

# Impact of spectral nudging and domain size in studies of RCM response to parameter modification

Leo Separovic · Ramon de Elía · René Laprise

Received: 21 October 2010 / Accepted: 3 April 2011 / Published online: 20 April 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** The paper aims at finding an RCM configuration that facilitates studies devoted to quantifying RCM response to parameter modification. When using short integration times, the response of the time-averaged variables to RCM modification tend to be blurred by the noise originating in the lack of predictability of the instantaneous atmospheric states. Two ways of enhancing the signal-to-noise ratio are studied in this work: spectral nudging and reduction of the computational domain size. The approach followed consists in the analysis of the sensitivity of RCM-simulated seasonal averages to perturbations of two parameters controlling deep convection and stratiform condensation, perturbed one at a time. Sensitivity is analyzed within different simulation configurations obtained by varying domain size and using the spectral nudging option. For each combination of these factors multiple members of identical simulations that differ exclusively in initial conditions are also generated to provide robust estimates of the sensitivities (the signal) and sample the noise. Results show that the noise magnitude is decreased both by reduction of domain size and the spectral nudging. However, the reduction of domain size alters some

sensitivity signals. When spectral nudging is used significant alterations of the signal are not found.

**Keywords** Regional climate models · Parameter perturbations · Internal variability · Spectral nudging · Domain size

## 1 Introduction

Nested limited-area Regional Climate Models (RCMs) are models that dynamically downscale global General Circulation Model (GCM) simulations or objective analyses to high-resolution computational grids, using a high-resolution representation of the surface forcing and model dynamics. RCMs require the information on some prognostic variables as their lateral boundary conditions (LBC). The choices of integration domains and nesting techniques are free parameters of RCMs. The optimal integration domain depends on the particular situation, although there are some general recommendations that can facilitate user's judgment (e.g., Laprise et al. 2008). For example, Leduc and Laprise (2008) showed that the use of a too small domain could result in the simulations being deficient in fine-scale variance. It has been also noted that in large continental-scale domains RCM large-scale variables can considerably drift from the driving fields, which can then result in appearance of large spurious gradients in the vicinity of the outflow boundaries. Spectral nudging (SN; Von Storch et al. 2000; Biner et al. 2000) has been employed to ensure that the model solution remains close to the large-scale components of the driving fields over the entire domain. However, the use of spectral nudging remains an open issue. Alexandru et al. (2009) raised concern that the application of the SN could suppress the

---

L. Separovic (✉) · R. de Elía · R. Laprise  
Centre pour l'Étude et la Simulation du Climat à l'Échelle Régionale (ESCER), Université du Québec à Montréal,  
201 avenue Président-Kennedy, local PK-2315,  
Montréal, QC H2X 3Y7, Canada  
e-mail: separovi@sca.uqam.ca

L. Separovic · R. Laprise  
Université du Québec à Montréal (UQAM),  
Montréal, QC, Canada

R. de Elía  
Consortium Ouranos, Montréal, QC, Canada

proper generation of fine-scale features. However, Colin et al. (2010) did not find SN to be detrimental on the modelling of extreme precipitation.

The choice of the integration domain and the use of spectral nudging can have a large impact on the RCM internal variability. Internal variability arises due to the non-linear, chaotic nature of atmospheric models: any perturbation; however, small it is in magnitude, provokes the trajectories of the model solution in the phase space to diverge in time. In autonomous Global Circulation Models (GCMs) the difference between two simulations conducted with the same model but departing from initially slightly different states is on average as large as the difference between two randomly chosen GCM states, given a specific season. Internal variability also emerges in RCMs but, typically, it is smaller than in GCMs; the advection of information prescribed as the LBC keeps the evolution of the RCM internal variability somewhat bounded (e.g., Giorgi and Bi 2000; Caya and Biner 2004). However, intermittently in specific areas of the integration domain it can achieve values as large as in GCMs (Alexandru et al. 2007). Its time evolution appears to depend on the synoptic situation enforced by the driving fields (e.g., Lucas-Picher et al. 2008b; Nikiema and Laprise 2010) and is scale selective (Separovic et al. 2008). Reduction of domain size or the application of spectral nudging can both considerably reduce internal variability in RCMs (Alexandru et al. 2009). Thus, the average amplitude of internal chaotic variations appears to be in RCMs, to a certain extent, a controllable parameter. This fact may be of particular interest in studies oriented to RCM testing and modification.

The sensitivity of a RCM to any change in its structure and configuration, such as a modified parameterization or a perturbation of its tuneable parameters, generally consists of the response of the simulated variables to the modification (signal), as well as of internal variability noise. Since the work of Weisse et al. (2000) it has been widely acknowledged that estimation of the signal in the temporal evolution of the RCM variables requires ensemble simulations that can be generated, for example, by imposing perturbations to the initial conditions of both the control and the modified model versions. Internal variability deviations are partly filtered in the ensemble mean depending on the ensemble size, as the variance of the sample mean of a collection of independent and identically distributed random variables is inversely proportional to the sample size (e.g., Von Storch and Zwiers 1999). When the signal is small or the internal variability is large, ensembles of large size are needed in order to obtain statistically significant estimates of the simulation differences resulting from the model modifications. For sufficiently long integration times, internal variability deviations are

substantially reduced in the time average. However, estimation of the time averages computed over shorter periods from years to a decade also necessitates sampling of the internal variability deviations, since it can be still non-negligible in the time average of the single model run, especially for fine-scale variables such as precipitation (de Elia et al. 2008; Lucas-Picher et al. 2008a, b). When considering the difference between the time averages in the control and a modified model version, the variance introduced by the internal variability is twice as large as that in the time average in each model version, due to the aggregation of error through the difference terms.

Providing statistically significant estimates by means of ensemble simulations or longer integration periods for the control and modified model versions is hence computationally time consuming. While this issue might be of little relevance when the RCM is to be tested for a single modification, it can represent a hindrance in studies that require multiple testing of RCM response to modifications of a large number of parameters. This would typically be the situation in deliberate model tuning or in studies that address uncertainty originating in the RCM's adjustable parameters wherein it is essential to identify in a high-dimensional parameter space the plausible parameter perturbations that produce the largest response of the model (e.g., Sexton and Murphy 2003). The underlying methodological issue in such RCM studies is thus to optimize the use of computational resources by finding an appropriate test bed configuration (prototype simulation) that would be as inexpensive as possible in terms of the number of computational points and integration time and that can provide robust estimates of the model response to the modifications.

Our working hypothesis is that suppressing the internal RCM variability by means of domain size reduction or application of SN would allow for quantifying the signal with a smaller ensemble size and help to reduce the computational cost (Alexandru et al. 2007, Weisse and Feser 2003). The application of these methods to reduce internal variability noise requires better understanding of the ways they might alter the signal of RCM sensitivity to modification, e.g., by suppressing its magnitude. Too small domains are generally non-recommended for climate simulations and sensitivity studies because of the spurious effects of the proximity of the lateral boundaries, fine-scale variance deficiency and lack of continental-scale interactions and feedback among the RCM variables (e.g., Jones et al. 1995; Seth and Giorgi 1998; Laprise et al. 2008). Results obtained in such domains are likely to be less realistic and difficult to extrapolate to the operational RCM simulations. However, when studying uncertainties originated in adjustable RCM parameters, a very large number of tests are required and the user may wish to conduct

preliminary tests in a computationally inexpensive small domain. Outside this context the reduction of domain size and SN should not be considered as competing techniques to improve the signal-to-noise ratio since the SN has not been shown to involve similar difficulties.

The manuscript is organized as follows. The model and the modifications performed on the model parameters in order to produce modified model versions and the experiments are described in Sect. 2. The analysis of model sensitivity to modification of parameters within different simulation configurations is carried out in Sect. 3. Summary and conclusions are provided in Sect. 4.

## 2 Experimental design

### 2.1 Model description

The model used in this study is the fifth-generation Canadian Regional Climate Model (CRCM5; Zadra et al. 2008). It is a limited-area version of the Canadian weather forecast model GEM (Côté et al. 1998); the model has a non-hydrostatic option, although this feature is not exploited here. GEM is a grid-point model based on a two-time-level semi-Lagrangian, semi-implicit time discretization scheme. The model includes a terrain-following vertical coordinate based on hydrostatic pressure (Laprise 1992) with 58 levels in the vertical, and the horizontal discretization on an Arakawa C grid (Arakawa and Lamb 1977) on a rotated latitude-longitude grid with a horizontal resolution of approximately 55 km and time step of 30 min. The nesting technique employed in CRCM5 is derived from Davies (1976); it includes a gradual relaxation of all prognostic atmospheric variables toward the driving data in a 10-point sponge zone along the lateral boundaries. The lateral boundary conditions (as well as the initial conditions) are derived from ERA40 reanalysis (Uppala et al. 2005). Ocean surface conditions are prescribed from Atmospheric Model Intercomparison Project (AMIP) data (Fiorino 2004).

### 2.2 Experiments

Modified model versions are obtained by perturbing the CRCM5 physics parameters. Three different model versions are considered: the control version (denoted hereafter as M00) and two perturbed-parameter versions (denoted by M01 and M10) obtained by perturbing one at a time, the following two parameters:

P01—Threshold vertical velocity in the trigger function of the deep convection parameterization (Kain and Fritsch 1990).

P10—Cloud water to precipitation conversion time scale in the large-scale condensation parameterization for stratiform precipitation (Sundqvist et al. 1989; Pudykiewicz et al. 1992).

The values of parameters used in the three model versions are given in Table 1. Two experts that participated in CRCM5 development judged the perturbations as being moderate to strong with respect to their range of variation, given the horizontal resolution.

Three sets of experiments are carried out in this study, all based on simulations conducted over a single year. For every model version multiple perturbed initial-condition ensemble simulations were performed. The initial conditions were perturbed initializing the model from November 01 1992 at 00UTC onward, 24 h apart. All the simulations, regardless of model version and initialization time, end on December 01 1993 at 00UTC. November 1992 is not considered in order to allow the spin-up of the initial differences, thus leaving a 1-year period for the analysis. The number of ensemble members is the same in all three sets; there are 10 members for the standard model version M00 and 5 members per each of the two perturbed-parameter versions M01 and M10; the last column in Table 1 shows the ensemble size per each model version.

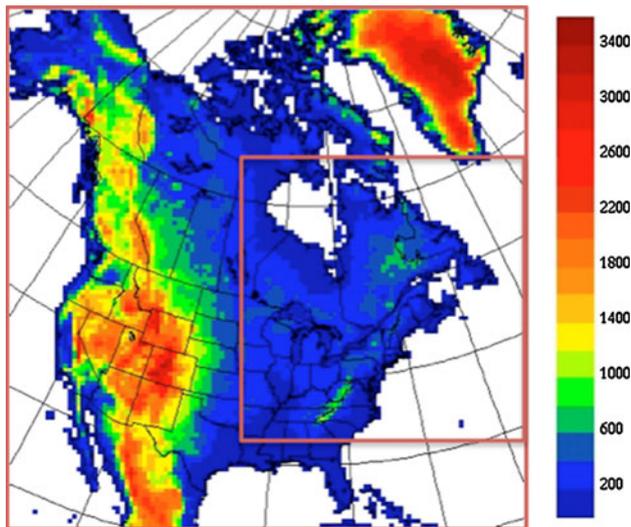
In the first set, denoted as SYNA, the simulations were performed with the three model versions (M00, M01 and M10) over the large continental-scale domain, referred to as NA, consisting of  $120^2$  grid points, and shown in Fig. 1 including the 10-point relaxation zone at the perimeter of the lateral boundaries.

