Dear Dr. Adler:

Thanks for reviewing this paper and providing thoughtful comments. We also thank the two independent reviewers for their time and insight. We will respond to comments in-line using indentation, immediately after the questions/comments/concerns are raised.

Editors’ comments:

In my view, the best feature of this study is the design. I really like the space vs time comparison, and the selection of sites across orthogonal precipitation and temperature gradients. But my biggest disappointment is that the analysis does not seem to take advantage of the design as well as it could or should. As far as I can tell, all transplant turfs are treated equally in the analysis, ignoring the fact that some turfs have been transplanted into sites with very different climate, while other turfs have been transplanted into sites with climates fairly similar to their origin. It seems like it should be possible to take the climate distance between source and transplant site into account, potentially giving your conservative analysis more power. For example, the magnitude of shifts in a CWM trait should correlate with the climate distance. I am not sure how the null expectations should be incorporated, but I will leave that problem to you to solve (perhaps you could analyze deviations of observed shifts from the null expectations?). You would also need to think about how to incorporate Reviewer 2’s comments about climate means vs. weather during the period of the experiment.

Incorporating climate distance in the analyses would be an excellent idea; unfortunately, the design does not include variation in climate distance. That is, the sites were selected so that they differ consistently in temperature and/or precipitation (Figure 1, Table S2) and transplants were always moved only a single “step” in temperature and/or precipitation. Thus, all turfs were either not moved at all or moved to a site that was approximately 3C warmer in mean summer temperature and/or ~800mm/year wetter. Therefore, explicitly addressing climate distance may not add much to the study. Please let us know if we are misinterpreting your comment.

We address Reviewer 2’s comments about climate means vs. weather below.

The approach you use to compare spatial vs temporal shifts in CWM traits could also be used to compare shifts in abundances of important taxa. Do you have a separate manuscript focused on those species-level results? If not, could I talk you into including that as part of this paper?

Several of the authors on this manuscript are indeed writing a separate manuscript on species-level results, with a particular focus on colonization/extinction dynamics. That said, your comment brings up an important question: are the shifts in CWMs in transplanted turfs driven by responses of a few abundant taxa, or many small shifts in abundance across many taxa? To answer this question, we have added a new figure (S4) to the supplementary material that shows how deviation from null expectations varies by species ordered along trait axes. Instead of observing just one or a few large deviations, we see moderate deviations in many species. Evidently, the shifts we observed in our transplanted turfs are driven by small changes in many taxa. Does this at least partially address your concern/curiosity of species-level responses?

Finally, you conclude that “the strengths of trait patterns along gradients are not necessarily proportional to their influence on community response to rapid climate change.” Following up on one of Reviewer 1’s questions, at what time scale would you expect the temporal response to match the spatial response? As you point out in the Introduction, spatial and temporal patterns reflect processes operating at different times scales. I would be interested in some Discussion material about which of these processes and time scales are most relevant to understanding responses to climate change.

This comment was very helpful. We have changed our language to reflect the fact that we can’t directly assess whether community response to climate change was strictly proportional to the strengths of spatial trait gradient patterns. Instead, we now more accurately state that we simply determine whether traits with significant spatial trends also show significant deviations from null expectations of community response to climate change over time.

Although we don’t feel confident in using our results to predict specific time frames for community response to climate change, we did add a detailed discussion of the processes we think will be important for community response to climate change in the real world (Line 376 – Line 384). We hope our focus on processes addresses your core concern.

Details:

Line 193: you mention a weighted multiple linear regression. What are the weights? Where do they come from?

Thank you for identifying this omission. The multiple linear regression was weighted by species abundance. The text has been updated.

Line 215: Is the “site species pool” a list of species (which is what I think of when I read “species pool”) or does it also contain information on # of individuals or percent cover (a list of species and their abundances)? I think you mean the latter, in which case I would use a different term, perhaps “site-level community.”

You are correct. We have applied your suggestion throughout the manuscript.

Line 226: Please provide computer code. Your description of the model is pretty good, but I would have much more confidence that I (and readers) really understand what you are doing if you provide the code.

We \*thought\* we did include the R code as supplementary material with the manuscript submission. Regardless, we will be sure to include R code for all analyses and figures in our resubmission.

