Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

11-Jul-2014

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

MYRIOPHYLLUM SPICATUM TOXICITY TEST: RESULTS OF AN INTER-LABORATORY RING TEST USING A SEDIMENT-FREE TEST SYSTEM - RING TEST REPORT

Series on Testing and Assessment

No. 205

This document is only available in PDF format.

JT03360393

ENV/JM/MONO(2014)21
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 205

MYRIOPHYLLUM SPICATUM TOXICITY TEST: RESULTS OF AN INTER-LABORATORY RING TEST USING A SEDIMENT-FREE TEST SYSTEM - RING TEST REPORT

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2014

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

A project proposal for the development of a sediment-free *Myriophyllum spicatum* toxicity test was submitted to the OECD Working Group of National Coordinators of the Test Guidelines Programme (WNT) by Germany in 2011, and included in the Test Guidelines work plan in 2012. A ring-test was conducted in 2010-2011. The proposed method is to evaluate the effects of chemicals on a dicotyledonous macrophyte – *Myriophyllum spicatum* – in a single-phase (sediment-free water) test system.

A draft Test Guideline for "Sediment-free *Myriophyllum spicatum* Toxicity Test" and the associated ring test report were circulated twice to the WNT for review in 2013-2014. The draft Test Guideline for a "Sediment-free *Myriophyllum spicatum* Toxicity Test" and ring test report were approved by the 26th WNT meeting in April 2014. The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to the declassification of the ring test report on 7th July, 2014.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

Myriophyllum spicatum toxicity test: Results of an inter-laboratory ring test using a sediment-free test system.

FKZ: 363 01 294; Final report

Organization of an inter-laboratory ring test - Methodological development and optimization of a test method with higher aquatic plants.

Contact person: Dirk Maletzki, Ecotoxicology Laboratory
Federal Environment Agency (UBA)

Contract period of the project: September 2010 to July 2011

Initiator:

Federal Environment Agency;
Division IV: Chemical and Biological Safety

Contractor:

Dr. Monika Ratte, ToxRat Solutions GmbH

Prof. Dr. Toni Ratte, RWTH Aachen University

Contents

Summary

So far, no suitable mono-species standard bioassays using rooted dicotyledonous macrophytes have been available. This is why a standard test system using the dicotyledon, *Myriophyllum spicatum* (Eurasian water milfoil) has been developed by the Ecotoxicology Laboratory of the Federal Environment Agency (UBA), based on ASTM Standard E 1913-04. In a sediment-free single-phase test system and under sterile conditions, the plants are singly exposed over a 14-day period.

During the period from October 2010 to July 2011, the test system was thoroughly examined in the context of an inter-laboratory comparison. Altogether, 30 test runs were performed by 12 laboratories, using three different test substances, namely 3,5-dichlorophenol (3,5-DCP), 2,4-dichlorophenoxy acetic acid (2,4-D) and isoproturon (IP) (10 control replicates, 8 treatments with 5 replicates each). The parameters recorded included the main shoot length, fresh and dry weight, the number of whorls, the number and length of lateral branches and number and length of roots. These parameters served as a basis to calculate the total shoot length, yield parameters and growth rates. The validity criteria for the inter-laboratory comparison included a doubling of the main shoot length of the control plants and at least 50% control replicates remaining without any apparent contamination by foreign organisms.

The high expenditure in terms of time and technology required to conduct the present inter-laboratory comparison is not representative of the use as a standard test system in the future. Rather, the high number of measurement times and variables measured served to generate solid data as a basis for selecting suitable test conditions, endpoints and measurement intervals. In the present evaluation, the general practicability and reproducibility of the test system are investigated. The sensitivity, reproducibility and general suitability of variables with regard to the different test substances are studied and proposals are derived for a further optimization and standardization of the test system.

The validity criteria are complied with in 23 of the 30 test runs. Four tests fail to achieve a doubling of the shoot length and three have shown apparent contamination with foreign organisms. As a result, the test system can be regarded as principally practicable.

The number and length of lateral branches are considered unsuitable for use as single endpoints since lateral branches are produced in only 54 % of control replicates and because their absolute number is too low. The intra-laboratory variability of the other variables is 10 - 30%. Based on the given test design, inhibition rates of 15% - 30% can be detected by means of Williams' test (%MDD, minimal detectable difference). A higher variability and / or lower number of evaluable replicates will increase the %MDD.

Toxicological parameters determined included the EC_{50} with a 95 % confidence interval, and the NOEC. 3,5-DCP, a substance with a non-specific narcotic mode of action, does not result in an extraneous impairment of the test organisms and enables a clear-cut determination of the toxicity parameters, EC_{50} and NOEC, for all variables. 2,4-D, a substance interfering with growth, affects the evaluability of shoot lengths and weight parameters and induces the formation of rudimentary roots in the leaf axils of the entire main shoot at intermediate concentration levels. Isoproturon, a substance inhibiting photosynthesis, causes elongation growth, which in turn impairs the evaluability of the shoot length parameters.

There were no obvious differences between the EC_{50} and NOEC values from tests fulfilling the validity criteria of sterility and from tests apparently contaminated with foreign organisms. On exposure to 2,4-D for some variables, tests with reduced growth of the controls resulted in higher EC_{50} and NOEC values as compared to tests with a doubling of shoot lengths among the controls. However, the available data do not suffice to confirm this finding.

The longer the duration of the test, the lower becomes the EC_{50} on exposure to 3,5-DCP (the comparison referred to EC_{50} values calculated after 7 days, 10 days and 14 days). This is why a test duration of 14 days with measurements on day 0, day 7 and day 14 is recommended. Culturing may be performed at constant temperature conditions (23°C $\pm$ 2°C). Replacement of the culture medium after 7 days can be omitted if it is certain that the test substance will remain sufficiently stable over the test duration (if necessary, this should be checked by preliminary testing).

Yields will, on principle, result in higher inhibition values leading to lower EC_{50} values than the underlying values from the original measurements. Therefore yields should be preferred for EC determination. The NOEC values are not affected by this phenomenon. On principle, there is a good correlation between EC_{50} and NOEC for the yields. The same is true for comparable variables such as the number and total length of roots.

The most sensitive variables regarding exposure to all three test substances include total root length, yield fresh weight (if evaluable), yield whorl number and number of roots. The yield and growth rate of shoot lengths occupy an intermediate to non-sensitive position. The most sensitive variables also show the best reproducibility for EC_{50} and NOEC.
2,4-D was found to produce the most outstanding toxic effect. It is followed by isoproturon ranking in the second position and 3,5-DCP, in the third. The more sensitive the variable, the greater is the difference in toxicity between the three test substances.

Parameters recommended for measurement and/or evaluation in the context of standard testing include: *Total root length* (also requires recording of the number of roots); *yield fresh weight*, *yield dry weight* (optional, since depending on the test substance, not being generally correlated with fresh weight; where applicable, the more sensitive parameter should be chosen); *growth rate of main shoot length, yield of main shoot length* (as a basis for the determination of the growth rate, optionally also as an independent endpoint); *yield of whorl number* (optional, if shoot length parameter is difficult to evaluate).

The validity criteria applied in the present inter-laboratory comparison have proved to be successful and should be maintained. 3,5-DCP is recommended for use as a reference substance. To determine the length, both digital evaluation based on photographs and manual determination by means of a ruler may be used. However, within one and the same test, a uniform method should be followed. Photographic archiving is preferred since it will enable a better documentation and evaluation of additional parameters at a later time.

Based on the present results and the experience gained by the test participants, proposals for optimization are developed, aiming at establishing the test system as a standard procedure: The SOP should be completed by additional practical information to ensure sterile conditions. The question of (visual) assessment of sterility should be explained in more detail. To this aim, further studies should be performed to arrive at a better understanding of the test system. The SOP should also reflect the evaluation and consequences of possible morphological changes in the test organism induced by the action of substances (rudimentary roots, elongation growth, deformation). It appears that a furthermore improved standardization by means of the most exact instructions possible with regard to handling may be achieved (handling of test organisms, determination of fresh weight), resulting in a further minimization of inter-laboratory variability.

The results of the inter-laboratory comparison have shown the general practicability, reproducibility and suitability of the sediment-free test system for *Myriophyllum spicatum*, enabling the determination of meaningful framework conditions for its use as a standard test system in the future.

1. Introduction

In the context of the enforcement of plant protection legislation, the performance of bioassays using Lemnaceae is envisaged for herbicide approval. It has been doubted to what extent the pesticide sensitivity spectrum of Lemnaceae, being monocotyledons, can be extrapolated to that of dicotyledons.
This is why a test with the dicotyledon, *Myriophyllum spicatum* has been developed by the Federal Environment Agency (UBA), which is based on ASTM Standard E 1913-04. The development of the test was aimed at ensuring the most unequivocal exposure possible of the test organism to the substances to be tested Therefore, the core properties of the new test with *Myriophyllum spicatum* include: Single, sediment-free culture of the test organism, and sterile conditions.

The details of the testing procedure are described in the following documents developed by UBA which are enclosed with the present report as annexes: Standard Operating Procedures, SOPs; 1) SOP Axenic and Sediment-free Stock Culture of *Myriophyllum spicatum*, Umweltbundesamt, 9.07.2009; 2) SOP Sediment-free *Myriophyllum spicatum* Toxicity Testing, Umweltbundesamt, 13.07.2009

In the period from October 2010 to July 2011, an inter-laboratory comparison was organized by UBA in order to thoroughly examine the test system. The objectives of the inter-laboratory comparison included

- Examination of the practicability and reproducibility of the test system;
- Identification of suitable endpoints (also with regard to different mechanisms of action of the substances to be tested); and
- Optimization and standardization of the test system.

UBA canvassed 12 laboratories to participate in the inter-laboratory comparison. All of them where invited to attend an initial workshop in October 2010. The participants where comprehensively informed about the details of the test procedure and the practice of handling plants. They were supplied with pre-cultured *Myriophyllum spicatum* organisms by the UBA Ecotoxicology Laboratory enabling them to establish in their laboratories a stock culture and then pre-cultures for the respective tests according to the above-mentioned SOPs.
The practical part of the inter-laboratory comparison was held from November 2010 until April 2011. Three substances were tested: 3,5-dichlorophenol (3,5-DCP), 2,4- dichlorophenoxy acetic acid (2,4-D) and isoproturon (IP).
The participants submitted their results to the company, ToxRat Solutions GmbH where a uniform evaluation of data was performed. Toxicity parameters determined included the EC_{50} and EC_{20} with 95 % confidence interval, as well as the NOEC and LOEC.

The present report summarizes the evaluation results and the results and experiences discussed at the final workshop in June 2011. It also gives recommendations for a further optimization of the test, aiming at establishing the sediment-free test system with *Myriophyllum spicatum* as a standard test system in the future.

2. Framework conditions of the inter-laboratory comparison

2.1. Timetable

24 Sep. 2010	Starting talks
18-19 Oct. 2010	Starting workshop
29 Nov. 2010	Presentation of progress report by the expert team
1 Oct. 2010 – 31 Apr. 2011	Conduction of tests in the participating laboratories
Submission of raw data to the expert team	
21 June 2011	Final workshop
15 July 2011	Presentation of the final report by the expert team

2.2. Participants

Altogether, 12 laboratories participated in the inter-laboratory comparison (see Table 3). Except the Ecotoxicology Laboratory of the Federal Environment Agency, none of the participating laboratories had previous experience with the test system.

Tab. 1 Participating laboratories and contact persons involved. Alphabetical order, i.e. not identical with the order of the internal laboratory code numbering

Participating laboratories	Contact persons	E-Mail
BASF SE	Johanna Kubitza	johanna.kubitza@basf.com
BayerCropScience AG	Eric Bruns Daniela Schrader	eric.bruns@bayer.com daniela.schrader@bayer.com
BioChem agrar	Daniela Juckeland Moritz Meiselbach	daniela.juckeland@biochemagrar.de
Dr. Noack Laboratorien	Lisa Bach Lidia Wincek	info@noack-lab.de lidia.wincek@noack-lab.de
Eurofins Agroscience Services	Guido Gonsior	guido.gonsior@eurofins-gab.com
Harlan Laboratories	Helene Eckenstein Anja Liedtke	heckenstein@harlan.com aliedtke@harlan.com
Helmholtz-Zentrum für Umweltforschung GmbH – UFZ Leipzig	Mechthild Schmitt-Jansen, Silke Aulhorn	Mechthild.Schmitt@ufz.de silke.aulhorn@ufz.de
IBACON	Katrin Hoffmann Igor Eckert	katrin.hoffmann@ibacon.com igor.eckert@ibacon.com
IME Schmallenberg	Andrea Wenzel Pamela Schulte	andrea.wenzel@ime.fraunhofer.de pamela.schulte@ime.fraunhofer.de
Mesocosm GmbH	Klaus-Peter Ebke Laszlo Dören	ebke@mesocosm.com doeren@mesocosm.de
Ökotoxikologielabor UBA Berlin	Carola Kussatz Dirk Maletzki Ines Hübner	carola.kussatz@uba.de dirk.maletzki@uba.de ines.huebner@uba.de
SGS Institut Fresenius	Herbert Lebertz Stefanie Stremlau	herbert.lebertz@sgs.com stefanie.stremlau@sgs.com

2.3. Test organism and conditions of testing

Pre-cultured *Myriophyllum spicatum* organisms were supplied to the participants by the Ecotoxicology Laboratory of the Federal Environment Agency. These served as a basis to establish stock cultures and pre-cultures for the respective tests according to the UBA SOP (see Annexes).

The test design and SOP envisage a 14-day test duration and a photoperiod of LD 16:8 (100 - 150 µE/m^2s or 6000-9000 Lux). Culture media are replaced after seven days. As suggested by the participants of the inter-laboratory study, and by derogation from the SOP data, the tests were conducted at a constant temperature of 23°C $\pm$ 2°C, instead of a circadian temperature regimen. The resulting advantages include a reduction of the expenditure on technology required for performing the tests and the possibility to conduct *Myriophyllum* testing in parallel to algae or *Lemna* testing. A test with a diurnal temperature regimen conducted simultaneously by UBA served to investigate possible differences among the test results.

For a survey of the variables to be determined or calculated, see Table 2. If possible, measurements were performed on days 0, 3, 7, 10 and 14 of testing.

Tab. 2 Variables measured and calculated (and abbreviations used)

Measured variables	**Abbreviations**	**Measurement time**
Fresh Weight	FW	0d*, 14d
Dry Weight	DW	0d*, 14d
Shoot Length	SL	on days 0, 3, 7, 10 and 14 of testing
Lateral Branches	LB	14d
Lateral Branches Length	LBL	14d
Whorls	W	on days 0, 3, 7, 10 and 14 of testing
Roots	R	14d
Root Length	RL	14d
Calculated variables	***I***	***I***
Yield Fresh Weight	YFW	14d
Yield Dry Weight	YDW	14d
Yield Shoot Length	YSL	on days 0, 3, 7, 10 and 14 of testing
Total Shoot Length	TSL	14d
Yield Total Shoot Length	YTSL	14d
Yield Whorls	YW	on days 0, 3, 7, 10 and 14 of testing
Total Root Length	TRL	14d
Growth Rate Shoot Length	GrSL	on days 0, 3, 7, 10 and 14 of testing
Growth Rate Total Shoot Length	GrTSL	on days 0, 3, 7, 10 and 14 of testing

* representative plants used for measurement

2.4. Test substances and test design

Criteria for selection of the three different test substances included:

- Different mechanisms of action;
- Availability of comparative data and experience; and
- Stability of the candidate substance for at least 7 days.

The substances selected were:

1. 3,5-dichlorophenol (3,5-DCP), reference substance OECD 221 and 201, pesticide, non-specific mechanism of action
2. 2,4-dichlorophenoxy acetic acid (2,4-D), herbicide active against dicotyledons, auxin, growth inhibitor
3. Isoproturon (IP); herbicide, photosynthesis inhibitor

Testing of 3,5-DCP was mandatory for all participants. If a second substance was tested, 2,4-D was used, and a possible third test was to use IP.

UBA ordered one batch each of the test substances, which were then distributed among the participating laboratories. A concomitant analytical determination was preferred, it was, however, performed on a voluntary basis. Due to the stability of test substances, the results could be expressed for nominal concentrations.

The test design is shown in Table 3. Test runs included one control each and eight treatments, in ascending concentrations. In the control, 10 replicates were used, and per treatment, five replicates.

Tab. 3: Survey of the test design: Test substances and concentrations, number of replicates

Test substances		**Number of replicates**
3,5-DCP [mg/L]	2,4-D + IP [µg/L]	
control	control	10
0.4	1.2	5
0.8	3.2	5
1.6	9,6	5
3.2	40	5
6.4	120	5
9.6	320	5
12.8	960	5
19.2	2000	5

2.5. Validity criteria

The tests conducted in the context of the inter-laboratory comparison were rated as valid if the following conditions were fulfilled:

- At least doubling of the main shoot length in the controls by day 14, *and*
- At least 50 % of control replicates being sterile on day 14 (or “apparently free from foreign organisms”, see section 4.2.1).

The question whether a replicate had remained sterile was assessed by the participants themselves on a purely visual basis and classified accordingly in the raw data. The assessment of samples as “non-sterile”, or “apparently contaminated with foreign organisms” referred to test samples if:

- The test medium appeared milky and turbid;
- Filamentous, cotton wool-like clusters were seen;
- Whitish coating on the plants was seen with *simultaneous* turbidity of the medium.

For details on the sterility assessment, please refer to Section 4.2.1.

3. Procedure for evaluation

3.1. Data considered for evaluation

The raw data sets were edited by removing obvious typing errors or other mistakes made during data input (e.g. length in cm instead of mm, decimal separator omitted etc.).

Subsequently, all data sets were subjected to the Hampel test for outliers. However, no values suspected of being outliers were excluded from evaluation (for details refer to Section 3.3 Treatment of statistical outliers).

Tests not complying with the "sterility controls" validity criterion [1]where nevertheless evaluated, including the data from replicates considered as non-sterile both from controls and treated samples. However, the data and results from these tests were not taken into account for further evaluation. They were only included in the presentations and tables to be compared with the results from tests assessed as sterile (marked accordingly).
Also test results not complying with the "shoot length doubling" validity criterion were included in the evaluation (after deletion of possible data from replicates apparently contaminated with foreign organisms). However, these results were not taken into account for further evaluation and used in presentations as described above.

These results used only for presentation purposes have not been included in the summarizing calculation of mean values and dispersion.

In all other data sets, the data from control and treatment samples apparently contaminated with foreign organisms were deleted prior to statistical evaluation.

3.2. Statistical evaluation

The data describing the test results were statistically evaluated by means of the ToxRat Professional ® software. To this aim, version 2.10.05 was expanded to include a routine enabling evaluation of the "*Myriophyllum*" workbook specifically developed for the inter-laboratory comparison. This routine consists of an EXCEL file containing, on 10 different worksheets, the raw data for all test parameters measured, as well as a classification indicating which of the replicates were classified as sterile/non-sterile.
The participants of the inter-laboratory comparison received this EXCEL file together with the raw data protocol to be kept in writing as a template for raw data input.

All variables measured and calculated (see 2.3., Variables measured) were evaluated as follows:
The EC_{20} and EC_{50} were determined by means of adjustment of the sigmoid normal distribution function (probit analysis for metric data). As a rule, the fits are based on mean values unless a significant fit was achieved only on the basis of the single data from replicates (marked accordingly in the result tables).

Normal distribution was checked by means of Shapiro-Wilk's test, and variance homogeneity was assessed by means of Levene's test.

[1] Cf. Section 4.2.1 Sterility as a validity criterion

As a rule (given a normal distribution and variance homogeneity), the NOEC and LOEC were determined by Williams' test or, given a normal distribution but no variance homogeneity, by Welch's t-test with Bonferroni correction. In cases where neither variance homogeneity nor normal distribution were found, Mann-Withney- U-test with Bonferroni-Holm correction was used (marked accordingly in the result tables).

Evaluation of controls

In order to compare the laboratories with one another and to assess reproducibility of the results, the arithmetic mean with its 95 % confidence interval and the standard deviation *s* of the control replicates of the respective laboratory was calculated for each variable *per laboratory and per test,* and based on these, the coefficient of variation, CV_r%, was determined as a measure for *intra-laboratory variability* (CV_r% "repeatability") according to

$$CV_r\ \% = (s_L / \bar{x}_L)*100$$

where

CV_r% = Coefficient of variation intra-laboratory variability
$\bar{x}_L$ = Laboratory-specific mean value
s_L = Standard deviation of laboratory-specific mean value

Based on the individual laboratory-specific coefficients of variation, CV_r%, the mean coefficient of variation, CV_r% per variable was determined.
In the present inter-laboratory comparison, the latter is based on 23 valid data sets (see Section 4.1 Data used).

To determine the *inter-laboratory variability* (CV_R% ("reproducibility"), the arithmetic mean and standard deviation *s of all* control replicates (i.e. of all laboratories) were calculated, and based on these, the coefficient of variation, CV_R%, was determined according to

$$CV_R\ \% = (s_{all} / \bar{x}_{all})*100$$

where

CV_R% = Coefficient of variation for inter-laboratory variability
$\bar{x}_{all}$ = Mean value of all valid control replicates
s_{all} = Standard deviation of all valid control replicates

In the present inter-laboratory comparison, the latter is based on 218 control replicates each (see Section 4.1. Data used).

Based on the mean values of all valid tests, the overall mean value and the overall standard deviation *s* of the control mean values were calculated for each variable as well as the 95 % prediction interval (PI) of this overall mean value according to

95% PI = Overall mean value $\pm$1,96 *s*

Evaluation of toxicity parameters

In order to compare the results of the different laboratories with one another and thus, assess the reproducibility of the results, and to compare the toxicity of the three test substances, the toxicity parameters EC_{50} including 95% confidence interval as well as the NOEC / LOEC were determined for each substance tested and for *each laboratory and variable.*

From the toxicity parameters of the valid tests, the respective geometric mean with its 95% prediction interval (PI) was calculated.

Geometric mean:	Taking the logarithm of EC_{20} / EC_{50}/ NOEC, calculating the mean, taking the antilogarithm;
95% prediction interval:	Calculating the geometric mean and standard deviation *s*, calculating 95% PI on log scale as geometric mean $\pm$1,96*s*, taking the antilogarithm for 95% PI.

The 95% prediction interval of the EC_{50} and the NOEC for the different variables serves as a *measure of reproducibility* of the respective toxicity parameters. The narrower the 95% prediction interval, the higher is the reproducibility of the toxicity parameters.

3.3. Treatment of statistical outliers

All data sets were tested for statistical outliers with the aim to identify possible errors in measurement and eliminate these prior to the actual evaluation. The Hampel outlier test was used, which does not require normal distribution and is considered as robust to outliers.

The test identified a number of suspected outlier values in almost every data set (both in controls and treatments). Due to this unusual accumulation of potential “outliers”, the data were subjected to a more detailed analysis:
In fact, a relatively high variability including a number of extreme values was shown e.g. by all weight variables. On the one hand, it is difficult to methodologically standardize the determination of fresh weight and on the other, factors such as the number and length of lateral branches and the number and length of roots are implicitly included in the weight parameters. Therefore, it may be assumed that at least the weight values suspected of being outliers are by no means due to errors in measurement but rather, represent solid, albeit extreme, measured values.
However, it would mean to act arbitrarily if suspected outlier values were excluded for some variables and for others not. In addition, excluding extreme values would mean to artificially minimize the natural variability of the test system, being one of the subjects of the present evaluation.
This is why on principle, all (valid) data were included in the present evaluation.

The resulting control mean values of the individual variables were subsequently also subjected to an outlier test (see Annex, Tables 1 and 2).

During this procedure, only two control mean values were identified as suspected outliers:

- L09 2,4-D LB max (number of lateral branches) in test of L09 with 2,4-D
- L03 IP TRL max (total root length) in test of L03 with IP

Below, the data sets affected are subjected to a detailed analysis.

L09 2,4-D: In the respective control, *all* replicates developed an *unusually large number* of lateral branches. Hence, the high mean value of 3.6 is not to be attributed to a single extreme value. As a consequence, the data on lateral branches could be excluded from evaluation as a whole because they are not within the 95% prediction interval. However, this would not appear to make sense because the data set shows consistent measured values and no abnormal results with regard to other variables. Unless it is possible to identify a concrete factor triggering the development of unusually large numbers of lateral branches (and deviating from the defined culture conditions), the respective sample, albeit rare, has to be considered as a sample positively belonging to the same sample population.

L03 IP: Here, the explanations given with regard to the lateral branches apply analogously. *All* replicates of the control uniformly show a pronounced root growth; the mean value for the number of roots is very high (6.0), leading to a high value for total root length. This is why there is no reason to assume an occurrence of single outliers. Rather, albeit rare, the sample has to be considered as positively belonging to the same sample population.

Based on the above facts, there is no reason for excluding the data sets mentioned from the determination of toxicity parameters. Both the control variabilities determined and the values obtained for reproducibility and dispersion of EC and NOEC values may be considered to represent upper limits.

4. Results

4.1. Data used

A total number of 30 tests were conducted: Seven participants tested all three test substances, four participants tested two test substances each, and one laboratory performed one test with one test substance.
Five participants could obtain three valid data sets each, two participants conducted two valid tests each, four participants produced one valid test each, and in one laboratory, none of the data sets produced complied with both validity criteria. Table 4 provides an overview of the data obtained during the inter-laboratory comparison.

Altogether, the results of 12 test runs with 3,5-DCP, 10 test runs with 2,4-D and 8 test runs with IP are available. Of the 30 test runs conducted, four fail to comply with the validity criterion of shoot length doubling. Three other runs had to be classified as invalid due to sterility problems.

Consequently, the onward evaluation could refer to eleven valid data sets for 3,5-DCP and six valid data sets each for 2,4-D and IP. In the present report, the results from the invalid tests have also been listed in the respective graphic representations and tables for comparative purposes (and marked accordingly).

Tab. 4 Data obtained from the inter-laboratory comparison

	3,5-DCP	**2,4-D**	**IP**	**Sum of all tests**
Number of tests performed	12	10	8	30
Number of valid tests	11	6	6	23
Number of not valid tests	1 x sterility problems	1 x sterility problems 3 x shoot length doubling	1 x sterility problems 1 x shoot length doubling	7

4.2. Practicability

4.2.1 Sterility as a validity criterion

A special characteristic of the sediment-free test system using *Myriophyllum spicatum* consists in the sterile conditions. For assessing the general practicability of the test, it is therefore of particular interest to what extent this validity criterion could be met.
Whether a test sample had remained sterile was assessed visually, by optical criteria. Since no microbiological analysis was carried out, the assessment of replicates during the test will be described below as “apparently free from foreign organisms” and “apparently contaminated with foreign organisms” instead of “sterile” and “non-sterile”.
The inter-laboratory comparison has demonstrated that for assessing the sterility factor, a distinction has to be made between controls and treated samples since the type and concentration of the test substance may have an influence on the visual appearance of the plant.
The observations made in the controls were as follows:

1. Plant appearing extraneously unimpaired, without any foreign bodies in clear medium (Fig. 1);
2. Plant entirely or partly wrapped in whitish, filamentous clusters (Fig. 2);
3. Plant with whitish aufwuchs, medium turbid (Fig. 3);
4. Plant with red-brown particles, medium clear (Fig. 4);

Fig. 1
Control replicate on day 14 L10 35DCP d14 ctrl repl 3
No foreign bodies, medium clear

Fig. 2
Control replicate on day 14 L06 35DCP d14 ctrl repl 6
Filamentous clusters

Fig. 3
Control replicate on day 14 L06 35DCP d14 ctrl repl 1
Whitish aufwuchs, medium turbid

Fig. 4
Control replicate on day 14 L02 35DCP d14 ctrl repl 9
Red-brown particles, medium clear

Cases 1, 2 and 3 can be classified without doubt: The control replicate in Fig. 1 is apparently free from foreign organisms, the foreign bodies observed in Figs. 2 and 3 are presumably fungal hyphae or bacteria. These replicates are apparently contaminated with foreign organisms.

In the case of Fig. 4, however, the nature of the red-brown particles has not yet been elucidated. Possible explanations would include contamination with foreign organisms (of unknown identity), plant secretions or something completely different. In the present inter-laboratory comparison, the brown particles occurred in three tests (L02 2,4-D and IP as well as L09 3,5-DCP), with the medium remaining clear in all cases. For evaluation, the replicates affected were classified as "apparently free from foreign organisms", i.e. as valid. For comparison of the controls from all valid tests, particular attention will be paid to any type of special characteristics to be observed in the corresponding results.

In the treated samples, another phenomenon occurred in addition to the observations already mentioned above: In many tests with 2.4-D and IP, whitish coating was seen on the plants while the medium always remained clear (Figs. 5, 6 and 7).

