

1 Response to reviewer 1

General Comment

This manuscript presents a well-integrated pipeline connecting CRNS observations, CLM5-PDAF data assimilation, ECMWF ensemble forecasts, and an accessible display of soil-moisture conditions. The application at Selhausen is clearly motivated, and the comparison with independent SoilNet measurements is a strength. The study provides a useful demonstration of the technical feasibility of this workflow. My recommendation is minor revision, principally to align the interpretation with the evidence presented. The results establish improved agreement of the assimilated model state with CRNS observations and illustrate one 10-day forecast. They do not yet establish general forecast skill or the effectiveness of irrigation decisions. Please find below some comments:

Response:

Thank you for your supportive and constructive comments and suggestions! We agree that some interpretations in the original version of the manuscript sometimes extended beyond what can be demonstrated from the present site-specific case study. We will revise the manuscript to better align the conclusions with the available evidence and clarify the methodological assumptions and limitations.

Comment 1

Figure 9 shows a forecast initialized on 1 July 2022, but the abstract and conclusions make general claims about lead-time skill (abstract, lines 11–17; lines 410–426, 518–532). The single example cannot establish when DA improves forecasts or how long that benefit persists. In case you wish to make a general claim, you shall run forecasts from multiple dates spanning wet and dry periods. Report error against subsequent observations by lead day for both DA- and open-loop-initialized forecasts, alongside ensemble coverage or a proper probabilistic score.

Response:

We agree that a forecast initialized on a single date cannot establish how long the benefit of data assimilation (DA) persists. We also agree that a more systematic evaluation requires forecasts initialized across both wet and dry periods. Although the system supports daily forecast initialization, a complete archive of forecasts covering the full season was not retained. Therefore, a full season evaluation cannot be conducted and would require rerunning the ensemble forecasts. Given the computational cost of these additional runs, we will limit the paired DA- and OL-initialized forecast evaluation to July 2022. We acknowledge that this predominantly dry period does not cover the suggested wet–dry contrast, and we will restrict the claims in the abstract and conclusions to the conditions and period evaluated.

For both DA-initialized and open-loop (OL)-initialized forecasts, we will report the RMSE of the ensemble mean against subsequent CRNS observations by forecast lead day. To evaluate the full ensemble forecast distribution, we will additionally report the continuous ranked probability score (CRPS) following [ECMWF's guidance](#), calculated at each verification dates t as: $\frac{1}{N} \sum_{i=1}^M |\theta_{\text{sim},t}^i - \theta_{\text{obs},t}| - \frac{1}{2M^2} \sum_{i=1}^M \sum_{j=1}^M |\theta_{\text{sim},t}^i - \theta_{\text{sim},t}^j|$

where M is the number of ensemble members, $\theta_{\text{sim},t}^i$ is the soil moisture forecast from member i , mapped to the CRNS observation space, and $\theta_{\text{obs},t}$ is the corresponding CRNS observation. CRPS evaluates the predictive distribution, has the same units as soil moisture, and is better when smaller. We will average CRPS across initialization dates separately for each forecast lead day, using the same verification cases for the DA and OL forecasts.

We will flag verification dates potentially affected by suspected irrigation, such as 5–6 July, and report metrics both including and excluding these dates to assess their influence.

Comment 2

Figure 7 shows modest RMSE improvements at 5–50 cm and a deterioration at 100 cm. The substantial shallow-depth discrepancy after SoilNet sensors were removed and reinstalled also limits interpretation. A brief, balanced summary of these findings should accompany the stronger CRNS-based result. Please clarify that point measurements and the spatially integrated CRNS observation have different spatial support.

Response:

Thank you for raising this concern. In the revised manuscript, we will:

- 1) Provide a clearer description of the quality-control procedure applied to the SoilNet profiles and mark the period affected by field operations in Figure 7.
- 2) Recalculate the validation statistics separately for the cropping period and the bare-soil periods. This will allow us to assess the influence of the disturbed measurement period more explicitly.
- 3) Add a brief, balanced summary of the SoilNet results and analyze the effect of sensor removal and reinstallation.
- 4) Clarify that the point measurements and the spatially integrated CRNS observations have different spatial support. This point is addressed in more detail in our response to Comment 4.

Comment 3

The site was irrigated in 2022, but its timing and amount are unknown, while the model is configured as rainfed potato (lines 206–213). This is an unmeasured water input and a serious limitation for both water-budget interpretation and forecast verification. The possible irrigation-related wetting around 5–6 July cannot have been predicted by a forecast initialized on 1 July unless the future input was specified. Later daily assimilation may correct the subsequent analysis, but that does not demonstrate forecast recovery (lines 359–365). Separate those two statements and, if records cannot be recovered, evaluate periods with and without suspected management inputs.

