

Title					
Color coding>>	YELLOW = critical considerations to be addressed	impacting the validity of the science; NO COLOR = suggested to improve the science but not mandatory; GREEN = suggested editorial corrections, not mandatory.			
Comment topic	Reviewer	Location: Page/Line in original doc	Reviewer comment	Author response	Reviewer response to author action
No metadata for 20yFHP	1	Q1	the construction of the best model, the 20yFHP, is effectively a mystery. The manuscript would benefit from the inclusion of the methodology used for creating this map	Yes, agreed. The methods of the 20yFHP are somewhat murky in reality. All available documentation of methods for its construction are somewhat vague and reference an ILAP update by "Henderson 2012". The document for the Henderson 2012 is not available in the literature. Official reports of the ILAP map construction and use are available on the USFS website: <a href="https://research.fs.usda.gov/treearch/47221">https://research.fs.usda.gov/treearch/47221</a> . The relevant information has been included in the Data Suitability section of Phase 1, step 2. We have also included more description of why our exploratory analysis of the 20yFHP was so extensive (because its methods of construction are so unclear). Link has been updated to direct reader directly to methods and metadata documents.	
Reconnaissance data	1	Q1	Convenience sampling is not statistically rigorous as it creates inferential ambiguity. The 20yFHP map interpreted as "truth" based on this road survey and other maps/models were derived from it. Retrospectively, the map should have been rigorously tested or a new map developed. Given this sampling approach, the manuscript would be more complete if it included the analysis of the reconnaissance data that confirms the degree to which these reconnaissance data agree with the 20yFHP map category types or match their assigned category type	Yes. Agreed. However, rigorous validation of the 20yFHP map was outside of the scope of our project. We were tasked with developing an alternative framework using available data. This issue is addressed in the discussion section, and we recommend a more formal evaluation of the map for future directions. This is why we discuss these limitations throughout the document. We recognize these methods are not ideal, but given our objectives, it was the best possible approach. We have removed the phrase "truth".  The reconnaissance data, as R1 has noted, is not a rigorous or scientifically sound sampling approach, we recognize these shortcomings as well. Because of the shortcomings of the recon data set, we are apprehensive to use this dataset in a formal evaluation of the 20yFHP, because it would contain these potential biases. This was not the purpose of the recon data. The recon data was collected over a relatively short period of time, to gather as much information as possible about the types of habitat covering as much of the study area as we could. Yes, this does amount to "convenience sampling", but we needed some actual field data with researcher observations of actual habitats present in the study area. This step is a classic forest ecosystem classification approach to begin developing forest type categories that can be adjusted and refined as new data is collected using a specific scientific design (i.e. hypothesis development). We did not collect the recon data with any intention of using it in a scientific analysis, its sole purpose was to give the researchers an increased confidence in the available datasets that were being appraised in Phase 1.  Building on this question, the appendix contains a representation of how we used the recon data to visualize patterns of tree species dominance and associations (See the visual ordination Figure S1). We did explore the data with NMDS as well, however, again, we are apprehensive to include this in the report as we feel it would be misleading given the "quick and dirty" approach to how the data was gathered. Nevertheless, the data was useful for investigating patterns and understanding how forest groups could fit into those available in the 20yFHP.	
Variables & ML derived from sampling 2	1	Q1	the researchers unfortunately expended a great deal of effort to achieve a foreseeable conclusion. Much of the manuscript (anything derived sampling of the 20yFHP) is unfortunately not necessary and can, in my estimation, be removed	The conclusion was not as foreseeable until we tested it. We were tasked with developing a framework with the best available dataset. We explored several different datasets. We determined the 20yFHP map to have the best coverage and most relevant forest groupings. We developed refined frameworks from the 20yFHP map using several methods. We recognize that one of our approaches included building a model that was trained on modeled data (the 20yFHP). We recognize that the single factor frameworks were themselves a coarsening of forest groups that already existed in the 20yFHP. However, it was an opportunity to explore options for different framework parameters. We considered these "preliminary frameworks" or hypotheses that had the potential for refinement or adjustment once actual field data was collected (Phase 2). We found that our manipulations of the 20yFHP mapped categories, while somewhat useful, did not exceed the accuracy or utility of the original 20yFHP map. No, this was not surprising, but again, it did give us a better understanding of the 20yFHP map and data. It also created a discussion with the scientific advisory groups for the justification of future work in framework development and refinement.  See comments that address the 20yFHP map's development. Because its methods of construction were unknown, the in-depth exploratory analysis was necessary.	
upland and riparian distinction 20yFHP	1	Q1	It seems essential to know whether the forest compositions of riparian areas in the Eastside are the same as upland habitat elsewhere, or whether the categories are ill-suited for defining riparian vegetation. Additional field sampling to test the similarities of 20yFHP within-classification upland and riparian areas may be required to address this issue, or the eastside riparian forest type classifications could be created afresh from field efforts.	See updated paragraphs describing the 20yFHP in the Data Suitability section starting on line 567. While the forest classifications within the riparian areas are not explicitly labelled as "riparian" nor are they exclusive to riparian areas, the map does show categories that change with and follow riparian corridors. Some of these categories extend into the upland, while others show a transition into different classifications. Yes, considering riparian function, the neighboring upland forest types likely impact the riparian forest function. However, this was outside of the scope of the study. We were tasked with identifying riparian forests, their estimated functional responses, and their logical grouping based on these responses.	
spatial autocorrelation	1	Q1	If the machine learning model is kept in the manuscript, the authors should assess and control spatial autocorrelation as well as incorporate more appropriate model validation procedures to produce reliable models that avoided overprediction. See Ploton et al. 2020 and Roberts et al. 2017.	The ML model is removed from the manuscript. We only reference its use in the exploratory analysis of the relationships between the independent and dependent variables. To be clear, the map generated was a byproduct of the exploration of relationships between independent and dependent (forest types at multiple levels of aggregation) factors. This map was added to our list of potential frameworks worth evaluating in Phase 2.  The model development does account for spatial autocorrelation using an autologistic approach. The methods have been updated with relevant references. In terms of validation, yes, moving forward, if this approach is explored further, then much more field data would be needed and a more rigorous validation.	
small sample size of field sites	1	line 884	"This selection process is appropriate for calibration and assessment of the ML Map and ensures objective coverage of all forest type categories developed for the precipitation-based and elevation-based frameworks." This is an overly strong statement that needs to be explained and backed up with citations (the Hemingway and Kimsey citation does not exist in the lit cited section). Given the few points that ended up in specific categories and the small sample size overall, it seems like this sample size was too small for a robust assessment of any of the maps.	The single-factor frameworks have been removed from body and into an appendix. They were an experiment in maximizing utility at the expense of some accuracy. Unfortunately, the accuracy was far too low to justify. Hemingway and Kimsey citation updated. In response to the second part of this comment. Yes, in general, the sample size was too small for a robust assessment of the maps. We do not claim it is adequate, but appropriate. We discuss this in the discussion as one of the limitations in this study. We recommend a more formal and robust assessment of the 20yFHP map and have updated this section of the discussion for recommendations.	

