HESSD-2014-354 Revision report
Local nutrient regimes determine site-specific environmental triggers of cyanobacterial and microcystin
Som Cit Sinang, Elke S. Reichwaldt and Anas Ghadouani

Dr. Christian Stamm
Editor HESS
Environmental Chemistry
Eawag
Ueberlandstr. 133
CH-8600 Dübendorf, Switzerland

Dear Dr Stamm,
Thank you very much for providing very helpful and detailed comments to further improve our manuscript. We would especially like to thank you for pointing out the use of linear mixed models instead of linear regressions. We have now spent a significant amount of time revising the manuscript and have consulted with Pierre Legendre, a colleague and well-known statistician on how to best statistically analyse our data. In our new version of the manuscript, we are still using Pearson's correlations in the descriptive phase of the study, but then use linear mixed models to calculate regressions between explanatory and dependent variables for each lake separately (using sampling site and sampling date as random factors) and for all lakes (adding lake as random factor). For this, please see our replies to your comments #2, #4, and #28, in particular. We have also added the results of the RDA for cyanobacterial biomass as suggested by you.
Most importantly, we would like to point out, that the use of different statistical methods has not led to major changes in the results and has not led to any changes in the conclusion of the manuscript.

Please find below a point-by-point revision report along with the revised manuscript for your consideration. We have highlighted the sections in the manuscript which have been amended or re-written.

	Comments	Response	Location in text
1	Reviewer 1 comment #4 On this issue, I don't agree. First, I cannot find the answers you provide in your response in the text (neither on L. 409 – 417 nor between 437 and 445).	We agree that our reply wasn't clear in the earlier version. We now clarified this by adding the following sentence to the discussion section: "Although the lakes in our study were not limited in phosphorus *per se*, the differences in phosphorus levels could have been responsible for the differences in the phytoplankton communities between lakes."	Lines 477-479
	Second, I think the RDA you have carried out should be included. Whether you include it into the main text or provide it as Supporting Information is up to you to decide.	We agree and now included the second RDA in the following places: a) method section (section 2.3 Data processing and statistical analyses) b) results section (3.5) c) abstract and discussion sections, where appropriate	a) Lines 284-287 b) Lines 408-423 c) Lines 36-40; 436-437; 521-528
	Third, the issue of absolute cyanobacteria biomass and relative fractions seems more important to me than what is appears from the current paper. I ask myself the simple question whether it is worse to have a lower biomass of cyanobacteria (and related microcystin) that contributes a larger fraction to total biomass as compared to a lower fraction but higher absolute biomass. Perhaps, I missed that point. If yes, please indicate where to find it. Otherwise, provide a discussion of this issue in the manuscript (or explain	We agree that it is important to mention this and now included a discussion of your question, whether "it is worse to have a lower biomass of cyanobacteria (and related microcystin) that contributes a larger fraction to total biomass as compared to a lower fraction but higher absolute biomass" at various places: 1) Introduction: "In addition, microcystin production by cyanobacteria is a complex issue that might depend on their competition with other	Lines 65-70

	why the question does not make sense).	phytoplankton (e.g., Huisman and Hulot, 2005; Jang et al., 2006). From these earlier studies it can be concluded that the toxin concentration produced by a certain cyanobacterial biomass level might differ, depending on the level of competition (i.e. cyanobacterial fraction) indicating that management should consider biomass and cyanobacterial fractions concurrently." 2) Introduction: "Furthermore, the occurrence of cyanobacterial toxins in a system is the result of a complex interaction between abiotic and biotic factors, including the competition with other phytoplankton. It therefore remains an important challenge for bloom management to fully understand the mechanisms behind toxic cyanobacterial bloom development and the drivers for biomass development, cyanobacterial dominance (fraction) and toxin production." 3) Discussion: "Based on the RDA results, the measured environmental factors were able to better predict the variability of cyanobacterial fraction than the variability of absolute cyanobacterial biomass in two out of three lakes (Yangebup and Jackadder Lakes). Both descriptors are important indicators for management. The competition with other phytoplankton, described by the cyanobacterial fraction in this study can affect the toxin production within a cell through allelopathy (Huisman and Hulot, 2005). Therefore, understanding the importance of site-specific drivers of both, biomass	Lines 94-99 Lines 521-528

