Reviewer 1

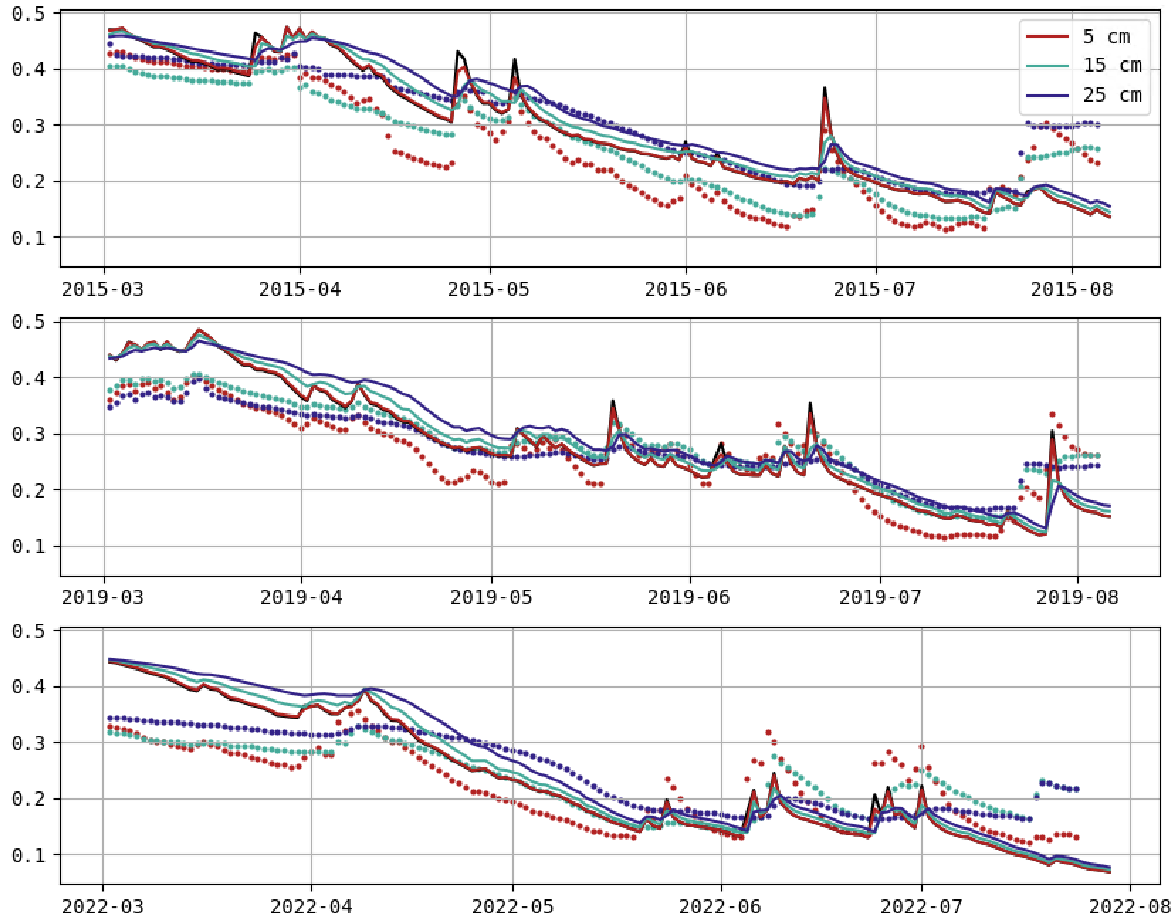
In my opinion, the main limitation of the study is the lack of data on soil water dynamics in different soil layers. It would have been very informative to evaluate how well the model simulated soil drying, both due to root water uptake and soil evaporation. Although different combinations of soil physical properties were tested in the sensitivity analysis and some root parameters were calibrated, the selected parameter set may not have been optimal. Such data could also have helped to identify the causes of the discrepancies between simulated and observed evapotranspiration. This aspect is missing from the discussion. A similar issue applies to the lack of leaf area data, which is only represented indirectly through biomass measurements of the different plant organs.

We understand your concern on the lack of data on soil water dynamics and the estimation of soil physical properties. Regarding the soil physical properties, we made ku-pF measurements (with HYPROP and WP4C instruments) in order to get soil moisture curves parameters. The GSA was then conducted by considering the uncertainties around these estimations. As a result, most of these parameters were not influential, thus set to their measured values, and $K_{sat,1}$ was included in the calibration process with a range also determined by the measurement uncertainty.

This is a relevant remark to bring up the soil water content. We didn't add this perspective because we have multiple problems with the SWC captors. We have a problem during the 2017 season (captors were changed around this time), and also with the deeper measurements (55 cm and 85 cm). These were replaced recently (2023), but we unfortunately cannot rely on these measurements. So, we are left with the measurement at a 5-cm, 15-cm and 25-cm depth for the remaining seasons (see below figure). The dynamics is similar, but we need to some remarks and observations:

1. Observed SWC near saturation is lower than predicted SWC. This is expected as the SWC captors have a default calibration (only based on texture classification) which we think is not adapted. Actually, it's partly because of the observed profile of SWC near saturation that we decided to conduct a soil properties measurement campaign.
2. There seems to be a lack of reactivity in the surface layers compared to observations. It's consistent with the daily evolution of LE (Figure 4 from the manuscript) that showed lesser variability than observed. At this point, it is hard to tell if it's only due to the parameters or also to the structural limitations highlighted in the discussion. We will come back to it in the specific comments.
3. If we focus on April 2019, you are right to assume that the soil evaporation is higher than predicted, meaning that SWC decreases more rapidly. However, it only regards the 5-cm depth and has a total magnitude of 1% per day. If we consider that this loss represents the first 10 cm of the profile, this is approximately 26W/m^2. Therefore, it plays a part in the underestimation of LE by Daisy, but it doesn't explain the main difference.

