Dr. Domenech,

Thank you for your thoughtful review. Please see our point-by-point response (in italics) below.

——————— General comments: ———————

The manuscript describes the validation activities performed to assess the accuracy of the CERES angular distribution models (ADMs) described in Su et al., 2010. This paper is of paramount importance to the scientific community in order to understand the uncertainties and limitations of the CERES flux new product Ed4SSF. This manuscript follows the studies developed in Loeb et al., 2003, 2006, 2007 and explores new approaches such as assessment of parallax effects or scene misidentification errors.

In my opinion, the manuscript is interesting and well written. This paper, once published, should be cited in all new papers using CERES derived fluxes, and justifies the publication of the previous Su et al., 2010 paper; which shows the new proposed methodology for the CERES ADM construction.

In general, I would recommend to publish the manuscript. However, I would like to see the authors response to some critical points. In particular, I am not very convinced about the methodology selected to assess the scene ID errors.

Please find below my comments.

Section 2.1

In my opinion, the methodology behind the direct integration is not sufficiently explained. I know that the method has already been discussed in Loeb et al. 2007 and presented in Loeb et al., 2003 (for the non sun-synchronous TRMM satellite), but I still think that a brief mathematical description is necessary.

The general idea is clear, but the details of the method are confusing. For instance, it is not explicitly stated in the text that the ADM predicted radiance for every angular bin within the study region is obtained using the DI-derived flux and the CERES ADM (anisotropic factor) corresponding to the observed scene. More details would be welcome.

We rewrote section 2.1 and included three equations (Eq. 1-3). Equations 2 and 3 describe how the two sets of direct integration ADMs are constructed which will help readers better understand the methodology used.

According to Fig. 1, it seems that the results show a positive TOA SW flux bias over the ITCZ in January and a negative bias over the ITCZ in July (not so clear though). It would be nice if the authors could discuss it in the text.

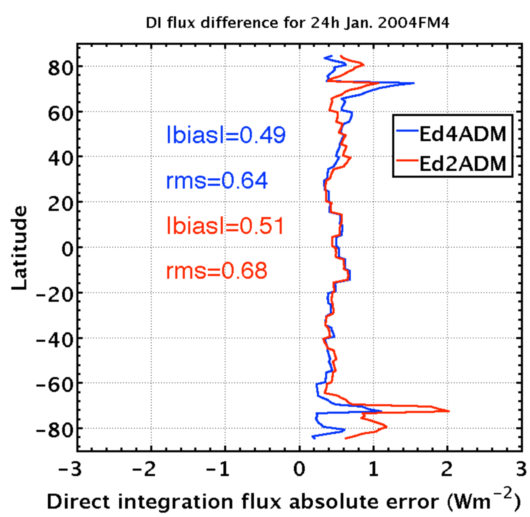
This is probably associated with the viewing geometry of the CERES instrument. The CERES instrument is on a sun-synchronous orbit which introduces a strong correlation between latitude and solar zenith angles and relative azimuth angles.

