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 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.
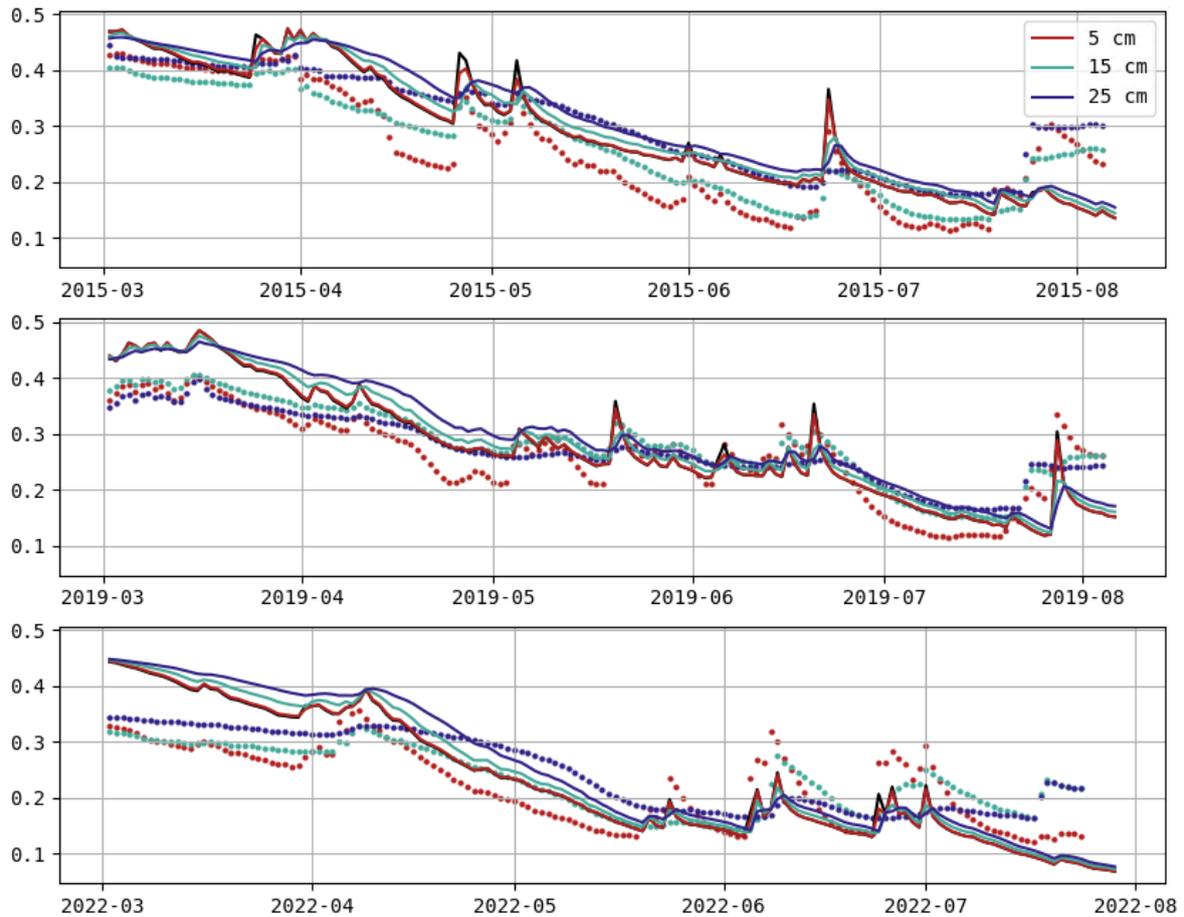
For the LAI (called CAI after as it may include other organs), we also decided to not include this in the manuscript, in the interests of clarity/synthesis, but also because it's directly linked to biomass through the model equation. The trade-off of CAI-NEE is similar to the trade-off of DM-NEE. In 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."
