A Comparison of Two Approaches for Generating Spatial Models of Growing-Season Variables for Canada

JOHN H. PEDLAR, DANIEL W. MCKENNEY, KEVIN LAWRENCE, AND PIA PAPADOPOL

Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste. Marie, Ontario, Canada

MICHAEL F. HUTCHINSON

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

DAVID PRICE

Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada, Edmonton, Alberta, Canada

(Manuscript received 4 February 2014, in final form 3 November 2014)

ABSTRACT

This study produced annual spatial models (or grids) of 27 growing-season variables for Canada that span two centuries (1901–2100). Temporal gaps in the availability of daily climate data—the typical and preferred source for calculating growing-season variables—necessitated the use of two approaches for generating these growing-season grids. The first approach, used only for the 1950–2010 period, employed a computer script to directly calculate the suite of growing-season variables from existing daily climate grids. Since daily grids were not available for the remaining years, a second approach, which employed a machine-learning method called boosted regression trees (BRT), was used to generate statistical models that related each growing-season variable to a suite of climate and water-related predictors. These BRT models were used to generate grids of growing-season variables for each year of the study period, including the 1950–2010 period to allow comparison between the two approaches. Mean absolute errors associated with the BRT-based grids were approximately 30% higher than those associated with the daily-based grids. The two approaches were also compared by calculating trends in growing-season length over the 1950–2010 period. Significant increases in growing-season length were obtained for nearly all ecozones across Canada, and there were no significant differences in the trends obtained from the two approaches. Although the daily-based approach tended to have lower errors, the BRT approach produced comparable map products that should be valuable for periods and regions for which daily data are not available.

1. Introduction

Length of growing season is an important determinant of plant growth and distribution. In agriculture, it is a key variable in crop-growth simulations that project spatiotemporal patterns in crop yields (e.g., Liu et al. 2007; Reidsma et al. 2009). In a similar way, many models of forest processes such as net primary productivity and carbon cycling benefit from accurate growing-season start and end dates (e.g., Richardson et al. 2009, 2010).

Moreover, growing season is recognized as an important indicator of climate change, and numerous studies have documented a trend toward longer growing seasons over the past 50–100 years (e.g., Qian et al. 2012; Christidis et al. 2007; Schleip et al. 2006; Beaubien and Freeland 2000). Growing-season-related variables have also been included in recent efforts to generate spatial climate models at regional (Wang et al. 2012), continental (McKenney et al. 2011), and global scales (Kriticos et al. 2012), reflecting a strong interest in these variables by various end users.

Growing-season length has been defined as the number of days in the year that a plant can grow (Wang 1963). Although simple in concept, rules for calculating growing-season start and end dates vary across regions and plant groups. For North American agricultural applications,
Growing season is often considered to be the number of days between the last occurrence of a critical daily minimum temperature (henceforth \( T_{\text{crit}} \)) in spring and the first occurrence of \( T_{\text{crit}} \) in autumn; \( T_{\text{crit}} \) is often set at 0°C, although temperatures of \(-2.2^\circ\text{C}\) and \(-4.4^\circ\text{C}\) have also been used (Brinkmann 1979).

From this definition, daily minimum temperature data are required to accurately estimate growing-season variables, but such data may not be available for all regions and time periods. For instance, standardized and quality-controlled daily climate data are scarce in Canada prior to 1950 (Hopkinson et al. 2011). Furthermore, future climate projections at the daily time step are not widely available [but see World Bank Group (2013)] and involve significant data-management issues. As a result, methods have been developed to estimate growing-season variables from widely available climate and environmental variables using a variety of statistical (e.g., Wang et al. 2012, 2006; Schlenker et al. 2007) and interpolation techniques (van Bussel et al. 2011; Nonhebel 1994).

