General comments

Thank you to the two reviewers and the editor for taking the time to provide feedback for our work. It was all very helpful and has resulted in what we believe to be a much improved manuscript. We accepted all but one of the minor editorial suggestions, as well as implementing all of the major changes suggested. Most notably, we re-ran the JSDM with post-seed treatment climate data as suggested and updated the text accordingly.

Editor comments

Both reviewers had valuable comments for the improvement of your manuscript, addressing them may be important and significantly improve the manuscript. In addition, my main concern relates to the assessment of only past climatic conditions on the final outcome. Given the availability of climate data for the first year of seedling establishment, I suggest evaluating the effect of that climate on the final outcome as well. This is because seedling establishment critically determines long-term restoration outcomes, and potentially also species interactions.

We re-ran the JSDM with post-climate variables as well as adding whether or not a species was part of the seeding treatments as a “trait”, as Reviewer #2 suggested. The new model ran with 5 times the iterations as the previous (7.5M total per chain) and converged well. New climate variables were:

- Pre-treatment
    - September-November
    - December-February
    - March-April
- Post-treatment:
    - May-June
    - July-September
    - October-December

Results for pre-treatment variables were very similar to the previous results, and the added post-treatment variables improved the model. We split figure 2 into panels to accommodate the additional variables, and added additional text to describe the post-treatment coefficients on **lines 265-274**.

Reviewer #1 (Comments to the Author):

This manuscript reports on a field experiment that aimed to understand the role of antecedent (planting-year) climate on grassland restoration outcomes. Plant communities were measured in the 10th and 9th growing season after restoring strips of agricultural land to grassland using the same seeding mix in two consecutive years (2013 and 2014). Total precipitation was similar among the two planting years, but the timing of precipitation differed. The 12-month standardized precipitation evaporation index (SPEI) was exceptionally low the prior to the first restoration. In addition to climate, topographic and edaphic variables were used to predict species, functional group, and community outcomes and associations. Results from this study demonstrate that different planting-year conditions differentially affect presence/absence of species, richness, community composition, dominant species and functional groups a decade after planting. Correlations among species abundances and associated traits imply facilitation and competition underlie different community outcomes. The results are nicely interpreted in the context of implications for restoration. In fact, the Discussion is the most well-written part of the manuscript

Thank you!

Moderate-to-major concerns & suggestions:

1. A potential confounding factor with the study design is that the first planting was sown into winter wheat that was harvested and the second planting was sown into winter wheat that was not harvested due to low grain yield. The second planting had more soil moisture, which was attributed to climate. But could the residual biomass of unharvested winter wheat have contributed to higher soil moisture in addition to climate? To this end, the results should be interpreted in the context of different antecedent environmental conditions or climate and soil moisture conditions throughout the manuscript (as stated in the first sentence of the discussion).

We agree that different management preceding planting may have some confounding effect on soil moisture at planting (approx. May 1), but that is difficult to assess. We added a paragraph to the discussion on lin**es 310-322:**

"There is some uncertainty about why soil moisture was different for the two seed applications. We believe weather was the main driver, but it is possible that this was complicated by differential management pre-planting. The first planting was preceded by wheat harvest in July 2012 and subsequent shallow tillage to maintain the fallow state, and the wheat stubble and residue may have provided some mulching effect, limiting evaporative losses. For the second planting, the wheat crop was unproductive and not harvested, with less stem density than the prior year's wheat residue and stubble, but with greater standing biomass. It is possible that shading/mulching was greater in these strips, but we argue these effects would have been minimal. We included strip identity and planting year as random and fixed effects, respectively, in the JSDM and those terms had minimal effects. Planting year had neutral effects on every species and explained 4.9% of the variation for the average species, while strip identity explained 3.5% on average. These potential confounding effects would not invalidate the model since we used direct measurements of soil moisture rather than downscaled climate data.**"**

2. The opening paragraph needs major revision: This paragraph does not set the stage for the objectives of this research. I anticipated that the manuscript would address the role of

interannual variability in climate on restoring species diversity on climate mitigation through carbon sequestration in soil.