Line 242: I think your definition of replacement rate ignores self-replacement. For example, if cover of species i does not change, it could reflect zero replacement (your assumption), or it could reflect the loss of a number of ramets balanced by recruitment of an equal number of ramets. By ignoring self-replacements, you might be underestimating replacement rate, and that could influence your estimate of the immigration rate (which depends on the replacement rate). I am not sure if this is a serious problem or not, but you need to address it.

You are correct, our definition of replacement rate ignores self-replacement, which likely leads to an underestimate. We updated the text to explicitly state this limitation (Line 244). However, underestimates in replacement rate do not appear to qualitatively affect our results. More specifically, it is visually evident in Figure S3 that vertical shifts upward in the positions of the black dots (i.e. increases in replacement rates) would not lead to significantly different deviations from observed data. We included this point in the manuscript (Line 245). We also feel that self-replacement is unlikely to be that common in experimental turfs. Species richness in turfs ranges from 10-40, thus any individual would be more likely to be replaced by an individual from another species.

It is also important to note that our Bayesian method of estimating immigration rate is actually independent from replacement rate; it depends on the relative abundances of species in the turf vs. site communities. This lack of dependency means that any potential underestimate in replacement rate would not indirectly affect model behavior through the immigration (*m*) (Line 247).

Line 264: It is OK to leave the details to the Appendix, but at least give a brief description of the method, and how it differs from you first approach, here.

After careful consideration we excised this second method of estimating immigration rate from the manuscript. It was our first attempt at estimating immigration and was less sensible and had more dubious assumptions than the Bayesian method we eventually settled on. Furthermore, model simulations from this second method were not qualitatively different from the Bayesian model. Thus, we felt that it made the description of our analysis more complicated (and longer) without adding any value or additional insights.

Line 280 and Fig. 3: I spent quite some time trying to understand exactly what is shown on Fig. 3. Formulas showing exactly what is on the x and y axes would help (these could go in Methods). For example, is the y-axis a comparison of 2013 to 2009 turfs, or a comparison of control vs transplant turfs in dissimilarity in 2009 vs control vs transplant turfs in dissimilarity in 2013?

Thanks for bringing this up - your second guess is correct. We have tried to clarify. First, we changed the axes labels in the figure (x-axis: Initial Dissimilarity to Controls 2009; y-axis: Change in Dissimilarity to Controls 2013). We also added this sentence to the figure legend: “The x-axis shows the Bray-Curtis dissimilarity between turfs and the centroid of control turfs in 2009, and the y-axis shows the shift observed in that dissimilarity by 2013.”

Line 313 and 314: Your description of “shifts in trait values” and “Increases in leaf area and SLA” imply that the trait values are changing. But if I understand your approach, you are assuming one, fixed value of each trait per species (no site specific trait data). If that is correct, you need to be much clearer that shifts in species composition are causing increases in CWM of leaf area and SLA.

Yes, you are right. We changed our language to emphasize this (Line 83, Line 189).

Line 345: “consistently influenced” implies causality, which you have not demonstrated. Please replace with wording like “…consistently associated with…”

Agreed. The language has been changed here and elsewhere.

Line 363: If the number of ramets (tillers) of grasses not meaningful? Or just not practical (too hard to measure)? I would argue the latter. I could do a lot with tiller counts!

Agreed – we also could do a lot with tiller counts! We now focus on its impracticality (Line 357)

Liner 371: This is a minor comment, but why use the word “mechanistic” to describe your approach? I would just remove that word here and in the abstract—it is not necessary and some readers, who think of mechanism differently, might be put off. If you really want to keep this, you need to explain what your model is mechanistic with respect to.

Agreed. The word ‘mechanistic’ is not necessary and was removed throughout the manuscript.

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

This manuscript uses a novel statistical technique to examine the potential for existing spatial gradients in plant functional traits to predict the success of taxa transplanted into novel environments. The authors transplant entire communities (small turfs) into new locations within an environmental gradient, as a proxy for understanding future response to climate change. They then examine how traits which covary with spatial environmental gradients are predictive of species performance in the new environment. I thought it was an interesting study, of broad relevance to the readership of Ecology. The manuscript is well written. I had some specific questions about the analyses, which are outlined below.