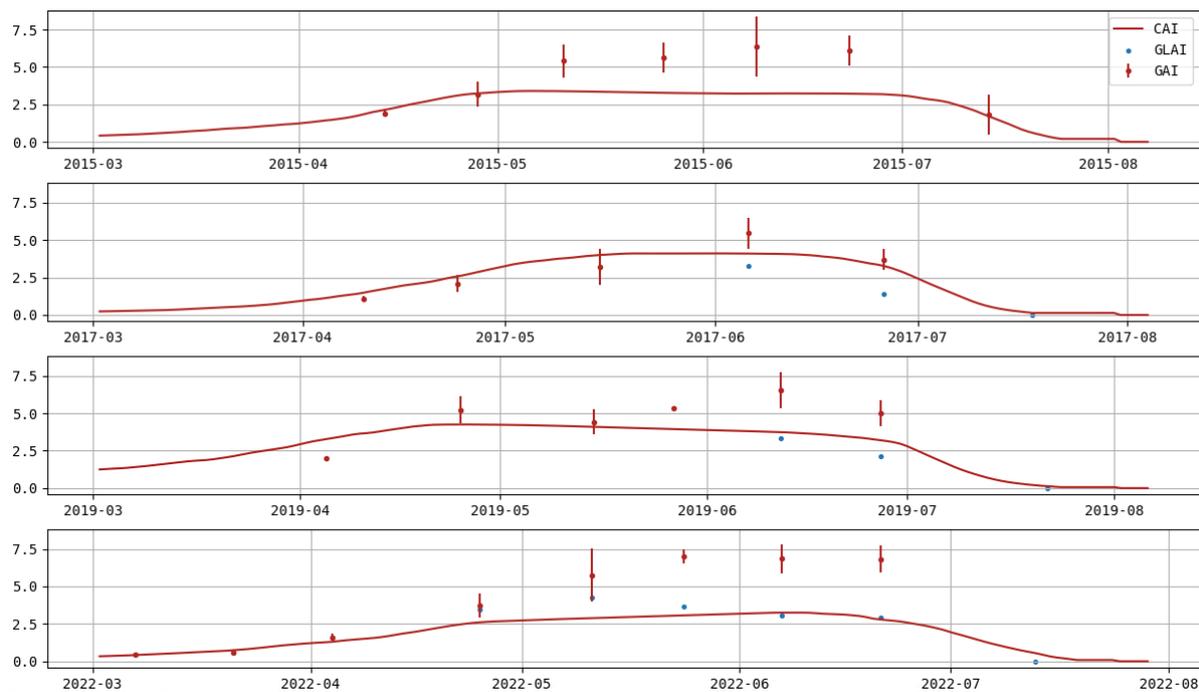
Specific comments:

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 CO_2 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.
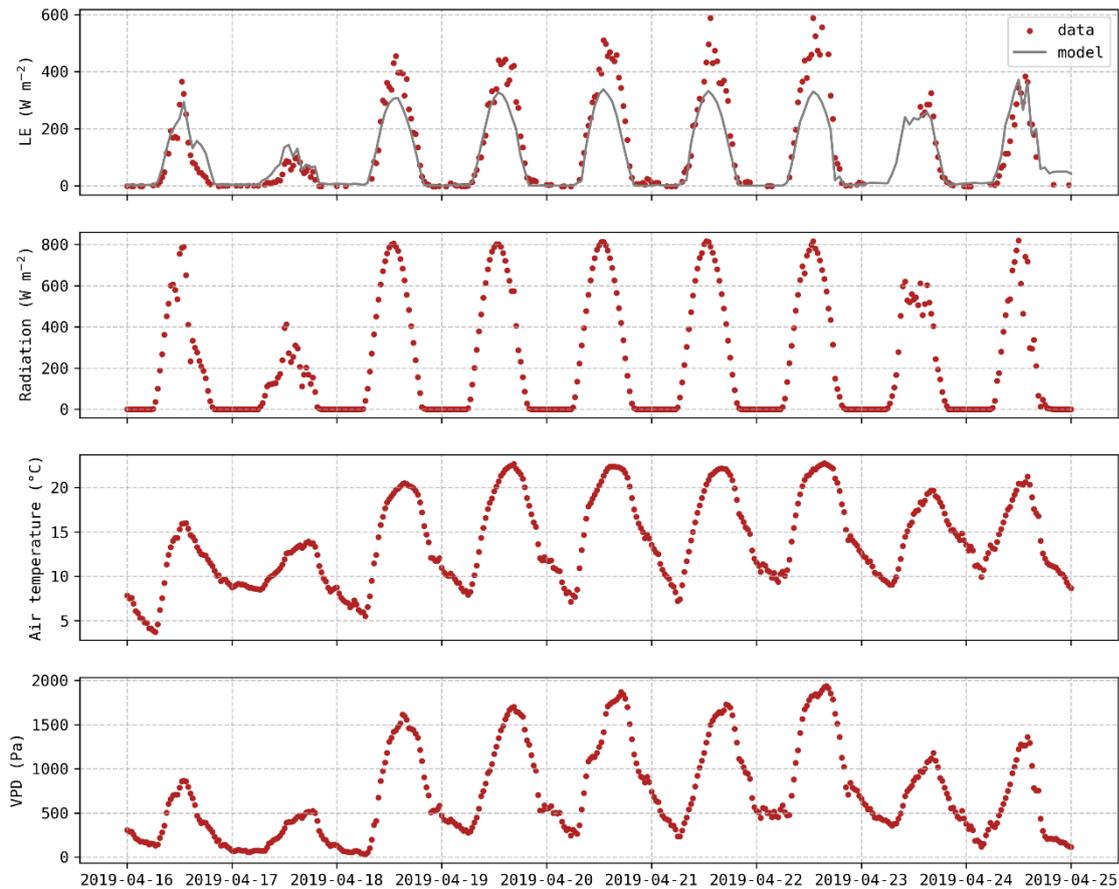
Section 4.2: The issue of calibrating the model separately for each cultivar grown in only one season is not discussed. For instance, the large differences observed in some parameters may be difficult to explain purely on a genetic basis. The large range in parameters such as *SOrgPhotEff* and *stemPhotEff* 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.