Paragraph 1 now reads:

“Understanding how species diversity develops in novel ecosystems such as post agricultural landscapes is one of most important questions in plant science today (Armstrong et al., 2023; Bell et al., 2023). Globally, about 15 Mkm2 (10% of global land area) are currently managed as croplands, resulting in an estimated 133 Pg in cumulative carbon (C) emissions throughout human history (Sanderman, Hengl and Fiske, 2017). Reestablishing native perennial plant cover in post agricultural landscapes is an important way to restore these systems to C sinks and enhance soil stabilization, biodiversity and other ecosystem functions and services. However, in order to realize these benefits, native perennial plants need to be successfully established, and this has proven difficult for many restoration practitioners in arid and semi-arid grasslands. The task is even more difficult in areas managed for long periods as intensive agriculture, because often the native seed bank is depleted, soil fertility has declined, non-native plants are abundant, and ecohydrological function is altered (Turnbull et al., 2012; Shackelford et al., 2021). Furthermore, restoration practitioners are often guided by mean annual climate conditions when selecting species for seed mixes. But a typical year will tend to have a combination of high, low or average seasonal values of temperature, precipitation and other climate variables, and precipitation in particular can be very difficult to predict. Therefore, “mean conditions” rarely capture climate extremes that drive community responses, especially in drylands.”

I also not agree with the statement “restoration practitioners have a difficult task in re-establishing native plants in areas managed for long periods as intensive agriculture... (Lines 52-53).” In mesic grassland, e.g. tallgrass prairie, it is relatively easy to restore native grasses and many forbs, although achieving diversity representative of the historic system has proved challenging. Insert "in arid grasslands" after "restoration practitioners.

Accepted suggestion.

3. Last paragraph of Introduction: This paragraph needs to state the objectives and hypotheses of the research, and all methods should be removed. The objectives and hypotheses need to be congruent with response variables and modeling approaches in the methods. Moreover, not all results are aligned with objectives and methods. For example, diversity response was not listed as an objective and methods to calculate diversity (richness?) were not described. Similarly, objectives for the interspecific analyses were not presented and the analysis seems ad-hoc.

We rewrote that paragraph. Now reads “Seed germination and seedling establishment are especially sensitive to short-term fluctuations in climatic conditions (Larson et al., 2015; Shriver et al., 2018), even if mean annual temperature or precipitation is suitable for a given species’ climatic niche.” **on lines 80-83.**

Minor editorial suggestions:

1. Lines 87-88: This is an incomplete sentence.

We fixed that. Now reads “Seed germination and seedling establishment are especially sensitive to short-term fluctuations in climatic conditions (Larson et al., 2015; Shriver et al., 2018), even if mean annual temperature or precipitation is suitable for a given species’ climatic niche.” **on lines 80-83.**

2. Line 90: Insert “seedling” before establishment.

Accepted suggestion

3. Line 94: What is a seeding treatment? The seed mixes were the same, so this could confuse readers. I recommend changing the wording to “After seeding a former agricultural field to native species, ..”

Accepted suggestion

4. Line 97: Delete “treatment”

Declined suggestion: we prefer “restoration treatment”

5. Lines 98-101: Add a comma after “abundant” and delete the comma after “perennials.”

Accepted suggestion

6. Line 102: Add “In arid grasslands,”

Accepted suggestion

7. Line 110: Revise to “antecedent environmental conditions.”

Accepted suggestion

8. Line 146: How were the seed mixes sown? This is where the species and seeding rates of each species should be described, as well as whether seeding rates were based on live seed (PLS).

Lines 142-143 now read “The existing CRP treatment was implemented via drill seeding in two stages, with the same seed mix of 8 species (Table S1 has seed application rates).”

9. Lines 148-154: This paragraph needs to move forward because it describes study site climate conditions.

Accepted suggestion

10. Lines 156-164: This paragraph contains methods to measure topographic attributes that should be moved to the section on “Ancillary Data.” The soil classification should be moved to the first paragraph of the Methods.

Accepted suggestions

11. General: There is no indication of how the fields were managed after seeding. Were they left alone or were weeds managed in any way?

Added “ with no post-seeding management interventions” to lines **144-145**

12. Lines 170-172: The number of strips and plots per strip in each planting need to be specified. The alignment with soil sampling is for total N (not C) as indicated in Line 190

Added “Each of the 12 strips had about 7 plots, with 49 plots in the strips seeded in 2014, and 39 plots sampled for the strips sampled in 2013 (Figure S1)”. Replaced “soil C” with “soil macronutrients” (**lines 156-158**). Figure S1 now shows the botany sampling locations.

13. What was the rationale for using such a small sampling area (one tenth of a meter) and what were the dimensions of the sampling area (i.e., x by x meters or centimeters)?

One tenth of a square meter is 31.6 cm x 31.6 cm. This is a commonly used quadrat size. We added “Within each plot, we established 4, 0.1 m2 subplots (31.6 cm x 31.6 cm) at random locations.” **to lines 158-159**

15. Line 225: How is “direct competition” observable? I recommend deleting this sentence and simply providing the methods that relate to the objective.

