

Review of “Evaluating sources of an apparent cold bias in MODIS land surface temperatures in the St. Elias Mountains, Yukon, Canada”

Overview:

This manuscript describes analysis of in-situ air temperature data relative to the MODIS LST product (MYD21) with supplemental data from ASTER, Landsat, and additional meteorological measurements at two sites in the St. Elias Mountains. The central focus of the manuscript is to determine the causes of difference between in-situ air temperature measurements and MODIS LST data (called the “MODIS offset”). Three main hypotheses are explored: 1) The large footprint of the MODIS pixels results in the offset; 2) The lack of constraint in surface emissivity results in the offset; 3) Near-surface temperature inversions lead to this offset. The work demonstrates that the first two hypothesis are unlikely to play a large role in the offset, and that the near-surface inversions are likely a reason why there is a difference between in-situ air temperature and MODIS LSTs. While this has been shown in previous studies, the authors indicate that this is the first time that inversions have been identified as a cause of this offset in this type of environment. The manuscript makes a compelling case that the system studied is an important one given the role of Alaskan glaciers in sea level rise and the fact that the surface temperatures in this region are understudied. Understanding how representative MODIS LSTs (and other remote sensing surface temperature products) are of actual surface conditions is critical to effectively monitoring this region.

Overall, this is important work and a useful contribution. I think that the work can be improved by increased clarity and discussion about the fact that a comparison of 2m air temperature and MODIS LST is not a 1:1 comparison. I see that it is addressed in the text, but I believe it needs to be a bit more explicit. For example: in Section 4.4, I think the language of “corrected” MODIS LST is a bit misleading because there isn’t a true error in MODIS LSTs identified; it is simply determined that it is not equivalent to 2m air temperature. The linear regression implemented is a tool to adjust the LST to an air temperature. Furthermore, in the abstract and conclusion section, I think it is important to clarify that this analysis doesn’t rule out the possibility that there may be measurement errors in the MODIS algorithm for determining surface temperature because there was not a direct comparison to in-situ skin temperature in this study. While convincing evidence is provided that inversions are likely a source of difference between MODIS LST and in-situ 2m air temp, it is still possible that MODIS LSTs have error relative to in-situ skin temp too. Lastly, in the title, I would be hesitant to call this a cold bias without explicitly stating that the “bias” is relative to air temperature measurements.

I have several specific comments to improve clarity and increase detail about methodology, but the manuscript is overall well-written, the figures are clear, and the methods are sound. This is a very rich dataset, and the analysis and results provide an important contribution to our understanding of near surface temperatures in understudied snow and ice covered regions.

Specific Comments:

Lines 4-5: In the statement that MODIS LSTs are offset from AWS data, is this referring to prior studies or the current work? Please clarify.

Lines 15-17: I agree that understanding near-surface physical properties is critical to convert MODIS LSTs from a surface temperature measurement to an air temperature measurement, but this is different

from improving the accuracy of the MODIS LST, which this work doesn't directly address because it compares in-situ air temp (and not in-situ skin temp) to MODIS LSTs. Please revise this statement accordingly.

Line 41: The statement that all in-situ temperatures are air temperatures is key here. I think it requires a bit of further discussion. Given the Guillevic et al. (2017) report, validation should be done with skin temperature wherever possible. I understand that in-situ skin temperature data is not available in this case, and I believe comparison to in-situ air temperature is a worthy endeavor, but I think it should be explicitly addressed here (or elsewhere) that this is different from a standard validation of the MODIS LST product.

Table 1: Consider adding to this table further information about the footprint of each sensor, the specific products/instruments used, and the uncertainty associated with each measurement (if available).

Lines 118-122: In the window when the two sensors at Divide overlap, which dataset is used?

Lines 123-125: Was the container containing the iButton sensor ventilated? Was it a light color to limit absorbed solar radiation? I know that there can be issues with iButton sensors heating up during periods of high incoming solar radiation.

Lines 125-126: It looks in Figure 3 like the Eclipse Weather Station and Eclipse iButton datasets do not overlap, so it's not clear to me how it was determined that the records were consistent. Please clarify and provide data if needed, perhaps in a supplement.

Lines 135-143: I think that choosing to look at a nearby MODIS pixel that does not include the darker nunatak surfaces is probably a good idea, but the implications of this choice should be further explored. Perhaps the air temperature above the dark surface actually is higher than the air temperature above the nearby ice/snow. Comparing the MODIS pixels and providing that information in Table 2 is great, and I think that in the discussion of results, the manuscript should come back to this and address what the implications of this choice are on the results.

Line 145: I see that later in the manuscript there is a justification provided for choosing to only consider data from the time window from 11:00 – 1:30. I think it would be appropriate to address this here in the methods.

Lines 145-147: Just to clarify, the 742 images at Divide span 20 years of data, and the 100 images at Eclipse span ~2 years of data. Is that correct? Please state that clearly here. Also, please specify how the MODIS images and AWS data are paired in time. Is it simply the closest hourly measurement that is paired with the MODIS temperature, or is it some kind of average of multiple AWS measurements?

