Response to Reviewer 2

General comments

This study proposes modifications to surface albedo representation in the ORCHIDEE-CROP model during leaf pigmentation changes and crop residue phases. The impacts of these changes on surface energy balance are analysed, and potential implications for Soil Water Content (SWC) under present-day and drier climate conditions are inferred.

While the modeling choices are generally reasonable, the proposed modifications do not convincingly demonstrate consistent improvements compared to the original ORCHIDEE-CROP formulation. In some cases, model performance appears to degrade, and the added complexity does not appear to translate into robust or statistically significant improvements. As currently presented, the manuscript would benefit from clearer methodological justification, more rigorous evaluation against the original model, and a more quantitative and transparent assessment of improvements and limitations.

I revised the manuscript up to Section 3.3 and recommend major revisions. Should the revised evaluation become more convincing, I would be happy to provide further feedback on the remaining sections.

[Response] Thank you so much for your time in reviewing our manuscript and for your valuable feedback on our work. Following your suggestions, the major revision of this manuscript is summarized in the following. We have also reorganized all corresponding figures and tables in the main text and supplementary information to follow their order of first appearance. For more details, please find our point-by-point responses to all reviewer comments below (the line numbers refer to the tracked text).

1. We improved the methodology and provided additional clarifications & justifications. This includes:

(1) Adding information on the quality control strategy for albedo retrievals. (see reply to comment(s) #1)

(2) Clarifying the construction of drying scenarios. (see reply to comment(s) #2)

(3) Adding justification for the use of MODIS LAI and the residue-related soil evaporation reduction parameterization. (see reply to comment(s) #3,5,20,28)

(4) Clarifying the calibration strategy for wheat phenology. (see reply to comment(s) #4,12,13)

(5) Unifying the meteorological forcing across all sites. (see reply to comment(s) #9)

(6) Replacing the von Kármán constant from 0.35 to a more realistic value of 0.41 and re-running all affected simulations. (see reply to comment(s) #11)

2. We extended the evaluation against the original model in order to provide a quantitative and transparent assessment of the improvements and limitations of the updated model. This includes:

(1) Explicitly comparing the performance of simulated surface albedo with and without the new developments. Extending the evaluation framework used for surface albedo to surface temperature, latent and sensible heat fluxes, and soil water content with and without the new developments. (see reply to comment(s) #6,7)

(2) Introducing formal statistical testing using paired t-tests across sites for all evaluated variables and periods (sections 3.2 and 3.3), and revising the text to avoid over-interpretation of non-significant changes, and summarizing the results in the added Table S4. (see reply to comment(s) #8,14)

Below, the authors may find some suggestions for methodological improvements. However, the authors are also encouraged to explore additional paths to improve model performance beyond the specific suggestions listed here.

[Response] Thanks for the helpful suggestions. Following the reviewer's suggestions, we are convinced that the revised manuscript includes a robust methodology.

1. *L122-136: It is said that 2 sites use a different methodology for calculations of albedo. The authors should try this methodology on the other 8 sites, compare the results and make sure that it is a reasonable approximation – this information should be added in the SI.*

[Response] Following this recommendation, we applied the satellite-based albedo retrieval approach to the eight sites where in situ shortwave radiation measurements are available, and compared the resulting albedo estimates with those derived from site radiation observations. The comparison is now presented in the supplementary information (Figure S15), providing a direct evaluation of the consistency between the two methodologies.

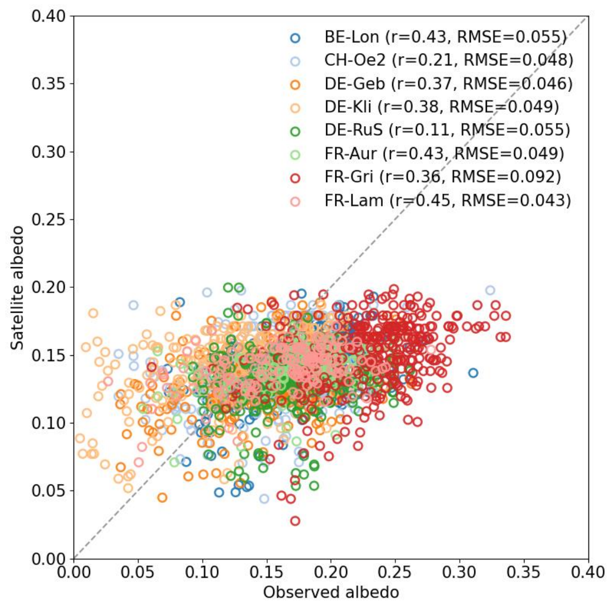

Figure S15. Comparison of surface albedo derived from site measurements (observed albedo) and the 5-day, 300-m Sentinel-2 albedo product (satellite albedo) at eight cropland sites. Site-level surface albedo was calculated as the ratio of outgoing to incoming shortwave radiation. The values of r and RMSE shown in the legend denote the Pearson correlation coefficient and the root mean square error, respectively. The gray dotted line indicates the 1:1 relationship.

Overall, the satellite-derived albedo shows a moderate but consistent correlation with site-based estimates, with correlation coefficients ranging from 0.36-0.43 at most sites, except of two sites (DE-RuS, CH-Oe2). Part of this discrepancy is likely related to the inherent spatial scale mismatch between the satellite product (300 m), and the in-situ sensor, which has a footprint of a few square meters. Nevertheless, the level of agreement is sufficient to support the use of satellite-based albedo as a reasonable approximation of field-scale surface albedo variability. For a more detailed analysis of the albedo product, please refer to our previous paper (attached as supplementary material<Spatiotemporal dynamics and drivers of bare soil albedo in European croplands.pdf>).

Given the limited number of available wheat sites, we decided to retain the two ClieNFarm sites where only satellite-based estimates were available in order to maximize the spatial coverage. Importantly, these two sites comprise six individual wheat fields, effectively adding six independent datasets to the analysis (Table S2). Retaining these sites thus substantially enhances the spatial sampling and representativeness of wheat surface albedo conditions. The associated uncertainty and its potential implications for the analysis are now explicitly discussed and acknowledged in Section 4.5 of the revised manuscript:

Line 756-767: "First, while site-based meteorological forcing (e.g., air temperature, wind speed, humidity) ensures high representativeness, **meteorological observations and direct radiation measurements were unavailable at two ClieNFarm sites comprising six farms (UK-Rookery, UK-Winwick, UK-Allerton, BE-AH, BE-BM, and BE-LB; Table S2). For these sites, simulations were** instead driven by the **hourly 0.25° ERA-interim product and surface albedo was derived from Sentinel-2 products.** This substitution inevitably limits the ability of our model to reproduce site-specific water and energy dynamics. **A comparison between satellite-derived albedo at 300 m spatial resolution and albedo calculated from shortwave radiation measurements representative for a few m^2 at eight sites reveals systematic differences (Fig. S15). These differences are likely attributable to spatial scale mismatch, cloud screening effects, and surface heterogeneity, which together dampen albedo variability in the satellite product. While satellite-based albedo enables inclusion of the six field sites, these substitute values introduce additional uncertainty in the simulated surface energy fluxes.**"

Also, it is mentioned that data are quality-controlled and gapfilled, and if I remember correctly these dataset come with quality flags for the quality of the measurement or for signaling gapfilling. Did the authors filtered the data based on this information? I wonder whether using only measurements flagged for good quality could improve the final performance.

[Response] In response to the request for clarification regarding the filtering and use of quality-controlled observations, we now provide a more detailed and explicit description of the full filtering strategy in the supplementary information (Text S1.1):

Line 14-26: "**Text 1.1 Quality filtering and preprocessing of surface albedo data**

Surface albedo estimates were derived after a multi-step quality filtering procedure adapted from Cescatti et al. (2012). Half-hourly records of incoming and reflected shortwave radiation (SW_IN and SW_OUT) were selected only when measurements were available or gap-filled and marked as reliable according to the corresponding quality control (QC) flag (values of 0, 1 and 2; Pastorello et al., 2020). We further required SW_IN to exceed 20 W m^{-2} and discarded cases where SW_OUT was equal to or greater than SW_IN.

To limit contamination from clouds, aerosols, and snow, we applied additional screening based on MODIS ancillary products (MCD43A2 and MOD(Y)D04_L2, https://ladsweb.modaps.eosdis.nasa.gov/). Daily albedo values were retained only when cloud cover remained below 60%, aerosol loading below 40%, and no snow presence was detected. Although applied at a coarse spatial resolution, this screening effectively reduces biases associated with periods of very weak shortwave radiation, for which albedo retrievals are inherently uncertain.

