

EGU Biogeosciences

Prof. Lishan Ran

Associate editor

Dear Prof. Ran,

Thank you for handling the review of our manuscript, 'Satellite-derived management indicators improve modeling of water and greenhouse gas fluxes in Swiss agroecosystems' (egusphere-2026-3522). We also thank the reviewers for their comments. We have made several major revisions to the manuscript.

We clarified the source and processing of the Earth observation products and reorganized the methodological framework to explain the derivation of the management indicators and the rationale for the four modeling scenarios, as suggested by both reviewers.

In response to Reviewer #2, we extended the DII evaluation using available harvest dry-matter records across the study sites and reassessed the background-window selection using KGE*, which jointly accounts for correlation, variability, and standardized mean bias.

We also added site-specific performance tables and extended the evaluation of models using non-interpolated drivers to multiple sites and years, addressing Reviewer #2's requests for a broader assessment of model performance.

We clarified the proof-of-concept scope, the limits of model transferability, temporal-scale matching, and gaps in satellite observations. We also revised ambiguous statements and figure captions.

Please find our detailed point-by-point responses below. Reviewer comments are reproduced in dark blue. Line numbers in our responses refer to the clean revised manuscript. Line numbers in the reviewer comments refer to the original submission.

Thank you for considering the revised manuscript.

Sincerely,

Dr. Aolin Jia, on behalf of the co-authors

aolin.jia@usys.ethz.ch

===== Reviewer #1 =====

Overall assessment. The manuscript (egusphere-2026-3522) "Satellite-derived management indicators improve modeling of water and greenhouse gas fluxes in Swiss agroecosystems" describes the development of an RS-assisted modeling framework to estimate daily latent heat flux (LE), net ecosystem CO₂ exchange (NEE), nitrous oxide (N₂O) and methane (CH₄) fluxes across six Swiss FluxNet sites (two croplands and four grasslands) between 2016 and 2025.

The framework is based on Sentinel-2 time series to derive leaf area index and RS-based field management indices (RS-FMIs), allowing to detect mowing events, quantifying defoliation intensity, and identifying crop rotation and bare soil periods. Four scenarios were tested to predict LE, NEE and GHG; 1) meteo only, 2) Meteo and field management, 3) meteo and RS +vegetation performance index (VPI), and 4) Meteo, RS+VPI and field management indices (FMI).

Several indicators were combined with meteorological drivers to train XGBoost models for each ecosystem type (crop and grass) and target variable separately, and driver contributions were evaluated using SHapley Additive exPlanations (SHAP) analysis in order to assess the relative importance.

Analyses showed that RS-FMIs effectively captured in situ recorded management events and enabled improved rebuilding of daily flux variability. Model performances were strong for LE ($R^2 \approx 0.89\text{--}0.90$) and NEE ($R^2 \approx 0.59\text{--}0.71$), whereas N₂O and CH₄ fluxes were reproduced with moderate accuracy ($R^2 \approx 0.37\text{--}0.55$). Models using RS-FMIs performed similarly to those using well-compiled in situ management records, supporting the ability of RS-derived vegetation and management indicators to represent management effects. Likewise the study showed that the incorporation of satellite-derived vegetation dynamics and disturbance signals, have a broader potential of remote sensing-informed management representations for monitoring agricultural greenhouse gas fluxes.

I read the manuscript with interest. It presents a novel methodology for predicting fluxes and analysing their drivers in agroecosystems. The manuscript is well written and well illustrated. Although the methodology involves complex modelling and data processing, the manuscript is generally easy to follow. In my opinion, the manuscript is suitable for publication after minor revisions

Response: We thank you for your careful reading and assessment of our manuscript. We address your general and specific comments point by point below.

General comments.

General Comment 1. Introduction: I suggest adding a few lines discussing the temporal heterogeneity of the different measurements.

Response: We have revised the Introduction to identify the temporal mismatch among half-hourly flux records, intermittent management events, and satellite observations (Introduction Section, Lines 48–50):

“Furthermore, half-hourly flux records, intermittent management events, and satellite observations obtained on discrete acquisition dates represent different temporal scales. This temporal heterogeneity complicates the alignment of management events and satellite-derived indicators with observed flux variability.”

General Comment 2. Using EO images for modelling is not straightforward because of several limitations that are not mentioned, such as cloud cover, satellite revisit frequency, pixel size and image quality (see also L123ff).

Response: The revised Introduction discusses how satellite revisit intervals, cloud cover, image-quality screening, and spatial resolution constrain field-scale management monitoring. We describe Sentinel-2’s nominal revisit interval and explain how cloud cover and image-quality screening can lengthen gaps between usable observations. We also explain that MODIS provides more frequent observations at a coarser spatial resolution, with potential mixed-pixel effects (Introduction Section, Lines 43–54).

“The Sentinel-2 constellation provides a nominal 5-day revisit interval at 10–20 m spatial resolution, supporting the detection of rapid management-induced surface changes at the field scale (Claverie et al., 2018; Liang et al., 2024). Satellite time series have increasingly been used to identify the timing and frequency of mowing, crop harvest, and other defoliation events (e.g., grazing and tillage), benefiting from the improved spatial and temporal resolution of recent EO products (Andreatta et al., 2022; De Vroey et al., 2022; Reinermann et al., 2020; Schwieder et al., 2022; Zhou et al., 2021). Furthermore, half-hourly flux records, intermittent management events, and satellite observations obtained on discrete acquisition dates represent different temporal scales. This temporal heterogeneity complicates the alignment of management events and satellite-derived indicators with observed flux variability. For all optical sensors, cloud cover and image-quality screening can reduce the number of usable observations. Gaps may therefore exceed the nominal revisit interval, making rapid management-induced changes harder to capture. MODIS offers near-daily temporal sampling, but at a coarser spatial resolution of 250–1000 m (Xiong et al.,

2019). Potential mixed-pixel effects therefore require evaluation for field-scale management monitoring in Switzerland.”

General Comment 3. Furthermore, fluxes are measured every 30 minutes (aggregated to daily values), whereas management events occur at much lower temporal frequency. Matching these different temporal scales is therefore an important challenge and deserves some Intro and discussion.

Response: We thank you for highlighting temporal-scale matching. We now introduce this challenge explicitly in the Introduction (Lines 48–50), as described in our response to General Comment 1 above.

As described in Methods, all four target fluxes were aggregated from half-hourly records to daily means (Lines 107–108). Days with more than 50% gap-filled half-hourly records were excluded, and measured and gap-filled records were combined for the retained days (Lines 108–110). For models using management records, individual operations were represented by daily time-since-event variables (Lines 98–100). Continuous Sentinel-2 predictors were interpolated to a daily time step (Appendix E, Lines 612–618), whereas crop type was assigned to each day from the nearest classification result (Methods, Lines 167–168). These steps aligned the predictors and target fluxes at a daily time step.

We have also introduced this comparison in the Modeling Configuration Section (Lines 256–258):

“We also evaluated model performance using non-interpolated continuous Sentinel-2 predictors and compared the results with those of the daily analysis using interpolated predictors (Appendix E). The two analyses differed in sample composition and therefore did not constitute a paired comparison on identical dates.”

We have condensed the related Discussion text into one paragraph on daily temporal matching, interpolation limitations, and possible sub-daily applications. Appendix E compares model performance using non-interpolated continuous satellite predictors with the daily analysis using interpolated continuous satellite predictors (Lines 619–637). The revised Discussion text reads (Discussion Section, Lines 494–503):

“We aggregated half-hourly flux records to daily means and combined them with daily time-since-event variables, interpolated continuous Sentinel-2 predictors, and daily crop-type assignments based on the nearest classification result. Daily flux modeling follows previous analyses of carbon and energy fluxes and management-related N₂O and CH₄ flux dynamics (Feigenwinter et al., 2023a; Tramontana et al., 2016). These processing steps aligned the model inputs and targets at a daily time step, while temporal interpolation does not provide independent observations of rapid changes

between satellite acquisitions. Sub-daily modeling would need to represent the pronounced diurnal cycles of LE and NEE. Combining GEO and polar-orbiting imagery could support sub-daily LE and NEE modeling; however, kilometer-scale GEO pixels limit field-level management monitoring (Geng et al., 2022; Rains et al., 2024). Cloud cover would still limit the availability of usable optical and thermal infrared observations. In comparison, N₂O and CH₄ fluxes show less regular diurnal patterns in managed grasslands (Feigenwinter et al., 2023a; Hörtnagl and Wohlfahrt, 2014).”

General Comment 4. Materials and Methods: More clarification is needed regarding the EO products. It is not always clear where the products originate from (Copernicus, Sentinel processing chain, published datasets, etc.), how they were generated, whether they were produced by the authors, and whether existing products or previously published methods were used. Please clarify this around L121ff and provide appropriate references (DOIs where available).

Response: Thank you for requesting clarification of the product sources and processing methods. The revised Sentinel-2 Satellite Products Section now identifies Agroscope as the producer of the LAI, fractional cover, and crop-type products. It describes how published retrieval and classification methods were adapted to Swiss vegetation, soil, and agricultural conditions. We have added supporting references with DOIs for these products and their generation methods (Sentinel-2 Satellite Products Section, Lines 130–169). We have also clarified that national-scale daily maps are being produced by Agroscope and that samples of the data used in this study are available upon request (Lines 523–525). The revised passages read as follows:

“Fractional vegetation cover (FVC) was estimated by Agroscope using a spectral mixture analysis framework (Ledain et al., 2026a). Satellite surface reflectance was represented as a linear combination of photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and bare soil. Neural network models were trained to predict the fractional covers of these three components. The method followed Lobert et al. (2025), with adaptations to Swiss vegetation and soil conditions.”

