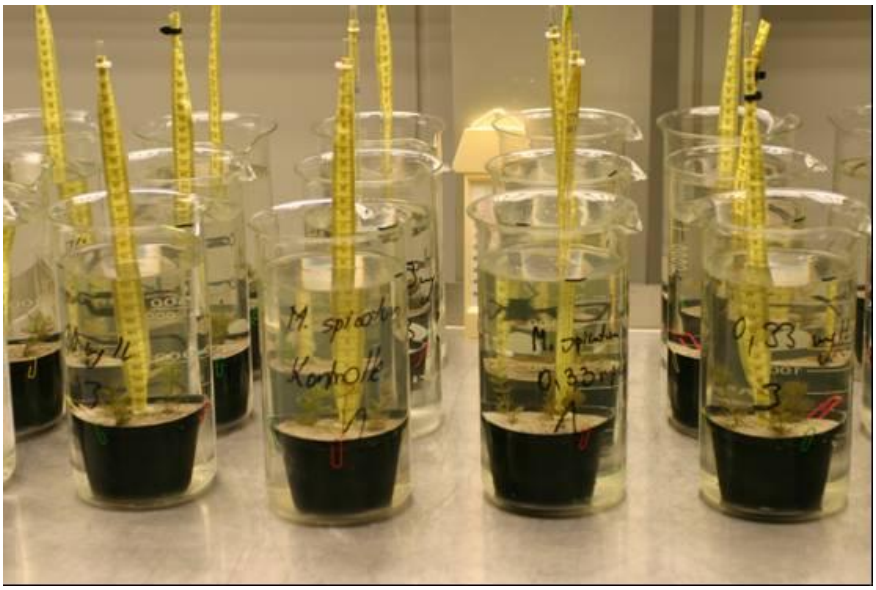


Myriophyllum Toxicity Test

Results of a Ringtest Using M. aquaticum and M. spicatum Grown in a Sediment-Water-System

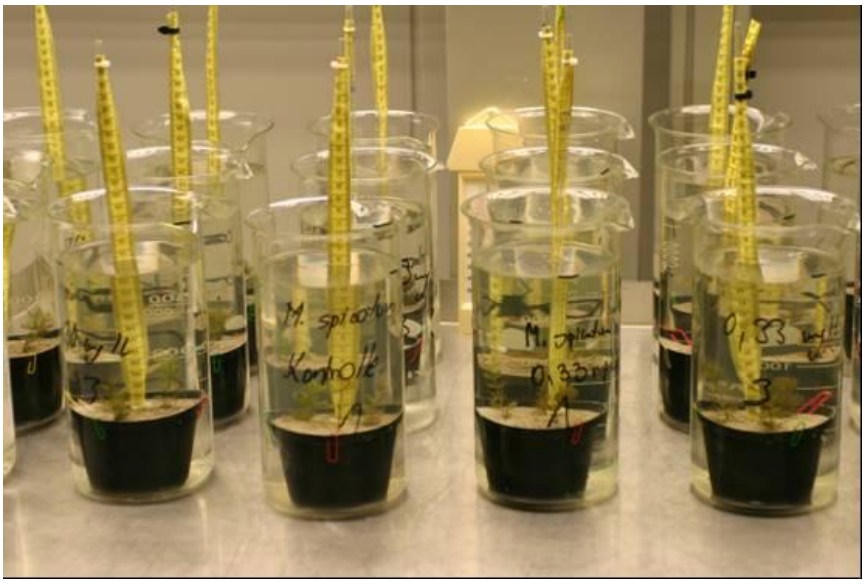
Organised under the Auspices of the
SETAC Aquatic Macrophyte Ecotoxicology Group (AMEG)

Final Report
October-31-2012

Myriophyllum Toxicity Test

Results of a Ringtest Using M. aquaticum and M. spicatum Grown in a Sediment-Water-System

Organised under the Auspices of the
SETAC Aquatic Macrophyte Ecotoxicology Group (AMEG)



Organization of the Ring Test	Dr. Peter Dohmen, Johanna Kubitza (BASF SE, Limburgerhof, Germany)
Financial Support for Evaluation and Report	Dow AgroSciences GmbH, Makhteshim Agan, Syngenta Ltd., Bayer CropScience AG, Monsanto Company, DuPont CropProtection-DuPont de Nemours (Deutschland) GmbH
Statistical Evaluation and Report	Dr. Monika Ratte, Dr. Hans Toni Ratte ToxRat Solutions GmbH, Alsdorf, Germany

Final Report
October-31- 2012

Summary

In the context of the plant protection legislation, the risk of pesticides with herbicidal action to aquatic plants is assessed from endpoints from toxicity tests with unicellular algae and the monocotyledonous lemnaceans. It has been questioned as to whether the pesticide sensitivity of lemnaceans can be extrapolated to that of dicotyledonous species. To reduce this uncertainty, the dicotyledonous water milfoils (*Myriophyllum* sp.) are considered suitable candidates to be used as an additional macrophyte species in the regulatory context for herbicide approval. Organized by SETAC Aquatic Macrophyte Ecotoxicology Group, a test method with *M. aquaticum* or *M. spicatum* grown in a sediment-water-system has been developed and ring tested to assess toxicity of substances via the water- and sediment-phase (method described in Maltby and al. 2010), subsequently called "AMRAP-test"[1]. The present report presents and discusses the results of the AMRAP-test.

AMRAP Ring Test – General

The AMRAP ring test was conducted between January and October 2011. Fifteen laboratories performed 51 tests in total using two test species (*M. aquaticum* and *M. spicatum*) and three test substances (3,5-dichlorophenol (3,5-DCP), isoproturone (IP) and trifluralin (TF). After a rooting phase of 3 days (*M. aquaticum*) or 7 days (*M. spicatum*) the test was started with three test plants per pot (=replicate) and 6 replicates in the control and 3 replicates in 5 treatments, each. Exposure time was 7 days (*M. aquaticum*) and 14 days (*M. spicatum*) whilst they were submerged. Total test duration was 10 days for *M. aquaticum* and 21 days for *M. spicatum*. Measured variables were total shoot length (TSL), fresh weight (FW), dry weight (DW) and number and total length of lateral branches (LB, TLBL). For length and weight variables, yields (Y) and growth rates (Gr) were calculated as endpoints. Root development was assessed visually.

Results from analytical measurements were provided from 7 laboratories for 12 tests. IP proved to be relatively stable during the test (mean recovery 78,1 % (DAT 7, *M. aquaticum*) and 75,2 % (DAT 14, *M. spicatum*). For 3,5-DCP mean recovery rates were 66% (DAT 7, *M. aquaticum*) and 47% (DAT 14, *M. spicatum*). Dissipation of the test item was highest in TF (mean recovery rates 9,5% (DAT 7, *M. aquaticum*) and 1,8% (DAT 14, *M. spicatum)*. Percentages of the test items found in the sediment (pore water, concentrations related to nominal concentrations in water phase at DAT 0) were 19,9% (IP), 8,2% (TF) and 6,2% (3,5-DCP), at DAT 7 in test with *M. aquaticum*. Analytical measurements in sediment were not available for tests with *M. spicatum*. In the present analysis always nominal concentrations were used to estimate endpoints. In several cases the data base for calculation is relatively small, and the number of values did not always meet the requirements for ring-test statistics, so that these results have to be taken with care. This is valid for the calculation of the overall mean EC50 and NOEC, and of the mean minimum detectable difference (MDD).

In the absence of defined validity criteria all toxicity-test-data sets were included in the initial statistical analysis. As a result of the evaluation process and as a compromise between minimizing variability and maximizing the number of remaining data sets, the following

[1] AMRAP = Aquatic Macrophyte Risk Assessment for Pesticides

preliminary validity criteria were proposed: growth rate fresh weight $\geq$ 0,07 d^{-1} (corresponding to a doubling time of 9,90 d), and coefficient of variation in yield of fresh weight $\leq$ 35%. Further evaluation was performed using only valid data based on these criteria.

AMRAP Ring Test – Results for *M. aquaticum*

The mean growth rate of total shoot length was 0,092 d^{-1} (all data, corresponding to a doubling time of 7,5 days) and 0,099 d^{-1} , respectively, (valid data, corresponding to a doubling time of 7,0 days). The mean growth rate of fresh weight was 0,078 d^{-1} (all data, corresponding to a doubling time of 8,9 days) and 0,114 d^{-1} , respectively, (valid data, corresponding to a doubling time of 6,1 days).

50% of the data sets fulfilled the validity criteria. Generally, in *M. aquaticum* lateral branches occurred only infrequently and therefore proved to be not appropriate for evaluation.

The repeatability CV_r% (i.e. the within-laboratory variability of the control data) of the measured variables, ranged from 18% to 55% (all data), respectively from 14% to 41% (valid data); the reproducibility CV_R% (i.e. the between-laboratory-variability of the control data) ranged from 35% to 74% (all data), respectively from 35% to 62% (valid data). The weight parameters exhibited the highest variability, while the shoot length parameters showed the lowest variability. Growth rates were less variable than yields. Based on all data, mean MDDs between 21% (GrTSL) and 75% (GrDW) where obtained, while based on valid data, mean MDDs between 13% (GrTSL) and 74% (YDW) were obtained.

3,5-DCP was less toxic than IP and TF. However, due to the limited number of data, the following statements about the sensitivity of *M. aquaticum* should be considered carefully (in the following, only means based on at least two data sets are reported). For 3,5-DCP, the mean EC50 ranged between 3,8 and 5,0 mg/L (all data, n=2-8), respectively between 4,6 and 6,1 mg/L (valid data, n=2-5), with only small differences between variables. IP and TF showed similar toxicity on a markedly lower concentration level than 3,5-DCP. For IP, mean EC50 values ranged between 0,06 and 0,77 mg/L (all data, n= 2-8), respectively between 0,37 and 0,86 mg/L (valid data, n=2-3). Dry weight parameters were most sensitive, however, in view of the small data base, the order of the sensitivities seems to be questionable. For TF, only EC50 values for the shoot length parameters were available. They ranged between 0,60 and 0,82 mg/L (both all data and valid data, n= 3-5). In the presence of IP (and to a lower extent of TF) growth promotion related to weights was observed at lower concentrations. Mean NOECs for *M. aquaticum* were as follows: 3,5-DCP: 2,9 - 5,0 mg/L (all data), 2,4 - 5,6 mg/L (valid data); IP: 0,069 - 0,147 mg/L (all data), 0,047 - 0,100 mg/L (valid data); TF: 0,042 – 0,057 (all data), 0,032 mg/L (valid data).

AMRAP Ring Test – Results for *M. spicatum*

The mean growth rate of total shoot length was 0,107 d^{-1} (all data) and 0,106 d^{-1} (valid data), respectively, both corresponding to a doubling time of 6,5 days. The mean growth rate of fresh weight was 0,093 d^{-1} (all data, corresponding to a doubling time of 7,5 days) and 0,103d^{-1} (valid data), respectively, corresponding to a doubling time of 6,7 days.

For *M. spicatum,* 86% of the data sets fulfilled the validity criteria. Lateral branches were produced regularly and accounting for approximately 50% of final total shoot length. Hence number of lateral branches (LB) and total length of lateral branches (TLBL) were evaluated as

additional variables.
M. spicatum exhibited a relatively low variability of the control data within the laboratories (CV_r%: 10,2% - 24,2% (all data), 9,9% to 24% (valid data)), but higher variability of the control data between the laboratories (CV_R: 29,6% - 57,5% (all data), 21% - 58%(valid data)). Systematic differences in variability between length and weight parameters were not obtained, but growth rates were less variable than yields. Based on all data, mean MDDs between 13% (GrTSL) and 44% (YFW) were obtained, while based on valid data, mean MDDs between 14% (GrTSL and GrFW) and 33% (TLBL) were obtained.
3,5-DCP was less toxic than IP and TF. However, due to the limited number of data the following statements about the sensitivity of *M. spicatum* should be considered carefully (only means based on at least two data sets are reported). For 3,5-DCP, the mean EC50 ranged between 4,2 and 6,1 mg/L (all data, n=4-7) and between 4,7 and 6,3 mg/L (valid data, n=4-6) with only small differences between the variables. IP and TF showed similar toxicity on a markedly lower concentration level than 3,5-DCP. For IP, mean EC50 values ranged between 0,05 and 0,34 mg/L (all data, n=5-7), respectively between 0,05 and 0,32 mg/L (valid data, n=4-6). A uniform order in sensitivity of weight and length parameters was not obtained, but growth rates were less sensitive than the corresponding yields. For TF, mean EC50 values ranged between 0,19 and 0,33 mg/L (all data, n=2-4), respectively between 0,16 to 0,32 mg/L (valid data, n= 2-4), i.e. only small differences between the variables were obtained. In the presence of IP (and to a lower extent of TF) growth promotion related to weights was observed at lower concentrations. Mean NOECs for *M. spicatum* were as follows: 3,5-DCP: 3,0 - 3,8 mg/L (all data), 2,9 - 3,9 mg/L (valid data); IP: 0,018 - 0,064 mg/L (all data), 0,015 - 0,076 mg/L (valid data); TF: 0,032 – 0,101 (both for all data and valid data).

Comparison of the Results for *M. aquaticum* and *M. spicatum*
Test duration was 10 days for *M. aquaticum* (rooting phase 3 days, exposure time 7 days) and 21 days for *M. spicatum* (rooting phase 7 days, exposure time 14 days). Regarding the overall means of all control data, *M. aquaticum* showed slightly lower growth rates than *M. spicatum* (GrTSL 0,092 d^{-1} compared to 0,107 d^{-1}; GrFW 0,078 d^{-1} compared to 0,093 d^{-1}). Regarding only valid data, again *M. aquaticum* showed a slightly lower GrTSL than *M. spicatum* (0,099 d^{-1} compared to 0,106 d^{-1}), but a slightly higher GrFW (0,114 d^{-1} compared to 0,103 d^{-1}).
In view of the obtained growth rates, the prescribed exposure time of 7 days for *M. aquaticum* on average was not sufficient for doubling fresh weight and hardly sufficient for doubling shoot length. In contrast, for *M. spicatum* growth rates and exposure time resulted in an increase of shoot length by a mean factor of 4,7 and an increase in fresh weight by a mean factors of 2,4.
M. aquaticum showed only infrequently production of lateral branches whereas in *M. spicatum* regular production of lateral branches was found.
M. aquaticum showed both higher between-laboratory-variability and higher within-laboratory-variability of the control data than *M. spicatum*. If preliminary validity criteria were applied, generally both species allowed statistical testing at minimal detectable differences up to about 30%, except for the dry weight variables in *M. aquaticum*. Therefore, from a statistical point of view, both species are appropriate for assessment of toxic metrics (ECx, NOEC), and in so doing, for *M. aquaticum* fresh weight and shoot length parameters should be preferred in contrast to dry weight parameters. Generally, for both species growth rates exhibited lower

variability and better statistical power than yields.
For 3,5-DCP, similar responses were obtained in both species. In contrast, *M. aquaticum* was less sensitive to IP than *M. spicatum* by a factor of 2 up to 10 depending on the variable. These differences were statistically significant in three out of four comparisons. For TF, only one EC50 is available for comparison (YTSL), which indicated that *M. aquaticum* appears to be less sensitive than *M. spicatum*.
A general effect on root development was confirmed by the semi-quantitative assessment of root development for both species. Root development was affected by all test items, with obvious inhibitions for 3,5-DCP and IP, and only low effects of TF. For *M. aquaticum*, it could be shown by means of two tests with 3,5-DCP and one test with IP, that also a quantitative assessment of root length in principal is possible. Sensitivity of root length was found to be either in the same range as the other variables (3,5-DCP) or lower (IP). The limited data base is not sufficient for final conclusions, but gives evidence that roots are as sensitive as the other parameters or even less.

Recommendations Addressing the AMRAP-test
The analysis of the ring-test data for *M. aquaticum* and *M. spicatum* revealed that the toxicity test in principle is practicable. Several laboratories delivered data sets of acceptable quality. However, there is a need for further methodical standardization and training of the experimenters in order to reduce variability, especially with *M. aquaticum*.
There might be several reasons for the observed variability (besides laboratory experience), amongst others: (1) adhering sediment particles to roots might generally skew weight determination; due to the lower absolute level of weights, this would mainly affect the results for *M. aquaticum;* (2) differences of more than 20% between length of "representative plants" (used for determination of initial weights) and test plants caused over- or underestimation of weight yields and growth rates; (3) test conditions deviating from test protocol (duration of rooting phase, temperature, light); (4) different plant quality and preculturing conditions and -durations.
Several specifications concerning initial plant length and weight and minimum requirements for variability of initial weights were already recommended in the test protocol. It should be clarified whether they simply were not always followed or whether they need to be enhanced.

In any case, the definition of appropriate validity criteria is essential. Preliminary validity criteria have been proposed (growth rate of fresh weight $\geq$ 0,07 d^{-1} (corresponding to a doubling time of 9,90 d), coefficient of variation in yield of fresh weight $\leq$ 35%), but were adjusted such, to get at least 50% valid tests in the present ring test. Hence the nature and value of final validity criteria have to be discussed among experts.

Based on the results of the ring test, the exposure time of 7 days for *M. aquaticum* given in the current test protocol was not sufficient for doubling fresh weight and hardly sufficient for doubling shoot length, even in tests which fulfilled the proposed validity criteria. So the question arises, as of whether the chosen time period for *M. aquaticum* generally was too short, or if the measured growth rates were unusual low due to any methodological shortcomings (see above). If both *Myriophyllum* species will remain candidates for the toxicity test in future, it is

recommended to prescribe experimental periods enabling at least doubling of shoot length and fresh weight in each species. Moreover, it remains to be clarified whether the shorter test duration is the reason for the lower sensitivity of *M. aquaticum* against two of the three test items.

Major and systematic differences in the sensitivity of the different variables were not obtained, except the finding that for both species growth rates were less sensitive than yields. So, it could be advisable to focus on a subset of variables in further testing. Candidates would be the growth rates based on total shoot length or weight since growth rates showed trends towards lower variability and - assuming the plants stay in exponential growth - are time independent in contrast to e.g. the yield. In addition, one of the weight variables should be included, in order to ensure capture of effects that are not manifest on shoot length, which is a function of cell elongation, not necessarily of biomass accumulation. Additionally, effects on roots biomass are captured by total plant weight.

The variables considered in the ring test were mainly related to growth, but depending on the mode of action also effects on the differentiation of milfoils cannot be excluded. Hence, to get comprehensive information about the intrinsic toxicity of test substances, especially those with herbicidal action, also effects on variables related to differentiation should be included in the future toxicity test, e.g. a quotient of weight and total shoot length or the number of lateral branches.

From the statistical viewpoint, the test design could be improved with respect to both the needs of a NOEC and of an EC-design. A regression design and the computation of ECs would benefit from more concentrations (e.g. eight treatments), especially if non-linear regression models should be used. Concerning NOEC determination, in view of the possibly high variability, more replicates would enhance the statistical power.

Table of Contents

1. Introduction

Up to now, the Alga Growth Inhibition Test (OECD 201) and *Lemna* sp. Growth Inhibition Test (OECD 221) are used to assess the risk of pesticides with herbicidal action to aquatic plants. But the floating macrophyte species "Lemna" is seen as not representative for rooted macrophytes and there is increasing doubt, as of whether the sensitivity of the monocotyledon *Lemna* sp. against herbicidal substances is representative also for higher aquatic dicotyledons. To address this uncertainty, the dicotyledonous water milfoils (*Myriophyllum* sp.) are considered suitable candidates to be used as an additional macrophyte species in the regulatory context for herbicide approval. Milfoils have been found to be sufficiently easy to work with and to provide reproducible results with reasonable effort.

Up to now three test methods have been developed and ring tested: (1) a test with *M. aquaticum,* using only sediment without water phase, developed for testing of sediments by the Federal Institute of Hydrology (BfG, Koblenz, Germany), ring tested in 2011, subsequently called "ISO-test" (ISO 16191, Feiler et al. 2012); (2) a test with *M. spicatum* grown in the water phase of a sediment-free system (Federal Environment Agency (UBA; Berlin, Germany, ring tested in 2010; method based on ASTM Standard E 1913-04; subsequently called "UBA-test") (UBA Report FKZ36301294); and (3) a test with *M. aquaticum* and *M. spicatum* grown in a sediment-water-system , organized by SETAC Aquatic Macrophyte Ecotoxicology Group, subsequently called "AMRAP-test"[2]. With respect to the latter it is argued that a water sediment test is preferable because the cultivation in a water-sediment system is a more realistic set-up.

The AMRAP ring test was conducted between January and October 2011. Fifteen laboratories performed 51 tests in total using two test species (*M. aquaticum* and *M. spicatum*) and three test substances (3,5-DCP, isoproturone (IP) and trifluralin (TF)) as representative substances showing differing modes of herbicidal action. After a rooting phase of 3 days (*M. aquaticum*) or 7 days (*M. spicatum*) the test was started with three test plants per pot (=replicate) and 6 replicates in the control and 3 replicates in 5 treatments each. Exposure time was 7 days (*M. aquaticum*) and 14 days (*M. spicatum*) whilst they were submerged. From the measured biological data the following variables were calculated and used for evaluation of toxicity metrics: fresh weight, dry weight and total shoot length (each with yield and growth rate); moreover number and total length of lateral branches. Additionally root development was assessed semi-quantitatively.

The present report describes the results of the statistical analysis of the AMRAP ring test data and addresses the following objectives:

- to characterize the variability of the measured variables and thus the general practicability of the test

[2] AMRAP = Aquatic Macrophyte Risk Assessment for Pesticides

- to determine the toxicity metrics (EC50, NOEC) for the different variables and test species for different test substances by the same statistical methods, each
- to identify appropriate variables to derive the toxicity metrics by means of sensitivity and reproducibility,
- to compare the results obtained for the two Myriophyllum species for different test substances.

A comparative ring test statistic according to DIN ISO 5725-2 & 5 was performed in order to answer these questions.

2. Basic Conditions and Details of the Ring Test

2.1. Participants

In total, 15 laboratories participated in the ring test (see Table 1). Not all of the participating laboratories had experience with the test system. An internal lab code was randomly assigned to each laboratory (L01…L15) and used in all figures and tables of the present report. The laboratory code differs from the sequence used in Table 1.

Table 1: Participating laboratories in alphabetical order

Laboratory
Alterra, Wageningen, NL
BASF SE, Limburgerhof, D
BayerCropScience LP, Stilwell, USA
BioChem agrar GmbH, Gerichshain, D
Chemex Environmental International Ltd, Cambridge, UK
Dr. U. Noack Laboratorien, Sarstedt, D
ECT Ökotoxikologie GmbH, Flörsheim, D
Eurofins Agroscience Services GmbH, Niefern-Öschelbronn , D [1)]
Federal Environmental Agency (UBA), Berlin, D
Fraunhofer IME, Schmallenberg, D
Harlan Laboratories Ltd, Itingen, CH
Ibacon GmbH, Rossdorf, D
Institute of Industrial Organic Chemistry (IPO), Pszczyna, Pl
University of Novi Sad, Lecotox Lab., , SRB
Smithers Viscient, Wareham, USA

1) Meanwhile name changed to Eurofins Agroscience Services EcoChem GmbH

2.2. Time table

Prior to 2011	Preparation of the ring test by AMRAP; organized by BASF SE, Limburgerhof, Germany
Jan 2011 – Sep 2011	Conduct of tests by the participating laboratories; see the overview in Table 2
Oct 2011 – May 2012	Preliminary data analysis and presentation of first results by BASF SE
31.05/4.06./5.06. 2012	Submission of the raw data to the contractor by BASF SE
03-July 2012	Submission of the Draft Report by the contractor

Table 2: Date of tests performed in the participating laboratories; all tests performed in 2011

Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Oct

Lab Code	M. aquaticum			M. spicatum			No of tests performed
	3,5-DCP	IP	TF	3,5-DCP	IP	TF	
L01				14-Jun			1
L02	18-Mar	18-Mar	8-Apr	10-May	10-May		5
L03	7-Mar	16-Mar	9-Sep	30-Aug	28-Jul	15-Sep	6
L04				7-Jul	4-Aug	1-Sep	3
L05		28-Mar					1
L06				2-Mar	30-Mar	15-May	3
L07	23-May	3-Jun	19-Aug				3
L08	18-Feb	11-Mar	25-Mar	24-Jun	22-Jul	8-Jul	6
L09	31-Mar	9-May					2
L10				15-Apr	1-Jun		2
L11	15-Jan	15-Jan		27-Apr	27-Apr		4
L12				9-Jun	14-Jan	26-Oct	3
L13	5-Aug	26-Aug	13-Sep				3
L14	7-Mar	12-Aug	23-Sep	7-Mar	12-Aug	23-Sep	6
L15	21-Mar	3-May	16-Jun				3
	9	**10**	**7**	**10**	**9**	**6**	**51**

2.3. Test Items, Test Species and Number of Conducted Tests

Three test items with herbicidal activity were selected, which differed in mode of action (MoA) and physic-chemical properties:

1. 3,5-dichlorophenol (3,5-DCP), reference substance OECD 221 and 201, pesticide, non-specific MoA

2. Isoproturone (IP); herbicide, photosynthesis inhibitor

3. Trifluralin (TF); pre-emergence herbicide; microtubule assembly inhibitor, interrupts mitosis (root inhibitor).

The complete ring test protocol is given in the annex. In the following the main issues are summarized.
According to the ring-test protocol test batches of 3,5-DCP had to be purchased and should have had at least 97% purity; isoproturone and trifluralin were provided to the participating laboratories by BASF SE. The ring-test protocol recommended testing the substances in the order shown above in case participating laboratories could not perform tests with all compounds. At least 3,5-DCP should have been tested with both species or all test substances with one of the species.

Analytical measurements of test substance concentrations in water at test initiation and termination was recommended, but was performed on a voluntary basis.

Tests were performed with two species: *Myriophyllum spicatum* and *Myriophyllum aquaticum*. This was to test options for different test species as in other guidelines (e.g. OECD 201, 202, or 203).
M. spicatum (Eurasian watermilfoil) is a submerged aquatic plant, and grows in still or slow-moving water in Europe, Asia, and North Africa. It is distributed by water flow, watercraft or other objects carrying fragments of it. Smallest fragments of watermilfoil can grow in to fully developed plants. It may be difficult to obtain *M. spicatum* of a suitable quality during winter time. Therefore the ring-test protocol recommended starting with testing of *M. aquaticum* first (Jan 2011).
M. aquaticum is commonly called parrot´s feather or parrot´s feather watermilfoil and inhabits the Amazon River in South America. From there it has colonized the warmer regions of every continent. As with *M. spicatum* new plants grow from fragments of already rooted plants.

Potential suppliers of *Myriophyllum* sp. were communicated to the participating laboratories. The source of the plants used in this ring had to be documented. For other details concerning the purity and health of plants as well as the maintenance culture consider the ring-test protocol.

The plants had to be acclimatized under conditions similar to those in the test for an adequate period before the study. The culture conditions had to be documented (temperature, day length, light conditions, media details).

Table 3 gives an overview of the conducted tests, Table 4 summarizes the sources of the plants, if reported.

Table 3: Overview of the tests conducted by the participating laboratories for each test item; [a]: laboratory provided results of analytical measurements

Laboratory	M. aquaticum			M. spicatum			Tests performed
	3,5-DCP	IP	TF	3,5-DCP	IP	TF	
L01				x			1
L02	x	x	x	x	x		5
L03	x	x	x	x	x	x	6
L04				x [a]	x [a]	x [a]	3
L05		x [a]					1
L06				x	x	x	3
L07	x [a]	x [a]	x [a]				3
L08	x	x	x	x	x	x	6
L09	x	x					2
L10				x [a]	x [a]		2
L11	x	x		x	x		4
L12				x [a]	x	x	3
L13	x [a]	x [a]	x [a]				3
L14	x [a]	x	x	x [a]	x	x	6
L15	x	x	x				3
Total tests	**9**	**10**	**7**	**10**	**9**	**6**	**51**

Table 4: Overview of the source of the *Myriophyllum* species used with testing; empty cells: test not performed; x: test performed but source of plants not stated. Please note: numbers of tests performed do not always correspond to those given in Table 2 and 3 in order to hamper identification of Laboratories.

Laboratory	M. aquaticum			M. spicatum		
	3,5-DCP	IP	TF	3,5-DCP	IP	TF
Alterra	Germany 2008			Sinderhoeve test facility, Renkum, NL 3.3.2011		
BASF	Petrowsky, Eschede, 23.2.11	Petrowsky, Eschede, 23.2.11	own stock culture	own stock culture	"small mesocosm"	Petrowsky, Eschede 2.9.11
BCS				Dupont May 02 2005		
BioChem agrar				UBA, March 2011		
Chemex				Paul Bromfield Aquatics, Maydencroft Lane, Gosmore, Hitchin, herts, SG4 7QD 12. May 2011		x
Dr. U. Noack	own stock culture					
ECT	emersed stock culture	x	x			
Eurofins	submersed Stock culture Source: Ökotox GmbH Stuttgart (2004), since then own stock culture			Source: Ökotox GmbH Stuttgart (2004), since then own stock culture		
UBA				BfG Koblenz	BfG Koblenz 8.12.2010	
IME	Mesocosm GmbH 19.11.2010					
Harlan				UBA 13.12.10		
Ibacon	Petrowsky, Eschede 8.2.2011	Mesocosm GmbH 14.2.2011	Petrowsky 16.3.2011	Petrowsky, Eschede (date?)		
IPO	Sklep zoologiczny. Wytwarzanie i Naprawa Urządzeń Akwarystycznych. Zofia Chałupka, Łokietka 1G, 43-200 Pszczyna, Polan 28. June 2011 cultivation adaption time: 300 l aquarium					
University of Novi Sad	Bavarian Environment Agency, Dec 2009 cultivation / adaption time: 2,5 month		x			
Smithers Viscient	Florida Aquatic Nurseries, Fort Lauderdale, USA; 3,5 DCP: Jan. 2011; IP and TF: May 2011 [1)]			Bayer Crop Science, Stilwell, KS, USA, previously received from DuPont, 2 May 2005, previously from Univ. of Toronto, culture collection (UTCC), Toronto, Ontario, Canada.[1)]		

1) Plant Source of Smithers Viscient: not stated in the ring test documents; personal information 19 Oct. 2012

2.4. Test Design and Test Conditions

2.4.1. Experimental design

Details of test vessels, the composition of the artificial sediment (according to OECD 218) and culture medium are laid down in the ring-test protocol (see annex). A short summary of the test procedure derived from the test protocol is given below.
Three replicate test vessels were prepared for each treatment (5 test concentrations arranged in a geometric series) and six replicate test vessels for the control. Each vessel contained one plant pot with three shoots. The individual test vessels had to be randomly assigned to the treatments.

The following concentrations were tested:

3,5-DCP:	control, 1.0, 1.8, 3.2, 5.6, 10 mg/L
IP:	control, 0.01; 0.032; 0.10; 0.32; 1.0 mg/L
TF:	control, 0.01; 0.032; 0.10; 0.32; 1.0 mg/L

2.4.2. Test procedure

Healthy, non-flowering, shoot tips from the culture plants were clipped off at a length of 6 cm (+/- 1 cm) and weighed individually. Five shoot tips were planted into each pot containing the sediment such that the lower 3 cm, covering two nodes, were beneath the sediment surface. Shoots were then maintained for 3 days for *M. aquaticum* or 7 days for *M. spicatum* in a nutrient-poor water to induce root development. Thereafter, two of the five plants are removed to leave individuals that were as uniform in size and appearance as possible. The shoot length of test plants was measured above sediment at DAT 0 using a ruler inside the test vessels. Representative plants from five additional pots were used to determine the initial plant biomass (wet and dry weight) and length for DAT 0. The exposure period was 7 days for *M. aquaticum* and 14 days for *M. spicatum*.

2.4.3. Temperature, Light, pH, and O_2 Conditions

The ring-test protocol required a light intensity of about 140 (+/- 20) µE m^{-2} s^{-1} (equivalent to about 8.8 -11.8 klux) at the water surface and a light:dark ratio of 16:8 h. Any differences from the selected light intensity over the test area should have not exceeded the range of ± 15 %. The temperature in the test vessels had to be 20 ± 2°C.

Other provisions of the ring-test protocol include the control of water level, measurement of the pH and oxygen concentration as well as the visual inspection of plant health. Table 6 shows the time schedule for recording the environmental conditions.

Annex Table 2 gives detailed information about the measured environmental factors. Table 5 gives an overview of those tests in which the prevailing environmental conditions did not meet the prescriptions of the ring-test protocol or in which irregularities occurred.

Table 5: Overview over temperature and light conditions (black and red) and unusual visual observations or facts (blue) (reported by participants). Empty cells: test not performed. Reference value for temperature was 20 ± 2°C, i.e. 18-22°C, indicated in red: tests exceeding the (enhanced) temperature range of 18-23°C;
Reference value light intensity was 140 ± 20 µE m² s⁻¹ / 8800-11800 lux and "differences should not exceed ± 15%"; indicated in red, italics: light intensity low, but within range of + 15% deviation; red, normal letters: light intensity below the prescribed values.

Laboratory	*M. aquaticum*			*M. spicatum*		
	3,5-DCP	IP	TF	3,5-DCP	IP	TF
L01				T=21-23.4 light 124-126µE		
L02	T=19-20 light 9143-10929lx	T=19.6-22 light 9054-10530lx	T=18.7-22.7 light 9893-11420lx	T=19,8-21,9 light 8923-10456 lx	T=20,6-21,9 light 8865-9883 lx	
L03	T=16.9-26.3 light 122-139 µE	T=21.2-22.5 light 132-140µE	T=20.9-21.3 light 129-140µE rooting phase 7d	T=20.5-21.8 light 129-140µE	T=20.7-23 light 130-145µE	T=20.8-22 light 130-137µE
L04				T=20.2-21.1 light 8380-11080	T=20-20.3 light 7980-11980	T=19.9-21.7 light 9210-11140
L05		T=19.1-21.7 light 7300-10500lx				
L06				T=19.6-25.3 light 109-148 µE	T=20.4-24.3 light 109-136 µE	T=21.2-24.5 light 110-143 µE
				visible black spots on leafs	visible black spots on leafs	test concentrations unclear
L07	T=19-25.8 *light 97-115* µE rooting phase 6d	T=19.4-20.6 light 87-114 µE	T=22.2-24.8 light: data not available			
L08	T=20 *light 8062-9164lx*	T=21-22 *light 8050-9010lx*	T=19-20 *light 8170-8900lx*	T=20-21 *light 8010 – 8320lx*	T=21 *light 8530 – 880lx*	T=20-21 *light 8210 – 8800lx*
				several measurements disabled due to excessive contaminations with algae, bacteria and snails		
L09	T=19 light data not available Plants in 10 mg/L dead	T=19-19.5 light 7517-7989 lux				
L10				T=21-22.8 light: data not available	T=19-22.9 light 119-163 µE	
					intensive algae growth on sediment, medium slightly turbid	
L11	T=19-20 light 134-153 µE rooting phase 4d	T=19-20 light 140-153 µE rooting phase 4d		T=19.1-20 light 123-145 µE	T=20-21.6 light 123-145µE	
L12				T=22.5-23 8620-10500lx	T=21-22.5 light 8500-12500	T=20-22 light data not available
L13	T=20-24 light 9900-11500lx	T=19-22.8 light data not available	T=18,8-21 light 9880-10520lx			
L14	T=23-24 light 95-141 µE Plants in 10 mg/L dead	T=20.4-22.4 light 120-154 µE	T=20.3-21.4 light122-158 µE	T=23.3-26.3 light 100-193µE	T=20.3-23.2 light 114-157µE	T=19.2-21.7 light 122-154µE
L15	T=19.5-25 light 83-118 µE	T=18.7-22.3 light 95-119µE	T=19.3-24.7 light data not available			

Existing standards about the conduct of ring-tests (DIN ISO 5725-2 & 5) stipulate exclusion of those data sets from tests that are not in line with the ring test protocol from evaluation. However, in this case, removal of test data would drastically reduce the data base. Therefore, all tests were initially included in the statistical analysis with one exception, i.e. the tests with *Myriophyllum spicatum* of laboratory L08. These datasets were excluded due to excessive contamination with algae, bacteria and snails which prohibited several measurements and resulted in the remaining data being questionable.

2.4.4. Biological Measurements

The time schedule of biological measurements is given in Table 6. In addition to the parameters listed therein, lateral branches (if present) were counted and their length measured and roots were assessed semi-quantitatively by assigning each plant to one out of four categories: "very good root development", "moderate root development", "few roots", "no roots". Weight measurements included fresh weight and dry weight determination. The length of the main shoot and lateral branches were combined and named "total shoot length" (TSL). Only TSL and number and total length of lateral branches (TLBL) were reported in the provided EXCEL-templates. The main shoot length at DAT 0 was not explicitly given and could not be recalculated from TSL since TLBL at DAT 0 was not separately reported in the Excel-templates. The statistical analysis therefore focuses on the TSL.

Table 6: Measurement schedule for the environmental and biological variables as given in the test protocol. – measurements were not made on these occasions

Day after treatment (DAT)	*Myriophyllum spicatum*			*Myriophyllum aquaticum*		
	Shoot length	**Shoot weight**	**pH, O_2 visual inspection**	**Shoot length**	**Shoot weight**	**pH, O_2 visual inspection**
0	X	(X) [1)]	X	X	(X) [1)]	X
3	-	-		X	-	X
5	X	-	X	-	-	X
7	-	-		X	X	X
10		-	X	-	-	-
14	X	X	X	-	-	-

1) Representative plants measured

It should be noted that the interim measurements were not intended for quantitative statistical analysis, i.e. a lower precision was acceptable. Generally several laboratories did not report visual observations, it cannot be stated whether in these cases no abnormalities were observed or if they were not reported.

Measured and calculated variables for statistical analysis are listed in Table 7 with the acronyms used in the text.

Table 7: Overview over the measured and calculated variables; italic: variable was subjected to statistical analysis

Measured	**Akronym**
Fresh Weight	FW
Dry Weight	DW
Total Shoot Length	TSL
Number of Lateral Branches	LB
Total Lateral Branches Length	TLBL
Root development (semiquantitative assessment)	R
Calculated	
Yield Fresh Weight	*YFW*
Yield Dry Weight	*YDW*
Yield Total Shoot Length	*YTSL*
Growth Rate Total Shoot Length	*GrTSL*
Growth Rate Fresh Weight	*GrFW*
Growth Rate Dry Weight	*GrDW*

From the measured variables yields (i.e. gain of weight or length) and growth rates (weight or length gain per time unit, on logarithmic scale) were calculated (see below). For all variables yields were evaluated instead of absolute values for the following reason: in cases where the biomass parameter at the start has already a considerable magnitude as in *Myriophyllum,* the treatment-caused inhibition is generally lower than that of the yield. For example, if the control biomass increases by factor two over the exposure duration and a given treatment prevents any growth (i.e. no change in biomass), the inhibition of biomass is 50% at maximum whereas inhibition of yield is 100%. Furthermore, biomass and yield are influenced by time of measurement, which can affect also the toxicity metrics (ECx, NOEC). In contrast, provided that plants exhibit exponential to linear growth over the experimental period and are not limited by nutrient deficiencies or space constraints, the growth rate provides a more consistent reflection of effect that is independent of the timing of underlying assessments. Moreover, the growth rate and its toxic inhibition (ECx, NOEC) can be directly compared between different species.

The yield variables were calculated using Eq. 1 and the growth rates using Eq. 2. In case of FW and DW the biomass at the start of the experiment was derived from a sample of representative plants. The growth rates and their toxicity metrics (ECx; NOEC) allow an easy comparison with other test organisms.

$$Y = B_t - B_0 \quad (1)$$

where
Y: yield; B_t, B_0: biomass (FW, DW, TSL) at start (0) and end (t) of experiment

$$r = \frac{\ln B_{t_2} - \ln B_{t_1}}{t_2 - t_1} \quad (2)$$

where
r: growth rate; B_{t2}, B_{t1}: biomass (FW, DW, TSL) at the end (t2) and another time t1; t1<t2) of experiment; in the present evaluation t1 is DAT 0.

2.5. Data Processing and Consolidation

The experimental data and observations were filled in by the participants into EXCEL-templates provided by the organizers of the ring test. Some of the laboratories also provided the handwritten protocols. In a first step the Excel templates were inspected and obvious inconsistencies were corrected, such as

- use of different decimal separators (comma vs. decimal point),
- deletion/insertion of columns and rows to return to the data structure fixed by the ring-test organizers
- entry of zeros in case no entry was found but zero was the correct value
- entry of missing data in case they were not filled in but were available from the handwritten protocols
- removal of text additions (e.g. cm) in cells with measurement values
- movement of data in case these were entered in wrong place (e.g. if tare and gross columns were interchanged)
- adaption of units to a uniform format (mg/g, cm/mm)

All submitted EXCEL template files were renamed according to a uniform nomenclature composed of abbreviations for the species and test items as well as the lab code. In addition, the laboratory code was inserted in the first sheet “General” of the EXCEL-Templates. A computer program was developed to transfer the data from the Excel templates into a Workbook which can be processed by the ToxRat® Professional software.

In the next step, all data were inspected in order to identify irregularities in the data values such as unusual high or low values. If the hand-written protocols were available, they were inspected in order to verify typing errors, which were then corrected. Several apparent irregularities occurred in fresh weight and dry weight data. However, these irregularities were not altered unless there was evidence of an error from the hand-written protocols.

