

We would like to thank all of the reviewers for the thorough and insightful suggestions and comments. We made substantial changes to the manuscript, replaced one figure, and completed an additional model simulation in response to the feedback we received. We feel that the manuscript has improved significantly as a result of these thoughtful reviews. Please find our detailed responses to the reviewer's comments below.

Please note that the **reviewer comments are shown in black** and **our author responses are in blue**. Where changes have been made in the manuscript, the page and line number(s) are given. In some cases, to highlight changes to passages in the manuscript, these sections are copied and pasted from the manuscript.

Reviewer #4:

General comments:

This is an interesting paper on the evaluation of an irrigation scheme within a land surface modeling framework. This is an area that needs research and I see this a potentially valuable contribution on the matter.

While generally well written, the structure and organization of the Background (particularly Section 2.3) and the Methods sections needs to be improved to ensure a better flow and enhanced readability. The study region, models, input datasets and evaluations should be described in a more logical and orderly manner with less intermixing. These issues are described in more detail in the specific comments below. The discussion section is very short and would benefit from more elaboration and high quality insights on the limitations and challenges as well as opportunities for irrigation modeling.

[A new section has been added to the methods and more information has been added to the Discussion. Please see specific comments below.](#)

Some of the used input datasets need more justification. GVF is an important dataset for the irrigation modeling but is reported at coarse resolution (3 and 16 km) inconsistent with the resolution of the LSM (1 km). Not clear to me why a 1 km based version isn't used here. The MODIS phenology product (produced at 500 m resolution) would probably be more useful in this context for establishing the start and duration of the growing season.

[Please see specific comments 10 and 12 for detailed responses.](#)

I'm also a bit concerned that 1 km isn't the most appropriate scale to do irrigation modeling and accuracy assessments as you will inevitable run into mixing of rainfed and irrigated fields given the characteristic size of the fields. LSM runs at 500 m resolution would probably have been

more appropriate, also considering the scale of the CRNP validation dataset, and feasible using widely available surface inputs generated at consistent resolutions.

Mixing of rainfed and irrigated fields is certainly an issue that arises in irrigation modeling, even at 1 km, which is considered high resolution for land-atmosphere interactions and regional weather modeling studies. However, 1 km is the highest resolution we can run, while still being appropriate and relevant to our broader goals.

The spatial resolution of 1 km is the most appropriate scale for this study for two main reasons:

- 1) The highest resolution input datasets we have are 1 km, so running at 500 m would not improve our results in this study; it would simply give the same information broken up into more gridcells.
- 2) The broader context goal of evaluating this irrigation scheme is for its later use in land-atmosphere interaction studies (Page 2, paragraph 1; Section 2.1). It is difficult and typically not advisable to run a coupled atmospheric model at 500 m, especially for land-atmosphere interaction studies. The behavior of the planetary boundary layer (PBL) in atmosphere/mesoscale models, such as WRF, is determined by the PBL parameterization. These parameterizations are not recommended for use at 500 m as some of their assumptions break down at such fine scales.

Specific comments:

- 1) Page 1 L14: Please define the scale associated with “high resolution”

High resolution is 1 km in this case.

- 2) Page 1 L19: What precisely does the “human practice data” consist of?

Human practice data is the irrigation timing and amount. This has been clarified in the newly added section 3.2 – Evaluation Data.

- 3) Page 1 L21: “two irrigated fields” – what irrigated fields are you referring to here (soybean and maize)?

Yes, this is clarified in the newly added section on Evaluation Data (3.2).

- 4) Page 2 L21 and L25: Please define what you mean by coarse and high resolution here.

- 5) Page 6 L1-7: This paragraph reads a bit confusing with mentioning of all the different temporal and spatial resolutions. A bit unclear what product version is used for the evaluation. Does the 12x12 km survey area correspond to the 15x15 km domain of this study? Why the domain difference?

