Spatio-temporal variability of air temperature biases in regional climate models over the Greenland Ice Sheet

Federico COVI,[1,2,3] Regine HOCK,[2,3] Åsa K. RENNERMALM,[4] Xavier FETTWEIS,[5] Brice NOËL[5]

[1]*British Antarctic Survey, Cambridge, United Kingdom*

[2]*Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA*

[3]*Department of Geosciences, University of Oslo, Oslo, Norway*

[4]*Department of Geography, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA*

[5]*Department of Geography, University of Liège, Liège 4000, Belgium*

Correspondence: Federico Covi <fedovi@bas.ac.uk>

ABSTRACT. Regional climate models are fundamental tools in understanding and quantifying the contribution of the Greenland ice sheet to sea-level rise. We perform an extensive evaluation of the daily air temperature simulated by two regional climate models, MARv3.12 and RACMO2.3p2, and a global atmospheric reanalysis, ERA5, at 35 locations across the ice sheet over the period 1995 – 2020. We compare model results to weather station data from two climate networks, focusing on the spatial and temporal variability in mean biases. All three models perform well at low elevations (<1500 m a.s.l.) with a mean bias of $0.16°$C (MAR), $0.36°$C (RACMO), and $0.41°$C (ERA5), while warm biases ($>1.70°$C) are found at high elevations (>1500 m a.s.l.). Temperature biases exhibit a strong seasonality, being more pronounced during winter and much smaller during summer ranging from $0.11°$C to $0.59°$C. No interannual variability is found in the biases of all three datasets. Daily variability within each month is captured well by both climate models and the reanalysis at most locations. Finally, all three models perform overall better in the ablation zone during summer, i.e., where and when considerable melt production occurs.

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

INTRODUCTION

The Greenland ice sheet is a major contributor to sea-level rise with 0.65 ± 0.09 mm yr^{-1} ($\pm$90% confidence interval) over the period 1993 – 2018 (Frederikse and 10 others, 2020), or about 15% of the global mean sea-level rise (Cazenave and 94 others, 2018). Projections indicate further contributions of 0.11 – 0.25 m by 2300 under low emission scenarios (RCP2.6/SSP1-2.6) and 0.31 – 1.74 m under high emission scenarios (RCP8.5/SSP5-8.5) (Fox-Kemper and 17 others, 2021).

Regional climate models (RCMs), such as the Modèle Atmosphérique Régionale (MAR, Fettweis and 8 others (2017)) and the Regional Atmospheric Climate Model (RACMO, Noël and 11 others (2018)), are important tools to understand and quantify the ice sheet's contribution to sea level rise (Fettweis and 40 others, 2020). Forced by global atmospheric reanalysis data or general circulation models, RCMs are used to model the surface energy balance and melt over the entire Greenland ice sheet, allowing for present-day simulations, past reconstructions and future projections of its contribution to sea level rise.

Evaluation of RCMs across the entire ice sheet is a fundamental step to ensure that the spatio-temporal variability of the simulated processes is modeled accurately. RCMs are typically compared to in-situ observations such as near-surface meteorological data from automatic weather stations (AWSs), surface mass balance measurements, or remote sensing data (e.g., used to derive the extent of bare ice area and/or estimates of mass changes through gravimeters and altimeters). However, previous efforts in evaluating RCMs are often limited, both spatially and temporally, mostly due to scarcity of available in-situ observations (Noël and 6 others, 2016; Fettweis and 8 others, 2017; Reeves Eyre and Zeng, 2017; Noël and others, 2019; Delhasse and 6 others, 2020; Zhang and 6 others, 2022). Models are often evaluated with observational data aggregated over the whole ice sheet and/or the entire study period for which data are available. Of these efforts, only Reeves Eyre and Zeng (2017) tried to systematically investigate temporal trends in near surface temperature biases and spatially distinguished their analysis between low ($<$ 1500 m a.s.l.) and high ($>$ 1500 m a.s.l.) elevations sites. Unfortunately, this study mostly focused on global reanalyses rather than RCMs, including only MARv3.5.2 in its comparison. While these approaches are straightforward, much more information on the skill of RCMs can be obtained from evaluating results at finer temporal and spatial scales, e.g. at single sites or during different time periods. Furthermore, RCMs and global reanalysis datasets are in continuous evolution, with new releases every year.

In this study, we focus on two state of the art regional climate models specifically used to produce estimates and projections of sea level rise from the Greenland ice sheet, MARv3.12 (Mankoff and 14

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

others, 2021) and RACMO2.3p2 (Noël and others, 2019). We assess the spatial and temporal variability of these RCMs based on near-surface daily temperature mean biases over the entire Greenland ice sheet between $1996-2020$ by comparing them to observations from the PROMICE (van As and 11 others, 2011) and GC-Net climate networks (Steffen and Box, 2001). For completeness, we also include in our analysis the global reanalysis product ERA5 (Hersbach and 41 others, 2020), which is used to force both MAR and RACMO. We examine the spatial variability in mean model biases evaluating seasonal and inter-annual changes as a function of latitude, longitude and elevation, where we use 1500 m a.s.l as a threshold to separate observations in the ablation zone from the accumulation zone.

DATA AND METHODS

MAR

The Modèle Atmosphérique Régionale (MAR) is a regional climate model based on the atmospheric model by Gallée and Schayes (1994) and fully coupled with the soil-ice-snow energy balance vegetation model SISVAT by Gallée and others (2001). Detailed descriptions of the MAR model and its surface and subsurface scheme SISVAT are given in Fettweis and 8 others (2017) and Reijmer and others (2012). In this study we use MARv3.12 at a spatial resolution of 10 km and forced with ERA5 reanalysis data (Hersbach and 41 others, 2020) every 6 hours. The dataset, including changes from previous versions of the model, is presented in Mankoff and 14 others (2021) while a detailed general description of MAR is given in Fettweis and 8 others (2017). Finally, it is important to note that each new version of MAR is calibrated using PROMICE-based surface mass balance observations along with satellite-derived melt extent data. However, near surface temperature in the accumulation zone does not impact significantly these fields and is therefore not considered a key field used to calibrate the model (Haacker and others, 2024). It is afterwards only used to validate the model (Fettweis and 8 others, 2017; Fettweis and 40 others, 2020).

RACMO

The Regional Atmospheric Climate Model (RACMO) is a regional climate model based on the High Resolution Limited Area Model (Undén and 23 others, 2002) and the physics of the European Centre for Medium-Range Weather ForecastsIntegrated Forecast System (ECMWF, 2009), including a snow module that accounts for subsurface processes (Ettema and others, 2010). In this study we use RACMO2.3p2 (where p stands for polar) at a spatial resolution of 5.5 km and forced with ERA5 reanalysis data (Hersbach

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

and 41 others, 2020) every 6 hours. The dataset is presented in Noël and others (2019) while a detailed description of the model is given in Noël and 11 others (2018). It is important to note that RACMO2.3p2 is not calibrated for improved representation of near-surface temperature.

ERA5

The ERA5 is the fifth and most recent generation of reanalysis products made available by the European Centre for Medium-Range Weather Forecasts (ECMWF). A full description of ERA5 model and improvements compared to its predecessors are listed in Hersbach and 41 others (2020). Because of its higher vertical and spatial resolution ($\sim$ 15 km over Greenland), it has been questioned whether this product could replace the use of RCMs like MAR or RACMO. However, Delhasse and 6 others (2020) found out that RCMs forced with ERA5, like the two used in this study, are still better performing at downscaling near-surface climate over the Greenland ice sheet.

Weather station observations

We use daily mean weather station observations from two Greenland ice sheet wide climate networks (Fig. 1 and Table 1): the Greenland Climate Network (GC-Net), and the Programme for Monitoring of the Greenland Ice Sheet (PROMICE). GC-Net and PROMICE data are neither assimilated in MAR nor in RACMO guaranteeing the independence between modeled and observed air temperatures. However, it is important to note that GC-Net data is assimilated in the production of the ERA5 global reanalysis.

