Dear PCI editors,

Please find attached the revision of our manuscript entitled “Partitioning the phenotypic variance of reaction norms”. First of all, apologies for the long time it took us to revise the manuscript. The reason is that we wanted to implement somewhat substantial changes suggested by the reviewers, which aligned with our own interest, but which we had omitted for simplicity in the first round. Most notably, we now derive the component of genetic variation caused by genetic variance in plasticity, which we then partition into contributions from different parameters of the reaction norm, using a formalism that applies to both the character state and curve parameter approach. We also introduce standardised unitless parameters, including different components of heritability that matter for the evolution of the trait or its plasticity. In addition, we have put together a now fully independent package (Reacnorm) and tutorial that allows users to apply the methods we highlight in different contexts, including in continuously distributed environments as found in natural populations (another addition to our manuscript). Overall, we are more satisfied by this version of the ms, and thank the reviewers for their constructive criticisms and comments.
We therefore hope you will find this ms suitable for recommendation in PCI, but will be happy to address any further suggestions for its improvement.

Sincerely,
Pierre de Villemereuil and Luis-Miguel Chevin

Reviewer 1: Jarrod Hadfield

In this paper the authors detail a new way of decomposing phenotypic variance into parts due to i) average plasticity, ii) genetic variation in reaction norms and iii) residual variation. As outlined in the introduction, a major motivation for this work is to develop metrics for comparing plasticity across traits. However, some form of standardisation is still required to achieve this as the metrics are in units of traits squared. Some discussion of alternative standardisations (for example, dividing by the mean squared or total variance) should be made.
We understand this comment by the reviewer. We have now introduced variance-standardised estimates of the variance components in Eqs. 8–11, which we use throughout the manuscript and in the figures.

This issue aside, I had mixed feelings about the novelty/utility of the approach. Component i) (V_{Plas}) is straightforward and sensible yet it is essentially an (adjusted) R^2 from the relevant fixed effect part of the model and so doesn’t, in my opinion, require a lengthy paper discussing it.
Similarly, metrics to distinguish between a curve-parameter function and a character state approach seemed a little ad-hoc given well-known tests such as likelihood-ratio tests or F-tests could be employed. Despite this lack of novelty, I do think the question is interesting, and I would advocate the approach if I was trying to characterise the environmental sensitivity of a trait.
We hope this criticism will be alleviated by the substantial additions we have made to the ms. We agree that the approach we propose can be implemented based on existing statistical methods, but our aim is to provide carefully designed and theoretically justified variance decomposition estimates from such existing models (in the same vein as multiple regression existed long before Lande & Arnold showed the meaning of β and ɣ for measurements of selection, keeping things in proportion).
Regarding V_{Plas} more specifically, note that it looses its direct connection with an adjusted R^2 in the non-linear case, because $Eg|\varepsilon(z)$ is then not equal to $f(\varepsilon,bar(\theta))$. Our aim was not to perform model selection between the curve-parameter and character-state approaches, but instead to gather them within a same framework, and use them in combination where relevant. Beyond this,

our interest is not so much in V_{Plas} per se (of which we now prefer its standardised counterpart), but in its decomposition in terms of components of reaction norm shape, which we hope people interested in plasticity will find useful. We now make these points clearer in the ms, and thank the reviewer for pointing to these lacks of clarity. Especially, we change the name of the R^2 quantity to M^2 to avoid any confusion about it being used for comparing goodness-of-fit or performing model selection.