All of the evaluation data is explicitly defined in the newly added Evaluation Data section (3.2). The 12x12 km survey area is contained entirely within the 15 x 15 km domain area of the study. The grid projection (UTM) used in the Franz et al. (2015) study is not directly compatible with the grid definitions in LIS. Therefore, since we couldn't recreate the exact grid, we made a slightly larger domain to ensure that the entirety of the Franz domain was contained within the LIS simulation domain.

6) Page 6 L8-16: This Section adds to the confusion by repeating some of the statements above and also adding additional evaluation datasets (human practice data etc.) not related to the CRNP (although that is the title of the Section). Differences between the CRNP and COSMOS datasets should be clarified, if any. The finishing paragraph relates the overall objectives and novelty of the work, which don't belong here. This Section requires some revision – the evaluation components might be more appropriately positioned in the method section. You may need a completely separate section for describing the additional datasets mentioned here.

COSMOS is the observing network of stations and rovers, while CRNP refers to the observing instrument. The first sentence of Section 2.3 makes the distinction:

“A potential solution to fill the gap between point and remote sensing observations of soil moisture is the Cosmic-Ray Neutron Probe (CRNP) **method**, organized through the Cosmic Ray Soil Moisture Observing **System** (COSMOS), which **has ~200 probes operating globally since 2011.**”

The source of the human-practice data is Franz et al. 2015, which is described in this section, and thus why the human-practice data is mentioned here.

We have shortened this paragraph by removing details of the evaluation data and have instead incorporated these details into a new section 3.2, called Evaluation Data, as suggested by the reviewer. The novelty statement has been moved to the last paragraph of the introduction.

7) Section 2.3: The CRNP data description is currently part of the introduction/background part of the manuscript. While it makes sense to mention and introduce the data as a useful validation source in this context, I feel that the detailed description of the actual dataset used here for evaluation purposes should be moved to a separate section in the Methods section (or Methods and data section). Here you could appropriately describe all the datasets used in the study.

Details about the CRNP data from the Franz et al. study used for evaluation in our study have been moved to a new sub-section of the Methods, called Evaluation Data (3.2), as per the reviewer's suggestion.

8) Section 3: I would start this with a description of the study area and domain to set the stage.

9) Section 3.1: I find this section quite confusing to read as it includes both modeling and

evaluation details and references to elements described in Section 3.2. I think you need to rethink the organization of the Method section adopting a more logical organization for improved flow and readability. Personally, I would prefer to have all model descriptions first before the description of experiments and evaluations to be performed.

This section does not include any evaluation details. It describes the land surface model and modeling framework (paragraph 1), the time period for the simulations (paragraph 2), introduces the four simulation experiments (paragraph 3), and then details the important distinctions between the four simulation experiments (remaining paragraphs).

10) Page 8 L1-5: So why isn't the GVF datasets provided at 1 km to be consistent with the LSM resolution?

The resolution of the NCEP climatological GVF used in this study is 1 km. The statement about the 16 km GVF dataset was included as part of the summary of results from Case et al., (2014); their study used 16 km climatological GVF. Admittedly, it did read like the climatological GVF used in this study is also 16 km. We removed these extra details from the Case et al., 2014 study description as they are unnecessary and added confusion. We also added the resolutions of the GVF datasets when introducing them. Page 8 Lines 2-5 now read (bold is newly added):

“The SPoRT run makes use of the GRIPC irrigation intensity dataset, like the Standard run, but uses a real-time GVF product at **3 km spatial resolution** from NASA-MSFC's Short Term Prediction, Research, and Transition Center (SPoRT; Case et al., 2014). This is in contrast to the other runs that use climatological GVF at **1 km** from the National Centers for Environmental Prediction (NCEP).”

With respect to the resolution of input datasets more generally, we always use the best available, most appropriate input datasets for our application. Although we like to use high-resolution whenever possible, the highest resolution is not always the best available. This is the situation with our SPoRT dataset. Although the SPoRT GVF dataset is produced at 0.01 degree (~ 1 km), there was a change in the Continental US grid in Feb 2012 that impacted the 1 km dataset. We used the 3 km dataset instead of 1 km to avoid potential inconsistencies resulting from the grid change in 2012 (in the middle of our long-term spinup).