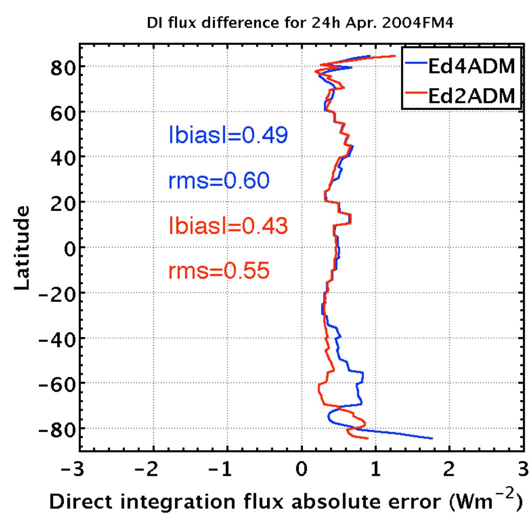
Section 2.2

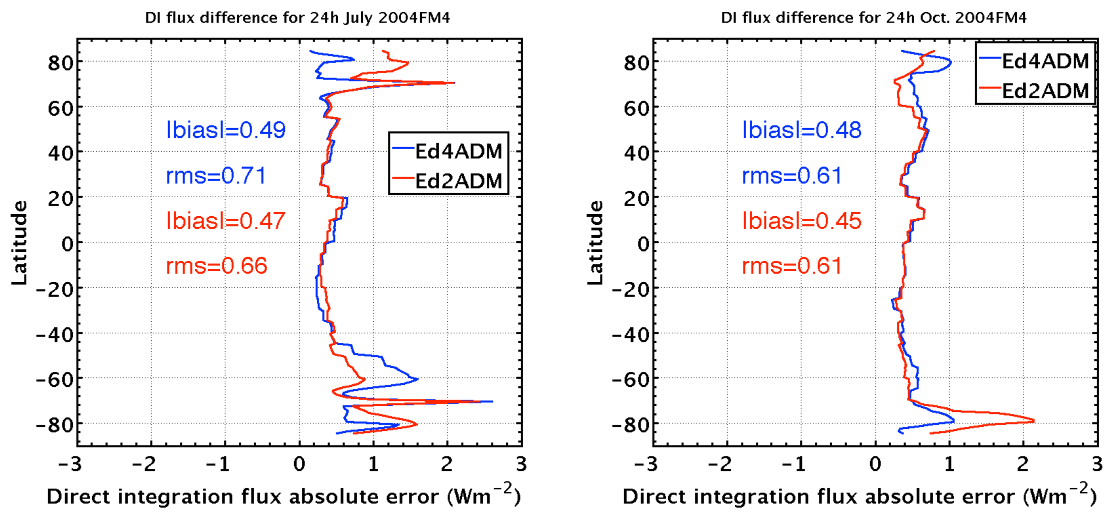
Both RMS and bias flux errors obtained with the Ed.4 LW ADMs are higher than the errors obtained using the previous Ed. 3. That's weird since the Ed. 4 ADM is an improved version of Ed. 3.
How do the authors justify this error increase? In particular considering that the LW flux consistency check reports a reduction of 2-3 Wm-2 over the errors obtain with Ed.3 (Loeb et al., 2007). These test quantify the viewing geometry dependence of the ADMs. Thus, could it be due to a worse scene ID (scene misidentification) for some new identified regions?

The RMS errors determined from the direct integration for Ed4 and Ed3 are essentially the same if we round the RMS error to a tenth of a Wm^{-2}. We agree that the biases associated with the Ed4 are slightly higher than those associated with Ed3, but this is mostly due to compensating errors in Ed3. As shown in the figures below that the mean absolute biases between these two editions are quite similar (to a tenth of a Wm^{-2}). The most noticeable differences are over 50S-70S, where the mean absolute biases using the new ADMs are higher for April and July. The new method used to construct LW ADMs over cloudy snow/ice scenes takes the cloud emissivity into account via cloud optical depth (see section 5.4 in Su et al (2015)). This could mean that cloud optical depth retrieval over sea ice under large solar zenith angles (>60°) is less reliable, but further study is needed to answer this question. Some discussion is added in Section 2.2.

The instantaneous LW flux consistency errors were derived using along track CERES observations and collocated MODIS data. The method and data used for the consistency test are very different from those used for direct integration. We stated in Section 3 and it was also pointed out by the reviewer that the consistency test only quantifies the viewing geometry dependent error of ADMs, it is possible that scene identification error caused the regional LW flux error to increase slightly, even though the anisotropic characterization is improved.







Page 4495, Line 21: I do not understand the meaning of that sentence. ???

"Table 3 summarizes the global monthly mean TOA WN flux biases and RMS errors for the CERES instrument on Terra in 2002 and for CERES on Aqua in 2004."

Section 3
Why do not the authors use the same approach of section 4 to determine the ADM error in the study with MODIS measurements? You could employ Eq. 7 to deduct the error derived from the nb2bb conversion from the total error.

We actually thought about subtracting the narrowband-to-broadband error from the total error for the MODIS-CERES test. But as we only use those regions that have a RMS error less than 3% in SW narrowband-to-broadband conversion in the SW analysis, and only use those regions that have a RMS error less than 0.5% in LW narrowband-to-broadband conversion in the LW analysis, the narrowband-to-broadband conversion errors are very small (generally about 1% for different cloud types for the SW, and even smaller for the LW). Thus, subtracting the narrowband-to-broadband conversion errors from the total errors won't change the results. We described the narrowband-to-broadband conversion error used for MODIS on page 4479 from lines 5-11 in the original submitted version.