Response:

Thanks for the comment. We opted for the rainfed configuration because the timing and amounts of irrigation applied at the site in 2022 were unavailable, though irrigation occurred during the 2022 growing season. Activating the irrigation scheme in CLM5 would generate synthetic irrigation based on the simulated soil water deficits and predefined thresholds, which does not represent the actual site-specific management and can also introduce biases in irrigation amounts (Cheng et al., 2021; Yao et al., 2022).

We also thank the reviewer for pointing out this ambiguity. Our intention in lines 359–365 was to refer to the continuous DA experiment shown in Figure 6, where assimilating CRNS measurements after an unrepresented irrigation event can help correct the soil moisture discrepancies, not to suggest that a forecast initialized on 1 July could predict or recover from an unknown future irrigation event. We will clarify this in the revised manuscript.

As described in our response to Comment 2, we will also evaluate forecast performance both including and excluding verification dates potentially affected by suspected irrigation such as 5–6 July.

Comment 4

A single 100 m × 100 m column represents the model, whereas the stated CRNS footprint extends roughly 150–240 m in radius and has moisture-dependent depth sensitivity (lines 92–98, 203–205, 279–288). Explain what land covers and soil conditions lie within that footprint, how horizontal weighting is handled, and why one column is a defensible observation counterpart. Likewise, SoilNet nodes within about 30 m of the instrument provide independent measurements, but they do not sample the full CRNS footprint (lines 227–235). These scale differences affect the observation error and interpretation of Figure 7.

Response:

We thank the reviewer for raising this important point regarding the spatial representativeness of the different data sources. We agree that these details need to be described more explicitly.

The CLM5 setup consists of a single horizontally homogeneous column, whereas the CRNS integrates soil moisture over a larger and moisture-dependent support volume. No explicit horizontal CRNS weighting is applied to the model state. We assume that the single model column is representative of the dominant conditions of the CRNS footprint for this specific field and setting.

Several site characteristics support this approximation. During the study period, the CRNS and centre of the 100 × 100 m box are located in the center of a large 9.7 ha potato field. This covers the large majority of the CRNS support volume. Soil conditions within approximately 100 m of the sensor are relatively homogeneous, with predominantly silty sediments extending beyond 2 m depth. This nearby area contributes strongly to the CRNS signal because sensitivity decreases with distance from the sensor (Köhli et al., 2015).

Similarly, we will also clarify that the SoilNet nodes are within approximately 30 m of the instrument sample. Although averaging these measurements reduces the influence of an individual point sensor, the network still samples only a limited part of the much larger CRNS footprint. We will clarify these points in the Discussion section.

Comment 5

The observation weights depend on soil moisture and sensing depth, yet Equations (2)–(3) present a fixed linear H_t ? (lines 279–288). State whether the weights are recomputed for each member and cycle, frozen during an analysis, or otherwise approximated. Describe exactly which CLM layers are updated, how increments are distributed, and how bounds on water content are enforced. Report cumulative DA water increments in mm alongside precipitation, evapotranspiration, and drainage. Without this accounting, an improved moisture fit could conceal substantial artificial water additions or removals.

Response:

Thanks for the comment. We agree that the wording was a bit ambiguous. We will expand the Methods section to describe the construction of the observation operator and how the state update is applied. We will also add a water-budget estimate of the cumulative DA-induced water increments (in mm) to show that the improved moisture fit does not conceal substantial artificial water additions or removals.

Comment 6

The CRNS estimates are calculated using a 24-hour rolling mean and assimilated daily at 12:00 UTC with an assumed uncertainty of $0.02 \text{ cm}^3 \text{ cm}^{-3}$ (lines 220–224, 279–288). Please specify whether each averaging window ends at the assimilation time; a centered window would include observations unavailable at that time. Clarify whether the model observation operator uses a corresponding 24-hour average or an instantaneous state. Finally, explain what the prescribed uncertainty includes, particularly whether it accounts for calibration and differences between the CRNS footprint and the modeled column. A brief discussion of how this assumption may affect the results would suffice.

Response:

Thank you for pointing out these issues. We will clarify the temporal support of both the CRNS observations and the model observation operator in the revised manuscript. We will also add explanations regarding what the prescribed observation uncertainty includes, and add a brief discussion of how this assumption may affect the DA results.

Comment 7

Figure 10 classifies 10 cm soil moisture below 22 vol% as “irrigation needed” (lines 380–407). This surface-layer threshold is useful for communicating drying risk, but irrigation timing for potato depends on water availability throughout the root zone, crop stage, and expected rainfall. FAO-56 likewise frames irrigation scheduling in terms of root-zone depletion and readily available water. Please explain why the 10 cm threshold is appropriate for this display and describe the categories as illustrative indicators unless their link to irrigation decisions has been evaluated. Also clarify whether the reported 84% favourable and 16% drying shares apply to one forecast day or are aggregated across the forecast period.

