

# Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain

Philippe Lucas-Picher · Daniel Caya ·  
Ramón de Elía · René Laprise

Received: 18 July 2007 / Accepted: 15 February 2008 / Published online: 14 March 2008  
© Springer-Verlag 2008

**Abstract** Previous investigations on regional climate models' (RCM) internal variability (IV) were limited owing to small ensembles, short simulations and small domains. The present work extends previous studies with a ten-member ensemble of 10-year simulations performed with the Canadian Regional Climate Model over a large domain covering North America. The results show that the IV has no long-term tendency but rather fluctuates in time following the synoptic situation within the domain. The IV of mean-sea-level pressure (MSLP) and screen temperature (ST) show a small annual cycle with larger values in spring, which differs from previous studies. For precipitation (PCP), the IV shows a clear annual cycle with larger values in summer, as previously reported. The 10-year climatology of the IV for MSLP and ST shows a well-defined spatial distribution with larger values in the northeast of the domain, near the outflow boundary. A comparison of the IV of MSLP and ST in summer with the transient-eddy variance reveals that the IV is close to its maximum in a small region near the outflow boundary. Same analysis for PCP in summer shows that the IV reaches its maximum in most parts of the domain, except for a small region on the western side near the inflow boundary. Finally, a comparison of the 10-year climate of

each simulation of the ensemble showed that the IV may have a significant impact on the climatology of some variables.

**Keywords** Internal variability ·  
Regional climate models · Ensemble of simulations ·  
North American climate · Ten-year simulations

## 1 Introduction

Global climate models (GCMs) are the most complete tools to simulate present climate and its evolution resulting from natural or anthropogenic modifications in the climate forcings (IPCC 2007). These models are built on basic conservation laws expressed as mathematical equations and solved numerically on a three-dimensional grid covering the globe. The evolution of the variables over time, as simulated by GCMs, is sensitive to small perturbations in the definition of the initial state. This sensitivity to small perturbations, often referred to as internal variability (IV), results from the dynamic and thermodynamic non-linear relations governing the climate system. In addition, the different time-responses of the climate system components associated to the numerous feedback processes also contribute to the climate system IV. Because of such a large spectrum of time scales in the climate system, and the strong feedbacks between the components, a transient state persists in the climate system even in the absence of modifications in the external forcings. This sensitivity to initial conditions limits our ability to predict the detailed evolution of the weather to about 2 weeks (IPCC 2001). However, its influence on the estimation of climate statistics is thought to be limited, at least for climate means.

---

P. Lucas-Picher (✉)  
Danish Meteorological Institute, Lyngbyvej 100,  
2100 Copenhagen E, Denmark  
e-mail: plp@dmi.dk

P. Lucas-Picher · D. Caya · R. de Elía · R. Laprise  
Canadian Network for Regional Climate Modelling  
and Diagnostics, UQAM, Montreal, QC, Canada

D. Caya · R. de Elía  
Ouranos Consortium, Montreal, QC, Canada

Today's computer power allows GCMs to be run for climate simulations, conducted over several decades, at a resolution of few hundreds of kilometres. However, this resolution is still too coarse to generate regional climate information applicable to most climate change impact studies. One popular approach to obtain high-resolution climate simulations is to dynamically downscale a GCM simulation with a regional climate model (RCM). The latter model simulates the evolution of the climate variables over a limited-area domain and is fed at its lateral boundaries by large-scale data taken from a GCM simulation or from an objective reanalysis. By concentrating the computer power over a limited area, the resolution can be efficiently increased to the order of tens of kilometres.

At the time of their establishment in the early 1990s, it was thought that RCMs were almost totally constrained by the lateral boundary forcing and admitted only one solution. However, recent studies have shown that RCMs keep a certain level of freedom and have significant IV despite being controlled at their boundaries by large-scale atmospheric flow (Giorgi and Bi 2000; Weisse et al. 2000; Rinke and Dethloff 2000; Christensen et al. 2001; Caya and Biner 2004; Rinke et al. 2004; Wu et al. 2005; Vanitsem and Chomé 2005; Alexandru et al. 2007; de Elía et al. 2008). Thus, like GCMs, RCMs are sensible to perturbations in the initial conditions, so that different solutions can be generated using the same set of lateral boundary conditions (LBC). However, the level of IV generated by RCMs would be smaller than by GCMs due to the restrictions on the large-scale atmospheric flow imposed by the lateral boundary conditions (Christensen et al. 2001).

Giorgi and Bi (2000) were amongst the first to study the IV using an RCM. They randomly perturbed the initial conditions or the lateral boundary conditions in a set of seasonal RCM simulations, and compared the solutions to those generated over the same period in a 13-month reference simulation. They showed that the level of IV is not sensitive to either the magnitude or the source of the perturbation, but is mostly conditioned by the synoptic circulation, the season (IV stronger in summer than in winter), the region and the model configuration. Giorgi and Bi (2000) also noted that the model's response to perturbations modified the day-to-day solution, but did not significantly affect the domain-wide average 3-month climatology.

Rinke et al. (2000) studied RCM's IV over the Arctic using four 2-month perturbed simulations compared to a reference. They raised the hypothesis that, for a given domain size, the forcing from the driving field on the RCM simulations is weaker over the Arctic than the mid-latitude. Rinke et al. (2000) also identified a strong influence of the domain's dimension on the IV and showed that a small

domain displays a weak dependence on the initial perturbations. Christensen et al. (2001) examined the problem from a different angle by running a 7-year RCM simulation driven repeatedly by the same annual set of atmospheric lateral boundary conditions, while allowing the soil simulated by the RCM to evolve freely. The experience involved an ensemble of seven 1-year simulations started with different initial conditions of soil variables. They found larger IV in summer than in winter. Furthermore, the IV estimated with the RCM was smaller than estimated with a GCM ensemble over the same area.

Caya and Biner (2004) compared three 1-year RCM simulations initiated with different atmospheric and/or surface initial conditions. They detected a clear annual cycle in the RCM's IV, with small values in winter and large values in summer associated to large discrepancies between the members of their small ensemble. The climate statistics of each simulation were similar even for summer period, despite the larger IV. In contrast, Rinke et al. (2004) found larger IV for the autumn/winter than for summer in the Arctic region. They suggested that this behavior might be explained by the polar vortex in winter, which impedes the migration of the perturbations out of the domain, and therefore reduces the lateral boundary control over the simulation.

With the growing of computational power over the years, the RCM's limited-area domain sizes have increased considerably. For example, the largest domain used in the ICTS [Inter CSE (Continental Scale Experiments) Transferability Study] now has  $201 \times 181$  grid cells with a 60-km resolution (Takle et al. 2007). In another recent study, Plummer et al. (2006) recently used a domain of  $201 \times 193$  grid cells at a 45-km resolution to study climate change over North America. As domains expand, LBC's control on RCM simulations is reduced and RCMs have more freedom to develop their own circulations, thereby increasing the IV. This weaker control by the LBC allows the RCM to modify large-scale atmospheric circulation, which creates problems at the outflow boundary when the RCM attempts to connect with the driving solution imposed by the one-way nesting scheme (Miguez-Macho et al. 2004). This also violates one basic assumption behind the use of RCMs as a physical interpolator of the existing driving data as input at the lateral boundaries (Jones et al. 1995). To prevent these discrepancies at the outflow boundary, certain modelling centres use large-scale nudging in the interior of their RCM domains (von Storch et al. 2000, Biner et al. 2000, Riette and Caya 2002), which keeps the large-scale of the RCM circulation close to the one of the driving field.

As mentioned by Vanitsem and Chomé (2005), the one-way nesting procedure introduces a free parameter, the size of the domain, which governs in a decisive way the

solutions generated by the RCMs. They showed that domain size influences the sensitivity of RCM simulations to the initial conditions. Hence, different trajectories were obtained for simulations started with different initial conditions over a large domain, while solutions were not so different from one another over a small domain.

Alexandru et al. (2007) explored RCM's seasonal IV using a time-lag ensemble of 20 simulations. Their study suggests that a minimum of ten simulations is required to obtain a robust estimation of the IV. Also, their study, performed with five domain sizes, showed a general increase of the IV with domain size. Finally, de Elía et al. (2008) analyzed the sources of uncertainty in RCM simulations over a 20-year period. Two simulations differing in their initial conditions showed that the IV impact on seasonal averages is relatively important in magnitude and that this magnitude decreases as the averaging period increases.