		and the cyanobacterial fraction is of highest importance to develop successful and sustainable management strategies." In addition to the RDA, we now also included Pearson's correlation analysis and linear mixed models for cyanobacterial biomass.	 Table 2, Fig. 8E, F, Lines 367-369, 390-392
2	Reviewer 2 comment #7 Here, it is not clear how the GLM was actually set up. On L. 260 – 261, it is stated that the three replicates per site and sampling date have been used for calculating an average. Fig. 5 and Fig. 6 and Tab. 2 however, indicate that all single measurements have been used (N = 48, which does not correspond to 3 lakes times 6 sampling dates). Because there were 3 distinct sampling points in each lake (L. 152 – 153), I assume you have also used the location as a random factor (see also comment on top and at the end). Please specify your model precisely by providing the equation. This may go into the Supporting Information. On the other hand, in Fig. 7 only 5(?) data points are depicted for Lake Bibra. Why that?	This is an important comment and we would like to thank the editor for pointing out that we should have used linear mixed models (LMM). We have revised the manuscript as follows: a) Deleted the GLM including the discussion of its results b) We clarified our sampling design. "The lakes were sampled twice a month between January and March 2010, leading to 6 sampling days. Three samples were collected from the same three points at each lake on every sampling occasion. As Bibra Lake dried up in late February no samples were taken from this lake in March, leading to only 4 sampling days." c) We explain why only five data points for Bibra lake are shown in Fig. 7 (now Fig. 8) in the method section and the figure legends:	 a) previously in the following sections: methods, results, and discussion b) Line 165-168

		Methods: “All dependent variables were ln-transformed. As extracellular microcystins were only detected in five out of twelve samples in Bibra Lake, this resulted in only five data points for this dependent variable in Bibra Lake…”. Table 3: “Extracellular microcystin fraction was zero in seven cases, leading to an N = 5 only” d) Instead of using simple linear regressions for earlier Figs. 5-7 (now Fig. 6-8), we now calculated linear mixed models as specified in the method sections: “We used linear mixed models to identify correlations between environmental variables and cyanobacterial fraction, cyanobacterial biomass, cellular microcystin concentration and extracellular microcystin fraction in each lake using sampling site and sampling date as random factors, and for all lakes combined adding lake as random factor (SPSS 21.0). All dependent variables were ln-transformed. As extracellular microcystins were only detected in five out of twelve samples in Bibra Lake, this resulted in only five data points for this dependent variable in Bibra Lake, making the calculation of linear mixed models for this explanatory variable impossible.”	c) Lines 281-284 Table 3 Lines 277-284
3	Reviewer 2 #8 Again, this is not sufficiently clear: did you analyse autocorrelations for data of each location? The data in Fig. 5 to 7 strongly suggest that there were strong	We agree and have resolved this problem by using linear mixed models which take the sampling date as a random factor into account. Therefore, we believe	Previous version: methods

	temporal changes in the chemical status of the lake. If this was true, how could distinguish autocorrelation from a trend with such a short time series? Please explain and show some actual data (as supporting information) for illustration.	that the autocorrelation analysis is now redundant and we have deleted it.	
4	Comments to the Editor #7 There is a question of what the regression for all lakes actually means: does this regression correspond to the fixed effect if you consider the three lakes and the three sampling locations in each lake as random effects (in a mixed model)? To be honest: the regression line depicted for all lakes in Fig. 5 – 7 often looks like it was calculated based on the assumption that all 48 data points were independent (which they aren't). As a consequence the slope seems often to be controlled by the large range observed in Lake Bibra. This can be illustrated by Fig. 6A or 6B.	We confirm that in the previous version we have used simple linear regressions on the assumption that all 48 data points were independent and we would like to thank the editor for pointing this out. We have substituted this calculation now with linear mixed models that take lakes, sampling sites and sampling date into account. We have specified the model in the method section and the figure legends. Please also see our reply #d to your comment #2). Regarding the question what the "regression for all lakes actually means": We think that this is a very important point and we added a clarification of why we show regressions for each lake separately and a combined regression for all lakes: "Currently, in the absence of lake-specific information, cyanobacterial management strategies are based on knowledge derived from general trends of the relationship between environmental factors and cyanobacteria or their toxins. Our study clearly indicates that the environmental variables explaining the variability in cyanobacteria and their toxins might be lake-specific and, more importantly, that these lake-specific correlations might also be different to	Lines 277-284 Figure legends: Fig. 6-8 Lines 464-471

