

# Response to reviewers

March 4, 2021

We are pleased to present a significantly revised version of our manuscript titled “Cutting out the middleman: Calibrating and validating a dynamic vegetation model (ED2-PROSPECT5) using remotely sensed surface reflectance” for consideration for publication in *Geoscientific Model Development*. We are very grateful to the two reviewers for their constructive feedback on our analysis, and we firmly believe addressing their suggestions has made our manuscript significantly better. In Section R1, we provide a summary of the most significant changes in this revision. We provide a point-by-point response to each reviewer’s comments in Sections R2 and R3. All line numbers refer to the highlighted changes version of the manuscript.

## R1 Summary of revisions

### R1.1 Revisions to the inversion algorithm

Reviewer T. Quaife remarked that we did not mention how our inversion algorithm accounted for correlations between closely-spaced hyperspectral channels. In our original submission, we did not account for this spectral autocorrelation, and our likelihood expression assumed all bands were independent. However, we recognize that this was flawed. In our revision, we have updated our inversion algorithm to account for autocorrelation in the residuals by imposing an order-1 autoregressive (AR1) correlation structure on the residual covariance matrix in our likelihood expression (calculated exogenously, with the AR1 correlation coefficient calculated from the residuals from the original simulation). This is described in our revised methods (lines 329–337).

While implementing this, we made several other changes to our inversion algorithm as well (all documented in the revised methods). For one, rather than prescribing a common set of illumination conditions (direct vs. diffuse radiation and solar zenith angle) for all observations, we now fully account for variability in both across all observations by extracting the solar geometry from the AVIRIS metadata and prescribing radiation conditions (direct-diffuse fraction) based on the hourly MERRA-2 meteorological reanalysis product (lines 307–310). In addition, we have replaced our original lognormal likelihood penalty for LAI predictions with a uniform distribution that guarantees that LAI predictions do not exceed a reasonable threshold but also does not differentiate between LAI values below that threshold (lines 345–348). We have updated all of our results figures and text accordingly. The impact of these methodological changes on our results was relatively minor, and none of the core conclusions have changed.

### R1.2 Discussion of reflectance fluxes

Reviewer T. Quaife remarked that our original manuscript incorrectly downplayed the important differences between directional reflectance factor (which is observed by most optical remote sensing instruments) and hemispherical reflectance (a.k.a., “albedo”; the quantity simulated by EDR and related models). We fully agree that these are two distinct (albeit related) physical quantities that, in general, cannot be compared. We also recognize that our original discussion of this topic was flawed and, as stated, misleading. To address this comment, we have completely rewritten our discussion of reflectance fluxes, taking care to emphasize that surface reflectance factors estimated by AVIRIS and similar sensors cannot be directly compared to hemispherical reflectance simulations from EDR and similar models without additional processing steps (lines 564–594).

That said, because we were using AVIRIS classic imagery that underwent additional processing for the NASA Forest Functional Types (FFT) project (Singh et al., 2015), we feel that our analysis is still valid. In addition to the standard atmospheric correction and orthorectification conducted by NASA JPL, the AVIRIS data we used for model calibration were also cross-track illumination corrected, as well as BRDF-corrected, following the procedure of Lucht et al. (2000). Briefly, the latter approach estimates “intrinsic surface albedo” — the precise quantity that is simulated by EDR — from angular reflectance data through application of a polynomial approximation to the Ross-Li semiempirical BRDF model. The full AVIRIS processing pipeline for the AVIRIS data (including the BRDF approximation) we used is described in Singh et al. (2015). This information was absent in our original submission, but now appears in our expanded “Site and data description” section (lines 293–302).

### **R1.3 Structural issues in EDR**

Reviewer 1 noted that several of our parameters exhibited edge-hitting behavior and suggested that we discuss this behavior and the extent to which it might be symptomatic of structural issues with EDR. At the same time, Reviewer T. Quaife suggested that we compare EDR and PRO4SAIL — another two-stream canopy radiative transfer model popular in the remote sensing community due to its ability to predict both directional and hemispherical reflectance fluxes — to explore the differences between hemispherical and directional reflectance. In our revision, we added sensitivity analyses of EDR to some of its parameters (Figures A1, A2, and A3), and a comparison of EDR and equivalently parameterized PRO4SAIL (Figures 10 and A16).

An important takeaway from this analysis is that, compared to PRO4SAIL, EDR systematically underpredicts hemispherical reflectance. This can be attributed to a difference in how EDR and SAIL define the direct radiation backscatter coefficient (see also discussion in Yuan et al. 2017). This underestimation of hemispherical reflectance by EDR may help explain some of the calibration issues identified by Reviewer 1. Specifically, both increasing canopy clumping (which increases effective leaf area index) and increasing the leaf orientation factor (more horizontal leaves) increase the simulated canopy albedo, so pushing these parameters up against their physical upper bounds may be a way for our inversion to compensate for EDR’s underestimation of albedo. We have included a discussion of this in our revised Discussion section (lines 544–557).

