

Spatial Models of Adjusted Precipitation for Canada at Varying Time Scales

HEATHER MACDONALD,^a DANIEL W. MCKENNEY,^a XIAOLAN L. WANG,^b JOHN PEDLAR,^a PIA PAPADOPOL,^a
KEVIN LAWRENCE,^a YANG FENG,^b AND MICHAEL F. HUTCHINSON^c

^a Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Sainte Marie, Ontario, Canada

^b Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada

^c Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia

(Manuscript received 28 February 2020, in final form 28 January 2021)

ABSTRACT: This study presents spatial models (i.e., thin-plate spatially continuous spline surfaces) of adjusted precipitation for Canada at daily, pentad (5 day), and monthly time scales from 1900 to 2015. The input data include manual observations from 3346 stations that were adjusted previously to correct for snow water equivalent (SWE) conversion and various gauge-related issues. In addition to the 42 331 models for daily total precipitation and 1392 monthly total precipitation models, 8395 pentad models were developed for the first time, depicting mean precipitation for 73 pentads annually. For much of Canada, mapped precipitation values from this study were higher than those from the corresponding unadjusted models (i.e., models fitted to the unadjusted data), reflecting predominantly the effects of the adjustments to the input data. Error estimates compared favorably to the corresponding unadjusted models. For example, root generalized cross-validation (GCV) estimate (a measure of predictive error) at the daily time scale was 3.6 mm on average for the 1960–2003 period as compared with 3.7 mm for the unadjusted models over the same period. There was a dry bias in the predictions relative to recorded values of between 1% and 6.7% of the average precipitations amounts for all time scales. Mean absolute predictive errors of the daily, pentad, and monthly models were 2.5 mm (52.7%), 0.9 mm (37.4%), and 11.2 mm (19.3%), respectively. In general, the model skill was closely tied to the density of the station network. The current adjusted models are available in grid form at ~2–10-km resolutions.

KEYWORDS: North America; Precipitation; Model evaluation/performance; Model output statistics

1. Introduction

Until 20 years ago, much of Canada's historical precipitation record had not been adjusted for documented shortcomings related to undercatch, wetting losses, unrealistic snow water equivalence conversion (Mekis and Brown 2010), and trace amounts of precipitation (see, e.g., Chvřla et al. 2005). Since then, multiple generations of adjusted precipitation datasets have incorporated such corrections, thereby rehabilitating Canadian precipitation datasets (Mekis and Hogg 1999; Mekis and Vincent 2011; Vincent et al. 2015; Wang et al. 2017). These studies suggest that the adjustments implemented to address multiple measurement and gauge-related issues represent important improvements to the raw data. As outlined in Wang et al. (2017), underestimates of more than 25% are present in Canada's archived historical total precipitation data for stations in northeastern Canada and 10%–15% in much of southern Canada.

The spatial models (thin-plate splines) presented in this study were built using the Adjusted Daily Rainfall and Snowfall dataset, version 2016 (AdjDlyRSv2016), which includes 3346 in situ observing stations with at least one year of data over the 1840–2016 period (Wang et al. 2017). AdjDlyRSv2016 opens the door for new perspectives and analyses, including the development of spatial models and gridded precipitation datasets at multiple time scales including daily, pentad (5 day), and monthly, which is the subject of the current study.

Spatial estimates of precipitation are critical for many users, with optimal time scales often dependent on the type of application. For instance, gridded precipitation datasets (derived from fitted spatial models) are important inputs to water budget/hydrological modeling (e.g., Akinremi et al. 1996; Islam et al. 2019; Persaud et al. 2020), wildfire projections (e.g., Ying et al. 2019), horticultural and plant hardiness studies (e.g., Bloomfield et al. 2019; McKenney et al. 2014), conservation/species distribution mapping (e.g., Leclerc et al. 2019), among others (e.g., Rezsőhazy et al. 2020). While for some applications a monthly time scale is suitable, point-location estimates or gridded data of daily precipitation are typically required for applications such as water budget models. However, a substantial number of stations are required to grid daily precipitation for reasons including the prevalence of zero values, difficulties measuring snow, and short correlation length scales (Hutchinson et al. 2009). For example, 8619 gauge locations were used to map precipitation extreme statistics over Australia (Johnson et al. 2016). In Canada, the gauge network is much more limited and often inadequate to capture spatial precipitation patterns at shorter time scales, particularly north of 55°N where stations may be 500–700 km apart (Wang and Lin 2015). Further challenges in Canada exist in the form of station closures (noted in Mekis and Vincent 2011). The Canadian weather/climate station network peaked in 1990 with almost 900 (manual) stations on average per day; this number decreased to 600 by 2001 and 300 by 2010. Although the current dataset is based on manual stations only, future versions of the adjusted

Corresponding author: Heather MacDonald, heather.macdonald4@canada.ca

precipitation dataset will be augmented with automatic stations (Wang et al. 2017) and/or other suitable station networks.

The purpose of this paper is to introduce new spatial models and gridded datasets of adjusted total daily, mean pentad, and monthly total precipitation for the period of 1900–2015. These spatial models were developed using a 60 arc s (approximately 2 km) and 300 arc s (~10 km) digital elevation model (see Lawrence et al. 2008). We compare metrics from this effort with the models fitted previously to the unadjusted daily and monthly precipitation data (Hutchinson et al. 2009; Hopkinson et al. 2011; McKenney et al. 2006).

2. Data and methods

a. Adjusted precipitation data

For full details on the station data used for the current work see Wang et al. (2017). In brief, AdjDlyRSv2016 contains adjusted daily total precipitation (unit: mm) created using manual observations of daily rainfall and snowfall at 3346 stations as extracted from the National Climate Data Archive of Environment and Climate Change Canada (ECCC). AdjDlyRSv2016 expands on previous efforts to adjust Canadian precipitation records. Specifically, Mekis and Hogg (1999) released a dataset of adjusted precipitation records for 798 stations in which corrections were implemented for wind undercatch and evaporation, gauge-specific wetting losses, trace precipitation amounts, and snow water equivalent (SWE) conversions. Subsequently, Mekis and Brown (2010) used coincident Nipher gauge and snow ruler measurements to develop a spatially varying SWE ratio to replace the static 10:1 SWE ratio used to calculate total precipitation in the ECCC archive. Further, Mekis and Vincent (2011) released a dataset that also incorporated the SWE ratio map of Mekis and Brown (2010).

Published studies have presented trend analyses using a second-generation adjusted precipitation dataset (Vincent et al. 2015), based on 464 station locations (Mekis and Vincent 2011). In this and earlier generation datasets, many stations were not included because of challenges such as short periods of record and the need to establish a baseline for anomaly calculations. However, for producing spatial models, typically more stations are better as in the case of the current study, and each record is linked to the exact observation location. Station joining is simply not needed for producing a gridded dataset, although it is needed for studying the local trend at that particular station location. Focusing on producing gridded precipitation datasets, Wang et al. (2017) applied similar correction procedures to those described above to all stations with appropriate metadata, regardless of station record length. As a result, two versions of the adjusted daily rainfall and snowfall (AdjDlyRS) dataset were produced: versions 2007 and 2016 of AdjDlyRS (i.e., AdjDlyRSv2007 and AdjDlyRSv2016). AdjDlyRSv2007 incorporated 2146 stations and covered 1840–2007; it was used to produce a gridded pentad precipitation dataset using blended gauge and satellite data at 2.5° resolution (Wang and Lin 2015). AdjDlyRSv2016 was developed in the same way as AdjDlyRSv2007, except that additional stations were incorporated.

AdjDlyRSv2016 includes 3346 stations of manual observations of precipitation over Canada, with 908 stations of more than 40 years of data, and 975 stations of 21–40 years of data (see Fig. 1 of Wang et al. 2017). Unlike the adjusted datasets of Mekis and Vincent (2011) and Mekis and Hogg (1999), AdjDlyRSv2016 station records were not joined, but instead the exact new station coordinates for relocated stations were used in the spatial modeling.

The current work makes use of daily total precipitation, pentad mean precipitation, and monthly total precipitation amounts derived from AdjDlyRSv2016. Note that, following standard practice (Wang and Lin 2015), pentad mean precipitation (mm day^{-1}) was calculated from AdjDlyRSv2016, and spatially modeled for each pentad period. In leap years, the pentad beginning on 25 February is a 6-day average of daily precipitation values, consistent with Weaver et al. (2009). The difference in units has implications for the magnitude of errors associated with these products (which are significantly lower than would be expected if 5-day totals were being modeled).

When deriving the monthly data, the monthly total precipitation (mm) was set to missing if there was one or more missing daily value in the month (zero tolerance for missing). The same zero tolerance for missing observations was applied in the calculation of the pentad average precipitation rates (mm day^{-1}).

b. Spatial modeling

This analysis makes use of the Australian National University Spline (ANUSPLIN) package, a widely used spatial modeling tool for climate variables (e.g., Fick and Hijmans 2017; Kriticos et al. 2012; New et al. 2002; McKenney et al. 2011) that has been under development for close to three decades (see Hutchinson and Xu 2013). Specifically, ANUSPLIN consists of a suite of Fortran programs designed to fit thin-plate smoothing splines to noisy climate station data and widely applied globally (e.g., Fick and Hijmans 2017; Kriticos et al. 2012; New et al. 2002). Here we provide only a brief description of thin-plate splines; readers are referred to Wahba (1990) for a mathematical explanation. A typical work-flow process is described in Hutchinson and Xu (2013).

