

Evaluation of CORDEX-Arctic daily precipitation and temperature-based climate indices over Canadian Arctic land areas

Emilia Paula Diaconescu¹  · Alain Mailhot¹ · Ross Brown^{2,3} · Diane Chaumont²

Received: 4 November 2016 / Accepted: 15 May 2017 / Published online: 22 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract This study focuses on the evaluation of daily precipitation and temperature climate indices and extremes simulated by an ensemble of 12 Regional Climate Model (RCM) simulations from the ARCTIC-CORDEX experiment with surface observations in the Canadian Arctic from the Adjusted Historical Canadian Climate Dataset. Five global reanalyses products (ERA-Interim, JRA55, MERRA, CFSR and GMFD) are also included in the evaluation to assess their potential for RCM evaluation in data sparse regions. The study evaluated the means and annual anomaly distributions of indices over the 1980–2004 dataset overlap period. The results showed that RCM and reanalysis performance varied with the climate variables being evaluated. Most RCMs and reanalyses were able to simulate well climate indices related to mean air temperature and hot extremes over most of the Canadian Arctic, with the exception of the Yukon region where models displayed the largest biases related to topographic effects. Overall performance was generally poor for indices related to cold extremes. Likewise, only a few RCM simulations and reanalyses were able to provide realistic simulations of precipitation extreme indicators. The multi-reanalysis ensemble provided superior results to individual datasets for climate indicators related to mean air temperature and hot extremes, but not for other indicators. These results support

the use of reanalyses as reference datasets for the evaluation of RCM mean air temperature and hot extremes over northern Canada, but not for cold extremes and precipitation indices.

Keywords Precipitation climate indices · Temperature climate indices · Regional climate model evaluation · CORDEX experiment · Daily precipitation · Canadian Arctic

1 Introduction

Trends in temperature and precipitation and their extremes in observations and Global Climate Models (GCMs) have been a subject of extensive study over the past decade because of the potential impacts on human society and ecosystems (e.g. Alexander et al. 2006; Kharin et al. 2007, 2013; Sillmann and Roeckner 2008; Donat et al. 2013; Sillmann et al. 2013a, b). There are a number of challenges in carrying out these studies (e.g. Zwiers et al. 2013; Alexander 2016) especially over data sparse regions such as the Canadian Arctic which is the focus of this study. In addition to limited surface observations, a key challenge in model evaluation is the scale mismatch (e.g. Booij 2002; Fowler et al. 2005; Zhang et al. 2011) between surface observations and climate models with resolutions ranging from ~25 to 50 km for the Regional Climate Models (RCMs) to ~100–300 km for the GCMs participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al. 2012). Generally, higher precipitation intensities are recorded at station (especially for more intense precipitation) than reported on gridded datasets since simulated gridded products are usually interpreted as mean values over the grid (for a discussion of this point see e.g. Chen

✉ Emilia Paula Diaconescu
Emilia_Paula.Diaconescu@ete.inrs.ca

¹ Institut national de la recherche scientifique, Eau Terre Environnement, 490 de la Couronne, Québec, QC G1K 9A9, Canada

² Ouranos, Montréal, QC, Canada

³ Environment and Climate Change Canada, Montréal, QC, Canada

and Knutson 2008). Consequences of scale mismatch are especially obvious for GCMs and reanalyses with coarse resolution. Different interpolation methods have been proposed in order to aggregate station information to the GCM scale. However, these methods can only be used with some confidence in regions with good spatial coverage. The Canadian Arctic (which includes the Yukon Territory, Northwest Territories and Nunavut) is a vast region covering approximately 3,600,000 km² of land. The number of stations reporting daily precipitation and daily temperature over extended periods of time in this region is very limited (see Sect. 2.3). Northern Canada is also a region of complex topography with a large number of lakes and 94 major islands. Interpolating information to common grids over this region of complex topography and ice/water boundaries can have a huge impact on the resulting fields, especially for extremes (e.g. Hofstra et al. 2010; Gervais et al. 2014). One approach for dealing with the scale-difference issue is to dynamically downscale GCM simulations using high-resolution RCM simulations. This has been shown to provide more realistic simulations of precipitation and precipitation extremes, with intensities and frequencies comparable to those recorded at surface stations (Chan et al. 2013).

The Arctic Coordinated Regional Downscaling Experiment (ARCTIC-CORDEX; Giorgi et al. 2009; Jones et al. 2011; Gutowski et al. 2016; <http://www.climate-cryosphere.org/activities/targeted/polar-cordex/arctic>) is part of the CORDEX initiative and there are three experimental streams consisting of: (1) RCM simulations driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim (ERA-I) reanalysis (Dee et al. 2011), (2) RCM historical simulations and (3) RCM projections driven by GCMs participating in the CMIP5 program. The RCM simulations obtained in the ARCTIC-CORDEX domain have not previously been evaluated over the Canadian Arctic, and previous RCM studies of Arctic climate have tended to focus on individual models (e.g. Saha et al. 2006; Matthes et al. 2010; Glisan and Gutowski 2014a, b; Koenigk et al. 2015). The spatial coverage of the multi-model ensemble of the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al. 2009) only partially covered the Canadian Arctic, which makes this analysis of the CORDEX multi-model ensemble novel.

The main motivation for this study was the need to provide decision-makers in northern Canada with information about the ability of current RCMs and reanalysis to simulate a range of commonly used temperature and precipitation indices. The need for this information is underscored by the rapid warming observed over the Arctic in the past several decades (Hansen et al. 2010; Vincent et al. 2015), which recent studies suggest may be underestimated (Way

et al. 2016; Cowtan and Way 2014). Climate extremes are expected to change more rapidly than mean warming (Fischer and Knutti 2015) with non-linear impacts that can pose challenges for adaptive capacity (Knutti et al. 2016). This study represents an important addition to previous regional climate based change projections provided for the Canadian Arctic (Allard and Lemay 2012; Stern and Gaden 2015) which did not include any analysis of climate extremes.

The aim of the present study is to evaluate the ability of CORDEX regional climate models to simulate key temperature and precipitation-based climate indices over the Canadian Arctic land areas. Fifteen temperature and ten precipitation climate and extremes indices were selected based on Arctic climate characteristics. Both the historical CMIP5 GCM- and ERAI-driven CORDEX simulations were evaluated. This allowed us to assess the RCM structural biases as well as the effect of GCM errors on the RCM simulations (Šeparović et al. 2013; Laprise et al. 2013). The evaluation was carried out by comparing simulated values to station records across the Canadian Arctic. In spite of their limitations (see Sect. 2.3), station observations remain the most reliable and the primary source of information for the historical climate of the region and are located where most human-related activities take place. This also avoids introducing potential errors into the evaluation from interpolation associated with gridded surface datasets, or from the numerous sources of errors associated with reanalyses. A second goal of the study was to evaluate how well recent reanalyses perform at reproducing the observed climate indices, and to determine if they can be used in model evaluation to complement station data in data-sparse regions.

The remainder of the paper is organized as follows: the RCMs, reanalyses and observed datasets used in the model evaluation are described in Sect. 2, while the evaluation methodology (i.e. computation of climate indices and evaluation metrics) is presented in Sect. 3. Section 4 presents the results of evaluating the reanalysis and GCM-driven simulations over a reference period of 25 years, while Sect. 5 presents the comparison of the GCM-driven and ERAI-driven simulations over a common period of 17 years. The final section (Sect. 6) summarises results and presents conclusions.

2 Dataset descriptions

2.1 RCMs

This study used four RCMs (AWI-HIRHAM5, CCCma-CanRCM4, SMHI-RCA4 and UQAM-CRCM5) that are part of the ARCTIC-CORDEX experiment (<http://www.>

climate-cryosphere.org/activities/targeted/polar-cordex/arctic). Table 1 provides information on the RCM simulations, their characteristics and relevant references for each model with the ARCTIC-CORDEX domain shown in Fig. 1. The simulations are driven by CMIP5 GCMs over a historical period, mainly from 1951 to 2005 and by ERAI over a more recent period, mainly from 1989 to 2008. Some models used spectral nudging while others did not. For SMHI-RCA4 both configurations, with (identified as SMHI-RCA4SN) and without (identified as SMHI-RCA4) spectral nudging, were considered. The CCCma-CanRCM4 simulations were provided on a rotated grid at two horizontal grid spacings, 0.44° and 0.22° , while the other model simulations were only available at the 0.44° (about 50 km) horizontal resolution. The CORDEX-ARCTIC ensemble was supplemented with three 0.44° horizontal-resolution simulations carried out with the UQAM-CRCM5 model over a North America domain, which completely covers Canada (Fig. 1), unlike the North America CORDEX simulations. These three simulations are identified as UQAM-CRCM5NA (NA for North America) to differentiate them from the runs carried out over the CORDEX-ARCTIC domain. The CCCma-CanRCM4 simulations at 0.22° are identified as CCCma-CanRCM4-022. In total this provided 12 GCM-driven simulations and six ERAI-driven simulations for analysis. The analysis was carried out over the Canadian Arctic land areas contained in the ARCTIC-CORDEX domain represented in Fig. 1 by the dark green shading.

2.2 Atmospheric reanalyses and gridded surface observations

The second goal of the study was to evaluate how well reanalyses and gridded surface observation datasets capture the observed local climate from station records to determine if they can be used in model evaluation to complement the observed station data. Six datasets were considered comprising four recent reanalyses (CFSR, ERAI, JRA55, MERRA), and one product that corrects the NCEP-R1 reanalysis with observed temperature, precipitation and radiation (GMFD). Dataset descriptions and relevant references are presented in Table 2. For the Arctic region and the period used in this study, the temperature observations used by GMFD are mainly from the CRU TS3.0 gridded ($0.5^\circ \times 0.5^\circ$) dataset, precipitation is corrected for undercatch and is downscaled based on relationships developed with the Global Precipitation Climatology Project daily product. Corrections were also made to high-latitude wintertime rain day statistics to remove a spurious wavelike pattern (Sheffield et al. 2006). The gridded surface observation dataset (NRCan) was included in the study to investigate the potential impacts of interpolating data from sparse surface observations on climate indices.

The NRCan dataset provides daily precipitation and temperature from Environment and climate change Canada stations gridded on a horizontal grid of ~ 10 km using the thin plate smoothing spline implemented in the ANUSPLIN climate modelling software (Hutchinson et al. 2009). A two-stage approach was applied for interpolating daily precipitation by estimating the spatial domain where precipitation occurred prior to carrying out the interpolation of observed precipitation amounts. Trace precipitation amounts for solid precipitation were assigned values from 0.03 to 0.07 mm that varied inversely with latitude following Mekis and Vincent (2011).

CFSR, MERRA and ERA-Interim monthly mean temperature and precipitation data were previously evaluated over the Canadian Arctic by Rapaić et al. (2015) and over the entire Arctic by Lindsay et al. (2014). Both papers indicate that MERRA and ERA-Interim have relatively small warm and wet biases compared to other reanalyses, while CFSR was found to have particularly large positive precipitation biases.

2.3 Surface observations

As previously mentioned, the number of stations with long-term daily temperature and precipitation records across the Canadian Arctic remains small. The interpolation errors in gridded products based on sparse station networks can be high, especially for extremes and climate indices (e.g. Hofstra et al. 2010; Contractor et al. 2015; Way et al. 2016). As a consequence, climate model evaluations over such regions are often carried out using station observations or reanalyses (e.g. Lindsay et al. 2014; Glisan and Gutowski 2014a, b; Matthes et al. 2010, 2015; Wilson et al. 2012). In this study, we use surface observations of mean, minimum and maximum daily temperature (Tmean, Tmin and Tmax, respectively) and daily precipitation (Pr) from the climate stations included in the National Climate Data and Information Archive at Environment and Climate Change Canada (<http://ccds-dscc.ec.gc.ca/index.php?page=download-obs>) that have at least 15 valid years in the 1980–2004 validation period of GCM-driven RCM simulations. Where stations were adjusted and corrected for known systematic errors, we used the corrected station data from the Adjusted Historical Canadian Climate Data set (AHCCD; Vincent et al. 2002, 2012; Mekis and Vincent 2011). The AHCCD takes account of systematic errors from changes in observing programs, instrumentation and station moves (the latter for temperature data only), and also includes station joining to produce longer records. Precipitation records in the AHCCD have undergone rigorous quality control and adjustments to account for known measurement issues such as wind-induced undercatch, evaporation

Table 1 Main characteristics of CORDEX and RCM simulations used in this study

Regional climate model (reference)	Simulation name	Lateral boundary conditions (reference)	Period	Main characteristics
AWI-HIRHAM5 (Christensen et al. 2007; Dethloff et al. 1996)	AWI-HIRHAM5-MPI-ESM-LR	MPI-M-MPI-ESM-LR (Giorgetta et al. 2013)	1949–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
CCCma-CanRCM4 (Scinocca et al. 2016)	CCCma-CanRCM4-CanESM2-022	CCCma-CanESM2 (Arora et al. 2011)	1950–2005	0.22° Resolution with spectral nudging; ARCTIC-CORDEX domain
	CCCma-CanRCM4-CanESM2	CCCma-CanESM2 (Arora et al. 2011)	1950–2005	0.44° Resolution with spectral nudging; ARCTIC-CORDEX domain
	CCCma-CanRCM4-ERA-Interim-022	ERA-Interim (Dee et al. 2011)	1989–2009	0.22° Resolution with spectral nudging; ARCTIC-CORDEX domain
	CCCma-CanRCM4-ERA-Interim	ERA-Interim (Dee et al. 2011)	1989–2009	0.44° Resolution with spectral nudging; ARCTIC-CORDEX domain
UQAM-CRCM5 (Takhsha et al. 2017)	UQAM-CRCM5-MPI-ESM-LR	MPI-M-MPI-ESM-LR (Giorgetta et al. 2013)	1949–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
	UQAM-CRCM5-ERA-Interim	ERA-Interim (Dee et al. 2011)	1979–2014	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
UQAM-CRCM5NA (Martynov et al. 2013; Šeparović et al. 2013)	UQAM-CRCM5NA-MPI-ESM-LR	MPI-M-MPI-ESM-LR (Giorgetta et al. 2013)	1949–2005	0.44° Resolution without spectral nudging; North America domain
	UQAM-CRCM5NA-CanESM2	CCCma-CanESM2 (Arora et al. 2011)	1949–2005	0.44° Resolution without spectral nudging; North America domain
	UQAM-CRCM5NA-ERA-Interim	ERA-Interim (Dee et al. 2011)	1979–2014	0.44° Resolution without spectral nudging; North America domain
SMHI-RCA4 (Samuelsson et al. 2011)	SMHI-RCA4-CanESM2	CCCma-CanESM2 (Arora et al. 2011)	1951–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4-NorESM1-M	NCC-NorESM1-M (Bentsen et al. 2013)	1951–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4-EC-EARTH	ICHEC-EC-EARTH (Hazeleger et al. 2012)	1951–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4-MPI-ESM-LR	MPI-M-MPI-ESM-LR (Giorgetta et al. 2013)	1951–2005	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4-ERA-Interim	ERA-Interim (Dee et al. 2011)	1980–2010	0.44° Resolution without spectral nudging; ARCTIC-CORDEX domain
SMHI-RCA4SN (Berg et al. 2013)	SMHI-RCA4SN-EC-EARTH	ICHEC-EC-EARTH (Hazeleger et al. 2012)	1951–2005	0.44° Resolution with spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4SN-MPI-ESM-LR	MPI-M-MPI-ESM-LR (Giorgetta et al. 2013)	1951–2005	0.44° Resolution with spectral nudging; ARCTIC-CORDEX domain
	SMHI-RCA4SN-ERA-Interim	ERA-Interim (Dee et al. 2011)	1980–2010	0.44° Resolution with spectral nudging; ARCTIC-CORDEX domain

Fig. 1 Spatial domains for the UQAM-CRCM5NA simulations (in *light green*) and ARCTIC-CORDEX simulations (in *cyan*). The Canadian Arctic analysis domain is shown in *deep green*

