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 #2:

Summary:

The authors provide a useful and clearly-written evaluation of irrigation simulated by an advanced Land Surface Model. These types of evaluation are in short supply, and the use of CRNP in model evaluation is, to my knowledge, novel and potentially quite useful. I believe that the Discussion Paper is of sufficient interest and quality for publication in HESS. That said, the numerical experiments presented in the study are rather limited. Sensitivity to GVF dataset and irrigation intensity factor are evaluated, but none of the many other factors that the authors list are explored. This may lead to the wrong impression that the tested factors are the most important when simulating irrigation, when I see no evidence presented by the authors that this is in fact the case. Ideally, the authors should present a more inclusive set of sensitivity tests to inform future modeling studies about the relative importance of different factors. If this is not possible, or if the authors view it as unnecessary, then a more convincing justification for the choice of experiments is required.

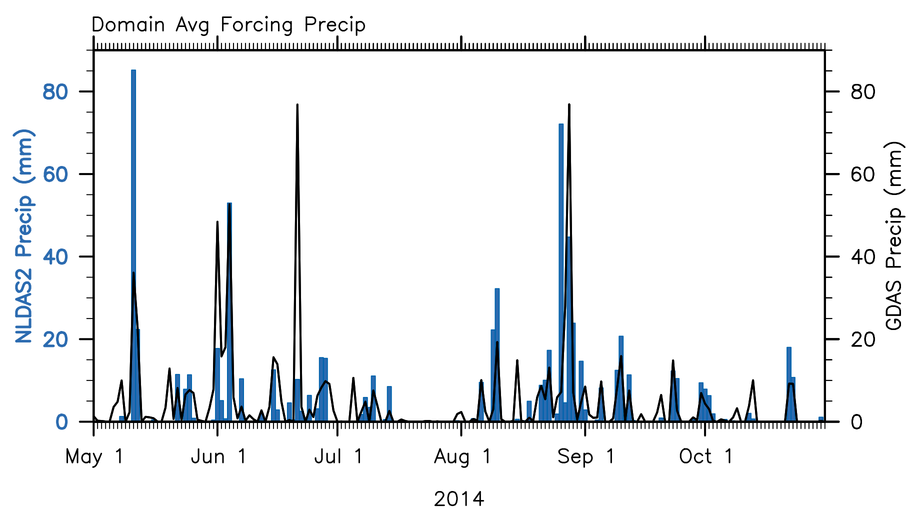
General Comments:

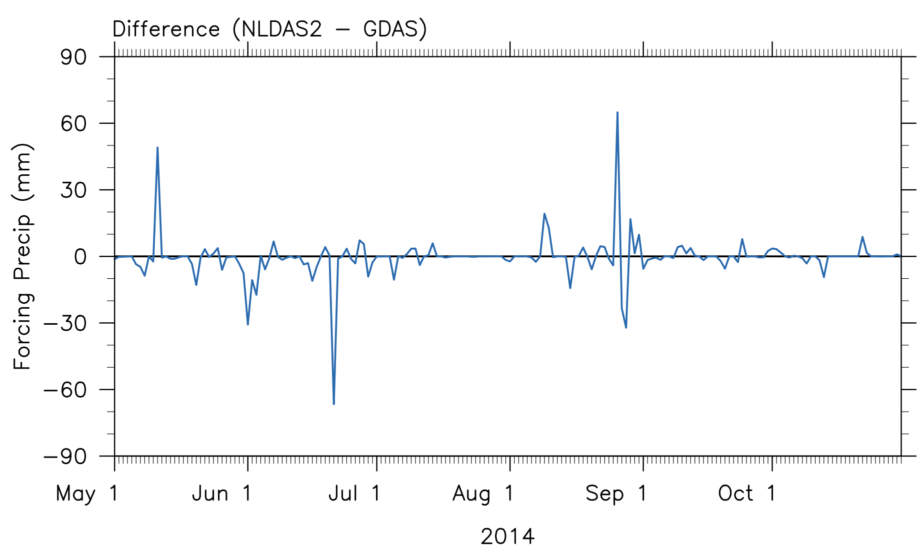
1. Meteorological Forcing: In the abstract and at several other passages in the text the authors emphasize the importance of high quality meteorological forcing data for accurate simulation of irrigation. Their results suggest that NLDAS is high quality, as shown most convincingly by the temporal match of simulated irrigation to spikes in observed soil moisture. I believe that NLDAS is high quality and that these results show impressive performance at local scale. But I'm not sure that the authors can actually make any conclusions about the importance of forcing data to irrigation simulations, given that they do not compare NLDAS simulations to simulations with any lower quality forcing dataset. Yes, it its intuitive that simulations with NLDAS will be better, but the numerical experiments don't demonstrate this, and they don't show us *how* important it is. This is particularly the case when one considers spatial or temporal scale. The authors nicely demonstrate that simulations are more realistic at larger and longer scales than they are at local and shorter scales. How important is meteorological forcing if we are concerned with large and long time scales? Additional simulations with an alternative, poorer quality meteorological

forcing dataset would be the obvious way to test this, but the authors might find other ways to make the point.