The second set of experiments, denoted as SYSN, is identical to SYNA in terms of its domain (NA; Fig. 1), model versions and number of ensemble members per every model version (Table 1); the only difference is that the spectral nudging (SN) was used. The nudging was only applied to the horizontal wind components, with the truncation at non-dimensional wavenumber 4 ( $\sim 1,500$  km). The SN strength is set to zero below the level of 500 hPa and increases linearly with height, reaching 10% of the amplitude of the driving fields per time step at the top level.

**Table 1** Parameters' settings used in different model versions

Model version	P10 (h)	P01 (m/s)	No. of ensemble simulations
M00	2.8	3.4E−2	10
M01	2.8	6.0E−2	5
M10	10.0	3.4E−2	5

P10 is the time scale of conversion from cloud to precipitable water in the large-scale condensation parameterization; P01 denotes the large-scale vertical velocity threshold in the Kain-Fritsch deep convection trigger function. Model version M00 is used as reference



**Fig. 1** Topography of the two CRCM5 computational domains, including the lateral boundary relaxation zone. The large domain is used in the SYNA and SYSN experiments and the smaller domain in the SYDS experiment

The choices of the truncation wavelength and the vertical profile of the nudging strength reflect the intention not to interfere with the model own interior dynamics at fine and intermediate spatial scales and in the lower half of the model's atmosphere.

The third set of experiments, denoted as SYDS, consists in reducing the domain size. For every model version, the single-year ensemble simulations are generated again, but over a domain of reduced size centred over the province of Quebec (without SN). The domain for the SYDS experiment consists of  $70^2$  grid points and is shown in Fig. 1, including the 10-point sponge zone.

### 3 Results

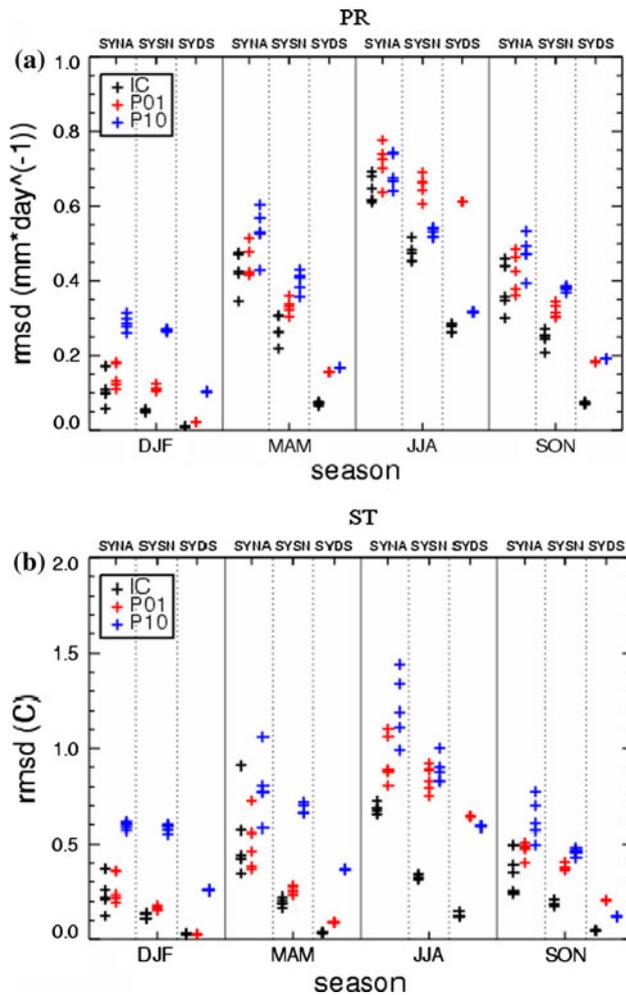
The variables selected for the analysis of results are seasonal-average precipitation and 2 m-temperature. The analysis is focused on the influence of SN and domain size reduction on the model sensitivity to perturbations, internal variability noise and signal-to-noise ratio. This section is organized as follows. Section 3.1 briefly reviews the sensitivities of CRCM5 seasonal averages to perturbations of the initial conditions and parameters, as a function of season and experimental configurations SYNA, SYSN and SYDS. Section 3.2 presents the spatial distribution of the internal variability noise in the three configurations. Sections 3.3 to 3.5 examine the spatial patterns of the sensitivity of CRCM5 seasonal averages to the parameter perturbations (signals), estimated with the difference of ensemble means of the control and modified model versions; these sections also provide the statistical significance

of the sensitivity estimates and compare the signal patterns in the three simulation configurations. Section 3.6 examines the computational cost associated with different simulation configurations in terms of the minimum ensemble size necessary to achieve significant estimates.

#### 3.1 Spread of differences excited by perturbations

We begin the analysis with a brief review of the magnitude of the response of the CRCM5 seasonal averages to the applied parameter perturbations, as a function of the simulation configuration (SYNA, SYSN and SYDS) and season (DJF, MAM, JJA, and SON). For this purpose the square root of the spatially averaged square differences (denoted as rmsd) is computed for the pairs of seasonal averages obtained from the simulations that differ either in the parameters settings (signal) or initial conditions (internal variability). The rmsd excited by the perturbations of parameters are calculated using the pairs of seasonal averages, such that each pair consists of one realization of the control ensemble M00 and one realization of the perturbed-parameter ensemble (M01 or M10). Since the latter have 5 members (see Table 1), 5 pairs were randomly chosen from the 10 members of the reference model, and hence 5 pairs of difference were computed for each parameter perturbation. The rmsd are displayed in Fig. 2 with the 5 “plus” marks coloured in red for the perturbation of the deep convection parameter and the 5 marks in blue for large-scale condensation parameter, for seasonal-average precipitation (a) and 2 m-temperature (b). All rmsd are computed for each configuration over its own domain exclusive of the 10-point wide sponge zone; thus in the SYNA and SYSN experiments, the rmsd is computed over the large domain, while for the SYDS over the small domain in Fig. 1. The rmsd displayed with coloured marks in Fig. 2 are a result of the model response to the parameter perturbations. Internal variability is displayed with black marks in Fig. 2. They represent the rmsd excited by different initial conditions of simulations with otherwise identical model configurations. The rmsd are assessed from the 10 ensemble members of the control model version M00 that are organized in five pairs on a random basis.

Figure 2 shows that all rmsd exhibit an annual cycle with the maximum in summer and minimum in winter. The magnitude of the rmsd illustrates the physical significance of the model response to perturbations. The range of responses for precipitation and 2 m-temperature is 0–0.3 mm/day and 0–0.7°C in winter and 0.3–0.8 mm/day and 0.6–1.5°C in summer, respectively. Also the rmsd are in general the largest in the SYNA set and the smallest in the reduced domain size SYDS set. This holds for the three kinds of perturbations. The SYSN reduces internal variability noise (black marks) but it is less efficient in that



**Fig. 2** The RMS difference between CRCM5 individual simulations for seasonal-average **a** precipitation and **b** 2 m-temperature as a function of the experimental setup and season. The *black marks* display the rmsd in seasonal averages among the ensemble members of the model M00 (Table 1); they are triggered by internal variability and are obtained as follows: from 10 ensemble members 5 pairs of seasonal averages are selected, for each pair the rmsd is plotted. The *coloured marks* show the realizations of the rmsd between ensemble members of M00 and M01 (M10); they are triggered by parameter perturbations (*red*) P01 and (*blue*) P10

than the reduction of domain size (SYDS); this being true for this case and different configurations of both spectral nudging and domain size could yield different results. The plots in Fig. 2 also provide a rule of thumb for the statistical significance of the response of the seasonal averages to the parameter perturbations: if differences between the control and perturbed-parameter model version (red or blue marks) tend to lie above the maximum rmsd due to internal variability noise (black marks), given a season and simulation setup, this suggests the statistical significance of the corresponding model response to the parameter perturbation. As of precipitation (Fig. 2a), all signal rmsd in the SYNA setup are barely above noise level, except for

condensation-related parameter P10 in winter. The SN and reduction of domain size reduce the noise rmsd considerably but also the rmsd due to the parameter perturbations generally decreases. Thus, for precipitation in the SYSN and SYDS sets, the situation with statistical significance is not considerably changed. The exception is in summer when the convection-related parameter P01 produces significant rmsd, especially in the SYDS set. For 2 m-temperature (Fig. 2b) the responses to parameter perturbations are generally more statistically significant. Despite that, when the signal is weak, as P01 in winter, or noise very high, as in spring and summer, the parameter-induced rmsd appear not to be statistically significant. This also implies that the signal-to-noise ratio varies for different CRCM5 variables.

It is difficult to infer from Fig. 2 whether the model response to parameter perturbations is on average smaller in the SYSN and SYDS sets or whether the lower rmsd in this set are a sole effect of reducing internal variability. We investigate this issue more thoroughly in the next subsections. Further, it can be seen that in winter (DJF), the perturbation P10 produces considerable and significant signals for both precipitation and temperature, while P01 produces a smaller response that is difficult to distinguish from internal variability. Perturbation P01 is related to the deep convection parameterization that is rarely active in winter over land. This perturbation produces a considerable and significant response over land only in the warmer half of the year.

The spatially averaged square differences may hide important information on the local behaviour of the CRCM5 response to the perturbations. In the following we begin the analysis of spatial patterns by first examining the noise level and then the spatial patterns of the model response to parameter perturbations are compared in the three experimental sets as a function of the parameter perturbation and season.

### 3.2 Noise level in the differences

Instead of using a standard measure of noise in seasonal averages (e.g., ensemble standard deviation in the control model M00) that would quantify the internal variability in CRCM variables, we rather analyze the internal variability of the model responses to the perturbations of parameters. This way, every difference computed between an ensemble member of a perturbed-parameter model (M01 or M10) and a member of the control model ensemble M00 is a sample of the model response to the parameter perturbation. Internal variability noise in estimates of the CRCM5 response can be measured with the variability in that sample. Since the variance of the difference of the two mutually independent identically distributed (iid) random

variables is equal to the sum of the variances of the two variables, the standard deviation of the sample of differences can be estimated as

$$\sigma = \left[ \frac{1}{M_x - 1} \sum_{m=1}^{M_x} (\bar{x}_m - \langle \bar{x} \rangle)^2 + \frac{1}{M_y - 1} \sum_{m=1}^{M_y} (\bar{y}_m - \langle \bar{y} \rangle)^2 \right]^{1/2}, \quad (1)$$

where the overbar denotes the time average over a three-month season, the angle brackets denote the ensemble average,  $M_x$  and  $M_y$  denote the number of ensemble realizations of a CRCM5 variable in the control ( $x$ ) and a modified model version ( $y$ ), respectively, and are given in Table 1. This specific measure of noise is employed to stress the fact that the ensemble variance of the difference between the two model versions is equal to the sum of the variances of the control and the modified model ensembles.

