RCS ISPR comments of Study Design					
Comment topic	**Reviewer**	**Location: Page/Line in original document**	**Reviewer comment**	**Author response**	**Reviewer response to author action**
Definition of shade	**All reviewers**		There is a lack of clarity and consistency in the use of the term "shade" and its quantification. AE recommendations: The authors should compute what R1 called "FAR" and what R2 called "effective shade" in addition to their proposed index based on pixel counting, which R2 has noted is actually a canopy density for a portion of the sky dome. R1 recommended using a densiometer to estimate canopy density in addition to determining shade indices from the hemispherical photographs. This is a worthwhile idea and should not require an inordinate amount of extra time in the field. If densiometer measurements are found to correlate with effective shade, that could be a useful tool for operational situations given the relative ease of making densiometer measurements compared to hemispherical photography.	Please see changes to text and related author responses throughout this matrix. Note that we are interested in change in shade in response to harvests and blocking by site.	R3: Line 548: the first analysis is labeled "Determine how treatments affect **shade**". Should 'shade' be changed to 'changes in shade over initial conditions' ?
Definition of shade	**3**		Effective shade is actually not shade but the proportion of sky occupied by canopy. I wondered if the scientific community would all agree that canopy cover could be labeled shade. I note this only because I wondered about it and also wondered if it would be more accurate to rename your response to more accurately reflect what it is and then discuss why the response is a shade metric. But I'm not the expert in shade so leave that up to the authors.	Refinements have been made to the text accordingly.	Thank you
Research approach	**3**	Pg. 2-4	I appreciated the overview of the investigation on pages 2-4 prior to the more detailed explanation in the rest of the document. But some important points are missing in the section which led me to be confused as I continued to read. I suggest including a more explicit description of the temporal progression of the no-harvest-zone-width levels and the timing of the data collection (photographs) for each treatment combination in the beginning. A more explicit description of the chronological implementation of the harvest treatments would make the description of the process more clear. Specific edits suggested by reviewer were added to the study design doc in track changes by PM.	Agreed. Specific edits suggested by reviewer were incorporated. The figure (Figure 4) has been updated to more clearly show the harvest and photo sequence.	Thank you - the revision is MUCH more clear.
Research approach	**3**	Pg. 2 first paragraph under research approach section	Study is identified as a 'split plot' design with 2 factors and the levels of the factors are identified. The paragraph refers to Table 1. Further in the document (Analysis section), it is correctly noted that the design is not technically a split plot design since the split plot factor (zone width) is not randomly assigned to each whole plot (the plot that receives a thinning intensity). Zone width is assigned in the same order every whole plot. When the 'split plot factor' is applied in the same order in all whole plots, the design is not a split plot design but is a 'strip plot design'. Most experimental design texts will have this identified. However, this design is not exactly this either given the imposition of the ecoregion factor. I have included more details when reviewing the Analysis section.	See text and author responses throughout for changes to split plot language and descriptions.	Thanks - changes are helpful.
Research approach	**3**	Pg. 3 Table 1	Table 1 incorrectly identifies the 100' zone width as only existing in the plots assigned to the clear-cut thinning intensity. This is not correct. As shown in Figure 1, the 100' width occurs in every plot assigned to a thinning intensity. This particular level of width only exists when no thinning intensity has been implemented within the 100' strip of land so that may need clarification in the text as well. But technically the 100' width occurs in every "whole plot" that receives a thinning intensity. This could be interpreted as the 100' width being applied not to a plot but to the entire site. The LMM analysis could reflect that (replication for the 100' width is achieved through sites, not plots) but it's a complication to the analysis and likely requires a statistician's involvement in order to be that contrasts and standard errors are correctly estimated. An alternative approach could be: (a) be sure that photos are taken in in all 3 plots after Step 1 (see later comment about describing timing of photos) but before Step 2 and (b) Be sure that measurements for the 100' width are made independently in each plot at each site. Then treat the study as though there are 12 treatment combinations – each level of zone width with each level of thinning intensity. What will be true is that 3 of the treatment combinations (100' with moderate thinning, 100' with heavy thinning and 100' with clear-cut) are all equivalent. The analysis section of the manuscript notes that they can be averaged to estimate the effect of the 100' width. Then the analysis could proceed as a fully factorial design with all 12 treatment combinations with the understanding that 3 of them are identical. The advantage of having the 100' in each plot is that comparisons involving the 100' width and other widths will use the 100' width value from the specific plot which may be less variable than if the comparison involved an average over the 3 plots.	The table has been revised to indicate that the 100 ft treatment will be analyzed separately. Otherwise, no change needed because what Reviewer 3 describes regarding the photo sequence matches what is being proposed. The table depicts the different riparian *treatment level combinations* that will be included in this study, not individual plot configurations (the table does not specifically depict Plots 1, 2, and 3). Figure 4 shows how the treatments will be sequentially applied within each plot. Figure 4 has been expanded to clarify that photos will be taken after each treatment in each plot.	Thanks, the revisions clarify this.
Research approach	**3**	Research approach section	In this section of the manuscript, the timing of the photographs (when they are taken) is not described. I suggest including a more explicit description of the temporal progression of the no-harvest-zone-width levels and the timing of the data collection (photographs) for each treatment combination on page 3 in the Harvest sequence section. It is never explicitly described here and I was left fairly confused as I read through the rest of the document.	Comments incorporated. Figure 4 (formerly Figure 1) has been expanded to clarify that photos will be taken after each treatment in each plot.	Thank you - the revision is MUCH more clear.
Research approach	**3**	Pg. 3 Table 1	Please include a description of the timing of the photographs in the caption. Figure 1. I found the phrase "three harvest sequences" in the caption to be confusing – there are 4 rows in your figure but you say '3 sequences' and you want the reader to read 'down' the rows to see the sequences. One suggestion is to say 'Site layout and the progression from 100 foot no-harvest width (1) to 25-foot no-harvest-width (4) for 3 plots along one stream in this study. A compass rose could be helpful since the legend says that the figure is for an east/west orientation, but I think what is meant is that the top of the page is north.	Comments incorporated. Figure 4 (formerly Figure 1) has been expanded to clarify that photos will be taken after each treatment in each plot.	Thank you - the revision is MUCH more clear.
