Response to Reviewers' Comments

Referees' responses are in black; authors' responses are in blue.

Response to Anonymous Referee #1

Summary: The authors present an application of SEBAL to semi-arid riparian regions in the US West. The study design is appropriate to the problem and the presentation of methods and results is clear throughout. I am not a SEBAL user, but my impression is that this paper will present a useful case study and a number of useful application tips for users of that system. Since SEBAL is a widely used model and semi-arid riparian zones are of particular interest for ET analysis I believe that publication of this paper in HESS is justified.
Answer: Thank you.

That said, I admit that I find the paper disappointing in some respects, as indicated in my major comments below. My overall impression is that the paper as it is currently written provides a technical resource for SEBAL users, but that the authors have missed the opportunity to put their analysis into the broader context of semi-arid riparian hydrology, or at least the broader context of available ET methods and datasets in these regions.

Major comments:

1. I find it odd that so many ET methods and products are listed in the introduction, yet the paper only presents internal analyses. It would be very interesting to know how the calibrated SEBAL estimates from this study compare to available ET estimates from MODIS (MOD16 and/or the UW product for CONUS), ALEXI, NLDAS, or other available datasets. Indeed, I had really hoped that in reading the paper I would learn whether calibrated SEBAL performs substantially better than other methods or publicly available products in these regions.
Answer: We fully agree that it would be interesting to know how the calibrated SEBAL estimates compare to those by operational ET products such as MOD16, ALEXI, NLDAS or other ones. However, the goal of this paper is "to conduct a thorough evaluation of the performance of SEBAL in arid riparian areas." Other international research groups have evaluated SEBAL for irrigated areas, forested catchments, cropped soil and desert surfaces. There are many scientists that believe SEBAL is designed for irrigation systems, hence it is an excellent idea to synthesize the multi-year research work in semi-arid riparian regions. The latter is an ecosystem requiring more global attention in water accounting frameworks of river basins. Comparison between SEBAL and METRIC at the one hand with other remote sensing based ET models at the other hand has been done before (TSEB, DISALEXI, IWMI Turkey experiment, Caren Jarmain South Africa). References to that could be provided
Co-author Hendrickx is involved with a statewide assessment of ET in New Mexico and evaluation of the MOD16, ALEXI and SSEBop operational ET products. The study started last fall and publication of its results will take a while to complete.

As it stands I find it difficult to interpret the reported error estimates, since I have no reference for what constitutes a good or bad estimate of turbulent heat fluxes for these locations.
Answer: Whether an ET estimate is "good or bad" depends on one's needs, and in particular on the given space and time scales. A farming operation requiring ET data for Variable Rate Irrigation Application needs localized data and for every single day. The basin agency working on water transfers or groundwater exploration plans, requires the regional ET data to be available at monthly or seasonal time scale. One can probably define also different type of users of water balances in riparian corridors. The best solution is to present the SEBAL accuracy of semi-arid riparian regions at a range of temporal scales, and then the user can understand the range of plausible standards and decide to use or reject it. Co-author Bastiaanssen just published a paper that gives a reference frame for interpretation of ET error estimates. [Karimi, P., and W. G. M. Bastiaanssen (2015). Spatial evapotranspiration, rainfall and land use data in water accounting – Part 1: Review of the accuracy of the remote sensing data, Hydrol. Earth Syst. Sci., 19, 507-532.] This reference has been added to the manuscript.