The noise measured with the standard deviation (Eq. 1) is displayed in Fig. 3 for the three single-year sets (SYNA, SYSN and SYDS) as a function of the parameter perturbation, season and CRCM5 variable. It is computed for differences between the members of the control ( $x$ ) and a modified M10 version ( $y$ ); similar patterns are obtained when M01 is used instead of M10 (not shown). Note that the same colour bar is used for precipitation and temperature. In winter the noise in precipitation in the SYNA set (Fig. 3a) is rather low in absolute terms, with values up to 0.3 mm/day over the southeastern portion of the continent and up to 0.7 mm/day off the East Coast of North America. However, these values are considerable in relative terms because the precipitation rates in winter are generally low, especially over the continent. The spectral nudging and reduced domain size (Fig. 3b, c) help to reduce noise level for precipitation in winter to fairly low values. The patterns of the 2 m-temperature in winter (Fig. 3d–f) are similar to precipitation; noise locally attains 0.6°C over the northern Canada in the SYNA set and is almost entirely suppressed in the SYDS set. However, in summer, the standard deviation of the differences between the control and modified model versions attains striking values in the SYNA set. For precipitation (Fig. 3g) it locally attains 2.5 mm/day over the southern and eastern coastal regions of the continent. Spectral nudging (Fig. 3h) is not very efficient in reducing noise. The domain size reduction (Fig. 3i) reduces noise but locally it is still up to 0.6 mm/day. As of 2 m-temperature in summer (Fig. 3j–l), noise levels are barely higher than 1°C. Spectral nudging suppresses the noise below 0.6°C and the reduction of domain size below 0.2°C.

The above considerations emphasize the need for ensemble integrations when studying RCM response to modification using single-year simulations. It is not likely that any reasonable modification performed on the state-of-the-art RCMs would produce larger differences in summer

precipitation than the values of the noise-induced standard deviation of the differences displayed in Fig. 3g. This implies a relative error of 100% in the estimates of the CRCM5 sensitivity to the parameter perturbations obtained without ensemble integrations. Time averaging over a season is not sufficient to ensure filtering of internal variability noise, and averaging over an ensemble or a longer period is required to assess the signal.

### 3.3 Signal P10 in winter

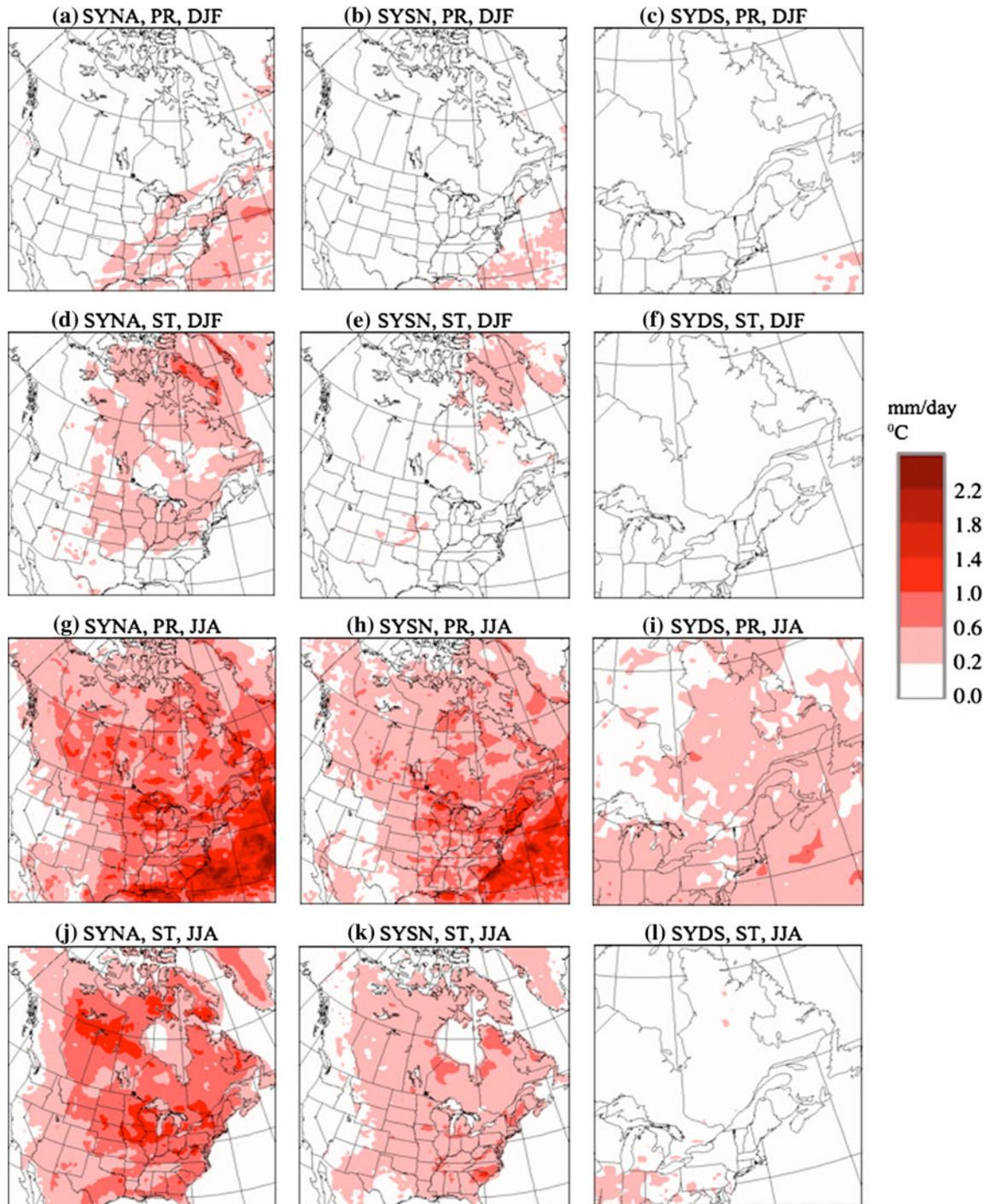
In this subsection we examine the change in seasonal averages due to the perturbation in the large-scale condensation parameter P10 (Table 1). As before, we denote the CRCM5 variable obtained in an individual simulation in the control model ensemble M00 with  $x$  and the same variable in the modified model ensemble M10 with  $y$ . The change in the CRCM5 seasonal averages due to the perturbation P10 is quantified by the difference of time-average ensemble averages of  $y$  and  $x$ ; the difference is computed in each simulation setup (SYNA, SYSN and SYDS) and will be referred to as the signal. Because of the internal variability in seasonal averages, especially in summer, and the relatively small number of available ensemble members for the two modified model versions M10 and M01, the ensemble averages are also prone to the noise-induced sampling error. In order to avoid erroneous interpretation of internal variability residuals in the ensemble averages as the model sensitivity to the parameter perturbations, statistical significance of the responses is also evaluated using the test for differences of means (Von Storch and Zwiers 1999). For the purpose of testing, the true ensemble variances of the control ( $x$ ) and modified model version ( $y$ ) are assumed to be equal, as we believe that the differences between these variances in model versions considered here are reasonably small with respect to the sampling error of their estimates. Under this assumption, the test statistic for the null hypothesis of no difference between the two model versions, is given as

$$t = \frac{\langle \bar{y} \rangle - \langle \bar{x} \rangle}{\sqrt{(1/M_x + 1/M_y)S_w^2}}, \quad (2)$$

where the overbar denotes seasonal average, the angle brackets ensemble average and  $M_x$  ( $M_y$ ) are the ensemble sizes corresponding to  $x$  and  $y$ . The quantity

$$S_w^2 = \frac{\sum_{m=1}^{M_y} (\bar{y}_m - \langle \bar{y} \rangle)^2 + \sum_{m=1}^{M_x} (\bar{x}_m - \langle \bar{x} \rangle)^2}{M_x + M_y - 2} \quad (3)$$

is the pooled estimation of the ensemble variances of the control and modified model version. Here,  $M_x = 10$  and  $M_y = 5$ , as shown in Table 1. The ensemble size of the control version is doubled in order to increase the signal to



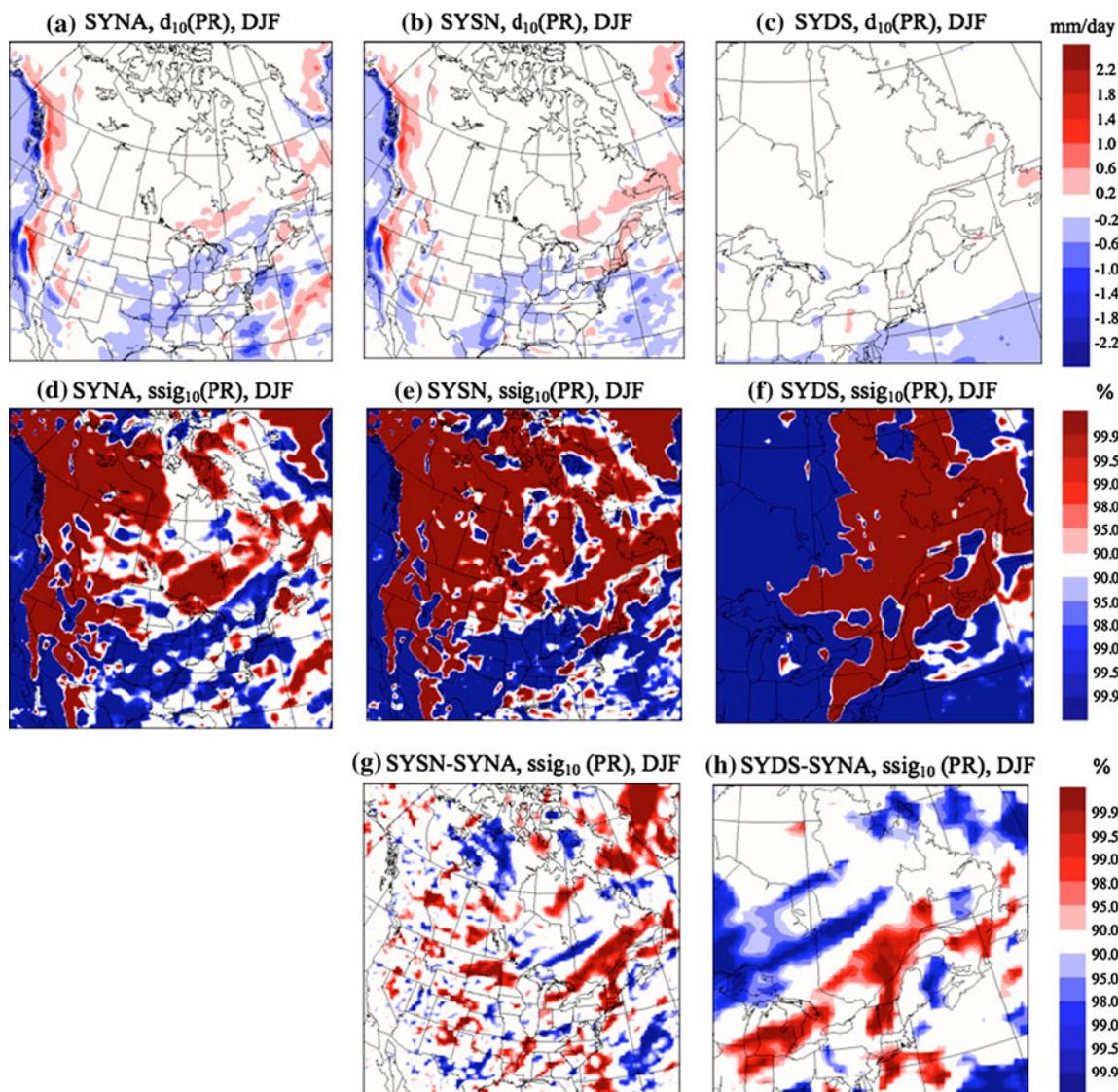
**Fig. 3** Sample standard deviation (Eq. 1) of the sensitivity of the CRCM5 seasonal average to parameter perturbation P10 (Table 1). The sensitivities are measured as the differences between members of the perturbed-parameter model M10 and members of the control

model M00 ensembles, as a function of experimental setup, variable and season: (a, d, g, j) SYNA (b, e, h, k) SYSN, (c, f, i, l) SYDS, (a, b, c, g, h, i) seasonal precipitation, (d, e, f, j, k, l) 2 m-temperature, (a–f) DJF, (g–l) JJA

noise ratio and to estimate well the ensemble variance for at least one model version. Appendix A provides a discussion on how to select the number of ensemble realizations  $M_x$  and  $M_y$  in order to optimize the signal-to-noise ratio (Eq. 2). Under the null hypothesis of equal means of

the two model versions,  $t$  follows the Student's distribution with  $f = M_x + M_y - 2$  degrees of freedom ( $f = 13$  here).

