

Wolfgang Kandler  
Rainer Schuhmacher  
Susanne Roch  
Patricia Schubert-Ullrich  
Rudolf Kraska

## Evaluation of the long-term performance of water-analyzing laboratories

Received: 16 June 2003  
Accepted: 6 August 2003  
Published online: 11 October 2003  
© Springer-Verlag 2003

W. Kandler (✉) · R. Schuhmacher  
S. Roch · P. Schubert-Ullrich · R. Kraska  
Institute for Agrobiotechnology  
(IFA-Tulln),  
Centre for Analytical Chemistry,  
Konrad-Lorenz-Straße 20, 3430 Tulln,  
Austria  
e-mail: kandler@ifa-tulln.ac.at

**Abstract** After 7 years of successful operation, the database of the proficiency testing (PT) scheme for water analysis organized by the Institute for Agrobiotechnology (IFA), Tulln contains nearly 4,000 data sets from over 300 interlaboratory comparison samples. About 70 analytical parameters (major ions, metals, trace ions, herbicides, volatile halogenated hydrocarbons, organochlorine pesticides and polycyclic aromatic hydrocarbons) were covered by the PT scheme. The data were evaluated using robust statistics in order to determine a set of coefficients of variation

(CV) for each analytical parameter. Concentration and time dependence of the CV were checked. The CV were combined to obtain standard deviations for proficiency assessment (Z-score criteria). Furthermore, a viewer program was designed to facilitate monitoring of the analytical performance of participating laboratories.

**Keywords** Interlaboratory comparisons · Proficiency testing · PT scheme · Z-scores · Water analysis

### Introduction

In the early 1990s a countrywide monitoring network for ground- and surface water quality was built up on a legally regulated basis in Austria. Since then about 2,000 groundwater and 240 river water sites have been tested regularly and the water samples have been analyzed for numerous inorganic and organic parameters [1, 2]. Such a permanent and comprehensive measurement program requires reliable and accurate results of analysis. Thus, a sophisticated quality assurance program had to be implemented [3] and the compulsory participation in interlaboratory comparisons was an integrative part of this quality assurance program. Due to the rather infrequent organization of suitable interlaboratory comparisons, the Federal Ministry for Forestry and Agriculture, Environment and Water Management decided to initiate a regional PT scheme. In 1995, the IFA-Tulln Centre for Analytical Chemistry started to develop a PT scheme called the "IFA-Test Systems PT Scheme" for external quality assurance in water analysis. The scheme started

with regular interlaboratory comparisons for inorganic parameters in water (major ions, heavy metals, and trace elements) and was successively extended to a multiplicity of inorganic and organic parameters (trace ions, herbicides, volatile halogenated hydrocarbons, organochlorine pesticides, polycyclic aromatic hydrocarbons (PAHs) and, recently, methyl-tert-butylether (MTBE). Depending on the parameter group, the number of series amounted to 1 to 6 per year and group. The operation of the scheme has closely followed the framework given in ISO/IEC Guide 43 [4]. The series reports provide data that allow evaluation of laboratories' performance, but do not contain any performance evaluation. Calculation of performance statistics using Z-scores and evaluation of performance will be implemented in the near future. This implementation will make the PT scheme fully consistent with [4]. The series reports are published on the web site of the PT scheme. Preparation of samples, uncertainties of the target values, and behavior of the samples for the interlaboratory comparisons are described elsewhere [5, 6, 7, 8] or will be published in the near

future. Participation in the PT scheme has always been open to all interested laboratories. Thus after 1 year of operation, the number of voluntarily participating laboratories superseded the number of compulsory participants. In 2002, the ratio of voluntary to compulsory participants was 4:1. Over 100 different Austrian and foreign laboratories have participated in the IFA-Test PT Scheme.

