The manuscript reports an interesting study whereby different combinations of experimental data are used for parameterisation of a 3-species microcosm model in an effort to identify appropriate testing strategies for generating ecotoxicological data that avoids unnecessary testing. Furthermore, this particular experimental and modelling setup aims to tease apart direct and indirect effects of chemical exposure. The principle of using models and experiments in an integrative, iterative process for optimisation is not new, but this manuscript presents a nice illustration of how this approach could be used in small, multi-species microcosms. The authors finish by recommending a testing strategy for future microcosms studies. Below are a few comments:

Experimental setup:

P6 L114 Experiment 1 is referred to as "Experiment without sediment". This suggests that all other experiments were performed with sediment, however, it is not clear whether this is the case when reading through the design of the other experiments in the manuscript. Clarification is required. This is especially pertinent to the experiments with cadmium, as the fate of cadmium ions will depend upon processes such as absorption, precipitation etc and result in both suspended and dissolved forms partitioned across a number of chemical species. Due to the low solubility of cadmium, much of it may precipitate and bind to the sediment which may alter the bioavailability to (and the effect on) the organisms. Whilst the authors state that analytical confirmation of cadmium was performed throughout the experiments, presenting mean measured values, depending on the presence of sediment (or not) may necessitate some additional discussion on the fate of cadmium in the system and any implications of its toxic effect and modelling outcomes (e.g. this may further inform the discussion on the NEC for *Daphnia* survival when compared with that in the wider literature as discussed on P17 L361-377).

Table 2 Why are there no 2-species experiments including *Daphnia* and algae in the cadmium exposed studies? P15 L 311 states that algae-*daphnid* interaction (grazing) parameter wasn't well supported in the model fit. This is quite a critical parameter if exploring multi-species interactions and could the 2-species study have helped to address this? Inclusion of these 2-species experiments would also have provided more interesting results in the modelling using only partial experimental results (as the *Daphnia* results would also have changed i.e. the "super-imposed" look of the daphnia results in figure 4 between the partial datasets and reference data would potentially have differed).

Modelling assumptions:

P11 L218 "*we make the assumption that the contaminant toxicokinetics is fast*" Given the potential importance of this assumption on the modelling results and effects observed, can the authors provide further justification for the validity of this assumption.

P13 L269 The reference data were considered the "*best possible estimates in the present case study in view of the model and all available data*". The authors have been clear that what they are comparing is one estimate against another (better) estimate (the best possible estimate given the limitations of the data available). It would be interesting to know whether the additional 2-species tests mentioned above would have improved this reference estimate and if so, what implications that would have on the results and conclusions (i.e. is the experimental plan incomplete thereby potentially undermining the reference estimate).

Results and discussion:

I generally found these sections an appropriate reflection of the study outcomes. How to integrate indirect effects as seen in this study into regulatory chemical risk assessment globally may be a

challenge ahead of us, as well as how to deal with other species and chemicals in a microcosm-modelling approach. However, models are likely to only become more important in the future for application in risk assessment so studies such as this one which optimise testing in line with modelling will only become more essential. Given the results obtained in the study, the authors' recommendation for future testing strategies is fair, even if the aim of reduced testing may not always be realised in practice given the remaining difficulty of knowing in advance what data is most essential.

Whilst generally well written, the meaning of a few sentences are unclear and should be addressed.

Note. I have not evaluated the R script and model code, though these are available for readers as a pre-requisite for submission to PCI