The first GC-Net stations were deployed in 1995, making it the longest-running network over the ice sheet with more than 25 years of data (Steffen and Box, 2001). Most of these stations are located in the accumulation zone (e.g. at elevations > 1500 m a.s.l.). Air temperature is measured at two levels above surface, roughly between 0.5 and 4 m, with a Vaisala CS-500 ($\pm 0.1^\circ$ C) and a Type-E Thermocouple ($\pm 0.1^\circ$ C) at each level, both of which are unventilated. We use the GC-Net augmented level-1 (L1) dataset from Vandecrux and 28 others (2023) archived at Steffen and 35 others (2023) which provides air temperature at 2 m linearly interpolated from the observations at two levels. Furthermore, this dataset has been extensively quality controlled.

The PROMICE weather station network started in 2007 and it currently includes 25 sites mostly located in the ablation zone (e.g. at elevations < 1500 m a.s.l.) of outlet glaciers (van As and 11 others, 2011). Air temperature is measured at one level above the surface, at roughly 2 m, with a Rotronic MP100H and a

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

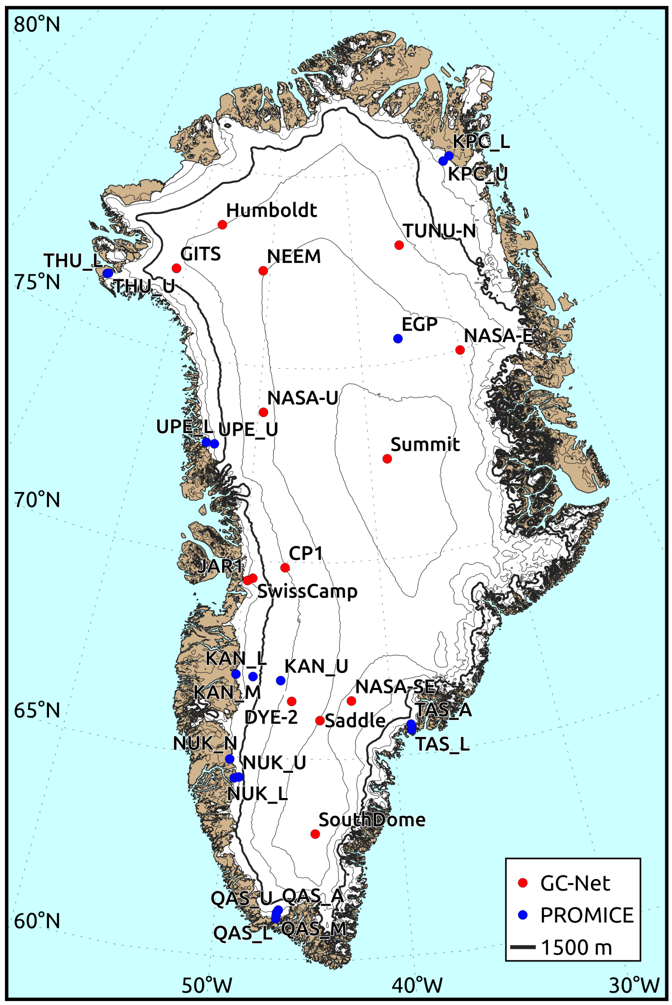

Fig. 1. Map of the Greenland ice sheet with the GC-Net and PROMICE weather stations used in this study. Elevation contours based on the ArcticDEM 1 km v3.0 product by the Polar Geospatial Center (Porter and 29 others, 2018) are shown at 500 m intervals (thin black lines) with the 1500 m contour highlighted by thick lines. The ice sheet extent (white) is based on Howat and others (2014).

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

Rotronic HygroClip S3 both mounted in an artificially ventilated Rotronic assembly. We use the MP100H measurements, which are provided quality-checked.

We exclude from our analysis all stations that are located on glaciers outside the ice sheet (three PROMICE stations) and the stations for which the difference between the site elevation and the one interpolated from either model exceeds 100 m (two PROMICE stations in East Greenland). These stations are not shown in Fig. 1 nor listed in Table 1. The average difference ($\pm$ RMSE) between the model grid elevation and the actual elevation derived from on site GPS measurements is -14 ± 43 m for MAR and -16 ± 42 m for RACMO. A total of 35 stations are used in this study (Table 1), 14 from the GC-Net network and 21 from the PROMICE network.

Data analysis

For simplicity, in this study when we refer to models we include both RCMs, MAR and RACMO, and the global reanalysis ERA5. At each station daily mean 2 m air temperatures from the models were compared to the weather station data. Model data were interpolated to each weather station site following a linear distance-weighted average of the four nearest grid point values. We computed the mean bias (MB: model − observed temperatures), root-mean-square error (RMSE), correlation coefficient (r), and the p-value (p). This approach is similar to that used in previous studies (Fettweis and 8 others, 2017; Zhang and 6 others, 2022) and often considered a better approach to using the nearest model grid-cell as often done in other validation studies (Noël and others, 2019). However, we investigated the spatial and temporal patterns in air temperature bias with particular attention to altitudinal, latitudinal and longitudinal trends as well as annual and interannual variability. We refer to the four seasons in a year as follows: March, April, May (MAM); June, July, August (JJA); September, October, November (SON); and December, January, February (DJF).

RESULTS

Combining the data from all sites and over the entire study period (138,361 daily means; Fig. 2 and Table 2) reveals that all three models show a warm bias compared to observations with a mean of 0.89° C for MAR, 1.04° C for RACMO, and 1.11° C for ERA5. The correlation between modelled and measured temperatures is very strong ($r > 0.95$) and statistically significant ($p < 0.01$) for all models. However, the root-mean-square error (RMSE) is large for all three models ($3.0 \pm 0.3°$ C). When considering the JJA

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

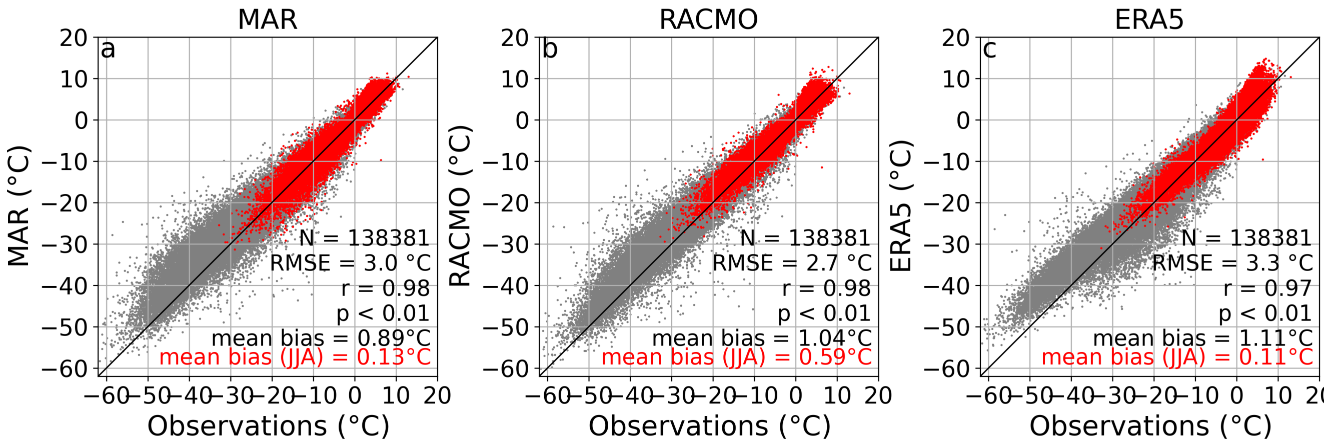

Fig. 2. Mean daily 2 m air temperatures from (a) MAR, (b) RACMO, and (c) ERA5 versus weather station observations over the entire study period (1996-2020) and for all the sites. Data for the June to August (JJA) period are shown in red and the 1:1 line is given in black. N is the number of samples, RMSE the root mean square error, r the correlation coefficient, and p the p-value.

period in isolation (red dots in Fig. 2), both mean bias and RMSE for the summer period is considerably lower for all models, by 0.74° C and 1.10° C on average respectively.