Fig. 5
Test with 2,4-D replicate with 2000 µg/L, d14 L02 2,4-D img 46

Fig. 6
Test with 2,4-D replicate with 2000 µg/L, d14 L02 2,4-D img 49

Fig. 7
Test with IP replicate with 2000 µg/L, d14 L10 IP img 50

Fig. 8
Test with 3,5-DCP replicate 19.6 mg/L, d14 L10 3,5-DCP img 50

Arguments supporting the assumption that the phenomena observed were indications of substance-related decomposition of the plants rather than contamination include:

a. The effect occurred only on exposure to 2,4-D and to IP (where it was less pronounced) (Figs. 5, 6 and 7), however, not on exposure to 3,5-DCP (Fig. 8).
b. The effect was observed only at high concentrations; and
c. The effect was also observed in tests whose controls did not show any special visual characteristics (in the case of contamination caused by handling etc., changes were expected to spread over all replicates).

For the present evaluation, the replicates affected were classified as "apparently free from foreign organisms", i.e. as valid.

Fig. 9 demonstrates that in 25 out of 30 tests, the sterility criterion was met in at least 9 out of 10 control replicates. In two more tests, 6 out of 10 control replicates were apparently free from foreign organisms resulting in the tests to comply with the validity criterion. It becomes obvious that contamination did not occur widely dispersed but, as a rule, either not at all or to a rather massive extent. Altogether, 259 out of 300 control replicates remained apparently free from foreign organisms, corresponding to a rate of 86 %.

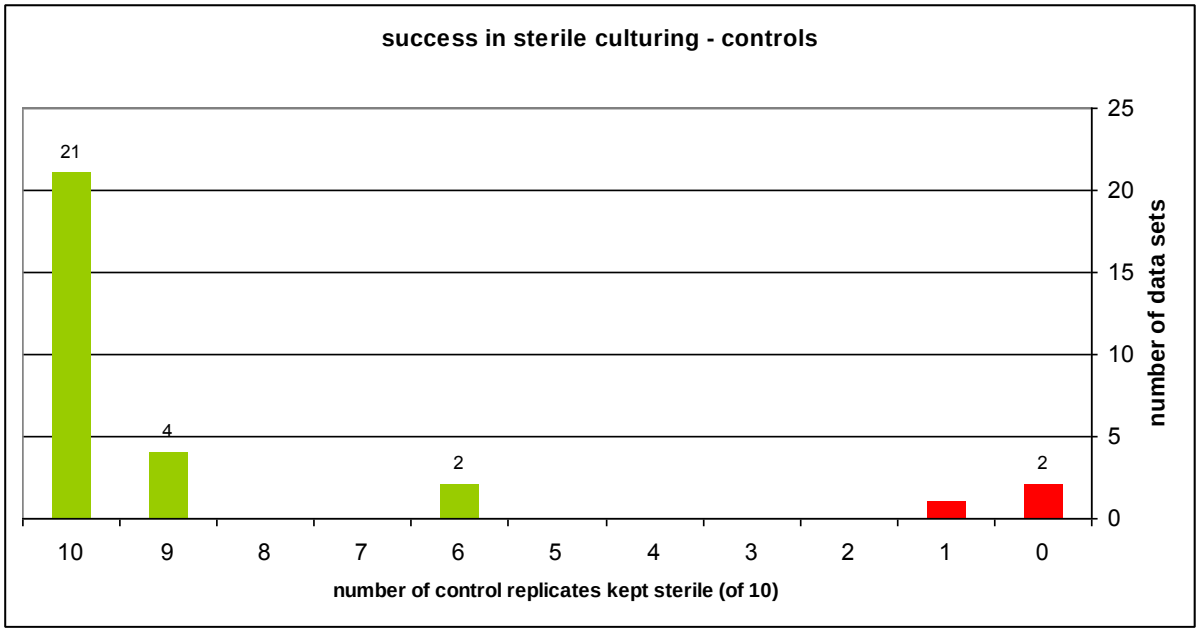


Fig. 9: Number of control replicates apparently not contaminated with foreign organisms in the tests conducted. Example of how to read the graphic representation: In 21 (out of 30) test runs conducted, all control replicates (10 out of 10) were apparently free from foreign organisms.

4.2.2 Shoot length doubling as a validity criterion

In the present inter-laboratory comparison, the shoot length yield factors measured varied between 1.8 and 3.8 (see Annex, Table 3). 26 out of 30 test runs complied with the minimum requirement of shoot length doubling, corresponding to a rate of 86%. In four test runs, the growth of controls showing shoot length yield factors of 1.8 to 1.96 was insufficient to meet this validity criterion. Of the valid test results, the shoot length yield factor was between 2.0 and 2.9 (mean value 2.43, s = 0.28) in 21 cases. In five cases, this factor was between 3 and 3.8 in the controls.

To find out the cause of reduced growth in a number of test runs, an analysis of the light and temperature conditions was carried out. The temperature recording revealed that the envisaged temperature regimen of 23 °C $\pm$ 2°C was complied with in all tests.
In contrast, considerable differences were found regarding the light intensities prevailing in the different tests. The rated light intensities were between 6000 and 9000 lux or 100-150 µE/m^{2}s. The actual light intensities measured in the tests were between 3000-4000 lux (minimum) and 8200-8400 lux (maximum), or 80 -100 µE/m^{2}s (minimum) and 127-150 µE/m^{2}s (maximum), i.e. the prevailing mean light intensities were between 47% and 110% of the target value.

This could suggest to assume the factor of light intensity to have been causative of reduced growth leading to invalidity in four tests, the more so since due to a technical failure, one of the invalid tests had to be moved from a climatic chamber to a shelf (where light intensity was no longer measured). In one case, randomization of replicates had been omitted on a scheduled date, and in another, the light intensity was only 72% of the target value.

However, no direct correlation could be found to exist between the light intensity and the shoot length yield factor: Yield factors between 2.2 and 2.9 were also achieved in four tests with only 50-60% of the target light intensity while the invalid tests also included one each with 100% and 110%, respectively, of the target light intensity.

In further analyses, the range of light sources used could be included in the evaluation and possible shading effects by lids, labels etc. on the test vessels could be examined.
Another potential influencing factor that may be considered is the mean starting length of the plants used: In the present inter-laboratory comparison, the starting lengths showed very little dispersion on the intra-laboratory level, however, with very different laboratory-specific starting values varying between a minimum of 19 – 22 mm and a maximum of 29 - 33 mm. Among these plants, those with higher starting values tended to show lower growth factors (for comparable light intensity values).

Due to the great number of influencing factors, it is impossible to clearly identify a cause for the reduced growth observed in some of the test runs conducted.

4.2.3. Assessment of practicability

11 of the 12 participants had no experience with the test system. Nevertheless, 86% of the control replicates remained apparently free from foreign organisms, and at least shoot length doubling was achieved in the controls in 26 out of 30 test runs.

Although it has been assumed that a number of test runs were stopped and repeated by some participants due to apparent contamination with foreign organisms or foreseeable poor growth, the test system can be classified as generally practicable, given the existing data and experience.

4.3. Reproducibility of controls

Initially, in an exemplary way, the mean values of the control replicates for selected variables (main shoot length SL, yield fresh weight YFW, number of roots, total root length TRL, and growth rate of shoot length GrSL), are compared both on the intra-laboratory and inter-laboratory levels (Figs. 10 to 14). [2]

Subsequently, the reproducibility of all variables is quantified by means of the coefficients of variation (Figs.15 and 16). The intra-laboratory variability determines the detectable magnitude of effect and thus, constitutes a measure for the statistical power of the NOEC that can be derived. The relationship between inter-laboratory and intra-laboratory variability is a measure of the current degree of standardization of the test system.

Finally, the count variables, number of root and number of lateral branches are characterized in detail (Figs. 17 and 18).

On this basis it is examined which of the variables are generally suitable for a further evaluation of toxicity parameters in the context of the present inter-laboratory comparison.

4.3.1. Comparison of mean values and prediction intervals

Depending on the laboratory and the variable considered, the resulting intra-laboratory reproducibility will be higher or lower. On principle, the values are always well reproducible. While the values for main shoot length, root number and growth rate are very close to each other (Figs. 10, 12, 14), variables for parameters such as yield fresh weight and total root length are scattered to a larger extent (Figs.11 and 13). This may have both intrinsic reasons (root lengths possibly more variable, on principle, than other endpoints, or root length possibly highly sensitive to different growth conditions) and methodological causes (determination of fresh weight difficult to standardize, influence of number and length of lateral shoots and roots on fresh weight).

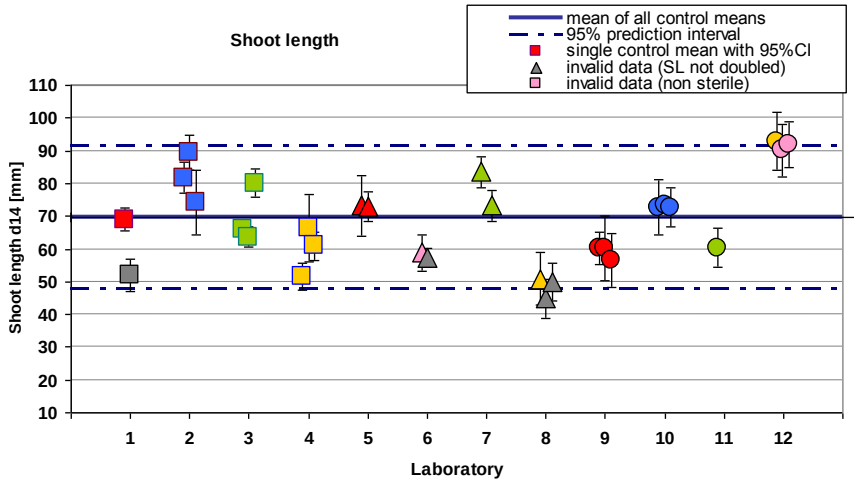


Fig. 10 Main shoot length of controls in all tests performed: Mean value with 95% confidence interval per laboratory (coloured symbols) and the overall mean value from valid tests with a 95% prediction interval (solid and broken lines); invalid tests included for comparison (marked in colour, grey: shoot length not doubled; pink: apparent contamination with foreign organisms).

[2] The corresponding Figures and data for *all* variables are found in the file "summary valid controls_fig_means_all.xls".

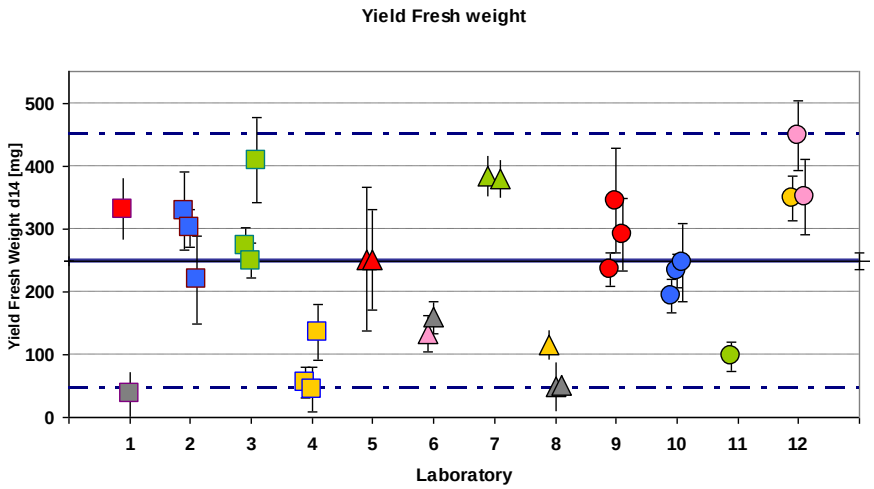


Fig. 11 Yield fresh weight of controls in all tests performed: Mean value with a 95% confidence interval per laboratory (coloured symbols) and overall mean value from valid tests with a 95% prediction interval (solid and broken lines); invalid tests included for comparison (marked in colour, grey: shoot length not doubled; pink: apparent contamination with foreign organisms).

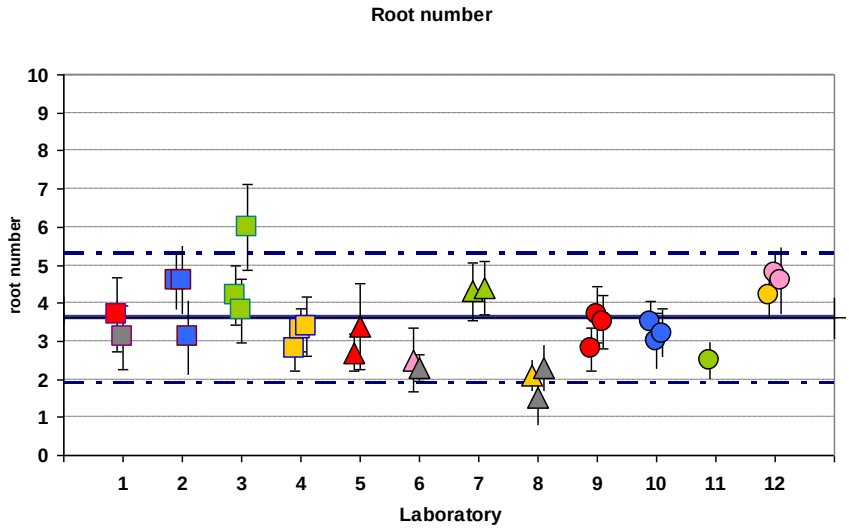


Fig. 12 Number of roots in controls in all tests performed. Mean value with a 95% confidence interval per laboratory (coloured symbols) and overall mean value from valid tests with a 95% prediction interval (solid and broken lines); invalid tests included for comparison (marked in colour, grey: shoot length not doubled; pink: apparent contamination with foreign organisms).

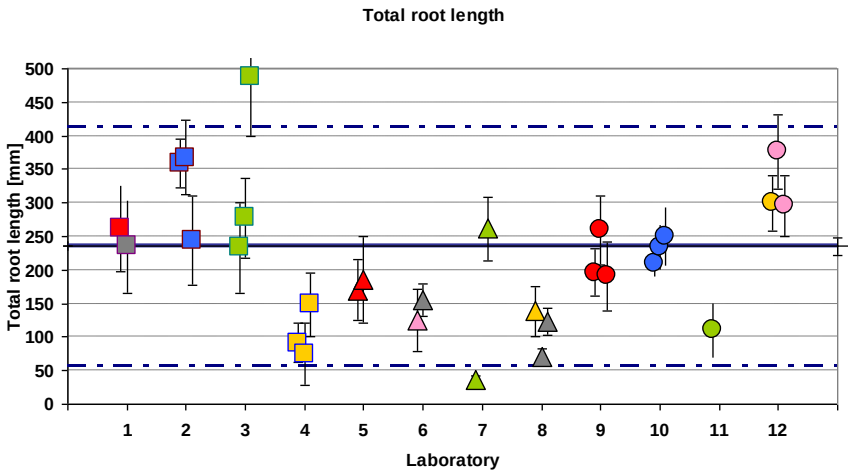


Fig. 13 Total root length of controls in all tests performed. Mean value with a 95% confidence interval per laboratory (coloured symbols) and overall mean value from valid tests with a 95% prediction interval (solid and broken lines); invalid tests included for comparison (marked in colour, grey: shoot length not doubled; pink: apparent contamination with foreign organisms).

On principle, intra-laboratory variability is lower than inter-laboratory variability. This becomes particularly evident for the yield fresh weight (Fig. 11) and the total root length (Fig. 13), both showing a high inter-laboratory variability and correspondingly wide prediction intervals for the mean value. This points to a potential for further standardization of the method.

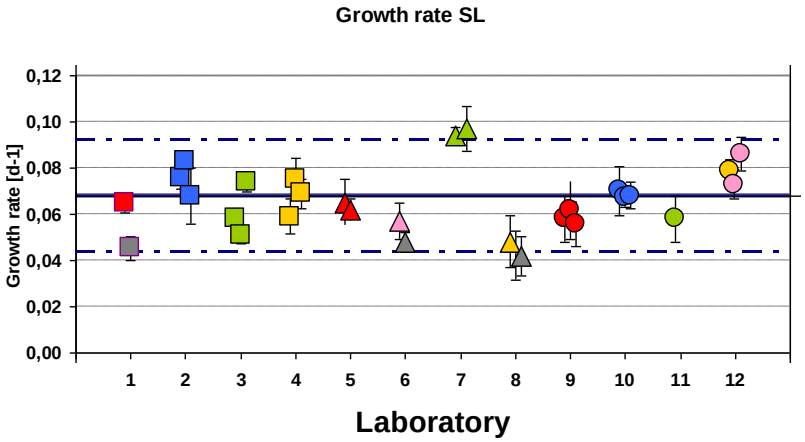


Fig. 14 Growth rate of the main shoot length of controls in all tests performed. Mean value with a 95% confidence interval per laboratory (coloured symbols) and overall mean value from valid tests with a 95% prediction interval (solid and broken lines); invalid tests included for comparison (marked in colour, grey: shoot length not doubled; pink: apparent contamination with foreign organisms).

Altogether, the results of all participating laboratories are, as a rule, within the 95 % prediction interval. Hence, they may be classified as representative of the test system (a certain number of deviant values has to be statistically expected since the prediction interval comprises only 95% of the mean values).

The controls “apparently contaminated with foreign organisms“ do not show any significant findings concerning the mean values for the controls; the results obtained are comparable with those for the controls from tests not contaminated with foreign organisms.
The control replicates that failed to achieve shoot length doubling naturally show lower values for all growth parameters. However, a major part of these are still within the 95% prediction interval for the respective mean value.

The test runs which were characterized by a formation of red-brown particles of unknown origin but nevertheless rated as “apparently free from foreign organisms“ and therefore assessed as valid (L02 2,4-D and IP as well as L09 3,5-DCP, cf. Section 4.2.1) likewise do not show any special features concerning the mean values of controls.

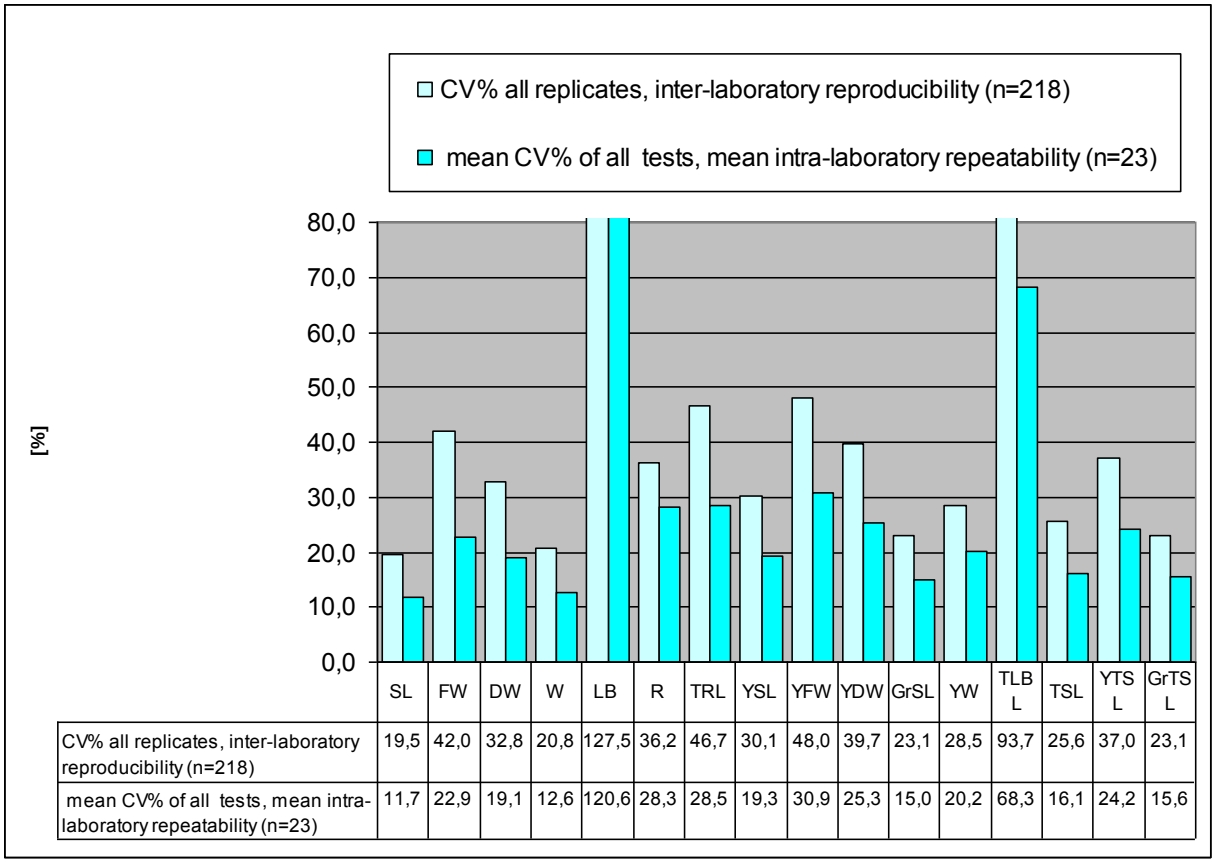
4.3.2. Coefficients of variation and effects of detectable magnitude

Fig. 15 depicts the mean intra-laboratory and the inter-laboratory coefficients of variation for all parameters. First of all, it becomes evident that the values for the variables, number and length of lateral branches are extremely variable both on the inter-laboratory and intra-laboratory levels. Given coefficients of variation ranging from 68% to more than 100%, a determination of toxicity parameters (such as EC and NOEC) with sufficient statistical power is impossible. The causes of the high variability are discussed in detail in Section 4.3.3.

Irrespective of the absolute values, inter-laboratory variability is always higher than intra-laboratory variability. In other words, reproducibility is worse than repeatability. This has demonstrated that the standardization of test conditions and handling can be furthermore enhanced. However, for the yields of fresh and dry weight, a higher variability cannot be avoided for methodological reasons because no individual starting weights can be determined but instead, the yields calculated are based on mean values for representative plants.

The best relationship between inter-laboratory and intra-laboratory variability is found for the variables, root number and yield of lateral branches (factors of 1.3 and 1.4, respectively). On the one hand, these variables can be determined more or less without errors of measurement, on the other, it appears that they are relatively robust to stand slightly varying test conditions.
For the other parameters, the relationship between inter-laboratory and intra-laboratory variability is characterized by factors between 1.5 and 1.8[3].

[3] Cf. the file “summary valid controls_fig_CV.xls”.



	SL	FW	DW	W	LB	R	TRL	YSL	YFW	YDW	GrSL	YW	TLBL	TSL	YTSL	GrTSL
CV% all replicates, inter-laboratory reproducibility (n=218)	19,5	42,0	32,8	20,8	127,5	36,2	46,7	30,1	48,0	39,7	23,1	28,5	93,7	25,6	37,0	23,1
mean CV% of all tests, mean intra-laboratory repeatability (n=23)	11,7	22,9	19,1	12,6	120,6	28,3	28,5	19,3	30,9	25,3	15,0	20,2	68,3	16,1	24,2	15,6

Fig. 15 Inter-laboratory and mean intra-laboratory coefficients of variation for all variables; based on valid control replicates

Fig. 16 lists the mean intra-laboratory coefficients of variation for all variables in an ascending order. Except for the number and lengths of lateral branches, all coefficients of variation are between 10 and 30 %. The lowest dispersion (11% to 15%) was observed for the variables, shoot length, number of whorls and growth rates, while the highest dispersion (20% - 30%) occurred for fresh weight, yield fresh and dry weight, yield total shoot length, root number and total root length.

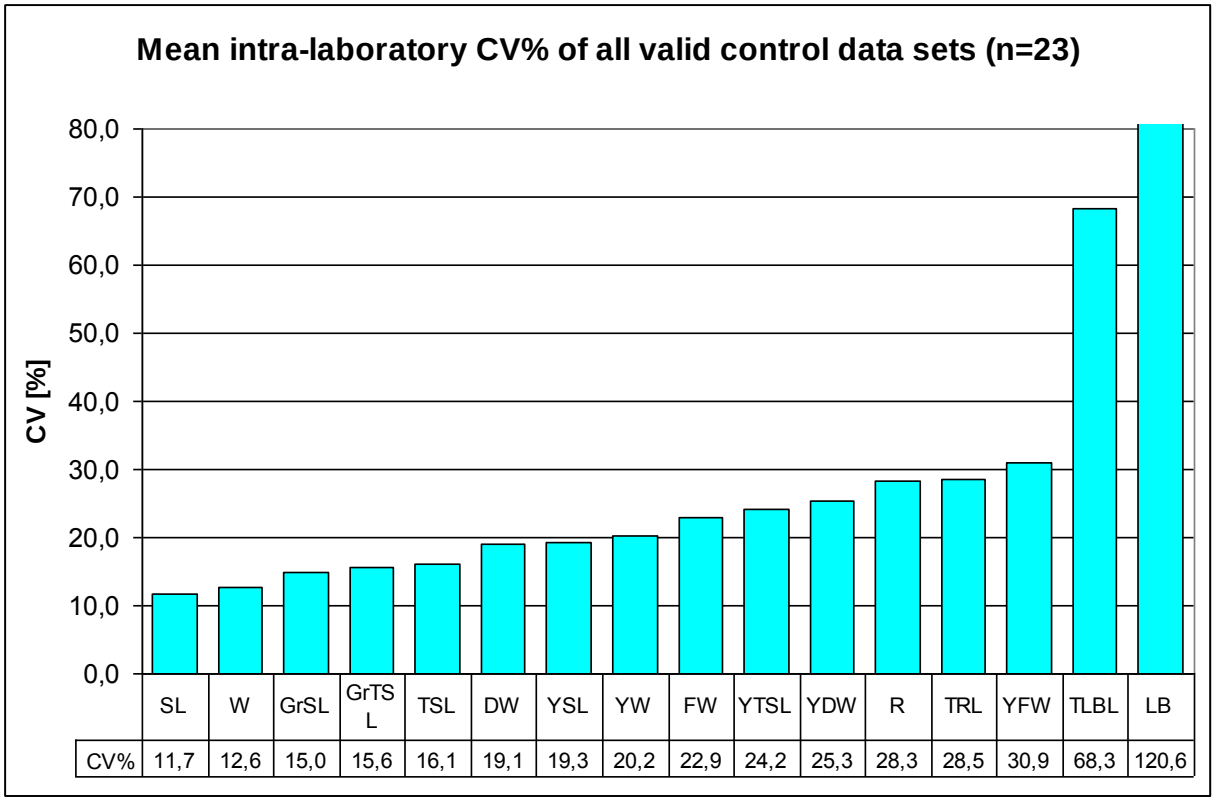


Fig. 16 Mean intra-laboratory coefficients of variation for all variables, in ascending order

Based on the coefficients of variation and the given test design (10 control replicates, 8 treatments with 5 replicates each), it is possible to derive the minimum detectable difference (%MDD), i.e. the level of the minimum detectable effect to be expected for the NOEC / LOEC determination (Table 5).
For the given design and a variable with a coefficient of variation of 30%, an effect has to be at least as great as ca. 25% to become detectable on a statistically significant level by means of Williams' test. Since ideally, the NOEC should not exceed the EC_{20} , it is indispensable for a test result with sufficient statistical power that the variability of parameters measured is minimized as much as possible and/or highly reproducible variables are selected.

Tab. 5 Detectable effect levels (MDD %) to be expected under the given test design (control n=10; 8 treatments with n=5), by using Williams' test for different variables, based on the mean values obtained in the present inter-laboratory comparison and mean standard deviations.

Variables	CV%	MDD%
SL, Shoot Length	11.7	10.9
W, Whorls	12.6	11.7
GrSL, Growth Rate Shoot Length	15.0	14.0
Gr TSL, Growth Rate Total Shoot Length	15.6	15.1
TSL, Total Shoot Length	16.1	16.0
DW, Dry Weight	19.1	17.8
YSL, Yield Shoot Length	19.3	17.5
YW, Yield Whorls	20.2	17.9
FW, Fresh Weight	22.9	19.3
YTSL, Yield Total Shoot Length	24.2	23.7
YDW, Yield Dry Weight	25.3	23.1
R, Roots	28.3	27.4
TRL, Total Root Length	28.5	26.7
YFW, Yield Fresh Weight	30.9	24.5

The minimum detectable effect levels listed in Table 5 represent a theoretical orientation. Depending on the concrete level of the respective variances in the individual test run and on whether the number of evaluable replicates deviates from the given test design, the MDDs actually resulting from a concrete test run can be lower, but also higher. The detectable effect levels actually obtained in the present inter-laboratory comparison are presented in Chapters 4.4 to 4.6.

4.3.3. Characterization of count variables

Both the number of roots and that of lateral branches are count variables. Their frequency distribution is a Poisson distribution with unknown mean value μ, with μ being estimated by the mean value of the data obtained. Under certain conditions, the Poisson distribution can be approximated by assuming a normal distribution.

The number of roots produced was between 1 and 9, which appears to approximate a normal distribution (Fig. 17, mean value 3.6). The assumption of normal distribution is supported by the Shapiro-Wilk's test (p(W) = 0.743). For the number of roots, the Poisson distribution can thus be approximated by assuming a normal distribution so that mean value comparison tests according to Dunnett or Williams may be used for NOEC determination.