Here, we report on an effort to produce grids of growing-season variables for Canada for the 1901–2100 period. Growing-season grids for the 1950–2010 period were generated directly from recently developed daily climate grids (Hutchinson et al. 2009; Hopkinson et al. 2011). Because daily grids were not available prior to 1950 and for future time periods, however, growing-season variables were estimated from a suite of predictors using a modern statistical technique called boosted regression trees (BRT; Friedman 2002). We summarize and compare errors associated with these grid products, and, to provide an illustrative application, we present trends across Canadian ecozones for growing-season length. The objective of this work is to contribute to an ongoing effort to generate—and make available—spatial climate data for Canada (McKenney et al. 2011). We anticipate a high demand for these growing-season products in the fields of forestry, agriculture, and a variety of other disciplines.

2. Data and related terminology

Here, we introduce the datasets and related terminology that are used in the current study. The term grid (or gridded data) refers to values of a climate variable that are arranged in a regular pattern across a geographic area; these can be observed as maps when viewed using appropriate software. All grids used in the current project were at a resolution of 0.1667 arc s (approximately 2 km). We employ the term model to refer to a mathematical relationship between a response variable and one or more predictor variables. Models can be gridded (or mapped) by supplying the predictor variables in gridded form and resolving the model at each grid-cell location. We also make use of the term climate-station data, which refers to the values of various weather/climate variables (temperature, precipitation, etc.) that are recorded at meteorological stations.

A key dataset in the current study consisted of recorded daily temperature and precipitation observations at 4267 climate stations across Canada for each day over the 1950–2010 period (Fig. 1a)—herein referred to as the daily climate-station data. These data were obtained from Environment Canada and were carefully checked for errors and inconsistencies, including a concerted effort to align observation times across stations and time periods (Hopkinson et al. 2011). Previously developed grids of daily minimum and maximum temperature and precipitation for each day over the 1950–2010 period—herein referred to as daily climate grids—were also used in the current study [see Hutchinson et al. (2009) for full details]. These daily climate grids were generated using the daily climate-station data described above. Station values were interpolated as a function of latitude, longitude, and elevation using thin-plate smoothing splines as implemented in the Australian National University “ANUSPLIN” software (Hutchinson 2004). The complexity of the fitted spline surface depends on the amount of smoothing; in ANUSPLIN this is optimized by minimizing the overall predictive error of the fitted surface as measured by the generalized cross-validation statistic [see Wahba (1990) for further details]. The resulting spline models (one for each day and climate variable over the 1950–2010 period) had mean absolute error (MAE) estimates of approximately 1.0°C for maximum temperature, 1.3°C for minimum temperature, and 1.0 mm for precipitation (Hutchinson et al. 2009). These models were resolved to a regular grid using a 0.1667 arc s (approximately 2 km) digital elevation model.

Generating grids of growing-season variables from the BRT models required grids of the predictor variables for past and future time periods. Annual grids of historical monthly and annual climate variables—herein referred to as historical climate grids—were generated in previous work (McKenney et al. 2006a). For this effort, a dataset of monthly climate-station values was obtained from Environment Canada for the 1901–2010 period. The number of climate stations in this dataset varied from as few as 133 at the start of the century to nearly 3000 at the peak of station coverage in the 1970s. As described above for the daily climate grids, the ANUSPLIN software (Hutchinson 2004) was employed to interpolate climate-station values of temperature and
FIG. 1. Locations of (a) weather stations used in generating the BRT models and (b) Canadian ecozones.
precipitation for each month over the period of interest. Withheld errors associated with these models were approximately 1.2°C for maximum temperature, 1.4°C for minimum temperature, and 30% for precipitation (McKenney et al. 2006a). The models were gridded as described above.

For the future climate grids [see McKenney et al. (2006b) for details], monthly time series of temperature and precipitation projections were downloaded for four GCMs [CanESM2, CESM1 (CAM5), HadGEM2-ES, and MIROC-ESM; see http://cmip-pcmdi.llnl.gov/cmip5/docs/ CMIP5_modeling_groups.pdf and http://www.ametsoc. org/PubsAcronymList for more information] and three representative concentration pathways (RCP 2.6, 4.5, and 8.5; van Vuren et al. 2011) from phase 5 of the Coupled Model Intercomparison Project (CMIP5) for the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. These GCM outputs were downscaled using a delta approach (Hay et al. 2000), in which projected changes (or deltas) in climate were combined with normal values recorded at North American climate stations over the 1961–90 period. These modified station values, which incorporated both established site-to-site variation in climate as well as the broad-scale average changes predicted by the GCMs, were then interpolated and gridded as described above. Here, we present results for RCP 8.5 and CanESM2 (Chylek et al. 2011).

