Validation and inter-comparison of three methodologies for interpolating daily precipitation and temperature across Canada

Nathaniel K. Newlands1*†, Andrew Davidson2, Allan Howard3 and Harvey Hill4

1 Environmental Health Research Branch, Agriculture and Agri-Food Canada, 5403 - 1 Avenue South, Lethbridge, AB T1J 4B1, Canada
2 National Land and Water Information Service, 960 Carling Avenue, Agriculture and Agri-Food Canada, Ottawa, ON K1A OC6, Canada
3 National Agro-climate Information Service, Agriculture and Agri-Food Canada, 1800 Hamilton St., Regina, SK S4P 4L2, Canada
4 Climate Impacts and Adaptation, Agriculture and Agri-Food Canada, 1011, Innovation Blvd, Saskatoon, SK S7V 1B7, Canada

SUMMARY

The use of daily climate data in agriculture has increased considerably over the past two decades due to the rapid development of information technology and the need to better assess impacts and risks from extreme weather and accelerating climate change. While daily station data is now regularly used as an input to biophysical and biogeochemical models for the study of climate, agriculture, and forestry, questions still remain on the level of uncertainty in using daily data, especially for predictions made by spatial interpolation models. We evaluate the precision of three models (i.e., spline, weighted-truncated Gaussian filter, and hybrid inverse-distance/natural-neighbor) for interpolating daily precipitation and temperature at 10 km across the Canadian landmass south of 60° latitude (encompassing Canada’s agricultural region). We compute daily, weekly, and monthly-aggregated bias and root-mean-square (RMSE) validation statistics, examining how error varies with orography and topography, and proximity to large water. Our findings show the best approach for interpolating daily temperature and precipitation across Canada requires a mixed-model/Bayesian approach. Further application of interpolation methods that consider non-stationary spatial covariance, alongside measurement of spatial correlation range would aid considerably in reducing interpolation prediction uncertainty. Copyright © 2010 Crown in Right of Canada.

KEY WORDS: agro-climatic interpolation; daily; precipitation; temperature; validation

1. INTRODUCTION

The interpolation of daily temperature and precipitation data is used in managing urban and agricultural landscapes as an input to decision-making tools, biophysical models, and helping to further climate change research (Changnon and Kunkel, 1999; Changnon, 2004). Interpolation models provide information where observational data is unavailable and provide a quantifiable measure of precision in
estimation and prediction (Held, 2005; Morss et al., 2005; O'Connor et al., 2005). The comparison of models serves to elucidate uncertainties in model predictions themselves. These models also provide continuous information in time and space at an appropriate scale for input to ecosystem models (Wang et al., 2005), and to re-construct, predict, and forecast climate trends and feedbacks (Weaver and Zwiers, 2000; Marnane et al., 2002; Held, 2005; Lobell et al., 2006). Reliable daily interpolated temperature and precipitation is required, for example, in quantifying atmospheric water vapor pressure (WVP) (Yin, 1999), in assessing crop or stand productivity (Fries et al., 2000) and the risk of forest fires (Bond-Lamberty et al., 2007), and in predicting soil carbon sequestration potential (Price et al., 1999). Considerable effort has been focused on climate interpolation across large geographical extents based on reference historical meteorological data interpolated across a range of spatial resolution (20 arcsec to 0.5 deg.) for Canada (Hopkinson, CICS, 2001), United States (Thornton et al., 1997), and South America (Liebmann and Allured, 2005). Effort has also been focused to span temporal resolution, for example at the monthly scale across 1961–1990 (New et al., 2000; New et al., 2002), and daily scale (Piper and Stewart, 1996). While the testing of interpolated models and climate surfaces has increased, so too has their application at finer temporal (e.g., daily, hourly) and spatial (e.g., <10 km) resolution. Interpolated data can be used in assessing agricultural risk/crop insurance claims, the quality, supply, and sustainability of storm- and ground-water flows, the hydrology of forests, flood-plains, and watersheds, and the operation, planning and design of urban networked infrastructure. In the agricultural sector, the use of daily interpolated climate offers information to spur new insights on spatial and temporal patterns in crop growth and productivity, soil erosion and leaching, greenhouse gas emission, flooding and drought impacts on farming operations, and production. Currently, there are a wide variety of end-users for interpolated data. End-users range from private producers and “aggregators” in the emerging carbon credit industry and growing bio-processing/bio-product industries to public agencies at the local, regional, provincial and federal level, integrated watershed, land-use and disaster management planners, ecological economists, and statisticians/modelers. Statistical indicators and measures drawn from daily, interpolation climate information, for example, can be directly used to generate risk quantile functions for better assessing the relative impacts of extreme events. Such information enhances the reliability of statistical information for decision makers engaged in identifying the best ways to mitigate against harmful impacts, based on when and where extreme conditions may occur.

There are many different methods for interpolating temperature and precipitation observations. In part, the diversity of methods reflects the multivariate nature of climate that varies with elevation, slope, and terrain transition, proximity to water, vegetation/land cover, and wind conditions. Spatial climate patterns are most affected by terrain and water bodies at spatial scales less than 10 km, primarily through the direct effects of elevation, terrain-induced climate transitions, cold air drainage and inversions, and coastal effects (Daly, 2006). Functional and statistical approaches include: (a) universal kriging and co-kriging (Chiles and Delfiner, 1990; Host et al., 1995), (b) spline-fitting (Wahba, 1990; Hutchinson, 1995; Hutchinson, 1998a, b), (c) distributional approaches (Thornton and Running, 1999; Thornton et al., 1997; Thornton et al., 2000), (d) simpler methods (e.g. nearest neighbor assignment, NN, inverse-distance weighting, IDW), (e) “hybrid” combinations of these simpler methods (Ninyerloa et al., 2000; Shen et al., 2001; Hasenauer et al., 2003; Shen and Shen, 2005a; Shen et al., 2005b), and (f) more complex expert rule-based methods (Daly et al., 2000; Johnson et al., 2000). By definition, an interpolation problem involves approximating an unknown function by an interpolation function whose form is postulated in advance either explicitly (e.g., second-order polynomial), or implicitly (e.g. under a condition of minimum curvature). Parameters of the interpolation function may be optimized under deterministic (i.e., exact fit at points) or stochastic criterion (i.e., least-squares).
Unlike the classic interpolation problem, classic kriging starts with a statistical model rather than postulating an interpolation function. Kriging represents a family of statistical interpolation techniques in which correlation or covariance functions are specified to allocate weights to minimize variance and bias in interpolated estimates (Webster and Oliver, 2007). Thin-plate splines are polynomial functions that fit a smooth surface through the data points with minimum curvature and are a generalization of a multivariate linear regression model where a non-parametric function is specified. Recent reviews of the strengths and weakness of a select set of interpolation methods identify that the performance of existing methods varies, primarily, according to the relative influence of key forcing factors at different spatial and temporal scales (Hartkamp et al., 1999; Jolly et al., 2005; Daly, 2006). As the 10 km spatial scale marks an identified transition whereby terrain and water-bodies dominate climate spatial patterning (Daly, 2006), in our study here, we examine these forcing factors in connection with our validation work.