In contrast, the absolute figures for the lateral branches with a maximum of 5 are clearly lower, and the mean value of 1.0 is too close to zero. Hence, it becomes obvious already on a visual basis that the frequency distribution cannot be approximated by a normal distribution (Fig. 18). Consequently, in the case of the number of lateral branches, it is necessary to use Fisher's exact test (with Bonferroni correction of the significance level) for NOEC determination.

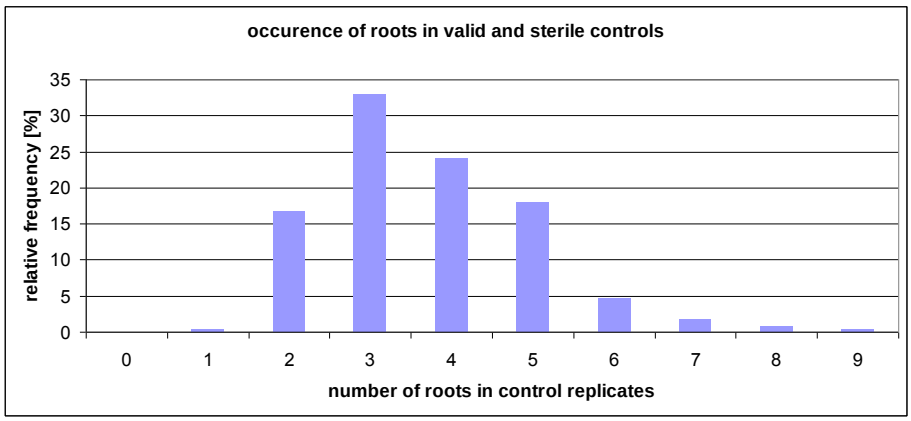


Fig. 17 Root formation in valid control replicates. Example of how to read the graphic representation: 33% of all control plants produced three roots each, 5% produced six roots each.

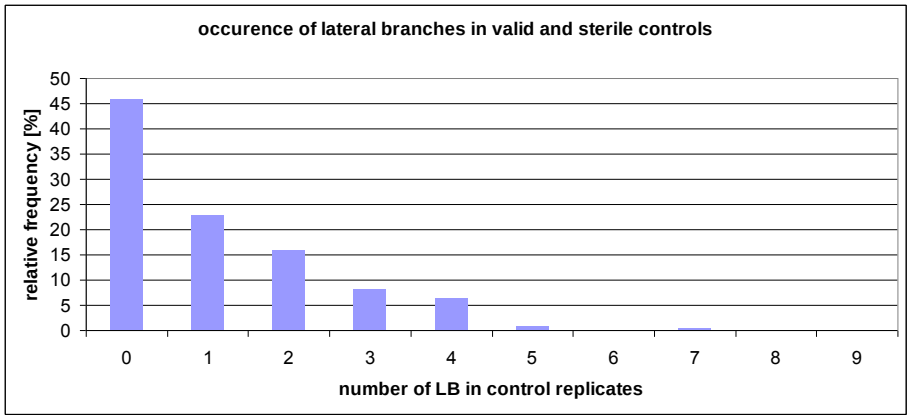


Fig. 18 Lateral branch formation in valid control replicates.
Example of how to read the graphic representation: 46% of all control plants did not produce lateral branches.

A more detailed analysis reveals that for several reasons and irrespective of the type of distribution, the number of lateral branches is not a suitable endpoint in the present inter-laboratory study:

1. Roots formed in 100 % of all control replicates, which means that the production of roots is a typical and reproducible differentiation process in the context of a normal growth of *Myriophyllum spicatum* in the present test system. Consequently, reduced or completely missing root production can be unequivocally rated as an effect of treatment.
   In contrast, only 54% of control replicates produced lateral branches at all. Hence, lateral branches are not produced as a general characteristic but apparently by coincidence. This is why differences observed between control and treatment cannot be clearly attributed to the treatment.

2. The number of roots shows an acceptable mean coefficient of variation of 28% (mean value 3.6; s = 1.3). In contrast, the mean number of lateral branches being 1.0 is extremely low (only 30% of control replicates produced more than one lateral branch) and with a coefficient of variation of more than 100% very poorly reproducible (s = 1.1).

4.3.4. Selection of variables for further evaluation

The majority of variables has shown intra-laboratory variability to be between 10% and 30% and thus, can be considered as generally suitable for the determination of toxicity parameters.
The number and length of lateral branches are the only variables unsuitable to serve as independent endpoints due to extremely high coefficients of variation. This is why they are exempt from the calculations of toxicity parameters (EC, NOEC) and sensitivity analysis presented below.
Nevertheless, determination of the total shoot length by means of the number and length of lateral shoots is required if (as in the present inter-laboratory comparison) the total shoot length is to be evaluated as an independent parameter.

4.4 Results for 3,5-DCP

4.4.1. 3,5-DCP: Mode of action and data used

3,5-DCP is characterized by a non-specific narcotic mode of action. The experimental plants remain morphologically unimpaired so that on principle, all endpoints to be recorded can be easily measured or counted.
Altogether, all 12 participating laboratories performed a test involving 3,5-DCP. 11 out of the 12 data sets submitted complied with the validity criteria; in one test, the control replicates were apparently contaminated with foreign organisms.

At concentrations ranging from 0.4 to 19.2 mg/L, all variables evaluated showed a pronounced dose-response relationship so that as a rule, significant fits were achieved on the basis of mean values. In few cases only (root number in three out of eleven data sets, total root length in four out of eleven data sets), fitting was performed on the basis of the replicates in order to achieve significant results[4].

On principle, the *yield values* determined by calculation for a defined variable show better dose-response relationships than the *original values* determined by measurement, and the EC_{50} derived is lower by a factor of ca. two, also showing a better reproducibility. This has to be attributed to the fact that for the yield values determined by calculation, inhibition of up to 100% is possible, while for the original values determined by measurement, inhibition values reach a mere 60% because for the latter, the starting values are always included. Below, this is demonstrated, using the main shoot length from data set 03 as an example (Figs. 19 and 20): In the control, the main shoot showed a growth factor of 2.27 (29.0 mm on d0 to 65.8 mm on d14, yield = 36.8 mm). At the highest concentration, the shoot length remained almost unchanged (28.4 mm on d0 to 29.3 mm on d14, yield = 0.9 mm), corresponding to almost no growth at all. This would correspond to an inhibition of the *yield* of 98 % (0.9 mm as opposed to 36.8 mm). In contrast, the original shoot length values resulted in a maximum inhibition of 55% because the initial shoot length was included (29.3 mm as opposed to 65.8 mm).

This results in a steeper course and better determination of the fit for the yield values (narrower confidence intervals) and therefore, a lower EC_{50} .

[4] Summary in file „35DCP EC50 fig_reprod labs + geom mean.xls“.

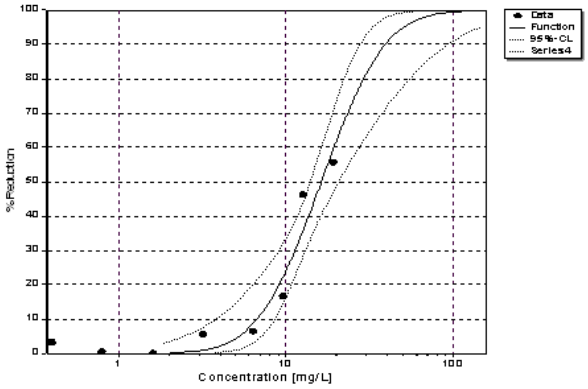


Fig. 19 Inhibition values and dose-response relationship original measurements main shoot length (L03 3,5-DCP)

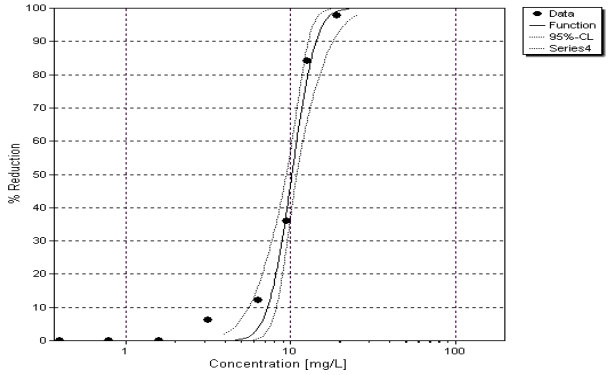


Fig. 20 Inhibition values and dose-response relationship yield main shoot length (L03 3,5-DCP)

The variables, number of roots and total root length are "yields" by definition because, on principle, at the starting of the test, no roots have formed yet.

4.4.2. 3,5-DCP: EC_{50} – Sensitivity of variables – reproducibility

Figs. 21 – 24 show the EC_{50} with 95% confidence intervals obtained by the different laboratories for selected variables (yield main shoot length, yield fresh weight, total root length, growth rate main shoot length), including the respective overall mean value with its 95% prediction interval.[5]
Fig. 25 lists the EC_{50} values for all variables by sensitivity.
Fig. 26 and Table 6 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the EC_{50} .

With one exception (L04, yield fresh weight), all EC_{50} values obtained by the different laboratories are within the 95% prediction interval. The EC_{50} values obtained by laboratory 04 are extremely low for almost all variables, a fact for which no cause could be identified.
The EC_{50} values obtained by laboratories 01 and 08 show comparatively wide 95% confidence intervals and hence, are relatively uncertain. In general, however, the EC_{50} values from all laboratories show a fairly good coincidence. This also applies to variables with high inter-laboratory variability, such as the fresh weight: In spite of yields varying by up to 70 %, the EC_{50} values differ by no more than 30 % (yield fresh weight L03: 272 mg/L; EC_{50} = 5.66 mg/L; L07: 383 mg/L; EC_{50} = 5.99 mg/L; L08: 114 mg/L; EC_{50} = 4.21 mg/L).
It may be concluded that laboratory-specific differences in test conditions and handling result in a higher variability of the measurements obtained for the different growth parameters (cf. Section 4.3.1.), however, affecting both controls and treatments to the same extent. Consequently, also tests with a quite different growth among the controls may result in similar EC_{50} values.

The EC_{50} value from the invalid test (L06, control replicates apparently contaminated with foreign organisms) is highly consistent with those from valid tests. Only the confidence intervals are comparatively large, which, however, was also observed in several other data sets (see above).

[5] The corresponding Figures for *all* variables are found in file „35DCP EC50 fig_reprod labs + geom mean.xls".

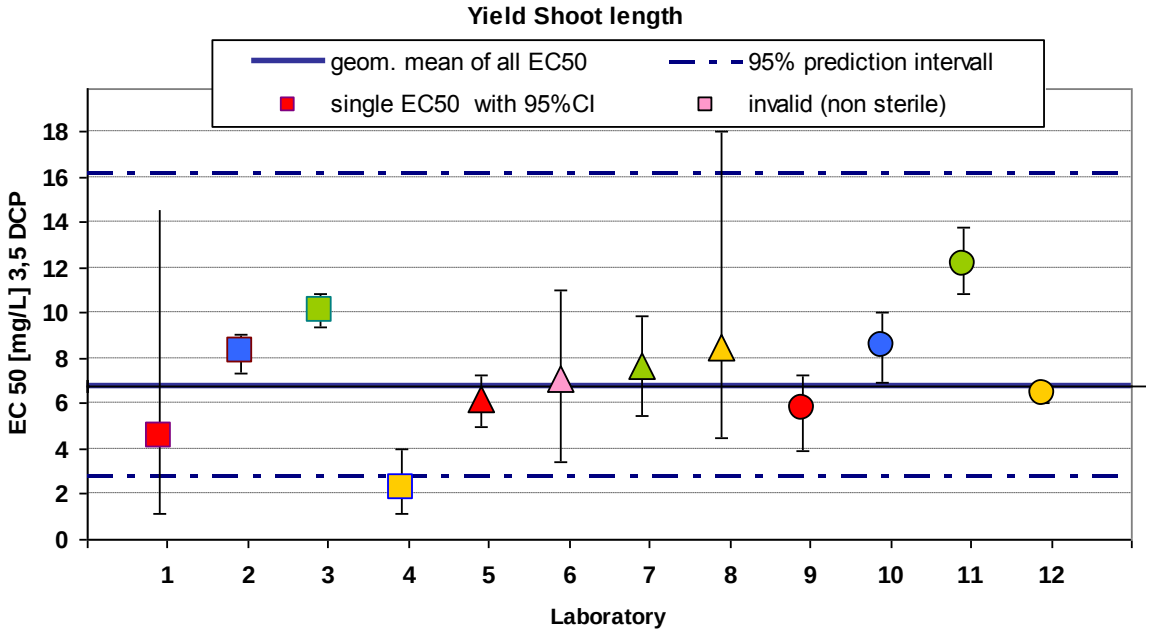


Fig. 21 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, yield main shoot length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

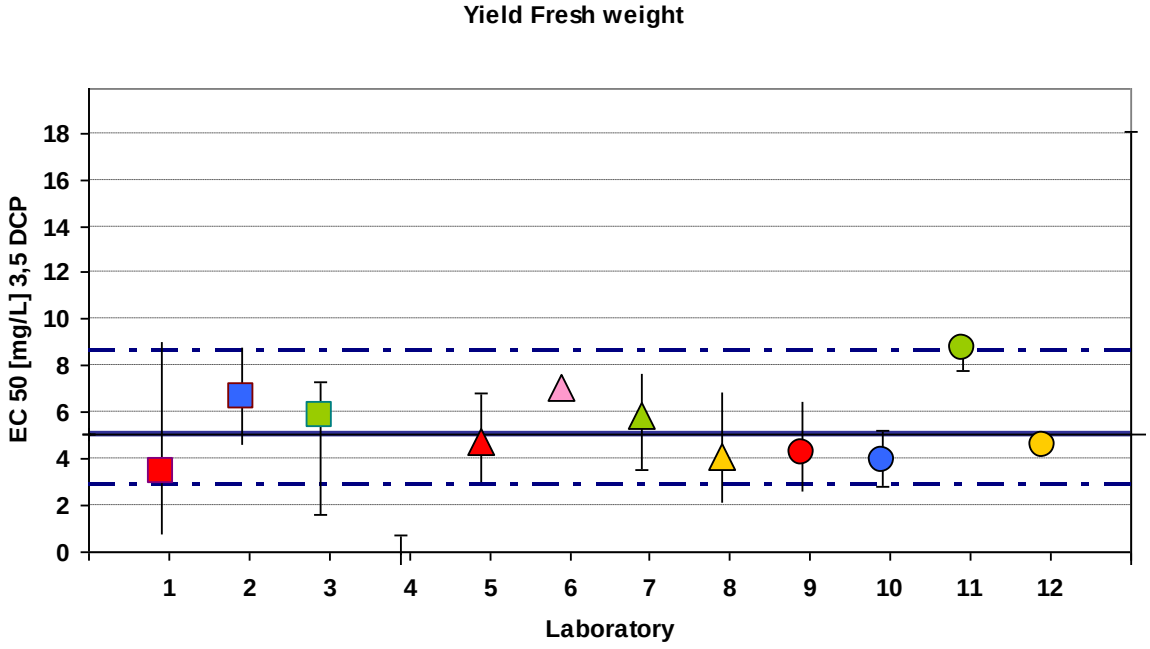


Fig. 22 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, yield fresh weight under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

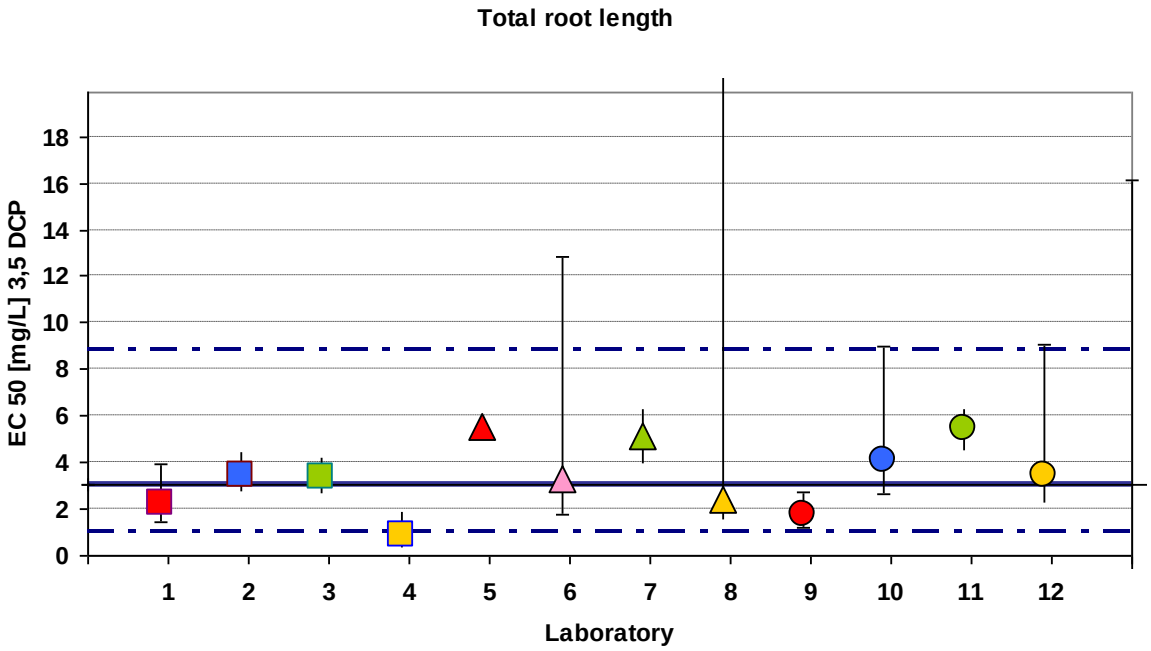


Fig. 23 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, total root length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

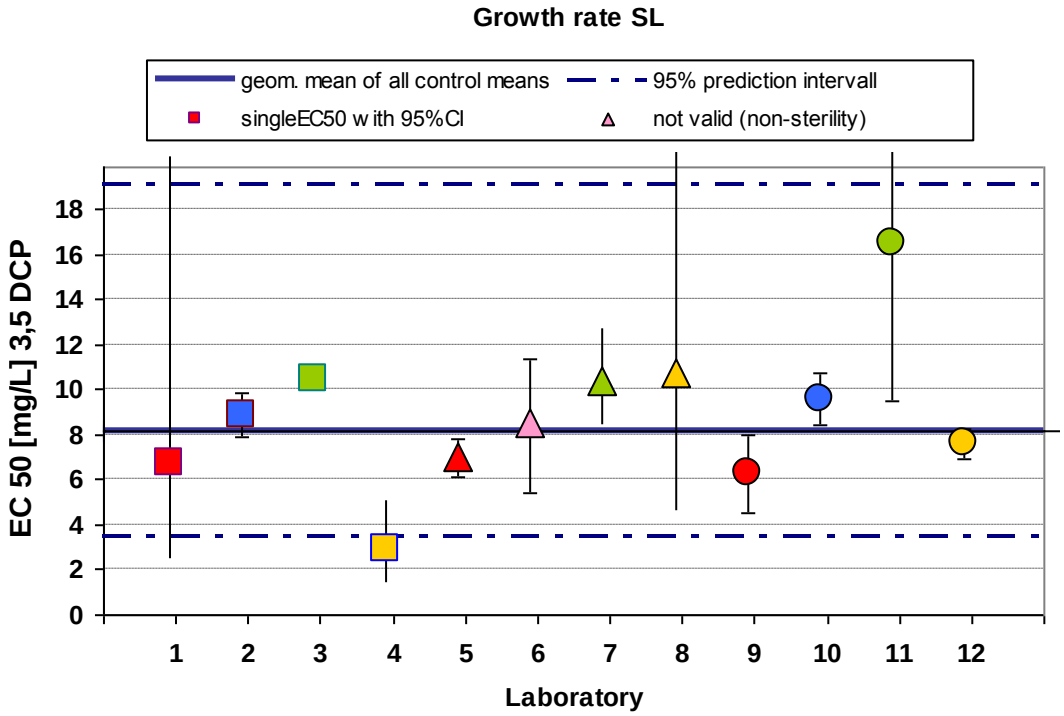


Fig. 24 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, growth rate main shoot length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

The mean EC_{50} values for the different variables on exposure to 3,5-DCP are between 3.2 and 15.9 mg/L (Fig. 25). The EC_{50} levels for the original values obtained by measurement are, on principle, about twice as high as those for the yield values calculated from these. The difference between the most sensitive and the least sensitive variable corresponds to a factor of ca. five. If only the exclusively yield-based variables are considered, the difference in sensitivity is smaller (factor of 2.7). The most sensitive variables are total root length (EC_{50} = 3.2 mg/L) and yield dry weight (EC_{50} = 3.8 mg/L), the least sensitive ones, the growth rates for main shoot length and total shoot length (EC_{50} = 8.3 and 8.6 mg/L, respectively).

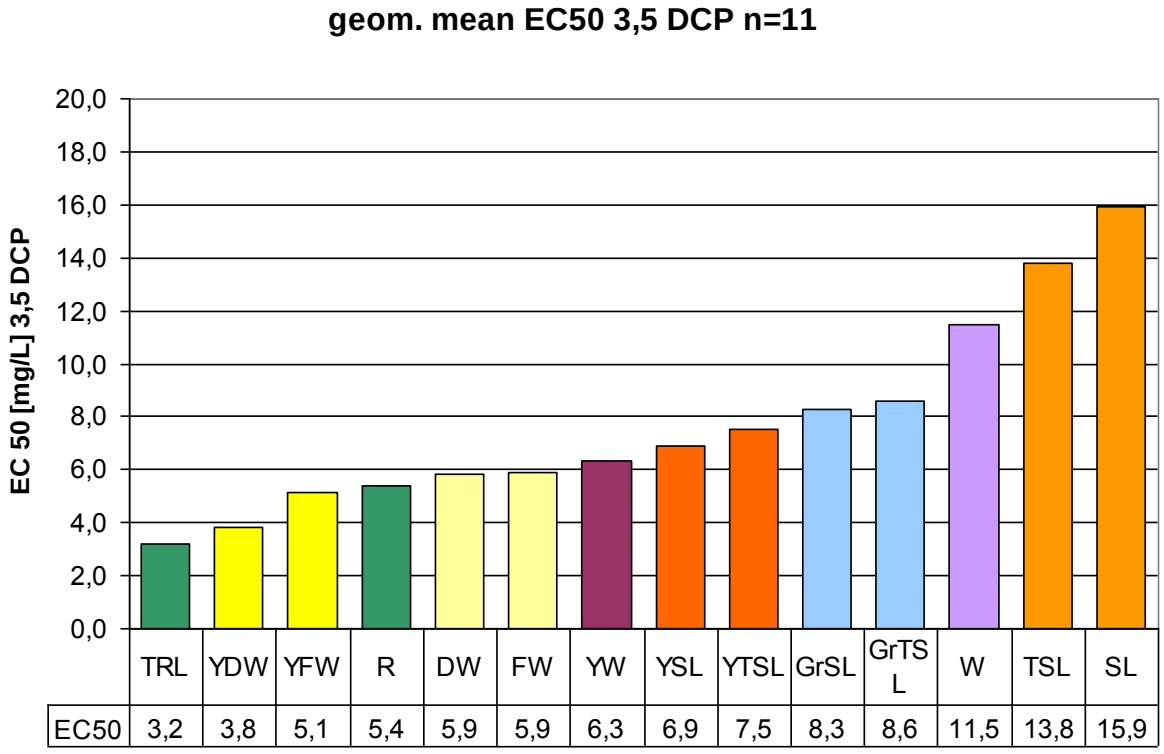


Fig. 25 Sensitivity of the different variables to 3,5-DCP: mean EC_{50} (based on 11 valid tests)

The 95% prediction interval of the mean values is used as a measure of reproducibility of the EC_{50} measured (Fig. 26). The narrower the prediction interval, the better will be the reproducibility.
Except for the dry weight, the narrowest prediction intervals are shown by those variables which also represent the most sensitive ones, namely root number, total root length and yield fresh weight.
The relationship between the width of the prediction interval and EC_{50} is considered as a measure of the relative quality of the prediction interval. Showing values between 0.9 and 3.4, this relationship is acceptable for all variables (Table 6).

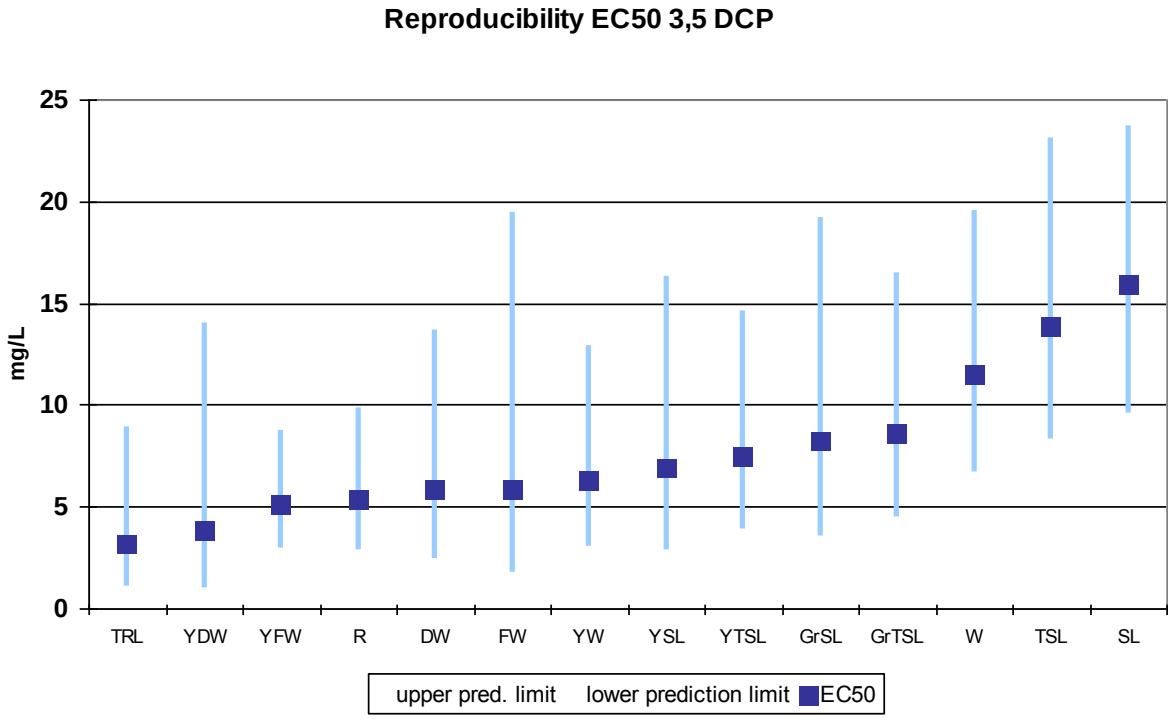


Fig. 26 Mean EC_{50} of the different variables on exposure to 3,5-DCP and their respective 95% prediction intervals as a measure of reproducibility (based on 11 valid tests).

Tab. 6 Relationship between width of 95% prediction interval and mean EC_{50} on exposure to 3,5-DCP (ratio)

	SL	TSL	W	YFW	R	GrTSL	YTSL	YW	GrSL	DW	YSL	TRL	FW	YDW
Ratio	0.9	1.1	1.1	1.1	1.3	1.4	1.4	1.6	1.9	1.9	1.9	2.4	3.0	3.4

4.4.3. 3,5-DCP: NOEC – Sensitivity of variables – Reproducibility

Figs. 27 – 30 show the NOEC determined by the different laboratories for selected variables including the respective overall mean value with its 95% prediction interval.[6]
In Fig. 31, the NOEC for all parameters are listed by sensitivity. The statistical power of the NOEC is examined based on the minimum detectable difference (%MDD) and by comparison with the EC_{20} (Table 7).
Fig. 32 and Table 8 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the NOEC.

In the majority of cases, the NOEC values from the different laboratories vary between two or three concentration levels around the mean value. Due to single upward variations towards the next concentration step(s), the 95% prediction intervals of the mean values become wider upwards than downwards in an asymmetric way.

The NOEC values from the invalid test (L06, control replicates apparently contaminated with foreign organisms) are highly consistent with those from tests not contaminated with foreign organisms.

