Interactive comment on “Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces” by H. Diedrich et al.

Anonymous Referee #1

Received and published: 5 September 2014

General comment

The paper presents a method to retrieve water vapour columns from MODIS over land. The focus of the paper fits well to the topics covered by AMT and, therefore, this paper should be considered for publication after addressing the comments listed below. I think it is important to publish these results. The structure of the manuscript makes it a good read. The content, however, still needs a few major and some minor revisions.
Major comments

1) The TCWV correction algorithm is described in Sect. 4. It uses ground-based MWR data in order to derive a correction for the TMWV retrieval from MODIS. The same MWR data, however, is used to "validate" the algorithm results in Section 5. I do not believe this is appropriate because independent data is needed to provide a meaningful assessment of the algorithm output. Hence, the "validation" exercise using MWR data should be removed and/or replaced. Furthermore, calling this an "extensive" validation is not justified. Calling it a "comparison" would be more appropriate.

In order to solve this issue and to improve the manuscript, the authors may consider to include further datasets into their study. For example, algorithms from many different satellite sensors are mentioned in the introduction. Many of their results are freely accessible or accessible upon request. Why not inter-comparing the results from the new TCWV retrieval from MODIS with those? This information would be of high value for the reader.

Furthermore, the MODIS algorithm is proposed as a "gap filler" (page 7771, line 26) after MERIS died. What I miss in the manuscript is a showcase that the proposed is indeed feasible. I would like to ask the authors to consider including a comparison of their data with MERIS data from Lindstrot et al., 2012, which should be fairly easy to obtain. The readers should at least know, when this will be done or that it at least should be done.

2) The manuscript mentions a cloud filter at several places. The only information I could find about this filter algorithm is "only […] 100

2) This paper describes the retrieval of MODIS TCWV and, apparently, at least the entire Aqua data-set has been processed to perform the "validation". Now you have this excellent data set, why not show it? Satellite measurements are a lot about imaging and I found it puzzling that there is not a single map included. Please consider including
TCWV maps, e.g. showing - a single orbit from MODIS Terra besides a MERIS orbit, - a monthly average, or - averages of the surroundings of the three ARM sites used for determining the correction.

4) Again on the "validation". in lines 17f on page 7769 it is stated: "Additionally, all outliers (deviation > 3\(\sigma\)) were rejected." I wonder whether applying this filter is actually required. If it is necessary, please explain how it works and why it is really, really necessary. I would prefer including all data in order to increase the significance of the comparison. I believe it is very dangerous to compute correlations using filtered data.

**Minor comments**

Generally, the verb "see" may always be removed in links: "(see Sect. 3.4)" -> "(Sect. 3.4)" Everybody knows, what to to with references. Please avoid clutter.

page 7754:

line 5: "at spatial" -> "at a spatial" line 7: "introducing correction" -> "introducing empirical correction" line 8: "that a" -> "that, a" line 9: What is an "extensive" validation versus a not extensive one. See comment above. line 22: Delete "hile on short time scales, w" or rephrase.

page 7755:

line 7: "over land is" -> "over land, however, is" line 8: "However, t" -> "T" line 9ff: Please include some further references, at least:


line 12: What do you mean with "very" suitable? line 13: "all surface types are sufficiently bright" -> the surface brightness only weakly depends on surface type line 13: C2492
delete " in this spectral interval" line 14: let "Lindstrot et al. (2012) introduced a procedure" begin a new paragraph line 15: delete "extensive" line 17: Is the other band literally without "any" absorption? Is there the possibility of interference? Please specify! line 22: insert "applying MERIS measurements" after "time series" line 28: Give figures/reasons why "the accuracy of this product" is limited! For which applications would it be sufficient?

page 7756:

line 5: Please specify, why an overestimation is unsatisfactory when investigating trends. In principle, a constant offset does not change the derivative. line 10: "in comparison" -> "compared" line 21: "pixel by pixel" -> "pixel-by-pixel" to be consistent with the rest of the text line 23: "However, an empirical factor accounting for ... is needed."