For both Eastern and Western Canada, Price et al. (2000) compared two statistical methods for interpolating 30-year monthly precipitation, minimum and maximum temperature ANUSPLIN (thin-plate smoothing splines) and GIDS (Gradient Inverse-Distance-Squared), and found that both methods performed equally well in Ontario and Quebec where topographic gradients were smoother. The GIDS method can be viewed as an “exact” interpolator within a local neighborhood, assuming that spatial autocorrelation is most sensitive to local station density gradients. However, in British Columbia (BC) and Alberta (AB), ANUSPLIN out-performed GIDS in interpolating precipitation due to abrupt changes in orography, sensitive to edge-effects. Nonetheless, both GIDS and ANUSPLIN have been shown in previous work to match the performance of universal kriging (Hutchinson and Gessler, 1994; Nalder and Wein, 1998). The main advantage of thin-plate splines (i.e., ANUSPLIN model) over competing geostatistical techniques is that splines do not require prior estimation of spatial covariance structure (Hutchinson, 1995). Likewise, a shortcoming of kriging is that it relies on (parametric) isotropic assumptions (Banerjee et al., 2005; Le and Zidek, 2006). Previous work that has compared ANUSPLIN with multivariate regression (MLR) and a rule-based method (PRISM—Parameter-elevation regressions on independent slopes model) have yielded precipitation estimates very close to expected values on the basis of stream-flow gauge measurements and the net-balance of river run-off flow in winter and summer for mountainous regions in southeastern BC and AB (Milewska et al., 2005). Lapse rate is the decline of temperature with increase in elevation. While on average the decline is 6.5°C per 1000 m (i.e., adiabatic portion of observed lapse rate), it can vary substantially according to local atmospheric characteristics (i.e., environmental portion of observed lapse rate). Stahl et al. (2006) has further examined the effects of orography in BC, comparing 12 interpolation methods that included GIDS, nearest-neighbor (NN), MLR, and LWR-G (lapse-rate by weighted regression with truncated Gaussian filter). Their study highlighted the importance of the calculation of local lapse rates when interpolating across mountainous terrain. Results of their work showed that such methods out-performed others, because they were better able to account for the high variability in daily lapse rate during winter that are known to reflect temperature inversions and rapidly changing circulation patterns (Laughlin, 1982). Recently, additional variables of flow accumulation, distance from stream, and distance from urban population have also been considered in reducing residual error in daily minimum temperature estimation (Choi et al., 2003; Lookingbill and Urban, 2003). The sensitivity of interpolation predictions of temperature has been identified to originate from both elevation differences and topographic diversity impacting solar irradiance pattern on sloping surfaces, requiring a consideration of sensible heat deficit (Chung and Yun, 2004). Geographic extent and resolution both affect the measurement and assessment in the relative importance of these variables (Jarvis and Stuart, 2001). For example, the most important factors at coarse (>100 km), medium (10–100 km), and fine
(<10 km) resolutions are the proximity to water (i.e., oceans or lakes), topography and orography, respectively (Daly, 2006). In summary, while simpler interpolation methods (e.g., NN and MLR) are able to retain variability in climate data, they often require considerable variation in model parameters due to inherent spatial heterogeneity in climate data, and are often not optimal or robust. Also, standard kriging methods retain spatial covariance, but cannot retain spatial and temporal non-stationarity in daily climate data. Spatial interpolation models must trade-off accuracy in preserving temporal non-stationarity against spatial non-stationarity or heterogeneity.

We report on findings of a verification and validation study of generating reliable daily climate surfaces for Canada using three widely-used interpolation models (i.e., ANUSPLIN, HYBRID, and DAYMET) using available historical records of daily station time-series data. As part of a broader methodology for adaptive interpolation and grid refinement, our modeling work has included examination of mass conservation of monthly and annual precipitation totals, spatial and temporal variance in precipitation, preservation of temporal trends, regional precipitation frequency during the growing season, and preservation of distribution extremes. Model validation includes the quantification of model precision (i.e., bias and variance error) and regional effects due to orography, topology, and climatology and presence of large bodies of water. Here, we provide selected results on the effect of regional orography and topography, proximity to water, coastal boundaries and seasonal variability in daily precipitation and temperature. Our findings provide an evaluation of the precision of the interpolation models and compare their strengths and weaknesses for interpolating daily precipitation and temperature at a fine spatial resolution of 10 km across Canada (south of 60°N) based on historical station data (1961–1990). Our findings reveal aspects where the models, individually and collectively, could be improved. We also discuss our results in the context of key challenges faced in developing reliable interpolation for non-homogeneous, non-stationary climate fields (Karl and Williams, 1987; Piper and Stewart, 1996; Daly et al., 2000; Janis et al., 2004; Mitchell and Jones, 2005). These are: (1) an inconsistent/decreasing number of stations over time and highly variable station density over space in the historical data set, requiring the selection of a narrower reference time-period (i.e., 1961–2003) for interpolation modeling, (2) large discrepancies between measured and digital-elevation model estimates of station elevation in some cases, (3) spatially-dependent discontinuities in station records with a previously reported annual measurement bias of 0.6–0.8°C for eastern Canada and appreciable smaller bias for western provinces, (4) temporal discontinuities in observed trends that interpolation models for Canada must track such as: (a) stronger decadal trends in annual precipitation amount, with larger shifts in minimum temperature than maximum temperature; (b) increasing frequency of heavy precipitation events; and (c) more and longer warm spells in western Canada and cold spells in the east, (5) separation of systematic uncertainties in data from process uncertainties in quantifying model interpolation error, (6) the need to resolve confounding effects of orography, land–water interaction and other variables on temperature and precipitation at the daily scale, (7) the need for customized Geographic Information System (GIS) scripts and computer code to manipulate, process, format, cross-reference, and store large data sets (i.e., >2000 monitoring stations and 47 years of 115 daily climate surfaces each containing 59741 interpolation points), and (8) logistical and operational transfer issues of interpolation surfaces to end-users within a reasonable time-frame to support their application work and decision making.

2. METHODS

An overview of our methodology for adaptive interpolation and grid refinement for the reference period data is shown in Figure 1. Verification is concerned with assessing the assumptions of a model given the
best quality data available, whereas validation is concerned with assessing the fit of a model to data and addresses questions such as: (1) are statistical assumptions of a model (e.g., normality, stationarity, and isotropy) satisfied? (2) how well does a model describe the data? and (3) how sensitive is a model to specific observations? (Haining, 1990). By assessing the accuracy of different models, one can identify which model provides a more accurate estimate of its own uncertainty.

2.1. Precipitation, temperature, and elevation data

Daily precipitation (mm), minimum and maximum temperature (°C), and elevation data were obtained from Environment Canada’s (EC) Meteorological Service of Canada (MSC) for the full historical record (1891–2004) (DLY04) consisting of 7514 stations. A reference period was identified as 1961–1990 consisting of 6616 stations. This data was then filtered for stations lying on or outside the national boundaries of Canada, stations with no measured elevation, and duplicate records, yielding a data set with 6600 stations. There were between 1200 and 2400 reporting stations during 1961–1990. Less than 10% of these stations were located at elevations greater than 1500 m. Elevation estimates at stations’ point locations were obtained from the EC reference data set having 0.001 m accuracy. Estimates were obtained from a Canadian 3 arc second digital elevation model (DEM) of Canada’s National Topographic Series 1:250 000 scale topographic data constructed by McKenney et al., using the ANUDEM algorithm (Hutchinson, 2004; Hutchinson and Gallant, 2000). Measured estimates of station elevations were in relatively good agreement with digital elevation model estimates at the 10 km spatial resolution ($R^2 = 0.971$), with larger discrepancies at coarser resolutions (i.e., 50 and 100 km).
The source of such discrepancies is attributed to large error in station elevation measurements, a varying definition of climate day and other systematic uncertainty changes due to incremental adjustments and improvements in monitoring equipment. A list of model parameters is provided in Table 1. Hereafter, we refer to daily maximum temperature (°C) as $T_{\text{max}}$, daily minimum temperature (°C) as $T_{\text{min}}$. The HYBRID model was the only model that did not utilize elevation input.

The ANUSPLIN model assumes a logistic model for precipitation occurrence, whereas the HYBRID and DAYMET models assume a binomial function. There were also differences in the maximum spatial correlation range specified in the HYBRID and DAYMET models for interpolating precipitation and temperature station data. These are discussed in the next section associated with details of each model. Trace amount was also specified differently in applying the models. The definition of trace amount was changed several times during the history of the monitoring network. Trace amount means that daily precipitation values less than this amount were assigned to be zero, and those greater, assumed their positive interpolated value. Measurement units for trace amount also changed from the imperial to metric system and introduced inconsistencies that were adjusted for. Reportedly, trace amount varied between 0.2 and 0.3 mm over history (Mekis, 2005). Trace amount can range, in station climate data, between 0.1–0.7 mm depending on station latitude and longitude and

### Table 1. Model parameters specified for gridding of daily temperature and precipitation. (−) denotes dimensionless units ($T_{\text{max}}$—maximum temperature, °C, $T_{\text{min}}$—minimum temperature, °C, PcPn—precipitation, mm)

