



Article

Cite this article: Zikan KH, Adolph AC, Brown WP, Fausto RS (2023). Comparison of MODIS surface temperatures to in situ measurements on the Greenland Ice Sheet from 2014 to 2017. *Journal of Glaciology* 69(273), 129–140. <https://doi.org/10.1017/jog.2022.51>

Received: 22 July 2021

Revised: 31 May 2022

Accepted: 6 June 2022

First published online: 8 August 2022

Key words:

Ice temperature; remote sensing; snow/ice surface processes

Author for correspondence:

Karina H. Zikan,

E-mail: karina.zikan@gmail.com

Comparison of MODIS surface temperatures to in situ measurements on the Greenland Ice Sheet from 2014 to 2017

Karina H. Zikan^{1,2} , Alden C. Adolph¹ , Wesley P. Brown¹ and Robert S. Fausto³

¹Department of Physics, St. Olaf College, Northfield, Minnesota, USA; ²Department of Geoscience, Boise State University, Boise, Idaho, USA and ³Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark

Abstract

Remotely sensed land surface temperature (LST) data, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) LST thermal infrared products, are useful for monitoring surface processes on the Greenland Ice Sheet in remote areas but must be validated to ensure accuracy. Using data from the Programme for Monitoring the Greenland Ice Sheet (PROMICE), we conducted a MODIS LST validation (MOD/MYD11 C6 swath level product) using radiometric in-situ skin temperature records from 2014 to 2017 over 17 PROMICE sites mostly in the ice sheet's ablation zone. There is a significant cold bias in MODIS LST when compared to PROMICE skin temperature, particularly when PROMICE records temperatures below 0°C (mean bias: $2.4 \pm 0.01^\circ\text{C}$ mean \pm standard error, RMSE = 3.2°C). Multiple linear regression analysis reveals the difference between MODIS LST and PROMICE skin temperature is larger at lower temperatures, lower latent heat fluxes and higher specific humidity. Our results confirm the presence of a progressive cold bias in the MODIS LST that should be considered in use of this product, and we identify and corroborate areas for ongoing algorithm development.

1. Introduction

In the Arctic, near-surface temperatures are rising at more than twice the rate of other regions of the globe due to Arctic amplification (Graversen and others, 2008; Serreze and Barry, 2011). Efforts to limit temperature increases globally to $+2^\circ\text{C}$ will result in an Arctic environment that has an average annual temperature that is $+4^\circ\text{C}$ higher than the preindustrial period (Overland and others, 2019). These warming trends increase ice melt and polar instability, which influence global climate and weather systems directly and indirectly through changes to the surface energy balance exacerbated by the snow albedo feedback (e.g. Déry and Brown, 2007; Box and others, 2012; Thackeray and Fletcher, 2016). The Greenland Ice Sheet (GrIS) is highly sensitive to temperature changes because many locations on the ice sheet are already at or near the melting point of ice, so further temperature increases are expanding the areas over which melt occurs (Colosio and others, 2021). The GrIS is losing an estimated average of 121 Gt a^{-1} of ice mass, relating to 0.33 mm a^{-1} of sea level rise (Vaughan and others, 2013), with that value increasing over the past several decades (Shepherd and others, 2020). The rate of ice mass loss of the GrIS is expected to increase in future years. Temperatures over the GrIS are rising (Shuman and others, 2001; Vaughan and others, 2013), with winter temperatures estimated to have increased by 4.4°C between 1991 and 2019, and summer temperatures increasing by 1.7°C during the same period (Hanna and others, 2021). Surface melt has caused just above 50% of mass loss between 1992 and 2018 (Shepherd and others, 2020) and is the cause for an increasingly large fraction of total ice mass loss (Enderlin and others, 2014; Mouginot and others, 2019). Land surface temperature (LST) is a measurement of the thermodynamic temperature of the outermost (skin) layer of a surface (Guillevic and others, 2017) and is a key determinant in analyzing surface energy processes and identifying and predicting ice melt. The GrIS's active role in the global climate and climate impacts makes accurate LST measurements of the ice surface essential to monitoring the GrIS and its contribution to our changing climate.

Due to the harsh climate and remote nature of the GrIS, in situ temperature measurements are difficult to obtain over a large spatial and temporal scale. Rapid surface changes in the ablation zone due to melt and enhanced surface velocity can make measurements and station maintenance even more challenging. Satellite-based LST measurements provide consistent and expansive data from the GrIS that can be used in a variety of climate analyses; however, these data must be validated and assessed on a regular basis due to issues of sensor accuracy and orbital drift among other concerns (e.g. Jin and Treadon, 2003). There are a number of remote-sensing instruments that produce thermal infrared LST products including the Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Landsat. The National Aeronautics and Space Administration (NASA) launched the Aqua and Terra satellites in the early 2000s, each carrying a MODIS instrument which is used to calculate many environmental variables, across the globe. All three instruments (MODIS, ASTER and Landsat) are on satellites with

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<https://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

cambridge.org/jog

solar synchronous orbits, but offer different resolutions and different revisit cycle times due to swath width. While ASTER and Landsat provide higher spatial resolution than MODIS (ASTER: 90 m; Landsat: 100 m; MODIS: 1 km), they have lower temporal resolution with a nominal 16-day revisit cycle for ASTER and Landsat versus a nominal daily revisit cycle or greater for MODIS. Due to the orbital pattern and swath size, MODIS collects data multiple times per day over the GrIS, providing significantly more data for comparison to in situ measurements. Because of this extensive coverage, MODIS data are widely used, yet there remain uncertainties in the validation that require further investigation.

MODIS LST data have been used to estimate temperature trends and determine GrIS melt extent (Hall and others, 2013), to estimate near-surface air temperature over ice-covered locations (Williamson and others, 2014; Zhang and others, 2018; Qie and others, 2020), and for modeling ice sheet or snowpack surface processes (Fréville and others, 2014; Shamir and Georgakakos, 2014; Navari and others, 2016). However, MODIS LST data are limited to clear sky conditions because cloud cover obstructs the path of radiation from the surface and therefore the ability to calculate surface temperature. Additionally, LST measurements can be affected by high concentrations of atmospheric aerosols (Wan, 2014). While there have been efforts to expand coverage during periods of cloud cover with interpolation algorithms for treating cloudy pixels (Jin, 2000; Jin and Dickinson, 2000), this issue of cloud cover limitation has led to a cold bias when LST measurements are averaged over weeks or months. The cold bias occurs because cloud cover typically creates warmer ground conditions that are undetected by MODIS (Westermann and others, 2012). Previous validation studies have also identified a cold bias in individual MODIS LST measurements when compared to in situ ground-based measurements (Hall and others, 2008; Koenig and Hall, 2010; Østby and others, 2014; Shuman and others, 2014; Williamson and others, 2017; Kindstedt and others, 2021).

When considering the topic of ‘surface’ temperature, we must carefully define the terms that we use for various temperature measurements, as ‘surface’ temperature in prior studies can be used to refer to near-surface air temperature instead of the true surface or skin temperature. In this manuscript, we will use the terms ‘land surface temperature (LST)’ or ‘skin temperature’ to refer to the radiometric temperature of the snow or ice surface, typically representative of the top several microns to millimeters of the surface given that the measurements are made in the thermal infrared wavelengths (Warren and Brandt, 2008). We will use the term ‘air temperature’ to refer to temperature measurements made ~2 m above the snow or ice surface, a typical measurement on a meteorological station.

Given the importance of remotely sensed MODIS LST products and their limitations, it is vital to validate data accuracy in polar regions. Previous studies have found a significant cold bias (findings range from 0.7°C to as high as 7°C) present in MODIS LST data (Hall and others, 2008, 2018; Koenig and Hall, 2010; Hachem and others, 2012; Westermann and others, 2012; Østby and others, 2014; Shuman and others, 2014; Williamson and others, 2017). Shuman and others (2014) observed that this cold bias became more pronounced at lower temperatures, which we refer to as a ‘progressive cold bias’. Several past validation studies have compared MODIS LST to ~2 m air temperature measurements. However, recent studies on the temperature difference between 2 m and skin measurements indicate that MODIS data should be validated using skin temperatures (Adolph and others, 2018; Guillevic and others, 2017). The need to directly compare MODIS LST to skin temperature instead of validating remote-sensing LST with air

temperatures is due to the presence of near-surface inversions in snow-covered regions. One Arctic-wide study of snow- and ice-covered surfaces found that skin temperatures were 0.65–2.65°C lower than air temperatures on average, and inversions were present 85% of the time (Nielsen-Englyst and others, 2019). When comparing MODIS directly to in situ infrared measured skin temperatures, Adolph and others (2018) did not find evidence of a cold bias of the MODIS LST at Summit, Greenland, while earlier studies that compared MODIS LST to air temperature did find a cold bias (Hall and others, 2004; Koenig and Hall, 2010; Shuman and others, 2014). However, the study by Adolph and others (2018) was limited to a single site and a small range of surface temperatures that did not include temperatures below –35°C or above –5°C.

The study presented here compares the MODIS MOD/MYD11_L2 Collection 6 swath-level LST product from both the Aqua and Terra satellites to skin temperature measurements calculated from upwelling and downwelling longwave radiation measurements collected by Automatic Weather Stations (AWS) installed primarily in the ablation zone of the GrIS by the Programme for Monitoring the Greenland Ice Sheet (PROMICE). This work builds on the findings of past validation studies to determine if there is a MODIS LST cold bias over the GrIS. We also investigate the progressive cold bias identified by Shuman and others (2014). We analyze data from the years 2014 to 2017 and explore the effect of environmental influences on the identified differences between the MODIS and PROMICE skin temperatures.