Response:

We thank the reviewer for this comment. We agree that the threshold-based display should not be read as a complete irrigation-scheduling scheme. Figure 10 is intended to show how a probabilistic soil-moisture forecast can be translated into a simple, management-oriented indicator, not to prescribe irrigation.

We used the 10 cm layer because it is the layer most directly constrained by the CRNS signal and the most responsive to short-term drying, which makes it suitable for communicating near-surface drying risk. We also agree that root-zone moisture is more relevant for irrigation decisions. We will therefore evaluate a root-zone SM-based display as an alternative and choose the more appropriate option for Figure 10 in the revised manuscript.

We also agree that FAO-56 scheduling is formulated in terms of root-zone depletion and readily available water. However, their integration into a crop-stage-dependent irrigation scheduling method and evaluation have not been undertaken in this study and are beyond the scope of the demonstrated forecasting workflow.

In the revised manuscript, we will describe the categories as illustrative indicators, and state that their link to actual irrigation decisions has not been evaluated. We will also clarify that the reported numbers are aggregated across the forecast period.

Specific comment 1

The pronounced post-reinstallation discrepancy at 5–20 cm complicates validation. Report metrics separately before and after sensor removal, and explain how disturbed observations were screened.

Response:

Please see our response to [General comment 2].

Specific comment 2

Figure 9 spans approximately 10 forecast days, whereas the decision display and its discussion end on 6 July. Explain why only this shorter interval is classified. Define precisely what the shaded “spread” denotes in Figure 9.

Response:

Thank you for pointing this out. Figure 9 presents the full forecast of approximately 10 days, whereas Figure 10 was intended as a short demonstration of how the ensemble forecast can be translated into a farmer-oriented decision display over the first several forecast days. In the revised manuscript, we will explain this choice in the text. We will also define the shaded ensemble spread precisely in the captions of Figures 7–9.

Specific comment 3

Give the measured or fitted retention-curve parameters behind the 10 and 42 vol% thresholds and evidence connecting the wet threshold to trafficability. Soil matric potential alone does not validate machinery access.

Response:

We thank the reviewer for this comment. We will add the ROSETTA class-average van Genuchten parameters for silt loam and their source: $\theta_r = 0.065$, $\theta_s = 0.439$, $\alpha = 0.00506 \text{ cm}^{-1}$, and $n = 1.663$, with $m = 1 - 1/n$. These estimated parameters yield approximately 8.5 vol% at $pF = 4.2$ and 41.9 vol% at $pF = 1.8$. For the illustrative display, we simplified the lower threshold to 10 vol% and rounded the upper threshold to 42 vol%.

We agree that soil water content or matric potential alone does not validate machinery trafficability. In the revised manuscript, we will describe the wet category as an illustrative indicator of very wet soil conditions with potentially reduced trafficability, and we will update Figure 10 accordingly.

Specific comment 4

Bulk-density units should be g cm^{-3} , not g cm^3 .

Response:

Thank you. We will correct this.

2 Response to reviewer 2

General Comment

I reviewed the manuscript entitled “From Sensors to Irrigation Decisions: An Automated Data-Model Pipeline for Real-Time Soil Moisture Forecasting in Agriculture” from H. Zhao and co-authors. I would like to thank the authors for their work. The manuscript is well written and well presented. I have however some major concerns, mainly related to the robustness of the conclusions of the paper. I therefore suggest a major revision before this manuscript can be accepted for publication. These comments rely on my own expertise from a modeling/DA background at larger scales, and should be taken as suggestions to improve the robustness of the manuscript and its overall understanding.

Response:

We thank the reviewer for the helpful comments and suggestions. Our detailed responses to each point follow.

Main comment 1

As I understand it, the main aim is to develop a decision-support tool for irrigation and field management. The underlying science hypothesis is that assimilating CRNS soil moisture (SM) into CLM5 gives better initial conditions, and therefore better forecasts to support the decision planning. This is currently based on one example forecast, initialized on 1 July 2022 (Fig. 9-10). A more thorough and complete analysis would make the manuscript more robust and valuable. Section 3.5 states that forecasts are initialized once per day, so different initialization times could be evaluated (e.g. all days of the 2022 growing season). I suggest reporting the performance of the OL-initialized and DA initialized forecasts per lead time, averaged over all initialization times. The evaluation can be done against the CRNS SM, as in Fig. 9 (noting that this is the assimilated dataset), and/or against the SoilNet sensors. This would add material to the paper, but other sections could be shortened (see next comment).

Response:

We agree that a forecast initialized on a single date is not sufficient to demonstrate the benefit of CRNS data assimilation (DA) for forecast initialization, and that a more systematic evaluation would make the manuscript more robust. Although the system supports daily forecast initialization, a complete archive of forecasts covering the full season was not retained, so the missing forecasts would need to be rerun. Each initialization requires separate OL- and DA-initialized ensemble forecasts driven by the meteorological forecasts issued for that date, with no further assimilation during the forecast period; therefore, a full season evaluation requires multiple forecasts. Given this additional computational cost, we will restrict the evaluation to July 2022 and acknowledge its predominantly dry conditions as a limitation.