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.
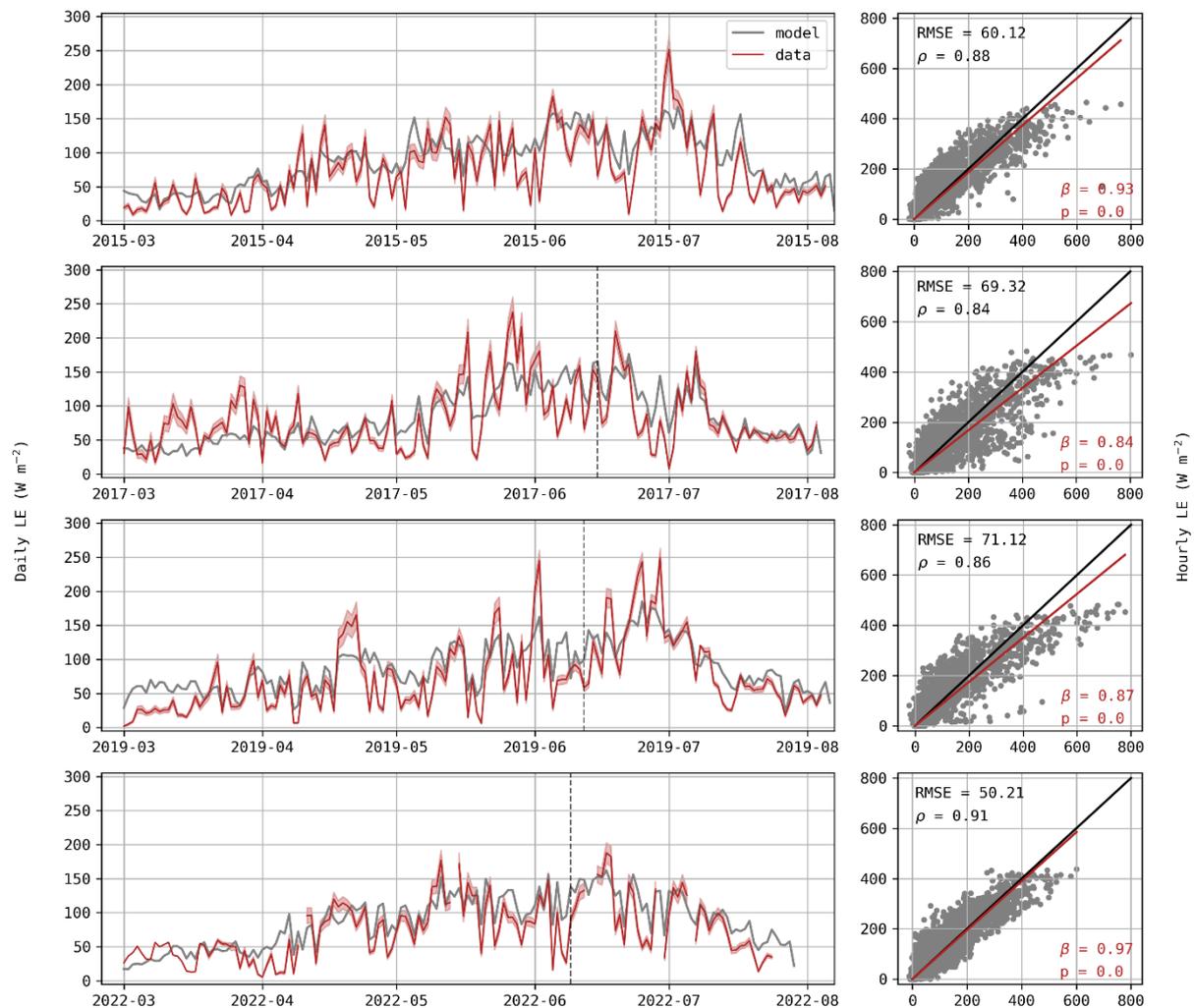
Technical corrections

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.