### Course of PT rounds

Within the IFA-Test PT Scheme every round is a complete small (5 to 50 participants) interlaboratory comparison study. Normally, two different synthetic water samples with well-defined target concentrations are sent out in a round. The samples are identified by a letter code for the parameter group and the series number, e.g., H36A and H36B for the two samples from the 36th PT round on herbicides. In total, 190 PT rounds were conducted by June 2003: 56, 40, 19, 9, and 8 rounds on inorganic parameters, herbicides, volatile halogenated hydrocarbons, organochlorine pesticides and PAHs, respectively. At the end of each round the laboratories' measurement results are transmitted and collected in a database. The measurement results of a certain parameter in one specified sample together with its basic statistical evaluation is called a data set in this paper. An example of a typical data set is given in Fig. 1. Not all substances of a certain parameter group are present in every sample, so the number of data sets available for every parameter may vary widely. Each PT round is completed by the release of a series report. The reports are designed to meet the requirements of the laboratories. They illustrate graphically all recoveries of each participating laboratory, which allows evaluation of the analytical performance of the laboratory. The reports also contain diagrams for each analytical parameter with a brief statistical evaluation: laboratory mean, standard deviation between the laboratories, and confidence intervals (Fig. 1). This allows identification of possible analytical problems for some parameters and yields a rough estimate of the accuracy that can be expected from an analytical laboratory. Thus, the series reports facilitate the comparison of the analytical performance of a selected laboratory with the performance of other laboratories.

### Evaluation of long-term performance

For both the end-users of the analytical results and the laboratories, the answers to the following questions may be of interest: How does the analytical performance of a selected (contract) laboratory behave over a longer period of time? Do significant deviations of results from the target values occur often or rarely? What deviations be-

Lab code	Result	Out	+/-	Unit	Recovery
A	7.9	*	2	mg/L	25%
B	31		1.0	mg/L	100%
C	32.7		0.1	mg/L	105%
D	33.6	*		mg/L	108%
E	30.6		2.4	mg/L	98%
F	30.0		1.5	mg/L	96%
G				mg/L	
H	33.2		1.4	mg/L	107%
I	31.1		0.4	mg/L	100%
J				mg/L	
K	30.7		0.5	mg/L	99%
L	31		1.6	mg/L	100%
M	29.6		1.5	mg/L	95%
N				mg/L	
O	31.311		0.375	mg/L	101%
P	31.3		1.07	mg/L	101%
Q	30.4		0.49	mg/L	98%
R	30.7		1.02	mg/L	99%
S	31.8			mg/L	102%

	All results	Outliers excl.	Unit
Mean +/- CI (99%)	29.8 ± 4.4	31.1 ± 0.8	µg/l
Recov. +/- CI (99%)	95.8 ± 14.1	100.0 ± 2.5	%
SD between labs	5.9	1.0	µg/l
RSD between labs	19.9	3.1	%
n for calculation	16	14	

**Fig. 1** Example data set: results for nitrate in water sample N52B with basic statistical evaluation. The target value was 31.1 mg/L nitrate

tween a measurement result and the assigned reference concentrations can be expected for each parameter and which parameters or analytical techniques cause most problems? The present project was initiated to answer these questions on the basis of the collected results of all previous PT rounds. It was also designed to be a helpful tool for monitoring the performance of the contracted laboratories.

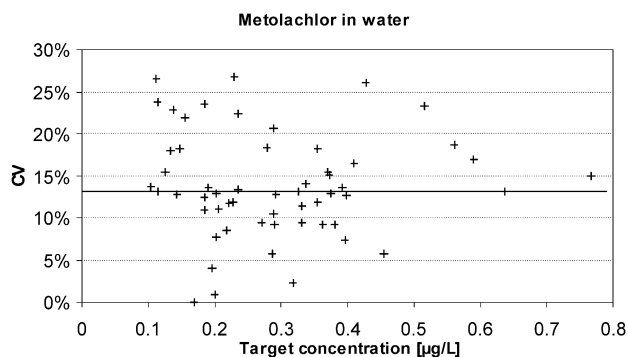
### Aim of this study

The combination of scores, like the sum of (squared) scores and running scores, is a common technique for the assessment of long-term trends [9]. This paper describes the determination of performance criteria (Z-score criteria  $\sigma$ ) derived from the results of multiple PT rounds. The results of 190 PT rounds were evaluated and coefficients of variation (CV) were calculated from every data set as estimates for the between laboratory precision ("estimates of dispersion"). The CV were visualized for every parameter over the target concentrations and the operation time of the PT scheme in order to check the influence of concentration and time. The CV of every analytical parameter were combined to a single Z-score criteria  $\sigma$  representing the average laboratory performance over a given concentration range.