3. Methods

a. Growing-season calculations for periods with daily climate grids

For the 1950–2010 period, nine growing-season variables for each of three \( T_{\text{crit}} \) values were calculated annually from the daily climate grids described above. A computer script was developed that processed the daily climate grids for a given year and identified, for each grid cell, the last day in spring and the first day in autumn on which \( T_{\text{crit}} \) values of 0°C, -2.2°C, and -4.4°C were encountered. The growing season was considered to start on the day following the last crossing of \( T_{\text{crit}} \) in spring and to end on the day preceding the first crossing of \( T_{\text{crit}} \) in autumn. Total precipitation was calculated by summing the daily precipitation values over the growing season. Mean temperature was calculated by averaging daily maximum and minimum temperature values over the growing season, and temperature range was taken as the difference between the lowest daily minimum temperature and the highest daily maximum temperature during the same period. Growing-degree-days were calculated by first subtracting a base temperature from the average temperature for each day during the growing season; daily values greater than zero were then summed. Base temperatures of 0°C, 5°C, and 10°C were employed, which are generally considered to be representative of the lower limits on growth of hardy (e.g., cabbage and asparagus), semihardy (e.g., peas and carrots), and non-hardy (e.g., tomatoes and corn) crops, respectively (Gordon and Bootsma 1993).

As noted above, the daily climate grids used here to calculate the growing-season variables have associated errors. To assess the impact of these errors on the resulting growing-season grids, we retrieved the estimate of each growing-season variable from the grid cell nearest the location of each of the 4267 climate stations in the daily climate-station data described above. We then calculated each growing-season variable as described above using the recorded daily climate data at each station; only station-years with fewer than 10 missing days were considered, resulting in a total of approximately 52 000 station-years (totals varied slightly for each growing-season variable). Errors were calculated as the difference between the gridded growing-season estimate and the value calculated using recorded climate-station data. Note that these are not predictive errors per se, because the stations at which these errors were calculated were used to generate the daily climate grids from which the growing-season grids were generated. Nonetheless, these errors do indicate the impact of spatial interpolation errors on the resulting growing-season grids.

b. Growing-season calculations for periods without daily climate grids

Daily climate grids were not available prior to 1950 or for future time periods; therefore, we generated statistical models that related each of the growing-season variables to a suite of predictor variables. These models were developed using the daily climate-station data at 4267 climate stations for the 1950–2010 period described above. At each of these stations, the nine growing-season variables were calculated for each of the three \( T_{\text{crit}} \) values; the resulting 27 variables constituted the response variables in the statistical analysis described below. Fifty-seven predictor variables were also estimated at each climate station, including monthly and annual temperature and precipitation variables and two water-related variables (see Table 1 for the list of variables). The monthly and annual climate-station variables were obtained from the historical climate grids described above. The water-related variables were distance to water and percentage of water in a 5-km radius around the station, estimated using a geographic information system. The impacts of different radii were explored, but 5 km yielded the strongest response. These
57 predictor variables were selected because they were readily available through our ongoing climate-modeling efforts and they all, to varying degrees, have an impact on heat and water budgets at a given location.