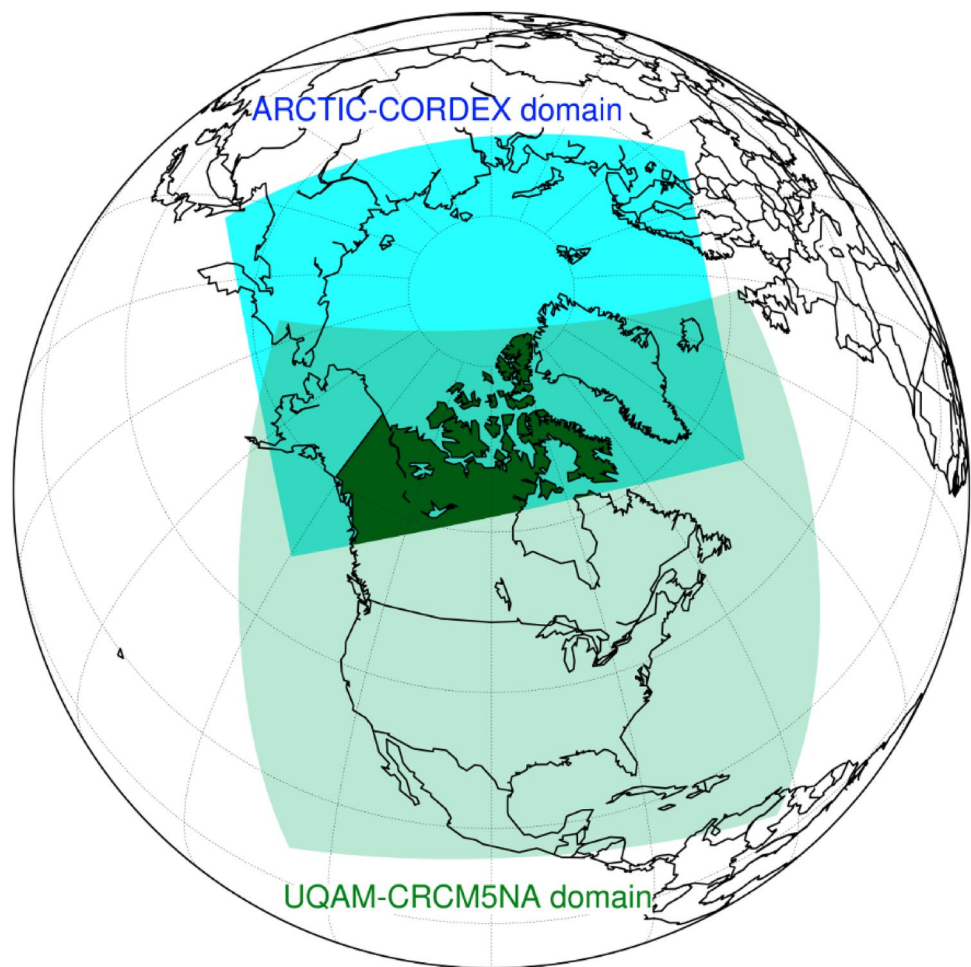


Table 2 Main characteristics of reanalyses and gridded observation datasets used in this study

Dataset (reference)	Download	Period	Type of dataset and spatial resolution
Global Meteorological Forcing Dataset for Land Surface Modeling (GMFD; Sheffield et al. 2006)	http://rda.ucar.edu/datasets/ds314.0/	1948–2008	Reanalysis, 0.25° (~28 km)
National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR; Saha et al. 2010)	http://rda.ucar.edu/datasets/ds093.1/ and http://cfs.ncep.noaa.gov/cfsr/	1979–2010	Reanalysis, T382 (~38 km)
Japanese 55-year Reanalysis (JRA55; Kobayashi et al. 2015)	http://rda.ucar.edu/datasets/ds628.0/	1958–2010	Reanalysis, TL319 (~60 km)
Modern Era Reanalysis for Research and Applications (MERRA; Rienecker et al. 2011)	http://gmao.gsfc.nasa.gov/merra/	1979–2010	Reanalysis, 0.667° longitude × 0.5° latitude (~65 km)
ECMWF ERA-Interim Reanalysis (ERA-Interim; Dee et al. 2011)	http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/	1979–2010	Reanalysis, T255 (~79 km)
Natural Resources Canada gridded surface data (NRCan; Hutchinson et al. 2009)	ftp://ftp.nrcan.gc.ca/pub/outgoing/canada_daily_grids	1950–2013	Gridded observations ~10 km

loss, and adjustments for trace observations that are particularly important for the frozen and light precipitation regimes that dominate the Canadian Arctic (Devine and Mekis 2008; Mekis 2005). The water equivalent of the

snowfall was adjusted based on climatological estimates of fresh snowfall density obtained from stations where both Nipher gauge (solid precipitation) and snowfall were measured (Mekis and Brown 2010). Trace snowfall

amounts were assigned values between 0.03 and 0.07 mm that varied inversely with latitude.

This resulted in a total of 47 stations for the air temperature indices evaluation and 78 stations for the precipitation indices evaluation, with the spatial distribution shown in Fig. 2. For the 1989–2005 period used in the comparison of the GCM-driven and ERAI-driven simulations, these conditions were relaxed to at least 10 valid years, which led to similar numbers of stations: 48 stations for temperature and 79 for precipitation. Details concerning the station selection criteria for climate indices computation are provided in Sect. 3.

It is important to note that none of the reanalyses considered in this study assimilated surface precipitation data and that CFSR, MERRA, and JRA55 do not assimilate surface air temperature. Surface air temperatures from Canadian synoptic stations are assimilated in ERA-Interim through an analysis based on the Optimal Interpolation procedure, while GMFD is using a post-processing procedure based on the CRU TS3.0 gridded ($0.5^\circ \times 0.5^\circ$) monthly dataset that contains Canadian climate station data. Consequently, surface air temperature fields from ERA-Interim and GMFD are not completely independent from the temperature field of the AHCCD dataset.

3 Methodology

The five reanalyses and the GCM- and ERAI-driven simulations cover different periods of time. In the first part of this study, the reanalyses and GCM-driven simulations were evaluated over a common reference period of 25 years (1980–2004) while, in the second part, GCM- and ERAI-driven RCM simulations were evaluated over a common period of 17 years (1989–2005).

3.1 Climate indices computation

The Expert Team on Climate Change Detection and Indices (ETCCDI; Klein Tank et al. 2009) proposed a large number of climate indices that characterize many aspects (not only extremes) of daily temperature and precipitation distributions relevant for climate change detection and climate-related applications. Also the ArcticNet Network of Centres of Excellence of Canada has identified several climate indices relevant for climate change impacts and adaptation studies in the Canadian Arctic (Allard and Lemay 2012; Stern and Gaden 2015). The reported indices were selected based on regional interests after due consultation of local representative involved in natural, health, social, and Inuit organizations, northern communities and federal and

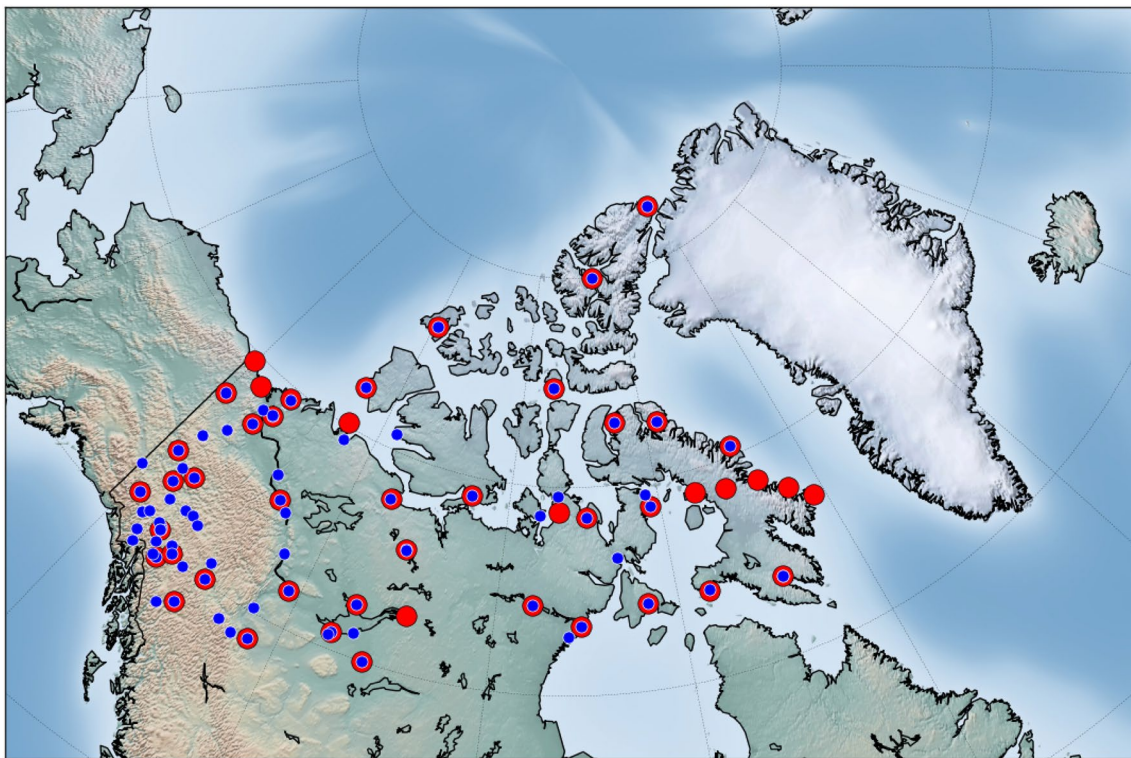


Fig. 2 Topography and location of stations with at least 15 valid years over the 1980–2004 period. The *blue points* show stations used in precipitation indices evaluation, while the *red points* show the location of stations used in temperature indices evaluation

provincial agencies. In our study, 15 daily temperature and ten daily precipitation climate indices associated with key characteristics of the precipitation and temperature regimes of the Arctic region were selected from those proposed by the ETCCDI and Arctic impacts-relevant indices provided by the Canadian ArcticNet program (Table 3).

It was however, necessary to adapt the nomenclature of certain ETCCDI indices to take account of Arctic conditions. A large portion of the analysis domain is located north of the Arctic Circle (66°32'N) that experiences polar days and nights with a less distinct diurnal cycle. Because the notion of night and day is different for the Arctic, the term “day” will be used to denote a calendar day of 24 h and not the period of sunlight, and the indices were named as in Table 3. The annual “cool” temperature indices were related only to the variable “daily minimum temperature”

(TNn, TN10 and TN10p), while the annual “warm” temperature indices are related only to the variable “daily maximum temperature” (TXx, TX90 and TX90p).

While total precipitation is generally low over much of the Canadian Arctic (the Arctic islands being known as Arctic desert) (Serreze and Barry 2005; Stern and Gaden 2015), increases in extreme precipitation events and precipitation intensity have been observed over the Arctic (Cohen et al. 2014; Ye et al. 2015; Wan et al. 2015) and further increases are expected in response to warming and increasing atmospheric humidity from reductions in sea ice extent. Therefore part of this paper is devoted to evaluating model skill in simulating the upper-tail of daily-precipitation annual distributions (Rx1day, Rx5day, R95ptot and R99ptot) and one related to the number of wet days (R1mm), with a wet day defined as a day with

Table 3 Definition of annual climate indices and bin sizes used for the estimations of the Perkins skill score (see Sect. 3.2)

Climate index	Name (description)	Unit	Bin size
Annual Tmean	Annual mean of daily mean temperature	°C	0.5 °C
DJF Tmean	December, January and February mean daily temperature	°C	0.5 °C
JJA Tmean	June, July and August mean daily temperature	°C	0.5 °C
Nthaw	Winter thaw events (number of days with Tmax > 0 °C during a winter period defined by −5 °C crossing dates from a 40-day centered moving average of daily Tmean)	Days	1 day
TDD	Thawing degree days (annual cumulative sum of daily degrees of Tmean above 0 °C)	°C	50 °C
FDD	Freezing degree days (annual cumulative sum of daily degrees of Tmean below 0 °C, over a winter-centred year)	°C	100 °C
HDD	Heating degree days (annual cumulative sum of Tmean daily degree days below 17 °C)	°C	300 °C
GDD	Growing degree days (annual cumulative sum of Tmean daily degree days above 5 °C)	°C	20 °C
SU15	Arctic summer days (annual count of days with Tmax > 15 °C)	Days	1 day
TXx	Warmest annual temperature (annual maximum value of Tmax)	°C	0.5 °C
TX90	90th percentile of Tmax over the year	°C	0.5 °C
TX90p	Warm days (annual percentage of time when Tmax is > 90th percentile of Tmax, which is calculated for a 5-day window centered on each calendar day of the reference period)	%	1%
FD	Frost days (annual number of days with Tmin < 0 °C)	Days	2 days
TNn	Coldest annual temperature (annual minimum value of Tmin)	°C	0.5 °C
TN10	10th percentile of Tmin over the year	°C	0.5 °C
TN10p	Cold days (annual percentage of time when Tmin is < 10th percentile of Tmin, which is calculated for a 5-day window centered on each calendar day in the reference period)	%	1%
Annual Pr	Annual mean of daily precipitation	mm/day	0.2 mm/day
DJF Pr	December, January and February mean daily precipitation	mm/day	0.2 mm/day
JJA Pr	June, July and August mean daily precipitation	mm/day	0.2 mm/day
R1mm	Wet days (annual count of days when daily precipitation is ≥ 1 mm/day)	Days	5 days
RX1day	Maximum 1-day precipitation amount (annual maximum of daily precipitation)	mm/day	2 mm/day
RX5day	Maximum 5-day precipitation amount (annual maximum of 5-day accumulated precipitation)	mm/day	3 mm/day
R95pTOT	Precipitation due to very wet days (annual total precipitation from days with daily precipitation > 95th percentile of wet-day precipitation over the reference period)	mm	20 mm
R95p	Very wet days (annual number of days with daily precipitation > 95th percentile of wet-day precipitation over the reference period)	Days	0.6 day
R99pTOT	Precipitation due to extremely wet days (annual total precipitation from days with daily precipitation > 99th percentile of wet-day precipitation over the reference period)	mm	20 mm
R99p	Extremely wet days (number of annual days with daily precipitation > 99th percentile of wet-day precipitation over the reference period)	Days	0.6 day

precipitation ≥ 1 mm/day as these may be impacted by climate change.

Some annual indices were defined using thresholds associated to percentiles computed over a reference period (e.g. TN10p, TX90p, R95ptot and R99ptot). The 1980–2004 period (25 years) was used as reference period for stations, reanalysis and GCM-driven simulations. For the comparison of GCM- and ERAI-driven simulations, these indices were estimated over the 17-year period 1989–2005.

In addition to these ETCCDI indices, annual, summer (June, July and August) and winter (December, January and February) mean daily precipitation, annual, summer and winter mean daily temperatures, heating degree days (HDD), growing degree days (GDD), freezing degree days (FDD), thawing degree days (TDD) and number of winter thaw events (Nthaw) were also estimated.

The GDD index is important in the Arctic for studies related to the growth, timing of vegetation green-up onset, insect development and migration (Herms 2004; Sridhar and Reddy 2013), which are key variables for caribou population, a major resource for local communities (Moerschel and Klein 1997; Stern and Gaden 2015). The TDD index is closely related to snow melt processes, the depth of the permafrost active layer and the initialisation of snowpack ablation (Stern and Gaden 2015), while the FDD index is related to ice growth and the depth of ground frost penetration (Stern and Gaden 2015) which are relevant for transportation (e.g. ice roads) and infrastructure. The Nthaw index is related to the production of ice layers within or under the snowpack that may limit caribou access to forage.

The climate indices were computed at each point of the original grid for each simulation/reanalysis as suggested in Diaconescu et al. (2015). For stations, annual indices were computed only for years with less than 20% missing values, defined as a valid year.

