Fugitive road dust alters annual plant physiology but perennial grass growth appears resistant
Response to Review

We appreciate the opportunity to revise our submission to *Plant Ecology*. Each reviewer clearly gave considerable thought to our manuscript, and we are certain the paper has been improved through the incorporation of their helpful comments.

Major points of revision include:

- More references to ecological impacts of dust throughout

- A specific paragraph at the end of the Discussion delineating three scales of dust impacts relevant to plant ecology

- Clarification on the statistical framework and justification for the modelling approach

- Revision of random effects term to include pot term to account for repeated measures

- Edits & additional materials in the Supporting Information

We trust the editor will find these revisions address the important reviewer concerns. Responses to each reviewer comment follow inline below, as colored text.

Reviewer 1

This MS describes a greenhouse study of physiological and growth responses of several plant species to application of dust from unpaved roadways. The study was prompted by development of the Bakken oil fields in North Dakota, which involved heavy truck traffic on unpaved roads, generation of substantial dust, and heavy dust loads on nearby plants. The authors chose several common annual crop plants and perennial forage grass species, and developed a protocol to apply road dust in controlled greenhouse conditions, with appropriate dust-free controls. Based on previous studies they anticipated a number of changes to leaf gas exchange, temperature, and photosynthesis of the crop species. The perennial grasses, in turn were clipped to simulate grazing, and the authors anticipated that dust would inhibit the ability to tolerate this damage through recovery of lost biomass. Some of the results were as expected, and others were surprises; for instance, increased photosynthetic activity in annual crops and striking tolerance of the grasses to defoliation even when dusty.

We appreciate the accurate synopsis, indicating that the reviewer engaged the paper exactly as we intended.

The MS is written clearly and efficiently, and the results are valuableperhaps especially the results that surprise rather than confirming previous published studiesin adding to our understanding of whole-organism dust effects. This seems to me to be very timely given anthropogenic increases in dust generation (e.g., reviewed by Field et al. Front Ecol Environ 2010; 8(8): 423-430, doi:10.1890/090050).

We appreciate the support, as well as the Field et al. reference. This one had slipped past us, and we've included it here. If the reviewer has not done so lately, he ought to forward-search this citation in Google Scholar. I expect he will share our surprise of how few of the 260-some

My major concern is that whereas the journal focus is ecology, neither the ecological context of the author's work, nor the possible ecological implications, are clear. By context I mean that an ecologist reading the MS does not immediately form a picture of the extent of the croplands or grasslands involved, or of any of their natural history and functioning. By implications I think of how leaf physiology and regrowth might scale to plant demography and plant community change (or lack of change).

We appreciate the reviewer's concern and trust he will find that additional literature review addresses this point; see more discussion below. To address the extent of rangelands and croplands impacted, we've added a reference to the footprint calculated by Allred et al. (2015).

On the final page of the MS the authors clearly acknowledge at least the second of these limitations. The greenhouse study was of relatively short duration (10-14 days); the authors are aware that initial and even final responses that they observe might not scale up to whole-plant life-cycle responses that influence population or community ecology. Likewise they are aware that plants in the field interact with factors other than dust. Whereas they posit that dust and other insults (e.g., herbivory) might reinforce each other, in fact the interactions could be even more complex (example: in our recent studies with a native wildflower we find that road dust deters a major insect seed predator, even though it also reduces receipt of pollen. These effects seem to counterbalance; dusty plants do not suffer lower seed yields than controls, which the seed predator attacks more heavily).

Can the authors strengthen the ecology here, by adding more on the ecological context, and perhaps by suggesting a bit more completely what the ecological implications might be, in part through some suggestions as to what studies ought to be done next? Or, would a more physiological journal be a better outlet and audience for this interesting work?

We very much appreciate this comment, especially the suggestions of how to take the discussion beyond our data. We've addressed this at two scales: Firstly, we've incorporated broader ecological literature in several places of both the Introduction and Discussion. Secondly, we've added a final paragraph that addresses the ecological context specifically, in which we delineate three scales at which we expect dust to have ecological impacts and clearly identify the scales to which our results apply. We hope that this delineation both places our results in a more clear ecological context and helps describe how the basic results shown here might be expanded upon to develop greater ecological understanding, ideally at multiple scales.