2. The title and introduction indicate that this paper is motivated by the problem of estimating ET in semi-arid riparian areas. The selection of study sites is consistent with this goal, and in some sections the presentation of results touches on matters relevant to riparian areas. But overall the very lengthy results section and the conclusions have very little to say about riparian zones. Instead various details of SEBAL calibration and bias correction are explored without any explanation as to how or why the results are specific to / informative of / generalizable across semi-arid riparian zones. Instead the paper becomes a list of specific lessons learned and recommendations for SEBAL, some of which are semi-arid specific but others of which seem not to be. I would urge the authors to present a more compelling synthesis of their results as they inform study of semi-arid riparian zones. Alternatively, if the results are more generalizable then the authors could consider removing the semi-arid focus and reframing the paper in terms of its technical contribution to SEBAL applications.
Answer: We share the ambivalent feeling of this comment: focus on all the little technical aspects of an extreme-condition-inverse calibration for mapping ET in riparian areas or skip those and show with a few figures how well the calibration works in riparian areas and what can be done with those reliable ET data. Further to satellite images and routine weather data, there is no additional information needed to apply models such as SEBAL and METRIC. This makes this method potentially attractive for heterogeneous landscapes and ungauged basins. Many other spatially distributed ET models require additional aerodynamic information and data on the atmospheric constituents during the moment of satellite overpass. It is in our view proper to explain some of these fundamental mechanisms, and provide a logical framework why SEBAL could work where other models fell short.
We definitively have opted for a focus on all technical aspects because the "devil is in the details". We want to take our readers by the hand and explain all the important details needed to arrive at accurate and reliable ET estimates with our method. There are too many articles that ignore those details. Often this may be justified but the disadvantage is that the interpretation of the error estimates becomes impossible. Almost every application of ET algorithms by individual investigators has its own quirks and often is different from other applications using the same method. In our opinion, there is also a need for papers that give all technical details as does our study. Because each application depends on its environment (semi-arid or sub-humid) we opt to keep the semi-arid focus. In addition, the assessments of performance are in the context of

riparian systems. This context should provide insights and "comfort" to users of remote sensing of ET information in riparian areas.

Minor comments:

Introduction p. 13482: The inclusion of NLDAS and LIS in the discussion of satellite derived ET estimates is misleading. NLDAS ET estimates are the product of land surface models that simulate ET prognostically, while LIS is a software framework that supports LSM simulations with data assimilation. Neither is really a satellite-derived ET product in the way that the other listed analyses are. If the review of "ET products" is to include prognostic modeling systems alongside diagnostic energy balance methods then the authors should make a clear distinction between the two.
Answer: We fully agree with your comment. Reviewer 2 made a similar comment. We have removed any reference to NLDAS and LIS in the manuscript.

Introduction p. 13484: The statement that "If SEBAL performs well under these challenging conditions, it is likely to perform well in most arid and semi-arid regions" requires further justification. I understand that short fetch and sub-pixel thermal contrasts make riparian areas difficult, but riparian areas also present a strong ET signal that is absent in most semi-arid regions and that might make RS detection easier.
Answer: We agree with these insights. We have expanded on this statement to suggest that good performance should be expected from other types of moderate to high ET systems that are surrounded by relatively dry land uses. The above quotation has been replaced by "A good SEBAL performance under these challenging conditions would be a strong indication that satisfactory performance should be expected from other types of moderate to high ET systems that are surrounded by relatively dry land uses [*Compaoré et al.*, 2008]".

Response to Anonymous Referee #2

The authors present an application and evaluation of SEBAL for the semi-arid riparian regions in the Rio Grande Basin of New Mexico. This is an interesting repetitive application and evaluation of SEBAL model in the riparian areas.
Answer: Thank you.

Major Comments:

Page 13483-line 1: Unlike NLDAS and LIS, SEBAL and METRIC do not require land cover maps: Please could you discuss how you will estimate the surface roughness without landuse and land cover maps.
Answer: We have inserted in the paper that an alternative for estimating surface roughness without land use and land cover is to use expressions that relate the NDVI to the momentum roughness length. See, for example, Eqs. 30-34 in Allen et al. (2007) or Eq. 24 in Bastiaanssen et al. (1998). Note that the ASCE guidelines use Leaf Area Index as an indication for surface roughness. This solution became widespread, and is now considered also by FAO in their international guidelines of crop ET.

Allen, R. G., M. Tasumi, and R. Trezza (2007), Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - Model, Journal of Irrigation and Drainage Engineering, 133, 380-394.
Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag (1998), A remote sensing surface energy balance algorithm for land (SEBAL): Part 1. Formulation, J. Hydrol., 212-213, 198-212.

I don't see any relation of LIS and NLDAS with this paper. Author added unnecessary NLDAS and LIS in this paper. Methodology, application and science behind NLDAS and LIS is way different than SEBAL and METRIC.
Answer: We fully agree with your comment on NLDAS and LIS. Reviewer 1 made a similar comment. We have removed any reference to NLDAS and LIS in the manuscript.

