Response letter for 'ESSD-2024-535 Normalized Difference Vegetation Index Maps of Pure Pixels over China's mainland for Estimation of Fractional Vegetation Cover'.

In the following, we addressed the specific points of the reviewers. Reviewer comments are black font and our responses are blue. We also use red highlights to mark changes in the revised manuscript.
We use codes to the Reviewers' comments, for example R1C2 means Reviewer 1 Comment 2.

Reviewer #2:

Review of ESSD-2024535 Normalized Difference Vegetation Index Maps of Pure Pixels over China's mainland for Estimation of Fractional Vegetation Cover, by Zhao and others.

The authors have addressed an important issue in the use of NDVI for monitoring foliage cover. The end members of the linear transform from NDVI to cover need to be specified and this is commonly ignored. A robust method for routinely identifying these end members across diverse ecosystem types is needed, and is provided in this work. The *Vv* and *Vs* data generated will be valuable. The methods for generating the 500 m version of these variables are sound; the methods for downscaling these to 30 m need improvement. Further, the methods used by the authors for validating these surfaces are not robust and also need to be revised. I recommend a major revision.

Re: **Done.** Thank you for recognizing the effectiveness of our algorithm and for providing constructive suggestions regarding the downscaling logic. In response to the concerns about potential downscaling errors in the 30 m product and limitations in the validation methodology, we have made the following improvements:

1) To address the lack of clarity in the downscaling method, we have provided the complete set of downscale equations (Equation 9), and updated the corresponding textual description to enhance transparency and eliminate possible misunderstandings;
2) We have reorganized the description of the downscaling process, explicitly analyzed the potential errors introduced by the underlying assumption, and acknowledged the possible limitations of the approach. It is worth noting that validation results show good FVC estimation accuracy at 30 m resolution, indicating that the potential errors caused by the downscaling process are acceptable in practice;
3) We have improved the statistical method by using a longer time series (2010-2020) Landsat data to derive the endmember NDVI values, thereby enhancing its representativeness;
4) We have cited the recent work by Donohue and Renzullo (2025), which demonstrates that more sophisticated statistical approaches can also achieve

accurate FVC estimates. In addition, we revised our manuscript to avoid overgeneralized descriptions of the limitations of statistical methods.

R2C1: A significant concern I have is with the downscaling of *Vv* and *Vs*. The method for calculating 500 m *Vs* and *Vv* are sound and the 500 m data are an excellent product. The logic of the downscaling step, and uncertainty about how this downscaling was performed, significantly weakens the quality of the 30 m product. The downscaling introduces the assumption that *Vv* and *Vs* are the same within a given land cover type (line 249). This assumption rarely holds true as soil types (the main driver of *Vs* if one ignores the effects of soil moisture) can vary within single landcover types, or, conversely, different landcovers can share the same soil type. This assumption opens the authors up to the same criticism that they have applied to traditional statistical methods (line 395).

Re: **Done.** Thanks for your insightful comments. Aiming at analyzing and clarifying the potential uncertainty caused by downscaling step, we have reorganized the method (Section 3.2) and discussion (Section 5), including the following points:

1) We appreciate your affirmation of the 500 m *Vv* and *Vs* data. We also published the 500 m *Vs* and *Vv* as supplement, which can facilitate the coarse-resolution FVC estimation. The datalink has been added in Section 7.
2) Unlike traditional statistical methods that often assume a uniform *Vs* value for the same land cover type across large spatial extents (e.g., national or eco-regional scales), our downscaling approach applies this assumption only within a localized 3×3 window of 500 m MODIS pixels. This assumption may introduce uncertainty only when substantial soil type variation exists within the 3×3 MODIS window (1.5km × 1.5km). This implies that the error introduced by assuming homogeneity within the same land cover type is likely to be limited. The relevant description has been clarified in Section 3.2.
3) We fully acknowledge the dependence of *Vs* on soil type. However, the *Vs* values are influenced not only by mineral soil reflectance, but also by non-photosynthetic vegetation (NPV) and biological components such as mosses or lichens. A previous study reported that the NDVI difference between bare soil and NPV endmembers can reach up to 0.2 (Tian et al. 2021), indicating that *Vs* may vary even within the same soil type. Figure 7 further shows that the retrieved *Vs* values deviate from soil NDVI in humid regions, likely due to the influence of surface litter and biological residues. Land cover classification can partly account for such heterogeneity, we used land cover data as a practical proxy for disaggregating *Vs*. Corresponding revisions have been made in the manuscript to clarify this rationale and discuss potential uncertainties in Section 5.
4) Despite the simplifications involved, the comparison with 500 m results shows that the downscaled 30 m *Vs* values achieve comparable accuracy (Song et al., 2022). The statistical comparison shows that the downscaling process introduces minimal changes to the endmember values (Song et al., 2022). This suggests that the downscale process preserves the overall spectral characteristics of the original