In some data sets the complete data for fresh weight or dry weight obviously seemed to be erroneous (i.e. zero increase in weight in the control in spite of normal increase in shoot length; or dry weights at DAT 7 threefold lower than at DAT 0). This was the case for the following data sets:

M. aquaticum	L02	3,5-DCP	DW
	L08	3,5-DCP, IP	FW, DW
	L15	3,5-DCP	DW
M. spicatum	L11	3,5-DCP, IP	DW

The named data sets for weights were completely excluded from further evaluation but the corresponding data sets for shoot length of the same test were still regarded for statistical analysis.

All changes and corrections on the very raw data were performed with copies of the original raw data files. The process of data consolidation was documented and the protocol can be obtained upon request.

2.6. Validity Criteria

Validity criteria such as minimum growth rate of biomass factor were not defined for evaluation, since they are intended to be fixed according to the experience and results from the ring-test. Hence all consolidated data sets were included in the initial statistical analysis, regardless of fulfilling any minimum growth rate, biomass increase factor or shoot length increase factor. Depending on the results, preliminary validity criteria were proposed and a second evaluation run was performed based only on valid data. Details of the validity criteria and their impact on data base and results are given in the corresponding section.

2.7. Statistical Analysis

Statistical evaluation of toxicological endpoints was done using the software ToxRat® Professional (section 2.7.1 - 2.7.2). The specific ring-test statistics were calculated according to DIN ISO 5725-2 and DIN ISO 5725-5 using MS-EXCEL (section 2.7.3). MS-EXCEL served also to calculate the probability of two-sample t-test to compare results of the two test species. The EC50 comparisons were performed on the base of the logarithms.

2.7.1. Outlier Testing

Outlier testing for single data, i.e. for a single plant of a treatment in a data set, was performed using the Dixon-Grubbs outlier test ($\alpha = 0{,}01$). When having investigated several data sets, it came out that the test identified a number of outlier-suspicious 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: a relatively high variability including a number of extreme values was found e.g. by all weight variables, although obvious typing errors etc. already had been corrected (see above). Therefore, it may be assumed that the weight values suspected of being outliers are by no means errors in measurement but rather represent methodological difficulties. Excluding extreme values would artificially minimize the variability of the test system. Hence all data that passed the consolidation process was included in the first run of the statistical analysis.

In contrast, outlier testing was performed for *the whole data sets of a laboratory*, i.e. for the arithmetic mean value, EC50 or NOEC for a certain variable. This assessment was done using 99% prediction intervals of the mean values, for details see section 2.7.3 .

2.7.2. Calculation of Simple Statistics, NOEC and EC50

For all data sets of controls and treatments, the arithmetic mean, $\bar{x}$, with its 95%- confidence interval and the standard deviation, s, was calculated for each variable per laboratory and per test. The coefficient of variation, CV%, was then determined as a measure for intra-laboratory variability.

Normal distribution was checked by means of Shapiro-Wilk's test, and variance homogeneity was assessed by means of Levene's test ($\alpha = 0{,}01$). Normal distribution and variance homogeneity of data could be seen as fulfilled throughout the whole data set. Thus, all NOEC and LOEC could be determined by a parametric test. Since it was observed that – against the reasonable theoretical assumption - in some cases the responses were not monotonously decreasing probably due to sampling errors, Williams' test procedure was applied which works with a smoothed order of maximum-likelihood means.

The EC50 was determined by means of a maximum likelihood regression, i.e. iterative re-weighted linear regression of the normal cumulative distribution function (normal CDF) according to Finney (1978), Christensen (1984) and Christensen & Nyholm (1984). In so doing, the normal CDF was linearized by the probit transformation of the observed inhibitions (I = 1 – treatment mean/control mean). In some cases, where a significant fit (checked by an ANOVA) could not be achieved or the confidence limits appeared very wide, the fitting was performed on the basis of the replicate values (I = 1 – treatment replicate/control mean), which increased the number of data points and thus sometimes led to smaller confidence intervals.

In the present report, only EC50 and confidence intervals are presented, that came from significant dose response relations. EC50 values were only calculated within the tested

concentration range, only very moderate extrapolation was performed (for EC50 calculation, highest inhibition should be at least 40% / lowest not higher than 60%).

Both EC50 and NOEC values were calculated using nominal concentrations.

2.7.3. Ring-test Statistics: Repeatability and Reproducibility

According DIN ISO 5725-2 & 5, the statistical analysis of a ring-test aims at providing information about the variance within a single laboratory – the so-called repeatability variance – and the variance between laboratories – the so-called reproducibility variance.

The intra-laboratory variance s^2_r together with the inter-laboratory variance, s^2_L constitutes the so-called reproducibility variance (Eq.3).

$$s^2_R = s^2_r + s^2_L \qquad \text{Eq. 3}$$

Assessment of repeatability necessitates that the whole experiment is repeated by a single laboratory several times. In the present ring-test, only the controls are repeated three times when testing the effects of the three substances. Therefore, the full ring-test statistics is done using the data from controls.

For each variable, based on the individual laboratory-specific standard deviations and coefficients of variation, the mean of s and of CV% was computed which is the repeatability standard deviation, s_r, and repeatability coefficient of variation CV_r%.

With respect to the toxicity metrics (EC50; NOEC) it was not possible to estimate the repeatability variance by the method shown for the control data sets, since each laboratory generated only one EC50 and NOEC per variable and substance. Thus for the EC50 and NOEC the between-laboratory variance could not be calculated under repeatability conditions, it has a different character but it can be seen also as a measure of reproducibility – to keep it simple it will be termed reproducibility, too.

The inter-laboratory variability was calculated as standard deviation of the arithmetic mean of each laboratory from the overall mean after Eq. 4.

$$s_L = \sqrt{\frac{\sum_{i=1}^{p} (\bar{x}_i - \bar{\bar{X}})^2}{p-1}} \qquad \text{Eq. 4}$$

where:

$\bar{\bar{X}}$	overall mean of single laboratory means
i	index of laboratory
$\bar{x}_i$	arithmetic mean of laboratory
p	number of laboratories

Based on Eq. 3, per variable and test substance the reproducibility standard deviation, s_R, and the reproducibility coefficient of variation CV_R% was calculated. The latter is defined by Eq. 5:

$$CV_R\% = \frac{s_R}{\overline{X}} 100 \qquad \text{Eq. 5}$$

In the present ring test, the latter is based on 120 up to 225 control replicates each.

In order to identify possibly outlying data sets the 95-% and 99%-prediction interval (PI) was computed after Eq. 6:

$$p\%PI = \overline{X} \pm z \cdot s_L \qquad \text{Eq. 6}$$

where

$\overline{X}$	overall mean of the arithmetic means or toxicity metrics
p%	95%|99%
z	(1-p) quantile of the standard normal distribution (95%: z = 1,96 | 99%: z = 2,58)

The reproducibility variance of EC50 and NOEC could be estimated via the deviation of a single EC50 and NOEC from the respective overall mean. Since the EC50 and NOEC were assumed to be log-normal distributed, the calculation of $\overline{X}$ and PIs was done using their logarithms. For visualization on the linear scale the antilog of the overall mean of the EC50 and of the NOEC as well as the PIs were used. By this the PIs become asymmetric.

The 95%-|99%-PI predicts that in 100 new toxicity tests 95 | 99 the arithmetic mean of variables as well as the EC50 and NOEC will fall between these limits. Outlying data sets can be easily identified by a chart, in which the overall mean and these two prediction intervals are plotted together with the single laboratory means. If the mean of a data set falls beyond the 99%-PI, DIN ISO-5725 terms it as “statistical outlier”. Data sets, the mean of which is located between the two PIs are termed “stragglers” (nearly outlier).

Data sets were excluded in case the laboratory mean of the data set was beyond the 99%-PI. If after recalculation without the outlying data sets, other outliers were observed, the exclusion procedure was repeated until no further outliers were found.

3. Analytical Results

Analytical results were provided by 7 laboratories. Table 8 gives an overview of the findings. Detailed information is shown in Annex Table 1. In view of the small data base, the conclusions about stability and dissipation of the test item during the experimental period have to be considered with care. Nonetheless, with mean analytical recovery rates of 78,1 % (DAT 7, *M. aquaticum*) and 75,2 (DAT 14, *M. spicatum*) isoproturone proved to be relatively stable during the test period. For 3,5-dichlorophenol mean amounts of 66% (DAT 7, *M. aquaticum*) and 47% (DAT 14, *M. spicatum*) were measured at test termination. Dissipation of the test item was highest in trifluralin (mean recovery rates 9,5% (DAT 7, *M. aquaticum*) and 1,8% (DAT 14, *M. spicatum)).* Concentrations of the test items in the sediment (pore water) were much lower than in the water: Percentages found in sediment (related to nominal concentrations at DAT 0 were 19,9% (IP), 8,2% (TF) and 6,2% (3,5-DCP), at DAT 7 in the test with *M. aquaticum*, while data were not available for the test with *M. spicatum*.

It is out of the scope of the present report to assess these findings with respect to the regulatory use of the present test system. However, the analytical results indicate that, for some test items, measured concentrations are considerably different from nominal concentrations. However, since the ring test statistics focuses on the variability of endpoints and toxicity metrics within and between the laboratories rather than the absolute magnitude of the measured values, it is justified to use nominal concentrations in the present report. Apart from that, measured values are not available from all laboratories.

Table 8: Mean analytical recovery rates [%] calculated from the analytical results of Annex Table 1; s: standard deviation; L: number of laboratories which provided analytical data; n: number of measurements

	Mean Analytical Recovery Rates [%]				
	Myriophyllum aquaticum			*M. spicatum*	
	Culture media		Sediment	Culture media	
Test Item	DAT 0	DAT 7	DAT 7	DAT 0	DAT 14
3,5-DCP	93,7	66,0	6,2	97,2	47,1
s	4,73	15,91	3,13	11,36	23,55
L	3	3	1	4	3
n	13	11	3	20	15
IP	96,6	78,1	19,9	94,3	76,2
s	12,18	19,86	14,68	3,46	5,63
L	3	3	1	2	2
n	11	11	5	10	10
TF	90,4	9,5	8,2	71,5	1,8
s	14,00	2,87	6,18	35,12	2,68
L	1	1	1	1	1
n	5	5	5	4	5

4. Results for *Myriophyllum aquaticum* and Discussion

4.1. Analysis of Control Data

4.1.1. General

The analysis of the control data sets had the following goals:

- to characterize the variability of the measured variables and thus the general practicability of the test
- to characterize the qualification of the measured variables to derive the toxicity metrics
- to identify possible outliers which then need to be excluded from further evaluations.

Up to three sets of control plants per laboratory – one for every test substance – were available to provide ring-test statistics including estimation of the repeatability variance and the reproducibility variance.
The variability aspect is already important at the start of the single toxicity tests, because start values are needed for the calculation of yields and growth rates. Since the weight variables could not be determined in the test plants themselves, a sample of representative plants was weighed (fresh and dry weight) at exposure initiation. Figure 1 shows the arithmetic means of initial fresh weights (representative plants) and lengths (test plants) for each laboratory.

In most cases, the variability of the mean weights of representative plants from different tests within the same laboratory was relatively low. However, there were cases of large variability of initial weights within the same laboratory, namely in the tests of L09 and L14. Variation between laboratories was also large such that the 99%-prediction interval became lower than zero.
Similar patterns of variability were apparent in initial TSL data for test plants (controls). In most cases the initial TSL showed low variability within the same laboratory and in a number of laboratories was not much different. However, in a small number of laboratories the initial TSL of test plants were considerable different in different tests (namely in the tests of L03, L08 and L09). Variation between laboratories was also large.

Possible explanations are (1) general differences in plant quality, (2) methodological difficulties with measurement of fresh weight (possibly due to particles of sediment adhering to roots and skewing the measurement) and – assuming that the initial start length was 6 cm $\pm$ 1 cm as prescribed in the test protocol – (3) variable growth during the rooting phase.

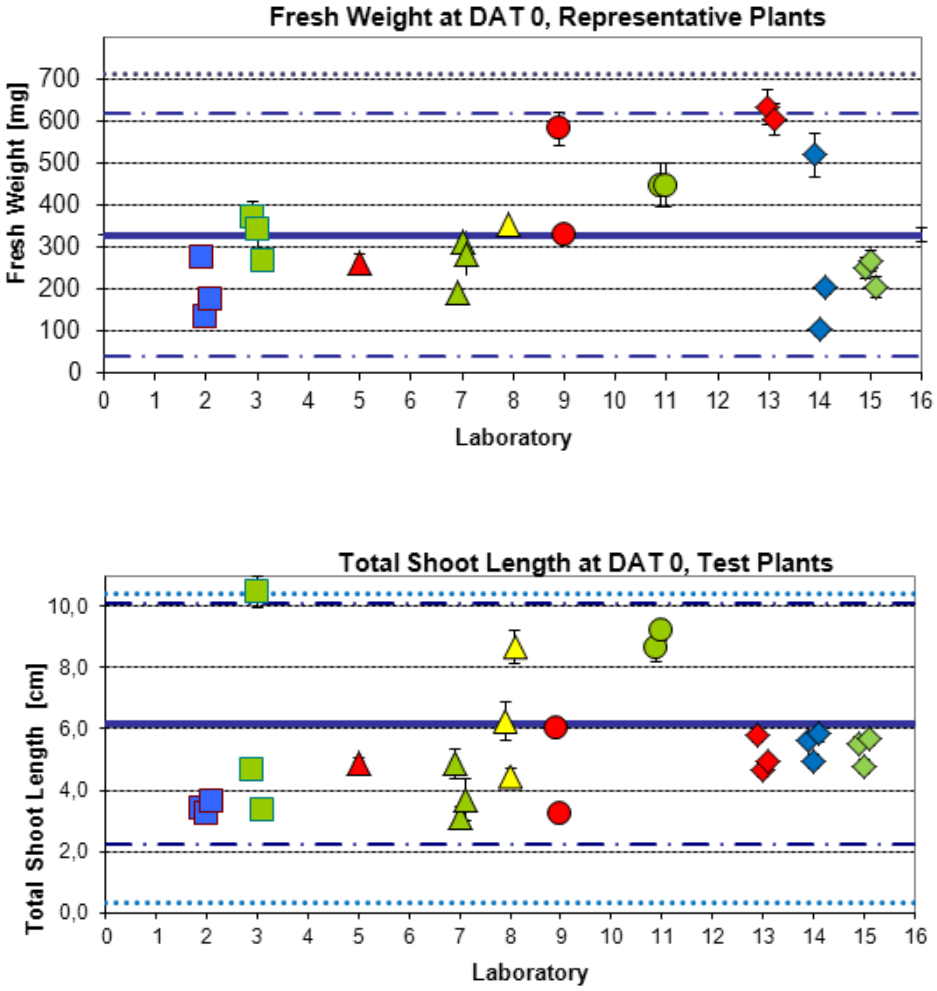


Figure 1: Arithmetic means (symbols) and 95%-confidence interval (whiskers) for the fresh weight of representative plants (top) and total shoot length of test plants in *M. aquaticum* at DAT 0 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line); lower prediction margins below zero are not shown. Data base: all data.

Regardless of the absolute value of initial fresh weight and length it should be ensured that the representative plants from which the initial weights for calculating yields and growth rates are obtained are really "representative", i.e. the ratio of TSL DAT 0 representative plants to TSL DAT 0 test plants should be close to one.
As can be seen from Table 9, only in 13 out of 25 data sets (in one test lengths of representative plants at DAT0 were not provided) ratios close to 1 were found; 3 data sets revealed ratios between 0,8 and 0,9 and in 9 data sets ratios between 1,25 and 2,07 were measured. The latter case leads to overestimation of the start weights of the test plants and thus underestimation – or in worse case even negative - yields and growth rates.

Table 9: Ratio between total shoot length of representative plants and test plants at DAT 0 in *M. aquaticum*; ratios deviating more than 20% from 1 in bold.

Ratio TSL repr.plants / test plants DAT 0			
	3,5-DCP	IP	TF
L02	**2,07**	**1,65**	**1,68**
L03	0,84	0,86	1,00
L05		**1,59**	
L07	0,83	**1,25**	0,95
L08	1,12	**1,44**	0,90
L09		0,97	
L11	1,03	0,97	
L13	**1,52**	1,00	1,00
L14	1,09	1,03	0,97
L15	**1,60**	1,11	**1,55**

4.1.2. Reproducibility and Repeatability

The control data of eight variables (YFW, YDW, YTSL, GrFW, GrDW, GrTSL, LB, TLBL) were analyzed in detail. Figure 2 shows the relative arithmetic means of the control data sets of each laboratory with respect to the overall mean – the estimate of the true mean of the respective variable. Since with *M. aquaticum* only a low number of control plants produced lateral branches at all, arithmetic means are not shown for number of lateral branches and total length of lateral branches (but see below for repeatabilities and reproducibilities).
Means outside of the 99%-prediction intervals (dotted lines in Figure 2) were identified as outliers and the data set for the corresponding variable was excluded from further evaluation. Outlier exclusion was done in three steps:

Step1 : 5 outlier data sets were removed (marked with a red circle in Figure 2):
(read the list below as follows, e.g. first line: Lab 03, test with 3,5-DCP, data set for variable Yield Fresh weight (YFW), Yield Total Shoot length (YTSL) and Growth rate Total Shoot length (GrTSL)

L03	3,5-DCP	YFW, YTSL, GrTSL
L09	3,5 DCP	YDW, GrDW

Step 2: recalculation of the overall mean and the 95%- and 99%-PIs without the above outliers (not shown); removal of 3 new outliers which were located outside of the new narrower 99%-PI:

L03	3,5-DCP	YDW
L03	IP	YFW
L14	TF	GrDW

Step 3: same as step 2 (not shown); again 1 outlier was found and removed:

L03	IP	YDW

Summarizing, 9 out of 144 control-data sets for certain variables were classified as outliers and excluded from further analyses (6 of L03, 2 of L09, 1 of L14; in 7 cases weight parameters were concerned).

The outlier-free ring-test-data sets of the yield and growth rate variables are shown in Figure 3 and Figure 4, respectively (for numerical data see Table 10). Nonetheless, the variability of the remaining laboratory means is high and some of the means are located outside the 95%-prediction interval and thus contributing a lot to variability.

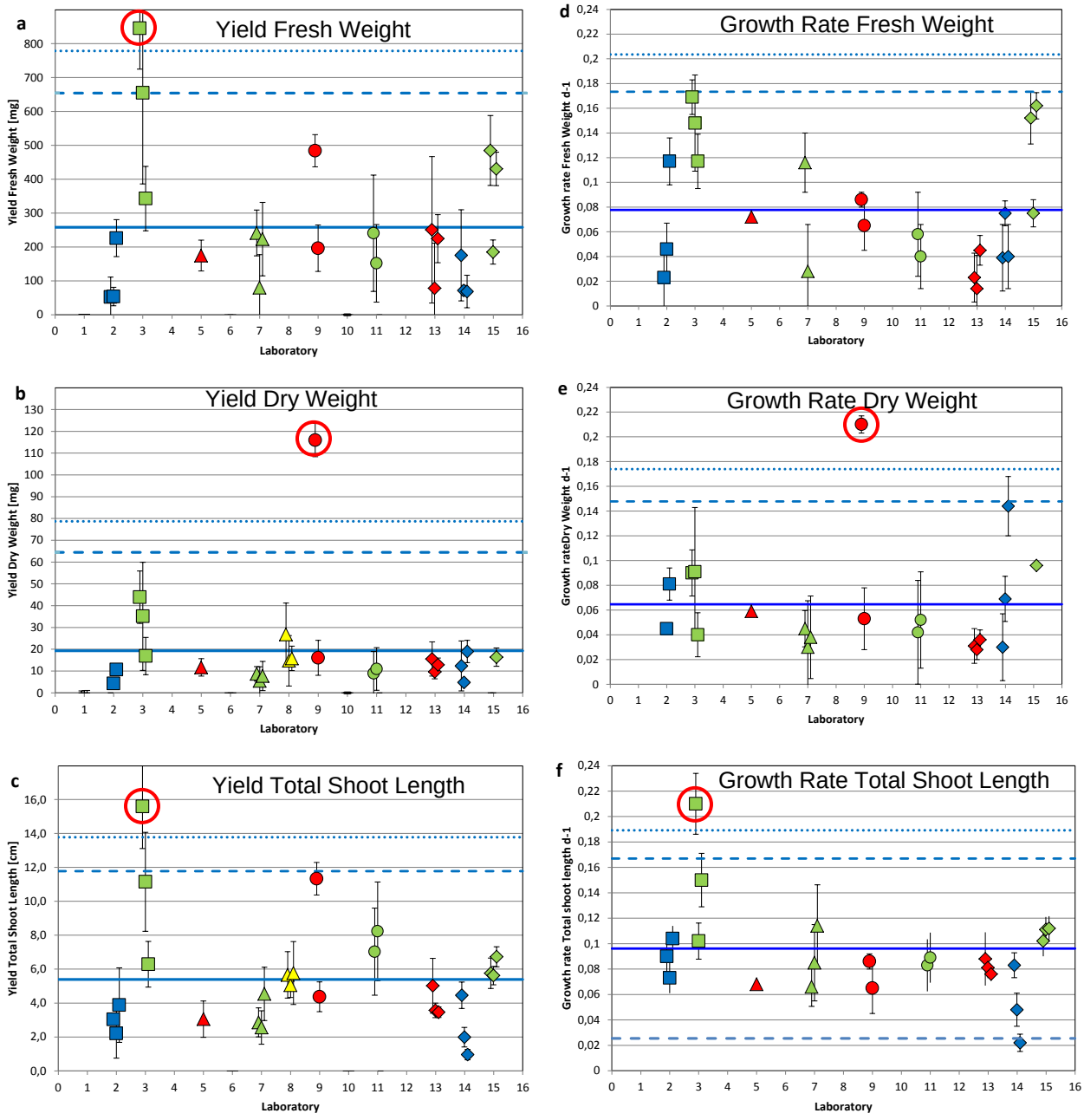


Figure 2: Arithmetic means (symbols) and 95%-confidence interval (whiskers) obtained for yield variables (a, b, c) and growth rates (d, e, f) in *M. aquaticum* in control data sets of each laboratory at DAT 7 together with the overall mean (thick blue line), the upper 95%-prediction margin (dashed line) and the upper 99%-prediction interval (dotted line); outlier data sets are included and marked with a red circle; lower prediction margins below zero are not shown. Data base: all data.

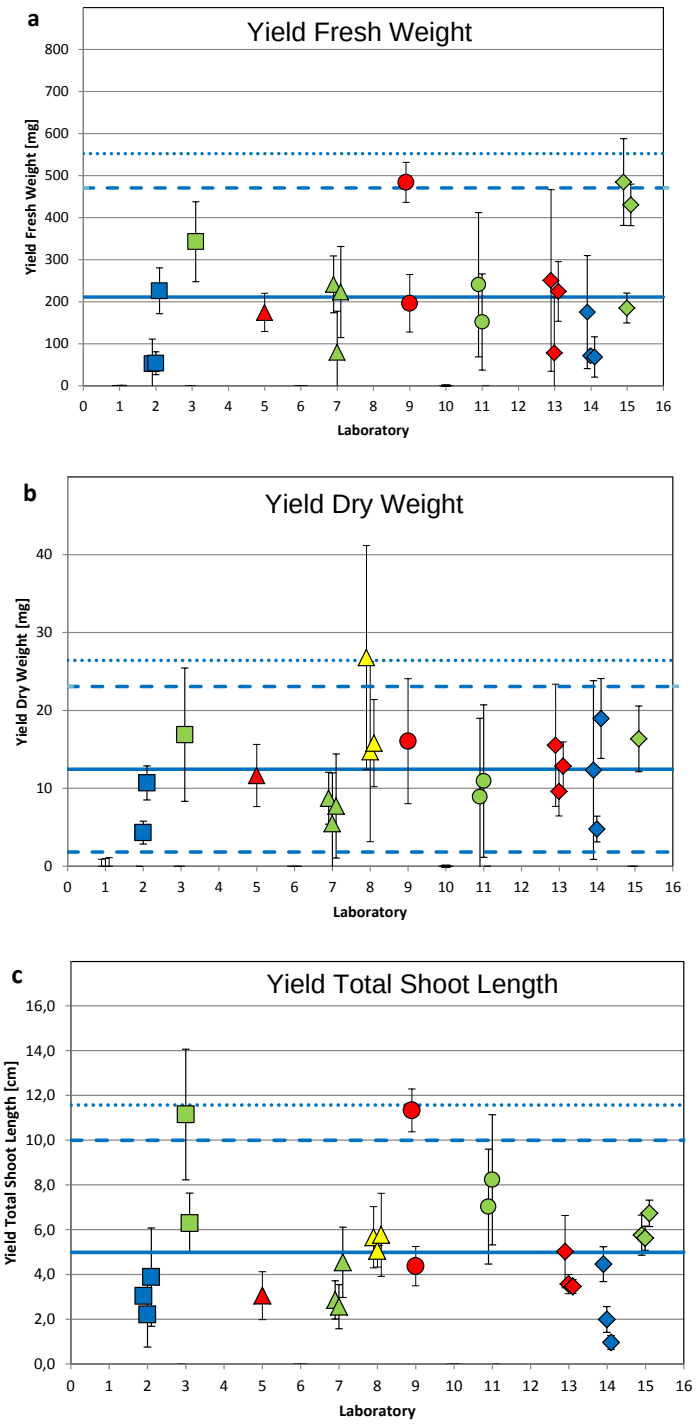


Figure 3: Arithmetic means (symbols) and 95%-confidence interval (whiskers) obtained for yield variables in *M. aquaticum* in the control data sets of each laboratory at DAT 7 together with the overall mean (thick blue line), the 95%-prediction margin (dashed line) and the 99%-prediction margin (dotted line) after exclusion of statistical outliers; lower prediction margins below zero are not shown. Data base: all data.

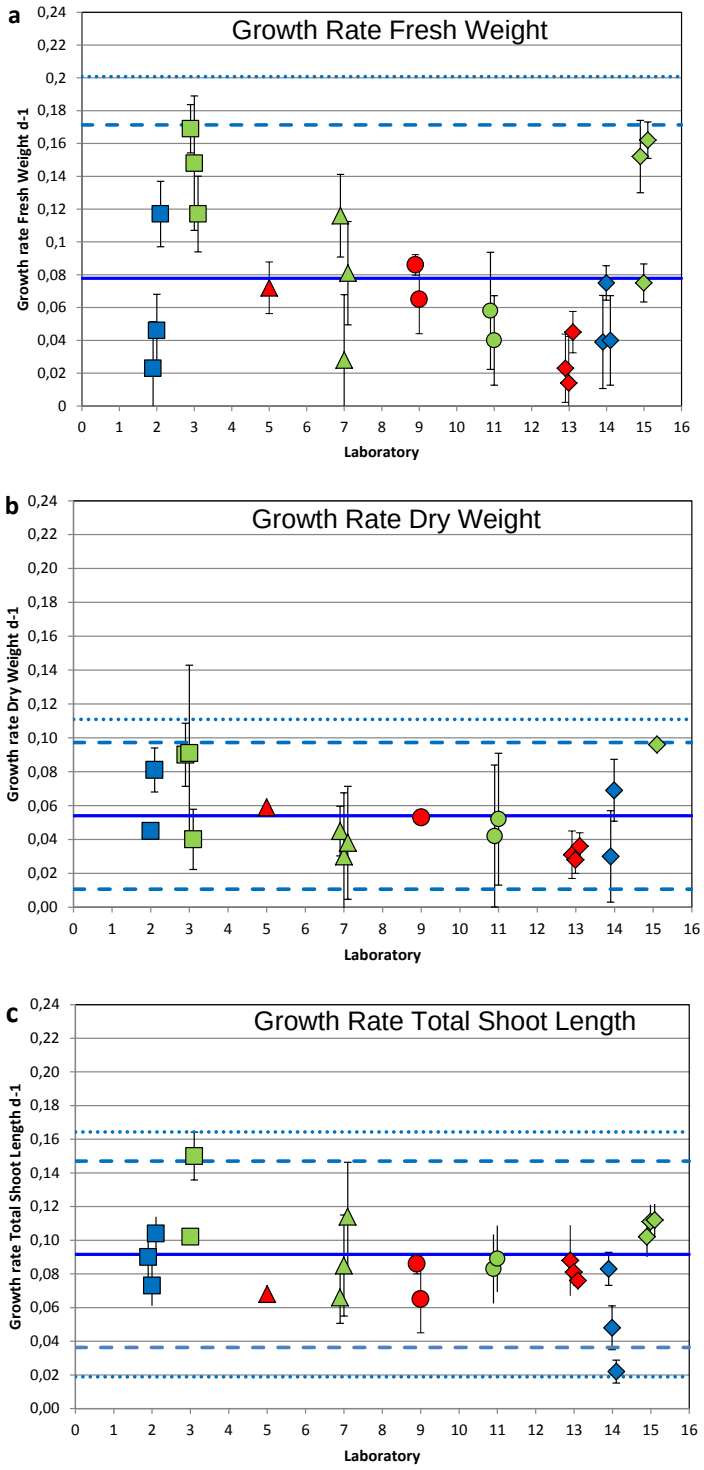
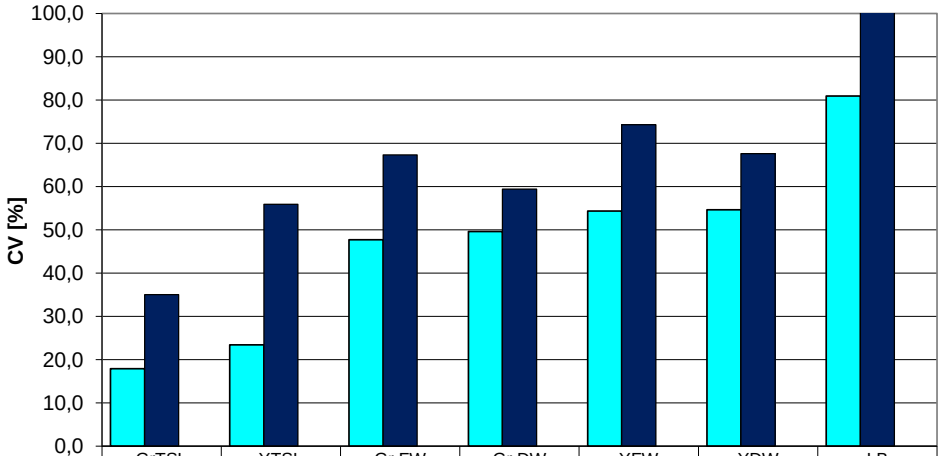


Figure 4: Arithmetic means (symbols) and 95%-confidence interval (whiskers) for the growth rates in *M. aquaticum* in the control data sets of each laboratory at DAT 7 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line) after exclusion of statistical outliers; lower prediction margins below zero are not shown. Data base: all data.

The reproducibility (inter-laboratory variability) and repeatability (intra-laboratory variability) of all variables are quantified by means of coefficients of variation (Figure 5 and Table 10). The repeatability determines the detectable magnitude of effect and thus, constitutes a measure for the statistical power of the NOEC that can be derived (see below). The relationship between reproducibility and repeatability is a measure of the current degree of standardization of the test system.

The values for the number of lateral branches (LB) are extremely variable in terms of repeatability and reproducibility. As a consequence, the same was true for the lateral branches length (TLBL, not shown). Coefficients of variation ranging from 81% to far more than 100% do not enable determination of toxicity metrics (such as EC and NOEC) with sufficient statistical power. The high variability was due to the fact that lateral branches were only occasionally produced, i.e. only in a small number of control plants lateral branches occurred at all. Therefore LB and TLBL were excluded from further evaluation of toxic metrics for *M. aquaticum*.

For the remaining variables, irrespective of the absolute values, reproducibility variability is always higher than repeatability variability. This demonstrates that the standardization of test conditions, plant quality and handling can be furthermore improved. Table 5 shows that in several tests deviations from the protocol were found. This might explain part of the variability observed. However, for the yields and growth rates 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 of representative plants. In view of this handicap, the comparability of representative plants to test plants should be even more ensured.



	GrTSL	YTSL	Gr FW	Gr DW	YFW	YDW	LB
CVr%	17,9	23,4	47,7	49,6	54,3	54,6	80,9
CVR%	35,0	55,9	67,3	59,4	74,3	67,6	127,5
n	25	25	23	21	21	20	23

Figure 5: Coefficients of variation for repeatability (CV_r%, light blue) and reproducibility (CV_R%, dark blue) of all control data sets in *M. aquaticum* ordered by magnitude; n: number of control data sets. Data base: all data (except outliers).

Overall, the variability appears considerably high in view of averaging the values of three plants of a single pot which lowers variability. Reproducibility/repeatability variability factors of 2,0 and 2,4 in GrTSL and YTSL , respectively, are possibly caused by variable experimental conditions (source of the plants, temperature and light conditions, season). As will be shown below, this will not markedly influence the computation of the toxicity metrics since all findings are related to the respective controls.
Coefficients of variation repeatability in the weight variables were around 50%. It appears that these results are caused both by methodical problems in determining the weight and by the highly variable starting conditions (i.e. ratios between weights of representative plants and of test plants, see above).

Table 10: Overview over the ring-test statistics of controls in *M. aquaticum*; overall mean: mean of all data from all controls and all laboratories; CV_r%: repeatability coefficient of variation; CV_R%: reproducibility coefficient of variation; 95%-/99%-PI: 95%-/99%-prediction interval; p: number of control data sets. Data base: all data (except outliers).

	Overall mean	CV_r%	CV_R%	CV_R%/ CV_r%	95% -PI	99%-PI	p
YFW [mg]	211,2	54,3	74,3	1,4	-48,3 – 470,7	-129,8 - 552,2	21
YDW [mg]	12,5	54,6	67,6	1,2	1,8 - 23,1	-1,5 - 26,4	20
YTSL [cm]	5,0	23,4	55,9	2,4	-0,019 – 10,0	-1,6 - 11,6	25
GrFW d^{-1}	0,078[3]	47,7	67,1	1,4	-0,016 – 0,171	-0,045 - 0,201	23
GrDW d^{-1}	0,054[4]	49,6	59,4	1,2	0,011 – 0,097	-0,003 - 0,111	21
GrTSL d^{-1}	0,092[5]	17,9	35,0	2,0	0,036 – 0,147	0,019 - 0,164	25

4.1.3. Minimal Detectable Difference and Statistical Power

The repeatability determines the detectable magnitude of effect and thus, constitutes an indicator for the statistical power of the NOEC that can be derived via the minimum detectable difference (MDD). Based on the repeatability coefficients of variation and the given test design (6 control replicates, 5 treatments with 3 replicates each), it is possible to calculate the expected minimum percentage difference to the control (%MDD) that can be detected by a statistical test.

[3] Corresponding to a doubling time of 8,9 days
[4] Corresponding to a doubling time of 12,8 days
[5] Corresponding to a doubling time of 7,5 days

Since the real ring test data were found to be in line with the prerequisites of parametric tests (normal distribution, variance homogeneity), MDD calculations were performed for Williams'-test procedure being one of the most powerful ones.

The calculated minimum-detectable effect-levels based are listed in Table 11. Since they are based on mean repeatabilities and a theoretical test design, they represent a theoretical orientation. Depending on the individual variances and the number of replicates in an individual test, the MDDs actually obtained in a certain test run can be lower, but also higher than those given in Table 11. The detectable effect levels actually obtained in the present ring test are presented together with the NOECs (see sections 4.2.2.2, 4.2.3.2, and 4.2.4.2).

Table 11: Detectable effect levels (MDD%) for different variables in *M. aquaticum* using Williams´test. MDDs calculated assuming the given test design (control n = 6; 5 treatments with n = 3) and mean control variances (i.e. repeatabilities) as obtained in the present ring test (see Table 10). CV_r% = repeatability coefficient of variation P = number of control data sets from which mean CV_r% was calculated

	CV_r%	MDD%	p
YFW	54,3	69,6	21
YDW	54,6	69,9	20
YTSL	23,4	29,9	25
GrFW	47,7	61,0	23
GrDW	49,6	63,4	21
GrTSL	17,9	22,9	25

For the given design and the variable with the lowest coefficient of variation (GrTSL, 18%), an effect has to be at least as great as ca. 23% to become statistically significant by means of Williams' test. Since ideally, the NOEC should not exceed the EC20, in order to ensure that the test has sufficient statistical power the variability of the measurement parameters should be minimized as much as possible and /or highly reproducible variables should be selected for measurement.

Overall, all MDDs are relatively high. All weight variables show MMDs greater than 60%, which is not desirable for sound NOEC determination. Several specifications concerning initial plant length and weight and minimum requirements for variability of initial weights were already recommended in the test protocol. It should be clarified whether they simply were not always followed or whether they need to be enhanced. In any case, the definition of appropriate validity criteria is essential (see section 4.2.6).

4.2. Reproducibility of Toxic Metrics and Sensitivity of Variables

4.2.1. General

This chapter deals with the reproducibility of the toxic effects of 3,5-DCP, IP and TF on the six variables YFW, YDW, YTSL, GrFW, GrDW and GrTSL. The repeatability cannot be determined since each laboratory only performed one toxicity test per substance. Root length was assessed qualitatively and will be considered in the last section of this chapter.

Please note that the EC50 and NOEC follow a log-normal distribution. Therefore, the calculation of the overall mean and the prediction intervals were performed on the logarithmic scale and then transformed to the linear scale. On a linear scale, the overall mean of the toxicity metric is a geometric mean and the prediction interval becomes asymmetric.

Due to the high variability and the selected concentration ranges the EC50 and NOEC could not always be determined. According to DIN ISO 5725-2 and DIN ISO 5725-5 at least 8 results are necessary for the assessment of reproducibility. Nonetheless, even if this requirement is not fulfilled all results will be reported.

Since validity criteria were not defined prior to the initial assessment, all data, except the outliers described in the previous section, were included in this evaluation.

4.2.2. Effects of 3,5-dichlorophenol

4.2.2.1. EC50

Figure 6 presents the results for the EC50 of 3,5-DCP. A significant concentration/response relationship was not observed in all datasets therefore for some variables only a limited number of EC50 are shown. Except for the YTSL and GrTSL, in which 8 test data sets delivered a valid EC50, the number of EC50s in the remaining variables in a strict sense was too low to the EC50-reproducibility (Table 12). It appears that the supposed methodical problems in weight determination causing high variability affected also the successful computation of the EC50.
No outlying results were found and all EC50s were located within the 95%-prediction interval. Obviously, the high inter-laboratory variability of the variables did not visibly affect the EC50-level. A reason for this is seen the fact that the response is expressed as inhibition relative to the controls. Table 12 and Figure 7 present the overall means and 95%-prediction intervals for each assessment parameter in order of sensitivity. Figure 7 also shows the confidence intervals of the single EC50 (whiskers). The validity of this order, however, is questionable, since only YTSL and GrTSL fulfill the requirement for the minimum number of results. Moreover, the sensitivity differences between variables appear to be small and the mean EC50 ranged between 3,8 and 5,0 mg/L.
The fresh weight parameters showed the best reproducibility (i.e. the smallest prediction intervals of the mean EC50). Regarding also the number of evaluable data sets at all and the confidence intervals of the single EC50 (see whiskers in Figure 6), also the yield and growth rate of total shoot length appear to be suitable variables.

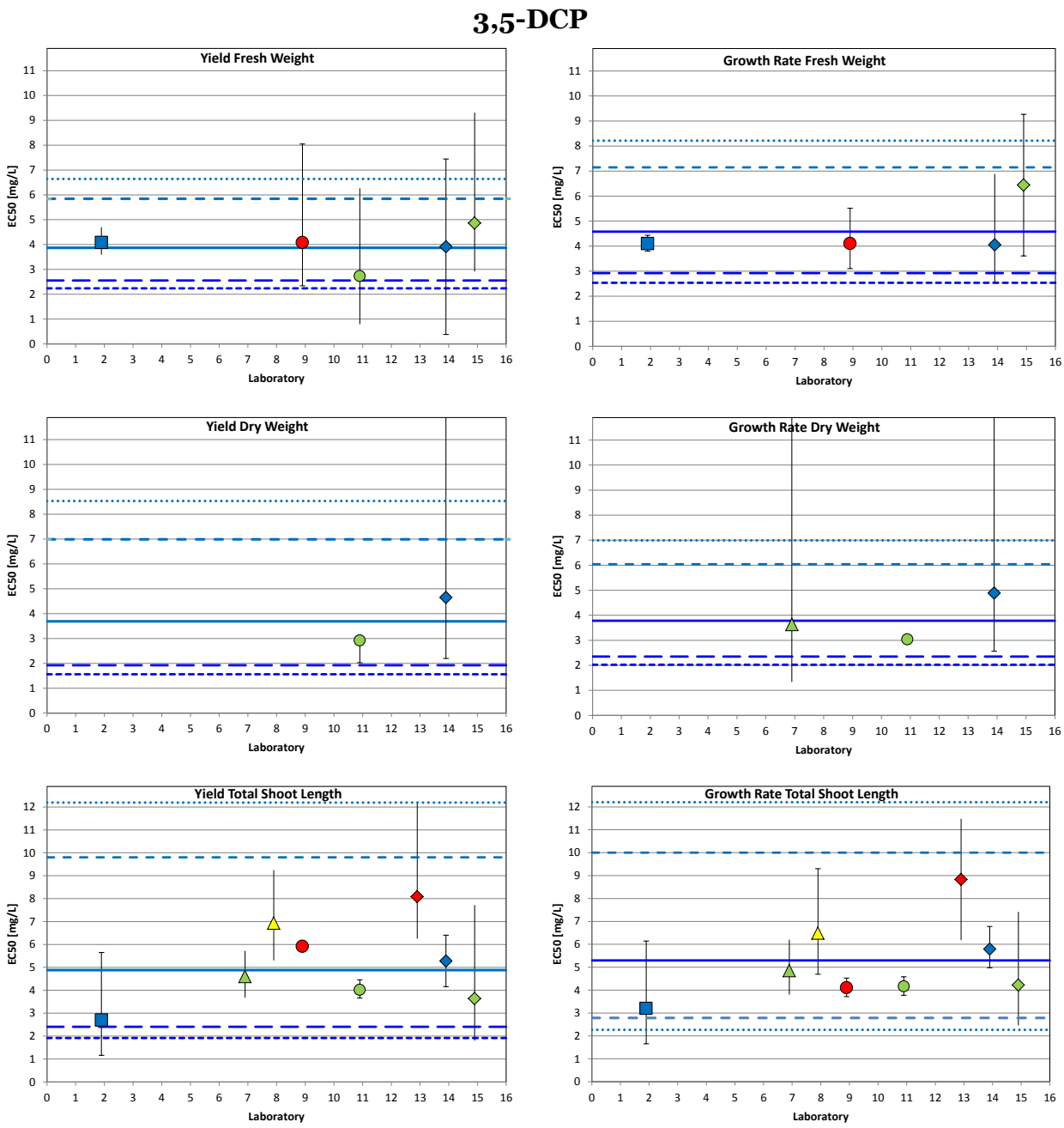


Figure 6: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield variables (left hand) and the growth rates (right hand) in presence of 3,5-DCP in *M. aquaticum* at DAT 7 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).