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Units</th>
<th>Description</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
</table>
| ANUSPLIN | $\rho$ | — | Smoothing parameter | All | min GCV
| | $f$ | — | Roughness penalty | All | min GCV
| | $\text{POP}_{\text{crit}}$ | mm | Daily precipitation trace amount | PcPn | 0.5 |
| HYBRID | $d_{ij,T}$ | km | Spatial correlation length scale | $T_{\text{max}}, T_{\text{min}}$ | 200 |
| | $d_{ij,P}$ | km | Spatial correlation length scale | PcPn | 60 |
| | $\text{POP}_{\text{crit}}$ | mm | Daily precipitation trace amount | PcPn | 0.2 |
| DAYMET | $I$ | — | Number of station density iterations | All | 3 |
| | $R$ | km | Truncation radius | All | 400 |
| | $\alpha$ | — | Gaussian shape parameter | $T_{\text{max}}, T_{\text{min}}$ | 3.5, 3.5 |
| | $N$ | — | Average number of stations with non-zero weights | $T_{\text{max}}, T_{\text{min}}$ | 25, 25 |
| | $S_S$ | km | Spatial smoothing width for elevation regressions | PcPn | 15 |
| | $S_T$ | days | Temporal smoothing width for elevation regressions | PcPn | 3.5 |
| | $\text{POP}_{\text{crit}}$ | mm | Critical precipitation occurrence parameter | $T_{\text{max}}, T_{\text{min}}$ | 1, 1 |
| | $f_{\text{max}}$ | — | Maximum value for precipitation regression extrapolations | PcPn | 5 |
| | | | | PcPn | 0.2 |
| | | | | PcPn | 0.8 |
temperature. A lowest value of 0.2 mm from 1977 and 0.3 mm before 1977 is well supported for this
data set (Hopkinson, personal communication). The ANUSPLIN model was run with a trace amount of
0.5 mm, while the HYBRID and DAYMET models were run with a lower value of 0.2 mm. The higher trace amount for the ANUSPLIN model was selected to ensure that days without precipitation did not adversely affect fitting of its spline across larger spatial regions. This 0.3 mm adjustment in trace amount between model runs was deemed not to make a difference to the outcome of the assessment of these models, given that the data itself experienced such adjustment in trace amount.

2.2. ANUSPLIN model

The ANUSPLIN model has been implemented in a software package available from the Centre for Resources and Environmental Studies (CRES) at The Australian National University in Canberra (http://cres.anu.edu.au/outputs/anusplin.php). The software contains scientific numerical routines for generating regular grid and point estimates of various climatic and weather variables, having been applied by different research groups around the world (Hutchinson, 2004). This method fits thin-plate smoothing spline surfaces to noisy data that can be multivariate with a multiplicity of responses at each location, denoted \( y = (y_1, \ldots, y_n) \) for \( n \) monitoring points in space, where \( e = (e_1, \ldots, e_n) \) is assumed to be uncorrelated, normally-distributed error with zero-mean and unknown variance \( \sigma^2 \), such that covariance is \( V \sigma^2 \), for \( V \) as a known \( n \times n \) matrix that is positive-definite. Thin-plate smoothing splines are a generalization of standard multivariate linear regression where a parametric function is replaced by a smooth, non-parametric function. Smoothing splines in ANUSPLIN are fit to observational data by minimizing (implicitly) the generalized cross-validation (GCV) as its goodness-of-fit statistic under variation in the degree of data smoothing (smoothing parameter \( \rho \), as an inverse ratio of signal-to-noise) (Bates et al., 1987; Wahba, 1990; Marcotte, 1995). The function, \( f \), (in Equations 1a and b) is a known “smooth” function with \( m - 1 \) continuous derivatives, where generally \( m = 2 \) and second-order derivatives are used to specify boundary conditions.

\[
y_i = f(x_i) + e_i, \quad i = 1, 2, \ldots, n
\]  

\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \rho \int_a^b (f^{(m)}(x))^2 \, dx \quad \rho > 0
\]  

This method permits a degree of flexibility in specifying the functional form of the smoothing spline and for incorporating additional dependencies and covariates. The spline smoothing interpolation method is derived from co-kriging that assumes the mean of the dependent variable is varying and unknown, and error covariance is independent (i.e., uncorrelated), varying and unknown. Spline smoothing typically assumes that random errors are independent, i.e., no temporal or spatial correlated random error, correlation in data can affect signal to noise ratios (i.e., smoothing parameter) (Wang, 1996; Wang, 1998). However, spatially-correlated error can be estimated by the ANUSPLIN model by specifying a non-diagonal covariance matrix or removed (Hutchinson, 1995). Spline smoothing is a statistical approach and does not require specification of spatial correlation lengths and assumes a global rather than a local neighborhood.

To reduce the effect of skewness in the input data, a square-root transformation was applied. While this transformation of the data introduces small positive correlation in residual error and positive bias, it has been shown to be reduced by up to 10%, even though a degree of systematic bias is introduced.
Hutchinson, 1998b). Final model estimates were then squared after the square root transformation was applied. Spline functions are commonly characterized by differing dimensions of their functional argument. In this way, tri-variate spline functions have three dimensions in their argument, whereas a bi-variate function has two. In this study, a spline model that consisted of a second-order piecewise polynomial joined by “knots” (explained further in the next section) as a tri-variate thin-plate spline was specified. This spline considers a spatially-varying dependence of a univariate on elevation. This assumption was made rather than a bi-variate form that assumes no topographic dependence and a tri-variate partial spline that assumes a constant dependence on elevation (Hutchinson, 1998b). For the second-order polynomial spline function ($p = 2$) and three spatial dimensions ($\lambda = 3$) (i.e., latitude, longitude, elevation), the corresponding spatial covariance function, $C(r)$, for a separation distance $r$ has the continuous form, as per co-kriging under isotropic assumptions (i.e., space is homogeneous) (Chiles and Delfiner, 1990),

$$
\frac{C(r)}{C(0)} = \frac{1}{\lambda^p} |r|^\lambda \log(r), \quad \lambda = 3, p = 2
$$

(1c)

The daily climate data is highly non-homogeneous and does not strictly obey the above power-law decay equation above (Equation 1c). However, in diagnostic testing, this isotropic function was useful for fitting to the data and testing of stationarity assumptions. For the daily climate data, spatial variance depends on more than just the distance separating two locations.

Knots are points that overlap with station locations and the number of knots specified was determined by sampling independently the range of each spline variable equally. The number of knots reflects the degree of spatial heterogeneity. The ANUSPLIN method selects the largest residuals and adds them as knot points (i.e., largest 20–50 residuals) and then re-fits a surface until a stable solution is obtained, or the variance of estimates are in agreements with expected values. Also, if a signal is within 10% of the number of knots (i.e., the maximum possible signal) then knot number are also increased. A condition to re-fit to a maximum of 2–3 times was used to prevent serious over-fitting, as this method can suffer from over-smoothing in areas where there is high curvature in topography and orography, missing localized pronounced peaks, while under-smoothing in flatter areas. The interpolation scheme in the ANUSPLIN method was adapted in close consultation with the developers. A two-stage approach was added to the ANUSPLIN model to deal with data gaps in daily precipitation. The first-stage consisted of determining spatial occurrence probability, whereby precipitation data was converted to binomial values (i.e., 0 or 1) according to a logistic regression function. Interpolated surfaces were then generated in the second-stage based on occurrence probability being greater than a critical or trace amount (\(\text{POF}_\text{crit}\)). This “ad hoc” procedure does introduce spatial discontinuities in the daily surface, but removes negative interpolation predictions as well as removal of values that are zero (i.e., zeros introduced by greater than 3-day evaporation loss), and values close to zero (i.e., trace) in the input data. An alternative to this extension of the ANUSPLIN model could involve a more formal treatment via disjunctive kriging (see Le and Zidek, 2006).

2.3. **HYBRID** model

The HYBRID model was developed by Shen et al. (2001) and is not available as a freely distributable software package. The method is a “hybrid” of inverse-distance weighting (IDW) and the natural-neighbor assignment that is also known as the Thiessen polygon method. For station $i$ of a total number
of stations, $M_j$ in relation to a grid point $j$, the estimated temperature at the interpolated grid point, $\hat{T}_j$, follows an inverse-distance weighting function for $i \neq j$ given below,

$$\hat{T}_j = \begin{cases} 
T_i & i = j \\
\left( \sum_{i=1}^{N} \frac{1}{d_{ij}} \right)^{-1} \sum_{i=1}^{M_j} \frac{T_i}{d_{ij}} & i \neq j (i = 1, 2, \ldots, M_j)
\end{cases}$$

