Authors' response to referee #1 comment (hessd-10-C341-2013) on **"Inter-comparison of four remote sensing based surface energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate"**. MS No.: hess-2012-545, by Chirouze et al.

The authors would like to give their thanks to anonymous referee #2 for his advised comments that will greatly help to improve the manuscript's scientific quality.

In the following document, bold & italic text is extracted from the referee's comment file; normal text corresponds to authors' response.

3.2.1 - I understand that EC Sites have budget closure issues. In your case, the budget closure seems to be substantial. You do mention towards the end of the paper that the lack of long-term measurements means that you didn't have the time to correct the mistakes that a long-term network might have. Is there any additional information about your site that you could provide that addresses why the budget closure problems might be so high? I'm not looking for definitive statements, but trying to understand why it is that poor given the environment would add value to the interpretation.

It is true that we did not mention many reasons to explain energy budget closure at each one of the stations. We will try to be more exhaustive in the next version of the paper. From what is known about the field experiment, the following assumptions can reasonably made in an attempt to explain those problems:

- Some problems were well identified malfunctioning of the instruments.
- Differences in behavior between stations can be firstly explained by the variety in observed crop (high row cropping for chickpea, homogeneous and relatively low for wheat, low row cropping for chili pepper, etc). Precision in estimating available energy can greatly differ from one site to the other depending on this typical surface heterogeneities.
- In addition to that, all the instruments were not from the same brand and were lent by different institutions. Instruments of a different making and with a different maintenance history will tend to have different behaviors. Also, only automatic data postprocessing could be carried out, thus excluding careful spectral analysis.
- The soil heat flux was measured both at the top and bottom of the furrows for sites with furrow irrigation. The soil heat flux used in this paper is a simple average of those two values.
- The CNR1 sensor mounted at the east wheat station presented a negative bias on incoming shortwave radiation (in comparison with measurements at the other stations).

Also, your choice of closure adjustment method seems very subjective. Could you give a little more insight into why you choose the method for each site? It's hard for

me to believe that towers that are that close together would not act the same way in terms of budget closure.

In the same way, we will try to add some information about the choice of the correction method, summarizing the following points.

Provided that one trusts the estimates of both net radiation and soil heat flux, the "Bowen ratio" method is widely used in the literature in order to correct energy budget closure issues.

Except for the east wheat station, the estimations of available energy seemed trustful. The R_n estimated at this wheat station will be corrected in the paper's next version by using the global incoming solar radiation measured at the meteorological station instead of the local one.

For three of the seven stations (east wheat, broccoli and potatoes-sorghum), the slope of the regression between available energy and the sum of the turbulent fluxes is larger than 0.65, therefore both fluxes can be considered as consistent, and the Bowen ratio method has been applied.

For the west wheat and chickpea stations, the equipped KH20 sensor had big failures during the whole season. Both measurements of H and λE were poor at the beginning of the season at the west wheat station, so their values were discarded for this period. For the rest of the measurements on both stations, H values were trustworthy. Therefore, we chose to assess λE as the residual of the energy balance at the station.

4.7 – When I started reading the spatial variability I got excited to see how the models compared spatially but I was let down. You only spend 3 paragraphs on this topic where you could spend much longer. I think this is a lost opportunity, please expand.

Given the lack of validation of the spatial variability, outside the flux station network, we do not want to extend the spatial standpoint analysis beyond what can be inferred from the existing dataset. For instance, most differences between the various patterns derived from the model outputs are within the accepted uncertainty and show consistency with what is expected at the various locations given the irrigation history and the vegetation. A paragraph addressing this more extensively is added in the revised document.

5. The results do support the conclusions and interpretation. However, I think there a missed opportunity to explain what is going on. By this I mean that you should dig a little deeper into why the models can be failing. You touch on the senescence issue but could you go a little further in proving this point? What about the time period in which you don't have these problems? I don't know how they perform then and if they improve substantially. There is a lot more going on there that you don't mention. I'm not looking for a complete explanation of the errors but more than just the senescence argument would be good (i.e. Input data uncertainties, model

uncertainty, parameter uncertainty…). You don't need to redo the analysis completely, but it would be good to add a little more context to saying that the main problems are due to the senescence. I would hypothesize that most of the errors come from the model parameterization and the remote sensing uncertainties. Feel free to prove me wrong.

