AMTD 8, 4049-4093, 2015

Item-by-item reply to Referee #3

Referee: yellow text
Authors: plain text

Kalman Filter physical retrieval of surface emissivity and temperature from SEVIRI infrared channels: a validation and inter-comparison study

This article presents a Kalman filter approach for retrieval of surface temperature and emissivity from SEVIRI and validation with in situ and satellite observations. The article is presented well and there is sufficient detail to describe the results, however there are a number of shortcomings which will have to be addressed if this is to be considered for publication in the future.
We thank the referee for the in-depth review.

Major Comments:

My major concern is that the validation is incomplete and unbalanced. For land, only two in situ sites are used for surface temperature (of which Evora in my opinion is associated with too large of an uncertainty to be used for validation) with no intercomparison with any other heritage products. Considering the SEVIRI temperature/emissivity products are mostly used for land surface studies, I'm not sure what benefit the comparisons with ECMWF/MODIS/AVHRR over water have other than to possibly demonstrate stability of the retrieval. When so few in situ sites are available for temperature validation over land, other approaches should be sought out, for example the Radiance-based validation method which is currently used extensively to validate the NASA MODIS and VIIRS LST products, for example. There are a number of sites over Africa/Arabian Peninsula where field emissivity data are currently available to enable the R-based method - for example from the KIT and JPL thermal infrared groups.
There exist only very few in situ measurements that a) are suitable for validating spatially coarse satellite LST products and b) lie in the FOV of MSG/SEVIRI. Most notable in this respect are the KIT stations in Africa and Europe, which represent a broad range of surface and climatic conditions. In-situ LST from Evora have been used to validate the absolute accuracy of operational LST products (e.g. Ermida et al., 2014) and the site offers more favorable conditions than the SURFRAD sites (USA), which were not designed to validate LST. A comparison with heritage products would allow a relative assessment of product accuracy but would be less conclusive than the validation results presented in our paper. So-called 'radiance-based validation' completely relies on correct information on land surface emissivity and atmospheric state: therefore, it is not a 'validation method' in a classical sense but a method for product evaluation (http://lst.nilu.no/Portals/73/Docs/Reports/UL-NILU-ESA-LST-LVP-Issue1-Rev0-1604212.pdf). We also like to highlight an important aspect of our KF approach: the KF parameters, i.e., T_s-ε state vector, are simultaneously retrieved with a physically-based approach, that is by inverting the radiative transfer equation, therefore Observed radiances are is in perfect agreement with the calculated radiances through the Forward Model (see e.g. Fig. 1 in this reply).

Why was AVHRR used over ocean to demonstrate the seasonal cycle and not ECMWF? AVHRR simply introduces more uncertainty since you are dealing with a bulk temperature estimate, not skin. Along the same lines, I'm not sure it makes sense to compare with ECMWF skin temperature when the atmospheric state vector in the KF approach is already driven by ECMWF fields? The author needs to justify that there is no correlation between skin temperature and atmospheric fields in the ECMWF assimilation approach.
Bulk–Skin surface temperature difference is below 0.3 K (e.g., Schluessel et al JGR, doi: 10.1029/JC095iC08p13341). Our results for the ocean give a *comparison* with AVHRR, not a validation: the aim of this comparison is to check the consistency of the retrieval. The difference of 0.3 K cannot mask the amplitude (several K degrees) and phase of the seasonal cycle, which we wanted to inter-compare. Also, the referee should consider that the surface temperature is not assimilated within the ECMWF model. This parameter is based on the OSTIA operational model (Donlon et al, Remote Sensing of Environment 116 (2012) 140–158), which uses satellite data

alone.

Past studies (e.g. Guillevic et al. and Ermida et al.) have showed large uncertainties associated with the Evora site due to spatial heterogeneity and shading, and this paper does not do enough to convince the reader that results from this site should be valid for a primary validation study - in fact the Guillevic and Ermida studies only further expose the difficulties of using Evora as a validation site for temperature, particularly during daytime.

The authors agree that Evora site is challenging and that is also what makes it so interesting: the landscape surrounding the station is fairly constant over several SEVIRI pixels; the endmembers within a single scene may present high temperature differences. It should, however, be taken into account that the site is equipped with a number of radiometers that allow the characterization of such surface heterogeneity – a feature that can hardly be found in other ground stations.

