Reviewer 1:

General comments:

The paper explores a novel approach for estimating ET, or more specifically, several parameters required to model ET using satellite imagery, from remote sensing data. While the approach has limitations, which are acknowledged by the authors, it has the potential to be a useful tool for enhancing ET estimates over large scales by providing observation-based estimates of the parameters needed for ET modelling. Although the paper presents the initial evaluations of the approach and additional studies are needed to refine and confirm the utility of the methods presented, the potential benefits of the technique to the modelling community are sufficient to merit publication. Nonetheless, there are several aspects of the study that need clarification of further explanation (see below); the authors are also urged to work with a grammarian to improve the English syntax and structure of the paper.

We thank Reviewer 1 for their supportive and helpful comments and will modify the text accordingly.

Specific

1. P2, L42: The “scales relevant to decision making” should be defined more rigorously. What are these scales?

These scales range from 0.01° to 1 degree. This is now mentioned in line 47 – 48.

The AIRS sounder scales are 100 km but geostationary sounders will come down to the 1 km scale. As a result, we speculate that this method could in principle furnish information from the field to the landscape scale and it is these scales we believe that are relevant to decision making on water resources.

2. P4, L89: These “counter arguments” need to be explained (justified) more fully. For example, the Bowen ratio is determined in terms of gradients in Eqs. 5 and 6. It is not clear how well measurements of temperature integrated over a volume of the atmosphere will represent the true gradient, particularly if these quantities vary nonlinearly with height? Also, relationships the authors present are based on assumptions of similarity (e.g. the eddy diffusivity terms are not shown in Eq. 5) given the course horizontal and vertical spatial scales of the soundings data, the authors need to justify the assumption that the measurements of the turbulent transport of heat and water are within the surface boundary layer and it conforms to similarity theory.

The ‘counter arguments’ are now mentioned in line 95 – 100. **These same arguments can be made for towers in that any sensor measurement contains some element of vertical integration of the observation. So, although the scale at which we are applying the method is unusual, the principle of using observations as point estimates is no different. However, sounders integrate signal horizontally over scales of thousands of square kilometres and hence benefit from stronger horizontal averaging characteristics in the measurement relative to the vertical. We accept they also suffer from ambiguities in the vertical integration of signal but the effective vertical sensor separation is large. The point about nonlinearities in the vertical is**

important though. We correct for the environmental lapse rate using the geopotential height although we accept there could be other things we are not correcting for.

Regarding the equation 5, based on the similarity hypothesis the eddy diffusivities are cancelled as mentioned in line 130 – 131 and the implications of the similarity hypothesis assumptions are also mentioned in the discussion section line 385 – 394.

3. P6, L129: The authors use the 1000 mb and 925 mb levels to estimate the temperature and pressure gradients. Since especially the latter is unlikely to be within the constant flux layer, it is questionable that the underlying assumptions of similarity, etc. are valid. Or, why they should be expected to be valid? While the authors point to studies such as Swinbank and Dyer (1967), those analysis are based on the assumption that the sources and sinks for heat and moisture identical and uniformly distributed. The large spatial scales corresponding to the sounding measurements place that assumption in doubt.

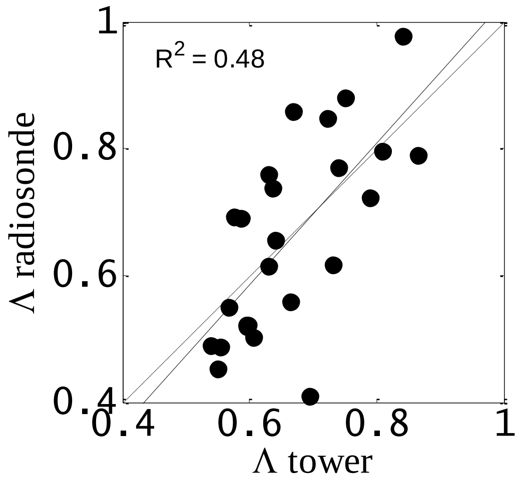
[In response to Reviewer 2 also on related points] We accept that sampling the boundary layer at this scale is likely to lead to non-idealised conditions in terms of the way the Bowen ratio was originally conceived (although such idealised conditions seldom exist at any scale), so the question is, to what extent do they impede the method? Obviously, this is a non-trivial question but we propose the following defense (*The details are explained in section 2.1 from line 147 – 197*).

