Reviewer #1

This is an interesting paper addressing the hot topic of providing high quality and timely agricultural drought forecasts information in a region hit by frequent droughts and famines. The methodology follows and compares with earlier work and similar approaches,
although the exact objective and added value of this approach as compared to existing and similar ones should be given some more attention. However setion 2.3 is extremely difficult to follow and clearly needs more explanations (see detailed comments
below). By reading the methodology again it is not completely clear whether the main envisaged result of the paper are the soils moisture forecasts or the derivation of daily rainfall amounts from cumulated seasonal forecasts and despite the complete introduction a clear statement of the objectives is missing. And whatever is the main objective, the benefit of the expected results on agricultural drought forecasting has to be better explained as well. Eg. are we talking about drought occurrence only or also drought impact? Finally it could be interesting to compare the results with some remote sensing derived Soil Moisture product.

Response: Thank you for your valuable comments! We have rewritten the section 2.3 in the revised version of the manuscript and also specifically mentioned the objectives of this study. The main focus of this manuscript is indeed SM forecasts however rainfall forecast is a crucial piece of that. We also now better explain VIC based soil moisture forecasts can be useful for agricultural drought forecasting and compare the results with a multi-satellite based soil moisture product.

Specific comments:

1. P3051 L10-14 I think it's not completely appropriate to put the need for early warning
   systems and early response following the 2011 famine on the same level. In general there is a high consensus that early warning systems worked relatively well while the lack of appropriate and timely response was one of the main reasons that lead to the famine.
   Response: We agree with you and have revised that statement accordingly.

2. Fig 1. CHIRPS appears here first, introduce the acronym.
   Response: Done!

3. P 3052 L5-15 I understand that you take SM (for crop areas only?) as a direct predictor
   of agricultural drought, as opposed to rainfall only which would be meteorological drought. Can this be stated more clearly? Also later on SM is compared directly with the WRSI, but could SM be used to improve the WRSI model? Also maybe worth to mention that, with the exception of limited areas in Southern Somalia, the study area is predominantly a pastoral area.

Response: We have revised those lines based on the reviewer's suggestion. Yes, SM could be used to improve the WRSI model. This was the focus of a study led by one of the co-authors (Dr. McNally) and has been described in McNally et al., 2014 (current in review in JHM special issue "SMAP early adopters"). We now mention that above the study area being predominantly a pastoral area.

4. P 3052 L29 and following. What is the link between the statement "Reliable rainfall forecasts over .." and the following one? What rainfall forecasts are meant? Long term climatological forecasts or seasonal forecasts? Is the debate concerning only the causes of the decline in rainfall or the decline itself? If seasonal forecasts are meant this would have a negative impact on the relevance of the paper since it is not clear how soil moisture forecasts based on debatable rainfall forecasts are expected to improve the final drought forecast. So please specify.

   Response: We refer to seasonal rainfall forecasts here (We have now revised that sentence to clarify that) here however that statement stands true for decadal scale rainfall forecasts as well. Through that statement we wanted to reiterate what previous studies focusing on MAM seasonal rainfall forecasts have concluded. The rainfall forecasting approach used in this manuscript does result into skillful rainfall forecasts so we expected SM rainfall forecasts to be skillful as well.

5. P 3053 L13 It would be nice to briefly outline the main differences with the mentioned
   approaches and also explain what is exactly the expected benefit of introducing an additional similar approach.

   Response: Good suggestion. We have added a few sentences describing the differences between our approach and previous similar approaches and also the need of developing such system in section 5 (please see lines: 472-498). The two primary differences between our approach and others are:
   (1) We use dynamical forecasts over Indo-Pacific region (as shown in Fig. 3) to bias correct dynamical precipitation forecasts over EA region, whereas in other approaches dynamical forecast over the domain itself is downscaled and bias corrected. Since the skill of dynamical forecast for the MAM season and EA region is negligible, forecast that is directly downscaled also have negligible skill. Whereas we show that through our approach useful precipitation forecast skill (0.67 as shown in Fig. 7) can be attained over the EA region for MAM season.
   (2) We use a rainfall dataset that has been recently developed and takes advantage of both satellite based precipitation estimates and stations data.