Until very recently, most studies on RCMs' IV were limited to small ensembles, short simulations and small domains due to limited computing capacity. The present work extends these limits by using a larger ensemble (ten-member) of multi-year (10-year) simulations over a larger domain (North American domain). The large ensemble will allow a more robust estimation of the IV, while the multi-year simulations will allow to investigate the dependence of the IV over time (time series and time means) and particularly the estimation of the interannual variability of the IV. The 10-year time series will also be useful to identify the spatial distribution of the IV with its 10-year climatology and to determine the long-term influence of the IV on the climatology of meteorological variables. Finally, the analysis of the IV over a large domain will provide a better estimation of the IV for domain size currently used over North America. The size and location of the selected domain are similar to the one used by the North American Regional Climate Change Assessment Program (NARC-CAP) (Mearns 2004).

The paper is organized as follows: the experimental set-up and the ensemble of simulations are presented in Sect. 2. The results and analyses follow in Sect. 3. Finally, conclusions appear in Sect. 4.

## 2 Experimental set-up

The present investigation employs the Canadian Regional Climate Model (CRCM: Caya and Laprise 1999) as modelling tool. The CRCM uses a semi-implicit semi-Lagrangian scheme to solve the fully elastic non-hydrostatic Euler equations. Its grid is projected on polar stereographic coordinates with a 45-km grid mesh (true at 60°N). This study uses CRCM 3.7, which differs in many aspects from CRCM 3.4 used by Caya and Biner (2004). We refer

the reader to Plummer et al. (2006) for a description of CRCM 3.7.

An ensemble of ten 10-year simulations (1980–1989) was performed with the CRCM 3.7 on the North American domain shown in Fig. 1. All simulations share exactly the same experimental setup (model configuration, LBC, land surface scheme, sea surface temperature and sea-ice spatio-temporal distribution), with the exception of the atmospheric initial conditions. The atmospheric initial conditions were perturbed by either modifying the starting time or by adding random or fixed perturbations in some of the atmospheric fields. It was found that the source or the magnitude of the perturbations has no impact on the level of internal variability 15 days after the initiation of the simulations, in agreement with the findings of Giorgi and Bi (2000). The number and the length of the simulations in the ensemble were limited by the time required to run the CRCM over the large domain with the available computing power. However, a ten-member ensemble is in agreement with the lower limit required to obtain a robust estimation of the IV, as suggested by Alexandru et al. (2007) with their season-long simulations.

The analyses were performed over the period of 1 January, 1980 to 31 December, 1989. No spin-up period was removed in order to keep the longer available time series comparison. However, time series analysis showed consistency in the behaviour of the IV over the years. The selected domain covers most of North America and contains  $193 \times 145$  grid points (Fig. 1). This domain contains high topography over the Rocky Mountains and the inflow boundary is mostly located in the Pacific and Arctic Oceans. Unlike the usual mode of operation of CRCM over comparable large domain, large-scale nudging was not employed in this study in order to allow the model to



**Fig. 1** North American domain uses for the CRCM ensemble of simulations. This domain contains  $193 \times 145$  grid cells at a 45-km resolution. Topographic heights are contoured every 500 m

evolve freely and to not interfere with the intrinsic IV of the model.

For all simulations, lateral boundary forcing and initial atmospheric states (horizontal winds, temperature, pressure and specific humidity) are taken from the NCEP/NCAR re-analysis data (Kalnay et al. 1996) at a resolution of  $2.5^\circ \times 2.5^\circ$  on 28 levels in the vertical. These atmospheric data are available at each 6 h and linear interpolation in time provides information on the CRCM boundaries at each 900-s time step. The nudging scheme of Davies (1976) is applied on the horizontal wind over a relaxation zone of nine grid-points on the periphery of the domain. The initial conditions for the land surface variables (surface temperature, liquid and frozen soil water contents, snow cover and snow age) are taken from a climatology of the Canadian GCM2 (McFarlane et al. 1992). Finally, the CRCM time-dependent surface boundary conditions for ocean-surface variables (sea surface temperature and sea-ice spatio-temporal distribution) are interpolated in time and space from the  $1^\circ \times 1^\circ$  resolution AMIP monthly means (Gates et al. 1999).

### 3 Results and analysis

#### 3.1 Climate statistics of the internal variability

The approach of Alexandru et al. (2007) is used to estimate the IV by computing the inter-member variance  $\sigma_X^2$  defined as

$$\sigma_X^2(i, j, k, t) = \frac{1}{M-1} \sum_{m=1}^M (X(i, j, k, t, m) - \langle X \rangle(i, j, k, t))^2. \quad (1)$$

The term  $X(i, j, k, t, m)$  refers to the value of  $X$  at a position  $(i, j, k)$  within the three-dimension grid, at the archival time  $t$  (archival interval of 6 h) and for member  $m$  of the ensemble.  $M$  corresponds to the total number of members (simulations) in the ensemble. While Alexandru et al. (2007) used the biased variance estimator, we used the unbiased variance estimator to prevent an artificial bias in the estimation of  $\sigma_X^2$  in our small ensemble size of ten members. The term  $\langle X \rangle(i, j, k, t)$  designates the ensemble mean defined as

$$\langle X \rangle(i, j, k, t) = \frac{1}{M} \sum_{m=1}^M X(i, j, k, t, m) \quad (2)$$

where  $X(i, j, k, t, m)$  is the value of the variable  $X$  at the coordinate  $(i, j, k)$ , at time  $t$  and for member  $m$ .

The inter-member variance  $\sigma_X^2(i, j, k, t)$  was computed for all cells of the three-dimensional grid and at each six-hourly archived time step of the simulated 10 years. The

inter-member variance provides an estimation of the IV for each variable analyzed. Since the IV fluctuates in space and time in different ways for each meteorological variable, we analysed its evolution using spatial and temporal averaging as for any other meteorological variable.

The time evolution of the IV is obtained with the domain average of  $\sigma_X^2$  computed as

$$\overline{\sigma_X^2}^{xy}(k, t) = \frac{1}{I \times J} \sum_{i=1}^I \sum_{j=1}^J \sigma_X^2(i, j, k, t) \quad (3)$$

where  $I$  and  $J$  designate the number of grid cells in the  $x$ - and  $y$ -direction of the horizontal plane over the domain of interest.

To describe the spatial distribution of the IV, we computed the 10-year climatology with the time average of  $\sigma_X^2$  defined as

$$\overline{\sigma_X^2}^t(i, j, k) = \frac{1}{N} \sum_{t=1}^N \sigma_X^2(i, j, k, t) \quad (4)$$

where  $N$  is the number of archived time steps over the period of interest. This expression represents the “climate” of the IV or its expected value over a given period of time and at a given location  $(i, j, k)$ .

The long-term impact of the IV on the climate of the meteorological variables is estimated by computing the variance between the climate of each member of the ensemble as

$$\sigma_X^2(i, j, k) = \frac{1}{M-1} \sum_{m=1}^M \left( \overline{X}^t(i, j, k, m) - \langle \overline{X}^t \rangle(i, j, k) \right)^2 \quad (5)$$

where  $\overline{X}^t(i, j, k, m)$  is the time average (climate) of each member  $m$  and  $\langle \overline{X}^t \rangle(i, j, k)$  is the ensemble mean of the time average.

It is important to appreciate the differences between Eqs. (4) and (5). The former defines the “climate” of the variance between each member of the ensemble while the latter defines the variance of the “climate” of individual members of the ensemble.

#### 3.2 Time evolution of the IV

We begin the characterization of the IV by analyzing its 10-year time series. Figure 2a presents the square root of the domain-average inter-member variance for the mean-sea-level pressure (MSLP) ( $\sqrt{\sigma_{\text{mslp}}^2}^{xy}$ ) from 1980 to 1989 computed using Eq. (3) with ten members. The square root of the variance is used to recover the original unit of the variable (e.g. hPa for MSLP). Also shown in this figure is the spatial root-mean-square-difference (RMSD) between

