
6	**Terms and definitions** All terms should be defined by the research laboratory in order that other laboratories (e.g. a transferee laboratory) may sufficiently understand what is meant by them. Compliance with terminology of international standards is not mandatory, but it is highly recommended at this stage of validation.
7	**Principle** This clause indicates the essential steps in the method used, the basic principles and the properties of which use is made and, if appropriate, the reasons justifying the choice of certain procedures
8	**Reactions** Defined with sufficient detail in order that other laboratories (e.g. a transferee laboratory) may sufficiently discern what is meant by them
9	**Reagents, materials, organisms, media** In general, all reagents, materials and organisms (and the source of each) should be listed in section 9 and its sub-sections. With respect to the degree of detail it is recommended that the respective recommendations of V3 level be followed, although not all the information may be available at the V2 level. This applies to all following sub-sections of section 9
9.1	Products used in their commercially available form
9.2	Products to be prepared by the laboratory
9.2.1	Solutions of defined concentration
9.2.2	Test organisms
9.2.3	Nutrients and Food
9.3	Reference substances
10	**Apparatus** A description of the key equipment necessary for the correct application of the method is required. In general, minor items of equipment and apparatus should be apparent from the method description. Where specific characteristics or performance requirements for the apparatus and equipment are critical, these should be clearly stated. The use of diagrams should facilitate the visualisation of equipment configurations.
11	**Sampling**
11.1	Sampling procedure Key requirements and advice on sampling should be given where appropriate or appropriate references cited.
11.2	Preparation of the test sample All the steps in the preparation shall be stated (e.g., drying, crushing, sieving, etc.) together with appropriate information on the required characteristics of the sample thus prepared (e.g., particle size distribution, approximate mass). If necessary, details of any containers to be used for storage, and the storage conditions shall be given.
12	**Procedure** The procedure to carry out the method shall be described in sufficient detail to enable another user having a suitable technical background and expertise to carry out the method with an acceptable level of reproducibility.

No.	Chapters, their content and the required degree of detail
	The method should not be open to misinterpretation and should describe required quality assurance and quality control processes. These shall address areas identified as critical to the performance of the method. Any cited references shall be readily available and required reagents and equipment shall also be widely available. Ideally, all sub-clauses recommended at the V3 level should be addressed within this section, to provide a consistent format for the transfer of methods, even if for a specific method some sections will require only limited text or be marked not applicable at the V2 level.
12.1	Preparation of the test portion
12.2	Preparation of growth medium
12.3	Preparation of pre-culture and inoculum
12.4	Preparation of test batches
12.5	Blank Test or control batches
12.6	Incubation
12.7	Preliminary test or check test
12.8	Determinations, measurement or tests
12.9	Calibration If the method requires any apparatus to be calibrated, this shall be the subject of a separate sub-clause located at the most appropriate point in the “procedure” clause. This sub-clause shall describe in sufficient detail all necessary operations.
13	**Calculation** This section shall describe all issues of data treatment, including the procedures to calculate the final result. It shall also describe whether comprehensive data treatment procedures need to be performed prior to the calculation (e.g., plotting of growth curves or selection and/or correction of chromatographic signals or peaks). In particular, information shall be given on: • the units in which the result is to be expressed • the equation(s) used for the calculation.
14	**Interpretation of results** If the result of the method needs specific interpretation steps, guidance on the interpretation should be given here.
15	**Performance characteristics** This section should include any existing validation results from the V1 and the V2 level where appropriate, i.e. precision data, information on measurement uncertainty or comparison/transferability tests. For detailed information, reference should be made to the documentation of the validation work, which may be part of an annex (see section No 19).
16	**Quality assurance and control, validity criteria** Information should be provided on • measures to be taken by laboratories to ensure the equipment used remains under control • requirements for a laboratory quality system necessary to perform a proper transferability study • validity criteria which may assist in the further promotion of the method to the V3 level
17	**Special cases** Optional element. Information on special cases that has not been given in the preceding sections may be placed here.

No.	Chapters, their content and the required degree of detail
18	**Test report** This section should specify minimum requirements for reporting results which should facilitate an audit to be carried out.
19	**Annexes** Optional element. The annex should be used to provide supporting information, e.g. on the history of the method and its validation maturity.
20	**Bibliography** Informative references may be given at the point in the text at which they are referred to or (if there are many) in a separate bibliography at the end of the document

8.2 Module C: Intra-laboratory performance

The validation at the intra-laboratory level has been done at the V1 level. Nevertheless, the transferee laboratories should also carry out a basic internal validation of the method in order to participate successfully in the transferability study at the V2 level. This should be performed and documented according to the appropriate parts of the V1 protocol, in particular chapter 7.3 and its sections. As the transferee laboratories have to adhere to the method description provided by the organising laboratory, several sections of chapter 7.3 may be skipped by the transferee laboratories (e.g., calibration procedures and treatment of raw data will usually be prescribed by the method description).

8.3 Module D: Inter-Laboratory Transferability

An important step towards the implementation of a method for European monitoring purposes is the demonstration of its transferability to another laboratory. In the context of this guidance, a test method can be regarded as being transferable if at least one other (expert) laboratory can produce similar or better results to the one that undertook the initial development (and successful internal validation).

A transferability study is usually designed with the aim of minimising the effect of within-laboratory variation on the measures used to characterise and evaluate the performance characteristics of the test method. Provided that these effects of within-laboratory variation have been minimised successfully, the agreement of the results and method performance characteristics between the initial laboratory and the transferee laboratory is an indicator of the transferability of the method.

The measure for the similarity of results between the laboratories and the level of acceptability can differ from case to case (depending on the type of method and measurand), and may even be prescribed by external authorities or stakeholders. Nevertheless, the general approach of an inter-laboratory transferability study to demonstrate the basic transferability of a method is similar in all cases, and is described in the following sections of this chapter.

Table 11 provides an overview of the requirements for such a transferability study, of the information to be compiled, the tasks to be performed, and the type of results that should be documented. The structure of this table shall also be used as a template for the documentation of the validation process. In the following text, the sections given in the table are discussed in more detail, and guidance is given on minimum requirements or recommended procedures for specific aspects and tasks of a V2 transferability study.

Table 11 Module D – Requirements for the transferability study

Module D – Inter-Laboratory Transferability		
D.1	**General set-up of the transferability study**	
D.1.1	Organising laboratory	Who is responsible for the organisation of this transferability study? Does the organiser meet the requirements for such a study?
D.1.2	Participating laboratories	How many laboratories are involved? Contact details of the laboratories
D.1.3	Criteria for participation	What criteria for participation have been defined? Have they been fulfilled completely?
D.1.4	Time frame	Start and End dates of the inter-laboratory study (including the complete schedule, i.e. dates of the announcements, sample delivery, completion of analysis, submission of results, finalisation of the evaluation).
D.1.5	Number of investigated alternatives/options	Were any variants/options of the method investigated? How many (by how many pairs of participants)? In this case, a separate documentation and evaluation of each variant is necessary
D.2	**Training phase**	
D.2.1	Participants	With how many participants has the training phase been undertaken?
D.2.2	Type of sample material	What type of sample material was used for the training phase; refer to B2 if necessary?
D.2.3	Analysed parameters	Full scope of the method, or only a representative subgroup of compounds?
D.2.4	Examined concentration levels	How many concentration levels were examined? What were the expected or assigned values, what are the actual data from participants?
D.2.5	Standards	Have calibration solutions been provided? How many solutions or samples were used, and were concentrations known or unknown to the participants?
D.2.6	Evaluation	Criteria for successful participation in the training phase; number of participants that were not successful in meeting these criteria.
D.3	**Transferability study**	
D.3.1	Materials	How many materials were included in the study? Are all relevant matrices covered? Are all relevant compounds and the whole application range covered? Were materials provided by the organiser or prepared by the participants? What measures have been taken to ensure sufficient homogeneity and stability of the samples?

Module D – Inter-Laboratory Transferability		
D.3.2	Replicates	Number and type of replicates (e.g. known replicates or coded blind replicates) used in the study
D.3.3	Performance of the study	Has a supplementary standard material or sample been provided? Has a tolerance level for correctness of the calibration of the participants been pre-set? If so, what is it? How has transport & storage of samples been performed? What was the timeframe for carrying out the analysis?
D.3.4	Reference data	How have the reference values of the materials and target (or reference) values for performance measures been determined (e.g., from V1 results, or fitness-for-purpose requirements, or by use of model calculations)?
D.4	**Calculation of the Results**	
D.4.1	Data Pre-Treatment	Have invalid or obviously erroneous results been reported and eliminated? Have outlying results or laboratories been identified? How have outliers been identified and treated?
D.4.2	Calculation of the Results	Calculate and present the final results a) of each laboratory b) of the whole inter-laboratory study for each material or concentration level
D.5	**Evaluation of the Transferability**	
D.5.1	Trueness	Are the requirements for the trueness met? Does the method have a significant bias? Is the fitness for purpose put at risk by the bias?
D.5.1	Precision	Are the precision measures derived from the transferability study within an acceptable range (cf. pre-set requirements)?
D.5.3	Measurement Uncertainty	If requirements have been defined in terms of measurement uncertainty, can values obtained by the method fulfil the requirements for measurement uncertainty?
D.5.4	Application Range	Have other requirements on the method (as documented in templates A and B) been met by all participants (if there are any that are not covered by D.5.1 to D.5.3)? Are there other limitations to the use of the method that had not been foreseen at the lower validation level (e.g., exclusion of specific matrices, or applicability to a limited number of compounds in case of a multi-compound method)?
D.5.6	Conclusion	Final conclusion on the transferability of the method

8.3.1 General Set-up of the transferability study (D.1)

D.1.1 – Requirements for the organising laboratory

The organiser is responsible for a pre-announcement within reasonable time limits of the inter-laboratory comparison test and its conditions. The organising laboratory will be the laboratory that has developed the method and/or performed the initial internal validation according to the Validation 1 protocol. It is therefore the responsibility of the organising laboratory to provide a method description (protocol) which meets the requirements outlined in 8.1. The organising laboratory should also be a participant in the inter-laboratory comparison.

D.1.2 – Participating laboratories

The main requirement at this level of validation is to demonstrate the transferability of a method, i.e. that another laboratory can obtain similar or better results within defined limits. Therefore the minimum number of participating laboratories is two, but more are preferable. One of the participating laboratories should be the one having performed either the development of the method or at least the initial internal validation, preferably according to the procedure outlined in the V1 protocol.

D.1.3 – Criteria for participation

The laboratory attempting to demonstrate the transferability of the method shall have sufficient experience with the type of test method that is to be validated. It is difficult to define a set of universally applicable formal criteria to evaluate the expertise of a laboratory with respect to test methods for monitoring or bio-monitoring of emerging pollutants in general. At least one of the following criteria should be fulfilled by the candidate laboratory:

- appointed or commonly recognised as a (national) reference laboratory undertaking analyses using methods similar to the method to be validated
- recognised expertise in methods similar to the candidate method due to its own research and development activities in this field, proven by peer-reviewed scientific publications

The participating laboratories shall ensure that the method description (protocol) is strictly adhered to, and all technical conditions described in the protocol shall be fulfilled. Tools for statistical evaluation of precision and trueness measures (preferably according ISO 5725-2) shall be available.

As the main criterion for the selection of a participating laboratory at the present level of validation maturity is its excellence and experience with the type of test method to be validated, no compulsory requirements for the geographic location of the participating laboratory should be made. Nevertheless, if there are several options, a laboratory from a member state where the particular emerging pollutant is an issue should be favoured.

D.1.5 – Number of investigated alternatives or options

In order to enable a transferability study of a method between two laboratories to be carried out, it is essential that both laboratories perform the method in an identical way. If optional or alternative procedures are to be validated as well, it shall be ensured that each of the options or alternatives (e.g. different combination of an extraction or derivatisation technique) is covered by at least two laboratories, with one laboratory being the one that has done the initial V1 validation.

8.3.2 The training phase (D.2)

D.2 – Training phase

The conduction of a training phase before the actual transferability study is highly recommended. This training phase is a precondition for a successful performance of the participating laboratories in the study. The objective of the training phase is to enable the participating laboratories to gain sufficient experience to be able to perform the test method adequately and successfully. Furthermore, the training phase can reveal whether the participating laboratories actually have sufficient expertise with the test method to be validated. However, at this stage of validation maturity, a failure of the participant(s) in the training phase may also indicate an insufficient performance of the method itself, or deficiencies in the method description.

Depending on the application range of the test method or the range of effects of the chemical, at least two samples of different concentrations (in the lower and upper range of application or effect) should be examined by the participating laboratories in the training phase. The training phase can be accomplished either with spiked sample material, or with standard solutions (provided by the organiser), which are to be added to a sample matrix by the participant(s). It is not sufficient to perform a training phase with standards only. In this phase, calibration standards should also be provided by the organising laboratory. Together with the exercise samples, information on the concentration of the analyte(s) and target values for the performance characteristics (e.g. precision), or on the expected effect shall be provided. In the training phase, the organising laboratory shall be prepared to willingly give advice on technical details of the method upon request from the participant(s). Requests for advice shall initiate a review of the method description with respect to whether it is complete and unambiguous.

8.3.3 The transferability study (D.3)

D.3.1 – Materials: selection, preparation and pre-testing of samples

A single chemical standard in a specified (simple) matrix may be adequate to demonstrate transferability. The sample matrix shall be representative of the intended application of the method, i.e. it shall be of the same type as the matrix used for the internal validation at the V1 level.

As a minimum requirement, a single concentration known to give the required response in the V1 validation level may be sufficient. However, it is recommended to use at least two concentration levels (close to the lower and the upper limits of application and/or effect). If a multi-compound test method is to be validated, the type and number of compounds in the sample(s) shall be representative for real scenarios to which the method will be applied in a regulatory environmental context for monitoring or bio-monitoring of emerging pollutants.

The samples can be provided by the organising laboratory or prepared internally by the laboratories (e.g. by adding standard solutions to matrix samples, or by exposing the organisms to the studied chemical). If samples are to be provided by the organising laboratory, it is the responsibility of this laboratory to provide information on the homogeneity and stability of the samples. If necessary, the samples shall be preserved and stabilised in a way that assures their homogeneity and stability up to the agreed period of study. Homogeneity and stability testing of the samples should be carried out to recognised procedures, for example ISO 13528 Annex B; ISO 5667-16 or IUPAC 2006, Appendix I & II. The sample quantity or amount shall be adjusted to satisfy the required number of tests to be carried out (see section D.3.2) with a sufficient margin of safety.

For biological materials, two different scenarios need to be considered:

- studies are conducted on bio-indicators
- or organisms are exposed to chemicals in field or laboratory experiments

In the first case, sampling is part of the method, and each participating laboratory should collect its own biological material on a single selected contaminated site. In the second case, each participating laboratory should conduct the test with its own biological material. Otherwise, for example if it has been shown that there are sensitivity differences between clones of the same animals or cells, the organising laboratory can provide the biological material in order to reduce the variability of the results.

D.3.2 – Replicates

The minimum number of replicate measurements is dependent on the number of participating laboratories and materials analysed. Due to the relatively small number of participants in a transferability study (usually n=2 or slightly more), a large number of replicate measurements is required at the V2 level to reduce the uncertainty component originating from within-laboratory variability.

The most common practice is to use known replicates, i.e. the participating laboratories are requested to perform the analysis of the same material in several independent runs. Guidance on the relationship between the number of participants, repeat measurements and the resulting uncertainty is given in the ISO 5725 series of standards and in the associated ISO TS 21748. However, these standards do not consider collaborative studies with fewer than 5 laboratories. Additional guidance is therefore given in the following. The main requirement for the number of replicates is that the uncertainty of the intra-laboratory repeatability study is small enough to enable inter-laboratory bias to be observed. For a single laboratory attempting to replicate a method, the number of replicates should be chosen such that the following criterion for the resulting uncertainty is fulfilled:

$$\sqrt{\frac{s_W^2}{n}} < 0.2 s_R \qquad (1)$$

where s_W^2 is the intra-laboratory standard deviation of the n repeated measurements of the reference material and
s_R is the method reproducibility standard deviation, determined by either collaborative study or estimated at the V1 validation stage.

In a limited study, the uncertainty of the estimate of s_R may be very high (particularly in the case of studies with fewer than 5 participants) and it may be more appropriate to replace s_R by a criterion based on the target uncertainty of the method. In this case, the equation to estimate the required number of replicates can be modified as follows:

$$\sqrt{\frac{s_W^2}{n}} < 0.2 u \qquad (2)$$

where u is the (target) uncertainty of the method, but without a coverage factor.

If the target uncertainty has been given as an expanded uncertainty (U), this value should be divided by 2 to obtain u.

Where V1 data are available, these may be used to plan the study, using s_W and s_R (or equivalent uncertainty information) determined by the V1 laboratory in Equation (2) to calculate n. However, participating laboratories should also demonstrate that they have carried out sufficient repeats to meet these requirements for their own value of s_W.

If only one material is investigated and repeatability was the principle source of uncertainty in the method, up to 25 replicate measurements may be necessary in a V2 transferability study.

D.3.3 – Performance of the study

As in the training phase, the participating laboratories should be provided with supplementary standards to evaluate the calibration. A tolerance level for the correctness of the calibration shall be pre-set in advance of the study. In the case of multi-compound methods and standards, a minimum number of compounds that have to be calibrated, identified and quantified correctly shall also be defined in advance for detection methods.

The transport and storage of the samples before analysis shall conform to the respective recommendations outlined in the method description (protocol).

A fixed time period for carrying out the measurements and reporting all results shall be agreed. Whilst specific reporting forms for the recording of results and experimental details are advantageous, they are not essential due to the (usually) low number of participating laboratories in a transferability study.

The reporting format of the results should be defined, and data aggregation (e.g. calculation of a mean) should not be carried out by the participants.

