Climate Extremes across the North American Arctic in Modern Reanalyses

ALVARO AVILA-DIAZ,[a] DAVID H. BROMWICH,[b] AARON B. WILSON,[b] FLAVIO JUSTINO,[a] AND
SHENG-HUNG WANG[b]

[a] *Department of Agricultural Engineering, Universidade Federal de Viçosa, Viçosa, Brazil*
[b] *Polar Meteorology Group, Byrd Polar and Climate Research Center, The Ohio State University, Columbus, Ohio*

ABSTRACT: Atmospheric reanalyses are a valuable climate-related resource where in situ data are sparse. However, few studies have investigated the skill of reanalyses to represent extreme climate indices over the North American Arctic, where changes have been rapid and indigenous responses to change are critical. This study investigates temperature and precipitation extremes as defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) over a 17-yr period (2000–16) for regional and global reanalyses, namely the Arctic System Reanalysis, version 2 (ASRv2); North American Regional Reanalysis (NARR); European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis; Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2); and Global Meteorological Forcing Dataset for Land Surface Modeling (GMFD). Results indicate that the best performances are demonstrated by ASRv2 and ERA5. Relative to observations, reanalyses show the weakest performance over far northern basins (e.g., the Arctic and Hudson basins) where observing networks are less dense. Observations and reanalyses show consistent warming with decreased frequency and intensity of cold extremes. Cold days, cold nights, frost days, and ice days have decreased dramatically over the last two decades. Warming can be linked to a simultaneous increase in daily precipitation intensity over several basins in the domain. Moreover, the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) distinctly influence extreme climate indices. Thus, these findings detail the complexity of how the climate of the Arctic is changing, not just in an average sense, but in extreme events that have significant impacts on people and places.

KEYWORDS: Arctic; North America; Extreme events; Reanalysis data

1. Introduction

Weather observations and climate data are sparse across the Arctic compared to the midlatitudes and tropics (Zhang et al. 2011; Lader et al. 2016; Diaconescu et al. 2017). To overcome this limitation, retrospective analyses (i.e., reanalyses) synthesize a wide variety of surface, atmospheric, and satellite data into model-based gridded datasets. A fundamental advantage of reanalyses is that they provide continuous data coverage across the globe or region of interest, allowing researchers to investigate climate change and variability over recent decades (Dee et al. 2014).

Reanalyses and gridded observations demonstrate that average temperatures are warming and precipitation is increasing, with rapid changes occurring to the physical Arctic environment (Boisvert and Stroeve 2015; Rapaić et al. 2015). An improved understanding of extreme climate events, such as the warm Alaskan winter (October–April) of 2015/16 when the mean temperature was more than 4°C above average (Walsh et al. 2017), is needed. Accurate representation of extremes across the Arctic is essential to the development of public policies, proper management of hydrological resources, and mitigation of impacts from human activity on the environment, as future projections over North America indicate a significant decrease in cold extremes and increase in warm extremes by the end of the twenty-first century (Schoof and Robeson 2016; Lader et al. 2017; Sheridan and Lee 2018; Wazneh et al. 2020).

Arctic-focused reanalysis studies have shown limited ability to reproduce daily observed extremes of temperature and precipitation (Lindsay et al. 2014; Donat et al. 2014; Koyama and Stroeve 2019; Przybylak and Wyszyński 2020; Zhang et al. 2011; Rapaić et al. 2015; Lader et al. 2016; Diaconescu et al. 2017). Lindsay et al. (2014) evaluated temperature and precipitation from seven reanalyses at a monthly scale across the Arctic and found significant biases compared to observations, especially in summer for temperature and winter for precipitation. Lader et al. (2016) used data from five atmospheric reanalyses for Alaska to evaluate the representation of mean and extremes of temperature and precipitation and found that all reanalyses overestimate temperature variability during summer, are too wet over the North Slope of Alaska, and tend to underestimate winter rainfall in southeastern Alaska. Diaconescu et al. (2017) reported that reanalyses can be used with confidence to accurately represent observed hot extremes (e.g., warm days) and precipitation-based frequency indices (e.g., number of wet days) but struggle to reproduce cold extremes (e.g., cold days/night) or precipitation intensity extremes (e.g., very wet days) over northern Canada.

The aim of this study is to assess the performance of a selection of modern global/regional reanalyses in estimating observed climate extremes and trends over the North American Arctic for a recent 17-yr period (2000–16). This period was chosen as it matches the current Arctic System Reanalysis, version 2 (ASRv2; Bromwich et al. 2018), coverage period and represents an era of rapid, amplified, and well-observed Arctic change. This relatively short analysis period limits the ability to fully explore the causal mechanisms for

Corresponding author: Aaron Benjamin Wilson, wilson.1010@osu.edu

TABLE 1. Characteristics of reanalyses and gridded observation datasets. Letters of the second column designate gauge data (G), reanalysis (R), and combined sources (C; i.e., gauge, satellite, and reanalysis).

Product name and download link	Type	Period	Horizontal resolution	Spatial coverage
DAYMET (version 3); https://daac.ornl.gov/	G	1980–2018	1 km	North America
The Arctic System Reanalysis (ASRv2); https://rda.ucar.edu/datasets/ds631.1/	R	2000–16	15 km	~40°–90°N
North American Regional Reanalysis (NARR); https://rda.ucar.edu/datasets/ds608.0/	R	1979–present	31 km	North America
ECMWF ERA5 Reanalysis (ERA5) https://rda.ucar.edu/datasets/ds633.0/	R	1979–present	31 km	Global
The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2); https://disc.sci.gsfc.nasa.gov/	R	1980–present	~65 km ($0.5° \times 0.625°$)	Global
Global Meteorological Forcing Dataset for Land Surface Modeling (GMFD); https://rda.ucar.edu/datasets/ds314.0	C	1948–2016	~28 km ($0.25°$)	Global

these trends but allows an assessment of the relative performance of trends across multiple reanalysis products.

The North American Arctic was selected to take advantage of a readily available gridded observation dataset over this region and recent studies that have explored the linkages between Arctic changes and impacts in the midlatitudes (e.g., Cohen 2016; Overland and Wang 2018) as well as the ever-growing threat of wildfires and other climate-sensitive ecological impacts (e.g., Melvin et al. 2017; Young et al. 2019). This is done by utilizing the framework recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI). This analysis will help researchers evaluate the performance of the contemporary reanalyses, proving their value in the study of Arctic climate change and variability, including the occurrence of extreme temperature and precipitation events. Also, these results provide a breakthrough in understanding by decreasing the uncertainty about the patterns of extreme temperature and precipitation events over the Arctic where observational information is sparse.

This paper is structured as follows. Section 2 describes the observations, reanalyses, and statistical tests performed. Section 3 discusses the results for the reference period 2000–16. Section 4 presents a summary and concluding remarks.

2. Data and methods

a. Data

We applied a multiple dataset approach to increase the confidence in using reanalyses to estimate temperature and precipitation extremes following previous climate and hydro-meteorological studies (Lindsay et al. 2014; Yin et al. 2015; Diaconescu et al. 2017; Wong et al. 2017). Table 1 summarizes the characteristics of the observations and reanalyses. All gridded products provide daily temperature and precipitation at horizontal spatial resolutions ranging from 15 to 65 km across the North American Arctic (Fig. 1).

Daily observed meteorological data were acquired from DAYMET (version 3) supported by the National Aeronautics and Space Administration (NASA; Thornton et al. 1997, 2018). DAYMET provides a 1-km spatial resolution of weather parameters for North America including Canada, Mexico, and the United States for 1980–2016. DAYMET inputs are a digital elevation model and in situ weather observations of daily maximum temperature, minimum temperature, and precipitation from the Global Historical Climatology Network (GHCN-Daily; Thornton et al. 2018).

To our knowledge, there are no studies that have evaluated the performance of temperature and precipitation extremes in newer reanalyses. We include reanalysis data from two regional reanalyses: ASRv2 (NCAR 2017; Bromwich et al. 2018) and the North American Regional Reanalysis (NARR; Mesinger et al. 2006). Global products include the ERA5 produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; Hersbach et al. 2020); Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017); and the Global Meteorological Forcing Dataset for Land Surface Modeling (GMFD; Sheffield et al. 2006). Studies have shown that outputs from these reanalyses (e.g., ASRv2, MERRA-2, and ERA5) are more consistent with climate observations (Gelaro et al. 2017; Koyama and Stroeve 2019; Wang et al. 2019; Bromwich et al. 2018; Tarek et al. 2020) than are the earlier generation of reanalyses (e.g., ERA-40 and NCEP–NCAR) because of higher horizontal and vertical resolutions and improved model physics.

ERA5 and MERRA-2 are considered "full input" reanalyses in that they assimilate surface and upper-air conventional and satellite data (Fujiwara et al. 2017). In addition to the conventional surface and upper-air data, ASRv2 assimilates QuikSCAT, and Special Sensor Microwave Imager sea surface winds, satellite radiances, and GPS data (Bromwich et al. 2018). Like ASRv2, NARR does not assimilate satellite temperatures retrievals but includes satellite radiances.
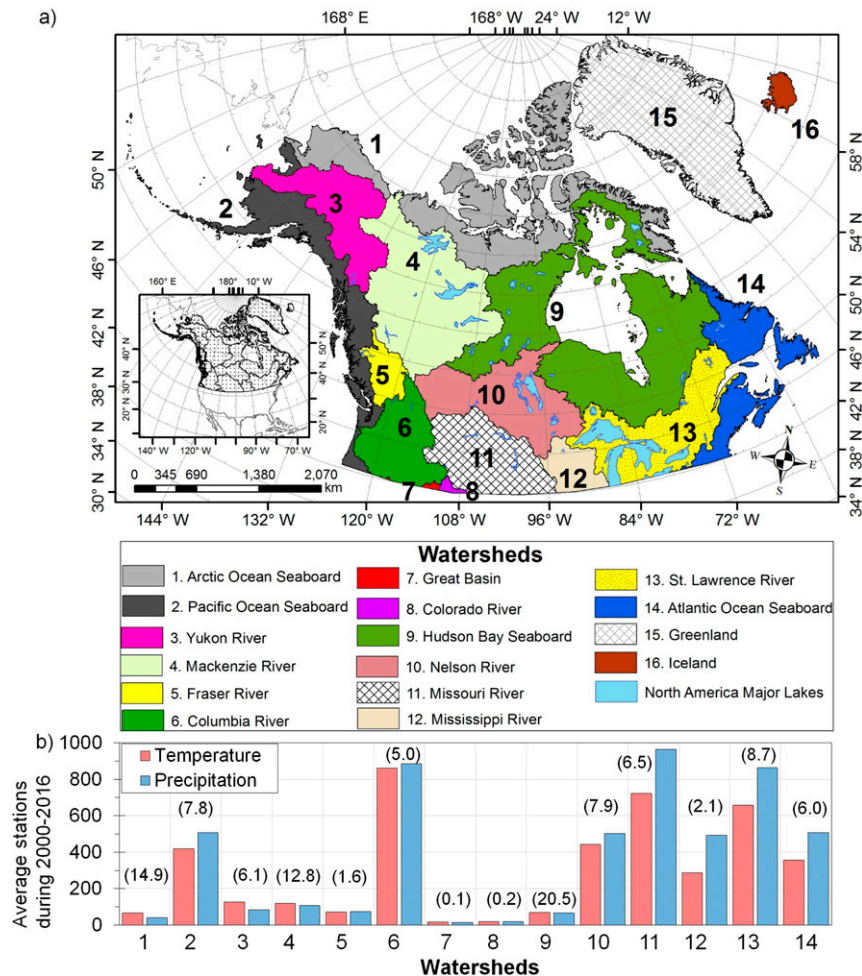
FIG. 1. (a) The domain of the North America region considered in this study. The colors depict the hydrological basins according to the U.S. Geological Survey (https://www.usgs.gov/). (b) The average number of weather and rainfall stations in each watershed. The number in parentheses indicates the percentage of the area considering the DAYMET domain.

ERA5 assimilates precipitation from ground-based radar from 2009 onward, MERRA-2 uses observation-based precipitation data to correct the precipitation over land surfaces outside of the high latitudes (Reichle et al. 2017), and NARR assimilates precipitation (mostly over lower latitudes with limited input across Canada) (Mesinger et al. 2006). GMFD is considered a reanalysis product, because the daily outputs of temperature and precipitation are constructed using the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis and global observations compiled by the Climate Research Unit (CRU), namely the monthly CRU TS3.24.01 gridded at 0.5° horizontal spatial resolution.

The period studied matches the current ASRv2 period (2000–16; 17 yr). All datasets are regridded to a common horizontal resolution of the 0.25° latitude/longitude covering the whole domain (Fig. 1a) using a bilinear remapping algorithm with the Climate Data Operators (CDO; https://code.mpimet.mpg.de/projects/cdo/). Note, however, that native resolution is important for the detection and accurate representation of extreme events, and we are not dismissing the benefits that higher resolution may provide. In fact, previous ASR studies have shown the impact that higher horizontal resolution has on localized processes such as wind events and surface fluxes across the Arctic (Moore et al. 2015; Bromwich et al. 2016, 2018; Justino et al. 2019). Roberts et al. (2018) thoroughly detail the spectrum of benefits related to higher-resolution models. However, some have shown that, despite a general improvement in capturing, for example, precipitation extremes with higher resolution, it does not always guarantee better skill and may require better model physics or tuning (e.g., Bador et al. 2020). A systematic analysis of incrementally finer resolution reanalyses is beyond the scope of this paper but should be considered for future assessments.

TABLE 2. Definition of temperature and precipitation indices selected for analysis of extreme climate indices recommended by ETCCDI. Note that TNn, TX10p, and TN10p are cold extreme indices and TXx, TX90p, and TN90p are hot extreme indices. A wet or dry day is defined when precipitation (PR) is ≥ 1 or < 1 mm. Also, note that 2TM (index 10) is not an ETCCDI index.

Index	Indicator name	Indicator definitions	Units
Extreme indices derived from daily max and min temperature series			
1. TXx	Hottest day	Annual max value of daily max temperature	°C
2. TX10p	Cold days	Percentage of days when daily max temperature < 10th percentile	%
3. TX90p	Warm days	Percentage of days when daily max temperature > 90th percentile	%
4. TNn	Coldest night	Annual min value of daily min temperature	°C
5. TN10p	Cold nights	Percentage of days when daily min temperature < 10th percentile	%
6. TN90p	Warm nights	Percentage of days when daily min temperature > 90th percentile	%
7. DTR	Diurnal temperature range	Annual mean difference between daily max and min temperature	°C
8. ID	Ice days	Annual count of days when daily max temperature < 0°C	Days
9. FD	Frost days	Annual count of days when daily min temperature < 0°C	Days
10. 2TM	Annual mean temperature	Annual daily mean temperature	°C
Extreme indices derived from daily precipitation series			
11. PRCPTOT	Annual total wet-day precipitation	Annual total precipitation (PR) in wet days (PR ≥ 1 mm)	mm
12. RX1day	Max 1-day precipitation amount	Annual max 1-day precipitation	mm
13. RX5day	Max 5-day precipitation amount	Annual max consecutive 5-day precipitation	mm
14. R95p	Very wet days	Annual total precipitation from days > 95th percentile	mm
15. SDII	Simple daily intensity index	The ratio of annual total precipitation to the no. of wet days (≥1 mm)	mm day^{-1}
16. R1mm	No. of wet days	Annual count of days when PR ≥ 1 mm	Days
17. CWD	Consecutive wet days	Max no. of consecutive days with PR ≥ 1 mm	Days
18. CDD	Consecutive dry days	Max no. of consecutive days with PR < 1 mm	Days

b. Extreme climate indices

We have analyzed the performance of the reanalyses using 17 extreme climate indices (defined in Table 2) proposed by ETCCDI (http://etccdi.pacificclimate.org/). Zhang et al. (2011) and Donat et al. (2013) provide additional details on each index. The indices are based on daily minimum temperature (TMIN), maximum temperature (TMAX), and precipitation (PRCP). We use nine temperature and eight precipitation indices relevant to the Arctic and evaluate based on annual scale (Lader et al. 2016; Diaconescu et al. 2017). These indices include absolute indices [e.g., hottest days (TXx), ice days (ID), and frost days (FD)], percentile-based indices [e.g., warm days (TX90p) and very wet days (R95p)], intensity indices [e.g., maximum 5-day precipitation amount (RX5day)], frequency indices [e.g., number of wet days (R1mm)], and indices based on the duration of the wet or dry events [e.g., consecutive wet

days (CWD) and consecutive dry days (CDD)]. The package "climdex.pcic.ncdf" (version 0.5-4) was used to calculate extreme climate indices, which is freely available to run on R software (https://github.com/pacificclimate/climdex.pcic.ncdf).

