



Performance of a 179-year high-resolution climate simulation of Southern Alaska

Sandra K. Königseder¹ · Timothy T. Barrows² · Jenny A. Fisher^{1,3} · Jason P. Evans⁴ · Youngil Kim⁴ · Chesley McColl⁵

Received: 9 March 2025 / Accepted: 13 June 2025 / Published online: 27 June 2025
© The Author(s) 2025

Abstract

Alaska has one of the world's largest glaciated areas and is highly sensitive to climate change. Alaskan glaciers currently contribute about a third of the global sea level rise, with tidewater glaciers playing a significant role through rapid retreat. Meteorological observations in this region are insufficient to assess climatic influences on the tidewater glacier cycle, and existing model datasets are either too coarse or cover too short a period. This study reconstructs the regional climate of southern Alaska by downscaling the NOAA-CIRES-DOE 20th Century Reanalysis (20CRv3) from 1836–2015 using the Weather Research and Forecasting model (WRF) to produce a high-resolution 4-km dataset. The new downscaled dataset (20CRv3-WRF) was validated for 1981–2015 against observational records (GSOD) and the Parameter-elevation Regression on Independent Slopes Model (PRISM) datasets and compared to European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5). Compared to the observational records, 20CRv3-WRF performed well for annual mean temperature ($0.61 \leq r \leq 0.96$) and moderately for annual precipitation ($0.16 \leq r \leq 0.76$). For temperature, 20CRv3 downscaling output was more consistent with PRISM than with the coarser resolution ERA5, suggesting a more accurate representation of temperature than the reanalysis. Precipitation was mostly overestimated in comparison to observations. The spatial variability of precipitation was better represented in 20CRv3-WRF than ERA5. The results demonstrate that 20CRv3-WRF provides a reliable high-resolution dataset to assess the influence of climate on southern Alaskan tidewater glaciers, enabling future studies requiring long-term atmospheric data.

1 Introduction

Global climate models (GCMs) are important for understanding climate processes, variability, change, and for performing climate experiments. GCMs are used to predict future climate and simulate past climate. Historical climate reanalyses have been developed for the period of observed climate history. They solve the problem of inconsistently

available and unevenly distributed observational records by using models to provide spatially and temporally continuous data. To do this, they use a method that combines forecast models with data assimilation from available observational records (Choudhury et al. 2023; Hersbach et al. 2020; Slivinski et al., 2019; Trenberth et al., 2008). Furthermore, reanalysis datasets are used to evaluate climate model simulations in conjunction with observational records. Reanalysis datasets (e.g., ERA5 (Hersbach et al. 2020), ERA5-Land (Muñoz-Sabater et al. 2021)) generally have higher resolution than GCMs. However, most climate reanalyses only start when the first upper air observations became available (1948), when radiosonde observation became regular (1958) or at the start of the satellite era (1979) (Compo et al., 2011; Kobayashi et al., 2015; Slivinski et al., 2019). The low resolution of GCMs and the short temporal coverage of reanalysis datasets makes studying long-term local climate difficult. In this work, we address this problem using the longest available reanalysis, the NOAA-CIRES-DOE 20th Century Reanalysis version 3 (20CRv3) spanning 1836 to

✉ Sandra K. Königseder
sandrak@uow.edu.au; koenigseder.sandra@gmail.com

¹ Environmental Futures, University of Wollongong, Wollongong, NSW, Australia

² Chronos Radiocarbon Laboratory, University of New South Wales, Sydney, NSW, Australia

³ College of Science and Engineering, James Cook University, Townsville, QLD, Australia

⁴ Climate Change Research Centre and ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia

⁵ NOAA Physical Sciences Laboratory, Boulder, CO, USA

2018 (Slivinski et al., 2019) at a spatial resolution of ~ 0.7 degree, to study the local climate of southern Alaska.

Global warming has raised mean surface temperatures by 0.99 ± 0.15 °C from 1850–1900 to 2011–2020 (Gulev et al. 2021). The high latitudes are more affected by climate change because of Arctic amplification (including feedbacks such as ice-albedo). For the period of 1979–2021 the Arctic experienced warming of four times as much as the global average (Rantanen et al. 2022). During 1957–2021, Alaska was shown to have a rate of warming more than a third higher than the warming rate for the contiguous United States (Ballinger et al. 2023). Local climate affects natural systems, and the effects of climate change vary with latitude. Warming causes glaciers to melt and retreat, consequently leading to sea level rise. During the period 1961–2016, melting glaciers contributed 27 ± 22 mm to global mean sea level rise, with a notable large contribution from Alaskan glaciers (Zemp et al., 2019). High latitude and remote regions typically lack long observational climate records because weather stations are difficult to maintain in these environments (Bieniek et al. 2012; De La France and McAfee 2019; Molnia 2007; Tangborn 1997). Homogenous high-elevation data covering at least 30 years is currently not available for this region (Kane and Stuefer 2015). Hence, with the current limitations of GCMs and reanalyses, changes in ice mass are often studied on either a global scale or a short-term local scale through field research, without knowledge of long-term local-scale changes.

Many glacierized regions are poorly covered by observations. Alaska has one of the largest temperate and subarctic glaciated areas in the world, which is highly sensitive to climate change, but with few long-term settlements. From 1961 to 2016, Alaskan glacier mass loss contributed about one third to the global mean sea level rise (Zemp et al., 2019). Southern Alaska accommodates ice fields and mountain, lake-terminating, and tidewater glaciers (Pfeffer et al., 2014). Changes in climate variations (especially 2-m air temperature and precipitation) have immediate effects on glaciers (Rasmussen et al., 2011b; Yde and Paasche, 2010). Tidewater glaciers in Alaska are retreating (Arendt et al., 2013; Black and Kurtz, 2023; Rasmussen et al., 2011a), but the processes triggering their retreat are not well understood. They are mainly influenced by factors such as glacier geometry (internal dynamics) and the underlying land (fjord bathymetry) (Enderlin et al., 2018; Meier and Post, 1987). However, the overall retreat of Alaskan tidewater glaciers that started during the last two centuries (Pfeffer, 2007) indicates that climate plays a larger role than previously thought (Enderlin et al., 2018; Post et al., 2011). The temporal cycle length of the tidewater glacier cycle is on the scale of centuries (Meier and Post, 1987). However, the weather station network in Alaska is sparse, with a limited time range of

meteorological observations, inhomogeneous time records (Bieniek et al., 2012; De La France and McAfee, 2019; Molnia, 2007; Tangborn, 1997) and a scarcity of observations at higher altitudes (Kane and Stuefer 2015). Therefore, using observations alone to investigate the climate influence on the tidewater glaciers of southern Alaska is not sufficient.

Dynamical downscaling as well as statistical downscaling have been previously applied over Alaska. Previous studies concentrated on northern Alaska (e.g., Cai et al. (2018); Poujol et al. (2020a)), the Alaskan interior (Mölders and Kramm, 2010), southeast Alaska (Lader et al. 2020), or Alaska as a whole (e.g., Bieniek et al. (2016); Cai et al. (2018); Hill et al. (2015); Lader et al. (2017); McAfee et al. (2014); Monaghan et al. (2018); Walsh et al. (2018)). The regional climate model used in those studies was the Weather Research and Forecasting model (WRF) (Powers et al., 2017; Skamarock et al., 2019), which is a commonly used downscaling regional climate model. WRF output for Alaska has been evaluated using either the limited available observational records or other reanalysis products (independent of the forcing data). These evaluations showed WRF was able to successfully reproduce variables such as 2-m air temperature, precipitation, and snowfall. Existing Alaskan downscaling studies focused on reconstructing modern-day climate or predicting future climate by downscaling GCM output. None attempted to reconstruct past climate prior to 1979, and therefore do not cover the time frame required to study the tidewater glacier cycle.

Before using model output for novel applications, it is necessary to quantify the skill and uncertainty of model output by evaluating it with observations or reanalyses (and considering their uncertainty as well). Downscaling and reanalysis products have been evaluated with observations (Bieniek et al. 2016; Cassano et al., 2001; Cassano et al., 2011; Evans et al., 2012; Jeworrek et al., 2021; Lavers et al. 2022; Maussion et al. 2011; Slivinski et al., 2021; Yu et al., 2021) and remote sensing data (Maussion et al. 2011; Monaghan et al. 2018). Regardless of which evaluation datasets are used, the statistical methods are usually similar (e.g., root mean square error, bias, difference, correlation).

This paper investigates the ability of a dynamical downscaling model to simulate high-resolution local climate over glacierized southern Alaska. To do so, we run the WRF model using the complete 179-year record of 20CRv3 reanalysis as forcing data. We then validate the model output with available observational records from the Global Surface Summary of the Day (GSOD) (NOAA 1999) and compare it to the ERA5 reanalysis (Hersbach et al. 2020) and PRISM (Parameter-elevation Regression on Independent Slopes Model) data. We focus our evaluation on 2-m air temperature and precipitation because of the importance of these variables for glacier mass balance.

2 Methodology

2.1 Simulation set up for 20CRv3-WRF downscaling

In this study, we downscale the 20CRv3 reanalysis product. 20CRv3 is the longest available climate reanalysis dataset (Slivinski et al., 2019), starting in 1836 and ending in 2015. We use the 20CRv3 ensemble mean as forcing data, rather than an arbitrarily chosen ensemble member. The ensemble mean of the 20CRv3 reanalysis was used to provide a consistent and representative boundary condition for long-term climate evaluation, minimizing biases from individual ensemble members while prioritizing stable forcing for the downscaling. The ensemble spread of the 20CRv3 is higher further back in time when fewer observational records for data assimilation are available (Slivinski et al., 2019). Because the observations are not equally distributed, the ensemble spread and, therefore the uncertainty, is higher in locations where fewer observations are available.

To conduct the 20CRv3 downscaling, we used WRF version 4.3, developed by the National Center for Atmospheric Research (NCAR) (Powers et al., 2017; Skamarock et al., 2019). WRF was run from 1836–2015 (with 1836 as spin up year). To run convection-permitting meteorology and improve the simulation of precipitation for regions with significant topography (Prein et al., 2013b), the resolution of the nest needed to be at least 4 km. Therefore, the domain set-up consisted of a 20-km resolution parent domain of 3300 km × 3600 km (165 × 180 grid cells) and a convection-permitting 4-km high-resolution nest over south-central/south-eastern Alaska with dimension 1744 km × 1164 km (436 × 291 grid cells) (Fig. S1). Both domains used a temporal resolution of 3 h. The full downscaling was conducted by running six 30-year simulations, with 1-year overlap for model spin-up, in parallel.

Prior to running the simulation, the most appropriate physics configuration was identified by forcing WRF with the 20CRv3 for 2010 using five different configurations (Table S1). The simulated temperature and precipitation of those five configurations were statistically evaluated against GSOD observations (r^2 , RMSE, normalized mean error, difference) and showed almost no difference. The WRF physics configuration used to downscale the 20CRv3 was based on the configuration used by Monaghan et al. (2018). The Thompson scheme was employed for microphysics, and Grell 3D scheme was used for the cumulus parameterization with the nest switched off. The RRTMG scheme was applied as shortwave and longwave radiation scheme, and the Yonsei University (YSU) scheme was used for the planetary boundary layer (PBL). MM5

Similarity was applied as the surface layer, and the land-surface model Noah-MP was used. The 20CRv3-WRF simulation used the default Noah-MP settings for snow, which includes snowpack representation in three layers. Additionally, the Noah-MP glacier treatment option was activated, which includes phase changes by improving the snow physics (Niu et al., 2011). This means that melting seasons are considered for glaciers.

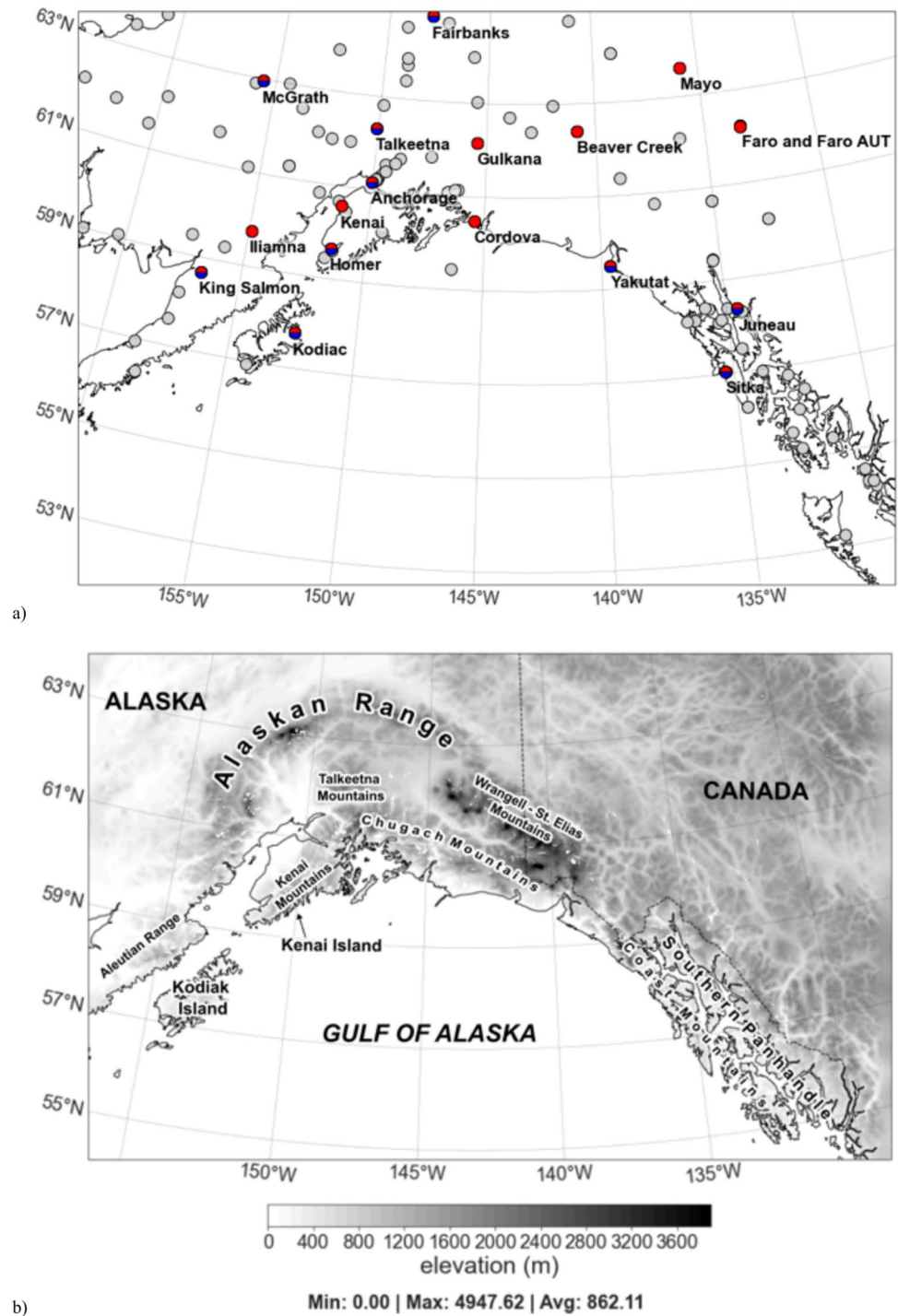
2.2 Evaluation datasets

Daily average 2-m air temperature (TEMP) and daily accumulated precipitation (PRCP) from NOAA GSOD were used to evaluate the simulation. For the study area, 22 GSOD stations recorded data during the years 1986–2015. 104 stations did not record for the full period, with at least one year of missing data (Fig. 1). One station was within a water grid cell and was excluded from the analysis. For temperature, stations that recorded data for at least 85% of each year were included (18 stations; see Table S2). For precipitation, stations that recorded data for at least 70% of each year were used for the analysis (10 stations; see Table S2).