D.3.4 – Use of reference data

An essential requirement for undertaking a successful transferability study is the availability of reliable reference data, preferably supported by quantitative information on the degree of reliability or variability (e.g. in terms of a standard uncertainty). For a method transferability study, there are two types of reference data:

a) reference data related to the tested materials
b) reference data related to the method

Reference data related to the materials used in an inter-laboratory study can be determined in several ways (see Section E.3.4 in Chapter 9.3.3), for instance using of certified reference values or consensus values from expert laboratories. However, at the V2 level, usually only the following approaches may be possible (owing to the lack of certified reference materials and existing data from expert laboratories):

1. Values derived by the originating laboratory (at the V1 or even the V2 level)
2. Formulation and calculation from the amounts or quantities used

The selection of the most appropriate approach will depend on the requirements and characteristics of the method that is to be validated. In any case, the selection should be agreed by the participating laboratories in the transferability study, and should be justified and documented. If a value generated by the originating laboratory (usually the laboratory that has done the validation at the V1 level and/or is the organiser of the transferability study) is used as a reference value, this value shall be based on a sufficient number of repeat tests (see Sections C.1, C.2 and D.3.2 for guidance). Since a transferability study may involve materials that are different from those used in the internal validation at the V1 level, the reference values according to the first option (i.e. option 1) will usually be determined during the V2 transferability study.

As regards reference data related to the method, the performance characteristics obtained by the originating laboratory, or values obtained on reference chemicals for biological methods, shall be regarded as reference values at this level of validation maturity.

8.3.4 Calculation of the Results (D.4)

D.4.1 – Data pre-treatment

In a first step, all reported data shall be screened for invalid or obviously erroneous data (e.g. data that are obviously impossible to obtain, either for logical, technical, chemical or biological reasons, for example negative concentrations, mortality rates > 100%, values outside the measuring range of the instrument). Such invalid data shall not be included in the calculation of the results.

If a sufficient number of participants (≥ 5) have submitted valid results, procedures for the identification and elimination of outliers may need to be applied (for example see Section E.4.1 in Chapter 9.3.4).

D.4.2 – Calculation of the Results

The following statistics shall be calculated and tabulated for each material.

a) Results of each laboratory

- Number of valid replicate measurements
- Laboratory mean
- Within laboratory standard deviation

b) Results of the transferability study

- Number of valid results
- Number of laboratories (after outlier elimination)
- Number of eliminated outliers (if any)
- Number of eliminated laboratories
- Mean
- Assigned or reference value and its standard uncertainty, if known (see D.3.4)
- Recovery rate (Ratio of mean to assigned value, in %; may not be applicable for biological methods) or bias
- Repeatability standard deviation (s_r), usually the arithmetic mean of the within-laboratory standard deviations
- Repeatability relative standard deviation (RSD_r)
- Reproducibility standard deviation s_R
- Reproducibility relative standard deviation (RSD_R)

Guidance on calculations is given in the ISO 5725 series of standards and in ISO TS 20281. Owing to the likelihood of the low number of participating laboratories at the V2 level (n=2 or slightly more), several of the statistical measures described in the above standards may not be able to be calculated or will be associated with a high degree of uncertainty and may therefore not always be suitable for drawing firm conclusions.

Alternative procedures for the calculations and statistical analyses have been developed (Cofino et al., 2000; De Boer & Cofino, 2002) and have several advantages in that

- they make use of the uncertainty of the individual laboratory data,
- the procedures are more robust for outliers and skewed data distributions than the ISO methods, and
- they can cope with multimodal distributions.

For either approach, the statistical procedures need to be documented in detail.

8.3.5 Evaluation of the Transferability of the Method (D.5)

Any transferee laboratory needs to demonstrate that it can obtain results that

a) are similar to, or better than, the results obtained by the originating laboratory
b) conform to the external or pre-set requirements for the method.

In particular, this should be demonstrated for the accuracy (trueness and precision) and application range of the method. In most cases, the closeness of the results to those obtained by the originating laboratory can be checked by selected statistical tests, e.g. t-test for trueness and F-test for precision measures with, for example a level of significance $\alpha = 0.05$. Depending on the type of method and the nature of the resulting data, other statistical tools such as e.g., the analysis of variance (ANOVA), or Chi-square test, may also be required. The type of statistical test to be applied in any given situation will depend mainly on the form of the specific variability measures, which in turn depend on the approach that has been selected to determine the reference values and the related uncertainty. It is therefore critical that the organiser of the V2 study has sufficient expertise available to enable the correct selection and application of the most appropriate statistical tests to be made in any given situation.

Compliance with the external or pre-set requirements can usually be checked by comparing the performance characteristics derived from the results (e.g., bias or precision at a given concentration level) with the requirements.

D.5.1 – Trueness

For each material (i.e. for all matrix / compound / concentration level combinations), it shall be evaluated whether the results from the transferee laboratories are significantly different from the reference values (see Section D.3.4). If, for a given transferee laboratory, the statistical tests indicate a significant difference between the mean of the replicates of a given material and the reference value(s), it should be evaluated whether the fitness-for-purpose of the method is put at risk by this bias. If the trueness (or bias) of the transferee laboratory is still within acceptable limits with regard to the intended purpose (e.g., compliant with the pre-set requirements), this shall be clearly documented.

D.5.2 – Precision

The same principle as for evaluation of the trueness (Section D.5.1) should be applied in evaluating the precision data of the method. In a first step, the precision of the transferee laboratories should be compared to the values obtained by the originating laboratory for each material.
In a second step, a comparison with pre-set requirements (if exist any) for the precision of the method (e.g., as documented in Section A.1 of Template A) should be carried out. This can usually be undertaken without applying any statistical tests, but instead as a simple decision whether the value of the respective precision measure is larger or smaller than the required precision.

D.5.3 – Measurement Uncertainty

If requirements for the method have been defined in terms of target values for measurement uncertainty (e.g. for a specific concentration level), the measurement uncertainty should be recalculated, e.g. as described in ISO/TS 21748, to provide the input from the results of the transferability study. This approach revisits the uncertainty sources determined in validation stage V1 (see section C.8 in Chapter 7.3), and replaces these with those that have been addressed by the transferability study. To compare the MU to pre-set requirements, it is

important that the MU be expressed at the same stated level of confidence. If the preset requirements do not explicitly state otherwise, this should be assumed to be with a level of confidence of 95%, implying a coverage factor of k=2.

D.5.4 – Application range

It may be the case that the required similarity of results or compliance of performance characteristics with the requirements may not be achievable for all compounds or concentration levels that shall be covered by the method. If this is the case, the actual application range and applicability domain for which the transferability has been shown to be successful shall be documented in this section. Any limitations to the use of the method that had not been anticipated or documented at the V1 level but which becomes apparent during the V2 transferability study shall be documented in this section. Evaluation of other performance criteria (such as limits of detection and quantification) should have been carried out at the V1 level. Nevertheless, after the completion of the transferability study at the V2 level, it should be checked whether the validation results from V1 are consistent with the results from the V2 study.

D.5.6 – Conclusion

The results of the transferability study (in particular the information in Sections D.4.1 to D.5.5) shall be summarized and evaluated with regard to the transferability of the method.

If only a partial transferability has been achieved (e.g., for a limited application range or only some of the investigated compounds or matrices), the internal validation data of the transferee laboratories should be checked for discrepancies to the V1 data obtained in the originating laboratory. It should also be checked whether the unsuccessful parts of the transferability study are due to insufficiencies in the method description.

Any limitations with regard to the desired applicability domain or method performance shall also lead to an update of the respective information in Templates A and B.

In general, a method can be regarded as validated at the V2 level when the results of at least one of the transferee laboratories conform to the pre-set or external requirements for the method. This can even be the case when statistical tests indicate that the results are significantly different from the results obtained by the originating laboratory. This may be the case for instance when the numerical values of the precision measures are small compared to those of the trueness measures. Nevertheless, significant differences between the results of the laboratories indicate that considerable limitations to data comparability may exist. It should therefore be checked whether a development of the method from the V1 level to the V2 level and ultimately to the V3 level is reasonable or justifiable.

8.4 Documentation, record-keeping and publication of data

The organiser of the transferability study shall be responsible for recording the results of this study and any associated information. Based on the results from the transferability study and the feedback from the participating laboratories, the method description shall be revised where necessary. The results of the validation should be published, preferably in electronic form on a web-server that the research, expert and routine laboratories involved in the respective monitoring task have access to. The templates for modules A to D shall be used for documentation and publication of the validation activities and the results.

If the development of the validated method from the V2 level to the V3 level is desired, appropriate measures should be taken to initiate an inter-laboratory study according to the Validation 3 protocol.

9 Protocol V3 – Inter-laboratory Validation (Routine Level)

The V3 validation protocol covers the completion of the external validation of a test method. If a method is intended to be used at the level of routine laboratories, the variability aspects of a method across a number of routine laboratories needs to be fully evaluated. This is carried out by means of an inter-laboratory study with the focus on method validation. This inter-laboratory study shall be performed under conditions that are representative for monitoring a pollutant by routine laboratories, in order to enable a realistic assessment of the method performance under routine conditions.

The inter-laboratory validation at V3 level also encompasses an evaluation of the usability of the method. This requires a close examination of aspects such as “ease of use”, robustness, completeness of the method description, and sufficient description of the required QA/QC measures.

9.1 Method definition and description

The objective of the V3 validation protocol is to facilitate and accelerate the validation and establishment of methods that are not only fit for the desired purpose but also suitable for harmonisation or standardisation across Europe. An important pre-condition for a method to be applicable to routine laboratories, and to become a European standard, is a detailed and unambiguous method description. The method description should ensure clarity in procedural detail and minimise or eliminate the risk of misinterpretation.

Detailed requirements on the structure and degree of detail of a method description at V3 level are defined in Table 12. These requirements mainly originate from ISO 78-2, a standard that defines the requirements of international standards describing chemical test methods. The requirements from ISO 78-2 have been modified to make them also applicable to biological methods. The numbering of chapters in Table 12 serves as an example only and may be changed if chapters have to be merged or omitted.

The method description from the V2 level (together with the information in the documentation templates A to D) should be used as a starting point for the preparation of the method description at the V3 level.

Based on the outcome of the transferability study at V3 level, the method description should be revised as necessary.

Table 12 Requirements on method description at the routine level

No.	Chapters, their content and the required degree of detail
1	**Title** The title shall express clearly and unambiguously (i) the test objects to which the method can be applied, (ii) the substances (analytes) or the effects to be measured, and (iii) the nature or principle of the determination.
2	**Introduction** Additional information on the technical content of the method description or any other background information on the method (or its “history” with respect to the development of the method) should be included in this chapter
3	**Warnings** If any of the reagents, samples or organisms used in this method are known to be hazardous either to human health or to the environment, these hazards shall be clearly identified here. Appropriate precautions and safety measures shall also be

No.	Chapters, their content and the required degree of detail
	described.
4	**Scope** This section should state succinctly the chemical or biological method and specifically the test objects to which it applies. If applicable, it shall state the detection limit and/or the limit beyond which the method can no longer be relied upon. The information in this section should enable the user to judge quickly whether the method is applicable to the task or purpose for which it is intended, or whether certain restrictions exist. These restrictions shall take into account the potential presence and extent of other components in the types of samples to be investigated, and of their limiting contents. Relevant information regarding possible interferences shall also be provided. If it is necessary to provide modifications to the basic method e.g., to ensure the elimination of certain interfering factors, these modifications should preferably be treated as special cases. These special cases shall be indicated in the "Scope" clause, and the corresponding modifications shall be described in the "Special Cases" clause (see chapter 16).
5	**References** This clause shall list those references which are necessary for the proper application of the method. Documents that have served as references in the preparation of the method description should be listed in the bibliography, at the end of the document.
6	**Terms and Definitions** This clause shall give any definitions of terms used in the text that facilitate its understanding. At this level of method validation, the terminology should as far as possible conform to the terminology of European or international standards (CEN, ISO), and reference should be made to existing ISO or CEN definitions.
7	**Principle** This clause indicates the essential steps in the method used, the basic principles and the properties of which use is made and, if appropriate, the reasons justifying the choice of certain procedures.
8	**Reactions** If knowledge about the essential reactions is necessary to understand the method description or for the calculation of the results, these reactions shall be indicated here (supported by reaction equations, if possible). Reactions can be (bio)-chemical reactions or physiological effects/mechanisms
9	**Reagents, materials, organisms, media** This section shall list (with a sequential reference number) all reagents, materials, organisms and media used during the test, together with their essential characteristics (concentration, density, species, strain etc.). In addition, this section shall specify, if necessary, their degree of purity (for chemicals) and/or other relevant details such as the sex and age of organisms. If they exist, Chemical Abstract Service Registry numbers (CAS numbers) of all chemicals should be given. If necessary, any precautions and conditions to be taken / applied in storing the reagents or holding acclimating organisms, and the time period for which they may, or should, be stored / acclimated, should also be specified. All necessary preliminary test procedures (e.g., to verify the absence of an interfering component in a reagent, or to verify viability of a culture or batch of organisms) should also be defined and described in this section
9.1	Products used in their commercially available form In the list of reagents, materials, organisms and media, products used in their commercially available form shall be described unambiguously, giving the particulars necessary for their identification (e.g., the chemical name, the chemical

No.	Chapters, their content and the required degree of detail
	formula, the concentration, the CAS number). For organisms (if standardised cultures or strains are to be used), it may be appropriate to provide contact data for suppliers from which the required cultures or strains can be obtained.
9.2	Products to be prepared by the laboratory
9.2.1	Solutions of defined concentration The concentration of all solutions which are to be prepared by the laboratory shall be given in an unambiguous form. Solvents for the preparation and/or dilution of solutions shall be clearly defined. Requirements on the quality and/or purity of solvents shall also be defined. If a solution is prepared by dilution of another specified solution, the conventions outlined in ISO 78-2 how to describe the dilution procedure shall be observed.
9.2.2	Test organisms Provide all relevant information on the organisms that are to be used. If organisms need to be collected or sampled by the investigating laboratory, give reference to the sampling section. Provide unambiguous taxonomic information on the organisms, information on specific subspecies or strains (if necessary), information on size, age, sex, maturity/development stage or any other criteria which are critical for the performance of the test and which shall be fulfilled by the organisms used. Specific information on the maintenance of the cultures or organisms, minimum acclimation periods and conditions, maximum tolerable storage times and conditions, requirements on the frequency of sub-culturing should also be given (if applicable).
9.2.3	Nutrients and Food All nutrients needed for maintaining (cultures or batches of) test organisms should be completely and unambiguously defined. If nutrients are needed in the form of dilute solutions these can be described in section 9.2.1. If pure salts are used to prepare nutrient (stock) solutions, the exact sum formula (including water of crystallisation, if necessary) of the substance to be used shall be given. If special food needs to be prepared (e.g. for higher organisms) this should be described in detail, with exact composition, source, or procedure of preparation.
9.3	Reference substances Any reference substances or reference materials that are required or recommended should be listed, and appropriate details given, see chapters 9.1 and 9.2
10	**Apparatus** This clause shall list (with a sequential reference number) the names and significant characteristics (e.g. material properties) of all the apparatus and equipment (other than standard laboratory apparatus) to be used during the analysis or test. If appropriate, reference shall be made to existing European or international standards e.g., concerning laboratory glassware and related apparatus, or to other relevant international standards or internationally acceptable documents. It is advisable to illustrate, by means of a diagram, special types of apparatus and to indicate the way in which they are assembled. Special requirements on any apparatus that is critical to the method shall be given in this section, especially if they play a significant role in the procedure or if they constitute an important factor in the safety, precision and/or trueness of the method. Pre-treatment or cleaning procedures of the apparatus should also be described in this section. Any checking of the functioning of the (assembled) apparatus shall be described in the “Procedure” section, preferably in a sub-clause titled “preliminary test” or “check

No.	Chapters, their content and the required degree of detail
	test".
11	**Sampling**
11.1	Sampling procedure For many occasions, it may be sufficient to refer to the relevant European or international standard dealing specifically with the sampling. If no appropriate standard exists, the sampling clause may include a sampling plan and a sampling procedure, giving guidance on the following issues: - how to obtain a representative sample that can be used for the intended test method - how to avoid or minimise undesirable changes occurring to the sample - required minimum number, mass or volume of sample(s) - sampling equipment - handling of samples - characteristics and material of the containers for sample collection and storage
11.2	Preparation of the test sample. This clause shall give all relevant information necessary for the preparation of the test sample from which the test portions will be drawn. This test sample is usually prepared from the laboratory sample or field sample as specified in 11.1. For details on the sample terminology (Laboratory sample, test sample, test portion) see ISO 78-2. In each case, all the steps in the preparation shall be stated (e.g., drying, crushing, grinding, sieving etc.) together with appropriate information (e.g., particle size distribution, approximate mass or volume) on the required characteristics of the sample thus prepared. If necessary, details of any containers used for storage, and the storage conditions shall be given.
12	**Procedure** The "procedure" section may be divided into as many sub-sections or clauses as there are operations or sequences of operations to be carried out. Each operation or sequence of operations shall be described unambiguously and concisely. If the number of steps in the procedure is large, it is recommended to use subdivisions in the sub-clauses (point numbering system), with each element corresponding to a given operation and including all indispensable preliminary operations. If the method or a specific sequence of operations within the method is already given in a European or international standard, this should be indicated. In such cases, it may be sufficient to indicate modifications of or deviations from the standard operations. If there are risks or hazards during the procedure for which special precautions are necessary, a statement shall be included at the beginning of the clause. If necessary, more detailed advice on safety procedures and first-aid measures can be given in an annex.
12.1	Preparation of the test portion Describe how the test portion is prepared from the test sample (or the laboratory sample, if the two are the same). It shall state the method of determining the mass or volume of the test portion (e.g., weighing). It shall state the mass or volume or amount of other discrete units (e.g. number of cells, organisms), and the tolerance with which this needs to be measured.
12.2	Preparation of growth medium Describe how the growth medium needs to be prepared. All components of the medium should have been described in section 9; and specific reference to the

