We would like to thank the editor and the reviewers for constructive feedback which has led to the improvement of the quality of our study. Please find our answers to the comments (indicated in *italic*) below. In **bold** we indicate where in the manuscript the comment has been addressed. Those locations (and locations of other significant changes) are also highlighted in yellow in the manuscript.

Dr. Conradt

Page 5911, middle of the lower paragraph: "The locations of the rain gauges and climate stations in relation to the study area are presented in Fig. 2 of Stisen et al. (2011a)."
I think it should be no problem to add these locations as red dots to Fig. 1 and I would like to see them there. There are still many occasions in this manuscript where references to other publications may be unavoidable, especially regarding lengthy model descriptions, but here there is no need to keep readers busy searching other sources.

We will add the locations of rain gauges and climate stations.

Addressed in Figure 1.

Page 5916, end of first paragraph and Page 5917, end of first paragraph: "The modelled fluxes are output at 1km resolution."
The resolution(s) of the original input pixels of the satellite scans should be given, and it should be added how they were georeferenced and interpolated to this 1 km output grid.

MODIS data is provided by NASA in a georeferenced grid with 930 m resolution and Sinusoidal projection. This was bilinear resampled to 1 km resolution grid and reprojected to UTM32-WGS84 projection.

The Envisat ATS_TOA_1P, AATSR Gridded Brightness Temperature and Radiance, product, is a full resolution dataset resampled to a 1km grid for both the nadir and forward views by the European Space Agency (Scarpino and Cardaci, 2009). The split-window brightness temperatures (11 and 12 µm) for both forward and nadir were then reprojected to UTM32-WGS84 and resampled to the same 1km resolution using a bilinear interpolation. LST at the two AATSR observation angles was then retrieved by the quadratic dual-channel split-window algorithm proposed by Coll et al. (2006) for AATSR.

The above clarification has been added to the relevant sections in the revised manuscript.

Addressed on lines 249-250, 404-408 and 469-481.

Page 5917, second paragraph: "all the days between 2003 and 2010"
There are several instances (cf. the figure captions) where you wrote that you used seven years of data. However, there are only six years between 2003 and 2010, and including these years it would be a timespan of eight years.

2003 and 2010 were also included in the study and therefore the correct number is 8 years.

Addressed throughout the manuscript.

Page 5921, end of first paragraph and Page 5923, third line from the bottom: "not shown."
This wording should be avoided. You may write about observations you made without visualising the data, but just giving the relevant numbers or showing the extra figure would be better, of course. If the latter makes no sense in your eyes, O.K., but then you may at most explain why you cannot show everything. (Read: just leave the "not showns" away!)

The statement has been removed.

Finally, all the figures illustrating the correlation of spatial patterns
These are entirely density scatter plots. I would like to see one or two triplets of maps showing the actual spatial distribution of the heat fluxes over the catchment. This would allow for analysing the deviations in specific locations with respect to special soil or surface characteristics.

Maps of energy fluxes on individual, cloud-free days might not provide much additional information. However, we have added maps showing correlation, RMSD and bias of modelled LE, averaged over the whole study period.

Addressed in Figure 10 and lines 808-834.

Reviewer 2

If the final goal is to assimilate ETR maps derived from remote sensing in a distributed model, one expects from this paper to answer at least partially the following question: "what is the added value of the surface energy balance models in an Observing System whose state-space model is a coupled hydrology-SVAT model ?". The results only briefly tackle this issue; it should be emphasized in a revised version.

The goal of this paper was to look at the question: "is there any additional (spatial) information present in outputs of remote-sensing based surface energy balance models that is missing from physically-based distributed hydrological model". From the study it appears that there is, and therefore in the conclusion we pose the question which (if any) of the outputs should be assimilated. Assessing the added value of such assimilation by, for example, comparing outputs of hydrological model runs with and without assimilation is a topic of ongoing research. We have clarified this in the revised manuscript.

Addressed on lines 29-36, 190-194 and 1072-1044.

For instance, all results are expressed as instantaneous fluxes at different times of the day (different overpass times). Instantaneous fluxes at 11:30AM and 13AM local time can differ from more than 100W/m2, which makes me doubtful about the relevance of such an intercomparion. In order to intercompare the models, results should be translated in, at least, daily totals in mm/day. Why not use classical EF extrapolation and interpolation methods to convert instantaneous ETR estimates to daily and seasonal ETR values ?