The model response to the perturbation of the large-scale condensation parameter P10 (signal), as estimated by the difference of the ensemble means of the model versions



**Fig. 4** Difference of the ensemble mean winter-average (DJF) precipitation (signal) due to the perturbation P10 (Table 1) in a SYNA, b SYSN and c SYDS experiments and statistical

significance of the responses (d, e, and f, respectively); statistical significance of the difference of the signals g between SYSN and SYNA and h between SYDS and SYNA experiments

M10 and M00, is presented in Fig. 4a–c for winter-average (DJF) precipitation in the SYNA, SYSN and SYDS experimental sets, respectively. The corresponding fields of statistical significance are shown in Fig. 4d–f. The regions of high significance (above the 90% level), corresponding to the positive (negative) values of the signal, are coloured red (blue). In the SYNA set (Fig. 4a) the strongest and also highly significant (Fig. 4d) signal is aligned with the entire Pacific Coast. It reaches locally up to  $\pm 2$  mm/day. The signal is negative over the eastern Pacific Ocean off the West Coast and more precipitation is brought inland over the Rocky Mountains region by the westerly flow that dominates this area in winter. The imposed perturbation implies that the time scale for conversion of cloud to

precipitable water in the parameterization of the large-scale (stratiform) condensation in the version M10 is longer than in the reference version M00. It is worth noting that this perturbation is independent of the parameterization of deep convection in CRCM5 and thus should have no direct effect on convective precipitation, although indirect effects are possible. Another noticeable feature in the SYNA set (Fig. 4a) is a mainly negative signal over the southeast portion of the domain, significant at 95% level. Also note that in several regions in Fig. 4d over the central part of the continent the signal is highly significant, but its magnitude is too low to make a fingerprint with the contour interval used in Fig. 4a. This illustrates the fact that statistical significance does not imply a physically relevant signal.

When the spectral nudging is applied (Fig. 4b,e) the statistical significance of the winter precipitation signal P10 is noticeably enhanced over the entire domain; it remains low only in the areas where the signal changes sign. The signal in the SYSN simulation is almost identical to that in SYNA over the west portion of the domain (Fig. 4a, b); these regions are closer to the inflow boundary and the SN is not likely to have a considerable impact on the large-scale dynamics. Some differences between Fig. 4a, b appear over the eastern portion of the domain. When the SYDS setup is considered (Fig. 4c, f) a further increase in significance occurs: an almost 100% significance level can be seen over the entire domain. However, there is no signal of a magnitude larger than 0.2 mm/day in the SYDS domain, unlike in the other two setups over these regions.

Now we examine whether the use of SN or reduced domain size can produce a significant change in the signal induced by the perturbation P10. Thus, we aim at finding physically and statistically significant differences between the signals in the SYSN (SYDS) displayed in Fig. 4b, c and the signal in the SYNA set shown in Fig. 4a. The fact that at a given location the signals in the SYNA (SYDS) and SYSN are statistically significant does not imply that their difference is also statistically significant. To quantify the statistical significance of the difference of the signals we again apply the test for differences of means, but this time on the difference between the signals in the SYSN (SYDS) and SYNA (see Appendix 2 for details). The resulting fields of statistical significance of the signal's differences are shown in Fig. 4g (for SYSN-SYNA) and Fig. 4h (SYDS-SYNA). The differences between the signals are not shown since they can be inferred from subtracting values from Fig. 4b, c from Fig. 4a. In Fig. 4g it can be seen that the SN yield statistically significant differences alterations of the signal at 90% level or higher only in small patchy areas; the exception is the north eastern part of the continent where the regions of significance occupy somewhat larger regions. The difference of the signals between the SYDS and SYNA sets (Fig. 4h) is similar to that between the SYSN and SYNA. From Fig. 4a, b it can be seen that the magnitudes of these alterations are not of large physical importance. It is also worth to note that even if the null hypothesis of no difference between the signals is true, it can be accidentally rejected. For the significance level of 90% the nominal rejection rate is 10% but larger rates are not unlikely; because of spatial correlation of the atmospheric variables, the nearby grid points tend to yield similar test results and the points that appear statistically significant only by chance can cluster, resulting in larger areas of apparent significance (Von Storch 1982; Livezey and Chen 1983).

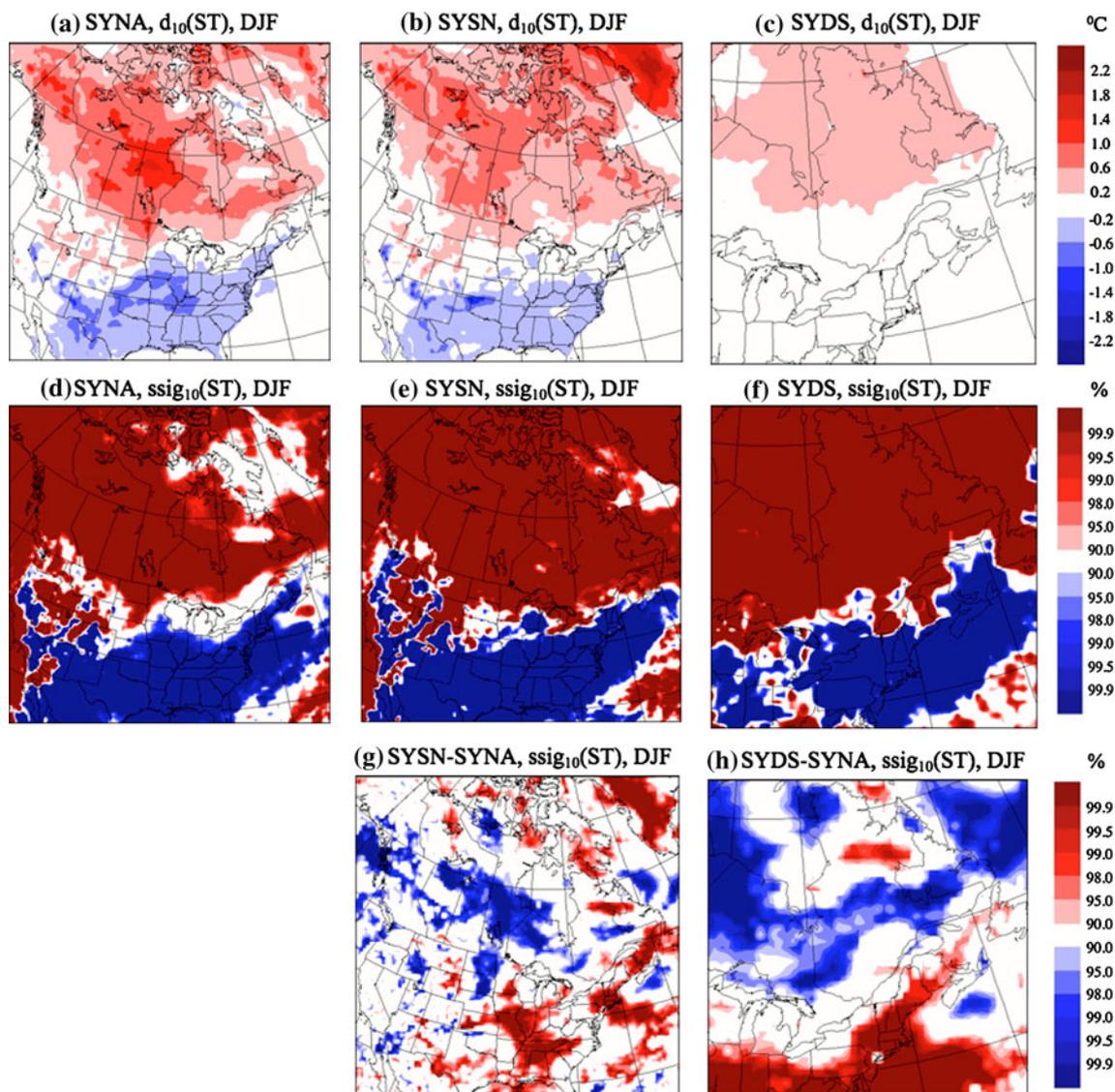
The same approach as above is adopted in order to analyze the winter-average 2 m-temperature response to

the perturbation of the large-scale condensation parameter P10 (Fig. 5a–c). It can be seen that the statistical significance levels for the signal are as high as 99% in all configurations (Fig. 5d–f). The signal in the SYNA set shows a dipole consisting of warming over the northern half of the domain and a slight cooling over the southern half, with magnitudes between  $-0.6$  and  $1.4^{\circ}\text{C}$ . The signal patterns in the SYSN and SYDS experiments are generally similar to those in the SYNA set, having, however, somewhat smaller magnitudes in the SYSN set. The test of the difference of the SYSN and SYNA signals (Fig. 5g) displays high local significance levels over the southern and north-central parts of the continent. In the small SYDS domain (Fig. 5c) the estimated magnitude of the signal is also somewhat reduced. Figure 5h shows that this decrease of the sensitivity with respect to the control run of the CRCM5 2-m winter temperatures to the perturbation P10 in the smaller domain is statistically significant at very high levels, especially near the southern boundary of the SYDS domain.

### 3.4 Signal P10 in summer

For summer (JJA) precipitation despite a physically relevant magnitude of the model response to the perturbation P10 over many regions, the response is generally statistically insignificant, which is the major difference with respect to the winter case. This happens because the noise is very large in summer precipitation (as shown in Fig. 3g–i) and strong signals are required for significance, given our ensemble size. Because of the lack of significance the analysis of the summer precipitation response to P10 will not be presented. It is worth reminding that the lack of statistical significance is always a function of sample size and hence a consequence of the small sample used here. The smaller the signal-to-noise ratio, the larger the sample needed to achieve significance.