No.	Chapters, their content and the required degree of detail
	relevant sub-section of section 9 should be given for any components used in the preparation of the growth medium. If it is essential that a specific sequence of operations needs to be carried out in the preparation of growth medium, this should be clearly described. All relevant details on any other treatment steps, such as equilibration times, autoclaving, sterilisation, filtration, or stabilisation measures as well as storage times and conditions etc should be given.
12.3	Preparation of pre-culture and inoculum If necessary, all relevant information on cultures of organisms required to carry out a test shall be described. This information shall specify pre-culture conditions (medium, duration, physico-chemical parameters such as temperature, initial concentration, and organism generation). If required, control of activity or of the quantity needed to start the test shall be described (e.g. absorbance control to evaluate cell density).
12.4	Preparation of test batches Techniques to prepare the test batch shall be described, e.g. information shall be given whether a controlled quantity or specific size or age of the test organisms are required to initiate the measurements.
12.5	Blank test or control batches Indicate whether a blank test is necessary or advisable to verify the purity of the reagents or the cleanliness of the laboratory environment or apparatus. If this is the case, this sub-clause shall indicate all the conditions for carrying out this blank test. The blank test should usually be carried out in parallel with and under the same conditions as the actual determination, following the same procedure, using the same quantities of all the reagents and using the same apparatus as in the determination, but without any test portion.
12.6	Incubation The conditions of the test incubation shall be clearly described, in particular all the information allowing the biochemical reactions or the development of the organisms (physico-chemical conditions, temperature, type of vessel, light, duration…)
12.7	Preliminary test or check test If it is necessary to perform any preliminary checks e.g., of the apparatus or the viability/vitality of the (culture of) test organism(s), all details necessary to carry out these checks should be given in this sub-clause.
12.8	Determinations, measurement or tests Each sequence of operations shall be described adequately and unambiguously. The test shall be set out in an easily readable form in suitable sub-clauses and paragraphs, in order to facilitate the description, the understanding and the application of the procedure. If the product resulting from one of the steps is to be retained and used as a test portion in a later procedure, this shall be clearly stated and identified..
12.9	Calibration If the method requires any apparatus to be calibrated, this operation shall be the subject of a separate sub-clause located at the most appropriate point in the “procedure” clause. This sub-clause shall describe all necessary operations to be carried out in detail, including requirements on traceable reference materials and calibration artefacts. The frequency of calibration and QA/QC criteria for the calibration (e.g., acceptability criteria or performance criteria) shall also be defined in this sub-clause. If several steps in the calibration procedure are identical to those of the determination procedure, one of the two sub-clauses shall make reference to

No.	Chapters, their content and the required degree of detail
	the other in order to avoid the duplication of redundant information.
13	**Calculation** This section shall describe all issues of data treatment, including the procedures to calculate the final (reported) result. If comprehensive procedures of data treatment need to be performed prior to the calculation (e.g., plotting of growth curves or selection and/or correction of chromatographic signals or peaks), detailed guidance on these steps shall be given. If the application of complex procedures like sophisticated mathematical or statistical models is required (e.g., fitting a non-linear model by regression analysis), reference can also be made to external sources where these procedures are described in detail (preferably references which are wide-spread and easily accessible (e.g., international standards). In particular, information shall be given on: - the units in which the result is to be expressed - the equation(s) used for the calculation - the meaning of the algebraic symbols used in the equation(s) - the units in which all used quantities are expressed - the number of decimal places or significant figures to which the result is to be given
14	**Interpretation of results** If the result of the method needs specific interpretative steps (which may be the case e.g. for toxicological data such as EC_x values), guidance on the interpretation should be given here.
15	**Performance characteristics** This section should include information on all performance characteristics of the method derived from validation work at all three levels (V1, V2 and V3). For detailed information, reference should be made to the documentation of the validation work, which may be part of an annex (see section 19).
16	**Quality assurance and control, validity criteria** This section should provide a full description of - expected QA/QC procedures - verified validity criteria - control measures and remedial actions to take if these measures indicate that the method is not under control.
17	**Special cases** Essential information on special cases that has not been given in the preceding sections may be placed here.
18	**Test report** This section should specify in detail the reporting requirements for the method which fully describe the results and supporting QA/QC information enabling an audit trail to be carried out (see the requirements of ISO 17025:2005).
19	**Annexes** Optional element. The annex should be used to provide supporting information, e.g. the completed forms A to E as a history of the method and its validation maturity.
20	**Bibliography** References may be given at the point in the text at which they are referred to, or in section 5, or in a separate bibliography at the end of the document. Recommendations of ISO 690 should be followed.

9.2 Module C: Intra-laboratory performance

The validation at the intra-laboratory level has been done at the V1 level. Nevertheless, the participating laboratories should also carry out a basic internal validation of the method in order to participate successfully in the transferability study at the V2 level. This should be performed and documented according to the appropriate parts of the V1 protocol, in particular chapter 7.3 and its sections. As the participating laboratories have to adhere to the method description provided by the organising laboratory, several sections of chapter 7.3 may be skipped by the participating laboratories (e.g., calibration procedures and treatment of raw data will usually be prescribed by the method description).

9.3 Module E: Inter-laboratory performance

In order to be applicable for large-scale European monitoring programmes, a test method for emerging pollutants shall show sufficient inter-laboratory performance (preferably across Europe) at the level of routine laboratories.

The principal tool used to evaluate the inter-laboratory performance of a method is an inter-laboratory comparison involving the analysis of identical test items across all participating laboratories. A collaborative study to evaluate inter-laboratory performance at the V3 level requires a considerably higher number of participating laboratories (and a broader geographical coverage) than the investigation of an inter-laboratory transferability. These V3 level inter-laboratory studies are usually designed with the aim of minimising the effect of within-laboratory variation on the measures used to characterise and evaluate the performance characteristics of the test method. Details of the tools and procedures to establish the value of measures for inter-laboratory performance criteria may be different depending on the type of method and measurand. However, the general approach is similar in most cases, and is outlined in this chapter and its sections.

The focus of an inter-laboratory study at this level of validation is to validate the method and to assess its applicability by routine laboratories, and not to evaluate the proficiency or capability of the participating laboratories. Nevertheless, the objective of such an inter-laboratory performance study requires the integration of a number of elements from proficiency testing schemes. Therefore, some of the references that are given in this chapter deal with the design or evaluation of inter-laboratory trials for proficiency testing.

Table 13 provides an overview of the requirements on such an inter-laboratory study, on the information to be compiled and on the tasks to be performed. The structure of this table shall also be used as a template for the documentation of the validation process. The sections given in the table are discussed in more detail, and guidance is given on minimum requirements for specific aspects of a V3 inter-laboratory study.

Table 13 Module E - Requirements on the study for inter-laboratory validation

Module E - Inter-laboratory performance		
E.1	**General set-up of the inter-laboratory study**	
E.1.1	Organising party	Who is responsible for the organisation of this inter-laboratory study?
		Does the organiser meet the requirements for such an inter-laboratory study?
E.1.2	Announcement/Dissemination	How was the information on this study disseminated?
		Which measures have been taken to address a representative cross-section of routine-laboratories

Module E - Inter-laboratory performance		
		that are active in the specific field?
E.1.3	Participating laboratories	Number of laboratories involved Contact data of the participating laboratories
E.1.4	Criteria for participation	How many of the V1 and V2 laboratories are involved?
		How many countries are involved?
		Does this reflect a cross section of those states where the particular pollutant that is to be measured by the validated method is an issue?
		Have other criteria / conditions for participation been pre-set, and were they fulfilled?
E.1.5	Time frame	Start and end dates of the inter-laboratory study (including the complete schedule, i.e. dates of the announcements, sample delivery, completion of analysis, submission of results, finalisation of the evaluation)
E.1.6	Number of investigated alternatives/options	Were any variants/options of the method investigated?
		How many (by how many participants)?
		In this case, a separate documentation and evaluation of each variant is necessary
E.2	**Training phase**	
E.2.1	Participants	With how many participants?
E.2.2	Type of sample material	What type of sample material was used for the training phase; refer to B2 if necessary?
E.2.3	Analysed parameters	Full scope of the method, or only a representative subgroup of compounds?
E.2.4	Examined concentration levels	How many concentration levels were examined? What were the expected or assigned values, what are the actual data from participants?
E.2.5	Standards	Have calibration solutions been provided?
		How many solutions or samples with concentrations known or unknown to the participants?
E.2.6	Evaluation	Criteria for successful participation in the training phase; number of unsuccessful participants
E.3	**Inter-laboratory study**	
E.3.1	Materials	How many materials are included in the study? Are all relevant matrices covered?
		Are all relevant compounds and the whole application range covered?
		Are materials provided by the organiser or prepared by the participants?
		What measures have been taken to ensure sufficient homogeneity and stability of the samples?
E.3.2	Replicates	Number and type of replicates (e.g. known replicates or coded blind replicates) used in the study

Module E - Inter-laboratory performance		
E.3.3	Performance of the study	Has a supplementary standard been provided? Has a tolerance level for correctness of the calibration of the participants been pre-set? How has transport & storage of samples been performed? What was the timeframe for carrying out the analysis?
E.3.4	Reference data	How has the assigned value been determined?
E.4	**Statistical analysis and calculation of the results**	
E.4.1	Statistical Analysis	What statistical tools & approaches have been used for the statistical analysis of the data? Have outlying results or laboratories been identified? How have outliers been treated?
E.4.2	Calculation of the results	Calculate and present the final results a) of each laboratory b) of the whole inter-laboratory study for each material or concentration level
E.5	**Evaluation of the fitness-for purpose**	
E.5.1	Trueness	Are the requirements on the trueness met? Does the method have a significant bias? Is the fitness for purpose put at risk by the bias? Can the bias be traced to a particular issue, i.e. can it be shown to be systematic?
E.5.1	Precision	Are the precision measures derived from the inter-laboratory within an acceptable range (cf. pre-set requirements)?
E.5.3	Measurement Uncertainty	If requirements have been defined in terms of measurement uncertainty, can values obtained by the method fulfil the requirements on measurement uncertainty?
E.5.4	Application range	Are the other requirements on the method (as documented in templates A and B) met by all routine laboratories (if there are any that are not covered by E.5.1 to E.5.3)? Are there stronger limitations to the use of the method that had not been foreseen at the lower validation level (e.g., exclusion of specific matrices, or applicability for a limited number of compounds in case of a multi-compound method)?
E.5.5	Usability	How many outlying results and/or outlying laboratories have been identified? Is the method fully applicable (with results meeting all requirements) by the majority of the participating routine laboratories?
E.5.6	Conclusion	Final conclusion on the fitness for purpose of the method

9.3.1 General set-up of the inter-laboratory study (E.1)

E.1.1 – Organizing Party

The inter-laboratory study shall be organised by specialists familiar with all the requirements relating to the design, execution and evaluation of inter-laboratory tests, and also with the method to be tested. The organising party shall be independent, impartial and clearly identifiable legally, and shall not operate in any way that might prejudice the assessments of the method to be validated. If such an interest can not be ruled out, it shall be ensured that at least the evaluation of the data is performed by an independent party, and that any potential conflicts of interest are clearly declared and defined. The same requirements shall apply to subcontractors who may deal with subordinate tasks, such as sample preparation or data evaluation. The overall responsibility for the inter-laboratory test shall remain with the organiser subcontracting any task.

The premises, staffing and equipment of the organising party shall meet the requirements of the fields covered by the inter-laboratory test. In particular, the organising party shall have accommodations and equipment that meet all the inter-laboratory test requirements of this protocol. The equipment for preparing test samples and for performing measurements to determine the assigned value (including its standard uncertainty) and the data processing equipment shall meet the requirements of the test method to be validated.

The proficiency test provider shall provide evidence of the trueness of the assigned values with respect to traceability of measurement results to national and international standards, and of the determination of measurement uncertainties.

With respect to the conduction of inter-laboratory studies, a quality management manual based on the criteria of ISO/IEC 17025 shall be available, and requirements of ISO Guide 43-1 shall be met.

The results of the validation studies performed according to the V1 and V2 protocols as outlined in this document (chapter 7 and 8) shall be available to the organising party.

E.1.2 – Announcement and dissemination

Potential participants should be informed about the intended inter-laboratory study by a general pre-announcement. This pre-announcement should contain details about the method to be validated, the analytes, matrices, organisms or effects to be examined and the anticipated time-scale.

Dissemination of the pre-announcement should be performed in a way to ensure a European-wide perception of the inter-laboratory study, especially in those member states where the particular emerging pollutant is considered to be an issue, and requires some form of monitoring.

Subsequently, a detailed announcement should be sent to all interested parties who have reacted on the pre-announcement by stating their interest in the study. This announcement should take place within reasonable time-scale prior to the planned delivery of samples and standards for the training phase (see comments on E.2).

Information provided at this stage should contain contact details of the organiser, the protocol of the method to be validated, and registration modalities. In addition, information about the exact extent of the study (e.g., number of samples or sample types to be investigated, concentration range etc), and a schedule of the study and its phases should also be included.

E1.3 – Participating laboratories

The inter-laboratory study shall only be performed if a sufficient number of laboratories can participate. The number of participating laboratories increases the reliability of the statistically based conclusions. The minimum number, n, of valid data sets required for a statistically reliable analysis of the full scope of the inter-laboratory variability of a method is usually recognised as being $n \geq 8$. In order to allow for potential outliers and those laboratories unable, for whatever reason, to submit data, it is recognised that n should be at least in the range of 10 - 12. For guidance on the relationship between the number of participants and replicate measurements, see comments on E.3.2.

E1.4 – Criteria for participation

The laboratories participating in an inter-laboratory study at the V3 level of validation shall be laboratories involved in routine monitoring or bio-monitoring of environmental pollutants. These laboratories shall work according to accredited schemes (e.g. ISO/IEC 17025) and/or possess comparable national independent and impartial certification for analytical methods similar to the method to be validated.

The participating laboratories shall assure that the method description (protocol) is strictly adhered to, and all technical conditions described in the protocol are fulfilled. Tools for statistical evaluation of precision and trueness measures (preferably according ISO 5725-2) shall be available. Participating laboratories shall disclose, at the latest during the training phase (see chapter 9.2), existing information on the internal validation and use of the method to be validated at the V3 level.

Participating laboratories should represent a cross section of member states where the particular environmental pollutant is considered an issue.

9.3.2 Training phase (E.2)

E.2 – Training phase

The inclusion of a training phase before the actual inter-laboratory study is highly recommended. This training phase is a pre-condition for the successful performance of the laboratories participating in the study. The objective of the training phase is to enable the participants to gain sufficient experience to enable the adequate and successful performance of the test method. According to the application range of the test method, at least two samples of different concentration (in the lower and upper range of application) should be examined by the participating laboratories in the training phase. The training phase can be accomplished either with spiked sample material, or with standard solutions (provided by the organiser), which are added to a sample matrix by the participant(s). If available, the use of (certified) reference materials shall be preferred. It is not sufficient to perform a training phase with standards only.

In this phase, calibration standards can be provided by the organiser. Together with the exercise samples, information on the concentration of the analyte(s) and on target values for internal performance characteristics (e.g. precision data) shall be provided. At least one sample with a concentration unknown to the participants should be included. In the training phase, the organising laboratory shall be prepared to provide technical support for the method upon request from the participant(s). Requests for advice shall immediately initiate a review of the method description with respect to the clarity and coverage of the protocol. If, in the training phase, laboratories do not achieve the required internal performance characteristics of the method, this suggests a problem within the method. In this case, sufficient efforts should be made to rectify the problem prior to the actual inter-laboratory study.

9.3.3 The inter-laboratory study (E.3)

E.3.1 – Materials: selection, preparation and pre-testing of samples

A number of five materials shall be investigated. A "material" is a compound/concentration level/matrix combination. If blind duplicates are used as replicates (see E.3.2), these shall be considered as one material only (they are not independent). Also blanks or negative controls in a given matrix shall be considered as a "material".

At the V3 level, the inter-laboratory variability with respect to all potential routine applications of the method shall be evaluated. The study shall therefore encompass all compounds and matrices for which the method is intended to be used. At the V3 level, the study shall be performed with samples that are representative of the actual sample composition under realistic routine conditions, i.e. at this level it is not sufficient to work with simplified matrices.
Furthermore, all of the intended application range (from the lower to the upper concentration limit) shall be covered. If pre-set requirements exist with respect to the lowest concentration that is to be determined (within a certain target error) at least one material in the inter-laboratory study shall cover this concentration. The same principle shall be applied if there are any requirements or target values for an upper (concentration) limit. If a multi-compound test method is to be validated, the type and number of compounds in the samples shall be representative of actual scenarios in which the method will be applied in an environmental context (e.g. for monitoring or bio-monitoring of emerging pollutants).

The samples shall be provided by the organising party (in case of quantitative chemical methods) or prepared internally by the participating laboratories (the latter approach may be reasonable for certain methods including the exposure of organisms to a specific compound). If samples are to be provided by the organiser, information on homogeneity and stability of the samples shall also be provided. If necessary, the samples shall be preserved and stabilised in a way that ensures homogeneity and stability up to the agreed period of the study. Homogeneity and stability testing of the samples should be undertaken by the organiser, e.g. according to [ISO 13528 Annex B; ISO 5667-16 for aqueous samples tested with biological methods] or [IUPAC 2006, Appendix I & II]. The sample quantity or amount should be adjusted to accommodate the required number of analytical and/or biological replicates (see E.3.2).

For biological materials, two different scenarios need to be considered:

- either studies are conducted on organisms sampled in the field (e.g. for biomarker assessment)
- or organisms are exposed to chemicals in field or laboratory experiments

In the first case, sampling is part of the method and each participant has to collect its own biological material on a single selected contaminated site. In the second case, each participant can conduct the test on its own biological material, with its own specificity. Otherwise, for example if it has been shown that there are sensitivity differences between clones of the same animals or cells, the organising laboratory should provide the biological material in order to reduce variability of results.

E.3.2 – Replicates

The most common practice is to use known replicates, i.e. the participating laboratories should be requested to perform the analysis of the same material in several independent runs. In this case, a minimum of three replicate measurements should be made.
An alternative strategy to obtain precision data is to repeat the measurement of randomly coded blind duplicate (or triplicate) test samples. In this case, only one measurement per test sample is required, and it is more effective to utilize resources for the analysis of more levels and/or materials rather than for increasing the number of replicates for the individual material.