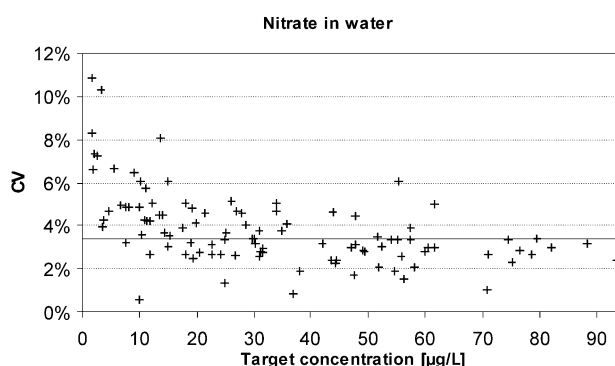
These performance criteria are much more reliable than those calculated from single rounds. They are highly suitable for assessment of long-term trends via Z-scores and combination scores.

## Realization

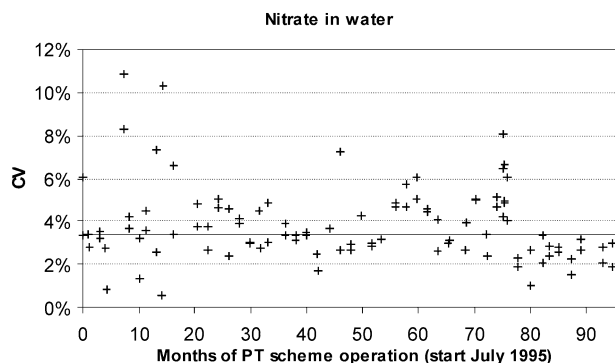
First, the laboratories' results of all interlaboratory comparisons were transformed into a uniform format and collected in a single database. At this stage of the project, 130 different laboratories had participated in the scheme. The total number of analytical parameters was 67 and the database contained nearly 4,000 data sets containing more than 60,000 analytical results from about 300 interlaboratory comparison samples. All data sets were checked by the Hampel outlier test, which is often referred to as "Huber test". This outlier test is based on the use of robust statistics. Davies suggested using it for the statistical evaluation of interlaboratory tests in 1988 [10]. The Hampel test has been used in the IFA-Test PT Scheme since its first year of operation. Laboratory mean values, mean recoveries, standard deviations, and CV were calculated from both the whole sets of data and the outlier-free sets of data. The agreement of laboratory mean values and target concentrations is permanently observed in the PT scheme. At this stage it could be shown that even for the evaluation of parameters over consecutive rounds no statistically significant deviations between target concentrations and outlier-corrected laboratory mean values were found. This is a basic requirement for a reliable estimation of dispersion. Furthermore, robust estimates of central tendency  $\mu'$  ("robust average") and dispersion  $\sigma'$  ("robust standard deviation") were calculated from all data sets according to "algorithm A" of the international standard ISO 5725-5 [11]. Both the Hampel test and algorithm A start with the median as "robust average"  $\mu'_0$  and a "robust standard deviation"  $\sigma'_0$  that is calculated from the median absolute deviation (MAD) multiplied by a factor (1.483 for  $n \rightarrow \infty$ ). The Hampel test simply rejects all results that deviate by more than  $3 \sigma'_0$  from  $\mu'_0$ . Algorithm A progressively transforms the original data by replacing results that deviate by more than  $1.5 \sigma'_0$  from  $\mu'_0$  and recalculating  $\mu'$  and  $\sigma'$ . According to this approach, the values are updated several times until the process converges. This process is called winsorization [12]. The data presented in this paper were obtained by algorithm A. CV were calculated from the estimates  $\mu'$  and  $\sigma'$  of all data sets with three or more measurement results. These CV were graphically depicted against the target concentrations for each of the analytical parameters. Two examples are shown in Figs. 2 and 3. In order to check whether there were visible changes during the operation time of the scheme, the CV were graphically depicted versus time, too. An example is shown in Fig. 4.