		the correlation derived from combining all data (e.g., 6A, 8A, B). This strongly supports the conclusion that site-specific conditions have to be taken into account for managing lakes with cyanobacterial blooms."	
	The slopes for all three lakes individually are steeper than for the entire data set. How can that be? Please specify the model used for the regression for the individual lakes and for the entire data set (see also comment on top and at the end).	In the new version of the manuscript, we now specify how regressions were calculated as follows: • Methods: "We used linear mixed models to identify correlations between environmental variables and cyanobacterial fraction, cyanobacterial biomass, cellular microcystin concentration and extracellular microcystin fraction in each lake using sampling site and sampling date as random factors, and for all lakes combined using lake as random factor (SPSS 21.0). All dependent variables were ln-transformed. As extracellular microcystins were only detected in five out of twelve samples in Bibra Lake, this resulted in only five data points for this dependent variable in Bibra Lake, making the calculation of linear mixed models impossible." • In the figure legends: "Regression curves for each individual lake were calculated by linear mixed models with site and date as random factors on data from each lake (broken lines) while all data points were combined for the overall regression using a linear mixed model adding lake as random factor (solid line). Only significant ($p<0.05$) regressions are shown. "	Lines 277-284 Figure legends: Fig. 6-8

5	Abstract: The text is not very elegant. For example, the first sentence ends with "…management strategies." The second one starts with "In the management …". For that section I have tried to come up with a eeworded version like "Toxic cyanobacterial blooms in urban lakes present serious health hazards to humans and animals and require effective management strategies. Managing such blooms requires a sufficient understanding of the controlling environmental factors. A range of them has been proposed in the literature as potential triggers for cyanobacterial biomass development and microcystins formation. …".	We re-wrote the abstract accordingly: "Toxic cyanobacterial blooms in urban lakes present serious health hazards to humans and animals and require effective management strategies. Managing such blooms requires a sufficient understanding of the controlling environmental factors. A range of them has been proposed in the literature as potential triggers for cyanobacterial biomass development and cyanotoxin (e.g., microcystin) production in freshwater systems."	Line 27-31
6	L. 36: Please insert "… in three urban lakes in Western Australia."	We inserted this as suggested and it now reads as: "In this study, we investigated the site-specificity of environmental triggers for cyanobacterial bloom and microcystin dynamics in three urban lakes in Western Australia."	Line 34-36
7	L. 41 – 42: Please explain a bit more in detail.	We added the following sentences to the abstract to explain this in more detail: "For instance, we found no correlation between cyanobacterial fraction and total phosphorous (TP) in the lake with the highest TP concentration, while correlations were significant and negative in the other two lakes. In addition, our study indicates that the difference of the correlation between TFe and the cyanobacterial fraction between lakes might have been a consequence of differences in the cyanobacterial community structure, specifically the presence or absence of nitrogen-fixing species."	Lines 41-47

8	L. 54: Do all urban lakes suffer from this problem? Be more careful with the wording.	We agree that not all urban lakes suffer from this problem and have revise the sentence and it is now read as: "To date, many urban lakes continue to deteriorate due to increased anthropogenic activities and often face water quality problems including toxic cyanobacteria blooms."	Line 57
9	L. 65: spatial basis: do you mean within a lake or within a geographical region? Be more precise.	We actually meant both and are now more precise. The sentence now reads: "Cyanobacterial biomass and the amount of microcystins being produced during toxic cyanobacterial blooms can vary significantly on a spatial basis within and between lakes (Reichwaldt et al., 2013; Sinang et al., 2013; Thi Thuy et al., 2014; Waajen et al., 2014)."	Lines 72-74
10	L. 68: Do you mean populations or planktonic communities here?	We clarified this by changing the sentence to: "Past studies have found large variations in the percentages of potentially toxic cyanobacteria and microcystin concentration between spatially isolated phytoplankton communities."	Line 76
11	L. 77, 80: phosphorus/iron concentrations	We have added the word "concentrations".	Lines 86, 89
12	L. 97 – 99: Be more specific why you expect site-specific relationships. Is it because of additional factors (not explicitly accounted for) that distinguish these lakes or is it because of the interplay of the factors that are explicitly considered?	We are now more specific and added the following information: " By taking into account the contrasting findings of earlier studies, including inconsistent outcomes of nutrient reduction strategies, we suggest that the main environmental triggers of cyanobacterial and microcystin variability may vary between water	Lines 105-108