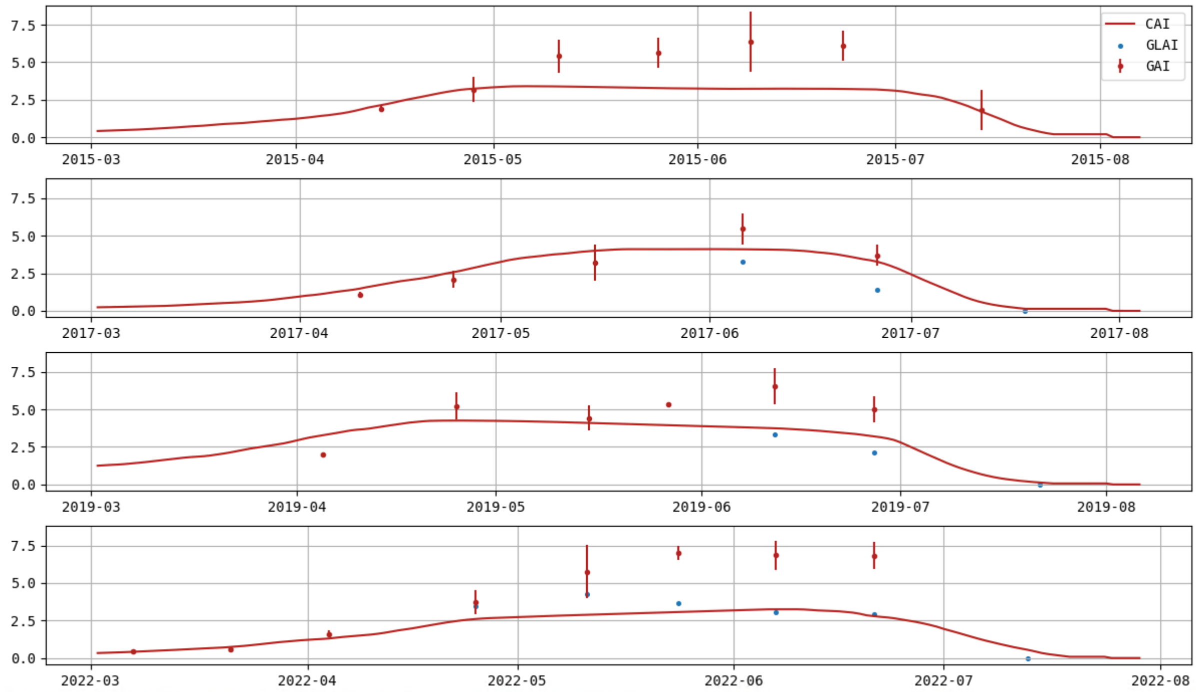
We can incorporate this last reflection in the section 3.3, when focusing on April 2019, as well as add the SWC of these three soil layers in figure 5.



For the LAI (called CAI after as it may include other organs), we also decided to not include this in the manuscript, in the interests of clarity/synthesis, but also because it's directly linked to biomass through the model equation. The trade-off of CAI-NEE is similar to the trade-off of DM-NEE. In the below figure, you have the predicted CAI (plot line) as well as the measured GAI (Green Leaf Area, including total stem and ear) and the measured leaf dry weight divided by the measured Leaf Mass Area (= 1/SLA), corresponding to the GLAI (Green Leaf Area Index). Depending on the real photosynthetic efficiency of the stem and storage organs, the predicted CAI should be closer to the blue points (no photosynthetic efficiency) or the red points (high efficiency). We can see that the CAI is more underestimated in the 2015 and 2022 seasons, when the biomass was also more underestimated. Similarly, the temporal evolution of the 2017 season is identical to the DM evolution, where both are slightly overestimated at the beginning of the season and underestimated at the end.

To answer this comment, we can include a sentence at the end of section 3.2:

"Although not showed, predictions Crop Area Index (CAI) follow a similar pattern, with greater underestimations near the end of the seasons."



Specific comments:

It would be relevant to mention the work of Delhez et al. (2025) already in the introduction. Currently, this study is cited only in the Materials and Methods section as the source of data, management and soil information, and the internal sensitivity analysis. However, it may be useful to clarify that the present study is a follow-up to that work, focusing on the calibration of the most relevant parameters identified previously. This could be stated in the final part of the introduction.

Line 55: It could be useful to explain why Daisy was selected for this study compared to the other models mentioned earlier (lines 25–30). It is unclear whether Daisy is unique in coupling Richards' equation for soil water dynamics with a Farquhar-based photosynthesis model, or whether the choice was driven by other factors. For example, Daisy being open source and implemented in C++ may have facilitated the implementation of the Pareto-based modelling framework.

