We would once again like to thank the reviewers for their help in improving this manuscript. We have already replied to the reviewers' comments during the open discussion phase of the review process. Therefore below we repeat those answers with the location of the relevant changes in the revised manuscript indicated in **bold** font.

Review by Dr. Kalma

It is noted (p. 4860, lines 3-5) that the ALEXI and DTD models are based on TSEB. Similarly, on p. 4867 (lines 8-9) it is stated that ALEXI and DTD have been developed from the TSEB. The TSEB model has been presented in Appendix A1. Where is the TSEB version referred to in Sect. 3.1 and Appendix A1 different from the versions described in Norman et al. (1995), Norman et al. (2000), Anderson et al. (1995), and Anderson et al. (1997)? Similarly: where does the DTD model described in Appendix A2 differ from that described by Norman et al. (2000), except for the formulation of the flow resistance network. Do we need these two Appendices? Also: is the disaggregation algorithm used here essentially the same as DisALEXI described by Anderson et al. (2004)?

The TSEB and DTD models used in the study are the same as presented in the listed papers, with the exception of the series resistance network in case of DTD. However we still think that the appendices are warranted. Their purpose is not to describe new model developments but to clearly show the model implementation used in this study. It is often the case that when model equations come from a number of different papers it is unclear which formulation was actually used, thus making it harder for others to replicate the results. We will add a clarification at the beginning of the appendices explaining this point.
P27L8 – P27L11

Regarding the disaggregation algorithm, it is different from the algorithm described in Anderson et al. (2004) in two main respects: 1) the original algorithm used instantaneous H estimates during disaggregation, while the current approach uses the constant ratio; and 2) the original algorithm kept the ALEXI-derived air temperature constant and adjusted the high resolution Land Surface Temperature (LST) data, while the current approach keeps LST constant and instead adjust air temperature at lower resolution scale. This makes it more similar to the approach described in Cammalleri et al. (2013). However in the current study we go a bit further by evaluating 3 constant ratios for performing the disaggregation. In addition we do not perform the disaggregation on daily H fields, as is done in Cammalleri et al. (2013), and which would require multiplying the daily ratio by averaged daily net radiation or incoming solar radiation. Instead we use the constant ratio itself as a disaggregation factor. Finally, we apply the disaggregation to DTD derived fluxes, and not ALEXI derived ones, thus we are not able to set the initial blending height air temperature estimation to the ALEXI-derived blending height air temperature.
P13L25 – P14L6 and P14L23 – P15L2

The general idea of weighing individual pixels according to their contribution to the EC measured flux (p. 4870, lines 24-25) makes sense. However, turning to lines 22-28 on p. 4872, it is not clear whether this paragraph relates to missing Landsat pixels in the MODIS pixel and/or to missing Landsat pixels across the footprint of the EC measured flux. This impacts on the two data sets S75 and S100. It is also not made clear on p. 4876 (lines 14-28) and on p. 4877 (lines 1-10).

Set S75 contains dates when Landsat pixels contributing to at least 75% of the EC measured flux are present. The percentage of Landsat pixels missing within the modelled MODIS pixels can be smaller or larger than 75%. Set S100, which is a subset of S75, consists of points where additionally all the Landsat pixels within the modelled MODIS pixels are present. This does not necessarily mean that pixels representing 100% of EC measured flux are present, since sometimes the measurement footprint extends slightly beyond the modelled MODIS pixels. However, in practice all the dates in S100 set contain at least 98% of EC measured flux footprint. This will be clarified in the manuscript.

P17L26 – P18L9

The paper omits to describe explicitly whether results such as the 30-min flux values used in tables and figures relate to instantaneous observations and simulations or are average daily flux values. (I note that Norman et al. (2000) have shown how the DTD approach may be used to simulate sensible heat flux H throughout the day). Does one disaggregate instantaneous flux values or mean daily fluxes? The issue of noncongruent times of MODIS and Landsat observations needs to be discussed more extensively than is done at present (see lines 1-10, p. 4869). This is also important when discussing the use of the three ratios in the upscaling of instantaneous estimates to daily values.