Shade measures	**3**	Pg. 5, first paragraph	Paragraph 1 was confusing and lacks clarity. I suggest integrating the information from the more clear description beginning on page 13. The page 2 description inaccurately says that shade is estimated for each treatment – it's really estimated for each plot. The first paragraph says that there will be 5 photo collection intervals and refers to Figure 1. But Figure 1 shows the 5 photo locations, not intervals. The paragraph should describe the start, end and length of the intervals.	Figure 1 (now Figure 4) has been expanded for clarity. There are in fact 5 photo intervals and 5 photo points per plot, so no change is needed. Shade will in fact be estimated for each of the 10 treatment level combinations.	Thanks. Having the figure include the intervals helps.

Shade measures	**3**	Pg. 5, third paragraph	I'm not an expert in photo analysis so my comments may reflect my ignorance. I did not understand this paragraph at all when I first read it and also after reading pages 16- 17; so, I think that there is likely not enough detail/context. The first sentence "The same set of photos will be used to estimate effective shade for streams with (1) east-west and (2) north-south orientations, regardless of actual stream orientation in the field" could be more clear by explaining the process rather than the outcome – i.e., "Regardless of actual stream orientation in the field, in the software, the stream will be reoriented to lie in an east-west and a north-south orientation". But I am still puzzled by this notion of 'reorientation' and why this is needed. I think that more description of why this is an appropriate technique is needed for non-experts. Here is my (possibly erroneous) thinking. The path of sun across the stream, and relative to the stream, in situ, will change depending on the actual stream orientation. The stream-adjacent vegetation, which may then be present in pixels within the photo, is also likely a function of the actual stream orientation. So, the measurement (proportion of canopied pixels) is a function of the actual stream orientation even if you change the virtual orientation within the software. That is, virtually reorienting the stream direction, redefining the solar path based on the reorientation will indeed give you a different value for your shade response, but it is STILL a biological function of how the plants around the stream grew based on the actual orientation in the field. So, I really don't understand how the virtual reorientation will give you a measurement of shade that is biologically defensible. As I said, I may be all wrong here so will defer to the experts but given my confusion, some clarifications are	See text for expanded explanation and updated figure.	The expanded explanation is helpful.
Re-orienting photographs to simulate different stream azimuths	**3**		I am still puzzled by this notion of 'reorientation' and why this is needed. I think that more description of why this is an appropriate technique is needed for nonexperts. Here is my (possibly erroneous) thinking. The path of sun across the stream, and relative to the stream, in situ, will change depending on the actual stream orientation. The stream-adjacent vegetation, which may then be present in pixels within the photo, is also likely a function of the actual stream orientation. So, the measurement (proportion of canopied pixels) is a function of the actual stream orientation even if you change the virtual orientation within the software. That is, virtually reorienting the stream direction, redefining the solar path based on the reorientation will indeed give you a different value for your shade response, but it is STILL a biological function of how the plants around the stream grew based on the actual orientation in the field. So, I really don't understand how the virtual reorientation will give you a measurement of shade that is biologically defensible.	See expanded explanation in the text and responses throughout this matrix. The stand inventory data will help us explain differences in stand composition and structure across sites.	It's much clearer now, thanks.
Re-orienting photographs to simulate different stream azimuths	**3**	Pg. 22, end of the 2nd full paragraph	On p. 22 at the end of the 2nd full paragraph it says that the re-orientation of the streams to standard orientations (N/S, E/W) "will ensure that shade response to the treatments is not influenced by differences in stream orientation across sites". I've noted this earlier, but it comes up here too - I question whether this is biologically reasonable. The under and over-story growth on a site is a function of the site and environmental conditions at a site and that includes the orientation of the stream relative to the tree stands. Re-orienting the image doesn't change the ***in situ*** conditions that affect plant growth and vigor. So, while some conditions may be standardized by the re-orientation, others are not. I suggest you clarify what conditions can and cannot be standardized.	Stream orientation will be standardized *in terms of solar geometry - the solar path for the specified study period* . We acknowledge that vegetation conditions will not be standardized across sites and stream orientations. The 100% stand inventory data will provide detailed information stand composition and structure in each plot/treatment, so these variables will be known. Understory vegetation influences will be greatly reduced in this study due to the solar altitudes being analyzed (40° or greater) and the elevation of the camera (1 m).	Ok. Thanks for the explanation.
Re-orienting photographs to simulate different stream azimuths	**3**	Pg. 16/17, data analysis procedures, effective shade section	In other comments I tried to explain some of my confusion with the re-orientation of photos. I suggest consolidating information on Page 5 with this section together so a reader can obtain the whole picture at one time with consistent language.	Agreed, the document has been substantially restructured accordingly.	Thanks, it's much better
Re-orienting photographs to simulate different stream azimuths	**3**		I suggest adding additional language to explain how (or if) the effective shade measurements change under reorientation.	An expanded explanation has been added to the text.	Thank you.
Re-orienting photographs to simulate different stream azimuths	**3**	Pg. 5	It is not clear to me how reorientation of the stream in the software "*…will standardize estimates of effective shade according to stream orientation*". Effective shade is the "percentage of pixels occupied by canopy within the portion of the photograph where the sun path crosses the sky during the period from 1 June to 1 September for solar altitudes 40° or greater" (p 5). My understanding is that the 'sun path' depends on the true cardinal directions at the site. The text reads as if the reorientation will redefine the sun path so that a different value for effective shade will be determined under different orientations of the photo. Otherwise, why bother to reorient?	"The text reads as if the reorientation will redefine the sun path so that a different value for effective shade will be determined under different orientations of the photo." - This is a feasible assumption. This comment is in alignment with what we are proposing.	If this is an assumption that is implied it should probably be added to the text
Re-orienting photographs to simulate different stream azimuths	**3**		Another suggestion is to give an example photo and identify the sun path, the canopy and an effective shade measurement. The figure on page 17 isn't too helpful since it simply shows the direction of the stream being changed but the sun's path is not present.	The figure has been updated to improve clarity.	Thanks!
Re-orienting photographs to simulate different stream azimuths	**3**		If the idea is that the sun's path changes in the photo with reorientation, some explanation about why this is biologically meaningful will be needed. See previous comments.	See responses throughout. Relevant changes have been made to the text.	The changes helped, thanks.