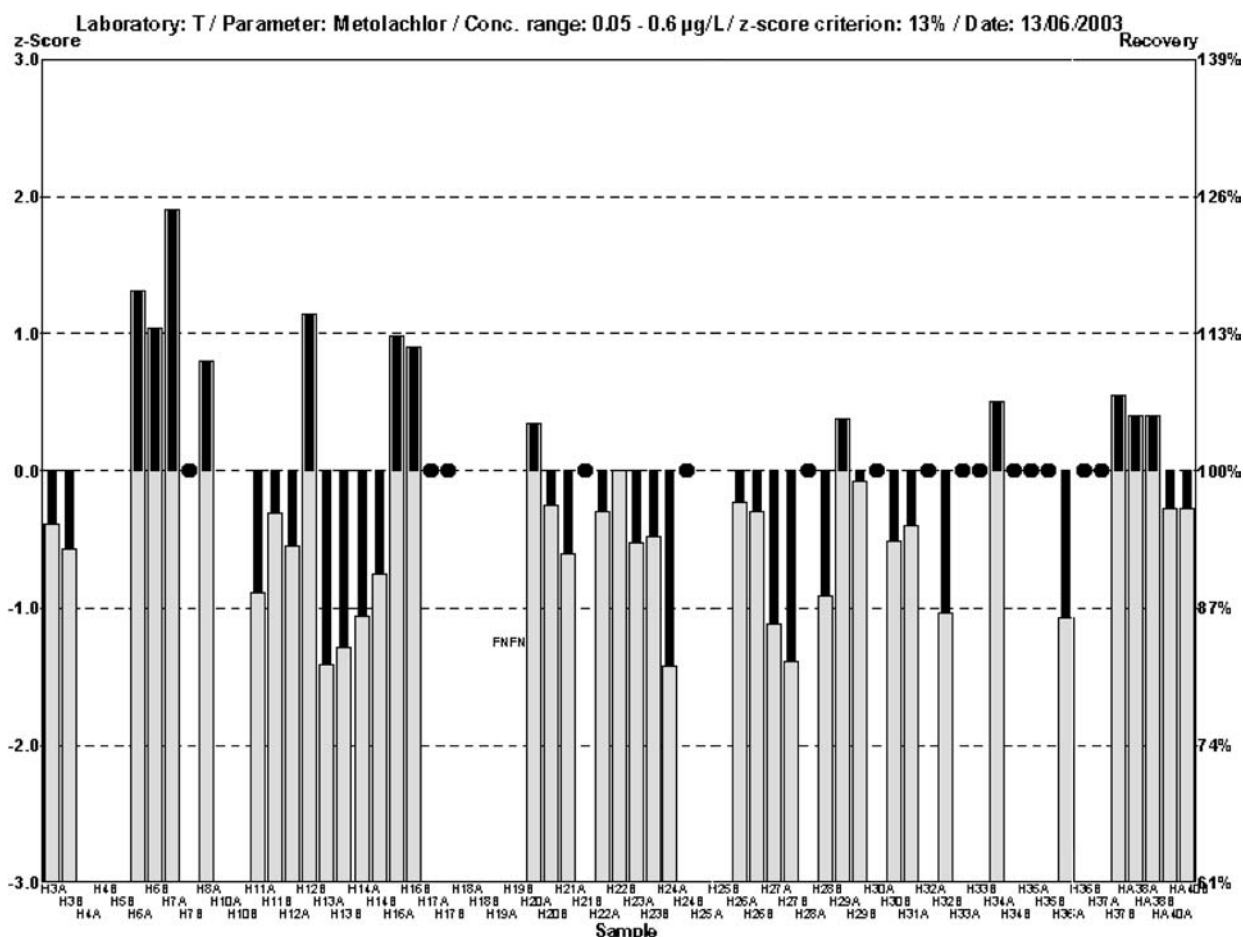
**Fig. 2** Determination of metolachlor in 56 water samples sent out in 40 proficiency testing (PT) rounds. The diagram shows the coefficients of variation (CV) between laboratories calculated by algorithm A versus target concentrations. The line at 13% indicates the median of the CV



**Fig. 3** Determination of nitrate in 103 water samples sent out in 50 PT rounds. The diagram shows the CV between laboratories calculated by algorithm A versus target concentrations. The line at 3.4% indicates the median of the CV. The concentration range of application of the Z-score criteria was limited to  $>3$  mg/L



**Fig. 4** The data of Fig. 3 versus operation time of the PT scheme. The diagram shows the CV between laboratories calculated by algorithm A versus months from July 1995. Each "+" represents a dispersion estimate for one water sample. The line at 3.4% indicates the median of the CV



**Fig. 5** Z-score diagram of a selected laboratory which regularly participated in the PT scheme. The diagram combines the recoveries of Metolachlor in samples from 40 PT rounds between September 1996 and May 2003. ● Indicates that no Z-score could be calculated. Missing bars indicate that no results were submitted. “FN” refers to false negative results