All the results compare instantaneous fluxes modelled by DTD, TSEB or disaggregation algorithm against 30-minute averaged EC flux measurements containing the time at which the fluxes are modelled (i.e. the satellite overpass time). The disaggregation is also performed using instantaneous flux estimates. The use of mean daily fluxes is avoided altogether in this study by using the constant ratio factors directly during the disaggregation procedure (see the first answer), with the assumption that those factors do not change between the times of MODIS and Landsat overpasses. To obtain mean daily fluxes from the constant ratios, the ratios would have to be multiplied by the mean daily net radiation (if EF is used) or mean daily incoming solar radiation (for the other factors). Since the same mean daily net/solar radiation would be used for estimating mean daily fluxes from the instantaneous MODIS and Landsat fluxes, the mean daily net/solar radiation would just serve as a scaling factor. Therefore its use is redundant. An explanation will be added to the manuscript.

P13L25 – P14L6

On p. 4889 (lines 4-10): The original DTD formulation used the "parallel" resistance network (as expressed by A36) and in the new DTD formulation this "parallel" resistance network has been replaced with a "series" resistance" network (as expressed by A39). This is NOT what has been written in lines 6-10. This problem can be resolved by interchanging series and parallel in line 7 and replacing "latter" with "former" in line 8.

The sentence refers to the original DTD model using many of the TSEB equations presented in appendix A1, but replacing the series resistance network presented in A1 with parallel resistance network. The new (series) DTD formulation is mentioned for the first time on line 1 of page 4890. The text will be reworded to avoid confusion.

P34L23 – P34L25

Tables and Figures: State more explicitly in headings and captions of all figures and tables whether the data refer to instantaneous values at the time of the various satellite overpasses or to mean daily flux values Rewrite the Table headings for Tables 4 and 5 by stating more clearly that these results have been obtained *after replacing tower-based meteorological data inputs with ERA-Interim Reanalysis data inputs. ERA does appear in the heading but needs to be used in an explanatory sentence. The headings of Tables 4 and 5 refer to three approaches used for estimating the constant ration, whereas results for only one (EF) are presented. I assume that the headings for Tables 1-3 where just copied.*

All the results shown in tables and figures show refer instantaneous flux values. This will be clarified. Headings for Tables 4 and 5 will be rewritten to make it clear that ERA data inputs were used and to remove the reference to the three approaches of estimating constant ratio.
Table and figure captions

4861 13: Explain "which are not at the extremes of their distribution". Can this be rephrased?

This will be rephrased to "…, especially for estimates which are not either very high or very low.
P5L9 – P5L10

4861 25: "with certain heterogeneity". Is it important or not important?

A sentence will be added: This allowed us to evaluate the performance of the disaggregation algorithm in different ecosystems and at different spatial scales of heterogeneity.
P5L21 – P5L23

4861 3:"in most cases obtaining satisfactory results". A few references to recent studies are needed.

A reference will be added to Guzinski et al., 2013
P4L28

4861 29: "between the modeled canopy and soil fluxes". Explain the type of (energy) fluxes we are concerned with.

The interaction is between soil and canopy sensible heat fluxes. This will be clarified.
P5L26

4865 13: "Emissivity was linear scaled". You mean it was scaled linearly between NDVI =0.15 and NDVI =0.70?

Yes, it was linearly scaled between fractional vegetation cover of 0 (when NDVI <=0.15) and fractional vegetation cover of 1 (when NDVI >= 0.7).
P9L19 – P9L21

4865 18: explain "LST of 0 K"

When estimating the upwelling atmospheric radiance and transmittance, the LST was set to 0 K and emissivity to 1 to avoid any emitted or reflected long wave radiation signal from the surface. This will be clarified.
P9L24 – P9L25

4865 21: What is meant with "of the Corinne land cover"?

It's a pan-European land cover map. A reference describing the data set will be added.
P9L28

4866 7: "that it can be treated as blending height temperature". How does spatial resolution impact on vertical mixing.

At regional scale (which is the spatial resolution of ERA data set) the air temperature at blending height is more uniform than the 2 meter air temperature. Therefore it makes more sense to treat the ERA air temperature as a regional temperature at blending height.
P10L12

4867 9: What are "thermal-based energy balance models"? Do you mean "energy-balance models in which each of the terms of energy balance can be expressed as a function of temperature"?