Section 3.1
I was just wondering, which average flux values do you use to obtain the "absolute" RMS errors?

I used the oblique-view fluxes to calculate the "absolute" RMS errors, but the mean oblique-view and nadir-view fluxes are close enough that using either one does not change the "absolute" RMS errors much.

Page 4498, Line 8: Since the authors are employing CERES AT data for your consistency check they could use CALIPSO data to discriminate thin clouds from aerosols with high AOD.

That would be an interesting check, but unfortunately CERES on Aqua did not take any along track observations after 2005. We do not have concurrent CERES and CALIPSO data to carry out this analysis.

Section 3.4
I appreciate the effort of the authors to provide an estimate of the CERES flux uncertainties.
That information is highly demanded by the scientific community. However, I have doubts in the approach described in this section.

You assume that your methodology to obtain the "scaling factor" is independent on the employed ADM. But results of the consistency check obtained from the idealized ADM (which I assume is obtained from the ratio between RT radiances and corresponding fluxes) is likely giving pretty different relative RMS errors than the Ed.4 ADM.
In addition, how do the authors compute the theoretical fluxes from the 3D RT models? How do they define the simulated scenes? 3D models usually provide a TOA flux (e.g. 100 km) obtained from limb-to-limb radiance observations; which is different from a ADM-derived flux. ADM-derived fluxes are dependent on the observed scene within the simulation domain, however modeled fluxes are obtained for the whole simulated domain. It would be nice if you could describe better the methodology.

The "scaling factor" was derived in Loeb et al. (2003, 2007). In this study we did not redo the calculation but simply applied the ratio derived by Loeb et al. (2003, 2007). We rewrote section 3.4 to make the fact clear and added some more details regarding how the ratio is derived. Loeb et al. (2007) provided the relative RMS errors and TOA flux error calculated using five idealized ADMs (Table 8). The relative RMS errors ranged from 6.9% to 18.7% (excluding the 26.5% RMS error derived using Lambertian ADMs for liquid clouds), which are similar to the RMS errors that we got in this study. Although the relative RMS errors changed by a factor of three, the ratio between TOA flux error and RMS error remained quite constant (from 0.54 to 0.65). Kato et al. (2006) used 3D radiative transfer (SHDOM) calculations to simulate the process of retrieving cloud optical depth and effective radius from radiances measured by satellite. Figures 5 and 8 in their paper implicitly indicated that the flux consistency is larger than the flux error. However, a definitive ratio was not provided. To avoid confusion, we have decided not to cite the 3D radiative transfer calculations done by Kato et al (2006). SHDOM was run in 1-D mode for flux calculation in Kato et al (2006).

Section 4
In order to check the dependence of the ADMs on the instrument measurements, it would be nice to compare the ADM errors obtain in the MSIR consistency check against the results obtained from MODIS.

We stated that the SW flux uncertainties based upon the MISR consistency test are consistent with the SW flux uncertainties from the MODIS-CERES consistency test at the end of Section 4.

Since the CERES-MISR-MODIS data is available, the authors could reproduce the study of section 3 using only the MODIS data corresponding to the cases studied here.

For the CERES-MODIS consistency test we used 137 days of CERES along track observations, and for the MISR consistency test we used 107 days of the CERES along track observations because MISR data was not available for some of the CERES along track observations. We could redo the CERES-MODIS consistency test using the same data as the MISR consistency test, but we don't think this will change the results.

During the TOA SW flux consistency study, which footprint size do you use to create the CERES-like MISR database? CERES 20x20 nadir spatial resolution?

The SSFM consists of MISR measurements within the collocated CERES footprint. For nadir footprints the resolution is indeed 20x20km, for off-nadir footprint the resolution will increase.

Section 4.2
I would like to point out that the RLRA is not always appropriate for collocation of multiangular broadband radiances. Semi-transparent cirrus or broken clouds are difficult to handle with the RLRA method, and usually the most significant radiatively layer of clouds is not at the top of the cloud.

We agree that the RLRA is not ideal for collocating broadband radiances under certain cloud conditions. It is, as the reviewer pointed out, possible to see radiances originating from points above or below the RLRA. However, we are simply using it as a means to estimate the parallax effect in the SSFM dataset. Because we applied strict criteria on the footprints we select we believe using RLRA is appropriate in this situation.

Has been addressed the parallax due to clouds in the optical path of oblique observations collocated at the RLRA determined for a nadir cloud?