### **R1.4 Discussion of equifinality**

Reviewer 1 noted that, while we mentioned in our Introduction the equifinality between leaf biochemistry and canopy structure in canopy radiative transfer, we did not revisit this important issue in the rest of our manuscript. We agree that the topic of equifinality in canopy radiative transfer is worth additional discussion. In our analysis, equifinality was reduced by using informative priors for the leaf parameters. We note in our revised discussion that, in future work, this prior information could also come from other kinds of remote sensing measurements (lines 512–514) as well as through the ecophysiological mechanisms embedded in the model itself — i.e., some combinations of model parameters and states that are consistent with a given surface reflectance may be excluded because they imply ecologically unrealistic states in previous time steps or result in ecologically unrealistic outcomes in future time steps (lines 475–482).

### **R1.5 Revised model description**

In the process of revising our manuscript (and particularly, of investigating the discrepancy between EDR and PRO4SAIL), we realized that while we correctly described canopy radiative transfer coefficients in EDR (e.g., inverse optical depth; leaf angle distribution functions) and correctly implemented EDR in our code, we did not correctly describe the multi-layer canopy radiative transfer solution actually used by EDR. In addition, we felt that our model description was not well-organized. Therefore, in this revision, we have completely re-written the Model Description to true to what the EDR code is actually doing, and, in our opinion, to be more clear (Section 2.1). Importantly, we now split this section into two sub-sections: The first subsection describes the radiative transfer coefficients and their derivations (which should make it easier to cross-reference against the descriptions of its parent model — Sellers 1985 — and other radiative transfer

models). The second subsection describes EDR’s unique solution to these coefficients for a multi-layered heterogeneous canopy.

## R2 Reviewer 1

General Comments (overall quality): This manuscript uses remotely-sensed surface reflectance observations to calibrate/constrain a series of ecosystem parameters from the ED2-PROSPECT (EDR) model characterizing the land surface at 54 forested sites related to leaf biochemistry, canopy radiative transfer, and soil characteristics. An important innovation is the introduction of the radiative transfer model PROSPECT to the biosphere model ED2 to provide an improved spectrally resolved simulation of surface reflectance. This is done, in part, to bring the observed surface reflectance closer to what the model actually predicts, helping to reduce the impact of observational uncertainty as well as more effectively constraining multiple components of the model. The authors find that through the assimilation of surface reflectance EDR can provide better simulations of the surface reflectance spectra and leaf area. The findings suggest that this approach could be used to better constrain surface energy balance as well as overall ecosystem functioning. Given many other ecosystem models include a two-stream radiative approach they contend their results should be widely applicable.

Scientific Questions/Issues: The authors bring up the issue of equifinality in the introduction, which presents a challenge for this application in that the surface reflectance can be a function of leaf biochemical properties, leaf structure, and canopy and soil radiative transfer characteristics. To some degree, equifinality was reduced in that the prior parameter distributions for biochemical leaf properties were tightly restricted, and limited to the extent the surface reflectance observations could influence them. In contrast the canopy radiative transfer parameters and LAI (through SLA) were simultaneously being adjusted by the optimization. Was hoping the authors could comment more about how equifinality of the surface reflectance influenced their results.

Please see Section R1.5 in our main response.

As a follow up question to the equifinality question above – ED2 is a dynamic vegetation model with the ability to simulate competition amongst cohorts thus providing a simulation of co-existing dominant PFTs. It wasn’t clear how well the simulation of cohort competition influenced the final distribution of PFTs and to what extent this matched the site level observed vegetation state. Given that the parameter optimization was PFT specific, the precise vegetation PFT distribution could have a large impact. Was the PFT distribution prescribed?

In this study, the vegetation composition at each site (including the PFT distribution and size-age structure) was prescribed in detail based on data from the NASA Forest Functional Types (FFT) field campaign. In our revision, we have clarified this point in the “Site and data description” (lines 287–289).

Was there any attempt to withhold some site reflectance data and apply the calibrated model parameters at those sites? It seems the optimized surface reflectance simulations were performed at sites that were already calibrated. The fact that the authors performed an across-site joint assimilation may in part account for this, but was interested how the calibrated parameters would perform at sites outside the calibration sites.

We acknowledge that cross-validation or out-of-sample validation are useful tests of model performance, and in our revised discussion, we recommend these activities as future directions for this work. However as the reviewer points out, because our calibration was joint across all sites, we did not feel that a separate validation at other sites not used in the calibration was necessary. With 54 sites in our calibration, any single site represents < 2% of the data, and for a joint calibration without site random effects, we have every reason to believe that the calibration is not overfitting to any individual site; trying to fit any one site well would cause others to do worse (especially given the large observed variability in forest structure) unless the EDR model structure was reasonable and the parameters chosen were genuinely good choices. We have added this information to our revised discussion (lines 483–489).