They are the energy-balance models which utilize LST and air temperature measurements as the main inputs for deriving the land surface energy fluxes. "Thermal-based" will be changed to "temperature-based" to avoid confusion.
P11L13

4870 24-25: How does one "weigh each modeled pixel according to its contribution to the overall measured flux"? This needs rephrasing/more explaining.

The footprint model of Detto et al., (2006) estimates the source of the EC measured fluxes as a 2D grid of pixels, representing the relative contribution of each pixel to the total EC measurement footprint, with the sum of all pixels being 1. When evaluating the high resolution fluxes, each modeled pixel is scaled according to how strong is the contribution of its location to the EC measurement. This will be clarified in the manuscript.
P15L26 – P16L4

4872 10: What is meant with "native" MODIS resolution?

It means that the MODIS data was not resampled after obtaining it form the provider. The word "native" will be removed to avoid confusion.
P17L15

4872 23-24: "containing at 75% of the flux footprint weights" needs rephrasing/more explaining.

4872 28: Are we concerned about missing Landsat coverage within the MODIS pixel or about missing Landsat coverage within the tower flux footprint? Or both?

Please see the answer to the second question at the beginning of this document.
P17L26 – P18L6

26:"scaled by the fraction of the missing footprint" is not clear

This means that if, for example, pixels representing only 80% of EC footprint are present then the aggregated modeled flux is divided by 0.8 before it is compared with EC measured flux. This will be clarified.
P18L6 – P18L9

4873 10-11: "This is true for the dates both in S75 and S100" and "errors for dates in S100" You mean days here, but instantaneous data or daily averages?

Instantaneous data is used in all the comparisons.
Section 4.2 was rewritten and clarification added to table and figure captions.

4878 25: Explain "Once again points to underestimation of net radiation"

The strong underestimation bias in the modelled turbulent fluxes could indicate underestimation of net radiation. In accordance with suggestions of Reviewer 3, in the revised paper we will also evaluate the accuracy of modelled LE and Rn. Therefore this text will be rewritten.
Section 4.5 was rewritten and this statement removed.

4879 11: Can this be rephrased?

The statement will be rephrased to: Finally, we discuss the impact on estimated fluxes of using model meteorological data instead of measured data as input.
P24L25 – P24L26

Figures

The figure captions will be rewritten and S75 and S100 referred to whenever applicable.
Figure captions.

General editorial comments and corrections

Thank you for a very thorough review of the manuscript. We will reread the manuscript and implement all the listed corrections.
Throughout the manuscript

Review by reviewer 2

Section 4.3.1 : air temperature seems to be one of the main factors affecting the accuracy/ uncertainty in high-resolution modeled fluxes. Could the sensitivity of TSEB-DTD to initial air temperature be assessed ? Since the initial air temperature is modified in a two-step procedure (Section 3.3 page 5) 1/ by the physically based TSEB-DTD coupling scheme (step 5) and 2/ by an empirical smoothing filter (step 6), one may wonder how far the resulting air temperature is from the initial (step 3) and intermediate (step 5) value. When setting the initial value to ERA Interim (instead of tower measured) air temperature, would a lower difference between the resulting and initial value support the better accuracy in flux estimates ? What is the impact of the smoothing filter on the disaggregated fluxes?

This is an interesting suggestion and will be investigated in the revised paper. We will provide a plot of the average air temperature within the flux tower footprint at three different stages of the disaggregation procedure: 1) the initial air temperature (tower measured or from ERA Interim); 2) air temperature after step 5 of the disaggregation procedure, 3) air temperature after step 6 (smoothing) of the disaggregation procedure. This should help to answer the two questions posed by the reviewer: the impact of the difference between initial and resulting air temperature values and the impact of the smoothing filter on the accuracy of modelled fluxes. However, it should be noted that the value of the derived air temperature does not necessarily reflect the actual air temperature. This is because the value is derived to compensate any errors in the Landsat LST and to ensure consistent fluxes between the DTD and disaggregated estimates.