3.2 Evaluation metrics

Evaluation of models and reanalyses is usually carried out by comparing area-averaged statistics (e.g. mean values, variances) or by comparing the distributions of values of co-located pairs. The latter approach is difficult in our situation due to the limited number of years of data. Hansen et al. (2012) and Hansen and Sato (2016) proposed to pool data from different stations within a region into a single sample in order to increase the sample size. This approach is reasonable as long as the value of interest over that region can be considered homogeneous, i.e. that the local values can be described by a unique distribution. A similar method was adopted by Alexander et al. (2006) in the analysis of global climate annual indices from a subset of 200 temperature and 350 precipitation stations across the

globe. Donat and Alexander (2012), Hansen et al. (2012), and Hansen and Sato (2016) used probability density distributions of temperature anomalies over specific regions rather than absolute-value distributions because the anomalies have a higher spatial correlation, while absolute mean temperatures can vary over short spatial scales (Hawkins and Sutton 2016). Consequently, even a few stations can capture the temporal characteristics of anomalies over a large region. We therefore chose to use this approach and pooled the anomalies to compare the distributions.

Consequently, each local climate index value $Y(lon, lat, t)$ was decomposed into the local climatological mean, $\bar{Y}(lon, lat)$, and the local anomalies $Y'(lon, lat, t)$:

$$Y(lon, lat, t) = \bar{Y}(lon, lat) + Y'(lon, lat, t) \quad (1)$$

The local climatological mean corresponds to the mean value over the reference period at a specific site. The local anomalies were obtained by subtracting the local climatological mean from the local annual index value. The spatial distribution of the climatological mean and the distribution of anomalies were then estimated. Model and reanalysis skill in simulating the spatial patterns of the climatological mean were evaluated (Sect. 4.1.1) as well as the distribution of anomalies (Sect. 4.1.2).

For the spatial pattern evaluation, the Mean Squared Skill Score (MSSS; see e.g., Murphy 1988; Murphy and Epstein 1989) was used. It compares the Mean Square Error (MSE) of a forecast (a given RCM simulation or reanalysis) with respect to the observations ($f; o$) to the MSE of a selected reference dataset with respect to the observations ($r; o$):

$$MSSS = 1 - \frac{MSE(f, o)}{MSE(r, o)} \quad (2)$$

A positive value indicates that the forecast has a greater skill than the reference, with $MSSS = 1$ indicating a perfect forecast skill and $MSSS \sim 0$ similar forecast and reference skills. For instance if the forecast values more closely match observed values than the reference dataset then $MSE(f, o) < MSE(r, o)$ and $0 < MSSS \leq 1$ (otherwise $MSSS \leq 0$).

In this study, two versions of the MSSS were used to evaluate reanalysis and RCM skill in simulating the spatial patterns of climatological mean of indices over the Canadian Arctic:

1. MSSS with reference to the variance of observations also known as the Reduction of Variance (RV) metric. For a given climate index, this score compared the MSE between the mean for a given dataset (RCM simulation or reanalysis) to the spatial variance of the cli-

matological mean of observations over the region

$$\left(s_o^2 = \frac{1}{N} \sum_{i=1}^N (\bar{O}_i - \langle \bar{O} \rangle)^2 \right);$$

$$RV_k = 1 - \frac{MSE(k, O)}{s_o^2} = 1 - \frac{\sum_{i=1}^N (\bar{Y}_{ki} - \bar{O}_i)^2}{\sum_{i=1}^N (\bar{O}_i - \langle \bar{O} \rangle)^2} \quad (3)$$

with \bar{Y}_{ki} representing the climatological mean of dataset k (a RCM simulation or a reanalysis) at the grid point closest to station i , \bar{O}_i the recorded observational climatological mean at station i , $\langle \bar{O} \rangle$ the spatial mean over the analysis region (see Fig. 1) of observational climatological mean at stations and N the number of stations in the region. The model/reanalyses MSE is computed with respect to observations

$$\left(MSE(k, O) = \frac{1}{N} \sum_{i=1}^N (\bar{Y}_{ki} - \bar{O}_i)^2 \right).$$

A value of $RV_k > 0$

means that dataset k has a smaller MSE value than the spatial variance in observations.

- Another version of the MSSS compares the MSE of the index climatological mean of a given dataset to the MSE value averaged over all reanalyses:

$$MSSS_k = 1 - \frac{MSE(k, O)}{\frac{1}{5} \sum_{r=1}^5 MSE_r} \quad (4)$$

with $MSE(k, O)$ representing the MSE of a given dataset k (either model simulation or reanalysis) and MSE_r representing the MSE of reanalysis r . The denominator sum in Eq. 4 is over all five reanalysis datasets. This version is particularly useful because it compares individual RCM performance to mean reanalysis performance (a value of $MSSS_k > 0$ means that dataset k outperformed the mean reanalysis performance). It also provides information regarding the use of reanalyses as reference datasets for model evaluation in the Arctic region.

Evaluation of anomaly distributions, for each given index, was performed by pooling all annual anomalies at stations and corresponding grid-point values over the Canadian Arctic Region. Corresponding empirical distributions were then constructed.

The ability of reanalyses and models in simulating station anomaly distributions was assessed with the Kuiper goodness-of-fit metric (Kuiper 1960) and the Perkins metric (Perkins et al. 2007). The two metrics were considered to check whether they provided consistent conclusions. Both metrics have the advantage of not depending on the shape of the underlying distribution and they can be applied to any variable.

The Kuiper metric, inspired from the Kolmogorov–Smirnov test, is one of the most commonly used methods to compare distributions from two samples (Smirnov 1939; Stephens 1970). It measures the distance between the two empirical cumulative distributions and is defined as the sum of the absolute values of the maximum positive and negative distances between the two empirical cumulative distributions:

$$D_K = \max_{-\infty < x < \infty} [ECDF_k(x) - ECDF_o(x)] + \max_{-\infty < x < \infty} [ECDF_o(x) - ECDF_k(x)] \quad (5)$$

with $ECDF_k$ and $ECDF_o$ representing the empirical cumulative distributions of dataset k and of recorded datasets respectively. D_K values range between zero and one, with zero indicating a perfect overlap of the two distributions while a value of one corresponds to no overlapping distributions.

The Perkins metric is defined as the overlap between the two empirical probability density functions (EPDF) and has been used in evaluating temperature and precipitation series simulated by GCMs (Perkins et al. 2007, 2012; Maxino et al. 2008; Pitman and Perkins 2009; Perkins 2009) as well as RCMs (Kjellstrom et al. 2010; Kabela and Carbone 2015; Boberg et al. 2010). Here we extend its use by evaluating climate index anomaly distributions. Normalised histograms of indices from reanalyses and simulations were compared with corresponding index histograms from recorded datasets. The size of the bin used for each index is presented in Table 3. The bins were selected to cover the whole range of values of both datasets. The common area between the two distributions was computed as:

$$PSS_k = \sum_{x=1}^n \min(imum(EPDF_k(x), EPDF_o(x))) \quad (6)$$

where n is the number of bins used to calculate the normalised histograms, $EPDF_k(x)$ is the frequency in bin x for the dataset k , $EPDF_o(x)$ is the corresponding frequency for the recorded dataset in bin x . PSS values range from zero to one, a zero value corresponding to no overlap between the two histograms and one to identical histograms. To ease the comparison with Kuiper metric, the $(1 - PSS)$ metric will be used in the following.

In summary, the RV and MSSS metrics were used to compare the performance of the climatological mean, while the Kuiper and Perkins metrics were used to compare the performance of the distribution of annual anomalies (i.e. the annual departures from the climatological mean).

4 Evaluation of GCM-driven RCM and reanalyses indices

4.1 Climatological means

Index climatological means are presented first at stations over the Canadian Arctic (Sect. 4.1.1). Next the skill of models and reanalyses at reproducing the observed values

are evaluated using the RV metric, biases and the MSSS metric (Sects. 4.1.2, 4.1.3 and 4.1.4).

4.1.1 Observed climatological means

The spatial coherence of the observed climate indices at stations was first examined visually (see Figs. 3, 4 for the climatological means of selected indices). Most of the temperature indices showed important spatial gradients

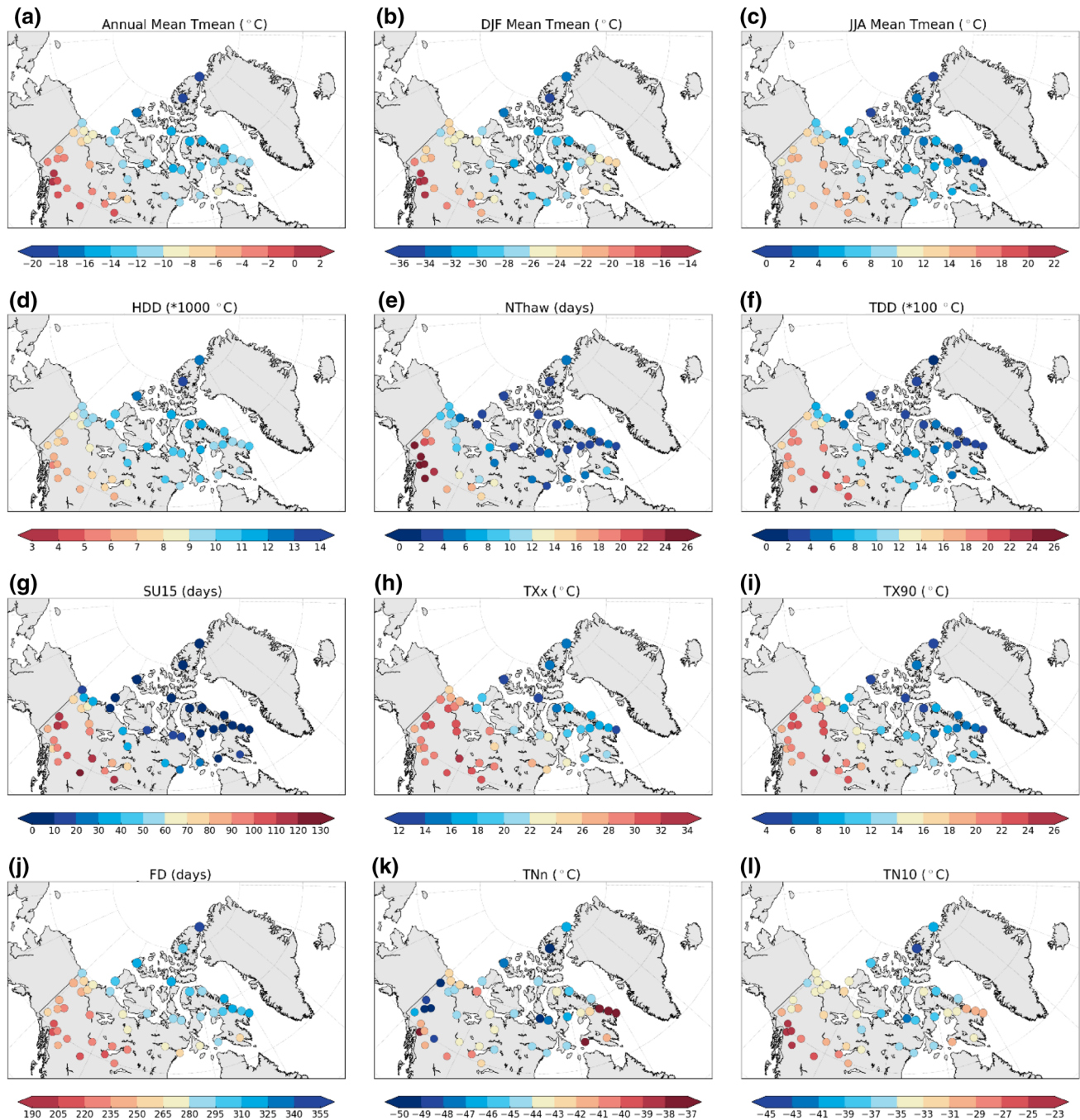


Fig. 3 Climatological mean of some daily temperature indices at stations over the 1980–2004 period

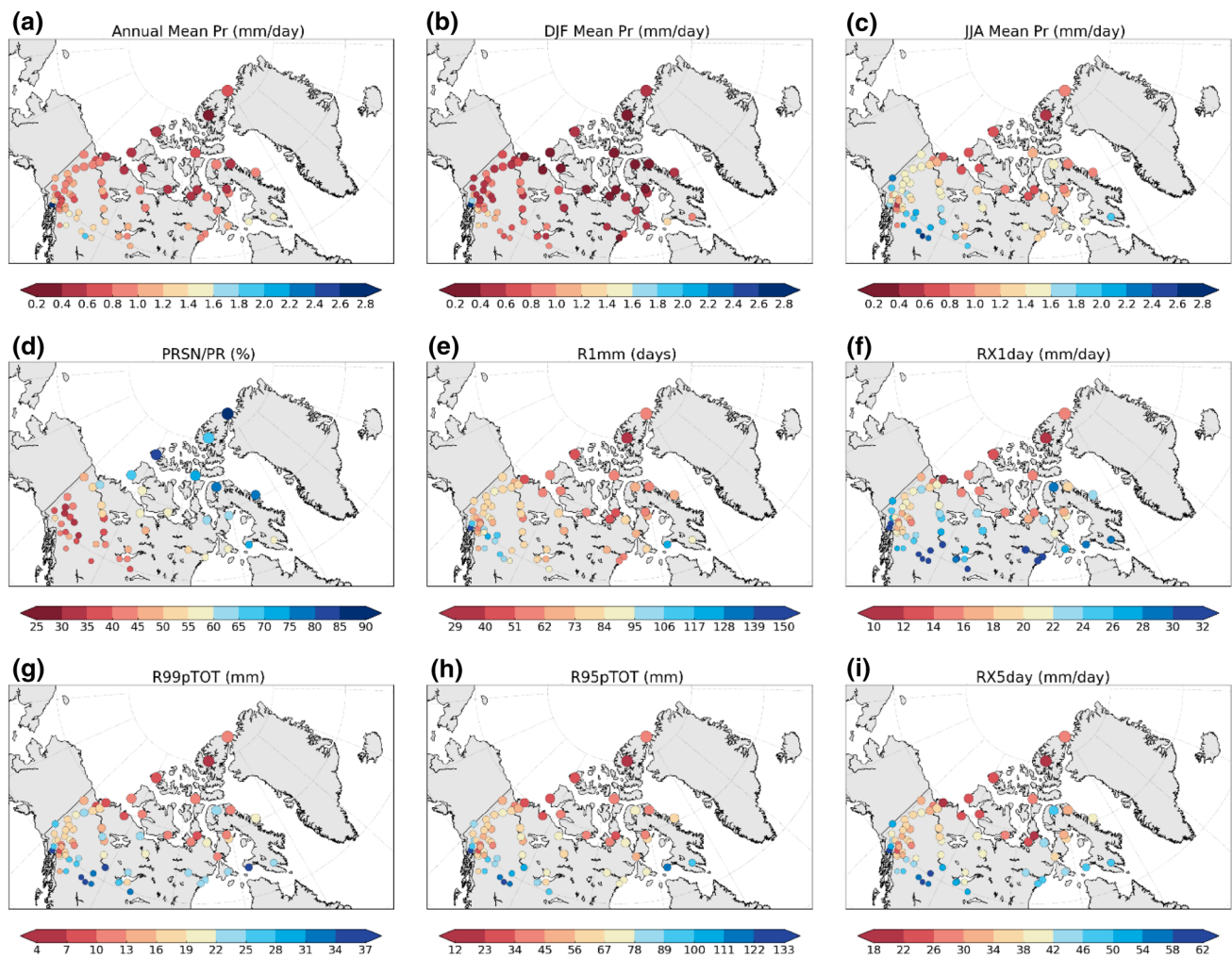


Fig. 4 Climatological mean of daily precipitation indices at stations over the 1980–2004 period. In order to improve the readability of the map, the stations in the South–West region are represented by *smaller dots* than the stations in the North

over the Arctic, consistent with the radiative forcing. The only exception was the annual coldest temperature (TNn; Fig. 3k), which is influenced by local scale factors such as proximity to open water and topography that influence inversion layer formation (Rapać et al. 2015).