Now reads: “We observed in the field that throughout much of the field, there was either high abundance of P. smithii or B. tectorum, but usually not both, and so we hypothesized that perhaps P. smithii was competing directly with B. tectorum.” **on lines 218-220.**

16. Lie 238 and throughout: I recommend referring to “CRP seed mix application” as “restoration” to broaden appeal to the audience.

Accepted suggestion

17. Line 290: Italicize B. tectorum

Accepted suggestion

Reviewer #2 (Comments to the Author)

This manuscript investigates the differences in plant composition between two (probably) identical seedings done in basically the same place but in two different years with markedly different antecedent weather conditions. The authors use the plant composition data to draw conclusions about the effects of those weather conditions on species establishment and on potential mutualistic or facilitative interactions among species, drawing implications from those conclusions for seed mix design. In my opinion, the strength of the manuscript is the first set of conclusions (though see my concerns about soil moisture and temperature readings, and the omission of some of the antecedent weather, below), the weakness is the second set of conclusions. Unfortunately, despite the title of the paper, most of the discussion is on residual species associations and their potential implications for planting, independent of environmental conditions. As the authors basically admit ("residual correlations … should be viewed skeptically"), their evidence for mutualism or facilitation is weak at best. Overall, I think the authors would better use their data from the relatively rare repetition of planting across years to focus on how weather conditions may impact seeding results. For example, the authors focus on antecedent conditions because managers will know those when putting seeds in the ground. But if weather conditions that follow the seeding are more strongly related to outcomes, how would that affect managers' choices of seed mixes?

We re-ran the JSDM with post-climate variables as well as adding whether or not a species was part of the seeding treatments as a "trait". See our response to the editor above for more details.

First, the title is misleading given that the paper is based on a single study site, two years of planting, and one year of data collection. Narrow down from "Western Great Plains". Similarly, with just two years of data, wider inference regarding the results are limited.

Title is now "Climate before and after planting drive divergent outcomes ten years after restoration of a wheat field to grassland"

Second, I had many questions about the methods: How many strips were planted each year?

Added "Each of the 12 strips had about 7 plots, with 49 plots in the strips seeded in 2014, and 39 plots sampled for the strips sampled in 2013 (Figure S1)" to **lines 156-158.**

Were the same batches of seeds used in both years (i.e., could viability have varied)?

Added "purchased from the same supplier" to **line 143.**

How were the 88 plots distributed among those strips? State here, don't just refer to other paper.

Figure S1 now has the plot locations and the probe locations.

What is the sample unit – plot or subplot? Fig. S2 caption suggests subplot, number of points in Fig. 2 suggests plot. But JSDM was apparently done on the subplot level (which makes sense for accounting for interspecific interactions).

The NMDS and diversity analyses were done at the plot scale. Added " at the plot scale." to lines 190, and "(4 per plot)" to line **200.**

The JSDM was done at the subplot level. see "we created a joint species distribution model (JSDM) in a Bayesian hierarchical framework (Tikhonov et al., 2020) for the occurrence of all species at the 0.1 m2 quadrat scale (4 per plot)." on lines **198-200.**

How were the 18 locations for soil moisture sensors distributed among strips? (A supplemental figure shows the locations, but doesn't provide the logic behind them.)

Figure S1 now has the veg plot and sensor locations. Added " Sixteen of these sensors were installed in two lines in adjacent strips in the eastern side of the field that captured the range of topographic variability of the field, and 2 more were placed at the far western edge of the field."

Why is the shallowest soil moisture (and only soil T) reading below the majority of plant roots? This greatly undermines my confidence in their relevance in describing plant-relevant environmental conditions.

Added: "The 30cm sensors represent soil moisture from 25-35 cm. Sensors needed to be buried at this depth to allow the farmer to apply shallow tillage while the field was in wheat/fallow rotation." to line**s 174-176.**

Why don't any of the time periods capture the November and December preceding seeding? Things that germinate in September-October can be killed by dry and/or cold during those months.

The prior model did indeed include November before planting, but we erroneously forgot to include that in the methods. The updated model now has September preceding planting to December after planting and everything in between. Line**s 177-187** now read:

"We used the topographic layers as predictors in a spatial process model (Nychka et al., 2021) to estimate seasonal averages of 30 cm soil temperature and moisture at a 5 m resolution (Fig. S1) for three time periods preceding seeding, and three post-seeding. Pre-seeding time periods were March and April to capture the conditions immediately preceding seeding, December, January and February to capture winter freezing conditions, and September-November, since many species actually germinate in fall and overwinter before growing in the spring. Post-seeding time periods were May and June to capture the early summer when most plants are actively growing and peak greenness occurs, July-September to capture the hottest part of the summer when most drought stress occurs, and October-December to capture the onset of cold after seeding."