Section 3.2 closure forcing:
Please could you put reference or criteria for 65%-110% energy balance closure, or maybe refer this paper http://www.sciencedirect.com/science/article/pii/S0168192302001090.
Answer: Thank you for steering us to this excellent reference. We amended this section as follows: "If the sum of *H* and LE, before correction, was less than 65% or greater than 110%of the available energy (*Rn* -*G*), the data were not used in our analysis. Wilson et al. (2002) found the average energy balance closure at FLUXNET sites to be between 53 to 99%. Since their numbers represent average closures and since data points at the lower end of the range raise greater concerns for data quality, we chose to shift the range up."

Section 3.4 page 13495-line 4-5:
There are several types of footprint models. Initially, simple two-dimensional analytical footprint models for neutral atmospheric conditions were developed. Have you developed or used in this study?
Answer: This section was misleading and did not contribute to the paper. Therefore, the following sentences were deleted: "There are several types of footprint models. Initially, simple two-dimensional analytical footprint models for neutral atmospheric conditions were developed (Gash, 1986; Schuepp et al., 1990). Later, the analytical footprint model was improved to account for atmospheric stability conditions (Horst and Weil, 1992; Hsieh et al., 2000)."
We did not develop, but use a footprint model as explained in this sentence: "The footprint flux, $F(x,zs)$ [–], along the upwind direction, x [m], measured at the height zs [m], suggested by (Hsieh et al., 2000) is used in this study."

Therefore, the use the average H and LE values of the 25 pixels surrounding the EC tower pixel is considered to be the best option for the comparison of daily ground measurements and SEBAL estimates. Please could you discuss how you come up with exact 25 pixels surrounding the EC tower?
Answer: Thank you for catching this inaccuracy. We have made a correction: "the 24 pixels surrounding the EC tower pixel".

There are major problems in the paper. 1. Energy balance closure 2. Comparison of Energy balance parameters consistently shows underestimation and overestimation of parameters (Rn, G, H, and LE) what should the reader believe 65% energy balance closure or the remote sensing model (15% error).
Answer: The 15% error is between the closed energy balance terms (corrections made for H and LE) and the remote sensing model. We believe the reviewer misinterpreted our analysis. Energy balance closures – while not desirable and not explainable – are commonly observed (Wilson et al. 2002). On many days the closure error was small (see Figure 1). The commonly agreed upon method to correct energy balance is to force energy balance closure using the Twine method that is used by many researchers. Because of the sometimes rather large energy balance error we changed the original title of our article from "Validation of …" to "Evaluation of …"

Response to Anonymous Referee #3

This paper presents a study of using the SEBAL to estimate daily energy balance components in arid riparian areas based on satellite (Landsat) imagery. Authors evaluated the model performance by comparing the model outputs to the ground-based measurements at three sites in the southwestern U.S. In general, the methodology (including problem setup, modeling, and data analyses) appears to be appropriate and the study could potentially be valuable addition to the hydrologic studies over arid/semi-arid riparian regions.
Answer: Thank you.

Although my overall opinion of the paper is positive, the paper in its present form is not publishable in the HESS and requires substantial revisions. Listed below are my major concerns as well as suggestions for the authors to consider.

1. My major concern is whether the application of the SEBAL approach could be extended to other or more general arid riparian regions. In the Introduction, authors state that the study areas are very "heterogeneous" and if the approach works "under these challenging conditions, it is likely to perform well in most arid and semi-arid regions." In the manuscript, however, I have read many explanations and discussion on why certain bias/errors occurred in this study associated with the three selected areas and what specific techniques were used. One of the examples is that, in Lines 19-23 (Page 13499), "these differences area about two to three times larger than those typically reported for SEBAL. The much larger than usual MRD is attributed to the heterogeneity of the riparian sites…" If the "heterogeneity" is the cause, how could readers believe that the approach would perform well in other heterogeneous riparian areas?
Answer: Thank you for this important feedback because it shows us that we need to reformulate this paragraph. We changed the statement on SEBAL performing well in other regions. The much larger than usual MRD should not be attributed to the heterogeneity of the riparian sites but to the fact that the ground measured net radiation did not capture the true net radiation averaged over a pixel consisting of patches of bare soil and vegetation. The reason is that only one net radiometer was used and it was directed towards the canopy. As a consequence the measured net radiation became biased towards the net radiation of vegetation instead of the true pixel-scale net radiation consisting of bare soil and vegetation. A remedy would have been to install several net radiometers on each site. In addition, it still would be difficult to place these