The second dataset used for model evaluation was the PRISM data from the PRISM Climate Group, Oregon State University (PRISM 2023b). This spatial climate dataset for the United States is based on statistical methods that interpolate observational records on a digital elevation model, combined with human expertise (Daly et al. 2002, 2018). Gridded 30-year mean temperature and 30-year accumulated precipitation data from 1981–2010 (PRISM 2023a) were transformed into netcdf format and re-gridded to the WRF grid. PRISM only provides 30-year means, and so annual and seasonal analysis was not performed with the PRISM dataset.

The model output was also compared to reanalysis data. This is an approach often used if observational records are not independent (Slivinski et al. 2021) or non-existent (Choudhury et al. 2023). The reanalysis datasets enable a gridded evaluation at annual and sub-annual timescales. For the reanalysis comparison, the ERA5 reanalysis was used (Hersbach et al. 2020). Annual and 30-year mean 2-m air temperature were calculated for 1986–2015 from the monthly mean 2-m air temperature. Monthly mean precipitation data was processed to monthly accumulated precipitation and summed to annual accumulated precipitation. From this, the 30-year mean annual accumulated precipitation for 1986–2015 was calculated. These processes were conducted on the native ERA5 grid. The data was then re-gridded to the WRF grid using bilinear interpolation to match the downscaling data.

Fig. 1 (a) GSOD station locations. Stations that recorded temperature for at least 85% of each year from 1986–2015 are shown in red (note: Faro has two stations), and stations that also recorded precipitation for at least 70% of each year from 1986–2015 are shown in blue (half circle). Stations in grey had incomplete records and were excluded from the analysis. (b) Map of mountain ranges and islands referred to in this chapter, with the Alaska-Canada border shown as the dashed black line



2.3 Evaluation method

Annual mean 2-m air temperature and annual accumulated precipitation from the downscaling simulation were evaluated. From here onwards, these variables are referred to as ‘temperature’ and ‘precipitation’, respectively.

The evaluation metrics used in this study are root mean square error (RMSE), square of the Pearson correlation

coefficient (r^2), normalized mean error (NME), time-averaged difference (DIFF) and standard deviation (σ). RMSE is defined in Eq. 1:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - o_i)^2} \quad (1)$$

where N is the number of data points m_i represents an individual model (WRF) value at time i , and o_i represents an individual observation at time i .

r was used to assess the linear relationship of day-to-day or year-to-year correlation of the values simulated by the model to the values from the evaluation datasets. r ranges from -1 to 1 and is ideal at 1 . Negative values show anti-correlation. r is defined in Eq. 2:

$$r = \frac{\sum_{i=1}^N (m_i - \bar{m})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (m_i - \bar{m})^2} \sqrt{\sum_{i=1}^N (o_i - \bar{o})^2}} \quad (2)$$

where $\bar{m}$ is the mean of the model (WRF) dataset sampled at observation locations and $\bar{o}$ is the mean of the observations. r^2 was used to assess the proportion of variation of the values simulated by WRF in relationship to the evaluation dataset. Values range from 0 to 1 , with 1 being ideal.

NME was calculated to identify the extent to which the simulated values differ from the evaluation dataset. NME is a positive value that, when equal to 0 , means the simulated value is the same as the evaluation dataset value. The higher the NME, the less similarity between the simulated and evaluation dataset values. NME is defined in Eq. 3:

$$NME = \frac{\sum_{i=1}^N |m_i - o_i|}{\sum_{i=1}^N |\bar{o} - o_i|} \quad (3)$$

DIFF was used to assess the absolute magnitude of the similarity between the values simulated for a variable by WRF and the values of that variable from the evaluation dataset, defined in Eq. 4:

$$DIFF = \frac{1}{N} \sum_{i=1}^N (m_i - o_i) \quad (4)$$

σ explains the variability of the values that are simulated for a given variable. The standard deviation of the simulated values produced by WRF was compared with that of the values from the evaluation datasets. σ was defined in Eq. 5:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (m_i - \bar{m})^2}{N}} \quad (5)$$

Downscaling output was evaluated at the point locations of the GSOD stations (Fig. 1a and Table S2) by comparing the 20CRv3-WRF with available GSOD observation station over the period 01-Jan-1986 to 31-Dec-2015. For the available locations, the model performance was assessed using RMSE, r^2 , NME, DIFF and standard deviation (σ) relative to the observations.

For ERA5, the modeled recent climate (01-Jan-1986–31-Dec-2015) was evaluated by comparing the 20CRv3-WRF downscaling output to gridded ERA5 reanalysis data. The

time-averaged difference in 30-year mean temperature and accumulated precipitation was calculated by subtracting the re-gridded ERA5 from the downscaling output. r was calculated for annual temperature and precipitation in each grid cell. For PRISM, the difference between the 20CRv3-WRF and PRISM for 30-year (01-Jan-1981–31-Dec-2010) mean temperature and 30-year mean accumulated precipitation was calculated.

3 Results

3.1 Evaluation using GSOD station data

3.1.1 Temperature

Available station records are evenly distributed across the domain (Fig. 1). Figure 2 compares the downscaling output with the input reanalysis data (20CRv3), ERA5 and GSOD observational data for four selected sites near tidewater glacier locations. The figure shows that the 20CRv3-WRF is an overall improvement relative to the 20CRv3. The 20CRv3 exhibited a larger difference compared to the GSOD data, but this was reduced by the 20CRv3-WRF downscaling, especially for Juneau and Anchorage.

The 20CRv3-WRF temperatures are highly correlated with the GSOD station records (Fig. S2, Fig. 3 and 4). This can be seen on different time scales and shows that weather as well as climate are realistically predicted by the downscaling (Fig. S2). The model performs best for temperatures at eastern interior locations (Mayo, Faro, Beaver Creek, Gulkana) (Fig. 3 and 4). At these locations, the 20CRv3-WRF produces on average a warm bias, but with low normalized mean errors and high correlation. Overall, the correlation is similar across all stations.

Figure 3 (a–g) shows the annual temperature for 20CRv3-WRF and the observations for a subset (seven representative stations) of the 18 analyzed continuous recording stations from 1986–2015 (Fig. S3 and S4 includes the remaining stations), along with the annual mean difference between them. This Figure also shows the average over the 1986–2015 time period for each day of the year to provide an annual cycle of 30-year mean daily temperature from 20CRv3-WRF and the observations, along with the difference between model and observations (Fig. 3h–n). The differences in average temperatures can mainly be explained by over- or underestimation by the model in specific seasons. For example, for Anchorage, the low annual mean temperature produced by the model is caused by a cold bias of temperatures in summer and autumn (Fig. 3c and j). However, in some cases, for example Sitka, the model shows a temperature bias in all seasons, which results in the lower annual mean temperature.

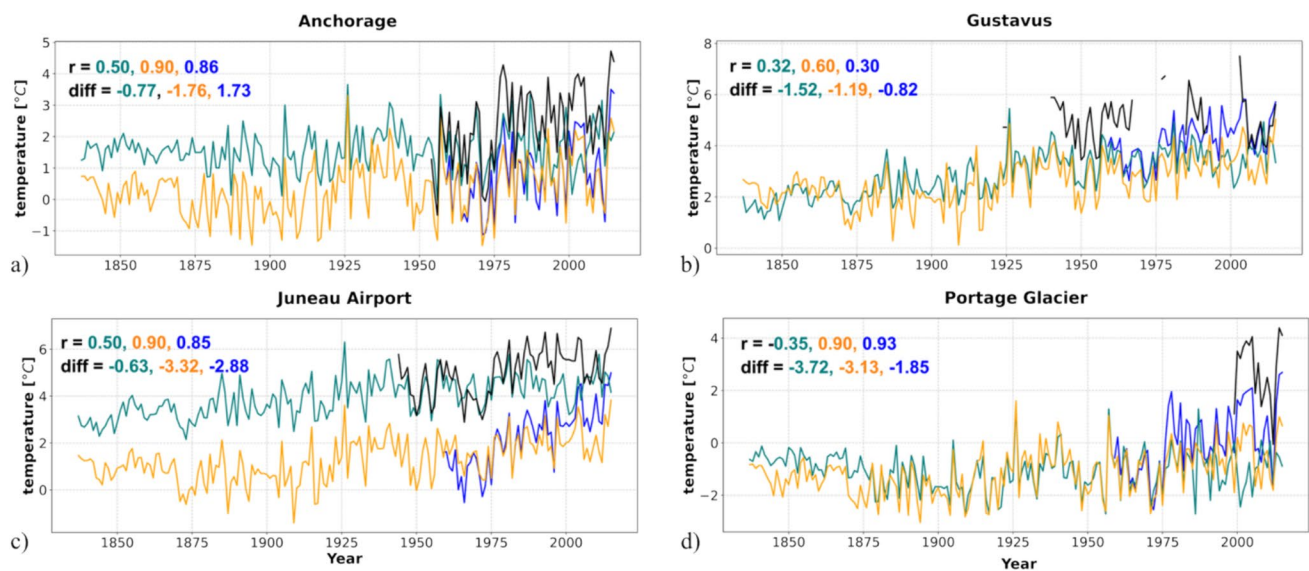


Fig. 2 Comparison of 20CRv3-WRF downscaling output (teal) to its 20CRv3 input data (orange), ERA5 (blue), and GSOD observational data (black) for four selected stations close to tidewater glacier sites:

(a) Anchorage, (b) Gustavus, (c) Juneau Airport and (d) Portage Glacier. The observational records cover different time spans and contain some missing data

The downscaling simulates a cold bias for coastal regions (Iliamna, King Salomon, Cordova, Anchorage, Yakutat, Sitka with the highest negative bias, see Fig. 3). On the other hand, the downscaling overestimates temperature at eastern interior stations (Mayo, Beaver Creek and Faro) during winter (Fig. 3n and Fig. S4). Central interior stations (Gulkana, Talkeetna and McGrath) show no particular seasonal pattern of bias.

Figure 4 presents Taylor diagrams that summarize the performance of the 20CRv3-WRF for simulating daily and annual temperatures. The diagrams include the statistics that are used to evaluate the model output with the observations (r , RMSE, and the standard deviation of the model with a referenced to the mean standard deviation of the observations). Each point represents one station and is colored by NME. Daily temperature correlations (Fig. 4a) show less spread than annual temperatures correlation (Fig. 4b), indicating better agreement in the timing of weather events. However, the standard deviation shows greater spread than annual temperatures, reflecting differences in magnitude. This is expected because the reanalysis assimilates pressure anomalies, which are directly related to synoptic weather systems. For Faro, located in the eastern interior of the domain, the model shows the worst performance for annual temperature (Fig. 4b). For some of the coastal locations (Yukata, Sitka, Juneau, Kenai), the model performs less realistically on both daily and annual time scales (except Kenai) (Fig. 4b). However, for other coastal stations (Homer, Kodiak), it performs better for annual than for daily temperature correlations. At other stations within the east (Iliamna, King Salmon,

Cordova, Anchorage), 20CRv3-WRF performs better (daily and annual). Overall, the model performs better in the area around the west coast compared to the east coast stations. Moderate performance for annual temperature is observed in stations of the interior (Fairbanks, Talkeetna, McGrath, Beaver Creek, Mayo).

The 20CRv3 downscaling performs the best for temperature in winter (Fig. 4). The performance of each station varies with the seasons. For example, the 20CRv3-WRF simulates winter temperatures well for Kenai but for the autumn and summer, this location is one of the worst performing. In contrast, the model performs moderately at simulating temperatures for Iliamna throughout all seasons, with the best performance in autumn. This pattern is similar to that at King Salmon, where the best results are seen in summer and performance is moderately good throughout the other seasons. Mayo performs best for spring and winter, but poorly in autumn and summer.