“Leaf area index (LAI) was retrieved from Sentinel-2 imagery by Agroscope using a neural network (NN) trained for this retrieval. The approach followed the development of operational global LAI products (Weiss and Baret, 2016), with adaptations to Swiss vegetation and soil conditions (Ledain et al., 2026b).”

“Crop type maps were produced by Agroscope following Turkoglu et al. (2025, 2021) from Sentinel-2 time series imagery using a U-Net Temporal Attention Encoder (U-TAE) (Sainte Fare Garnot and Landrieu, 2022). This deep

learning architecture was designed for multi-temporal satellite image segmentation. The classification approach was further adapted to Swiss agricultural conditions, as described by Lauber et al. (2026).”

The Data Availability statement reads (Lines 523–525):

“National-scale daily maps are currently being produced by Agroscope; samples of the LAI, fractional cover, and crop-type data used in this study are available upon request.”

General Comment 5. The section describing the Sentinel-2 products contains a lot of information and is sometimes difficult to follow. I wonder whether a few italic subheadings would help distinguish between existing methods/products and the developments introduced by the authors. Alternatively, a small summary table could be useful, listing existing products, processing steps, improvements, training, and outputs.

Response: We adopted your suggestion to use subheadings. This study evaluates RS-derived indicators for flux modeling, rather than developing new retrieval or classification algorithms. We therefore describe the product sources, underlying methods, and available accuracy assessments, with references for further methodological details. We have added four italic, unnumbered subheadings aligned with the left margin to organize the Sentinel-2 Satellite Products Section: *Sentinel-2 surface reflectance*, *Fractional vegetation cover*, *Leaf area index*, and *Crop type* (Lines 123–169). These subheadings distinguish the input reflectance data from the Agroscope-derived products.

General Comment 6. Regarding Sections 2.2.2 to 2.2.4 onwards, I may simply have misunderstood the workflow, but this could indicate that L165–173 would benefit from additional explanation. Personally, I would move the model framework section after the descriptions of remotely sensed defoliation event detection, defoliation intensity estimation, and bare soil period detection, which would make the methodological flow easier to follow.

Response: We thank you for this suggestion. We have reorganized the Methods Section to introduce the RS-FMIs before their application in flux modeling.

Specifically, we added a brief opening paragraph explaining the motivation for deriving management indicators from remote sensing (Methods Section, Lines 171–173):

“Field management records are often unavailable or inaccessible, limiting their use in flux modeling. We therefore derived RS-based field management indicators (RS-FMIs) of defoliation timing, disturbance intensity, and bare soil duration. The following subsections describe how the indicators were calculated and used in flux modeling.”

We then moved the framework description and its figure after the three subsections on defoliation event detection, defoliation intensity estimation, and bare soil period detection, and renamed it “Methodological Framework for Multiple Modeling Scenarios” (Methods Section, Lines 229–242).

General Comment 7. Section on sentinel-2 products, is a lot information and reader has difficulties to follow. I wonder if some sub-titles (in italic) may help to identify what was taken over from existing work/methods and what was added here I wonder if a small table may be useful... eg 3x columns existing products and work/improvement/choices/training... ect by authors

Response: As described in our response to General Comment 5, we have added four italic subheadings and clarified the provenance of the products and their underlying methods to make the section easier to follow (Sentinel-2 Satellite Products Section, Lines 123–169).

General Comment 8. Sections 2.2.2 to 2.2.2, might that I did not well understand, which might also show that L165-173 would need more detail (which arrive in L244), personally I would move the model framework after the Remotely Sensed Defoliation Event Detection, Defoliation Intensity Estimation and Bare Soil Period.

Response: As described in our response to General Comment 6, we have moved the renamed Methodological Framework for Multiple Modeling Scenarios Section after the three RS-FMI subsections and before the Modeling Configuration Section, bringing the framework and its implementation together (Methods Section, Lines 229–242).

General Comment 9. discussion and Conclusion: In the last part of the Discussion (around L494 and L504) and in the Conclusions, the manuscript ends without a very clear take-home message. I encourage the authors to be more explicit about the readiness of the proposed framework. Is this mainly a proof-of-concept or first step, or is it already a mature and operational product? What are the main remaining limitations? What further developments would be needed before applying the framework more broadly or operationally?

Response: We thank you for suggesting a clearer statement of the framework’s readiness and main contribution. We have clarified that the present study provides a proof of concept evaluated within the Swiss FluxNet site network and identified two priorities for further development: enhancing spatial transferability and improving the representation of high-emission events (Discussion Section, Lines 507–510). The revised text reads (Conclusions Section, Lines 517–521):

“The study provides a proof of concept for using these indicators in agroecosystem flux models. Further development could support regional flux monitoring and the reconstruction of historical management regimes.

Operational application requires better spatial transferability and more accurate predictions of high-emission events. Such developments could support the assessment of agricultural climate mitigation measures, sustainable management practices, and greenhouse gas emissions reporting.”

We have also clarified the proof-of-concept scope in the Abstract (Lines 18–20):

“These results provide a proof of concept for using RS-FMIs to reconstruct management information and improve the representation of management effects in agroecosystem flux models.”

Specific comments.

Specific Comment 1. L5 abstract though the work is clear the why the variables where chosen to be important is a little bit unclear I suggest to: “Here, we developed an RS-assisted modeling framework to estimate to determine water and GHG fluxes (daily latent heat flux (LE), net ecosystem CO₂ exchange (NEE), nitrous oxide (N₂O) and methane (CH₄) fluxes) across six Swiss FluxNet sites (two croplands and four grasslands) between 2016 and 2025.

Response: We have revised the sentence as you suggested to introduce water and GHG fluxes before listing the four target variables (Abstract, Lines 5–8):

“Here, we developed an RS-assisted machine learning framework to estimate daily water and GHG fluxes across six Swiss FluxNet sites (two croplands and four grasslands) between 2016 and 2025. Target variables were latent heat flux (LE), net ecosystem CO₂ exchange (NEE), and nitrous oxide (N₂O) flux and methane (CH₄) flux.”

Specific Comment 2. L45-50 mention models DNDC and DayCent however the manuscript does not work with these models... I wonder if the sentence can be removed. L43 “To move beyond site-level observations, regional studies and modelling frameworks have been developed to characterize GHG flux dynamics. Process-based models (Smith et al., 2008) such as the Denitrification-Decomposition model (DNDC) or DayCENT, enable large-scale simulation of carbon and nitrogen fluxes, but....their performance strongly depends “

Response: We have removed the passage introducing DNDC and DayCENT and the preceding transition sentence, keeping the Introduction focused on the motivation for the RS-assisted modeling framework used in this study (Introduction Section, Lines 33–62).

Specific Comment 3. L51 and L55ff: This section is somewhat vague. I suggest mentioning that the availability and spatial resolution of satellite products have greatly improved in recent years. For example: "Satellite time series have increasingly been used to identify the timing and frequency of mowing, crop harvest and other defoliation events, benefiting from the improved spatial and temporal resolution of recent EO products."

Response: The revised sentence links management-event detection to improvements in the spatial and temporal resolution of EO products. We retained this statement in the Introduction (Lines 45–48). The nominal Sentinel-2 revisit interval is described immediately before it (Lines 43–45). The retained sentence reads:

"Satellite time series have increasingly been used to identify the timing and frequency of mowing, crop harvest, and other defoliation events (e.g., grazing and tillage), benefiting from the improved spatial and temporal resolution of recent EO products (Andreatta et al., 2022; De Vroey et al., 2022; Reinermann et al., 2020; Schwieder et al., 2022; Zhou et al., 2021)."

Specific Comment 4. L63 I suggest to reword. EO images in modelling is not so easy task due to many points not mentioned here, clouds, satellite passage, pixel size and quality.... so I suggest to reword "In this study aimed to examine whether incorporating high-resolution, satellite-derived management and vegetation indicators can enhance the simulation and driver attribution of...."

-It would also be useful here to briefly acknowledge some of the practical limitations of EO-based modelling (cloud cover, revisit time, spatial resolution, etc.).

-EO images in modelling is not so easy task due to many points not mentioned here, clouds, satellite passage, pixel size and quality (see L123ff)

-fluxes being measured on 30min (daily) and management events on monthly so to find the matching moment is a challenge

Response: We thank you for these suggestions. We have addressed the points as follows:

1. Study objective. We have adopted your suggested wording to frame the study objective as examining whether satellite-derived management and vegetation indicators could improve flux simulation and driver attribution (Introduction Section, Lines 63–71).

2. Practical limitations of EO-based modeling. The Introduction now describes the constraints imposed by spatial resolution and revisit time. It also explains how cloud cover and image-quality screening create gaps that make rapid management-induced changes harder to capture, as discussed in General Comment 2 (Introduction Section, Lines 43–54).

3. Temporal-scale matching. We introduce this challenge in the Introduction (Lines 48–50). As explained in our response to General Comment 3, daily processing aligns inputs and

targets but interpolation does not provide independent observations between satellite acquisitions (Discussion Section, Lines 494–498).

Specific Comment 5. L121: Please clarify where the EO products originate from. Were they obtained from Copernicus or another provider? Were they processed by the authors? What spatial resolution was used?