Citations	1	Q2	WA DNR 2020 is a bizarre citation. It is RCW 76.06.200. I have a feeling the authors meant to cite the Box repository for the 20yFHP map (which, as I mentioned above, I did not find useful). Regardless, the methodology should be provided in this document	The references have been fixed for easier search and find, the RCW 76.06.200 tag is in the reference. For the 20yFHP data link in the data description section, we have updated the link to take it directly to the GIS_Data_Dictionary. This document is in the Box repository but it is the document which describes the methods for PVT development. Hopefully, much more useful.	
Reconnaissance data	1	Q2	the actual analysis methodology and results of the Phase I data are not presented. These are needed to understand how well the reconnaissance data align with the pre-defined forest type categories.	The reconnaissance data was not used for this purpose. As mentioned, the data was only used to help increase confidence in the available datasets and to get approval to move forward with the 20yFHP. Because this type of sampling is not thorough, and represents a potential bias, we did not use it in statistical analyses with a purpose of validating the map. The visual ordination available in the supplemental appendix show common forest grouping and timber species dominance. These were cross referenced with the groupings found in the 20yFHP and showed general agreement. This step was done as a preliminary investigation of actual field sites for the purpose of increasing confidence in these groupings. We include it here, because it was work done for the project and should be included. It holds little statistical importance and we are apprehensive to treat it as such. Thus, it is not a focus of the Phase 1 methods.	
4-category framework	1	Line 1262	In the discussion the authors come to the conclusion that "Ultimately, we interpret these results as evidence that a 4-category framework (PP, DRY MXD, MOIST MXD, COLD) may be the most functionally meaningful". However, the authors present findings that contradict this conclusion. The PP category had a 75% classification error rate (Figure 16) and the COLD/SE category has only 3 to 4 plots and a 50% classification error rate. I strongly urge the authors to reconsider the inclusion of these categories until they have more data for COLD and better validation that PP is at all a useful category based on field data-based confusion matrices	This is a statement about how the different forest types function (i.e. LW recruitment potential and shade). Yes, the predictive accuracy of the 20yFHP for PP showed high error, but please remember our objectives of creating a framework that separates groups based on function. The results from the LW and Shade assessment showed divergence of the PP group in both assessments (Shade and LW) from the other "DRY" forest types. We interpret these results as evidence that PP dominated stands likely provide effective shade and LW at different magnitudes and frequencies than other groups. In other words, PP should likely NOT be combined with the DRY MXD and DRY DF as one large "DRY" group when considering function. Included a call back to these figures in the text to help clarify (Figures 19 A and C, and 20 A) and updated the language to be less assertive.	
sampling 20yFHP map	1	Line 1300	The authors state around line 1300 that creating the ML map from the 20yFHP map forest groups was "not a traditional approach, and is not recommended". The stated conclusions would be stronger if they said what a recommended, traditional approach could look like	Yes, good point. This paragraph has been updated to include traditional tree-based model training methods with an added reference from Liu et al. 2018.	
Citations	1	Q5	I recommend that the authors carefully validate all references.	Note taken. All citations have been accounted for and validated.	
No metadata for 20yFHP	1	Q5	Since this map (which is a model) was the best at classifying forest types, and better than the THT approach, it needed to be explained. How was it constructed? What assumptions went into it?	The description of the 20yFHP construction has been updated in the Data Suitability section at the end of Phase 1, Step 1.	
Clarification	1	Line 636	If communities were not homogenous, why were they excluded? Perhaps they represent a riparian type not captured by the model.	We have clarified this in the text, please see edits. If a community showed discontinuity then it was likely a transitional area and not a distinct forest type that repeats with considerable size. We did allow for a homogenous heterogeneity (e.g. moist forest types dominated by a mixture of species such as WRC and WH as codominants), but continuity was important from a functional perspective (e.g. effective shade and LW potential). Also, this framework will advise timber harvest based on common types (i.e. repeatable units). Finally, this data was not used to definitively create groups, but to hypothesize testable groups (for our Phase 2 approach) and to cross reference with available datasets to understand how well they translate to what we can observe on the ground.	
Clarification	1	Line 974	How do you know what forest category to put a plot in? Isn't that another model type? What if the fit isn't great, how do you decide?	The description of our forest typing methods has been updated in the Accuracy Assessment section. Yes, some plots did not fit perfectly in one type showing changes in dominance as the transect moved away from the stream. This is discussed later in that section "line 1080. We evaluated these split types separately, changing our methods (as described in the appendix). However, considering the purpose of the framework (Timber harvest) we conserved the forest typing that was most prevalent beyond the 30' "core-zone" as they were the areas with most potential for being harvested.	
Clarification	1	Line 1107	Previously stated as 5x10'. Meters or feet? Also, where is this subplot, since it seems like w/in 30' of the stream can be very different than outside of 30'.	Updated. Thank you for the note. We actually made two mistakes in this description, it should be 10' x 5' as the length is 10 feet along the central axis with a 2.5' width on either side. The language has been updated for clarification of the subplot layouts and a figure included.	
Phase II field sites	1	Q2	description of how the Phase II field site data were categorized as informed by the data in Step 1 would be useful. To what extent were there sites that would adequately fit in multiple categories? And what fraction of sites poorly fit any categories	Clarified in the methods section. See response to earlier comment.	
Uncertainties-limitations	1	Q6	It would be very useful to understand uncertainties and limitations arising from approaches mentioned in Question 2	Yes. this has been clarified in the methods section.	
Clarification	1	Q6	A discussion about the likely origin of the error in field data classifications, overlap in histogram distributions of classifications, and lack of NMDS separation of groups. I imagine the classifications are missing historical context (fire suppression, succession, plantings, climate change) that may matter a great deal	This is discussed on line 1069. We do not know the error in the field classifications but we discuss that it is potentially due to unaccounted for site factors. Inserted a parenthetical for an example.	
Clarification	1	Q8	Line 698: The claim is made that the model showed over 80% of the variation could be predicted using the 10 most important variables listed in Table 5. The percentages sum to 78%.	Good catch. Changed to "nearly 80%"	
Bias	1	Q8	Unbiased: I have some concern that there is bias from pigeonholing sites into predefined categories, but a greater explanation of the Phase 1 reconnaissance sampling methodology and addressing the points raised in the answers to question 2 would probably assuage these concerns.	Yes, thank you. We have updated the description and purpose of the reconnaissance data.	
Clarification	1	Line 362	It would be nice if an accuracy goal was provided (maybe it will be), a benchmark to know whether a framework was sufficient or insufficient. Hopefully, when Objective 2 is explained, this sort of goal will be described.	This is not for us to judge in the policy making context that this study is conducted. We were not given a goal for accuracy. Outside of scope.	

Clarification	1	Line 409	Drop "The High Elevation category" because it "contains only approximately 0.01% of FPA managed stream length."	But this is a description of the THT categories, and the High Elevation exists at this percentage. Disagree with dropping as it helps illustrate one of the shortcomings of the THT	
fire regimes	1	Q2	glad to see that the fire regime for sites is considered, although I recommend that the procedure for doing so be expanded upon	The fire regimes were retrieved from the LANDFIRE dataset (line 979).	