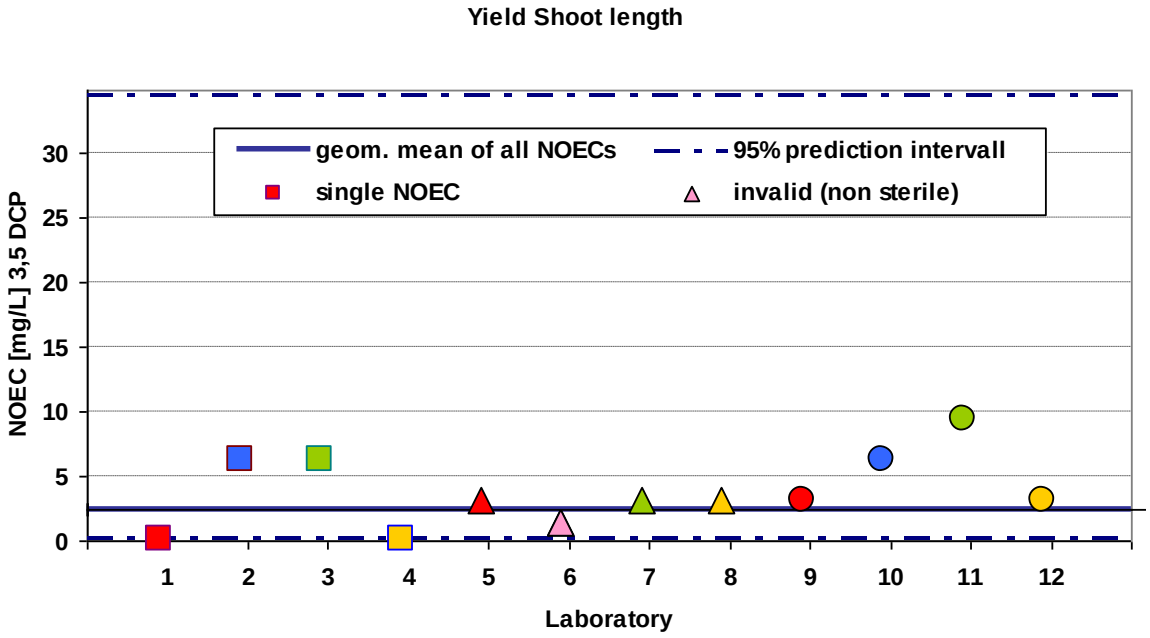


Fig. 27 Laboratory-specific NOEC for the variable, yield main shoot length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

[6] The corresponding Figures for *all* variables are found in file „35DCP NOEC fig_reprod labs + geom mean.xls".

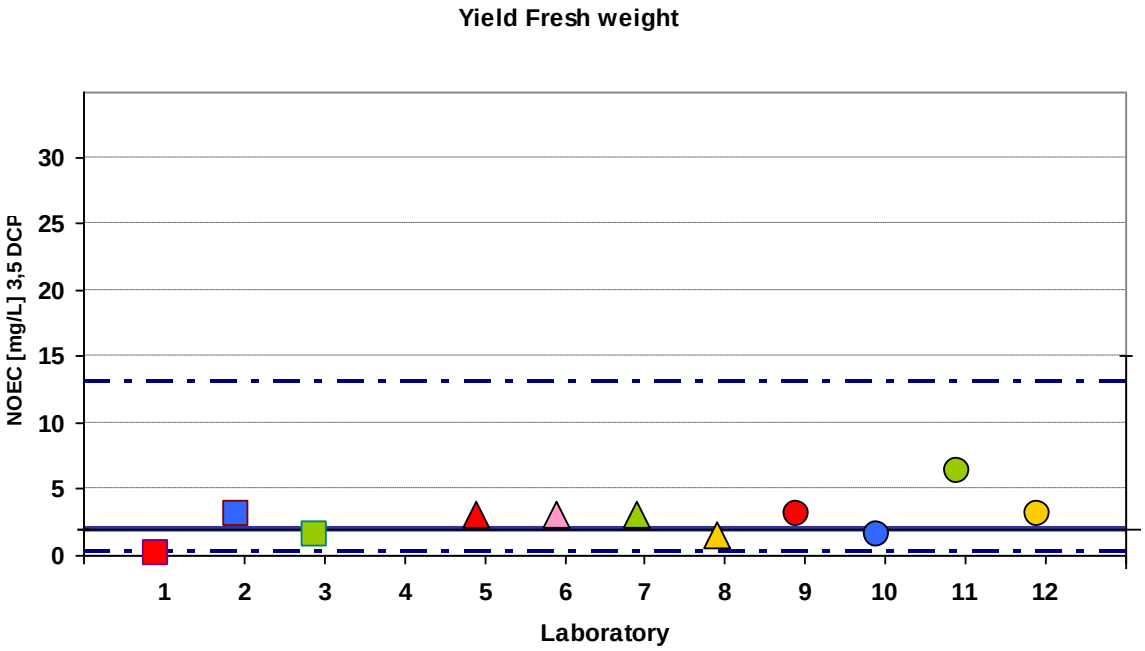


Fig. 28 Laboratory-specific NOEC for the variable, yield fresh weight under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

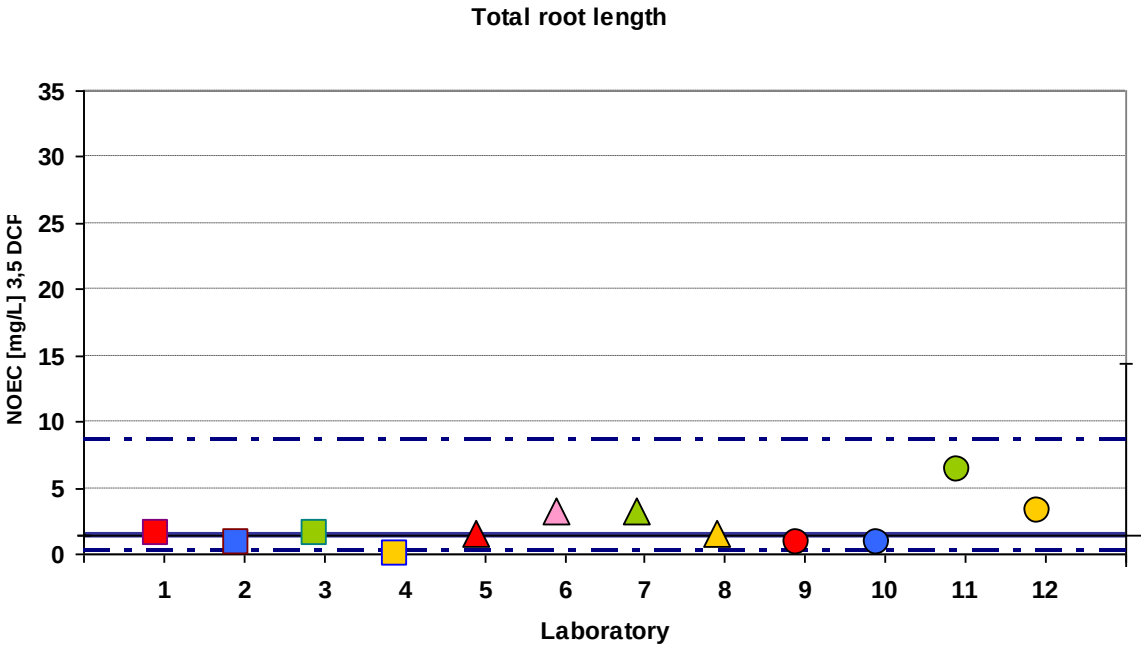


Fig. 29 Laboratory-specific NOEC for the variable, total root length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

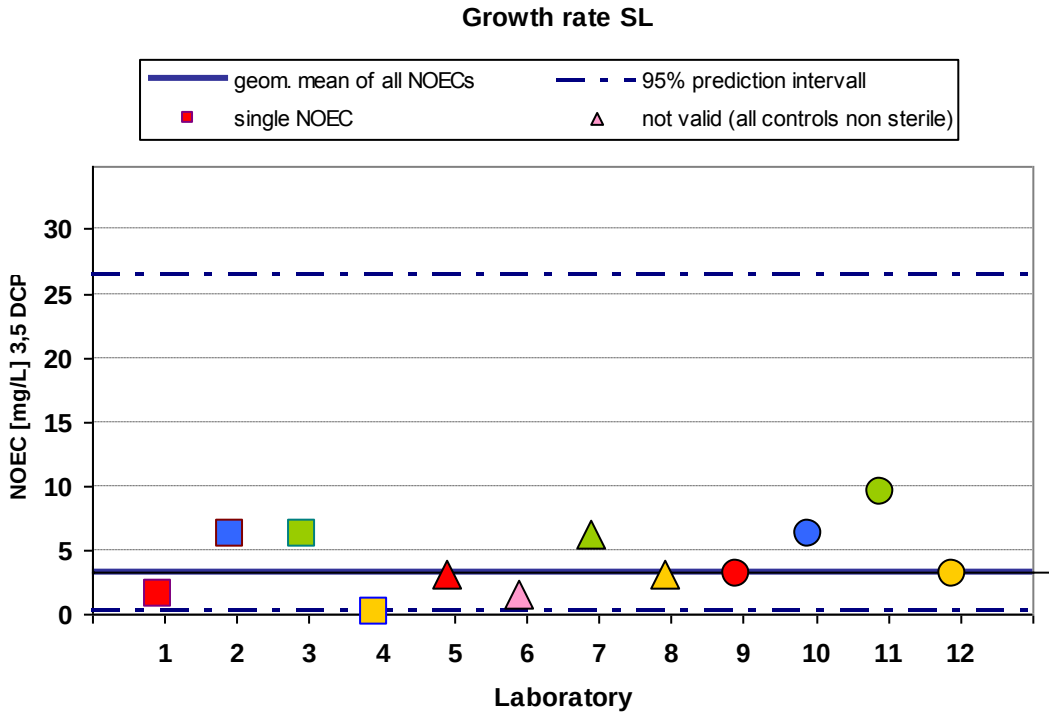


Fig. 30 Laboratory-specific NOEC for the variable, growth rate main shoot length under exposure to 3,5-DCP (coloured symbols); solid and broken lines: Mean NOEC with a 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

The mean NOEC values for the different variables on exposure to 3,5-DCP are between 1.4 and 4.1 mg/L (Fig. 31). Naturally, unlike the EC_{50} values, the NOEC values for original values and yield values do not show any systematic differences. The difference between the levels of the most sensitive and the least sensitive variable corresponds to a factor of ca. three.
The most sensitive variable is the total root length (NOEC = 1.4 mg/L), followed by all weight parameters (NOEC 2.1 – 2.2 mg/L). The yield total shoot length (NOEC = 3.9 mg/L), and the original total shoot length (NOEC = 4.1 mg/L) are found to be the least sensitive variables.

The EC_{50} and NOEC values reveal a comparable sensitivity of the variables. An exception is seen for the root number, whose NOEC assumes a less sensitive position than the EC_{50}. However, this finding may be coincidental, considering the possible discrete concentration steps the NOEC may assume and the generally very small ranges of the mean NOEC determined.

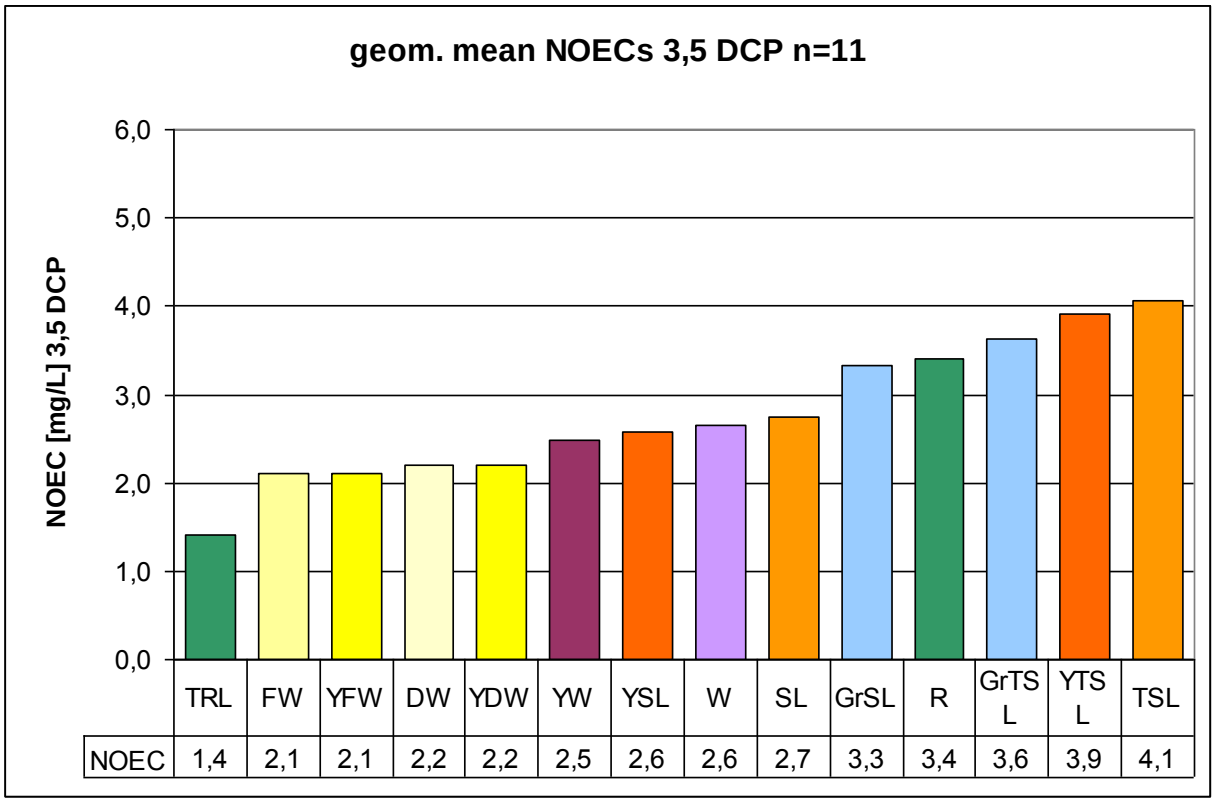


Fig. 31 Sensitivity of the different variables to 3,5-DCP: Mean NOEC (based on 11 valid tests)

A NOEC is of sufficient statistical power only if it corresponds to a sufficiently small effect. As a rule, the NOEC should not exceed the EC_{20}. This is demonstrated by the %MDD (minimum detectable difference) of the multiple test used: The %MDD represents the minimum level of an effect expressed as per cent of the control value which is required for the effect to be detectable on a statistically significant level. Table 7 provides an overview of the quality of the NOEC values established (evaluated only for yield values and growth rates).

The mean NOEC either corresponds to an approximate mean EC_{20} (root number, total root length, yield fresh weight) or is lower (yield shoot length and total shoot length, yield dry weight, yield number of whorls, growth rates). The minimum detectable differences in Williams' test were 20% on average, somewhat lower for the growth rates (14%-16%), and somewhat higher for the total root length (28%).

Altogether, the mean %MDDs actually obtained in the multiple Williams' tests performed correspond to those theoretically calculated. This supports the conclusions regarding the general suitability of the different variables for NOEC determination that were made on the basis of the coefficients of variation measured.

Tab. 7 Comparison of the NOEC and EC_{20} for 3,5-DCP; Frequency of Williams' test used as a multiple test for NOEC determination, resulting mean %MDD (minimum detectable difference). Evaluation of yields and growth rates.
Entered for comparison (last column): %MDD theoretically calculated, based on the test design and the mean coefficients of variation for the controls (cf. Table 5).

Variables	NOEC [mg/L] geom. mean n=11	EC_{20} [mg/L] geom. mean n=11	Williams test x times of 11	mean MDD [%] Williams-Test	theoretically calculated %MDD
R	3.4	3.8	6	22	*27.4*
TRL	1.4	1.3	10	28	*26.7*
YSL	2.6	3.9	9	16	*17.5*
YFW	2.1	1.7	5	20	*24.5*
DW	2.2	2.9	2	23	*23.1*
YTSL	3.9	5	8	21	*23.7*
YW	2.5	3.7	10	18	*17.9*
GrSL	3.3	4.4	7	14	*14.0*
GrTSL	3.6	5.5	6	16	*15.1*

The 95% prediction interval of the mean values is used as a measure of reproducibility of the calculated NOEC (Fig. 32). The narrower the prediction interval, the better is the reproducibility.

Analogous to the reproducibility of EC_{50} values, the narrowest absolute prediction intervals are shown by the number of roots and the total root length, followed by the weight parameters. In absolute terms, the NOEC values resulting from these variables are characterized by the best reproducibility.

The relationship between the width of the prediction interval and the mean NOEC is considered as a measure of the relative quality of the prediction interval. Showing values between 2.2 and 13.2, this relationship is highly variable (Table 8). The best reproducibility is seen for the NOEC values resulting from the root number, the total shoot length and the yield total shoot length. The relatively widest prediction intervals are seen for the main shoot length, yield main shoot length and the number of whorls, as well as the growth rates.

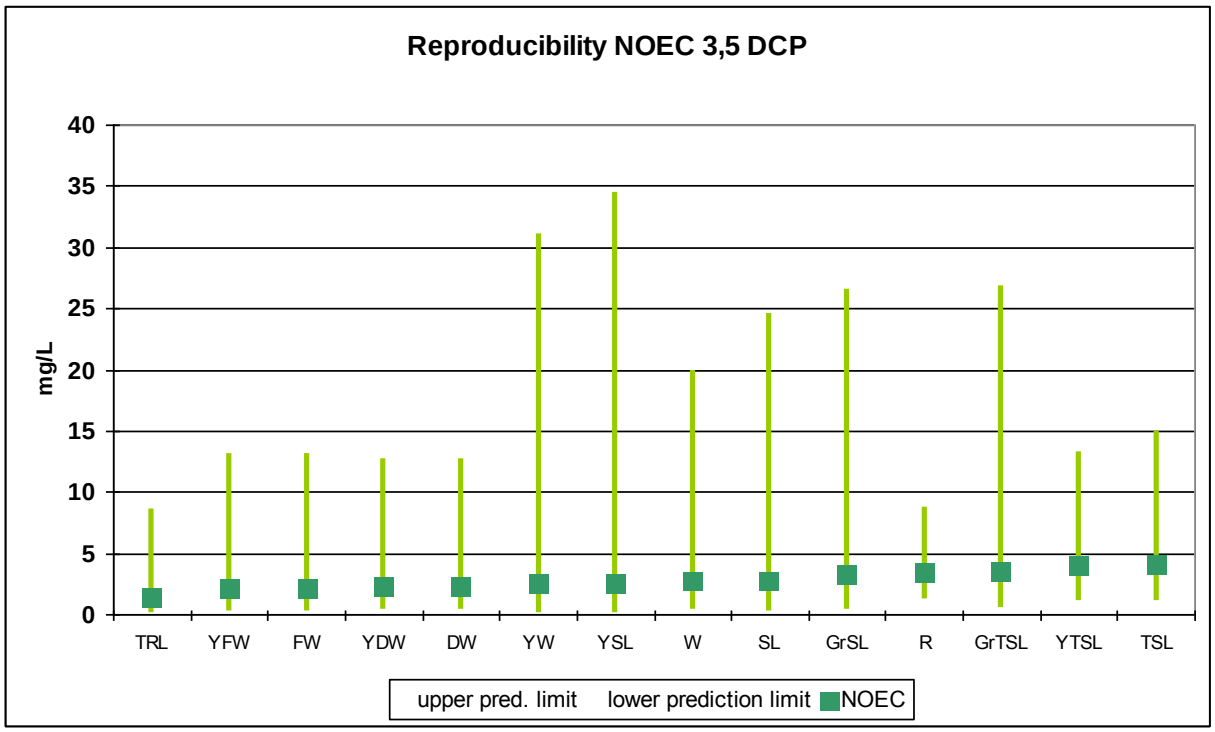


Fig. 32 Mean NOEC of the different variables on exposure to 3,5-DCP and their respective 95% prediction intervals as a measure of reproducibility (based on 11 valid test runs).

Tab. 8 Relationship width of 95% prediction interval and mean NOEC on exposure to 3,5-DCP (ratio)

	R	YTSL	TSL	DW	YDW	TRL	FW	YFW	GrTSL	W	GrSL	SL	YW	YSL
Ratio	2.2	3.1	3.4	5.6	5.6	6.0	6.1	6.1	7.3	7.4	7.9	8.8	12.4	13.2

4.5 Results for 2,4-D

4.5.1. 2,4-D: Mode of action and data used

2,4-D is an auxin herbicide, i.e. a substance interfering with growth, causing for example shoot deformation, twisted stalks, chlorophyll degradation and wilting. The measurability and statistical evaluability of a number of variables on exposure to 2,4-D is impaired by the factors described below:

Fig. 33 Test with 2,4-D, replicate with 2000 µg/L, d14; L02 2,4-D img 46

1. With increasing concentrations, the experimental plants become increasingly deformed (Fig. 33) so that it may even become impossible to count and measure, respectively, parameters such as the number of whorls or shoot length. This leads to gaps in some data sets which impair the evaluability (from low number of replicates to total dropout of a concentration for evaluation).

2. At low and moderate concentrations, growth-promoting effects are seen. The corresponding concentrations cannot be evaluated for establishing the dose-response relationship because they do not show inhibition. EC determination is possible, however, with confidence intervals increasing due to the reduced data volume (Fig. 34).

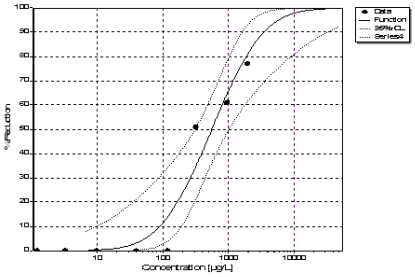


Fig. 34 Test with 2,4-D, dose-response relationship for yield main shoot length; L02

3. Due to masking of growth-promoting and growth-inhibiting effects (Table 9), the adjustment of a dose-response relationship is impossible so that neither an EC_{50} can be determined nor a definite NOEC derived.

Tab. 9 Results for the variable, fresh weight on exposure to 2,4-D (L03). The inhibition values calculated allow neither a significant dose-response analysis nor a clear-cut determination of NOEC values.

Treatm.[µg/L]	Mean	Std. Dev.	n	%Reduction
Control	318,290	39,2943	10	0,0
1,200	258,960	45,4145	5	18,6
3,200	292,980	52,2668	5	8,0
9,600	234,000	49,4698	5	26,5
40,000	201,460	53,1990	5	36,7
120,000	178,580	41,9874	5	43,9
320,000	235,660	32,6480	5	26,0
960,000	335,120	65,8036	5	-5,3
2000,000	285,580	67,6191	5	10,3

4. In the present inter-laboratory comparison, at mean concentrations on exposure to 2,4-D, ”root growth“ is induced in the leaf axils of the entire main shoot. “Rudimentary roots“ of 1 mm length were produced over the entire shoot length (Fig. 35).

In the corresponding replicates, up to 17 rudimentary roots of 1 mm length were counted. Hence, promotion rather than inhibition was measured, as compared to the controls. Even if at higher concentrations, inhibition of up to 100 % is observed, the corresponding dose-response curves become shifted to the right (towards higher concentrations). The resulting ECx values are clearly higher (by a factor of 20) than those that would have to be expected without the promotion seen at medium concentrations. Due to the root-inducing effect, the root number parameter thus appears to be a less sensitive endpoint.

Fig. 35 Test with 2,4-D, 40 µg/L, repl 4, d14; L09 2,4-D, induction of rudimentary roots in the leaf axils

In contrast, the total root length is hardly affected by the root-inducing effect because due to their small size, the additional roots can be neglected for this parameter.

Although observed in all tests with 2,4-D, the additional rudimentary roots were identified and counted in only three out of the ten data sets recorded in the present inter-laboratory comparison. The possible formation of rudimentary roots in the leaf axils should be explicitly pointed out in an envisaged standardization of the test procedure because otherwise, the rudimentary roots could escape consideration in the evaluation.

Altogether, ten of the participating laboratories performed a test with 2.4-DCP. The validity criteria were fulfilled by six of the ten data sets submitted. In one test, the control replicates were apparently contaminated with foreign organisms. In three other tests, the criterion of shoot length doubling failed.
In the remaining six data sets, no significant dose-response relationships could be achieved for the variables, fresh weight, dry weight and yield dry weight due to the framework conditions described above so that no EC_{50} values could be determined in these cases. For the variable, yield fresh weight, the EC_{50} could be determined in one case only, for the root number, in two cases. Since the data available vary from one variable to the other, the number of underlying data sets has been stated for the results which are presented below.

4.5.2. 2,4-D: EC50 – Sensitivity of variables – Reproducibility

In analogy to the results for 3,5-DCP, Figs. 36 – 39 show the EC_{50} with 95% confidence intervals for selected variables (since only one significant EC_{50} is available for the yield fresh weight, the yield total shoot length is shown instead), including the respective overall mean value with its 95% prediction interval.[7]
In Fig. 40, the EC_{50} values for all variables are listed by sensitivity.
Fig. 41 and Table 10 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the EC_{50} .

In general, the calculation of significant dose-response relationships on exposure to 2,4-D is difficult for many variables. Therefore, the volume of data available for comparison is less comprehensive than that for 3,5-DCP. All of the EC_{50} values obtained by the different laboratories are within the 95% prediction interval.

Some EC_{50} values show comparatively wide 95% confidence intervals and hence, are relatively unreliable. In general, however, the EC_{50} values from all laboratories have shown relatively little deviation from one another.

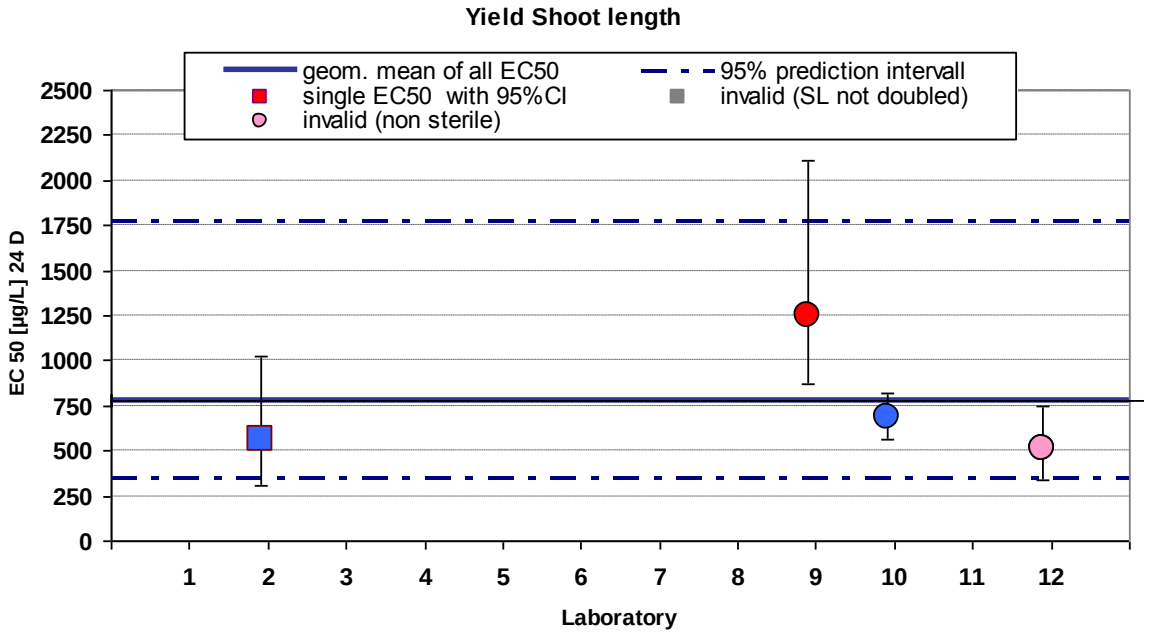


Fig. 36 Laboratory-specific EC_{50} with 95% confidence interval for the variable, yield main shoot length under exposure to 2.4-D (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

[7] The corresponding Figures for *all* variables are found in file „24D EC50 fig_reprod labs + geom mean.xls“.

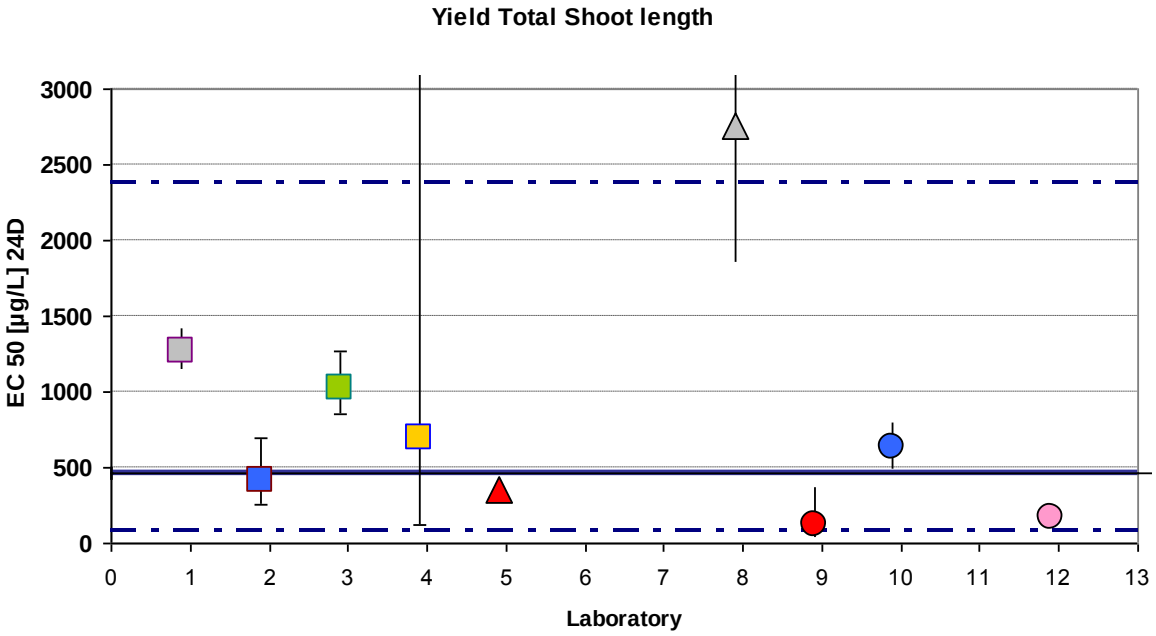


Fig. 37 Laboratory-specific EC_{50} with 95% confidence interval for the variable, yield total shoot length under exposure to 2.4-D (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbols: Test invalid due to lack of doubling of shoot length.