Mentioning the previous GSA and the reasons for using Daisy are a good suggestion. Indeed, that's a choice driven by multiple factors. The most important was the integration of a Farquhar-based photosynthesis model as well as the computation of the surface energy budget (within the SVAT module), in order to have coupled carbon, water and energy fluxes. In that regard, Daisy is similar to GECROS, SiBcrop, ORCHIDEE-STICS, etc. Richard's equation and being open source were indeed the additional criteria that finally lead to Daisy (being implemented in C++ was not really considered).

To account for your comments and the second reviewer's, we suggest updating the last paragraph of the introduction as follows:

"In this case study, five growing seasons of winter wheat, cultivated on a Belgian site, were simulated using Daisy soil–plant–atmosphere model (Hansen et al., 2012). In addition to the stomatal coupling implemented with Farquhar biochemical model (Plauborg et al., 2010), this open-source model explicitly simulates the surface energy budget. Soil water movement is

described by Richards' equation while accounting for drainage pipes, which are widely installed in agricultural fields. Together, these features make Daisy suitable for studying crop yield as well as coupled carbon and water flux dynamics under current and future conditions.

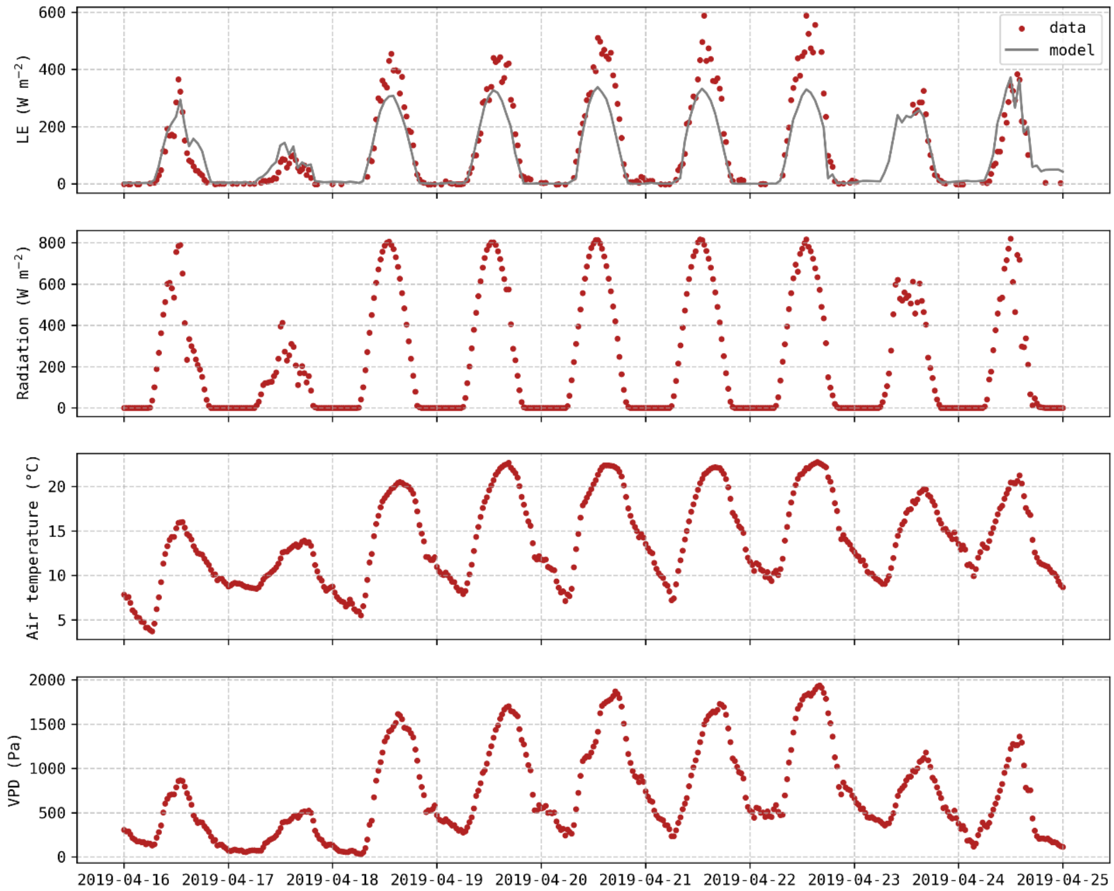
In a previous paper, we have identified key parameters controlling crop yield, energy and carbon fluxes and discussed the implications of contrasting water conditions on sensitivity analysis results (Delhez et al., 2025). The objective of this follow-up study was twofold: to improve the simulation of biomass, carbon and water fluxes through calibration and to explore the trade-offs between these outputs. To this end, we applied a swarm-based MOO algorithm called Speed-constrained Multi-objective Particle Swarm Optimisation (SMPSO) targeting (i) dry matter of vegetation organs (DM), (ii) Net CO2 Ecosystem Exchange (NEE) and (iii) latent heat flux (LE). Accurate simulation of these variables is essential for food production assessments and for quantifying crop carbon and water budgets. Furthermore, analysing trade-offs, particularly between carbon (NEE) and water (LE) fluxes, could reveal potential limitations in their coupled representation within the model."

Lines 90–95: Is the Daisy setup similar to that used in Delhez et al. (2025)? If so, it may be relevant to mention that a drainage system based on the Hooghoudt equation is included. Alternatively, this information could simply be referenced to Delhez et al. (2025), where it is described. Otherwise, there is a risk of soil profile flooding when an aquitard layer is present and no drainage system is implemented.

Indeed, this is the same setup, we will mention that and add a reference to the previous study.