(2a,b)

where $d_{ij}$ is the station-to-grid point distance and $T_i$ is observed temperature at a station $i$. $M_j$ is selected as the first $n$th natural neighbors that lie with temperature and precipitation correlation length scales, typically, $M_j$ varies up to eight neighboring stations for $d_{ij,T} \leq 200$ km and $d_{ij,P} \leq 60$ km, respectively. Because the total number of stations varies and is determined within a search radius, this model inherently adjusts to changes in station density. When a station and grid point coincide ($i = j$) the interpolated value is the observed value. When no stations are situated within the respective climate variable correlation length scale to a grid point (i.e., $M_j = 0$), then station assignment is applied. This method assigns the interpolated value as the value of the first nearest station, one of the hybrid features of this model. Inverse distance method is used only for temperature, while the hybrid method is used for precipitation. In this context, the “hybrid” feature of this model is centered on predicting precipitation. Because inverse-distance weighting typically over-estimates the number of precipitation days, while under-estimating daily precipitation amount, this approach is generally not able to represent observed temporal and spatial variance in precipitation (Shen et al., 2001). For this reason, inverse-distance and natural-neighbor methods are combined in estimating daily precipitation amount and frequency of day with trace amounts of precipitation. In the hybrid approach, the precipitation of a polygon’s centroid defines whether a polygon has precipitation on a given day. If the centroid of a polygon has precipitation, then the polygon is defined to also have precipitation that day. The nearest station to a centroid of a polygon is taken as the best indicator for a centroid’s precipitation and which days receive precipitation, as a centroid of a polygon is rarely the location of a station. Monthly total precipitation of a polygon ($P_{m,polygon}$) is the sum of the daily polygon precipitation ($P_t$) determined by the inverse-distance method. The precipitation frequency is then calculated by neighbor assignment, where $P_{t,centroid}$ is the precipitation of the station nearest to the centroid for the given day, $t$, and $P_{m,centroid}$ is the monthly total precipitation of the station(s) nearest to the centroid. $P_{t,polygon}$ is the precipitation over a polygon for a given day, $t$ (Shen et al., 2001).

$$\hat{P}_{m,polygon} = \sum_t \left[ \left( \sum_{i=1}^{N} \frac{1}{d_{ij}} \right)^{-1} \sum_{i=1}^{M_j} \frac{P_t}{d_{ij}} \right]_{i \neq j} (i = 1, 2, \ldots, M_j)$$

(2c,d)

Precipitation occurrence for this model is determined by a binomial interpolation function weighted by observed occurrence of surrounding stations and trace amount (POPcrit). For this model, isotropic spatial covariance is represented as the contribution of two terms—the first from inverse-distance decay and the second due to station assignment as,

$$\frac{C(r)}{C(0)} = e^{-\left( \frac{r}{a} \right)^2} - \mu re^{-2\mu r}$$

(2e)

Apart from inhomogeneities in the climate variable fields, also introduced by changes in station density, Equation 2c leads to substantially different spatial variance from the power-law decay isotropic field described by Equation 1c. This is especially apparent as the station to grid separation distance $r$. 

Copyright © 2010 Crown in Right of Canada
increases. A spatial correlation length scale of 200 and 60 km was used for temperature and precipitation, respectively. These values were consistent with a previous application and validation testing of this model within Alberta, Canada (Shen et al., 2001). These values can be compared to an spatially-averaged empirical correlation function derived from information of the entire North American landmass and approximated at the monthly scale that varies from 250–1050 km, with correlation extending in winter compared to summer months (Groisman and Easterling, 1994). The values specified here at the daily time-step were previously estimated from a maximum value of 1200 km at the monthly-scale slightly above the empirical range determined by Groisman and Easterling, under the assumption that daily temperature anomalies are independent from each other (Shen et al., 2001).

2.4. DAYMET model

The DAYMET model scheduled for release as a web-portal in 2008–2009 by the U.S. Numerical Terradynamic Simulation Group at the University of Montana (www.daymet.org). For a grid point \( j \) and a total of \( M_j \) stations with station-to-grid distances \( d_{ij} \leq d_{ij,T} \), temperature is predicted by exponential decay weighting of observed temperature at a station location \( i \), regressed by observed station-to-grid point elevation \( (z_j - z_i) \) differences,

\[
\hat{T}_j = \frac{\sum_{i=1}^{M_j} e^{-\left(d_{ij}/d_{ij,T}\right)} \left(T_i + \beta_0 + \beta_1 (z_j - z_i)\right)}{\sum_{i=1}^{M_j} e^{-\left(d_{ij}/d_{ij,T}\right)}}
\]  

(3a)

where \( \beta_0, \beta_1 \) are coefficients obtained by regressing temperature with elevation. In Equation 3a, \( z_j \) and \( z_i \) are the elevations of stations \( j \) and \( i \), respectively. Similarly, daily precipitation, conditional on precipitation occurrence (PO) at station and grid points, is estimated also according to a weighted regression (i.e., precipitation–elevation) as,

\[
\hat{P}_j = \left(\frac{\sum_{i=1}^{M_j} (e^{-\left(d_{ij}/d_{ij,T}\right)} - e^{-\alpha}) (\frac{T_i}{P})^\text{PO}_i}{\sum_{i=1}^{M_j} e^{-\left(d_{ij}/d_{ij,T}\right)} (\frac{T_i}{P})^\text{PO}_i}\right)\left(\frac{\hat{T}_j}{\hat{T}_j < \hat{T}_\text{max} < 1}
\]  

(3b,c)

As with the HYBRID model, precipitation occurrence for the DAYMET model is determined by a binomial interpolation function weighted by distances to surrounding stations and by trace amount of precipitation (POPcrit). To reiterate, in the ANUSPLIN model, precipitation occurrence is predicted based on a logistic function. Similar to the HYBRID model, the isotropic form of DAYMET’s spatial covariance is the contribution of two terms—the first from exponential-distance decay and the second due to the model’s truncation dependent on the shape parameter, \( \alpha \).

\[
\frac{C(r)}{C(0)} = e^{-\left(\frac{r}{\alpha}\right)^2} - e^{-\alpha}
\]  

(3d)

Equation 3d reveals that the DAYMET model, in comparison to the HYBRID model (Equation 2e) contains an additional parameter, \( \alpha \). This parameter scales the dependence of climate variables

Copyright © 2010 Crown in Right of Canada

Environmetrics (2010)

DOI: 10.1002/env
spatially. This scaling represents a distance-independent contribution to spatial interpolation variance. DAYMET interpolates daily values using data from stations located within the truncation radius. This radius depends on the density of stations around an interpolation location and on the shape parameter. To parameterize the model, an iteration procedure starts with an initial truncation radius, a given number of iteration steps and target number of stations required for interpolation for temperature and precipitation. Thus, if station density is high, truncation radius is low and vice versa. Based on changes in station density, the truncation radius is varied from the initial value to ensure the targeted number of stations for interpolation achieving the lowest mean absolute error (MAE). The initial Gaussian truncation radius \( R_p \) was fixed at 400 km and shape parameter, \( \alpha \), was 3.50 for \( T_{\text{min}} \) and \( T_{\text{max}} \) and 6.25 for precipitation. These same parameter values were used previously for daily interpolation in U.S. (Thornton et al., 1997) with a target number of 25 stations for temperature, and 15 for precipitation interpolation, but differ from values in recent validation testing of DAYMET for complex terrain in Austria (Hasenauer et al., 2003). For Austria, an initial truncation radius of 150 km was specified and using the same number of target stations, variable-specific values of 3.31, 4.12, and 3.87 for the shape parameter, and mean truncation radii of 46, 44, and 31 km for \( T_{\text{min}} \), \( T_{\text{max}} \), and precipitation, respectively, were estimated by DAYMET’s iterative parameter-adjustment procedure. Note that with this shape parameter generally varying between 1 and 7, the parameter settings for precipitation of 6.25 for Canada was considerably larger than estimated for Austria. Such differences likely arise due to the dependence of precipitation on changes in station density.