P23L9 – P23L24 and Figure 12

Review by Reviewer 3

General comments:

1. The adopted methodologies are sound and described in detail, but I do miss a better overview of the different model implementations and model runs being evaluated – a table or flow diagram with a clear outline of the tested configurations would be very helpful.

A flow diagram and the description of the model configurations used will be added.
P15L6 – P15L16 and Figure 3

2. I'm also missing some details on the Landsat data used; specifically number of valid acquisitions, their frequency and distribution over the growing seasons for the studied period (again a table/figure showing this along with e.g. LAI dynamics would have been helpful)

We will provide a figure showing the temporal distribution of Landsat scenes and the magnitude of LAI in those scenes.
P9L18 – P9L19 and Figure 2

3. Landsat-scale fluxes are extremely useful for applications that require sub-field scale resolution but the temporal frequency of valid acquisitions is an issue. This point should at least be mentioned in the paper with reference to recent work that tries to capitalize on the spatial detail of Landsat and high temporal frequency of MODIS for producing daily surface energy fluxes at the Landsat spatial scale (e.g. Cammalleri et al., 2013; Gao et al., 2006).

We are aware of the above studies and will add a mention in the discussion section of the paper.
P26L29 – P27L6

4. A key concern is the lack of LE validation at the Landsat scale. Only the sensible heat fluxes are currently being compared against the tower observations all though for many applications the latent heat fluxes are of most interest. I would highly recommend including LE in the Landsat-scale flux validation. A full evaluation of the surface energy fluxes (H, LE, G and Rn) would have been even better. This would also provide a better basis for diagnosing the cause of the estimation uncertainties.

We have decided to validate only the H fluxes in the original manuscript to improve the clarity of graphs and the discussion. Since LE is estimated as a residual, the accuracy of H estimates should translate into the accuracy of LE estimates. However, we are modelling all the fluxes and will present validation of all of them against ground measurements in the revised paper.
Sections 4.2 – 4.5

5. Validation of vegetation inputs (LAI) would also be useful for this purpose (if possible)

Unfortunately field measurements of LAI were only undertaken in 2013 and only in one of the fields covered by the agricultural flux tower footprint. We will assess whether this data can be used for validation of satellite based LAI estimates. We will also add a discussion of uncertainties introduced by LAI estimates.
P24L19 – P24L22

6. The paper would benefit from a careful read through to correct for some language issues and several sentences should be shortened and rewritten for improved clarity and flow.

The paper will be edited to improve the language.
Entire manuscript

Specific comments:
1. Abstract: The abstract should also mention TSEB as it forms the basis for the Landsat scale fluxes.

A mention of TSEB will be added to the abstract.
P2L3

2. Page 4858 L18: Firstly, only the surface heat fluxes are being evaluated which should be made clear here (or seriously consider including latent heat fluxes in the validation). Secondly, the accuracy very much depends on the dataset used (S75, S100) and the source of the meteorological input (Table 2, 4). It is important to mention this in the abstract.

Latent heat fluxes will be validated in the revised paper (see general comment 4). We will also clarify that the statistics mentioned in the abstract are coming from the optimal case.
P2L17 – P2L19 and section 4.2 – 4.5

3. Page 4859 L1-2: A full evaluation of the surface energy fluxes (LE, H, G and Rn) would have been welcome, and is typically done (e.g. Kustas et al., 2012, Advances in Water Resources, 50, 120-133). As a minimum LE validation results should be included.

Please see answer to general comment 4.
Section 4.2 - 4.5

4. Page 4859 L11: Should also mention thermal resolution of L8 (100m).

The resolution of L8 will be added to the sentence.
P3L11 – P3L12

5. Page 4860 L3: "more robust TSEB" – more robust compared to what?

The sentence tries to convey that ALEXI and DTD are more robust than pure TSEB, since they use changes in the observed temperatures instead of the absolute values. This will be clarified.
P4L1 – P4L4

6. Page 4864 L24: thermal observations resampled from 120 (L5), 60 (L7) and 100 (L8) to 30 m by…

The sentence will be updated.
P8L27 – P8L28

7. Page 4865 L1: The spatial and temporal resolution of the MOD08 gridded products used should be mentioned.}

The sentence will be updated to include the spatial and temporal resolution.
P9L5 – P9L6

8. Page 4865 L10-13: Resulting time-series of Landsat/MODIS scale LAI over the study sites should be displayed and validated (if possible). Large uncertainties are associated with the MODIS LAI in some cases and these will translate into the Landsat scale LAI as Feng Gao's regression tree approach produces MODIS consistent LAI. Given the importance of accurate LAI for surface flux mapping, uncertainties in LAI over the sites should be quantified (if possible), or at least mentioned (with reference to appropriate literature on MODIS LAI uncertainties).