Figure 5. A latent heat flux of 600 w/m2 in late April seems extreme, since the theoretical potential clear sky radiation is 750 w/m2. Would have been relevant to see the data on air temperature, humidity, and wind speed during this event.

We were also surprised by the high observed values at first, but this consistent with weather conditions (high air temperature, VPD and radiation). There was no technical problem reported. Including these data in the Figure 5 might make this figure too busy, but we can add an additional figure (see below) in the appendix C, where the computation of the stomatal conductance during this period is explained. Radiation is the incoming shortwave radiation; this will be specified in the figure legend.



Section 4.2: The issue of calibrating the model separately for each cultivar grown in only one season is not discussed. For instance, the large differences observed in some parameters may be difficult to explain purely on a genetic basis. The large range in parameters such as *SOrgPhotEff* and *stemPhotEff* could potentially be caused by seasonal stress factors—such as disease or water or nitrogen stress—not explicitly represented in the model, rather than by genetic differences. It might have been more robust to calibrate a single cultivar across all seasons, given that modern wheat cultivars generally do not differ substantially in yield potential or growth patterns.

We agree that some parameter might be different not just because of genetic, but because of other factors. For instance, we know that $V_{c,max}$ can be impacted by water stress, which is not taken into account in the model, but also has genetic variation for wheat (https://doi.org/10.1093/jxb/erx421, https://doi.org/10.1093/jxb/eraa077). Similarly, ShldResC can also impacted by genotype and water stress (https://doi.org/10.2135/cropsci2006.01.0013). As the observed difference between cultivars can be due to a combination of factors (GxE), and as some factors are not explicitly considered in the model, we decided to take them as cultivar specific. One of our goals was to understand model limitations, thus by considering them as cultivar-specific, we can identify which parameters or processes must be better described in the future. This is apparently the case for StemPhotEff and SOrgPhotEff.

We didn't discuss the calibrated values of the parameters, nor the difference between the cultivars to keep the manuscript concise and focus on the trade-offs. However, we can add a paragraph about this reflection in section 4.1, between the two existing paragraphs:

“The calibrated parameters are all within the explored range, without relying on extreme values (Table B1), indicating a physically plausible solution. It is, however, worth noting that some cultivar-specific parameters have very different values between seasons, such as δ describing the hydraulic signalling for stomatal conductance or the photosynthetic efficiency of non-leaf organs ($SOrgPhotEff$ and $StemPhotEff$). Such differences may reflect genotypic variability, but they may also be due to environmental factors not represented in the model. This is, for example, the case for the remobilisation reserve $ShldResC$ which can be impacted by genotype and water stress (Ehdaie et al., 2006). The absence or misrepresentation of environmental factors in the model can be compensated by wider variations between cultivars. Depending on the modeller’s objective, these parameters might be kept as cultivar-specific while investigating the environmental influences, or constrained across cultivars to improve model parsimony and robustness.”

Lines 330–340: It could be added that models based on Richards’ equation tend to overestimate soil evaporation. One reason is the difficulty in obtaining accurate hydraulic parameters for the surface soil layers, which is further complicated by soil water hysteresis. As a result, the hydraulic conductivity curve used to calculate potential matrix exfiltration may be too high, leading to an overestimation of soil evaporation in Daisy. This issue will also be relevant for a model with a more mechanistic coupling between the surface energy and water balance, simulating microclimate effects on soil EP as a fully coupled approach.