As a final step, we removed extreme values identified as exceeding five standard deviations from the distribution computed over the full analysis period, which predominantly captures residual measurement artifacts and uncharacterized sensor errors."

For clarity and readability, we chose to put this information in the SI and now refer to it in the methodology section. We are open to moving it to the main text if you or the editor thinks that would be more appropriate. This description was intentionally kept concise in section 2.1.1 in the methodology:
Line 125-126: "Data were quality-controlled and gap-filled using uniform methods (Pastorello et al., 2020). **Details of quality filtering and preprocessing of α_{surf} are shown in Text S1.1.**"

***2.** L147-150: I disagree with this approach. Using rainfall from the driest year, while the other variables vary means that decoupled quantities are being used, which is not physically realistic. Meteorological fields should be allowed to change coherently, and then the effects that the authors find interesting should be isolated via statistical methods.*

[Response] We apologized for the unclear descriptions in our original manuscript. This is a misunderstanding; we have not used only the rainfall, but all variables from the driest year to ensure consistency among meteorological variables. We agree with you on this critical point. We have revised the text in section 2.1.2 to avoid this misunderstanding:
Line 148-152: "**For the drying climate scenario, the driest year at each site was identified based on the lowest annual precipitation. All meteorological variables from this year, including precipitation, air temperature, wind speed, humidity, and radiation, were then used to force the simulation. This ensures that the drying climate scenario is driven by a coherent and physically realistic set of meteorological conditions. Yearly-averaged meteorological variables in the driest year for each site are presented in Table S1.**"

***3.** L162-167: It is true that LAI measurements are sparse at ICOS site. However, I suggest that the authors use the locally available information to correct the MODIS values and match the locally shown variability and range (after all, a max LAI=2.5 is quite low for wheat in my opinion)*

[Response] We agree that LAI measurements at ICOS sites are sparse, and that MODIS LAI can have biases in their absolute values. Nonetheless, we argue that the use of MODIS LAI is justified.
Our analysis focuses on the temporal evolution of LAI to identify the timing of maximum LAI, rather than on the absolute peak LAI value itself. For this purpose, a temporally continuous LAI time series is required, which is only available from MODIS. While site-level LAI observations are valuable for validation purposes, their limited temporal coverage, typically consisting of only a few observation dates, prevents their use for correcting MODIS LAI or reconstructing seasonal LAI dynamics.
Nevertheless, comparison between MODIS-derived LAI time series and all available site observations show that, despite the sparsity of the in-situ data, the timing and approximate magnitude of peak LAI are generally consistent (Fig. S2).
Following this comment, we updated Fig. S2 and Table S3 to include LAI comparisons for all site-years with available LAI measurements (following comment #28) and revised Section 2.1.3 to emphasize that MODIS LAI is used to identify the timing of peak LAI rather than its absolute values:

Line 167-169: "Due to the absence or sporadicity of leaf area index (*LAI*, m^2 m^{-2}) measurements at ICOS sites (Table S3), we used continuous satellite-based *LAI* datasets to **identify the timing of peak LAI** at those sites."

Line 171-172: "**MODIS-derived LAI was used primarily to constrain the onset of the maturity period, rather than to reproduce absolute LAI values.**"

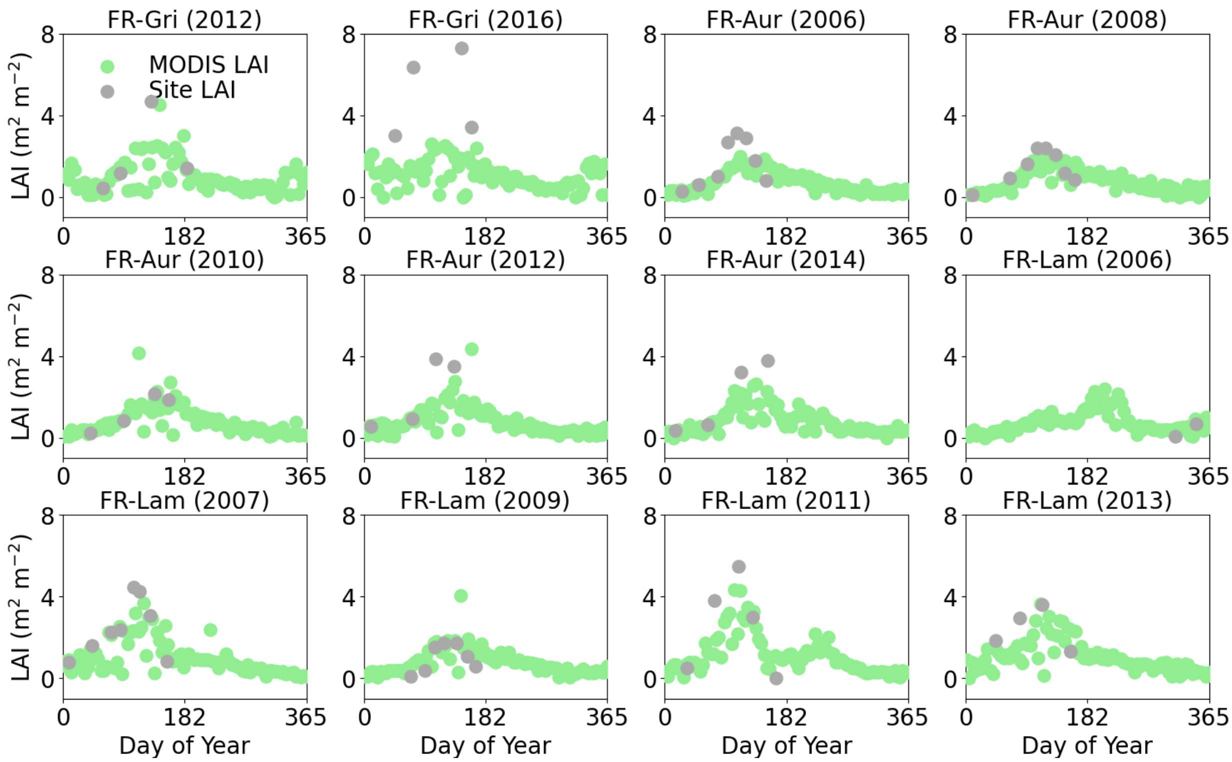

Figure S2. Comparison of leaf area index (LAI) between daily satellite-based product (4-day, 500-m MCD15A3H) (green dots) and site measurements (grey dots) at three ICOS sites in different years.

***4.** L257-258: This sentence is unclear. Does it mean that sowing and maturity dates are determined via GDD and then a number of days after maturity is added so that the observed/modeled harvesting dates match? If that is the case, it is then not guaranteed that modeled sowing and maturity days match with the observed ones. An evaluation of this aspect is also required, as I imagine this would affect calibration of the parameters related to leaf pigmentation (and also beta_4 and Z_0). Is there a reason why the authors didn't force STICS to use observed dates instead of the ones determined via GDD? An alternative could be to retrain STICS so that sowing and maturity provided via GDD match the observed sowing and maturity dates.*

[Response] In our modeling framework, the sowing date is prescribed using observed site-specific management information, ensuring consistency between the model and observations at each site. The subsequent phenological development is determined by accumulated growing degree days (GDD) following the phenology scheme.

The GDD thresholds used to determine phenological stages were not arbitrarily chosen, but were carefully calibrated using extensive phenological observations, as described in Su et al. (2025). For the convenience

of the reviewer, we here detail the approach: Specifically, ORCHIDEE-CROP was calibrated using phenological records of wheat across Germany, including location, planting year, planting date, flowering date, and harvest date. The calibration dataset comprises observations from over 1941 sites for wheat. During the calibration, parameters controlling the GDD requirements for key phenological stages were optimized with the objective of minimizing the root mean square errors (RMSEs) of simulated flowering and harvest dates. The calibration focused on regions where wheat is predominantly rainfed, in order to avoid the confounding effects of irrigation on phenological development.
As a result, the modeled maturity and harvest dates derived from GDD are, on average, consistent with observed phenology. Therefore, although maturity dates are not explicitly forced to match observations at each individual site available in this study, the calibrated GDD-based phenology provides a robust and internally consistent representation of crop development across a wide range of environmental conditions.
To ensure consistency with site observations, the delay between maturity and harvest was further optimized at each site using reported harvest dates, as described in Section 2.3 (lines 267-272) and reported in Table S2. The period between physiological maturation and harvest varies among sites due to practical constraints. We note that direct observations of the exact maturity date are not available at the sites, and therefore the calibration described above represents the best possible approach. This clarification has been added to the revised sections 2.2.5 and 2.3.
Line 249-255: "**In our simulations, sowing dates were prescribed using observed, site-specific management information to ensure consistency with field observations. Subsequent phenological development, including flowering, maturity, and harvest, is driven by growing degree days following the ORCHIDEE-CROP phenology scheme which was calibrated using extensive wheat phenology records from 1941 DWD climate stations across Germany (Su et al., 2025). This approach ensures that modeled flowering and harvesting dates are on average, consistent with observed phenology across a wide range of environmental conditions. We further adjusted the delay between crop maturation and harvest at each site using actual harvest dates at each site which vary depending on practical reasons and weather (Table S2).**"
Line 267-268: "In addition, **to ensure consistency between simulated and observed crop development periods** (from sowing to harvest) at each site, [...]".
Line 272-274: "It means that the **site-specific** durations of foliar yellowing before harvesting vary across all sites in the new model, compared to the default 14 days in the old model."