Response: We thank you for requesting this clarification.

1. Product origin and processing. As detailed in our response to General Comment 4, the LAI, FVC, and crop-type products were generated by Agroscope from Sentinel-2 imagery. We have clarified their origin and described the retrieval and classification methods used to produce them (Sentinel-2 Satellite Products Section, Lines 130–169).

2. Spatial resolution. We have specified the 10 m output pixel size of the LAI, FVC, and crop-type products and distinguished it from the native spatial resolutions of the Sentinel-2 input bands (Sentinel-2 Satellite Products Section, Line 169). The added sentence reads:

“The LAI, FVC, and crop-type products used in this study had a spatial resolution of 10 m.”

Specific Comment 6. L165 Consider a more explicit title, for example: “methodological framework for multiple modelling scenarios”.

Please also explain the rationale behind selecting the four modelling scenarios. At present, the section is very brief and the progression from Scenario 1 to Scenario 4 is not entirely obvious. There is more detail in L244 but this is much later. Please think about moving sections ect

Response: We thank you for these suggestions.

1. Section title and organization. We have adopted your suggested title, “Methodological Framework for Multiple Modeling Scenarios”. As described in our response to General Comment 6, this section has been moved immediately before the Modeling Configuration Section, bringing the framework and its implementation together (Methods Section, Lines 229–242).

2. Rationale for the four scenarios. We have moved the scenario definitions into the framework description and explained the purpose of each configuration. We now explicitly distinguish the progressive addition of RS information in Scenarios 1-3 from the separate management-record benchmark in Scenario 4 (Methods Section, Lines 230–238). The revised passage reads:

“The RS-FMIs described above and LAI were integrated with meteorological drivers (METEO) to construct four modeling scenarios (Figure 2):

- (1) METEO only provided the meteorological baseline.
- (2) METEO + RS-VPI assessed the additional contribution of vegetation information.
- (3) METEO + RS-VPI + RS-FMIs assessed the additional contribution of remotely-sensed management indicators beyond meteorological and vegetation information.
- (4) METEO + MGMT provided a separate benchmark incorporating site management records.

Scenarios 1–3 therefore progressively added remotely-sensed information, whereas Scenario 4 evaluated the contribution of recorded management information relative to the same meteorological baseline.”

Specific Comment 7. Figure 3 legend: Please explain in the legend what the bold blue line represents compared to the pale blue line.

Response: We have clarified that the wider dark-blue bars in panel (a) indicate grazing events recorded on consecutive days, whereas the thin light-blue lines mark individual recorded defoliation events. The revised caption reads (Evaluation of Remotely-Sensed Defoliation Intensity Index Section, Figure 3 caption, p. 13):

“In panel (a), thin light-blue lines mark individual recorded defoliation events, whereas wider dark-blue bars indicate grazing events recorded on consecutive days.”

Specific Comment 8. Figure 5 legend: Modeled versus observed daily fluxes for four grassland sites under four scenarios (vertical panels).

Response: We have clarified that columns correspond to the four modeling scenarios and rows correspond to the four flux variables. The revised caption begins (Model Evaluation Section, Figure 4 caption, p. 14):

“Modeled versus observed daily fluxes for four grassland sites under four modeling scenarios arranged in columns. Rows show latent heat flux (LE, a–d), net ecosystem exchange of CO₂ (NEE, e–h), nitrous oxide flux (N₂O flux, i–l), and methane flux (CH₄ flux, m–p).”

Specific Comment 9. Figure 8 Legend. The legend would benefit from additional explanation to the phrase “SHAP values representing positive effects on the predicted fluxes compared to the overall mean, and vice versa for negative SHAP values. Points are colored by driver value.”

At present, it is difficult to interpret the figure because each driver contains both red and blue points of varying density. For example, for grassland CH₄, the "day of defoliation" variable contains both red and blue points. Does this indicate that high values can have both positive and negative effects depending on the context? A few additional sentences in the legend would greatly help readers interpret these plots correctly.

Response: We thank you for pointing out that the graphical conventions of the SHAP plots were insufficiently explained.

1. SHAP values. SHAP values on the horizontal axis indicate each driver’s contribution to the model prediction. Positive values increase the prediction relative to the model baseline, whereas negative values decrease it. Values farther from zero indicate larger contributions.

2. Point colors. Each point represents one sample for a given driver. Point colors represent driver values, from low (blue) to high (red). A red point far to the right represents a sample in which a high driver value makes a large positive contribution to the predicted flux.

3. Variation in point density. Points with similar SHAP values are displaced vertically to reduce overlap. Greater vertical thickness at a given SHAP value indicates more samples with similar contributions. The vertical position itself has no additional quantitative meaning.

4. The grassland CH₄ example. The variable is Days since Defoliation, representing the elapsed time since the most recent defoliation event. Red points indicate relatively longer intervals and blue points relatively shorter intervals. In panel (g), longer intervals generally correspond to negative SHAP contributions and shorter intervals to positive contributions, although exceptions occur. Similar driver values can have contributions of different signs across samples, depending on the values of other predictors. These patterns describe model behavior and should not be interpreted as causal effects of defoliation on CH₄ flux.

We have revised the caption to clarify the graphical conventions while keeping the detailed example in this response (Driving Factor Analysis Section, Figure 7 caption, p. 19):

“Each point represents one sample for a given driver. SHAP values on the horizontal axis indicate each driver’s contribution to the model prediction. Positive values increase the prediction relative to the model baseline, whereas negative values decrease it. Values farther from zero indicate larger contributions. Point colors represent driver values, from low (blue) to high (red). Greater vertical thickness at a given SHAP value indicates more samples with similar contributions. A red point far to the right represents a sample in

which a high driver value makes a large positive contribution to the predicted flux.”

===== Reviewer #2 =====

Overall assessment. The manuscript “Satellite-derived management indicators improve modeling of water and greenhouse gas fluxes in Swiss agroecosystems” estimates latent heat and greenhouse gas (GHG) fluxes using a machine learning (ML) framework that applies meteorological (METEO) and other remotely sensed vegetation (RS-VPI) and field management indices (RS-FMI) as predictor variables. The authors show that by including RS based predictors in their ML framework, they improve flux estimation, especially the GHG fluxes. Overall, the study shows the potential offered by high resolution RS at detecting poorly documented land management events/practices, which in turn contribute to GHG emissions. The paper is properly written and structured, and is well suited for this journal. However, the modeling framework was only applied over a few (six) Swiss (2) crop- and (4) grass-sites, with its spatial generalization potential remaining untested. Remarks:

Response: We thank you for evaluating our manuscript and the RS-derived predictors. We address your concerns regarding model transferability, DII evaluation, and presentation below.

General Comment 1. The limitations regarding transferability and generalizability of the ML models developed in this study (which appear to only be applicable over their pre-selected Swiss sites) need to be addressed/clarified. First, the authors present two frameworks for the different land covers in their study sites. Why not train one model (i.e., one for each flux) that could be applied for both cover types—or even one that could be applied on other land covers such as forests? Second, GHG gas emissions occur globally, and it has generally been understood that such emissions contribute to climate change at global scales. GHG modeling frameworks that would be transferable or applicable at larger scales should therefore be preferred in place of localized ones. The authors need to discuss these limitations and clarify whether (and/or how) they intend to evaluate a more generalized framework over diverse cover types and broader spatial scales.

Response: Thank you for asking us to clarify model transferability and the study scope. The primary goal of this study was to evaluate whether remote sensing-based field management indicators (RS-FMIs) and the remote sensing-based vegetation performance indicator (RS-VPI) capture management-related variability and improve daily flux modeling. This proof of concept provides a basis for future work on upscaling Swiss FluxNet observations to national-scale flux estimates for grasslands and croplands. We clarify this distinction in our response to Reviewer 1, General Comment 9.

1. Observational coverage and spatial transferability. The site coverage reflected the eddy covariance (EC) flux records accessible to us, particularly N₂O and CH₄ flux measurements accompanied by management records. Flux responses to management may differ among sites, particularly among croplands with different crop types. Given the limited number of sites, our evaluation focused on comparing predictor combinations within the observed site network rather than assessing transferability across crop types and management conditions. Leave-one-site-out (LOSO) validation was not performed, and predictive performance at unseen sites remains untested. We explain the observational constraints, geographical scope, and steps toward broader application in our response to Specific Comment 3.

2. Separate model fitting. We used a common XGBoost framework with separate fitting for grasslands and croplands to compare their relative driver contributions. The rationale and supporting evidence are provided in our response to Specific Comment 10.

3. Forests. Natural forest ecosystems generally experience less frequent direct management interventions than managed grasslands and croplands. Our study focuses on agroecosystems, where recurrent management provides the context for evaluating RS-FMIs. These indicators represent vegetation disturbance and exposed-soil dynamics associated with mowing, grazing, harvest, and cultivation. Applying the approach to forests would require indicators suited to forest management and disturbance regimes.

Please find detailed explanations and corresponding manuscript revisions in our responses to Specific Comments 3 and 10.

General Comment 2. This study proposes an interestingly simple yet visibly effective method to infer defoliation events from remotely sensed variables, specifically EVI. The authors show that the DII index is capable of timing defoliation events relatively well. However, the DII evaluations/sensitivity analyses performed are insufficient. E.g.: selection of window size(s) of surrounding background(s) appears somewhat arbitrary; why did the authors prefer EVI to some other variable like NDVI, which requires relatively less spectral information to estimate?; ...etc.