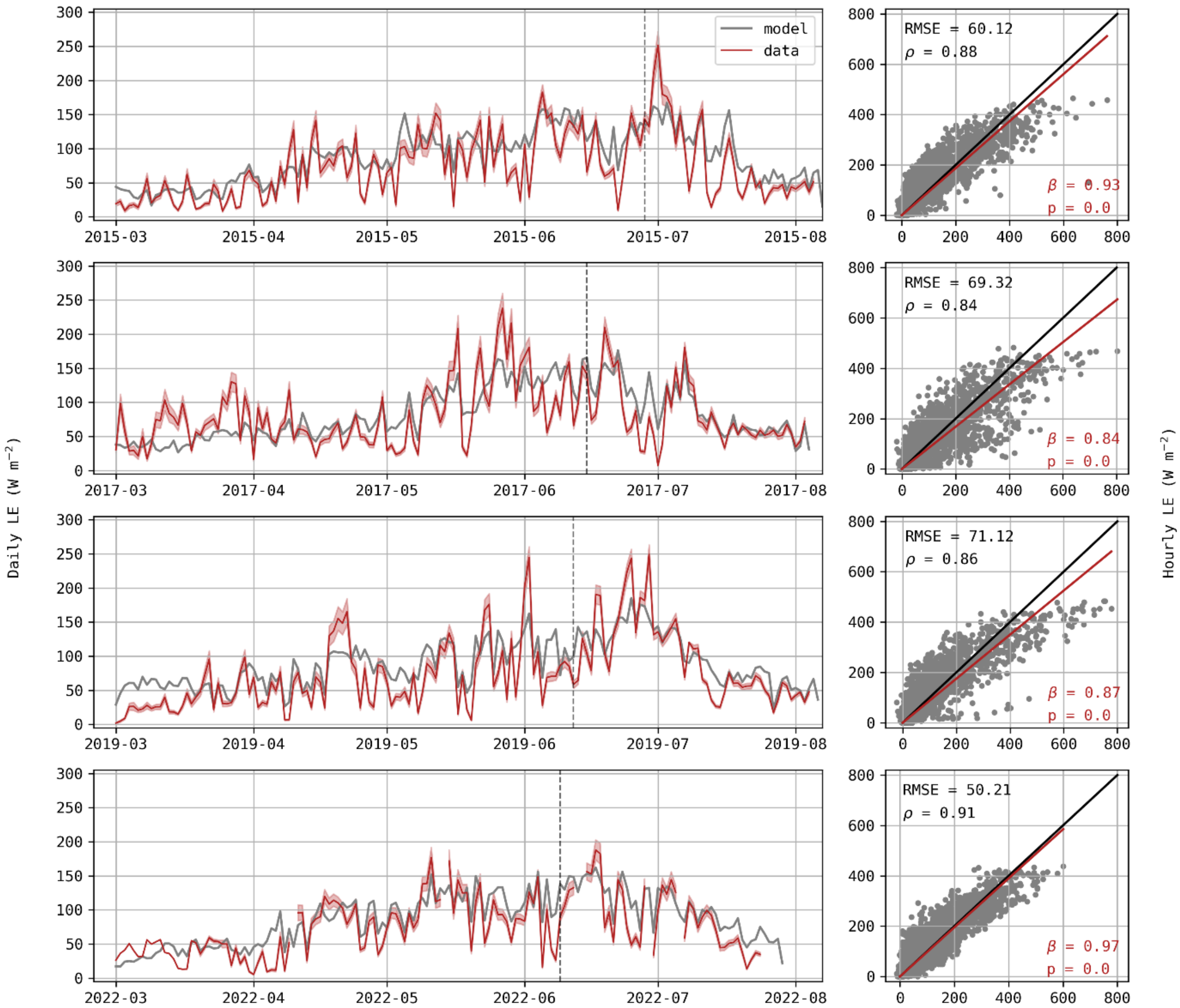
Thank you for that valid remark. Now that you can look at the soil water content at different depths, you can see that predicted soil evaporation might be more often underestimated rather than overestimated.

Lines 330–340 (continued): Daisy also includes a transfer function controlled by the *EpInterchange* coefficient, which allows energy transfer from a dry soil surface to the canopy (default value β = 0.6 [–]). This function could potentially explain the relatively high simulated latent heat flux (LE) during periods of low leaf area at the beginning and end of the time series. In theory, this parameter could convert some soil water into LE under dry surface conditions. The parameter was not included among the 200 parameters in the initial sensitivity analysis by Delhez et al. (2025), and its omission may have resulted in an overestimation of transpiration as a starting point for the SSOC iterations. Again, this interpretation assumes that Daisy uses accurate hydraulic conductivity curves for the surface layers, as discussed above.

You are right, we did not take EpInterchange into account in the GSA, but we should have. We were really confused about the physical meaning behind this process. However, we just checked the soil evaporation outputs, and soil_ea was always equal to soil_ep (except for about ten time steps where the difference is equal to 1e-17, so we assume it comes from a numerical approximation). Although surprising, it means that the energy transfer and thus EpInterchange were never solicited.

If soil_ep is aways reached by soil_ea, it also means that soil water movement is not limiting for the upward flux. Thus, the observed quick decreases in SWC at 5-cm depth might not be due to misparameterisation, but wrong estimation of soil_ep. Regarding the downward flux, however, we cannot tell if that’s an issue due to soil parameters, leaf interception, or model limitations.

When focusing on the period you highlighted, especially at the beginning of March 2019 where the difference is the biggest, we saw that the hourly observations were not as low as it appeared to be when looking at daily values. It seems that some data were discarded due to a data flag, so it impacted the daily-aggregated LE as well as the uncertainties. We corrected the daily observations, but note that it doesn't change the calibration results, as we calibrated against hourly LE. Only the figure 4 (left, with daily dynamics) needs to be updated, this doesn't apply to the figure 5 as it's hourly observations, and neither to the NEE results. We thank you for this remark, as it would have gone unnoticed otherwise! You can find the updated figure right below. The difference in March 2019 is now lower, but still there. When checking the outputs, soil evaporation is the main part of total evapotranspiration, so soil_ep might be overestimated during this period.



Technical corrections

Line 70: The following sentence is confusing: "(Meza et al., 2018; 2023). As the same cultivar was sown for VAL and SAH, the VAL season was set aside for validation." This is unclear. Instead of naming growing seasons after cultivars, it might be clearer to refer to them by year.

We can denote them as S11, S15, S17, S19 and S22. All the manuscript will be updated accordingly.

Furthermore, as noted above, it is not clearly stated whether a specific calibration was performed for each cultivar/season or whether a single parameterised cultivar was used across all seasons in the text. This information is only apparent from Table B1 in Appendix B.

This can be understood from Table 1 and 2, as we have different cultivars (Table 1) and cultivar-specific parameters are denoted by 'a' in Table 2, with an explanation below the table. However, to avoid any misunderstanding, we can add this information in the main text, and, in section 2.3.3, we can explicitly mention it:

"For each parameter sets (i.e. potential solution) generated by SMPSO, the parameter values were transcribed into Daisy setup files. These include cultivar-specific parameters, with possibly different values for each season, and site-specific parameters that are shared (Table 2). Hence, the model was executed four times, specifically for S15, S17, S19 and S22. The three objective functions (rRMSE) were then aggregated and computed from these four runs and passed back to SMPSO algorithm."