2. Methods

2.1 MODIS land surface temperature product

This study uses the MOD/MYD11_L2 (MxD11) Collection 6 swath product from 2014 to 2017. The MxD11 product calculates LST with a 1 by 1 km pixel footprint using a split window technique; brightness temperature measurements from band 31 (T_{31} , bandwidth: 10.780–11.280 µm, spectral radiance: 9.55 (300 K)) and band 32 (T_{32} , bandwidth: 11.770–12.270 µm, spectral radiance: 8.94 (300 K)) are implemented together to calculate the surface temperature, taking advantage of the difference between the two bands to account for atmospheric effects (Wan and Dozier, 1996; Wan and others, 2002; Wan, 2014). MxD11 uses the following equation to calculate LST (Wan, 2014):

$$\text{LST} = b_0 + \left(b_1 + b_2 \frac{1 - \varepsilon}{\varepsilon} + b_3 \frac{\Delta\varepsilon}{\varepsilon^2} \right) \frac{T_{31} + T_{32}}{2} + \left(b_4 + b_5 \frac{1 - \varepsilon}{\varepsilon} + b_6 \frac{\Delta\varepsilon}{\varepsilon^2} \right) \frac{T_{31} - T_{32}}{2}, \quad (1)$$

where ε is the mean emissivity and $\Delta\varepsilon$ is the difference of the emissivity of bands 31 and 32, and b_0 through b_6 are a series of coefficients that are calculated using viewing zenith angle, surface air temperature and atmospheric water vapor content (Guillevic and others, 2017). In the MxD11 product, emissivity values for each pixel are estimated based on land cover type. In the dataset we use, emissivity in band 31 ranges from 0.972 to 0.994 (mean = 0.994), and emissivity in band 32 ranges from 0.970 to 0.990 (mean = 0.990).

Because of the high latitude of the Greenland sites, there are multiple passes of the Terra and Aqua satellites overhead in a given day. Each swath is time stamped and images are within 5 min of the stated time. When clouds are not present, this can provide sub-daily measurements of surface temperature. We use the built-in cloud filtering algorithm in the LST product which is

drawn from the MOD/MYD35 cloud mask product and only includes pixels that are calculated to have a 95% or greater probability of being cloud free. The MODIS LST data are additionally filtered using a cloud cover estimate from the PROMICE AWS measurements (described below) with an exclusion of >0.3 estimated cloud cover fraction as the cutoff.

There are other MODIS LST products including MOD/MYD21 (MxD21) and MOD/MYD29 (MxD29). In contrast to the split window technique implemented in MxD11, the MxD21 product uses a temperature/emissivity separation (TES) technique to calculate surface temperature (Hulley and Hook, 2017), while MxD29 uses an algorithm developed by Key and Haeffliger (1992). Comparisons between the MxD11 and MxD21 product in non-glaciated areas have shown that the MxD21 product more closely matches in situ measurements, potentially due to overestimation of emissivity in the MxD11 product (Yao and others, 2020; Li and others, 2021). A validation comparison between the two products has not been tested in snow-covered regions to our knowledge. The MxD29 product is incorporated into a multilayer regional product MODGRNLD that was developed specifically for the GrIS (Hall and others, 2018; Hall and DiGirolamo, 2019). Both the MxD29 and MxD11 products were compared to the same in situ skin temperature dataset measured at Summit, Greenland in summer 2015; the mean bias for the MxD29 product was 0.98°C while it was −0.4°C for the MxD11 product, and the RMSE was also slightly larger for the MxD29 product (1.3 vs 1.0°C for MxD11) (Adolph and others, 2018; Hall and others, 2018). We decided to use the MxD11 product in this study because the results from Adolph and others (2018) indicated that there was not a cold bias in the MxD11 product in the accumulation zone of the GrIS, but the study was limited in temporal and spatial scope. Comparing the MxD11 product to the large in situ dataset from PROMICE provided an opportunity to expand on the results from Adolph and others (2018). Validating the MxD21 or MODGRNLD LST products, or other higher spatial resolution LST products (e.g. ASTER and Landsat), with in situ measurements from the PROMICE AWS datasets would be a worthwhile endeavor but is beyond the scope of this paper.

2.2 PROMICE data

Skin temperature, near-surface air temperature, specific humidity, albedo, latent and sensible heat flux, incoming and outgoing longwave radiation and estimated cloud cover fraction data were extracted from 2014 to 2017 from the PROMICE database (Ahlstrøm and others, 2008; van As and Fausto, 2011; Fausto and van As, 2019). We used data from 17 automatic weather stations in eight locations on snow and ice mostly in the GrIS ablation zone, with the exception of KAN_U which is in the accumulation zone. Figure 1 shows 2019 satellite imagery of each site (MacGregor and others, 2020) as well as the location of each site. The PROMICE sites (Table 1) used are the Kangerlussuaq upper, middle and lower sites (KAN_U, KAN_M, KAN_L), the Crown Prince Christian Land upper and lower sites (KPC_U, KPC_L), the Nuuk upper and lower sites (NUK_U, NUK_L), the Qassimiut upper and lower sites (QAS_U, QAS_L), the Scoresbysund upper and lower sites (SCO_U, SCO_L), the Tasiilaq alternate and lower sites (TAS_A, TAS_L), the Thule upper and lower sites (THU_U, THU_L) and the Upernavik upper and lower sites (UPE_U, UPE_L). As seen in Figure 1, the complexity of the terrain surrounding each of these sites varies. We excluded PROMICE sites on independent glaciers, sites that do not have coverage for multiple years of the analysis, and sites near the center of the ice sheet to focus on the ablation area.

PROMICE skin temperature (in °C) is derived from incoming and outgoing longwave radiation measurements using the following equation,

$$\text{LST} = ((\text{LR}_{\text{out}} - (1 - \epsilon) \times \text{LR}_{\text{in}}) / (\epsilon \times 5.67 \times 10^{-8}))^{0.25} - 273.15, \quad (2)$$

where LR is longwave radiation (in W m^{-2}), the Stefan–Boltzman constant has units of $\text{W m}^{-2} \text{K}^{-4}$, and ϵ (surface emissivity) is set to 0.97 for consistency with publicly available PROMICE data (Fausto and van As, 2019; Fausto and others, 2021). Surface conditions (and thus surface emissivity) vary in the ablation zone of the GrIS, and implications of the choice of emissivity value are explored in Section 3.6. The four-component radiation measurements are made using a Campbell Scientific net radiometer (model CNR4 or CNR1) manufactured by Kipp & Zonen. The instrument's pyrgeometer measures longwave radiation in the wavelength range from 4.5 to 42 μm with an operating temperature range of −40 to 80°C. The reported maximum uncertainty from the manufacturer is $\pm 10\%$, but in practice the maximum uncertainty in daily total longwave radiation is $\pm 5\%$ (van den Broeke and others, 2004; Fausto and others, 2016). The PROMICE AWSs report values for each variable every 10 min. In implementing Eqn (2), temperatures are frequently calculated to be above 0°C during summer months. However, because the AWS sites are all on snow and ice, the surface cannot be above 0°C. Therefore, temperatures above the freezing point do not make physical sense. These temperatures calculated above zero may be due in part to the error of the longwave measurements, due to the effect of the longwave emissions from the atmosphere between the surface and the sensor, or due to the presence of liquid water on the surface in some instances. Both surface temperature calculations (PROMICE and MODIS) report values above zero. We compare MODIS values to the PROMICE values that are directly calculated from Eqn (2) (called 'unclipped') and values that are corrected such that temperatures above 0°C (from Eqn (2)) are reported as 0°C so that they make physical sense (called 'clipped'). Note that the PROMICE data reported online (<https://www.promice.org/PromiceDataPortal/>) already have this 'clipped' correction.

The emissivity value chosen to calculate PROMICE skin temperatures is fixed at 0.97 (Fausto and others, 2021). Emissivity of the ice-sheet surface varies directionally, with wavelength, and with surface type. Hori and others (2006) report values as high as 0.997 for dendritic snow at nadir angle at a wavelength of 10.5 μm and as low as 0.949 for bare ice at 12.5 μm . At an off-nadir angle of 75°, the emissivity of bare ice was as low as 0.709; however, this angle is larger than we would expect from an AWS measurement. Nielsen-Englyst and others (2019) modeled snow emissivity with the spectral response function for the type of sensor used by PROMICE stations which cover the range from 4 to 40 μm . They reported an emissivity of 0.997, however this was for a 'typical snow surface', and surface type at the PROMICE sites varies throughout the year from snow to bare ice. Results presented throughout the current work use an emissivity of 0.97, but in the Supplementary material we present data showing the sensitivity of our results to varying the emissivity between 0.7 and 0.99 (Appendix A, Table 5).

2.3 Additional environmental variables

To aid in the analysis of the difference between MODIS LST and PROMICE skin temperature measurements, we also used the following variables: solar zenith angle, air temperature, MODIS viewing angle, relative humidity, specific humidity, MODIS emissivity at band 31, MODIS emissivity at band 32, error of MODIS LST,

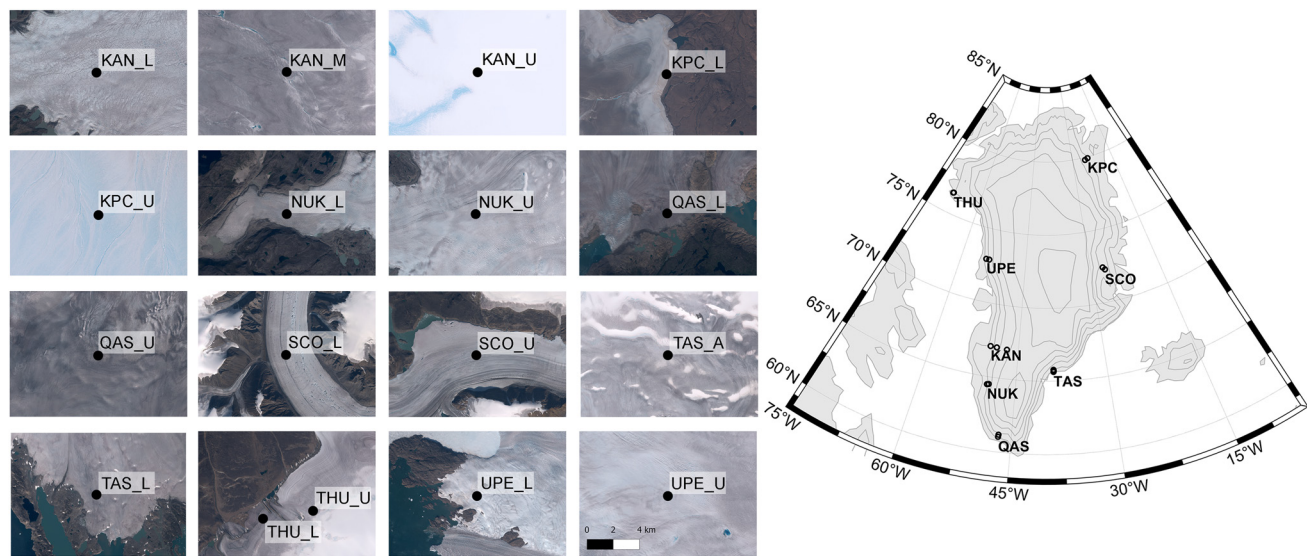


Fig. 1. PROMICE AWS locations. The figure at the left shows 10 m resolution 2019 satellite imagery from Sentinel-2 of each of the PROMICE sites. Imagery is from MacGregor and others (2020), and the figure was created using QGreenland (Moon and others, 2021). The scale for all imagery is shown in the bottom right image. Map at the right shows the location of each PROMICE site, noting that each location has multiple AWS sites at different elevations which sometimes appear as overlapping markers.