The soundings we utilize are for a 13:30 overpass time. Although not universally so, the turbulent boundary layer tends to be at its most mature around or soon after this time of day and the average depth of the turbulent boundary layer should extend well beyond the 925 mb level in all but high latitudinal environments (Fisch et al., 2004). Therefore, although the system we are sampling is not the constant flux region near the surface, in affect we have a surface source region (sampled by the 1000 mb sounding) exchanging with a well-mixed volume (sampled by the 925 mb sounding). The flux exchange between these two should be approximately linear and equivalent in the concentration differences between the two providing we are near dynamic equilibrium (i.e. the turbulent boundary layer isn't growing/contracting excessively) and that additional fluxes into and out of the boundary layer (including phase changes) are small relative to the surface sourced fluxes of heat and water vapour. The 1000 mb level would correspond to heights of approximately 10 m and the 925 mb level to approximately 500 m, thus Δz = 500 m and the effective measurement height is, therefore, at approximately 300 m. This contrasts with horizontal scales of the soundings of 100 km i.e. nearly three and a half orders of magnitude larger. Therefore, although advective fluxes affect the vertical profile, they should be small relative to the effects of vertical exchange on these scales. The principle difficulty as far as we can ascertain is the effect of phase changes associated with cloud formation, producing latent warming of the boundary layer whilst removing water vapour. Providing this happens above the 925 mb sounding we anticipate it being less of a problem, but if it happens below it then clearly this is problematic. Of course, this also impacts on the estimation of the net available energy.

In an attempt to reassure the reader about the validity of the assumptions we are making we have also tested the same methodology over a surface flux measurement site in the central United States where both the radiosonde measurements and eddy

covariance flux observations were available. Bowen ratio was estimated from the air temperature and dewpoint temperature measurements of the radiosonde observations using the same methodology as described in the manuscript. The Bowen ratio was then converted to the evaporative fraction and these were then compared with the eddy covariance derived evaporative fractions. The figure of this comparison is given below and shows a fair degree of correspondence between the two. This analysis produces a modest correlation (R^2 = 0.48), reasonably low RMSE (0.11) and mean absolute percent deviation (14%) between radiosonde derived evaporative fraction and tower observed evaporative fraction.

This figure is now added in the manuscript (Figure 2) and necessary explanations are elaborated in the result section (section 3.1) from line 269 to 282.



4. P8, L187: With the possible exception of tall tower data, the source area of EC flux measurements is typically much less than 10 km2 so the mismatch in the scale of the source areas between the tower and soundings data likely exceeds three orders of magnitude. The smaller source area of EC systems would tend to mask the impacts of spatial heterogeneity that would be seen at the coarser resolution of satellite data.

We agree. This shouldn't impact on the flux measurements themselves for the horizontal vs. vertical scales arguments made above. The most important implications for spatial heterogeneity in the present context is that, in addition to complicating comparison with tower data, relating these observations to unique surface characteristics is likely to be problematic.

5. P10, L219: A correlation of 0.34 may be statistically significant, but it suggests only a modest relationship between the satellite and tower-derived estimates of evaporative fraction. Moreover, from Fig. 2 it appears that the maximum tower measurements of latent heat flux are about 350 Wm-2; in that case, an error of 79 Wm-2 would be an error of 20% to 25%.

Point taken, although we are not claiming high accuracy and precision for the method at this stage. Indeed, such things may be unattainable with this method - time will tell.