There is no specific seasonal pattern between the performance of western and eastern stations. For autumn and winter, the central and eastern coastal stations (Anchorage, Kenai, Cordova, and King Salmon, Iliamna, Homer, Kodiak) are part of the better performing stations with the central coastal stations performing well for winter. For autumn and winter the 20CRv3-WRF simulates temperatures least reliably for the central interior stations (McGrath, Fairbanks, and Gulkana). Kenai, Cordova and Beaver Creek (spread over the domain) show the poorest temperature performance in spring. This spatially unrelated performance is similar in summer, with the worst performing stations being Yakutat (west coast), Kenai (central east coast) and Mayo (western

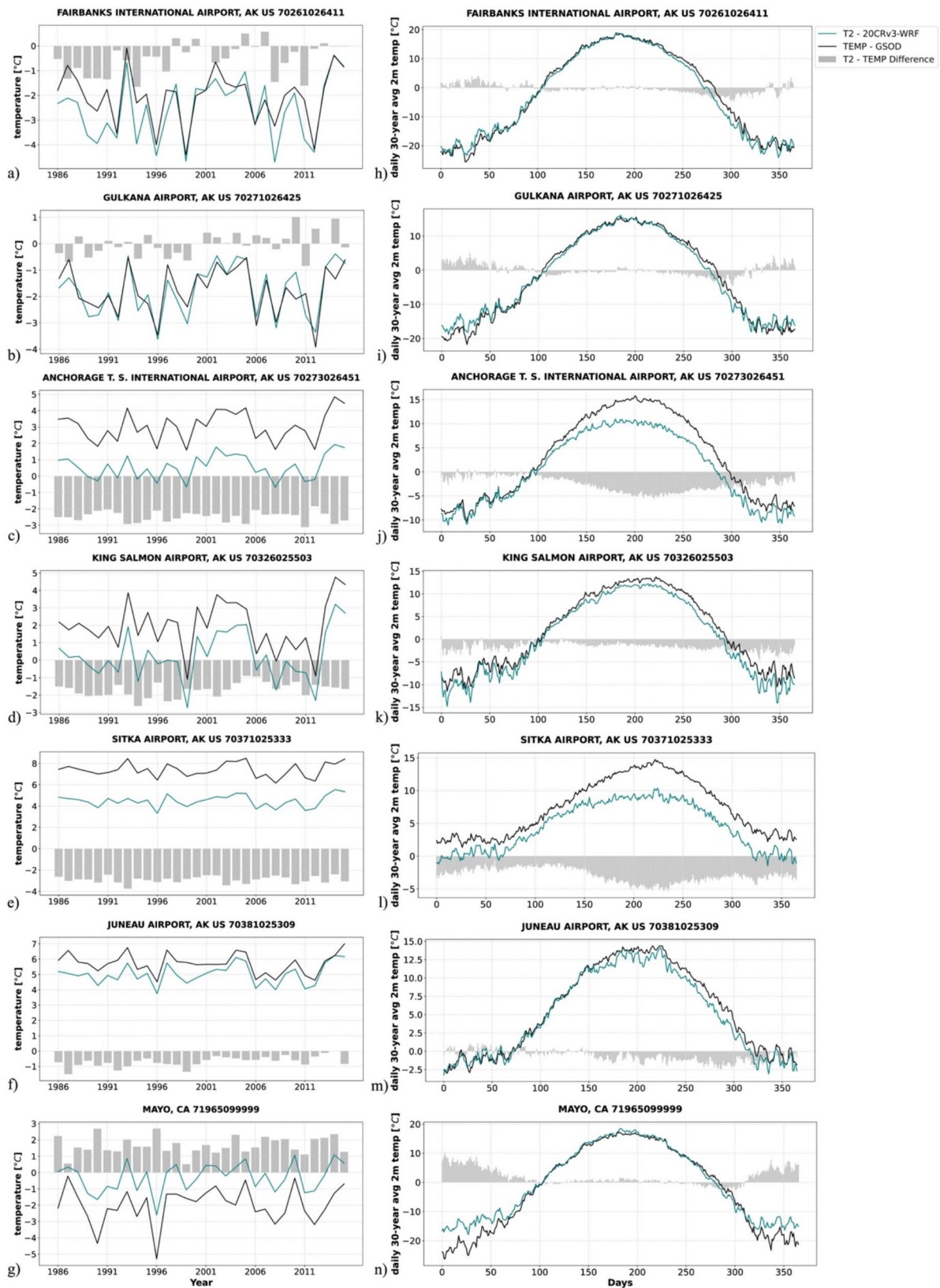


Fig. 3 Left (a–g)—annual mean temperature for the 20CRv3-WRF downscaling (teal lines) and GSOD observations (black lines), and the difference (model – observations, grey bars). Right (h–m)—daily 30-year (1986–2015) mean temperature for 20CRv3 downscaling modelled data (teal) and GSOD observations (black), and the difference (model – observations, grey bars). For the selected stations: (a,h) Fairbanks International Airport, AK US; (b,i) Gulkana Airport, AK US; (c,j) Anchorage Ted Stevens International Airport, AK US; (d,k) King Salmon Airport, AK US; (e,l) Sitka Airport, AK US; (f,m) Juneau Airport, AK US; (g,n) Mayo, CA. AK, US for Alaska, United States of America, and CA for Canada

interior). For autumn, the set of worst performing stations are different, but again, show a spread over the domain (Sitka, Kenai, Mayo) while for winter the worst performing stations are all within the interior (McGrath, Fairbanks).

For the coastal and interior region, a pattern for standard deviation could be identified, however not for the other metrics. For spring, the coastal stations (apart from King Salmon) show lower standard deviation, which means the range of daily temperatures is less than the range in observations. This is a similar pattern in summer, with King Salmon being the only coastal station that exceeds the diversity in daily temperatures. In autumn, most stations exhibit greater variability in simulated temperature than in observational records compared to other seasons. For interior stations, simulated temperatures generally show more variability than observed values during spring (except Talkeetna and Gulkana), summer, and autumn. For winter, the stations differ and some show higher (McGrath, Fairbanks, Gulkana and Beaver Creek) and lower variability (Talkeetna, both Faro stations, and Mayo).

3.1.2 Precipitation

The modeled precipitation shows relatively poor performance with the observations in comparison to temperature. Agreement improves when the data is averaged over longer timescales. Precipitation shows higher correlation with GSOD for annual accumulated precipitation than for daily accumulated precipitation (Fig. S5). The inaccuracy in simulating daily precipitation can be related to temporal location error (Barros and Lettenmaier 1993; Cassano et al. 2016), where the precipitation at the end of the day is simulated a few hours too late and assigned to the next day, causing both days to be inaccurate. Better correlation occurs for annual accumulated and 30-year average annual accumulated (Fig. 5) precipitation. For nearly all stations, the model produces a positive precipitation bias.

Figure 5 shows the annual accumulated precipitation for the 20CRv3-WRF and GSOD as well as its annual difference. Some stations generally overestimate precipitation (McGrath, Fairbanks, King Salmon and Juneau), whereas

other stations (Kodiak, Yakutat and Sitka) underestimate precipitation. The 30-year daily accumulated precipitation for each station is also presented in Fig. 5. Some coastal stations (Yakutat, Sitka, and Juneau) show a dry bias in summer, and all stations produce a wet bias in winter and spring (Fig. 5).

The performance of the 20CRv3-WRF precipitation modeling is shown in Fig. 6 at daily (a) and annual (b) resolution. On a daily timescale, 20CRv3-WRF performs least well at Yakutat, Juneau, and Sitka. The annual performance shows no regional patterns, with the worst performance at Yakutat, Anchorage, and Homer. The daily variability is lower in the model than in the observations for regions that are furthest from the ocean (McGrath, Talkeetna, Fairbanks, Anchorage, King Salmon and Homer). In contrast, Fig. 6 shows model variability is higher than observed at Kodiak, Yakutat, Sitka, and Juneau stations. On annual timescales, the model produces lower variability compared to the observational records for most stations (apart from Yakutat and Sitka) which may be related to the use of ensemble mean boundary conditions.

Overall, the evaluation of 20CRv3-WRF with station-based observational records shows that for precipitation, the downscaling performs the best in spring and winter (Fig. 6). There is no consistent pattern for a given station's performance in simulating precipitation. For example, Kodiak is one of the best simulated stations for nearly all seasons (apart from autumn) while Anchorage is one of the worst locations in spring, winter and summer. However, McGrath shows moderate performance throughout all seasons. The central interior stations (McGrath, Talkeetna, Fairbanks) consistently produce lower variation in precipitation in the model than observations for all seasons. Central interior stations are amongst the better performing stations for autumn and winter but less in spring and summer. The easternmost coastal stations, Sitka and Juneau, perform best in spring but worst in winter. Kodiak is the best performing station for spring and summer but performs worse in autumn. Yakutat differs most from the other eastern coastal stations (Sitka and Juneau) by having a high standard deviation.

3.2 Evaluation using PRISM and ERA5 gridded data

3.2.1 Temperature

The 20CRv3-WRF downscaling produces slightly lower temperatures than PRISM (Fig. 7). The mountainous areas show a mixed bias (Fig. 7), possibly explained by the higher resolution PRISM data (800 m). Within the Alaskan Ranges, some 20CRv3-WRF grid cells have warmer temperatures than PRISM and surrounding ones show colder temperatures. Higher resolution enables better representation of colder

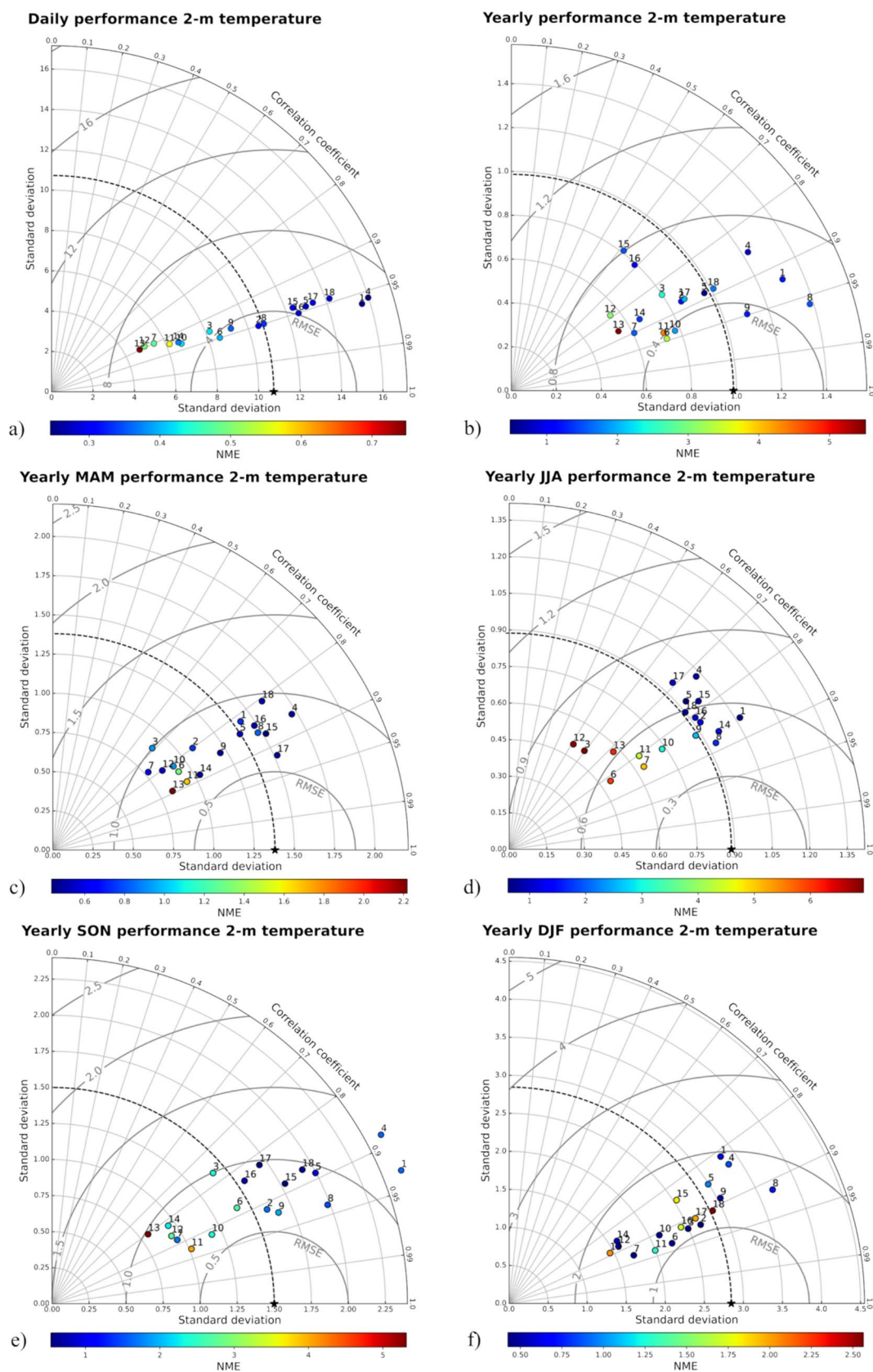


Fig. 4 Daily (a), annual (b) and seasonal ((c) spring—MAM, (d) summer—JJA, (e) autumn—SON, (f) winter -DJF) performance for 2-m temperature—1. McGrath Airport, AK US; 2. Talkeetna Airport, AK US; 3. Kenai Airport, AK US; 4. Fairbanks International Airport, AK US; 5. Gulkana Airport, AK US; 6. Anchorage Ted Stevens International Airport, AK US; 7. Cordova Airport, AK US; 8. King Salmon Airport, AK US; 9. Iliamna Airport, AK US; 10. Homer Airport, AK US; 11. Kodiak Airport, AK US; 12. Yakutat Airport, AK US; 13. Sitka Airport, AK US; 14. Juneau Airport, AK US; 15. Faro AUT YT, CA; 16. Faro, CA; 17. Mayo, CA; 18. Beaver Creek Airport, CA. AK, US for Alaska, United States of America, and CA for Canada

temperatures at high elevation, a result that was also observed in the comparison between the downscaling and ERA5.

Figure 8 compares the 30-year (1986–2015) mean temperature from 20CRv3-WRF to ERA5. Differences based on resolution are clearly visible in the figure. The 20CRv3-WRF, with its higher resolution and more detailed model topography than ERA5, presents more temperature variations due to elevation. The downscaling, with its complex topography along the coast that includes fjords and valleys, produces higher temperatures in the lowlands of the coast than ERA5. Colder temperatures in the higher altitude mountain regions along the coast are also simulated by the downscaling. On average, the 20CRv3-WRF produced noticeably lower temperatures (about $-6\text{ }^{\circ}\text{C}$) than ERA5 within the Chugach Mountains, Talkeetna Mountains, Alaska Range and Kenai Mountains. In most of the area around the mountain ranges, 20CRv3-WRF produced higher temperatures than ERA5. This is because the coarser 31-km resolution of ERA5 averages the colder temperature of the mountains with the warmer temperatures of the foothills. Therefore, differences between the 20CRv3-WRF and ERA5 are largely explained by resolution and model topography. The comparison also shows other regional differences between 20CRv3-WRF and ERA5. The downscaling produced higher temperature in the interior and east, except in the Southern Panhandle which, especially in winter, shows lower temperatures simulated by the 20CRv3-WRF. Temperatures in the Southern Panhandle are less consistent between 20CRv3-WRF and ERA5 than they were between 20CRv3-WRF and PRISM. In general, the 20CRv3-WRF simulates lower temperature in the west and over water. Within the area of Sitka, the downscaling shows lower temperatures in comparison with ERA5 and GSOD. On average, the difference over the domain is $-0.71\text{ }^{\circ}\text{C}$; however, when excluding water, the difference between 20CRv3-WRF and ERA5 reduces to $0.07\text{ }^{\circ}\text{C}$.

