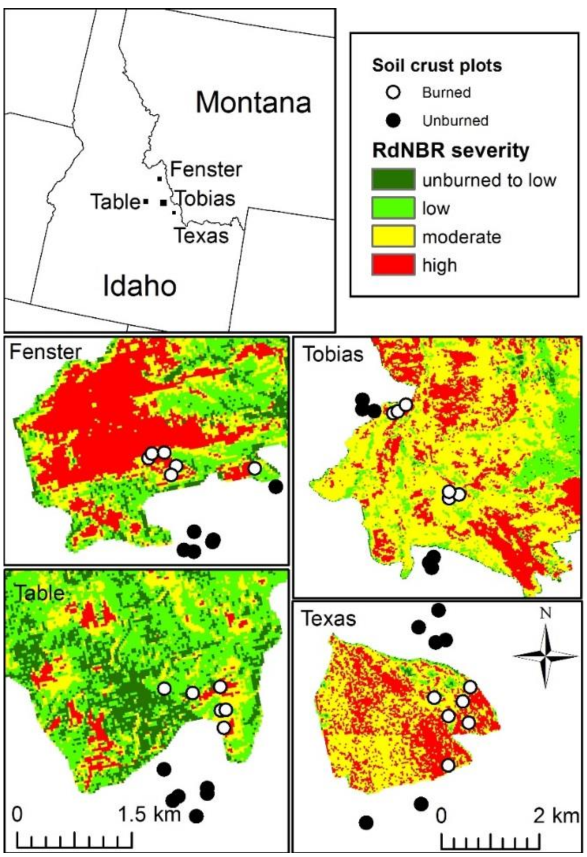
Thanks to both reviewers for careful consideration of our manuscript. We've responded below.

Both reviewers commented on how it appears from Fig. 1 that we sampled a small range of fire severity, suggesting that this is a concern that other readers will also have. The categorical dNBR shown is not especially useful because of the arbitrary cut-offs, and the effect of initial greenness (Kolden et. al. 2015). Because of that, we used continuous dNBR in analysis, which showed considerably more variability (from -32 to 321; p. 4 line 11). We also analyzed RdNBR, which is more appropriate for sparse vegetation or fires with high variability in pre-fire vegetation because it is relativized by pre-fire greenness (Miller and Thode 2007 Remote Sensing of Environment). RdNBR, which we think is the best severity metric available from MTBS, showed a much larger range of severities (-227 to 2629). We remade the fire severity maps with RdNBR based on severity classed used in a recent regional study (Reilly et al. 2017, Ecosphere) with unburned to low as RdNBR < 0, low severity as RdNBR < 235, Moderate severity as RdNBR 235-649, and high severity as RdNBR > 649. Here is our new version of Figure 1, which makes it much clearer that we measured quite a wide range of severities after relativizing by pre-fire greenness:



To respond to each reviewer, we've included below their original comments in black, our responses to reviewer #1 are in red and to reviewer #2 in blue.

Reviewer #1:

The authors document species of biological soil crusts at four sites in Idaho on the border with southwestern Montana in an effort to see their recovery 12-16 years following fire. The topic of biological soil crust recovery after fire is important and one that is understudied in North America in general. Also, I appreciate that the authors have attempted to account for secondary disturbances such as grazing. However, there appears to be a lack of understanding about the ecology of the system that they are working within as well as a lack of understanding of these two disturbances. I suggest some reworking of the paper that uses the strength of this dataset, the species identities and cover along gradients of elevation and precipitation. The authors could present which species were present at each site both inside and outside of the fire. It appears that there are some obvious compositional differences amongst sites given the ordinations.

This study was designed to examine the effects of wildfire on BSCs communities, not to evaluate the effects of precipitation and elevation gradients. Although that would be an interesting study, it would be best addressed with a dataset that broadly spanned those gradients across the region as has been done in some other papers (for example, Root & McCune 2012, Journal of Arid Environments). To focus on the effects of wildfire, we specifically sampled inside and adjacent to burned areas (clustered sampling around burned areas), and our design does not have sufficient replication across elevation and precipitation gradients to explain why the differences among fires was observed.

Although it may be too late for this manuscript, I suggest that the authors consider adding a coauthor on future projects to fill the knowledge gaps that are apparent in regards to the disturbances and ecology of this system.

See specific responses below.

My issues with the documentation of disturbances include only noting cow pies in 1m2 subplots. Cattle move across a landscape so this measure of cow pie cover, at this scale, is not appropriate when looking at animals that trample and graze over landscapes. Many folks use a piosphere approach for this reason.