Table 12: Overall mean of the EC50 for 3,5-DCP, number of determined EC50 (n) and the 95%-prediction interval (PI) for the yield and growth rate variables in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	3,8	2	2,0 - 7,1
GrDW	3,9	3	2,5 - 6,1
YFW	4,0	5	2,7 - 5,9
GrFW	4,7	4	3,0 - 7,2
YTSL	5,0	8	2,5 - 9,9
GrTSL	5,0	8	2,9 - 10,1

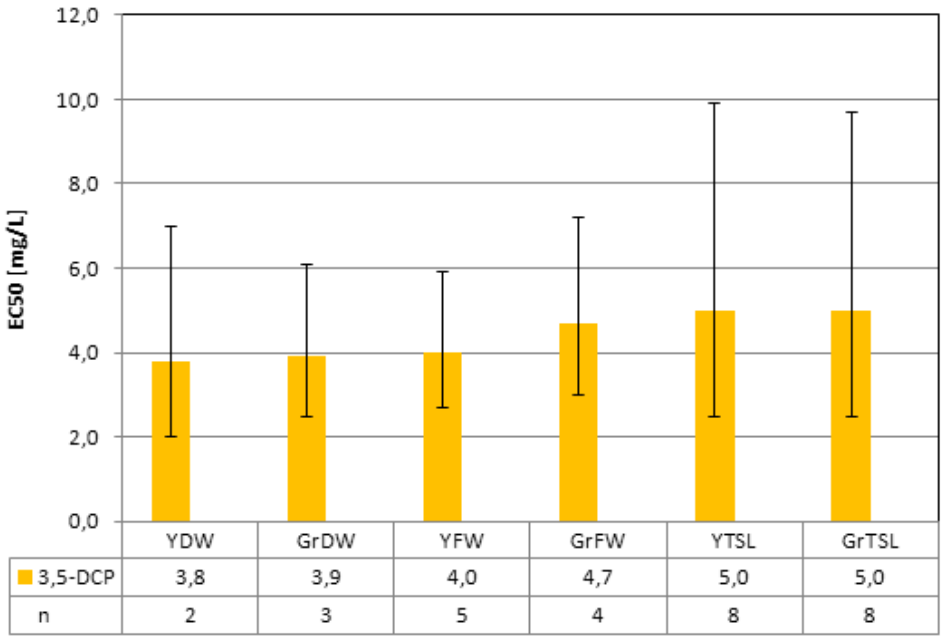


Figure 7: Mean EC50 (bars) and 95%-prediction interval (whiskers) for the variables studied in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

4.2.2.2. NOEC

Figure 8 presents the NOECs obtained for 3,5-DCP. All NOECs were computed using Williams' test, i.e. normal distribution and variance homogeneity were fulfilled (α=0,01). Note that the NOEC can only be one of the applied exposure concentration values. NOEC testing is very sensitive to variability, thus in several data sets, MDDs were 100% (Table 13) - 100% MDD means "statistical detection of any effect impossible". The highest number of NOECs were determined in the shoot length variables, nevertheless the number of NOEC per variable was too low to assess the NOEC-reproducibility according to the ISO criteria (Table 13).
Outlying results were not found; all NOECs were located within the 95%-prediction interval and ranged over three test concentration at the most. As with the EC50, the high inter-laboratory variability of the variables did not visibly affect the NOEC-level. Table 13 and Figure 9 present the overall means and 95%-prediction intervals in order of sensitivity. The NOEC ranged between 2,9 and 5,0 mg/L. As with the EC50, the sensitivity differences appear to be small and the importance of this order is questionable.
The MDDs averaged from the test laboratories (Table 13) correspond well with the expected values (Table 11). On average, again the shoot length variable YTSL and GrTSL showed the most acceptable mean MDD values of 25% and 21%, respectively. Three laboratories even achieved low MDDs between of 9% and 15% for TSL parameters. This result shows that, in principle, the problems with high variability can be reduced.

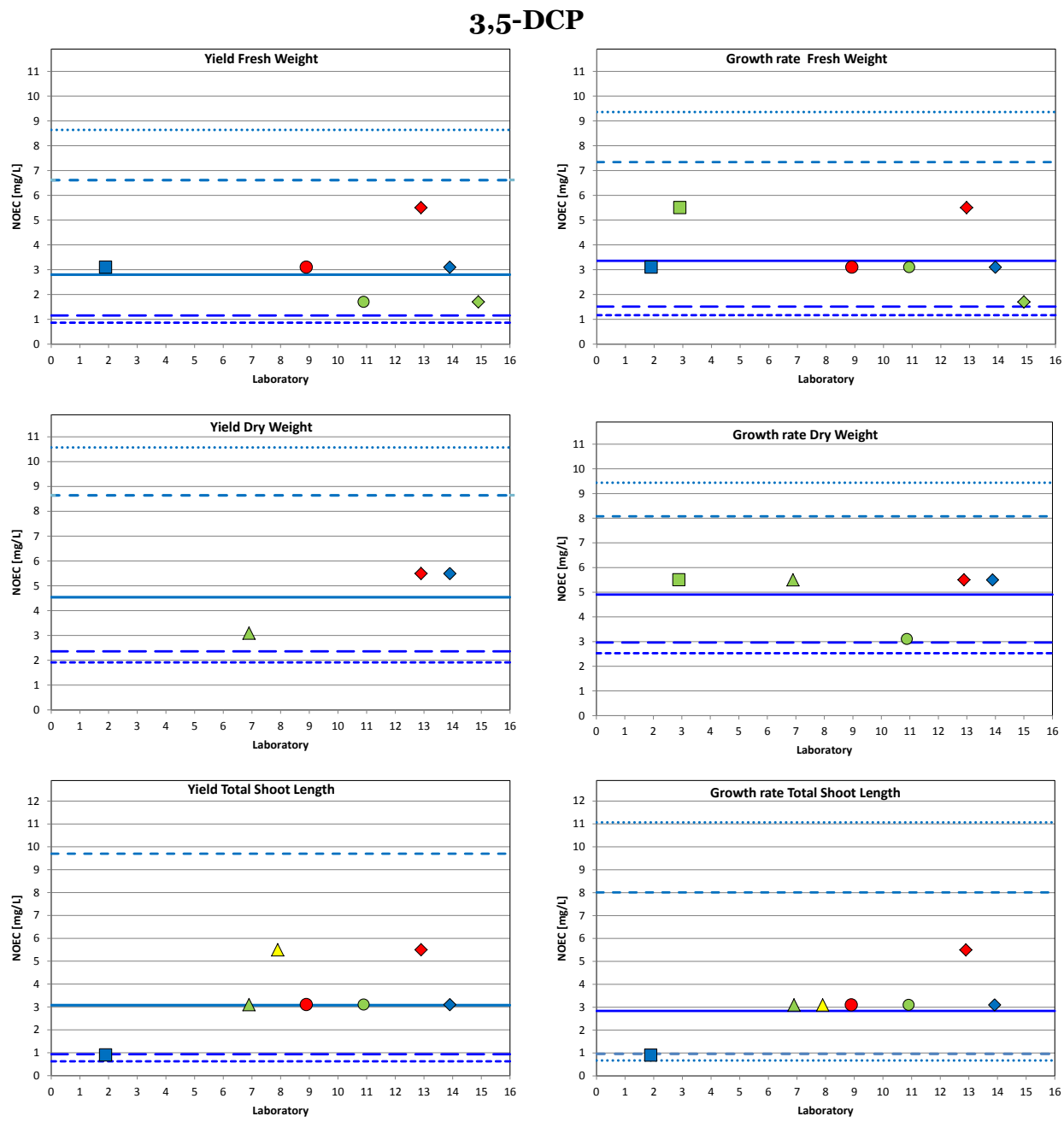


Figure 8: NOEC (symbols) obtained for the yield variables (left hand) and the growth rates (right hand) for 3,5-DCP in *M. aquaticum* at DAT 7 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).

Table 13: Overall mean of the NOEC for 3,5-DCP, number of determined NOEC (n), the 95%-prediction interval (PI) for the yield and growth rate variables in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

3,5 DCP							
Variable	NOEC [mg/L]			MDD[1]			remarks
	Geom. mean[2]	n	95% PI	mean	range	n	
YFW	2,9	6	1,3 - 6,7	57%	22% - 100%	7	1 x NOEC ≥ 10 mg/L 1 x MDD > 100%
GrTSL	2,9	7	1,1 - 8,1	21%	9% - 33%	8	1 x NOEC < 1 mg/L
YTSL	3,2	7	1,0 - 9,8	25%	12% - 40%	8	1 x NOEC < 1 mg/L
GrFW	3,5	7	1,6 - 7,4	51%	15% -100%	8	1 x NOEC ≥ 10 mg/L 1 x MDD > 100%
YDW	4,6	3	2,5 - 8,8	81%	48% - 100%	5	2x NOEC ≥ 10 mg/L 2 x MDD > 100%
GrDW	5	5	3,1 - 8,2	75%	46% - 94%	6	1 x NOEC ≥ 10 mg/L

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")
2) NOECs ≥ 10 mg/L and < 1 mg/L not included

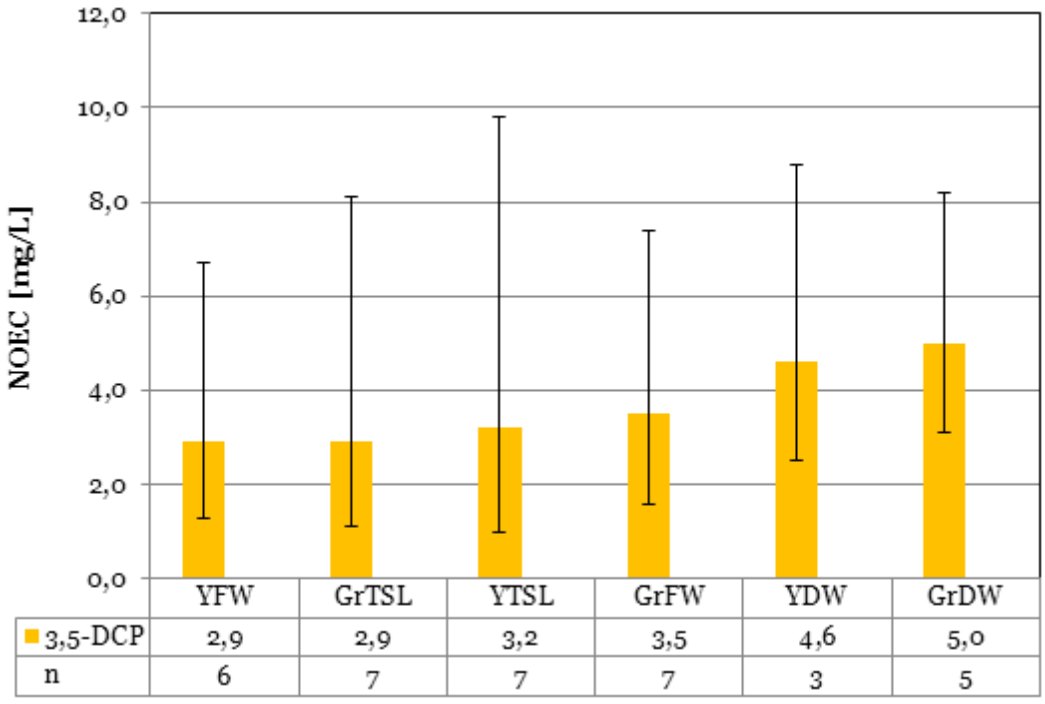


Figure 9: Mean NOEC (bars) and 95%-prediction interval (whiskers) for the variables studied for 3,5-DCP in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

4.2.3. Effects of Isoproturone

4.2.3.1. EC50

Table 14 and Figure 11 present the results for the EC50 of IP. The impression of the calculated values is very similar to 3,5-DCP and similar descriptions and discussions concerning the data base can be made here and need not to be repeated. Differences include the EC50-level, which was lower by around one order of magnitude, and the sensitivity range of the variables, which extend to more than factor 10 (lowest EC50: YDW, 0,06 mg/L; highest EC50: GrFW, 0,77 mg/L). But, as for 3,5-DCP regarding the low numbers for the weight variables, the validity of the mean EC50 and the sensitivity order is questionable.
Some of the problems to find an EC50 were due to growth promotions being observed in the three lowest concentrations in some cases. This was only observed in the weight parameters, not in the shoot length parameters.
It appears that for a regression design the selection, spread and number of concentrations was not optimal. These factors combined with methodological problems in root and, therefore, total plant weight determination caused high variability and affected the successful computation of the EC50.

Table 14: Overall mean of the EC50 for IP, number of determined EC50 (n) and the 95%-prediction interval (PI) for the yield and growth rate variables for IP in M. aquaticum at DAT 7. Data base: all data (except outliers).

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,06	2	0,01 - 0,33
GrDW	0,07	2	0,01 - 0,37
YFW	0,62	4	0,19 - 1,97
GrFW	0,77	4	0,22 - 2,73
YTSL	0,32	8	0,11 - 0,98
GrTSL	0,47	8	0,17 - 1,25

IP

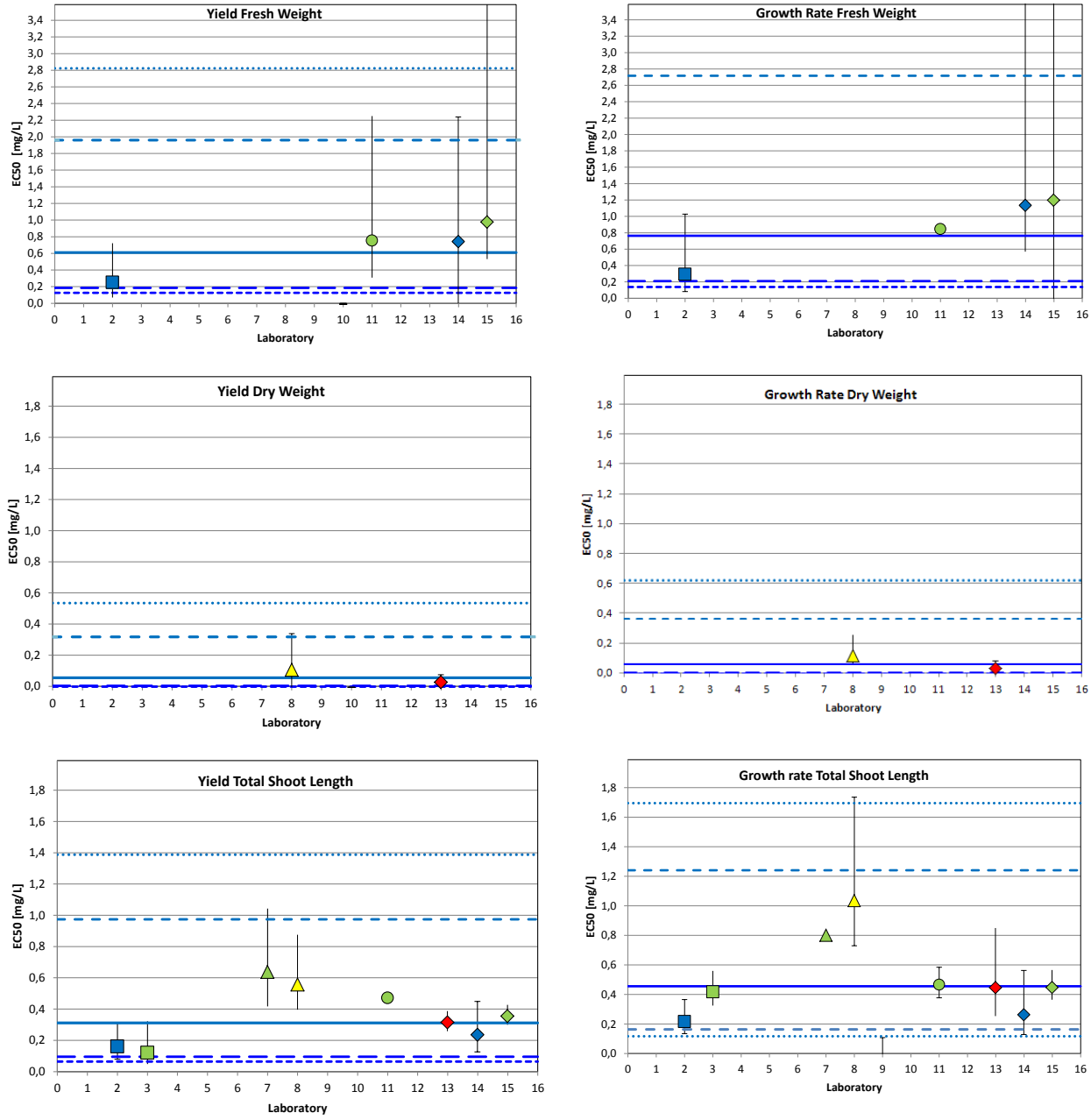
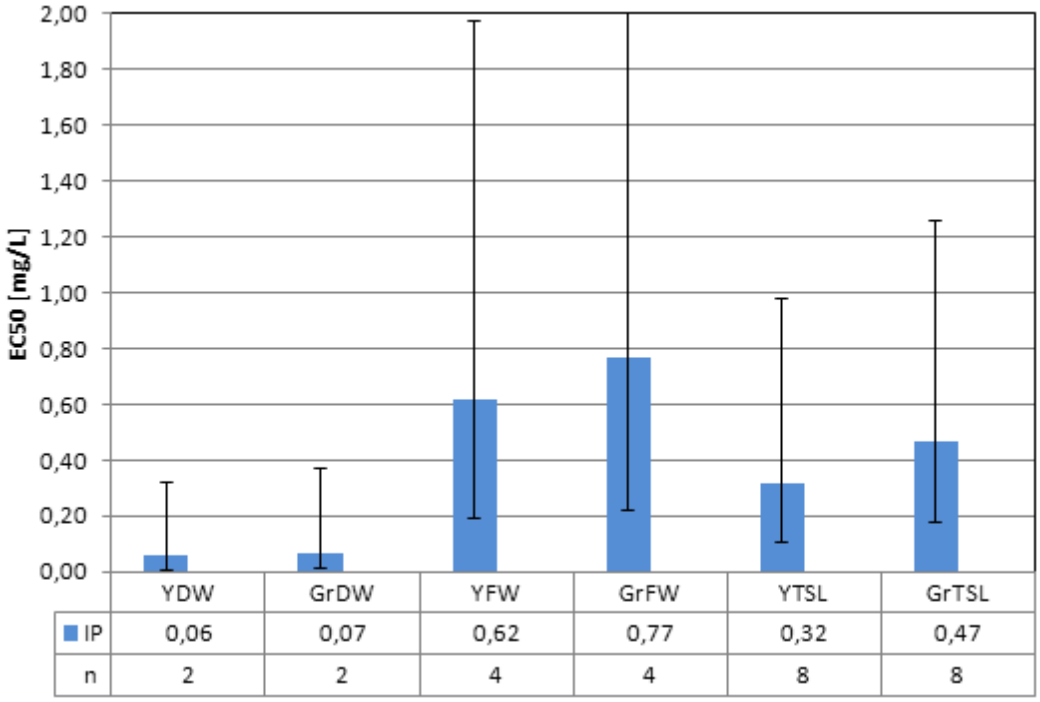


Figure 10: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield variables (left hand) and the growth rates (right hand) for IP in *M. aquaticum* at DAT 7 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).



	YDW	GrDW	YFW	GrFW	YTSL	GrTSL
IP	0,06	0,07	0,62	0,77	0,32	0,47
n	2	2	4	4	8	8

Figure 11: Mean EC50 (bars) and 95%-prediction interval (whiskers) for the variables studied for IP in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

4.2.3.2. NOEC

Figure 12, Table 15 and Figure 13 present the results for the NOEC of IP. Descriptions and discussions related to applied test and obtained MDDs are similar than that for 3,5-DCP and shall not to be repeated again. Differences between the NOEC s of IP and 3,5-DCP include the NOEC-level (about 20fold lower) and the sensitivity range of the variables (difference by around factor 2) (lowest NOEC: YTSL, 0,069 mg/L; highest NOEC: GrFW, 0,147 mg/L). Again, in view of the low numbers of weight variables the results should be regarded carefully. The parameters YTSL and GrTSL fulfill the ISO-requirement for the minimum number of results necessary for ring-test statistics, produced the lowest NOECs and provided the highest statistical power (Table 15).

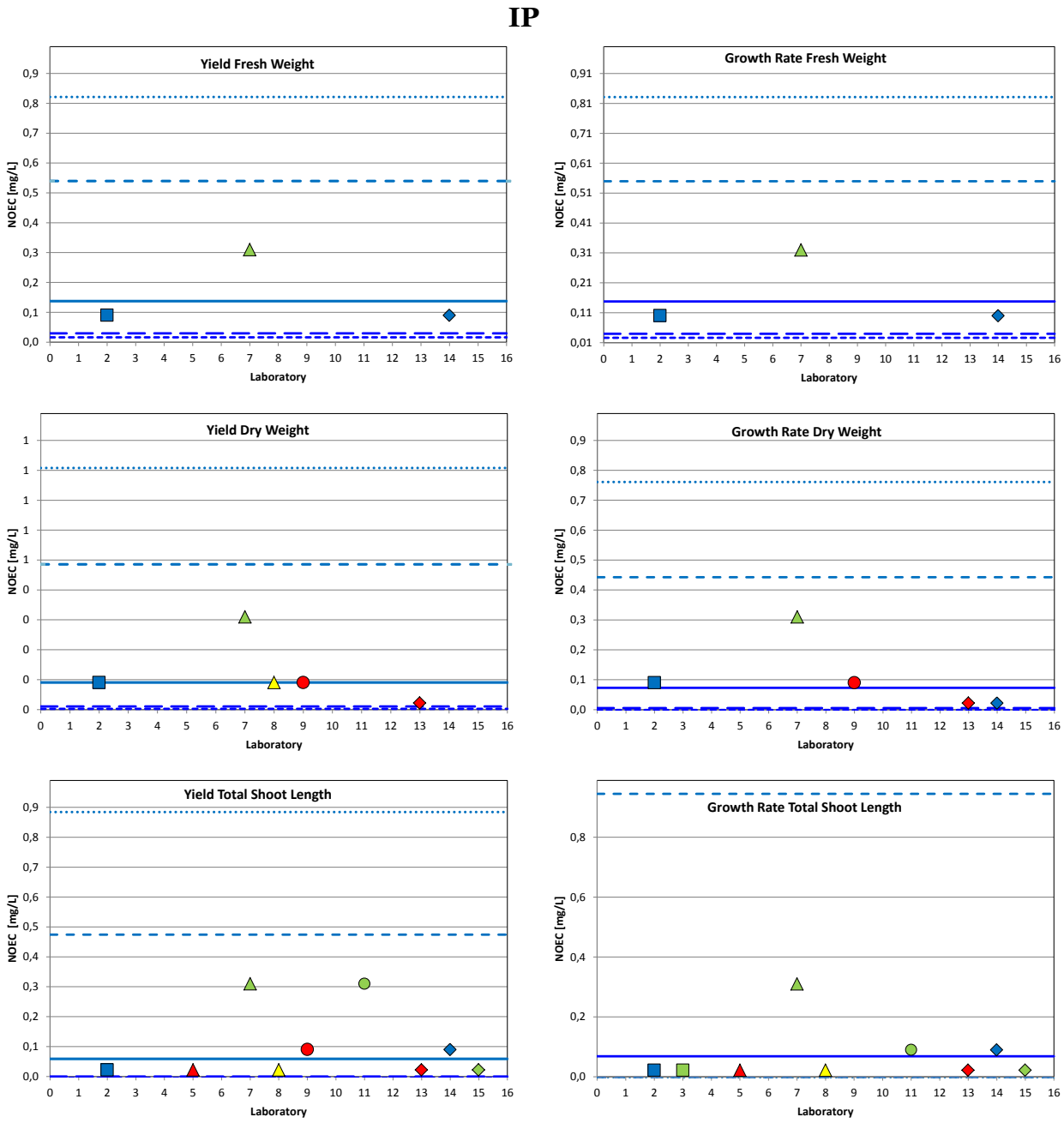


Figure 12: NOEC (symbols) obtained for the yield variables (left hand) and the growth rates (right hand) for IP in *M. aquaticum* at DAT 7 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).

Table 15: Overall mean of the NOEC for IP, number of determined NOEC (n), the 95%-prediction interval (PI) for the yield and growth rate variables for IP in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

IP							
Variable	NOEC [mg/L]			MDD[1]			remarks
	geom. mean[2]	n	95% PI	mean	range	n	
YTSL	0,069	9	0,01 - 0,48	27%	11% - 43%	9	1 x NOEC < 0,01
GrTSL	0,079	9	0,006 - 0,96	23%	19% - 23%	9	
GrDW	0,083	6	0,02 - 0,45	60%	34% - 100%	8	1 x NOEC ≥ 1 mg/L 2 x NOEC < 0,01 mg/L 2 x MDD > 100%
YDW	0,100	5	0,02 - 0,49	65%	36% - 100%	8	2 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L 3 x MDD > 100%
YFW	0,147	3	0,04 - 0,55	63%	32% - 100%	9	4 x NOEC ≥ 1 mg/L 3 x MDD > 100%
GrFW	0,147	3	0,04 - 0,55	54%	24% - 100%	8	2 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L 2 x MDD > 100%

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")
2) NOECs ≥ 1 mg/L and < 0,01 mg/L not included

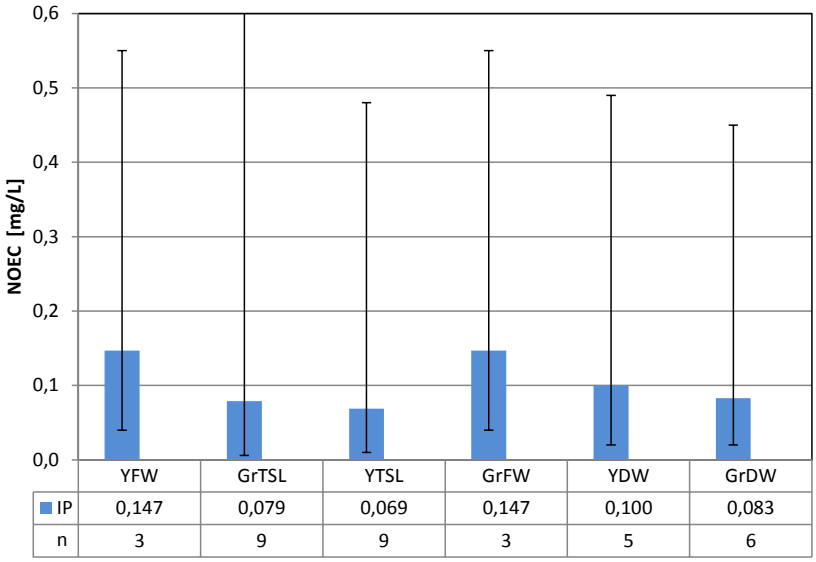


Figure 13: Mean NOEC (bars) and 95%-prediction interval (whiskers) for the variables studied for IP in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

4.2.4. Effects of Trifluralin

4.2.4.1. EC50

Figure 14, Table 16 and Figure 15 present the results for the EC50 of TF. Several data sets failed to provide EC50 values due to the absence of a consistent dose-response-relationship (inconsistent order of inhibitions rather than a monotonous increase , i.e. inhibitions subsequent to treatments with growth promotion). Moreover calculated maximum inhibitions often were below 50%. EC50 values were only obtained for a small number of tests for shoot length variables (0,60 mg/L and 0,67 mg/L). Sensitivity differences of the variables cannot be stated. In addition to the methodological problems already discussed with TF also the dissipation of the substance was probably responsible for these results (see the analytical results in section 3).

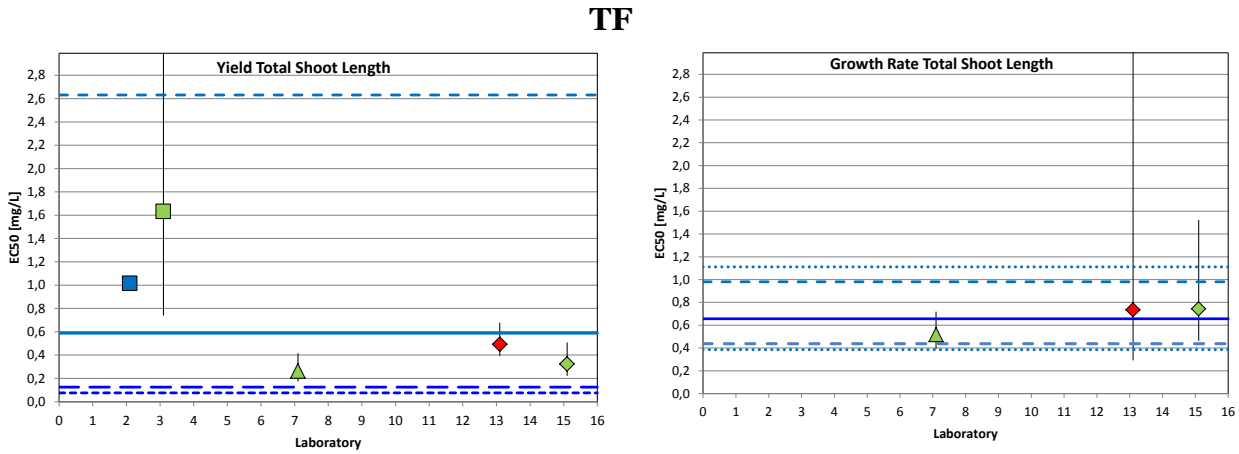


Figure 14: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield of TSL (left hand) and the growth rate of TSL (right hand) for TF in *M. aquaticum* at DAT 7 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).

Table 16: Overall mean of the EC50 for TF, number of determined EC50 (n) and the 95%-prediction interval (PI) for the yield and growth rate of TSL for TF in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to missing concentration/response relation or not significant due to too few / low inhibition values		
GrDW			
YFW			
GrFW			
YTSL	0,60	5	0,14 - 2,64
GrTSL	0,67	3	0,45 - 0,99

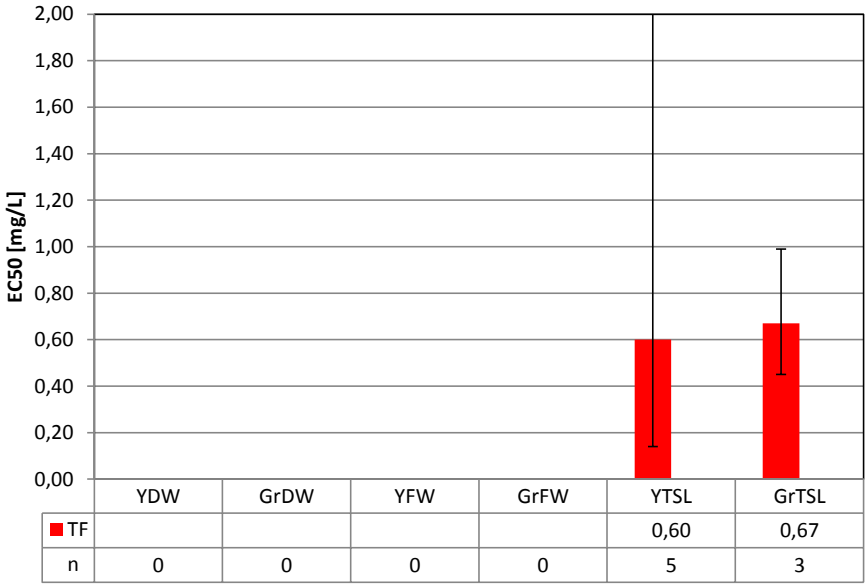


Figure 15: Mean EC50 (bars) and 95%-prediction interval (whiskers) for the shoot length variables studied in *M. aquaticum* at DAT 7 . Data base: all data (except outliers).

4.2.4.2. NOEC

Figure 16, Table 17 and Figure 17 present the results for the NOEC of TF. In several cases either all concentrations were significant or no significances were found at all, hence no concentration could be determined as NOEC. Therefore the validity of the mean NOEC as well as the sensitivity order is highly questionable due to the low numbers ($1 \leq n \leq 4$). The parameters YTSL and GrTSL produced the lowest NOECs (0,042 mg/L) and the smallest MDDs (24% - 28%), indicating the highest statistical power.

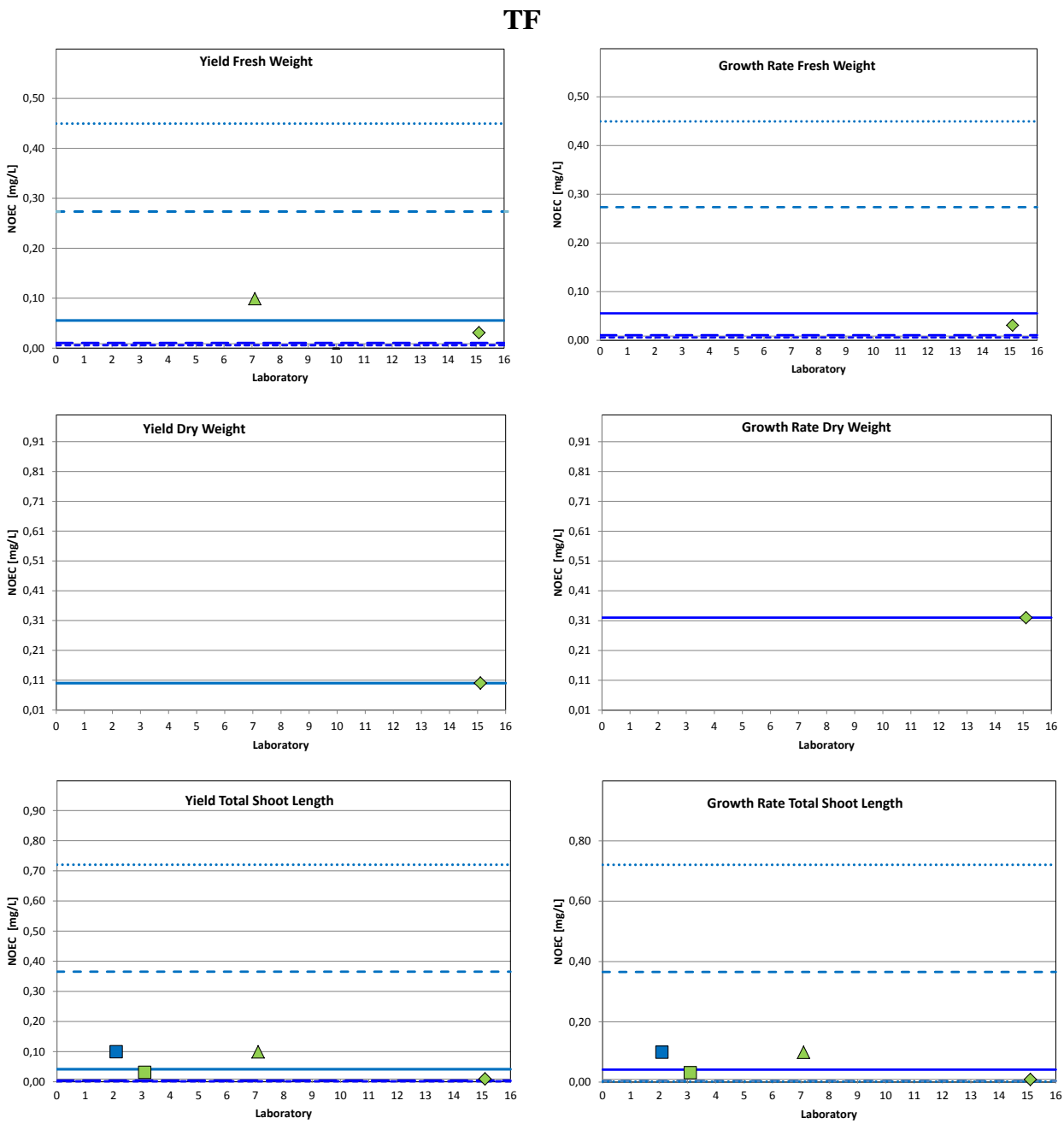


Figure 16: NOEC (symbols) obtained for the yield variables (left hand) and the growth rates (right hand) for TF in *M. aquaticum* at DAT 7 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data (except outliers).

Table 17: Overall mean of the NOEC for TF, number of determined NOEC (n), the 95%-prediction interval (PI) for the yield and growth rate variables for TF in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

TF							
Variable	NOEC [mg/L]			MDD[1]			remarks
	geom. mean[2]	n	95% PI	mean	range	n	
YTSL	0,042	4	0,005 - 0,36	28%	11% - 70%	6	1 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L
GrTSL	0,042	4	0,005 - 0,36	24%	11% - 62%	6	1 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L
YFW	0,057	2	0,012 - 0,27	47%	16% - 90%	6	4 x NOEC ≥ 1 mg/L
GrFW	0,057	2	0,012 - 0,27	38%	11% - 80%	6	4 x NOEC ≥ 1 mg/L
YDW	0,100	1	--	54%	29% - 100%	5	3 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L
GrDW	0,32	1	--	46%	21% - 94%	4	3 x NOEC ≥ 1 mg/L

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")
2) NOECs ≥ 1 mg/L and < 0,01 mg/L not included

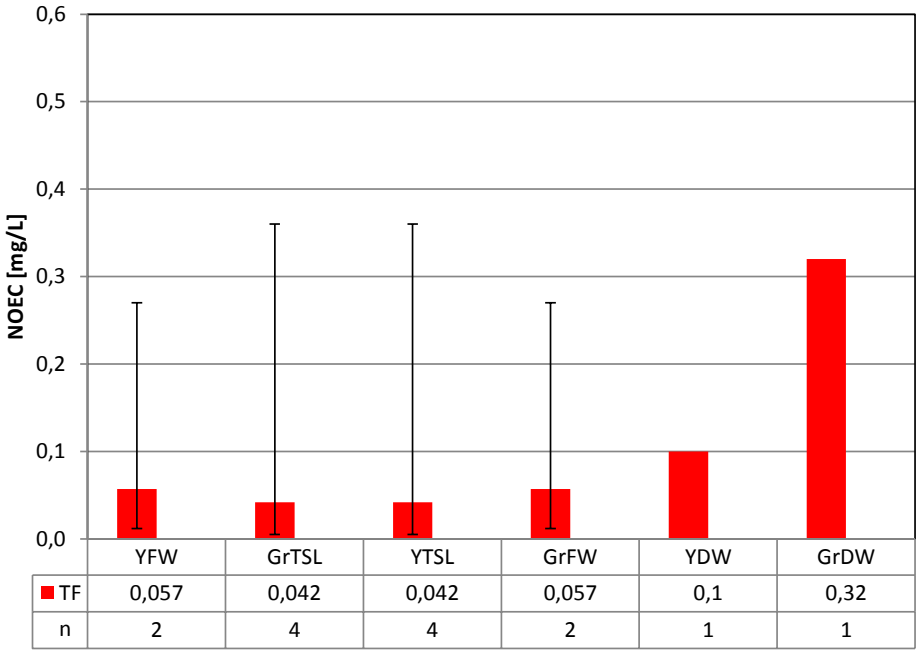


Figure 17: Mean NOEC (bars) and 95%-prediction interval (whiskers) for the variables studied for TF in *M. aquaticum* at DAT 7. Data base: all data (except outliers).