Technical

1. P2, L36: It’s less managing climate change than its effects.

This is line 38 now and is changed as ‘to manage the effects of this change’

2. P3, L53: The sentence beginning here is awkward and confusion. It needs to be rewritten.

This is line 55 to 59 and changed as ‘The most common approach centres on assuming a physical model of evaporation given many of the variables required to compute evaporation using these models are available directly as satellite products (e.g., land surface temperature, vegetation index, albedo etc.) (Choudhury and Di Girolamo, 1998; Mu et al., 2007, 2011).’

3. P5: The authors use a number of non-standard symbols (e.g. using phi to represent available energy and P to represent water vapour pressure). Using more typical symbology would avoid confusion for the reader. Also, the authors need to confirm that all of the symbols are consistent. For example, the authors inter-mingle “z” and “Z” to represent height.

Symbols are made consistent now and our general notation is in line with some classic studies in this area (e.g., Raupach, 1995).

4. P6, L121: The word “plain” is misspelled.

Corrected now.

Reviewer 2:

General comments:

The authors propose a novel application of an old method. Instead of applying the Bowen Ratio as a micrometeorological method used at the field-scale, they make use of satellite soundings to apply the Bowen Ratio method to large spatial scales. They test the method against flux tower measurements.

Specific

1. Given the rather large leap in spatial scale of the application of the Bowen Ratio, it might be informative to more systematically evaluate why the original assumptions (and typical assumptions of most subsequent uses) of the Bowen Ratio Energy Balance method remains valid or nearly valid in this modified use of the method. Typically, there are three basic assumptions layed out for the Bowen Ratio method to work: i. 1-D vertical transport, ii. homogeneous land surface, iii. steady-state conditions (Fritschen and Simpson, 1989, Journal of Applied Meteorology, 28: 680-689).

This is a very good point, and not something the original paper sets out explicitly. To this we could add similarity in the vertical transport mechanisms for heat and water vapour. We set our position on these assumptions in quite some detail (we hope) in our response to R1 so rather than repeat them here we direct the reader to that response. The revised manuscript contains a much more explicit handling of these core assumptions. These are described in *in section 2.1 from line 147 – 197*.

2. The dramatic difference in spatial scale between the flux tower footprint and this satellite based method would seem to inherently limit the ability of the flux towers to validate the method (a point made on Lines 350-355). Lack of energy balance closure at the flux tower sites is also an issue. Thus, it could be worthwhile to consider other methods to estimate actual ET that would be more consistent in terms of scale. Namely why not use an annual water balance method (P-Q=ET assuming minimal storage) for gaged watersheds at the approximately 1° scale? One would look at differences in ET across multiple years instead of monthly differences within the same year. Given the sometimes questionable relationship between this Bowen Ratio method and flux tower measurements at certain sites (Figure 3), this additional method of validation would provide a further check on the soundness of the method.

We have considered at length using the annual water balance method (and others) to validate the sounding approach. However, in our judgement the cumulative uncertainties associated with deriving annual budgets from uncertain and often biased precipitation and flow data we feel prevents this from being a robust test. We have however developed an additional test we believe is quite stern, namely exploiting radiosonde data in conjunction with tower fluxes (please see our response to R1). Again, we stress we envisage the need for significant further evaluation of the proposed methodology but believe enough is offered in the current manuscript to warrant publication at this stage.