   Please also see the last two paragraphs of section 4.

6. P 3054 L 10-11. Was a land cover classification used to assign some fixed values to each vegetation type? It is not mentioned in the following description.

   Response: Yes. The vegetation parameters that we used was based on UMD AVHRR vegetation classes. We now mention that in the manuscript.

7. Fig 3. Can you explain why the correlation figure is covering nearly the whole globe? It is not clear how the correlations outside the study area, Eg. in the ENSO area, are used for the following steps of the analysis

   Response: We look for the correlation between the MAM observed rainfall and CFsv2 precipitation forecast across the globe to find a large scale teleconnection pattern. We eventually mask out the grid cells that have small correlation with EA MAM observed rainfall and use the first principal component of the correlation scaled CFSv2 precipitation forecasts over rest of the grid cells for generating weights as described in step 3 of the section 2.3.
   There are two main reasons for using a larger area: (1) strong teleconnection between precipitation over Indo-Pacific region and East Africa rainfall over MAM season (2) higher skill of dynamic forecast models over Indo-pacific ocean than over terrestrial regions of East Africa.

8. Section 2.3. It is really hard to a fully understand the method presented in this section. I suggest that this section is deeply revised.

   8.1 Some ideas: - Introduce the section by stating what is the general purpose of the production of seasonal climate scenarios (producing daily sequences of rainfall forecast from CFSv2 seasonal forecast - I assume it's a single map updated time to time,

   Response: Done.

   8.2 Explain clearly what time of forecast you get from CFSv2 and how you treat the fact that they are dynamical forecasts).

   Response: Done.

   8.3 EA MAM rainfall is compared to CFSv2 precipitation forecasts at global scale? - Point 2 and 3.
   Response: Yes, we now specify that in the manuscript.

   8.4 Always describe on which domain the computations are performed (EA or Global).
   Response: Done.

   8.5 Explain why you use the absolute value. Negative correlations are considered equally important as positive ones?

   Response: Yes, negative correlations are equally important as positive ones. We wanted to focus on those grid cells that in the analog domain that had strong correlation with the EA MAM rainfall regardless of the sign.

8.6 Explain which similarity metric you use.

Response: We now explain this in the step 3 and 4 of the section 2.3. The metric used was the distance between the forecast and observed seasonal precipitation total. The inverse of these distances were used to produce a set of sampling frequencies that summed to 1. Please see step-3 and 4 for further details.

8.7 Explain clearly how the daily sequence is produced staring from all other years daily sequences.

Response: We have now clarified that in the manuscript. Please see steps 5 and 6.

1. To generate daily climate scenarios we start with the final weights W_f mentioned in step 4. We use these weights to guide the probability of selection during the bootstrapping process (following the methods described in Husak et al., 2013) from the observed MAM precipitation over the EA domain during the hindcast years (1993-2012). The years with higher weights get selected more often than other years because the frequency of selection is proportionate to the weights. We first perform this bootstrapping process for the first dekad of MAM, comprised of 10 daily values of precipitation and temperature maximum and minimum. In order to build the scenarios for the first dekad of the MAM season for any target year, we sampled the first dekad of the MAM season from all years (1993-2012, except the target year) as described previously.
2. We then repeat this process for subsequent dekads of the MAM season. For example, Fig. 4 shows the frequency of years in the available record (1993-2012) picked in generating 100 climate scenarios for the MAM season of the year 2011, which was a drought year. Based on our estimates, year 2011 was most similar to the years 2009, 1999, and 2000, which were all drought years. Beyond the MAM season our bootstrapping selection is based on the equiprobable weights (similar to climatological forecasts).