2.5. Model validation

A list of 368 “high-quality” Reference Climate Stations (RCS) were provided by Environment Canada (Milewska and Hogg, 2001). From this RCS set, 150 stations were selected having: (1) at least 27 years of data, (2) at least 90% temporal coverage, and (3) a location south of the boundary line determined by: \( \text{lat} = -0.15 \times \text{long} + 42.0 \) (Mike Hutchinson, personal communication). As 95% of the historical climate monitoring stations are situated south of this line, stations north of this latitude line have few if any neighbors, and were not included because withholding such stations would artificially lead to misleading large residuals/outliers as a result of the low station density in the data. Of these 150 stations with long-term, high-quality daily precipitation data, a module/routine called SELNOT, in the ANUSPLIN model, was applied. Using the SELNOT routine, 53 stations were selected. This module approximately samples the three-dimensional space spanned by the data, in which longitude, latitude, and elevation are each scaled to have unit variance (Hutchinson, 1995). Three extreme stations were then removed because they had few (if any) neighboring stations. The remaining 50 withheld stations had nearly continuous or complete daily precipitation records and sampled the full range of latitude, longitude, and elevation, south of the previously described northern line of latitude and lie within the agricultural boundary region of Canada (Agricultural Census, 2001). The spatial coverage of these 50 withheld stations for the 30-year reference period (1961–1990) for validating the models is shown in Figure 2. It is important to note that a higher number of validation stations will not necessarily improve the reliability of validation testing, but the quality of such selections does. While, ultimately, the reliability of such a test increases with the more repeated, independent data available to validate against, in the case where independent data is not available, the level of representativeness of the validation stations is most important (as per the 50 stations that were selected). This is because when stations are withheld they should have the largest impact on the ability of interpolation models to track and reproduce spatio-temporal trends. Also, having validation results for a consistent sample size of stations facilitates robust interpretation and inter-comparisons. For example, a total of 50 stations and a
Figure 2. Long-term monitoring sites withheld for cross-validation testing (n = 50)
similar procedure was used previously to validate and compare the performance of ANUSPLIN with the GIDS model (Price et al., 2000), and 100–200 stations in recent validation of historical climate surfaces generated by the ANUSPLIN model at the monthly scale for Canada and U.S. (McKenney et al., 2006).

The most effective method now commonly used to assess the error of climate data estimation is cross-validation (Cressie, 1993). Leave-one-out cross validation (LOOCV) was applied for cross-validating the interpolation models. This method involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is the same as a $K$-fold cross-validation, with $K$ being equal to the number of observations in the original sample. The LOOCV cross-validation method (also known as the withholding method) works well for a station-dense region, since the withheld station does not alter the station distribution significantly. However, deleting one station may significantly alter the station distribution in the station-sparse regions. Further, if the spatial distribution of the precipitation has large gradient or noticeable discontinuity, the deletion of one station even in a station-dense region can still alter the interpolation results significantly from the true interpolation that uses all the stations (S. Shen, personal communication). Validation statistics that were computed are bias and variance error, range and $t$-statistics on the mean, minimum, and maximum values in daily temperature and precipitation across weekly, monthly, and annual time periods. Error statistics on daily temperature and precipitation distribution percentiles and percentage of correct precipitation occurrence were also computed. Model validation statistics commonly reported in the literature are standard residual error (SE), mean absolute (MAE) and bias error (MBE), and root-mean-squared error (RMSE). Here we use MBE and RMSE to elucidate differences in the ability of each model to estimate and predict temperature and precipitation as

$$\text{MBE} = \frac{1}{n} \sum_{j=1}^{n} (f(x_i(t_j)) - y_i(t_j))$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (y_i(t_j) - f(x_i(t_j)))^2}$$

where $n$ is the total number of days for cross-validation purposes, $i$ denotes spatial location, and $j$ denotes a unit of time interval (i.e., day), $f(x)$ is a model-estimated climate variable (i.e., precipitation, minimum temperature, maximum temperature). The MBE statistic retains both the magnitude and sign of the difference between measured and model estimates: positive and negative values indicate over-and under-estimation, respectively. More generally, the statistics MAE and MBE estimate the degree of bias in a variable’s mean, whereas SE and RMSE measure statistical precision due to both variance and bias and measure the central tendency and detects extreme errors, being sensitive to the presence of outliers in data. Validation statistical analysis was conducted using SPLUS Version 6.2 (Insightful Corp.). Canada-wide spatial maps of daily mean, bias, and error variance for all climate variables for selected years provided visualization of regional contouring, including point, localized, and regionalized gradients of the validation results, and were generated using ArcGIS 9 (ESRI, Canada). Also, the proximity of withheld stations to closest water body were calculated according to the Haversine (i.e., half-versed sine) formula (see Sinnott, 1984) considered accurate to within 1 m and well-conditioned for numerical computation even at small distances, unlike calculations based on the spherical law of cosines, and with the aid of ArcGIS Spatial Analyst© toolbox.

Model validation yielded daily means and statistics for temperature and precipitation at weekly, monthly, and annual aggregated intervals of time. Weekly and monthly time intervals were selected for each season. The annual period is affected by seasonal variation, but the other periods are not, given that they are contained within separate seasons. The monthly periods were selected mid-season, i.e.,
period #1: January (winter), #2: April (spring), #3: July (summer), and #4: October (fall/autumn). The mid-season weeks were then selected as: #1: days 15–22 (all years), #2: days 106–112 (non-leap year) or 107–113 (leap year), #3: days 198–204 (non-leap year), 199–205 (leap year), #4: days 289–295 (non-leap year), 290–296 (leap year). For any particular period (i.e., weekly, monthly, annual), validation statistics were computed for each of the three interpolation models, with reference to the 50 validation (i.e., withheld) stations, using the following steps. (1) For each station, calculate daily means in $T_{\text{min}}$, $T_{\text{max}}$, and precipitation for the chosen period for both the observed (OD) and interpolated (ID) values (i.e., 7 averages for a week, 30/31 day-averages for a month, and 365/366 averages for a year). (2) RMSE and MBE differences between OD and ID values for each station. (3) Calculate means of these differences over the 50 validation stations. (4) Over all stations, calculate the parametric (paired $t$-test) and associated probability of exceedance as an estimate of the statistical significance of model differences for precipitation, $T_{\text{min}}$, and $T_{\text{max}}$. Where normality conditions were not satisfied, a non-parametric (ranked-Wilcox paired test) was applied. (5) For each station, over all years and for both OD and ID values, calculate measures of extremes as the lower 5th percentiles and minimums (for $T_{\text{min}}$) and upper 95th percentiles and maximums (for $T_{\text{max}}$ and precipitation) of all daily values (non-averaged) for the time periods. (6) For OD and ID values, calculate differences in the percentiles, calculating the mean of these differences over all stations and the associated $t$-statistics and probability of exceedance. (7) The steps 1–6 were repeated for a subset of validation stations within high elevation, complex topography ($n = 6$), and low elevation, flat topography ($n = 6$). (8) For each time period and station, over all 30 years, calculate the percentage of correct precipitation predictions over all 30 years. This percentage was calculated as the number of correct model predictions of precipitation that were also observed (termed as occurrence) added to the number of correct model predictions that were not observed (termed as non-occurrence) in the numerator and then divided by the total number of days of precipitation in the denominator. Percentage correct precipitation occurrence is a more reliable test than just frequency of precipitation days, as the latter can hide significant errors of commission and omission (M. Hutchinson, personal communication). Also, daily precipitation occurrence was defined as daily precipitation greater than or equal to the trace amounts of 0.2 mm. (9) Calculate percentage precipitation occurrence across all stations. (10) Calculate daily average precipitation (i.e., averages not totals) for all periods for each station and year for both OD and ID values, including both zeros and non-zero values, MBE and RMSE statistics. (11) Calculate daily mean precipitation across all stations. (12) Calculate the proximity (separation distance) of each station to nearest large water body (i.e., ocean or lakes) associating RMSE statistics in mean daily rainfall for each period.

2.6. Interpolation region and boundaries

Spatial point coordinates (i.e., grid nodes) were generated in projected space (Albers Conical Equal Area) having an equi-distant (i.e., regular) spacing of 10 km and clipped to the east/west coastline and southern national boundaries. Projections are classified according to the properties they preserve. In producing spatial maps of the grid, the Albers Conical Equal Area projection was selected because it preserves area. Although scale and shape are not preserved, distortion is minimal between the standard parallels. This projection is the standard projection for British Columbia (see http://geobc.gov.bc.ca/ and information provided therein). Previous examination of the spatial representativeness of this dataset of Canadian surface climate conditions identifies that the data set does not provide adequate climatological information north of 60°N latitude, but that southern regions of the country is adequately represented, except for some areas around Hudson Bay (Milewska and Hogg, 2001). Based on this, interpolation was a performed south of this northern...
boundary 60°N and covers the main agricultural region of Canada as per the Agriculture Boundary Census (2005). Geo-referencing was performed using the United States Geological Survey (USGS)'s General Cartographic Transformation Package (GCTP) (available at ftp://edcftp.cr.usgs.gov/pub/software/gctpc/). Distance to closest water body (i.e., proximity) was determined using numerical routines provided by ArcGIS—Spatial Analyst Extension, Version 9 (ESRI, Canada 2004). For each of the three models, the total size of files is 33 GB (GeoTIFF file format) and considering all three primary variables, there are 47,115 total daily climate surfaces were generated (i.e., each containing 59,741 interpolation points/grid nodes).