Minor comment: there is a slight typo on line 220 (a spurious "70").

Thank you, we've removed it.

Also, are more key words allowed? More might help to flag this for intended readers.

Good point, we've added keywords.

I am going to upload our first paper on road dust and wildflower pollination success and reproduction in case it is of interest; the more recent finding of deterrence of the seed predator is in manuscript form almost ready to submit for publication.

We've incorporated the reference. Its omission in the first draft was an oversight on my part. We've discussed the paper since it was first published and it was meant to be included, but you know how it goes with Master's students sometimes...

Reviewer 2

In this study, the authors examined physiological responses of seven species of annual crop plants and accumulation of aboveground biomass in eight species of perennial grasses after exposing the plants to the road dust. Negative effect of the dust was not found in either annual crop plants or perennial grasses.

We appreciate the accurate synopsis.

I am writing this set of comments after reading the entire manuscript once. I decided against reading the manuscript over because I felt that it might benefit greatly from a revision and reanalysis of the data. After making that decision, I did read the methods section and supplementary information several times until I felt that I had a reasonable understanding of what had been done.

Thank you for the detailed description of your review process.

A big concerns is that the methods section is not written precisely and completely enough, and consequently, I am having a difficult time understanding exactly what has been done.

We apologize for the lack of precision and completeness. We trust the following comments assist the reviewer in having a less difficult time understanding exactly what has been done.

Would you please spell out the physical layout of the pots on each of the two benches, for example?

As for the physical layout, the pots were placed beside one another on benches. It is difficult to determine what is unclear about this and what additional information is necessary or even possible to expound upon.

Here is my understanding from the manuscript: For the perennial grasses, 8 species x 2 trt x 33 replicates = 528 pots.

As the manuscript clearly states, "We used 32 pots of each species, 16 in both the dusted and undusted treatments." $32 \cdot 8 = 256$. We've changed this to say "each of" instead of "both"; hopefully this adds clarity.

The pots were grouped in sixes. Therefore, $528/6 = 88$ groups. There were 44 groups on each bench. From Fig 1a, I guessed that not all 6 pots in a group are of the same species. Am I correct? There are eight species, but there are only six pots in a group. This means that different groups of six pots have different combination of species. Am I correct?

As we described, pots were randomly assigned to the six-pot groups. We've added another sentence that specifically states clusters were of mixed species.

For the annual species, 7 species x 2 trt x 16 replicates = 224 pots. The pots were grouped in sixes. Therefore, $224/6 = 37.33333$ This probably means that some groups have had less than six pots. Also, as with the perennials, different groups of six pots have different combination of species. Am I correct?

We don't follow the reviewer's math that results in fractions. All groups had six pots.

The authors may ask why I am so concerned about the layout of pots. The answer is that I think that the physical layout of pots (= experimental design) determined the statistical model to be used in the analysis. I feel that the models used in the analysis by the authors are not exactly appropriate. I am not sure how inappropriate they are (or whether the conclusions may change), but please see below for my issues.

The reviewer is correct. We did wonder why they were so concerned about the layout of the pots, and we appreciate the explanation. We remain confused about the threshold for ''exactly appropriate'' and do not believe such a standard it relevant or achievable.

In this type of study, I would have used a split-plot design with 1 pot x 8 species in each group (= subplot). To enable this, I would have made the tents that would be big enough for eight pots. In this way, the analysis would have been a lot simpler and more robust, I think.

We appreciate the reviewer's insight and suggestions for future work. We are confident the reviewer understands the constraints of pot sizes, bench sizes, and the limited sizing options in the marketplace for tents that fit on bench tops and also meet other criteria for successfully completing the experiment.

However, the experiment was done differently, and we need to find a good way to analyze the data. I feel that the models that the authors used have two types of weaknesses in particular. First one is simple: the authors need to use models for repeated measurements because measurements were taken repeated on the same set of plants. At least in theory, this problem is relatively easy to fix, and the authors may already know how to. For example, for a given species, concentration ˜ 0 + date + (date—pot) Of course, we will have to include the unnecessary fixed effect of the date (of measurements). I understand that the authors have tried to remove the effect of the date by including the "round" or "date" as random effects, but this does not account for the temporal autocorrelation of the measurements. The models that the authors used treat measurements on different dates as measurements taken on different plants. As such, these models are not appropriate for the data.