Table 1. AWS location information

AWS	Latitude (°N)	Longitude (°W)	Elevation (m)
KPC_L	79.9108	24.0828	370
KPC_U	79.8347	25.1662	870
SCO_L	72.2230	26.8182	460
SCO_U	72.3933	27.2333	970
TAS_L	65.6402	38.8987	250
TAS_A	65.7790	38.8995	890
QAS_L	61.0308	46.8493	280
QAS_U	61.1753	46.8195	900
NUK_L	64.4822	49.5358	530
NUK_U	64.5108	49.2692	1120
KAN_L	67.0955	49.9513	670
KAN_M	67.0670	48.8355	1270
KAN_U	67.0003	47.0253	1840
UPE_L	72.8832	54.2955	220
UPE_U	72.8878	53.5783	940
THU_L	76.3998	68.2665	570
THU_U	76.4197	68.1463	760

PROMICE AWS information within each of the eight locations (Fig. 1).

latitude, longitude, latent heat flux, sensible heat flux, albedo and cloud cover. Latent heat flux, sensible heat flux, relative humidity, specific humidity, albedo and cloud cover were extracted from the PROMICE AWS database. PROMICE calculates specific humidity from the relative humidity and air temperature following Goff and Gratch (1946). Latent and sensible heat fluxes are calculated using the Monin–Obukhov similarity theory based on vertical gradients in wind speed, specific humidity and potential temperature. Albedo is calculated based on incoming and outgoing shortwave radiation measurements that are tilt-corrected (Fausto and others, 2021). Cloud cover fraction is estimated from incoming and outgoing longwave radiation measurements (Ahlström and others, 2008; van As and Fausto, 2011). Solar zenith angle data were extracted from the MODIS MxD11 product. Solar zenith angle indicates angle of the direct incoming radiation and often correlates to the amount of incoming solar radiation. Emissivity at band 31 and band 32, viewing angle and error of MODIS LST were also extracted from the MODIS MxD11 product.

2.4 Comparison of MODIS and PROMICE data

To conduct comparative analyses that validate the swath level product, discrete temperature measurements from the MODIS

and PROMICE datasets must be paired spatially and temporally, described below. We analyzed data for 4 years, 2014–2017, seeking to optimize between time for data extraction and ensuring we used more than a single year of data in case that year was anomalous in some way. While both data products extend temporally before and beyond these years, 4 years of data provide sufficient variability to identify relationships in the comparison between datasets. Because we are not looking to identify temporal trends, adding more data to the analysis was not necessary. In order to pair the temperature measurements, for each of the 17 PROMICE sites, we extracted any MODIS swath LST data within 10 km (calculated using the distance algorithm by Sohrabinia (2021)) that were available over the 4 years of the study. Each MODIS data point is time stamped with a resolution of 5 min, and we paired each with the temporally closest PROMICE measurement, requiring that it be within 30 min of the MODIS measurement. If there was not a match that met this requirement, then the data point was not included. A sensitivity study altering the matching distance radius from 10 km down to 1 km showed that variation in matching radius caused negligible change in the MODIS LST bias statistics (see Appendix Table 6). Within each pair of data points, we subtracted the MODIS LST measurement from the PROMICE skin temperature measurement to calculate the difference between the MODIS and PROMICE measurements, called the ‘MODIS/PROMICE difference’. Positive values of this difference denote a MODIS LST lower than the reference PROMICE skin temperature, and we refer to this as a ‘cold bias’. To determine if a calculated ‘cold bias’ was significantly different from 0°C, we conducted a *t*-test with a 5% significance level.

In addition to analyzing this dataset as a whole, separate analyses were conducted for two periods (1) instances when PROMICE skin temperature was 0°C or higher (referred to throughout as above freezing period) and (2) instances when the PROMICE skin temperature was below 0°C (referred to throughout as below freezing period) to determine if systematic differences exist between these different periods. Because calculated surface temperatures from radiometric measurements are often above zero in the above freezing period, the above freezing period has more possible sources of complication due to temperatures that do not always seem physically reasonable. We believe analyzing these two periods separately gives clearer insight into MODIS LST behavior over the GrIS.

The MODIS/PROMICE difference in the below freezing period was analyzed using robust multiple linear regression using Student's t distributed errors (Lange and others, 1989) via the `mvsregress` function from Piche (2022). Using this model, we investigated the amount of variance in the MODIS/PROMICE difference that can be explained by selected environmental variables. Our initial models explored the influence of day of year, solar zenith angle, MODIS LST, PROMICE air temperature, MODIS viewing angle, relative humidity, specific humidity, air temperature, emissivity at band 31, emissivity at band 32, error of MODIS LST, latitude, longitude, latent heat flux, specific heat flux and albedo on the difference between paired MODIS LST and PROMICE skin temperature. MODIS viewing angle, sensible heat flux, relative humidity, specific humidity, emissivity at band 31, emissivity at band 32, error of MODIS LST, latitude, longitude, albedo and solar zenith angle all had p -values above 0.05 and were removed from later models. The multiple linear regression model assumes that the predictor variables are not highly correlated, and that the residuals express homoscedasticity and are normally distributed. Based on an examination of the residuals using quantile–quantile diagnostic plots, we found that the residuals were symmetric and homogeneous but heavy-tailed. They were fitted well by a Student's t distribution with 6–7 degrees of freedom. To correct our regression analyses for the heavy tails, we used robust multiple regression with Student's t distributed errors (Lange and others, 1989). A benefit of using this distribution class is that the standard F -tests for regression models are unaffected by the Student's t heavy-tailed error distribution (Zellner, 1976; Qin and Wan, 2004). Thus, we can present the familiar F -tests for our regression results. A multiple linear regression analysis was also explored for the above freezing period; however, the residuals for the above freezing period were poorly behaved (non-homogeneous, non-symmetric) so these results are not included in this paper.

The final multiple regression model of the MODIS/PROMICE difference included MODIS LST, specific humidity and latent heat flux as explanatory variables. To determine the relative influence (effect size) of each environmental variable on the modeled difference between MODIS LST and PROMICE skin temperature, we first calculated the overall variance explained by the multiple linear regression model (the adjusted R^2 value). Then the relative explained variance of each variable is calculated as follows. The sum of squared residuals of the i -th variable (SSR_i) is divided by the total sum of squared residuals of the model (SST). This gives the fraction of variance in the response variable (temperature difference) that can be explained by the i -th variable. The fraction of variance for the i -th variable is then divided by the sum of all fractional variances. Finally, this value is multiplied by the adjusted R^2 to give the fraction of the overall variance that is attributed to the i -th variable.

3. Results and discussion

3.1 Comparison of MODIS and PROMICE data

Table 2 shows the number of skin temperature data points from MODIS and PROMICE that are available at each of the 17 AWS sites over the span of the 4 years (2014–2017) of the study, along with the number of paired data points in the two distinct periods: above freezing and below freezing. The number of PROMICE data points varies as sites were established and decommissioned over this 4-year period and due to some temporary outages that occurred during the measurement windows that could not be repaired until the following field season. Overall, data collection of the variables of interest had few outages during the selected years (see Fausto and others, 2019 Appendix C). The

Table 2. Number of data points in the PROMICE skin temperature, MODIS LST, below freezing period paired measurements, above freezing period paired measurements and total paired measurements datasets for each AWS site

AWS	PROMICE	MODIS	Paired (below freezing)	Paired (above freezing)	Paired (total)
KAN_L	13 793	3376	1444	1011	2455
KAN_M	11 369	4236	2124	631	2755
KAN_U	9957	3438	1610	346	1956
KPC_L	12 631	8823	3254	1246	4500
KPC_U	9414	10 769	3146	1210	4356
NUK_L	13 258	2976	1294	947	2241
NUK_U	10 964	3048	1403	615	2018
QAS_L	13 867	2402	1072	539	1611
QAS_U	10 531	3209	1117	605	1722
SCO_L	12 587	5744	2167	1255	3422
SCO_U	15 000	6206	2938	1537	4475
TAS_A	8567	3158	969	752	1721
TAS_L	5906	2615	480	338	818
THU_L	12 922	7224	3698	1253	4951
THU_U	13 060	7559	4099	1058	5157
UPE_L	11 149	5943	2306	1241	3547
UPE_U	32 752	6556	4543	3407	7950

availability of MODIS data fluctuates with cloud presence, and therefore the number of data points varies across sites and periods (see Appendix Fig. 8 for a plot showing when paired MODIS LST and PROMICE skin temperature data were available). Additionally, the number of days in the above freezing or below freezing period also reflects local climate effects. Figure 2 shows the MODIS LST data and PROMICE skin temperature data over the year of 2015 at the THU_U site as a representative sample of the time series seen in other years. The MODIS LST data often fall below the PROMICE skin temperature data in the spring, fall and winter. During the summer (which largely overlaps with the above freezing period), the PROMICE skin temperature data should have an upper boundary at 0°C, as all sensors are placed over ice; however, as mentioned above, calculated PROMICE skin temperatures are sometimes >0°C. In the above freezing period, we observe the MODIS LST data fluctuate both above and below 0°C. In Figure 3, a 1:1 plot of PROMICE skin temperature and MODIS LST is shown for the same THU_U site data in 2015. We note that the data are further from the 1:1 line at lower temperatures, suggesting the presence of a progressive cold bias that will be investigated in our multiple regression analysis. Additionally, the large deviation from the 1:1 plot when temperatures are at or above freezing (see Fig. 3) led us to approach the analysis of temperatures above 0°C by analyzing the 'clipped' and 'unclipped' data as described in Section 2.2.