3. RESULTS AND DISCUSSION

The total percentage of daily precipitation occurrence (hereafter, DPO), correctly predicted by each of the three interpolation models, falls between 67–80% (Figure 3a). This percentage decreased in winter months (i.e., January mth1 and wk1 time periods). The seasonal variation across time at the weekly and monthly time periods is consistent, with no significant improvements as the length of time period increases. The ANUSPLIN model predicted DPO the best of all three models, roughly 2% better than the DAYMET model, and 2–5% better than the HYBRID model. These percentages are not significantly different given the random variation involved. Overall, the models correctly predict the occurrence of no precipitation much better than days having precipitation, where values range between 84–94% (not shown). A regional comparison of DPO for six validation sites shows that for spring (period 2) and autumn (period 4), the models predicted DPO for stations situated at lower elevation and flat terrain in Alberta (AB) (Figure 3c) up to 10% better than high elevation/complex terrain in British Columbia (BC) (Figure 3b). This suggests that the models can better track seasonal (i.e., temporal) variation in daily precipitation under smoother spatial variation (i.e., topography and orography). Moreover, this result highlights a spatial versus temporal trade-off that can significantly affect interpolation prediction of daily precipitation. Mean surfaces generated by unbiased nearest-station contouring of model interpolated estimates at the 50 validation stations for 2003 at the monthly time periods (Figure 4) provide a profile of model seasonal variation in daily precipitation and dominant trends between winter and summer precipitation across Canada. For January/winter, precipitation is high in the BC Rockies and localized further south along the Eastern and Western coastal regions, with dry conditions in AB Prairies. In contrast, in July/summer, precipitation is greatly reduced in the BC Rockies with dry conditions persistent in the AB Prairies, and high precipitation extending far inland from either coasts. Also, higher (lake-effect) precipitation is modeled on the downwind/leeward side of the Great Lakes in Ontario in January/winter as compared to July/summer where the lakes cause a downwind decrease in precipitation of 10–20%. Lake-effects are the combined result of changes in wind, temperature, moisture, and the stability of “meso-scale,” synoptic air flow. Intuitively, the modeled lake-effect could be explained on the basis of changing land surface roughness, whereby a reduction in roughness reduces lake-effect precipitation (see Lofgren, 2006). The smoothness conditions for spline-fitting (ANUSPLIN model), even with the inclusion of DEM elevation input are shown to generally produce smoother gradients and transition in the precipitation-level boundaries. This can be compared to the HYBRID model daily mean precipitation surface, which has no elevation input and produces sharper gradients, with more piece-wise linearity in precipitation boundary gradients and higher frequency of localized gradient (i.e., patchiness), but still tracking spatial variation and seasonal shifts in daily precipitation reasonably well at 10 km resolution.
The seasonal effect of water bodies evident from model interpolation output is revealed by spatial changes in daily precipitation around the Great Lakes. This is a well-known regional climatological forcing effect, and also is found to have a measurable impact on the estimation of precipitation by the models. Figure 5 shows RMSE estimation error with respect to the proximity of validation stations to

Figure 3. Percentage of correct model prediction of daily precipitation occurrence for: (a) Top inset: all validation sites \( (n = 50) \), (b) Middle inset: high elevation, complex topography \( (n = 6) \), (c) Bottom inset: low elevation, flat topography \( (n = 6) \). Note: week and month time periods are mid-season in the following numerical order: 1 (January/winter), 2 (April/spring), 3 (July/summer), and 4 (October/Autumn)
large bodies of water (i.e., both lakes and oceans), for all three models and all years, across monthly
periods. The insets in each of these plots refer just to validation stations in proximity to the Great Lakes/
Great Lakes Region in Ontario ($n = 12$) and profile just lake-effects on RMSE error. These profiles of
estimation error for January, April, and October show similar patterning whereby daily precipitation
highly variable nearest to water, ranging between 1 and 14 mm and then sharply decreases to
approximately 200 km in spatial range, where it again sharply increases within a range of 3–6 mm,
decreasing further away. The large range of error in close proximity to water, could, in part, be
attributable to station density effects of the monitoring network that extend between 40 and 64 km
(Milewska and Hogg, 2001). During summer (i.e., July) the range of error close to water is much
reduced and more uniform as the proximity to water decreases. As none of the models explicitly
considered lake-effect spatial correlation in daily precipitation, additional prediction error is
introduced because underlying spatial correlation is not accounted for. During winter, spring, and fall,
correlation extends more locally to roughly 160 km, but in summer this error correlation profile
disappears, a uniform range of error whether close or far away from water bodies. This may be induced
by a more non-local control of lake-effect precipitation in summer, but certainly this effect is seen to
extend to 200 km, by its nonlinear forcing effect on model estimation error. Our results here profile
topography, orography, and lake-induced forcing affects on interpolation error for daily precipitation.
We now supplement these results by highlighting forcing effects on the mean, range and extremes in
daily minimum and maximum temperature and precipitation.

Model estimates of daily $T_{\text{min}}$ deviate greatest during spring and summer, with all models under-
estimating for mid-season weekly, monthly periods, leading to net annual under-estimation (Figure 6).
For cooler climate in autumn and winter, all models tend to over-estimate $T_{\text{min}}$ (except HYBRID in the autumn). For spring–summer, the HYBRID model produces estimates of $T_{\text{min}}$ that are the least biased, however for autumn–winter, the model estimates $T_{\text{min}}$ with the greatest bias. Bias in $T_{\text{min}}$ ranges between $0.4-0.8^\circ\text{C}$. This is similar to the daily mean of maximum temperature ($T_{\text{max}}$), except that the models over-estimate $T_{\text{max}}$ more than $T_{\text{min}}$ during the colder seasons, while under-estimating $T_{\text{max}}$ more than $T_{\text{min}}$ during the warmer ones. Across an annual time period all models under-estimate daily mean $T_{\text{min}}$ and over-estimate mean $T_{\text{max}}$, as seasonal bias changes across the warmer and colder seasons are combined. Error variance (RMSE) in the model interpolation predictions of temperature vary between 1.5 and 3.5 $^\circ\text{C}$, with model estimates differing by only 0.5 $^\circ\text{C}$ consistently for the weekly, month, and annual periods. For $T_{\text{min}}$, variance error is greatest where bias is smallest, whereas for $T_{\text{max}}$, variance error is greatest where bias is greatest, such that all models have the lowest precision in estimating $T_{\text{max}}$ during the winter. Hereafter, we refer to daily precipitation (mm) as PcPn. Daily PcPn is under-estimated by ANUSPLIN in spring and summer, with bias varying between $0.2\text{ mm}$, whereas other models consistently over-estimate within $0.3\text{ mm}$. Except for winter, HYBRID model produces the least biased PcPn estimates. RMSE values for daily PcPn range between 3 and 5 mm for all time periods, with the highest error occurring in summer (i.e., roughly 1 mm increase in error compared to other seasons), for all models.
Model bias and error variance statistics for daily means are summarized in Table 2 across all years and stations (region A) are compared to conditions of high elevation/complex terrain (region B) and low elevation/flat topography (region C). In the case of $T_{\text{min}}$, the ANUSPLIN model yields the small bias and error variance for all of Canada, and also regionally for high elevation/complex terrain, slightly better than DAYMET. Differences here may be attributed to differing model assumptions and methodology, but also differences in inputs, whereby DAYMET performs a lapse-rate regression on elevation input, and HYBRID did not utilize station elevation information. Unexpectedly, the DAYMET model out-performed the other models in estimating $T_{\text{min}}$ across low elevation/flat topography. Given that HYBRID model did not utilize elevation information, it performs remarkable well. If bias and error differences between HYBRID and the other models are attributed to orography alone, then orography introduces $-0.17$ biases and an error of 0.49°C for $T_{\text{min}}$. Likewise, if comparison

Figure 6. Cross-validation statistics (RMSE—Root mean squared error, MBE—mean biased error) for daily precipitation, minimum, and maximum daily temperature averaged over weekly, monthly, and annual time periods across 30 year period (1961–1990, $n = 50$)
between these statistics between regions B and C is attributed to topography alone, then topography introduces −0.09 biases and 0.04 °C error, considerably less than orography. These estimates provide a useful gauge on orographic and topographic effects on the model interpolation error. While the models under-estimate mean daily $T_{\text{min}}$ consistently, they over-estimate mean daily $T_{\text{max}}$. As with $T_{\text{min}}$, ANUSPLIN provided the best estimation across Canada, as well as within both the low and high elevation regions. Elevation and topological differences introduce roughly the same magnitude of bias and error in the estimating $T_{\text{max}}$, as with $T_{\text{min}}$. In the case of mean daily PcPn, ANUSPLIN provides the best overall model precision. Unlike the temperature variables, the model precision varies little under topographic and orographic change. DAYMET and ANUSPLIN’s use of elevation information for high elevation/complex terrain (region C) provide roughly 20% improvement in RMSE error. The DAYMET model yielded two-fold higher levels of bias than the other models for regions A–C.

