

Interactive comment on “Kalman filter physical retrieval of surface emissivity and temperature from SEVIRI infrared channels: a validation and inter-comparison study” by G. Masiello et al.

Anonymous Referee #1

Received and published: 12 May 2015

GENERAL COMMENTS

The manuscript is a follow-up paper to Masiello et al (2013), who presented a Kalman Filter approach for retrieving land surface temperature and emissivity from SEVIRI data. In the current paper the authors now focus on validating their KF-based retrieval against in situ observations at two stations and perform inter-comparisons with other data sources. The manuscript in general is well organized, and provides a coherent description of the performed work, although some copy-editing by a native English speaker would be beneficial - I remark on some language issues in the specific comments but was not able to address all of them. Overall the manuscript presents a

C1105

quite interesting new approach for retrieving LST and emissivity from geostationary instruments and the presented validation and inter-comparison exercises are important for showing the capabilities of the proposed approach. The general topic of the manuscript fits well within the scope of AMT and I highly recommend the manuscript to be published once the following issues have been taken into account.

By far the strongest aspect of the paper is the validation of the retrieved LST products against in situ measurements at the sites in Gobabeb and Evora. It shows that the KF-based algorithm presented by the authors is capable of delivering SEVIRI-based LST with an accuracy which is more or less comparable to that of the existing LSA-SAF retrieval methodology. There are, however, several weaknesses in the current manuscript that should be addressed before a publication can be considered.

The overall weakest point of the manuscript is the inter-comparison against other data products of surface temperature and emissivity and I think the paper requires major revisions along those lines. In some cases these sections describe what are in my opinion apple-versus-orange comparisons such as for example the comparison between the KF-based retrieval of SEVIRI SST (skin temperature) with the bulk temperature from the OI-based Reynolds AVHRR product. Furthermore the inter-comparisons are being carried out over monthly (e.g. Figure 13) and even yearly (Figure 14) averages. When using such long averaging periods the random error in the products is reduced considerably and the resulting statistics appear much more impressive than they are in reality. I think the authors should perform a proper spatial inter-comparison between the various products using swath-level data (for the LEO instruments). The 15 minute temporal sampling of SEVIRI makes the time-matching to the overpass time of MODIS/AVHRR quite straightforward to do. This is generally how sensor inter-comparisons are being carried out in validation studies and it will provide a much more realistic estimate of the algorithm performance.

I also had the feeling while reading through the paper that the authors in some cases described only those those statistics that appeared favourable for their algorithm while

C1106

ignoring others that would show their retrieval in a less positive way. In similar sections a bit further on, the statistics then showed a slightly different picture and those statistical metrics that previously were considered important are not mentioned anymore. I urge the authors to be more consistent in showing and describing all relevant statistical metrics in the text and to openly discuss them, even if they do not necessarily support the hypothesis that the suggested retrieval is superior to existing ones.

There are also several statements in the current manuscript that are not properly supported by the results (see specific comments). I encourage the authors to go over the manuscript again and make sure that every claim is backed up by real data and analysis (and that adequate evidence is provided in the text particularly when making bold claims).

Finally, while surface temperatures are validated against in situ station data and compared against other data sources, the KF-based emissivity products are not validated to the same extent. I think it would be beneficial for the paper to perform a direct spatial comparison of the full-disk emissivity maps (Figures 18-20) against existing emissivity maps from other data sources, e.g. from MODIS or ASTER GED.

SPECIFIC COMMENTS

P4050L12: "yields an accuracy of". What is this metric? RMSE? Please be specific.

P4050L17: "we have that emissivity". Something is superfluous here. Please revise this sentence.

P4051L7: "exemplified" -> "demonstrated" or similar

P4051L16: "is a breakthrough". Isn't that a bit grandiose? Especially since the KF algorithm provides similar accuracy as the LSA-SAF one? Please be modest and let the community decide on the merit of your approach.

P4051L22: "boosting" This sounds odd, please consider replacing with "complementing" or similar

C1107

LP4051L29: Namib -> Namibia

P4052L4: "analyses for sea surface" Please be specific about what products you are using.

P4052L7: "reference"? What makes it reference? Consider dropping this word.

P4052L8: Better write "The results from the validation exercise are reported in Sect. 4"

P4052L9: Section 2 needs a section on the actual SEVIRI data you used and how it was pre-processed before applying the Kalman filter. You need to provide information on how you performed cloud-masking and similar processing issues.