Line 164. The authors state 'Eight species were identifiable only to genus but treated as species in downstream analyses'. Question: how much total cover did these species represent and which constituent species traits were used for them? An average of the species within the genus?

These species represented only 3.8% of local cover. Leaf traits were taken from our field measurements; seed mass and maximum potential height were estimated by referring to locally-present congeners and using the median trait value. Clonal traits were left blank unless all locally-present congeners exhibited the same categorical trait values. The text has been updated to reflect this.

Line 168: Clarify number of offspring per parent - do you mean number of offspring per parent \*per year\*?

Yes. The text has been updated.

Line 181: Clarify: Species-level and community-level trait correlations were relatively low (Appendix A) and thus we treated traits as independent. What does this mean?

It means that trait-based responses (e.g. panels in Figures 2-4) can be interpreted as independent results, rather than a result of covariation of traits. It should be noted that we removed Table S2 (community-level trait correlations) from the manuscript because it added nothing the text; trait correlations at the community-level are interesting in that they underscore possibly relationships between traits and species strategies, but they do not relate directly to our central questions or discussion.

Line 241: Why is it necessary to set d to a constant? Did you try allowing it to be a free parameter or will the models not converge if both d and m are estimated from the bayesian model?

This is correct. Because immigration rate and replacement rate influence community change in a compensatory way, the model cannot converge unless one was held constant.

Line 251: Clarify the Why C\_{i,t-1} is multiplied by 1-m? Is it assumed that migrating to elsewhere causes losses here? That doesn't seem right for rooted organisms like plants, that don't really lost ground one place when they gain it elsewhere. Instead, I think m should be the mortality rate, which I don't think biologically would equal the migration rate.

C{i, t-1} is multiplied by 1-m because we are modeling the probability that each new individual (actually percent cover units, in this case) originates from outside the turf (*P{i,t}*) as opposed to from genets already inside the turf (i.e., C{i,t-1}). In other words, because new individuals must either originate from within the turf or from outside of the turf, C and P must vary inversely; we use m and 1-m to impose this inverse relationship.

Line 260. In addition to gelman rubin to check convergence, other model diagnostics should be reported (either in main test or supp mat) that address model fit, not just convergence.

We have added an additional step in the Bayesian model to determine the mean posterior estimates of λ*i,t.* We compare these estimates of percent cover to the actual values to determine an R2 value. The mean and median R2 value at the site level is ~0.63. This information has been added to the text (Line 267).

Reviewer: 2

Comments to the Author

This study provides a timely test of whether trait-environment relationships can be used to predict responses to climate change. The turf communities studied provide a model experimental set up for testing key assumptions of this space-for-time substitution common in modelling approaches. Also, the ‘climate grid’ afforded by the site topography is another unique component of this experimental study, making this both a novel and innovative experiment. I make major and minor suggestions on areas that need improvement and/or clarification.

The study is conducted for a total of 5 years but there is little information on whether this is enough time to detect a change in species composition or trait plasticity during this time. Including citations for previous studied related to temporal change (like succession) in species composition would partially address this uncertainty.

As we mentioned above, our analysis is blind to any changes in CWMs driven by trait plasticity (a fact that we clarified further in the text; Line 189); we assume static trait values for species, and therefore our analysis focuses strictly on compositional changes. For the point about sufficient time, the fact that we did indeed observe consistent shifts in community composition over time is strong evidence that our experiment was conducted at an appropriate time scale.

Likewise, it is not clear whether the plot size of 25x25cm is large enough for significant amount of recruitment to occur or be detectable. The manuscript mentions that 10s-100s of individuals are found within each turf plot so a reader assumes that these are all small sized individuals. Including the range of heights of plants found within each turf plot would clarify this doubt.

The “vegetation canopy height” for each turf community was measured at each census point; the mean value was 9.1 cm +/- 6.2 cm (1 SD). This detail was added to the text (Line 140). In addition, the fact that new species appear in the plots makes it clear that recruitment is indeed occurring.

There have been few studies that have determined whether trait values in global databases are reliable estimates of trait values for the same species collected from a different location. If the traits used in this study are responsive to temperature across space, the use of LEDA trait values for sites that differ from the ones used in this study may possibly misrepresent trait values for a particular site. It may not be possible to disentangle in the present study but even a small demonstration that species collected in situ have similar trait values to those reported in LEDA would address this minor concern.