ANUSPLIN employs a generalization of a multivariate linear regression model in which the parametric model is replaced by a nonparametric function. The fitted nonparametric function provides a relationship between the dependent and independent variables. The method optimizes the amount of data smoothing to minimize the predictive error, as measured in this case by the generalized cross validation (GCV). The model is robust to varying underlying spatial statistical models of the data (Hutchinson 1993). Further, the process is automated, making it well suited to processing a large volume of data at the daily time scale. The method has no underlying “range parameter,” makes use of all available stations in the surface fitting, and can adjust automatically to the uneven spatiotemporal density of the station network employed here.

The ANUSPLIN software package produces model diagnostics that give insights into the predictive quality away from the station locations. For our adjusted precipitation models,

the amount of data smoothing was determined by minimizing predictive error as measured by the GCV diagnostic. The GCV is calculated for each value of the smoothing parameter ρ by implicitly removing each data point and calculating the residual from the omitted data point of a surface fitted to all other data points (Wahba 1990). The new version of the software used in this study produces cross-validated estimates for each and every station used in the modeling effort.

Hutchinson (1995) gives a general model for a thin-plate spline function f fitted to n data values z_i at position x_i :

$$z_i = f(x_i) + \varepsilon_i, i = 1 \dots, n \quad (1)$$

in which the x_i refer to a vector of independent variables (in this case, longitude, latitude, and elevation multiplied by a factor of 100). Multiplying elevation by a factor of 100 reflects the generally accepted relative scale of elevation with respect to horizontal coordinates in atmospheric dynamics (Hutchinson et al. 2009; Daley 1991). The ε_i are mean random errors that represent both measurement error as well as model deficiencies, for example, reflecting local effects below the resolution of the model and the source data. For instance, Hutchinson (1998b) illustrated that the impact of slope diminishes at lower resolution, and therefore, to some extent the effects of slope are below the 2- and 10-km resolution of the current models. In terms of independent variables, Hutchinson (1998b) compared thin-plate spline models with varying topographic dependencies, including the inclusion of slope and aspect as predictors. Other unpublished tests using the adjusted precipitation data incorporating 30-yr precipitation averages as additional predictors were unsuccessful due to insufficient station data on which to calculate long-term averages. As a result, we based the current models on predictors (latitude, longitude, and scaled elevation), which are widely available and well tested.

We applied a square root transformation to the station precipitation values, which has been reported to reduce interpolation errors by about 10% (Hutchinson 1995, 1998a). The square root transformation reduces the skewness of the precipitation variable, making smoothing more consistent between small and large precipitation values (Hutchinson 1998b). Tait et al. (2006) have confirmed that the square root transformation can yield a significant reduction in daily precipitation interpolation error.

For daily models, a two-step process is used, in which both a probability of occurrence and a positive precipitation estimate are produced by ANUSPLIN, and then integrated to produce a final interpolated result. The probability of occurrence is estimated using precipitation presence or absence from the recorded station observations. If the probability of occurrence is less than or equal to 0.5, the final interpolated precipitation grid value is set to zero. If the occurrence surface is estimated to be greater than 0.5, the final interpolated value is taken from the positive precipitation estimate.

For large datasets, an approximate spline function is defined in terms of a set of knots. Knots are a subset of stations chosen to match the complexity of the fitted surface (Hutchinson and Xu 2013). Work by Johnson et al. (2016) iteratively increased the number of knots from 2040 to 3570 (of 8074 gauges), which resulted in a steady reduction in error rates as the number/proportion of

knots was increased—particularly for surface residuals as compared with cross-validation residuals. To investigate the impact of knot selection on model errors, we tested the selection of 40%, 80%, and 100% of 1) daily and 2) monthly data points as knots. All knot-related testing was done on data from 1990 to 1999; this decade was selected because of the large number of daily observations during this time period. PROC ANOVA in the SAS software, version 9.4, of SAS System for Windows was used to conduct paired t tests on absolute error rates for December–February (DJF) and June–August (JJA) by knot option.

ANUSPLIN provides Studentized (i.e., normalized) residuals, enhancing user ability to systematically detect outliers (see Kutner et al. 2004). Station values flagged as having high internally Studentized residuals (i.e., “NFLAGS”) were reviewed as part of the quality control process. NFLAGS were used to identify and resolve such problematic values. To assist with evaluation of flagged values, residual lists were linked to data from proximate stations.

Once knot testing was concluded, final models were constructed at the daily, pentad, and monthly time scales and then resolved at 60 arc s or ~2-km (monthly) and 300 arc s or ~10-km (daily, pentad) resolutions using a Canada-wide digital elevation model (see Lawrence et al. 2008).

c. Model assessments

1) STATISTICS OF FITTED SURFACES

Model quality was evaluated initially via the root generalized cross validation (RTGCV), the square root of the GCV described earlier. The RTGCV is an estimate of standard predictive error, although it is a conservative one, because this measure includes data error (Hutchinson and Xu 2013).

In the course of model development, we reviewed the “signal,” a diagnostic metric also produced by ANUSPLIN that ranges between zero and the number of knots, as well as the ratio of signal to number of knots (NK) (the S:NK ratio). The signal is a measure of the complexity of the fitted surface that ranges between a small positive integer and the number of stations used to create the model (Wahba 1990). Models with signals approaching the number of knots suggest an overreliance on available data points. Exact interpolations could suggest a perfect model fit but instead reflect a model that is less reliable, particularly in regions with few stations because models based on fewer data points may have too few stations to disprove estimates.

2) ASSESSMENT OF MODEL PREDICTIONS

The quality of model predictions was assessed for bias and accuracy using the mean error (ME) and the mean absolute error (MAE) of the individual cross-validated residuals (calculated as the estimated value at the station location minus the recorded value) at the station locations. ANUSPLIN outputs the individual station cross-validation (CV) estimates as well as their summary ME and MAE residual statistics to enable an in-depth analysis of model errors.

The quality of final surfaces was evaluated using errors calculated as the difference between the CV values and the recorded station values at 60 “test” stations shown in Fig. 1.



FIG. 1. Locations of the 60 stations used for testing model predictions.

This sample was selected to better reflect the range in latitude, longitude, and elevation across the country as compared with the full set of stations, which tend to be concentrated in southern Canada. The representative sample contained 10 stations in British Columbia; 15 in the prairie provinces of Alberta, Saskatchewan, and Manitoba; 9 in Ontario; 8 in Quebec; 9 in the territories (Yukon, Northwest, and Nunavut); and 9 in Atlantic provinces (New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland). We calculated MAEs and MEs for the entire period from 1960 to 2015 for all time scales and expressed these errors as a percentage of the long-term average precipitation for each season over the same period. This time period is consistent with the one selected for [Hutchinson et al. \(2009\)](#).

Accuracy of daily precipitation model predictions was assessed in two stages; prediction of occurrence (all cases) and positive precipitation (stations where precipitation is predicted to occur). SAS version 9.4 was used to 1) implement a chi-square test on the association between a false negative with respect to the occurrence layer and daily precipitation amount and 2) apply *t* tests to assess whether model biases were significantly different from zero using alpha of 0.05.

To see the spatial pattern of uncertainty (prediction error) in the ANUSPLIN gridded precipitation datasets, for each of the 3346 stations, the individual cross-validated ANUSPLIN predictions for each 5-yr period from 1900 to 2014 (e.g., 1900–04, 1905–09, . . . , 2010–14) were pooled together to calculate the root-mean-square error (RMSE). The resulting RMSEs are expressed as a percentage of the corresponding 5-yr average observed precipitation amount, obtaining the relative RMSEs. Root-mean-square errors (RMSEs) were calculated and mapped at the daily and monthly time scales.

d. Seasonal analysis

For seasonal assessments, the winter season included DJF; the spring season included March, April, and May (MAM); the summer season included JJA; and the autumn season included

September, October, and November (SON). Note that, for the winter season, December of the previous year (e.g., 2007) was grouped with January and February of the following year (e.g., 2008). Pentads 1–12 and 68–73 were classified as winter, pentads 13–30 were grouped into “spring,” pentads 31–49 were considered “summer,” and pentads 50–67 were defined as “autumn,” following [Wang and Lin \(2015\)](#).

e. Comparison with the previous (unadjusted) ANUSPLIN-gridded data

Regional mean series of annual total precipitation (mm) derived from the version 2009 [v2009, unadjusted developed by [Hutchinson et al. \(2009\)](#)] were compared with version 2018 (v2018, adjusted) of ANUSPLIN-gridded daily precipitation datasets over the period of 1950 to 2013. The regional means were gridbox-area weighted. Long-term mean values and Mann–Kendall trend estimates of the corresponding series were also calculated and presented. The results are presented and discussed in [section 3e](#).

Further, the differences between the v2018 (adjusted) and v2009 (unadjusted) gridded values for each grid cell were summed over each day from 1960 to 2010 for each season and then divided by the total unadjusted precipitation over the same period. The percentage differences between the gridded adjusted and unadjusted precipitation were summed, mapped, and discussed in [section 3e](#).