In order to combine the CV for each parameter, the medians of the CV were taken as Z-score criteria. Thus, one Z-score criterion was defined for every parameter. In some cases the concentration range had to be limited at the lower end to make the criteria applicable, e.g., 3 mg/L for nitrate (Fig. 3). The criteria and their concentration limits were added to the database. In parallel, a computer program (a viewer) was developed which allowed convenient access to the laboratory results' database. The viewer creates Z-score diagrams and tables after selection of laboratory, parameter group, one or more parameters, and a time interval. As the viewer was designed to facilitate monitoring of the analytical performance of contracted laboratories, sub-databases that contained exclusively results of these laboratories were created for this purpose. An example of the viewers' capability is given in Fig. 5. A corresponding table

contains target concentrations, results of the laboratory, recoveries, and the Z-scores. The viewer program supports the calculation of the sum of the Z-scores (SZ) [9], too.

## Results and discussion

### Estimates of dispersion

Our comparison of the Hampel test and algorithm A showed both procedures yielded very similar estimates,  $\mu'$  and  $\sigma'$ , when they were applied on the data sets of the interlaboratory comparisons. Thus, the Hampel test and algorithm A can be considered to be equivalent for this task. Algorithm A was chosen for calculating CV because of its “more official character” as this procedure is recommended by ISO standards [11, 13].

### Precision of interlaboratory comparisons

The procedure described above yielded a set of CV for every parameter (examples are shown in Fig. 2 and Fig. 3). It is possible to calculate CV from these sets.



These “CV of the CV” obtained from different samples for each parameter were typically in the range of 40% to 50%. This is well above the theoretical value that can be expected for random samples of one population. It can be calculated by:

$$\sqrt{\frac{1}{2n}}$$

where  $n$  is the number of results per sample [14]. This would be, e.g., 16%, for 20 results per sample. It was surprising too that 40%–50% were found in all parameter groups. As the average number of results ranged from 7 (herbicides) to 20 (major ions), it can be concluded that the observed “CV of the CV” were not dependent on the actual number of results per sample. There may be several reasons for this: The field of participants in the IFA-Test PT Scheme is not homogenous. The laboratories vary widely with the PT rounds, as the frequency of the series (3 or 6 times a year) is higher than the participation frequency of the laboratories. The observed CV are most probably not normally distributed. It may also be possible that the sample-to-sample deviation of the CV is partially caused by the use of the estimation procedure for  $\mu'$  and  $\sigma'$ , the algorithm A. This will be the subject of further investigations. Anyhow, it should be widely considered that the accuracy of the CV determined from results of single interlaboratory comparisons might be low, even if the number of results is above 30, which is the number that is generally recommended for determination of CV [15]. Thus, for the definition of Z-score criteria it might be most useful to combine the results of multiple (periodical) interlaboratory comparisons, as presented in this paper.

### Combination of CV

In the present study, the between-laboratory dispersion estimates of the results of many interlaboratory comparisons (6–106 data sets per parameter, on average 44 data sets) were combined. As the distribution of the single CV was not clearly identified, the median was chosen to determine the Z-score criteria from the CV. There may be doubts about the usefulness of applying consecutively robust statistics to a dataset, so the mean CV was calculated as square root of the arithmetic mean of the variances, too. The latter method yielded slightly higher Z-score criteria. All results are summarized in Table 1.

### Dependence of CV on concentration levels

In general no significant dependence of the CV on the actual concentration of the parameter in the sample was observed. One reason may be the relatively wide spread of the CV (see above). Another reason was the narrow

concentration ranges, which were covered in the PT system. The target concentrations were within 1 order of magnitude for most parameters in the PT scheme. For some parameters obviously higher CV were found for low target values which were near the limits of detection that are demanded in the Austrian water monitoring program [16]. An example is given in Fig. 3. In this case the concentration range, in which the Z-score criteria can be applied, was limited. The lower limits given in Table 1 refer to the applicability of the determined Z-score criteria. They consider the legal limits of determination and the observed concentrations, too. In our experience, the introduction of variance functions as suggested, e.g., in [14], cannot be justified in a concentration range within 1 order of magnitude. Besides, single Z-score criteria are much more transparent to the participants.