It is known that radiative transfer models are challenged in simulating evergreen species in part because of

the irregular and open space canopy structure. Many of the figures in the supplement demonstrate stronger biases in simulated surface reflectance exist for evergreen sites as compared to deciduous. Was hoping the authors could comment on this, and recommendations for getting around this.

The reviewer brings up a great point about conifer canopies historically being harder to capture. Although some conifer-dominated sites did demonstrate significant biases in reflectance predictions, our analysis of reflectance bias by composition, structure, and spectral region (Figures 4, 5, A6–A15) shows that these biases are not systematic (though they may drive greater predictive variance). In our revision, we describe this analysis in more detail in the Methods (lines 372–378) and Results (lines 424–434).

Whereas leaf level biochemistry related parameters were well constrained by the priors, many of the radiative transfer posterior parameters seemed to be edge hitting parameters. To what extent do the authors believe this behavior was caused from structural error in canopy radiation transfer and/or mismatch caused in part from radiation directionality differences between the simulated and observed canopy reflectance? The authors devote a considerable amount of general discussion regarding this topic, but don't directly address how this might have effected their own results.

Please see Section R1.4.

The manuscript begins by justifying the inclusion of the PROSPECT model in to ED2 to bring the model closer to the observations, to, in part help reduce the uncertainty of the observation that are assimilated into EDR. More explanation on how the surface reflectance observational uncertainty was quantified here, and what it represents, and to what extent overconfidence in uncertainty may have led to the posterior edge hitting parameters.

As shown in equation 48, observation error in the reflectance data was not estimated *a priori* based on the instrument itself, but was modeled as the residual error between the model and the data, analogous to what is done for any linear or nonlinear regression model. A key difference, however, is that the error model accounts for the known heteroskedasticity in spectral data (i.e., the size of the variance increases with the magnitude of the reflectance). In terms of random spectral errors, there is no reason to expect this variance to be overconfident for inferences made on these landscapes, especially as the study sites were not all imaged on the same day under the same atmospheric conditions, though we'd agree that one might not want to apply this variance to entirely different ecoregions. Furthermore, because the variance slope and intercept are fit parameters, whose parametric uncertainty is being quantified and propagated, this makes it even less likely that our uncertainty estimate is overconfident. That said, the current approach does not formally account for any possible systematic errors in the observations, which could have a more serious impact on inferences. However, we would note that we are unaware of any derived data products that account for these systematic errors either. Furthermore, in addition to the same uncertainties about reflectance that we face, those products additionally contain numerous uncertainties about model structure, parameters, and covariate data whose uncertainties are rarely fully propagated, meaning that the alternative approach (calibration to derived data products) is more likely to result in overconfidence in uncertainties than the approach taken here.

That said, we feel it is highly unlikely that overconfidence in surface reflectance estimates contributed to the edge-hitting behavior of some of our posteriors. Random errors would not result in a systematic parameter bias and, given the long history of the instrument and maturity of atmospheric correction approaches, any systematic errors in AVIRIS are likely to be quite small relative to the structural uncertainty in EDR (i.e., biologically implausible parameters should not be necessary to capture observational data biases of the magnitude likely to be present).

Detailed Comments: Abstract and manuscript in general: Need more discussion on what we hope to gain by this. We don't really care about surface reflectance (although energy balance is important), but we do care about how LAI, chlorophyll, pigments and water status influence ecosystem functioning through carbon and water exchange. I think this needs to be emphasized more, and provide evidence that this sort of setup can accomplish this.

We agree that additional emphasis on the implications of this work is warranted. In our revisions, we emphasize several important implications in the abstract (lines 1–7) and introduction (lines 74–93).

First, as you say, energy balance is important, and the contribution of vegetation to land surface albedo is a critical mechanism by which vegetation influences regional and global climate (Bonan, 2008). Therefore, ensuring the accuracy of model simulations of albedo, including its sensitivity to vegetation structure and composition, is essential to accurately projecting the effects of climate- and land use-driven changes to terrestrial ecosystems on future climate.

Second, canopy radiative transfer directly affects many physiological, ecological, and physical processes included in complex demographically-enabled vegetation models like ED2 (Viskari et al., 2019). Light availability and absorption is a first-order control on photosynthesis, and ability to survive under different light levels is an essential component of a tree species' position in forest succession. Meanwhile, temperature—which is strongly influenced by albedo—directly affects the rates of both enzyme-kinetic physiological processes and evaporation.

Finally, ED2 and similar models are highly sensitive to many of the leaf traits constrained by this analysis (Raczka et al., 2018; Dietze et al., 2014; Shiklomanov et al., 2020). Given the large variability of these traits through space and time, remote sensing is an essential data source for model parameterization, and our work provides a useful approach for doing so.

Line 1: Remove 'derived'. The fact that they are 'data products' and not 'observations' gets across the point.

While we agree that “data products” should imply a difference from observations, treatment of remote sensing data products as true observations without accounting for uncertainties or biases is widespread in the Earth science community. Therefore, we think it is important to emphasize that these products are derived.