When was the last shallow tillage applied? Was shallow tillage used in the strips planted in 2014? I ask because if the last tillage was applied after any of the three antecedent weather condition periods, it would alter mechanisms for the relevance of these periods to later plant community composition.

Sentence now reads "Prior to the CRP plantings, the first set was planted to winter wheat on Oct. 3, 2011, which was harvested on July 5, 2012, then maintained in fallow prior to CRP seed application using shallow tillage" on lines **145-147.**

In the JSDM, why not include whether the species was planted? Also, the emphasis on species occurrences rather than cover is somewhat concerning, as abundance is more important for a species' impact on ecosystem structure and functioning

Presence in the seed mix was added as a "trait" to the updated model. It did not come out as an important variable, but we kept it in there as something to account for. **We added "presence in the seed mix" to lines 209-210.**

Third, much of the text refers to traits included in the model, which are supposedly shown in Table 1, but traits aren't shown in this table. Perhaps something went wrong with the submission software converting to PDF. Luckily, I know most of these species and didn't have to look all their traits up to assess the statements in the text. Without this information though, it's not clear to the reader how the authors classified each species in terms of traits, what authority they used to do so, and how the number of each species with each trait may affect the ability of the analyses to detect trait-environment relationships. For example, the first sentence describing Fig. 3 says caespitose perennial native grasses are strongly positively associated with spring soil moisture but there's only one non-caespitose native perennial grass and its association with spring SM is also pretty clear. Further down, the authors say "rhizomatous occurrence had negative associations with spring soil temperature", but there were very few rhizomatous species (PASM, BRIN, COAR, CIAR) compared to non-rhizomatous.

We added life history traits used in the model as **table SX**, and changed the table reference accordingly.

"Taller plants had negative associations with bare ground, as well as grasses" – taller plants had negative associations with grasses?

The trait results changed a bit from the previous model and that paragraph now reads as such on **line 274-281**

“There were several associations between life history traits and conditions around the time of planting (Fig. S5). Fall soil temperature after planting was positively associated with C4 species and negatively associated with height. High spring soil moisture before planting was associated positively with Perennials. High spring soil temperature before planting was associated with graminoids. High summer soil moisture after fire was negatively associated with C4 species and positively associated with height. Summer soil temperatures after planting were positively associated with height. High topographic wetness index was associated negatively with height.“

“C4 plants were more likely at sites with high bare ground cover” doesn’t jive with Fig. 3.

We included bare ground cover measured in 2022 more to account for it rather than to use it to interpret results. Therefore we did not include the effect of bare ground on the traits in the new Figure S5 or in Figures 3 and 4.

Overall, use of the term “group” for analyses is misleading, especially in the Fig. 3 caption and the text referring to it where the term “functional group” is used, given that most of the “groups” consist of a single species and that traits do not correspond with groupings

Changed “functional group” to “functional type” on lin**e 255.** It is explained that only a few species with sample rarity were grouped for the JSDM on lines **203-206:** “Because the model is estimating the occurrence of species, it has difficulty with species that only occur at one or two plots. Therefore, we grouped locally rare species with other functionally similar species (Table S1).”

(e.g., perennials and non-perennials can be in the same group, as in LACT).

That example is the only time we grouped plants together with different durations. *Taraxacum officinale* (3 occurrences) and *Podospermum laciniatum* (2 occurrences) are biennial to perennial, and *Tragopogon dubious* (1 occurrence) is an annual to biennial. Duration is only one life history trait of many. There are other ways to group plants by function, for example the position of the meristem, root type, dispersal mechanism etc, and here, these Chicoridae tribe members are all rosette-forming, wind-dispersed, non-native species. For these reasons, we felt as though it was reasonable to group them for the JSDM analysis. We added the following text to lines **251-253**:

“The “Introduced Cichorioideae” group contains three species, Tragopogon dubius Scop., Taraxacum officinale L., and Scorzonera laciniata L. which are all rosette-forming, wind-dispersed, non-native species in the Cichorioideae subfamily of Asteraceae.”

And we also added a note explaining that nuance in the new **table S4.**

Minorly, table 1 says Bromus secalinus is grouped with Bromus tectorum, but the text often refers to B. tectorum, not “Introduced Annual Bromus”; were they grouped or not?