The minimum numbers of participants and replicate measurements are dependent on the acceptable uncertainty of the estimates for repeatability and reproducibility standard deviations. These estimates can differ considerably from their true values if only a small number of laboratories (e.g., ≈ 5) take part in the collaborative study. An increase in the number of laboratories by 2 or 3 yields only a small reduction in the uncertainties if the number of participants is already larger than 20.

More detailed guidance on this issue is provided in the ISO 5725 series, in particular in ISO 5725-1 and ISO 5725-4. These documents provide equations and tables that can be used to adjust the number of replicates and participants to the specific requirements.

E.3.3 – Performance of the inter-laboratory study

As in the training phase, each participating laboratory should be provided with a supplementary standard to evaluate the calibration carried out within each laboratory. At this validation level, the concentration of the standard shall not be disclosed to the participants. A tolerance level for the correctness of the calibration shall be pre-set in advance. In case of multi-compound methods and standards, a minimum number of compounds that have to be calibrated, identified and quantified correctly shall also be defined in advance.
Transport and storage of the samples before analysis shall be compliant with the respective recommendations outlined in the method description. A fixed time-scale for carrying out the measurements shall be agreed. Forms for the transmission of results and experimental details shall be provided by the organising party.

E.3.4 – Reference data

One of the most essential requirements for a successful validation study is the availability of a reliable assigned value, preferably supported by some quantitative information on the degree of reliability or variability (e.g. in terms of a standard uncertainty).
The assigned value at the V3 level can be determined in one of five ways [cf. ISO 13528 and ISO Guide 43-1 A1.1], as described in order of priority:

1) certified reference values;
2) reference values (measured in a real sample matrix, calibrated against a certified reference material);
3) consensus values from expert laboratories (e.g. from transferability study at V2 level);
4) formulation and calculation from the amounts used;
5) consensus values from participants (e.g. mean of the valid results from all participants at V3 level, preferably calculated by robust statistical algorithms).

Guidance on the estimation of the standard uncertainty of the assigned value for the approaches given above can be found in ISO 13528 and in chapter 7.3 (section C.8). If an assigned value from option number 5 above is used, then it will not be possible to investigate a systematic bias of the method.

9.3.4 Statistical analysis and calculation of the results (E.4)

E.4.1 – Statistical analysis

There are several appropriate ways for a statistical evaluation of the inter-laboratory data. Recognising the broad range of methods that may be applied in the monitoring and bio-monitoring of emerging pollutants (or their effects) it is not possible to cover all scenarios by a single protocol.
Therefore, the basic principles of the most common approaches are outlined, and some guidance is given on the advantages and disadvantages of the presented approaches.

Calculation of accuracy in terms of precision and trueness measures should follow the recommendations of the ISO 5725 series, in particular ISO 5725-2. These “classical” statistical procedures are widely disseminated and accepted, and are applicable to data from a wide range of method types. The ISO 5725 series of standards also provides guidance on statistical outlier tests.

Robust statistical methods, such as the algorithms given in ISO 5725-5, ISO 13528 or the Q-method outlined in ISO/DIS 20612, may also be applied to calculate some of the statistical measures given below. These methods have the advantage of being less influenced by single outliers or anomalous results. Therefore, less effort is required for outlier testing when applying these algorithms.

An alternative approach (Cofino et al., 2000; De Boer & Cofino, 2002) is based on the concept that the data points are replaced by so-called laboratory measurement functions (LMFs), which are used to calculate inter-laboratory measurement functions (IMFs). This approach has a number of advantages that include:

i) the method makes use of the uncertainty of the individual laboratory data,
ii) the method is more robust than the ISO standards for outliers and skewed distributions of data, and
iii) it can cope with multi-modal distributions.

A disadvantage is that these alternative statistical tools are more complex and not as widely disseminated or recognised as the ISO series of standards. In any case, the statistical procedures applied shall be documented in detail.

For methods of quantitative chemical and biological analyses, the following approach (which is mainly based on the procedures of ISO 5725-2) may be appropriate (see also Figure 3).

For each (concentration) level the following evaluation steps should be performed:

Step	Procedure
1	Pre-Screening for (and elimination of) invalid (obviously erroneous) data, e.g. data outside the range of the measuring instrument or data which are impossible for logical, technical, chemical or biological reasons (e.g. mortality rate > 100%, negative concentration)
2	Preliminary calculation of laboratory mean, standard deviation and number of replicate measurements (at a single laboratory level)
3	Check for single outliers at laboratory level (e.g. by Grubbs’ test, usually applying the 1% critical value for rejection)
4	If outliers have been identified: remove outliers and re-enter loop at 2. Otherwise proceed to 5.
5	Preliminary calculation of the mean and the standard deviation of the laboratory means.
6	Check the laboratory means for outliers (e.g. by Grubbs’ test); Remove those outliers
7	If outliers have been identified and removed, re-enter the loop at 5. Otherwise proceed to step 8
8	Check the within laboratory standard deviations for outliers (e.g. by Cochran test), and remove those outliers.
9	If outliers have been identified and removed: re-enter the loop at 5. Otherwise the calculation of the results can be performed (see section E.4.2)

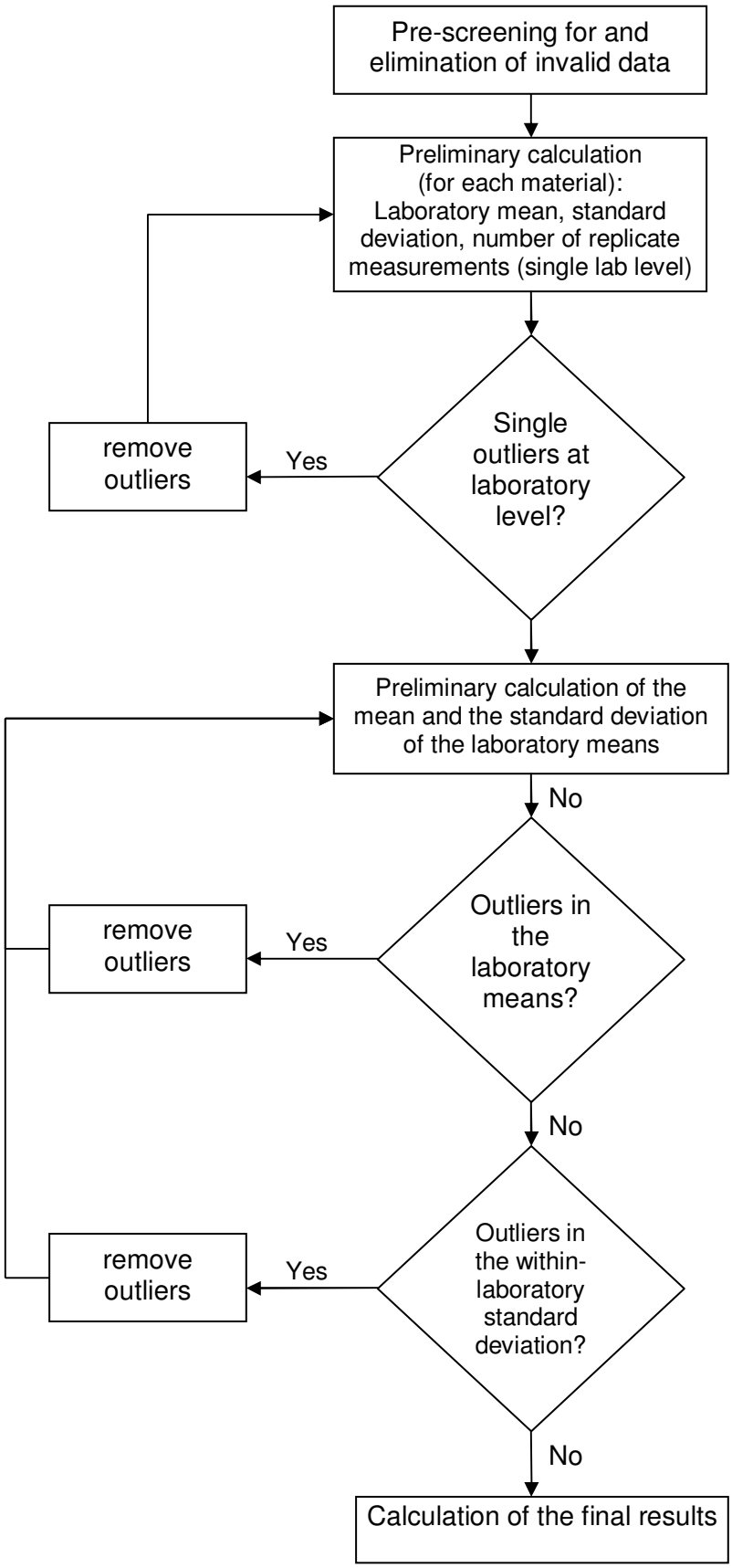

Figure 3 Flowchart of the main steps of the statistical analysis

E.4.2 – Calculation of the final results

a) Results of each laboratory

The following results shall be calculated for each laboratory (and each material)

- Number of valid replicate measurements
- Laboratory mean
- Within-laboratory standard deviation

b) Results of the inter-laboratory study

The following data should be calculated and tabulated for each material. Detailed guidance for calculation is given in the ISO 5725 series.

- Number of valid results
- Number of laboratories after outlier elimination
- Number of eliminated outliers
- Number of eliminated laboratories
- Mean
- Assigned value and its standard uncertainty, if known (see comment on E.3.4)
- Recovery rate (Ratio of mean to assigned value, in %; may not be applicable for biological methods) or bias
- Repeatability standard deviation (s_r), usually the arithmetic mean of the within-laboratory standard deviations
- Repeatability relative standard deviation (RSD_r)
- Reproducibility standard deviation s_R
- Reproducibility relative standard deviation (RSD_R)

9.3.5 Evaluation of the fitness for purpose (E.5)

E.5.1 – Trueness

The trueness of the method shall be evaluated in order to investigate the potential for systematic bias in the method. Statistical tools can be used to compare the mean (and its variability measures; usually the reproducibility standard deviation s_R) to the assigned value (and its variability measures). The type of statistical test to be applied in any given situation will depend mainly on the form of the specific variability measures, which in turn depend on the approach that has been selected to determine the assigned value and its uncertainty.
It is therefore critical that sufficient expertise in the selection and application of the appropriate statistical tests to be applied in any relevant situation is available to the organiser of the study.

If the statistical test indicates a significant difference between the mean from the validation study and the assigned value, it should be evaluated whether the fitness-for-purpose of the method is put at risk by this bias. If the bias is within acceptable limits with regard to the intended purpose, this should be clearly documented. Otherwise, the method in question fails to fulfil the requirements at the V3 level, and needs to be improved by modification or optimisation of the some or all of the procedures.(which means a downgrading of the method to the V2 or even the V1 level).

The evaluation given above shall be performed for each concentration level investigated in the inter-laboratory study. Requirements on a method are often expressed for a specific minimum concentration level, above which the method shall fulfil the respective criteria. The lowest concentration level at which the method fulfils the requirements on trueness (or bias) shall be clearly identified.

E.5.2 – Precision

The same principle for evaluating the trueness (Section E.5.1) should be applied in evaluating the precision data of the method, which should have been defined in advance (and documented in the respective template of module A, section A.1). The reproducibility standard deviation, s_R, and where appropriate, also the repeatability standard deviation, s_r, shall be compared with the requirements on precision measures. Usually this can be carried out without applying any statistical test (simply by observing whether the respective standard deviation is larger or smaller than the required precision).

E.5.3 – Measurement uncertainty

If requirements on the method have been defined in terms of target values for measurement uncertainty (e.g. for a specific concentration level), the measurement uncertainty (MU) should be recalculated, as described in ISO/TS 21748 to provide the input from the inter-laboratory results.

This approach revisits the uncertainty sources determined in earlier validation stages (e.g., V2), replacing these with those that have been addressed by this V3 inter-laboratory study. For example, if the inter-laboratory study can be considered to have covered a suitable range of conditions for a given influence factor (for example an extraction stage in the analysis of a sample) which has been determined individually in a sensitivity study, then this component of the MU can be excluded as it will have been covered within the trueness and precision studies of the inter-laboratory studies. The resultant calculation then reduces to the equation given in ISO TS 21748 Section 5.3 which in simple terms combines, as a sum of squares, the reproducibility standard deviation calculated from the terms determined in ISO 5725-2 for the collaborative study with any terms addressing uncertainty sources not covered within the scope of the inter-laboratory study.

If the inter-laboratory studies have fully covered the potential sources of uncertainty, and the calculation of MU has properly estimated the range of influence quantities, then it may be expected that the uncertainty observed in a collaborative inter-laboratory study will be less than that previously determined in the V1 level. This is because it is likely that the real range of influence factors observed during a specific test will probably not encompass the assumed ranges of influence quantities used for the calculation in V1.

If the MU determined using the collaborative study results is significantly greater than the V1 uncertainty, then it may be indicative of the presence of a source of uncertainty not previously considered. In this case, it is recommended that further laboratory studies be undertaken to identify and include this missing 'uncertainty' in the intra-laboratory based uncertainty calculation. Similarly, if the inter-laboratory study results in a significantly smaller MU than that determined in V1 it may be that the study did not include the variation of a significant influence factor, and it should be confirmed that all potential sources were either varied in the study or have been included in the overall calculation of MU as additional terms.

To compare the MU to preset requirements it is important that the MU be expressed at a stated level of confidence – if the preset requirements do not explicitly state otherwise this should be assumed to be with a level of confidence of 95%, implying a coverage factor, k, of k=2.

E.5.4 – Application Range

Evaluation of other performance criteria (such as limits of detection and quantification, application range etc) should have been carried out at the lower validation levels. Nevertheless, after the completion of the inter-laboratory study at V3 level, it should be checked whether the validation results from V1 and V2 are consistent with the results from

the V3 study. It may be the case that the required performance may not be achievable by all the routine laboratories. If this is the case, this may indicate a clear limitation on the applicability of the method at the level of routine laboratories. Therefore, discrepancies between the internal validation data of the participants and the data obtained at V1 and V2 levels should be listed in this section, and should be regarded as a clear indication for a limited usability of the method at routine level.

E.5.5 – Usability of the method at the routine level

If the respective ratio of outlier values (single laboratories) or outlying laboratories is more than 25% of the total data (or of the number of participants) this may indicate that the method is not yet applicable at the level of routine laboratories. Reports where participating laboratories were unable to submit results or to comply with the requirements of the method shall be included in the calculation of the outlier ratio.

Depending on the specific statistical methods that have been used to calculate the statistical data, outlying values or laboratories with insufficient performance may not have been eliminated, and therefore no number or ratio of eliminated values or laboratories may be available. In this case, a calculation of z-scores (or z_u-scores) should be performed (according to ISO 13528). Error target values are needed to calculate the required laboratory z-scores, and these should be derived from either the pre-set requirements on the method or using (uncertainty) data that have been generated in V1 and refined in V2 activities. The latter approach is to be preferred, especially in cases where there is significant uncertainty in the reference material used for the study. This can be taken into account by including a term related to the uncertainty of the reference material in the target standard deviation.

If necessary, it may also be possible to calculate error targets using an appropriate model, e.g. the Horwitz function for quantitative chemical methods (see section C.3 in chapter 7.3). In general, z-scores outside the range -2 to 2 should be used to provide an indication of the ratio of eliminated data. This approach is also applicable to biological methods provided that the assigned value(s) and error targets are derived from a source independent of the current study.

E.5.6 – Final conclusion

A statement on the fitness for purpose of the method shall be given in this section, summarising any relevant restraint and limitation of the method. The results of the inter-laboratory study (in particular the information in Sections E.4.1 to E.5.5) shall be summarized and evaluated with regard to the fitness for purpose of the method (on the basis of the preset requirements).

If only a partial validation of the method has been achieved (e.g., only for a limited application range or only a part of the investigated compounds or matrices), this shall be documented in this section. Any limitations with regard to the desired applicability domain or method performance shall also lead to an update of the respective information in Templates A and B. If such limitations exist, the internal validation data of the participating laboratories should be checked for discrepancies to the V1 and V2 data. Furthermore, it should be checked whether some of the limitations are due to any insufficiencies in the method description, in order to enable a targeted refinement of the method or the method description and eventually a recurrence of the validation activities where appropriate.

9.4 Documentation, publication and standardisation

The organiser of the inter-laboratory is responsible for the controlled record-keeping of the documents and results of this study.

Based on the results from the inter-laboratory study and the feedback from the participants, the method description shall be revised where necessary. The results of the validation should be published, preferably in electronic form on a web-server to which laboratories involved in the respective monitoring task(s) have access to.

If standardisation of the validated method is desired, the organiser of the inter-laboratory study should contact its respective national representative in the appropriate CEN or ISO technical committee, in order to check the possibilities for launching the method as a new work item proposal.

10 Sampling and handling of samples

Sampling is a crucial step in the whole analytical process, and sometimes an inherent element of the test method itself. Several steps in the validation process depend on the proper application of suitable sampling methodologies, e.g. for the preparation of test materials (reference materials) for recovery experiments and inter-laboratory studies. However, it is not the aim of this document to provide procedures for all issues related to environmental sampling, but rather to outline guidance on the key issues related to sampling the main environmental areas and matrices. Moreover, the aim is to present reliable references which provide guidance and further details for specific sampling tasks.

Preference has been given to those references which have undergone a thorough international process of review, harmonisation and dissemination. Each of the main chapters addressing a specific matrix or compartment contains a table which can be used as a quick reference for specific sampling issues.

The recommendations and guidance presented in this chapter (and the references recommended therein) are often of general nature. The properties (e.g. volatility, sorption, stability) and potential sources of contamination of the analyte(s) under consideration must always be taken into account, and this may result in procedures for some pollutants that differ considerably from the recommendations below.

