

This manuscript examines how shortwave radiation modulates VPD effects on forest carbon exchange across European flux sites. The topic is timely and the multi-method approach is potentially valuable, but the validation strategy, threshold interpretation, and mechanistic attribution need additional clarification and revision.

Response: We sincerely thank the reviewer for the careful reading of our manuscript and for the valuable and constructive comments and suggestions, which will help us improve the robustness and interpretation of the study. We will revise the manuscript accordingly and address each comment in detail as described below.

Major comments

Lines 136–143: The XGBoost model uses a random 80/20 split of half-hourly observations. Because neighboring observations from the same site and event can enter both subsets, temporal autocorrelation may inflate validation performance. A temporally blocked and/or leave-one-site-out validation would be more appropriate.

Response: We agree that a random split of half-hourly observations may lead to optimistic estimates of model performance because observations from the same site can occur in both the training and validation subsets. In the revised manuscript, we will therefore conduct leave-one-site-out (LOSO) cross-validation as a more stringent robustness test of model generalizability across sites. We will report the LOSO model performance and compare it with the performance obtained using the current random-split validation.

Lines 203–229: G_s is derived by inverting the Penman–Monteith equation, in which VPD already appears explicitly, and is then regressed against VPD. Thus, part of the negative G_s –VPD relationship is mathematically imposed. The authors should quantify this dependence or weaken the mechanistic interpretation of Fig. 7.

Response: We agree with the reviewer that, because VPD explicitly enters the inverted Penman–Monteith equation used to derive G_s , part of the observed G_s –VPD relationship may be structurally imposed. In the revised manuscript, we will explicitly acknowledge this mathematical dependence and weaken the mechanistic interpretation of Fig. 7. We will present the G_s analysis as being consistent with rather than providing independent evidence for stomatal regulation of the VPD–NEE relationship.

Lines 159–167 and 294–296: The very narrow confidence intervals for the reported thresholds likely underestimate uncertainty because the bootstrap resamples individual half-hourly observations and does not account for temporal dependence, site clustering, or uncertainty in the XGBoost model. A site- or site-year-level bootstrap would be more convincing.

Response: We agree that sampling individual half-hourly observations may underestimate threshold uncertainty because observations are temporally correlated and clustered within sites. In the revised manuscript, we will therefore use a site-level cluster bootstrap to provide a more conservative assessment of threshold uncertainty.

Lines 168–198: The SEM is based on observational data and predefined directional paths. Its coefficients should therefore be interpreted as conditional associations rather than direct causal effects.

Response: In the revised manuscript, we will clarify this limitation and that the SEM results should be interpreted as conditional associations consistent with the hypothesized pathways, rather than as evidence of direct causal effects. We will also revise causal wording throughout the Methods, Results, and Discussion where appropriate.

Lines 91–103: The use of surface-layer SWC could underestimate soil-moisture effects, especially in forests, because surface SWC may not adequately represent root-zone water availability. Therefore, the relatively weak SWC contribution should be

interpreted cautiously.

Response: We agree that surface-layer SWC may not fully represent the root-zone water availability experienced by trees, particularly in forests with relatively deep rooting systems. In the revised manuscript, we will explicitly acknowledge this limitation and interpret the relatively weak contribution of SWC more cautiously. In particular, we will clarify that the weak contribution of surface SWC in our analysis should not be interpreted as evidence that soil-moisture limitation is generally unimportant for forest carbon uptake.

Lines 99–116 and 475–497: The main analysis is based on daytime growing-season NEE rather than annual carbon balance or changes in carbon storage. The manuscript should therefore refer more cautiously to carbon exchange or carbon uptake rather than “carbon sequestration,” particularly in the title and conclusions.

Response: We agree that “carbon uptake” or “carbon exchange” provides a more precise description of the quantity analyzed because our analysis focuses on daytime growing-season NEE rather than annual ecosystem carbon balance or long-term changes in carbon storage. In the revised manuscript, we will replace “carbon sequestration” with “carbon uptake” where appropriate, including in the title, Abstract, Discussion, and Conclusions.

Lines 100–108: Please report site-level sample sizes and assess whether sites with longer records disproportionately influence the pooled analyses. The dataset includes 23 sites, but the available record length differs substantially among sites.

Response: Thank you for this suggestion. In the revised manuscript, we will report the number of available years and observations for each site to make the contribution of individual sites more transparent. We will also assess whether sites with longer records disproportionately influence the pooled analyses, including through a site-balanced sensitivity analysis in which the influence of individual sites is more equally

represented. We will report the results of this sensitivity analysis and discuss any implications for the main conclusions.

Lines 292–307: The 3.90 hPa threshold should be interpreted cautiously as a model-derived SHAP threshold rather than a directly observed physiological threshold.

Response: We agree with the reviewer. In the revised manuscript, we will make this distinction explicit throughout the Results and Discussion and will interpret the value more cautiously as a model-derived SHAP threshold or response-regime breakpoint.

Lines 240–247 and 271–275: The manuscript should avoid describing SHAP importance as “variability explained.” SHAP values quantify contributions to model predictions rather than variance partitioning.

Response: In the revised manuscript, we will replace expressions such as “variability explained” with more accurate terms such as “relative SHAP importance” or “contribution to model predictions” as appropriate throughout the manuscript.

Lines 428–442: The potential circularity in the GPP analyses due to temperature- and VPD-dependent flux partitioning should be emphasized more clearly when interpreting the GPP results.

Response: Thank you for pointing this out. In the revised manuscript, we will emphasize this limitation more clearly when interpreting the GPP results. We will also clarify that the GPP analyses will be treated as complementary evidence rather than as independent validation of the VPD effects or threshold estimates derived from the directly observed NEE data.