Each response variable was related to the suite of 57 predictor variables using BRT (Friedman 2002; Elith et al. 2008). BRT is an extension of simple regression trees, which typically employ a series of decision rules to recursively partition the observations in a dataset into subsets (or leaves) that are relatively homogeneous with respect to a given response variable. With BRT, however, a series of trees is generated rather than just one. The analysis proceeds by first fitting a regression tree to the data. The residuals from the first tree are then fed into the second tree, which attempts to reduce the error. This process is repeated through a series of successive trees. The final predicted value is formed by adding the weighted contribution of each tree. The individual trees are usually fairly small (typically three levels deep), but the full series may consist of thousands of these small trees. Since each tree in the series is fitted to the residuals of the preceding trees, there is an increasing focus on poorly modeled observations as the fitting proceeds. As a result, BRT models are generally highly accurate, performing as well or better than other modeling techniques (Elith et al. 2006). Overfitting, which is a problem with single regression trees, is avoided by using a random subset of the full dataset when fitting each successive tree in the series and by setting a slow “learning rate,” which reduces the influence of any given tree in the series. Another important consideration with BRT is the number of trees to include in the series; this choice is a trade-off among model accuracy, learning rate, and computing efficiency.

We implemented BRT using the “gbm” package in the R software (Ridgeway 2013). For all response variables, a series of 2000 trees was determined to be a reasonable trade-off between cross-validation error and acceptable computing time (Fig. 2). We used a 50% random subset of the data for fitting each tree in the series and a learning rate of 0.01. Smaller learning rates (e.g., 0.001) were tested but resulted in increased computing time with negligible improvement in model accuracy. Model predictive errors were estimated using fivefold cross validation. This approach, which is detailed in Elith et al. (2008), involved 1) randomly dividing the data into

<table>
<thead>
<tr>
<th>Code</th>
<th>Variable name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMT</td>
<td>Annual mean temperature (°C)</td>
</tr>
<tr>
<td>MDR</td>
<td>Mean diurnal range (°C)</td>
</tr>
<tr>
<td>ISO</td>
<td>Isothermality</td>
</tr>
<tr>
<td>TSEAS</td>
<td>Temperature seasonality</td>
</tr>
<tr>
<td>MAXTHM</td>
<td>Max temperature of hottest month (°C)</td>
</tr>
<tr>
<td>MINTCM</td>
<td>Min temperature of coldest month (°C)</td>
</tr>
<tr>
<td>TAR</td>
<td>Temperature annual range (°C)</td>
</tr>
<tr>
<td>MEANTWOQ</td>
<td>Mean temperature of wettest quarter (°C)</td>
</tr>
<tr>
<td>MEANTDQ</td>
<td>Mean temperature of driest quarter (°C)</td>
</tr>
<tr>
<td>MEANTHQ</td>
<td>Mean temperature of hottest quarter (°C)</td>
</tr>
<tr>
<td>MEANTCQ</td>
<td>Mean temperature of coldest quarter (°C)</td>
</tr>
<tr>
<td>ANNP</td>
<td>Annual precipitation (mm)</td>
</tr>
<tr>
<td>PWM</td>
<td>Precipitation of wettest month (mm)</td>
</tr>
<tr>
<td>PDM</td>
<td>Precipitation of driest month (mm)</td>
</tr>
<tr>
<td>PSEAS</td>
<td>Precipitation seasonality (mm)</td>
</tr>
<tr>
<td>PWQ</td>
<td>Precipitation of wettest quarter (mm)</td>
</tr>
<tr>
<td>PDQ</td>
<td>Precipitation of driest quarter (mm)</td>
</tr>
<tr>
<td>PHQ</td>
<td>Precipitation of hottest quarter (mm)</td>
</tr>
<tr>
<td>PCQ</td>
<td>Precipitation of coldest quarter (mm)</td>
</tr>
<tr>
<td>MINT_01, MINT_12</td>
<td>Monthly mean min daily temperature (°C)*</td>
</tr>
<tr>
<td>MAXT_01, MAXT_12</td>
<td>Monthly mean max daily temperature (°C)*</td>
</tr>
<tr>
<td>PCP_01, PCP_12</td>
<td>Monthly total precipitation (mm)*</td>
</tr>
<tr>
<td>WAT_DIST</td>
<td>Distance to large water body (km)</td>
</tr>
<tr>
<td>WAT_AREA</td>
<td>Amount of water in 5-km radius (%)</td>
</tr>
</tbody>
</table>

* Note that there are 12 variables for this table entry, one for each month.
five subsets, 2) developing a BRT model using four of the subsets, 3) testing the model predictions from step 2 using the remaining data subset, and 4) repeating steps 2 and 3 until each data subset had been used once to test the model predictions. Overall predictive error was summarized as the mean error across the five data subsets.