The correlation (r) for temperature between the 20CRv3-WRF and ERA5 is presented in Fig. 9. In the east and in mountainous areas (especially the Alaskan Ranges and Chugach Mountains), the correlation is the lowest. The highest correlation is found in the west of the domain, especially on Kodiak Island and the surrounding King Salmon area. This is also the case for Glacier Bay and around Gustavus, where

20CRv3-WRF temperatures correlate highly with the reanalysis. Lower correlation in the east interior where Faro is located is consistent with the GSOD comparison. Within the fjords of the Southern Panhandle, the correlation is also low.

3.2.2 Precipitation

The comparison with PRISM precipitation data is shown in Fig. 10. The difference is highest within the coastal regions, which show both positive and negative biases. This contrasts with the interior, where 20CRv3-WRF overestimates precipitation compared to PRISM. This result is consistent with the comparison with GSOD (Sect. 3.1.2) where the 20CRv3-WRF overestimates precipitation for the interior stations McGrath, Fairbanks and Talkeetna.

20CRv3-WRF and ERA5 have similar spatial patterns (Fig. 11), with the biggest differences found in areas of complex topography. Both show high accumulation of precipitation in the mountain region adjacent to the Gulf of Alaska, consistent with the precipitation pattern in PRISM. This is due to the mountains blocking the moisture transfer from the ocean to the land interior leading to orographic precipitation. However, ERA5, with its coarse resolution, only shows very high precipitation accumulation ($> 5000\text{ mm yr}^{-1}$) in the western Chugach Mountains and the northern end of the Coast Mountains (the mountain range along the coast of the Southern Panhandle). In the 20CRv3-WRF, accumulation this high is present along all the mountain ranges. 20CRv3-WRF-simulated precipitation in the Alaskan Ranges agrees better with PRISM than with ERA5. The downscaling product and ERA5 show an extreme drop in precipitation on Kenai Peninsula resulting from the Kenai mountains blocking the moist air, but 20CRv3-WRF simulates much higher precipitation amounts.

As a result of the coarse resolution, ERA5 predicts high precipitation over the water in coastal regions adjacent to the mountain ranges (Aleutian Range, Kenai mountains, Chugach Mountains, Wrangell-St. Elias Mountains, Coast Mountains). 20CRv3-WRF shows increased precipitation beginning once the air masses hit the mountain ranges, with no extension over the coastal waters. ERA5 has higher precipitation around mountain ranges, while precipitation in the downscaling is more localized, again explained by the resolution difference. ERA5 lacks representation of precipitation within regions of complex topography and overall has lower precipitation amounts than the 20CRv3-WRF.

Overall, 20CRv3-WRF simulates more precipitation relative to ERA5 (Fig. 11). ERA5 produces higher precipitation along the coast surrounding the mountains and coast, which is caused by model resolution. On average over the domain, including the ocean and water, the difference maps show that 20CRv3-WRF produces slightly lower precipitation (-5 mm yr^{-1}) than ERA5. Differences are the highest in autumn and winter (Fig. 11).

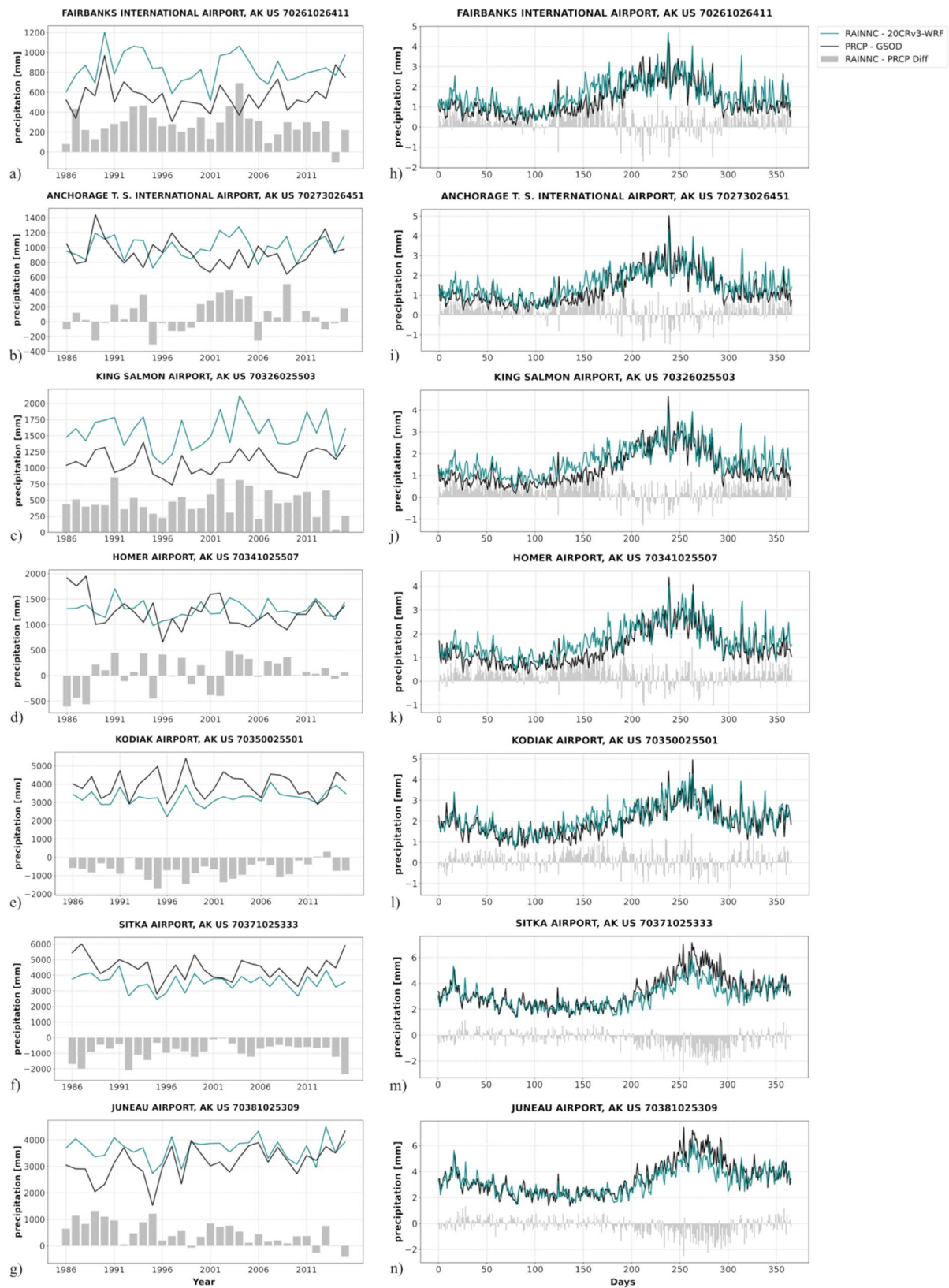


Fig. 5 Left (a–g)—yearly accumulated precipitation for the 20CRv3-WRF downscaling (teal) and GSOD observations, with the difference (model – observation) for each year presented in the grey bars. Right (h–n)—daily 30-year (1986–2015) average precipitation for 20CRv3 downscaling modelled data (teal) and GSOD observations (black), and the difference (model – observations, grey bars). For the stations: (a,h) Fairbanks International Airport, AK US; (b,i) Anchorage Ted Stevens International Airport, AK US; (c,j) King Salmon Airport, AK US; (d,k) Homer Airport, AK US; (e,l) Kodiak Airport, AK US; (f,m) Sitka Airport, AK US; (g,n) Juneau Airport, AK US. AK, US for Alaska, United States of America

Annual as well as seasonal accumulated precipitation between the ERA5 and the 20CRv3-WRF are overall positively correlated (Fig. 12). However, they do not agree as well for precipitation as they did for temperature, and are anti-correlated in some locations, for example on the leeward side of the Kenai Peninsula and in parts of the eastern interior. Precipitation is least correlated between the 20CRv3-WRF and ERA reanalysis in summer and most correlated in winter. Precipitation from the models is less correlated on land than over water (Fig. 12 and Fig. S8). The interior shows patches of non-statistically significant results, mainly in spring and summer. For autumn, the results are more robust, and for winter, only the Kenai Peninsula correlation is not statistically significant. Precipitation in southern Alaska is the highest in autumn and winter, the seasons for which 20CRv3-WRF produces the best results. In all seasons, correlation is low on the leeward side of Kenai Peninsula. This feature can be explained by the detailed topography of the downscaling, which creates a boundary for moist air on the peninsula. Although temperature has low correlation in the Southern Panhandle in spring, this is the opposite for precipitation. The similarity between 20CRv3-WRF and reanalysis in this region is highest in spring.

4 Discussion

This study presents an evaluation of downscaling 20CRv3 using WRF over southern Alaska. It is important to note that the evaluation should be considered in the context of the limited available station records and the limitations of the reanalysis product. Overall, the downscaled model reproduced station data (GSOD) well with little bias. Spatially, the model reconstructed gridded climate data (PRISM) well and was most different from the ERA5 reanalysis results. This shows that the 20CRv3-WRF produced temperature and precipitation realistically.

Topography explains much of the difference between reanalysis and observations. The 20CRv3-WRF downscaling resolves topography in finer detail (4 km) than the coarser (31 km) ERA5 reanalysis. Given that temperature is a function of elevation, disparities in modelled elevation as

compared to station elevation can contribute to variations in simulated temperatures (Monaghan et al. 2018). PRISM, with a spatial resolution of 800 m, largely preserves the elevation of the stations with minimal smoothing. The average temperature appears to be warmer than that obtained through downscaling, potentially due to differences in resolution and, therefore, elevation.

Precipitation is often overestimated by climate models in general, including reanalyses such as 20CRv3 and ERA5. Precipitation in 20CRv3 is overestimated at high altitudes relative to datasets based on station and satellite measurements (Slivinski et al. 2021). However, precipitation is not an input variable for the downscaling, and so the 20CRv3 precipitation biases do not directly affect the downscaling results. Precipitation modelled by WRF is, however, still dependent on other input variables that are related to the modelled 20CRv3 precipitation. Consequently, this contributes to a higher precipitation in the 20CRv3-WRF downscaling than seen in the observations.

The downscaling also inherits the uncertainty of the input data, which provides the lateral boundary conditions (Bruyère et al. 2013; Errico et al. 1993). 20CRv3 has been evaluated, and its overall performance compared to observations, satellite products and other reanalysis products was assessed by Slivinski et al. (2021). The authors concluded that the 20CRv3 reanalysis successfully estimates mass and circulation fields with decadal variability and produces fields that are more accurate in the Northern Hemisphere than in the Southern Hemisphere. It is worth noting that 20CRv3 assimilates surface pressure observations, and these have been found to better constrain the climate during winter than during summer, resulting in higher ensemble spread (an indicator of uncertainty) during the warm season (Yu et al., 2022). Hawkins et al. (2023a, 2023b) demonstrated how the incorporation of additional historical observations improved the representation of a historical windstorm in the 20CRv3 reanalysis. As a result, the storm Ulysses was accurately captured in the improved dataset. This improvement underscores the value of assimilating a greater number of high-quality observations into the reanalysis system. Therefore, the inclusion of standardized, long-term climate records from Alaska (as done for Antarctica Wang et al. (2021) and Wang et al. (2023)) would help fill observational gaps and enhance the accuracy of reanalysis input for future downscaling efforts in this region.

The 20CRv3-WRF downscaling better represents observed temperature than the original 20CRv3. Zhang et al. (2021) found that for mainland China, temperature from 20CRv3 is less accurate than from ERA5; however, their results show that 20CRv3 performed better at higher elevation (3000–5000 m) than ERA5. It is therefore possible that our finding of lower correlation between 20CRv3-WRF and ERA5 in the Alaskan Ranges (which includes areas above

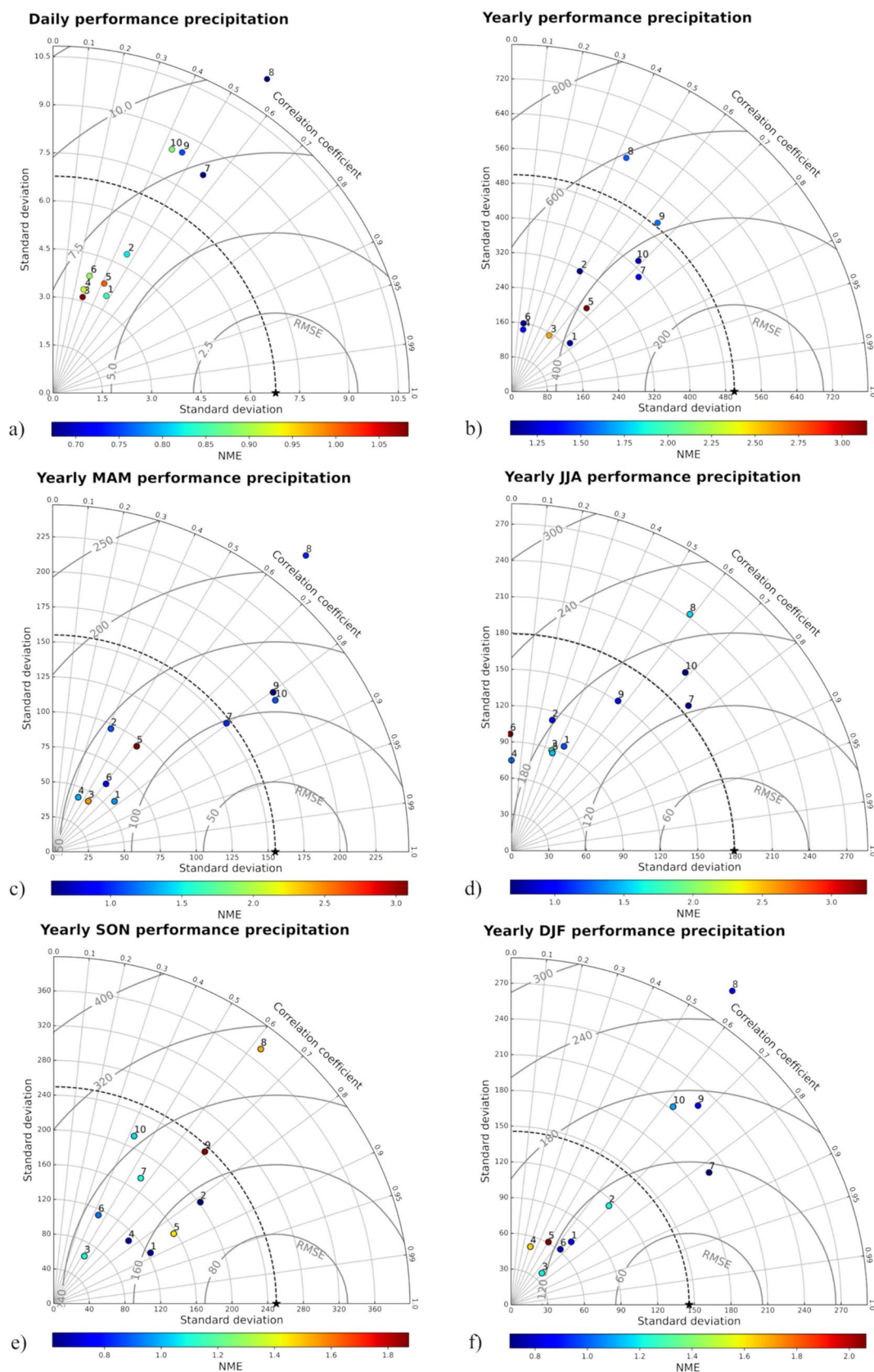


Fig. 6 Same as Fig. 5 for daily (a), annual (b) and seasonal ((c) spring—MAM, (d) summer—JJA, (e) autumn—SON, (f) winter—DJF) accumulated precipitation—1. McGrath Airport, AK US; 2. Talkeetna Airport, AK US; 3. Fairbanks International Airport, AK US; 4. Anchorage Ted Stevens International Airport, AK US; 5. King Salmon Airport, AK US; 6. Homer Airport, AK US; 7. Kodiak Airport, AK US; 8. Yakutat Airport, AK US; 9. Sitka Airport, AK US; 10. Juneau Airport, AK US. AK, US for Alaska, United States of America