To make the DII index applicable at larger (even global) scales across diverse land management practices and land-cover types, further analyses are required.

Response: We thank you for these comments. We explain the choice of EVI here and address the evaluation questions under the corresponding specific comments.

1. Choice of EVI. EVI was developed to reduce canopy background effects, including soil reflectance influences, and to improve sensitivity to vegetation variation under high-biomass conditions where NDVI tends to saturate (Huete et al., 2002). These properties motivated its use for tracking vegetation changes associated with defoliation and

subsequent regrowth. The required blue, red, and near-infrared bands were available in our Sentinel-2 inputs.

We have clarified the rationale for selecting EVI at its first use in the Remotely-Sensed Defoliation Event Detection Section (Lines 177–178):

“EVI was selected because of its reduced sensitivity to soil background influences and greater sensitivity to canopy changes under dense vegetation conditions compared with NDVI (Huete et al., 2002).”

2. Further evaluation and spatial scope. We agree that further evaluation of DII is needed. We have extended the event-level evaluation by compiling all available harvest dry-matter records across the six study sites and pooling the available records by agroecosystem. Please find the responses concerning background-window selection and evaluation criteria under Specific Comment 8, and validation-site selection and differences in DII performance between agroecosystems under Specific Comment 12.

Please see our response to Specific Comment 3 for clarification of the study scope. Our focus is on evaluating the contribution of RS-FMIs and RS-VPI to GHG flux modeling in Swiss agricultural ecosystems. Assessing spatial transferability beyond Switzerland is a direction for future work. We also identify evaluation of RS-FMIs beyond Switzerland in the Mechanistic Interpretation and Scaling Implications Section (Lines 509–510):

“Future work could evaluate the performance of RS-FMIs in other regions and their contribution to GHG flux modeling.”

General Comment 3. The manuscript is sometimes too wordy. To improve flow and readability, rephrase/break long sentences, especially in the introduction and methods section.

L38-...: “Long-term eddy-covariance ... Merbold et al., 2014)” For example, is too long and can be broken into at least 2 parts to make it easier to follow.

Response: Thank you for this suggestion. We have reviewed the entire manuscript, split long and complex sentences, simplified wording, and clarified technical statements. These revisions include the Introduction Section (Lines 22–73), the Methods Section (Lines 77–299), and the Results and Discussion Sections (Lines 302–510). We also revised the Conclusions Section and appendices (Lines 512–521 and 531–637).

For the sentence you specifically identified, we separated the descriptions of mowing-related CO₂ flux changes and post-harvest N₂O fluxes and removed the comparison with meteorological forcing. The revised text in the Introduction Section (Lines 37–39) reads:

“Long-term eddy-covariance (EC) observations indicate that mowing-induced disturbances can drive CO₂ source–sink shifts through abrupt defoliation and reduced photosynthetic capacity (Feigenwinter et al., 2023b). In croplands, post-harvest bare soil periods are associated with strong but short-lived N₂O emission pulses (Maier et al., 2025).”

Specific comments:

Specific Comment 1. L5-6: “... we developed an RS- assisted modeling framework ...” – the authors should clearly indicate that the modeling performed here is based on applying machine learning modeling (i.e. statistical learning algorithms) and not physically based process modeling.

Response: We have explicitly identified the framework as a machine learning framework at its first mention in the Abstract (Lines 5–8):

“Here, we developed an RS-assisted machine learning framework to estimate daily water and GHG fluxes across six Swiss FluxNet sites (two croplands and four grasslands) between 2016 and 2025.”

Specific Comment 2. L13-15: “... LE ($R^2 \ll 0.89-0.90$)” – for completeness, maybe also include the respective rmse or bias values here.

Response: We have added the corresponding RMSE values alongside R^2 in the Abstract (Lines 12–15):

“The models achieved R^2 values of 0.89–0.90 for LE fluxes (RMSE = 13.26–14.36 W m⁻²) and 0.59–0.71 for NEE (RMSE = 1.68–1.76 μmol CO₂ m⁻² s⁻¹). Performance was lower for N₂O fluxes (R^2 = 0.54–0.55, RMSE = 0.72–0.73 nmol N₂O m⁻² s⁻¹) and CH₄ fluxes (R^2 = 0.37–0.50, RMSE = 2.68–14.56 nmol CH₄ m⁻² s⁻¹).”

Specific Comment 3. L72-74: “While spatial upscaling to unobserved regions remains outside the scope of this work, this ...” – again, transferability/generalization of your ML models should be discussed a bit more than this. Could a similarly trained RS-driven ML model be easily applied in croplands and/or grasslands in other regions with similar climate or even globally?

Response: Thank you for this comment. We have expanded the discussion of spatial transferability and clarified the study’s objectives, observational basis, and implications for applications in other regions.

1. Study objective and scope. As clarified in our response to Reviewer 1, General Comment 9, this study provides a proof of concept for incorporating RS-FMIs and RS-VPI into agroecosystem flux modeling. Its primary objective is to evaluate whether remotely sensed management and vegetation information improves daily flux prediction and how their contributions differ among fluxes and agroecosystem types. Spatial upscaling is a direction for future work.

2. Availability of agricultural flux observations. Long-term eddy covariance (EC) measurements of N₂O and CH₄ fluxes remain scarce in agricultural ecosystems (Hörtnagl et al., 2018). These measurements require specialized analyzers and substantial investment in instrumentation, calibration, and maintenance (Nemitz et al., 2018). Public access to agricultural flux observations and their corresponding management records is still limited. These constraints make it difficult to assemble geographically extensive datasets that capture both flux variability and the management activities associated with it.

Swiss FluxNet combines long-term ecosystem flux measurements across grasslands of different management intensities with targeted management experiments (Buchmann et al., 2019). These observations allow us to examine management-related flux variability and evaluate RS-FMIs in the Swiss agricultural systems studied here. The availability of corresponding management records also allows the remotely sensed indicators to be evaluated against recorded field operations and compared with models using those records.

We assembled the Swiss agroecosystem EC flux records accessible to us, with particular emphasis on N₂O and CH₄ flux measurements. The resulting dataset comprised six sites for LE and NEE, of which four also provided N₂O and CH₄ flux observations. We welcome your suggestions for additional agricultural N₂O and CH₄ flux datasets with accompanying management records. We would consider using them in future model development and evaluation.

We have clarified this observational coverage in the Swiss FluxNet Section (Lines 81–83):

“We assembled the available Swiss agroecosystem EC flux records accessible to us, with particular emphasis on N₂O and CH₄ flux measurements. The resulting dataset comprised six sites for LE and NEE, of which four also provided N₂O and CH₄ flux observations (Table 1).”

3. Application in other regions. The model comparison approach can be applied in other regions with flux observations and corresponding meteorological and satellite data. Applying the trained models elsewhere requires further evaluation, because similar climates may still differ in soil conditions, crop composition, and management practices.

The need for broader training-data coverage has also been recognized in field-scale ML studies of agricultural N₂O fluxes. Hamrani et al. (2020) compared models using five growing seasons of observations from a single agricultural site in Canada. Their evaluation demonstrated predictive performance within the observed conditions, while leaving

applicability to other sites unresolved. Dhaliwal et al. (2025) analyzed 17 years of observations from different cropping systems at a long-term experimental site. Despite this temporal and management coverage, they explicitly identified the need for more diverse soil, climate, crop, and management data before applying their models in other regions. Similarly, Gnisia et al. (2025) evaluated models across seven management treatments and identified training with observations from multiple locations and land uses as a requirement for larger-scale application.

These studies show that long time series and diverse management treatments alone do not establish spatial transferability.

We have clarified this in the Discussion Section 4.1 (Lines 417–421):

“This study provided a proof of concept for using RS-FMIs and RS-VPI in daily agroecosystem flux modeling. The model comparison approach can be applied in other regions with flux observations and corresponding meteorological and satellite data. Future work could evaluate RS-FMIs and their contribution to flux modeling in other regions. Leave-one-site-out (LOSO) validation could assess the spatial transferability of trained models, accounting for differences in management conditions and emission events among sites (Dhaliwal et al., 2025; Gnisia et al., 2025).”

Specific Comment 4. L95: “No surface energy balance closure correction was applied to LE flux.” - Why? Is the energy imbalance at all 6 sites too little to warrant any corrections?

Response: Thank you for asking this question. Ground heat flux was not measured at some sites, preventing consistent energy balance closure correction across the dataset. Our analysis focused on daily fluxes, for which aggregation can reduce energy imbalance associated with diurnal heat storage and release (Grachev et al., 2020). We therefore retained uncorrected LE observations for daily flux modeling, following Nelson et al. (2024).

We have clarified these reasons in the Swiss FluxNet Section (Lines 95–98):

“No surface energy balance closure correction was applied to LE flux because ground heat flux was not measured at some sites. We focused on daily fluxes, for which aggregation can reduce energy imbalance associated with diurnal heat storage and release (Grachev et al., 2020). We therefore retained uncorrected LE observations for daily flux modeling, following Nelson et al. (2024).”

Specific Comment 5. L116-119: “Due to the limited availability of gridded soil moisture and soil temperature data, these variables were not included in the analysis. Precipitation-based temporal features partially capture moisture variability.” - what about soil temperature? Satellite surface temperature (an important variable for inferring surface turbulent flux exchanges) can be used as a proxy for soil temperature..

