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 #3:

I. Summary

This manuscript examines the issue of developing and validating realistic irrigation schemes for use in land surface models (LSMs). In this study, the authors utilize observation-based datasets of irrigation intensity and green vegetation fraction (GVF) to tune the LSM irrigation amounts, which are validated against data obtained from Cosmic Ray Neutron Probes (CRNP). The main conclusion of the authors is that the timing, amount, and spatial spread of irrigation are more sensitive to the choice of irrigation scheme at smaller spatiotemporal scales than at larger, more typical scales for regional climate models. Given the balance of evidence presented and the use of a novel dataset (CRNP) for addressing this issue, it seems that the authors have arrived at robust and meaningful conclusions that would be worthwhile additions to HESS and to the field of hydrology, in general. While I have no major qualms with the content or substance of the manuscript, I do present below some more minor comments for improving the robustness and presentation of the results.

II. General comments

A. NLDAS-2 – The authors mention several times throughout the manuscript the need for "high-quality" meteorological forcing and point out repeatedly the accuracy of the precipitation data from NLDAS-2 for their domain. While it certainly seems that NLDAS- 2 provides accurate forcing over this domain (and is a high-quality dataset, in general), I echo Reviewer 2 in cautioning against drawing far-reaching conclusions about NLDAS- 2 from this limited study. The entire study domain is 15 x 15 km, very small even for typical regional climate model simulations; the entire domain would fit in 4 grid cells of NLDAS-2 (1/8 degree horizontal resolution). Is there evidence that NLDAS-2 would provide equally accurate data for a different domain within the same region, or in a different region or year? If so, then I would provide a sentence or two explaining the skill of NLDAS-2 over the general region (e.g., Great Plains/Midwest) during the growing season (perhaps from the Xia study). If not, then please temper the language emphasizing the high quality of NLDAS-2 with the understanding that the spatial domain of this study is extremely limited and that NLDAS-2 may not be as accurate in other agricultural regions in North America.

As per reviewer 2's suggestion, an additional run with GDAS forcing was completed and a brief description of the results is now included in the Discussion section (Page 15, Lines 9-15 to Page 16 Lines 1-5). In this newly added paragraph, we've taken care to add qualifiers to the NLDAS2 statements. The section now reads as follows, with bolded words to emphasize the tempered language about NLDAS2:

> "**For this small domain**, the NLDAS2 precipitation proved to be sufficiently accurate, matching well that given by the nearby York, Nebraska AWDN. **However, for other regions reliable meteorological forcing may not be available.** To further explore the impact of the forcing precipitation on the irrigation triggering, an additional simulation was completed that is equivalent to the Standard irrigation run in all aspects (e.g., GRIPC irrigation intensity, climatological GVF) except that the Global Data Assimilation System (GDAS) meteorological forcing is used rather than NLDAS2. In contrast to NLDAS2, GDAS is coarser resolution (1/4 degree) and does not include rain-gauge corrections. Results show that GDAS supplied a greater amount of total of precipitation in May through July 2014, creating a wetter soil column and prohibiting irrigation triggering in mid-to-late July, in contrast to observations and the other irrigation simulations. As a result, the soil moisture dynamics of the GDAS simulation at the maize site differ substantially from the CRNP observations and the NLDAS2-forced simulations. **These results underscore the need for highest quality datasets available for the area of interest, which for this region and time frame was NLDAS2."**

III. Specific comments

A. Page 14, line 10 – "These results suggest that if this domain were one gridcell in a larger, coarser resolution domain (e.g. 15 km spatial resolution), the variation in the gridcell soil moisture (given here by the domain average) over the growing season would be representative of observations."

It would be interesting to see a supplemental model analysis with coarser-resolution grid cells (either in this paper or a future one) that validates this hypothesis. For example, what is the spatial threshold at which large-scale forcings begin to dominate the changes in the soil moisture signal?

We agree that this is an interesting question and appreciate the suggestion! This will certainly be an area of future study using the flexibility of the LIS system (resolution, forcing, and inputs).

B. Page 15, line 9 – ". . .indicating that the model is quite insensitive to the maximum root depth change. . ."

Some common irrigated crops, such as alfalfa, have max root depths of 2+ meters. Though irrigated alfalfa is much less common in Nebraska when compared to corn and soybeans, it would be instructive to not make the above claim about the insensitivity of the model to max root depths unless other crops with much larger or smaller max root depths have been tested.