Measuring grazing was not our primary purpose; several other studies have examined the effects of grazing on BSCs (see citations in the introduction). We sought to focus on the effects of wildfire, and included grazing only to make sure that our primary objective (wildfire) was not compromised by confounding with another disturbance (grazing). Several lines of reasoning lead us to conclude that the effects observed were primarily due to wildfire, including the cowpies, and the consistent presence of cattle freely moving between burned and unburned sites. It is still possible that cattle are selecting one habitat over the other, but there did not appear to be a difference in livestock management that would confound our interpretation of the fire effects. All sites in this study were grazed by cattle, the difference is some were burned and some were not burned. We will explicitly add that burned and unburned areas were in the same grazing allotments and that cattle moved freely across the sites if we are invited to revise the manuscript.

Table Mountain was a prescribed fire and prescribed burning is generally conducted outside of wildfire season. Its ecological effects should be different from the others and it looks like you see a wide range of BSC species at this site.

Table Mountain's burn date is noted in Table 1 as April 27; the next earliest fire was Tobias on July 12, so Table did burn a little earlier in the season than the others; it also was generally lower in severity. There could be other confounding factors here, and among the other fires. Since there is only one prescribed burn in the study we can't draw any conclusions about how that may have factored in our results. This is intriguing, and a possible area for future research.

We will clarify that Table Mountain is a prescribed burn in the study site description and Discussion.

My issues with the ecological knowledge used in this study applies to the authors identifying Artemisia as the dominant shrub, which does not resprout following fire. So why do the burned plots at the Fenster Creek fire have greater shrub cover than the unburned plots? Burned and unburned plots at Table Mountain have similar amounts of shrub cover. This does not make sense. Did these plots actually burn? Did something happen with the data? Was the dominant shrub species misidentified?

See response from Erich Dodson clarifying that we did indeed sample sagebrush steppe, not forest (with photos). There were no trees anywhere close to plots at most of the sites and absolutely no evidence of forest (live, dead, stumps, logs) within at least 50 m of any plots. We were careful about identification of the sagebrush species; they were tricky on our plots, and we brought samples to other botanists, particularly Roger Rosentreter (retired, Idaho BLM).

Resprouting (by *Artemesia* spp. and other shrub species) likely played a role in the higher shrub cover than expected by the reviewer for some sites. Recovery of high big sagebrush (*A. tridentata*) is possible on favorable sites within 1-2 decades (Nelson et al. 2014 IJWF; Shinneman and McIlroy 2016, IJWF). The most common shrubs at Fenster were *Artemisia tripartita*, *Gutierrezia sarothrae* and *Ericameria* spp. Schultz 2012 "Pocket Guide to Sagebrush" page 73 mentions of *A. tripartita* that "plants usually resprout after disturbance." Every burned plot that we sampled showed clear signs of having burned, with charred bunchgrass and/or sagebrush stumps.

At Table Mountain, it seemed possible that some of the sagebrush survived the burn, although as with the other fires, every plot had clear signs of having burned. There was a mix of shrub species, including: *A. tridentata* (ssp. *wyomingensis* and spp. *vaseyana*), *A. arbuscula* and *A. tripartita.* Like *A. tripartita, A. arbuscula* can resprout after disturbance, but Schultz (2012) states that it is uncommon (page 16). *A. tridentata* ssp. *vaseyana* can regenerate profusely from seed following wildfire (Champlin 1982).

If we are invited to revise the manuscript, we can add study site photos and greater detail on the variety of *Artemisia* species that were sampled at our study sites and explicitly state that several of these species are capable of resprouting following wildfire.

If these are sage steppe sites as the authors seems to imply, the expected fire return interval would be at least 100 years but we are led to believe that 12-16 years is a "long" time since fire.

This is a good point, we will add information about fire return intervals to the introduction if we are invited to revise the manuscript.

We were considering this time frame as long not in reference to the return interval, but in reference to the fact that most studies of wildfire effects are done very shortly (1-5 years) following the burn. A related point that we would add is that BSCs are generally considered to take several decades to recover.

Given that the authors had to look for patches free of tree cover, I'm guessing that this is not sage steppe and maybe something with a more frequent fire return interval such as ponderosa pine? The reader isn't told what the dominant tree is.

We will clarify this in the methods. Most of these fires were quite large and spanned a wide elevation range, and the upper parts of the fire were forested. By sampling at lower elevations (places without trees), we were solidly in sagebrush steppe habitat. We didn't have to look for patches without trees so much as focus on the lower elevation habitat. Most of our plots were far from trees, only at Tobias did we have trees within 100 m of two of our plots. We weren't close enough to trees at most sites to notice what they were. At Tobias, we observed *Pseudotsuga menziesii*. Only Fenster is within the geographic range of ponderosa pine.