For both DA-initialized and open-loop (OL)-initialized forecasts, we will report the RMSE of the ensemble mean against subsequent CRNS observations by forecast lead day. To evaluate the full ensemble forecast distribution, we will additionally report the continuous ranked probability score (CRPS) following [ECMWF's guidance](#), calculated at each verification dates t as: $\frac{1}{N} \sum_{i=1}^M |\theta_{\text{sim},t}^i - \theta_{\text{obs},t}| - \frac{1}{2M^2} \sum_{i=1}^M \sum_{j=1}^M |\theta_{\text{sim},t}^i - \theta_{\text{sim},t}^j|$

where M is the number of ensemble members, $\theta_{\text{sim},t}^i$ is the soil moisture forecast from member i , mapped to the CRNS observation space, and $\theta_{\text{obs},t}$ is the corresponding CRNS observation. CRPS evaluates the predictive distribution, has the same units as soil moisture, and is better when smaller. We will average CRPS across initialization dates separately for each forecast lead day, using the same verification cases for the DA and OL forecasts.

We will flag verification dates potentially affected by suspected irrigation, such as 5–6 July, and report metrics both including and excluding these dates to assess their influence.

Main comment 2

The first part of the results (Section 4.1, Fig. 6) is basically a sanity check of the EnKF, because the DA is evaluated against the CRNS SM that it assimilates. This should be stated more explicitly. If the first suggestion is taken up, Fig. 6 could be moved to the supplementary material and discussed in a short paragraph. The comparison against the SoilNet sensors (Fig. 7) is the independent validation and should stay in the main text.

Response:

Thank you for this comment. We agree that Section 4.1 is essentially a sanity check of the EnKF, since the DA results are evaluated against the same CRNS soil moisture that is assimilated. In the revised manuscript, we will state this and reformulate the section.

Methodological clarifications 1

Observation operator and update (Section 2.3.3 and L282-288). The text explains how the model layers are weighted to match the CRNS measurement depth (D_{86} , Schrön et al., 2017), but the following is not clear:

- The whole soil column is updated, but how? L287-288 states that the layer-specific weights are used "to distribute the observation increments across the soil column", whereas Eq. (2)-(4) imply that the increments follow from the covariance matrix derived from the ensemble. Which of the two is used, and which soil layers are part of the state vector?
- There seems to be a seasonal bias between the CRNS SM and the OL SM (overestimation in wet periods, underestimation in summer [but this could be due to the irrigation!], L308-309), but no strong overall bias (MBE = 1.43 vol%). Is any bias correction or rescaling applied, in the observation operator or before assimilation?
- Even if these aspects are described in Strebel et al. (2022), they should be briefly summarized here.

Response:

Thank you for pointing out these unclear aspects. We will clarify the observation operator and analysis update in Section 2.3.3.

At each daily assimilation cycle at 12:00 UTC, the effective measurement depth D_{86} is recalculated at horizontal distances of 1, 20, and 85 m using the current CRNS soil moisture observation and the mean dry bulk density of the upper eight CLM5 soil layers, extending to 0.92 m. For each distance, an unnormalized vertical weighting profile is calculated at 1 mm intervals as $q_r(z_j) = \exp\left[-\frac{2z_j}{D_{86}}\right]$, where $z_j = j/10$ cm for ($j=1, \dots, 920$), and D_{86} is expressed in the same units as z_j . Each profile is first normalized over this depth interval, then the three normalized profiles are then averaged with equal weights. The resulting depth-specific weights are summed within each of the upper eight model soil layers and normalized to obtain the layer observation weights:

The depth-dependent weights define the observation operator:

$$\mathbf{H}_t \mathbf{x}_{i,t}^f = \sum_{k=1}^8 w_{k,t} \theta_{i,k,t}^f \quad (1)$$

where $\theta_{i,k,t}^f$ is the forecast soil moisture of ensemble member i in layer k , and $w_{k,t}$ is the corresponding normalized observation weight. The weights influence the Kalman gain through the observation operator, but are not used to distribute the analysis increments. The increments depend on each layer's ensemble covariance with the simulated CRNS-equivalent soil moisture. Layers below 0.92 m have zero observation weights, but can still be updated through these cross-covariances.

No bias correction or rescaling was applied, either before assimilation or within the observation operator. The annual MBE was 1.43 vol.%, indicating a limited offset, although we acknowledge that compensating seasonal differences may contribute to this value. Also, rescaling observations to the model climatology is not universally beneficial. Its effectiveness depends on the relative accuracy of the model and observations, and it can degrade assimilation when the observations are more accurate (Yilmaz et al., 2016). Moreover, where irrigation is insufficiently represented in the model, rescaling may suppress useful irrigation-related information in the observations (Kwon et al., 2022).