As indicated in the main text, Ermida et al. (2014) presented a methodology to fully explore the in situ observations leading to ground LST estimates at pixel scale with an uncertainty of about 0.6ºC. Furthermore, Ermida et al (2014) showed that surface heterogeneity at Evora presents a strong daily (lower at night-time) and seasonal variability (very low in winter). The results presented in the manuscript suggest that the performance of the KF (and LSA SAF product) is very stable, and independent of time of day or year. This is a strong argument to support the quality of the three datasets (ground and two SEVIRI LST estimates), obtained through completely independent processes. The difficulty with the Evora site has been fairly presented and discussed in the paper. In contrast, Gobabeb station is also used to demonstrate the performance of the products over a much more homogeneous terrain.

Why was the Dahra LSA-SAF site not used for temperature validation?

Dahra data were not available at the time of the analysis.

For emissivity there is some discussion of emissivity comparison at Gobabeb and some qualitative discussion on full disk map, but this does not constitute a validation. Emissivity spectra should be compared with either in situ emissivity measurements or from lab measurements of field samples. At the minimum, at least an intercomparison should be made with other physical-based emissivity products such as MOD11B1 v4.1 or the ASTER Global Emissivity Database available from the LPDAAC. There are meth- ods already demonstrated to adjust the emissivity from either the MODIS or ASTER bands to any other sensor’s spectral response (e.g. Goettsche and Hulley 2012).

A comparison with the UW-BFEMIS data base has been provided at the global scale in the revised version of the paper. The authors stress that this exercise aims at demonstrating the feasibility and stability of the KF method applied to the whole SEVIRI disk. The comparison of estimated emissivity values, as compared with MODIS and ASTER, and lab measurements for Gobabeb is mentioned in the main text. A thorough validation of land surface emissivity over the disk will be the subject of a future work. In this case, the high resolution global emissivity from ASTER will be used as one of the main references.

Diurnal emissivity variation is highlighted at the Gobabeb site (Fig. 11) using 3-hr moving average, but at the Evora site a daily average is used, possibly smoothing out any diurnal variations. In order to convince the reader that the diurnal variations seen at Gobabeb are in fact due to a physical phenomenon, the emissivity variations should be smoothed over the same time window at both sites. In fact if these kind of diurnal variations are seen over all surface types then the close correlation with temperature variation indicates it could be some kind of retrieval artifact instead. Without in situ measurements, implying the variations observed are due to dew or vapor adsorption is purely conjecture.

A 3h moving average for Evora has now been provided in the revised version of the paper. Regarding the diurnal emissivity variations and water vapor adsorption as a possible cause, we were already careful and stated that *it is consistent…* We agree that further evidence is needed to rule out a possible retrieval artifact. However, for the presented Gobabeb data it is unlikely that the surface temperature goes below the dew point temperature for condensation to occur, i.e. condensation can be ruled out. It should be also stressed that in case there is a poor emissivity contrast, and dry conditions occur (e.g., the case of urban areas in winter time) the phenomenon of emissivity daily variation is not observed at all (see e.g. Fig. 2 in this report). We have not yet conclusive results on these day-night emissivity variations.

It is not evident that the approach used by the author to extend the UW/BFEMIS emissivity

so SEVIRI was implemented correctly. You cannot simply interpolate the UW/BFEMIS emissivity to the central bandwidth of another broadband sensor. The High Spectral Resolution (HSR) algorithm has to be first applied to generate a high spec resolution version and then this should be convolved to the sensors (e.g. SEVIRI) spectral response.
Of course, this is exactly what we did. We have clarified and added proper references, e.g., Masiello et al 2013b, 2014

Minor Comments:
Abstract: Not sure about AMT but typically acronyms are spelt out in the Introduction otherwise they can be distracting in the abstract.
We will leave this decision to the editor.

P4052, L2: ’Here’ is confusing. Do you mean here as in the current study or here as in the previous study.
We mean in the previous study, we have clarified.

P4052, L10-15: For readers seeing this for the first time at least a basic description of the principles used in the KF to separate temperature/emissivity should be given.
Since we are trying to focus on the validation and inter-comparison aspect, we decided to simply state in the introduction that we use a ‘physically based simultaneous retrieval approach’. However, interested readers are then referred to our previous AMT publication describing the KF physical retrieval method in detail (doi:10.5194/amt-6-3613-2013).

2. Data: At least some basic background on SEVIRI should be given. 2.1.1: ’..the whole year round’ should be ’an entire year’
We have expanded section 2. The section now begins with a description of SEVIRI data used in the analysis. Changed to *an entire year*

P4054, L24: Was a linear or quadratic approach used for time interpolation P4056, L4: Express pixel numbers as a percentage instead.
Linear in time. We have clarified the matter in the revised version.

