**Referee: 1**  
  
Comments to the Author  
  
The main idea is interesting and has a good potential for practical application. But I still have some questions.  
My main concern is that there is a large difference between the spatial resolution of DMC and MODIS reflectance products yet the uncertainty caused by the difference was not addressed. Why not use a finer spatial resolution, perhaps Landsat data as reference given that one scene of Landsat 7/8 ETM+/OLI covers 185km by 170km?  
In page 14 line 25-27 the author claims that despite the big spatial resolution gap between aerial images and the MODIS MCD43A4 product, the final mosaic do not need to further reduce seam line (feathering) and the overlapping areas can be chosen from any of the overlapping images, from my experience, this is highly doubtable.  
  
The following are some detailed comments and questions.  
Page 2  
Line 3: I am not a native English speaker, but may be the title should be "by calibrating with"  
Line 17: “It is shown that ...” what shows? Please rephrase the sentence.  
Line 24: maybe “lease squares regression”? Please clarify.  
Line 25: the acronym DMC should be in parenthesis, please check the journal style.  
Line 27: BRDF used before defined in line 38.  
Line 35-39: the sentence is not clear to me, please rephrase this sentence.  
Line 55: VHR, maybe need to be defined before use in main text, please check the journal style.  
  
Page 4  
Line 23-28: as I understand, aerial image such as DMC usually took in a very small yaw and roll angle, why there are large view angles? Or perhaps large field of view（FOV）? Why solar varies? This sentence is confusing, please clarify?  
  
Page 7  
Line 19: the acronym DN should be in parenthesis, please check the journal style.  
Line 30: Equation (2) should be  , where d is the distance between the sun and the earth in astronomical units.  
  
Line 34:   should be described as “TOA reflectance” rather than “reflectance”, since the reflectance and TOA reflectance are quite different.  
  
Page9  
Line 36: The size of sliding window should be specifically defined. The BRDF and RSR are both related to the type of the targets. It is quite important to define the window size to make sure that the pixels in the window represent the same targets.  
  
Figure 9：Since the effects of BRDF and RSR are both related to the type of the targets, the linear relationships between the reflectance of the DMC and MODIS should be different for different target type. The results should be showed independently for each sampling type.  
  
  
Page 14  
In my opinion, section 2.4 should be placed as 3.1, or 2.1, please check the journal manuscript guidance.  
  
Page 15  
Line 23: since “except NIR”, then not good in all bands. Please phrase？  
  
Page 26  
Line 57: “and mosaic normalization techniques to reduce seam lines” sentence is not complete.  
  
Page 27  
Line 34-57: you point out that the varying size of the sliding window should be investigated and a higher spatial resolution reference such as Landsat OLI, yet in my opinion, they need to be discussed in the manuscript, or at least the varying size of the sliding windows should be discussed if there were no concurrent Landsat data available.

**Referee 2**

This manuscript is aiming to perform the radiometric normalization of aerial images by collocated and concurrent, well-calibrated satellite images. The content shows no novelty but is somehow useful for those experiments without radiometric calibration of aerial camera. However, the description is too poor to be accepted for the publication. Detailed comments are as follow:

1. Section 2.1 and 2.2, in my opinion, is almost useless. However, this part occupied 6 pages.

For example, equation 5 is a standard relationship between DN and reflectance for any optical camera, which means equation 1-4 is unnecessary. In the algorithm, the authors assume that effect caused by RSR difference is linear. They didn’t take this effect into account. If so, it is unnecessary to give detailed description here.

2. In homogenization procedure, if we perform step (2) and (3) directly at course resolution without step (1), what’s the difference, please clarify.

3. The purpose of this manuscript is to increase the radiometric accuracy of the aerial images. However, in whole manuscript, the authors talked too less about the uncertainty. What is the accuracy of the MODIS MCD43A4? What is accuracy of the algorithm? The accuracy impacted mostly by atmospheric correction, geometry, and RSR difference. But I cannot find any explanations.

4. The validation by SPOT 5 is meaningless, which gave only the comparison between MODIS and SPOT 5. I suggest the authors conduct the comparison between with and without BRDF correction.

5. The most important thing is the description of aerial experiment and images, as well as surround conditions. In this manuscript, however, I almost can find nothing. Where is the study area? What are the specifications of DMC? When did the images take? How many days? What is the AOD in those days? Is it possible that the impact from no atmospheric correction is much larger than BRDF correction?