Spatial variability

With most stations' data coverage spanning more than ten years (only three stations have less than five years of data, Table 1), we have a sufficiently long temporal range to estimate meaningful mean biases at each individual site (Fig. 3 and Table S1). Maps of mean bias (Fig. 3a, b, c) show that for all models the annual mean bias is smaller at sites located at low elevations (< 1500 m a.s.l.) near the ice sheet margin, e.g. in the ablation zone. Here, the annual mean bias ranges from $-1.20°$ C to 1.59° C in MAR, from $-1.08°$ C to 1.36° C in RACMO, and from $-2.63°$ C to 2.15° C in ERA5 (Table S1) with most values between $\pm 0.5°$C. However, at high elevations (> 1500 m a.s.l.), e.g. in the accumulation zone, large positive annual mean biases can be found, with values up to 2.61° C in MAR, 2.95° C in RACMO, and 2.76° C in ERA5.

We further investigate the spatial variability by plotting the mean bias against elevation, latitude and longitude (Fig. 4, where blue and red circles indicate low and high elevations sites). All models indicate that the mean bias increases with increasing elevation, with a trend of 0.81° C km^{-1} in MAR, 0.75° C km^{-1} in RACMO, and 1.01° C km^{-1} in ERA5, respectively (Fig. 4a, b, c), with linear regressions showing a moderate but statistically significant correlation ($p < 0.01$) for all models ($r = 0.79$ for MAR, $r = 0.73$

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

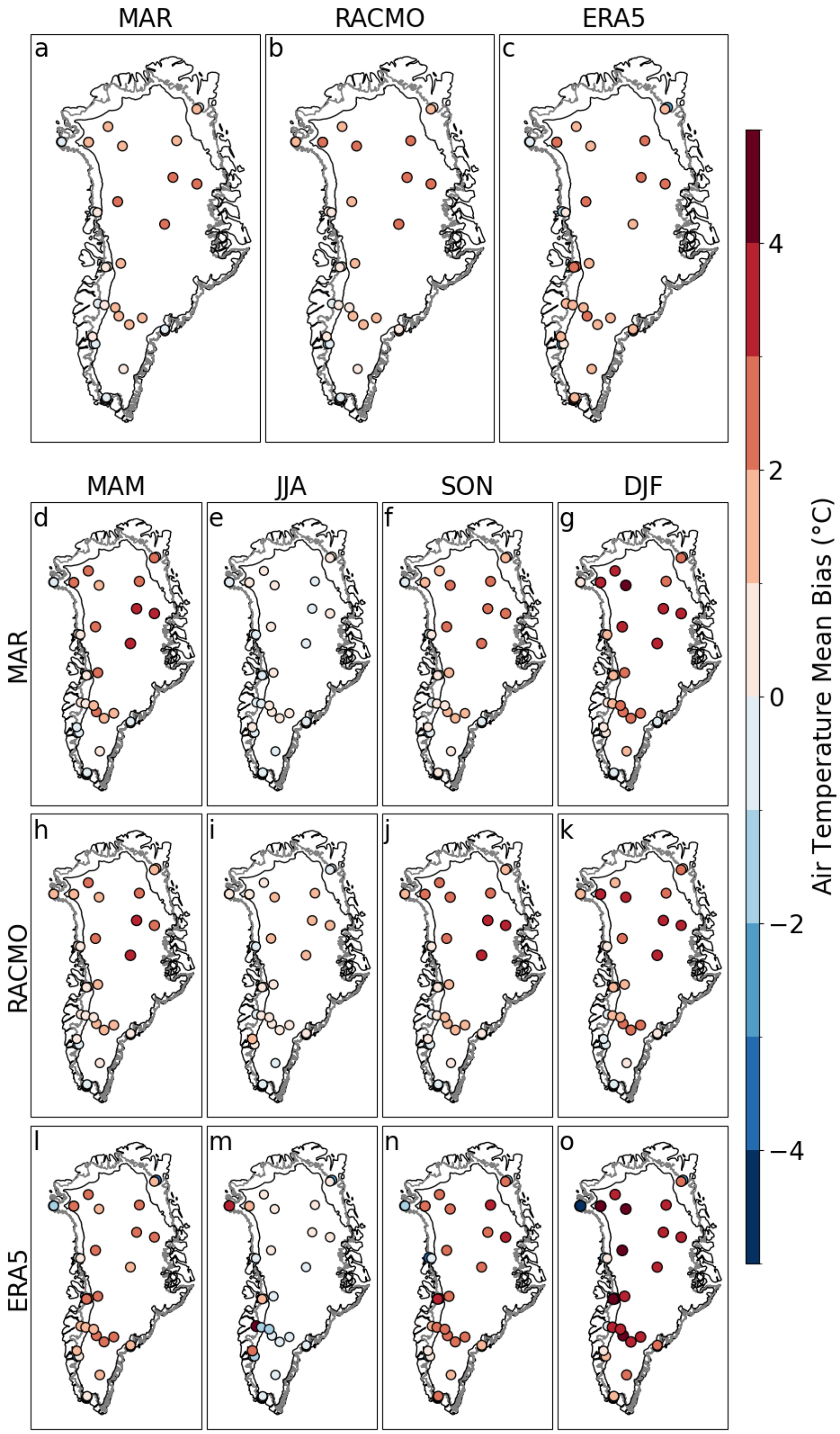

Fig. 3. Maps of mean bias in 2 m air temperature in (a, d-g) MAR, (b, h-k) RACMO, and (c, l-o) ERA5 compared to daily observations at 35 sites (a-c) over the entire study period and (d-o) for four seasons. The 1500 m contour (black line) is from the ArcticDEM (Porter and 29 others, 2018) and the ice sheet extent (gray line) is based on Howat and others (2014).