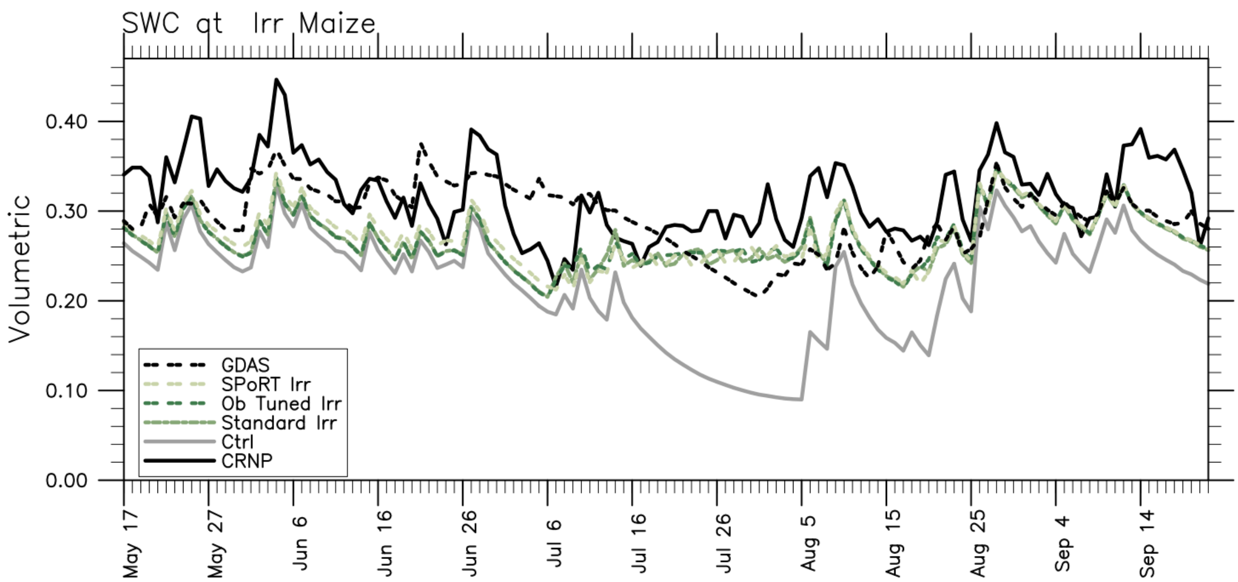
The foundational study for this work, Ozdogan et al., (2010), evaluated this scheme at larger (continental U.S.) and longer (yearly) time scales with annual water withdrawals and county level data. For this study, the primary interest is in evaluating the scheme performance at smaller and shorter timescales, so a robust evaluation of the meteorological forcing at large and long timescales is beyond the scope of this work. With respect to the support for the NLDAS2 conclusions, however, the reviewer raises some good and justified questions.

In response, we have completed an additional run equivalent to the Standard irrigation run in all aspects (e.g., GRIPC irrigation intensity, climatological GVF) except that we used GDAS meteorological forcing instead of NLDAS2. GDAS is coarser resolution (1/4 degree) and does not include rain-gauge corrections. GDAS supplies a greater total amount of precipitation in the May through July time period. See figures:





The greater total amount of precipitation from GDAS results in a wetter soil column leading up to and throughout the mid-to-late July rain-free period, delaying the onset of irrigation triggering by the scheme. As a result, the soil moisture starts out wetter in mid-July than the other irrigation simulations (forced with NLDAS2) and even the CRNP, then dries out to a level below that of the other schemes (as a result of moisture being sustained in the root zone and prohibiting irrigation). The irrigation is finally triggered at the beginning of August, a few days prior to the return of precipitation to the area. See figure below (top layer soil moisture):



This simulation adds support to the conclusion that accurate precipitation data is essential to constrain the irrigation triggering. A brief description of this additional run has been added to the discussion section.

The newly added part of the Discussion (Page 15, Line 19 – Page 16 Lines 1-5) reads:

> "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."

