

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.

Main comments:

- 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).
- 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.

Methodological clarifications:

- 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 (D86, Schrön et al., 2017), but the following is not clear:
 - o 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.
- 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.

Minor comments:

- 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.
- 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.
- 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.
- 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.
- 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?
- Please make clear that only state updating (SM) is done, and no parameter or forcing updates. L163-164 could be read otherwise.
- 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?
- 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.
- 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.
- 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.
- 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 +/- 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".
- L343: "model coupling". The phrasing "model propagation" is also commonly used.
- Section 4.1: the authors could also note the reduced spread because of the DA (visible in Fig. 7).

- 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?
- 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.
- 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.
- Fig. 10: I cannot see the precipitation mean in the lower panel.
- L478: the reference should be "De Lannoy et al., 2022". The entry in the reference list also starts with "Lannoy".