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

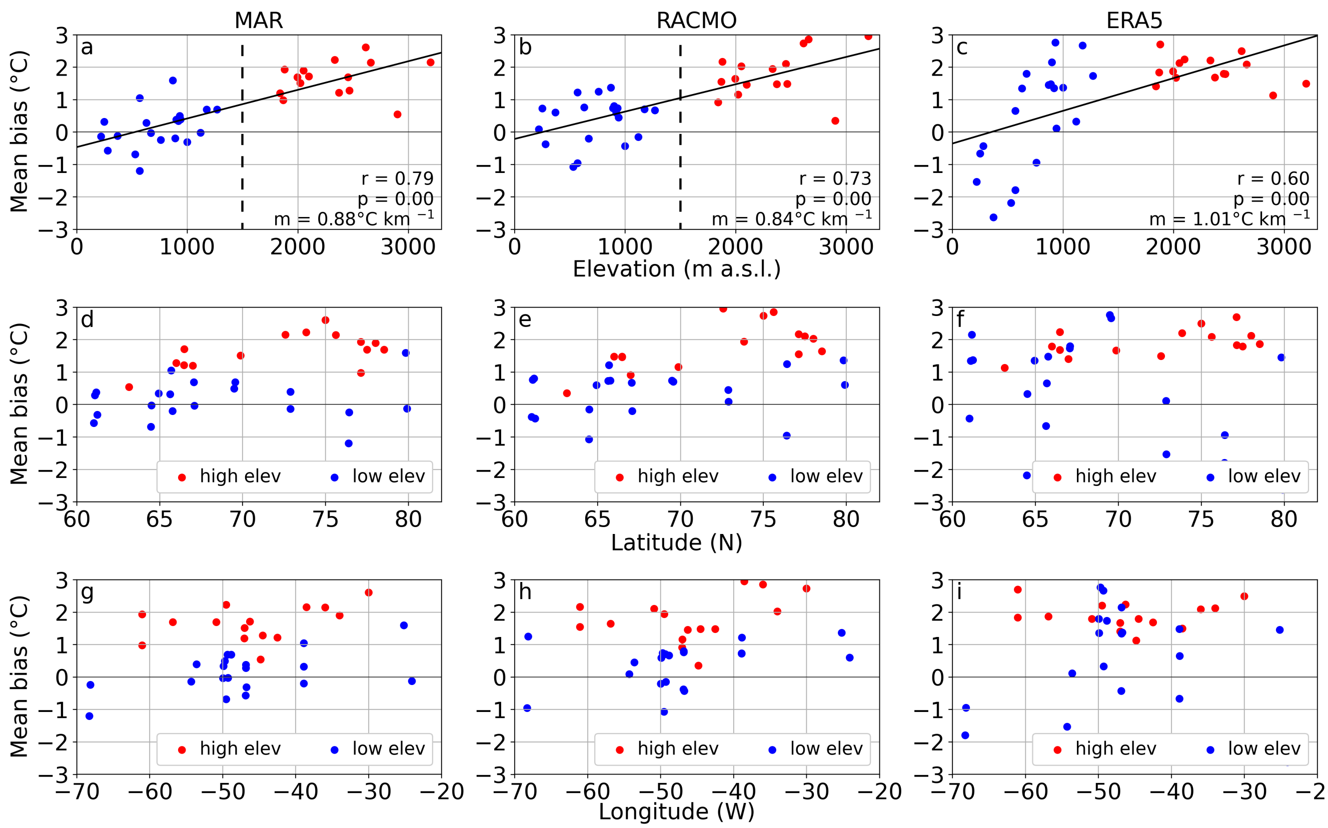

Fig. 4. Mean bias in 2 m air temperature between models, (a, d, g) MAR, (b, e, h) RACMO, and (c, f, i) ERA5, and daily observations plotted against (a, b, c) elevation, (d, e, f) latitude, and (g, h, i) longitude for each site over the entire study period (1996-2020). Low elevation sites (< 1500 m a.s.l.) are shown in blue and high elevation sites (> 1500 m a.s.l.) in red. In (a, b, c) linear regressions are shown in solid lines, r is the correlation coefficient, p the p-value, and m the slope. The dashed black line in (a, b) highlights the 1500 m elevation.

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

for RACMO, $r = 0.60$ for ERA5). While for MAR and RACMO the mean bias increase with elevation is steady and constant, for ERA5, the transition from a negative to a positive mean bias is notably steep below 1000 m a.s.l., stabilizing at approximately 2° C above 1000 m a.s.l. (Fig. 4c). No significant trend is found with either latitude (Fig. 4d, g, f) or longitude (Fig. 4g, h, i). The scatter is considerably larger than for the elevation dependency, and the magnitude of the mean bias is again controlled by its elevation rather than by its latitude or longitude as shown by the blue (low elevation sites) and red (high elevation sites) circles in Fig. 4.

Temporal Variability

Mean biases computed for each of the four seasons (Fig. 3d-o) reveal a strong annual seasonality in all models. Both mean seasonal bias and root-mean-square error computed over all the sites are larger during the winter time (DJF) and smaller during the summer (JJA) (Table 2). Because of the strong elevation dependency in mean bias found above, we further distinguish between sites at high elevation ($>$ 1500 m a.s.l.) and low elevation ($<$ 1500 m a.s.l.). While seasonal variations are absent at low elevations, the seasonality is amplified at high elevations (Fig. 5), with monthly mean bias ranging from -0.30° C to 3.28° C in MAR, from 0.66° C to 2.52° C in RACMO, and from -0.44° C to 3.96° C in ERA5. The amplitude of the annual seasonality is the largest in ERA5, followed by MAR and then RACMO (Fig. 3 and Fig. 5).

Interpreting the spatial difference in seasonality requires care since the two weather station networks used in this study cover different time periods and different areas of the Greenland ice sheet. GC-Net data cover almost all of the study period and these stations are almost entirely located at high elevation ($>$ 1500 m a.s.l., Table 1). PROMICE data are available starting from 2007 and these stations are almost entirely located at low elevation ($<$ 1500 m a.s.l., Table 1).

Fig. 6 shows the annual mean bias and standard deviation in 2 m air temperature between models and observations. When considering years with more than $\sim$ 10 stations available for the calculations, the annual mean bias (Fig. 6) does not show a trend as evident as the seasonality shown by the monthly mean bias. At high elevation annual mean bias ranges from 1.23° C to 2.30° C in MAR, from 1.20° C to 2.02° C in RACMO, and 1.23° C to 2.50° C in ERA5 (Fig. 6b). At lower elevations annual mean bias ranges from -0.26° C to 0.41° C in MAR, from -0.02° C to 0.65° C in RACMO, and from 0.00° C to 0.59° C in ERA5 (excluding years prior to 2008 when only 2 GC-Net stations were operational, Fig. 6c). These

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

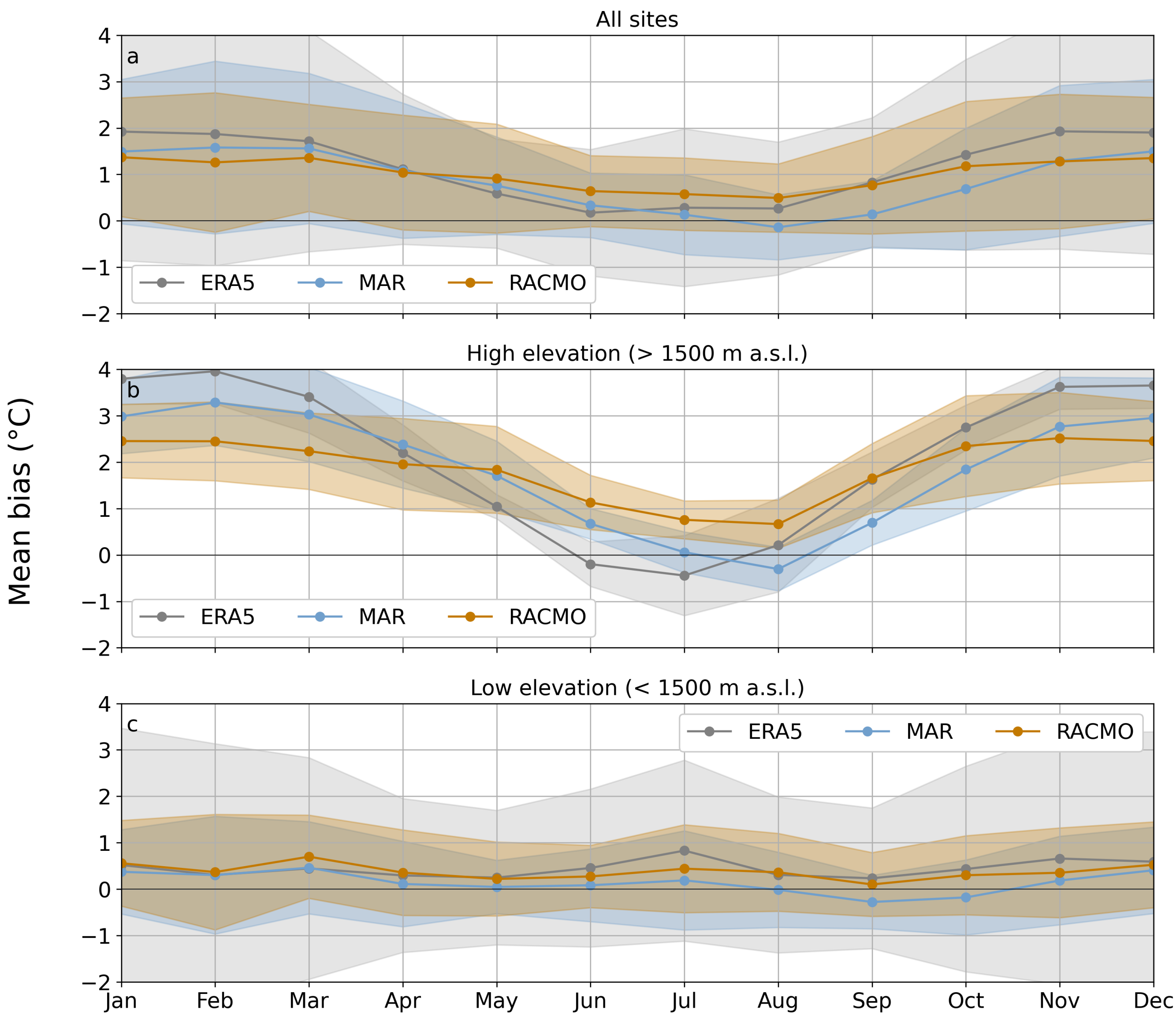

Fig. 5. Monthly mean bias in 2 m air temperature and standard deviation (shaded) between models (MAR, RACMO, and ERA5) and observations at (a) all sites, (b) high elevation sites, and (c) low elevation sites.