It is true that Aqua and Envisat satellites have overpass times at different times during the day. However this difference is already accounted for, and discussed in the study. As explained on P5918 L4-13, when the DTD or TSEB-2ART fluxes are compared with MIKE SHE fluxes, the MIKE SHE estimate is interpolated from two hourly steps bracketing the satellite overpass time. For example, if satellite overpass was at 11:48, MIKE SHE estimates from 11:00 and 12:00 would be used. Therefore the issue of fluxes being modelled at different times is only applicable when DTD and TSEB-2ART fluxes are compared and it is discussed in results in section 6.1.3.

In addition EF is also used during the comparison (Table 2 and Figs 2-4). Extrapolating EF to daily values would just involve multiplying it by the daily available energy (or net radiation assuming negligible daily G). Therefore the multiplicative factor would be the same (if tower observations were used) or very similar (if modelled, as seen in the available energy intercomparison from figs 2-4 and Table 2) for the three models. Hence, we consider that no additional information or insight would be gained by extrapolating to daily values. Additionally, as mentioned on P5923 L1-9 the self-preservation of EF might not always hold over the whole study area due to frequently cloudy conditions, bringing additional complications and errors when extrapolating to daily values.

Addressed on lines 523-525 and 530-542.

The cumulative values should also be compared to the information amount already provided by the potential ET which is a good reference to assess the added value of any ETR model (this is also true for instantaneous values).

Potential ET was added to the comparison of temporal pattern modelled in the catchment.

Addressed in Figure 5 and lines 1039-1040.

Abstract is not conclusive enough and the introductive part of the abstract (roughly half of it) is too long. You should state clearly the outcomes of the study. I don't see, in its present state, what insight the paper brings to the hydrological modelers to consider assimilating ETR maps in their models. After reading the paper, if I was a modeler, I'd keep the current calibration method based on the sole available data incl. surface temperature! Actually, whether one must directly assimilate surface temperature instead of ETR maps derived from surface temperature in a distributed model is still an open question, and the answer lies in a careful analysis of the various uncertainties and the consistency of the spatial covariances, not only on the average catchment outputs. This should be commented.

The abstract has been rewritten to better represent the outcome of the study and to shorten the introductory part. However, in this paper we are not trying to convince the hydrological modeller to assimilate remotely sensed ET maps into their models. We are merely exploring if there is any extra information in those maps that could be assimilated. In fact, in the conclusion (P5930 L20-25) we are stating, in accordance with the reviewer's comment, that whether ET or LST (or another parameter) should be used during the assimilation is still an open question.

Addressed on lines 1-36.

The structure of the energy budget (dual source model with series resistances) is essentially the same between MIKE_SHE and both SEB models but the formulation of the aerodynamic resistances (soil to aerodynamic level, vegetation to aerodynamic level) are different (Shuttleworth Wallace vs original Norman et al.). What part of the current findings comes from this difference ? It would be easy to implement the same resistance scheme in the 3 models, this would ensure that the differences come from the methods (constraint on ETR by surface temperature vs by the water budget) and not the algorithmic choice.

This is a very valid comment and in the revised manuscript we have also run the models with the same formulations of the aerodynamic resistances.

Addressed on lines 962-995 and Table 6.

The reference mentioned for TSEB-ART is a conference abstract, I think that for the sake of comprehension of the paper the model should be better described in the manuscript so that the reader can grasp the difference between both SEB models. Actually, the presentation of the SEB models should focus on the main difference between the two models: both use two surface temperatures to constraint the dual source energy budget. However, in DTD, the constraint on ETR is the same as the original TSEB model, the day/night surface temperature data being used only to get rid of systematic errors in the surface temperature estimates. The vegetation is transpiring at the potential rate unless there is an inconsistency in the soil surface temperature retrieval. In TSEB-2ART on the other hand, it seems to me that the directional temperatures are used to derive the soil and vegetation component temperatures, and that the transpiration rate is therefore computed directly as a residual term. It's not clear how the resulting inconsistencies are treated. It seems from P5917 L13 that inconsistencies are not corrected for but simply ignored. The lack of robustness of most ETR retrieval algorithms based on the use of surface temperature data is an open issue for hydrological applications and should be dutifully commented in the paper.

A full paper describing TSEB-2ART is currently in the final stages of being written and will be submitted to peer review soon. However we have elaborated on the description of TSEB-2ART in the revised manuscript.