The ETCCDI indices have been previously used for monitoring changes in climate extremes across different parts of the world (Zhang et al. 2000; Aguilar et al. 2005; Skansi et al. 2013; Ávila et al. 2016, 2019; Lader et al. 2016). They have been widely used to evaluate the accuracy of the Earth system models and reanalyses in simulating observed temperature and precipitation extremes (Diaconescu et al. 2017; Lader et al. 2017; Avila-Diaz et al. 2020; Rapaić et al. 2015).

c. Evaluation metrics

The reanalyses are evaluated using annual values of climate indices and spatial patterns compared with DAYMET using

the modified Kling–Gupta efficiency [KGE; Eq. (1) below] (Gupta et al. 1999; Kling et al. 2012). The KGE metric has been used by many authors to evaluate the accuracy of atmospheric reanalyses (Chaney et al. 2014; Siqueira et al. 2018; Acharya et al. 2019; Beck et al. 2019a,b; Bhuiyan et al. 2019). The KGE approach combines three components. The Pearson correlation coefficient [CORR; Eq. (2) below] measures the degree of linear relationship between variables that are simulated and measured. CORR values of 1.0 and −1.0 respectively indicate a perfect positive and negative linear relationship between the measured and simulated values, and CORR = 0 implies no linear relationship (Moriasi et al. 2007). The second component of KGE is the bias ratio [BR; Eq. (3) below], which indicates the difference between mean simulated and observed values (Walther and Moore 2005). The best value of BR is 1.0, and magnitudes that are less than or greater than 1.0 indicate underestimation or overestimation, respectively, by simulated values. The last component is the relative variability [RV; Eq. (4) below], which delivers a measure of dispersion for data series. The perfect value of RV is 1.0; RV > 1 indicates overdispersed simulated data, and RV < 1 indicates underdispersed simulated data. Thus, the optimal value of KGE, BR, and RV is 1.0. The definitions are

$$KGE = 1 - [(1 - CORR)^2 + (1 - BR)^2 + (1 - RV)^2]^{1/2}, \quad (1)$$

$$\text{correlation (CORR)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}, \quad (2)$$

$$\text{bias ratio (BR)} = \overline{y}/\overline{x}, \quad \text{and} \quad (3)$$

$$\text{relative variability (RV)} = CV_y/CV_x. \quad (4)$$

In Eqs. (2) and (3), x_i is the gridded observation value, y_i is the gridded product being evaluated, and $\overline{x}$ and $\overline{y}$ are the observed and estimated means. In Eq. (4), CV_x and CV_y are the coefficient of variation of observed and estimated values, respectively.

Skill is evaluated over the watersheds within the study domain (Fig. 1). It is important to note that the gridded observations (DAYMET) do not cover Greenland and Iceland. We also limit the evaluation to areas north of approximately 42°N to avoid boundary discontinuities with ASRv2. For this reason, we consider 14 watersheds: 1) Arctic Ocean Seaboard, 2) Pacific Ocean Seaboard, 3) Yukon River, 4) Mackenzie River, 5) Fraser River, 6) Columbia River, 7) Great Basin, 8) Colorado River, 9) Hudson Bay Seaboard, 10) Nelson River, 11) Missouri River, 12) Mississippi River, 13) St. Lawrence River, and 14) Atlantic Ocean Seaboard. To avoid the long name of watersheds, we restrict our usage to just the first name (e.g., Hudson basin). It is important to know that that the southwestern Great Basin and Colorado basin were considered in this study even though they are not in the North American Arctic; however, they are part of the domain (Fig. 1a).

d. Trends of extreme climate indices

We used the Mann–Kendall (MK; Kendall 1975; Mann 1945) test and Sen's slope (Sen 1968) method to calculate the

statistical significance and magnitude of trends in extreme climate indices. We assessed the trend significance at the 90% confidence level (*p* value ≤ 0.1). For a full discussion of the MK and slope estimator methods and their advantages in climate series, see Yue et al. (2002). These methods were adopted, because a nonparametric approach is less sensitive to outliers in time series of climate extremes series than linear regression (Skansi et al. 2013; Zhang et al. 2000). The MK and Sen's slope methods are widely used to detect trends in climate indices based on daily climate data (Cornes and Jones 2013; Skansi et al. 2013; Rapaić et al. 2015; Ávila et al. 2016, 2019).

It must be stressed that the presence of autocorrelation influences the statistical significance of the MK test (El Kenawy et al. 2011; Croitoru et al. 2016; Li et al. 2018). Consequently, we performed an autocorrelation function to confirm serial correlation in the climate extreme series using the Box–Pierce (Box and Pierce 1970) test at the 95% confidence level (*p* < 0.05). For the series that presented autocorrelation, we used the modified MK method proposed by Hamed and Rao (1998) to assess the significance and detrend the time series. The autocorrelation analysis revealed that 94% of the series show insignificant serial correlation at the lag-1. This indicates that most of the temperature and precipitation series are free from serial correlation.

3. Results

a. Temperature indices

1) PERFORMANCE EVALUATION

Figure 2 shows the evaluation metrics between the reanalyses and observational data for the hottest days (TXx) and Fig. 3 displays the regional performance for each watershed. According to the KGE scores (Fig. 2, first column), ASRv2 and ERA5 show the best performance, while MERRA-2 and GMFD perform the worst for the 2000–16 period, as demonstrated by KGE larger than 0.7 for ASRv2 and ERA5 (Fig. 3a). For KGE values of the key variables described here, see Fig. A1 of the appendix. In general, all reanalyses show consistently lower skill in capturing TXx over the Arctic (basin 1 in Fig. 1), northeastern parts of the Hudson (basin 9), and northern Atlantic basins (basin 14; Figs. 1 and 2). This is supported by lower correlation coefficients [CORR; Fig. 2 (second column) and Fig. 3b], warm biases over northeastern Hudson, and northern Atlantic basins [BR; Fig. 2 (third column) and Fig. 3c], and a general underestimation of the coefficient of variation [RV; Fig. 2 (fourth column) and Fig. 3d].

Except for GMFD, all datasets strongly demonstrate the ability to reproduce the coldest night index (TNn; Fig. 3). All products tend to slightly underestimate the minimum temperature (BR < 1.0) except for NARR, which overestimates TNn values (Fig. 3c). Furthermore, the worst performance among the reanalyses is demonstrated by GMFD, especially over the Arctic, Yukon, Great, Colorado, Nelson, and Mississippi basins (basins 1, 3, 7, 8, 10, and 12, respectively) with KGE values between 0.46 and 0.58 (Fig. 3, fourth column; also see Fig. A1 of the appendix).
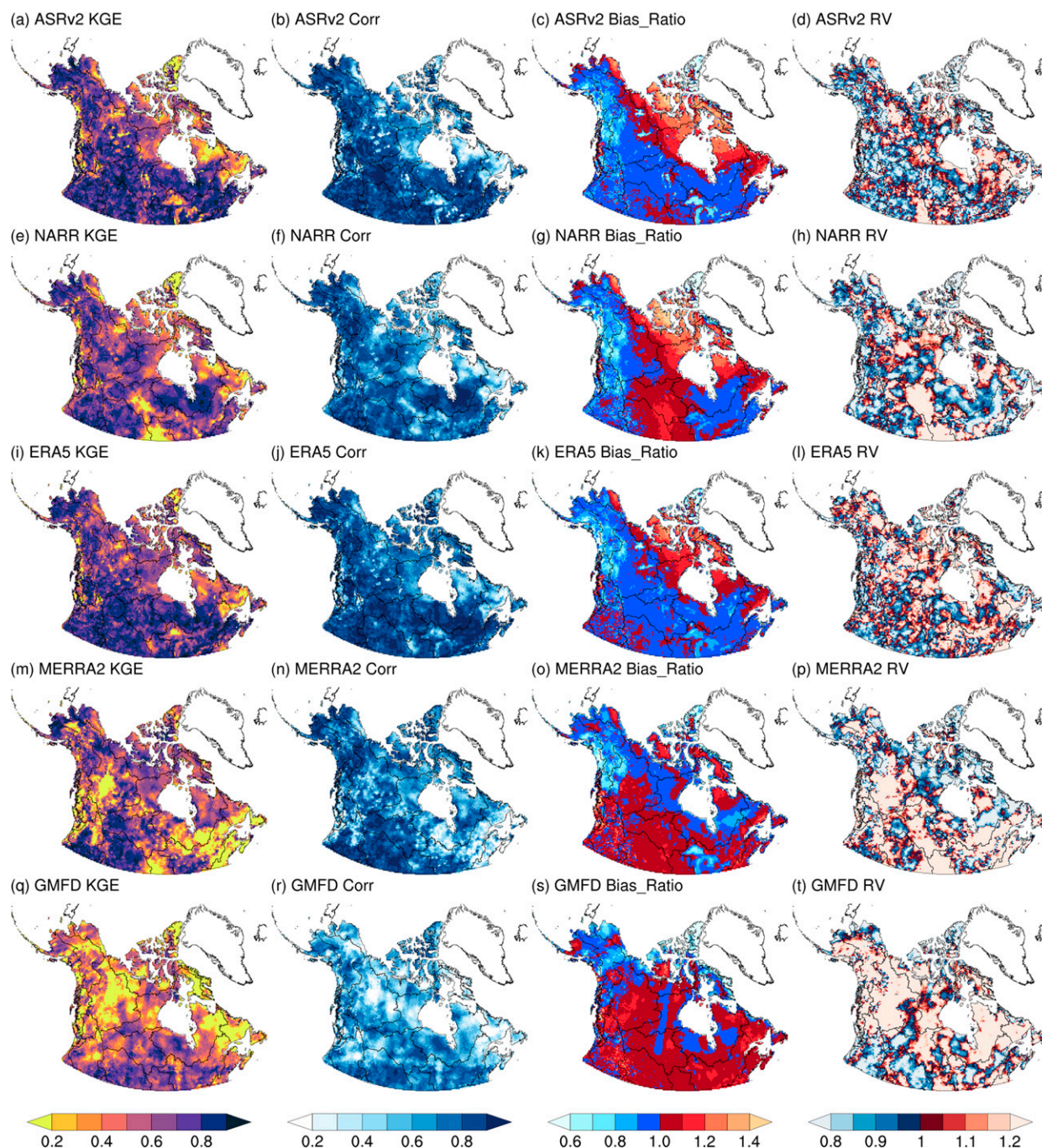
FIG. 2. Performance metrics for warmest annual temperature (TXx) using the DAYMET dataset as a reference over the 2000–16 period for (left) KGE, (left center) correlation (CORR, labeled here as Corr), (right center) bias ratio, and (right) RV scores for (a)–(d) ASRv2, (e)–(h) NARR, (i)–(l) ERA5, (m)–(p) MERRA2, and (q)–(t) GMFD. The optimal value for all metrics is 1.0 indicated by dark colors on the right side of the color bars for KGE and CORR and by the middle of the color bars for bias ratio and RV. Lighter colors illustrate weaker values.

All datasets adequately reproduce the annual percentile indices (TX10p, TX90p, TN10p, and TN90p), with KGE values above 0.7 and CORR above 0.6, but with a moderate and consistent warm bias. However, the lowest scores of KGE are found for warm days (TX90p) over the Colorado basin (basin 8) in all datasets (Fig. 3a, third column). Furthermore, the GMFD performed worse in the west (watersheds 1–8) than the other datasets for percentiles indices derived from daily minimum temperature (e.g., TN10p and TN90p).

The diurnal temperature range (DTR; Fig. 3, seventh column) shows relatively poor performance over the Arctic and
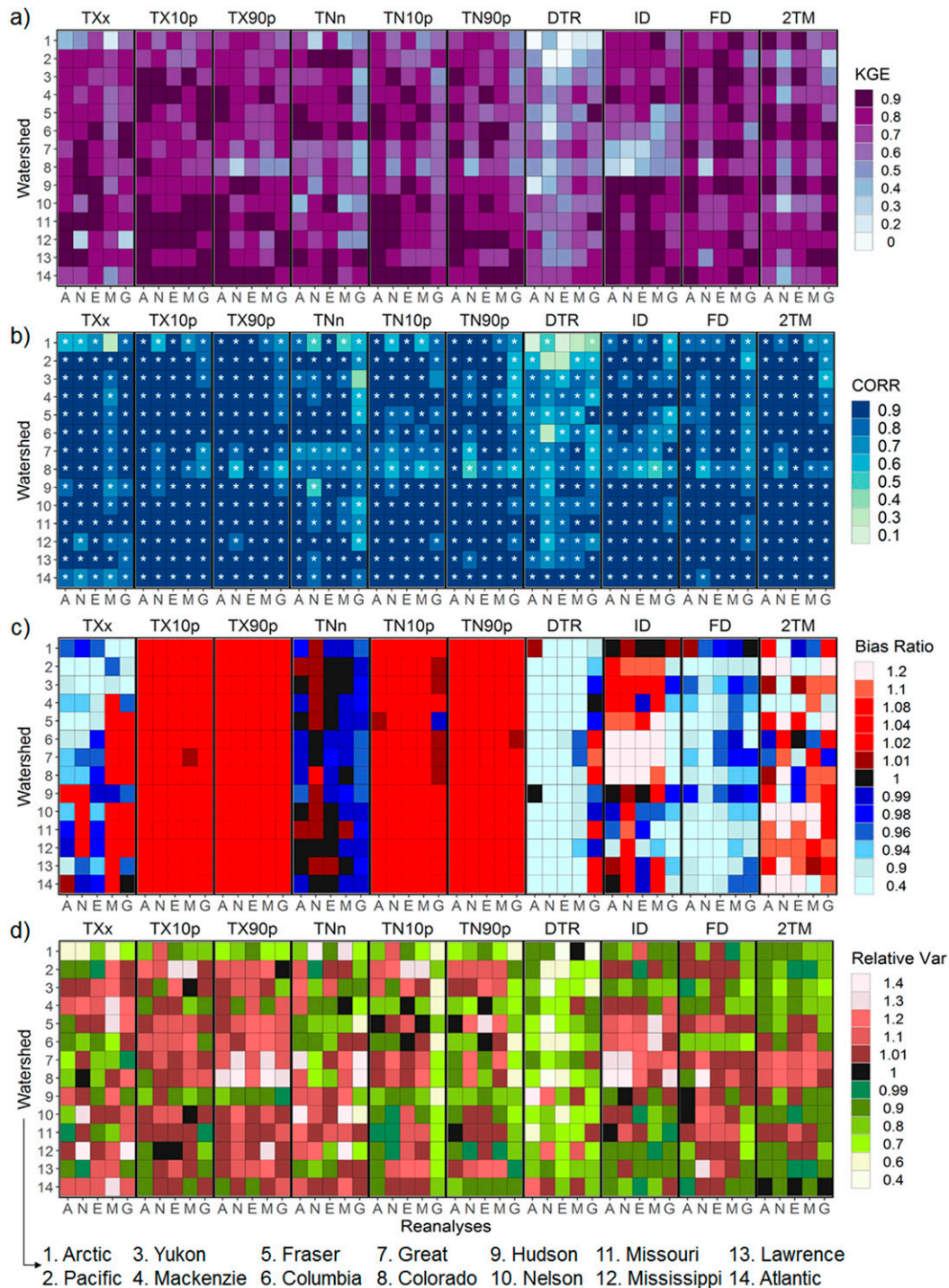
FIG. 3. Evaluation metrics for temperature indices using the DAYMET dataset as a reference over 14 watersheds (Fig. 1) for (a) KGE, (b) CORR, (c) bias ratio, and (d) RV. The period of the analysis is 2000–16. The letters on the x axis represent ASRv2 (A), NARR (N), ERA5 (E), MERRA-2 (M), and GMFD (G). Asterisks indicate a significant correlation at the 95% confidence level. The optimal value for all metrics is 1.0 (dark colors).
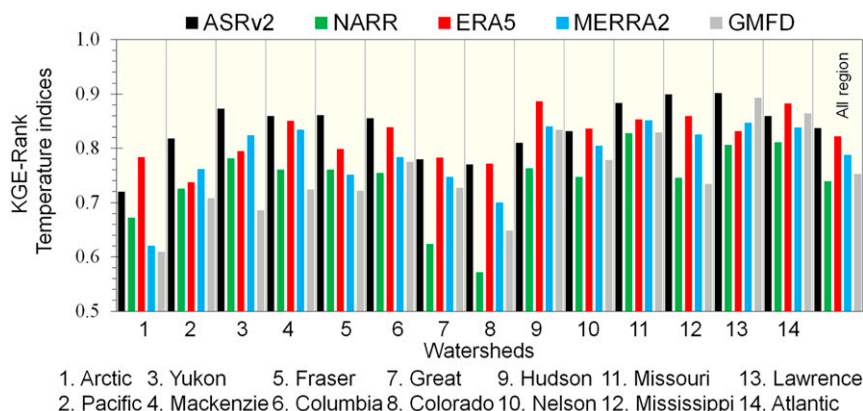
FIG. 4. KGE values for reanalyses in climate temperature indices. The rank is based on the mean of KGE's coefficients in each extreme climate index between the gridded observations (DAYMET) and the reanalysis (ASRv2, NARR, ERA5, MERRA-2, and GMFD).

