This study proposes a new framework for estimating reaction norms and their variations due to inter-individual differences in genetic background. This objective is especially laudable in that it aims to facilitate comparisons of phenotypic plasticity across different studies, different organisms and different traits, which is a very real problem, notably for meta-analyses. Overall, I find the manuscript clear and to the point. I especially appreciated the explanations and illustrations that accompany the presentation of the variance partition itself.

The framework presented by the authors appears sound, and I found their tutorial on R well explained and with sufficient comments to be understood by biologists knowing about the software (and perhaps familiar with tidyverse). However, I find it unfortunate that the files "model_p2_ds.rds", "model_cs_ds.rds" and "model_nl_ds.rds" are missing from the repository linked in the manuscript (at the time I got it, at least), even though the phrasing of the comments suggests that they should have been included. I was able to rerun the models using some commented parts of the provided script, but I think the inclusion of these files would make this tutorial easier to use, especially for biologists less familiar with R.

The influence of the ways the environment is sampled is briefly addressed regarding its impact on the estimated values of "trendiness" or "curviness" (as per the authors) of the reaction norms. I would find interesting to expand on the role of sampling size and distribution on the efficiency of the framework presented.

Firstly, I would have liked for the authors to tackle more comprehensively the efficiency of their method when applied on more scarce datasets. As they point out in their parameter estimation on a simulated dataset, they considered a substantial amount of data points (4,000). The question arises as to whether and to what extent the accuracy of parameter estimation degrades for smaller data sets, which can be several orders of magnitude smaller, particularly for data collected in nature. The authors report that reducing the number of environments considered (from 10 to 4, a reduction to 1,600 points) did not qualitatively affect the result. I would be particularly interested in a reduction in the number of samples per genotype, which is set at 20 in both cases.

Secondly, I would have found it interesting to develop a bit more the role of the distribution of observations across the range of possible environments. Practical constraints may limit the number of observations on certain parts of the environmental gradient considered, typically the most extreme values, resulting in a non-uniform sampling of the different possible environments. Would this type of data affect the estimates of all the different parameters equally, or would the precision of some of them be more severely degraded? This aspect is briefly addressed in the manuscript, where the author talks about the impact of a Gaussian/uniform distribution on the values of π_b and π_c, but I think this question could be considered more thoroughly.

Going a step further, how would this framework behave with incomplete experimental designs (e.g. observations at different values over a continuous environmental gradient for the different genotypes)?

The authors do a good job of presenting the robustness of their framework to various distributions of the actual reaction norm, and show in particular that polynomial functions can be appropriate for approximating other reaction norm shapes. Yet, the examples presented have notably led me to wonder about the robustness of fitting a sigmoid reaction norm by (essentially) a linear function. Would the results presented by the authors remain as satisfying for other parameter values of the reaction norm? For certain extreme values, a sigmoid can be akin to a jump in trait value beyond a given threshold, whose description with a linear function seems rather counter-intuitive.

Besides, the authors address the usage of polynomials of higher order to account for more of the unexplained variance, while noting that R^2_{mod} (the variance explained) does not account for over-fitting. I don't know if there's a simple way to account for the number of parameters and identify the degree of the polynomial for which the reaction norm is best and most parsimoniously approximated, but if there is, I'm sure it would also be of great interest to potential readers of this article.

In addition to these remarks, here are some additional minor comments I have on the manuscript:

L 117: to get the message clearer, it might be useful to explain what ε_k from Eq. 3 is right there.

L.127: similarly, ε is not defined prior to this sentence, although it can be inferred from the context.

L.216: As Eq 18. is in a different section of the manuscript from Eq.9, I would recommend recalling the reader that $p(\varepsilon)$ denotes the probability density function of the environmental variable.

L.360: although I may understand why the first letters of "Performance Curve" are in upper case, I would rather suggest putting the name of the scenario in quotes. At least, this should be consistent throughout the manuscript (it is for instance also the case at L.417 but not at L.373)

L.400: the first letter of "Independence" is in upper case.

L.423: I guess that the part in brackets was not not updated to include the said references.