Table 2: Not all parameters listed can be found in the Daisy documentation. For example, it is unclear what *k_net* refers to in the Daisy reference manual. It would be helpful to include the exact name from the setup files in this list as a separate column.

We specified the daisy name when the chosen name was different, but you are right, we missed k_net! It corresponds to EPext. As suggested, we will add another column, that would be clearer.

Reviewer 2

Specific comments

1. Scientific framing and justification of tool choice.

The introduction would benefit from a clearer articulation of the scientific questions motivating this work. Parts of the current manuscript read as if improving DAISY model performance is the primary objective, rather than a means to address broader bio-ecological questions. Explicitly clarifying what real-world processes or uncertainties this model improvement aims to resolve, and why DAISY is an appropriate tool for this purpose, would strengthen the scientific contribution and better justify the modelling effort.

In addition, the choice of output variables lacks clear justification. The importance and relevance of NEE and LE should be explained more explicitly. These concepts may be straightforward for the authors, but additional explanation would benefit general readers and clarify their scientific relevance.

To account for your remarks as well as the first reviewer's, we suggest updating the last paragraph of the introduction as follows:

"In this case study, five growing seasons of winter wheat, cultivated on a Belgian site, were simulated using Daisy soil–plant–atmosphere model (Hansen et al., 2012). In addition to the stomatal coupling implemented with Farquhar biochemical model (Plauborg et al., 2010), this open-source model explicitly simulates the surface energy budget. Soil water movement is described by Richards' equation while accounting for drainage pipes, which are widely installed in agricultural fields. Together, these features make Daisy suitable for studying crop yield as well as coupled carbon and water flux dynamics under current and future conditions.

In a previous paper, we have identified key parameters controlling crop yield, energy and carbon fluxes and discussed the implications of contrasting water conditions on sensitivity analysis results (Delhez et al., 2025). The objective of this follow-up study was twofold: to improve the simulation of biomass, carbon and water fluxes through calibration and to explore the trade-offs between these outputs. To this end, we applied a swarm-based MOO algorithm called Speed-constrained Multi-objective Particle Swarm Optimisation (SMPSO) targeting (i) dry matter of vegetation organs (DM), (ii) Net CO2 Ecosystem Exchange (NEE) and (iii) latent heat flux (LE). Accurate simulation of these variables is essential for food production assessments and for quantifying crop carbon and water budgets. Furthermore, analysing trade-offs, particularly between carbon (NEE) and water (LE) fluxes, could reveal potential limitations in their coupled representation within the model."

2. Model simplification, dynamic processes, and interpretation

The authors provide a clear and well-reasoned discussion of structural limitations of the current model (e.g. constant SLA). To strengthen this section, it would be helpful to clarify whether introducing additional dynamic processes is expected to significantly improve predictions (e.g. of LE and NEE) for answering the research questions, and what will be the trade-offs by making the model more complex.

You are right to bring that up. Having a dynamic SLA (or specific area of other organs) could help us to better fit biomass, for example, but would worsen NEE. If we increase SLA after anthesis, photosynthesis would be higher as well as dry matter. Therefore, dry matter predictions would more likely be closer to observations, but NEE would be even more negative. This was not an appropriate example. However, we think it's important to discuss the model assumptions, but also the lack of data leading to these assumptions. For instance, dynamic SLA is already implemented in Daisy, but we assumed it was constant as we have only one measurement per season.

To correct this part, we suggest combining this discussion with the previous paragraph about dry matter partitioning:

"Wrong modelling assumptions. *Modelling real-world system comes with simplifications. For instance, in this study, the partitioning schemes were derived from S22 since root measurements were only made during that season, as well as early distinction between leaf and stem. These partitioning coefficients were also applied to the other cultivars, but they might be inadequate and partly contribute to this trade-off. If too much assimilated carbon is allocated to the roots at the expense of the stem, the model would underestimate stem DM which is crucial for grain filling during the reproductive stage. Conversely, as leaves are not directly affected and are the major contributor to photosynthesis, the influence on NEE would be attenuated. Besides this example, many assumptions could play a role in these results, such as the organ death rates depending only on development stages or the non-dynamic leaf abscission rate. However, accounting for these within the model would be, in most cases, ineffective as model users hardly have enough data to support this level of complexity."*

3. Linking model limitations to real-world mechanisms (LE and NEE).

The discussion of limitations in predicting LE and NEE is thoughtful and balanced. It would be particularly insightful to briefly expand on how these diagnosed limitations relate to real-world mechanisms. For example, are the relevant processes already well understood from experimental studies but not yet parameterised in models, or do the results point to areas where targeted experiments or observational designs are needed? Even a short reflection along these lines would enhance the scientific relevance of the work.

We rephrased the section 4.3 to better reflect the advances in experimental studies and advances in modelling:

"Stomatal behaviour plays a central role in plant regulation of water loss and carbon uptake, and has therefore been extensively studied (Damour et al., 2010). While it is well established that mild edaphic drought induces a decrease in carbon assimilation by reducing the stomatal aperture (Beauclaire et al., 2024), plant responses to heatwaves and high atmospheric demand are less evident. Most studies report reduced g_s under high VPD (Grossiord et al., 2020; Bourbia and Brodribb, 2024), aligning with the theoretical predictions of stomatal models (Sabot et al., 2022). On the other hand, high temperatures seem to trigger stomatal opening to promote transpiration cooling (Urban et al., 2017), but reported results vary greatly among species (Moore et al., 2021). Under the combined effect of high temperature and VPD, Marchin et al. (2022) observed a significant increase in g_s for two well-watered species, interpreting this as a strategy to prevent leaf overheating via enhanced LE loss without any increase in carbon assimilation. They later confirmed that this behaviour, referred to as stomatal decoupling, is not restricted to well-watered species (Marchin et al., 2023).