		bodies due to the complex, lake specific interplay of environmental conditions."	
13	L. 145: For which period have these mean values been reported?	We added the period for which these values have been reported: "… earlier study conducted between November 2008 and July 2009 (Sinang et al. 2013), these lakes …"	Line 157
14	L. 177: Be consistent with the spelling: is it Fluoroprobe or FluoroProbe?	We corrected the spelling mistake.	Line 197
15	L. 246: Knowing the temperature range in the water implies actual measurements. Why haven't you used them directly?	We clarified our rationale for using air temperature instead of water temperature as follows and additionally including a citation of a study by Yen et al 2007 who found that this method is reliable for shallow lakes: "Surface water temperatures were between 19.9 and 28.7°C during the study period. However, the onsite measurements of surface water temperatures were dependent on the time of sampling and varied by up to 3.9°C over the course of a day. Therefore, maximum air temperature on each sampling day recorded by weather stations located nearest to the studied lakes was used as a substitute for surface water temperature in all analyses (Yen et al 2007)."	Lines 185-189
16	L. 244 – 248: This is not really data processing or statistical analysis. Please move to 2.2.	We agree, and we have moved the sentence to section 2.2.	Lines 185-189
17	L. 248 – 252: This part is not sufficiently clear. Please provide the actual model equation in the SI and show some of the corresponding data (see also comment above (Reviewer 2) and below regarding Sec. 3.1).	We believe that the autocorrelation calculation is redundant in the latest manuscript, as we now use lakes, sampling sites and dates as random factors in linear mixed models. Therefore, we deleted this section. Please also see our reply to #3.	Previous version: methods
18	L. 252 – 257: Please move to 2.2.	We have moved the sentence to section 2.2.2 as we think it is more suitable.	Lines 263-268

19	According to Fig. 5 – 8 and Table 3, you have used the individual data points and not the average values. Please clarify in the text.	We agree and confirm that all 48 data points were used in all analyses. Due to this, we have deleted the sentence. Please also see our reply to comment #2 above.	Previous version: Lines 258-261
20	Section 3.1: I think this part contains important information for understanding the situation in the three lakes. Tab. 1 demonstrates that many variables had a large range. I assume that this implied severe changes over time (see also comment by Reviewer 2). Please describe here the most prominent temporal patterns in the data (including covariance of water chemistry parameters). In addition, provide explanations why the nutrient levels change so dramatically. This may be important for understanding the context of algae dynamics as well.	We agree and therefore we added a discussion of temporal changes of physicochemical parameters to section 3.1 and also included time plots (new Figure 3). We agree and in the new version of the manuscript we included a discussion of the fact that the temporal variability of salinity and macronutrient concentration (especially TP and TN) in Bibra Lake could be due to the lake drying up during the study.	Lines 296- 319 Fig. 3 Lines 298-300 Lines 307-309
21	L. 326 - 328: You make a general statement regarding nutrient levels. However, you did not include N species in this list. A positive correlation with the TN:TP ratio can be due to relatively low TP concentrations or due to relatively high TN concentrations. Please specify.	We agree and therefore re-wrote the sentence: “In contrast, cyanobacterial fraction was positively correlated with TN:TP ratio, potentially due to relatively lower TP concentrations in comparison to TN concentrations.”	Lines 366-367
22	L. 341: Consistency can take different forms: there could be a consistent pattern in that the correlations for all lakes are either positive or negative for a given independent variable. Having the same slope and/or intercept is a much stronger criterion for consistency. Perhaps you may differentiate here because for some correlations even the sign of the slope is different between lakes (see Fig. 5B or C).	We agree and this section has been rewritten considerably due to the fact that we now calculated linear mixed models instead of GLM. Please note, that section is now included in section 3.3.	Lines 379-397