This sentence has been rephrased to emphasize that the root depth sensitivity tested was only for a small change to a specific crop:

> "The results of this analysis showed little difference between this simulation and the others, indicating that the model is insensitive to small changes (up to 20%) in the maximum root depth. However, land surface models that have a more complex treatment of crops, study areas with greater heterogeneity of crop types, or experiments that replace a particular crop with one that has a vastly deeper root system, are examples beyond the scope of this study that could potentially result in a greater sensitivity of the model results to crop root depth."

C. Page 15, line 22 – ". . .a growing number of options for irrigation intensity datasets in the coming years".

A new global irrigation dataset (the Historical Irrigation Dataset) was published through HESS rather recently (S. Siebert, M. Kummu, M. Porkka, P. Döll, N. Ramankutty, and B. R. Scanlon (2015), "A global dataset of the extent of irrigated land from 1900 to 2005," Hydrology and Earth System Sciences. DOI: 10.5194/hess-19-1521-2015). It may deserve a citation here because of its recent development and global coverage.

It is certainly appropriate and has been added.

D. Figure 1 – Are the spotty areas of low irrigation intensity in the Tuned plot over urban areas? A brief explanation of this in the text may be warranted.

The spotty areas indicate the irrigation intensity has been reduced due to the presence of roads, wetlands, rainfed fields, and/or buildings. Of the three gridcells with 0% irrigation intensity, two contained mixed-use land, small buildings, and roads (though, not built up enough to really be considered 'urban'). The remaining 0% irrigation intensity gridcell contains the rainfed site given in the CRNP observations.

The figure caption has been updated as follows:

> "**Figure 1.** (a) GRIPC irrigation intensity (percent) given by Salmon et al. (2015) used in the Standard and SPoRT simulations and (b) the observationally tuned irrigation intensity used in the Tuned simulation. **The spotty nature of Tuned indicates irrigation intensity has been reduced due to the presence of roads, wetlands, rainfed fields, and/or buildings**. Also shown is the average greenness vegetation fraction (unitless) in July 2012

given by (c) NCEP climatology used in the Standard and Tuned simulations and (d) SPoRT real-time dataset used in the SPoRT run."

E. Figure 2 – It would be helpful to mention in the figure caption that SPoRT uses the climatological GVF in years 2009 and 2010 (as is already mentioned in the text) to avoid confusion.

The caption has been updated to include this information (bolded) and now reads:

> Figure 2. Domain and monthly averaged GVF from the NCEP climatological GVF dataset, used in the Standard run, the SPoRT GVF dataset used in the SPoRT run, and the difference between the two (SPoRT – Climatology). **As the SPoRT dataset is not available prior to 2010, the long-term SPoRT simulation uses climatological GVF for 2009-2010, and the SPoRT GVF dataset is incorporated in December 2010 and used throughout the rest of the simulation."**

F. Figure 4 – I don't believe that IRR was ever defined (in either the main text or the figure caption).

For all captions, "IRR - Ctrl" has been replaced with (i.e., each irrigation run minus Controll). The legend have been updated in Figures 7 and 9 so that they don't include IRR and are consistent with the other legend labels.