Line 5: 'compared against airborne and satellite data' Technically, this is still data and not observations in that even reflectance data requires RTM models, I believe. But it is more direct relationship

We agree that additional nuance is required here (especially when also considering reviewer T. Quaipe's comments). Therefore, we have revised this section to capture the fact that although these data are still derived (processing steps include atmospheric correction and orthorectification), they rely on fewer assumptions (especially about the land surface), have fewer processing steps, and therefore are closer to observations (lines 61–64).

Line 23: add 'to' calibrate or constrain

We have revised this accordingly.

Line 25: I know exactly what you mean by 'constrain', but could you use 'calibrate' or 'inform' in this context?

We have replaced “constrain” with “inform”.

Line 32: “More sophisticated approaches for estimating vegetation properties based on physically-based radiative transfer models face issues of equifinality, whereby many different combinations of vegetation and soil properties can ultimately produce the same modeled surface reflectance (Combal et al., 2003; Lewis and 35 Disney, 2007).” This is important I think – and raises a key point for this analysis – is there not equifinality when trying to constrain leaf structure vs. leaf status? I would think equifinality could be a problem here, and I think you need to acknowledge this and how you might address this – Strong priors? Demonstration that surface reflectance can tease apart these two things. . .

Please see Section R1.4.

Line 52: Awkward topic sentence. Simplify “Some land surface models already include there own . . . that allowd for a more direct comparison to remotely sensed surface reflectance.”

We have revised this sentence according to the reviewer's suggestion.

Line 57: “Canopy radiative transfer plays a particularly important role in the current generation of demographically-enabled dynamic vegetation models, where differences in canopy radiative transfer representations and parametrizations have major impacts on predicted community composition and biogeochemistry

(Loew et al., 2014; Fisher et al., 2018; Viskari et al., 2019).” Seems weird to word it this way. Isn’t it the other way around, community composition and biogeochemistry impact the RTM?

It is true that composition and structure impacts the RTM, but all of these studies show that the opposite is true as well! These studies illustrate that because the RTM determines the overall energy balance of the ecosystem and the distribution of light within the canopy, RTM formulation and parameters profoundly impact the predicted biogeochemical fluxes and vegetation dynamics. For example, Viskari et al. (2019) show that uncertainties in canopy RT can cascade and impact a number of associated processes, including photosynthesis, energy balance, internal competition, and demography. We have revised the sentences here to more clearly and explicitly convey this idea (line 74–90).

Line 72: “...will significantly constrain model parameters related to canopy structure.” So the goal all along was to constrain canopy structure with surface reflectance, not necessarily foliar biochemistry? Maybe talk a bit more about the differences in sensitivity of surface reflectance to canopy structure vs foliar biochemistry.

Our objective was to evaluate which parameters could be constrained. The list of candidate parameters included parameters related to both structure and biochemistry. However, we hypothesized that, because of the informative priors on foliar biochemistry, the constraint would be relatively greater for canopy structural parameters (as stated here).

Per our earlier responses (Section R1.3), we have included an additional parameter sensitivity analysis of EDR and additional discussion of equifinality.

Section 2.3: Can you provide a sense of scale? For example for the 54 sites, what spatial range was the inventory data taken, and what spatial resolution did AVIRIS cover? Trying to get a feel for spatial mismatch, etc. Were sites chosen because they were rather homogeneous for certain PFTs?

In response to this and Reviewer T. Quaife’s comments, we have elaborated on the methods behind the data used in this analysis. Specifically, each of the 54 sites here consisted of a  $60 \times 60$  m transect within which forest inventory data were collected. AVIRIS-Classic data were extracted as the average of a  $3 \times 3$  pixel array (each pixel is 15–20m, depending on aircraft altitude) centered on the site transect center, resulting in a single composite spectrum for the  $60 \times 60$  m area. Additional details on the sampling methodology are described in Singh et al. (2015) and references therein.

Figure 1: Was a little surprised to see many sites so close to Lake Superior. No issues with interference from nearby water reflectance?

Although sites do appear very close to Lake Superior in the map we provide, all sites are sufficiently inland (several kilometers) that contributions from water reflectance of nearby pixels can safely be assumed to be negligible. In the revision, this figure has been broken up into two figures (Figures 1 and 2).

Line 200: Could you provide a bit more explanation of what including the EDR predicted LAI term within your probability function accomplishes? EDR response becomes saturated to LAI so is this an artificial way to account for increased reflectance?

Yes, the LAI penalty in the likelihood function accounts for the saturating effect of increasing LAI on reflectance. This effect is not unique to EDR—rather, it is a well-known consequence of the exponential extinction of light through a medium, following Beer’s Law. Therefore, a canopy with an unrealistically high LAI like 15 has virtually the same reflectance as a canopy with a high but more feasible LAI like 6 (all else being equal), and therefore a likelihood calculation that does not penalize excessively high LAI values would consider both outcomes equally likely.

