Response:

(1) General Comments from the Reviewers

1) The two reviewers of this paper expected something quite different than they found in this experiment. Reviewer #1 objects to the term "Observing Systems Simulation Experiment" because it does not fit the accepted definition of an OSSE which derives from the use of the term by Arnold and Dey (BAMS, 67, 687-695, 1986). It is unfortunate that we have no term (other than a "Proving Ground" or "testbed") by which to describe a process where synthetic data is fed to a planned satellite algorithm to test it but does not involve an assimilation step into a numerical model. Reviewer #1 is correct that is not an assimilation experiment and to that end we are removing that terminology. Readers are probably aware that assimilating aerosol optical depth (AOD) into numerical aerosol forecast models is still rare after over a decade of real AOD data availability. It was certainly beyond the scope of this project to begin to assimilate GOES-R Advanced Baseline Imager data into an aerosol model. In fact, such an OSSE was done in 2009 in Europe (Timmermans et al., JAOT, 26, 2673-2682, 2009). Their conclusion was that the OSSE suffered from the same imprecision in the known science between AOD and PM2.5 as discussed in Hoff and Christopher(2009). They did not reach a quantitative comparison in that study because of this. Since the term, OSSE, is a diversion from the point of the paper, we will revert to simply using the term "Proving Ground" as suggested by reviewer #1.

   The second expectation seems to be that the output of the ABI algorithm should have been compared on a case by case basis with ground data generated by another instrument such as MODIS or ground monitors. Figure 2 shows exactly that comparison, comparing apples to apples (Reviewer #1 correctly pointed out some confusion in this Figure). Using MODIS radiances modified for the viewing geometry for the GOES-R ABI, those radiances in the ABI spectral channels were put through the ABI processor and the results show regression values of about 0.8-0.9. It would be moot to compare the additional hourly imagery through the day on a regression basis from the WRF-CMAQ generated data to a single MODIS image since the aerosol moves and the regression will obviously degrade. We have no other imager to get these radiances throughout the day, which is why this process was carried out in the first place. Quantitative comparison to surface monitors has been studied in detail over the last decade by our group (Engel-Cox et al., *Atmos. Environ.* **38**, 16, 2495-2509 doi:10.1016/j.atmosenv.2004.01.039., 2004; Engel-Cox et al., *J.A.W.M.A.*, **55**, 1389-1397, 2005; Engel-Cox et al., *Atmos. Environ.* **40**, 8056-8067, 2006; Weber et al., *J. Air & Waste Manage. Assoc.* **60**, 574-585, 2008; Kondragunta et al., *J. Appl. Met*., **47**, 425–442, 2008; Zhang et al. Atmos. Chem. Phys., 11, 11977-11991, doi:10.5194/acpd-11-11977-2011., 2011; Hoff and Christopher, *J. Air & Waste Manage. Assoc.* **59**,645–675, DOI:10.3155/1047-

3289.59.6.645, 2009) and regressions of AOD to PM2.5 at the surface is not a "calibration" for the ABI processor.

Comparison with another geostationary satellite such as SEVIRI was suggested and there is an explanation (pg 3 line 23 to pg 4 line 7) why this would not have been realistic. SEVIRI has no 2.1 μm band so the retrieval of a "clear sky" surface reflectance at that wavelength is not relevant. For this reason, SEVIRI has a fundamentally different retrieval concept than the MODIS/VIIRS/ABI referencing to this 'aerosol free' (actually not quite) channel. Additionally, the "users" of this Proving Ground experiment were air quality forecasters in the United States and Canada. They don't know SEVERI nor will they ever use it for their daily work. ABI will be a tool that they will use and this Proving Ground was a training step for those who had used MODIS and now VIIRS extensively in forecasting, but had never seen AOD imagery on an hourly basis during the day. The suggestion that the "real-time" aspect of this experiment was not relevant really seems to miss the point of the paper. The point was to have these users see a simulacrum of the data that they would see ON THAT DAY they are forecasting from a GOES-R ABI output. The ability to compare the real ABI output from the imager to SEVIRI's current performance will come after launch, but since the bands are so different, forcing the ABI to compare on a different continent with SEVIRI radiances was not this experiment. In fact, the algorithm team did use SEVIRI data in comparison with the ABI codes, but that was not our work.