Pacific in all datasets. For instance, the correlations between datasets and reanalyses are insignificant and weaker (CORR < 0.4) in these northern basins. The reanalyses tend to underestimate the DTR index magnitude except for GMFD, which represents the best performance by reasonably reproducing CORR, BR, and RV (Figs. 3b–d).

Nevertheless, all reanalyses exhibit weaker performance for the DTR index, predominantly at high latitudes (e.g., the Arctic basin), attributed to the relative lack of observational data. Similar to our results, Koenigk et al. (2015) also found that ERA-Interim (Dee et al. 2011) underestimates the Arctic diurnal temperature range during 1980–2005. Note that the number of stations used in DAYMET to derive gridded fields of maximum and minimum temperature during each year, on average, is 67 stations in an area of more than 2.0 million km^2 (Figs. 1a,b). For DTR, reproducing spatial patterns is a particular challenge, because in situ data are influenced by their proximity to water (Wilson et al. 2011), which can lead to a smaller diurnal temperature ranges in reanalyses.

Ice days (ID) and frost days (FD) show that most of the reanalyses perform reasonably well for both indices. However, over the Columbia, Great Basin, and Colorado watersheds (basins 6–8), reanalyses display KGE values sometimes less than 0.4 with substantial overestimation in the ID index. All datasets tend to underestimate the FD index, with fewer days than observed when the daily minimum temperature is below 0°C.

An important point to note is that, as shown by Beck et al. (2019b), CORR is the most important factor in getting greater values of KGE. For example, the CORR from some temperature indices (e.g., TXx, TNn, TN10p, TN90, DTR, and ID) are generally lower in the Arctic, Great Basin, and Colorado watersheds (basins 1, 7, and 8), relative to what is observed for the CORR from the other indices (Fig. 3b).

Over most basins, ASRv2, ERA5, and GMFD demonstrate the smallest 2-m temperature (2TM) warm biases (Fig. 3c). The worst performance among all metrics is

found for NARR, especially over the Pacific, Mackenzie, Colorado, Nelson, and Atlantic basins (basins 2, 4, 8, 10, and 14). For those basins, the KGE varies between −0.33 and 0.57 (Fig. 3c, last column; see also Fig. A1 of the appendix). Betts et al. (2009) showed that ERA-Interim has a cold bias in 2TM over the Mackenzie basin, and our results indicate that the new generation of this reanalysis (ERA5) underestimates 2TM over several basins (basins 1, 2, 4, 7, 8, and 9).

Figure 4a shows the ranking based on KGE scores for reanalyses in climate temperature indices over the 14 watersheds considered in this study. To that end, 10 temperature indices were used to evaluate the ability of the five reanalyses to reproduce climatic extremes for air temperature over a given basin (Fig. 1). The best performance is noted in the ASRv2, followed by ERA5, MERRA-2, GMFD, and NARR over the whole region (Fig. 4a, far right).

In general, reanalyses are similar for temperature extremes based on Fig. 3. We identified some discrepancies across various reanalyses over the Arctic basin in the TXx, TNn, and DTR indices. Our evaluation demonstrates that the use of gridded data with sparse observations must be done carefully, and reanalyses may be more adequate since they merge several sources of data such as satellite data, weather stations, and radiosonde data (Hoffmann et al. 2019; Wilson et al. 2011; Dee et al. 2014). ASRv2 and ERA5 have shown greater ability to reproduce observational temperature extreme patterns compared to earlier products.

However, reanalyses are not without their faults and must also be viewed with caution due to the relatively sparse observational network from which to assimilate data. Other approaches, such as the multiple reanalyses ensemble (REM), could be a useful product for study of the polar regions (Uotila et al. 2019; Diaconescu et al. 2017). However, Diaconescu et al. (2017), found good performance of the REM for near-surface temperature and hot extremes (TXx, TX90, and TN90p) but not for other climate extreme indices [e.g., TNn, annual total wet-day precipitation (PRCPTOT), R1mm, and R95p] over the Canadian Arctic during 1980–2004. For instance, the poor
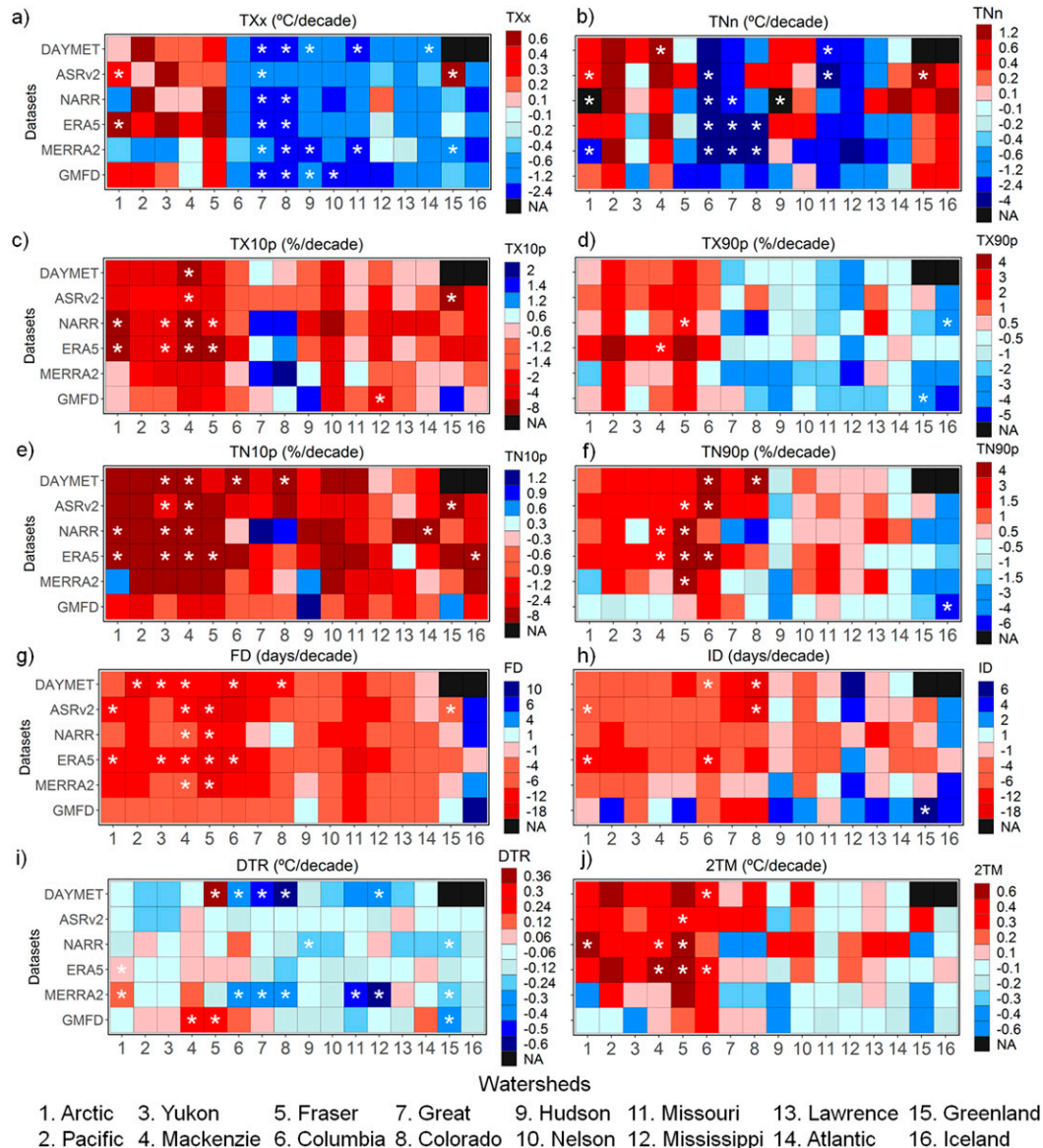
FIG. 5. Decadal trends in temperature indices calculated from the DAYMET, ASRv2, NARR, ERA5, MERRA-2, and GMFD over 16 watersheds (Fig. 1) for the period 2000–16. The asterisks indicate a statistically significant trend at the 90% confidence level; NA (black boxes) indicates that no data value is stored for the variable in the observation (DAYMET).

performance (low KGE values) of REM is similar in northern basins (e.g., the Arctic and Hudson, basins 1 and 9) for all climate indices (see Figs. A1 and A2 of the appendix).

2) TRENDS IN TEMPERATURE INDICES

Figures 5 and 6 illustrate the decadal trends of temperature indices from 2000 to 2016. The warm extremes indices (TXx, TX90p, and TN90p) show positive trends for the northern and western part of the study domain, principally in the Arctic, Pacific, Yukon, Mackenzie, Fraser, and Columbia basins (basins 1–6). For those basins, regional trends of TXx are positive between 0.07 and

$0.77°C$ decade^{-1} (Figs. 5a and 6). The maximum magnitudes of TXx are found over the Fraser basin, especially for NARR ($0.82°C$ decade^{-1}), ERA5 ($0.69°C$ decade^{-1}), GMFD ($0.48°C$ decade^{-1}), and observations ($0.45°C$ decade^{-1}) (Figs. 5a and 6).

On the other hand, TXx (Figs. 5a and 6) and TX90 (Fig. 5d) display west to east negative trends for basins 7 to 16, the Great Basin to Iceland. In general, there is reasonable agreement in the positive trends for most products, except for GMFD that shows a remarkable decrease in the percentage of warm nights (TN90p; Fig. 5f) over the Pacific, Yukon, and Mackenzie basins (basins 2–4) between −0.1% to −0.7% of days decade^{-1}.
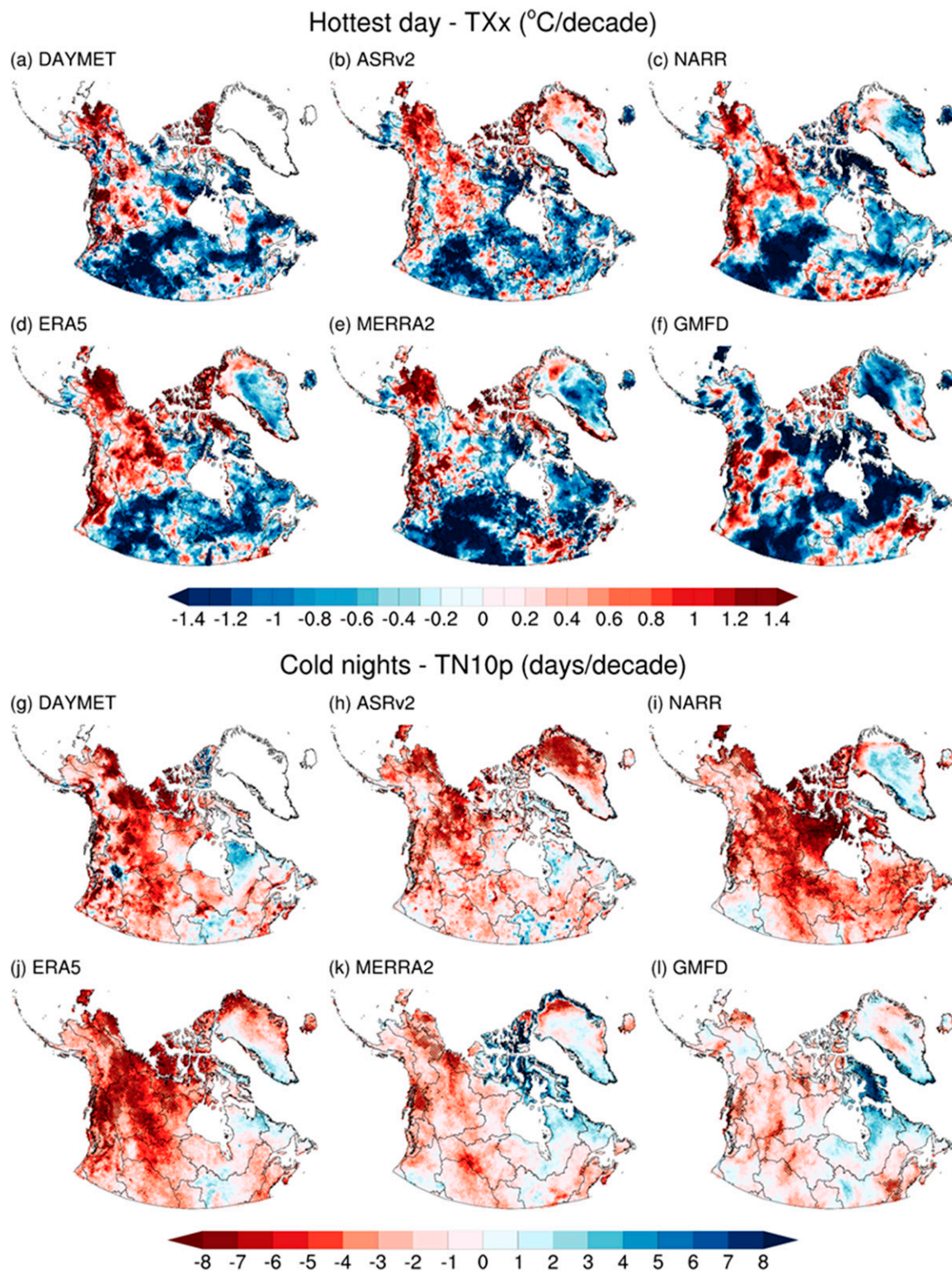
Hottest day - TXx (°C/decade)

(a) DAYMET   (b) ASRv2   (c) NARR

(d) ERA5   (e) MERRA2   (f) GMFD

-1.4 -1.2 -1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1 1.2 1.4

Cold nights - TN10p (days/decade)

(g) DAYMET   (h) ASRv2   (i) NARR

(j) ERA5   (k) MERRA2   (l) GMFD

-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8

FIG. 6. The 2000–16 decadal trends for the (a)–(f) TXx and (g)–(l) TN10p indices calculated from observations (DAYMET) and reanalyses (ASRv2, NARR, ERA5 MERRA-2, and GMFD). The areas with hatching show significant trends at the 90% confidence level.

Figures 5b, 5c, and 5e show the trends in extremes cold indices (TNn, TX10p, and TN10p). Notably, TNn (Fig. 5b) displays positive trends and good agreement in spatial patterns between observations (DAYMET) and reanalyses over the Pacific, Mackenzie, and Iceland basins (basins 2, 4, and 16) and downward trends in the southwest and southeast parts of the study domain (e.g., Columbia, the Great Basin, Missouri, and Mississippi, basins 6, 7, 11, and 12). A good match is also indicated for TX10p (Fig. 5c) and TN10p (Figs. 5e and 6g–l); both exhibit a decrease in the frequency of cold days over all watersheds. Similar results were found by Grotjahn et al. (2016) over North America between 1950 and 2012.

The frequency of indices related to cold conditions, such as frost days (Fig. 5g) and ice days (Fig. 5h) highlight negative trends for the majority of basins, except for the GMFD dataset that shows a mixed pattern for ID. However, positive trends are found in GMFD for the ID index over the Mississippi, Lawrence, Atlantic, Greenland, and Iceland basins (basins 12–16), with magnitudes varying between 2 and 6 days decade^{-1} (Fig. 5h). On the other hand, GMFD show a remarkable decrease in ID over the southwestern portion of the domain (e.g., the Columbia, Great, and Colorado basins, basins 6–8) between −2 and −12 days decade^{-1}.

The annual mean temperature (2TM; Fig. 5j) shows a warming pattern for the Arctic, Pacific, Yukon, Mackenzie, Fraser, and Columbia basins (basins 1–6, respectively). This agrees well with those obtained by Lindsay et al. (2014) and Simmons and Poli (2015), who found positive trends in 2TM for the Arctic region over the 1981–2010. Also, Rapaić et al. (2015) observed positive trends in 2TM over the Mackenzie basin and Great Basin (basin 7) from 1981 to 2010, reflecting that this pattern has continued during the last decades, especially in northwestern watersheds. Exceptionally, GMFD delivers opposite (negative) of the 2TM trends over the Arctic, Pacific, and Yukon basins.