P4052L12: "Diverges"? Better write "increases" or similar

P4052L16: "reasonable" Says who? Have you tested this? How did you arrive at the 70 degree threshold? This needs justification.

P4054L4: "sequence of data" -> "periods". Also, this entire sentence reads very awkward. Please revise or consider splitting up into two sentences.

P4054L9: Please be specific. Which MODIS product did you use? MOD28?

P4054L9: I think this OI-based product is not suitable for comparison with your SEVIRI retrieval as it provides bulk temperature not skin temperature

P4054L24: This section needs more information about how the interpolation in space and time was carried out.

P4055L2: Please make a note here that when you refer to full disk you actually mean all pixels with ZVA < 70 deg

P4056L7: (T, Q, O) Please define those parameters

P4056L12: Again, please provide information about how the interpolation was carried out

C1108

P4056L24: Replace the semicolon with "and"

P4057L7: Please provide a short explanation about why you perform a logit-transform.

P4058L4: Either "posterior" or "a posteriori"

P4060L14: "clearness" -> "clarity"

P4060L23: Please also provide the RMS difference also in Figures 4 and 8

P4061L3: "very few data points". Please provide the number of data points going into each month (maybe in the Figure? or at least as a rough range in the text). Also, describe what do you consider too few data points to be significant and how did you arrive at that number?

P4062: "we now see a striking improvement". This is worded a bit strangely - it's not really an "improvement" as such since the plot is for an entirely different station. Also it does not really appear "striking" to me. Maybe write something like "we can observe a closer agreement of the linear fit line with the 1:1 line than for Figure 4" or similar

P4062L10: "slope close to unity". The slope was not used as a metric for Evora, maybe because there it worse than for LSA-SAF? Please be consistent when describing statistical metrics.

P4062L14: I believe you mixed up KF and LSA-SAF here? I fail to see how KF is supposed to "perform slightly better" than LSA-SAF. The bias for KF is twice as high as for LSA-SAF??

P4064L11: Why monthly averages?

P4064L18: A yearly average? This reduced the random error so much that the resulting metrics are not indicative of the true performance of the algorithm.

P4065L5 and this entire section: Again, I think you need to use an actual SST (skin temp) retrieval to perform such a comparison. You cannot validate a skin temperature

C1109

product with an OI-based bulk temperature product.

P4066L4: "really run" -> "really be run"

P4066L14: But in the introduction you say that "the limit of 70 degree is reasonable". I think you will need to perform a proper analysis of this.

P4066L20: "seas of sand". This is very poetic but probably not too suitable for a scientific article.

P4067L6: "exemplified". Better write "demonstrated" or similar

P4067L10/11: But the behaviour of the algorithm in the presence of data gaps has not been tested specifically. I think this statement is too bold without actually testing this in detail. Please consider revising.

P4067L13: "better than 1.5C" This needs to be qualified. Over which surfaces? Based on which comparison?

P4067L14: "bias less" -> "bias of less"

P4067L23: Again, this has not been tested properly. Either add some text or figures in the results section specifically studying the impact of the KF-based approach in the presence of clouds or consider revising this statement.

P4068L1: "safely applied to the SEVIRI full disk". This statement is not supported by evidence. In fact in Section 4.4. the authors themselves suggest that unrealistically low temperature values in South America could be due to the "SEVIRI point spread function at this angle". Without specifically testing the angle-dependence of the algorithm properly I do not think that this statement holds.

Figure 4 Caption: R^2 is not the linear correlation coefficient but the coefficient of determination

Figures 5-7: The caption should mention that this is for Evora.

C1110

Figure 6b: Move the legend so it doesn't cover the time series

Figure 7: I think these plots should show both the actual data (maybe in a light color) and over it the time series smoothed with the moving average filter. Also, please show the corresponding emissivity from other sources (e.g. UW and/or LSA-SAF) for reference.

Figure 8 Caption: R^2 is not the linear correlation coefficient but the coefficient of determination

Figure 11a: It could be helpful to show here also the time series of the difference between KF and in situ temperature, in order to be able to discern where they differ (this is hard to see in Figure 11a).

Figure 12: This histogram would be a lot more useful if you limited the x axis limits to something like the range of -10 C to 10 C or so and increased the number of classes/bins.

Figure 13 Caption: Note that this is for 2013

Figures 18-20. These emissivity maps are nice. You should perform a spatial comparison (e.g. difference maps) of these maps against other emissivity database, e.g. from MODIS or ASTER.

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 4049, 2015.