We have added additional text to this effect to the methods (lines 340–347), and have added the attached figure to the supplement demonstrating the saturating effect of LAI on reflectance (Figure A1). However, as noted in Section R1.1, in our revision, we have changed this penalty from a lognormal distribution to a uniform distribution between 0 and 10.

Line 227: So, to evaluate the model you compared the EDR-spectra against the AVIRIS observations at sites that were used to calibrate the model? Was there any attempt to withhold some site data and apply the calibrated model at those sites?

As we stated above, because our calibration was joint across all sites, we did not feel that a separate validation at other sites not used in the calibration was necessary.

Line 232: Can you quantify what ‘informative’ means

We have clarified this sentence to say these are “leaf parameters whose prior distributions were already independently constrained by an earlier analysis”. These prior distributions are shown in Figure 3.

Line 233: I cannot see in figure where PROSPECT N parameter is?

In our experience, labelling PROSPECT’s “N” parameter as such is confusing to readers unfamiliar with PROSPECT because it suggests leaf nitrogen content. Therefore, we prefer the more explicit name “# mesophyll layers”. We have revised the methods and results text in a few places to clarify this.

Section 3 Results: Although Figure 2 was very informative, I found the Results section in general, relatively vague, perhaps some sense of % reductions in 95% credible interval.

We have revised the results (lines 380–388) to include more precise statistics, including the suggested relative reductions in the width of the 95% credible interval.

Figure 2: Hopefully, the authors comment on some of the apparent edge-hitting parameters specifically related to canopy RTM parameters such as leaf orientation, canopy clumping and water. I worry that the information from the relatively strong and defensible leaf biochemistry prior parameters leading to relatively self contained and PFT differentiated posteriors for the leaf biochemistry parameters is lost or made irrelevant due to biases between model simulated and observed surface reflectance that are corrected by ‘fitting’ the RTM parameters. The context of this manuscript doesn’t indicate how sensitive the surface reflectance is for the suite of parameters calibrated here... perhaps included in one of cited manuscripts.

Please see our response in Section R1.3.

Figure 3: A map would be helpful with site codes provided. Perhaps a zoomed in map that demonstrates where these sites are spatially? Also is the stem diameter plot on the right simulated or observed? In fact, doesn’t that have a large impact on the assimilation, the PFT distribution and stem diameter distribution?— has it been demonstrated that ED2 can properly simulate the competition of PFTs at the site providing the correct vegetation state, such that the parameter optimization reflects the observed vegetation state ?? Or has the vegetation state been prescribed in this case?

We have split up the original figure into two separate figures: one showing a larger map with sites labelled (Figure 1) and one showing the density vs. diameter plot (Figure 2).

As stated in the main response, the vegetation composition at each site (including the PFT distribution and size-age structure) was prescribed in detail based on NASA Forest Functional Types (FFT) campaign field data. We have clarified that these are the observed stand structures used as prescribed EDR inputs in the methods (lines 287–289).

Figure 5: “The observed vs. predicted line had a slope of 0.37 and an intercept of 2.80, indicating that EDR calibration underpredicted LAI on average but overexaggerated across-site LAI variability.” What do you attribute this clear structure in residuals between observed-simulated LAI?

We agree the clear mismatch between predicted and observed LAI was not given sufficient attention in our original draft. In our revision, we now devote more space to this in the Results (lines 437–441) and Discussion (lines 533–542).

Line 258-270: I like the overview explanation of bringing observations closer to models or alternatively bring models closer to the observations. In the end, it’s a bit of semantics, especially for using information from satellites we will always need some sort of transfer function or forward operator to convert from what a satellite observes and what a model predicts. I don’t think one way or the other should take precedent. The advantage in your approach, however, is the potential for the observed surface reflectance to constrain multiple model components, whether that be leaf structure or biochemistry or water status. I think that is potentially extremely valuable although I am not sure it has been demonstrated, yet, that this is the case. I feel more could be ‘learned’ about what information surface reflectance could provide if you could

prescribe the LAI and PFT-distribution in ED2, then you could really get a grasp on what it can inform, leaf biochemistry? Within-canopy RTM parameters? Etc.

We disagree that the difference between bringing models closer to observations and vice-versa is a purely semantic one—the fact that all observations need some transformation does not mean that all transformations are the same. Instead, we argue that comparing model output to a highly derived data product (e.g., MODIS GPP) is closer to a model intercomparison than a model validation, since generating those data products invariably requires their own models (whether statistical or process-based) that make specific assumptions about different processes, often very different assumptions than the model the products are compared with. Thus, it is often really an “apples to oranges” comparison. In other words, in addition to the *shared* observational uncertainties (e.g., atmospheric correction, instrument calibration) that are present in both approaches, derived data products include numerous additional assumptions and uncertainties that are avoided in a spectra-to-spectra comparison. Meanwhile, as you point out, the approach of bringing models closer to observations is advantageous precisely because of the model’s emergent covariance across many disparate variables rooted in specific assumptions about biophysical and ecological processes. In practice, for an observation operator that is almost entirely disconnected from the model (for example, if we just took the total LAI from ED2 and used a standalone RTM like SAIL to simulate the reflectance), we agree that there’s relatively little advantage relative to a derived product, but the more tightly an observation operator is integrated into a model (as in our analysis, where canopy RTM parameters and outputs are used elsewhere in ED2), the greater its value.