Yes. We addressed this in the third paragraph of section 4.2. When using RLRA to re-project MISR radiances, some pixels from the oblique angles are obscured if clouds in the surrounding pixels are higher. These pixels are flagged as missing in the convolution process, reducing the percentage coverage. To get a comparable estimate of the error using both MISR Level 1 and Level 2 data, we use a subset of the Level 1 data by requiring that for each CERES footprint at least five of the same MISR cameras have valid radiances for both Level 1 and Level 2 data, and both Level 1 and Level 2 data have greater than 99.9% coverage. This matched Level 1 and Level 2 dataset is used to estimate the parallax effect. This matching criteria bias the footprints to homogenous scenes, therefore the parallax effect reported here should be considered as the lower *bound of the parallax effect. By selecting scenes that have a sufficiently high MISR coverage for all of the cameras then we remove the problem of obscured radiance values.*

Yes. In our analysis we only consider footprints in the SSFM that have greater than 99% coverage of at all of the 9 MISR angles. When MISR sets its RLRA, if any cameras are blocked from seeing that pixel then no radiance value is provided for that camera. By selecting scenes that have a sufficiently high MISR coverage for all of the cameras then we remove the problem of obscured radiance values.

Section 5
I think the use of C3M-enhanced data to improve the scene ID will result in the introduction of biases in the retrieval of ADM-derived fluxes unless the flux inversion is performed for ADMs constructed with that scene ID definition.
The C3M-enhanced product uses radar and lidar A-train measurements to "improve" the cloud detection. If you define your scene ID based on this new retrieval you should "re-train" your ADM models to be consistent with the new scene definition.

The Ed. 4 models rely on the cloud detection provided by MODIS (Minnis et al., 2010). Clouds that are, by definition, not captured by the CERES-MODIS cloud-mask are part of the intra-scene variability inherent to the clear-sky ADMs. Even if you are able to detect them with an improved scene ID the cloudy ADM cannot provide a better estimation of the flux because no ADM has been produced to model the angular behavior of such clouds. The same can be applied to the rest of cloud retrievals that determine the anisotropic factor selection of cloudy scenes.
Thus, I would not say that biases in the cloud retrievals determination of MODIS, i.e. limitations of imager-based cloud retrievals, result in misidentification errors in CERES derived fluxes. Those limitations are part of the CERES approach. I consider misidentification errors scenes that are occasionally not well determined, but MODIS, in general, identifies them correctly. Those are the scene ID errors that affect the selection of anisotropic factors for flux conversion.
I my opinion the authors comment in page 4512, line 5 the most relevant study to perform on this topic, that is to assess the error caused by misclassification of scenes during the ADM construction.

We agree with the reviewer that the ideal way to assess the scene identification error would be to use the same lidar-radar enhanced cloud detection to construct the ADMs and to use the same cloud detection to select the ADMs. However, it is not possible to construct ADMs with the lidar-radar enhanced cloud detection, as CALIPSO and CloudSat do not overlap with the CERES RAP measurements and they only provide near-nadir observations.

We also agree with the review that cloud retrievals from MODIS observations in general are correct. This is also supported by the validation study of the CERES cloud working group, see presentation by Pat Minnis (http://ceres.larc.nasa.gov/documents/STM/2014-10/TUESDAY/CERES/Clouds.CERES.STM.10.14.pdf). When we construct the ADMs we use the averaged radiances (sorted into angular bins) for most of the scene types, so if *the misidentifications happen occasionally and have small or no effect on the mean radiance for each angular bin and scene type, thus it will have small or no effect on the derived ADMs. The goal here is to evaluate the flux uncertainty associated with the scenes that are difficult for the passive sensor to identify by assuming that these occasional misidentifications do not have (or have very small) effect on the anisotropy characterization.*

——————- Technical corrections: ——————-

Page 4503, Line 8: Table number is missing.

We added the Table number.

Figures 6-9: These plots are difficult to read. The idea of the "thin" and "thick" bars is terrible, please try to use another way to represent them.

We redid the figures 6-9 to show each surface type separately. For example, we split figure 6 into 3 figures (one for ocean surface, one for land surface, and one for snow/ice surface). We also used "hatched" bars for multi-layer clouds to distinguish them from single-layer clouds. Doing so, we think the legibility is much improved.