When reading the introduction (for example, the main paragraph of P3) I thought the main innovation would be in developing methods for comparing genetic variation in plasticity. This is a difficult problem since plasticity is always in units that is a function of the trait and environment units, an issue the authors point out in the context of the study by Murren et al. (2014). However, component ii) (V_{Gen}) does not really characterise genetic variation in plasticity in my view: it is simply the average genetic variation within environments. The limitations of this are most obvious in the character state approach where the metric is completely insensitive to the genetic correlations between environments. Most people would consider the sign and magnitude of these correlations to completely represent the magnitude of genetic variation in plasticity (GxE).
We understand this comment, as V_{Gen} is computed as the average genetic variance across environments. It is the total genetic variation in the reaction norm, and as such, does not separate between average genetic variation in the trait and information about genetic variation in plasticity. Our main interest in defining V_{Gen} was to later partition it into different components of genetic variation in reaction norms. However we agree that it is more satisfying and enlightening to first isolate a component of genetic variation in plasticity (i.e., genotype-by-environment interactions), and then partition the latter into components due to different reaction norm parameters. We now show how to separate, for any kind of modelling approach, this genetic variance in the reaction norm (V_{Gen}), into the marginal genetic variance of the trait (V_G) and the genetic variance in plasticity (V_{GxE}). We believe these metrics reveal simple, and we believe illuminating, relationships with the basic parameters of reaction norms on one hand, and the influence of the environment on the other hand.

Attempts have been made to develop metrics for quantify GxE in discrete environments (most recently, Albecker et al. (2022)) and I wonder if these ideas could be fruitfully extended to reaction norm approaches?
Thanks for this reference. The approach that we developed to compute V_{GxE} has a superficial relationship with Albecker et al.'s ΔGxE, but appear to us as more general and more theoretically justified (most of all because it is based on a carefully laid out variance decomposition and reasoning about the genotypic and breeding values, as explained in our Appendix). A main difference, of course, is that V_{GxE} is a variance, while ΔGxE is based on absolute value. Nonetheless, we now cite this reference to shortly compare their approach to ours.

• L56: It's not clear why standard statistical methods cannot be used to assess fit in this context (e.g. AIC or likelihood-ratio tests).
We suggested a ratio akin to a R^2 to quantify how much of the phenotypic variance due to the average plasticity can be recovered by a given curve parameter approach. This is analogous to the other constructs of the manuscript, which are all expressed as fractions of total variance, but our aim was not to suggest that this metric should be used to select between alternative models, for which approaches such as AIC comparison or likelihood-ratio tests would be more appropriate. We now clearly warn the readers about this (l.492-495):

> It is important to note here that M^2_{Plas} is just a convenient way to quantify the amount of V_{Plas} explained by the parameters included in the curve parameter approach and should not be used to perform model

selection. Model selection is a complex matter and we refer the readers to published reviews on this subject (e.g. Johnson & Omland 2004, Tredennick et al. 2021).

• L116: ϵ_k is not defined but presumably stands for a (single) environmental variable.
Yes. We now write the equation without the k, and this notation is properly introduced when tackling the matter of the character-state approach (l.120).

• L118-119 'such approach' should read 'such an approach'.
Thanks, corrected

• L124 'function-values traits' should read 'function-valued traits'.
Thank you, this was corrected.

• L139 The law of total variance given is the one for conditioning on a single variable, not two. For two variables, it would be:
$V(\hat{z}) = E[Var(\hat{z}|\epsilon, g)] + E[Var(E[\hat{z}|\epsilon, g)]|\epsilon) + Var(E[\hat{z}|\epsilon])$
I guess the simplification to one variable comes about because $\hat{z}$, as opposed to z, is fully defined by ϵ and g such that $Var(\hat{z}|\epsilon, g) = 0$. Whether this is worth mentioning in the text, I'm not sure – it might unduly complicate the paper?
The reviewer is right in both statements. To remove ambiguity on the subject, we removed the reference to the law of total variance. We now justify our variance decomposition using a more explicit formalism as to what these variances are, notably in terms of genotypic or breeding values.