Clarification	1	Q6	paragraph at 1078-1082. An identification of which figures in Appendix III this paragraph relates to (I think it is S4 and S5	Inserted	
Clarification	1	Q8	Figure 16 DRY DF has 4 plots (84 plots total) but Figure 15 has 3 plots for DRY DF (83 plots total). Why the discrepancy?	Typo, corrected	
Clarification	1	Line 668	This is a weird definition of pseudoreplication. Spatial autocorrelation should be of greater concern.	Yes. removed. Incorrect usage.	
Clarification	1	Line 752	i.e., matched the 20yFHP map? This could stand to be clarified. Were the 20yFHP map categories deemed accurate from the driving around data collection? Is that what's meant here?	We make no assessment of the 20yFHP. The purpose of the recon was to understand range of conditions. Description of the recon methods has been updated to address.	
Clarification	1	Line 884-886	Seems a bold statement. Why?	Unclear what statement is being addressed.	
Clarification	2	Q1	I found the format of the document with methods interspersed made following the study a little difficult. It was hard to see how things fit together. Perhaps more detail in the flow diagram (Figure 2) would increase transparency and show how the different steps and components of the study fit together	Yes, this is understandable. However, this format was chosen to illustrate the sequential nature of our work. This is also the purpose for showing Phases and steps for each objective.	
Spatial autocorrelation	2	Q1	There may be issues with spatial autocorrelation when creating point grids in riparian areas to sample the maps of interest, but I don't suspect these would change results	Yes, this was accounted for in the methods. Methods have been updated in response to earlier comments.	
Statistics PCA	2	Q1, Line 588	PCA doesn't seem appropriate here, but you do use NMS which is appropriate. I think you should get rid of the PCA and not mention it. It is odd that your first mention of an appendix is Appendix III. Why isn't this referred to as Appendix I?	Good points. Yes, the PCA has been removed. For clarification, it was only used in the exploratory analysis of the variability of the environmental factors. The PCA used a different dataset that covered a larger portion of the study area than the available vegetation data that was used in the NMDS. However, yes, it does not add much to the results or our interpretation of the available data.	
Clarification	2	Line 224	How is usefulness determined here? Does it differ from accuracy?	Described in Phase 2, step 2. This involved discussions with scientific advisory groups about the ability of the framework to be used practically in the field and if the groupings were functionally meaningful (described in the modeling riparian function). I understand this is a vague statement here, but this is the executive summary section. Details are provided throughout the document.	
Clarification	2	Line 227	Functional relationships with what?	Riparian function. This is described in the first paragraph	
Clarification	2	Line 273	Although it did not improve accuracy, it does seem to have shed light on error which is valuable and potentially worth mentioning.	Fair comment. We talk about this in the discussion	
Clarification	2	Line 290	add space after (fish-bearing)	Done, thank you.	
Clarification	2	Line 340	Nice figure but the text is difficult to read and should be clarified.	Yes, good note. The vegetation groups have been added to the figure description. The image taken from the publication is itself of low resolution. We have attempted to sharpen the figure as much as possible. Vegetation groups have been described in the figure caption to account for low visibility.	
Clarification	2	Line 345	Thus far there has been no mention of Type-5 streams except in the beginning. How are these different from Type F streams in terms of their ecological setting and associated relationships with elevation and topography?	Type 5 is mentioned throughout the document (e.g., Site Selection line 866, within figure captions). But, the note of discrepancy is understandable. Updated the first description of Type 5 (i.e., fish-bearing streams that are also classified as shorelines of the state). The Type 5 streams are included in the analyses and field collection protocol. Any further mention of "fish-bearing" streams includes Type 5.	
Clarification	2	Line 362	Characterize what? I believe you are focused exclusively on composition. How does development fit in to the frameworks? Same for the next two objectives.	These objectives and questions are verbatim from the scoping document and thus, cannot be changed. However, the paragraph following the Objectives "line 389 expands, stating to characterize site, vegetation, and landscape features.	
Clarification	2	Line 402	I see you give more details on these datasets down the line, but you could give a little more info here and refer to them as existing predictive maps.	Sentence included at the end of this paragraph to give more information on what we took from these datasets.	
Clarification	2	Line 432	It would be valuable to include any information on accuracy assessments and method of prediction in these datasets. Also, what are the specific variables from these that you use? Are some more accurate than other? Are there known strengths or limitations that would make them more or less useful for your purposes? I think a table with data source in the first column, and then subsequent columns for methods, specific variables that are predicted, and accuracy (if known) would be very helpful for you. What if your map corresponds well with one of them, but it has low predictions accuracy?	Thank you for the note. This is a good point and is something we will consider in the manuscript moving forward.	
Clarification	2	Line 436	These 350,000 plots are distributed across the entire United States, not "the landscape" as stated here. Most do not fall in your study area.	Corrected. Thank you.	
Clarification	2	Line 442	This definition of Ecological Systems is good, I suggest a similar statement upfront in the introduction in regards to what a FHT refers to. I am guessing they are pretty similar?	Good note. Yes, we have updated the introduction with a more clear description of "function" and how they are used to delineate FHTs.	
Clarification	2	Line 457	Does this dataset have information on species? If it's strictly biomass it's hard for me to see how it is useful here.	Yes, your comment is correct. However, these are among the datasets we were asked to evaluate (i.e. are a part of our project scope). Agreed, this is not a useful data set, but we were asked to consider it.	
Clarification	2	Line 481	I don't see a need for multiple emission scenarios. Is there some link to climate change that I missed earlier? My understanding is that you are only interested in current vegetation, not future.	No, you are not mistaken. But again, this was within the list of datasets to evaluate in our scoping document.	
Clarification	2	Line 525	What is LW?	Large Wood. Corrected.	
Clarification	2	Line 554	Aren't these Landsat based at 30 m resolution?	Yes, that is true for the LEMMA and TreeMap datasets, but the ITSP has a 240 m resolution. This is shown in the table below.	
Clarification	2	Table 2	My understanding was that LEMMA and BIGMAP were spatially continuous across all forested lands. I am surprised to see here they only cover 50% of the area. Is this correct?	Yes, this is correct. In much of the lower elevational state, private, and industrial timber lands the upland area contains very few or no trees and are not designated as "forests" in these datasets (No data/NA). However, these areas do contain small disjunct riparian forest stands that can extend somewhat into the uplands and are subject to FPA timber harvest rules (i.e. our study area). This is one of the reasons the 20yFHP dataset wins over the others, it has been updated with a larger forest mask to include these lands.	

sampling 20yFHP map	2	Line 603	Why a 10 m by 10 m point grid if you are working with 30 m resolution data? Doesn't his mean a single 30 m pixel gets sample multiple times and you are essentially using the same information in the model multiple times? I think it would be much more defensible if you sampled points at greater distances to avoid spatial autocorrelation.	Most data were in a 30 x 30 resolution, but our DEM and then TPI were 10 x 10, so we developed our point grid from the DEM resolution. We also resampled (downscaled) the 30 m datasets. Our modeling approach accounts for spatial autocorrelation. See response to earlier comments and updated description of BRT methods.