Decreasing trends are observed in the DTR index throughout most of the domain, with most reanalyses demonstrating similar trends in the central (e.g., 8–12) and northeastern (e.g., 15 and 16) watersheds (Fig. 5i). This could be related to the increase in the annual mean temperature and minimum temperature, which are greater than trends in maximum air temperature, which has a slightly positive trend. A similar pattern was found by Qu et al. (2014) from observational datasets over the continental United States from 1911 to 2012.

Temperature indices show annual warming trends during 2000–16 over northwestern basins, and the overall trend is spatially consistent among the reanalyses (e.g., ASRv2, NARR, and ERA5) and previous studies (Lindsay et al. 2014; Rapaić et al. 2015; Simmons and Poli 2015; Matthes et al. 2016; Shepherd 2016; Hu and Huang 2020). However, it is interesting to note that NARR's performance in reproducing the observed trends in temperature indices over the Great Basin and Colorado watersheds (basins 7 and 8) is subpar compared to observations (see Figs. 5c–g and 5j). The regional trends gradually decrease from western to eastern areas such as the Atlantic and Greenland basins (14 and 15) in maximum temperature and annual mean temperature. Furthermore, Donat et al. (2016) found that warm extremes (e.g., TN90p and TX90p) and cold extremes (e.g., TN10p and TX10p) delivered warming patterns over high latitudes of North America during the last six decades (1951–2010). They used several gridded datasets between

observations (e.g., HadEX2) and reanalyses (e.g., ERA-20C and ERA-20CM) and demonstrated that those percentiles indices show stronger global warming in the early 2000s than the early 1980s.

b. Precipitation indices

1) PERFORMANCE EVALUATION

Figure 7 shows the performance metrics for the wet-days index (R1mm) compared to DAYMET for 2000–16. Figure 8 displays the KGE, CORR, BR, and RV over the 14 watersheds. The overall performance score (KGE) shows that reanalyses demonstrate less ability to reproducing extreme precipitation indices over northern basins (e.g., the Arctic and Hudson basins, basins 1 and 9) for almost all extreme precipitation indices (Figs. 7 and 8).

The PRCPTOT shows that over the southwest, south, and southeast watersheds, KGE values are between 0.5 and 0.94 (Fig. 8); the best results are found over the Nelson, Missouri, and Mississippi Rivers (basins 10–12). Nevertheless, the reanalyses show the lowest values over the Arctic, Pacific, Yukon, and Hudson basins (basins 1, 2, 3, and 9, respectively).

Bromwich et al. (2016) found dry annual total precipitation biases for Arctic stations between observations and two reanalyses (ASRv1 and ERA-Interim) from December 2006 to November 2007. According to our results, the new generations of these reanalyses (ASRv2 and ERA5) still have dry annual biases over the Arctic and Hudson watersheds (Fig. 8c, first column). Furthermore, similar to our results, Wong et al. (2017) found dry bias from 2002 to 2012 over eastern Canada in the NARR and GMFD precipitation products when compared with the precipitation-gauge stations.

Murdock et al. (2013) found a low correlation between gridded observations and NCEP2 Reanalysis (NCEP–DOE AMIP-II Reanalysis) in the Canadian Columbia Basin (basin 6) for 1980–2000. It should be stressed that we found significant correlation coefficients between 0.92 and 0.99 (Fig. 8b, line 6) over the Columbia watershed. Thus, our results deliver strong evidence that an increase in horizontal spatial resolution can lead to better performance in regions with complex topography such as the Columbia basin located in the Rocky Mountains.

It is important to note that northern and western basins have low station density as well, which can induce errors in estimating values of precipitation in DAYMET (Fig. 1b). The average of the total number of rainfall observing stations between 2000 and 2016 for the Arctic, Yukon, and Hudson (basins 1, 3, and 9) are 41, 85, and 66 over areas with 2.09, 0.85, and 2.87 million km^2, respectively (Figs. 1a,b). This makes the performance comparison very difficult because of limitations in the DAYMET dataset (Daly et al. 2008; McEvoy et al. 2014; Timmermans et al. 2019).

Precipitation intensity indices such as the maximum 1-day precipitation (RX1day; Fig. 8, second column), maximum 5-day precipitation (RX5day; Fig. 8, third column), very wet days > 95th percentile (R95p; Fig. 8, fourth column), and daily intensity index (SDII; Fig. 8, fifth column), show the lowest values of KGE and insignificant correlations (CORR < 0.2) over northern watersheds (e.g., the Arctic and Hudson).
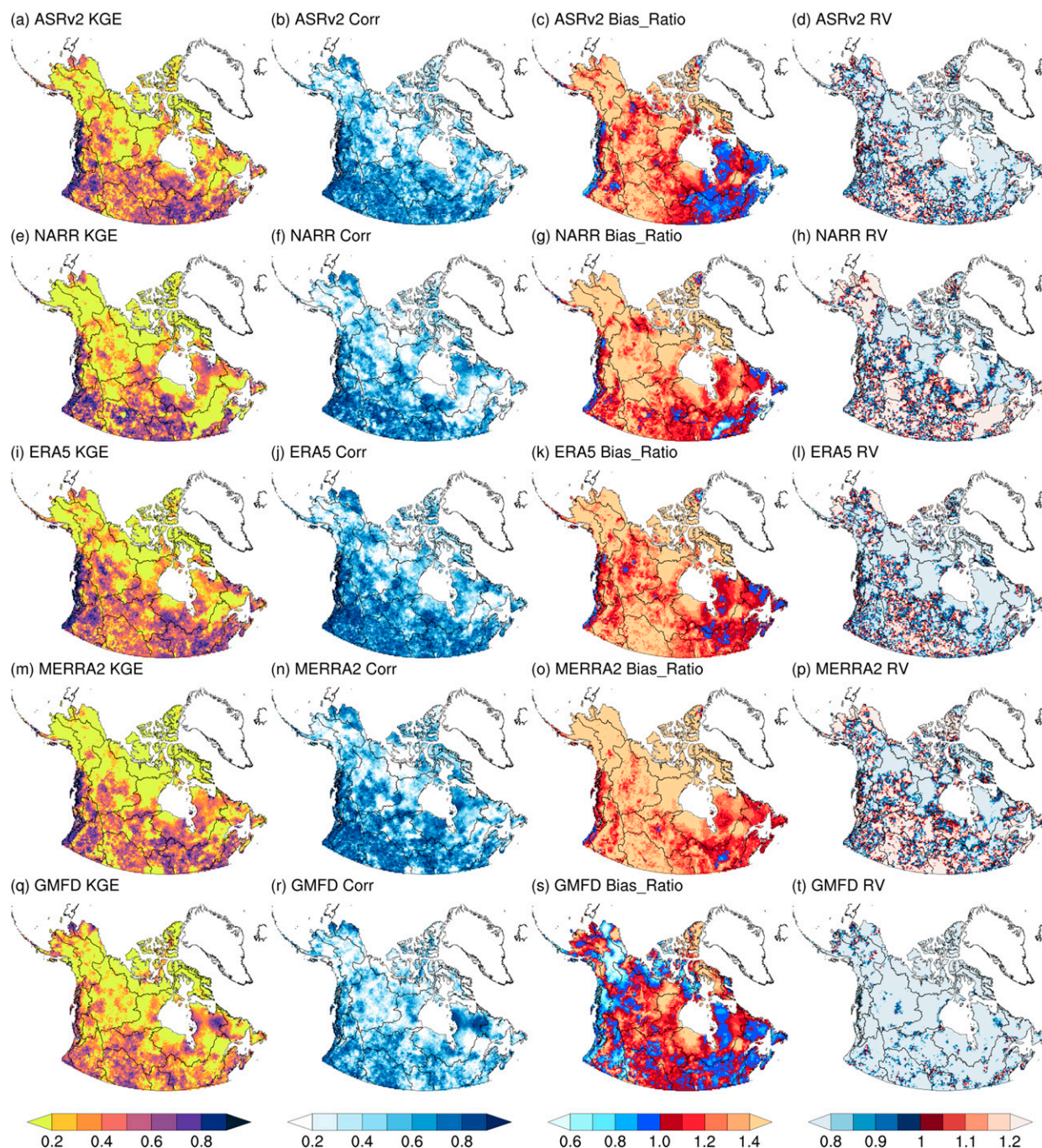
FIG. 7. Performance metrics for the wet days (R1mm) using the DAYMET dataset as a reference over the 2000–16 period for (left) KGE, (left center) CORR, (right center) bias ratio, and (right) RV scores for (a)–(d) ASRv2, (e)–(h) NARR, (i)–(l) ERA5, (m)–(p) MERRA2, and (q)–(t) GMFD. The optimal value for all metrics is 1.0, indicated by dark colors on the right side of the color bars for KGE and CORR and by the middle of the color bars for bias ratio and RV. Lighter colors illustrate weaker values.

Results show that reanalyses generally underestimate RX1day and RX5day and overestimate R95p in most watersheds (Fig. 8c). Particularly, ASRv2 does not underestimate the observation for the RX1day index in almost all watersheds except for the Mississippi basin (basin 12).

Similar performance is found for the frequency index R1mm (see Fig. 7 and the sixth column of Fig. 8). Reanalyses have low KGE values below, 0.5 over the Arctic, Mackenzie, Colorado, Hudson, and Atlantic basins (basins 1, 4, 8, 9, and 14; see Fig. A2 of the appendix). However, the ASRv2 and ERA5
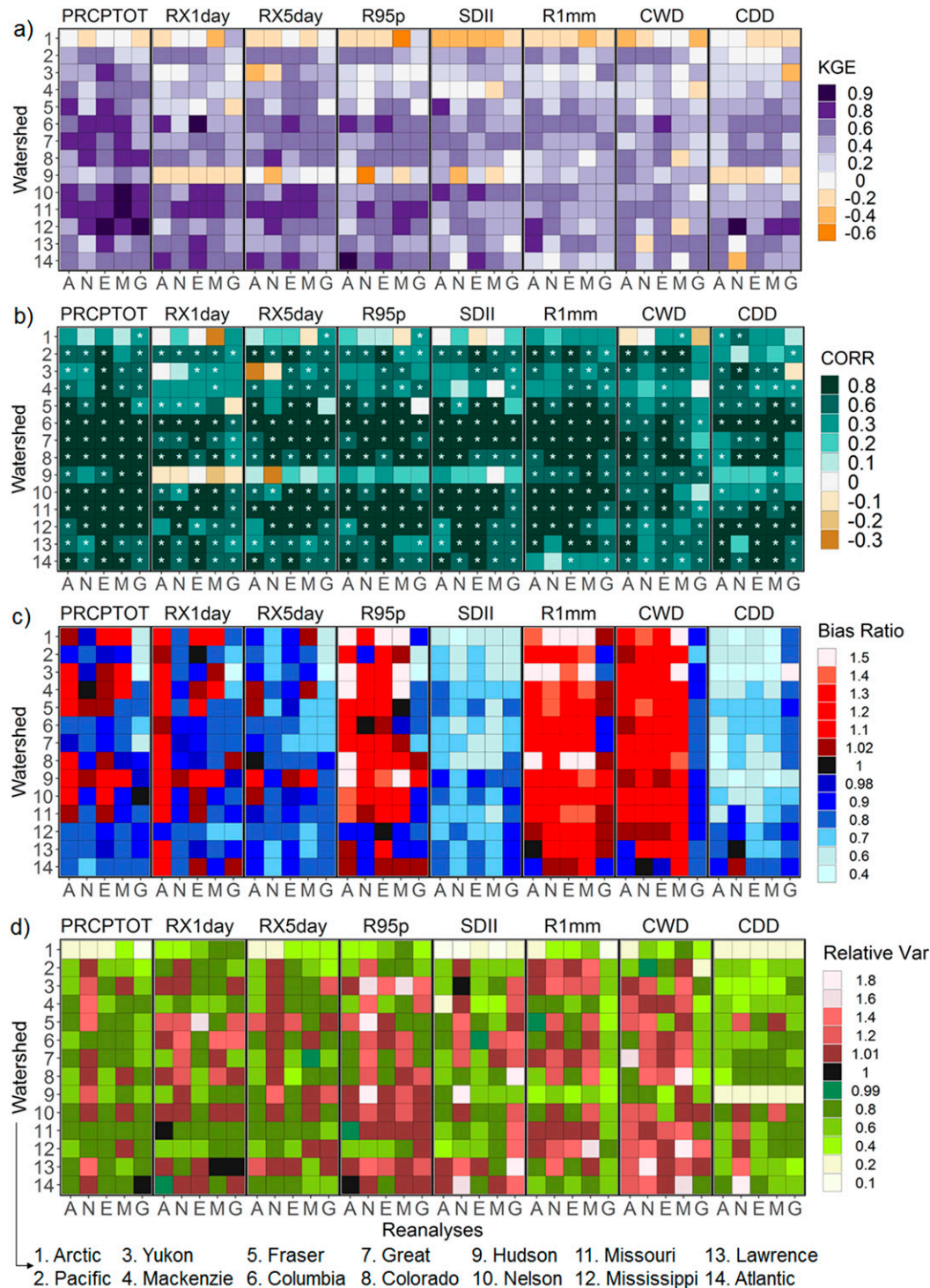
FIG. 8. Evaluation metrics for precipitation indices using the DAYMET dataset as a reference over 14 watersheds (Fig. 1) for (a) KGE, (b) CORR, (c) bias ratio, and (d) RV. The period of the analysis is 2000–16. The letters on the x axis represent ASRv2 (A), NARR (N), ERA5 (E), MERRA-2 (M), and GMFD (G). Asterisks indicate a significant correlation at the 95% confidence level. The optimal value for all metrics is 1.0 (dark colors).
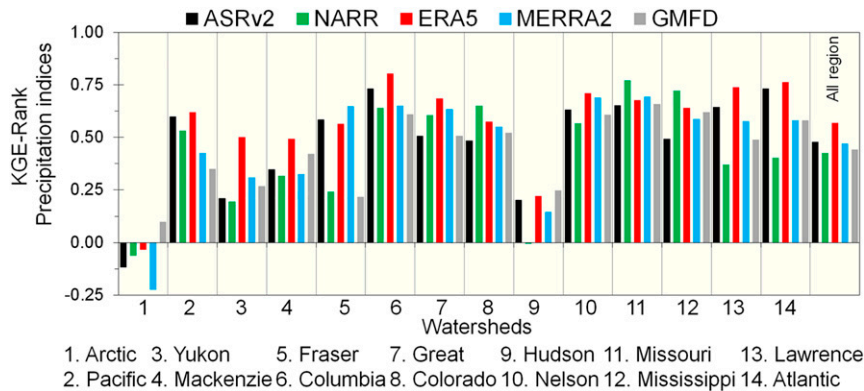
FIG. 9. KGE values for reanalyses in climate precipitation indices. The rank is based on the mean of KGE's coefficients in each extreme climate index between the gridded observations (DAYMET) and the reanalysis (ASRv2, NARR, ERA5, MERRA-2, and GMFD).

display the greatest skill in capturing the interannual variation of the number of wet days over the Nelson, Missouri, Mississippi, and Lawrence basins (basins 10–13). For those basins, the KGE scores varied between 0.56 and 0.82 and between 0.47 and 0.79 for the ASRv2 and ERA5, respectively. Furthermore, the bias findings for the R1mm index agree with those obtained by Diaconescu et al. (2017) over the Canadian Arctic land areas. According to Diaconescu et al. (2017), the bias of the R1mm index reveals a wetter pattern in almost all reanalyses (e.g., CFSR, ERA-Interim, JRA-55, and MERRA-2) over 25 years (1980–2004).

For consecutive wet days (CWD) and consecutive dry days (CDD), reanalyses show KGE values between 0.4 and 0.8 (Fig. 8a), with the weakest values for the Arctic and Hudson basins (basins 1 and 9). However, for the rest of the study domain, reanalyses have significant correlations for CWD and CDD (see the last two columns of Fig. 8b).