3000 m) could be due to the forcing data, in addition to the resolution effects discussed previously. Colder temperatures in 20CRv3-WRF in the mountainous areas where glaciers are located might also be explained by a temperature overestimate in ERA5, as has been identified by He et al. (2022) for glaciers in China. In the northeast US, previous work showed the coarse resolution 20CRv3 reanalysis was unable to capture warming in coastal regions, where topographic variations include both lowlands and higher elevations (Karmalkar and Horton 2021). The coastline of the Gulf of Alaska has a similar topography.

The evaluation of 20CRv3-WRF against PRISM supports the conclusion that the downscaling produces realistic results in simulating temperature. In addition, the complex topography and differences between datasets in model elevation likely contribute to the low correlation in areas of complex terrain such as the Southern Panhandle. This could mean that, due to the higher resolution, the downscaling is an improvement in the coastal mountain area relative to ERA5, even though it inherits the coastal disadvantage of 20CRv3.

Previous ERA5 evaluations in both the study region and similar geographic locations (He et al. 2022; Song et al. 2021; Zhang et al. 2021) found inaccurate temperature on large glaciers (He et al. 2022), overestimated snow accumulation (Song et al. 2021) and poorer performance in simulating temperature at high elevation compared to 20CRv3 (Zhang et al. 2021). Our evaluation of 20CRv3-WRF against ERA5 presented above shows lower correlation within coastal areas with more temperature variation, and the difference shows that 20CRv3-WRF produces higher temperatures than ERA5 in some coastal areas. However, compared to GSOD observations, the downscaling underestimates summer and autumn temperatures in coastal locations. Within the Southern Panhandle, temperatures of ERA5 are higher than in the downscaling. This region shows low correlation between the datasets, which might be explained by a bias in ERA5. The comparison to observations in this region is limited to two stations (Sitka and Juneau). For Sitka, 20CRv3-WRF produces a consistent cold bias, while for Juneau, the average cold bias is smaller and concentrated in winter. The Juneau region is cooler in ERA5 than in 20CRv3 and GSOD. The results of 20CRv3-WRF and PRISM within the Southern panhandle are more consistent.

Precipitation spatial variability is improved by the 20CRv3-WRF downscaling compared to ERA5, even though the downscaling simulation overestimates precipitation in comparison to GSOD observations. Li et al. (2020) showed that WRF downscaling produces higher but more realistic precipitation than reanalysis products in high mountain glacierized areas. Regions of wet bias are especially common within the mountain regions in the 20CRv3-WRF. This pattern might be explained by the resolution, as the bordering grid cells show lower precipitation in comparison to the coarser reanalysis (with high precipitation variability averaged out by the reanalysis). Another explanation could be that ERA5 is known to overestimate solid precipitation in the foothills of mountains in southern Alaska (Song et al. 2021). The 20CRv3-WRF downscaling produces higher precipitation within the mountains compared to PRISM; however, the difference is less than ERA5, which is known to have a wet bias (Lavers et al. 2022). Lavers et al. (2022) evaluated ERA5 precipitation globally and found a general wet bias, but the evaluation was dependent on available observation stations, which are rare in Alaska. They found that the Spearman's correlation of ERA5 precipitation with observations from the few stations available in Alaska ranged from 0 to 1 with seasonal differences, and biases varied between -2.5 and 5 mm d^{-1} (for the studied months January, April, July, and October) for the period from 2001 to 2020 (Lavers et al. 2022). Further evaluation of ERA5 precipitation would be needed to better understand its accuracy in Alaska. The fact that 20CRv3-WRF produces higher precipitation than PRISM but less than ERA5 suggests that it might still have a wet bias, though potentially less severe than ERA5's.

On average, precipitation simulated by the 20CRv3-WRF downscaling model in the study domain is slightly lower in comparison to ERA5 but higher than PRISM. However, accuracy improves when annually accumulated. Weather stations that use gauge data underestimate solid precipitation because of wind blowing it, a phenomenon known as gauge undercatch (Goodison et al. 1998). This is the case in Alaska, where most precipitation in winter is snow (Bieniek et al. 2016; Kane and Stuefer 2015). The wet bias in winter and spring for nearly all stations likely reflects not only overestimation by the model but also low biases in the measurements from gauge undercatch of solid precipitation. The fact that the 20CRv3-WRF downscaling overestimates precipitation compared to the GSOD station observations and PRISM, but with on average lower precipitation than ERA5, implies that the downscaling improves the estimation of precipitation due to the higher spatial resolution, but does not fully correct model biases. The physics parameterization could also play a role in the overestimation of precipitation and

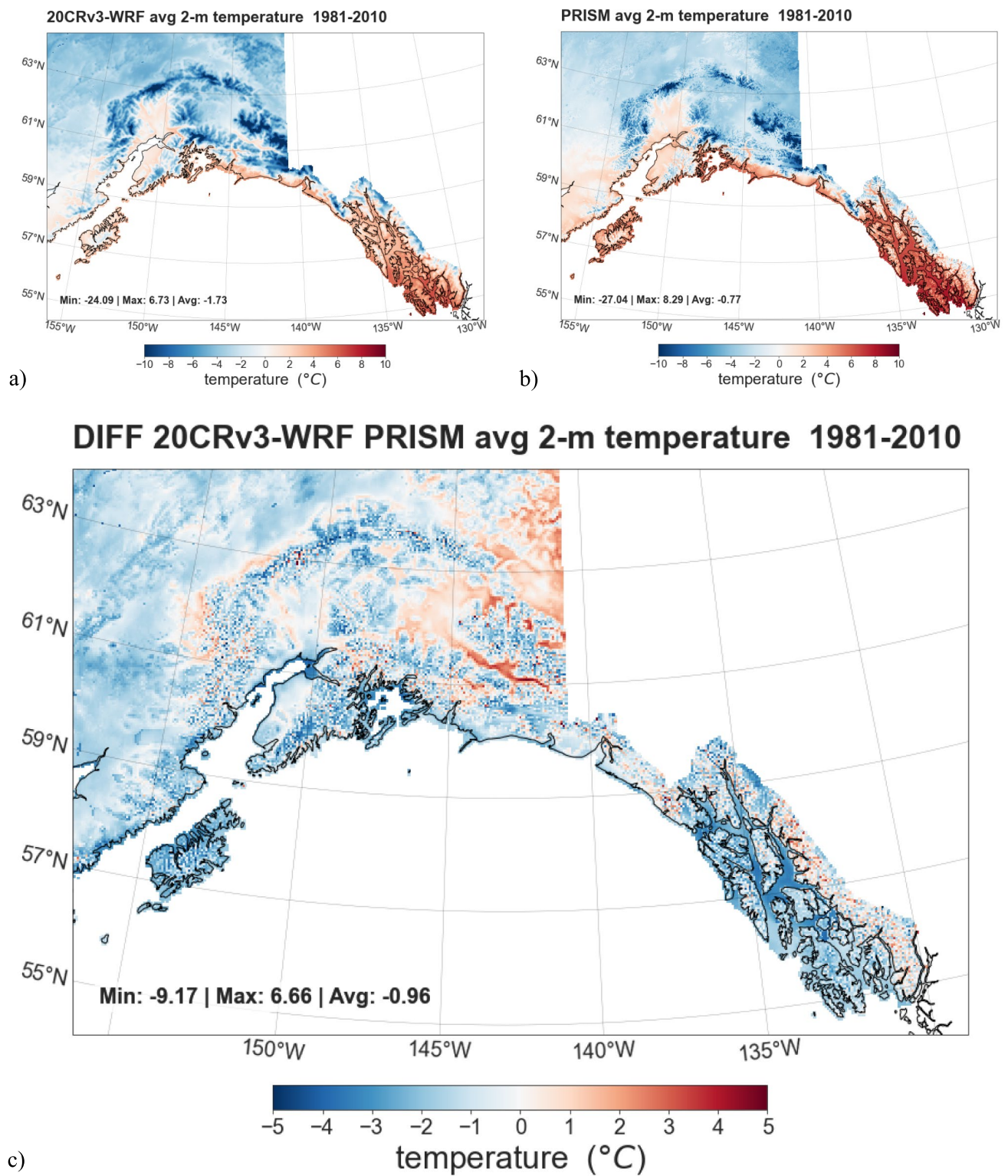


Fig. 7 30-year (1981–2010) mean 2-m temperature from (a) 20CRv3-WRF downscaling, (b) PRISM, and (c) the difference (calculated as 20CRv3-WRF minus PRISM). Minimum, maximum and mean difference values, in °C, are shown inset

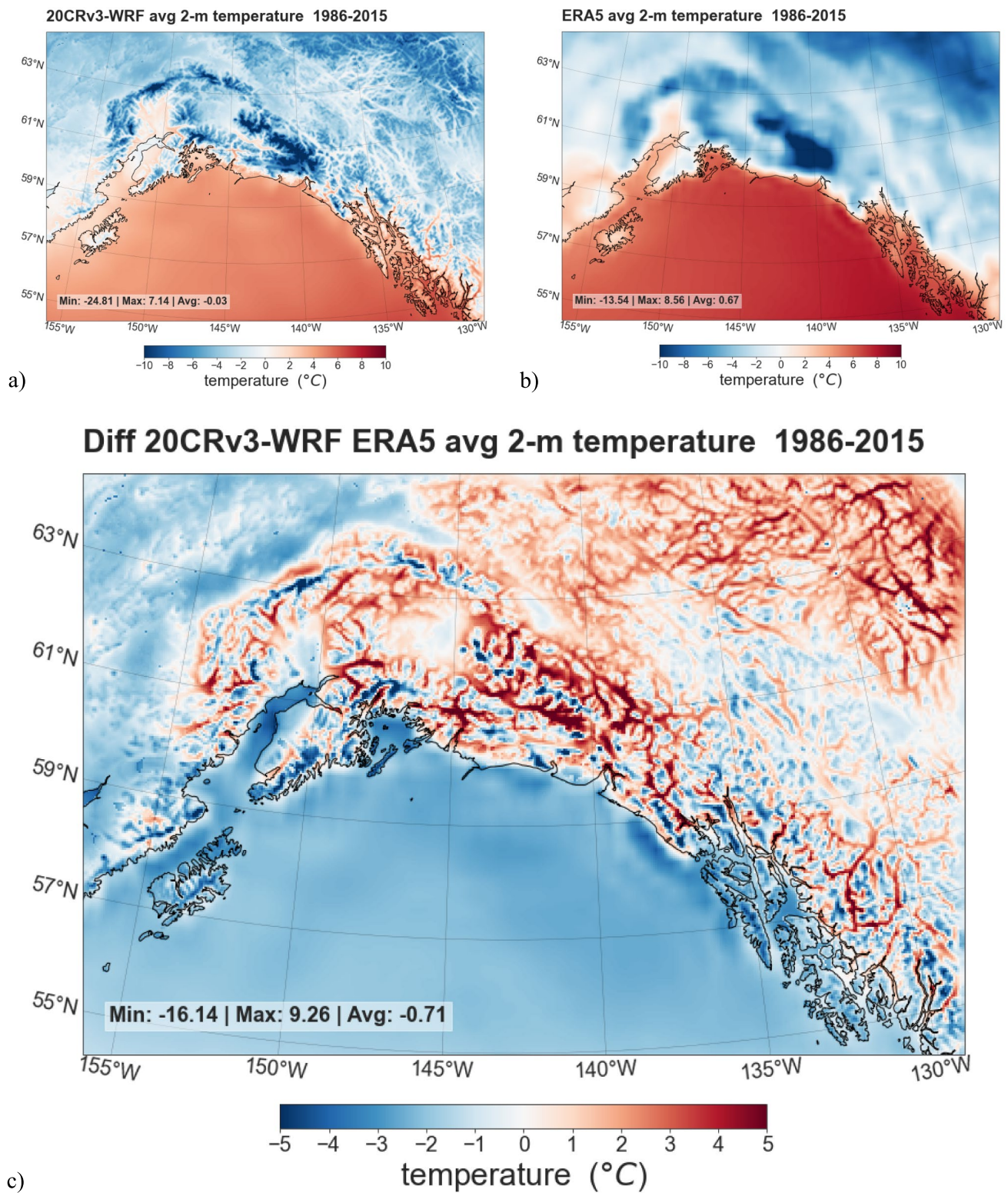
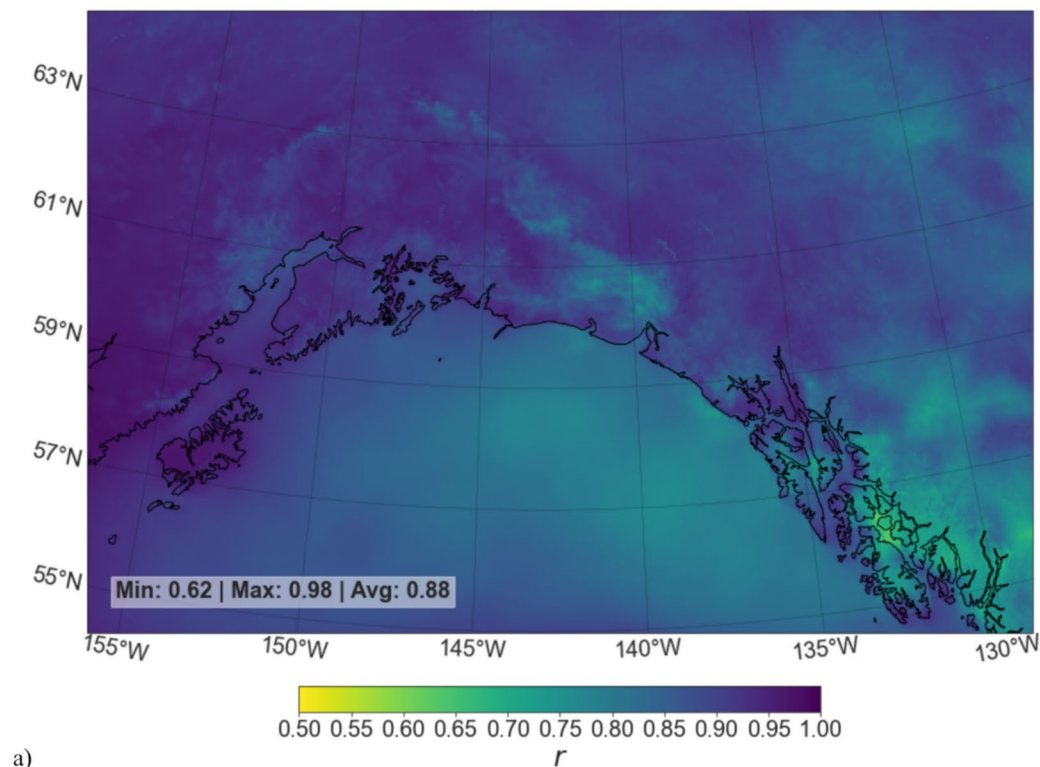
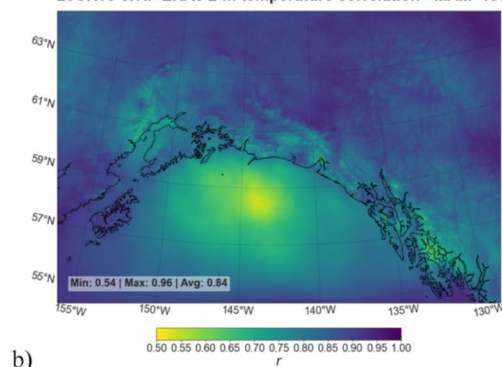