The final BRT models (one for each response variable) were used to generate grids of growing-season variables for all time periods of interest using the “predict.gbm” function in the gbm R package (Ridgeway 2013). This required grids of the 57 predictor variables—the historical climate grids (described above) were used for each year over the 1901–2010 period and the future climate grids (described above) were used for each year over the 2011–2010 time period.

c. Application of growing-season grids

Trends in growing-season length were calculated for Canadian ecozones (Fig. 1b; Marshall et al. 1999) over the period of 1950–2010 using data generated from the daily climate grids and the BRT approach. The trends were calculated using an autoregressive model to correct for serial correlation in the error term (SAS Institute 2004). The order of the autoregressive model used for each climate variable was determined using backward elimination starting with a 12th-order autoregressive model.

Analysis of covariance (ANCOVA; SAS Institute 2004) was used on each ecozone to test for differences between the trends generated using the daily and BRT approaches. This method typically models a response variable (GROWING SEASON LENGTH), as a function of both a continuous variable (YEAR) and a categorical variable (APPROACH—i.e., daily grids vs BRT approach); an interaction term (YEAR*APPROACH) is specified to test for differences in the slopes associated with the different levels of the categorical variable. Note that our primary interest with ANCOVA is the effect of APPROACH (which tests for differences in mean growing-season length between the two approaches) and the interaction term (which tests whether the two approaches have different temporal trends in growing-season length); the effect of YEAR (i.e., the temporal trend) is most appropriately addressed with the autoregressive model described above.

4. Results

A total of 1647 annual growing-season grids (9 variables × 3 \(T_{\text{crit}}\) values × 61 yr) were generated directly from daily climate grids for each year over the 1950–2010 period. MAEs associated with these grids averaged 5–6 days for the growing-season start and end variables, 10–11 days for growing-season length, 5%–15% for the degree-day variables, 6%–9% for precipitation, and less than 1°C for temperature mean and range (Table 2). Thirty-year averages (i.e., normals) were also calculated for each growing-season variable at each grid cell and climate-station; errors associated with the 1971–2000 normal period were about 30% less than annual errors (Table 2).

A total of 2970 annual growing-season grids (9 variables × 3 \(T_{\text{crit}}\) values × 110 yr) were generated using BRT models for the 1901–2010 period. A further 29,160 annual growing-season grids (9 variables × 3 \(T_{\text{crit}}\) values × 90 yr × 3 emissions scenarios × 4 GCMs) were generated by resolving the BRT models using grids of projected future climate. MAEs associated with the BRT models for the 1950–2010 period (Table 3) were generally higher than those associated with the products generated directly from daily grids for the same time period (Table 2): 9–10 days for the growing-season start
and end variables, 15–17 days for growing-season length, 5%–12% for the degree-day variables, 14%–20% for precipitation, and less than 1.5°C for temperature mean and range. Errors associated with the 1971–2000 normal period were 35% less than annual errors on average (Table 3).

These models allow growing-season variables to be mapped and analyzed over a significant time span. For example, maps of growing-season length across six consecutive normal periods clearly illustrate a trend toward longer growing seasons (Fig. 3). Changes were relatively subtle over the three historical normal periods, although increases of more than 20 days can be observed in western Canada and significant declines are not apparent in any region of the country. Substantial increases are projected by the end of the current century, with growing seasons that are 20–60 days longer than those of the early twentieth century across much of the country.