***5.** L291: why is beta_4 assumed at 50% and not calibrated on observations? Picking a value in the upper range can lead to exaggerating the effects on Esoil. If I remember correctly, the selected sites should provide measured LE fluxes, which could be used to calibrate this parameter.*

[Response] We disagree that the parameter is in the upper range, as it cannot be directly observed. In ORCHIDEE-CROP, soil evaporation (E_{soil}) is a function of multiple interacting variables, including air density, wind speed, aerodynamic resistance, surface moisture availability, and atmospheric humidity (see Section 2.2.3). The parameter β_4 represents an empirical reduction factor for soil conductance induced by residue cover, but it is only one of many parameters controlling E_{soil}.
The generic calibration of ORCHIDEE-CROP is not capable of capturing site-specific hydrological properties. Calibrating β_4 against observed latent heat fluxes would therefore require simultaneous recalibration of several other interacting hydrological parameters and risks compensating for existing model biases rather than isolating residue effects. Given the limited number of site observations and the

complexity of the water-energy partitioning, such site-level optimization was considered beyond the scope of this study.

Instead, we choose to test a range of β_4 values and assess the resulting relative reduction in simulated E_{soil} against experimental evidence reported in the literature (Fig. S3, Table S4). This sensitivity analysis allowed us to select a value that yields a residue-induced reduction in E_{soil} consistent with the 20-50% range documented by field studies compiled in this work. This approach targets a realistic mean effect size expected among European croplands and focuses on improving the representation of albedo properties of winter wheat during the yellowing and residue covering periods, while associated impacts on surface energy fluxes are evaluated qualitatively rather than through a full recalibration of the hydrological scheme.

We now explicitly clarify both the scope of this work in section 2.3:

Line 275-278: "**The objective of this model development targets a realistic mean effect size expected among European croplands, rather than capturing in detail the site-specific effects. We focus on improving the representation of albedo properties of winter wheat during the yellowing and residue covering period, while associated impacts on surface energy fluxes are evaluated qualitatively rather than through a full recalibration of the hydrological scheme.**"

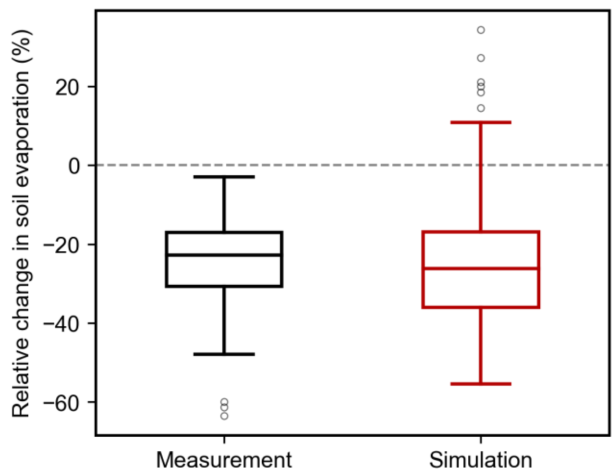

Figure S3. The comparison of relative change in soil evaporation (%) between model simulation and eight field experiments during residue covering periods. Values from model simulation derive from ORC-AE and ORC-D. ORC-AE adjusts surface albedo, soil conductance and surface roughness together in the ORCHIDEE-CROP model. ORC-D is the initial model version. The solid black and red lines within each box indicate the mean values, respectively. The box spans the interquartile range, covering the 25th percentile (lower bound) to the 75th percentile (upper bound). Black crossbars at the top and bottom represent the dataset's minimum and maximum values. The gray hollow circles are outliers. The gray horizontal dotted line represents zero on the y-axis.

Table S4. Site experiments to analyze changes in soil evaporation due to residue coverage.

Residue covering period	**Site location**	**Latitude, Longitude**	**Relative change (%)**	**Reference**

March-July in 1997-1998	Losses Plateau, China	35°33’N, 104°35’E	-6.05	Huang et al., 2005
May-August in 2012	Chengcheng, China	37.89° N, 114.69° E	-26.14	Li et al., 2013
June-September in 2007-2019	Heyang, China	35°19’N, 110°4’E	-30.09	Li et al., 2022
April-June in 2014	Beijing, China	39°37′N, 116°25′E	-31.31	Wang et al., 2018
June-July in 1999	Luancheng, China	37°53’N, 114°40’E	-61.30	Zhang et al., 2005
October 1988-June 1989	Merredin, Australia	31 °29’S, 118°12’E	-2.98	Yunusa et al., 1994
September-October, December 2008	Londrina, Brazil	23º18’S, 51º09’W	-17.48	Gava et al., 2013
June 2018-September 2018	Ahvaz, Iran	31.30° N, 48.65° E	-22.72	Yusefi et al., 2020
November 2023-January 2024	Henan, China	35°14’N,113°76’E	-18.5	Wang et al., 2025
-	Losses Plateau, China	34°17’N, 108°04’E	-19.46	Xing et al., 2017
June-Septemble 2002-2004	Heyang, China	35°19’37”N, 110°4’57”E	-17.22	Zhang et al., 2016

June-August 2010-2011	Beijing, China	39.6° N, 116.8° E	-21.67	Li et al., 2015

6. *L375-380: The results section should start with the inferred parameter values, and their 95% CIs. Evaluation on albedo predictions should also consider the performance of the old model: how far do the improvements go? I suggest to modify Figure 2 to include both the predictions from the new model and from the old model (with another color), so that the evaluation can be presented coherently.*

[Response] We agree on all points and revised the manuscript accordingly.
First, we now include the inferred parameter values (i.e., k_1, k_2, k_3 and k_4) and their 95% CIs:
Line 41.-414: "**During the albedo calibration phase, the estimated parameter values of $\boldsymbol{k_1}$, $\boldsymbol{k_2}$, $\boldsymbol{k_3}$ and $\boldsymbol{k_4}$ (Eqs.12 and 14) are 0.013 (95% CI: 0.012-0.013), 1.005 (95% CI: 0.998-1.017), 0.895 (95% CI: 0.889-0.902) and 0.0022 (95% CI: 0.0017-0.0026), respectively. Using these calibrated parameters, the relationship between observed surface albedo (α_{surf}) and the parameterized α_{surf} representation used to constrain the model shows a moderate correlation with an R^2 of 0.54 during the foliar yellowing period, and of 0.62 during the residue covering period for five testing sites.**"
Second, we have updated our evaluation to explicitly compare the performance of the original and modified model versions across all relevant variables. For albedo, Figure 2 now shows predictions from both the original model (without albedo parameterizations) and the new model (with albedo parameterizations in foliar yellowing and residue cover processes), thereby highlighting the importance of considering these two processes.

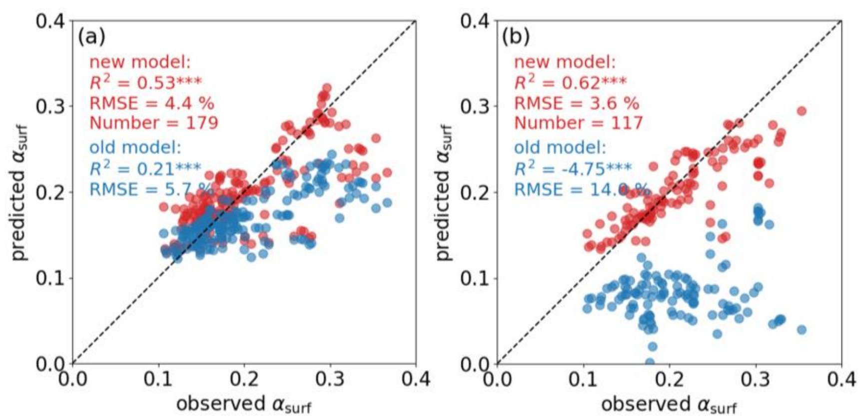

Figure 2. The comparison of daily surface albedo (α_{surf}) predictions **derived from calibration models with and without explicit albedo parameterizations against** independent observations from five sites in Europe during (a) foliar yellowing and (b) residue covering periods. **Coefficient of determination (R^2) and root mean square error (RMSE) are shown in the upper-left corners of each panel. '***' shows the comparison passes the significance level of 0.01.** The dotted black line is the 1:1 line.