For precipitation indices, the entire domain is characterised by relatively small amounts of precipitation, with most stations having mean annual precipitation rates of less than 1.0 mm/day (Fig. 4a). More precipitation arrives in the summer period than winter (Fig. 4b, c) because warmer temperature in summer and the presence of ice-free water (lakes and Arctic Ocean) increase atmospheric moisture. The fraction of total precipitation falling in solid form is varying from 30 to 90% over the region (Fig. 4d). Solid precipitation (PRSN/PR between 60 and 100%) dominates in the north where the mean annual precipitation is very low (less than 1 mm/day),

while liquid precipitation is more important over the southern regions (PRSN/PR between 40 and 20%) where mean annual precipitation has values between 1.2 and 2.8 mm/day. Analysis of daily-precipitation distributions shows that higher intensity daily precipitation can occur especially in the southern part of the domain and in some coastal regions with open water during the summer. Also, in summer, extratropical cyclones can penetrate further north and can reach the Hudson Bay, in winter such trajectories are unlikely (Reitan 1974). Consequently, precipitation is generally less extreme over the northern regions of the study area (RX1day/RX5day less than 20/30 mm/day; R99pTOT less than 16 mm and R95pTOT less than 45 mm) and more extreme over southwestern regions and south of Baffin Island (RX5day between 30 and 70 mm/day and R99pTOT with values between 20 and 40 mm).

4.1.2 RV metric

The performance of the models and reanalysis in simulating the spatial patterns of the climatological means of the precipitation and temperature indices was first evaluated using the RV metric (Eq. 3), which was computed using the station records as a reference as described in Sect. 2.3 and presented in Sect. 4.1.1. Heat maps in Fig. 5 summarize the estimated RV values with columns representing the indices, and rows the datasets. The last two columns denoted “All T” and “All Pr” present the average performance of each data set in simulating all temperature indices (All T) and all precipitation indices (All Pr), and corresponds to the mean RV value over all temperature or precipitation indices of the corresponding dataset. The top two rows of Fig. 5 present the RV of the ensemble mean of reanalyses (identified as REM and corresponding to the RV values of the corresponding index averaged over all five reanalyses) and of the model ensemble mean (identified as MEM and corresponding to the RV values of the corresponding index averaged over all simulations). In the following, for ease of

interpretation, ‘good skill’ (or ‘good performance’) will be associated with positive RV values (blue boxes in Fig. 5), for which the squared differences between model/reanalysis and station indices were smaller than the recorded spatial variance among stations, while ‘poor skill’ (or ‘poor performance’) will be associated with negative RV values (red boxes in Fig. 5), for which the MSE of a model/reanalysis were greater than the spatial variance among stations.

Figure 5 shows that the overall model and reanalysis performance was better for the temperature indices (first 14 columns) than for the precipitation indices (last eight columns). Ten of the temperature indices (annual Tmean, DJF Tmean, JJA Tmean, Nthaw, TDD, FDD, HDD, GDD, SU15 and TX90) were well simulated by all reanalyses and almost all RCM simulations. The reanalyses demonstrated a good ability at simulating the FD and TXx indices (only ERAI has a negative RV for TXx), while many RCM simulations were less effective for these two indices. The TNn index was poorly reproduced by most of the reanalysis and RCM simulations. The reanalyses’ performance was better than the RCMs for 11 of the 14 temperature indices, as also

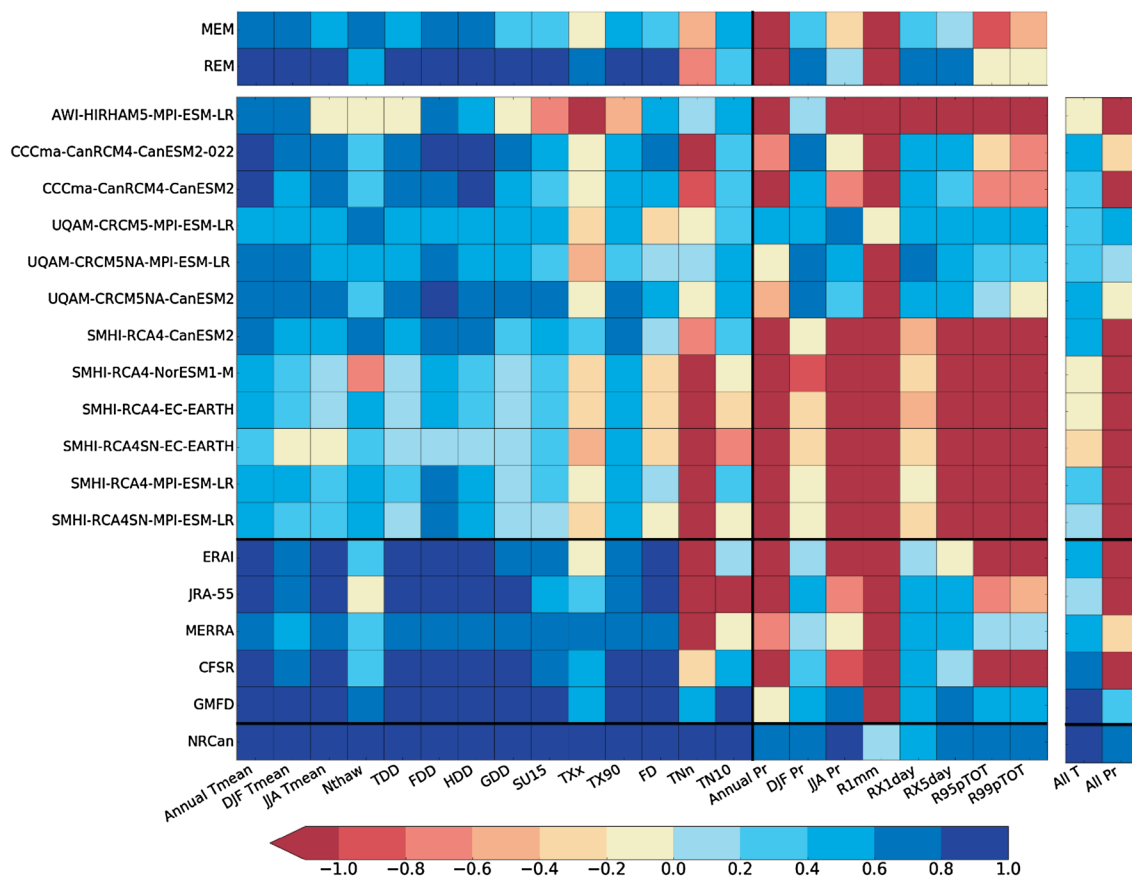


Fig. 5 Heat map showing the reduction of variance metric for the mean temperature and precipitation indices over the 1980–2004 period for the Canadian Arctic. *MEM* model ensemble mean, *REM* reanalysis ensemble mean. All T corresponds to the RV value averaged over all temperature indices and All Pr to the RV value averaged over all precipitation indices. The RV metric is computed against the station observations described in Sect. 2.3 and presented in Figs. 3 and 4

indicated by the REM and MEM performances. For Nthaw and TN10, MEM was better than REM, while for TNn both MEM and REM had negative RV values.

The climatological means of DJF Pr (16th column of Fig. 5), RX1day (19th column of Fig. 5) and RX5day (20th column of Fig. 5) indices were well reproduced by most reanalysis and by the CCCma-CanRCM4 and UQAM-CRCM5 simulations. The ability to reproduce these indices was also reflected by the positive RV of REM and MEM. Only two reanalysis (GMFD and MERRA) and the UQAM-CRCM5 simulations were capable of reproducing the climatological mean of R95pTOT (21th column of Fig. 5) and R99pTOT (22th column of Fig. 5). Also, all reanalysis and most RCM simulations showed poor performance for Annual Pr (15th column of Fig. 5), JJA Pr (17th column of Fig. 5) and R1mm (18th column of Fig. 5) indices. Therefore, the reanalyses and some models were able to effectively reproduce the annual extremes of daily precipitation (R1days and RX5days) but show poor performance for the annual and summer mean of daily precipitation (annual Pr and JJA Pr) and the annual number of wet days (R1mm) indicating a frequency bias in the occurrence of precipitation. While RCMs have similar skill in simulating the DJF Tmean and JJA Tmean, reanalyses show better skill for the summer mean temperature. Almost all RCMs and reanalyses have better scores in simulating DJF Pr than JJA Pr.

Figure 5 also compares the performances of different configurations of the same RCM. For the SMHI-RCA model, configurations with (SMHI-RCASN) and without (SMHI-RCA) spectral nudging were evaluated and show that the configuration without spectral nudging had superior skill for some temperature indices, while performances were similar for precipitation indices.

For the CCCma-CanRCM4 model, simulations with two spatial resolutions were compared. Figure 5 shows that the model configuration at 0.22° resolution (CCCma-CanRCM4-CanESM2-022) had better performance than the 0.44° configuration (CCCma-CanRCM4-CanESM2) for DJF Tmean, FDD, GDD, SU15, FD, Annual Pr, DJF Pr, RX5day and R95pTOT indices. Conversely, the other indices had similar performances.

4.1.3 Biases

To examine the potential influence of geography in model performance, the spatial patterns of the differences in climatological mean values between MEM/REM and observations for the DJF Tmean, JJA Tmean and RX1day indices are presented in Fig. 6. These patterns are representative of most RCMs and reanalysis and most indices: DJF Tmean for cold and winter temperature indices (Nthaw, DJF Tmean, TN10, TNn), JJA Tmean for warm temperature and

degree-day indices, and RX1day for precipitation indices. The analysis of the spatial distribution of the RCM/reanalysis - observations differences for all temperature indices showed different patterns for cold and winter-time indices and all other temperature indices. Most reanalyses and especially RCMs present a larger bias over the Yukon high-topography region for the warm indices and the annual daily degree indices (Fig. 6a, b), while the cold indices don't display any particular pattern (Fig. 6c, d). The larger negative errors observed over the mountain region in indices incorporating the summer temperatures (Fig. 6a, b) can be due to the difference in topographic elevation in RCMs/reanalyses and stations, stations being usually located in valleys. Consequently, RCMs/reanalysis temperatures are usually colder than temperatures recorded at stations since mean grid-point elevations of RCMs/reanalysis are typically higher than station elevations.

These results suggest that some bias correction based on the temperature climatological lapse rate for this region should be applied to correct station temperatures. However, the difference between station altitude and corresponding RCM/reanalysis mean grid-point elevation does not seem to affect cold and winter indices (Fig. 6c, d), for which no structured spatial distribution of differences was observed. In Fig. 6, mean temperature from reanalyses are warmer than observed ones (Fig. 6d) while models simulate colder mean temperatures over most part of the Arctic (Fig. 6c). This difference in spatial pattern of the bias between warm and cold temperature indices can be explained by the high frequency of surface and upper tropospheric temperature inversions during the cold period (December–March) caused by the radiative surface cooling or by the warmer air advection over the arctic cold air masses (Lovatt 2009; Serreze and Barry 2005; Przybylak 2016). Important «semi-permanent» inversions are present especially in the deep valleys of the Yukon and the Alaska mountainous regions (Lovatt 2009; Przybylak 2016). Consequently, local processes have a high impact on cold indices and a bias correction based on the climatological lapse rate would not be appropriate to correct cold daily temperature series. Statistical methods or more complex physical methods that incorporate local processes conditioning the seasonal evolution of the local lapse rate would be needed to adequately correct daily minimum, maximum and mean temperature indices over this specific region. Precipitation indices do not present a particular pattern in the bias (see for example RX1day bias in Fig. 6e, d) with the exception of higher bias values for more southerly points and those close to the Pacific Coast (especially for annual Pr and R1mm indices—not shown) where mean and extreme precipitation amounts are larger (see Fig. 4).

In summary, for temperature indices, reanalyses and RCMs generally performed well in simulating the

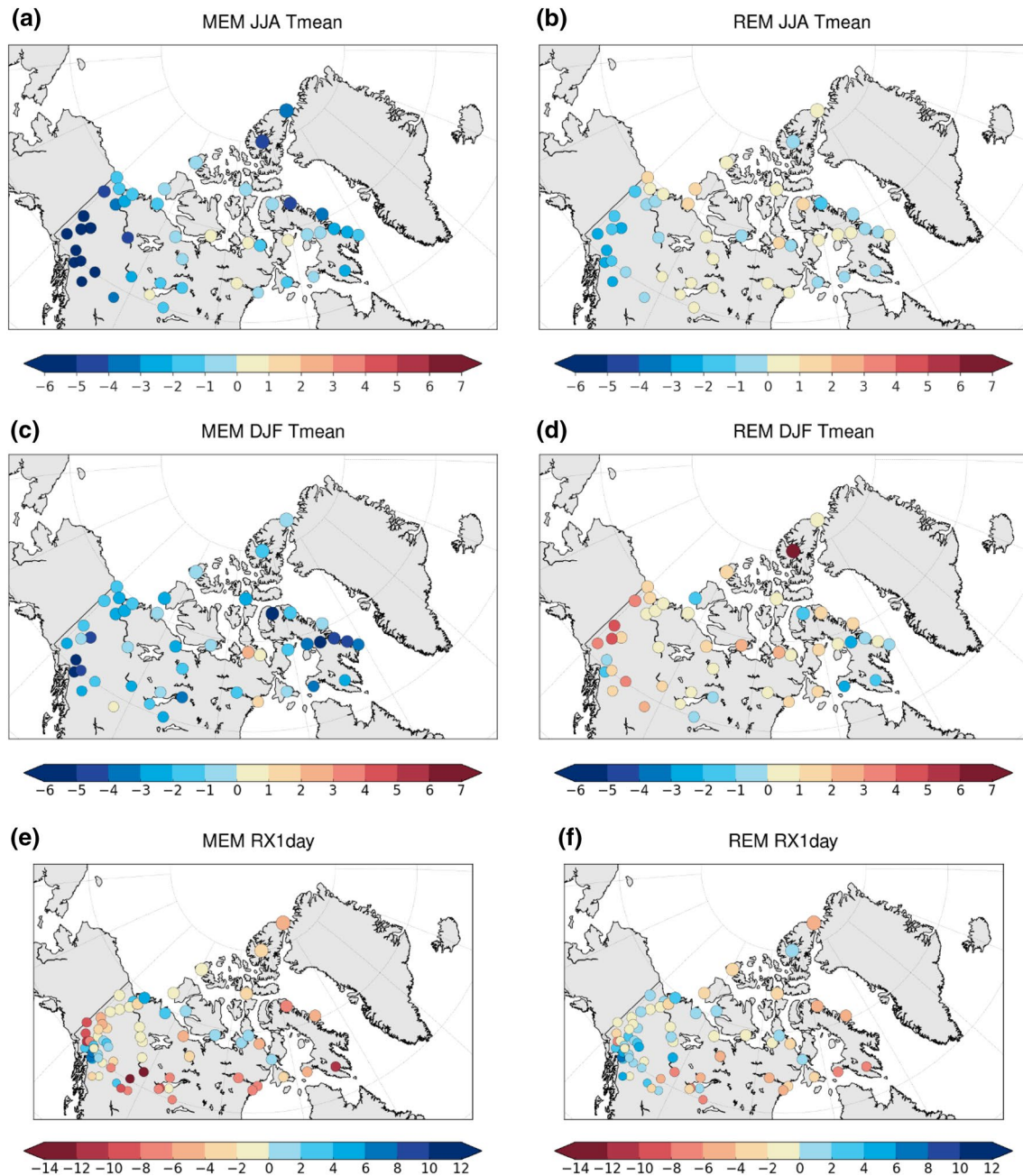
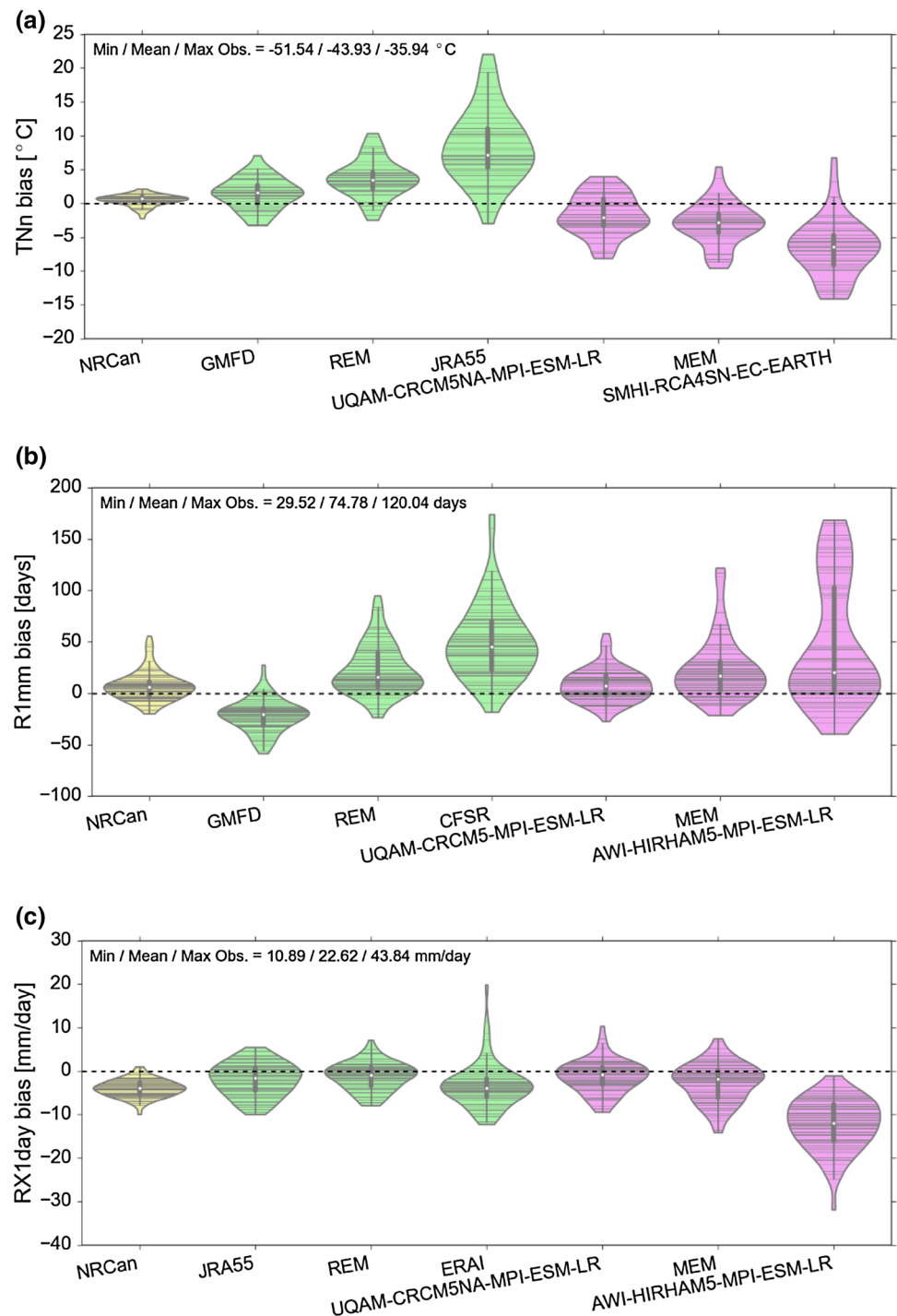