### Time dependence

No statistically clear trend to lower CV over the operation time of the PT scheme could be observed. One reason may be the relatively wide spread of the CV (see above). The CV of “nitrate in water” (Fig. 4) were all below the median line in the last year. Most probably the laboratories are improving their analytical performance by various measures like modernization of instrumentation and improving their quality assurance systems. So this will be seen in the future development of the CV values.

## Conclusion and outlook

The collection of all measurement results in a single database in combination with computer-program controlled SQL queries appears to be a very powerful tool for the statistical evaluation of the collected data. The estimates of dispersion obtained by robust statistics have an unexpected low precision with typical CV of 40%–50% between the data sets of a selected parameter. Thus, Z-score criteria should never be calculated from a single PT round even if the number of participants is considerably high. No clear influence of concentration and time on the dispersion estimates was observed. Higher CV between laboratories were only found for concentrations close to the determination limits of the analytical methods. In our experience, a single Z-score criterion is applicable over a concentration range of 1 order of magnitude and we do not recommend the use of variance functions. Performance criteria evaluated from a number of PT series are the most suitable and transparent for the assessment of long-term performance. The Z-score criteria of Table 1 represent the average performance of (mainly) Austrian laboratories for the specified parameters and concentration ranges between 1997 and 2003. This limits their applicability and should be considered if

**Table 1** Overview of the evaluated parameters. The medians were used as Z-score criteria: CV, coefficients of variation

Parameter	Number of data sets n	Lower limit	Target value			Number of results in the data sets			Averaged CV	
			Min	Max	Unit	Min	Max	Mean	Median %	Mean (by variances) %
<b>Major ions</b>										
Ammonium	72	0.01	0.012	0.6	mg/L	5	43	19	14	20
Boron	88	0.02	0.02	1.1	mg/L	5	31	15	9.3	13
Calcium	106	5	4	71	mg/L	6	46	20	4.5	5.9
Chloride	105	3	1.1	82	mg/L	7	46	21	4.0	7.2
DOC	87	1	0.4	7	mg/L	5	23	14	11	16
Total hardness	82	0.2	0.06	3	mmol/L	12	45	23	3.6	4.1
Hydrogen carbonate	104	30	14	270	mg/L	5	44	19	4.0	6.2
Potassium	104	2	0.5	20	mg/L	7	41	19	7.2	8.9
El. conductivity	106	10	80	1000	µS/cm	6	49	22	1.9	2.4
Magnesium	106	2	1.2	24	mg/L	6	46	20	5.0	6.5
Sodium	106	1	1.6	93	mg/L	7	43	19	5.1	5.4
Nitrate	103	3	1.5	94	mg/L	7	46	22	3.4	4.2
Nitrite	87	0.01	0.01	0.2	mg/L	6	42	21	7.0	9.8
Orthophosphate	66	0.02	0.02	0.4	mg/L	4	41	19	10	13
Alkalinity	48	0.1	0.3	4	mmol/L	10	40	22	3.6	4.7
Sulphate	105	3	3	140	mg/L	7	44	22	3.5	4.3
<b>Metals</b>										
Aluminium	33	0.01	0.02	0.09	mg/L	10	22	16	13	15
Arsenic	47	0.001	0.0014	0.04	mg/L	5	25	14	13	15
Lead	49	0.005	0.003	0.08	mg/L	6	26	16	10	15
Cadmium	49	0.0005	0.0004	0.007	mg/L	6	28	15	11	14
Chromium	46	0.002	0.0012	0.1	mg/L	6	27	16	9	11
Iron	52	0.02	0.02	0.4	mg/L	4	25	15	10	14
Copper	35	0.01	0.006	0.13	mg/L	10	27	19	9	11
Manganese	49	0.02	0.008	0.12	mg/L	7	27	16	7	10
Nickel	32	0.001	0.005	0.14	mg/L	10	29	19	9	10
Mercury	43	0.0002	0.0003	0.006	mg/L	5	25	14	16	20
Zinc	34	0.02	0.012	0.18	mg/L	10	26	17	11	11
<b>Herbicides</b>										
2,4-D	34	0.05	0.08	0.7	µg/L	3	7	4	20	29
Alachlor	48	0.09	0.09	0.7	µg/L	4	12	8	15	18
Atrazine	74	0.05	0.06	2.3	µg/L	4	28	9	13	15
Bentazon	58	0.05	0.06	0.8	µg/L	3	7	4	13	21
Cyanazine	69	0.05	0.05	1.8	µg/L	3	16	8	16	17
Desethylatrazine	75	0.05	0.07	1.5	µg/L	3	19	8	16	18
Desisopropylatrazine	69	0.05	0.07	2.0	µg/L	3	17	8	19	23
Diuron	30	0.05	0.07	0.5	µg/L	3	17	5	20	22
Hexazinone	32	0.05	0.06	0.5	µg/L	3	12	5	18	24
Linuron	29	0.05	0.08	0.5	µg/L	3	15	5	19	26
Mecoprop	31	0.05	0.08	0.4	µg/L	3	7	4	14	22
Metalaxyl	33	0.05	0.10	0.4	µg/L	3	12	4	13	20
Metolachlor	56	0.05	0.10	0.8	µg/L	3	13	8	13	15
Pendimethalin	49	0.05	0.06	0.7	µg/L	4	14	8	22	27
Prometryn	69	0.05	0.07	0.8	µg/L	4	18	9	15	18
Propazine	74	0.05	0.06	1.7	µg/L	3	19	8	13	15
Pyridate	35	0.05	0.14	1.0	µg/L	3	5	4	14	32
Sebuthylazine	62	0.05	0.06	1.9	µg/L	4	12	8	12	14
Simazine	72	0.05	0.05	1.9	µg/L	3	28	9	13	16
Terbuthylazine	76	0.05	0.06	4	µg/L	3	22	9	14	15
Terbutryn	69	0.05	0.05	0.7	µg/L	3	16	8	17	20
<b>Volatile halogenated HC</b>										
1,1,1-Trichloroethane	35	0.1	0.3	3	µg/L	6	36	20	18	22
1,1-Dichloroethene	30	1.0	1.0	10	µg/L	11	24	18	24	27
1,2-Dichloroethane	28	1.0	0.7	35	µg/L	8	29	17	17	20
Bromodichloromethane	25	0.1	0.2	2	µg/L	5	35	19	17	16
Dibromochloromethane	29	0.1	0.4	4	µg/L	9	35	20	14	15
Dichloromethane	28	1	0.9	37	µg/L	9	27	17	19	21