10.1 Sampling of biota

Table 14 Quick reference table for biota sampling issues

Compartment		**Problem / issue / task**	**Find guidance in**
General	*specific*		
Freshwaters	Running rivers	Fish sampling, handling, preservation	ISO/DIS 23893-1 chapter 4.4 and 5.3
		Macrophyte sampling	EN 14184 chapter 7
		Benthic diatoms sampling and preservation	EN 13946 chapter 6.3 and 6.4
	Lentic waters	Sampling and preservation of macroinvertebrates	EN 27828
	Shallow waters	Sampling of macroinvertebrates	EN 27828 EN 28265 chapter 4 and 5
	Deep waters	Sampling of macroinvertebrates	EN ISO 9391
	Sediment / stony substrate	Sampling and preservation of benthic macroinvertebrates	ISO 8265 EN 28265 EN 27828
Marine waters	Soft bottom	Sampling of invertebrates and sample fixation	EN ISO 16665 chapter 4, 5.2 and 5.3
	Water body	Fish sampling, handling, preservation	ISO/DIS 23893-1 chapter 4.4 and 5.3

Compartment		Problem / issue / task	Find guidance in
General	*specific*		
Soils	Invertebrates	Sampling and preservation of macroinvertebrates	Anderson and Ingram,1993 Krell (undated)
		Sampling and preservation of earthworms	ISO 23611-1 chapter 7
		Sampling and preservation of microarthropods	ISO 23611-2 chapter 6
		Sampling and preservation of enchytraeids	ISO 23611-3 chapter 7
	Plants	Epiphytic lichen sampling	Giordani et al., 2001
		Epilithic lichen sampling	Insarov et al., 1999
		Macrophyte sampling	Ling, 2003
		Moss sampling	Couto et al., 2003 Fernandez et al., 2002
All	Choice of sites	Selection of reference sites	EN 14184 chapter 6.3 EN ISO 16665 ISO/DIS 23893-1
		Selection of representative sites	EN 27828 EN 14184 chapter 6.5

10.1.1 Sampling methodology

10.1.1.1 General considerations

The choice of the sampling species is crucial, and depends on the biological endpoints that is studied and on the objective that is addressed. The selected species shall be representative for the investigated aquatic environment and, when necessary, it shall be possible to sample the organs of interest in sufficient quantity (USGS 1999). Species for which the chosen biomarkers have been at least partially described shall be preferred, as there are references to which the obtained results can be compared.

For fish and invertebrate biomarker studies (e.g. Vtg, AChE, EROD activity measurement), biometric measurements such as sex, size (length or/and weight), sexual maturity (i.e. gonadal weight or secondary sexual characters) have to be selected and accurately checked. Fish or organisms with visible external lesions and parasites should be excluded from the analysis. These factors mentioned above may increase the variance of some of the biological effects measurements. Physico-chemical data, including temperature, which can influence some enzymatic activities, should be checked.

There is a great importance of the representativeness of the different habitats in the sampling area regarding for example stream speed and type of substrate (EN 27828).
Locations of the sampling sites should therefore be determined by the objectives, which are usually related to the location of point sources of pollution. A suitable number of sites should be placed in a gradient from the local discharge point, or at sites that should be protected from disturbances (ISO 23893-1).

Reference sites should be as close as possible to natural conditions with respect to their species composition and the abundance of each species. Parameters to be taken into account are (EN 14184; EN ISO 16665; ISO/DIS 23893-1):

- condition of substrate
- water depth
- flow type
- sediment type
- ecological status

Reference stations should also be used in surveys where special circumstances demand direct comparison of the fauna with that beyond the distributed or affected area, or where knowledge of the extent of natural variation is required. Multiple reference stations are particularly important in heterogeneous areas.

10.1.1.2 Plant sampling

For benthic diatoms sampling, cobbles are the preferred substratum for sampling. Pebbles and boulders can also be used. At least 5 cobbles should be sampled. An area of approximately 10 cm^2 or more should be scraped. The following microhabitat conditions should be fulfilled:

- areas of heavy shade or very close to the bank should be avoided;
- the substrata shall be submerged for long enough to ensure that assemblages are in equilibrium with their environment (at least 4 weeks);
- samples should be collected from within the main flow of the river at the sample site: zones of very slow flow rate should be avoided (EN 13946);
- Sample size should be about 300 to 500 units counted (units are either valves or frustules) (EN 14407);

For epiphytic lichen sampling, a sampling grid of 30 cm x 50 cm, split up into 10 rectangles measuring 15 cm x 10 cm each should be used (according to Giordani et al. 2001). This grid has to be positioned on the part of the bole with the highest lichen coverage, at a height of 120 cm, on trees with trunks that are neither damaged nor irregular, and having a circumference greater than 70 cm (for olive-trees, the circumference should be greater than 50 cm).

For epilithic lichen, some properties of the rocks from where lichens are sampled have to be similar (Insarov et al., 1999). These are rock type, surface characteristics (roughness, slope, exposure), and shading conditions. A linear sampling strategy should be preferred instead of the squares one, as it has been demonstrated to be more efficient in lichen monitoring

Moss should also be sampled on defined area. For example, the size of the area can be between 35 m * 35 m and 50 m * 50 m, and contain 30 subsamples separated by at least 50 cm (Couto et al., 2003). Subsamples should have a similar weight and be distributed homogeneously, avoiding collection of concentrated mops (Fernandez et al., 2002).

Macrophyte surveys should be undertaken between late spring and early autumn, when macrophyte growths will be at an optimum (EN 14184). Comparative surveys in subsequent years should be undertaken at the same time of the year as in previous years. This will ensure that changes resulting from different seasonal growth patterns are minimised.

Terrestrial macrophytes can be sampled using the protocol described by Ling (2003): squares of area 2 m^2 (1.41 m x 1.41 m) are used in a grid of regularly spaced points, 35 m apart.

10.1.1.3 Invertebrate sampling

Concerning aquatic macroinvertebrate sampling, the choice of sampler design depends on the species to be sampled (pelagic, benthic, size of organisms) and on the type of sediment (rocks, fine substrate), and on the depth of water (EN 28265; EN ISO 9391).

The sampling strategy for terrestrial macro-invertebrates can be based on the basic field transect of 40*4 m recommended by Anderson and Ingram (1993). In each transect, 8 to 10 monoliths should be sampled, from which invertebrates are extracted. Monoliths should be between 5 to 30 cm deep, depending on the type of soil and the species studied (ISO 23611). For the extraction of invertebrates from the monoliths, several solutions are possible:

- The portions of soil can be by placed in water (Krell, undated; ISO 23611-3). Animals will then float on the water surface.
- A solution of formalin is added to the soil portion and animals are sampled by hand (ISO 23611-1).
- A gradient of temperature around 30 to 35 °C is created between the upper part and the lower part of the soil sample, and the organisms are collected in a bottle (ISO 23611-2).

The sampling program of marine soft-bottom macro fauna should be developed with regard to local topographical and hydrographical conditions in the survey area. For monitoring purposes, sampling stations should preferably be positioned in areas of sandy/muddy bottom sediments (EN ISO 16665).

Positioning of sampling stations in marine environment can be as follows (EN ISO 16665):

Station network
Sampling stations are arranged in a regular grid-like pattern. This arrangement is appropriate for overview surveys and for mapping distribution of factors of interest. The survey area should be one of topographic homogeneity, but some adjustments can be made according to local conditions (e.g. in fjords and coastal waters with small variations in depth).

Random or scattered sampling
This may be possible in special circumstances (e.g. when no previous knowledge of the area is available), or when an unbiased value for a whole area is desired.

Stratified sampling
Sampling stations are arranged within locally homogeneous subdivisions of the survey area, delineated according to depth, sediment types or other factors that vary across the survey area. Stratification is appropriate in cases where habitat variability can confound patterns of interest.

Transect sampling
In order to trace effects of point source discharges by establishing the transect in the main current direction from the source, the stations can be placed along a known gradient in a sub-area of minimum habitat variability. When it is not feasible to work in strata, the stations can be placed across possible habitat gradients.

Single-spot sampling
This applies when a small number of stations are placed according to individual assessment. For example, when a specific chemical contamination is suspected, sampling stations may be positioned in the deepest parts of the survey area.

10.1.1.4 Fish sampling

Important natural factors which have to be considered prior to fish sampling for biochemical and physiological measurements are

- abiotic factors: climate, temperature, hydrology, oxygen and salinity
- biotic factors: age, size, sex, maturation, nutritional status, parasites and diseases.

All these factors can contribute to the overall variability of the measured response variables (ISO/DIS 23893-1). Therefore, sampling campaigns have to be planned and conducted taking into account the following parameters and considerations:

Frequency and season for sampling

Fish are sampled once a year during the autumn period in order to avoid the effects of rapid changes in physiological conditions due to the reproduction season. During the autumn most species of fish are not reproducing and the conditions to get enough fish by stationary gear like gill nets and fyke nets are still good because the fish are still active.

Sampling procedures

The number of fish should be sufficient in order to detect a predetermined change in the response variable within a certain number of years.

In order to avoid unnecessary stress on the fish, when the fishes are being caught and sacrificed for sampling of tissues, all fishes are first brought to a wooden fish chest and kept there for 2 days to 4 days before they are being sacrificed. This stabilises stress sensitive response variables like blood glucose, blood lactate and hematocrit. In cases where this procedure can not be followed (e.g. on a cruise vessel), consistency in handling between different stations should be a minimum requirement.

10.1.2 Sample pre-treatment for biological purpose, and stability

For most of biological analyses, which like biochemical measurements cannot be determined on site, tissue or organ samples from fish or macroinvertebrates shall be kept frozen below -70°C until analysed (ISO 23893-1).

Blood samples are processed directly to determine hematocrite by centrifugation, haemoglobin spectrophotometrically, and to produce blood plasma by centrifugation. Samples of blood plasma, bile and other tissues are placed in containers and frozen directly in liquid nitrogen or in contact with solid carbon dioxide. Samples for histological examination and determination of fish age are treated as required by the methods that are used. The sampling of fish tissues shall be made in a locality that is less than 100 m from the wooden fish chest. Moreover, all tissue samples (liver and muscle) are processed further immediately by taking sub samples for different analysis. The sub samples shall be taken from the same part of the organ for each variable to be analysed.

Diatoms

The cell division of diatoms and decomposition of organic matter should be stopped. No preservative is necessary if the sample is to be processed within a few hours of collection. Lugol's iodine can be used for short-term storage. Buffered ethanol or formaldehyde are recommended for long-term storage of samples. Samples can also be deep-frozen (EN 13946).

Algae and plankton

Samples of algae or plankton should be placed in a cool, dark place and stand for at least 24h for sedimentation. Alternatively, the sample can be centrifuged. Conservatives can be added, if necessary, to the samples before or after sedimentation or centrifugation (EN 13946).

Invertebrates
Samples of invertebrates (aquatic and terrestrial) can be conserved in formalin or ethanol (EN 27828; Krell, undated; ISO 23611).

Samples of marine macroinvertebrates should be washed and sieved with seawater.

Polychaetes, amphipods and oligochaetes are particularly fragile. Sieving is complete as soon as the fine material is washed out of the sample. Long sieving times should be avoided because small animals may actively pass through the sieve.

Animals that produce slime, large or heavy ones, and predators should be removed from the samples and placed in separate containers.

Fragile animals may be carefully washed or picked out of the sample during sieving.

All that can damage the sample material during transport (e.g., large stones, shells, sticks) should be discarded.

Samples should be fixed as soon as possible after sieving using formalin. Samples that should be kept for a long period can be transferred in ethanol after having being rinsed. In this case, no study on biomass can be done (EN ISO 16665).

10.1.3 Sample homogeneity

In specific cases such as fish sampled for biomarkers determination, sample homogeneity is very important, notably in terms of size, weight and sex. For example, for EROD determination, only immature or sexually mature fish of one sex (e.g. females for perch and eel pout, and males for chub and zebrafish) within a certain size interval are used for each species in order to minimize the influence of sex and size (ISO/DIS 23893-1).

Otherwise, samples should be as most representative as possible of the sampling area and therefore do not need to be homogeneous.

10.2 Water Sampling

Table 15 Quick reference table for water sampling issues

Compartment		**Problem / issue / task**	**Find guidance in**
General	*specific*		
All		design of sampling programmes	ISO 5667-1
		sampling techniques	ISO 5667-1
		preservation & handling of water samples	ISO 5667-3
		sampling of water used for biotesting	ISO 5667-16
		quality assurance measures in sampling and sample handling	ISO 5667-14
		sample homogeneity	ISO 13528 ISO Guide 35
Environmental Waters	Surface waters	suspended particulate matter	ISO 5667-17
	Surface freshwaters	lakes (natural and man-made)	ISO 5667-4
		Rivers & streams	ISO 5667-6
	Groundwater	general guidance on sampling	ISO 5667-11
		sampling at contaminated sites	ISO 5667-18
	Wet deposition	general guidance on sampling	ISO 5667-8
	Marine waters	Water sampling	ISO 5667-9
		sediment sampling	ISO 5667-19
Waters in or stemming from technical systems	boiler plants		ISO 5667-7
	waste waters		ISO 5667-10
	Drinking water		ISO 5667-5

10.2.1 Sampling methodology

The basic purpose of the sampling of water for analysis of emerging pollutants is usually to collect samples whose composition represents the quality of the water from which they have been taken, in order that subsequent examination of these samples may provide information on water quality that satisfies the objective of the sampling programme.

For the selection of adequate sampling procedures the specific characteristics of both, the selected analytes and the sampled water source (water body) must be taken into account. For example, for the determination of trace concentrations of organic compounds the sampling methodology is often guided by information on persistence and physico-chemical properties of the substance, e.g. K_{oc} and K_{ow}, as well as vapour pressures (Barceló and Hennion 1997a).

These are important to consider for overcoming problems such as adsorption on sampling tubes, bottles, filters, and suspended material, or evaporation and biological or photochemical degradation. Therefore, sample containers should be adapted to the

requirements of the analyte in question. In particular, this requires consideration of the following properties of the container:

- material (chemical composition, transparency, sorption and diffusion properties)
- type of sealing / stopper (e.g. gas tight, chemically inert, no air above the sample)
- cleaning procedure

Furthermore, the specific analyte in question may require the addition of specific preservatives or suitable changes in the sampling procedure.

Various kinds of water bodies create specific sampling situations. For instance surface waters include a wide range of different types (surface run-off, ditches, creeks, rivers, lakes, estuaries, seas, industrial areas, effluents, and piped water) and there is no single procedure or device that is adequate for sampling such a variety of situations without modification. As it is usually necessary to collect representative samples, an understanding of the inherent temporal and spatial variability in the water body from which the samples are to be taken is indispensable, and the limitations in taking representative samples from this water body have to be known. This affects strongly the selection of sampling points and sampling frequency, which have to be adjusted to the objective of the specific monitoring activity.

Guidance on sampling methodology is provided by numerous standards, guidelines, handbooks and scientific articles dealing with

- design of sampling plans and programmes:
  ISO 5667-1, Burton and Pitt 2001, Boulding and Ginn 2004
- sampling techniques:
  US EPA 1980, US EPA 1982, ISO 5667-2
- preservation and handling of water samples:
  US EPA 1980, US EPA 1982, Uzoukwu 2000, ISO 5667-3, Barceló and Hennion 1997b
- quality control in water sampling:
  ISO 5667-14

Detailed instructions for specific sampling situations are content of parts of international standard series ISO 5667 (see Table 14).

Valuable literature sources with a wide scale of information can also be obtained on the World Wide Web. Selected references for processing of water samples are given at
http://water.usgs.gov/owq/FieldManual/chapter5/pdf/selected.pdf or
http://nepis.epa.gov/pubtitleORD.htm.

10.2.2 Sample pre-treatment

Sample pre-treatment procedures for water samples are determined by the characteristics of analytes of interest and also by the type of analysed water, mainly by the content of suspended particulate matter (SPM), and the requirements on the analysis.

Total determinand concentration
If the total determinand concentration shall be analysed, analysis is performed with the whole water sample, i.e. when solid matter and the liquid phase have not been separated.

Dissolved determinand concentration
In order to analyse the dissolved determinand concentration, an appropriate separation step (e.g. filtration) must be applied to remove the suspended particulate matter fraction from the whole water sample.

SPM determinand concentration
In order to investigate or quantify the fraction of the analyte which is bound to the SPM, an appropriate separation to separate SPM from the liquid phase needs to be performed.

Preconcentration:
In the analysis of trace concentrations of organic compounds in water a preconcentration of determinants to enhance the sensitivity of the determination is in most cases an inevitable part of the analytical procedure. The preconcentration can be accomplished *in situ* (on sampling site) using solid sorbents (SPE disks or cartridges) and is in this case called "sorbent sampling". However, usually the preconcentration is completed in the laboratory using various kinds of sample handling procedures (e.g., purge-and-trap, liquid-liquid extraction, solid phase (micro) extraction) (Barceló and Hennion 1997b, Loconto 2001).

10.2.3 Sample homogeneity

Homogeneity of water samples is usually not a problem, because the solutes are quickly reaching their equilibrium distribution in water. If considerable amounts of SPM are present in the sample, the parent sample source must be mixed well before taking a sample aliquot for further processing. Some guidance on possible techniques for evaluation of homogeneity can be found in references dealing with homogeneity testing of reference materials (IUPAC [2006], ISO 13528, ISO Guide 35).

10.2.4 Sample stability

After the sampling, chemical and biological changes are taking place in water samples, which can affect the stability of the analyte. To maintain the stability and integrity of analytes until analysis can be performed, appropriate selection and pre-treatment of containers, selection of suitable preservation methods and short holding times (the time interval between collection and analysis) are necessary. Methods of preservation are relatively limited and are generally intended to inhibit or at least to retard processes such as biological activity, hydrolysis of chemical compounds and complexes, and volatilisation of constituents. Preservation methods are usually limited to pH control, chemical addition, refrigeration, filtration, and freezing. No single method of preservation is entirely satisfactory. The preservative has to be chosen with due regard to the determination to be made (it must not interfere with the analysis being made). Addition of preservatives may be inadequate when applied to samples containing considerable amounts of SPM (US EPA 1980, US EPA 1982, US EPA 1983, Uzoukwu 2000, ISO 5667-3).