Furthermore, this section should include the evaluation of Tsurf, H and LE: instead of presenting Figure S9-S10, the authors should conduct an evaluation similarly to the one conducted in Figure 2. This one should also quantify performances of the original model (ORC-D) and of the modified version (ORC-AE).

[Response] We have extended the evaluation to water and heat variables (surface temperature, latent and sensible heat fluxes, and total soil water content) following the same approach as Figure 2. For each variable, we present daily relative changes (normalized to the first day of each period, %) during the foliar yellowing and residue covering periods across all available sites, along with quantitative metrics of model performance (R and RMSE) for ORC-D and ORC-AE against site observation (Fig. S4).

This comparison demonstrates that the modified model achieves clear improvements in simulated surface albedo and surface temperature during both foliar yellowing and residue covering periods (Fig. S4a-d), reflecting the direct impact of the introduced albedo and roughness parameterizations. Improvements in latent and sensible heat fluxes are more moderate and are mainly evident during the residue covering period (Fig. S4e-h), while no systematic improvement is found for total soil water content (Fig. S4i-j).

We now include the result of daily relative change of α_{surf} (normalized to the first day of each period, %) during the foliar yellowing and residue covering periods (Fig. S4a,b) to section 3.1:

Line 418-422: "**The refined ORCHIDEE-CROP configuration (ORC-AE), represents the new model version with effects of the modified α_{surf} and the refined soil conductance (β_4) and surface roughness (*Z_0*), was used to simulate α_{surf} dynamics at all 33 site years, which were evaluated independently against observations (Fig. S4a,b). This independent evaluation shows a substantially stronger agreement, with significantly higher Pearson correlation coefficients compared to the default model configuration (ORC-D).**"

In addition, we explicitly discuss these contrasting responses of all variables in Section 4.5:

Line 708-719: "**α_{surf} shows the most consistent improvement across sites, reflecting its direct control by the new parameterizations for foliar yellowing and residue covering (Fig. 2, Fig. S4a,b). However, degraded performance is observed at two sites (DE-Geb in 2014 and FR-Aur in 2019; Fig. S4a,c). At these sites, the refined configuration simulates higher α_{surf} during foliar yellowing and residue covering than estimated by site radiometer measurements. Surface conditions shown by the site photos suggest that the discrepancy may be related to limitations in the spatial representativeness of point-scale observations, which is further investigated in Text S1.2. T_{surf} also significantly improves, especially during the foliar yellowing period (Fig. S4c,d). In contrast, *LE* and *H* exhibit moderate improvement during the residue covering period and little to no improvement during the crop growth and foliar yellowing periods (Fig. S4e-g). This performance is expected, as *LE* and *H* during the growing season are primarily governed by transpiration processes not recalibrated in this study. During the residue covering period, remaining biases in *LE* and *H* likely arise from unrepresented processes associated with residue cover, such as water interception and retention by residues and soil hydraulic heterogeneity. Limitation in the representation of land-atmosphere coupling in ORCHIDEE also contributes to persistent uncertainties in simulated turbulent fluxes (Sect. 4.3).**"

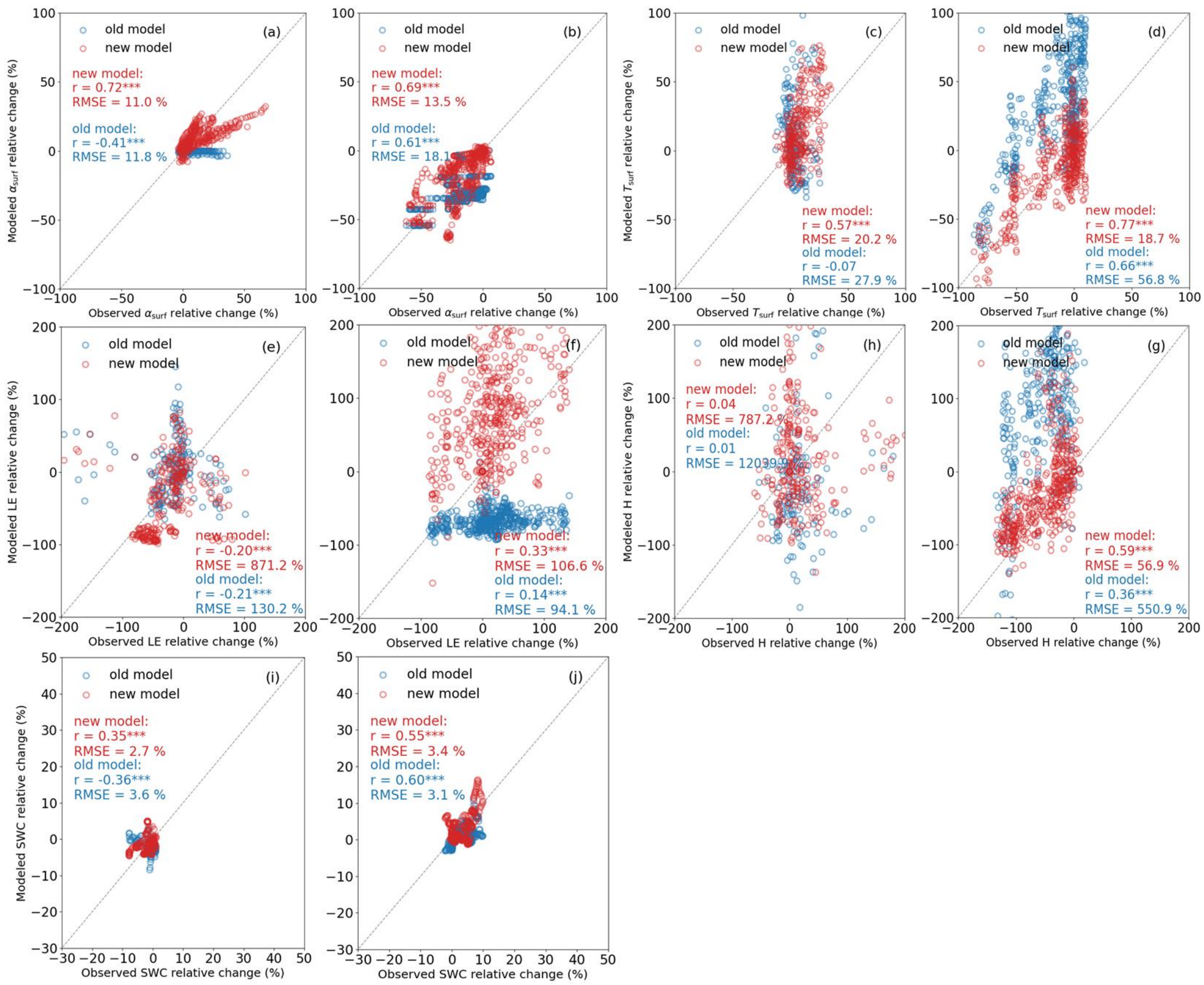

Figure S4. Daily relative changes (normalized to the first day of each period, %) in surface albedo (α_{surf}), surface temperature (T_{surf}), latent (LE) and sensible (H) heat fluxes, and total soil water content (SWC) during the foliar yellowing and residue covering periods across all available sites. The y-axis shows relative changes simulated by ORC-AE (red dots) and ORC-D (blue dots), while the x-axis shows the corresponding relative changes derived from site observations. ORC-AE represents the new model version with effects of the modified α_{surf} and the refined soil conductance (β_4) and surface roughness (Z_0), while ORC-D represents the initial version of the model. The values of r and RMSE shown in the legend denote the Pearson correlation coefficient and the root mean square error, respectively. '***' shows the comparison passes the significance level of 0.01. The gray dotted line indicates the 1:1 relationship.