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

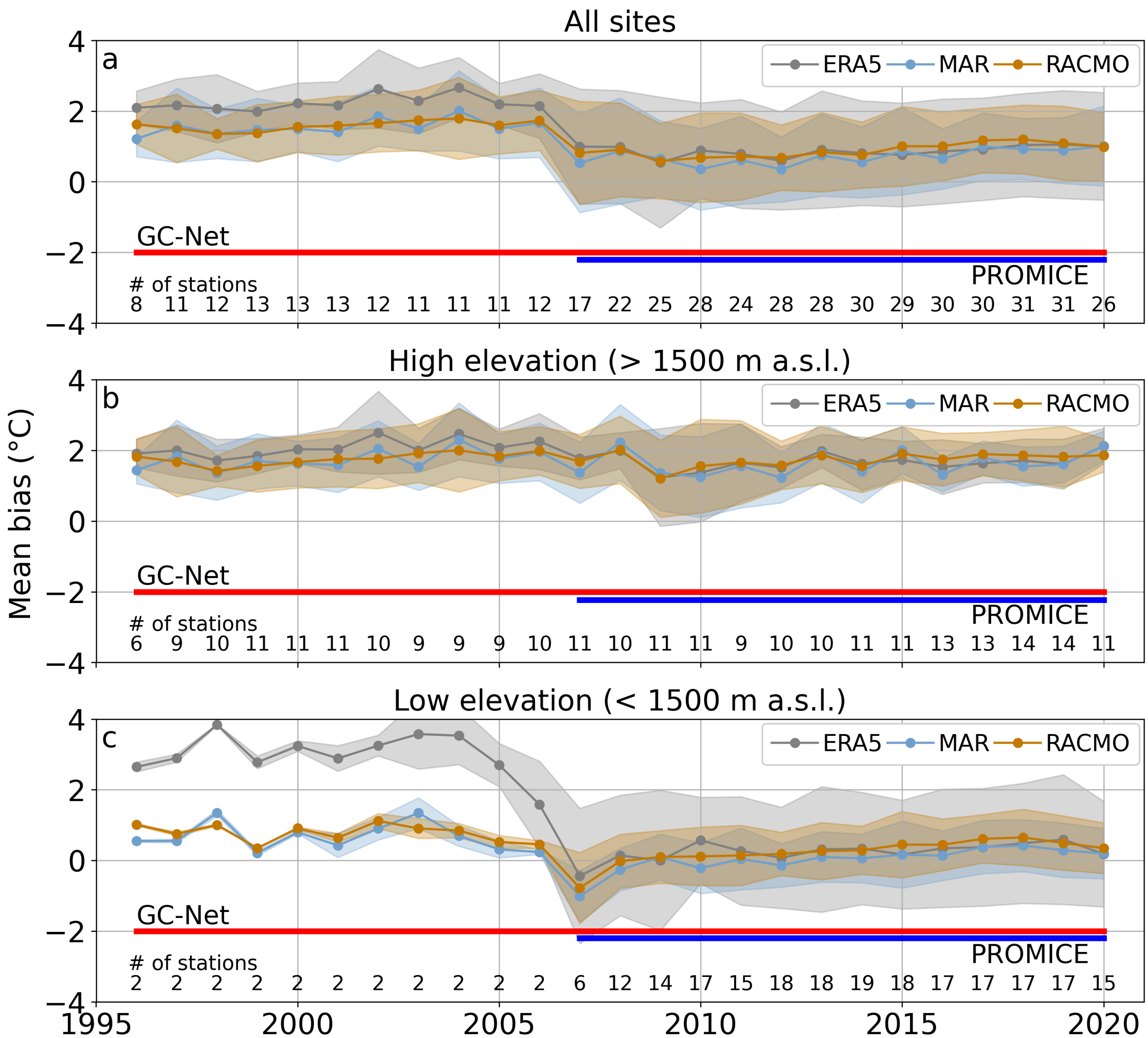

Fig. 6. Annual mean bias in 2 m air temperature and standard deviation (shaded) between models (MAR, RACMO, and ERA5) and observations at (a) all sites, (b) high elevation sites, and (c) low elevation sites. Data coverage is shown for the GC-Net (red) and PROMICE (blue) weather stations network. The number of stations from which each annual mean is computed is also shown (# of stations).

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

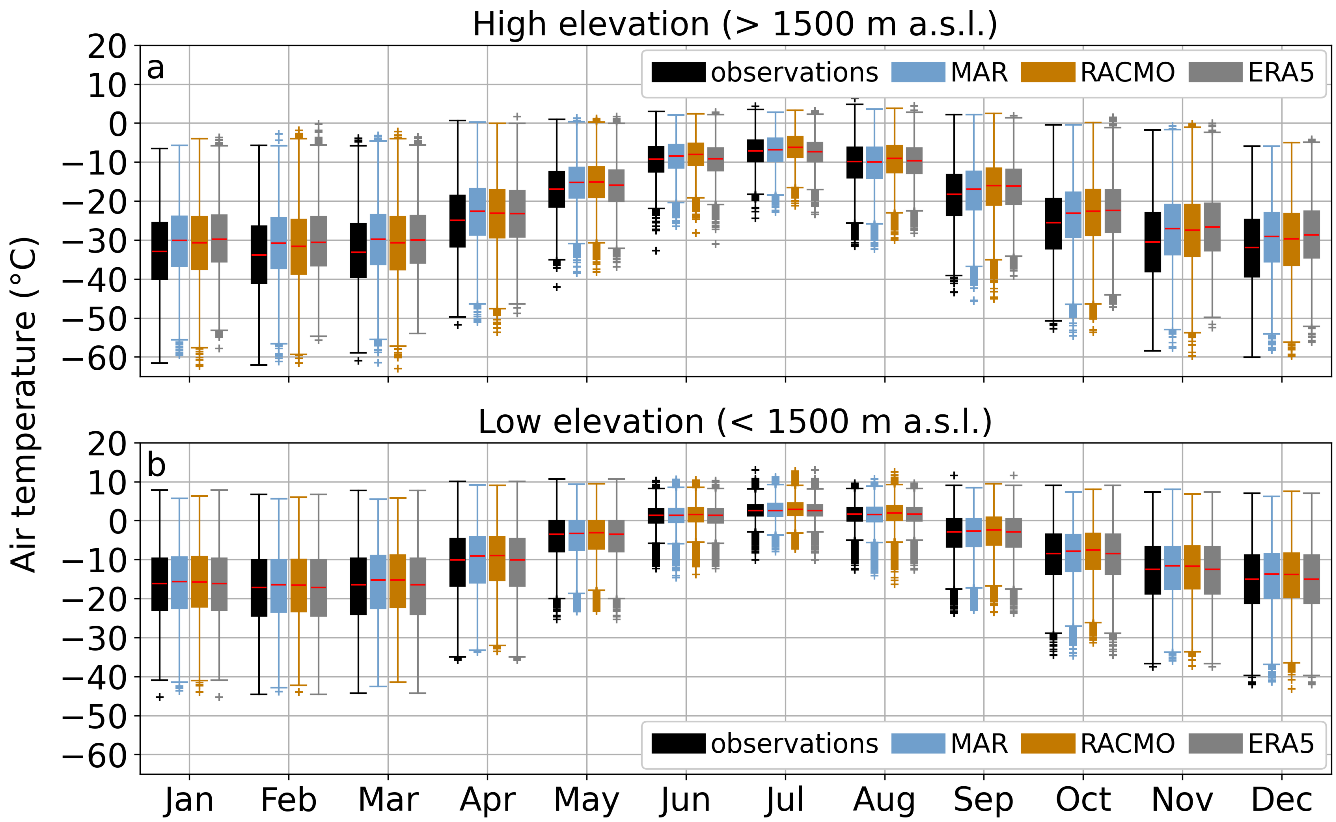

Fig. 7. Boxplot of monthly 2 m air temperature from observations and models (MAR, RACMO, and ERA5) at (a) high elevation sites, and (b) low elevation sites. Median is shown with a red line, 1^{st} and 3^{rd} quartiles with a box, lower and upper whiskers with colored lines, and outliers as a cross.

results confirm once more the strong elevation dependency of the air temperature mean bias, with greatest biases at the higher elevations.