When MODIS LST is compared to PROMICE skin temperature, we observe an average cold bias of 1.8°C and RMSE of 3.9°C. To put our analysis in context of past validation studies, we also analyzed the difference between MODIS LST and PROMICE air temperature data. When analyzing the aggregated differences between MODIS LST and PROMICE air temperature data, we identified an average cold bias of 4.6°C and RMSE of 5.2°C (Table 3). We observe that the distribution of the difference between MODIS LST and PROMICE air temperature data is centered above the 0°C line indicating the MODIS LST data are colder than the PROMICE air temperature data in the majority (over 50%) of cases (Fig. 4d). The histograms presented in Figure 4 and the data in Table 3 suggest that MODIS LST measurements more accurately match skin temperature measurements than the near-surface air measurements. The same conclusion was reached by Adolph and others (2018), showing that MODIS LST more closely matched in situ skin temperatures than in situ air temperature due to near-surface inversions at Summit, Greenland. Additionally, the calculation of the MODIS

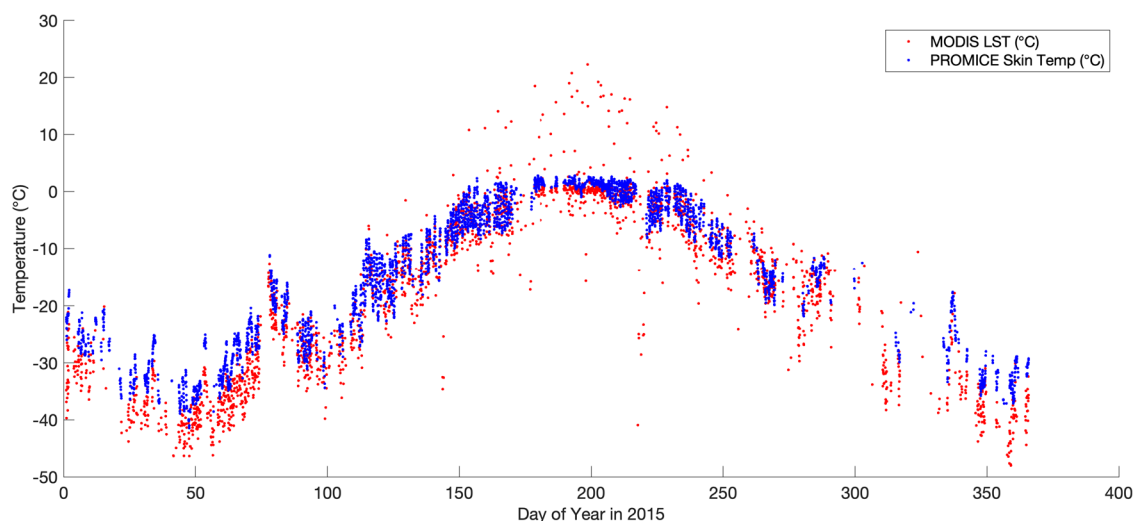


Fig. 2. Time series of PROMICE skin temperatures and MODIS LST at the THU_U site for 2015. MODIS LST is shown in red and PROMICE skin temperature is shown in blue.

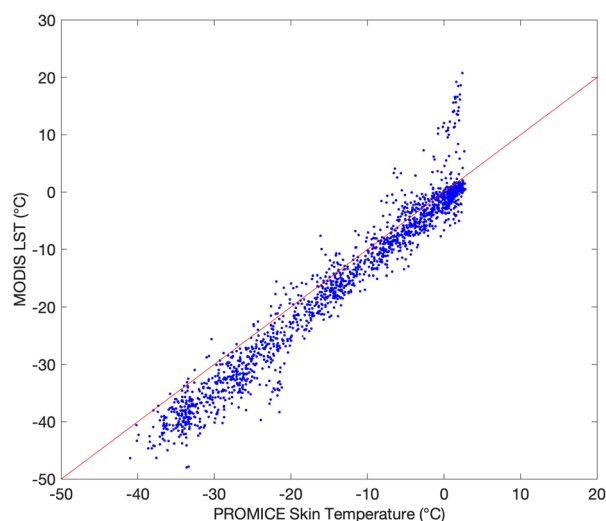


Fig. 3. MODIS LST plotted against PROMICE skin temperature readings for the THU_U site for the year 2015. The red line marks the one-to-one line. Points above the red line indicate instances when MODIS recorded higher temperatures than PROMICE, while points below indicate instances when MODIS recorded lower temperatures than PROMICE.

LST is fundamentally a skin temperature and not a near-surface air temperature, and Guillevic and others (2017) indicate that IR surface temperature measurements should be validated with skin temperature measurements for that reason.

In order to assess if aggregating the data into monthly periods might result in closer alignment between the datasets, we calculated monthly averages of MODIS LST and PROMICE skin temperature. We only included data in the monthly average when there were both MODIS and PROMICE surface temperatures available so as to avoid introducing a bias by eliminating cloudy days from the MODIS dataset but not the PROMICE dataset. We find that monthly aggregation does not eliminate the difference between the MODIS LST and PROMICE skin temperature (see Fig. 5). We do find that there is a larger difference between the monthly averages during the winter months.

3.2 Above freezing period

As described above, the above freezing period analysis includes data points when PROMICE skin temperature is reported as

$\geq 0^{\circ}\text{C}$. To analyze the difference between MODIS LST and PROMICE skin temperature, the data were aggregated across all 17 sites and all 4 years in the above freezing periods. In this analysis, we consider both ‘clipped’ and ‘unclipped’ PROMICE skin temperature datasets, where temperatures calculated above zero are set to zero or they are left as is, respectively. MODIS LST values are also often above the melting point during the summer; therefore, we hypothesized that better agreement between the in situ and remote temperatures would occur if the PROMICE data were unclipped. Because the unclipped PROMICE temperatures were higher than the clipped values (which are fixed at 0°C when the unclipped value is above 0°C), we see an expected cold bias in the above freezing period MODIS data that we do not observe when MODIS LST is compared to clipped PROMICE skin temperature data. Statistics for this comparison are presented in Table 3. Comparing MODIS LST to the unclipped PROMICE skin temperatures, we identified high variability between datasets ($\text{RMSE} = 4.9^{\circ}\text{C}$) and a mean bias of 0.34°C (Table 3). The distribution of the aggregated above freezing data for all years is shown in Figure 4b, and individual histograms for each site do not show a consistent bias or distribution (Appendix A, Fig. 10). When MODIS LST was compared to the clipped PROMICE skin temperature data where values above 0°C are corrected down to 0°C (clipped), we observe a mean bias of -1.0°C and RMSE of 4.9°C .

3.3 Below freezing period

Considering all below freezing period data from all sites, the distribution of difference in the aggregated MODIS LST and PROMICE skin temperature data during the below freezing period is distributed around a point above the zero line (Fig. 4a). Individual histograms for each site consistently show this distribution centered above the zero line (Appendix A, Fig. 9). This supports the existence of a cold bias in the MODIS LST data during the below freezing period, and this cold bias is present across all ablation zones of the GrIS in this study. Analyzing the difference between MODIS LST and PROMICE skin temperature data aggregated across all sites and all years in the below freezing periods, we identified an average cold bias of 2.4°C in the MODIS LST data (Table 3).

We also observed that the paired MODIS LST and PROMICE skin temperatures fall further beneath the one-to-one line at lower

Table 3. MODIS and PROMICE comparison statistics including mean bias, median bias, root mean square error and std dev. for the above freezing period, the below freezing period, the below freezing period when temperatures are above -25°C , the below freezing period when temperatures are below -25°C , and both above freezing and below freezing periods combined

Temperature range	MODIS dataset	PROMICE dataset	Mean bias ($^{\circ}\text{C}$)	Standard error ($^{\circ}\text{C}$)	Median bias ($^{\circ}\text{C}$)	RMSE ($^{\circ}\text{C}$)	Std dev. ($^{\circ}\text{C}$)	Observations
$-25^{\circ}\text{C} < x < 0^{\circ}\text{C}$	-	-	2.1	0.01	1.8	2.6	3.3	35 154
$-25^{\circ}\text{C} > x$	-	-	4.6	0.1	4.3	3.5	5.8	8659
Below 0°C	-	-	2.4	0.01	2.0	3.2	4.0	37 025
Above 0°C	Unclipped	Clipped	-1.0	-0.01	-0.1	4.9	5.0	16 045
Above 0°C	Unclipped	Unclipped	0.3	0.002	1.3	4.9	4.9	16 045
All	Unclipped	Clipped	1.4	0.01	1.2	4.1	4.3	53 070
All	Unclipped	Unclipped	1.8	0.01	1.7	3.9	4.3	53 070
All	Clipped	Clipped	2.3	0.01	1.7	3.0	3.8	44 409
All	Unclipped	Air temp	4.6	0.02	3.9	5.2	5.2	80 502

These statistics are reported for both unaltered MODIS LST data ('Unclipped') and MODIS LST data with values above 0°C set to 0°C ('Clipped') compared to both PROMICE skin temperature data with values above 0°C set to 0°C ('Clipped') and uncorrected PROMICE skin temperature data ('Unclipped'). All mean biases are significantly different from 0°C at a 5% significance level.

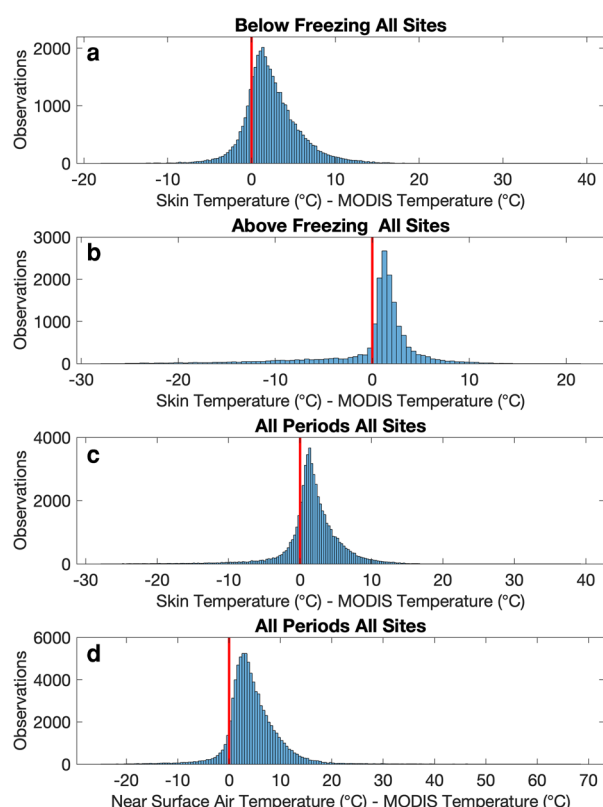


Fig. 4. Histograms of the MODIS difference calculated as PROMICE temperature minus MODIS LST at all 17 sites for years 2014–2017. Zero is indicated by a red vertical line. Figures (a–c) compare MODIS LST to PROMICE skin temperature in the (a) below freezing period, (b) above freezing period and (c) all data. Figure (d) compares MODIS LST with PROMICE 2 m air temperature.

temperatures in Figure 3, suggesting a progressive cold bias in the MODIS LST data. We find that when the MODIS LST temperature is between 0 and -25°C , the average bias is 2.1°C , and when the MODIS LST temperature is below -25°C , an average cold bias of 4.6°C in the MODIS LST data exists (Table 3). This clearly indicates a cold bias is present in MODIS LST data during the below freezing period that is larger at lower temperatures.