Please see answer to general comment 5.
P9L17 – P9L19, P24L19 – P24L22 and Figure 2

9. Page 4865 L21: Reference for Corine land cover missing.

The reference will be added.
P9L28

10. Page 4865 L24-26: Radiation inputs may also be supplied by geostationary satellites (GOES, Meteosat).

That is true but other required inputs, such as wind speed or air temperature, cannot be reliably estimated from geostationary satellites over land surface. In addition the study area is located at quite high latitude (around 56 degrees north) and therefore estimates derived from geostationary satellite observations might not be accurate due to longer atmospheric path.

11. Page 4866 L20. Duplicate symbols used for the view zenith angle (VZA or sigma) – choose one.

VZA is an acronym, while sigma represents the value of VZA in model equations. We think that the acronym and the symbol play a different role in the paper but will review the text to reduce any confusion caused by this, if present.
Appendices

12. Page 4868 L15: The ALEXI pixel resolution depends on the resolution of the geostationary satellite and may range from 3 – 10 km.

The text will be updated.
P12L16 – P12L17

13. Page 4868 L25-27: Not sure this is true; I believe a fairly coarse resolution air temperature prediction is needed for approximating a regional blending height temperature (as the authors mention themselves in the discussion – page 4879).

Blending height air temperature is certainly more spatially uniform than near-ground air temperature and we make use of this in the discussion on page 4879 and in other parts of the paper (e.g. section 2.3). However, in highly heterogeneous landscapes, especially when strong LST contrast is present, the blending height temperature might not be constant due to advection and localized couplings between the surface and the atmosphere (Anderson et al., 2004). This will be clarified in the manuscript.
P12L26 – P13L1

14. Page 4869 L20-23: Not sure I understand correctly. In any case, validating LE would help determine if this is the case.

Since LE is estimated as a residual of the other fluxes, the errors present in the other fluxes estimates will contribute to error in the LE estimate. H is calculated directly from model inputs, and so the errors in other flux estimates do not contribute directly to errors in H estimate. Therefore it can be assumed that errors in modelled H will be smaller than the errors in modelled LE. This will be clarified. Validation of LE will also be performed (see general comment 4).
P13L21 – P13L24 and Sections 4.2 – 4.5

15. Page 4870-4869: Consider creating a flow diagram of some sort depicting the required steps and model configurations used. That would be very helpful.

Please see general comment 1.
P15L6 – P15L16 and Figure 3

16. Page 4807 L20-21: It says here that disaggregated fluxes of both sensible and latent heat are being evaluated against the tower observations. As far as I can tell only sensible heat flux is being validated (Table 2 – 5). LE validation is only done at the MODIS scale using DTD (Table 1).

Please see general comment 4.
Sections 4.2 – 4.5

17. Page 4871 footprint model: Mention if the footprint model is considering atmospheric stability.

The Hsieh et al. (2000) footprint model does take atmospheric stability into account. This will be clarified in the text.
P15L24

18. Page 4871 section 4.1: The DTD results are characterized by really high RMSE and low correlations even for the series implementation. It may partly be a footprint issue (as shown for H) but unfortunately the effect of the disaggregation on LE has not been assessed. Errors in the vegetation inputs and H/ET partitioning are other plausible causes. I think the authors mentioned this briefly in the text, but a fix should be implemented (or screened out if it can be justified) to avoid the large number of pixels with zero LE (modeled) (Fig. 2).