**Table 1** (continued)

Parameter	Number of data sets	Lower limit	Target value			Number of results in the data sets			Averaged CV	
			Min	Max	Unit	Min	Max	Mean	Median %	Mean (by variances) %
Tetrachloroethene	32	0.1	0.2	3	µg/L	12	37	20	17	19
Tetrachloromethane	34	0.1	0.2	3	µg/L	6	35	19	21	24
Tribromomethane	32	0.1	0.3	3	µg/L	5	33	17	16	16
Trichloroethene	33	0.1	0.2	3	µg/L	12	36	20	19	21
Trichloromethane	32	0.1	0.5	4	µg/L	12	35	12	17	18
PAH										
Benzo(a)pyrene	15	0.01	0.02	0.18	µg/L	4	26	15	28	31
Benzo(b)fluoranthene	16	0.02	0.02	0.14	µg/L	4	26	15	19	21
Benzo(ghi)perylene	16	0.01	0.012	0.15	µg/L	4	24	15	31	36
Benzo(k)fluoranthene	15	0.01	0.03	0.14	µg/L	4	25	14	19	26
Fluoranthene	16	0.01	0.016	0.3	µg/L	4	26	16	16	19
Indeno(1,2,3-cd)pyrene	16	0.01	0.015	0.12	µg/L	4	26	15	41	42
Organochlorine pesticides										
1,2,3-Trichlorobenzene	12	0.05	0.07	0.9	µg/L	5	8	6	19	19
1,2,4-Trichlorobenzene	12	0.05	0.08	0.9	µg/L	5	8	6	17	21
1,3,5-Trichlorobenzene	12	0.05	0.05	0.9	µg/L	5	8	6	24	27
Aldrin	11	0.05	0.11	1.2	µg/L	7	12	9	24	25
alpha-HCH	13	0.05	0.13	0.9	µg/L	5	7	9	12	13
beta-HCH	6	0.05	0.14	0.4	µg/L	5	9	7	15	15
cis-Chlordan	13	0.05	0.08	0.7	µg/L	6	12	9	20	23
cis-Heptachlorepoxyde	8	0.05	0.11	0.4	µg/L	6	10	8	18	24
Dieldrin	16	0.05	0.13	0.9	µg/L	3	12	9	15	17
Endrin	15	0.05	0.10	1.0	µg/L	3	10	8	17	17
gamma-HCH	15	0.05	0.05	0.7	µg/L	3	12	8	12	13
Hexachlorobenzene	12	0.05	0.07	0.6	µg/L	8	12	10	22	22
Isodrin	11	0.05	0.09	1.3	µg/L	5	8	6	17	19
Methoxychlor	11	0.05	0.12	0.9	µg/L	6	11	9	21	27
o,p-DDE	12	0.05	0.08	0.7	µg/L	7	11	8	25	26
o,p-DDT	10	0.05	0.09	0.7	µg/L	6	12	8	34	35
p,p-DDE	12	0.05	0.12	0.9	µg/L	7	12	9	25	27
p,p-DDT	12	0.05	0.10	0.5	µg/L	6	12	9	33	33
Pentachlorobenzene	12	0.05	0.09	0.8	µg/L	5	8	6	16	17
Pentachloronitrobenzene	12	0.05	0.14	1.0	µg/L	5	8	6	30	27
trans-Chlordane	14	0.05	0.13	0.5	µg/L	6	11	9	24	27
trans-Heptachloroepoxyde	14	0.05	0.07	0.9	µg/L	7	11	8	20	23