P4056, L7: I’m not sure ’scheme’ is the right choice of word, should it not be algorithm or approach?
Rephrased in KF approach

Section 3. Again there is no brief description (even a sentence or two) to describe the basic physics principle used in the KF approach to separate temperature and emissivity. It is difficult for the reader to attempt to understand the approach by looking at equations alone. It is only at the end of this section that the link to an Optimal estimation approach is stated.
As said, we use a physical approach, which perform a simultaneously mathematical inversion of the radiative transfer equation. We have clarified the matter beginning of section 3 and refer to our previous AMT publication (doi:10.5194/amt-6-3613-2013).

P4060, L25: ’slightly superior’. Given the large uncertainties associated with the in situ measurement made at this site, I don’t think you can claim any one product is ’superior’ of another, unless there is a significantly large and systematic bias.
Fair comment. We have rephrased.

P4061, L22: At least describe what this difference is between green and dry grass emissivity, and a possible physical explanation. E.g. French et al. 2000 (Discrimination of Senescent Vegetation Using Thermal Emissivity Contrast).
We have expanded and better described this part. Reference to French has been quoted.

P4062, L14-15. I think you meant the opposite here, former performs slightly worse not better.
Yes, the referee is right we have adjusted.

P4064, L2: How is direct adsorption justified? It could also be from surface dew.
In arid, very warm regions, dew formation is unlikely because surface temperature rarely goes below dew point temperature. However, we have no firm evidence and this is why we chose the wording *it is consistent..*

P4064, L17: No description of which MODIS temperature product was used, at least the product name and a description is required here.
MODIS data have been described in the revised version. The MODIS product used in the present study is MOD28 Sea Surface Temperature 5-Minute L2 Swath 1 km.

Section 4.4: This is really just a qualitative justification for the emissivity product. An objective intercomparison with other heritage products is needed, and would be very interesting (e.g. ASTER GED, MOD11B1)
A comparison with UW-BFEMIS maps will be now included for the year 2007.

P4068, L1-25: This is reading more like a discussion, not conclusion, since this is the first time processing speed of the algorithm is discussed. Furthermore this sounds more like a justification for including this approach in real-time LSA SAF processing and is not necessarily relevant and interesting to the reader.
As we said in section 3, *..the KF approach has been implemented in such a way to provide a suitable prototype for use at a data satellite data processing centre..* Giving the computing time is completing the technical information about the approach, and for many readers this will be interesting in its own right.

Figures:
Figure 2. The word descriptions within the images are near impossible to read.
The images are of high quality and all text can be read by zooming in.

Figures 2 and 3. Google Earth reference is needed for both figures since using Google Earth imagery is copyrighted.
We agree - references are now provided,

Figure 4 and 8. x/y labels and (a) and (b) fonts are too small. Figure 5. No description in caption of the location of these results, or the meaning of SD. Figure6. Very difficult to see or learn anything from the delta T time serious shown in the bottom figure. Either smaller range or moving average should be shown.
SD is standard deviation; the acronym is given in the text. AMT is an online journal; all text and details of the plots can be seen by zooming in..

Figure 17. Why is emissivity from 1400-2600 shown if only spectral responses in the 800-1200 range are shown? Figure 20: Low emissivites (greenish values below 0.94) at 12 micron over eastern Africa and parts of Madagascar seem to be a reason for concern. The 12 micron emissivity values should very seldom go below 0.95, unless the surface consists of maffic rock types (e.g. basalt), or possibly dry grass.
The comparison with UW-BFEMIS, based on MODIS, (now added in the paper) show a good agreement also in that region.