To address this, we have revised the discussion to discuss the relative advantage of building models with remote sensing-friendly observation operators, especially when such operators are tightly integrated with other processes in the model (lines 475–482). For the latter, we now discuss some future directions for ED2 and optical remote sensing specifically, such as using the PROSPECT leaf model to couple the model’s representations of plant hydraulics (for leaf water content) to canopy radiative transfer. Similar approaches that leverage the model’s internal RTM can also be applied to other remotely-sensed data, such as lidar (canopy structure), radar/microwave (structure and water content), thermal (canopy traits and water use), and fluorescence (photosynthetic traits). Indeed, this paper is an important first step towards the ultimate goal of leveraging internally-consistent process-based models to make joint inferences across multiple remotely-sensed data constraints simultaneously.

But, as you brought up in the introduction, this brings up equifinality issues. Not so much in this case for your leaf biochemistry parameters because of strong priors, but it does seem to be the case for canopy RTM parameters and predicted LAI (SLA). I think you need to caveat or address this concern.

Please see Section R1.4.

Lines 286-310: It seems like you are pointing out sources of structural error within the radiative transfer of EDR or, to the extent, mismatch between what EDR simulates vs. what AVIRIS-Classic observes. Could this help explain why the calibration caused some posterior RTM parameters to be edge-hitting against the bounds of the priors? Also, this work was in part motivated by being better able to quantify the uncertainty in the surface reflection observations, but I am not sure I saw a clear explanation of the uncertainty that was used or provided for the AVIRIS and how was this quantified. It seems parameters have the potential to be overfit, if the observation uncertainty is not realistically quantified. May have missed this.

Please see Section R1.3

Line 320: Really it’s the power to upscale that remote sensing products provide. However, this comes at a cost, they ‘observe’ reflected radiation from the land surface which indirectly characterizes things that we care about like, like leaf biochemistry, albedo etc.

We agree the reflectance data collected by passive optical remote sensing are affected by many different surface features and processes, often in confounding ways, which makes them challenging to use in isolation. Fully leveraging the power of remote sensing requires additional sources of information from both in situ data and the understanding of biophysical and ecological processes embedded in our vegetation models. We have revised this text and expanded this text accordingly.

Line 326: ‘accurately reproduce surface reflectance and leaf area index’ I think this is a bit of an over-



statement, especially because the figures demonstrate systematic mismatch between the optimized surface reflectance and observed reflectance (figure 3), and strong residual error structure in LAI (Figure 5). Perhaps this approach provided ‘improved’ reflectance and LAI is better wording.

We agree that we overstated our LAI prediction accuracy here, and have revised this text to temper our conclusions accordingly. However, as we pointed out in earlier comments, although our reflectance predictions were not perfect everywhere, the quantitative analysis of surface reflectance revealed no systematic biases by PFT or stem density.

Line 330: I think you also need to say where this work can lead — this is of interest for those that are concerned with ecosystem functioning and that this could provide improved estimates of both biomass and land-atmosphere carbon and water exchange. Also some discussion of the differences between evergreen and deciduous forests would be helpful. Generally RTM’s have more difficulty with more open canopy evergreen species and not really discussed in this manuscript.

Our revised discussion (see earlier comments) and conclusions (lines 632–634) include more text on ways that ED2 (and other vegetation models) can be further enhanced to better take advantage of passive optical and other remote sensing techniques.

Appendix:

Figure A2: What would be really helpful is to use the same color-coding in Figure A1 to show deciduous vs evergreen sites. Also this brings about the question – why did you choose the sites that you did to put in the manuscript itself?

Figures A5-A13 provide the requested breakdown of plots by PFT composition. For the main figure, we selected sites that, collectively, span the geographic, stand-structure, and PFT composition space of our study. We have added this information to the figure caption, highlighted these plots in our site map, and mention this in the revised Results section.

Figure A3: This sort of gets at the hardwood vs conifer performance as well. I think it would be helpful to comment on this distinction in performance within the results/discussion.

As mentioned above, we have added additional text about our analysis of performance by PFT to the methods, results, and discussion. As this figure clearly shows, there is no systematic bias in reflectance simulations for conifer species.

Figure A13: This is also a very compelling figure that gets at the increased bias in spectra for conifer. Worth discussing in main manuscript.

See earlier comments about this. We note that even this figure shows no systematic bias—three sites have underpredicted reflectance (AK06, AK60, OFO1), two sites have overpredicted reflectance (MN02, MN04, BH02), and the remaining 2 sites have relatively accurate reflectance predictions (SF01, SF04).