Fig. 6 Differences between the model ensemble mean (**a**, **c** and **e**) and reanalysis ensemble mean (**b**, **d** and **f**), and observations for climatological mean of JJA Tmean in °C (**a**, **b**), DJF Tmean in °C (**c**, **d**) and RX1day in mm/days (**e**, **f**) over Arctic

climatological means of mean-daily and maximum-daily temperature indices over most Canadian Arctic. The bias is generally larger over the Yukon region in indices based on summer temperatures. RCMs and reanalyses were less effective at simulating the TNn index. Regarding precipitation indices, most reanalyses and some RCMs were able to reproduce climatological means of DJF Pr, RX1day and RX5day indices effectively, but were less effective in simulating the annual Pr and the R1mm indices.

Violin plots (Hintze and Nelson 1998; computed using the Seaborn Python package: <https://stanford.edu/~mwaskom/software/seaborn/index.html>) were used in Fig. 7 to examine the climatological mean biases of TNn, R1mm (that were poorly represented by reanalyses and models) and RX1day (that was well simulated by most reanalyses and RCMs). Violin plots show the density of the distribution of biases over the Canadian Arctic based on a kernel smoother. The more “squashed” the violin and the

Fig. 7 Violin plots showing mean bias kernel distributions of selected reanalysis and models over the 1980–2004 period for the Canadian Arctic for: **a** TNn, **b** R1mm, and **c** RX1day indices. Each horizontal line on the violins represents the bias at one station, while the envelope of the violin represents the kernel density estimation of the underlying bias distribution at all stations. Violin plots in green correspond to the reanalyses and in magenta to the RCMs. Mean, minimum and maximum values recorded over all stations are presented on the top left corner of each graph. Box plots inside the violin plots (in gray) display the 1st and 3rd quartiles of the bias distribution as well as the median values (white dot). Each violin is scaled to have the same width



closer the median (white dot) is to zero, the closer the values are to the observations. The first violin of each graph corresponds to the NRCan dataset and shows the biases between grid-point index values and corresponding values at the closest stations. The next three violins, in green, capture the range of results from the reanalyses i.e. the reanalysis with the lowest RV metric (left), the reanalysis ensemble mean (REM) (middle), and the reanalysis with

the largest RV value (right). The range in RCM simulations is similarly represented by the three magenta violins.

For the TNn index, the interpolation process in NRCan smoothed the minimum daily temperature values resulting in an overall small positive bias on most of the grid points close to stations, with the median bias equal to 0.63 °C. All reanalyses displayed an overall warm bias in agreement with previous studies (e.g. Rapačić et al. 2015)

with a median bias of $+3.4^{\circ}\text{C}$ for REM. Contrary to reanalyses, RCMs had a cold bias in TNn with MEM bias of approximately -2.8°C . A warm bias in reanalyses and a cold bias in simulations was also observed for the 10th percentile of minimum daily temperatures (TN10 index), but with smaller amplitude than for the TNn index (not shown). The poor performance of RCMs and reanalyses in simulating the cold extremes, is attributed in part to a lack of local-scale information related to inversion layer development. The representation of atmospheric humidity profiles and clouds in models/reanalyses would also play roles in radiative cooling.

Four of the five reanalyses and all RCM simulations displayed positive R1mm biases, and therefore a larger number of wet days than recorded at the stations, with maximum biases in the southwestern region of the domain characterised by high topography (see Fig. 2). Small negative biases were observed for the northern regions. An exception was the GMFD product with a smaller number of wet days than observations over a large part of the domain (median bias of -20 days per year) and the UQAM-CRCM5-MPI-ESM-LR simulation, which had the smallest model-and-reanalysis median bias (approximately 7 days per year) similar to the NRCan median bias (approximately 6 days per year).

The interpolation procedure has an important impact on the estimated NRCan RX1day index (see Fig. 7c; Hutchinson et al. 2009). The interpolation process introduces a smoothing of daily precipitation that results in a negative bias for RX1day (underestimation of RX1day values estimated from recorded series) for almost all NRCan grid points next to stations (only two stations in the north had positive biases and these stations were probably not integrated with the NRCan product). NRCan biases ranged from -9.9 to $+1.0$ mm/day with a median bias of -4.0 mm/day, a value larger than the REM (-0.9 mm/day) and MEM median biases (-1.9 mm/day). REM good performance was also illustrated by the RV metric presented in Fig. 5, REM has a greater RV value than NRCan for this index. The best reanalysis performance for this index was obtained for JRA55 with a median bias of -1.6 mm/day. This value was larger than the REM median bias, the good REM performance being in part due to bias compensation in the average process across the five reanalyses (positive and negative biases compensating each other). The best performance for RX1day index was obtained by UQAM-CRCM5NA-MPI-ESM-LR simulation that has a median bias of -0.79 mm/day, which was better than all five reanalyses and the NRCan dataset. Overall, the median biases for REM and MEM were similar and close to the best RCM/reanalyses value.

4.1.4 MSSS metric

The skillfulness of the RCMs and reanalyses simulations were also compared by using the mean MSE of reanalyses as a reference dataset in the MSSS metric (Eq. 4). The corresponding heat maps are plotted in Fig. 8 where, as in Fig. 5, each row corresponds to a dataset and each column to an index. The top two lines compare REM and MEM MSE to the mean MSE of the five reanalyses, while the last two columns show the average value of the MSSS metrics for all temperatures ('All T') and precipitation ('All Pr') indices for a given dataset. The positive MSSS values, in blue, signify that the dataset outperformed the mean reanalysis performance.

The performances of the RCM simulations were below the mean performances of reanalyses for all temperature indices (negative values for MSSS metric), except for the Nthaw and the cool extreme indices TNn and TN10. For the mean-temperature and maximum-temperature indices (the first 12 columns), all reanalyses demonstrated good performances (e.g. Fig. 5), with GMFD performance well above the mean reanalysis value (positive values for MSSS metric), CFSR performance close to the mean reanalysis performance, and ERAI and MERRA performances generally below the mean reanalysis value.

The overall excellent RV scores obtained for the reanalyses in Fig. 5 suggest that these can be used as reference dataset for RCM evaluation for mean and maximum temperature indices. However, most reanalyses were not effective in simulating the TNn and TN10 indices. Moreover, Fig. 8 shows that most RCM simulations have errors smaller than or similar to the mean reanalysis MSE for these indices.

For precipitation indices, the performances of the five reanalyses were similar, with GMFD displaying an overall higher performance and with CFSR and ERAI having performances below the average reanalysis value. The five simulations with CCCma-CanRCM4 and UQAM-CRCM5 had simulation performances superior to the reanalysis average. This suggests that reanalyses should not be used as reference datasets for RCM evaluation for precipitation indices over the Arctic.

As for the individual models, the performance of the MEM for the warm and mean temperature extremes was below the average reanalysis performance, while for the cold extremes (TNn and TN10), the Nthaw, the R1mm, R95pTOT and R99pTOT indices, it was above. Bias compensation was more effective for REM presenting positive MSSS metric for all indices. Nevertheless, the performances of UQAM-CRCM5 simulations were better than the REM performance for TNn, Mean Pr, R1mm, R95pTOT and R99pTOT.

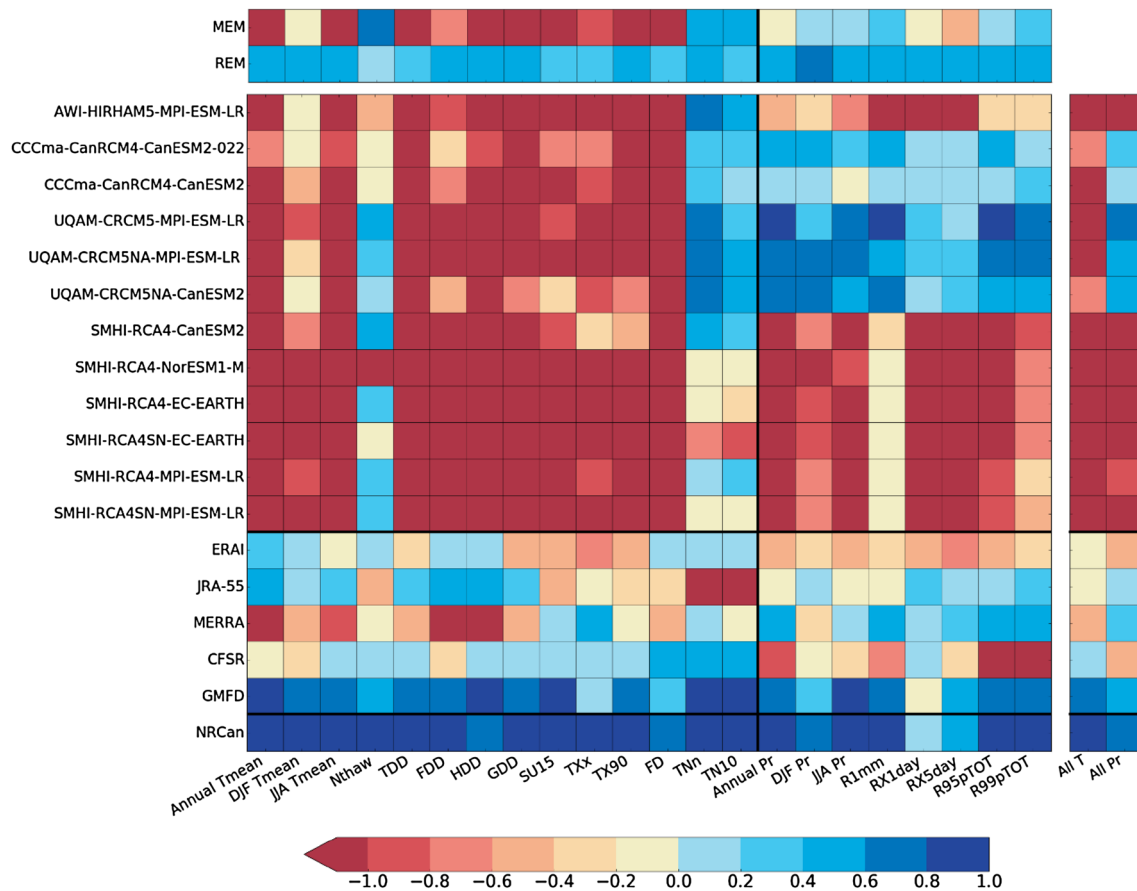


Fig. 8 Heatmap showing the MSSS metric for mean temperature and precipitation indices over the 1980–2004 period for the Canadian Arctic. *MEM* model ensemble mean, *REM* reanalysis ensemble mean. All T and All Pr as defined in Fig. 5

In summary, reanalyses outperformed RCM simulations for mean and warm daily temperature indices and the best performances were obtained by GMFD and CFSR. This suggests that they could be used as a reference in the RCM evaluation of daily temperature indices over the region. However, for daily precipitation indices, the performances of reanalyses were lower and some RCM simulations even outperformed all reanalyses. Caution is therefore recommended when using reanalyses as reference datasets when evaluating RCM performance for daily precipitation indices.

4.2 Anomaly distributions

Empirical distributions of anomalies from RCMs/reanalyses and surface observations were compared to determine the skill of reanalyses and models in simulating the observed distributions. Anomalies from all station records in the Canadian Arctic were pooled and ECDF and EPDF were computed. The skills of reanalyses and models in simulating the observed distribution were evaluated using Kuiper metric (D ; Eq. 4) and Perkins skill scores (PSS; Eq. 5).

The values obtained for each climate index and for each dataset are plotted in Fig. 9 using a diagram inspired by the Performance Portrait diagram of Gleckler et al. (2008).