Clarification	2	Line 606	How big was this subsample? Is this intended to be a cross-validation with an independent dataset? Again, if your 10 m grid is sampling the same 30 m pixel multiple times this doesn't seem like it would be an independent validation	No, this was not intended for a cross-validation of other datasets. This was a method used for extracting data from multiple datasets (the 20yFHP for vegetation data, and several other datasets for physiography), which could then be used in further investigation. We did not sample the same 30 m grid multiple times, because with the overlap of multiple environmental factors, many of these points are unique, those that weren't were removed (Machine Learning). We used the 10 m grid because that was the resolution of our DEM, all other datasets were downscaled.
Clarification	2	Line 632	How does one decide if the areas are a "continuous vegetation association"? Is this based on structure or composition or both?	See response to earlier comment. The text has been updated with clarifications of homogeneity.
Clarification	2	Line 646	So you "hypothesized", or assigned groups based on your ordination? Can you give more details so this would be repeatable?	Hypothesized is probably the wrong word. Updated to "potential"
Clarification	2	Line 671	It is important to use consistent language when referring to "forest type". For example in this line you use "vegetation type", then on line 682 you use "forest type". Then on line 688 you use "riparian forest type". Are these all the same as FHTs?	Yes. Corrected. Riparian forest types is most accurate.
Clarification	2	Line 694	I assume the data in this table all have different resolutions which is an important consideration for predictions. While the topographic variables are likely at 30m, I assume the climate grids are much coarser. Also, what are the sources of the climate grids – PRISM?	Correct. There was a range of resolutions with the 30 m being the most common. Updated the climate grids to define original source.
Clarification	2	Line 751	Remind the reader how these were identified – BRT?	BRT included. Good note.
Clarification	2	Line 764	Significant differences in what?	Clarified.
Clarification	2	Line 791	It would be more powerful and useful if you looked at these in multiple dimensions. How do they separate along multiple gradients? I think ideally you would have used a direct gradient analysis approach (Canonical Correspondence Analysis) as opposed to NMS since you are interested in specific variables. What are the percentages following the forest types in this table?	Yes, agreed. CCA would be more powerful and could have shown separation in more dimensions. This approach was simply building on the information learned during BRT. The percentages following these forest types are the inverse of the expected errors listed in Figures 8 and 9. Note that the single-factor frameworks have been moved to the appendix. These were developed as an exercise in maximizing utility (i.e., single factor frameworks are the easiest to be applied).
Clarification	2	Line 799/804	Did you consider using some sort of decision tree approach? This would be compatible with the BRT analysis, and you could include multiple variable and would avoid doing this subjective thresholding.	Yes, we did use TWINSpan in early exploration of the recon dataset. Not included in this report because it showed similar groupings as the ordinations which are a better visual representation.
Clarification	2	Line 1029	NMS analysis is missing details on implementation and results. For guidelines see: McCune, B. and Grace, J.B. (2002). Analysis of Ecological Communities. MjM Software Design, Gleneden Beach, Oregon	Description of NMDS implementation and results has been included. McCune & Grace has been added to the references. Thank you.
Clarification	2	Line 1046	Much of the text in this and the next figure is too small and illegible. I suggest reviewing all figure to ensure they are legible.	Attempted. The bigger fonts had more overlap and made it even harder to read.
Clarification	2	Line 1060 & 1090	Is the framework actually "improved"? You might increase accuracy, but you lose resolution. I wouldn't call that an improvement, more of a trade-off. By coarsening and losing resolution, are the new categories more meaningful?	Fair point, it is only improved in the context of our objectives and relative to the THT but not to itself. Language changed to "more accurate".
Comment	2	Line 1189	Seems like fire is a pretty important disturbance.	included.
Clarification	2	Figure SM1	I don't see any cold forest in the Forest Types – are these missing or just so limited in extent they don't show up?	Yes, in the area shown there are no Cold areas.
Clarification	2	Figure S1	This isn't really an ordination, what exactly is the x-axis representing? It's an interesting way to view the data but what is it depicting?	This is a visual ordination technique used to explore patterns of tree associations between plots. No, it is not a true ordination in the sense that we are not applying multiple iterations and comparing to random assortment. This is a classic first step in ECS approach to begin grouping vegetation. From Pfister & ARno (1980).
General Comment	3		General forest composition and structure can be indeed constructed from data such as the 20yFHP and MSDIM, in concert, with the MSDIM map providing carrying capacity (in a stocking management sense) and the 20y FHP providing modeled composition and generalized effective environment class (dry, moist, cold). However, there are missing components that are relevant to actual management: Degree of departure with respect to historic structure as maintained by fire regime and Habitat structure such as down woody debris, etc.	Agree in principle. However, we were not tasked with this and the rules don't allow for typing in that manner. It would be very unreliable to task foresters with typing something they can't see. Rather, foresters are to determine what they are working with through direct observation and apply rules accordingly. I believe the next phase, rule-making for silvicultural prescriptions can address this concern to move stands towards a desired future condition. This is complicated by climate change, which I also believe can be addressed in the next phase.
Question	3		Major question- will these modeled results be used as a regulatory benchmark, or will local project-scale site data be used?	This is for the policy makers to decide. This is where scale of application becomes an issue. We are presenting results of our findings based on the scope of the project. What Policy decides to do with this information is beyond our scope.
Question	3		Is the described work actually developing and presenting "a system that can be used to inform and guide management prescriptions that support the goals and objectives of the FPHCP"? This is brought up in a few places (e.g., 348-351), but the document is eventually silent on prescription development.	We are not tasked with prescription development; rather, we are tasked with identifying/developing a framework within which prescriptions can be considered. The key word in this sentence is "guide". We are presenting results inline with the scope of the project that will help inform prescriptions.
Question	3		Riparian vegetation can be highly site-specific, with a strong degree of azonal occurrence of species, even dominant tree species. What processes will there be for incorporating the observations of on-the-ground practitioners, such as loggers, foresters, and DNR regulatory personnel? How will that be incorporated along with the predictive modeling here presented?	This is for the policy makers to decide. This is where scale of application may become an issue. It is a fair statement but is outside the scope of this project.
General Comment	3	Line 322-325	A good example is how Engelmann spruce ( <i>Picea engelmannii</i> ) can be found at very low elevations (~2500 feet) in riparian corridors otherwise dominated by zonal vegetation, such as <i>Pinus ponderosa</i> .	Yes. The 20yFHP map actually demonstrates sensitivity to these types of azonal vegetation distributions. This is discussed further in Phase 1.

General Comment	3	344-377	The focus here seems to be very much on forests. I am concerned that the forest focus may exclude critically important non-forest vegetation. Some portions of the riparian zone along an entire reach perhaps should remain open to provide habitat for critically important non-tree vegetation, including sedges ( <i>Carex</i> ), rushes ( <i>Juncus</i> ), willows ( <i>Salix</i> ), alders (especially shrubby alders such as mountain alder, <i>Alnus incana</i> ), ninebark ( <i>Physocarpus</i> ), birch ( <i>Betula</i> ), and more.	Yes, eastside riparian areas contain many different non-forest types. These types are typically a) not treated/not operated in/avoided, and/or b) often have site conditions (like seeps or wet soils) that are covered by other aspects of the rules. In any case, they are not a situation where riparian harvest would occur; hence, not of interest to this study.	
