We thank the reviewers for their helpful comments and suggestions. Our point-by-point responses are provided below (in blue italic).

Comments by Anonymous Referee #1

This paper addresses the sensible and latent heat fluxes over the Tibetan Plateau, where ground-based measurements are very sparse. As the important role of the Tibetan Plateau on Asian monsoon, discussion on this issue is important and encouraging. The authors made their contribution by fusing ground-based measurements, reanalysis, and satellite data. The paper was well organized and written. The Introduction is very good.

There are many evidences showing that current estimates of sensible and latent heat fluxes over the Tibetan Plateau have huge uncertainty (i.e., Zhu et al., 2012), including those as input of this study. The author fused the input data by a linear regression. This regression did not improve the estimates physically. However, the author obtained nice "validation" results. The major reason for this is the "cross-validation" used by the authors (Page 30359). It is not a real validation. The author used ground-based measurements to calibrate the regression method and validate the regression at same time. By this mathematical calculation, one can always obtained good validation results no matter how good the input data are. However, this does not means that the estimates of sensible and latent heat flues by this regression are good elsewhere except for the validation sites.

Zhu X Y, Liu Y M, Wu G X. An assessment of summer sensible heat flux on the Tibetan Plateau from eight data sets. Sci China Ear th Sci, 2012, doi: 10.1007/s11430-012-4379-2

Since previous studies have suggested that there are uncertainties of the sensible and latent heat fluxes over the Tibetan Plateau from existing datasets (Zhu et al., 2012), it is thus of great *importance to integrate ground observations, reanalysis, and remote sensing datasets to characterize the sensible and latent heat fluxes over the Tibetan Plateau.*

We did not explain the cross-validation method in detail in the manuscript, which led to the major comment on the method from the reviewer. We agree that the cross-validation is not a real validation since the ground measurement is used to both derive the regression method and validate the accuracy. Therefore, we apply the leave-one-site-out cross-validation method (Shi and Liang, 2013) instead of the conventional cross-validation method to quantify the accuracy of the estimated sensible and latent heat fluxes.

We have replaced the description of the leave-one-site-out cross-validation method by the detailed descriptions: "The leave-one-site-out cross-validation was conducted by applying the fusion model after removing a site and validating with the unused site. The root mean square error (RMSE) from cross-validation (RMSE_CV) was computed as the average RMSE after applying this process multiple times (omitting one site for validation each time). RMSE_CV was used to validate the fused SEB components."

Review by C. Jimenez (referee)

General comments

This paper describes a data fusion approach to merge heat fluxes from in situ, reanalyses, and remotes sensing products over the Tibetan Plateau (TB), followed by an analysis of temporal and spatial variations of the fused fluxes over the TB for 1984-2007, including an attempt to link these variations to variations of other meteorological parameters, mainly by looking at correlations between the different parameters and the fluxes.

The paper reflects a large amount of work. Condensing all this information in the paper makes some times the paper difficult to read, and perhaps 2 companion papers may have been a more focused choice (a first one focused on the data fusion technique and evaluation of the fluxes, and a second one with the regional analysis).

Considering that the objective of this paper is to analyze the spatiotemporal characteristics of the surface sensible and latent heat fluxes over the TP by integrating ground observations, reanalysis, and remote sensing datasets, we have not separated it into 2 companion papers to ensure the integrity of the presented study. The multiple linear regression has already been applied as a data fusion method to generate the fused surface radiation budget over the Tibetan Plateau (Shi and Liang, 2013), so the focus of this paper is to quantify the surface sensible and latent heat fluxes over the TP in the most recent two decades, rather than the evaluation of the data fusion method.

My main concerns related to the methodology are:

1. The data fusion assumes that the in situ observations are the "truth", and uses the other fluxes to spatially extrapolate the in situ observations. This is an approach as valid as others (e.g., to spatially extrapolate by using flux related forcings as in Jung, 2009), but the part of the evaluation based on the in situ data is a bit misleading as the fused data by construction tries to "mimic" the in situ fluxes. The "leave-one-out cross validation" does not change this basic fact. I think this should be better conveyed in the paper.