8.8 If you are doing a weighted average of all annual profiles, you might have small rain contributions from a large number of days. How is this deal with? More information is needed to really understand what is done here

Response: No, we are not doing weighted average of all annual profiles, we instead use the value of similarity matrix to assign the probability of selection from different annual profiles. For example if an year X has higher probability than an year Y then X was selected in the boot-strapping scenarios more often than the year Y. Since we picked one full dekad at a time the frequency of rainy days is not from its climatological values. We have now clarified this in the manuscript as well (see section 2.3)

8.9 Fig 4. Bottom panel refers to a uniform distribution? If this is the case it is not really adding information, I suggest to omit it. In addition, if uniform, why is the frequency not exactly the same? - Fig 4. Revise English of the caption.

Response: We have removed the bottom panel and also revised the caption of Figure 4.

9. Fig 5. This figure adds very little to the simple statement in the text. I strongly suggest
to omit it. A more meaningful one should be constructed with actual model runs.

   Response: We have now removed that figure and since figures 8 and 9 already broadly convey our message about the variability of forecast skill during a season, we have decided to not replace this figure with a figure based on actual model runs. Please note that we have added 3 additional figures in the revised manuscript already.

10. Section 3. I suggest to emphasize that the initial comparison with WRSI is made with VIC-SM in retrospective mode (not using forecast). I suggest this because the title of the section is focused on forecast.

    Response: Great suggestion. Done!

11. P 3059 L7-10. Here I am a bit confused by the terminology. Are the SMs forecasted by
VIC?
Response: Yes. SM forecasts were generated by forcing the VIC model with the climate scenarios. We have now clarified that in the manuscript.

12. Fig 7 Why are the results presented here as spatial aggregates and later on they are pixel based?

    Response: As described in the section 2.3 the rainfall forecasts were for aggregated rainfall over the entire focus domain and entire season. During the bootstrapping step the forecasts are disaggregated into gridded forecasts at daily scale for forcing the VIC model. Therefore we thought it would be appropriate to provide assessment of rainfall forecast skill aggregated over entire region. SM forecasts evaluation were done at pixel scale because that is how they are used by the decision makers. A spatial map of soil moisture forecasts helps in demonstrating areas with higher skill during a given time of season and we now clarify this point in the manuscript as well.

13. P 3059 L 15. Referring to SM estimates (using the VIC model when all info are available) as SM observations sounds confusing to me. The exercise of comparing the two is meaningful but the nomenclature is confusing. Would it make sense to refer to "SM forecast" and "SM a posteriori estimates"?

Response: We have incorporated reviewer's suggestion.

14. Section 4. There is no discussion in this section (move the text to conclusion and introduction). The discussion is mostly in section 3. When referencing Rojas et al. 2011 consider referencing Meroni et al. 2014 (Early detection of biomass production deficit hot-spots in semi-arid environment using FAPAR time series and a probabilistic
approach. Remote Sensing of Environment, 142, 57-68) more focused on drought forecasting instead of drought monitoring.

    Response: We have now moved the content of the discussion section to conclusions and summary section. We now cite Meroni et al., 2014. Thanks for your suggestion.

15. P 3062 L23. Where are the station data described in section 2 used? Do you mean the station data used in CHIRPS?
Response: Yes. We mean the station data used in CHIRPS. We have now removed that sentence from the revised version of the manuscript.

16. P 3063 Point 1. Please explain what is the benefit of transferring the system to LDAS with regards to the problem stated above. What observation do you plan to assimilate?

    Response: In lines 1-4 of the page 3064 of the original manuscript we explain benefit of transferring the system to LDAS. The primary benefit of that is it allows us to use more than one model to get the estimate of initial conditions and NASA's Land Information System (LIS) has inbuilt capabilities to work with ensembles of forcings and implement data assimilation.
We now mention the observations that we plan on assimilating (i.e. soil moisture and total water storage)

17. P 3064 Point 3. The point is very relevant given the complexity of the paper and also the difficulty of representing forecasts in an easily understandable way. However the sentence "We recognize .." does not add much. It would be better to specify how exactly you plan to improve the presentation of the forecasts.