10.2.5 Water sampling for biotesting

For water samples used for biotests, it is recommended that the sampling, transportation and storage of the samples should be carried out in accordance with the general procedures described in ISO 5667-16. Samples should be collected in bottles made from chemically inert materials.
The toxicity test should be carried out as soon as possible, ideally within 12 h of collection. If this time interval cannot be observed, the sample shall be cooled (0 to 4 °C) and tested within 24 h. If testing cannot be carried out within 72 h, the sample, where the characteristics are known to be unaffected by freezing, may be frozen (below –18°C) for testing within 2 months after collection. At the time of testing, the sample should homogenised by shaking manually and (if necessary) be allowed to settle during 2 h in a container, and subsequently sampled by drawing off with a pipette the required quantity of supernatant, maintaining the end of the pipette in the centre of the section of the test tube and half way between the surface of the deposited substances and the surface of the liquid.
In the case where the raw sample or the decantation supernatant is likely to interfere with the organism to be tested (micro-crustaceans, residual suspended matter, protozoa, micro-organisms, etc.), a filtration through a sieve with a 0,45 µm mesh or centrifuging of the raw or decanted sample may be performed.

10.3 Soil and sediment sampling

Table 16 Quick reference table for soil and sediment sampling issues

Problem / issue / task	Find guidance in
control samples	IAEA (2004): TECDOC-1415 chapter 7 US EPA (1989): EPA/600/8-69/046 chapter 9
equipment	ISO 10381-2 US EPA (1997): SOP env 3.13 chapter 5,6,8 ISO 4364
containers	ISO 10381-2: US EPA (1997): SOP env 3.13 chapter 8.4 US EPA (1989): EPA/600/8-69/046 chapter 11
cross-contamination	Karstensen (1996), chapter 2.2
decontamination	US EPA (2001a),EISOPQAM appendix B US EPA (1999) SOP env 3.15 March 1999
documentation	US EPA (1989) EPA/600/8-69/046 chapter 11 US EPA (2001a) EISOPQAM, chapter 3
pre-treatment	ISO 14507:2003
preservation	US EPA (1997) SOP env 3.13, chapter 8.4
sieving	ISO 11464:1994; ISO 11277:1998 US EPA (1989) EPA/600/8-69/046, chapter 11
drying	ISO 11464:1994 ISO 11465:1993
crushing / milling	ISO 11464:1994 Karstensen (1996), chapter 4-10
sub-sampling	ISO 11464:1994 US EPA (2003) EPA/600/R-03/027
homogenisation	US EPA (1990) EPA/600/X-90/043 ASTM (2002) D422-63
compositing samples	US EPA (1992) EPA/600/R-92/128, chapter 7 IAEA (2004) TECDOC-1415, chapter 7
transport	US EPA (2001b) EPA SOP env 3.16
storage	ISO 10381-2:2002 US EPA (1997) EPA SOP env 3.13, chapter 8.4
sampling of VOC	US EPA (1992) EPA /600/R-92/128, chapter 5 Karstensen (1996), chapter 7

10.3.1 Sampling methodology

10.3.1.1 Field spiked samples and field blanks

To be able to evaluate errors in the sampling procedure at least two types of control samples are recommended: field spiked samples and field blanks.
Field spiked samples are samples to which a known quantity of the pollutant is added in conjunction with sampling in the field. They are used to identify errors in the sample transport.
Field blanks are samples consisting of pure material of a type similar to the sampled material that is treated in the same manner as the soil samples. They are used to determine whether contaminants have been introduced during sample preparation, transport etc.

10.3.1.2 Sampling equipment

The selection of sampling equipment requires consideration of many factors such as site limitations, soil and sediment characteristics and depth, the required accuracy, and the ease of use. The sampling equipment should not contaminate the samples. For volatile organic compounds (VOC) it is important that sample handling and containerizing should minimize losses through volatilisation.

For soft surface soil sampling, a scoop or trowel will be appropriate. For harder soil a spade or shovel is a better choice. If the sampling objective is to analyse each soil horizon, a soil coring device must be used. Depending on depth and type of soil different equipments are more or less suitable. A Shelby tube sampler can be used for soft soil while a Split-spoon sampler is a better choice for hard soils. When sampling at depth, the use of different kinds of augers in connection with a sample collector is recommended. Several augers are available such as continuous-flight, hand-operated power type and bucket type. With a power auger a depth of at least 5 m can be reached.

For sediment, several techniques are available from simple mechanical devices, such as grab and core samplers, to more sophisticate optical and electronic techniques. Often other environmental parameters are measured during sample collection (e.g., water level, turbidity, pH, electrical conductivity). Many samplers have been developed for sampling the sediment bed. Basic grab samplers are shovels, scoops and pipe dredging. But also excavation enclosures and resuspension techniques can be used. The choice of sampling technique depends on the objective of the study, and issues to consider are:

i) Is a stratified sediment core needed?
ii) Is fine matter needed?
iii) What is the required sample size?
iv) What analytes should be determined?

Suspended sediment can be sampled using for example a continuous-flow centrifuge, tangential-flow system or large diameter tubes. More information on sediment sampling can be found in ISO 4364.

10.3.1.3 Sampling containers

When analysing soils or sediments for metals, suitable materials for storage containers are borosilicate glass, polyethylene and polypropylene. Polyvinylchloride (PVC) and polystyrene should be avoided as they might contain cadmium, tin or lead. Before use the container should be cleaned with 10% nitric acid and rinsed with purified water.
In cases where the sample is to be analysed for non-volatile organic compounds, glass containers with aluminium foil or Teflon-lined caps are recommended. As a rule, no plastic containers should be used (although there are exceptions). Soil or sediment samples to be analysed for VOCs can be collected in glass jars, which should be completely filled. It is important to immediately seal the container.

10.3.1.4 Cross-contamination

Contaminants can be transferred from one sample to another because of inadequate cleaning of sampling equipment, vessels etc. Pollutants may also be spread between different soil or sediment strata if inappropriate sample devices are used. Cleaning procedures for the sampling equipment must be described in the quality control programme.

10.3.1.5 Documentation

A sampling plan should include information about how to identify the samples, date of sampling, sample location, sample equipment, sample storage containers, sampling depth, sample amount, sieving, sample preservation, prevention of contamination, labelling, control samples, transportation and storage. All activities in the field must be recorded in a logbook and/or in specific record forms.

10.3.2 Sample pre-treatment

Sample preservation should be performed immediately upon sample collection. For metals and non-volatile organic compounds, preservation at 4°C is recommended. For volatile organic compounds (VOC), preservation in methanol, followed by storage at 0 - 4°C is recommended.

10.3.3 Sample homogeneity

Those components of the sample that are not relevant for characterizing the sample (e.g. stones, roots) are first of all identified and discarded. Sieving is recommended for soils to be analysed for metals (2 mm) and non-volatile organic compounds (2-8 mm) but not for VOC due to possible losses of this fraction. Before sieving the soil is oven-dried at <40°C. Depending on the type of analyte, non-volatile organic compounds can be freeze- or air-dried. If sub-samples are to be analysed, the soil must be homogenized to provide a uniform distribution of the contaminants (not for VOC). Sediments for the analysis of non-volatile organic compounds are also often sieved, if the sand fraction (> 63 µm) is high.

10.3.4 Sample stability

Samples destined for metal analysis should be transported and stored at +4°C in containers of borosilicate glass, polyethylene, polypropylene or Teflon. The plastic and glass should be colourless.

Samples to be analysed for non-volatile organic compounds should preferably be transported under cool conditions and protected from direct sunlight. It is recommended that they be stored at +4°C or frozen if the storage period is longer than 2 days. The containers should be glass. Depending on the analytes, storage in the dark might be necessary.

Samples to be analysed for VOC should be transported in an ice chest but not under -10°C. Storage in the dark at 0 to +4°C is recommended.

10.4 Air sampling

Sampling of air for analysis can be split into two broad categories, in situ measurement and sampling for subsequent analysis. In addition two classes of measurand can be defined, particulate and aerosol phase (potentially including nanomaterial), and gas phase.

10.4.1 In situ measurement

Many measurements of airborne pollutants can be made in situ. With respect to emerging pollutants the most likely techniques will be either optical – spectroscopic techniques, gas chromatographic techniques, mass spectrometry or one of the related hyphenated techniques. With such instrumental techniques the issue is to collect a representative sample of the air mass to be monitored, and to transport the compounds of interest to the analyser. In most cases this involves the use of a sampling manifold to draw air into the system. Obviously such manifolds must not affect the concentration of the measurand, and requirements on characteristics of the sampling system such as the residence time are often specified in the methods.

In order to improve the sensitivity of these techniques sample conditioning systems are often employed, these range from cryogenic pre-concentrators on GC systems to membrane technologies often used with mass spectrometers. Such systems are often designed to remove potential interferents, particularly water vapour. Often the choice of sampling system will affect the scope for which the technique is valid, for example tenax may emit benzene at high temperatures, and a common drying system, the nafion dryer removes all polar compounds in addition to the target water molecules.

Online systems for particulate monitoring often need size selective sample heads to fraction the material collected to a relevant region (i.e. the typical PM10 or PM2.5 sampling systems routinely used for air quality monitoring).

Size partitioning may also be achieved using cascade impactors or cyclones. Various systems have been deployed which transport the collected material from such samplers, generally in a liquid wash, into an analyser to enable the measurement of chemical or biological composition of the particulate matter.

A class of optical techniques can be deployed in an ‘open-path’ or remote sensing configuration. This removes the issue of sample handling completely, as the measurement is carried out in the air mass itself. Such systems exhibit their own range of performance characteristics and influence factors, and current standardisation activies in CEN are addressing open-path FTIR (Fourier transform infrared) and DOAS (differential absorption spectroscopy) techniques and in the future DIAL (differential absorption lidar). Because of their wide applicability to a range of compounds multi-species techniques require validation when used for compounds for which they have not been previously been validated.

10.4.2 Sampling for subsequent analysis

Sampling systems employed in the air compartment consist of either active or passive systems. The simplest approach is to collect whole air samples in passivated cylinders. In general however such sampling does not collect enough material to detect low levels of pollutants. Active systems draw air through a sampling medium, either a solid sorbent (i.e. tenax, carbosieve etc), a liquid impinger, or filter systems (mainly for aerosol sampling). Such sorbents may collect material by mechanical or chemical processes. Passive samplers, which are routinely deployed for air quality assessment, rely on uptake onto a sorbent from the ambient air. The key parameter in such systems is the uptake rate, which must be determined on a measurand specific basis and is affected by a range of ambient conditions.

Key performance characteristics of all such sampling systems are the sampling efficiency, break through volume (i.e. the point at which collected material starts to be drawn off the sampler), the effect of ambient conditions, the time period for which the sampler can be used, and the recovery efficiency.

For collected particulate matter analysis requires some form of extraction from the filter media, this may be by washing (often with sonification), or digestion.

Recovery from sorbents is usually either by thermal or solvent desorption, though soft ionisation techniques may be used for mass spectrometry.

One of the key issues with extending the scope of methods which rely on sampling is the impact of the sampling media, either because it does not have the required efficiency for the new pollutants or because the media itself contains significant levels of contaminants.

11 References

Anderson JM & Ingram JSI (1993): Tropical Soil Biology and Fertility – A Handbook of Methods. 2nd ed. CAB International, Wallingford.

ASTM (2002). Standard Test Method for Particle-Size Analysis of Soils, ASTM D422-63.

Barceló D & Hennion M-C (1997a,b): Trace determination of pesticides and their degradation products in water, Elsevier, Amsterdam, a: Chapter 2; b: Chapter 4.

Barwick VJ & Ellison SLR (2000): VAM Project 3.2.1 Development and Harmonization of Measurement Uncertainty Principles. Part (d): Protocol for uncertainty evaluation from validation data.

Burton GA & Pitt RE (2001): Stormwater effects handbook: A toolbox for watershed managers, scientists, and engineers, CRC/Lewis Publishers, Boca Raton, FL, Ch. 5.

Cofino WP, van Stokkum IHM, Wells DE, Ariese F, Wegener JWM, Peerboom RAL (2000): A new model for the inference of population characteristics from experimental data using uncertainties. Application to interlaboratory studies. Chemometrics Intelligent Laboratory Systems 53: 37-55.

Couto JA, Fernandez JA, Aboal J. & Carballeira A (2003): Annual variability in heavy-metal bioconcentration in moss: sampling protocol optimisation, Atmospheric Environment, 37: 3517-3527.

De Boer J, Cofino WP (2002): First world-wide interlaboratory study on polybrominated diphenylethers (PBDEs). Chemosphere 46: 625-633.

EN 13946 (2003): Water quality - Guidance standard for the routine sampling and pre-treatment of benthic diatoms from rivers.

EN 14184 (2003): Water quality - Guidance standard for the surveying of aquatic macrophytes in running waters.

EN 14407 (2004): Water quality - Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters.

EN 27828 (1994): Water quality - Methods of biological sampling - Guidance on handnet sampling of aquatic benthic macroinvertebrates.

EN 28265 (1994): Water quality - Design and use of quantitative samplers for benthic macroinvertebrates on stony substrata in shallow freshwaters (ISO 8265:1988).

EN ISO 16665 (2006): Water quality - Guidelines for quantitative sampling and sample processing of marine soft-bottom macro fauna.

EN ISO 9391 (1995): Water quality – Sampling in deep waters for macroinvertebrates – Guidance on the use of colonization, qualitative and quantitative samples.

Eurachem (1998): EURACHEM Guide: The fitness for purpose of analytical methods – a laboratory guide to method validation and related topics.

Eurachem (2000): EURACHEM/CITAC Guide CG 4: Quantifying Uncertainty in Analytical Measurement, 2nd Edition.

Fernandez JA, Aboal JR, Couto JA & Carballeira A (2002): Sampling optimisation at the sampling-site scale for monitoring atmospheric deposition using moss chemistry. Atmospheric Environment 36: 1163-1172.

Giordani P, Brunialti G & Modenesi P (2001): Applicability of the lichen biodiversity method (L.B.) to a Mediterranean area (Liguria, NW Italy), Part 2: Sampling and extraction of microarthropods (Collembolla and Acarina), Cryptogamie Mycol. 22: 193-208.

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi A, Prieto P, Sabioni E, Scott L, Worth A & Zuang V (2004): A modular approach to the ECVAM principles on test validity. Atla 32: 467-472.

Huber L (1998): Validation and qualification in analytical laboratorties, Interpharm, East Englewood, CO, USA.

IAEA (2004): Soil Sampling for Environmental Contaminants, IAEA-TECDOC-1415, Vienna.

ICH-Q2a (1995): Guideline for industry: Text on validation of analytical procedures. htpp://www.fda.gov/cder/guidance/index.

ICH-Q2B (1996a): Guidance for industry: Validation of analytical procedures: Methodology. htpp://www.fda.gov/cder/guidance/index.

ICH-Q6B (1996b): Harmonised tripartite guideline. Specifications: Test procedures and acceptance criteria for biotechnological/biological products. htpp://www.fda.gov/cder/guidance/index.

Insarov GE, Semenov SM & Insarova ID (1999): A system to monitor climate change with epilithic lichens, Environ. Monit. Assess. 55: 279-298.

ISO 10381-2 (2002): Soil quality – Sampling Part 2: guidance of sampling techniques.

ISO 11277 (1998): Soil quality - Determination of particle size distribution in mineral soil material - method by sieving and sedimentation.

ISO 11464 (1994): Soil quality – Pretreatment of samples for physico-chemical soil analysis.

ISO 11465 (1993): Soil quality - Determination of dry matter and water content on a mass basis -Gravimetric method.

ISO 11843 series (1997): Capability of detection. Parts 1 – 5.

ISO 13528 (2005): Statistical methods for use in proficiency testing by interlaboratory comparisons.

ISO 14507 (2003): Soil quality - Pretreatment of samples for determination of organic contaminants.

ISO 14956 (2002): Air quality. Evaluation of the suitability of a measurement procedure by comparison with a required measurement uncertainty.

ISO 23611-1 (2006): Soil quality – Sampling of soil invertebrates – Part 1: Hand-sorting and formalin extraction of earthworms.

ISO 23611-2 (2006): Soil quality – Sampling of soil invertebrates – Part 2: Sampling and extraction of microarthropods (Collembolla and Acarina).

ISO 23611-3 (2005): Soil quality – Sampling of soil invertebrates – Part 3: Sampling and extraction of enchytraeids.

ISO 3534-1 (2006): Statistics - Vocabulary and symbols - Part 1: General statistical terms and terms used in probability.

ISO 4364 (1997): Measurement of liquid flow in open channels – Bed material sampling. [plus Technical Corrigendum 1 – (2000)]

ISO 5667 (1987-2006): Water Quality – Sampling.
Part 1: Guidance on the design of sampling programmes and sampling techniques.
Part 2: Guidance on sampling techniques
Part 3: Guidance on the preservation and handling of water samples
Part 4: Guidance on sampling from lakes, natural and man-made
Part 5: Guidance on sampling of drinking water from treatment works and piped distribution systems
Part 6: Guidance on sampling of rivers and streams
Part 7: Guidance on sampling of water and steam in boiler plants

Part 8: Guidance on the sampling of wet deposition
Part 9: Guidance on sampling from marine waters
Part 10: Guidance on sampling of waste waters
Part 11: Guidance on sampling of groundwaters
Part 12: Guidance on sampling of bottom sediments
Part 13: Guidance on sampling of sludges from sewage and water treatment works.
Part 14: Guidance on quality assurance of environmental water sampling and handling
Part 15: Guidance on preservation and handling of sludge and sediment samples
Part 16: Guidance on biotesting of samples
Part 17: Guidance on sampling of suspended sediments
Part 18: Guidance on sampling of groundwater at contaminated sites
Part 19: Guidance on sampling of marine sediments

ISO 5725-1 (1997): Accuracy (trueness and precision) of measurement methods and results. Part 1: General principles and definitions.

ISO 5725-2 (1994): Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method (plus Technical Corrigendum 2002).

ISO 5725-3 (1994): Accuracy (trueness and precision) of measurement methods and results. Part 3: Intermediate measures of the precision of a standard measurement method (plus Technical Corrigendum 2001).

ISO 5725-4 (1994): Accuracy (trueness and precision) of measurement methods and results. Part 4: Basic methods for the determination of the trueness of a standard measurement method.