We agree that the data fusion method assumes that the in situ observations are the "truth". The data fusion method in this paper directly integrates multiple state-of-art surface energy budget products (reanalysis datasets, remote sensing product, and in situ observations), which is different from other empirical methods that estimate the surface energy budget components using statistical models trained by multiple explanatory variables and ground observations (Jung et al., 2009; Wang et al, 2008). The "leave-one-out cross validation" is actually a "leave-one-site-out *cross validation", which ensures the independency between the site used for validation and the sites used for developing the fusion coefficients. Please see also our reply to Anonymous Referee #1.*

2. The choice of products to merge may be questioned. If the reanalyses are well covered, the reader may question the choice of one single remote sensing flux product. Why not having others? This is an important part of the paper, as the reader my start thinking how the spatial-temporal variations and trends may be affected by a different choice of products. In my opinion, this fusion product would have been more focused if it were only based on in situ and reanalysis. If remote sensing products want to be added, only one does not seem the best choice.

We have added the justification of the choice of the remote sensing dataset. The reviewer is correct that using multiple high-quality remote sensing products will improve the result. We have included a recent global remote sensing-based evapotranspiration product by Miralles et al., (2014) (hereafter GLEAM-LE) in the data fusion. Another remote sensing dataset that cover the similar temporal period (1984-2007) by Vinukollu et al., (2011) are not included in the fusion of latent heat flux because of its relative low accuracy (results not shown in the manuscript). We excluded the sensible and latent heat flux product by Jung et al., (2009), which used the AsiaFLUX site over the TP (HBM) in the method development. The validation shows that the remote sensing-based latent heat flux by Zhang et al., (2010) (hereafter Zhang10) and GLEAM-LE achieve comparable RMSE, MBE, and R^2 *to that of reanalysis datasets, including MERRA, ERA-Interim, and JRA-25 (Table 3). We have developed two sets of fused latent heat flux, one using ground measurement and reanalysis datasets, and the other using ground measurement, reanalysis, and remote sensing datasets (Table 4). -We have added the following sentence in section 4.1: "Compared with the fused result by only using only in situ and reanalysis data*

(MLR_RA in Table 5), the MLR model lowered the RMSE_CV by including the two remote sensing products." Therefore, the spatiotemporal analysis was based on the fused sensible and latent heat flux from in situ, reanalysis, and remote sensing products. Since we have included the GLEAM-LE in the data fusion scheme, the result has been updated throughout the manuscript.

3. The choice of not getting a fused sensible heat flux (H) product, but deriving one by closing the SEB with another fused net radiation product, requires to be better explained as it is a very important choice in the methodology. For instance, does the choice imply that we "trust" more the latent heat fluxes (LE) from the different products, compared with the Hs? Or, the choice of closing the surface balance with the radiation product, does it mean that we assume that the difference between the radiation product and the LEs are less uncertain that a fused H product? Perhaps fused LE, H, and ground fluxes (G) and the comparison of their sum to the net radiation product could have been useful to identify problematic periods-regions of this methodology (i.e., lack of the expected closure as a sign of not well characterized periods-regions).

We have added justification to the choice of the method to estimate the sensible heat flux. In the approach of estimating the sensible heat flux based on surface energy balance, we assumes that the latent heat flux and the radiative fluxes are more reliable than that of the sensible heat flux from multiple reanalysis and remote sensing. We have added the comparison result in section 4.1: "An alternative approach to estimate the sensible heat flux by fusing in situ, reanalysis, and remote sensing datasets directly has a higher RMSE_CV (by 21.2 W m^{-2}) than that by computing it from the surface energy balance." We agree with the reviewer that it could be useful to diagnose the problem of the methodology by using the lack of the expected closure. However, the energy balance of the fused datasets is closed since the proposed method estimates the sensible heat flux by subtracting the fused latent and ground heat flux from the previously developed net *radiation (Shi and Liang, 2013). Therefore, the suggested diagnostic analysis by using the lack of the expected closure was not performed in this study.*

Regarding the presentation of the findings, the paper is not badly written, but sometime the sentences are long and difficult to interpret. The discussion section seems more focused on the methodology (errors, sampling, other methods, etc), while the discussion about the flux variations and trends is mainly given in the results section. Perhaps "discussion" is not the proper name of that section. I think a bit more of care to say CMA stations (and not just stations) may help the reader (very easy to confuse the words sites and stations). The figure and table captions need to be improved, they are too concise to allow a basic interpretation of the plots.

Effort has been made to improve the presentation level, especially the clarity of sentences. We agree with the reviewer that the discussion section is not clearly constructed. Accordingly, we have moved the data fusion method descriptions in the discussion to the methodology (Section 3.2), merged the comparison to other fusion methods and other datasets in the discussion with the validation results (Section 4.1), and added the rest part of the discussion of as Section 4.3 (Uncertainties of the fused surface energy fluxes). We have changed the Section 4 to "Results and Discussions", because the discussions of methods and flux spatiotemporal characters have already been included in this section. We have changed to "CMA stations" throughout the manuscripts to avoid confusions. We have revised the figure and table captions by adding detailed information to facilitate interpreting the result.

