

# Fugitive road dust alters annual plant physiology but perennial grass growth appears resistant

## Response to Review

We appreciate the opportunity to continue the revision process for our submission to *Plant Ecology*. Please find below responses addressed to each reviewer in turn.

### Reviewer 1

We appreciate the comments on the additions to the ecological context.

Each of the line comments have been corrected as the reviewer suggested.

We appreciate the reviewer's interest in properly reporting results. Indeed, we share the critique of p-values and null hypothesis testing, which in fact motivated our use of confidence intervals. In light of the reviewer's confusion about how CIs are presented and discussed, we've edited for clarity.

### Reviewer 2

In the period since receiving this review we've consulted extensively with colleagues familiar with mixed-effect regression models and trained specifically in the use of the *lme4* package. We appreciate the scrutiny the reviewer had given our models, and concede that steps could be taken to more precisely model our data, especially with respect to variance structures in the random effects term. The reviewer asks frequently why we did not do  $x$ ,  $y$ , or  $z$ , which is a difficult position from which to respond. Indeed, there are likely several other approaches, and as we conceded above, some might be more precise. At the same time, based on our conversations with even more statistically-minded colleagues and our own scrutiny of the analysis as prompted by the reviewer, we have confidence in the framework that we developed and have described it in the best detail we can, hopefully with improved clarity following the reviewer's comments. In this revision, we stick to describing what we did and why, and not exhausting all other options.

We've structured our response by the themes addressed in the review:

**Random effect terms** The reviewer has provided a thorough tutorial on building robust random effects terms. In our paper, we state that we developed our random effect terms to encompass known sources of potential random variation. The reviewer appears to at least agree with us that the relevant sources have been identified. In conducting these analyses, I have found that different formats of random effect construction do change how variation is allocated among the included terms, but these variations do not affect the amount of variation

In my experience with mixed-effect models, I am most interested in ensuring that random variation is accounted for, but not parsing the error structure within the random effect. I believe this applies here—we are not interested in which of the terms contribute more or less to the overall variation, just that the overall variation is accounted for. If we were in fact

interested in parsing the error structure, we would need to give much greater scrutiny to how the random error is structured, not just that it has been accounted for. This explains our use of the random effect term given in our script: It appropriately accounts for random variation contributed by the modeled sources, but does not provide the unnecessary information about which term contributed what proportion of that variation.

The reviewer suggests accounting for repeated measures by adding a random intercept term, which is quite different from my own training and my conversations with other users of *lme4*. Our understanding is that it is sufficient for the objectives of this study to model repeated measures by nesting the temporal term within the spatial term, an approach proven adequate in previous applications.

**Fixed effects** The reviewer asked why we fit and subsequently compare intercept-only ( $y\tilde{1}$ ) models, and what these models represent. This is how we were trained to build candidate model sets for model selection and averaging. If  $y\tilde{x}$  has an acceptable lower AIC value than  $y\tilde{1}$ , then one can assume the independent variable has a meaningful relationship with the dependent variable and should thus be included in the model. One interprets no difference between the model with the term and the null model with the intercept alone as an indication that the term is not meaningful in explaining variation in the response variable. This is a very common convention, especially when users of *lme4* wish to calculate a  $P$  value despite the package's default not to do so.

The reviewer was absolutely correct in questioning why models with and without the intercept term were included in candidate model sets. Prompted by the reviewer's attention to this detail we found a more parsimonious way to retrieve the desired information and have as such revised the model sets in question to be internally consistent in whether they include an Intercept term.

As for model comparisons and the results given in Tables 2 & 3, based on the high degree of variability in the raw data over the repeated measurements, we believe it is informative for readers to see how models that consider single terms, multiple additive terms, and multiple terms combined with interactive effects compare against each other. Readers should know when main effects are tempered by significant interactions, each of which we have considered as to whether or not they negate main effects, or simply condition them (e.g., flipped signs in responses between two species vs. moderated magnitude). The reviewer is correct in suggesting that folks most want to know about the main effects of dust, which we emphasize in the confidence interval graphs but substantiate with the model comparisons results because there is clearly variability at the species level.

While on the topic of model selection, models were fit with Maximum Likelihood rather than REML because functions in the *AICcmodavg* package convert *lme4* objects to ML anyway to facilitate model comparison, specifically warning against REML.

The reviewer mentioned:

This suggests that the comparisons made are between the following two models:  
 $scale(lconc) \ 0 + species + dust + species : dust + (1|block : date : pot)$  and  
 $scale(lconc) \ spp + (1|block : date : pot)$

However, the correct comparison should have been as follows:

$scale(lconc) \ 0 + species + dust + (1|block : date : pot)$  and  
 $scale(lconc) \ 0 + species + (1|block : date : pot)$

And, the result of this comparison reveals the effect of dust (NOT species).

The reviewer is correct—a comparison of these two models returns the significance of the dust term (if the model with the dust term has substantially lower AIC, the dust term is meaningful and there is a general effect of foliar dust on the response variable).

We go beyond simply testing the dust effect. We do not extract regression coefficients from the species-only model, but from the *species*  $\times$  *dust* interaction, which with the intercept removed provides us species-level estimates for the effect of dust *on each species*. This is meaningful and appropriate given how frequently the *species*  $\times$  *dust* interaction ranks competitively in the model selection.

We understand that this is different than conventional null hypothesis testing and even alternative model comparisons, but we believe it is more informative especially with so much species-level variability in response to dust. Furthermore, based on our conversations with several colleagues, we are confident that even though this might not be the conventional approach (and clearly isn't how the reviewer would proceed), it is legitimate.

We've declined to include residual plots for all models. While we appreciate (and share) the reviewer's interest in having more folks "understand importance of checking residuals whenever running data analysis using statistical models" we don't believe filling our Supplementary Information with plots will advance this goal. We've stated in the Methods section that we confirmed all models fir the assumptions of the distributions with which they were fit.

**Etc.** Regarding axis and legends for SI Fig. 2: We see the reviewer's issue with the "bottom" of the figure—this last set of panes shows specific leaf area, which is unique from the others in that it is a destructive measurement taken only once at the end of the study. As the reviewer observed, it is not at all well-explained by the legend information given for the other figures. We apologize for this oversight and have corrected it accordingly.

**Discussion** The reviewer is absolutely right about calling attention to  $\frac{F_V}{F_M} < 0.8$ , and we apologize for overlooking this in response to the previous review. We've made reference to this condition twice in the Discussion when interpreting possible environmental impacts when compared to a study such as ours conducted in the greenhouse.

We appreciate very much the suggestion to incorporate dust effects on herbivory, and have made explicit reference to this in the Discussion.