The perturbation P10 produces a statistically significant response in the JJA 2 m-temperatures (see Fig. 6). A widespread cooling is notable over most of the continent, with magnitudes up to  $2.2^{\circ}\text{C}$  (Fig. 6a–c), and the signal is robust at significance levels higher than 95% over most parts of the domain for all the three configurations SYNA, SYSN and SYDS (Fig. 6d–f). In the SYSN and SYDS setups the rejection levels are almost 100% in the entire domain, which imply that only a few ensemble simulations might be required to adequately assess the temperature signal in these configurations. Over the ocean the 2 m-temperatures are strongly constrained by the imposed SST variations, so that the response to parameter perturbations is small. It is worth noting that the model displays a considerable increase in the cloud cover and relative humidity

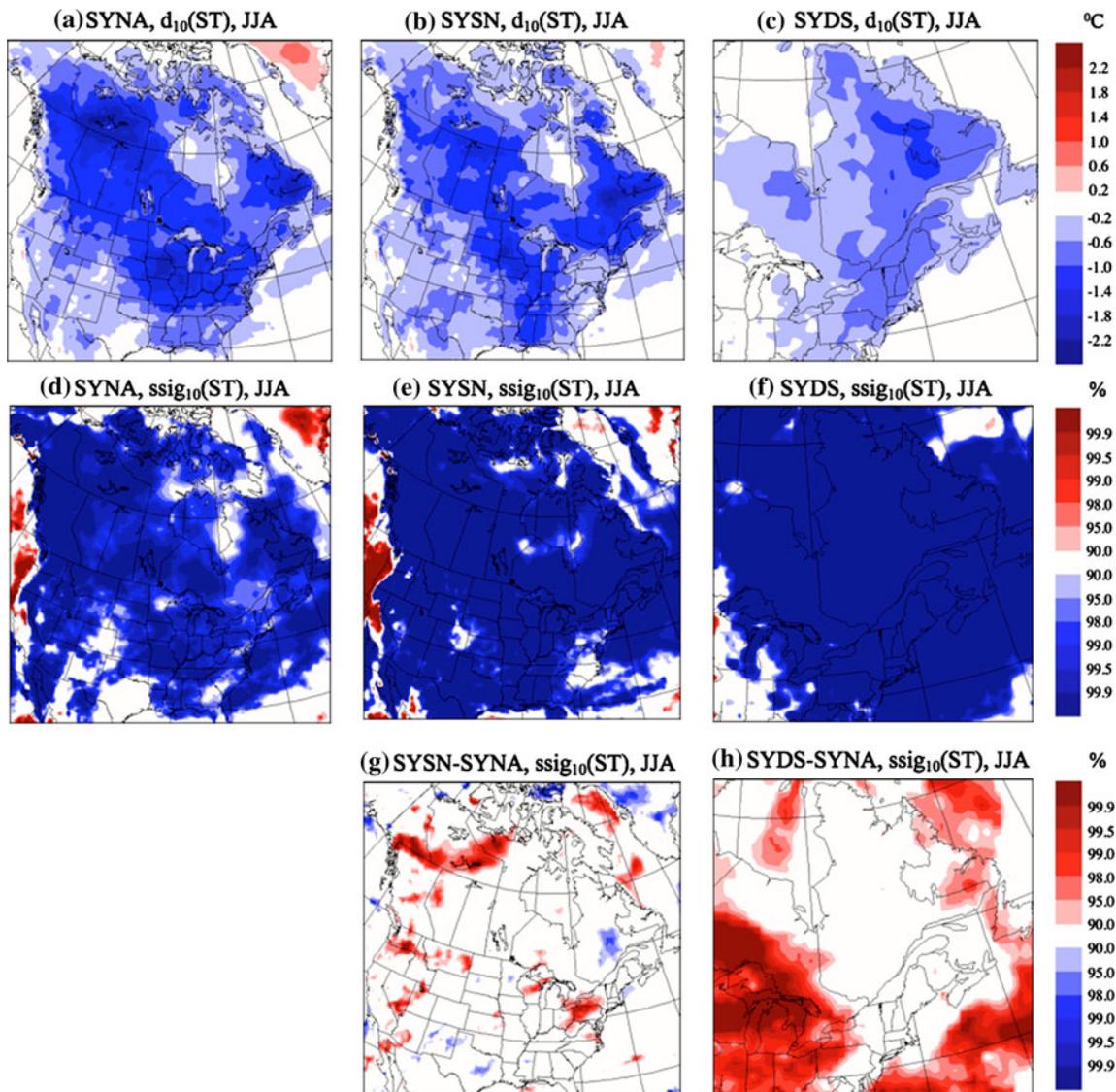


**Fig. 5** Same as in Fig. 4 but for winter 2 m-temperature (DJF)

at altitudes below 500 hPa (not shown). Increase in low clouds might reduce the solar heating at the ground in summer resulting in cooling but also might reduce the IR emission over high latitudes in winter resulting in warming, as in Fig. 5a, b. Further, despite that the temperature signal has a smaller magnitude in the SYSN experiment than in SYNA (the area in which the cooling is stronger than  $-1^{\circ}C$  occupies more than a half of the continent in the SYNA, extending from the Pacific to the Atlantic coast unlike in SYSN), this difference is not significant when tested in Fig. 6g. The absence of significant differences between the SYSN and SYNA signal does not mean that there is no change but that it was small enough to go below our capability to detect it. In the SYDS set (Fig. 6c, f) the magnitude of the signal is

heavily reduced with respect to the SYNA experiment in the southwest portion of the SYDS domain, which might be an artefact of the proximity of the lateral boundaries. Figure 6h shows that this alteration of the signal in the SYDS domain is statistically significant. These results seem to favour the use of SN as a viable tool to study parameter perturbation.

We proceed to examine the model's response to the perturbation of the threshold parameter for the onset of deep convection (P01 in Table 1). In winter, deep convection activity is at its minimum and is likely absent in higher latitudes of the domain. For this reason, the perturbation P01 produces almost no significant signal in winter (Fig. 2). Hence, for this perturbation, we focus on the summer months.



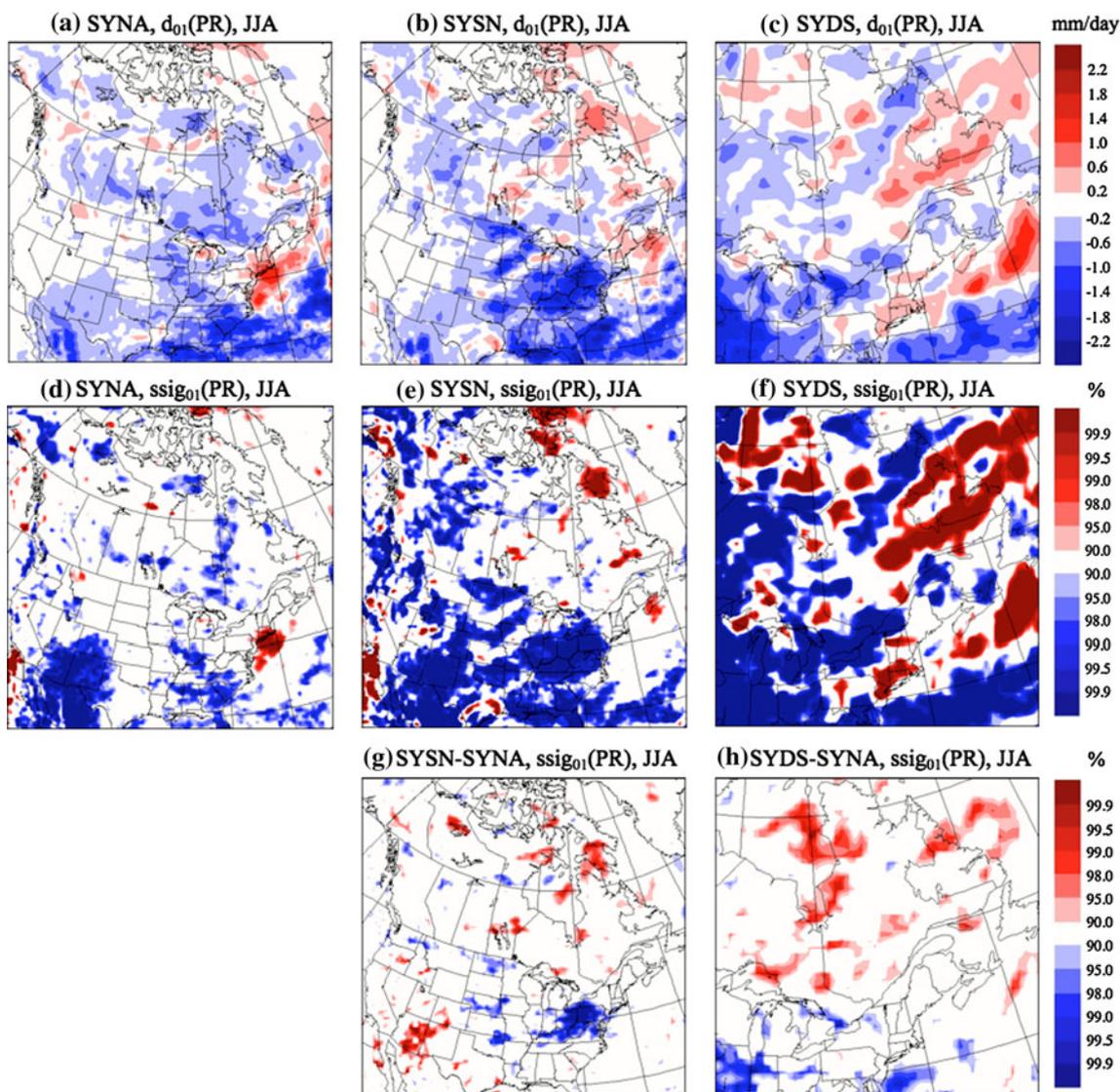
**Fig. 6** Same as in Fig. 4 but for summer 2 m-temperature (JJA)

### 3.5 Signal P01 in summer

Figure 7 displays the analysis of the difference between the summer (JJA) averages of the model versions M01 and M00, for precipitation in the SYNA (panels a and d), SYSN (b, e) and SYDS (c, f) experimental setups. Also shown is the statistical significance of the signals' difference SYSN-SYNA (g) and SYDS-SYNA (h). For precipitation in the SYNA set, the signal P01 is mainly negative, with magnitudes reaching 2 mm/day in the southeast part of the domain. However, the signal is in general not statistically significant, except in relatively small areas. The region where the signal is robust is the US southwest and northern Mexico, where the convective precipitation dominates the total precipitation. The signal is also significant in scattered

areas over the eastern half of the continent, a region with important convective precipitation in summer.