The reviewer is correct about the issue of repeated measures. It was an oversight on our part to have not included the pot term in our random effects. We have updated all models, which had the effect of returning more interesting results for the immediate effect of Photosynthetic Yield, so we appreciate the reviewer's attention to ensuring that the statistical models best represent trends in the data.

Second problem is the inappropriate way to attempt to account for the spatial autocorrelation in the data. The authors used the "block (= bench)" as a random effect in an attempt to deal with some spatial autocorrelation. However, the physical layout of the pots is such that including the bench as a random effect may not be effective in taking out the spatial autocorrelation. I wonder if the authors have checked to see if the "block" was really needed in the model. One way of dealing with this problem relatively effectively is to specify the physical position of each pot in the greenhouse, using the x- and y-coordinate. If there was the record of position of each pot, the authors just need to include the x-, y-, and xy-position of each pot as random effects. I would probably try including the "group (of six pots that are fitted under a tent)" as another random effect to see if it would have a significant effect on the model. Including these terms in the model should, in theory, help dealing with the fact that different groups had different combination of species. All of these seem easy in theory, but including many random effects may destabilize the model. My personal approach is to use a robust design so that I do not have to rely on complex mixed models. Please consider taking this approach in the next experiment.

We appreciate the reviewer's suggestion for future experiments. There is no evidence of spatial autocorrelation in the data. The greenhouse room was as homogenous of a growing environment a plant ecologist could ask for, and all pots were randomly assigned and two rounds were conducted for each trial.

If trying to deal with the temporal and spatial autocorrelations properly in the model makes the model too unstable, then I would analyze the data for one sampling date at a time. If trying to deal with the spatial autocorrelation properly using the x-, y-, and xy-position of

each pot proves to be too much for the model, then try using the x-, y-, and xy-position of each group next. If this was also too much for the model, then try using the group and bench as random factors.

Again, we see no evidence that some sort of spatial autocorrelation occurred on the greenhouse benches and have never heard of such a spatial analysis of pots on benches.

The second example of the description needing to be precise and complete is the use of the term "overall effect". I think that the authors use the term "overall effect" to describe two different things. In the first way, authors use the term to mean the overall effect across all seven species of annual plant species. In the second way, the term is used to mean the overall effect across all seven species of annual plant species and over all four physiological responses. Please distinguish them appropriately.

It isn't clear what the reviewer finds confusing here. Specifically, we don't see two uses. Figs. 2 & 3 clearly only use overall once, for the response of each trait across all species, yes. If this is in fact an issue, we'd appreciate specific line numbers for confusing instances.

Here is the third example of the description needing to be precise and complete. It is not clear how the dust effect has been calculated. For the immediate effect, it is explained that the dust effect to be the difference between the measurements taken before and after each dusting event (lines 149-150). In lines 153-155, the authors explain that to estimate the overall dust effect, the "response" was put into the model as the fixed effect. According to the supplementary information, the model in concern is as follows: diff ~ 0 + response + (1—block:round:spp). I have never analyzed data in the way that the authors have, so it is possible that I am confused due to lack of my understanding. However, I will explain what I can figure out. The "diff" is the difference between the pre- and post-dusting measurements, as explained in lines 149-150. The estimated coefficient for the fixed effect (the "response") is used as the overall dust effect. BUT, what is this "response" term in the fixed effect? Please explain. [By the way, why is the "species" treated as a random effect? How can this be justified? I do not think that one can justify this.] I think that a part of my confusion is that the authors use the term "response" to mean different things in one paragraph. In line 151, "each response" is probably the "each response variable" (such as chlorophyll concentration). In line 152, "dust response" is probably the "dust effect". I am less certain of the true meaning of the "response" in "we fit response as a fixed effect" (line 154), but I feel that the "response" is meant to be the "response variables (= the four physiological variables)". It may sound odd to include the response variables as a fixed effect, but my guess is that the authors used the names (rather than measurements) of the response variables as the fixed effect. If I have described in this paragraph is correct so far, then I think that the authors' description is misleading. What is described in lines 153-155 (the model diff ~ 0 + response + (1—block:round:spp)) has nothing to do with what is described in lines 155-156 (the estimate of the fixed effect). The overall effect across all species seems to be estimated using the model: scale(response variable) ~ 0 + spp + treatment + (1—block:date). Is this what has been done? If not, the annotation on the R script needs improvement. If yes, then I am not sure how the above model can be used to estimate the overall dust effect for a given response variable. This leads me to conclude that my understanding of what has been done is not sufficiently precise and complete enough for me, and I think that most readers of any scientific papers will not spend as much time and energy to try to understand a paper as I, and most readers will finish reading this paper without understanding the experimental methods or data analysis. So, the authors have the responsibility to write clearly, precisely, and completely.