Section 2.3: Please clarify if only ASTER and MODIS images from the 11-1:30 window that had paired AWS data are used in this comparison.

Lines 168-169: Please explain how Landsat images are "examined for cloud cover." Is there a particular algorithm used, and are particular thresholds implemented?

Equations 2+3: I believe it might be appropriate to use \approx instead of $=$ in these expressions.

Line 200: Provide a bit more information about the atmospheric emissivity from ERA5.

Line 202: Please provide a bit of description of the surface types that are present at the two sites. Is surface melt common? This doesn't necessarily need to be right here, but the point about albedo raised the question of if this is a reasonable assumption or not.

Section 3.1 (and elsewhere): When comparing different seasons, I think that the results of the Wilcoxon rank sum tests can be referenced in the tables and don't need to be repeated in each paragraph. The many parentheticals make it a bit challenging to read and the statistics are clearly presented in the tables. Maybe the journal or author preferences dictate that it should also be in each paragraph, but I think it interferes a bit with clarity.

Figure 4: 1:1 plots of in-situ air temperature vs. MODIS LST are very instructive, and should potentially be included in addition to these box plots.

Lines 247-248: Please indicate if it was tested that the data are normally distributed prior to linear regression, and normalize data if necessary.

Figure 8: Adding a horizontal line at 0 C would be helpful in figures a and b; Is this data for all 20 years when Divide AWS data are available?

Figure 9: Adding a horizontal line at 0 C would be helpful in figures a and b

Lines 287-290: Is it possible that the lack of difference between offsets in MODIS brightness and surface temperatures in the fall and winter is also because of increased inversions relative to other seasons?

Lines 294-295: I would potentially expect a wider variability in surface conditions (and thus emissivity) in the summer, but this may be site specific. Perhaps explain more why winter could result in more rapidly changing emissivity.

Lines 304-305: The analyses exploring links between accumulation and the LST offset seem to have covered a wide array of options. Was "days since last accumulation" one of the ways this data was analyzed? I couldn't quite tell from the descriptions and that might be a reasonable metric to consider.

Lines 314-315: Why used different emissivity values for the different sites, or is this a typo?

Line 321: Is it a discrepancy because it is not accounted for given the assumptions used in the T_{surface} calculations presented here?

Lines 323-338: Consider also comparing to 2m near surface inversions investigated in Nielsen-Englyst et al. (2019).

Line 342-343: Perhaps I'm missing something, but why wouldn't winter surface temperatures also be high then?

Lines 370-375: Please explain in a bit more detail how the data are processed before going into the linear regression. I suggest framing the linear fit with the actual variables desired (instead of generic x and y). The corrected MODIS temperatures have a low mean error because of how they are calculated, but how does the standard deviation compare in magnitude to the interannual variability? Is a better fit to the data achieved if it is not averaged annually but instead each paired data point (of 2 m air temp and MODIS LST) is part of the regression?

Minor Comments:

Line 42: Consider starting a new paragraph with the sentence that begins “Instrumentation...”

Line 68: Consider phrasing as “apparent MODIS LST offset” here and elsewhere?

Line 86: Consider adding a definition of a nunatak

Lines 88-89: Consider reversing sentence that begins with “Here...” so that it discusses inversions first, which were just being discussed. I found the sentence a bit confusing on the first read because I was unsure which hypotheses were being referenced.

Line 221 (and elsewhere): forgotten degrees C units on reported median.

Line 222: “MODIS temperature difference data spaces >10 degrees C” – is this saying the maximum offset is greater than 10 degrees C? Please clarify.

Line 249: Perhaps an “overwhelming majority” or some other language would be better suited to describe 95%.

Line 256: “modeled temperature difference data is greater than 60 degrees C” – does this mean that the modeled inversions are sometimes up to 60 degrees? Please clarify.

Lines 275-280: Since the manuscript already revealed that the brightness temp and surface temp do show the same pattern, going through the hypothetical scenarios in the way they are framed confused me a bit. Consider rephrasing to make it clear that one of them we already know to be what was observed.

Citations:

Guillevic, P., Göttsche, F., Nickeson, J., Hulley, G., Ghent, D., Yu, Y., Trigo, I., Hook, S., Sobrino, J.A., Remedios, J., Román, M. & Camacho, F. (2017). Land Surface Temperature Product Validation Best Practice Protocol. Version 1.0. In P. Guillevic, F. Göttsche, J. Nickeson & M. Román (Eds.), Best Practice for Satellite-Derived Land Product Validation (p. 60): Land Product Validation Subgroup (WGCV/CEOS), doi:10.5067/doc/ceoswgcv/lpv/lst.001

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R., Dybkjær, G., & Alerskans, E. (2019). In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic. *The Cryosphere*, 13(3), 1005-1024.