General Comment	3	Table 2	The review of the datasets suggests that DNR at some point may wish to collaborate with WDFW, the USFS, tribes, private forest industry, and non-industrial private forest owners (e.g., family forest landowners) to establish a state-wide standardized data collection effort that covers: Both riparian and non-riparian forest, Overstory and understory vegetation, Fuelbed structure and composition (woody fuels, duff, litter), Habitat elements such as snags and down woody debris (can overlap with data collection for fuelbed concerns), Departure from historic conditions (see work by Paul Hessburg and others), Seral stage and stand age distribution. These data could be then used to generate a terrain- and remote sensing-derived modeled raster dataset at the desired resolution (30 m seems appropriate for management) with multiple layers for different modeled attributes	In effect, that's what the 20yFHP data set is. It is a common basis from which multiple agencies and landownerships evaluate landscape level needs. This point has been more emphasized and clarified in the Data Suitability section with a more in-depth description of the 20yFHP.	
Clarification-error	3	Lines 429-514	ITSP link no longer functions. Could not find maps from MSFP site. MSDIM an excellent spatial basis for site evaluation. These could be linked to different species. TreeMap promising, but not current	The link for ITSP has been fixed. The MSFP site links to a data library, the maps have to be selected based on desired source and location. Yes, the MSDIM is an excellent resource but would require more field data points than we were able to obtain.	
General Comment	3	lines 575-593	An additional approach might be to have a simple protocol deployed to regulatory foresters who are engaged in compliance review in the field. The data they take on composition and structure could be automatically fed into a constantly updating spatial predictive model for riparian forest attributes. This could be similar to the logistic model employed by Fransen et al. (2006, North American Journal of Fisheries Management 26:960-975) to describe end-of-fish distribution, based on surveys of hundreds of streams. The beginning and end of the main forest zones could be similarly modeled, as it seems this is what is proposed.	This is actually a very cool idea. If foresters and managers could continuously upload riparian forest data, then a riparian forest map could be generated with much higher accuracy. The only issue would be differences in sampling methods... I believe the data collection would need to be standardized, may cause an increase in time spent at each stand by the foresters and may cost the DNR quite a bit more money but the product created would be incredibly useful. This is, however, beyond the scope of our study and is something for Policy to consider for future applications.	
Ordinations	3	Lines 579-607	I would worry that the PCA, with its requirements for linearity and orthogonality, would be messy to apply to these environmental data. And furthermore, why use both NMS and PCA? They are both indirect ordination approaches. A better explanation is needed for why both are used. They seem to have returned very similar results.	PCA was removed. The PCA and NMDS in Phase 1 had a different spatial coverage of the study area. The PCA randomly sampled points extending across the entire study area. The NMDS used the LEMMA data set for vegetation cover. However, the LEMMA dataset does not cover the entire study area (clarified in the data description section of Phase 1). Thus, this was only used to increase our confidence and corroborate with the results of the machine learning approaches results in the relationships between physiography and forest types found in the 20yFHP.	
General Comment	3	Figure 4	shows an example of field data collection points and their classification results from field data. One interesting effect that is apparent is that dry forest areas sometimes occur upstream of moist forest areas, indicating that there may be increasing mesic qualities moving downstream (i.e., with increasing moisture reception as a result of landform), and not just upstream with elevation (overall reduction in temperature).	Yes, the 20yFHP dataset appears to represent this phenomenon. The description of the 20yFHP dataset has been updated in the Data Suitability section to emphasize this point.	
General Comment	3	Lines 641-645	It appears like a relatively stable seral endpoint was the goal, which is good for modeling potential vegetation. However, much of the landscape may be in a disturbed condition- which is why the ability to model potential vegetation is critical for accurate application of riparian protection rules.	Yes, agreed. However, recall this framework only informs on stands that are harvestable based on FPA rules and regulations with a focus on protecting the FPCHP objectives. That means the stand will be mature forest dominated by timber species. Highly disturbed forests containing non-timber species or only very few mature stems will not be harvestable and thus not managed by this framework. This framework is a coarse tool that defines the best groupings of timber types to protect specific functions in the FPCHP.	
Ordinations	3	Lines 646-650	I still believe that NMS is sufficient (and appropriate) for this task of developing sets of point clouds in ordination space that will reflect these. Did the indirectly-generated NMS axes seem to reflect moisture and temperature, as one might hope?	The PCA has been removed. Yes, the NMDS axes reflect moisture and temperature as the most important vectors correlating with the species distributions.	
General Comment	3	Lines 682-683	The six categories seem very well suited to characterizing riparian vegetation in eastern Washington. The ponderosa and dry Douglas-fir types may have significant broadleaf representation, it should be noted, due to the frequent openness of the conifer canopy. This allows black cottonwood, willow, and other species to frequently occur.	Don't disagree that six classification is better than three for reasons stated. But can they be reliably determined? The answer is not very well, in which case there is a high probability the wrong prescription could be applied yielding unintended consequences. I would argue 3 is a minimax solution in that broad silvi guidelines can be developed that would be more reliably applied across the landscape. Optimal? No. Less error? Yes. Less blows? Probably.	
General Comment	3	Table 4	many of the variables will contain the same information (e.g., 'Heatload' and 'Hillshade'). It is good that the modeling indicated that most of the variation is due to a reduced number of these variables, as this generates confidence in the model results.	Comment noted.	
General Comment	3	Lines 700-702	this captures the classical concepts of 'available biota, moisture, and temperature'- in other words, the effective environment paradigm of Daubenmire and others.	Comment noted. Agreed.	
General Comment	3	Line 704	The Bailey's ecoregion classification is a bit coarse when viewed at the scale of our region- M242 Cascade Mixed Forest, 342 Intermountain Semi-Desert, and M333 Northern Rocky Mountain Forest-Steppe- but when the continuous axes of moisture and temperature are superimposed upon these, it is likely very informative.	Comment noted.	
General Comment	3	Figure 6	It appears in the ML Model panel that there are aspect-driven differences in predicted regulatory zone, in that 'Dry' type may occur on south-facing facets, and 'Moist' may occur on north-facing facets, directly opposing each other on the same reach. That is a likely outcome in real landscapes, particularly deeply incised drainages.	Agreed. Glad this dynamic could be illustrated and interpreted by the reader.	
General Comment	3	Lines 729-735	The point about AET is well-taken; I do wonder if NDVI might not be useful. Being a differenced ratio metric at the pixel scale, it has the beneficial property of 'cancelling' terrain-driven patterns in reflectance. One must, however, be careful to use satellite images from the growing season only.	We considered NDVI; however, it is an ephemeral variable changing not only within year--as noted--but also between years (e.g. drought v wet years).	