We agree that this is an important point. To partially answer your question, we have now added comparisons of field-collected trait values (SLA and leaf area only) to LEDA-collected values (Lines 168 – 170). The Pearson correlation coefficients are now included in the methods section (SLA: 0.69 and leaf area: 0.73).

As noted by the authors, within-turf recruitment is expected to be high due to the predominance of clonal growth. As a result, the probabilities of m and (1-m) can be given informed priors allowing m to never exceed (1-m). Yet m is given a uniform prior that varies from 0 to 1. Please comment on how this (giving m an informed prior) changes the reported outcomes.

This is a good suggestion. We changed the m prior to a uniform distribution from 0 to 0.5 (i.e., we gave m an informed prior such that m could not exceed 1-m) and it had no effect on model estimates. In retrospect, this is perhaps unsurprising given that estimates for m at each site never exceeded 0.5, but we have included a sentence about this in the text (Line 261).

Also, the authors use a relaxed assumption of zero-sum replacement and instead force simulated percent cover to match observed percent cover over the duration of the experiment (p13,L39). Yet plant biomass is known to respond (sometimes dramatically) to changes in climate. A discussion of how this assumption is either unavoidable (due to the nature of the Bayesian analyses) or doesn’t change the reported outcome would address this concern.

We agree that plant biomass can respond dramatically to changes in climate, and we feel that our decision to force simulated percent cover to match observed percent cover accounts for this phenomenon. For example, if total percent cover of a turf increased from 90 to 110 after transplantation (i.e., if turf plant biomass increased), the simulated turf percent cover would also increase from 90 to 110, thereby accounting for any change in total coverage.

The experiment relies heavily on a natural grid of sites that differ in temperature and precipitation. Although mean annual precipitation and temperature were compared between sites it will be important to consider temperature and precipitation differences in the years of the experiment. Perhaps the lack of responses to transplantation is an artefact of unusually cool/wet years during the experiment so even warm/dry sites look more similar to cool/wet sites thus masking any potential response. It could be that local site conditions were not measured during the experiment, limiting the ability to determine this relationship but some discussion related to this seems necessary.

Because we found trends in trait values along temperature gradients, not precipitation gradients, we focus on temperature in our response to this concern. While a particularly warm/cold year at a particular site (or all sites) could have affected transplant responses, this did not appear to be the case. We have added two figures to our supplementary information to substantiate this argument. The first (Figure S1) shows mean summer temperatures at 2m and 30cm aboveground over the duration of the experiment. Figure S1 shows that the sites maintain a consistent order in temperature, with roughly consistent differences between them (from alpine to intermediate to lowland). The second figure (Figure S2) shows climate conditions by year as experienced specifically by transplanted turfs relative to the meanclimate of their origin sites over the duration of the experiment. Figure S2 shows that turfs always experienced a warmer climate than they would have in their origin sites. You may notice that the temperatures experienced by transplanted turfs in 2012 deviated less from their origin climates than in other years; 2012 was an exceptionally cold year at all of the sites. The lower temperatures in 2012 (which were still ~1 degree warmer than the origin sites) were probably of secondary importance as the most dramatic trait-based community responses to turf transplantation occurred from 2009 to 2011. It is possible these cooler temperatures stalled the rate of community response in 2012, but this does not qualitatively alter our results or conclusions.

The results highlight that only maximum height and bud number responded to transplantation to warmer sites. While there is much trait ecology scattered throughout the discussion, incorporating more would broaden the applicability of these intriguing results. For example, an explanation of why SLA did not respond to transplantation (possibly because SLA is a poor predictor of growth strategies in herbaceous species where photosynthesize can occur in green stems) would strengthen the paper.

We have added some discussion about this topic, and have included your interesting hypothesis about the potential for species with photosynthetic stems to deviate from traditional SLA-based expectations of growth strategy (Line 345 – Line 348).

In sum, this is an innovative and timely experiment that addresses major assumption in climate change modelling. A more in-depth discussion on the above comments would strengthen the impact and significance of this paper to ongoing unresolved issues in climate change biology as well as trait-based ecology.