**Author Rebuttal**

We thank the reviewers for their detailed reviews which led to significant addition of new material in the paper and clarification of its motivations. The changes are highlighted in the text.

**Response to reviewer 1:**

As we now mention in the paper more than once, our objective is not to project the LOO cross-validation as a cure-all approach to all validation problems. Our aim is to portray the severe inadequacies of external validation for small samples of high-dimensional data, and to foster a more involved outlook in the QSAR community regarding the use of proper validation method.

We expanded the data analysis to 300 simulated datasets, and the outputs corroborate our previous findings on the previous simulation scenario as well as the real dataset. Keeping in mind the length of the paper we stop here to keep the succinctness of our message intact. However, we do plan to extend our approach to several diverse datasets, modelling and validation methods and sample-splitting techniques through future studies: we mention this in the conclusion. Our paper is meant to serve as a starting point of a body of research into case-specific best practices for validation of QSAR models, and we sincerely hope that the paper achieves this objective.

We also mention the caveats of LOO and the need for research and application of intelligent splitting methods in the intro and conclusion.

**Response to reviewer 2:**

As we now mention in the paper more than once, our objective is not to project the LOO cross-validation as a cure-all approach to all validation problems. Our aim is to portray the severe inadequacies of external validation for small samples of high-dimensional data, and to foster a more involved outlook in the QSAR community regarding the use of proper validation method.

We expanded the data analysis to 300 simulated datasets, and the outputs corroborate our previous findings on the previous simulation scenario as well as the real dataset. Keeping in mind the length of the paper we stop here to keep the succinctness of our message intact. However, we do plan to extend our approach to several diverse datasets, modelling and validation methods and sample-splitting techniques through future studies: we mention this in the conclusion. Our paper is meant to serve as a starting point of a body of research into case-specific best practices for validation of QSAR models, and we sincerely hope that the paper achieves this objective.

We have given references to previous and well-known works on external validation.

Monte carlo cross validation is the same as our multi-split validation. We clarify that in the manuscript with proper reference.