4.2.5. Effects on Root Development

As stated in the ring test protocol, root development was qualitatively judged as "very good root development", "moderate root development", "few roots" and "no roots". Classification into these categories is subjective (except "no roots"), and was open to different interpretation between laboratories.
To evaluate and present the ratings, the number of individual plants (not replicates) allocated to each category in each data set were summed and the sum was expressed as percentage of total plants exposed to a certain treatment. So, even in case of the minimum number of available data sets (n=4, test with trifluralin), the number of plants assessed per treatment was 36 (4 x 9), while the minimum number of control values was 72 (4 x 18). In the case of 3,5-DCP, root assessment was performed in 5 data sets, corresponding to a total of 90 control plants and 45 plants per treatment. In the 10 mg/L treatment the overall sum of plants assigned to level 3 ("few roots") was 18 out of 45, corresponding to 40% of all plants being exposed to this concentration. The relatively high sample size may compensate for possible differences in assignment between the laboratories and lead to an overall sound picture.
The summary of the available data for all substances (not all laboratories reported root development data) is presented in Figure 18. There is a clear trend that root development was affected by all test items, with obvious inhibitions for 3,5-DCP and IP and minor effects following exposure of TF.

Two laboratories also performed a quantitative root length measurement in two tests with 3,5-DCP and one test with IP. These data were used to estimate EC50 and NOEC values (Table 18). A resulting low MDD indicates low variability, which allows robust NOEC-testing in this variable. For 3,5-DCP, the EC50 is similar to that of the other variables, albeit slightly enhanced indicating marginally lower sensitivity than the weight and shoot length parameters. For IP only one EC50 was obtained, which was three-fold higher than the EC50 values based on shoot length and weight parameters. This result suggests that root length is less sensitive than the other variables. However, the maximum inhibition obtained for root length was 32%, so that an EC50 could only be derived by extrapolation. So the results should mainly be used as a proof that quantitative measurement of toxic effects on roots is possible in the sediment-test.

Table 18: Statistical analysis of the root length in *M. aquaticum*, performed by two laboratories for 3,5-DCP and IP; 95%-LCL/UCL: lower and upper 95%-confidence limit; n.d.: not determined. Data base: all data (except outliers).

Test item	Lab Code	EC50 [mg/L]	95%-LCL	95%-UCL	NOEC[mg/L]	MDD%
3,5-DCP	L09	5,19	4,096	6,765	< 1mg/L	13,5
	L15	5,78	5,007	6,723	3,2	13,0
IP	L09	2,044	1,431	4,197	0,1	11,1

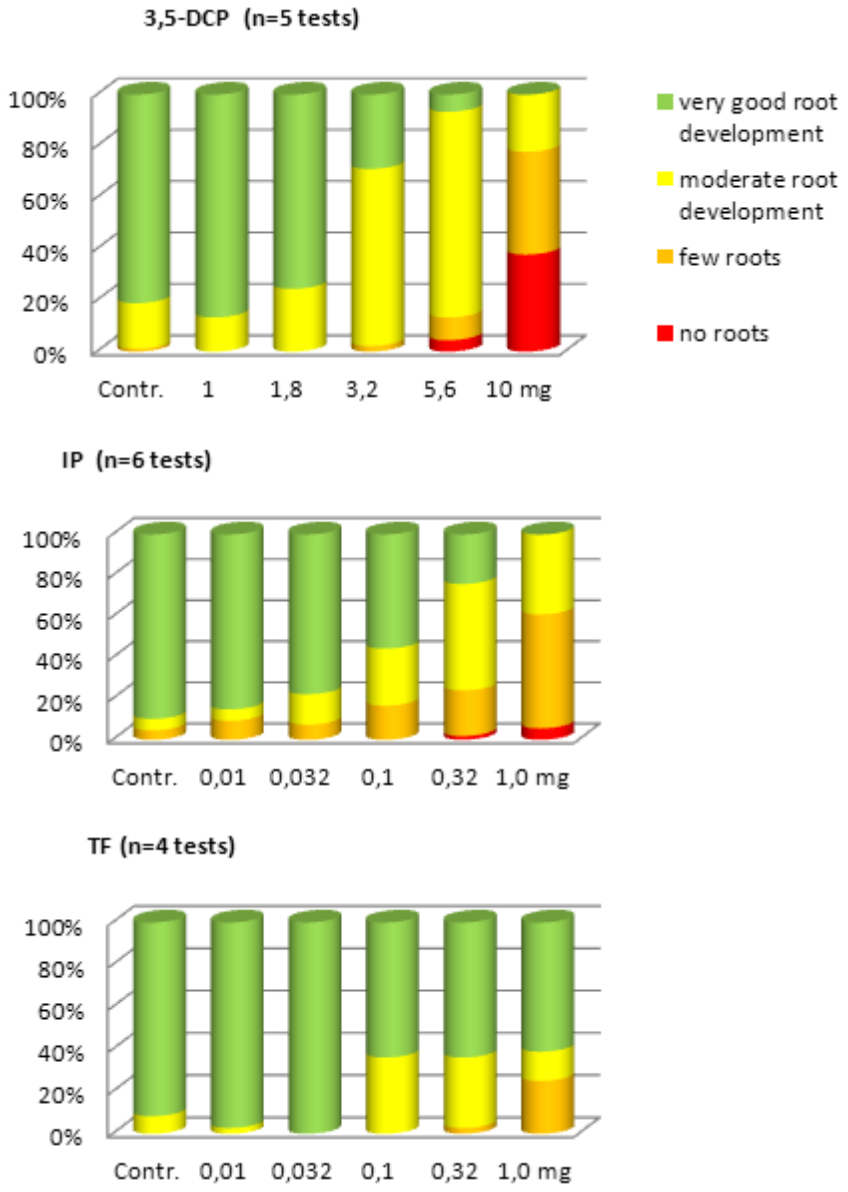


Figure 18: Result of the qualitative assessment of root development in *M. aquaticum* at DAT 7 for all test items. Please note that percentages are related to total number of plants assessed, i.e. in case of four tests, 100% correspond to 72 (4 x 18) control plants and 36 (4 x 9) plants per treatment. Data base: all data (except outliers).

4.2.6. Effects of Validity Criteria on the Results

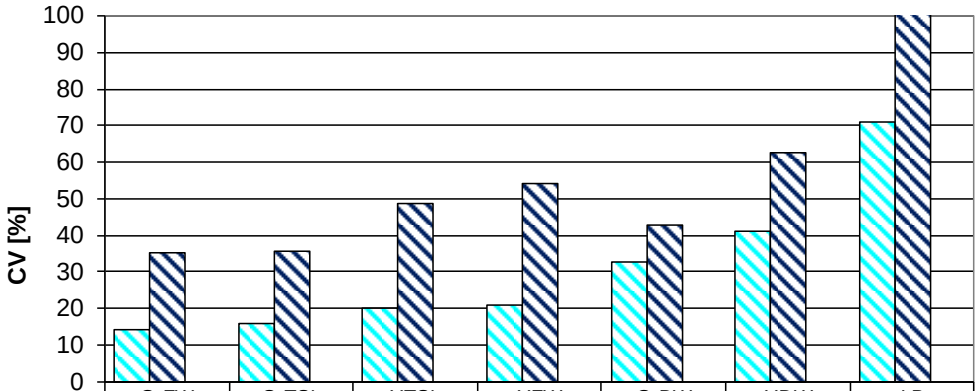
As shown above, the data for *M. aquaticum* showed a high variability, which restricted the calculation of EC50 and NOECs in several data sets. The high variability in some data sets clearly supports the need for validity criteria. Evidence for defining validity criteria comes from a closer examination of the data sets which failed to reveal an EC50 or NOEC. Most of them exhibited both very low growth rates for fresh weight (GrFW) and very high coefficients of variation (CV%) for yield fresh weight (YFW). Since the test protocol states that "only plants within a weight range of 30% should be utilized", a maximum CV% for YFW of 35% in the control plants was defined as one validity criterion. A second validity criterion related to growth was selected. If shoot length doubling were required (as was done in the UBA test), 15 out of 25 data sets failed to fulfill this criterion (i.e. only in 40% of the tests with *M. aquaticum* the shoot length was at least doubled). Another hint for a possible validity criterion comes from ISO 16191: here for *M. aquaticum* a minimum growth rate for FW of 0,09 d^{-1} in the control plants is defined as validity criterion corresponding to a doubling time of 7,70 days (exposure time according to ISO 16191 is 10 days). However, applying this criterion to the present data set again lead to the exclusion of 16 out of 25 data sets and thus appeared to be too strict for the present ring test. As a compromise a minimum growth rate of 0,07 d^{-1} in FW was defined, corresponding to a minimum doubling time of 9,90 days.
Table 19 lists the data sets that passed or failed the validity check. Eleven data sets failed to fulfill both the minimum growth rate FW of 0,07 d^{-1} and the maximum CV% for YFW of 35%. In two more data sets only the maximum CV% for YFW was exceeded (L03, IP: 41%; L07 TF: 48%). Because of the impact of the variability especially on the MDD and thus the NOEC, these data sets were also excluded from further evaluation. In total, 13 out of 26 data sets were found to be invalid (50%). The second evaluation run included only the valid data sets given in Table 19.

In view of the fact that the applied validity criteria were adjusted such to get at least 50% valid tests in the present ring test, type and value of final validity criteria have to be discussed among experts.

The exclusion of invalid data sets led to a marked decrease in both the repeatability (CV_r%) and reproducibility coefficient of variation (CV_R%;Figure 19 and Table 20). The lowest CV_r% were found in the GrFW (14%; before 48%) and the GrTSL (16%; before 18%). YTSL (20%; before 23%) and YFW (21%; before 54%) are found in the mid range. Also, the dry weight CV_r% values changed from 50% to 33 % (GrDW) and from 55% to 41% (YDW). The CVs of the number of lateral branches remained high since *M. aquaticum* generally does not produce considerable numbers of submerged lateral branches. Consequently, this variable was not considered for further statistical analysis. Also, with the valid data sets, the reproducibility coefficient of variation (CV_R%) remained relatively high and exceeded the CV_r% by factors up to 2,6. These factors are greater in the valid data indicating that the within- laboratory variability was reduced to a greater extent than the between-laboratory variability. Overall, following the implantation of validity criteria, the general level of CV_r% appears acceptable for all parameters except for YDW and LB.

Table 19: Overview of all valid (black, x) and invalid (red) data sets and the reason for invalidity in *M. aquaticum:* Gr = GrFW < 0,07 d^{-1} ; CV = CV YFW > 35%.

Laboratory	M. aquaticum			No of tests performed	No of tests valid
	3,5-DCP	IP	TF		
L01				-	-
L02	CV, Gr	CV, Gr	x	3	1
L03	x	CV	x	3	2
L04				-	-
L05		x		1	1
L06				-	-
L07	x	CV, Gr	CV	3	1
L08	x	x	x	3	3
L09	x	CV, Gr		2	1
L10				-	-
L11	CV, Gr	CV, Gr		2	0
L12				-	-
L13	CV, Gr	CV, Gr	CV, Gr	3	0
L14	CV, Gr	x	CV, Gr	3	1
L15	x	x	x	3	3
Total tests	**9**	**10**	**7**	**26**	**13**
Valid tests	**5**	**4**	**4**	**13**	



	GrFW	GrTSL	YTSL	YFW	GrDW	YDW	LB
CVr%	14,1	15,9	20,0	21,0	32,9	41,3	71,3
CVR%	35,2	35,4	48,5	54,2	42,6	62,4	111,1
n	10	12	12	9	10	9	12

Figure 19: Coefficients of variation for repeatability (CV_r%, light blue) and reproducibility (CV_R%, dark blue) of all valid control data sets in *M. aquaticum* ordered by magnitude; n: number of valid control data sets.

Table 20: Overview of the ring-test statistics obtained before and after exclusion of invalid data sets of controls in *M. aquaticum*; overall mean: mean of all data from all controls and all laboratories (except outliers); CV_r%: repeatability coefficient of variation; CV_R%: reproducibility coefficient of variation; 95%-/99%-PI: 95%-/99%-prediction interval; p: number of control data sets

	Overall mean	CV_r%	CV_R%	CV_R%/ CV_r%	95% -PI	99%-PI	p
Valid Data Sets							
YFW [mg]	293	21,0	54,2	2,6	2,2 – 584,5	-89,2 - 676,0	9
YDW [mg]	14,0	41,3	62,4	1,5	1,8 – 26,3	-2,0 – 26,3	9
YTSL [cm]	5,3	20,0	48,5	2,4	0,61 - 10,1	-0,87 - 11,6	12
GrFW d^{-1}	0,114[6]	14,1	35,2	2,5	0,041 - 0,187	0,019 - 0,210	10
GrDW d^{-1}	0,066[7]	32,9	42,6	1,3	0,025 - 0,106	0,012 - 0,119	10
GrTSL d^{-1}	0,099[8]	15,9	35,4	2,3	0,036 - 0,161	0,016 - 0,181	12
All Data Sets							
YFW [mg]	211,2	54,3	74,3	1,4	-48,3 – 470,7	-129,8 - 552,2	21
YDW [mg]	12,5	54,6	67,6	1,2	1,8 - 23,1	-1,5 - 26,4	20
YTSL [cm]	5,0	23,4	55,9	2,4	-0,019 – 10,0	-1,6 - 11,6	25
GrFW d^{-1}	0,078[9]	47,7	67,1	1,4	-0,016 – 0,171	-0,045 - 0,201	23
GrDW d^{-1}	0,054[10]	49,6	59,4	1,2	0,011 – 0,097	-0,003 - 0,111	21
GrTSL d^{-1}	0,092[11]	17,9	35,0	2,0	0,036 – 0,147	0,019 - 0,164	25

[6] Corresponding to a doubling time of 6,1 days
[7] Corresponding to a doubling time of 10,5 days
[8] Corresponding to a doubling time of 7,0 days
[9] Corresponding to a doubling time of 8,9 days
[10] Corresponding to a doubling time of 12,8 days
[11] Corresponding to a doubling time of 7,5 days

Also, the MDDs derived from the CV_r% was reduced following the elimination of invalid datasets (Table 21). The growth rates of FW and TSL revealed the lowest MDDs, enabling the statistical detection of an effect size of around 20%.

Table 21: Detectable effect levels (MDD%) for different variables in *M. aquaticum* using Williams´test. MDDs calculated assuming the given test design (control n = 6; 5 treatments with n = 3) and mean control variances (i.e. repeatabilities) as obtained in the present ring test before and after exclusion of invalid data sets of controls (see Table 20). Validity criteria were a minimum growth rate for Fresh Weight of 0,07 d^{-1} and a maximum CV for Yield Fresh Weight of 35%. CV_r% = repeatability coefficient of variation P = number of control data sets from which mean CV_r% was calculated

	Valid Data Sets			All Data Sets		
	CV_r%	MDD%	p	CVr%	MDD%	p
YFW	21,0	26,9	9	54,3	69,6	21
YDW	41,3	52,9	9	54,6	69,9	20
YTSL	20,0	25,6	12	23,4	29,9	25
GrFW	14,1	18,0	10	47,7	61,0	22
GrDW	32,9	42,1	10	49,6	63,4	21
GrTSL	15,9	20,3	12	17,9	22,9	25

Excluding invalid data sets, the number of remaining EC50 and NOECs was below the acceptable values for ring-test statistics. Therefore, the results of the second evaluation run will be presented only by overview tables. Table 22 to Table 25 give the result for the EC50 NOEC and observed MDD.
Throughout the results, the prediction intervals (PI), the CV and the MDD became markedly smaller. The MDD decreased by a factor up to around two, indicating a marked increase in the statistical power of the single tests.
For 3,5-DCP the mean EC50 values increased by 5% to 22%, for IP the EC50 values increased by 8% to 53%, for IP the EC50 values increased by 36% (only means based on at least 2 single EC50 considered for comparison). This increases were possibly due to the fact that the plants of the

invalid data sets often showed very low growth rates and were more sensitive – probably an artifact of the bad performance of the respective plants.

As a consequence of the decreased variability and in contrast to the EC50, the NOEC was smaller in 12 out of 18 cases. In contrast to 3,5-DCP, where the NOEC 3 times increased and 3 times decreased, lower NOECs were observed for IP and TF. This mainly indicates that in the first evaluation run with all data included reasonable NOECs could not be derived due to high MDDs. The smallest MDDs were obtained for the shoot length parameters.

Overall, the results presented in this section clearly indicate that the test system is generally practicable and variability can be adjusted to a level that enables robust EC50 and NOECs when applying validity criteria (except for dry weight). The high percentage of invalid tests probably could be lowered by further standardization of the experimental methods. The validity criteria that were applied in the present evaluation (minimum growth rate FW of 0,07 d^{-1} and maximum CV for YFW of 35%) provide a starting point but require further debate.

Table 22: Overview of the EC50 obtained for the valid data sets versus all data sets in *M. aquaticum* at DAT 7

Valid Data Sets

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW		0	
GrDW	3,7	1	-
YFW	4,6	2	3,6 - 5,8
GrFW	5,2	2	2,8 - 9,7
YTSL	5,7	5	3,1 - 10,6
GrTSL	6,1	5	3,5 - 10,8

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,12	1	--
GrDW	0,13	1	--
YFW	0,86	2	0,59 - 1,25
GrFW	1,18	2	1,09 - 1,27
YTSL	0,37	3	0,16 - 0,84
GrTSL	0,51	3	0,13 - 1,91

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to no clear dose-response relation or not significant due to too few / low inhibition values		
GrDW			
YFW			
GrFW			
YTSL	0,82	3	0,16 - 4,1
GrTSL	0,75	1	--

All Data Sets

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	3,8	2	2,0 - 7,1
GrDW	3,9	3	2,5 - 6,1
YFW	4,0	5	2,7 - 5,9
GrFW	4,7	4	3,0 - 7,2
YTSL	5,0	8	2,5 - 9,9
GrTSL	5,0	8	2,9 - 10,1

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,06	2	0,01 - 0,33
GrDW	0,07	2	0,01 - 0,37
YFW	0,62	4	0,19 - 1,97
GrFW	0,77	4	0,22 - 2,73
YTSL	0,32	8	0,11 - 0,98
GrTSL	0,47	8	0,17 - 1,25

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to no clear dose-response relation or not significant due to too few / low inhibition values		
GrDW			
YFW			
GrFW			
YTSL	0,60	5	0,14 - 2,64
GrTSL	0,67	3	0,45 - 0,99

Table 23: Overview of the NOEC for 3,5-DCP obtained for the valid data sets versus all data sets in *M. aquaticum* at DAT 7

Valid Data Sets

3,5 DCP							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	2,4	2	1,1 - 5,3	28%	22% - 31%	3	1 x NOEC ≥10 mg/L
GrFW	3,2	3	1,0 - 9,7	23%	15%-30%	4	1 x NOEC ≥ 10 mg/L
YDW	3,2	1	--	74%	48% - 100%	2	1 x NOEC ≥ 10 mg/L 1 x MDD > 100%
GrDW	5,6	2	5,6 - 5,6	63%	46% - 79%	3	1 x NOEC ≥ 10 mg/L
YTSL	3,9	3	2,0 - 7,3	24%	12% - 40%	4	1 x NOEC < 1 mg/L
GrTSL	3,2	3	3,2 - 3,2	18%	9% - 28%	4	1 x NOEC < 1 mg/L

All Data Sets

3,5 DCP							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	2,9	6	1,3 - 6,7	57%	22% - 100%	7	1 x NOEC ≥ 10 mg/L 1 x MDD > 100%
GrFW	3,5	7	1,6 - 7,4	51%	15%-100%	8	1 x NOEC ≥ 10 mg/L 1 x MDD > 100%
YDW	4,6	3	2,5 - 8,8	81%	48% - 100%	5	2 x NOEC ≥ 10 mg/L 2 x MDD > 100%
GrDW	5,0	5	3,1 - 8,2	75%	46% - 94%	6	1 x NOEC ≥ 10 mg/L
YTSL	3,2	7	1,0 - 9,8	25%	12%-40%	8	1 x NOEC < 1 mg/L
GrTSL	2,9	7	1,1 - 8,1	21%	9% - 33%	8	1 x NOEC < 1 mg/L

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")

2) NOECs ≥ 10 mg/L and < 1 mg/L not included

Table 24: Overview of the NOEC for IP obtained for the valid data sets versus all data sets in *M. aquaticum* at DAT 7

Valid Data Sets

IP							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	0,100	1	--	40%	32% - 53%	2	1 x NOEC ≥ 1 mg/L
GrFW	0,100	1	--	33%	28% - 43%	2	1 x NOEC ≥ 1 mg/L
YDW	0,100	1	--	60%	40% - 100%	3	2 x NOEC < 0,01 mg/L 1 x MDD > 100%
GrDW	0,056	2	0,01 - 0,27	58%	36% - 100%	3	1 x NOEC < 0,01 mg/L 1 x MDD > 100%
YTSL	0,047	4	0,01 - 0,17	27%	11% - 43%	4	
GrTSL	0,047	4	0,01 - 0,17	24%	11% - 38%	4	

All Data Sets

IP							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	0,147	3	0,04 - 0,55	63%	32% - 100%	9	4 x NOEC ≥ 1 mg/L 3 x MDD > 100%
GrFW	0,147	3	0,04 - 0,55	54%	24% - 100%	8	2 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L 2 x MDD > 100%
YDW	0,100	5	0,02 - 0,49	65%	36% - 100%	8	2 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L 3 x MDD > 100%
GrDW	0,083	6	0,02 - 0,45	60%	34% - 100%	8	1 x NOEC ≥ 1 mg/L 2 x NOEC < 0,01 mg/L 2 x MDD > 100%
YTSL	0,069	9	0,01 - 0,48	27%	11% - 43%	9	1 x NOEC < 0,01
GrTSL	0,079	9	0,006 - 0,96	23%	19% - 23%	9	

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")
2) NOECs ≥ 1 mg/L and < 0,01 mg/L not included

Table 25: Overview of the NOEC for TF obtained for the valid data sets versus all data sets in *M. aquaticum* at DAT 7

Valid Data Sets

TF							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	0,032	1	--	29%	16% - 38%	3	2 x NOEC ≥ 1 mg/L
GrFW	0,032	1	--	19%	11% - 24%	3	2 x NOEC ≥ 1 mg/L
YDW	0,100	1	--	47%	30% - 80%	3	2 x NOEC ≥ 1 mg/L
GrDW	0,320	1	--	39%	23% - 70%	3	2 x NOEC≥ 1 mg/L
YTSL	0,032	3	0,003 - 0,303	17%	11% - 21%	3	
GrTSL	0,032	3	0,003 - 0,303	13%	11% - 16%	3	

All Data Sets

TF							
Variable	NOEC [mg/L]			MDD [1]			remarks
	geom. mean [2]	n	95% PI	mean	range	n	
YFW	0,057	2	0,012 - 0,27	47%	16% - 90%	6	4 x NOEC ≥ 1 mg/L
GrFW	0,057	2	0,012 - 0,27	38%	11% - 80%	6	4 x NOEC ≥ 1 mg/L
YDW	0,100	1	--	54%	29% - 100%	5	3 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L
GrDW	0,320	1	--	46%	21% - 94%	4	3 x NOEC ≥ 1 mg/L
YTSL	0,042	4	0,005 - 0,36	28%	11% - 70%	6	1 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L
GrTSL	0,042	4	0,005 - 0,36	24%	11% - 62%	6	1 x NOEC ≥ 1 mg/L 1 x NOEC < 0,01 mg/L

1) MDDs > 100% were set to 100% for calculation of mean and range (100% MDD means "statistical detection of any effect impossible")
2) NOECs ≥ 1 mg/L and < 0,01 mg/L not included

4.3. Summary of the Results for *M. aquaticum*

The data sets of *M. aquaticum* exhibited a high variance both within and between the laboratories. Nine out of 144 of the control data sets proved to be statistical outliers which were removed from the analysis. As a consequence, the variability based on all control data except outliers was reduced, but was still considered high (repeatability 18% - 55%, reproducibility 35% - 74%). The weight parameters exhibited the highest variability, the shoot length parameters the lowest. Growth rates were less variable than yields. Mean MDDs between 21% (growth rate total shoot length) and 75% (yield dry weight) were obtained (data base: all data).
The reason for the high variability might be seen in different plant sources, different preculturing conditions, test conditions deviating from the test protocol and / or differences between test plants and representative plants used for determining initial weights. Especially for weight parameters also methodological problems are seen as particles of sediment adhering to roots might skew the measurements.
Generally, in *M. aquaticum* lateral branches occurred only occasionally and therefore proved to be not appropriate for evaluation. Low growth rates for fresh weight were obtained in several tests: only 42% of the tests with *M. aquaticum* showed a growth rate fresh weight of 0,09 d-1 or higher (according to a doubling time of 7,5 days, i.e. enabling doubling of fresh weight within the exposure time of 7 days).
In an initial evaluation including all data sets (except outliers), the EC50 and NOECs could not be determined in every case. This was not only due to high variability, but also due to growth promotion in presence of IP and TF being observed in lower concentrations and leading to increased weights. Especially the determination of NOECs based on weight variables was prohibited by the high variability leading to MDDs > 100%.
In a second evaluation, two preliminary validity criteria were applied: a minimum growth rate of the fresh weight of 0,07 d^{-1} (corresponding to a doubling time of 9,9 days – this was set in order to get at least 50% valid tests) and a maximum coefficient of variation in the yield of fresh weight of 35%. Fifty percent of the data sets failed to fulfill both criteria. Further statistical evaluation was performed with the remaining 50% of data sets. If based on only valid data, EC50 values were observed to be higher in a number of cases as compared to EC50 values based on all data, indicating that the very slow growing and excluded plants were among more sensitive ones. Due to the limited number of data sets the obtained EC50 should be considered carefully. In the following, only mean EC50 values based on at least two data sets are reported.
For 3,5-DCP, the mean EC50 ranged between 3,8 and 5,0 mg/L (all data, n=2-8), respectively between 4,6 and 6,1 mg/L (valid data, n=2-5) with only small differences between variables. IP and TF showed similar toxicity on a markedly lower concentration level. For IP, mean EC50 values ranged between 0,06 and 0,77 mg/L (all data, n= 2-8), respectively between 0,37 and 0,86 mg/L (valid data, n=2-3). Dry weight parameters were most sensitive, however, in view of the small data base, the order of the sensitivities seems to be questionable. For TF, only EC50 values for the shoot length parameters were available. They ranged between 0,60 and 0,82 mg/L (both all data and valid data, n= 3-5).
The application of the validity criteria led to a bisection of the CV_r% and MDD in most cases. The repeatability CV_r% based on valid data sets ranged from 14% - 41%, the reproducibility

CV_R% from 35% - 62%. As a consequence and in contrast to the EC50, by trend the NOEC proved to be lower after the exclusion of invalid data sets. NOECs of valid data were between 2,4 and 5,6 mg/L in 3,5-DCP (all data: 2,9 to 5,0 mg/L), in IP between 0,047 and 0,1 mg/L (all data; 0,069 to 0,147 mg/L), and in TF were 0,032 mg/L, respectively 0,042 – 0,057 mg/L (all data).

5. Results for *Myriophyllum spicatum* and Discussion

5.1. Analysis of Control Data

5.1.1. General

The analysis of the control data sets had the same goals as in *M. aquaticum*: characterization of repeatability variance and reproducibility variance and identification of outliers. In addition to the six variables, YFW, YDW, YTSL, GrFW, GrDW and GrTSL the number of lateral branches (LB) and the total length of lateral branches (TLBL) were analyzed since lateral branches were regularly produced by the test plants and their total length accounted to around 50% of the total shoot length in *M. spicatum* (Figure 20).

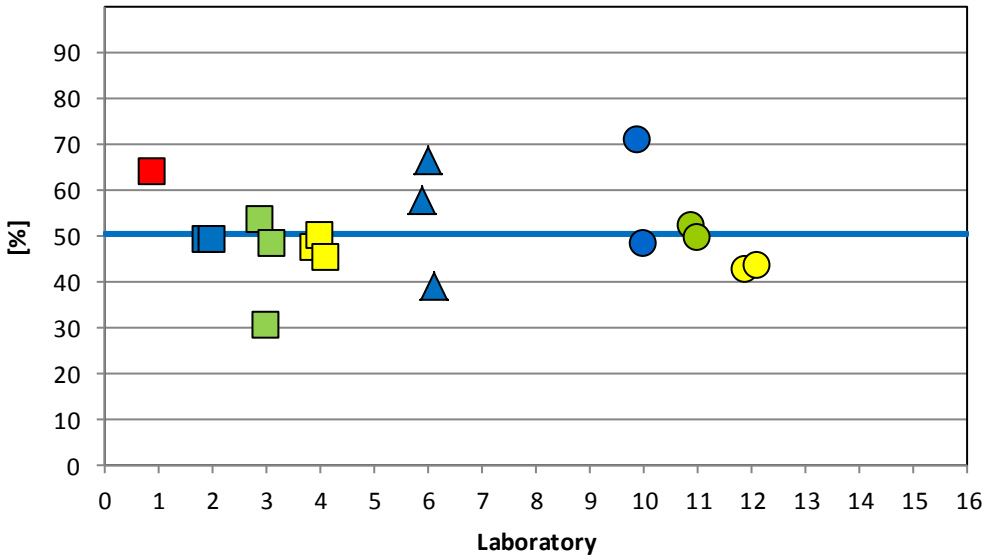


Figure 20: Contribution of lateral branches to total shoot length of *M. spicatum* (controls, DAT 14): Percentage total length lateral branches of total shoot length. Data base: all data.

Although a total of 25 toxicity tests with *M. spicatum* were performed, the data base for evaluation was reduced for various reasons:

Lo8	3,5-DCP, IP, TF	Extensive contamination with bacteria, algae and snails was reported for all tests. Several plants could not be evaluated at all leading to a huge number of missing data.

		Consequently, the remaining results seem highly questionable. Therefore only initial values of the controls were included in evaluations.
L06	TF	Reported concentrations of treatments were two fold higher than intended concentrations stipulated in the test protocol. While this treatment error should not affect the toxic metrics, the EC50 and NOEC were found to be 2-8 fold higher than in other tests and so it was concluded that concentrations were incorrectly reported by the laboratory. However, this point could not be verified, the results were excluded from statistical evaluations.
L14	3,5-DCP	All plants died in 10 mg/L treatment. Measurements of weights and lengths were not performed with this treatment at the end of the test. The remaining data mostly did not reveal significant dose response relationships.
L14	3,5-DCP, IP, TF	Lateral branches and roots were not recorded.

In some of the remaining data sets control growth rates based on fresh weight or dry weight were close to zero (< 0,01 d^{-1}) with large CVs or even below zero resulting in negative yields. These data were excluded from the statistical analysis (L03 3,5-DCP, L11 3,5-DCP and IP, L12 IP).

As for *M. aquaticum*, validity criteria were not applied in initial evaluation and all available data sets were included in the statistical analysis.

Figure 21 presents the arithmetic means and 95%-confidence intervals of the initial lengths of the test plants and initial weights of representative plants for each laboratory. In most cases, the initial total shoot length of test plants showed relatively low variability within the same laboratory (95%-confidence interval not visible since covered by the symbol). In contrast, there is a large variability in initial lengths between the laboratories (4,8 to 18, 1 cm). Assuming that all tests were started with an initial plant length of 6 cm $\pm$ 1 cm as prescribed in the test protocol, the discrepancy is probably due to variation in growth during the rooting period. The data set with the initial length of 18 cm was identified as an outlier and thus excluded from further evaluations.

The initial fresh weight also showed a high variability. Maximum weights of L08 and L10 led to lower margins of the 95%- and 99%-prediction intervals below zero. Besides the different initial length of the plants, the main reason for these differences probably is the difficulty of measuring total weights including roots, which need to be retrieved from the sediment.

Regardless of the method it should be ensured that the representative plants from which the initial weights for calculating yields and growth rates are obtained are really “representative”, i.e. the relation of TSL DAT 0 representative plants to TSL DAT 0 test plants should be close to one. As can be seen from Table 26, this was the case in most data sets. 16 out of 24 data sets showed ratios between 0,90 and 1,10. Four tests showed ratios between 0,81 and 1,12, i.e. in total 20 out of 24 tests showed ratios of 1 $\pm$ 20%. Only in four data sets of 24 tests ratios of 1,21

to 1,44 were found. The latter led to overestimation of the start weights of the test plants and thus underestimation of yields and growth rates or – in worst case – to negative values. To summarize, although the absolute level of the fresh weight appeared not to affect resulting yields, growth rates or inhibitions, it appears that more standardization is needed.

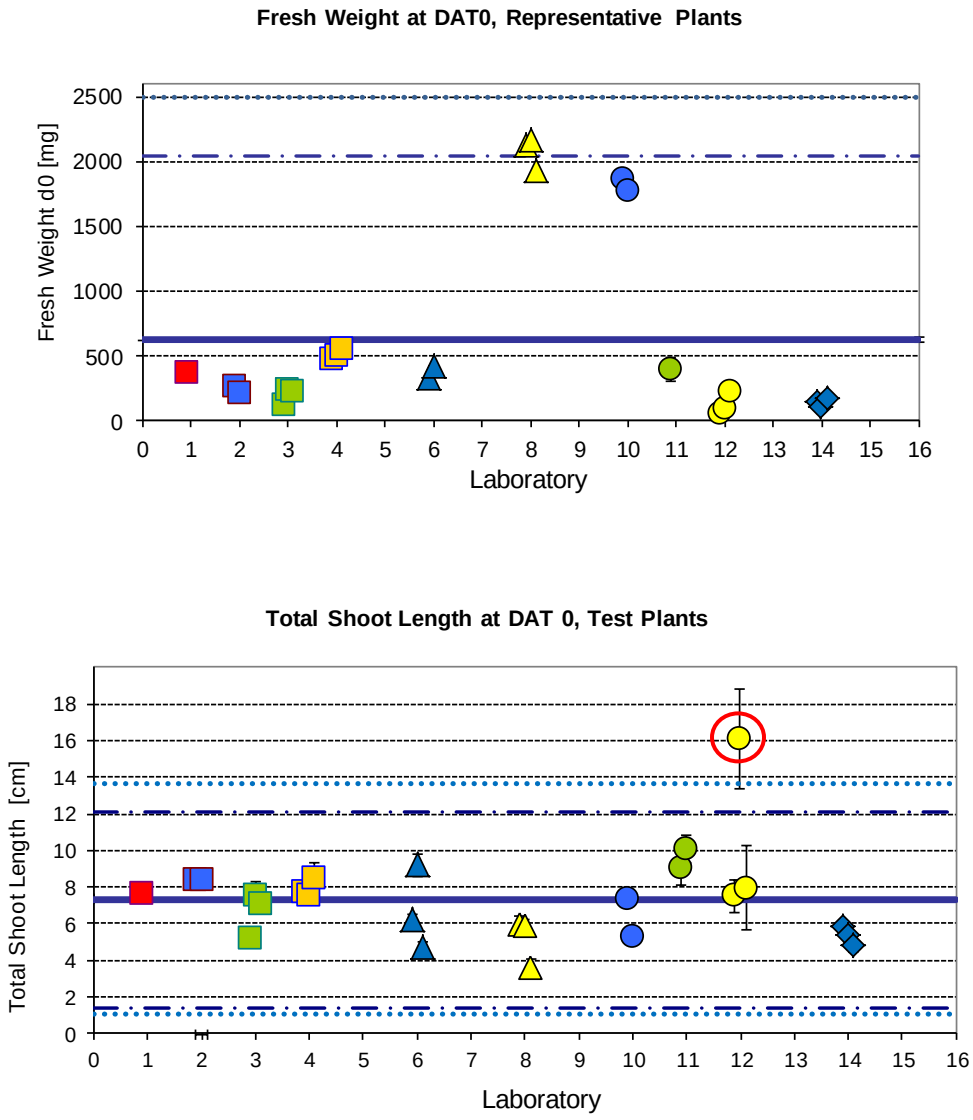


Figure 21: Arithmetic means (symbols) and 95%-confidence interval (whiskers) for the fresh weight of representative plants (top) and total shoot length of test plants in *M. spicatum* at DAT 0 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line); outlier marked with a red circle; lower prediction margins below zero aren't shown. Data base: all data.

Table 26: Ratio between total shoot length of representative plants and test plants at DAT 0 in *M. spicatum*; ratios deviating ≥ 20% from 1 in bold (total number of ratios = 24, in one test no start lengths of representative plants were provided). Data base: all data.

Ratio TSL repr.plants / test plants DAT 0			
	3,5-DCP	IP	TF
L01	1,02		
L02	**1,44**	**1,35**	
L03	1,04	1,11	**1,21**
L04	0,93		**1,35**
L06	1,06	0,88	1,09
L08	0,90	1,00	1,02
L10	0,81	0,90	
L11	1,01	0,90	
L12	0,97	1,12	0,92
L14	0,99	1,00	1,04

5.1.2. Reproducibility and Repeatability

The control data of eight variables (YFW, YDW, YTSL, GrFW, GrDW, GrTSL, LB, and TLBL) were analyzed in more detail. Figure 22 and Figure 23 show the arithmetic means of the control data sets of each laboratory relative to the overall mean – the estimate of the true mean of the respective variable. Outlier data sets were not identified, except for an outlier candidate in the GrTSL (L12). Laboratory means did not fall outside of the 99%-prediction intervals (dotted lines in Figure 22 and Figure 23; for numerical data see Table 27) and thus all data sets were included in the subsequent analysis.

The reproducibility (inter-laboratory variability) coefficients of variation as well as those for repeatability (intra-laboratory variability) of all variables are presented in Figure 24 and Table 27. Figure 24 gives also the order of variability: the growth rates (in front GrTSL) showed the lowest variability (CV_r: 10,2%, CV_R: 30,6%) whereas the LB and TLBL showed the highest variability (CV_r: 24,2%, CV_R: 57,5%). For all parameters, the variability level was markedly lower than in *M. aquaticum*. Since the TLBL accounts for 50% of the TSL (Figure 20) the variability of this variable will influence that of all weight and shoot length variables.

The relationship between reproducibility and repeatability is a measure of the current degree of standardization of the test system. Again, the reproducibility variability (CV_R%) was always higher than repeatability variability (CV_r%). The factors between CV_R% and CV_r% are higher than in *M. aquaticum*. In *M. spicatum* the highest factor (3,1) was found in YTSL (in *M. aquaticum*: maximum 2,4 in YTSL). However, in view of the fact that the reproducibility CV_R% in *M. spicatum* is comparable to that of *M. aquaticum*, the higher factors found here probably are due to the lower repeatability variability of *M. spicatum*. This may be due to the overall higher levels of length and weight values, hence the relative effect of varying lengths and weights due to methodological difficulties or inaccuracies are of lower consequences for the resulting variability.

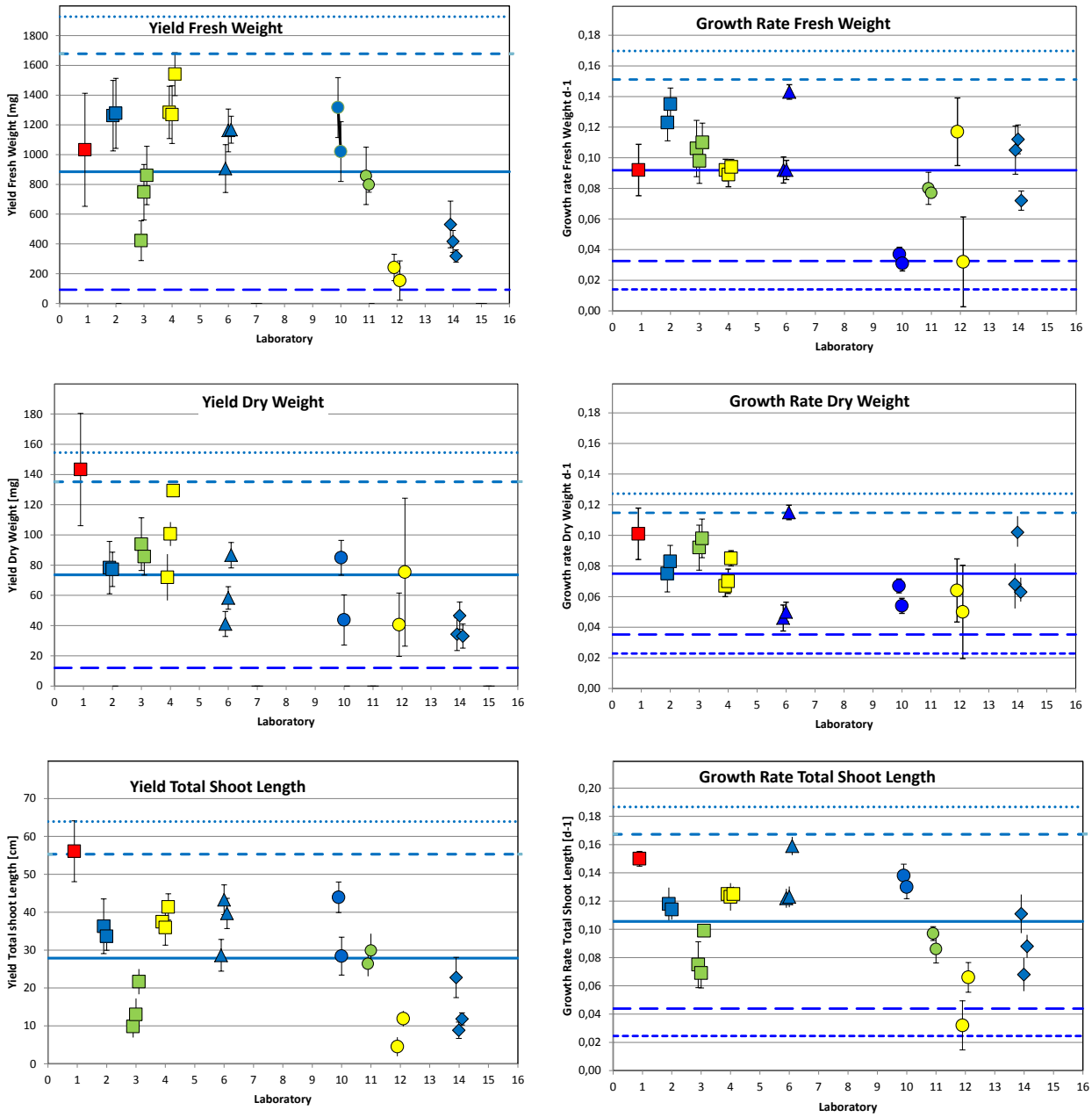


Figure 22: Arithmetic means (symbols) and 95%-confidence interval (whiskers) for yield variables (a, b, c) and growth rates (d, e, f) in *M. spicatum* in control data sets of each laboratory at DAT 14 together with the overall mean (thick blue line), the upper 95%-prediction margin (dashed line) and the upper 99%-prediction interval (dotted line); statistical outliers were not encountered; lower prediction margins below zero are not shown. Data base: all data.