We added “The “Introduced Annual Bromus” group (Table S1) from the model groups together B. tectorum, which was very common and abundant, with B. secalinus, of which we encountered one individual plant. Hereafter we will refer to this group as simply B. tectorum.” to lines **248-251**

Finally, the two “groups” described supposedly based on results in Fig. 4 are not convincing. NAVI, a native perennial grass, appears more positively associated with BROM, and more negatively associated with other major perennial grasses (PASM, BOGR), than does BASC. Should it be included in Group 1? The authors themselves seem a little confused, as they switch how the refer to G1 and G2 after line 308.

We fixed the typos mixing up G1 and G2. We do not think that NAVI should be in G1. We explain on lines **282-287** that we defined the two groups based on positive intra-group correlations. NAVI was positively correlated with BROM, but negatively correlated with SECE, neutral with BASC in the initial model. After following the reviewer’s excellent suggestion to redo the model with post-seeding variables, the grouping comes out stronger in the updated model, and should be a little more convincing with the updated figure 5.

In the discussion of Medicago sativa, are the authors saying that it is a native forb? As an (introduced) agricultural crop, its establishment via seed is much more assured than establishment of most native forbs, so it is not a good analogy for them.

We see the potential for confusion. Added “M. sativa is an introduced agricultural crop, but it typically does not persist in high abundance without supplemental irrigation, making it an effective cover crop.” **on lines 353-355.**

Minor issues:
“Dryer” is a noun. “Drier” is the adjective you want.

Corrected

87: Incomplete sentence.

Lines **80-83** now read “Seed germination and seedling establishment are especially sensitive to short-term fluctuations in climatic conditions (Larson et al., 2015; Shriver et al., 2018), even if mean annual temperature or precipitation is suitable for a given species’ climatic niche.”

151: Should “sampling” be “planting”?

Accepted suggestion

162: Table 1 in this paper is about precip; reword citation to make clear you're referring to Table 1 in the reference.

Added "therein"

239-240, 240, 267: Punctuation needs fixing.

Lines 230-232 now read: "Across the field, three introduced species, *B. tectorum*, *Bassia scoparia*, and *Salsola tragus*, along with and the native *P. smithii*, were ubiquitous regardless of strip number and year of seed application."

Lines 256-259 now read The most prevalent annual introduced grasses, (S. cereale & B. tectorum) were insensitive to spring soil moisture but positively associated with spring soil temperature, while the most prevalent introduced forbs, B. scoparia and S. tragus were more strongly associated with fall and winter conditions before seeding."

262-263: These two species are also strongly associated with winter conditions.

Added "and winter" on line **259**

267-270: Sentence is redundant with others in the paragraph.

Deleted the sentence

Table 1 caption: Says "drake farm", but that location isn't mentioned anywhere in the main text.

Drake farm is now "study site".

Clarify what's meant by "High values are bold" – highest month in each year? If so, 2010 is incorrectly bolded.

Thanks for catching that. Fixed the 2010 bolding. Caption now reads "Highest monthly values are bold".

Figure 3 caption: Species are ordered by origin, then prevalence, but what "functional groups" are you referring to (presumably by the dotted lines)?

We changed it to simply grass vs forb, and altered the figure captions accordingly.

Table S1. Check names on species – some are outdated (e.g., Podospermum, Bassia). Explain "n" and "I" as used in "origin" column, or just write out the whole word in the table.

Accepted suggestion

The species seem to be ordered according to traits (though B. tectorum and B. secalinus aren't together), which only a reader familiar with the species would be able to figure out.

The table is sorted first by native status and second by family.

Figure S2: Just spell out the groups in the graph titles instead of making the reader translate the abbreviations. Refer to Table S1 for species codes. Prevalence ("number of plots in which a species was encountered") goes up to 200 in some graphs, but there are only 88 plots. Do you mean subplots?

That's a great suggestion. We updated Figure S2 to have species fully spelled out on the y axis, with prevalence by plot as the new x axis label.

Figure S3: What do the dashed red lines indicate? Grammar needs fixing in last sentence of caption.

The caption now reads: "Model convergence diagnostics. The potential scale reduction factor (PSRF) measures the convergence among chains, and being closer 1.0 is ideal. The vertical red dashed line lies at 1.001, a sensible target for most values to be less than. Effective sample size (ESS) measures autocorrelation between successive iterations within each chain, and higher values are better. Ideally ESS matches the number of posterior samples (indicated by the vertical red dashed line), but that is not fully necessary. Betas are the parameters for environmental filters, gammas are the parameters for the traits, and omegas are the parameters for species associations."