Methodological clarifications 2

Perturbation and ensemble setup (Sections 3.2 and 3.5):

- How many ensemble members are used in the OL and DA experiments? N in Eq. (4) is never given.
- When are the forcing perturbations applied (at every time step? Daily?), and is there a temporal autocorrelation? Table 1 gives the cross-correlations between variables, but no information on temporal autocorrelation, which is important for meteorological forcings in case they are perturbed at subdaily time steps.
- For the forecasts, how are the ensemble members combined with the 50 ECMWF ENS members? Is each member restarted with one ENS member, or with all ENS members (which would increase the number of ensemble members)?
- Are perturbations still applied in the forecasts? L367-368 suggests that the perturbed soil properties are kept and that the forcing perturbations are replaced by the ENS members. Please state this explicitly in Section 3.5.

Response:

Thank you for pointing these out. Both the OL and DA experiments used 50 ensemble members, matching the number of ECMWF ENS forecast members. We will state this explicitly in the revised manuscript. Forcing perturbations are applied daily without temporal autocorrelation. At each forecast initialization, each model ensemble member is driven by one corresponding ENS member, retaining an ensemble size of 50. No additional forcing perturbations are applied during the forecast, while the perturbed soil properties are retained.

Minor comment 1

Acronyms: IC (Fig. 2c), EC (EC station, from L242) and eCLM (L264) are not defined. In general, please make sure all acronyms are spelled out.

Response:

We will define those acronyms at first use and will also check the full text to make sure that all acronyms are spelled out at first mention.

Minor comment 2

How is the spread defined? The variance of the ensemble, the standard deviation, the min-max range? Fig. 7-8 use "span" in the legend and Fig. 10 uses percentile ranges. Please define it once and use it consistently. However, I understand that percentiles might be given to the end user for clarity.

Response:

In Figures 7-8, the shaded area denotes the min-max range of the ensemble. We will define this and use the term consistently in the legends and captions of all figures. For Figure 10, where percentile ranges are shown for the end user, we will state this explicitly in the caption and explain why they are used there.

Minor comment 3

L207-210: the potato parameterization of Boas et al. (2021) is used instead of the default CLM5 parameters. How sure are we about this parameterization for this site and year? In general, it would be good to have an idea of the performance of the vegetation estimates (e.g. LAI, biomass, ...), as they also influence the soil moisture.

Response:

Thank you for the comment. The potato parameterization of Boas et al. (2021) Developed by combining literature values with field observations, and was evaluated for potato seasons at Selhausen, Germany (2019), and Lonzée, Belgium (2011, 2014, and 2018). Their Fig. 6 shows improved agreement with observed LAI and crop development compared with the default

spring-wheat parameterization used to represent potatoes in CLM5. This provides a basis for selecting these parameters (as in Fig. 1).

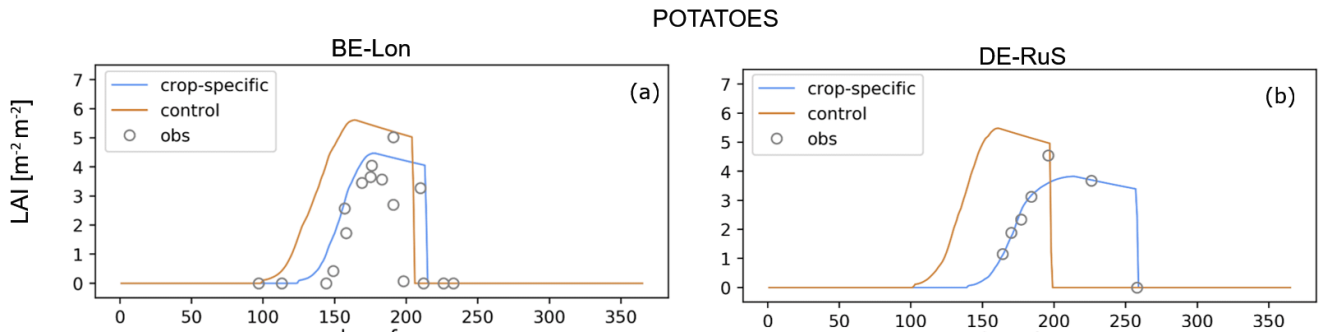


Figure 1: Evaluation of the potato parameterization at Lonzée (left column) and Selhausen (right column). Panels show (a–b) leaf area index (LAI). The default parameterization (orange) and the potato-specific parameterization (blue) are compared with observations (grey). LAI observations are shown individually, whereas fluxes are monthly averages across the evaluated potato years. Reproduced from Figure. 6 of Boas et al. (2021), distributed under the Creative Commons Attribution 4.0 License.