<u>***7.*** *L386: ORC-AE does not capture alpha_surf evolution at all sites. It is quite clear that the new model leads to a degradation of albedo estimates in two out of five instances (De-Geb in 2014, FR-Aur in 2019). The authors should state this clearly in the paragraph and should better quantify improvements and shortcomings.*</u>

[Response] We agree that ORC-AE does not improve surface albedo simulations at all sites, and that albedo estimates are degraded at two site-year combinations (DE-Geb in 2014 and FR-Aur in 2019). We have now explicitly stated this in the manuscript and quantified both improvements and shortcomings.

At these two sites, the model simulates an increase in surface albedo during the foliar yellowing and residue covering periods, whereas the observed albedo derived from radiometer measurements remains anomalously low (Fig. S5a,c). To better understand this discrepancy, we inspected site photographs taken around the dates of maximum LAI, maturity, and harvest (Figure S16 below). These photos indicate visibly bright surfaces during yellowing and residue cover compared with dark-green periods when LAI reaches maximum, which would be expected to increase surface albedo and are consistent with the model behavior. This suggests that the mismatch is likely related to measurement issues.

Despite these limitations at specific sites, ORC-AE generally improves both the timing and magnitude of albedo changes across all sites (Fig. S4a,b). We explicitly clarified these two cases in Supplementary Text 1.2.

Line 28-45: "**Text S1.2 Site-specific degradation of surface albedo simulations**

While the modified ORCHIDEE-CROP configuration (ORC-AE) improves the simulation of surface albedo at most sites, degraded performance is observed for two site-years (DE-Geb in 2014 and FR-Aur in 2019). For these site-years, ORC-AE simulates an increase in surface albedo during the foliar yellowing and residue covering periods, whereas the albedo derived from shortwave radiation measurements remains comparatively low (Fig. S5a,c).

To investigate the origin of this mismatch, we examined photographs taken on site around key phenological stages, including the dates of maximum leaf area index (LAI), crop maturity, and harvest. These images show visibly bright surface conditions during both foliar yellowing and residue covering periods, possibly associated with high surface reflectivity (Fig. S16). This qualitative evidence suggests that the model captures the expected surface state during these periods at the field scale.

The persistent low albedo values derived from radiometer observations at these two sites may therefore reflect limitations in observational representativeness rather than deficiencies in the model formulation. Radiometer measurements sample a limited spatial footprint, typically on the order of 8-1000 m^2 (Eichler, 2022), and may therefore be influenced by local heterogeneity in residue distribution, sensor viewing geometry, or the partial sampling of surface patches that are not representative of the surrounding field. Such effects are particularly relevant during post-harvest periods, when residue cover can be highly heterogeneous at small spatial scales.

These cases highlight the challenges of evaluating field-scale albedo dynamics using point-scale observations and underscore the importance of combining local measurements with spatially distributed remote sensing data when assessing crop albedo processes."

This clarification is also explicitly mentioned in Section 4.5:

Line 708-713: "[...]. α_{surf} **shows the most consistent improvement across sites, reflecting its direct control by the new parameterizations for foliar yellowing and residue covering (Fig. 2, Fig. S4a,b). However, degraded performance is observed for two sites (DE-Geb in 2014 and FR-Aur in 2019) (Fig. S4a,c). At these sites, the refined configuration simulates higher** α_{surf} **during foliar yellowing and residue covering than estimated by site radiometer measurements. Surface conditions shown by the site photos suggest that the discrepancy may be related to limitations in the spatial representativeness of point-scale observations, which is further investigated in Text S1.2.**"

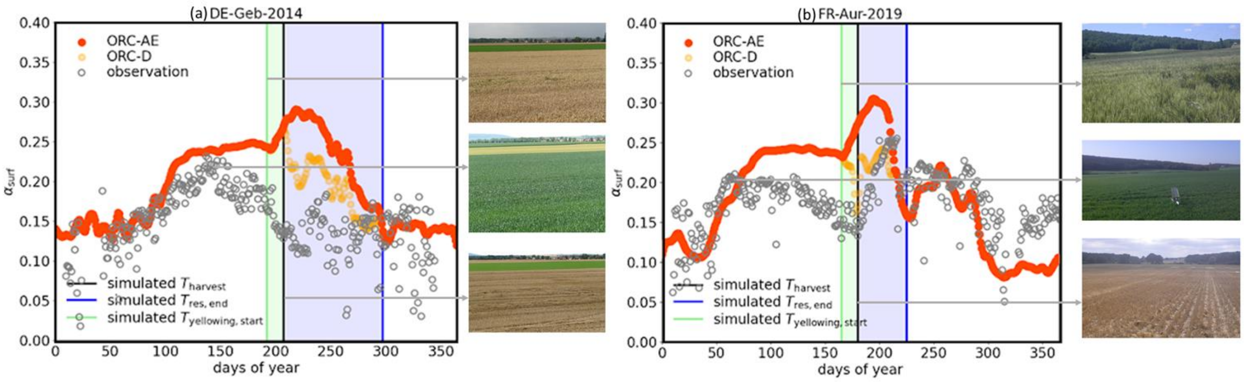

Figure S16. Surface conditions in different winter wheat growing periods at (a) DE-Geb and (b) FR-Aur sites in 2014 and 2019, respectively. Plots represent surface albedo evolutions predicted from the old (orange dots, ORC-D) and new (red dots, ORC-AE) ORCHIDEE-CROP models and observations (gray dots). The observed surface albedo is derived from site radiation measurements through the Integrated Carbon Observation System (ICOS) Data Portal. The green, black and blue solid lines are the simulated start of the foliar yellowing period (shallow-green area), harvesting date and the end of residue covering period (shallow-blue area), respectively. The surface conditions are identified based on site photos obtained from farm managers.

***8.** L405-436: It seems to me that there is virtually no change in any variable during foliar yellowing. What the phrase as "slightly lower" should be better phrased as "no change", given also the large variability shown. Residue covering shows some larger effects, but it wouldn't seem like it's significant. The authors are requested to better quantify statistically the inferred changes (statistical significance), in order to understand better the extent of the effects of the changes they made in ORCHIDEE-CROP. Furthermore, this section is quite number intensive and hard to follow – I suggest that the authors find alternative phrasings to present their finding. Also, I believe that percentages are seldomly relevant in this context.*

[Response] We agree that during the foliar yellowing period, the changes induced by the refined ORCHIDEE-CROP configurations are small relative to the overall variability and we revised the interpretation. In addition, to provide a quantitative assessment, we now explicitly report the results of statistical significance tests. A t-test was applied across sites for each variable and period (Table S5). The results show that changes in T_{surf} and LE show significant changes across sites during either the foliar yellowing or residue covering periods ($p<0.05$) for ORC-AE, ORC-A and ORC-E compared with ORC-D. While SWC generally does not exhibit statistically significant changes ($p>0.05$). During the foliar yellowing period, E_{soil} is not significantly affected in ORC-A, while in ORC-E H is significantly affected ($p>0.05$). This indicates that mean differences of targeted variables can be observed when averaging across sites, except for SWC.

Following this, we have revised Sections 3.2 and 3.3 to avoid over-interpreting insignificant mean differences during the foliar yellowing and residue covering period, with consideration of the significance test. To improve clarity and readability, detailed numerical values are now consolidated in Table S5, and the main text focuses on the description of the magnitude and direction of changes simulated by ORC-AE, which represent the combined effects of the newly implemented processes:

Line 451-536: "

3.2 The impact of improved surface albedo, surface roughness and soil conductance on soil evaporation, surface temperature and soil water content

Compared with the standard version of ORCHIDEE-CROP (ORC-D), statistically significant changes in averaged surface temperature (T_{surf}) are detected across sites in ORC-AE and ORC-E (p<0.05 in t-test), resulting from modifications in surface albedo (α_{surf}), surface roughness (Z_0), and soil conductance (β_4) (Table S5). In contrast, improving albedo alone (ORC-A) does not lead to statistically significant changes in soil evaporation (E_{soil}) during the foliar yellowing period (p>0.05). Soil water content (SWC) generally does not exhibit statistically significant changes (p>0.05) in any of the three model configurations. This indicates that, except for SWC, site-averaged differences in the targeted variables are detectable and exhibit consistent signs and magnitudes across sites.

ORC-AE **leads to a statistically significant reduction in** E_{soil} **relative to ORC-D** during foliar yellowing period, with an averaged **decrease of -0.18±0.08** mm d^{-1} (Fig. 4a). **A slightly smaller but consistent reduction of E_{soil} (-0.17±0.14 mm d^{-1}) occurs during the residue covering period. Changes in β_4 and Z_0 (ORC-E) also exert significant impacts on E_{soil} that are comparable in magnitude to those in ORC-AE during both periods**, **whereas changes in** α_{surf} alone (ORC-A) cause **relatively minor and statistically insignificant** reductions in E_{soil} for the same period. The influence of crop residues on E_{soil} disappears within **approximately** 40 days after residue coverage (Fig. S6a, Fig. S7a).