3.4 Effect of environmental variables on the bias

The multiple linear regression analysis showed that MODIS LST, specific humidity and latent heat flux can explain variability in the MODIS/PROMICE difference. Figure 6 shows what percent of variance in the MODIS/PROMICE difference is explained by each of the explanatory variables at a given site from all years

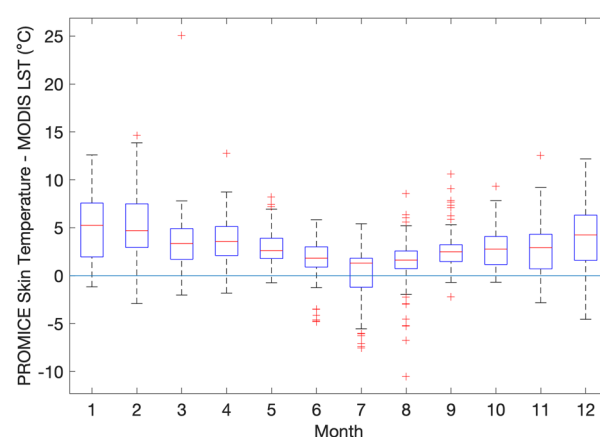


Fig. 5. Temperature difference between monthly average MODIS LST and PROMICE skin temperature shown for each month of the year. Each boxplot shows the median for all 17 sites over the 4 years in a given month, with bounds showing the 75% quartile and 25% quartile ranges. A horizontal line is included at 0°C that would represent an exact match between the monthly average MODIS LST and PROMICE skin temperature. There is more agreement between the datasets during summer months than the rest of the year. The mean difference when all months are grouped together is 2.94°C , indicating MODIS LST monthly averages are lower than PROMICE monthly averages.

of data (see Appendix A, Table 4 for maximum, minimum and average percentages of variance explained by each variable). These relationships are significant with p -values < 0.001 meaning that each of the predictor variables in the model is significant in improving the model performance; there is a significant relationship between predictor variables and temperature difference, with the exception of specific humidity at KPC_U which is not significant. Our analysis of the variance explained by each variable showed that the explanatory variables (MODIS LST, specific humidity and latent heat flux) can collectively explain 7.7–70.9% of variation in the MODIS/PROMICE difference in the below freezing period. As shown in Figure 6, temperature typically explains the most variance followed by specific humidity at most sites. The variation in ability to explain the bias means that at some sites the difference between MODIS LST and PROMICE skin temperature is not correlated with our explanatory variables. Less variance can be explained at sites KAN_M and KPC_U compared to other sites. This same multiple linear regression with Student's t distribution was conducted with all data (both above freezing and below freezing periods) and is presented in Appendix A, Figure 11.

Our multiple regression analysis of the difference between MODIS LST and PROMICE skin temperature data shows the cold bias becomes stronger at lower temperatures, higher specific humidity and when latent heat flux is negative, drawing energy

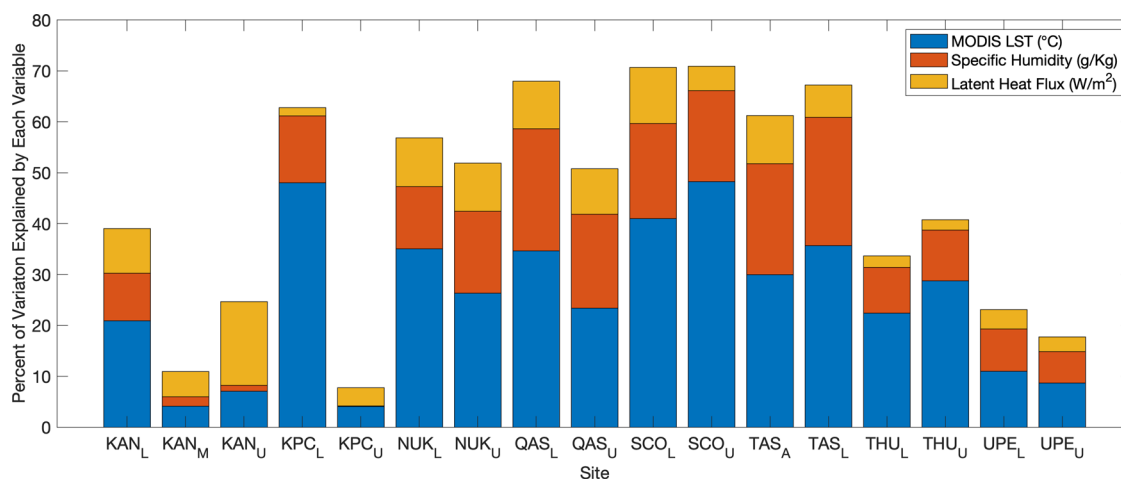


Fig. 6. Stacked histograms showing percentage of variance in the MODIS/PROMICE difference explained by MODIS LST, specific humidity and latent heat flux in our multiple regression model for all 17 sites for the below freezing period.

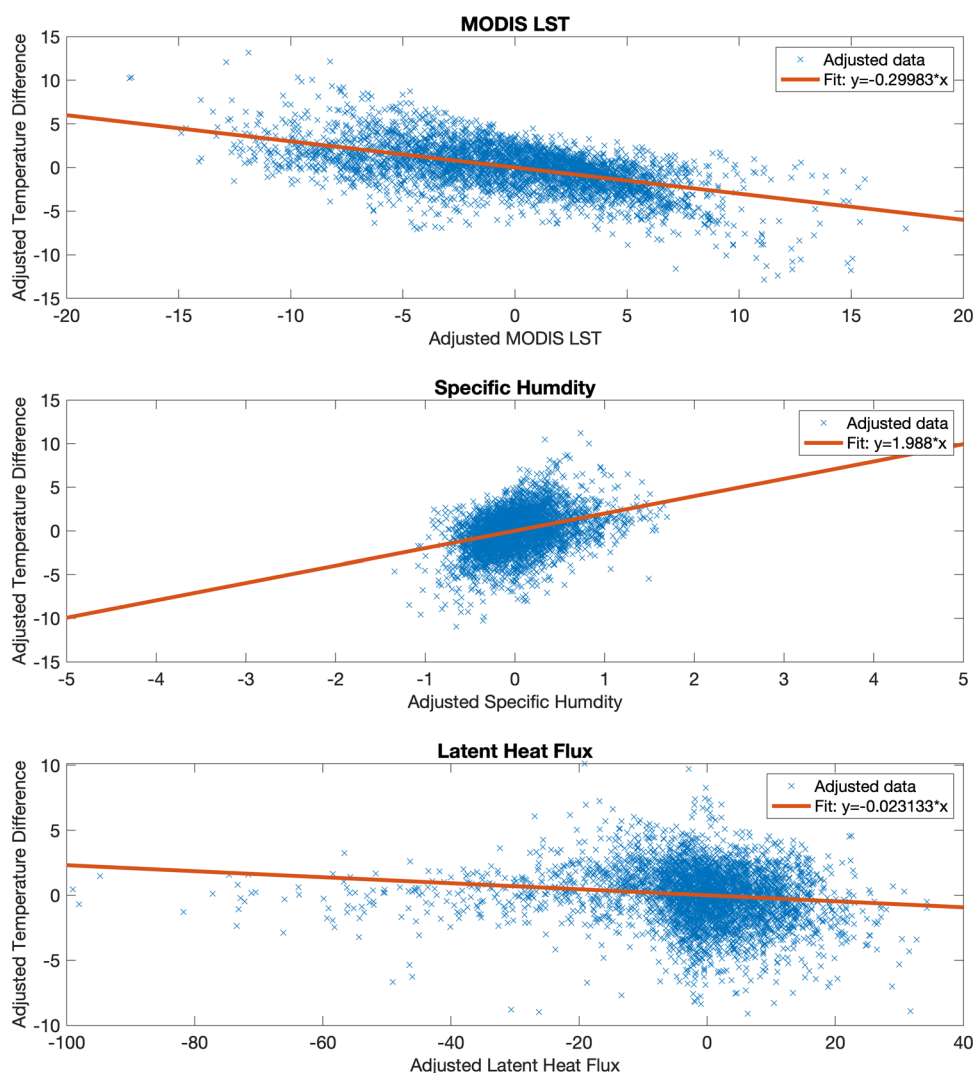


Fig. 7. Leverage plots from our multiple linear regression analysis of the MODIS/PROMICE difference at the THU_U site for the below freezing period for all years. In each plot 'adjusted temperature difference' represents the residuals of MODIS/PROMICE temperature difference regressed against the two predictor variables not included in the plot (i.e. in the MODIS LST plot, adjusted temperature difference is the residuals of the MODIS/PROMICE difference regressed against specific humidity and latent heat flux). The adjusted MODIS LST is the residual of MODIS LST regressed against specific humidity and latent heat flux, adjusted specific humidity is the residual of specific humidity regressed against MODIS LST and latent heat flux, and adjusted latent heat flux is the residual of latent heat flux regressed against MODIS LST and specific humidity.

from the surface (Fig. 7). Representative leverage plots (e.g. Sall, 1990) resulting from the multiple linear regression analysis from a single site are shown in Figure 7, indicating that the MODIS LST cold bias is larger at lower temperatures and higher specific humidity. Data presented in Figure 7 are from the THU_U site, as a representative example of leverage plots seen at the array of sites. In the below freezing period, we found regression coefficients of $-0.4^{\circ}\text{C }^{\circ}\text{C}^{-1} \pm 0.2$ associated with MODIS LST, $2.9^{\circ}\text{C }(\text{g kg}^{-1})^{-1} \pm 1.6$ associated with specific humidity and $-0.05^{\circ}\text{C }(\text{W m}^{-2})^{-1}$ associated with latent heat flux averaged across sites. The fact that MODIS LST has a larger difference with in situ skin temperature at lower temperatures affirms the progressive cold bias in MODIS data over Greenland found in Shuman and others (2014).