Specific comments

P30355-L5. Interesting to see this quoted 5-20% error figure, when apparently the EC measurements have on average a 20% lack of closure when compared with the measured net radiation (ref). It may be worth commenting on this here (later on in the article a correction is applied to the EC fluxes).

It has been added to illustrate that a correction has been applied to the EC fluxes since EC measurements have approximately 20% lack of closure.

P30358-L25. As the remote sensing and reanalysis has full time coverage, were the monthly means of the in situ data matched in time to those ones (i.e., all datasets use the same number of days for each monthly mean calculated)?

We have added detail descriptions in the methodology of computing monthly means of in situ data. To ensure sufficient temporal coverage of valid in situ data (to derive monthly means), daily mean value of the in situ data was calculated from daily hourly mean values with less than half missing, and the monthly mean value was calculated from the daily mean covering more than 15 days in a month.

P30360-L20. The fact that the MSE and RMSE of the LE and H is larger than the G is expected, giving the much larger LE and H absolute values in general.

We thank the reviewer for highlighting this point. This sentence has been removed.

P30361-L14. This lowest RMSE_CV is not unexpected at all, as commented above. Same comment applies to P30362-L12.

The lowest RMSE_CV was by comparing the fused result to individual datasets rather than comparing the RMSE_CV of the latent or sensible heat flux to that of the ground heat flux. We have modified this sentence: "the fused sensible heat, latent heat, and ground heat fluxes exhibited the lowest RMSE_CVs over those by using individual datasets by 14.3 Wm^{-2}, 10.3 Wm^{-2}, and 2.6 Wm^{-2}, respectively (Table 5)"

P30362-L13. Some of the input-forcings in MPI-ET and PU-ET are also part of the inputs-forcings used to derive the reanalysis and Zhang, 2010 fluxes fused here, why should we trust more the trend and variability derived in this study? The following sentence (L16) is more convincing (turned? You meant tuned?).

We agree that this sentence is not correct since multiple reanalysis and remote sensing datasets do have similar inputs. We have removed this sentence. In the following sentence, "turned" has been changed to "tuned". The accuracy assessment shows that the fused result lowered the RMSE_CV than those using individual datasets (Table 5). Therefore, we would argue that the trend and variability derived from our study could better approximate the actual trend and variability of the surface energy budget over the Tibetan Plateau than using individual reanalysis or remote sensing dataset alone.

P30362-L23. More consistent seasonal cycles, but compared with what?

Compared with ground observations, latent heat and the ground heat fluxes are more consistent than that of the sensible heat flux from reanalysis and remote sensing datasets. The sentence has been revised to:" the comparison with the ground observations of the monthly cycle averaged over all available stations showed more consistent seasonal cycles of latent heat and the ground heat fluxes than that of the sensible heat flux from reanalysis and remote sensing datasets (Fig. 3)."

P30362-L25. Better to take a consistent approach for figures and text: or we use seasons (as in the figures), or months (as in the text).

It has been changed by using seasons throughout the manuscript.

P30363-L15. True for the CMA stations, but also true for all other product fluxes for all years and pixels in the TB where there are no in situ data.

This point is well taken. The challenges of validating the retrieved sensible and latent heat fluxes both at stations and at pixels over the Tibetan Plateau are the extreme-sparse in situ measurements. Therefore, this sentence highlighted that we compare the fused fluxes to the Yang11 datasets over CMA stations rather than validate it by using Yang11 datasets.

P30364-L13. As commented above, if the fused datasets were not the winners compared with in situ data, I'd be really worried about the methodology. Accuracy may not be the best word here, the smallest differences with the in situ data?

We have modified this sentence for clarification. The accuracy is based on the cross validation result that is presented in the previous section. The sentence has been changed to: "overall, the fused sensible and latent heat flux anomalies (averaged over CMA stations) are consistent with that from Yang11 in addition to the highest accuracy according to the leave-one-site-out cross-validation (Section 4.1), which were used in the spatiotemporal analysis of the sensible and latent heat fluxes (sections 4.4 and 4.5)."

P30365-L3. I may be missing something, but I cannot find in the paper the source of the meteorological data used for the correlations and analyses. The CMA stations (but then this specific analysis is not over the whole TB)?

We have revised this sentence to avoid the confusion. This sentence is based on published literature. The data source of the simulated sensible heat flux, the observed wind speed, and the estimated downward shortwave irradiation from CMA stations over the TP can be found in Tang et al., (2011) and Yang et al. (2011).