The reduced E_{soil} in ORC-AE leads to a **modest but not significant** increase in *SWC* **over 2 m soil depth** during the residue covering period among all sites **(13.30±8.02 kg m^{-2})** (Fig. 4c, **Table S5**). The combined effect is **comparable to** the sum of the **individual contributions from** α_{surf} (ORC-A) and **from** β_4 and Z_0 (ORC-E), suggesting an approximately additive effect. **Although statistically insignificant, this weak water conservation signal persists toward** the end of year, with gradual infiltration into deeper soil layers (Fig. S8).

Incorporating changes in α_{surf}, β_4 and Z_0 in ORC-AE results in a **statistically significant** reduction in T_{surf} during foliar yellowing and residue covering periods with averaged cooling of **-1.55±0.93 °C** and **-1.90±0.88 °C** over all sites, respectively, **relative to** ORC-D simulation (Fig. 4b). The factorial simulation experiments show that this cooling effect is **mainly driven by** changes in E_{soil} and Z_0 (ORC-E), **whereas increased** α_{surf} alone **produces a weaker and statistically insignificant** T_{surf} reductions. Interaction among α_{surf}, β_4 and Z_0 (ORC-AE) lead to a dampened overall effect compared to individual effects of these factors on T_{surf}. The reduction in T_{surf} tends to weaken over a**pproximately** 60 days of decreasing residue coverage (Fig. S6d, Fig. S7d).

3.3 The impact of improved surface albedo, surface roughness and soil conductance on latent and sensible heat flux

Relative to ORC-D simulation, statistically significant changes in latent heat fluxes (*LE*) induced by modifications in surface albedo (α_{surf}), surface roughness (Z_0), and soil conductance (β_4) are detected across sites for ORC-AE, ORC-A, and ORC-E configurations (p<0.05 in t-test) (Table S5). Changes in sensible heat fluxes (*H*) are also statistically significant for ORC-AE and ORC-A, whereas modifications in Z_0 and β_4 alone (ORC-E) do not produce statistically significant changes in *H* (p>0.05). This suggests that the modified model configurations capture systematic variations in surface water and energy exchanges when averaged across sites.

Modified α_{surf}, β_4 and Z_0 in ORC-AE **leads to statistically significant reductions in** *LE* relative to ORC-D during foliar yellowing and residue covering periods, with the averaged changes of **-6.84±1.41 W m^{-2} and -5.15±4.03 W m^{-2}**, respectively **(Fig. 4e)**. **This process is primarily controlled by changes in β_4 and Z_0, represented by comparable reductions in** *LE* **by ORC-E. While α_{surf} alone (ORC-A) plays a more limited role, the effect remains statistically significant.**

The corresponding changes in the *H* **are also statistically significant and of similar magnitude to those in *LE* induced under the combined effect of α_{surf}, β_4 and Z_0 during the foliar yellowing and residue covering periods,** with averaged decreases of **-8.10±4.18W m^{-2} and of -7.44±3.98** W m^{-2}, compared to the ORC-D simulation (Fig. 4f). **In both periods**, direct radiative effects associated with increased α_{surf} **(ORC-A) contribute most strongly to the reduction in** *H*, **whereas changes in the surface-atmosphere energy exchange (ORC-E) leads to only a slight, statistically not significant increase on** *H***.** Consistent with the T_{surf} **response**, the influence of crop residues on *H* **weakens over the approximately** 60 days of residue cover (Fig. S6c, Fig. S7c)."

Minor comments:

9. *L144-145: I don't see any reason behind using CRU-JRA at the two ClieNFarm sites rather than ERA-interim – why did the authors decided on a different product in this case instead of using ERA-interim data corresponding to those two locations? The authors should uniform the data across all sites, but if they decide to keep this different meteorological input, they should justify their choice adequately.*

[Response] We agree with the reviewer and we have updated our simulations at the two ClieNFarm cropland sites by replacing the CRU-JRA forcing with the hourly 0.25° ERA-Interim meteorological dataset. The model simulations for these two sites were fully re-run using ERA-Interim forcing.

We have revised Section 2.1.2 accordingly, correcting and clarifying the description of the meteorological input data:

Line 144-146: "At two ClieNFarm sites with no flux observations, the meteorological variables from the **hourly 0.25° ERA-interim product** were utilized as climate forcing data **by linearly resampling to 0.5°** in this model."

10. *L151-167: I suggest aggregating subsections 2.1.3, 2.1.4, 2.1.5, because they are very short. I'd instead merge them in a unique section (called, for instance, "Management practice, bare soil albedo and LAI").*

[Response] We agree and we have merged sections 2.1.3-2.1.5 into a single subsection titled "Management practices, bare soil albedo, and LAI". This restructuring improves readability and reduces fragmentation in the methods section:

Line 155-174:

"**2.1.3 Management practice, bare soil albedo and LAI**

Management practices for winter wheat years at all sites (Table S1) were obtained from a crop management database provided by site managers. Key practices, including sowing, harvesting, and tillage dates, were prepared as input for the ORCHIDEE-CROP model.

Bare soil albedo (α_{soil}, unitless) was prescribed using a high-resolution 5-day, 300-m European dataset derived from 10 m S2 α_{surf} **observations covering the period** 2018-2020 (Yu et al., in review). This dataset demonstrated sufficient skills in resolving daily soil dynamics at the field scale in European fragmented croplands. **To align with the spatiotemporal resolution of the ORCHIDEE-CROP model,** the new soil dataset was linearly aggregated to 0.5° and a daily time step.

Due to the absence or sporadic leaf area index (LAI, m^2 m^{-2}) measurements found at ICOS sites (Table S3), we used continuous satellite-based LAI datasets **to identify the timing of the LAI peak** at those sites. **LAI at each site was extracted from** the 4-day, 500-m MODIS LAI product (MCD15A3H) (Myneni et al., 2015). Daily LAI dynamics were derived through linear interpolation combined with a Savitzky-Golay filter (window length=15) to smooth the trend and remove the outliers. **MODIS-derived LAI was used primarily to constrain the onset of the maturity period, rather than to reproduce absolute LAI values.** The LAI comparison between MODIS product and site measurements is shown in Fig. S2. We used the MODIS-derived but not S2**-derived LAI product, because the temporal coverage of the latter is insufficient to match the available site observations.**"

11. *L220: I thought that the von Karman constant was more around 0.40-0.41. Why using a value this different from what is typically used?*

[Response] We agree that the standard value of the von Kárxmán constant is often cited as 0.40-0.41 (Tennekes 1973). In response to this comment, we have updated the von Kármán constant to 0.41 throughout the model and rerun all relevant simulations.
This modification leads to a clearer and more statistically robust response of surface processes during the foliar yellowing period when comparing the refined configurations (ORC-AE, ORC-A and ORC-E) with the default version (ORC-D). Specifically, the impacts on soil evaporation, latent and sensible heat fluxes, and surface temperature become more pronounced and statistically significant across sites (Fig. 4).
We also found that the updated simulations affect the relative contributions identified in the factorial analysis (ORC-A and ORC-E). The combined effects of surface roughness and soil conductance become more important than the surface albedo impact, especially for surface temperature responses (Fig. 4b,d). Based on the new results, all corresponding figures and tables have been updated accordingly (Figs 4 and 5, Figs. S6-S9, Fig. S11), and the discussion parts (Sect. 4.1 and 4.2) have been revised to reflect these strengthened and rebalanced effects.

12. *L316: does the value reported indicate mean +- SD? Looks like quite a wide range. Does this come from only 33 values, correct? The authors should not indicate SD, but instead the 95% CI for the mean: I imagine the distribution would not be normal, so I suggest to compute bootstrap CIs. Same goes for the value in L320.*

[Response] The originally reported value indeed corresponded to the mean±standard deviation derived from the 33 site observations. Following the recommendation, we have revised the manuscript to report the mean onset timing together with the 95% confidence interval of the mean, estimated using non-parametric bootstrap resampling (10,000 iterations), which does not require a normal distribution. Results show that the duration between the onset of foliar yellowing and harvest is 22.3 days (95% bootstrap CI: [20.6-23.6] days, n=33), while the duration between harvest and the end of α_{surf} increase is 13.4 days (95% bootstrap CI: [10.5-15.9] days, n=33). We have revised section 2.4.1:
Line 339-346: "α_{surf} dynamics were parameterized by identifying critical temporal thresholds. The minimum between the *LAI* peak and harvest marked the onset of foliar yellowing, which is approximately **22.3 days (95% bootstrap CI: [20.6-23.6] days, n=33)** pre-harvest computed by site observation. This delay is longer than the assumed delay between maturity date and harvest in ORCHIDEE-CROP of 14 days. The maximum α_{surf}, constrained to occur within 30 days post-harvest, divided residue-driven α_{surf} into an increase phase (harvest to maximum) and a decline phase from post-maximum α_{surf} to tillage. Without tillage, the residue covering period persists for 90 days post-harvest (Kriauciuniene et al., 2012). On average,

the maximum α_{surf} occurs **13.4 days (95% bootstrap CI: [10.5-15.9] days, n=33)** after harvest based on site identification."