Taking hemispherical photographs in the field and post-processing the images	**3**	Pg. 14, second paragraph	One bullet point says, "Hemispherical photographs will only be collected under uniform sky conditions when the sun is not directly in view, and according to the camera manual" and there is a somewhat different sentence on page 14 that says, "Photographs will be taken when no direct sunlight is visible, at pre-dawn, post-sunset, or under an evenly overcast sky." I suggest you decide which is most specific and use common language for both. There are conditions under which the sentence could contradict each other.	Deleted redundant information in QA/QC section and retained more detailed description earlier in document.	Thanks.
Study design - site selection and study population	**3**	Pg. 10-11, Figure 4a and 4b	I don't fully understand how the hemispherical photos 'work' so my comments may be irrelevant but want to provide them to illustrate how a reader may be confused. The figure shows shadow length of a tree located at the photo-taking location near the plot. I thought I understood that the photos would be of the canopy over the stream. Is the reader supposed to infer that a tree located along one of the shadow lengths in the figure (not at the photo point) would be captured by the photo taken along the stream? It would help to explain why imagining a tree located at the photo-taking point is relevant.	No change needed. "Is the reader supposed to infer that a tree located along one of the shadow lengths in the figure (not at the photo point) would be captured by the photo taken along the stream?" - Yes. The figure does not ask the reader to imagine a tree located at the photo-taking point. The 'imaginary tree' is at the outer extent of the lines relative to the central photo point.	Thanks for the explanation

Forest and site characteristics	**3**	Pg. 5, Stand Characteristics section, first sentence	Paragraph 1, first sentence: "Stand structural and compositional metrics known to influence stream shade will be measured in each plot for each harvest treatment and treatment zone (stream-adjacent no-harvest zone or adjacent-stand harvest zone)." For clarity and reproducibility, please explicitly identify the metrics "known to influence stream shade". If these metrics are the only tree metrics identified later in the paragraph, just say so. Please use language about factors, levels of factors and treatments carefully and accurately. Levels of factors are NOT treatments. Factors have levels and combinations of levels of factors comprise treatments. This error occurs in other places in the document and will need attention there too. The language in the first sentence of this section confuses these definitions. The sentence says 'measured in each plot for each treatment' which is confusing because (a) a plot has been defined as the large area which receives multiple no-harvest-zone-width levels (is a plot or subplot measured?) and (b) it does not explain WHEN the metrics will be obtained with respect to the timing of the implementation of the no-harvest-zone-widths. Something along the lines of 'measured in each plot after the implementation of each of the levels of zone width' would be clearer.	Text has been revised to remove confusing language.	Well done!
Forest and site characteristics	**3**	Pg. 5, stand characteristics section	It's stated that trees and tree measurements are 100% inventoried. Is this 100% inventory conducted once in each of the 3 plots on a stream or is it conducted once after each of the 4 implementations of the no-harvest-zone-width level within a plot? A suggestion for your consideration is to measure the tree density in the field after the various harvests. Variation in harvesting or even errors in the field can lead to tree densities that are not what you intended. Having a field measurement could be helpful if you find you have to 'adjust' for the actual density in the plot.	What Reviewer 3 describes matches what is being proposed. As stated in the document, "The tag number of each harvested tree at each treatment interval will be recorded so that stand characteristics (e.g., basal area by species) can be computed for the harvest and no-harvest zones for each interval." .... "After each thinning treatment, follow-up inspections will be conducted to ensure that all trees marked for harvest were harvested or at least felled. Additional harvest may be required to meet this standard. Additionally, any unintended tree falling or damage that occurred during the harvest activities will be recorded by tag number."	The revision on the new page 9 is more explicit - thanks.
Forest and site characteristics	**3**	Pg. 5, stand characteristics section	"Stand data will be used to help account for changes in shade in response to the treatments, variation in shade response among ecoregions, and the magnitude of statistical model variance. Stand data will be used to improve the fit of a Linear Mixed-effects Model (LMM; explained in the Data Analysis Procedures section) and control for site-specific conditions. Stand data will also be investigated independently of the LMM in relation to shade and treatment combinations." This description is written in a way that is not uniquely replicable. Is this a separate analysis from the LMM analysis that will "test" for differences among treatments? That is, is this a post-hoc analysis looking for potential explanations? Or is this a description of what will be done in the single LMM analysis that will formally test treatment effectiveness? If the latter, then the process for deciding which variables, and the function of each variable (linear, quadratic, log-scale) to be included (or not) should be described.	Under a newly-inserted subsection, Model Selection, describes the procedure for creating and selecting among models. This model selection procedure allows for comparison of different LMMs. We are not advocating for a post-hoc analysis that searches for potential explanations nor is this a description of what will be done within a single LMM.	Lines 607-608: "If they are important, they will assist with overall model fit....." I am unsure if this was meant as a statement about the unknown 'truth' (aka unknown statistical parameters) or a statement about the data analysis and estimated model. The statement is true if the authors meant the unknown truth but it's not true about the estimated model. In the analysis, adding any variable to the estimated model will always improve the fit - even if the variables are not important - so using model fit is not always the best way to choose a model. Lines 612-628 provides context but would be stronger with addtional details. The new section doesn't provide specifics about the decisions that will be made to identify the models that will be compared (Lines625 to 628) On what basis will models be proposed or the covariates to be included in the core model identified in the a priori process? This isn't included in the text. Since *a priori* hypothesized models/covariates are based on previous research so the models could potentially be identifed now and included in the proposal. If that's not possible some description of **how** they will be identified prior to examining the data would be best -right now it just says they will be identified. I refer the authors to a later Burham, Anderson and Huyvaert paper : Behav Ecol Sociobiol (2011) 65:23–35 that describes both *a priori* and *post hoc* use of AIC. If models will be proposed after looking at the data that process should also be described. With respect to proposing a prior models, the idea is to find a subset of models smaller than the set that includes all possible models or even most models. I appreciate the last sentence in this section - it's very informative, thanks.