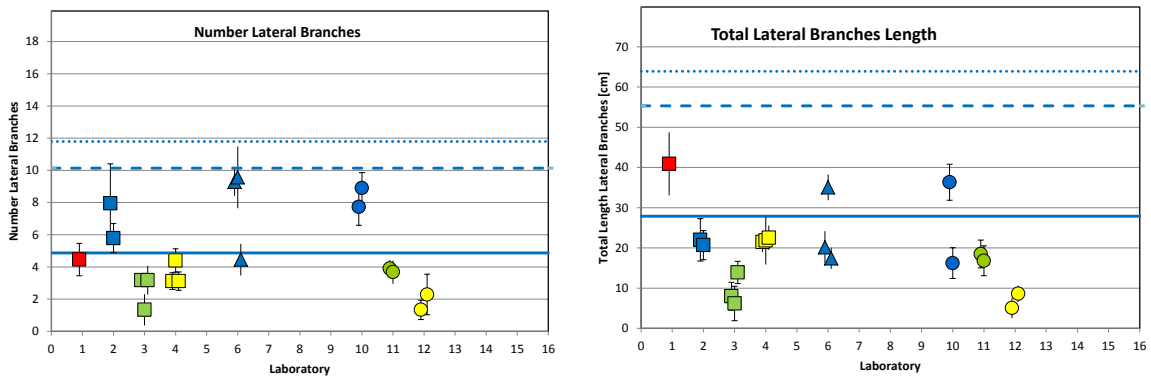
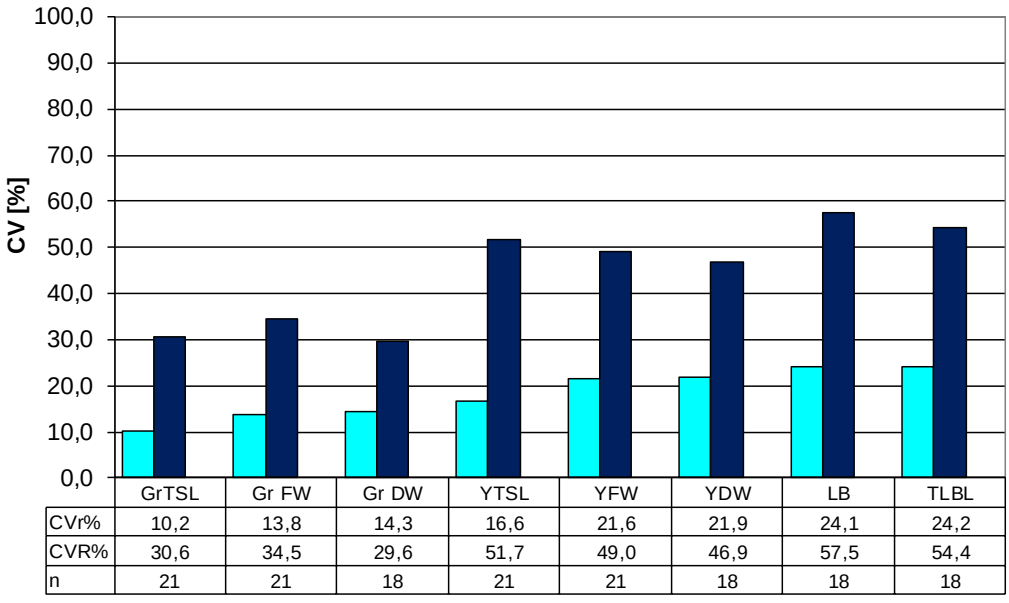


Figure 23: Arithmetic means (symbols) and 95%-confidence interval (whiskers) for the Lateral Branches Number and Lateral Branches Length as well as the percentage fraction of TLBL from TSL in *M. spicatum* in the control data sets of each laboratory at DAT 14 together with the overall mean (thick blue line), the 95%-prediction margin (dashed line) and the 99%-prediction margin (dotted line); statistical outliers were not encountered; lower prediction margins below zero are not shown. Data base: all data.



	GrTSL	Gr FW	Gr DW	YTSL	YFW	YDW	LB	TLBL
CVr%	10,2	13,8	14,3	16,6	21,6	21,9	24,1	24,2
CVR%	30,6	34,5	29,6	51,7	49,0	46,9	57,5	54,4
n	21	21	18	21	21	18	18	18

Figure 24: Coefficients of variation for repeatability (CV_r%, light blue) and reproducibility (CV_R%, dark blue) of all control data sets in *M. spicatum* ordered by magnitude; n: number of control data sets. Data base: all data.

Table 27: Overview of the ring-test statistics of controls in *M. spicatum*; overall mean: mean of all data from all controls and all laboratories; CV_r%: repeatability coefficient of variation; CV_R%: reproducibility coefficient of variation; 95%-/99%-PI: 95%-/99%-prediction interval; p: number of control data sets

	Overall mean	CV_r%	CV_R%	CV_R%/ CV_r%	95% -PI	99%-PI	p
YFW [mg]	885,1	21,6	48,9	2,3	92,2 – 1678,1	-157,0 - 1927,3	21
YDW [mg]	73,6	21,8	46,9	2,1	12,0 - 135,2	-7,3 - 154,5	18
YTSL [cm]	28,0	16,6	51,7	3,1	0,55 - 55,4	-8,1 - 64,0	21
LB [cm]	4,97	24,1	57,4	2,4	-0,29 - 10,24	-1,9 - 11,9	18
TLBL [cm]	19,6	24,2	54,4	2,3	0,028 - 39,3	-6,1 – 45,4	18
GrFW d^{-1}	0,093[12]	13,8	34,5	2,5	0,033 – 0,152	0,015 – 0,171	21
GrDW d^{-1}	0,076[13]	14,3	29,7	2,1	0,036 – 0,116	-0,024 - 0,128	18
GrTSL d^{-1}	0,107[14]	10,2	30,6	3,0	0,045 – 0,168	0,026 - 0,188	21

5.1.3. Minimal Detectable Difference and Statistical Power

Note that the minimum detectable effect levels listed in Table 28 represent a theoretical orientation extrapolated from the variability of the controls. Depending on the level of the respective variances in the individual test run and on cases where the number of evaluable replicates deviates from the given test design, the MDDs actually resulting from a certain test run can be lower, but also higher. The detectable effect levels obtained in the present ring test are presented with the NOECs in Sections 5.2.2.2, 5.2.3.2 and 5.2.4.2.

For the given ring-test design and GrTSL being the variable with the lowest coefficient of variation of 10,2%, already an effect 13,1% becomes detectable on a statistically significant level by means of Williams' test. This MDD, and also the others are markedly lower than in *M. aquaticum*. Except for the lateral branches variables, MDDs were found to be around 20% allowing robust determination of the NOEC.

[12] Corresponding to a doubling time of 7,5 days
[13] Corresponding to a doubling time of 9,1 days
[14] Corresponding to a doubling time of 6,5 days

Table 28: Detectable effect levels (MDD%) for different variables in *M. spicatum* using Williams´test. MDDs calculated assuming the given test design (control n = 6; 5 treatments with n = 3) and mean control variances (i.e. repeatabilities) as obtained in the present ring test (see Table 27). CV_r% = repeatability coefficient of variation P = number of control data sets from which mean CV_r% was calculated.

	CV_r%	MDD%	p
YFW	21,6	27,7	21
YDW	21,9	28,0	18
YTSL	16,6	21,3	21
LB	24,1	30,8	18
TLBL	24,2	30,9	18
GrFW	13,8	17,6	21
GrDW	14,3	18,3	18
GrTSL	10,2	13,1	21

5.2. Reproducibility of Toxic Metrics and Sensitivity of Variables

5.2.1. General

This chapter deals with the reproducibility of the toxic effects of 3,5-DCP, IP and TF on YFW, YDW, YTSL, GrFW, GrDW, GrTSL, LB and TLBL. The repeatability cannot be determined since each laboratory performed the toxicity test only once per substance. As for *M. aquaticum*, in the last section also the root development will be additionally considered.

Due to some shortcomings with the provided data sets (see above, Section 5.1.1) the EC50 and NOEC could not always be determined for all variables. According to DIN ISO 5725-2 and DIN ISO 5725-5 at least 8 datasets are required for the assessment of reproducibility. Nonetheless, even though this requirement is not always fulfilled all results will be reported.

As for *M. aquaticum*, validity criteria were not initially applied and all data were initially considered for evaluation.

5.2.2. Effects of 3,5-dichlorophenol

5.2.2.1. EC50

Figure 25 and Figure 26 present the results for the EC50 for 3,5-DCP. Outlying results were not found and all EC50s were located within the 95%-prediction interval. The highest and lowest EC50 differed by maximum of 5-fold across all variables. As for *M. aquaticum,* the high inter-laboratory variability of the variables did not visibly affect the absolute EC50-level. This point is because the response is expressed as inhibition relative to the controls which compensates for the differences in the absolute values of the variables. Table 29 and Figure 27 present the overall means and 95%-prediction intervals in order of the sensitivity of the variables. The validity of this order should be considered with care since the variables did not fulfill the requirement for the minimum number of eight results. Anyway, the sensitivity differences appear to be small and the EC50 values ranged between 4,7 and 6,1 mg/L.

5.2.2.2. NOEC

Figure 28 and Figure 29 present the NOECs obtained for 3,5-DCP. All NOECs were computed using the Williams' test, i.e. normal distribution and variance homogeneity were found to be fulfilled (α=0,01). Due to the generally lower within-laboratory variability and thus lower MDDs, NOECs were found in more cases than for *M. aquaticum*. The NOEC-reproducibility in YFW, GrFW, YTSL, and GrTSL fulfilled the ISO criteria (Table 30).

No outlying results were found; all NOECs were located within the 95%-prediction interval and ranged over three test concentration at maximum for most variables and over only two concentrations for YTSL, GrTSL, LB and TLBL. As for the EC50, the high inter-laboratory variability of the variables did not have a major effect on the absolute NOEC-level. Table 30 and Figure 30 present the overall means and 95%-prediction intervals in order of the sensitivity of the variables. The sensitivity differences were small and the NOEC ranged between 3,0 and 3,8 mg/L.

The MDDs obtained from the test laboratories (Table 30) correspond well with the expected values (Table 28). Regarding the high maximum values of the MDDs it should be considered that most of these MDDs given in Table 30 were derived from one data set (L12), exhibiting untypically high variability for all MDDs. GrTSL and GrFW showed MDDs lower than 20% (minimum 6 and 9%, respectively). All minimum MDDs ranged between 6 and 18%, indicating that in principle the toxicity test with *M. spicatum* can be conducted such that NOECs can be determined with robust statistical power.

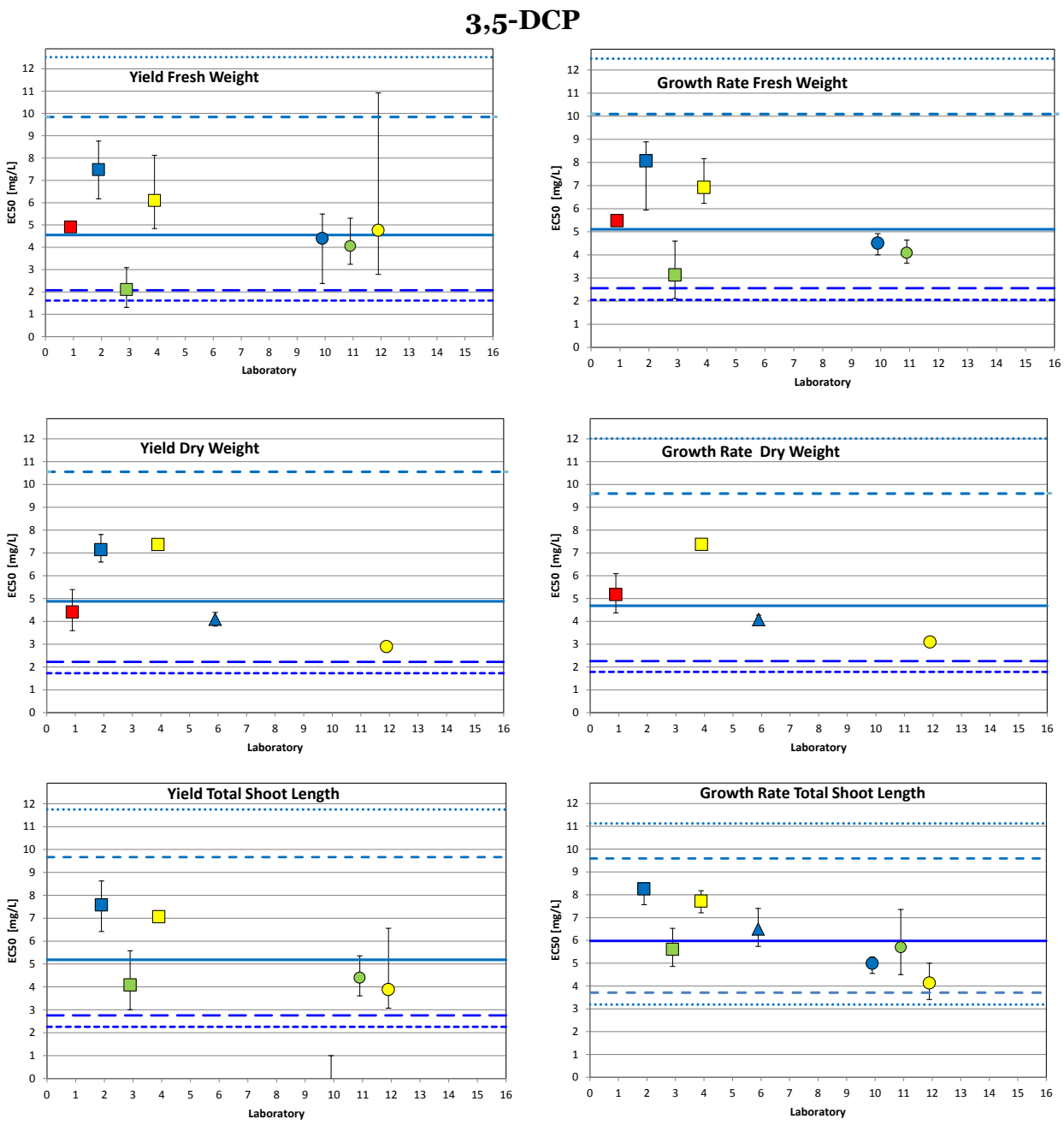


Figure 25: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield variables (left hand) and the growth rates (right hand) for 3,5-DCP in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

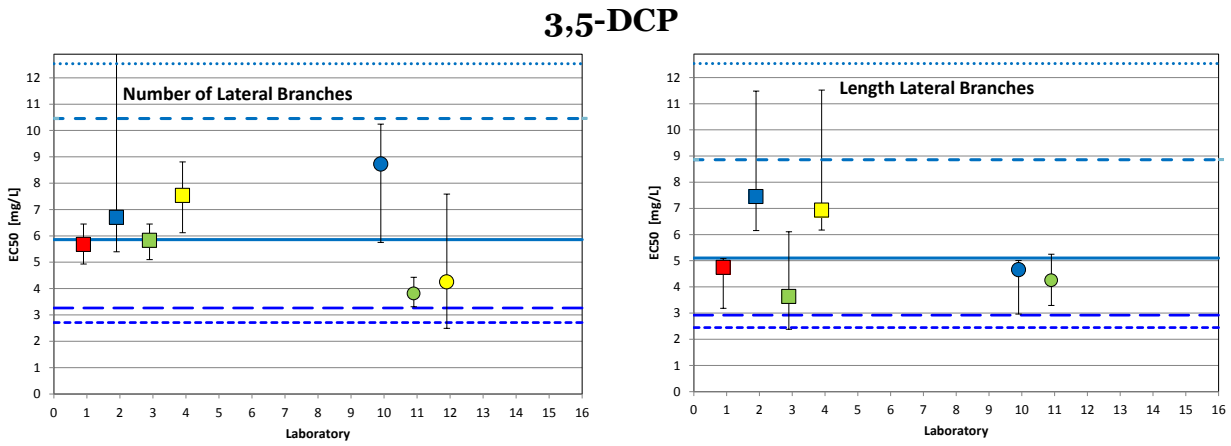
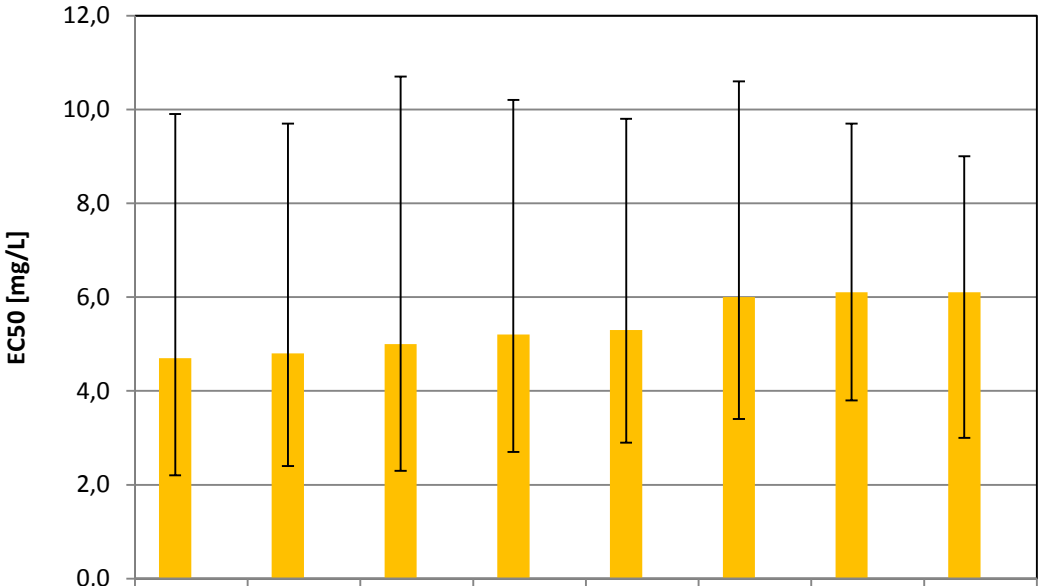


Figure 26: EC50 (symbols) and 95%-confidence interval (whiskers) for the lateral branches for 3,5-DCP in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line) . Data base: all data.

Table 29: Overall mean of the EC50 for 3,5-DCP, number of determined EC50 (n) and the 95%-prediction interval (PI) for the variables in *M. spicatum* at DAT 14. Data base: all data.

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YFW	4,7	7	2,2 - 9,9
GrDW	4,8	4	2,4 - 9,7
YDW	5,0	5	2,3 - 10,7
GrFW	5,2	6	2,7 - 10,2
YTSL	5,3	5	2,9 - 9,8
LB	6,0	7	3,4 - 10,6
GrTSL	6,1	7	3,8 - 9,7
TLBL	6,1	6	3,0 - 9,0



	YFW	GrDW	YDW	GrFW	YTSL	LB	GrTSL	TLBL
3,5-DCP	4,7	4,8	5,0	5,2	5,3	6,0	6,1	6,1
n	7	4	5	6	5	7	7	6

Figure 27: Mean EC50 (bars) and 95%-prediction interval (whiskers) for 3,5-DCP for the variables studied in *M. spicatum* at DAT 14. Data base: all data.

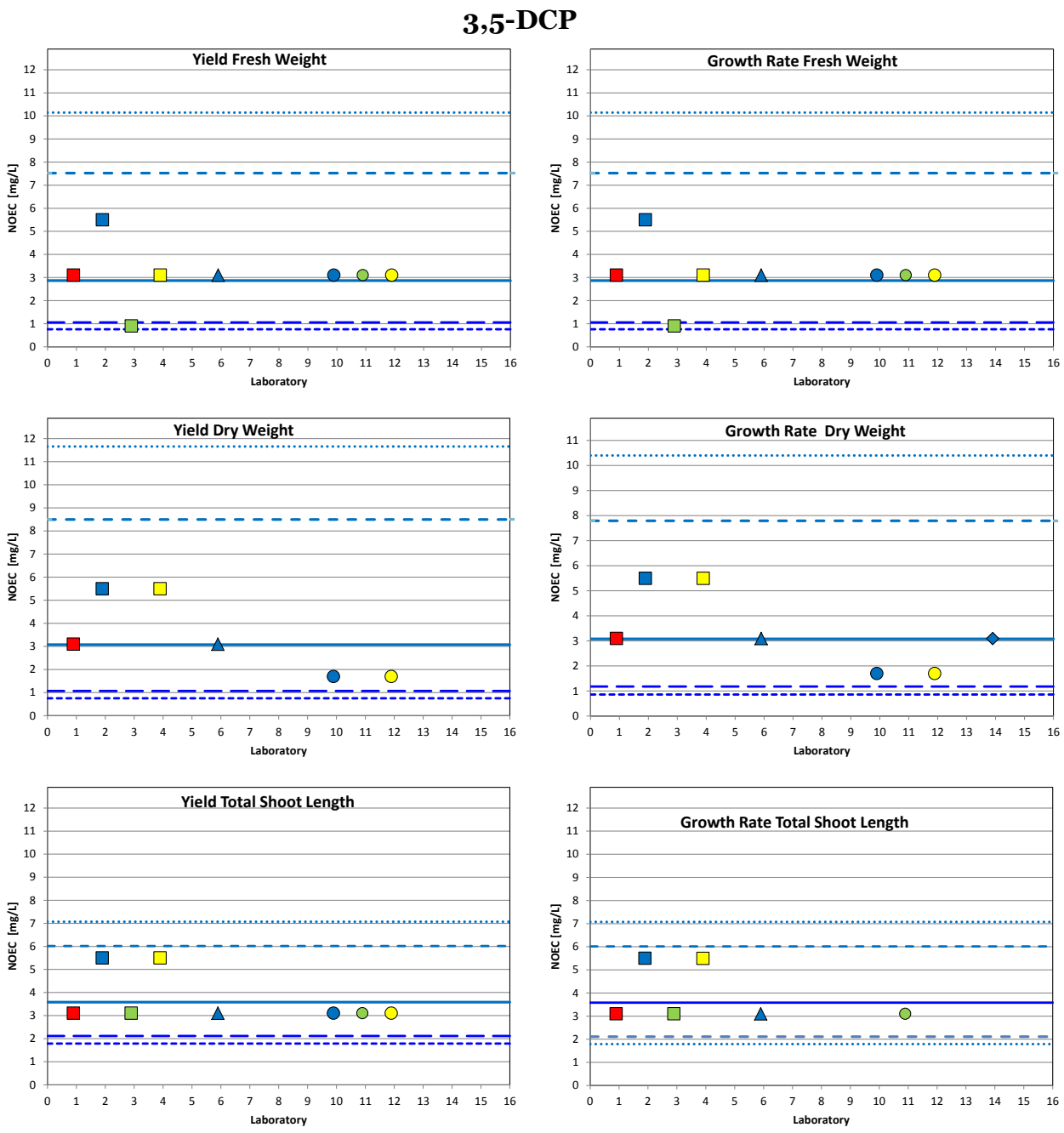


Figure 28: NOEC (symbols) for the yield variables (left hand) and the growth rates (right hand) for 3,5-DCP in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

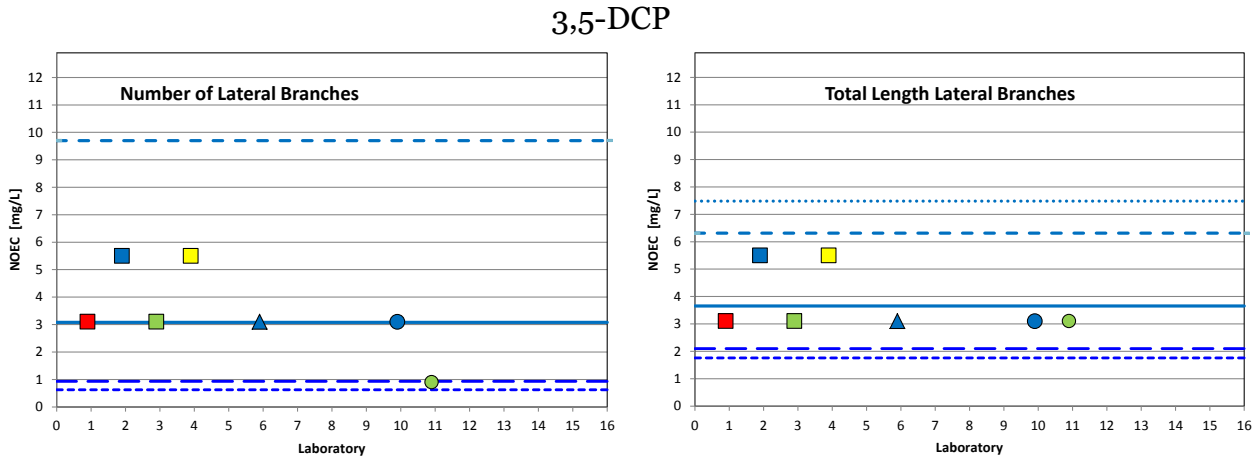
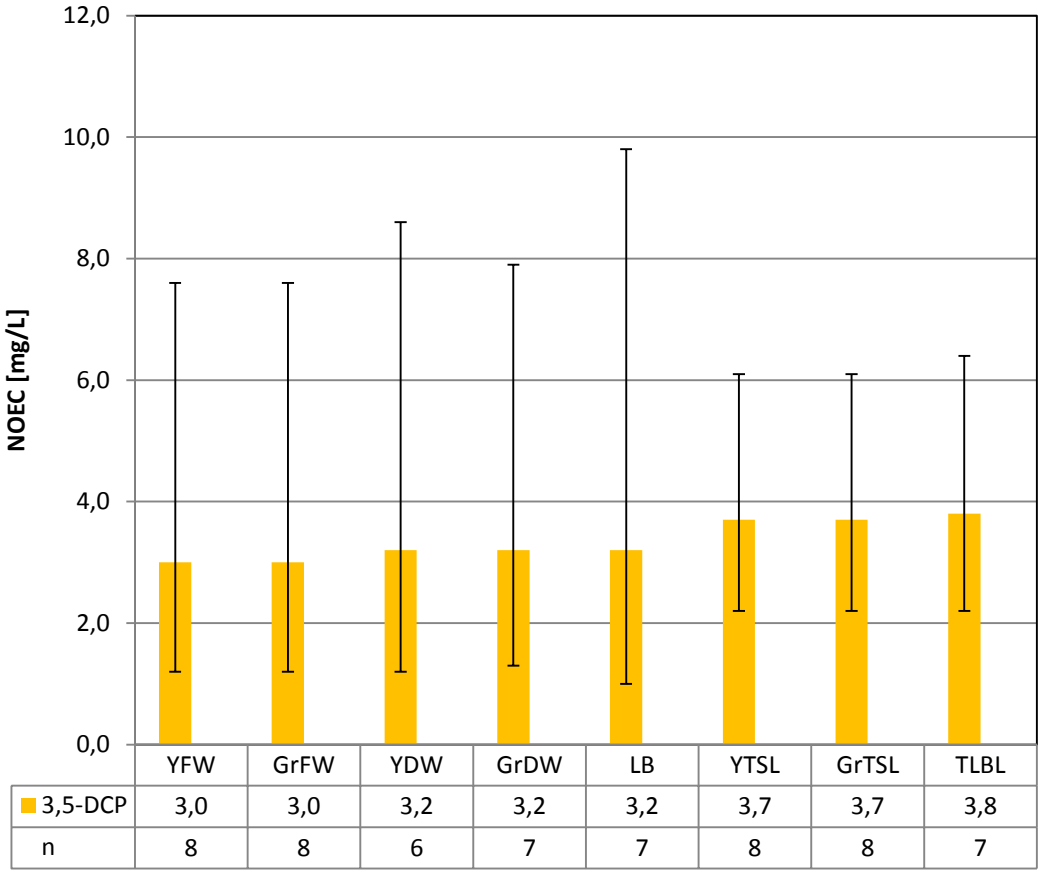


Figure 29: NOEC (symbols) for the lateral-branches variables for 3,5-DCP in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

Table 30: Overall mean of the NOEC for 3,5-DCP, number of determined NOEC (n), the 95%-prediction interval (PI) for the variables in *M. spicatum* at DAT 14. Data base: all data.

3,5 DCP							
Variable	NOEC [mg/L]			MDD [2]			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
YFW	3,0	8	1,2 - 7,6	26%	15% - 46 %	9	1 x NOEC < 1 mg/L
GrFW	3,0	8	1,2 - 7,6	19%	9% - 33%	9	1 x NOEC < 1 mg/L
YDW	3,2	6	1,2 - 8,6	28%	18% - 48%	7	1 x NOEC < 1 mg/L
GrDW	3,2	7	1,3 - 7,9	26%	14% - 49%	7	
LB	3,2	7	1,0 - 9,8	30%	15% - 74%	8	1 x NOEC ≥ 10 mg/L
YTSL	3,7	8	2,2 - 6,1	25%	13% - 58%	8	
GrTSL	3,7	8	2,2 - 6,1	17%	6% - 47%	8	
TLBL	3,8	7	2,2 - 6,4	31%	16% - 65%	8	1 x NOEC ≥ 10 mg/L

1) NOECs ≥ 10 mg/L and < 1 mg/L not included
2) 6 of 8 maximum MDDs are obtained in only one and the same data set (L12)



	YFW	GrFW	YDW	GrDW	LB	YTSL	GrTSL	TLBL
3,5-DCP	3,0	3,0	3,2	3,2	3,2	3,7	3,7	3,8
n	8	8	6	7	7	8	8	7

Figure 30: Mean NOEC (bars) and 95%-prediction interval (whiskers) obtained for the variables studied for 3,5-DCP in *M. spicatum* at DAT 14. Data base: all data.

5.2.3. Effects of Isoproturone

5.2.3.1. EC50

Figure 31, Figure 32, Table 31 and Figure 33 present all results for the EC50 for IP (5 ≤n ≤7). In contrast to M. *aquaticum*, growth promotion in low concentrations was hardly observed, if it occurred at all, weight parameters and total length lateral branches were affected. As for *M. aquaticum,* the EC50-level was markedly lower than that for 3,5-DCP by factors of 18 (LB, GrTSL) to 94 (YDW) and the sensitivity range of the variables was found to be greater (factor about 6). Growth rates generally were less sensitive than yields. The YDW and GrDW were the most sensitive variables whereas GrTSL and LB were the least sensitive ones.

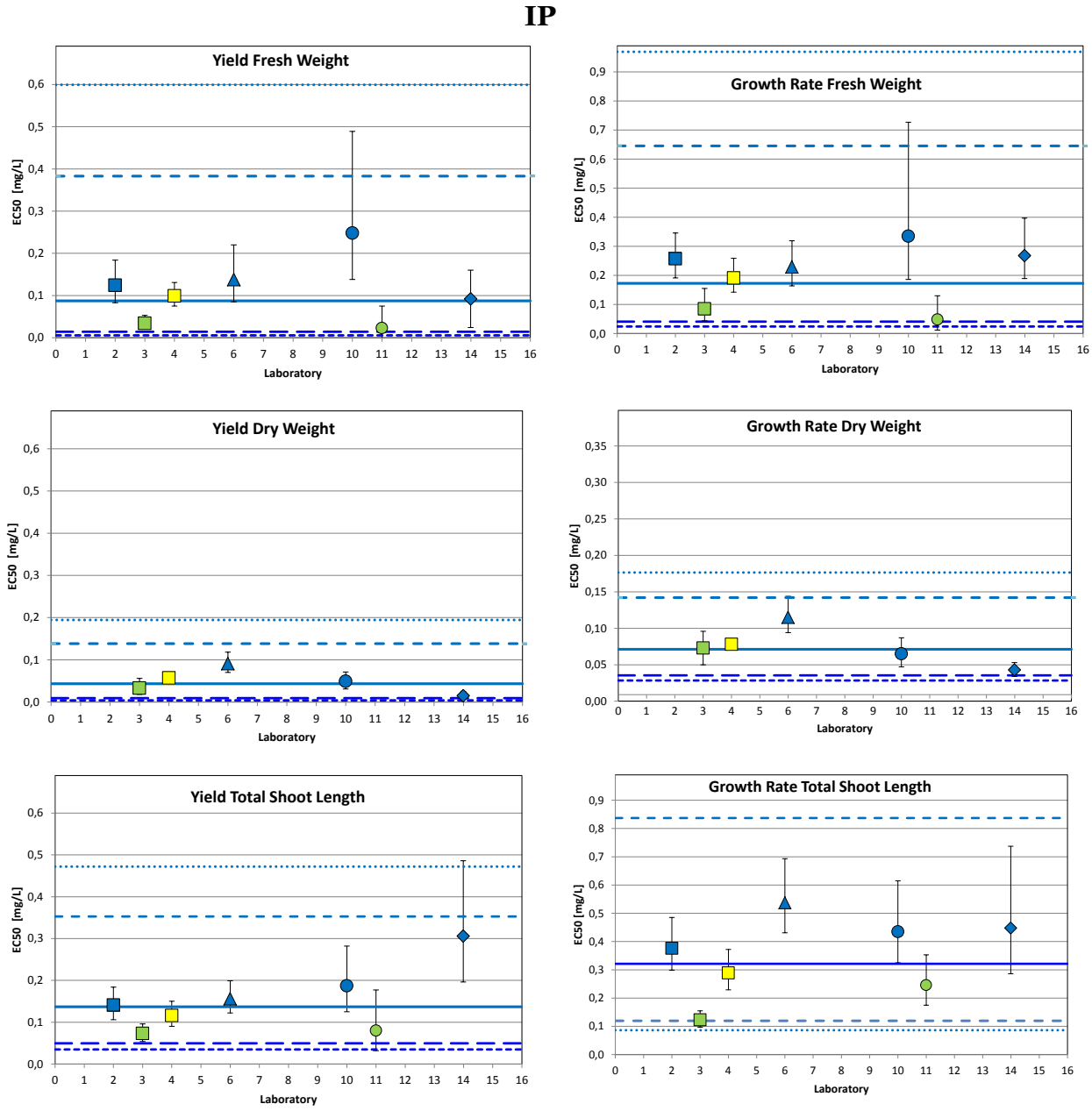


Figure 31: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield variables (left hand) and the growth rates (right hand) for IP in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

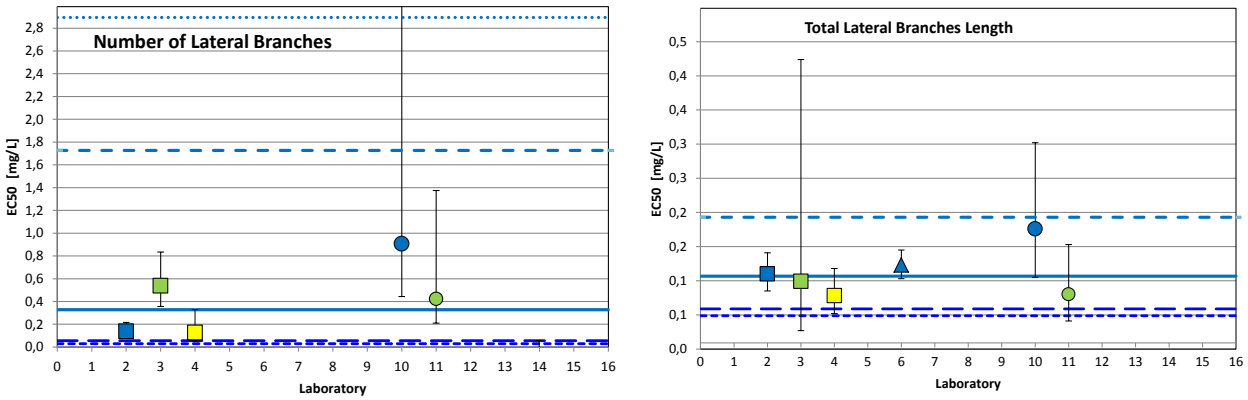
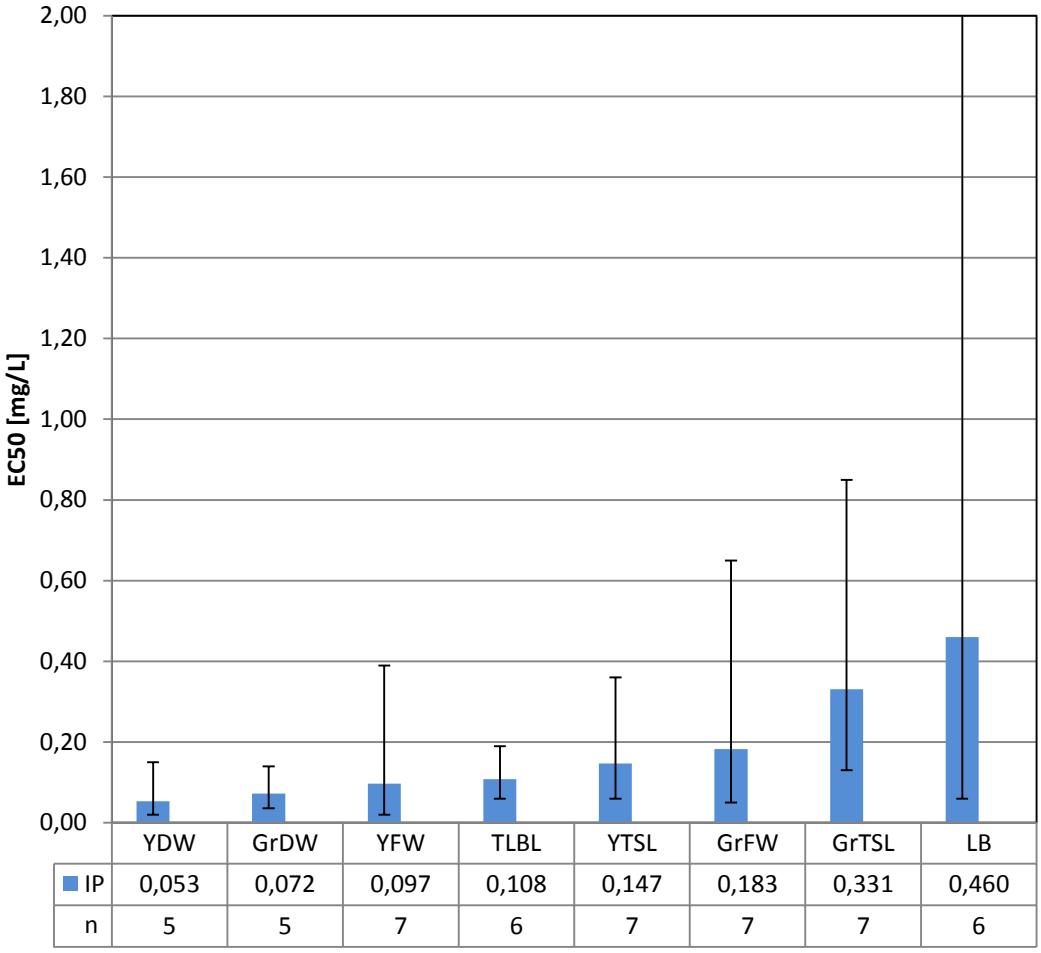


Figure 32: EC50 (symbols) and 95%-confidence interval (whiskers) for the lateral branches for IP in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

Table 31: Overall mean of the EC50 for IP, number of determined EC50 (n) and the 95%-prediction interval (PI) for the variables for IP in *M. spicatum* at DAT 14. Data base: all data.

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,053	5	0,02 - 0,15
GrDW	0,072	5	0,04 - 0,14
YFW	0,097	7	0,02 - 0,39
TLBL	0,108	6	0,06 - 0,19
YTSL	0,147	7	0,06 - 0,36
GrFW	0,183	7	0,05 - 0,65
GrTSL	0,331	7	0,13 - 0,85
LB	0,338	6	0,06 - 1,74



	YDW	GrDW	YFW	TLBL	YTSL	GrFW	GrTSL	LB
IP	0,053	0,072	0,097	0,108	0,147	0,183	0,331	0,460
n	5	5	7	6	7	7	7	6

Figure 33: Mean EC50 (bars) and 95%-prediction interval (whiskers) for the variables studied for IP in *M. spicatum* at DAT 14. Data base: all data.

5.2.3.2. NOEC

Figure 34, Figure 35, Table 32 and Figure 36 show the results for the NOEC obtained for IP. As with the EC50, differences from 3,5-DCP include the absolute NOEC-level which was markedly lower by factors of between 50 (LB) and 206 (TSL). The sensitivity range of the variables was lower than that of the EC50 (difference of a factor of 3,6). The MDDs obtained from the test laboratories correspond well with the expected values (Table 30 and Table 32). The most powerful statistical test results are to be expected in GrTSL and GrFW whereas the lateral branches variables showed the highest MDD. However, most MDDs are relatively low and minimum MDDs below 10% in GrDW, GrFW, GrTSL (7% !) and YDW indicate that some laboratories were able to reduce sources of variance and thus achieved a sound statistical power.