The reviewer correctly notes the main difference between DTD and TSEB-2ART. When TSEB is applied with single angle LST, some assumptions are needed based on the expectation that plants transpire at their potential rate. This assumption may lead to significant errors in cases when plants are stressed, or when the potential canopy transpiration is not well defined. For that reason, the green fraction of vegetation

($f_{\$\_g\$}$) is an important parameter within the model since it improves TSEB accuracy in forested ecosystems and during senescence by taking into account the phenological development of the vegetation (Guzinski et al. 2013). However there does not exist yet an established method of estimating $f_{\$\_g\$}$ using remote sensing data. To overcome this issue, dual angle LST can be used for retrieving soil and canopy temperatures without employing any assumptions based on the canopy transpiration (Chehbouni et al. 2001; Kustas and Norman 1997) (Nieto et al. 2010a; Nieto et al. 2010b). Simple models for such retrieval have been proposed based on the proportion of vegetation and soil observed at two different viewing angles (Chehbouni et al. 2001; Kustas and Norman 1997). However, since plant canopies are composed of finite leaves, multiple scattering of energy occurs within the canopy and therefore a more physically complex methods for retrieving soil and canopy temperatures may be needed when using dual angle LST measurements (François 2002) (Nieto et al. 2013).

However, we do not believe that the differences between DTD and TSEB-2ART lead to inconsistency in the study. On the contrary, we think that it contributes to the study's validity since the two remote-sensing models use both different data (MODIS and Envisat) and different approaches to estimate the fluxes. Therefore, if the flux patterns produced by the remote-sensing models are more similar to each other than to the flux patterns produced by the hydrological model, it indicates that the remote sensing models might contain some additional information that is not currently present in the hydrological model. It would also indicate some level of robustness in the remote-sensing models. We have elaborated on this point in the revised manuscript.

Finally, P5917 L13 was a technical note indicating that sometimes TSEB-2ART produced unrealistic resistance values and that fluxes from those model runs were not used in the comparison. However this condition was removed in the revised manuscript.

Addressed on lines 411-435 and 1067-1074.

- P5909 L8: This statement is too strong. Both approaches (single and dual source with series resistances) compute an aerodynamic temperature as a common source of heat for the whole surface. Whether this temperature has a physical meaning is not certain. I'd rather say "the two-source models have the advantage of explicitly representing the separate contribution of the soil and the vegetation and avoiding the need of an excess resistance whose value differs largely from one reference to the other"; provide a more recent reference for the "excess" resistance (kB-1) issue (e.g. (Matsushima 2005; Kustas and Anderson 2009; Boulet, Olioso et al. 2012))
- P5913 L20: Same as above (Norman et al., 1995 is a rather old reference on the kB-1 issue)

The statements and references have been updated following the reviewer's suggestions.

Addressed on lines 130-136 and 307-313.

- P5917 L12-15: What is the percentage of discarded pixels/dates that don't meet each of the 3 conditions ?

The first condition (pixel not classified as urban or water) depends on the land cover map and is the same for all dates. According to the land cover map used in the study, water and urban areas together comprise 97 pixels within the catchment area, which represents less than 4% or all pixels.

The other condition (valid modelled fluxes) changes from date to date and depends on model formulations (including the equations used for resistances). It is most often caused by noise in the input data or invalid model parameterization and it leads to removal of 15% of output pixels.

Addressed on lines 490-494.

- P5915: I find this part confusing and actually do not understand how the various Rs values are derived. If b and c values for short crops (LAI<2) are different in temperate and arid regions then Rs is an empirical relationship. It seems to me that the Shuttleworth-Wallace (SW) expression of this resistance gives satisfaction in all climate conditions with the same amount of input data. It's actually the expression used in the Mike-SHE model. Why not use this resistance formulation then ?

In both DTD and TSEB models, b and c are constant for all LAIs and for all landcover classes. However, since DTD aspires to use just time differential temperature measurements, it was originally decided to remove the $(T_S - T_C)^{1/3}$ term from the resistance equation and instead to replace it by a LAI-dependant constant. For dense canopies T_S-T_C was assumed to be 5 K, while for sparse canopies it was assumed to be 15 K (Norman et al 2000). This assumption is empirical and might not hold in all climatic regions. Therefore in the current study we decided to replace $(T_S - T_C)^{1/3}$ by $((T_{R,1} - T_{R,0}) – (T_{A,1} – T_{A,0}))^{1/3}$. This formulation avoids the use of non-time differential temperature estimates while also avoiding the empirical assumption. This has been clarified in the revised manuscript. In addition, the resistance formulations used in MIKE SHE have been implemented in DTD as mentioned in a previous answer.

Addressed on lines 366-372 and 390-392.

- P5931: why use a uniform and constant vegetation height in the SEB models rather than the formulation used by Mike-SHE ?