Forest and site characteristics	**3**	Pg. 5, stand characteristics section	*"Stand data will be used to improve the fit of the LMM and control for site specific conditions"*. Explaining changes in shade with stand data is not the same thing as improving the model fit. Model 'fit' (observed data minus modelled response) will always be improved by the inclusion of any additional variable, so specifying the amount of allowable lack of fit is needed if you're trying to improve model fit. But I don't think you really intend to find a 'best fitting' model so much as want to know if stand data are correlated (to what extent?) with shade. This step is exploratory and it may be that graphical analysis is adequate compared to attempting to add stand variables to the LMM (which is complicated). By including such variables you may reduce the variation but induce bias and overfitting the bias/precision tradeoff in modelling. You may need to revise your wording here.	The text did appear to indicate that a post-hoc analysis will be performed, which is not our intent. The text has been changed to "Stand data will be used to control for site specific conditions." Also, please see the new text referenced in the above response. The model selection procedure avoids overfitting models and post-hoc data exploration.	See my comment on line 62. There are not enough details in the new model selection section to identify on what basis or which models will be chosen a priori. It just say they will be chosen. The revised sentence is better but it suggest that ALL stand data will be used. If there are specific site conditions which need to be controlled for they should be identified. 'Site'' is in your LMM as a random effect and will account for some of that variation. It would be helpful to identify what what conditions should be separately accounted for over and above that.
Forest and site characteristics	**3**	Pg. 5, stand characteristics section	Please be specific about how you will 'control for site specific conditions'. This is generic, there are many options for how this could be done, so this language is not specific enough for someone to repeat what you plan to do. Please also specify what you are controlling for. There will always be differences among the sites and plots that exist by chance so identifying what needs controlling for ***a prior*** i is important. Otherwise, it borders on data snooping.	Agreed. Please see above.	The text is more clear but I don't see explanation for which variables will control for which site specific conditions, or alternatively for how it will be determined if a condition has to be controlled for.
Forest and site characteristics	**3**	Pg. 5, stand characteristics section	The sentences refer to 'statistical model variance' (singular). In the LMM there will be many types of variances – variance among sites, among plots and among the smaller units within plots that receive zone width levels. I don't understand what is meant by accounting for model variance using stand data. Please explain what you intend to do.	Changing to: "Stand data will be used to help account for changes in shade in response to the treatments~~,~~ **and** variation in shade response among ecoregions~~., and the magnitude of statistical model variance~~	Excellent - thanks.
Forest and site characteristics	**3**	Pg. 7, general predictions and figure 3	Figure 3 and the text for the general predictions are not congruent. There are multiple versions of Figure 3 that would match the general predictions. I don't know which prediction (text or Fig 3) is the one you intend. Are the magnitudes of the differences among the zone widths for each harvest intensity also what you are hypothesizing (e.g. an exponential decrease in shade with over zone widths)? Please clarify.	Figure 3 and related text have been deleted.	Ok.

Forest and site characteristics	**3**	Pg. 7, Figure 3	Figure 3 is hypothetical, so I assume the error bars are also hypothetical. Are you trying to depict that you predict that the standard errors of the estimated treatment means will be constant? They may not be. If you estimate the mean for the 100' zone width it will have 3 times the replication of the other widths so its SE will be narrower. If this is hypothetical and really is a depiction of the TRUE means, no error bars are needed. Please clarify.	Figure 3 and related text have been deleted.	Ok
Study design - site selection and study population	**3**	Pg. 8, Site selection criteria	The listed site selection criteria on page 8 leave out a very important criteria - the landowner must agree to participate in the study. Please add this to your list. There are many reasons why a landowner may decline to participate – some of which could bias your sample away from the set of all streams in the targeted ecoregions that meet the criteria.	The site selection criteria is based on technical specifications. The text acknowledges that we must have landowner cooperation for access to study sites.	An important assumption, and likely an accurate assumption is that landowner cooperation is not correlated with shade conditions.
Study design - site selection and study population	**3**		Since you note that it's possible that there could only be 5 such sites in an ecoregion that match the selection criteria, the scope of inference could be limited – at least by the site selection. Conclusions that infer that results apply broadly would be incorrect and therefore statements about the presence or absence of effects should be carefully worded.	See below. Text has been added about the scope of inference.	Thanks.
Study design - site selection and study population	**3**	Pg. 15, QA/QC section, first bullet. Pg. 20, project risk analysis section, study scope paragraph	I suggest you make the scope of inference explicitly clear, that is, to what population the results of this study will apply. On page 15 it says that field inspection will confirm that sites are representative of the study population, but this population isn't clearly and completely identified in one location in the text. On page 20 it says "The findings may be interpolated within the range of the treatments but cannot be extrapolated outside of that range with great confidence" but this is not strictly correct. The findings will only apply to the environmental conditions **and** treatments in this study. Pages 20-21 do not refer to the limitations imposed by the site conditions and sampling of sites – it probably should.	See Study Area/Study Scope/Risk Analysis sections for related changes to site selection process and scope of inference.	Thanks for the addtion.
Study design - site selection and study population	**3**		Please clarify if there is a minimum distance between sites, whether or not different sites can be on the same stream, or within the same watershed. GRTS will help spatially disburse if you have many potential sites but what will you do if the geographic extent of potential sites is severely limited? Will you allow sites to be very close together? Some description of that situation would be helpful.	The text has been modified to clarify the randomized site selection process. There is no minimum distance between sites. The Project Risk section addresses your remaining questions.	Additions addded clarity - thanks.
Study design - site selection and study population	**3**		I appreciate the discussion of all the ways that sites can vary. It's important to acknowledge that as you constrain the sites to be more similar you are reducing the range of variation over which you can draw conclusions. This is a never-ending tension between enough variation to allow broad conclusions with associated large sample sizes versus narrow variation with narrower scope and smaller sample sizes. So, making sites similar will improve your precision but will decrease your scope of inference.	Agreed, thank you. Relevent text has been added.	Thanks. Having the figure include the intervals helps.