Fig. 8 30-year mean 2-m temperature (1986–2015) from (a) 20CRv3-WRF downscaling; (b) ERA5, (c) difference (calculated as 20CRv3-WRF—ERA5). Minimum, maximum, and average difference, in °C, are provided inset

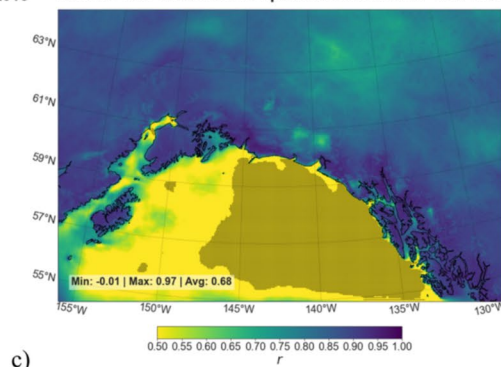
20CRv3-WRF ERA5 2-m temperature correlation 1986-2015



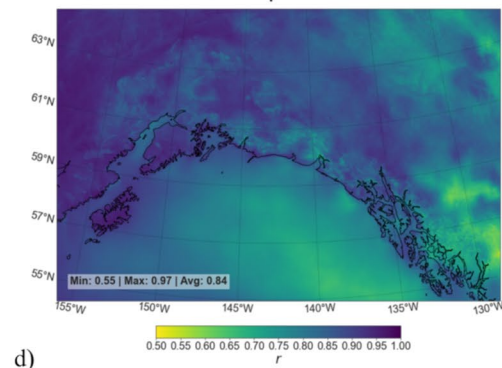
20CRv3-WRF ERA5 2-m temperature correlation - MAM 1986-2015



20CRv3-WRF ERA5 2-m temperature correlation - JJA 1986-2015



20CRv3-WRF ERA5 2-m temperature correlation - SON 1986-2015



20CRv3-WRF ERA5 2-m temperature correlation - DJF 1986-2015

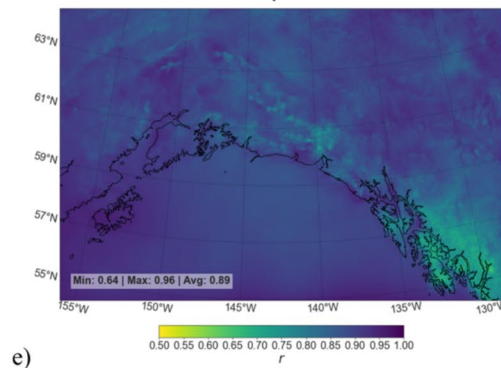


Fig. 9 Pearson correlation (r) between of 20CRv3-WRF downscaling and ERA5 for (a) annual mean (whole domain $= p < 0.05$) and seasonal (b) spring—MAM, (c) summer—JJA, (d) autumn—SON, (e) winter—DJF) 2-m temperature, calculated over the period 1986–2015 for each WRF grid cell. Shaded areas indicate $p > 0.05$, showing no statistical significance. Minimum, maximum, and average values of r are provided in inset. Further statistical analysis for station locations is provided in Table S3

further work is required to explore the influence of different physics schemes that influence precipitation direct or indirect.

Our results are consistent with findings from previous downscaling over Alaska. We found here that the 20CRv3-WRF downscaling tends to overestimate precipitation inland. This is similar to the finding of Bieniek et al. (2016) that the forcing data potentially introduces a positive bias. Bieniek et al. (2016) compared a downscaling model output with GSOD observations and a gridded dataset from statistical downscaling by Hill et al. (2015) and assessed similarity to the forcing data (ERA-Interim). Studies such as Bieniek et al. (2016) and Monaghan et al. (2018) have shown that downscaling reanalysis products improves precipitation representation compared to the original reanalysis products, consistent with our results. As in our analysis, Bieniek et al. (2016) also found that the higher elevation Alaskan Ranges area has higher precipitation in the downscaling output than in the reanalysis (ERA-Interim). Another similarity is higher precipitation in the model output than for the station observations at Juneau and Fairbanks. For temperature at Fairbanks, they reported a slight warm bias in winter and better model performance in summer, as seen in our results. Lader et al. (2020) presented a negative precipitation bias for Yakutat. Their downscaling dataset over southern Alaska (1981–2010, using the Climate Forecast System reanalysis as input data) performed most poorly for Yakutat, with RMSE above 14 and a negative bias in precipitation. This is consistent with our precipitation results for Yakutat that showed RMSE of 14.62 mm the simulated precipitation of the 20CRv3-WRF was lower than the observed. Lader et al. (2020) attributed this bias to the 4 km resolution of their downscaling product, which is too coarse to capture the topographic changes within the grid cell for the location of Yakutat. The station location is close to the ocean, and the model underestimates the upward motion in the flow pattern of the vertical motion. Because the 20CRv3-WRF has the same resolution and uses the same land surface model (Noah-MP) as Lader et al. (2020), we suspect the discrepancy between observed and simulated precipitation is similarly due to the model's handling of vertical motion at the ocean-land interface. For temperature, the downscaling model of Lader et al. (2020) performed better at Juneau than Yakutat, as seen here for the

20CRv3-WRF downscaling. The better performance of simulating temperature than precipitation seen by Lader et al. (2020) is also reflected in our results.

Monaghan et al. (2018) used a gridded dataset produced by statistical downscaling (SNAP dataset), along with satellite data and observational records, to validate snow cover, temperature, and precipitation for Alaska. Their WRF downscaling of Alaska performed more realistically for temperature than for precipitation, which was also seen in our results. In comparison with SNAP, their modelled annual temperature was colder and precipitation higher in the interior, with mixed biases (both negative and positive) on the coast. This is consistent with our comparison between the 20CRv3 downscaling and the PRISM dataset, in which the downscaling model also showed a negative bias for temperature and a positive bias in the interior but mixed biases along the coast. In general, our results are consistent with previous downscaling studies for Alaska, and we find similar magnitudes and patterns of bias for both temperature and precipitation.

5 Conclusions

This study evaluated the performance of the last four decades of a new 179-year 20CRv3-WRF downscaling over Alaska, with focus on its ability to predict 2-m temperature and precipitation. The model output was compared to observational records (GSOD) for 1986–2015, interpolated observational data (PRISM) for 1981–2010 and reanalysis (ERA5) for 1986–2015.

For temperature, the 20CRv3 downscaling was highly correlated with the gridded datasets (ERA5, PRISM) and observations ($0.61 \leq r \leq 0.96$). Differences between 20CRv3-WRF and ERA5 were larger than the differences between 20CRv3-WRF and PRISM, which suggests that the higher resolution contributes to a more realistic simulation of temperature. Higher accuracy was identified in the western part of the domain. However, ERA5 has weaknesses in reproducing soil temperatures for the eastern part of the domain, which could mean that 2-m air temperature is also biased, explaining the lower correlation between the 20CRv3-WRF and ERA 2-m temperatures. The evaluation with observational records indicates realistic performance of the downscaling. Seasonal biases were comparable to previous downscaling studies for Alaska (Bieniek et al. 2016; Monaghan et al. 2018).

Evaluating precipitation continues to be challenging. The performance is worse than for temperature (correlation with GSOD of $0.16 \leq r \leq 0.76$). Precipitation in the downscaling shows more spatial variation than seen in ERA5, with a lower correlation than for temperature. The 20CRv3-WRF overestimates precipitation, especially

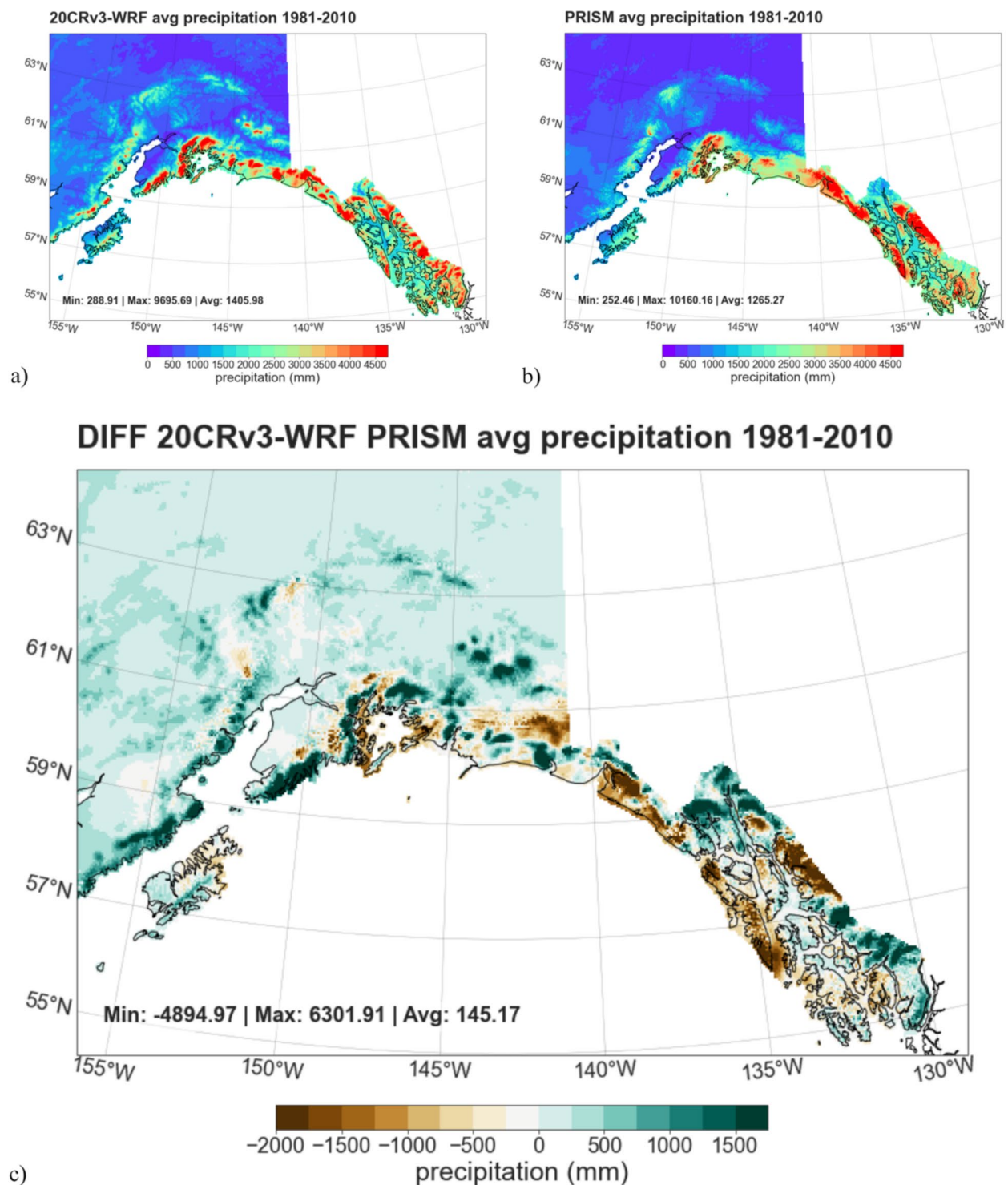


Fig. 10 30-year (1981–2010) mean precipitation from (a) 20CRv3-WRF downscaling, (b) PRISM, and (c) the difference (calculated as 20CRv3-WRF minus PRISM). Minimum, maximum and mean difference values, in mm, are shown inset