Figure 9 displays the overall performance ranking based on KGE values for reanalyses in precipitation obtained by averaging of KGE in each watershed, using the set of eight indices. The greatest overall skill over the study domain was revealed by ERA5, followed by ASRv2, MERRA-2, GMFD, and NARR (see the last column of Fig. 9a). However, the rankings over the Colorado, Missouri, and Mississippi basins (basins 8, 11, and 12) demonstrate that NARR is the best reanalysis for capturing precipitation variability and extremes as is MERRA-2 for the Fraser watershed (basin 5; Fig. 9). The dataset choice should be made carefully, because each reanalysis shows varying accuracy over each region (Beck et al. 2017; Yin et al. 2015; Timmermans et al. 2019; Rapaić et al. 2015).

2) TRENDS IN PRECIPITATION INDICES

Figure 10 shows the decadal trends for precipitation indices in DAYMET and reanalyses from 2000–16. Also, to illustrate the spatial patterns of trends (Fig. 11), we chose two precipitation indices (PRCPTOT and R1mm). There is a spatially homogenous increase in annual total wet-day precipitation index (PRCPTOT; Figs. 10a and 11a–f) over southwestern and southern basins such as the Fraser, Columbia, Great Basin,

Colorado, Nelson, Missouri, and Mississippi watersheds (basins 5, 6, 7, 8, 10, 11, and 12, respectively) with trends varying from 16 to 150 mm decade^{-1}. The patterns of changes in PRCPTOT are very similar to the RX1day, RX5day, and R95p (Figs. 10b–d), showing an increase in intense events; also, observations and reanalyses display good agreement in the direction and magnitude of the trend over the southwestern part of the domain, especially in the Columbia, Great Basin, Colorado, Nelson, and Missouri basins.

The PRCPTOT index displays a nonhomogenous pattern that was found for northern watersheds (e.g., the Arctic, Mackenzie, and Hudson basins, basins 1, 4, and 9; Figs. 11a–f. Rapaić et al. (2015) and Yang et al. (2019) found similar results for the Canadian Arctic from 1981 onward and all of Canada during the 1950–2012 period. The limited agreement in wet/dry patterns among observations and reanalyses can be explained by the short period for trend analysis and uncertainties in gridded precipitation observations (Diaconescu et al. 2017; Rapaić et al. 2015).

The SDII index (Fig. 10e) delivers better agreement between gridded observational data and reanalyses, and an increase in daily intensity occurs for all basins except for the Arctic and Mackenzie basins (basins 1 and 4) where the trend is not clear. According to Booth et al. (2012), SDII is an indicator of the intensification of the hydrologic cycle; in our study, the upward trend in the SDII index matches with the warming pattern shown in several temperature indices (e.g., TN10p, TX10p, FD, and ID).

The number of wet days (R1mm; Fig. 10f) shows an increase in most datasets over southwest parts and central parts of the domain (e.g., the Columbia, Great Basin, Colorado, Nelson, and Missouri watersheds, basins 6, 7, 8, 10, and 11), ranging between 2 and 18 days decade^{-1}. Also, consecutive wet days (CWD; Fig. 10g) delivered negative trends in the Pacific, Mackenzie, Hudson, Nelson, Lawrence, and Atlantic basins (basins, 2, 4, 9, 10, 13, and 14). On the other hand, the Arctic, Pacific, Yukon, and Hudson watersheds (basins 1, 2, 3 and 9), as well as Greenland and Iceland (basins 15 and 16) have increases in consecutive dry days (CDD; Fig. 10h).
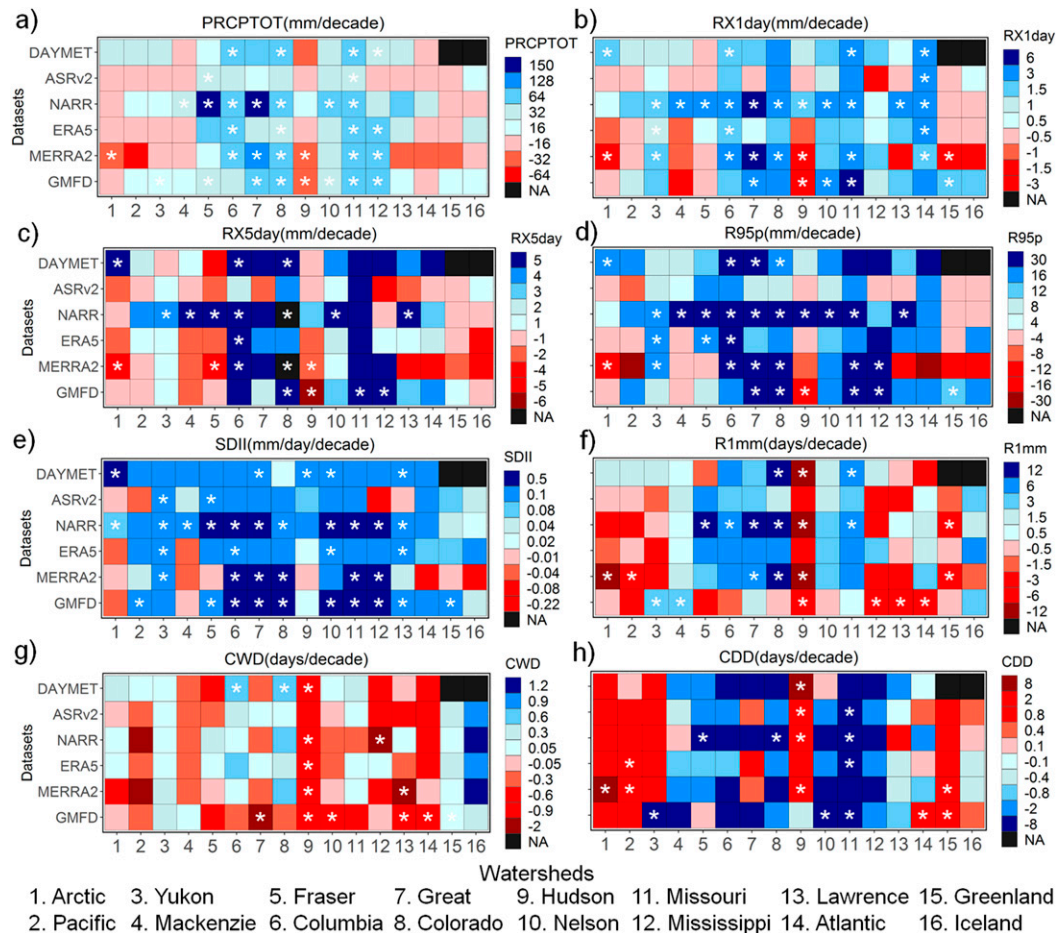
FIG. 10. Decadal trends in precipitation indices calculated from the DAYMET, ASRv2, NARR, ERA5, MERRA-2, and GMFD over 16 watersheds (Fig. 1) for the period 2000–16. The asterisks indicate a statistically significant trend at the 90% confidence level; NA (black boxes) means that no data value is stored for the variable in the DAYMET (gridded observation).

We conclude the temperature trends deliver better consistency in terms of the trend signals across all datasets than those trends in precipitation indices. Nevertheless, the GMFD dataset performed poorly in terms of capture the observed trends for intensity/frequency indices (Figs. 10b–h). Notably, we have demonstrated an increase in intensity and frequency of extreme precipitation events during 2000–16 over southwestern and southern parts of the study domain, especially over the Columbia, Great Basin, Colorado, Nelson, Missouri, and Mississippi watersheds (basins 6, 7, 8, 10, 11, and 12).

c. Extreme winters over the North American Arctic region

Having established the ability of ASRv2 and ERA5 to reproduce extreme climate indices, we investigated their monthly maximum (TXx), minimum (TNn), total wet-day precipitation (PRCPTOT), and number of wet days (R1mm) related to the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO) for 2000–16 (Figs. 12–14). While the NAO and AO are significantly correlated (CORR = 0.86; $p < 0.01$) during winter [December–March (DJFM)] over the 2000–16

period, the negative and positive phases of the AO/NAO produce different responses in the climate extremes across the Arctic (Wanner et al. 2001; Cohen et al. 2010; Dai and Tan 2017; Luo et al. 2020). It is important to note that this exercise aims not to create a new analysis based on a seasonal scale but to evaluate the capacity of these reanalyses to estimate ETCCDI indices during extreme winters. Further details about North American extreme temperature and precipitation events and related large-scale meteorological patterns are found in Grotjahn et al. (2016) and Barlow et al. (2019).

The DJFM period was selected to maximize the influence of NAO/AO on the North American Arctic climate (Kunkel and Angel 1999; Ning and Bradley 2015; Rivière and Drouard 2015; Dai and Tan 2017). The NAO and AO are generated by projecting the lower-level geopotential height anomalies onto the empirical orthogonal function loading vectors of the NAO and AO mode, respectively, details of which can be found online (https://www.cpc.ncep.noaa.gov).

The evaluation considers temperature and precipitation responses to NAO (AO) indices larger than 0.71 (0.96) standard

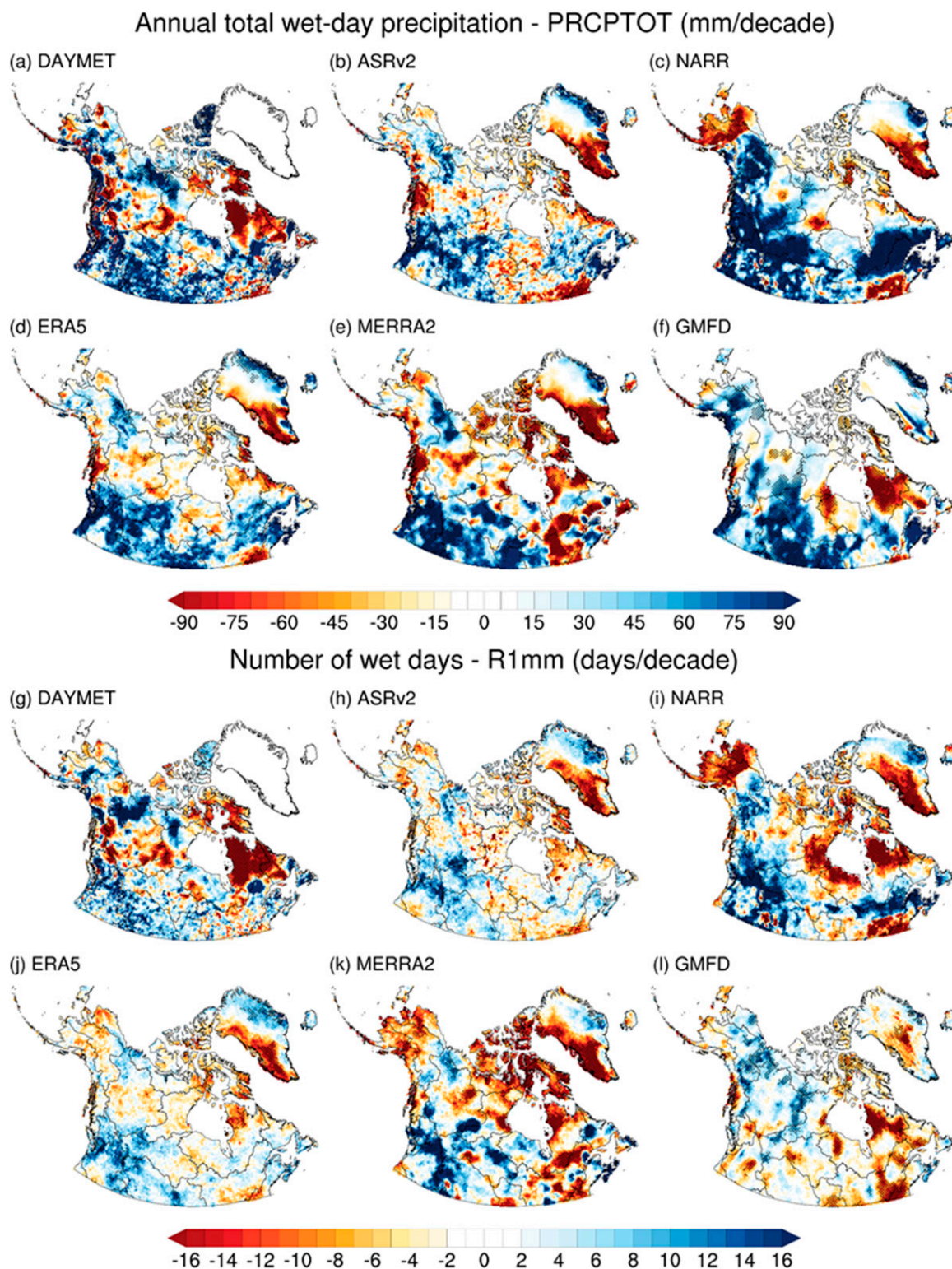## Annual total wet-day precipitation - PRCPTOT (mm/decade)

FIG. 11. The 2000–16 decadal trends for the (a)–(f) PRCPTOT and(g)–(l) R1mm indices calculated from Observations (DAYMET) and reanalyses (ASRv2, NARR, ERA5 MERRA-2, and GMFD). The areas with hatching show significant trends at the 90% confidence level.
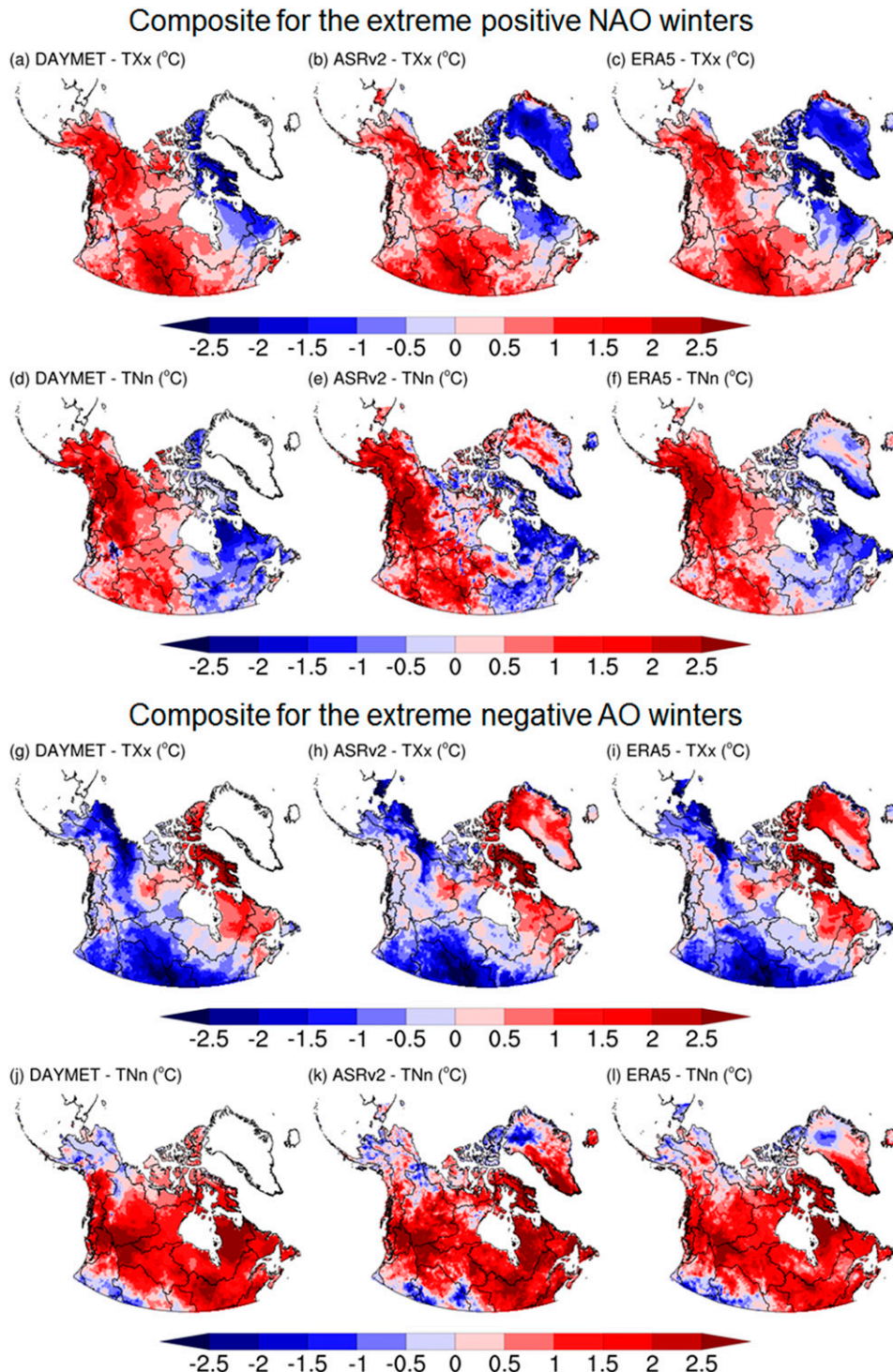
Composite for the extreme positive NAO winters