Study design - site selection and study population	**3**	Pg. 8-9, Experimental unit section	The plots are identified as the 'experimental units'. In this study plots are experimental units for the levels of thinning intensity and sections of plots are the experimental unit for the zone widths. In a strip or split plot there are multiple types of experimental units. The first 2 paragraphs of the section are about the plots, so my suggestion is to not refer to experimental units at all. See the figure at the end of this document that schematically identifies the different sized experimental units within this study. A defining feature of split and strip plot designs is the existence of different sized experimental units.	Agreed, revised and reorganized related text accordingly.	Thank you. The revision is much clearer
Study design - site selection and study population	**3**	Pg. 12, sample size section	The text does a good job explaining numbers of physical 'things' (regions, sites and plots) in the study and this is useful. The number of treatments needs revision because the document confuses the fact that the 100' width doesn't have an adjacent thinning intensity with the fact that all levels of width are assigned with plots assigned to a thinning intensity. So, there is a combination of the 100' width inside the plot with a thinning intensity of, for example 40. If the analysis will be done by with one LMM that includes ecoregion, then it may be helpful to accumulate replications across ecoregions. Note that the number of physical units isn't the same as the number of values (replications of an effect, aka a difference) used in the statistical analysis. So, the statistical replication (what is reflected in degrees of freedom) will be different.	The text has been changed to reflect that the main LMM analysis will not include the 100' no-harvest buffer data; those will be analyzed in an ancillary and simplified LMM. It is true that the number of physical units does not equal the number of values in the LMM. If the project proceeds with a strip-plot design then LMMs with the strip-plot analysis structure will consider the data set as having effectively three sample sizes based on the random effects structure.	Thanks, revisions are good.
Study design - site selection and study population	**3**	Pg. 12, last sentence in site layout section	Last sentence in Site layout section on page 12: *"The three different plot-level treatments (harvest sequences) will be randomly assigned to plots 1, 2, and 3 within a site, and will not necessarily be assigned in the order depicted in Figure 1.* " I found this sentence confusing (treatments confused with levels and not clear relationship between thinning intensity and zone widths. I suggest "One of each level of thinning intensity (40, 20, 0) is randomly assigned to each plot at each site. Within each plot, the specified level of thinning intensity is applied sequentially in time to create a sequence of the no-harvest zone levels. The sequence of no-harvest zone levels is in the same order in all plots." I also suggest adding this sentence to the caption in Figure 1.	Revised text accordingly to improve clarity.	Thanks, the revision is much clearer
Study design - site selection and study population	**3**	Pg. 13, sentence that begins "Bankfull width"	"*Bankfull width and bankfull depth will be measured at the midpoint of each plot*". Please rephrase this. The midpiont of a plot is a 2-dimensional measurement and not in the stream.	Revised text accordingly.	Thanks.
Study design - site selection and study population	**3**	Pg. 13, understory Veg cover section, second paragraph	"*A set of four oblique digital photos will be taken from the central photo point associated with each plot (Figure 1) to provide a visual record of site attributes, including understory vegetation cover (Table 3). Four photos will be taken from each point at 90° intervals (upstream, downstream, left bank, and right bank).* " Please include the temporal context for the photos as well – WHEN are the photos taken? This is an example of attending to spatial issues without needed temporal context.	Revised text accordingly.	Thanks.
Study design - site selection and study population	**3**	Pg. 13/14, effective shade section	"*The photo points will be located at a consistent distance from the plot boundary at a manageable depth (~<1 foot deep), to be determined after study sites are selected* " Please clarify what is meant by 'depth' or 'deep'. Depth is usually measured vertically but I think this is referring to a horizontal measure but that is unclear. Sentences further along describe the height above the stream at which photo points occur so I'm assuming this is not a measure in the vertical direction.	Thanks for catching that. Text has been revised to clarify this refers to vertical water depth.	Thanks.

Study design - site selection and study population	**3**	Pg. 14, stand characteristics section	"*After boundary marking……"* Would more clearly be "after plot boundary marking…… The information here should be coordinated with the information about the 100% inventory on page 5. See my other comments regarding this.	Revised text accordingly.	Thanks.
Study design - site selection and study population	**3**	Pg. 15, harvest layout and implementation section	A suggestion for your consideration is to measure the tree density in the field after the various harvests. Variation in harvesting or even errors in the field can lead to tree densities that are not what you intended. Having a field measurement could be helpful if you find you have to "adjust" for the actual density in the plot.	What Reviwer 3 describes matches what is being proposed. As stated in the document, "The tag number of each harvested tree at each treatment interval will be recorded so that stand characteristics (e.g., basal area by species) can be computed for the harvest and no-harvest zones for each interval." …. "After each thinning treatment, follow-up inspections will be conducted to ensure that all trees marked for harvest were harvested or at least felled. Additional harvest may be required to meet this standard. Additionally, any unintended tree falling or damage that occurred during the harvest activities will be recorded by tag number."	Thanks for pointing me to that.
Quality Assurance and Quality Control	**3**		Please include direction on how to record missing data on the data sheets.	No change needed. As stated, "Field staff will be instructed to take detailed notes and photographs to document any anomalous situations." More detailed instructions are reserved for the forthcoming Implementation Plan/Field Manual/SOPs based on this Study Design.	Good to know that there will be formal SOP documentation proces and instructions that cover this.
Quality Assurance and Quality Control	**3**	Pg. 14, second paragraph	One bullet point says, "Hemispherical photographs will only be collected under uniform sky conditions when the sun is not directly in view, and according to the camera manual" and there is a somewhat different sentence on page 14 that says, "Photographs will be taken when no direct sunlight is visible, at pre-dawn, post-sunset, or under an evenly overcast sky." I suggest you decide which is most specific and use common language for both. There are conditions under which the sentence could contradict each other.	Deleted inconsistent/redundant sentence and kept the earlier, more specific sentence.	Thanks.
Quality Assurance and Quality Control	**3**	Pg. 16, eight bullet	*"Field datasheets (digital or hard copy) and digital photos will be duplicated and stored in a secure location as soon as possible following completion of each field survey"*. This is a fairly generic statement which may suggest that this process is not well-defined. More details could generate greater confidence that is critically important step will be well-executed.	Deleted. This detail will be expanded upon in the forthcoming Implementation Plan/Field Manual.	Thanks. It wasn't clear to me that there would be a formal plan and process for documenting procedurs.