meters in a way that a representative net radiation would be measured. Instead we argue that under such circumstances the SEBAL observed net radiation is likely a better alternative for energy balance assessment than the use of a biased net radiation measurement on the ground. Earlier tests between net radiation from remote sensing and in-situ measurements of homogeneous land use classes showed an excellent agreement. There are many papers (SEBAL as well as non-SEBAL models) showing that net radiation can be accurately estimated from remote sensing.

2. I am a modeler, but not a SEBAL user. I am curious about the acceptable level of accuracy for SEBAL estimates. According to the results (e.g., Figures 4 and 6), the deviations (from 1:1) are obvious and R2 values are not very impressive, ranging from 0.56 to 0.75 (Table 4) and from 0.32 to 0.78 (Table 8).
Answer: The R2 values do not give information about the accuracy of the SEBAL ET values but only about their correlations with the ground measurements. Therefore, it is better to inspect the values for MRD; for LE those are well within ±10% that is considered a pretty good accuracy.

3. I understand that the selection of "hot" and "cold" pixels is very essential to the success of SEBAL application. But how sensitive is this procedure to the quality of ET estimation? In other words, if a less accurate or a bad selection of "hot" pixel is made, how would it impair the model performance in estimating the ET? It is not clear according to the manuscript and some quantitative analysis in this respect would be interesting.
Answer: You are right. The selection of the cold and hot pixel is the most critical aspect of the extreme-condition-inverse calibration. If done well, it yields seasonal ET values for watersheds with a bias of around ±5%. For cropland seven papers have reported a mean absolute percentage error of 1%. All these papers used an extreme-condition-inverse calibration (see Karimi and Bastiaanssen, 2015). Morton et al (2013) demonstrates that these hot and cold pixel selection can be fully automated on the basis of years of manual experiences (that can be spelled out in terms of mathematical steps). They also demonstrated that the uncertainty was the highest for fields with low ET levels and lowest for fields with high ET levels, with a seasonal mean uncertainty of approximately 5% for all fields. In a blind comparison, automated daily and seasonal ET estimates compared well with flux tower measurement ET data at multiple sites.

Morton, C.G., J.L. Huntington, G.M. Poll, R.A. Allen, K.C. McGwire and S.D. Basset, 2013. Assessing calibration uncertainty and automation for estimating evapotranspiration from agricultural areas using METRIC, J. American Water Resources Association (JAWRA) 49(3): 549-562

4. The manuscript is too long, containing too many details that are not necessary.
Pages 13485-13487: if the descriptions of the method are not new to the SEBAL users, many contents of these pages could be removed. Sections 3.2, 3.3, and 3.5: have too many details and could definitively be shortened.
Answer: Please, see our response to the second major comment of reviewer 1. At the request of the editor we removed all SEBAL method information that is easily found in the literature but we kept study specific information. Unfortunately, there are many critical details that we cannot remove without shortchanging readers who want to implement SEBAL themselves.

5. Locations of the study areas. Please, provide map(s) showing the locations of the three areas. It is very important to include information such as where/how large the study areas are and where the field-based measurements were made.
Answer: We have added a map.

6. The manuscript needs careful editing, e.g.: “has” (Line 9, Page 13482) “Peters Lidard et al., 2004” (Line 14, Page 13482) “exists” (Lines 11-12, Page 13482) “others suggest” (Line 13, Page 13487) “Since,” (Line 20, 13488)
Answer: We have checked our editing.

7. "Table 5” (Line 2, 13491): mentioned/used in the context before Tables 3 and 4. Please re-arrange tables in sequence.
Answer: We have removed the reference to Table 5 in order not to violate the table sequence.

8. Figures 4, 6, 8, 9, 11, 12 and 13: The fonts and symbols of these figures are too small.
Answer: We have increased the size of these figures.