3. Results and discussion

a. Knot testing

We completed knot testing on 3558 Canadian stations at the monthly time scale between 1990 and 1999 inclusive (10 years). Because of the large number of values for the daily time scale, we tested the effects of knot selection from 1990 to 1999 but assessed model predictions for one year only. We chose 1993 as the study year for daily because the number of stations was at its approximate peak at this time. Reviewing model diagnostics, we found the S:NK ratio to be problematic (i.e., >0.8 or <0.2 ; [Hutchinson and Xu 2013](#)) for 11.6% of monthly models using 100% knots. An S:NK ratio in this range indicates lower prediction skill for areas with limited station data. No problematic surfaces were identified for the other knot selections. Because of the number of problematic surfaces using 100% knots, we dropped this option from further consideration.

At the daily time scale, the predicted number of precipitation occurrences was virtually the same for the various knot selections ([Table 1](#), columns for 1993). Knot density did not affect the quality of the occurrence layer. This is consistent with the greater spatial coherence of the occurrence layer relative to positive daily precipitation noted by [Hutchinson et al. \(2009\)](#).

We also examined daily prediction errors by recorded precipitation thresholds ([Fig. 2](#)). Our analysis indicated that on average, the prediction error was greater for the model with 40% positive precipitation knots, particularly for large precipitation amounts. The bias was different depending on the precipitation amount, with positive (wet) biases associated

TABLE 1. Number of precipitation occurrences predicted for 40% and 80% knot selection options for 1993, and monthly MAEs by knot option using DJF and JJA from 1990 to 1999, with *t* statistics and accompanying *p* values for specified paired tests comparing the MAE for the 40% vs 80% knots option.

Knot option	1993		DJF		JJA	
	No precipitation	Precipitation	MAE (mm)	<i>t</i> statistic (probability)	MAE (mm)	<i>t</i> statistic (probability)
40%	411 699	341 296	17.93		15.60	
80%	411 863	341 132	16.54	21.12 ($p < 0.0001$)	14.80	30.75 ($p < 0.0001$)

with precipitation amounts less than 5 mm and negative (dry) biases associated with larger amounts. Statistical testing of the differences in monthly MAEs revealed statistically smaller errors for the models based on 80% knots as compared with 40% knots (testing of DJF and JJA from 1990 to 1999; see Table 1). Because of its greater predictive ability relative to 40% knots, we chose 80% knots for the final full model parameterization.

b. Full model development: Anomaly detection

As described in the methods section, the quality control process included an analysis of observations flagged by ANUSPLIN as outliers. This process helps to identify spatially anomalous observations. Table 2 provides the number of flagged values at each time scale from 1900 to 2015, along with the accompanying flag threshold specifying the probability of exceeding such a value in a normal distribution. Of 39 388 715 daily precipitation occurrence values from 1900 to 2015, there were some 36 288 (0.09%) values flagged as anomalous relative to values at neighboring stations. From 1950 to 2015, there were 27 852 688 daily precipitation values, of which 30 831 (0.11%) were flagged. The higher percentage of anomalous values in the later time period does not imply that records during the later years are necessarily worse but rather that, with a lower density of observations, fewer neighboring stations were available with which to detect spatial anomalies. For instance, there were no flagged cases for the pentad time scale from 1900 to 1910, but this is due to the sparsity of observations during this time period.

Also shown in Table 2, a total of 2604 (0.04%) pentad values were flagged of 6 485 591 observations in total. At a monthly time scale, 805 (0.06%) adjusted precipitation values of 1 273 788 observations over the 1900 to 2015 period were flagged as having a high Studentized residual. Following a review of the list of NFLAGS, we removed station values with Studentized residuals of 3.719 or greater from the modeling process but retained flagged values in testing the quality of model predictions with the set of 60 stations.

c. Number of models

Following a review of flagged values, we developed 42 331 daily precipitation occurrence surfaces and the same number of positive precipitation surfaces for 116 years from 1900 to 2015. In comparison, 8395 pentad and 1392 monthly surfaces were generated over the same period. Figure 3 gives the number of climate stations that were available by year over this period.

d. Model diagnostics

As shown in Table 3, root GCVs were highest at the daily time scale (69.7% of the surface mean) as compared with

pentad (49.9%), and monthly (26.5%) models. Root GCVs were lowest in the autumn across all time scales, consistent with the large well-organized synoptic systems that prevail in this season. In comparison, root GCVs were generally higher in summer months, reflecting greater spatial complexity of convective rainfall compared to frontal precipitation occurring in winter months. However, root GCVs were also elevated at the monthly time scale in winter as a result of high residuals at coastal stations linked to winter storms.

At the daily time scale, root GCVs varied through time, averaging 3.83 mm for models over the 1950 to 2015 period (Table 3, Fig. 4a), or 69.7% of the average surface mean. Errors were highest in the summer (4.43 mm, or 77.5%). The root GCVs at the daily time scale are comparable to statistics presented by Hutchinson et al. (2009), who reported a root GCV of 3.7 mm for the 1961–2003 period. Over this same period, the current daily models exhibit an average root GCV of 3.6 mm. That there is an improvement, even a slight one, is perhaps surprising given that the current study includes only manual observations with enough metadata to make adjustments, compared to the Hutchinson et al. (2009) study, which included all Canadian stations (different manual and automated gauges) without any adjustments for the observation type or gauge specific biases. This may be due to the data preparation phase and the exclusion of possibly anomalous data via Studentized residuals from the model construction phase.

The average pentad root GCV was 1.27 mm from 1950 to 2015 (Table 3, Fig. 4c), or 49.9% of the average pentad surface

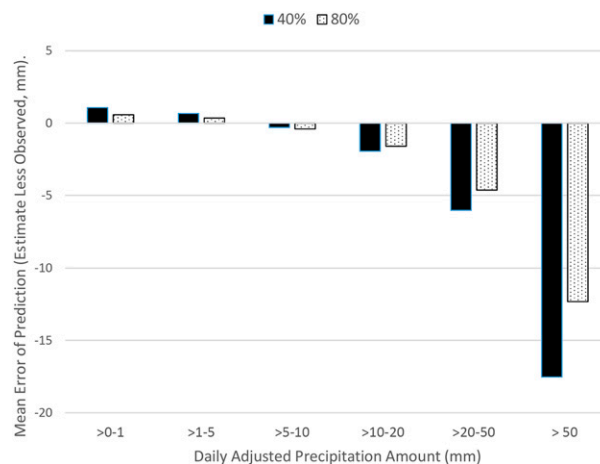


FIG. 2. Daily MEs by knot option (40% vs 80% of data points selected as knots) for 1993.

TABLE 2. Distribution of NFLAGS from ANUSPLIN by time scale and by approximate Studentized residual threshold. The number of observations exceeding the Studentized residual threshold (count) is provided, as well as the percent of the number of cases. Station values with Studentized residuals of greater than 3.719 were excluded from spatial modeling but retained in the assessment of model results.

Approximate Studentized residual threshold	Daily 1900–2015		Daily 1950–2015		Pentad		Monthly	
	Count	%	Count	%	Count	%	Count	%
3.090	20 376	0.05%	17 010	0.06%	1576	0.02%	388	0.03%
3.290	8606	0.02%	7326	0.03%	602	0.01%	210	0.02%
3.481	3867	0.01%	3390	0.01%	244	0.004%	113	0.01%
3.719	3439	0.01%	3105	0.01%	182	0.003%	94	0.01%
Total flags	36 288	0.09%	30 831	0.11%	2604	0.04%	805	0.06%
Cases	39 388 715		27 852 688		6 485 591		1 273 788	

mean. Errors were relatively balanced between the seasons, with an average root GCV for the summer of 1.37 mm (or 55.0%) as compared with 1.38 mm (49.5%) for the winter. Lower root GCVs were evident in the autumn seasons (1.23 mm, or 44.4%).

At a monthly time scale, the average root GCV was 20.6 mm (or 26.5% of the surface mean) from 1950 to 2015 (Table 3, Fig. 4e). Errors were highest in winter (23.6 mm or 28.2%) and lowest in spring (18.1 mm or 26.9%). Previously published studies of historical monthly precipitation (unadjusted) reported an average root GCV of 27% based on data from 1950 to 2000 (McKenney et al. 2006). This previous study was based on automated as well as manual stations for North America, whereas the current study represents only Canadian manual gauges. When we limit our assessment to the 1950–2000 period, the period of greater station numbers, the average root GCVs for the adjusted monthly surfaces averaged 19.6 mm (25.8%), slightly smaller than that of McKenzie et al. (2006).