Question	3	Lines 747-749	Why not also NMDS? Also, temperature and elevation are correlated (roughly). Would not also topographic exposure of some form (TOPEX, etc.) be important? Or is that too fine-scale a variable for use here?	Yes, we did corroborate these results with an NMDS (in appendix). We did not evaluate TOPEX specifically, but we did look at TPI, slope, aspect, and heatload which incorporates exposure. But, yes, good note, in retrospect, this may have been a good addition to our analysis. It may have been too fine-scale, which is what we found in the TPI.	

Question	3	Lines 756-763	These tend to be gradient-derived mixtures (see Whittaker's pioneering work), not so much hard transitions. How was that accounted for?	By trying to find the variables that best discriminated, minimized overlap. Yes, there will definitely be gradient-derived mixtures, but the line needs to be somewhere when creating a framework for timber prescriptions.
Statistics	3	Lines 764-771	Clearly specific (spp. specify?) the model, i.e. that you are using a categorical predictor (e.g., your derived categorical timber types) to predict a continuous "response variable" (even though the mechanism is precisely the opposite). ANOVA is fairly robust to violation of normality - but you could have explored alternative generalized linear models where the variance is not assumed to be homoscedastic (e.g., double GLMs or gamma GLMs, etc.).	Clarified in the text. Yes, I see how this was confusing as written. We used continuous data (MAP/Elevation) sampled within each forest category (e.g., PP, DRY MXD.) and used ANOVA/KW to compare their differences in distributions based on MAP/Elevation (not the other way around). This was to add confidence in our separation/lumping of categories based on visual inspection of the histograms. ANOVA methods have been updated. Note: single factor frameworks have been moved to an appendix.
Clarification-error	3	Figure 7	The peaking of timber types in a given region of the precipitation range is evidence that it is not necessarily only the amount of precipitation that falls, but the relative moisture deficit due to topographic convexity and heat lead. This may be where aspect and slope position are important to model. Disturbance and seral dynamics may also play a role, as fire exclusion may have permitted mesic species to encroach into formerly fire-maintained ponderosa woodlands, for example. For the northeast, the elevation largely explains the spruce-fir category, which makes sense, but the histogram for precipitation is a little interesting, as it suggests that much of the spruce-fir type receives less than 1000 mm of precipitation. Again, there may be factors like wind redistribution of snow that cause localized areas of relatively high "effective precipitation". I also see here that the ILAP (related to the Bailey ecoregions?) categories are comprised of the Blue Mountains, the Columbia Basin, the Northeast, and the East Cascades. Is the Northeast a combination of the Okanogan and the Rockies categories mentioned earlier? This should be explained, perhaps earlier in the document.	Good point. We agree that encroachment in absence of fire likely accounts for a high degree of the overlap we are seeing in moist and dry. Reliably characterizing fire history for the modeling data is problematic, however. We discussed this in several meetings with SAGE and CMER, and we agreed that while disturbances of fire suppression or fire history are likely impacting the species distributions there are no reliable datasets with a fine enough resolution to effectively characterize this effect. This would require more available field data with verification. This is outside the scope of work.  Yes, the ILAP zones modification of Bailey's ecoregion (combining Northeast and Okanogan) is defined in the earlier in this section "line 710.
Clarification	3	Lines 787-807	The histograms and the "keys" of Figures 7, 8, and 9 are good tools, especially in light of your process. However, an additional graphic that would help immensely would be a direct gradient graph with elevation and precipitation as axes, and the different types represented with point clouds of different colors, or perhaps convex hull polygons that may (or may not) overlap to various degrees	This is a good suggestion. The single-factor frameworks have been moved to the appendix, but we will explore this option when developing the manuscript. Thank you.
General Comment	3	Lines 820-821	This is an admirably low error rate, given the number of terrain-related, substrate-related, and other variables that affect vegetation. Maps thus generated could be given to the field forester or timber operator as a useful starting point, e.g., "the riparian forest here is expected to belong to Type X" - verify and proceed with RMZ delineations.	This is what we were striving for.
Clarification	3	Line 844-849	Was there effort put forth to model bankfull width?	No. As an input variable? Output variable? Agree that it could help explain vegetation distribution; however, we do not have regionwide data to consider.
Clarification	3	Lines 854-857	"Therefore, we chose a sampling design that would cover the maximum possible RMZ width (130 feet) with a 30-foot extension (160 feet total) to compensate for any inaccuracies. The 160-foot sampling distance was applied to the site selection, and field data collection protocols described below." It is not clear what you are saying here about inaccuracies. Are you just saying that you measured attributes of the vegetation out to the maximum possible legal RMZ width to capture vegetation variability within that topographic gradient?	This was to account for any spatial inaccuracies in the geospatial datasets we used in the site selection. This has been added to the text for clarification.
Clarification	3	Lines 873-875	this explanation could come earlier.	Added it text in Phase 1. It is also described in Table 4.
Replication	3	Lines 878-883	I understand the "experimental blowup" when one multiplies categorical variables of even low 'n'. That said, no replication within each "cell" of the design is a little worrisome.	Yes, agreed, but it was a limitation of the scope. We did not find usable existing field data in Phase 1, and our Phase 2 budget and timeline only allowed for the limited sample size. That is why we raise this point in the discussion and recommend further validation.
Clarification	3	Lines 899-901	Is this a transect with variable-radius sample points at intervals, or is this a transect Relaskop approach applied to live trees? This needs to be explained for the reader to properly interpret the resulting data.	Yes, this is a transect with a continuously variable radius plot determined with a relaskop. It applied to living and standing dead trees (snags). The methods have been updated.
Comment	3		the word <i>data</i> is plural for <i>datum</i> , meaning a single value of measured information. Wherever the word <i>data</i> is used, it should be accompanied by references in the plural- for example, "data are" (not "data is").	Thank you. I believe we have found all verb disagreements and have corrected them.
Question	3	Lines 901-903	Was the transect just on one side of the stream? As noted earlier, riparian vegetation can differ between valley walls as a function of aspect, particularly as one moves up away from the floodplain or channel.	Yes, the transect was on one side of the stream. Yes, vegetation can differ on either side of the stream. These differences are accounted for in our sampling design when selecting for Heatload categories. We sampled on whichever side of the stream the randomly selected point fell.
Clarification	3	Lines 922-954	please provide the frequency type (percentage, count, etc.)	Confused by this comment. The frequency (percent and count) is listed further down in the paragraph and in the figures.
Comment	3	Lines 955-991	All reasonable. I do believe you should not censor the <i>Populus</i> plots, but incorporate them into the error calculations.	They are outside the population of interest. We are defining timber habitats
Comment	3	Figure 13	An overall error matrix for all forest types, collapsing the ILAP zones together (i.e., an error matrix agnostic of ILAP zone), might be a better presentation of overall user's error. Otherwise, these are rather sparse numbers upon which to base a real understanding of percent error.	Yes, but note that the forest type categories (i.e., collapsed groups) differ from one ILAP zone to another. We cannot combine datapoints into one large error matrix if they are describing different plant associations.