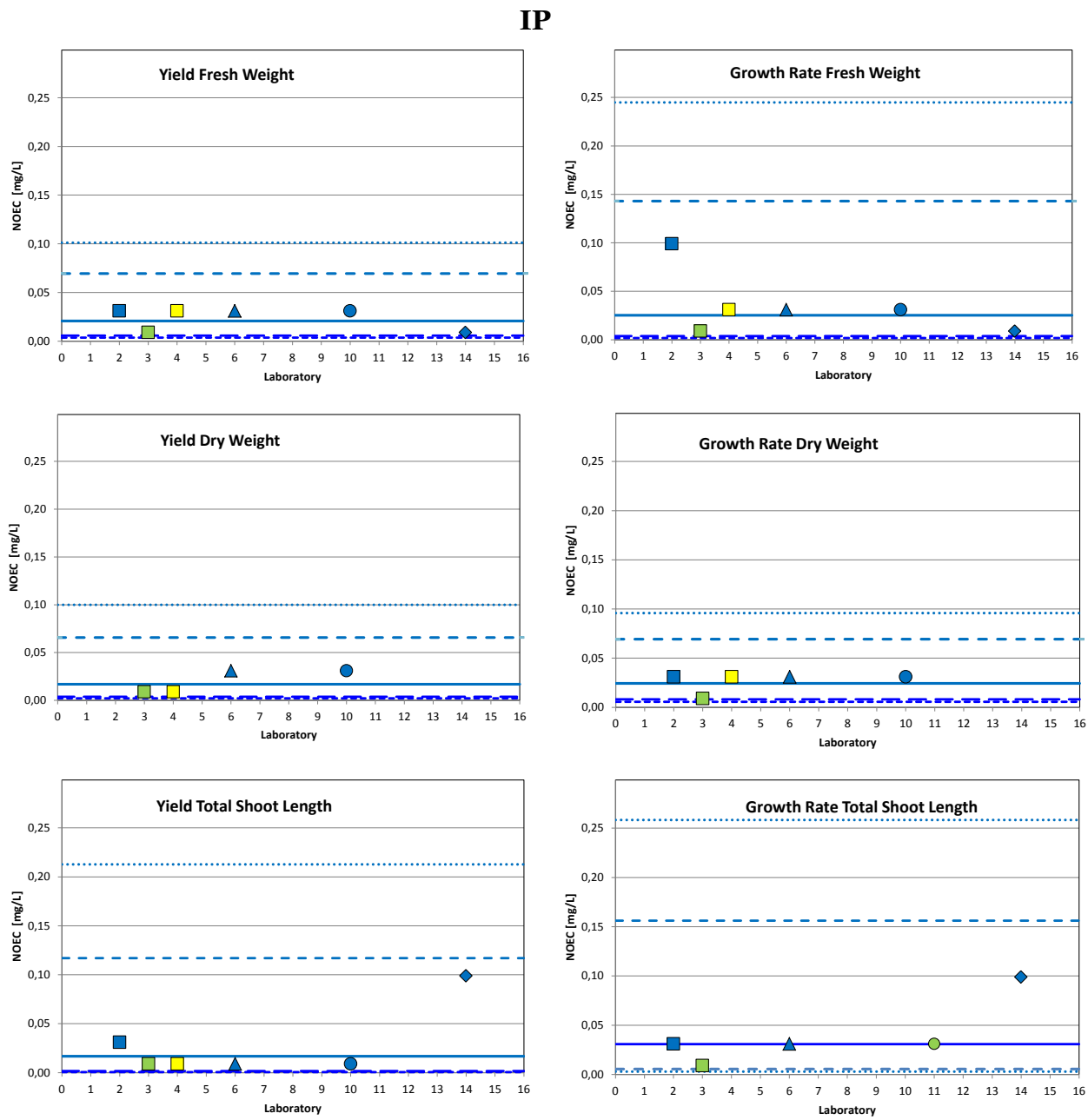


Figure 34: NOEC (symbols) for the yield variables (left hand) and the growth rates (right hand) for IP in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

IP

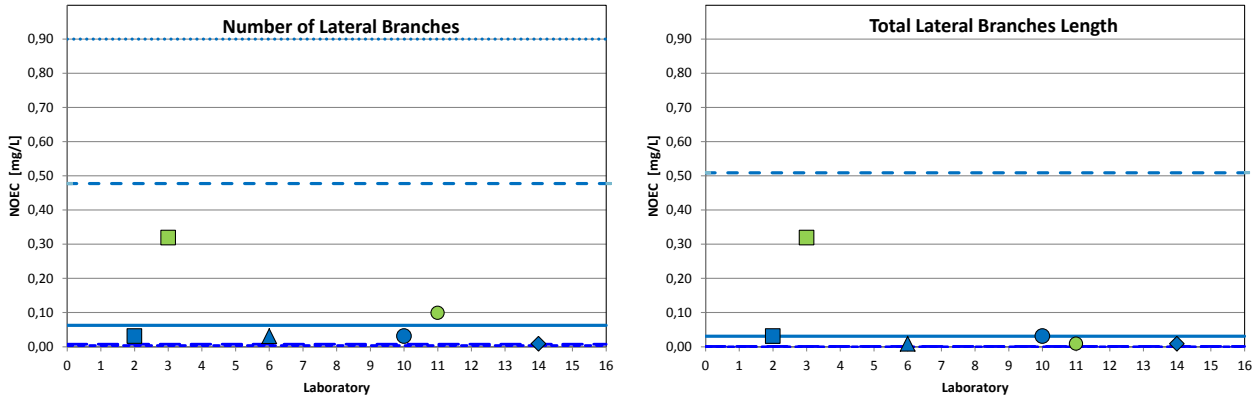
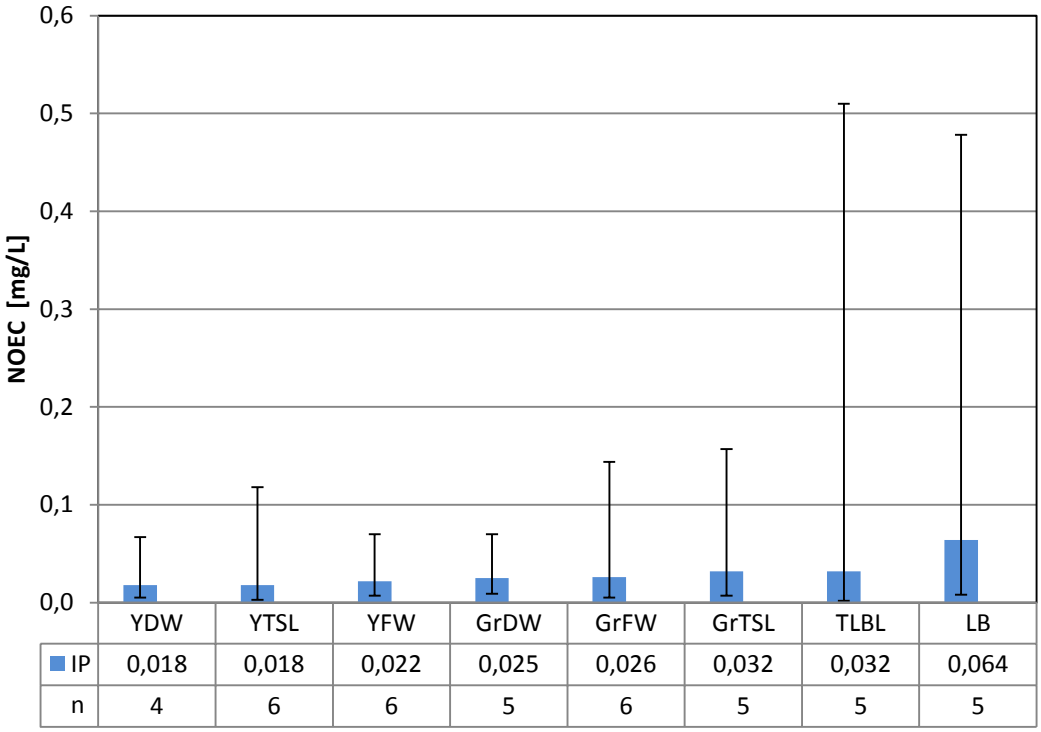


Figure 35: NOEC (symbols) obtained for the lateral-branches variables for IP in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

Table 32: Overall mean of the NOEC, number of determined NOEC (n), the 95%-prediction interval (PI) for the variables for IP in *M. spicatum* at DAT 14. Data base: all data.

IP							
Variable	NOEC [mg/L]			MDD			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
YDW	0,018	4	0,005 - 0,067	20%	9% - 36%	5	1 x NOEC < 0,01 mg/L
YTSL	0,018	6	0,003 - 0,118	19%	10% - 36%	7	1 x NOEC < 0,01 mg/L
YFW	0,022	6	0,007 - 0,070	18%	13% - 22%	7	1 x NOEC < 0,01 mg/L
GrDW	0,025	5	0,009 - 0,070	18%	8% - 31%	6	1 x NOEC < 0,01 mg/L
GrFW	0,026	6	0,005 - 0,144	14%	9% - 22%	7	1 x NOEC < 0,01 mg/L
GrTSL	0,032	5	0,007 - 0,157	13%	7% - 25%	7	2 x NOEC < 0,01 mg/L
TLBL	0,032	5	0,002 - 0,510	25%	11% - 63%	6	1 x NOEC < 0,01 mg/L
LB	0,064	5	0,008 - 0,478	26%	12% - 64%	6	1 x NOEC < 0,01 mg/L

1) NOECs $\geq$ 1 mg/L and < 0,01 mg/L not included



	YDW	YTSL	YFW	GrDW	GrFW	GrTSL	TLBL	LB
IP	0,018	0,018	0,022	0,025	0,026	0,032	0,032	0,064
n	4	6	6	5	6	5	5	5

Figure 36: Mean NOEC (bars) and 95%-prediction interval (whiskers) for the variables studied for IP in *M. spicatum* at DAT 14. Data base: all data.

5.2.4. Effects of Trifluralin

5.2.4.1. EC50

Figure 37, Figure 38, Table 33 and Figure 39 give the results for the EC50 for TF. As for *M. aquaticum* growth promotion at lower concentrations and an inconsistent dose response were observed. Maximum inhibitions were below 50% in some cases. This resulted in failure to estimate EC50 values. For the dry weight variables (YDW, GrDW) no EC50 values were obtained at all and only 1 to 4 EC50 values were obtained for the remaining variables. The reason for these results may be seen in the dissipation of the test substance during exposure time and / or in the selected dose range. The sensitivity differences between the variables were small (min YTSL: 0,19 mg/L (n = 4); max GrFW: 0,84 mg/L (n = 1)) with growth rates being less sensitive than yields.

The results of the data set in which unclear test concentrations were used (L06) cannot be shown in Figure 37 and Figure 38 since the EC50 values obtained exceeded the scale of the figures (0,9 mg/L to 2,7 mg/L). As discussed, this gives evidence that wrong concentrations were reported.

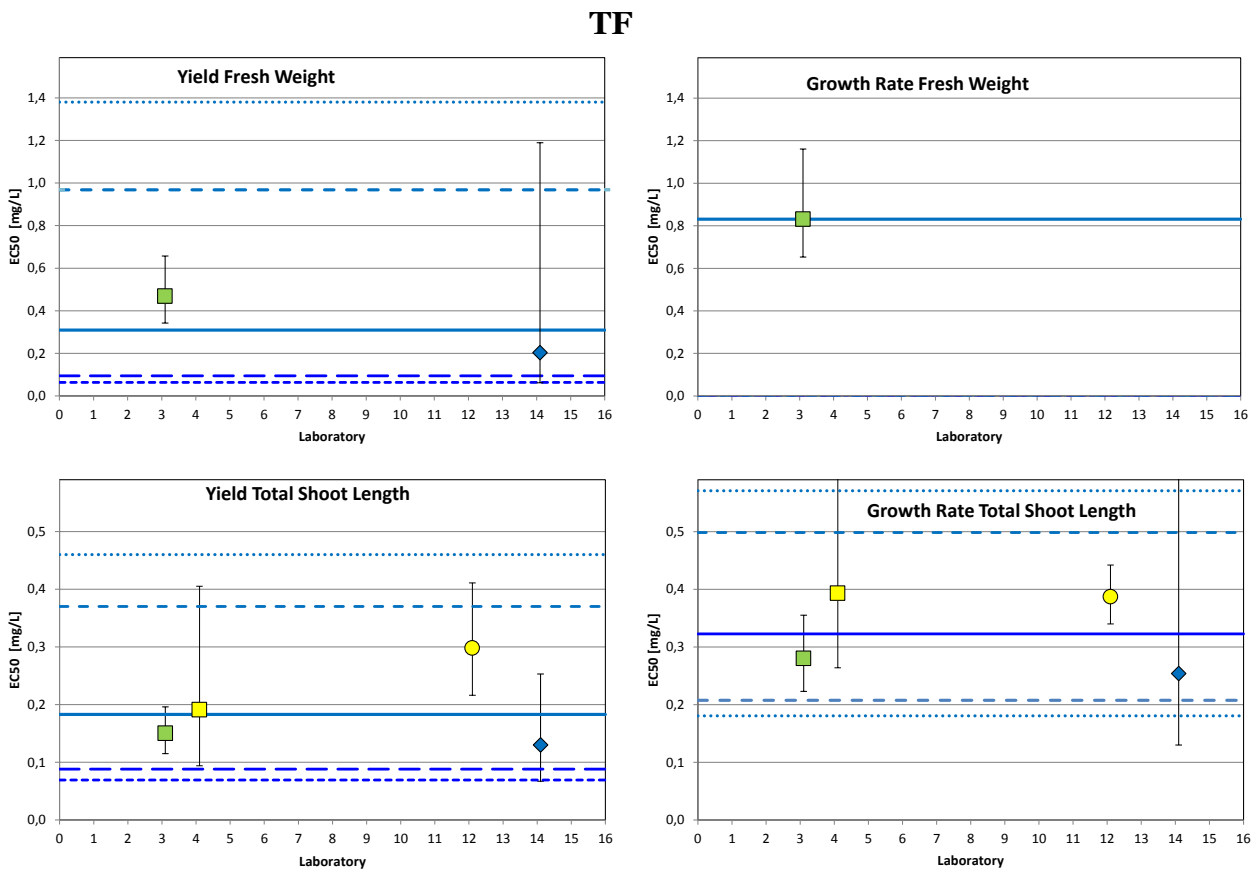


Figure 37: EC50 (symbols) and 95%-confidence interval (whiskers) for the yield variables (left hand) and the growth rates (right hand) for TF in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

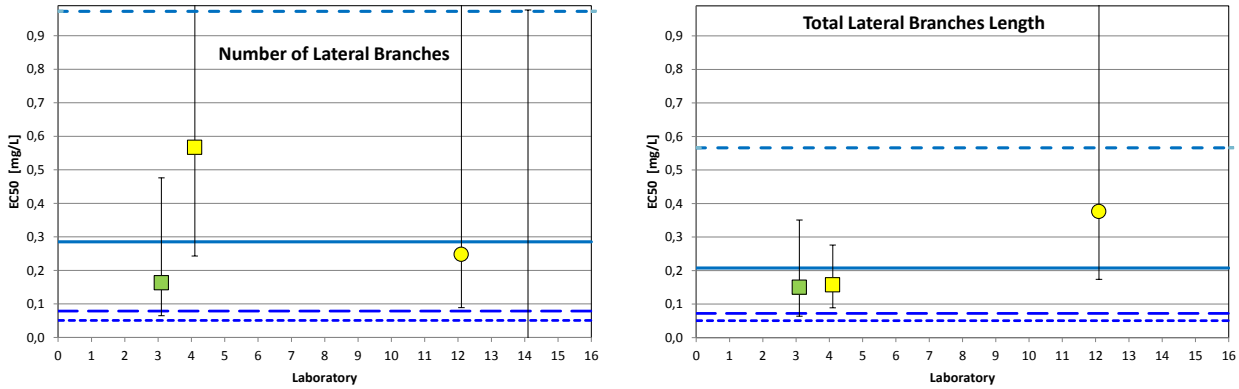


Figure 38: EC50 (symbols) and 95%-confidence interval (whiskers) for the lateral branches for TF in *M. spicatum* at DAT 14 together with the overall geometric mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

Table 33: Overall mean of the EC50, number of determined EC50 (n) and the 95%-prediction interval (PI) for the variables for TF in *M. spicatum* at DAT 14

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to no clear dose-response relation or not significant due to too few / low inhibition values		
GrDW			
YTSL	0,193	4	0,10 - 0,38
TLBL	0,218	3	0,08 - 0,58
LB	0,295	3	0,09 - 0,98
YFW	0,319	2	0,10 - 0,98
GrTSL	0,333	4	0,22 - 0,51
GrFW	0,840	1	--

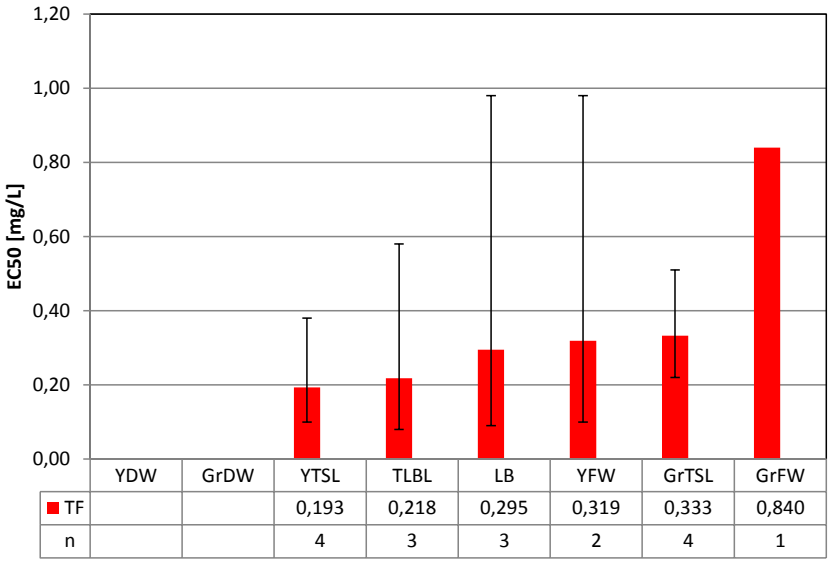


Figure 39: Mean EC50 (bars) and 95%-prediction interval (whiskers) for the variables for TF in *M. spicatum* at DAT 14. Data base: all data.

5.2.4.2. NOEC

Figure 40, Figure 41, Table 34 and Figure 42 present the results for the NOEC for TF. Figure 40 and Figure 41 also give values of a data set, in which two-fold higher concentrations were reported (L06, indicated by pink color). As discussed above, the results were not in line with the reported concentration range and, therefore as a precaution, were not considered for the overall statistical analysis.

TLBL, YDW and GrDW appear to be the most sensitive variables. The smallest mean MDDs were to be observed for GrTSL (16%), YTSL (19% and GrDW (20%), indicating an expected acceptable statistical power for determination of NOECs.

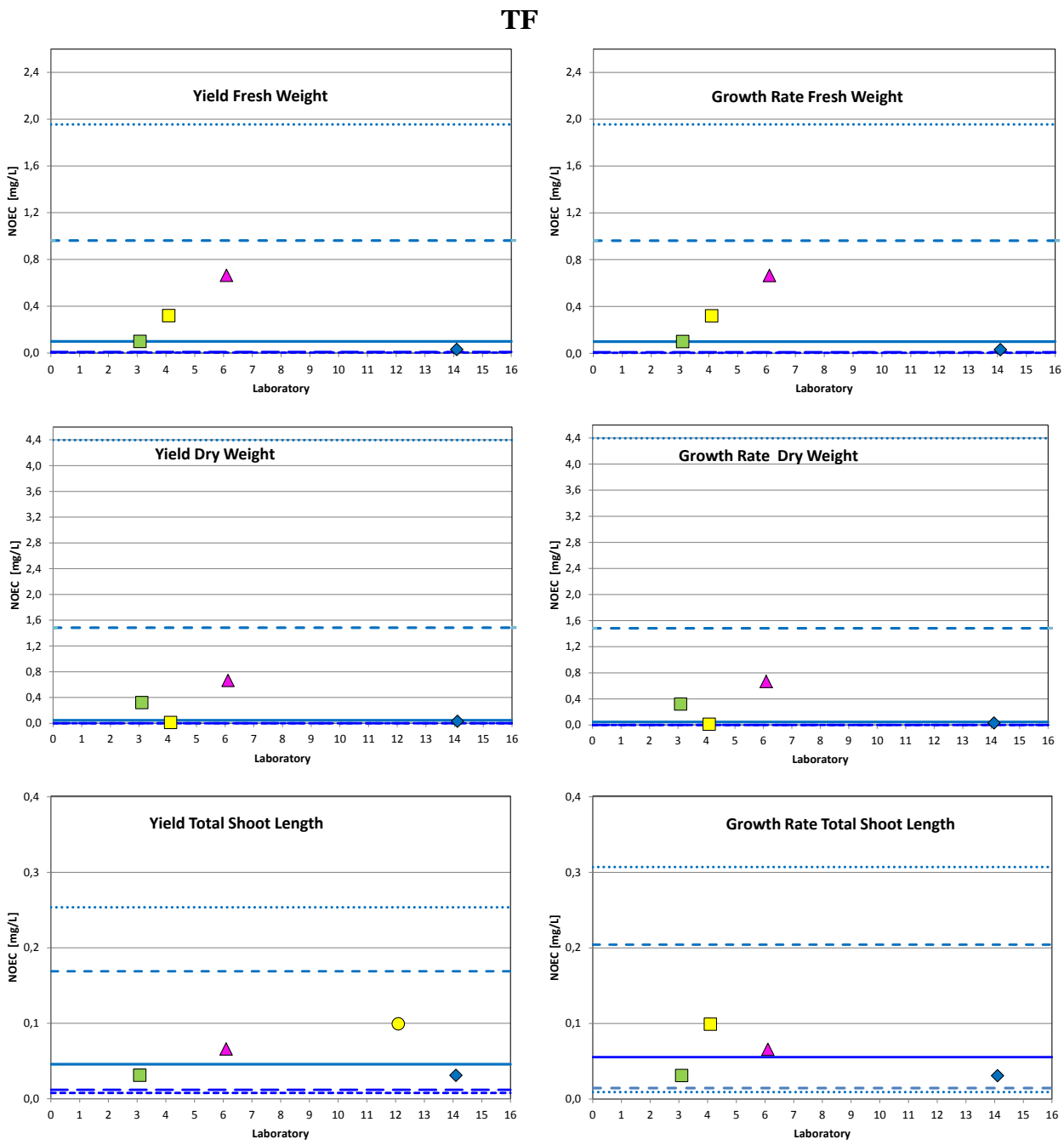


Figure 40: NOEC (symbols) for the yield variables (left hand) and the growth rates (right hand) for TF in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Pink symbol: test with two-fold higher concentrations reported. Data base: all data.

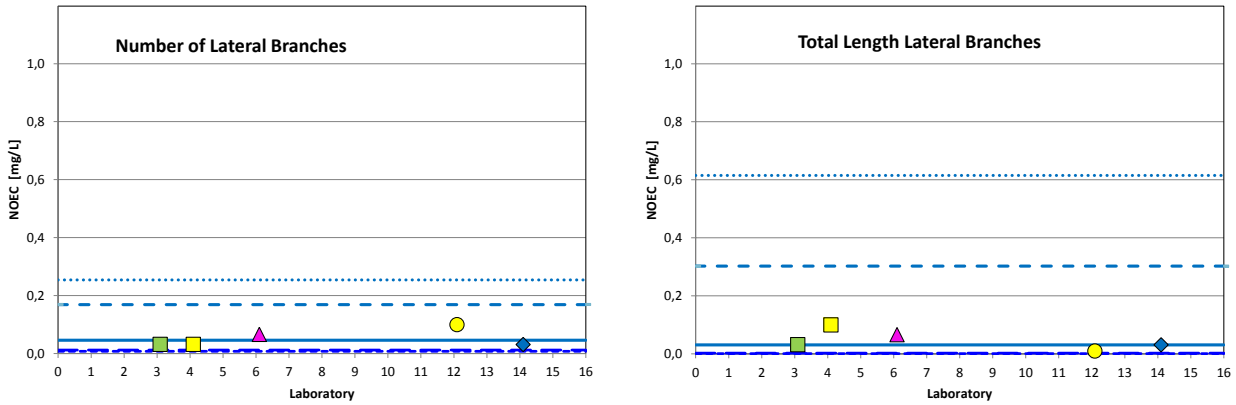
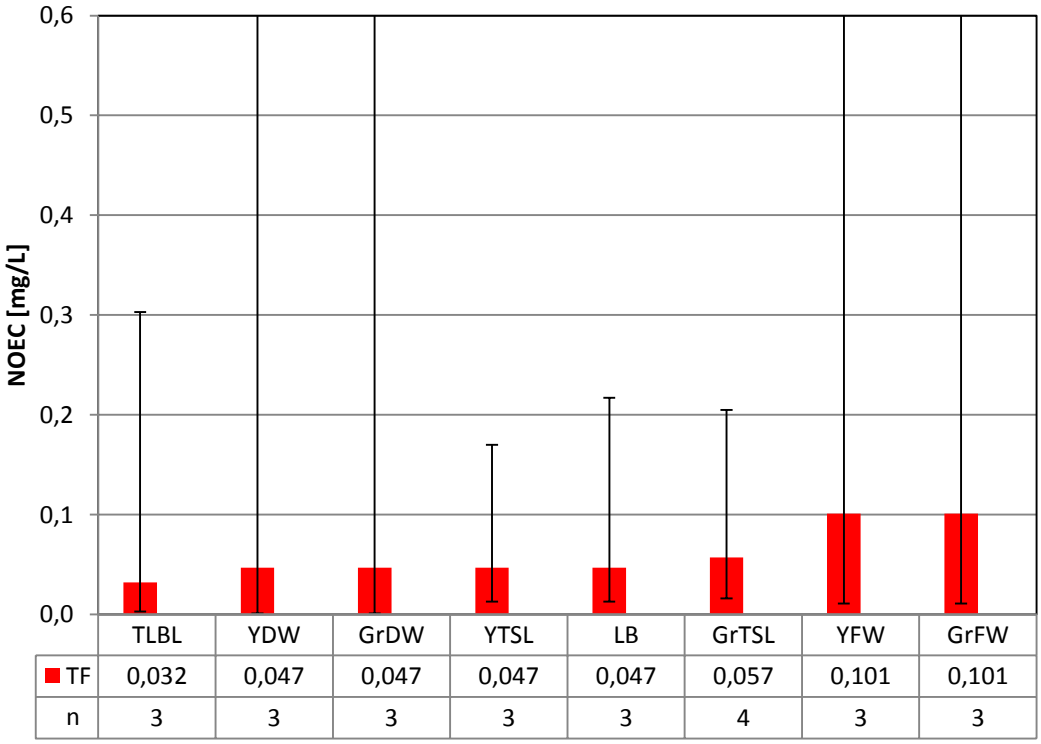


Figure 41: NOEC (symbols) for the lateral-branches variables for TF in *M. spicatum* at DAT 14 together with the overall mean (thick blue line), the 95%-prediction interval (dashed line) and the 99%-prediction interval (dotted line). Data base: all data.

Table 34: Overall mean of the NOEC, number of determined NOEC (n), the 95%-prediction interval (PI) for variables for TF in *M. spicatum* at DAT 14. Data base: all data.

TF							
Variable	NOEC [mg/L]			MDD [2]			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
TLBL	0,032	3	0,003 - 0,303	32%	25% - 45%	3	
YDW	0,047	3	0,001 - 1,484	34%	11% - 68%	4	1 x NOEC ≥ 1 mg/L
GrDW	0,047	3	0,001 - 1,484	20%	7% - 56%	4	1 x NOEC ≥ 1 mg/L
YTSL	0,047	3	0,013 - 0,170	19%	14% - 23%	4	1 x NOEC < 0,01 mg/L
LB	0,047	3	0,013 - 0,170	33%	23% - 43%	3	
GrTSL	0,057	4	0,016 - 0,205	16%	9% - 26%	4	
YFW	0,101	3	0,011 - 0,963	44%	20% - 100%	4	1 x NOEC ≥ 1 mg/L
GrFW	0,101	3	0,011 - 0,963	37%	13,5% - 94%	4	1 x NOEC ≥ 1 mg/L

1) NOECs ≥ 1 mg/L and < 0,01 mg/L not included
2) All maximum MDDs are obtained in only one and the same data set (L12)



	TLBL	YDW	GrDW	YTSL	LB	GrTSL	YFW	GrFW
TF	0,032	0,047	0,047	0,047	0,047	0,057	0,101	0,101
n	3	3	3	3	3	4	3	3

Figure 42: Mean NOEC (bars) and 95%-prediction interval (whiskers) for the variables studied for TF in *M. spicatum* at DAT 14. Data base: all data.

5.2.5. Effects on Root Development

Figure 43 presents the results of the semi-quantitative root assessment. As for *M. aqua*ticum, the numbers of individual plants (not replicates) assigned to a certain level of root development ("very good root development", "moderate root development", "few roots", "no roots") in each data set were summed up and the sum was expressed as percentage of total plants exposed to a certain treatment. For exposure to 3,5-DCP, root assessments were performed in 8 data sets, corresponding to a total of 144 control plants (8 x 18) and 72 plants (8 x 9) per treatment. In the 5,6 mg/L treatment, the overall sum of plants assigned to level 1 ("very good root assessment") was 18 out of 72, corresponding to 25% of all plants being exposed to this concentration.

As for *M. aquaticum* there is a clear trend that root development was affected by all test items. For all test substances, a successive increase of the fraction exhibiting lower level of root development was observed with increasing concentrations. Effects of 3,5-DCP and IP were found to exceed those of TF.

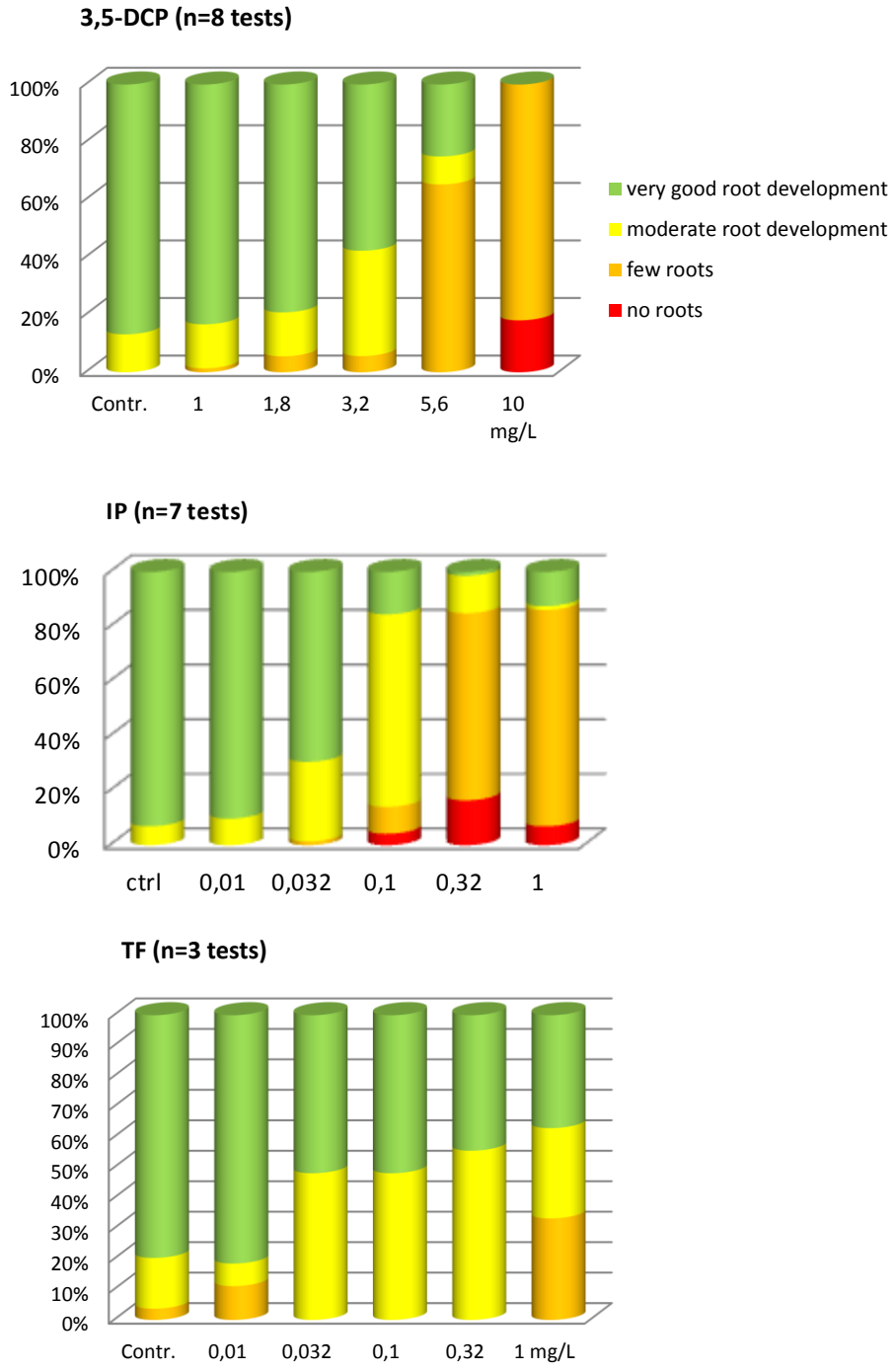


Figure 43: Result of the qualitative assessment of root development in *M. spicatum* at DAT 14 for all test items. Please note that percentages are related to total number of plants assessed, i.e. in case of three tests, 100% correspond to 54 (3 x 18) control plants and 27 (3 x 9) plants per treatment. Data base: all data.

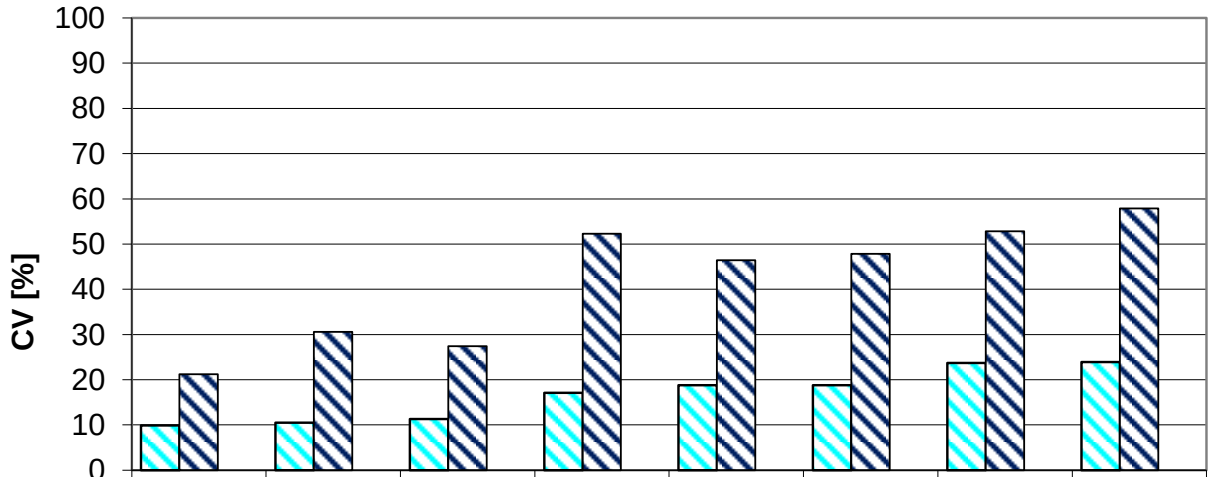
5.2.6. Effects of Validity Criteria on the Results

In *M. spicatum,* the variability was relatively low. Two tests (10%) failed to fulfill doubling of shoot length, four tests (18%) failed to fulfill doubling of fresh weight. To ensure consistency with the results of *M. aquaticum,* a second evaluation run was performed applying the same validity criteria as in *M. aquaticum*: a minimum growth rate of 0,07 d^{-1} in GrFW (corresponding to a doubling time of 9,9 days) and a CV% in YFW not greater than 35% . As can be depicted from Table 35, only 3 out of 21 data sets (14%) failed to fulfill the validity criteria, i.e. 86% of the datasets were valid. One of the excluded data sets is L12 TF, where in view of the large variability it was obvious that something went wrong with the test. In contrast, in the two invalid tests of L10 as well relatively strong growth (Shoot length increase by factor 6 -7) and low variability was to be observed. The results below show that exclusion of the L10 data sets even resulted in increased mean MDDs for NOEC determination. Nevertheless, in both tests the growth rates for fresh weight were only 0,038 d^{-1} and 0,032 d^{-1}, respectively (corresponding to doubling times of 18,2 days and 21,7 days, respectively). Regarding the outlying high absolute fresh weights at the test start (see Figure 21), there may have been a methodological problem with fresh weight measurement. For reasons of consistency both data sets were left out of the second evaluation run.

Table 35: Overview of tests performed, tests evaluable, tests valid (x) and tests invalid (red) and the reason for invalidity in *M. spicatum*; Gr: Growth rate for Fresh weight < 0,07 d^{-1} ; CV: CV% for Yield Fresh Weight > 35%.

Laboratory	M. spicatum			Tests performed	Tests valid
	3,5-DCP	IP	TF		
L01	x			1	1
L02	x	x		2	2
L03	x	x	x	3	3
L04	x	x	x	3	3
L05					
L06	x	x	x 1)	3	3
L07					
L08	contamin.	contamin.	contamin.	3	0
L09					
L10	Gr	Gr		2	0
L11	x	x		2	2
L12	x	TSL d0 17 cm	Gr, CV	3	1
L13					
L14	x	x	x	3	3
L15					
Tests performed	10	9	6	25	
Tests evaluated	9	7	5	21	
Tests valid	**8**	**6**	**4**	**18**	
1) test concentrations questionable					

Figure 44 and Table 37 demonstrate that the exclusion of the three invalid data sets had only minor effects on the repeatability (CV_r%) and reproducibility coefficient of variation (CV_R%). The lowest CV_r% were found in the GrFW (10%; all data 14%), the GrTSL (11%; all data 10 %) and the GrDW (11%; all data 14%). For the remaining variables the CV_r% were around 20%.



	GrFW	GrTSL	GrDW	YTSL	YFW	YDW	TLBL	LB
CVr%	9,9	10,5	11,3	17,1	18,8	18,8	23,7	23,9
CVR%	21,2	30,6	27,4	52,3	46,4	47,8	52,8	57,9
n	18	18	15	18	18	15	15	15

Figure 44: Coefficients of variation for repeatability (CV_r%, light blue) and reproducibility (CV_R%, dark blue) of all valid control data sets ordered by magnitude in *M. spicatum;* n: number of valid control data sets available.

As a consequence, the calculated MDDs were mostly reduced (Table 37) but the range remained similar: 12,7% to 30,6% (all data; 13,1% to 30,9%). Although the differences are small, the results indicate that the application of validity criteria can increase the statistical power of the toxicity test.

Regarding the toxic metrics, excluding the three invalid data sets did not markedly affect the overall results except some calculated MDDs. Therefore, the resulting EC50 and NOECs of the second evaluation run based only on valid data sets is presented only by overview tables rather than by tables and figures. Also, for root development the results of the second evaluation run show the same qualitative trends as were observed using all data and will not be shown again.

Table 38 gives the results for the EC50, Table 39 to Table 41 those for the NOEC and observed MDD. The EC50 fell between 4,7 and 6,3 mg/L for 3,5-DCP (all data: 4,7 and 6,1 mg/L), between 0,052 and 0,315 mg/L for IP (all data: 0,053 and 0,338) and between 0,163 and 0,840 mg/L for TF (all data: 0,193 and 0,840 mg/L).