page 7757:

line 4: Is global coverage gained in one or two days? On what does this depend upon? line 14 move "in band 18" between "pronounced" and "and"

page 7758:

line 7: What is a "good" background? Please provide a reference. line 12f: "approximated as (simplified form) :" -> "simplified to:" line 16: use abbreviation: "top-of-atmosphere solar flux" -> "TOA solar flux" line 23: What is the exact meaning of "air mass"? Please specify its definition (there are several "air masses") and why it is dimensionless. line 25: insert "a" after "of"

page 7759:

line 7: "actual" Specify, that the surface temperature needs to be inferred from a meteorological model. line 18: insert "only by water vapour" after "pure absorption" to specify "pure" line 21: "f is increased" -> "f increases"
line 3: insert "without H2O absorption" after bands. to specify "window"
line 7: Please specify, why the "secant method" is no longer applicable.
line 12: How was this issue solved by Lindstrot et al, 2012? The list is intended to describe the differences.
line 12: Is this an improvement?
line 20ff: Please provide already defined symbols yet again in order to improve readability of formula (4).
line 23: How is the simulation of $T_{\text{noscat}}$ performed anyway? What model is used to calculate the LUTs? Is the choice of atmospheric parameters sufficient? Is the number of simulated atmospheric conditioned justified? Please give more details here! Nobody will be able to reproduce your results otherwise.

line 2: How is the interpolation in the LUTs performed? Is it linear? Is the method sufficient?
line 4: "highly" -> "sufficiently"
line 7: Why not taking the surface pressure from ECMWF as well? It may also be possible to infer the surface pressure on the more detailed GTOPO30 model using our knowledge on adiabatic expansion. If not, please explain why different sources are used for the different surface meteorology parameters.

line 15: insert "," after values

line 6ff: Please provide enumerated formulas.

line 14f: Please define, what $c_i$ are and in which formula they appear. Please be more specific.

line 15: erase "-" before optical, or replace it by "aerosol"
line 19: Please justify the parameters of the default aerosol profile chosen. How does it look like? What is the presumed phase function?
line 7: insert "," after "Finally"

page 7765:

line 6: insert plural "s" after "author" line 11: Please provide citation/URL of the DWD archive.

page 7766:

line 14: "has a built-in so-called" -> "features a"

page 7767:

line 2: omit " the following" line 3ff: Maybe use an enumerated list to improve readability of the outcomes. line 5: "Additionally" will be obsolete, then. line 16f: The sentence starting with "Here, ..." is unclear. Please rephrase and be more specific. line 22: What is "fairly constant"? Is it "sufficiently constant"? What would "sufficient" then be?

page 7768:

line 1: "not shown here" -> Why? line 16: omit "beforehand" line 16: "optimizing" -> How is the optimization performed? What is the cost function? I presume minimisation taking place?

page 7769:

line 1: omit "separately" after "done" line 1: "separately." -> ", respectively" line 2: "Table" -> "Tables" line 2: "5 and 6" -> "5 and 6 for Aqua and Terra instruments, respectively." line 13: What kind of cloud data is used? What are the thresholds? Please specify. line 16: "limited accuracy of the MODIS geolocation" -> What is the accuracy? Why is it a limitation? It is certainly better than 20km. line 21: "Section." -> "Sect."

page 7770:

line 5: "of" -> probably "less than" line 9: How large is the number of samples? The text only states "very" high which is not very satisfactory. line 16ff: "In order to account
for cloud contamination but also preserve a sufficient high number of data-points, only cases with less than 50

page 7771:

line 2f: "Whereas the bias between GNSS and and MODIS is shifted to negative values (meaning that MODIS values are generally higher than GNSS)." -> "Whereas, MODIS values are generally higher than GNSS."

line 5f: Why is the std. dev. so high? How high is it? Can you provide an upper limit for the maximum amplitude?

line 10: "which" -> "whose"

line 11f: "the surface reflectance, transmittance due to atmospheric water vapour and the shortening" -> "(1) the surface reflectance, (2) the transmittance due to atmospheric water vapour, and (3) the shortening."

line 13f: "where measurement-and forward-model-uncertainties are considered." -> "where the uncertainties of measurement and forward-model are considered, respectively."

Furthermore, please extend the summary and outlook:
- Compare to operational MODIS TCWV validation. Please discuss!
- What about the retrievals in other spectral ranges mentioned in the introduction? Please discuss: accuracy, precision, differences, advantages and disadvantages!

Figures (pages 7782 to 7792):

Figures 1, 4, 6, 7, and 10 are black and white. It would be nice to switch to colors and provide a colorbar.

page 7786:

Figure 5: It is hard to make out any information. Why not copy the style of Figures 4, 6, and 10? Please improve the readability of this plot!

page 7792:

Figure 11: Please improve the readability of this plot. The caption is confusing and needs to be rephrased. Please also increase the contrast, preferably with col-
ors stronger than just black and blue. Another option would be, to use whiskers to better illustrate the statistical dimension of the displayed data.