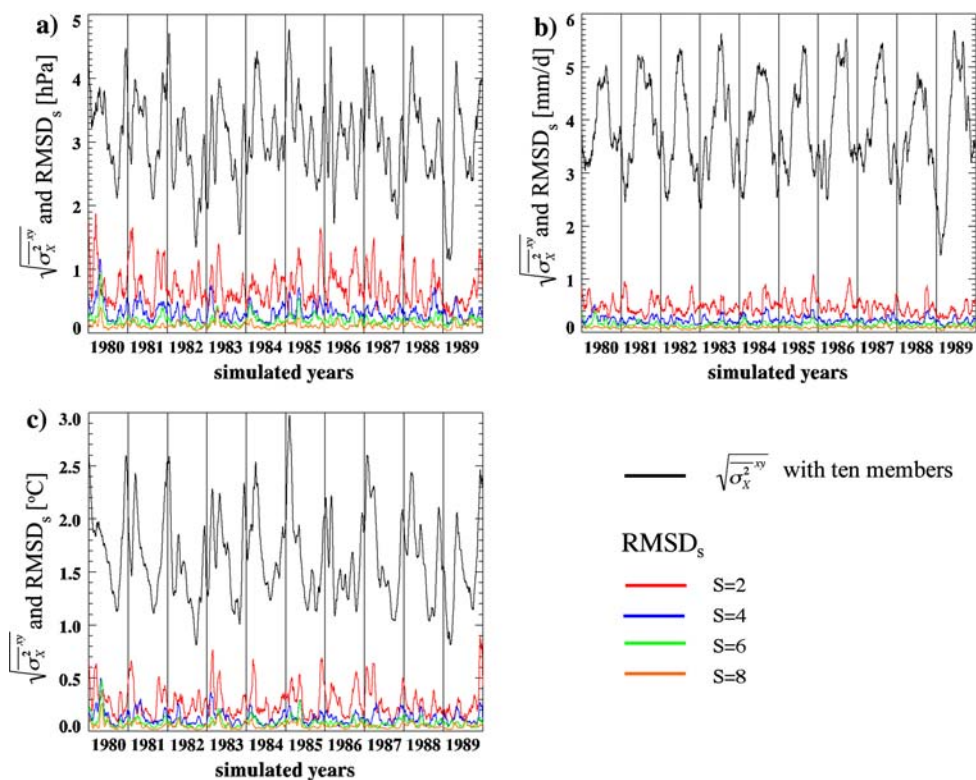
the estimation of  $\sqrt{\sigma_{X,10}^{2^{xy}}}$  with a ten-member ensemble, versus 2, 4, 6, 8 members indicated by  $S$ :

$$\text{RMSD}_S(k, t) = \sqrt{(\sigma_{X,10}(i, j, k, t) - \sigma_{X,S}(i, j, k, t))^{2^{xy}}}. \quad (6)$$

This latter computation allows the evaluation of the IV estimation error using a small number of members compared to ten members, which is considered to be the minimum ensemble size for a robust estimation of  $\sigma_X^{2^{xy}}$  (Alexandru et al. 2007).

Figure 2a shows that the IV fluctuates in time, but behaves similarly with different ensemble sizes according to the small  $\text{RMSD}_S$  values. We can see that the  $\text{RMSD}_S$  decreases as the ensemble sizes increase. In few occasions, the estimation of the IV with two members follows a distinct path (e.g. the beginning of 1981). Overall, the small values and the similar time evolution of the  $\text{RMSD}_S$  errors computed with eight and six members indicate that the computation with six members is generally robust. For precipitation (PCP; Fig. 2b) and screen temperature (ST; Fig. 2c), the same conclusions can be drawn, whereby the  $\text{RMSD}_S$  error on the estimation of the IV with six members is similar to the ones acquired with eight members. It is worth mentioning that no long-term tendency is visible in the IV time series contrary to the study of Wu et al. (2005), which showed that the impact of the initial conditions decreases as simulation time increases.

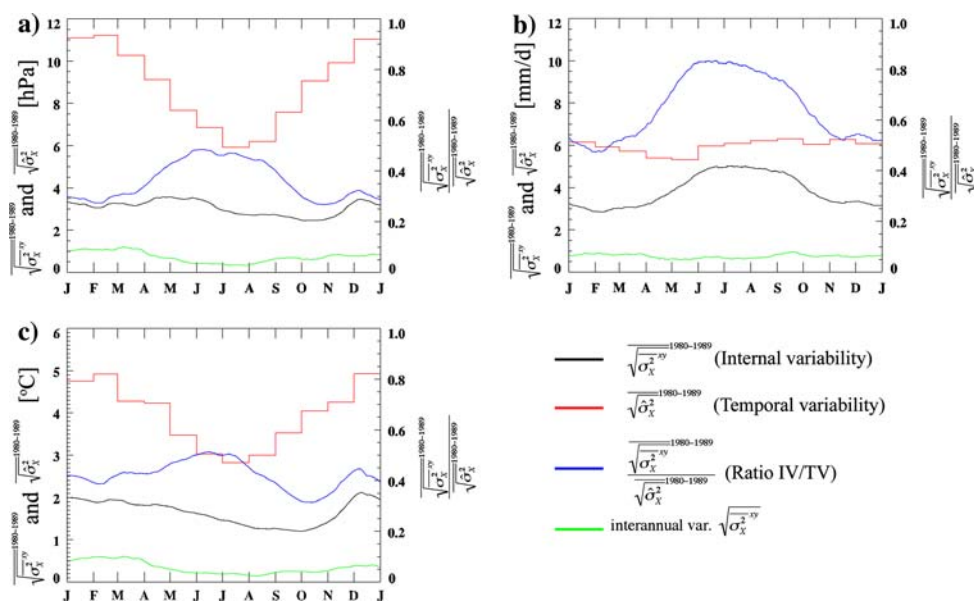
**Fig. 2** 1980–1989 square root of the domain-average inter-member variance ( $\sqrt{\sigma_X^{2^{xy}}}$ ) computed with ten members for **a** the mean-sea-level pressure (MSLP; hPa), **b** precipitation (PCP, mm/day) and **c** screen temperature (ST, °C). The colored curves present the spatial root-mean-square-differences ( $\text{RMSD}_S$ ) between the estimation of the ( $\sqrt{\sigma_X^{2^{xy}}}$ ) using ten members and those with two, four, six and eight members. A 30-day moving average is applied to each curve



### 3.3 Annual cycle and interannual variability of the IV

In order to investigate the annual cycle of the IV, Fig. 3 presents the 1980–1989 mean annual cycle of  $\sqrt{\sigma_X^{2^{xy}}}$  for MSLP, PCP and ST computed with the ten-member ensemble. The interannual variability is also presents in order to estimate the variability around the mean annual cycle of  $\sqrt{\sigma_X^{2^{xy}}}$ . In Fig. 3a, a weak 1980–1989 mean annual cycle of  $\sqrt{\sigma_X^{2^{xy}}}$  is observed for MSLP, with larger values in spring and smaller values in fall. The interannual variability of the IV exhibits a pronounced annual cycle for MSLP, with smaller values in summer and up to four times larger values in winter. The weak IV annual cycle for MSLP differs from previous studies conducted over smaller mid-latitude domains (Giorgi and Bi 2000; Caya and Biner 2004), which showed a clear annual cycle with larger values in summer. Lucas-Picher et al. (2004), based on pair of 2-year simulations, suggested that the size of the domain could explain this different annual cycle. Their results showed larger values of IV in winter with a large domain (similar to the one in the present study) and larger values of IV in summer with a smaller domain (similar to the one used in Caya and Biner 2004).

Two factors might contribute to the time evolution of the IV. The first factor is related to the decorrelation between the members within the ensemble due to the chaotic nature



**Fig. 3** Mean annual cycle of  $\sqrt{\sigma_X^{2-xy}}$  (internal variability) computed with ten-members over 1980–1989 in *black* for **a** mean-sea-level pressure (MSLP; hPa), **b** precipitation (PCP; mm/day) and **c** screen temperature (ST; °C). The *green* line shows the interannual variability of  $\sqrt{\sigma_X^{2-xy}}$ . The mean annual cycle of  $\sqrt{\hat{\sigma}_X^2}$  (monthly temporal

variability) over 1980–1989 of the first member is shown by the *red* line. The *blue* line shows the relative internal variability (*right-hand side scale*), computed as the ratio of the internal to temporal variabilities. A 30-day moving average is applied for the values associated to the *black*, *blue* and *green* lines

of the simulated climate system. This decorrelation might be dependent on the season or on the atmospheric conditions inside the limited area-domain. The second factor comes from the relation between the IV and the transient-eddy variability ( $\sigma_t^2$ ) as Caya and Biner (2004) showed in a pair of simulations using different initial conditions. As explained in Caya and Biner (2004), a pair of totally uncorrelated simulations can reach a maximum IV of  $\sqrt{2}\sigma_t$  computed with the RMSD, if there is no bias between the simulations and if both simulations share the same transient-eddy variance, conditions satisfied if the same model is used to run the simulations. By analogy, for an ensemble of simulations, the maximum IV computed with the inter-member variance of a large ensemble is  $\sigma_t^2$  when the members of the ensemble are unbiased, uncorrelated and share the same transient-eddy variance. The transient-eddy (or temporal) variance ( $\sigma_t^2$ ) is the natural variability of a meteorological field over time, which is mainly due to the travel of the weather systems along the storm track. As an estimation of the transient-eddy variability over time, Caya and Biner (2004) computed the domain-average temporal variance  $\hat{\sigma}_X^2$  of a variable  $X$  for every month as

$$\hat{\sigma}_X^2(i, j, k) = \overline{\left( X(i, j, k, t) - \overline{X(i, j, k)^t} \right)^2}^{xyt} \quad (7)$$

where the  $^{-xyt}$  operator refers to a domain monthly time average and  $^{-t}$  operator refers to a monthly time average. Since  $\hat{\sigma}_X^2$  is similar for each member due to the control of

the driving field on the RCM simulation,  $\hat{\sigma}_X^2$  was computed only for the first member of the ensemble. The ratio between the inter-member variance ( $\sigma_X^2$ ) and the domain-average monthly temporal variance ( $\hat{\sigma}_X^2$ ) normalize the IV. A ratio close to one tells that the IV of the RCM is close to its maximum value, which corresponds to the IV of a GCM. A ratio close to one also tells that the driving field have very limited control on the RCM simulation and that the RCM behave almost as a GCM.