Response: Thank you for this suggestion. We agree that satellite-derived surface temperature could provide useful information on thermal conditions relevant to ecosystem fluxes. Our decision to exclude it reflected the trade-off between spatial resolution and temporal coverage.

Common daily LST products, such as MODIS, have a spatial resolution of approximately 1 km, which may mix fields with different vegetation and management conditions. Landsat provides finer spatial resolution, but its less frequent acquisitions and cloud screening limit the availability of valid observations. Restricting model development and evaluation to days with valid Landsat surface temperature retrievals would have substantially reduced the available flux samples. Given the limited training data, we therefore did not include satellite-derived surface temperature in this study.

Specific Comment 6. Section 2.1.2: Restructure section for coherence. Too wordy.

Response: We have organized the Sentinel-2 Satellite Products Section under four italic subheadings: Sentinel-2 surface reflectance, Fractional vegetation cover, Leaf area index, and Crop type. We have also clarified product provenance and simplified lengthy sentences (Lines 123–169). These revisions also address Reviewer 1’s General Comments 4 and 5.

Specific Comment 7. L203: “... is the corresponding spatial standard deviation.” – ‘corresponding’ is not very clear here. spatial standard deviation of ? add ...of EVI.

Response: We have added “of EVI” to clarify the definition of the spatial standard deviation (Remotely-Sensed Defoliation Intensity Estimation Section, Line 203).

“...and $\sigma_{\text{win}}(d)$ is the corresponding spatial standard deviation of EVI.”

Specific Comment 8. L213-...: “Sensitivity analysis indicated that a 1 km window best captured management-related disturbance signals in grasslands, whereas a 2 km window was more appropriate for croplands (Appendix B) “ – You come to this conclusion after basing your analysis on R^2 , which is maybe not the best measure. Did you test with multiple objective functions, e.g. RMSE, bias ...?

Response: Thank you for this suggestion. We extended the sensitivity analysis using the revised Kling–Gupta efficiency (KGE^* ; Tang et al., 2021), which jointly evaluates correlation, variability, and mean bias. Although RMSE and bias allow comparisons within each flux, their differing units and magnitudes limit direct comparisons across fluxes. We therefore used KGE^* as a dimensionless criterion that explicitly incorporates standardized mean bias.

For grasslands, N_2O achieved its highest KGE^* at 800 m, with little variation between 800 m and 1 km, whereas CH_4 peaked at 1 km. We therefore retained a common 1 km window for grasslands. For croplands, both fluxes achieved their highest KGE^* at 2 km, supporting the original 2 km window selection. LE and NEE showed little sensitivity to window diameter. We have added the metric definition and updated the results in Appendix B (Lines 549–580). The revised Figure B1 is reproduced below.

“To evaluate the sensitivity of DII to spatial window size, we varied the window diameter from 200 to 5000 m, using 200 m increments up to 1000 m and 500 m increments thereafter (Figure B1). For each diameter, DII was recalculated and the METEO + RS-VPI + RS-FMIs model was retrained for each flux and agroecosystem type. All other predictors, hyperparameters, training and validation samples, and data partitions were held constant across diameters.

Predictive performance was evaluated using the revised Kling–Gupta efficiency (KGE^* ; Tang et al., 2021):

$$KGE^* = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta^*)^2}$$

where r is the Pearson correlation between predicted and observed fluxes, $\alpha = \sigma_{pred}/\sigma_{obs}$ is the variability ratio, and $\beta^* = (\mu_{pred} - \mu_{obs})/\sigma_{obs}$ is the standardized mean bias. Here, μ and σ denote the mean and standard deviation, respectively. KGE^* combines these three dimensionless components, with a maximum value of 1 indicating perfect agreement. Standardizing mean bias by the observed standard deviation avoids division by an observed mean that may approach zero, allowing the same formulation to be applied to LE, NEE, N_2O , and CH_4 fluxes.”

“Sensitivity to spatial window size differed among fluxes and ecosystem types. LE and NEE showed limited sensitivity to the selected window size, whereas N_2O flux in croplands showed the strongest response. For LE, model performance remained nearly constant across all tested diameters, reflecting

the dominant role of meteorological forcing in controlling latent heat flux. CH₄ flux showed smaller variations across window diameters.

For N₂O flux in croplands, KGE* increased with window diameter up to 2 km, where it reached 0.54, and then gradually declined. Grassland N₂O flux showed a less pronounced, non-monotonic response, with the highest KGE* of 0.35 at 800 m and little variation between 800 m and 1 km. CH₄ flux reached its highest KGE* at 1 km in grasslands and 2 km in croplands, with values of 0.40 and 0.60, respectively. Given the limited sensitivity of grassland N₂O performance near its optimum and the CH₄ optimum at 1 km, a common 1 km window was retained for grasslands. A 2 km window was retained for croplands, where both fluxes reached their highest KGE*."

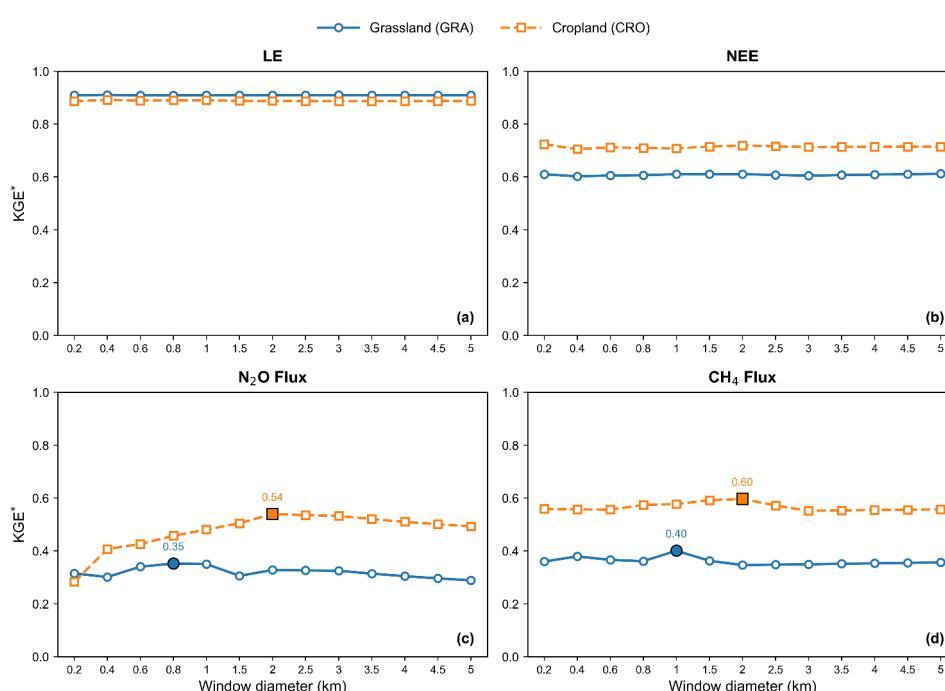


Figure B1. Sensitivity of model performance, evaluated using KGE*, to the spatial window diameter used to calculate DII for grassland (GRA) and cropland (CRO) systems. Panels show (a) LE, (b) NEE, (c) N₂O flux, and (d) CH₄ flux. Filled markers indicate the highest KGE* for N₂O and CH₄ fluxes within each agroecosystem type.

Specific Comment 9. L226-228: "A field was classified as bare when the bare soil fraction exceeded 0.5" – 0.5 bare soil fraction would mean 50% vegetation cover fraction ...is 0.5 fraction of bare soil not too low to consider the surface totally bare?

Response: The threshold was intended to identify periods dominated by exposed soil, rather than completely vegetation-free surfaces. A bare soil fraction greater than 0.5 indicates that exposed soil covers more than half of the surface. The remaining fraction includes both photosynthetic and non-photosynthetic vegetation.

Rapid vegetation recovery can occur in grasslands during the growing season, whereas winter crops may retain low vegetation cover for extended periods. We selected 0.5 as a consistent and interpretable criterion for identifying periods dominated by exposed soil.

We have revised “classified as bare” to “classified as bare-soil dominated” and replaced “phases of canopy absence” with “phases dominated by exposed soil” to clarify this interpretation (Remotely-Sensed Bare Soil Period Section, Lines 222–228).

“A field was classified as bare-soil dominated when the bare soil fraction exceeded 0.5.”

Specific Comment 10. L249-...: “Grassland and cropland systems were modeled separately to enable a comparison of their sensitivities to environmental and management drivers within a unified modeling framework.” - the authors need to better clarify why separate ML frameworks (as per AppendixC: Table C1) were needed for the 2 cover types. Why not have a generalized ML model (per target flux) that could be applied over both land cover types?

Response: Thank you for sharing this concern. We fitted grassland and cropland models separately within a common XGBoost framework to compare relative driver contributions (Figure 8).

The two systems did not share all management predictors. Crop rotation, for example, applied only to croplands. Shared indicators, such as DII, could also have different meanings in grasslands and croplands. Separate fitting allowed the models to learn how these indicators related to fluxes within each agroecosystem type.

Comparable designs include vegetation-type-specific XGBoost–SHAP models for summer net ecosystem productivity (Ren et al., 2024) and separate RF–SHAP analyses of contrasting coastal ecosystems (Ren et al., 2026). Voigt et al. (2023) likewise fitted separate random forest models for three vegetation types to compare the importance of environmental predictors of CH₄ fluxes.