MODIS-derived endmembers. Considering the increasing demand for high spatial and temporal resolution applications, we believe that providing 30 m endmember products is of practical significance. Corresponding explanations and references have been added to the manuscript in Section 5.

Newly added Reference: Tian, J., Su, S., Tian, Q., Zhan, W., Xi, Y., & Wang, N. (2021). A novel spectral index for estimating fractional cover of non-photosynthetic vegetation using near-infrared bands of Sentinel satellite. *International Journal of Applied Earth Observations and Geoinformation*, 101, 102361. https://doi.org/10.1016/j.jag.2021.102361

R2C2: Further, it is difficult to understand how this downscaling was performed as the methods do not currently describe a proper unmixing method. Equation 8 apportions *Vs* (or *Vv*) solely according to landcover type proportion, regardless of which landcover type occupies that proportion. As currently described, for a hypothetical 500 m pixel with a calculated *Vs* value and which has 10% area of forest and 10% bare ground (in the surrounding 3x3 window), the method would apportion the same *Vs* value to the forest and bare pixels. Can the authors better explain the method used?

Re: **Done.** Thank you for raising this important point. We have clarified the disaggregation process in the revised manuscript and provide the following step-by-step explanation to address your concern:

1) For each target MODIS pixel, we define a 3×3 window centered on it (i.e., covering 9 MODIS pixels). We assume that within this local window, each land cover type *k* has a consistent endmember value *Vv,k* or *Vs,k*.
2) We construct a system of linear equations (as illustrated in Equation 9), where the known variables are the MODIS-scale *Vv* (or *Vs*) values for the 9 pixels and the land cover fractions *fk,x,y* of each type within each MODIS pixel. The unknowns are the land cover-specific values *Vv,k* (or *Vs,k*) within the window.

$$\begin{cases} V_{v,\text{ modis },x-1,y-1} = \sum_{k=1}^{m} f_{k,x-1,y-1} V_{v,k,x,y} \\ V_{v,\text{ modis },x-1,y} = \sum_{k=1}^{m} f_{k,x-1,y} V_{v,k,x,y} \\ V_{v,\text{ modis },x-1,y+1} = \sum_{k=1}^{m} f_{k,x-1,y+1} V_{v,k,x,y} \\ V_{v,\text{ modis },x,y-1} = \sum_{k=1}^{m} f_{k,x,y-1} V_{v,k,x,y} \\ V_{v,\text{ modis },x,y} = \sum_{k=1}^{m} f_{k,x,y} V_{v,k,x,y} \\ V_{v,\text{ modis },x,y+1} = \sum_{k=1}^{m} f_{k,x,y+1} V_{v,k,x,y} \\ V_{v,\text{ modis },x+1,y-1} = \sum_{k=1}^{m} f_{k,x+1,y-1} V_{v,k,x,y} \\ V_{v,\text{ modis },x+1,y} = \sum_{k=1}^{m} f_{k,x+1,y} V_{v,k,x,y} \\ V_{v,\text{ modis },x+1,y+1} = \sum_{k=1}^{m} f_{k,x+1,y+1} V_{v,k,x,y} \end{cases} \tag{9}$$

3) We solve this overdetermined system to obtain the optimal values *Vv,k* or *Vs,k*. The value corresponding to each land cover type in the center MODIS pixel is then assigned to all the 30 m pixels within that MODIS pixel that share the same land cover type.
4) This 3×3 window is then moved across the MODIS grid to repeat the estimation for each MODIS pixel.