Studies have shown MODIS algorithms can fail to identify cloud cover, particularly in polar regions (Ackerman and others, 2008; Williamson and others, 2013), and failure to effectively screen out all cloudy pixels can contaminate MODIS LST data (Westermann and others, 2012; Østby and others, 2014). It is possible that there is some undetected cloud contamination (or fog at the surface level) that is correlated with low MODIS LST and high specific humidity that explains some of the MODIS LST bias. The relationship between specific humidity and the magnitude of the MODIS LST bias may be indicative of issues with the implementation of corrections for atmospheric water vapor content, a variable in the MxD11 LST algorithm. Malakar and Hulley (2016) present a water vapor scaling model which improves the atmospheric correction in thermal infrared bands and therefore can improve LST calculations that implement the TES algorithm. Because the MxD11 product uses the split window technique and a fixed lookup table for emissivity, some of the variability in emissivity may be lost and that could be correlated with instances of high specific humidity.

Conversely, it is possible that the MODIS/PROMICE difference is larger at higher specific humidities because of increased likelihood of hoar frost growing on the PROMICE AWS sensors, causing the PROMICE skin temperature data to become inaccurate. This measurement issue could create a false cold bias in the MODIS LST data when conditions are favorable for frost growth. We filter out data where relative humidity is above 99% to remove data suspected of frost contamination to mitigate this potential impact, but it is an imperfect attempt to eliminate this issue.

During the below freezing period, the latent heat flux would be a result of vapor transport through sublimation or condensation. Positive fluxes are indicative of deposition or condensation in the form of surface hoar growth or riming of the snow surface. Negative fluxes would result from sublimation, though at low temperatures, the capacity of air to hold water vapor is low (Albert and Hawley, 2000). The slope of the correlation between the latent heat flux and the MODIS/PROMICE difference indicates that when latent heat flux is negative, there is a larger temperature difference. A potential mechanism to account for this effect is unclear.

3.5 Comparison to previous MODIS validation studies

Several prior studies have compared near-surface air temperature to MODIS LST in an effort to validate MODIS LST products when air temperatures were the only available in situ measurements. These studies found that MODIS LST was often lower than the near-surface air temperature (Koenig and Hall, 2010; Østby and others, 2014; Shuman and others, 2014; Williamson and others, 2017); however, near-surface inversions may explain some of the identified MODIS LST cold bias. Adolph and others (2018) did not identify a cold bias in the MxD11 LST product when compared to in situ skin temperature measurements at

the Summit station on the GrIS on 8 June to 18 July 2015. Using the same MODIS product, in the below freezing period, our analysis clearly identifies a cold bias in the MODIS LST data corroborating Koenig and Hall (2010), Østby and others (2014), Shuman and others (2014) and Williamson and others (2017). The RMSE calculated when considering all data in our analysis is 3.9°C , which is also comparable to RMSE values found in prior studies that did use skin for comparison; RMSE = 3.0°C in Østby and others (2014), and RMSE = 3.1°C in Koenig and Hall (2010). These studies had fewer data points and were more limited in spatial and temporal extent. Our multiple linear regression findings support a progressive cold bias, as proposed by Shuman and others (2014). This cold bias that is more pronounced at low temperatures could be related to issues identified in MODIS instrument calibration at temperatures below -20°C due to the MODIS sensors long surpassing their design lifetime of 6 years (Wenny and others, 2012; Xiaoxiong and others, 2015).

When MxD11_L2 Collection 6 LST data are used, accounting for this progressive cold bias will be important to ensure that any spatial or temporal trends that are identified are still present when this source of error is considered alongside other sources of error or uncertainty in the analyses.

3.6 Uncertainty in measurements and analysis

There are several spatial and temporal differences between the MODIS LST measurements and the PROMICE in situ skin temperature measurements. The MxD11 LST product averages data over a 1 by 1 km area while the PROMICE in situ AWSs collect data from an area on the scale of a meter. Additionally, the MODIS LST measurements can occur up to 10 km and 30 min away from the corresponding PROMICE skin measurement. These differences and the inherent error of both measurements certainly will cause some differences between any two paired measurements; however, we do not anticipate that this introduces systematic bias for the following reasons. First, we are looking at the average difference between the PROMICE and MODIS measurements, and we have typically 1000–4000 data points at each site in each period. Second, there is no embedded systematic directional (north/south, east/west) bias in the MODIS versus PROMICE data points; the closest (temporally and spatially) non-missing MODIS data point was chosen. There is also no embedded systematic temporal offset, either MODIS or PROMICE data may lead or lag the other, so temperature variability on the timescale of <1 h should not lead to a directional offset. These chosen methods follow precedent from prior studies comparing MODIS and in situ measurements (e.g. Westermann and others, 2012; Adolph and others, 2018). Since several of the AWS are within 10 km of the edge of the ice sheet, it is probable that some reported MODIS LSTs are for non-snow/ice surfaces, particularly in the above freezing period comparison. Additionally, terrain complexity surrounding the sites varies (see Fig. 1) and could be a source of error in our analyses. We conducted a sensitivity study of the ‘pairing’ radius, altering it from 10 km down to 1 km so that the MODIS pixel was in closer proximity to the PROMICE AWS. At the 1 km distance, all MODIS LSTs would be on the ice. Constricting the paired radius did reduce the mean bias slightly, but also had the effect of increasing the RMSE (see Appendix Table 6). Therefore, while the choice of a 10 km radius cutoff may introduce some error to the analysis, it does not affect the overall conclusions.

Since MODIS can only calculate LST in clear skies, average MODIS LST is lower than average in situ temperatures over a given time period (Westermann and others, 2012), though techniques have been developed to interpolate near-surface

temperatures that are missing due to cloud cover, and more research on the most effective aggregation methods is needed (Williamson and others, 2014). Because we compare data at a given point in time, this issue does not directly affect any given comparison point within the dataset under consideration; however, it is important to consider the combined implications of this known clear sky cold bias with the progressive cold bias presented in this work. Given the observed progressive nature of the MODIS cold bias, the clear sky bias would amplify the average cold bias of the MODIS LST data. Additionally, any misidentified cloudy days would introduce inaccuracy into the MODIS LST data, most often resulting in temperatures lower than ground-based temperatures. Westermann and others (2012) were able to identify an average MODIS cold bias effect from misidentified cloudy days, and Adolph and others (2018) were able to reduce the MODIS cold bias with additional cloud screening. The atmosphere between the radiation sensor and the ice/snow surface can also influence the outgoing longwave measurement, especially when the air temperature is high (above 5°C) modulating outgoing radiation measurement as a combination of both surface radiation and radiation from the warm air (Giesen and others, 2014; Fausto and others, 2016). This might lead to increased error in PROMICE skin temperature measurements during the above freezing period, but is unlikely to affect results in the below freezing period.

Using the reported error of $\pm 5\%$ for the PROMICE AWS longwave radiation measurements from the pyrometers, we propagated error of the calculated PROMICE skin temperature using Eqn (2) and determined an average uncertainty in the PROMICE skin temperature of $\pm 3.4^\circ\text{C}$. This is a conservative estimate of the error, and we anticipate that actual error is smaller. Using the error reported in the MxD11 product, we calculated MODIS LST data to have an average error of $\pm 0.2^\circ\text{C}$. As a result, the combined potential measurement error in the MODIS/PROMICE difference is 3.41°C , which is an upper bound for the magnitude of this error. The average bias when the MODIS temperatures are below -25°C is 4.6°C , indicating a cold bias that is certainly beyond potential measurement errors. The systematic nature of the cold bias that increases as temperature decreases in the below freezing period is also unexplained by the measurement uncertainty. Additionally, the average MODIS/PROMICE difference in the above freezing, below freezing and all periods was all found to be significantly different from 0°C at a 5% significance level using a one-sided *t*-test.

Because emissivity is known to vary with surface type, we tested the effect of altering the emissivity of the surface in the calculations of PROMICE skin temperature. We tested emissivity values ranging from 0.7 to 1, and all results presented in the manuscript implement a PROMICE emissivity of 0.97. We particularly expected that when separating the data into the above freezing period and the below freezing period, we might see the effect of the changing emissivity. When emissivity is decreased to 0.7, the mean difference, median difference and RMSE all increase, and when the emissivity is increased up to 1, these metrics all decrease. The only exception is that the RMSE in the above freezing period does not decrease due to the increase in emissivity within the range from 0.95 to 1. This is likely because the surface type (and thus emissivity) is most variable in the above freezing period, so all emissivities within that range provide similar results. While reported statistics do change slightly as we alter the emissivity, the progressive cold bias is evident for all emissivities tested (Appendix A, Table 5 and Fig. 12). To investigate if using Eqn (2) to calculate surface temperature was a source of error, we directly compared MODIS LST to upwelling longwave radiation yielding a similar R^2 value (0.92) to the direct comparison between MODIS LST and PROMICE skin temperature ($R^2 =$

0.93, see Appendix A, Fig. 13). The use of look up values for emissivities implemented in the MxD11 product is a known limitation of the product (Malakar and Hulley, 2016). This issue did not result in a cold bias in the validation by Adolph and others (2018) where the MxD11 product was used at Summit, Greenland. It is possible that emissivity variability was not a factor in that study because there was significantly less variability in surface type due to a shorter time window and generally less surface variability in the accumulation zone as compared to the ablation zone.

Other potential sources of uncertainty in the PROMICE dataset include the possibility of liquid water at the surface during warm periods, surface roughness and potential off-nadir measurements due to sloping of the surface or tilt in the instruments. The tilt of the weather station varies and the maximum tilt recorded at each site and each year is reported in Appendix A, Table 7. The maximum tilt varies from 2.8 to 26.3° in one year at one site in the most extreme case. The tilt of the AWS will change the field of view of the instrument. Shortwave radiation data and calculated albedo are corrected to account for this tilt; however, other measurements are not corrected.