Also, there is a lot of discussion about invasive annual grasses and cheatgrass- fire cycles for a study on the edge of a forested system where only one site has annual grasses. Cheatgrass is shade intolerant. See work by McGinness and Keeley. Again, being on the edge of a forested system, I think that you are in a different ecosystem. It is great to add to what we know about BSCs where there is not a lot of cheatgrass but this is a different system and that should be discussed.

Since we were not in forested habitat, we were actually quite surprised that there wasn't more cheatgrass. It looked like habitat that, if it were in the Snake River Plain (ID), would have been completely overgrown with cheatgrass. A 2016 paper (Brummer et al. http://scholarworks.montana.edu/xmlui/bitstream/handle/1/12644/Lavin_Ecosystems_2016.pdf), which we will add to the discussion if we are invited to revise the manuscript, suggests that cheatgrass dominance is limited by climate, and is less of a problem in our study area. Because many of the wildfire studies in sagebrush steppe have a major cheatgrass component, this piece of the story seemed important to discuss. Given that there is a substantial amount sagebrush steppe habitat mapped by Brummer et al. (2012) where cheatgrass isn't a major problem, our study adds important insight into fire effects in these areas.

I'm also thinking that fire severity was not accurately captured in this study. See work by Kolden related to the ecological meaningfulness of MTBS. MTBS data is 30m2 pixels so the average value is given for the change in reflectance of vegetation at that scale. You are comparing this with your 1m2 subplots on which biocrusts were measured. You would have to assume that the fire burned along the soil surface evenly over a 30m2 pixel for this to be the right scale at which to assess fire severity on BSCs.

We acknowledge that in the discussion (p. 8 line 25 and 31-34), and agree that there are concerns about the ecological meaningfulness. MTBS pixels are based on LANDSAT, which is 30 m x 30 m pixels (900 m^2 not 30 m^2). This is clearly mismatched with the mm scale of BSCs, which is why we pooled the 8 subplots per plot for the analysis of burn severity (which we could clarify in a revision). This is also a scale mismatch with shrubs and even trees, and considerable within-pixel variability is likely in rangelands where the average plant is less than 1 m^2.

Our point isn't that we don't think that severity matters, but rather that severity as measured by MTBS doesn't impact the BSCs (p. 8 lines 28-29). This means that if management decisions are made based on MTBS data, which they are, they won't reflect what's happening on the ground for soil crusts and the biogeochemical processes they provide. It would be interesting to see what the data would look like when the true severity on microplots is known, but that would require much more information than is generally available.

Specific Comments

p.1 Line 9- Saying that BSC richness is 65% greater when comparing burned to unburned plots seems to conflict with your telling us throughout the paper that there were dramatic differences in richness amongst sites. This kind of statement needs to include the range in differences or what they were for each site since there were only four.

There can be both big differences among sites and a large effect of fire at the same time (what we found). There was not a statistically significant interaction between fire (site) and burn treatment for richness. If there were (there was for cover, which we summarized differently), it is absolutely true that we would need to clarify the effect size for each fire. Since there was not, it is entirely appropriate to summarize the main effect, though we can clarify the lack of an interaction in a revision.

Line 29- Be more specific about which lichen crusts. Also "thick crusts" needs a definition.

We can revise this to be more specific.

p. 2 Line 9 It would be nice to see examples of the different BSC groups described. For example, what are "tall growth forms"?

This is a good idea, and is a point also brought up by reviewer #2. We would like to add a figure with examples of different growth forms in the revision. This may also address the previous point about "thick crusts."

Line 16-17 Need a citation.

We aren't sure what this refers to.

p.4 Line 8- How were your points randomly selected?

Using http://www.geomidpoint.com/random/, which we will add.

Line 14- You say that dNBR and RdNBR varied but from your maps it looks like you only surveyed the low severity end of the spectrum. I don't see how this range tells us anything about the meaningful of the data.

See the comments for both reviewers at the top. The low severity in the maps was based on the categorical dNBR which was not used in any analyses but is a standard format for mapping. Thresholds for classes in dNBR are somewhat arbitrary and vary from fire to fire (Kolden et. al. 2015). The continuous dNBR used in analysis showed considerably more variability (from -32 to 321; p. 4 line 11). RdNBR is more appropriate for sparse vegetation or fires with high variability in prefire vegetation (e.g., nonforest vegetation; Miller and Thode 2007 Remote Sensing of Environment) showed a much larger range of severities (-227 to 2629). In a revision we will elaborate on the use of dNBR and RdNBR as fire severity measures in both the Introduction and Discussion, including why RdNBR is preferable for our study site. We will also remake the fire severity maps with categorical RdNBR severity to help reduce confusion and focus on RdNBR.

p.5 Line 7- You should state your reasons for using generalized linear mixed models. Was there something different about the distributions of your response variable? Please clarify this.