When the data from all sites are included (Fig. 6a), the annual mean bias shows a clear step-like drop in 2007 and remains relatively constant thereafter. This drop coincides with the first year in which the first PROMICE stations were deployed, complementing the only 2 GC-Net stations located at elevations $<$ 1500 m a.s.l.. The decrease in annual mean bias is due to the overall smaller biases at PROMICE sites which are located at low elevations.

Daily variability

To assess the skill of the models in simulating daily variability throughout the year we examine the distribution of daily mean air temperature in each month for MAR, RACMO, ERA5, and the observations for high elevation and low elevation sites (Fig. 7). Two patterns in both observations and models emerge

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

from this analysis: first, air temperature variability is larger during winter and smaller during summer; and second, variability is larger at high elevation (Fig. 7a) than at low elevation (Fig. 7b). All models capture well the daily variability in air temperature at both high and low elevations and in all months of the year. Median values reflect what is shown by the mean bias analysis above, i.e. the models have a warm bias especially at high elevations and during the winter (Fig. 7a). Both the interquartile range, which contains 50% of the data, and the lower to upper whisker range, which contains 99.3% of the data, compare very well between the models and the observations (Fig. 7).

DISCUSSION

Air temperature biases show a strong dependency on elevation in all models (MAR, RACMO, and ERA5), while no dependency is found with latitude or longitude (Fig. 4ab). The mean model bias is 0.16, 0.36, and 0.41°C at elevations $<$ 1500 m a.s.l. and 1.71, 1.79, and 1.89°C at elevations $>$ 1500 m a.s.l., for MAR, RACMO, and ERA5, respectively. This dependency on elevation recurs in all the statistical analyses performed. A strong seasonality in mean bias is found only at high elevations (Fig. 5), and annual mean biases are higher at elevations $>$ 1500 m a.s.l. (Fig. 6).

When compared to previous studies, our results reveal in general greater biases for all three models. For MARv3.5, forced with ERA-Interim reanalysis data, Fettweis and 8 others (2017) found a negative mean bias of −0.29°C (validation at 12 PROMICE stations over the period 2008 – 2010), while Reeves Eyre and Zeng (2017) found a positive mean bias of 1.38°C (validation at PROMICE, GC-Net and other available stations over the period 1958 – 2015, e.g. coastal stations from the Danish Meteorological Institute). For MARv3.9, forced with ERA5 reanalysis data, Delhasse and 6 others (2020) found a mean bias of 0.06°C (validation at 21 PROMICE stations over the period 2010 – 2016) compared to the bias of 0.98°C in this study. For RACMO2.3p2, forced with ERA-Interim reanalysis data, Noël and others (2019) found a mean bias of 0.14°C (validation at 18 PROMICE and 5 IMAU stations over the period 2007 – 2016), while Zhang and 6 others (2022) found a mean bias of of 1.0°C using monthly data (validation at 20 PROMICE stations over the period 2007 – 2020) compared to the 1.04°C in this study. For ERA5 global reanalysis Delhasse and 6 others (2020) found a mean bias of 0.01°C (validation at 21 PROMICE stations over the period 2010 – 2016), while Zhang and 6 others (2022) found a mean bias of 2.0°C using monthly data (validation at 20 PROMICE stations over the period 2007 – 2020) compared to the 1.11°C in this study. However, the comparison is not straightforward and our results should not be interpreted as the models'

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

skill has deteriorated from previous studies. Instead the reason for the higher bias in our study can be explained by differences in study design, since previous studies often only analyzed the PROMICE stations, used different model versions forced with different datasets, and validations were performed over different time periods. If we consider only PROMICE stations in our analysis, we find mean biases ($\pm$ RMSE) of 0.23 ± 2.29°C for MAR, 0.44 ± 2.15°C for RACMO, and 0.22 ± 2.94°C for ERA5 (Fig. 8), which are still warmer than previous studies but much smaller than the results including GC-Net stations.

A key difference between the two AWS networks is that PROMICE stations are mostly located at low elevations ($<$ 1500 m a.s.l.) while GC-Net stations are at high elevations ($>$ 1500 m a.s.l.). This explains why, when GC-Net stations are removed from the analysis, the biases decrease as most of the warm biases are found at higher elevations. A possible reason for the greater warm bias at higher elevations is the daily variability in air temperature. Larger variability at high elevations during the winter (Fig. 7) could in fact lead to higher biases. However, all models capture well the daily variability (Fig. 7). Another possible explanation is that models are continuously developed to improve surface mass balance representation compared to in-situ measurements (Fettweis and 8 others, 2017; Noël and 11 others, 2018), which are mostly located at lower elevations, in the ablation zone where most of the melt occurs. However, model tuning typically does not involve air temperature calibration. Furthermore, GC-Net stations data are assimilated in the ERA5 reanalysis; yet the mean bias at GC-Net stations for ERA5 is larger than for MAR and RACMO (Fig. 8a). Finally, warm biases could be explained by systematically erroneous colder air temperature observations at the GC-Net stations, e.g. by riming or sensor burial by snow during winter. However, the augmented L1 GC-Net dataset was carefully quality controlled in order to avoid erroneous data due to external factors like sensors burial and riming (Vandecrux and 28 others, 2023). Furthermore, while warmer air temperature observations could be physically explained (e.g. by the use of unventiled sensors, discussed in detail in the next section), it is hard to explain consistently colder air temperature measurements observed in this study.

When discussing possible bias sources in the models, the most obvious one is that both MAR and RACMO are forced with ERA5 data, hence the biases in the two RCMs could be directly inherited by the forcing dataset (e.g. Fig. 3 and 5). However, both MAR and RACMO include boundary layer parameterizations which are independent of the forcing dataset, making them sensitive to ERA5 biases only in the free atmosphere. It could be the case that near surface biases are corrected by RCMs if the free atmosphere is well represented in ERA5.

When considering model resolution, all three models differ substantially with a resolution of 5.5 km for RACMO, 10 km for MAR, and 15 km for ERA5 (over Greenland). When looking at the whole ice sheet, model resolution doesnt seem to affect the mean bias, however RMSE decreases with finer resolution (2.7°C for RACMO, 3.0°C for MAR, and 3.3°C for ERA5). Model resolution becomes more important at lower elevations, where the topography is more complex, compared to higher elevations where the ice sheet is generally flatter and more uniform. This is evident in Fig. 4c, where ERA5 clearly underestimates the 2 m air temperature at very low elevations. In contrast, both MAR and RACMO do not exhibit significant sensitivity to model resolution at these elevations.

Another source of biases in the 2 m air temperature could potentially be traced to the model parameterizations. While this is not the scope of this work a few important speculations are listed below to encourage further validation efforts. For example, clouds representation directly affects the near-surface air temperature via the surface energy budget, by enhancing or reducing shortwave and longwave downward radiation. A proper representation of clouds is thus required to reduce biases. Furthermore, snow models used to parameterize snow and ice processes affect the near-surface air temperature via the albedo effect and a proper representation of snow extent and properties is essential to reduce temperature biases.

Unventilated observations

While PROMICE weather stations use ventilated air temperature sensors, GC-Net does not. It is known that non-ventilated sensors over snow surfaces tend to overestimate air temperature during summer, when solar radiation reaching the sensor shield is the strongest and typically wind speed, providing natural ventilation, is low (Arck and Scherer, 2001). This might be of concern for this study, especially because GC-Net stations are predominantly located at high elevation. We hypothesize that an overestimation of observed air temperatures during the summer at these sites might in fact be responsible for the strong seasonality in mean bias found at high elevation, given that the models warm bias drops considerably during the summer months.

To investigate this hypothesis we computed mean bias and RMSE over the JJA period using only days with average wind speed greater than 2.5, 5.0, and 7.5 m s^{-1} and separating sites belonging to the GC-Net and PROMICE networks (Fig. 8). Our reasoning is to verify whether the bias is affected by wind speed during summer, proving that non-ventilated sensors may introduce a systematic bias in air temperature measurements.