• L143 In the introduction it is pointed out that plasticity measures will have different units and scales depending on the trait and environment, and this makes it hard to make cross-trait comparisons. Based on the introduction, the reader is expecting this issue to be resolved, yet V_{Plas} is not unitless - it is in the units of the trait squared. What would the recommendation be? Standardise by the phenotypic variance or the mean or something else?
We had in mind the unitless π and γ metrics in this sentence, but but we have now opted for variance-standardised estimates for V_{Plas} and V_{gen} too for consistency.

• L145 I think it would be easier to simply refer to V_{Gen} as the average genetic variance within environments throughout the manuscript. While I think $V_\epsilon(E_{g|\epsilon}(\hat{z}))$ has the straightforward meaning attached to it, I think $E_\epsilon(V_{g|\epsilon}(\hat{z}))$ does not, as is pointed out briefly in the paper. Imagine the case where there is no plasticity whatsoever, only genetic variation (in the intercept). Then, $E_\epsilon(V_{g|\epsilon}(\hat{z}))$ is simply the additive genetic variance in a standard non-plasticity model and it would be odd to refer to this as the genetic variance in reaction norms or genetic variance in plasticity. Consequently, I don't think this would be my choice of metric for comparing, say, levels of genetic variance for plasticity across traits. Similarly, imagine a discrete character state model - the metric is insensitive to the genetic correlations between environments (i.e. Equation 17), despite this being the major signal of genetic variance in plasticity (i.e. GxE). In addition, in most cases (where the average reaction norm is modelled using fixed effects) V_{Plas} is simply the variance explained by the terms, which when scaled by the total variance, would be the R^2 of the model. I think this simple fact will be lost on the less mathematical readers and should be stated.
Thanks for this important comment, which should be largely addressed in our revised ms, where we now makes the distinction between the total (additive) genetic variance in the reaction norm, the marginal (additive) genetic variance in the trait and the (additive) genetic variance in plasticity. This distinction should clarify the biological meaning of each of these variances for the reader. Regarding V_{Plas}, see our response above.

• L184 This partition changes depending on arbitrary choices of what constitutes the reference environment, and in fact by chosing the reference environment to be the mean, this minimises the covariance and maxisises the variances (at least when ϵ is symetric, not sure otherwise). I think this is fine, but it would be good to explicitly state this.
Indeed, this is a common issue in the study of reaction norms (and probably the reason why mean-centring is widespread, as it makes for a natural reference environment most of the time). We now explicitly state so (l.260-263):

> Crucially, we chose to express this partition using the mean environment as the reference environment (as commonly practiced, e.g. Morrissey & Liefting 2016), but any other choice of a reference environment would result in a different π-partition of V_{Plas}, notably due to a non-null value for $Cov_{\epsilon}(\epsilon, \epsilon^2)$.

• L196 'includes all exponentiation levels (up to n) of the environmental variable ϵ' is perhaps better stated as 'includes the environmental variable ϵ taken to all powers from 0 to n
Thank you. This sentence was removed from the manuscript. It has an equivalent in the Appendix (l.1010), which we believe is now clearer.

• L203 Sometimes orthogonal polynomials (e.g. Legendre) are used in statistical analyses. This would get rid of these issues - would the authors recommend them?
True, but at the cost of loosing the "geometric" interpretation of the coefficients, as we now state in the ms (l.290-292):
"Using orthogonal polynomials would solve this issue of covariances, but at the cost of a more complex interpretation of the coefficients."

• L216 It may be worth noting that this is equivalent to Equation 11 in (Johnson 2014) although there the variance is conditional on observed values of ϵ rather than considering a probability distribution for ϵ. Note also, that Johnson's (2014) aim is to develop an R^2 metric for random-regression models. However, the result is inexact depending on how you want to define the R^2 because $E[V_{explained}]/E[V_{total}]$ (Johnson's (2014) metric where the expectation is over ϵ) is not equal to $E[V_{explained}/V_{total}]$ (see here).
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fmec.15394&file=mec15394-sup-0002-AppendixS2.html
Thanks for this comment, we now cite Johnson (2014) where suggested (l.494). We agree that the choice between ratio-of-average or average-of-ratio is complicated, and depends a lot on the biological question. In this precise case of computing the variance V_{Gen} and getting a variance-standardised estimate (now called H^2_{RN}), we decided to use the ratio-of-average, as it is the most common practice (as can be seen from Johnson, 2014). Note that this question becomes even more difficult when averaging over posterior distributions is involved in a Bayesian context, as the reviewer is aware.

