Waterborne Environmental

Aquatic System Models: a ring study

Testing streambugs: Results of the calibration to control data

Project Team

Dr. Isabel O'Connor, EBP
Dr. Simon Spycher, EBP
PD. Dr. Nele Schuwirth, Eawag

EBP Schweiz AG
Mühlebachstrasse 11
8032 Zürich
Switzerland

EAWAG
Überlandstrasse 133
8600 Dübendorf
Switzerland

Version 1.2
Zürich, 22.6.2022

1. Overall results

The approach taken to calibrate the streambugs model is described in the TRACE documentation of streambugs.

As the result of the streambugs calibration, we present a set of streambugs simulations that corresponds to the top 0.1% of the simulations ("set of top runs", corresponding to 299 out of 280231 runs, as of 07.03.2022). In addition, for information purposes we also show the results of the single best simulation.

The set of top runs fulfill at least 83% of the calibration criteria, the best simulation itself 86% of the calibration criteria (Table 1). Comparing the performances in the plots in appendix 2 (plots of the best run) and 4 (plots of set of top runs) with appendix 6 (plots of the simulation based on the prior) demonstrates that the performance of the streambugs model improved during the calibration procedure.

Overall, the calibrated model performs well for macroinvertebrates and periphyton, while the performance is weaker for zooplankton and phytoplankton. The latter two groups are new to the streambugs model, and the weaker performance may indicate that the pragmatic approaches to represent the new groups within this project could be improved with future work.

Group	Best run	Set of top runs
Overall	86%	83%
Macroinvertebrates	94%	97%
Zooplankton	87%	75%
Phytoplankton	50%	59%
Periphyton	100%	83%

Table 1: Overview of the model performance of the best run and the set of top runs according to the calibration criteria.

All detailed results are provided in the appendix. Table 2 gives an overview of the plots and table in the appendixes.

Notes:

— As agreed in the consortium meeting, the model comparison in the ring study is based on comparing the performance of a single simulation. Therefore, this documentation focuses on discussing the performance of the best run and provides the performance of the set of top runs as additional information (indication of uncertainty).

— There are different ways to calculate the performance. This document shows the performance how it was initially calculated and deviates slight from how the performance was calculated in the final version of the ring study comparing all model comparisons. The excel file with the results lists both the performance according to the initial approach and according to the final approach. This documents focuses on the initial approach to calculate the performance.

Filename	Figure Legend
Results as Table *Model_control_calibration_results_stream-bugs.xlsx*	Excel file with overview on the performance of the streambugs model. Shown are for each species and study the performances of the best run and of the set of top runs, respectively. Values in the table of the best run reflect the values from simulations (e.g., minimum abundance, as required by the template). Values in the table with the set of top runs reflect the share of simulations in the set that fulfill the specific criterion.
Plots for individual species *Model_control_calibration_results_stream-bugs_1_plots_algea.pfd* *Model_control_calibration_results_stream-bugs_1_plots_invertebrates.pdf*	The plots show for each species and study the measured abundance (dots), simulated abundance (blue line), the calibration criteria regarding minimum and maximum abundance (black horizontal lines) and the peak date criterion (grey shaded area).
Plots with additional results per species *Model_control_calibration_results_stream-bugs_2_plots_additional_results_algae.pdf* *Model_control_calibration_results_stream-bugs_2_plots_additional_results_invertebrates.pdf*	The plots show for each species and study the plots as in appendix 2 (see above), combined with information on underlying rates and limiting factors. Top: measured abundance (dots), simulated abundance (blue line), the calibration criteria regarding minimum and maximum abundance (black horizontal lines) and the peak date criterion (grey shaded area). Middle. Underlying rates Bottom: limiting factors.
Spaghetti plots per species *Model_control_calibration_results_stream-bugs_3_spaghetti_plots_algae.pdf* *Model_control_calibration_results_stream-bugs_3_spaghetti_plots_invertebrates.pdf*	The plots show for each species and study the measured abundance (dots), simulated abundances of the set of top runs (blue lines), the calibration criteria regarding minimum and maximum abundance (black horizontal lines) and the peak date criterion (grey shaded area).
Plots per species with priors *Model_control_calibration_results_stream-bugs_4_plots_prior_algae.pdf* *Model_control_calibration_results_stream-bugs_4_plots_prior_invertebrates.pdf*	The plots show for each species and study the measured abundance (dots), simulated abundance using the priors (blue line), the calibration criteria regarding minimum and maximum abundance (black horizontal lines) and the peak date criterion (grey shaded area).
Parameter distributions *Model_control_calibration_results_stream-bugs_5_plots_parameter_distributions.pdf*	The plots show for each parameter the distribution as defined a priori ("priors", black line), the distribution of the parameters in the set of top runs (blue shaded area) and the parameter value used in the best run (blue vertical line)

Table 2: Overview of the files in the appendix.