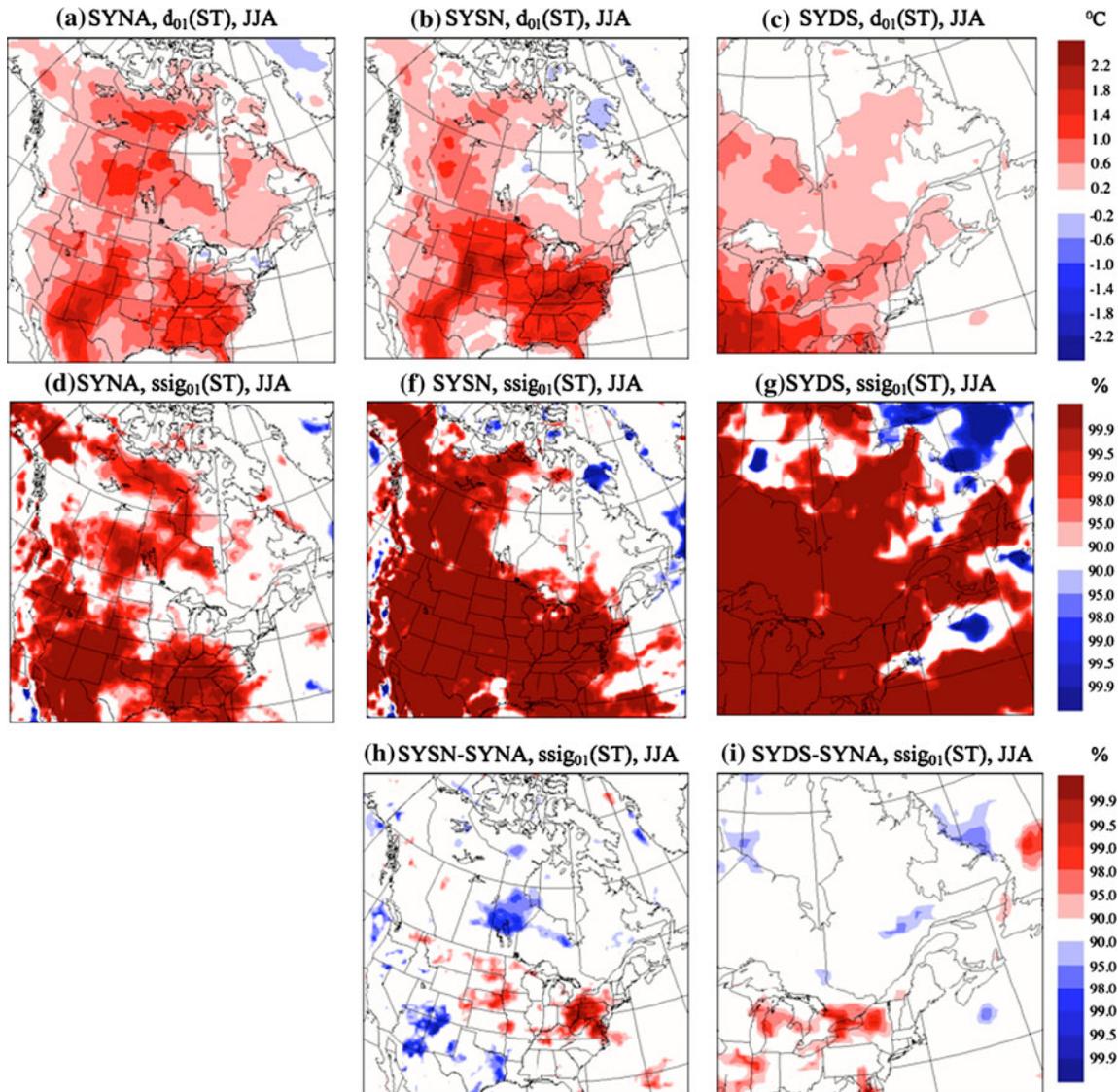
The results in the SYSN configuration show a substantial gain in statistical significance when SN is applied. The SYSN experiment reveals that the perturbation P01 mainly leads to a decrease in summer precipitation that varies from  $-0.2$  in the northwest to below  $-2.0$  mm/day in the southeast portion of the domain (Fig. 7b). The perturbation P01 also exhibits a strong effect on summer precipitation in the small SYDS domain (Fig. 7c); the model response is negative with values as small as  $-1.8$  mm/day south of the Great Lakes. Further, the signal in the SYDS set is quite similar to the SYSN case, with somewhat smaller magnitudes. In other parts of the small domain, such as over the province of Quebec and off the



**Fig. 7** Same as in Fig. 4 but for the signals induced by the perturbation P01 for summer precipitation (JJA)

East Coast, the signal is spatially variable, despite being highly statistically significant (Fig. 7e) and with considerable magnitudes of up to 1 mm/day. Since there are no remarkable topographic features in the small domain it can be argued that they are rather fingerprints of instantaneous weather patterns (storm tracks) that are not filtered out in 3-month averages because of insufficient sample of the instantaneous atmospheric states and small variability between the ensemble members. This points to the fact that in such a small temporal sample, the ensemble means of the control M00 and perturbed-parameter model M01 are dependent on the particular year. Figure 7g, h show that internal variability in summer is too large to permit the detection of the effect of the SN and domain size reduction on summer precipitation (if there is any) given the actual ensemble sizes.

To complete the analysis of the model response to the parameter perturbations we consider the differences of means for the CRCM summer 2 m-temperatures induced by the perturbation P01, displayed in Fig. 8. The model response to the present parameter perturbation has the same sign in the three experiments over the entire domain. In the SYNA set the perturbation P01 produces a warming of 0.2–3.0°C over almost all land points. This warming signal is statistically significant over most of the southern half of the domain, while over the northern half either the signal has a small magnitude or the internal variability of the difference renders the signal difficult to estimate. In addition, the magnitude of the signals in the SYSN and SYDS sets do not appear to be reduced with respect to that in SYNA, unlike the case of the JJA temperature response to P10 already shown Fig. 6a, b. Figure 8g shows that the



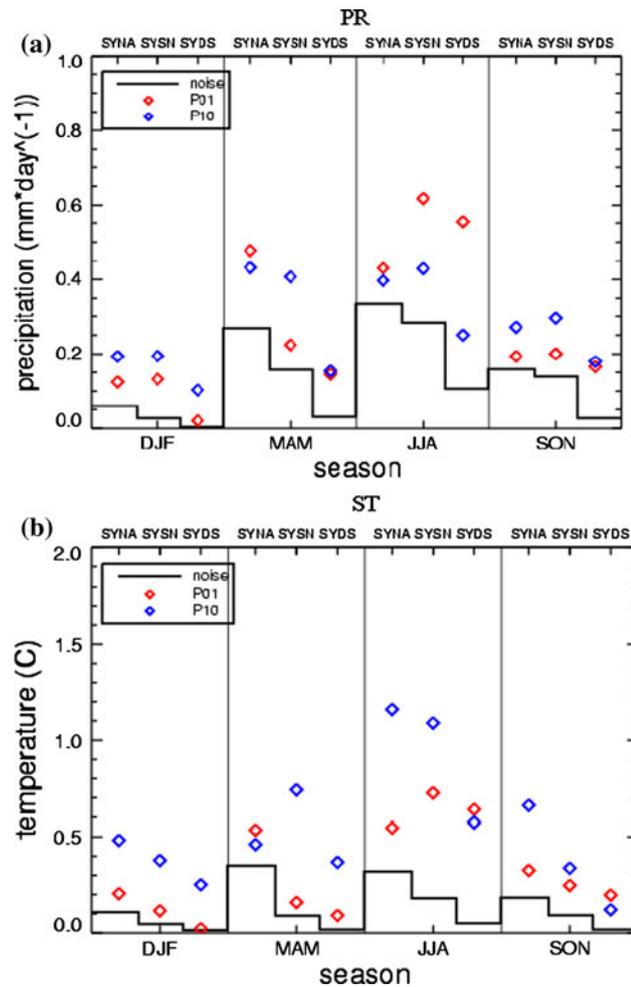
**Fig. 8** Same as in Fig. 4 but for the signals induced by the perturbation P01 for summer 2 m-temperature (JJA)

high statistical significance of the difference between responses is confined to a few rather small regions, which can be also result of chance. Further, when the statistical significance of the difference SYDS-SYNA is examined in Fig. 8h, the difference of means is mostly non-significant, and therefore, no evidence is found against the reduction of domain size in the signal P01.

### 3.6 Rule of thumb for the minimum ensemble size

The findings in the previous subsections as obtained from the analysis of the differences of means and their statistical significance are summarized in Fig. 9, for precipitation (a) and temperature (b). The plots in Fig. 9 represent the rms values of the signal and its standard deviation. They are obtained with the help of the test statistic for the

difference of means. Note that  $t$  in (Eq. 2) is in fact the signal-to-noise ratio. The numerator in (Eq. 2) is the signal estimated with the difference of ensemble means of the control and perturbed model version and the denominator represents the standard deviation of this estimate due to insufficient sample size. Figure 9 displays the square root spatially averaged (rms) values of these quantities. The rms values are computed only over an area common to all the experiments. The evaluation area consists of  $50^2$  grid points and corresponds to the central part of the small-domain SYDS simulations (Fig. 1), exclusive of the 10-point sponge zone. Only land points are accounted for in the computation of the rms. In Fig. 9 the red (blue) diamonds represent the rms difference of means triggered by the perturbation P01 (P10), i.e., the spatially averaged magnitude of the signal, as a function of season and



**Fig. 9** The rms signal (diamonds) and noise (step-like line) as a function of experimental setup and season for seasonal-average **a** precipitation and **b** 2 m-temperature. Signal is estimated as the rms difference of ensemble means of the perturbed-parameter (red) M01 and (blue) M10 model and control model M00 (Table 1). Noise is measured with the standard deviation of the difference of ensemble means

simulation configuration. The black step-like line in Fig. 9 represents the standard deviation of the difference of ensemble means. It is computed as the rms of the denominator in (Eq. 2).

It can be seen in Fig. 9 that the reduction of domain size (SYDS) is more efficient in reducing the noise level than the SYSN. It is worth reminding here that the SN parameters in the SYSN experiment were adjusted so that the SN forcing be rather weak and applied only in the upper levels. Alexandru et al. (2009) showed that a stronger nudging of large scales, applied at all levels, could substantially reduce internal variability noise. Whether this would change the magnitude of the signal cannot be inferred from the experiments considered here. The signal in the small SYDS domain has in most of the cases smaller magnitude than

those in the large domain experiments SYNA and SYSN. Exceptions such as for parameter P01 for summer precipitation (Fig. 9a), could happen due to the contamination of the SYNA signal with noise, since the noise can alter the estimate of the magnitude of the signal in both ways—decreasing and increasing it. Similarly, the smaller signal magnitudes in SYDS domain in Fig. 9 do not prove that the small domain suppresses the signal but rather indicate that this could sometimes be the case. On the other hand, the SN is fairly efficient in reducing noise, while there is not much evidence that the model response is smaller.