It can be seen that only in the rare case where all the nine MODIS pixels in a 3×3 window have identical land cover proportions (e.g., the same ratio of forest to bare ground), the resulting estimates would be the same. To avoid potential misunderstanding, we have revised the text to clarify the downscaling logic and provided a complete formulation of the equations used in the method (see revised Section 3.2 and Equation 9).

R2C3: I have two significant concerns about the data used to validate/assess their products. The first is the rather unsophisticated way the authors have applied the 'statistical' method. They have only used 3 years of data to derive statistics about *Vv* and *Vs*. What if that period was continually wet, or continually dry, or was fire affected? The derived statistics cannot be assumed to be representative of that site. The authors have the ability to use a much longer time series and should do so. Also, the authors have applied the method with the expectation that it will work everywhere, which it is known not to. The method cannot return reliable *Vs* values in heavily vegetated areas nor *Vv* in sparsely vegetated areas. While the authors acknowledge this in the conclusion, this knowledge hasn't been applied in their design of the derivation of the statistically derived *Vs Vv* data. And so it is no surprise that this product performs poorly in these respective situations. This led the authors to conclude that (line 442)
"Traditional statistical methods are impractical to achieve this goal due to their reliance on pure pixels."
This is not universally true. More sophisticated implementations of the statistical method can be quite effective. Can the authors at least provide some more context to the reader about the simplicity of their approach relative to alternative approaches? Or maybe the authors could restrict the application of their statistical method to where it is known to be valid and hence avoid reporting values where it quite rightly doesn't work. None of this will change the excellent result that the multi-VI method is superior.
Re: **Done.** Thank you for your thorough and constructive comments. We have carefully revised the manuscript to address your concerns regarding the statistical method used for deriving *Vv* and *Vs*.

1) To improve the robustness of the statistical endmembers, we have recalculated *Vv* and *Vs* pixel by pixel using a longer time series of Landsat data. Specifically, the maximum and minimum NDVI values over the period 2010–2020 were used to represent *Vv* and *Vs*, respectively. After this update, the rationality of the statistical endmembers has improved: *Vv* values increased in humid region, while *Vs* values decreased in arid area. The corresponding description has been updated in Section 2.2.1.
2) We clarified that the statistical method used in this study yields reasonable FVC estimation in most regions, except in evergreen forest areas and extremely arid zones. The updated content demonstrates that, outside of these extreme regions, the statistical endmembers provide reliable FVC (Figure 9). The practicability of the statistical method has now been explicitly stated, and the corresponding analysis has been supplemented in Sections 4.2 and 4.3.

3) We fully acknowledge the practicality of statistical methods, especially their simplicity and reasonable accuracy in suitable area. In fact, the MultiVI model proposed in this study incorporates statistical endmembers as boundary for inversion. We have revised the manuscript in the introduction and discussion to avoid overgeneralized or dismissive statements regarding statistical methods, and have expanded our discussion to better reflect their strengths and appropriate use cases. In addition, we have cited the recent work by Donohue and Renzullo (2025), which demonstrates that improved statistical implementations—when used with appropriate constraints—can achieve reliable results.

R2C4: The second concern I have about the data used to validate/assess their products relates to how the field data at Heihe were derived. In scaling the field observations from 10 x 10 m to 90 x 90 m, the authors have effectively turned the field observations into a modelled product with its own errors. I would expect that a direct comparison between the 10 m field data and the 30 m *Vs Vv* data would provide a more robust comparison than upscaling the field data.

Re: **Done.** Thank you for your insightful comment. Following your suggestion, we have revised the validation approach for the Heihe site. Specifically, we directly compared the 10 × 10 m field-measured FVC with the 30 m MultiVI and statistical FVC estimates, instead of upscaling the field data. The updated validation results are now presented in Figure 9.

R2C5: One last comment is that some recent work is of direct relevance to this Mutil-VI paper (Donohue and Renzullo, 2025; https://doi.org/10.1071/BT24060). I expect this would have been published after the current manuscript's submission; however, it may be of interest to the authors. In making this statement I should also disclose that this is my paper (it's Randall Donohue here).

Re: **Done.** Thank you for sharing your recent work and for disclosing your authorship. We appreciate the valuable contribution of your study, which proposes improvements to traditional statistical methods and demonstrates their effectiveness in estimating FVC over Australia. We have cited this reference in the revised manuscript to acknowledge that the statistical method adopted in our study is relatively simple and that more advanced implementations, such as yours, can achieve high estimation accuracy.