• L266 Eq. 26 It would be nice to relate this to the adjusted-R^2 which also gets a 'variance explained' without the bias.
We now mention this (l.417).

• L296 It may be worth mentioning the V_{Tot} as calculated from the data and V_{Tot} as calculated from the model may be quite different if the random effect structure is highly unbalanced. Which do you authors think is most relevant? The latter, presumably?
Indeed, we would consider the latter to be the most relevant, following the practice in most estimations in quantitative genetics. We now state this in the ms (l.418-421):

> "[…] we advise using the sum of all estimated components rather the raw sample variance. The former is common practice in most quantitative genetics inference to account for potential imbalance in the experimental or sampling design (Wilson et al. 2010; de Villemereuil et al. 2018).

• L309 'we offer to rely' should read 'we often rely'.
Thanks, replaced simply by "we rely on…".

• L324 I think it is obvious that the character state approach will be unbiased and a misspecified curved-parameter approach will be biased. The reason for choosing a curved-parameter approach is not lack of bias, but because it has fewer parameters and so is more precisely estimated. I think it would be more interesting to show the mean-squared error in V_{Plas} and V_{gen}. However, my guess is that the relative magnitudes of the mean-squared error will favour the character state approach when sample sizes are large, as in these simulations.
In our view, the main reason for choosing a curve-parameter approach is that it yields more actionable information, as we now state more explicitly in the manuscript. To our surprise, however, the character-state approach has **almost exactly** the same sampling distribution as the curve parameter approach in our "perfect modelling" scenario (see Figure 3), meaning that both approaches have the same precision when the true reaction norm shape is well fitted. As a consequence, they share **almost exactly** the same MSE in this scenario:

Scenario	Source	VPlas	VGen	VTot
Nenv = 10, Ngen = 20, Nrep = 20	Character-State	0.0288	0.0138	0.0436
Nenv = 10, Ngen = 20, Nrep = 20	Curve-Parameter	0.0288	0.0137	0.0435
Nenv = 4, Ngen = 20, Nrep = 20	Character-State	0.0539	0.0273	0.0773
Nenv = 4, Ngen = 20, Nrep = 20	Curve-Parameter	0.0539	0.0273	0.0772
Nenv = 4, Ngen = 5, Nrep = 5	Character-State	0.24	0.135	0.333
Nenv = 4, Ngen = 5, Nrep = 5	Curve-Parameter	0.235	0.134	0.329

This means that even when the curve-parameter could have the advantage of being a perfect fit to the data, it does not (substantially) outperform the character-state approach. There would thus be little point in comparing MSE in contexts where the curve-parameter approach is not a perfect fit.

• L338 I would not use the word 'robust' here as it has a precise statistical meaning: perhaps use 'unbiased '
We changed for a less technical wording, using "widely applicable" rather than "robust" or "unbiased", neither of which would be technically correct here indeed (l.522).

• L341 I'm not convinced R^2_{mod} should be used to distinguish the models as it doesn't have known statistical properties. If a character state approach is set up for each unique value of ϵ then a curve-parameter model is nested within it, and a straight forward likelihood ratio test could be used. Alternatively, for the fixed effects, both the curve-parameter model and the character states can be fit simultaneously (although clearly some number of the character state coefficients will be aliased) and sequential F-tests used.
We understand this comment, but we never intended for R^2_{mod} to be a tool for model selection, for which we agree it would be poorly designed. We only use the character-state as a reference to quantify the amount of V_{Plas} missed by the chosen curve parameter approach. The user can compute the π metrics of their reaction norm curves, relating to their average slope/curvature, and at the same time quantify how close the estimated curve is to a character-state model that would capture all plasticity. To make sure this misunderstanding does not persist in the final text

(especially since this point was also raised by the second reviewer), we decided to remove all references to "goodness-of-fit" and we renamed R^2_{mod} as $M_{Plas}{}^2$.