The 10 most important variables in the BRT models for growing-season start, end, and length (for $T_{crit} = 0$) are provided in Table 4. The variables identified as important in each model were highly plausible: spring minimum temperature variables dominated the growing-season-start model, autumn minimum temperature variables dominated the growing-season-end model, and a combination of spring, autumn, and annual temperature variables dominated the growing-season-length model. The water-related variables were identified as important in all three models. This is due to the fact that large bodies of water tend to moderate climate, resulting in longer growing seasons in these areas.

Because of the large number of iterations involved in a BRT analysis, complex relationships can be fitted between the response and predictor variables. Figure 4 shows the response curves for the 10 most important variables in the growing-season-start model. The shapes of the curves are relatively similar for all of the monthly minimum temperature variables, with little change in growing-season start date followed by a steep decline (i.e., earlier start date) across a critical range of temperatures. For example, locations with average minimum May temperature of 8°C are predicted to have growing-season start dates around yearday 130 (early May) while locations with average minimum May temperature of 2°C are predicted to start growing around yearday 150 (late May). The majority of climate stations in Canada are located in areas where the growing season commences in May and June; hence, these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis. The relationships shown for March and July minimum temperature are driven by the relatively few climate stations in Canada are located in areas where the growing season commences in May and June; these variables were identified as the two most important in the analysis.

Growing-season length increased across nearly all ecozones over the 1950–2010 period (Table 5; Fig. 5). The largest increases [of $\sim$5 days (10 yr)$^{-1}$] were in the Pacific Maritime and Montane Cordillera ecozones of western Canada. Conversely, the Hudson Plain and Taiga Plain ecozones showed little or no change in growing-season length. Trends were relatively consistent across the two approaches (i.e., daily grids vs BRT). An analysis of covariance for each ecozone revealed no significant differences (i.e., significance level $p > 0.05$) between the daily and BRT-derived products with respect to mean growing-season length (tested by the term APPROACH) and temporal trend (tested by the term APPROACH*YEAR).

---

**Table 3.** MAEs and MEs from BRT analyses that relate each growing-season variable (calculated from daily data recorded at 4267 weather stations for each of three critical temperatures $T_{crit}$) to 57 predictor variables composing annual climatic, monthly climatic, and water-related measures (see text for details and Table 1 for list of variables). Errors were calculated using fivefold cross validation.

<table>
<thead>
<tr>
<th>Growing-season variable</th>
<th>$T_{crit} = 0°C$</th>
<th>$T_{crit} = -2.2°C$</th>
<th>$T_{crit} = -4.4°C$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual*</td>
<td>Normal**</td>
<td>Annual*</td>
</tr>
<tr>
<td></td>
<td>MAE ME</td>
<td>MAE ME</td>
<td>MAE ME</td>
</tr>
<tr>
<td></td>
<td>MAE ME</td>
<td>MAE ME</td>
<td>MAE ME</td>
</tr>
<tr>
<td>Start (days)</td>
<td>10.4 0.7</td>
<td>6.7 0.2</td>
<td>9.5 0.5</td>
</tr>
<tr>
<td>End (days)</td>
<td>9.6 -0.8</td>
<td>6.2 -0.3</td>
<td>9.4 -0.5</td>
</tr>
<tr>
<td>Length (days)</td>
<td>16.8 -1.0</td>
<td>12.5 -0.7</td>
<td>15.2 -0.7</td>
</tr>
<tr>
<td>Degree-days over 0°C (%)</td>
<td>11.1 0.8</td>
<td>8.7 0.4</td>
<td>5.5 0.9</td>
</tr>
<tr>
<td>Degree-days over 5°C (%)</td>
<td>10.2 -1.2</td>
<td>8.2 -0.6</td>
<td>5.3 -0.4</td>
</tr>
<tr>
<td>Degree-days over 10°C (%)</td>
<td>10 -1.3</td>
<td>8.3 -0.4</td>
<td>6.8 -0.2</td>
</tr>
<tr>
<td>Precipitation (%)</td>
<td>20.5 -0.9</td>
<td>14.1 -0.2</td>
<td>13.9 0.0</td>
</tr>
<tr>
<td>Mean temperature (°C)</td>
<td>0.6 0.0</td>
<td>0.3 0.0</td>
<td>0.6 0.0</td>
</tr>
<tr>
<td>Temperature range (°C)</td>
<td>1.5 0.0</td>
<td>0.7 0.0</td>
<td>1.4 0.0</td>
</tr>
</tbody>
</table>