**To Do**

* Respond to reviewers comments (reviewer2 to be taken with a pinch of salt).
* Investigate if we can use Landsat as a reference, if not, we should explain why we are limited to coarse resolution references. How is it calibrated? Can scan line be fixed? Note that MODIS bands may be more similar to DMC, Landsat data revisit time is cloud-free concurrency issue, Landsat scanline off prevents use as reference, Landsat surface reflectance is not BRDF corrected (?)
* NB: Check the SPOT validation – that SPOT image seems to be in a different projection and when I hack it, it does not match well to Landsat / NGI
* We could produce results with different window sizes, then compare eg overall error with SPOT5 image
* How do we better justify no seam line smoothing? Rev1 has a kind of pt – just because the MODIS pixels are continuous / smooth, doesn’t mean the calibrated aerial pixels will be, as they are only being corrected on average / with interpolated gains
* Come up with a generic measure of homogeneity and use this to quantify mosaic before / after
* AOD on days of aerial survey?
* Show an image of M / C
* Can 2.1 and 2.2 to summarised and moved to an appendix? Perhaps also the stuff about RSR and viewing angle?
* What about a difference in overlapping pixels before and after calibration? Hmm hmmm? This could be both an accuracy / consistency measure and a justification for no seam line removal. We just need to be careful about how much overlap we have after calibration…. Ref Gehrke 2016, Gehrke 2010. Possible exclude shadow variations. This could be combined with a de-emphasis of the whole absolute reflectance thing. It is only a relative before/after comparison and has the same problems as SPOT comparison but does serve as a new measure of “homogeneity” that can be linked to surface reflectance. We could also potentially create a mosaic / spatial distribution of these errors. I think we need to look more carefully at how each mosaic method evaluated their accuracy. This would also be a good way of comparing effects of window sizes, offset params etc. To make this error nicer, we should downsample the image as in Gehrke 2016
* Overall homogenisation factor of Lopez et al 2011
* Choose some key images, apply normal theoretical atmospheric and brdf corrections, then compare to our method. Perhaps also do the overlap consistency check as above !!!
* To save space, rather than store everything at 0.5,0.5m, it can be stored at 10m. Eg each ngi image is accumulated into a mosaic as it is generated. Then it is deleted
* We could further justify offset only model by fitting a global linear model between MODIS and 500m NGI mosaic.
* The sliding window does not have to be square – a circle would make better sense for larger win sizes
* Another way of evaluating could be histogram comparison / distance eg between 500m NGI and MODIS

Summary of evaluations

* Chandelier et al 2009: it is only a mosaicking technique not necessarily intended for RS. There is no quantitative evaluation, only visual
* Downey et al 2010: Only qualitative visual evaluation
* Collings et al 2011: They evaluated on placed known targets. They also compare this error to an error from a theoretical model and show it is mildly increased. Exact details unk. Non-target accuracy is unknown.
* Lopez et al 2011: “A mesh was created in the overlapping area with randomly located points within each cell, providing a total number of 500 comparison points. The differences in these comparison points were evaluated in terms of standard deviation (STD) and expressed as a factor of homogenization in percentage as:” std\_orig – std\_correct / std\_orig . So note that this is for a comparison of methods not an absolute accuracy measure. They also check with actual field reflectances and also with the radiometric control points i.e. overlapping points.
* Gehrke 2016: not 100% on this but they seem to do a comparison to absolute reflectance at specific points and they also plot a kind of before and after contrast and brightness factor along a certain line.

**Summary To Do**

* Look into reviewer1 seamline comment – it has some validity. I think it is sufficient to explain what happens at borders of rasters and how discarding pixels improves this. But that in some circumstances there will be visible seamlines.
* Justify non-use of Landsat (check out validation for fun) for reviewer1 – there is only a Landsat 7 image with clouds and scanline defect. In general, MODIS revisit time is beneficial for use as reference.
* Simulate effect of sliding window size (and m and c paramaters?) Issues here are data size (~500GB per data set) and how to evaluate the effects of changing the size. We could compare against SPOT for each window. But a homogeneity measure would also be helpful.
* Relook at similar paper for ideas on seamlines and validation.
* Relook at my seamline notes.