This is just a mixed model with a normal distribution (LMM, which is a subset of GLMM). We will clarify in the revision that it is a linear mixed model instead of generalized linear mixed model.

Figure 2- It would be valuable to see the sites differentiated here. How much is the change in cover driven by one or two sites? Given the ordinations, it looks like two sites are driving this relationship.

This is a good idea, we can add it to the figure. Yes, Table and Texas had a much stronger pattern than Tobias and Fenster (as described in Section 3.1), even though Table was a prescribed burn.

Figure 4. Define the biocrust growth forms presented here. A superscript of which sites each species occurred on would be really interesting.

We cited the literature that our growth forms followed (p. 4 lines 31-32) as well as how we defined each species (appendix). We don't think it is necessary to add specific definitions here, however, a better sense for them would be added with a figure showing examples. We can add information on sites where species occurred to the appendix table with all the species.

Reviewer #2:

In this manuscript, the authors address the long-term influence of fire on biotic soil crust communities in a sagebrush steppe system. Wildfires are increasing in frequency and severity in this region and we know very little about how this disturbance influences soil crust communities which play an integral role in structuring these systems. This research helps to fill that gap by comparing vascular and nonvascular communities in 12-16 year old burned sites with adjacent communities in unburned areas. While the functional group distinctions are interesting and appear to be very important, many readers like myself are likely to not know how these nonvascular groups actually differ in function, the importance of those functions or the consequences for burned systems which might experience a loss of some functional group. Discussion on this in the introduction and then later in the discussion would add more clarification on the functions of those groups and how changes in functional group diversity might impact ecosystem function. I agree that the purpose and findings of this study increase our understanding of the resilience of this system and of the BSC communities to wildfires. A return to this concept in the discussion could strengthen that connection for your reader.

This is very helpful feedback, we will add a paragraph and figure to further describe and illustrate the growth forms.

Page 3, line 14: This wording is slightly confusing. Are you referring to the end product of recovery, the process of recovery, or that different communities will recover in different ways (or all of the above?)?

We mean that different communities will recover in different ways; we will clarify this in the revision.

Page 3, lines 15-17: Is this something you can parse out with your data? Were there remnant populations within the burned areas or do you not know? If you think that remnant populations may have contributed to the patterns you find it should be added into the discussion. If not, this information may be extraneous here.

No, we can't really tease this out, but we can come back to the idea in the discussion. The reason that it is relevant is that if low severity sites retain their BSCs better, then they act as patches for recovery. However, we didn't find that low severity plots retained more of their BSCs. This is related to the severity/patchiness comments of reviewer #1, if we had data regarding post-fire BSC survival and how that relates to severity we might be able to answer this. With one visit over a decade later it was not possible to tell if the patches seen are remnants or regenerants.

Page 3, paragraph 2: Information on the negative impacts of fire and recovery go back and forth a bit in a manner that it not easy to follow. You may consider separating the topics and even integrating some of the previous paragraph on disturbance and recovery from mechanical disturbances.

Good idea, we will re-work these two paragraphs.

Page 3, line 20: This second mention of algal and cyanobacterial community recovery makes me wonder if this is something you looked at or are you assuming that since your BSC communities were dominated by lichen and moss that they are later successional BSC communities?

We can clarify this. We didn't look at it, because it requires other methods – you can't just look and see these communities. It isn't necessary safe to assume full recovery of those communities, even though they are considered early successional; most studies are finding that these communities are more complex than we thought. We'll add a little more about that and a citation here.

Page 3, line 23: It would be interested to understand how these recovering ecosystem functions might be connected with the recovering BSCs functional groups if it is known.

It is less well-known than would be ideal, but we will add what is known with citations from the relevant literature.

Page 4, lines 12-14: It would be useful if these parameters could be clarified, defined or at least contextualized. For instance, does -32 dNBR indicate a greater severity burn than 381 dNBR?

Good point, we will clarify these metrics.