The S:NK ratios for positive daily precipitation ranged from 0.02 to 0.99 (Fig. 4b), with 462 (1.9%) daily positive precipitation models from 1950 to 2015 exceeding 0.8 and 948 (3.9%) daily models less than 0.2. These problematic cases were associated with days with a smaller number of station records: the mean number of station records associated with models with a ratio of greater than 0.8 or less than 0.2 was 296, as compared with an average of 524 station records on days with favorable S:NK values. Notably, Fig. 4b shows an increase in the range of S:NK ratios in 2010 and 2015; 169 of 1825 pentad surfaces had problematic S:NK ratios over this time period (9.3%) as compared with less than 1.0% from 1950 to 2015. Models with S:NK ratios less than 0.2 and greater than 0.8 should be used with caution. In contrast to daily models, the pentad and monthly time scales produced very few problematic surfaces. The S:NK ratio ranged from 0.17 to 0.86 at the pentad time scale (Fig. 4d), as compared with 0.30–0.86 for the monthly time scale (Fig. 4f).

e. Quality of model predictions using 60 test stations

1) TEMPORAL QUALITY ASSESSMENT

The set of 60 test stations generated 1 240 695 daily observations from 1950 to 2015 (14% of possible daily observations over 66 years were missing). Of these cases, there were 716 084 days with recorded precipitation of greater than zero

(57.7%). Of the 524 611 days with zero recorded precipitation, ANUSPLIN correctly predicted no precipitation 85.2% of the time (Table 4). For stations with positive recorded daily precipitation of less than 1 mm, ANUSPLIN correctly predicted the presence of precipitation 51.7% of the time. The high rate of false negatives (i.e., 48.3%) could be partly due to inter-station differences with respect to recording practices for small amounts of precipitation, but also underlines the difficulty in correctly distinguishing no precipitation from very small amounts of precipitation. Work is under way to test the value of trace flags assigned through reanalysis. In comparison, daily precipitation from 1 to less than 5 mm was correctly identified by ANUSPLIN as having positive precipitation 82.8% of the time. For daily precipitation of 5 mm or greater, ANUSPLIN correctly predicted positive precipitation 95.1% of the time. Cases in which ANUSPLIN predicted no precipitation on days for which significant precipitation amounts were recorded may reflect highly localized precipitation events or even observer errors related to the magnitude and/or timing of the precipitation event (e.g., recorded on the wrong day). Across all daily precipitation amounts, models correctly predicted the occurrence of precipitation 77.2% of the time. The probability of a

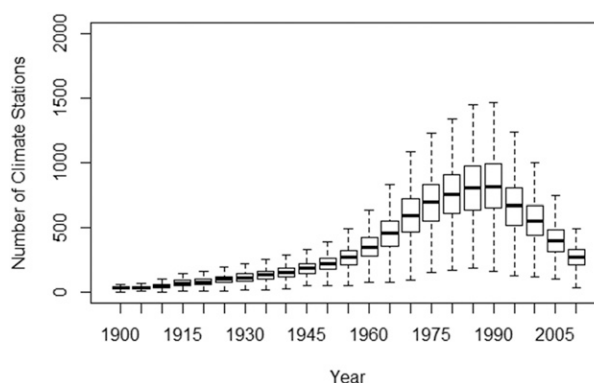


FIG. 3. Number of manual climate stations in Canada available for the current spatial modeling at the daily time scale for the 1900–2015 period. Symbols show the median (center line), quartile (box), and minimum/maximum (whiskers) values for each 5-yr interval (shown at the start of the 5-yr period).

TABLE 3. Seasonal summaries of daily, pentad, and monthly precipitation surface diagnostics from 1950 to 2015. The S:NK ratio is the ratio of the signal of the fitted thin-plate spline to the number of knots. The root-mean GCV is an estimate of the overall standard predictive error of the fitted surface.

Season	Daily positive precipitation				Pentad				Monthly			
	Mean No. of climate stations	Mean S:NK ratio	Root-mean GCV: mm (%)	Root-mean GCV: mm day ⁻¹ (%)	Mean No. of climate stations	Mean S:NK ratio	Root-mean GCV: mm day ⁻¹ (%)	Mean No. of climate stations	Mean S:NK ratio	Root-mean GCV: mm (%)	Mean No. of climate stations	Mean S:NK ratio
Spring	492	0.48	3.40 (69.4%)	1.12 (51.1%)	1133	0.51	1.12 (51.1%)	1095	0.64	18.1 (26.9%)		
Summer	496	0.48	4.43 (77.5%)	1.37 (55.0%)	1156	0.55	1.37 (55.0%)	1127	0.63	20.3 (26.6%)		
Autumn	531	0.50	3.81 (62.3%)	1.23 (44.4%)	1136	0.44	1.23 (44.4%)	1105	0.67	20.2 (24.1%)		
Winter	576	0.42	3.68 (70.0%)	1.38 (49.5%)	1117	0.50	1.38 (49.5%)	1077	0.62	23.6 (28.2%)		
Avg	524	0.47	3.83 (69.7%)	1.27 (49.9%)	1136	0.50	1.27 (49.9%)	1101	0.64	20.6 (26.5%)		

false negative was statistically significantly associated with the amount of daily precipitation ($\chi^2 = 122\,371$; $p < 0.0001$).

For positive daily precipitation events (716 084) recorded at the 60 test stations from 1950 to 2015, the average prediction bias was dry by 0.06 mm, which represents -1.3% of the average daily precipitation (Table 5). For the 1950–2015 period, MAE values averaged 2.5 mm, or 52.7% expressed as a percentage of the average annual daily precipitation over the same period. The highest errors occurred in the summer at 3.3 mm day⁻¹, or 62.9% of the average daily summer precipitation, with similar percentage errors in autumn (2.5 mm day⁻¹, or 51.3%), spring (2.3 mm, or 51.6%), and winter (2.2 mm, or 49.9%). All of these biases were statistically significant, as detailed in Table 5. Despite the improvement in the RTGCV for the daily models compared to unadjusted models, the analysis of model errors identifies significant challenges in modeling precipitation at the daily time scale. As a result, where it is feasible, it may be advantageous to consider the pentad models, which are also expressed in millimeters per day.

The direction of the bias, similar to that reported in Fig. 2, depended on the amount of daily precipitation. Specifically, there was a positive (wet) bias of approximately 1.2 mm for daily precipitation amounts less than 1 mm and a wet bias of 0.7 mm for daily precipitation amounts ranging from 1 to 5 mm. In contrast, there was a dry bias of 2.1 mm for daily precipitation amounts of greater than 5 mm. The bias for small precipitation amounts may reflect recording errors, particularly in cases where a large disparity between the modeled and recorded values are evident. Recorded values with a high Studentized residual were removed from modeling, so the possible effect of measurement errors in ameliorated. However, all recorded values were retained in testing model accuracy. Because these high-disparity cases have a disproportionate effect on error rates, further review of these cases for the next version of the dataset is advisable.

At the pentad time scale, there were 3504 observations generated from the 60 test stations over the 1950–2015 period. The average MAE was 0.87 mm day⁻¹ annually, which represents 37% of the average pentad adjusted precipitation. The highest errors were in the summer (1.07 mm day⁻¹ or 42.2%; Table 6). The average annual bias at the pentad scale was -0.16 mm day⁻¹ (or 6.7%), indicating that pentad model predictions were dry by approximately seven percent compared to the recorded value. All biases reported in Table 6 at the pentad time scale were statistically significantly different from zero.

At the monthly time scale, there were 3447 observations assessed over the 1950–2015 period. In this case, the average MAE value was 11.2 mm (19.3%) annually (Table 6), with seasonal errors ranging from 12.9 mm (23.0%) in winter to 9.8 mm (20.4%) in spring. Model predictions were again slightly dry at the 60 test stations (-1.8 mm or -2.6%), with greatest bias in winter (-2.9 mm or -4.1%) and smallest bias in autumn (-0.5 mm or -0.7%). Seasonal biases were significantly different from zero, with the exception of autumn.

2) SPATIAL QUALITY ASSESSMENT

Monthly errors were highest on the Pacific and Atlantic coasts. For example, station 1026270 (Port Hardy, British Columbia, on

