Anonymous Referee #1

The study addresses an important issue of enlarging the data sets available for LSM validation by estimating AET from SWC measurements. Also the underlying idea that recharge and AET data are generally more valuable to society than SWC alone justifies this field of research. The manuscript is very well written. However:

1. The inverse methodology description is very weak. There is no description of which search method is used! What is the combined objective function? A detailed sensitivity analysis has to be given, especially in light of the mentioned problems of equifinality. It is extremely unlikely that all 24 parameters are sensitive and justify optimization. Also inverse modelling offers the opportunity to provide the reader with an estimate of the confidence intervals for each estimated parameter, which will also reveal the sensitivity and associated uncertainty.

   Thank you so much for your comments. The central theme of the paper was to employ a standard publicity available model to test our hypothesis, not to devise new algorithms for inversion, and that is why we did not get into the inverse modeling details in great depth. As it was mentioned in the paper, more description about inverse modeling can be found in Mualem (1976), van Genuchten (1980), and Turkeltaub et al. (2015).

   Moreover, Wang et al (2009) have done a detailed sensitivity analysis of groundwater recharge and evapotranspiration for soil hydraulic parameters in a single layer. We respect your concerns and have undertaken a sensitivity analysis of all 4 layers (24 parameters) extending the original work of Wang et al (2009).

   Wang, T., V. A. Zlotnik, J. Simunek, and M. G. Schaap. 2009. Using pedotransfer functions in vadose zone models for estimating groundwater recharge in semiarid regions. Water Resources Research 45: 12. doi:10.1029/2008wr006903.

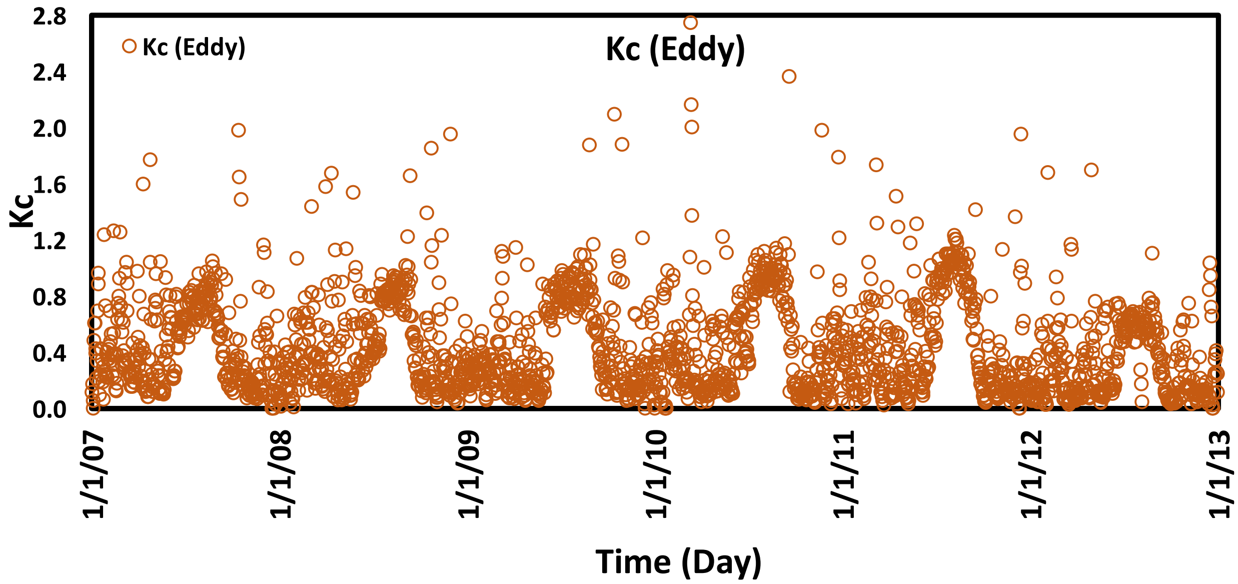
2. The results of simulated SWC seems to be reasonable from a SWC perspective, but it's important to also address the certainty/robustness and likelihood of the estimated soil parameters. Are they random parameter picks from an equifinal problem or are they physically reasonable and do their mutual differences fit into field/lab measurements (I assume soil samples exists from the sites)? The author has attempted to validate the spatial distribution of the estimated soil parameters based on a soil map, which is highly appreciated. However, it would have been interesting to utilize this information for regionalizing the soil parameters and thereby limiting the number of free parameters in the calibration. Likewise, the soil map could have been used to upscale the AET simulations to the field scale by including the soil map instead of a simple average of the four points.