(a) DAYMET - TXx (°C)   (b) ASRv2 - TXx (°C)   (c) ERA5 - TXx (°C)

-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

(d) DAYMET - TNn (°C)   (e) ASRv2 - TNn (°C)   (f) ERA5 - TNn (°C)

-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

Composite for the extreme negative AO winters

(g) DAYMET - TXx (°C)   (h) ASRv2 - TXx (°C)   (i) ERA5 - TXx (°C)

-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

(j) DAYMET - TNn (°C)   (k) ASRv2 - TNn (°C)   (l) ERA5 - TNn (°C)

-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5

FIG. 12. Composite anomalies of TXx and TNn during extremes winters (DJFM) from 1999/2000 to 2015/16 (no data for December 1999) in (a)–(f) the positive phase of the NAO and (g)–(l) the negative phase of the AO.

deviations from the long-term means (1950–2016). We composited the winters of 2000, 2012, 2014, 2015, and 2016 according to the positive phases of NAO and the 2001, 2006, 2010, and 2013 winters for the negative phases of AO. Note the lack of significant NAO negative events during the period (see Fig. A3 of the appendix). The criteria used to determine intense NAO/AO events ignored the presence of NAO (AO) negative (positive) values to illustrate the most negative and
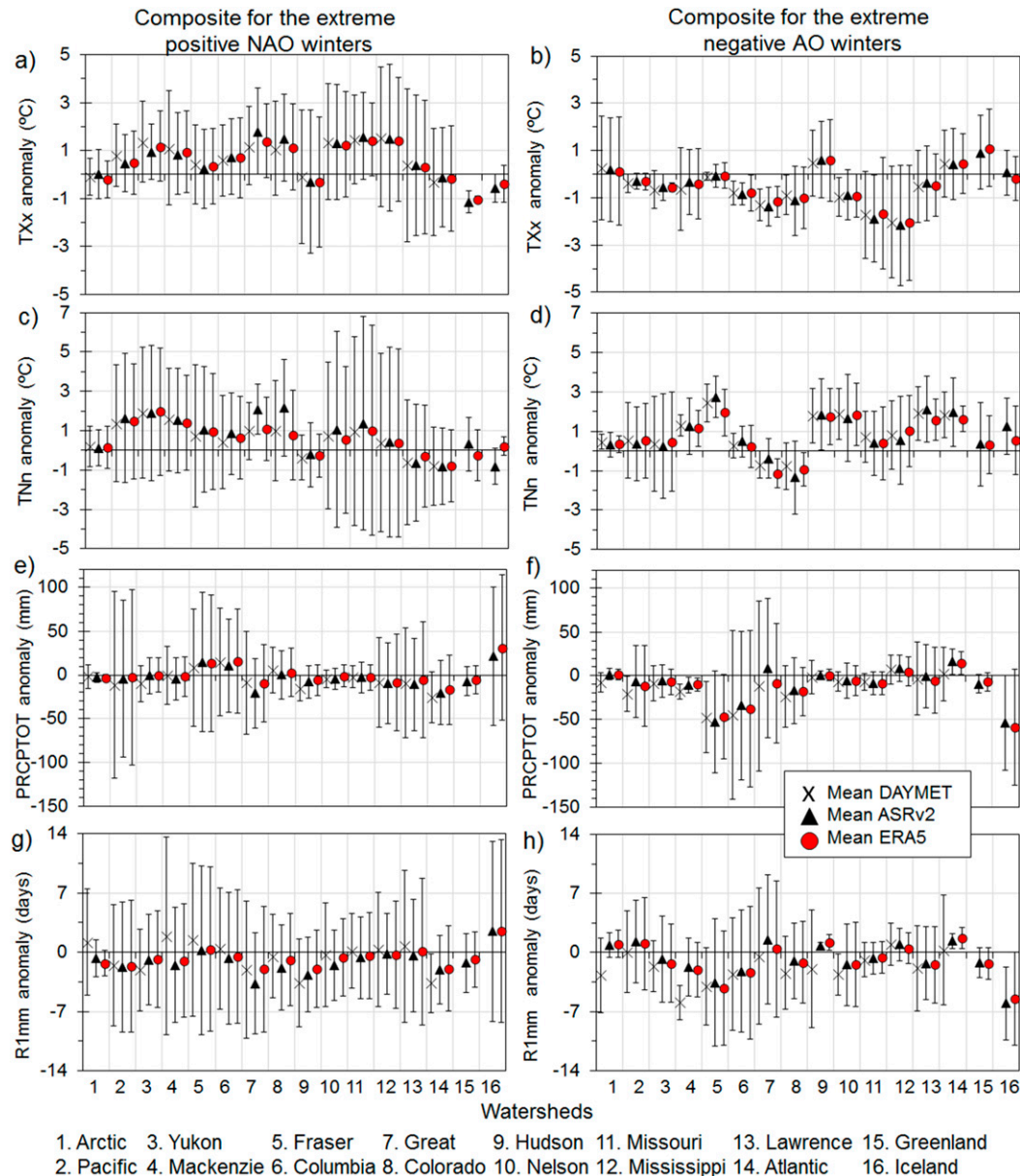
FIG. 13. Composite of the extreme winters (DJFM) over 16 watersheds (Fig. 1) for the (left) North Atlantic Oscillation (NAO) and (right) Arctic Oscillation (AO) for (a),(b) TXx; (c),(d) TNn; (e),(f) PRCPTOT; and (g),(h) R1mm. Error bars represent the standard deviations of the mean in DAYMET, ASRv2, and ERA5.

positive magnitudes recorded in the NAO/AO series (see the asterisks in appendix Fig. A3).

Figures 12a–f show that extreme winters driven by the positive phase of the NAO typically result in below-average temperatures across northeastern North America and Greenland and above-average temperatures across the United States in response to changes in atmospheric flow (Hurrell et al. 2003; Hurrell and Deser 2009; Ning and Bradley 2015). Enhanced northerly winds on the backside of a strengthened Icelandic low induce northerly wind anomalies that produce colder air intrusions across northeast North America, while a stronger

subtropical high in the Atlantic leads to southerly, clockwise flow across much of the United States and western North America. Note the strong agreement between ASRv2 and ERA5 when compared with DAYMET.

Figure 13 shows that the watershed-averaged DJFM anomalies support Fig. 12, with strong positive anomalies in TXx and TNn over the Yukon, Mackenzie, Great Basin, Colorado, Nelson, Missouri, and Mississippi watersheds (basins 3, 4, 7, 8. 10, 11, and 12, respectively) and negative anomalies over the Hudson, Atlantic, and Greenland basins (basins 9, 14, and 15; Figs. 12 and 13a,c). In some parts of the Greenland, Lawrence
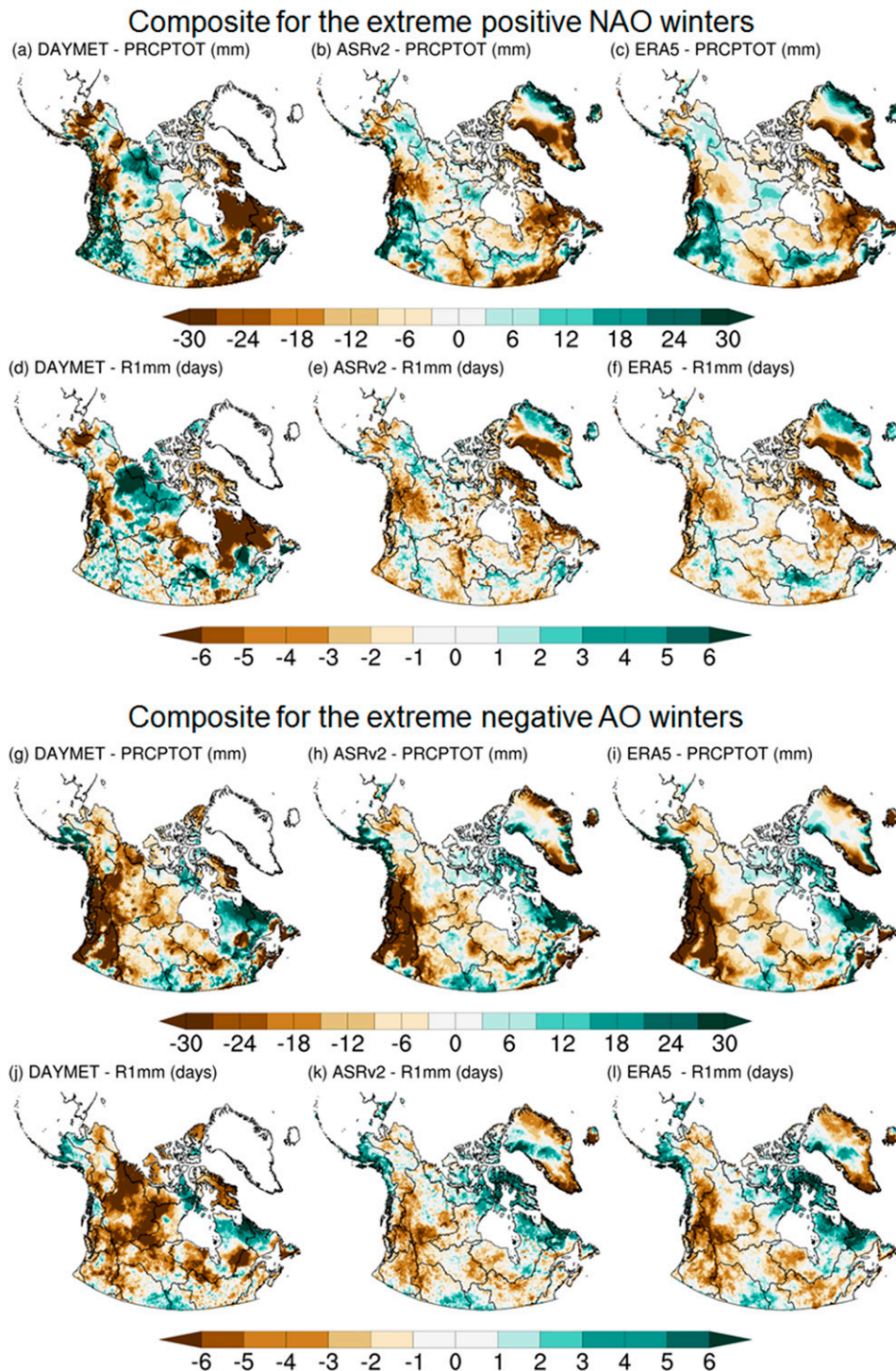
Composite for the extreme positive NAO winters

(a) DAYMET - PRCPTOT (mm)   (b) ASRv2 - PRCPTOT (mm)   (c) ERA5 - PRCPTOT (mm)

-30 -24 -18 -12 -6 0 6 12 18 24 30

(d) DAYMET - R1mm (days)   (e) ASRv2 - R1mm (days)   (f) ERA5 - R1mm (days)

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Composite for the extreme negative AO winters

(g) DAYMET - PRCPTOT (mm)   (h) ASRv2 - PRCPTOT (mm)   (i) ERA5 - PRCPTOT (mm)

-30 -24 -18 -12 -6 0 6 12 18 24 30

(j) DAYMET - R1mm (days)   (k) ASRv2 - R1mm (days)   (l) ERA5 - R1mm (days)

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

FIG. 14. Composite anomalies for PRCPTOT and number of wet days (R1mm) during extremes winters (DJFM) from 1999/2000 to 2015/16 (no data for December 1999) in (a)–(f) the positive phase of the NAO and (g)–(l) the negative phase of the AO.

(basin 13), and Atlantic watersheds, TNn decreases by up to 2.5°C during positive NAO winters (Figs. 12d–f).

Insofar as temperatures are concerned, differences are induced by each mode of climate variability (Figs. 12 and 13). For

instance, positive NAO increases TXx and TNn in the western basins (basins 2–8 and 10–13), while the negative AO induces the opposite, including anomalously cold patterns of TXx over the same watersheds and especially over the central and

southern basins. Like the NAO, a negative AO pattern reflects a weakened westerly flow across the Northern Hemisphere Arctic, allowing colder air to penetrate deeper into the lower latitudes. Interestingly, however, a negative AO also delivers anomalously positive TNn temperatures over much of the domain, matching the same response in TNn to positive NAO. The reduction in the extreme minimum temperature (positive TNn anomalies) is similar to one presented by Wettstein and Mearns (2002), who describe larger diurnal temperature difference under positive AO conditions. Our results indicate smaller diurnal temperatures differences over much of the western and southern stretches of the domain for the AO phase (Figs. 12g–l). We have long known about the regional impacts of the AO on climate variability (Thompson and Wallace 1998, 2000a,b), but a recent modeling study points to changes in the Northern Hemisphere storm tracks during winter under Arctic amplification (Wang et al. 2017). Their results suggest a weakening of the North Atlantic storm track leads to anomalous equatorward moisture flux that enhances the warming upstream due to changes in downward infrared radiation. Increased moisture is likely to lead to warmer overnight low temperatures, consistent with the TNn results shown in Fig. 12.

Turning to changes in precipitation, the positive phase of the NAO is associated with increased precipitation (PRCPTOT; Figs. 13e and 14a–c), with positive anomalies over the Fraser, Columbia, northwestern Lawrence, northeastern Greenland, and Iceland basins (basins 5, 6, 13, 15, and 16). However, there is reduced precipitation over the Great Basin (basin 7), eastern Hudson (basin 9), eastern Lawrence, Atlantic (basin 14), and western Greenland watersheds (Figs. 13e and 14a–c). Conversely, NAO positive events are not shown to have a strong influence on R1mm (Figs. 13g and 14d–f). This indicates no significant changes in the number of wet days during extremes winters, except for the western part of Greenland, which shows a reduction in precipitation both in intensity and frequency. These reductions in total winter precipitation over the eastern domain are consistent with a strengthened Icelandic low under positive NAO conditions, one that induces stronger northerly winds and reduced northward moisture flux into this region.

The negative phase of AO delivers similar mean conditions in all watersheds across DAYMET, ASRv2, and ERA5 (Figs. 13f,h and 14g–l). Still, dry conditions are noted in the western domain (e.g., watersheds 4–7) and and the east coast of Greenland and in Iceland (watersheds 15 and 16). Interestingly, wetter conditions in PRCPTOT and R1mm are presented during negative AO events over northeastern parts of the Hudson (basin 9) and Atlantic (basin 14), and in the central region of Greenland, with homogeneous spatial variability indicated by Figs. 13f and 13h.

In general, compared to observed gridded data (DAYMET), both ASRv2 and ERA5 show similar precipitation patterns. However, extreme precipitation indices for the northern domain of the Yukon basin (basin 4) and northwestern part of the Arctic basin (basin 1) show larger positive (negative) anomalies in the NAO (AO) phase for DAYMET compared to ASRv2 and ERA5 (Figs. 14 and 13). These differences may be associated with the scarcity of stations or the regionalization of precipitation within DAYMET.

4. Discussion and conclusions

We have compared the performance of the two regional (ASRv2 and NARR) and three global reanalyses (ERA5, MERRA-2, and GMFD) with DAYMET (a gridded observationally based dataset) in reproducing extreme climate indices of temperature and precipitation over North America including the Arctic. The comparison was performed at a 0.25° horizontal spatial resolution during the 17-yr ASRv2 period (2000–16) using the Kling–Gupta efficiency, which is a combination of the three components, namely correlation, bias ratio, and variability.

Observations and reanalyses show a consistently warm pattern with a decrease in frequency and intensity of cold extremes over most of the study area. Cold days, cold nights, frost days, and ice days have decreased during the last two decades. These changes in cold extremes are linked to changes in wintertime atmospheric circulation and synoptic conditions (e.g., Wu 2017; Papritz 2020). Furthermore, the hottest day, warm days, and warm nights have increased over the northern and western parts of the domain, especially across the Arctic, Pacific, Yukon, Mackenzie, Fraser, and Columbia watersheds (basins 1–6 in Fig. 1). Not only do warmer days increase the impact on sea ice loss (Bliss et al. 2019; Brennan et al. 2020; Peng et al. 2020), with a positive feedback for further environmental changes in the Arctic, but they also increase the likelihood and intensity of wildfires across the Arctic landscape (Wang et al. 2020). These changes in climate extremes have significant impacts on Arctic residents (e.g., Vogel and Bullock 2020).