The relationship used by MIKE SHE is empirically derived in croplands and therefore it underestimates the vegetation height in, for example, forest. It was used in all land cover classes because MIKE SHE is rather insensitive to this parameter. In the remote sensing models we wanted to assign more realistic vegetation height to each land-cover class. However in the revised manuscript we have used the MIKE SHE vegetation height formulation in all the models in order to avoid confusion and to make sure that the different models are run with the same ancillary inputs.

Addressed in Table 1. Additionally LAI correction factor was introduced as explained on lines 242-248.

- P5940 and 5941: all findings indicate that MIKE-SHE simulates a lot more stressed vegetation than both SEB models(lower ETR and higher vegetation temperature). This should be commented more largely in the text. Given your knowledge of the crops and the climatic conditions, do you think that the stress level simulated by MIKE-SHE is realistic ? If not, this would help answering the question about the added value of SEB models in an Observing System.

Due to the introduction of LAI correction factor the individual statistics have changed (although the overall findings of the study remain the same). Therefore, the explanations are spread throughout the paper.

- P5943: merging all seven years in a single Figure is confusing. I don't understand why the line (catchment average) is so far below the circles (even though those circles represent clear sky days only). Please indicate the catchment average potential ET in a second line in order to assess the overall catchment water stress.

It was decided to combine all the years in a single figure since this way the temporal (seasonal) patterns of catchment-wide ET modelled by the different models can be better illustrated. On average there were around 10 days per year satisfying the selection criteria mentioned on P5919 L19-21. Therefore, placing the years next to each other would have resulted in a drawn out, less clear figure. The line represents an average ET for a particular DOY for all the years under study, regardless of whether it was a cloudy or sunny day. Since, as discussed on P5929 L1-9, the ET is mostly limited by energy availability, it is expected that clear-sky ET will be higher than an average of mostly cloudy-sky ETs. We have clarified this in the revised manuscript. In addition we have added a line to the graph showing a 8-year averaged potential ET for each day.

Addressed on lines 1034-1040 and caption of figure 5.

References

Chehbouni, A., Nouvellon, Y., Lhomme, J.P., Watts, C., Boulet, G., Kerr, Y.H., Moran, M.S., & Goodrich, D.C. (2001). Estimation of surface sensible heat flux using dual angle observations of radiative surface temperature. *Agricultural and Forest Meteorology, 108*, 55-65

Coll, C., Caselles, V., Galve, J., Valor, E., Niclos, R., & Sanchez, J. (2006). Evaluation of split-window and dual-angle correction methods for land surface temperature retrieval from Envisat/Advanced Along Track Scanning Radiometer (AATSR) data. Journal of Geophysical Research-Atmospheres, 111, D12105.

François, C. (2002). The potential of directional radiometric temperatures for monitoring soil and leaf temperature and soil moisture status. *Remote Sensing of Environment, 80*, 122 - 133

Guzinski, R., Anderson, M.C., Kustas, W.P., Nieto, H., & Sandholt, I. (2013). Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day–night MODIS observations. *Hydrol. Earth Syst. Sci., 17*, 2809-2825

Kustas, W.P., & Norman, J.M. (1997). A two-source approach for estimating turbulent fluxes using multiple angle thermal infrared observations. *Water Resources Research, 33*, 1495-1508

Nieto, H., Guzinski, R., Jensen, R., Sandholt, I., & Jensen, K.H. (2013). TSEBRTM: Coupling a canopy Radiative Transfer model with a Two Source Energy Balance model for the estimation of surface energy fluxes with dual-angle land surface temperature. In, *TR32-HOBE International Symposium*. Bonn, Germany

Nieto, H., Sandholt, I., Rasmussen, M.O., Herbst, M., Ringgaard, R., & Friborg, T. (2010a). Optical and thermal multiangular measurements for the estimation of evapotranspiration on an eddy covariance flux tower. In J.A. Sobrino (Ed.), *3rd International Symposium on Recent Advances in Quantitative Remote Sensing*. Valencia (SPAIN)

Nieto, H., Sandholt, I., Ringgaard, R., Herbst, M., Stisen, S., & Friborg, T. (2010b). Combining ENVISAT AATSR and MERIS imagery for the retrieval of evapotranspiration in a two-source resistance energy balance model. In, *Remote Sensing and Hydrology 2010 Symposium*. Jackson Hole, Wyoming

Norman, J. M., Kustas, W., Prueger, J., and Diak, G.: Surface flux estimation using radiometric temperature: a dual temperature-difference method to minimize measurement errors, Water Resour. Res., 36, 2263–2274, 2000

M.G.Scarpino, M. Cardaci (2009) ENVISAT-1 PRODUCTS SPECIFICATIONS - VOLUME 7: AATSR PRODUCTS SPECIFICATIONS, Issue 4 / A