### R3 Reviewer 2 (T. Quaife)

The manuscript “Cutting out the middleman: Calibrating and validating a dynamic vegetation model (ED2-PROSPECT5) using remotely sensed surface reflectance” by Shiklomanov et al. illustrates the assimilation of reflectance data observed from aircraft into a component of the ED2 model. In many respects this is an excellent paper. The Bayesian approach taken is state-of-the-art and the use of forward modelling of reflectance for assimilation purposes is desirable for many reasons (and yet surprisingly little progress has been made in this area for land surface studies, making this paper especially welcome). The text is also well written on the whole.

Unfortunately I do have one rather significant concern about this paper, which may at first seem like a subtlety, but really is not. The authors are not “cutting out the middleman” so much as choosing to ignore them. I am concerned that the take home message of the paper for people less familiar with the underlying physics will be that it’s OK to take this approach.

1. Assimilation of BRF and HDRFs.

[Note: I am using the definitions of reflectance quantities from Schaepman-Strub et al. (2006; hereafter S06), which the authors have also done.]

My major concern with this paper is the authors' misuse of different reflectance quantities. They are not comparing like with like and I do not agree that it is OK to assimilate quantities that are not physically consistent with those being modelled. The Sellers two-stream model predicts reflectances (BHR and DHR), whereas the AVRIS data are reflectance factors (BRF and HDRF). Although they are both unitless they are fundamentally different quantities and have different scales.

The authors do devote a paragraph to discussing this, but it is misleading and I am not convinced by the arguments they make. The statement that the ED radiative transfer model predicts BHR is only partly correct. It also simulates DHR and the predicted reflectance is a weighted mean of the two. The authors go on to state the AVRIS observations are most related to HDRFs. This would only be true for overcast skies. AVRIS observations are best represented as a mix of HDRFs and BRFs (although the reality is more complex of course as the down-welling diffuse flux is rarely isotropic). I argue that for most cases the AVRIS data will be most closely related to BRFs as most flight campaigns are under relatively clear skies.

I think part of the problem here arises because S06 define anything with more than 0

Using the Hogan et al (2018) paper to defend this position is disingenuous. One of the things that paper shows quite clearly is that solar geometry effects are not well dealt with by classic two-stream formulations for complex canopies. The adjustments to the two-stream scheme made in that paper are required to make the model predicted DHR match the reference 3D model.

I am prepared to accept that in the specific experiment in this paper the limited angular sampling of AVRIS may mean that the overall effect will be small. However, when the authors make statements such as the one at Line 300 ("We therefore conclude that additional computational and conceptual challenges (as well as parameter uncertainties) associated with treatment of angular effects in similar models are unwarranted") it is very misleading: the take home message is that it is, in general, OK to take this approach. It is most definitely not.

On a related note I also don't accept the statements on lines 269 and 299 that seem to imply that it's necessary to have additional parameters to predict directional reflectances. This is not the case and hence also not a valid defence of the approach taken. The set of parameters required to define a two stream model can be used also to define a BRDF model derived from the same underlying assumptions (i.e. semi-infinite, plane parallel turbid media with point scatterers).

Please see Section R1.2.

I propose the following modifications to address these issues:

a) Change the title to remove "cutting out the middleman".

We respectfully disagree. The "middleman" in the title refers to the modeling activities typically used to derive level 3 and 4 products such as GPP and LAI, and in targeting lower-level reflectance data products, we have successfully "cut this middleman out". Taking into account the significant re-framing of the discussion (Section R1.2) in this revision, we feel this title is still accurate.

b) Modify the discussion around this point, including stating clearly that the reflectances and reflectance factors are physically different things and that, in general, it is not appropriate to do assimilate one into a model that predicts the other.

See Section R1.2.

c) Quantify the differences between the BRFs and the modelled BHRs. I noticed, poking around in the github repository, that the SAIL model is included and hence, presumably, it is trivial to take representative posterior parameters values and model both BRF (from SAIL) and BHRs. If the authors' assertion that it shouldn't make any difference is correct then this will help to defend that. The observed reflectance factors should also be compared against SAIL predictions. I would be happy to iterate on the experimental procedure with the authors.

We have added to the results a comparison of simulations between EDR and PRO4SAIL, parameterized identically (Figures 10; lines 442–445). Lower albedo predictions from EDR than SAIL can be attributed primarily to differences in how each model defines the direct radiation backscatter coefficient in the radiative

transfer equation — see our main response (Section R1.3).

The reason that we ultimately did not include SAIL simulations is that there is no obvious way to convert the complex, multi-PFT canopies simulated by EDR into the single homogeneous inputs expected by SAIL. Even if we performed an analogous multi-site calibration of PRO4SAIL, the only calibrated parameters we could compare directly would be the total LAI (and even that would not be a fully apples-to-apples comparison because a PRO4SAIL would optimize that parameter directly for each site, whereas in EDR, we are targeting multiple parameters that interact to give LAI)—for the remaining parameters, we would still have to aggregate the multi-PFT EDR parameters in a non-trivial, non-obvious way.