2. Analysis of the model simulations concerning the calibration criteria

2.1 Macroinvertebrates

The best simulation fulfills 94% of the calibration criteria for macroinvertebrates and the set of top runs fulfill 97%. The slightly lower performance of the best run is due to the fact that for some criteria, the best run just falls beyond the calibration criteria while most other simulations are within the defined range.

The following groups fulfill all calibration criteria[1]:

— Asellus

— (Chironomidae: best run fulfills all criteria)

— Cloeon

— Hirudinea

— Nepomorpha

— Odonata

— Oligochaeta

— Tanypodinae

The following groups do not fulfill one or more criteria in at least in some runs:

— **Chaoborus**: In the best simulation, the minimum abundance in three studies (14a, 16a and 19b with 8.2 – 9.0 individuals / sample) is just below the minimum abundance criterion (≥ 10.6 individuals / sample). The reason of the failure of the best run is the decreasing abundance with time that drops below the minimum abundance criterion towards the end of the studies. The underlying processes reveal that chaoborus are limited by phytoplankton availability. This issue is rather unique to the best run, as 90-95% of the top runs fulfill this criterion.

— **Chironomidae**: The best simulation fulfills all calibration criteria. However, 31% of the top simulations do not meet the maximum abundance criterion in study 16a. 18% of those simulations start already above the criterion (i.e., the simulated abundance at t=0 is higher than the maximum abundance criterion) and the abundances subsequently either increase or decrease. The remaining 82% of the simulations start below the maximum abundance criterion but increase subsequently, such that maximum abundance criterion is exceeded. A systematic analysis of identifying the main reasons why simulations increase or decrease is considered out of scope. Analyzing selected examples (best run & simulation 522335 & 543630) suggest that small differences can lead to passing the tipping point and thus different behaviors. For example, in the selected examples the self-inhibition is slightly more limiting where the abundances are decreasing, while the abundance of the predators and food are similar as a first proxy.

— **Gammarus**: The best simulation fails the criterion regarding peak date in the studies 16a and 18c. The streambugs simulates increasing abundances while the data show a decrease

[1] i.e., where the best run fulfills all criteria, and the set of top runs fulfill for each species and study the criteria in at least 90% of the runs

in abundance with time. Apparently, the gammarus are limited in the studies which is not represented by the simulation. The peak date criterion in 16a and 18c is met in 20-38% of the top simulations.

— **Gastropoda**: The best simulation in study 14a starts just above the minimum abundance criterion and subsequently drops below the criterion. This issue is rather unique to the best run, as 90% of the top runs stay above the minimum abundance criterion and therefore fulfill this criterion. Overall, the simulations represent the observed abundance well. 16% of the top simulations fail the minimum abundance criterion in study 18c for the same reason (i.e., start above the minimum abundance criterion and drop just below it with time).

2.2 Zooplankton

The best simulation fulfills 87% of the calibration criteria for zooplankton and the set of top runs fulfill 75% of the criteria.

The following groups always fulfill all calibration criteria[1]:

— Cladocera small

— Copepoda herbivorous (even though in study 14a it shares the same issue as with daphnia: the first measurement reported no abundances, while later on some organisms were observed. The model simulates no abundance during the whole study, as we did not include inputs of organisms into the system. However, since the calibration criterion included zero, the calibration criteria regarding minimum and maximum abundances were met).

— (Rotifera herbivorous 1: best run fulfills all criteria)

The following groups do not fulfill one or more criteria in at least in some runs:

— **Cladocera phytophilous**: The best simulation fulfills the minimum and maximum abundance criterion in all studies, but the simulation does not fulfil the criterion regarding peak date in three studies with measured abundances (16a, 18c and 19b). The same pattern is observed in most top simulations. The measured abundances start low and increase in 16a and 18c, while the simulation does not predict such a behavior. Instead, the underlying processes reveal that Cladocera phytophilous is strongly limited by the availability of periphyton such that the abundance start low and remain low.

— **Copepod predatory**: The best simulation fails the minimum abundance criterion in the study 19b, because the simulated abundance decreases towards the end of the study and drops below the criterion. The reason is the limited food availability towards the end of the study. This criterion is failed by 92% of the top simulations and 94% and 60% of the top simulations in the studies 16a and 18c.

— **Daphnidae:** The best simulation fails the peak date criterion in all studies. In the studies 16a, 18c and 19b, the peak occurred in all simulations at the beginning of the study, while the peak was measured in the second half of the studies. The underlying processes reveal that daphnia are increasingly limited by food availability as the study proceeds, while the limitation by self-inhibition decreases. The food availability may mainly be determined by the decreasing availability of the phytoplanktonic blue greens. All best simulations fail this criterion.