P30365-L16. I see clearly the 2004 peaks, but the 1990, 1993 are not that clear, are those the right years?

It has been changed to major peak (drops) of sensible heat flux was consistent with that from water vapor in 2004 (1990 and 1993).

P30366-L18. Table 6, should it be Table 7?

The reviewer is correct. It has been changed to Table 7.

P30366-L19. What figure are the anomalies plotted?

The Figure number has been added in this sentence.

P30366-L24. The point here is that this is an artefact in the data due to the inclusion (or exclusion for the previous years) of GMS data, rather than a real trend/transition?

The point is that the inclusion of GMS data is a possibly factor of the transition of the fused latent heat flux in the late 1990s. Therefore, more evidence is needed to verify whether a real transition exists as well as the magnitude of the trend.

P30367-L15. This paragraph may be better as part of the discussions.

This paragraph has been moved to the discussion section as suggested.

P30369-L6. What is it meant by "accuracies constraints"?

It referred to the uncertainty from inputs and methods that estimate the sensible and latent heat flux from stations, reanalysis, and remote sensing datasets. It has been replaced by "uncertainties".

P30369-L25. Referred to? I do not follow this sentence.

This sentence states that the latent heat flux is regulated by not only the surface energy balance but also the surface water balance, which could be related to a better accuracy than that of the sensible heat flux as shown in the validation result. It has been changed to: "The validation result also proves that accuracy of the latent heat flux is significantly higher than the sensible heat flux from reanalysis or remote sensing, which is possibly attributed to the estimation of the *latent heat flux is constrained from not only the energy but also water balance over the TP (Mueller et al., 2013).”*

P30370-L3. No idea about what the authors try to say in this sentence.

This sentence has been removed. In section 3.2, we have added the assumption that the latent heat flux and the net radiation from the previous study (Shi and Liang 2013a) are more reliable than that of the sensible heat flux from multiple reanalysis and remote sensing. Therefore, the proposed method estimate the sensible heat flux form the energy balance that achieve lower RMSE_CV than that by applying the MLR to fuse the sensible heat flux directly (Table 5).

P30370-L9. What is the MLR close Bowen ratio? This needs to be explained or suppressed from the text.

The MLR close Bowen ratio stands for a fused approach where MLR is applied to generate the fused sensible and latent heat fluxes, and the energy balance was closured by adjusting fused sensible and latent heat fluxes assuming Bowen ratio not change. We have removed the result of the MLR close Bowen ratio, and added a comparison with the fused sensible heat flux by applying the MLR directly by using the ground measurement, reanalysis, and remote sensing sensible heat flux datasets (Table 5). The detail can be found in our response to the general comment 3.

P30370-L15. I am not sure this makes any sense. The land surface modelling is part of this specific fused approach (reanalyses and remote sensing products are used), so their “large uncertainties” (as described in the text) should propagate into the fusing methodology. Why this should result in “fewer uncertainties”?

We agree with the reviewer that this sentence is not correct that the uncertainties of reanalysis and remote sensing products as (inputs) could propagate into the fused data. The proposed data *fusion approach improved RMSE_CV than each calibrated input dataset (that uses land surface modeling) because of the advantage of constraining the uncertainty by integrating ground measurement, reanalysis, and remote sensing products. We have reworded this sentence to:" Nevertheless, the data fused approach is an applicable method for this regional application because of constraining the uncertainty by integrating ground measurement, reanalysis, and remote sensing products, which lowered the RMSE_CV than that by using each calibrated input dataset (Table 5)."*

P30370-L25. I may be missing something, but I do not understand this "higher heterogeneity" of the LE, compared with the H (spatial I guess). Do you mean that Hs are spatially less heterogeneous than LEs? For a given area and instant, due to the SEB we always have LE+H=available radiation (we assume G=0 for simplication). So if LE changes X, H should also change X, in absolute terms the same amount. So same spatial integrations of those variations, should not be giving the same integrated variations?

At the monthly scale, the available radiation may display heterogeneity from the varied surface net radiation, especially over area with complex terrain (assuming the ground heat flux is negligible), which could lead to different heterogeneity of the sensible and latent heat flux. For instantaneous measurement, previous research suggests that there is a potential that the different footprint might exist for the surface sensible and latent heat fluxes (Kustas et al., 2006). In addition, the estimated instantaneous net radiation and ground heat flux also showed heterogeneity over the Tibetan Plateau (Ma et al., 2006). Since our study is not capable of comparing the heterogeneity between the sensible and latent heat fluxes but is an effort to quantify the variation of the RMSE_CV from different resample scales, we have deleted "which *supports the expectation of higher heterogeneity of latent heat flux and the potential for different footprint to exist for latent and sensible heat fluxes".*

P30371-L5. This section seems more part of the methodology than of the discussions.