ISO 5725-5 (1998): Accuracy (trueness and precision) of measurement methods and results. Part 4: Alternative methods for the determination of the precision of a standard measurement method (plus Technical Corrigendum 2005).

ISO 78-2 (1999): Chemistry – Layout for standards – Part 2: Methods of chemical analysis.

ISO 8466-1 (1990): Water quality -- Calibration and evaluation of analytical methods and estimation of performance characteristics -- Part 1: Statistical evaluation of the linear calibration function.

ISO 8466-2 (2001): Water quality -- Calibration and evaluation of analytical methods and estimation of performance characteristics -- Part 2: Calibration strategy for non-linear second-order calibration functions.

ISO 9000 (2005): Quality management systems – Fundamentals and vocabulary.

ISO Guide 35 (2006): Reference materials. General and statistical principles for certification, 3rd edition.

ISO Guide 43-1 (1997): Proficiency Testing by Interlaboratory Comparisons - Part 1: Development and Operation of Laboratory Proficiency Testing Schemes.

ISO Guide 98 (1995): Guide to the expression of uncertainty in measurement (GUM).

ISO/DIS 20612 (2006): Water quality - Interlaboratory comparisons for proficiency testing of analytical testing laboratories.

ISO/DIS 23893-1 (2006): Water Quality – Biochemical and physiological measurements on fish – Part 1: Sampling of fish, handling and preservation of samples.

ISO/IEC 17025 (2000): Conformity assessment - General requirements for the competence of testing and calibration laboratories.

ISO/IEC Guide 30 (1992): Terms and definitions used in connection with reference materials.

ISO/TS 20281 (2006): Water Quality – Guidance on Statistical Evaluation of Ecotoxicity Data.

ISO/TS 21748 (2004): Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.

IUPAC (1997): The compendium of analytical nomenclature ("IUPAC orange book"). http://www.iupac.org/publications/analytical_compendium/

IUPAC (2006): The international harmonized protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure & Applied Chem. 78(1): 145-196.

JAMP (2003): JAMP Guidelines for contaminant-specific Biological Effect Monitoring. OSPAR Commission Monitoring Guidelines. Ref N° 2003-10.

Johnson I (1994): A procedure to select appropriate ecotoxicological methods to meet the operational needs of regulators. Draft R&D note 494 – UK Environment Agency.

Karstensen KH (1996): Nordic Guidelines for Chemical Analysis of Contaminated Soil Samples, Nordtest Project 1143-93 / Nordtest Technical Report 329.

Krell FT (undated): Dung Beetle Sampling Protocol. 1. Comparing Dung Beetle Assemblages – without traps, Scarab Research Group, London.

Ling KA (2003): Using environmental and growth characteristics of plants to detect long-term changes in response to atmospheric pollution: some examples from British beechwoods, The Sci.ence of the Total Environ.ment, 310:, 203-210.

Loconto, PR (2001): Trace environmental quantitative analysis: Principles, techniques and applications, Marcel Dekker, New York, Ch. 3.

Magnusson B, Näykki T, Hovind H, Krysell M (2003): Handbook for calculation of measurement uncertainty in environmental laboratories. Nordtest Report TR 537, 1-41.

Melancon MJ (1995) Bioindicators used in aquatic and terrestrial monitoring. In: Hoffman DJ., Rattner BA, Burton GA, Cairns J. Editors. Handbook of ecotoxicology. Boca Raton (FL): CRC press. 220-239.

Peakall D.W. (1994) Biomarkers: the way forward in environmental assessment. Toxicol. Ecotoxocol. News 1: 55-60.

SWIFT VG (2005): Guidelines for laboratories carrying out measurements where the results will be used to implement the Water Framework directive (2000/60/EC). E. Prichard. 48p.

Tas JW & Van Leeuwen CJ (1995): Glossary. In: van Leeuwen CJ & Hermens JLM [eds.], Risk Assessment of Chemicals: An Introduction, Dordrect, Netherlands, Kluwer Academic Publishers, pp. 339-361.

Traverniers I, De Loose M, Van Bockstaele E (2004): Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance. TRAC 23 (8): 535-552.

US EPA (1980): Samplers and sampling procedures for hazardous waste stream, EPA 600/2-80-018, Cincinnati, OH.

US EPA (1982): Handbook for sampling and sample preservation of water and wastewater, Environment monitoring and support laboratory, EPA 600/4-82-029, Cincinnati, OH.

US EPA (1983): Methods for chemical analysis of water and wastes, EPA 600/4-79-020, Cincinnati, OH, p.xv-xx.

US EPA (1989): Soil Sampling Quality Assurance User's guide, EPA/600/8-69/046, Environmental Monitoring Laboratory, Las Vegas NV.

US EPA (1990): A comparison of Soil Homogenization Techniques, EPA/600/X-90/043, Las Vegas, Nevada.

US EPA (1997): Soil Sampling, EPA SOP env 3.13, Ecology and Environmental Inc, New York.

US EPA (1999): Sampling Equipment Decontamination, EPA SOP env 3.15, Ecology and Environmental Inc, New York.

US EPA (2001a): Environmental Investigations Standard Operating Procedures and Quality Assurance Manual, EPA EISOPQAM, Athens, Georgia.

US EPA (2001b): Sample Packaging, EPA SOP env 3.16, Ecology and Environmental Inc, New York.

US EPA (2003): Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples, EPA/600/R-03/027.

USGS (1999) Biomonitoring of Environmental Status and Trends (BEST) Program: Field Procedures for Assessing the Exposure of Fish to Environmental Contaminants.

USGS (2000): Biomonitoring of environmental status and trends (BEST) Program: selected methods for monitoring chemical contaminants and their effects in aquatic ecosystems. U.S. Geological Survey Biological Resources Division, Columbia, (MO): Information and Technology Report USGS/BRD-2000-0005, 81p.

Uzoukwu BA (2000) Methods of water preservation, in: Nollet LML (ed.): Handbook of water analysis, Marcel Dekker, New York, Ch. 2.

VIM (2004): International vocabulary of basic and general terms in metrology, Draft Guide.

Von Holst C, Muller A, Bjorklund E, Anklam E (2001): In-house validation of a simplified method for the determination of PCBs in food and feedingstuffs. Eur. Food Res. Technol. 213: 154.

12 Annex

12.1 Definitions – Glossary

There is abundant literature on issues of validation and QA/QC procedures. Unfortunately, there is no consensus among these documents on a common terminology. Among the disciplines involved (e.g., metrology, statistics, chemistry, biology), there are considerably diverging concepts of the fundamental steps, tools, measures, elements and objects of the validation process. In order to establish a guideline with a consistent use of an unambiguous terminology, a glossary is indispensable. The most promising way to find a terminology which draws a consensus among the different faculties involved is to use definitions which have undergone a thorough international process of review and harmonisation. Therefore, definitions in this document were selected from overarching international standards or guidelines wherever possible or available (e.g., ISO 3534 and 5725 series; IUPAC „Orange book";draft version of VIM [2004]). Some definitions were taken from a validation guideline developed in the EU funded R&D project SWIFT (SWIFT VG).

Term	Definition	Reference
Accepted reference value	A value that serves as an agreed-upon reference for comparison, and which is derived as: a) a theoretical or established value, based on scientific principles; b) an assigned or certified value, based on experimental work of some national or international organisation c) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group d) when a), b) and c) are not available, the expectation of the (measurable) quantity, i.e. the mean of a specified population of measurements	ISO 3534-1
Accuracy	The closeness of agreement between a test result and the accepted reference value. NOTE: The term accuracy, when applied to a set of test results, involves a combination of random components (usually expressed by a precision measure) and a common systematic error or bias component (usually expressed by a measure for trueness). The technical term 'accuracy' must not be confused with the term 'trueness' (cf. definition of 'trueness').	ISO 3534-1
Adaptation	Deliberate modification of a test method with the aim to extend its scope or applicability, e.g. to make it applicable for a new matrix or organism	
Analyte	The substance subject to analysis	
Bias	The difference between the expectation of the test results and an accepted reference value. NOTE: Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value	ISO 3534-1

Term	Definition	Reference
	is reflected by a larger bias value	
Biomarker	A change in biological response, ranging from molecular through cellular and physiological responses to behavioural changes, which can be related to exposure to or toxic effects of environmental chemicals	Peakall 1994
Blank	Type of sample or test scheme without the analyte known to produce the measured signal. Use of various types of blanks enable assessment of how much of the measured signal or effect is attributable to the analyte and how much to other causes. Various types of blank are available.	
Blank – Reagent Blank	Reagents used during the analytical process (including solvents used for extraction or dissolution) are analysed in isolation in order to see whether they contribute to the measurement signal. The measurement signal arising from the analyte can then be corrected accordingly.	Eurachem 1998
Blank – Sample Blank	These are essentially matrices with no analyte. They are difficult to obtain but such materials are necessary to give a realistic estimate of interferences that would be encountered in the analysis of test samples	Eurachem 1998
Calibration	Operation establishing the relation between quantity values provided by standards and the corresponding indications of a measuring system, carried out under specified conditions. This relation can be expressed by calibration diagrams, calibration functions, or calibration tables. NOTE: The standards referred to here are measurement standards rather than written standards	VIM 2004
Certified Reference Material	Reference material, accompanied by a certificate, one or more whose property values are certified by a procedure, which establishes its traceability to an accurate realisation of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence	ISO/IEC Guide 30 – 1992
Development	In this guide, development is used in the meaning of extending the applicability of a method, e.g. to a new matrix, compound or organism.	
Fitness for Purpose	Degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose	IUPAC (1997)
Intermediate precision (within-lab reproducibility)	Precision under conditions where independent test results are obtained with the same method on identical test items in the same laboratory by different operators using different equipment on different days.	SWIFT VG
Limit of Detection	The lowest concentration of analyte in a sample that can be detected, but not necessarily quantitated under the stated conditions of the test.	Eurachem (1998)
Limit of Quantitation	The lowest concentration of an analyte that can be determined with acceptable precision under the stated conditions of the test	Eurachem (1998)
Linearity	The ability of the method to obtain test results proportional to the concentration of analyte.	Eurachem (1998)

Term	Definition	Reference
	NOTE: The linear range is by inference the range of analyte concentrations over which the method gives test results proportional to the concentration of the analyte	
Material	A compound / concentration level / matrix combination.	
Measurand	Quantity intended to be measured	VIM (2004)
Measurement	Process of experimentally obtaining information about the magnitude of a quantity	VIM (2004)
Measurement Uncertainty	Parameter that characterises the dispersion of the quantity values that are being attributed to a measurand, based on the information used.	VIM (2004)
Outlier	A member of a set of values which is inconsistent with the other members of that set. NOTE: ISO 5725-2 specifies the statistical tests and the significance level to be used to identify outliers in trueness and precision experiments	ISO 5725-1
Precision	The closeness of agreement between independent test results obtained under stipulated conditions. NOTES: Precision depends only on the distribution of random errors and does not relate to the true value or the specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation. „independent test results“ means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.	ISO 3534-1
Proficiency Testing	A periodic assessment of the performance of individual laboratories and groups of laboratories that is achieved by the distribution by an independent testing body of typical materials for unsupervised analysis by the participants	IUPAC (1997)
QA/QC	see: Quality Assurance and Quality Control	
Quality Assurance	The assembly of all planned and systematic actions necessary to provide adequate confidence that a product, process, or service will satisfy given quality requirements. A major part of Quality Assurance is Quality Control	ISO 9000
Quality Control	The operational techniques and activities that are used to satisfy quality requirements (e.g, in terms of method performance criteria that have to be met)	ISO 9000
Quantity	Property of a phenomenon, body, or substance, to which a magnitude can be assigned	VIM (2004)
Quantity values	Magnitude of a quantity represented by a number and a reference	VIM (2004)
Range	The difference between the largest and the smallest observed value of a quantitative characteristic	ISO 3534-1
Recovery	If a known amount of analyte is added to a test sample and the test sample is then analysed for that analyte by a particular method, the recovery is that fraction of the	SWIFT VG

Term	Definition	Reference
	amount of analyte added which is found by the method	
Reference Material	Material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials	ISO/IEC Guide 30
Repeatability	Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time	ISO 3534-1
Reproducibility	Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.	ISO 3534-1
Residual	Difference between the observed response and that predicted by a calibration function.	SWIFT VG
Robustness	The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage	ICH Q2A
Sample	The totality of a homogeneous analysis material having an identical composition or quality (similar to term batch)	ISO/DIS 20612
Sample - Field Sample	For example, the bulk of water collected from the river	SWIFT VG
Sample - Laboratory Sample	Primary sample material delivered to the laboratory	SWIFT VG
Sample - Subsample	A defined portion of a sample obtained by suitable sample division and identical in terms of composition	ISO/DIS 20612
Sample - Test sample	The aliquot of the laboratory sample taken for processing into a test portion	SWIFT VG
Sample - Test portion	The portion of the laboratory sample taken for analysis or testing	SWIFT VG
Sample handling	The manipulation to which samples are exposed during the measurement process, from the selection from the original material through to the disposal of all samples	SWIFT VG
Sample preparation	The procedures followed to select the test portion from the sample (or sub-sample). These include in-laboratory processes such as: homogenisation, mixing, and filtering. It often includes chemical operations such as: extraction, separation, derivatisation, and concentration	SWIFT VG
Selectivity	The ability of a method to determine accurately and specifically the analyte of interest in the presence of other components in a sample matrix under the stated conditions of the test	SWIFT VG
Sensitivity	The change in the response of a measurand divided by the corresponding change in the stimulus	
Specificity	The ability of a method to measure only what it is intended to measure	
Toxicity test	Determination of the effect of a substance on a group of selected organisms under defined conditions	Tas & Van Leeuwen

Term	Definition	Reference
Traceability	Property of the result of a measurement or the value of a standard whereby it can be related with a stated uncertainty, to stated references, usually national or international standards (i.e. through an unbroken chain of comparisons all having stated uncertainties) NOTE: The standards referred to here are measurement standards rather than written standards	(1995) ISO/IEC Guide 30 – 1992
Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. NOTE: The measure of trueness is usually expressed in terms of bias. Trueness must not be confused with the term 'accuracy' (cf. definition of 'accuracy').	ISO 3534-1
Validation	Method validation is the process of verifying that a method is fit for its intended purpose, i.e. to provide data suitable for use in solving a particular problem or answering a particular question. This process includes: • establishing the performance characteristics, advantages and limitations of a method and the identification of the influences which may change these characteristics, and if so to what extent, and • a comprehensive evaluation of the outcome of this process with respect to the fitness for purpose of the method.	
Working range	The interval between the upper and lower concentration (amounts) of analyte in the sample for which it has been demonstrated that the analytical procedure has a suitable level of uncertainty.	SWIFT VG

12.2 Detailed guidance on measurement uncertainty

The following sections will address the steps to estimate uncertainty of measurement as indicated in section C.8 of chapter 7.3, in more detail and with specific reference to the issues affecting environmental monitoring methods of the kinds being addressed by the Validation Protocol.

12.2.1 Overview of approach

The overall approach to estimating the uncertainty of a measurement is outlined in Figure 4. This provides a five-step process for determining the measurement uncertainty. These steps will be described in more detail in the following sections:

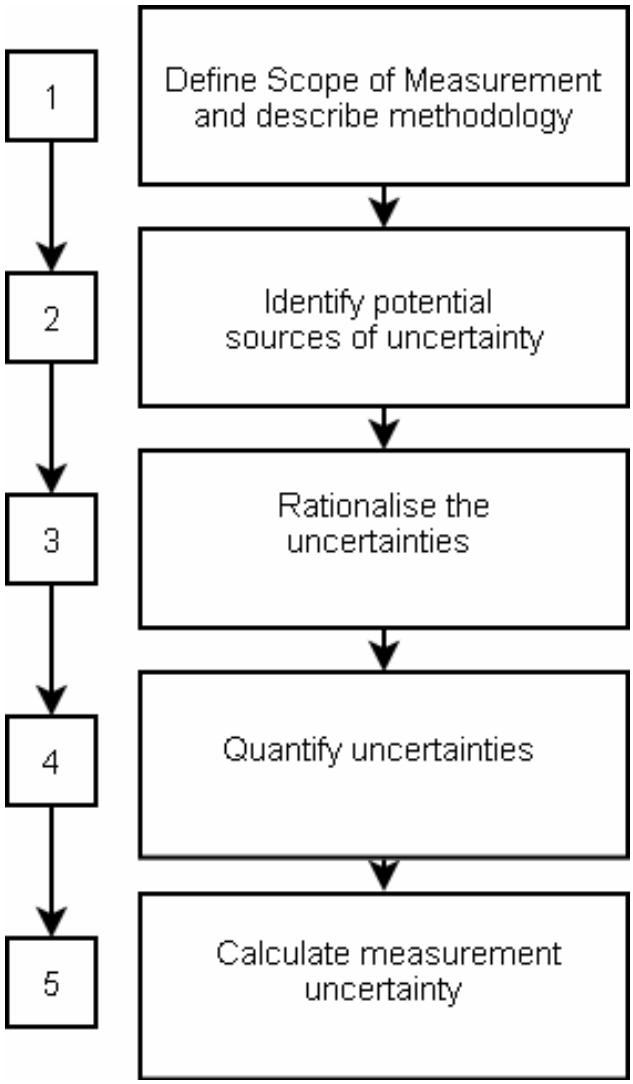

Figure 4 Steps to determine the measurement uncertainty

Step 1: Define scope of measurement and describe the methodology.

This is arguably the most important stage in determining the uncertainty of a measurement. It is necessary to fully understand the scope of the measurement in order to be able to assess all potential causes of error, and therefore calculate the measurement uncertainty. In order to allow the uncertainties to be determined, it is necessary to fully describe the measurement steps. This should follow naturally from module A (see Table 7) in the V1 validation protocol. All measurements, which are used to calculate the result, should be included, including any necessary calibration and QA/QC steps. Care will need to be taken in cases where the direct result of measurement is not the final quantity being assessed, but is in effect a surrogate indicator for this. In these cases, the results of previous validation studies may need to be assessed to determine potential uncertainty contributions introduced as a result of this.