Newly added Reference: Donohue, R. J. and Renzullo, L. J.: An assessment of the accuracy of satellite-derived woody and grass foliage cover estimates for Australia, Aust. J. Bot., 73, BT24060, https://doi.org/10.1071/BT24060, 2025.

R2C6: Lines 42 and 49. The VI-based mixture model referred by the authors is specifically the NDVI-based mixture model. It is not a generic model that can use *any* vegetation index.

Re: **Done.** Thank you for your helpful comment. While the current study focuses on the NDVI-based mixture model, we would like to clarify that both the traditional VI-based mixture model and the proposed MultiVI method are applicable to other vegetation indices, such as EVI. Previous studies applying MultiVI to EVI have also reported high estimation accuracy (Song et al., 2022a). To avoid confusion, we have added a sentence in the introduction to clarify that the model framework is not limited to NDVI.

R2C7: Line 173. Doesn't look like the UAV data were used for anything at the Hebei site. Do they need to be mentioned at all?
Re: **Done.** Thank you for pointing this out. At the Hebei site, the grassland data were indeed acquired using UAV, and these data were included in the validation analysis. We have revised the text to clarify the role of the UAV data and avoid potential confusion.

R2C8: Line 200. It is a misconception that the NDVI has a saturation effect. When compared to foliage cover (which it what is has been shown to be linearly related to) there is no 'saturation'. This misconception arises when NDVI is incorrectly expected to bear some relationship with leaf area.
Re: **Done.** Thanks for your suggestion. Instead of using the term "saturation," we have revised the text to emphasize the nonlinear relationship between NDVI and FVC. Although NDVI and FVC exhibit an approximately linear relationship in pure pixel assumptions, several studies have shown that, in practice, NDVI often displays a nonlinear response to FVC in mixed pixels due to the influence of multiple factors (Montandon and Small, 2008). These include soil background variability, sub-pixel shadow fractions, viewing geometry, terrain effects, and especially the spatial scale of observation (Mu et al. 2024). We revised the sentence in the manuscript to avoid misunderstanding.
Newly added Reference: Mu, X., Yang, Y., Xu, H., Guo, Y., Lai, Y., McVicar, T. R., Xie, D., & Yan, G. (2024). Improvement of NDVI mixture model for fractional vegetation cover estimation with consideration of shaded vegetation and soil components. Remote Sensing of Environment, 314, 114409. https://doi.org/10.1016/j.rse.2024.114409

R2C9: Line 229. Calling the values derived from a single year (2014) the 'historical' values is counterintuitive. They are not representative of site history.
Re: **Done.** Re: Thank you for your comment. To avoid the misunderstanding, we have revised the wording to clarify that the minimum and maximum NDVI values were derived from all available observations within the year 2014.

R2C10: Line 229. How much does using statistics derived from only one year of data (2014) limit the accuracy of the method when applied to other years? I would think it important to derive these 'historical' values from as long a time series as possible (which would be 23 or so years for MODIS).

Re: **Done.** Thank you for your comment. In response, we have revised the manuscript and addressed this issue from the following perspectives:

1) We acknowledge that assessing the representativeness of single-year data is important. As detailed in our response to R1C1, we conducted a supplementary analysis comparing *Vv* and *Vs* from 2014 with those from 2018 and 2022. The results show minimal interannual differences. We also clarified that the NDVI values of pure vegetation and bare soil pixels are generally stable across years unless affected by abrupt disturbances. Please refer to our response to R1C1 for full justification and supporting evidence.
2) The MultiVI algorithm estimates *Vv* and *Vs* by solving equations derived from two angular observations with significantly different NDVI values. As long as the differences of angular observations are sufficient, a valid solution can be obtained. A single year of MODIS data recorded a complete vegetation growth cycle, ensuring the availability of angular NDVI pairs with sufficient contrast for reliable inversion. Our experiments show that introducing too many angular observations can lead to overfitting, which degrades the estimation accuracy of *Vv* and *Vs*. Therefore, selecting a representative set of well-separated angular observations from one year is an effective strategy to ensure solution quality while avoiding overfitting.