The implication of the increasing warmth is reflected in the increase of daily intensity precipitation over the domain as water vapor provides an important passive link between a warming yet wetter world (Bengtsson 2010). Furthermore, the increase in intensity indices is more consistent in RX1day, RX5day, and R95p, which may potentially produce locally adverse effects (e.g., storms and floods), especially in the Columbia, Great Basin, Colorado, Nelson, and Missouri watersheds (basins 6, 7, 8, 10, and 11). Also, temperature extremes over Greenland have been linked to the occurrence of clouds as well (Gallagher et al. 2020), whose variability is directly linked to moisture availability.

The overall comparison reveals that reanalyses demonstrate better skill in reproducing temperature than precipitation climate indices. This is especially true for precipitation magnitude (PRCPTOT), intensity (RX1day and RX5day), frequency (R1mm), and duration (CDD and CWD) over northern parts of the domain including the Arctic and Hudson basins (basins 1 and 9). Evaluating these regions is still a challenge because of uncertainties associated with lower gauge coverage and measurement inconsistency.

ASRv2 and ERA5 performed better than NARR, MERRA-2, and GMFD compared to DAYMET. However, the best choice in capturing the spatiotemporal extreme precipitation patterns for the Colorado, Missouri, and Mississippi basins (basins 8, 11,
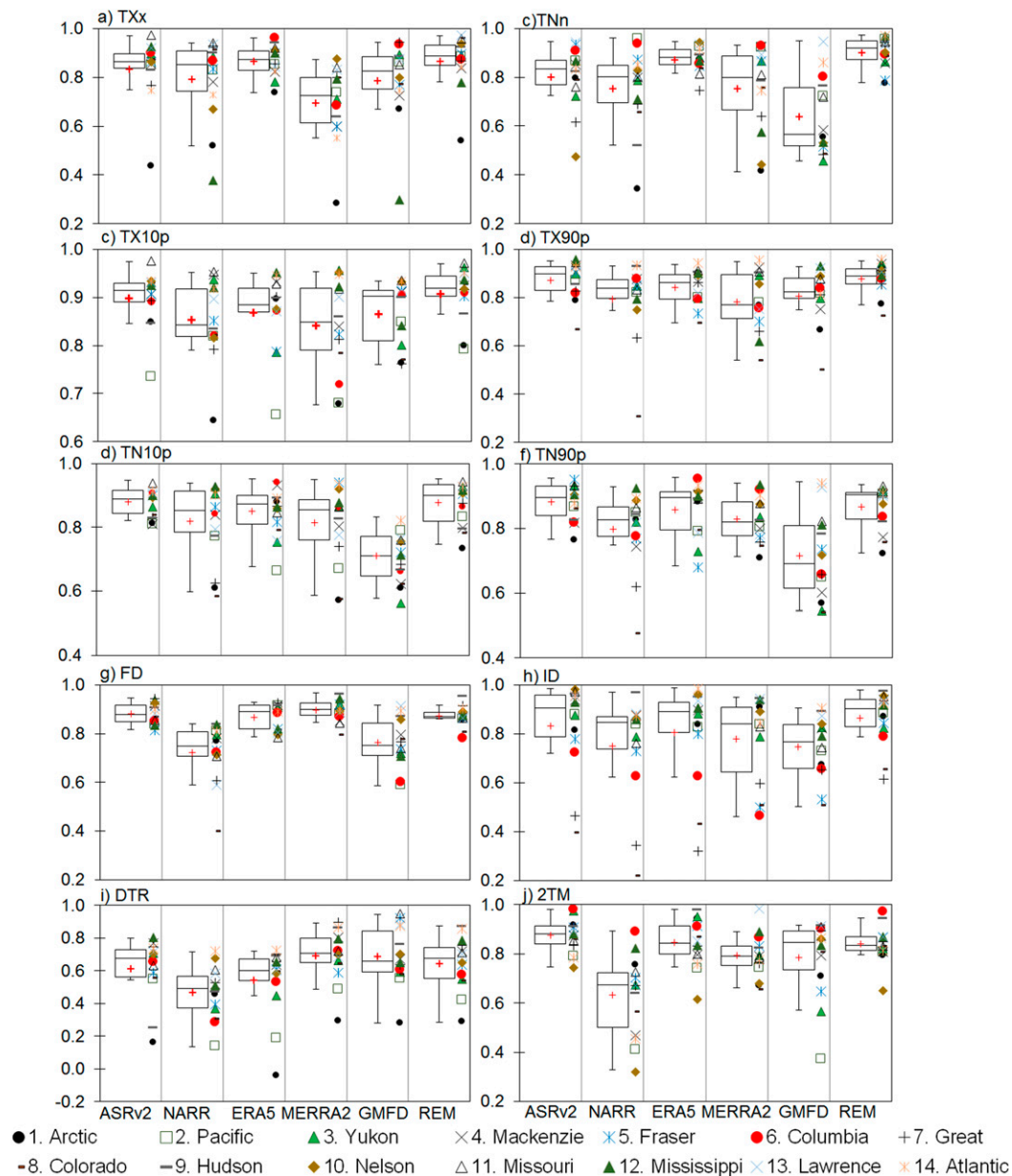
FIG. A1. KGE values for reanalyses (ASRv2, NARR, ERA5, MERRA-2, and GMFD) and the ensemble mean of reanalyses (REM) in climate temperature indices. The boxplot displays the 25th, 50th (median), and 75th percentiles for each dataset as well as the mean values (plus signs). The symbols next to each boxplot represent the KGE coefficients for each basin (Fig. 1).

and 12) is the NARR and for the Fraser basin (basin 5) is MERRA-2. In most cases, ASRv2 and ERA5 appear to be accurate representations for North America and the Arctic. Both are new reanalyses with high spatial (ASRv2 with 15 km and ERA5 with ~0.25°) and temporal (ASRv2 with 3-hourly and ERA5 with 1-hourly) resolutions. Despite these results, future studies should further investigate the impacts of horizontal resolution of the reanalyses on the estimation of climate

extremes, principally in the middle-to-high latitudes and basins with a complex topography (e.g., the Pacific watershed).

Using ASRv2 and ERA5, it has been found that the North Atlantic (NAO) and the Arctic Oscillation (AO) exert distinct influences on extreme climate indices. The evaluation of both the NAO and AO is important because the NAO impacts more strongly the eastern part of the North American pan-Arctic. In contrast, the AO displays a
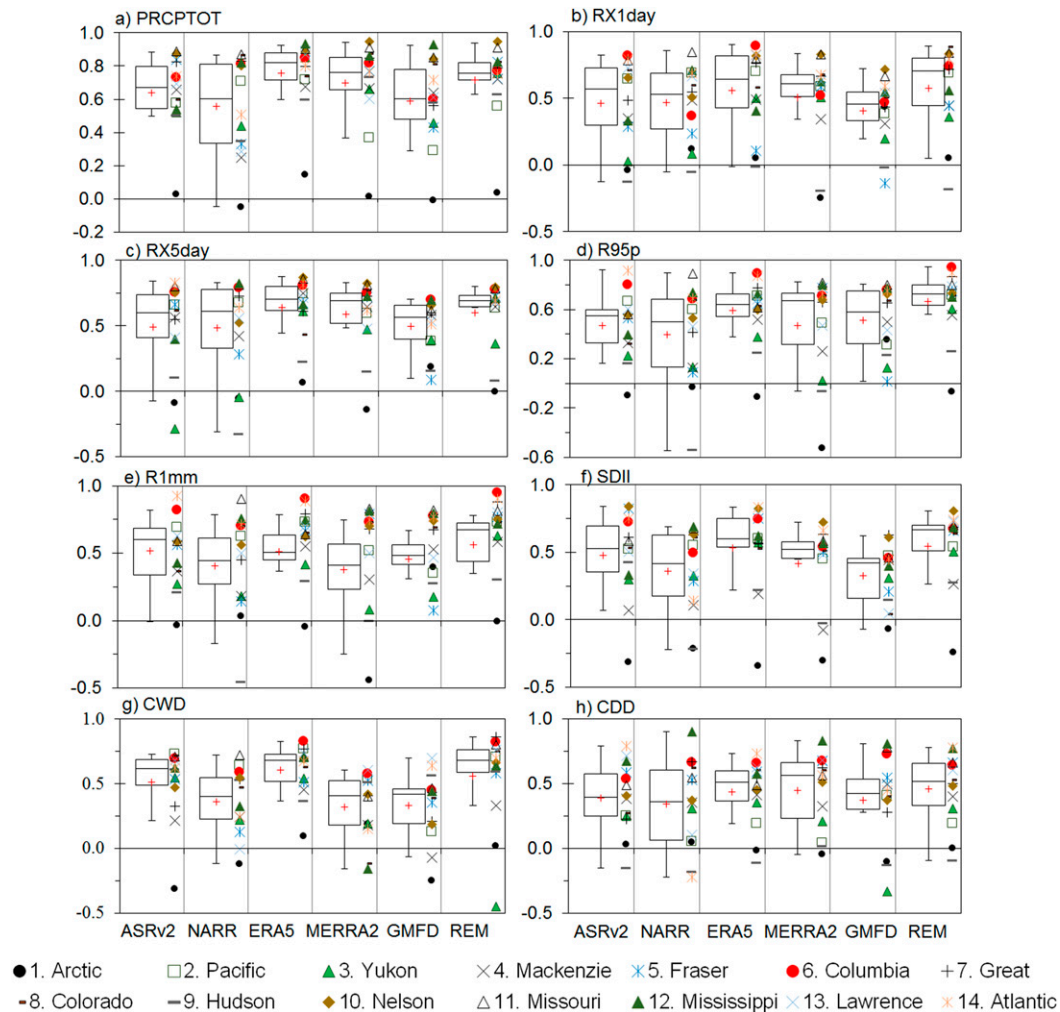
FIG. A2. As in Fig. A1, but for climate precipitation indices.

In short, results here provide a guide as to the relative performance of regional/global reanalyses for the community of climate modelers and observational scientists to assess the robustness of reanalysis datasets in the North American Arctic. This study has also synthesized the recent amplified warming across high-latitude North America, helping to quantify changes in extremes so that they may be linked to the impacts experienced by the people and environment of this region.

APPENDIX

KGE and Index Values

Figures A1 and A2 give the KGE values for the temperature and precipitation indices, respectively. Boxplots are given for each of the 14 watershed basins. Figure A3 provides the
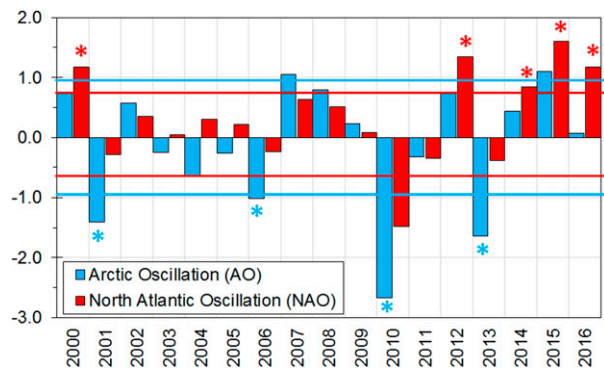
FIG. A3. The NAO and AO indices from 1999/2000 to 2015/16 winters. The lines indicate ±1 standard deviation from the long-term mean of the NAO index (0.71) and AO (0.96) from 1950 to 2016; an asterisk indicates the winters used in the composite analysis.

North Atlantic Oscillation and Arctic Oscillation indices for the winters from 1999/2000 to 2015/16.

REFERENCES

Acharya, S. C., R. Nathan, Q. J. Wang, C.-H. Su, and N. Eizenberg, 2019: An evaluation of daily precipitation from a regional atmospheric reanalysis over Australia. *Hydrol. Earth Syst. Sci.*, **23**, 3387–3403, https://doi.org/10.5194/hess-23-3387-2019.

Aguilar, E., and Coauthors, 2005: Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *J. Geophys. Res.*, **110**, D23107, https://doi.org/10.1029/2005JD006119.

Ávila, A., F. Justino, A. Wilson, D. Bromwich, and M. Amorim, 2016: Recent precipitation trends, flash floods and landslides in southern Brazil. *Environ. Res. Lett.*, **11**, 114029, https://doi.org/10.1088/1748-9326/11/11/114029.

——, F. Guerrero, Y. Escobar, and F. Justino, 2019: Recent precipitation trends and floods in the Colombian Andes. *Water*, **11**, 379, https://doi.org/10.3390/w11020379.

Avila-Diaz, A., G. Abrahão, F. Justino, R. Torres, and A. Wilson, 2020: Extreme climate indices in Brazil: Evaluation of downscaled Earth system models at high horizontal resolution. *Climate Dyn.*, **54**, 5065–5088, https://doi.org/10.1007/s00382-020-05272-9.

Bador, M., and Coauthors, 2020: Impact of higher spatial atmospheric resolution on precipitation extremes over land in global climate models. *J. Geophys. Res. Atmos.*, **125**, e2019JD032184, https://doi.org/10.1029/2019JD032184.

Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **53**, 6835–6875, https://doi.org/10.1007/s00382-019-04958-z.

Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

——, E. Wood, M. Pan, C. Fisher, D. Miralles, A. van Dijk, T. McVicar, and R. Adler, 2019a: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.

——, and Coauthors, 2019b: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, https://doi.org/10.5194/hess-23-207-2019.

Bengtsson, L., 2010: The global atmospheric water cycle. *Environ. Res. Lett.*, **5**, 025202, https://doi.org/10.1088/1748-9326/5/2/025202.

Betts, A., M. Köhler, and Y. Zhang, 2009: Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *J. Geophys. Res.*, **114**, D02101, https://doi.org/10.1029/2008JD010761.

Bhuiyan, M., E. Nikolopoulos, and E. Anagnostou, 2019: Machine learning-based blending of satellite and reanalysis precipitation datasets: A multi-regional tropical complex terrain evaluation. *J. Hydrometeor.*, **20**, 2147–2161, https://doi.org/10.1175/JHM-D-19-0073.1.

Bliss, A. C., M. Steele, G. Peng, W. N. Meier, and S. Dickinson, 2019: Regional variability of Arctic sea ice seasonal change climate indicators from a passive microwave climate data record. *Environ. Res. Lett.*, **14**, 045003, https://doi.org/10.1088/1748-9326/aafb84.

Boisvert, L., and J. Stroeve, 2015: The Arctic is becoming warmer and wetter as revealed by the Atmospheric Infrared Sounder. *Geophys. Res. Lett.*, **42**, 4439–4446, https://doi.org/10.1002/2015GL063775.

Booth, E., J. Byrne, and D. Johnson, 2012: Climatic changes in western North America, 1950–2005. *Int. J. Climatol.*, **32**, 2283–2300, https://doi.org/10.1002/joc.3401.

Box, G. E. P., and D. A. Pierce, 1970: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Stat. Assoc.*, **65**, 1509–1526, https://doi.org/10.1080/01621459.1970.10481180.

Brennan, M. K., G. J. Hakim, and E. Blanchard-Wrigglesworth, 2020: Arctic sea-ice variability during the instrumental era. *Geophys. Res. Lett.*, **47**, e2019GL086843, https://doi.org/10.1029/2019GL086843.

Bromwich, D., A. Wilson, L. S. Bai, G. Moore, and P. Bauer, 2016: A comparison of the regional Arctic System Reanalysis and the global ERA-Interim reanalysis for the Arctic. *Quart. J. Roy. Meteor. Soc.*, **142**, 644–658, https://doi.org/10.1002/qj.2527.

——, and Coauthors, 2018: The Arctic System Reanalysis, version 2. *Bull. Amer. Meteor. Soc.*, **99**, 805–828, https://doi.org/10.1175/BAMS-D-16-0215.1.

Chaney, N., J. Sheffield, G. Villarini, and E. Wood, 2014: Development of a high-resolution gridded daily meteorological dataset over sub-Saharan Africa: Spatial analysis of trends in climate extremes. *J. Climate*, **27**, 5815–5835, https://doi.org/10.1175/JCLI-D-13-00423.1.

Cohen, J., 2016: An observational analysis: Tropical relative to Arctic influence on midlatitude weather in the era of Arctic amplification. *Geophys. Res. Lett.*, **43**, 5287–5294, https://doi.org/10.1002/2016GL069102.

——, J. Foster, M. Barlow, K. Saito, and J. Jones, 2010: Winter 2009–2010: A case study of an extreme Arctic Oscillation event. *Geophys. Res. Lett.*, **37**, L17707, https://doi.org/10.1029/2010GL044256.

Cornes, R., and P. Jones, 2013: How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *J. Geophys. Res. Atmos.*, **118**, 10 262–10 276, https://doi.org/10.1002/jgrd.50799.