Figure 9 shows that the two metrics gave similar results in all cases (as a reminder a value of zero corresponds to a perfect match of the two distributions). Reanalyses and RCMs were very effective in simulating the anomalies' distribution for most indices. Lower scores were obtained by RCM simulations for GDD and SU15 indices and by RCM simulations and reanalyses for the R99p index. Among the RCM simulations, the AWI-HIRHAM5-MPI-ESM-LR simulations poorly reproduced anomaly distributions for R99p, Rx1day, Rx5day and SU15 indices, but had excellent scores for DJF Tmean, Nthaw, FDD, TXx, TX90p, TNn and R1mm indices. Figure 10 presents an example of the observations (in blue) and AWI-HIRHAM5-MPI-ESM-LR (in red) Rx1day and TNn anomaly ECDFs and EPDFs. The figure highlights the similarity of simulated and observed anomaly distributions for TNn index, characterised by a Kuiper distance of 0.05 and $(1 - \text{PSS})$ value of 0.06 (Fig. 10b). For the RX1day index, the range of values of model anomalies is

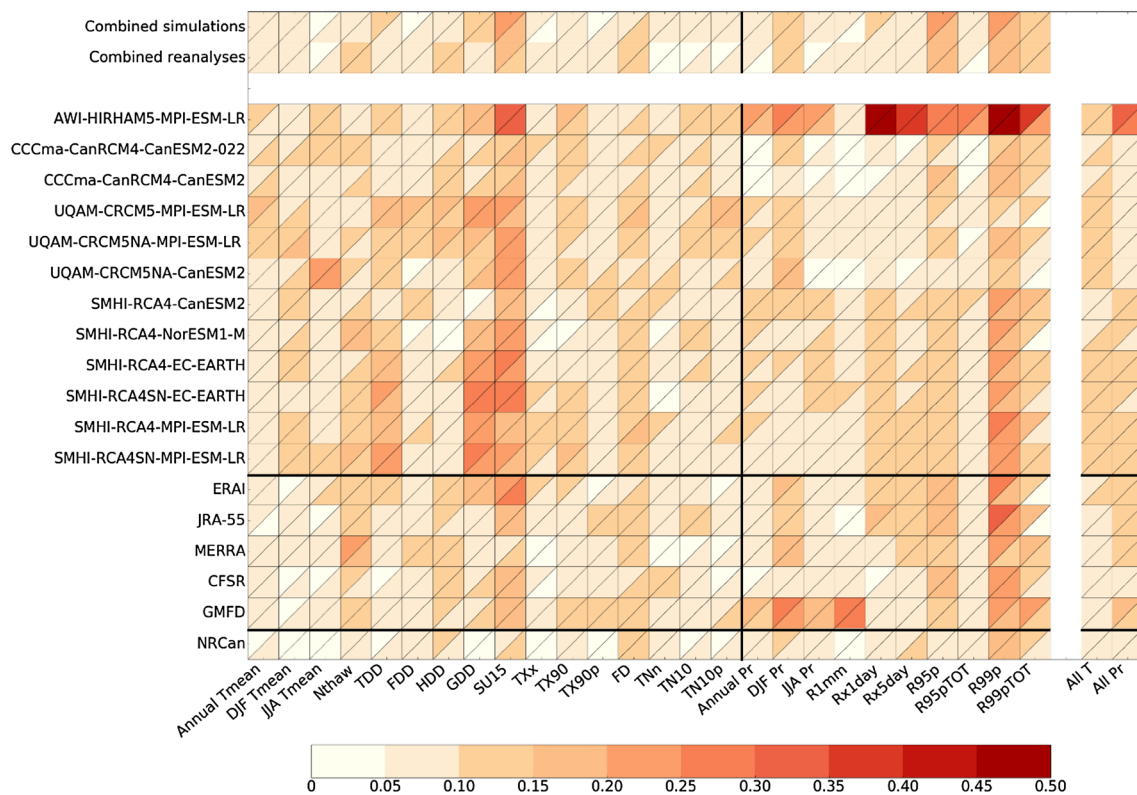


Fig. 9 Heatmaps showing the Kuiper distance values (*upper-left triangles*) and the (1-PSS) values (*bottom-right triangles*) between RCM or reanalysis anomaly distributions and the observed anomaly distributions for the Canadian Arctic over the 1980–2004 period. The *central part* of the figure presents results obtained for the individual

datasets. The ‘Combined simulations’ (*top row*) and ‘Combined reanalysis’ (*second top row*) correspond to the cases where the anomaly distributions were constructed by pooling the index anomalies from all simulations or all reanalyses. All T and All Pr as defined in Fig. 5

smaller than that of observations, resulting into a smaller inter-annual variability. This is added to an already poor performance for the climatological mean RX1day index (12th column in Fig. 5).

When anomalies from all reanalyses or all simulations were pooled together (top two rows of Fig. 9), the resulting distributions resembled closely the recorded anomaly distributions ($D < 0.1$ and $1 - \text{PSS} < 0.1$) for a majority of the indices. Therefore, for reanalyses and RCM climate indices over the Canadian Arctic, the main errors are related to bias in the mean index values, while these datasets showed an overall good performance in simulating the anomaly distribution over the region. These results suggest that, once the climatological mean of indices have been bias corrected, some datasets, displaying good performance in reproducing the observed anomalies over the Arctic region for a given index, can be used to assess its mean value over the Arctic, but also its inter-annual variability.

5 Comparison of GCM-driven and ERAI-driven simulations

The skill of GCM-driven and ERAI-driven RCMs in simulating climate indices were compared over a common period of 17 years (1989–2005). Most of the published RCM evaluation studies have been conducted using RCM simulations driven by reanalyses, which presumably represents the most realistic lateral boundary conditions. However, GCM-driven simulations are used to develop climate projections and therefore the errors introduced by GCMs in these simulations should be assessed. The comparison of errors in simulations driven by GCM with those driven by reanalyses provides some information on the contributions of RCM structural bias and GCM bias to the total errors.

Figure 11 presents the results of this comparison for six RCM configurations (CCma-CanRCM4 at 0.22° and at 0.44° resolution; UQAM-CRCM5 over Arctic and North-America domains; SMHI-RCA with and without spectral nudging). Some RCM configurations were driven only by

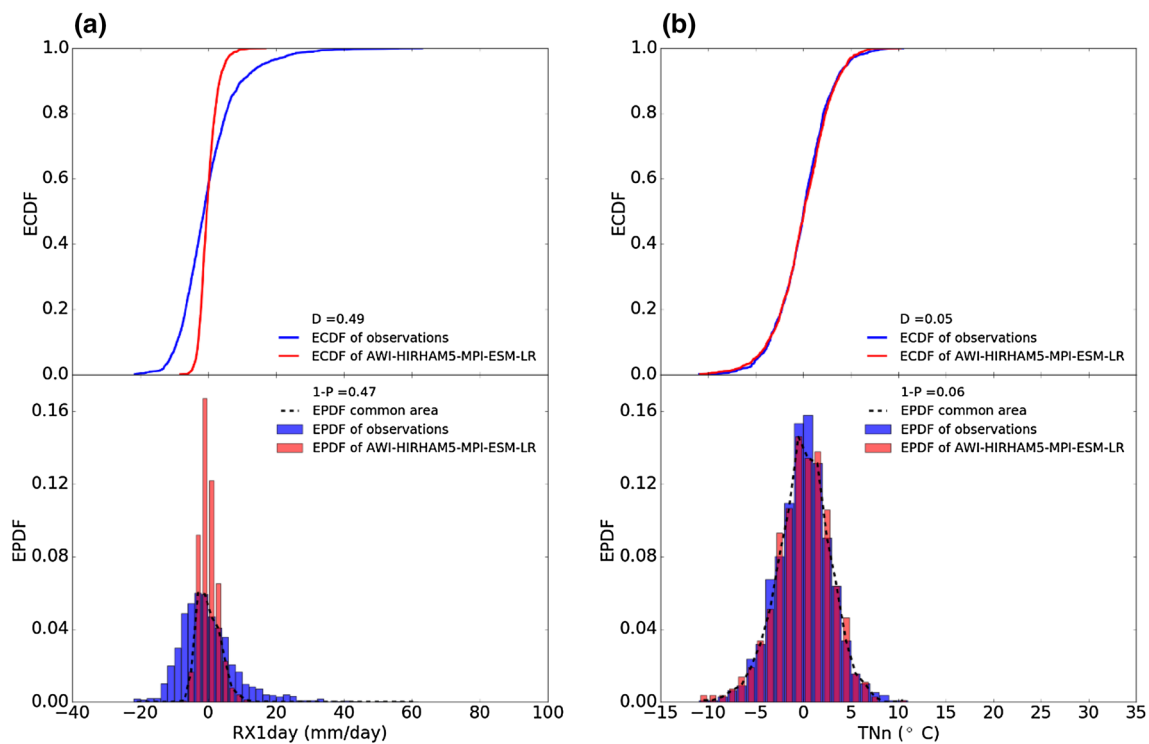


Fig. 10 Comparison of the ECDFs and EPDFs for AWI-HIRHAM5-MPI-ESM-LR simulation (in red) and observations (in blue) for RX1day (a) and TNn (b) indices for the Canadian Arctic over the

1980–2004 period. The *dashed lines* appearing in the EPDF graphs highlight the common area of the two EPDFs

one GCM (CCCma-CanRCM4; CCCma-CanRCM4_022; UQAM-CRCM5), while some others were driven by two GCMs (UQAM-CRCM5NA, SMHI-RCASN) or four GCMs (SMHI-RCA). Figure 11 compares, for each RCM configuration, the mean performance of the GCM-driven simulations to that of the ERAI-driven simulation. The upper-left triangle in Fig. 11 presents the mean performance for all simulations from a given RCM configuration driven by GCMs, while the bottom-right triangle represents the performance of the same RCM configuration driven by ERAI. The effectiveness of simulating the climatological mean (Fig. 11a) was measured using the reduction of variance as in the previous section, while the performance in simulating the anomaly distribution (Fig. 11b) was measured using Kuiper distance.

Figure 11 shows that differences in performance between GCM- and ERAI-driven simulations were globally more pronounced for climatological means than for anomalies. In general, anomaly distributions were very well reproduced by RCMs whether they were driven by GCMs or by ERAI (in most cases D was smaller than 0.2). For a majority of indices, ERAI-driven simulations showed better performance at simulating climatological means. SMHI-RCA4 and SMHI-RCA4SN had negative RV metrics for TXx, FD and RX1day when driven by GCMs, but

positive values when driven by ERAI. These two RCM configurations poorly simulated TNn, Annual Pr, JJA Pr, R1mm, R95pTOT and R99pTOT mean indices even when driven by ERAI. Note that ERAI did not effectively simulate either of the latter indices (see Fig. 5).

Negative RV values were also obtained for the CCCma-CanRCM4 simulations for TNn, Annual Pr, JJA Pr and R1mm indices when driven by GCM or ERAI. Nevertheless, for R95pTOT and R99pTOT indices, CCCma-CanRCM4_022, CCCma-CanRCM4, UQAM-CRCM5 and UQAM-CRCM5NA, driven by GCMs and by ERAI, performed well, while ERAI did not (see Fig. 5 for RV of ERAI). Surprisingly, UQAM-CRCM5 and CCCma-CanRCM4 produced better simulations when driven by a GCM than when driven by ERAI for these two precipitation indices. This suggests that the simulated fields within the domain can be improved by these RCMs, therefore adding value to the representation of ERAI for these indices.

The RCM-added value, with respect to ERAI, can also be assessed through the ratio of RCM MSE and ERAI MSE as presented in Fig. 12 for each mean climate index. CCCma-CanRCM4_022, CCCma-CanRCM4, UQAM-CRCM5, and UQAM-CRCM5NA produced smaller MSE than ERAI for precipitation indices. The better representation of precipitation indices by some RCMs is probably due

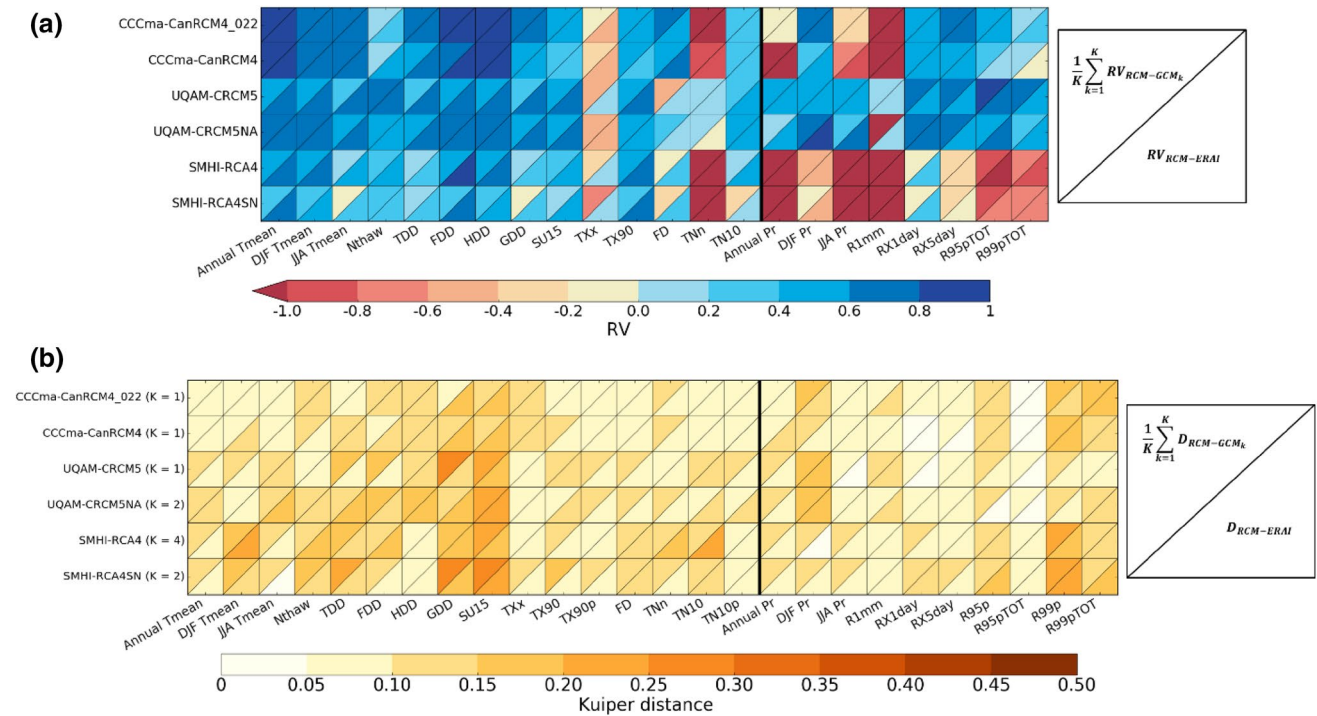


Fig. 11 Heat maps comparing the mean performance of RCMs driven by GCM simulations (*upper-left triangles*) with the performance of RCM driven by ERAI simulation (*bottom-right triangles*) for **a** climatological mean of indices based on the RV metric; and **b** anomaly distribution over Canadian Arctic based on the Kuiper dis-

tance (D). The equations are presented on the right-hand side diagram. The period of analysis is 1989–2005. K next to simulation names, corresponds to the number of GCMs used as lateral boundary conditions for each RCM configuration

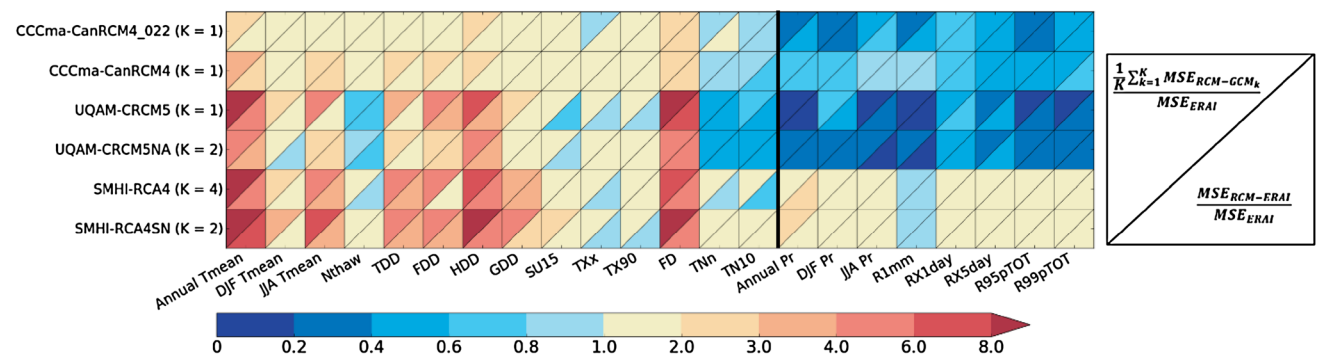


Fig. 12 RCM added-value diagram for the climatological mean of indices over Canadian Arctic. The *upper-left triangles* correspond to the mean MSE of RCM simulations driven by GCMs ($MSE_{RCM-GCM_k}$) divided by the MSE of ERAI (MSE_{ERAI}), with the corresponding equation in the *upper-left triangle* of the right-hand side diagram. The corresponding *bottom-right triangles* cor-

respond to the MSE of the corresponding RCM driven by ERAI ($MSE_{RCM-ERAI}$) divided by the MSE of ERAI, with the corresponding equation in the *bottom-right triangle* of the right-hand side diagram. The period of analysis is 1989–2005 (only valid years are considered). K as defined in Fig. 10

to a better representation of the physics in these models as well as to their higher spatial resolution. The comparison of the CCCma-CanRCM4_022 MSE with the CCCma-CanRCM4 MSE also demonstrated that, for the Annual Pr, DJF Pr, JJA Pr, R1mm, R95pTOT and R99pTOT indices,

the higher resolution simulation provided a better simulation of mean climate values.