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

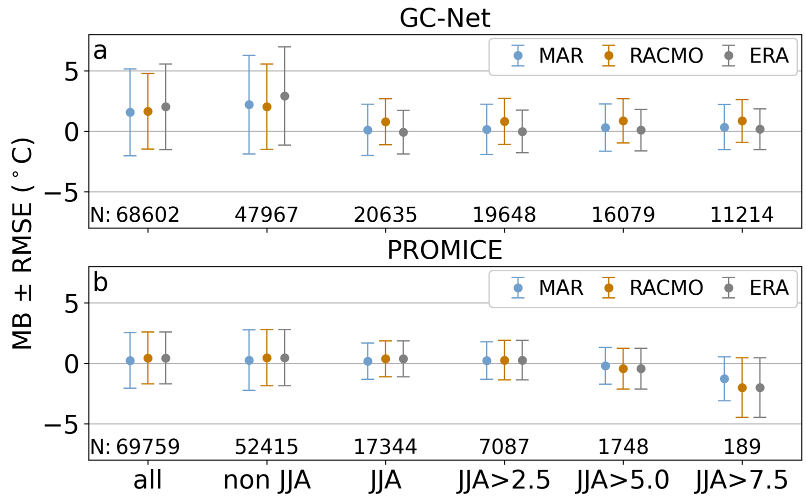

Fig. 8. Mean bias (MB) and root-mean-square error (RMSE) in 2 m air temperature between models (MAR, RACMO, and ERA5) and the (a) GC-Net and (b) PROMICE climate networks computed over the whole study period (all), over the period between September and May (non JJA), over the summer (JJA), and over the summer but using only days with average wind speed greater than 2.5, 5.0, or 7.5 m s^{-1} (e.g. JJA > 2.5 m s^{-1}, etc). Sample number (N) is shown at the bottom of each plot.

This analysis reveals that at GC-Net sites the mean bias becomes slightly more positive when days with low wind speed are excluded from the calculation (Table S2). However, even when only days with wind speed > 7.5 m s^{-1} are considered, the mean bias is far from the annual average for all three models (0.33° C versus 1.57° C for MAR, 0.85° C versus 1.65° C for RACMO, and 0.16° C versus 2.02° C for ERA5 (Fig. 8, Table S2)).

At PROMICE sites the situation is the opposite, with mean biases becoming slightly more negative as the wind speed threshold increases. It has to be noted that the number of days with high wind speed (N in (Fig. 8)) is much smaller at low elevation (PROMICE sites) than at high elevation (GC-Net sites). In fact, there are only 1748 and 189 days with wind speeds > 5.0 and > 7.5 m s^{-1}, respectively, at all PROMICE sites. We therefore urge caution when interpreting and comparing statistics computed from such differently sized samples. In summary, this analysis reveals that while there is an indication that GC-Net stations might be overestimating air temperature during summer due to the usage of unventilated sensors this alone cannot explain the strong annual seasonality in mean bias found at high elevations.

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

CONCLUSION

We performed an extensive evaluation of air temperature simulated by two regional climate models, MARv3.12 and RACMO2.3p2, and a global reanalysis, ERA5, over the entire Greenland ice sheet. We computed the mean bias in air temperature over the period $1996-2020$ at 35 sites where weather station data are available from two climate networks (GC-Net and PROMICE) showing that focusing on spatial and temporal variability of mean bias, can provide useful information.

All models perform well at low elevations, in the ablation zone ($<$ 1500 m a.s.l.), where most of the melt occurs. However a warm bias in air temperature is consistently found in all models at high elevations ($>$ 1500 m a.s.l.). The warm bias does not vary interannually but shows a strong seasonal variability, with higher warm biases during the winter and biases approaching 0°C during the summer. The seasonality of the temperature bias is stronger in ERA5, followed by MAR and then by RACMO. However, the source of the warm bias at high elevations remains unclear, as it is not affected by unventilated temperature measurements at GC-Net stations nor by sensor burial. A detailed analysis of high wind speed conditions reveals that the mean bias is only slightly affected by wind speeds. Furthermore, the warm bias at high elevations during the winter is also not explained by daily variability in air temperature since all models capture it well.

Regional climate models and global reanalysis are important tools in understanding and quantifying the contribution of the Greenland ice sheet to sea level rise. Our study shows that models are able to reproduce air temperature well in the ablation zone (MAR MB = 0.16°C, RACMO MB = 0.36°C, ERA5 MB = 0.41°C) and in the summer also at higher elevations (MAR MB JJA = 0.15°C, RACMO MB JJA = 0.87°C, ERA5 MB JJA = -0.21°C), which is where and when melt occurs the most. Although RCMs and global reanalysis show low biases on an annual basis, significant biases remain at high elevation in winter ($>$ 2° C). However, these biases are not likely to significantly affect modeled surface mass balance of the GrIS as air temperature remain well below the melting point in elevated regions during winter time.

ACKNOWLEDGEMENTS

Funding for this work was provided by the US National Science Foundation (NSF) (Grants OPP-1603815 and OPP-1604058). RH was also supported by Norwegian Research Council Project 324131 and the ERC-2022-ADG grant agreement 101096057 (GLACMASS) We thank Martin Truffer and Matthew Sturm for

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

comments on the manuscript. We thank scientific editor Joseph Shea and three reviewers for their valuable reviews.

DATA AVAILABILITY

The stable version of the GC-Net level-1 dataset is available at `https://doi.org/10.22008/FK2/VVXGUT`, (Steffen and 35 others, 2023). The PROMICE weather stations' data are available at How and 23 others (2022). MAR outputs were provided by Xavier Fettweis. RACMO outputs were provided by Brice Noël. ERA5 reanalysis data can be found at the ECMWF Climate Data Store at `https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5`.

AUTHOR CONTRIBUTIONS

FC and RH conceived the study. FC designed and performed the analysis and wrote the manuscript with input from RH. XF and BN contributed substantially to the interpretation of the MAR and RACMO models respectively. All co-authors reviewed and contributed to the editing of the manuscript.

REFERENCES

Arck M and Scherer D (2001) A physically based method for correcting temperature data measured by naturally ventilated sensors over snow. *Journal of Glaciology*, **47**(159), 665–670 (doi: DOI:10.3189/172756501781831774)

Cazenave A and 94 others (2018) Global sea-level budget 1993-present. *Earth System Science Data*, **10**, 1551–1590 (doi: 10.5194/essd-10-1551-2018)

Delhasse A and 6 others (2020) Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet. *The Cryosphere*, **14**(3), 957–965 (doi: 10.5194/tc-14-957-2020)

ECMWF (2009) IFS Documentation CY33R1 - Part IV: Physical Processes. In *IFS Documentation CY33R1*, IFS Documentation, ECMWF

Ettema J, Van Den Broeke MR, Van Meijgaard E, Van De Berg WJ, Box JE and Steffen K (2010) Climate of the Greenland ice sheet using a high-resolution climate model - Part 1: Evaluation. *The Cryosphere*, **4**(4), 511–527 (doi: 10.5194/tc-4-511-2010)

Fettweis X and 40 others (2020) GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet. *The Cryosphere*, **14**(11), 3935–3958 (doi: 10.5194/tc-14-3935-2020)

Fettweis X and 8 others (2017) Reconstructions of the 1900-2015 Greenland ice sheet surface mass balance using the regional climate MAR model. *The Cryosphere*, **11**(2), 1015–1033 (doi: 10.5194/tc-11-1015-2017)

Fox-Kemper B and 17 others (2021) Ocean, cryosphere and sea level change. In V Masson-Delmotte and 18 others (eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1211 1362, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (doi: 10.1017/9781009157896.011)

Frederikse T and 10 others (2020) The causes of sea-level rise since 1900. *Nature*, **584**(7821), 393–397 (doi: 10.1038/s41586-020-2591-3)

Gallée H and Schayes G (1994) Development of a Three-Dimensional Meso-γ Primitive Equation Model: Katabatic Winds Simulation in the Area of Terra Nova Bay, Antarctica. *Monthly Weather Review*, **122**(4), 671–685 (doi: 10.1175/1520-0493(1994)122<0671:DOATDM>2.0.CO;2)