4. Conclusion

Our findings suggest that a progressive cold bias is present in the MxD11 LST product when temperatures are below freezing across the ablation zone of the GrIS. An individual MODIS LST data point could be above or below the true skin temperature, but on average researchers should expect a cold bias in the MODIS LST data over the GrIS, and studies utilizing this MODIS product should incorporate this error into their analyses. In the below freezing period, we found an average cold bias of $2.4 \pm 0.01^\circ\text{C}$ (mean \pm standard error) in the MODIS LST data. For MODIS LST between 0 and -25°C , we found an average cold bias of $2.1 \pm 0.01^\circ\text{C}$, and for MODIS LST below -25°C , we found an average cold bias of $4.6 \pm 0.1^\circ\text{C}$. In the above freezing period, we observed an average cold bias of $0.3 \pm 0.002^\circ\text{C}$ in the MODIS LST data. More detailed study of the above freezing period is needed as the behavior of temperature data in the above freezing period is inconsistent in this study and ice surface temperature measurements above 0°C likely do not reflect the true ice surface temperature. Aside from temperature, specific humidity and latent heat flux are the main environmental variables that correlate with the MODIS LST bias. This conclusion supports ongoing work to improve MODIS LST algorithms, with a particular focus on atmospheric water vapor corrections and low temperature calibration (Xiaoxiong and others, 2015; Malakar and Hulley, 2016). However, it is also possible that higher specific humidities increase the likelihood of hoar frost growing on the PROMICE AWS sensors causing the PROMICE skin temperature data to become inaccurate creating a false cold bias in the MODIS LST data. While remote AWS stations provide invaluable data for monitoring climate, similar studies conducted at continuously-maintained stations would allow us to ensure instruments are operating properly and to rectify measurement issues quickly. Such validation studies would be beneficial in definitely determining the existence and extent of the MODIS LST cold bias.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/jog.2022.51>.

Data availability. AWS data from the Greenland Ice Sheet, including skin and 2 m air temperature and humidity, were acquired from the Programme for Monitoring the Greenland Ice Sheet (<https://doi.org/10.22008/promice/data/aws>). MODIS surface temperature data were acquired from NASA LP DAAC (https://doi.org/10.5067/MODIS/MYD11_L2.006).

Acknowledgements. We are grateful to Caitlin Glennon, Trevor Stewart, and Elizabeth Eli Holmes for their contributions to this project. We thank the St. Olaf College Collaborative Undergraduate Research and Inquiry Program and the Adam S. Thomas Endowment Fund for supporting this work. Data from the Programme for Monitoring the Greenland Ice Sheet (PROMICE) and the Greenland Analogue Project (GAP) were provided by the Geological Survey of Denmark and Greenland (GEUS) at <http://www.promice.dk>.

Conflict of interest. None.

References

- Ackerman SA and 5 others (2008) Cloud detection with MODIS. Part II: validation. *Journal of Atmospheric and Oceanic Technology* **25**(7), 1073–1086. doi: [10.1175/2007JTECHA1053.1](https://doi.org/10.1175/2007JTECHA1053.1)
- Adolph AC, Albert MR and Hall DK (2018) Near-surface temperature inversion during summer at Summit, Greenland, and its relation to MODIS-derived surface temperatures. *The Cryosphere* **12**(3), 907–920. doi: [0.5194/tc-12-907-2018](https://doi.org/10.5194/tc-12-907-2018)
- Ahlström AP and PROMICE project team (2008) A new programme for monitoring the mass loss of the Greenland ice sheet. *Geological Survey of Denmark and Greenland (GEUS) Bulletin* **15**, 61–64. doi: [10.34194/geusb.v15.5045](https://doi.org/10.34194/geusb.v15.5045)
- Albert MR and Hawley RL (2000) Seasonal differences in surface energy exchange and accumulation at Summit, Greenland. *Annals of Glaciology* **31**, 387–390. doi: [10.3189/172756400781820129](https://doi.org/10.3189/172756400781820129)
- Box JE and 5 others (2012) Greenland ice sheet albedo feedback: thermodynamics and atmospheric drivers. *The Cryosphere* **6**(4), 821–839. doi: [10.5194/tc-6-821-2012](https://doi.org/10.5194/tc-6-821-2012)
- Colosio P, Tedesco M, Ranzi R and Fettweis X (2021) Surface melting over the Greenland ice sheet derived from enhanced resolution passive microwave brightness temperatures (1979–2019). *The Cryosphere* **15**(6), 2623–2646. doi: [10.5194/tc-15-2623-2021](https://doi.org/10.5194/tc-15-2623-2021)
- Déry SJ and Brown RD (2007) Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback. *Geophysical Research Letters* **34**(22), L22504. doi: [10.1029/2007GL031474](https://doi.org/10.1029/2007GL031474)
- Enderlin EM and 5 others (2014) An improved mass budget for the Greenland ice sheet. *Geophysical Research Letters* **41**(3), 866–872. doi: [10.1002/2013GL059010](https://doi.org/10.1002/2013GL059010)
- Fausto RS, van As D, Box JE, Colgan W and Langen PL (2016) Quantifying the surface energy fluxes in South Greenland during the 2012 high melt episodes using in-situ observations. *Frontiers in Earth Science* **4**(82), 1–9. doi: [10.3389/feart.2016.00082](https://doi.org/10.3389/feart.2016.00082)
- Fausto RS and van As D (2019) Programme for monitoring of the Greenland ice sheet (PROMICE): automatic weather station data. Version: v03, Dataset published via Geological Survey of Denmark and Greenland. doi: [10.22008/promice/data/aws](https://doi.org/10.22008/promice/data/aws)
- Fausto RS and 29 others (2019) The Programme for Monitoring of the Greenland Ice Sheet: PROMICE Annual Report 2019. Geological Survey of Denmark and Greenland. Available at https://promice.org/PromiceDataPortal/api/download/92acb911-e952-4128-9030-ce34ab284715/PROMICE_annual_report_2019.pdf
- Fausto RS and 16 others (2021) Programme for Monitoring of the Greenland Ice Sheet (PROMICE) automatic weather station data. *Earth System Science Data* **13**(8), 3819–3845. doi: [10.5194/essd-13-3819-2021](https://doi.org/10.5194/essd-13-3819-2021)
- Fréville H and 7 others (2014) Using MODIS land surface temperatures and the Crocus snow model to understand the warm bias of ERA-Interim reanalyses at the surface in Antarctica. *The Cryosphere* **8**(4), 1361–1373. doi: [10.5194/tc-8-1361-2014](https://doi.org/10.5194/tc-8-1361-2014)
- Giesen RH, Andreassen LM, Oerlemans J and van den Broeke MR (2014) Surface energy balance in the ablation zone of Langfjordjøkelen, an Arctic, maritime glacier in northern Norway. *Journal of Glaciology* **60** (219), 57–70. doi: [10.3189/2014JGJ13J063](https://doi.org/10.3189/2014JGJ13J063)
- Goff JA and Gratch S (1946) Low-pressure properties of water from –160 to 212°F. *Transactions of the American Society of Heating and Ventilating Engineers* **52**, 95–122.
- Graversen RG, Mauritsen T, Tjernström M, Källén E and Svensson G (2008) Vertical structure of recent Arctic warming. *Nature* **451**(7174), 53–56. doi: [10.1038/nature06502](https://doi.org/10.1038/nature06502)
- Guillevic P and 11 others (2017) Land Surface Temperature Product Validation Best Practice Protocol. Version 1.0. In Guillevic P Götsche F Nickeson J and Román M eds. *Best Practice for Satellite-Derived Land Product Validation*. Land Product Validation Subgroup (WGCV/CEOS), p. 60. doi: [10.5067/doc/ceoswgcv/lpv/lst.001](https://doi.org/10.5067/doc/ceoswgcv/lpv/lst.001)
- Hachem S, Duguay CR and Allard M (2012) Comparison of MODIS-derived land surface temperatures with ground surface and air temperature measurements in continuous permafrost terrain. *The Cryosphere* **6**(1), 51–69. doi: [10.5194/tc-6-51-2012](https://doi.org/10.5194/tc-6-51-2012)
- Hall D and DiGirolamo N (2019) Multilayer Greenland ice surface temperature, surface albedo, and water vapor from MODIS, Version 1. doi: [10.5067/7THUWT9NMPDK](https://doi.org/10.5067/7THUWT9NMPDK)
- Hall DK, Key JR, Casey KA, Riggs GA and Cavalieri DJ (2004) Sea ice surface temperature product from MODIS. *IEEE Transactions on Geoscience and Remote Sensing* **42**, 1076–1087. doi: [10.1109/TGRS.2004.825587](https://doi.org/10.1109/TGRS.2004.825587)
- Hall DK and 5 others (2008) Comparison of satellite-derived and in-situ observations of ice and snow surface temperatures over Greenland. *Remote Sensing of Environment* **112**(10), 3739–3749. doi: [10.1016/j.rse.2008.05.007](https://doi.org/10.1016/j.rse.2008.05.007)
- Hall DK and 5 others (2013) Variability in the surface temperature and melt extent of the Greenland ice sheet from MODIS. *Geophysical Research Letters* **40**(10), 2114–2120. doi: [10.1002/grl.50240](https://doi.org/10.1002/grl.50240)
- Hall DK and 5 others (2018) A multilayer surface temperature, surface albedo, and water vapor product of Greenland from MODIS. *Remote Sensing* **10**(4), 555. doi: [10.3390/rs10040555](https://doi.org/10.3390/rs10040555)
- Hanna E and 8 others (2021) Greenland surface air temperature changes from 1981 to 2019 and implications for ice-sheet melt and mass-balance change. *International Journal of Climatology* **41**, E1336–E1352. doi: [10.1002/joc.6771](https://doi.org/10.1002/joc.6771)
- Hori M and 10 others (2006) In-situ measured spectral directional emissivity of snow and ice in the 8–14 µm atmospheric window. *Remote Sensing of Environment* **100**(4), 486–502. doi: [10.1016/j.rse.2005.11.001](https://doi.org/10.1016/j.rse.2005.11.001)
- Hulley G and Hook S (2017) MOD21 MODIS/Terra land surface temperature/3-band emissivity 5-min L2 1 km V006. [Data set]. NASA EOSDIS Land Processes DAAC. doi: [10.5067/MODIS/MOD21.006](https://doi.org/10.5067/MODIS/MOD21.006)
- Jin M (2000) Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 2. Cloudy-pixel treatment. *Journal of Geophysical Research: Atmospheres* **105**(D3), 4061–4076. doi: [10.1029/1999JD901088](https://doi.org/10.1029/1999JD901088)
- Jin M and Dickinson RE (2000) A generalized algorithm for retrieving cloudy sky skin temperature from satellite thermal infrared radiances. *Journal of Geophysical Research: Atmospheres* **105**(D22), 27037–27047. doi: [10.1029/2000JD900318](https://doi.org/10.1029/2000JD900318)
- Jin M and Treadon RE (2003) Correcting the orbit drift effect on AVHRR land surface skin temperature measurements. *International Journal of Remote Sensing* **24**(22), 4543–4558. doi: [10.1080/0143116031000095943](https://doi.org/10.1080/0143116031000095943)
- Key J and Haeffliger M (1992) Arctic ice surface temperature retrieval from AVHRR thermal channels. *Journal of Geophysical Research: Atmospheres* **97**(D5), 5885–5893. doi: [10.1029/92JD00348](https://doi.org/10.1029/92JD00348)
- Kindstedt I and 6 others (2021) Evaluating sources of an apparent cold bias in MODIS land surface temperatures in the St. Elias Mountains, Yukon, Canada. *The Cryosphere Discussions*, 1–24. doi: [10.5194/tc-2021-211](https://doi.org/10.5194/tc-2021-211)
- Koenig LS and Hall DK (2010) Comparison of satellite, thermochron and air temperatures at Summit, Greenland, during the winter of 2008/09. *Journal of Glaciology* **56**(198), 735–741. doi: [10.3189/002214310793146269](https://doi.org/10.3189/002214310793146269)
- Lange KL, Little RJA and Taylor JMG (1989) Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881–896. doi: [10.2307/2290063](https://doi.org/10.2307/2290063)
- Li H and 8 others (2021) Temperature-based and radiance-based validation of the collection 6 MYD11 and MYD21 land surface temperature products over barren surfaces in Northwestern China. *IEEE Transactions on Geoscience and Remote Sensing* **59**(2), 1794–1807. doi: [10.1109/TGRS.2020.2998945](https://doi.org/10.1109/TGRS.2020.2998945)
- MacGregor JA and 5 others (2020) The age of surface-exposed ice along the northern margin of the Greenland Ice Sheet. *Journal of Glaciology* **66**(258), 667–684. doi: [10.1017/jog.2020.62](https://doi.org/10.1017/jog.2020.62)
- Malakar NK and Hulley GC (2016) A water vapor scaling model for improved land surface temperature and emissivity separation of MODIS thermal infrared data. *Remote Sensing of Environment* **182**, 252–264. doi: [10.1016/j.rse.2016.04.023](https://doi.org/10.1016/j.rse.2016.04.023)
- Moon T, Fisher M, Harden L and Stafford T (2021) QGreenland (v1.0.1). Available at <https://doi.org/10.5281/zenodo.4558266>
- Mouginot J and 8 others (2019) Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. *Proceedings of the National Academy of Sciences* **116**(19), 9239–9244. doi: [10.1073/pnas.1904242116](https://doi.org/10.1073/pnas.1904242116)