To assess the applicability to our study period, we compared LAI from the open-loop simulation with the available field observations for 2022. This comparison revealed a timing mismatch: simulated canopy development and the end of the growing season occurred earlier than observed. The planting records help explain this discrepancy. In 2019, the model's temperature-based planting criteria were satisfied on 23 April, close to the recorded planting date of 26 April. In 2022, however, the same criteria were already satisfied on 1 April, whereas actual planting took place on 30 April, a difference of 29 days. It should be noted here that using the default parameterization would have resulted in an even earlier simulated plantig date, and thus a larger mismatch with the actual date. These findings suggest that the temperature-based planting rule, which does not fully capture actual management decisions, contributed to the early simulated crop development in 2022. We will therefore acknowledge this uncertainty in the simulated vegetation dynamics for this year. The LAI timing mismatch may affect the timing and magnitude of soil water depletion and should be considered when interpreting the soil moisture results.

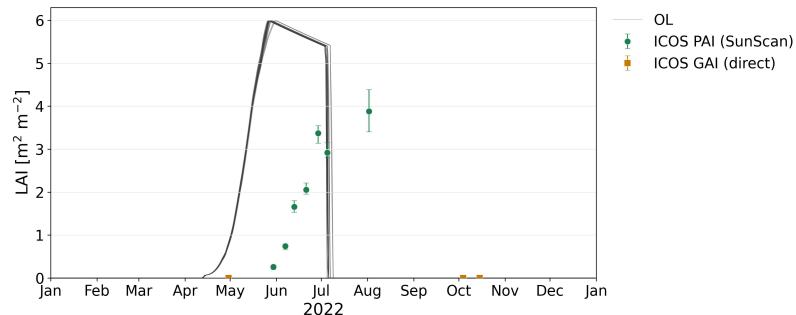


Figure 2: Seasonal development of simulated leaf area index (LAI) and observed canopy area indices at Selhausen in 2022. Grey lines represent individual open-loop (OL) ensemble members. Green circles show ICOS plant area index (PAI) measured using SunScan, and orange squares show ICOS green area index (GAI) from direct measurements.

Minor comment 4

Fig.2:

- Panel b: what does the inverted triangle stand for? It is typically used for the groundwater table depth. Is that the case here?
- Panel a: "Enduser" is written in one word, whereas the text uses "end users" (L184) and "end-users" (L45-46). Please harmonize.

Response:

Thank you for these remarks. In Panel b, the inverted triangle indicates the groundwater table depth. We will add this to the Figure 2: Schematic overview of the monitoring, modelling, and forecasting framework developed for site-specific soil moisture prediction at the Selhausen field site. a) Data acquisition and processing chain; b) Conceptual representation of the modelled soil column; c) Temporal workflow of the modelling system.

We will also harmonize the spelling to “end users” in Panel a of Figure 2 and throughout the text.

Minor comment 5

Perturbations (Section 3.2): uncertainty in soil parameters and meteorological forcings is considered via the perturbations. The authors could also consider perturbing the SM directly, as is typically done in SM DA, although the spread was probably large enough. Was any kind of ensemble verification performed to check that the spread is correctly designed for assimilation?

Response:

We thank the reviewer for this suggestion. Direct perturbation of soil moisture states is an established option. In our implementation, soil moisture spread was generated through perturbations to soil properties and meteorological forcings, whose effects propagate through the CLM5 dynamics.

We calculated the OL ensemble against CRNS observations using the same vertical weighting as the assimilation operator. The mean ensemble standard deviation was $0.035 \text{ cm}^3 \cdot \text{cm}^{-3}$, compared with a RMSE of $0.042 \text{ cm}^3 \cdot \text{cm}^{-3}$. The ensemble range contained 91.8% of the observations, and the ensemble mean showed a slight wet bias of $0.013 \text{ cm}^3 \cdot \text{cm}^{-3}$. These diagnostics indicate substantial ensemble spread, though some discrepancies remain insufficiently represented.

Minor comment 6

Please make clear that only state updating (SM) is done, and no parameter or forcing updates. L163-164 could be read otherwise.

Response:

Thanks. We will clarify this in the revised manuscript.

Minor comment 7

Fig. 5: HRES is introduced at L151-153, but it is not used in the pipeline and its comparison with the ENS in Fig. 5 is not discussed in the text. Please explain why it is shown. Out of curiosity: why is there such a large deviation in Pair between ENS and HRES, already at the start of the forecast?

Response:

We included HRES as a deterministic reference to highlight the additional uncertainty information provided by the ENS ensemble. We will clarify this purpose in the text.

The surface-pressure offset between HRES and the ENS control may reflect differences in model grid elevation associated with their respective horizontal resolutions of approximately 9 and 18 km used in 2022. Selhausen lies at approximately 103 m above sea level, and surrounding terrain features may be represented differently by these two grids. Under representative summer conditions of virtual temperature ($T_v = 295$) K and pressure ($p = 1000$) hPa, the hydrostatic relationship indicates that a pressure difference of 2.5 hPa corresponds to an elevation difference of approximately 21.6 m, with lower pressure corresponding to higher elevation. This provides a plausible explanation for the offset. However, confirming it would require comparing surface geopotential from the two model configurations. We have not performed this comparison and therefore present this explanation as a hypothesis.