We have clarified this rationale (Modeling Configuration Section, Lines 263–266):

“Grasslands and croplands were modeled separately using the same XGBoost framework and scenario design to compare relative driver contributions between agroecosystem types. The two systems did not share all management predictors, as crop rotation applied only to croplands. Shared indicators, such as DII, also differed in interpretation between grasslands and croplands. Separate fitting allowed the models to learn how these indicators related to fluxes within each agroecosystem type.”

Specific Comment 11. L271: “of individual variables” – since you also call LE and GHG fluxes “target” variables’, you need to clarify what these ‘individual variables’ are – ‘...individual predictor variables (i.e. METEO, RS-VPI, RS-FMIs)’?

Response: We have replaced “individual variables” with “individual predictor variables from the METEO, RS-VPI, and RS-FMIs groups” to distinguish the predictors from the target flux variables (Driver Analysis Section, Lines 282–283):

“We used SHAP values from the optimized XGBoost models to quantify the contributions of individual predictor variables from the METEO, RS-VPI, and RS-FMIs groups to flux variability (Lundberg and Lee, 2017).”

Specific Comment 12. L295-...: “... The remotely sensed DII was evaluated against recorded management events at representative long-term grassland and crop-land sites before being used as an RS-FMI in the flux modeling framework. At the CH-Cha grassland ...” - what was the consideration here for selecting the two sites as the representative sites. Indeed, DII peaks appear to generally coincide well with the real defoliation events over the grassland but not so well over the selected crop-site. Is this the case over the other 4 sites as well?

Response: We thank you for these comments. CH-Cha and CH-Oe2 were selected for the time-series analysis (Figure 3) because they had the longest and most complete records and represented grassland and cropland agroecosystems, respectively.

For the supplementary event-level evaluation (Appendix A2, Figure A2), we compiled all available harvest dry-matter records across the six study sites and pooled the available records by agroecosystem. The updated comparison showed a positive association in grasslands ($r = 0.49$, $p < 0.001$), whereas the positive relationship in croplands was not statistically significant ($r = 0.36$, $p = 0.147$). We placed the DII–harvested DM scatterplot in Appendix A2 because it provides a supporting evaluation of the index, while the main text focuses on the flux-modeling results. We refer to the comparison in the Results and Discussion. We have clarified the site selection in Results Section 3.1 (Lines 303–305):

“CH-Cha and CH-Oe2 were selected for the time-series analysis because they had the longest and most complete records and represented grassland and cropland agroecosystems, respectively.”

Regarding differences among land-cover types, DII represents a standardized vegetation deviation from a local same-land-cover background. Its relationship with biomass removal depends on the vegetation and management context. Its relationship with harvested biomass was more consistent in grasslands, whereas crop rotation, crop identity, harvested components, and post-harvest surface conditions affected its interpretation in croplands, as described in Appendix A2 and the DII discussion.

The same DII value does not necessarily represent the same disturbance state in grasslands and croplands. We therefore fitted the DII–flux relationships separately, although crop-specific variability still needs to be considered. We have made this distinction explicit (Defoliation Intensity Index as a Proxy for Management Intensity Section, Lines 443–445):

“The vegetation disturbance represented by a given DII value can therefore differ between grasslands and croplands. Separate model fitting allowed its relationship with fluxes to be learned within each agroecosystem type, although crop-specific variability within croplands remained.”

Specific Comment 13. L329..., Figure5, Figure6: Can you also provide the evaluation performance metrics per site separately, in addition to the combined analyses (maybe in the appendix or supplements).

Response: Thank you for suggesting a site-specific evaluation. We added a new Appendix D with four tables presenting site-specific R^2 , RMSE and bias for the four fluxes under each modeling scenario. The tables group grassland and cropland sites separately and supplement the combined analyses in Figures 4 and 5. The interpolation appendix is now Appendix E. We also added a brief note in Appendix D to acknowledge site-specific differences in agreement with the management benchmark (Table D3). The CH-For results should be interpreted cautiously given the relatively short observational record. We added the following sentence to the Model Evaluation Section (Line 322):

“Site-specific evaluation metrics are provided in Appendix D.”

Table D1. Site-specific model evaluation for LE fluxes of the different scenarios using meteorological variables (METEO), management records (MGMT), remotely-sensed vegetation performance index (RS-VPI), and remotely-sensed field management indices (RS-FMIs). RMSE and bias are given in $W\ m^{-2}$.

Site	Scenario	R ²	RMSE	Bias
<i>Grassland</i>				
CH-Cha	METEO	0.90	11.31	4.00
	METEO + RS-VPI	0.91	10.70	3.38
	METEO + RS-VPI + RS-FMIs	0.92	10.55	3.11
	METEO + MGMT	0.91	11.07	3.89
CH-Fru	METEO	0.89	13.04	1.09
	METEO + RS-VPI	0.89	13.00	1.05
	METEO + RS-VPI + RS-FMIs	0.89	12.92	0.49
	METEO + MGMT	0.90	12.66	0.20
CH-Aws	METEO	0.90	14.57	-2.57
	METEO + RS-VPI	0.91	14.23	-2.18
	METEO + RS-VPI + RS-FMIs	0.92	13.04	-1.28
	METEO + MGMT	0.91	13.82	-1.64
CH-For	METEO	0.66	20.50	-2.52
	METEO + RS-VPI	0.67	20.20	-2.57
	METEO + RS-VPI + RS-FMIs	0.69	19.40	-2.37
	METEO + MGMT	0.66	20.27	-2.54
<i>Cropland</i>				
CH-Oe2	METEO	0.86	15.73	0.07
	METEO + RS-VPI	0.90	13.48	-0.55
	METEO + RS-VPI + RS-FMIs	0.90	13.26	-0.38
	METEO + MGMT	0.88	14.92	0.23
CH-Tnk	METEO	0.75	18.99	-1.35
	METEO + RS-VPI	0.87	13.69	-2.46
	METEO + RS-VPI + RS-FMIs	0.88	13.34	-1.06
	METEO + MGMT	0.78	17.90	-3.03

Table D2. Site-specific model evaluation for NEE fluxes. RMSE and bias are given in $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$.

Site	Scenario	R ²	RMSE	Bias
<i>Grassland</i>				
CH-Cha	METEO	0.42	1.93	0.22
	METEO + RS-VPI	0.57	1.66	0.14
	METEO + RS-VPI + RS-FMIs	0.63	1.55	0.12
	METEO + MGMT	0.70	1.38	0.15
CH-Fru	METEO	0.42	2.04	-0.06
	METEO + RS-VPI	0.48	1.93	0.02
	METEO + RS-VPI + RS-FMIs	0.50	1.88	0.08
	METEO + MGMT	0.49	1.91	0.06
CH-Aws	METEO	0.55	1.69	-0.15
	METEO + RS-VPI	0.63	1.53	-0.14
	METEO + RS-VPI + RS-FMIs	0.69	1.41	-0.19
	METEO + MGMT	0.57	1.65	-0.21
CH-For	METEO	0.36	2.08	-0.06
	METEO + RS-VPI	0.46	1.92	-0.02
	METEO + RS-VPI + RS-FMIs	0.50	1.84	-0.02
	METEO + MGMT	0.53	1.78	-0.03
<i>Cropland</i>				
CH-Oe2	METEO	0.48	2.36	0.19
	METEO + RS-VPI	0.68	1.86	0.20
	METEO + RS-VPI + RS-FMIs	0.70	1.78	0.14
	METEO + MGMT	0.60	2.08	0.09
CH-Tnk	METEO	0.32	2.59	1.41
	METEO + RS-VPI	0.74	1.60	0.34
	METEO + RS-VPI + RS-FMIs	0.73	1.63	0.25
	METEO + MGMT	0.50	2.22	1.20

Table D3. Site-specific model evaluation for N₂O fluxes. RMSE and bias are given in nmol N₂O m⁻² s⁻¹.

Site	Scenario	R ²	RMSE	Bias
<i>Grassland</i>				
CH-Cha	METEO	0.41	0.82	0.13
	METEO + RS-VPI	0.46	0.79	0.11
	METEO + RS-VPI + RS-FMIs	0.55	0.72	0.12
	METEO + MGMT	0.56	0.71	0.09
CH-For	METEO	<0.1	0.84	0.29
	METEO + RS-VPI	<0.1	0.82	0.17
	METEO + RS-VPI + RS-FMIs	<0.1	0.91	0.21
	METEO + MGMT	0.48	0.59	0.48
<i>Cropland</i>				
CH-Oe2	METEO	<0.1	1.52	-0.19
	METEO + RS-VPI	0.10	1.45	-0.20
	METEO + RS-VPI + RS-FMIs	0.55	1.02	-0.02
	METEO + MGMT	0.59	0.98	-0.17
CH-Tnk	METEO	<0.1	0.73	0.34
	METEO + RS-VPI	<0.1	0.67	0.32
	METEO + RS-VPI + RS-FMIs	0.14	0.39	0.12
	METEO + MGMT	<0.1	0.42	0.23

Table D4. Site-specific model evaluation for CH₄ fluxes. RMSE and bias are given in nmol CH₄ m⁻² s⁻¹.