    Response: We have now added a few sentences giving an example of how we might improve our forecast presentation. Specifically, we plan to improve the presentation of our forecasts by incorporating the feedback of the end users (FEWS NET's food analysts) on our forecasts. Thus far we have learned that providing the forecasts in terms of the chances of drought onset/persistence/recovery and best analogs is well receipted by the FEWS NET analysts.

Reviewer #2

Seasonal drought forecasting in food-insecure regions such as East Africa is important for reducing drought risks in terms of decision making. This paper is targeted at augmenting the Famine Early Warning System Network (FEWS NET) through incorporating dynamical climate forecast models and a physically-based large-scale land surface hydrologic model. It is an interesting topic and it will benefit local agencies for drought vulnerable regions. I think the paper will finally fit HESS, but currently it suffers from insufficient validation and inappropriate presentation on its difference against previous studies. I would recommend for its publication after the comments below are addressed.

Response: We thank the reviewer for the thoughtful comments that certainly improved the manuscript. Please see our response to your comments below.

Major comments:

1. The title has to be changed, given that Sheffield et al. 2014 and Yuan et al. 2013 already introduced an African drought forecasting system based on CFSv2 and VIC which are also used in this paper. We know that climate forecast model and land surface hydrologic model (with non-trivial calibration) are the most important component for a dynamical-model-based seasonal hydrologic forecasting system, although an update of observation climatology with CHIRPS data is not trivial. Actually the novelty of the paper, in my opinion, is to assess CFSv2/VIC system for growing season in East Africa, and is more targeted at agricultural/crop management. I would suggest changing the title as "Seasonal forecasting of agricultural drought for food-insecure regions of East Africa" to avoid using "system development".

Response: We respectfully disagree with the reviewer's comment. It is correct that we used the same temperature and wind forcings and model parameters as in Sheffield et al., 2014 and Yuan et al., 2013 (we acknowledge this in the manuscript as well). However we believe that this forecasting system has enough differences with Princeton's Africa Drought Monitor to be recognized as a separate system. Primary differences between both systems are:

(1) The primary focus of this system is to forecast agricultural drought in East Africa.
(2) Our approach for bias correction CFSv2 forecasts and generating climate scenarios is indeed an unique addition and sets us apart from the approaches of Sheffield et al., 2014 and Yuan et al., 2013. We use dynamical forecasts over Indo-Pacific region (as shown in Fig. 3) to bias correct dynamical precipitation forecasts over EA region, whereas in in Africa Drought Monitor the dynamical forecast over the domain itself is downscaled and bias corrected. Since the skill of dynamical forecast for the MAM season and EA region is negligible, forecast that is directly downscaled also have negligible skill. Whereas we show that through our approach useful precipitation forecast skill (0.67 as shown in Fig. 7) can be attained over the EA region for MAM season.
(3) We use a rainfall dataset that has been recently developed and takes advantage of

both satellite based precipitation estimates and stations data.
(4) Finally the future directions that we mention for this system will further set it apart from Africa drought monitor.

Please also see section 4 where we describe the differences between our approach and others in detail.

2. Validation. I was excited when I was looking at the title because I was supposed that the paper will address the application of seasonal hydrologic forecasting in crop management in a food-insecure region. But I finally realized that, as pointed out by the authors, the paper is a first step toward augmenting the FEWS NET. It's a reasonable argument because we have to validate the system before application. But I could not find any reference forecast to compare with the CFSv2/VIC forecast throughout the paper. While comparison with ESP/VIC (although straightforward) might be a huge task for revising the paper, at least the comparison with the FEWS NET seasonal climate outlooks (no matter precipitation or soil wetness) would be beneficial to show the rationale of implementing such CFSv2/VIC system.

Response: Thank you for your suggestions. We have now added some figs showing the comparison of VIC-SM with an independent multi-satellite based soil moisture product. We have also added the comparison between ESP-VIC and CFSv2-VIC and highlight the added skill in CFSv2/VIC SM forecasts with respect to ESP-VIC forecasts.