The referee's guess is good in saying that most of the errors are due to model parameterization and remote sensing uncertainties. Indeed, RMSD are quite high in this study and the major part of those errors can be explained by these two arguments plus the quality of the in-situ dataset. Given those high error levels, and with the aim of using infra-red data into an assimilation scheme, we focused our study more in interpreting behaviors and reasons for high punctual errors than in explaining the global reasons for such high RMSD.

In the revised manuscript, more elements of comparison between phenological stages of plants (senescence, green growing, bare soil) will be added.

6. One suggestion is to discuss how you fit parameters to a lot of the parameterizations that you use (i.e. LAI relationship). If I were given your original data could I reproduce the input data and parameters by using this paper as a guideline?

We agree that some information is missing in the LAI=f(NDVI) relationship paragraph. The fitting of the logarithmic curve has been done in (Fieuzal et al., 2011) but the reference is missing in the paragraph. It will be added in the next version of the paper. In this relationship, the three parameters k, $NDVI_{soil}$ and $NDVI_{\infty}$ are calibrated simultaneously to achieve a minimum RMSD. No fitting has been done by the authors in the albedo computation from FORMOSAT-2 bands.

8. That being said, I would suggest shortening the title if at all possible.

We will shorten the title.

9. The abstract does summarize well the main points of the paper. However, it could use some refinement. What you say in the abstract could be said in 75% of the number of words that you use. I would give preference to work on making it more concise before adding any new information.

The abstract will be shortened.

10. The introduction (State Of the art + Context And Objectives) should be put into one section and should be shorter.

As specified in the response to the referee n°1, these two parts will be merged into one.

11. I don't have any complaints about the grammar. My major complaints are the very long sentences and odd phrasing. I would comment them here but there are too

many and a lot of the information would be redundant. I would suggest having a native English speaker read through it to point out the main issues.

Additional attention will be put to the language in the next version of the paper.

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

0. ***Abstract - Too wordy.***

Abstract will be shortened.

1. ***The introduction should be shorter and more concise.***

The introduction will be shortened.

2. ***You need to work on the size of the figures. They seem to be all over the place. Figure 8 Is massive. The reader will get the message with a smaller version. Along those same lines, could you include other plots and metrics to compare the surface fluxes? I think it would be much more informative.***

Figures will be modified as suggested. Metrics will be added to plots to make comparison of fluxes easier.

3. ***The Spacing in Figure 11 Looks odd.***

A less odd spacing will be applied to figure 11.

TAKE HOME MESSAGE
Overall, I think the paper does a great job at using high---resolution remote sensing information with a suite of state of the art remote sensing ET models. You get that message across clearly and well. My main suggestions are the following:

1. ***Make everything much more concise. You Ramble sometimes and lose your audience. This could help in your introduction, abstract, and certainly your results.***

Efforts will be put to make the text clearer as suggested.

2. ***Work on your phrases. Some just seem odd in English.***

More care will be put in the language of the revised manuscript.

3. ***Expand on the results and discussion section. This is a missed opportunity. This is where you should prove to me that you understand the models***

thoroughly and given the assumptions the argument of why I should you these models in the future.

As said in the response to comments 4.7 and 5, results and discussion part will be expanded as suggested.

4. ***I don't see a definitive statement of which model I Should use. Please Make the case for one of them and explain why. The Answer could always be that it depends on the scenario. You touch on this briefly, but it would be good to expand."***

The answer could be that TSEB is the most consistent one (SEBS has too much bare soil issues) but needs to be improved to be more accurate for senescent pixels. Another point is that simplified methods can be very interesting to use (especially because they are less computer-time-consuming and auto-calibrated) given that a sufficient variety of hydric and vegetative states are present in the same image.