Project Risk Analysis	**3**	Pg. 21, study design assumptions	This section correctly identifies why the design is not a split plot design and the issues with assuming that it is. The strip plot analysis may help.	Reviewer 3's recommendation of a split plot design does not address the issue described in the section on study design assumptions. Both split-plot and strip-plot designs, conducted as intended, require randomized application of the two treatment levels. A strip-plot involves the application of one treatment, thinning density, within a plot. The second treatment, buffer width, must be applied by first starting with a 100' no-harvest zone and sequentially harvesting inward. We cannot stand trees back up and randomize the order in which we observe the effect of no-cut harvest width. This violation of randomization occurs regardless of whether we consider the study design a strip-plot or split-plot design and has important implications for interpreting study results. Reviewer 3 is correct that the study could also be analyzed as a strip-plot design. This situation of being able to choose between a split-plot or a strip-plot design is an artifact of our inability to randomize the buffer width treatment order. Since buffer width could not be randomized, it becomes a philosophical debate about whether the true form should be a split plot or strip plot design. Both have merit. For strip plots we sacrifice power for estimating both treatments (main effects) instead of one and improve our estimates of the interaction effects of the two treatments. We appreciate the recommendation and will adopt the strip-plot design, but the adoption of the strip-plot design does nothing to ameliorate the	Agreed. The split plot approach is sometimes chosen because analysts know how to do it and are familiar with it. It can be a reasonable approximation.
Project Risk Analysis	**3**	Pg. 21, Site availability and sample size section	Small sample size is a possibility. Please provide ***a priori*** i determined biological or management-relevant differences. Statistical significance is a function of the study design (larger sample size will always decrease the p-values) so it is possible that non-significant results are obtained even if the absolute magnitude of the change is relatively large or that a tiny change in shade could be statistically significant due to a large sample size. Therefore, it is advisable to ***a priori*** determine important changes and state that in the anticipated results. This way even a statistically non-significant value that is quite large can be identified and discussed. More importantly, if confidence intervals are wide and encompass both important and non-important values, it will be important to discuss the reasons for this and what is needed for future work.	The purpose of this study is not to determine whether change has occurred (based on an arbitrary p value level or a pre-defined management/biological difference) but instead to estimate the magnitude of observed change and associated uncertainty around estimates. This study represents neither effectiveness monitoring nor compliance monitoring. This study estimates the effects of riparian harvest treatments on stream shade. There are few similar studies and none with this level of replication. As such, there is uncertainty about how precise estimates will be and the site features that drive levels of precision. We are in agreement that we should discuss (and analyze, based on the data collection efforts described in the document) sources of observed error around estimates.	AIC will be used. Similarly to p-values, AIC can be affected by small sample sizes - that is, fail to detect differences among models when the models really are different. the proposed apriori model identification will help but identifying what makes a model be practically or biologically relevantly different could be helpful. No changes requested here but important to consider as you do the model selection.
Project Risk Analysis	**3**	Pg. 21, Site availability and sample size section	I appreciate the contingencies for adding additional sites. See my earlier comment about required spacing between sites. The same goes for plots. Please carefully consider how some of potential alternations might change your scope of inference and hence the usefulness of the study. Adjusting site criteria 'on the fly' can be particularly problematic if not carefully considered so I appreciate the comment that such changes will be carefully considered. But in the interest of reproducibility, further criteria should be identified ***a priori*** .	It is typically very difficult to locate study sites and landowner cooperators within our research program. A further limitation is that we are unable to select sites until after the study design is finalized and approved by the presiding committees. Thus, we intentially do not want to limit our options at this stage, to prevent falling short of our sample size requirements. However, we do plan to carefully select sites based on our criteria. Note that the 100% stand inventory data will allow us to describe the overstory conditions in detail.	Thanks for the clarification.
Project Risk Analysis	**3**	Pg. 21/22, variation in site conditions	I appreciate the discussion of all the ways that sites can vary. It's important to acknowledge that as you constrain the sites to be more similar you are reducing the range of variation over which you can draw conclusions. This is a never-ending tension between enough variation to allow broad conclusions with associated large sample sizes versus narrow variation with narrower scope and smaller sample sizes. So, making sites similar will improve your precision but will decrease your scope of inference.	Thank you, agreed.	Thank you.
Project Risk Analysis	**3**	Pg. 22, second paragraph	On p. 22 at the end of the 2nd full paragraph it says that the re-orientation of the streams to standard orientations (N/S, E/W) *"will ensure that shade response to the treatments is not influenced by differences in stream orientation across sites"*. I've noted this earlier but it comes up here too - I question whether this is biologically reasonable. The under and overstory growth on a site is a function of the site and environmental conditions at a site and that includes the orientation of the stream relative to the tree stands. Re-orienting the image doesn't change the *in situ* conditions that affect plant growth and vigor. So, while some conditions may be standardize by the re-orientation, others are not. I suggest you clarify what conditions can and cannot be standardized.	The text has been re-worked throughout to clarify what non-treatment factors will be reduced/eliminated by photo post-processing procedures. The goal is to measure change in shade, all things being equal (by blocking by site), for our stream orientations of interest. The 100% stand inventory data will allow us to account for and explain differences in shade response that are related to overstory composition and structure, regardless of actual stream bank direction in the field.	Thanks.
Statistical analysis	**3**		The study design better fits a strip-plot analysis than a split-plot analysis. See schematic submitted on last page of reviewer 3 review comments.	See comments above.	Yes, thanks.

Statistical analysis	**3**		The references to factors, levels of factors and treatments are not always accurate	The document has been edited to provide better clarity. The word "treatment" unfortunately has different meanings when used to describe statistical and operational considerations.	Thanks for the clarifications. The new text is more consistent in its use of the words. I found it to be more clear.