Page 6, lines 15-21: Could this also be interpreted as when there were more crusts present there were more crusts to be damaged? You might play with how you present these results. To me, the part you are presenting is the common sense part (e.g. there are more crusts in unburned plots and they have a greater richness and cover). Isn't the question though about how fires influence long term BSC cover and richness (or recovery) and whether or not the variation present in the BSC community before fire influences that recovery? I actually think all of this information is there in your results but I am having a hard time picking it out. You might consider reworking the presentation of these results to make that more clear.

All of these comments are true, and we will re-write to clarify this in a revision. It seems the connection between the nuts and bolts of our results to the broader themes mentioned (long term recovery, pre-fire communities) was not as clear as it could be to this reviewer. Part of the problem is that fact that this was not a long-term study that could clearly draw a straight line from 1) pre-fire community, through 2) disturbance severity, to 3) recovery. We can only piece the story of these sites together with incomplete information, but the story we are laying out is the same. We have elements in our results that pertain to each of the three steps but in revision, we will be more explicit about connecting the dots.

Page 7, line 7: This is great! It is crazy that burn severity does not influence BSC cover or richness?

Yeah! But possibly because the remotely sensed data are too coarse.

Page 7, BSC functional groups: It would be interesting to know which functional groups were indicators of burned sites for a within and across site understanding of the influence of wildfires on BSC functional groups. I wonder if it is possible for you to parse apart if any groups were more strongly influenced by the high severity fires? This could help lay out those ecological expectations of changing BSC functional groups after fire and as BSCs recover from fire.

No functional groups indicated burned sites (we can add that explicitly). Great idea about functional groups and burn severity, we have the data to explore those potential patterns in a revision.

Page 8, lines 1-5: I think this is one of the most important findings of this research and it makes me see that the connection between fire intensity and invasive annual grass cover was not clear in the introduction. Knowing that well developed BSC communities are vulnerable to wildfire damage is important. Well done!

Thanks!

Page 8, lines 10-13: I think these findings are really interesting!

Thanks!

Page 8, line 14: recovery to recover

Will do.

Page 8, lines13-15: Is this contrary? Since the speed of recovery with or without vascular plants wasn't tested in this study, I'm not sure that the two are comparable but there is a good chance that I am simply misunderstanding.

This needs to be clarified. What we mean is that our study is different from other studies in cheatgrass dominated ecosystems and provides insight into a different kind of habitat that hasn't been as well studied.

Page 8, lines 25-27: My understanding was that the fire severity metrics were quantitative values associated with the fire intensity. My interpretation then of these sentences is that they were derived post-fire (sorry I realize that is obvious but I had to spell it out for myself), based on reflectance, and that when those reflections were off of crusts the metrics appear to be incorrect. I assumed that the severity metrics would be used to compare recovery across a severity gradient but the conclusion of this paragraph states that these metrics may not be able to predict long term effects of fire on BSC communities. Was this the goal? You should consider stating it more plainly. From the maps, it does look like your sites did not venture into high severity conditions. Could the discrepancy be simply that the fires were not severe enough?

We were hoping to examine how severity affected recovery, but we found that there wasn't a relationship with severity measured by MTBS. This could be either because severity doesn't affect recovery or because the metric we have for severity isn't very good. We will state this more plainly in the revision.

It is true that in dNBR we didn't go into the highest severity category, this is because those sites are all forested. So, from the perspective of severity in sagebrush steppe ecosystems, we spanned the range of what is possible – you can't get that severe of a dNBR value without having trees to start with, which is why RdNBR is likely a more useful metric in sagebrush steppe. See also our response to both reviewers, and our response to reviewer #1 on the same topic.

This question brings up an interesting point about the severity measures that we would like to expand in a revision about how severity measures are used to initially evaluate the impact of disturbance, but may or may not be relevant to long-term recovery.

Figure 4: I stared at this figure for a long time trying to come up with other ways to present it that might be easier to interpret. Ultimately though, I think you are right on. I'm sure you played with making paired ordinations of burned and unburned? I just wondered if that might help at all. There is a lot of information here that isn't discussed. I'm not sure that it is necessary to do so but there appear to be some neat patterns that might help parse apart some of the site specific differences in BSC growth forms, cover, and richness. Wonderful!

It is a juicy figure; we think we've picked out the key patterns in section 3.3, but we will consider adding to that interpretation. Paired ordinations were harder to interpret because they made it harder to see how the burned and unburned plots within the same fire were similar to each other; we concluded that this presentation best made our point that the functional communities differed among sites in ways that were patterned by climate. It is difficult to pick apart which reasons contribute to the differences among fires because so many of the climate variables are interrelated. Our study focused on the fire effects, and didn't sample the other gradients as broadly. However, it does suggest that there is a site-specific component to the effects of wildfire.