Figure 3 presents the 1980–1989 mean annual cycle of the  $\sqrt{\hat{\sigma}_X^2}$  and the ratio between  $\sqrt{\sigma_X^{2-xy}}$  and  $\sqrt{\hat{\sigma}_X^2}$ . We can see on Fig. 3a that the  $\sqrt{\hat{\sigma}_X^2}$  has a large annual cycle for MSLP with larger values in winter than in summer due to the more intense cyclonic activity in winter. Thus, the ratio between  $\sqrt{\sigma_X^{2-xy}}$  and  $\sqrt{\hat{\sigma}_X^2}$  for MSLP shows higher values in summer than in winter meaning that each member are more uncorrelated from one another in summer than in winter. The ratio is equal to 0.5 in summer and 0.3 in winter. We think that it is the larger domain, which reduces the control of the driving field on the RCM simulations, and the larger variability in winter, that explain the larger IV of this ensemble in winter compared to previous works (Giorgi and Bi 2000; Caya and Biner 2004).

For PCP in Fig. 3b, the IV showed a clear 1980–1989 mean annual cycle with larger values in summer and smaller values in winter, in agreement with the results of

Giorgi and Bi (2000) and Caya and Biner (2004). This result may be dependent on the amount and type of precipitations, which is more convective in summer than in winter. The interannual variability of the IV for PCP in Fig. 3b is stable all year long (~1 mm/day). Due to the small annual cycle of  $\sqrt{\hat{\sigma}_X^2}$ , the ratio has a similar annual cycle than for the inter-member variance. The ratio is close to one in summer meaning that the inter-member variance is close to its maximum and that the generation of the PCP in the RCM depends little on the driving field. For ST in Fig. 3c, the small annual cycle of the IV is close to the one of MSLP in Fig. 3a, with large values in winter/spring and small values in fall. Other comments for the MSLP can be shared with the ST.

In winter, the atmospheric circulation of the mid-latitude, is characterized by a strong jet stream and fast moving weather systems. This intense atmospheric circulation imposes a strong forcing from the LBC on the RCM because of the large “flux of information” through the boundaries, which therefore reduces the IV. In summer, the atmospheric circulation is usually weaker and the parameterized processes (radiation, convection, etc.) are more active. The weaker flow reduces the information flux through the boundary and the stronger subgrid-scale processes of the model, which are more stochastic in their behavior, are believed to enhance the IV. These factors were used to explain the strong summer IV values reported in previous studies (Giorgi and Bi 2000; Caya and Biner 2004). The present study suggests that the large domain is responsible for the small annual cycle of the IV and the larger values of the IV in winter for MSLP.

### 3.4 Spatial distribution of the IV

Figure 4a, b present the 1980–1989 climatology of summer and winter IV for MSLP, as estimated using the square root of the time-average inter-member variance ( $\sqrt{\sigma_{\text{mslp}}^2}$ ) computed with a ten-member ensemble following Eq. (4). One can see that the IV of MSLP has similar spatial distribution and magnitude for both seasons, which is not uniformly distributed with larger values over the northeast region. As for the time evolution of the IV, two factors may contribute to the IV spatial distribution. First, the general easterly flow makes information from the driving fields entering by the western and northern boundaries for the mid-latitude domain. Therefore, on the west side of the domain, the RCM simulation is strongly conditioned by the driving data, resulting in a weak IV (low inter-member variance). Moving eastward in the domain, the chaotic nature of the flow acts to increase the IV. The IV reaches its maximum value in the northeast region (Fig. 4a, b) just before the outflow boundary. At the northeastern boundary, the RCM

solution is forced back to the driving flow by the one-way nesting. Therefore, the IV has to reach zero at the boundary where the driving circulation is imposed on the RCM.

The second factor derived from the relation between the IV and the transient-eddy variability ( $\sigma_t^2$ ) as showed by Caya and Biner (2004). As discussed in Sect. 3.3, for an ensemble of simulations, the maximum IV computed with the inter-member variance for a large ensemble is  $\sigma_t^2$  when members of the ensemble are unbiased, uncorrelated and share the same temporal variance. The transient-eddy variance ( $\sigma_t^2$ ) is the natural variability of a meteorological field with time, which mainly results from the travel of the weather systems along the storm track. It can be estimated as follow for each member  $m$  of the ensemble as

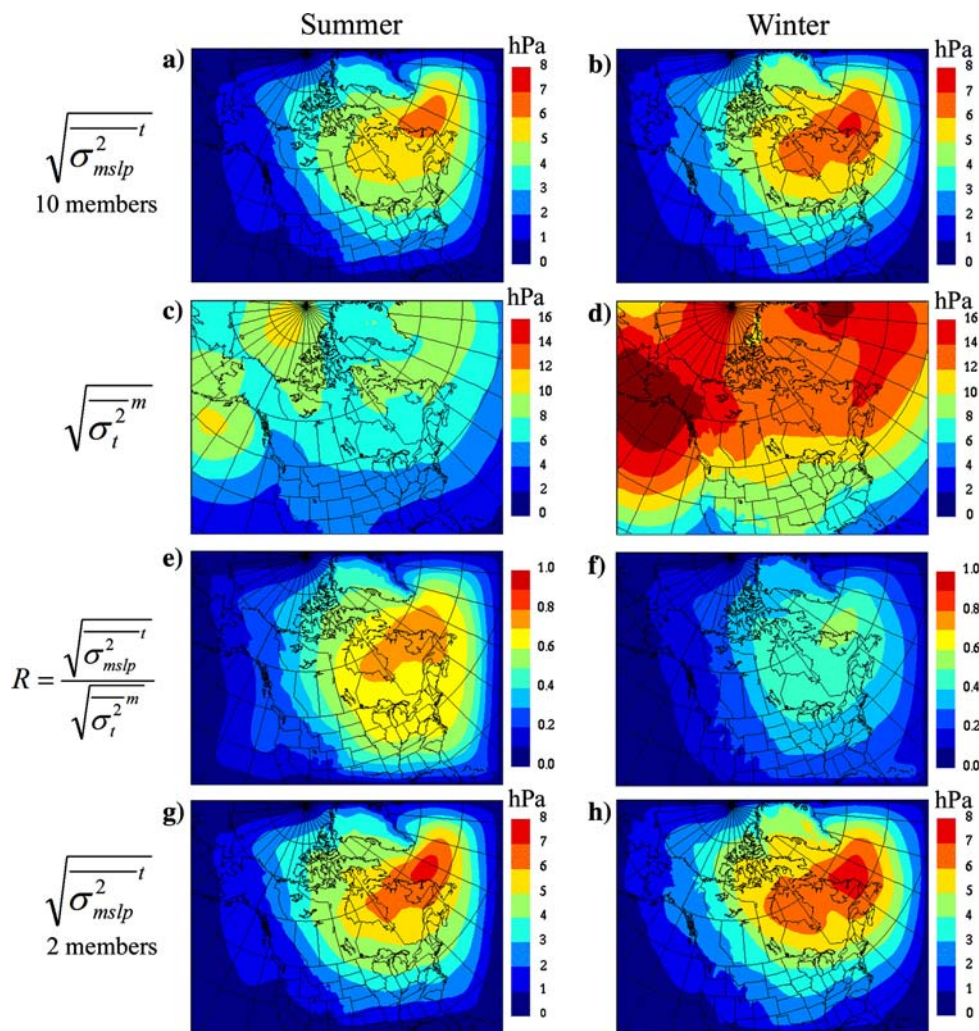
$$\sigma_t^2(i, j, k, m) = \frac{1}{N-1} \sum_{t=1}^N \left( X(i, j, k, t, m) - \overline{X(i, j, k, m)} \right)^2 \tag{8}$$

where the  $^{-t}$  operator refers to a climate time average and  $N$  is the number of archived time steps over the period of interest. The ensemble mean of  $\sigma_t^2$  can be computed to take in consideration the temporal variance of all members  $M$  simultaneously.