13. *L317: "This delay is close … of 14 days." Is this to state what the old model is doing? Please better clarify within the paragraph what are the settings for the old model and which are the settings for the new one.*

[Response] We apologized for the unclear descriptions in our original manuscript. The sentence referring to the 14-day delay was intended to describe the default assumption used in the original ORCHIDEE-CROP model, which was not clearly stated.

In the revised manuscript, we now explicitly distinguish between the default model configuration, which assumes a fixed 14-day interval between maturity and harvest, and the new parameterization introduced in this study, which derives site-specific maturity-harvest duration based on observed α_{surf} trajectories. The paragraph has been rewritten to clearly separate the old and new model settings and to avoid ambiguity:

Line 339-342: "The minimum α_{surf} between the *LAI* peak and harvest marked the onset of foliar yellowing, which is approximately **22.3 days (95% bootstrap CI: [20.6-23.6] days, n=33)** pre-harvest computed by site observation. This delay is **longer than the default 14-day assumption** between maturity date and harvest **in the old ORCHIDEE-CROP model**."

Line 345-347: "On average, the maximum α_{surf} occurs **13.4 days (95% bootstrap CI: [10.5-15.9] days**, n=33) after harvest based on site identification, **while the configuration in the old ORCHIDEE-CROP model does not consider the timing of maximum $\boldsymbol{\alpha_{surf}}$.**"

14. *L375-376: the authors claim that R2 of 0.54 is a "good performance". The adjective is not appropriate and should be reconsidered. Good compared to what? Other models in the literature? Compared to the old model? One could argue that 0.54 is far from being optimal.*

[Response] We agree and have revised the sentence:

Line 412-422: "Using the calibrated parameters, the relationship between observed surface albedo (α_{surf}) and the parameterized α_{surf} representation used to constrain the **model shows a moderate correlation with an R^2 of 0.54 during the foliar yellowing period and of 0.62 during the residue covering period for five testing sites.** The corresponding RMSE has values of 0.04 and 0.03, respectively (Fig. 2). The least-squares regression in these two fitting models captures the mean tendency of the data unlike **models that consider them constant**, but falls short of reproducing full variation, as it minimizes average errors rather than explaining outliers.

The refined ORCHIDEE-CROP configuration (ORC-AE), represents the new model version with effects of the modified $\boldsymbol{\alpha_{surf}}$ and the refined soil conductance (β_4) and surface roughness (*Z_0*), was used to simulate $\boldsymbol{\alpha_{surf}}$ dynamics at all 33 site years, which were evaluated independently against observations (Fig. S4a,b). This independent evaluation shows a substantially stronger agreement, with significantly higher Pearson correlation coefficients compared to the default model configuration (ORC-D)."

15. *L386-387: is FR-Gri2018 in the testing dataset? I thought there were 5 sites and a total of 5 years for testing, but I see that there are 5 years being presented in Figure S4 (and one of them is not FR-Gri2018,* *but FR-Gri2020). The authors should clarify this aspect. If FR-Gri2018 is in the calibration dataset, Figure 3 can be safely removed, and Figure S4 should take its place instead (with all panels).*

[Response] We apologize for the confusion caused by the inconsistent labeling. FR-Gri in 2020, not FR-Gri in 2018, belongs to the testing dataset. We have corrected this mistake by updating Figure 3, replacing the FR-Gri-2018 panel with the corresponding FR-Gri-2020 results.

Style comments:

***16.** L22: The sentence "However, the latter aspect is not yet covered by existing cropping system models" sounds a bit too final. I don't have enough knowledge on the current state of crop modeling regarding this feature, but I suggest that the authors leave this sentence only if they are 100% sure about it.*

[Response] We agree that the original sentence was overly definitive. In the revised manuscript, we now rephrase this statement to avoid implying a comprehensive assessment of all existing cropping system models:
Line 22: "However, the latter aspect **remains only partially represented in current** cropping system models."

***17.** L56-57: there's no need to start a new paragraph here*

[Response] The sentence is now incorporated into the preceding paragraph.

***18.** L70: "Winter wheat … reflectivity". This sentence doesn't sound right, as it suggests causality and intentionality where there is none. A sentence that simply describes the species would be enough.*

[Response] The sentence has been revised to provide a neutral description of winter wheat phenology and associated changes in canopy reflectance: **Line 70-71**: "Winter wheat, as a major cereal crop, undergoes distinct phenological stages **that are accompanied by changes in canopy pigmentation and surface reflectance** (Zhang et al., 2022)."

***19.** L150-152: Table S2 is mentioned before Table S1. The authors should switch the tables in the SI so that the ordering is correct.*

[Response] The tables in the Supplementary Information have been renumbered so that Table S1 now appears before Table S2, and we have also corrected the table citation in the main text.

***20.** L166-167: What does "but not S2" mean? Does it refer to site measurements?*

[Response] The phrase "but not S2" refers to Sentinel-2-derived LAI products, which were not used because their temporal coverage prior to 2018 is insufficient to match the available site observations. We have revised the manuscript to make this clear:
Line 173-174: "We used **MODIS-derived** but not **S2-derived LAI products, because the temporal coverage of the latter is insufficient to match the available site observations.**"

21. *L281-284: Equation formatting needs to be improved*

[Response] All equations have been reformatted in the revised manuscript to improve clarity and consistency.

Line 295-296:

$$\begin{cases} \alpha_{res}\,(t) = \alpha_{yellowing,start} \times (k_1 \times t + k_2) & t\ in\ albedo\ increase\ phase; \\ \alpha_{res}\,(t) = \alpha_{max} \times \left(k_3 \times e^{-k_4 \times t}\right) & t\ in\ albedo\ decrease\ phase; \\ \alpha_{res}\,(t) = 0 & otherwise; \end{cases} \quad (14)$$

22. *L304: Equation formatting needs to be improved*

[Response] All equations have been reformatted in the revised manuscript to improve clarity and consistency.

Line 302-303:

$$\begin{cases} f_{res}(t) = f_{crop,yellowing} & t\ in\ albedo\ increase\ phase; \\ f_{res}(t) = f_{crop,yellowing} \times \left(1 - \frac{t}{RC}\right) & t\ in\ albedo\ decrease\ phase; \\ f_{res}\,(t) = 0 & otherwise; \end{cases} \quad (15)$$

23. *L311: "between these two phases" instead of "between them"?*

[Response] We have revised this sentence according to the suggestion:

Line 337-338: "At the site scale, we assumed that foliar yellowing (i.e., α_{surf} increase) begins when maturity is reached due to the absence of observational data supporting a temporal gap **between these two phases.**"

24. *L330: not sure if "parametrized" is the right term. Do the authors mean that the new values were included in the new ORCHIDEE-CROP version?*

[Response] To clarify, the calibrated values of k_1, k_2, k_3 and k_4 were incorporated into the updated ORCHIDEE-CROP model to represent albedo dynamics affected by foliar yellowing and residue cover. We have revised the manuscript to replace "parameterized" with "incorporated" to avoid the potential ambiguity:

Line 357: "Finally, calibrated k_1, k_2, k_3 and k_4 were **incorporated into** the ORCHIDEE-CROP model to simulate the albedo dynamics affected by foliar yellowing and residue covering for all ten sites."

25. *L337: what does "-D" stands for here? What about the other acronyms? Please specify*

[Response] While the different model configurations were described, the meaning of the acronyms was not explicitly stated. We have revised the manuscript to clearly define the naming convention for all configurations, explicitly indicating what each suffix (A, E, Z_0, β_4, and D) refers to.