Site	Scenario	R ²	RMSE	Bias
<i>Grassland</i>				
CH-Cha	METEO	0.20	18.34	1.71
	METEO + RS-VPI	0.35	16.47	1.95
	METEO + RS-VPI + RS-FMIs	0.50	14.52	1.26
	METEO + MGMT	0.52	14.14	0.59
CH-For	METEO	0.16	18.36	2.06
	METEO + RS-VPI	0.31	16.66	2.40
	METEO + RS-VPI + RS-FMIs	0.46	14.65	1.65
	METEO + MGMT	0.49	14.34	1.01
<i>Cropland</i>				
CH-Oe2	METEO	<0.1	3.16	1.20
	METEO + RS-VPI	0.12	3.05	0.92
	METEO + RS-VPI + RS-FMIs	0.27	2.79	0.40
	METEO + MGMT	0.13	3.03	1.05
CH-Tnk	METEO	0.10	2.73	0.00
	METEO + RS-VPI	0.18	2.61	0.26
	METEO + RS-VPI + RS-FMIs	0.18	2.61	0.54
	METEO + MGMT	0.18	2.61	0.03

Specific Comment 14. L331: “the NEE range, with underestimation of both strong negative and high positive fluxes” - from the scatterplots, strong negative NEE fluxes appear to be higher than the observations so not sure how that is an ‘underestimation’

Response: For strongly negative NEE, the predictions were less negative than the observations and therefore numerically higher. The sign indicates the direction of net CO₂ exchange, whereas the absolute value indicates its magnitude. Our original wording referred to the underestimated magnitude of these fluxes. We have revised the sentence to state the direction of the prediction errors explicitly (Model Evaluation Section, Lines 331–333):

“Scatterplots further showed compressed variability at both ends of the NEE range, with predictions that were less negative than observations for strong negative fluxes and lower than observations for high positive fluxes.”

Specific Comment 15. L381-383: "... The grouped SHAP results showed increased contributions from RS-derived vegetation and management indicators in croplands compared to grasslands (Figure 9), indicating a stronger role of vegetation structural dynamics." – the authors should clarify that this statement generally only applies to the LE and NEE fluxes (see figure 9)

Response: We have explicitly restricted this statement to LE and NEE (Driving Factor Analysis Section, Lines 381–383):

"For LE and NEE, the grouped SHAP results showed increased contributions from RS-derived vegetation and management indicators in croplands compared to grasslands (Figure 8), indicating a stronger role of vegetation structural dynamics."

Specific Comment 16. L407: "... A LOSO " - LOSO acronym needs to be described prior to use.

Response: We have expanded the acronym to "leave-one-site-out (LOSO)" and revised the sentence to describe a direction for future evaluation (Discussion Section 4.1, Lines 419–421):

"Future work could evaluate RS-FMIs and their contribution to flux modeling in other regions. Leave-one-site-out (LOSO) validation could assess the spatial transferability of trained models, accounting for differences in management conditions and emission events among sites (Dhaliwal et al., 2025; Gnisia et al., 2025)."

Specific Comment 17. Appendix D: ...the analyses here (Figure D1-noninterpolated vs Figure7-interpolated) only consider one [grass] site (CH-Cha) in 2021. Can the same conclusions be reached when all sites and/or other periods are considered?

Response: Thank you for suggesting that we extend this evaluation to other sites and periods. We added a pooled scatterplot covering the available sites and test years in both grasslands and croplands (Appendix E). The new figure shows fewer samples and fewer high-emission observations than the daily analyses in Figures 4 and 5, with narrower sampled ranges and lower R^2 for N_2O fluxes and CH_4 fluxes. We retained the CH-Cha time series (Figure E2) to illustrate the temporal gaps and event-level behavior. The new scatterplot precedes that example in Appendix E (Lines 619–637; Figure E1, p. 35). We have also clarified in Discussion Section 4.1 that the lower predictive performance in the pooled non-interpolated analysis reflects differences in both temporal coverage and sample composition.

The following paragraph was added to Appendix E (Lines 619–624):

“The evaluation was based on pooled samples across the available sites and test years in both agroecosystem types (Figure E1). Compared with the daily analyses in Figures 4 and 5, the non-interpolated analysis contained fewer samples and represented fewer high-emission observations, narrowing the sampled ranges of N₂O and CH₄ fluxes. For the METEO + RS-VPI + RS-FMIs scenario, R² for N₂O fluxes was 0.35 in grasslands and <0.1 in croplands, compared with 0.54 and 0.55 in the daily analyses. For CH₄ fluxes, the corresponding values were <0.1 and 0.11, compared with 0.50 and 0.37. The METEO + MGMT scenario also showed lower R² for both fluxes in the non-interpolated analysis.”

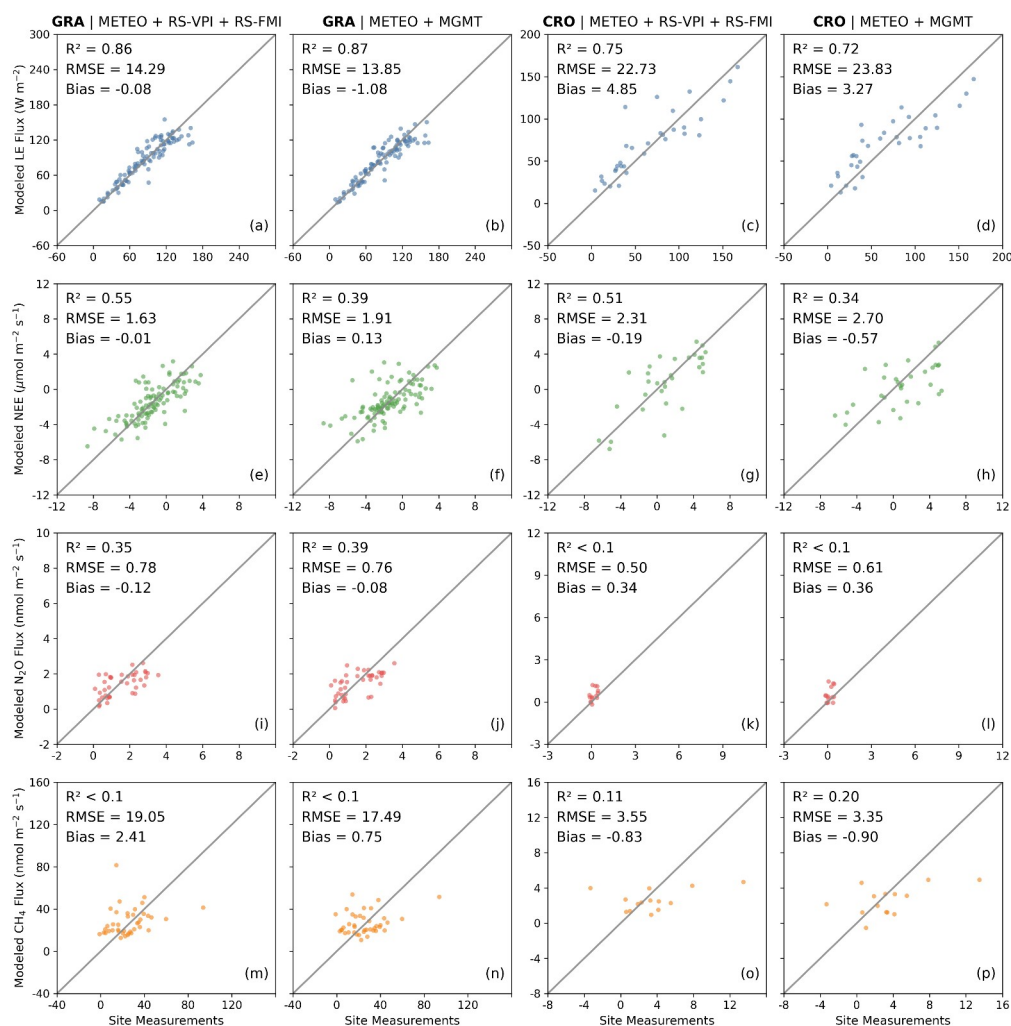


Figure E1. Modeled versus observed fluxes in the non-interpolated analysis, pooled across available sites and test years. The first two columns show grassland results and the last two show cropland results, each for the METEO + RS-VPI + RS-FMIs and METEO + MGMT scenarios. Rows show LE (a–d), NEE (e–h), N₂O fluxes (i–l), and CH₄ fluxes (m–p). The grey line denotes the 1:1 relationship. R², RMSE and bias are shown in each panel. The non-interpolated samples differ from those in the daily analyses in Figures 4 and 5.