23	L. 352 – 353: There is a negative correlation between TDP and cyanobacterial fraction in Lake Bibra.	Thanks for pointing this out. After calculating a linear mixed model, the negative correlation was not significant anymore and this was deleted in this section. This does not change the conclusion of our manuscript. Please also see our reply to comment #22.	Lines 379-397
24	L. 430, 451: Skip "variability of": the cyanobacterial fraction is correlated directly with the mentioned water quality parameters. If you actually refer to the variability then you need to show actual data.	We agree and deleted the word "variability".	Line 434, 439
25	Table 1: Please explain the ANOVA that you list in the last column: What was actually compared?	We agree and rephrased our previous explanation (section 2.3) to make it clearer. We now also include a brief explanation in the caption of this table: "Differences in physicochemical factors, cyanobacterial biomass and microcystin between lakes were analysed with one-way ANOVA (SPSS 17.0) with post hoc test (Least Significance Difference; LSD) as all assumptions for an ANOVA were met (homogeneity of variances, normality)." "Physical and chemical properties of the three lakes throughout the sample period (Jan – March 2010), including analysis of differences between lakes (one-way ANOVA)."	Lines 271-274 Table 1
26	Table 2: The caption seems to be at odds with the description on L. 266 – 271.	We agree and changed the caption to: "Table 2. Pearson's correlation coefficients (R) between the environmental factors and cyanobacterial fraction (%) or cyanobacterial biomass (µg chl-a L^{-1}) analysed for each lake and for all lakes combined using bivariate correlation	Table 2

		analysis.”	
	Please specify the model used for this analysis.	As we are not using the GLM in this latest version of the manuscript, we believe that this comment is now redundant.	
	Make sure that you properly take into account that not all data are independent: that’s what is your manuscript about if you consider the three lakes! Additionally, you may also have to account for the three sampling locations in each lake. Accordingly, a mixed model would be an appropriate approach.	In the new version of the manuscript we took this into account by using linear mixed models to analyse the relation between explanatory and dependent variables using sampling site, sampling date and lake as random factors. This is explained in the methods. Please also see our reply to comment #2.	Lines 277-284
	Such a model would directly yield the input you need for Table 3.	Any significant results of the linear mixed models are shown in Figures 6-8. We decided to still use Pearson’s correlations for the descriptive phase of the manuscript (Tables 2, 3). However, please note, that instead of using Table 2 for ‘all-lakes-combined’ correlations and Table 3 for the ‘each-lake-separately’ correlations, we now provide data for all analyses for cyanobacterial fraction and cyanobacterial biomass in Table 2 and data for all analyses for cellular microcystin concentration and extracellular microcystin fraction in Table 3.	Table 2, 3
27	Table 3: Units in Column 1 can be skipped.	We agree and have deleted the units; also in Table 2	Table 2, 3
28	Fig. 5 – 7: It seems that the regression for all lakes is calculated with an ordinary linear regression using all data points (see comment # 7) without taking the lake-specific dependence into account. Based on Table 3 I assume that this holds true also for the regression lines of the single lakes.	We agree and would like to thank the editor for pointing this out to us. Please also see our reply to comments #2 and #4. We have now used linear mixed models instead of simple regressions to calculate the relationship	Figs. 6-8

	There are two comments regarding these regressions: First, describe more precisely what how the regression you show have actually been derived. Second, use a regression approach that correctly accounts for the dependencies of the data. Because your data demonstrate that the data are influenced by the lakes, pooling all data for an ordinary regression is not appropriate. A mixed model would be a valid option.	between explanatory and dependent variables in these Figures, which are now Figs 6-8. The models are explained in the method section as follows: "We used linear mixed models to identify correlations between environmental variables and cyanobacterial fraction, cyanobacterial biomass, cellular microcystin concentration and extracellular microcystin fraction in each lake using sampling site and sampling date as random factors, and for all lakes combined using lake as random factor (SPSS 21.0). All dependent variables were ln-transformed. As extracellular microcystins were only detected in five out of twelve samples in Bibra Lake, this resulted in only five data points for this dependent variable in Bibra Lake, making the calculation of linear mixed models impossible." We also added a short version of this in the respective figure legends.	Lines 277-284 Figure legends: Fig. 6-8