Line 363-367: "In the following, we **refer to ORC-D as the default ORCHIDEE-CROP model configuration prior to any modifications, where "D" denotes the default setup.** To test the respective impacts of the five model modifications, we used five different configurations of the improved model. Configuration ORC-AE **includes modifications of** α_{surf}, β_4 and Z_0 during foliar yellowing and residue covering periods, **where "A" refers to surface albedo and "E" to energy-related parameters.**"

26. *L341: maybe use ORC-beta_4 instead of just ORC-beta*

[Response] We replaced ORC-β_4 with ORC-β consistently throughout the manuscript.

27. L370: y overbar is missing

[Response] We have corrected the explanation of equation (20): **Line 402**: "$\bar{y}$ is the mean of the observed data."

Figures and tables:

28. *Figure S2. In the supplementary, there are no limitations. The authors are invited to show more of the years and sites available (or, even better, a summarized evaluation)*

[Response] According to the management information available to us, continuous in situ LAI measurements for winter wheat are not available at all analyzed sites. Only FR-Aur, FR-Lam, and FR-Gri provide sporadic LAI observations, with limited temporal coverage. For example, at FR-Aur, an average of approximately nine LAI measurements per year is available for five individual years (2006, 2008, 2010, 2012, and 2014).

We initially considered using boxplots as a summarized evaluation of LAI differences between MODIS and site observations, as suggested by the reviewer. However, the primary objective of our LAI analysis is to examine the temporal evolution of LAI, rather than its aggregated distribution. Boxplots have no temporal information, so we thought they were not suitable for identifying phenological transition points and chose to retain the time-series representation.

We have revised Figure S2 to include LAI dynamics from both MODIS and site observations for all years in which in situ measurements are available at each site (see comment #3).

29. *Table S1. Wrong date format for tillage*

[Response] We have revised the date format for tillage to ensure consistency.

30. *Table 1. This table needs to be simplified greatly. Authors should indicate with just a tick whether there are modifications with respect to the original version or not, and then add in the caption or in footnotes what these changes mean only if they differ with what is described in the methods text.*

[Response] We have simplified Table 1 following the reviewer's suggestion. The table now only indicates whether a given parameter is modified relative to the control simulation using '-' or '√'. Detailed descriptions of the parameter changes were explained in section 2.5.

Table 1. Simulation experiments

Names of experiments	**Periods**	**Surface albedo (α_{surf})**	**Parameter of soil conductance (β_4)**	**Surface roughness (Z_0)**	**Crop & residue fraction (f_{crop}, f_{res})**	**Test purpose**
ORC-D		-	-	-	-	Control simulation
ORC-AE	Foliar yellowing	√	-	-	√	Combined impacts of surface albedo, soil resistance and surface roughness
	Residue covering	√	√	√	√	
ORC-A	Foliar yellowing	√	-	-	√	Isolated impact of surface albedo
	Residue covering	√	-	-	√	
ORC-E	Foliar yellowing	-	-	-	√	Combined impact of soil resistance and surface roughness
	Residue covering	-	√	√	√	
ORC-β_4	Foliar yellowing	-	-	-	√	Isolated impact of soil resistance
	Residue covering	-	√	-	√	
ORC-Z_0	Foliar yellowing	-	-	-	√	Isolated impact of surface roughness

Residue covering	-	-	√	√

* '-' means there is no change, whereas '√' denotes that the corresponding parameter is modified relative to ORC-D;

*The temporal evolution of foliar and residue albedo, soil resistance parameter, surface roughness, as well as changes in crop and residue fractions, follows the parameterizations described in Section 2.3, unless stated otherwise.

Reference

Cescatti, A., Marcolla, B., Santhana Vannan, S. K., Pan, J. Y., Román, M. O., Yang, X., Ciais, P., Cook, R. B., Law, B. E., Matteucci, G., Migliavacca, M., Moors, E., Richardson, A. D., Seufert, G., and Schaaf, C. B.: Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network, Remote Sens. Environ., 121, 323-334, https://doi.org/10.1016/j.rse.2012.02.019, 2012.

Eichler, S. E.: Portable apparatus for high spatial and temporal resolution of in situ real-time surface albedo measurement in agricultural fields, Environ. Sci. Adv., 1, 297-304, https://doi-org.insu.bib.cnrs.fr/10.1039/d1va00051a, 2022.

Gava, R., de Freitas, P. S. L., de Faria, R. T., Rezende, R., and Frizzone, J. A.: Soil water evaporation under densities of coverage with vegetable residue, Eng. Agric., 33, 116-127, https://doi.org/10.1590/S0100-69162013000100010, 2013.

Huang, Y., Chen, L., Fu, B., Huang, Z., and Gong, J.: The wheat yields and water-use efficiency in the Loess Plateau: straw mulch and irrigation effects, Agric. Water Manage., 72, 209-222, https://doi.org/10.1016/j.agwat.2004.09.012, 2005.

Pastorello, G., Trotta, C., Canfora, E., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci. Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Li, S. X., Wang, Z. H., Li, S. Q., Gao, Y. J., and Tian, X. H.: Effect of plastic sheet mulch, wheat straw mulch, and maize growth on water loss by evaporation in dryland areas of China, Agric. Water Manage., 116, 39-49, https://doi.org/10.1016/j.agwat.2012.10.004, 2013.

Li, H., Zhang, Y., Sun, Y., Zhang, Q., Liu, P., Wang, X., Li, J., and Wang, R.: No-tillage with straw mulching improved grain yield by reducing soil water evaporation in the fallow period: A 12-year study on the Loess Plateau, Soil Tillage Res., 224, 105504, https://doi.org/10.1016/j.still.2022.105504, 2022.

Li, Y., Liu, H., and Huang, G.: Modeling resistance of soil evaporation and soil evaporation under straw mulching, T. Chinese Soc. Agric. Eng., 31, 98-106, https://doi.org/10.3969/j.issn.1002-6819.2015.01.015, 2015.

Su, Y., Lauerwald, R., Makowski, D., Viovy, N., Guilpart, N., Zhu, P., Gabrielle, B., and Ciais, P.: Future warming increases the chance of success of maize–wheat double cropping in Europe, Eur. J. Agron., 170, 127723, https://doi-org.insu.bib.cnrs.fr/10.1016/j.eja.2025.127723, 2025.

Tennekes, H.: The logarithmic wind profile, J. Atmos. Sci., 30, 234-238, https://doi.org/10.1175/1520-0469(1973)030<0234:TLWP>2.0.CO;2, 1973.

Wang, S., Lei, L., Gao, Y., and Zhan, E.: Effects of Straw Decomposition on Soil Surface Evaporation Resistance and Evaporation Simulation, Plants, 14, 434, https://doi.org/10.3390/plants14030434, 2025.

Wang, J., Zhang, Y., Gong, S., Xu, D., Juan, S., and Zhao, Y.: Evapotranspiration, crop coefficient and yield for drip-irrigated winter wheat with straw mulching in North China Plain, Field Crops Res., 217, 218-228, https://doi.org/10.1016/j.fcr.2017.08.021, 2018.

Xing, X., Li, Y., and Ma, X.: Effects on Infiltration and Evaporation When Adding Rapeseed-Oil Residue or Wheat Straw to a Loam Soil, Water, 9, 700, https://doi.org/10.3390/w9090700, 2017.

Yusefi, A., Farrokhian Firouzi, A., and Aminzadeh, M.: The effects of shallow saline groundwater on evaporation, soil moisture, and temperature distribution in the presence of straw mulch, Hydrol. Res., 51, 720-738, https://doi.org/10.2166/nh.2020.010, 2020.

Yunusa, I. A. M., Sedgley, R. H., and Siddique, K. M. H.: Influence of mulching on the pattern of growth and water use by spring wheat and moisture storage on a fine textured soil, Plant Soil, 160, 119-130, https://doi.org/10.1007/BF00010145, 1994.

Zhang, X., Chen, S., Liu, M., Pei, D. and Sun, H. Improved Water Use Efficiency Associated with Cultivars and Agronomic Management in the North China Plain. Agron. J., 97: 783-790. https://doi-org.insu.bib.cnrs.fr/10.2134/agronj2004.0194, 2005.

Zhang, S., Yang, X., and Lovdahl, L.: Soil Management Practice Effect on Water Balance of a Dryland Soil during Fallow Period on the Loess Plateau of China, Soil Water Res., 11, 64–73, https://doi.org/10.17221/15/2015-SWR, 2016.

Zhang, Y., Ma, J., & Cao, Z.: The von Kármán constant retrieved from CASES-97 dataset using a variational method. Atmospheric Chemistry and Physics, 8, 7045-7053, https://doi.org/10.5194/acp-8-7045-2008, 2008.