Comment	3	Figure 15	The spruce-fir that was predicted as dry-mixed likely had cold air drainage from an adjacent massif of higher elevation. That is also a variable that can be considered for a valley of a given size and elevation- is it connected topographically to a higher elevation area that produces down valley cold air drainage, or not? The amount of dry-mixed that turned out to be moist-mixed is a bit concerning, and I wonder if local moisture collection and soil depth is a dominant factor. A topographically or artesian-driven wet spot in an otherwise dry zone might be the answer, but harder to model, of course. In general, I see that the off-diagonal values trend to zero values, as one would hope, and there is more mass in the upper-left to lower-right diagonal that shows inerrant prediction. I would only surmise that with more data, the true distribution would become clear, and more reliable. I think the real issue is simply low numbers- that a greater sampling effort would produce the expected result.	We tend to agree with your possible reasoning for moist in mapped dry. In my experience, this is fairly common across the eastside landscape and, because water flows downhill, it is common in riparian areas. As noted, this is difficult to model. However, it could be a discussion point and something for policy to consider in implementation. As we noted in earlier comments, however, our study lacks sufficient data to account for these unique or special cases.

Comment	3	Figure 16	47% error is not that different from 58% error, especially with the low plot numbers used to generate this error matrix. It is encouraging that they are similar, even at these low numbers.	Agreed.	
Comment	3	Lines 1060-1076	The Dry DF type really is a subset of the Dry Mixed group. It is often accompanied by an understory of <i>Calamagrostis rubescens</i> or <i>Festuca idahoensis</i> , and nowhere is it a truly common or widespread type. It really ought to be subsumed under the Dry Mixed type. Likewise the Thuja plicata group is likewise uncommon as a pure type- at minimum <i>Pseudotsuga</i> or <i>Abies grandis</i> will be there, and often several other species.	Agreed on all points. That said, I have run across large areas of pure WRC in moist soils due to seeps and springs and subsurface groundwater. But even then, WH and GF take hold on hummocks and islands within.	
Comment	3	Lines 1102-1104	Some of this methodology could have appeared earlier.	See other comments about spreading methods across the document. Essentially, because of the serialization of our procedures (i.e., step 2 is built on step 1), this was the approved representation of our methods as each step builds on the previous.	
Question	3	Lines 1183-1212	It seems more useful to have fewer categories that are more broadly useful. Will clear rules of separation (or guidelines, more likely) be provided for practitioners?	Yes, fewer categories are more useful, however, with the loss of some detail. These are the trade offs we attempt to account for in our estimation of riparian functions. These purposes are updated in the text in response to earlier comments. How Policy decides to enforce or detail how these lines of separation are implemented is outside of our scope.	
Clarification	3	Line 1200	one variable is not autocorrelated with another- it is correlated with another. Autocorrelation refers to correlation of values within a variable to other values within a variable on some basis (usually spatial or temporal). Simply say that temperature and elevation are correlated variables (generally, and with pronounced local exception, such as with cold air drainage!).	Fair, misnomer, corrected.	
Question	3	Lines 1252-1264	I am not surprised that the four groups eventually recommended fell out this way- "PP, DRY MXD/DRY DF, MOIST MXD/WR, and SF"- but would you not want to then circle back and provide the error matrix, shade, and large wood results specific to those categories? And you are correct, a <i>Pinus ponderosa</i> stand that has been invaded by <i>Pseudotsuga</i> behaves very differently than a pure <i>P. ponderosa</i> stand with respect to shade, fuels, etc.!	We are not directly recommending this. We are interpreting this from the results and presenting it as a possible further delineation of the 20yFHP map. It will be the decision of the Policy group to decide if it is worth exploring.	
Question-clarification	3	Figure S1	Why are the types presented in this left-to-right order? Clearly this is intended to mimic an elevational gradient, with PIPO on the left and Spruce-Fir on the right- but why do redcedar and moist-mixed conifer precede Dry-Mixed Conifer and PSME-Dry?	Suggestion taken. Figure updated.	
Clarification	AE	Line 299-297	This was already stated, no need to repeat it.	Not seeing where this is previously stated in the document. Further, this statement also provides a link to the actual study design document. It is necessary.	
Clarification	AE	Line 310,312	Are these two acronyms supposed to be FPHCP?	Yes. Typo. Corrected. Thank you.	
Clarification	AE	Line 324	Should this be replaced with "forest type"?	Yes. changed to "dominant riparian forest vegetation"	
Clarification	AE	Line 345	This is the third time this sentence appears	This is the second time this sentence appears. The first is in the executive summary, which summarizes the entire document. This is the first time it is formally stated in the working document. It is necessary here as this is the official Purpose Statement provided in the scoping document.	
Objectives	AE	Page 11, Objectives	This first objective is unnecessary because you already know what type and quality of data are needed based on the three questions asked of the geospatial datasets. Step 1 evaluates existing datasets relative to these 3 questions.	This comment is confusing. Step one of phase one was developed to answer the first question of objective 1, not the other way around. These objectives come directly from the scoping document, we cannot change or delete them.	
Clarification	AE	Fig 3, caption	If orange streams are also F and S then the suggested wording is clearer.	Okay. Accepted	
Clarification	AE	Line 523	It isn't clear what this means exactly...	clarified "evaluate how forest type categories develop and whether they maintain distinct from other forest type categories over time."	
Objectives	AE	Line 525	This seems to be a new objective not mentioned above.	No, this is building on the questions from objective 1 listed above, this was how we investigated the datasets to satisfy objective 1. Revised: "This involves evaluating not only distinctions in dominant riparian forest vegetation but also evaluating distinctions in associated riparian function."	
Clarification	AE	Line 527	An explanation why estimating large wood potential is necessary and how it will be used to achieve the objectives, is missing.	It is one of the key riparian functions. Clarifying sentence added to Background line 355	
Clarification	AE	Line 543	Not sure how a question answers a question	clarified.	
Clarification	AE	Line 552	This sentence should be followed by sentences that explain what "conducted at several levels of aggregation" means for this dataset. It seems out of place because what follows instead is discussion of other datasets. This sentence and paragraph should be improved, moved or deleted. If this sentence is used elsewhere it should be modified to list and or explain the levels of aggregation instead of just stating that there are several (which is really vague) levels of aggregation.	Removed. Described in further detail in the next paragraph.	
Clarification	AE	Lines 563-565	This sentence needs some work, sloppy!	Revised. Stricken.	
Clarification	AE	Line 571	It isn't clear what "all the conditions covered by the Step 2 frameworks" refers to specifically. Or should this be reworded by referencing the conditions to the 4 questions-criteria for datasets to be used with FVS?	Revised.	
Clarification	AE	Line 574	Replace "degraded" with "exact location of plot centers are not publicly available" I was expecting to read that there are not enough FIA plots within riparian areas are to be useful.	Revised. Suggestion accepted.	
Clarification	AE	Line 584	There isn't an explanation of why the MSDIM would be used. Including it here seems to contradict the first sentence in the Data Suitability section.	The description of the MSDIM and its use has been moved up to the Data Suitability section.	
Statistical jargon	AE	Line 590 & beyond	"Machine learning" is ambiguous and could be a number of different analytical procedures. Replace "machine learning" with the name of actual analyses conducted, BRT. The use of "machine learning" adds unnecessary jargon.	Okay. Replaced throughout.	