You also need to specify precisely what the GVF product is used for, when first introduced. From what I can read later in the manuscript it is predominantly used to determine the start and end of the growing season; couldn't you use the MODIS phenology product (see comment 12) more appropriately for this purpose? In addition, this product is available for the full duration of the study.

The GVF dataset is used in irrigation scheme in two main ways:

- 1) It is involved in the determination of the irrigation season, as the reviewer notes. This is a central feature of the Ozdogan et al. (2010) irrigation algorithm. While it is certainly

possible to use a different method, such as the MODIS phenology for determining the irrigation season, this would be a considerable deviation from the irrigation scheme and therefore would be counter to the goals of the study, which are to evaluate this particular scheme.

- 2) GVF is used to define the crop root zone, which impacts the amount of water applied by the irrigation scheme. The maximum root zone for each crop type is defined by a lookup table; the GVF is multiplied by the maximum root zone to determine the crop root zone. In this way, the scheme mimics the season cycle of crop root growth. More water is applied for greater crop root depth. Therefore, GVF is important for defining the irrigation season, triggering irrigation, and for determining the amount of irrigation water applied by the irrigation scheme.

The land surface model does not explicitly use a phenology dataset, such as MODIS EVI or NDVI, but rather uses proxies of Greenness Vegetation Fraction (GVF) and Leaf Area Index. The SPORT GVF dataset is based on NDVI, and therefore in essence translates the MODIS NDVI information into a form that the model can use (GVF).

11) Page 8 L12-15: You need to mention the resolution of these input datasets.

The resolution of each dataset has been added to this paragraph. It now reads as follows, with the additions shown in bold italics:

“Additional datasets common to all simulations include MODIS – International Geosphere Biosphere Program (MODIS-IGBP) land cover **at 1 km**, State Soil Geographic (STATSGO?) soil texture **at 1 km**, University of Maryland crop type **at 1 km**, and National Land Data Assimilation System – Phase 2 (NLDAS2, Xia et al., 2012) meteorological forcing at **1/8th degree (approximately 12 km)** that includes bias corrected radiation and gauge-based precipitation.”

Is the UMD crop type product static or is a separate classification provided for each year? The annual Cropland Data Layer (https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php) product (provided at 30 m) is updated for each year to account for crop rotations and changing crop type patterns and might be a more correct source to use for something like this.

The UMD crop type product is static. We agree that the Cropland Data Layer is a great improvement on static crop maps and we have discussed integrating the CDL into LIS. However, for this study, because of the small domain and the detailed ground observations we have, the CDL would not have added value beyond the ground truth provided by the Franz group. We completed the default crop type run and an additional crop type run with an observationally tuned map (detailed in the Discussion section) and found no significant differences. As a result, we believe a run with the CDL would not have differed significantly from either of these two runs.

12) Page 9 L4-5: The GVF product is used for establishing the length and timing of the growing season. A more appropriate source for this would be the MODIS global vegetation phenology product (MCD12Q2) currently produced at 500 m resolution that is also more consistent with the LSM resolution and the CRNP validation dataset (and the scale of irrigation effects). Reasons for not using something like this should be addressed.

As discussed above, a main feature of the Ozdogan et al. (2010) irrigation scheme is the determination of the irrigation season based on a threshold of the GVF. While it is certainly possible to use a different method, such as the MODIS phenology for determining the irrigation season, this would be a considerable deviation from the scheme and therefore would be counter to the goals of the study in evaluating this particular scheme.

The land surface model does not explicitly use a phenology dataset, such as MODIS EVI or NDVI, but rather uses proxies of Greenness Vegetation Fraction (GVF) and Leaf Area Index. The SPORT GVF dataset is based on NDVI, and therefore essentially translates the MODIS NDVI information into a form that the LSM can use (GVF).

13) Page 9 Section 4: A brief intro statement would be useful here.

We don't believe an intro statement is necessary here as the previous paragraph sets up the organization of this section.