2. Thresholds: The authors appropriately emphasize the importance of selecting proper thresholds for soil moisture and GVF at several points in the text. But the manuscript does not offer any evaluation of either. In both cases a single threshold is applied and attributed to previous studies. It would be quite interesting to know how the impact of using different GVF datasets compares to differences caused by small changes in GVF threshold. And how does a modest change in threshold impact total water use, as compared to the tested sensitivity to prescribed irrigation intensity?

The sensitivity of the irrigation scheme to the soil moisture and GVF thresholds has already been examined in the Ozdogan et al., (2010) for a larger area that includes our study region. The 50% of field capacity soil moisture triggering threshold was selected by their study as being most appropriate based on discussions with local experts, including some in Nebraska, as well as through trial and error (Ozdogan et al., 2010). As this is the same scheme used here, we didn't consider it necessary to re-test the SM threshold and instead accepted it as being the best for this region based on current literature. The accurate timing of irrigation triggering shown in the results supports that this threshold was reasonable.

Although the gridcell GVF ***value*** is used to calculate the crop root zone and to scale the amount of water applied, the GVF ***threshold*** is only used to determine the start and end of the irrigation season. As a result, a small change in the GVF ***threshold*** would only increase or decrease slightly the length of the irrigation season. The GVF threshold for our region gives an appropriate irrigation season of June – September, so we didn't consider it necessary to change this threshold at all.

I understand that no study can be comprehensive on all parameters, but I don't fully understand why the authors chose to look only at GVF dataset in GRIPC irrigation intensity when other subjective modeling decisions might have as large or larger impacts on the simulations. If possible I would encourage the authors to expand their sensitivity test in order to justify the selection of these two factors as the focus of study.

The main objectives of this study were not necessarily to turn every knob, but instead to take the best available collection of default datasets we have (e.g., those that someone new to model would probably choose) and to see how well it performs (i.e., the Standard run). Then secondarily, to determine if it is possible to improve upon that standard model performance by either 1) incorporating additional information to tailor the datasets to our study area (Tuned irrigation intensity), or 2) by using a new and improved GVF dataset (SPoRT) that detects vegetation response to soil stress. Rather than a blanket sensitivity study, these were targeted in areas where we knew we could improve the model/datasets based on solid information.

The focus on irrigation intensity and GVF datasets for potential improvement to model performance is two-fold:

1) Irrigation intensity and GVF are critical to **both** the ***triggering*** of irrigation and the calculation of the ***amount*** of irrigation water applied. As a result, flaws in the scheme could be made more apparent by switching out these datasets. Additionally, these two datasets (SPoRT GVF and GRIPC irrigation intensity) are brand new and have not been used with an irrigation scheme until now.

2) The other datasets that play a role in irrigation triggering, (i.e., landcover, soil texture, soil type, crop type) were by default homogeneous across the study area and were appropriate for the area based on the ground truth we had. For example, the landcover for every grid cell in the domain was 'croplands'. At 1 km resolution, there is not a better classification of these gridcells than cropland (e.g., even the gridcells that contain small buildings or roads still occupy < 50%; croplands is dominant land use). Similarly, we didn't have additional information to be able to improve upon the default soil type or texture. With regards to crop type, the data from Franz et al., (2015) showed 81% maize and 19% soybean, in contrast to 100% maize in the default crop type map. As a result, we did an additional run with tuned crop type and altered max root depth. The results of this run are presented as a note in the Discussion (Page 16, Lines 12-21) rather than featured prominently. This was done with the intention of simplicity (i.e., to minimize confusion that could be caused by introducing another iteration) because this run was not significantly different than the other irrigation runs.

Minor Comments:

Page 3, line 20: This list of options misses flood irrigation simulation (unless it's supposed to be covered by #1). Several studies have employed flood irrigation, including Yilmaz et al. (2014), Leng, and Evans & Zaitchik (2008).

The intent was for flood to be covered by #1. The sentence has been edited to clarify:

> Page 3, Line 20: "1) defined increases to soil moisture in one or more soil layers (Kueppers and Snyder, 2011; de Vrese et al. 2016)., sometimes referred to as flood (Evans and Zaitchik, 2008),"

Section 2.3: It would be useful to include a sentence or two on why CRNP measure- ments are sensitive to soil moisture. Many readers (myself included) are not deeply familiar with this technique.

An additional sentence has been added to section 2.3 (Page 5 Lines 18-20) addressing this point (bolded):

> "**The theoretical basis for the CRNP method follows that fast neutrons injected into the soil by the CRNP will be slowed more effectively by collisions with hydrogen atoms present in soil water than by collisions with any other element (Visvalingam and Tandy,**

1972). Thus, the neutron density measured by the probe is inversely correlated with soil moisture…”