The calculations of the rms differences of means and noise levels can be used to derive a rule of thumb for the minimum ensemble sizes that need to be generated for the control and modified model versions in order to achieve, on average, a given level of statistical significance. For this purpose we define the effective signal-to-noise ratio as

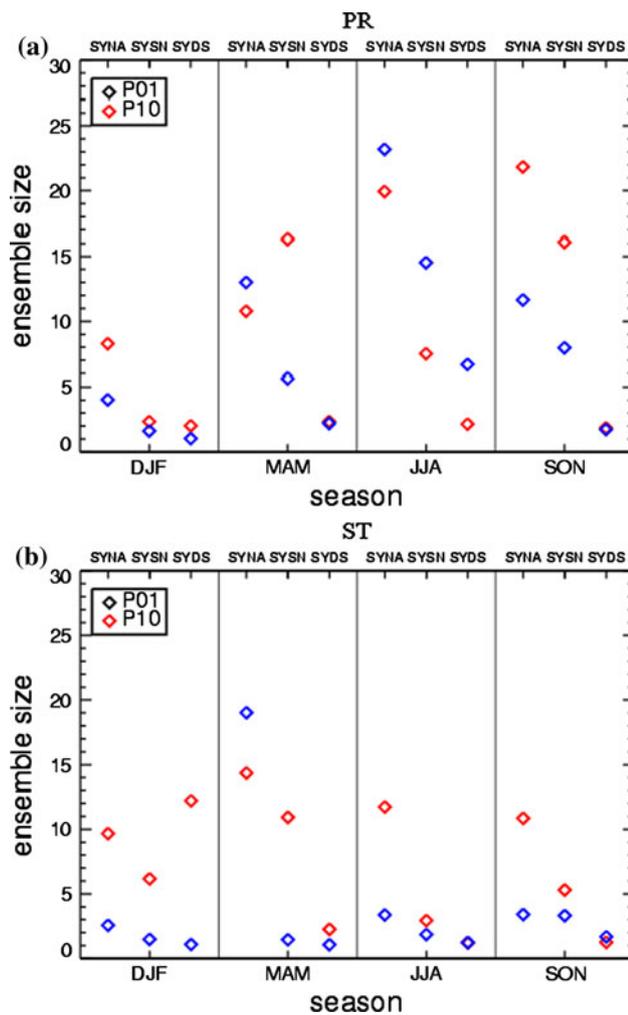
$$t_{\text{eff}} = \frac{\text{RMS}[\langle \bar{y} \rangle - \langle \bar{x} \rangle]}{\sqrt{\frac{2}{M}} \text{RMS}[S_w]}, \quad (4)$$

where an equal ensemble size  $M$  is assumed for the both control ( $x$ ) and perturbed-parameter model version ( $y$ ). The pooled variance  $S_w$  is as given in (Eq. 3). For a significance level of 95% the  $t$ -statistic (Eq. 8) is required to be larger than  $t_0 = 1.96$  for the two-sided test and for infinite number of degrees of freedom. The latter is correct for very large ensemble sizes. Substituting 1.96 for  $t_{\text{eff}}$  in (Eq. 4) and solving for  $M$  gives the proposed rule of thumb for the minimum ensemble size as follows:

$$M_{\text{min}} = 2 \times 1.96^2 \times \left\{ \frac{\text{RMS}[S_w]}{\text{RMS}[\langle \bar{y} \rangle - \langle \bar{x} \rangle]} \right\}^2, \quad (5)$$

Note that due to the properties of the Student's distribution, if a small number of degrees of freedom was assumed instead of infinite number, the required critical value  $t_0$  that corresponds to the 95% significance would be larger, resulting in a more conservative (higher) demand for  $M_{\text{min}}$ . Due to some vagueness of the concept we rather intend to use  $M_{\text{min}}$  in relative terms, to compare the required sizes among different perturbations, simulation setups and seasons, than to recommend it in absolute terms for achieving specified significance levels.

The rule of thumb for the minimal ensemble size is displayed in Fig. 10 for seasonal precipitation (a) and 2 m-temperature (b), for the perturbations of the deep-convection (P01) and stratiform condensation (P10) parameters, as a function of season and simulation configuration. In winter (DJF) the computational cost of providing significant estimates for the model response to P10 (blue diamonds) is fairly low for both precipitation and 2 m-temperatures in all configurations, as the signal is non-negligible and the noise level is at its minimum. However, the same does not



**Fig. 10** Minimal number of ensemble members needed to achieve significant estimates at 95% level for the signals induced by the perturbations (red) P01 and (blue) P10, as a function of season and experimental setup, as derived from the rule of thumb in Eq. (5)

hold for the response to P01. This perturbation produces very little response of the model in winter, especially for temperature; in order to provide significant estimates relatively large ensembles of 10 members or more would be necessary, despite that internal variability is very small. On the other hand, to find out that the signal P01 is small it is sufficient to estimate the maximum rmsd between two simulations that differ in initial conditions (this can be done from the control ensemble M00) and to generate a single simulation of the modified model M01. Then the rmsd between this simulation conducted with M01 and a randomly chosen member of the control ensemble M00 would indicate that the signal is small. This can be clearly seen in Fig. 2b in winter: all rmsd between pairs of individual members formed from M01 to M00 are below or at the maximum noise level, indicating that the signal is negligible with respect to internal variability.

In summer (JJA) the ensemble sizes required for the significant estimates of seasonal precipitation signals (Fig. 10a) are much larger. In the large domain with no SN (SYNA) the minimum number of members is about 25 for the perturbation P10 and 20 for P01, despite the latter exciting locally high sensitivities (see Fig. 7a). The spectral nudging (SYSN) almost halves the number of ensemble members needed to achieve statistical significance, while the reduction of domain size (SYDS) reduces the minimum number of members almost 5 times. Both methods of noise reduction appear to be very efficient for precipitation in summer. When summer 2-m temperatures are considered (Fig. 10b) the SYSN and SYDS configuration are still efficient in reducing the minimum ensemble sizes but appear less sensitive to reduction of noise. This is due to the fact that signals in the SYNA configuration in summer temperatures are relatively strong (see Figs. 6a, 8a); so in that case the need for ensemble calculations is low in all the three configurations, as compared to the case of precipitation. In fact, in the case of 2 m-temperature the season that is associated with the largest computational cost of significant estimates is spring when the minimum ensemble sizes are 20 (Eq. 15) for the response to P10 (P01), respectively. Also the noise level in the SYNA setup in spring is slightly higher than in autumn.

#### 4 Summary and conclusions

Development of RCMs and study of uncertainty related to the choices that must be made in constructing and applying RCMs often requires multiple testing of model response to a large number of modifications, which imposes a high demand on computational resources. A high-resolution RCM simulation configuration, less computationally demanding than the operational RCM runs (in terms of the integration period, computational domain and internal variability noise), if used as test bed for RCM modification, would allow the allocation of the computational resources to testing a larger number of modifications. The objective of this work was to study the model response to RCM parameter perturbations using computationally less demanding configurations than the operational runs and eventually select an optimal configuration as a result of the trade-off between the representativeness of results it may provide and its computational cost. The approach followed consisted of analysing sets of RCM simulations conducted for the three parameters' settings, here referred to as the model versions: the control (unperturbed) model version and two modified versions in which two parameters that control deep convection and stratiform precipitation, respectively, were perturbed one at a time. These three model versions were used to generate RCM simulations

within three setups, all with the integration period of a single year.

In the first setup, denoted as SYNA, we performed ensemble simulations with perturbed initial conditions over a large continental-scale domain with spectral nudging turned off. The parameter perturbations produced fairly large differences of ensemble means in 2 m-temperature, especially in summer. These differences were statistically significant in a large part of the domain. On the other hand, for precipitation the results in all seasons in the largest part of the domain were statistically insignificant, with exception of the topographically rich regions along the West Coast of North America.

In order to reduce internal variability noise—a nuisance at the time of quantifying the signal—, we performed perturbed parameter RCM simulations using two additional setups: (1) SYSN in which we used the same domain and number of ensemble members as in the previous two configurations but applied a weak spectral nudging (SN) at upper levels, and (2) SYDS in which the domain size is reduced. The main concern with these two configurations was that they might alter or even suppress the model sensitivity to parameter perturbations along with reduction of internal variability. However, the results of these two experiments when compared to the SYNA configuration showed that this concern was only justifiable in the case of a reduced domain. Not surprisingly, in the case of the large-scale condensation parameter perturbation, the SYDS signal exhibited deviations of considerable magnitude from its counterpart in the SYNA set that is taken as reference here. These changes were statistically significant over larger areas near the inflow lateral boundaries. The use of the very small domains, such as SYDS, is known to be associated to several flaws, which was discussed in the Introduction section. The alteration of the responses to perturbations by the proximity of the lateral boundaries, noted in the SYDS, is in accord with the previous evidence. The SYDS domain may, however, be attractive for conducting fast and computationally inexpensive RCM sensitivity tests at the development stage of the model. The reduction of the computational cost when using the small SYDS domain is twofold: the integration area is much smaller (and hence computational cost) and the internal variability is low (hence potentially contributing to increasing statistical significance or reducing the need of large ensembles).

The model response to parameter perturbations in the SYNA and SYSN configurations was rather similar in pattern as well as in magnitude, and statistically significant only in rather small, scattered areas (which could be also a result of internal variability in case the null hypothesis of equal responses is true). Results did not provide evidence that the spectral nudging altered the mean model response

to parameter perturbations. However, this should not be understood as a proof of SN not affecting the signal but rather as a consequence of the fact that the number of ensemble members was insufficient to identify the differences. In addition, the SN configuration used here was designed to minimally force the large-scale flow and this only at upper levels. It is not known to the authors whether a stronger SN (that would better constrain internal variability deviations) would still exhibit little or no effects on the signal, as it is the case with the SN configuration used here.

**Acknowledgments** The authors want to thank Ms Katja Winger for maintaining the CRCM5 at UQAM and for her technical help in performing simulations and the subsequent analysis of results. We thank Mr Mourad Labassi for maintaining a local computing environment at the Ouranos Consortium and Mr Alejandro Di Luca for his valuable comments. We also want to thank the two anonymous reviewers for their constructive suggestions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix 1: Optimization of sample sizes for the test of the difference of means

In this appendix we assume the specific situation when limited resources are available for sampling while, at the same time, a large family of random variables need to be sampled, each variable in a separated experiment, and compared to some control variable using the test of the difference of means. This corresponds to a situation when a control climate model is compared with a large number of perturbed model versions. The objective of this appendix is to optimize the allocation of sample lengths between the control variable and those that are to be tested.

For this purpose we assume a control random variable  $x$  and a family of random variables  $\{y_k\}_{k=1}^K$ . We assume that  $x$  will be sampled  $M_x$  times while each member of the family  $\{y_k\}_{k=1}^K$  will be sampled an equal number of  $M_y$  times. The total sample length is then  $L = M_x + KM_y$ , and it needs to fit some non-negotiable constant imposed by the resources, given as the maximum sample length. Under these assumptions increasing  $M_y$  by one would increase the total sample length  $L$  by  $K$ , while increasing the control sample size  $M_x$  by one would increase the total sample size  $L$  also by one. This shows that  $M_y$  needs to be decided straightforwardly from the maximum allowed value of  $L$ . Once  $M_y$  is decided such that it leaves some space for the control size  $M_x$  to fit within the maximum total length  $L$ , which value of  $M_x$  will optimize the statistical significance

of the estimates of the differences of means between  $y_k$  and  $x$ ?