Thank you for the suggestion. Upscaling the AET by the SSURGO soil map will be interesting. We also note that Professor Franz uses a range of hydrogeophysical mapping techniques (i.e. electromagnetic induction, cosmic-ray neutron rover) to understand soil patterns and properties. We have mapped this site several times and will consider adding some of the maps to this manuscript or a companion manuscript.

3. The results of the AET simulations seem to be very poor. I miss a critical view on the results regarding lacking ability to simulate even inter-annual variability (fig 11) and perhaps more importantly the apparently complete lack of predictive capability on the daily scale. The performance metrics in Table 6 indicate good R^2 and NSE, but that correlation is intrinsically given by the seasonality of the climate. The real test is if the model has any predictive power on estimating the evaporative fraction AET/PET. If you normalize the AET on a daily timescale by the daily PET and then calculate the R^2 and NSE, you probably get no explanation of variance. This can also be somewhat illustrated in table 6, if you add a column of RMSE in % of average daily AET, then you see that the RMSE is in the order of 50-80% of the daily AET (see attached table). In comparison most Remote sensing AET methods can, with calibration, achieve results in the order of RMSE of 25-30% of the daily mean AET.

We compared the results with EC measured ET in this study just as a simple comparison as there was no other relatively accurate measured ET data available in the study area. As it was mentioned in the paper there are always different uncertainties involve in the Eddy-Covariance (EC) measurements. EC measurements can be bias by up to 20% or even more. Considering this we cannot easily say since the simulated ET values are not perfectly matched with EC ET measured data that "the AET simulations seem to be very poor".

Your suggestions are appreciated and if we had access to more accurate measured ET data, like Lysimeter measured ET, we could investigate such analysis. Because of the nature of EC ET measurements (which is not based on Kc values, but instead based on the flux measurements) such comparison may not be useful. As an example obtained Kc values from EC (2007-2012) are shown below. According to the graph, most of the times during mid-growing season, obtained Kc values from EC are less than 1 (we usually expect to have values of 1-1.2 during the mid-growth season). The average EC Kc value during July and August (2007-2012) is 0.81 with a minimum average Kc value of 0.58 in 2012 and maximum Kc value of 0.99 in 2011. On the other hand, sometimes Kc values exceed 4 while in the real world such Kc values do not exist. In addition, Kc values do not usually change suddenly during growing season and it is rarely possible to have Kc value of 1 in one day and Kc value of 0.4 for the next day, but according to the graphs in some of the days we can see this case in the EC Kc values. The inherent noise seen in Kc makes this comparison challenging without temporal smoothing.



4. Given the very little detail available on the AET model used (Feddes 1978) I can only speculate, but perhaps the simulated SWC is not accurate enough at the critical moments when AET is limited by water availability, or the AET model is not appropriate or the climate data are poor. But overall I do not find the results on simulated daily AET encouraging. An uncertainty analysis of the different model components would be appropriate (see comment below).

We appreciate the reviewer's thoughts, but most of the comments are made based on the apparent difference between the EC measured ET and simulated ET. The Hydrus model is a widely used method based on a solution to the Richards Equation. The Mead Site 3 flux tower is a long standing Ameriflux tower and continues to be a part of the core network. In order to address the comments, we will perform a sensitivity analysis of all 24 soil hydraulic properties building on Wang et al. (2009). A full sensitivity analysis of the root model parameters is beyond the current scope of the paper and we refer the reviewer to Guswa (2012).

Guswa, A. J. 2012. Canopy vs. Roots: Production and Destruction of Variability in Soil Moisture and Hydrologic Fluxes. Vadose Zone Journal 11:3. doi:10.2136/vzj2011.0159.

Q: footprint analysis? EC footprint of 250 m radius is very large, what is the height of the EC mast?

The height on the EC mast varies with crop height. According to Suyker et al. (2004):

"To have sufficient upwind fetch (in all directions) representative of the cropping system being studied, eddy covariance sensors were mounted at 3.0 m above the ground while the canopy was shorter than 1.0 m, and later moved to a height of 6.2 m until harvest."

The footprint of the tower will there change over the season, ~100 times the tower height. This is a long running Ameriflux site and the variable footprint is a part of the method and its inherent uncertainty.

Suyker, A. E., S. B. Verma, G. G. Burba, T. J. Arkebauer, D. T. Walters, and K. G. Hubbard. 2004. Growing season carbon dioxide exchange in irrigated and rainfed maize. Agric. For. Meteorol. 124:1-2: 1-13. doi:10.1016/j.agrformet.2004.01.011.

5. Please explain the reasoning behind eq. 2 and 3?

   We needed to introduce potential evaporation (Ep) and potential transpiration (Tp) values to the Hydrus model. By using Beer's law we were able to divide ETp to Ep and Tp. Based on LAI values, with equation 2 we can calculate the Ep values and then by having the Ep value we can use equation 3 to calculate the Tp values. More information can be found in Šimunek et al, (2013).