Minor comment 8

L272: "accelerated deposition", do you mean "accelerated decomposition" (the AD spin-up mode of CLM)? Also, are all ensemble members spun up separately, or is one deterministic spin-up split into an ensemble at a later stage? L276 suggests the former. Please clarify.

Response:

Thank you for pointing this out. Yes, “accelerated deposition” was a typo, and we will correct it to “accelerated decomposition”. We will also state in the revised manuscript that each ensemble member was spun up separately.

Minor comment 9

Fig. 6: it would be useful to also plot the OL and DA spread in panel a, to show the expected decrease in uncertainty for the DA compared to the OL.

Response:

Thank you. In the revised manuscript, we will add the OL and DA ensemble spread to panel a of Figure 6.

Minor comment 10

L320-321: "but the improvements are more limited within the 5-20 cm layers". Please rephrase. It reads as if the improvements are smaller at those depths, which is not the case: the improvements are limited to those depths.

Response:

Thank you for the comment. We will remove it and rephrase the statement. To better illustrate the results and assess the influence of sensor removal and re-installation, we will also calculate the statistics separately for the field management period and for the periods outside it and discuss the results in the revised text.

Minor comment 11

Fig. 7:

- Could you also explain the jumps in the observations around May and October in panel d-1? L331-333 attributes the decline in May to the removal and re-installation of the sensors during ploughing. I suspect the jump in October is also due to field management?
- Use "OL spread" and "DA spread" instead of "span" (also in Fig. 8), and clarify how the ensemble spread is defined, as requested above. I suspect it is the min-max range, because you have the output of all the ensemble members. Other models (only outputting the ensemble mean and standard deviation) typically plot the ensemble mean $\pm$ 1 standard deviation.
- The boxes with the metrics in the scatter plots sometimes overlap with the data (e.g. panel a-3). Consider placing the box in the lower right corner in those cases.
- Caption: remove the space between "(" and "5".

Response:

Thank you for your comments. Our previous quality control relied on the SoilNet quality flags, but we found that these flags did not identify all the artifacts related to field operations. We will apply a more rigorous quality-control procedure to the SoilNet data and mark the affected periods in Figure 7. We will also replace “span” in the legends and define the ensemble spreads (see our response to the earlier comment on the spread definition). We will also move the metric boxes in the scatter plot and remove the space in the caption.

Minor comment 12

L343: "model coupling". The phrasing "model propagation" is also commonly used.

Response:

We appreciate the comment. We will replace “model coupling” with “model propagation” in L343 of the revised manuscript.

Minor comment 13

Section 4.1: the authors could also note the reduced spread because of the DA (visible in Fig. 7).

Response:

Thank you for the suggestion. We will discuss the reduced ensemble spread after DA in Section 4.1. The mean ensemble standard deviation decreased from 0.035 (OL) to 0.012 $\text{cm}^3 \cdot \text{cm}^{-3}$, while the RMSE between the ensemble mean and CRNS observations decreased from 0.042 to 0.021 $\text{cm}^3 \cdot \text{cm}^{-3}$. Ensemble-range coverage decreased from 91.8% to 79.7%. The lower coverage suggests that the reduced DA ensemble spread may not fully represent uncertainty.

Minor comment 14

Fig. 8: there is a large deviation between the observed and the simulated soil temperature (OL and DA) in July, August and September. L347-349 attributes this to a too high thermal conductivity or a too low LAI in the model. Could it also be due to the irrigation, which is unknown (L212-213), maybe indirectly through a larger LAI in reality compared to the model?

Response:

Thank you for this suggestion. As discussed in our response to [Minor comment 3] on the potato parameterization, the comparison shows that simulated LAI declined to zero in early July, whereas the in situ observations indicated large canopy area into August. Although the observed PAI is not directly equivalent to modelled LAI, this timing mismatch supports insufficient simulated canopy cover as a plausible contributor to the summer soil temperature bias. We agree that unrecorded irrigation could have contributed to the soil temperature discrepancy. Irrigation could have helped maintain the observed canopy during dry conditions, but its contribution cannot be quantified because the timing and amount of irrigation are unknown. However, the available canopy observations do not establish whether this explanation extends through September. We will expand the discussion to include the unrecorded irrigation and its possible indirect effect via canopy development.