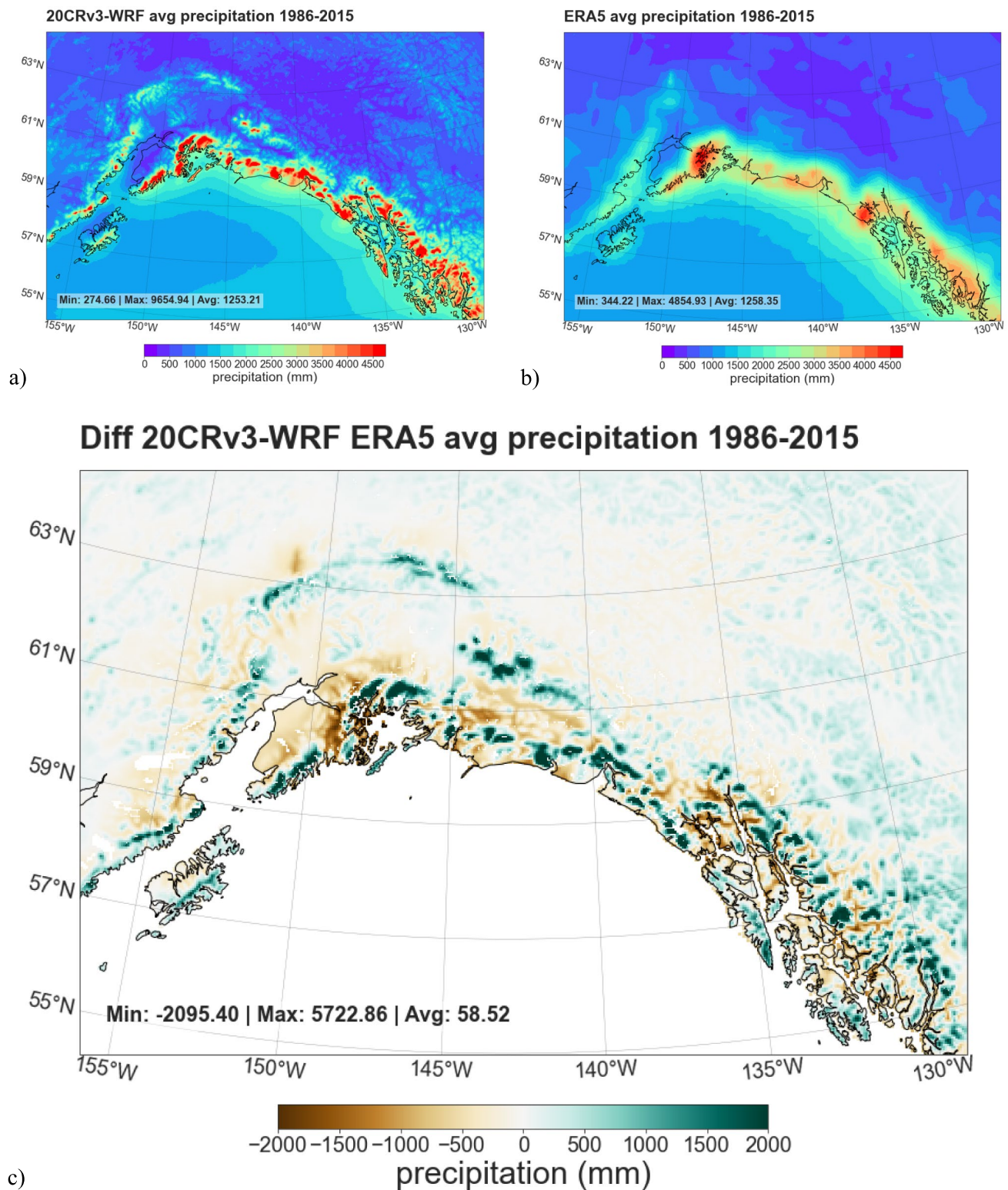
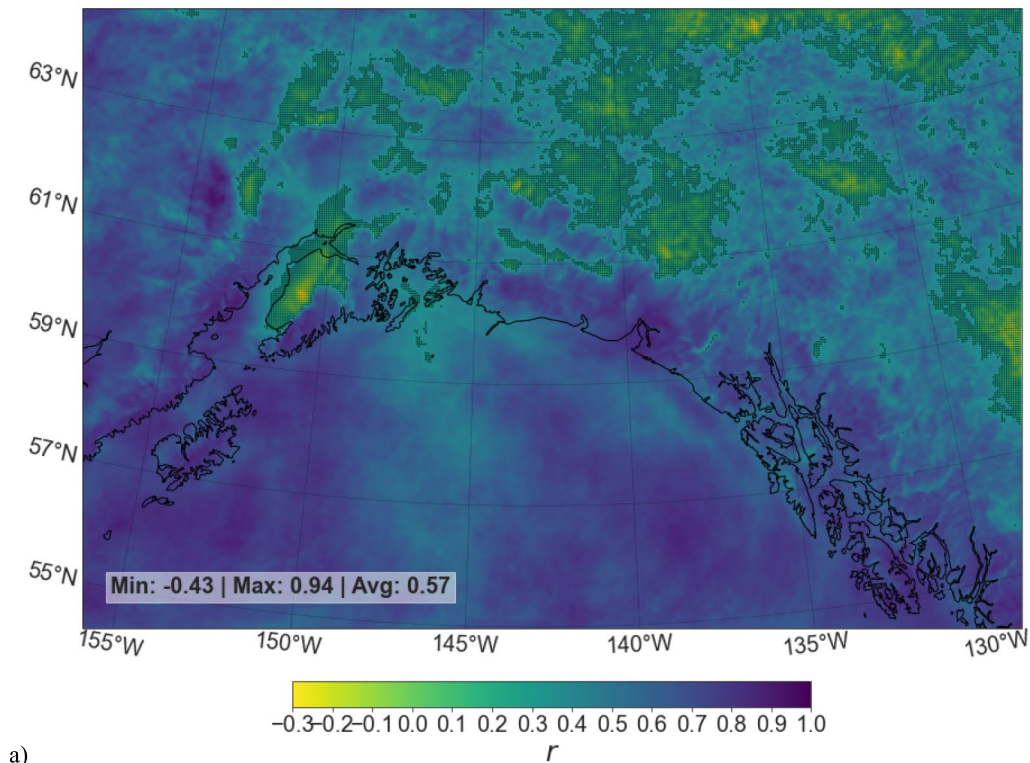


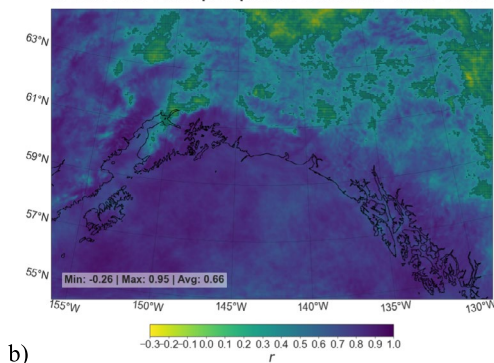
Fig. 11 30-year (1981–2010) mean annual accumulated precipitation from (a) 20CRv3-WRF downscaling, (b) ERA5, and (c) the difference (calculated as 20CRv3-WRF – ERA5). Minimum, maximum and mean difference values, in mm, are shown inset

20CRv3-WRF ERA5 precipitation correlation 1986-2015



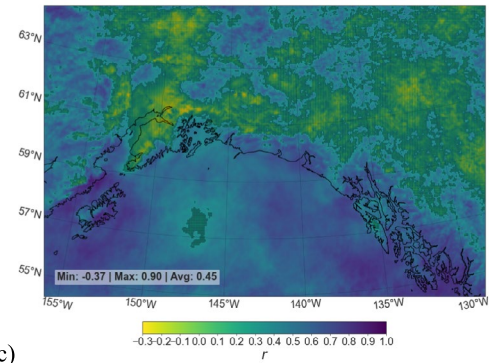
a)

20CRv3-WRF ERA5 precipitation correlation - MAM 1986-2015



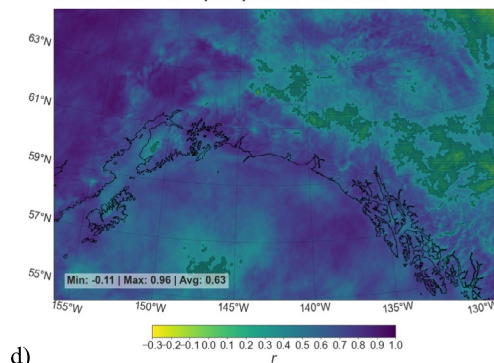
b)

20CRv3-WRF ERA5 precipitation correlation - JJA 1986-2015



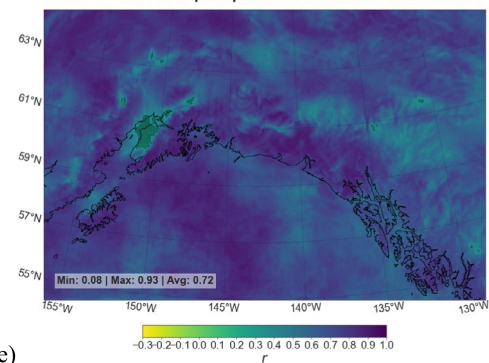
c)

20CRv3-WRF ERA5 precipitation correlation - SON 1986-2015



d)

20CRv3-WRF ERA5 precipitation correlation - DJF 1986-2015



e)

Fig. 12 Pearson correlation (r) between of 20CRv3-WRF downscaling and ERA5 for (a) annual mean and seasonal ((b) spring—MAM, (c) summer—JJA, (d) autumn—SON, (e) winter—DJF) accumulated precipitation, calculated over the period 1986–2015 for each WRF grid cell (shaded area means $p > 0.05$ showing no statistical significance)—minimum, maximum, and average values of r are provided inset. Further statistical analysis for station locations is provided in Table S4

in comparison to observations. However, measurements of solid precipitation in Alaska are known to be biased because of gauge undercatch (Bieniek et al. 2016; Kane and Stuefer 2015). The precipitation results found here are consistent with previous downscaling for Alaska (Bieniek et al. 2016; Lader et al. 2020; Monaghan et al. 2018), and with the overestimation of precipitation in WRF simulations for other parts of the world (Li et al. 2020; Maussion et al. 2011).

The results presented here provide a necessary precursor for further use of this dataset. We have shown that the 20CRv3 can successfully be used as forcing dataset for WRF downscaling, resulting in reasonable performance for the modern era (1981–2015) when observation-based evaluation datasets are available. However, given the 20CRv3 uncertainty in the nineteenth century and the lack of evaluation datasets before 1981, the full 179-year 20CRv3-WRF downscaling must be used with care. Although bias correction would improve the accuracy of the dataset, this is challenging because of the limited observational records and uncertainty in existing datasets, especially for precipitation. The 20CRv3-WRF does not perform equally well for temperature and precipitation. Despite the challenges, the accuracy of the modern-era temperature and precipitation outputs, as well as the length and high resolution of the dataset, provide the unique possibility to use these data to investigate long timescale climate phenomena. Future work will apply this downscaling product to the tidewater glacier cycle in southern Alaska.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00704-025-05612-x>.

Acknowledgements This research was supported by the Australian Research Council (grant FT170100220). The work was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme and the NCI Adapter Scheme supported by the Australian Government, as well as the University of Wollongong Partner Share Scheme (projects tl59, wj07, w28, n81 and q90). Data and software storage was utilized from projects rt52 (ERA5), sx70 (WRF), and hh5 (Python) at NCI. Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office, and by the NOAA

Physical Sciences Laboratory (PSL). The ERA5 reanalysis (ECMWF) was accessed from the storage at NCI (European Centre for Medium-Range Weather Forecasts, 2020). Observations were downloaded from the website of the Global Surface Summary of the Day (GSOD) prepared by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). The 20CRv3-WRF simulation was also evaluated with the Parameter-elevation Regression on Independent Slopes Model (PRISM) data developed by the PRISM Climate Group, Oregon State University (<https://prism.oregonstate.edu>). Technical assistance with HPC computing and Python was gratefully received from the NCI help team, CLEX CMS, and Sophie Grunau. Appreciation is also extended to Dr. Christopher Thomas for his help with the WRF variable post-processing.

Author contribution The ARC project was conceived by T.B.. All authors contributed to the study design. Model input data preparation and transfer was performed by C.M. and S.K., model design and simulation were conducted by S.K. advised by J.E., and analysis was performed by S.K. with the support of J.F., J.E. and Y.K.. The first draft of the manuscript was written by S.K. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This research was supported by an Australian Research Council Future Fellowship awarded to TB (FT170100220). This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), the University of Wollongong partner share, the Australian National University through the National Computational Merit Allocation Scheme and the NCI Adapter Scheme, with computational resources provided by NCI Australia, an NCRIS-enabled capability, which are supported by the Australian Government.