2. Is ED2 actually used in this paper? It seems that it is only a relatively small component of the model (i.e. the canopy radiative transfer scheme) that has been extracted. I think the title is slightly misleading in this respect. More important, I am not sure calling the code ED2-PROSPECT5 is appropriate. Perhaps EDR-PROSPECT5 would be better?

No, ED2 is not used directly in this analysis. However, the results of this analysis can be used directly to parameterize ED2. The R version of the ED2 canopy radiative transfer model is mathematically identical to the actual Fortran version and was only extracted for analytical convenience and computational efficiency (since running the original version of ED2, which is what we initially implemented, incurred significant computational overhead for initializing ED2’s many variables).

We have kept the title as-is, but have added additional text to the discussion highlighting how this approach could be used directly with the ED2 model (lines 475–482).

3. How are correlations between spectral channels dealt with? Do the authors use all of the AVIRIS observations in the domain 400-1300nm? The errors in spectrally adjacent bands will be very highly correlated and the overall information content will be much lower the same number of independent observations.

Please see Section R1.1.

Minor comments:

L27 The MacBean et al. (2018) paper referenced here does not assimilate a derived product in the sense the authors mean it. The assimilated data in that paper is solar induced fluorescence which, whilst it is “derived” in the sense that it is not what is being measured by the sensor, is no more derived than the surface reflectance data used in the manuscript under review. I suggest finding another reference here.

We have revised this accordingly.

L35 “Meanwhile, the estimating” — “Meanwhile, estimating”

We have revised this accordingly.

L61 Surface reflectance is not assimilated in Zobbitz et al., (2014). The title is misleading (and I regret not standing my ground on that issue when we submitted that paper!); in fact fAPAR data is assimilated.

We have revised this accordingly.

L85 Not sure this line should finish with a colon.

We have revised this accordingly.

Eqn 3 What is meant by  $r_{n+1}$  (and other variables with that subscript)? If I have understood correctly this is there are  $n$  layers, so what is the reflectance of the  $n + 1$  layer? (I am sure I have just missed something here, but it was not obvious).

EDR models fluxes not at canopy layers, but at the boundaries between layers. Boundary  $i$  is defined as the air space immediately below the  $i^{th}$  canopy layer (counting from the bottom of the canopy). For  $n$  canopy layers,  $i = n$  refers to the space immediately below the tallest cohort layer, while  $i = n + 1$  is the (imaginary) boundary above the tallest cohort (i.e., between the canopy and the atmosphere). Therefore,  $r_{n+1}$  refers to the reflectance at the top of the canopy. We have added this clarification to the methods (lines 221–228).

L102 “tau” should presumably be the Greek symbol  $\tau$ .

In the revised methods, this comment no longer applies.

Eqn 7 I am confused by this, why is forward scattering defined by the sum of R+T? This is just the total scattering isn't it?

In the revised methods, this comment no longer applies.

L188 This a different definition of X from earlier in the paper. I suggest finding a different symbol.

We have replaced the earlier definition of  $\mathbf{X}$  (in the EDR matrix equation  $\mathbf{MX} = \mathbf{Y}$ ) with the symbol  $\mathbf{A}$  to eliminate this ambiguity.

L273 “DALEC-predicted foliar biomass, which required introducing an additional fixed parameter (grams of leaf carbon per leaf area) present in neither model.” This statement is incorrect – that parameter already existed DALEC.

We have removed this sentence from our discussion.

## References

- Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, 320(5882):1444–1449.
- Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., Kooper, R., LeBauer, D., Mantooh, J., McHenry, K., and Wang, D. (2014). A quantitative assessment of a terrestrial biosphere model’s data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, 119(3):2013JG002392.
- Lucht, W., Schaaf, C. B., and Strahler, A. H. (2000). An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):977–998.
- Raczka, B., Dietze, M. C., Serbin, S. P., and Davis, K. J. (2018). What limits predictive certainty of long-term carbon uptake? *Journal of Geophysical Research: Biogeosciences*, 123(12):3570–3588.
- Shiklomanov, A. N., Bond-Lamberty, B., Atkins, J. W., and Gough, C. M. (2020). Structure and parameter uncertainty in centennial projections of forest community structure and carbon cycling. *Global Change Biology*, 26(11):6080–6096.
- Singh, A., Serbin, S. P., McNeil, B. E., Kingdon, C. C., and Townsend, P. A. (2015). Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications*, 25(8):2180–2197.
- Viskari, T., Shiklomanov, A., Dietze, M. C., and Serbin, S. P. (2019). The influence of canopy radiation parameter uncertainty on model projections of terrestrial carbon and energy cycling. *PLOS ONE*, 14(7):e0216512.
- Yuan, H., Dai, Y., Dickinson, R. E., Pinty, B., Shangguan, W., Zhang, S., Wang, L., and Zhu, S. (2017). Reexamination and further development of two-stream canopy radiative transfer models for global land modeling. *Journal of Advances in Modeling Earth Systems*, 9(1):113–129.