$$\overline{\sigma_t^2}^m(i, j, k) = \frac{1}{M} \sum_{m=1}^M \sigma_t^2(i, j, k, m). \tag{9}$$

Figure 4c, d present the square root of the ensemble-mean transient-eddy variance ( $\sqrt{\overline{\sigma_t^2}^m}$ ) for 1980–1989 for MSLP in summer and in winter, respectively. Large values of  $\sqrt{\overline{\sigma_t^2}^m}$  are generated north of the 50th parallel with the values largest near the coasts, along the storm track. In addition, winter values of  $\sqrt{\overline{\sigma_t^2}^m}$  are almost twofold larger than in summer. The ratio of the RCM IV (Fig. 4a, b) over the transient-eddy variability (Fig. 4c, d) should tend toward a value of 1 when the IV in the RCM is close to its maximum value. In such a situation, a RCM behaves in a similar way to a GCM, which means that the evolution of the RCM is independent from its lateral boundary forcing and that the correlation between the ensemble members is close to zero. This ratio is closer to 1 in summer (Fig. 4e) than in winter (Fig. 4f), reaching a maximum of nearly 0.8 in the north of the Québec Province in summer. These results suggest that the members are less controlled by the LBC in summer than in winter, and that each member is more uncorrelated from one another in summer than in winter. Even if the absolute values of winter IV (Fig. 4b) are slightly larger than in summer (Fig. 4a), the relative IV is smaller in winter (Fig. 4f) than in summer (Fig. 4e). The larger values of the IV in winter on Fig. 4b are caused by the larger temporal variance in winter (Fig. 4d) than in summer (Fig. 4c).

**Fig. 4** Square root of the time average inter-member variance ( $\sqrt{\sigma_{mslp}^2}$ ) for 1980–1989 with ten members for the mean-sea-level-pressure (hPa) in **a** summer and **b** winter. Square root of the ensemble-mean transient-eddy variance  $\sqrt{\sigma_t^m}$  from 1980–1989 in **c** summer and **d** winter. Ratio between  $\sqrt{\sigma_{mslp}^2}$  (**a–b**) and  $\sqrt{\sigma_t^m}$  (**c–d**) in **e** summer and **f** winter.  $\sqrt{\sigma_{mslp}^2}$  with two members in **g** summer and **h** winter



Another tool that can estimate the IV is the time correlation between the members that measure the independence between the members. Since our ensemble contains ten members and that time correlation can only be estimated from a pair of members, a suitable coefficient consists in the average time correlation of five pairs of member. For MSLP, this coefficient (not shown) has a spatial distribution similar to the relative internal variability estimated with the ratio describes above (Fig. 4e, f). The average time correlation for the five pairs of member for MSLP start from one at the boundaries, meaning that the IV is low, and decreases from west to east. It reaches a minimum value of 0.3 in summer and 0.7 in winter in the north of Quebec. In this region in summer, the RCM simulation behave close to a GCM where each simulation is almost independent from the driving field forcing, shared by each RCM simulation, which tend to increase the time correlation between simulations.

The 1980–1989 IV climatology ( $\sqrt{\sigma_{mslp}^2}$ ) has also been estimated from a two-member ensemble (Fig. 4g, h) instead

of ten (Fig. 4a, b). Randomly selected pairs showed similar results (not shown). The IV computed with two members (Fig. 4g, h) is similar to the one estimated using a ten-member ensemble (Fig. 4a, b). These results differ from the study of Alexandru et al. (2007), where a minimum of ten members was required to obtain a robust estimation of the spatial distribution of the IV for one season. This can be explained by the longer recording time ( $10 \times 3$  months) we used to compute the spatial distribution, compared to the  $1 \times 3$  months of Alexandru et al. (2007). It seems that a pair of 10-year members is as effective as a ten-member ensemble for a single season (both 30 months per season) to provide a good estimation of the IV. The longer recording time that we used increases the sample size and filters the intermittent inter-member departures. A two-member ensemble of 10 years seems therefore sufficient to obtain a good estimation of the spatial distribution of the IV for MSLP for summer and winter seasons.

The 1980–1989 climatology of IV for PCP ( $\sqrt{\sigma_{pcp}^2}$ ) is highly conditioned by the amount and frequency of

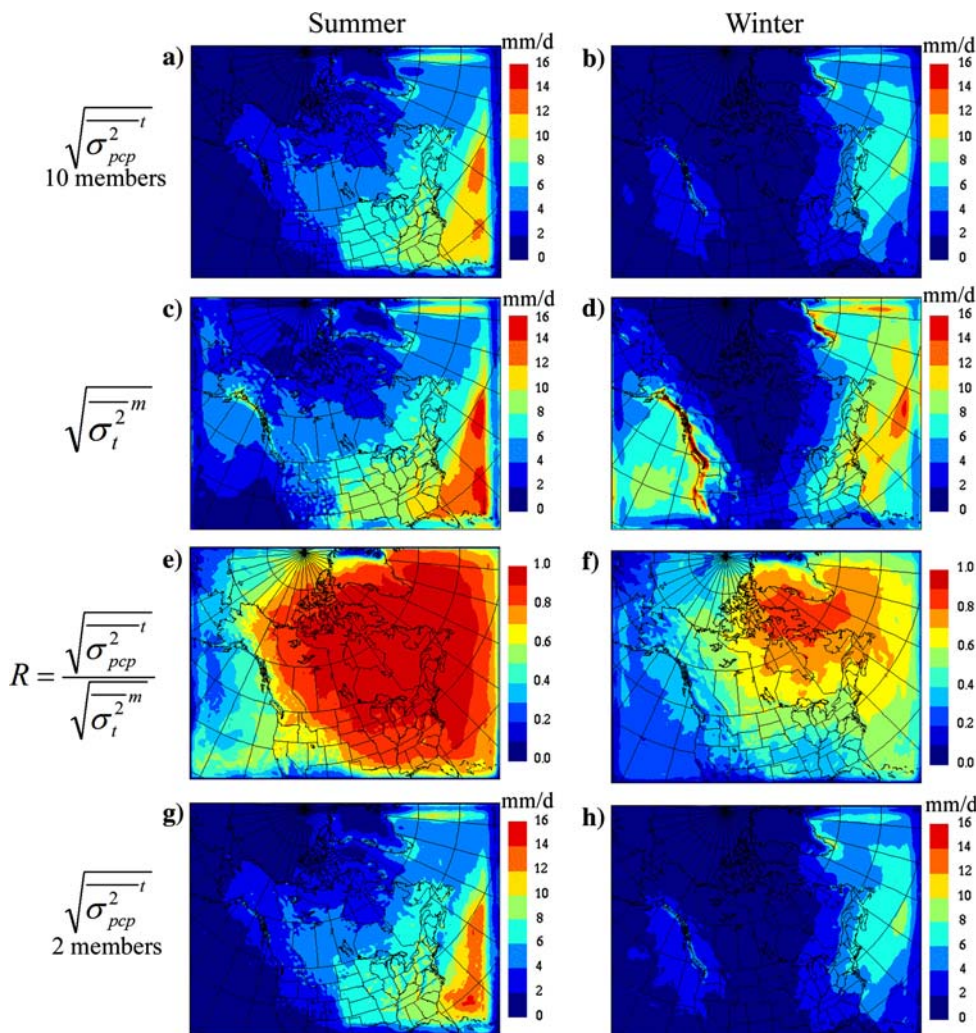


weather events taking place during each season. In summer (Fig. 5a), the strong PCP events in the southeast of the United States show a large variability of solutions in each member (Fig. 5a). The other maximum in the northeast of the domain seems to be related to the lateral boundary forcing where the RCM generates PCP when brought back to the driving field at the boundary, as imposed by the nesting. This artificial PCP seems to behave differently in each simulation and could explain the large IV near the northeast boundaries. The IV in winter (Fig. 5b) is smaller than in summer (Fig. 5a) due to the weaker PCP. There is an IV maximum in winter on the west coast of Canada where large precipitations occur. The IV is weak in the periphery of the domain in summer and in winter according to the nine grid-points nudging zone where each simulation are forced back to the driving field.