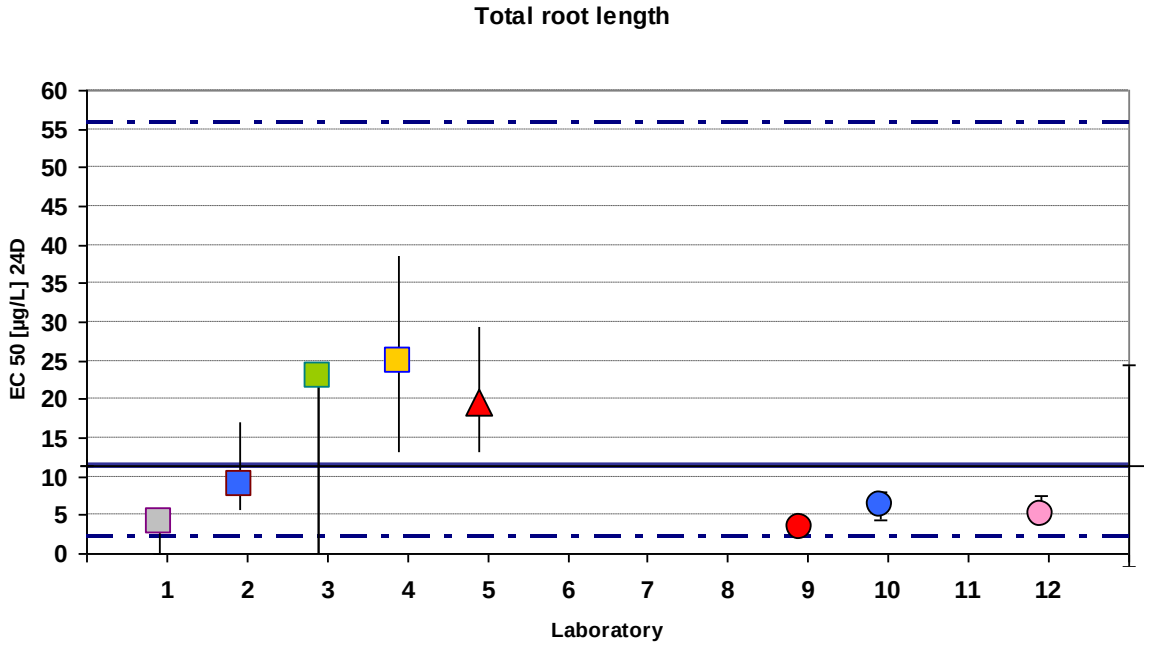


Fig. 38 Laboratory-specific EC_{50} with 95% confidence interval for the variable, total root length under exposure to 2.4-D (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbol: Invalid due to lack of doubling of shoot length.

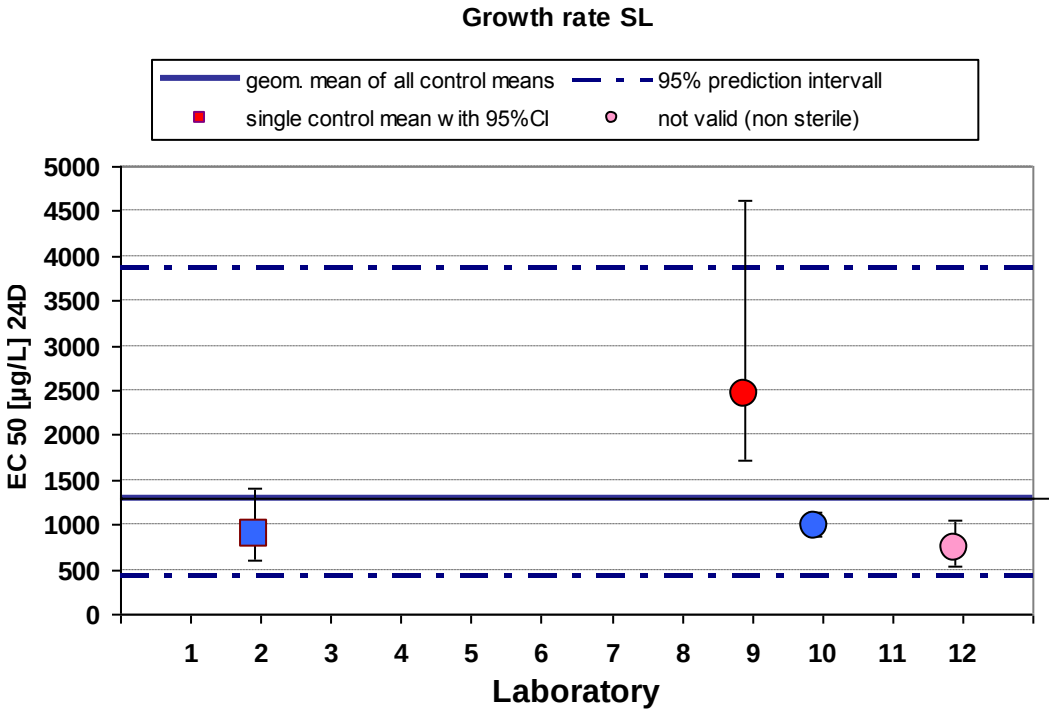
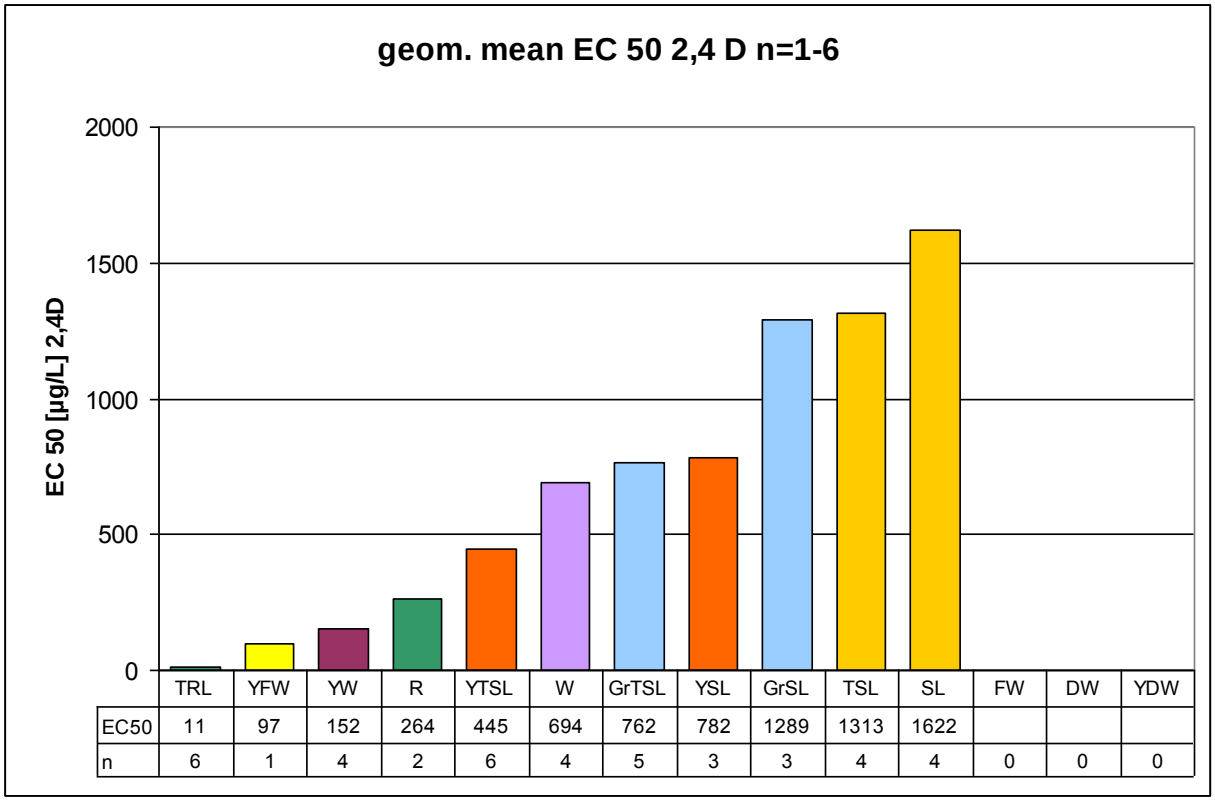


Fig. 39 Laboratory-specific EC_{50} with 95% confidence interval for the variable, growth rate main shoot length under exposure to 2.4-D (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

The EC_{50} values from the invalid test run (L12, control replicates apparently contaminated with foreign organisms) are consistent with those from tested samples not contaminated with foreign organisms. No clear trend is seen for the EC_{50} values from tested samples with reduced growth. In some cases, they are consistent with those from valid tests, in others, obvious upward deviations are observed (corresponding to a reduced sensitivity).

The mean EC_{50} values of the evaluable variables on exposure to 2.4-D range from 11 to 1 622 µg/L (Fig. 40). In analogy to the effect of 3,5-DCP, the EC_{50} values for the original values obtained by measurement are at least twice as high as those for the yield values calculated from these. The difference between the most sensitive variable and the least sensitive one correspond to a factor of ca.150.
If only the yield-based variables are considered, the difference in sensitivity is smaller, but still high (a factor of 117). This means that the sensitivity differences shown by the variables evaluated are more pronounced on exposure to 2,4-D than to 3,5-DCP.
The by far most sensitive variable is that of total root length (EC_{50} = 11 µg/L), followed by the yield fresh weight (EC_{50} = 97 µg/L) and the yield whorl number (EC_{50}= 152 µg/L). For the yield fresh weight, there is only one evaluable EC_{50} , which, however, is supported by the corresponding NOEC of 16.5 µg/L, which is based on as many as 4 data sets (see below, Fig. 46).
The fourth position is occupied by the number of roots with a mean EC_{50} of 264 µg/L, based on two test runs where the rudimentary roots were identified and counted as roots. Without taking into account the rudimentary roots, the resulting EC_{50} would have corresponded to ca. 15 µg/L and thus, to the same order of magnitude as that based on the total root length.
The least sensitive variables (among those based on yields) are the growth rate and the yield main shoot length (EC_{50} = 1289 and 782 µg/L, respectively). The fact that effects were well observed irrespective of the shoot length is demonstrated for the variable, yield whorl number: Showing an EC_{50} of 152 µg/L, it reacts obviously in a more sensitive way than the shoot length parameters.



	TRL	YFW	YW	R	YTSL	W	GrTSL	YSL	GrSL	TSL	SL	FW	DW	YDW
EC50	11	97	152	264	445	694	762	782	1289	1313	1622			
n	6	1	4	2	6	4	5	3	3	4	4	0	0	0

Fig. 40 Sensitivity of the different variables on exposure to 2.4-D: Mean EC_{50} (the number of underlying laboratory-specific EC_{50} levels stated)

The 95% prediction interval of the mean values served as a measure of reproducibility of the EC_{50} levels measured (Fig. 41). The narrower the prediction interval, the better is the reproducibility.
The narrowest prediction intervals are seen for those variables also showing the highest sensitivity, namely total root length, number of roots and yield whorl number (for the variable, yield fresh weight, which has also been classified as a sensitive one, it is impossible to determine the reproducibility because only one data set could be evaluated).
The relationship between the width of the prediction interval and the mean EC_{50} is considered as a measure of the relative quality of the prediction interval (Table 10). Showing values between 1.0 and 3.3, this relationship resembles the good one obtained under exposure to 3,5-DCP for many variables. However, some variables reveal a higher variability of EC_{50} values (ratios between 4.4 and 6.0). This may be attributed also to the fact that the mean value is based on fewer data (three to five EC_{50} values instead of 11 for 3,5-DCP).

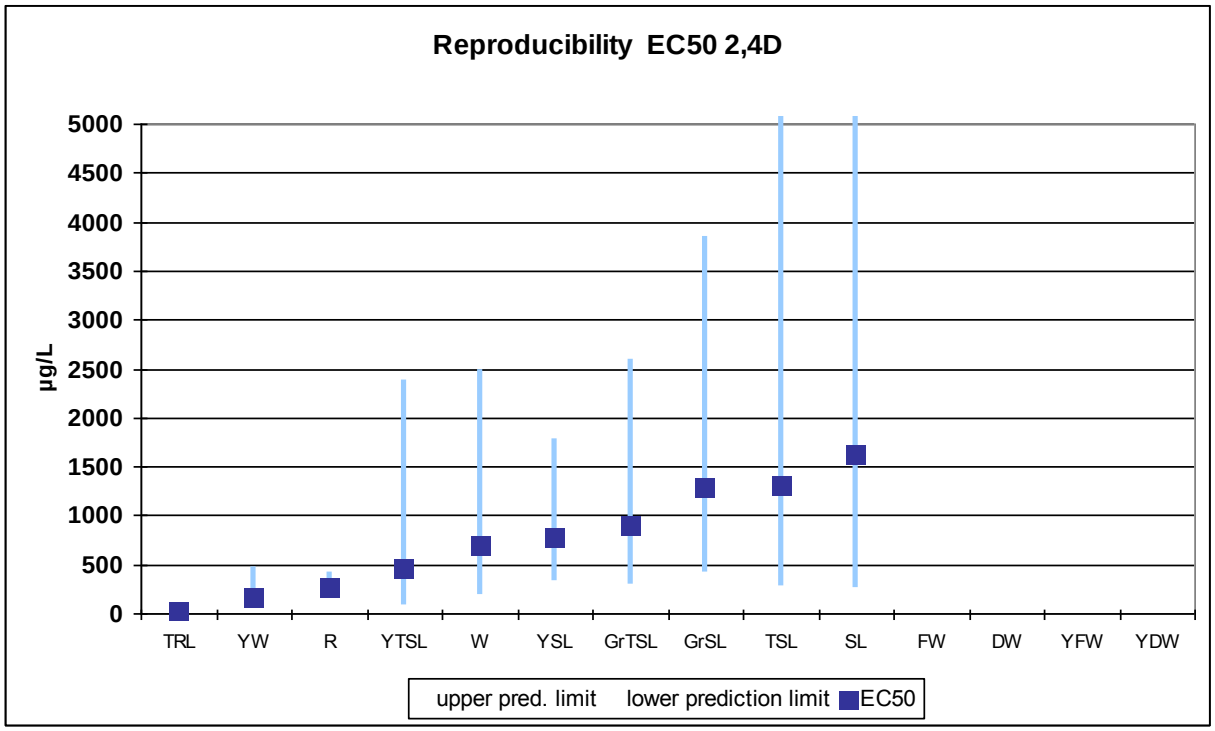


Fig. 41 Mean EC_{50} for the different variables on exposure to 2.4-D and their respective 95% prediction intervals as a measure of reproducibility (number of underlying laboratory-specific EC_{50} as in Fig. 40).

Tab. 10 Relationship between width of 95% prediction interval and mean EC_{50} on exposure to 2.4-DCP (ratio) n.e. = not evaluable. Number of underlying laboratory-specific EC_{50} as in Fig. 40.

	R	YSL	GrTSL	GrSL	YW	W	TSL	YTSL	TRL	SL	FW	DW	YFW	YDW
Ratio	1.0	1.8	2.6	2.7	2.9	3.3	4.4	4.9	4.9	6.0	n.a.	n.a.	n.a.	n.a.

4.5.3. 2,4-D: NOEC – Sensitivity of variables – Reproducibility

Figs. 42 – 45 show the NOEC values determined by the different laboratories for selected variables including the respective overall mean value with its 95% prediction interval.[8]
In Fig. 46, the NOEC values for all variables are listed by sensitivity. The statistical power of the NOEC values is examined based on the minimum detectable difference (%MDD) and by comparison with the EC_{20} (Table 11).
Fig. 47 and Table 12 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the NOEC.

The comparability of the NOEC from the different laboratories varies to a considerable extent depending on the variable. For the depicted variables, yield main shoot length, growth rate main shoot length and total root length, the results from the laboratories are largely consistent. For the yield fresh weight, this is, on principle, also the case. However, the test from L04 resulted exclusively in promotion effects, which is why this NOEC value was designated “larger than highest concentration”. As a result, the 95% prediction interval for the mean value becomes extremely wide (Fig. 43).

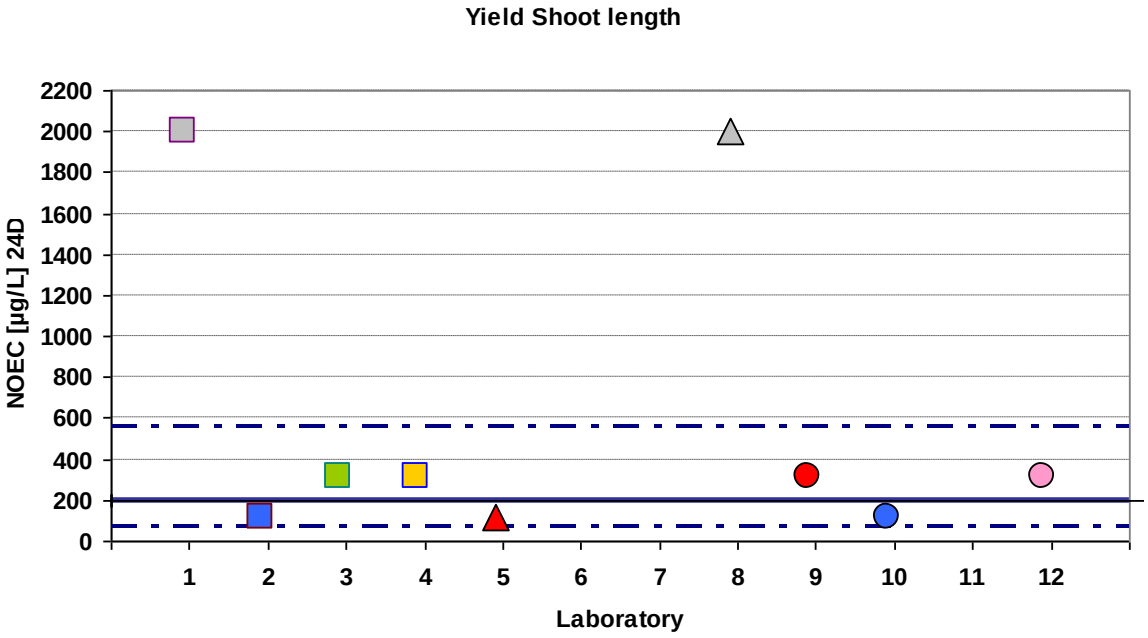


Fig. 42 Laboratory-specific NOEC for the variable yield main shoot length under exposure to 2,4-D (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbols: Test invalid due to lack of doubling of shoot length.

[8] The corresponding figures for *all* variables are found in file „24D NOEC fig_reprod labs + geom mean.xls“.

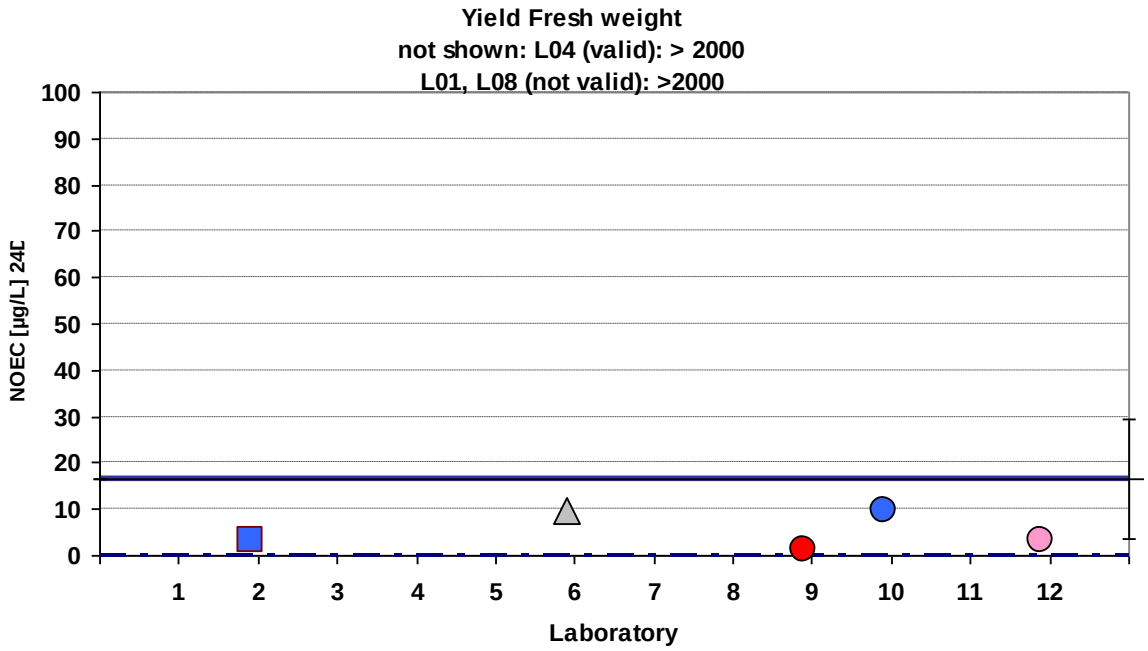


Fig. 43 Laboratory-specific NOEC for the variable yield main shoot length under exposure to 2,4-D (coloured symbols); not shown: L04: > 2000 µg/L; solid and broken lines: Mean NOEC with 95% prediction interval (upper limit not shown because value = 10800); pink symbol: Test invalid due to apparent contamination; grey symbols: Test invalid due to lack of doubling of shoot length, of these, not shown: L01 and L08: > 2000 µg/L.

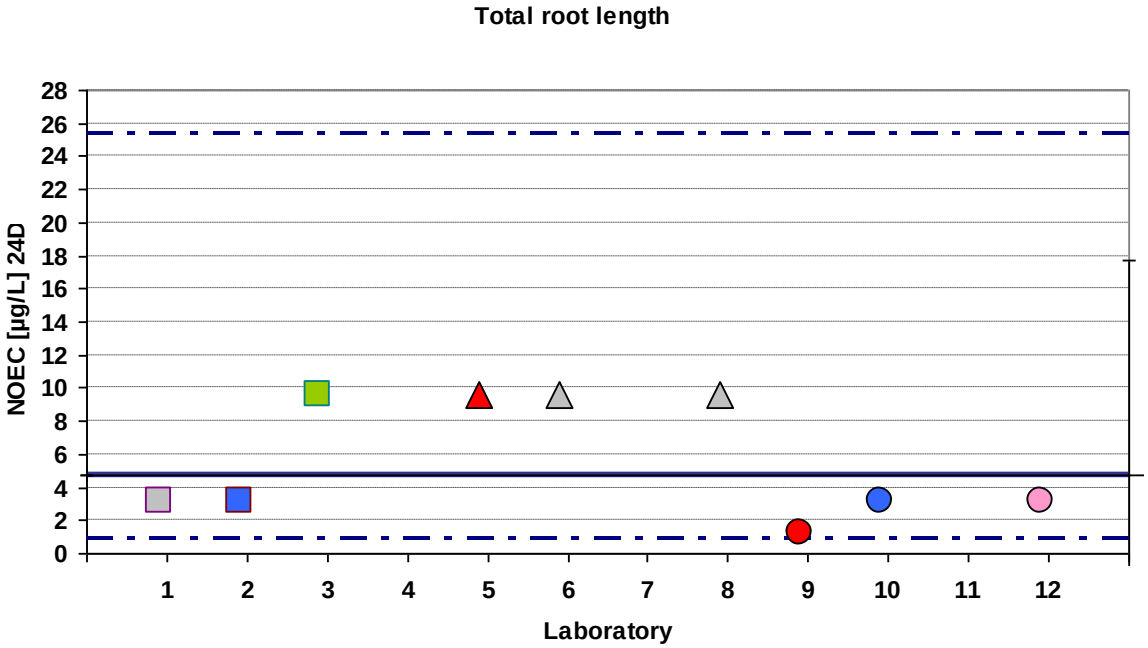


Fig. 44 Laboratory-specific NOEC for the variable total root length under exposure to 2,4-D (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbols: Test invalid due to lack of doubling of shoot length.

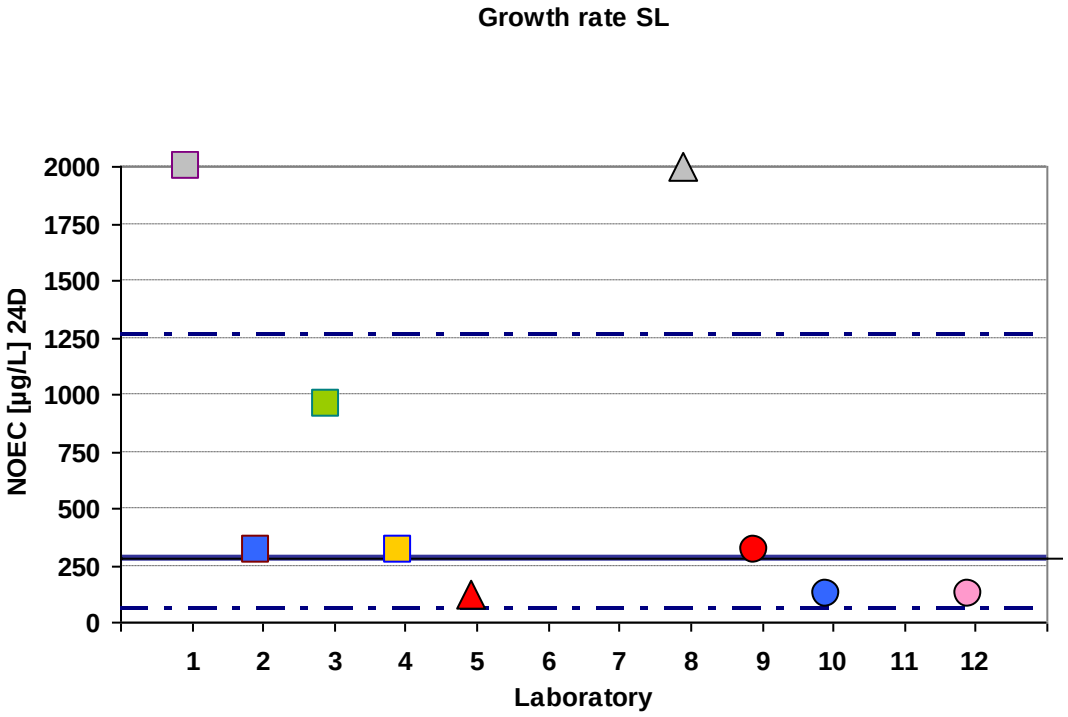


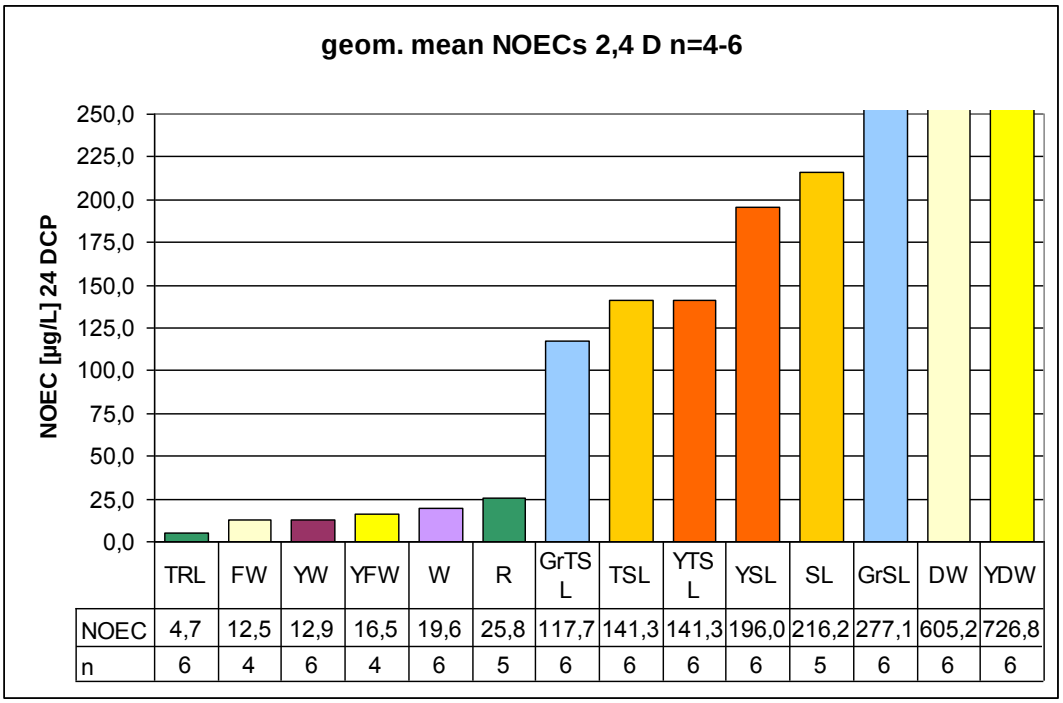
Fig. 45 Laboratory-specific NOEC for the variable growth rate main shoot length under exposure to 2,4-D (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbols: Tests invalid due to lack of doubling of shoot length.