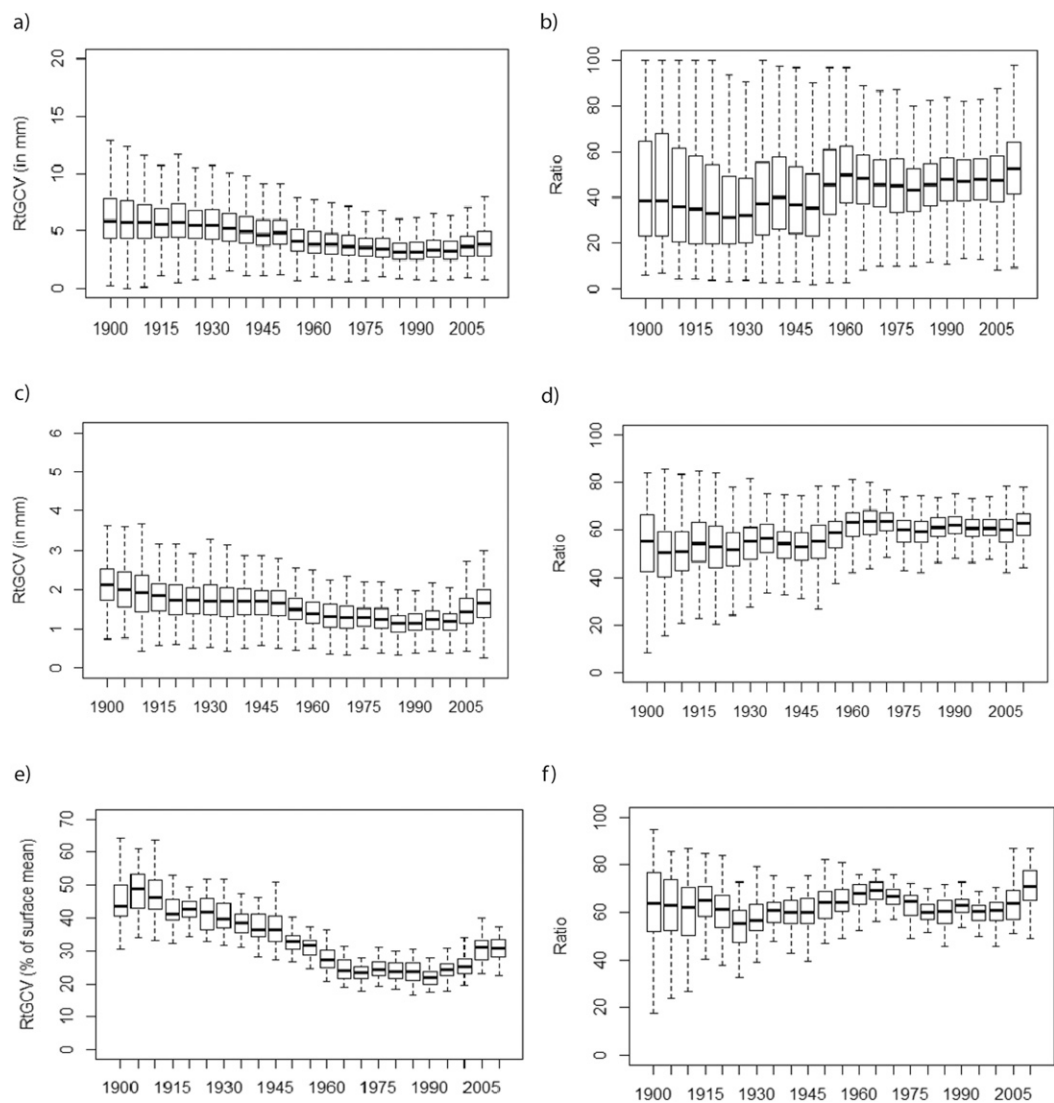


FIG. 4. Root GCV and S:NK ratio for 5-yr intervals over the 1900–2015 period at the (a),(b) daily, (c),(d, pentad, and (e),(f) monthly time scales (shown at the start of the 5-yr period).

Vancouver Island, one of the 60 test stations) had over 200 mm of monthly precipitation recorded in winter on average, with an absolute cross-validation error of over 50 mm, which is approximately 25%. Previous studies have noted that higher error is related to higher precipitation amounts during winter storms along

the Pacific and Atlantic coasts and limited station numbers (McKenney et al. 2006). However, note also that, even with the aforementioned adjustments (Wang et al. 2017), some measurement errors can remain and impact the quality of precipitation data particularly in windy conditions (Mahmood et al. 2017).

TABLE 4. ANUSPLIN predictive accuracy of daily precipitation occurrence by amount of recorded adjusted daily precipitation (1950–2015, using 60 test stations).

Recorded precipitation amount (mm day ⁻¹)	Correct predictions (%)	False positives (%)	False negatives (%)	N
0	446 867 (85.2%)	77 744 (14.8%)	—	524 611
>0–1	171 362 (51.7%)	—	160 053 (48.3%)	331 415
>1–5	180 122 (82.8%)	—	37 337 (17.2%)	217 459
>5	158 943 (95.1%)	—	8267 (4.9%)	167 210
Total	957 294 (77.2%)	77 744 (6.3%)	205 657 (16.6%)	1 240 695

TABLE 5. Seasonal and annual summaries of prediction errors from 1950 to 2015 for daily precipitation recorded at the 60 test stations shown in Fig. 1. The N denotes the number of daily precipitation values; bias denotes the mean difference between the modeled surface values and observed values, calculated as CV estimate less recorded; and t -statistic values reflect a test of whether the bias is statistically different from zero.

Season	N	Avg recorded precipitation (mm day ⁻¹)	Bias: mm day ⁻¹ (%)	t statistic for $H_{0\text{Bias}} = 0$ (p value)	MAE: mm day ⁻¹ (%)
Spring	166 608	4.46	-0.07 (-1.6%)	-6.4 (<0.0001)	2.3 (51.6%)
Summer	162 410	5.25	-0.19 (-3.6%)	-12.8 (<0.0001)	3.3 (62.9%)
Autumn	179 133	4.88	-0.04 (-0.8%)	-3.6 (<0.0004)	2.5 (51.3%)
Winter	207 933	4.41	0.03 (0.7%)	2.9 (<0.004)	2.2 (49.9%)
Annual	716 084	4.75	-0.06 (-1.3%)	-10.6 (<0.0001)	2.5 (52.7%)

Figure 5 shows the relative RMSEs in the ANUSPLIN models for two 5-yr time periods (including 1985–89, the period of the highest number of stations). Clearly, the uncertainty is much lower for the monthly models than for the corresponding daily models and is much lower in the south than in the north for the same time scale (Figs. 5b,d,f). Particularly, there is a high degree of uncertainty in the north, including a complete lack of stations in this region during the early part of the twentieth century. As shown in Fig. 6, there is a clear negative relationship between the relative RMSEs and the number of stations with observed data: the higher the number of stations is, the smaller is the model prediction error. Such a relationship can also be seen in the corresponding RMSEs (not shown). The regional average uncertainty (RMSEs and relative RMSEs) is much higher in the early period; it reached the lowest level around 1970 and stayed at the lowest level during 1970–2005 when the total number of stations over Canada is over 1000, with over 100 stations in the north (Figs. 6a,c). Because of a decrease in station numbers, the uncertainty in the recent decade is comparable to that of the 1950s (Fig. 6c). The uncertainty is also higher in the cold seasons (DJF and MAM) than in the warm seasons, especially in the north (Fig. 6a). The regional averages of the relative RMSEs of the daily models (not shown) lead to similar conclusions. Because of the complete lack of stations in this region during the early part of the twentieth century, we have not released gridded precipitation estimates for the region north of 55°N prior to 1935 and for the region north of 70°N prior to 1950.

f. Results of comparison of adjusted and unadjusted gridded values

As shown in Fig. 7, v2009 (unadjusted) has lower total precipitation amount and slower increasing trends than v2018

(adjusted). The long-term mean annual total precipitation averaged over Canada, north Canada, and south Canada is 72.9 mm (12.2%), 67.1 mm (16.4%), and 81.3 mm (9.3%) lower in v2009 than in v2018, respectively. The 1950–2013 trend of the regional mean series for Canada, north Canada, and south Canada is estimated to be 0.33, 0.57, and 0.15 mm yr⁻¹ in v2009, respectively, and 1.20, 1.30, and 0.90 mm yr⁻¹ in v2018.

Maps of long-term (1960–2010) fractional mean differences between adjusted and unadjusted gridded daily precipitation amounts for each season are shown in Fig. 8. The overall effect of the adjustments was to increase precipitation amounts, with the exception of the snow water equivalent adjustments, which caused slightly negative net adjustments in some areas of British Columbia and Alberta. These effects are consistent with the effect of previously adjusted datasets (e.g., Mekis and Vincent 2011). In percentage terms, the effect of the adjustments relative to unadjusted gridded values is greatest in northern Canada, where the effect of the adjustments in winter was to double the climatological (1960–2010) mean values of the gridded daily precipitation amounts compared to earlier unadjusted precipitation grids.

Importantly, these findings reflect those of Wang et al. (2017), who also reported that the greatest relative increases due to precipitation adjustments were found in northern Canada. The comparisons in Wang et al. (2017) present net effects of greater than 25% in northern Canada and 5%–25% at most stations in southern Canada. Wang et al. (2017) also noted that the adjustments resulted in decreased precipitation totals for some stations in the Rocky Mountains. Our analysis differs from that of Wang et al. (2017) in that we are comparing gridded estimates rather than values at specific station locations and we used the unadjusted daily precipitation grids of

TABLE 6. Seasonal and annual summaries of prediction errors for pentad and monthly precipitation surfaces among a set of 60 test stations from 1950 to 2015. The N denotes the number of daily precipitation values, bias denotes the mean difference between surface values and observed values, and t statistics and p values for the null hypothesis denote that the bias equaled zero. The annual value is a mean of the four seasonal values.