14) Page 10 L7: The relationship used to compute the root zone length from GVF should be provided in the methodology.

The root zone length calculation, as it applies to the irrigation scheme, is described on Page 10, Lines 7-8.

“...,while the root zone is the produce of the maximum root depth (as defined by crop type) scaled by the GVF to mimic a seasonal cycle of root growth.”

15) Page 12 L6: This is the first mentioning of a rainfed validation site within the study domain. Details like this should be provided in the method section (preferably in a dedicated study region section).

The rainfed site was mentioned in Section 2.3 but has been moved to the new Evaluation Data section (3.2).

16) Page 13 L8-13: This should be moved to the methodology section. A shorter summary of the CRNP would suffice here.

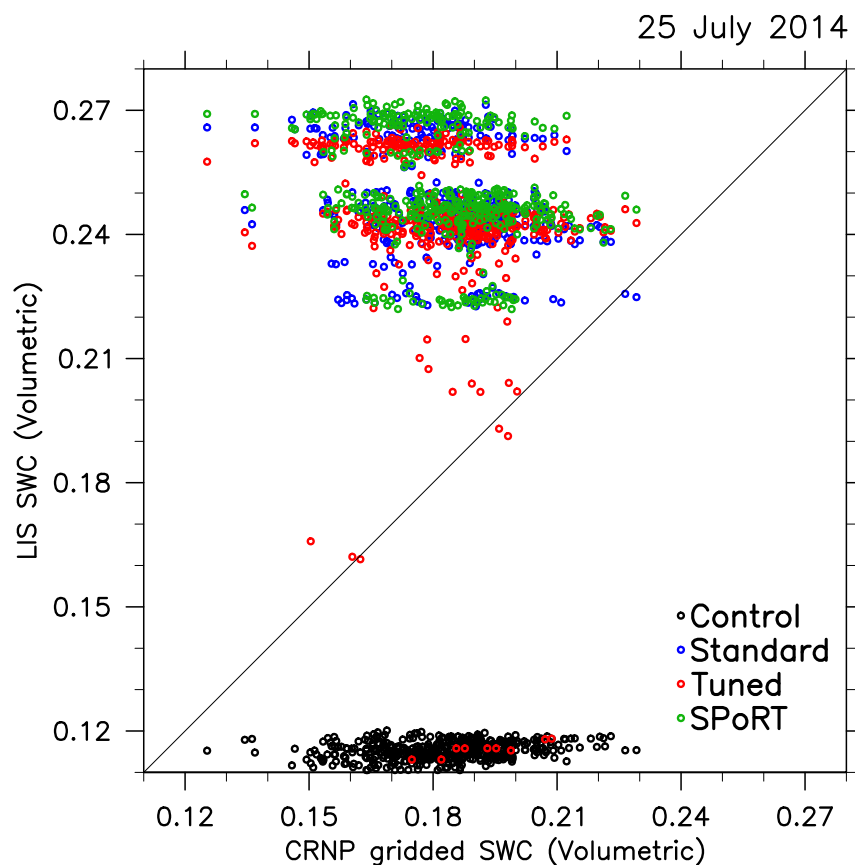
The description of the CRNP gridded soil moisture product and the alterations made to the regression for this study have been moved to the new Evaluation Data section (3.2)

17) Page 13 L15: Not clear what modifications were made to the COSMOS product; provide a section reference or more details here. Also a bit confused about the references to both CRNP and COSMOS as they are presumably the same thing?

COSMOS is the observing network, CRNP is the instrument. COSMOS was a typo here and has been corrected to 'CRNP'. More description has been added about the changes to the spatial regression and they've been moved to the new Evaluation Data section (3.2).

18) Page 13 L14-15: I wonder if a non-cumulative PDF wouldn't be better in this context?

This comment echoes that of reviewer 3 in that this information could be presented in a more effective manner. This figure has been changed to a scatterplot:



19) Page 14 L6: I believe that the GVF is provided at 3 km (and 16 km) rather than 1 km resolution, correct?