The reviewer is correct, much of this confusion probably relates to not having analyzed data in this manner. We used here a proven statistical framework developed for plant traits and

This is a minor point, but please explain what scale(response variable) means. I am assuming that a response variable was standardized. Please describe exactly how it was done in the M & M section.

There are a few other (but still important) points. The authors mentioned that the response variables did not deviate from normality. So, the authors used LMMs instead of GLMMs, which was good. What linear models (i.e., LM, GLM, LMM, and GLMM) require is the normality of residuals. To check whether the residual are more or less normally distributed in R, please do the model checking: For example, One will get four graphs, and by examining the graphs, one can get a feel for whether the assumption of the model is met. If the residuals are not normally distributed, then the results of the model may not be trusted.

Throughout the analysis of the short-term responses, the authors used the model "diff0" to estimate the overall responses and the second model "diff1" to estimate the species specific responses. Am I correct? I think that more appropriate approach would be to test significance of the "species" effect using something like Another (and equally valid) approach is to assume from the beginning that species are different, and do not look at the overall responses at all. Please take one of these approaches and do not look at both species specific responses and overall response automatically.

Maybe I belong to an old school, but I feel uncomfortable seeing the AIC-based comparisons of models with different fixed effects. I understand that Bates does not like to calculate p-values for significance of fixed effects in mixed models. I do respect his opinion, but I still feel that AIC-based comparisons do have weaknesses of its own. I would like to see the following as a comparison. mod1mod2mod3anova(mod2, mod3) anova(mod1, mod2) Assuming that mod2 is the best model of the three, then finally anova(mod2)

Please apply the approach described in the paragraph above to the analyses of SLW and biomass.

Would you please add more explanations to figure captions in the supplementary information? It took me a long time to figure out what different colors in these figures meant. Also, what is the color difference in Fig 2 (bottom) in supplementary information?

In Fig. 1 (bottom) in the supplementary material, the caption says "photosynthetic yield", and the title for the y-axis is deltaF/Fm. In line 123, it says quantum yield (Fv/Fm). What is the difference? If the variable in Fig. 1 (bottom) is not different from Fv/Fm, then the plants are stressed (meaning Fv/Fm ¡ 0.8). What might be the reason for the stress and what effect the stress might have on the results of other response variables (and overall conclusions of the study)?

The terms are synonymous. We've clarified this in the expanded caption explanations.

I feel that the discussion section can be improved, but I do not want to start making comments until issues to do with the analysis and results are all sorted.

Improvements to the Discussion have been made in response to Reviewer 1's comments.

This is an aside: I noticed that most models did not have the treatment (dust vs. no dust) as a fixed effect, and this puzzled me. After staring at figures in the supplementary information for some time, I understood the reason. For a given date of measurements, the physiological measurements were taken only once on plants in the "no dust" treatment, while the measurements were taken twice on plants in the "dust" treatment (i.e., before and after the application of the dust). When analyzing the data, authors probably realized that they should have taken the measurements twice on all plants.

This is also an aside: For measuring SLW, I personally think it is better to weigh leaf discs. Measuring leaf discs is simpler, more efficient and cost effective than measuring whole leaves. Moreover, measuring whole leaves have an additional issue of dealing with differences in the proportion of leaves occupied by major veins. Please consider this in the future.

I look forward to going over the revised manuscript. Sincerely, Mamoru Matsuki

We appreciate your close attention to our work and your helpful comments and suggestions.