Season	N	Pentad			Monthly			
		Bias: mm day ⁻¹ (%)	t statistic (p value)	MAE: mm day ⁻¹ (%)	N	Bias: mm (%)	t statistics (p value)	MAE: mm (%)
Spring	3462	-0.16 (-8.1%)	-17.11 (<0.0001)	0.73 (37.4%)	3454	-2.1 (-3.6%)	-8.04 (<0.0001)	9.8 (20.4%)
Summer	3446	-0.19 (-7.4%)	-21.34 (<0.0001)	1.07 (42.2%)	3439	-1.7 (-2.1%)	-6.32 (<0.0001)	11.2 (17.3%)
Autumn	3446	-0.11 (-4.5%)	-11.23 (<0.0001)	0.84 (33.9%)	3444	-0.5 (-0.7%)	-1.82 (0.07)	10.7 (16.6%)
Winter	3460	-0.17 (-7.4%)	-13.9 (<0.0001)	0.84 (35.7%)	3452	-2.9 (-4.1%)	-8.07 (<0.0001)	12.9 (23.0%)
Annual	3504	-0.16 (-6.7%)	-20.904 (<0.0001)	0.87 (37.4%)	3447	-1.8 (-2.6%)	-7.95 (<0.0001)	11.2 (19.3%)

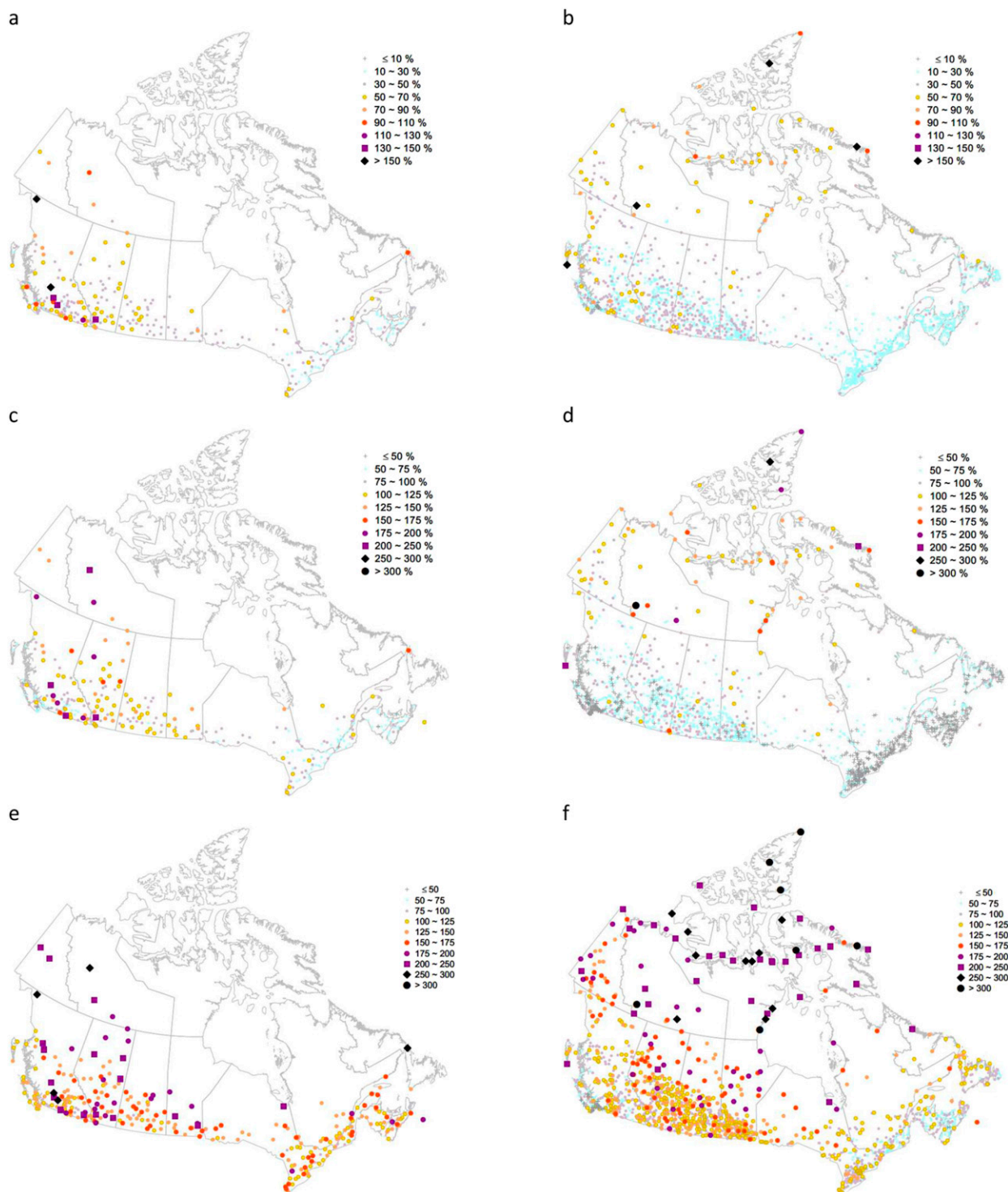


FIG. 5. Relative RMSEs (in percentage of the average observed amount over the indicated period) of the (a),(b) monthly; (c),(d) pentad; and (e),(f) daily precipitation models at the individual stations for (left) 1925–29 and (right) 1985–89. Note the different plotting scales for monthly, pentad, and daily models.

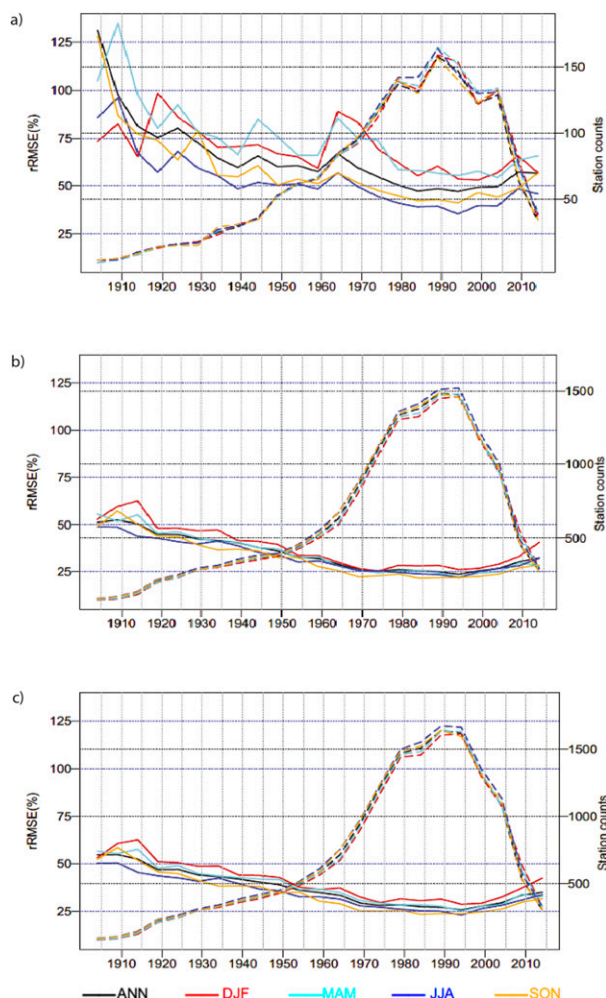


FIG. 6. Time series of regional averages of relative RMSEs of the ANUSPLIN monthly precipitation models (solid lines; left axis) and the total number of manual stations with observations (dashed lines; right axis) for the indicated seasons (ANN for annual). Values are shown at the middle of the 5-yr period.

Hutchinson et al. (2009; see also Hopkinson et al. (2011), which involved both automated and manual stations and thus more stations than those in the Wang et al. (2017) study.

In summary, the differences between v2009 and v2018 are primarily results of the adjustments for known problems as discussed by Wang et al. (2017), although different versions of ANUSPLIN modeling process may have resulted in small differences in the results.

4. Usage notes

The current spatial models and gridded datasets can be used as inputs to a wide variety of ecological, agricultural, hydrological, and economic applications. Other potential applications include comparison with precipitation estimates from satellite data (Lin and Wang 2011). Of note is that users typically interested in daily models may want to consider the possible utility of the pentad time scale given its higher predictive accuracy.

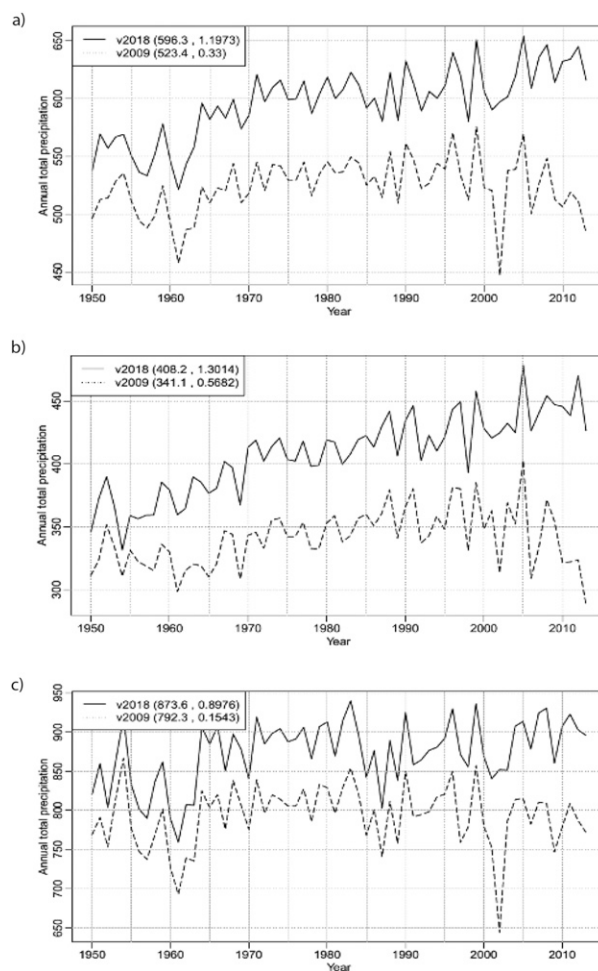


FIG. 7. Comparison of regional mean series of annual total precipitation (mm) derived from v2009 (unadjusted) and v2018 (adjusted) of ANUSPLIN-gridded daily precipitation datasets over the common period 1950–2013 for (a) Canada, (b) north Canada—north of 55°N, and (c) south Canada—south of 55°N. The regional means are gridbox-area weighted. The numbers in the parentheses are the long-term mean values and Mann-Kendall trend estimates of the corresponding series.