The NOEC values from the invalid test L12 (control replicates apparently contaminated with foreign organisms) are highly consistent with those from valid tests.
In contrast, the tests classified as invalid due to reduced growth show a reduced sensitivity to 2,4-D for the variables related to shoot length and weight increase.
This might be attributed to a reduced effect of the growth-impairing action of the test substance under poor growth conditions. It would be an indication of the importance of the shoot length doubling as an efficient validity criterion. For a definite statement, however, onward studies are required.

Unlike the EC_{50} values, the NOEC values derived from original values and yield values naturally do not show any systematic differences. The mean NOEC values from the different variables on exposure to 2,4-D show extremely wide ranges of variation (Fig. 46).
There is a group of variables that prove to be highly sensitive (NOEC between 4.7 and 25.8 µg/L). These include the total root length, fresh weight and yield fresh weight, number of whorls and yield whorl number as well as number of roots.
A second group of variables show a marked leap upwards in sensitivity (mean NOEC between 118 and 727 µg/L). These include all shoot length parameters plus growth rates and the dry weight variables.

Within the respective groups, the difference in sensitivity corresponds to a factor of 5-6. As referred to all variables, the difference in sensitivity corresponds to a factor of 154. The most sensitive variable is the total root length (NOEC = 4.7 µg/L), followed by the fresh weight parameters (NOEC 12.5 and 16.5 µg/L), the number of whorls and the yield whorl number (NOEC 12.9 and 19.6 µg/L, respectively) and the number of roots (NOEC 25.8 µg/L). The least sensitive variables are represented by the growth rate main shoot length (NOEC = 277 µg/L) and the dry weight parameters (NOEC = 605 and 727 µg/L).

Thus, the variables show, on principle, a comparable sensitivity regarding the EC_{50} and NOEC. The fresh weight, for which no determination of an EC_{50} was possible (with one exception), shows a rather low NOEC. Consequently, it has proved to be a definitely sensitive endpoint (however, with limited reproducibility in the present inter-laboratory comparison, see below). In contrast, the dry weight is the least sensitive parameter. Hence, the dry weight and the fresh weight are anticorrelated parameters in this case: The inhibition of dry weight increase is less pronounced than that of fresh weight increase. In other words, with increasing concentrations, there is an excess proportion of plant parts that contain less water than usual (which is later extracted during the drying process). Possibly, the rudimentary roots formed at medium concentrations are a reason for this result. Also an invisible thickening of cell walls or similar processes could provide an explanation.



	TRL	FW	YW	YFW	W	R	GrTSL	TSL	YTSL	YSL	SL	GrSL	DW	YDW
NOEC	4,7	12,5	12,9	16,5	19,6	25,8	117,7	141,3	141,3	196,0	216,2	277,1	605,2	726,8
n	6	4	6	4	6	5	6	6	6	6	5	6	6	6

Fig. 46 Sensitivity of the different variables to 2,4-D: Mean NOEC (based on 4-6 laboratory-specific NOEC values)

In analogy to the evaluation of the tests with 3,5-DCP, Table 11 provides a summarizing view of the quality of the determined NOEC for the tests with 2,4-D, based on the %MDD and the EC_{20} (evaluated for the yields and growth rates only).

The mean NOEC either corresponds approximately to the mean EC_{20} (total root length) or is lower (number of roots, yield shoot length, yield whorl number, growth rates). For the weight variables, a comparison is thwarted by the fact that no EC_{20} could be evaluated. Only the yield total shoot length results in a mean NOEC which is twice as high as the mean EC_{20}, and hence, cannot stand up to critical scrutiny of its statistical power.

When comparing the average %MDDs actually obtained with those theoretically calculated it becomes clear that (unlike in the case of 3,5-DCP) for some variables, the MDDs obtained are markedly higher than expected (a phenomenon particularly pronounced for the yields dry weight and fresh weight and the number of roots). This has probably to be attributed to the special mode of action of 2,4-D discussed above. As a result, replicates could in part not be evaluated or showed higher dispersion, i.e. n was in fact lower and/or the variance higher than theoretically assumed.

Tab. 11 Comparison of the NOEC and EC20 for 2,4-D; frequency of Williams' test as a multiple test for NOEC determination (including invalid test runs), and resulting mean %MDD (minimum detectable difference). Evaluation for yields and growth rates.
Entered for comparison (last column): %MDD theoretically calculated, based on the test design and the mean coefficients of variation of the controls (cf. Table 5).

Variables	NOEC [mg/L] geom. mean n=2-6	EC_{20} [mg/L] geom. mean n=2-6	Williams test x times of 10	mean MDD [%] Williams-Test	theoretically calculated %MDD
R	25.8	155	8	36.1	*27.4*
TRL	4.7	6	8	31.9	*26.7*
YSL	196	238	6	24.8	*17.5*
YFW	16.5	EC_{20} = n.a. EC_{50} = 97	9	41.3	*24.5*
YDW	727	EC_{20} = n.a. EC_{50} = n.a.	8	34.6	*23.1*
YTSL	141	73	8	29.1	*23.7*
YW	12.9	27	8	14.0	*17.9*
GrSL	277	308	9	16.1	*14.0*
GrTSL	118	181	9	17.1	*15.1*

Again, the 95% prediction interval of the mean values is used as a measure of reproducibility of the calculated NOEC (Fig. 47). The narrowest absolute prediction intervals are observed for the total root length, whorl parameters and main shoot length / yield main shoot length. In absolute terms, these variables render the NOEC values characterized by the highest reproducibility.

The relationship between the width of the prediction interval and mean NOEC is considered as a measure of the relative quality of the prediction interval (Table 12). While the NOEC values show an acceptable level of reproducibility for the shoot length variable and total root length (ratios of 1.7 to 5.2), the mean NOEC values for the variables whorl number and root number are clearly less reproducible (ratios of 13-45). For the fresh weight variables, the mean NOEC obtained is virtually insignificant in view of the width of the prediction intervals (ratios of 656 and 807).

The latter is a direct consequence of the specific action of 2,4-D: Promoting effects and deformations of the experimental plants increase the variability. As a result affecting the present inter-laboratory comparison, the test of L04 could only determine a NOEC “larger than highest concentration“, i.e. larger than 2000 µg/L, while three other laboratories determined maximum NOEC values of 9.6 µg/L.

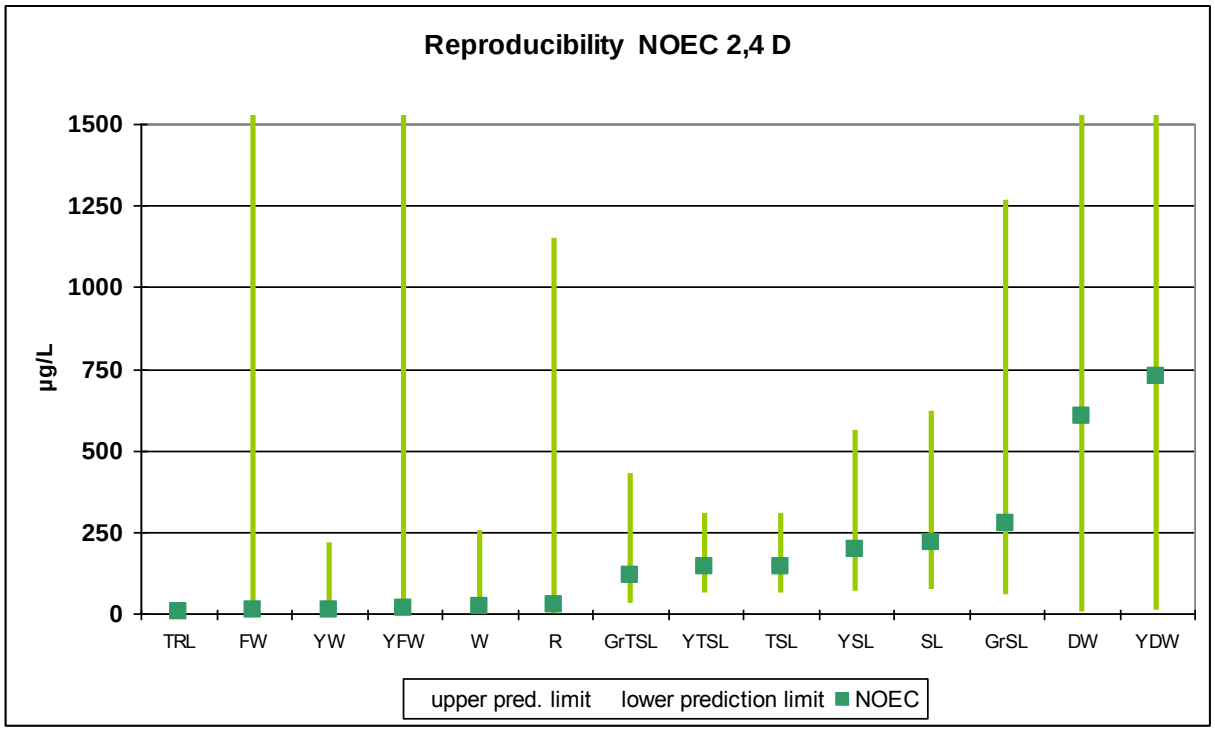


Fig. 47 Mean NOEC of the different variables on exposure to 2,4-D and their respective 95% prediction intervals as a measure of reproducibility (based on 4 - 6 laboratory-specific NOEC values each, cf. Fig. 46).

Tab. 12 Relationship between width of 95% prediction interval and mean NOEC on exposure to 2;4-D (ratio). Number of underlying single values as in Fig. 46.

	TSL	YTSL	YSL	SL	GrTSL	GrSL	TRL	W	YW	R	YDW	DW	YFW	FW
Ratio	1.7	1.7	2.5	2.5	3.4	4.4	5.2	13	17	45	66	159	656	807

4.6 Results for IP

4.6.1. IP: Mode of action and data used

Isoproturon acts as a photosynthesis inhibitor. The experimental plants react with elongation growth (Fig. 49), i.e. the length of internodes increases compared with the unimpaired control plants (Fig. 48).

Fig. 48 Unimpaired control plant on day 14

Fig. 49 Plant in highest IP concentration (2000 µg/L) on day 14

As a result, because of a missing dose-response relationship or low inhibition rates, variables referring to the shoot length either cannot be evaluated at all (main shoot length, total shoot length) or evaluation is only possible for a few tests (yield main shoot length and yield total shoot length, growth rates).

Altogether, eight of the participating laboratories performed a test with isoproturon. Six out of the eight data sets submitted fulfilled the validity criteria. In one test, the control replicates were apparently contaminated with foreign organisms. In another one, the criterion of shoot length doubling failed.
In the remaining six data sets, no significant dose-response relationships could be achieved for the variables, main shoot length and total shoot length due to the effect of the test substance as described above so that no EC_{50} values could be determined in these cases. For the variables, yield main shoot length and yield total shoot length, EC_{50} determination succeeded only in a single and two cases, respectively, for the growth rate, in one case only. For these variables, the data available are insufficient for a reliable statement. The number of underlying data sets has been stated in the results presented below.

4.6.2. IP: EC50 – Sensitivity of variables – Reproducibility

Figs. 50 – 53 show the EC_{50} values and 95% confidence intervals obtained by the different laboratories for selected variables including the respective overall mean value with its 95% prediction interval.[9] Since for the shoot length variables and the growth rates shown for 3,5-DCP and 2,4-D in this position, only one or two EC_{50} values each are available, the variables, yield whorl number and number of roots are shown instead below.

In Fig. 54, the EC_{50} for all variables are listed by sensitivity.
Fig. 55 and Table 13 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the EC_{50} values.

All of the valid EC_{50} values obtained by the different laboratories are within the 95% prediction interval, however, with a high uncertainty in some single cases (wide 95% confidence intervals).

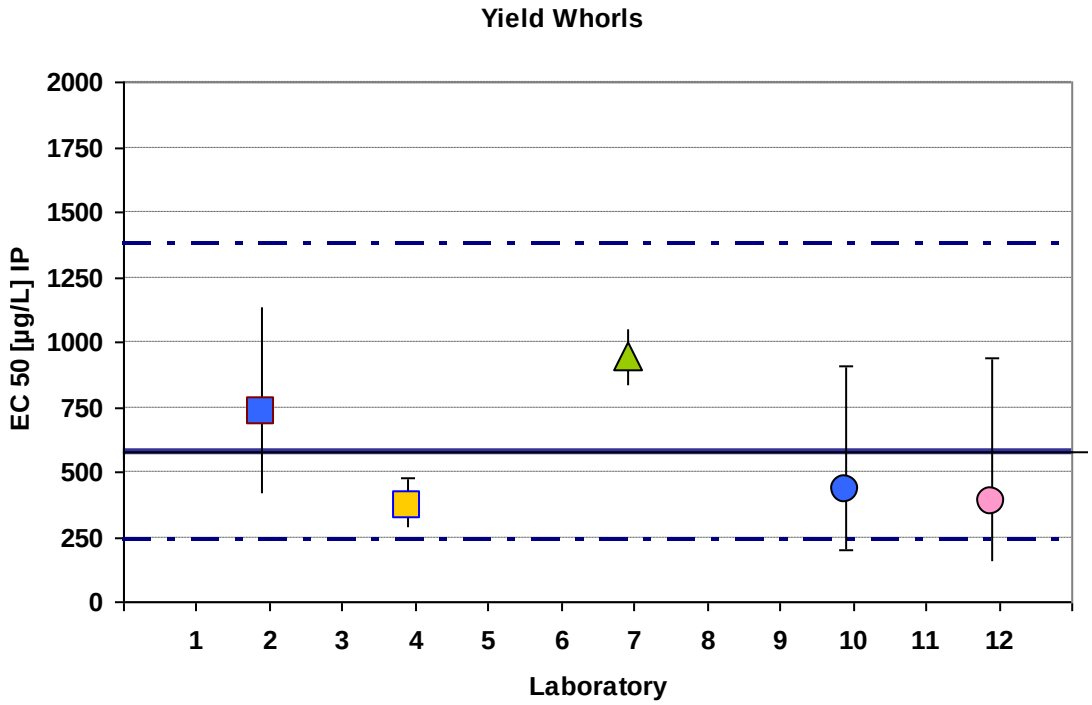


Fig. 50 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, yield whorl number under exposure to IP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

[9] The corresponding Figures for *all* variables are found in file „IP EC50 fig_reprod labs + geom mean.xls“.

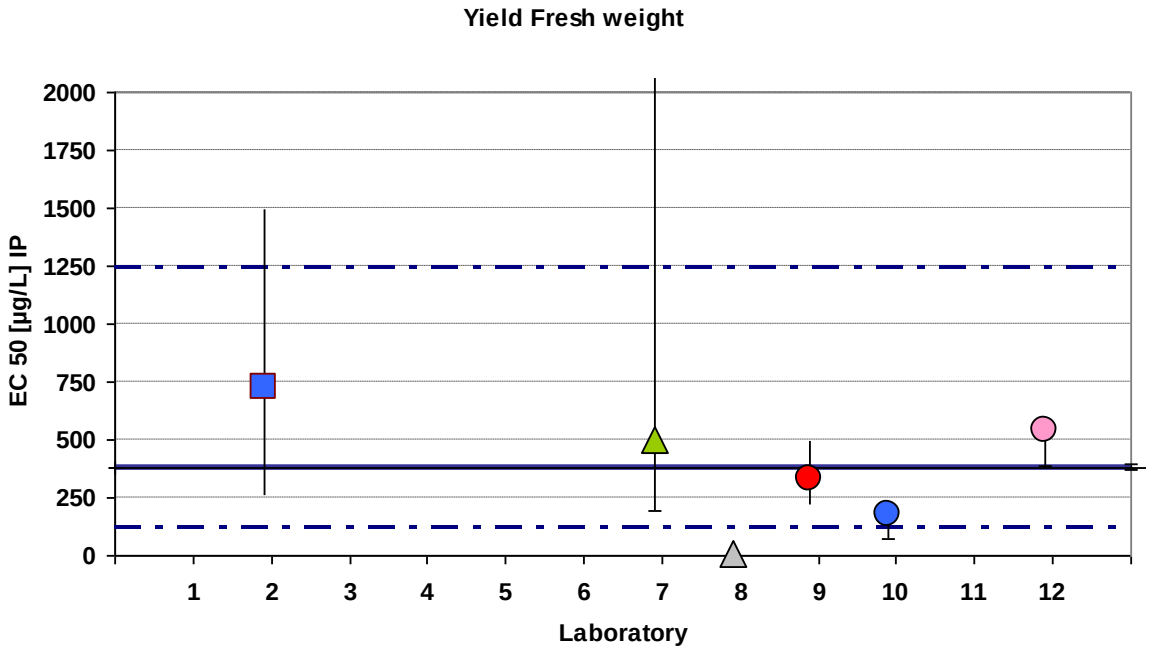


Fig. 51 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, yield fresh weight under exposure to IP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbol: Test invalid due to lack of doubling of shoot length (EC_{50} = 9.2 µg/L).

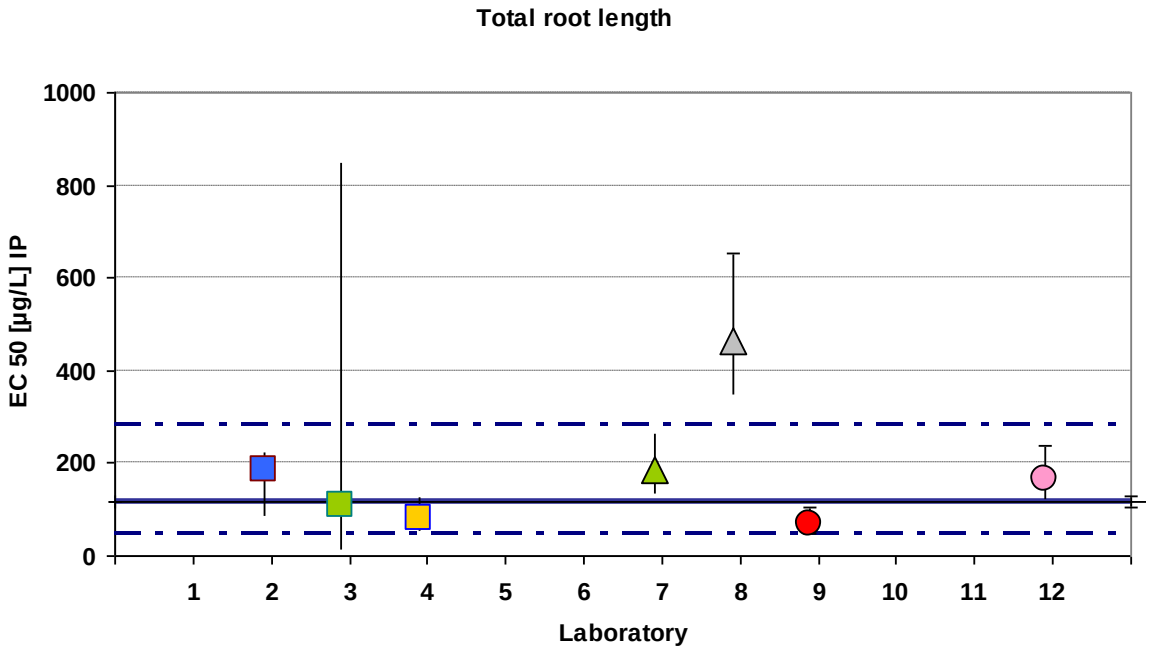


Fig. 52 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, total root length under exposure to IP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbol: Invalid due to lack of doubling of shoot length.

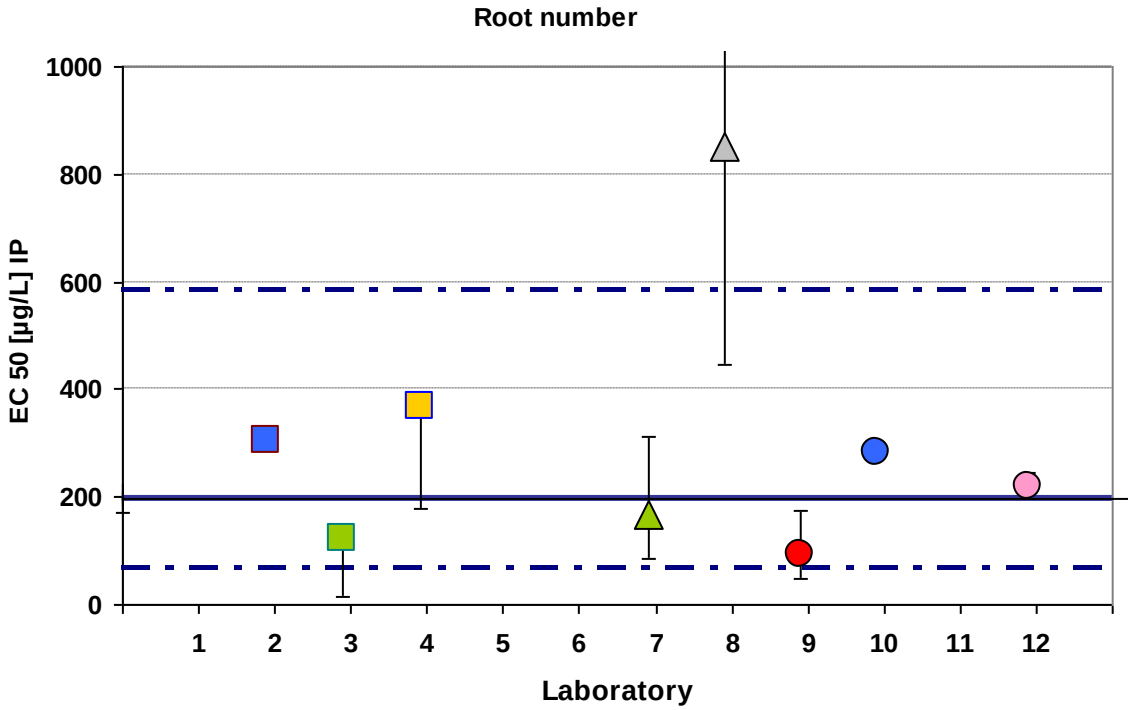
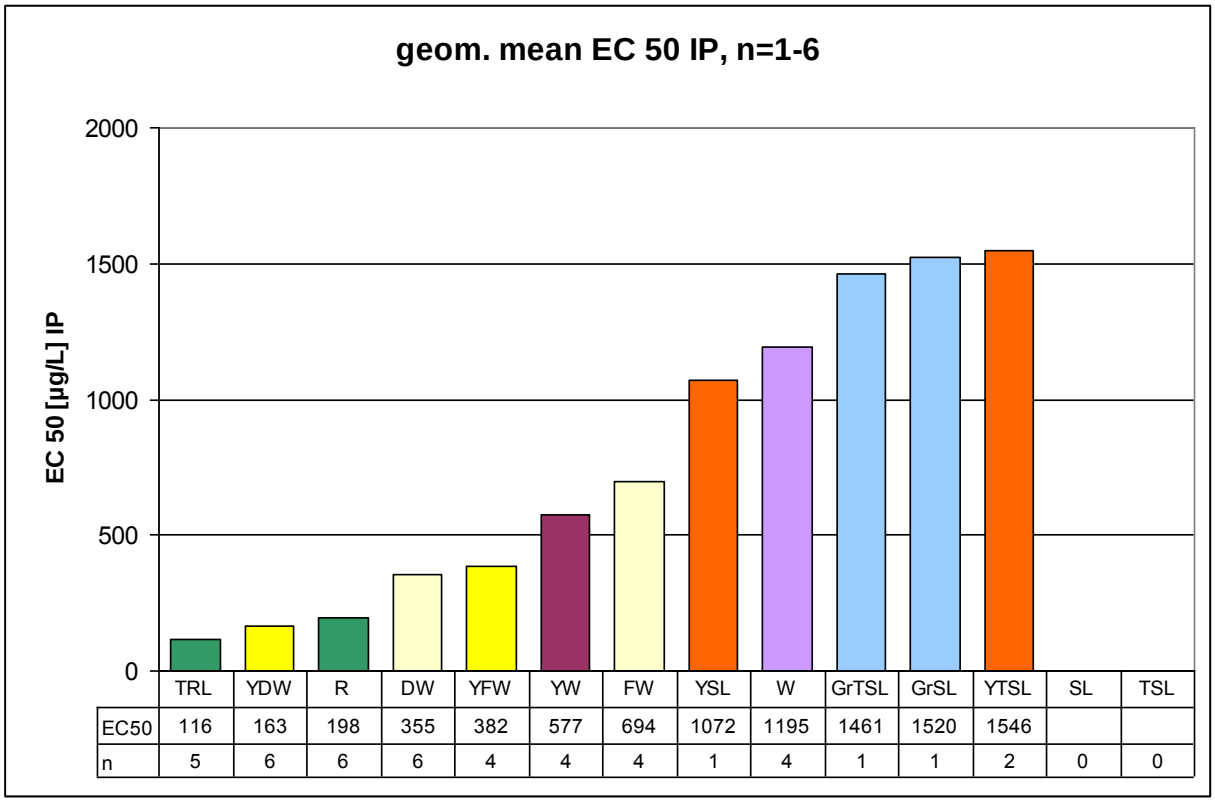


Fig. 53 Laboratory-specific EC_{50} with 95% confidence intervals for the variable, number of roots under exposure to IP (coloured symbols); solid and broken lines: Mean EC_{50} with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbol: Invalid due to lack of doubling of shoot length.

The EC_{50} values from the invalid test (L12, control replicates apparently contaminated with foreign organisms) are consistent with those from valid tests. As a rule, the EC_{50} values from the tests with reduced growth are not within the 95% prediction interval of the mean EC_{50}, with deviations occurring both upwards and downwards.

Sensitivity and reproducibility of EC_{50}

The mean EC_{50} values of the evaluable variables on exposure to IP are between 116 and 1546 µg/L (Fig. 54). In analogy to the action of 3,5-DCP and 2,4-D, the EC_{50} values for the original values obtained by measurement are about twice as high as those for the yield values calculated from these. The difference between the most sensitive and the least sensitive variable corresponds to a factor of ca. 13.
Thus the variables evaluated show more pronounced differences in sensitivity to IP than in sensitivity to 3,5 DCP. They are, however, less pronounced than that in sensitivity to 2,4-D.
The variable being by far the most sensitive also on exposure to IP is total root length (EC_{50} = 116 µg/L), followed by yield dry weight (EC_{50} = 163 µg/L), root number (EC_{50}= 198 µg/L) and yield fresh weight (EC_{50} = 382 µ/L). The shoot length parameters show, on principle, higher EC_{50} values: namely 1072 µg/L (yield main shoot length) to 1546 µg/L (yield total shoot length) than the other variables, corresponding to a lower sensitivity. This can presumably be attributed to a masking effect of the elongation growth induced by isoproturon. This is confirmed by comparison with the considerably lower EC_{50} of the yield whorl number (EC_{50} = 577 µg/L).



	TRL	YDW	R	DW	YFW	YW	FW	YSL	W	GrTSL	GrSL	YTSL	SL	TSL
EC50	116	163	198	355	382	577	694	1072	1195	1461	1520	1546		
n	5	6	6	6	4	4	4	1	4	1	1	2	0	0

Fig. 54 Sensitivity of the different variables to IP: Mean EC_{50} (stating the number of underlying laboratory-specific EC_{50})

The 95% prediction interval of the mean values is used as a measure of reproducibility of the measured EC_{50} (Fig. 55). The narrower the prediction interval, the higher is the reproducibility. The narrowest absolute prediction intervals are shown by those variables which at the same time represent the most sensitive ones, namely the total root length and number of roots. However, the 95% prediction intervals for the respective mean values are narrow for the remaining variables as well. Only the 95% prediction interval for the yield total shoot length is very wide, but based on no more than two EC_{50} values. Moreover, this impression is put into perspective in view of the absolute EC_{50} level (cf. ratio, see below). For the growth rates and the other shoot length parameters, no prediction intervals can be evaluated because no EC_{50}, or only one value each, could be determined.