Response to Anonymous Referee #4

The authors present work that compares SEBAL’s estimation of energy budget components (mostly latent heat exchange) for 3 arid riparian areas, (Owens valley, CA, Middle Rio Grande Valley, NM, and San Padro, AZ) with ground based measurements. This work could be used to identify potential issues regarding the calibration and validation of remote sensing energy budget data specific to SEBAL against ground based measurements. However, as currently written, the manuscript appears unfocused due to either unnecessary detail and/or indirect writing. Due to the unfocused nature of the manuscript it is difficult for the reader to interpret the results or the validity of the study.
Answer: See our response to the second major comment of reviewer 1. However, we have inspected and improved our writing.

Moreover, it is hard to discern what the primary objective of the paper is. For example, section 3.3 (Comparison of SEBAL flux predictions to ground measurements) seems to be (at least partly) the heart of the paper, but is under the methods section and is currently written as part methods and part discussion that appears to repeatedly fault ground based measurements for being inadequate to verify SEBAL results. However, from the title and introduction, I thought the purpose of the paper was to evaluate the performance of SEBAL in arid riparian areas, and not ‘address the issues of comparing satellite based energy budget data to ground based data’. Furthermore, at L4-5 on page 13496 the aim is restated to ‘evaluated the challenges of SEBAL flux perditions in arid riparian areas using a validation approach’, which is more in line with sections with section 3.3’s discussion. However, the approach discussed in section 3.3 leads me to believe that ground based measurements may not be an appropriate method to validate SEBAL and that comparisons to other satellite based methods are warranted given the scale differences between ground based measurements and satellite observations. However, the decision of whether or not to include further comparisons hinges on what the specific purpose of the paper is, which needs to be better defined by the authors.

Answer: We have replaced and restated portions of the manuscript to hopefully make the descriptions and discussions more clear and consistent.
The overall goal of this study is to conduct a thorough evaluation of the performance of SEBAL in arid riparian in New Mexico, Arizona and California as stated in the introduction. Our original intent was a straightforward comparison between ground measurements and satellite based energy budget data. Comparison with other remote sensing methods is not an option since there are no other remote sensing ET algorithms available that have the high spatial resolution and accuracy of the SEBAL and METRIC approach (see Karimi and Bastiaanssen, 2015). However, it turned out that such a simple straightforward method would not be adequate because of issues with the energy balance closure and the representativeness of the net radiation measurements. We do respect your opinion and valuable comments but also strongly believe that it is high time to inform the hydrologic community about the challenges of using energy balance measurements for validation of remote sensing ET algorithm even if this leads to complex papers.

Major Comments:

Page 13481 L13-L14: It is actually not clear to me that the work presented here provides evidences that SEBAL yields reliable estimates for actual evapotranspiration rates in riparian areas of the southwestern United States. Primarily, because of the issues presented in section 3.3 as well as the results section that painstakingly point out the issues with ground based measurements, which are the only set of validation data used in this study.
Answer: Issues with ground measurements and remote sensing observations should be painstakingly pointed out. We conclude use a combination of both to come up with the best possible energy balance and LE. We have implemented what we understand to be the more accepted means of data analysis and error correction for both ground-based measurement and RS-based measurements. Comparison of ground and RS-based measurements is consistent with accepted practice.

Page 13487 L4-L6: Why is it safe to assume that soil moisture is constant? Especially for arid environments were ET usually is a large part of the water budget? What are the implications of this assumption?
Answer: We refer to a daily soil moisture change that typically will not cause a large effect on total daily ET. We have changed the text from “Where soil moisture does not significantly change” to “Where daily soil moisture does not significantly change”.

Page 13487 L17-L18: Support for the assumption that G24 = 0 should also be stated here rather than later at Page L13492, L26-L27. Regardless this to me seems to be a rather large assumption that has consequences as the land surface and soil column in these environments will experiences seasonal (larger then daily) scale warming and cooling.
Answer: We write “close to zero”. This is a common assumption for soils that has been backed up by measurements. Of course, although we can ignore the daily changes of soil heat flux for the daily energy balance, we cannot ignore the sum of the daily changes totaled over a season. In this study we deal with the daily energy balance.