* Models were developed using annual values over the 1950–2010 period.
** Models were developed using values averaged over the 1971–2000 period.
5. Discussion

This study produced more than 30,000 growing-season-related grids as part of an ongoing effort to produce spatial climate data for Canada (McKenney et al. 2011). These grids provide a suite of growing-season variables for historical and future time periods (on the basis of outputs from four GCMs and three RCPs) that span two centuries. Of particular note are the grids generated using daily climate grids...
for each year over the period of 1950–2010; these products had relatively low errors when compared with values generated using actual daily data at climate stations. The errors associated with the BRT approach were on average 30% higher than those associated with the daily climate grids. This result is not surprising given that the BRT approach uses monthly and annual climate variables to estimate growing-season metrics that are ideally calculated using daily climate data. Although the daily-based approach tended to have lower errors, the BRT approach produced comparable map products that should be valuable for periods and regions for which daily data are not available. Users can access these data online (http://cfs.nrcan.gc.ca/projects/3) or make specific data requests by contacting the corresponding author.

We are not aware of other efforts that present comparable results for annual models in Canada. Wang et al. (2012) generated a suite of growing-season variables for western North America. Similar to our BRT approach, they used a multiple-regression model to estimate growing-season start and end dates as a function of a suite of readily available monthly and annual climate variables. They reported withheld MAE estimates for the 1961–90 normal period of 8, 5, and 12 days for growing-season-start, growing-season-end, and growing-season-length models, respectively. These values are comparable to the errors of 6.7, 6.2, and 12.5 days presented here for our BRT models over the 1971–2000 normal period.

Our experience suggests that many users are not aware of the magnitude and sources of errors associated with spatial data products such as these. For this study, errors were calculated by comparing values generated using derived daily data with those generated using actual daily data recorded at climate stations. It is important to recognize that recorded daily data also contain errors and inconsistencies, however, including observer error, instrument bias, inconsistent timing of observations, missing data, and transcription errors (see Mekis and Hogg 1999; Vincent and Gullet 1999). Efforts are ongoing to correct these issues (e.g., Mekis and Vincent 2011; Hopkinson et al. 2011), but there will always be a certain degree of error inherent in these data. Such errors may reduce or increase the growing-season errors reported here at any given location but likely average out over the full set of stations used here for model development and testing.

We found a trend toward longer growing seasons across all ecozones in Canada. This finding supports previous efforts that have reported positive growing-season trends [of ~1–4 days (10 yr)−1] at various locations throughout North America (McKenney et al. 2006a; Robeson 2002; Qian et al. 2012; Zhu et al. 2012; Terando et al. 2012). Despite significant differences in the method used to generate the daily-derived versus BRT-derived growing-season grids, there was little difference in the trends calculated using the two approaches. This result lends further credibility to the BRT-derived models, which constitute our only growing-season products outside the 1950–2010 period.

As computing capacity increases, climate data products are becoming available at increasingly fine spatial and temporal resolutions (Overpeck et al. 2011). For example, global downscaled GCM projections at the daily time step have recently been made available to the general public (World Bank Group 2013). Furthermore, historical daily data are continually being reassessed to improve data quality and availability (e.g., Hopkinson et al. 2011). As a result, it