In the study 14a, the initial measurements report no abundance of daphnidae. Since we did not model any input of daphnidae into the system, the simulation stays at zero, while in the experiment individuals of daphnidae were observed later on. As a consequence of the simulation being zero throughout the study 14a, the minimum peak criterion is not calculated and therefore failed. Apparently, the first measurements did not catch daphnidae even though they were present (though only few organisms). These uncertainties in the measurements were not further addressed in this project but should be discussed for future work. All best simulations fail this criterion.

— **Rotifera herbivorous 1:** The best simulation fulfills all calibration criteria, however, 56 - 71% of the top runs fail the minimum abundance criterion in the four studies. In those simulations the abundance decreases with time and drops below the minimum abundance criterion. The reason may be food limitation (not further analyzed).

— **Rotifera herbivorous 2**: The best simulation fails the minimum abundance criterion in study 19b, because the simulation starts in the range of the measurements but decreases stronger than the measured abundance and therefore drops below the calibration criterion towards the end of the study. Nearly all best simulations share this pattern and show this pattern also in the three remaining experiments. The underlying rates and limiting factors clearly reveal the strong dependence of the herbivorous rotifera on the phytoplankton availability and dynamics.

— **Simocephalus**: The best simulation fails the minimum abundance criterion in study 16a. The simulation starts at an abundance that is below the calibration criterion but very similar to the measured abundance. The simulated abundance increases with time but to a lower extent than observed in the measured data. All top simulations fail this criterion in study 16a. The minimum abundance is also not met in 49% of the top runs in study 14a and 25% in study 19b. In the studies 14a and 19b the simulations start above the minimum abundance criterion, decrease with time, and drop below the minimum abundance criterion.

2.3 Phytoplankton

The simulations of the top set typically fulfill the maximum abundance criteria, while the minimum abundance criteria are typically failed. The only exception are the greens, however, the reason is that the calibration interval for greens includes 0 and thus the minimum abundance criterion is always fulfilled.

The plots of the set of top simulations demonstrates that simulations for phytoplankton either approach zero within the very first timesteps and therefore are below the minimum abundance criteria or they explode and exceed the maximum abundance criteria. This demonstrates that the parameter sets for phytoplankton are extremely difficult to find and were apparently not found with the approach used in this project.

As such, the **best run** is a “nice” example where the blue greens explode and exceed the maximum abundance criterion, while the remaining phytoplankton groups approach zero within the very first time step. The underlying processes and limiting factors reveal that the phytoplankton are strongly limited by nitrogen availability. The diatoms are also limited by light, while the blue greens are also limited by self-shading.

2.4 Periphyton

In the best simulation, all calibration criteria are fulfilled. However, in the set of top simulations, the lower abundance criterion is in around 30% of the runs not fulfilled. Limiting factors are nitrogen and light availability.

2.5 Conclusion

The results show that with the calibration procedure used in this project, the simulations fulfill 83% of the calibration criteria (best run: 86%). While the simulation performed well for macroinvertebrates and periphyton, the new groups phytoplankton and zooplankton appear more difficult. The reason is that some species are extremely sensitive to small changes in parameter sets and it appears difficult to find a suitable set of parameters.

There are three types of improvements that future work could address (beyond the scope of this project)

— Improvements in the model structure: especially including nutrient cycle may further help to improve model performance as it restricts the parameter space during the calibration procedure. This aspect is more relevant in closed systems like mesocosm as in open systems like small streams.

— Calibration method: Streambugs is typically calibrated using the Bayesian inference method. However, the Bayesian inference method is a) not applicable for the binary calibration criteria defined in this project and b) relies on a likelihood function and information on the sampling error that was not available in this project. Furthermore, it is resource- and time-intensive to conduct and there are no guarantees that it is successful. Nevertheless, it has shown to be successful in most applications of streambugs so far. Due to the timeline and set up of this project the project team proposed to use the opportunity to test an alternative calibration method that a) fits the binary calibration criteria that were provided and b) test whether robust results can be obtained with a simple method. The results show however, that it may be worth to invest in the Bayesian inference method and using assumptions where relevant information is missing

— Experimental methods: Linking the streambugs model with the experimental data of the mesocosm could be further improved if mesocosm studies measure nutrient concentration (especially nitrogen) with sufficient low LOQ. We are aware that this information is beyond what is needed for regulatory purposes, however, for modelling purposes this information would be very valuable. Furthermore, standardized methods to sample invertebrates could better enable the extrapolation of number of individuals in samples to the abundance in the whole mesocosm.