References

- Andreatta, D., Gianelle, D., Scotton, M., Vescovo, L., and Dalponte, M.: Detection of grassland mowing frequency using time series of vegetation indices from Sentinel-2 imagery, *GIScience & Remote Sensing*, 59, 481–500, 2022.
- Buchmann, N., Fuchs, K., Feigenwinter, I., and Gilgen, A. K.: Multifunctionality of permanent grasslands: ecosystem services and resilience to climate change, *Grassland Science in Europe*, 24, 19–26, <https://doi.org/10.3929/ethz-b-000472190>, 2019.
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, *Remote Sensing of Environment*, 219, 145–161, <https://doi.org/10.1016/j.rse.2018.09.002>, 2018.
- De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., and Defourny, P.: Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring, *Remote Sensing of Environment*, 280, 113145, 2022.
- Dhaliwal, J. K., Panday, D., Robertson, G. P., and Saha, D.: Machine learning reveals dynamic controls of soil nitrous oxide emissions from diverse long-term cropping systems, *Journal of Environmental Quality*, 54, 132–146, <https://doi.org/10.1002/jeq2.20637>, 2025.
- Feigenwinter, I., Hörtnagl, L., and Buchmann, N.: N₂O and CH₄ fluxes from intensively managed grassland: The importance of biological and environmental drivers vs. management, *Science of the Total Environment*, 903, 166389, <https://doi.org/10.1016/j.scitotenv.2023.166389>, 2023a.
- Feigenwinter, I., Hörtnagl, L., Zeeman, M. J., Eugster, W., Fuchs, K., Merbold, L., and Buchmann, N.: Large inter-annual variation in carbon sink strength of a permanent grassland over 16 years: Impacts of management practices and climate, *Agricultural and Forest Meteorology*, 340, 109613, <https://doi.org/10.1016/j.agrformet.2023.109613>, 2023b.
- Geng, X., Li, B., Pan, X., Wei, Z., Yang, X., Liao, K., and Cui, Y.: Generating hourly continuous evapotranspiration using Fengyun-2F geostationary satellite data, *Earth and Space Science*, 9, e2021EA002090, <https://doi.org/10.1029/2021EA002090>, 2022.
- Gnisia, G., Weik, J., Ruser, R., Essich, L., Lewandowski, I., and Stein, A.: Machine learning-based prediction of nitrous oxide emissions from arable farming: Exploring management practices as predictor variables, *Ecological Indicators*, 172, 113233, <https://doi.org/10.1016/j.ecolind.2025.113233>, 2025.
- Grachev, A. A., Fairall, C. W., Blomquist, B. W., Fernando, H. J. S., Leo, L. S., Otárola-Bustos, S. F., Wilczak, J. M., and McCaffrey, K. L.: On the surface energy balance closure at different temporal scales, *Agricultural and Forest Meteorology*, 281, 107823, <https://doi.org/10.1016/j.agrformet.2019.107823>, 2020.

- Hamrani, A., Akbarzadeh, A., and Madramootoo, C. A.: Machine learning for predicting greenhouse gas emissions from agricultural soils, *Science of the Total Environment*, 741, 140338, <https://doi.org/10.1016/j.scitotenv.2020.140338>, 2020.
- Hörtnagl, L. and Wohlfahrt, G.: Methane and nitrous oxide exchange over a managed hay meadow, *Biogeosciences*, 11, 7219–7236, <https://doi.org/10.5194/bg-11-7219-2014>, 2014.
- Hörtnagl, L., Barthel, M., Buchmann, N., Eugster, W., Butterbach-Bahl, K., Díaz-Pinés, E., Zeeman, M., Klumpp, K., Kiese, R., Bahn, M., Hammerle, A., Lu, H., Ladreiter-Knauss, T., Burri, S., and Merbold, L.: Greenhouse gas fluxes over managed grasslands in Central Europe, *Global Change Biology*, 24, 1843–1872, <https://doi.org/10.1111/gcb.14079>, 2018.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sensing of Environment*, 83, 195–213, [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2), 2002.
- Lauber, T., Turkoglu, M. O., Ledain, S., and Aasen, H.: SwissCrop25: A national multi-year benchmark for operational crop mapping, *arXiv preprint arXiv:2608.09497*, <https://doi.org/10.48550/arXiv.2608.09497>, 2026.
- Ledain, S., Gilgen, A., and Aasen, H.: Large-scale, high-resolution fractional cover mapping from Sentinel-2 for agri-environmental monitoring, *EGU General Assembly 2026*, EGU26-17433, <https://doi.org/10.5194/egusphere-egu26-17433>, 2026a.
- Ledain, S., Gilgen, A., and Aasen, H.: Soil-informed PROSAIL modelling improves scalable retrieval of leaf area index: evidence from multi-year, multi-country winter wheat observations, *SSRN preprint*, <https://doi.org/10.2139/ssrn.6626101>, 2026b.
- Liang, S., He, T., Huang, J., Jia, A., Zhang, Y., Cao, Y., Chen, X., Chen, X., Cheng, J., and Jiang, B.: Advances in high-resolution land surface satellite products: a comprehensive review of inversion algorithms, products and challenges, *Science of Remote Sensing*, p. 100152, 2024.
- Lobert, F., Schwieder, M., Alsleben, J., Broeg, T., Kowalski, K., Okujeni, A., Hostert, P., and Erasmí, S.: Unveiling year-round cropland cover by soil-specific spectral unmixing of Landsat and Sentinel-2 time series, *Remote Sensing of Environment*, 318, 114594, <https://doi.org/10.1016/j.rse.2024.114594>, 2025.
- Lundberg, S. M., and Lee, S.-I.: A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, 30, 2017.
- Maier, R., Hörtnagl, L., and Buchmann, N.: Large nitrous oxide emissions from arable soils after crop harvests prior to sowing, *Nutrient Cycling in Agroecosystems*, 130, 161–175, <https://doi.org/10.1007/s10705-024-10395-0>, 2025.
- Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., et al.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, *FLUXCOM-X*, *Biogeosciences*, 21, 5079–5115, <https://doi.org/10.5194/bg-21-5079-2024>, 2024.

- Nemitz, E., Mammarella, I., Ibrom, A., Aurela, M., Burba, G. G., Dengel, S., Gielen, B., Grelle, A., Heinesch, B., Herbst, M., Hörtnagl, L., Klemedtsson, L., Lindroth, A., Lohila, A., McDermitt, D. K., Meier, P., Merbold, L., Nelson, D., Nicolini, G., Nilsson, M. B., Peltola, O., Rinne, J., and Zahniser, M.: Standardisation of eddy-covariance flux measurements of methane and nitrous oxide, *International Agrophysics*, 32, 517–549, <https://doi.org/10.1515/intag-2017-0042>, 2018.
- Rains, D., Trigo, I., Dutra, E., Ermida, S., Ghent, D., Hulsman, P., Gómez-Dans, J., and Miralles, D. G.: High-resolution (1 km) all-sky net radiation over Europe enabled by the merging of land surface temperature retrievals from geostationary and polar-orbiting satellites, *Earth System Science Data*, 16, 567–593, <https://doi.org/10.5194/essd-16-567-2024>, 2024.
- Reinermann, S., Asam, S., and Kuenzer, C.: Remote sensing of grassland production and management—A review, *Remote Sensing*, 12, 1949, 2020.
- Ren, L., Yang, H., Yang, H., Gu, R., Zhu, Z., Zhao, N., Wang, Y. P., and Hou, L.: Modulating mechanisms of marine hydrodynamics on global coastal carbon burial, *Nature Communications*, 17, 8369, <https://doi.org/10.1038/s41467-026-75374-1>, 2026.
- Ren, Y., Qiu, J., Zeng, Z., Liu, X., Sitch, S., Pilegaard, K., Yang, T., Wang, S., Yuan, W., and Jain, A. K.: Earlier spring greening in Northern Hemisphere terrestrial biomes enhanced net ecosystem productivity in summer, *Communications Earth & Environment*, 5, 122, <https://doi.org/10.1038/s43247-024-01270-5>, 2024.
- Sainte Fare Garnot, V. and Landrieu, L.: Panoptic segmentation of satellite image time series with convolutional temporal attention networks, *arXiv preprint arXiv:2107.07933*, <https://arxiv.org/abs/2107.07933>, 2022.
- Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., and Hostert, P.: Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series, *Remote Sensing of Environment*, 269, 112795, 2022.
- Tang, G., Clark, M. P., and Papalexiou, S. M.: SC-Earth: A station-based serially complete Earth dataset from 1950 to 2019, *Journal of Climate*, 34, 6493–6511, <https://doi.org/10.1175/JCLI-D-21-0067.1>, 2021.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences*, 13, 4291–4313, <https://doi.org/10.5194/bg-13-4291-2016>, 2016.
- Turkoglu, M. O., D’Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., and Wegner, J. D.: Crop mapping from image time series: Deep learning with multi-scale label hierarchies, *Remote Sensing of Environment*, 264, 112603, 2021.
- Turkoglu, M. O., Ledain, S., and Aasen, H.: Model-agnostic, temperature-informed sampling enhances cross-year crop mapping with deep learning, *arXiv preprint arXiv:2506.12885*, <https://arxiv.org/abs/2506.12885>, 2025.

- Voigt, C., Virkkala, A.-M., Hould Gosselin, G., Bennett, K. A., Black, T. A., Detto, M., Chevrier-Dion, C., Guggenberger, G., Hashmi, W., Kohl, L., Kou, D., Marquis, C., Marsh, P., Marushchak, M. E., Nesic, Z., Nykänen, H., Saarela, T., Sauheitl, L., Walker, B., Weiss, N., Wilcox, E. J., and Sonnentag, O.: Arctic soil methane sink increases with drier conditions and higher ecosystem respiration, *Nature Climate Change*, 13, 1095–1104, <https://doi.org/10.1038/s41558-023-01785-3>, 2023.
- Weiss, M. and Baret, F.: S2ToolBox level 2 products: LAI, FAPAR, FCOVER, European Space Agency, http://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf, 2016.
- Xiong, X., Angal, A., Twedt, K. A., Chen, H., Link, D., Geng, X., Aldoretta, E., and Mu, Q.: MODIS reflective solar bands on-orbit calibration and performance, *IEEE Transactions on Geoscience and Remote Sensing*, 57, 6355–6371, <https://doi.org/10.1109/TGRS.2019.2905792>, 2019.
- Zhou, W., Rao, P., Jat, M. L., Singh, B., Poonia, S., Bijarniya, D., Kumar, M., Singh, L. K., Schulthess, U., and Singh, R.: Using Sentinel-2 to track field-level tillage practices at regional scales in smallholder systems, *Remote Sensing*, 13, 5108, 2021.