Other considerations for users could include that unlike other publicly available gridded adjusted precipitation datasets (as described in Vincent et al. 2015), our station data have not been homogenized for the purposes of trend analysis. ECCC's Canadian Gridded Temperature and Precipitation Anomalies (CANGRD) dataset of gridded precipitation anomalies is built for broadscale trend analysis using around 460 long-term series of adjusted precipitation records from manual stations (with station joining). However, ECCC could no longer update this dataset due to the closure or automation of the majority of these stations. Currently, ECCC is working on joining manual and automated station data records to develop a long-term adjusted and homogenized precipitation dataset for Canada, which will be used as input to produce a new version of the CANGRD dataset. ECCC is also working on comparison of

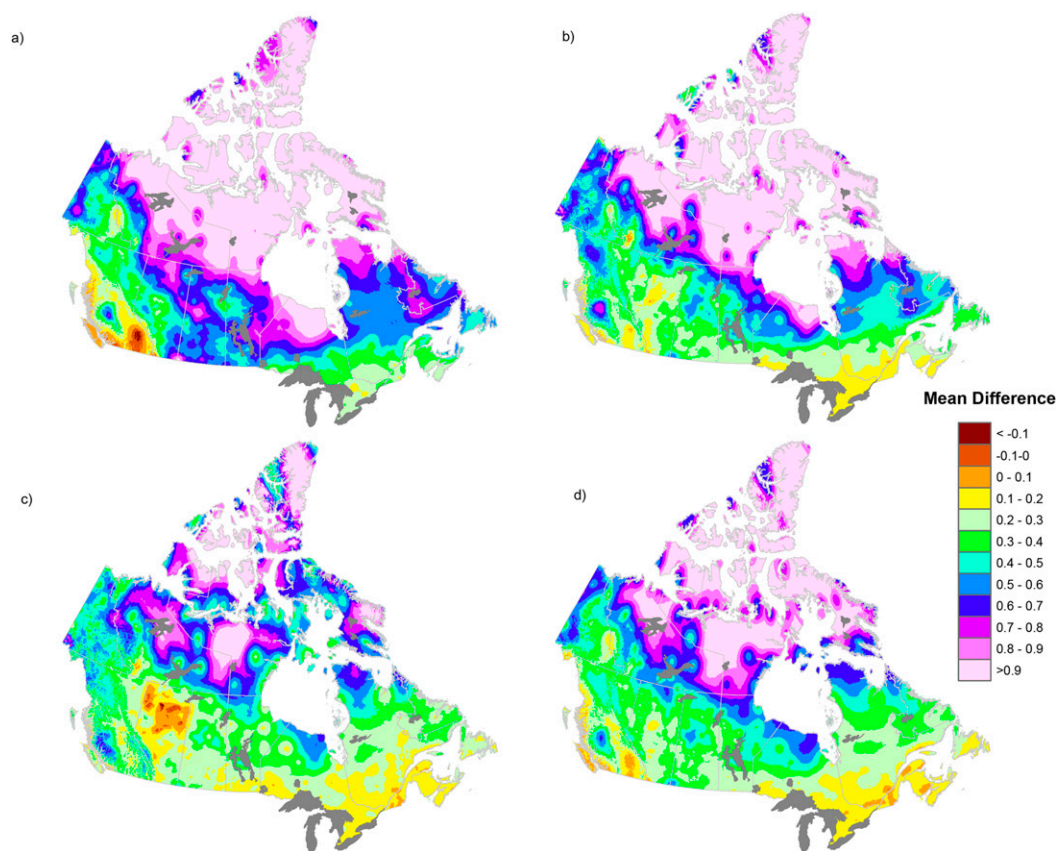


FIG. 8. Maps of long-term (1960–2010) fractional mean differences between adjusted and unadjusted gridded precipitation amounts for (a) winter, (b) spring, (c) summer, and (d) autumn. The differences were calculated as (adjusted – unadjusted)/unadjusted. Seasonal precipitation values (adjusted and unadjusted) were obtained by summing daily values.

the CANGRD method with other gridding methods and will report the results in a separate study. That said, the current models will be of interest to users that desire 1) actual adjusted precipitation estimates as opposed to anomalies, 2) precipitation estimates for any location desired, and 3) gridded data at finer resolutions (2–10 km vs CANGRD’s 50-km resolution).

5. Conclusions

The spatial models of adjusted precipitation reported here represent a significant improvement over previous products, making them more appropriate for applications such as water balance and runoff modeling. The adjustments reflect substantial effort over the past 20 years to reduce inconsistencies and inhomogeneity and reduce bias for time series analysis (Mekis and Hogg 1999; Mekis and Vincent 2011; Wang et al. 2017). More specifically, the adjustments diminish significant underestimation of the precipitation amount and of the long-term trend therein (Fig. 7). The adjustment is particularly significant in the Canadian Arctic, where frequent trace occurrences are common, which “makes the accumulation of these miniscule amounts significant to the annual total precipitation and the water balance of the region” (Mekis and Hogg 1999, p. 60).

Studies prior to Wang et al. (2017) also found that adjustments tended to cause an increase in gauge-measured precipitation. Yang et al. (2005) reported an increase of 80%–120% due to adjustments in winter precipitation. Mekis and Vincent (2011) concluded that the impact of their adjustments was largest in the far north where the frequency of trace measurements was greatest. For example, the cumulative impact of all adjustments was a 48.8% increase at Resolute, Nunavut (Mekis and Vincent 2011). However, similar to our findings, Mekis and Vincent (2011) reported adjustment-related precipitation reductions in western mountainous regions, which they attributed to the lower adjustment required for snowfall in this area.

For the most part, these spatial models exhibited error rates that were in the range of previous studies (e.g., Hutchinson et al. 2009; McKenney et al. 2006). The performance of the current daily models compares favorably to those developed by Hutchinson et al. (2009), which is perhaps surprising given that the current study includes only manual observations with enough metadata to make adjustments, compared to the Hutchinson et al. (2009) study, which included all Canadian stations (automated and manual). We have put considerable effort into describing model quality. Even though the current

adjustments address important measurement issues, measurement error is a source of noise in the spatial models, which ultimately affects model performance (Mahmood et al. 2017). The current models exhibit a slight dry bias. With respect to daily precipitation occurrence, daily models were more likely to predict an absence of precipitation on days with small amounts of precipitation. For positive precipitation, the direction of the bias in positive precipitation surfaces is related to the recorded precipitation amounts, with wet biases for small precipitation amounts, and dry biases for recorded precipitation of greater than 5 mm. At the monthly time scale, there is a slight dry bias as well, particularly for the winter months (average ME of -4.1%). The bias is greater for large precipitation amounts in part because neighboring stations not on the coast have highly variable precipitation amounts particularly during winter storms, making estimates from neighboring stations challenging. The largest monthly errors occurred in the winter for stations on the Pacific and Atlantic coasts, suggesting that adding ocean proximity or distance to large water bodies as an independent variable could improve estimates. This is a consideration for future research although previous (unpublished) efforts suggest that station numbers will again be a limiting challenge.

While these models are now available, users should be cognizant of the accuracy of spatial models/surfaces for their particular applications. Mekis and Hogg (1999) suggest 25% as an upper limit for monthly precipitation error rates. Seasonally and annually summarized error rates are below this limit and root GCVs average 25% for monthly models from 1950 to 2015. Greater model errors post 2000 reflect the declining number of in situ stations (particularly manual stations). The magnitude of errors was clearly tied to the density of the station network and ongoing efforts are aimed at adjusting automated station records in order to increase the number of available stations. While the current dataset does include some station networks other than those that have data in the ECCC digital archive, additional efforts to assess and incorporate other station networks (such as agricultural or fire weather monitoring networks) may also be beneficial—despite the fact that these networks may not operate during all seasons and may have different operating standards.

Future work is planned to produce adjusted North American precipitation gridded datasets. Precipitation station values from the United States have not been adjusted due to shortcomings as considered here. As a result, our previous efforts to incorporate stations from the United States produced a distinct border effect. Current work is focused on collaborations to incorporate adjustments into U.S. station values. Future work could also focus on integration of satellite data (e.g., Lin and Wang 2011; Wang and Lin 2015). Successful incorporation of satellite data requires appropriate calibration from gauges, including the nontrivial task of generating totals or aggregate amounts over the desired time scale (e.g., daily, pentad, monthly). For this reason, Bisselink et al. (2016) conclude that it is difficult to construct hydrological models in ungauged or inadequately gauged areas. Another area for future work is adding appropriately adjusted automatic gauge data to the input dataset, which will improve station coverage in locations

where manual gauges were replaced with automatic ones. In the interim, testing will be undertaken on the usefulness of the pentad time scales for hydrological models.

Acknowledgments. We acknowledge the assistance provided by Kaitlin de Boer in producing some of the maps/plots for the current paper. Funding to support these models to the Geospatial Tools and Economic Analysis Group of Natural Resources Canada/Canadian Forest Service was provided in part by Environment and Climate Change Canada.