The relationship between the width of the prediction interval and the EC_{50} value is considered as a measure of the relative quality of the prediction interval (Table 13). Showing values between 1.5 and 2.9, this ratio mostly resembles the low one obtained under exposure to 3,5-DCP. Only the EC_{50} for the yield dry weight shows a lower but nevertheless acceptable level of reproducibility (ratio 5.1).

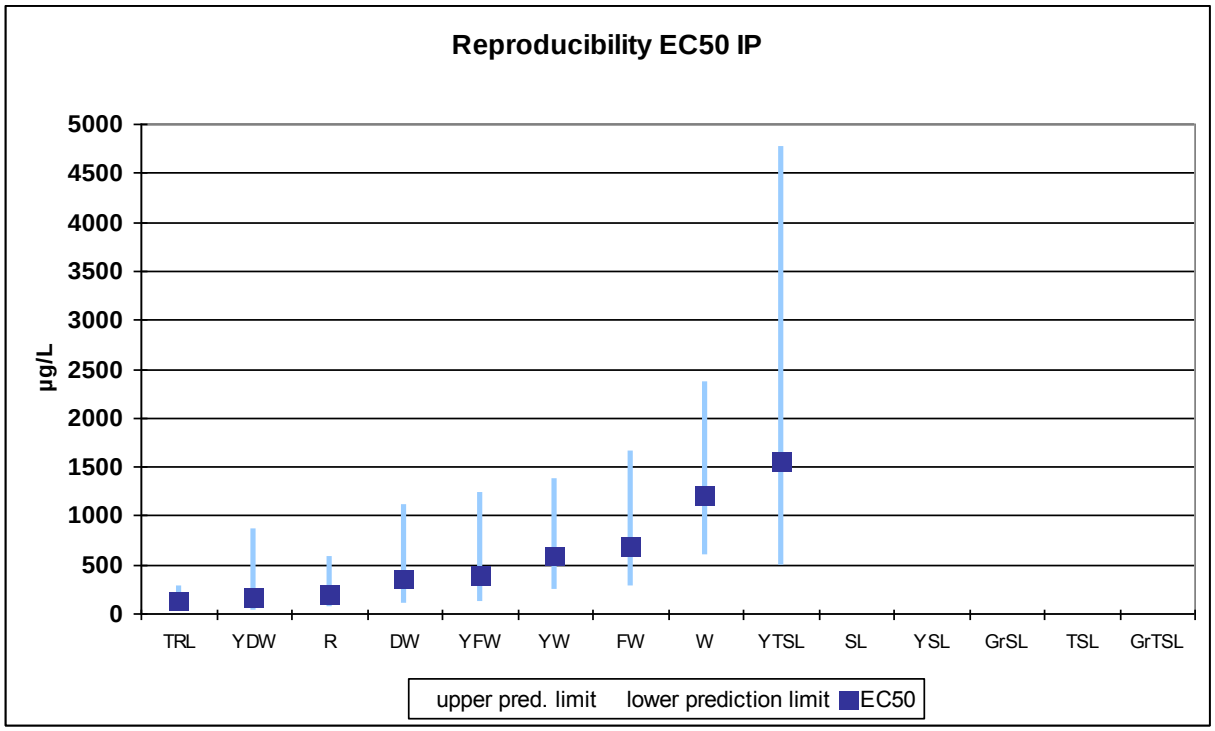


Fig. 55 Mean EC_{50} values of the different variables on exposure to IP and their respective 95% prediction intervals as a measure of reproducibility (number of underlying laboratory-specific EC_{50} as in Fig. 54).

Tab. 13 Relationship between width of 95% prediction interval and mean EC_{50} on exposure to IP (ratio). n.e. = not evaluable. Number of underlying EC_{50} values as in Fig. 54.

	W	YW	FW	TRL	R	YTSL	DW	YFW	YDW	SL	YSL	GrSL	TSL	GrTSL
Ratio	1.5	2.0	2.0	2.0	2.6	2.8	2.8	2.9	5.1	n.a.	n.a.	n.a.	n.a.	n.a.

4.6.3. IP: NOEC – Sensitivity of variables – Reproducibility

Figs. 56 – 59 show the NOEC results for selected variables including the respective overall mean value with its 95% prediction interval.[10]
In Fig. 60, the NOEC values for all variables are listed by sensitivity. The statistical power of the NOEC values is examined based on the minimum detectable difference (%MDD) and by comparison with the EC_{20} values (Table 14).
Fig. 61 and Table 15 show the respective absolute and relative 95% prediction intervals as a measure of reproducibility of the NOEC.

The NOEC on exposure to isoproturon determined by the different laboratories are scattered over three or four concentration steps. For the shoot length parameters (for example Figs 56 and 59), the NOEC is often determined as "larger than highest concentration" (depicted as „2000µg/L", marked accordingly in the result tables). This leads to asymmetric 95% prediction intervals with high upper limits for the mean NOEC.

The NOEC obtained from invalid tests do not show any deviations against the valid tests.

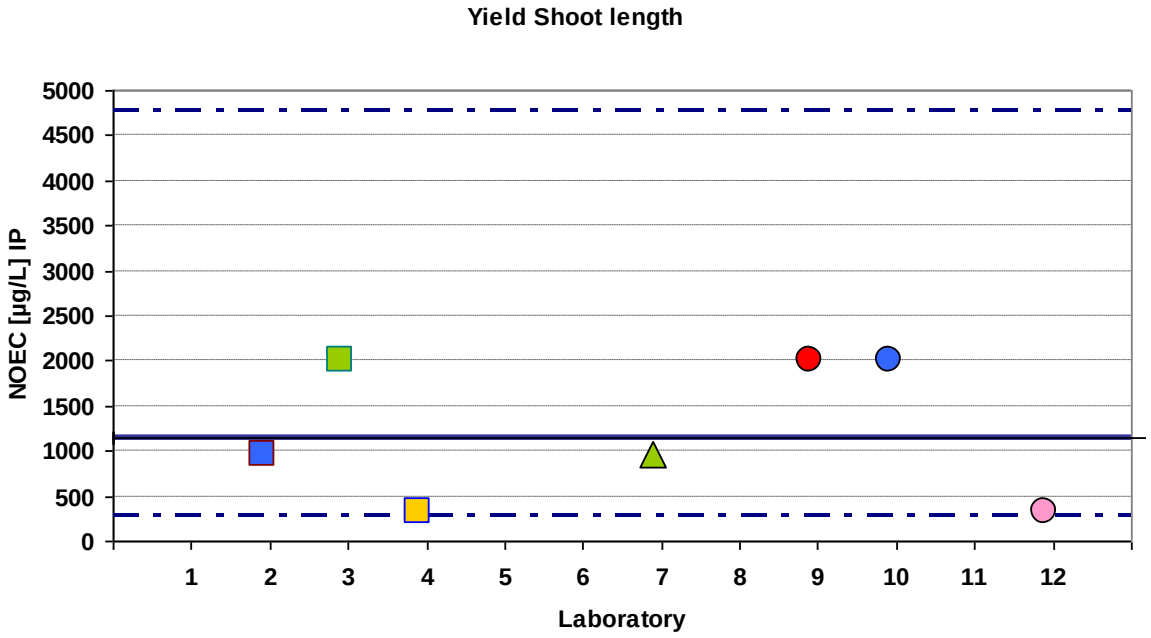


Fig. 56 Laboratory-specific NOEC for the variable yield main shoot length under exposure to IP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

[10] The corresponding Figures for *all* variables are found in file „IP NOEC fig_reprod labs + geom mean.xls".

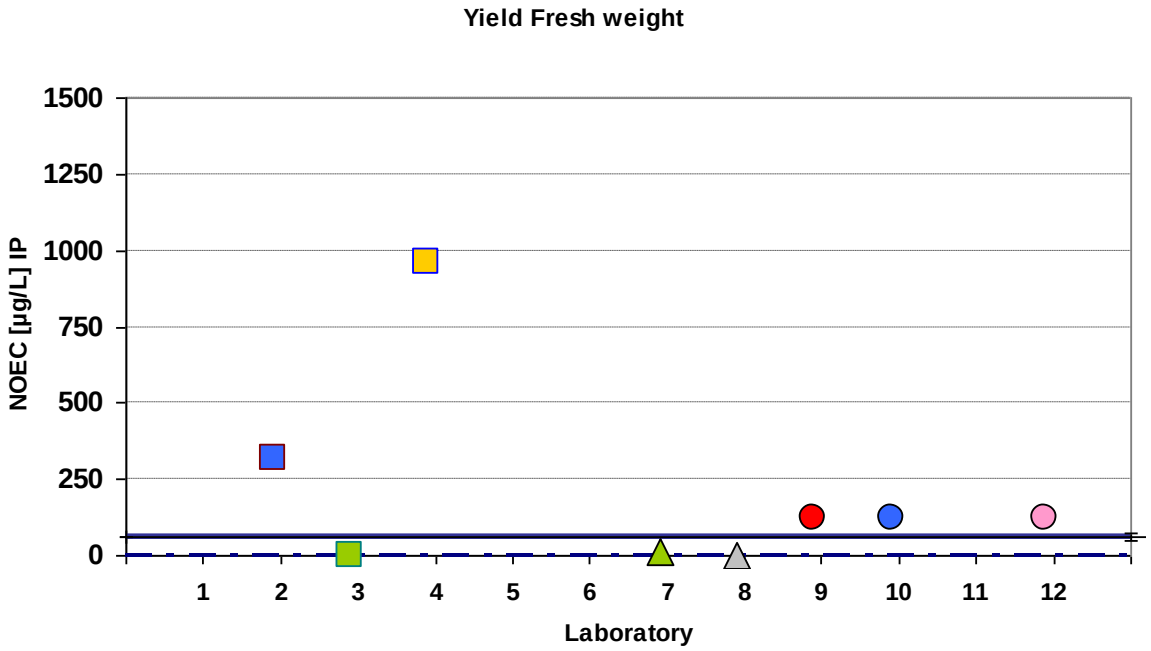


Fig. 57 Laboratory-specific NOEC for the variable yield fresh weight under exposure to IP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval (upper limit not shown because value = 7400); pink symbol: Test invalid due to apparent contamination; grey symbol: Test invalid due to lack of doubling of shoot length.

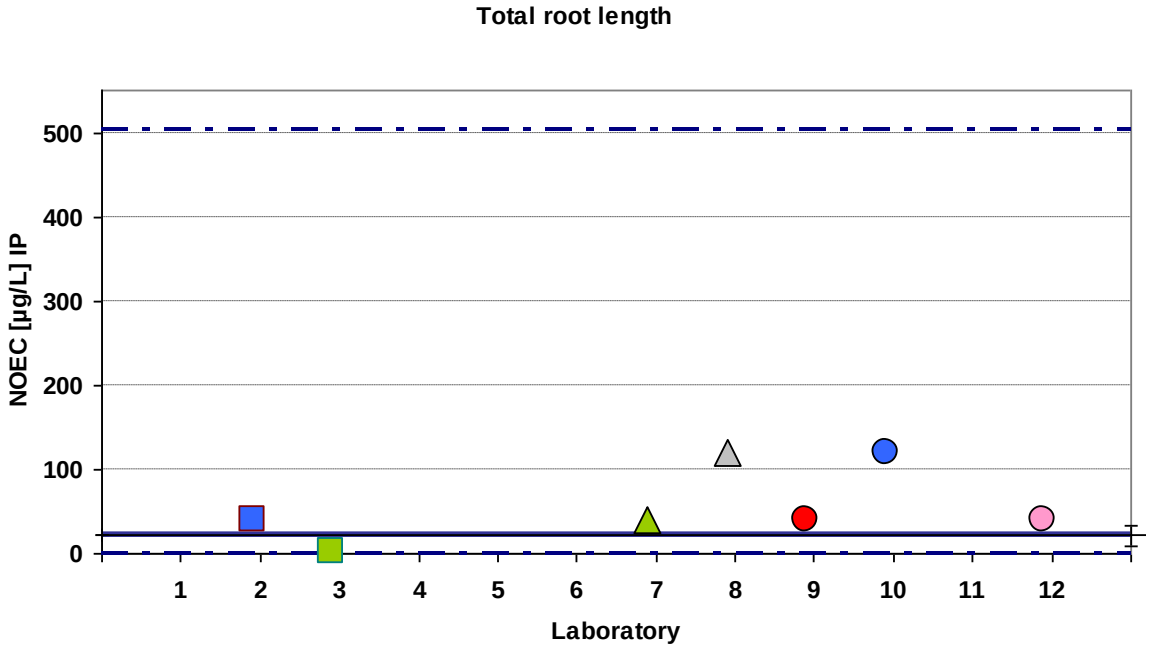


Fig. 58 Laboratory-specific NOEC values for the variable total root length under exposure to IP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination; grey symbol: Tests invalid due to lack of doubling of shoot length.

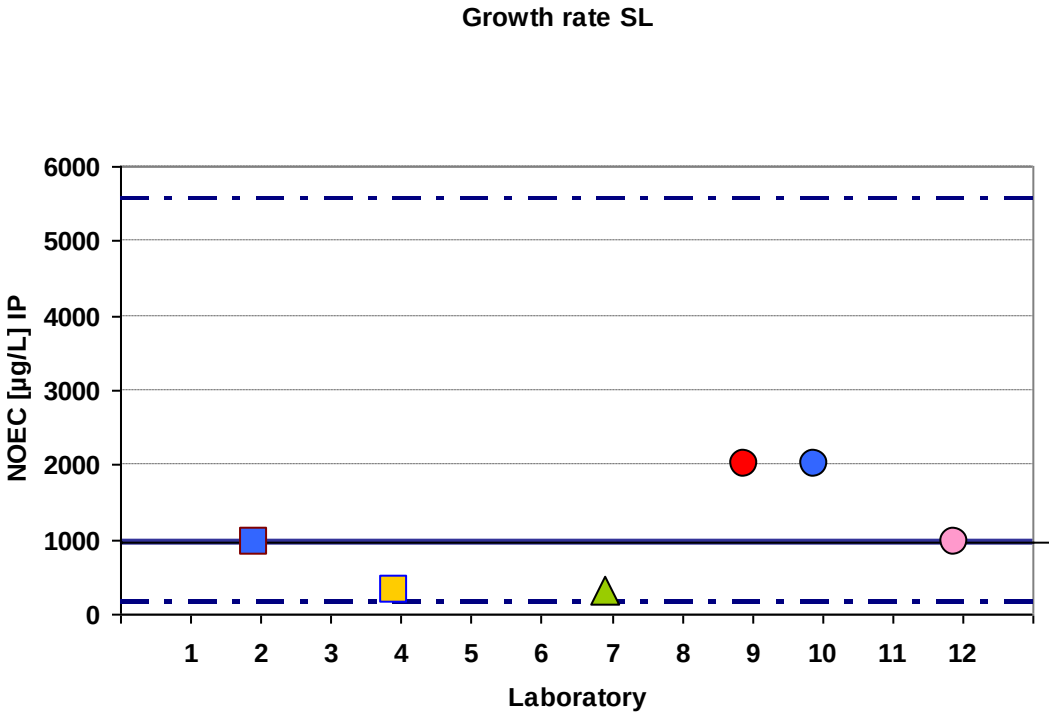


Fig. 59 Laboratory-specific NOEC for the variable, growth rate main shoot length under exposure to IP (coloured symbols); solid and broken lines: Mean NOEC with 95% prediction interval; pink symbol: Test invalid due to apparent contamination.

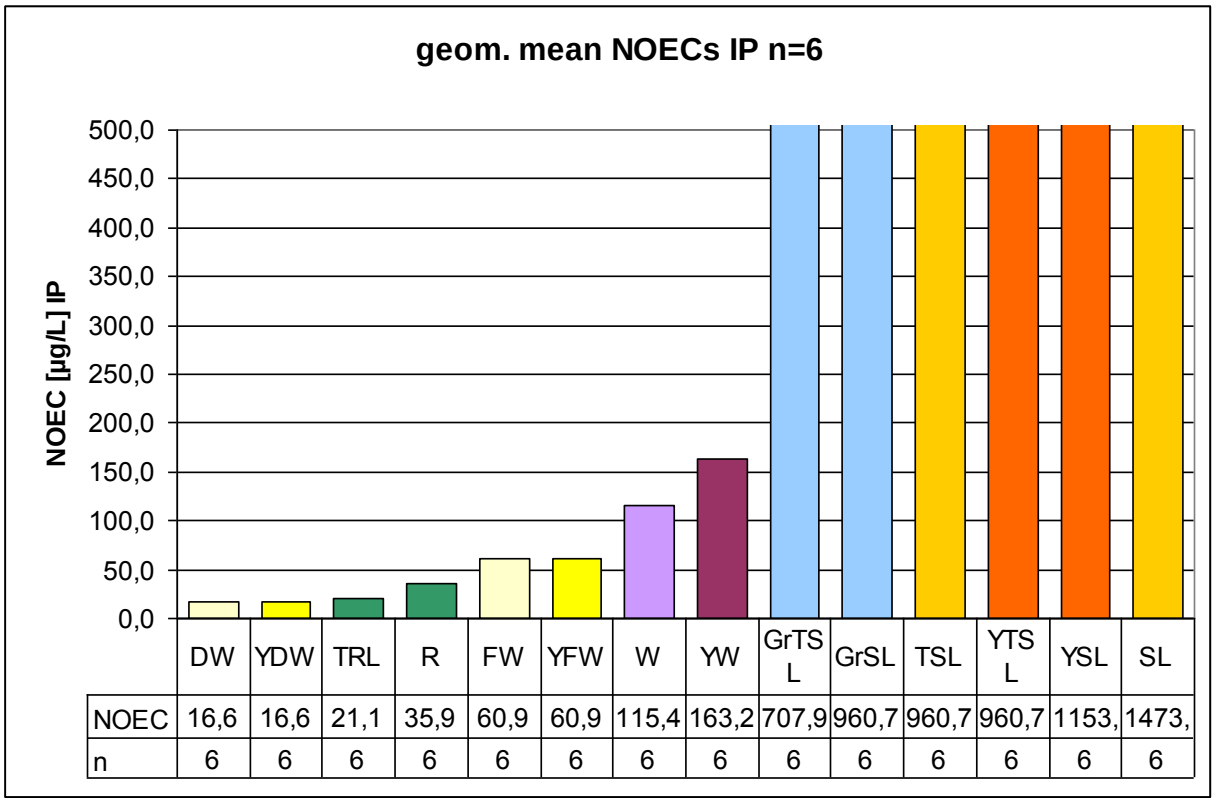
Sensitivity and reproducibility of NOEC

As far as the mean NOEC values on exposure to IP are concerned, two groups of variables can be distinguished:

In the concentration range tested, all shoot length parameters show a low sensitivity with mean NOEC values of 708 µg/L to 1470 µg/L. These concentrations have to be interpreted as lower limits because several findings of „NOEC > highest concentration“ are included, which have numerically been taken into account by entering a figure of 2000 µg/L. Hence, the mean NOEC for the shoot length parameters are in fact even higher.

In contrast, the remaining variables show a high sensitivity with mean NOEC values between 16.6 µg/L (yield dry weight) and 163 µg/L (yield number of whorls). The difference in sensitivity within this group corresponds to a factor of 9.8. The variables most sensitive also to isoproturon include the number of roots and the total root length, in addition to the weight parameters.

The order of mean NOEC values confirms the sensitivity findings based on the EC_{50}.



	DW	YDW	TRL	R	FW	YFW	W	YW	GrTSL	GrSL	TSL	YTSL	YSL	SL
NOEC	16,6	16,6	21,1	35,9	60,9	60,9	115,4	163,2	707,9	960,7	960,7	960,7	1153,	1473,
n	6	6	6	6	6	6	6	6	6	6	6	6	6	6

Fig. 60 Sensitivity of the different variables to IP: Mean NOEC (based on 6 laboratory-specific NOEC values each)

By means of a comparison NOEC-EC_{20} and analysis of the %MDD, Table 14 provides an overview of the quality of the NOEC values established (evaluated only for yields and growth rates).

As a rule, the mean NOEC values approximately correspond to the mean EC_{20} or are lower than these. Only for the growth rate shoot length, the mean NOEC is higher than the EC_{20}. In view of the fact that in general, the evaluation of shoot length parameters on exposure to isoproturon is difficult, this may be due to a single additional test result "greater than highest concentration", i.e. should not be overestimated.

As a rule, the mean %MDD values actually obtained correspond to those theoretically expected. Only the yield shoot length shows lower real %MDD values. This could be caused by the variability of the yield shoot length, which due to the special mode of action of isoproturon is higher than theoretically expected.

Tab. 14 Comparison of the NOEC and EC20 for IP; Frequency of Williams' test as a multiple test for NOEC determination (including invalid tests), and resulting mean %MDD (minimum detectable difference). Evaluation for yields and growth rates.
Entered for comparison (last column): %MDD values theoretically calculated, based on the test design and the mean coefficients of variation of controls (cf. Table 5).

Variables	NOEC [mg/L] geom. mean n=5-6	EC_{20} [mg/L] geom. mean n=2-6	Williams test x times of 8	mean MDD [%] Williams-Test	theoretically calculated %MDD
R	35.9	39.6	6	22	*27.4*
TRL	21.1	34	7	28	*26.7*
YSL	1153	(n=1)	7	30	*17.5*
YFW	60.9	98.5	7	23	*24.5*
YDW	16.6	23.2	8	22	*23.1*
YTSL	462	799	5	23	*23.7*
YW	163	294	7	24	*17.9*
GrSL	960	663	5	17	*14.0*
GrTSL	708	720	6	15	*15.1*

Fig. 61 provides information on the reproducibility of NOEC values, based on the 95% prediction intervals of mean NOEC values. The narrower the prediction interval, the higher is the reproducibility.
The narrowest absolute prediction intervals are shown by the variables rendering the lowest NOEC values, namely total root length and dry weight parameters. The highest uncertainty is seen for the NOEC values for the fresh weight and shoot length parameters. In relative terms, the prediction intervals of the NOEC on exposure to isoproturon are higher than on exposure to the other test substances for all variables (lowest ratios: 3.9 to 6.1; highest ratios: 23.5 to 122.5).

The ratios for the 95% prediction intervals of the NOEC on exposure to isoproturon (at least 3.9) are generally higher than those obtained for the EC_{50} and also higher than those for the NOEC on exposure to other test substances.
Strikingly, the variables, total root length and number of roots, which before always showed narrow 95% prediction intervals for the mean NOEC *and* low ratios of prediction value/mean NOEC, result in less well reproducible NOEC values in this case.

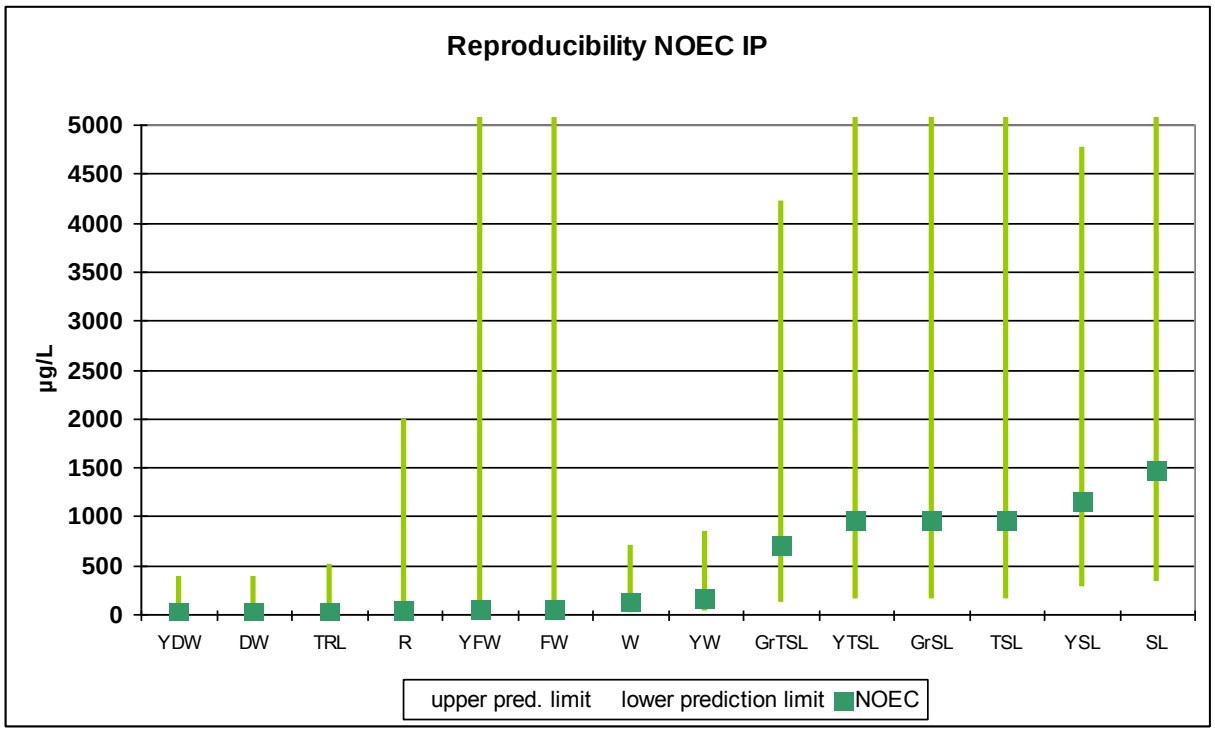


Fig. 61 Mean NOEC of the different variables on exposure to IP and their respective 95% prediction intervals as a measure of reproducibility (based on 5 - 6 laboratory-specific NOEC values each, cf. Fig. 60).

Tab. 15 Relationship between width of 95% prediction interval and mean NOEC on exposure to IP (ratio).

	YSL	SL	YW	GrSL	TSL	YTSL	GrTSL	W	DW	YDW	TRL	R	FW	YFW
Ratio	3.9	4.1	5.0	5.6	5.6	5.6	5.8	6.1	23.5	23.5	24.1	55.3	122.5	*122.5*

4.7 Comparison of the toxicity of all test substances

After the presentation of the individual results for each substance tested, the toxicity of the three test substances is now compared based on the respective EC_{50} and NOEC values (Figs. 62 and 63).

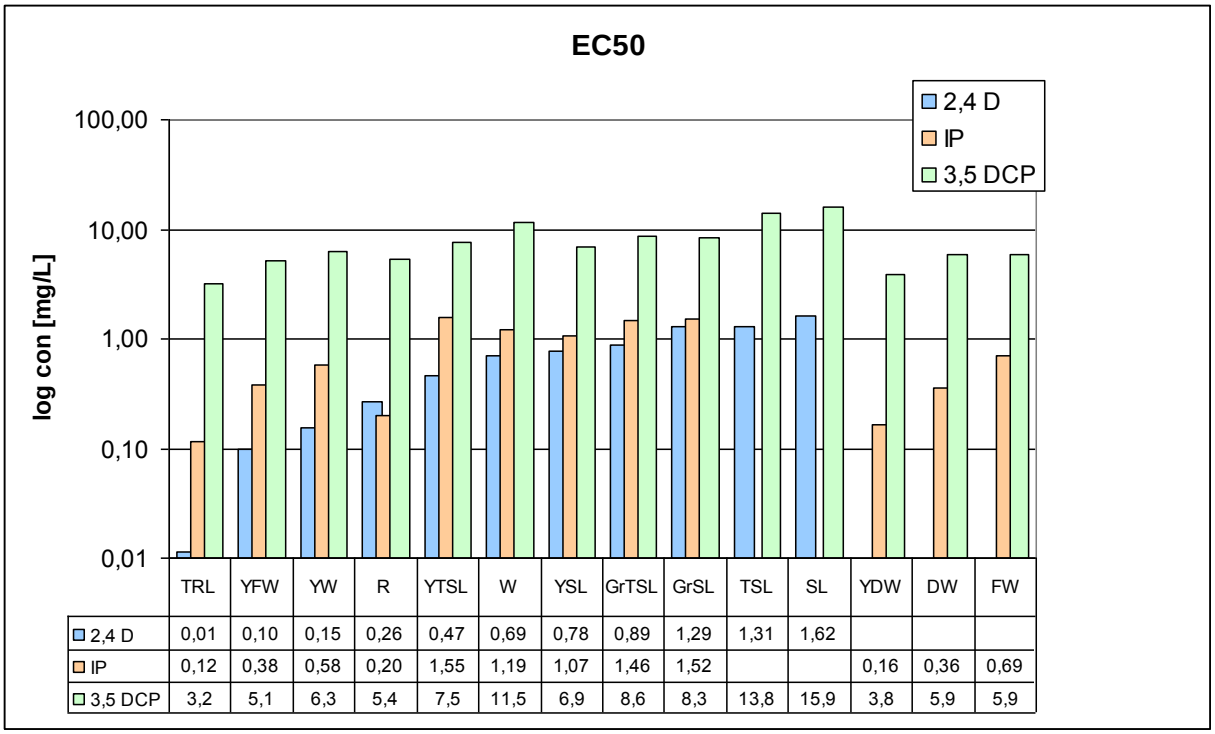
By definition, the yields will, on principle, result in higher inhibition values and therefore, lower EC_{50} values than the underlying original measurements. The NOEC values are not affected by this phenomenon.
EC_{50} and NOEC are correlated, on principle, except that, as a rule, unlike the respective EC_{50} values, the NOEC values for yields and original measurements show identical results.
On exposure to 2,4-D, often no EC_{50} can be determined for the weight parameters due to specific action of the substance, and especially regarding the dry weight, the NOEC is often greater than the highest test concentration. In analogy, this applies to the shoot length parameters on exposure to isoproturon.
In contrast exposure to 3,5-DCP induces obvious dose-related inhibition effects for all variables and results, on principle, in clearly defined test concentrations representing the NOEC values.