The spatial distribution of model bias and error variance for each of the models is profiled in Figure 7. Also, bias and error variance for weekly mid-season periods across Canada, averaged over all 30 years of record are profiled in Figure 8. Regions where the models under-estimate daily mean PcPn appear orange and red, and over-estimation occurs in regions that are green and blue. For January, the models under-estimate on the Eastern coast with a northerly extension, and at single high-elevation validation station in BC (Figure 7). For all model surfaces, in April (spring), under-estimation occurs for two sites in northern Ontario and Quebec. A further examination of a larger set of surfaces (i.e., for all time periods and for both bias and error variance) reveals that these two sites, together with an additional one in Alberta exhibit the highest bias and error, appearing as distinct (i.e., localized) outliers. These validation stations are Nordegg, Alberta (1320 m, ID: 3054845), North Bay, Ontario (370 m, ID:6085700), and Causapscal, Quebec (168 m, ID: 7051200). All the models over-estimate

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MBE</td>
<td>RMSE</td>
<td>MBE</td>
</tr>
<tr>
<td>Minimum temperature ($T_{\text{min}}$) (°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>−0.06</td>
<td>2.29</td>
<td>−0.11</td>
</tr>
<tr>
<td>ANUPLIN</td>
<td>−0.02</td>
<td>2.26</td>
<td>−0.25</td>
</tr>
<tr>
<td>HYBRID</td>
<td>−0.14</td>
<td>2.50</td>
<td>−0.25</td>
</tr>
<tr>
<td>Mean</td>
<td>−0.07</td>
<td>2.35</td>
<td>−0.21</td>
</tr>
<tr>
<td>Maximum temperature ($T_{\text{max}}$) (°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>0.07</td>
<td>1.84</td>
<td>0.25</td>
</tr>
<tr>
<td>ANUPLIN</td>
<td>0.07</td>
<td>1.77</td>
<td>0.11</td>
</tr>
<tr>
<td>HYBRID</td>
<td>0.05</td>
<td>2.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Mean</td>
<td>0.06</td>
<td>1.90</td>
<td>0.15</td>
</tr>
<tr>
<td>Precipitation (PcPn) (mm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>0.15</td>
<td>3.69</td>
<td>0.04</td>
</tr>
<tr>
<td>ANUPLIN</td>
<td>−0.05</td>
<td>3.47</td>
<td>0.02</td>
</tr>
<tr>
<td>HYBRID</td>
<td>0.05</td>
<td>4.32</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean</td>
<td>0.05</td>
<td>3.83</td>
<td>0.03</td>
</tr>
</tbody>
</table>
daily PcPn north of the Great Lakes in winter, and along the West Coast during other seasons (refer to Figure 7). High error variance in mean daily PcPn for all models occurs during the summer across a considerably wide expanse and along the coasts. During winter, the models deviate most from observed data predominately across the Prairies (Figure 8).

Figure 7. Spatially-distributed model bias (MBE) for daily precipitation (weekly mid-season periods) across 30 years, 1961–1990, $n = 50$). Dots indicate the spatial location of validation stations. For these maps, contrary to Table 2 and Equation 4a, negative and positive bias values indicate over- and under-estimation, respectively. This figure is available in color online at www.interscience.wiley.com/journal/env

Figure 8. Spatially-distributed model prediction error (RMSE) for daily precipitation (weekly mid-season periods) across 30 years, 1961–1990, $n = 50$). Dots indicate the spatial location of validation stations. For these maps, contrary to Table 2 and Equation 4a, negative and positive bias values indicate over- and under-estimation, respectively. This figure is available in color online at www.interscience.wiley.com/journal/env
Measured differences in interpolated and observed weekly, monthly, and annual distribution percentiles for all stations and years are shown in Figure 9. Similar to the results obtained for temperature variable means, all the models over-estimate the lower percentile of $T_{\text{min}}$ and under-estimate the upper percentile of $T_{\text{max}}$. During the spring there is a clear tendency for under-estimation in the lower extremes of $T_{\text{min}}$, but this is also a time when distribution percentiles are the least biased. In general, the models estimate upper daily temperature extremes better than lower ones. For the upper extreme of daily PcPn, ANUSPLIN consistently under-estimates, and HYBRID consistently over-estimates. The largest and smallest deviations in estimating the upper percentile in daily PcPn occurs with the ANUSPLIN and DAYMET model, respectively. These results indicate a greater level of

![Figure 9](image_url)
disparity between in models in estimating the upper extreme of daily PcPn than its mean. The models also preserve mean temperature values better than the extremes across Canada (refer to Table 3). Most apparent, is the increase in model bias in estimating the lower extreme of daily \( T_{\text{min}} \) and upper extreme in daily PcPn within high elevation/complex terrain (i.e., region C).

4. CONCLUSIONS

The total percentage of correct daily precipitation is estimated reasonably well by all models. They identify the lack of occurrence in daily precipitation much better than its occurrence. Nonetheless, high percentages achieved in both cases suggest that these models provide sufficient accuracy in matching days when precipitation is historically observed to occur. This result indicates that the generated historical interpolated time-series of daily mean PcPn are sufficiently accurate to be utilized for calibration purposes by, for instance, daily precipitation forcing of agricultural crop and drought models. However, appropriate selection of the nearby stations/locations where interpolated estimates are in highest agreement with observed data are still likely given the high mean model error variance in daily mean PcPn ranging between 3 and 5 mm across Canada. In many cases, applications that use the interpolated gridded data may not require high precision and/or can account for such error. Furthermore, the occurrence of regional forcing effects of water in close proximity to the Great Lakes and along the Coasts, of orography and topography between mountainous and Prairie landscape, introduce added heterogeneity in model estimation error. Our findings indicate that the models are most inaccurate in estimating daily mean PcPn during the summer, when there is high variability in PcPn due to convective control of local rainfall events and this inference is supported by previous validation work (Shen et al., 2005b; McKenney et al., 2006). The results obtained from a regional comparison of validation statistics indicates that orography and topology may explain roughly 0.5°C and 1 mm error in interpolated estimates of mean daily PcPn. Other recent validation of tri-variate thin plate smoothing spline at 0.05° across New Zealand (1960–2004) indicates that including topography in the model as an

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum temperature ( T_{\text{min}} ) (°C) (5th percentile)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>0.70</td>
<td>−0.10</td>
<td>1.50</td>
</tr>
<tr>
<td>ANUSPLIN</td>
<td>0.33</td>
<td>−0.34</td>
<td>1.05</td>
</tr>
<tr>
<td>HYBRID</td>
<td>0.51</td>
<td>−0.10</td>
<td>1.79</td>
</tr>
<tr>
<td>Maximum temperature ( T_{\text{max}} ) (°C) (95th percentile)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>0.20</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>ANUSPLIN</td>
<td>−0.15</td>
<td>0.14</td>
<td>0.36</td>
</tr>
<tr>
<td>HYBRID</td>
<td>−0.31</td>
<td>−0.06</td>
<td>−0.99</td>
</tr>
<tr>
<td>Precipitation (mm) (PcPn) (95th percentile)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYMET</td>
<td>0.20</td>
<td>0.60</td>
<td>−0.10</td>
</tr>
<tr>
<td>ANUSPLIN</td>
<td>−1.27</td>
<td>−0.18</td>
<td>−0.99</td>
</tr>
<tr>
<td>HYBRID</td>
<td>0.35</td>
<td>0.54</td>
<td>0.76</td>
</tr>
</tbody>
</table>
independent variable (rainfall surface) reduced interpolation error in PcPn more than orography alone (Tait et al., 2006). High error variance in mean daily PcPn for all models occurs during summer across a considerably wide expanse and along the coasts, whereas during winter, the models deviate most from observed data predominately across the Prairies. It is expected that inclusion of elevation input in the HYBRID model would improve its ability to estimate mean daily PcPn. As well, only for the ANUSPLIN model was a square-root transformation applied to PcPn observed values, so that some of the HYBRID’s resultant error variance, in comparison to the other models, can be attributed to this difference.