Step 2: Identify potential sources of uncertainty.

Based on the description of the method, all sources of uncertainty should be identified. This is an extremely useful process to go through, as it provides a chance for a systematic review of the measurement process, enabling the laboratory to identify potential sources of error. In many cases, expert judgement can be used to quickly discount a step as a potential source of uncertainty. For example, certain QA/QC activities, although important for the correct application of the method, may not impact directly on the measurement result in a quantified way. Sources of uncertainty that should be identified include any measurements made and any other input quantities used in calculating the result (Note: this result is the quantity for which the uncertainty budget will be valid).

Step 3: Rationalise the uncertainty sources

This step involves a review of the identified sources of uncertainty – the aim of which is to identify any instances of double counting, to assess possible systematic (covariance) effects, and to group uncertainty sources together in ways which may facilitate their quantification. It may also be possible at this point to discount uncertainty sources which can be demonstrated to be insignificant. In many cases, a large proportion of the uncertainty sources may be combined in such a way that they may be assessed by validation studies (inter-laboratory and intra-laboratory studies).

Step 4: Quantify uncertainties

In this step, information on all the sources of uncertainties should be obtained in order to quantify their contribution to the measurement uncertainty. Much of this information may be derived from the validation studies carried out within the Validation Protocol.

Step 5: Calculate measurement uncertainty

The measurement uncertainty should be calculated, for example by combining the uncertainty contributions quantified in Step 4, in accordance with the rules of uncertainty propagation as described in the ISO Guide 98 (1995). These calculations should take into account correlation effects where appropriate. It is proposed that this should be simplified to either the no-correlation case or the fully correlated (in the case of any suspected significant level of correlation).

12.2.2 Guidance on the steps

12.2.2.1 Description of the methodology (Step 1)

A full method description is a requirement of the validation protocol, however it is also the first step in determining the measurement uncertainty. In terms of the requirements for the estimation of uncertainty, the description of the method should contain full experimental details of the method. However, in addition, it should define the scope of the method as comprehensively as possible. This includes the external conditions in which the method is valid, for example the range of ambient temperatures, and other external factors such as the matrices for which the method may be applicable. Some of these parameters may not be known at this stage, and may be determined during method validation. In addition it should be clear if the method actually involves the measurement of a surrogate indicator for the quantity which is desired to be measured.

An important aspect of the description of the method is the description on how traceability is achieved. If traceability for the entire measurement is obtained by one or more calibrations, traceable back to the SI, then the uncertainty of the method, including the uncertainty due to the calibration chain, will be determined. It is more common however that the calibration process only controls certain parts of the measurement chain, with certain aspects only controlled by QA/QC measures, or not controlled at all. Sampling aspects often fall into this category, and in such cases it may be that while the measurement will be made of a quantity in the environment, the measurement process itself can only really be considered to be a measurement of that quantity in the sample tested in the laboratory. A further issue arises where the calibration is to a CRM which is itself poorly determined, not traceable to SI, or is not representative of the measurand (and commonly not the matrix). In some cases suitably designed validation studies and QA/QC processes will enable the uncertainties due to these issues to be determined. Considerations such as these are important in determining the scope of the uncertainty evaluation.

As described in the ISO Guide 98 (1995), the validity of the uncertainty evaluation is directly related to the level of understanding and detailed knowledge of the method.

Once a review of the measurement scope and the methodology has been made, the measurement process should be described, either as a documented description or as a measurement equation. As well as detailing all calculations, and measurements and other input parameters which are directly used in determining the result, the measurement equation should also contain terms which represent potential influences on the measurement.

12.2.2.2 Identification of potential sources of uncertainty (Step 2)

In identifying these terms one approach is to take the documented description of the method, and ensure that all steps are represented in the measurement equation. In addition, any influences that may not have been directly accounted for in the measurement equation should be included. The Eurachem 'Guide to uncertainty in analytical measurements' (Eurachem, 2000) provides guidance on identifying uncertainty sources.

On way to identify potential uncertainty sources is to consider all conditions which the method statement defines, for example a range of injection temperatures. The effect of varying the injection temperature across this allowable range is a potential uncertainty source.

For example, consider measurement in which a sample of soil or dust is collected, digested and then analysed for metal content and the concentration in the soil determined. The measurement equation, at one level, is simply the mass of metal determined from the analysis over the mass of the sample. Each step and measurement should be assessed for potential terms in the measurement equation. For example this may include purity requirements on reagents, calibrations, and the measurement steps included in generating

these. In addition, terms which may not be numerically included in the calculation of the measurement result, but which are considered to potentially have an unassessed impact on the result, should be added. For example this may include a conjectured effect of the ambient temperature of the analysis laboratory. The measurement equation then will have additional terms of the form $f(x_i)$ where f is a function of the influence factor x_i. The function, f, determines the effect a given change in the influence factor will have on the result, and is determined as the sensitivity coefficient to that influence effect, c_i, and can be determined during validation from ruggedness tests. At this stage, the list of potential uncertainty sources should be based on expert judgement and knowledge of the likely sources of uncertainty in the method. It is appropriate for this exercise to err on the side of caution and include as many sources of uncertainty as possible.

12.2.2.3 Review and rationalise the uncertainty sources (Step 3)

Once the uncertainty sources have been identified, the next stage is to review these. The aim of this process is to identify the most effective and efficient way to evaluate the uncertainties. The first stage of the review process should be to identify and exclude any potential sources of uncertainty which expert judgement and knowledge can reasonably be used to justify them as being regarded as insignificant. If it is known a particular analytical technique is not susceptible to, for example, ambient temperature variation, then this source of uncertainty can be excluded. The justification for such decisions should be recorded.

The uncertainty sources are then grouped into combined uncertainties which can be assessed by validation experiments. As an example, intra-laboratory reproducibility experiments may cover a number of uncertainty sources, including sample recovery, analytical repeatability and the variability in laboratory conditions. If the initial uncertainty assessment is carried out before the validation studies, the range of conditions which can be included in the validation studies may be tailored to optimise the information available for uncertainty evaluation. If this is not the case, the range of uncertainty sources covered by the validation results may not address all uncertainty sources. If this is the case, additional studies may be required, or the scope of the validity of the uncertainty calculation may be limited. In practice, this means that the validation of the method and the consequent measurement uncertainty will be limited to a range of conditions, and use of the method outside these tested conditions will not be possible without further validation.

The VAM report by Barwick & Ellison (2000) provides a description of an approach for listing uncertainty sources and then reviewing them to identify double counting, possible groupings and other issues. It also provides some detailed methods for using validation studies such as precision, trueness and ruggedness tests as inputs into the uncertainty evaluation. In addition, ISO 2088 provides details of the uncertainty contributions for a number of typical validation experiments, compliant with ISO Guide 98 (1995), including repeated measurements of reference materials and comparisons with reference methods.

12.2.2.4 Quantify uncertainties (Step 4)

Within the validation process there are a number of method performance characteristics that should be determined. These studies cover a number of potential uncertainty sources. Figure 5 provides an overview of the likely uncertainty information which may be provided from the validation studies undertaken during intra-laboratory validation.

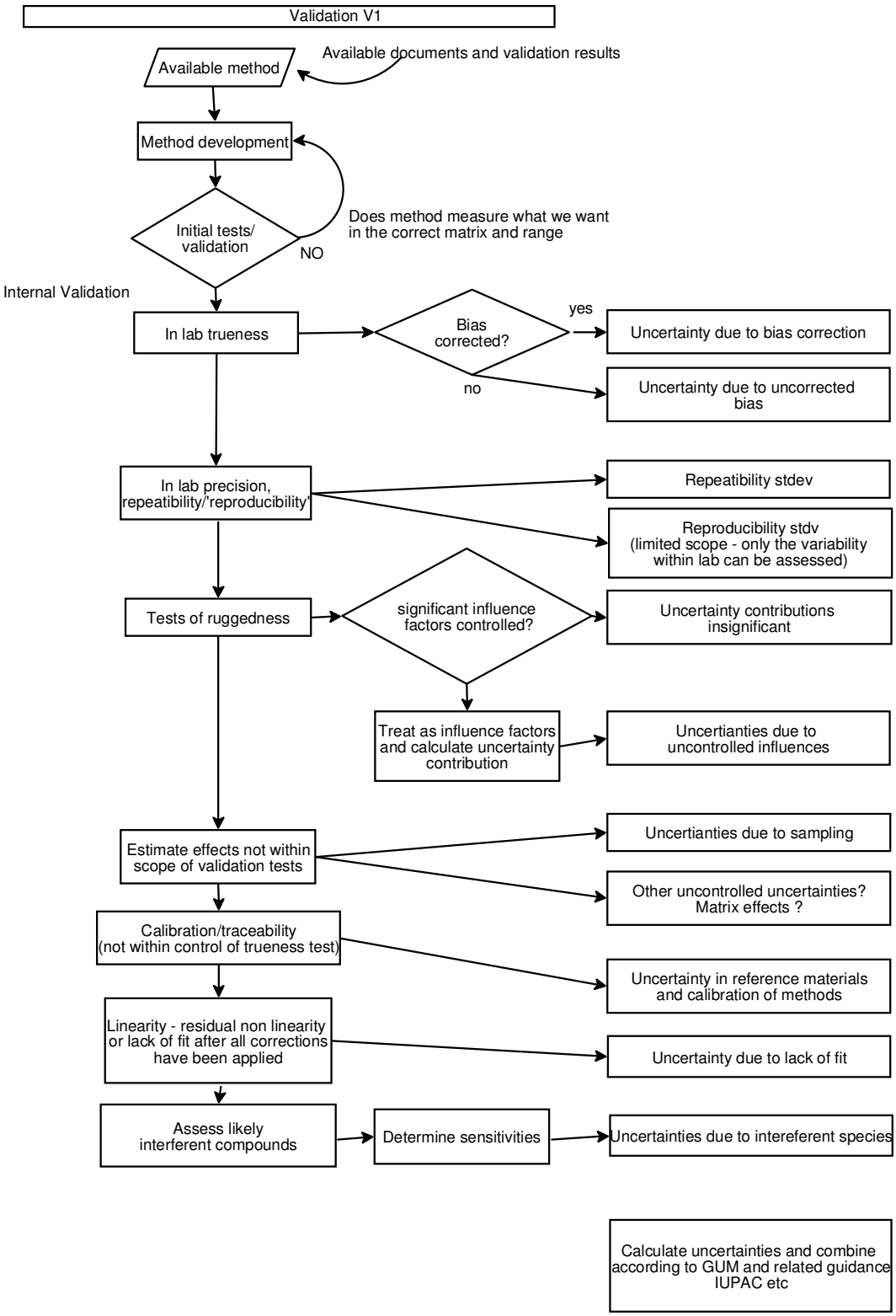

Figure 5 Possible sources of uncertainty from an intra-laboratory validation study

The following sections provide procedures for obtaining uncertainty contributions from these validation steps.

In reviewing which uncertainty terms are covered by each of the validation experiments it is necessary to have detailed knowledge of the conditions which were varied during the trials. For example, a reproducibility study in which the same matrix is used for all samples will not include uncertainty terms due to matrix effects

For environmental methods, the trueness study will often be primarily a measurement of the recovery of the sample. Where this has been determined out using repeated measurements of a CRM, the uncertainty of the CRM will have been included in the measured trueness. The uncertainty of the CRM value should however be stated on a certificate provided by the supplier of the CRM. It is impossible to remove this factor, unless multiple independent CRMs are available.

If trueness studies have been used to determine a method bias correction, which is subsequently applied to measurement results, the bias is removed by this correction, however there will be an uncertainty in the value of this bias term. This uncertainty will be due to the (necessarily) finite number of tests used to determine the bias. Where the bias is not corrected, then the value of the observed bias should also be included in the uncertainty term, as described in Section 4.2 of Barwick & Ellison (2000).

In the case of bias correction, the uncertainty in measuring the correction, u_b, is given by

$$u_b = \sqrt{\frac{u^2{}_r}{n} + \frac{u_c^2}{n_{cm}}}$$

where u_r is the repeatability standard deviation determined below and
n is the number of replicates in the bias factor determination,
u_c is the uncertainty of the CRMs, expressed as a standard uncertainty and
n_{cm} is the number of CRMs used.

To obtain the standard uncertainty from an uncertainty stated on a CRM certificate expressed as an expanded uncertainty, divide by the coverage factor used.

If no correction for bias is made the best estimate of the uncertainty due to bias u_b must also include the value of the uncorrected bias. As a best estimate, it can be treated as a rectangular distribution and the standard uncertainty calculated by dividing the mean bias, b, by the square root of 3 as described in the ISO Guide 98 (1995).

$$u_b = \sqrt{\frac{u^2{}_r}{n} + \frac{u_c^2}{n_{cm}} + \left(\frac{b}{\sqrt{3}}\right)^2}$$

where b is the mean bias.

Precision studies provide uncertainty values for a number of uncertainty sources which have been included in the experiment, i.e. those conditions which have been varied during the study. In general, the precision determination in which the most parameters have been varied (i.e. a reproducibility study) should be used. In these cases, the information from a repeatability study should not be used, as this would result in double counting. However, the result of the repeatability study, i.e. the basic variability of the method when nothing is varied, may be put to use. It can be used to facilitate the design of other tests, to ensure sufficient repeat measurements are used so that the repeatability has insignificant impact on the average values (i.e. the repeatability of the mean is small enough that double counting of the uncertainty due to repeatability does not occur). An example of this was given above in the determination of the bias correction factor. It can also be instructive to compare the repeatability with reproducibility, to gain an insight into the effect on the uncertainty of the additional parameters which have been varied during the reproducibility study. Such investigations can inform further studies with the aim of method optimisation or improvement.

The uncertainty due to repeatability, u_r, is determined from the standard deviation of the series of repeatability measurements.

$$u_r = s_r$$

where s_r is the standard deviation of the results.

The uncertainty due to reproducibility, u_R, can be estimated from a collaborative study by

$$u_R = \sqrt{s_L^{\,2} + s_r^{\,2}}$$

where s_L is the inter-laboratory standard deviation.

If u_R is determined by the V1 laboratory by an intra-laboratory study, where as many parameters are varied as possible, then an estimate of u_R is

$$u_R = s_R$$

where s_R is the standard deviation of the repeated measurements.

Ruggedness tests can be used to assess the sensitivity of the method to external influences not covered in the reproducibility studies. In order to assign an uncertainty contribution for a given influence parameter it is necessary to define over what range the parameter will vary. For instance, if it is determined that the sensitivity of a given method to ambient temperature is 1 µg/m^3 change per 1°C change in temperature, then the uncertainty contribution from this influence will depend on the expected change in temperature during the measurements. Where a method is controlled by calibration, the actual influence effect will be the variation in the ambient temperature from the conditions under which the calibration was carried out. This can lead to a systematic uncertainty contribution (bias) if, for example the method is always calibrated at room temperature, but analyses are then carried out in the field at say a different temperature of 5°C. In many cases the method statement will prescribe the range of conditions over which the method is valid, and the uncertainty can be determined using this range. The standard uncertainty terms u_i are expressed as standard deviations:

$$u_i^{\,2} = \sum \left| c_i u(x_i) \right|^2$$

where c_i is the sensitivity coefficient for the influence factor i, and
$u(x_i)$ is the variation expected in the influence quantity, expressed as a standard deviation.

In the example above, if the ruggedness test showed that there was a 1 µg/m^3 change per 1°C change in temperature then c_i would be 1. If the effect was a 0.1 µg/m^3 change per 1°C change in temperature then c_i would be 0.1. The value x_i would depend on the range of temperature variability which is allowed in the method. If the method were to be used with ± 10 degree temperature range (i.e., from the temperature at which it is calibrated) then x_i would be $10/\sqrt{3}$, assuming the variation in x as a rectangular distribution.

A specific set of influence quantities relate to compounds or matrix elements which act as interferent quantities on the analysis of the measurand. These can be assessed by spiking or direct measurement, to determine their sensitivity coefficients. Once again the expected range of these influence quantities needs to be defined in order to calculate their uncertainty contributions. One aspect of these, discussed in ISO 14956, is that often the levels of these interfering compounds in the sample will be correlated. This can lead to a significant covariance between them, and ISO 14956 suggests they should be treated as covariant, and

their individual uncertainty contributions summed directly, rather than summed in quadrature as for other uncorrelated uncertainty contributions. The individual uncertainty contributions of each interferent compound can be determined as above for an influence quantity, by multiplying the sensitivity coefficient for that compound by its expected variation. The combined uncertainty due to interferents, u_{int}, is then calculated as the direct sum of these terms rather than summing in quadrature, for example the uncertainty due to a number, j, of interferent compounds would be

$$u_{\text{int}} = \sum u_j$$

where u_j is the uncertainty for each interferent compound j determined as for other influence quantities.

The final set of uncertainty contributions are those not covered by the trueness, precision and ruggedness tests. These can include external uncertainties invariant during the tests, such as the traceability chain back to SI or other reference measurements and other influence quantities which are determined from expert knowledge.

12.2.2.5 Calculate measurement uncertainty (Step 5)

The uncertainty contributions quantified in the previous steps should then be combined to determine the overall uncertainty. This is carried out in accordance with the process defined in the ISO Guide 98 (1995), by combining the uncertainties in quadrature. The combined standard uncertainty is therefore:

$$u = \sqrt{u_b^2 + u_R^2 + u_i^2 + u_{\text{int}}^2}$$

The final step in the uncertainty evaluation is to expand the uncertainty to a given confidence level. This is usually to a 95% level of confidence. This enables uncertainties, determined by different approaches, and the uncertainties of different measurement methods, to be compared. To determine the uncertainty at 95% confidence the combined uncertainty should be multiplied by a coverage factor (k). To determine this rigorously, requires an assessment of the degrees of freedom of the full uncertainty assessment. However, if the evaluation studies carried out to determine the performance characteristics of the method have followed the guidance in this protocol, and have provided results which are statistically valid, then it may be assumed that, to a good approximation, the coverage factor will be k= 2. The measurement uncertainty U is therefore given by:

U = k u = 2 u