Clarification	AE	Line 589	"this" is ambiguous. Be very specific about how the map of expected forest types was mapped.	Revised.	

Ordinations	AE	Line 591	Both the PCA and NMDS are unnecessary. There is no explanation to justify why these ordinations are conducted, how the comparisons would be conducted, and what metrics would be used to use for validation-refutation. This amounts to generating two ordinations for the heck of it.	Disagree. We used these approaches to explore patterns of species distributions across the study area and their relationships to environmental factors. These were conducted to review the data through alternative means yielding different, though related, perspectives on the parameters most influential to describing the distribution of dominant riparian forest vegetation on lands managed under the FCPH. We wanted to know the similarity between randomly selected points based on timber species relative frequencies and how they correlate with the underlying ecological conditions. Recall the scope and purpose of this project described in the background section. NMDS methods have been revised to clarify. The PCA has been removed.
sampling 20yFHP map	AE	Line 608	This number of points is excessive with no explanation of what an appropriate resampling size should be	This number of points covers all areas of the study area. We needed to ensure that we were covering all areas of the study area to investigate and explore the relationships between predictor variables and forest types. Yes, it is excessive, but we used it in our exploratory analysis (BRT) and the map generated was only a byproduct of this exploration. This was done because the construction of the 20yFHP map is murky, and mixed multiple methods to reflect what was actually observed on the ground by forest managers. See updated description of the 20yFHP. Other analyses used a randomly sampled subset of this dataset.
Clarification	AE	Line 646	How is an ecotone of trees distinct from mixed forest categories?	An ecotone is a transitional area between two distinct habitats. A mixed conifer is a distinct habitat itself showing consistent heterogeneity in large patches that provide similar functions throughout, and are repeatable patches of considerable size. Recall the scope and purpose of this study. Forest types are combined and separated based on our understanding of their function in the riparian area (e.g. shade and LW).
Clarification	AE	Line 672	This is removing duplicate records not avoiding "psuedo-replication." Misusing statistical jargon.	corrected.
Clarification	AE	Line 651	Is this a conclusion and what is it based on?	Revised, showed evidence that it represented commonly found (i.e. repeatable forest type categories of considerable size). Because the reconnaissance data was collected based on ease of accessibility (e.g., accessible by road) it was not used for any tests of statistical significance; it was only used for hypothesis development and cross-referencing with available data
Clarification	AE	Line 705	No, the ft categories are informed by the information that was used for their determination. It should not be surprising that the categories are correlated with environmental gradients across the area of analysis. The results of the model show that it is less accurate than the original mapped data.	Statement stricken.
Clarification	AE	Line 716	Is more "nuance" the objective? Nuance is an ambiguous term, poorly defined and an indication someone doesn't know what they are talking about.	Okay. Revised.
Clarification	AE	Line 758	There wasn't an accuracy assessment conducted.	fair. revised. Clause stricken.
Statistics	AE	Line 771	ANOVA is poorly executed. What is the hypothesis being tested? What is the structure of the ANOVA model, exactly what data were used, and how are you modeling the error? Without this and other fundamental information for conducting and ANOVA the analysis is meaningless.	The ANOVA was used to give add a more objective analysis to the separation of forest groups supplemental to the visual inspection of the histograms. The term "statistically significant" has been removed. The null hypotheses and methods have been clarified.
Comment	AE	786-788	Which basically leaves precip and elev which have low percentage values in Table 5.	correct.
Elevation Framework	AE	Line 789	Creating another framework on elevation in addition to the THT seems redundant.	Yes, but it isn't that elevation itself was a poor choice of factors for a framework, it is how the cutoffs and groupings are applied. The THT framework used elevational cutoffs across the entire study area based on the observed frequency of upland forest types. We attempted to make this framework more relevant to riparian forests (using the data we had, only sampled in riparian areas), and we changed the forest groupings, and finally, we separated it by ecoregion as that also appeared to have an effect of forest groupings.
Citations	AE	Line 819-823	This should be no surprise and should be accompanied by the references such as Daubenmire (1968), Franklin and Dyness (1973)	included.
Statistical jargon	AE	Line 882	Need to include definition of orthogonal as it is used here, otherwise, this is meaningless statistical jargon.	Updated. new references included.
Citations	AE	Line 900	Hemingway and Kimsey, 2020 Not in bibliography	added.
Clarification	AE	Line 908	Radius or diameter?	length and width. Updated in response to earlier comments.
Clarification	AE	Line 929	What was the criteria used to designate these categories? Was it subjective or based on a key or field guide?	updated.
Clarification	AE	Fig 13	Rate column is unnecessary because it is redundant with the % Error column.	Fair, but it shows a different expression of the values. Our advisory groups like them.
Clarification	AE	Line 1078	Not very distinct if it falls within another group	Removed.
Clarification	AE	Line 1095	And why is separation important for this shade analysis?	Recall the purpose and scope of this study. Riparian function (Shade, LW, Bank stability..) is the driving factor for forest habitat separation or combination in this framework. Definitions of these functions have been included in the background.
Clarification	AE	Line 1096-1098	Need to clarify what functional response and convergence means in this sentence	Function and functional response definitions and descriptions have been updated in the background and introduction. Sentence has been revised.
Clarification	AE	Line 1150	"Stand" is a level of abstraction that hasn't been defined in the simulation. Should this be changed to "transect" or sampling site?	I think site is best. Updated.
Clarification	AE	Line 1174	Why is separation important here?	updated. Function.
Clarification	AE	Line 1202-1203	More meaningful how?	updated.
Clarification	AE	Line 1208	Species distribution modeling was not part of the analysis. Stating it here is misleading.	Modeling is not mentioned or implied in this sentence. We are describing actual distribution of species as they have been observed in the field data.
Clarification	AE	Line 1221	How much is "enough".	to be comparable to classifications based on the elevation-based rules for the current three THT categories
Clarification	AE	Line 1239	This makes me wonder how useful is a dataset with unknown error-accuracy for use and for a means of comparison?	it should, that is why it is mentioned and we recommend further investigation of this dataset.
Clarification	AE	Line 1233	So why do it?	This section has been updated to describe a traditional and appropriate approach. It ties in with the subsequent sentences and updated description of the 20yFHP dataset in response to previous comments. This was an exploration of the 20yFHP given the unknown specifics of its construction.
Comment/Suggestion	AE	Line 1251-1253	More data, randomly sampled, used to generate a suitability model with Maxent would produce a more reliable output and map that could be used to demarcate different types of forest.	Yes, but again, please understand the scope and objective of the project described in the background and introduction sections. We were tasked with developing and testing alternative frameworks with existing data sets. Phase 1, step 1 was an evaluation of all available datasets to choose which was the most appropriate dataset available (20yFHP). Step 2 was development of potential frameworks from that existing dataset. Phase 2 was the collection of data to assess the potential frameworks developed in Phase 1. We were not tasked with developing a whole new suitability model. Considering the scope of the project, collection of enough data to develop a model was not possible.
Clarification	AE	Line 1301	In terms of what measure of evaluation?	Language updated. "performance" removed from this sentence and defined in the following.