This stomatal decoupling aligns with ecophysiological studies pointing out the role of non-stomatal factors during photosynthesis. Inside the leaf, CO_2 diffuses through air spaces and membranes to reach the sites of carboxylation inside chloroplasts, whereas H_2O moves from the *xylem network to the stomata (Flexas et al., 2012; Sack and Holbrook, 2006). These different pathways can be affected independently, where carbon assimilation can be restricted by non-stomatal factors, often dominated by mesophyll conductance (Flexas et al., 2012). Over the last decade, these non-stomatal limitations have received increased attention, especially under water-stressed conditions (Nadal and Flexas, 2018; Gago et al., 2020). Based on leaf gas exchange measurements, experimental studies proved that these limitations should be considered in leaf models linking photosynthesis and transpiration (Yang et al., 2019; Beauclaire et al., 2024; Chen et al., 2025). Despite that, these complex processes are still not fully understood, and hence barely integrated into crop or terrestrial models (Vidale et al., 2021). As an example, an empirical equation for the mesophyll conductance was implemented in the land surface model JSBACH (Knauer et al., 2019a). The authors extensively discussed the environmental drivers, the modelling implications as well as the effects on the predicted fluxes in a following paper (Knauer et al., 2019b).*

These non-stomatal limitations, or stomatal decoupling, are not modelled in Daisy. Based on stomatal coupling (Eq. 1), the increase in VPD reduced the predicted slope between carbon assimilation and g_s. Consequently, the model tended to underestimate g_s and LE fluxes and overestimated sensible heat, leading to a misrepresentation of energy partitioning under such conditions. On the other hand, observations show that plants likely opened their stomata, promoting water loss, while carbon assimilation did not increase accordingly. This stomatal decoupling was particularly visible in April 2019, as the heatwave lasted several days, but also appeared during shorter warm periods, suggesting a recurrent plant strategy. Future work should (i) continue the effort in better understanding the mechanisms of non-stomatal limitations affecting carbon assimilation and (ii) incorporate these mechanisms in process-based models, particularly under such conditions."

4. Clarity and presentation

Overall presentation is good, but several instances of vague wording (e.g. "good accuracy"), long sentences, and **undefined acronyms** reduce clarity. Addressing these issues would improve readability.

We have sent it to be edited (English native speakers) for lengthy phrasing and global clarity. Besides CLM as undefined acronym (please refer to our below response for this particular acronym), they didn't find any.

Technical corrections

Abstract

Line 12: "good accuracy" is vague. The term "good" lacks context for readers unfamiliar with the subject. Consider removing it or replacing it with quantitative information.

Is it because readers might be unfamiliar with magnitudes of DM, NEE and LE, and thus the RMSE values are hard to interpret? Or do you think RMSE values are not an indicator for the specific term of accuracy? Depending on your answer, we can replace good accuracy by "satisfactory results" or add rRMSE in addition to RMSE values.

In the meantime, we used the first solution, please do not hesitate to reach out for this.

Line 14: "wrong parameterisation" is a strong claim. The discussion suggests that simplification may contribute to bias, but no quantitative evidence is provided to demonstrate that the parameterisation is incorrect. Model simplification is a design choice; the issue is appropriateness rather than right or wrong. Rephrasing would improve accuracy.

Wrong parameterisation doesn't only refer to the simplifications, but also the underestimation of heterotrophic respiration by Daisy. In section 4.2, we can add this sentence at the end of the corresponding paragraph to make it clearer: "Adjusting soil respiration parameters would require long-term simulations and respiration data from this study site."

Introduction

Line 35: The sentence introducing g_1 is abrupt. It moves directly from physiology to a specific parameter without transition. A brief link from physiological concepts to their model formulation would improve readability.

Indeed, we can add a generic equation linking carbon assimilation (An) and stomatal conductance (gs) as follows:

"In these couplings, carbon and water fluxes between vegetation and atmosphere are assumed to only be regulated by stomatal behaviour. Widely used models of stomatal conductance g_s can be generally expressed as:

$$g_s = g_0 + g_1 \cdot f(E) \cdot A_{net}$$

Where A_{net} is the net leaf CO2 assimilation rate, $f(E)$ is a function of environmental influences such as Vapour Pressure Deficit (VPD) and CO2 concentration at the leaf surface, and g_0 and g_1 are parameters. When water is not limiting, plants open their stomata in order to maximise their carbon uptake. Conversely, they reduce their stomatal aperture under water stressed conditions, compromising between carbon assimilation and water loss (Manzoni et al., 2011). This response is often represented by a decrease of the slope parameter g_1(Zhou et al., 2013; Liu et al., 2009)."

The Equation numbers would be updated for the entire manuscript and the variable description in lines 101-102 will be reduced, as some have already been introduced in this new equation.

Line 51: The term equifinality is introduced without explanation. A brief definition here, similar to the one later in the discussion, would improve accessibility.

Good point, we will add a similar explanation.