The analysis with the IV and transient-eddy variance is repeated for the PCP. The spatial distribution of the square root of the ensemble-mean transient-eddy variance

$(\sqrt{\sigma_t^m})$  for summer (Fig. 5c) is close to the climatology of the IV  $(\sqrt{\sigma_{pcp}^t})$  (Fig. 5a). This is more obvious in Fig. 5e, where the normalized IV (square root of the inter-member variance divided by the square root of the transient-eddy variance) is close to 1 over most of the domain, meaning that the domain is large enough for the PCP IV to reach its maximum value. Therefore, the time evolution of PCP in summer in a large RCM domain can become totally uncorrelated between the members of the ensemble. In summer, the boundaries seem to have very weak influence on the simulation of PCP everywhere in the domain, except near the inflow boundary. However, Fig. 5f shows a different behaviour in winter, where the ratio is much smaller. This can be explained by the generation of winter PCPs by large-scale synoptic systems, which are well correlated in each member. As for the MSLP, the spatial distribution of the average time correlation between the five pairs of members for PCP (not shown) is similar to the spatial

**Fig. 5** Square root of the time average inter-member variance  $(\sqrt{\sigma_{pcp}^t})$  for 1980–1989 with ten members for the precipitation (mm/day) in **a** summer and **b** winter. Square root of the ensemble-mean transient-eddy variance  $\sqrt{\sigma_t^m}$  from 1980–1989 in **c** summer and **d** winter. Ratio between  $\sqrt{\sigma_{pcp}^t}$  (**a–b**) and  $\sqrt{\sigma_t^m}$  (**c–d**) in **e** summer and **f** winter.  $\sqrt{\sigma_{pcp}^t}$  with two members in **g** summer and **h** winter



distribution of the ratio (Fig. 5e, f). It reaches rapidly 0 in the interior of the domain in summer and has a minimum of 0.2 in winter, south of the Greenland. Still as for the MSLP, the estimation of the IV using two members (Fig. 5g, h) closely reproduced the values obtained using ten members (Fig. 5a, b). Again, it seems that a pair of members of 10 years is sufficient to estimate the spatial distribution of the IV for PCP.

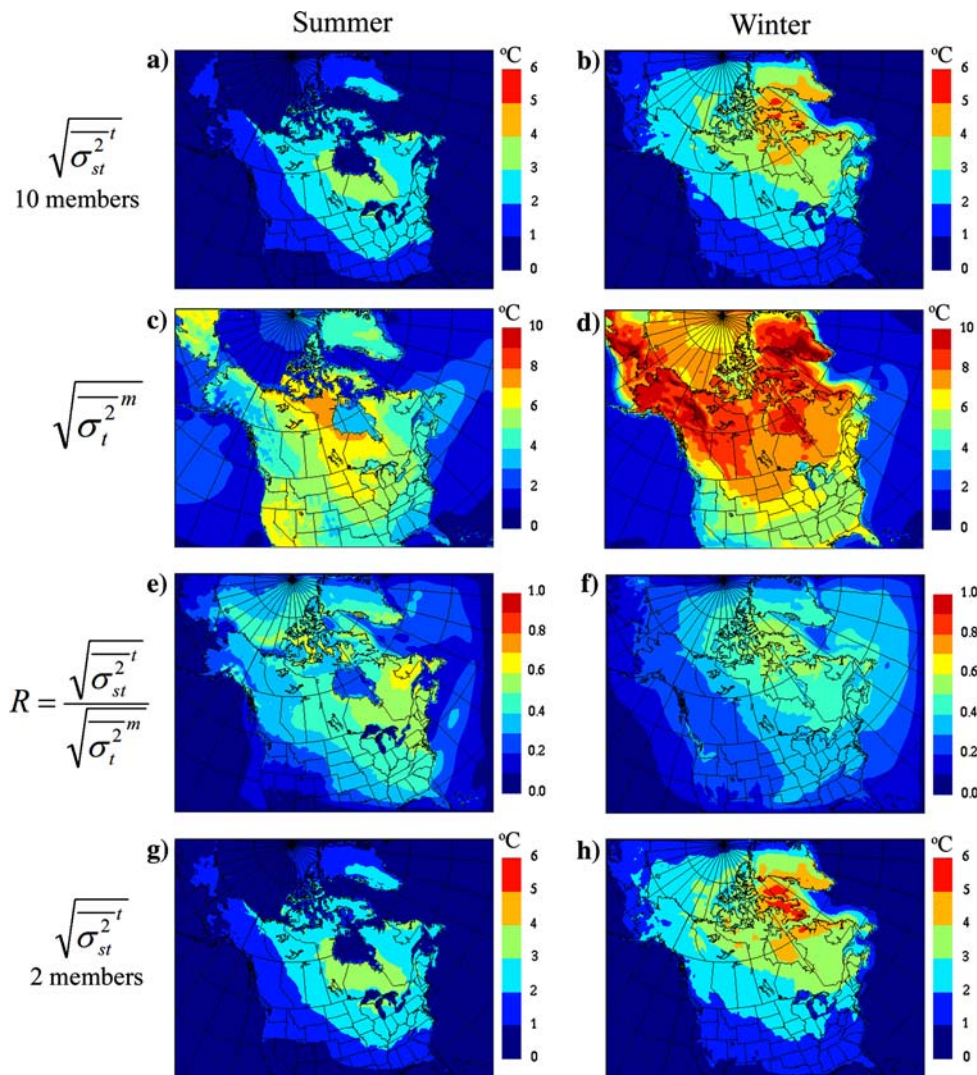
The 1980–1989 summer climatology of the IV for ST shows large values over northern Ontario and around the Hudson Bay (Fig. 6a). In winter, large values are seen over the Canada Arctic (Fig. 6b). In summer, the prescribed sea-surface temperature (SST) over the Hudson Bay from the AMIP data limits the IV of the ST. In winter, when the Hudson Bay is ice covered, the ST is prognostic in the CRCM and is therefore subjected to IV. Larger values of the transient-eddy variance in summer (Fig. 6c) are located over California and around the Hudson Bay. The strong

values of the transient-eddy variance in winter, located in North Canada and on the West Coast, result from the strong cyclonic atmospheric activity (Fig. 6d). As for the two previous variables, the ratio of the IV over the temporal variability is larger in summer (Fig. 6e) than in winter (Fig. 6f) and shows the largest values over the North of Québec. As for MSLP, this region is weakly controlled by the driving field, being far from the inflow boundary, thus allowing large variability between the members. As for the other two variables, computation of the 1980–1989 IV climatology using two members (Fig. 6g, h) showed similar results to the ones computed with ten members (Fig. 6a, b).

### 3.5 Influence of the IV over climate estimations

Alexandru et al. (2007) showed that the IV could have an impact on seasonal averages. However, it has also been

**Fig. 6** Square root of the time average inter-member variance ( $\sqrt{\sigma_{st}^{2t}}$ ) for 1980–1989 with ten members for the screen temperature (°C) in **a** summer and **b** winter. Square root of the ensemble-mean transient-eddy variance  $\sqrt{\sigma_t^{2m}}$  from 1980–1989 in **c** summer and **d** winter. Ratio between  $\sqrt{\sigma_{st}^{2t}}$  (**a–b**) and  $\sqrt{\sigma_t^{2m}}$  (**c–d**) in **e** summer and **f** winter.  $\sqrt{\sigma_{st}^{2t}}$  with two members in **g** summer and **h** winter



suggested that even if RCM simulations generate different time evolution in their solution, the net effect of the IV on the computed climate is small (Giorgi and Bi 2000; Caya and Biner 2004). In a recent work, de Elía et al. (2008) established a relationship between the differences in the simulated climate caused by the IV and the length of the averaging period. They estimated that, for variables uncorrelated in time and showing a weak spatial correlation, the root-mean-square-difference between two time-average fields decreases with the square root of the averaging period. From this perspective, the IV could be associated to white noise where its effect diminishes with the period length over which the climate is computed. This is also similar to the estimation of two time averages, which tend to converge as the sample size increase with the averaging period.

To evaluate the global impact of the IV on the 1980–1989 climatology of a meteorological variable, we computed the square root of the variance between the climate of each member of the ensemble ( $\sqrt{\sigma_X^2}$ ) following Eq. (5).