the data is used for calculating Z-scores. The performance criteria will be implemented in the scheme for calculating performance scores for all participants on a voluntary basis in the near future. Assuming that there is, in fact, progress in the laboratories' average per-

formance, the Z-score criteria will have to be adapted periodically.

**Acknowledgements** The authors gratefully acknowledge the financial support of the Federal Ministry for Forestry and Agriculture, Environment and Water Management.

## References

- Philippitsch R, Schimon W (2003) Oesterr Wasser Abfallwirtsch 55(3-4):41-53
- Kralik M, Grath J, Philippitsch R, Gruber D, Vogel W (1999) Groundwater quality in Austria. A unique groundwater monitoring system. In: Extended abstracts of international conference on quality, management and availability of data for hydrology and water resources management. Koblenz, pp 89-90
- Philippitsch R, Wegscheider W (1999) Quality assurance program for the assessment of water quality in Austria. In: Extended abstracts of international conference on quality, management and availability of data for hydrology and water resources management, Koblenz, pp 115-118

- 
4. ISO (1997) Proficiency testing by interlaboratory comparisons—Part 1: Development and operation of proficiency testing schemes, ISO/IEC Guide 43-1:1997. ISO, Geneva
  5. Linsinger TPJ (1997) Development of quality assurance systems for the determination of trace components in water under consideration of natural constituents. Dissertation, Vienna University of Technology
  6. Kandler W (1999) Aufbau und Betrieb eines Kontrollprobensystems zur Qualitätssicherung in der Wasseranalytik. Dissertation, Vienna University of Technology
  7. Apfalter S, Krska R, Linsinger T, Oberhauser A, Kandler W, Grasserbauer M (1999) Fresen J Anal Chem 364:660–665
  8. Apfalter S, Kandler W, Krska R, Grasserbauer M (2001) Intern J Environ Anal Chem 81:1–14
  9. Analytical Methods Committee (1992) Analyst 117:97–104
  10. Davies PL (1988) Fresenius Z Anal Chem 331:513–519
  11. ISO (1998) Accuracy (trueness and precision) of measurement methods and results—Part 5: Alternative methods for the determination of the precision of a standard measurement method, ISO 5725-5:1998. ISO, Geneva
  12. Analytical Methods Committee (1989) Analyst 114:1489–1495
  13. ISO (2002) Statistical methods for use in proficiency testing by interlaboratory, Draft international standard ISO/DIS 13528. ISO, Geneva
  14. DIN (2003) German standard methods for the examination of water, wastewater and sludge—Part 45: Interlaboratory comparisons for proficiency testing of laboratories (A45), DIN 38402–45. DIN, Berlin
  15. Funk W, Damman W, Donnevert G (1992) Qualitätssicherung in der Analytischen Chemie. VCH, Weinheim
  16. Bundesgesetzblatt für die Republik Österreich (1991) 338. Verordnung: Wassergüte-Erhebungsverordnung—WGEV, Vienna