A more pointed criticism is that this paper is better suited to another journal and not Atmospheric Measurement Techniques. I would have thought that decision would have been made by the editor on accepting the paper for discussion. We went through a round of review with the Editor prior to the discussion paper being made available and it was accepted at that point. If the Editor chooses to change his mind based on this review comment, we will go elsewhere.

(2) Specific Comments from the Reviewers

Reviewer #1 (AMTD-7-C3493-2014):

The manuscript lacks a scientific focus or problem and merely attempts to describe the process to produce simulate ABI imagery used in the retrieval of air quality products from simulated GOES-R ABI imagery.

We disagree. We think that training a user community who will in fact use GOES-R ABI AOD results in their daily work is a scientific problem and we describe how that was done.

Feedback provided by end users viewing the simulated products over a web page was cosmetic (in terms of changes in the display) and did not address the utility of the products.

The (now) three workshops in which feedback was obtained from the user group was in

fact referenced in the article by Huff et al. (2012) but perhaps the reference journal is a bit obscure. The feedback conclusions have been broadened here and a web reference is given to the reports from the user group (pg 14, lines 4-14).

The reader is expecting an evaluation of the utility of the GOES-R ABI baseline air quality products as viewed from simulated or proxy imagery. They do not get this. No comparison is made between products derived from simulated imagery and ground observed, not to a similar product derived from MODIS or SEVIRI. Authors need to present an end user evaluation of the baseline products from several case studies.

As mentioned in the general comments above, we do not believe that a comparison with non-optical measures such as PM2.5 at the surface is a realistic test of the performance of the ABI algorithm. Sticking with optical intercomparisons, Figure 2 does discuss in a quantitative manner the performance of the GOES-R ABI algorithm fed with MODIS radiances (comment also made below by this reviewer). Alternatively, one could evaluate AOD from ground radiometers such as AERONET against the synthetic GOES-R ABI AOD predictions. The results were not good as WRF-CMAQ underestimated the aerosol mass for some of these major AOD events during this period. This is not really a test, however, of the ABI algorithm as much as it is a test of WRF-CMAQ's ability to predict aerosol mass and sizes. This is now explained further in the text.

The authors call this approach an OSSE, but merely simulate radiances (for subsequent product retrievals) from prescribed conditions. OSSEs use a data assimilation approach to evaluate the impact of a new observation on weather prediction and the impact is compared to a nature run generated by a separate system that pushes the state of the art capabilities. The authors are performing more of a "proving ground" activity preparing potential end users for the new data products. Removing references to this terminology throughout the paper is recommended.

Agreed.

The research and assessment described in the manuscript seems to be cast in a near real-time environment needlessly constraining the approach. There is no need for this constraint given that the evaluation of the utility was seemingly not done in real time (there was no description of this real-time evaluation process). Therefore higher resolution simulations which better mimic ABI characteristics could have been done, reducing the impact of course resolution imagery on continuous cloud fields depicted by the model, and enhancing the conclusions about the utility of the products.

The evaluation process is described in Huff et al. (2012) and was provided by reference. The funding and rationale for this work was based on assessment of the real-time usage of the ABI products (as is done for the other Proving Grounds listed). Suggesting that we use another period in an *ex-post facto* analysis would never have happened in this project. Suggesting that we use higher resolution model runs is a supposition that those would provide better simulacra of reality. The practicality is that the experimental requirements did not allow for such higher resolution runs. It is now a moot point as the instrument will

launch and a real evaluation of the ABI's capabilities will occur before such a different experiment could be carried out.

The manuscript also suffers from poor organization and often mixes grammatical tenses, describers and shows examples of several approaches to generate a simulated or proxy products when one is ultimately used, includes an irrelevant discussion of SBDART, and describes the generation of natural color RGB imagery from simulated ABI imagery when it is not even shown.