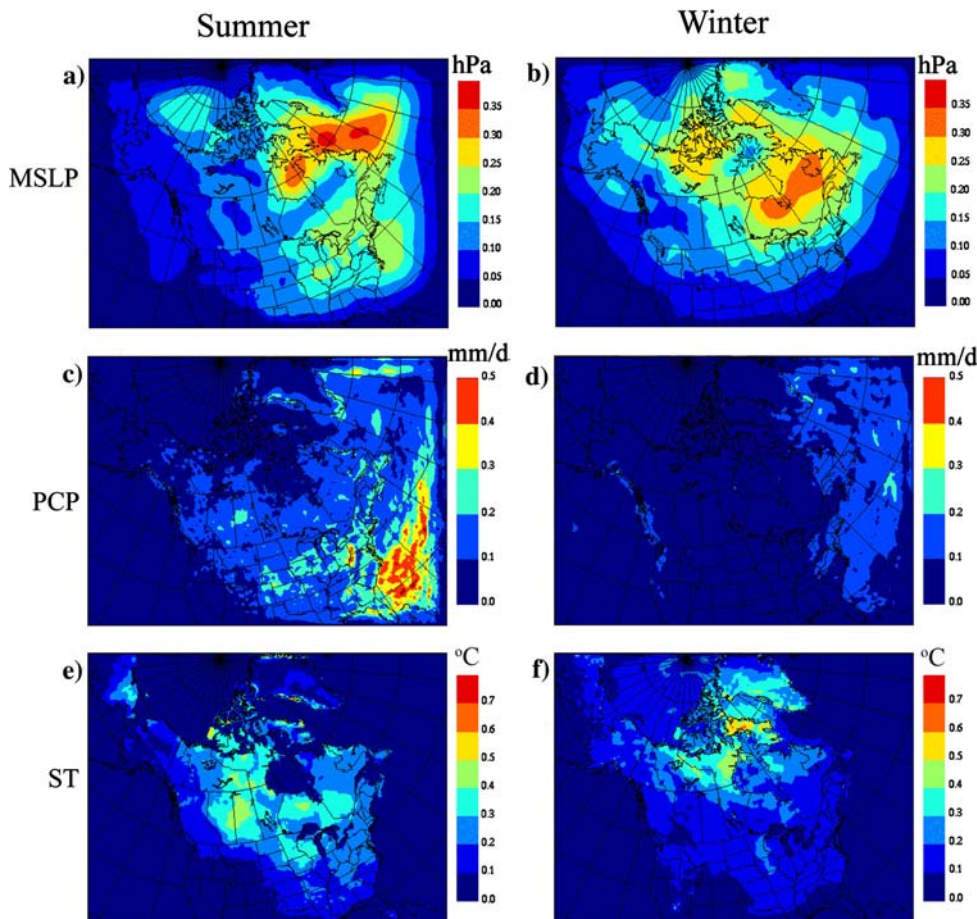
Figure 7a shows that the values of  $\sqrt{\sigma_X^2}$  for MSLP are larger on the Hudson Bay and the Labrador Sea during summer, with a second maximum over New England. It is

important to note that the variance of the climate ( $\sqrt{\sigma_X^2}$ ) (Fig. 7a, b) is very small with respect to the climate of variance ( $\sqrt{\sigma_X^2}$ ) (Fig. 4a, b). In winter, the maximum variance in the climate of MSLP is found over the Canadian Shield (Fig. 7b). The spatial distribution of  $\sqrt{\sigma_X^2}$  (Fig. 7a, b) is different from  $\sqrt{\sigma_X^2}$  (Fig. 4a, b).

For PCP, the large values of  $\sqrt{\sigma_X^2}$  estimated in the southeast USA during summer (Fig. 7c) are co-localized with the large values of  $\sqrt{\sigma_X^2}$  (Fig. 5a), but they are noisier. The large values of  $\sqrt{\sigma_X^2}$  estimated in summer for the ST are located over Saskatchewan (Fig. 7e), where small values of  $\sqrt{\sigma_X^2}$  are seen over that same region (Fig. 6a). Altogether, these data suggest that the longer time scale of the deep soil, associated with feedback processes, could drive some members away from the ensemble mean over a long period of time.

Climate change simulations and observations statistics are usually based on 30 years. Since our simulations span only 10 years, a possible way to estimate the variance

**Fig. 7** Square root of the variance between the 10-year climate of each member of the ensemble ( $\sqrt{\sigma_X^2}$ ) from 1980 to 1989 for the mean-sea-level pressure (MSLP; hPa) with ten members in **a** summer and **b** winter. Computation is repeated for **c–d** the precipitation (PCP; mm/day) and **e–f** the screen temperature (ST; °C)



between 30-year climates values in our analysis could be through extrapolation. We use here an approach similar to that used by de Elía et al. (2008). They showed that using the variance of a sample mean ( $S_z^2$ ) for a collection of independent and identically distributed random variables (von Storch and Zwiers 2001), we can write

$$S_z^2 = \frac{1}{N} S_z^2 \quad (10)$$

where  $N$  is the number of members in the sample and  $S_z^2$  is the variance of the independent variable  $z$ . This means that the sample mean, as an estimator of the population mean, has an uncertainty that is proportional to the population variance and inversely proportional to the size of the sample.

From Eq. (10), we can get

$$S_z^2 = 30 \times S_{z_{30}}^2 = 10 \times S_{z_{10}}^2 \quad (11)$$

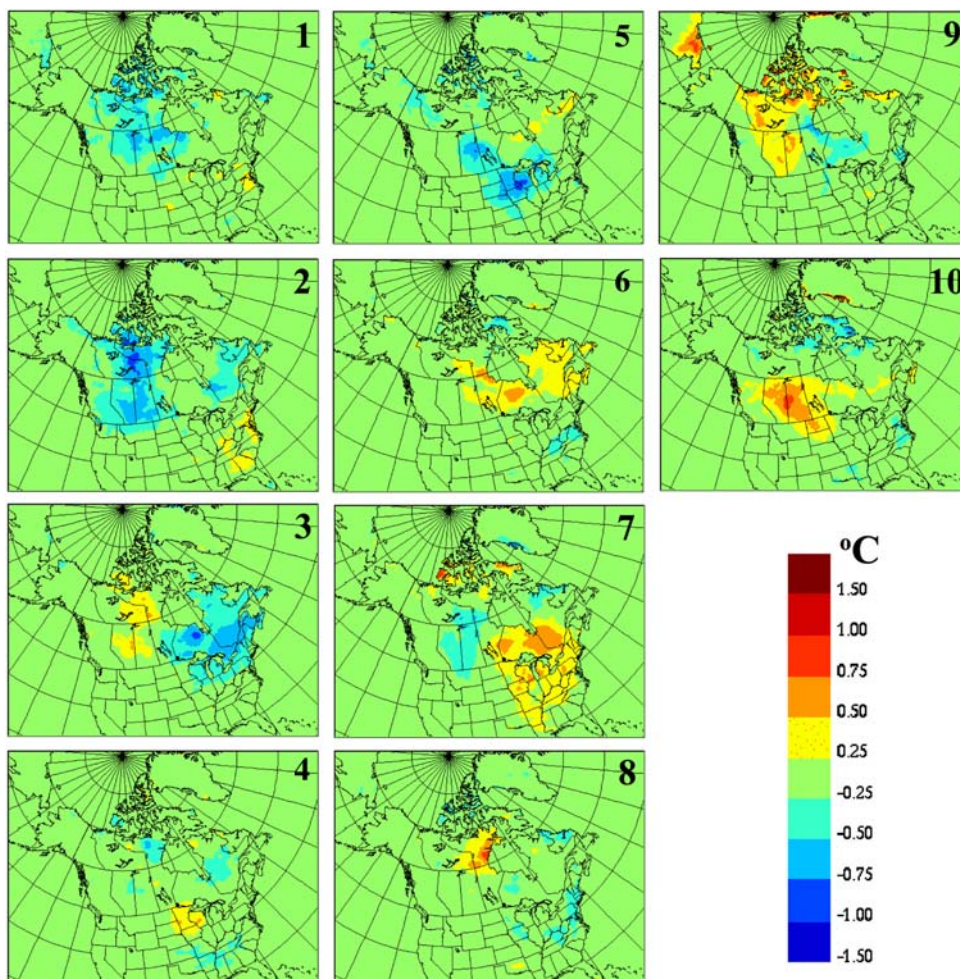
$$S_{z_{30}} = \frac{1}{\sqrt{3}} \times S_{z_{10}}. \quad (12)$$

The values computed in Fig. 7 correspond to  $S_{z_{10}}$ . Thus, an approximation of the square root of the variance with a

30-year climate ( $S_{z_{30}}$ ) can be obtained by multiplying the values on Fig. 7 by  $\frac{1}{\sqrt{3}}$ . Since our ensemble contains only a sample of ten 10-year climates, the spatial distribution is not very robust and might change with more members or with 30-year climates.

In Fig. 8, we plotted the departure of the 1980–1989 climate of each member from the 1980–1989 climate of the ensemble mean for summer ST. In summer, members 1, 2, 5 and 7 are below the ensemble mean over Saskatchewan, while members 3, 6, 9 and 10 are above the ensemble mean. In this region, the large values of  $\sigma_X^2$  not only result from one extreme member, but seem to oscillate between two modes. The soil water content, having a longer time response, might create some memory that can extend from year to year, enhancing the variability between members for sensible regions, such as Saskatchewan which is dry and close to the Rocky Mountains. The departures from  $\pm 1$  °C observed in certain regions, like Saskatchewan for members 2 and 10 (Fig. 8), are not negligible for 10-year averages in summer. One should keep in mind that these departures are generated from very small perturbations in the initial conditions and that each member is a plausible

**Fig. 8** Departure of 1980–1989 time average for each member (indicated in the *top right* of the figures) from the 1980–1989 time average of the ensemble mean for screen temperature (°C) in summer



solution for the same set of LBC forcing. According to our experimental setup, which uses the perfect-model approach, the anomaly observed should be considered as the minimum uncertainty to take into account for a 10-year simulation. Finally, we showed that the IV could have a significant impact on the 10-year climatology of meteorological variables, which differs from previous studies (Giorgi and Bi 2000; Caya and Biner 2004). The large IV estimated due to the large domain size, especially in winter, should be responsible for this different result.