Statistical analysis	**3**		Attention needs to be paid to the assumptions underlying the application of linear mixed-effects models, particularly diagnostic and remedial approaches (the latter in the case that the assumptions are not met).	Agreed. Text has been added: LMM assumptions will be tested following tests described in Pinheiro and Bates (2000). If assumptions are violated we will strive to correct them.	Thanks.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	A general comment is that much of what is here has already be said earlier in the document but with different language and different emphasis and some details missing in previous sections. Earlier sections on analysis did not include all details and thus left the reader with gaps in understanding about the analysis and flipping back and forth between sections was needed. I suggest writing a single analysis section that incorporates all the parts using consistent language and terminology and then writing a summary of it if needed for the introduction. That would improve consistency in the language and help make sure that relevant details are provided earlier.	Agreed, the document has been substantially restructured accordingly.	Yes, revisions do a great job of clarifying
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	This section identifies the response as the CHANGE in effective shade. This is not consistent with earlier text or what is stated in the Objectives or Critical Questions on p. 1 which identifies effective shade as the response. Please reword to accurately describe the questions as understanding if the difference in effective shade responds to harvest intensity or zone width etc.	No change needed. The current wording is compatible with what is being proposed - 'change' and 'response' have very similar meanings in this context. For the purposes of readability, the purpose statement/objectives/critical questions use the more general 'response' and the exact definition of this ('change' in shade) is described as appropriate later in the document (along with a descriptive math equation)	I understand the authors' response. Since it can be confusing for readers to have to parse different meanings to 'shade response', 'shade','change in shade' and 'response' so my suggestion was intended to help the reader by removing the need for them to figure out precisely what was meant.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	Given that the response is a difference from pre-harvest conditions, it is important to have measurements for the 100' width associated with EACH plot at each site as I noted in comments regarding Research Approach.	That is the plan. See difference in effective shade (ΔES) equation (individual plots denoted by 'j')	agreed.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	Shade values will not be "normally distributed". Please explain why you are assuming this and why it matters. Proportions or percentages around 50% can be approximately normally distributed and the LMM doesn't require that the response be normally distributed, only that the various estimated residuals be normally distributed. Be sure you know how to check this assumption because the assumptions could fail even using the differences.	The text states, "Shade values will not be normally distributed; however, the differences in shade values will be approximately normally distributed." The text has been changed such that a mean, not median, of the five shade measurements will be taken for each plot. Means are normally distributed and differences in means are normally distributed.	Ok. But technically what is required is that the model residuals are normally distributed - the response is not required to be.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	The LMM is described as having a fixed effect of ecoregion. This is reasonable. But it contradicts what is stated about analyses on p. 6 (Data will be summarized and analyzed according to ecoregion) and on p. 22 (Data will be analyzed according to ecoregion). Please correct or add to describe the multiple types of LMM's that will be used.	Deleted earlier sentence: "~~Data will be summarized and analyzed according to ecoregion.~~"	Thank you.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	How will the constant variance assumption of the LMM be checked? The LMM analysis is going to a priori assume that the variation is constant over ecoregions or over sites. This needs to be investigated and the model should be altered to accommodate heteroscedasticity if that occurs.	The constant variance assumption will be checked by examining residual plots. If it appears that the assumption is violated, the models can be adjusted to account for heterscedasticity (see Pinheiro and Bates 2000). Relevant text has been added to the Analysis section.	Thank you.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	For reproducibility of analysis, what is the analysis plan if assumptions for the linear mixed model fail?	If the assumptions are violated we will strive to alter the analysis so that they are addressed. This may include modeling heterscedasticity, splitting data, inclusion of variables, etc.	Thanks for thinking ahead.
Statistical analysis	**3**	Pg. 18, Analysis section, third paragraph	Paragraph 3 refers to the design as a split plot with blocking. That is a bit misleading because this entire design is more complicated than a text-book example of a blocked split plot design and as noted earlier it's not a split plot design. 1)Ecoregion is a fixed effect and the sites within ecoregion serve as the replication of the ecoregion (that is, 4 degrees of freedom to test differences among ecoregions). The ecoregion fixed effect exists in the LMM at a scale in the model higher up than the random effect referred to in the document as the 'block' and there is no blocking factor associated with the fixed effect of ecoregion. 2) The whole plot factor (thinning intensity) is randomly assigned to each plot within a site and the split/strip plot factor (zone width) is sequentially applied within each whole plot (plot). So yes, sites serve as a random effect (aka grouping factor, block, 'error' term) for the whole plot factor while the plots serve as the random effect (aka grouping factor, block, 'error' term) for the strip/split plot factor. 3) As mentioned earlier, the split plot factor (zone width) is 'stripped', not 'split' since the strip/split plot factor is applied in the same order in every plot. In a strip plot design there would be a random effect that accounts for the grouping of all plots at the time intervals (since the zones occur in the same order in all plots) which forms that basis for the test of the main effect of width, and there would also be a random effect (the final 'error' term in the model) on which the interaction of the zone and intensity is assessed.	We have changed the language of the document to indicate that models will include a strip-plot model (split plot would work too) parameterization with additional variables such as ecoregion included. Please see new text. The purpose of the second point is unclear, except to indicate that we are in agreement regarding the assignment of random effects (at least to a split-plot design). We believe the structure of the strip plot described in the third point is addressed in the new analysis text.	Thanks. My comments were only for clarity. Please see the sentence on line 524 to 535. It says that the model will include a random effect of the thinning treatment nested within the site (in addition to a random effect for plots nexted within sites). Thinning is a fixed effect - there doesn't need to be a random effect for thinning nested within site.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	This is a complicated analysis. There is also the potential that some random effects may not be estimable if they are numerically confounded with other terms. So, I suggest involving a statistician to help program and interpret the analysis.	Thanks. A statistician has been consulted during the development of this study design, and has helped write and revise the analysis section. Continued assistance from a statistician will be considered for the analysis phase.	Great - glad to know.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	This section suggests that ecoregion will be included in the LMM so that ecoregion and its interactions with zone and intensity are analyzed in one statistical model. However, on page 6 it says "*Data will be summarized and analyzed according to ecoregion*". This is different than what is in this section. Please determine which you will do and consistently explain that.	Deleted earlier sentence, "~~Data will be summarized and analyzed according to ecoregion.~~"	Thanks.
Statistical analysis	**3**	Pg. 18, analysis section, last paragraph	The last paragraph on p. 18 says that you will 'test' if ecoregion improves fit. You can test if there is a statistically significant effect of ecoregion but that is not the same test as the test for model fit. I suggest you don't test model fit. The LMM is generated based on *a priori* identified design structure (including ecoregion) which means the LMM has inherent orthogonality build into it (that's why it's a 'design') and you should not remove any of the design variables from the LMM even if they are not statistically significant. This is not a setting in which model fit is an issue. Specifically, you say "Contrasts will be examined to statistically compare different treatments and treatments by area" - by which I think 'area' means ecoregion. It could be that the author used the word 'fit' when they actually meant statistical significance.	The sentence states: "[…] analysis will test whether including ecoregions in the model improves model fit by comparing models that do and do not include the ecoregion variable." Please see the response above and new text discussing model comparisons. It is absolutely feasible to compare the performance of several models which have different parameterization so long as they utilize the same dependent variable. The "test" is less a classical p-value driven test than an evaluation of model performance relative to other models. A model comparison utilizing AIC (or AICc, or QAIC, or BIC) takes into account degrees of freedom and model performance. Contrasts are a separate tool, available for estimating within models the difference between factor levels.	We agree. My points were (a) text says you'll test if adding ecoregion improves fit and that's a different test than testing if there's support in the model for ecoregion. I think you want the latter. Current statistical thinking is that variables that are part of the design should always be included in the model since they usually imply a grouping (correlation among observations in one group) or a restriction to the randomization.