It would be useful to state somewhere that the character state approach can always be recovered using a polynomial with sufficient order (de Jong 1995).
We already referred to de Jong (1995) when first mentioning the equivalence between the "parameter curve" and "character state" approaches, but we now cite it to state more explicitly that a polynomial of sufficient order is exactly equivalent to the character-state (l.143-145).

• L359 Here, or elsewhere, I think it would be good to restate the important result in de Villemereuil et al. (2016) that an additive genetic model on some scale produces non-additive genetic variance on the transformed scale when the transform is non-linear. Hence, the genetic variation on the transformed scale may need partitioning into additive and non-additive variance even when the model is purely additive on the original scale. If this is not restated, I think few readers will understand why the authors are having to distinguish between broad-sense and narrow sense genetic variance.
We now mention this matter much earlier in the manuscript (l.183-190), and explicitly added a V_{NonAdd} (see e.g. Eq. 7 and Figure 1) in our variance decomposition to highlight this matter. More generally, we decided to focus more on the **additive** genetic variances throughout the manuscript. A main reason is that working with the additive genetic variance requires to compute what we call in the article "reaction norm gradients", which provides a simplified and unifying formalism across models, including for linear models such as character-state and polynomial ones.

• L411 Eq 29 I've always found this way of looking at evolutionary change in phenotype when there is plasticity a little awkward. I prefer to think of a multivariate system with the phenotype z and the reaction norm parameters as traits. Evolutionary change is then determined by the genetic covariances between z and the reaction norm parameters (which are a function of ϵ). In the example here, the selection vector would then be zero except for the element pertaining to z, but in reality there is likely to be a cost to plasticity which can be easily accommodated by having other elements of the selection vector be non-zero. This approach produces a function $\Delta\bar{z}(\epsilon)$ which needs to be averaged over the distribution of ϵ and the γ's in this paper are essentially doing this averaging, I believe. However, if there is a cost to plasticity does the approach advocated here work, and what if there is environment specific selection (the main interest from a GxE perspective) - can the two β's be replaced with βi and βj respectively?
Thanks for this comment. The approach suggested by the reviewer (considering evolutionary change in reaction norm parameters as correlated responses to selection on the trait) should lead to the same selection response as the more usual approach of working directly at the level of the reaction norm parameters themselves. This also holds when there is a cost of plasticity, which simply adds another term to the selection gradient in the latter approach. Initially, our main interest was in responses to selection across environments, but our first explorations of this topic convinced us that it would require more thorough and careful investigation, worthy of another full manuscript. This is why we only touched on this topic in passing and in the discussion in this ms.

• L423 problems with reference formatting.
We fixed this issue, thank you.

• L474 should be 'de Jong' not 'De Jong'.
We fixed this issue, thank you.

• L435 I would argue that if you were interested in the genetic variation in plasticity, rather than genetic variation per se, then the approach advocated here would fail to provide insight. It is in fact hard for me to see how a (meta) analysis of genetic variation in plasticity would not at some point have to focus on the variance in reaction norm parameters.
Agreed, and we proposed the distinction between h^2 and h^2_I, as well as the π, φ, γ and ι metrics, to that intent. Note that, from an evolutionary perspective for the trait, V_{Add} and the γ-decomposition are also quite important.