3. The recent 2011 East of Horn Africa drought is a severe drought, which has been addressed in terms of seasonal forecasting by several papers (Dutra et al., HESS, 2013; Sheffield et al., 2014). Given that the hindcast period in this paper is 1993-2012 that also covers 2011, I would suggest adding a figure to show the system's performance on the prediction of 2011 drought for comparison with other studies.

Response: Agreed. We have added a figure showing the performance of this system for 2011 drought event.

Minor comments:

4. P3053. For the introduction of experimental/operational seasonal hydrologic fore-casting system, the Princeton's CONUS seasonal drought forecast system that is based CFSv2 and VIC (http://hydrology.princeton.edu/forecast/current.php; Yuan et al.,
J Climate, 2013) would also be relevant.

   Response: Agreed. We now cite that website and the reference.

5. P3056, generation of seasonal climate scenarios. The hindcast period is 1993-2012, while CFSv2 became operational in 2011 where different numbers of ensemble are generated: there are 24 ensemble members during CFSv2 hindcast period (1982-2010), while up to 124 members in the real-time forecast from 2011 to present. I am

wondering how to handle them in post-processing CFSv2 forcings in this paper. Do you
use all real-time members or just the 5-day gap members that are exactly the same as the hindcast? Is there any significant difference between them?

Response: Great point! From the real-time CFSv2 forecasts we only used ensembles that were initialized on the same days (i.e. 5-day gap members) as in the hindcasts, keeping the number of ensembles consistent with the hindcasts. We mention this in the manuscript as well.

6. P3057. It is not clear how the bias correction is carried out. Some key equations should be introduced. Although the authors mentioned that the general method was introduced and validated in previous study, it will be useful for the readers to understand
the paper if the authors could introduce that in this paper by showing a bias correction example with CFSv2 data. The bias correction might be another unique feature of the system and so it needs to be addressed clearly.

   Response: We have now revised the section 2.3 to make our description of the process of bias-correcting CFSv2 forecasts for the focus domain, clearer. Our approach for bias correction CFSv2 forecasts and generating climate scenarios is indeed an unique addition and sets us apart from the approaches of Sheffield et al., 2014 and Yuan et al., 2013. The MAM precipitation forecast skill of CFSv2 over East Africa is negligible at best so using the forecasts over East Africa itself would not have provided much skill beyond the climatology. Therefore we used CFSv2 forecasts over Indo-Pacific Ocean to get bias-corrected forecasts over the focus domain. In doing so our approach benefits from the strong teleconnection between Indo-Pacific precipitation and East Africa precipitation during MAM as well as the high skill of CFSv2 over tropical Indo Pacific region.

7. P3064. For the NMME/drought topic, Yuan and Wood, GRL, 2013 is also relevant.

   Response: Thank you for your suggestion. We cite that reference now.

8. Figure 3. How to explain the big negative correlations over west tropical Pacific? Are they reasonable?

   Response: Yes, they are reasonable. This recent (post 1999) negative relationship between EA MAM rainfall and west tropical pacific precipitation and SST has been documented in a few recent studies, such as Lyon and Dewitt 2012, Lyon et al., 2013 and Hoell and Funk 2013 (please see the reference below). We also cite these references in the manuscript now.

Reference:

McNally, A., Husak, G., Brown, M.E., Carroll, M., Funk, C., Yatheendradas, S., Arsenault, K., Peters-Lidard, C.D., & Verdin, J.P.: Calculating crop water requirement satisfaction in the West Africa Sahel with remotely sensed soil moisture. In review *Journal of Hydrometeorology, SMAP early adopters' special collection, 2014.*

Hoell A and Funk C.:Indo-Pacific sea surface temperature influences on failed consecutive rainy seasons over eastern Africa *Clim Dyn* 1–16. 2013.

Lyon, B. and DeWitt, D. G.: A recent and abrupt decline in the East African long rains, Geophys. Res. Lett., 39, L02702, doi:10.1029/2011GL050337, 2012.

Lyon B, Barnston A G and DeWitt D G.: Tropical pacific forcing of a 1998–1999 climate shift: observational analysis and climate model results for the boreal spring season *Clim Dyn* 1–17, 2013.