- Navari M and 5 others (2016) Feasibility of improving a priori regional climate model estimates of Greenland ice sheet surface mass loss through assimilation of measured ice surface temperatures. *The Cryosphere* **10**(1), 103–120. doi: [10.5194/tc-10-103-2016](https://doi.org/10.5194/tc-10-103-2016)
- Nielsen-Englyst P and 5 others (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. doi: [10.5194/tc-13-1005-2019](https://doi.org/10.5194/tc-13-1005-2019)
- Østby TI, Schuler TV and Westermann S (2014) Severe cloud contamination of MODIS land surface temperatures over an Arctic ice cap, Svalbard. *Remote Sensing of Environment* **142**, 95–102. doi: [10.1016/j.rse.2013.11.005](https://doi.org/10.1016/j.rse.2013.11.005)
- Overland J and 9 others (2019) The urgency of Arctic change. *Polar Science* **21**, 6–13. doi: [10.1016/j.polar.2018.11.008](https://doi.org/10.1016/j.polar.2018.11.008)
- Piche R (2022) mvsregress. Available at <https://www.mathworks.com/matlab-central/fileexchange/38212-mvsregress>.
- Qie Y, Wang N, Wu Y and Chen A (2020) Variations in winter surface temperature of the Purog Kangri Ice Field, Qinghai–Tibetan Plateau, 2001–2018, using MODIS data. *Remote Sensing* **12**(7), 1133. doi: [10.3390/rs12071133](https://doi.org/10.3390/rs12071133)
- Qin H and Wan ATK (2004) On the properties of the t - and F -ratios linear regressions with errors. *Econometric Theory* **20**(4), 690–700. doi: [10.1017/S0266466604204042](https://doi.org/10.1017/S0266466604204042)
- Sall J (1990) Leverage plots for general linear hypotheses. *The American Statistician* **44**(4), 308–315. doi: [10.2307/2684358](https://doi.org/10.2307/2684358)
- Serreze MC and Barry RG (2011) Processes and impacts of Arctic amplification: a research synthesis. *Global and Planetary Change* **77**(1), 85–96. doi: [10.1016/j.gloplacha.2011.03.004](https://doi.org/10.1016/j.gloplacha.2011.03.004)
- Shamir E and Georgakakos KP (2014) MODIS land surface temperature as an index of surface air temperature for operational snowpack estimation. *Remote Sensing of Environment* **152**, 83–98. doi: [10.1016/j.rse.2014.06.001](https://doi.org/10.1016/j.rse.2014.06.001)
- Shepherd A and 88 others (2020) Mass balance of the Greenland Ice Sheet from 1992 to 2018. *Nature* **579**(7798), 233–239. doi: [10.1038/s41586-019-1855-2](https://doi.org/10.1038/s41586-019-1855-2)
- Shuman CA, Steffen K, Box JE and Stearns CR (2001) A dozen years of temperature observations at the Summit: Central Greenland automatic weather stations 1987–99. *Journal of Applied Meteorology and Climatology* **40**(4), 741–752. doi: [10.1175/1520-0450\(2001\)040<0741:ADYOTO>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0741:ADYOTO>2.0.CO;2)
- Shuman CA, Hall DK, DiGirolamo NE, Mefford TK and Schnaubelt MJ (2014) Comparison of near-surface air temperatures and MODIS ice-surface temperatures at Summit, Greenland (2008–13). *Journal of Applied Meteorology and Climatology* **53**(9), 2171–2180. doi: [10.1175/JAMC-D-14-0023.1](https://doi.org/10.1175/JAMC-D-14-0023.1)
- Sohrabinia M (2021) LatLon distance. Available at <https://www.mathworks.com/matlabcentral/fileexchange/38812-latlon-distance>.
- Thackeray CW and Fletcher CG (2016) Snow albedo feedback: current knowledge, importance, outstanding issues and future directions. *Progress in Physical Geography: Earth and Environment* **40**(3), 392–408. doi: [10.1177/0309133315620999](https://doi.org/10.1177/0309133315620999)
- van As D and Fausto RS (2011) Programme for Monitoring of the Greenland Ice Sheet (PROMICE): first temperature and ablation records. *Geological Survey of Denmark and Greenland (GEUS) Bulletin* **23**, 73–76. doi: [10.34194/geusb.v23.4876](https://doi.org/10.34194/geusb.v23.4876)
- van den Broeke M, van As D, Reijmer C and van de Wal R (2004) Assessing and improving the quality of unattended radiation observations in Antarctica. *Journal of Atmospheric and Oceanic Technology* **21**(9), 1417–1431. doi: [10.1175/1520-0426\(2004\)021<1417:AAITQO>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<1417:AAITQO>2.0.CO;2)
- Vaughan DG and 13 others (2013) Observations: cryosphere. In Stocker TF ed. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: IPCC, 5–14.
- Wan Z (2014) New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment* **140**, 36–45. doi: [10.1016/j.rse.2013.08.027](https://doi.org/10.1016/j.rse.2013.08.027)
- Wan Z and Dozier J (1996) A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing* **34**(4), 892–905. doi: [10.1109/36.508406](https://doi.org/10.1109/36.508406)
- Wan Z, Zhang Y, Zhang Q and Li Z (2002) Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment* **83**(1), 163–180. doi: [10.1016/S0034-4257\(02\)00093-7](https://doi.org/10.1016/S0034-4257(02)00093-7)
- Warren SG and Brandt RE (2008) Optical constants of ice from the ultraviolet to the microwave: a revised compilation. *Journal of Geophysical Research: Atmospheres* **113**, D14220. doi: [10.1029/2007JD009744](https://doi.org/10.1029/2007JD009744)
- Wenny BN, Xiong X and Madhavan S (2012) Evaluation of Terra and Aqua MODIS thermal emissive band calibration consistency. *Proc. SPIE, Sensors, Systems, and Next-Generation Satellites XVI* **8533**, 853317. doi: [10.1117/12.974230](https://doi.org/10.1117/12.974230)
- Westermann S, Langer M and Boike J (2012) Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway. *Remote Sensing of Environment* **118**, 162–167. doi: [10.1016/j.rse.2011.10.025](https://doi.org/10.1016/j.rse.2011.10.025)
- Williamson SN, Hik DS, Gamon JA, Kavanaugh JL and Koh S (2013) Evaluating cloud contamination in clear-sky MODIS Terra daytime land surface temperatures using ground-based meteorology station observations. *Journal of Climate* **26**(5), 1551–1560. doi: [10.1175/JCLI-D-12-00250.1](https://doi.org/10.1175/JCLI-D-12-00250.1)
- Williamson SN, Hik DS, Gamon JA, Kavanaugh JL and Flowers GE (2014) Estimating temperature fields from MODIS land surface temperature and air temperature observations in a sub-Arctic alpine environment. *Remote Sensing* **6**(2), 946–963. doi: [10.3390/rs6020946](https://doi.org/10.3390/rs6020946)
- Williamson SN and 6 others (2017) Spring and summer monthly MODIS LST is inherently biased compared to air temperature in snow covered sub-Arctic mountains. *Remote Sensing of Environment* **189**, 14–24. doi: [10.1016/j.rse.2016.11.009](https://doi.org/10.1016/j.rse.2016.11.009)
- Xiaoxiong X and others (2015) Terra and Aqua MODIS thermal emissive bands on-orbit calibration and performance. *IEEE Transactions on Geoscience and Remote Sensing* **53**(10), 5709–5721. doi: [10.1109/TGRS.2015.2428198](https://doi.org/10.1109/TGRS.2015.2428198)
- Yao R and others (2020) A detailed comparison of MYD11 and MYD21 land surface temperature products in mainland China. *International Journal of Digital Earth* **13**(12), 1391–1407. doi: [10.1080/17538947.2019.1711211](https://doi.org/10.1080/17538947.2019.1711211)
- Zellner A (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error terms. *Journal of the American Statistical Association* **71**(354), 400–405. doi: [10.1080/01621459.1976.10480357](https://doi.org/10.1080/01621459.1976.10480357)
- Zhang H and 5 others (2018) Daily air temperature estimation on glacier surfaces in the Tibetan Plateau using MODIS LST data. *Journal of Glaciology* **64**(243), 132–147. doi: [10.1017/jog.2018.6](https://doi.org/10.1017/jog.2018.6)