Statistical analysis	**3**	Pg. 19, analysis section, #2 and #3	The analyses identified at the top of page 19 (#'s 2 and 3) make use of the site metrics (continuous variables). This analysis should also include the random effects used in the #1 analysis to accurately account for the various 'groupings' for sites within ecoregions, plots within sites and subplots (for zone width) within plots. Again, this could be complicated by numerical effects so involving an experienced statistician could be helpful.	This comment is interpreted as meaning that the same random effects structure as was used for analysis 1 is employed here, and not that the actual random effects are reused. The random effects structure for Analysis 2 may differ from Analysis 1. Analysis 2 seeks to examine how stand metrics relate to shade, not the different treatment levels. Therefore the random effects structure could be a random effect for site and another for plot nested within site. Analysis 3 is strictly exploratory and will make use of random effects as fit the questions being asked.	I think we're on the same page. My comment was interpreted correctly. If the response variable for analysis 2 is measured on the same 'unit' (plot with site) as the response in analysis 1 than the random effect structure would be similar. But if the response in analysis 2 is measured on a different type of unit (e.g. a subplot within a plot, or at the site level) the random effect structure would differ from analysis 1.

Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	This section refers to 'predictive' equations suggesting that you intend to apply these results to streams and locations beyond the ones in your study. It is important to clarify if you intend to predict estimates that pertain to the streams you sampled, estimates for other very similar streams represented by your scope of inference (including owner permissions) or estimates for other streams not represented in your sample of 20 plots. The 'predictions' you generate, and more importantly the standard errors are a function of the estimated random effects in the model. Random effects are known to be poorly estimated unless the sample size is very large so concluding that the standard errors, confidence intervals or prediction intervals from the analysis apply beyond your sampled sites could be tenuous. You may want to consider not 'predicting' anything but simply reporting the estimates from your study.	Random effects within LMMs are assumed to be normally distributed and centered around zero. The analysis will verify these assumptions. Random effects estimates are specific to levels of nonindependence that the model accounts for and are not useful outside of the dataset at hand. We are in agreement on this point. However, the fixed effects estimates can be predictive. At the conclusion of the study the analysis may very well provide information on expected (predicted) responses for stands in certain regions and having specific characteristics. We strongly disagree with Reviewer 3 on this point. This study is part of the adaptive management program. The results of this study will indeed result in one or several potentially useful statistical models. These models form a basis of understanding system function. As the adaptive management program proceeds, these statistical models can be used to predict outcomes, and those predicted outcomes can be compared against future observations. Reviewer 3 suggests we consider "not 'predicting' anything but simply reporting estimates". We counter that by ignoring the information provided by this study CMER would squander an opportunity to learn from this study and refine future monitoring efforts.	The measure of precision for your estimate will depend on the type of prediction you are doing - different types will use different random effect variances in the estimation of the standard error. The random effect variance is poorly estimated with small sample sizes so the precision of some fixed effects could be small.
Statistical analysis	**3**	Pg. 18/19, Statistical analysis section	For reproducibility, please describe how the assumption that changes don't differ by initial shade will be tested, and also how the 'model will be adjusted accordingly' since there are multiple ways to do this.	We have changed the text to address these points.	Thanks - revisions are more clear.
General comment	**3**		Thank you for the opportunity to review this study design. The investigation incorporates a complex study design with multiple spatial scales and temporal considerations, responses based on hemispheric photography, a proposed sophisticated statistical analysis (mixed model methodologies) and additional covariates. It is reasonably well-described but would benefit from revisions to more clearly explain the methodology with consistent terminology for the wide range of experts who are likely to read the description.	Thank you, revised accordingly.	The revisions are all excellent and the document is much more clear now, thanks.
General comment	**3**		This study is spatially and temporally complicated. There's a tendency to focus on either space or time in the various sections of the manuscript but not both in any one section. An overriding suggestion is to be sure that both space and time are incorporated in each section.	Agreed and thank you for catching that. The document has been reorganized to clarify the time and space components.	The revisions are all excellent and the document is much more clear now, thanks.
General comment	**3**		Please use accurate language around factors, levels of factors and treatments accurately - it is not always done in the document. A treatment is a combination of levels of multiple factors. I have more explanation the specific comments sections.	Revised throughout accordingly.	The revisions are all excellent and the document is much more clear now, thanks.
General comment	**3**		The response is identified as 'effective shade' but in the analysis section it is identified as a change in effective shade relative to pre-treatment values. This needs to be clarified and made consistent throughout the document. Most designed studies only estimate "changes", that is differences between treatments or between levels of factors. But if the response truly is a change from pre-treatment values then the language needs to reflect that differences between treatments are estimated after accounting for pre-treatment values.	Edits have been made throughout to improve clarity. The specific definitions are detailed in the Analysis section: "The main analysis response variable will be the difference between shade values for the nine different treatment level combinations (three no-harvest zone widths [the 100-foot no-harvest distance will be excluded] and all three thinning levels) and the original pre-harvest plot-level effective shade values. Shade values will not be normally distributed; however, the differences between mean shade values will be approximately normally distributed. The treatment level combination values will be subtracted from the original shade values to control for the initial differences in shade among sites." Difference in effective shade (ΔES) will be computed as: Δ⟦ES⟧_hijk=⟦ES⟧_hij0- ⟦ES⟧_hijk"	Thanks for the clarifications. The first of the 3 analyses (line 548) is to determine how treatments affect shade (not change in shade) and the description says that means will be estimated from the LMM. So should this section be revised to refer to change in shade rather than shade? Since the idea is that shade is not normally distributed but change in shade is, estimating shade from the LMM might not be appropriate.