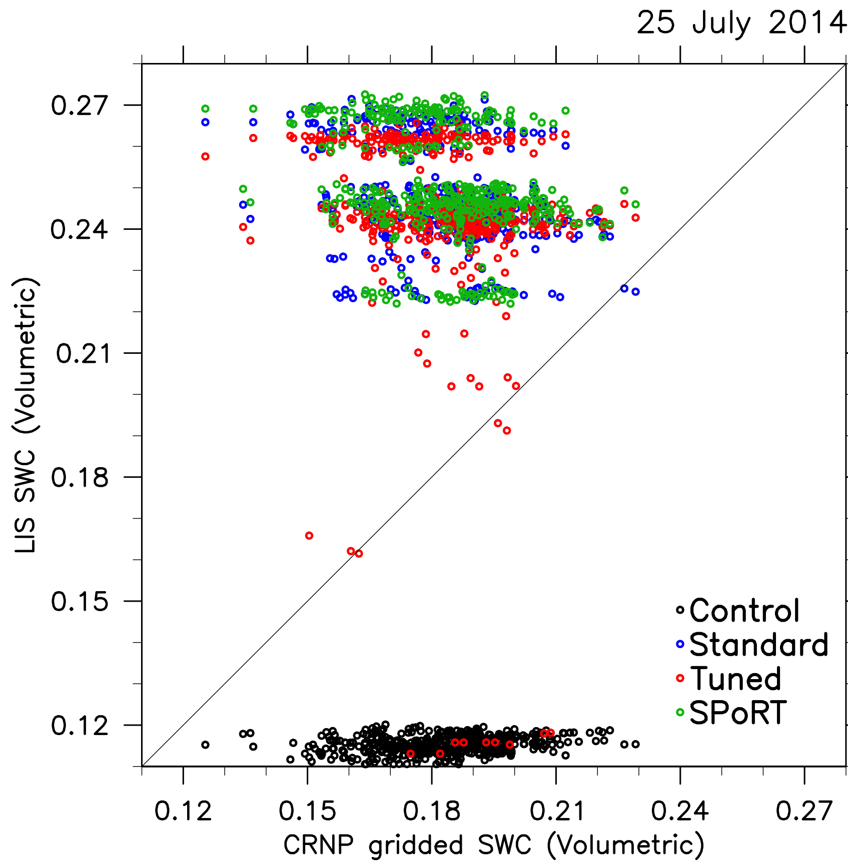
G. Figure 4 – The boundaries of Layer 4's soil depths are only mentioned here, not in the main text. Since crop roots barely extend into this layer (max root depths of 1 or 1.2 m), perhaps this further explains why there seems to be much more variability in soil moisture between irrigation simulations in Layer 3 than in Layer 4.

Yes, the reviewer is correct. To call attention this fact, Page 11, Lines 8-9 have been updated to read:

> "Increases in the third soil layer, **which includes the root zone**, are quite consistent annually with a near doubling of the soil moisture when irrigation is turned on."

H. Figure 8 – I think that the presentation of "spatial" CDFs in this figure is rather non- intuitive. To me, it would be much more intuitive to see the differences in the spatial distributions of soil moisture within the domain using a histogram, especially since each CDF is plotted for only a single time step and thus there is no "accumulation" of data over time. In this figure, since data is accumulated spatially (in two dimensions) rather than temporally (in one dimension), the shape of the CDF would be rather arbitrary and would partly depend on the order in which you spatially sample the grid cells.

Thanks much for this suggestion. This figure has been changed to a scatterplot of the gridded observations versus the LIS simulations:



The text has been updated accordingly and all mentions of 'temporal CDF' have been changed to 'CDF':

> "The LIS-simulated soil moisture variability in time and space is evaluated against the CRNP gridded soil moisture product, described in Sect. 3.2. The spatial variability is assessed first with a scatterplot generated using all gridcell soil moisture values from the LIS simulations and the modified CRNP product aggregated at 4, 12, and 20 UTC on 25 July 2014 (Fig. 8). Next, the temporal variability is assessed using a CDF of the domain-averaged soil moisture values from May 5 to Sept 22 at 8-hour intervals (Fig. 9).
>
> Figure 8 shows that the Control simulation does not match the observations in magnitude or variability, instead showing uniformly dry soil across the domain (e.g., range of 0.01 versus more than 0.1 in observations). The spatial variability is increased in the irrigated simulations, but these runs exhibit jumps between clusters of values as a result of irrigation triggering and dry down across the domain. The different levels of clustering shown by the irrigated simulations are a result of the input parameter datasets, as triggering and timing are dependent on these datasets. Although the Control

simulation is too dry, the irrigation overcompensates and increases the soil moisture to levels beyond that shown in the gridded observations. These results suggest that the model, even with the irrigation algorithm turned on, is not able to accurately simulate the small-scale (i.e., field scale) heterogeneity in soil moisture that is present in the CRNP data…"

I. Figure 8 – Neither the figure nor the figure caption explain what is being plotted in the figure. Units would also be appreciated (even if unitless).

Please see notes above notes about the updated Figure 8.

IV. Technical corrections

All of the following technical corrections have been made as suggested. We thank the reviewer for his/her attention to detail.

With respect to comment "S" below, the text has been changed to:

"In this study, we modified the spatial regression technique to treat irrigated and non-irrigated areas differently by using the CRNP rainfed values in the regression for non-irrigated gridcells and the average of the irrigated CRNP probes for the irrigated gridcells.