Page 13491 L3-L8: Why do you exclude data were the sum of the H an LE is 65% less then or 110% greater than of the available energy? This criteria seems to eliminate much of the available data. Is this an indication of poor observational conditions, such as days that are not clear or have variable weather?
Answer: A riparian area with its high variability in roughness length, mesoscale temperature heterogeneity, short fetch distances, and maybe heat storage in the woody parts of trees and bushes is arguably one of the most challenging locations for EC measurements. The poor energy balance closure is therefore not a surprise. Since satellite images are only useable under clear sky conditions, the meteorological conditions should generally be consistent and variable weather would not be expected to impact the analysis. The poor closures may be caused by large low frequency eddies and horizontal heat advection that are both triggered by the mesoscale temperature variability between the desert and the riparian area.
We amended the text as follows: "Wilson et al. (2002) found the average energy balance closure at FLUXNET sites to be between 53 to 99%. Since these numbers represent average closures and since data points at the lower end of the range raise greater concerns for data quality, we chose to shift the range up.

Section 3.4 Footprint model: There are important concepts that are partly presented here regarding the problems of validating SEBAL using ground based measurements, which I believe contributes to a lot of the issues of calibrating and validating SEBAL. Mostly reconciling scale issues between satellite observations and point scale(ish) measurements. These issues are cryptically mentioned in sections 3.3.2 and 3.3.3. For example the difference between the heat plate scale of 0.001 m2 and 900m2 landsat pixel. Unfortunately, the scale of measurements and the scale of the SEBAL pixel is never clearly or coherently presented, rather some information about the scale is peppered throughout the paper.
Answer: The scale of SEBAL is largely tied to the 30 to 60 to 120 m scale of Landsat, due to the strong impact of thermal signals on the energy balance. We have attempted to employ this scale in the manuscript.

Minor Comments:
Page 13483 L13-L22: These two sentences have unnecessary detail that can be summed up as field measurements are slow and costly, in contrast satellite measurements are fast. Please be more direct.
Answer: We have shortened this section.

Page 13483 L20: Change '…86000 ha of the office…' to "…86000 ha from the office…"
Answer: Done.

Page 13483 L22: The phrase 'expert months' is not clear.
Answer: This phrase has been eliminated.

Page 13484 L6-L8: "Another difference with previous studies is our focus on all components of the energy balance during the instant of satellite overpass …" Is this the only difference? Did the other validation studies not focus on arid riparian areas? Also, did those validation studies have the same problems with ground based measurements discussed in section 3.3?

Answer: Actually, we don't know of any other studies in arid riparian lands that offer the amount of detail included in this study. Many studies conduct a simple comparison between eddy covariance ground measurements and remotely sensed fluxes without the scrutiny of this study. Not taking into account the common well-known problems with ground based measurements will lead to less accurate remote sensing ET algorithm.

Page 13485 Equation 2 and 3: Is H instantaneous or daily? It appears the notation is not consistent.
Answer: H is used for both the instantaneous and daily sensible heat flux but in the caption of Tables and Figures we inform the reader whether H, G, Rn or LE is instantaneous or daily. Using subscripts would have deteriorated the readability of the transcript.

Page 13487 L17: Why is Cef set to 1.0, which would then have no effect on equations 5 and 6?
Answer: We have now included Cef as a variable in the equation for 24-hour ET. This is consistent with the testing of Cef = 1.0 and 1.1 later in the paper.

Page 13493 L19 – Page 13494 L10: This paragraph is an example of indirect writing. I assume the point of the paragraph is the last sentence, "Therefore, in this study rather then trying . . ." In scientific writing, the point should be stated up front and supporting details follow the main point.
Answer: We have restructured this section.

Page 13500 L19: 'Incoming short and longwave radiation' and for that matter outgoing short and longwave radiation. These terms can be separated from Rn in your equations and in many energy balance equations can be calculated/measured separately. Perhaps the terms should be presented as separate components of the energy balance equation (equation 1).
Answer: Good suggestion. We have done so.

Page 13503 L22: The phrase 'traditional SEBAL' is awkward.
Answer: We agree and change to "original SEBAL".