Table 36: Overview of the ring-test statistics before and after exclusion of invalid data sets of controls in *M. spicatum*; overall mean: mean of all data from all controls and all laboratories; CV_r%: repeatability coefficient of variation; CV_R%: reproducibility coefficient of variation; 95%-/99%-PI: 95%-/99%-prediction interval; p: number of control data sets

	Overall mean	CV_r%	CV_R%	CV_R%/ CV_r%	95% -PI	99%-PI	p
Valid Data Sets							
YFW [mg]	894,3	18,8	46,4	2,5	137,7 - 1651	-100,0 - 1889	18
YDW [mg]	74,7	18,8	47,8	2,5	9,0 - 140,4	-11,6 - 161,1	15
YTSL [cm]	27,9	17,1	52,3	3,1	0,217 - 55,7	-8,5 - 64,4	18
LB [cm]	4,68	24,0	57,9	2,4	-0,30 - 9,67	-1,9 - 11.2	15
TLBL [cm]	19,5	23,7	52,8	2,2	0,66 - 38,3	-5,2 - 44,2	15
GrFW d^{-1}	0,103[15]	9,9	21,2	2,1	0,065 - 0,140	0.053 - 0,152	18
GrDW d^{-1}	0,080[16]	11,3	27,4	2,4	0,040 – 0,119	-0,028 - 0,131	15
GrTSL d^{-1}	0,106[17]	10,5	30,6	2,9	0,044 – 0,167	0,025 - 0,186	18
All Data Sets							
YFW [mg]	885	21,6	48,9	2,3	92,2 – 1678	-157,0 - 1927	21
YDW [mg]	73,6	21,8	46,9	2,1	12,0 - 135,2	-7,3 - 154,5	18
YTSL [cm]	28,0	16,6	51,7	3,1	0,55 - 55,4	-8,1 - 64,0	21
LB [cm]	4,97	24,1	57,4	2,4	-0,29 - 10,24	-1,9 - 11,9	18
TLBL [cm]	19,6	24,2	54,4	2,3	0,028 - 39,3	-0,62- 6,13	18
GrFW d^{-1}	0,093[18]	13,8	34,5	2,5	0,018 – 0,173	0,034 - 0,15	21
GrDW d^{-1}	0,076[19]	14,3	29,7	2,1	0,036 – 0,116	-0,024 - 0,128	18
GrTSL d^{-1}	0,107[20]	10,2	30,4	3,0	0,045 – 0,168	0,026 - 0,188	21

[15] Corresponding to a doubling time of 6,7 days
[16] Corresponding to a doubling time of 8,7 days
[17] Corresponding to a doubling time of 6,5 days
[18] Corresponding to a doubling time of 7,5 days
[19] Corresponding to a doubling time of 9,1 days
[20] Corresponding to a doubling time of 6,5 days

Table 37: Detectable effect levels (MDD%) for different variables in *M. spicatum* using Williams´test. MDDs calculated assuming the given test design (control n = 6; 5 treatments with n = 3) and mean control variances (i.e. repeatabilities) as obtained in the present ring test before and after exclusion of invalid data sets of controls (see Table 36). Validity criteria were a minimum growth rate for Fresh weight of 0,07 d^{-1} and a maximum CV for Yield Fresh Weight of 35%. CV_r% = repeatability coefficient of variation P = number of control data sets from which mean CV_r% was calculated

	Valid Data Sets			All Data Sets		
	CV_r%	MDD%	p	CVr%	MDD%	p
YFW	18,8	24,1	18	21,6	27,7	21
YDW	18,8	24,1	15	21,8	28,0	18
YTSL	17,1	21,9	18	16,6	21,3	21
LB	23,9	30,6	15	24,1	30,8	18
TLBL	23,7	30,3	15	24,2	30,9	18
GrFW	9,9	12,7	18	13,8	17,6	21
GrDW	11,3	14,5	15	14,3	18,3	18
GrTSL	10,5	13,4	18	10,2	13,1	21

The NOEC ranged from 2,9 to 3,9 mg/L for 3,5-DCP (all data: 3,0 to 3,8), from 0,015 to 0,032 mg/L for IP (all data: 0,018 to 0,64 mg/L) and from 0.032 to 0,101 mg/L for TF, which is the same range as in all data. Hence, the qualitative trend was similar as for *M. aquaticum*.

Also for root development, the results of the second evaluation run showed the same qualitative trends as using all data (not shown).

Table 38: Overview of the EC50 for the valid data sets as compared to those for all data sets in *M. spicatum* at DAT 14

Valid Data Sets

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	5,0	5	2,3 - 10,7
GrDW	4,8	4	2,4 - 9,7
YFW	4,7	6	2,0 - 10,7
Gr FW	5,3	5	2,5 - 11,2
YTSL	5,3	5	2,9 - 9,8
GrTSL	6,3	6	3,9 - 10,1
LB	5,6	6	3,4 - 9,2
TLBL	5,3	5	2,9 - 9,7

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,052	4	0,016 - 0,168
GrDW	0,074	4	0,034 - 0,161
YFW	0,083	6	0,024 - 0,280
Gr FW	0,164	6	0,046 - 0,581
YTSL	0,140	6	0,054 - 0,362
GrTSL	0,315	6	0,117 - 0,849
LB	0,263	4	0,065 - 1,074
TLBL	0,097	5	0,066 - 0,143

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to no clear dose-response relation or not significant due to too few / low inhibition values		
GrDW			
YFW	0,319	2	0,104 - 0,978
Gr FW	0,840	1	-
YTSL	0,165	3	0,115 - 0,236
GrTSL	0,314	3	0,203 - 0,485
LB	0,316	2	0,060 - 1,677
TLBL	0,163	2	0,154 - 0,173

All Data Sets

3,5 DCP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	5,0	5	2,3 - 10,7
GrDW	4,8	4	2,4 - 9,7
YFW	4,7	7	2,2 - 9,9
Gr FW	5,2	6	2,7 - 10,2
YTSL	5,3	5	2,9 - 9,8
GrTSL	6,1	7	3,8 - 9,7
LB	6,0	7	3,4 - 10,6
TLBL	6,1	6	3,0 - 9,0

IP			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	0,053	5	0,02 - 0,15
GrDW	0,072	5	0,04 - 0,14
YFW	0,097	7	0,02 - 0,39
Gr FW	0,183	7	0,05 - 0,65
YTSL	0,147	7	0,06 - 0,36
GrTSL	0,331	7	0,13 - 0,85
LB	0,338	6	0,06 - 1,74
TLBL	0,108	6	0,06 - 0,19

TF			
Variable	EC50 [mg/L]		
	Overall mean	n	95% PI
YDW	not evaluable due to no clear dose-response relation or not significant due to too few / low inhibition values		
GrDW			
YFW	0,319	2	0,10 - 0,98
Gr FW	0,840	1	--
YTSL	0,193	4	0,10 - 0,38
GrTSL	0,333	4	0,22 - 0,51
LB	0,295	3	0,09 - 0,98
TLBL	0,218	3	0,08 - 0,58

Table 39: Overview of the NOEC for 3,5-DCP for the valid data sets as compared to those for all data sets in *M. spicatum* at DAT 14

Valid Data Sets

3,5 DCP							
variable	NOEC [mg/L]			MDD [2]			remarks
	geom mean [1]	n	95% PI	mean	range	n	
YFW	2,9	7	1,1 - 8,1	28%	15% - 46 %	8	1 x NOEC < 1 mg/L
GrFW	2,9	7	1,1 - 8,1	20%	9% - 33%	8	1 x NOEC < 1 mg/L
YDW	3,6	5	1,4 - 9,0	29%	20% - 48%	6	1 x NOEC < 1 mg/L
GrDW	3,5	6	1,5 - 8,1	28%	14% - 49%	6	
YTSL	3,8	7	2,2 - 6,4	26%	13% - 58%	7	
GrTSL	3,8	7	2,2 - 6,4	18%	6% - 47%	7	
LB	3,2	6	0,9 - 10,9	31%	15% - 74%	7	1 x NOEC ≥ 10 mg/L
TLBL	3,9	6	2,2 - 6,8	33%	20% - 65%	7	1 x NOEC ≥ 10 mg/L

All Data Sets

3,5 DCP							
Variable	NOEC [mg/L]			MDD [2]			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
YFW	3,0	8	1,2 - 7,6	26%	15% - 46 %	9	1 x NOEC < 1 mg/L
GrFW	3,0	8	1,2 - 7,6	19%	9% - 33%	9	1 x NOEC < 1 mg/L
YDW	3,2	6	1,2 - 8,6	28%	18% - 48%	7	1 x NOEC < 1 mg/L
GrDW	3,2	7	1,3 - 7,9	26%	14% - 49%	7	
YTSL	3,7	8	2,2 - 6,1	25%	13% - 58%	8	
GrTSL	3,7	8	2,2 - 6,1	17%	6% - 47%	8	
LB	3,2	7	1,0 - 9,8	30%	15% - 74%	8	1 x NOEC ≥ 10 mg/L
TLBL	3,8	7	2,2 - 6,4	31%	16% - 65%	8	1 x NOEC ≥ 10 mg/L

[1]) NOEC values ≥ 10 mg/L and < 1 mg/L not included

[2]) All maximum MDDs are obtained in only one and the same data set (L12)

Table 40: Overview of the NOEC for IP for the valid data sets as compared to those for all data sets in *M. spicatum* at DAT 14

Valid Data Sets

IP							
variable	NOEC [mg/L]			MDD			remarks
	geom mean [1]	n	95% PI	mean	range	n	
YFW	0,020	5	0,006 - 0,070	18%	13% - 22%	6	1 x NOEC < 0,01 mg/L
GrFW	0,025	5	0,004 - 0,167	14%	9% - 22%	6	1 x NOEC < 0,01 mg/L
YDW	0,015	3	0,004 - 0,055	16%	9% - 19%	4	1 x NOEC < 0,01 mg/L
GrDW	0,024	4	0,008 - 0,075	16%	8% - 26%	5	1 x NOEC < 0,01 mg/L
YTSL	0,020	5	0,003 - 0,151	20%	10% - 36%	6	1 x NOEC < 0,01 mg/L
GrTSL	0,032	5	0,006 - 0,157	14%	7% - 25%	6	1 x NOEC < 0,01 mg/L
LB	0,076	4	0,009 - 0,655	29%	18% - 64%	5	1 x NOEC < 0,01 mg/L
TLBL	0,032	4	0,001 - 0,782	26%	11% - 63%	5	1 x NOEC < 0,01 mg/L

All Data Sets

IP							
Variable	NOEC [mg/L]			MDD			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
YFW	0,022	6	0,007 - 0,070	18%	13% - 22%	7	1 x NOEC < 0,01 mg/L
GrFW	0,026	6	0,005 - 0,144	14%	9% - 22%	7	1 x NOEC < 0,01 mg/L
YDW	0,018	4	0,005 - 0,067	20%	9% - 36%	5	1 x NOEC < 0,01 mg/L
GrDW	0,025	5	0,009 - 0,070	18%	8% - 31%	6	1 x NOEC < 0,01 mg/L
YTSL	0,018	6	0,003 - 0,118	19%	10% - 36%	7	1 x NOEC < 0,01 mg/L
GrTSL	0,032	5	0,007 - 0,157	13%	7% - 25%	7	2 x NOEC < 0,01 mg/L
LB	0,064	5	0,008 - 0,478	26%	12% - 64%	6	1 x NOEC < 0,01 mg/L
TLBL	0,032	5	0,002 - 0,510	25%	11% - 63%	6	1 x NOEC < 0,01 mg/L

[1]) NOEC values $\geq$ 1 mg/L and < 0,01 mg/L not included

Table 41: Overview of the NOEC for TF for the valid data sets as compared to those for all data sets in *M. spicatum* at DAT 14

Valid Data Sets

TF							
variable	NOEC [mg/L]			MDD			remarks
	geom mean [1]	n	95% PI	mean	range	n	
YFW	0,101	3	0,011 - 0,963	25%	21% - 28%	3	
GrFW	0,101	3	0,011 - 0,963	18%	13,5% - 23%	3	
YDW	0,047	3	0,001 - 1,484	23%	11% - 31%	3	
GrDW	0,047	3	0,001 - 1,484	16%	7% - 25%	3	
YTSL	0,032	2	0,032 - 0,032	17%	14% - 21%	3	1 x NOEC < 0,01 mg/L
GrTSL	0,047	3	0,013 - 0,170	12%	9% - 16%	3	
LB	0,032	2	0,032 - 0,032	28%	23% - 33%	2	
TLBL	0,057	2	0,012 - 0,274	26%	25% - 27%	2	

All Data Sets

TF							
Variable	NOEC [mg/L]			MDD [2]			remarks
	geom. mean [1]	n	95% PI	mean	range	n	
YFW	0,101	3	0,011 - 0,963	44%	20% - 100%	4	1 x NOEC ≥ 1 mg/L
GrFW	0,101	3	0,011 - 0,963	37%	13,5% - 94%	4	1 x NOEC ≥ 1 mg/L
YDW	0,047	3	0,001 - 1,484	34%	11% - 68%	4	1 x NOEC ≥ 1 mg/L
GrDW	0,047	3	0,001 - 1,484	20	7% - 56%	4	1 x NOEC ≥ 1 mg/L
YTSL	0,047	3	0,013 - 0,170	19%	14% - 23%	4	1 x NOEC < 0,01 mg/L
GrTSL	0,057	4	0,016 - 0,205	16%	9% - 26%	4	
LB	0,047	3	0,013 - 0,170	33%	23% - 43%	3	
TLBL	0,032	3	0,003 - 0,303	32%	25% - 45%	3	

[1]) NOEC values ≥ 1 mg/L and < 0,01 mg/L not included

[2]) All maximum MDDs are obtained in only one and the same data set (L12)

5.3. Summary of the results for *M. spicatum*

M. spicatum exhibited relatively low variability of the control data within the laboratories (CV_r%: 10,2% to 24,2%), and higher variability of the control data between the laboratories (CV_R: 29,7% - 57,4%) (based on all data sets). Probably the same difficulties with weight measurements of total plants including roots occurred as for *M. aquaticum*; however, due to the higher absolute values of weights, measurement inaccuracies are of lower consequences for variability than in *M. aquaticum*.
Only one of the control data sets proved to be a statistical outlier. Further data sets could not be considered for analysis for other reasons (e.g. negative values of yield and growth rate; contamination, questionable concentrations etc.). Growth promotion was observed in some tests with IP und TF which prevented the EC50 computation (no hormesis available in the model used in this report). Hence, for the calculation of the overall means of EC50, NOEC and MDD, not in all cases the minimum number of values was in accordance to the requirements of a ring-test statistics.
Lateral branches were produced regularly and accounting for approximately 50% of final total shoot length.
Also for *M. spicatum,* a second statistical evaluation procedure was performed using the same validity criteria as for *M. aquaticum* (minimum growth rate of the fresh weight 0,07 d^{-1} and a maximum tolerable coefficient of variation in the yield of fresh weight of 35%). Eighteen out of 21 data sets (86%) fulfilled these validity criteria. Exclusion of three invalid data sets did not markedly affect the resulting variabilities, control means and toxic metrics.
The mean growth rate for fresh weight was 0,093 d^{-1} (all data, corresponding to a doubling time of 7,5 days), respectively 0,103 d^{-1} (valid data, corresponding to a doubling time of 6,7 days). The mean growth rate for total shoot length was 0,107 d^{-1} and 0,106 d^{-1} , both according to a doubling time of 6,5 days. Hence, doubling of fresh weight and total shoot length within the exposure time of 14 days was fulfilled.
Based on all data, mean MDDs between 13% (GrTSL) and 44% (YFW) were obtained, while based on valid data, mean MDDs between 14% (GrTSL and GrFW) and 33% (TLBL) were obtained.
Due to the limited number of data, the following statements about the sensitivity of *M. spicatum* should be considered carefully (only means based on at least two data sets are reported). For 3,5-DCP, the mean EC50 ranged between 4,2 and 6,1 mg/L (all data, n=4-7) and between 4,7 and 6,3 mg/L (valid data, n=4-6) with only small differences between the variables. IP and TF showed similar toxicity on a markedly lower concentration level. For IP, mean EC50 values ranged between 0,05 and 0,34 mg/L (all data, n=5-7), respectively between 0,05 and 0,32 mg/L (valid data, n=4-6). A uniform order in sensitivity of weight and length parameters was not obtained, but growth rates were less sensitive than the corresponding yields. For TF, mean EC50 values ranged between 0,19 and 0,33 mg/L (all data, n=2-4), respectively between 0,16 to 0,32 mg/L (valid data, n= 2-4), i.e. only small differences between the variables were obtained. Mean NOECs for *M. spicatum* were as follows: 3,5-DCP: 3,0 - 3,8 mg/L (all data), 2,9 - 3,9 mg/L (valid data); IP: 0,018 - 0,064 mg/L (all data), 0,015 - 0,076 mg/L (valid data); TF: 0,032 – 0,101 (both for all data and valid data).

6. Overall Comparison of the Different Test Species

6.1. General

Before going into detail with comparisons of results obtained for *M. aquaticum* and *M. spicatum,* main differences between the two test species will be emphasized. These differences are possibly relevant for the interpretation of differences in the results (all details are summarized Table 47 at the end of this chapter):

- The rooting period preceding the exposure period is 3 days for *M. aquaticum* and 7 days for *M. spicatum*.
- Exposure time is 7 days for *M. aquaticum* and 14 days for *M. spicatum*.
- Total test duration is 10 days for *M. aquaticum* and 21 days for *M. spicatum*.
- Regarding the overall means of all data (except outliers), *M. aquaticum* showed slightly lower growth rates (TSL, FW) than *M. spicatum*. The corresponding doubling times obtained for fresh weight were 8,9 days (*M. aquaticum*) and 7,5 days (*M. spicatum*), for shoot length the doubling times were 7,5 days (*M. aquaticum*) and 6,5 days (*M. spicatum*)
- On average, in the ring test the exposure time of 7 days for *M. aquaticum* was not sufficient for doubling fresh weight and hardly sufficient for doubling shoot length.
- Lateral Branches were produced only infrequently in *M. aquaticum*. In contrast, in *M. spicatum* lateral branches were produced regularly and the total length of lateral branches contributed to about 50% of total shoot length.

First of all, the two species will be generally characterized by their reproducibility and repeatability coefficients of variation as well as the resulting minimum detectable differences (MDD) based on the results of the controls (Section 6.2). Subsequently the effects of the test substances on the biological variables will be evaluated by means of the EC50 obtained for the various variables (Section 6.3). Detailed information about number of data sets and exact values for a certain variables can be obtained from Table 42 to Table 46. Figure 45 to Figure 49 provide a visual impression of agreement or differences between results of the different test species. *M. aquaticum* will be symbolized in blue, *M. spicatum* in green color.

Finally, there will be an overall characterization of the test species (Section 6.4, Table 47).

The statistical comparisons of the EC50 and the data given in the tables and figures are based on the results of the valid data for *M. aquaticum* and *M. spicatum* (applied validity criteria were a minimum growth rate of fresh weight of 0,07 d^{-1} and a maximum coefficient of variation for yield fresh weight of 35%). However, to investigate, whether the comparison would lead to different results if it were based on all data (i.e. including invalid data sets, except outliers), the t-tests were also performed for the results obtained from all data. The resulting two sided probabilities were slightly different, but the conclusions (significant / not significant) were

exactly the same than those obtained for the valid data base. Therefore in the following only the results of the valid data sets are shown.

6.2. Repeatability, Reproducibility and MDD

Table 42 and Table 43 as well as Figure 45 and Figure 46 show the results of a comparison of the reproducibility and reproducibility coefficients of variation as well as of the mean minimum detectable differences (MDD) of the two test species. The following statements are based on 9 up to 18 control data sets, each.
Since in *M. aquaticum*, there was only marginal lateral-branches production, these variables showed extremely high coefficients of variation. Therefore, these variables were not appropriate for further evaluation. If lateral branches were formed regularly (*M. spicatum*), both the reproducibility and repeatability for this variable was only slightly higher than that of other variables.
Generally, reproducibility was worse than repeatability. This is probably due to different sources of plants, differences in pre-cultures, handling etc. and to a certain extent is inevitable. Nevertheless, *M. aquaticum* showed the highest inter-laboratory variability, indicating the need for further standardization and experience.

Regarding the intra-laboratory coefficients of variation, both test species showed quite acceptable results with values up to 20%, except in the dry weight variables in *M. aquaticum* which showed higher variability (33% for the growth rate of dry weight (GrDW) and 40% for the yield of dry weight (YDW)). This is probably due to a methodological difficulty of the AMRAP-tests in general: To obtain the weights at test end, plants have to be completely removed from the sediment and any sediment particles need to be rinsed off with water before weighting. Thereby roots could be lost and/or sediment particles could remain on the plant. Both effects probably increased the variability in the resulting weights. In view of the very low absolute dry-weight values in *M. aquaticum* (mean YDW: 14 mg), dry weights would be especially influenced by this effect. In the ISO-test with *M. aquaticum* this effect was also observed and led to exclusion of total plant dry weight parameters from further evaluation (Ute Feiler, personal communication).

Generally both test species allowed statistical testing at minimal detectable differences of up to about 30%, except for the dry weight variables in *M. aquaticum* (Table 43 and Figure 46). Therefore, from a statistical point of view both test species are appropriate for assessment of toxic metrics (ECx, NOEC), and in so doing, fresh weight and shoot length parameters should be preferred in contrast to total plant dry weight parameters (i.e. roots and shoots). Generally, growth rates exhibited lower variability and better statistical power than yields.

Table 42: Overview of the repeatability and reproducibility coefficients of variation found in the considered variables for the two test species based on valid tests.

	Repeatability CVr%		Reproducibility CVR%	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*	AMRAP *M. aquaticum*	AMRAP *M. spicatum*
	n = 9-13	n = 15-18	n = 9-13	n = 15-18
GrFW	14,1	9,9	35,2	21,2
GrDW	32,9	11,3	42,6	27,4
GrTSL	15,9	10,5	35,4	30,6
YFW	21,0	18,8	54,2	46,4
YDW	41,3	18,8	62,4	47,8
YTSL	20,0	17,1	48,5	52,3
LB	71,3	23,9	111,1	57,9
TLBL	n.d.	23,7	n.d.	52,8

Table 43: MDD% as calculated from the CVr% given Table 42 and the given test design (6 / 3 replicates, 5 treatments); n.d. = not determined

	Calculated MDD%	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*
GrFW	18,0	12,7
GrDW	42,1	14,5
GrTSL	20,3	13,4
YFW	26,9	24,1
YDW	52,9	24,1
YTSL	25,6	21,9
LB	n,d,	30,6
TLBL	n.d.	30,3

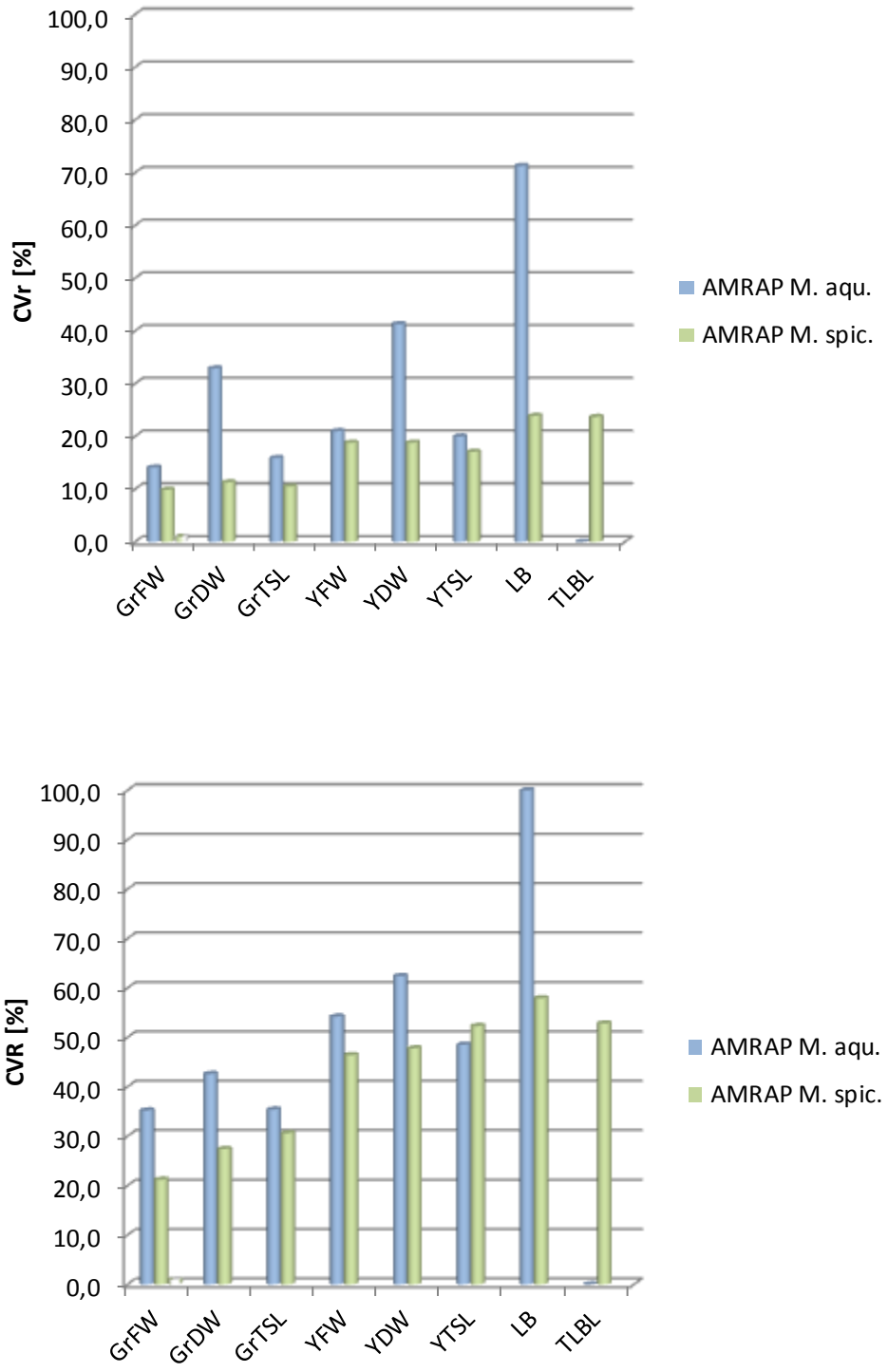


Figure 45: Repeatability (top) and reproducibility coefficients of variation (bottom) of all considered variables for *M. aquaticum* and *M. spicatum*. Zero = no data available. Data base: valid data.

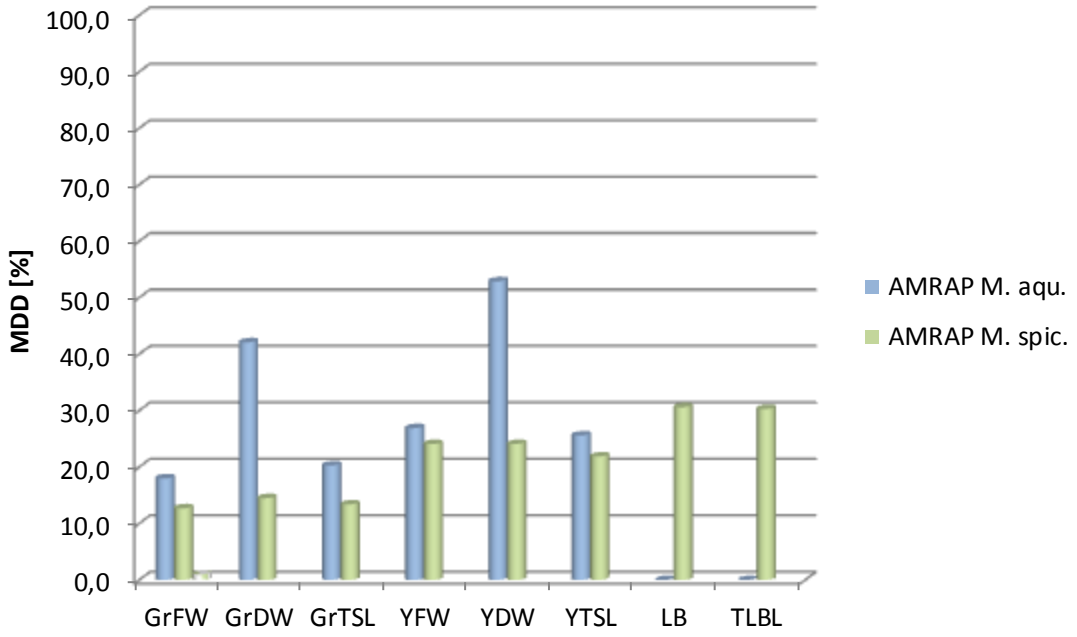


Figure 46: Minimum detectable differences (MDD%), calculated from the CVr% given in Table 42 (i.e. derived from valid data) and for the given test design (6 / 3 replicates, 5 treatments). Zero = no data available.

6.3. Sensitivity of Test Species

6.3.1. General

A statistical comparison of toxic metrics was performed in order to determine whether there is a significant difference in sensitivity between *M. aquaticum* and *M. spicatum* in the AMRAP-test. Results are shown in Table 44 and in Figure 47 to Figure 49. Mean EC50s for each species were compared in two-sample t-tests. Thereby two-sided probabilities were computed since *a priori* it cannot be assumed whether an EC50 was higher or lower depending on test species. It should be stated upfront that statistical testing of the obtained EC50 does not provide uniform trends in all cases, which may be due to low numbers of data sets. Nevertheless, the sum of EC50s available for the considered variables provide a weight of evidence for distinct trends in sensitivity observed for the two test species, even if the result of a single two-sample comparison is not in line with the results of the comparisons for other variables.
It should be kept in mind that the analytical recovery rates at the test end of each test differed considerably depending on the test substance and test duration (see Chapter 3). Therefore, the following statements about sensitivity are only valid for the test substances under consideration and should not be extrapolated to other substances.

Table 44: Overall mean EC50 for different test species. Data base: valid data.

	3,5-DCP [mg/L]	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*
	n = 2-5	n = 4-6
GrFW	5,2	5,3
GrDW	*(3,7) n=1*	4,8
GrTSL	6,1	6,3
YFW	4,6	4,7
YDW	-	5,0
YTSL	5,7	5,3
LB	-	5,6
TLBL	-	5,3
R	semi-quantitative assessment; effects obvious	semi-quantitative assessment; effects obvious
TRL		

	IP [µg/L]	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*
	n = 2-3	n = 4-6
GrFW	1180	164
GrDW	(130) n=1	74
GrTSL	510	315
YFW	860	83
YDW	(120) n=1	52
YTSL	370	140
LB	-	263
TLBL	-	97
R	semi-quantitative assessment; effects obvious	semi-quantitative assessment; effects obvious
TRL		

Table 43 (continued)

	TF [µg/L]	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*
	n = 3	n = 2 - 3
GrFW	-	*(840) n=1*
GrDW	-	-
GrTSL	(750) n=1	314
YFW	-	319
YDW	-	-
YTSL	820	165
LB	-	316
TLBL	-	163
R	semi-quantitative assessment; effects obvious	semi-quantitative assessment; effects obvious
TRL		

6.3.2. Sensitivity to 3,5-dichlorophenol

The EC50 for 3,5-DCP was about 5 to 6 mg/L for most variables in both species (Table 44; Figure 47), indicating a similar sensitivity of the species. As can be derived from Table 45, significant differences between the EC50 of all variables in both test species were not detected. Different analytical recovery rates of 3,5-DCP in the water phase at the end of the test were reported: 67% (*M. aquaticum*) and 47% (*M. spicatum*). Data for 3,5-DCP concentrations in sediment were only available for *M. aquaticum* (6%, one measurement), no analytical data for recovery rates in sediment were available for *M. spicatum*.

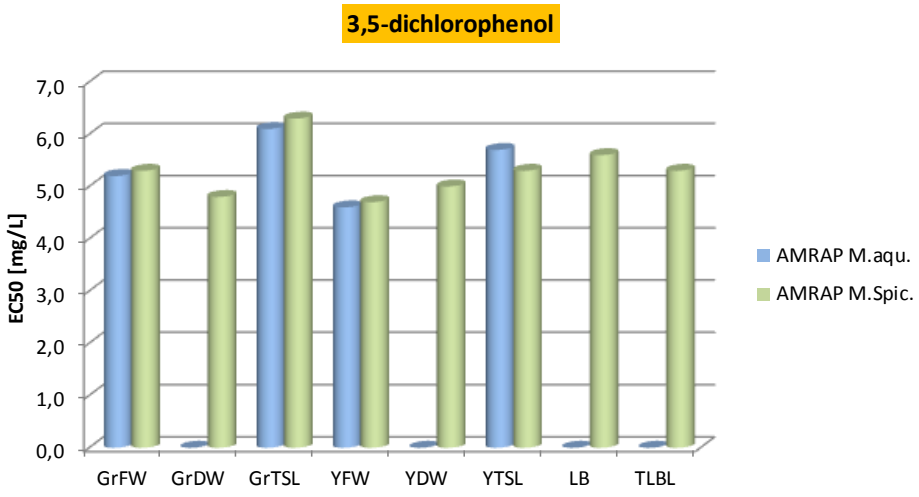


Figure 47: Overall mean EC50 for 3,5-DCP for different test species. Zero = no data available. Data base: valid data.

Table 45: Statistical comparison of the EC50 for 3,5-DCP between test species in the AMRAP-test. Results of two-sample t-tests for selected variables. Data base: valid data.

Tests between species	EC50 [µg/L]		t-test	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*	p (two-sided)	Sign. (α = 0,05)
GrFW	5,2	5,3	0,991	-
GrTSL	6,1	6,3	0,901	-
YFW	4,6	4,7	0,891	-
YTSL	5,7	5,3	0,703	-

6.3.3. Sensitivity to Isoproturone

Statistical testing of the EC50 for IP for *M. aquaticum* compared to *M. spicatum* was performed for four variables (Table 46). The EC50 for *M. spicatum* was lower than the EC50 for *M. aquaticum* for all variables and for three variables, these differences were found to be significant. This provides evidence that *M. spicatum* is more sensitive than *M. aquaticum* to IP. Similar analytical recovery rates of IP at the end of the test were reported: 78% (*M. aquaticum*) and 76% (*M. spicatum*).

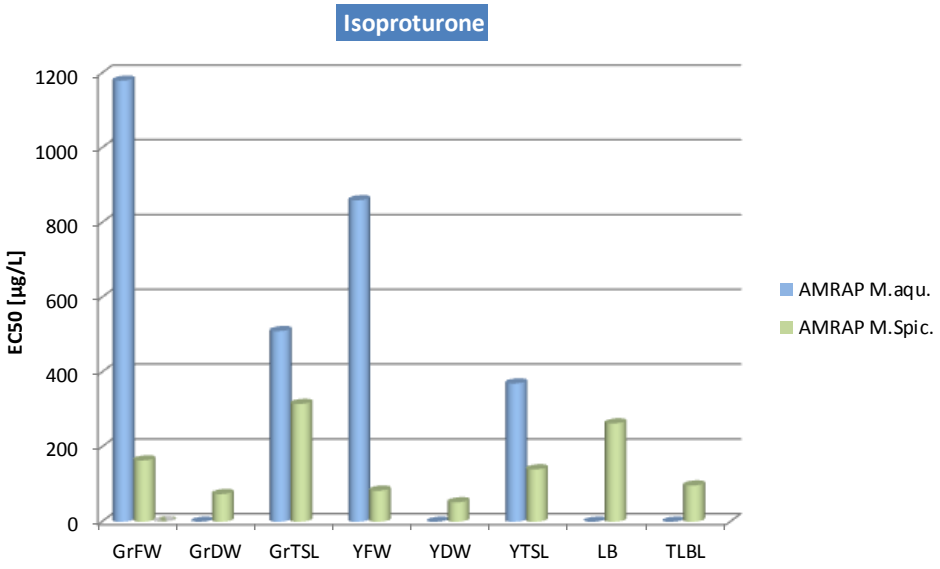


Figure 48: Overall mean EC50 for isoproturone for different test species. Zero = no data available. Data base: valid data.

Table 46: Statistical comparison of the EC50 for IP between species (*M. aquaticum* and *M. spicatum* in the AMRAP-test. Results of two-sample t-tests for selected variables. Data base: valid data.

Tests between species	EC50 [µg/L]		t-test	
	AMRAP *M. aquaticum*	AMRAP *M. spicatum*	p (two-sided)	Sign. (α = 0,05)
YTSL	370	140	0,017	+
YFW	860	83	< 0,001	+
GrTSL	510	315	0,311	-
GrFW	1180	164	< 0,001	+

6.3.4. Sensitivity to Trifluralin

The number of EC50s obtained for TF was low for several reasons (see Section 4.2.4). So, overall it can only be stated that the EC50s derived for *M. spicatum* are in the same order of magnitude than those for IP (Table 44, Figure 49). The only EC50 available for *M. aquaticum* (yield of total shoot length) is significantly higher than the corresponding EC50 for *M. spicatum* (820 µg/L, 165 µg/L; p = 0,029). Hence, *M. aquaticum* appears to be less sensitive than *M. spicatum* to TF. The analytical recovery rates for TF at the test end were much lower than those for the other test substances (water phase: 9,5% measured for TF in *M. aquaticum* test at DAT 7 and 1,8% in *M. spicatum* test at DAT 14; sediment: 8,2% in *M. aquaticum* test at DAT 7). Also with TF, few analytical data were available.

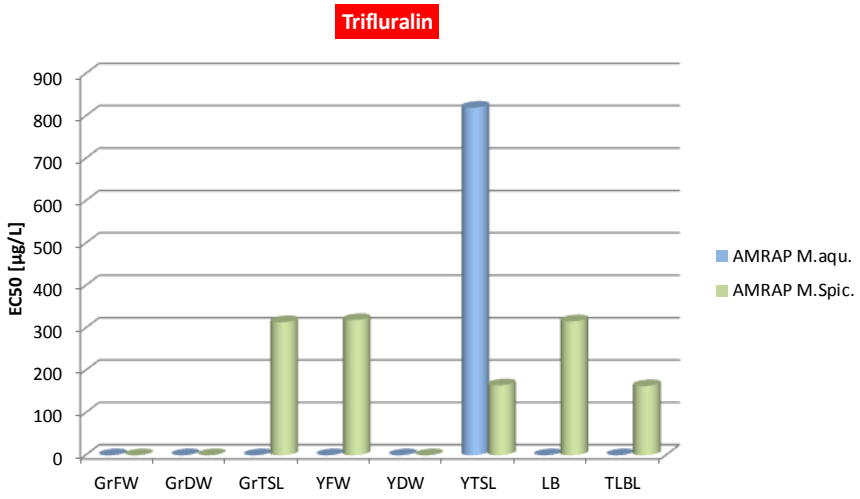


Figure 49: Overall mean EC50 for trifluralin using different test species. Data base: valid data. Zero = no data available.

Possible factors determining the species sensitivity differences in *Myriophyllum* are (1) different effective concentrations of test substance, (2) the general growth characteristic of test plants, (3) duration of exposure and (4) extent of exposure path way via roots.
(ad 1) The recovery rates measured in the water phase at the end of the tests with *M. aquaticum* and *M. spicatum*, respectively, were different for 3,5-DCP and TF and similar for IP.
(ad 2) Generally, the measured growth rates for *M. aquaticum* and *M. spicatum* were comparable or only slightly different, however, due to the different exposure times, only *M. spicatum* was able to double shoot length and weight during exposure time.
(ad 3) Exposure time and thus the time for uptake and metabolism of the toxic substance of *M. spicatum* in the AMRAP test was twice that of *M. aquaticum*.
(ad 4) Rooting phase was 7 days for *M. spicatum* compared to 3 days for *M. aquaticum*. Hence *M. spicatum* possibly had got more roots at test start.

To sum up the results, there is evidence that with the AMRAP-test method *M. aquaticum* and *M. spicatum* showed similar sensitivity to 3,5-DCP and different sensitivity to IP and TF. *M. spicatum* was more sensitive to IP and TF than *M. aquaticum*. In view of the few analytical data and the multiple factors possibly involved in the resulting toxicity, no clear explanation for the results obtained can be given so far.

6.4. Overall Characterization of the Test Species

Table 47 summarizes different aspects for characterization of the two test species. The main differences between the test species are (1) the duration of rooting phase and exposure time (2) the production of side shoots (infrequent / regular) and (3) the absolute increase in shoot length and weight of the test plants resulting from growth rates and exposure times.

Rooting phases were 3 days for *M. aquaticum* and 7 days for *M. spicatum*, followed by exposure times of 7 days for *M. aquaticum* and 14 days for *M. spicatum*. Hence total test duration was 10 days for *M. aquaticum* and 21 days for *M. spicatum*.
M. aquaticum showed only infrequent production of lateral branches, while *M. spicatum* regularly produced lateral branches, which in total contributed to 50% of the total shoot length.

Measurement variables for *Myriophyllum* were fresh weight, dry weight, total shoot length, root length and lateral branch length (if present). A general effect on root development was confirmed by the semi-quantitative assessment of root development for both species.

Depending on test item and test duration, the analytical recovery rates at test end were 2% to 73%.

Regarding the overall means of all control data, *M. aquaticum* showed slightly lower growth rates (TSL, FW) than *M. spicatum*. In view of the obtained growth rates, the prescribed exposure time of 7 days for *M. aquaticum* on average was not sufficient for doubling fresh weight and hardly sufficient for doubling shoot length. The obtained yields for TSL and FW were 4-5 fold higher for *M. spicatum* compared to *M. aquaticum*.

Preliminary validity criteria applied were a minimum growth rate fresh weight of 0,07 d^{-1} and / or a maximum coefficient of variation in yield of fresh weight > 35%. This resulted in 50% of valid tests for *M. aquaticum* and 86% of valid tests for *M. spicatum*.

M. aquaticum showed markedly higher variability than *M. spicatum*. When applying (preliminary) validity criteria, for both test species acceptable coefficients of variation and MDDs were obtained (i.e. up to 30%), except for dry weight parameters with *M. aquaticum*.

Regarding the toxicity metrics, 3,5-DCP evoked similar toxicity in both species. For IP, *M. aquaticum* was less sensitive than *M. spicatum* (by a factor 2 up to 10 depending on the variable). For TF the data base for comparison was poor, but toxicity appeared to be similar to IP (based on the nominal concentrations).

Table 47: Summary of characteristics of the different test species. Both results from all data and from valid data are shown.