• L441 To use numerical integration it is assumed the environment follows a known distribution, and the parameters of that distribution are known without error. I would think conditioning on the observed environmental variables in the data, as is typically done when calculating an R^2, would be simpler and more robust to miss-specification of the environmental distribution?
The Reacnorm package was coded so that the user has the choice between considering the environment "fixed" or assuming a distribution of choice. If the model is linear in its parameters, then it is possible to not use numerical integration in some instances, but numerical integration is necessary even if the environment is considered "fixed" if the model was not linear. We now discuss more carefully and in more details the implications of the distribution of the environment, especially when it is not normally distributed for the π-decomposition.

Reviewer 2: Thibaut Morel-Journel

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.
We thank the reviewer for his kind words about our manuscript.

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.
This was an oversight from our part, we apologise for the inconvenience. We thank the reviewer for having been thorough and looked into the online tutorial as well. Since we rewrote the code into a fully independent package named Reacnorm, we have now re-written the tutorial as a PDF vignette for the package. We are afraid that the tutorial will require users to run the models on their machine, but the upside is that the package does now most of the heavy lifting after running the models.

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.

Thanks for these suggestions. We have now included simulations using a lower sample size (4 environments, 5 genotypes and 5 repeats per genotype, resulting in 100 data points). The precision of the estimates was of course affected, although to a very reasonable degree (estimates were “only” twice as imprecise in these degraded conditions).

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.

The manuscript now tackles the subject of the distribution of the environment in more details, especially in relation with the applicability of the π-decomposition when the environment is not normally distributed. In particular, we now explicitly distinguish between the π-decomposition explicitly based on the slope/curvature (π_Sl and π_Cv, accessible only if the environment is normally distributed, and/or the reaction norm is well-approximated by a quadratic function), and the less interpretable ω-decomposition based on the parameters (ω _a and ω _b). This point made us realised that continuous environments were too much of second-class citizens in the manuscript, and they are now fully addressed (e.g. the non-linear simulation scenario was switched to a continuous environment case).

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)?

We now tackle this scenario explicitly in the non-linear simulation scenario, where unique environmental values are drawn from a normal distribution, resulting in exactly this kind of scenario. We show that our framework works well in this case (it is possibly even one of the easiest scenario due to the nice properties of the normal distribution described in the manuscript). However, studying further the limits of various sampling designs would rather start testing the properties of (non-)linear mixed modelling which is beyond the scope of our paper.

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.

The sigmoid reaction norm was used as an illustration of our framework, rather than to advocate approximating it as linear (if anything, we would advocate using an explicit sigmoid function in that case, as we do later in the paper). But yes, another choice of parameters or environmental distribution might have found higher curvature in the reaction norm, when fitting a quadratic polynomial. That the sigmoid lacks curvature in our example occurs because (i) it is perfectly symmetrical with regard to the inflexion point and (ii) the mean environment happens to be at the inflexion point (as would be the case even for a step function). A third-order effect certainly improve the fit. We now make all these points clear in the ms.

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.
There are indeed multiple ways to perform model selection to assess the best order of a polynomial function to fit a dataset, from nested likelihood-ratio tests to information criteria such as AIC. However this metric was not introduced for model selection purposes, but only to quantify how much of the total variance in plasticity is captured by a chosen reaction norm model (see our response to a similar point by Reviewer 1 above).

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.
Thank you for noticing this oversight. We changed how k is used throughout the manuscript: it now only serves for the character-state model and is properly introduced, see the response to reviewer #1.

L.127: similarly, ε is not defined prior to this sentence, although it can be inferred from the context.
Thanks again. The variable ε is now properly introduced (l.103).

L.216: As Eq 18. is in a different section of the manuscript from Eq.9, I would recommend recalling the reader that p(ε) denotes the probability density function of the environmental variable.
For the sake of simplicity, we decided to remove integrals, and only use expected value and variance instead.

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)
The styling for the scenario names has been changed and is now more consistent, thanks for pointing this out.

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

L.423: I guess that the part in brackets was not not updated to include the said references.
Thanks, the references were correctly added.