The higher (lake-effect) PcPn along the downwind/leeward side of the Great Lakes in Ontario in January/winter as compared to July/summer, where the lakes cause a downwind decrease in precipitation of 10–20%, matches the observed error range reported to be associated the impacts of the Great Lakes on regional seasonal climate conditions (Scott and Huff, 1996), and is in good agreement with previous findings from the simulation of lake–atmospheric interactions using the regional (MM4) simulation model (Bates et al., 1993). Spatially, lake-induced forcing affects the precision of all models, particularly during summer when error in daily PcPn is introduced by underlying spatial correlation <160 km proximity to large water bodies, above any station density effects within 40–64 km (Milewska and Hogg, 2001). These additional errors across time and space that are introduced by spatially-correlation driven by dominant climate forcing affects (i.e., at the daily scale and 10 km resolution), emphasizes the importance of a trade-off between tracking temporal (i.e., seasonal) variation and tracking spatial variation in mean PcPn. We find that this trade-off can affect interpolation predictions by up to 10%.

Increasing the number of high-quality, long-term stations for validation purposes from 50 to 100–200 would likely overall lead to better estimation of model error, but still would not address areas far away from monitoring stations (e.g., Northern Quebec) or across high elevation areas where high quality, historical data is far more sparse, and extrapolation, rather than interpolation is instead required. It is not clear why all three models consistently identified three validation stations with the highest error (i.e., Nordegg, Alberta (1320 m, ID: 3054845), North Bay, Ontario (370 m, ID:6085700), and Causapscal, Quebec (168 m, ID: 7051200). With all three models showing high error for these stations, either all the models neglect important trend behavior occurring at these locations, or more plausibly, measurement error is, in some way, contributing to such high error. If the high error at this location was due to just a station–density effect than the cross-validation station selection procedure should have eliminated the station due to low number of neighboring stations/low station density, relative to mean inter-station distance. Across complex terrain in Austria (1960–1998) using 23 validation stations, testing of the DAYMET model indicates that larger interpolation error in daily PcPn estimates at altitudes >1800 m is mainly attributable to measurement error (Hasenauer et al., 2003). MAE values for daily $T_{\text{min}}$, $T_{\text{max}}$, and PcPn were 1.17°C, 1.01°C, and 3.0 mm, with error varying with elevation (up to 1500 m) by 0.99–1.64°C, 0.76–1.73°C, and 2.5–5.0 mm, respectively. Our mean model error estimates for interpolation across Canada have half as less bias in temperature and a factor of ten less bias in PcPn, and with orography explaining 0.5°C and 1 mm error, our estimates appear surprisingly close to validation statistics for interpolation of daily temperature and precipitation across New Zealand.

Overall, the models tend to under-estimate $T_{\text{min}}$ for warmer periods and over-estimate $T_{\text{max}}$ for colder ones, indicating some difficulty in tracking seasonality (i.e., bias error about ±0.5°C) As identified previously from an analysis of observed time-series in mean daily temperatures and associated extremes across Canada, largest daily temperature changes are observed in winter and early spring, when substantial warming occurs (Bonsal et al., 2001), and in the summer, increases are
observed only for \( T_{\text{min}} \), alongside model estimates of daily \( T_{\text{min}} \) that deviate greatest also during spring and summer, at weekly, monthly periods, and annual periods. Clearly, the models track the observed seasonality trends in temperature and the largest bias and errors are introduced during seasons where the variables themselves exhibit the sharpest warming or cooling trend. How well models generally need to track seasonal trends in daily means to provide useful information depends on how the interpolated estimates and associated model surfaces are used.

In general, the models estimate upper daily temperature extremes better than lower ones. The models differ greatly in their ability to estimate upper extremes in daily \( \text{PCPn} \), compared to its mean. In particular, the ANUSPLIN model estimates the upper percentile in daily mean \( \text{PCPn} \) noticeably worse than the other models, but estimates the means, in many instances, better. Likely, the ANUSPLIN method may not reproduce extremes in precipitation well because it assumes spatial covariance is \( V \sigma^2 \), where \( V \) is known (see Method section), and suffers from the same deficiency as kriging in having a predictive error interval that is too small. While further improvements in estimating daily \( \text{PCPn} \) extremes (i.e., ANUSPLIN), the inclusion of orography as an input (i.e., HYBRID), and topography and the effects of water-bodies (i.e., all models) may improve their overall precision, at the current time, the best approach for interpolating daily temperature and precipitation across Canada requires a mixed-model approach. Based on our validation findings, the use of the ANUSPLIN model for interpolation daily mean values at 10 km is preferable, with variance in the mean estimated as a mixed-model average (i.e., all three models) across space and time.

Based on the findings of our assessment regarding interpolation of climate across Canada at the daily scale, we expect that the non-stationary models considered here would perform just as well in other regions. Localized environmental climate controls arising from variation in orography, topography, and lake-effects are expected to generate approximately the same range of prediction error as findings reported here. As long as various local controls were accounted for, we would expect the models to perform equally well in point estimation. However, broader climate trends are expected to vary substantially and noticeably differences between the models are expected when estimating extremes in temperature and precipitation. Performance of the models in other regions would clearly depend strongly on the quality of available data, the density and coverage of monitoring stations and their length of record. In applying these methods to other regions, we identify a particular need to ensure that there are sufficient stations of sufficient quality spanning the range of elevation. In particular, the HYBRID model needs to be further tested within Canada with elevation input.

It would be worthwhile to compare the performance of the models applied here to daily climate data against kriging methods. In this way, kriging would serve as a baseline for comparison and aid also in the comparison of validation statistics between different climate data sets. Applying kriging approaches would enable the assessment of stationarity assumptions. Kriging geostatistical approaches are associated with a broad set of algorithms/libraries of software that facilitate and simplify geostatistical analysis and visualization. In particular, the \( R \) libraries, “gstat” and “geoR” contain a large inventory of freely available algorithms (see www.r-project.org). In terms of cost, ease of use and availability and software support, interpolating using \( R \) libraries is likely the best route for future work. Using \( R \) statistical libraries in future interpolation work would enable testing of many different types of interpolation models with relative ease. Both simple/disjunctive kriging as well as a new model that combines the strengths of the three models considered here by employing a hierarchical, Bayesian methodology could be tested using this available software.

To improve the estimation of daily extremes over sufficiently long reference periods, such as the 30 years considered here, interpolation models may have to consider decadal-scale trends and variability. Incorporating spatially- and temporally-dependent adjustments in smoothing and
correlation range parameters could better capture seasonal, topography, orography, and proximity and boundary effects improving model precision. Such changes would likely require extension of validation and withholding procedures by requiring a broader examination of how model precision varies with number of validation stations. Comparison of our results with 1, 20, 50 km grids would extend the validation of daily interpolation model surfaces, especially where: the total number of possible validation stations is small, there is large variance in station density, and other historical data sets are not available to validate against. Empirical estimates of correlation range would improve the reliability of future model reconstructions of daily precipitation and temperature patterns and trends in space and time from climate monitoring networks.

ACKNOWLEDGEMENTS

This work was funded in part by Agriculture and Agri-Food Canada (AAFC): National Agro-climate Information Service (NAIS), Climate Impacts and Adaptation Branch, the National Land and Water Information Service (NLWIS), and Research Branch. Historical daily climate station data for all of Canada was provided by The Meteorological Service of Canada (MSC), Environment Canada for this study. The authors thank Ron Hopkinson for custom climate services as part of this validation study under contract KM040-05-6052R. The authors also thank for processing data and numerical modeling: Dan McKenney, Kevin Lawrence, John Pedlar, Pia Papadopol (Natural Resources Canada-Canadian Forest Service), Mike Hutchinson (Australian National University, Australia), Peter Thornton, Craig Hartsough (U.S. National Center for Atmospheric Research, CO), Samuel Shen, Bin Shen (University of Alberta, AB), and AAFC technical personnel: Huamei Yin, Liu Lin (Tracy) Yang, Prabha Virah Sammy, and Ken Maclean. Data, model analyses and production of graphical output presented here was prepared in part using ArcGIS 9 (ESRI Canada) software. The authors are grateful for the helpful input, feedback, and insights of many Canadian and international climate analysis and statistical experts who participated in a project workshop (June 2006, Edmonton, AB) where model interpolation results were first evaluated and compared. They also thank reviewers for their comments and insights that improved this manuscript.

REFERENCES