Line 57: NEE and LE are introduced without justification. Please briefly clarify why these outputs were selected and what ecological or biophysical significance they represent.

Lines 58–59: The flow may improve if concepts are introduced in a consistent order (e.g. DM, then NEE, then LE), rather than switching between topics.

We took your two comments into account when rephrasing the last part of the introduction (see at the beginning of this document).

Methods

Line 69: TOB, SKY, SAH, and SMA are unclear. Please clarify whether these refer to varieties or growing seasons. If they represent seasons, it may be clearer to use digits.

We can denote them as S11, S15, S17, S19 and S22. All the manuscript will be updated accordingly.

Line 102: Unsure whether g_1 here refers to the same parameter introduced in the introduction.

Yes it is, it should be clearer with the generic equation in the introduction. We can also rephrase lines 99-100 as: "…a stomatal conductance empirical model (Eq. 1). In this version, the stomatal conductance gs is based on Leuning (1995) and adjusted to account for hydraulic signalling (Plauborg et al., 2010):"

Line 104: It would be beneficial to clarify why model modifications were required.

This is specified for the diffuse radiation in the appendix A (considering cloudy skies), but not explicitly for the others, we can add it in the appendix as well, or do you think in the main text would be better?

In the meantime, we added in the appendix.

Line 136: The description of "individual exploration" and "social learning" is unclear. Does this refer to local versus global optima?

These terms are defined and commonly used with particle swarm-based techniques. We agree that it might be unclear for unfamiliar users, we can add a reference to local and global optima between brackets, rather than "(its own experience)".

Line 154: The description of "once for each growing season" seems to suggest all seasons were used.

We rephrased this to account for your comment and the first reviewer's: *"For each parameter sets (i.e. potential solution) generated by SMPSO, the parameter values were transcribed into Daisy setup files. These include cultivar-specific parameters, with possibly different values for each season, and site-specific parameters that are shared (Table 2). Hence, the model was executed four times, specifically for S15, S17, S19 and S22. The three objective functions (rRMSE) were then aggregated and computed from these four runs and passed back to SMPSO algorithm."*

Line 157: Please clarify how swarm size and total simulation number were chosen. Were these based on empirical experience or hyperparameter tuning? Showing convergence behaviour (e.g. in an appendix) would increase confidence.

The number of simulations was chosen based on the convergence. We can indeed add the convergence behaviour in appendices. Regarding the swarm size, it was more empirical. Common values are around 100, but after looking at the LE-NEE trade-off, which was really narrow, we increased it to have a better spread and be sure that the shape is not due to local optima. This increases calculation time but improves the exploration.

Line 165: Please clarify what is meant by EF being "easily interpreted." Is this relative to other metrics, or related to its physical meaning for the output variables?

This is because EF has no dimension and the threshold for considering a result acceptable is known. We can write the "which is easily interpreted" at the end of line 169.

Results

Line 176: "Most solutions" is vague. Please consider reporting this as a percentage.

With this sentence, we are not talking about the black solutions but the grey ones, saying that grey solutions don't diverge far from the black solutions (contrary to LE-NEE). We understand that it can be confusing, and if we were talking about the black solutions, we could have reported a percentage. We can rephase this as:

"…where the curved Pareto front extends over a large part of the objective space. This 2D front portrays 29% of the solutions from Fig. 1a, and few other solutions (grey points) diverge far from this front."

Line 202: "Good overall agreement" lacks quantitative support. Please provide numerical values or remove the qualitative descriptor.

Numerical values are provided for hourly NEE, as we calibrate against these data, but not for daily values. We can remove this part.

"As depicted in Fig. 3, simulated temporal variations generally follow observed trends, although predictions tend to be more negative (i.e. more CO2 uptake) during the vegetative stage."

Line 204: Figure 3 also shows underestimation toward the end of the season; this could be mentioned explicitly.

Indeed, it is relevant. When reaching maturity, Daisy stops the plant processes. Therefore, there is no longer any autotrophic respiration until harvest, which could explain this difference. We will add a few lines about this.

Lines 245 and 253: Table numbering appears inconsistent. Table 1a and 1b may need to be Table 3a and 3b?

Indeed, thank you for the comment!

Discussion

Line 301: The statement that constant values may lead to significant bias is plausible, but quantitative evidence is not presented. Please clarify whether incorporating dynamic SLA (e.g. dependent on leaf age or water status) is expected to substantially improve predictions, or whether this remains a hypothesis.

This is answered in detail under one of your specific comments (3).

Lines 303–305: The sentence describing EC data filtering and uncertainty estimation could be rewritten for clarity.

This was rewritten, we hope this is clearer.

"To account for these, EC data are thoroughly filtered, resulting in low data coverage as in Table 1. Moreover, turbulence conditions are accounted for in the uncertainty estimation rather than affecting the data itself, with an uncertainty related to the friction velocity, an indicator of turbulence strength (Pastorello et al., 2020)."

Line 335: CLM is introduced without prior explanation. Please define it when first mentioned.

As for STICS or JULES, that's a model name commonly accepted so we didn't describe the acronym. "such as the CLM model" might make it clearer?

Line 374: Please rewrite for clarity. It is unclear what is meant by "This decoupling behaviour." Does this refer to increased transpiration while assimilation remains unchanged? Also, please clarify whether "April 19" refers to a date April 19th or the year 2019.

This is taken into account in the updated paragraph, in response to one of your specific comments.