Added value was also observed in GCM-driven and ERAI-driven UQAM-CRCM5 and UQAM-CRCM5NA simulations of Nthaw, TNn and TN10 indices compared

to ERAI. For SU15, TXx and TX90, slightly better scores were obtained by some ERAI-driven RCMs than for ERAI. For these indices, the RCM's higher resolution may explain the slight improvement with respect to ERAI. The fact that the corresponding GCM-driven simulations had lower performance than ERAI for SU15, TXx and TX90 indices, is explained by the presence of larger errors in GCM temperature fields than in the ERAI fields applied at the RCM lateral boundary conditions.

As expected, the overall performance in reproducing temperature indices depended on lateral boundary datasets, ERAI-driven RCMs tending to have generally better scores than GCM-driven RCM for these indices.

6 Summary and conclusions

The present study used climate station observations across the Canadian Arctic to evaluate five daily precipitation and ten daily temperature indices simulated by (1) an RCM ensemble of 12 GCM-driven simulations and six ERAI-driven simulations participating in the CORDEX experiment, (2) five reanalyses products (GMFD, CFSR, MERRA, JRA-55 and ERAI) and (3) one high-resolution gridded observational product (NRCan). The reanalysis and GCM-driven simulations were first evaluated over a reference period of 25 years (1980–2004). Records from 47 stations for temperature and 78 stations for precipitation were used. Climatological means and annual anomaly distributions were evaluated separately for each index and compared to corresponding index values estimated from recorded datasets. The impact of lateral boundary conditions was also analyzed by comparing the GCM-driven simulations to the corresponding ERAI-driven simulations over a common period of 17 years (1989–2005). In this case, records from 48 stations for temperature and 79 stations for precipitation were used.

The analysis of mean climate indices over the reference period showed that:

- The RCMs, and especially the reanalyses, demonstrate good performance at simulating the mean and warm daily temperature indices over most Canadian Arctic.
- Few RCMs and reanalyses performed well in simulating the TN10, RX1day, RX5day, R95pTOT and R99pTOT indices. UQAM-CRCM5 simulations outperformed the five reanalyses for these precipitation indices.
- Most RCM simulations and reanalyses performed poorly for the TNn, Mean Pr and R1mm indices.
- The gridded product NRCan showed similar values to observations (for grid points next to stations) for temperature indices, but the interpolation procedure appeared to have an impact on precipitation indices. The

most impacted index was RX1day for which NRCan had biases similar to those of reanalyses and some RCM simulations.

The evaluation of the spectral nudging impact on SMHI-RCA4 simulations has shown that the configuration without spectral nudging had better performances for most temperature indices. However, both configurations poorly performed for the precipitation indices. Similar results were obtained for the climatological means of mean daily temperature and precipitation in agreement with Berg et al. (2013).

The impact of increasing the spatial resolution from 0.44 to 0.22 was analysed for CCCma-CanRCM4. Some improvement was observed for the GDD, SU15, FD and RX5day indices in higher-resolution simulation. In a recent analysis of daily precipitation indices using CCCma-CanRCM4 simulations at 0.22° and 0.44° resolutions, but integrated over the North-America CORDEX domain, Diaconescu et al. (2016) also found that the simulation at 0.22° resolution was more effective for some summer-time indices over Canada south of 60°N, but not for winter-time indices, the skill in simulating the winter indices being already good in the 0.44° version of the model. Prein et al. (2016) also showed that added value can be obtained in higher resolution simulations for regions with complex orography.

Empirical anomaly distributions from RCMs/reanalyses and recorded series were compared using two metrics: the Kuiper and Perkins metrics. Anomalies from all station records and corresponding grid-points in the Canadian Arctic were pooled and corresponding empirical distributions were computed. Kuiper and Perkins metrics produced similar results and demonstrated, for most reanalyses and RCM simulations, very good agreement between simulated and recorded empirical anomaly distributions for most indices (metrics values smaller than 0.2). Comparing the relative performances of RCM/reanalysis in simulating climatological mean of the selected indices and their anomaly distribution suggests that the main errors are associated with mean climate estimation of indices. Therefore a simple bias correction (post-treatment) of the mean of indices could result in a good representation of the analysed indices across the Arctic. It must be noted that this conclusion concerns only the indices presented in the study. Also, since metrics that were used to compare annual index anomaly distributions (i.e. Kuiper and Perkins metrics) are not tailored to assess specifically the performance in distribution tails, a more detailed analysis would be needed to assess the performance for more extreme index values (e.g. anomalies associated with large return periods).

The present study also demonstrates that the temperature indices were impacted by the type of lateral boundary

conditions, ERAI-driven RCM having in general better performances than GCM-driven RCM for these indices. For precipitation indices, the comparison of UQAM-CRCM5, UQAM-CRCM5NA, CCCma-CanRCM4 and CCCma-CanRCM4_022 simulations to ERAI demonstrated that these RCMs, driven by ERAI or GCM, outperformed ERAI. The accuracy of the simulated precipitation indices depend on the RCM physics and added value can be obtained in some simulations by improving the precipitation representation. Some ERAI-driven RCMs have also slightly better scores than ERAI for SU15, TXx and TX90 indices. This added value may be attributed to the higher resolution of RCM compared to ERAI, RCM resolution being closer to the point resolution of observations than ERAI.

In conclusion, the very good performance of reanalyses for mean and warm temperature indices supports their use as reference dataset for RCM evaluation. However, the performance of reanalyses for the TN10, Mean Pr and R1mm indices was poor and some RCM simulations even produced better scores than reanalyses in simulating extreme precipitation. Therefore, we do not recommend that reanalyses be used as a reference dataset in RCM evaluations of these indices. Additionally, given the observed impact of the interpolation procedure on estimated NRCan indices and especially extreme indices even at grid-points adjacent to stations, we do not support the use of this dataset as a reference for data-sparse regions such as the Arctic.

Acknowledgements This study was financially supported by the ArcticNet research program. The authors are grateful to the modelling centers that contributed to the simulation, reanalysis and observation datasets used in this paper. The CFSR dataset was developed by NOAA's National Centers for Environmental Prediction (NCEP). The MERRA dataset was provided by the Global Modeling and Assimilation Office (GMAO) and the GES DISC for the dissemination of MERRA. The ERA-Interim, JRA55 and GMFD datasets were retrieved from the NCEP Research Data Archive (RDA) (rda.ucar.edu). The NRCan dataset was developed by Natural Resources Canada, and is available at <http://cfs.nrcan.gc.ca/projects/3/4>. The AHCCD and DAI datasets were developed by Environment Canada and are available at http://ccrp.tor.ec.gc.ca/pub/EC_data/AHCCD_daily/ and <http://climat-quebec.qc.ca/CC-DEV/trunk/index.php/pages/dai> respectively. The authors acknowledge the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modelling, former coordinating body of CORDEX and responsible panel for CMIP5. They also thank the climate modelling groups (listed in Table 1 of this paper) for producing and making available their model output. AWI-HIRHAM5 datasets were provided by Annette Rinke and Ines Hebestadt from Alfred-Wegener Institut. The UQAM-CRCM5 dataset was provided by René Laprise and Oumarou Nikiema from Université du Québec à Montréal. The UQAM-CRCM5NA dataset was provided by Katja Winger and Laxmi Sushama from Université du Québec à Montréal. CCCma-CanRCM4 dataset was provided by CCCma Environment Canada at http://www.cccma.ec.gc.ca/french/data/canrcm/CanRCM4/index_cordex.shtml. The SMHI-RCA4 dataset was downloaded from the Earth System Grid Federation Infrastructure an international effort led by the US Department of Energy's Program for Climate Model

Diagnosis and Intercomparison, the European Network for Earth System Modelling and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP). We would finally like to thank the two anonymous reviewers for all their comments and suggestions that significantly improved the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alexander LV (2016) Global observed long-term changes in temperature and precipitation extremes: a review and update since IPCC AR5. *Weather Clim Extremes* 11(2016):4–16. doi:[10.1016/j.wace.2015.10.007](https://doi.org/10.1016/j.wace.2015.10.007)
- Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res Atmos* 111:D05109. doi:[10.1029/2005JD006290](https://doi.org/10.1029/2005JD006290)
- Allard M, Lemay M (2012) Nunavik and Nunatsiavut: from science to policy. An Integrated Regional Impact Study (IRIS) of climate change and modernization. ArcticNet Inc., Quebec. http://www.arcticnet.ulaval.ca/pdf/media/iris_report_complete.pdf. Accessed Jan 2017
- Arora VK, Scinoca JF, Boer GJ, Christian JR, Denman KL, Flato MG, Kharin VV, Lee WG, Merryfield WJ (2011) Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys Res Lett* 38:L05805. doi:[10.1029/2010GL046270](https://doi.org/10.1029/2010GL046270)
- Bentsen M, Bethke I, Debernard JB, Iversen T, Kirkevåg A, Seland Ø, Drange H, Roelandt C, Seierstad IA, Hoose C, Kristjansson JE (2013) The Norwegian earth system model, NorESM1-M. Part 1: Description and basic evaluation of the physical climate. *Geosci Model Dev* 6:687–720. doi:[10.5194/gmd-6-687-2013](https://doi.org/10.5194/gmd-6-687-2013)
- Berg P, Doscher R, Koenigk T (2013) Impacts of using spectral nudging on regional climate model RCA4 simulations of the Arctic. *Geosci Model Dev* 6:849–859. doi:[10.5194/gmd-6-849-2013](https://doi.org/10.5194/gmd-6-849-2013)
- Boberg F, Berg P, Thejll P, Gutowski WJ, Christensen JH (2010) Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. *Clim Dyn* 35:1509–1520. doi:[10.1007/s00382-009-0683-8](https://doi.org/10.1007/s00382-009-0683-8)
- Booij MJ (2002) Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. *Int J Climatol* 22:69–85. doi:[10.1002/joc.715](https://doi.org/10.1002/joc.715)
- Chan SC, Kendon EJ, Fowler HJ, Blenkinsop S, Ferro CAT, Stephenson DB (2013) Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? *Clim Dyn* 41(5–6):1475–1495. doi:[10.1007/s00382-012-1568-9](https://doi.org/10.1007/s00382-012-1568-9)
- Chen CT, Knutson T (2008) On the verification and comparison of extreme rainfall indices from climate models. *J Clim* 21(7):1605–1621. doi:[10.1175/2007jcli1494.1](https://doi.org/10.1175/2007jcli1494.1)

- Christensen OB, Drews M, Christensen JH, Dethloff K, Ketelsen K, Hebestadt I, Rinke A (2007) The HIRHAM regional climate model version 5 (b). Technical report 06–17, Copenhagen. <http://www.dmi.dk/fileadmin/Rapporter/TR/tr06-17.pdf>. Accessed July 2016
- Cohen J, Screen JA, Furtado JC, Barlow M, Whittleston D, Coumou D, Francis J, Dethloff K, Entekhabi D, Overland J, Jones J (2014) Recent Arctic amplification and extreme mid-latitude weather. *Nat Geosci* 7(9):627–637. doi:10.1038/ngeo2234
- Contractor S, Alexander LV, Donat MG, Herold N (2015) How well do gridded datasets of observed daily precipitation compare over Australia? *Adv Meteorol*. doi:10.1155/2015/325718
- Cowtan K, Way RG (2014) Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q J R Meteorol Soc* 140:1935–1944. doi:10.1002/qj.2297
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Holm EV, Isaksen L, Kallberg P, Kohler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavalato C, Thépaut J-N, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597. doi:10.1002/qj.828
- Dethloff K, Rinke A, Lehmann R, Christensen JH, Botzet M, Machenhauer B (1996) Regional climate model of the Arctic atmosphere. *J Geophys Res* 101(D18):23401–23422. doi:10.1029/96JD02016
- Devine KA, Mekis E (2008) Field accuracy of Canadian rain measurements. *Atmos Ocean* 46(2):213–227. doi:10.3137/ao.460202
- Diaconescu EP, Gachon P, Laprise R (2015) On the remapping procedure of daily precipitation statistics and indices used in regional climate model evaluation. *J Hydrometeorol* 16:2301–2310. doi:10.1175/JHM-D-15-0025.1
- Diaconescu EP, Gachon P, Laprise R, Scinocca JF (2016) Evaluation of precipitation indices over North America from various configurations of regional climate models. *Atmos Ocean*. doi:10.1080/07055900.2016.1185005
- Donat MG, Alexander LV (2012) The shifting probability distribution of global daytime and night-time temperatures. *Geophys Res Lett* 39(14):L14707. doi:10.1029/2012GL052459
- Donat MG, Alexander LV, Yang H, Durre I, Vose R, Dunn R, Willett K, Aguilar E, Brunet M, Caesar J, Hewitson B, Jacks C, Klein Tank AMG, Kruger A, Marengo J, Peterson TC, Renom M, Oria Rojas C, Rusticucci M, Salinger J, Sekele S, Srivastava A, Trewin B, Villarreal C, Vincent L, Zhai P, Zhang X, Kitching S (2013) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J Geophys Res Atmos* 118(5):2098–2118. doi:10.1002/jgrd.50150
- Fischer EM, Knutti R (2015) Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat Clim Change* 5(6):560–564. doi:10.1038/nclimate2617
- Fowler HJ, Ekström M, Kilsby CG, Jones PD (2005) New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. *J Hydrol* 300:212–233. doi:10.1016/j.jhydrol.2004.06.017
- Gervais M, Gyakum JR, Atallah EH, Tremblay LB (2014) How well are the distribution and extreme values of daily precipitation over North America represented in the community climate system model? A comparison to reanalysis, satellite, and gridded station data. *J Clim* 27:5219–5239. doi:10.1175/JCLI-D-13-00320.1
- Giorgetta MA, Jungclaus J, Reick C, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmid H, Schnur R, Segschneider J, Six KD, Stockhause M, Timmreck C, Wegner J, Widmann H, Weiners K, Claussen M, Marotzke J, Stevens B (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J Adv Model Earth Sys* 5:572–597. doi:10.1002/jame.20038
- Giorgi F, Jones C, Asrar G (2009) Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bull* 58:175–183
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res* 113(D6):D06104. doi:10.1029/2007JD008972
- Glisan JM, Gutowski WJ Jr (2014a) WRF summer extreme daily precipitation over the CORDEX Arctic. *J Geophys Res Atmos* 119:1720–1732. doi:10.1002/2013jd020697
- Glisan JM, Gutowski WJ Jr (2014b) WRF winter extreme daily precipitation over the North American CORDEX Arctic. *J Geophys Res Atmos* 119:10738–10748. doi:10.1002/2014JD021676
- Gutowski WJ Jr, Giorgi F, Timbal B, Frigon A, Jacob D, Kang H-S, Raghavan K, Lee B, Lennard C, Nikulin G, O'Rourke E, Rixen M, Solman S, Stephenson T, Tangang F (2016) WCRP coordinated regional downscaling experiment (CORDEX): a diagnostic MIP for CMIP6. *Geosci Model Dev* 9:4087–4095. doi:10.5194/gmd-9-4087-2016
- Hansen J, Sato M (2016) Regional climate change and national responsibilities. *Environ Res Lett* 11(2016):034009. doi:10.1088/1748-9326/11/3/034009
- Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. *Rev Geophys* 48(4):RG4004. doi:10.1029/2010RG000345
- Hansen J, Sato M, Ruedy R (2012) Perception of climate change. *Proc Natl Acad Sci USA* 109(37):E2415–E2423. doi:10.1073/pnas.1205276109
- Hawkins E, Sutton R (2016) Connecting climate model projections of global temperature change with the real world. *Bull Am Meteorol Soc* 97(6):963–980. doi:10.1175/BAMS-D-14-00154.1
- Hazeleger W, Wang X, Severijns C, Ștefănescu S, Bintanja R, Sterl A, Wyser K, Semmler T, Yang S, van den Hurk B, van Noije T, van der Linden E, van der Wiel K (2012) EC-Earth V2.2: description and validation of a new seamless Earth system prediction model. *Clim Dyn* 39:2629. doi:10.1007/s00382-011-1228-5
- Hermes DA (2004) Using degree-days and plant phenology to predict pest activity. In: Krischik V, Davidson J, IPM (Integrated Pest Management) of Midwest Landscapes (eds) Minnesota Agricultural Experiment Station Publication 58-07645. <http://cues.cfans.umn.edu/old/ipmbook.htm>. Accessed Jan 2017
- Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. *Am Stat* 52:181–184. doi:10.1080/00031305.1998.10480559
- Hofstra N, New M, McSweeney C (2010) The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Clim Dyn* 35(5):841–858. doi:10.1007/s00382-009-0698-1
- Hutchinson MF, McKenney DW, Lawrence K, Pedlar JH, Hopkinson RF, Milewska E, Papadopol P (2009) Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. *J Appl Meteor Climatol* 48:725–741. doi:10.1175/2008JAMC1979.1
- Jones C, Giorgi F, Asrar G (2011) The Coordinated Regional Downscaling EXperiment CORDEX, an international downscaling link to CMIP5. *CLIVAR Exch* 16:34–40
- Kabela ED, Carbone GJ (2015) NARCCAP model skill and bias for the Southeast United States. *Am J Clim Change* 4:94–114. doi:10.4236/ajcc.2015.41009

- Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J Clim* 20:1419–1444. doi:[10.1175/JCLI4066.1](https://doi.org/10.1175/JCLI4066.1)
- Kharin VV, Zwiers FW, Zhang X, Wehner M (2013) Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim Change* 119:345–357. doi:[10.1007/s10584-013-0705-8](https://doi.org/10.1007/s10584-013-0705-8)
- Kjellstrom E, Boberg F, Castro M, Christensen JH, Nikulin G, Sanchez E (2010) Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Clim Res* 44:135–150. doi:[10.3354/cr00932](https://doi.org/10.3354/cr00932)
- Klein Tank AMG, Zwiers FW, Zhang X (2009) Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Climate data and monitoring WCDMP-No. 72, WMO-TD No. 1500, pp 56. http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP_72_TD_1500_en_1.pdf. Accessed May 2017
- Knutti R, Rogelj J, Sedláček J, Fischer EM (2016) A scientific critique of the two-degree climate change target. *Nat Geosci* 9(1):13–18. doi:[10.1038/ngeo2595](https://doi.org/10.1038/ngeo2595)
- Kobayashi S, Ota Y, Harada Y, Ebata A, Moriwa 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* 93:5–48. doi:[10.2151/jmsj.2015-001](https://doi.org/10.2151/jmsj.2015-001)
- Koenigk T, Berg P, Döscher R (2015) Arctic climate change in an ensemble of regional CORDEX simulations. *Polar Res* 34:24603. doi:[10.3402/polar.v34.24603](https://doi.org/10.3402/polar.v34.24603)
- Kuiper NH (1960) Tests concerning random points on a circle. *Indagat Math Proc* 63:38–47. doi:[10.1016/S1385-7258\(60\)50006-0](https://doi.org/10.1016/S1385-7258(60)50006-0)
- Laprise R, Hernández-Díaz L, Tete K, Sushama L, Šeparović L, Martynov A, Winger K, Valin M (2013) Climate projections over CORDEX Africa domain using the fifth-generation Canadian regional climate model (CRCM5). *Clim Dyn*. doi:[10.1007/s00382-012-1651-2](https://doi.org/10.1007/s00382-012-1651-2)
- Lindsay R, Wensnahan M, Schweiger A, Zhang J (2014) Evaluation of several different atmospheric reanalysis products in the Arctic. *J Clim* 27(7):2588–2606. doi:[10.1175/JCLI-D-13-00014.1](https://doi.org/10.1175/JCLI-D-13-00014.1)
- Lovatt A (2009) Characteristics of low-level temperature inversions, Whitehorse Yukon Territory: 1956–2003. Master Thesis. Department of Geography. University of Ottawa, Canada. <https://www.ruor.uottawa.ca/handle/10393/28207>. Accessed Jan 2017
- Martynov A, Laprise R, Sushama L, Winger K, Šeparović L, Dugas B (2013) Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian regional climate model, version 5: model performance evaluation. *Clim Dyn* 41(11–12):2973–3005. doi:[10.1007/s00382-013-1778-9](https://doi.org/10.1007/s00382-013-1778-9)
- Matthes H, Rinke A, Dethloff K (2010) Variability of extreme temperature in the Arctic—observation and RCM. *Open Atmos Sci J* 4:126–136. doi:[10.2174/1874282301004010126](https://doi.org/10.2174/1874282301004010126)
- Matthes H, Rinke A, Dethloff K (2015) Recent changes in Arctic temperature extremes: warm and cold spells during winter and summer. *Environ Res Lett* 10:114020. doi:[10.1088/1748-9326/10/11/114020](https://doi.org/10.1088/1748-9326/10/11/114020)
- Maxino CC, McAvaney BJ, Pitman AJ, Perkins SE (2008) Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int J Climatol* 28:1097–1112. doi:[10.1002/joc.1612](https://doi.org/10.1002/joc.1612)
- Mearns LO, Gutowski WJ, Jones R, Leung L-Y, McGinnis S, Nunes AMB, Qian Y (2009) A regional climate change assessment program for North America. *EOS* 90:311–312. doi:[10.1029/2009EO360002](https://doi.org/10.1029/2009EO360002)
- Mekis E (2005) Adjustments for trace measurements in Canada. In: 15th Conference on applied meteorology. American Meteorological Society, Savannah
- Mekis E, Brown R (2010) Derivation of an adjustment factor map for the estimation of the water equivalent of snowfall from ruler measurements in Canada. *Atmos Ocean* 48(4):284–293
- Mekis E, Vincent LA (2011) An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmos Ocean* 49(2):163–177. doi:[10.1080/07055900.2011.583910](https://doi.org/10.1080/07055900.2011.583910)
- Moerschel FM, Klein DR (1997) Effects of weather and parasitic insects on behavior and group dynamics of caribou of the Delta Herd, Alaska. *Can J Zool* 75:1659–1670. doi:[10.1139/z97-793](https://doi.org/10.1139/z97-793)
- Murphy AH (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon Weather Rev* 116:2417–2424. doi:[10.1175/1520-0493\(1988\)116<2417:ssbotm>2.0.co;2](https://doi.org/10.1175/1520-0493(1988)116<2417:ssbotm>2.0.co;2)
- Murphy AH, Epstein ES (1989) Skill scores and correlation coefficients in model verification. *Mon Weather Rev* 117:572–582. doi:[10.1175/1520-0493\(1989\)117<0572:SSACCI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2)
- Perkins SE (2009) Smaller projected increases in 20-year temperature returns over Australia in skill-selected climate models. *Geophys Res Lett* 36:L06710. doi:[10.1029/2009GL037293](https://doi.org/10.1029/2009GL037293)
- Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate model's simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J Clim* 20:4356–4376. doi:[10.1175/jcli4253.1](https://doi.org/10.1175/jcli4253.1)
- Perkins SE, Irving DB, Brown JR, Power SB, Moise AF, Colman RA, Smith I (2012) CMIP3 ensemble climate projections over the Western Tropical Pacific based on model skill. *Clim Res* 51:35–58. doi:[10.3354/cr01046](https://doi.org/10.3354/cr01046)
- Pitman AJ, Perkins SE (2009) Global and regional comparison of daily 2-m and 1000-hPa maximum and minimum temperatures in three global reanalyses. *J Clim* 22:4667–4681. doi:[10.1175/2009JCLI2799.1](https://doi.org/10.1175/2009JCLI2799.1)
- Prein AF, Gobiet A, Truhetz H, Keuler K, Goergen K, Teichmann C, Maule CF, Van Meijgaard, E, Déqué M, Nikulin G, Vautard R (2016) Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits? *Clim Dyn* 46(1–2):383–412. doi:[10.1007/s00382-015-2589-y](https://doi.org/10.1007/s00382-015-2589-y)
- Przybylak R (2016) The climate of the Arctic, 2nd edn. Atmospheric and Oceanographic Sciences Library 52, Springer International Publishing, Berlin. doi:[10.1007/978-3-319-21696-6](https://doi.org/10.1007/978-3-319-21696-6)
- Rapaic M, Brown R, Markovic M, Chaumont D (2015) An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. *Atmos Ocean* 53:283–303. doi:[10.1080/07055900.2015.1045825](https://doi.org/10.1080/07055900.2015.1045825)
- Reitan CH (1974) Frequencies of cyclones and cyclogenesis for North America, 1951–1970. *Mon Weather Rev* 102(12):861–868. doi:[10.1175/1520-0493\(1974\)102<0861:FOCACF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0861:FOCACF>2.0.CO;2)
- Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim G-K, Bloom S, Chen J, Collins D, Conaty A, Da Silva A, Gu W, Joiner J, Koster, RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. *J Clim* 24(14):3648. doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1)
- Saha SK, Rinke A, Dethloff K (2006) Future winter extreme temperature and precipitation events in the Arctic. *Geophys Res Lett* 33:L15818. doi:[10.1029/2006GL026451](https://doi.org/10.1029/2006GL026451)
- Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou YT, Chuang HY, Juang H-MH, Sela J, Iredell M, Treadon R, Kleist D, Delst PV, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Dool HVD, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm JK, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou

- CZ, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M (2010) The NCEP climate forecast system reanalysis. *Bull Am Meteorol Soc* 91(8):1015–1057. doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1)
- Samuelsson P, Jones CG, Willen U, Ullerstig A, Gollvik S, Hansson U, Jansson C, Kjellstrom E, Nikulin G, Wyser K (2011) The Rossby Centre regional model RCA3: model description and performance. *Tellus A* 63(1):4–23. doi:[10.1111/j.1600-0870.2010.00478.x](https://doi.org/10.1111/j.1600-0870.2010.00478.x)
- Scinocca JF, Kharin VV, Jiao Y, Qian MW, Lazare M, Solheim L, Dugas B (2016) Coordinated global and regional climate modelling. *J Clim* 29:17–35. doi:[10.1175/JCLI-D-15-0161.1](https://doi.org/10.1175/JCLI-D-15-0161.1)
- Šeparović L, Alexandru A, Laprise R, Martynov A, Sushama L, Winger K, Tete K, Valin M (2013) Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model (CRCM5). *Clim Dyn* 41(11–12):3167–3201. doi:[10.1007/s00382-013-1737-5](https://doi.org/10.1007/s00382-013-1737-5)
- Serreze MC, Barry RG (2005) The Arctic climate system. Cambridge atmospheric and space science series. Cambridge University Press, Cambridge. doi:[10.1017/CBO9780511535888](https://doi.org/10.1017/CBO9780511535888)
- Sheffield J, Goteti G, Wood EF (2006) Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J Clim* 19(13):3088–3111. doi:[10.1175/JCLI3790.1](https://doi.org/10.1175/JCLI3790.1)
- Sillmann J, Roeckner E (2008) Indices for extreme events in projections of anthropogenic climate change. *Clim Change* 86:83–104. doi:[10.1007/s10584-007-9308-6](https://doi.org/10.1007/s10584-007-9308-6)
- Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013a) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J Geophys Res Atmos* 118:1716–1733. doi:[10.1002/jgrd.50203](https://doi.org/10.1002/jgrd.50203)
- Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D (2013b) Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J Geophys Res Atmos* 118:2473–2493. doi:[10.1002/jgrd.50188](https://doi.org/10.1002/jgrd.50188)
- Smirnov NV (1939) Estimate of deviation between empirical distribution functions in two independent samples. *Bull Moscow Univ* 2(2):3–16 (**Russian**)
- Sridhar V, Reddy PVR (2013) Use of degree days and plant phenology: a reliable tool for predicting insect pest activity under climate change conditions. In: *Climate-resilient horticulture: adaptation and mitigation strategies*. Springer, Berlin, pp 287–294. doi:[10.1007/978-81-322-0974-4_25](https://doi.org/10.1007/978-81-322-0974-4_25)
- Stephens MA (1970) Use of the Kolmogorov–Smirnov, Cramer–Von-Mises and related statistics without extensive tables. *J R Stat Soc Ser B* 32:115–122
- Stern GA, Gaden A (2015) From science to policy in the Western and Central Canadian Arctic: an Integrated Regional Impact Study (IRIS) of climate change and modernization. ArcticNet, Quebec
- Takhsha M, Nikiéma O, Lucas-Picher P, Laprise R, Hernández-Díaz L, Winger K (2017) Dynamical downscaling with the fifth-generation Canadian Regional Climate Model (CRCM5) over the CORDEX Arctic domain: effect of large-scale spectral nudging and of empirical correction of sea-surface temperature. *Clim Dyn* (**submitted**)
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498. doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- Vincent LA, Zhang X, Bonsal BR, Hogg WD (2002) Homogenization of daily temperatures over Canada. *J Clim* 15(11):1322–1334. doi:[10.1175/1520-0442\(2002\)015<1322:HODTOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2)
- Vincent LA, Wang XL, Milewska EJ, Wan H, Yang F, Swail V (2012) A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J Geophys Res* 117:D18110. doi:[10.1029/2012JD017859](https://doi.org/10.1029/2012JD017859)
- Vincent LA, Zhang X, Brown RD, Feng Y, Mekis E, Milewska EJ, Wan H, Wang XL (2015) Observed trends in Canada's climate and influence of low-frequency variability modes. *J Clim* 28(11):4545–4560. doi:[10.1175/JCLI-D-14-00697.1](https://doi.org/10.1175/JCLI-D-14-00697.1)
- Wan H, Zhang X, Zwiers F, Min SK (2015) Attributing northern high-latitude precipitation change over the period 1966–2005 to human influence. *Clim Dyn* 45:1713–1726. doi:[10.1007/s00382-014-2423-y](https://doi.org/10.1007/s00382-014-2423-y)
- Way RG, Oliva F, Viau AE (2016) Underestimated warming of northern Canada in the Berkeley earth temperature product. *Int J Climatol Early View*. doi:[10.1002/joc.4808](https://doi.org/10.1002/joc.4808)
- Wilson AB, Bromwich DH, Hine KM (2012) Evaluation of polar WRF forecasts on the Arctic system reanalysis domain: 2. Atmospheric hydrologic cycle. *J Geophys Res* 117:D04107. doi:[10.1029/2011JD016765](https://doi.org/10.1029/2011JD016765)
- Ye H, Fetzer EJ, Wong S, Behrangi A, Yang D, Lambrigtson BH (2015) Increasing atmospheric water vapor and higher daily precipitation intensity over northern Eurasia. *Geophys Res Lett* 42:9404–9410. doi:[10.1002/2015GL066104](https://doi.org/10.1002/2015GL066104)
- Zhang X, Alexander L, Hegerl GC, Jones P, Klein Tank AMG, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim Change* 2:851–870. doi:[10.1002/wcc.147](https://doi.org/10.1002/wcc.147)
- Zwiers FW, Alexander LV, Hegerl GC, Kossin JP, Knutson TR, Naveau P, Nicholls N, Schär C, Seneviratne SI, Zhang X, Donat M, Krueger O, Morak S, Murdock TQ, Schnorbus M, Ryabin V, Tebaldi C, Wang XL (2013) Challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events. In: Asrar GR, Hurrell JW (eds) *Climate Science for Serving Society: research, modelling and prediction priorities*. Springer, Dordrecht, p 389 (**ISBN 978-94-007-6691-4**)