Data availability The dataset generated and analysed during the current study are available in the NSF NCAR Research Data Archive (RDA) (<https://doi.org/10.5065/VP7Z-YG52>) repository: <https://rda.ucar.edu/datasets/d010080/>

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arendt A, Luthcke S, Gardner A, O'Neel S, Hill D, Moholdt G, Abdalati W (2013) Analysis of a GRACE global mascon solution for Gulf of Alaska glaciers. *J Glaciol* 59:913–924

- Ballinger TJ, Bhatt US, Bieniek PA, Brettschneider B, Lader RT, Littell JS, Thoman RL, Waigl CF, Walsh JE, Webster MA (2023) Alaska terrestrial and marine climate trends, 1957–2021. *J Clim* 36:4375–4391
- Barros AP, Lettenmaier DP (1993) Dynamic modeling of the spatial distribution of precipitation in remote mountainous areas. *Mon Weather Rev* 121:1195–1214
- Bieniek PA, Bhatt US, Thoman RL, Angeloff H, Partain J, Papineau J, Fritsch F, Holloway E, Walsh JE, Daly C (2012) Climate divisions for Alaska based on objective methods. *J Appl Meteorol Climatol* 51:1276–1289
- Bieniek PA, Bhatt US, Walsh JE, Rupp TS, Zhang J, Krieger JR, Lader R (2016) Dynamical downscaling of ERA-Interim temperature and precipitation for Alaska. *J Appl Meteorol Climatol* 55:635–654
- Black T, Kurtz D (2023) Maritime glacier retreat and terminus area change in Kenai Fjords national park, Alaska, between 1984 and 2021. *J Glaciol* 69:251–265
- Bruyère CL, Done JM, Holland GJ, Fredrick S (2013) Bias corrections of global models for regional climate simulations of high-impact weather. *Clim Dyn* 43:1847–1856
- Cai L, Alexeev VA, Arp CD, Jones BM, Liljedahl AK, Gädeke A (2018) The polar WRF downscaled historical and projected twenty-first century climate for the coast and foothills of arctic Alaska. *Front Earth Sci* 5
- Cassano JJ, Box JE, Bromwich DH, Li L, Steffen K (2001) Evaluation of Polar MM5 simulations of Greenland's atmospheric circulation. *J Geophys Res Atmos* 106:33867–33889
- Cassano JJ, Higgins ME, Seefeldt MW (2011) Performance of the weather research and forecasting model for month-long pan-arctic simulations. *Mon Weather Rev* 139:3469–3488
- Cassano JJ, Cassano EN, Seefeldt MW, Gutowski WJ Jr, Glisan JM (2016) Synoptic conditions during wintertime temperature extremes in Alaska. *J Geophys Res: Atmospheres* 121(7):3241–3262
- Choudhury D, Ji F, Nishant N, Di Virgilio G (2023) Evaluation of ERA5-simulated temperature and its extremes for Australia. *Atmosphere* 14(6):913
- Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The twentieth century reanalysis project. *Q J R Meteorol Soc* 137:1–28
- Daly C, Gibson W, Taylor G, Johnson G, Pasteris P (2002) A knowledge-based approach to the statistical mapping of climate. *Clim Res* 22:99–113
- Daly C, Smith J, Halbleib M (2018) 1981–2010 High-resolution temperature and precipitation maps for Alaska. Oregon State University, PRISM Climate Group
- De La France A, McAfee SA (2019) Evaluation of synoptic-scale patterns during extreme temperature and precipitation events in Alaska. *Int J Climatol* 39:3134–3146
- Enderlin EM, O'Neel S, Bartholomaeus TC, Joughin I (2018) Evolving environmental and geometric controls on Columbia glacier's continued retreat. *J Geophys Res Earth Surf* 123. <https://doi.org/10.1029/2017JF004541>
- Errico RM, Vukićević T, Raeder K (1993) Comparison of initial and lateral boundary condition sensitivity for a limited-area model. *Tellus A* 45:539–557
- European Centre for Medium-Range Weather Forecasts (2020) ERA5 Replicated datasets. NCI Australia
- Evans JP, Ekström M, Ji F (2012) Evaluating the performance of a WRF physics ensemble over South-East Australia. *Clim Dyn* 39:1241–1258
- Goodison BE, Louie PY, Yang D (1998) WMO solid precipitation measurement intercomparison
- Gulev SK, Thorne PW, Ahn J, Dentener FJ, Domingues CM, Gerland S, Gong D, Kaufman DS, Nnamchi HC, Quaas J, Rivera JA, Sathyendranath S, Smith SL, Trewin B, von Schuckmann K, Vose RS (2021) Changing state of the climate system. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (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*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp 287–422. <https://doi.org/10.1017/9781009157896.004>
- Hawkins E, Brohan P, Burgess SN, Burt S, Compo GP, Gray SL, Haigh ID, Hersbach H, Kuijjer K, Martínez-Alvarado O, McColl C, Schurer AP, Slivinski L, Williams J (2023a) Rescuing historical weather observations improves quantification of severe windstorm risks. *Nat Hazards Earth Syst Sci* 23:1465–1482
- Hawkins E, Compo GP, Sardeshmukh PD (2023b) ESD Ideas: Translating historical extreme weather events into a warmer world. *Earth Syst Dynam* 14:1081–1084
- He Y, Chen C, Li B, Zhang Z (2022) Prediction of near-surface air temperature in glacier regions using ERA5 data and the random forest regression method. *Remote Sens Appl: Soc Environ* 28:100824
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049
- Hill DF, Bruhis N, Calos SE, Arendt A, Beamer J (2015) Spatial and temporal variability of freshwater discharge into the Gulf of Alaska. *J Geophys Res Oceans* 120:634–646
- Jeworrek J, West G, Stull R (2021) WRF precipitation performance and predictability for systematically varied parameterizations over complex terrain. *Weather Forecast* 36:893–913
- Kane DL, Stuefer SL (2015) Reflecting on the status of precipitation data collection in Alaska: A case study. *Hydrol Res* 46:478–493
- Karmalkar AV, Horton RM (2021) Drivers of exceptional coastal warming in the northeastern United States. *Nat Clim Chang* 11:854–860
- Kobayashi S, Ota Y, Harada Y, Ebata A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *J Meteorol Soc Jpn Ser II* 93:5–48
- Lader R, Walsh JE, Bhatt US, Bieniek PA (2017) Projections of twenty-first-century climate extremes for Alaska via dynamical downscaling and quantile mapping. *J Appl Meteorol Climatol* 56:2393–2409
- Lader R, Bidlack A, Walsh JE, Bhatt US, Bieniek PA (2020) Dynamical downscaling for Southeast Alaska: Historical climate and future projections. *J Appl Meteorol Climatol* 59:1607–1623
- Lavers DA, Simmons A, Vamborg F, Rodwell MJ (2022) An evaluation of ERA5 precipitation for climate monitoring. *Q J R Meteorol Soc* 148:3152–3165
- Li D, Yang K, Tang W, Li X, Zhou X, Guo D (2020) Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas. *Int J Climatol* 40:5114–5127

- Maussion F, Scherer D, Finkelnburg R, Richters J, Yang W, Yao T (2011) WRF simulation of a precipitation event over the Tibetan Plateau, China – an assessment using remote sensing and ground observations. *Hydrol Earth Syst Sci* 15:1795–1817
- McAfee SA, Walsh J, Rupp TS (2014) Statistically downscaled projections of snow/rain partitioning for Alaska. *Hydrol Process* 28:3930–3946
- Meier M, Post A (1987) Fast tidewater glaciers. *J Geophys Res Solid Earth* 92:9051–9058
- Mölders N, Kramm G (2010) A case study on wintertime inversions in interior Alaska with WRF. *Atmos Res* 95:314–332
- Molnia BF (2007) Late nineteenth to early twenty-first century behavior of Alaskan glaciers as indicators of changing regional climate. *Glob Planet Change* 56:23–56
- Monaghan AJ, Clark MP, Barlage MP, Newman AJ, Xue L, Arnold JR, Rasmussen RM (2018) High-resolution historical climate simulations over Alaska. *J Appl Meteorol Climatol* 57:709–731
- Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut JN (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst Sci Data* 13:4349–4383
- NOAA (National Oceanic and Atmospheric Administration) (1999) Global surface summary of the day - GSOD. 1.0, viewed 28/01/2021. <https://www.ncei.noaa.gov/access/metadata/landingpage/bin/iso?id=gov.noaa.ncdc:C00516>
- Pfeffer WT, Arendt AA, Bliss A, Bolch T, Cogley JG, Gardner AS, Hagen JO, Hock R, Kaser G, Kienholz C, Miles ES, Moholdt G, Mölg N, Paul F, Radić V, Rastner P, Raup BH, Rich J, Sharp MJ, Andreassen LM, Bajracharya S, Barrand NE, Beedle MJ, Berthier E, Bhambri R, Brown I, Burgess DO, Burgess EW, Cawkwell F, Chinn T, Copland L, Cullen NJ, Davies B, De Angelis H, Fountain AG, Frey H, Giffen BA, Glasser NF, Gurney SD, Hagg W, Hall DK, Haritashya UK, Hartmann G, Herreid S, Howat I, Jiskoot H, Khromova TE, Klein A, Kohler J, König M, Kriegel D, Kutuzov S, Lavrentiev I, Le Bris R, Li X, Manley WF, Mayer C, Menounos B, Mercer A, Mool P, Negrete A, Nosenko G, Nuth C, Osmonov A, Pettersson R, Racoviteanu A, Ranzi R, Sarikaya MA, Schneider C, Sigurdsson O, Sirguey P, Stokes CR, Wheate R, Wolken GJ, Wu LZ, Wyatt FR (2014) The Randolph glacier inventory: a globally complete inventory of glaciers. *J Glaciol* 60:537–552
- Post A, O'Neil S, Motyka RJ, Streveler G (2011) A complex relationship between calving glaciers and climate. *Eos Trans Am Geophys Union* 92:305–306
- Poujol B, Prein AF, Newman AJ (2020a) Kilometer-scale modeling projects a tripling of Alaskan convective storms in future climate. *Clim Dyn* 55:3543–3564
- Powers JG, Klemp JB, Skamarock WC, Davis CA, Dudhia J, Gill DO, Coen JL, Gochis DJ, Ahmadov R, Peckham SE, Grell GA, Michalak J, Trahan S, Benjamin SG, Alexander CR, Dimego GJ, Wang W, Schwartz CS, Romine GS, Liu ZQ, Snyder C, Chen F, Barlage MJ, Yu W, Duda MG (2017) The weather research and forecasting model: overview, system efforts, and future directions. *Bull Am Meteorol Soc* 98:1717–1737
- Prein AF, Gobiet A, Suklitsch M, Truhetz H, Awan NK, Keuler K, Georgievski G (2013b) Added value of convection permitting seasonal simulations. *Clim Dyn* 41:2655–2677
- PRISM (2023a) Alaska - PRIMS data 30-year mean 2-m air temperature and 30-year mean accumulated precipitation. In: Group PC (ed) PRISM (2023b) PRISM Climate data
- Rantanen M, Karpechko AY, Lipponen A, Nordling K, Hyvärinen O, Ruosteenoja K, Vihma T, Laaksonen A (2022) The Arctic has warmed nearly four times faster than the globe since 1979. *Commun Earth Environ* 3:168
- Rasmussen LA, Conway H, Krimmel RM, Hock R (2011a) Surface mass balance, thinning and iceberg production, Columbia glacier, Alaska, 1948–2007. *J Glaciol* 57:431–440
- Rasmussen R, Liu C, Ikeda K, Gochis D, Yates D, Chen F, Tewari M, Barlage M, Dudhia J, Yu W, Miller K, Arseneault K, Grubišić V, Thompson G, Gutmann E (2011b) High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: a process study of current and warmer climate. *J Clim* 24:3015–3048
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, Wang W, Powers JG, Duda MG, Barker DM, Huang XY (2019) A description of the advanced research WRF model version 4. Ncar Technical Notes, No. NCAR/TN-556+ STR. <https://doi.org/10.5065/1dfh-6p97>
- Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, Allan R, Yin X, Vose R, Titchner H, Kennedy J, Spencer LJ, Ashcroft L, Brönnimann S, Brunet M, Camuffo D, Cornes R, Cram TA, Crouthamel R, Domínguez-Castro F, Freeman JE, Gergis J, Hawkins E, Jones PD, Jourdain S, Kaplan A, Kubota H, Blancq FL, Lee T-C, Lorrey A, Luterbacher J, Maugeri M, Mock CJ, Moore GWK, Przybylak R, Pudmenzky C, Reason C, Slonosky VC, Smith CA, Tinz B, Trewin B, Valente MA, Wang XL, Wilkinson C, Wood K, Wyszyński P (2019) Towards a more reliable historical reanalysis: improvements for version 3 of the twentieth century reanalysis system. *Q J R Meteorol Soc* 145:2876–2908
- Slivinski LC, Compo GP, Sardeshmukh PD, Whitaker JS, McColl C, Allan RJ, Brohan P, Yin X, Smith CA, Spencer LJ, Vose RS, Rohrer M, Conroy RP, Schuster DC, Kennedy JJ, Ashcroft L, Brönnimann S, Brunet M, Camuffo D, Cornes R, Cram TA, Domínguez-Castro F, Freeman JE, Gergis J, Hawkins E, Jones PD, Kubota H, Lee TC, Lorrey AM, Luterbacher J, Mock CJ, Przybylak RK, Pudmenzky C, Slonosky VC, Tinz B, Trewin B, Wang XL, Wilkinson C, Wood K, Wyszyński P (2021) An evaluation of the performance of the twentieth century reanalysis version 3. *J Clim* 34:1417–1438
- Song Y, Broxton PD, Ehsani MR, Behrangi A (2021) Assessment of snowfall accumulation from satellite and reanalysis products using SNOTEL observations in Alaska. *Remote Sens* 13(15):2922
- Tangborn W (1997) Using low-altitude meteorological observations to calculate the mass balance of Alaska's Columbia glacier and relate it to calving and speed. *Byrd Polar Clim Res Cen* 15:141–161
- Trenberth KE, Koike T, Onogi K (2008) Progress and prospects for reanalysis for weather and climate. *Eos* 89:234–235
- Walsh JE, Bhatt US, Littell JS, Leonawicz M, Lindgren M, Kurkowski TA, Bieniek PA, Thoman R, Gray S, Rupp TS (2018) Downscaling of climate model output for Alaskan stakeholders. *Environ Model Softw* 110:38–51
- Wang Y, Ding M, Reijmer CH, Smeets PCJP, Hou S, Xiao C (2021) The AntSMB dataset: a comprehensive compilation of surface mass balance field observations over the Antarctic Ice Sheet. *Earth Syst Sci Data* 13:3057–3074. <https://doi.org/10.5194/essd-13-3057-2021>
- Wang Y, Zhang X, Ning W, Lazzara MA, Ding M, Reijmer CH, Smeets PCJP, Grigioni P, Heil P, Thomas ER, Mikolajczyk D, Welhouse LJ, Keller LM, Zhai Z, Sun Y, Hou S (2023) The AntAWS dataset: a compilation of Antarctic automatic weather station

- observations. *Earth Syst Sci Data* 15:411–429. <https://doi.org/10.5194/essd-15-411-2023>
- Yde JC, Paasche Ø (2010) Reconstructing climate change: not all glaciers suitable. *Eos* 91:189–190
- Yu Y, Xiao W, Zhang Z, Cheng X, Hui F, Zhao J (2021) Evaluation of 2-m air temperature and surface temperature from ERA5 and ERA-I using buoy observations in the arctic during 2010–2020. *Remote Sens*
- Yu B, Wang XL, Feng Y, Chan R, Compo GP, Slivinski LC, Sardeshmukh PD, Wehner M, Yang XY (2022) Northern hemisphere extratropical cyclone activity in the twentieth century reanalysis version 3 (20CRv3) and its relationship with continental extreme temperatures. *Atmosphere* 13
- Zemp M, Huss M, Thibert E, Eckert N, McNabb R, Huber J, Barandun M, Machguth H, Nussbaumer SU, Gärtner-Roer I, Thomson L, Paul F, Maussion F, Kutuzov S, Cogley JG (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* 568:382–386
- Zhang S-Q, Ren G-Y, Ren Y-Y, Zhang Y-X, Xue X-Y (2021) Comprehensive evaluation of surface air temperature reanalysis over China against urbanization-bias-adjusted observations. *Adv Clim Chang Res* 12:783–794
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.