For the sake of simplicity we assume that the variables  $y_k$  and  $x$  are normally distributed and that their true variances are known and equal to some constant  $S^2$ . Then the test statistic for the test of the difference of means is given as

$$z_k = \frac{\langle x \rangle - \langle y_k \rangle}{S \sqrt{\frac{1}{M_x} + \frac{1}{M_y}}} \tag{6}$$

Under the null hypothesis that the two variables have equal true means follows the normal distribution  $N(0,1)$ . Upon defining the ratio of sample sizes of the control and tested variables as

$$b = M_x/M_y, \tag{7}$$

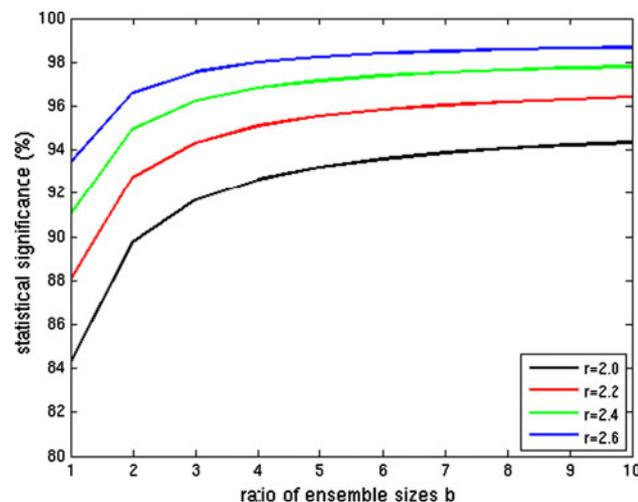
Equation (6) can be expressed as

$$z_k = r_k \left( \frac{b}{b+1} \right)^{1/2}, \tag{8}$$

where

$$r_k \equiv \frac{\langle x \rangle - \langle y_k \rangle}{S/\sqrt{M_y}} \tag{9}$$

is the signal-to-noise ratio expressed independently of the control sample size  $M_x$ . Fig. 11 displays the plots of the probabilities (statistical significances) that correspond to the critical values of the normally distributed test statistics  $z_k$  for the two-sided test, as a function of the sample-size ratio  $b$  and signal-to-noise ratios  $r_k$ . Values of  $b$  are given



**Fig. 11** Statistical significance derived from the two-sided test of the difference of means of two samples of unequal sample sizes; the size of the first sample is kept constant and the size of the second is increased by the factor  $b$  (abscissa), as in Eq. (7). Plots are drawn for selected signal-to-noise ratios (Eq. 9)

on the abscissa and each plot represents a given signal-to-noise ratio. The plots show that the largest increase in the statistical significance occurs when  $b$  is increased from 1 to 2 (that is, when the control sample is twice as large as the samples of the tested variables) and somewhat less from 2 to 3 (when it is three times larger). Further investment of the resources in the control sample size would result in no considerable gain in significance, since for values of  $b$  larger than 3 all curves quickly saturate, implying that an optimal value of  $b$  is to be chosen from the interval  $[1,3]$ . In addition, the jump in statistical significance between  $b = 1$  and  $b = 3$  is larger for lower signal-to-noise ratios.

### Appendix 2: Test for the differences of signals

In this appendix we describe the estimation of the difference between the signals produced with a single parameter perturbation in two separate simulation setups, namely SYSN or SYDS and SYNA. In every setup ensemble simulations are generated for the control model version M00 and the perturbed-parameter model (M01, M10) by varying initial conditions. Let us denote the model variable sampled with the control model ensemble with  $x$  and the same variable in the perturbed-parameter model ensemble with  $y$ . The signal is defined as

$$\delta = \langle \bar{y} \rangle - \langle \bar{x} \rangle, \tag{10}$$

where the overbar denotes seasonal average and the angle brackets ensemble average. The number of ensemble members of the control (perturbed) model version is  $M_x = 10$  ( $M_y = 5$ ) for all setups, as given in Table 1. It is assumed that  $\bar{x}$  and  $\bar{y}$  are independent and have equal variances. Under these assumptions it can be shown that the variance of  $\delta$  can be estimated as

$$S^2 = \frac{1/M_x + 1/M_y}{f} \left[ \sum_{m=1}^{M_y} (\bar{y}_m - \langle \bar{y} \rangle)^2 + \sum_{m=1}^{M_x} (\bar{x}_m - \langle \bar{x} \rangle)^2 \right], \tag{11}$$

where  $f$  denotes the degrees of freedom and is given as

$$f = M_x + M_y - 2. \tag{12}$$

Now we turn our attention to the difference between the signal  $\delta_2$  obtained in the SYSN (or SYDS) and  $\delta_1$  in SYNA setup. The assumption that the variances of  $\delta_1$  and  $\delta_2$  are equal is not suitable; the SN and smaller domain size can considerably reduce the internal variability in RCM simulations (e.g., Weisse and Feser 2003; Alexandru et al. 2009). This is also implied by results displayed in Fig. 3. Thus, the test for the difference of means of variables with unequal variances (Von Storch and Zwiers 1999) has to be used to test the difference of the signals

obtained in two configurations. The test statistic can be written as follows:

$$t' = \frac{\delta_2 - \delta_1}{\sqrt{S_1^2 + S_2^2}}, \quad (13)$$

where  $S_i^2$  is the estimator of the variance of  $\delta_i$  in the setup  $i = 1, 2$ , as defined in (Eq. 11). Because of unequal variances, the variable  $t'$  does not have  $t$  distribution under the null hypothesis that  $\delta_1$  and  $\delta_2$  have equal means. To solve this problem we employ the Welch's approximate solution (Scheffé 1970) that consists in approximating  $t'$  with a  $t$  distribution whose degrees of freedom  $F$  are estimated from the data as follows:

$$F = \frac{(S_1^2 + S_2^2)^2}{S_1^4/f_1 + S_2^4/f_2}. \quad (14)$$

The statistical significance is estimated from the local value  $t_0$  as the probability that the absolute value smaller than  $|t_0|$  would be obtained under the null hypothesis of no differences. Since the difference of means has no preferred sign, the “two-tailed” test is used. The statistical significance is computed using the cumulative distribution function of  $t$  distribution as:

$$\Pr(|t| \leq t_0) = 1 - I_{\frac{F}{t_0^2 + F}}\left(\frac{F}{2}, \frac{1}{2}\right), \quad (15)$$

where  $I$  is the regularized incomplete beta function (Press et al. 1992).

## References

- Alexandru A, de Elia R, Laprise R (2007) Internal variability in regional climate downscaling at seasonal scale. *Mon Weather Rev* 135:3221–3238
- Alexandru A, de Elia R, Laprise R, Separovic L, Biner S (2009) Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Mon Weather Rev* 137:1666–1686
- Arakawa A, Lamb VR (1977) Computational design of the basic dynamical processes of the UCLA general circulation model. In: Chang J (ed) *Methods in computational physics*. Academic Press, London, pp 174–267
- Biner S, Caya D, Laprise R, Spacek L (2000) Nesting of RCMs by imposing large scales. *Research activities in Atmospheric and Oceanic Modelling, WMO/TD 987, Report 30:7.3–7.4*
- Caya D, Biner S (2004) Internal variability of RCM simulations over an annual cycle. *Clim Dyn* 22:33–46
- Colin J, Dequé M, Radu R, Somot S (2010) Sensitivity study of heavy precipitation in limited area climate simulations: the influence of the size of the domain and the use of spectral nudging technique. *Tellus A* 62:592–604
- Côté J, Gravel S, Méthot A, Patoine A, Roch M, Staniforth A (1998) The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation. *Mon Weather Rev* 126:1373–1395
- Davies HC (1976) A lateral boundary formulation for multilevel prediction models. *Q J R Meteor Soc* 102:405–418
- De Elia R, Caya D, Côté H, Frigon A, Biner S, Giguère M, Paquin D, Harvey R, Plummer D (2008) Evaluation of uncertainties in the CRCM-simulated North American climate. *Clim Dyn* 30:113–132
- Fiorino M (2004) AMIP II sea surface temperature and sea ice concentration observations. <http://www-pcmdi.llnl.gov/projects/amip/AMIP2EXPDSN/BCS/amip2bcs.php>
- Giorgi F, Bi X (2000) A study of internal variability of a regional climate model. *J Geophys Res* 105:29503–29521
- Jones RG, Murphy JM, Noguer M (1995) Simulation of climate change over Europe using a nested regional-climate model. Part I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Q J R Meteor Soc* 121:1413–1449
- Kain JS, Fritsch JM (1990) A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J Atmos Sci* 47:2784–2802
- Laprise R (1992) The Euler equations of motion with hydrostatic pressure as independent variable. *Mon Weather Rev* 120:197–207
- Laprise R, de Elia R, Caya D, Biner S, Lucas-Picher Ph, Diaconescu E, Leduc M, Alexandru A, Separovic L (2008) Challenging some tenets of regional climate modelling. *Meteor Atmos Phys* 100:3–22
- Leduc M, Laprise R (2008) Regional climate model sensitivity to domain size. *Clim Dyn* 32:833–854
- Livezey RE, Chen WY (1983) Statistical field significance and its determination by Monte Carlo techniques. *Mon Weather Rev* 111:46–59
- Lucas-Picher Ph, Caya D, de Elia R, Laprise R (2008a) Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain. *Clim Dyn* 31:927–940
- Lucas-Picher Ph, Caya D, Biner S, Laprise R (2008b) Quantification of the lateral boundary forcing of a regional climate model using an aging tracer. *Mon Weather Rev* 136:4980–4996
- Nikiema O, Laprise R (2010) Diagnostic budget study of the internal variability in ensemble simulations of the Canadian Regional Climate Model. *Clim Dyn*, doi:10.1007/s00382-010-0834-y
- Press WH et al (1992) *Numerical recipes in Fortran 77*. Cambridge University Press, Cambridge
- Pudykiewicz J, Benoit R, Mailhot J (1992) Inclusion and verification of a predictive cloud-water scheme in a regional numerical weather prediction model. *Mon Weather Rev* 120:612–626
- Scheffé H (1970) Practical solutions of the Behrens-Fisher problem. *J Am Stat Assoc* 65(332):1501–1508
- Separovic L, de Elia R, Laprise R (2008) Reproducible and irreproducible components in ensemble simulations with a regional climate model. *Mon Weather Rev* 136:4942–4961
- Seth A, Giorgi F (1998) The effects of domain choice on summer precipitation simulations and sensitivity in a regional climate model. *J Clim* 11:2698–2712
- Sexton D, Murphy J (2003) Design of coupled atmosphere/ocean mixed-layer model experiments for probabilistic prediction. Technical Report HCTN45, Hadley Centre
- Sundqvist H, Berge E, Kristjánsson JE (1989) Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon Weather Rev* 117:1641–1657
- Uppala SM, Kållberg PW, Simmons AJ et al (2005) The ERA-40 reanalysis. *Q J R Meteor Soc* 131:2961–3012
- Von Storch H (1982) A remark on Chervin-Schneider's Algorithm to test significance of climate experiments with GCM's. *J Atmos Sci* 39:187–189
- Von Storch H, Zwiers FW (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge
- Von Storch H, Langenberg H, Feser F (2000) A spectral nudging technique for dynamical downscaling purposes. *Mon Weather Rev* 128:3664–3673

- Weisse R, Feser F (2003) Evaluation of a method to reduce uncertainty in wind hindcasts performed with regional atmosphere models. *Coast Eng* 48:211–225
- Weisse R, Heyen H, Von Storch H (2000) Sensitivity of a regional atmospheric model to a sea state-dependent roughness and the need for ensemble calculations. *Mon Weather Rev* 128:3631–3642
- Zadra A, Caya D, Côté J, Dugas B, Jones C, Laprise R, Winger K, Caron LP (2008) The next Canadian regional climate model. *Phys Can* 64:74–83