   Šimunek, J., Šejna, M., Saito, H., Sakai, M., van Genuchten, M.T. (2013). The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water,Heat, and Multiple Solutes in Variably-Saturated Media, Version 4.17.Department of Environmental Sciences, University of California Riverside, Riverside, California, USA, 307 pp.

6. L168: The Actual Transpiration is calculated using Feddes 1978 based on Tp and root density distribution. That must be a key component of this approach, please give more details on the application of the Feddes model.

   According to Šimunek et al, (2013), S is a sink term and has been defined as plant water uptake:

$$S(h) = \alpha(h)S_p$$

   where α(h)is a dimensionless function varies between 0 and 1 depending upon soil water pressure head and S_p is the potential water uptake rate assumed to be equal to Tp. More information can be found in Šimunek et al, (2013), and Wang et al. (2016). We will add a better description to the manuscript. See Guswa (2012) for a more in depth look.

7. L198-204: Optimized against which objective function? What was the calibration target? Which optimization algorithm (gradient based/global etc.) is used? That has to be clear up front? Also what was the result of the sensitivity analysis? Which type of sensitivity analysis, was it necessary to optimize all parameters? And why not calibrate all four layers simultaneous?

We could not optimize all the layers simultaneously because the maximum number of parameters that we can be optimized by the Hydrus-1D model is 15. We have followed the same procedure as Turkeltaub et al. (2015) and Wang et al. (2015, 2016). We used RMSE as our objective function and will clarify this more in the manuscript. Finally, a sensitivity analysis of all 24 parameters will be presented.

8. L220-224: It might be obvious, but please state clearly, which observation data the performance metrics are based on.

   The performance metrics are based on the soil water content data and it will be added to the modified manuscript.

9. L230: How are the best defined, what are the weights and how was your combined objective function defined?

   We chose the selected optimized sets of soil parameters values based on the RMSE but the other objective functions were performed in order to double check the optimization process. We will clearly state in the modified manuscript that the soil hydraulic parameters were chosen based on the RMSE.

10. L265-266: Of course the upper layers are better you calibrated them first and then kept them fixed while calibrating the lower layers, so they have had significantly more freedom in the optimization. Try to calibrate the lower first and then fix them and calibrate the upper, then you might get different results.

    The Hydrus-1D can just optimize up to 15 parameters simultaneously and we decided to optimize the upper 2 layers first and then the 2 lower layers. The SWC data in the 2 upper layers has more dynamics than the 2 lower layers. Again a sensitivity analysis of all 24 parameters will be presented. Interestingly, preliminary results indicate the deepest layer n value is the most sensitive.

11. L336: "the various ETa estimation techniques performed well." I disagree.

    Each method, measurements, inverse modelling, are prone to uncertainties. Unfortunately, no true ETa estimate exists at the field scale. We will try to soften the language here.

12. L337: "In fact, it is difficult to identify which is the clear solution if any." Please rephrase.

    Thank you, we will rephrase it.

13. Fig 9: How come the simulated values cannot go down to 0.20-0.25 for the Cosmic ray calibration, when that is possible for the TP calibrations?

    We pointed out in the paper during the growing season when crops extract water from deeper soil layers comparison between simulated and observed values deteriorates due to the fact that the CRNP observational depth is limited to near surface layers (~20 cm).

14. Fig 11: The proposed method seems to not capture the inter-annual variability, try to plot the annual values of EC against simulated annual values in a scatterplot to see if there is any correlation on an annual basis?

    Thank you for your suggestion. We will try to plot the annual measured EC ET values versus simulated annual values in a scatterplot to see if that seems more informative.

15. Fig 13: You need to plot the daily obs vs. simulated AET in a scatterplot, the accumulated curves gives no indication of the performance of the daily model simulations! The bias of the Scatter plot will however give you the same information as the offset in accumulated values.

    We have investigated the plots and will provide the 1:1 in the full response to reviewers.

16. Table 6: Needs units.

    Thank you, we will add them to the table.

17. I suggest resubmission of a new manuscript after major development of the inverse modelling and careful rethinking about the quality of the daily AET simulation results and reasons for the insufficient performance (AET model concept, upscaling, SWC simulations at critical stages, uncertainty in soil parameters, climate data etc.). Here I would suggest some uncertainty analysis of the relative importance of these factors for the final AET results. E.g. how important are changes in soil parameters to the final result? And how important are the assumptions in the model (e.g. root depth, soil profile depths etc.).

Thank you for the very insightful comments. Obviously the reviewer is well versed in this type of analysis. We will try our best to resolve the above issues. A sensitivity analysis of all 24 soil parameters in combination with the work of Wang et al. (2009) will be instructive.

Good luck