#### 4 Conclusions

This work extends previous studies on RCM's internal variability (IV) that were limited by small ensembles, short simulations and the use of small domains. To push these limits, a ten-member ensemble of 10-year simulations was constructed over a large domain covering North America. To generate the ensemble, the simulations were launched with perturbations in their initial conditions. All members of the ensemble used the same set of time-dependent lateral boundary conditions taken from the NCEP reanalyses and the same prescribed ocean surface boundary conditions (SSTs and sea ice) taken from the AMIP data.

The IV was estimated as the variance between the ensemble members. In a first analysis, the time evolution of the IV was investigated using the domain average inter-member variance over the 10-year period. The results showed that the IV has no long-term tendency and seems to fluctuate in time according to the synoptic situation within the domain. The IV did not exhibit a distinct annual cycle for mean-sea-level pressure and screen temperature, a conclusion at variance with previous studies over mid-latitude that showed a clear annual cycle in the IV with small values in winter and large values in summer. It seems that the increase domain size reduces the control of the driving field on the RCM simulations and enhances the IV, especially in winter. The annual cycle for PCP, with large values in summer and small values in winter, is in agreement with previous studies.

In a second analysis, we examined the spatial distribution of the IV with its 10-year climatology. The analysis shows that the IV is not uniformly distributed within the domain, with larger values for mean-sea-level pressure in the northeast of the domain near the outflow boundary. Small values of the IV were found on the western side of the domain near the inflow boundary. The normalization of the time-average inter-member variance with the transient-eddy variance (which is an estimation of the maximum value of the IV) showed that the relative IV is closer to its maximum in summer than in winter. In the region of larger IV for summer period, the RCM behaves similarly to a

GCM in the sense that the meteorological events are not synchronized despite the forcing applied at the lateral boundaries. The higher control from the driving fields in winter explains why similar absolute IV values for MSLP are estimated for summer and winter periods, despite the larger transient-eddy variance in winter fields.

Finally, the influence of the IV on the 10-year climatology was examined using the inter-member variance between the climates of each member. The largest variances for the climate of each member were not always located in the region with largest climatological IV. The small size of our ensemble and/or feedback processes associated to long time responses of the soil variables could be responsible for the differences in the climate of the ensembles members. A larger ensemble is required to fully address these questions.

This work looked at the IV using a specific RCM and a specific experimental configuration. We must be careful in generalizing the conclusions of this work for other RCMs or configurations. It would be interesting to test whether large spreads between the members of an ensemble are associated with a drift in the RCM circulation, with respect to that of the driving data. It would also be interesting to identify the spatial scales affected by the internal variability. Finally, this study could be repeated using large-scale nudging, which is widely used on such large domains.

**Acknowledgments** This work is part of the Ph.D. thesis of Philippe Lucas-Picher in Environmental Sciences at Université du Québec at Montréal. The authors wish to thank Claude Desrochers and Mourad Labassi for maintaining a user-friendly local computing environment at the Ouranos Consortium. A special thank is directed to Sébastien Biner and Samuel Somot, which generously devoted time to the discussion of some sections of the manuscript. This work was carried out as part of the research programs of the Canadian Climate Variability Research Network (CLIVAR), the Canadian Network for Regional Climate Modelling and Diagnostics (CRCMD) and the Ouranos Consortium. It was financially supported by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), the National Science and Engineering Research Council of Canada (NSERC) and the Ouranos Consortium. Finally, the authors would like to thank Dr. Maryse Picher for her careful revision that has improved the readability of the manuscript.

#### References

- Alexandru A, de Elía R, Laprise R (2007) Internal variability in regional climate downscaling at the seasonal time scale. *Mon Weather Rev* 135:3221–3238
- Biner S, Caya D, Laprise R, Spacek L (2000) Nesting of RCMs by imposing large scales. In: Ritchie H (ed) *Research activities in atmospheric and oceanic modelling*. WMO/TD, No. 987, Report No. 30, pp 7.3–7.4
- Caya D, Biner S (2004) Internal variability of RCM simulations over an annual cycle. *Clim Dyn* 22:33–46
- Caya D, Laprise R (1999) A semi-implicit semi-lagrangian regional climate model: the Canadian RCM. *Mon Weather Rev* 127:341–362

- Christensen OB, Gaertner MA, Prego JA, Polcher J (2001) IV of regional climate models. *Clim Dyn* 17:875–887
- Davies HC (1976) A lateral boundary formulation for multilevel prediction models. *Q J R Meteorol Soc* 102:405–418
- de Elía R, Caya D, Frigon A, Côté H, Giguère M, Paquin D, Biner S, Harvey R, Plummer D (2008) Evaluation of uncertainties in the CRCM-simulated North American climate. *Clim Dym* 30:113–132. doi:10.1007/s00382-007-0288-z
- Gates WL et al (1999) An overview of the results of the Atmospheric Model Intercomparison Project. *Bull Am Meteorol Soc* 80:29–55
- Giorgi F, Bi X (2000) A study of IV of regional climate model. *J Geophys Res* 105:29503–29521
- IPCC, Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (2001) *Climate change 2001: the scientific basis*. Cambridge University Press, Cambridge, p 881
- IPCC, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (2007) *Climate change 2007: the physical science basis*. In: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, New York, NY, USA, p 996
- Jones RG, Murphy JM, Noguer M (1995) Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Q J R Meteorol Soc* 121:1413–1449
- Kalnay E et al (1996) The NCEP/NCAR 40-year Reanalysis Project. *Bull Am Meteorol Soc* 77:437–471
- Lucas-Picher P, Caya D, Biner S (2004) RCM's IV as function of domain size. In: Côté J (ed) *Research activities in atmospheric and oceanic modelling*. WMO/TD, vol 1220, no. 34, pp 7.27–7.28
- McFarlane NA, Boer GJ, Blanchet JP, Lazare M (1992) The Canadian climate centre second-generation general circulation model and its equilibrium climate. *J Clim* 5:1013–1044
- Mearns L (2004) North American Regional Climate Change Assessment Program (NARCCAP): a Multiple AOGCM and RCM climate scenario project over North America, AGU Fall Meeting, San Francisco, USA
- Miguez-Macho G, Stenchikov GL, Robock A (2004) Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *J Geophys Res* 109:D13104. doi:10.1029/2003JD004495
- Plummer D, Caya D, Côté H, Frigon A, Biner S, Giguère M, Paquin D, Harvey R, de Elía R (2006) Climate and climate change over North America as simulated by the Canadian regional climate model. *J Clim* 19:3112–3132
- Riette S, Caya D (2002) Sensitivity of short simulations to the various parameters in the new CRCM spectral nudging. In: Ritchie H (ed) *Research activities in atmospheric and oceanic modelling*. WMO/TD, No. 1105, Report No. 32, pp 7.39–7.40
- Rinke A, Dethloff K (2000) On the sensitivity of a regional Arctic climate model to initial and boundary conditions. *Clim Res* 14:101–113
- Rinke A, Marbaix P, Dethloff K (2004) IV in Arctic regional climate simulations: case study for the Sheba year. *Clim Res* 27:197–209
- von Storch H, Zwiers FW (2001) *Statistical analysis in climate research*. p 484
- von Storch H, Langenberg H, Feser F (2000) A spectral nudging technique for dynamical downscaling purposes. *Mon Weather Rev* 128:3664–3673
- Takele ES, Roads J, Rockel B, Gutowski WJ, Arritt RW, Meinke I, Jones CG, Zadra A (2007) Transferability intercomparison: an opportunity for new insight on the global water cycle and energy budget. *Bull Am Meteorol Soc* 88:375–384
- Vannitsem S, Chomé F (2005) One-way nested regional climate simulations and domain size. *J Clim* 18:229–233
- Weisse R, Heyen H, von Storch H (2000) Sensitivity of a regional atmospheric model to a sea state dependent roughness and the need of ensemble calculations. *Mon Weather Rev* 128:3631–3642
- Wu W, Lynch A, Rivers A (2005) Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions. *J Clim* 18:917–933