Gallée H, Guyomarc'h G and Brun E (2001) Impact Of Snow Drift On The Antarctic Ice Sheet Surface Mass Balance: Possible Sensitivity To Snow-Surface Properties. *Boundary-Layer Meteorology*, **99**(1), 1–19 (doi: 10.1023/A:1018776422809)

Haacker J, Wouters B, Fettweis X, Glissenaar IA and Box JE (2024) Atmospheric-river-induced foehn events drain glaciers on novaya zemlya. *Nature Communications*, **15**(1), 7021 (doi: 10.1038/s41467-024-51404-8)

Hersbach H and 41 others (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**(730), 1999–2049 (doi: 10.1002/qj.3803)

How P and 23 others (2022) PROMICE and GC-Net automated weather station data in Greenland (doi: 10.22008/FK2/IW73UU)

Howat IM, Negrete A and Smith BE (2014) The Greenland Ice Mapping Project (GIMP) land classification and surface elevation data sets. *The Cryosphere*, **8**(4), 1509–1518 (doi: 10.5194/tc-8-1509-2014)

Mankoff KD and 14 others (2021) Greenland ice sheet mass balance from 1840 through next week. *Earth System Science Data*, **13**(10), 5001–5025 (doi: 10.5194/essd-13-5001-2021)

Noël B and 11 others (2018) Modelling the climate and surface mass balance of polar ice sheets using RACMO2 - Part 1: Greenland (1958-2016). *The Cryosphere*, **12**(3), 811–831 (doi: 10.5194/tc-12-811-2018)

Noël B and 6 others (2016) A daily, 1 km resolution data set of downscaled Greenland ice sheet surface mass balance (1958-2015). *The Cryosphere*

Noël B, van de Berg WJ, Lhermitte S and van den Broeke MR (2019) Rapid ablation zone expansion amplifies north Greenland mass loss. *Science Advances*, **5**(9), 2–11 (doi: 10.1126/sciadv.aaw0123)

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

Porter C and 29 others (2018) ArcticDEM, Version 3 (doi: 10.7910/DVN/OHHUKH)

Reeves Eyre JEJ and Zeng X (2017) Evaluation of greenland near surface air temperature datasets. *The Cryosphere*, **11**(4), 1591–1605 (doi: 10.5194/tc-11-1591-2017)

Reijmer CH, Van Den Broeke MR, Fettweis X, Ettema J and Stap LB (2012) Refreezing on the Greenland ice sheet: A comparison of parameterizations. *The Cryosphere*, **6**(4), 743–762 (doi: 10.5194/tc-6-743-2012)

Steffen K and 35 others (2023) GC-Net Level 1 historical automated weather station data (doi: 10.22008/FK2/VVXGUT)

Steffen K and Box J (2001) Surface climatology of the Greenland ice sheet: Greenland Climate Network 1995-1999. *Journal of Geophysical Research Atmospheres*, **106**(D24), 33951–33964 (doi: 10.1029/2001JD900161)

Undén P and 23 others (2002) The, HIRLAM-5 Scientific Documentation. Technical report

van As D and 11 others (2011) Programme for Monitoring of the Greenland Ice Sheet (PROMICE): First temperature and ablation records. *Geological Survey of Denmark and Greenland Bulletin*, **23**, 73–76

Vandecrux B and 28 others (2023) The historical greenland climate network (gc-net) curated and augmented level-1 dataset. *Earth System Science Data*, **15**(12), 5467–5489 (doi: 10.5194/essd-15-5467-2023)

Zhang Q and 6 others (2022) Temporal and spatial variability in contemporary greenland warming (19582020). *Journal of Climate*, **35**(9), 2755 – 2767 (doi: 10.1175/JCLI-D-21-0313.1)

Table 1. Weather stations from the GC-Net and PROMICE networks used in this study. Start and End denote the years with the first and the last temperature observation used in this study, respectively. Days refers to the number of days used in the analysis and Years to the equivalent number of years. Elevation (Elev) for each site is taken from the networks metadata (van As and 11 others, 2011; Steffen and 35 others, 2023). Source refers to: 1 Vandecrux and 28 others (2023) and 2 van As and 11 others (2011).

Name	Lat (° N)	Lon (° W)	Elev (m a.s.l.)	Start	End	Days	Years	Source
			GC-Net					
SwissCamp	69.5647	-49.3308	1176	1996	2022	5660	15.5	1
CP1	69.8783	-46.9967	2022	1996	2020	4435	12.2	1
NASA-U	73.8414	-49.5069	2334	1997	2021	5878	16.1	1
GITS	77.1378	-61.0400	1869	1996	2019	2023	5.5	1
Humboldt	78.5267	-56.8306	1995	1996	2022	5350	14.7	1
Summit	72.5794	-38.5053	3199	1996	2019	6463	17.7	1
TUNU-N	78.0164	-33.9833	2052	1996	2019	5937	16.3	1
DYE-2	66.4806	-46.2831	2099	1996	2022	7486	20.5	1
JAR1	69.4950	-49.7039	932	1996	2019	4568	12.5	1
Saddle	65.9997	-44.5017	2467	1998	2021	4383	12.0	1
SouthDome	63.1489	-44.8172	2901	1997	2020	4733	13.0	1
NASA-E	75.0006	-29.9972	2614	1997	2022	6716	18.4	1
NASA-SE	66.4750	-42.4986	2373	1998	2012	3181	8.7	1
NEEM	77.5022	-50.8744	2454	2006	2020	3169	8.7	1
			PROMICE					
KPC_L	79.9108	-24.0828	370	2008	2022	4004	11.0	2
KPC_U	79.8347	-25.1662	870	2008	2022	4909	13.4	2
EGP	75.6247	-35.9748	2660	2016	2022	2075	5.7	2
TAS_L	65.6402	-38.8987	250	2007	2022	4558	12.5	2
TAS_U	65.6978	-38.8668	570	2008	2015	2488	6.8	2
TAS_A	65.7790	-38.8995	890	2013	2022	2983	8.2	2
QAS_L	61.0308	-46.8493	280	2007	2022	5104	14.0	2
QAS_M	61.0998	-46.8330	630	2016	2022	1893	5.2	2
QAS_U	61.1753	-46.8195	900	2008	2022	4789	13.1	2
QAS_A	61.2430	-46.7328	1000	2012	2015	571	1.6	2
NUK_L	64.4822	-49.5358	530	2007	2022	5151	14.1	2
NUK_U	64.5108	-49.2692	1120	2007	2022	4290	11.8	2
NUK_N	64.9452	-49.8850	920	2010	2014	1372	3.8	2
KAN_L	67.0955	-49.9513	670	2008	2022	4881	13.4	2
KAN_M	67.0670	-48.8355	1270	2008	2022	4642	12.7	2
KAN_U	67.0003	-47.0253	1840	2009	2022	4517	12.4	2
UPE_L	72.8932	-54.2955	220	2009	2022	4506	12.3	2
UPE_U	72.8878	-53.5783	940	2009	2022	4497	12.3	2
THU_L	76.3998	-68.2665	570	2010	2022	3777	10.3	2
THU_U	76.4197	-68.1463	760	2010	2021	3457	9.5	2
CEN	77.1333	-61.0333	1880	2017	2021	1480	4.1	2

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press

Table 2. Mean bias (MB, $^\circ$ C) and root-mean-square error (RMSE, $^\circ$ C) in 2 m air temperature between models (MAR, RACMO, and ERA5), and daily observations at all sites for the whole study period (all) and the four seasons.

	all	MAM	JJA	SON	DJF
			MAR		
MB	0.89	1.27	0.13	0.82	1.52
RMSE	3.01	3.14	1.86	3.19	3.72
			RACMO		
MB	1.04	1.19	0.59	1.16	1.29
RMSE	2.68	2.79	1.72	2.93	3.17
			ERA5		
MB	1.11	1.16	0.11	1.46	1.89
RMSE	3.25	3.03	2.11	3.49	4.21

https://doi.org/10.1017/jog.2025.38 Published online by Cambridge University Press