We agree with the reviewer. We have moved most parts of this section (descriptions of various data fusion method) to Section 3 Methodology. The remainder of this section has been moved to Section 4.1.

P30372-L18. I do not think that we need a reference to backup the fact that ground observations cannot alone characterise a continuous spatial scale.

We agree with the reviewer. The reference has been removed.

Table 3. Mentioning that the validation is a comparison with in situ data.

We have added the validation is a comparison with in situ data in the caption of Table 3.

Table 5. Adding something about the different Fusion schemes in the caption, these are referred to in the text a few pages after the Table 5 is presented first, so the reader may get confused.

We have added different fusion schemes in the caption of Table 5.

Table 6 and Table 7. It will help to indicate that these are the LE and H for the fused product.

We have added fused sensible and latent heat fluxes in the captions of Table 6 and Table 7.

Figure 1. Black circle? You mean red large and small circles (instead on red and black)?

It has been changed to red large and small circles from red and black circles.

Figure 2. It would have been very interesting to see also a "real" Fused H scatterplot (the fused H here is in reality coming from closure with a radiation product). I will not be surprised if the correlation of a "real" Fused H would be much larger than 0.61.

We have added a scatterplot of the fused sensible heat flux by applying the data fusion to multiple reanalysis and remote sensing datasets directly in Figure 2. We have added the *description in section 4.1:“The R^2 of the fused sensible heat flux was significantly improved (0.67), which was higher than that of the applying MLR to multiple sensible heat flux products directly.”*

Figures 6 to 9. Perhaps for clarity adding in the caption fused latent flux, fused sensible flux, etc.

We have added fused sensible heat flux, fused latent heat flux in the captions of Figures.

References

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001-2013, 10.5194/bg-6-2001-2009, 2009.

Kustas, W. P., Anderson, M. C., French, A. N., and Vickers, D.: Using a remote sensing field experiment to investigate flux-footprint relations and flux sampling distributions for tower and aircraft-based observations, Adv. Water Resour., 29, 355-368, 10.1016/j.advwatres.2005.05.003, 2006.

Ma, Y., Zhong, L., Su, Z., Ishikawa, H., Menenti, M., and Koike, T.: Determination of regional distributions and seasonal variations of land surface heat fluxes from Landsat-7 Enhanced Thematic Mapper data over the central Tibetan Plateau area, J. Geophys. Res., 111, D10305, 10.1029/2005jd006742, 2006.

Miralles, D. G., van den Berg, M. J., Gash, J. H., Parinussa, R. M., de Jeu, R. A. M., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., Dorigo, W. A., Teuling, A. J., and Johannes Dolman, A.: El Niño–La Niña cycle and recent trends in continental evaporation, Nat. Clim. Change, 4, 122-126, 10.1038/nclimate2068, 2014.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K. C., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707-3720, 10.5194/hess-17-3707-2013, 2013.

Tang, W. J., Yang, K., Qin, J., Cheng, C. C. K., and He, J.: Solar radiation trend across China in recent decades: a revisit with quality-controlled data, Atmos. Chem. Phys., 11, 393-406, 10.5194/acp-11-393-2011, 2011.

Shi, Q., and Liang, S.: Characterizing the surface radiation budget over the Tibetan Plateau with ground-measured, reanalysis, and remote sensing data sets: 1. Methodology, J. Geophys. Res.-Atmos., 118, 9642-9657, 10.1002/jgrd.50720, 2013.

Vinukollu, R. K., Meynadier, R., Sheffield, J., and Wood, E. F.: Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends, Hydrol. Process., 25, 3993-4010, 10.1002/hyp.8393, 2011.

Wang, K., and Liang, S.: An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture, J. Hydrometeorol., 9, 712-727, 10.1175/2007jhm911.1, 2008.

Yang, K., Ye, B., Zhou, D., Wu, B., Foken, T., Qin, J., and Zhou, Z.: Response of hydrological cycle to recent climate changes in the Tibetan Plateau, Climatic Change, 109, 517-534, 10.1007/s10584-011-0099-4, 2011.

Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Water Resour. Res., 46, W09522, 10.1029/2009wr008800, 2010.

Zhu, X., Liu, Y., and Wu, G.: An assessment of summer sensible heat flux on the Tibetan Plateau from eight data sets, Sci. China Ser. D., 55, 779-786, 10.1007/s11430-012-4379-2, 2012.