The SPoRT GVF is provided at 3 km, but the climatological GVF is provided at 1 km. Please see comment #10.

20) Section 5: The discussion is very brief and lacks more substantial and high quality discussion elements on limitations, challenges and opportunities.

A paragraph has been added to the Discussion that addresses the concerns of reviewers 2 and 3 related to the choice of meteorological forcing dataset. An additional paragraph has been added discussing the potential limitations of the uncoupled configuration used in this study.

Other limitations of the study are presented in the discussion Page 16, paragraph 2 and 3. Challenges are discussed extensively in the Background section. The future of irrigation intensity datasets is detailed in Page 17 paragraph 2.

21) Page 15 L3-8: These are useful details that should have been provided in the methodology or result sections

A description of the triggering datasets and exactly how they impact triggering is included in the methodology section (Page 10 Lines 1-17). The relative importance of the triggering datasets is included here, not in the methodology, because this is a main finding of the study.

22) Page 15 L9-12: Not sure I understand this correctly, particularly the part about the scaling by GVF being more important than changes in rooting depth.

The logic here is as follows. First, the maximum crop root zone is multiplied by the GVF (non-dimensional number 0-1) to mimic a seasonal cycle of root growth. The amount of water added by the irrigation scheme is then dependent on the depth of the crop root zone (more water applied for crops that have deeper roots). To determine the potential impact of crop rooting depth specification, we completed an additional run where we used an observationally tuned crop map and changed the maximum root depth of maize and soybeans. It was concluded that the impacts of the crop root depth on irrigation amounts and fluxes were insignificant compared to the influence of the scaling of the crop root zone.

23) Page 15 L13: The method for determining the start and end of the growing season hasn't been described anywhere, but it must be. Justifications for adopting that methodology (rather than relying on existing phenology products for instance) should also be provided.

The details of the irrigation season have been added to the method section when first introduced. Page 10, Lines 3-4 now reads:

“The growing season, addressed in question three, is a function of the gridcell GVF (i.e., 40% annual range in climatological GVF; Ozdogan et al. 2010)..”

This method is used as it is a main feature of the Sprinkler irrigation algorithm. Please see comment 12.

Technical corrections:

1) Page 4 L1: “with a two different..” - should be “with two different..”
This has been changed.

2) Page 4 L23: “..water resources region. . .”?

This has been reworded to:

“reproduce irrigation water usage within counties and water resource regions, respectively”

3) Page 5 L14: use “high resolution” rather than “high-resolution”

4) Figure 5: I would also show the irrigation amounts here as done in Figure 7. Why is the impact of irrigation high when no irrigation is applied (e.g., during rain events)?

The observed irrigation amounts are not shown because this figure is used to analyze only model results/datasets, not observations. It would be possible to show simulated irrigation amounts for all irrigation runs, but that would make the figure much more confusing/busy without contributing additional information. We feel that the combination of forcing precipitation and flux changes due to irrigation already make it readily apparent when irrigation is being triggered.

As compared to the rain-free periods, the impact of irrigation is dramatically reduced during rain events. There is still some impact to fluxes during rain events in the summer because the soil column in the irrigated simulation is generally wetter than control due to the memory of previous irrigation, even if irrigation does not occur on that day.

5) Figure 5: Issue with the legends – they are not consistent with what is shown; currently I can only distinguish two different line styles.

This figure shows changes from control in each model configuration for latent and sensible heat fluxes. Latent heat flux changes are shown in blue and sensible heat flux changes are shown in red. The line style corresponds to the model configuration. Therefore, the change from Control in latent heat flux when using irrigation and the SPoRT GVF dataset is shown in the blue dotted line. Only two lines are distinguishable because the Tuned and Standard configurations do not differ enough from each other at this scale to be distinguishable. This is a main conclusion shown in the figure.

6) Figure 5: a and b rather than top and bottom should be used for more precise figure referencing in the manuscript. This also applies to the other figures.

All figures have been updated to use the (a),(b), etc.