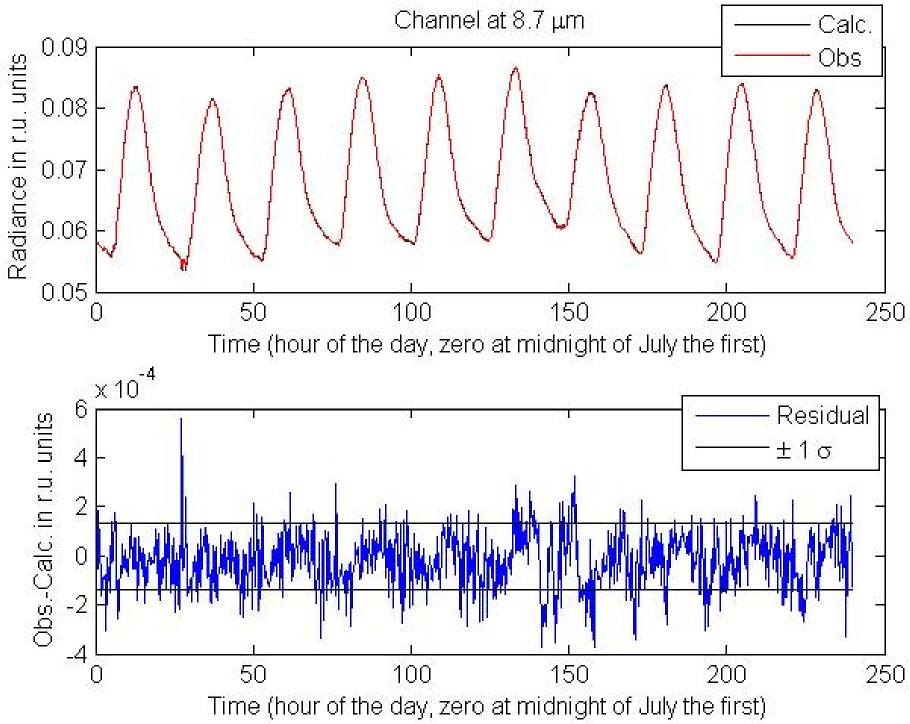

Fig. 1 (not shown in the paper). Typical example of Calculated (Calc.) and Observed (Obs.) SEVIRI radiance and related spectral residual (Obs.-Calc.) for the channel at 8.7 μm). The example refers to ten consecutive, clear sky, days of July 2010 for a pixel over the Sahara desert. Similar results are obtained for the other window channels at 12 μm and 10.8 μm (e.g. Serio et al 2013, doi: 10.13140/RG.2.1.2017.7442). The error confidence interval shown in figure refers to the SEVIRI NEDN. 1 r.u.=1 W m^{-2} sr^{-1} $(cm^{-1})^{-1}$.

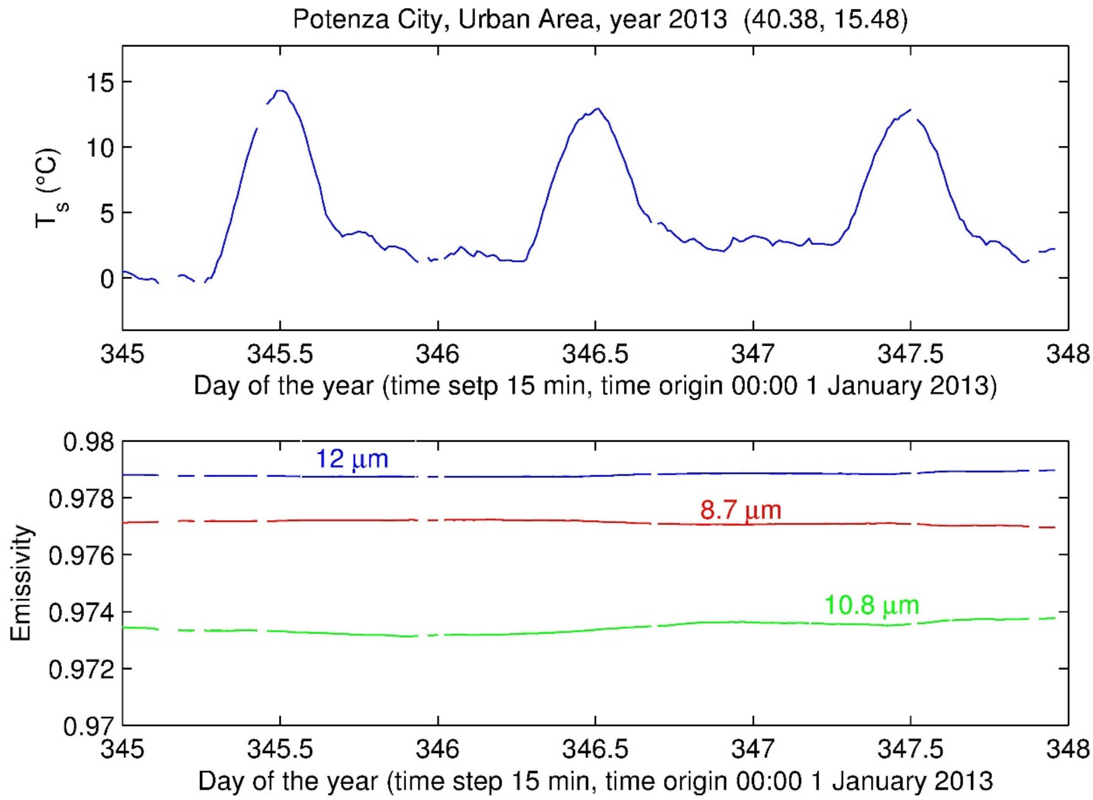

Fig. 2 (not shown in the paper). Surface temperature and emissivity time sequence for an urban area for three clear sky consecutive days (emissivity has been smoothed with a moving average filter with a window of 3 hours).