Characteristics of the Test System		AMRAP *M. aquaticum*	**AMRAP *M. spicatum***
Data Set	Date of ring test	Jan - Oct 2011	Jan - Oct 2011
	Number of Tests performed	26	25
	Number of laboratories	10	10
	Test substances	3,5-DCP, IP, TF	3,5-DCP, IP, TF
Test Design and Characteristics	Replicates	6 / 3	6 / 3
	Treatments	5	5
	Replication	1 replicate = 3 plants	1 replicate = 3 plants
	rooting phase	3 d	7 d
	culturing	with sediment, submersed	with sediment, submersed
	Test conditions	20°C ± 2°C, 16:8, 120 - 160 µE m-1 s-1	20°C + 2°C, 16:8, 120 - 160 µE m-1 s-1
	Exposure time	7 d	14 d
	Total test duration	10 d	21 d
Variables	measured	FW, DW, TSL, LB, TLB	FW, DW, TSL, LB, TLBL
	qualitative assessment	Root development	Root development
	calculated	YFW, YDW, YTSL GrFW, GrDW,GrTSL	YFW, YDW, YTSL GrFW, GrDW,GrTSL
	evaluated	YFW, YDW, YTSL GrFW, GrDW,GrTSL	YFW, YDW, YTSL, GrFW, GrDW,GrTSL, LB, TLBL
Recovery Test Substances at Test End (water phase)		3,5-DCP 68% IP 73% TF 10%	3,5-DCP 47% IP 76% TF 2%
Validity	Criteria applied	GrFW ≥ 0,07 d-1 (doubling time ≤ 9,9 d) CV% YFW ≤ 35%	GrFW ≥ 0,07 d-1 (doubling time ≤ 9,9 d) CV% YFW ≤ 35%
	percentage of valid tests	50%	86%
Practicability	SL Factor ≥ 2	40%	91%
	CV% YFW ≤ 35%	58%	95%
	GrFW ≥ 0,07 d^{-1} = doubling time ≤ 9,9 d	58%	86%
	GrFW ≥ 0,09 d^{-1} = doubling time ≤ 7,7 d	42%	71%

Table 47 (continued)

Characteristics of the Test System		AMRAP *M. aquaticum*	AMRAP *M. spicatum*
Growth characteristics (all data / only valid data)	Lateral Branches	infrequent LB growth	regular LB growth TLBL ~ 50% of TSL
	Mean GrTSL	0,092 / 0,099 d-1	0,107 / 0,106 d-1
	Mean doubling time TSL	7,5 / 7,0 d	6,5 / 6,5 d
	Mean GrFW	0,078 / 0,114 d-1	0,093 / 0,103 d-1
	Mean doubling time FW	8,9 / 6,1 d	7,5 / 6,7 d
	Mean SL increase Factor	2,0 / 2,1	4,7 / 4,6
	Mean YTSL	5,0 / 5,3 cm	28,0 / 27,9 cm
	Mean YFW	211 / 293 mg	885,1 / 894,3 mg
	Mean YDW	12,5 / 14 mg	73,6 / 74,7 mg
Statistical characteristics and Variability (all data / only valid data)	$CV_{R\,\%}$	35,0% - 74,3% [1)] 35,2% - 62,4% [1)]	29,6% - 57,5% 21,2% - 57,9%
	CVr%	17,9% - 54,6% [1)] 14,1% - 41,3% [1)]	10,2% - 24,2% 9,9% - 23,9%
	MDD calculated based on CVr% and test design	22,9% - 69,9% 18,0% - 52,9%	31,1% - 30,8% 13,4% - 30,6%
Overall mean EC50 (range for all variables evaluated) (all data / only valid data)	3,5-DCP [mg/L]	3,8 - 5,0 (n=2-8) 4,6 - 6,1 (n=2-5)	4,7 – 6,1 (n=4-7) 4,7 - 6,3 (n=4-6)
	IP [µg/L]	60 - 770 (n=2-8) 370 - 860 (n=2-3)	53 - 338 (n=5-7) 52 - 315 (n=4-6)
	TF [µg/L]	600 - 670 (n=3-5) 820 (n=3)	193 - 840 (n=1-4) 163 - 319 (n=2-3)
Overall mean NOEC (range for all variables evaluated) (all data / only valid data)	3,5-DCP [mg/L]	2,9 - 5,0 (n=3-7) 2,4 - 5,6 (1-3)	3,0 – 3,8 (n=6-8) 2,9 - 3,9 (n=5-7)
	IP [µg/L]	60 - 147 (n=3-9) 47 - 100 (n=1-4)	18 - 64 (n=4-6) 15 - 76 (n=3-5)
	TF [µg/L]	42 - 100 (n=1-4) 32 - 100 (n=1-3)	32 - 101 (n=3-4) 32 - 101 (n=2-3)

[1)] without LB and TLLB

7. Summary of the AMRAP-Test and Recommendations

The analysis of the ring-test with *M. aquaticum* and *M. spicatum* revealed, that the toxicity test in principle is practicable. Several laboratories delivered data sets of acceptable quality. But the observed high variability, especially with *M. aquaticum*, highlights the need for further methodological standardization (e.g. plant quality, preculturing, duration of rooting phase, light and temperature conditions during exposure, weight measurement, comparability of representative plants and tests plants).

Since one objective of the ring test was to define validity criteria depending on the outcomes, in an initial evaluation all data sets (except outliers) were considered for evaluation. Based on the results preliminary validity criteria were set and applied for evaluation in a second evaluation: a minimum growth rate of fresh weight of 0,07 d^{-1} (corresponding to a doubling time of fresh weight of 9,9 days) and a maximum variability of yield fresh weight of 35%. Thereby it should be kept in mind that the proposed validity criteria was adjusted such to have at least 50% of the tests valid, regardless of the fact, that the applied minimum growth rate for fresh weight was not sufficient to enable fresh weight doubling within the exposure time for *M. aquaticum*. With respect to a further guideline development, the variable, type and value of validity criteria have to be discussed among experts.
With *M. aquaticum*, 50% of the tests performed were found to be valid, with *M. spicatum* 86% of the tests fulfilled the proposed validity criteria.

Regarding the overall means of all control data, *M. aquaticum* showed slightly lower growth rates than *M. spicatum* (GrTSL 0,092 d^{-1} compared to 0,107 d^{-1}; GrFW 0,078 d^{-1} compared to 0,093 d^{-1}). Regarding only valid data, again *M. aquaticum* showed a slightly lower GrTSL than *M. spicatum* (0,099 d^{-1} compared to 0,106 d^{-1}), but a slightly higher GrFW (0,114 d^{-1} compared to 0,103 d^{-1}).
In view of the growth rates obtained, for *M. aquaticum* the prescribed exposure time of 7 days on average was not sufficient for doubling fresh weight and hardly sufficient for doubling shoot length. In contrast, for *M. spicatum* growth rates and exposure time resulted in an increase of shoot length by a mean factor of 4,7 and an increase in fresh weight by a mean factor of 2,4. So the question arises as of whether the chosen time period for *M. aquaticum* generally is too short, or if the measured growth rates were unusual low due to bad plant performance or any methodological shortcomings.

The EC50 of 3,5-DCP were comparable in both *M. aquaticum* and *M. spicatum* (range EC50 *M. aquaticum*: 4,6 – 6,1 mg/L; *M. spicatum*: 4,7 – 6,3 mg/LM based on valid data). In contrast, *M. aquaticum* seems to be less sensitive towards IP and TF than *M. spicatum*: EC50 *M. aquaticum*, IP: 0,370 – 0,86 mg/L; *M. spicatum*, IP: 0,052 – 0,315 mg/L, means of at least two EC50 based on valid data; EC50 *M. aquaticum*, TF: 0,820 mg/L (only one variable available); *M. spicatum*, TF: 0,163 – 0,316 mg/L). This could be confirmed by statistical testing of the EC50 (Section 6.3). The same statement cannot be derived from the NOECs obtained for both species, since these usually were equal. Since the NOEC can only be one of the chosen test concentrations the accuracy of this toxicity measure is lower.

The measured variables will be compared by means of sensitivity, variability and capability to detect certain effects.
For 3,5-DCP, major differences in the sensitivity of the studied variables (i.e. weights or length) were not observed. For IP, EC50 values differed up to a factor of two. However, the data base was too small to support a systematic difference in sensitivity of weight and length variables. For TF, only small differences between the EC50 values for the different variables were obtained. Growth rates based on total shoot length or weight show trends towards lower variability and – assuming the plants stay in exponential growth - are time independent in contrast to e.g. the yield. Weight variables ensure capture of effects that are not manifest on shoot length, which is a function of cell elongation, not necessarily of biomass accumulation. Additionally, effects on roots biomass are captured by total plant weight. If directly quantitative assessment of effects on roots is desired, root length was shown to be an appropriate variable.

The following aspects concerning test design should be discussed during the process of future guideline development. From the statistical viewpoint the test design could be improved with respect to both the needs of a NOEC and of an EC-design. Because measurements have to be done with each single plant (high effort anyway) one would markedly increase the statistical power if the plants would be grown singly in smaller pots. A regression design and the computation of ECs would benefit from more concentrations (e.g. eight treatments), especially if non-linear regression models should be used. Concerning NOEC determination, in view of the possibly high variability, more replicates would enhance the statistical power.

References

Christensen (1984): Dose-response-functions in aquatic toxicity testing and the Weibull model. Water Res. 1984, 18, 213-221

Christensen & Nyholm: Ecotoxicological assays with algae: Weibull dose-response curves. Environ Sci Technol 18:713-718).

Finney (1978; Statistical Method in Biological Assay. 3rd Edition , Cambridge University Press, London)

Feiler et al (2012): Sediment contact test with *Myriophyllum aquaticum* (ISO/DIS 16191: first results of an international ring test. Poster presented at the 6th SETAC World Congress, Berlin, May 2012

ISO 5725-2 (1994) : Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method (lSO 5725-2:1994 including Technical Corrigendum 1:2002)

lSO 5725-5 (1998): Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method (lSO 5725-5:1998), Corrigenda to DIN ISO 5725-5:2002-11)

ISO/ DIS 16191 (2012). Water quality - Determination of the toxic effect of sediment on the growth behaviour of *Myriophyllum aquaticum*.

Maltby, L., Arnold, D., Arts, G., Davies, J., Heimbach, F., Pickl, C., Poulsen, V. (eds.), (2010). Aquatic Macrophyte Risk Assessment for Pesticides. Guidance from the AMRAP workshop in Wageningen (NL), 14-16 January 2008.

Myriophyllum spicatum toxicity test: results of a ring test using a sediment free test system (2011). Final Report, Federal Environment Agency, Germany, FKZ 36301294

Annex

Analytical results

Annex Table 1: Analytical data as provided by participants of the ring test

		Myriophyllum aquaticum						
		Culture medium						Sediment
3,5-DCP	**Nominal**	**DAT 0**			**DAT 7**			**DAT 7**
	[mg/L]	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**
		L13	**L14**	**L07**	**L13**	**L14**	**L07**	**L13**
	1	89,9	98,1	96	37,3	59,0	72	2,7
	1,8	90,6	98,4			62,1		
	3,2	91,2	98,4	84,4	45,8	64,8	62,5	7
	5,6	88,5	99,2			82,8		
	10	90,9	97,8	94	65,7	89,3	85	8,8
IP	**Nominal**	**DAT 0**			**DAT 7**			**DAT 7**
	[mg/L]	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**
		L13	**L05**	**L07**	**L13**	**L05**	**L07**	**L13**
	0,01	72,4	116	100	69,1	106	95	0
	0,032	87,6			53,9			10,1
	0,1	102,8	101	100	53,9	88	98	22,8
	0,32	83,5			62,1			33,1
	1	95,2	94	110	54,5	84	95	33,6
TF	**Nominal**	**DAT 0**			**DAT 7**			**DAT 7**
	[mg/L]	**[%]**			**[%]**			**[%]**
		L13			**L13**			**L13**
	0,01	87,5			7,6			10,1
	0,032	80,9			8,7			3,3
	0,1	84,2			7			3,6
	0,32	84,3			10,1			5,8
	1	115,1			14,2			18,1

Annex Table 1 (continued)

Myriophyllum spicatum									
3,5-DCP	**Nominal**	**DAT 0**				**DAT 14**			
	[mg/L]	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**		**[%]**	**[%]**
		L04	**L10**	**L14**	**L12**	**L04**	**L10**	**L14**	**L12**
	1	107,0	98	98,1	73,8	61,0	not measured	28,1	23,9
	1,8	94,0	96,7	98,4	128,3	64,0		24,3	13,8
	3,2	93,0	97,8	98,4	86,8	66,0		29,7	17,8
	5,6	96,0	97,3	99,2	112,2	68,0		55,2	58,7
	10	96,0	98,4	97,8	75,7	83,0		80,5	33,0
IP	**Nominal**	**DAT 0**				**DAT 14**			
	[mg/L]	**[%]**	**[%]**			**[%]**	**[%]**		
		L04	**L10**			**L04**	**L10**		
	0,01	99,0	94,5			77,0	70		
	0,032	94,0	94,8			73,0	90		
	0,1	86,0	94,9			73,0	73,9		
	0,32	93,0	94,8			73,0	77,7		
	1	98,0	93,9			80,0	74,6		
TF	**Nominal**	**DAT 0**				**DAT 14**			
	[mg/L]	**[%]**				**[%]**			
		L04				**L04**			
	0,01	121,0				0,0			
	0,032	238,0*)				0,0			
	0,1	68,0				0,0			
	0,32	58,0				3,0			
	1	39,0				6,0			

*) probably typing error; not included in the mean and standard deviation of

Environmental Conditions during testing

Annex Table 2: Details about the prevailing environmental conditions during testing as provided by the participating laboratories

Labora-tory	Factor	*M. aquaticum*		
		3,5-DCP	**IP**	**TF**
L02	**T°C**	DAT 7:19 – 20	19,6 - 22	DAT 0 + DAT 3: 21,7;DAT 5: 18,7 DAT 7: 20,2 - 21,4
	O_2 mg/L	DAT 0: 107%; marked decrease at 5,6 and 10 mg/L	DAT 0: 100%; decrease at 1 mg/L	DAT 0: 100%; strong increase
	pH	DAT 0: 7,6; minor changes	DAT 0: 7,8; minor changes	DAT 0: 7,5; strong increase
	Light	9143 -10927 lux; LD 16:8	9054 - 10530 lux LD 16:8	9893-11420 lx
L03	**T°C**	DAT 0: 17 - 18.5 DAT 7: 24 - 26	21,2 - 22,5	20,9 -21,3
	O_2 mg/L	DAT 0: 9,0; strong decrease at 5,6 and 10 mg/L	not measured at DAT 0:, remaining time points 100% - 70%, decrease at higher concentrations	DAT 0: 90%; throughout strong increase
	pH	DAT 0: 7,7; increase except at 5,6 and10 mg/L	DAT 0: 7,7; increase until DAT 7 up to 8,7 except at 10 mg/L	DAT 0: 7,6; throughout strong increase
	Light	132 - 139 µE; LD 16:8	132 -140 µE; LD 16:8	129 -140 µE; LD 16:8
L05	**T°C**		19,1 - 21,7	
	O_2 mg/L		DAT 0: ca. 100%; hardly changing; slight decrease at 1 mg/L	
	pH		DAT 0: 7,9; hardly changing	
	Light		7300 -10500 lux LD 16:8	
L07	**T°C**	DAT 0: 25.8; DAT 7: 19 - 20	19,4 - 20,6	DAT 0: 24,8; other DAT 22,2 - 23,7
	O_2 mg/L	DAT 0: 107%;strong decrease except at 5,6 and 10 mg/L	DAT 0: 105%; strong increase in control and at 0,01 mg/L; decrease at 1 mg/L	DAT 0: 108%; throughout increase
	pH	DAT 0: 7,6; hardly changing		DAT 0: 7,7; nearly constant or slight increase
	Light	97 - 115 µE	87 - 114 µE	?

Environmental Conditions during testing

Annex Table 2 (continued)

Laboratory	Factor	*M. aquaticum*		
		3,5-DCP	IP	TF
L08	T°C	constant 20	21 - 22	19 - 20
	O_2 mg/L	DAT 0: 98%,; slight decrease at 5,6 an 10 mg/L	DAT 0: 98%; strong decrease at 0,32 und 1 mg/L	DAT 0: 92%; throughout increase
	pH	DAT 0: 7,5; hardly changing	DAT 0: 7,9; minimum decrease	DAT 0: 7,4;throughout slight increase
	Light	8062 - 9164 lux; LD 16:8	8050 - 9010 lux; LD 16:8	8170 - 8900 lux; LD 16:8
L09	T°C	constant 19; measured? Perfect constant cabinet?	19 -19,5	
	O_2 mg/L	O2 DAT 0: 98%, slight decrease at 5,6 and 10 mg/L	DAT 0: 97%; hardly changing	
	pH	DAT 0: 8,3; hardly changing except at 1 mg/L, repl 2 DAT 7: 5,6	DAT 0; 7,8; hardly changing	
	Light	?	7517 -7989 lux; LD 16:8	
L11	T°C	19 - 20	DAT 0: not reported; remaining days: 19,1 - 21,0	
	O_2 mg/L	not measured at DAT 0; decrease at higher concentrations	DAT 0: not reported; remaining DAT: 12 - 9	
	pH	decrease at higher concentrations	DAT 0: not reported; remaining DAT: 7,5-8,1	
	Light	134 - 153 µE; LD 16:8	140 - 153 µE; LD 16:8	
L13	T°C	20 - 24	DAT 0: 22,8: until DAT 7: 19 - 20	18,8-21
	O_2 mg/L	O2 DAT 0: 100-120%; strong decrease at 5,6 and 10 mg/L	DAT 0: 100%; hardly changing	DAT 0: 98%; throughout increase
	pH	DAT 0: 7,8; throughout increase except at 10 mg/L	DAT 0: 7,7; increase up to 8,7 (control), less in treatments	DAT 0: 7,7; throughout increase
	Light	9900 -11500 lux; LD 16:8	9880 - 10520 lux; LD 16:8	?

Environmental Conditions during testing

Annex Table 2 (continued)

Laboratory	Factor	*M. aquaticum* 3,5-DCP	IP	TF
L14	T°C	23 - 24	DAT 0: 22,4; DAT 3 until DAT 7: 20,4 - 21,7	20,3 - 21,4
	O_2 mg/L	O2 DAT 0: 95%; strong decrease at 10 mg/L	DAT 0: 89%; slight increase except at higher concentrations	DAT 0: 90%; throughout increase
	pH	DAT 0: 7,5; throughout increase except at 10 mg/L	DAT 0: 7,9; slight increase except at higher concentrations	DAT 0: 7,9; throughout increase
	Light	95 - 141 µE	120 - 154 µE	122 - 158 µE
L15	T°C	DAT 0: 19,5; DAT 3: 22-25; DAT 5: 24; DAT 7: 19,6	19 - 22	DAT 0: 24,7, decreases until DAT 7 to 19,3 - 20,3
	O_2 mg/L	DAT 0: 70%; increase until DAT 7	DAT 0: 86%; slight decrease	O2 DAT 0: 80%; remaining at the same level
	pH	DAT 0: 7,7; remained nearly constant	DAT 0: 7,3; slight increase	DAT 0: 7,6; remained nearly constant
	Light	83 - 118 µE; LD 16:8	95 - 119 µE; LD 16:8	?
Total no of tests		**9**	**10**	**7**
		M. spicatum		
L01	T°C	DAT 0: 23,4, DAT 5-DAT 14 = 21-21,9		
	O_2 mg/L	DAT 0: 7,8; increase except at 10 mg/L		
	pH	DAT 0: 7,8; increase except at 10 mg/L		
	Light	124 -126 µE; LD 16:8		
L02	T°C	19,8 -21,9	20,6 - 21,9	
	O_2 mg/L	DAT 0: 96%; extreme increase except at 10 mg/L	DAT 0: 97 %; increase except at 1 mg/L	
	pH	DAT 0: 7,4; increase except at 10 mg/L	DAT 0: 7,5; slight increase except at 1 mg/L	
	Light	8923 – 10456 lux, LD 16:8	8865 – 9883 lux, LD 16:8	

Environmental Conditions during testing

Annex Table 2 (continued)

Laboratory	Factor	*M. spicatum* 3,5-DCP	*M. spicatum* IP	*M. spicatum* TF
L03	T°C	20,5 - 22,8	20,7 - 23	20,8 -22
	O_2 mg/L	DAT 0: 97%; strong increase except at 5,6 and 10 mg/L	DAT 0: 85%; strong increase except at 1 mg/L	DAT 0: 85%; throughout strong increase
	pH	DAT 0: 7,6; throughout strong increase	DAT 0: 7,9; slight increase except at 1mg/L	pH DAT 0: 6,7; throughout strong increase
	Light	129 -140 µE; LD 16:8	130 -145 µE; LD 16:8	130 -137 µE; LD 16:8
L04	T°C	20,2 - 21,1	20 - 20,3	19,9 - 21,7
	O_2 mg/L	DAT 0: 94%; increase except at 10 mg/L	DAT 0: 90 %; increase except at 1 mg/L	DAT 0: 97%; throughout slight increase
	pH	DAT 0: 8, Increase except at 10 mg/L	DAT 0: 7, 9; slight increase except at 1 mg/L	DAT 0: 7,9; throughout increase up to 9,6
	Light	8380-11080 lux LD 16:8	7980-11980 lux LD 16:8	9210-11140 lux LD 16:8
L06	T°C	DAT 0: 19,6; DAT 5-DAT 14: 23,8 - 25,3	DAT 0: 20,4; increase up to 24,3 until DAT 14	DAT 0: 21,2; increase up to 23,7 and 24,5 until DAT 9 and DAT 14, respectively
	O_2 mg/L	DAT 0: 98%; throughout extreme increase except at 10 mg/L	DAT 0: 98 %; throughout increase except at 1 mg/L	O2 DAT 0: 100 %; throughout marked increase
	pH	DAT 0: 7,7; throughout increase except at 10 mg/L	DAT 0: 7,7; throughout increase except at 1 mg/L	DAT 0: 8,3; throughout marked increase
	Light	109 -148µE	109 - 136 µE	110 - 143 µE
L08	T°C	20 - 21	21	20 - 21
	O_2 mg/L	DAT 0: 95; decrease especially strong at 5,6 and 10 mg/L; microbial contamination!?	DAT 0: 100%; decrease especially strong at 0,32 + 1 mg/L; microbial contamination!?	DAT 0: 100-110%; remained on the same level
	pH			DAT 0: 7,6; increase up to 8,9
	Light	8010 - 8320 lux; LD 16:8	8530 - 8800 lux; LD 16:8	8210 - 8800 lux; LD 16:8
L10	T°C	21 - 22,8	19 - 22,9	
	O_2 mg/L	DAT 0: 61%; throughout moderate increase	DAT 0: 75%; throughout increase except at 0,32 and 1 mg/L	
	pH	not reported	DAT 0: 8,0; throughout increase except at 0,32 and 1 mg/L	

	Light	not reported	119 - 163 µE; LD 16:8	

Environmental Conditions during testing

Annex Table 2 (continued)

Labora-tory	Factor	*M. spicatum* 3,5-DCP	*M. spicatum* IP	*M. spicatum* TF
L11	**T°C**	19,1 - 21	20 - 21,6	
	O_2 mg/L	DAT 0: 110%; increase except at 10 mg/L	DAT 0: 95-120%; increase except at 1 mg/L	
	pH	DAT 0: 8,6; increase except at 10 mg/L	DAT 0: 7,8 - 9; remmained nearly constant	
	Light	123 - 145 µE; LD 16:8	123 - 145 µE; LD 16:8	
L12	**T°C**	22,5 - 23	21 - 22,5	DAT 0: 20; DAT 9 and DAT 14: 21,5 - 22
	O_2 mg/L	DAT 0: 90%; throughout increase except at 10 mg/L	DAT 0: 93 %; throughout increase except at 1 mg/L	DAT 0: 99 %;throughout increase
	pH	DAT 0: 7,8; throughout increase except at 10 mg/L	DAT 0: 7, 5; throughout slight increase except at 1 mg/L	DAT 0: 8; throughout slight increase
	Light	8620 - 10500 lux; LD 16:8	8500 - 12500 lux; LD 16:8	?
L14	**T°C**	23,3 - 26,3	DAT 0: 22,5 - 23,2; afterwards 20,3 - 21,7	19,2 - 21,7
	O_2 mg/L	DAT 0: 93%; throughout moderate increase	DAT 0: 85-95%: throughout moderate increase except at 0,32 and 1 mg/L	DAT 0: 90%; throughout increase
	pH	DAT 0: 7,6; throughout moderate increase	DAT 0: 7,8-8,4; throughout moderate increase except at 0,32 and 1 mg/L	DAT 0: 8; throughout increase
	Light	100 - 193 µE	114 - 157 µE	122 - 154 µE
Total no of tests		**10**	**9**	**6**

Ring-test protocol

Standardized method for investigating test substance impact on rooted aquatic macrophytes

Background, general information

Standard test guidelines are available for aquatic plants, for alga species (OECD 201) and for *Lemna* sp. (OECD 221) as a representative of higher aquatic plants. These methods can be used to generate data to address the risk of substances (in particular substances with herbicidal activity) to aquatic non-target plant species. However, in some cases, these studies may not be sufficient and information on additional macrophytes may be required. This may be the case, when contamination via sediment is a relevant exposure pathway requiring testing of a sediment rooting macrophyte or if there is indication that the standard species (alga and *Lemna*) may not be considered representative for other potentially more sensitive aquatic plant species (Maltby et al., 2010).
Based on current understanding and experience , *Myriophyllum* sp. were considered suitable additional macrophyte species for the above mentioned purposes.

As in other guidelines as well (see OECD 201, 202, or 203), there are options for different test species. In this case, Myriophyllum aquaticum and M. spicatum have been found to be suitable standard test organisms (they have been found to be sufficiently easy to work with and can provide reproducible results with reasonable effort.

Participants of the ring test should preferably test both species. It may be difficult to obtain M. spicatum of a suitable quality during winter time. Therefore it is recommended that labs start testing M. aquaticum first. Results should be provided by end of March 2011.
M. spicatum results are expected to be handed in not later than June 2011.

Three substances are used in this ring test:
- 3,5-dichlorophenol (3,5 DCP)
- isoproturon
- trifluralin

Test batches of 3,5-DCP should have at least 97% purity (may be obtained for example from Sigma-Aldrich); isoproturon and trifluralin will be provided to the participating labs.

Substances should be tested in the order shown above in case participating labs cannot perform tests with all compounds. Participating labs should at least test one of those substances (3,5 DCP) with both species or all test substances with one of the species.

Principle of the test

The objective of the test is to assess substance-related effects on the vegetative growth of the genus *Myriophyllum* in standardised media (water, sediment and nutrients) containing different concentrations of the test substance over certain test periods. For this purpose, individual shoot apices of healthy, non flowering plants are potted in artificial standard sediment fertilized to ensure adequate plant growth and are maintained in a nutrient formulated water. After an

establishment period, the plants are exposed to a series of test concentrations added to the water column. Alternatively, it is also possible to simulate exposure via the sediment by spiking the artificial sediment with in the required concentration range. The growth of the plants is evaluated for a period sufficient to allow a robust assessment of growth. At the end of the test, the plants are harvested and their biomass, length and other relevant observations are recorded.
To quantify substance-related effects, growth in the test solutions is compared to that of the controls, and the concentration causing a specified x% inhibition of growth (e.g. 50 %) is determined and expressed as the ECx (e.g. EC_{50}).

Plant material

Potential suppliers of Myriophyllum sp. are provided in the Appendix. The source of the plants used in this ring test needs to be documented. Species identification must be verified (in North America there is evidence of hybridization between *M. spicatum* and related species, (Moody, M.L. & Les, D.H., 2002)).

The plants should be visibly free of any other species (particularly snails or filamentous algae, in some regions also eggs or larvae from the small moth *Paraponyx stratiotata* or larve or adults from the curculionidae *Eubrychius velutus* can be a problem; some level of epiphytes - such as diatoms, but no filamentous algae - may often not be avoidable and will generally not be a problem). Only visibly healthy plants, without flowering shoots, should be used for the study; a photographic documentation of the plants used for the ring test should be provided .

If the plants are kept within the laboratory before the test as a maintenance culture, temperature, light and nutrient conditions should be at a low level, i.e. nutrient concentrations reflecting oligotrophic to mesotrophic systems. For this, often standard tap water may be useful. If culturing/maintenance conditions differ significantly from lab conditions (i.e. if plants are taken from outdoor systems at a time when temperature and day length differ significantly from those in the lab or when plants have just arrived from the supplier) then, in order to support good growth, plants should be cultured/acclimatized under conditions similar to those in the test for an adequate period before the study. The culture conditions need to be documented (temp, day length, light conditions, media details).

Test vessels

The study using *Myriophyllum* as test species is conducted in 2-L glass beakers (tall form) (approximately 24 cm high and 11 cm in diameter). Other vessels may be suitable, but they should guarantee a suitable depth of water to allow unlimited growth and keep the plants submersed throughout the study. Small plant pots (approx. 9 cm diameter and 8 cm high and 500 mL volume) are used as containers for potting the plants into the sediment.

The sediment surface coverage should be > 70 % of the test vessel surface; the minimum overlaying water depth should be 12 cm.

Sediment

The following formulated sediment, based on the artificial sediment used in OECD Guideline 219, is used in this test; the sediment is prepared per the guideline except for the additional nutrients as described below:

(a) 4-5 % peat (dry weight, according to 2 +/- 0.5% organic carbon) as close to pH 5.5 to 6.0 as possible; it is important to use peat in powder form, finely ground (particle size < 1 mm) and only air dried.

(b) 20 % (dry weight) kaolin clay (kaolinite content preferably above 30 %).

(c) 75-76 % (dry weight) quartz sand (fine sand should predominate with more than 50 per cent of the particles between 50 and 200 µm).

(d) An aqueous nutrient medium is added such that the final sediment batch contains 200 mg/kg sediment (dry wgt) of both ammonium chloride and sodium phosphate and the moisture of the final mixture is in a range of 30% - 50%.

(e) Calcium carbonate of chemically pure quality ($CaCO_3$) is added to adjust the pH of the final mixture of the sediment to 7.0 +/- 0.5.

The source of peat, kaolin clay and sand should be known and documented. If the origin is unknown or gives some level of concern, then the respective components should be checked for the absence of chemical contamination (e.g. heavy metals, organochlorine compounds, organophosphorous compounds).

The dry constituents of the sediment should be mixed homogenously; afterwards the aqueous nutrient solution should be mixed thoroughly into this sediment. The moist sediment should be prepared at least two days before use to allow proper soaking of the peat (to prevent hydrophobic peat particles floating to the surface when the sediment is overlaid with media). The moist sediment should be stored in the dark.

For the test, the sediment is filled into a suitable size container, such as standard planting pots of a diameter which just fit into the glass vessels (the sediment should cover a minimum of 70% of the vessel bottom surface). In cases where the container has holes at the bottom, a piece of filter paper in the bottom of the vessel will help to keep the sediment within the container. The pots are filled with the nutrient containing sediment. This is covered with a very thin layer (≤ 5 mm) of an inert material such as quartz sand (or crushed corral) to assist in keeping the sediment in place.

Water medium

The test is conducted using Smart & Barko medium, which has been shown to provide good plant growth (but not containing phosphate or nitrogen and thus avoiding unwanted alga growth) and which is easy to produce (see Appendix 1 for its composition). The pH at test initiation should be between 7.5 and 8.0.

Experimental design

Three replicate test vessels are prepared for each treatment group (5 test concentrations arranged in a geometric series) and six replicate test vessels are prepared for the control.

The following concentrations will be tested:

3,5-DCP: control, 1.0, 1.8, 3.2, 5.6, 10 mg/L

IPU: control, 0.01; 0.032; 0.10; 0.32; 1.0 mg/L

Trifluralin: control, 0.01; 0.032; 0.10; 0.32; 1.0 mg/L

Each test vessel contains one plant pot with three shoots. The individual test vessels should be impartially (respectively randomly) assigned to the different treatment groups.

A randomized design for the location of the test vessels in the growth chamber is required to minimize the influence of spatial differences in light intensity or temperature. A repositioning of the vessels in an impartial way needs also be taken into account after observations are made.

Test procedure

Due to the different inherent growth rates of *M. spicatum* and *M. aquaticum*, the test procedure varies according to the species selected for testing.

Healthy shoot tips from the culture plants are clipped off at a length of 6 cm (+/- 1 cm). As the shoot tips may differ significantly in weight, all clippings should be weighed individually and, for the test, only plants within a 30% weight range should be utilized. Five shoot tips are planted into each pot containing the sediment such that the lower 3 cm, covering two nodes, are beneath the sediment surface. Shoots are then maintained for 3 days for *M. aquaticum* or 7 days for *M. spicatum* in a nutrient-poor water to induce root development (pots may be maintained either in the standard glass beakers or easier in aquaria during the adaptation phase). Thereafter, two of the five plants are removed to leave three uniform (size, appearance) individuals.

Plants from five additional pots are harvested at test initiation (again, only using the three most homogenous individuals) and plant biomass (wet and dry weight) and length is determined to obtain respective mean biomass data for DAT 0.

The pots with the three plants are placed into the test vessels (one pot per vessel). Bulk solutions are set up for each treatment level. 1.8 L per replicate of these solutions will be added very carefully (i.e. via a funnel) in order to avoid any disturbance of the sediment.

The shoot length above sediment is measured thereafter (DAT 0) using a ruler inside the test vessels.

The water level should be marked. Beakers may be loosely covered (by a transparent cover such as a thin glass plate) during the study to prevent evaporation (and contamination with alga spores). If water evaporates during the test by more than 10%, the water level should be adjusted with distilled water.

Test conditions

Warm and/or cool white fluorescent lighting should be used to provide a light intensity in the range of about 140 (+/- 20) $\mu E \cdot m^{-2} * s^{-1}$ when measured as a photosynthetically active radiation (400-700 nm) (equivalent to about 8.8 -11.8 klux)[21] at the water surface and using a light:dark ratio of 16:8 h. Any differences from the selected light intensity over the test area should not exceed the range of ± 15 %. The temperature in the test vessels should be 20 ± 2°C.

[21] The method of light detection and measurement, in particular the type of sensor, will affect the measured value. Spherical sensors (which respond to light from all angles above and below the plane of measurement) and "cosine" sensors (which respond to light from all angles above the plane of measurement) are preferred to unidirectional sensors, and will give higher readings for a multi-point light source of the type described above.

Biological assessments

The exposure period is 7 days for *M. aquaticum* and 14 days for *M. spicatum*. During this time, visual inspections on plant growth, morphological changes or any other unusual observations are recorded at least twice during the exposure period (i.e. on days 3 and 5 for *M. aquaticum* and on days 5 and 10 for *M. spicatum*). In addition, shoot length is determined once during the test (day 3 for *M. aquaticum* and day 5 for *M. spicatum*), e.g. using a ruler positioned within the vessel close to the plant to be measured. It may be necessary to straighten shoots (obviously this needs to be done carefully without inflicting any damage to the plant) for more accurate length measurements. If side shoots are present, their numbers and length should also be measured (the value of the interim measures is mainly to show constant plant growth over time, respectively to indicate lag phases or potential recovery phases; to that extent a lower precision of the interim measurements using a ruler or visual observations within the system is comprehensible and acceptable; in general, these data are not intended for quantitative statistical evaluations).

At the end of the test, all plants are measured again (shoot length above sediment) and any growth anomalies are recorded; thereafter the whole plants are harvested. Any symptoms (such as chlorosis or necrosis) or other observations are recorded. Total plant wet weights (after carefully blotting off remaining test medium) and subsequently, total plant dry weights are determined. A visual assessment of the roots is made and any unusual findings should be recorded. A summary of the minimum biological assessments required over the test duration is provided in Table 1.

Table 1: Assessment schedule

Day after treatment (DAT)	***Myriophyllum spicatum***			***Myriophyllum aquaticum***		
	Shoot length	**Shoot weight**	**pH, O_2 visual inspection**	**Shoot length**	**Shoot weight**	**pH, O_2 visual inspection**
0	X	(X) [1)]	X	X	(X) [1)]	X
3	-	-		X	-	X
5	X	-	X		-	X
7	-	-		X	X	X
10		-	X	-	-	-
14	X	X	X	-	-	-

X indicates that measurements are made on these occasions; "-" indicates that measurements are not required

1) no direct measurement of tested plants; instead based on information of average weight from the representative sample taken in the additional test vessel after adaptation phase

Environmental assessments

pH, oxygen levels and conductivity of the water are determined in bulk solutions of each treatment level at test initiation. Light conditions are measured at several random places between test vessels at the hight of the water surface. Temperature in the test medium of several random vessels places at different places (and within the room) should be monitored over the whole test period. The pH and oxygen concentration of the test medium (water) should

be checked at test initiation (bulk solutions, see above), twice during the study and at the end of the study in all replicate vessels.

Analytical measurements of test substance

The correct application of the test substance should be supported by analytical measurements of test substance concentrations in water at test initiation (in bulk solutions) and termination (minimum in combined replicates per treatment)at least in the lowest, the medium and the highest concentration group. Where significant losses occur during the study (> 30%) concentrations in sediment should additionally be determined at test termination.

Several methods for 3,5-DCP analysis are available in the literature; it is left to the participating labs to select a method suitable for the purpose of the study and the analytical options available. A proposal for an analytical method for isoproturon (trifluralin) will be supplied (with the test substance).

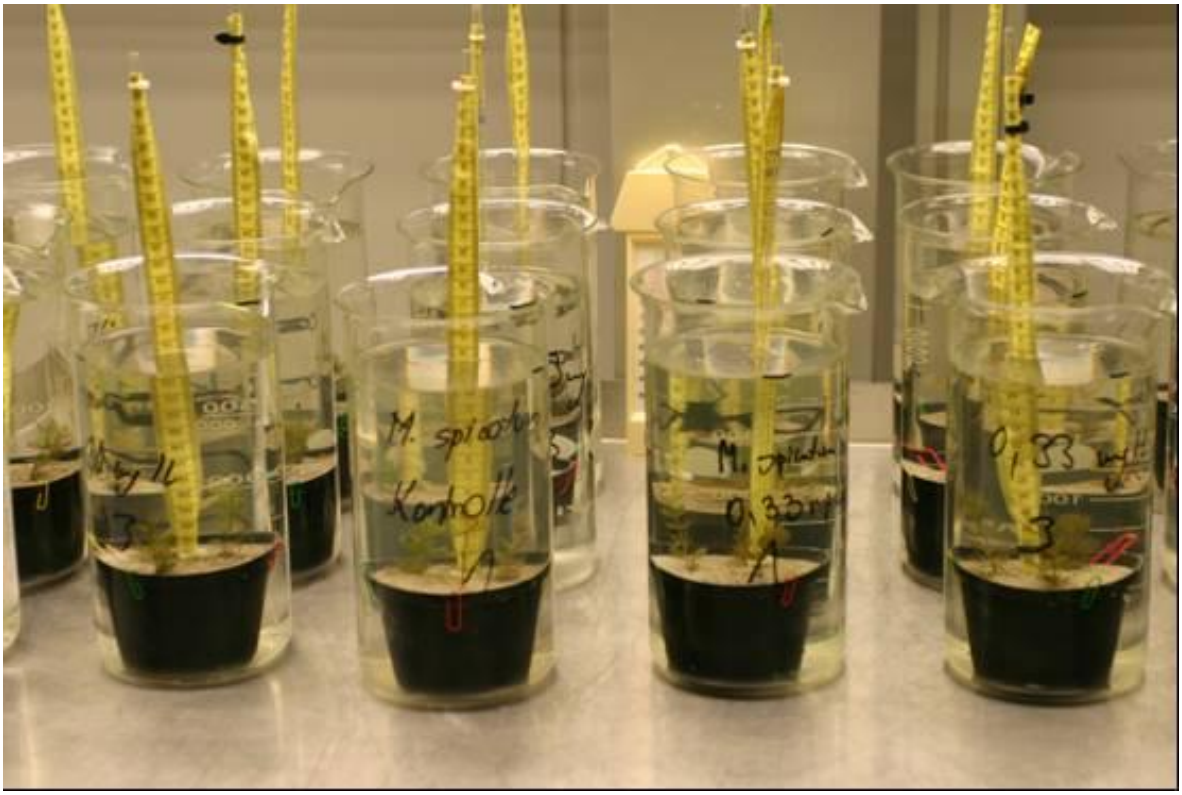
Appendices

Appendix 1
Preparation of **Smart & Barko** Medium (1985)

	Amount of reagent added to water
$CaCl_2 \cdot 2 H_2O$	**91.7 mg/L**
$MgSO_4 \cdot 7 H_2O$	**69.0 mg/L**
$NaHCO_3$	**58.4 mg/L**
$KHCO_3$	**15.4 mg/L**

pH (air equilibrium) – 7.9

Appendix 2, Pictures of set up



Appendix 3, Potential plant suppliers (in Europe):

Jörg Petrowsky, Anschauteiche, D-29348 Eschede, Germany
(petrowsky.wasserpflanzen@t-online.de, www.repo-pflanzen.de)

Zoo Schnack Genin, Wakmühlenweg 1-3, D-23560 Lübeck, Germany
(info@zoo-schnack.de)

Aquaflora B.V. , Uitweg 25, 3645 TA Vinkeveen, The Netherlands
Telelefon: +31-297-286-709, Fax: +31-297-289-256, E-mail: info@aquaflora.nl,
www.aquaflora.nl.

Institut für Gewässerschutz, Mesocosm GmbH; Neu-Ulrichstein 5; 35315 Homberg (Ohm);
Germany; www.mesocosm.de; ebke@mesocosm.de