A. Page 1, line 17 – "at the interannual scale, but become. . ." – Remove the comma.
B. Page 2, line 23 – "previous evaluation efforts, and introduces..." – Remove the comma.
C. Page 3, line 14 – e.g., "de Vrese et al. 2016" – Please be consistent with placing commas after "et al." in internal citations.
D. Page 4, line 1 – "with a two different. . ." – Remove "a".
E. Page 4, line 2 – "in the U.S. Central Great Plains. . ." – "Central" should be lowercase.
F. Page 4, line 5 – "Tuinenburg et al., 2014), or in. . ." – Remove the comma after "2014)"
G. Page 4, line 15 – No need for commas surrounding "such as these".
H. Page 4, line 23 – ". . .to reproduce county and water resource region irrigation water usage. . ." – Change to ". . .to reproduce irrigation water usage within counties and water resource regions. . .".
I. Page 5, line 17 – Change "c.f." to "cf.".
J. Page 5, line 19 – "reliable, area-average soil water content" – Throughout the manuscript, please change to "area-averaged" or "domain-averaged" (as in the above example) when being used as an adjective and "area average" and "domain average" when being used as a noun.
K. Page 6, line 9 – Change to "Sect. 3".
L. Page 7, line 22 – "i.e. observationally tuned" – Change all instances of "i.e." and "e.g." to "i.e.," and "e.g.,".
M. Page 8, line 8 – "more sophisticated, but computationally expensive. . ." – Remove the comma.
N. Page 8, line 8 – "such a dynamic. . ." – Change to "such as".
O. Page 8, line 14 – Change to "bias-corrected".

P. Page 11, line 14 – "the SPoRT run increases latent heat flux by more than 100 W mˆ-2 more than Standard" – Change to "latent heat flux in the SPoRT run is more than 100 W mˆ-2 greater than Standard".
Q. Page 12, line 15 – Add a space between "mm dayˆ-1" and "(not shown)".
R. Page 12, line 25 – Add a comma after "(e.g., satellite)".
S. Page 13, line 11 – "CRNP (irrigated) rainfed data. . ." – I would discourage this par- enthetical style (it already seems to have confused other reviewers). If you must use it, I would recommend putting the parenthetical expression second, e.g., "CRNP irrigated (rainfed) data". However, I would instead prefer this and related sentences to be written as: "by using the CRNP irrigated and rainfed data in the regression for irrigated and non-irrigated gridcells, respectively".
T. Page 13, line 23 – Add a period after "dependent on these datasets".
U. Page 13, line 25 – Change "exhibit" to "exhibits".
V. Page 14, line 5 – Hyphenate "deficit based".
W. Page 14, line 11 – Hyphenate "coarser-resolution".
X. Page 16, line 3 – Remove the comma after "LSM framework".
Y. Page 16, line 4 - Remove the comma after "latent heat flux".
Z. Page 16, line 21 – Remove the comma after "soil moisture".
AA. Page 16, line 23 – Change to "USDA Census of Agriculture".
BB. Page 17, line 1 – Hyphenate "satellite based".
CC. Page 17, line 2 – Add period after "(Kumar et al., 2015)".
DD. Page 17, line 4 – Change "premiere" to "premier".
EE. Page 17, line 23 – Capitalize "a" after Myhre, and ditto for all other instances of mixed case for author names in the reference list.
FF. Page 18, line 8 – Be consistent with italicizations: Either italicize all journal names or keep them all as plain text.
GG. Page 18, line 8 – Change "hess" to "HESS".
HH. Page 18, line 28 – What does "Received" mean?
II. Page 19, line 3 – Be consistent with capitalization of the article titles: Either capitalize only the first word and proper nouns (standard practice) in every title or capitalize all words in every title.
JJ. Page 20, lines 5-9 – I think that these lines are in a slightly different font than the other references.
KK. Page 21, lines 2-3 – See above comment.
LL. Page 21, line 23 – What is "Artn"? Article number?
MM. Fig. 1 caption – Please define the units of irrigation intensity (even if unitless).
NN. Fig. 4 caption – Add a colon after "LSM default layers".
OO. Fig. 4 caption – Be consistent with parenthetical notes: Delta Z is included for the middle layers but not for the top or bottom layers.