However, in an attempt to reassure the reader about the promise of the method we have also tested the same methodology over a surface flux measurement site in the

central United States where both the radiosonde measurements and eddy covariance flux observations were available. Bowen ratio was estimated from the air temperature and dewpoint temperature measurements of the radiosonde observations using the same methodology as described in the manuscript. The Bowen ratio was then converted to the evaporative fraction and these were then compared with the eddy covariance derived evaporative fractions. The figure of this comparison is added as Figure 2 and detail explanation is given in *in the result section (section 3.1) from line 269 to 282.*

3. On Lines 355 to 380, the manuscript reports RMS errors in latent heat fluxes from other studies. It suggests that errors observed in this study are comparable to those in other studies. However, nearly all these other studies use daily fluxes or less. It would seem that the RMS errors for daily values represent something somewhat different, namely the ability to replicate daily variations in fluxes and maybe (but not necessarily) consistent bias in estimates. At a monthly scale, RMS error would seem to more strongly suggest bias since daily variations would cancel out over the longer averaging window. In essence, I might expect that a suitable RMS error for a monthly averaging window would be lower than that for a daily averaging window (much the same way it is often easier to model monthly streamflows relative to daily streamflows). Are there any additional studies that report RMS error for monthly fluxes?

In the present case, the reported RMS error is the RMS error of monthly averages of 13:30 hours ET. There are additional studies that reported RMS error of monthly fluxes (for example, Cleugh et al., 2007; Mu et al., 2011). In these studies daily ET was modeled using daily radiation and meteorological variables and monthly fluxes were generated from the daily averages. Cleugh et al. (2007) reported RMS error of 27 W m^{-2} over two contrasting sites in Australia using tower meteorology and MODIS vegetation index over the eddy covariance footprints. Mu et al. (2007, 2011) reported RMS error 8 – 180 W m^{-2} on eight-day average ET and 12 mm on monthly average ET. *These are now cited in the manuscript in line 443 to 448.*

<u>Minor Comments:</u>

a. Lines 89 to 91 – Could this statement be explained in greater depth. Not enough context for most readers to understand the exact reasoning. May be a citation to something else?

The description is expanded as **'Firstly, the degree of signal integration going on at the scale of the satellite sounding should help relax the requirement on signal resolution. Sounders integrate signal horizontally over scales of thousands of square kilometres and hence benefit from strong spatial averaging characteristics in the measurement, despite suffering from ambiguities in the vertical integration of signal. However, this later drawback is aided by an effectively large sensor separation in the vertical (Thompson and Hou, 1990).' *This is now line 95 to 100.***

b. Eqn's 4a, 4b, 5 and adjoining text. Inconsistent notation in terms of T, P, and Z. Sometime capitalized, sometimes not.

Necessary corrections are incorporated.

c. Line 224 – Make clear that the values in this expression came from the relationship shown in the caption of Figure 2. If one does not immediately look at the Figure 2 caption, it is confusing were this expression came from.

Figure 2 is now Figure 3. Necessary corrections in the descriptions are made.

d. It would be helpful to repeat a full description of biome types in Figure 2. The abbreviations are not totally obvious and somewhat awkward to flip back to Table 2 depending how paper is layed out.

Figure 2 is now Figure 3 and biome types are elaborated as suggested.

e. What is N in Table 2? Is this the number of months of observation across all sites within each biome? Please clarify.

N is the number of datasets in Table 2. This is now clarified in the revised version.

References:

Cleugh, H. A., R. Leuning, Q. Mu, and S. W. Running (2007), Regional evaporation estimates from flux tower and MODIS satellite data, Remote Sens. Environ., 106, 285 – 304.

Fisch, G., Tota, J., Machado, L. A. T., Silva Dias, M. A. F., da F Lyra, R. F., Nobre, C. A., Dolman, A. J., and Gash, J. H. C.: The convective boundary layer over pasture and forest in Amazonia. Theoretical and Applied Climatology, 78, 47 – 59, 2004.

Mu, Q., F. A. Heinsch, M. Zhao and S. W. Running (2007), Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519 – 536.

Mu, Q., M. Zhao and S. W. Running (2011), Improvements to a MODIS Global Terrestrial Evapotranspiration Algorithm, Remote Sens. Environ., 115, 1781 – 1800.

Raupach, M.R. (1995). Vegetation-atmosphere interaction and surface conductance at leaf, canopy and regional scales. Agricultural and Forest Meteorology, 73, 151 – 179.