Examples of the egregious grammar could be called out. We don't see them. The rest of the sentence however says that SBDART is irrelevant as is the natural color RGB imagery. If the reviewer had read Christopher's 2014 paper (which evolved from this project, as well), he or she would get the relevance. That was an alternate method of gaining the radiances (tried and published). The RGB capability that the ABI does not have was also discussed in the Christopher paper. We removed a sentence about what SBDART is.

It would be helpful if the authors provided more information on the utility of AOD from SEVIRI and demonstrated how using the additional 2.25 micrometer channel information improves AOD retrievals over land and how the time continuity of geostationary view adds significant value over the twice daily product from polar orbit. Dust is included in the input data but not included from the simulation – The impact of this needs to be discussed.

We believe that the discussion quoted above deals adequately with SEVIRI. We did not set out to assess AOD from SEVIRI which views the Europe and Africa. The utility of the 2.13 μm channel in the ratio-type algorithm has been widely discussed in the literature. The utility of the continuity of the view of AOD was in fact discussed on lines 16-18 of page 3. While dust was included in the simulation, it had no discernable affect on AOD for this study period. The dust/smoke mask was not discussed in detail here as it was judged to be not "ready" for operational use.

3.1 MODIS proxy imagery â˘A ´c This paragraph and its relation to figure 2 is unclear. The description of Fig 2 in the text is different than in figure caption causing confusion. The figure needs more description. Maybe separate land from water statistics. Clarify times of MODIS data used (obviously more than one swath used to get coverage, but caption says 1645UTC). â˘A ´c Unclear the value of this section when it is not used for evaluation.

This discussion has been clarified. The description of the figure was inadequate. A sentence describing panels (c) and (d) had been dropped from the caption. This method was indeed used for part of the evaluation of the ABI AOD product.

3.2 Modeling approach. In this research, computer time and latency to generate the simulated imager is not an issue. When GOES-R ABI is operational, the need to simulate the imagery is not present, and therefore documenting the time to produce this product is irrelevant. The simulated imagery should be generated at the spatial resolution closest to that of ABI, regardless of how much computer time it takes in this per-operational study, or a convincing justification needs to be presented (which has not been). Availability of near real time simulated products does not relevant to this study.

We don't understand this comment at all. The time to do a simulation would not be an issue if you don't care about the need to present this data in real-time (which this reviewer obviously doesn't but he did not pay for the project). When the GOES-R ABI is operational, a different evaluation which has nothing to do with models such as CRTM and WRF-CMAQ will be carried out. We disagree with this comment entirely.

Tables 2-4 only briefly discussed. They do not add value to the manuscript. Figures 4 and 5 only briefly discussed. The authors should enhance the scale on the nitrate and dust images to show detail, or exclude them because of low values and show only the other four parameters.

We disagree. Prior to submission the editor asked for clarification of these Tables. Table 2 explains where this Proving Ground falls in the tests of the GOES-R products, Tables 3 and 4 explain the configuration of WRF-CMAQ as run.

Were we to change scale on the figures in the panels, it would provide a false image of how dust and nitrate do NOT contribute significantly to AOD.

Reviewer #2 Comments (AMTD-7-C3791-2014):

Thank you for an interesting and useful project to prepare the operational forecasting community for the new data sources that will become available with the next generation of GOES geostationary satellites. This work is very important for maximizing exploitation of the new capabilities as soon as possible after launch.

We were happy to hear this.

I recommend this paper be withdrawn and submitted to a journal more appropriate to describe the operational proving ground nature of this experiment. The emphasis of the paper is not on hypothesis testing, but rather on the generation of the synthetic products and their evaluation in the proving ground program. While I sympathize with the authors that validation of the synthetic products is challenging with existing data, the absence of this validation makes this work unsuited for Atmospheric Measurement Techniques which will provide visibility in that community, such as the NWA Journal of Operational Meteorology: http://www.nwas.org/jom/. The performance of these future sensor products is of great scientific interest, but the scientific community requires a more rigorous analysis that the methodology employed here cannot provide.

Ah, the Marcus Antonius part of the review..... "but they are honorable men"....

We will leave the discussion of whether this paper is appropriate to the Editor who accepted it for discussion. If he agrees with this comment, we will go elsewhere.