Whenever DTD cannot obtain a plausible solution at the end of model run (LE < 0), it is assumed that this is due to dry conditions and therefore there is no evapotranspiration. In those cases LE is set to 0 and the net radiation is split between H and G (see Appendix for details). This is a backup behavior of the model and can lead to inaccurate results when the lack of convergence in the model was not due to low ET but due to, for example, wrong inputs. This is happening much more frequently in the forest site than in the agricultural site which could point to inaccurate characterization of MODIS LAI or LST in the forest. However series resistance network leads to reduction in the number of those cases. In the revised manuscript we will show the results both with those cases included and excluded, since this is not the primary model behavior. Disaggregated LE will also be evaluated.
P16L15 – P16L24, Table 1 and Figure 4

19. Page 4873 L13-15: It would be helpful having a figure showing the frequency and distribution of acquired Landsat scenes over the growing seasons, overplotted on LAI time-series for example.

Please see general comment 2.
Figure 2

20. Page 4873 L27 to 4874 L9: Please rewrite this part for better clarity, some parts (e.g. less points in panel (d)…) I don't fully understand.

The section will be rewritten to improve clarity.
P19L13 – P19L18

21. Page 4875 L13: "reasonable accurate" I guess it depends on what you define as reasonable: : :the statistics in Table 3 indicate some major model performance issues.

In this sentence "reasonably accurate" refers to EF approach producing most balanced results at GLU, as is elaborated on page 4878 L4-11. Once LE validation is performed this should, hopefully, become clearer.
P20L25 – P20L27

22. Discussion section 4.4: There are some very good points and observations here but try to write it in a more concise (short sentences) and to the point manner, as far as possible.

This section, like the rest of the paper, will be edited to improve the language.
Section 4.5

23. Page 4878 L25: Validation of net radiation could have been more integrated in the result section (along with LE and potentially G), for a full evaluation of surface energy fluxes (as the manuscript title suggests).

Please see general comment 4.
Sections 4.2 – 4.5

24. Page 4880 L26-27: The RMSE and bias depends on the dataset used S75 versus S100), and a bias of -14 W m-2 (S100) is not negligible and the correlation (0.65-0.94) does not constitute a perfect (1) agreement. Please be precise in your description of performance metrics.

This sentence referred to the optimal case which is S100 dataset with ERA interim meteorological inputs, as shown in Table 4. However, the description of results will be made more factual and quantitative.
P26L9 – P26L12

25. Appendix A: It is very detailed and most of the equations have already been listed in other publications (Norman et al., 2000), but on the other hand it is also helpful with a repetition and complete list. Please check that all parameters and symbols have been properly defined.

The purpose of the appendix is not to describe new model developments but to clearly show the model implementation used in this study. It is often the case that when model equations come from a number of different papers it is unclear which formulation was actually used, thus making it harder for others to replicate the results. We will add a clarification at the beginning of the appendices explaining this point. We will also double check the presented model formulations.
P27L8 – P27L11

Page 4882: Try to avoid having duplicate symbols for the same variable (LAI and F, VZA and sigma).

Please see specific comment 11.
Appendices

How is the vegetation height parameterized (measured, seasonally variable, empirical function of NDVI)?

Vegetation height at the coniferous forest was kept constant, while at the agricultural site it varied between a minimum and maximum limits based on an empirical function of LAI. This will be clarified.
P28L17 – P28L22

P4882 L21: "vegetation width" – do you mean leaf width or row spacing?

Vegetation width refers to the plant crown width. This will be clarified.
P28L11

26. Figures: very small text size – difficult to read properly

The figures will be redone will larger text size.
Figures

Language issues (just a few selected issues listed. Careful review needed to correct for many, mostly minor, language and flow issues):

Please see general comment 6.
Entire manuscript

Other changes

A number of additional changes were implemented in the study, which impacted on the results presented in the manuscript. The most significant were:

- at GLU LAI was not divided by fraction of vegetation that is green (**P8L14 – P8L24)**
- 8 day MODIS LAI product was used instead of 4 day to improve LAI accuracy
- downscaling of LAI and albedo was redone with better settings in the regression tree
- points where measured H or closed LE had implausible values were removed (**P8L1 – P8L3)**
- S75 become S70 to increase the number of points
- models were run with full ERA-Interim meteorological inputs and with ERA-Interim meteorological inputs except for incoming shortwave radiation which was obtained from tower measurements (**Section 4.4**)