2,4-D is by far the substance having the highest toxic effect. It is followed by isoproturon ranking in the second position and 3,5-DCP, in the third.

The order of sensitivity of the variables is almost independent of the substance tested. The variables identified as the principally most sensitive ones include: total root length, yield fresh weight, yield whorl number and number of roots. The yield growth rate of shoot length occupy a moderate to rather non-sensitive position.
Under exposure to 3,5-DCP and IP, also the yield dry weight proves to be a very sensitive variable. In the presence of a growth-inhibiting test substance such as 2,4-D, however, specific effects on the experimental plants are observed which result in the dry weight variable to react in a very non-sensitive way.
The most sensitive variables also show the best reproducibility values (Table 16).

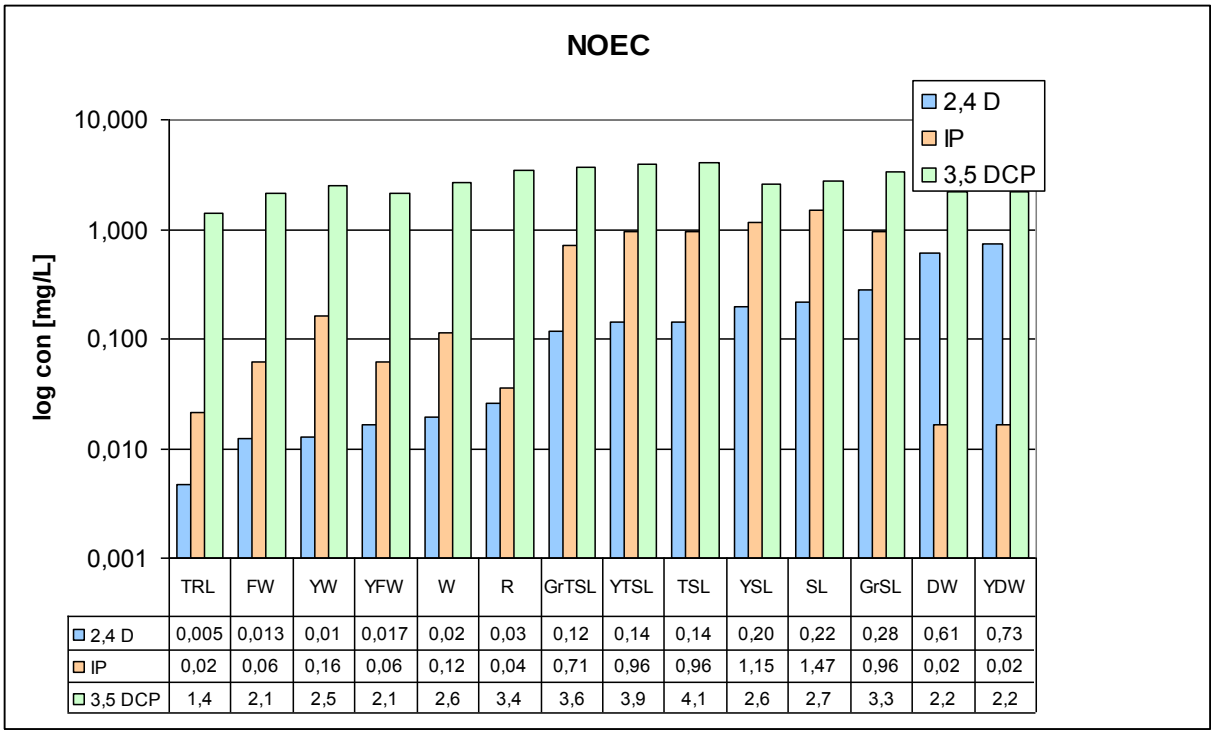
The more sensitive the variable, the greater is the difference in toxicity between the three test substances:
The mean EC_{50} of the total root length being the most sensitive parameter is for 2,4-D 12 times lower than for IP and 320 times lower than for 3,5-DCP. For the yield fresh weight being the second most sensitive parameter, the difference in toxicity between 2,4-D and IP corresponds to a factor of 3.8, that between 2,4-D and 3,5-DCP, to a factor of 50, and for the generally less sensitive variable, yield main shoot length, to factors of no more than 3.3 and 8.8, respectively.

Accordingly, the mean NOEC of the total shoot length as the most sensitive parameter is 280 times lower for 2,4-D than for 3,5-DCP. For the variables, fresh weight and yield fresh weight as the second most sensitive parameters, the difference in toxicity between 2,4-D and 3,5-DCP still corresponds to factors of 160 and 120, respectively, and for the generally less sensitive variable, yield main shoot length, no more than a factor of 13.



	TRL	YFW	YW	R	YTSL	W	YSL	GrTSL	GrSL	TSL	SL	YDW	DW	FW
2,4 D	0,01	0,10	0,15	0,26	0,47	0,69	0,78	0,89	1,29	1,31	1,62			
IP	0,12	0,38	0,58	0,20	1,55	1,19	1,07	1,46	1,52			0,16	0,36	0,69
3,5 DCP	3,2	5,1	6,3	5,4	7,5	11,5	6,9	8,6	8,3	13,8	15,9	3,8	5,9	5,9

Fig. 62 Comparison of the mean EC_{50} values of all variables evaluated for all three test substances. Variables listed by sensitivity to 2,4-D, which has shown the highest toxicity. Concentrations plotted logarithmically because EC_{50} values differ by several logs.



	TRL	FW	YW	YFW	W	R	GrTSL	YTSL	TSL	YSL	SL	GrSL	DW	YDW
2,4 D	0,005	0,013	0,01	0,017	0,02	0,03	0,12	0,14	0,14	0,20	0,22	0,28	0,61	0,73
IP	0,02	0,06	0,16	0,06	0,12	0,04	0,71	0,96	0,96	1,15	1,47	0,96	0,02	0,02
3,5 DCP	1,4	2,1	2,5	2,1	2,6	3,4	3,6	3,9	4,1	2,6	2,7	3,3	2,2	2,2

Fig. 63 Comparison of the mean NOEC values for all variables evaluated for all three test substances. Variables listed by sensitivity to 2.4-D, which shows the highest toxicity. Concentrations plotted logarithmically because NOEC different by several logs.

Tab. 16 Reproducibility of the mean EC_{50} and NOEC values on exposure to the three test substances. The table lists all variables whose mean EC_{50}, respectively NOEC values show the narrowest absolute 95% prediction interval.

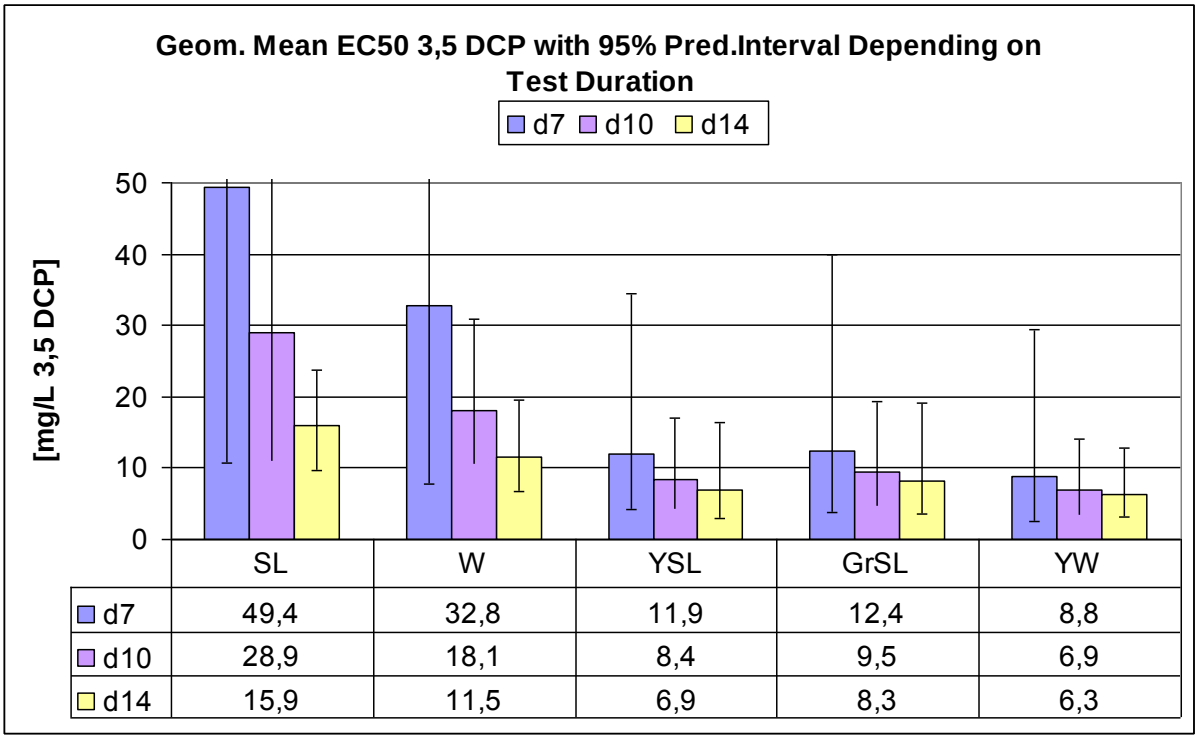
	EC_{50}		NOEC	
	narrowest 95% prediction interval	Factor prediction interval / EC_{50}	Narrowest 95% prediction interval	Factor prediction interval / EC_{50}
3,5-DCP	R, TRL YFW	1,3 2,4 1,1	R TRL	2,2 6,0
2,4-D	R, TRL YW	1,0 4,9 2,9	TRL TSL, YTSL SL, YSL	5,2 1,7 2,5
IP	R TRL	2,6 2,0	TRL DW YDW	24,1 23,5 23,5

4.8 Influence of test conditions on toxicity parameters

In order to decide on a possible modification of test conditions, the influence of test duration and temperature conditions on the results was examined.

4.8.1. Influence of the test duration

In the context of the inter-laboratory comparison, measuring of the main shoot length and whorl number was carried out on days 3, 7, 10 and 14. Thus, the endpoints for which the toxicity parameters can be compared depending on the test duration include main shoot length, yield main shoot length, number of whorls, yield number of whorls and growth rate main shoot length. The comparison was performed for the test substance, 3,5-DCP because for the latter, no effects on *Myriophyllum spicatum* were observed that would have impaired the evaluability of single variables.
Since a photograph was archived for each replicate and each measurement time, it would be possible to subsequently evaluate also the root number and total root length for shorter test periods. This was, however, omitted in the framework of the present report because the existing measurement data have already provided unequivocal results (Fig. 64).



	SL	W	YSL	GrSL	YW
d7	49,4	32,8	11,9	12,4	8,8
d10	28,9	18,1	8,4	9,5	6,9
d14	15,9	11,5	6,9	8,3	6,3

Fig. 64 EC_{50} on exposure to 3,5-DCP for selected variables as a function of the test duration. Depicted: geometric mean of EC_{50} from all valid tests and 95% prediction interval

Result of a comparison of the EC_{50} for days 7, 10 and 14: The longer the test period, the lower the EC_{50} and the better the reproducibility of the EC_{50} (i.e. the narrower the 95% confidence intervals). This effect is less pronounced for the yields and the growth rate than for the originally measured starting variables but nevertheless, can be clearly demonstrated. Thus, a shorter test duration is ruled out.

4.8.2. Influence of the temperature conditions

Originally, a circadian change in temperatures had been envisaged for the test with *Myriophyllum spicatum* (light phase (16h): 25°C; dark phase (8h): 20°C). In order to reduce the technological expenditure and ensure comparability of the conditions with other bioassays, the tests performed in the context of the inter-laboratory comparison took place under constant temperature conditions (23°C ± 2°C).

Below, it is examined whether such different temperature conditions have an influence on the resulting EC_{50} values. Fig. 65 shows the mean EC_{50} values on exposure to 3,5-DCP and their 95% prediction intervals from the present inter-laboratory comparison for all variables. The comparative values originate from a test with a temperature regimen which was conducted by the Ecotoxicology Laboratory of the Federal Environment Agency simultaneously with the tests for the inter-laboratory study. Owing to the vast experience of the UBA Ecotoxicology Laboratory with the test system, this test can be considered as representative.

The comparison shows the EC_{50} to coincide for all variables under both temperature conditions.

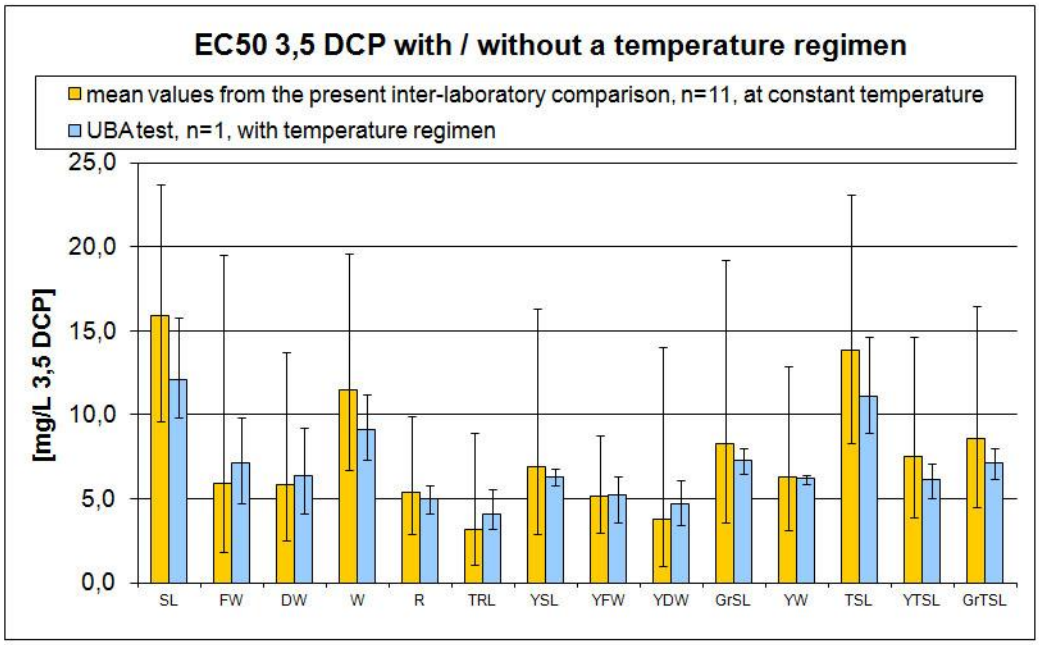


Fig. 65 EC_{50} on exposure to 3,5-DCP for all variables from tests at constant temperature (mean values from the present inter-laboratory comparison) and from a test with a temperature regimen conducted simultaneously at UBA.

5. Discussion and proposals for test optimization

5.1. Practicability, validity criteria, handling

Given the requirements of sterility, five measurement times, eight measurement variables, eight variables calculated in addition and three test substances, the present inter-laboratory comparison meant a great temporal and technological challenge for the participants.
The high number of measurement times and variables served to generate a solid backing by pertinent data for selecting suitable endpoints and generating meaningful measurement intervals. This is why the temporal effort required for the practical performance of these tests is not representative of its use as a standard test system in the future. Rather, based on the experience gained and the results obtained, proposals for optimization and simplification are presented below.

The requirement of **sterility** is an essential characteristic of the test system whose practicability was to be tested in the context of the inter-laboratory comparison.
Although eleven of the participants had no previous experience with the test system, no more than three of 30 tests submitted had to be evaluated as invalid due to apparent contamination with foreign organisms. Also taking into account the fact that – as reported by the participants – test runs were stopped and freshly prepared or repeated, the success rate can be considered as an indication of the general practicability of the test system in its present form. Altogether, 259 out of 300 control replicates (86%) apparently remained free from foreign organisms, among these where 21 tests showing no contamination with foreign organisms at all. Hence, contamination with foreign organisms was not a consistent and widespread problem but was limited to a few test runs. However, it occurred to a relatively massive extent in many of the replicates. This suggests that the cause has to be attributed to shortcomings in working under sterile conditions which, once identified, can be avoided.

Thus, in the final discussion, the participants rated the requirement of sterility as “demanding but feasible”. Participants who had performed tests with all three substances reported that increasing routine had facilitated working under sterile conditions. Particularly the permanent cultures can be kept sterile more easily if cultured at reduced light intensity and temperature. Nevertheless, due to the sugar content of the medium, the requirements to be made for working under sterile conditions are particularly high, which applies both to the pre-cultures and the tests proper. Based on their experience, the participants agree that the use of a FLAMEBOY provides an efficient remedy to prevent contamination.
It was suggested to omit replacement of the medium during the test in order to avoid an additional source of contamination and to essentially reduce the working efforts. In addition, this would prevent plants from becoming injured during the transfer (some participants reported that the transfer of large-grown plants had been difficult). UBA findings have shown that the concentration of nutrients is sufficient for a stable growth over the entire testing period. Replacement of the medium is carried out for the single purpose of ensuring a constant exposure to the test substance. Hence, replacement of the medium can be omitted on condition that the test substance will remain sufficiently stable (if necessary, this should be checked by preliminary testing).

The inter-laboratory comparison has shown that during the visual inspection to assess whether a replicate was contaminated with foreign organisms or not, in addition to the obvious cases (turbid medium and/or filamentous clusters were seen), a number of observations were difficult to interpret.

1. At higher concentrations of 2,4-D (and to a minor extent, also under IP), whitish coatings were seen on the plants while the medium remained clear. These were interpreted as indications of substance-related decomposition. However, a final explanation is still pending.
2. In some replicates, reddish brown particles were observed to adhere to the plants, however, without resulting in a clouding of the medium. Possible causes include both contamination or plant secretions, or something completely different. The participants reported unanimously that reddish brown particles were also observed in some of the permanent cultures (in part on the plants, in part as a red-brown deposit at the bottom of the culture vessel). Their origin and cause have remained unclear. It was reported that motility was seen on microscopic examination. However, culturing on agar remained unsuccessful. A final clarification of the nature of this phenomenon remains a subject for further examination.
The participants agreed that clear instructions should be included in the SOP regarding the assessment of sterility and proper working under sterile conditions.

Four out of the 30 tests failed to comply with the validity criterion of **shoot length doubling**. As a result, the number of tests invalid due to reduced growth was greater than that of tests invalid due to contamination problems. It is therefore required to take a closer look at the factors influencing growth.
While the required temperature was achieved in all tests, the light intensities measured varied between 47 and 110% of the target value. At the same time, the control replicates showed great differences in the growth factors that varied between 1.8 and 3.9. Hence, a relationship could have been assumed to exist between growth factor and light conditions. However, no direct correlation could be found between the light intensity and the shoot length yield factor. In further analyses, the range of light sources used and possible shading effects by lids, labels etc. on the test vessels might be examined. Another potential (additional) influencing factor to be considered is the mean starting length of the plants used. In the present inter-laboratory comparison, the starting lengths showed very little dispersion on the intra-laboratory level, however, with most different laboratory-specific starting values varying between a minimum of 19 – 22 mm and a maximum of 29 - 33 mm. Among these plants, those with higher starting values tended to show lower growth factors (at comparable light intensity values).
In the majority of cases, no difference is seen between the EC_{50} and NOEC values from tests with and without shoot length doubling. However, on exposure to 2,4-D, the tests without shoot length doubling showed a distinctly lower sensitivity to the test substance for some variables. This could be interpreted as an indication that the criterion of shoot length doubling was essential for the validity of the test, at least for the testing of substances interfering with growth. Although this cannot be rated as confirmed because of the low numbers of cases observed, the validity criterion of shoot length doubling should be considered as basically indispensable because on principle, the potential growth-interference action of a substance cannot be assumed to be known.

Evaluation of the variability of the controls has demonstrated that the inter-laboratory variability is always higher than the intra-laboratory variability by a factor of ca. 1.5. It has been demonstrated that robust toxicity parameters with sufficient statistical power can be determined under the present conditions. Nevertheless, it is assumed that the standardization of **handling** and **framework conditions** can be enhanced. Particularly the determination of fresh weights has a potential for further standardization by means of detailed instructions to be included in the SOP.

The variables, number and length of lateral branches proved to be inappropriate endpoints because their formation appears to be coincidental and was observed only in 50% of the control replicates. The discussion on triggering factors mentioned a possible relationship with minimal injuries of the shoot tip. This would have to be verified in further studies. In the interest of a high

degree of standardization of the test conditions, the SOP should at any rate recommend experimental plants always to be held with a (flattened) forceps or featherweight forceps by the stalk, not by the shoot tip.
According to reports by participants, also the age of the pre-culture seems to have an influence on the formation of lateral branches: Plants originating from older pre-cultures (> three weeks) apparently show a higher tendency to produce lateral branches. This would be another field for further studies. Already in its presently existing form, the SOP specifies a 14 – 21 day period for pre-culture. This rule could be supplemented by a reference to the findings described above.

Specific action of substances
In the present inter-laboratory comparison, the formation of rudimentary roots in the leaf axils of the main shoot under the influence of 2,4-D complicated the evaluation of the root number parameter because many participants did not identify these rudimentary roots as roots and therefore, failed to enter them in the test record. However, the question whether or not the rudimentary roots are counted as roots is of considerable influence on the resulting EC_{50} and NOEC. Since the rudimentary roots with a maximum length of 1 mm have a clear influence on the number of roots but virtually none on the total root length, the latter should be preferred to the number of roots as a parameter.
It may be disputable whether these are actually functioning roots. Nevertheless, this phenomenon indicates a morphological change of the experimental organism which should be documented just like other visually distinguishable effects. Therefore, these root-like structures should be defined and their evaluation explained in the SOP.

The same applies, on principle, to all potential morphological changes in the test organism due to the test substance, if known (e.g. deformation due to 2,4-D, elongation growth due to IP). In cases where due to the specific action of the substance certain endpoints can no longer be evaluated or it is impossible to derive toxicity parameter values with sufficient statistical power, alternative or additional endpoints can be recommended (e.g. yield number of whorls in cases of elongation growth).

The argument was discussed that the root parameters, although being by far the most sensitive to all test substances, are inappropriate as endpoints in the present sediment-free test system because these organs are not normally exposed to light. However, other participants experienced in testing with sediment reported that also in this type of test, the root variables are the ones to react most sensitively although the roots were not exposed to light. In the sediment contact tests with *Myriophyllum*, the endpoint root growth has been evaluated only in qualitative terms so far because the plants produce far more roots than in the present sediment-free test system. As a result, quantification becomes complicated or impossible. In the sediment-free test system, in contrast, quantification of the root number and length is possible. In addition, given the fact that irrespective of the test substance, the root parameters are by far the most sensitive endpoints showing the best reproducibility and evaluability, these endpoints have proved to be indispensable. Of course, it is always required for the respective endpoints to be interpreted and evaluated in the context of the other results.

The participants discussed the **photographic documentation** and subsequent digital measurement of lengths as compared to the manual measurement of lengths by means of a ruler. Some of the participants considered the photographic recording as very complex and time-consuming. Eventually, the choice of the method for measuring the length (digital or manual) can be left to the laboratories. It should only be recommended to use the same method within one test (unless the measurement of lengths is on principle carried out digitally but in the case of seriously deformed plants only possible in the manual way by means of a ruler, if at all). Beyond practical arguments (practicable or not, depending on plant morphology), it is a general

advantage of the photographic method that results can be permanently archived and verified at any time. Moreover, evaluation of additional parameters at a later time is possible if this turns out to be meaningful (e.g. number of whorls, see above). Routine use could be facilitated by the installation of a stationary camera set-up in the laboratory.

5.2. Suitability and selection of variables and test design

Criteria for selecting variables suitable for the determination of the toxicity parameters, EC_{50} and NOEC, include

1. A good reproducibility in the controls, i.e. sufficiently low coefficients of variation;
2. A good statistical evaluability (i.e. inhibition values up to 100 % possible);
3. A good evaluability irrespective of the mode of action of the test substance, i.e. minor impairment by substance-specific side effects (morphological changes in the test organism);
4. Sensitivity;
5. A good reproducibility of toxicity parameters; and
6. Comparability with other test systems (e.g. *Lemna*).

On item 1: With the exception of lateral branches as a single endpoint, the results regarding reproducibility of control replicates (CV% 10-30%) have proved all variables tested to be suitable endpoints for the determination of toxicity parameters.

On item 2: The yield variables (which also include the root parameters, since roots are on principle not produced before test start) will result by definition in higher inhibition values than the originally measured values and therefore, are more suitable for a dose-response analysis and EC determination.

On item 3: Variables that may be difficult to evaluate due to side effects of the test substances include: Root number, yield fresh weight, yield dry weight, main shoot length, growth rate main shoot length, yield number of whorls.

On items 4, 5 and 6: The variables that proved to be most sensitive both regarding the EC_{50} and the NOEC, irrespective of the test substance, included the number of roots, total root length, yield fresh weight and yield whorl number. These variables have also shown the best reproducibility for EC_{50} and NOEC. The yield and growth rate of shoot length assume a moderate to non-sensitive position both regarding sensitivity and reproducibility. However, the growth rate ensures comparability with the results from other toxicity tests and is therefore indispensable.

Given the unresolved problem of lateral branch formation (see above), the total shoot length appears to be less meaningful as an endpoint than the shoot length. Consequently, also the growth rate total shoot length can be omitted.

Of the remaining variables, those recommended for measurement and evaluation include: *Total root length* (also requires recording of the root number), *yield fresh weight*, *yield dry weight* (optional, since depending on the test substance, not generally correlated with fresh weight; where applicable, the more sensitive parameter should be chosen); *yield main shoot length* (at least as a basis for determining the growth rate, optionally also as an independent endpoint); *growth rate main shoot length, yield number of whorls* (optional, if shoot length parameter is difficult to evaluate). Thus, the number of variables evaluated in the inter-laboratory comparison is reduced from 16 to a maximum of 6.

Moreover, measures recommended to simplify the testing procedure as compared to that applied in this inter-laboratory comparison include: Omitting the replacement of the medium after 7 days (presupposing stability of the test substance), limiting the measuring times to three (day 0, day 7 and day 14; on day 7 only shoot length, root number and length, and, if required, whorl number), testing at a constant temperature (23°C $\pm$ 2°C).

The test duration of 14 days has to be maintained to ensure at least a doubling of the shoot length on the one hand and on the other, ensure a reliable determination of EC_{50} and NOEC values.
Also the test design with 10 control replicates and 5 replicates each in 8 treatments should be maintained since it is essential to achieve %MDD values similar to those resulting from the present inter-laboratory comparison.
The validity criteria of “shoot length doubling” and “at least 50% control replicates apparently free from foreign organisms” have proved to be successful and should also be maintained.
The existing comprehensive data set on 3,5-DCP and the consistently high evaluability and reproducibility of the toxicity parameters for all variables under exposure to 3,5-DCP speak in favour of envisaging 3,5-DCP to serve as a reference substance for the test system with *Myriophyllum spicatum*.

5.3. Outlook

The present inter-laboratory comparison has raised a number of questions that could be elucidated in further studies. Such questions include

- Cause and nature of the red-brown particles seen in a number of pre-cultures and test runs;
- Cause and nature of the whitish coating observed at high test concentrations of 2,4-D;
- Cause of reduced growth in controls; Influence of reduced growth on toxicity parameters;
- Cause of formation of lateral branches (injury of shoot tip, age of pre-culture?); and
- Cause of anticorrelation between fresh weight and dry weight on exposure to 2,4-D.

6. Conclusion

The results of the inter-laboratory comparison have shown the practicability, reproducibility and suitability of the sediment-free test system for *Myriophyllum spicatum*. Owing to the comprehensive volume of data obtained it is possible to determine framework conditions (test design, conditions of culture, measuring times) and select suitable variables as endpoints to determine toxicity parameters that can be fulfilled and generated, respectively, with justifiable effort and may also be expected to provide for results with sufficient statistical power.
Based on the experience of participants, practical proposals for further standardization are derived and suggestions made for further studies that may improve the understanding of the test system.