Croitoru, A.-E., A. Piticar, and D. C. Burada, 2016: Changes in precipitation extremes in Romania. *Quat. Int.*, **415**, 325–335, https://doi.org/10.1016/j.quaint.2015.07.028.

Dai, P., and B. Tan, 2017: The nature of the Arctic Oscillation and diversity of the extreme surface weather anomalies it generates. *J. Climate*, **30**, 5563–5584, https://doi.org/10.1175/JCLI-D-16-0467.1.

Daly, C., M. Halbleib, J. Smith, W. Gibson, M. Doggett, G. Taylor, J. Curtis, and P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, https://doi.org/10.1002/joc.1688.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

——, M. Balmaseda, G. Balsamo, R. Engelen, A. Simmons, and J. Thépaut, 2014: Toward a consistent reanalysis of the climate system. *Bull. Amer. Meteor. Soc.*, **95**, 1235–1248, https://doi.org/10.1175/BAMS-D-13-00043.1.

Diaconescu, E. P., A. Mailhot, R. Brown, and D. Chaumont, 2017: Evaluation of CORDEX-Arctic daily precipitation and temperature-based climate indices over Canadian Arctic land areas. *Climate Dyn.*, **50**, 2061–2085, https://doi.org/10.1007/s00382-017-3736-4.

Donat, M., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, https://doi.org/10.1002/jgrd.50150.

——, J. Sillmann, S. Wild, L. Alexander, T. Lippmann, and F. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, https://doi.org/10.1175/JCLI-D-13-00405.1.

——, L. Alexander, N. Herold, and A. Dittus, 2016: Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. *J. Geophys. Res. Atmos.*, **121**, 11 174–11 189, https://doi.org/10.1002/2016JD025480.

El Kenawy, A., J. I. López-Moreno, and S. M. Vicente-Serrano, 2011: Recent trends in daily temperature extremes over northeastern Spain (1960–2006). *Nat. Hazards Earth Syst. Sci.*, **11**, 2583–2603, https://doi.org/10.5194/nhess-11-2583-2011.

Fujiwara, M., and Coauthors, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017.

Gallagher, M. R., H. Chepfer, M. D. Shupe, and R. Guzman, 2020: Warm temperature extremes across Greenland connected to clouds. *Geophys. Res. Lett.*, **47**, e2019GL086059, https://doi.org/10.1029/2019GL086059.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Grotjahn, R., and Coauthors, 2016: North American extreme temperature events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **46**, 1151–1184, https://doi.org/10.1007/s00382-015-2638-6.

Gupta, H., S. Sorooshian, and P. Yapo, 1999: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.*, **4**, 135–143, https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135).

Hamed, K. H., and A. Rao, 1998: A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.*, **204**, 182–196, https://doi.org/10.1016/S0022-1694(97)00125-X.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoffmann, L., and Coauthors, 2019: From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos. Chem. Phys.*, **19**, 3097–3124, https://doi.org/10.5194/acp-19-3097-2019.

Hu, L., and G. Huang, 2020: The changes of high-temperature extremes and their links with atmospheric circulation over the Northern Hemisphere. *Theor. Appl. Climatol.*, **139**, 261–274, https://doi.org/10.1007/s00704-019-02970-1.

Hurrell, J. W., and C. Deser, 2009: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **78**, 28–41, https://doi.org/10.1016/j.jmarsys.2008.11.026.

——, Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, Geophys. Monogr., Vol. 134, Amer. Geophys. Union, 1–35.

Justino, F., A. B. Wilson, D. H. Bromwich, A. Avila, L.-S. Bai, and S.-H. Wang, 2019: Northern Hemisphere extratropical turbulent heat fluxes in ASRv2 and global reanalyses. *J. Climate*, **32**, 2145–2166, https://doi.org/10.1175/JCLI-D-18-0535.1.

Kendall, M., 1975: *Rank Correlation Methods*. 4th ed. Charles Griffin, 202 pp.

Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424-425**, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011.

Koenigk, T., P. Berg, and R. Döscher, 2015: Arctic climate change in an ensemble of regional CORDEX simulations. *Polar Res.*, **34**, 24603, https://doi.org/10.3402/polar.v34.24603.

Koyama, T., and J. Stroeve, 2019: Greenland monthly precipitation analysis from the Arctic System Reanalysis (ASR): 2000–2012. *Polar Sci.*, **19**, 1–12, https://doi.org/10.1016/j.polar.2018.09.001.

Kunkel, K., and J. Angel, 1999: Relationship of ENSO to snowfall and related cyclone activity in the contiguous United States. *J. Geophys. Res.*, **104**, 19 425–19 434, https://doi.org/10.1029/1999JD900010.

Lader, R., U. S. Bhatt, J. E. Walsh, T. S. Rupp, and P. A. Bieniek, 2016: Two-meter temperature and precipitation from atmospheric reanalysis evaluated for Alaska. *J. Appl. Meteor. Climatol.*, **55**, 901–922, https://doi.org/10.1175/JAMC-D-15-0162.1.

——, J. Walsh, U. Bhatt, and P. Bieniek, 2017: Projections of twenty-first-century climate extremes for Alaska via dynamical downscaling and quantile mapping. *J. Appl. Meteor. Climatol.*, **56**, 2393–2409, https://doi.org/10.1175/JAMC-D-16-0415.1.

Li, X., X. Wang, and V. Babovic, 2018: Analysis of variability and trends of precipitation extremes in Singapore during 1980–2013. *Int. J. Climatol.*, **38**, 125–141, https://doi.org/10.1002/joc.5165.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1.

Luo, B., D. Luo, A. Dai, and L. Wu, 2020: Combined influences on North American winter air temperature variability from North Pacific blocking and the North Atlantic Oscillation: Sub-seasonal and interannual timescales. *J. Climate*, **33**, 7101–7123, https://doi.org/10.1175/JCLI-D-19-0327.1.

Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13**, 245, https://doi.org/10.2307/1907187.

Matthes, H., A. Rinke, and K. Dethloff, 2016: Recent changes in Arctic temperature extremes: Warm and cold spells during winter and summer. *Environ. Res. Lett.*, **11**, 029501, https://doi.org/10.1088/1748-9326/11/2/029501.

McEvoy, D. J., J. F. Mejia, and J. L. Huntington, 2014: Use of an observation network in the Great Basin to evaluate gridded climate data. *J. Hydrometeor.*, **15**, 1913–1931, https://doi.org/10.1175/JHM-D-14-0015.1.

Melvin, A., J. Murray, B. Boehlert, J. Martinich, L. Rennels, and T. Rupp, 2017: Estimating wildfire response costs in Alaska's changing climate. *Climatic Change*, **141**, 783–795, https://doi.org/10.1007/s10584-017-1923-2.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Moore, G. W. K., I. A. Renfrew, B. E. Harden, and S. H. Mernild, 2015: The impact of resolution on the representation of southeast Greenland barrier winds and katabatic flows. *Geophys. Res. Lett.*, **42**, 3011–3018, https://doi.org/10.1002/2015GL063550.

Moriasi, D., J. Arnold, M. Van Liew, R. Binger, R. Harmel, and T. Veith, 2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, **50**, 885–900, https://doi.org/10.13031/2013.23153.

Murdock, T., S. Sobie, F. Zwiers, and H. Eckstrand, 2013: Climate change and extremes in the Canadian Columbia basin. *Atmos.–Ocean*, **51**, 456–469, https://doi.org/10.1080/07055900.2013.816932.

NCAR, 2017: Arctic System Reanalysis version 2. National Center for Atmospheric Research Computational and Information Systems Laboratory, accessed 5 June 2019, https://doi.org/10.5065/D6X9291B.

Ning, L., and R. S. Bradley, 2015: Winter climate extremes over the northeastern United States and southeastern Canada and teleconnections with large-scale modes of climate variability. *J. Climate*, **28**, 2475–2493, https://doi.org/10.1175/JCLI-D-13-00750.1.

Overland, J., and M. Wang, 2018: Arctic–midlatitude weather linkages in North America. *Polar Sci.*, **16**, 1–9, https://doi.org/10.1016/j.polar.2018.02.001.

Papritz, L., 2020: Arctic lower-tropospheric warm and cold extremes: Horizontal and vertical transport, diabatic processes, and linkage to synoptic circulation features. *J. Climate*, **33**, 993–1016, https://doi.org/10.1175/JCLI-D-19-0638.1.

Peng, H. T., C. Q. Ke, X. Shen, M. Li, and Z. De Shao, 2020: Summer albedo variations in the Arctic sea ice region from 1982 to 2015. *Int. J. Climatol.*, **40**, 3008–3020, https://doi.org/10.1002/joc.6379.

Przybylak, R., and P. Wyszyński, 2020: Air temperature changes in the Arctic in the period 1951–2015 in the light of observational and reanalysis data. *Theor. Appl. Climatol.*, **139**, 75–94, https://doi.org/10.1007/s00704-019-02952-3.

Qu, M., J. Wan, and X. Hao, 2014: Analysis of diurnal air temperature range change in the continental United States. *Wea. Climate Extremes*, **4**, 86–95, https://doi.org/10.1016/j.wace.2014.05.002.

Rapaić, M., R. Brown, M. Markovic, and D. Chaumont, 2015: An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. *Atmos.–Ocean*, **53**, 283–303, https://doi.org/10.1080/07055900.2015.1045825.

Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, https://doi.org/10.1175/JCLI-D-16-0570.1.

Rivière, G., and M. Drouard, 2015: Understanding the contrasting North Atlantic oscillation anomalies of the winters of 2010 and 2014. *Geophys. Res. Lett.*, **42**, 6868–6875, https://doi.org/10.1002/2015GL065493.

Roberts, M. J., and Coauthors, 2018: The benefits of global high resolution for climate simulation process understanding and the enabling of stakeholder decisions at the regional scale. *Bull. Amer. Meteor. Soc.*, **99**, 2341–2359, https://doi.org/10.1175/BAMS-D-15-00320.1.

Schoof, J., and S. Robeson, 2016: Projecting changes in regional temperature and precipitation extremes in the United States. *Wea. Climate Extremes*, **11**, 28–40, https://doi.org/10.1016/j.wace.2015.09.004.

Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, https://doi.org/10.1175/JCLI3790.1.

Shepherd, T. G., 2016: Effects of a warming Arctic. *Science*, **353**, 989–990, https://doi.org/10.1126/science.aag2349.

Sheridan, S., and C. Lee, 2018: Temporal trends in absolute and relative extreme temperature events across North America. *J. Geophys. Res. Atmos.*, **123**, 11 889–11 898, https://doi.org/10.1029/2018JD029150.

Simmons, A. J., and P. Poli, 2015: Arctic warming in ERA-Interim and other analyses. *Quart. J. Roy. Meteor. Soc.*, **141**, 1147–1162, https://doi.org/10.1002/qj.2422.

Siqueira, V. A., and Coauthors, 2018: Toward continental hydrologic–hydrodynamic modeling in South America. *Hydrol. Earth Syst. Sci.*, **22**, 4815–4842, https://doi.org/10.5194/hess-22-4815-2018.

Skansi, M., and Coauthors, 2013: Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global Planet. Change*, **100**, 295–307, https://doi.org/10.1016/j.gloplacha.2012.11.004.

Tarek, M., F. P. Brissette, and R. Arsenault, 2020: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.*, **24**, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300, https://doi.org/10.1029/98GL00950.

——, and ——, 2000a: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016, https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

——, and ——, 2000b: Annular modes in the extratropical circulation. Part II: Trends. *J. Climate*, **13**, 1018–1036, https://doi.org/10.1175/1520-0442(2000)013<1018:AMITEC>2.0.CO;2.

Thornton, P., S. Running, and M. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, https://doi.org/10.1016/S0022-1694(96)03128-9.

——, M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook, 2018: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, version 3. Oak Ridge National Laboratory Distributed Active Archive Center, accessed 25 June 2019, https://doi.org/10.3334/ORNLDAAC/1392.

Timmermans, B., M. Wehner, D. Cooley, T. O'Brien, and H. Krishnan, 2019: An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dyn.*, **52**, 6651–6670, https://doi.org/10.1007/s00382-018-4537-0.

Uotila, P., and Coauthors, 2019: An assessment of ten ocean reanalyses in the polar regions. *Climate Dyn.*, **52**, 1613–1650, https://doi.org/10.1007/s00382-018-4242-z.

Vogel, B., and R. C. L. Bullock, 2020: Institutions, indigenous peoples, and climate change adaptation in the Canadian Arctic. *GeoJournal*, https://doi.org/10.1007/s10708-020-10212-5, in press.

Walsh, J., P. Bieniekb, B. Brettschneidera, E. Euskirchenc, R. Lader, and R. Thomand, 2017: The exceptionally warm winter of 2015/16 in Alaska. *J. Climate*, **30**, 2069–2088, https://doi.org/10.1175/JCLI-D-16-0473.1.

Walther, B. A., and J. L. Moore, 2005: The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**, 815–829, https://doi.org/10.1111/j.2005.0906-7590.04112.x.

Wang, C., R. M. Graham, K. Wang, S. Gerland, and M. A. Granskog, 2019: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and evolution. *Cryosphere*, **13**, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019.

Wang, J., H. M. Kim, and E. K. M. Chang, 2017: Changes in Northern Hemisphere winter storm tracks under the background of Arctic amplification. *J. Climate*, **30**, 3705–3724, https://doi.org/10.1175/JCLI-D-16-0650.1.

Wang, J. A., D. Sulla-Menashe, C. E. Woodcock, O. Sonnentag, R. F. Keeling, and M. A. Friedl, 2020: Extensive land cover change across Arctic–boreal northwestern North America from disturbance and climate forcing. *Global Change Biol.*, **26**, 807–822, https://doi.org/10.1111/gcb.14804.

Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, and E. Xoplaki, 2001: North Atlantic oscillation—Concepts and studies. *Surv. Geophys.*, **22**, 321–381, https://doi.org/10.1023/A:1014217317898.

Wazneh, H., M. A. Arain, and P. Coulibaly, 2020: Climate indices to characterize climatic changes across southern Canada. *Meteor. Appl.*, **27**, 1–19, https://doi.org/10.1002/met.1861.

Wettstein, J. J., and L. O. Mearns, 2002: The influence of the North Atlantic–Arctic Oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada. *J. Climate*, **15**, 3586–3600, https://doi.org/10.1175/1520-0442(2002)015<3586:TIOTNA>2.0.CO;2.

Wilson, A. B., D. H. Bromwich, and K. M. Hines, 2011: Evaluation of polar WRF forecasts on the Arctic System Reanalysis domain: Surface and upper air analysis. *J. Geophys. Res.*, **116**, D11112, https://doi.org/10.1029/2010JD015013.

Wong, J. S., S. Razavi, B. R. Bonsal, H. S. Wheater, and Z. E. Asong, 2017: Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrol. Earth Syst. Sci.*, **21**, 2163–2185, https://doi.org/10.5194/hess-21-2163-2017.

Wu, B., 2017: Winter atmospheric circulation anomaly associated with recent Arctic winter warm anomalies. *J. Climate*, **30**, 8469–8479, https://doi.org/10.1175/JCLI-D-17-0175.1.

Yang, Y., T. Y. Gan, and X. Tan, 2019: Spatiotemporal changes in precipitation extremes over Canada and their teleconnections to large-scale climate patterns. *J. Hydrometeor.*, **20**, 275–296, https://doi.org/10.1175/JHM-D-18-0004.1.

Yin, H., M. G. Donat, L. V. Alexander, and Y. Sun, 2015: Multidataset comparison of gridded observed temperature and precipitation extremes over China. *Int. J. Climatol.*, **35**, 2809–2827, https://doi.org/10.1002/joc.4174.

Young, A., P. Higuera, J. Abatzoglou, P. Duffy, and F. Hu, 2019: Consequences of climatic thresholds for projecting fire activity and ecological change. *Global Ecol. Biogeogr.*, **28**, 521–532, https://doi.org/10.1111/geb.12872.

Yue, S., P. Pilon, and G. Cavadias, 2002: Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *J. Hydrol.*, **259**, 254–271, https://doi.org/10.1016/S0022-1694(01)00594-7.

Zhang, X., L. A. Vincent, W. D. Hogg, and A. Niitsoo, 2000: Temperature and precipitation trends in Canada during the 20th century. *Atmos.–Ocean*, **38**, 395–429, https://doi.org/10.1080/07055900.2000.9649654.

——, L. Alexander, G. Hegerl, P. Jones, A. Klein Tank, T. Peterson, B. Trewin, and F. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, https://doi.org/10.1002/wcc.147.