Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2016-268-RC2, 2016

Atmospheric Measurement Techniques Discussions

Interactive comment on "Deriving clear-sky longwave spectral flux solely from hyperspectral radiance: a case study with AIRS observations" *by* Xiuhong Chen and Xianglei Huang

Anonymous Referee #2

Received and published: 9 October 2016

GENERAL COMMENTS

The manuscript addresses a relevant technical question (radiance-to-flux conversion) in the retrieval process of products from AIRS satellite instrument data and instruments alike. As to my knowledge, the presented algorithm is a unique realisation of such a radiance-to-flux conversion, even though the concept of "deriving flux solely from radiance" is not new (e.g. doi:10.1007/s00376-015-5040-8). Since the derivation of the flux from such measurements generally contains several assumptions and is, by no means, straight-forward, I support a diversity of reported conversion algorithms.

The conclusion, i.e. the presented algorithm compares well with CERES OLR estimates, is sufficient for publication. The methods and assumptions are mostly sound

and clear; see "specific comments" for limitations. The manuscript is well structured and language is fluent and precise.

In the following, I will only address aspects listed at http://www.atmospheric-measurement-techniques.net/peer_review/review_criteria.html, if they give reason for concerns.

SPECIFIC COMMENTS

One major issue that concerns me is the clear-sky selection. It should be clearly outlined, why the algorithm is refined to clear-sky cases at all. Hints are given the very last paragraph of the manuscript. I suggest to add appropriate reasoning in the introduction around line 74; a short mentioning of the reason in the abstract could also be helpful.

The clear-sky selection process as described in section 3.1.1 of the manuscript seams reasonable. However, the assessment of the performance has substantial deficits:

The authors use "accuracy" as metric for the performance, which they define as "percentage of cases in which our algorithm can correctly classify the footprints". Given that only 9-10% of all observations are clear-sky (line 185, I assume this is the figure for CERES classification), there is no skill needed to reach an overall accuracy of 88.7% (Table 2: Near-globe Accuracy of Steps 1+2+3). Using the trivial approach "all observations are cloudy-sky" would yield an accuracy of 90-91%!

Furthermore, the "false positive" metric and the implications of the according numbers given in Table 2 is not addressed well enough in the performance assessment (section 3.1.2). The given false positive rate of 11.1% (Table 2: Near-globe FP of Steps 1+2+3) actually exceeds the occurence of the event (i.e. 9-10% clear-sky).

Say there were 1000 CERES cloudy-sky observations and 100 clear-sky:

FP=11.1% yields 111 AIRS clear-sky are CERES cloudy-sky (misclassified) and 889 AIRS cloudy-sky are CERES cloudy-sky (correct)

FN=13.9% yields ∼14 AIRS cloudy-sky are CERES clear-sky (misclassified) and ∼86 AIRS clear-sky are CERES clear-sky (correct)

In total, this is:

889+86 = 975 correct classifications (consistent with ACC=88.7% of 1100 observations)

111+86 = 197 AIRS clear-sky classifications

889+14 = 903 AIRS cloudy-sky classifications

So, essentially more than half of the AIRS spectra for which the radiance-to-flux algorithm shall be applied, are "misclassified" in the CERES-comparison-sense. I would not say that this rules out the suitability of the clear-sky detection algorithm. Also, most CERES clear-sky observations are captured by the detection algorithm, which proves skill. Since the CERES clear-sky fraction does not compare well with an average planetary cloud-cover of around 2/3, while the suggested AIRS-based detection comes closer to that, this needs some discussion in the assessment. Footprint-size is definitely an issue here, since classification seems to sort out pixels with very little cloud cover amount inside the pixel.

The reader should be clearly informed that the number of clear-sky observations used for AIRS- and CERES-derived flux comparison differs by a factor greater than two. Since this fact does not become apparent from the presented data in section 3.1.2 and Table 2, I would argue, that the applied metrics (FN, FP, Accuracy) are rather inappropriate and misleading here. Presenting and discussing actual numbers or other metrics and a more comprehensive discussion of this issue would improve the manuscript. Another possibility to improve comparability of results with CERES could be stricter threshold values for clear-sky detection.

I see a similar problem in the assessment of sub-scene type classification (section 3.2). The authors summarise, that the accuracy of their classification is "80% or even

higher" (line 204). I cannot support this conclusion from table 3, since common sub-scene types (e.g. "-1-") have substantially lower accuracy. More important though, the "accuracy" metric does not seem appropriate to judge the performance. Many sub-scene types in table 3 have a lower occurence than the misclassification rate (by which I mean the complement to "accuracy").

Minor comments:

* Mentioning of the word "satellite" in title, abstract, and/or key words would help readers to assign the article to the right discipline

* The term "solely" in title and abstract is slightly misleading, since ECMWF surface temperature analysis is used in the process (line 150)

* Mention footprint size/ pixel size of CERES SSF detection of clear-sky in section 2

* Does the algorithm account for attenuation due to the height of the satellite above the "top of atmosphere"?

* line 211: Is F_AIRS an integration over Theta? This is missing in the equation.

* line 212-213: Please provide a reference to the exact algorithm used. E.g., which spectral range is used? From lines 97-98 one could assume a greater spectral range than what Huang et al. (2008) uses.

* line 236: Reasoning why "These circumstances make it difficult" would be helpful here

TECHNICAL CORRECTIONS/ COMMENTS

* The paper uses wavelength and wavenumber for referring to certain parts of the IR-spectrum. This seems particularly confusing in two paragraphs: lines 96-101 and 135-156. Using one of the two consistently would improve readability.

* line 105, line 391-394: I could not find the referenced document, only a "Version

1" of it under http://ceres.larc.nasa.gov/collect_guide.php. Please provide an URL if possible.

* line 159: consider changing "another words" to "other words"

* line 168: consider changing "dash-dot" to "dashed"

* Figure 1: Equal bin sizes in all panels would improve comparability

* line 239: From Table 3, I would read "no more than 1%", not 2% (sum of sub-scene types "–4" and "–5")

* line 240 and Figure 2: These numbers are hard to read from the figure. Consider using a discrete colourbar, e.g. in steps of 0.005 Wm-2/10cm-1

* Figure 2: "Sub-scene type" labels should be moved slightly up to align with the corresponding colouring in the figure

* Figure 2: Marking the spectral bands of actual AIRS observations in the figure would help in judging the differences shown. This could easily be marked on the top or bottom axis.

* line 250: Are the numbers for RMS weighted by area of the respective 2°x2.5°-pixel?

* line 270-273: The reader gets the impression, that only the "false positive" cases were used here, but lines 268-269 state otherwise (all AIRS clear-sky, i.e. "false and correct positive"). Please clarify.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-268, 2016.