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Predictor variable</th>
<th>Importance value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growing-season start date</td>
<td>MINT_05</td>
<td>32.1</td>
</tr>
<tr>
<td></td>
<td>MINT_06</td>
<td>23.5</td>
</tr>
<tr>
<td></td>
<td>AMT</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>MINT_03</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>MINT_07</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>WAT_AREA</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>MINT_04</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>WAT_DIST</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>MINT_02</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>MINT_09</td>
<td>1.7</td>
</tr>
</tbody>
</table>

| Growing-season end date            | MINT_09            | 44.3             |
|                                    | MINT_10            | 8.9              |
|                                    | MINT_08            | 7.9              |
|                                    | MINT_11            | 5.5              |
|                                    | WAT_AREA           | 5.1              |
|                                    | AMT                | 4.1              |
|                                    | MINTCM             | 4.0              |
|                                    | WAT_DIST           | 3.0              |
|                                    | MDR                | 2.1              |
|                                    | MINT_12            | 2.0              |

| Growing-season length              | AMT                | 21.1             |
|                                    | MINT_09            | 16.6             |
|                                    | MINT_06            | 8.6              |
|                                    | MINT_08            | 8.4              |
|                                    | MINT_05            | 7.9              |
|                                    | MINT_10            | 7.6              |
|                                    | WAT_AREA           | 5.4              |
|                                    | WAT_DIST           | 3.5              |
|                                    | MINT_07            | 2.8              |
|                                    | MINT_04            | 2.7              |
FIG. 4. Relationship between growing-season start date and the 10 most important predictor variables in a BRT analysis.
could be argued that the BRT-based products presented here will have a limited life span as daily data become increasingly accessible for past and future time periods.

There are several reasons why this obsolescence may not prove to be the case. It is widely recognized that GCM skill and accuracy decline at increasingly fine spatial and temporal resolutions (Seneviratne et al. 2012; Terando et al. 2012; Hegerl et al. 2007). For example, Jiang et al. (2013) reported that GCMs were reasonably accurate at predicting long-term monthly mean precipitation amounts in the southwestern United States but failed to simulate flood-generating extreme events that occur on the scale of days or hours. Terando et al. (2012) similarly reported poor GCM performance in simulating twentieth-century North American agro-climate indices, citing limited model skill at the daily time step as a major reason for the result. Thus, the extent to which the use of daily GCM projections improves the accuracy of future growing-season estimates is likely to be limited by GCM skill for the foreseeable future. Furthermore, downscaling and processing of daily GCM data require considerably more skill and computing capacity (both CPU and storage space) than are required for monthly projections. For groups with modest computing capacity, the BRT approach represents a computationally efficient approach for generating reasonably accurate growing-season products. Also, despite improvements in the quality and availability of historical daily data, there are limits to the confidence that can be placed on retrieval and reconstruction of climate-station records, say for the early twentieth century in remote regions of Canada (and many other countries). Long-term (century scale) trends and variations in growing-season variables, constructed from the more reliable and complete records of monthly data using BRT, may be invaluable for explaining past changes in ecosystem dynamics and agricultural production and as validation of models used for projecting the future.

6. Conclusions

We generated a significant time series of growing-season grids for Canada using two analytical approaches: directly from daily climate grids and using BRT models. The BRT approach allowed growing-season grids to be generated for time periods that lacked daily climate grids. These BRT-based grids had modest errors and compared favorably to growing-season grids computed directly from daily climate grids. To be specific, trends in growing-season length across Canadian ecozones did not differ significantly between the two approaches. These BRT-based grids had modest errors and compared favorably to growing-season grids computed directly from daily climate grids. To be specific, trends in growing-season length across Canadian ecozones did not differ significantly between the two approaches. This represents a novel application of the BRT approach that may stimulate further usage in the climate-modeling community.

We anticipate significant demand for these products. Users requesting growing-season data for years falling within the 1950–2010 period will be provided with the grids generated directly from daily climate grids; requests for outside this time frame will be fulfilled with the BRT-based grids. As daily climate-station data (and related grids) become available for other time periods we intend to update these growing-season grids, but it may be some time before reliable and homogenized daily climate data are available for the early part of
the twentieth century and for future years. Until then, our BRT-based grids provide valuable and consistent estimates of growing-season variables for Canada over the 1901–2100 period.

Acknowledgments. We recognize the significant contributions of Kathy Campbell to various aspects of this work. Comments provided by three anonymous referees are gratefully acknowledged.

REFERENCES