Minor comment 15

Fig. 9:

- I would find it interesting to see both the OL and DA spread, in the hindcast and in the forecast (only the DA spread is shown now). This could be discussed to see if the DA also reduces the forecast uncertainty (but I expect a large part of the uncertainty comes from the ENS forecasts).
- Again, how is the spread defined here? It seems very large in general.
- In the hindcast period, SM, ST and ET follow a straight line from about 00:00 on 1 July until the forecast initialization at 12:00. Can the authors confirm and explain this?
- I cannot see the ECMWF ENS mean line in panel b.
- Caption: "ForecastOL" and "ForecastDA" are used here (with a bracket missing) and in the caption of Fig. 10, whereas the legend uses "OL forecast" and "DA forecast". Please improve the overall consistency.

Response:

Thank you for these detailed remarks. We initially omitted the OL spread to avoid overloading the figure. We will add the OL spread in addition to the DA spread for both the hindcast and the forecast period in Figure 9. The shaded spread here is the min-max range of the ensemble members, and we will state this in the caption.

We checked the original model output and found that the straight-line behavior of SM, ST, and ET is a plotting artifact. We will correct the figure. We will also make the ECMWF ENS mean in panel b clearly visible and harmonize the terminology in the legends and captions.

Minor comment 16

Section 4.3: why not also consider the critical soil moisture (and soil moisture in general) for the root zone or a deeper layer? The root-zone value of Lu et al. (2026) is given (L385) but not used, because the forecasts represent the upper 10 cm (L386-387). You have a model, and the DA also updates the other layers, so the authors could take advantage of this.

Response:

Thank you for this suggestion. We agree that root-zone soil moisture is more relevant for irrigation and field management. We used the 10 cm layer because it aligns better with the CRNS sensing depth, which is approximately 15-30 cm depending on soil moisture content, and because it shows the drying more clearly. We also agree that we can take better advantage of the model, since the DA updates the other soil layers as well. In the revised manuscript, we will therefore consider soil moisture integrated over the CRNS effective sensing depth and apply the critical soil moisture of Lu et al. (2026).

Minor comment 17

Fig. 10: I cannot see the precipitation mean in the lower panel.

Response:

Thank you for pointing this out. The precipitation mean was hidden by an overlapping marker. We will revise the lower panel of Figure 10 so that the mean is visible.

Minor comment 18

L478: the reference should be "De Lannoy et al., 2022". The entry in the reference list also starts with "Lannoy".

Response:

Thank you for catching this. We will correct the citation in L478 to "De Lannoy et al., 2022" and fix the corresponding entry in the reference list. We will also apply the same correction to all other references that include De Lannoy as a co-author.

References

- Boas, T., Bogen, H., Grünwald, T., Heinesch, B., Ryu, D., Schmidt, M., Vereecken, H., Western, A., and Franssen, H. J. H.: Improving the representation of cropland sites in the Community Land Model (CLM) version 5.0, Geoscientific Model Development, 14, 573–601, <https://doi.org/10.5194/GMD-14-573-2021>, 2021.
- Cheng, Y., Huang, M., Zhu, B., Bisht, G., Zhou, T., Liu, Y., Song, F., and He, X.: Validation of the Community Land Model Version 5 Over the Contiguous United States (CONUS) Using In Situ and Remote Sensing Data Sets, Journal of Geophysical Research: Atmospheres, 126, e2020JD033 539, <https://doi.org/10.1029/2020JD033539>, 2021.
- Kwon, Y., Kumar, S. V., Navari, M., Mocko, D. M., Kemp, E. M., Wegiel, J. W., Geiger, J. V., and Bindlish, R.: Irrigation characterization improved by the direct use of SMAP soil moisture anomalies within a data assimilation system, Environmental Research Letters, 17, 084 006, <https://doi.org/10.1088/1748-9326/AC7F49>, 2022.
- Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, Water Resources Research, 51, 5772–5790, <https://doi.org/10.1002/2015WR017169>, 2015.
- Lu, X., Groh, J., Graf, A., Pütz, T., Gerke, H. H., Gründling, R., Rupp, H., Kiese, R., Vogel, H. J., Javaux, M., Vereecken, H., and Franssen, H. J. H.: Robustness of critical soil moisture to curve-fitting methods and its variability with soil depth, soil texture, and climatic conditions: insights from lysimeter data in Germany, Journal of Hydrology, 668, 134 959, <https://doi.org/10.1016/J.JHYDROL.2026.134959>, 2026.

- Yao, Y., Vanderkelen, I., Lombardozzi, D., Swenson, S., Lawrence, D., Jägermeyr, J., Grant, L., and Thiery, W.: Implementation and Evaluation of Irrigation Techniques in the Community Land Model, *Journal of Advances in Modeling Earth Systems*, 14, e2022MS003074, <https://doi.org/10.1029/2022MS003074>, 2022.
- Yilmaz, M. T., Crow, W. T., and Ryu, D.: Impact of Model Relative Accuracy in Framework of Rescaling Observations in Hydrological Data Assimilation Studies, *Journal of Hydrometeorology*, 17, 2245–2257, <https://doi.org/10.1175/JHM-D-15-0206.1>, 2016.