REFERENCES

- Akinremi, O. O., S. M. McGinn, and A. G. Barr, 1996: Evaluation of the Palmer drought index on the Canadian prairies. *J. Climate*, **9**, 897–905, [https://doi.org/10.1175/1520-0442\(1996\)009<0897:EOTPDI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0897:EOTPDI>2.0.CO;2).
- Bisselink, B., M. Zambrano-Bigiarini, P. Burek, and A. de Roo, 2016: Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. *J. Hydrol. Reg. Stud.*, **8**, 112–129, <https://doi.org/10.1016/j.ejrh.2016.09.003>.
- Bloomfield, K. J., and Coauthors, 2019: The validity of optimal leaf traits modelled on environmental conditions. *New Phytol.*, **221**, 1409–1423, <https://doi.org/10.1111/nph.15495>.
- Chvřla, B., B. Sevruck, and M. Ondrás, 2005: The wind-induced loss of thunderstorm precipitation measurements. *Atmos. Res.*, **77**, 29–38, <https://doi.org/10.1016/j.atmosres.2004.11.032>.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, <https://doi.org/10.1002/joc.5086>.
- Hopkinson, R. F., D. W. McKenney, E. J. Milewska, M. F. Hutchinson, P. Papadopol, and A. L. A. Vincent, 2011: Impact of aligning climatological day on gridding daily maximum–minimum temperature and precipitation over Canada. *J. Appl. Meteor. Climatol.*, **50**, 1654–1665, <https://doi.org/10.1175/2011JAMC2684.1>.
- , M. F. Hutchinson, D. W. McKenney, E. J. Milewska, and P. Papadopol, 2012: Optimizing input data for gridding climate normals for Canada. *J. Appl. Meteor. Climatol.*, **51**, 1508–1518, <https://doi.org/10.1175/JAMC-D-12-018.1>.
- Hutchinson, M. F., 1993: On thin plate splines and kriging, Computing Science and Statistics. Proc. 25th Symp. on the Interface, San Diego, CA, Interface Foundation of North America, 55–62.
- , 1995: Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.*, **9**, 385–403, <https://doi.org/10.1080/02693799508902045>.
- , 1998a: Interpolation of rainfall data with thin plate smoothing splines: I. Two dimensional smoothing of data with short range correlation. *J. Geogr. Inf. Decis. Anal.*, **2**, 153–167.
- , 1998b: Interpolation of rainfall data with thin plate smoothing splines: II. Analysis of topographic dependence. *J. Geogr. Inf. Decis. Anal.*, **2**, 168–185.
- , and T. Xu, 2013: ANUSPLIN version 4.4 user guide. Australian National University, Fenner School of Environment and Society Doc., 55 pp., <https://fennerschool.anu.edu.au/files/anusplin44.pdf>.
- , D. W. McKenney, K. Lawrence, J. H. Pedlar, R. F. Hopkinson, E. Milewska, and P. Papadopol, 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, <https://doi.org/10.1175/2008JAMC1979.1>.
- Islam, U., C. L. Curry, S. J. Déry, and F. W. Zwiers, 2019: Quantifying projected changes in runoff variability and flow regimes of the Fraser River basin, British Columbia. *Hydrol. Earth Syst. Sci.*, **23**, 811–828, <https://doi.org/10.5194/hess-23-811-2019>.
- Johnson, F., M. F. Hutchinson, C. The, C. Beesley, and J. Green, 2016: Topographic relationships for design rainfalls over Australia. *J. Hydrol.*, **533**, 439–451, <https://doi.org/10.1016/j.jhydrol.2015.12.035>.
- Kriticos, D. J., B. L. Webber, A. Leriche, N. Ota, I. Macadam, J. Bathols, and J. K. Scott, 2012: CliMond: Global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol. Evol.*, **3**, 53–64, <https://doi.org/10.1111/j.2041-210X.2011.00134.x>.
- Kutner, M. H., J. Neter, C. J. Nachtsheim, and W. Li, 2004: *Applied Linear Statistical Models*. 5th ed. McGraw-Hill Irwin, 1415 pp.
- Lawrence, K. M., M. F. Hutchinson, and D. W. McKenney, 2008: Multi-scale digital elevation models for Canada. Natural Resources Canada, Great Lakes Forestry Centre Frontline Tech. Note 109, 4 pp., <https://d1ied5g1xfqpx8.cloudfront.net/pdfs/31499.pdf>.
- Leclerc, M., A. Tarroux, P. Fauchald, A. Stien, T. Tveraa, and M. St-Laurent, 2019: Effects of human-induced disturbances and weather on herbivore movement. *J. Mammal.*, **100**, 1490–1500, <https://doi.org/10.1093/jmammal/gyz101>.
- Lin, A., and X. L. Wang, 2011: An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada. *J. Geophys. Res.*, **116**, D21111, <https://doi.org/10.1029/2011JD016359>.
- Mahmood, R., and Coauthors, 2017: Mesonets: Mesoscale weather and climate observations for the United States. *Bull. Amer. Meteor. Soc.*, **98**, 1349–1361, <https://doi.org/10.1175/BAMS-D-15-00258.1>.
- McKenney, D. W., J. H. Pedlar, P. Papadopol, and M. F. Hutchinson, 2006: The development of 1901–2000 historical monthly climate models for Canada and the United States. *Agric. For. Meteorol.*, **138**, 69–81, <https://doi.org/10.1016/j.agrformet.2006.03.012>.
- , and Coauthors, 2011: Customized spatial climate models for North America. *Bull. Amer. Meteor. Soc.*, **92**, 1611–1622, <https://doi.org/10.1175/2011BAMS3132.1>.
- , J. H. Pedlar, K. Lawrence, P. Papadopol, K. Campbell, and M. F. Hutchinson, 2014: Change and evolution in the plant hardiness zones of Canada. *BioScience*, **64**, 341–350, <https://doi.org/10.1093/biosci/biu016>.
- Mekis, É., and W. D. Hogg, 1999: Rehabilitation and analysis of Canadian daily precipitation time series. *Atmos.–Ocean*, **37**, 53–85, <https://doi.org/10.1080/07055900.1999.9649621>.
- , and R. Brown, 2010: Derivation of an adjustment factor map for the estimation of the water equivalent of snowfall from ruler measurements in Canada. *Atmos.–Ocean*, **48**, 284–293, <https://doi.org/10.3137/AO1104.2010>.
- , and L. A. Vincent, 2011: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmos.–Ocean*, **49**, 163–177, <https://doi.org/10.1080/07055900.2011.583910>.
- New, M., D. Lister, M. Hulme, and I. Makin, 2002: A high-resolution data set of surface climate over global land areas. *Climate Res.*, **21**, 1–25, <https://doi.org/10.3354/cr021001>.
- Persaud, B. D., P. H. Whitfield, W. L. Quinton, and L. E. Stone, 2020: Evaluating the suitability of three gridded-datasets and their impacts on hydrological simulation at Scotty Creek in the southern Northwest Territories, Canada. *Hydrol. Processes*, **34**, 898–913, <https://doi.org/10.1002/hyp.13663>.
- Rezsöházy, J., H. Goosse, J. Guiot, F. Gennaretti, E. Boucher, F. André, and M. Jonard, 2020: Application and evaluation of the dendroclimatic process-based model MAIDEN during the last century in Canada and Europe. *Climate Past*, **16**, 1043–1059, <https://doi.org/10.5194/cp-16-1043-2020>.
- Tait, A., R. Henderson, R. Turner, and X. Zheng, 2006: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *Int. J. Climatol.*, **26**, 2097–2115, <https://doi.org/10.1002/joc.1350>.
- Vincent, L. A., X. Zhang, R. D. Brown, Y. Feng, E. Mekis, E. J. Milewska, H. Wan, and X. L. Wang, 2015: Observed trends in Canada's climate and influence of low-frequency variability modes. *J. Climate*, **28**, 4545–4560, <https://doi.org/10.1175/JCLI-D-14-00697.1>.
- Wahba, G., 1990: *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, 161 pp., <https://doi.org/10.1137/1.9781611970128>.
- Wang, X. L., and A. Lin, 2015: An algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale. *J. Geophys. Res. Atmos.*, **120**, 3728–3744, <https://doi.org/10.1002/2014JD022788>.
- , H. Xu, B. Qian, Y. Feng, and E. Mekis, 2017: Adjusted daily rainfall and snowfall data for Canada. *Atmos.–Ocean*, **55**, 155–168, <https://doi.org/10.1080/07055900.2017.1342163>.
- Weaver, S. J., A. Ruiz-Barradas, and S. Nigam, 2009: Pentad evolution of the 1988 drought and 1993 flood over the Great Plains: An NARR perspective on the atmospheric and terrestrial water balance. *J. Climate*, **22**, 5366–5384, <https://doi.org/10.1175/2009JCLI2684.1>.
- Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.*, **32**, L19501, <https://doi.org/10.1029/2005GL024057>.
- Ying, H., Y. Shan, H. Zhang, T. Yuan, W. Rihan, and G. Deng, 2019: The effect of snow depth on spring wildfires on the Hulunbuir from 2001–2018 based on MODIS. *Remote Sens.*, **11**, 321, <https://doi.org/10.3390/rs11030321>.